Article

Haplotype-resolved, gap-free genome assemblies provide insights into the divergence between Asian and European pears

🔴 Check for updates

Manyi Sun [1,2,8], **Beibei Cao**[1,2,8], **Kui Li** [3,8], **Jiaming Li**[1,2,8], **Jun Liu**[4,8], **Cheng Xue** [5,8], **Kaidi Gu**[5], **Shaozhuo Xu**[1,2], **Yuanjun Li**[6], **Qingyu Li**[6], **Meina Qu**[6], **Mingyue Zhang** [5], **Runze Wang**[1,7], **Yueyuan Liu**[1,2], **Chenjie Yao**[1,2], **Hang He** [3] & **Jun Wu** [1,2] ✉

Pears (*Pyrus* spp.) are self-incompatible crops with broad genetic diversity. High heterozygosity and technical limitations result in gaps within reference genomes. Our study presents the telomere-to-telomere, haplotype-resolved genomes of a representative Asian pear 'Dangshansuli' (*Pyrus bretschneideri*) and a European pear 'Max Red Bartlett' (*Pyrus communis*). Haplotype-specific genes exhibited notable differences from biallelic genes regarding transposable content, methylation patterns and expression levels. Allele-specific expression analysis suggested that the dominance effect is vital in the formation of fruit quality of pears. Population analysis of 362 accessions revealed that interspecific introgression increased pear diversity. We constructed a graph-based genome and identified structural variations associated with agronomic traits. A 286-bp insertion in the promoter region, along with differential expression of *PyACS1*, was identified between Asian and European pears, which exhibit distinct fruit-softening characteristics. Further experiments demonstrated the role of *PyACS1* in fruit softening. Overall, this study provided insights into genetic variation and will facilitate pear improvement.

Many vital crops display self-incompatibility (SI); examples are potato[1], cabbage[2] and pear[3]. SI prevents inbreeding depression and maintains the genetic diversity of species, which can lead to high levels of genomic heterozygosity. The hybridization of different parents can produce offspring that show superior performance for several traits[4]. However, this heterozygosity increases the complexity of the genome assembly and most assembled genomes represent a consensus consisting of haplotype mixtures that restrict investigations into allelic variations that underlie important traits at the whole-genome level. In recent years, more haplotype-resolved plant genomes have been assembled from species such as apple[5], tea[6], lychee[7], potato[1], cassava[8] and sugarcane[9]. These assemblies have facilitated the identification of structural variations (SVs) and allele-specific expression patterns between haplotypes

and have contributed to understanding the biological characteristics of highly heterozygous species.

Pear is an SI diploid species and is one of the most economically important fruit species worldwide. Generally, pears can be divided into two main groups, Asian and European pears, which showed distinct phenotypic difference including fruit shape, sugar, acid, flavor and softening[10]. The first Asian pear reference genome from the *Pyrus bretschneideri* cultivar 'Dangshansuli' was released in 2013 (ref. 11), followed by the European pear *Pyrus communis* 'Bartlett' genomes in 2014 (ref. 12) and 2019 (ref. 13). In 2019, a phased diploid Asian pear genome was assembled using single-pollen cell sequencing technology[14]. Due to sequencing and assembly technology limitations, these earlier pear reference genomes were incomplete and lacked a

large amount of genomic information. Recently, the nearly complete telomere-to-telomere (T2T) genomes of 'Dangshansuli' and 'Conference' (*P. communis*) were released using a mosaic assembly strategy[15]. However, the mosaic genome assembly restricted analyses to variations between two haplotypes. In addition, population structure and domestication in pear have been extensively discussed recently[10,16]. Nevertheless, the complex evolutionary history and uncertain phylogeny, especially the pattern of reticulate evolution with close wild relatives, remain to be examined. In addition, the high-quality T2T genomes provide the opportunity to construct a graph-based genome, which is a powerful tool for SV genotyping using next-generation sequencing (NGS) data[17].

In this study, we constructed haplotype-resolved, T2T genome assemblies for the cultivars 'Dangshansuli' (DS) and 'Max Red Bartlett' (MRB; red bud mutation of 'Bartlett'), which are representative of Asian and European pears, respectively. The assemblies are improved with regard to genome completeness and accuracy and provide opportunities to analyze allelic and interspecific variations. Resequencing data for 362 pear accessions were collected to investigate the evolutionary and demographic history of these pear groups. We constructed a graph-based genome by combining the two high-quality genomes with the resequencing data. SVs exhibiting significant frequency variations between Asian and European pears have been identified. We further explored the association between SVs and fruit traits and between SVs and gene expression. One 286-bp insertion sequence in the promoter region of *PyACS1* may have contributed to changes in fruit maturity and softening in European pear fruit. The results of this study advance our understanding of heterosis, evolutionary history and critical genetic variation in the pear phenotype and will provide valuable resources for pear breeding.

Results

Haplotype-resolved T2T assemblies of pears

In view of the heterozygosity of *P. bretschneideri* DS (1.23%) and *P. communis* MRB (1.33%), two haplotypes of the preliminary diploid genomes were assembled using PacBio HiFi, ONT (Oxford Nanopore Technologies) ultra-long, and Hi-C data with hifiasm[18]. Then, Verkko[19] and HiCanu[20] assembly results were used to fill the gaps and telomeric regions (Supplementary Table 1 and Methods). All gap-closure regions were supported with uniform coverage of ONT and HiFi reads (Supplementary Figs. 1, 2 and 3). Finally, each haplotype was assembled into 17 contigs corresponding to the 17 pear chromosomes with genome sizes ranging from 497.50 Mb to 505.55 Mb and contig N50 sizes ≥28.96 Mb (Table 1). The sequence 18.89 Mb to 24.13 Mb was not assembled in released *P. bretschneideri* and *P. communis* genomes (Supplementary Fig. 4a). The contig N50 size, mapping rate, Benchmarking Universal Single-copy Orthologs (BUSCO) and Long Terminal Repeat (LTR) Assembly Index supported the high quality of the assemblies (Supplementary Note 1). We identified 34 telomeres on the terminal ends of the 17 chromosomes of the haplotypes in DS and 33 in MRB, using a telomere repeat sequence (TTTAGGG) as the query (Fig. 1a,b). In summary, we successfully assembled T2T haplotype-resolved genomes for DS and MRB.

The DS and MRB haplotypes contained 262.86–272.23 Mb of transposable elements (TEs) (Table 1) and LTRs comprised the most abundant category of TEs (Supplementary Table 5). A total of 42,675–43,602 protein-coding genes were predicted in the haplotypes of DS and MRB (Table 1), with the BUSCO completeness ranging from 97.90% to 98.60% (Supplementary Fig. 7). The high-quality assembly and annotation results provided an opportunity to investigate the complex regions, including centromeres and S-locus alleles (Fig. 1c,d and Supplementary Note 2). In addition, 577 (563) genes were identified in newly assembled regions of haplotype A (HapA) (haplotype B (HapB)) in DS and 697 (804) were in HapA (HapB) of MRB. These genes were predominantly (77.33–84.06%) involved in the regions with high TE content and high

Table 1 | Summary of genome assembly and annotation of DS and MRB

Assembly	DS		MRB	
	Haplotype A	Haplotype B	Haplotype A	Haplotype B
Total length (Mb)	497.50	499.72	503.33	505.55
Contig number	17	17	17	17
Contig length (Mb)	497.50	499.72	503.33	505.55
Largest contig (Mb)	40.62	41.82	42.54	43.02
Contig N50 (Mb)	28.96	29.26	29.74	29.58
Length of Ns	0	0	0	0
Number of telomeres	34	34	33	33
GC content of the genome (%)	37.38	37.39	37.45	37.45
Mapping rate (%)	99.52	99.60	99.54	99.65
Genome coverage (%)	99.99	99.99	99.99	99.99
BUSCO completeness of genome assembly	98.90	99.00	98.80	98.90
Quasi-variance value	54.21	54.21	53.91	53.98
Homology SNP ratio	$5.80×10^{-6}$	$4.52×10^{-6}$	$6.73×10^{-6}$	$5.33×10^{-6}$
LTR Assembly Index	23.58	21.38	21.63	24.31
Haplotype switch error ratio (%)	0.30		0.69	
Annotation				
Number of predicted protein-coding genes	43,330	43,602	42,784	42,675
Masked repeat sequence length (Mbp)	262.86	265.14	270.33	272.23
Percentage of repeat sequence sequences	52.84	53.06	53.71	53.85
BUSCO completeness of annotation	98.40	97.90	98.60	98.00
Percentage of gene with functional annotation	92.60	92.48	93.26	93.22

divergence between the two haplotypes (Supplementary Fig. 4b). Within the newly assembled regions, we identified cell division cycle protein 48, a key regulator of plant development and stress response[21], and UDP-glucose flavonoid 3-*O*-glucosyltransferase 7, which is critical for sugar decomposition in strawberries[22].

Identifying haplotype-specific and ASE genes

A long hybrid history has resulted in extensive genetic variations (Supplementary Note 3) and alleles between the two haplotypes in the pear. Based on the presence of ortholog genes in corresponding haplotypes, 33,710–34,036 allelic genes without the same coding sequence (CDS) were defined as biallelic genes (Fig. 2a,b). Of these genes, 6,691–8,087 contained partners with the same CDS in corresponding haplotypes (alleles with the same CDS) (Supplementary Table 11). We identified 1,703–1,994 genes in a haplotype with no allelic gene (haplotype-specific genes). Haplotype-specific genes had higher abundances of TEs compared with biallelic genes in the 5-kb upstream and downstream regions and gene body regions (Fig. 2c). The mean expression levels of haplotype-specific genes were lower than those of biallelic genes (or alleles with the same CDS) in different fruit samples of DS and MRB (Supplementary Fig. 14). Our analysis revealed that haplotype-specific genes exhibited higher DNA methylation levels in
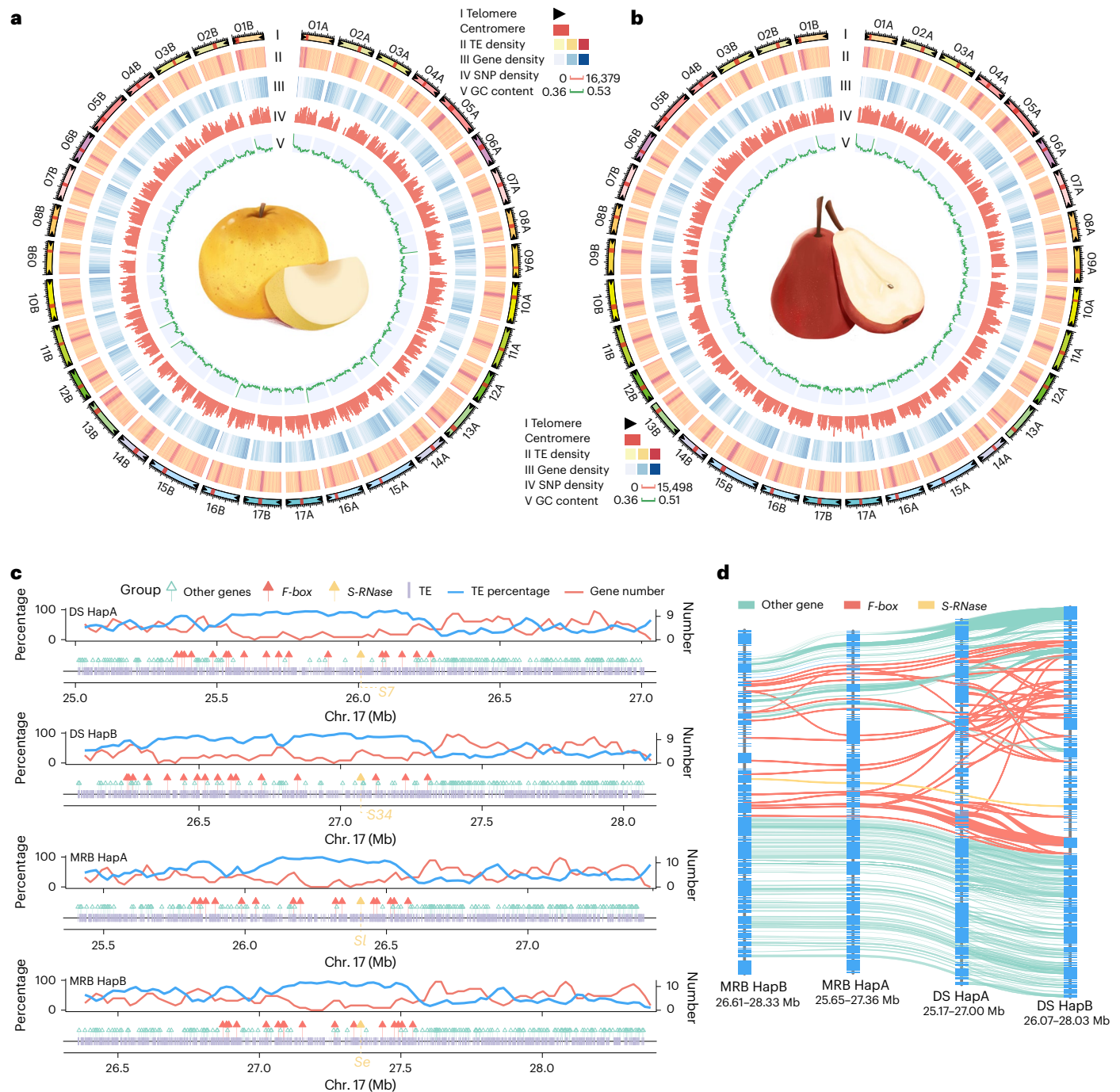
Fig. 1 | DS and MRB genome assemblies. a,b, Genomic features across haplotype A and B of DS (**a**) and MRB (**b**): (I) telomeres and centromeres; (II) TE density (number of TEs per 200 kb); (III) gene density (number of genes per 200 kb); (IV) SNP density (number of SNPs per 200 kb); and (V) GC content per 200 kb. 'A' means HapA and 'B' HapB. **c,** *S-RNase* and *SFB* (S-locus-related *F-box* gene) distribution in HapA and HapB of DS and MRB. The top graphs show gene number and TE percentage (per 50-kb window size and 20-kb step size) and the bottom graphs show the distribution of *S-RNase; SFB* and other genes and the genotype of *S-RNase* are marked. **d,** Orthologous gene pairs at an S-locus region across four haplotypes. The red line depicts orthologous *F-box* gene pairs, the yellow line the *S-RNase* gene pair and the green line other gene pairs.

both the gene body and the regions spanning ±5 kb from the gene in leaf and mature fruit tissues of DS and MRB (Supplementary Fig. 15). This suggested stronger repression by methylation in these areas. These findings were in line with previous research, which generally regarded DNA methylation as a regulatory factor for gene expression[23,24]. The haplotype-specific genes were enriched in 'Plant–pathogen interaction' and 'Environmental adaptation' in the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (Supplementary Fig. 16) and

the 'Response to biotic stimulus' gene ontology (GO) terms (Supplementary Fig. 17).

