Plant Communications
Research article

CellPress
Partner Journal

Gap-free genome assemblies of two *Pyrus bretschneideri* cultivars and GWAS analyses identify a CCCH zinc finger protein as a key regulator of stone cell formation in pear fruit

Yunpeng Cao[1,14,*], Xiaofeng Feng[2,14], Baopeng Ding[10], Heqiang Huo[11], Muhammad Abdullah[12], Jiayi Hong[2], Lan Jiang[4], Han Wang[5], Risheng Li[7], Yongping Cai[2], Xiaoxu Li[3], Zhichao Xia[6,*], Rajeev K. Varshney[13,*], Haifei Hu[7,*], Mengfei Lin[8,*] and Fei Shen[9,*]

[1]CAS Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, China

[2]College of Life Sciences, Anhui Agricultural University, Hefei 230036, China

[3]Beijing Life Science Academy, Beijing 102209, China

[4]Key Laboratory of Non-coding RNA Transformation Research of Anhui Higher Education Institution, Yijishan Hospital of Wannan Medical College, Wuhu 241000, China

[5]Key Laboratory of Horticultural Crop Germplasm Innovation and Utilization (Co-construction by Ministry and Province), Institute of Horticulture, Anhui Academy of Agricultural Sciences, Hefei 230000, China

[6]School of Forestry & Landscape Architecture, Anhui Agricultural University, Hefei 230036, China

[7]Rice Research Institute, Guangdong Academy of Agricultural Sciences & Key Laboratory of Genetics and Breeding of High-Quality Rice in Southern China (Co-construction by Ministry and Province), Ministry of Agriculture and Rural Affairs & Guangdong Key Laboratory of New Technology in Rice Breeding & Guangdong Rice Engineering Laboratory, Guangzhou 510640, China

[8]Jiangxi Provincial Key Laboratory of Plantation and High Valued Utilization of Specialty Fruit Tree and Tea, Institute of Biological Resources, Jiangxi Academy of Sciences, Nanchang, Jiangxi, China

[9]Institute of Biotechnology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China

[10]Engineering Research Center of Coal-Based Ecological Carbon Sequestration Technology of the Ministry of Education and Key Laboratory of National Forest and Grass Administration for the Application of Graphene in Forestry, Shanxi Datong University, Datong 037009, China

[11]Department of Environmental Horticulture, Mid-Florida Research and Education Center, University of Florida, Apopka, FL 32703, USA

[12]Queensland Alliance for Agriculture and Food Innovation, University of Queensland, 7 Brisbane, Brisbane, QLD, Australia

[13]Centre for Crop and Food Innovation, WA State Agricultural Biotechnology Centre, Food Futures Institute, Murdoch University, Murdoch, WA, Australia

[14]These authors contributed equally to this article.

***Correspondence: Yunpeng Cao (xfcypeng@126.com), Zhichao Xia (zhichaoxia0623@163.com), Rajeev K. Varshney (rajeev.varshney@murdoch.edu.au), Haifei Hu (huhaifei@gdaas.cn), Mengfei Lin (linmengfeilixi@163.com), Fei Shen (shenf1028@gmail.com)**

https://doi.org/10.1016/j.xplc.2024.101238

ABSTRACT

The Chinese white pear (*Pyrus bretschneideri*) is an economically significant fruit crop worldwide. Previous versions of the *P. bretschneideri* genome assembly contain numerous gaps and unanchored genetic regions. Here, we generated two high-quality, gap-free genome assemblies for 'Dangshansu' (DS; 503.92 Mb) and 'Lianglizaosu' (ZS; 509.01 Mb), each anchored to 17 chromosomes, achieving a benchmarking universal single-copy ortholog completeness score of nearly 99.0%. Our genome-wide association studies explored the associations between genetic variations and stone cell traits, revealing a significant association peak on DS chromosome 3 and identifying a novel non-tandem CCCH-type zinc finger gene, designated *PbdsZF*. Through genetic transformation, we verified the pivotal role of *PbdsZF* in regulation of both lignin biosynthesis and stone cell formation, as it transcriptionally activates multiple genes involved in these processes. By binding to the CT-rich motifs CT1 (CTTTTTTCT) and CT2 (CTCTTTTT), PbdsZF significantly influences the transcription of genes essential for lignin production, underscoring its regulatory importance in plant lignin metabolism. Our study illuminates the complex biology of fruit development and delineates the gene regulatory networks that influence stone cell and lignocellulose

formation, thereby enriching genetic resources and laying the groundwork for the molecular breeding of perennial trees.

Key words: *Pyrus bretschneideri*, gap-free genome, stone cell formation, GWAS, transcriptional regulation

Cao Y., Feng X., Ding B., Huo H., Abdullah M., Hong J., Jiang L., Wang H., Li R., Cai Y., Li X., Xia Z., Varshney R.K., Hu H., Lin M., and Shen F. (2025). Gap-free genome assemblies of two *Pyrus bretschneideri* cultivars and GWAS analyses identify a CCCH zinc finger protein as a key regulator of stone cell formation in pear fruit. Plant Comm. **6**, 101238.

INTRODUCTION

Pear, the third most significant temperate fruit species after grape and apple, is classified within the Pomoideae subfamily of the Rosaceae family. It encompasses over 22 species and more than 5000 accessions, each exhibiting distinct physiological, morphological, and adaptive traits (Li et al., 2022a, 2022b; Song et al., 2024a). *Pyrus bretschneideri* cv. Dangshansu (DS), a pear variety native to China and the world's most commercially important Asiatic pear cultivar, has been cultivated for over 500 years for its significant nutritional and medicinal benefits, with an annual production exceeding 4 million tons in China (Wu et al., 2013). The DS pear variety is characterized by a relatively high stone cell content, a critical determinant of fruit quality that significantly affects not only nutrient and sucrose levels but also flesh chewiness and hardness (Cai et al., 2010; Yan et al., 2014; Zhang et al., 2017; Cheng et al., 2020). The 'Lianglizaosu'(ZS) variety, derived from a natural bud sport of DS pear, has been observed over several years for trait stability (Zhang et al., 2017; Cheng et al., 2020). The stone cell content in ZS fruit is notably lower than that in DS, indicating that ZS is an ideal candidate for exploring the developmental mechanisms of pear stone cells.

Stone cell content significantly influences pear fruit flavor (Yan et al., 2014; Xue et al., 2019; Zhang et al., 2021a). Lignin synthesis, a key factor in stone cell development, directly affects the quality and market value of fruits (Yan et al., 2014; Zhang et al., 2021a). Lignin is composed of different phenylpropanoid components, including monolignols like p-coumaryl, coniferyl, and sinapyl alcohols, which are characterized by various levels of methoxylation. These monolignols are incorporated into the lignin structure as p-hydroxyphenyl (H), syringyl (S), and guaiacyl (G) units, respectively. The formation of monolignols involves a complex process with 11 enzyme-driven steps (Carocha et al., 2015; Cao et al., 2019), starting with the conversion of phenylalanine into simpler components. This process engages the broad phenylpropanoid pathway, culminating in the creation of hydroxycinnamoyl CoA derivatives. Key enzymes such as phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), and 4-coumarate:CoA ligase (4CL) play vital roles in these initial stages (Cao et al., 2023). Although hydroxycinnamoyl CoA esters are foundational for the predominant phenylpropanoid, lignin, they also serve as precursors for diverse biochemical substances like flavonoids, anthocyanins, and condensed tannins. The presence and variety of these substances are influenced by the plant species, specific cell type, and external environmental conditions. The generation of monolignols requires a series of biochemical transformations, including successive hydroxylation and O-methylation of the aromatic rings, facilitated by a set of enzymes: shikimate O-hydroxycinnamoyltransferase (HCT), caffeoyl shikimate esterase (CSE), p-coumarate 3-hydroxylase (C3H), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5-hydroxylase (F5H), and caffeate/5-hydroxyferulate O-methyltransferase (COMT). The process concludes with transformation of the carboxyl side chain into an alcohol, a reaction sequentially catalyzed by cinnamoyl CoA reductase (CCR) and cinnamyl alcohol dehydrogenase (CAD), which have been identified as crucial enzymes in the monolignol biosynthesis pathway.

Genome sequencing of the *P. bretschneideri* 'Suli' pear, completed in 2013, significantly advanced functional genomics and subsequently guided pear breeding efforts (Wu et al., 2013). However, technical limitations of next generation sequencing led to numerous gaps in the genome assembly, causing a loss of genetic information and constraining our understanding of pear genome structure and evolution. Recent advances in long-read sequencing technologies, such as PacBio HiFi sequencing, coupled with Hi-C sequencing methods, have greatly improved genome assembly (Jiang et al., 2025). This approach now facilitates the straightforward generation of gap-free, high-quality, chromosome-scale genome assemblies in diverse plant species. For example, Shi et al. (2023) used HiFi sequencing technology to produce the first gap-free genome of *Vitis vinifera* (Shi et al., 2023).

Here, we selected two *P. bretschneideri* varieties, 'Dangshansu' (DS) and 'Lianglizaosu' (ZS), to investigate the metabolic alterations and pivotal genes associated with the development of stone cells in pears. We successfully generated gap-free genomes of DS and ZS, each encompassing all 17 chromosomes as contiguous sequences with an N50 of 29.29 Mb and 29.41 Mb, respectively. Using this genomic data, we identified a core set of genes associated with fruit lignification in pears. Genome-wide association study (GWAS) analysis was then used to pinpoint candidate genes implicated in fruit lignin synthesis. One newly identified gene, *PbdsZF*, was found to significantly modulate lignin content and influence stone cell formation in pear fruits, and its role was confirmed through experiments in transgenic pear fruits. We constructed an interaction network involving *PbdsZF* and core fruit lignification-related genes, aiming to deepen our understanding of their regulatory mechanisms. The genomic resources and analyses presented in this study offer a vital foundation for comprehending the evolution of the Rosaceae family and provide crucial insights for future molecular breeding of pears and advancing our understanding of fruit biology.

RESULTS

Gap-free genome assemblies for two *P. bretschneideri* cultivars

To produce high-quality genome assemblies, we used GenomeScope2 to estimate the genome sizes of the

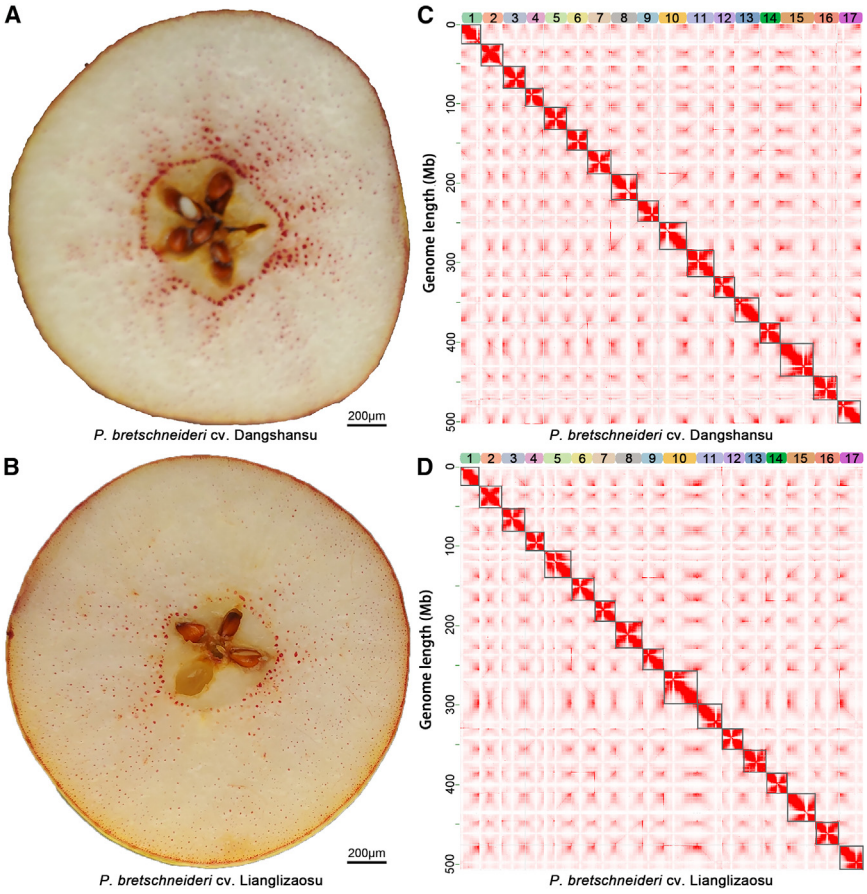

Figure 1. Genome assemblies and annotations of DS and ZS.
(A) Phenotypic characteristics of 'Dangshansu' (DS) fruit.
(B) Phenotypic characteristics of 'Lianglizaosu' (ZS) fruit.
(C) Heatmap showing genomic interactions among the DS chromosomes. Each gray box represents a chromosome. The x axis shows the chromosome number, and the y axis represents the genome size. Chr, chromosome.
(D) Heatmap showing genomic interactions among the ZS chromosomes. Each gray box represents a chromosome. The x axis shows the chromosome number, and the y axis represents the genome size. Chr, chromosome.

