



OPEN ACCESS

EDITED BY

Thomas Hartwig,
Heinrich Heine University of
Düsseldorf, Germany

REVIEWED BY

Panagiotis Ioannidis,
Foundation for Research and
Technology Hellas (FORTH), Greece
Larry Croft,
Deakin University, Australia

*CORRESPONDENCE

Qingzhong Liu
qzliu001@126.com

SPECIALTY SECTION

This article was submitted to
Functional and Applied Plant
Genomics,
a section of the journal
Frontiers in Plant Science

RECEIVED 20 September 2022

ACCEPTED 10 November 2022

PUBLISHED 28 November 2022

CITATION

Wang J, Hong P, Qiao Q, Zhu D,
Zhang L, Lin K, Sun S, Jiang S,
Shen B, Zhang S and Liu Q (2022)
Chromosome-level genome assembly
provides new insights into Japanese
chestnut (*Castanea crenata*) genomes.
Front. Plant Sci. 13:1049253.
doi: 10.3389/fpls.2022.1049253

COPYRIGHT

© 2022 Wang, Hong, Qiao, Zhu, Zhang,
Lin, Sun, Jiang, Shen, Zhang and Liu.

This is an open-access article
distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original
author(s) and the copyright owner(s)
are credited and that the original
publication in this journal is cited, in
accordance with accepted academic
practice. No use, distribution or
reproduction is permitted which does
not comply with these terms.

Chromosome-level genome assembly provides new insights into Japanese chestnut (*Castanea crenata*) genomes

Jiawei Wang¹, Po Hong¹, Qian Qiao¹, Dongzi Zhu¹,
Lisi Zhang¹, Ke Lin^{1,2}, Shan Sun¹, Shuna Jiang³,
Bingxue Shen^{1,3}, Shizhong Zhang³ and Qingzhong Liu^{1*}

¹Shandong Key Laboratory of Fruit Biotechnology Breeding, Shandong Institute of Pomology, Taian, Shandong, China, ²Department of Biology Science and Technology, Taishan University, Taian, Shandong, China, ³State Key Laboratory of Crop Biology, Shandong Agricultural University, Taian, Shandong, China

Japanese chestnut (*Castanea crenata* Sieb. et Zucc) is an economically and ecologically important chestnut species in East Asia. Here, we presented a high-quality chromosome-level reference genome of the Japanese chestnut cultivar 'Tsukuba' by combining Nanopore long reads and Hi-C sequencing. The final assembly has a size of 718.30 Mb and consists of 12 pseudochromosomes ranging from 41.03 to 92.03 Mb, with a BUSCO complete gene percentage of 97.6%. A total of 421.37 Mb repetitive sequences and 46,744 gene models encoding 46,463 proteins were predicted in the genome. Genome evolution analysis showed that Japanese chestnut is closely related to Chinese chestnut and these species shared a common ancestor ~6.5 million years ago. This high-quality Japanese chestnut genome represents an important resource for the chestnut genomics community and will improve our understanding of chestnut biology and evolution.

KEYWORDS

japanese chestnut, nanopore, Hi-C, genome sequencing, evolution

Introduction

Chestnut is the only domesticated nut species in family *Fagaceae*, and Japanese chestnut, Chinese chestnut and European chestnut have all been planted for a thousand years in Japan, China and European countries (Conedera and Krebs, 2008; Xing et al., 2019; Nishio et al., 2021). There are four members in the genus *Castanea*, which provide food for humans and other animals, including Chinese chestnut (*Castanea mollissima* Bl.), Japanese chestnut (*C. crenata* Sieb. et Zucc), European chestnut (*C. sativa* Mill.) and American chestnut (*C. dentate* Borkh.). Both Japanese chestnut and Chinese chestnut are

blight-resistant, in contrast to European chestnut and American chestnut (Shirasawa et al., 2021).

Japanese chestnut is a woody native plant of Japan and South Korea and is widely cultivated in Asian countries. Similar to Chinese chestnut, several Japanese chestnut cultivars have been selected for their large nut size (Nishio et al., 2021). Compared to Chinese chestnut, Japanese chestnut has larger nuts and better yields but a lower sugar content, and the pellicle is more difficult to remove. However, two newly released cultivars, ‘Porotan’ and ‘Porosuke’, exhibit an easily peeled pellicle, and several studies suggest that this trait may be controlled by one recessive gene (Kurogi and Uritani, 1966; Sato et al., 2007; Sakamoto et al., 2015; Nishio et al., 2021). More research is needed to clone this gene and uncover the molecular mechanism of this important trait in chestnut.

The genome is the foundation of genetic research and has provided many advantages in crop breeding, such as in rice, corn and cotton. Marker-assisted breeding and genomic selection can speed up the breeding cycle, but these methods require one or several high-quality reference genomes. Although seven Chinese chestnut genome assemblies are available publicly (Xing et al., 2019; Staton et al., 2020; Sun et al., 2020; Wang et al., 2020b; Hu et al., 2022), only two Japanese chestnut genome assemblies have recently been released by NCBI (GCA_019972055.1 and GCA_020976635.1); further research is still needed on the genome variations in Japanese chestnut. One of the released Japanese chestnut genomes is for cultivar ‘Ginyose’ and was generated at the scaffold to chromosome-level (Shirasawa et al., 2021); the other is a draft genome assembly that was generated from one 400-years-old native tree in South Korea. Studies have proven that several high-quality genomes help uncover functional genomic variation by direct comparative analysis (Valliyodan et al., 2019; Zhang et al., 2019a; Zhou et al., 2019). Therefore, a high-quality genome and annotation for Japanese chestnut are still needed, and direct comparative analysis between these two genomes may provide useful information for research on this crop.

In this study, Nanopore long-read sequencing and Illumina sequencing were used to assemble the genome of the Japanese chestnut cultivar ‘Tsukuba’, and Hi-C sequencing was used to generate the chromosome-level assembly. Transcriptomes from roots, stems, leaves, flowers and different developmental stages of Japanese chestnut were generated by RNA sequencing and used for genome annotation. Our study provides an opportunity to assess the genome variations between Japanese chestnut and other chestnuts.

Materials and methods

DNA extraction and sequencing

Leaf samples from the Japanese chestnut cultivar ‘Tsukuba’ grown in the Germplasm Resources Nursery of Shandong

Institute of Pomology, Taian, China, were collected and frozen in liquid nitrogen. Genomic DNA was extracted, size selected and sequenced on an Oxford Nanopore PromethION system by Wuhan Benagen Tech Solutions Company Limited (Wuhan, China). Sequencing adapters were removed from the raw reads, and then low-quality reads were filtered out. For Illumina sequencing, a paired-end library with an insert size of 350 bp was constructed and sequenced by Wuhan Benagen Tech Solutions Company Limited (Wuhan, China) following the manufacturer’s protocol using the Illumina HiSeq X Ten platform.

For Hi-C library construction, fresh young leaves from the same Japanese chestnut tree were fixed using formaldehyde at a concentration of 1%. The chromatin was cross-linked and digested using the restriction enzyme HindIII. The library construction and sequencing followed the method used in our previous research (Wang et al., 2020a).

Genome assembly and pseudochromosome construction

The primary assembly was generated by NECAT v0.01 and polished by medaka (<https://github.com/nanoporetech/medaka>) and NextPolish v1.3.1 with the Nanopore reads and Illumina short reads, respectively (Hu et al., 2020; Chen et al., 2021). To anchor the contigs of the primary assembly to the chromosomal-level scaffolds, duplicated contigs were first removed by using Purge Haplotigs v1.1.0 (Roach et al., 2018), and then ALLHiC v0.9.1214 was applied to construct the chromosomal-level scaffolds (Zhang et al., 2019b), Juicer and 3d-dna pipelines were used to adjust and polish the super scaffolds (Durand et al., 2016; Dudchenko et al., 2017). TGS-GapCloser v1.0.1 was used to close sequence gaps in the genome assembly (Xu et al., 2020).

