Check for updates

Beta-Amylase and Phosphatidic Acid Involved in Recalcitrant Seed Germination of Chinese Chestnut

Yang Liu[1,2†], Yu Zhang[1†], Yi Zheng[3], Xinghua Nie[1], Yafeng Wang[1], Wenjie Yu[1], Shuchai Su[2], Qingqin Cao[1], Ling Qin[1*] and Yu Xing[1*]

[1] Beijing Advanced Innovation Center for Tree Breeding by Molecular Design, College of Plant Science and Technology, Beijing University of Agriculture, Beijing, China, [2] Key Laboratory for Silviculture and Conservation of Ministry of Education, College of Forestry, Beijing Forestry University, Beijing, China, [3] Bioinformatics Center, Beijing University of Agriculture, Beijing, China

Chinese chestnut (*Castanea mollissima*), a species with recalcitrant seeds, is an important source of nuts and forest ecosystem services. The germination rate of recalcitrant seeds is low in natural habitats and decreases under conditions of desiccation and low temperature. The germination rate of cultivated Chinese chestnut seeds is significantly higher than that of wild seeds. To explore the reasons for the higher germination rate of cultivated seeds in Chinese chestnut, 113,524 structural variants (SVs) between the wild and cultivated Chinese chestnut genomes were detected through genome comparison. Genotyping these SVs in 60 Chinese chestnut accessions identified allele frequency changes during Chinese chestnut domestication, and some SVs are overlapping genes for controlling seed germination. Transcriptome analysis revealed downregulation of the abscisic acid synthesis genes and upregulation of the beta-amylase synthesis genes in strongly selected genes of cultivated seeds. On the other hand, hormone and enzyme activity assays indicated a decrease in endogenous ABA level and an increase in beta-amylase activity in cultivated seeds. These results shed light on the higher germination rate of cultivated seeds. Moreover, phosphatidic acid synthesis genes are highly expressed in seed germination stages of wild Chinese chestnut and may play a role in recalcitrant seed germination. These findings provide new insight into the regulation of wild seed germination and promote natural regeneration and succession in forest ecosystems.

Keywords: SVs, seed germination, recalcitrant seeds, amylase, phosphatidic acid

INTRODUCTION

Seeds act as an important vehicle by which angiosperms to disperse offspring, representing a key stage in the plant life cycle. Seeds are generally classified as "recalcitrant" and "orthodox" based on their tolerance of desiccation (Xia et al., 2014). Recalcitrant seeds die rapidly when stored under desiccation and low-temperature conditions (Walters et al., 2013). Numerous forest species, including *Castanea* (Vieitez et al., 2011), *Quercus* (Sghaier-Hammami et al., 2020), and *Aesculus* (Obroucheva et al., 2016), have recalcitrant seeds. Fagaceae species are the main trees with recalcitrant seeds used to maintain forest ecosystems and play important roles in the survival of

wild animals and forest ecological restoration (Walters et al., 2013; Wilf et al., 2019). However, most Fagaceae plant seeds have a low germination rate in the wild, and thus, increasing their germination rate is necessary for afforestation and ecological protection.

To improve the germination rate of recalcitrant seeds, their germination mechanism should be clarified. Several studies have focused on the sensitivity to dehydration and low temperature of recalcitrant seeds in terms of cell biology, physiology and molecular biology (Romero-Rodríguez et al., 2018; Li et al., 2021; Xia and Zhu, 2021), and the findings contribute to the conservation of recalcitrant seeds and the natural regeneration and ecological protection of forests. Seed germination is a complex adaptive trait of higher plants that is influenced by a large number of genes, endogenous hormones and environmental factors (Shu et al., 2016). In species with recalcitrant seeds, high levels of gibberellins (GAs) and low levels of ABA are observed in mature seeds, and the levels of auxin (IAA) and GAs increase during the seed germination process (Romero-Rodríguez et al., 2018). Moreover, exogenous ABA treatments of recalcitrant seeds results in a reduced germination rate (Wang et al., 2019). Genes related to hormone biosynthesis and signal transduction are significantly differentially expressed under desiccation (Dussert et al., 2018; Kijak and Ratajczak, 2020). The *ZEAXANTHIN EPOXIDASE* (*ZEP*), *NINE-CIS-EPOXYCAROTENOID DIOXYGENASE* (*NCED*), *PYRABACTIN RESISTANCE 1* (*PYR1*) genes related to ABA and *YUCCA, ADP-RIBOSYLATION FACTOR* (*ARF*), genes related to IAA have been found to regulate the seed desiccation tolerance acquisition in *Quercus variabilis* (Li et al., 2021). Additionally, paclobutrazol interferes with GA-related gene expression to induce desiccation tolerance in the seeds of *Citrus limon* (Marques et al., 2019). Hormones also regulate seed germination by controlling metabolism, and the amylase activity (Zaynab et al., 2021), lipid metabolism (Cui et al., 2020), glycerophospholipid metabolism (Chen et al., 2018) and reactive oxygen species (ROS) metabolism (Romero-Rodríguez et al., 2018) play key roles in recalcitrant seed germination. Compared to ungerminated seeds, germinated seeds show higher beta-amylase activity (Zaynab et al., 2021). Moreover, there is a close association between reduced seed viability and increasing phosphatidic acid (PA), and glycerophospholipid metabolism genes regulated desiccation sensitivity in recalcitrant seeds (Chen et al., 2018; Li et al., 2021). Nevertheless, a low germination rate of recalcitrant seeds in Fagaceae has not yet been addressed. Most Fagaceae plants grow in the wild, and the reproduction *via* seedlings in afforestation practices mainly relies on seeds. With the cultivation and utilization of forest trees, some cultivated species with high seed germination rates should represent an important source of seedlings for ecological restoration.

Fortunately, there are both wild and cultivated types of Chinese chestnut, which is in the *Castanea* genus. Compared with wild Chinese chestnut seeds, cultivated Chinese chestnut seeds have a higher levels of seed germination capacity under natural conditions. Chinese chestnut (*Castanea mollissima*) has a long history of commercial nut production and is widely cultivated in 26 provinc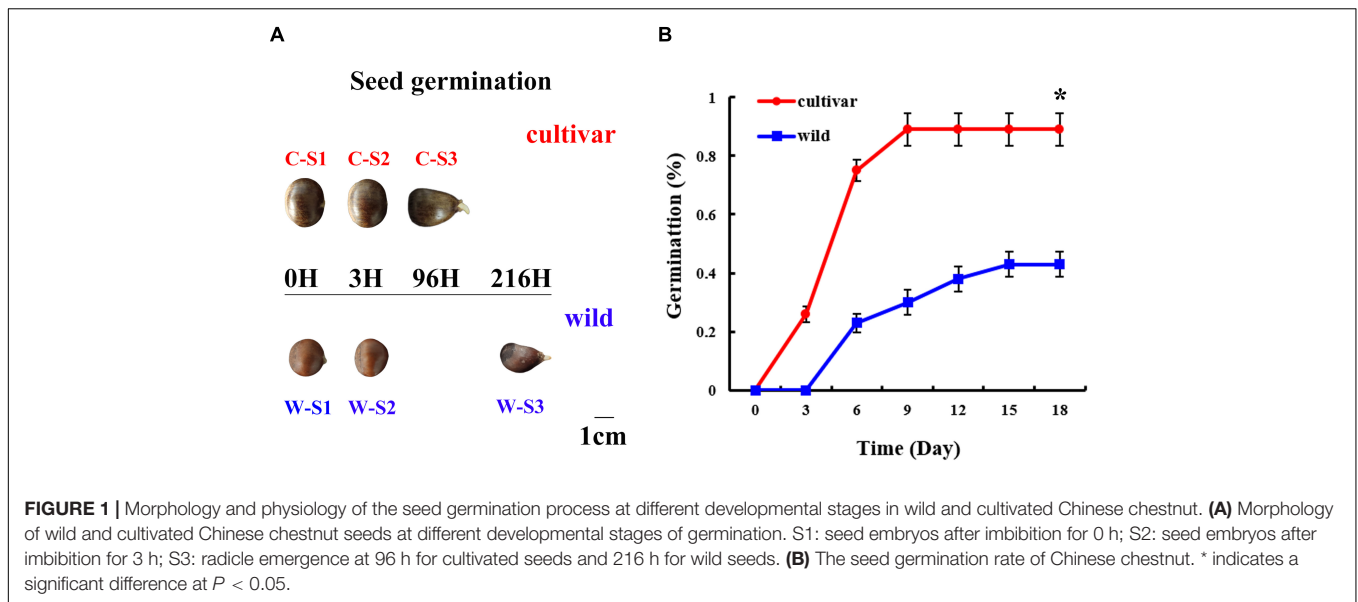es in China. As starchy nuts, the seeds (dry weight) contain 46～64% starch, 12～22% soluble sugars and 0.27～0.64% lipids (Chen et al., 2017). As the germination characteristics of cultivated seeds are different from those of wild seeds of Chinese chestnut, we explored the mechanisms underlying the seed germination characteristics of Chinese chestnut to provide new insights into recalcitrant seeds.

