

Article

The Chromosome-Level Genome of *Elaeagnus moorcroftii* Wall., an Economically and Ecologically Important Tree Species in Drylands

Xinxing Fu ¹, Jingjing Wu ², Xiaohui Ma ², Kunpeng Li ^{3,4}, Hui Zhang ¹, Shengdan Wu ^{2,*}  and Kun Sun ^{1,*}

¹ College of Life Sciences, Northwest Normal University, Lanzhou 730070, China; fuxinxing@nwnu.edu.cn (X.F.); zhanghui@nwnu.edu.cn (H.Z.)

² State Key Laboratory of Grassland Agro-Ecosystems, College of Ecology, Lanzhou University, Lanzhou 730000, China; wujj21@lzu.edu.cn (J.W.); maxh20@lzu.edu.cn (X.M.)

³ State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China; likunpeng@ibcas.ac.cn

⁴ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wusd@lzu.edu.cn (S.W.); kunsun@nwnu.edu.cn (K.S.)

Abstract: *Elaeagnus moorcroftii* Wall. (Elaeagnaceae) is an important tree species naturally growing in arid Northwest China that has great economic and ecological values in drylands. In this study, we de novo assembled a chromosome-level genome for *E. moorcroftii* by using PacBio's high-fidelity (HiFi) sequencing and Hi-C-assisted assembly technology. The assembled genome size was 529.56 Mb, of which 94.56% was anchored to 14 pseudochromosomes with a contig N50 up to 28.21 Mb. In total, 29,243 protein-coding genes were annotated, and 98.5% of the Benchmarking Universal Single-Copy Orthologs (BUSCOs) were captured in the genome. Evolutionary genomic analysis showed that *E. moorcroftii* split with *Elaeagnus mollis* 9.38 million years ago (Ma), and contrasted evolutionary trajectories of gene family expansion and contraction were observed for these two closely related species. Furthermore, we identified two successive whole genome duplication (WGD) events occurred in the genome of *E. moorcroftii*, in addition to the ancient *gamma* hexaploidization event shared by core eudicots. Together, the chromosome-level genome assembly for *E. moorcroftii* decoded here provides valuable genomic information for the further genetic improvement and molecular breeding of this indigenous species in drylands.

Keywords: *Elaeagnus moorcroftii* Wall.; PacBio's high-fidelity sequencing; Hi-C-assisted assembly; whole genome duplication; xerophyte; drylands



Citation: Fu, X.; Wu, J.; Ma, X.; Li, K.; Zhang, H.; Wu, S.; Sun, K. The Chromosome-Level Genome of *Elaeagnus moorcroftii* Wall., an Economically and Ecologically Important Tree Species in Drylands. *Diversity* **2022**, *14*, 468. <https://doi.org/10.3390/d14060468>

Academic Editors: Hong-Hu Meng and Yi-Gang Song

Received: 30 May 2022

Accepted: 10 June 2022

Published: 11 June 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Land degradation and desertification constitute one of the most serious environmental problems facing the world. Drylands cover about 41% of the global land area and are home to more than 38% of the world's population [1,2]. Ecosystems in drylands are fragile and vulnerable to climate change and human activities [3]. The degree of desertification in these drylands is likely to increase rapidly, the areas of drylands will continue to expand, and the risk of ecological degradation will be further exacerbated [2]. Desert plants play an important role in maintaining the stability of dryland ecosystems and provide ecological services for the production and life of people in drylands. Therefore, it is very meaningful to conduct scientific research on desert plants. With the rapid development of sequencing technologies, the genomes of some important desert plants have been deciphered recently, such as the sea buckthorn (*Hippophae rhamnoides*) with medicinal and edible value [4,5], wild and perennial legume forage *Medicago ruthenica* [6], and xerophytic plant *Haloxylon ammodendron* [7].

E. moorcroftii is a kind of deciduous tree of the Elaeagnaceae family, up to 10 m tall, mainly distributed in the desert areas of Northwest China, including the Xinjiang, Gansu,

and Inner Mongolia provinces. It has excellent characteristics of drought resistance and salt tolerance and is an important tree species for windbreaks, sand fixation, and soil and water conservation in Northwest China [8]. This species is a non-leguminous nitrogen-fixing plant, its roots are symbiotic with *Frankia* actinomycetes, which can play the role of biological nitrogen fixation and soil improvement [4,9], and its rhizosphere arbuscular mycorrhizal fungi can improve their resistance to salt stress [10]. Therefore, *E. moorcroftii* can be introduced to barren desert and saline-alkali land for soil improvement and afforestation, which has important ecological value. In addition, this species has high economic value because of its edible fruits, medicinal whole plants, and ornamental flowers. The species exhibits a high fruit yield; the fruit can be eaten directly and used in jam, vinegar, wine, pastries, livestock feed, etc. [11]. The branches, leaves, and flowers of the species have remarkable observed biological activities and are widely used to treat many health issues like aging, burns, dyspepsia, diarrheal, pain, bronchitis, and neurasthenia [11,12]. The flowers are attractive and aromatic and are also used for extracting aromatic oil [11]. However, the molecular-level study of *E. moorcroftii* is limited to taxonomic relationships [8]. The lack of genomic information hinders a comprehensive understanding of the evolutionary history and unique biology of *E. moorcroftii*.

In this study, we first report a chromosomal-level genome assembly of *E. moorcroftii* ($2n = 2x = 28$) with PacBio's long-read single molecule high-fidelity (HiFi) reads and high-throughput chromosome conformation capture (Hi-C) data, and then we used it to explore the evolutionary trajectories of *E. moorcroftii* and other Elaeagnaceae species, including *H. rhamnoides* [4,5] and *E. mollis* [13] as recently published, by comparative genomic analysis. The genome sequence of *E. moorcroftii* presented here will provide valuable genomic resources for further in-depth study and utilization of this indigenous species in drylands.

2. Materials and Methods

2.1. Plant Materials and Sequencing

The fresh young leaves used for genomic DNA sequencing were collected from an adult plant of *E. moorcroftii* growing in Minqin Desert Botanical Garden, Gansu Province, China. The total genomic DNA was extracted using a modified trimethylammonium bromide (CTAB) method [14]. For PacBio's HiFi sequencing, SMRTbell libraries were constructed using the protocol of Pacific Biosciences with 20 kb inserts and sequenced using circular consensus sequencing (CCS) mode on a PacBio Sequel II platform, generating a total of more than 28 Gb high-quality CCS reads. For Hi-C sequencing, cross-linked chromatin was first digested with *Dpn* II, end-labeled with biotin-14-dATP, and then used for in situ DNA ligation. The ligated DNA was sheared into 300–600 bp fragments, blunt-end repaired, purified through biotin–streptavidin-mediated pull-down, and then sequenced on the Illumina HiSeq 2500 platform, generating a total of more than 52 Gb raw sequencing data.

