

Brief Report

Genome Assemblies of Two *Ormosia* Species: Gene Duplication Related to Their Evolutionary Adaptation

Pan-Pan Liu ¹, En-Ping Yu ^{2,3,4}, Zong-Jian Tan ¹, Hong-Mei Sun ¹, Wei-Guang Zhu ^{2,3}, Zheng-Feng Wang ^{2,3,*} and Hong-Lin Cao ^{2,3,*}¹ Zhongshan Management Centre of the Natural Protected Area, Zhongshan 528400, China² Guangdong Provincial Key Laboratory of Applied Botany, Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China³ South China National Botanical Garden, Guangzhou 510650, China⁴ University of Chinese Academy of Sciences, Beijing 100049, China

* Correspondence: wzf@scib.ac.cn (Z.-F.W.); caohl@scib.ac.cn (H.-L.C.)

Abstract: *Ormosia* is a genus of the Fabaceae family that shows a distinct evolutionary history due to its typical Asian-American tropical disjunction distribution pattern. However, both its phylogeny and biogeographic mechanisms have not been fully resolved. In addition, *Ormosia* species have great economic and ecological potential in the wood and handicraft (using their attractive seeds) industries, reforestation, and folk medicine (due to their flavonoids, alkaloids, and terpenoids), making them highly valuable in research, especially from a genomic perspective. We report the genome assemblies of two common *Ormosia* species, *Ormosia emarginata* and *Ormosia semicastrata*, in South China, using both long and short sequencing reads. The genome assemblies of *O. emarginata* and *O. semicastrata* comprised 90 contigs with a total length of 1,420,917,605 bp and 63 contigs with a total length of 1,511,766,959 bp, respectively. Benchmarking Universal Single-Copy Orthologs (BUSCO) assessment revealed 97.0% and 98.3% completeness of the *O. emarginata* and *O. semicastrata* assemblies, respectively. The assemblies contain 48,599 and 52,067 protein-coding genes, respectively. Phylogenetic analyses using 1032 single-copy genes with 19 species indicated that they are closely related to *Lupinus albus*. We investigated genes related to plant hormones, signaling, the circadian rhythm, transcription factors, and secondary metabolites derived from the whole genome and tandem and proximal duplications, indicating that these duplications should play important roles in the growth, development, and defense of *Ormosia* species. To our knowledge, our study is the first report on *Ormosia* genome assemblies. This information will facilitate phylogenetic and biogeographic analyses and species breeding in the future.

Keywords: comparative genome; gene family; gene prediction; Fabaceae; tandem and proximal duplications; whole-genome duplication



Citation: Liu, P.-P.; Yu, E.-P.; Tan, Z.-J.; Sun, H.-M.; Zhu, W.-G.; Wang, Z.-F.; Cao, H.-L. Genome Assemblies of Two *Ormosia* Species: Gene Duplication Related to Their Evolutionary Adaptation. *Agronomy* **2023**, *13*, 1757. <https://doi.org/10.3390/agronomy13071757>

Academic Editor: Andras Cseh

Received: 25 April 2023

Revised: 24 June 2023

Accepted: 26 June 2023

Published: 29 June 2023



Copyright: © 2023 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

The genus *Ormosia* within the family Fabaceae contains about 150 species [1]. These species are mostly distributed in tropical regions, mainly in Southeast Asia, northern Australia, Papua New Guinea, Brazil, and the Caribbean [1,2]. Phylogenetic analysis using ITS, *matK*, and *trnL-F* markers indicate the *Ormosia* species are clustered into three clades: the Old World *Ormosia* clade 1, the Old World *Ormosia* clade 2, and the New World *Ormosia* clade [1]. Species of the Old World *Ormosia* clades are distributed from east-central China to southern India, the Solomon Islands, and northeastern Australia. Species of the New World *Ormosia* clade are mainly present on the mainland from southern Mexico to Bolivia and southernmost Brazil. Continental Asia is the origin of *Ormosia*, and the New World distribution might have occurred around the Oligocene or early Miocene [1]. In China, fossil analysis indicates that southern China was the *Ormosia* distribution center, and it

extended northwards during the Miocene due to drastic climate change in East Asia [3]. However, due to complicated morphological variations, the phylogeny and biogeography of *Ormosia* have not yet been fully resolved [1,2,4].

Many *Ormosia* species are high-quality timber species that have great economic benefits in production, and some species are valuable garden or medical trees [5,6]. The other distinct characteristics of *Ormosia* species are their attractive, brightly colored seeds (red or black) that are often used for jewelry and ornamental items.

Ormosia emarginata and *Ormosia semicastrata* are naturally distributed in southern China, but *O. emarginata* extends to Vietnam. *O. semicastrata* can be found in high latitudes to 27°, while the highest latitude for *O. emarginata* is 23° [3]. Both *Ormosia* species are evergreen trees, and *O. emarginata* sometimes occurs as a shrub. Phylogenetic analysis indicates that these two species are in different clades, with *O. emarginata* in the Old World *Ormosia* clade 1 and *O. semicastrata* in the Old World *Ormosia* clade 2 [1]. *Ormosia emarginata* is different from the other *Ormosia* species because its leaf tip forms a clear notch, a feature that underlies its species name (*emarginata*).

In this study, we obtained the whole genome assemblies for *O. emarginata* and *O. semicastrata*. To our knowledge, this is the first report of the genome assemblies of *Ormosia* species. The reported assemblies will provide a valuable resource for future evolutionary and ecological research.

2. Materials and Methods

2.1. Sample Collection and Sequencing

One *O. emarginata* and one *O. semicastrata* individual were collected from Wuguishan, Zhongshan City, China. Their genomic DNA and total RNA were isolated from leaf tissues. Next, long- and short-read whole-genome sequencing (WGS) and short-read RNA sequencing (RNA-seq) libraries were constructed. Long-read WGS sequencing was performed by using the Nanopore PromethION sequencer, while both short-read WGS and RNA-seq were performed by using the MGI DNBSEQ-T7 sequencer under the 150 base pair (bp) paired-end (insert size 300 bp) sequencing strategies. GrandOmics Biosciences (GB, Wuhan, China) completed the sequencing and library construction.

2.2. Data Pre-Processing

Most modern sequencing technologies produce different kinds of errors in the raw sequences, which influence the accuracy of downstream bioinformatics analyses. These low-quality reads then need to be trimmed and/or corrected. After sequencing, short WGS reads of *O. emarginata* and *O. semicastrata* were quality-trimmed using Sickle v1.33 [7], an efficient quality-trimming tool via a sliding window, by removing the reads with a base quality value of <30 and a length of <80 bp. The reads were further error corrected by using RECKONER v1.1 [8], an efficient error corrector using a *k*-spectrum-based algorithm. Using the error-corrected reads, the genome size of *O. emarginata* and *O. semicastrata* was estimated with KmerGenie v1.7044 [9], GenomeScope 2.0 [10], and findGSE [11]. All these programs are *k*-mer-based genome size estimators. The adapters for Nanopore long reads were removed using Porchop 0.2.4 [12].

