agronomy

MDPI

Communication

De Novo Assembly of an Allotetraploid *Artemisia argyi* Genome

Qiming Mei [1,2,3], Hanxiang Li [1,2], Yanbin Liu [4], Feng Wu [4], Chuang Liu [4], Keya Wang [1,2], Hongjun Liu [5], Cheng Peng [1,2,6], Zhengfeng Wang [1,2], Long Wang [1,2], Zhanfeng Liu [1,2], Junhua Yan [1,2] and Wei Zhang [1,2,*]

1 Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China
2 South China National Botanical Garden, Guangzhou 510650, China
3 Guangzhou Linfang Ecological Technology Co., Ltd., Guangzhou 510520, China
4 Tea Bureau of Tongbai County, Nanyang 474750, China
5 People's Government of Tongbai County, Nanyang 474750, China
6 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: zhangwei@scbg.ac.cn; Tel.: +86-20-37086136

Abstract: The Chinese mugwort (*Artemisia argyi* Lév. et Vaniot) is an important traditional Chinese medicine plant that is ubiquitously disturbed in Asia. However, the molecular mechanisms that reflect the natural evolution of *Artemisia argyi* remain unclear. In this study, a high-quality draft assembly of the allotetraploid *A. argyi* (ArteW1-Tongbai) was conducted utilizing PacBio long-read sequencing and Hi-C technologies. The assembly is about 7.20 Gb with a contig N50 length of 0.87 Mb. The allotetraploid genome of ArteW1-Tongbai is highly heterozygous and rich in repeat sequences (the heterozygous ratio is 1.36%, and the repeat rate is 86.26%). A total of 139,245 protein-coding genes were identified. The KEGG enrichment analysis revealed that 846 species-specific genes were related to the biosynthesis of secondary metabolites. The plants with allopolyploid genomes can potentially exhibit a better adaptive capacity to environmental stresses and accumulation of secondary metabolites. Therefore, the genome assembly serves as a valuable reference for *Artemisia*, the genus characterized by species richness and diverse specialized metabolites.

Keywords: *Artemisia argyi*; de novo assembly; Hi-C technology; allotetraploid

Citation: Mei, Q.; Li, H.; Liu, Y.; Wu, F.; Liu, C.; Wang, K.; Liu, H.; Peng, C.; Wang, Z.; Wang, L.; et al. De Novo Assembly of an Allotetraploid *Artemisia argyi* Genome. *Agronomy* **2023**, *13*, 436. https://doi.org/10.3390/agronomy13020436

Academic Editor: Mirosław Tyrka

Received: 5 December 2022
Revised: 18 January 2023
Accepted: 30 January 2023
Published: 1 February 2023

1. Introduction

The *Artemisia argyi* H. Lév. et Vaniot (Asteraceae), also known as Chinese mugwort or Aicao, is ubiquitously distributed in Asia, and is an important traditional Chinese medicine (TCM) herb. Tongbai County, Henan Province of Central China, is one of the main places for producing *A. argyi*. The cultivated and wild *A. argyi* growing in Tongbai are named "Tongbai Ai". Dried *A. argyi* leaves are the original material for moxibustion. Previous phytochemical studies reported that the main components present in *A. argyi* leaves, such as flavonoids, polysaccharides, terpenoid, polyketides, and phenolic acids, exhibit antioxidant, anti-cancer, antimicrobial, and neuroprotective activities [1]. Recently, the terpenoid biosynthesis pathway was analyzed by transcriptome profiling [2]. However, the molecular mechanisms underlying the biological synthesis of medicinal components of *A. argyi* are still largely unknown.

Polyploidy changes the quantitative and qualitative patterns of secondary metabolite production in plants [3]. Polyploidy, or whole-genome duplication (WGD), is one of the main evolutionary forces enhancing the adaptive potential of organisms [4]. There are two types of polyploidies in plants: allopolyploidy and autopolyploid. The formation of allopolyploidy involving interspecific hybridization and genome duplication is more prevalent than autopolyploidization in nature [5]. Allopolyploidation may be the most common mechanism of sympatric speciation and promotes adaptation in plants [6]. Many Asteraceae species are allopolyploid because of interspecific hybridization [7]. The *Artemisia* is a

large genus of the Asteraceae family that consists of about 380 species [8]. The chromosome number of *Artemisia* species was identified mainly to be x = 9, which is prevalent; or x = 8, which is less frequent [9,10].

In the last decade, a dozen *Artemisia* chloroplast genomes were completely characterized, which provide a valuable resource for phylogenetic analysis [11,12]. Recently, the genome assembly of *A. annua* set an excellent model system for studying the artemisinin synthesis and evolution of *Artemisia* [13]. In this study, the high-quality assembly and annotation of an allopolyploid genome of *A. argyi* were conducted utilizing PacBio long-read sequencing and Hi-C technologies. The results provide fundamental information for illustrating the structure and organization of the highly heterozygous genome. In addition, the genome assembly of *A. argyi* is a useful tool for dissecting phytochemical component synthesis pathways and drug discovery.

2. Materials and Methods

2.1. Plant Sample, Sequencing, and Genome Survey

In September 2019, a wild individual of *A. argyi* (ArteW1-Tongbai) was collected from the "Protected Area of Tongbai Ai" located at Huaiyuan Town, Tongbai County, Henan Province, China (32°29′201″ N, 113°15′12″ E) at altitudes of 287.67 m (Figure 1a,b). The samples were sent to the commercial genome research organization Frasergen (Wuhan, China) for DNA extraction and sequencing. The total genomic DNA was extracted from fresh leaves utilizing a modified CTAB method, in which LiCl and polyvinylpyrrolidone (PVP) were added to the extraction buffer to remove the high level of polysaccharides, polyphenolics, and secondary metabolites [14]. RNA contaminant was removed by RNaseA. Then, the quality of the DNA was checked using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) as well as agarose gel electrophoresis. Qubit 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) was used to quantify the DNA.