In this study, we used omics to explore the molecular mechanism of the difference in the germination of recalcitrant seeds of wild and cultivated Chinese chestnut. Recently, wild (HBY_2) and cultivated (N11_1) Chinese chestnut genome assemblies were generated by using Pacific Biosciences single-molecule sequencing technology (Xing et al., 2019; Wang J. et al., 2020), and such Chinese chestnut genomes will provide a platform for comparative genomics analysis to characterize functional and structural features. To identify the differential gene loci related to seed germination of wild and cultivated Chinese chestnut, structural variants (SVs) between wild and cultivated plants were detected through genome comparison and genotyping, and these SV overlapping genes may control seed germination in Chinese chestnut. Comparative transcriptome analysis revealed the related genes involved in the differences in wild and cultivated seed germination. The results help lay a theoretical foundation for improving the recalcitrant seed germination and promoting natural regeneration and succession in forest ecosystems.

MATERIALS AND METHODS

Germination Assay, Endogenous Hormone and Transcriptome Sequencing of Chinese Chestnut Seeds

Mature Chinese chestnut seeds were harvested from a wild tree (HBY_2) and a tree of cultivar 'Jingshuhong' growing in an orchard in Beijing city, China. Undamaged Chinese chestnut seeds were selected for germination in darkness at 22°C. The seed germination assay included wild and cultivar groups. A total of 270 seeds were equally divided among three biological replicates for each group (**Figure 1A**). The wild and cultivated seeds were soaked in sterile water for a seed imbibition experiment until the seed weight did not increase. The relative water content was determined as the initial relative water content minus the relative water content that after desiccation. Then, the seeds were planted in a planter (50 cm × 30 cm × 40 cm) containing sterilized sand in a climate chamber with a 16 h light (25°C): 8 h dark (25°C) cycle and 60% relative humidity. The seeds were considered to have successfully germinated when the emerged radicle was at least 5 mm long and the germination rate reached 30% (Roach et al., 2010). According to the germination assay, seed embryos were sampled at 0 (S1) and 3 h of imbibition (S2) and when the seed germination rate reached 30% (S3: 96 h of imbibition for 'Jingshuhong' seeds, 216 h for wild seeds), respectively. Three biological replicates were tested for each sample. Endogenous hormone [GA$_3$, GA$_4$, ABA, IAA, zeatin riboside (ZR), isopentenyladenoside (IPA), brassinosteroids (BR) and methyl jasmonate (MeJA)] contents of these samples were

FIGURE 1 | Morphology and physiology of the seed germination process at different developmental stages in wild and cultivated Chinese chestnut. **(A)** Morphology of wild and cultivated Chinese chestnut seeds at different developmental stages of germination. S1: seed embryos after imbibition for 0 h; S2: seed embryos after imbibition for 3 h; S3: radicle emergence at 96 h for cultivated seeds and 216 h for wild seeds. **(B)** The seed germination rate of Chinese chestnut. * indicates a significant difference at $P < 0.05$.

determined by enzyme-linked immunosorbent assays (Deng et al., 2008), and each sample was measured in parallel three times. Alpha-amylase and beta-amylase activities were measured as described previously (Gimbi and Kitabatake, 2002). Then, the RNA of the samples was extracted from the seed embryos for transcriptome sequencing using a Plant RNA Kit (OMEGA, GA, United States), and libraries from the RNA samples were sequenced using the Illumina HiSeq 4000 platform (Illumina, CA, United States) at Novogene Bioinformatics Technology Co., Ltd. (Tianjing, China). Clean data were aligned to the Chinese chestnut V2 and N11_1 genome using STAR (version 2.7.1a). Counts were generated using HTSeq, and DESeq2 was used to identify differentially expressed genes (DEGs).

Genome Assembly and Chinese Chestnut Genome Annotation

Recalcitrant seed germination is a complex process regulated by various factors. To reveal the basis of the higher seed germination rate of cultivated than wild Chinese chestnut, we used comparative genome analysis to identify related genes. Wild and cultivated genomes of Chinese chestnut have been assembled, but the wild genome still lacks a high-quality chromosomal assembly and annotation version. Therefore, we assembled a chromosome-scale genome of wild Chinese chestnut. Fresh leaf tissue from Chinese chestnut (HBY_2) was used to construct Hi-C libraries, which were sequenced using the Illumina HiSeq PE150 platform (Illumina, CA, United States) at Novogene Bioinformatics Technology Co., Ltd. (Tianjing, China). We utilized Burrows–Wheeler Aligner (BWA) (version 0.7.17-r1188) (Li and Durbin, 2009) to compare the raw Hi-C reads with the draft assembled sequence (Xing et al., 2019); and low-quality reads were removed by SAMtools (version 1.9) (Li et al., 2009). Valid interaction pairs were then applied to build interaction matrices and scale up the major contigs to chromosome-scale scaffolds with LACHESIS (version 201701) (Burton et al., 2013).

For annotation analysis, different tissue types, including leaves, mixed floral buds, leaf buds, stems, female flowers, staminate catkins, nuts, burs, inner and outer shells, globular stage embryos and roots, were collected and immediately frozen in liquid nitrogen. Equal amounts of RNA from these tissues were pooled and used for single-molecule real-time (SMRT) sequencing. Full-length cDNA was synthesized using SMARTer PCR cDNA Synthesis Kit (Clontech Laboratories, CA, United States), and the library was subjected to SMRT sequencing using the PacBio Sequel platform. At the same time, we collected 44 Illumina-based RNA-seq data sets from various tissues, including male flowers, female flowers, leaves, mixed floral buds, male floral buds, nut developmental stages, embryo developmental stages, ovule developmental stages, stems and roots.

Homology alignment and *de novo* annotation were used to identify repetitive elements in the Chinese chestnut genome. For homology-based prediction, the Repbase TE library[1] was utilized to search against the Chinese chestnut genome using RepeatMasker (version 4.0) (Tarailo-Graovac and Chen, 2009) and RepeatProteinMask[2] with default parameters. For *de novo* prediction, a *de novo* repetitive element database was constructed by LTR_FINDER[3], RepeatScout[4] and RepeatModeler[5] with default parameters. A hybrid approach of *de novo* prediction, homology-based prediction, and transcriptome-based prediction approaches was applied to identify protein-coding regions and to predict genes. For homology-based prediction, Exonerate (version 2.47.3) (Slater and Birney, 2005) was employed for mapping against *C. mollissima*, with protein sequences from *Arabidopsis thaliana*, *Oryza sativa*, *Vitis vinifera*,

[1]http://www.girinst.org/repbase/
[2]http://www.repeatmasker.org/
[3]http://tlife.fudan.edu.cn/ltr_finder/
[4]http://www.repeatmasker.org/
[5]http://www.repeatmasker.org/RepeatModeler.html

Populus trichocarpa, *Malus domestica*, *Cucumis sativus*, *Juglans microcarpa* × *Juglans regia*, *Casuarina equisetifolia* and *Quercus robur*. Homologous proteins were then aligned to the Chinese chestnut genome by TBLASTN (Altschul et al., 1997), which were then aligned to the assembly by GeneWise (v2.4.1) (Birney et al., 2004) to predict gene structures. For *de novo* prediction, Augustus (version 3.0.2) (Stanke et al., 2006), GlimmerHMM (version 3.0.2) (Majoros et al., 2004), SNAP (version 11-29-2013) (Korf, 2004), GeneID (version 1.4) (Parra et al., 2000), and GENSCAN (version 1.0) (Salamov and Solovyev, 2000) were used to predict coding regions in the repeat-masked genome. For transcriptome-based prediction, Cufflinks (version 2.1.1) (Trapnell et al., 2012) was used to link transcripts from the TopHat results to gene models, and then PASA (version 2.3.3) (Campbell et al., 2006) was applied on the basis of the assembled RNA-seq unigenes. Finally, an integrated gene set from the above three methods was produced by EVidenceModeler (EVM version 1.1.1) (Haas et al., 2008).