P. bretschneideri cultivars DS (Figure 1A) and ZS (Figure 1B). The genome size of DS was estimated to be approximately 523 Mb, with a heterozygosity rate of 1.57% (Supplemental Figure 1), and ZS was estimated to have a genome size of approximately 505 Mb, with a heterozygosity rate of 1.59% (Supplemental Figure 2). The preliminary DS genome assembly combined a total of 28.08 Gb (~56× coverage) of HiFi reads with 52.07 Gb (~100× coverage) of Hi-C data, processed using Genome Puzzle Master (GPM) (Supplemental Table 1). Using HapHiC software (version 1.0.6), the contigs from the HiFi genome assembly were anchored to 17 chromosomes, exhibiting strong interaction signals between these chromosomes, which attest to the high quality of the Hi-C-assisted assembly (Figure 1). The DS assembly contained only one gap (Supplemental Figure 3), and the ZS assembly contained only three gaps (Supplemental Figure 4). The different HiFi-contig assemblies and reads were then first integrated into this genome using quarTeT (version 1.2.1), TGS-GapCloser (version 1.2.0), and LR_Gapcloser (version 1.0). The remaining gaps were finally filled using a novel cyclic elongation process called DEGAP, successfully closing all gaps in the assembly. This resulted in a complete 503.92 Mb gap-free DS pear genome with a scaffold N50 of 29.29 Mb (Figure 1C). Using plant-specific telomere repeats (3′-TTTAGGG/5′-CCCTAAA) as queries, we identified 31 telomeres on the terminals of the 17 chromosomes in the DS pear genome. The ZS genome was assembled in a similar manner, combining 27.17 Gb of HiFi reads (~53× coverage) with 49.16 Gb of Hi-C data (~98× coverage), following the same protocol. This assembly produced a 509.01-Mb genome distributed across 17 gap-free chromosomes (Figure 1D), with a scaffold N50 of 29.41 Mb. In the ZS pear genome, 31 telomeres were identified among 17 chromosomes. The chromosome orientations and IDs of the DS and ZS genomes were aligned to match the gap-free *Pyrus pyrifolia* cv. Yunhong standards (Sun et al., 2023).

The completeness and accuracy of these two assemblies were evaluated through various methods. Specifically, we used 1614 benchmarking universal single-copy orthologs (BUSCOs) to assess the genome integrity. Our evaluation revealed that 99.2% of the BUSCO sequences were complete in the DS genome, and 98.9% were complete in the ZS genome, indicating a high level of completeness in genic regions (Supplemental Table 1). Merqury, a tool that assesses genome assembly quality from k-mer spectra, assigned average quality values (QVs) of 63.97 to the DS genome and 67.49 to the ZS genome with the HiFi data, indicating that the base error rate of both genomes was less than 10^{-6}. We also annotated long terminal repeat (LTR) elements within the genome using the LTR assembly index (LAI). The DS genome achieved an LAI value of 20.53, and the ZS genome reached 21.63, both exceeding the threshold of 20. This surpasses the criteria established for "gold quality" genomes as reported previously (Ou et al., 2018b). Collectively, these findings confirm the high accuracy, continuity, and completeness of the gap-free genome assemblies for the two pear species.

P. bretschneideri benefits from a high level of heterozygosity. With an estimated genomic heterozygosity of approximately 1.5% (Supplemental Figures 1 and 2), we were able to independently assemble and phase the two haplotypes, as a single 10-kb HiFi read contained over 150 heterozygous sites. *De novo* assemblies of the two haplotypes for DS and ZS were generated in addition to the primary assemblies, made possible by sufficient linkage information derived from heterozygous loci in the reads. Finally, using Hi-C data, we successfully phased the DS genome into two distinct haplotypes (Supplemental Figure 5), each consisting of 17 chromosomes: DS_Hap1

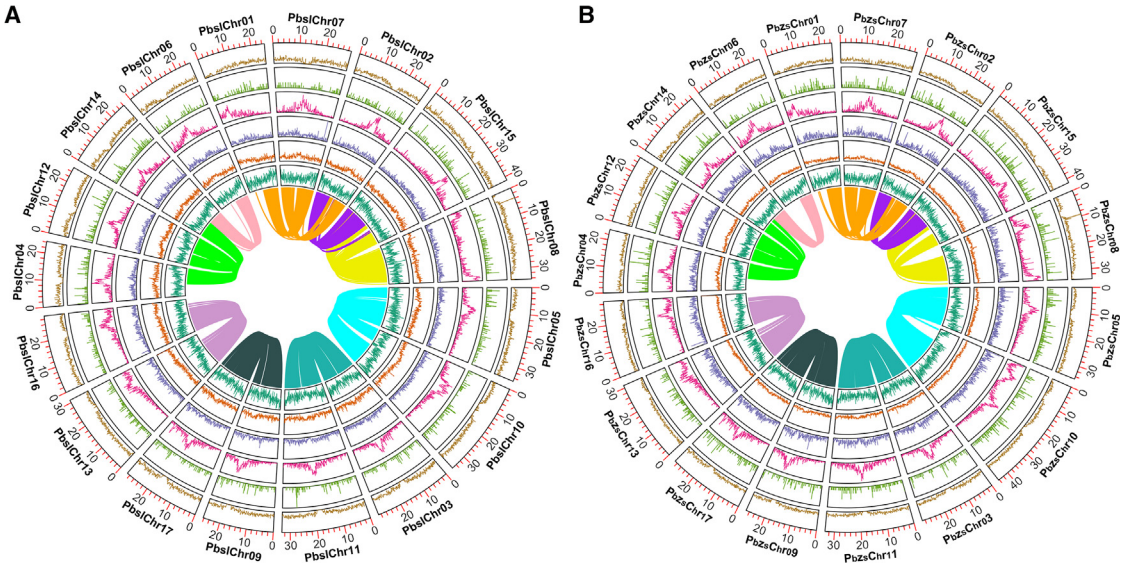

Figure 2. Genome assembly and annotation landscape of two pear cultivars

(A and B) Distribution of basic genomic elements in the DS **(A)** and ZS **(B)** genomes. Tracks (from outer to inner): TIR transposon density, LINE element density, Gypsy LTR retrotransposon density, Copia LTR retrotransposon density, GC content density, and gene density. Colored segments illustrate the genomic architecture consistent with that of the Rosaceae ancestor.

(~503.12 Mb) with a scaffold N50 of 29.10 Mb, and DS_Hap2 (~501.28 Mb) with an N50 of 29.00 Mb (Supplemental Table 2). The ZS genome was also resolved into two haplotypes (Supplemental Figure 6): ZS_Hap1 (~501.09 Mb) with a scaffold N50 of 29.13 Mb and ZS_Hap2 (~501.19 Mb) with an N50 of 29.07 Mb. Notably, the DS genome assembly exhibited just two gaps across two pseudomolecules, whereas 32 phased chromosomes were entirely gap-free (Supplemental Table 2). By contrast, the ZS haplotype assembly contained 11 gaps distributed across seven pseudomolecules.

Comparative analysis of two *P. bretschneideri* cultivar genome assemblies with a previously published *P. bretschneideri* genome assembly and the telomere-to-telomere *P. pyrifolia* genome assembly

Compared with the previously published *P. bretschneideri* 'Suli' genome (Wu et al., 2013), the DS and ZS genomes exhibited significant improvements in accuracy, continuity, and completeness, achieving a quality level comparable to that of the *P. pyrifolia* genome (Sun et al., 2023). The N50 scaffold sizes of the DS (29.29 Mb) and ZS (29.41 Mb) genomes were comparable to those of the *P. pyrifolia* genome (29.26 Mb) at the telomere-to-telomere level and significantly larger than those of the previously published *P. bretschneideri* 'Suli' genome (540.8 kb). However, the major improvement was the absence of gaps, 208 832 of which were found in the previously published *P. bretschneideri* 'Suli' genome (Supplemental Table 2). In our study, 31 telomeres were identified on the 17 DS chromosomes, and 31 telomeres were assembled on the 17 ZS chromosomes; by contrast, no telomeres were captured in the previously published *P. bretschneideri* 'Suli' genome. BUSCO values indicated that the DS genome had a gene-structure annotation completeness of 99.00%, and the ZS genome had a slightly higher completeness of 99.10%, both exceeding that of the previously published *P. bretschneideri* 'Suli' genome (83.7%) (Wu et al., 2013). Furthermore, Merqury analysis showed that the QVs of DS (63.97) and ZS (67.49) were approximately 1.3 times higher than that of *P. pyrifolia* (49.21) (Supplemental Table 2), confirming that the DS and ZS genomes had superior quality and higher accuracy. These results indicate that both genomes we assembled are high-quality *P. bretschneideri* reference genomes, with a level of quality comparable to that of the *P. pyrifolia* genome (Sun et al., 2023).

Genome annotation

Genomic analysis revealed a substantial number of transposable elements (TEs) within the DS and ZS genomes. These TEs accounted for approximately 48.68% (259.01 Mb) of the DS genome and 47.30% (250.14 Mb) of the ZS genome. Within the DS genome, Gypsy (12.75%), Copia (7.38%), and unknown LTR elements (12.79%) were the most abundant TE categories (Figure 2; Supplemental Tables 1 and 3). Similar trends were observed in the ZS genome, where Gypsy (12.80%), Copia (6.95%), and unknown LTR elements (12.20%) were most prevalent (Figure 2; Supplemental Tables 1 and 4).

For gene structure annotation, the repeat-masked genomes of DS and ZS were independently analyzed using a combination of *de novo*, homolog-based, and transcript-based prediction methods implemented in BRAKER3. This analysis identified 56 171 protein-coding genes in the DS genome and 56 027 in the ZS genome, covering 99.0% and 99.1% of the embryophyta_odb10 BUSCO gene set (1614 core plant conserved genes), respectively. In addition, an estimated 65 799 and 65 900 transcripts were predicted from the entire gene pools of DS and ZS, respectively, with an average of 1.17 and 1.18 splice variants per gene. Functional annotation of these genes using eggNOG-mapper revealed that 87.72% of the genes in the DS genome and 87.40% of those in the ZS genome showed homology to known proteins or functional domains in other species

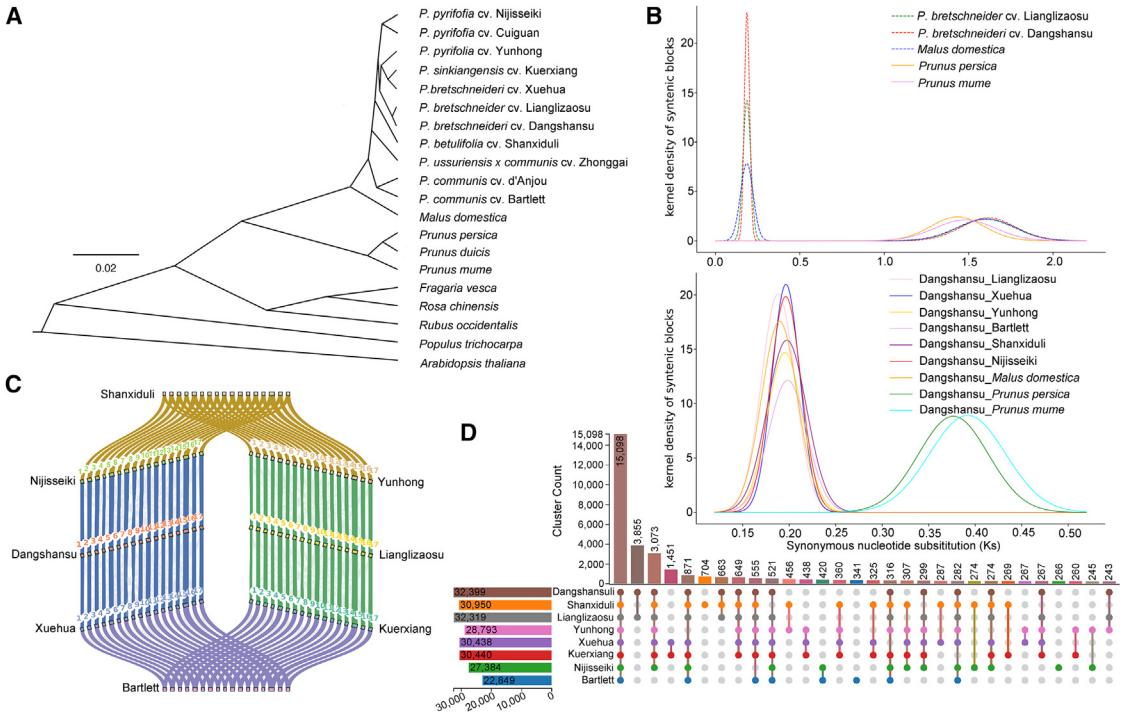

Figure 3. Comparative genomics, evolutionary events, and synonymous nucleotide substitutions per synonymous site (Ks) density analysis of pear.

(A) Phylogenetic tree illustrating divergence times among pear lineages.

(B) Distribution of Ks values for collinear genes.

(C) Intragenomic synteny of different pear varieties.

(D) UpSet plot showing the distribution of orthologous gene clusters across analyzed species. The plot categorizes genes into unique and commonly shared clusters among the various species. The top 30 cluster counts were used to generate the UpSet plot.

(Supplemental Tables 5 and 6). Collectively, these findings demonstrate the near-completeness and high fidelity of the gene models in our two assembled genomes.