We further investigated allele-specific expression (ASE) during the eight fruit stages of DS and MRB: 19.64% (8,511 of 43,330) and 26.21% (11,213 of 42,784) of genes showed ASE in at least one sample of DS and MRB (Fig. 2d). These ASE genes were further classified into three categories (Methods): 4,498 (4,953) haplotype dominance (HapDom, genes with expression biased toward one allele), 219 (527) Sub (sub- or
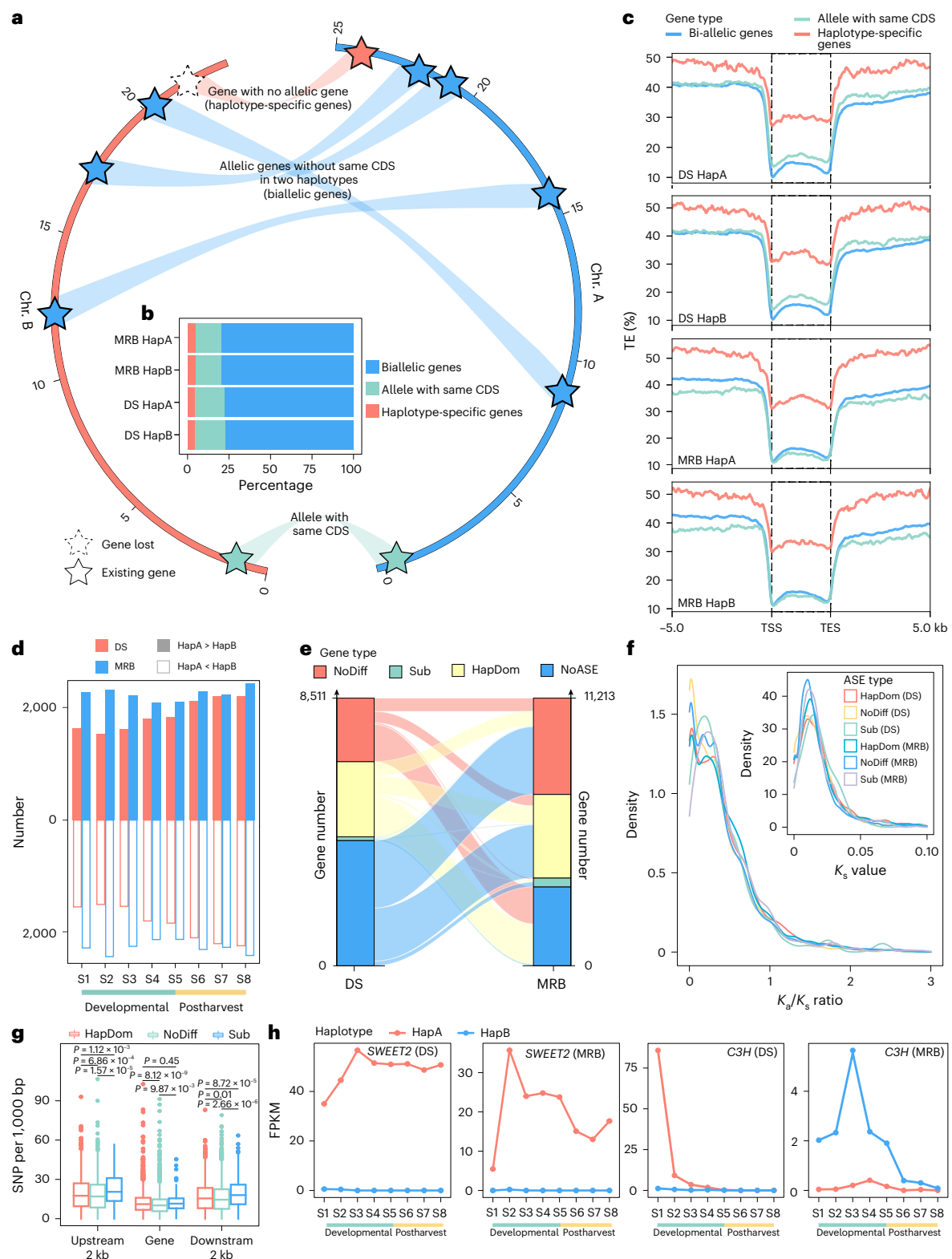
Fig. 2 | Allelic gene variation and expression pattern in DS and MRB. a, The layout for allele gene classification. The alleles were divided into the haplotype-specific genes (no paired genes in the corresponding haplotype), the alleles with the same CDS and biallelic genes. The star with the dotted line represents this gene not existing. **b**, Proportion of biallelic, allele with the same CDS and haplotype-specific genes in each haplotype. **c**, The TE distribution across gene body, 5-kb upstream and 5-kb downstream regions of biallelic, allele with the same CDS and haplotype-specific genes in HapA and HapB of DS and MRB. **d**, Number of ASE genes in eight fruit stages (including five developmental and three postharvest stages) of DS and MRB. **e**, The ASE gene type alteration between DS and MRB. HapDom indicates haplotype dominance gene pair. Sub denotes sub- or neofunctionalized pair and other ASE genes (Nodiff). NoASE denotes that the gene pair did not have ASE. **f**, The density of K_s and K_a/K_s values of ASE genes. **g**, The SNP density in gene features of HapA in DS ($nHapDom$ = 4,498, $nSub$ = 219, $nNodiff$ = 3,794). The center line is the median, the box limits the first and third quartiles and the whiskers 1.5× the interquartile range (IQR). The terms $nHapDom$, $nSub$ and $nNodiff$ refer to the counts of different ASE genes. The y axis represents SNP numbers every 1,000 bp. The two-sided Student's t-test was used for determining the significance. **h**, Expression pattern of the two alleles of *SWEET2* and *C3H* in fruit samples of DS and MRB. TES, transcription end site; TSS, transcription start site.

neofunctionalized pairs, alleles with switched expression in different samples) and 3,794 (5,733) other ASE genes (Nodiff) in DS (MRB). Only 44.52% of ASE genes in DS also exhibited an ASE pattern in MRB (Fig. 2e). Of ASEs in DS, 22.12% maintained a consistent category with ASEs in MRB and 22.40% ASEs in DS showed an inconsistent category with ASEs in MRB. Enrichment results showed that these inconsistent ASEs between DS and MRB were enriched in biosynthesis and metabolism pathways and consistent ASEs were enriched in response to stress and different stimulus GO terms (Supplementary Fig. 18).

We annotated 380 (in DS) and 450 (in MRB) ASE genes, including sugar-related fructokinases, α-glucosidases and sugar transporters, acid-related malate dehydrogenases, flavonoid-related flavanol synthases, lignin related *p*-coumarate 3-hydroxylase (C3H), 4-coumarate: coenzyme A ligase (4CL) and cinnamoyl-CoA reductase, which might contribute to important fruit agronomic traits at the allele level (Supplementary Fig. 20a). Among the ASE genes related to fruit traits, we observed that 95.26% (362 of 380) of the genes in DS and 88.89% (400 of 450) of the genes in MRB exhibited a haplotype dominance expression pattern, indicating that one allele had a greater influence on fruit quality and development than its counterpart. For example, SWEET2 (sugars will eventually be exported transporters 2) and CPK28 (calcium-dependent protein kinase 28) transport and increase sugar content[25,26] and showed a dominant expression pattern in DS and MRB (Fig. 2h and Supplementary Fig. 20b). *Pbre1_15G769200* and *Pbre1_13G264800*, which encode C3H and 4CL protein, respectively, are major structural genes involved in lignin biosynthesis[27,28]. These genes exhibited higher expression levels during the early fruit developmental stages (S1–S2) compared with their counterparts in DS HapB (Fig. 2h and Supplementary Fig. 20b).

Genomic evolution between Asian and European pears

To explore the evolutionary history of Asian and European pears, the protein-coding genes of DS, MRB and ten other species were used in comparative analyses (Fig. 3a). Phylogenetic analysis indicated that DS and MRB diverged approximately 3.60 million years ago (Ma). Compared with the common ancestor, 767 and 740 gene families were expanded in DS and MRB. The expanded gene families in DS were enriched in 'Biosynthesis of plant secondary metabolites' and 'Plant–pathogen interaction' pathways and 'Response to stress' GO terms (Supplementary Fig. 21a), which are important for plant development and adaptation[29,30]. The 'Fructose and mannose metabolism' and 'Carbohydrate metabolism' pathways and 'Secondary metabolic process' GO terms were featured in the expanded gene families in MRB.

The extensive evolutionary divergence between Asian and European pears implies their marked genomic variations (Fig. 3b). In this study, four haplotypes were used to construct a graph-based genome for SV identification and 103,069 SVs were successfully identified and genotyped. Of the SVs, 57.42% were in the gene body and gene-regulatory regions (±2 kb) and 61.27% overlapped with TEs (Fig. 3c). In addition, 12 large inversions (>100 kb) were identified, spanning 5.79 Mb, and all were supported by ONT reads or Hi-C maps (Fig. 3d and Supplementary Fig. 23). Notably, a 0.58-Mb inversion on chromosome 17 (chr. 17) was present between DS HapA and HapB at the S-locus regions (Fig. 3d). This inversion may suppress recombination and maintain the linkage of *SFB* and *S-RNase* genes for SI[31,32].

To reveal the SVs between Asian and European pears, NGS data from 362 pear accessions were collected (Supplementary Table 15) and mapped to the graph-based genome for SV genotyping. The number of genotyped SVs reached a plateau as more Asian and European samples were added (Fig. 3e). A total of 83,369 SVs were retained with minor allele frequency >0.05 and missing rate <0.2. The reference (DS HapA as reference) presence frequency of 76.82% SVs was >0.5 in the Asian pear group (Fig. 3f), and the frequencies of 42.73% SVs were >0.5 in the European pear group. There were 35,481 SVs that showed different presence frequency (false discovery rate (FDR) < 0.001 and fold-change >2) between the Asian and European pear groups.

Evolutionary history and genetic introgression

We further explored the introgression events, demographic history and domestication process. The SNP-based and SV-based phylogenetic tree showed that a division of the 362 pear accessions into 2 distinct clades, with clade I comprising most Asian pears and clade II consisting of European pears (Fig. 4a and Supplementary Fig. 27a). Clade I was further subdivided into four subclades—(1) *P. bretschneideri* and *Pyrus pyrifolia*; (2) Asian wild pears; (3) wild and cultivated *Pyrus ussuriensis*; and (4) *Pyrus sinkiangensis* and interspecific hybrids—and clade II was subdivided into two subclades—(5) *P. communis* and (6) European wild pears. High fixation statistics (F_{ST}) and principal component analysis results supported a long divergence history between Asian and European pears (Fig. 4b and Supplementary Fig. 27b,c). Cultivated *P. ussuriensis* (4.80×10^{-3}) had higher nucleotide diversity than wild *P. ussuriensis* (3.26×10^{-3}). The structure analysis revealed the historical hybridization of the cultivated pears (Fig. 4a and Supplementary Fig. 27d) and cultivated *P. ussuriensis* showed two genetic backgrounds, originating from wild *P. ussuriensis* and cultivated *P. pyrifolia* (or *P. bretschneideri*). The Treemix analysis detected gene flow from *P. bretschneideri* to *P. ussuriensis* (Supplementary Fig. 28), suggesting that an introgression event may have played a role in *P. ussuriensis* domestication.