2.2. Genome Assembly and Assessment

The high-quality HiFi reads were assembled into contigs using hifiasm v0.14 [15] with default parameters. The Hi-C data were aligned to the contig assembly using JUICER v.1.5 [16]. The contigs of the *E. moorcroftii* assembly were further clustered, ordered, and oriented onto chromosomes based on the contact frequency calculated from the mapped Hi-C read pairs by 3d-DNA pipeline [17] with parameters of ‘-m haploid -r 2’. Orientation and placement errors were manually corrected via Juicebox Assembly Tools (<https://github.com/aidenlab/Juicebox> (accessed on 17 November 2021)).

The completeness of the genome assembly of *E. moorcroftii* was assessed by transcript alignment and Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis [18]. RNA-seq reads were assembled and mapped to the genome by HISAT2 v2.1.0 [19]. BUSCO analysis of the final assembly and annotation was performed using BUSCO v4 [18] with the Embryophyta odb10 database to evaluate the completeness of the reference genome of *E. moorcroftii*.

2.3. Genome Annotation

Tandem repeats in the *E. moorcroftii* genome were annotated using GMATA v2.2 [20] and Tandem Repeats Finder v4.09 [21]. The transposable elements (TEs) in the genome were predicted using a combination of homology-based and de novo approaches. In the homology-based approach, we used RepeatMasker v4.1.0 [22] with the known repeat database Repbase v21.01 [23] to search for the TEs in the *E. moorcroftii* genome. For the de novo approach, we used MITE-Hunter [24] and RepeatModeler v2.0 [25] to construct a de novo repeat sequence database for *E. moorcroftii* and then used RepeatMasker v4.1.0 [22] to search for repeats in the genome. After removing overlapping repeats, the repeats identified by different methods were combined into the final repeat annotation.

Protein-coding genes were predicted based on the repeat masked genome using three approaches, including homology search, de novo prediction, and estimation from transcriptome evidence. Homology-based prediction was performed using GeMoMa v1.3.1 [26] with the protein sequences of four closely related species in eudicots, *E. mollis* [13], *Ziziphus jujuba* [27], *Arabidopsis thaliana* [28], and *Vitis vinifera* [29]. The de novo prediction was conducted using Augustus v3.2.3 [30] with default parameters. RNA-seq based gene prediction was performed using STAR v2.7.3a [31], Stringtie v2.0.1 [32], and PASA v2.0.2 [33], and the public available RNA-seq data for mixed tissue samples including the leaf, root, stem, and fruit were downloaded from the National Center for Biotechnology Information (NCBI) under the accession numbers of SRR12569922, SRR12569923, and SRR12569924. Finally, the outputs from the above three approaches were integrated into a final gene set by EvidenceModeler (EVM) v1.1.1 [34].

For gene function annotation, the predicted gene models were blasted against the SwissProt v5.3 (<https://www.expasy.org/resources/uniprotkb-swiss-prot> (accessed on 29 July 2021)), NCBI non-redundant (NR) (<https://www.ncbi.nlm.nih.gov/refseq/about/nonredundantproteins/> (accessed on 29 July 2021)), and Clusters of Orthologous Groups for Eukaryotic complete Genomes (KOG) v5.3 (<http://genome.jgi-psf.org/help/kogbrowser.jsf> (accessed on 29 July 2021)) databases for the best matches using BLASTP with an *E*-value cut-off of 1e-5. The protein motifs and domains were annotated using InterProScan v5.31 [35]. The Gene Ontology (GO) entries were searched using Blast2GO v2.5 [36]. Pathway information for each gene was assigned using the Kyoto Encyclopedia of Genes and Genomes (KEGG) v5.24 database [37].

In addition, the non-coding RNA genes were annotated. tRNAs were predicted by tRNAscan-SE v1.3.1 [38] with eukaryote parameters. MicroRNA, small nuclear RNA, and small nucleolar RNA were predicted using INFERNAL v1.1.2 [39] based on the Rfam [40] and miRbase databases [41]. The rRNAs and their subunits were predicted using RNAmmer v1.2 [42].

2.4. Comparative and Evolutionary Genomic Analysis

The protein sequences of *E. moorcroftii* and ten other sequenced plant species, *H. rhamnoides*, *E. mollis*, *Malus domestica*, *Prunus persica*, *Z. jujuba*, *Morus notabilis*, *Cannabis sativa*, *Populus trichocarpa*, *A. thaliana*, *V. vinifera*, were used for the phylogenetic analysis (see Table S1 for the summary of genomic information of these species). Single-copy orthogroups among the 11 species were identified using OrthoFinder v2.3.11 [43]. The amino acid alignments of each single-copy orthogroup were aligned by MAFFT v7 [44], and nucleotide alignments were generated according to the corresponding amino acid alignments using PAL2NAL [45]. A maximum likelihood phylogeny was constructed based on the concatenated alignments of all single-copy genes using IQ-TREE v1.6.12 [46]. The species divergence time was estimated using the program MCMCTREE in the PAML v4.9 package [47]. We selected three fossil calibration points from the TimeTree database (<http://www.timetree.org> (accessed on 1 March 2022)) for the split of: (1) *Malus-Prunus* at 30–61 million years ago (Ma); (2) *Morus-Cannabis* at 53–97 Ma; (3) *Populus-Arabidopsis* at 98–117 Ma. In addition, the time calibration of family Rhamnaceae (that is the split of *Ziziphus* with three Elaeagnaceae species) was set to >99 Ma based on an old fossil in the extant genus *Phyllica* of Rhamnaceae [48]. The gene

family expansion and contraction of 11 species were analyzed using CAFE v3.1 [49], and the expanded and contracted gene families in *E. moorcroftii* were subjected to GO enrichment to analyze their functions. The different modes of gene duplications were identified by DupGen_finder software [50].

2.5. Whole Genome Duplication Analysis

The protein-coding sequences of *E. moorcroftii*, *H. rhamnoides*, and *E. mollis* were self-aligned and aligned with each other using BLASTP with an *E*-value cut-off of $1e-5$. Syntenic blocks within a genome or between genomes were identified using MCScanX [51] with default parameters based on the above protein-sequence alignments. For each syntenic gene pair, the synonymous substitution rate (*K*_s) was calculated using the Nei-Gojobori method [52] by yn00 program of the PAML package [47]. The macro-syntenic relationships among the three species were visualized by the python version of MCScan software ([https://github.com/tanghaibao/jcvi/wiki/MCscan-\(Python-version\)](https://github.com/tanghaibao/jcvi/wiki/MCscan-(Python-version))) (accessed on 29 July 2021)).