2.3. Genome Assembly

The genomes of *O. emarginata* and *O. semicastrata* were assembled based on their Nanopore long reads with nextDenovo 2.3.1 [13]. NextDenovo is a string graph-based de novo assembler for noisy long reads. It applies a “correction then assembly” strategy and promises continuous assemblies containing fewer misassemblies. After assembly, Pseudohaploid [14] and Purge_Dups v1.2.6 [15] were applied to examine and remove duplications caused by heterozygosity in the assemblies. Considering the initial assembly still contained many errors, the assemblies were then polished by racon v1.5.0 [16], hapo-G v1.3.2 [17], and polypolish v0.5.0 [18], and possible misassemblies were further corrected using Depthcharge v0.2.0 [19]. Racon and Depthcharge were performed using the Nanopore

long reads of two species, and hapo-G and polypolish were performed using their error-corrected short WGS reads. Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.4.4 [20] with the database eudicots_odb10.2020-09-10 was used to evaluate the quality of each assembly. BUSCO assessments are based on matching the regions in the genome assembly to the lineage-specific sets of near-universal single-copy orthologous genes. Regions that contain potential matches to the orthologue set are classified as “complete” if they are full-length matches, “single-copy” if they are complete matches and only appear once, “fragmented” if 90% matches, and “missing” if there are no matches.

2.4. Repeat Annotation

Repeat sequences (Repeats) are DNA fragments with multiple copies in the genome and are important genetic material in the genomes of almost all species [21,22]. Repeats generally include tandem repeats (TRs, including satellites, minisatellites, and microsatellites) and transposable elements (TEs). TEs are categorized as retrotransposons and DNA transposons. Long-terminal repeat (LTR) retrotransposons and terminal inverted repeat (TIR) are two major components of retrotransposons and DNA transposons, respectively. Repeat sequences of *O. emarginata* and *O. semicastrata* assemblies were identified with EDTA v2.1.0 [21] and RED v2.0 [22], respectively, and their results were then combined by using the “merge” command in bedtools v2.29.2 [23]. Based on the combined repeated sequences, *O. emarginata* and *O. semicastrata* genome assemblies were masked using the “maskfasta” command in bedtools. EDTA is a precise and efficient TE detector, and RED is rapid and accurate in both TRs and TEs identifications.

2.5. Gene Prediction and Annotation

The primary gene sets of *O. emarginata* and *O. semicastrata* were obtained using de novo transcriptome assemblies with TransPi 1.1.0-Rc [24] using their RNA-seq reads. TransPi applies different transcriptome assemblers to produce a combined transcriptome assembly, which greatly improves the representation of the assembly. The over-assembled transcriptome is then reduced to output the results with fewer biases, duplication, and misassemblies using a consensus approach. The TransPi results and protein sequences from *Abrus precatorius*, *Aeschynomene evenia*, *Cicer arietinum*, *Glycine max*, *Lupinus albus*, and *Senna tora* (Table S1, used as protein evidence-based prediction) were then input into the funannotate pipeline v1.8.13 [25] to obtain the final integrated and consensus gene sets, using the commands “funannotate train”, “funannotate predict”, and “funannotate update”. For the “predict” and “update” steps, the parameters “-max_intronlen 100,000”, “-busco_db embryophyte”, and “-organism other” were applied. Funannotate is a user-friendly genome prediction, annotation, and comparison software package. It takes into account repeated elements and includes different evidence (ab-initio predictions, transcript, and protein sequences) to improve the gene predictions.

After gene prediction, the command “funannotate annotate” was used for functional annotation of the genes. The annotation databases used included dbCAN v10.0 [26], eggno-mapper v5.0.2 [27], Gene Ontology (GO) [28,29], Kyoto Encyclopedia of Genes and Genomes (KEGG) [30], InterPro v5.60-92.0 [31], MEROPS v12.0 [32], Pfam v35.0 [33], and UniProt v2022_05 [34].

2.6. Gene Families and Comparative Genomics

A gene family is a group of functionally related genes formed by duplication of a single original gene. The gene families for *O. emarginata* and *O. semicastrata* and 17 other species (Table S2) were identified with OrthoFinder 3.0.0 [35,36] by comparing their predicted protein-coding gene sequences. In this process, only the longest transcript for each gene in each species was used in the subsequent analyses unless otherwise mentioned. The orthologous groups generated by OrthoFinder were regarded as distinct gene families. After the identification of orthologous groups, 1032 single-copy orthologs were selected to generate the species trees by using STAG [37] and STRIDE [38] embedded in OrthoFinder.

Based on the inferred phylogenetic tree, MCMCTree [39] was used to obtain the dated tree, in which 13 species pairs were used as calibration points, and their estimated divergence time was derived from <http://timetree.org/> (accessed on 31 March 2023) (Table S3). In the MCMCTree run, the burn-in period, the sample frequency, and the sample number were 2,000,000, 10, and 4,000,000, respectively. Using the dated tree, CAFE v5 [40] was applied to identify the gene family (i.e., the orthologous gene group) that had potentially undergone expansion or contraction. CAFE is a tool for the statistical analysis of changes in gene family size among species along a phylogenetic tree using a stochastic birth and death process. For significantly expanded and contracted gene families, their corresponding enrichment analysis according to the GO and KEGG databases was conducted by using TBtools v1.115 [41].

2.7. Gene Duplications

Whole Genome Duplications (WGDs) are characterized by large-scale genomic duplication in an organism. Ancient WGD events in *O. emarginata* and *O. semicastrata* were detected using wgd v1.1.2 [42]. Doubletrouble v0.99.1 [43] was further used to examine how many duplications were derived from WGD or other duplications, including tandem duplications (TD), proximal duplications (PD), transposed duplications (TRD), and dispersed duplications (DD) [44]. *Lupinus albus* was used as outgroup species in the doubletrouble analysis. For WGD, TD, and PD genes, enrichment analysis according to the GO and KEGG databases was conducted using TBtools.

3. Results and Discussion

3.1. Genome Sequencing

Sequencing yielded approximately 150.7 Gb of long-WGS reads, 145.7 of Gb short-WGS reads, 32.5 Gb transcriptome reads for *O. emarginata*, and 132.8 Gb of long-WGS reads, 150.7 Gb of short-WGS reads, and 36.7 Gb transcriptome reads for *O. semicastrata*.

3.2. Genome Assembly

The estimated genome sizes varied from 1,049,123,288 to 1,676,161,399 bp for *O. emarginata* and from 1,117,552,851 to 1,436,238,535 bp for *O. semicastrata* under different K-mer sizes (Table S4). The genome assembly based on Nonopore long reads was 1,420,917,605 bp with 90 contigs and N50 of 28,195,512 bp for *O. emarginata*, and 1,511,766,959 bp with 63 contigs and N50 of 48,976,089 bp for *O. semicastrata* (Table 1).

