



Article

Dating Whole Genome Duplication in *Ceratopteris thalictroides* and Potential Adaptive Values of Retained Gene Duplicates

Rui Zhang^{1,2,†}, Fa-Guo Wang^{3,†}, Jiao Zhang^{1,2}, Hui Shang^{1,2}, Li Liu^{1,2}, Hao Wang^{1,2}, Guo-Hua Zhao^{1,2}, Hui Shen^{1,2} and Yue-Hong Yan^{1,2,*}

- ¹ Shanghai Chenshan Plant Science Research Center, Shanghai Chenshan Botanical Garden, Chinese Academy of Sciences, Shanghai 201602, China; zhangrui@csnbgsh.cn (R.Z.); zhangjiao@sibs.ac.cn (J.Z.); sh007-008@163.com (H.S.); 15609616871@163.com (L.L.); 1000441464@smail.shnu.edu.cn (H.W.); zhaoguohua1990@163.com (G.-H.Z.); shenhui@csnbgsh.cn (H.S.)
- ² Eastern China Conservation Center for Wild Endangered Plant Resources, Shanghai 201602, China
- ³ Key Laboratory of Plant Resources Conservation and Sustainable Utilization, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou 510650, China; wangfg@scib.ac.cn
- * Correspondence: yhyan@sibs.ac.cn; Tel.: +86-21-37792288-903
- † These authors are contributed equally.

Received: 15 March 2019; Accepted: 16 April 2019; Published: 19 April 2019



Abstract: Whole-genome duplications (WGDs) are widespread in plants and frequently coincide with global climatic change events, such as the Cretaceous–Tertiary (KT) extinction event approximately 65 million years ago (mya). Ferns have larger genomes and higher chromosome numbers than seed plants, which likely resulted from multiple rounds of polyploidy. Here, we use diploid and triploid material from a model fern species, *Ceratopteris thalictroides*, for the detection of WGDs. High-quality RNA-seq data was used to infer the number of synonymous substitutions per synonymous site (K_s) between paralogs; K_s age distribution and absolute dating approach were used to determine the age of WGD events. Evidence of an ancient WGD event with a K_s peak value of approximately 1.2 was obtained for both samples; however, the K_s frequency distributions varied significantly. Importantly, we dated the WGD event at 51–53 mya, which coincides with the Paleocene-Eocene Thermal Maximum (PETM), when the Earth became warmer and wetter than any other period during the Cenozoic. Duplicate genes were preferentially retained for specific functions, such as environment response, further support that the duplicates may have promoted quick adaption to environmental changes and potentially resulted in evolutionary success, especially for pantropical species, such as *C. thalictroides*, which exhibits higher temperature tolerance.

Keywords: whole genome duplication; *Ceratopteris thalictroides*; synonymous substitutions; transcriptome; evolution

1. Introduction

Whole-genome duplication (WGD), or polyploidy, has long been considered as an important evolutionary force and often drives plant speciation [1–5]. With more than 300,000 living species, angiosperms are currently the largest group of land plants, and most flowering plants have experienced multiple rounds of WGD. Even the small genome of *Arabidopsis thaliana* has undergone two recent WGDs and one whole genome triplication during the course of its evolution [3,6,7]. Two WGD events (ρ and δ) are indicated to have occurred early in the monocot lineage after its divergence from the eudicot clade [2,3]. Furthermore, substantial evidence shows that ancient WGDs occurred in the common ancestor of extant seed plants, ferns, and mosses [3,5,8,9].

Massive WGDs in flowering plants have occurred at specific times during extreme environmental changes or extinction periods, including the Cretaceous–Tertiary (KT) boundary, and Paleocene-Eocene Thermal Maximum (PETM) [5,10,11]. WGDs are thought to be related to dramatic global climate changes and unstable environments [5,10–12]. Polyploidy could accelerate adaptation to dramatically changing environments via genetic innovations or heterotic effects, as well as mutational robustness, subfunctionalization, and changed modes of reproduction [10,11,13]. Interestingly, gene retention after WGDs is not random; a general pattern has been observed indicating that regulators and signal transducers are preferentially retained in vertebrates and flowering plants [3,14,15]. Another type of biased retention following WGDs are involved in response to biotic and abiotic stress in flowering plants and are important for local adaptation [5,15]. Most *MADS-box* family genes were retained after WGDs, especially *MIKC*-type genes [13,16]. Biased gene retention is considered to have an important impact on innovation and diversification that might contribute to long-term survival [5]. However, little is known about the occurrence of WGDs for nonflowering plants.

Ferns (Monilophyta) constitute the second largest group of vascular plants and an early diverged lineage of land plants; they possess a higher frequency of polyploidy than seed plants. Previous studies have reported that the frequency of polyploidy in ferns varies from 31%–95%, while that in flowering plants varies by 15%–70% [17,18]. With a high base chromosome number, ferns are prone to polyploidization, which might derive from ancient polyploidization [17,19]. The average chromosome number in homosporous ferns ($n = 57.0$) greatly exceeds that in angiosperms ($n = 15.99$) and heterosporous ferns ($n = 13.62$) [20]. Moreover, polyploid species are commonly more prevalent in nature than initially suspected [5]. Several WGDs in the ancestor of extant ferns were detected based on chromosome counts analysis [21,22]. Li et al. (2018) identified two WGDs using phylogenomic analyses and *Ks*-based age distribution: one *Azolla*-specific and one at the base of core leptosporangiates in *Azolla filiculoides* and *Salvinia cucullata* [23]. WGDs were also found in *Equisetum giganteum* and *Ceratopteris richardii* [9,24]. However, little is known about the whole genome sequences of ferns.

Recent improvement in technology, such as transcriptome sequencing, has allowed for a relatively convenient and efficient alternative to evaluate paleopolyploid events because a large number of sequences can be obtained at a low cost [24]. To determine polyploidy events, one such method requires determining the number of synonymous substitutions per synonymous site (*Ks*) between paralogs, which has been widely performed in lots of species [6,9,12]. Without changes in protein sequences, synonymous substitutions are considered to be putatively neutral and accumulate changes at a constant rate. Duplicated (paralogous) pairs in a genome can be sorted in order and time as substitutions accumulate over time by plotting the *Ks* age distribution; this data can be used to detect whether paleopolyploidizations have occurred in plants [6,9]. Most of these duplicates include small-scale duplication and recent neopolyploidizations that were lost frequently that result in an exponential decrease and create an initial peak in the young age classes, and a long and flat tail is expected for older duplicates over time due to fewer genes retained, resulting in an L-shaped pattern [9,25]. However, large-scale duplication events, such as WGDs, are expected to show a secondary peak superimposed upon this L-shaped distribution because of the burst of large-scale duplicate genes at about the same time [6,24]. A WGD is predicted if a clear peak is present in the age distribution of paralogous pairs. To obtain more insight into fern evolution, it is important to examine the genes of fern species to see if there are any signals that indicate multiple genome duplications (e.g., multiple *Ks* peaks), determine the time at which WGDs occurred, and identify whether gene retention is biased towards specific functional classes.