Genome assembly quality evaluation

Bwa v0.7.17 was used to align the Illumina reads to the primary assembly (Li and Durbin, 2010), and SAMtools v0.1.9 was used to calculate the mapping rate (Danecek et al., 2021). The completeness of the genome assembly was assessed by BUSCO v4.0.2 with the Embryophyta_obd10 database (Seppey et al., 2019). LTR_FINDER_parallel was used to detect LTR-RTs in the genome, and LTR_retriever v2.9.0 was used to calculate LAIs for each genome (Ou et al., 2018).

Repetitive sequence annotation

For repetitive sequence annotation, RepeatModeler v2.0.1 was first applied as the *de novo* method to identify repetitive

elements in the genome (Flynn et al., 2020); then, RepeatMasker v4.0.9 was used as the homology-based tool to identify and annotate the repetitive sequences using Dfam v3.151 and Repbase library v20170127 (Zhi et al., 2006).

Protein-coding gene prediction and functional annotation

The Funannotate v1.7.4 pipeline was used to predict protein-coding genes and functionally annotate the predicted genes (Love et al., 2019). Briefly, RNA-sequencing data were used to train the pipeline first, which employs HISAT2 v2.1.0 (Kim et al., 2019), Trinity v2.8.5 (Grabherr et al., 2011) and PASA v2.4.1 (Haas et al., 2008). Then, multiple gene models were predicted by GeneMark-ES (Lomsadze et al., 2005), Augustus (Hoff and Stanke, 2019), CodingQuarry (Testa et al., 2015), GlimmerHMM (Majoros et al., 2004), SNAP (Korf, 2004) and PASA v2.4.1 by using the training parameters, and EVIDENCEModeler v1.1.1 (Haas et al., 2008) was used to combine the *ab initio* and evidence-based gene models. The tRNAs were predicted by tRNAscan-SE v2.0.6 (Chan and Lowe, 2019). After filtering out gene models with short lengths (<50 bp), spanning gaps and transposable elements, UTRs were added by the Funannotate update command. Finally, the functions of the proteins were annotated by the EggNOG v4.5.1 (Huerta-Cepas et al., 2017), Pfam v32.0 (El-Gebali et al., 2019), UniProt v2020-08-12 (Apweiler et al., 2004), KEGG (Kanehisa et al., 2014), Gene Ontology (Harris et al., 2004), COG (Koonin et al., 2004), BUSCO v2.0 (Seppey et al., 2019), MEROPS v12.0 (Rawlings et al., 2018), Phobius v1.01 (Kall et al., 2004), SignalP v4.1 (Nielsen, 2017), and CAZyme v8.0 (Lombard et al., 2014) databases or pipelines. We also used KofamKOALA and KofamScan to annotate the proteins (Aramaki et al., 2019) with KEGG Orthologs (KOs).

Gene family analysis and whole-genome duplication event identification

OrthoFinder v2.5.4 was used to identify gene orthologs and gene duplication events in Japanese chestnut genomes and other genomes (Emms and Kelly, 2019). CAFÉ v4.2.1 was employed to explore gene family size expansion and contraction based on the results of OrthoFinder (De Bie et al., 2006). Whole genome duplication analysis was conducted by using wgd (Zwaenepoel and Van de Peer, 2019), and the Ks distribution of one-to-one orthologs between species was plotted by R packages (ggplot2).

RNA sequencing analysis and KEGG pathway enrichment analysis

Total RNA was extracted from roots, stems, leaves, flowers, pellicles and three nut developmental stages (70, 80, and 90 days

after flowering) of the Japanese chestnut cultivar ‘Tsukuba’. There were three replicates for samples from the different nut developmental stages. The cDNA libraries were constructed and sequenced by Shanghai OE Biotech Co., Ltd. (Shanghai, China) on an Illumina HiSeq 2500 platform. RNA sequencing data were first aligned against the Japanese chestnut genome, and gene counts for each sample were generated by using RASflow (Zhang and Jonassen, 2020). TCC-GUI was used to normalize the expression data and detect the DEGs (Su et al., 2019). TBtools V1.098696 was used to perform the KEGG enrichment analysis based on the KEGG annotation of the proteins (Chen et al., 2020).

Genome synteny analysis

The Python module jcvl was employed to analyze the synteny between the Japanese chestnut genome, Chinese chestnut genome and oak genome (Tang et al., 2015). SyRI (Synteny and Rearrangement Identifier) was used for structural-variant detection between the two Japanese chestnut genomes (Goel et al., 2019). TBtools was used to show the syntenic genes between different genomes (Chen et al., 2020).

Results

Sequencing and assembly of the Japanese chestnut (*Castanea crenata* Sieb. et Zucc) genome

The Japanese chestnut cultivar ‘Tsukuba’ (*Castanea crenata* Sieb. et Zucc) was used for whole-genome sequencing and chromosome-scale assembly. After filtering out low-quality reads, a total of 93.48 Gb of Oxford Nanopore long reads and 80.77 Gb of Illumina short reads were obtained. The sequencing details are provided in Table S1. The primary assembly consisted of 469 contigs with a total length of 857.86 Mb and a contig N50 of 5.82 Mb (Table 1).

Three methods were used to assess the quality of the primary assembly. First, the Illumina short reads were mapped back to the primary assembly, and 96.91% (521.82 million reads out of 538.47 million reads) of the total reads were properly paired in the primary assembly, which is significantly higher than several Chinese chestnut varieties (84.24% (‘Yan-Hong’), 90.36% (‘Yan-Shan-Zao-Sheng’), and 89.98% (‘Hei-Shan-Zhai-7’)) (Hu et al., 2022). Second, Benchmarking Universal Single-Copy Orthologs (BUSCOs) analysis showed that 97.8% of the complete BUSCOs were generated in the primary assembly, even though 17.3% of these complete BUSCOs were duplicated (Table 1) (Seppey et al., 2019). Third, the LTR assembly index (LAI) score was generated following the method described by Ou et al. (Ou et al., 2018), and the primary assembly yielded a raw LAI = 8.73 and an LAI=

TABLE 1 BUSCO analysis results for the genomes of Japanese chestnut and Chinese chestnut.

Type	<i>C. crenata</i> .v. TsukubaPrimary	<i>C. crenata</i> .v. TsukubaChromosome	<i>C.</i> <i>crenata</i> .v. Ginyose	<i>C.</i> <i>mollissima</i> Wang et al	<i>C.</i> <i>mollissima</i> Xing et al	<i>C.</i> <i>mollissima</i> Sun et al
Total genome length (Mb)	857.86	718.30	721.17	688.93	785.53	773.99
Contig number	469	206	781	652	2707	422
Contig N50 (Mb)	5.82	6.36	1.59	2.83	0.94	5.88
Complete BUSCOs (C)	1579 97.8%	1576 97.6%	1560 96.6%	1483 91.9%	1578 97.8%	1546 95.8%
Complete and single-copy BUSCOs (S)	1300	1527	1500	1424	1486	1507
Complete and duplicated BUSCOs (D)	279	49	60	59	110	39
Fragmented BUSCOs (F)	13	17	25	34	14	18
Missing BUSCOs (M)	22	21	29	97	22	50
Total BUSCO groups searched				1614		

14.36. These results confirmed the high quality of the primary assembly of the Japanese chestnut genome.

The purged_duplicates assembly consisted of 206 contigs with a total length of 718.30 Mb and a contig N50 of 6.36 Mb. BUSCO analysis of the purged_duplicates assembly generated 97.6% complete BUSCOs, with only 3.04% duplicated BUSCOs. Hi-C sequencing generated a total of 302,781,503 read pairs (45.42 Gb). The statistics of the sequencing and mapping details for the Hi-C data are shown in Table S2. After mapping the Hi-C reads against the purged_duplicates assembly of Japanese chestnut, 87.11 million valid interaction pairs, accounting for 28.77% of the unique mapped read pairs, were used to construct the chromosomal-level scaffolds. Finally, 12 super-scaffolds were generated, with lengths ranging from 41.03 to 92.03 Mb, accounting for 99.72% of the purged_duplicates assembly. After gap closing by using TGS-GapCloser (Tanaka and Kotobuki, 1992), the final chromosome-scale genome assembly of Japanese chestnut was generated, with fewer 20 gaps per chromosome. The statistics of the pseudochromosomes of the Japanese chestnut genome of the two cultivars are shown in Table S3, and the genome profile and Hi-C contact map of the genome of cultivar ‘Tsukuba’ are shown in Figure 1.

