Check for updates

The chromosome-level genome of *Eucommia ulmoides* provides insights into sex differentiation and α-linolenic acid biosynthesis

Qingxin Du[1,2,3†], Zixian Wu[4†], Panfeng Liu[1,2,3†], Jun Qing[1,2,3], Feng He[1,2,3], Lanying Du[1,2,3], Zhiqiang Sun[1,2,3], Lili Zhu[5], Hongchu Zheng[6], Zongyi Sun[7], Long Yang[4*], Lu Wang[1,2,3*] and Hongyan Du[1,2,3*‡]

[1]Research Institute of Non-timber Forestry, Chinese Academy of Forestry, Zhengzhou, China, [2]Key Laboratory of Non-timber Forest Germplasm Enhancement and Utilization of National Forestry and Grassland Administration, Chinese Academy of Forestry, Zhengzhou, China, [3]Engineering Research Center of Eucommia ulmoides, State Forestry and Grassland Administration, Zhengzhou, China, [4]Agricultural Big-Data Research Center and College of Plant Protection, Shandong Agricultural University, Taian, China, [5]Academy of Chinese Medical Sciences, Henan University of Chinese Medicine, Zhengzhou, China, [6]Product Department, Henan Jinduzhong Agricultural Science and Technology Co., Ltd., Yanling, China, [7]Operation Department, Grandomics Biosciences Co., Ltd., Wuhan, China

Eucommia ulmoides Oliver is a typical dioecious plant endemic to China that has great medicinal and economic value. Here, we report a high-quality chromosome-level female genome of *E. ulmoides* obtained by PacBio and Hi-C technologies. The size of the female genome assembly was 1.01 Gb with 17 pseudochromosomes and 31,665 protein coding genes. In addition, Hi-C technology was used to reassemble the male genome released in 2018. The reassembled male genome was 1.24 Gb with the superscaffold N50 (48.30 Mb), which was increased 25.69 times, and the number of predicted genes increased by 11,266. Genome evolution analysis indicated that *E. ulmoides* has undergone two whole-genome duplication events before the divergence of female and male, including core eudicot γ whole-genome triplication event (γ-WGT) and a recent whole genome duplication (WGD) at approximately 27.3 million years ago (Mya). Based on transcriptome analysis, *EuAP3* and *EuAG* may be the key genes involved in regulating the sex differentiation of *E. ulmoides*. Pathway analysis showed that the high expression of ω-3 fatty acid desaturase coding gene *EU0103017* was an important reason for the high α-linolenic acid content in *E. ulmoides*. The genome of female and male *E. ulmoides* presented here is a valuable resource for the molecular biological study of sex differentiation of *E. ulmoides* and also will provide assistance for the breeding of superior varieties.

KEYWORDS

Eucommia ulmoides Oliver, genome, whole-genome duplication, MADS-box genes, sex differentiation, α-linolenic acid biosynthesis

Introduction

There has been a wide range of uses for *Eucommia ulmoides* Oliver, such as medicinal and natural rubber extraction purposes. Because there has been a long history of application of this plant in China (Zhu and Sun, 2018), it has aroused the interest of many researchers. In 2018, the first draft reference genome of *E. ulmoides* (Male V1) was released based on next-generation sequencing, and consisted of a total genome length of 1.18 Gb genome with contig N50 that was 17.06 kb in length, and the assembled genome size was larger than the 1.1 Gb predicted by 17-mer analysis. A large number of superscaffolds were assembled from contigs, but none was completely anchored to the chromosomes (Wuyun et al., 2018).

It was realized that the accuracy and completeness of this genome assembly could be further refined. A new version of the genome was subsequently published in 2020, whereby the authors first obtained the *Eucommia* haploid plant through parthenogenesis, and then sequenced and assembled it. The genome size was 947.86 Mb, accounting for 92.93% of the estimated genome size (1.02 Gb). Contig N50 was 13.16 Mb, and scaffold N50 was 53.15 Mb in length, and long scaffolds were further anchored to 17 pseudochromosomes by Hi-C (Li et al., 2020). Although two versions of the *E. ulmoides* genome have been released, a high-quality female *E. ulmoides* genome has yet to be produced, which is necessary for the study of evolution and sex differentiation.

E. ulmoides is a dioecious perennial woody plant, with male and female inflorescences that greatly vary, and thus, the quantities of important compounds in male and female plant are different (Figure 1). The utilization value of *E. ulmoides* also varies with the sex of the plant (Wang et al., 2020). In flowering plants, the sex difference is mainly manifested as different flower organs, and therefore, genes related to flower development may participate in the process of sex differentiation (Barrett and Hough, 2012; Massonnet et al., 2020). As the key regulators of almost all aspects of plant reproductive development, MADS-box genes play a particularly prominent role in flowering time control, inflorescence structure, floral organ identity determination, and seed development. However, the floral induction and floral organ regulation genes and regulatory networks of *E. ulmoides* are still unclear.

Based on the studies of model plants *Arabidopsis thaliana* (Bowman et al., 1989) and *Antirrhinum majus* (Schwarz-Sommer et al., 1990), the genetic regulation mechanisms that govern flower organ development have been preliminarily explored, and the famous ABC model of flower organ development has been proposed (Coen and Meyerowitz, 1991; Soltis et al., 2007), which was later supplemented by the ABCDE model (Theissen, 2001; Causier et al., 2010). Except for the *AP2* gene, all the genes involved in the ABCDE model were MADS-box genes belonging to different functional categories (Hu et al., 2021). Therefore, the MADS-box gene family may play an important role in regulating the sexual differentiation of flowers, and a large number of MADS-box genes and their functions have been found and verified in plants such as *Arabidopsis* (Di Marzo et al., 2020), *Orchidaceae* (Teo et al., 2019), *Petunia* (Morel et al., 2019), *Oryza sativa* (Yin et al., 2019) and *Zea*
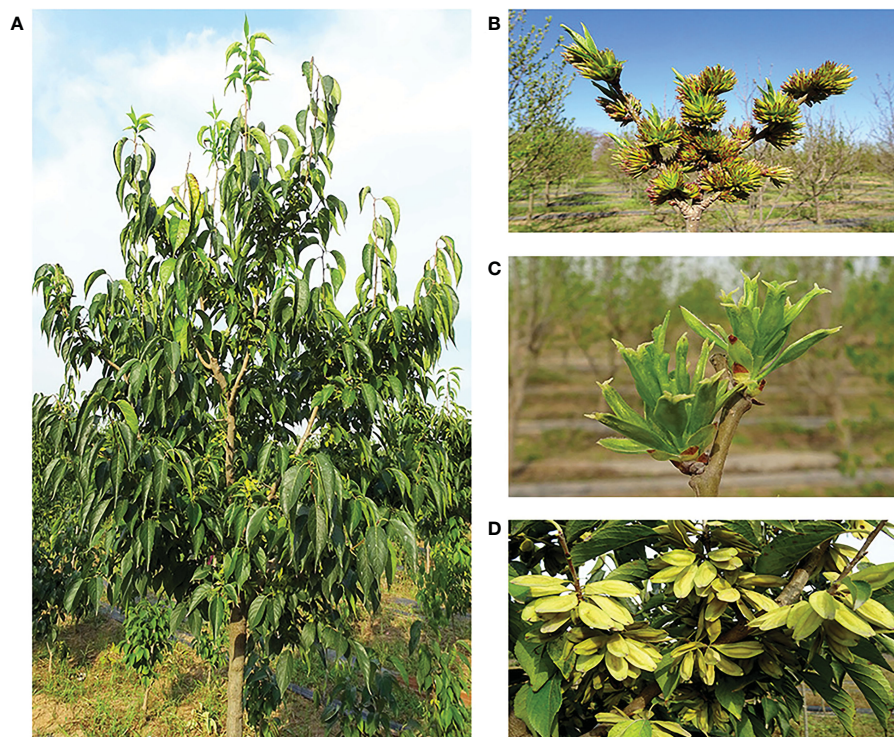

FIGURE 1
Morphological characteristic of *Eucommia ulmoides*. **(A)** Mature plant, **(B)** male inflorescence, **(C)** female inflorescence, **(D)** fruit.

mays (Abraham-Juárez et al., 2020). A previous study based on comparative transcriptome analysis of male and female *E. ulmoides* identified putative sex-associated genes (Wang and Zhang, 2017). Regrettably, only one gene with high homology to the *AP3* gene in *Arabidopsis* was detected, which was a B class organ identity gene. Although that study provides a reference for exploring the sex differentiation of *E. ulmoides*, there is currently a lack of information on the sex differentiation mechanism of *E. ulmoides*.