Ceratopteris is a pantropical genus of annual ferns. It is commonly used as a model for ferns because it is easy to culture and produces abundant spores. It also has an independent sporophyte, gametophyte generation, and short sexual cycle [26,27]. Here, we used two samples of *C. thalictroides* in previous published studies [28,29], performed chromosome counts on them, then detected WGD events using high-quality transcriptome data via *Ks* analysis. Additionally, we dated the WGD event using both *Ks* distribution peaks and the absolute dating approach. Furthermore, annotation of the

biased gene retention following the WGD events was also analyzed. Our results will help improve the understanding of the genome history and evolutionary impact of WGD events in ferns.

2. Results

2.1. Chromosome Counting

As the genetic variation in a single sample fails to adequately represent the variation present within a species, we performed a cytological study on the root tips of two collected sporophytes to reflect their genetic background. The *C. thalictroides* sample reported by Shen et al. (2018) was identified as a diploid with a chromosome number of $2n = 78$ (Figure 1A,B) [29]. The chromosome number of *C. thalictroides* published by Zhang et al. (2016) is $2n = 117$ (Figure 1C,D), thus it is a possible triploid [28].

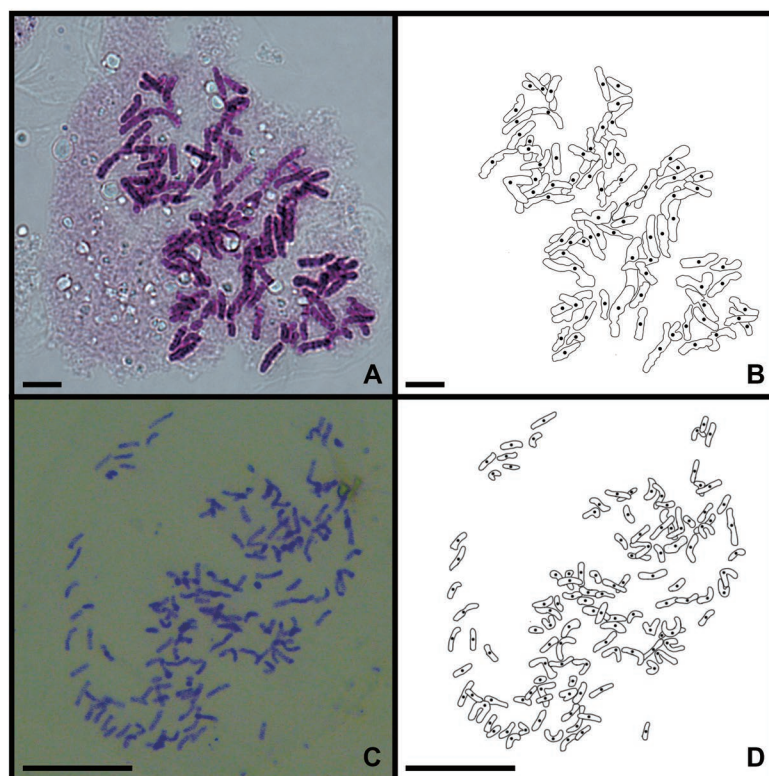


Figure 1. Chromosomes of *C. thalictroides* in mitotic root-tip cells (scale bars=20 μm). (A) Metaphase chromosome of the diploid, $2n = 78$; (B) Lined drawing of Figure A; (C) Metaphase chromosome of the triploid, $2n = 117$; (D) Lined drawing of Figure C.

2.2. Benchmarking Universal Single Copy Orthologs (BUSCO) Analysis

A total of 74,728 and 69,929 contigs with an N50 of 1,610 bp and 787 bp, respectively, were obtained in the diploid and the triploid transcriptomes of *C. thalictroides*. A total of 83,202 and 60,823 unigenes from the diploid and the triploid transcriptomes of *C. thalictroides*, respectively, were then used to predict open reading frames with details provided in Table 1 [28,29].

We performed a BUSCO analysis to assess the level of completeness of the transcriptome assembly (Table 2) using a plant species database containing 1440 ortholog groups. We were able to identify 63.7% of complete single-copy genes and 5.9% partial sequences in the diploid. In addition, 46.5% of the BUSCO orthologs (12.2% of which were fragmented genes) were identified in the triploid. These results indicate that the data were of a high quality and can be used for subsequent analyses.

Table 1. A summary of the sequencing and assembly for diploid and triploid samples of *C. thalictroides* and seven other fern species.

Species	Total Reads (Clean)	Number of Contigs	Total Number of Unigenes	N50 (bp)	Mean Length (bp)
Triploid ^a	35,528,634	69,929	60,823	787	576.00
Diploid ^b	31,741,082	74,728	83,202	1610	912.26
<i>Goniophlebium niponicum</i> ^b	38,786,214	54,152	58,494	1663	951.92
<i>Woodwardia prolifera</i> ^b	40,967,322	69,931	74,564	1557	859.72
<i>Dennstaedtia pilosella</i> ^b	45,618,446	84,813	89,185	1582	831.56
<i>Cheilanthes chusana</i> ^b	51,851,066	49,449	52,782	1727	1012.63
<i>Acrostichum aureum</i> ^b	43,422,574	46,189	50,594	1729	1043.2
<i>Osmolindsaea odorata</i> ^b	46,808,646	113,778	130,549	1521	845.96
<i>Alsophila podophylla</i> ^b	48,768,608	66,254	72,404	1580	904.62

Note: ^a refer to the triploid *C. thalictroides* [28], ^b refer to the diploid *C. thalictroides* and the other seven fern species [29].

Table 2. BUSCO results (genome completeness) for diploid and triploid samples of *C. thalictroides*.

Species	BUSCO Notation Assessment Results
Diploid	C: 63.7% [S:39.7%, D:24%], F:5.9%, M: 30.4%, n: 1440
Triploid	C: 34.3% [S:30.5%, D:3.8%], F:12.2%, M: 53.5%, n: 1440

BUSCO was used to assess the transcriptome data quality with 1440 conservative orthologs in plant species as reference. Abbreviation: C—Complete Single-Copy BUSCOs; S—Complete and Single-Copy BUSCOs; D—Complete Duplicated BUSCOs; F—Fragmented BUSCOs; M—Missing BUSCOs; N—Total BUSCO groups searched.