Annotation of the Japanese chestnut (*Castanea crenata* Sieb. et Zucc) genome

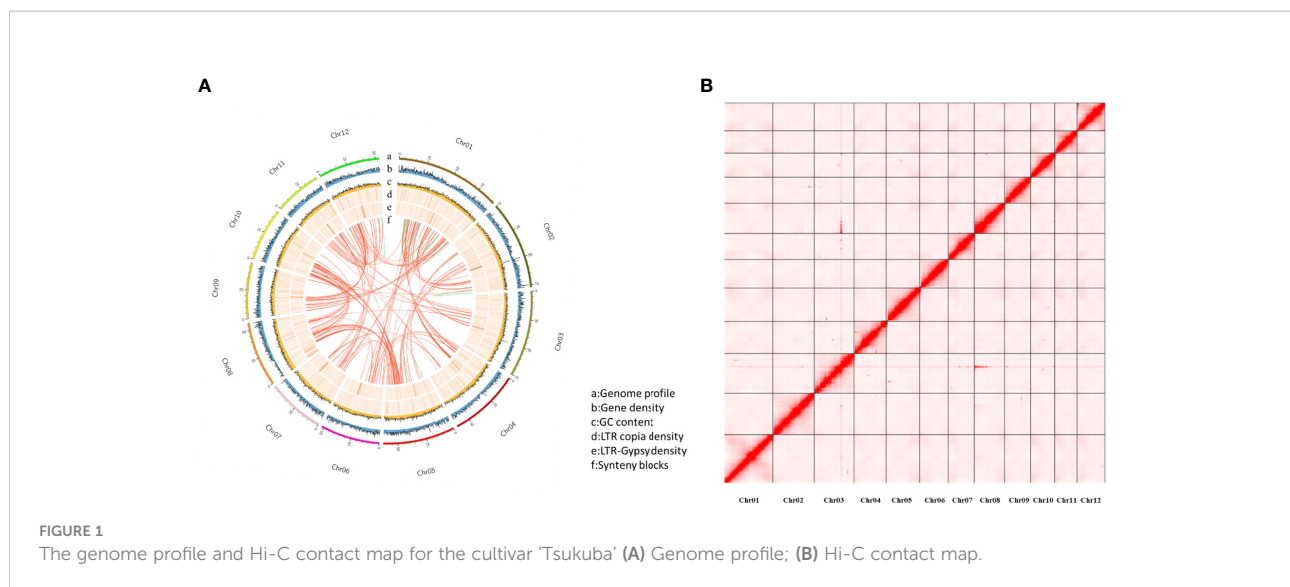
We used homology-based, *de novo* and RNA-seq methods for protein-coding gene prediction and functional annotation. A total of 46,744 gene models encoding 46,463 proteins were predicted in the Japanese chestnut genome (Table 2). A total

of 36,074 of 46,463 proteins (77.64%) were annotated by using several databases or pipelines (see methods).

A total content of 421.37 Mb of repetitive sequences was annotated in the final assembly of the Japanese chestnut genome, indicating that 58.78% of the genome was repetitive (Table S4). Among these repetitive elements, LTR retrotransposons (23.62%) were predominant (14.15% Gypsy, followed by 8.47% Copia), whereas L1 (3.64%) was the most abundant class of LINEs (long interspersed nuclear elements).

Synteny analysis between the genomes of Japanese chestnut, Chinese chestnut and oak

We performed a synteny analysis among the genomes of the Japanese chestnut cultivar ‘Tsukuba’, two cultivars of Chinese chestnut (Sun et al., 2020; Wang et al., 2020b) and oak (Plomion et al., 2018). The genome of the Japanese chestnut cultivar ‘Tsukuba’ shows good overall synteny with the two genome assemblies of Chinese chestnut (Figure 2). A poor collinearity result was observed between Japanese chestnut and oak. These results might be due to the method used by the previous authors to anchor the contigs to pseudochromosomes in the oak genome; specifically, they used the peach genome as a reference to scaffold the genome (Plomion et al., 2018). For the other two Chinese chestnut genome assemblies, Hi-C sequencing technology was employed to scaffold the genome. This result also suggests the good quality of our genome assembly of Japanese chestnut.



Structural-variant detection between Japanese chestnut genomes

Using the whole-genome comparison tool SyRI (Goel et al., 2019), we found 414–415 Mb of collinear regions between the genomes of cultivars 'Tsukuba' and 'Ginyose' (Figure 3). In addition, the SyRI analysis identified 207 inversions (accounting for 165.93 Mb in the genome of 'Tsukuba'), 2,247 translocations (83.27 Mb in the genome of 'Tsukuba'), and 15,916 duplications in the 'Tsukuba' genome (55.65 Mb) and 9,071 duplications in the 'Ginyose' genome (37.34 Mb) (Table S5). We also identified a

30.12 Mb sequence in the 'Tsukuba' genome and a 14.59 Mb sequence in the 'Ginyose' genome that were not aligned with each other.

Evolutionary history and whole-genome duplication events of the Japanese chestnut genome

For evolutionary analysis, we selected six other species whose genomes have been sequenced in *Fagales*: three were

TABLE 2 Genome annotation statics for the Japanese chestnut genome and three Chinese chestnut genome assemblies.

Type	<i>C. crenata</i> v. Tsukuba	<i>C. crenata</i> v. Ginyose	<i>C. mollissima</i> Wang et al	<i>C. mollissima</i> Xing et al	<i>C. mollissima</i> Sun et al
Gene number	46,744	69,980	33,612	36,479	45,661
Gene density (per 100 kb)	6.46	10.23	4.87	4.64	5.89
Average gene length (bp)	3880.57	2323.261	4694.974	5091.057	3283.947
Average exon number per gene	5.767	4.535	5.935	5.797	4.904
Average Exon length (bp)	291.70	277.038	295.956	259.149	227.969
Average intron mean length (bp)	881.549	434.626	752.789	1,156.915	790.527
Genome GC content	35.25%	35.14%	35.11%	36.07%	34.99%
Exon GC content	43.30%	44.56%	41.84%	43.36%	43.49%
Complete BUSCOs (C)	1416	1223	1513	1568	1481
	87.7%	75.8%	93.8%	97.1%	91.7%
Complete and single-copy BUSCOs (S)	1101	1165	1415	1466	1369
Complete and duplicated BUSCOs (D)	315	58	98	102	112
Fragmented BUSCOs (F)	131	202	33	31	85
Missing BUSCOs (M)	67	189	68	15	48
Total BUSCO groups searched			1614		