3. Results

3.1. Genome Assembly of *E. Moorcroftii*

To obtain a high-quality reference genome of *E. moorcroftii*, we used a combination of HiFi long reads and Hi-C data to construct a chromosome-level assembly. We generated a total of 28 Gb high-quality HiFi long reads, with a maximum and average read length of 64.20 and 14.78 kb, respectively (Table S2 and Figure S1, distribution of PacBio's HiFi sequencing data). De novo assembly of these HiFi reads with hifiasm v0.14 [16] yielded a preliminary assembly comprising 168 contigs with contig N50 size of 28.21 Mb, and the total length of the assembled genome was 529.56 Mb (Table 1). A total of 52 Gb Hi-C reads were generated to refine the assembly, resulting in 500.73 Mb (94.56%) of the contig sequences anchored onto 14 chromosomes (Figure 1, Table 1).

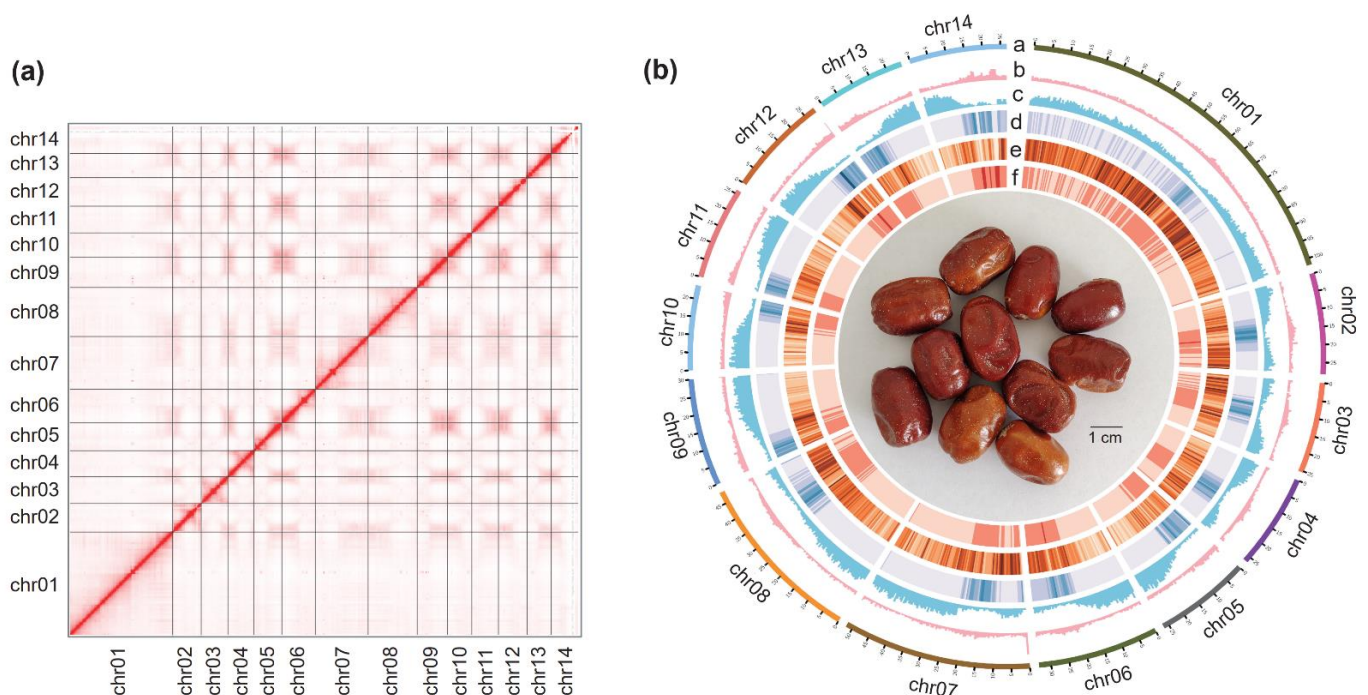


Figure 1. The genome features of *E. moorcroftii*. (a) Heatmap of Hi-C interactions for 14 pseudochromosomes. (b) Circos plot showing the genomic landscape of *E. moorcroftii*. The tracks from outer to inner circles indicate the following: a: 4 pseudochromosomes in megabases; b: GC content; c: gene density; d: density of *Gypsy* LTR retrotransposons; e: density of *Copia* LTR retrotransposons; f: LTR retrotransposons density. The center of the circos plot shows the fruits of *E. moorcroftii* (photo by X.F.).

Table 1. Assembly and annotation statistics of *E. moorcroftii* genome.

Assembly	
Length of genome assembly (Mb)	529.56
Anchored to chromosome (Mb)	500.73
Contig N50 (Mb)	28.21
Longest contig (Mb)	101.76
BUSCO score of assembly (%)	96.7%
Annotation	
GC content	30.39%
Percentage of repeat sequences (%)	60.95%
Number of protein-coding gene (%)	29243
Average gene length (bp)	4318.92
Average exon length (bp)	220.21
BUSCO score of annotation (%)	98.5%

To assess the quality and completeness of this assembled genome, we first mapped RNA-seq reads to the assembled genome, more than 93% of which were properly mapped (Table S3). Furthermore, BUSCO analysis for the genome assembly showed that 96.7% of the 1614 core plant genes was captured, including 95.8% complete BUSCOs and 0.9% fragmented BUSCOs (Table S4). These evidences together indicated that the *E. moorcroftii* assembly has high quality and completeness.

3.2. Annotation of the *E. Moorcroftii* Genome

We identified 322.78 Mb of non-redundant repetitive sequences in the *E. moorcroftii* genome, representing 60.95% of the genome assembly (Table S5). Long terminal repeat (LTR) retrotransposons were the most abundant type, accounting for 31.95% of the whole genome, of which *Copia* and *Gypsy* were the two most frequent LTR types, accounting for about 16.02% and 13.30% of the genome, respectively (Table S5).