Four libraries with insert sizes larger than 15 kb were constructed according to the SMRTbell Express Template Prep kit 2.0 (Pacific Biosciences, Menlo Park, CA, USA). Then, single-molecule DNA sequencing was conducted on the PacBio Sequel II platform (Pacific Biosciences, Menlo Park, CA, USA). A total of 570.70 Gb (80× of the estimated genome size) bases were generated. The raw reads were filtered by HTQC v.1.92.310 to remove adapters and low-quality sequences (read pairs with an average quality lower than 20 and with any end shorter than 75 bp) [15]. In addition, two libraries with an insertion length of 150 bp were generated and sequenced on the HiSeq X Ten platform (Illumina, San Diego, CA, USA). A total of 162.63 Gb of short-read sequences were generated for the genome survey. Then, the genome size, the level of heterozygosity, and repeat content of the genome were estimated utilizing the *k*-mer method. The 17-mer survey was conducted by GCE software [16].

2.2. Genome Assembly and Quality Evaluation

Clean data generated by the PacBio Sequel II platform were processed to genome assembly after filtering by fastp v.0.12.6 [17]. The draft genome was assembled using mecat2 with default parameters [18]. Then, the genome was polished by using the arrow pipeline from the SMRT link 4 toolkit to correct errors in the initial genome assembly. Finally, the short reads derived from the Illumina sequencer were utilized to correct the remaining errors by pilon v.1.22 [19].

In this study, the Hi-C technique was employed to construct chromosome-level assemblies. Fresh leaves of ArteW1-Tongbai were fixed in formaldehyde to create DNA–protein bonds. Then, the restriction enzyme MboI was utilized to digest the chromatin. After re-ligation, the DNA was sheared into 300–500 bp fragments by sonication. Then, the Hi-C library was prepared following a standard procedure and sequenced on the IlluminaHiSeq X Ten platform under the PE150 model. A total of 381 Gb data were generated from the Hi-C library (50× of the estimated genome size). Next, the raw data were filtered to remove self-ligation, non-ligation, and other invalid reads by fastp v.0.12.6 [17]. The

clean reads were mapped to the polished genome using BWAv.0.7.16a (r1181) with default parameters [20].

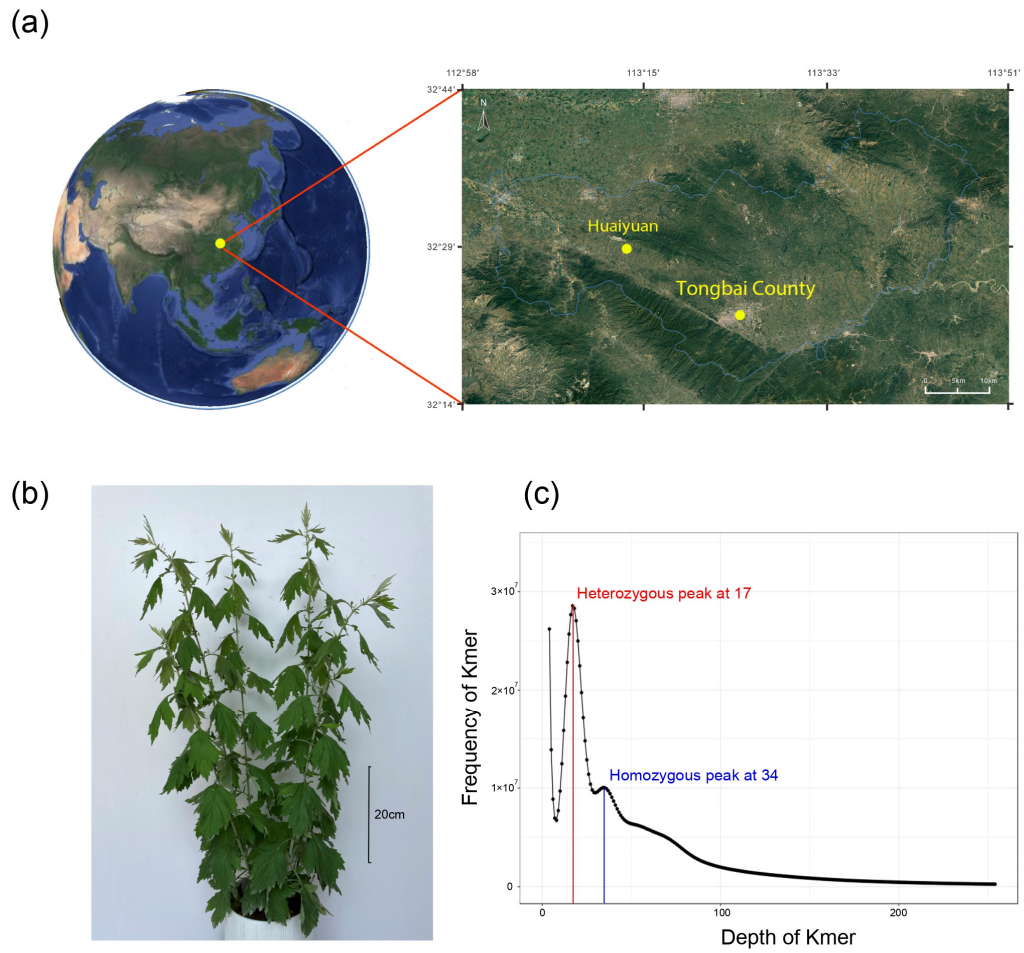

Figure 1. Location of the study site and genomic characteristic of the wild *A. argyi*. (**a**) Sampling site of ArteW1-Tongbai (Google Earth, earth.google.com/web/, accessed on 30 April 2020). (**b**) Picture of the ArteW1-Tongbai growing in "Protected Area of wild *A. argyi*". (**c**) Estimation of the genome size of ArteW1-Tongbai based on *k*-mer analysis.