Inferences about the demographic history revealed three effective declines in population size (N_e) in Asian (except wild *P. ussuriensis*) and EW pears (Fig. 4c and Supplementary Fig. 29a). The first decline event occurred at 2.5–1.7 Ma during the Gelasian epoch (2.59–1.81 Ma)[33], characterized by a dramatic temperature decrease. The second N_e decline event occurred 1.0–0.6 Ma with another period of temperature decrease[34] and the last decline occurred 250 thousand years ago (ka) to 160 ka. It was followed by a quick demographic expansion from 150 ka to 10 ka, consistent with increasing temperatures[35]. Only two N_e decline events were observed in wild *P. ussuriensis* (Supplementary Fig. 29b), with the second event occurring from 380 ka. to 180 ka A slight rebound was observed thereafter, resulting in the lower N_e of wild *P. ussuriensis* compared with Asian wild (AW) and European wild (EW) pears.

Selection signatures for fruit quality of pear

The cultivated pears, domesticated from wild progenitors, exhibit various phenotypical changes, including sugar, acid, stone cell and fruit size. In this study, we identified 21.31 Mb of selective sweep regions for AW versus Asian cultivated (AC, including *P. pyrifolia* and *P. bretschneideri*) pears and 14.59 Mb of selective sweep regions for EW versus European cultivated (EC; *P. communis*) pears (Fig. 4d,e); 123 selected genes related to fruit size, plant development, stone cell content, acids and sugars were identified in the selective sweeps (Supplementary Table 18). Notably, a robust selective sweep region was identified at chr. 16 for AW versus AC pears, and fruit acid-, size- and sugar-related genes were identified, including *Ma1* (*Pbre1_16G081600*), *fw2.2* (*Pbre1_16G040000* and *Pbre1_16G040200*) and *SWEET15* (*Pbre1_16G220200*) (Fig. 4d). The orthologs of *Ma1* and *fw2.2* also appeared in a selective sweep in cultivated apples[36], suggesting that apple and pear may experience convergent selection for fruit acid and size.

Stone cell content (SCC) is a crucial trait in pear domestication. SCC forms during the early stages of fruit development and is associated with lignin and cellulose formation[37]. In Asian pear domestication, 11 structural genes identified in selective sweeps of AW versus AC pears were related to lignin and cellulose formation (Supplementary Table 18). *PbrbZIP48* (*Pbre1_16G094200*), which regulates the expression of stone cell-related genes[38], was identified in the selective sweeps of AC pears (Fig. 4d). *PbrmiR397a*, identified in the selective sweep, can silence the expression of LAC genes, thereby reducing lignin and stone cell contents in pear flesh[39]. These results suggested that structural genes, transcription factors and microRNAs (miRNAs) related to stone
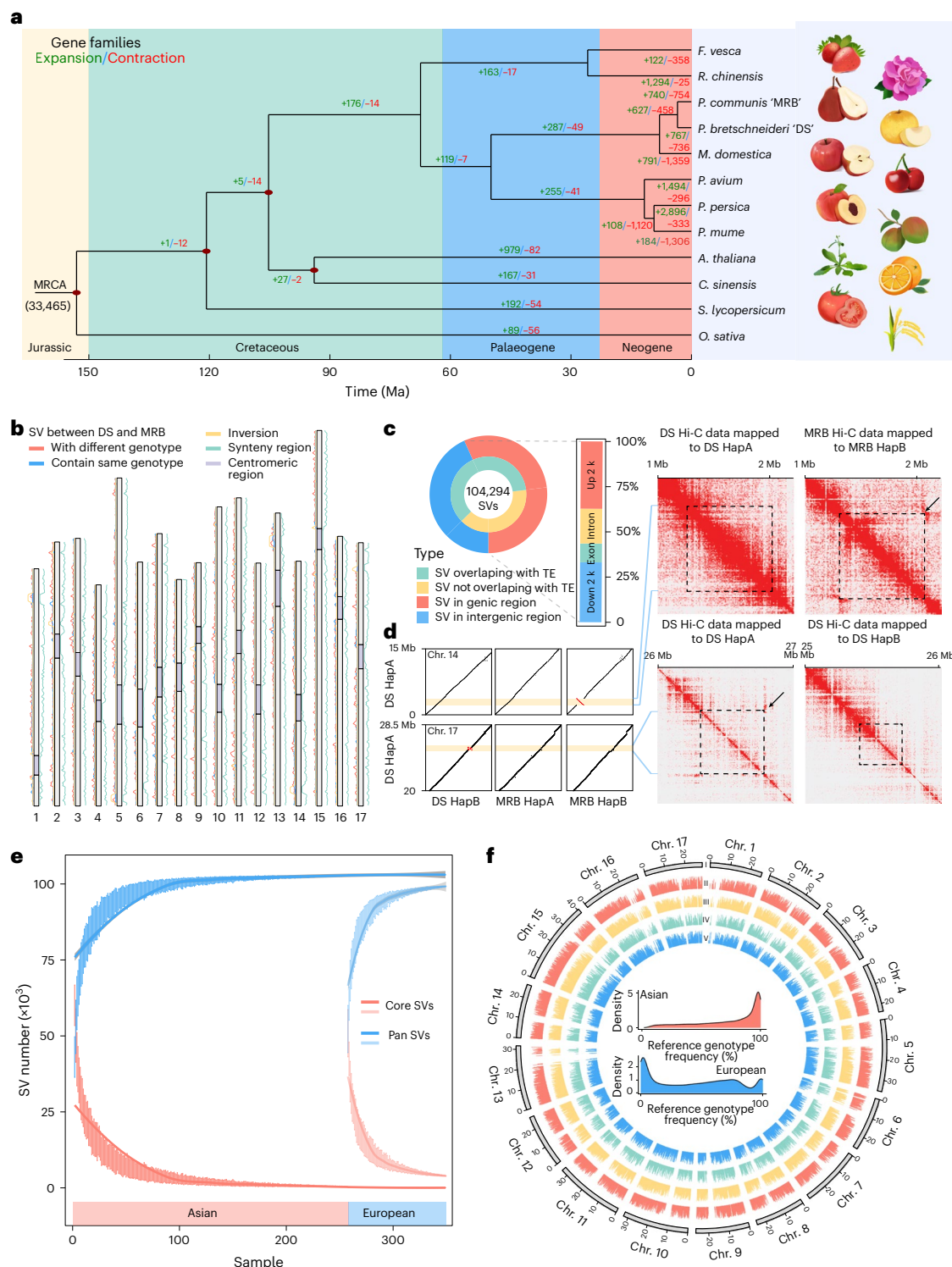
Fig. 3 | Divergence and SVs between DS and MRB. a, Phylogenetic, molecular clock dating and gene family analysis of the DS and MRB genome with ten other species. The red oval indicates the fossil calibration point. The gain and loss of total gene families in DS, MRB and the other ten species are highlighted in green and red, respectively. **b**, Ideogram with synteny regions, SVs and inversion distribution in 200-kb bins. Synteny regions, SV and inversion distribution are plotted to the side of each chromosome. Centromeric region is plotted in chromosome representations. **c**, The proportion of SVs overlapping with TEs or different genomic regions. In the gene body, upstream 2-kb and downstream 2-kb regions of the gene body are considered genic regions and other regions are considered intergenic regions. **d**, Large inversions between two haplotypes in DS or MRB. The yellow regions on the collinear blocks represent the inversions.

Hi-C heatmaps support the inversion. The dashed box represents the inversion and the arrows show the high interaction signal at the reverse diagonal. **e**, Pan-SV and core-SV count with Asian and European pear accessions. Accessions were sampled as 1,000 random combinations of each given number of accessions. **f**, The distribution and frequency density of SV genotype between Asian and European pears. The circos plot shows the distribution of the frequency variance between Asian and European pears: (I) the chromosomes of HapA of DS; and (II–V) frequency variance of SVs that overlapped with the upstream 2-kb (II), downstream 2-kb (III), exon (IV) and intron (V) regions. The density graph shows the distribution of SV frequency with reference genotype (HapA of DS as reference).
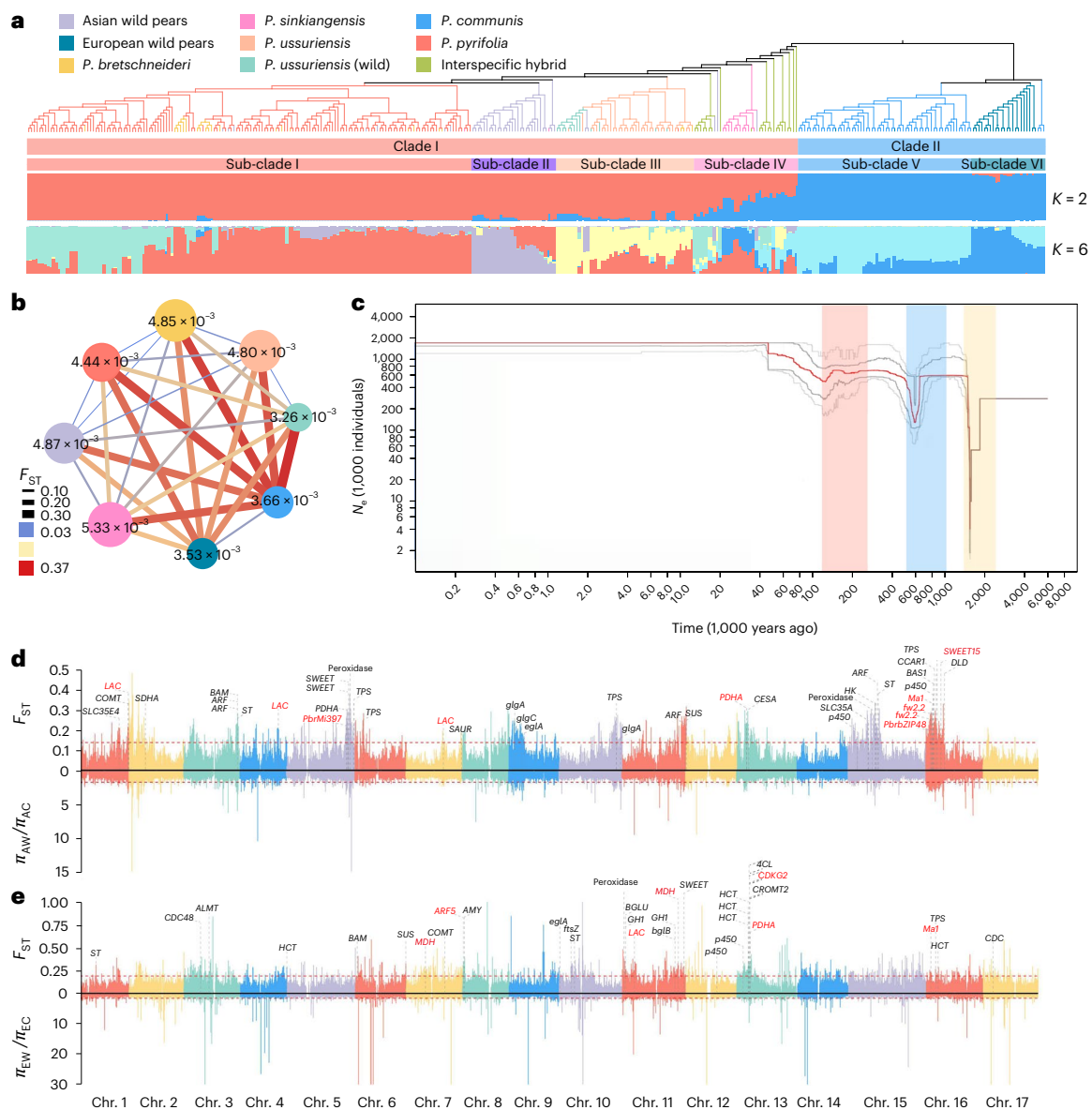
Fig. 4 | Phylogenetic relationship, demographic history and selective sweeps of different pear groups. a, Phylogenetic tree and population structure of 362 pear accessions, inferred using SNPs at fourfold degenerate sites. K is the the number of ancestry kinship. **b**, Mean nucleotide diversity in each group and population differentiation (F_{ST}) between groups. The line width and color of the edges represent F_{ST} values between the two groups. The circle size and value represent the nucleotide diversity of each pear group. The label of circle colors is the same with a phylogenetic tree plot. **c**, Historical effective population size N_e for Asian wild accessions (except wild *P. ussuriensis*) with a generation time of 7.5 years. Orange, blue and yellow shadows represent three bottlenecks. **d,e**, Genome-wide screening of selective sweeps in AW versus AC pears (including *P. pyrifolia* and *P. bretschneideri*) (**d**) and EW versus EC pears (*P. communis*) (**e**). The red dotted lines represent the threshold of the top 5% F_{ST} and 5% ratio of π_{wild} to $\pi_{cultivar}$, respectively.

cell formation may have been selected in parallel during Asian pear domestication to reduce the SCC in fruit.