Using genomic and transcriptomic data, we predicted 29,243 protein-coding gene models in the *E. moorcroftii* genome, with an average gene length of 4318.92 bp and an average exon length of 220.21 bp (Table 1). Among the predicted protein-coding genes, 95.10% were annotated through at least one of the following protein-related databases: the NCBI NR database (94.75%), the SwissProt protein database (81.38%), the KOG database (91.24%), the InterProScan database (86.10%), the GO database (64.91%), and the KEGG database (3.87%) (Table S6). In total, 95.10% of the protein-coding genes were functionally annotated by various databases. In addition, our annotation captured 98.5% of BUSCOs, including 97.4% complete gene models plus 1.1% fragmented gene models (Table S7).

We also predicted 183 micro RNAs (miRNAs), 2461 small-nuclear RNAs (snRNAs), 1636 transfer RNAs (tRNAs), and 20,568 ribosomal RNAs (rRNAs), with calculated average lengths of 124.86, 107.52, 75.21, and 238.47 bp, respectively (Table S8).

3.3. Evolutionary History of *E. Moorcroftii*

To investigate the evolutionary history of *E. moorcroftii*, we performed a gene family clustering using *E. moorcroftii* and ten other representative angiosperm species, including two related species of the Elaeagnaceae family (*E. mollis*, *H. rhamnoides*), five species of the same order Rosales (Rosaceae: *M. domestica*, *P. persica*, Rhamnaceae: *Z. jujuba*; Moraceae: *M. notabilis*; Cannabaceae: *C. sativa*), and three outgroup species of eudicot clade (*A. thaliana*, *P. trichocarpa*, and *V. vinifera*) (Table S1). We identified 175 single-copy orthogroups and used them to reconstruct the phylogenetic tree of *E. moorcroftii* and ten other plant species (Figure 2). The results showed that *E. moorcroftii* clustered a monophyletic group with its related species *E. mollis* and *H. rhamnoides*, which in turn formed sister to *Z. jujuba* of Rhamnaceae. The divergence time between *E. moorcroftii* and *E. mollis* was estimated to be around 9.38 million years ago (Ma), and the most recent common ancestor of two Elaeagnus species split with *H. rhamnoides* occurred at about 30.52 Ma (Figure 2).

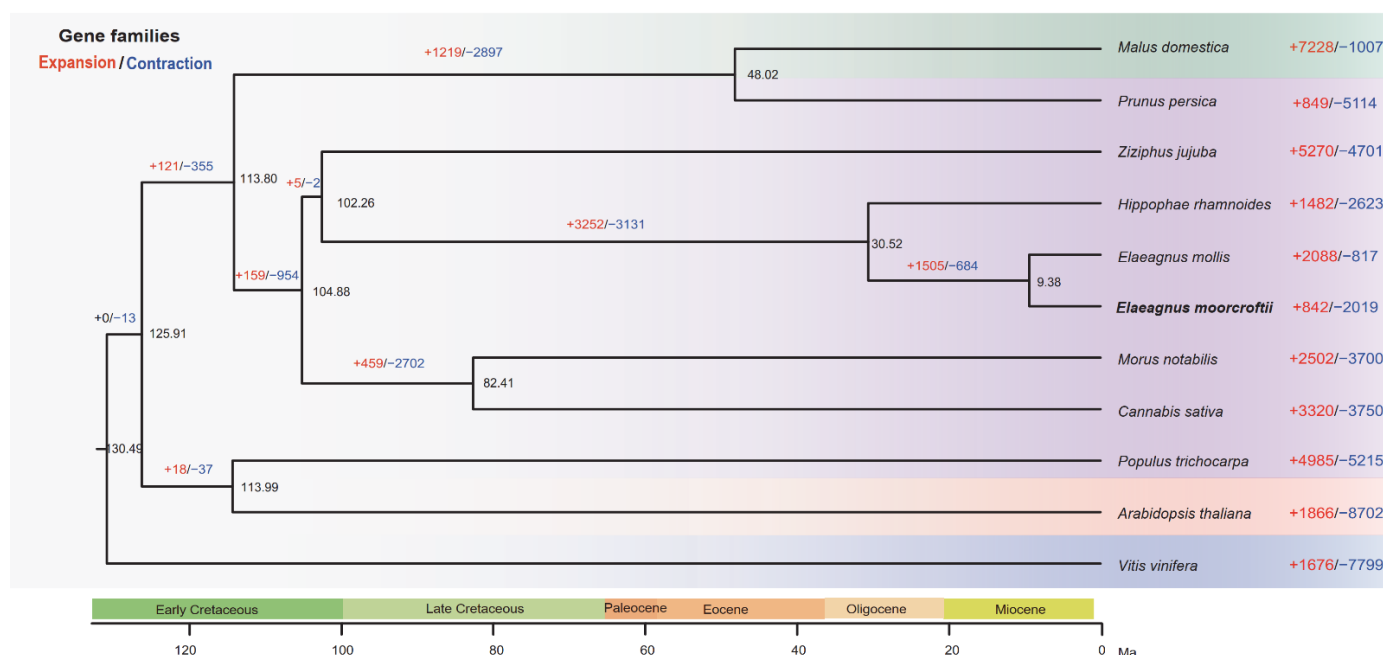


Figure 2. Comparative and evolutionary genomic analysis of *E. moorcroftii* and 10 other plant species. A phylogenetic tree among 11 species was reconstructed based on 175 single-copy genes, and their divergence times were also estimated. The numbers of expansion (blue) and contraction (red) gene families are shown above the branches.

Furthermore, to explore lineage-specific dynamic changes in gene families, the expansion and contraction of gene families based on the birth-and-death model were identified by CAFE v3.1 [49]. We detected 842 expansion and 2019 contraction gene families in *E. moorcroftii* genome, relative to the most recent common ancestor of *E. moorcroftii* and *E. mollis* (Figure 2), whereas an opposite trend was observed in its close relative *E. mollis*, which included more expanded (2088) than contracted (817) gene families. This difference might be partly resulted from the different levels of contribution from various modes of gene duplications. For example, dispersed duplications contributed to more in *E. mollis* than in *E. moorcroftii* with relation to the expanded gene families (16.45% versus 7.64%). In contrast, tandem duplications contributed to a higher percentage of expanded genes in *E. moorcroftii* than in *E. mollis* (14.34% versus 6.40%) (Figure S2). GO enrichment analysis revealed that the expanded gene families in *E. moorcroftii* were mainly involved in multiple biosynthetic processes (e.g., diterpenoid biosynthetic process and ATP biosynthetic process) and various metabolic process (e.g., ATP metabolic process, diterpenoid metabolic process, and purine-containing compound metabolic process), as well as immune responses (innate immune response and immune system process) (Figure S3), while contracted gene families were mainly related to protein depolymerization biological process and their enriched molecular function including ADP binding, ion channel activity, hormone binding, etc. (Figure S4).