Gene Functional Annotation

In a previous study, a standard pipeline was established for thoroughly annotating predicted protein-coding genes (Zheng et al., 2019). In brief, the predicted gene sequences were compared to the *Arabidopsis* protein (TAIR), NCBI non-redundant (nr), and UniProt (Swiss-Prot and TrEMBL) databases using the BLASTP command of DIAMOND (Buchfink et al., 2015), with an *E*-value cutoff of $1e^{-5}$. Furthermore, InterProScan was used to identify functional domains in all protein sequences by comparing them to the InterPro database (Mitchell et al., 2019). To assign GO terms for each protein-coding gene, the BLASTP results from the nr database and the discovered InterPro domains were input into the Blast2GO pipeline (Conesa and Götz, 2008). The BLASTP leads obtained from searches against the TAIR and UniProt databases were input into the AHRD program[6] to acquire accurate, succinct and useful gene functional summaries. The KEGG pathways encoded by each of the *Castanea* genomes were also predicted by KEGG Automatic Annotation Server[7]. The iTAK program was employed to identify TFs and TRs from the predicted protein-coding genes and to classify them into different families (Zheng et al., 2016).

Comparative Genomic Analysis and Detection of Structural Variants Between the V2 and N11_1 Genome

Changes in SV allele frequency across distinct Chinese chestnut populations are a consequence of domestication processes such as favorable trait selection and introgression from ancestral groups. We analyzed SV allele frequency changes from wild to cultivated materials for domestication to discover SVs under selection during Chinese chestnut domestication and breeding. The relation pipelines for SV detection and genotyping in Chinese chestnut were based on the study of Wang X. et al. (2020). SVs of Chinese chestnut genomes V2 and N11_1 were

---

[6]https://github.com/groupschoof/AHRD

[7]https://www.genome.jp/tools/kaas/

identified by Minimap with the parameter "-ax asm5" (Li, 2018). The resulting alignments were analyzed using Assemblytics (v1.1) (Nattestad and Schatz, 2016) for SV identification. SV calling was based on genome comparison, and then filtering by Illumina read mapping.

Structural Variants and SNP Genotyping in the Chinese Chestnut Population

A total of 60 accessions were collected, including 28 wild and 32 cultivated samples of Chinese chestnut in China (**Supplementary Data 1**). Young leaves were immediately frozen in liquid nitrogen and stored at −80°C; genomic DNA was extracted from young leaves using the DNeasy Plant Mini Kit (Qiagen, Hilden, Germany) and used to construct sequencing libraries. Paired-end sequencing libraries with an insert size of approximately 450 bp and a read length of 150 bp were sequenced using an Illumina HiSeq 2500 sequencer. The raw reads were filtered to obtain high-quality sequences by Trimmomatic (v0.39) (Bolger et al., 2014). The high-quality reference SVs identified between the V2 and N11_1 genome were genotyped in the 60 Chinese chestnut accessions based on the protocols of Gao et al. (2019) and Wang X. et al. (2020). Clean Illumina reads for each accession were aligned to the V2 and N11_1 genome by BWA-MEM (v0.7.17) (Li and Durbin, 2009), with no more than 3% mismatches allowed. SVs in each accession were classified as V2 (same as V2), N11_1 (same as N11_1), and heterozygous (carrying both V2 and N11_1 alleles) genotypes. The allele frequency of certain SVs was computed for each group, and the significance of the frequency differences between two compared groups was established using Fisher's exact test. Raw *P* values of SVs were adjusted using the false discovery rate (FDR) approach, and SVs with corrected *P* values < 0.05 were considered as those under selection.

In addition, we employed Illumina reads matched to the V2 genome for SNP calling. Variant calling was performed independently with Sentieon (Freed et al., 2017) with default parameters. GATK (v4.1.1.0) was used for hard filtering with the settings parameters "QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < −12.5 || ReadPosRankSum < −8.0" (McKenna et al., 2010). SNPs with at least 70% genotyped accessions, with a minor allele frequency (MAF) of no less than 0.03 and overlapping with known SNPs from the V2 and N11_1 genome alignment were preserved.

Population Genetic Analysis

IQ-TREE (v2.0.3) with maximum likelihood was used to build phylogenetic trees for the Chinese chestnut accessions using the full SNPs at fourfold degenerated sites (4DTV) in 1,000 bootstrap runs with *Castanea henryi* as the outgroup (Sun et al., 2020). Principal component analysis (PCA) was performed using the program Plink (v1.90) (Purcell et al., 2007). VCFtools (Danecek et al., 2011) was used to calculate F_{ST} and π across the genome with a 15 kb window based on the SNPs. XP-CLR analyses were implemented to detect selective sweeps based on SNPs across chestnut populations (Chen et al., 2010). Correlation coefficients (r^2) were calculated for all pairs of SNPs to measure linkage

disequilibrium (LD) decay using PopLDdecay (v3.41) (Zhang et al., 2019) with default parameters.
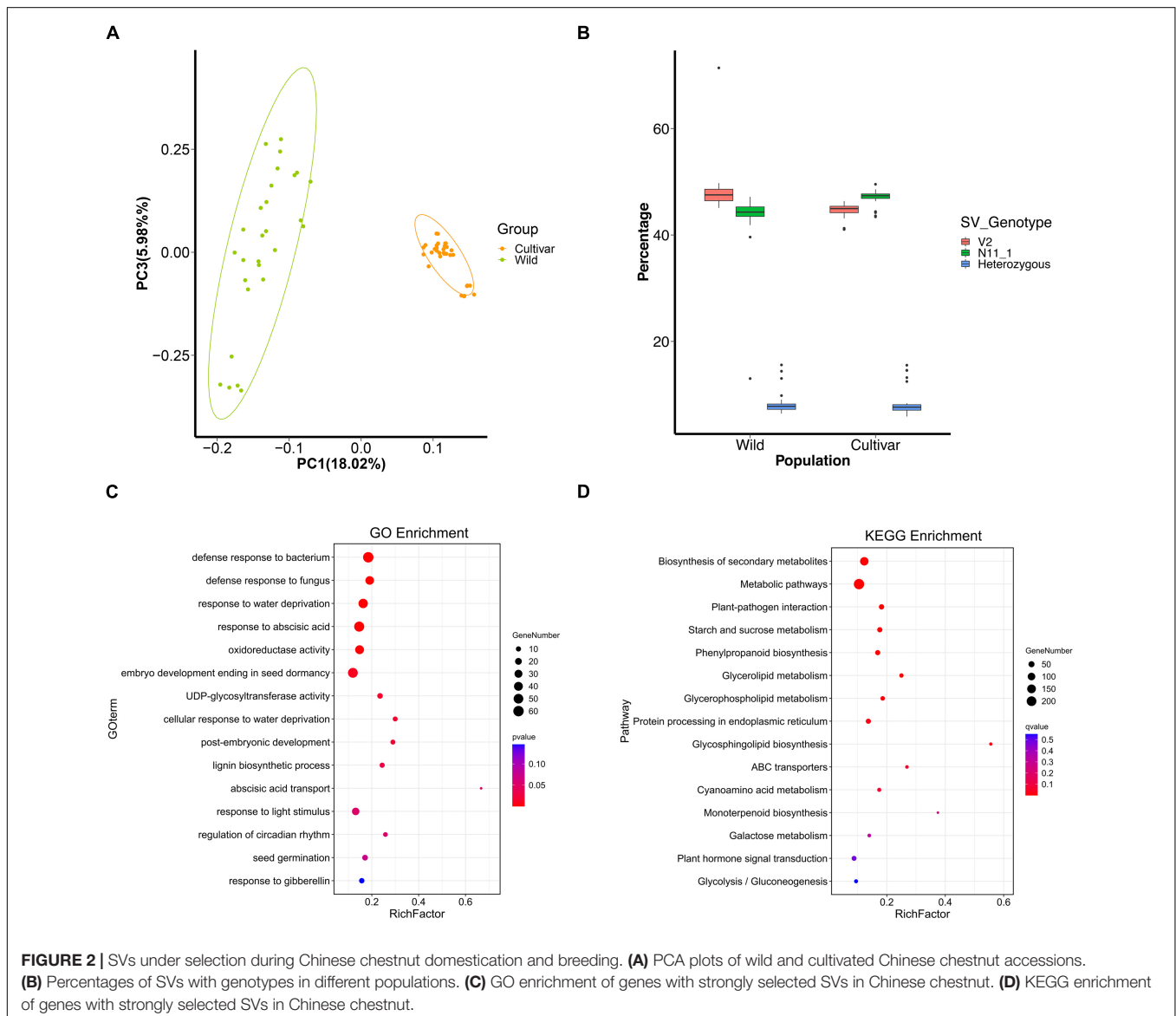
RESULTS

Physiological Features of Seed Germination in Wild and Cultivated Chinese Chestnut

As based on observation of the seed germination process, 0–12 h (0–12H) represented a rapid initial penetration phase; and 12–48 h (12–48H) represented a plateau phase in terms of water content (**Supplementary Figure 1**). There was a significant difference in the water content of seeds, with averages water contents of 46 and 52% for wild and cultivated Chinese chestnut seeds before the water uptake phase, respectively

(**Supplementary Figure 1**). Radicles started to emerge after imbibition for 96 h (4 days) and 216 h (9 days) in wild and cultivated Chinese chestnut seeds, respectively (**Figure 1A**). There were significant changes in the seed germination rate, with total germination rates of 43 and 89% for wild and cultivated Chinese chestnut seeds, respectively (**Figure 1B**). The cultivated and wild seed germination rates reached a maximum at 9 and 15 days, respectively.