Genome evolutionary analyses

OrthoFinder assigned 806 641 genes from 20 studied species and cultivars (DS, ZS, *Fragaria vesca,* Nijisseiki, Cuiguan, Xuehua, Kuerxiang, Yunhong, Shanxiduli, Zhonggai, d'Anjou, Bartlett, *Rosa chinensis*, *Rubus occidentalis*, *Malus domestica*, *Prunus persica*, *Prunus dulcis*, *Prunus mume*, *Populus trichocarpa*, and *Arabidopsis thaliana*) to 47 137 orthogroups (OGs), representing 95.18% of all genes analyzed. The G50 value is the number of genes in an OG such that at least 50% of the genes are in an OG of that size or larger, whereas the O50 denotes the minimum number of OGs needed to reach G50. In this case, 50% of the genes were included in the largest 7867 orthogroups (O50 = 7867), with these orthogroups containing 33 or more genes (yielding a G50 value of 33). Across all analyzed species, 8418 orthogroups were identified, 205 of which consisted exclusively of single-copy genes. A maximum-likelihood (ML) tree indicated that the *Malus* and *Pyrus* lineages were sister clades (Figure 3A), with DS showing the closest relationship to ZS (Figure 3A).

A Ks value analysis revealed a shared ancient whole-genome duplication (WGD) event across all Rosaceae species, with a peak Ks value of approximately 1.6 (Figure 3B and Supplemental Figures 7–12). This peak also coincided with the whole-genome triplication event (γ event) common to core eudicots. In addition, a more recent WGD event specific to the Maloideae subfamily was identified with a peak Ks value of approximately 0.19 (Figure 3B), indicating a unique evolutionary event within this subfamily. When eight related pear cultivars (DS, ZS, Kuerxiang, Shanxiduli, Nijisseiki, Xuehua, Bartlett, and Yunhong) were examined, significant collinearity was observed, suggesting a highly conserved genomic structure. For example, 2163 collinear blocks were identified between DS and Nijisseiki, 1837 between DS and Xuehua, 1407 between ZS and Yunhong, 1851 between ZS and Kuerxiang, and 2745 between DS and ZS (Figure 3C; Supplemental Table 7). This extensive collinearity underscores the genetic stability and shared evolutionary paths of these cultivars. The analysis also identified a conserved set of 15 098 genes across the cultivars, likely representing the core genome of pear species (Figure 3D).

Identification of bona fide genes encoding the core fruit lignification toolbox

The phenylpropanoid pathway, crucial for monolignol synthesis, involves key gene families like *4CL*, *COMT*, *CCR*, *CCoAOMT*, and *CAD*. These families are part of larger superfamilies, complicating genome-wide research and leading to frequent misannotations. To minimize these annotation errors, we used a comprehensive literature review to accurately identify genes within these 11 families that were experimentally verified as encoding bona

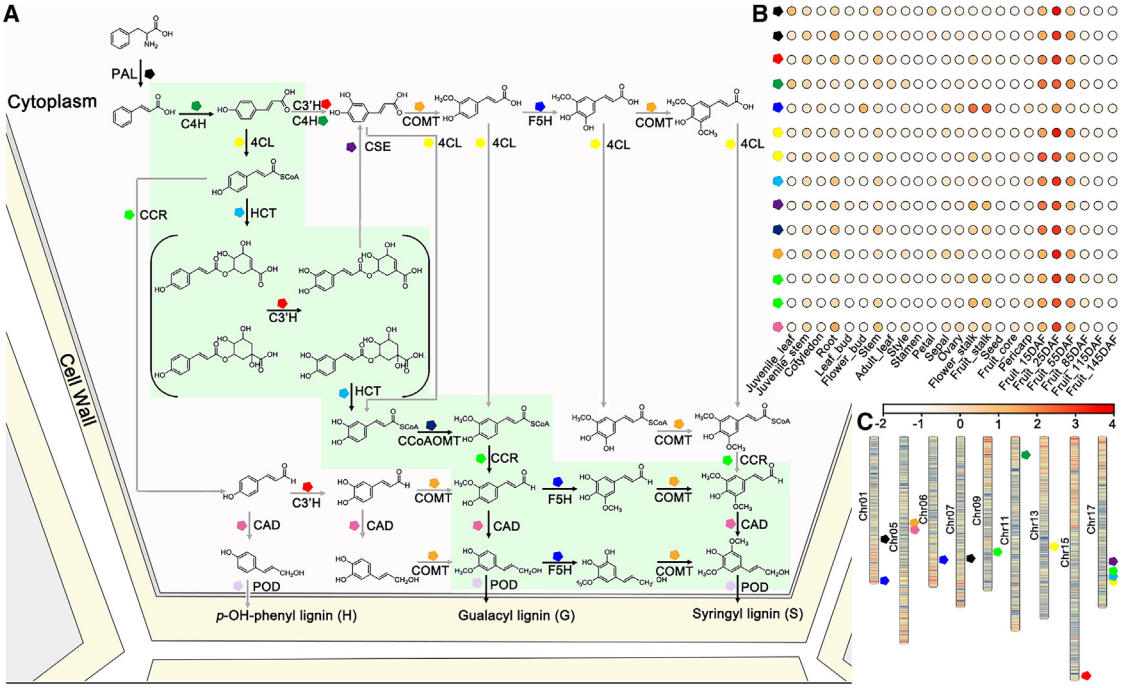

Figure 4. The core fruit lignification toolbox in the DS genome.

(A) Biosynthetic pathway diagram adapted from the framework established by Humphreys and Chapple (2002), with refinement to incorporate the recent advances reported in Vanholme et al. (2013) and Carocha et al. (2015).

(B) Heatmaps displaying the RNA-seq-derived transcript accumulation patterns of 14 genes in the core fruit lignification toolbox of the DS genome.

(C) Chromosomal distribution of 14 genes in the core fruit lignification toolbox of the DS genome.

fide enzymes (Figure 4A). This verification was supported by direct analysis of biochemical activity or genetic studies, including both forward and reverse approaches, as outlined in Supplemental Table 8. This meticulous approach helped us pinpoint 164 genes confirmed to encode functional enzymes from various plant species for phylogenetic analysis. We then compared these 164 genes with phenylpropanoid and lignin-associated genes from *P. trichocarpa*, *Vitis vinifera*, *A. thaliana*, and *Nicotiana tabacum*, along with 1845 predicted phenylpropanoid genes from the *Pyrus* genomes (Supplemental Table 9). Through phylogenetic analysis (Supplemental Figures 13–23), we established bona fide clades for each gene family, enabling us to identify a refined group of 640 bona fide enzymes in these *Pyrus* genomes (Supplemental Table 10; Supplemental Figure 24).

We focused on the DS genome to examine the expansion patterns of gene families related to monolignol synthesis. This analysis enabled us to infer their evolutionary trajectories across *Pyrus* genomes. A total of 63 bona fide enzymes were identified in the DS genome, including five PALs, five C4Hs, four C3Hs, 17 F5Hs, four 4CLs, two HCTs, two CSEs, three CCoAOMTs, 14 COMTs, two CCRs, and five CADs (Supplemental Table 10; Supplemental Figure 24). Further analysis revealed that several multigene families, such as *F5H*, *COMT*, *PAL*, and *C4H*, had expanded through tandem duplication or WGD (Supplemental Figures 25 and S26). Analysis of expression patterns across these gene families indicated that multiple coding genes were often associated with a single enzyme function, suggesting their complementary or redundant roles in the evolutionary development of the *Pyrus* species (Supplemental Figures 27–37).

To identify the key genes involved in lignification during fruit development, we integrated comparative phylogenetic analysis with expression profiling of individual genes. Specifically, we focused on 63 validated genes from 11 previously identified bona fide gene families. This combined analysis revealed a critical set of 14 genes (*PbdsPAL1*, *PbdsPAL4*, *PbdsC3H4*, *PbdsC4H5*, *PbdsF5H3*, *Pbds4CL2*, *Pbds4CL4*, *PbdsHCT2*, *PbdsCSE2*, *PbdsCCoAOMT1*, *PbdsCOMT8*, *PbdsCCR1*, *PbdsCCR2*, and *PbdsCAD1*) as pivotal for the lignification pathway during fruit maturation (Supplemental Table 11; Figure 4B and 4C). These genes exhibited specific and/or high expression at various stages of fruit development, with particularly significant expression peaks at about 25 or 55 days after flowering (DAF). The expression patterns of these genes coincided with the observed increase in lignin content in pear fruit (Xue et al., 2019), suggesting that our 14 identified genes play a vital role in lignin synthesis and accumulation during fruit development.

GWAS analysis of fruit stone cell content

Stone cell content is a crucial factor that influences pear fruit quality. Increased stone cell accumulation negatively affects fruit texture, especially in the flesh of Chinese white pears, where such cells form a mosaic pattern (Figure 1) (Yan et al., 2014; Xue et al., 2019; Zhang et al., 2021a). These cells, known as brachysclereids, exhibit secondarily thickened cell walls and

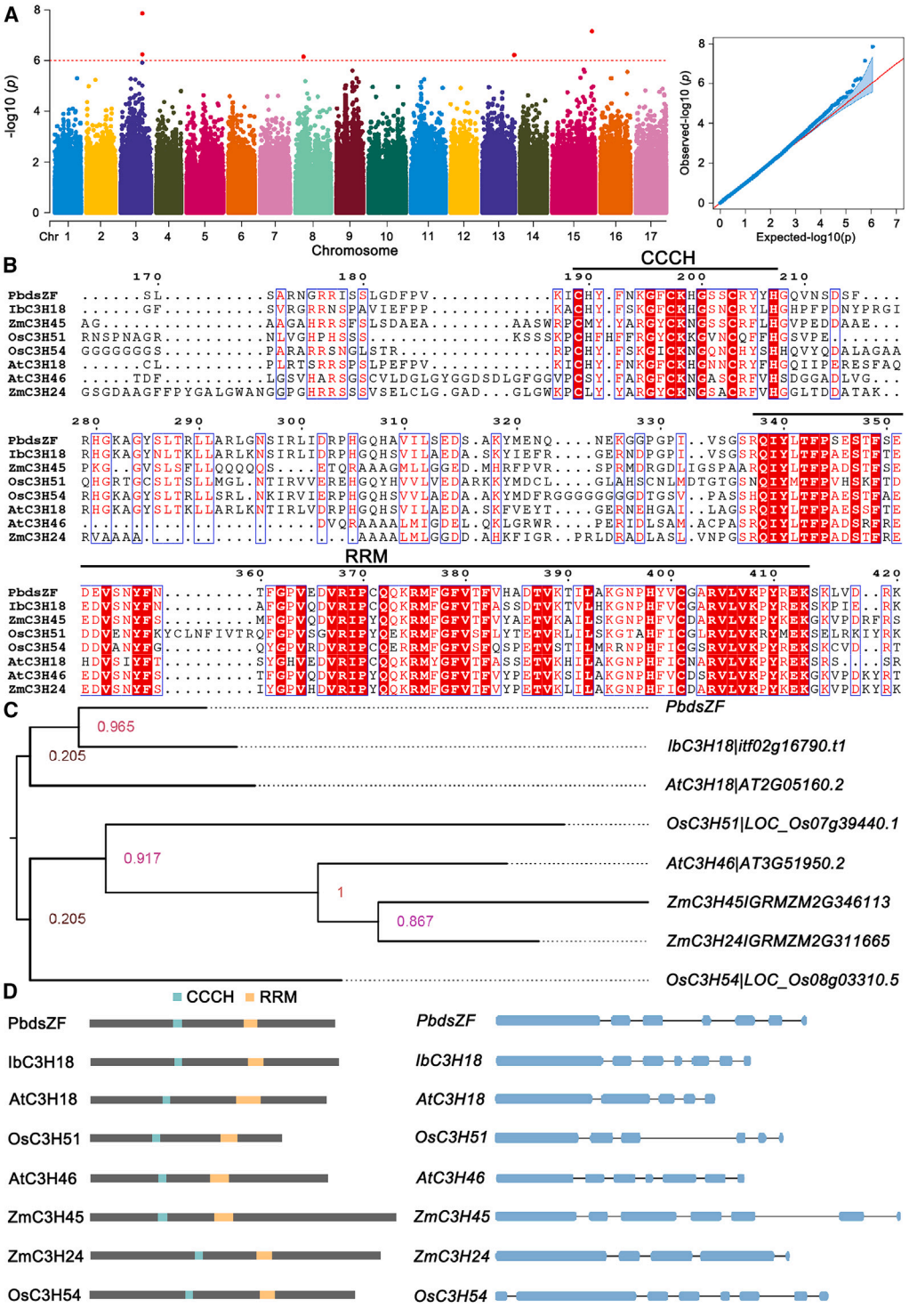

Figure 5. *PbdsZF*, a candidate gene associated with the stone cell trait in pear.

(A) Manhattan plot from GWAS using re-sequencing data and stone cell phenotype data, showing significant association peaks.

(B) Multiple sequence alignment of the PbdsZF protein and its closest homologs from sweet potato, maize, *Arabidopsis*, and rice. Black lines delineate the conserved RNA recognition motif (RRM) and the CCCH zinc finger motif.

(C) Phylogenetic tree constructed using IQ-TREE (Minh et al., 2020) (version 2.3.2) showing the evolutionary relationships between PbdsZF and its closest homologs in sweet potato, maize, *Arabidopsis*, and rice. Numbers at the nodes indicate bootstrap support percentages obtained from 1000 replicates.