2.3. Identification of Pairs of Paralogous and WGD Events Estimated on the Basis of *Ks* Age Distributions

In this study, the distributions of *Ks* within paralogous pairs were determined to detect potential WGD events (Figure 2 and Table 3). The difference in *Ks* frequency distributions between the diploid and the triploid was significant (Wilcoxon matched pairs test; $p = 0.035$). A total of 8364 paralogous pairs were identified in the diploid *C. thalictroides*, whereas 3088 paralogous pairs were identified in the triploid sample. We used Gaussian mixture models to confirm the WGD signature. We observed seven and five peaks in the diploid and the triploid, respectively. A significant *Ks* peak near 1.2 was detected by the fitted Gaussian mixture models for both samples, and several other smaller peaks were also fitted. We only focused on *Ks* values less than 2 because higher *Ks* values are uncertain due to *Ks* saturation and stochastic effects, which might provide misleading data for the mixture models. Thus, one ancient WGD event with a median *Ks* peak at approximately 1.2 occurred in the diploid and triploid *C. thalictroides* samples, indicating that the large peak present in *C. thalictroides* was a true WGD rather than stochastic variation.

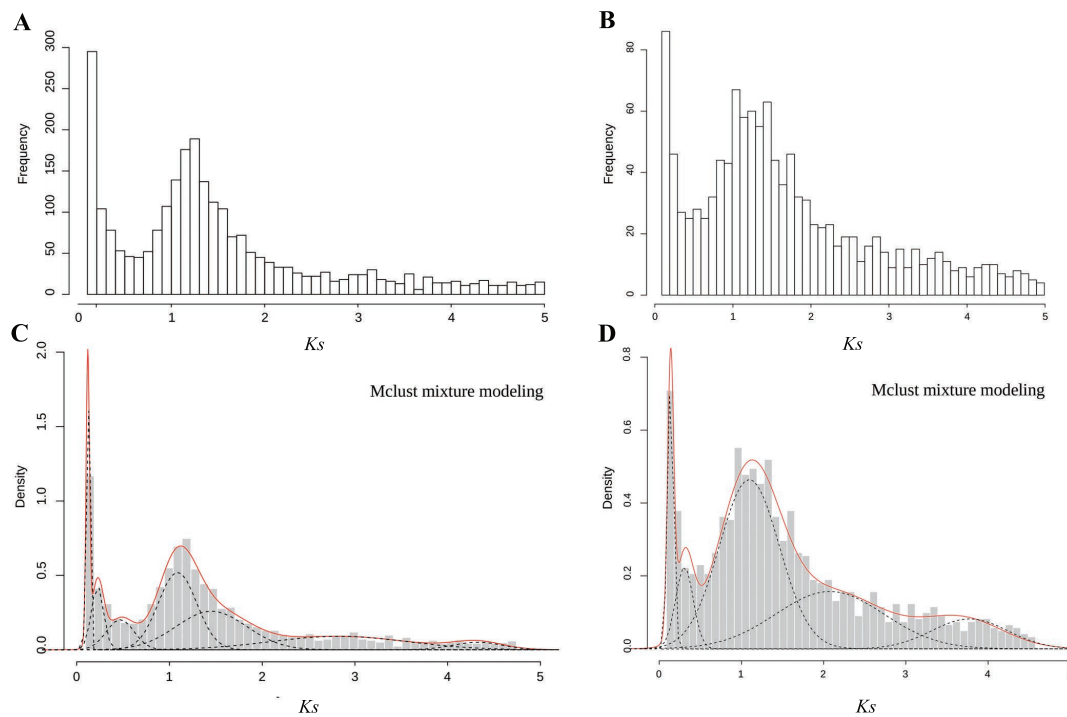


Figure 2. Frequency distributions of K_s values based on paralogous pairs of *C. thalictroides*. (A,C) Distributions of K_s values pairs of the diploid and triploid within 5. The x-axis represents the synonymous substitutions with a K_s cutoff of five in bins of 0.1, and the y-axis shows the number of retained duplicated paralogous gene pairs. (B,D) Mclust Gaussian mixture model analysis of (A,C), respectively. Optimal number of log-normal components overlaid on K_s distributions. The red line shows the sum of components.

Table 3. Mixture modeling of the age distribution of *C. thalictroides* presented in Figure 2.

No. of Duplicates	No. of Components	Bayesian Information Criterion	Mixture Means (K_s)	Variance (K_s)	Proportion
Diploid					
8364	7	5998.011	0.128	0.0004	0.091
8364	7	5998.011	0.238	0.0048	0.074
8364	7	5998.011	0.499	0.0248	0.079
8364	7	5998.011	1.148	0.0461	0.278
8364	7	5998.011	1.526	0.1619	0.265
8364	7	5998.011	2.989	0.6015	0.176
8364	7	5998.011	4.561	0.0790	0.036
Triploid					
3088	5	3380.834	0.154	0.0016	0.075
3088	5	3380.834	0.338	0.0135	0.066
3088	5	3380.834	1.199	0.1591	0.464
3088	5	3380.834	2.268	0.5133	0.282
3088	5	3380.834	4.019	0.2962	0.112

2.4. Functional Classifications of Retained Duplicates after WGD Events Revealed That Retained Genes Are Biased Rather Than Random

A gene ontology (GO) enrichment analysis was performed to explore the potential functions of retained duplicates following WGDs in the diploid with a higher quality for this study. The focus of the analysis was on the WGD event shared by the diploid and the triploid, with a K_s peak at approximately 1.2. The results showed that several GO terms were preferentially retained. A total of 111 GO terms were found to be significantly enriched (Table S1). The following GO terms were found to be enriched: GO terms of type I genes that have major contributions to biological processes (BP), especially

genes related to response to stimulus (response to bacterium, “GO:0009617”; cellular response to carbohydrate stimulus, “GO:0071322”; heat acclimation, “GO:0010286”; and cellular response to salt stress, “GO:0071472”) (Figure S1); GO terms of genes involved in signaling (signaling, “GO:0023052”; small GTPase-mediated signal transduction, “GO:0007264”; and sugar mediated signaling pathway, “GO:0010182”); and GO categories corresponding to various aspects of regulation (protein transport, “GO:0015031”; magnesium ion transport, “GO:0015693”; post-translational protein modification, “GO:0043687”; and protein amino acid autophosphorylation “GO:0046777”). Type II genes were related to cellular components (CC), such as cortical microtubules (GO:0055028) and Golgi apparatus (GO:0005794), and type III genes had specific functional roles that are associated with kinase activity, protein binding, transporter activity, and catalytic activity.

2.5. Dating the WGD Event in *C. thalictroides* Using *Ks* Distribution Peaks and Absolute Dating

We determined the age of the WGD events using both *Ks* distribution peaks and the absolute dating approach. Without changes of protein sequences, synonymous substitutions are considered to be putatively neutral and accumulate changes at a constant rate, which can be used to infer the age of WGDs. When a plant average *Ks*/year rate of 6.1×10^{-9} was used, the age of the WGD event with a peak value of approximately 1.2 was dated at 93–95 mya ($K_s = 1.13$ – 1.16) in the diploid and 96–101 mya ($K_s = 1.17$ – 1.23) in the triploid, with a confidence interval of 95%.