In plants, stearic acid is catalyzed by fatty acid desaturase to produce oleic acid and linoleic acid, and eventually α-linolenic acid (Li-Beisson et al., 2013), which is an essential fatty acid and is very important for human health (Kim et al., 2014). There has been prior study on α-linolenic in plants such as *Perilla frutescens* (Liao et al., 2018), *Plukenetia volubilis* (Hu et al., 2018), *Paeonia ostia* (Yu et al., 2021). *E. ulmoides* is a woody plant with a very high α-linolenic acid content, even more than 60% in its seed oil (Zhang et al., 2018), well above purslane which has the highest content of linolenic acid in green leafy vegetables (30.15%). Although the synthesis of α-linolenic acid in *E. ulmoides* has also been studied from the glycolytic pathway based on transcriptome analysis (Feng et al., 2016), the information in genes encoding key enzymes for linolenic acid synthesis and the mechanism of α-linolenic acid accumulation are still not clear.

In this study, we aimed to construct a female and male *E. ulmoides* genome and explore the molecular mechanism of sex differentiation and α-linolenic acid accumulation in *E. ulmoides*. Our results will provide essential resources for studies on *E. ulmoides* evolution and directional variety improvement, and also lay a foundation for the rapid identification of male and female plants, thus improving the utilization rate of *E. ulmoides*.

Materials and methods

Plant samples

Fresh leaves for genome sequencing were collected from a female *E. ulmoides* plant 'Huazhong No. 8', which was used to assemble the Female V1 genome. To observe the morphological structure of female and male flower bud development, flower buds were collected from female 'Huazhong No. 6' and the male 'Huazhong No. 11' during the entire development period. Samples were obtained every 5 days from the beginning of bud germination in late April to the end of September, once every 10 days from early October to January of the following year, and once every 3 days from February to early April of the following year, with 20 female and male flower buds being collected each time. The experimental materials for sex differentiation analysis were derived from the collected flower buds. Male and female flower buds at the floral organ induction stage (the inflorescence primordium formation stage), floral organ morphological differentiation initial stage (the pistil and stamen differentiation stage), and flower organ maturity stage were selected for transcriptome sequencing, with three biological repetitions for each sample. To analyze the expression of α-linolenic acid synthesis genes, four different tissues (stem, bark, leaf and fruit) of female 'Huazhong No. 8'

were extracted. All plant samples were collected from the Yuanyang Experimental Base of Research Insitute of Non-timber Forestry, Chinese Academy of Forestry. Plant materials were immediately immersed in liquid nitrogen, transported to the laboratory, and then stored in a -80℃ freezer for sequencing.

Morphological and structural observation of flower bud development

After collection, the materials were immediately immersed in FAA fixative solution (70% ethanol: glacial acetic acid: 38% formaldehyde =18:1:1), transported to the laboratory, and then stored in a -80° freezer. Materials preserved in FAA fixative solution were sliced using a conventional paraffin-sectioning method. Then the slices were observed and photographed using OLYMPUS optical microscope.

Sequencing and library construction

Total genomic DNA was extracted from fresh leaves using the QIAGEN® Genomic DNA extraction kit (Cat#13323, QIAGEN) for PacBio sequencing on a PacBio Sequel II instrument. For Hi-C library construction and sequencing, restriction endonucleases (HindIII/MboI) were used for chromatin digestion, and after biotin labeling, flat-end binding and DNA purification. Hi-C samples were prepared and sampled for DNA quality testing. The Hi-C fragment consisted of removed biotin that was fragmented by ultrasound, end-repaired, and after the addition of base A, the biotin-containing fragment was isolated and then sequenced to form the jointed product. Then, Polymerase Chain Reaction (PCR) conditions were tested, and DNA was amplified to obtain the library products (Kong and Zhang, 2019). An Illumina HiSeq X Ten sequencer was used for sequencing after the library quality control was confirmed.

RNA was extracted using a TRIzol kit (Invitrogen) according to the manufacturer's instructions. The RNA concentration was measured using a NanoDrop spectrophotometer, and RNA purity and integrity were determined by an Agilent 2100 Bioanalyzer and 1% agarose gel electrophoresis, respectively. After passing the quality inspection, a cDNA library was constructed according to the instructions accompanying the cDNA Library Construction Kit (NEB). Paired-end sequencing was performed on the cDNA library using the Illumina HiSeq X-10 platform, and 150 bp paired-end sequences were generated.

Genome assembly and evaluation

For the genome assembly of female *E. ulmoides* plant 'Huazhong No. 8', first, sub-sequences from PacBio sequencing were assembled using Falcon v2.0.5 (https://github.com/PacificBiosciences/FALCON/) with the parameter '–max diff 100 –max cov 100 –min cov 2 –min len 5000', the pure third-generation assembly software officially launched by PacBio. Then, Arrow (Chin

et al., 2013) was applied to align the third generation data to the preliminary assembly for correction with default parameters. After the initial genome assembly was completed, Pilon v1.22 (Walker et al., 2014) with the parameter '–mindepth 10 –changes –fix bases' was used for iterative polishing. The Hi-C data were aligned to the genome assembly *via* Juicer v1.5 software with the following parameter settings: -s DpnII -t 20 (Durand et al., 2016b). Finally, the draft genome of *E. ulmoides* was assembled onto chromosomes by 3D-DNA pipeline with the parameter '-m haploid -s 4 -c 17 -j 15' (Dudchenko et al., 2017). The contact matrix and heatmap of chromosomes were drawn with Juicebox (Durand et al., 2016a).

As for the male *E. ulmoides* tree (SNJ) genome reassembly, named Male V2 genome, we first downloaded the previous genome assembly data from the NCBI under accession number SRP095726. Next, the downloaded genome was refined using Hi-C data with the same method as that used for the female genome assembly. Benchmarking Universal Single-Copy Orthologs (BUSCO) v4.1.4 (https://busco.ezlab.org/) was used to assess the integrity and accuracy of the final genome assembly with the Eudicotyledons_odb10 database (Simão et al., 2015).

Genome annotation

TRF v4.07 (Benson, 1999) (tandem repeats finder) software (https://tandem.bu.edu/trf/trf.html) was applied to predict tandem repeats of the genome. Transposable elements (TEs) were identified by combining *de novo*-based and homology-based approaches. In the *de novo*-based approach, RepeatModeler v1.0.11 (Flynn et al., 2020) (http://www.repeatmasker.org/RepeatModeler/) and LTR-FINDER v1.05 (Xu and Wang, 2007) were used to build the repeat library (LTR length 100 to 5000 nt; length between two LTRs: 1000 to 20,000 nt), and then, RepeatMasker v4.0.7 (Tarailo-Graovac and Chen, 2009) (http://www.repeatmasker.org/) was used to identify and classify repeats under the parameter '-a -e ncbi -q -norna -nolow -div 30 -cutoff 225'. For the homology-based approach, the known repetitive sequences database Repbase (Bao et al., 2015) (https://www.girinst.org/repbase/) was searched by RepeatMasker to find homologous sequences in the *E. ulmoides* genome.

Gene prediction for the *E. ulmoides* genome was performed by integrating three different methods: *de novo* prediction, homology-based prediction and transcriptome-based prediction. *De novo* gene prediction was achieved using Augustus v3.3 (parameter: -strand=both -genemodel=partial -gff3=on -species= arabidopsis) (Stanke and Morgenstern, 2005) (http://bioinf.uni-greifswald.de/augustus/), SNAP (Korf, 2004) and GlimmerHMM v3.52 (Majoros et al., 2004). For the homology approach, BLASTP (Camacho et al., 2009) was used to map protein sequences onto the *E. ulmoides* genome, and GeneWise v2.4.1 (Birney et al., 2004) was used to align the homologous genome sequences with the matching proteins. RNA-seq sequences were mapped to the genome assembly using PASA v2.0.2 to identify putative exon regions and splice junctions. Finally, EVidenceModeler v1.1.1 (Haas et al., 2008) (https://evidencemodeler.github.io/) was employed to integrate the gene sets predicted by the three

methods. For gene annotation, EggNOG-mapper v4.5 (Huerta-Cepas et al., 2017) software was employed with emapper DB 4.5.1 to obtain the Clusters of Orthologous Genes (COG), eggNOG, Gene Ontology (GO) and KEGG pathway information for each gene. HMMER v3.2.1 (Prakash et al., 2017) was applied to search the Pfam database (http://pfam.xfam.org/) and obtain the protein domains. Functional annotation of protein coding genes was performed using the UniProt database (https://www.uniprot.org/).