The 3d-DNA pipeline was used to cluster, order, and orient the Hi-C contigs [21]. The 3d-DNA pipeline anchored and oriented 7.20 Gb (98.98%) contigs into 32 superscaffolds (named aar1 to aar32) according to the syntenic relationship (Supplementary Table S1). Then, a contact map was plotted using Juicer [22], and visualized and corrected by Juicebox [23]. Finally, the quality of pseudo-chromosome assembly was assessed by BUSCO v.3.0.2 with the embryophyta_odb10 dataset [24].

2.3. De Novo Genome Annotation

The gene structural annotation of *A. argyi* was conducted following three strategies: (1) de novo prediction performed by AUGUSTUS v.3.3.1 [25] and GENSCAN [26]; (2) homology-based annotation using Exonerate v.2.2.0 with the default parameters [27]; and (3) finding coding regions in RNA-Seq transcripts by Cufflinks v.2.2.1 [28]. Finally, the results of the three methods were combined into gene models by MAKER [29].

The predicted genes were further functionally annotated according to the best match of the alignments with *Lactuca saligna*, *Mikania micrantha*, *Cynara cardunculus* var. *scolymus*, *Artemisia annua*, and *Helianthus annuus*. The sequences were queried against databases including the National Center for Biotechnology Information (NCBI) Non-Redundant (NR) database [30], TrEMBL [31], Swiss-Prot [32], and the Kyoto Encyclopedia of Genes

and Genomes (KEGG) [33] database by blasstp (e-value = 1×10^{-5}) [34]. In addition, InterProScan v.5.35–74.0 [35] and PfamScan [36] were used to annotate the protein domains based on the InterPro [37] and Pfam [38] databases. Gene Ontology (GO) [39] IDs for predicted genes were obtained from Blast2GO [40]. Finally, Benchmarking Universal Single-Copy Orthologs (BUSCO) was used to validate the gene annotations with the embryophyta_odb10 and default parameters [24].

2.4. Evolutionary Analysis

The genomes of six Asteraceae species including *L. sativa*, *M. micrantha*, *C. cardunculus* var. *scolymus*, *A. annua*, and *H. annuus*, as well as *A. argyi* were included in the phylogenetic analysis. The *Solanum tuberosum* was used as the outgroup. OrthoFinder v.2.3.1 was used for the identification of orthologous groups of the *A. argyi* genome with an e-value of 10^{-5} [41]. Gene families unique for each species were extracted from the clustering result. Expansion and contraction analysis of gene families was conducted using CAFE v.2.2 [42]. The enrichment of KEGG and Gene Ontology terms for the expansion, contraction, and lineage-specific gene families was performed using goseq [43].

Orthologous groups (OGs) with single-copy genes were used in the phylogenetic analysis. Protein sequences from each OG were aligned using MUSCLE v.3.8.31 [44]. Ambiguous sites were removed manually by trimAI [45]. The final dataset was generated by concatenating alignments. The phylogenetic tree was constructed using IQ-TREE software with maximum likelihood (ML) algorithm. The divergence time of species using MCMCTree was implemented in PAML v.4.9 package [46]. The tree was calibrated by Asteraceae crown age (95–106 million years ago, Mya) and split time between *C. cardunculus* var. *scolymus* and *L. sativa* (27–40 Mya). The time-calibrating points were obtained from the timetree website accessed on 14 July 2020 (timetree.org).

We utilized the wgd software to calculate *K*s distribution (ranging from 0.05 to 3) among paralogs from *A. argyi* [47]. The paralogs were pruned on the basis of co-linearity analysis using i-ADHoRe [48]. Then, we fitted the *K*s distribution of the paralogs from each hypothesized WGD peak according to the fitted mixture model (BGMM in 'wgd').

3. Results

3.1. Genomic Characteristics of Wild A. argyi

The *A. argyi* (ArteW1-Tongbai) used in this study is native to the semi-arid hilly region of Huaiyuan Town, Tongbai County, Henan Province, China (32°29′201″ N, 113°15′12″ E) at altitudes of 287.67 m (Figure 1a,b). The karyotype analysis revealed that the nuclear genome of ArteW1-Tongbai comprises 34 chromosomes (2n = 34).

The genome size of ArteW1-Tongbai was estimated through *k*-mer analysis using Illumina short reads at a *k*-mer size of 17. As shown in Figure 1c, the frequency distribution of *k*-mer manifested two clear peaks at depths of 17 and 34, corresponding to the heterozygous and homozygous reads, respectively. Then, the haploid genome size of ArteW1-Tongbai was estimated to be about 3.81 Gb based on the homozygous peak depth, and the diploid genome size was predicted to be about 7.62 Gb heterozygous based on the heterozygous peak depth.

The *k*-mer results indicated that the heterozygous ratio of ArteW1-Tongbai is 1.36%, and the repeat rate is 86.26% (Figure 1c). Both the chromosome number and genome size were different from the diploid *A. argyi* (2n = 2x = 18, genome size = 4.3 Gb) [10], suggesting that ArteW1-Tongbai possesses an allotetraploid genome. The genome size of ArteW1-Tongbai is approximately four-fold higher than the closely-related species *A. annua* (1.74 Gb) [13].

3.2. De Novo Assembly and Annotation of the A. argyi Genome

In this study, an assembly of an allopolyploid *A. argyi* genome was conducted utilizing PacBio long-read sequencing and Hi-C technologies. The 7.20 Gb genome is assembled into eight main clusters of superscaffolds and further divided into 32 pseudochromosomes (numbered aar1–32) with contig N50 of 0.87 Mb and scaffold N50 of 215.81 Mb. The size of superscaffolds ranged from 175.86 Mb to 372.72 Mb. In addition, the 32 pseudochro-

mosomes formed eight main clusters, and each cluster contained two "haplotype-fused" homologous chromosomes (Supplementary Figure S1 and Supplementary Table S1). The number of pseudochromosomes (32) did not match the karyotype result (34), because four pairs of them are fused and very difficult to isolate. The Hi-C assembly contained 98.98% of the assembled sequences. The completeness of the genome assembly was qualified by BUSCO [24]. In total, 1572 out of 1614 BUSCOs were identified as complete (97.40%).