The *Ks* distribution peaks method can introduce inaccuracies because it relies on the assumption of a strict molecular clock and synonymous substitution rates, and synonymous substitution rates are usually variable across different species. To improve the accuracy of our results, we further performed absolute dating of the diploid via phylogenomic analysis to infer the age of the WGDs. Using this method, we obtained an absolute age distribution with a clear peak at 52 ± 1 mya with a confidence interval of 95% (Figure 3). We can also obtain the silent substitution rate r of 11.04×10^{-9} if combining *Ks* and absolute ages of *C. thalictroides*. As mentioned above, the evolutionary rate of *Ceratopteris* should be consistent, so we think the modified *Ks*/year rate is more accurate than a plant average *Ks*/year rate of 6.1×10^{-9} , which could be applicable in the calculation of the date for other *Ceratopteris* species. Thus, we dated the same WGD in the triploid at 54 mya.

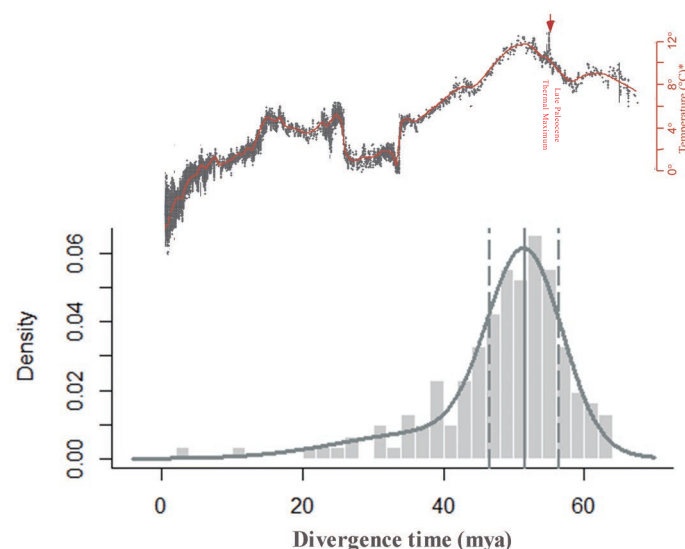


Figure 3. Absolute age distributions for the peak-based duplicates of *C. thalictroides* compared to the global climate changes during the Cenozoic. The vertical gray solid line represents its peak, viewed as the WGD age estimate, and the vertical gray dashed lines corresponded to 95% confidence intervals on the WGD age estimate. The parameters of statistically significant components identified using *mclust* were 52 mya ago and 0.76, which represent the inferred date and proportion, respectively. The global climate curve at the top comes from Zachos et al. (2001) with permission from AAAS [30].

2.6. Phylogenomic Analysis of MADS-box Family Genes

MADS-box family genes, especially MIKC-type proteins, are well known to be involved in the developmental and morphological novelties in the sporophytic generation [31]. We identified and characterized a total of 32 candidate MADS-box genes in *C. thalictroides* (Figure S2), 27 of which were type II MADS-box genes. Our phylogeny of type II MADS-box genes revealed that three genes were clustered together with MIKC*-type genes (Figure 4). CRM1-, CRM3-, and CRM6- formed separate subclades that clustered with MIKC^c proteins in *C. thalictroides*. The CRM6 subclade consisted of a high number of type II MADS-box genes in *C. thalictroides*; however, only two new genes were identified in the CRM3 subclade, and these grouped together with *AtAGL15* protein. A total of four genes, *c24583_g4_i1*, *c16672_g3_i1*, *c24583_g1_i2*, and *c29048_g3_i1*, were retained following the WGD event in *C. thalictroides*. These four genes grouped together with the previously reported CMADS3, CRM4, CMADS2, and CMADS4 proteins, respectively.

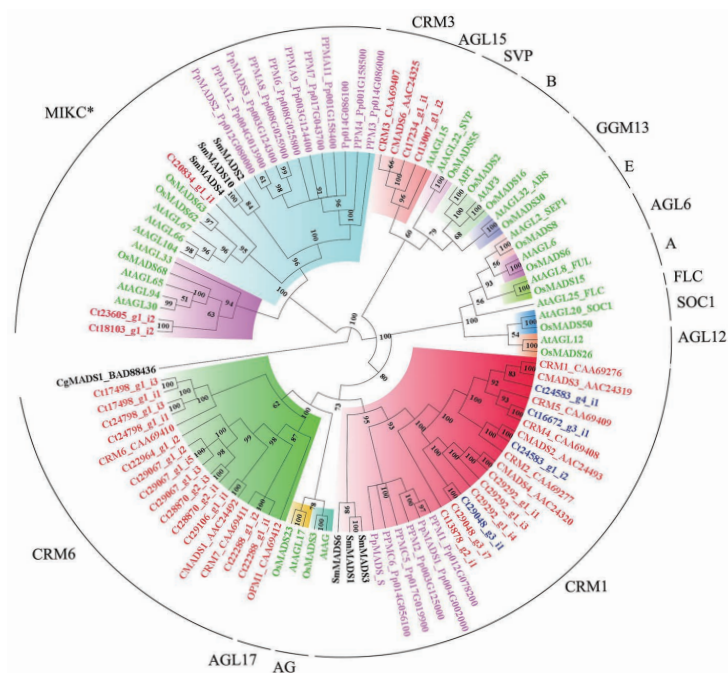


Figure 4. Phylogenetic tree of type II MADS-box proteins using MrBayes 3. The different clades are indicated by different colors, and *C. thalictroides* gene names are given in red, except for the four retained duplicates, denoted with blue, following WGD. Posterior probabilities are also indicated on the branches. The plant species included are as follows: *Arabidopsis*, *Oryza sativa* (Os), *Selaginella moellendorffii* (Sm), *Physcomitrella patens* (Pp), *Chara globularis* (Cg), and *C. thalictroides* (Ct).

3. Discussion

3.1. Inferring WGD from *Ks* Age Distribution

WGDs are common in plants and are considered major factors driving plant diversity [3,5]. Several methods for the detection of WGDs have been reported, including synteny and gene-family trees with a reference species tree, and are widely used in plants with complete genome information. Another commonly applied method uses transcriptome data with a partial expressed sequence tag (EST), which is typically used in determining WGD features on the basis of *Ks* age distribution [3,6,9,12]. However, the quality of transcriptome data assembly significantly influences the identification and understanding of WGDs. In this study, we performed *Ks* analyses to detect the WGD events that have occurred in *C. thalictroides* and discussed the effect of polyploidy level on WGDs.