Four types of non-coding RNAs (ncRNAs), tRNA, rRNA, miRNA, and snRNA, were annotated in the *E. ulmoides* genome. TRNAscan-SE v1.4 (Lowe and Chan, 2016) (http://lowelab.ucsc.edu/tRNAscan-SE/)software was employed to identify the tRNAs with tRNAscn-SE -i -q and eukaryote parameters. rRNAs were detected using BLASTN alignment of rRNA sequences known related species. Other ncRNAs were predicted using INFERNAL v1.1 (Nawrocki et al., 2009) software with the parameter 'cmsearch –ga –incE 0.01 -E 10.0' to search the Rfam v12.0 (Nawrocki et al., 2015) database (http://rfam.xfam.org/).

Detection of WGD events

To identify whole genome duplication (WGD) events in *E. ulmoides*, BLASTP was used to perform a homology search using the female and male genome, and MCScanX v0.8 (Wang et al., 2012) (https://sourceforge.net/projects/mcscanx/) was employed to detect syntenic blocks. The collinearity analysis of male and female was conducted by JCVI v1.2.1. Ks (synonymous substitution) values for paralogous blocks were calculated by KaKs_Calculator v2.0 with the parameter '-i test.axt -m MA -c -o' (Wang et al., 2010), and a frequency distribution graph of the Ks values was drawn to identify WGD events in *E. ulmoides*. The WGD event time was estimated by the formula ks/2r, with a r value of 8.25 e-9.

MADS-box genes analysis

To identify MADS-box genes, HMMER v3.2.1 software with the SRF-TF domain (PF00319) and K-box (PF01486) were used to search against the *E. ulmoides* genome proteins (–cut_ga, E value$<1\times10^{-5}$), which were obtained from Pfam. The search results containing MADS-box domain or K-box domain were retained, and incomplete functional domain sequences were removed. Using all Arabidopsis MADS-box genes as queries, the predicted *E. ulmoides* MADS genes were checked by BLASTP searches (E value$<1\times10^{-5}$). The phylogenetic tree was then constructed using RAxML v8.2.12 with the GTRGAMMA substitution model and 1000 bootstraps on the CIPRES website (https://www.phylo.org/portal2/home.action). The phylogenetic tree was visualized using FigTree v1.4.4. The combination of genome annotation, homology search and MADS-box gene expression in male and female flower buds, enabled the ABCDE model genes to be identified. Using all female *E. ulmoides* MADS-box genes as queries, BLASTP was used to map protein sequences onto the male genome. For these protein sequences, the

phylogenetic tree was constructed through MEGA-X using the maximum likelihood method and the JTT matrix-based model with 1,000 bootstraps. Multiple sequence alignments were performed by CLUSTALW program and viewed in GeneDoc. The gene structure of these MADS-box genes was demonstrated by TBtools. The data of *Nymphaea colorata* was downloaded from BIG Data Center with the accession number GWHAAYW00000000. And the phylogenetic trees of genes related to sex differentiation in *E. ulmoides* and *N. colorata* were constructed using MEGA7 with the Neighbor Joining method.

Identification of α-linolenic acid biosynthesis and metabolism genes

The KEGG database (https://www.kegg.jp/kegg/kegg1.html) was queried to identify genes involved in α-linolenic acid synthesis and metabolism. First, known genes related to α-linolenic acid biosynthesis and metabolic pathways in KEGG were selected and their protein sequences were downloaded. The downloaded protein sequences were aligned with the *E. ulmoides* genome using the NCBI online comparison function BLASTP (https://blast.ncbi.nlm.nih.gov/Blast.cgi), and the annotation information was combined to predict genes involved in α-linolenic acid synthesis and metabolism.

Transcription analysis

FastQC v0.11.9 was used to estimate the quality of RNA sequences. To obtain clean sequences, Trimmomatic v0.36 was used to remove adapter sequences and low-quality sequences (with a base quality value less than 30 or greater than 5% unknown bases). To estimate the gene expression level, clean sequences from RNA-seq were mapped onto the final Female V1 genome assembly using Tophat2 v2.1.1 with the parameters: –b2-sensitive -N 2 -p 6 -g 10 –read-edit-dist 2 –read-mismatches 3 –read-gap-length 2 (Kim et al., 2013), and then fragments per kilobase per million (FPKM) of each gene was performed by Cufflinks v2.2.1 with the following parameter: -p 6 -g file.gff -u –library-type fr-unstranded (Trapnell et al., 2010) (http://cole-trapnell-lab.github.io/cufflinks/). Finally, the values of $\log_2$(FPKM +1)-that represented the gene expression level and the differentially expressed genes (DEGs) were identified using Cuffdiff v2.2.1, with a p-value (FDR) < 0.05 and |log2 (fold change)| > 1.

Results

Morphological and structural observation of female and male flower bud differentiation

With the growth and development of female and male flower buds of *E. ulmoides*, the cells of the flower bud meristem continued dividing, and the internal and external morphological structure

gradually changed. The floral organ development of the female and male flower buds of *E. ulmoides* can be divided into four stages: inflorescence primordium formation, bract differentiation, pistil and stamen differentiation, and pistil and stamen morphological formation (Figure 2).

In the inflorescence primordium formation stage, the female and male flower buds exhibit similar morphological characteristics. The buds are young and green, approximately triangular, with white fluff on the surface (Figures 2A1, C1). The internal anatomical structure shows that the flower buds are apical and conical, with thick cytoplasm (Figures 2B1, D1).
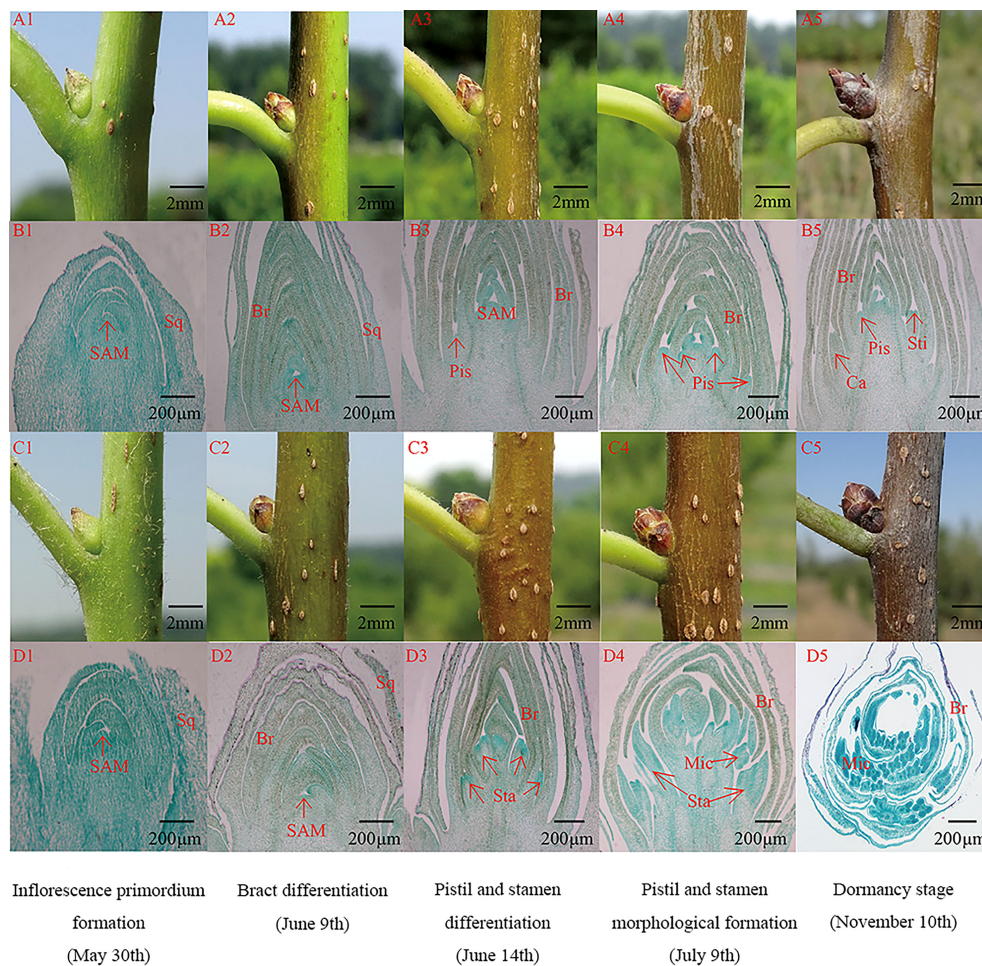
In the bract differentiation stage, the female and male flower bud morphological characteristics remain relatively similar, and males and females cannot be accurately distinguished. The outer bracts of female and male flower buds gradually lignify into scales, and the color of the flower buds gradually changes from green to light brown (Figures 2A2, C2). The number of bracts increases, and the meristem at the bud apex widens and flattens (Figures 2B2, D2).