Genes were predicted in the allotetraploid *A. argyi* genome by a pipeline integrating de novo, homology-based, and RNA-Seq methods. A total of 139,245 protein-coding gene models were identified. Next, the gene models were evaluated using the same methods as gene prediction. A total of 133,932 (96.18%) predicted genes were supported by two or three methods. In addition, BUSCO was utilized to validate the annotation results: 1607 of 1614 (99.6%) BUSCOs were complete. Then, the gene models were further functionally annotated by blasting against databases. A total of 134,956 genes were successfully annotated, corresponding to 96.92% of the predicted genes.

3.3. Comparative Genome Analysis of Asteraceae Species

In this study, the genome of *A. argyi* was compared to the other five Asteraceae species with genomic assemblies (*A. annua*, *H. annuus*, *L. sativa*, *C.* var. *scolymus*, and *M.micrantha*), and the asteroid *S. tuberosum* was used as the outgroup of the phylogenetic analysis. A total of 139,245 genes belonging to 27,279 families in the *A. argyi* genome were identified (Table 1).

Table 1. Comparison of gene number of *A. argyi* and other species.

Species	Genes	Families	Clustered Genes	Unclustered Genes	Specific Genes	Specific Families
A. argyi	139,245	27,279	123,512	15,733	18,576	4858
A. annua	66,918	23,935	61,118	5800	7547	2030
C. cardunculus	38,406	16,749	37,567	839	1695	333
H. annuus	44,144	16,818	40,078	4066	5909	1656
L. sativa	45,243	17,972	43,605	1638	5612	1000
M. micrantha	46,351	17,554	43,088	3263	8059	1522
S. tuberosum	37,967	15,602	35,723	2244	7563	1541

The phylogeny of *A. argyi* and other Asteraceae species was reconstructed using a concatenated sequence alignment of 24 single-copy genes shared by *A. argyi* and another six plant species (Figure 2a). The *A. argyi* was clustered with the closely related *A. annua*, as expected. The divergent time of *A. argyi* and *A. annua* was estimated to be around 18.28 Mya. When compared to *A. annua*, a total of 9119 gene families were expanded in *A. argyi*, whereas 401 were contracted (Figure 2a). moreover, 4697 gene families were shared by *A. argyi* and *A. annua*, whereas 4858 gene families were specific to *A. argyi* (Figure 2b). The KEGG enrichment analysis indicated that the number of genes involved in secondary metabolites biosynthesis (846) was notably higher than the species-specific genes of other pathways, for example, the biosynthesis of amino acids (256) (Figure 2c). Most expanded gene families were associated with secondary metabolites, such as isoquinoline alkaloid biosynthesis (ko00950), tropane, piperidine, and pyridline alkaloid biosynthesis (ko00960), monobactam biosynthesis (ko00261), betalain biosynthesis (ko00965), and glucosinolate biosynthesis (ko00966) (Figure 2c). Additionally, some key enzymes of secondary metabolites biosynthesis pathways were expanded in *A. argyi.* For example, the polyphenol oxidase (K00422) primary-amine oxidase (K00276) catalyzed the synthesis of main metabolic intermediates of various isoquinoline alkaloids. Isoquinoline alkaloids were enriched in herbal plants that have been used for their anti-inflammatory, antimicrobial, and analgesic effects [49].