A detailed comparative analysis among *E. moorcroftii*, *E. mollis*, *H. rhamnoides*, and *Z. jujuba* identified 12,118 common gene families shared by these four species, and 278 gene families uniquely appeared in *E. moorcroftii* (Figure S5). This number is comparable with the unique gene families in *E. mollis* (388) and *H. rhamnoides* (396), but it is much fewer than the unique gene families in *Z. jujuba* (2083) (Figure S5). The unique gene families are mainly involved in cellular respiration and oxidative phosphorylation biological processes (Figure S6).

3.4. Whole-Genome Duplication Events in *E. Moorcroftii*

The distributions of synonymous substitutions per synonymous site (K_S) of paralogous gene pairs in *E. moorcroftii* genome showed two recent clear peaks around 0.38 and

0.45 (Figure 3a). Similar K_S peaks also were identified in the genomes of its two closely related species, *E. mollis* and *H. rhamnoides* (Figure 3a). Moreover, the large-scale gene duplications in three species occurred earlier than the time of their divergence (Figure 3a). Therefore, these two closely occurring peaks might reflect two successive whole-genome duplication (WGD) events shared by three Elaeagnaceae species. To further confirm this, an intragenomic synteny analysis of *E. moorcroftii* identified one syntenic block corresponding to three homologous syntenic blocks (Figure S7), again supporting two relatively recent WGD events that occurred in *E. moorcroftii*. In addition, syntenic analysis among the genomes of *E. moorcroftii* and the most closely related species (*E. mollis* and *H. rhamnoides*) was performed to explore WGD history that occurred in the Elaeagnaceae family. By using a *Vitis vinifera*, a basal core eudicot lineage lacking any further WGD event after the ancient gamma event shared by core eudicots [53], genome as a reference, the intergenomic synteny comparisons between *V. vinifera* and *E. moorcroftii* revealed a clearly syntenic depth ratios of 1:4, indicating two WGD events occurred after the split of the two species (Figure 3b). Furthermore, the genomes of *E. moorcroftii* and *E. mollis* present highly conservative synteny, and consistent syntenic depth ratios were found for *E. mollis* and *H. rhamnoides* (Figure 3b). Thus, the integrated evidence showed that two successive WGD events might have occurred in the common ancestor of *E. moorcroftii*, *E. mollis*, and *H. rhamnoides*.

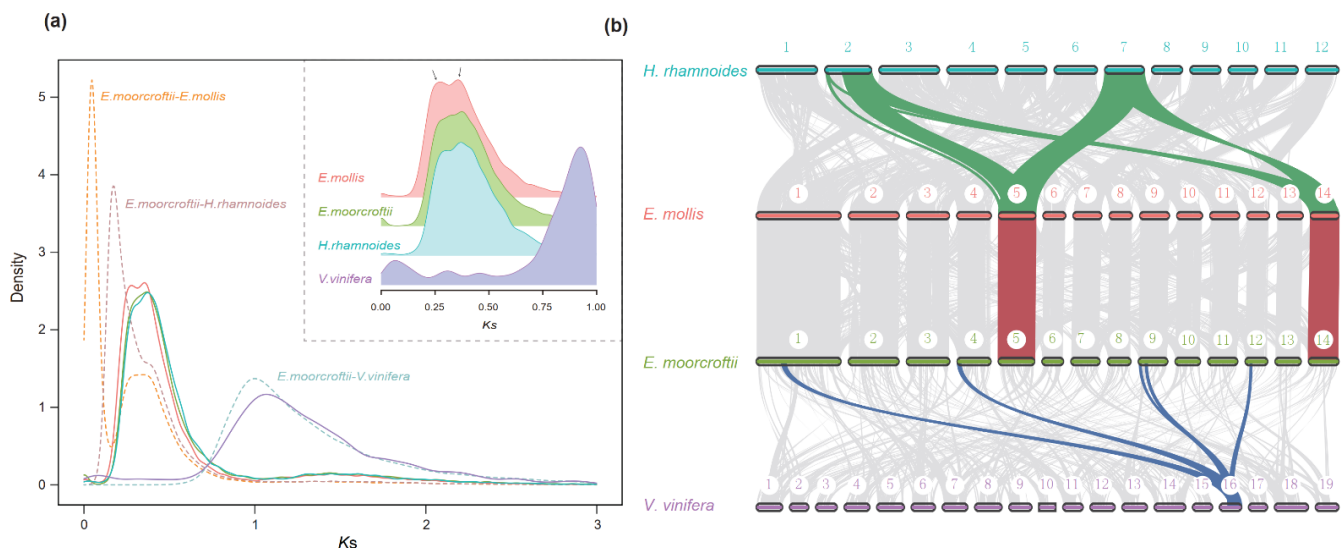


Figure 3. Integrating K_S and synteny analyses reveal two successive WGD events in *E. moorcroftii*. (a) The distributions of K_S of paralogous gene pairs of *E. mollis* (pink line), *E. moorcroftii* (green line), *H. rhamnoides* (blue line), and *V. vinifera* (purple line). The dashed lines represent K_S distributions related to the species divergence of the corresponding species pairs. (b) Macro-syntenic comparisons among three Elaeagnaceae species and *V. vinifera*.

4. Discussion

4.1. A High-Quality Dryland Tree Species Genome

Land degradation and desertification is an ongoing environmental problem that threatens the sustainable development of human beings, and this is particularly serious for lives in drylands [1–3]. A native or indigenous species naturally growing in drylands, to some extent, might hold the key to solve this problem. Recently, a growing number of studies have focused on decoding the genomic information of typical arid plants, for example, *Populus euphratica* [54], *H. rhamnoides* [4,5], *H. ammodendron* [7], etc. To our knowledge, besides the well-known desert poplar *P. euphratica*, our genome assembly for *E. moorcroftii* represents the second genome for high-tree species naturally growing in arid northwest China. Meanwhile, the chromosome-level genome of *E. moorcroftii* is of high quality, with contig N50 up to approximately 30 Mb (Table 1), which is much higher than

that of its two closest relatives (*H. rhamnoides* with N50 of 3.56 Mb in [4] and 2.15 Mb in [5], and *E. mollis* with N50 of 18.40 Mb) of Elaeagnaceae with published genomes.