(D) Schematic representation of PbdsZF and its homologs. Left: positions of RRM and CCCH motifs. Blue boxes represent the RRM motifs, and red boxes represent the CCCH motifs. Right: genomic structure of *PbdsZF* and its homologs. Blue boxes and black lines represent exons and introns, respectively.

ultimately differentiate into parenchyma cells characterized by the accumulation of lignin and cellulose (Yan et al., 2014; Xue et al., 2019; Zhang et al., 2021a). Using our high-quality DS genome along with previously published pear re-sequencing data, we performed a GWAS analysis to explore genetic variations associated with phenotypic traits of stone cells. We identified 28 potential loci on chromosomes 2, 3, 4, 5, 8, 9, 11, 13, 14, 15, and 17 (Supplemental Figure 38) that were linked to stone cell concentration using a criterion of $p < 1E{-}5$, as used by Zhang et al. (2021b). Most previously identified loci (9 out of 12) were also detected in our study, validating the accuracy of the GWAS identification in this research. By comparing the p values, the number of candidate sites was further reduced. With a stringent threshold of $p < 1E{-}6$, we identified four significant association peaks on chromosomes 3, 8, 13, and 15 (Supplemental Table 12; Figure 5A). Among the candidate genes identified, aside from previously characterized lignin-regulating genes such as those encoding cinnamate 4-hydroxylases (Anterola and Lewis, 2002), the *PbdsZF* gene (*PbslChr03G002183*) was of particular interest. Stone cells in pear are primarily produced during the early stages of fruit development, and *PbdsZF* from this candidate gene locus was highly expressed during the peak period of lignin synthesis in these early stages (Supplemental Figure 39A). In addition, a stone cell-associated SNP at 20 802 799 base pairs (bp) (C/G) on Chr3 was linked to *PbdsZF*. Haplotype analysis suggested that this SNP was associated with increased stone cell content, with the heterozygous variant genotype having the most significant impact (Supplemental Figure 39B). The PbdsZF proteins are characterized by a CCCH zinc finger motif of the C-X7-C-X5-C-X3-H type (CHYFNKGFCKHGSSCRYYH) and an RNA recognition motif (RRM) (Figure 5B; Supplemental Figure 40). We used PbdsZF as a query to identify its homologs in different *Pyrus* genomes and other Rosaceae species. Interestingly, the results showed that this gene is highly conserved across various species within the Rosaceae. Despite the relatively low overall amino acid sequence identity of PbdsZF with its orthologs in *Arabidopsis*, rice, sweet potato, and maize (ranging from ~23% to ~41%), the CCCH zinc finger and RRM motifs were remarkably consistent in their positions and structures. Further phylogenetic analysis revealed that *PbdsZF* is most closely related to *IbC3H18* (*itf02g16790.t1*) in sweet potato and *AtC3H18* (*AT2G05160*) in *Arabidopsis* (Figure 5C). Notably, *AtC3H18* is a unique single-copy gene in *Arabidopsis* and does not belong to any of the 11 established subfamilies (Wang et al., 2008). Our analysis, including multiple sequence alignment and phylogenetic and WebLogo analyses, demonstrated that the CCCH zinc finger motif of PbdsZF is distinct from all 152 *Arabidopsis* CCCH zinc finger motifs examined (Supplemental Table 13) and shows the highest similarity to the CCCH zinc finger motif of AtC3H18 (Supplemental Figure 41). In addition, the gene structure of *PbdsZF*, comprising six introns and seven exons, also differs from that of *AtC3H18*, which has four introns and five exons, yet it shares similarity with the structure of *IbC3H18* (Figure 5D). In conclusion, *PbdsZF* in the DS pear genome is classified as a non-tandem zinc finger (non-TZF) gene (Figure 5).

Validating the role of *PbdsZF* in regulating stone cell formation in pear fruit

To investigate the potential roles of *PbdsZF* in various tissues and during fruit development, we analyzed its transcript abundance using RNA-sequencing (RNA-seq) data across the DS pear genome. *PbdsZF* expression was relatively high in seeds, roots, reproductive organs (styles, stamens, flower buds, ovaries), and young fruits (25–55 DAF), whereas its expression was significantly lower in leaf buds, young leaves and stems, cotyledons, mature leaves, petals, sepals, flower stalks, fruit stalks, fruit cores, pericarp tissues, and fruits older than 55 DAF (Supplemental Figure 42). Interestingly, the expression pattern of *PbdsZF* throughout fruit development mirrored the observed pattern of lignin accumulation in stone cells, which is most pronounced between 25 and 55 DAF (Yan et al., 2014; Xue et al., 2019). This finding suggests a potential link between *PbdsZF* and the regulation of lignin biosynthesis during stone cell formation.

To better understand the role of *PbdsZF*, we investigated its subcellular localization using confocal microscopy. We analyzed the distribution of PbdsZF-GFP and observed that it localized exclusively to the nucleus. This localization, further validated by co-localization with DAPI, a dye that specifically stains nuclear DNA, strongly supports the nuclear localization of PbdsZF and suggests its involvement in nuclear-specific functions (Figure 6A). The primers used in this experiment are detailed in Supplemental Table 14. To further clarify the role of *PbdsZF* in regulating lignin biosynthesis and stone cell formation in pear fruit, we analyzed the distribution of *PbdsZF* transcripts within pear fruit. Using *in situ* hybridization with antisense probes for *PbdsZF* mRNA, we observed positive signals (blue-purple staining) specifically in regions associated with lignin deposition and developing stone cells (Supplemental Figure 43). These results indicate that *PbdsZF* mRNA is localized in areas of active lignification and stone cell development. Collectively, these findings underscore the specific role of *PbdsZF* in the formation of stone cells, suggesting its regulatory function in lignification and stone cell development in pear fruit.

To assess the effect of *PbdsZF* on stone cell development in pear fruit, we transiently overexpressed and silenced this gene in DS pear fruit at 39 DAF using agroinfiltration (Figure 6B). *PbdsZF* was highly expressed in fruits of the overexpression lines and expressed at significantly lower levels in fruits of silenced lines compared with flesh tissues around the EV-infiltrated sites (Supplemental Figure 44). Subsequent analysis of hand-cut sections, stained with phloroglucinol-HCl, revealed a significant increase in stone cell distribution and lignin content at the injection site for the *PbdsZF* overexpression construct (PbdsZF-OE) compared with the control vector. By contrast, silencing of *PbdsZF* with the PbdsZF-VIGS vector resulted in noticeably fewer stained regions, indicating lower levels of lignin and stone cells. Quantification assessments of these changes confirmed the significant effect of manipulating *PbdsZF* expression (Figure 6C). These results suggest that *PbdsZF* expression significantly affects the accumulation of stone cells and lignin in pear fruit. Furthermore, an RT–qPCR analysis showed that changes in *PbdsZF* expression affected the transcript levels of several key genes in the lignin biosynthetic pathway, including *PbdsPAL1*, *PbdsPAL4*, *PbdsC3H4*, *PbdsC4H5*, *PbdsF5H3*, *Pbds4CL2*, *Pbds4CL4*, *PbdsHCT2*, *PbdsCSE2*, *PbdsCCoAOMT1*, *PbdsCOMT8*, *PbdsCCR1*, *PbdsCCR2*, and *PbdsCAD1* (Figure 6D). Overall, these findings highlight the pivotal role of *PbdsZF* in regulating the biosynthesis of lignin, a key component of the stone cells in pear fruit.

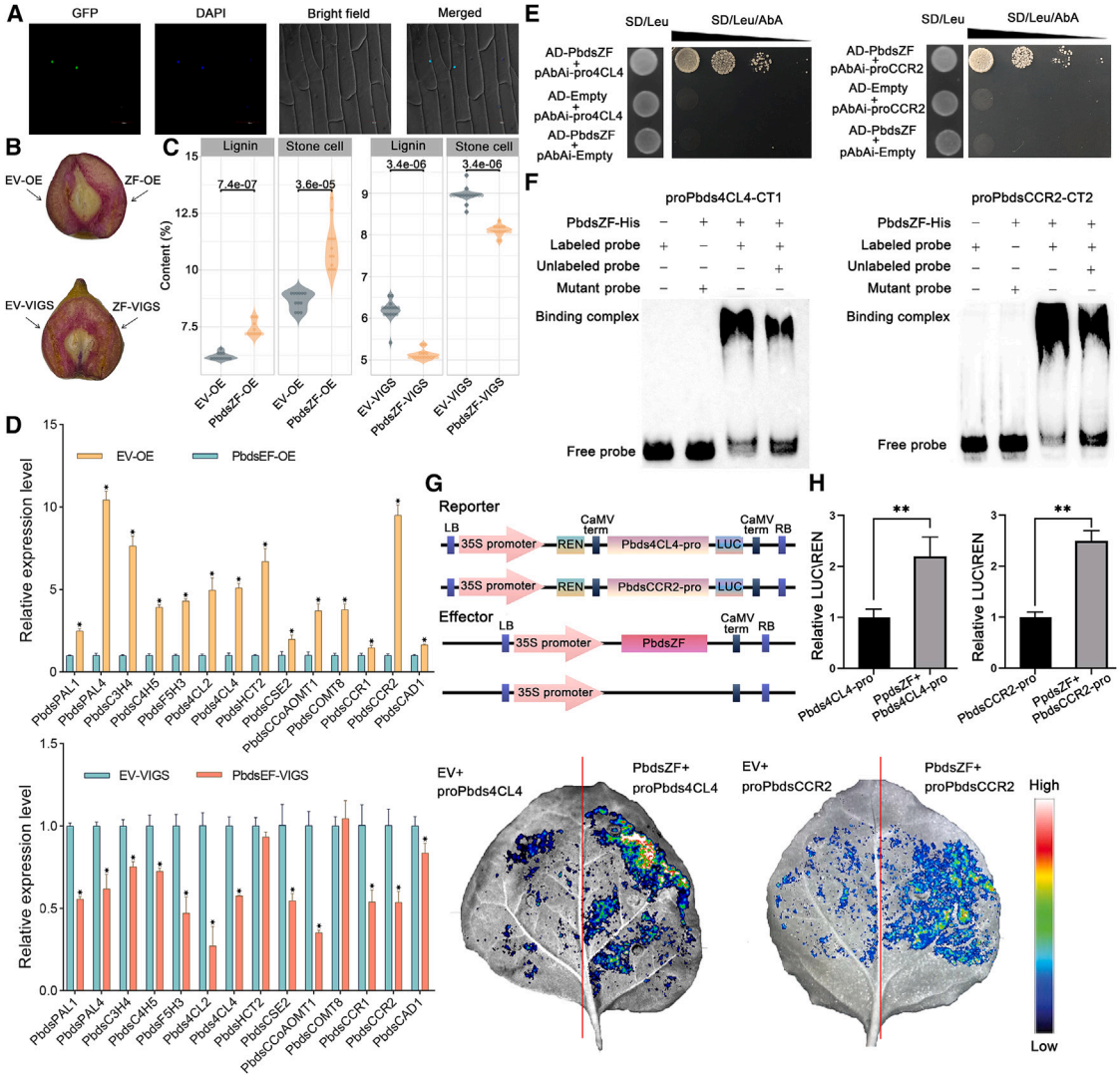

Figure 6. Functional validation of the role of PbdsZF in controlling stone cell content and lignin in pears, with PbdsZF activating key genes related to fruit lignification by binding directly to their promoters.

(A) Subcellular localization of PbdsZF fused with GFP in onion epidermal cells.

(B) Transient expression assays using *PbdsZF* overexpression and silencing constructs in 'Dangshansu' fruit at 39 days after flowering (DAF). Images were taken 10 days after agro-infiltration.

(C) Quantification of lignin content and stone cell content in the flesh tissue surrounding the infiltration sites. More than six fruits were injected with each construct in three independent experiments. EV, empty vector; OE, overexpression; VIGS, virus-induced gene silencing.

(D) Expression levels of genes encoding enzymes involved in monolignol biosynthesis and polymerization in the flesh tissue surrounding the infiltration site, as determined by RT–qPCR analysis.

(E) A Y1H assay showing the direct binding of PbdsZF to the promoters of *Pbds4CL4* and *PbdsCCR2*.

(F) An EMSA assay depicting the direct binding of PbdsZF to CT-rich motifs in the promoters of *Pbds4CL4* and *PbdsCCR2*.

(G) A dual-luciferase reporter assay in leaf epidermal cells of *N. benthamiana* showing significant enhancement of *Pbds4CL4 and PbdsCCR2* promoter activity by PbdsZF.

(H) Relative firefly luciferase/*Renilla luciferase* (LUC/REN) ratios in different leaf areas following co-expression of either PbdsCL4-promoter-LUC or PbdsCCR2-promoter-LUC with either the PbdsZF expression construct or the control construct. Data are presented as mean ± SD; *p*-values were determined by two-tailed Student's *t*-test (*$p < 0.05$, **$p < 0.01$).