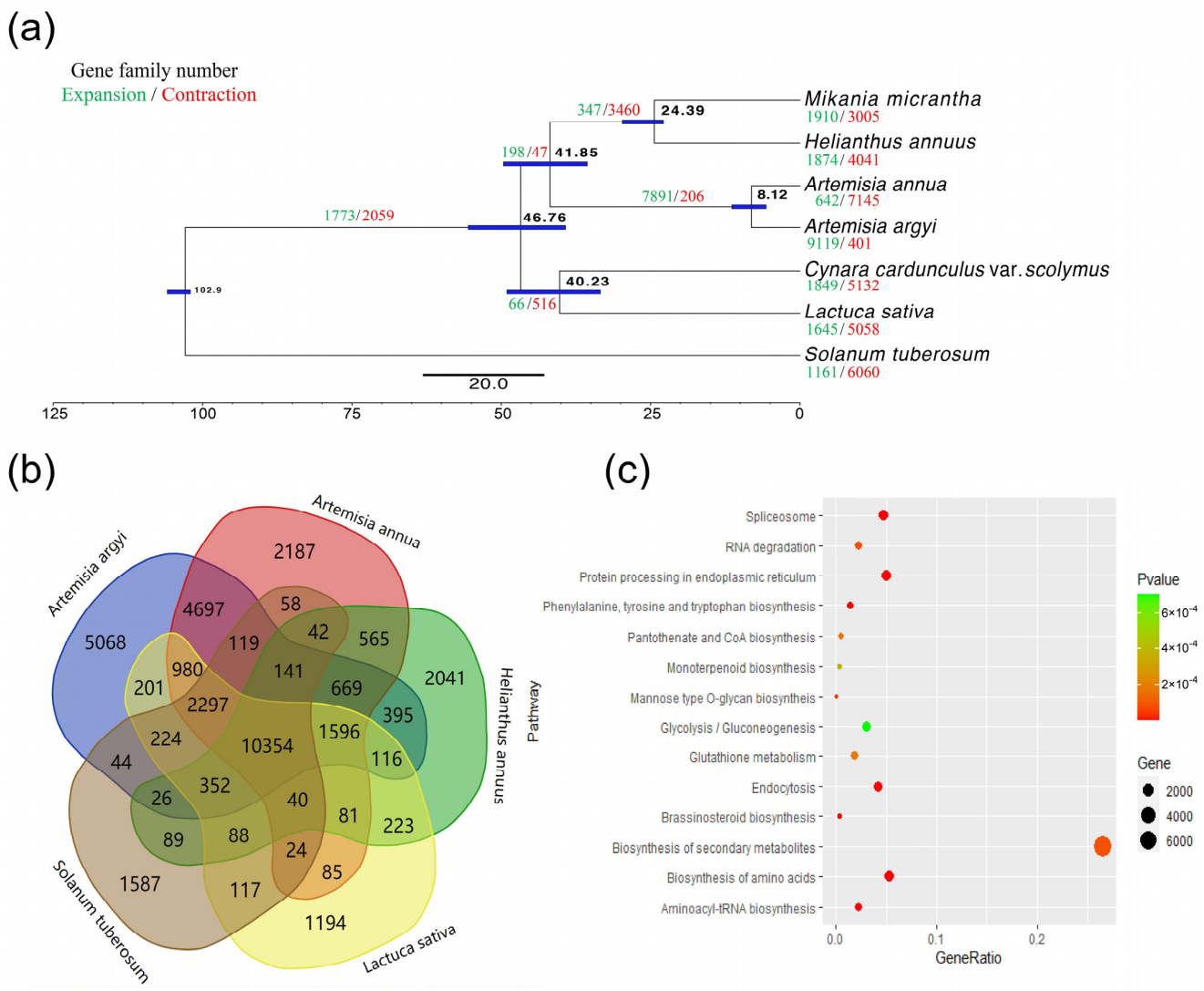

Figure 2. Evolutionary analysis of *A. argyi* genome. (**a**) Phylogenetic tree of Asteraceae species. Branch and taxa labels are numbers of gene families manifesting expansion (green) and contraction (red). Node labels are divergence times estimated by maximum likelihood (ML) algorithm. (**b**) Overlap of gene families of four Asteraceae species and *Solanum*. (**c**) KEGG enrichment analysis of expansion gene families in the *A. argyi* genome.

Moreover, the distribution of synonymous substitutions per synonymous site (*Ks*) revealed the whole-genome duplication (WGD) events that occurred in the *A. argyi* genome. The peak observed from the lg(*Ks*) = −3, demonstrating a most recent genome duplication, and peak at lg(*Ks*) = 0 represented an older WGD that occurred in the evolution of *Artemisia* (Figure 3).

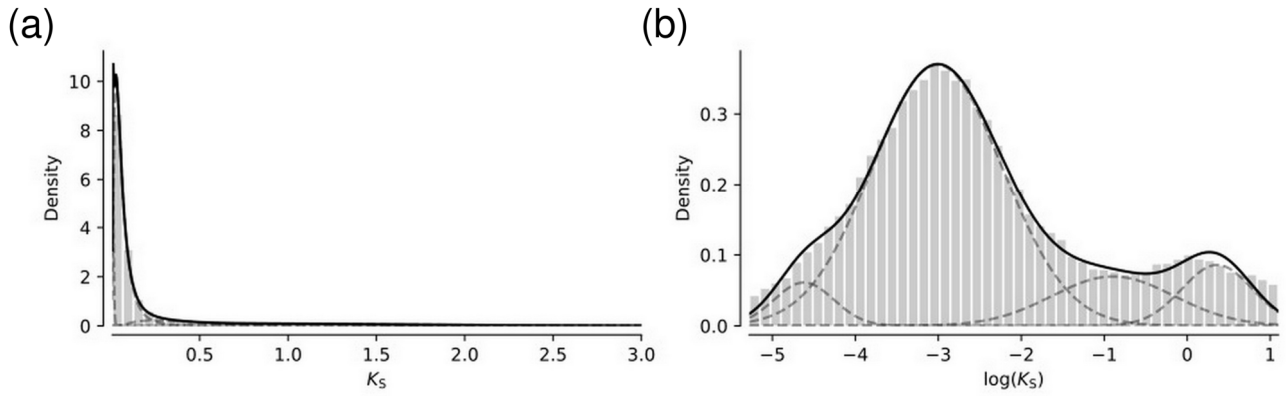

Figure 3. Distribution of synonymous substitutions per site (*Ks*) (**a**) and lg(*Ks*) (**b**) inferring WGD events.

4. Discussion

A. argyi is an important traditional Chinese medicine plant. In this study, we conducted high-quality assembly and annotation on an allopolyploid genome of wild *A. argyi*. The results showed that the allotetraploid genome of ArteW1-Tongbai is highly heterozygous and rich in repeat sequences. The *A. argyi* genome contains a larger number of gene families compared to other Asteraceae species (Table 1) [13,50–53]. The *Ks* results indicated that a duplication event occurred recently and led to the synthesis of the allotetraploid *A. argyi* genome (Figure S1). The plants with allopolyploid genomes tend to exhibit a better adaptive capacity to environmental stresses and the accumulation of secondary metabolites [6]. A total of 846 species-specific genes were related to the biosynthesis of secondary metabolites. The results of the genome characteristics are consistent with the results reported recently (high heterozygosity and high repetitive sequences), but vary in terms of gene numbers (62,844–279,294) and genome size (7.20 Gb in this study, and 7.44–7.87 Gb in previous reports) [54,55]. In addition, all three assemblies demonstrate chromosome fusion events by synteny analysis (Figure S1a).