In the pistil and stamen differentiation stage, the internal structure of female and male flower buds showed obvious morphological differences. Round small protuberances form between the axils of the primordia of the female buds, namely the pistil primordia, and several stamen primordia that form between the axils of the male buds are clustered between the bracts. The male and female buds can be distinguished according to the sectioning results (Figures 2B3, D3).

In the pistil and stamen morphological formation stage, female flower buds grow faster longitudinally and are approximately conical in shape, while the stamens are approximately spherical (Figures 2A4, C4). A single pistil primordium is differentiated from the base to the center on both sides of the female flower bud, while the male flower bud is differentiated from several stamens to form a stamen cluster inside the bract (Figures 2B4, D4). The pistil shape is pinnate, with a central depression at the top that forms two protuberances, namely the pistil stigma. After the pistil completes its morphogenesis, and stamen pollen sac differentiates into the obvious endothecium, middle layer, and tapetum (Figures 2B5, D5). Then, the internal and external morphology of the flower bud does not change, and the bud enters the dormancy stage.

Genome assembly and assessment

A combined strategy of PacBio and Hi-C sequencing technologies was used to construct the high-quality female *E. ulmoides* genome assembly (Female V1). The PacBio long sequences were corrected and assembled into the preliminary genome assembly using FALCON. The obtained genome size was 957 Mb, which consisted of 2,662 contigs with contig N50 that was 1.3 Mb in length. Finally, with the assistance of 184.7 Gb of Hi-C sequences, the assembled contigs were anchored to 17 pseudochromosomes of 1.01 Gb with superscaffold N50 that was 51.89 Mb in length, and the GC content of the genome was 35.14% (Table 1). The interactions of contigs on pseudochromosomes were detected to adjust their order and direction in the Hi-C heat map, which also indicated the high quality of Female V1 genome

FIGURE 2
Internal anatomical structure and external morphological characteristics of male and female flower buds at different differentiation stages. Br, Bract; Ca, Carp; Mic, Microsporocyte; Pis, Pistil, SAM, Stem Apical Meristem; Sq, Squama; Sta, Stamen; Sti, Stigma. **(A1-A5)**: The external morphology structure of female flower bud; **(B1-B5)**: The anatomic structure of female flower bud; **(C1-C5)**: The external morphology structure of male flower bud; **(D1-D5)**: The anatomic structure of male flower bud.

TABLE 1 Summary of the *Eucommia ulmoides* genome.

Type	Female V1	Male V2	Male V1
Genome Size	1.01 Gb	1.24 Gb	1.18 Gb
Contig N50	1.33 Mb	17.06 Kb	17.06 Kb
Scaffold N50	5.31 Mb	1.03 Mb	1.03 Mb
SuperScaffold N50	51.89 Mb	48.30 Mb	1.88 Mb
GC content	35.14%	35.19%	–
Complete BUSCOs	93.2%	92.1%	90%
Protein coding genes	31,665	37,998	26,732
Mean gene length	6,273 bp	6,199 bp	–
Mean coding sequence length	1,086 bp	1,007 bp	1,001
Mean number of exons per gene	5.18	4.45	4.74
Mean exon length	209 bp	226 bp	211 bp
Total repetitive sequence	68.26%	62.25%	61.24%
Total non-coding RNAs	2,488	2,865	3,201

assembly (Supplementary Figures 1, 2). The Male V2 genome was re-constructed using Hi-C technology. The genome size of the new version (Male V2) was 1.24 Gb, with 23 superscaffolds and the superscaffold N50 that was 48.30 Mb long, and the GC content was 35.19% (Table 1). Genomic alignments of chromosomes and superscaffolds between Female V1 and Male V2 are showed in (Figure 3C).

To assess the quality of the two newly assembled genomes, BUSCO was carried out. Among the 2,121 plant-specific orthologs, 2,024 (95.4%) and 2,026 (95.5%) BUSCOs were identified in the Female V1 and Male V2 assembly respectively. In addition, 1,975 (93.2%, Female V1) and 1,954 (92.1%, Male V2) of the identified orthologs in the two genomes were considered to be complete, which were higher than the 90.0% completed genes identified in the Male V1 genome (Supplementary Figure 3 and Supplementary Table 1). The BUSCO results suggested that both genomes were assembled with high quality.

Genome annotation

The Female V1 and Male V2 genome consisted of 68.26% (688.75 Mb) and 62.25% (770.88 Mb) repeats, respectively (Table 1 and Figure 4). In both genomes, LTR (long terminal repeat) was the most abundant type of repetitive element, accounting for 40.15% (405.08 Mb) and 36.6% (453.30 Mb) of the genome size, respectively. Additionally, 59.05 Mb (5.85% of the Female V1 genome) and 64.56 Mb (5.21% of the Male V2 genome) DNA transposons were identified, which were the second highest repeated sequence in both genomes (Supplementary Table 2). The

LTR-RT insertion time was estimated to provide a retrospective view of the expanded history of LTR-RTs in the two newly assembly *E. ulmoides* genomes. LTR-RTs gradually accumulated over 5 million years ago, and there was an insertion explosion approximately 2 million years ago in both genomes, with the Male V2 genome undergoing another peak at 0.15 Mya (Figure 3A).

A combination of ab initio prediction, RNA-seq, and homology-based search was used to predict the protein-coding genes in the *E. ulmoides* genome. Overall, 31,665 and 37,998 genes were predicted in the Female V1 and Male V2 genome, and the average protein-coding gene size was 6,273 bp and 6,199 bp respectively (Table 1). Among these predicted protein-coding genes, 24,049 (75.95%) genes from the Female V1 genome and 27,284 (71.8%) genes from the Male V2 genome were functionally annotated in the GO, KEGG, EggNOG, Pfam and UniProt databases (Supplementary Table 3). In addition, 2,488 and 2,910 noncoding RNAs were separately annotated in the two newly assembled genomes. The Female V1 genome contained 976 tRNAs, 178 rRNAs, 1,141 snRNAs and 193 miRNAs. Similarly, four types of noncoding RNAs were identified in the Male V2 genome, including 1,216 tRNAs, 214 rRNAs, 1,253 snRNAs and 182 miRNAs (Supplementary Figures 4, 5 and Supplementary Table 4).