Nakazato et al. (2006) via genetic linkage mapping indicate that no ancient WGD events were detected in *C. richardii* [32], or signals of WGDs have been masked by chromosomal rearrangements

and smaller-scale duplications. Nevertheless, multiple-copy genes and significant clustering observed in the linkage maps suggest that *C. richardii* may experience WGDs. Previous research has described a WGD event for *C. richardii*; however, the conclusion was built on low-coverage EST sequences with 631 duplicate pairs [24,33]. In our study on *C. thalictroides*, we used 8364 and 3088 paralogous pairs from diploid and triploid samples, respectively (Table 3), thus high-quality transcriptome assembly was possible. Our data is more reliable than previous studies because it covers a total of 69.6% (containing 63.7% complete orthologs and 5.9% partial sequence information) and 46.5% (with 34.3% complete orthologs and 12.2% fragmented genes) of the BUSCO sequences in the diploid and triploid, respectively (Table 2).

The K_s -based age distribution had an obvious peak near 1.2 in the diploid and the triploid samples of *C. thalictroides* (Figure 2), indicating that an ancient WGD event occurred in *C. thalictroides*. We also observed some additional small peaks after 1.2 for each sample. These results indicate that *C. thalictroides* might have undergone one or more WGD events. However, considering the stochastic effect and K_s saturation, which might lead to increasing uncertainty in K_s and artificial peaks in the distribution, we disregarded peaks with values larger than 2 [34], although there were seven and five peaks in the diploid and triploid, respectively (Table 3). The third additional component before 1.2 in the diploid might be the results of a small-scale duplication or stochastic variation because of its lower peaks. Peaks corresponding to young polyploidizations are often superimposed upon older large-scale duplication events such as WGDs [24]. The difference of the K_s frequency distribution and the peak value of 1.526 (diploid) and 2.268 (triploid) in Table 3 might be related to the effect of individual variability with the small sample size or the homoeologous recombination following hybridization and polyploidization during the formation of the triploid. The formation of triploids is still unknown, thus additional studies should be conducted in the future. In addition, it is unclear whether the WGD observed at the peak located at 1.526 and 2.268 are true WGDs or stochastic variations, thus additional research should be conducted. Additionally, although the K_s frequency distribution of the diploid significantly differed from that of the triploid, the ancient WGD event detected is strongly supported, suggesting that ancient occurrences of polyploidy had a larger impact on the K_s age distribution than on recently formed polyploids, which can be disregarded.

Our results reveal that *C. thalictroides* has indeed experienced an ancient WGD, as indicated by the obtained K_s -based age distributions. This event, which is prevalent but not limited to angiosperms, occurred more than once, and the recent polyploidy event has influenced K_s frequency distribution, but its weak interference on the detection of WGDs can be ignored.

3.2. Absolute Dating of the WGD Event in *C. thalictroides*

There are several different methods that can be used to estimate the age of a WGD event. In this study, we used both K_s distribution peaks and the absolute dating approach to infer the dates of WGD events of *C. thalictroides*. Our K_s distribution peak analyses dated the ancient WGD event at 93–101 mya based on the assumption that *C. thalictroides* experienced a similar evolutionary rate with angiosperms. However, our results may not be completely accurate because absolute ages converted by K_s value are more dependent on the assumption of a strict molecular clock and the synonymous substitution rates used. Synonymous substitution rates are variable among different plants, and the rates are unknown for fern species [10,35]. Although the K_s values and absolute ages of *E. giganteum* can be converted to synonymous substitution rates, the date estimation in *E. giganteum* is largely uncertain [9].

Therefore, to either confirm or refute the dating we obtained for the WGD event of the diploid based upon our K_s analysis, we further dated the WGD event using an absolute age distribution, which is a combination of the K_s -based relative age distribution and phylogenetic analysis. This method uses a relaxed clock model that assumes a lognormal distribution of evolutionary rates, is based on several fossil calibrations from a broad taxonomic sampling, and the age estimates for all actual duplicates present an approximate lognormal distribution, which is more likely to obtain accurate estimates than K_s distribution peaks. Absolute age distribution analyses have been widely used in many

studies [9,10,36]. Figure 3 shows that the WGD event occurred at 52 ± 1 mya (95% confidence interval), which is dated at 54 mya in the triploid based on the assumption that *C. thalictroides* shared similar evolutionary rates; the small confidence interval represents a small chance of error. Therefore, we trust the results from the absolute age distribution analysis and placed the WGD event at approximately 52 mya, which is much closer to the PETM than the KT boundary [30]. In previous studies, WGDs were also identified in other ferns, including *E. giganteum*, *C. richardii* (180 mya), *Azolla*, and an earlier WGD shared by core leptosporangiates [9,23,24]. A WGD event in *E. giganteum* was estimated to be 75.16–112.53 mya (with a large confidence interval via absolute dating), and the WGD event placed at 50–70 mya using a *Ks* distribution based on the assumption that *E. giganteum* shares similar evolutionary rates with angiosperms [9]. In the case of *Azolla* and core leptosporangiates, no information was provided for the age of the WGD [23]. We obtained the silent substitution rate r to be 11.04×10^{-9} for the combination of *Ks* and absolute ages of *C. thalictroides*, which is higher than the average rate for plants (6.1×10^{-9}) and cereals (6.5×10^{-9}) and slightly lower than the range for *Drosophila* (15.6×10^{-9}) and dicots (15×10^{-9}) [6,10,25]. The rapid substitution rate might be linked with the life history of *C. thalictroides*, which is an annual and has a short generation time [27,37].

3.3. WGDs Contribute to Evolutionary Success and Potential Adaptation of *Ceratopteris*

Polyploidy is usually considered to be an evolutionary dead end. However, WGDs might occur at specific times, such as during dramatic climate changes, providing novel opportunities for evolutionary success [5]. The occurrence of WGDs are often correlated with mass extinction events and global climate changes [5,10,12]. Many WGD events, including several seed plant species, *E. giganteum*, and *p. patens*, are indicated to cluster around the KT boundary (60–70 mya), which led to the extinction of approximately 60% of the plant species on Earth [9,10,36]. However, the WGD event that occurred in *C. thalictroides* was inferred at 51–53 mya, which is closer to the PETM than the KT boundary. As is common in other tropical species, *C. thalictroides* may be more susceptible to increases in temperature, which facilitates the occurrence of WGD during the PETM [11,13,30,38,39]. Cenozoic climate issues occurring during the PETM, Early Eocene Climatic Optimum, and Mid-Eocene Climatic Optimum have intrigued scientists for a long time because of the drastically fluctuating temperature and atmospheric CO₂ concentration. Viewed as a representative of extreme cases, PETM, also known as the Late Paleocene Thermal Maximum (55 mya), reached exceptionally warmer and wetter climates than any other period on Earth [30,40]. During the PETM, the global temperature increased by at least 5–10 °C due to massive carbon release.