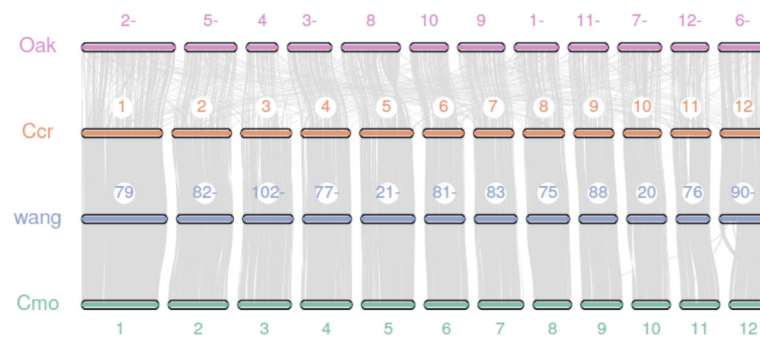


FIGURE 2

Synteny analysis between the genomes of Japanese chestnut, Chinese chestnut and oak. Oak: *Quercus* spp. Ccr: *Castanea crenata* 'Tsukuba', Cmo: *Castanea mollissima* (Sun et al., 2020), and wang: *Castanea mollissima* (Wang et al., 2020b).

from the *Fagaceae* family (Cmo=*Castanea mollissima*, Qrob=*Quercus robur*, Qmo=*Quercus mongolica*) (Nakamura, 1994; Plomion et al., 2018; Ai et al., 2022), two were from the *Juglandaceae* family (Jre=*Juglans regia*, Cil=*Carya illinoensis*) (Neale et al., 2020; Lovell et al., 2021), and one was from the *Betulaceae* family (Cma=*Corylus mandshurica*) (Li et al., 2021). The soybean genome (Gma=*Glycine max*) was used as an outgroup (Valliyodan et al., 2019). A phylogenetic analysis including 346 single-copy orthologous genes confirmed the close relationship between Japanese chestnut and Chinese chestnut (Figure 4). Using a reference divergence time of 47–89 million years ago (MYA) between *Fagaceae* and 86–108 MYA between fabids, we estimated that Japanese chestnut and Chinese chestnut shared a common ancestor ~6.5 MYA (Figure 4).

We calculated the gene *Ks* values in the genomes of Japanese chestnut, grape (*Vitis vinifera*) and Amborella (*Amborella trichopoda*) together with the *Ks* distribution of one-to-one orthologs between these species (Figure 5). The Amborella genome is known to be structurally conserved with no lineage-specific genome duplications since angiosperm diversification (Albert et al., 2013), and the grape genome only experienced paleohexaploidization events (Jaillon et al., 2007). We found that the Japanese chestnut genome and grape genome shared a shallow peak at 1.3–1.5, probably reflecting the paleopolyploidy WGD (γ) event in the angiosperm lineage. This result was also confirmed by the Ccr-Atr and Vvi-Atr gene pairs peaking at ~2.0. We showed evidence of one WGD event in Japanese chestnut compared with grape, reflected by the Ccr-Vvi gene pair peak at 0.8. These WGD events were also confirmed by the syntenic patterns within the Japanese chestnut genome (Figure S1).

Comparative genomics of gene families

Based on sequence homology, 302,394 proteins accounting for 91.5% of the total genes in these eight species were assigned

to 31,300 orthogroups. The statistics for the orthology analysis results are shown in Table S6. There were 657 species-specific orthogroups in the Japanese chestnut genome, consisting of 1,929 proteins. The gene duplication event analysis revealed that Japanese chestnut and Chinese chestnut shared 2,078 gene duplication events, whereas only 1,128 gene duplication events were shared by Mongolian oak (*Quercus mongolica*) and pedunculate oak (*Quercus robur* L.).

The expansion and contraction of the gene families in Japanese chestnut were analyzed by comparing the gene families within order *Fagales*. There were 987 expanded gene families comprising 3,981 genes and 1,204 contracted gene families comprising 1,377 genes in the Japanese chestnut genome. Japanese chestnut and Chinese chestnut shared 686 expanded gene families and 489 contracted gene families (Figure 4).

Of the 987 expanded gene families, 1,609 genes (40.42%) were annotated with KEGG orthology (KO) identifiers and assigned to 421 KEGG pathways (Table S7). KEGG enrichment analysis of these genes was conducted, and the ten most significant pathway categories were Vitamin B6 metabolism (17 genes, ko00750), Mismatch repair (25 genes, ko03430), Plant-pathogen interaction (87 genes, ko04626), Glutathione metabolism (44 genes, ko00480), Phenylpropanoid biosynthesis (70 genes, ko00940), Protein export (25 genes, ko03060), DNA replication (23 genes, ko03030), Nucleotide excision repair (23 genes, ko03420), Sesquiterpenoid and triterpenoid biosynthesis (19 genes, ko00909), and Starch and sucrose metabolism (42 genes, ko00500) (Figure S2).

In the Vitamin B6 metabolism pathways, 17 genes were identified as pyridoxine 4-dehydrogenase, and there were also 9 genes in the same gene family that were not annotated by the KEGG database. However, only 6 genes in the Chinese chestnut genome were assigned to the same orthogroups, whereas 25 genes in pedunculate oak and 26 genes in Mongolian oak were assigned to these orthogroups. In the Phenylpropanoid

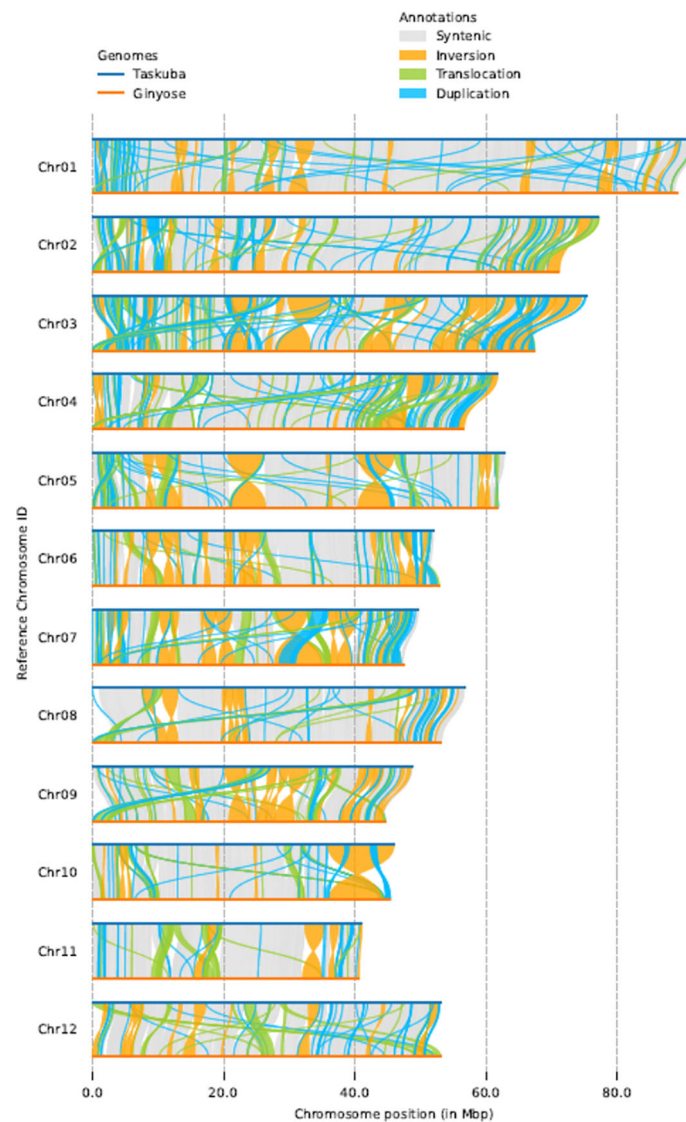


FIGURE 3
Structural-variant detection between Japanese chestnut genomes.

biosynthesis pathways, 28 genes in the Japanese chestnut genome were identified as peroxidase and assigned to 5 orthogroups, and the expansion of these genes may be caused by gene duplication events (Table S8). Several gene families involved in terpenoid biosynthesis, such as Sesquiterpenoid and triterpenoid biosynthesis and Terpenoid backbone biosynthesis, were also expanded in the Japanese chestnut genome.