Whole-genome duplication analysis

WGD analysis was performed by calculating synonymous substitutions per synonymous site (Ks) values in the two newly assembled genomes. The density distribution of Ks values exhibited two peaks both in Female V1 and Male V2, with Ks values of 0.45 and 2.29 (Figure 3B). After the early γ duplication event that affected
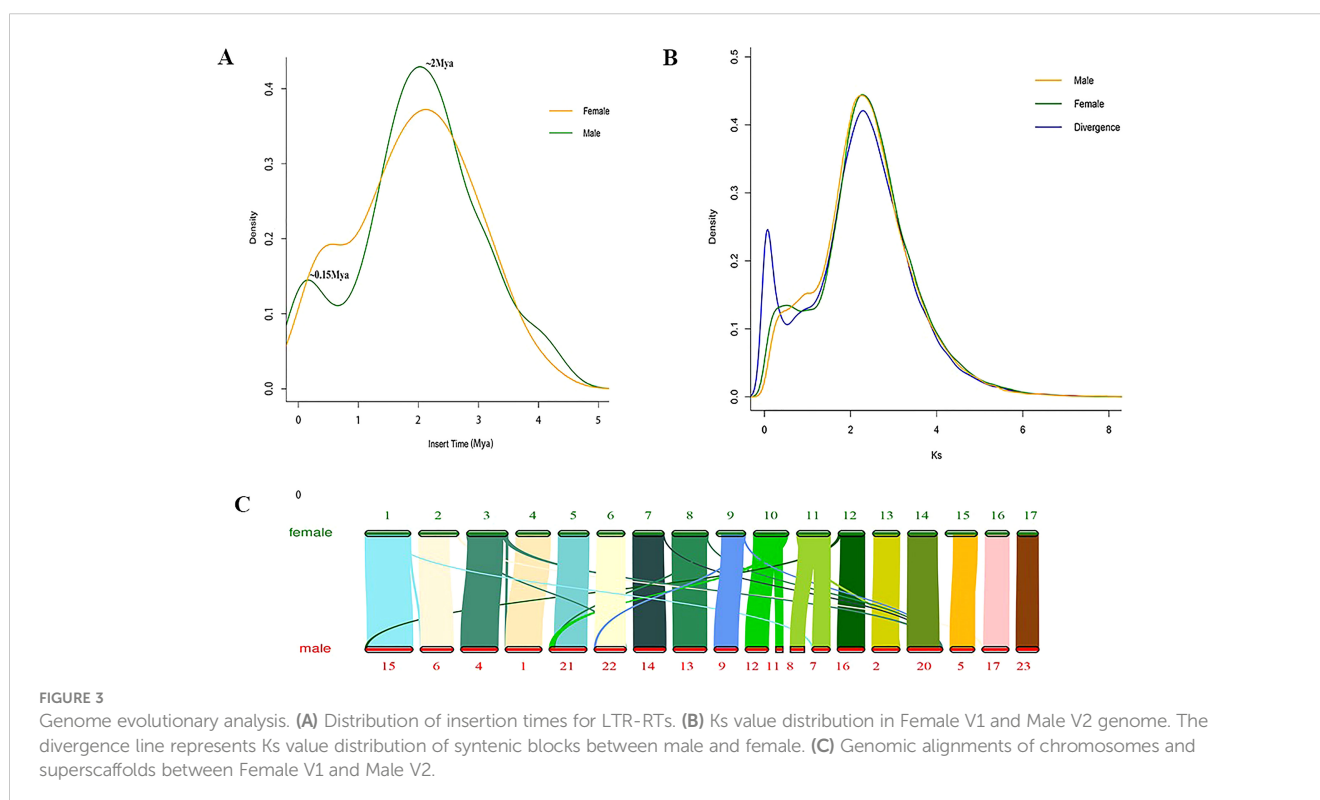

FIGURE 3
Genome evolutionary analysis. **(A)** Distribution of insertion times for LTR-RTs. **(B)** Ks value distribution in Female V1 and Male V2 genome. The divergence line represents Ks value distribution of syntenic blocks between male and female. **(C)** Genomic alignments of chromosomes and superscaffolds between Female V1 and Male V2.
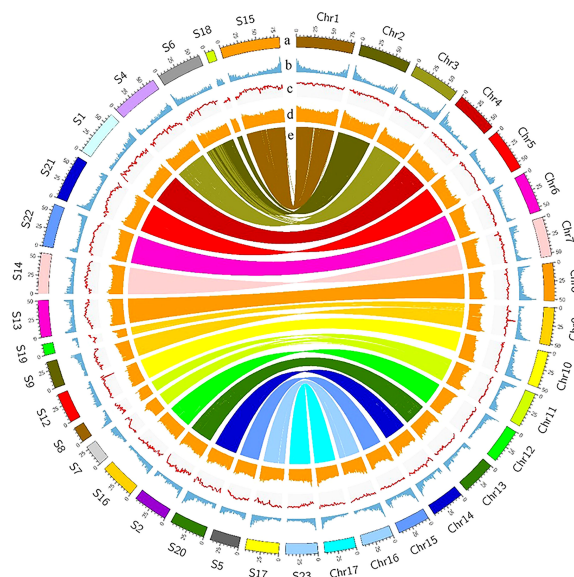
FIGURE 4
Landscape of Female V1 and Male V2 genome. The circle from outside to inside represents, **(A)** Chromosomes of Female V1 and Superscaffolds of Male V2, **(B)** gene density, **(C)** GC content, **(D)** repeat abundance, **(E)** synteny information. All distributions were drawn in a window size of 1 Mb.

eudicots, *E. ulmoides* experienced a more recent WGD event approximately 27.3 Mya. The collinearity analysis of male and female revealed a 1:1 synteny pattern (Supplementary Figure 6). The divergence time between female and male *E. ulmoides* was defined with a Ks peak value of 0.075 (Figure 3B).

MADS-box genes involved in sex differentiation

In the Female V1 and Male V2 genome, 76 and 81 MADS-box genes were separately identified, and further phylogenetic analysis of these genes was performed (Supplementary Figure 7). Based on the transcriptome analysis, 30 MADS-box genes of Female V1 were differentially expressed at different developmental stages of male and female flower buds. In addition, nine genes involved in the ABCDE model of floral development were identified in these MADS-box genes, which included three A-class genes (*EuAP1, EuFUL1, EuAGL6*), one B-class gene (*EuAP3*), one C-class gene (*EuAG*), two D-class genes (*EuSHP1, EuSHP2*), and two genes (*EuSEP1, EuSEP2*) that were associated with the E function (Supplementary Table 5). Nine MADS-box genes involved in the ABCDE model of floral development were identified in Male V2. We found high homology between male and female lineages (Supplementary Figure 8) and the gene structure and domain sequence of these genes were similar (Supplementary Figures 9, 10). In addition, we compared the obtained sex differentiation genes with the related sex differentiation genes in *Nymphaea colorata*, and the genes involved in sex differentiation in *E. ulmoides* and *N. colorata* showed high homology (Supplementary Figure 11).

The expression levels of the ABCDE genes at the floral organ maturity stage were higher than those in the floral organ induction stage and morphological differentiation initial stage. Except for the

EuFUL1 and *EuSHP2* genes, the other seven ABCDE genes were differentially expressed between the male and female flower buds. Seven ABCDE genes were differentially expressed during the floral organ maturity stage, with four differentially expressed genes in the floral organ morphological differentiation initial stage, and only one gene differentially expressed in the floral organ induction stage. The *EuAP3* gene was differentially expressed in three different
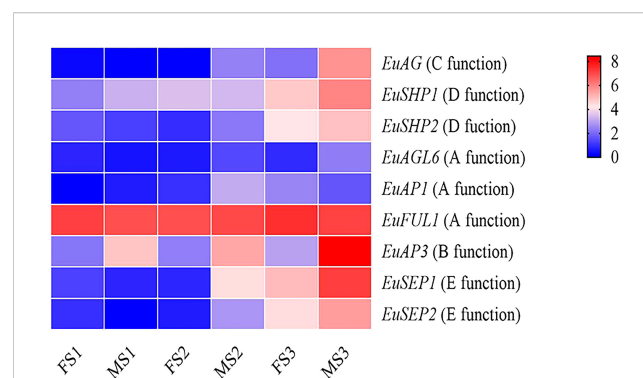
FIGURE 5
Heatmap of sex determination genes expression data in different stages of flower bud development in Female and male *E. ulmoides*. FS1: female floral organ induction stage flower bud, MS1: male floral organ induction stage flower bud, FS2: female floral organ morphological differentiation initial stage flower bud, MS2: and male floral organ morphological differentiation initial stage flower bud, FS3: female floral organ maturity stage flower bud, MS3: male floral organ maturity stage flower bud. Various color blocks represent the normalized gene expression levels of candidate genes involved in sex determination at different stages of flower bud development in Female and male *E. ulmoides*. The six boxes in one row of each heatmap (left to right) correspond to the expression levels in FS1, MS1, FS2, MS2, FS3, and MS3. Each row in the heatmap corresponds to one gene.

developmental stages of male and female flower buds (Figure 5 and Supplementary Figure 12).

Identification of the genes involved in α-linolenic acid synthesis

According to the functional annotation, 12 genes encoding key enzymes involved in the main synthesis and metabolic pathways of linolenic acid were identified in the Female V1 genome. Among them, six genes of four key enzymes related to the synthesis of α-linolenic acid were identified, including two acyl-ACP desaturase (FAB2) genes—*EU0119133* and *EU0120166*, which are homologues of SAD and perform the same function of converting stearoyl-ACP (C18:0-ACP) to oleoyl-ACP (C18:1-ACP) (Kazaz et al., 2020); one fatty acyl-ACP thioesterase A (FATA) gene—*EU0103200;* two omega-6 fatty acid desaturase (FAD2, FAD6) genes—*EU0105412* and *EU0128492;* and one omega-3 fatty acid desaturase (FAD7) gene—*EU0103017*. Six key enzyme genes for α-linolenic acid metabolism were also identified, including four lipoxygenase (LOX) genes—*EU0114412, EU0114414, EU0119724,* and *EU0119725*—and two alpha-dioxygenase (DOX) genes, *EU0107025, EU0131326* (Supplementary Table 6).