Several studies have suggested that polyploids often have a greater chance to survive an extreme environment compared to diploids because they can undergo rapid morphological innovation [5,22,41–44]. As a pantropical genus, *Ceratopteris* displays greater heat tolerance and intolerance to cooler temperatures than the temperate taxon [45]. In *C. thalictroides*, we discovered that genes were retained after WGDs and these were not random but preferentially retained. Genes preferentially retained after the WGD related to its tolerance to contribute to its adaption to extreme environments [5,36]. Figure S1 shows that genes related to response to stimuli, such as response to salt stress (GO:0071472) and heat acclimation (GO:0010286), were enriched, which is consistent with the climate changes during the PETM (55 mya), as well as higher temperature tolerance and the aquatic environments for extant *C. thalictroides* [30,38,40]. Previous reports have indicated that duplicates in the copy number alternation after WGDs have become subfunctionalized or neofunctionalized, contributing to the species diversity and evolutionary success [5,25]. We observed that four MADS-box proteins belonging to the type II CRM1 subclade that were retained following WGD in *C. thalictroides* (Figure 4), and a member of this gene family, CMADS4 (the homolog of c29048_g3_i1), might have acquired a novel function via neofunctionalization in root development to better survive in aquatic semi-aquatic environments. In addition, changes in copy number would lead to changes in gene expression, likely resulting in hybrid vigor and giving rise to genes and alleles available for selection, further leading to the development of an adaptive phenotype [5,46,47]. WGDs also contribute to

the formation of the MIKC-type proteins in seed plants, most of which have existed in the common ancestor of fern and extant seed plants with ubiquitous expression, and were recruited to specific tissue giving rise to novel function such as floral development [48–50]. Changes in gene expression have also been shown to occur in allopolyploid cotton, which indicated a higher potential adaption than its diploid progenitors [5,51].

4. Materials and Methods

4.1. Chromosome Counting

Two samples of *C. thalictroides* were used in this study. One was collected from Guangdong Province, China, and cultivated in a greenhouse. This sample was previously used for transcriptome analysis [29]. The other sample was obtained from a greenhouse in Sun Yat-sen University, Guangdong, South China, the transcriptome of which was sequenced by Zhang et al. (2016) [28]. The young root tips of the sporophytes were pretreated in 0.002 mol/L 8-hydroxyquinoline solution for 3–6 h and then fixed in Carnoy's solution (95% ethanol: glacial acetic acid = 3:1) for 12–24 h. Then, the samples were hydrolyzed at 37 °C with a mixture of 2% cellulase and pectinase for 1 h. They were then stained with carbol fuchsin. The chromosomes in the samples were counted and photographed using a Carl Zeiss Axio Scope A1 photomicroscope (Jena, Germany).

4.2. Data Collection

In total, transcriptome data from nine species was obtained. For the diploid sample of *C. thalictroides* and seven other fern species including *Goniophlebium niponicum*, *Woodwardia prolifera*, *Dennstaedtia pilosella*, *Cheilanthes chusana*, *Acrostichum aureum*, *Osmolindsaea odorata*, and *Alsophila podophylla*, where the candidate coding sequences and protein sequences were downloaded from GigaScience repository, GigaDB [29]. In the case of the triploid sample of *C. thalictroides*, the unigenes were obtained from the published data with accession number GEEK00000000 [28], and the candidate open reading frames of each putative unigene were predicted with TransDecoder (<http://transdecoder.sf.net>) (access on 1 August 2018) [52].

4.3. BUSCO Analysis

BUSCO analysis was employed for the evaluation of the completeness level of the transcriptome assembly [53]. Both sets of the *C. thalictroides* unigenes were blasted against a core set of conservative orthologs in plant species from the OrthoDB database (www.orthodb.org) (access on 10 August 2018), and the number of complete and partially matched genes were recorded.

4.4. Evaluation of WGD candidates on the Basis of Ks Age Distribution

To examine candidate WGD events in *C. thalictroides*, the Ks-based age distribution was performed as described previously [54]. An all-against-all BLASTP search for each transcriptome was performed using BLASTP version 2.2.29 (NCBI, Bethesda MD, USA) for the identification of gene families with a cutoff E-value of 1×10^{-5} . A best match was considered significant if the alignment length was >100 amino acids and the expected value (E) was $<1 \times 10^{-15}$. Then, all pairs of sequences within each gene family were aligned using MUSCLE (v3.8.31, EMBL-EBI, Hinxton, Cambridgeshire, UK) with default parameters [55]. Ks estimates were generated through maximum likelihood estimation with the CODEML program of PAML version 4.8 (University College London, London, UK) [56]. A gene family of n members originates from n – 1 retained single gene duplications, whereas the number of possible pairwise comparisons within a gene family is $n(n - 1)/2$. To correct for the redundant Ks values, all Ks estimates for a particular duplication event were added to the Ks distribution, while the total weight of a single duplication event sums up to one [6]. The applied Python script is available online (<http://github.com/EndymionCooper/KSPlotting>). Ks values from 0.1 to 5 were retained for subsequent analyses, and Gaussian mixture models were performed in the R package mclust (v5.3, University of

Washington, Seattle, WA, USA) for the fitting of log-transformed K_s distributions [9,57]. The age of a duplication event was inferred by using the relative divergence of the duplicates and the formula divergence date = $K_s/(2 \times r)$ if the number of silent substitutions increased approximately linearly with time. The plant average K_s /year rate of 6.1×10^{-9} was applied to the WGDs of *C. thalictroides* [25]. Additionally, the K_s frequency distributions of the diploids were compared with the triploid through the Wilcoxon matched pairs test [58].

4.5. Absolute Dating

Absolute dating for the WGD was performed to determine the age of the event, similar to the analysis described by Vanneste et al. (2014) [36]. A diagram of the work flow is provided in Figure S3. We determined the age of the WGD event using the diploid sample with a higher quality for this study. Duplicate pairs from the WGD peak with K_s values between 0.9 and 1.5 in the K_s distribution were collected. This resulted in 737 gene families. For each gene family, we retained the duplicates pair nearest the WGD peak boundaries as the representative homeologous pair, as multiple paralogous pairs might descend from the same gene duplication [36]. The orthologs were determined using a BLASTP analysis with the reciprocal best hit based on the protein dataset of peak-based duplicate pairs of the diploid *C. thalictroides* and seven other ferns, with an E-value of 1×10^{-10} [59]. Based on one homeologous pair, an orthogroup was obtained that contained the homeologous pair and several orthologs from other fern species (Figure S4), including *G. niponicum*, *W. prolifera*, *D. pilosella*, *C. chusana*, *A. aureum*, *O. odorata*, and *A. podophylla*. A total of 706 orthogroups were collected. Maximum-likelihood trees were constructed using RAXML (Heidelberg Institute for Theoretical Studies, Heidelberg, Germany) under the GTRGAMMA model with 100 bootstrap replicates, and the paralogous pair of *C. thalictroides* that grouped together was used for further analyses. After removal of the orthogroups with gene trees in conflict with the species tree according to the Pteridophyte Phylogeny Group I system [60], a total of 153 orthogroups were obtained for the dating using BEAST v1.8.4 (University of Auckland, Auckland, New Zealand; University of Edinburgh, Edinburgh, UK; University of California, California, USA) with an uncorrelated relaxed clock model and a JTT+G; (four rate categories; Jones-Taylor-Thornton model and a site heterogeneity model gamma) evolutionary model [61]. Fossil calibrations were employed as follows: *C. chusana*, as the Pteroids, was dated to 93.5 mya, and *W. prolifera*, as *Woodwardia* (family *Blechnaceae*), was dated to 55.8 mya [62–64]. The chain length of the Markov chain Monte Carlo for each orthogroup was set to five million, with sampling every 1000 generations, thus a total of 200 samples. Tracer v1.5 (build-in BEAST 1.8.4) was used to estimate the trace file for all orthogroups, and the orthogroup was accepted if the minimum effective sample size was at least 100 [61]. A total of 153 orthogroups were accepted and all age estimates for the node adding homeologous pair were grouped into one absolute age distribution. The mclust package in R was applied to fit a mixture model based on the given grouped WGD age estimates, and a 95% confidence interval for the significant component was also estimated [57].