Genomic Structural Variants Between Wild and Cultivated Genomes

We assembled a chromosome-scale genome of wild Chinese chestnut, and a total of 2,652,199 read pairs (90.55 Gb clean reads) were acquired by Hi-C sequencing, offering over 115-fold coverage (**Supplementary Tables 1, 2**). Then, to improve gene prediction accuracy, high-quality transcripts of 11 mixed

FIGURE 2 | SVs under selection during Chinese chestnut domestication and breeding. **(A)** PCA plots of wild and cultivated Chinese chestnut accessions. **(B)** Percentages of SVs with genotypes in different populations. **(C)** GO enrichment of genes with strongly selected SVs in Chinese chestnut. **(D)** KEGG enrichment of genes with strongly selected SVs in Chinese chestnut.
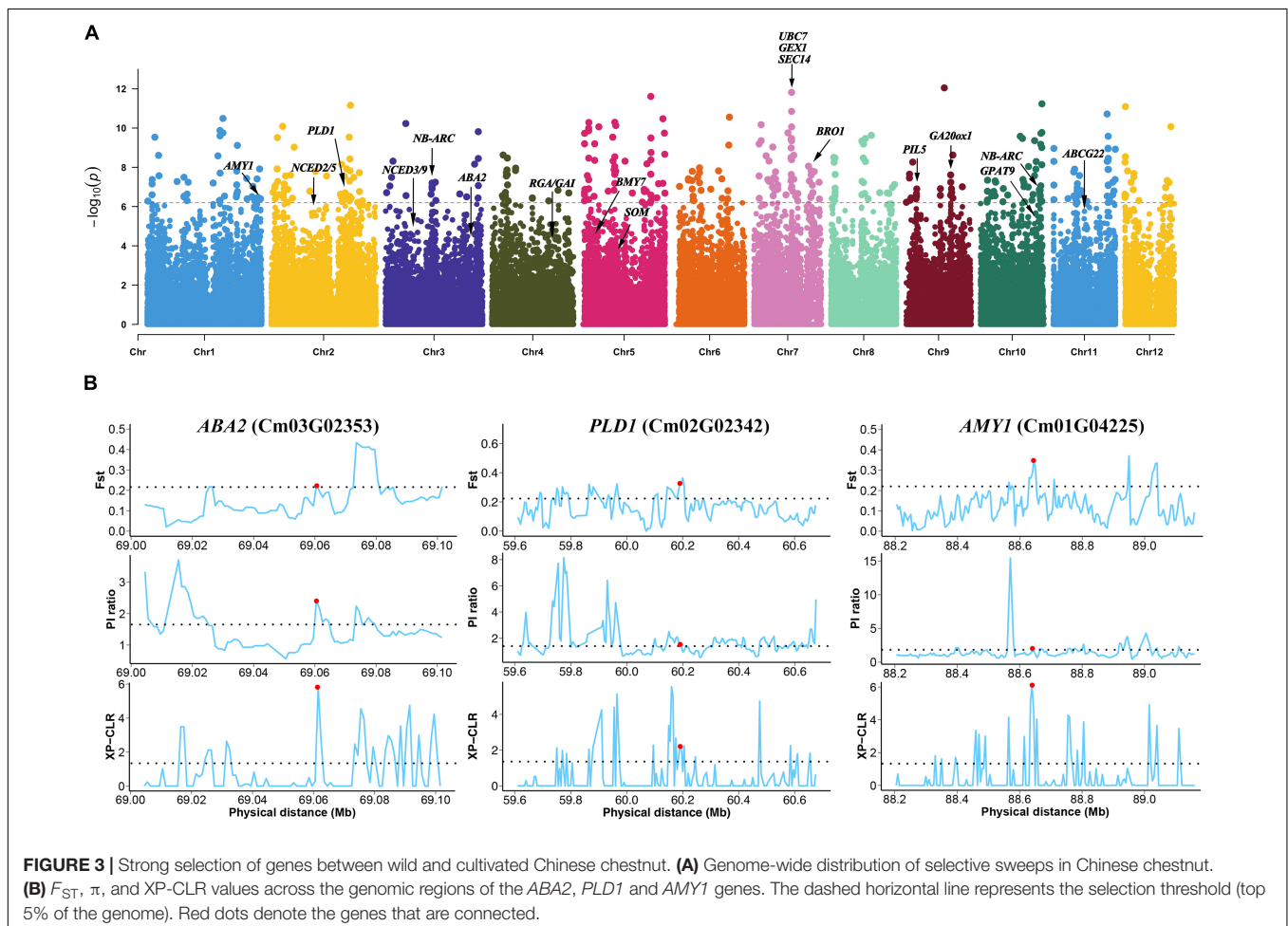
tissue types were sequenced using a single-molecule real-time sequencer from Pacific Biosciences (**Supplementary Table 3**). Finally, a total of 33,991 protein-coding genes were aligned to the wild Chinese chestnut genome (V2) (**Supplementary Figure 2** and **Supplementary Tables 4, 5**). The Chinese chestnut V2 genome assembly can be accessed at *Castanea* Genome Database (CGD[8]). Indels were discovered through direct comparison of the genomes of wild (V2) and cultivated (N11_1) Chinese chestnut with mapping of PacBio long reads. A total of 113,524 indels were detected, ranging in size from 10 to 20,094 bp, including intergenic, intronic, upstream, downstream, and exonic areas (**Supplementary Data 2** and **Supplementary Table 6**). Most of the indels are short, with 87.7% being less than 100 bp and only 0.6% exceeding 10 kb (**Supplementary Figure 3**). To further identify the SVs under strong environmental and artificial selection, 113,524 indels were genotyped in the Chinese chestnut population to investigate allele frequency changes, including 28 wild and 32 cultivated accessions (**Supplementary Data 1**). The short read sequencing data yielded 796 Gb of high-quality clean data, and the sequencing depths ranged from 10× to 42× in the wild and cultivated populations of Chinese chestnut. Genotyping

the SVs in two reference genomes using Illumina short reads further supported their high specificity (**Supplementary Table 6**). Phylogenetic analysis and PCA clearly separated the wild and cultivated groups of Chinese chestnut (**Figure 2A** and **Supplementary Figure 4**). LD values were also calculated among wild and cultivated populations, and higher LD was found in the cultivated population (**Supplementary Figure 5**).