In addition, the expression levels of 12 identified key enzyme genes were compared between four parts of the fruit, stem, bark, and leaf. A heatmap of gene expression levels was also drawn (Figure 6). The expression levels of α-linolenic acid synthesis-related genes were higher in fruits and leaves than that in bark and stems. Overall, the expression levels of α-linolenic acid synthesis-related genes were much higher than those of metabolism-related genes. In particularly, the gene *EU0103017*—encoding FAD7, a key enzyme for linolenic acid synthesis—was highly expressed.

Discussion

E. ulmoides is an important economic plant because its bark, leaf, flower, fruit, and other tissues are rich in active secondary metabolites with important medicinal and economic value (Wang et al., 2019). High-quality *E. ulmoides* genomes have been useful for exploring the mechanisms of sex differentiation, synthesis and metabolism of important compounds. We assembled high-quality female and male genomes using PacBio and Hi-C technologies. There were 31,665 protein coding genes identified in Female V1 as compared to the 37,998 predicted genes in Male V2, and the repetitive sequence size in Female V1 was higher than that in Male V2. These results indicated that some differences in the genomes of male and female *E. ulmoides* were produced during the evolutionary process. Compared to previously published genomes (Wuyun et al., 2018) (Male V1), the quality of the reassembled version in this study (Male V2) was significantly increased. For example, superscaffold N50 was 25 times higher in Male V2 as compared to Male V1, and there were 1,266 new protein coding genes in addition to the 26,732 identified in Male V1. All these results indicated that there was greater accuracy and completeness in the Male V2 genome
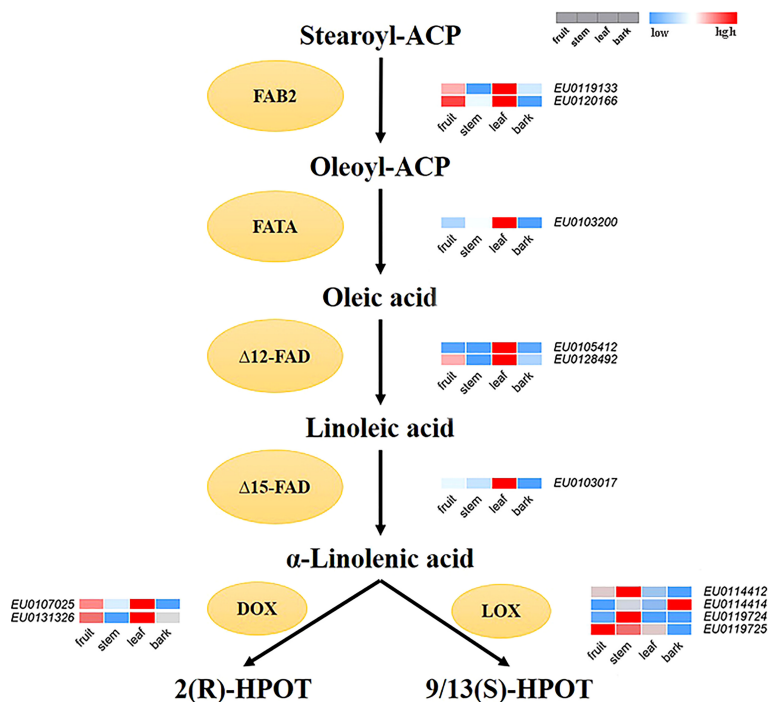
FIGURE 6

The reconstructed pathway of α-linolenic acid biosynthesis and metabolism in *E. ulmoides*. FAB2, Aacyl-ACP desaturase; FATA, acyl-ACP thioesterase; Δ12-FAD, omega-6 fatty acid desaturase; Δ15-FAD, omega-3 fatty acid desaturase; LOX, lipoxygenase; DOX, alpha-dioxygenase. Various color blocks represent the normalized gene expression levels of candidate genes related to α-linolenic acid biosynthesis and metabolism in *E. ulmoides*. The four boxes in one row of each heatmap (left to right) correspond to the expression levels in fruit, stem, leaf, and bark. Each row in the heatmap corresponds to one gene.

assembly as compared to the Male V1 genome. These two newly genomes provide a valuable resource for understanding sex differentiation and synthesis of important compounds in *E. ulmoides*, and they can be widely applied to genetics and genomics studies, such as the development of markers related to sex identification, detection of genes that synthesize important substances, and studies on the evolutionary history of *E. ulmoides*.

TEs is an important driving force to promote genome evolution, and the most important driving force is the accumulation of LTR-RTs (Sahebi et al., 2018; Lanciano and Cristofari, 2020). Further analysis showed that LTR-RTs erupted approximately 2 Mya, and cold tolerance may explain why *E. ulmoides* survived the Quaternary glaciation (Cros et al., 2020) approximately 2 Mya. Another insertion peak was identified in the Male V2 genome, this may be related to a greater number of repetitive sequences in the male genome. Whole genome duplication (WGD), also known as a paleopolyploidy event, provides abundant evolutionary materials for species and is considered as an important driving force for the formation and evolution of species (Soltis and Soltis, 2016; Clark and Donoghue, 2018). Most plants experience WGD, and many WGD events in angiosperms were detected during the Cretaceous-Tertiary period (145.5–65.5 Mya) (Vanneste et al., 2014). There is an obvious peak in the Ks distribution at the Ks value of 2.29375 in the two genomes, which indicated that males and females underwent a WGD event approximately 139 Mya (γ-WGT, 122–164 Mya) (Salse, 2016). Plants undergoing WGD events can enhance their environmental adaptability and stress resistance, thus promoting species evolution (del Pozo and Ramirez-Parra, 2015). Therefore, γ-WGT event may increase the adaptability of *E. ulmoides* to harsh environments, thus enabling *E. ulmoides* to resist the extinction event that occurred in the Cretaceous-Tertiary period. The peak at 0.45 in male genome is not as obvious as that in female, which might be the comparatively smaller continuity (Contig N50 17.06 kb) and lower quality of male genome than that of female genome. To detect whether there is a WGD event here, we have performed collinearity analysis. The 1:1 synteny pattern between male and female genomes also revealed that they have undergone the same number of WGD events. Besides, the divergence time between female and male *E. ulmoides* was identified with a Ks peak value of 0.075, while the recent WGD, specifically occurred in female, was with the Ks peak value of 0.45, which indicated that the recent WGD occurred before their differentiation.

There were nine MADS-box genes involved in the ABCDE mode in both female and male *E. ulmoides*, and their gene structure and domain sequence were conservative, suggesting their important role in floral development. During the development of male and female *E. ulmoides* flower buds, the greatest number of MADS-box differentially expressed genes was observed in the floral organ maturity stage, followed by the floral organ morphological differentiation initial stage, and the same was true for ABCDE genes. These results showed that the male and female flower buds of *E. ulmoides* were identical in the early developmental stage. However, with the development of female and male flower buds, the morphological differences gradually became obvious, and the differential genes increased. The ABCDE model of floral organ has been studied in *Nymphaea colorata* (Zhang et al., 2020), and the genes involved in floral organ in *E. ulmoides* and *N. colorata* showed high homology,

which indicated that the genes related to flower development identified in *E. ulmoides* were reliable. In the ABCDE flower development model, B, C, and E genes jointly control stamen development (Hu et al., 2021). During the floral organ maturity stage of *E. ulmoides*, a large number of MADS-box family genes were differentially expressed in male and female flower buds. In particular, we observed that the B-class gene *EuAP3* was differentially expressed during the three stages of male and female flower bud development, and the expression level in male flower buds was significantly higher than that in female flower buds, which further supports the role of *EuAP3* in regulating stamen development. The expression of the C-class gene (*EuAG*) was also upregulated in male flower buds. This indicated that similar to the AG gene function in *Arabidopsis* and rice, the *EuAG* gene was involved in the development of stamens. *EuAP3* and *EuAG* are the key regulatory genes for stamen development and may be candidate genes for regulating sex differentiation.