A high-quality reference genome assembly serves as a fundamental resource for research on polyploid plant genomes. On the other hand, the high ratio of repeat sequences and heterozygosity are the main obstacles to the assembly of polyploid plant genomes [56]. In this case, the *A. argyi* genome is highly heterozygous (heterozygosity = 1.36%) and contains a high ratio of repeat sequences (86.26%) (Figure 1c). Therefore, we could not derive a haplotype-resolved assembly from the allopolyploid genome despite utilizing long-read sequencing and Hi-C assembly technologies. For further study, we plan to conduct the genome sequencing of diploid *A. argyi* and provide an updated assembly of the allopolyploid genome.

Polyploidization is common in plants. The increase in gene numbers facilitates the production of secondary metabolites [3]. For example, the tetraploid *Echinacea purpurea* has a higher abundance of cichoric acid than diploid [57]. Moreover, hybridization enhances the variation of secondary metabolites and herbivore resistance [58]. The allopolyploidy contains two sets of genes involved in secondary metabolite biosynthesis inherited from parental species. Recent research reported that the hybrid cultivar of oolong tea (*Camellia sinensis*) exhibited structural variations and the expansion of terpene synthase gene families, which contributed to the high aroma and stress tolerance [59]. As a result, the allotetraploid *A. argyi* showed great potential for metabolic engineering.

The *A. argyi* genome contains a larger number of gene families compared to other Asteraceae species (Table 1). The *Artemisia* species synthesized diverse secondary metabolites [60]. In this study, the functional annotation and KEGG enrichment analysis revealed that many secondary metabolites synthesis gene families are unique to *A. argyi*. Our study provides fundamental resources for analyzing diverse specialized metabolites in *Artemisia*.

5. Conclusions

In summary, we assembled and annotated a high-quality genome of the genus *Artemisia*. The genome data can be used in studying gene functions, the genome evolution of Asteraceae, and *Artemisia* taxonomy. The results showed that *A. argyi* diverged from other *Artemisia* species very recently (at ~8.12 Mya) and experienced two rounds of WGD events. We also found that the genes of secondary metabolites biosynthesis expanded. These results can provide insight into thedrug discovery of *Artemisia* plants. The findings suggest that the allotetraploid *A. argyi* can be used as a useful tool for dissecting phytochemical component synthesis.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy13020436/s1, Figure S1: De novo assembly of the A. argyi genome; Table S1: Estimation of the genome size of ArteW1-Tongbai based on *k*-mer analysis.

Author Contributions: Conceptualization, Q.M. and W.Z.; methodology, H.L. (Hanxiang Li); software, Z.W.; validation, L.W. and Z.L.; formal analysis, Q.M.; investigation, Y.L. and C.L.; resources, F.W. and K.W.; writing—original draft preparation, Q.M.; writing—review and editing, Q.M. and W.Z.; visualization, C.P.; supervision, J.Y.; project administration, W.Z.; funding acquisition, F.W. and H.L. (Hongjun Liu). All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant from Tongbai County Government, Henan Province of China (Y841191001), and the National Natural Science Foundation of China (3160020257).

Data Availability Statement: The *A. argyi* genome assembly was submitted to the National Genomics Data Centerof China (accession number CRX552966). The raw sequencing reads and Hi-C data were deposited in the NCBI Sequence Read Archive (SRA) under the BioSample SAMN20447770. The annotation file is available at https://figshare.com/articles/dataset/Gene_annotation_of_Artemisia_argyi/16621711 (uploaded on 15 September 2021).

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

References

1. Shi, X.S.; Song, Y.P.; Meng, L.H.; Yang, S.Q.; Wang, D.J.; Zhou, X.W.; Ji, N.Y.; Wang, B.G.; Li, X.M. Isolation and characterization of antibacterial carotene sesquiterpenes from *Artemisia argyi* associated endophytic *Trichoderma virens* QA-8. *Antibiotics* **2021**, *10*, 213. [CrossRef]
2. Liu, M.; Zhu, J.; Wu, S.; Wang, C.; Guo, X.; Wu, J.; Zhou, M. De novo assembly and analysis of the *Artemisia argyi* transcriptome and identification of genes involved in terpenoid biosynthesis. *Sci. Rep.* **2018**, *8*, 5824. [CrossRef] [PubMed]
3. Madani, H.; Escrich, A.; Hosseini, B.; Sanchez-Munoz, R.; Khojasteh, A.; Palazon, J. Effect of polyploidy induction on natural metabolite production in medicinal plants. *Biomolecules* **2021**, *11*, 899. [CrossRef] [PubMed]
4. Van de Peer, Y.; Mizrachi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [CrossRef] [PubMed]
5. Qin, J.; Mo, R.; Li, H.; Ni, Z.; Sun, Q.; Liu, Z. The transcriptional and splicing changes caused by hybridization can be globally recovered by genome doubling during allopolyploidization. *Mol. Biol. Evol.* **2021**, *38*, 2513–2519. [CrossRef] [PubMed]
6. Alix, K.; Gerard, P.R.; Schwarzacher, T.; Heslop-Harrison, J.S.P. Polyploidy and interspecific hybridization: Partners for adaptation, speciation and evolution in plants. *Ann. Bot.* **2017**, *120*, 183–194. [CrossRef]
7. Qi, X.Y.; Wang, H.B.; Song, A.P.; Jiang, J.F.; Chen, S.M.; Chen, F.D. Genomic and transcriptomic alterations following intergeneric hybridization and polyploidization in the *Chrysanthemum nankingense* × *Tanacetum vulgare* hybrid and allopolyploid (Asteraceae). *Hortic. Res.* **2018**, *5*, 5. [CrossRef]
8. Lin, Y.; Ling, Y.; Humphries, C.J.; Gilbert, M.G. *Artemisia*. In *Flora of China*; Wu, Z., Raven, P.H., Eds.; Science Press: Beijing, China, 2011; pp. 20–21.
9. Garcia, S.; Canela, M.A.; Garnatje, T.; Mcarthur, E.D.; Pellicer, J.; Sanderson, S.C.; Valles, J. Evolutionary and ecological implications of genome size in the North American endemic sagebrushes and allies (*Artemisia*, Asteraceae). *Biol. J. Linn. Soc.* **2008**, *94*, 631–649. [CrossRef]
10. Pellicer, J.; Garcia, S.; Canela, M.A.; Garnatje, T.; Korobkov, A.A.; Twibell, J.D.; Valles, J. Genome size dynamics in *Artemisia* L. (Asteraceae): Following the track of polyploidy. *Plant Biol.* **2010**, *12*, 820–830. [CrossRef]
11. Kang, S.H.; Kim, K.; Lee, J.H.; Ahn, B.O.; Won, S.Y.; Sohn, S.H.; Kim, J.S. The complete chloroplast genome sequence of medicinal plant, *Artemisia argyi*. *Mitochondrial DNA Part B* **2016**, *1*, 257–258. [CrossRef]