In the Starch and sucrose metabolism pathway, 14 proteins were identified as trehalose 6-phosphate phosphatase (TPP) in the Japanese chestnut genome, and 14 proteins in the Chinese chestnut genome were identified as TPP (Table S9). However, only 2 proteins in pedunculate oak were identified as TPP, and no TPPs were identified in Mongolian oak. All these TPPs were

grouped into one orthogroup in OrthoFinder, and 12 of the 14 TPPs in the Japanese chestnut genome were located in one 450 kb region on Chr11. Synteny analysis of this region between the Japanese chestnut and Chinese chestnut genomes suggested that tandem gene duplication events occurred in these two genomes (Figure S3). Since trehalose-6-phosphate (T6P) is a signal of sucrose status and a powerful growth regulator, that coordinates plant growth and development with sucrose supply, TPPs reduced the concentration of T6P to regulate carbon partitioning from source to sink organs (Nuccio et al., 2015; Li et al., 2022). The expansion of these TPPs in the Japanese chestnut and Chinese chestnut genomes might contribute to the higher sugar or starch content in the nut of chestnut compared with that of oak.

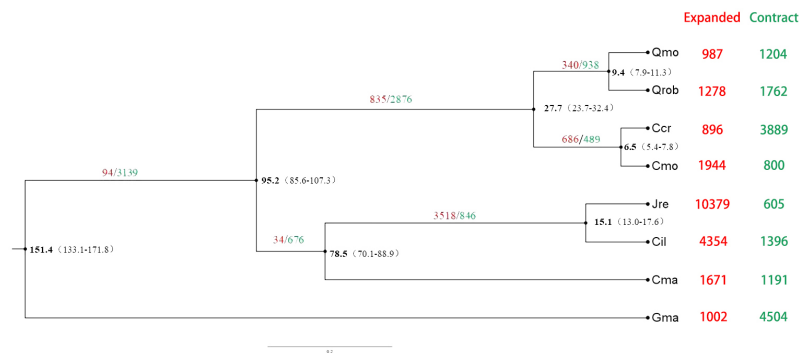


FIGURE 4

Evolutionary history of the Japanese chestnut genome. Qmo, *Quercus mongolica*; Qrob, *Quercus robur*; Ccr, *Castanea crenata* Sieb. et Zucc.; Cmo, *Castanea mollissima*; Jre, *Juglans regia*; Cil, *Carya illinoensis*; Cma, *Corylus mandshurica*; and Gma, *Glycine max*.

Nut development-related genes in Japanese chestnut

The transcriptomes at three nut development stages (70, 80, and 90 days after flowering) were generated to identify genes involved in nut development in Japanese chestnut. A total of 10,971 genes were identified as differentially expressed genes (DEGs, FDR < 0.05) during nut development. The KEGG pathway enrichment analysis results of these DEGs are shown in Figure S4 (Table S10). Among these pathways, Photosynthesis (Ko00195), beta-Alanine metabolism (Ko00410), Carotenoid biosynthesis (Ko00906), Starch and sucrose metabolism (Ko00500), Pentose phosphate pathway (Ko00030), Flavonoid biosynthesis (Ko00941), Pantothenate and CoA biosynthesis (Ko00770), Limonene and pinene degradation (Ko00903), Linoleic acid metabolism (Ko00591), and Propanoate

metabolism (Ko00640) were the ten most enriched metabolic pathways.

Among all these metabolic pathways, Starch and sucrose metabolism had the most gene hits, in which 93 genes were identified as SUS (sucrose synthase [EC: 2.4.1.13]), SS (starch synthase [EC: 2.4.1.21]), GBSS (granule-bound starch synthase [EC: 2.4.1.242]), GPA (glucose-1-phosphate adenylyltransferase [EC: 2.7.7.27]), PGM (phosphoglucomutase [EC: 5.4.2.2]), HK (hexokinase [EC:2.7.1.1]), GPI (glucose-6-phosphate isomerase [EC:5.3.1.9]), GBE (1,4-alpha-glucan branching enzyme [EC:2.4.1.18]), FK (fructokinase [EC:2.7.1.4]), and GPI (glucose-6-phosphate isomerase [EC:5.3.1.9]) (Table S11). Most of these genes showed an increased expression level at stage S2 and slight decrease at stage S3, indicating that starch was mainly synthesized at stage S2 in the nut development process of Japanese chestnut (Table S11).

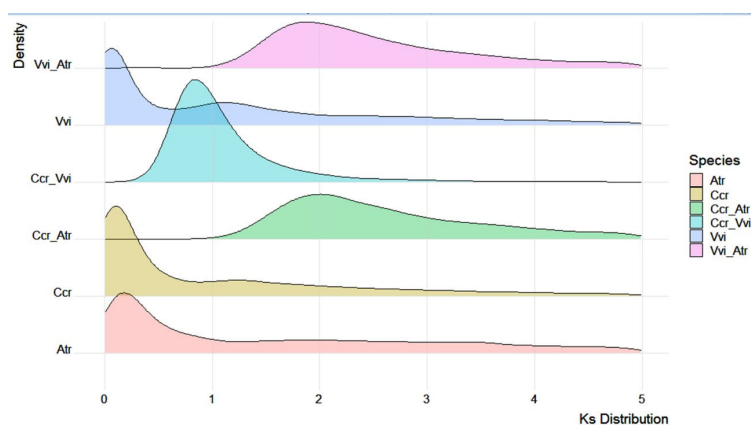


FIGURE 5

Genome duplication events of the Japanese chestnut genome. Vvi, *Vitis vinifera*; Atr, *Amborella trichopoda*; and Ccr, *Castanea crenata* Sieb. et Zucc.

Several genes involved in linoleic acid metabolism were also identified as DEGs in the nut development process, which included 11 genes identified as LOX2S (lipoxygenase [EC: 1.13.11.12]), 2 genes identified as LOX1_5 (linoleate 9S-lipoxygenase [EC: 1.13.11.58]), one gene identified as PLA2G (secretory phospholipase A2 [EC: 3.1.1.4]) and one gene identified as TGL4 (TAG lipase/steryl ester hydrolase/phospholipase A2/LPA acyltransferase [EC: 3.1.1.3/3.1.13/3.1.1.4/2.3.1.51]) (Table S12). Most of these genes showed a decreased expression level during stages S1 to S3, and the overall expression level was lower in the nut than in other tissues, which was consistent with the low content of linoleic acid in the nut (Table S12).

Discussion

Genome assembly quality evaluation of the Japanese chestnut cultivar ‘Tsukuba’

In this study, we generated a *de novo* whole genome assembly of the Japanese chestnut cultivar ‘Tsukuba’ by using Nanopore sequencing and Hi-C technology. The final assembly consisted of 206 contigs with a total length of 718.30 Mb and a contig N50 size of 6.36 Mb. Twelve pseudochromosomes were constructed by Hi-C scaffolding, with lengths ranging from 41.03 to 92.03 Mb, accounting for 99.72% of the final assembly.

There is one draft genome assembly of *Castanea crenata* (under accession number GCA_020976635.1) released in NCBI that was generated from one 400-year-old native tree in South Korea. Compared to this genome assembly, our primary assembly shows a longer contig N50 (2.70 Mb vs 6.36 Mb). The BUSCO score of this assembly was also lower than that of our assembly (95.8% vs 97.6%). Since this assembly is only at the scaffold to draft level, and no other information is available, further research is needed when this assembly improves to chromosome-level.