4.2. Differential Evolutionary Dynamics of Gene Families between *E. Moorcroftii* and *E. Mollis*

Our phylogenomic and molecular dating analyses showed that the divergence time between *E. moorcroftii* and its close relative *E. mollis* was about 9.38 Ma. After the split from their most recent common ancestor (MRCA), these two species experienced independent gene family evolutionary dynamics, which can be observed by the contrasted expansion and contraction trend (Figure 2). The number of contracted gene families is more than twice the expanded gene families in *E. moorcroftii*, while this trend is opposite in *E. mollis*, which have more expanded than contracted gene families (Figure 2). A detailed analysis suggested that these differential evolutionary dynamics of gene families between *E. moorcroftii* and *E. mollis* might partly result from contribution levels of different modes of gene duplications (Figure S2). Nevertheless, the opposite evolutionary trend of gene families might reflect their unique biology of these two species.

4.3. The Potential Evolutionary Significance of two Successive WGD Events in *E. Moorcroftii*

For Elaeagnaceae plants, a recent genome study of *H. rhamnoides* revealed it experienced a recent whole-genome duplication after the divergence between *H. rhamnoides* and the most closely related species jujube, with a signature *Ks* peak at c. 0.38 [4]. However, almost simultaneously, another study of *H. rhamnoides* genome indicated that it experienced two rounds of WGDs, one recent WGD (*Ks* peak at ~0.38) and one older WGD (*Ks* peak at ~0.45) [5]. In this study, our integrated *Ks* and synteny analyses provide strong evidence for the occurrence of two successive WGD events in *E. moorcroftii* genome (similar with the report by [5]), in addition, these two events possibly being shared by *E. moorcroftii* and its two close relatives (*E. mollis* and *H. rhamnoides*). It has been well-acknowledged that the frequent occurrence of WGD events has played an important role on the plant's adaptive evolution and diversification [55–59]. For example, the co-retained duplicated genes, after multiple independent WGDs in different angiosperms lineages were selected by environmental stresses during the Cretaceous–Paleocene (K-Pg) mass extinction, possibly enhanced the plant's survival and adaptation [59]. Furthermore, it was demonstrated that not only did the MRCA of angiosperms experience an ancient WGD event, but the MRCA of all extant Gymnosperms also shared an ancient WGD event, as revealed by a recent report of the *Cycas* genome, which may have contributed to seed-related morphology innovation [60]. Considering the evolutionary significance of the recognized WGD, it should be postulated that these two successive WGD events might have contributed to the evolution and adaptation of *E. moorcroftii*, although it needs to be further investigated.

5. Conclusions

As an indigenous species naturally growing in arid Northwest China, *E. moorcroftii* presents ideal tree species with great economic and ecological values that can be widely cultured in drylands for economic development and ecological restoration. In this study, the high-quality genome sequence of *E. moorcroftii* decoded here represents a small but essential step forward to understanding the evolution and adaptation of this species in drylands. Our results not only provided new insights into the genomic evolution of Elaeagnaceae species but also revealed clear evidence for two successive WGD events that occurred in the common ancestor of *E. moorcroftii*, *E. mollis*, and *H. rhamnoides*. Further, by using the *E. moorcroftii* genome as reference, more studies should be encouraged to investigate the unique biology of this species. For example, how could this species respond to long-term drought and salt stresses? Moreover, whole-genome resequencing analysis of samples with certain properties (such as resistance to abiotic factors) is helpful to screen for better germplasm resources and breed new varieties. In total, the genomic resource of *E. moorcroftii* paves ways for the further genetic improvement and widespread cultivation of this indigenous species in drylands.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/d14060468/s1>, Figure S1: The distribution frequency of PacBio's HiFi sequencing data; Figure S2: The different modes of gene duplications related to expanded and contracted gene families in *E. moorcroftii* and *E. mollis*; Figure S3: The enriched GO terms for the expanded gene families of *E. moorcroftii*; Figure S4: The enriched GO terms for the contracted gene families of *E. moorcroftii*; Figure S5: Venn diagram showing the shared and specific gene families among *E. moorcroftii*, *E. mollis*, *H. rhamnoides*, and *Z. jujube*; Figure S6: The enriched GO terms for the unique gene families of *E. moorcroftii*; Figure S7: Syntenic dot plot of the self-comparison of *E. moorcroftii*; Table S1: Information of plant genomes used in this study; Table S2: Basic statistics of PacBio's HiFi sequencing data; Table S3: Mapping rates of RNA-seq reads onto the assembly genome; Table S4: Genome assembly completeness evaluated based on BUSCO; Table S5: Information of different classes of repetitive sequences in *E. moorcroftii* genome; Table S6: Functional annotation of the predicted genes in *E. moorcroftii*; Table S7: *E. moorcroftii* genome annotation completeness evaluated based on BUSCO; Table S8: List of non-coding RNA genes in *E. moorcroftii* genome.

Author Contributions: Conceptualization, X.F., S.W. and K.S.; methodology, J.W., X.M. and K.L.; software, J.W. and X.M.; validation, X.F., S.W. and K.S.; formal analysis, J.W. and X.F.; investigation, X.F.; resources, S.W.; data curation, J.W.; writing—original draft preparation, X.F.; writing—review and editing, S.W., K.S. and H.Z.; visualization, X.F. and J.W.; supervision, K.S.; project administration, X.F.; funding acquisition, X.F. and K.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partly funded by the Shenzhen Key Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden (Grant No. SSTLAB-2022-01 to X.F.), 2022 Key Talent Project of Gansu Province (to K.S.), and the Fundamental Research Funds for the Central Universities (lzujbky-2021-49) (to S.W.).

Institutional Review Board Statement: Not applicable.

Data Availability Statement: All raw sequencing data and the assembled genome for this project were deposited in the National Genomics Data Center (NGDC; <https://ngdc.cnpc.ac.cn/> (accessed on 29 July 2021)) under accession number CRA007134.