Table 1. Statistics of genome assembly.

<i>Ormosia emarginata</i>		<i>Ormosia semicastrata</i>	
The Length of Sequence (bp)	The Order of Sequence Length	The Length of Sequence (bp)	The Order of Sequence Length
N10 = 81,285,628	L10 = 2	N10 = 89,031,100	L10 = 2
N20 = 63,464,384	L20 = 4	N20 = 79,796,434	L20 = 3
N30 = 43,593,171	L30 = 7	N30 = 73,253,298	L30 = 5
N40 = 37,463,220	L40 = 10	N40 = 56,807,054	L40 = 8
N50 = 28,195,512	L50 = 15	N50 = 48,976,089	L50 = 11
N60 = 25,800,464	L60 = 20	N60 = 45,239,136	L60 = 14
N70 = 20,527,781	L70 = 26	N70 = 31,722,207	L70 = 18
N80 = 13,438,452	L80 = 35	N80 = 22,051,163	L80 = 23
N90 = 7,895,810	L90 = 49	N90 = 12,933,450	L90 = 31
N100 = 173,104	L100 = 90	N100 = 128,272	L100 = 63
Total length	1,420,917,605 bp	1,511,766,959 bp	
Average length	15,787,973.39 bp	23,996,300.94 bp	
Largest length	84,853,091 bp	144,833,628 bp	
Minimum length	173,104 bp	128,272 bp	

BUSCO assessment of the assembled *O. emarginata* genome produced a 97.0% completeness score for the eudicot dataset (2326 core genes), including 2075 (89.2%) complete and single-copy genes, 182 (7.8%) complete and duplicated genes, 9 (0.4%) fragmented genes, and 60 (2.6%) missing genes. Assessment of the *O. semicastrata* genome yielded a 98.3% completeness score, including 90.3% complete and single-copy genes, 8.0% complete and duplicated genes, 0.2% fragmented genes, and 1.5% missing genes.

3.3. Repeat Annotation

Based on the EDTA and RED results, 69.87% (992,753,897 bp) and 62.47% (887,588,204 bp), respectively, of the *O. emarginata* genome assembly were repetitive regions and 71.54% (1,081,509,223) and 64.11% (969,217,860 bp) of the *O. semicastrata* genome assembly were repetitive regions. According to EDTA, the most repetitive sequences were long terminal repeats (LTRs) in both *Ormosia* species. In *O. emarginata*, they accounted for 60.99% (866,687,265 bp) of the genome size, and Gypsy-like LTRs were the most abundant, with 497,532,853 bp (35.01%). In *O. semicastrata*, LTRs accounted for 60.42% (913,370,786 bp) of the genome size, and again Gypsy-like LTRs in it were the most abundant, with 417,795,386 bp (27.64%) (Table S5). By combining the EDTA and RED results, a total of 1,053,199,477 bp (74.12%) of the assembled *O. emarginata* genome and 1,137,510,135 bp (75.24%) of the assembled *O. semicastrata* genome were annotated as repetitive components.

3.4. Gene Prediction and Annotation

Gene predictions via the Funannotate pipeline yielded a total of 46,117 genes coding 48,599 proteins in the *O. emarginata* genome and 49,301 genes coding 52,067 proteins in the *O. semicastrata* genome. Functional annotation of these genes showed that 35,339 (72.71%) and 37,533 (72.09%) of the genes, respectively, were annotated to at least one database from these two species (Table S6).

3.5. Gene Families

Orthofinder identified a total of 37,260 gene families among 641,470 genes from *O. emarginata*, *O. semicastrata*, and the other species (Table S2). For *O. emarginata*, 88.5% (40,829/46,117) of its genes were assigned to 56.6% (21,075/37,260) of the gene families, and 970 gene families composed of 2872 genes were specific to *O. emarginata* (Table S7). Enrichment analysis indicated that these specific genes are mainly functionally related to serine, trehalose, and disaccharide biosynthetic/metabolic processes in the GO biological process category (Table S8) and related to aminoacyl-tRNA biosynthesis, prenyltransferases, and mismatch repair in KEGG (Table S9).

For *O. semicastrata*, 87.2% (42,987/49,301) of its genes were assigned to 58.0% (21,605/37,260) of the gene families, and 1063 gene families composed of 3276 genes were specific to *O. semicastrata* (Table S7). Enrichment analysis indicated that these specific genes in *O. emarginata* are mainly functionally related to vesicle targeting/localization in the GO biological process category (Table S10) and related to terpenoid backbone biosynthesis and membrane trafficking in KEGG (Table S11).

The phylogenetic tree (Figure 1) shows that *O. emarginata* and *O. semicastrata* cluster together and are sisters to *L. albus*. The estimated divergence time between the two species is about 10.42 (95% confidence interval [CI] 3.73511–18.5831) million years ago, and the two species diverged from *L. albus* about 44.46 (95% CI 35.1582–52.6539) million years ago.

For *O. emarginata*, there were 1565 expanded families and 1152 contracted families, of which 24 were significantly expanded and 30 were significantly contracted ($p < 0.05$). Based on GO enrichment analysis, the significantly expanded gene families are related to terpenoid and glutathione biosynthetic/metabolic process, the defense response, and transcription (Table S12); KEGG annotation further indicated their major functions are flavone and flavonol biosynthesis (Table S13). The significantly contracted gene families are related to proteolysis and pyruvate metabolism (Tables S14 and S15).