Compared to the genome assembly of the Japanese chestnut cultivar ‘Ginyose’ (Shirasawa et al., 2021), the contig N50 of the genome assembly of cultivar ‘Tsukuba’ improved from 1.59 Mb to 6.36 Mb, and the cumulative length of these two assemblies also confirmed the continuity of the genome of cultivar ‘Tsukuba’ (Figure S5). There were also more sequences anchored to the pseudochromosomes in the genome assembly of cultivar ‘Tsukuba’. Ten out of 12 pseudochromosomes in the genome assembly of ‘Tsukuba’ were longer than the genome assembly of ‘Ginyose’ (Table S3), which resulted in a total of 34.5 Mb longer sequence in our final genome assembly. There were more complete BUSCOs identified in the genome assembly of cultivar ‘Tsukuba’ than in that of ‘Ginyose’ (97.6% vs. 96.6%) (Shirasawa et al., 2021).

Compared to the two other genome assemblies of Japanese chestnut, our genome assembly of the cultivar ‘Tsukuba’ shows longer continuity and better quality. This high-quality genome may help us identify genes in the Japanese chestnut genome and

investigate the domestication and evolutionary histories of Japanese chestnut.

Annotation and evolutionary history of the Japanese chestnut genome

A total of 46,744 gene models encoding 46,463 proteins were predicted in the genome of cultivar ‘Tsukuba’. However, 69,980 high-confidence genes were predicted in the genome of the cultivar ‘Ginyose’ (Shirasawa et al., 2021). Compared with the former genome annotation of Chinese chestnut (Table 2), our genome annotation of Japanese chestnut predicted a similar number of gene models as the research conducted by Sun et al. (Sun et al., 2020) and predicted more gene models than the research performed by Wang et al. (Wang et al., 2020b), Xing et al. (Xing et al., 2019) and Hu et al. (Hu et al., 2022). Compare the BUSCO scores of these predicted gene sets, more than 193 complete BUSCOs were predicted in the Genome of ‘Tsukuba’ than that in the Genome of ‘Ginyose’ (Table 2). There were also less missing BUSCOs in the genome of ‘Tsukuba’ than ‘Ginyose’. This result suggested a better annotation of the genome ‘Tsukuba’ than ‘Ginyose’. However, the other three Chinese chestnut genomes generated even better annotations, further research is needed to improve the annotation of the Japanese chestnut genome (Table 2).

A total of 421.37 Mb of the sequence was annotated as repetitive sequences in the genome assembly of cultivar ‘Tsukuba’, which was similar to the genome assembly of cultivar ‘Ginyose’ (Shirasawa et al., 2021). The two Japanese chestnut genome assemblies exhibit a repetitive rate similar to that of the genome assembly of Chinese chestnut (431.41 Mb, 55.74%; 437.75 Mb, 64.38%; 423.16 Mb, 53.49%; 442.76 Mb, 64.43%) described by Sun et al. (Sun et al., 2020) and Hu et al. (Hu et al., 2022), but slightly different from that found in research by Wang et al. (366.84 Mb, 53.24%) (Wang et al., 2020b) and Xing et al. (390 Mb, 49.69%) (Xing et al., 2019). This difference might be caused by the exclusion of duplicated contigs in our research or the use of different annotation methods.

Phylogenetic analysis based on single-copy orthologous genes from six other species in Fagales whose genomes have been sequenced, namely, *Castanea mollissima*, *Quercus robur*, *Quercus mongolica*, *Juglans regia*, *Carya illinoensis*, and *Corylus mandshurica* (Plomion et al., 2018; Neale et al., 2020; Sun et al., 2020; Li et al., 2021; Lovell et al., 2021; Ai et al., 2022), revealed that Japanese chestnut and Chinese chestnut diverged 6.5 MYA.

Nut development-related genes in Japanese chestnut

Chestnut, as a tree species that has been used to fight against hunger throughout history (Beccaro et al., 2019), should be given

more attention and studied. Japanese chestnut is one of the four major chestnut trees in the world. Compared to Chinese chestnut, Japanese chestnut has a larger nut size and better yields, which are advantageous for lessening hunger.

Starch is one of the most important components of chestnuts, and accounts for 50–80% of their dry matter content (Liu et al., 2015). Chestnut starch is considered a potentially functional component of dietary fiber, which may be a source of resistant starch, thus improving health (Liu et al., 2022). Chestnut starch has unique physicochemical properties, such as high swelling power, freeze–thaw stability, pasting viscosity, and low gelatinization temperature (Liu et al., 2015; Liu et al., 2019). In our study, in the Starch and sucrose metabolism pathway, 14 proteins were identified as trehalose 6-phosphate phosphatase (TPP) in the Japanese chestnut genome, and 14 proteins in the Chinese chestnut genome were identified as TPP. However, only 2 proteins in pedunculate oak were identified as TPP, and no TPPs were identified in Mongolian oak (Plomion et al., 2018; Ai et al., 2022). Synteny analysis of this region between the Japanese chestnut and Chinese chestnut genomes suggested that tandem gene duplication events occurred in these two genomes. The expansion of these TPPs in the Japanese chestnut and Chinese chestnut genomes might contribute to the higher sugar or starch content in the nut of chestnut compared with oak (Plomion et al., 2018). Meanwhile, the starch and sucrose metabolism pathway had the most gene hits. Most of these genes showed an increased expression level at 80 days after flowering and slight decrease at 90 days after flowering, indicating that starch was mainly synthesized at 80 days after flowering in the nut development process of Japanese chestnut. Therefore, water and fertilizer management during this period is extremely important.

Conclusion

In this study, we assembled the genome of the Japanese chestnut cultivar ‘Tsukuba’ by Nanopore long-read sequencing and Illumina sequencing and generated a chromosome-level assembly by Hi-C sequencing. The final assembly had a size of 718.30 Mb and consisted of 12 pseudochromosomes ranging from 41.03 to 92.03 Mb in length, with a BUSCO complete gene percentage of 97.6%. A total of 421.37 Mb of repetitive sequences and 46,744 gene models encoding 46,463 proteins were annotated in the genome. Genome evolution analysis showed that Japanese chestnut is closely related to Chinese chestnut, and they shared a common ancestor ~6.5 million years ago. There were 987 expanded gene families comprising 3,981 genes and 1,204 contracted gene families comprising 1,377 genes in the Japanese chestnut genome. Synteny analysis of this region between the Japanese chestnut and Chinese chestnut genomes suggested that tandem gene duplication events occurred in these

two genomes. The expansion of these TPPs in the Japanese chestnut and Chinese chestnut genomes might contribute to the higher sugar or starch content in their nuts of chestnut.

Data availability statement

The data that support the findings of this study have been deposited into the CNGB Sequence Archive (CNSA) of the China National GeneBank DataBase (CNGBdb) with accession number CNP0003446. Genome assembly and annotation data are available at Figshare: <https://doi.org/10.6084/m9.figshare.21391389>.

Author contributions

JW, QQ and QL conducted the experiments and analyzed the data. QQ and PH analyzed the data and prepared the manuscript. DZ, LZ, KL, SS, SJ, BS and SZ performed the collection and processing of samples and analyzed the data. All authors contributed to the article and approved the submitted version.

Funding

This work was financially supported by the Shandong Provincial Key Laboratory for Fruit Biotechnology Breeding, the Special Fund for Innovation Teams of Fruit Trees in Agricultural Technology System of Shandong Province (SDAIT-06-01), and National Germplasm Repository of Walnut and Chestnut (Tai’an).

Acknowledgments

We are grateful to the Wuhan Benagen Tech Solutions Company Limited (Wuhan, China) for providing technical support.