Selection of Structural Variants for Seed Germination

Structural variant loci with the V2 alleles were common in the wild population, accounting for 48.46% of the SVs in each accession, whereas 43.48% of SV loci showed the homozygous N11_1 genotype (**Figure 2B**). The allele frequencies of 4,892 SVs were significantly differentially changed between the wild and cultivated populations of Chinese chestnut (**Supplementary Data 4**), and these indels might affect 4,673 genes in the V2 genome. These genes are significantly involved in biosynthesis of secondary metabolites, plant-pathogen interaction, starch and sucrose metabolism, and glycerophospholipid metabolism according to KEGG enrichment function (**Figure 2D**), and response to abscisic acid, response to water deprivation, and seed germination according to GO
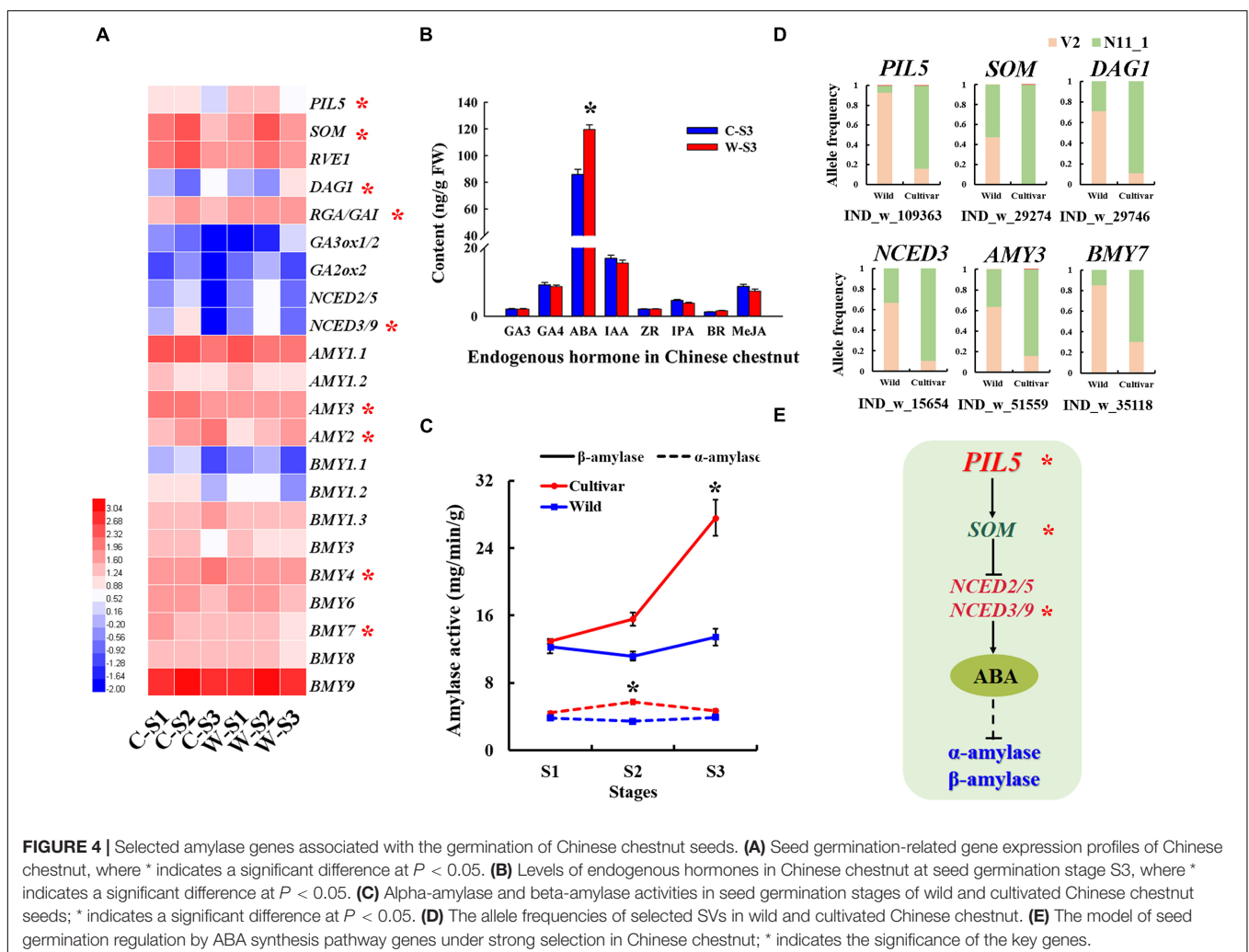

FIGURE 3 | Strong selection of genes between wild and cultivated Chinese chestnut. **(A)** Genome-wide distribution of selective sweeps in Chinese chestnut. **(B)** F_{ST}, π, and XP-CLR values across the genomic regions of the *ABA2*, *PLD1* and *AMY1* genes. The dashed horizontal line represents the selection threshold (top 5% of the genome). Red dots denote the genes that are connected.

enrichment function (**Figure 2C** and **Supplementary Data 3, 5**). These genes were also selected at the sequence level using F_{ST}, nucleotide diversity (π), and the cross-population composite likelihood ratio (XP-CLR) (**Figure 3**).

Interestingly, some genes involved in plant hormone signal transduction and starch and sucrose metabolism were detected, and these genes may influence the differences in seed germination between wild and cultivated Chinese chestnut. The regions included a 12 bp insertion in the exon of *PHYTOCHROME INTERACTING FACTOR 3-LIKE 5* (*PIL5*, Cm09G00427), and an 18 bp deletion in the intergenic region of *SOMNUS* (*SOM*, Cm05G01513), *ABA DEFICIENT 2* (*ABA2*, Cm03G02353) for ABA biosynthesis and *GIBBERELLIN 20 OXIDASE 1* (*GA20ox1*, Cm09G01563) for GA biosynthesis in plant hormone signal transduction and *ALPHA-AMYLASE-LIKE 1* (*AMY1*, Cm01G04225) and *BETA-AMYLASE 7* (*BMY7*, Cm05G00315) for starch degradation *via* alpha-amylase and beta-amylase activities (**Figure 3B**). Additionally, some genes involved in glycerophospholipid metabolism were detected in wild and cultivated Chinese chestnut, including a 17 bp deletion in phospholipase D (*PLD1*, Cm02G02342) and a 39 bp

deletion in *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 9* (*GPAT9*, Cm10G02185), which are involved in activating the synthesis of phosphatidic acid. The *ARABIDOPSIS THALIANA ATP-BINDING CASSETTE G22* (*ABCG22*, Cm11G01020) gene, an ABC transporter gene, is involved in drought susceptibility in seed germination. Moreover, the 1.61 Mb genomic region with a high selection score contains four genes on chromosome 7 (**Figure 3A**), including *UBIQUITIN CARRIER PROTEIN 7* (*UBC7*, Cm07G01553) and *SECRETION 14* (*SEC14*, Cm07G01557), which are involved in plant responses to multiple stress conditions (Ascencio-Ibáñez et al., 2008; Feng et al., 2020). *GAMETE EXPRESSED PROTEIN 1* (*GEX1*, Cm07G01555) contributes to gametophyte development (Alandete-Saez et al., 2011), suggesting strong selection of this region during domestication. In addition, there are several regions of genes involved in plant pathogen interactions, including a 33 bp deletion in the promoter of *NB-ARC* (Cm10G02154) and a 17 bp deletion in *NB-ARC* (Cm03G01169), were detected. These genes may contribute to differences in disease resistance and responses to stress conditions between wild and cultivated Chinese chestnut. The results demonstrated a common
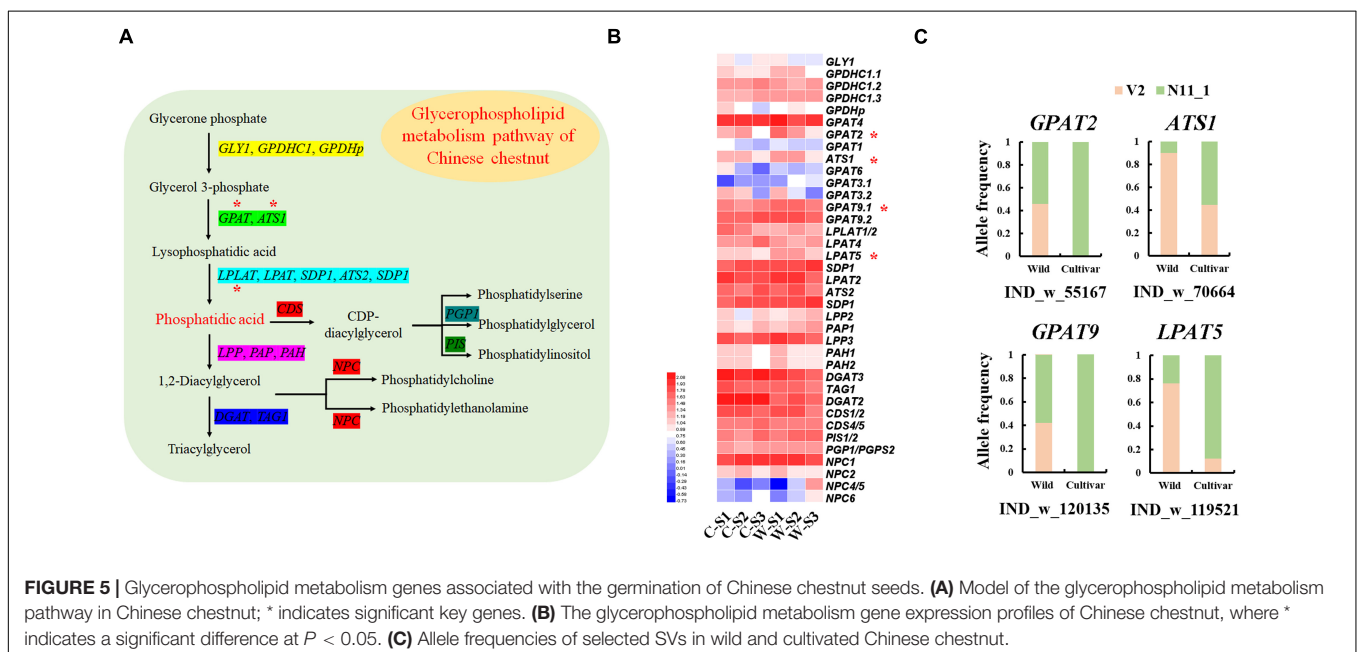

FIGURE 4 | Selected amylase genes associated with the germination of Chinese chestnut seeds. **(A)** Seed germination-related gene expression profiles of Chinese chestnut, where * indicates a significant difference at $P < 0.05$. **(B)** Levels of endogenous hormones in Chinese chestnut at seed germination stage S3, where * indicates a significant difference at $P < 0.05$. **(C)** Alpha-amylase and beta-amylase activities in seed germination stages of wild and cultivated Chinese chestnut seeds; * indicates a significant difference at $P < 0.05$. **(D)** The allele frequencies of selected SVs in wild and cultivated Chinese chestnut. **(E)** The model of seed germination regulation by ABA synthesis pathway genes under strong selection in Chinese chestnut; * indicates the significance of the key genes.