E. ulmoides is rich in α-linolenic acid, which is synthesized from stearic acid under the catalysis of a series of fatty acid desaturases (Liu et al., 2020). First, stearoyl-ACP (C18:0-ACP) is catalyzed by enzyme FAB2 to produce oleoyl-ACP (C18:1-ACP), which is then converted to oleic acid (OA, C18:1) by the FATA enzyme. Oleic acid is desaturated by the enzyme fatty acid desaturase 2 (FAD2) and fatty acid desaturase 6 (FAD6) to generate linoleic acid (LA, C18:2). The fatty acid desaturase 7 (FAD7) enzyme catalyzes the synthesis of α-linolenic acid (ALA, C18:3) from LA in *E. ulmoides*. DOX and LOX are two key enzymes that catalyze the first step of α-linolenic acid metabolism, and they catalyze α-linolenic acid to produce 2 (R)-HPOT and 9/13(S)-HPOT, respectively (Gfeller et al., 2010; Shimada and Hara-Nishimura, 2015).

Transcriptome analysis results showed that the expression of synthesis-related genes was high, and there was very high expression of the genes encoding the key enzyme ω-3 fatty acid desaturase (FAD7), which is involved in α-linolenic acid synthesis and allows *Eucommia* to synthesize sufficient α-linolenic acid. The expression of metabolism-related genes was lower or they were not expressed, and thus, α-linolenic acid could be efficiently accumulated in *E. ulmoides*. These results may explain why the α-linolenic acid content in *E. ulmoides* is higher than that in other plants. In addition, the expression level of genes related to α-linolenic acid synthesis in fruit and leaf was higher than that in bark and stem. Although the expression of genes related to α-linolenic acid synthesis was higher in leaf (*EU0119133, EU0120166, EU0103017*), the expression of genes related to the downstream α-linolenic acid metabolic pathway was also higher in leaf (*EU0107025, EU0131326*). This allows the synthesis of α-linolenic acid in the leaf to be greatly metabolized, while the α-linolenic acid in the fruit can be accumulated, which explains why the α-linolenic acid is mainly distributed in the fruit of *E. ulmoides*.

In the current study, two high-quality *E. ulmoides* genomes are provided, which are helpful for evolutionary and genomics research. Using these genomic data, we identified WGD events and evaluated the LTR outbreak time to provide a basis for future studies on genomic evolution and stress resistance of *E. ulmoides*. MADS-box family genes were also identified, and the genes related to flower development in *E. ulmoides* were predicted according to the expression patterns of genes in male and female flower buds, and finally the key candidate genes that regulate sex differentiation were identified. In addition, the key

regulatory genes for α-linolenic acid synthesis and metabolism were identified, and the mechanism of α-linolenic acid accumulation was revealed. According to these genes, *E. ulmoides* can be improved by directed genetic modification. Overall, our research results not only provide valuable information for the evolutionary analysis of *E. ulmoides* but also advance our understanding of the molecular mechanism of sex differentiation and α-linolenic acid accumulation. This study will accelerate the breeding of superior varieties and enable us to more thoroughly explore and use the value inherent in *E. ulmoides*.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://www.ncbi.nlm.nih.gov/, PRJNA792509.

Author contributions

HD and LW conceived this study; QD, ZW, and PL worked on the data analysis and wrote the manuscript; JQ, FH, LD, ZhS LZ, and HZ contributed to the sample preparation; ZoS provided sequencing service; LY and LW revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

Acknowledgments

Conflict of interest

Author HZ is employed by Henan Jinduzhong Agricultural Science and Technology Co., Ltd., Henan, China. Author ZoS is employed by Grandomics Biosciences Co., Ltd., Wuhan, China.

The remaining 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.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpls.2023.1118363/full#supplementary-material

References

Abraham-Juárez, M. J., Schrager-Lavelle, A., Man, J., Whipple, C., Handakumbura, P., Babbitt, C., et al. (2020). Evolutionary variation in MADS box dimerization affects floral development and protein abundance in maize. *Plant Cell* 32 (11), 3408–3424. doi: 10.1105/tpc.20.00300

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* 6, 11. doi: 10.1186/s13100-015-0041-9

Barrett, S. C. H., and Hough, J. (2012). Sexual dimorphism in flowering plants. *J. Exp. Bot.* 64 (1), 67–82. doi: 10.1093/jxb/ers308

Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* 27 (2), 573–580. doi: 10.1093/nar/27.2.573

Birney, E., Clamp, M., and Durbin, R. (2004). GeneWise and genomewise. *Genome Res.* 14 (5), 988–995. doi: 10.1101/gr.1865504

Bowman, J. L., Smyth, D. R., and Meyerowitz, E. M. (1989). Genes directing flower development in arabidopsis. *Plant Cell* 1 (1), 37–52. doi: 10.1105/tpc.1.1.37

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinf.* 10, 421. doi: 10.1186/1471-2105-10-421

Causier, B., Schwarz-Sommer, Z., and Davies, B. (2010). Floral organ identity: 20 years of ABCs. *Semin. Cell Dev. Biol.* 21 (1), 73–79. doi: 10.1016/j.semcdb.2009.10.005

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., et al. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat. Methods* 10 (6), 563–569. doi: 10.1038/nmeth.2474

Clark, J. W., and Donoghue, P. C. J. (2018). Whole-genome duplication and plant macroevolution. *Trends Plant Sci.* 23 (10), 933–945. doi: 10.1016/j.tplants.2018.07.006

Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353 (6339), 31–37. doi: 10.1038/353031a0

Cros, E., Chattopadhyay, B., Garg, K. M., Ng, N. S. R., Tomassi, S., Benedick, S., et al. (2020). Quaternary land bridges have not been universal conduits of gene flow. *Mol. Ecol.* 29 (14), 2692–2706. doi: 10.1111/mec.15509

del Pozo, J. C., and Ramirez-Parra, E. (2015). Whole genome duplications in plants: an overview from arabidopsis. *J. Exp. Bot.* 66 (22), 6991–7003. doi: 10.1093/jxb/erv432

Di Marzo, M., Roig-Villanova, I., Zanchetti, E., Caselli, F., Gregis, V., Bardetti, P., et al. (2020). MADS-box and bHLH transcription factors coordinate transmitting tract development in arabidopsis thaliana. *Front. Plant Sci.* 11. doi: 10.3389/fpls.2020.00526

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., Robinson, J. T., Shamim, M. S., Machol, I., Mesirov, J. P., Lander, E. S., et al. (2016a). Juicebox provides a visualization system for Hi-c contact maps with unlimited zoom. *Cell Syst.* 3 (1), 99–101. doi: 10.1016/j.cels.2015.07.012

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., Huntley, M. H., Lander, E. S., et al. (2016b). 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

Feng, Y., Zhang, L., Fu, J., Li, F., Wang, L., Tan, X., et al. (2016). Characterization of glycolytic pathway genes using RNA-seq in developing kernels of eucommia ulmoides. *J. Agric. Food Chem.* 64 (18), 3712–3731. doi: 10.1021/acs.jafc.5b05918

Flynn, J. M., Hubley, R., Goubert, C., Rosen, J., Clark, A. G., Feschotte, C., et al. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* 117 (17), 9451–9457. doi: 10.1073/pnas.1921046117

Gfeller, A., Dubugnon, L., Liechti, R., and Farmer, E. E. (2010). Jasmonate biochemical pathway. *Sci. Signal* 3 (109), cm3. doi: 10.1126/scisignal.3109cm3

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

Hu, J., Chang, X., Zhang, Y., Yu, X., Qin, Y., Sun, Y., et al. (2021). The pineapple MADS-box gene family and the evolution of early monocot flower. *Sci. Rep.* 11 (1), 849. doi: 10.1038/s41598-020-79163-8

Hu, X. D., Pan, B. Z., Fu, Q., Niu, L., Chen, M. S., and Xu, Z. F. (2018). *De novo* transcriptome assembly of the eight major organs of sacha inchi (Plukenetia volubilis) and the identification of genes involved in α-linolenic acid metabolism. *BMC Genomics* 19 (1), 380. doi: 10.1186/s12864-018-4774-y

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. Evol.* 34 (8), 2115–2122. doi: 10.1093/molbev/msx148

Kazaz, S., Barthole, G., Domergue, F., Ettaki, H., To, A., Vasselon, D., et al. (2020). Differential activation of partially redundant Δ9 stearoyl-ACP desaturase genes is critical for omega-9 monounsaturated fatty acid biosynthesis during seed development in arabidopsis. *Plant Cell* 32 (11), 3613–3637. doi: 10.1105/tpc.20.00554

Kim, K. B., Nam, Y. A., Kim, H. S., Hayes, A. W., and Lee, B. M. (2014). α-linolenic acid: nutraceutical, pharmacological and toxicological evaluation. *Food Chem. Toxicol.* 70, 163–178. doi: 10.1016/j.fct.2014.05.009