Fruit size is widely considered a domestication trait in domesticated crops[40,41] and it is known to be regulated by two key processes: cell division and cell expansion[41]. Cell division and cell expansion genes were identified, suggesting that they play a role in these processes and may affect fruit size (Supplementary Table 18). *Pbre1_13G194600* encodes CDKG2 (cyclin-dependent kinase G2), a protein that positively regulates cell growth in *Arabidopsis* spp.[42]. *Pbre1_08G029000* encodes ARF5 (auxin response factor 5) and the homology of *ARF5* has been proved to increase fruit size in tomatoes[43], suggesting its potential to increase fruit size in pears. In addition, five missense SNPs with different genotype frequencies between EW and EC pears were identified (Supplementary Fig. 30a).

SVs associated with agronomic traits

Previous studies have suggested that SVs are causative variants for agronomic traits[29]. The high-quality 83,369-SV set provided us a chance to reveal the trait-associated SVs for pear. We performed an SV-based genome-wide association study (GWAS) for fruit transverse diameter (FTD), soluble solid content (SSC), SCC (Supplementary Note 11) and single fruit weight (SFW) (Supplementary Note 11), which are important traits for pear fruit quality and production. We identified a 66-bp deletion at chr. 2 that was significantly associated with FTD (Fig. 5a). This deletion overlapped with the selective sweep of AC pears (Fig. 5b) and its occurrence in *P. pyrifolia* (Ppy) and *P. bretschneideri* (Pbr) was lower than in the AW group (Fig. 5c). In addition, pear accessions without deletion had significantly higher FTD compared with those with deletion (Fig. 5d). A candidate gene, *PyTPR*, encoding a tetratricopeptide
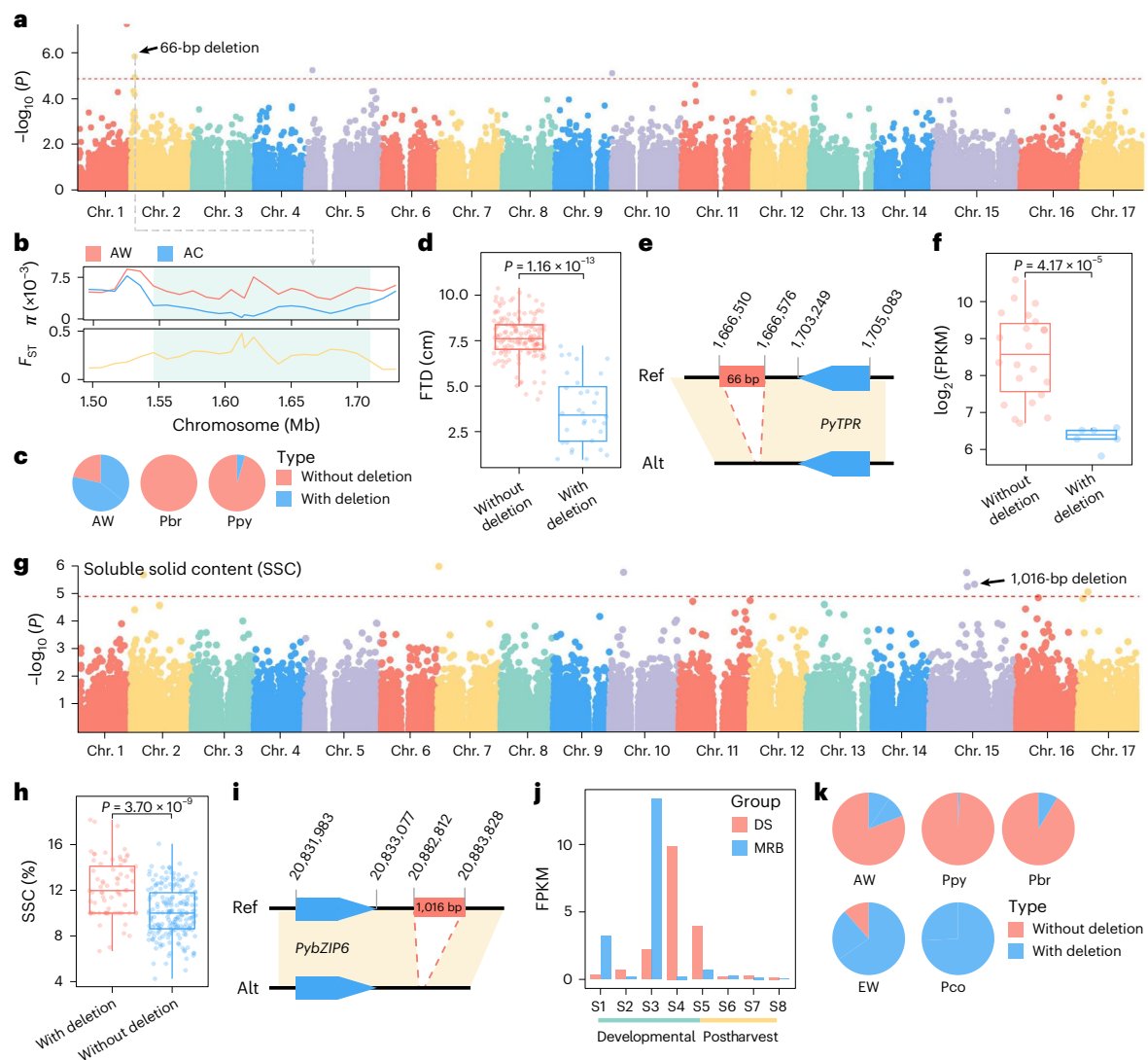
Fig. 5 | SV GWAS of FTD and SSC. a, Manhattan plot displaying the SV GWAS outcomes of the FTD. *P* values were computed from a two-sided *F* test. The red dotted lines indicate the significance thresholds ($1/n$, where *n* denotes the SV number) for SV-based GWASs. **b**, Distribution of nucleotide diversity (π) of AW and AC pears (including *P. pyrifolia* and *P. bretschneideri*) and F_{ST} between AW and AC pears in a 20-kb interval. The shaded region means the selective sweep of AW versus AC pears. **c**, The deletion appearance frequency in AW pears and Ppy and Pbr groups. **d**, Comparison of FTD between samples with ($n = 32$) and those without ($n = 150$) deletion. **e**, The significantly associated SV downstream of *PyTPR*. **f**, *PyTPR* expression pattern in samples with ($n = 6$) and without ($n = 22$) deletion. **g**, Manhattan plot displaying the SV GWAS outcomes of the SSC. **h**, Comparison of SSC between samples with ($n = 73$) and those without ($n = 199$) deletion. **i**, The significantly associated SV downstream of *PybZIP6*. **j**, *PybZIP6* expression pattern in DS and MRB at five fruit developmental (S1–S5) and three postharvest (S6–S8) stages. **k**, The deletion appearance frequencies in AW, Ppy, Pbr, EW and Pco groups. In **d**, **f** and **h**, *P* values were computed from two-sided Student's *t*-tests. In **d**, **f** and **h**, each box represents the center line (median), the box limits (first and third quartiles) and the whiskers (1.5× the IQR).

repeat-like superfamily protein, was identified downstream of this deletion (Fig. 5e). This gene is involved in plant development, elongation of cotton fibers[44] and *Arabidopsis* roots[45]. Accessions without deletion have significantly higher expression levels of *PyTPR* compared with accessions with deletion (mean number of fragments per kilobase of exon per million mapped fragments (FPKM): 497.33 versus 81.85) (Fig. 5f). These results suggested that *PyTPR* was under selection during Asian pear domestication and may have a potential role in pear FTD and fruit development.

Similarly, we identified a 1,016-bp deletion associated with SSC (Fig. 5g) and accessions with this deletion showed significantly higher SSC than those without it (Fig. 5h). *Pbre1_15G533000*, which encodes a bZIP transcription factor (PybZIP6), was identified (Fig. 5i). It has been reported that bZIP transcription factors are involved in SSC accumulation in pears[46–48]. PybZIP6 was highly expressed during the late fruit

development stages (S3–S5) in DS, MRB and four other representative pear accessions, suggesting its potential in SSC accumulation at late fruit developmental stages (Fig. 5j and Supplementary Fig. 32). Compared with the EW group, the deletion occurrence frequency was higher in the *P. communis* (Pco) group (Fig. 5k). Overall, our SV-based GWAS provided new markers that will be helpful to improve the fruit quality of pear cultivars.

The role of *PyACS1* in fruit maturity and softening

Nearby SVs can alter gene expression patterns due to the interruptions in regulatory or gene regions[49,50]. RNA-seq data from 110 samples (consisting of 62 Asian pears and 38 European pears) were generated to identify the SV effect on gene expression (Supplementary Table 19). A total of 4,040 SVs in the upstream 2-kb region and 568 SVs at the gene exon region were significantly correlated with gene expression

(Fig. 6a); 70.79% (2,860 of 4,040) SVs and 65.49% (372 of 568) SVs exhibited divergent occurrence (FDR < 0.001 and fold-change >2) between the Asian and the European groups, suggesting the potential of SVs in regulating intergroup gene expression. These genes were enriched in metabolism and biosynthesis process, including 'fructose and mannose metabolism' and 'citrate cycle' (Supplementary Figs. 34 and 35). A total of 40 SVs and 12 SVs were significantly correlated with 36 sugar-related and 10 acid-related genes, including previously reported, fruit sugar-related genes like *PpySTOP1*, *PpySPS8* and *PbVHA-A*[25,48] (Supplementary Table 20).

Compared with Asian pears, European pears are usually classified as climacteric fruit and ethylene production is responsible for the initiation and progression of fruit ripening and softening[51,52]. Five ethylene biosynthesis-related genes with SVs were identified (Supplementary Table 20) and one gene (*Pbre1_15G535800*) encoding 1-aminocyclopropane-1-carboxylic acid synthase (*PyACS1*) functions as the rate-limiting enzyme in ethylene synthesis[53]. Transcriptome analysis showed that *PyACS1* was significantly more highly expressed in European pears than in Asian pears (Fig. 6b). The interspecific hybrids have higher expression levels than both Asian and European pears, indicating that overdominance expression pattern may be observed in these hybrid accessions[54]. The *PyACS1* gene was only highly expressed during the late development fruit stages (S4 and S5) and three postharvest (S6–S8) stages in MRB and the FPKM value of *PyACS1* was <0.1 at all eight stages (S1–S8) of DS (Fig. 6c). We further explored the functions of *PyACS1* during fruit maturation and postharvest stages (S5–S8). Overexpressing *PyACS1* (OE-*PyACS1*) tomato lines showed accelerated ripening (Fig. 6d, e), higher ethylene (Fig. 6f) and lower firmness (Fig. 6g) compared with wild-type (WT) fruit, suggesting that *PyACS1* promotes fruit ripening and softening.

Furthermore, a 286-bp insertion sequence was identified in the promoter region of *PyACS1*, which was present in all European pears and interspecific hybrid accessions, but not in Asian pears (Fig. 6h–j). A PCR assay further confirmed that the 286-bp insertion sequence was present in European pears and absent in Asian pears (Fig. 6i and Supplementary Fig. 36). Sequence analysis of insertion showed that one E-box (CATTTG) and one G-box (CACGTT) *cis*-element were included (Supplementary Fig. 37a), which are the potential binding sites of basic helix–loop–helix transcription factors[55,56]. Thus, we conducted weighted co-expression network analysis and one basic helix–loop–helix transcription factor (PybHLH94) was identified in the *PyACS1* gene co-expression module (Supplementary Fig. 37b). Dual-luciferase (LUC) transcription assay in tobacco protoplasts revealed that PybHLH94 can significantly activate *PyACS1*pro^MRB (with a 286-bp insertion), but did not activate *PyACS1*pro^DS (without a 286-bp insertion), suggesting insertion-dependent transcriptional regulation (Fig. 6k). Promoter truncation results showed that activation occurred within the −1,221-bp to −830-bp region, which contained the 286-bp insertion sequence (−1,141 bp to −855 bp) (Fig. 6l and Supplementary Fig. 37c). This suggests that PybHLH94 can specifically activate the promoter of *PyACS1* by binding to the 286-bp insertion sequence. The electrophoretic mobility shift assay (EMSA) result showed that *PybHLH94* (His) binds directly to the G-box and E-box in the 286-bp insertion sequence (Fig. 6m).

Discussion

In this study, we generated T2T haplotype-resolved genome assemblies for DS and MRB, allowing us to investigate all genomic regions and the allelic variations. Based on the haplotype assemblies, haplotype-specific genes and biallelic genes were identified. TEs and methylation have been proven to regulate gene expression[57,58] and variation of methylation and TE between haplotype-specific genes and biallelic genes may be related to gene expression alteration in DS and MRB. HapDom ASE has an allele with biased expression in different fruit developmental stages of the pear, suggesting the dominance effect

on heterosis. Furthermore, a sub-ASE gene exhibits one allele-biased expression in several samples but shifts to another allele in other samples, suggesting the overdominance effect[59]. As with apple and kiwifruit[5,60], we observed that HapDom genes were more than subgenes, suggesting the major role of the dominance effect during fruit development. Our results showed that less than half of ASE genes were shared between DS and MRB and 22.12% of ASEs kept consistent categories, suggesting that ASE patterns were not conserved between different accessions. Studies showed that ASEs were related to the genetic and epigenetic variations[8,60]. Considering the long divergence history, variation accumulation between DS and MRB may have the potential for ASE alteration and further affect the adaptive capacity and plant development[59,61].