Selected Structural Variants Affecting Expression of Seed Germination Genes in Wild and Cultivated Chinese Chestnut

To further verify the strongly artificially selected genes and identify DEGs associated with seed germination in wild and cultivated Chinese chestnut, stage S1–S3 samples were used for transcriptomic sequencing analysis based on seed germination phases. There were 5,562, 6,741 and 9,105 DEGs in stages S1, S2 and S3 of cultivated Chinese chestnut seed germination, respectively. According to KEGG pathway analysis, these DEGs are significantly associated with the biosynthesis of secondary metabolites, metabolic pathways, starch and sucrose metabolism, and plant hormone signal transduction (**Supplementary Figure 6**). These genes may be related to recalcitrant seed germination in Chinese chestnut.

In general, seed germination was influenced by alpha-amylase and beta-amylase activity. Notably, the *AMY* and *BMY* genes exhibited allele frequency change patterns indicative of strong selection in the wild and cultivated populations of Chinese chestnut. For example, a 599 bp deletion in the *AMY3* gene promoter had a frequency of 63.64% in the wild population and 16.13% in the cultivated population in the V2 genome, and the 11 bp deletion in the *BMY7* gene had a frequency of 85.19% in the wild population and 30.00% in the cultivated population in the V2 genome (**Figure 4D**). Furthermore, the *AMY1*, *AMY2,* and *AMY3* genes were significantly more highly expressed during the S1–S2 stages of cultivated seed germination (**Figure 4A**) and may be involved in the differential of the alpha-amylase activity in the seed imbibition stage of wild and cultivated Chinese chestnut (**Figure 4C**). In the radicle emergence stage (S3), the beta-amylase activity of cultivated seeds was significantly higher than that of wild seeds, and there was no significant difference in alpha-amylase activity between cultivar and wild seeds (**Figure 4C**). The *BMY4* and *BMY7* genes were strongly selected in the wild and cultivar genomes and significantly highly expressed in the radicle emergence stage (S3) of the cultivated seeds; it may be involved in the differential of beta-amylase activity between of wild and cultivated seeds in the S3 stage (**Figures 4A,C**). Additionally, according to the SV divergence across the cultivated and wild genomes, GA and ABA signal transduction genes may be involved in seed germination differences. The *NCED2/5* and *NCED3/9* genes were strongly selected and showed specific allele frequency change patterns (**Figure 4D**). We also found that several genes of the ABA metabolic pathway to be significantly upregulated expression in stage S3 of wild Chinese chestnut, including *ABA4*, *ABA2*, *NCED2/5*, *NCED3/9* and *ABSCISIC ALDEHYDE OXIDASE 3* (*AAO3*) (**Supplementary Figure 8**). In particular, *NCED2/5* and *NCED3/9* were activated in wild seed germination stage S3. The *PIL5* gene was also detected in the SV divergence analysis of the genome and may be involved in seed germination. In terms of gene expression, the *PIL5* gene was significantly more highly expressed in wild seed germination stages, which directly activates transcription of the *SOM*, GIBBERELLIC ACID INSENSITIVE/*REPRESSOR OF GA* (*GAI/RGA*) and *DOF AFFECTING GERMINATION 1* (*DAG1*) genes to regulate the expression of genes related to GA and ABA biosynthesis, and was significantly more highly expressed in the wild seed germination stages (**Supplementary Figure 9**). Moreover, expression of the GA biosynthetic genes *GA3ox1/GA3ox2* was suppressed in stages S1–S2 and activated in stage S3 of wild seed germination, and *GA2ox2* was activated in wild seed germination stages S1–S3 (**Figure 4A** and **Supplementary Figure 7**). In addition, the content of the endogenous hormone ABA was significantly

FIGURE 5 | Glycerophospholipid metabolism genes associated with the germination of Chinese chestnut seeds. **(A)** Model of the glycerophospholipid metabolism pathway in Chinese chestnut; * indicates significant key genes. **(B)** The glycerophospholipid metabolism gene expression profiles of Chinese chestnut, where * indicates a significant difference at $P < 0.05$. **(C)** Allele frequencies of selected SVs in wild and cultivated Chinese chestnut.

higher in wild seeds than in cultivated seeds (**Figure 4B**), consistent with agreement with the role of *NCED2/5* and *NCED3/9* in the activation of this direct regulation of ABA levels. There were no significant changes in the contents of other endogenous hormone in the cultivated and wild seed germination stages (**Figure 4B**).

Although the lipid content of Chinese chestnut seeds was less than 1%, we found that glycerophospholipid metabolism genes were strongly selected in the wild and cultivated populations (**Figure 2D**). The *GLYCEROL-3-PHOSPHATE SN-2-ACYLTRANSFERASE 2* (*GPAT2*), *GPAT9*, *ACYLTRANSFERASE 1* (*ATS1*) and *LYSOPHOSPHATIDYL ACYLTRANSFERASE 5* (*LPAT5*) genes showed specific allele frequency change patterns. For example, a 45 bp deletion in the *LPAT9* gene had a frequency of 76.19% in the wild group and 12.50% in the cultivated group (**Figure 5C**). These genes, which are involved in the synthesis of phosphatidic acid (**Figure 5A**), were significantly highly expressed during wild seed germination (**Figure 5B**). Additionally, the other genes involved in phosphatidic acid synthesis showed high expression during wild seed germination, although they were not strongly selected in the genomes.

DISCUSSION

Seed germination is crucial for plant development and breeding programs. This process is affected by environmental conditions and genetic structure, and plants require large amounts of energy, such as starch, proteins, and lipids during germination (Zhao

FIGURE 6 | Model of the seed germination mechanism regulated by strongly selected genes in Chinese chestnut.

et al., 2018; Zaynab et al., 2021). The starchy nuts of Chinese chestnut have a dry weight of 46~64% (Chen et al., 2017), and the germination characteristics of these recalcitrant seeds differ between cultivated and wild trees. In general, seeds are unable to survive drying and chilling, as they rapidly lose their germination capacity and viability during storage (Pammenter and Berjak, 1999), and wild seeds have a strong dormancy capacity and low germination rates (Nakamura et al., 2017; Wang et al., 2018). Various factors are responsible for the low rate of seed germination, among which amylase has been recognized as an important factor in regulating seed germination (Damaris et al., 2019). In the present study, cultivated seeds displayed a stronger germination capacity than wild seeds. Some strongly selected alpha-amylase and beta-amylase genes were identified by comparing the genomes V2 and N11_1. *BMY4* and *BMY7* were significantly more highly expressed during the germination stages of cultivated seeds, and the activity of beta-amylase was significantly higher than that of alpha-amylase during the seed germination stage. These findings implied that beta-amylase may be play a key role in the high rate of seed germination processes of cultivated Chinese chestnut. Similarly, beta-amylase activity has been also verified to be an important factor in recalcitrant seed germination (Zaynab et al., 2021). Moreover, the amylase gene encodes an enzyme involved in starch degradation and is regulated by ABA and GA (Vanstraelen and Benková, 2012). ABA biosynthetic genes were strongly selected, with suppressed expression to decrease ABA levels in cultivated Chinese chestnut seeds. *PIL5*, a key negative regulator of seed germination, activates the expression of *SOM* by binding directly to its promoter, partially regulating expression of ABA and GA metabolic genes (Oh et al., 2004; Kim et al., 2008). Among the regulatory genes associated with Chinese chestnut seed germination (**Figure 4E**), high expression of the *PIL5* gene is also involved in the differential germination between wild and cultivated seeds.

Lipid degradation and conversion to sugars are the key factors for seed germination in oil palm (Cui et al., 2020). Membrane lipid analysis in relation to recalcitrant seed desiccation tolerance has a close relationship with a reduced seed germination rate (Chen et al., 2018), and phosphatidic acid synthesis genes are involved in the dehydration sensitivity of recalcitrant seed germination in cork oak (Li et al., 2021). Moreover, phosphatidic acid results in a response to ABA during seed germination (Katagiri et al., 2005). Regardless, it remains unclear whether the lipids are involved in the germination of starchy seed species. Although the lipid content of Chinese chestnut is only 0.27~0.64% (Chen et al., 2017), there are some strongly selected genes involved in phospholipid metabolism and phosphatidic acid synthesis in cultivated and wild genomes, and these genes were upregulated in wild seed germination stages. This indicates that phospholipids may be involved in the germination of Chinese chestnut seeds.