4.6. Functional Enrichment

As reported previously, genes are preferentially retained after WGDs [65,66]. We performed a functional enrichment analysis to determine whether biased functional retention in *C. thalictroides* occurs after a WGD and to understand whether the WGDs might be linked with extreme environmental changes. We obtained duplicate pairs with K_s values between 0.9 and 1.5, which lie under the WGD signature peak in the age distributions. GO enrichment analyses were performed using agriGO v2.0 (China Agricultural University, Beijing, China) with the singular enrichment analysis tool at default settings for the Fisher test ($p < 0.05$ signifying statistical significance) [67].

4.7. Identification of the MADS-box Gene Family in *C. thalictroides*

To identify putative MADS-box family genes in *C. thalictroides*, BLASTP was performed to search the *C. thalictroides* protein transcript database with the *Arabidopsis* MADS proteins as queries. An E-value

cutoff of 1×10^{-5} was used. Additionally, proteins with SRF-TF domains (PF00319) were obtained from the Pfam database, and the hidden Markov model (HMM) was also performed to identify new members of the MADS-box gene family. All candidate MADS-box genes were examined by using the Simple Modular Architecture Research Tool (<http://smart.embl-heidelberg.de/>) and conserved domain databases to verify and remove incomplete MADS-box domains. We also removed redundant sequences with an identity higher than 99%.

4.8. Phylogenetic Analysis

To estimate the evolutionary relationship of MADS-box family genes retained in *C. thalictroides* and others, all putative MADS-box family genes in *C. thalictroides* and *Arabidopsis* were aligned using ClustalX (v1.83, Toby Gibson EMBL, Heidelberg, Germany), and manually adjusted. A Bayesian phylogenetic tree was constructed using the program MrBayes (v3.2.6) with the following settings: mixed model, 10,000 generations, and a sampling frequency of ten. The run was stopped when the standard deviation of the split frequencies was below 0.01 through adding generations and sampling frequency continually. A total of ten million generations were run and tree sampling density was 10,000 generations [68]. The first 25% of samples were discarded as burnin. FigTree (v1.4.2, University of Edinburgh, Edinburgh, UK) was used to visualize and edit the consensus tree in Figure S2 [69]. In addition, a Bayesian phylogenetic tree was also constructed based on type II MADS-box genes from *Arabidopsis*, *Oryza sativa*, *Physcomitrella patens*, *Selaginella moellendorffii*, and *C. thalictroides* (Figure 4), including the type II MADS-box genes previously isolated from *Ceratopteris*. *CgMADS1* from the streptophyte green alga, *Chara globularis*, was used as the outgroup.

Supplementary Materials: Supplementary materials can be found at <http://www.mdpi.com/1422-0067/20/8/1926/s1>.

Author Contributions: Experiments were designed by Y.H.Y., R.Z., and H.S. (Hui Shen). Experiments were performed by R.Z., F.G.W., and G.H.Z. Bioinformatics analyses were performed by R.Z., J.Z., L.L., and H.W. R.Z., F.G.W., H.S. (Hui Shang), and Y.H.Y. drafted the manuscript. All authors read and approved the final manuscript.

Funding: This research was funded by the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA13020603), the Shanghai Landscaping & City Appearance Administrative Bureau of China, Scientific Research Grants (grant number G192421), and Guangdong Natural Science Foundation (grant number 2015A030308015).

Acknowledgments: We thank Renchao Zhou (Sun Yat-Sen University) for providing a sample of *Ceratopteris*, and Dongmei Jin and Jun Yang for their help and valuable advice. We are also grateful to Xiaofeng Zhu, Xiangxu Huang, Lei Xu, and Yuwen Cui for their assistance with the chromosome analysis.

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

Abbreviations

WGD	Whole-genome duplication
KT	Cretaceous–Tertiary
PETM	Paleocene-Eocene Thermal Maximum
Ks	synonymous substitutions per synonymous site
BUSCO	Benchmarking Universal Single Copy Orthologs
EST	expressed sequence tag

References

1. Soltis, D.E.; Albert, V.A.; Leebens-Mack, J.; Bell, C.D.; Paterson, A.H.; Zheng, C.; Sankoff, D.; Depamptilis, C.W.; Wall, P.K.; Soltis, P.S. Polyploidy and angiosperm diversification. *Am. J. Bot.* **2009**, *96*, 336–348. [[CrossRef](#)] [[PubMed](#)]
2. Tang, H.; Bowers, J.E.; Wang, X.; Paterson, A.H. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 472–477. [[CrossRef](#)] [[PubMed](#)]
3. 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](#)]