Previous studies showed that cultivated *P. ussuriensis* was directly derived from wild *P. ussuriensis*[10]. However, we found that cultivated *P. bretschneideri* (and *P. pyrifolia*) played an important role in *P. ussuriensis* domestication. Cultivated *P. ussuriensis* showed higher nucleotide diversity than wild *P. ussuriensis*, which was likely increased by interspecific introgression[6]. Wild *P. ussuriensis* has high cold tolerance and was mainly distributed in the frigid environment of China[10]. Through introgression from cultivated *P. bretschneideri*, cultivated *P. ussuriensis* may promote the improvement of fruit quality and preserve adaptability of wild *P. ussuriensis* to a cold environment at the same time. These findings provide a new perspective on the domestication history of pears and offer a new strategy for breeding cold-resistant pear varieties.

Fruit sweetness, acidity and SCC are well-known domestication traits that are still the main targets in pear-breeding programs. We found many previously unreported trait-associated genes in selective sweeps based on the gold standard reference genome. *PbrbZIP48* and *PbrmiR397a* were also identified in the selective sweeps of Asian pears and the products (a transcription factor and miRNA, respectively) regulate lignin biosynthesis-related genes and affect the SCC in fruit flesh[39]. These findings demonstrate that transcription factors, miRNAs and protein-coding genes may cause pear domestication-associated changes. *MiR397* also appears in the *Indica* rice domestication process[62], which suggests that *miR397* was widely selected during plant domestication due to its conserved function in crop improvement. In addition, we also found that *Ma1* and *fw2.2* were selected during the domestication of apples and pears. *Ma1* has been identified as a factor that can affect the fruit acidity of cultivated apples and a mutation (G to A) causes the *Ma1* protein truncation and is responsible for the low fruit acidity in apple[5]. Still, this mutation did not appear in pears, implying that other mutations may have been selected during pear domestication. These selected or co-selected genes provide helpful genetic resources for understanding the domestication process and facilitating pear breeding.

Based on the two high-quality genome assemblies for DS and MRB, we constructed a graph-based genome and used it for SV genotyping with next-generation sequencing (NGS) data and identified additional signatures using the SV-GWAS approach, including several potential candidate genes for agronomically important traits. Ethylene production is the response for the initiation and progression of fruit ripening and softening in European pears[51,52,63]. In this study, we identified that *PyACS1* has a high expression level in European pears and provided evidence of its role in fruit ripening and softening. We found one 286-bp insertion sequence from the promoter of the *PyACS1* in European pears and this insertion sequence contains two new *cis*-elements that strengthen transactivation by *PybHLH94*. Asian and European pears experienced long divergence history and independent domestication events[10] and their hybridization may form new germplasms. Overdominance gene expression patterns have been documented in hybrids of potato[64], rice[65–67] and *Arabidopsis* spp.[68], which have elevated gene expression levels compared with their parental lines. Similarly, *PyACS1* exhibits higher expression levels in interspecific hybrids compared with Asian and European pears, suggesting that, aside from
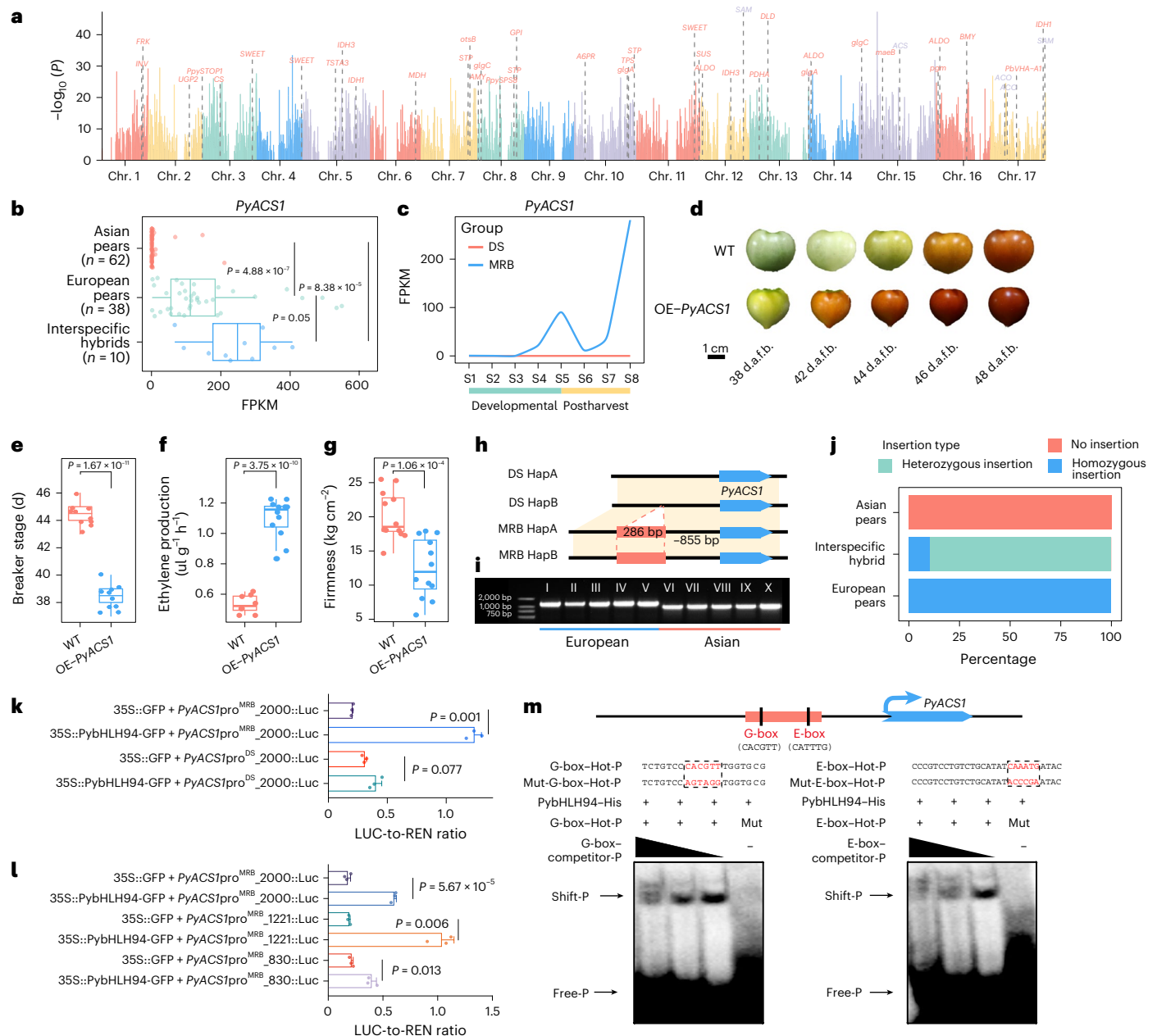
Fig. 6 | SVs associated with gene expression and functional impact on fruit ripening and softening of European pears. a, Manhattan plot displaying the SVs at exon or promoter regions significantly associated with gene expression. Sugar- and acid-related genes were marked in red and ethylene biosynthesis genes were marked in purple. **b**, The expression of *PyACS1* in Asian ($n = 62$), European ($n = 38$) and interspecific hybrid ($n = 10$) pears. The 'n' means the number of accessions (center line as median, box limits as first and third quartiles and whiskers as 1.5× the IQR). **c**, *PyACS1* expression pattern of DS and MRB at eight fruit stages. **d**, Phenotype of WT tomato fruit and tomato transformants overexpressing *PyACS1* (OE-*PyACS1*). **e–g**, Breaker stage ($n_{WT} = 10$, $n_{OE-PyACS1} = 10$), ethylene production ($n_{WT} = 8$, $n_{OE-PyACS1} = 12$) and firmness ($n_{WT} = 12$, $n_{OE-PyACS1} = 12$) in the WT and OE-*PyACS1* tomato fruit (center line as median, box limits as first and third quartiles and whiskers as 1.5× the IQR). The n_{WT} and $n_{OE-PyACS1}$ refer to the numbers of fruit samples used for WT and OE-*PyACS1*. **h**, A 286-bp insertion sequence identified at the promoter region of *PyACS1* in MRB. **i**, PCR-based screen showing

the absence (right) or presence (left) of the insertion sequence in the upstream of *PyACS1* in European (I–V represent 'Max Red Bartlett', 'Red comice', 'Spalding', 'Charneus' and 'Anjou', respectively) and Asian (VI–X represent 'Dangshansuli', 'Cuiguan', 'Cuiyu', 'Yunhong No. 1' and 'Yali', respectively) pears. **j**, The insertion appearance frequency in Asian, European and interspecific hybrid pears. **k**, TE assays showing significant differences in promoter activities for *PyACS1*pro^MRB and *PyACS1*pro^DS. The data are presented as mean ± s.e.m. **l**, Relative activity levels of luciferase of three truncated promoters (830 bp, 1,221 bp and 2,000 bp) of *PyACS1*pro^MRB. The data are presented as mean ± s.e.m. **m**, EMSA showing that PybHLH94 binds directly to the G-box and E-box within the insertion sequence at the promoter region of *PyACS1*. Hot-P and Mut-P represent biotin-labeled WT and mutation probes, respectively; competitor-P represents the unlabeled WT sequence. In **b**, **e–g** and **k–l**, *P* values were computed from two-sided Student's *t*-tests. The numbers of biologically independent samples used in **k–l** is 3. The full scans of all the gels are included in Source data.

the 286-bp insertion sequence, multiple factors in these hybrids may contribute to the overdominance expression pattern. Furthermore, accession-specific mutations may also influence gene expression by changing *cis*-elements like ABRE, STRE and G-box (Supplementary

Fig. 38)[69–71]. However, the direct evidence of a regulatory effect of the 286-bp insertion in *PyACS1* should be obtained in future research, once current technical limitations, such as the lack of a stable transgenic or genome-editing system in pears, has been overcome. Overall, our

results showed that a graph-based genome may be a powerful tool for accurate population-level SV genotyping and genomic studies, as well as the forward application of genes and SVs in pear genome-guided breeding.

In conclusion, we generated two T2T haplotype-resolved pear genomes in this study and provided insights into genome evolution, haplotype imbalance and interspecies variation in *Pyrus* spp. Combining the high-quality genomic and population resequencing data allowed us to show the extensive genetic variance between Asian and European pears and identify trait-associated SVs. These genomic resources will provide a basis for further molecular biology research and will accelerate the breeding process in commercial pears.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-025-02273-4.