The germination rate of the highly recalcitrant seeds of cultivated the Chinese chestnut was influenced by several factors, including natural and artificial selection. Finally, a possible working model for proposed for the regulatory mechanism of germination in recalcitrant seeds of Chinese chestnut is proposed
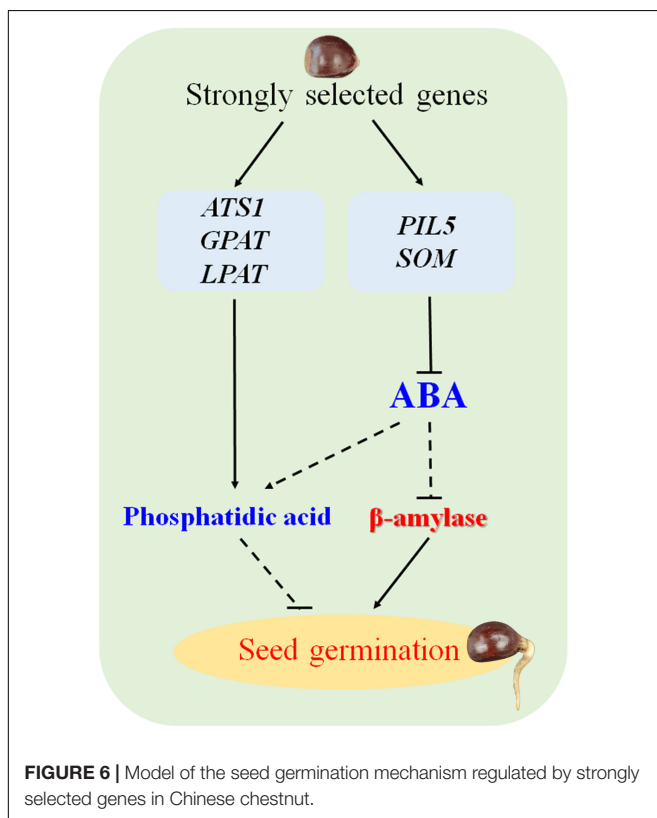
(**Figure 6**). These findings will contribute to improving the germination rate of recalcitrant seeds and provide insight into recalcitrant seed germination.

DATA AVAILABILITY STATEMENT

The sequencing datasets presented in this study can be found in online repositories. The Chinese chestnut genome assembly can be accessed at Castanea Genome Database (www.castaneadb.net). The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

AUTHOR CONTRIBUTIONS

FUNDING

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2022.828270/full#supplementary-material

Supplementary Data 4 | GO and KEGG enrichment results of were strongly selected genes.

Supplementary Data 5 | The allele frequencies of SVs in Chinese chestnut.

REFERENCES

Alandete-Saez, M., Ron, M., Leiboff, S., and McCormick, S. (2011). *Arabidopsis thaliana GEX1* has dual functions in gametophyte development and early embryogenesis. *Plant J.* 68, 620–632. doi: 10.1111/j.1365-313X.2011.04713.x

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–3402. doi: 10.1093/nar/25.17.3389

Ascencio-Ibáñez, J. T., Sozzani, R., Lee, T. J., Chu, T. M., Wolfinger, R. D., Cella, R., et al. (2008). Global analysis of *Arabidopsis* gene expression uncovers a complex array of changes impacting pathogen response and cell cycle during geminivirus infection. *Plant Physiol.* 148, 436–454. doi: 10.1104/pp.108.121038

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and Genomewise. *Genome Res.* 14, 988–995. doi: 10.1101/gr.1865504

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170

Buchfink, B., Xie, C., and Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176

Burton, J. N., Adey, A., Patwardhan, R. P., Qiu, R., Kitzman, J. O., and Shendure, J. (2013). Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125. doi: 10.1038/nbt.2727

Campbell, M. A., Haas, B. J., Hamilton, J. P., Mount, S. M., and Buell, C. R. (2006). Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis. BMC Genomics* 7:327. doi: 10.1186/1471-2164-7-327

Chen, H., Patterson, N., and Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.* 20, 393–402. doi: 10.1101/gr.100545.109

Chen, H., Yu, X., Zhang, X., Yang, L., Huang, X., Zhang, J., et al. (2018). Phospholipase Dα1-mediated phosphatidic acid change is a key determinant of desiccation-induced viability loss in seeds. *Plant Cell Environ.* 41, 50–63. doi: 10.1111/pce.12925

Chen, L., Lu, D., Wang, T., Li, Z., Zhao, Y., Jiang, Y., et al. (2017). Identification and expression analysis of starch branching enzymes involved in starch synthesis during the development of chestnut (*Castanea mollissima* Blume) cotyledons. *PLoS One* 12:e0177792. doi: 10.1371/journal.pone.0177792

Conesa, A., and Götz, S. (2008). Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* 2008:619832. doi: 10.1155/2008/619832

Cui, J., Lamade, E., and Tcherkez, G. (2020). Seed germination in oil palm (*Elaeis guineensis* Jacq.): a review of metabolic pathways and control mechanisms. *Int. J. Mol. Sci.* 21:4227. doi: 10.3390/ijms21124227

Damaris, R. N., Lin, Z., Yang, P., and He, D. (2019). The rice Alpha-amylase, conserved regulator of seed maturation and germination. *Int. J. Mol. Sci.* 20:450. doi: 10.3390/ijms20020450

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. doi: 10.1093/bioinformatics/btr330

Deng, A., Tan, W., He, S., Liu, W., Nan, T., Li, Z., et al. (2008). Monoclonal antibody-based enzyme linked immunosorbent assay for the analysis of jasmonates in plants. *J. Integr. Plant Biol.* 50, 1046–1052. doi: 10.1111/j.1744-7909.2008.00715.x

Dussert, S., Serret, J., Bastos-Siqueira, A., Morcillo, F., Déchamp, E., Rofidal, V., et al. (2018). Integrative analysis of the late maturation programme and desiccation tolerance mechanisms in intermediate coffee seeds. *J. Exp. Bot.* 69, 1583–1597. doi: 10.1093/jxb/erx492

Feng, H., Wang, S., Dong, D., Zhou, R., and Wang, H. (2020). *Arabidopsis* Ubiquitin-Conjugating Enzymes *UBC7, UBC13,* and *UBC14* are required in plant responses to multiple stress conditions. *Plants (Basel)* 9:723. doi: 10.3390/plants9060723

Freed, D., Aldana, R., Weber, J., and Edwards, J. (2017). The Sentieon Genomics Tools - A fast and accurate solution to variant calling from next-generation sequence data. *bioRxiv [preprint]* doi: 10.1101/115717

Gao, L., Gonda, I., Sun, H., Ma, Q., Bao, K., Tieman, D. M., et al. (2019). The tomato pan-genome uncovers new genes and a rare allele regulating fruit flavor. *Nat. Genet.* 51, 1044–1051. doi: 10.1038/s41588-019-0410-2

Gimbi, D. M., and Kitabatake, N. (2002). Changes in alpha-and beta-amylase activities during seed germination of African finger millet. *Int. J. Food Sci. Nutr.* 53, 481–488. doi: 10.1080/09637480220164361

Haas, B. J., Salzberg, S. L., Zhu, W., Pertea, M., Allen, J. E., Orvis, J., et al. (2008). Automated eukaryotic gene structure annotation using EvidenceModeler and the program to assemble spliced alignments. *Genome Biol.* 9:R7. doi: 10.1186/gb-2008-9-1-r7

Katagiri, T., Ishiyama, K., Kato, T., Tabata, S., Kobayashi, M., and Shinozaki, K. (2005). An important role of phosphatidic acid in ABA signaling during germination in *Arabidopsis thaliana. Plant J.* 43, 107–117. doi: 10.1111/j.1365-313X.2005.02431.x

Kijak, H., and Ratajczak, E. (2020). What do we know about the genetic basis of seed desiccation tolerance and longevity? *Int. J. Mol. Sci.* 21:3612. doi: 10.3390/ijms21103612

Kim, D. H., Yamaguchi, S., Lim, S., Oh, E., Park, J., Hanada, A., et al. (2008). SOMNUS, a CCCH-type zinc finger protein in *Arabidopsis*, negatively regulates light-dependent seed germination downstream of *PIL5*. *Plant Cell*. 20, 1260–1277. doi: 10.1105/tpc.108.058859

Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinform*. 5:59. doi: 10.1186/1471-2105-5-59

Li, D., Li, Y., Qian, J., Liu, X., Xu, H., Zhang, G., et al. (2021). Comparative transcriptome analysis revealed candidate genes potentially related to desiccation sensitivity of recalcitrant *Quercus variabilis* seeds. *Front. Plant Sci*. 12:717563. doi: 10.3389/fpls.2021.717563

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100. doi: 10.1093/bioinformatics/bty191

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760. doi: 10.1093/bioinformatics/btp324

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20, 2878–2879. doi: 10.1093/bioinformatics/bth315