PbdsZF binds to the CT-rich motifs in the promoters of core fruit lignification-related genes

Previous studies have shown that CCCH-type zinc finger proteins interact with CT-rich motifs to regulate downstream target genes (He et al., 2016; Li et al., 2022a; Lan et al., 2023). In line with these findings, we identified two CT-rich motifs, named CT1 (CTTTTTTCT) and CT2 (CTCTTTTT), within the promoters of *Pbds4CL4* and *PbdsCCR2*, two genes previously implicated in fruit lignification (Li et al., 2022a; Lan et al., 2023). Pear stone cells are primarily formed during the early stages of fruit development. Indeed, our study found that *Pbds4CL4* (Supplemental Figure 45) and *PbdsCCR2* (Supplemental Figure 46) were specifically and highly expressed in the early stages of pear fruit development, further confirming their role

in stone cell synthesis in pear fruit. We also examined RNA-seq data (Mamat et al., 2021; Gong et al., 2023; Zhu et al., 2023) for fruit development in four pear cultivars ('Cuiguan', 'Shannongsu', 'Korla' and 'Dangshansu') and found that *PbdsZF*, *Pbds4CL4*, and *PbdsCCR2* were highly expressed during the early stages of fruit development, but their expression decreased as the fruit matured (Supplemental Figure 47). We next performed yeast one-hybrid (Y1H) assays, which confirmed that PbdsZF could bind to the promoters of *Pbds4CL4* and *PbdsCCR2* (Figure 6E). An electrophoretic mobility shift assay (EMSA) further corroborated the Y1H findings, confirming the direct interaction between PbdsZF and the CT-rich motifs within the promoters of *Pbds4CL4* and *PbdsCCR2* (Figure 6F). This interaction was evidenced by distinct binding bands observed only when both PbdsZF and the biotin-labeled promoter probes were present. Unlabeled competitive promoter probes significantly reduced the intensity of these bands. By contrast, the absence of PbdsZF or the presence of mutated promoter probes resulted in no binding, suggesting a specific interaction between PbdsZF and the CT motifs within these promoters (Figure 6F).

To determine whether PbdsZF can activate the promoter activity of key genes in the lignin biosynthesis pathway, such as *Pbds4CL4* and *PbdsCCR2*, we performed a dual-luciferase reporter gene assay (Figure 6G). In *Nicotiana benthamiana* leaves, luminescence significantly increased when pGreenII0800-promoter-LUC was transiently co-expressed with pGreenII62-SK-PbdsZF, compared with its co-expression with the empty pGreenII62-SK vector. Further analysis of the LUC/REN ratio at these co-expression sites revealed a considerable increase ($P < 0.01$) in areas where PbdsZF and the promoter constructs were co-expressed, relative to co-expression with the empty vector (Figure 6H). These results indicate that PbdsZF acts as a transcriptional activator for *Pbds4CL4* and *PbdsCCR2*, directly binding to specific CT-rich motifs in their promoter regions and thereby regulating lignin biosynthesis in pear fruits.

DISCUSSION

Since the release of the first pear genome assembly in 2013 (Wu et al., 2013), numerous pear genomes have been published using various assembly techniques (Linsmith et al., 2019; Dong et al., 2020; Gao et al., 2021; Shirasawa et al., 2021; Sun et al., 2023; Ding et al., 2024); however, the assemblies of *P. bretschneideri* contained numerous gaps, leading to significant losses of genetic information. Such gaps have consistently impeded comprehensive genetic analyses and limited our ability to analyze gene functions and understand evolutionary pathways in pear species. In this study, we addressed this limitation by constructing two new, high-quality, gap-free *de novo* genome assemblies for *P. bretschneideri* cv. DS and ZS. We constructed these assemblies using PacBio HiFi reads and Hi-C data, which provided long-range information for accurate genome scaffolding. Compared with that of the existing *P. bretschneideri* genome, the scaffold N50 values of our new assemblies are at least 10 times higher (Wu et al., 2013). This substantial improvement in continuity significantly enhances our ability to perform detailed genetic analyses, potentially revolutionizing our understanding of disease resistance, stress tolerance, and quality traits in pear. These complete genomes provide a foundation for future research on the genetic mechanisms underlying key agronomic traits, thereby facilitating precision breeding and conservation efforts.

Integrative analysis to identify the lignification toolbox encoding bona fide enzymes

Pear quality is significantly influenced by stone cells, and lignin level plays a critical role in their development (Yan et al., 2014; Xue et al., 2019; Zhang et al., 2021a). Lignin is synthesized through the oxidative polymerization of monolignols: p-coumaryl, coniferyl, and sinapyl alcohols, which contribute to the formation of p-coumaryl, coniferyl, and sinapyl alcohols that yield p-hydroxyphenyl (H), guaiacyl (G), and syringyl (S) units, respectively (Yan et al., 2014). These monolignols are produced through 11 enzymatic reactions starting from the deamination of phenylalanine to hydroxycinnamoyl CoA esters, involving a series of hydroxylations and O-methylations. This pathway is catalyzed by enzymes encoded by the *PAL*, *C4H*, *C3H*, *4CL*, *HCT*, *CSE*, *CCoAOMT*, *F5H*, *COMT*, *CCR*, and *CAD* gene families. However, some of these gene families, such as *4CL*, *COMT*, *CAD*, and *CCR*, contain multiple members, adding complexity to the genetic control of lignin synthesis. Among the 11 gene families, three displayed no evidence of tandem or segmental duplications, consisting of low-copy or even single-copy genes. Single-copy genes are pivotal for essential housekeeping functions across species, preserved as unique entities by evolutionary pressures because of their vital roles. For example, *F5H* and *C4H* are crucial for the regulation of lignification and the composition of lignin monomers, highlighting the importance of specific genes in complex biological processes (Franke et al., 2000). The evolutionary pressures against gene duplication are particularly evident given the potential disruptions it can cause to protein functionality and stability within complex protein assemblies. The critical involvement of C4H in multiprotein complexes, essential for metabolic processes, highlights this phenomenon. Such insights not only shed light on the genetic regulation of lignin biosynthesis but also emphasize the evolutionary adaptations that maintain the integrity and functionality of vital metabolic pathways.

In pear genomes, tandem duplication has significantly influenced the expansion of specific gene families, notably affecting only a subset of lignification gene families (*PAL*, *HCT*, *C3H*, and *COMT*; Supplemental Figure 25). This selective impact contrasts with the broader impact observed in other gene families, suggesting a targeted evolutionary strategy to enhance lignification. This has led to the diversification of these gene families beyond genome-wide averages. However, such diversification does not result in uniform expression across all tissues; only a subset of these genes show preferential expression in lignified tissues, indicating a complex regulation tailored to specific functional needs. The morphological and functional diversification driven by this expression divergence plays a critical role in plant growth, adaptation, and defense. Furthermore, the *CAD* and *CCoAOMT* families, which have diversified through non-tandem duplication mechanisms (Supplemental Figure 26), exhibit functional redundancy and play significant roles in vascular lignification, as reported by Carocha et al. (2015). This redundancy, evident across different duplication mechanisms, highlights the complexity and adaptability of lignification

pathways, contributing significantly to the ecological success and adaptability of plants.

Identifying PbdsZF as a key transcription factor for stone cell regulation in pear fruits

We combined phylogenetic and RNA-seq analysis across 11 bona fide gene families to identify the key enzymes involved in lignin biosynthesis in pear fruit. We identified 14 genes that likely contribute to this process, collectively forming the "core fruit lignin toolbox" in the DS genome (Supplemental Table 11). Previous studies have shown that transcription factors regulate some of these key enzyme genes, influencing both lignin and stone cell content in pear fruits (Wang et al., 2021, 2024; Li et al., 2023; Xu et al., 2023; Xue et al., 2023; Zhu et al., 2023; Liu et al., 2024). In addition, genetic studies using GWAS and eQTL methods have revealed that genes such as *PbrNSC* and *PbrSTONE* are pivotal for the control of lignin accumulation in pear fruits (Zhang et al., 2021b; Wang et al., 2021). However, limitations of earlier sequencing technologies resulted in many gaps in previously published *P. bretschneideri* genomes (Wu et al., 2013), making it challenging to identify genes responsible for trait variations. Using our newly assembled gap-free *P. bretschneideri* genome for a GWAS, we identified a novel transcription factor, PbdsZF, that regulates lignin accumulation and may serve as a candidate gene for modification of stone cell content in pear fruits.

The PbdsZF protein contains a zinc finger_CCCH (Cys-Cys-Cys-His) domain and an RRM, categorizing it within the family of zinc finger protein transcription factors. CCCH-type zinc finger transcription factors are known for their diverse roles in plant development and stress responses. For instance, *AtC3H17* in *Arabidopsis* affects seed development by modulating the transcription of *AtOleo1*, *AtOleo2*, and *AtCRU3* (Seok et al., 2016). Similarly, IbC3H18 in tomato enhances salt and drought tolerance through its interaction with IbPR5 (Zhang et al., 2019), and the MaCCCH33-like2/MaEBF1/MaABI5-like complex in banana is crucial for regulation of fruit softening during ripening (Song et al., 2024b). In addition, overexpression of *OsDOS* in rice delayed leaf senescence (Kong et al., 2006), and upregulation of *PvC3H72* in switchgrass enhanced cold tolerance by modulating the expression of ABA-responsive genes (Xie et al., 2019). In pear, the zinc finger protein PbZFP1 has been characterized as a positive regulator of ABA biosynthesis in the flesh (Gu et al., 2024). Despite these insights, the specific roles of CCCH-type zinc finger transcription factors in lignin biosynthesis remain largely unexplored. Our study contributes to this knowledge gap by demonstrating that *PbdsZF*, which is phylogenetically close to *AtC3H18* and *IbC3H18*, can alter the transcript levels of genes within the core fruit lignification toolbox and influence lignin accumulation in pear fruits.

Understanding the interplay between transcription factors and *cis*-elements is crucial for deciphering the mechanisms of transcriptional regulation. CCCH proteins are known to bind to DNA sequences with CT-rich motifs. For example, Li et al. (2022a) found that the CCCH protein PuC3H35 from *Populus ussuriensis* binds directly to the CTCTCTTCT motifs in the *PuEARLI1* promoter region and the CTCACCTTC or CTCTTTT motifs in the *PuANR* promoter. This interaction leads to upregulation of these genes, enhancing drought tolerance in transgenic plants. Although the interactions between CCCH proteins and DNA motifs have been investigated in some plant species, the specific *cis*-elements targeted by CCCH transcription factors during stone cell formation remain poorly understood. In this study, a dual-luciferase reporter assay revealed that the transcription factor PbdsZF directly activates the promoters of key lignin biosynthesis genes such as *Pbds4CL4* and *PbdsCCR2* by interacting with CT-rich motifs within their promoter regions (Figure 7). This finding not only contributes to our understanding of lignin biosynthesis but also suggests potential targets for enhancing specific agronomic traits in pear through genetic engineering.

In summary, we present two gap-free genome assemblies for *P. bretschneideri* cultivars, offering significant resources for exploring genome evolution and understanding the impact of chromosome structure variations on rapid speciation. Our phylogenetic analysis identified 640 genes as bona fide contributors to lignification in *Pyrus*, including a core set of 14 genes that are specifically expressed in lignifying fruit tissues. Through a combined GWAS and gene expression analysis, we identified PbdsZF, a CCCH-type zinc finger transcription factor, as a key regulator of stone cell formation in pear fruits. In addition, our findings demonstrate that PbdsZF directly interacts with the CT-rich motifs in the promoters of *Pbds4CL4* and *PbdsCCR2*, thereby influencing both lignin biosynthesis and stone cell accumulation. These two high-quality, gap-free genomes represent powerful tools for the identification of candidate genes associated with fruit-quality traits. Furthermore, the present findings deepen our understanding of the regulatory mechanisms underlying stone cell formation and offer valuable genetic resources for future efforts aimed at enhancing pear fruit quality.