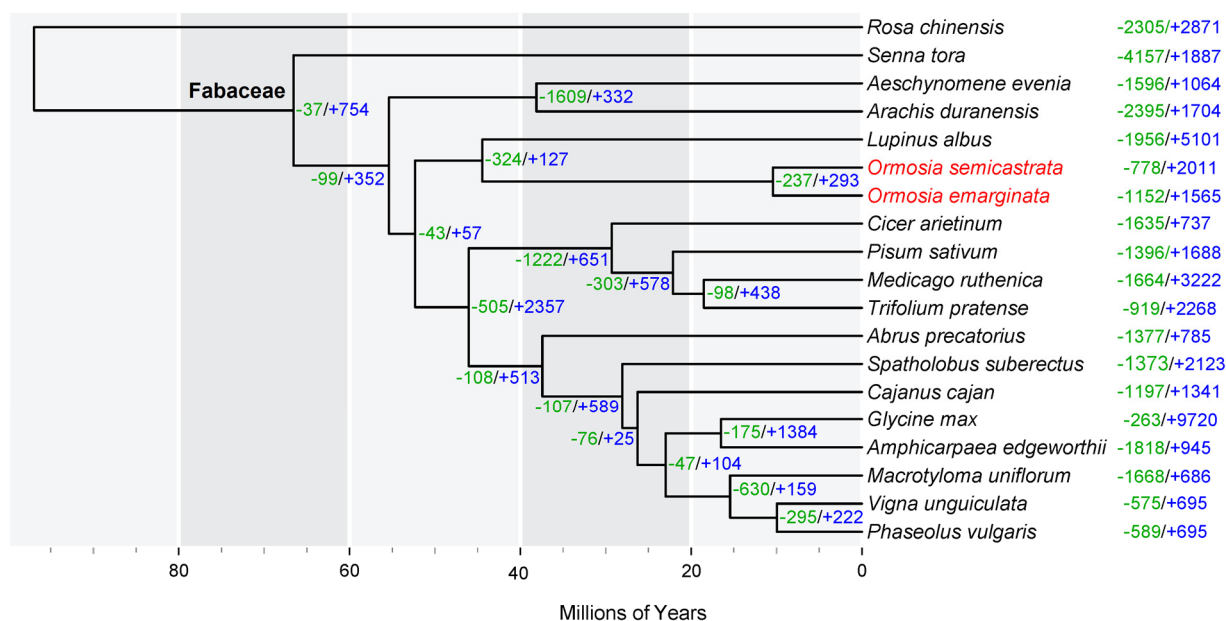


Figure 1. A phylogenetic tree and contracted (–) and expanded (+) gene family of two *Ormosia* species with the other Fabaceae species and outgroup.

For *O. semicastrata*, there were 2011 expanded families and 778 contracted families, of which 53 were significantly expanded and 18 were significantly contracted ($p < 0.05$). Based on GO enrichment, the significantly expanded gene families are related to DNA integration, cell recognition, the defense response, and recognition of pollen (Table S16); KEGG annotation further indicated that their major functions are biosynthesis of various alkaloids, monoterpene biosynthesis, tyrosine metabolism, and fatty acid degradation (Table S17). The significantly contracted gene families are related to oxidoreductase activity and alpha-linolenic acid metabolism (Tables S18 and S19).

Specific and expanded gene enrichment analysis indicated that terpenoids play an important role in the metabolism of *O. emarginata* and *O. semicastrata*. Terpenoids represent a highly diverse class of natural products [45–47]. They are involved in various stages of plant growth and development, but more basically, they serve as a chemical defense in the stressful abiotic and biotic environment of plants [46,48]. Although less studied in *O. emarginata* and *O. semicastrata*, it has been reported that *Ormosia* species are rich in terpenoids as well as alkaloids and flavonoids [49]. We found that the genes related to the latter two components were significantly expanded in *O. semicastrata* and *O. emarginata*, respectively.

The analysis identified 283 gene families that were specific to Fabaceae. In *O. emarginata* and *O. semicastrata*, these families contained a total of 521 and 535 genes, respectively. Enrichment analyses indicated these genes are associated with the regulation of response to alcohol and lipid, zeatin biosynthesis, tryptophan metabolism, ascorbate and aldarate metabolism, porphyrin metabolism, phenylpropanoid biosynthesis, and pentose and glucuronate interconversions in two species (Tables S20–S23). In legumes, alcohol and lipid metabolism play essential roles in nodule formation and development [50,51]. Similarly, by studying 13 legume species, Garg et al. [52] found legume-specific gene families were enriched in genes involved in nodulation and nitrogen fixation, and the others included defense response, TOR signaling, flavonol biosynthesis, calcium ion homeostasis, response to symbiotic bacterium, arbuscular mycorrhizal association, and gravitropism. Our results, therefore, improve the understanding of legume species evolution.

3.6. Gene Duplications

Both *O. emarginata* and *O. semicastrata* underwent the same ancient WGD events (Figure 2), which was consistent with the WGD shared by other legume species [53–55].

Gene duplications using the doubletrouble program revealed 12,489 WGD-type genes, 2410 TD-type genes, 2645 PD-type genes, 35 TRD-type genes, and 16,850 DD-type genes in *O. emarginata*, and 13,063 WGD-type genes, 2521 TD-type genes, 3027 PD-type genes, 49 TRD-type genes, and 17,803 DD-type genes in *O. semicastrata*.

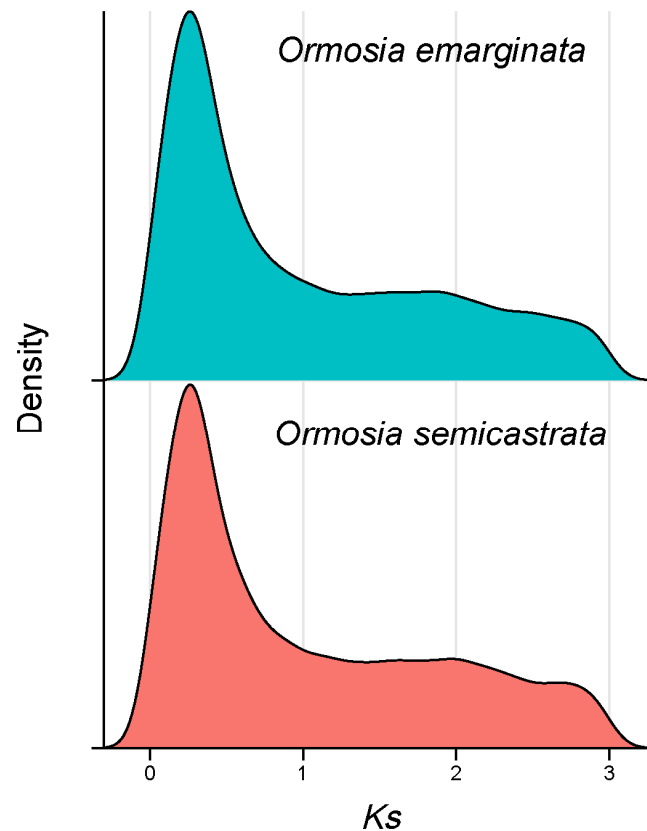


Figure 2. Whole genome duplication analysis comparing the synonymous substitutions per synonymous site (K_s) value distribution for two *Ormosia* species.

WGD, TD, and PD are central duplication processes related to the biotic and abiotic environmental adaptations of plant species [44]. Enrichment analysis indicated duplicated genes due to WGD in *O. emarginata* and *O. semicastrata* are mainly related to multiple aspects of development, growth, and defense—for example, glycosylphosphatidylinositol (GPI)-anchored proteins, plant hormone signal transduction, GTP-binding proteins, signal transduction, the circadian rhythm in plants, and transcription factors (Tables S24–S27). GPI-anchored proteins are important surface proteins that are anchored to the membrane or cell wall. These proteins can take part in many processes, including signaling, immune response, and cell wall development, which are essential for pathogen resistance in plants [56–58]. Plant hormones are messengers and signaling agents that control plant growth, development, and environmental stress responses [59]. They are classified into five major types: abscisic acid, auxins, cytokinins, ethylene, and gibberellins, while salicylic acid, jasmonates, polyamines, strigolactones, brassinosteroids, and nitric oxide (NO) are signaling molecules involved in hormonal responses. These molecules interact [60], and signal transductions among them are often carried out via GTP-binding proteins (G-proteins) [61]. Hormone signaling is also involved in many other processes in plants, such as photo-morphogenesis and carbon metabolism, which are linked to the circadian clock [62,63]. Transcription factors function as transcriptional regulation of gene expression, which is crucial to the plant's response to various abiotic and biotic stresses [64,65]. In *Astragalus sinicus*, functional insight of WGD retained genes showed they were also related to plant hormone signal transduction, accompanied by plant–pathogen interaction [66].