12. Kim, G.B.; Lim, C.E.; Kim, J.S.; Kim, K.; Lee, J.H.; Yu, H.J.; Mun, J.H. Comparative chloroplast genome analysis of *Artemisia* (Asteraceae) in East Asia: Insights into evolutionary divergence and phylogenomic implications. *BMC Genom.* **2020**, *21*, 415. [CrossRef]
13. Shen, Q.; Zhang, L.; Liao, Z.; Wang, S.; Yan, T.; Shi, P.; Liu, M.; Fu, X.; Pan, Q.; Wang, Y.; et al. The Genome of *Artemisia annua* provides insight into the evolution of Asteraceae family and artemisinin biosynthesis. *Mol. Plant* **2018**, *11*, 776–788. [CrossRef]
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]
15. Yang, X.; Liu, D.; Liu, F.; Wu, J.; Zou, J.; Xiao, X.; Zhao, F.; Zhu, B. HTQC: A fast quality control toolkit for Illumina sequencing data. *BMC Bioinform.* **2013**, *14*, 33. [CrossRef]
16. Liu, B.; Shi, Y.; Yuan, J.; Hu, X.; Zhang, H.; Li, N.; Li, Z.; Chen, Y.; Mu, D.; Fan, W. Estimation of genomic characteristics by analyzing *k*mer frequency in de novo genome projects. *arXiv* **2013**, arXiv:1308.2012.
17. Chen, S.; Zhou, Y.Q.; Chen, Y.R.; Gu, J. Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
18. Xiao, C.-L.; Chen, Y.; Xie, S.-Q.; Chen, K.-N.; Wang, Y.; Han, Y.; Luo, F.; Xie, Z. MECAT: Fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. *Nat. Methods* **2017**, *14*, 1072–1074. [CrossRef]
19. Walker, B.J.; Abeel, T.; Shea, T.; Priest, M.; Abouelliel, A.; Sakthikumar, S.; Cuomo, C.A.; Zeng, Q.D.; Wortman, J.; Young, S.K.; et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **2014**, *9*, e112963. [CrossRef]
20. Li, R.; Zhu, H.; Ruan, J.; Qian, W.; Fang, X.; Shi, Z.; Li, Y.; Li, S.; Shan, G.; Kristiansen, K.; et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **2010**, *20*, 265–272. [CrossRef]
21. 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]
22. Durand, N.C.; Shamim, M.S.; Machol, I.; Rao, S.S.P.; 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]
23. Durand, N.C.; Robinson, J.T.; Shamim, M.S.; Machol, I.; Mesirov, J.P.; Lander, E.S.; Aiden, E.L. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **2016**, *3*, 99–101. [CrossRef]
24. Seppey, M.; Manni, M.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness. *Methods Mol. Biol.* **2019**, *1962*, 227–245.
25. 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]
26. Burge, C.; Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **1997**, *268*, 78–94. [CrossRef]
27. Slater, G.S.; Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinform.* **2005**, *6*, 31.
28. Trapnell, C.; Williams, B.A.; Pertea, G.; Mortazavi, A.; Kwan, G.; van Baren, M.J.; Salzberg, S.L.; Wold, B.J.; Pachter, L. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **2010**, *28*, 511–515. [CrossRef]
29. Cantarel, B.L.; Korf, I.; Robb, S.M.C.; Parra, G.; Ross, E.; Moore, B.; Holt, C.; Alvarado, A.S.; Yandell, M. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **2008**, *18*, 188–196. [CrossRef]
30. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]
31. Boeckmann, B.; Bairoch, A.; Apweiler, R.; Blatter, M.C.; Estreicher, A.; Gasteiger, E.; Martin, M.J.; Michoud, K.; O'Donovan, C.; Phan, I.; et al. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **2003**, *31*, 365–370. [CrossRef]
32. Soudy, M.; Anwar, A.M.; Ahmed, E.A.; Osama, A.; Ezzeldin, S.; Mahgoub, S.; Magdeldin, S. UniprotR: Retrieving and visualizing protein sequence and functional information from Universal Protein Resource (UniProt knowledgebase). *J. Proteom.* **2020**, *213*, 103613. [CrossRef] [PubMed]
33. Kanehisa, M.; Goto, S.; Sato, Y.; Furumichi, M.; Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* **2012**, *40*, D109–D114. [CrossRef] [PubMed]
34. Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.H.; Zhang, Z.; Miller, W.; Lipman, D.J. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402. [CrossRef] [PubMed]
35. Jones, P.; Binns, D.; Chang, H.Y.; Fraser, M.; Li, W.Z.; 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] [PubMed]
36. Mistry, J.; Bateman, A.; Finn, R.D. Predicting active site residue annotations in the Pfam database. *BMC Bioinform.* **2007**, *8*, 298. [CrossRef]
37. Mitchell, A.; Chang, H.Y.; Daugherty, L.; Fraser, M.; Hunter, S.; Lopez, R.; McAnulla, C.; McMenamin, C.; Nuka, G.; Pesseat, S.; et al. The InterPro protein families database: The classification resource after 15 years. *Nucleic Acids Res.* **2015**, *43*, D213–D221. [CrossRef]