MATERIALS AND METHODS

Sample preparation and sequencing

We obtained samples of DS and ZS for sequencing from Anhui Agricultural University, Hefei, Anhui, China. High-quality genomic DNA was extracted using the Qiagen Genomic DNA Kit according to the manufacturer's protocols (Qiagen, https://www.qiagen.com). For PacBio sequencing, genomic DNA (gDNA) was used to construct SMRTbell libraries according to the standard protocol of Pacific Biosciences (CA, USA), which uses 20-kb preparation solutions. Construction of the SMRTbell library, including end repair, DNA shearing, and ligation of DNA fragments with hairpin adapters, facilitated the creation of circular templates for SMRT sequencing. The library was sequenced on a PacBio Revio System in circular consensus sequence (CCS) mode, and the raw data were processed into HiFi reads using the CCS workflow with parameters including –knrt-ada –pbdc-model –non-hifi-prefix fail –kestrel-files-layout–min-rq 0.9 –stderr-json-log –log-level INFO –streamed. CCS software (https://github.com/PacificBiosciences/ccs) was used for quality control of the raw data with the parameters "-min-length 100," "-min-rq 0.99," and "-min-passes 1." For Hi-C sequencing, fresh leaf samples were fixed with 2% formaldehyde to facilitate DNA–protein crosslinking. Hi-C library construction involved several steps: crosslinked DNA processing, digestion, biotin labeling, proximity ligation, and DNA purification, all performed according to the protocol. In brief, 400 units of MboI were used to digest the crosslinked material, biotin-14-dCTP was incorporated for labeling, and blunt-end ligation was performed. After purification, reverse crosslinking, and re-ligation, sonication was performed to shear the chromatin DNA into 200- to 600-bp fragments. Streptavidin magnetic beads were then used to enrich the biotin-labeled Hi-C fragments. The prepared Hi-C library was subjected to paired-end sequencing using a 150-bp read

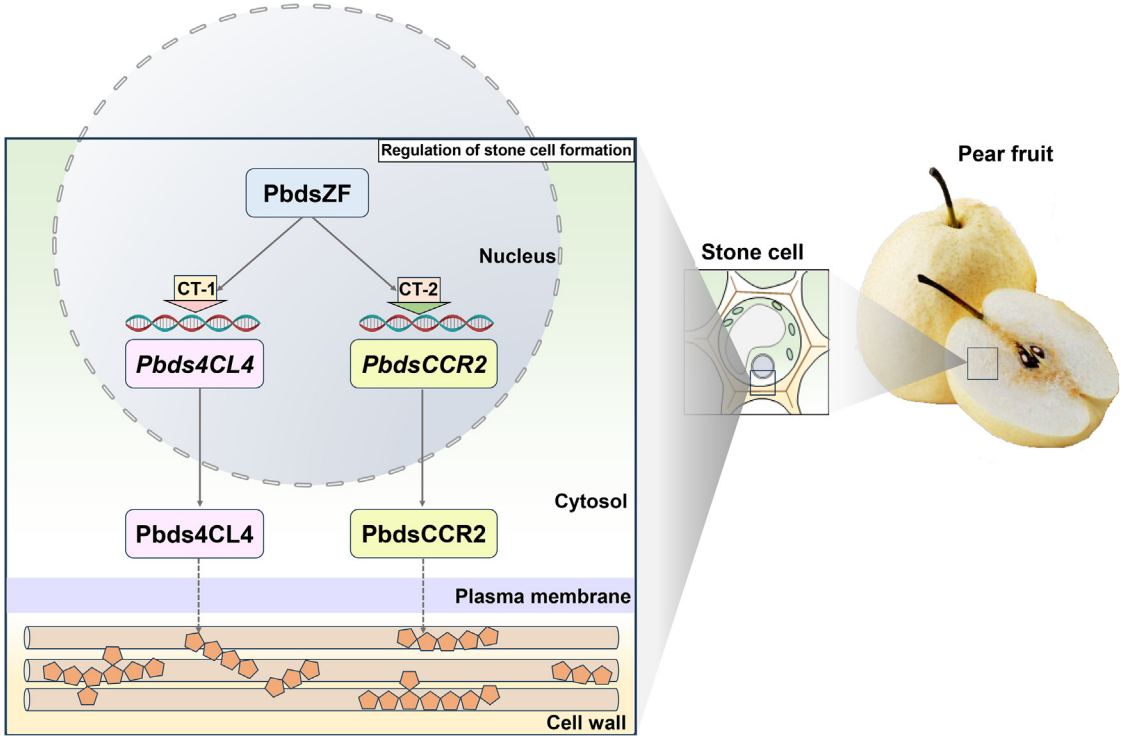

Figure 7. Hypothetical working model of PbdsZF regulation of stone cell formation in pear fruits.
PbdsZF activates the transcription of *Pbds4CL4* and *PbdsCCR2* genes by binding to the CT-1 element and CT-2 element in their promoter regions, respectively.

length on the DNBSEQ-T7 platform to capture spatial interactions between chromosomal regions.

Genome assemblies generated by GPM editing

A total of 28.08 Gb and 27.17 Gb of HiFi reads were generated from the DS and ZS libraries, respectively. Approximately 52.07 Gb of Hi-C data were used to anchor all contigs to the DS chromosomes, and 49.16 Gb were used for the ZS chromosomes. For the preliminary assembly, we used six programs (Wtdbg2 version 2.5; Canu version 2.0; Flye version 2.8.2-b1689; MECAT2 version 1.0; FALCON toolkit version 1.3.0; Hifiasm version 0.19.8) to assemble the HiFi reads for both the DS and ZS cultivars. Wtdbg2 (Ruan and Li, 2020) (version 2.5) was used for assembly with the parameters "-g 520m, -x ccs" to optimize the process. The genomes were assembled using Canu (Koren et al., 2017) (version 2.0) specified for CCS technology with the parameter "-pacbio-hifi"; Flye (Kolmogorov et al., 2019) (version 2.8.2-b1689) was used for assembly with "–pacbio-hifi -g 520m"; MECAT2 (Xiao et al., 2017) (version 1.0) was used to assemble the HiFi reads with the command "mecat.pl assemble" and the following parameters: FSA_OL_FILTER_OPTIONS = "–min_identity = −1 –max_overhang = −1," ASM_OVLP_OPTIONS = "-a 400-j 1 -u 0 -e 0.5 -n 100 -b 2000 -z 10," CNS_OUTPUT_COVERAGE = 30, TRIM_OVLP_OPTIONS = "-B," CNS_PCAN_OPTIONS = "-k 100 -p 100000," CNS_OVLP_OPTIONS = "-kmer_size 13," MIN_READ_LENGTH = 2000, GENOME_SIZE = 520m, and FSA_ASSEMBLE_OPTIONS = ""; FALCON toolkit (Chin et al., 2016) (version 1.3.0) was used with the parameters: overlap_filtering_setting = –min-cov 2 –max-cov 120 –max-diff 90, ovlp_DBsplit_option = -s200, ovlp_HPCdaligner_option = -v -M24 -B128, ovlp_daligner_option = -s100 -l3000 -h1024 -k24 -e.99, length_cutoff_pr = 10000, falcon_sense_option = –n-core 4 –max-n-read 100 –min-cov 3 –output-multi –min-idt 0.70, pa_DBsplit_option = -s400 -x500, pa_HPCREPmask_option = -s100 -e.8 -w8 -h480 -k18, pa_HPCdaligner_option = -v -M24 -B128, pa_daligner_option = -s100 -w8 -e0.8 -h70 -k18 -l1000 to assemble the genomes; Hifiasm (Cheng et al., 2021) (version 0.19.8) was used for genome assembly with its default settings. Finally, GPM (Zhang et al., 2016) was used to integrate and refine all the HiFi read assemblies, ensuring optimized results. For the phased haplotype assemblies, we used only Hifiasm software (Cheng et al., 2021) (version 0.19.8) for *de novo* contig assembly, integrating both Hi-C and HiFi reads using default parameters. Finally, the Hi-C reads were used to correct and scaffold contigs with default parameters of HapHiC (Zeng et al., 2024) (version 1.0.6) to assess and eliminate redundancy in unanchored contigs. Juicebox (version 1.11.08) was used to examine and refine the obtained scaffolds.

Closing the gaps with DEGAP

For the preliminary assembly, we first used the six contig assemblies and employed HiFi reads for gap filling using quarTeT (Lin et al., 2023) (version 1.2.1), TGS-GapCloser (Xu et al., 2020) (version 1.2.0), and LR_Gapcloser (Xu et al., 2019) (version 1. 0). The remaining gaps were closed using a cyclic elongation method, DEGAP, as described by Huang et al. (2024). To close the gaps, we selected optimal HiFi reads capable of spanning the gap boundaries. Before each elongation round, all HiFi reads were mapped to one edge of the gap (a 60 000-bp sequence) using Minimap2 with the parameters "-Y -ax asm20." We then identified a set of HiFi reads with high sequence identity and coverage that could extend the edge sequence. After selecting the best extension reads, these were assembled using Hifiasm with default parameters. Once extension contigs or HiFi reads were identified, we selected the sequences with the highest alignment to further extend the edge sequence and construct the extension. If the extended sequence successfully bridged the gap to the opposite side, the elongation process was considered complete. If not, the new extension sequence was used as input for the next elongation cycle. This iterative process continued until an alignment with high identity and long coverage fully spanned the gap (Huang et al., 2024). For the phased haplotype assemblies, we first used only HiFi reads for gap filling with quarTeT (Lin et al., 2023) (version 1.2.1), TGS-GapCloser (Xu et al., 2020) (version 1.2.0), and LR_Gapcloser (Xu et al., 2019) (version 1.0). We then used the same method as above to fill the remaining gaps using DEGAP (Huang et al., 2024).

Genome quality assessment

Multiple methods were used to evaluate the completeness and integrity of the assembled genomes. LTR_retriever (Ou and Jiang, 2018a) (version 5.22) was used to calculate the LAI with parameters -q -all. The genome completeness was evaluated using BUSCO (embryophyta_odb10: 1614 core plant conserved genes) (Simão et al., 2015) (version 5.22). To evaluate the overall base accuracy of the genome, we calculated the QV using Merqury (Rhie et al., 2020) (version 1.3) with default settings.

Gene annotation

RepeatModeler2 (Flynn et al., 2020) (version 2.0.1) was used for *de novo* identification and classification of repeats in the DS and ZS genomes. The genomes were then soft-masked using RepeatMasker (version 4.0.7) with their respective repeat libraries and the parameters -xsmall, -norna, and -nolow, and BRAKER3 (Gabriel et al., 2023) was used to annotate the soft-masked genome. First, trimmomatic (version 0.39) was applied to RNA-seq reads for quality filtering and trimming. HISAT2 (Pertea et al., 2016) (version 2.2.1) was then used to map the RNA-seq libraries, and the resulting mapping data were provided as input to the BRAKER pipeline, together with protein sequences from related species (*Pyrus betulifolia*, *Malus domestica*, *Pyrus communis*, *Arabidopsis thaliana*, and *Prunus persica*) from the GDR database (https://www.rosaceae.org/). This pipeline uses StringTie (Pertea et al., 2015) (version 2.2.0) to assemble RNA-seq reads, followed by gene prediction and training with AUGUSTUS (Stanke et al., 2006) (version 3.8.0) and GeneMark (Gabriel et al., 2023) (version 1.0). The gene sets were integrated using TSEBRA (Gabriel et al., 2021) (version 1.0.3). Finally, functional annotations were obtained using eggNOG-mapper (Huerta-Cepas et al., 2019) (version 5.0) to identify potential functions based on homology to sequences in the Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), Protein family (Pfam), and Swiss-Prot databases. For the phased haplotype assembly annotations, Liftoff (Shumate and Salzberg, 2021) was used to annotate genes from the preliminary assembly annotation.

Phylogenetic analysis

We used OrthoFinder (Emms and Kelly, 2019) (version 2.5.5) to identify OGs within the *P. bretschneideri* cv. DS and *P. bretschneideri* cv. ZS genomes, as well as those of 18 other species (*F. vesca*, *R. chinensis*, *R. occidentalis*, *M. domestica*, *P. pyrifolia* cv. Nijisseiki, *P. pyrifolia* cv. Cuiguan, *P. bretschneideri* cv. Xuehua, *Pyrus sinkiangensis* cv. Kuerxiang, *P. pyrifolia* cv. Yunhong, *P. betulifolia* cv. Shanxiduli, *P. ussuriensis* × *communis* cv. Zhonggai, *P. communis* cv. d'Anjou, *P. communis* cv. Bartlett, *P. persica*, *P. dulcis*, *P. mume*, *Populus trichocarpa*, and *A. thaliana*). Amino acid alignments for each single-copy OG were performed using MAFFT (Katoh and Standley, 2013) (version 7), and nucleotide alignments were produced using PAL2NAL (Suyama et al., 2006) (version 14.0). The concatenated alignments of all single-copy genes were then used to create a maximum-likelihood phylogeny with RAxML (Stamatakis, 2014) (version 7.2.8). In addition, species divergence times were estimated using MCMCTREE in PAML (Yang, 2007) (version 4.9). The calibration points were selected as follows: 102.0–112.5 million years ago (Mya) between *P. trichocarpa* and *A. thaliana*, 34.4–67.2 Mya between *P. persica* and *P. communis* cv. Bartlett, and 49.2–77.1 Mya between *F. vesca* and *M. domestica*.

Identification of core fruit lignification genes

These protein sequences were predicted and identified using a combined approach: a keyword-based search complemented by BLASTP analysis with query proteins from *P. trichocarpa*, *Eucalyptus grandis*, and *A. thaliana* (Carocha et al., 2015; Cao et al., 2019). The sequences were then used to construct a comprehensive phylogenetic tree with the best-fit model in IQ-TREE (Minh et al., 2020) (version 2.3.2). The tree also included protein sequences from additional species such as *Oryza sativa* and *Vitis vinifera*. To accurately delineate bona fide clades that contained homologs of genes experimentally verified as participating in fruit lignification, we incorporated the sequences of enzymes (PALs, C4Hs, C3Hs, F5Hs, 4CLs, HCTs, CSEs, CCoAOMTs, COMTs, CCRs, and CADs) that had been confirmed to exhibit true biological functions and enzymatic activity.