Marques, A., Nijveen, H., Somi, C., Ligterink, W., and Hilhorst, H. (2019). Induction of desiccation tolerance in desiccation sensitive *Citrus limon* seeds. *J. Integr. Plant Biol*. 61, 624–638. doi: 10.1111/jipb.12788

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., et al. (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 20, 1297–1303. doi: 10.1101/gr.107524.110

Mitchell, A. L., Attwood, T. K., Babbitt, P. C., Blum, M., Bork, P., Bridge, A., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res*. 47, D351–D360. doi: 10.1093/nar/gky1100

Nakamura, S., Pourkheirandish, M., Morishige, H., Sameri, M., Sato, K., and Komatsuda, T. (2017). Quantitative trait loci and maternal effects affecting the strong grain dormancy of wild barley (*Hordeum vulgare* ssp. spontaneum). *Front. Plant Sci*. 8:1840. doi: 10.3389/fpls.2017.01840

Nattestad, M., and Schatz, M. C. (2016). Assemblytics: a web analytics tool for the detection of variants from an assembly. *Bioinformatics* 32, 3021–3023. doi: 10.1093/bioinformatics/btw369

Obroucheva, N., Sinkevich, I., and Lityagina, S. (2016). Physiological aspects of seed recalcitrance: a case study on the tree *Aesculus hippocastanum*. *Tree Physiol*. 36, 1127–1150. doi: 10.1093/treephys/tpw037

Oh, E., Kim, J., Park, E., Kim, J. I., Kang, C., and Choi, G. (2004). PIL5, a phytochrome-interacting basic helix-loop-helix protein, is a key negative regulator of seed germination in *Arabidopsis thaliana*. *Plant Cell*. 16, 3045–3058. doi: 10.1105/tpc.104.025163

Pammenter, N., and Berjak, P. (1999). A review of recalcitrant seed physiology in relation to desiccation-tolerance mechanisms. *Seed Sci. Res*. 9, 13–37. doi: 10.1017/S0960258599000033

Parra, G., Blanco, E., and Guigó, R. (2000). GeneID in Drosophila. *Genome Res*. 10, 511–515. doi: 10.1101/gr.10.4.511

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet*. 81, 559–575. doi: 10.1086/519795

Roach, T., Beckett, R. P., Minibayeva, F. V., Colville, L., Whitaker, C., Chen, H., et al. (2010). Extracellular superoxide production, viability and redox poise in response to desiccation in recalcitrant *Castanea sativa* seeds. *Plant Cell Environ*. 33, 59–75. doi: 10.1111/j.1365-3040.2009.02053.x

Romero-Rodríguez, M. C., Archidona-Yuste, A., Abril, N., Gil-Serrano, A. M., Meijón, M., and Jorrín-Novo, J. V. (2018). Germination and early seedling development in *Quercus ilex* recalcitrant and non-dormant seeds: targeted transcriptional, hormonal, and sugar analysis. *Front. Plant Sci*. 9:1508. doi: 10.3389/fpls.2018.01508

Salamov, A. A., and Solovyev, V. V. (2000). *Ab initio* gene finding in Drosophila genomic DNA. *Genome Res*. 10, 516–522. doi: 10.1101/gr.10.4.516

Sghaier-Hammami, B., Hammami, S. B. M., Baazaoui, N., Gómez-Díaz, C., and Jorrín-Novo, J. V. (2020). Dissecting the seed maturation and germination processes in the non-orthodox *Quercus* ilex species based on protein signatures as revealed by 2-DE coupled to MALDI-TOF/TOF proteomics strategy. *Int. J. Mol. Sci*. 21:4870. doi: 10.3390/ijms21144870

Shu, K., Liu, X. D., Xie, Q., and He, Z. H. (2016). Two faces of one seed: hormonal regulation of dormancy and germination. *Mol. Plant* 9, 34–45. doi: 10.1016/j.molp.2015.08.010

Slater, G. S., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. *BMC Bioinform*. 6:31. doi: 10.1186/1471-2105-6-31

Stanke, M., Schöffmann, O., Morgenstern, B., and Waack, S. (2006). Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinform*. 7:62. doi: 10.1186/1471-2105-7-62

Sun, Y., Lu, Z., Zhu, X., and Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat. Commun*. 11:3375. doi: 10.1038/s41467-020-17111-w

Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinform*. 25, 4.10.1–4.10.14. doi: 10.1002/0471250953.bi0410s25

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc*. 7, 562–578. doi: 10.1038/nprot.2012.016

Vanstraelen, M., and Benková, E. (2012). Hormonal interactions in the regulation of plant development. *Annu. Rev. Cell Dev. Biol*. 28, 463–487. doi: 10.1146/annurev-cellbio-101011-155741

Vieitez, A. M., San-José, M. C., and Corredoira, E. (2011). Cryopreservation of zygotic embryonic axes and somatic embryos of European chestnut. *Methods Mol. Biol*. 710, 201–213. doi: 10.1007/978-1-61737-988-8_15

Walters, C., Berjak, P., Pammenter, N., Kennedy, K., and Raven, P. (2013). Plant science. Preservation of recalcitrant seeds. *Science* 339, 915–916. doi: 10.1126/science.1230935

Wang, J., Tian, S., Sun, X., Cheng, X., Duan, N., Tao, J., et al. (2020). Construction of pseudomolecules for the Chinese chestnut (*Castanea mollissima*) genome. *G3 (Bethesda)* 10, 3565–3574. doi: 10.1534/g3.120.401532

Wang, X., Gao, L., Jiao, C., Stravoravdis, S., Hosmani, P. S., Saha, S., et al. (2020). Genome of *Solanum pimpinellifolium* provides insights into structural variants during tomato breeding. *Nat. Commun*. 11:5817. doi: 10.1038/s41467-020-19682-0

Wang, M., Li, W., Fang, C., Xu, F., Liu, Y., Wang, Z., et al. (2018). Parallel selection on a dormancy gene during domestication of crops from multiple families. *Nat. Genet*. 50, 1435–1441. doi: 10.1038/s41588-018-0229-2

Wang, Y., Htwe, Y. M., Li, J., Shi, P., Zhang, D., Zhao, Z., et al. (2019). Integrative omics analysis on phytohormones involved in oil palm seed germination. *BMC Plant Biol*. 19:363. doi: 10.1186/s12870-019-1970-0

Wilf, P., Nixon, K. C., Gandolfo, M. A., and Cúneo, N. R. (2019). Eocene Fagaceae from Patagonia and Gondwanan legacy in Asian rainforests. *Science* 364:eaaw5139. doi: 10.1126/science.aaw5139

Xia, K., Hill, L. M., Li, D. Z., and Walters, C. (2014). Factors affecting stress tolerance in recalcitrant embryonic axes from seeds of four *Quercus* (Fagaceae) species native to the USA or China. *Ann. Bot*. 114, 1747–1759. doi: 10.1093/aob/mcu193

Xia, K., and Zhu, Z. Q. (2021). Characterization of physiological traits during development of the recalcitrant seeds of *Quercus serrata*. *Plant Biol*. 23, 1000–1005. doi: 10.1111/plb.13309

Xing, Y., Liu, Y., Zhang, Q., Nie, X., Sun, Y., Zhang, Z., et al. (2019). Hybrid de novo genome assembly of Chinese chestnut (*Castanea mollissima*). *Gigascience*. 8:giz112. doi: 10.1093/gigascience/giz112

Zaynab, M., Pan, D., Fatima, M., Sharif, Y., Chen, S., and Chen, W. (2021). Proteomics analysis of *Cyclobalanopsis gilva* provides new insights of low seed germination. *Biochimie* 180, 68–78. doi: 10.1016/j.biochi.2020.10.008

Zhang, C., Dong, S. S., Xu, J. Y., He, W. M., and Yang, T. L. (2019). PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* 35, 1786–1788. doi: 10.1093/bioinformatics/bty875

Zhao, M., Zhang, H., Yan, H., Qiu, L., and Baskin, C. C. (2018). Mobilization and role of starch, protein, and fat reserves during seed germination of

six wild grassland species. *Front. Plant Sci.* 9:234. doi: 10.3389/fpls.2018.0 0234

Zheng, Y., Jiao, C., Sun, H., Rosli, H. G., Pombo, M. A., Zhang, P., et al. (2016). iTAK: a program for genome-wide prediction and classification of plant transcription factors, transcriptional regulators, and protein kinases. *Mol. Plant* 9, 1667–1670. doi: 10.1016/j.molp.2016.09.014

Zheng, Y., Wu, S., Bai, Y., Sun, H., Jiao, C., Guo, S., et al. (2019). Cucurbit Genomics Database (CuGenDB): a central portal for comparative and functional genomics of cucurbit crops. *Nucleic Acids Res.* 47, D1128–D1136. doi: 10.1093/nar/gky944

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.