It has been found that hormones are crucial in the process of nodulation [66]. Therefore, WGD shared by legume species could contribute to the hormonal regulation of nodulation. In addition, in *Cyamopsis tetragonoloba*, WGD retained genes were mainly related to the cell wall and carbohydrate/galactose/mannan/monosaccharide/metabolic process [67]. The cell wall-related genes after WGD were enriched in *O. emarginata* (Table S25) but not in *O. semicastrata*. This indicated that the duplicated gene after WGD can be lost or retained for long periods depending on the gene evolution in different species [67].

The TD-related genes in *O. emarginata* are mainly related to phloem development, the defense response, and various (secondary) biosynthetic/metabolic processes (e.g., alkaloids, anthocyanins, butanoate, flavonoids, glutathione, phenylalanine, terpenoids, tyrosine, and zeatin) (Tables S28 and S29). The TD-related genes in *O. semicastrata* are mainly related to the defense response, root development, and various (secondary) biosynthetic/metabolic processes (e.g., alkaloids, brassinosteroids, cyanoamino acid, flavonoids, glucosinolate, glutathione, histidine, linoleic acid, pentose, terpenoids, tryptophan, and zeatin) (Tables S30 and S31). When considering PD duplication, enrichment analyses revealed that these genes are mainly related to the defense response and alkaloid/flavonoid/terpenoid biosynthesis/metabolism in *O. emarginata* and *O. semicastrata* (Tables S32–S35). Alkaloids are nitrogen-containing secondary metabolites. Unlike terpenoids, which are quite common in higher plants, only a few woody trees, mainly tropical species, contain alkaloids [68,69]. Alkaloids are toxic and, therefore, mainly act as defense compounds in plants [70,71]. Flavonoids are low-molecular-weight polyphenolic secondary metabolites in plants [72,73]. They are found in flowers, leaves, and seeds and participate in a variety of biological activities in plant growth, development, and responses to biotic/abiotic stress [73,74]. Flavonoids are also essential molecules that are partly responsible for the coloration in most flowers, fruits, seeds, and vegetative tissues, thus attracting pollinators and seed-dispersing animals [72–74]. Therefore, future studies are needed to explore the association between flavonoids and *Ormosia* seed coat colors.

4. Conclusions

Legumes are important components in natural and human ecosystems. *Ormosia* species are valuable for their timber as well as their secondary metabolites for medicine. Their colorful seeds have been used for ornamental purposes and have a rich cultural and historical significance. However, very few *Ormosia* species have been studied and cultivated. Therefore, our genome assembly and analysis of the genomes of *O. emarginata* and *O. semicastrata* provide a crucial resource to understand the biology, ecology, and adaptations of *Ormosia* and the other legume species. In particular, the assembled genomes revealed that WGD, TD, and PD of genes play a vital role in many biological function processes in *Ormosia* growth, development, and environmental adaptation. This information will facilitate future genomics-assisted breeding and genetic engineering applications in *Ormosia* species.

The assembled genomes of two *Ormosia* species are still fragmented. Further high-quality genome assemblies using Hi-C and other sequencing technologies are needed to generate their chromosome-scale assemblies. Using these assemblies, comparative genome analyses including more model legumes, such as *Medicago truncatula* and *Lotus japonicus*, will be conducted. In particular, the genes related to nodulation, which is a characteristic feature of legumes, will be compared in each legume species to investigate their different nodulation mechanisms.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/agronomy13071757/s1>. Table S1: Protein sequences of species used for gene prediction; Table S2: Species used for comparative genomics; Table S3: Species pairs and their estimated divergence times used for time calibration points to infer time-calibrated phylogeny of two *Ormosia* species; Table S4: Genome size (bp) estimation using different programs; Table S5: Repeat content of assemblies; Table S6: Summary of gene functional annotations of the *Ormosia* genome using different databases; Table S7: Summary of gene families; Table S8: GO enrichment results for *Ormosia emarginata* specific gene families; Table S9: KEGG enrichment results for *Ormosia*

emarginata specific gene families; Table S10: GO enrichment results for *Ormosia semicastrata* specific gene families; Table S11: KEGG enrichment results for *Ormosia semicastrata* specific gene families; Table S12: GO enrichment results for *Ormosia emarginata* significantly expanded gene families; Table S13: KEGG enrichment results for *Ormosia emarginata* significantly expanded gene families; Table S14: GO enrichment results for *Ormosia emarginata* significantly contracted gene families; Table S15: KEGG enrichment results for *Ormosia emarginata* significantly contracted gene families; Table S16: GO enrichment results for *Ormosia semicastrata* significantly expanded gene families; Table S17: KEGG enrichment results for *Ormosia semicastrata* significantly expanded gene families; Table S18: GO enrichment results for *Ormosia semicastrata* significantly contracted gene families; Table S19: KEGG enrichment results for *Ormosia semicastrata* significantly contracted gene families; Table S20: GO enrichment results for *Ormosia emarginata* Fabaceae specific gene families; Table S21: KEGG enrichment results for *Ormosia emarginata* Fabaceae specific gene families; Table S22: GO enrichment results for *Ormosia semicastrata* Fabaceae specific gene families; Table S23: KEGG enrichment results for *Ormosia semicastrata* Fabaceae specific gene families; Table S24: GO enrichment results for *Ormosia emarginata* WGD genes; Table S25: KEGG enrichment results for *Ormosia emarginata* WGD genes; Table S26: GO enrichment results for *Ormosia semicastrata* WGD genes; Table S27: KEGG enrichment results for *Ormosia semicastrata* WGD genes; Table S28: GO enrichment results for *Ormosia emarginata* TD genes; Table S29: KEGG enrichment results for *Ormosia emarginata* TD genes; Table S30: GO enrichment results for *Ormosia semicastrata* TD genes; Table S31: KEGG enrichment results for *Ormosia semicastrata* TD genes; Table S32: GO enrichment results for *Ormosia emarginata* PD genes; Table S33: KEGG enrichment results for *Ormosia emarginata* PD genes; Table S34: GO enrichment results for *Ormosia semicastrata* PD genes; Table S35: KEGG enrichment results for *Ormosia semicastrata* PD genes.