GWAS analysis for stone cell traits

The stone cell trait phenotypes of 312 pear accessions were obtained from a previously published paper (Zhang et al., 2021b). The genome of DS, which exhibits high lignin and stone cell contents, was used to identify candidate genes associated with fruit stone cell traits. The raw re-sequencing data for 312 pear accessions (PRJNA563813) were retrieved from the National Center for Biotechnology Information (NCBI) database (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA563813) (Zhang et al., 2021b). Four low-coverage accession lines (SRR13493728, SRR13493935, SRR13493944, and SRR13493954) were excluded from further analysis because they had a sequencing depth of less than 2×. Low-quality reads and adaptors were removed using fastx_toolkit (https://github.com/agordon/fastx_toolkit). Clean, paired-end short-read data from each accession were mapped to the DS genome using BWA–MEM (Li, 2013) (version 0.7.11) with default settings. The resulting data were processed by sorting with Picard and filtering with SAMtools (Li et al., 2009) (version 1.1071-gb298f29) to retain reads with a mapping quality greater than 20. Variant calling for each accession was performed using default parameters of the GATK pipeline (McKenna et al., 2010) (version 3.8-1-0), with GenotypeGVCFs and CombineGVCFs tools for calling population nucleotide variants and both VariantFiltration and SelectVariants for filtering the population genotypes. SNPs with less than 10% missing data and a minor allele frequency (MAF) greater than 0.05 were selected for GWAS analysis. GWAS was performed using a mixed linear model (MLM) that incorporated a kinship matrix and principal-component analysis, implemented in GAPIT (Tang et al., 2016) (version 2). The significance cutoff for this GWAS was set to $-\log10(p) < 6$. Manhattan plots illustrating these results were generated using the CMplot package (https://github.com/YinLiLin/R-CMplot) in R (version 3.5.0).

Gene expression

The RNA-seq dataset generated from six floral organ tissues (flower stalk, style, petal, stamen, ovary, and sepal), six fruit developmental stages (15, 25, 55, 85, 115, and 145 DAF), and various other tissues, including adult leaves, stems, flower buds, leaf buds, juvenile leaves, juvenile stems, cotyledons, roots, fruit stalks, seeds, fruit cores, and pericarps, was assessed using FastQC (version 0.11.7). Trimmomatic (Bolger et al., 2014) (version 0.39) was used to remove low-quality reads and bases, ensuring that only high-quality reads proceeded to the next stages of analysis. The clean reads were then aligned to annotated gene models using HISAT2 (Pertea et al., 2016) (version 2.2.1). Expression quantification was performed with StringTie (Pertea et al., 2015) (version 2.2.0), which normalized the data and estimated the transcripts per million (TPM) values for each gene model, as described by Cao et al. (2024a). The expression data were visualized on a pictorial representation of a pear using the Fancy Heatmap Browser (eFP Browser) in TBtools (Chen et al., 2020) (version 2.122).

Transient transformation of pear fruits

For transient overexpression in pear fruit, the full-length coding sequence (CDS) of *PbZF* was cloned into the pCAMBIA1301-35S binary vector. The recombinant plasmids were transfected into *Agrobacterium tumefaciens* strain GV3101 using the freeze–thaw technique. Pear fruit at 39 DAF, still attached to 20-year-old trees, were selected for *Agrobacterium* infiltration. One half of each fruit was injected with *A. tumefaciens* containing the gene-of-interest expression vectors, and the other half was injected with *A. tumefaciens* carrying empty pCAMBIA1301-35S vectors, serving as controls. All injections were performed using a syringe as described by Wang et al. (2024).

Histochemical staining and determination of lignin content in pear fruits

Pear fruit cross-sections were subjected to histochemical staining using standard phloroglucinol staining techniques as described by Wang et al.

(2024). Stone cell analysis was performed using procedures previously published by Yan et al. (2014) and Wang et al. (2024). For this analysis, a 5-g portion of pulp, extending from 2.0 mm beneath the peel to 0.5 mm outside the core, was sampled and frozen at −80°C for 24 h. Nine replicates of pear pulp were homogenized at 20 000 rpm for 3 min using water. The supernatant was decanted, and fresh water was added repeatedly until the upper layer became clear. The pear stone cells were then filtered onto paper and dried to a constant weight. Stone cell content was calculated using the formula: stone cell content (%) = (weight of stone cells in grams dry weight [g DW]/weight of pulp in grams fresh weight [g FW]) × 100. Lignin content was analyzed as described by Cai et al. (2010). In brief, stone cells were initially ground in 95% ethanol, and the resulting sediments were washed three times, first with 95% ethanol and then with a 1:2 (v/v) ethanol:hexane mixture, and dried. The dried sediment was digested in 2 ml of 25% (v/v) acetyl bromide in acetic acid at 70°C for 30 min. The reaction was quenched with 0.9 ml of 2 M NaOH, followed by the addition of 0.1 ml of 7.5 M hydroxylamine hydrochloride and 5 ml of acetic acid. The total volume was adjusted to 10 mL with acetic acid, and the absorbance at A280 was measured. A linear calibration curve was developed using commercial alkali lignin (Yuanye Biotech, Shanghai, China) and used to quantify lignin content.

Subcellular localization

The full-length CDS of *PbdsZF* without a stop codon was inserted into the pCAMBIA1305 vector for subcellular localization analysis. The construct was electroporated into *A. tumefaciens* GV3101 (pSoup-p19) competent cells. Transformed bacteria (OD600 = 0.8) were delivered into the epidermis of *N. benthamiana* leaves via injection. After a 3-day incubation in darkness, GFP fluorescence was visualized using a Leica TCS-SP8 SR fluorescence microscope (Leica Microsystems Ltd, Germany).

In situ hybridization

Fresh pear fruit segments were collected at 39 DAF and fixed overnight in a specialized *in situ* hybridization fixation solution from Servicebio (Wuhan, China). The probe sequence targeting the *PbdsZF* gene yielded a single hit in a BLAST search against our DS genome, indicating its high specificity. *In situ* hybridization probes were synthesized by Servicebio (Wuhan, China) and labeled with digoxigenin, and hybridization was performed as detailed by Wang et al. (2024).

Y1H assays

The CDS of *PbdsZF*, along with promoter fragments of *Pbds4CL4* and *PbdsCCR2* containing CT motifs, were cloned and inserted into the pGADT7 and pABAi vectors, respectively. The bait plasmid was linearized and then integrated into the Y1HGold yeast strain as described by Cao et al. (2024b). To determine the minimum inhibitory concentration of Aureobasidin A (AbA) for the pAbAi constructs, yeast strains harboring these plasmids were grown on SD medium lacking uracil. The pGADT7-PbdsZF construct was then introduced into these strains, which contained specific promoter regions. The efficacy of interactions between the pAbAi and pGADT7-PbdsZF constructs was assessed by observing colony growth on selective SD/-Ura/AbA medium. The empty pGADT7 vector served as a negative control for the corresponding recombinant pABAi. Images of the colonies were captured after 48 h of incubation.

LUC/REN activity analysis

The promoters of *Pbds4CL4* and *PbdsCCR2* were inserted into the pGreenII 0800-LUC vector as reporters to assess luciferase (LUC) activity. The full-length CDS of *PbdsZF* was cloned into the pGreenII62-SK vector to serve as an effector, with the empty vector used as a control. These recombinant vectors were transformed into *A. tumefaciens* GV3101 (pSoup-p19) competent cells, which were subsequently injected into *N. benthamiana* leaves. The transcriptional activation potential of *PbdsZF* was assessed by measuring the ratio of LUC to REN activity, detected 48 h post-transfection using a dual LUC assay kit (Promega, USA). The Tanon 4600SF Chemiluminescent Imaging System (Shanghai, China) was used to capture the fluorescence images.

DATA AND CODE AVAILABILITY

The raw sequence data reported in this paper have been deposited in the Genome Sequence Archive at the National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences (GSA: CRA016243 and CRA016227) and are publicly accessible at https://ngdc.cncb.ac.cn/gsa. The raw transcriptome sequencing data were obtained from NCBI (RJNA867130, PRJNA588520, and PRJNA825067) and the China National GeneBank Sequence Archive (CNSA: CNP0003413). The genetic variant file used for GWAS analysis can be accessed at https://data.pawsey.org.au/download/wcga-pangenome/share_file/SNPs_variants.vcf.tar.gz.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32201602, 32101486, 32102364), the Natural Science Fund of Hubei Province (Grant No. 2023AFB1036), the Funding of China Tobacco Genome Project (Grant No. 110202201012 [JY-12]), the Beijing Life Science Academy Project (Grant No. 2023200CC0270), and the Shanxi Provincial Higher Education Youth Academic Leader Project (Grant No. 2024Q031). All the authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Y. Cao supervised the project. Y. Cao designed the experiments and wrote the manuscript with input from all authors. R.K.V., X.F., B.D., H. Huo, M.A., J.H., L.J., H.W., R.L., Y. Cai, Z.X., X.L., H. Hu, M.L., and F.S. constructed the vectors and revised the paper. Y. Cao, X.F., and J.H. transformed the plants. Z.X., V.R., Y. Cao, and F.S. analyzed the data.

SUPPLEMENTAL INFORMATION

Supplemental information is available at *Plant Communications Online*.

Received: October 11, 2024
Revised: December 1, 2024
Accepted: December 30, 2024
Published: December 31, 2024

REFERENCES

Anterola, A.M., and Lewis, N.G. (2002). Trends in lignin modification: a comprehensive analysis of the effects of genetic manipulations/mutations on lignification and vascular integrity. Phytochemistry **61**:221–294.

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

Cai, Y., Li, G., Nie, J., Lin, Y., Nie, F., Zhang, J., and Xu, Y. (2010). Study of the structure and biosynthetic pathway of lignin in stone cells of pear. Sci. Hortic. **125**:374–379.

Cao, Y., Li, X., and Jiang, L. (2019). Integrative analysis of the core fruit lignification toolbox in pear reveals targets for fruit quality bioengineering. Biomolecules **9**:504.

Cao, Y., Chen, Y., Zhang, L., and Cai, Y. (2023). Two monolignoid biosynthetic genes 4-coumarate: coenzyme A ligase (4CL) and p-coumaric acid 3-hdroxylase (C3H) involved in lignin accumulation in pear fruits. Physiol. Mol. Biol. Plants **29**:791–798.

Cao, Y., Hong, J., Zhao, Y., Li, X., Feng, X., Wang, H., Zhang, L., Lin, M., Cai, Y., and Han, Y. (2024a). De novo gene integration into regulatory networks via interaction with conserved genes in peach. Hortic. Res. **11**:uhae252.

Cao, Y., Mo, W., Li, Y., Xiong, Y., Wang, H., Zhang, Y., Lin, M., Zhang, L., and Li, X. (2024b). Functional characterization of NBS-LRR genes reveals an NBS-LRR gene that mediates resistance against Fusarium wilt. BMC Biol. **22**:45.

Carocha, V., Soler, M., Hefer, C., Cassan-Wang, H., Fevereiro, P., Myburg, A.A., Paiva, J.A.P., and Grima-Pettenati, J. (2015). Genome-wide analysis of the lignin toolbox of E ucalyptus grandis. New Phytol. **206**:1297–1313.

Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant **13**:1194–1202.

Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods **18**:170–175.

Cheng, X., Zhang, J., Wang, H., Chen, T., Li, G., Yan, C., Jin, Q., Lin, Y., and Cai, Y. (2020). Effects of metaxenia on stone cell formation in pear (Pyrus bretschneideri) based on transcriptomic analysis and functional characterization of the lignin-related gene PbC4H2. Forests **11**:53.

Chin, C.-S., Peluso, P., Sedlazeck, F.J., Nattestad, M., Concepcion, G.T., Clum, A., Dunn, C., O'Malley, R., Figueroa-Balderas, R., Morales-Cruz, A., et al. (2016). Phased diploid genome assembly with single-molecule real-time sequencing. Nat. Methods **13**:1050–1054.

Ding, B., Hu, H., Cao, Y., Xu, R., Lin, Y., Muhammad, T.u.Q., Song, Y., He, G., Han, Y., Guo, H., et al. (2024). Pear genomes display significant genetic diversity and provide novel insights into the fruit quality traits differentiation. Horticultural Plant Journal **10**:1274–1290. https://doi.org/10.1016/j.hpj.2024.05.005.

Dong, X., Wang, Z., Tian, L., Zhang, Y., Qi, D., Huo, H., Xu, J., Li, Z., Liao, R., Shi, M., et al. (2020). De novo assembly of a wild pear (Pyrus betuleafolia) genome. Plant Biotechnol. J. **18**:581–595.

Emms, D.M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. **20**:1–14.

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

Franke, R., McMichael, C.M., Meyer, K., Shirley, A.M., Cusumano, J.C., and Chapple, C. (2000). Modified lignin in tobacco and poplar plants over-expressing the Arabidopsis gene encoding ferulate 5-hydroxylase. Plant J. **22**:223–234.

Gabriel, L., Hoff, K.J., Brůna, T., Borodovsky, M., and Stanke, M. (2021). TSEBRA: transcript selector for BRAKER. BMC Bioinf. **22**:1–12.

Gabriel, L., Brůna, T., Hoff, K.J., Ebel, M., Lomsadze, A., Borodovsky, M., and Stanke, M. (2023). BRAKER3: Fully automated genome annotation using RNA-Seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. Preprint at bioRxiv. https://doi.org/10.1101/2023.06.10.544449.

Gao, Y., Yang, Q., Yan, X., Wu, X., Yang, F., Li, J., Wei, J., Ni, J., Ahmad, M., Bai, S., et al. (2021). High-quality genome assembly of'Cuiguan'pear (Pyrus pyrifolia) as a reference genome for identifying regulatory genes and epigenetic modifications responsible for bud dormancy. Hortic. Res. **8**:197.