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.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their

affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Ai, W., Mei, M., Zhang, X., Zhang, L., Han, X., Zhan, H., et al. (2022). A chromosome-scale genome assembly of the Mongolian oak (*Quercus mongolica*). *Mol. Ecol. Resour.* 22 (6), 2396–2410. doi: 10.1093/aem/160216757.73285629/v1
- Albert, V. A., Barbazuk, W. B., dePamphilis, C. W., Der, J. P., Leebens-Mack, J., Ma, H., et al. (2013). The *Amborella* genome and the evolution of flowering plants. *Science* 342 (6165), 1241089. doi: 10.1126/science.1241089
- Apweiler, R., Bairoch, A., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32 (Database issue), D115–D119. doi: 10.1093/nar/gkh131
- Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., et al. (2019). KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36 (7), 2251–2252. doi: 10.1093/bioinformatics/btz859
- Beccaro, G., Alma, A., Bounous, G., and Gomes-Laranjo, J. (2019). *The chestnut handbook: Crop and forest management* (Boca Raton, FL, USA: CRC Press).
- Chan, P. P., and Lowe, T. M. (2019). tRNAScan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol. Biol.* 1962, 1–14. doi: 10.1007/978-1-4939-9173-0_1
- Chen, C., Chen, H., Zhang, Y., Thomas, H. R., Frank, M. H., He, Y., et al. (2020). TBtools: An integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant* 13 (8), 1194–1202. doi: 10.1016/j.molp.2020.06.009
- Chen, Y., Nie, F., Xie, S.-Q., Zheng, Y.-F., Dai, Q., Bray, T., et al. (2021). Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat. Commun.* 12 (1), 60. doi: 10.1038/s41467-020-20236-7
- Conedera, M., and Krebs, P. (2008). History, present situation and perspective of chestnut cultivation in Europe. *Acta Hortic.* 784, 23–27. doi: 10.17660/ActaHortic.2008.784.1
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., et al. (2021). Twelve years of SAMtools and BCFTools. *Gigascience* 10 (2), giab008. doi: 10.1093/gigascience/giab008
- De Bie, T., Cristianini, N., Demuth, J. P., and Hahn, M. W. (2006). CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* 22 (10), 1269–1271. doi: 10.1093/bioinformatics/btl097
- Dudchenko, O., Batra, S. S., Omer, A. D., Nyquist, S. K., Hoeger, M., Durand, N. C., et al. (2017). *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* 356 (6333), 92–95. doi: 10.1126/science.aal3327
- Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* 3 (1), 95–98. doi: 10.1016/j.cels.2016.07.002
- El-Gebali, S., Mistry, J., Bateman, A., Eddy, S. R., Luciani, A., Potter, S. C., et al. (2019). The Pfam protein families database in 2019. *Nucleic Acids Res.* 47 (D1), D427–D432. doi: 10.1093/nar/gky995
- Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* 20 (1), 238. doi: 10.1186/s13059-019-1832-y
- Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2: automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* 117 (17), 9451–9457. doi: 10.1073/pnas.1921046117
- Goel, M., Sun, H., Jiao, W. B., and Schneeberger, K. (2019). SyRI: finding genomic rearrangements and local sequence differences from whole-genome assemblies. *Genome Biol.* 20 (1), 277. doi: 10.1186/s13059-019-1911-0
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-seq data without a reference genome. *Nat. Biotechnol.* 29, 644. doi: 10.1038/nbt.1883
- 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 (1), R7. doi: 10.1186/gb-2008-9-1-r7
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The gene ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32 (Database issue), D258–D261. doi: 10.1093/nar/gkh036
- Hoff, K. J., and Stanke, M. (2019).). predicting genes in single genomes with AUGUSTUS. *Curr. Protoc. Bioinf.* 65 (1), e57. doi: 10.1002/cpbi.57
- Hu, G., Cheng, L., Cheng, Y., Mao, W., Qiao, Y., and Lan, Y. (2022). Pan-genome analysis of three main Chinese chestnut varieties. *Front. Plant Sci.* 13. doi: 10.3389/fpls.2022.916550
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., et al. (2017). Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evolution.* 34 (8), 2115–2122. doi: 10.1093/molbev/msx148
- Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long read assembly. *Bioinformatics* 36 (7), 2253–2255. doi: 10.1093/bioinformatics/btz891
- Jaillon, O., Aury, J. M., Noel, B., Polcristi, A., Clepet, C., Casagrande, A., et al. (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449 (7161), 463–467. doi: 10.1038/nature06148
- Kall, L., Krogh, A., and Sonnhammer, E. L. (2004). A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* 338 (5), 1027–1036. doi: 10.1016/j.jmb.2004.03.016
- Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2014). Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res.* 42 (Database issue), D199–D205. doi: 10.1093/nar/gkt1076
- Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37 (8), 907–915. doi: 10.1038/s41587-019-0201-4
- Koonin, E. V., Fedorova, N. D., Jackson, J. D., Jacobs, A. R., Krylov, D. M., Makarova, K. S., et al. (2004). A comprehensive evolutionary classification of proteins encoded in complete eukaryotic genomes. *Genome Biol.* 5 (2), R7. doi: 10.1186/gb-2004-5-2-r7
- Korf, I. (2004). Gene finding in novel genomes. *BMC Bioinf.* 5, 59. doi: 10.1186/1471-2105-5-59
- Kurogi, M., and Uritani, I. (1966). Isolation and identification of two coumarin derivatives from Japanese chestnuts. *Agric. Biol. Chem.* 30 (1), 78–82. doi: 10.1271/abb1961.30.78
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26 (5), 589–595. doi: 10.1093/bioinformatics/btp698
- Li, Y., Sun, P., Lu, Z., Chen, J., Wang, Z., Du, X., et al. (2021). The corylus mandshurica genome provides insights into the evolution of betulaceae genomes and hazelnut breeding. *Horticulture Res.* 8 (1), 54. doi: 10.1038/s41438-021-00495-1
- Liu, T., Ma, M., Guo, K., Hu, G., Zhang, L., and Wei, C. (2019). Structural, thermal, and hydrolysis properties of large and small granules from c-type starches of four Chinese chestnut varieties. *Int. J. Biol. Macromolecules* 137, 712–720. doi: 10.1016/j.ijbiomac.2019.07.023
- Liu, C., Wang, S., Chang, X., and Wang, S. (2015). Structural and functional properties of starches from Chinese chestnuts. *Food Hydrocolloids* 43, 568–576. doi: 10.1016/j.foodhyd.2014.07.014
- Liu, W., Zhang, Y., Wang, R., Li, J., Pan, W., Zhang, X., et al. (2022). Chestnut starch modification with dry heat treatment and addition of xanthan gum: Gelatinization, structural and functional properties. *Food Hydrocolloids* 124, 107205. doi: 10.1016/j.foodhyd.2021.107205
- Li, Z., Wei, X., Tong, X., Zhao, J., Liu, X., Wang, H., et al. (2022). The OsNAC23-Tre6P-SnRK1a feed-forward loop regulates sugar homeostasis and grain yield in rice. *Mol. Plant* 15 (4), 706–722. doi: 10.1016/j.molp.2022.01.016
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42 (Database issue), D490–D495. doi: 10.1093/nar/gkt1178

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1049253/full#supplementary-material>