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14 (4), R36. doi: 10.1186/gb-2013-14-4-r36

Kong, S., and Zhang, Y. (2019). Deciphering Hi-c: from 3D genome to function. *Cell Biol. Toxicol.* 35 (1), 15–32. doi: 10.1007/s10565-018-09456-2

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

Lanciano, S., and Cristofari, G. (2020). Measuring and interpreting transposable element expression. *Nat. Rev. Genet.* 21 (12), 721–736. doi: 10.1038/s41576-020-0251-y

Li, Y., Wei, H., Yang, J., Du, K., Li, J., Zhang, Y., et al. (2020). High-quality *de novo* assembly of the eucommia ulmoides haploid genome provides new insights into evolution and rubber biosynthesis. *Hortic. Res.* 7 (1), 183. doi: 10.1038/s41438-020-00406-w

Liao, B., Hao, Y., Lu, J., Bai, H., Guan, L., and Zhang, T. (2018). Transcriptomic analysis of perilla frutescens seed to insight into the biosynthesis and metabolic of unsaturated fatty acids. *BMC Genomics* 19 (1), 213. doi: 10.1186/s12864-018-4595-z

Li-Beisson, Y., Shorrosh, B., Beisson, F., Andersson, M. X., Arondel, V., Bates, P. D., et al. (2013). Acyl-lipid metabolism. *Arabidopsis Book* 11, e0161. doi: 10.1199/tab.0161

Liu, K., Zhao, S., Wang, S., Wang, H., and Zhang, Z. (2020). Identification and analysis of the FAD gene family in walnuts (Juglans regia l.) based on transcriptome data. *BMC Genomics* 21 (1), 299. doi: 10.1186/s12864-020-6692-z

Lowe, T. M., and Chan, P. P. (2016). tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 44 (W1), W54–W57. doi: 10.1093/nar/gkw413

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

Massonnet, M., Cochetel, N., Minio, A., Vondras, A. M., Lin, J., Muyle, A., et al. (2020). The genetic basis of sex determination in grapes. *Nat. Commun.* 11 (1), 2902. doi: 10.1038/s41467-020-16700-z

Morel, P., Chambrier, P., Boltz, V., Chamot, S., Rozier, F., Rodrigues Bento, S., et al. (2019). Divergent functional diversification patterns in the SEP/AGL6/AP1 MADS-box transcription factor superclade. *Plant Cell* 31 (12), 3033–3056. doi: 10.1105/tpc.19.00162

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., et al. (2015). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res.* 43 (Database issue), D130–D137. doi: 10.1093/nar/gku1063

Nawrocki, E. P., Kolbe, D. L., and Eddy, S. R. (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* 25 (10), 1335–1337. doi: 10.1093/bioinformatics/btp157

Prakash, A., Jeffryes, M., Bateman, A., and Finn, R. D. (2017). The HMMER web server for protein sequence similarity search. *Curr. Protoc. Bioinf.* 60, 3.15.1–3.15.23. doi: 10.1002/cpbi.40

Sahebi, M., Hanafi, M. M., van Wijnen, A. J., Rice, D., Rafii, M. Y., Azizi, P., et al. (2018). Contribution of transposable elements in the plant's genome. *Gene* 665, 155–166. doi: 10.1016/j.gene.2018.04.050

Salse, J. (2016). Ancestors of modern plant crops. *Curr. Opin. Plant Biol.* 30, 134–142. doi: 10.1016/j.pbi.2016.02.005

Schwarz-Sommer, Z., Huijser, P., Nacken, W., Saedler, H., and Sommer, H. (1990). Genetic control of flower development by homeotic genes in antirrhinum majus. *Science* 250 (4983), 931–936. doi: 10.1126/science.250.4983.931

Shimada, T. L., and Hara-Nishimura, I. (2015). Leaf oil bodies are subcellular factories producing antifungal oxylipins. *Curr. Opin. Plant Biol.* 25, 145–150. doi: 10.1016/j.pbi.2015.05.019

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31 (19), 3210–3212. doi: 10.1093/bioinformatics/btv351

Soltis, D. E., Chanderbali, A. S., Kim, S., Buzgo, M., and Soltis, P. S. (2007). The ABC model and its applicability to basal angiosperms. *Ann. Bot.* 100 (2), 155–163. doi: 10.1093/aob/mcm117

Soltis, P. S., and Soltis, D. E. (2016). Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* 30, 159–165. doi: 10.1016/j.pbi.2016.03.015

Stanke, M., and Morgenstern, B. (2005). AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* 33 (Web Server issue), W465–W467. doi: 10.1093/nar/gki458

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

Teo, Z. W. N., Zhou, W., and Shen, L. (2019). Dissecting the function of MADS-box transcription factors in orchid reproductive development. *Front. Plant Sci.* 10. doi: 10.3389/fpls.2019.01474

Theissen, G. (2001). Development of floral organ identity: stories from the MADS house. *Curr. Opin. Plant Biol.* 4 (1), 75–85. doi: 10.1016/s1369-5266(00)00139-4

Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M. J., et al. (2010). Transcript assembly and quantification by RNA-seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* 28 (5), 511–515. doi: 10.1038/nbt.1621

Vanneste, K., Baele, G., Maere, S., and Van de Peer, Y. (2014). Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-paleogene boundary. *Genome Res.* 24 (8), 1334–1347. doi: 10.1101/gr.168997.113

Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., et al. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS One* 9 (11), e112963. doi: 10.1371/journal.pone.0112963

Wang, Y., Tang, H., Debarry, J. D., Tan, X., Li, J., Wang, X., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* 40 (7), e49. doi: 10.1093/nar/gkr1293

Wang, C. Y., Tang, L., He, J. W., Li, J., and Wang, Y. Z. (2019). Ethnobotany, phytochemistry and pharmacological properties of eucommia ulmoides: A review. *Am. J. Chin. Med.* 47 (2), 259–300. doi: 10.1142/s0192415x19500137

Wang, W., Yang, G., Deng, X., Shao, F., Li, Y., Guo, W., et al. (2020). Molecular sex identification in the hardy rubber tree (Eucommia ulmoides Oliver) *via* ddRAD markers. *Int. J. Genomics* 2020, 2420976. doi: 10.1155/2020/2420976

Wang, W., and Zhang, X. (2017). Identification of the sex-biased gene expression and putative sex-associated genes in eucommia ulmoides Oliver using comparative transcriptome analyses. *Molecules* 22 (12), 2255. doi: 10.3390/molecules22122255

Wang, D., Zhang, Y., Zhang, Z., Zhu, J., and Yu, J. (2010). KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinf.* 8 (1), 77–80. doi: 10.1016/s1672-0229(10)60008-3

Wuyun, T. N., Wang, L., Liu, H., Wang, X., Zhang, L., Bennetzen, J. L., et al. (2018). The hardy rubber tree genome provides insights into the evolution of polyisoprene biosynthesis. *Mol. Plant* 11 (3), 429–442. doi: 10.1016/j.molp.2017.11.014

Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* 35 (Web Server issue), W265–W268. doi: 10.1093/nar/gkm286

Yin, X., Liu, X., Xu, B., Lu, P., Dong, T., Yang, D., et al. (2019). OsMADS18, a membrane-bound MADS-box transcription factor, modulates plant architecture and the abscisic acid response in rice. *J. Exp. Bot.* 70 (15), 3895–3909. doi: 10.1093/jxb/erz198

Yu, S. Y., Zhang, X., Huang, L. B., Lyu, Y. P., Zhang, Y., Yao, Z. J., et al. (2021). Transcriptomic analysis of α-linolenic acid content and biosynthesis in paeonia ostii fruits and seeds. *BMC Genomics* 22 (1), 297. doi: 10.1186/s12864-021-07594-2

Zhang, L., Chen, F., Zhang, X., Li, Z., Zhao, Y., Lohaus, R., et al. (2020). The water lily genome and the early evolution of flowering plants. *Nature* 577 (7788), 79–84. doi: 10.1038/s41586-019-1852-5

Zhang, Z. S., Liu, Y. L., and Che, L. M. (2018). Characterization of a new α-linolenic acid-rich oil: Eucommia ulmoides seed oil. *J. Food Sci.* 83 (3), 617–623. doi: 10.1111/1750-3841.14049

Zhu, M. Q., and Sun, R. C. (2018). Eucommia ulmoides Oliver: A potential feedstock for bioactive products. *J. Agric. Food Chem.* 66 (22), 5433–5438. doi: 10.1021/acs.jafc.8b01312