Author Contributions: Conceptualization, H.-L.C. and Z.-F.W.; methodology, H.-L.C. and Z.-F.W.; validation, H.-L.C. and Z.-F.W.; formal analysis, P.-P.L., E.-P.Y., Z.-J.T., H.-M.S., W.-G.Z., Z.-F.W. and H.-L.C.; investigation, P.-P.L., E.-P.Y., Z.-J.T., H.-M.S., W.-G.Z., Z.-F.W. and H.-L.C.; resources, P.-P.L., E.-P.Y., Z.-J.T., H.-M.S., W.-G.Z., Z.-F.W. and H.-L.C.; data curation, H.-L.C. and Z.-F.W.; writing—original draft preparation, P.-P.L., E.-P.Y., Z.-J.T., H.-M.S., W.-G.Z., Z.-F.W. and H.-L.C.; writing—review and editing, P.-P.L., E.-P.Y., Z.-J.T., H.-M.S., W.-G.Z., Z.-F.W. and H.-L.C.; project administration, H.-L.C.; funding acquisition, H.-L.C. All authors have read and agreed to the published version of the manuscript.

Funding: The study is supported by the Project from Financial Funds of Zhongshan City—Preliminary study on conservation ecology of *Ormosia* species in Wuguishan, “Conservation of endemic *Ormosia* species of Guangdong Province and *Ormosia* species center creation” from Forest Bureau of Guangdong Province—Plan on provincial plant ex-situ conservation system, propagation and expansion of national key protected plant in ex-situ conservation areas, and the “One Center and Three Bases” Project for Flora and Fauna Conservation of Guangdong Province.

Data Availability Statement: We deposited the sequenced reads to NCBI Sequence Read Archive under the accession number SRR22805615 for the Nanopore reads, SRR21047276 for the BGISEQ reads, and SRR21705960 for the RNA-seq reads in *Ormosia emarginata*; SRR22805496 for the Nanopore reads, SRR21047305 for the BGISEQ reads, and SRR21705959 for the RNA-seq reads in *Ormosia semicastrata*. The assembled *O. emarginata* and *O. semicastrata* genome assemblies were submitted to GenBank under the accession number JAQQTP000000000 and JAQQTO000000000. Genome assembly and gene annotation data are available at <https://doi.org/10.6084/m9.figshare.21225467.v3> (accessed on 3 April 2023).

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

References

1. Torke, B.M.; Cardoso, D.; Chang, H.; Li, S.-J.; Niu, M.; Pennington, R.T.; Stirton, C.H.; Xu, W.B.; Zartman, C.E.; Chung, K.-F. A dated molecular phylogeny and biogeographical analysis reveals the evolutionary history of the trans-pacifically disjunct tropical tree genus *Ormosia* (Fabaceae). *Mol. Phylogenet. Evol.* **2022**, *166*, 107329. [[CrossRef](#)]
2. Cardoso, D.B.O.S.; Stirton, C.H.; Torke, B.M. Taxonomy of South American *Ormosia* (Leguminosae, Papilionoideae): Recircumscription of *O. costulata*, reinstatement of *O. trifoliolata*, and the new species *O. lewisii* from the Brazilian Atlantic forest. *Syst. Bot.* **2014**, *39*, 1132–1141. [[CrossRef](#)]
3. Li, X.C.; Manchester, S.R.; Xiao, L.; Wang, Q.; Hu, Y.; Sun, B.N. *Ormosia* (Fabaceae: Faboideae) from the Miocene of southeastern China support historical expansion of the tropical genus in East Asia. *Hist. Biol.* **2021**, *33*, 3561–3578. [[CrossRef](#)]
4. Liu, H.; Su, Z.; Yu, S.; Liu, J.; Yin, X.; Zhang, G.; Liu, W.; Li, B. Genome comparison reveals mutation hotspots in the chloroplast genome and phylogenetic relationships of *Ormosia* species. *Biomed. Res. Int.* **2019**, *2019*, 7265030. [[CrossRef](#)]

5. Li, L.; Lei, M.; Wang, H.; Yang, X.; Andargie, M.; Huang, S. First report of dieback caused by *Lasiodiplodia pseudotheobromae* on *Ormosia pinnata* in China. *Plant Dis.* **2020**, *104*, 2551–2555. [CrossRef]
6. Liu, D.; Liu, D.-D.; Ma, L.; Yun, X.-L.; Xiang, Y.; Nie, P.; Zeng, G.-R.; Guo, J.-S. *Ormosia henryi* prain leaf extract alleviates cognitive impairment in chronic unpredictable mild stress mice. *Prog. Biochem. Biophys.* **2020**, *47*, 768–779.
7. Joshi, N.A.; Fass, J.N. Sickle: A Sliding-Window, Adaptive, Quality-Based Trimming Tool for FastQ Files (Version 1.33). Available online: <https://github.com/najoshi/sickle> (accessed on 24 August 2022).
8. Długosz, M.; Deorowicz, S. RECKONER: Read error corrector based on KMC. *Bioinformatics* **2017**, *33*, 1086–1089. [CrossRef]
9. Chikhi, R.; Medvedev, P. Informed and automated *k*-mer size selection for genome assembly. *Bioinformatics* **2014**, *30*, 31–37. [CrossRef]
10. Vurture, G.W.; Sedlazeck, F.J.; Nattestad, M.; Underwood, C.J.; Fang, H.; Gurtowski, J.; Schatz, M.C. GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **2017**, *33*, 2202–2204. [CrossRef] [PubMed]
11. Sun, H.; Ding, J.; Piednoël, M.; Schneeberger, K. findGSE: Estimating genome size variation within human and Arabidopsis using *k*-mer frequencies. *Bioinformatics* **2018**, *34*, 550–557. [CrossRef] [PubMed]
12. Porchop v0.2.4. Available online: <https://github.com/rrwick/Porchop> (accessed on 4 November 2022).
13. NextDenovo v2.3.1. Available online: <https://github.com/Nextomics/NextDenovo> (accessed on 24 January 2023).
14. Pseudohaploid. Available online: <https://github.com/schatzlab/pseudohaploid> (accessed on 26 January 2023).
15. Guan, D.F.; McCarthy, S.A.; Wood, J.; Howe, K.; Wang, Y.D. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **2020**, *36*, 2896–2898. [CrossRef] [PubMed]
16. Vaser, R.; Sović, I.; Nagarajan, N.; Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **2017**, *27*, 737–746. [CrossRef]
17. Aury, J.M.; Istace, B. Hapo-G, haplotype-aware polishing of genome assemblies with accurate reads. *NAR Genom Bioinform.* **2021**, *3*, lqab034. [CrossRef]
18. Wick, R.R.; Holt, K.E. Polypolish: Short-read polishing of long-read bacterial genome assemblies. *PLoS Comput. Biol.* **2022**, *18*, e1009802. [CrossRef]
19. Depthcharge v0.2.0. Available online: <https://github.com/slimsuite/depthcharge> (accessed on 30 January 2023).
20. Manni, M.; Berkeley, M.R.; Seppey, M.; Zdobnov, E.M. BUSCO: Assessing genomic data quality and beyond. *Curr. Protoc.* **2021**, *1*, e323. [CrossRef]
21. Ou, S.; Su, W.; Liao, Y.; Chougule, K.; Agda, J.R.A.; Hellinga, A.J.; Lugo, C.S.B.; Elliott, T.A.; Ware, D.; Peterson, T.; et al. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biol.* **2019**, *20*, 275. [CrossRef] [PubMed]
22. Girgis, H.Z. Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinform.* **2015**, *16*, 227. [CrossRef]
23. Quinlan, A.R.; Hall, I.M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef]
24. Rivera-Vicéns, R.E.; Garcia-Escudero, C.A.; Conci, N.; Eitel, M.; Wörheide, G. TransPi—a comprehensive TRanscriptome ANalysis Pipeline for de novo transcriptome assembly. *Mol. Ecol. Resour.* **2022**, *22*, 2070–2086. [CrossRef] [PubMed]
25. Funannotate v1.8.13. Available online: <https://github.com/nextgenusfs/funannotate> (accessed on 12 March 2023).
26. Zhang, H.; Tanner, Y.; Huang, L.; Entwistle, S. dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* **2018**, *46*, W95–W101. [CrossRef]
27. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Damian, P.C.; Szklarczyk, D.; Jensen, L.J. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [CrossRef]
28. The Gene Ontology Consortium. The gene ontology resource: 20 years and still GOing strong. *Nucleic Acids Res.* **2019**, *47*, D330–D338. [CrossRef] [PubMed]
29. 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. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef] [PubMed]
30. Kanehisa, M.; Soto, 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] [PubMed]
31. Mitchell, A.L.; Attwood, T.K.; Babbitt, P.C.; Blum, M.; Bork, P.; Bridge, A.; Brown, S.D.; Chang, H.Y.; El-Gebali, S.; Fraser, M.I.; et al. InterPro in 2019: Improving coverage, classification and access to protein sequence annotations. *Nucleic Acids Res.* **2019**, *47*, D351–D360. [CrossRef]
32. Rawlings, N.D.; Barrett, A.J.; Thomas, P.D.; Huang, X.S.; Bateman, A.; Finn, R.D. The merops database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Res.* **2018**, *46*, D624–D632. [CrossRef]
33. 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]
34. The UniProt Consortium. UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **2019**, *47*, D506–D515. [CrossRef] [PubMed]
35. Emms, D.M.; Kelly, S. OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **2015**, *16*, 157. [CrossRef]