- Lomsadze, A., Ter-Hovhannisyanyan, V., Chernoff, Y. O., and Borodovsky, M. (2005). Gene identification in novel eukaryotic genomes by self-training algorithm. *Nucleic Acids Res.* 33 (20), 6494–6506. doi: 10.1093/nar/gki937
- Lovell, J. T., Bentley, N. B., Bhattarai, G., Jenkins, J. W., Sreedasyam, A., Alarcon, Y., et al. (2021). Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. *Nat. Commun.* 12 (1), 4125. doi: 10.1038/s41467-021-24328-w
- Love, J., Palmer, J., Stajich, J., Esser, T., Kastman, E., Bogema, D., et al. (2019). Nextgenusfs/funannotate: funannotate v1.7.2. *Zenodo*. doi: 10.5281/zenodo.3594559
- Majoros, W. H., Pertea, M., and Salzberg, S. L. (2004). TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* 20 (16), 2878–2879. doi: 10.1093/bioinformatics/bth315
- Nakamura, M. (1994). Elongation of pollen tubes and degeneration of ovules in Japanese Chestnut (*Castanea crenata* sieb. et zucc.). *J. Japanese Soc. Hortic. Sci.* 63 (2), 277–282. doi: 10.2503/jjshs.63.277
- Neale, D. B., Salzberg, S. L., Dandekar, A., Timp, W., Leslie, C. A., Troggo, M., et al. (2020). High-quality chromosome-scale assembly of the walnut (*Juglans regia* L.) reference genome. *GigaScience* 9 (5), gaa050. doi: 10.1093/gigascience/gaa050
- Nielsen, H. (2017). Predicting secretory proteins with SignalP. *Methods Mol. Biol.* 1611, 59–73. doi: 10.1007/978-1-4939-7015-5_6
- Nishio, S., Takada, N., Terakami, S., Takeuchi, Y., Kimura, M. K., Isoda, K., et al. (2021). Genetic structure analysis of cultivated and wild chestnut populations reveals gene flow from cultivars to natural stands. *Sci. Rep.* 11 (1), 240. doi: 10.1038/s41598-020-80696-1
- Nuccio, M. L., Wu, J., Mowers, R., Zhou, H. P., Meghji, M., Primavesi, L. F., et al. (2015). Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. *Nat. Biotechnol.* 33 (8), 862–869. doi: 10.1038/nbt.3277
- Ou, S., Chen, J., and Jiang, N. (2018). Assessing genome assembly quality using the LTR assembly index (LAI). *Nucleic Acids Res.* 46 (21), e126. doi: 10.1093/nar/gky730
- Plomion, C., Aury, J.-M., Amselem, J., Leroy, T., Murat, F., Duplessis, S., et al. (2018). Oak genome reveals facets of long lifespan. *Nat. Plants* 4, 440–452. doi: 10.1038/s41477-018-0172-3
- Rawlings, N. D., Barrett, A. J., Thomas, P. D., Huang, X., Bateman, A., and Finn, R. D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* 46 (D1), D624–D632. doi: 10.1093/nar/gkx1134
- Roach, M. J., Schmidt, S. A., and Borneman, A. R. (2018). Purge haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinf.* 19 (1), 460. doi: 10.1186/s12859-018-2485-7
- Sakamoto, D., Inoue, H., Kusaba, S., Sugiura, T., and Moriguchi, T. (2015). The effect of nitrogen supplementation by applying livestock waste compost on the freezing tolerance of Japanese chestnut. *Horticulture J.* 84 (4), 314–322. doi: 10.2503/hortj.MI-046
- Sato, I., Kofujita, H., and Tsuda, S. (2007). Identification of COX inhibitors in the hexane extract of Japanese horse chestnut (*Aesculus turbinata*) seeds. *J. Veterinary Med. Sci.* 69 (7), 709–712. doi: 10.1292/jvms.69.709
- Seppely, M., Manni, M., and Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* 1962, 227–245. doi: 10.1007/978-1-4939-9173-0_14
- Shirasawa, K., Nishio, S., Terakami, S., Botta, R., Marinoni, D. T., and Isobe, S. (2021). Chromosome-level genome assembly of Japanese chestnut (*Castanea crenata* sieb. et zucc.) reveals conserved chromosomal segments in woody rosids. *DNA Res.* 28 (5), dsab016. doi: 10.1093/dnares/dsab016
- Staton, M., Addo-Quaye, C., Cannon, N., Yu, J., Zhebentyayeva, T., Huff, M., et al. (2020). A reference genome assembly and adaptive trait analysis of castanea mollissima ‘Vanuxem,’ a source of resistance to chestnut blight in restoration breeding. *Tree Genet. Genomes* 16 (4), 57. doi: 10.1007/s11295-020-01454-y
- Sun, Y., Lu, Z., Zhu, X., and Ma, H. (2020). Genomic basis of homoploid hybrid speciation within chestnut trees. *Nat. Commun.* 11 (1), 1–10. doi: 10.1038/s41467-020-17111-w
- Su, W., Sun, J., Shimizu, K., and Kadota, K. (2019). TCC-GUI: a shiny-based application for differential expression analysis of RNA-seq count data. *BMC Res. Notes* 12 (1), 133. doi: 10.1186/s13104-019-4179-2
- Tanaka, K., and Kotobuki, K. (1992). Analysis of histological and chemical factors related to the ease of pellicle removal in Japanese chestnut. *J. Japanese Soc. Hortic. Sci.* 61 (1), 1–6. doi: 10.2503/jjshs.61.1
- Tang, H., Krishnakumar, V., and Li, J. (2015). Jcvi: JCVI utility libraries (v0.5.7). *Zenodo*. doi: 10.5281/zenodo.31631
- Testa, A. C., Hane, J. K., Ellwood, S. R., and Oliver, R. P. (2015). CodingQuarry: highly accurate hidden Markov model gene prediction in fungal genomes using RNA-seq transcripts. *BMC Genomics* 16, 170. doi: 10.1186/s12864-015-1344-4
- Valliyodan, B., Cannon, S. B., Bayer, P. E., Shu, S., Brown, A. V., Ren, L., et al. (2019). Construction and comparison of three reference-quality genome assemblies for soybean. *Plant J.* 100 (5), 1066–1082. doi: 10.1111/tpj.14500
- Wang, J., Liu, W., Zhu, D., Hong, P., Zhang, S., Xiao, S., et al. (2020a). Chromosome-scale genome assembly of sweet cherry (*Prunus avium* L.) cv. tieton obtained using long-read and Hi-c sequencing. *Horticulture Res.* 7 (1), 122. doi: 10.1038/s41438-020-00343-8
- Wang, J., Tian, S., Sun, X., Cheng, X., Duan, N., Tao, J., et al. (2020b). Construction of pseudomolecules for the Chinese chestnut (*Castanea mollissima*) genome. *G3 Genes|Genomes|Genetics* 10 (10), 3565–3574. doi: 10.1534/g3.120.401532
- 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 (9), giz112. doi: 10.1093/gigascience/giz112
- Xu, M., Guo, L., Gu, S., Wang, O., Zhang, R., Peters, B. A., et al. (2020). TGS-GapCloser: A fast and accurate gap closer for large genomes with low coverage of error-prone long reads. *GigaScience* 9 (9), gaa094. doi: 10.1093/gigascience/gaa094
- Zhang, L., Hu, J., Han, X., Li, J., Gao, Y., Richards, C. M., et al. (2019a). A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nat. Commun.* 10 (1), 1494. doi: 10.1038/s41467-019-09518-x
- Zhang, X., and Jonassen, I. (2020). RASflow: an RNA-seq analysis workflow with snakemake. *BMC Bioinf.* 21 (1), 110. doi: 10.1186/s12859-020-3433-x
- Zhang, X., Zhang, S., Zhao, Q., Ming, R., and Tang, H. (2019b). Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-c data. *Nat. Plants* 5 (8), 833–845. doi: 10.1038/s41477-019-0487-8
- Zhi, D., Raphael, B. J., Price, A. L., Tang, H., and Pevzner, P. A. (2006). Identifying repeat domains in large genomes. *Genome Biol.* 7 (1), R7. doi: 10.1186/gb-2006-7-1-r7
- Zhou, Y., Minio, A., Massonnet, M., Solares, E., Lv, Y., Beridze, T., et al. (2019). The population genetics of structural variants in grapevine domestication. *Nat. Plants* 5 (9), 965–979. doi: 10.1038/s41477-019-0507-8
- Zwaenepoel, A., and Van de Peer, Y. (2019). Wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* 35 (12), 2153–2155. doi: 10.1093/bioinformatics/bty915