References

1. Zhou, Q. et al. Haplotype-resolved genome analyses of a heterozygous diploid potato. *Nat. Genet.* **52**, 1018–1023 (2020).
2. Murase, K. et al. Mechanism of self/nonself-discrimination in *Brassica* self-incompatibility. *Nat. Commun.* **11**, 4916 (2020).
3. Chen, J. et al. Phosphatidic acid counteracts S-RNase signaling in pollen by stabilizing the actin cytoskeleton. *Plant Cell* **30**, 1023–1039 (2018).
4. Huang, X. H. et al. Genomic architecture of heterosis for yield traits in rice. *Nature* **537**, 629–633 (2016).
5. Sun, X. P. et al. Phased diploid genome assemblies and pan-genomes provide insights into the genetic history of apple domestication. *Nat. Genet.* **52**, 1423–1432 (2020).
6. Zhang, X. T. et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*. *Nat. Genet.* **53**, 1250–1259 (2021).
7. Hu, G. B. et al. Two divergent haplotypes from a highly heterozygous lychee genome suggest independent domestication events for early and late-maturing cultivars. *Nat. Genet.* **54**, 73–83 (2022).
8. Hu, W. et al. Allele-defined genome reveals biallelic differentiation during cassava evolution. *Mol. Plant* **14**, 851–854 (2021).
9. Zhang, Q. et al. Genomic insights into the recent chromosome reduction of autopolyploid sugarcane *Saccharum spontaneum*. *Nat. Genet.* **54**, 885–896 (2022).
10. Wu, J. et al. Diversification and independent domestication of Asian and European pears. *Genome Biol.* **19**, 77 (2018).
11. Wu, J. et al. The genome of the pear (*Pyrus bretschneideri* Rehd.). *Genome Res.* **23**, 396–408 (2013).
12. Chagne, D. et al. The draft genome dequence of European pear (*Pyrus communis* L. 'Bartlett'). *PLoS ONE* **9**, e92644 (2014).
13. Linsmith, G. et al. Pseudo-chromosome-length genome assembly of a double haploid 'Bartlett' pear (*Pyrus communis* L.). *Gigascience* **8**, giz138 (2019).
14. Shi, D. Q. et al. Single-pollen-cell sequencing for gamete-based phased diploid genome assembly in plants. *Genome Res.* **29**, 1889–1899 (2019).
15. Qi, Y. et al. Telomere-to-telomere genome assembly of two representative Asian and European pear cultivars. *Sci. Data* **11**, 1170 (2024).
16. Chen, X. et al. Genome-wide genetic diversity and IBD analysis reveals historic dissemination routes of pear in China. *Tree Genet Genomes* **18**, 1 (2022).
17. Zhou, Y. et al. Graph pangenome captures missing heritability and empowers tomato breeding. *Nature* **606**, 527–534 (2022).
18. Cheng, H. Y. et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
19. Rautiainen, M. et al. Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat. Biotechnol.* **41**, 1474–1482 (2023).
20. Nurk, S. et al. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).
21. Ines, D. et al. CDC48 in plants and its emerging function in plant immunity. *Trends Plant Sci.* **29**, 786–798 (2024).
22. Griesser, M. et al. Multi-substrate flavonol *O*-glucosyltransferases from strawberry (*Fragaria* x *ananassa*) achene and receptacle. *J. Exp. Bot.* **59**, 2611–2625 (2008).
23. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39**, 457–466 (2007).
24. Stroud, H. et al. Comprehensive analysis of silencing mutants reveals complex regulation of the *Arabidopsis* methylome. *Cell* **152**, 352–364 (2013).
25. Li, J. et al. Natural variations in the PbCPK28 promoter regulate sugar content through interaction with PbTST4 and PbVHA-A1 in pear. *Plant J.* **114**, 124–141 (2023).
26. Lama, K. et al. Extreme sugar accumulation in late fig ripening is accompanied by global changes in sugar metabolism and transporter gene expression. *Physiol. Plant* **174**, e13648 (2022).
27. Xue, C. et al. PbrMYB169 positively regulates lignification of stone cells in pear fruit. *J. Exp. Bot.* **70**, 1801–1814 (2019).
28. Barros, J. et al. 4-Coumarate 3-hydroxylase in the lignin biosynthesis pathway is a cytosolic ascorbate peroxidase. *Nat. Commun.* **10**, 1994 (2019).
29. Yang, C. Q. et al. Transcriptional regulation of plant secondary metabolism. *J. Integr. Plant Biol.* **54**, 703–712 (2012).
30. Dodds, P. N. & Rathjen, J. P. Plant immunity: towards an integrated view of plant-pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548 (2010).
31. Zhu, S. et al. The snapdragon genomes reveal the evolutionary dynamics of the S-locus supergene. *Mol. Biol. Evol.* **40**, msad080 (2023).
32. Yang, Q. et al. Heterochromatic and genetic features are consistent with recombination suppression of the self-incompatibility locus in *Antirrhinum*. *Plant J.* **51**, 140–151 (2007).
33. Cohen, K. M. et al. The ICS International Chronostratigraphic Chart. *Episodes* **36**, 199–204 (2013).
34. Guo, J. F. et al. Genomic clines across the species boundary between a hybrid pine and its progenitor in the eastern Tibetan Plateau. *Plant Commun.* **4**, 100574 (2023).
35. Petit, J. R. et al. Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
36. Liao, L. et al. Unraveling a genetic roadmap for improved taste in the domesticated apple. *Mol. Plant* **14**, 1454–1471 (2021).
37. Xu, S. et al. Auxin inhibits lignin and cellulose biosynthesis in stone cells of pear fruit via the PbrARF13-PbrNSC-PbrMYB132 transcriptional regulatory cascade. *Plant Biotechnol. J.* **21**, 1408–1425 (2023).
38. Gong, X. et al. Multi-omics analyses reveal stone cell distribution pattern in pear fruit. *Plant J.* **113**, 626–642 (2023).
39. Xue, C. et al. PbrmiR397a regulates lignification during stone cell development in pear fruit. *Plant Biotechnol. J.* **17**, 103–117 (2019).
40. Li, Y. et al. Genomic analyses of an extensive collection of wild and cultivated accessions provide new insights into peach breeding history. *Genome Biol.* **20**, 36 (2019).
41. Azzi, L. et al. Fruit growth-related genes in tomato. *J. Exp. Bot.* **66**, 1075–1086 (2015).

42. Jiang, S. et al. The UBP14-CDKB1;1-CDKG2 cascade controls endore-duplication and cell growth in *Arabidopsis*. *Plant Cell* **34**, 1308–1325 (2022).

43. Liu, S. et al. Tomato AUXIN RESPONSE FACTOR 5 regulates fruit set and development via the mediation of auxin and gibberellin signaling. *Sci. Rep.* **8**, 2971 (2018).

44. Fang, D. D. et al. An EMS-induced mutation in a tetratricopeptide repeat-like superfamily protein gene (Ghir_A12G008870) on chromosome A12 is responsible for the li$_y$ short fiber phenotype in cotton. *Theor. Appl. Genet.* **133**, 271–282 (2020).

45. Zhang, M. et al. A tetratricopeptide repeat domain-containing protein SSR1 located in mitochondria is involved in root development and auxin polar transport in *Arabidopsis*. *Plant J.* **83**, 582–599 (2015).

46. Wang, H. et al. A pear S1-bZIP transcription factor PpbZIP44 modulates carbohydrate metabolism, amino acid, and flavonoid accumulation in fruits. *Hortic. Res.* **10**, uhad140 (2023).

47. Jia, L. et al. PbrbZIP15 promotes sugar accumulation in pear via activating the transcription of the glucose isomerase gene PbrXylA1. *Plant J.* **117**, 1392–1412 (2024).

48. Zhang, H., Tao, X., Fan, X., Zhang, S. & Qin, G. PpybZIP43 contributes to sucrose synthesis in pear fruits by activating PpySPS3 expression and interacts with PpySTOP1. *Physiol. Plant* **174**, e13732 (2022).

49. Shang, L. et al. A super pan-genomic landscape of rice. *Cell Res.* **32**, 878–896 (2022).

50. Chen, S. et al. Gene mining and genomics-assisted breeding empowered by the pangenome of tea plant *Camellia sinensis*. *Nat. Plants* **9**, 1986–1999 (2023).

51. Hiwasa, K. et al. Ethylene is required for both the initiation and progression of softening in pear (*Pyrus communis* L.) fruit. *J. Exp. Bot.* **54**, 771–779 (2003).

52. Paul, V., Pandey, R. & Srivastava, G. C. The fading distinctions between classical patterns of ripening in climacteric and non-climacteric fruit and the ubiquity of ethylene—an overview. *J. Food Sci. Technol.* **49**, 1–21 (2012).

53. Song, Q. et al. Diurnal down-regulation of ethylene biosynthesis mediates biomass heterosis. *Proc. Natl Acad. Sci. USA* **115**, 5606–5611 (2018).

54. Liu, C. et al. The OsWRKY72-OsAAT30/OsGSTU26 module mediates reactive oxygen species scavenging to drive heterosis for salt tolerance in hybrid rice. *J. Integr. Plant Biol.* **66**, 709–730 (2024).

55. Dombrecht, B. et al. MYC2 differentially modulates diverse jasmonate-dependent functions in *Arabidopsis*. *Plant Cell* **19**, 2225–2245 (2007).

56. Hu, D. G. et al. The regulatory module MdPUB29-MdbHLH3 connects ethylene biosynthesis with fruit quality in apple. *N. Phytol.* **221**, 1966–1982 (2019).

57. He, L. et al. DNA methylation-free *Arabidopsis* reveals crucial roles of DNA methylation in regulating gene expression and development. *Nat. Commun.* **13**, 1335 (2022).

58. Zhang, J. et al. Autotetraploid rice methylome analysis reveals methylation variation of transposable elements and their effects on gene expression. *Proc. Natl Acad. Sci. USA* **112**, E7022–E7029 (2015).

59. Shao, L. et al. Patterns of genome-wide allele-specific expression in hybrid rice and the implications on the genetic basis of heterosis. *Proc. Natl Acad. Sci. USA* **116**, 5653–5658 (2019).

60. Han, X. et al. Two haplotype-resolved, gap-free genome assemblies for *Actinidia latifolia* and *Actinidia chinensis* shed light on the regulatory mechanisms of vitamin C and sucrose metabolism in kiwifruit. *Mol. Plant* **16**, 452–470 (2023).

61. Hu, H. et al. Allele-specific expression reveals multiple paths to highland adaptation in maize. *Mol. Biol. Evol.* **39**, msac239 (2022).

62. Swetha, C. et al. Major domestication-related phenotypes in indica rice are due to loss of miRNA-mediated laccase silencing. *Plant Cell* **30**, 2649–2662 (2018).

63. Nham, N. T. et al. A transcriptome approach towards understanding the development of ripening capacity in 'Bartlett' pears (*Pyrus communis* L.). *BMC Genom.* **16**, 762 (2015).

64. Li, D. et al. The multi-omics basis of potato heterosis. *J. Integr. Plant Biol.* **64**, 671–687 (2022).

65. Ma, X. et al. An enhanced network of energy metabolism, lysine acetylation, and growth-promoting protein accumulation is associated with heterosis in elite hybrid rice. *Plant Commun.* **4**, 100560 (2023).

66. Zhang, J. et al. Extensive sequence divergence between the reference genomes of two elite indica rice varieties Zhenshan 97 and Minghui 63. *Proc. Natl Acad. Sci. USA* **113**, E5163–E5171 (2016).

67. Shen, G. et al. The regulatory network mediated by circadian clock genes is related to heterosis in rice. *J. Integr. Plant Biol.* **57**, 300–312 (2015).

68. Yang, M. et al. Genomic architecture of biomass heterosis in *Arabidopsis*. *Proc. Natl Acad. Sci. USA* **114**, 8101–8106 (2017).

69. Yang, Q. et al. ABA-responsive ABRE-BINDING FACTOR3 activates DAM3 expression to promote bud dormancy in Asian pear. *Plant Cell Environ.* **43**, 1360–1375 (2020).

70. Fu, J. et al. Genome-wide identification of *Petunia* HSF genes and potential function of PhHSF19 in benzenoid/phenylpropanoid biosynthesis. *Int. J. Mol. Sci.* **23**, 2974 (2022).

71. Zhang, N. et al. Changes to the core and flanking sequences of G-box elements lead to increases and decreases in gene expression in both native and synthetic soybean promoters. *Plant Biotechnol. J.* **17**, 724–735 (2019).

¹College of Horticulture, State Key Laboratory of Crop Genetics & Germplasm Enhancement and Utilization, Nanjing Agricultural University, Nanjing, China. ²Zhongshan Biological Breeding Laboratory, Nanjing, China. ³National Key Laboratory of Wheat Improvement, Peking University Institute of Advanced Agricultural Sciences, Shandong Laboratory of Advanced Agricultural Sciences in Weifang, Weifang, China. ⁴Institute of Forestry and Pomology, Beijing Academy of Agriculture and Forestry Sciences, Beijing, China. ⁵State Key Laboratory of Crop Biology, College of Horticulture Science and Engineering, Shandong Agricultural University, Tai'an, China. ⁶Yantai Academy of Agricultural Sciences, Yantai, China. ⁷School of Horticulture, Anhui Agricultural University, Hefei, China. ⁸These authors contributed equally: Manyi Sun, Beibei Cao, Kui Li, Jiaming Li, Jun Liu, Cheng Xue. ✉e-mail: wujun@njau.edu.cn

Methods

Samples and genome sequencing

The pear cultivars 'Dangshansuli' (DS) and 'Max Red Bartlett' (MRB) are two critical commercial cultivars grown in China and Europe, respectively. Therefore, these two cultivars were chosen for genome sequencing. We generated HiFi, Hi-C and ONT ultra-long 100-kb reads for DS and MRB. To assist in gene prediction, RNA-seq data were generated from young stems, mature leaves, young leaves, seeds, pollen, radicles and fruits at different stages with: S1: 35 d after flowering blooming (d.a.f.b.); S2: 63 d.a.f.b.; S3: 91 d.a.f.b.; S4: 120 d.a.f.b.; S5: 150 d.a.f.b. (for MRB, 135 d.a.f.b.); and S6–S8: 2, 4 and 8 d after harvest. For population analysis, 362 pear accessions were used in this study, including 137 newly collected accessions and 225 previously released accessions[10,72]. The fruit pericarp was collected from 137 pear accessions and NGS data were generated. This study provides a detailed description of the pear samples used for genome assembly and short-read sequencing (Supplementary Method 1).

T2T genome assembly

The two haplotype assemblies of DS and MRB were performed using hifiasm (v.0.19.0)[18] and Verkko software (v.1.4.1)[19] with PacBio HiFi, ONT ultra-long and Hi-C data, and HiCanu[20] with PacBio HiFi data. The Verkko assembly sequences were aligned to hifiasm assembly sequences for gap filling. If the assembly sequence can bridge the gap in the hifiasm result, the sequence will be merged to the genome sequence. HiCanu assembly sequences were used for telomeric region filling. ALLHIC (v.0.9.8) was used to anchor the contigs onto chromosomes with Hi-C data[73] and the resulting assembly was manually corrected. To check the quality of gap filling, PacBio HiFi and ONT reads were mapped to each haplotype using Minimap2 (v.2.23)[74]. To evaluate the completeness and quality of the assembly, NGS reads were mapped to each haplotype using Burrows–Wheeler alignment (BWA, v.0.7.17)[75]. SAMtools (v.1.14)[76] was then used to calculate the genome coverage and mapping rate. The LTR Assembly Index value was calculated using LTR_retriever[77]. Genome completeness was evaluated by BUSCO (v.5.2.2) using the embryophyta_odb10 database[78]. The accuracy of the genome is evaluated by calculating quasi-variance using Merqury (v.1.3)[79]. For haplotype switch error ratio estimation, we used the calc_switchErr pipeline (v.1.0)[6] (Supplementary Fig. 39 and Supplementary Method 2). Hi-C data were mapped to each haplotype of DS and MRB using HiC-Pro (v.2.7.0)[80]. The output matrix file was visualized using HiCPlotter (v.0.8.1)[81].