36. Emms, D.M.; Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **2019**, *20*, 238. [CrossRef]
37. Emms, D.M.; Kelly, S. STAG: Species tree inference from all genes. *bioRxiv* **2018**. Available online: <https://www.biorxiv.org/content/10.1101/267914v1.abstract> (accessed on 31 March 2023).
38. Emms, D.M.; Kelly, S. STRIDE: Species tree root inference from gene duplication events. *Mol. Biol. Evol.* **2017**, *34*, 3267–3278. [CrossRef]
39. dos Reis, M.; Zhu, T.; Yang, Z. The impact of the rate prior on Bayesian estimation of divergence times with multiple Loci. *System Biol.* **2014**, *63*, 555–565. [CrossRef] [PubMed]
40. Han, M.V.; Thomas, G.W.C.; Jose, L.M.; Hahn, M.W. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using cafe 3. *Mol. Biol. Evol.* **2013**, *30*, 1987–1997. [CrossRef]
41. Chen, C.J.; Chen, H.; Zhang, Y.; Thomas, H.R.; Frank, M.H.; He, Y.H.; Xia, R. TBtools- an integrative toolkit developed for interactive analyses of big biological data. *Mol. Plant.* **2020**, *13*, 1194–1202. [CrossRef] [PubMed]
42. Zwaenepoel, A.; de Peer, Y.V. Wgd-simple command line tools for the analysis of ancient whole-genome duplications. *Bioinformatics* **2019**, *35*, 2153–2155. [CrossRef]
43. Almeida-Silva, F.; Van de Peer, Y. Doubletrouble: Identification and Classification of Duplicated Genes. R package Version 0.99.1. 2022. Available online: <https://github.com/almeidasilvaf/doubletrouble> (accessed on 28 March 2023).
44. Qiao, X.; Li, Q.H.; 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] [PubMed]
45. Chen, F.; Tholl, D.; Bohlmann, J.; Pichersky, E. The family of terpene synthases in plants: A mid-size family of genes for specialized metabolism that is highly diversified throughout the kingdom. *Plant J.* **2011**, *66*, 212–229. [CrossRef] [PubMed]
46. Pichersky, E.; Raguso, R.A. Why do plants produce so many terpenoid compounds? *New Phytol.* **2016**, *220*, 655–658. [CrossRef]
47. Jiang, S.Y.; Jin, J.J.; Sarojam, R.; Ramachandran, S. A comprehensive survey on the terpene synthase gene family provides new insight into its evolutionary patterns. *Genome Biol. Evol.* **2019**, *11*, 2078–2098. [CrossRef]
48. Chen, A.-X.; Lou, Y.-G.; Mao, Y.-B.; Lu, S.; Wang, L.-J.; Chen, X.-Y. Plant terpenoids: Biosynthesis and ecological functions. *J. Integr. Plant Biol.* **2007**, *49*, 179–186. [CrossRef]
49. Zhang, L.-J.; Zhou, W.-J.; Ni, L.; Huang, M.-Q.; Zhang, X.-Q.; Xu, H.-Y. A review on chemical constituents and pharmacological activities of *Ormosia*. *Chin. Tradit. Herb. Drugs* **2021**, *52*, 4433–4442.
50. Zhang, G.; Ahmad, M.Z.; Chen, B.; Manan, S.; Zhang, Y.; Jin, H.; Wang, X.; Zhao, J. Lipidomic and transcriptomic profiling of developing nodules reveals the essential roles of active glycolysis and fatty acid and membrane lipid biosynthesis in soybean nodulation. *Plant J.* **2020**, *103*, 1351–1371. [CrossRef] [PubMed]
51. Piya, S.; Pantalone, V.; Zadegan, S.B.; Shipp, S.; Lakhssassi, N.; Knizia, D.; Krishnan, H.B.; Meksem, K.; Hewezi, T. Soybean gene co-expression network analysis identifies two co-regulated gene modules associated with nodule formation and development. *Mol. Plant Pathol.* **2023**, *24*, 628–636. [CrossRef] [PubMed]
52. Garg, V.; Dudchenko, O.; Wang, J.; Khan, A.W.; Gupta, S.; Han, P.K.K.; Saxena, R.K.; Kale, S.M.; Pham, M.; Yu, J.; et al. Chromosome-length genome assemblies of six legume species provide insights into genome organization; evolution; and agronomic traits for crop improvement. *J. Adv. Res.* **2022**, *42*, 315–329. [CrossRef]
53. Kreplak, J.; Madoui, M.A.; Cápál, P.; Novák, P.; Labadie, K.; Aubert, G.; Bayer, P.E.; Gali, K.K.; Syme, R.A.; Main, D.; et al. A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **2019**, *51*, 1411–1422. [CrossRef] [PubMed]
54. Liu, Y.; Zhang, X.; Han, K.; Li, R.; Xu, G.; Han, Y.; Cui, F.; Fan, S.; Seim, I.; Fan, G.; et al. Insights into amphicarpy from the compact genome of the legume *Amphicarpaea edgeworthii*. *Plant Biotechnol. J.* **2020**, *19*, 952–965. [CrossRef]
55. Chang, D.; Gao, S.; Zhou, G.; Deng, S.; Jia, J.; Wang, E.; Cao, W. The chromosome-level genome assembly of *Astragalus sinicus* and comparative genomic analyses provide new resources and insights for understanding legume-rhizobial interactions. *Plant Comm.* **2022**, *3*, 100263. [CrossRef] [PubMed]
56. Rittenour, W.R.; Harris, S.D. Glycosylphosphatidylinositol-anchored proteins in *Fusarium graminearum*: Inventory, variability, and virulence. *PLoS ONE* **2013**, *8*, e81603. [CrossRef] [PubMed]
57. Oliveira-Garcia, E.; Deising, H.B. The glycosylphosphatidylinositol anchor biosynthesis genes *GPI12*, *GAA1*, and *GPI8* are essential for cell-wall integrity and pathogenicity of the maize anthracnose fungus *Colletotrichum graminicola*. *Mol. Plant Microbe Interact.* **2016**, *29*, 889–901. [CrossRef]
58. Mei, J.; Ning, N.; Wu, H.; Chen, X.; Li, Z.; Liu, W. Glycosylphosphatidylinositol anchor biosynthesis pathway-related protein *GPI7* is required for the vegetative growth and pathogenicity of *Colletotrichum graminicola*. *Int. J. Mol. Sci.* **2022**, *23*, 2985. [CrossRef]
59. Waadt, R.; Sella, C.A.; Hsu, P.K.; Takahashi, Y.; Munemasa, S.; Schroeder, J.I. Plant hormone regulation of abiotic stress responses. *Nat. Rev. Mol. Cell Biol.* **2022**, *23*, 680–694. [CrossRef] [PubMed]
60. McSteen, P.; Zhao, Y. Plant hormones and signaling: Common themes and new developments. *Dev. Cell* **2008**, *14*, 467–473. [CrossRef] [PubMed]
61. Ku, Y.-S.; Sintaha, M.; Cheung, M.-Y.; Lam, H.-M. Plant hormone signaling crosstalks between biotic and abiotic stress responses. *Int. J. Mol. Sci.* **2018**, *19*, 3206. [CrossRef] [PubMed]
62. Bhattacharya, A.; Khanale, V.; Char, B. Plant circadian rhythm in stress signaling. *Ind. J. Plant Physiol.* **2017**, *22*, 147–155. [CrossRef]
63. Srivastava, D.; Shamim, M.; Kumar, M.; Mishra, A.; Maurya, R.; Sharma, D.; Pandey, P.; Singh, K.N. Role of circadian rhythm in plant system: An update from development to stress response. *Environ. Exp. Bot.* **2019**, *162*, 256–271. [CrossRef]