Acknowledgments: We would like to thank the staff in Minqin Desert Botanical Garden for their help with the plant materials' collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Reynolds, J.F.; Smith, D.M.S.; Lambin, E.F.; Turner, B.L., II; Mortimore, M.; Batterbury, S.P.J.; Downing, T.E.; Dowlatabadi, H.; Fernández, R.J.; Herrick, J.E.; et al. Global desertification: Building a science for dryland development. *Science* **2007**, *316*, 847–851. [CrossRef] [PubMed]
2. Huang, J.; Yu, H.; Guan, X.; Wang, G.; Guo, R. Accelerated dryland expansion under climate change. *Nat. Clim. Change* **2016**, *6*, 166–172. [CrossRef]
3. Huang, J.; Zhang, G.; Zhang, Y.; Guan, X.; Wei, Y.; Guo, R. Global desertification vulnerability to climate change and human activities. *Land Degrad. Dev.* **2020**, *31*, 1380–1391. [CrossRef]
4. Wu, Z.; Chen, H.; Pan, Y.; Feng, H.; Fang, D.; Yang, J.; Wang, Y.; Yang, J.; Sahu, S.K.; Liu, J.; et al. The genome of *Hippophae rhamnoides* provides insights into a conserved molecular mechanism in actinorhizal and rhizobial symbiosis. *New Phytol.* **2022**, *235*, 276–291. [CrossRef]
5. Yu, L.; Diao, S.; Zhang, G.; Yu, J.; Zhang, T.; Luo, H.; Duan, A.; Wang, J.; He, C.; Zhang, J. Genome sequence and population genomics provide insights into chromosomal evolution and phytochemical innovation of *Hippophae rhamnoides*. *Plant Biotechnol. J.* **2022**. [CrossRef]
6. Wang, T.; Ren, L.; Li, C.; Zhang, D.; Zhang, X.; Zhou, G.; Gao, D.; Chen, R.; Chen, Y.; Wang, Z.; et al. The genome of a wild *Medicago* species provides insights into the tolerant mechanisms of legume forage to environmental stress. *BMC Biol.* **2021**, *19*, 96. [CrossRef]
7. Wang, M.; Zhang, L.; Tong, S.; Jiang, D.; Fu, Z. Chromosome-level genome assembly of a xerophytic plant, *Haloxylon ammodendron*. *DNA Res.* **2022**, *29*, dsac006. [CrossRef]
8. Ma, W.; Wang, P.; Huang, J.; Zhang, D.; Yang, W.; Pan, B.; Shi, W. The taxonomic relationship between *Elaeagnus moorcroftii* and *E. angustifolia* (Elaeagnaceae) based on morphological similarities and simple-sequence repeat markers. *Nord. J. Bot.* **2021**, *39*, e03011. [CrossRef]

9. Kucho, K.; Yamanaka, T.; Sasakawa, H.; Mansour, S.R.; Uchiumi, T. Different dynamics of genome content shuffling among host-specificity groups of the symbiotic actinobacterium *Frankia*. *BMC Genom.* **2014**, *15*, 609. [\[CrossRef\]](#)
10. Liang, B.B.; Wang, W.J.; Fan, X.X.; Kurakov, A.V.; Liu, Y.F.; Song, F.Q.; Chang, W. Arbuscular mycorrhizal fungi can ameliorate salt stress in *Elaeagnus angustifolia* by improving leaf photosynthetic function and ultrastructure. *Plant Biol.* **2021**, *23*, 232–241. [\[CrossRef\]](#)
11. Nazir, N.; Zahoor, M.; Nisar, M. A review on traditional uses and pharmacological importance of genus *Elaeagnus* species. *Bot. Rev.* **2020**, *86*, 247–280. [\[CrossRef\]](#)
12. Wang, Y.; Zhao, P.; Wang, Y.L.; Zhang, Y. Nutritional composition of wild *Elaeagnus angustifolia* fruits. *J. Gansu Agric. Univ.* **2006**, *41*, 130–132.
13. Ren, B.; Ru, D.; Chen, L.; Duan, N.; Li, Y.; Shi, J.; Cao, J.; Liu, B. Genome Sequence of *Elaeagnus mollis*, the first chromosome-level genome of the family Elaeagnaceae. *Genome Biol. Evol.* **2021**, *13*, evab266. [\[CrossRef\]](#) [\[PubMed\]](#)
14. Allen, G.C.; Flores-Vergara, M.A.; Krasynanski, S.; Kumar, S.; Thompson, W.F. A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nat. Protoc.* **2006**, *1*, 2320–2325. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Cheng, H.; Concepcion, G.T.; Feng, X.; Zhang, H.; Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **2021**, *18*, 170–175. [\[CrossRef\]](#)
16. Durand, N.C.; Shamim, M.S.; Machol, I.; Rao, S.S.; Huntley, M.H.; Lander, E.S.; Aiden, E.L. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **2016**, *3*, 95–98. [\[CrossRef\]](#)
17. Dudchenko, O.; Batra, S.S.; Omer, A.D.; Nyquist, S.K.; Hoeger, M.; Durand, N.C.; Shamim, M.S.; Machol, I.; Lander, E.S.; Aiden, A.P.; et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **2017**, *356*, 92–95. [\[CrossRef\]](#)
18. Simao, F.A.; Waterhouse, R.M.; Ioannidis, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with singlecopy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [\[CrossRef\]](#)
19. Kim, D.; Langmead, B.; Salzberg, S.L. HISAT: A fast spliced aligner with low memory requirements. *Nat. Methods* **2015**, *12*, 357–360. [\[CrossRef\]](#)
20. Wang, X.; Wang, L. GMATA: An integrated software package for genome-scale SSR mining, marker development and viewing. *Front Plant Sci.* **2016**, *7*, 1350. [\[CrossRef\]](#)
21. Gary, B. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *2*, 573–580.
22. Tarailo-Graovac, M.; Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **2009**, *25*, 4.10.1–4.10.14. [\[CrossRef\]](#) [\[PubMed\]](#)
23. Bao, W.; Kojima, K.K.; Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **2015**, *6*, 11. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Han, Y.; Wessler, S.R. MITE-Hunter: A program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.* **2010**, *38*, e199. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Flynn, J.M.; Hubley, R.; Goubert, C.; Rosen, J.; Clark, A.G.; Feschotte, C.; Smit, A.F. REPEATMODELER2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9451–9457. [\[CrossRef\]](#)
26. Keilwagen, J.; Wenk, M.; Erickson, J.L.; Schattat, M.H.; Grau, J.; Hartung, F. Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res.* **2016**, *44*, e89. [\[CrossRef\]](#)
27. Huang, J.; Zhang, C.; Zhao, X.; Fei, Z.; Wan, K.; Zhang, Z.; Pang, X.; Yin, X.; Bai, Y.; Sun, X. The Jujube genome provides insights into genome evolution and the domestication of sweetness/acidity taste in fruit trees. *PLoS Genet.* **2016**, *12*, e1006433. [\[CrossRef\]](#)
28. Kaul, S.; Koo, H.L.; Jenkins, J.; Rizzo, M.; Rooney, T.; Tallon, L.J.; Feldblyum, T.; Niernan, W.; Benito, M.-I.; Lin, X.; et al. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **2000**, *408*, 796–815.
29. Jaillon, O.; Aury, J.M.; Noel, B.; Policriti, A.; Clepet, C.; Casagrande, A.; Choisne, N.; Aubourg, S.; Vitulo, N.; Jubin, C.; et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* **2007**, *449*, 463–467.
30. Stanke, M.; Keller, O.; Gunduz, I.; Hayes, A.; Waack, S.; Morgenstern, B. AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **2006**, *34*, W435–W439. [\[CrossRef\]](#)
31. Dobin, A.; Davis, C.A.; Schlesinger, F.; Drenkow, J.; Zaleski, C.; Jha, S.; Batut, P.; Chaisson, M.; Gingeras, T.R. STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **2013**, *29*, 15–21. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Kovaka, S.; Zimin, A.V.; Pertea, G.M.; Razaghi, R.; Salzberg, S.L.; Pertea, M. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **2019**, *20*, 1–13. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Campbell, M.A.; Haas, B.J.; Hamilton, J.P.; Mount, S.M.; Buell, C.R. Comprehensive analysis of alternative splicing in rice and comparative analyses with *Arabidopsis*. *BMC Genom.* **2006**, *7*, 327. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Haas, B.J.; Salzberg, S.L.; Zhu, W.; Pertea, M.; Allen, J.E.; Orvis, J.; White, O.; Buell, C.R.; Wortman, J.R. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **2008**, *9*, R7. [\[CrossRef\]](#)
35. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.; McAnulla, C.; McWilliam, H.; Maslen, J.; Mitchell, A.; Nuka, G.; et al. InterProScan 5: Genome-scale protein function classification. *Bioinformatics* **2014**, *30*, 1236–1240. [\[CrossRef\]](#)
36. Conesa, A.; Götz, S.; García-Gómez, J.M.; Terol, J.; Talón, M.; Robles, M. Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **2005**, *21*, 3674–3676. [\[CrossRef\]](#)

37. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [\[CrossRef\]](#)
38. Lowe, T.M.; Eddy, S.R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **1997**, *25*, 955–964. [\[CrossRef\]](#)
39. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [\[CrossRef\]](#)
40. Griffiths-Jones, S.; Bateman, A.; Marshall, M.; Khanna, A.; Eddy, S.R. RFAM: An RNA family database. *Nucleic Acids Res.* **2003**, *31*, 439–441. [\[CrossRef\]](#)
41. Kozomara, A.; Griffiths-Jones, S. miRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res.* **2014**, *42*, D68–D73. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Lagesen, K.; Hallin, P.; Rødland, E.A.; Staerfeldt, H.H.; Rognes, T.; Ussery, D.W. RNAmmer: Consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **2007**, *35*, 3100–3108. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 1–14. [\[CrossRef\]](#)
44. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software v.7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Suyama, M.; Torrents, D.; Bork, P. PAL2NAL: Robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* **2006**, *34*, W609–W612. [\[CrossRef\]](#) [\[PubMed\]](#)
46. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **2015**, *32*, 268–274. [\[CrossRef\]](#)
47. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [\[CrossRef\]](#)
48. Shi, C.; Wang, S.; Cai, H.H.; Zhang, H.R.; Long, X.X.; Tihelka, E.; Song, W.; Feng, Q.; Jiang, R.; Cai, C.; et al. Fire-prone Rhamnaceae with South African affinities in Cretaceous Myanmar amber. *Nat. Plants* **2022**, *8*, 125–135. [\[CrossRef\]](#)
49. De Bie, T.; Cristianini, N.; Demuth, J.P.; Hahn, M.W. CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **2006**, *22*, 1269–1271. [\[CrossRef\]](#)
50. Qiao, X.; Li, Q.; Yin, H.; Qi, K.; Li, L.; Wang, R.; Zhang, S.; Paterson, A.H. Gene duplication and evolution in recurring polyploidization-diploidization cycles in plants. *Genome Biol.* **2019**, *20*, 38. [\[CrossRef\]](#)
51. Wang, Y.; Tang, H.; Debarry, J.D.; Tan, X.; Li, J.; Wang, X.; Lee, T.H.; Jin, H.; Marler, B.; Guo, H.; et al. MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **2012**, *40*, e49. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Nei, M.; Gojobori, T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **1986**, *3*, 418.
53. Jiao, Y.; Leebens-Mack, J.; Ayyampalayam, S.; Bowers, J.E.; McKain, M.R.; McNeal, J.; Rolf, M.; Ruzicka, D.R.; Wafula, E.; Wickett, N.J.; et al. A genome triplication associated with early diversification of the core eudicots. *Genome Biol.* **2012**, *13*, R3. [\[CrossRef\]](#) [\[PubMed\]](#)
54. Ma, T.; Zhou, G.; Yue, Z.; Hu, Q.; Chen, Y.; Liu, B.; Qiu, Q.; Wang, Z.; Zhang, J.; Wang, K.; et al. Genomic insights into salt adaptation in a desert poplar. *Nat. Commun.* **2014**, *4*, 2797. [\[CrossRef\]](#)
55. Jiao, Y.; Wickett, N.J.; Ayyampalayam, S.; Chanderbali, A.S.; Landherr, L.; Ralph, P.E.; Tomsho, L.P.; Hu, Y.; Liang, H.; Soltis, P.S.; et al. Ancestral polyploidy in seed plants and angiosperms. *Nature* **2011**, *473*, 97–100. [\[CrossRef\]](#)
56. Wang, J.; Sun, P.; Li, Y.; Liu, Y.; Yang, N.; Yu, J.; Ma, X.; Sun, S.; Xia, R.; Liu, X.; et al. An overlooked paleotetraploidization in Cucurbitaceae. *Mol. Biol. Evol.* **2018**, *35*, 16–26. [\[CrossRef\]](#)
57. Soltis, P.S.; Soltis, D.E. Ancient WGD events as drivers of key innovations in angiosperms. *Curr. Opin. Plant Biol.* **2016**, *30*, 159–165. [\[CrossRef\]](#)
58. Van de Peer, Y.; Mizrahi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [\[CrossRef\]](#)
59. Wu, S.; Han, B.; Jiao, Y. Genetic Contribution of Paleopolyploidy to Adaptive Evolution in Angiosperms. *Mol. Plant* **2020**, *13*, 59–71. [\[CrossRef\]](#)
60. Liu, Y.; Wang, S.; Li, L.; Yang, T.; Dong, S.; Wei, T.; Wu, S.; Liu, Y.; Gong, Y.; Feng, X.; et al. The *Cycas* genome and the early evolution of seed plants. *Nat. Plants* **2022**, *8*, 389–401. [\[CrossRef\]](#)