Repeated sequences and gene annotation

TEs in the two haplotypes of DS and MRB were annotated by combing prediction from the start and homology searches. First, RepeatScout (v.1.0.6)[82], Piler (v.1.0)[83], LTR_FINDER (v.1.07)[84] and RepeatModeler (v.2.0.4)[85] were used to build the de novo repeat sequence library and RepeatMasker (v.4.1.2)[86] was used to annotate the repeat regions based on this library. Tandem Repeats Finder (v.4.09)[87] was used to find the tandem repeats. Finally, RepeatMasker was used to retrieve the divergent TEs.

The protein-coding gene structure was predicted using a combination method involving prediction from the start, RNA-seq data prediction and homology-based prediction. The protein sequences from five species (*Arabidopsis thaliana*, *P. communis*, *Malus domestica*, *Prunus persica* and *Pyrus betulifolia*) were downloaded, and TBLASTN (v.2.2.26)[88] searches and GeneWise (v.2.4.1)[89] were used to predict the gene structure using the corresponding genomic regions of each BLAST hit. Augustus (v.2.5.5)[90], GlimmerHMM (v.3.0.1)[91] and Genscan (v.1.0)[92] were used for gene predictions from the start. RNA-seq data were aligned to the haplotype sequences using HISAT2 (v.2.0.4)[93] and the output was used for genome-guided transcript assembly using Cufflinks (v.2.1.1)[94]. De novo transcriptome read assemblies were generated using Trinity (v.2.1.1)[95]. Evidence Modeler (v.2.0.0) was used to integrate

these prediction results to predict the gene model and PASA (v.2.5.3) added to predict the UTR region[96]. The functions of protein-coding genes were identified by eggNOG-mapper (v.2.1.10)[97].

Collinearity analysis, K_a and K_s calculations

Paralogous and orthologous gene pairs were identified in syntenic blocks using JCVI (v.1.4.18)[98] with the parameter '--cscore=0.99'. The protein sequences of each gene pair were aligned using MAFFT (v.7.49)[99] and then converted into CDS alignments based on the CDS using ParaAT (v.2.0)[100]. The number of nonsynonymous substitutions per synonymous site (K_a), synonymous substitutions per synonymous site (K_s) and the K_a-to-K_s ratios were calculated for these pairs using the NGS method implemented in PAML (v.4.9b)[101].

Gene family analysis and phylogenetic reconstruction

The protein sequences from 12 species, including *P. bretschneideri* (DS), *P. communis* (MRB), *Oryza sativa*, *A. thaliana*, *Solanum lycopersicum*, *Prunus avium*, *P. persica*, *Prunus mume*, *M. domestica*, *Fragaria vesca*, *Rosa chinensis* and *Citrus sinensis* were collected (http://www.phytozome.net). OrthoMCL (v.6.18)[102] was used to identify the gene families and MUSCLE (v.5.2)[103] to perform multiple sequence alignments. A maximum-likelihood tree was constructed using RAxML (v.1.1.0)[104] with the PROTGAMMAAUTO model. The MCMCTree program in PAML[101] was applied to estimate divergence times with five calibration values chosen from the Time Tree website (http://www.timetree.org). Expansions and contractions of orthologous gene families in the 12 species were determined using CAFE (v.2.2)[105].

SNP and indel calling and SV analysis

MUMmer4 (v.4.0.0)[106] was used to compare the haplotypes and a delta-filter was further used to filter the raw alignment results with the parameters -m -i 90 -l 100. The filtered alignment results were used to detect the SNP and insertion and deletion (indel) (≤50 bp) variations using the SyRI (v.1.6.3) pipeline with default parameters[107]. Minigraph (v.0.20-r559)[108] was used to integrate the haplotypes (HapA and HapB of DS or MRB and HapA of DS and MRB) into graphs with the option -xggs. The gfatools (v.0.4-r214)[108] was used to derive the SVs from the graph with the bubble-popping algorithm. The position information, reference and nonreference paths and path length of each bubble were extracted. The SVs were further divided into indels based on the lengths of the reference and nonreference paths. The variation was considered a deletion if the reference path was longer than the nonreference path. Each haplotype sequence was mapped to the graph-based genome for SV genotyping in DS and MRB using Minigraph (v.0.20-r559)[108]. The inversions were identified using the SyRI pipeline with the default parameters[107]. For SV validation, 100 SVs with lengths ranging from 51 bp to 11.24 kb and 20 SVs with lengths ranging from 22.40 kb to 181.19 kb were randomly selected. The ONT reads were aligned to the HapA of DS using Minimap2 (v.2.23)[74] to check the SVs. The alignment results were manually checked using the Integrative Genomics Viewer (v.2.15.2)[109].

Analysis of allele-specific gene expression

The allelic gene pairs between haplotypes were identified referring to a previous method used in sugarcane[110] and tea[6]. The allele identification was based on two strategies: (1) synteny-based and (2) coordinate-based approaches. First, JCVI was used to identify the synteny region with the parameter '--cscore=0.99' and all paired genes were considered allele A and B (biallelic gene). Then, gene models that were not present in syntenic blocks were mapped to genes in the corresponding haplotype using BLAST (v.2.2.26)[88] and GMAP (v.2021-08-25)[111]. Potential alleles were considered if two genes had >50% overlap on coordinates, >80% sequence identity and the alignment length was >80% of gene sequences. Gene models with the same CDSs (allele with same CDS) were considered as a single allele. The allele with tandem and proximal duplication genes was identified with the Tandem module in JCVI using

'--tandem_Nmax 10 --percent_overlap 80'. The remaining unpaired genes were considered to be haplotype-specific genes, which were mapped to the corresponding haplotype using GMAP (v.2021-08-25)[111]. Based on the mapping result, haplotype-specific genes were further classified into three categories, including: genes with start codon loss or stop codon gain; genes truly lost; and genes with high divergence (gene with alignment length <80% gene length or gene identity <80%). If biallelic genes were in one-to-one synteny blocks, they were used for further allele-specific gene expression analysis.

RNA-seq data of DS and MRB fruit at eight stages (S1–S8) were used for allele-specific gene expression analysis. RNA-seq reads were trimmed using Trimmomatic (v.0.39)[112] and mapped against allele-aware annotated gene models using STAR (v.2.7.9)[113] software with parameters '--alignIntronMax 20000 --alignMatesGapMax 25000 --outFilterMultimapNmax 1'. StringTie (v.2.1.7)[114] software was used for FPKM estimation based on unique mapping reads. ASE was determined using DEseq2 (ref. [115]) with more than a twofold change and adjusted $P < 0.05$. Based on a previous study[116], ASE genes were further divided into three categories: (1) haplotype dominance gene pair (Hap), in which one haplotype gene was more highly expressed than its partner allele in at least one-third of the tissues and was not significantly different in other tissues; (2) Sub (sub- or neofunctionalized pairs), in which each of the two allelic genes showed a significantly high expression level in at least one tissue; and (3) other ASE genes that were classified as no-difference pairs (NoDiff). Methylation analysis followed a previous study[117] in which Bismark (v.2.2.5)[118] was used to analyze the whole-genome bisulfite sequencing data of leaf and fruit of DS and MRB and deepTools (v.3.5.1)[119] to evaluate the methylation level of the upstream 5-kb gene body and the downstream 5-kb regions.

Graph-based genome construction and SV genotyping
To construct the graph-based genome, we used the two haplotypes of DS and MRB for graph generation using Minigraph (v.0.20-r559)[108] and DS HapA was the base linear genome (Supplementary Fig. 40). The output Graphical Fragment Assembly file was used to construct a graph-based genome with the vg toolkit (v.1.46.0)[120]. We further evaluated the performance of SV genotyping using NGS data with a graph-based genome as reference (Supplementary Method 4). To compare the differences in SV genotype frequency between the Asian and European pears, short sequencing reads from the 362 pear accessions were mapped to the graph-based genome using the above graph-based SV genotyping pipeline and SVs with missing data rate <0.2 and minor allele frequency >0.05 were retained. The SV genotype frequencies of the Asian and European pear populations were calculated using VCFtools (v.0.1.16)[121]. SVs were counted twice for homozygous alleles and once for heterozygous alleles.

SNP calling
For capturing more SNP variations, we constructed a linear pangenome sequence with four haplotype sequences using the psvcp (v.1.01) pipeline[122] and 47.82-Mb nonreference sequences were added to the reference genome (DS HapA). Raw sequencing reads were trimmed using Trimmomatic (v.0.39)[112] and the clean reads were then aligned to the reference genome sequence using BWA (v.0.7.17)[75]. The alignment result was sorted and duplicate reads were marked using Picard software (https://broadinstitute.github.io/picard). The Genome Analysis Toolkit (v.4.2.3)[123] was used for SNP calling with the HaplotypeCaller method and the SNP set was further filtered using VCFtools (v.0.1.16)[121] software with the parameters '-maf 0.05 -max-missing 0.9'. Finally, SNPs with minor allele frequency >0.05 and missing data rate <0.1 were retained for further analysis.

Phylogenetic analysis and population structure study
A total of 273,985 SNPs at fourfold degenerate sites (4D SNPs) were selected for phylogenetic and population structure analyses.

A maximum-likelihood tree was constructed using IQ-TREE (v.2.2.0)[124], with 1,000 bootstrap replicates, and the best substitution model 'PMB + F + R10' was selected with the ModelFinder function. Admixture (v.1.3.0)[125] was used for population structure analysis and principal component analysis was performed with PLINK (v.1.90)[126]. The genome-wide selective sweeps were identified between AW and AC pears (including *P. pyrifolia* and *P. bretschneideri*) and EW and EC pears (*P. communis*). Using VCFtools (v.0.1.16)[121], we calculated pairwise F_{ST} (wild versus cultivated populations) and nucleotide diversity (π) in 20-kb sliding windows (10-kb step). Regions ranking in the top 5% for both F_{ST} and π were classified as selective sweeps.

Ancestral state estimation and inference of demographic history
To obtain the ancestral states, genomic resequencing data of the outgroup species hawthorn, apple and loquat were aligned against the HapA assembly of DS using BWA (v.0.7.17)[75] and ancestral alleles were called using ANGSD (v.0.940)[127] with the parameters '-doFasta 2 -doCounts 1'. To infer the demographic history, we used the stairway plot method[128,129] and bam files for each pear accession were filtered and then used for calculating the site-frequency spectrum using ANGSD[127], with the '-doSaf 1 -minMapQ 1 -minQ 20' option. A stairway plot was performed with 200 bootstraps, 3.9×10^{-9} substitutions per site per year and 7.5 years per generation.

SV GWAS
We performed GWASs for SSC, FTD, SCC and SFW in 362 accessions using 83,369 high-quality SV markers with GAPIT (v.3.4)[130] and the first three PCs and kinship matrix as covariates. The genome-wide significance thresholds (1.20×10^{-5}) were determined using a uniform threshold of $1/n$, where n is the adequate number of SVs. RNA-seq data from four pear accessions at five fruit development stages (S1–S5) were downloaded for evaluating *PybZIP6* gene expression[131] and RNA-seq data of samples with SCC were downloaded for evaluating *PyC3H18* gene expression[132]. STAR (v.2.7.9)[113] and StringTie (v.2.1.7)[114] were used to align the reads and quantify the gene expression with the same parameters mentioned earlier (DS HapA as reference).

SV association with gene expression
A total of 110 samples (consisting of 62 Asian pears and 38 European pears) (Supplementary Table 19) were collected and RNA-seq data were generated to identify the SV effect on gene expression. The FPKM values were calculated using STAR (v.2.7.9)[113] and StringTie (v.2.1.7)[114] with the same parameters as above (DS HapA as reference). Then, 21,581 genes with a mean FPKM > 1 were retained for further analysis. The PanMarker program[50] was used to identify SVs significantly associated with gene expression. SVs that overlapped with CDSs or the upstream 2-kb region were selected. All samples were classified into three groups, with groups 0, 1 and 2 representing the reference (Ref), heterozygous (Het) and alternative (Alt) genotypes, respectively. The Grubbs' test was used to remove outliers ($P < 0.05$) and Levene's test was applied to assess the equality of variances for a variable calculated for each of the two groups ($P < 0.05$). In addition, the two-sided Student's t-test was used to determine whether there was a significant difference between the means of groups 0 and 2 ($P < 0.05$) and the locus was retained if the nonoutlier values in one of the two groups was strictly larger than that in another. We also compared the difference between group 1 and group 0 (or group 2) to evaluate the effect of heterozygous SV on the difference in gene expression.