64. Singh, K.B.; Foley, R.C.; Oñate-Sánchez, L. Transcription factors in plant defense and stress responses. *Curr. Opin. Plant Bio.* **2002**, *5*, 430–436. [[CrossRef](#)]
65. Strader, L.; Weijers, D.; Wagner, D. Plant transcription factors—being in the right place with the right company. *Curr. Opin. Plant Bio.* **2022**, *65*, 102136. [[CrossRef](#)]
66. Gaikwad, K.; Ramakrishna, G.; Srivastava, H.; Saxena, S.; Kaila, T.; Tyagi, A.; Sharma, P.; Sharma, S.; Sharma, R.; Mahla, H.R.; et al. The chromosome-scale genome assembly of cluster bean provides molecular insight into edible gum (galactomannan) biosynthesis family genes. *Sci. Rep.* **2023**, *13*, 9941. [[CrossRef](#)]
67. Shi, T.; Rahmani, R.S.; Gugger, P.F.; Wang, M.; Li, H.; Zhang, Y.; Li, Z.; Wang, Q.; de Peer, Y.V.; Marchal, K.; et al. Distinct expression and methylation patterns for genes with different fates following a single whole-genome duplication in flowering plants. *Mol. Biol. Evol.* **2020**, *37*, 2394–2413. [[CrossRef](#)]
68. Singh, S.; Pathak, N.; Fatima, E.; Negi, A.S. Plant isoquinoline alkaloids: Advances in the chemistry and biology of berberine. *Eur. J. Med. Chem.* **2021**, *226*, 113839. [[CrossRef](#)] [[PubMed](#)]
69. Carrera, J.C.; Ucella-Filho, J.G.M.; de Andrade, C.M.L.; Stein, V.C.; Mori, F.A. Research, inventiveness and biotechnological advances in the production of value-added alkaloids occurring in tropical species. *S. Afr. J. Bot.* **2022**, *150*, 1122–1137. [[CrossRef](#)]
70. Matsuura, H.N.; Fett-Neto, A.G. Plant Alkaloids: Main Features, Toxicity, and Mechanisms of Action. In *Plant Toxins; Toxinology; Gopalakrishnakone, P., Carlini, C., Ligabue-Braun, R., Eds.*; Springer: Dordrecht, The Netherlands, 2015; pp. 1–15.
71. Ali, A.H.; Abdelrahman, M.; El-Sayed, M.A. Alkaloid Role in Plant Defense Response to Growth and Stress. In *Bioactive Molecules in Plant Defense; Jogaiah, S., Abdelrahman, M., Eds.*; Springer: Cham, Germany, 2019; pp. 145–158.
72. Dias, M.C.; Pinto, D.C.G.A.; Silva, A.M.S. Plant flavonoids: Chemical characteristics and biological activity. *Molecules* **2021**, *26*, 5377. [[CrossRef](#)] [[PubMed](#)]
73. Shen, N.; Wang, T.; Gan, Q.; Liu, S.; Wang, L.; Jin, B. Plant flavonoids: Classification, distribution, biosynthesis, and antioxidant activity. *Food Chem.* **2022**, *383*, 132531. [[CrossRef](#)]
74. Wang, L.; Chen, M.; Lam, P.Y.; Dini-Andreote, F.; Dai, L.; Wei, Z. Multifaceted roles of flavonoids mediating plant-microbe interactions. *Microbiome* **2022**, *10*, 233. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.