Gong, X., Qi, K., Chen, J., Zhao, L., Xie, Z., Yan, X., Khanizadeh, S., Zhang, S., and Tao, S. (2023). Multi-omics analyses reveal stone cell distribution pattern in pear fruit. Plant J. **113**:626–642.

Gu, C., Pei, M.-S., Guo, Z.-H., Wu, L., Qi, K.-J., Wang, X.-P., Liu, H., Liu, Z., Lang, Z., and Zhang, S. (2024). Multi-omics provide insights into the regulation of DNA methylation in pear fruit metabolism. Genome Biol. **25**:70.

He, L., Ma, X., Li, Z., Jiao, Z., Li, Y., and Ow, D.W. (2016). Maize OXIDATIVE STRESS2 homologs enhance cadmium tolerance in Arabidopsis through activation of a putative SAM-dependent methyltransferase gene. Plant Physiol. **171**:1675–1685.

Huang, Y., Wang, Z., Schmidt, M.A., Su, H., Xiong, L., and Zhang, J. (2024). DEGAP: Dynamic elongation of a genome assembly path. Briefings Bioinf. **25**:bbae194.

Huerta-Cepas, J., Szklarczyk, D., Heller, D., Hernández-Plaza, A., Forslund, S.K., Cook, H., Mende, D.R., Letunic, I., Rattei, T., Jensen, L.J., et al. (2019). eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. **47**:D309–D314.

Humphreys, J.M., and Chapple, C. (2002). Rewriting the lignin roadmap. Curr. Opin. Plant Biol. **5**:224–229.

Jiang, L., Li, X., Lyu, K., Wang, H., Li, Z., Qi, W., Zhang, L., and Cao, Y. (2025). Rosaceae phylogenomic studies provide insights into the evolution of new genes. Horticultural Plant Journal **11**:389–405.

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

Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. **37**:540–546.

Kong, Z., Li, M., Yang, W., Xu, W., and Xue, Y. (2006). A novel nuclear-localized CCCH-type zinc finger protein, OsDOS, is involved in delaying leaf senescence in rice. Plant Physiol. **141**:1376–1388.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. **27**:722–736.

Lan, Y., Chen, F., Zhang, K., Wang, L., Zhang, S., Wu, M., and Xiang, Y. (2023). The CCCH zinc finger protein PeC3H74 of Moso bamboo (Phyllostachys edulis) positively regulates drought and salinity tolerances in transgenic plants. Ind. Crop. Prod. **206**:117683.

Li, D., Yang, J., Pak, S., Zeng, M., Sun, J., Yu, S., He, Y., and Li, C. (2022a). PuC3H35 confers drought tolerance by enhancing lignin and proanthocyanidin biosynthesis in the roots of Populus ussuriensis. New Phytol. **233**:390–408.

Li, G., Song, C., Manzoor, M.A., Li, D., Cao, Y., and Cai, Y. (2023). Functional and kinetics of two efficient phenylalanine ammonia lyase from Pyrus bretschneideri. BMC Plant Biol. **23**:612.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv. https://doi.org/10.48550/arXiv.1303.3997.

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

Li, J., Zhang, M., Li, X., Khan, A., Kumar, S., Allan, A.C., Lin-Wang, K., Espley, R.V., Wang, C., Wang, R., et al. (2022b). Pear genetics: recent advances, new prospects, and a roadmap for the future. Hortic. Res. **9**:uhab040.

Lin, Y., Ye, C., Li, X., Chen, Q., Wu, Y., Zhang, F., Pan, R., Zhang, S., Chen, S., Wang, X., et al. (2023). quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. Hortic. Res. **10**:uhad127.

Linsmith, G., Rombauts, S., Montanari, S., Deng, C.H., Celton, J.-M., Guérif, P., Liu, C., Lohaus, R., Zurn, J.D., Cestaro, A., et al. (2019). Pseudo-chromosome–length genome assembly of a double haploid "Bartlett" pear (Pyrus communis L.). GigaScience **8**:giz138.

Liu, D., Xue, Y., Wang, R., et al. (2024). PbrMYB4, a R2R3-MYB protein, regulates pear stone cell lignification through activation of lignin biosynthesis genes. Horticultural Plant Journal **11**:105–122.

Mamat, A., Tusong, K., Xu, J., Yan, P., Mei, C., and Wang, J. (2021). Integrated transcriptomic and proteomic analysis reveals the

complex molecular mechanisms underlying stone cell formation in Korla pear. Sci. Rep. **11**:7688.

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

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., Von Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. **37**:1530–1534.

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

Ou, S., Chen, J., and Jiang, N. (2018b). Assessing genome assembly quality using the LTR Assembly Index (LAI). Nucleic Acids Res. **46**:e126.

Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. **11**:1650–1667.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.-C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. **33**:290–295.

Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. **21**:1–27.

Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Methods **17**:155–158.

Seok, H.-Y., Woo, D.-H., Park, H.-Y., Lee, S.-Y., Tran, H.T., Lee, E.-H., Vu Nguyen, L., and Moon, Y.-H. (2016). AtC3H17, a non-tandem CCCH zinc finger protein, functions as a nuclear transcriptional activator and has pleiotropic effects on vegetative development, flowering and seed development in Arabidopsis. Plant Cell Physiol. **57**:603–615.

Shi, X., Cao, S., Wang, X., Huang, S., Wang, Y., Liu, Z., Liu, W., Leng, X., Peng, Y., Wang, N., et al. (2023). The complete reference genome for grapevine (Vitis vinifera L.) genetics and breeding. Hortic. Res. **10**:uhad061.

Shirasawa, K., Itai, A., and Isobe, S. (2021). Chromosome-scale genome assembly of Japanese pear (Pyrus pyrifolia) variety 'Nijisseiki. DNA Res. **28**:dsab001.

Shumate, A., and Salzberg, S.L. (2021). Liftoff: accurate mapping of gene annotations. Bioinformatics **37**:1639–1643.

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

Song, B., Yu, J., Li, X., Li, J., Fan, J., Liu, H., Wei, W., Zhang, L., Gu, K., Liu, D., et al. (2024a). Increased DNA methylation contributes to the early ripening of pear fruits during domestication and improvement. Genome Biol. **25**:87.

Song, Z., Chen, H., Lai, X., Wang, L., Yao, Y., Qin, J., Pang, X., Zhu, H., Chen, W., Li, X., et al. (2024b). The Zinc Finger Protein MaCCCH33-Like2 Positively Regulates Banana Fruit Ripening by Modulating Genes in Starch and Cell Wall Degradation. Plant Cell Physiol. **65**:49–67.

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

Stanke, M., Keller, O., Gunduz, I., Hayes, A., Waack, S., and Morgenstern, B. (2006). AUGUSTUS: ab initio prediction of alternative transcripts. Nucleic Acids Res. **34**:W435–W439.

Sun, M., Yao, C., Shu, Q., He, Y., Chen, G., Yang, G., Xu, S., Liu, Y., Xue, Z., and Wu, J. (2023). Telomere-to-telomere pear (Pyrus pyrifolia) reference genome reveals segmental and whole genome duplication driving genome evolution. Hortic. Res. **10**:uhad201.

Suyama, M., Torrents, D., and Bork, P. (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. **34**:W609–W612.

Tang, Y., Liu, X., Wang, J., Li, M., Wang, Q., Tian, F., Su, Z., Pan, Y., Liu, D., Lipka, A.E., et al. (2016). GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genome **9**:plantgenome2015.

Vanholme, R., Cesarino, I., Rataj, K., Xiao, Y., Sundin, L., Goeminne, G., Kim, H., Cross, J., Morreel, K., Araujo, P., et al. (2013). Caffeoyl shikimate esterase (CSE) is an enzyme in the lignin biosynthetic pathway in Arabidopsis. Science **341**:1103–1106.

Wang, D., Guo, Y., Wu, C., Yang, G., Li, Y., and Zheng, C. (2008). Genome-wide analysis of CCCH zinc finger family in Arabidopsis and rice. BMC Genom. **9**:1–20.

Wang, H., Zhang, Y., Feng, X., Hong, J., Aamir Manzoor, M., Zhou, X., Zhou, Q., and Cai, Y. (2024). Transcription factor PbMYB80 regulates lignification of stone cells and undergoes RING finger protein PbRHY1-mediated degradation in pear fruit. J. Exp. Bot. **75**:883–900. https://doi.org/10.1093/jxb/erad434.

Wang, R., Xue, Y., Fan, J., Yao, J.-L., Qin, M., Lin, T., Lian, Q., Zhang, M., Li, X., Li, J., et al. (2021). A systems genetics approach reveals PbrNSC as a regulator of lignin and cellulose biosynthesis in stone cells of pear fruit. Genome Biol. **22**:1–23.

Wu, J., Wang, Z., Shi, Z., Zhang, S., Ming, R., Zhu, S., Khan, M.A., Tao, S., Korban, S.S., and Wang, H. (2013). The genome of the pear (Pyrus bretschneideri Rehd.). Genome Res. **23**:396–408.

Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., Luo, F., and Xie, Z. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat. Methods **14**:1072–1074.

Xie, Z., Lin, W., Yu, G., Cheng, Q., Xu, B., and Huang, B. (2019). Improved cold tolerance in switchgrass by a novel CCCH-type zinc finger transcription factor gene, PvC3H72, associated with ICE1–CBF–COR regulon and ABA-responsive genes. Biotechnol. Biofuels **12**:224–311.

Xu, G.-C., Xu, T.-J., Zhu, R., Zhang, Y., Li, S.-Q., Wang, H.-W., and Li, J.-T. (2019). LR_Gapcloser: a tiling path-based gap closer that uses long reads to complete genome assembly. GigaScience **8**:giy157.

Xu, M., Guo, L., Gu, S., Wang, O., Zhang, R., Peters, B.A., Fan, G., Liu, X., Xu, X., Deng, L., et al. (2020). TGS-GapCloser: a fast and accurate gap closer for large genomes with low coverage of error-prone long reads. GigaScience **9**:giaa094.

Xu, S., Sun, M., Yao, J.L., Liu, X., Xue, Y., Yang, G., Zhu, R., Jiang, W., Wang, R., Xue, C., et al. (2023). 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.

Xue, C., Yao, J.L., Qin, M.F., Zhang, M.Y., Allan, A.C., Wang, D.F., and Wu, J. (2019). PbrmiR397a regulates lignification during stone cell development in pear fruit. Plant Biotechnol. J. **17**:103–117.

Xue, Y., Shan, Y., Yao, J.-L., Wang, R., Xu, S., Liu, D., Ye, Z., Lin, J., Li, X., Xue, C., et al. (2023). The transcription factor PbrMYB24 regulates lignin and cellulose biosynthesis in stone cells of pear fruits. Plant Physiol. **192**:1997–2014.

Yan, C., Yin, M., Zhang, N., Jin, Q., Fang, Z., Lin, Y., and Cai, Y. (2014). Stone cell distribution and lignin structure in various pear varieties. Sci. Hortic. **174**:142–150.

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

Zeng, X., Yi, Z., Zhang, X., Du, Y., Li, Y., Zhou, Z., Chen, S., Zhao, H., Yang, S., Wang, Y., et al. (2024). Chromosome-level scaffolding of haplotype-resolved assemblies using Hi-C data without reference genomes. Nat. Plants **10**:1184–1200.

Zhang, H., Gao, X., Zhi, Y., Li, X., Zhang, Q., Niu, J., Wang, J., Zhai, H., Zhao, N., Li, J., et al. (2019). A non-tandem CCCH-type zinc-finger protein, IbC3H18, functions as a nuclear transcriptional activator and enhances abiotic stress tolerance in sweet potato. New Phytol. **223**:1918–1936.

Zhang, J., Kudrna, D., Mu, T., Li, W., Copetti, D., Yu, Y., Goicoechea, J.L., Lei, Y., and Wing, R.A. (2016). Genome puzzle master (GPM): an integrated pipeline for building and editing pseudomolecules from fragmented sequences. Bioinformatics **32**:3058–3064.

Zhang, J., Cheng, X., Jin, Q., Su, X., Li, M., Yan, C., Jiao, X., Li, D., Lin, Y., and Cai, Y. (2017). Comparison of the transcriptomic analysis between two Chinese white pear (Pyrus bretschneideri Rehd.) genotypes of different stone cells contents. PLoS One **12**:e0187114.

Zhang, J., Li, J., Xue, C., Wang, R., Zhang, M., Qi, K., Fan, J., Hu, H., Zhang, S., and Wu, J. (2021a). The variation of stone cell content in 236 germplasms of sand pear (Pyrus pyrifolia) and identification of related candidate genes. Horticultural Plant Journal **7**:108–116.

Zhang, M.-Y., Xue, C., Hu, H., Li, J., Xue, Y., Wang, R., Fan, J., Zou, C., Tao, S., Qin, M., et al. (2021b). Genome-wide association studies provide insights into the genetic determination of fruit traits of pear. Nat. Commun. **12**:1144.

Zhu, Y., Wang, Y., Jiang, H., Liu, W., Zhang, S., Hou, X., Zhang, S., Wang, N., Zhang, R., Zhang, Z., et al. (2023). Transcriptome analysis reveals that PbMYB61 and PbMYB308 are involved in the regulation of lignin biosynthesis in pear fruit stone cells. Plant J. **116**:217–233.