RNA extraction, gene cloning and RT–qPCR assays
Total RNA was isolated using the Plant Total RNA Isolation Kit Plus (Foregene) following the manufacturer's instructions. First-strand complementary DNA was synthesized from 2 μg of total RNA using

the TransScript One-Step gDNA Removal and cDNA Synthesis Super-Mix (TransGen), according to the manufacturer's protocol. All oligonucleotide primers used are listed in Supplementary Table 21. Amplification for gene cloning was performed as follows: an initial denaturation step at 95 °C for 3 min, followed by 35 cycles of 95 °C for 15 s, 58 °C for 15 s and 72 °C for 90 s, with a final elongation step at 72 °C for 5 min.

Dual-luciferase reporter assay

The 2-kb promoter regions (5′-upstream of the ATG start codon) of *PyACS1* from DS and MRB were cloned and inserted into the pGreen II 0800-LUC binary vector for use as reporter genes. The *PybHLH94* CDS was amplified by PCR using a pair of gene-specific primers and inserted into the pCAMBIA1300-35S-EGFP vector by homologous recombination. The recombinant plasmids constructed in the pCAMBIA1300-35S-EGFP vector and pGreen II 0800-LUC vector were transformed into the GV3101 strain of *A. tumefaciens*. *A. tumefaciens* GV3101 cells containing these vectors or the control vector were resuspended in the infiltration solution containing 10 mmol of 2-(*N*-morpholino) ethanesulfonic acid hydrate, 10 mmol of $MgCl_2$ and 0.2 μmol of acetosyringone at an optical density at 600 nm = 1.0, then cultured at 25 °C with shaking at 70 rpm for 4 h before infiltration. The *Agrobacterium* cultures containing the pGreen II 0800-LUC and the pCAMBIA1300-35S-EGFP recombinant plasmids were mixed in a 1:9 ratio and mixtures were infiltrated into 4-week-old *Nicotiana benthamiana* leaves for transient gene expression assays. The infiltrated tobacco plants were placed in the dark for 12 h and then grown under normal conditions for 48 h, after which the firefly luciferase (LUC) and *Renilla* luciferase (REN) activities were assayed using dual-luciferase assay reagents (Promega). Determination of fluorescence values was performed using a microplate reader (SpectraMax Paradigm).

Electromobility shift assay analysis

The CDSs of *PybHLH94* were cloned into the pET-32a⁺ vector to generate His fusion protein constructs (PybHLH94–His). Purification of the PybHLH94–His fusion proteins was performed using HIS Sefinose resin (Sangon Biotech). Mutated probes and unlabeled probes were used for competition in the EMSA experiments. Gene promoter sequences containing putative *cis*-element-binding sites were labeled at the 5′-end using a biotin DNA-labeling kit (Beyotime Biotechnology), as directed by the manufacturer. EMSA was performed using a chemiluminescent EMSA kit (Beyotime Biotechnology).

Construction of transgenic tomato lines

For the construction of stable transgenic lines, the full-length CDS of *PyACS1* from MRB was inserted into the pBI101 vector by homologous recombination. The recombinant plasmids were transformed into *A. tumefaciens* strain GV3101 using the freeze–thaw method. Transgenic tomato plants 'Micro-Tom' were obtained using an improvement of the method used in ref. 133 (Supplementary Method 8). Fruit firmness was quantitatively evaluated using a Brookfield CT3 Texture Analyzer following standardized protocols. Ethylene production dynamics were analyzed using a closed chamber methodology (Supplementary Method 9).

Statistical analysis

The statistical tests used are described throughout the article and in the figure legends. The χ^2 test was performed using the chisq.test function in R. The two-sided Student's *t*-test was calculated using Student's *t*-test in R. Specifically, the χ^2 test that is FDR corrected was used for comparison of SV occurrence frequency between Asian and European pears.

Reporting summary

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

Data availability

Raw genome reads and RNA-seq data of DS and MRB can be downloaded from the National Center for Biotechnology Information (NCBI) BioProject (accession no. PRJNA1237550). Raw WGS data of 137 pear accessions and RNA-seq data of 110 pear accessions generated in this study can be downloaded from the NCBI BioProject (accession nos. PRJNA989672 and PRJNA1237442). Raw WGS data of previously released pear accessions were downloaded from the NCBI BioProject (accession nos. PRJNA563813 and PRJNA381668). The genome sequences were archived in the NCBI under accession nos. PRJNA1254919, PRJNA1254920, PRJNA1254911 and PRJNA1254912. The files for genome assembly and annotation in this study are also available in the Pear Genomics Database (http://pyrusgdb.sdau.edu.cn/). All data are publicly available. Source data are provided with this paper.

Code availability

Customized code and scripts supporting this work are available via GitHub at https://github.com/bioinformatic1996512/DS-MRB-genome and Zenodo at https://doi.org/10.5281/zenodo.15113930 (ref. 134).

References

72. Zhang, M. Y. et al. Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. *Nat. Commun.* **12**, 1141 (2021).
73. Zhang, X. T. et al. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat. Plants* **5**, 833–845 (2019).
74. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
75. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
76. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
77. Ou, S. J. et al. Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018).
78. Manni, M. et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
79. Rhie, A. et al. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
80. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
81. Akdemir, K. C. & Chin, L. HiCPlotter integrates genomic data with interaction matrices. *Genome Biol.* **16**, 198 (2015).
82. Price, A. L. et al. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, I351–I358 (2005).
83. Edgar, R. C. & Myers, E. W. PILER: identification and classification of genomic repeats. *Bioinformatics* **21**, I152–I158 (2005).
84. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
85. Flynn, J. M. et al. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl Acad. Sci. USA* **117**, 9451–9457 (2020).
86. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* **25**, 4.10.1–4.10.14 (2009).
87. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
88. Altschul, S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
89. Birney, E. et al. GeneWise and genomewise. *Genome Res.* **14**, 988–995 (2004).

90. Urnov, F. D. et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).

91. Majoros, W. H. et al. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).

92. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94 (1997).

93. Kim, D. et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).

94. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).

95. Haas, B. J. et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).

96. Haas, B. J. et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).

97. Cantalapiedra, C. P. et al. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).

98. Tang, H. et al. JCVI: A versatile toolkit for comparative genomics analysis. *Imeta* **12**, e211 (2024).

99. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

100. Zhang, Z. et al. ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments. *Biochem. Biophys. Res. Commun.* **419**, 779–781 (2012).

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

102. Li, L. et al. OrthoMCL: identification of ortholog groups for eukaryotic genomes. *Genome Res.* **13**, 2178–2189 (2003).

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

104. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

105. De Bie, T. et al. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).

106. Marcais, G. et al. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.* **14**, e1005944 (2018).

107. Goel, M. et al. SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* **20**, 277 (2019).

108. Li, H. et al. The design and construction of reference pangenome graphs with minigraph. *Genome Biol.* **21**, 265 (2020).

109. Thorvaldsdottir, H. et al. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

110. Zhang, J. et al. Allele-defined genome of the autopolyploid sugarcane *Saccharum spontaneum* L. *Nat. Genet.* **50**, 1565–1573 (2018).

111. Wu, T. D. & Watanabe, C. K. GMAP: a genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).

112. Bolger, A. M. et al. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

113. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

114. Pertea, M. et al. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295 (2015).

115. Love, M. I. et al. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

116. Lan, X. & Pritchard, J. K. Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science* **352**, 1009–1013 (2016).

117. Song, B. et al. Increased DNA methylation contributes to the early ripening of pear fruits during domestication and improvement. *Genome Biol.* **25**, 87 (2024).

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

119. Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).

120. Hickey, G. et al. Genotyping structural variants in pangenome graphs using the vg toolkit. *Genome Biol.* **21**, 35 (2020).

121. Danecek, P. et al. The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

122. Wang, J. et al. A pangenome analysis pipeline provides insights into functional gene identification in rice. *Genome Biol.* **24**, 19 (2023).

123. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).

124. Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).

125. Alexander, D. H. et al. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).

126. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).

127. Korneliussen, T. S. et al. ANGSD: analysis of next generation sequencing data. *BMC Bioinf.* **15**, 356 (2014).

128. Liu, X. & Fu, Y. X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47**, 555–559 (2015).

129. Liu, X. & Fu, Y. X. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biol.* **21**, 280 (2020).

130. Wang, J. & Zhang, Z. GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinform.* **19**, 629–640 (2021).

131. Zhang, M. Y. et al. Distinct transcriptome profiles reveal gene expression patterns during fruit development and maturation in five main cultivated species of pear (Pyrus L.). *Sci. Rep.* **6**, 28130 (2016).

132. Wang, R. Z. et al. A systems genetics approach reveals PbrNSC as a regulator of lignin and cellulose biosynthesis in stone cells of pear fruit. *Genome Biol.* **22**, 313 (2021).

133. Lin, T. B. Resveratrol production of tomato transformed with a vitis stilbene synthase gene. *Life Sci. Res* **7**, 262–266 (2003).

134. Sun, M. Bioinformatic1996512/DS-MRB-genome: code for DS and MRB genome (DSMRBv1.0). *Zenodo* https://doi.org/10.5281/zenodo.15113931 (2025).

Acknowledgements

Author contributions

Competing interests

Additional information

Corresponding author(s): Jun Wu

Last updated by author(s): Mar 19, 2025

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	Sequencing platforms used to generate the raw data are listed as followed: Pacbio Sequel II, OXFORD_NANOPORE PromethION, Illumina NovaSeq 6000.
Data analysis	The software used in this manuscript include Hifiasm v0.16, Verkko v1.4.1, HiCanu v2.2, ALLHIC v0.9.8, Minimap2 v2.23, SAMtools v1.14, BWA v0.7.17, Merqury v1.3BUSCO v5.2.2, LTR_retreiver v2.9.6, PEPPER-Margin-DeepVariant r0.8, calc_switchErr v1.0, WhatsHap v2.1, Nucmer v4.0.0rc1, HiC-Pro v2.7.0, HiCPlotter v0.8.1, RepeatModeler v2.0.4, RepeatMasker v4.1.2, Tandem Repeats Finder v4.09, TBLASTN v2.2.26, GeneWise v2.4.1, Augustus v2.5.5, GlimmerHMM v3.0.1, Genscan v1.0, HISAT2 v2.0.4, Cufflinks v2.1.1, Trinity v2.1.1, Evidence Modeler v2.0.0, PASA v2.5.3, eggNOG-mapper v2.1.10, cd-hit v4.8.1, HMMER v3.3.2, JCVI v1.4.18, MAFFT v7.49, ParaAT v2.0, PAML v4.9b, MUSCLE v5.2, OrthoMCL v6.18, RAxML v1.1.0, CAFE v2.2, HMMER v3.1b1, IQ-TREE v2.2.0, MUMmer4 v4.0.0, Minigraph v0.20-r559, SyRI v1.6.3, gfatools v0.4-r214, Integrative Genomics Viewer v2.15.2, GMAP v2021-08-25, Trimmomatic v0.39, STAR v2.7.9, StringTie v2.1.7, Bismark v2.2.5, deepTools v3.5.1, vg toolkit v1.46.0, manta v1.6, delly v0.76, lumpy v0.2.13, SURVIVOR v1.0.7, VCFtools v0.1.16, GENOME Analysis Toolkit v4.2.3, psvcp v1.01, PLINK v1.90, Admixture v1.3.0, ANGSD v0.940, GAPIT v3.4, TBtools v2.008. Custom code and scripts supporting this work are available at github.com (https://github.com/bioinformatic1996512/DS-MRB-genome) and and Zenodo at https://doi.org/10.5281/zenodo.15113930

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

Data

Raw genome reads, and RNA-seq data of DS and MRB can be downloaded from NCBI BioProject (PRJNA1237550). Raw WGS data of 137 pear accessions and RNA-seq data of 110 pear accessions generated in this study can be downloaded from NCBI BioProject (PRJNA989672 and PRJNA1237442). Raw WGS data of previously released pear accessions were downloaded from NCBI BioProject (PRJNA563813 and PRJNA381668). The genome sequence were archived in NCBI under accession number PRJNA1254919, PRJNA1254920, PRJNA1254911, PRJNA1254912. The files for genome assembly and annotation in this study are also available in Pear Genomics Database (http://pyrusgdb.sdau.edu.cn/). All data are publicly available. Source data are provided with this paper.

Research involving human participants, their data, or biological material

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

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

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No predefined statistical methods were used for experimental sample size determination. The study utilized 362 pear accessions spanning major cultivated species (Asian and European pears), sourced from globally representative regions, including China, Japan, Europe (Germany, France), and the United States. These accessions comprehensively capture the genetic diversity of pear. 'Dangshansuli' (Asian pear) and 'Max Red Bartlett' (European pear) were selected for de novo genome assembly and comparative analyses due to their agronomic and geographical significance. Comprehensive metadata and supporting results are provided in Supplementary Dataset and Figures.
Data exclusions	No data were excluded from the analysis in the article.
Replication	The qRT-PCR experiments were repeated independently for three times with similar results. Three independent T1 tomato lines over-expressing PyACS1 were used for analysis. For all experiments, at least three replications were repeated and succeed.
Randomization	For each accessions used for sequencing, the sampling process for genome DNA/RNA sequencing was randomly conducted.
Blinding	Blinding is not necessary for genome sequencing and assembly, since the investigators know the pear accessions that they were handing.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

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

Methods

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