38. El-Gebali, S.; Mistry, J.; Bateman, A.; Eddy, S.R.; Luciani, A.; Potter, S.C.; Qureshi, M.; Richardson, L.J.; Salazar, G.A.; Smart, A.; et al. The Pfam protein families database in 2019. *Nucleic Acids Res.* **2019**, *47*, D427–D432. [CrossRef]
39. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
40. Conesa, A.; Gotz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genom.* **2008**, *2008*, 619832. [CrossRef]
41. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef]
42. 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] [PubMed]
43. Young, M.D.; Wakefield, M.J.; Smyth, G.K.; Oshlack, A. Gene ontology analysis for RNA-seq: Accounting for selection bias. *Genome Biol.* **2010**, *11*, R14. [CrossRef]
44. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
45. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1972–1973. [CrossRef]
46. Yang, Z.; Rannala, B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol. Biol. Evol.* **2006**, *23*, 212–226. [CrossRef]
47. Zwaenepoel, A.; Van de Peer, Y. Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates. *Mol. Biol. Evol.* **2019**, *36*, 1384–1404. [CrossRef] [PubMed]
48. Proost, S.; Fostier, J.; De Witte, D.; Dhoedt, B.; Demeester, P.; VandePeer, Y.; Vandepoele, K. i-ADHoRe 3.0—Fast and sensitive detection of genomic homology in extremely large data sets. *Nucleic Acids Res.* **2012**, *40*, e11. [CrossRef] [PubMed]
49. Yun, D.; Yoon, S.Y.; Park, S.J.; Park, Y.J. The anticancer effect of natural plant alkaloid isoquinolines. *Int. J. Mol. Sci.* **2021**, *22*, 1653. [CrossRef]
50. Scaglione, D.; Reyes-Chin-Wo, S.; Acquadro, A.; Froenicke, L.; Portis, E.; Beitel, C.; Tirone, M.; Mauro, R.; Lo Monaco, A.; Mauromicale, G.; et al. The genome sequence of the outbreeding globe artichoke constructed de novo incorporating a phase-aware low-pass sequencing strategy of F1 progeny. *Sci. Rep.* **2016**, *6*, 19427. [CrossRef]
51. Staton, S.E.; Bakken, B.H.; Blackman, B.K.; Chapman, M.A.; Kane, N.C.; Tang, S.; Ungerer, M.C.; Knapp, S.J.; Rieseberg, L.H.; Burke, J.M. The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* **2012**, *72*, 142–153. [CrossRef]
52. Reyes-Chin-Wo, S.; Wang, Z.; Yang, X.; Kozik, A.; Arikit, S.; Song, C.; Xia, L.; Froenicke, L.; Lavelle, D.O.; Truco, M.-J.; et al. Genome assembly with in vitro proximity ligation data and whole-genome triplication in lettuce. *Nat. Commun.* **2017**, *8*, 14953. [CrossRef] [PubMed]
53. Liu, B.; Yan, J.; Li, W.; Yin, L.; Li, P.; Yu, H.; Xing, L.; Cai, M.; Wang, H.; Zhao, M.; et al. *Mikania micrantha* genome provides insights into the molecular mechanism of rapid growth. *Nat. Commun.* **2020**, *11*, 340. [CrossRef] [PubMed]
54. Miao, Y.; Luo, D.; Zhao, T.; Du, H.; Liu, Z.; Xu, Z.; Guo, L.; Chen, C.; Peng, S.; Li, J.X.; et al. Genome sequencing reveals chromosome fusion and extensive expansion of genes related to secondary metabolism in *Artemisia argyi*. *Plant Biotechnol. J.* **2022**, *20*, 1902–1915. [CrossRef] [PubMed]
55. Chen, H.; Guo, M.; Dong, S.; Wu, X.; Zhang, G.; He, L.; Jiao, Y.; Chen, S.; Li, L.; Luo, H. A chromosome-scale genome assemblyof *Artemisia argyi* reveals unbiased subgenome evolution and key contributions of gene duplication to volatile terpenoid diversity. *Plant Commun.* **2023**, *2*, 100516. [CrossRef]
56. Kyriakidou, M.; Tai, H.H.; Anglin, N.L.; Ellis, D.; Stromvik, M.V. Current strategies of polyploid plant genome sequence assembly. *Front. Plant Sci.* **2018**, *9*, 1660. [CrossRef]
57. Xu, C.G.; Tang, T.X.; Chen, R.; Liang, C.H.; Liu, X.Y.; Wu, C.L.; Yang, Y.S.; Yang, D.P.; Wu, H. A comparative study of bioactive secondary metabolite production in diploid and tetraploid *Echinacea purpurea* (L.) Moench. *Plant Cell Tissue Organ Cult.* **2014**, *116*, 323–332. [CrossRef]
58. Cheng, D.; Vrieling, K.; Klinkhamer, P.G. The effect of hybridization on secondary metabolites and herbivore resistance: Implications for the evolution of chemical diversity in plants. *Phytochem. Rev.* **2011**, *10*, 107–117. [CrossRef]
59. Wang, P.; Yu, J.; Jin, S.; Chen, S.; Yue, C.; Wang, W.; Gao, S.; Cao, H.; Zheng, Y.; Gu, M.; et al. Genetic basis of high aroma and stress tolerance in the oolong tea cultivar genome. *Hortic. Res.* **2021**, *8*, 107. [CrossRef]
60. Ivanescu, B.; Burlec, A.F.; Crivoi, F.; Rosu, C.; Corciova, A. Secondary metabolites from *Artemisia* genus as biopesticides and innovative nano-based application strategies. *Molecules* **2021**, *26*, 3061. [CrossRef]