4. Badouin, H.; Gouzy, J.; Grassa, C.J.; Murat, F.; Staton, S.E.; Cottret, L.; Lelandais-Briere, C.; Owens, G.L.; Carrere, S.; Mayjonade, B.; et al. The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature* **2017**, *546*, 148–152. [[CrossRef](#)]
5. Van de Peer, Y.; Mizrachi, E.; Marchal, K. The evolutionary significance of polyploidy. *Nat. Rev. Genet.* **2017**, *18*, 411–424. [[CrossRef](#)] [[PubMed](#)]
6. Blanc, G.; Wolfe, K.H. Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell* **2004**, *16*, 1667–1678. [[CrossRef](#)] [[PubMed](#)]
7. Mun, J.H.; Kwon, S.J.; Yang, T.J.; Seol, Y.J.; Jin, M.; Kim, J.A.; Lim, M.H.; Kim, J.S.; Baek, S.; Choi, B.S.; et al. Genome-wide comparative analysis of the *Brassica rapa* gene space reveals genome shrinkage and differential loss of duplicated genes after whole genome triplication. *Genome Biol.* **2009**, *10*, R111. [[CrossRef](#)] [[PubMed](#)]
8. Zhang, Y.; Xu, G.; Guo, X.; Guo, X.; Fan, L. Two ancient rounds of polyploidy in rice genome. *J. Zhejiang Univ. Sci.* **2005**, *6B*, 87–90. [[CrossRef](#)]
9. Vanneste, K.; Sterck, L.; Myburg, A.A.; Van de Peer, Y.; Mizrachi, E. Horsetails are ancient polyploids: Evidence from *Equisetum giganteum*. *Plant Cell* **2015**, *27*, 1567–1578. [[CrossRef](#)]
10. Fawcett, J.A.; Maere, S.; Van de Peer, Y. Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 5737–5742. [[CrossRef](#)] [[PubMed](#)]
11. Cai, L.; Xi, Z.; Amorim, A.M.; Sugumaran, M.; Rest, J.S.; Liu, L.; Davis, C.C. Widespread ancient whole genome duplications in Malpighiales coincide with Eocene global climatic upheaval. *New Phytol.* **2019**, *221*, 565–576. [[CrossRef](#)] [[PubMed](#)]
12. Ren, R.; Wang, H.; Guo, C.; Zhang, N.; Zeng, L.; Chen, Y.; Ma, H.; Qi, J. Wide-spread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Mol. Plant* **2018**, *11*, 414–428. [[CrossRef](#)] [[PubMed](#)]
13. Van de Peer, Y.; Maere, S.; Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* **2009**, *10*, 725–732. [[CrossRef](#)] [[PubMed](#)]
14. Kassahn, K.S.; Dang, V.T.; Wilkins, S.J.; Perkins, A.C.; Ragan, M.A. Evolution of gene function and regulatory control after whole-genome duplication: Comparative analyses in vertebrates. *Genome Res.* **2009**, *19*, 1404–1418. [[CrossRef](#)] [[PubMed](#)]
15. Li, Z.; Defoort, J.; Tasdighian, S.; Maere, S.; Van De Peer, Y.; De Smet, R. Gene duplicability of core genes is highly consistent across all angiosperms. *Plant Cell* **2016**, *28*, 326–344. [[CrossRef](#)]
16. Lee, H.L.; Irish, V.F. Gene duplication and loss in a MADS box gene transcription factor circuit. *Mol. Biol. Evol.* **2011**, *28*, 3367–3380. [[CrossRef](#)] [[PubMed](#)]
17. Wood, T.E.; Takebayashi, N.; Barker, M.S.; Mayrose, I.; Greenspoon, P.B.; Rieseberg, L.H. The frequency of polyploid speciation in vascular plants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 13875–13879. [[CrossRef](#)]
18. Soltis, P.S.; Marchant, D.B.; Van de Peer, Y.; Soltis, D.E. Polyploidy and genome evolution in plants. *Curr. Opin. Genet. Dev.* **2015**, *35*, 119–125. [[CrossRef](#)]
19. Wagner, W.H.; Wagner, F.S. *Polyploidy: Biological Relevance*; Lewis, W.H., Ed.; Plenum: New York, NY, USA, 1980; pp. 199–214.
20. Klekowski, E.J.; Baker, H.G. Evolutionary significance of polyploidy in the Pteridophyta. *Science* **1966**, *153*, 305–307. [[CrossRef](#)]
21. Clark, J.; Hidalgo, O.; Pellicer, J.; Liu, H.; Marquardt, J.; Robert, Y.; Christenhusz, M.; Zhang, S.; Gibby, M.; Leitch, I.J.; et al. Genome evolution of ferns: Evidence for relative stasis of genome size across the fern phylogeny. *New Phytol.* **2016**, *210*, 1072–1082. [[CrossRef](#)]
22. Schneider, H.; Liu, H.M.; Chang, Y.F.; Ohlsen, D.; Perrie, L.R.; Shepherd, L.; Kessler, M.; Karger, D.N.; Hennequin, S.; Marquardt, J.; et al. Neo- and Paleopolyploidy contribute to the species diversity of *Asplenium*—The most species rich genus of ferns. *J. Syst. Evol.* **2017**, *55*, 353–364. [[CrossRef](#)]
23. Li, F.W.; Brouwer, P.; Carretero-Paulet, L.; Cheng, S.; de Vries, J.; Delaux, P.M.; Eily, A.; Koppers, N.; Kuo, L.Y.; Li, Z.; et al. Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nat. Plants* **2018**, *4*, 460–472. [[CrossRef](#)]
24. Barker, M.S.; Wolf, P.G. Unfurling fern biology in the genomics age. *Bioscience* **2010**, *60*, 177–185. [[CrossRef](#)]
25. Lynch, M.; Conery, J.S. The evolutionary fate and consequences of duplicate genes. *Science* **2000**, *290*, 1151–1155. [[CrossRef](#)] [[PubMed](#)]

26. Leroux, O.; Eeckhout, S.; Viane, R.L.; Popper, Z.A. *Ceratopteris richardii* (C-Fern): A model for investigating adaptive modification of vascular plant cell walls. *Front. Plant Sci.* **2013**, *4*, 367. [[CrossRef](#)] [[PubMed](#)]
27. Sessa, E.B.; Banks, J.A.; Barker, M.S.; Der, J.P.; Duffy, A.M.; Graham, S.W.; Hasebe, M.; Langdale, J.; Li, F.W.; Marchant, D.B.; et al. Between two fern genomes. *GigaScience* **2014**, *3*, 15. [[CrossRef](#)] [[PubMed](#)]
28. Zhang, Z.; He, Z.; Xu, S.; Li, X.; Guo, W.; Yang, Y.; Zhou, R.; Shi, S. Transcriptome analyses provide insights into the phylogeny and adaptive evolution of the mangrove fern genus *Acrostichum*. *Sci. Rep.* **2016**, *6*, 35634. [[CrossRef](#)] [[PubMed](#)]
29. Shen, H.; Jin, D.; Shu, J.P.; Zhou, X.L.; Lei, M.; Wei, R.; Shang, H.; Wei, H.J.; Zhang, R.; Liu, L.; et al. Large scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* **2018**, *7*, 1–11. [[CrossRef](#)]
30. Zachos, J.; Pagani, M.; Sloan, L.; Thomas, E.; Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **2001**, *292*, 686–693. [[CrossRef](#)] [[PubMed](#)]
31. Gramzow, L.; Theissen, G. A hitchhiker’s guide to the MADS world of plants. *Genome Boil.* **2010**, *11*, 214. [[CrossRef](#)]
32. Nakazato, T.; Jung, M.K.; Housworth, E.A.; Rieseberg, L.H.; Gastony, G.J. Genetic map-based analysis of genome structure in the homosporous fern *Ceratopteris richardii*. *Genetics* **2006**, *173*, 1585–1597. [[CrossRef](#)]
33. Barker, M.S. Evolutionary genomic analyses of ferns reveal that high chromosome numbers are a product of high retention and fewer rounds of polyploidy relative to angiosperms. *Am. Fern J.* **2009**, *99*, 136–141.
34. Vanneste, K.; Van de Peer, Y.; Maere, S. Inference of genome duplications from age distributions revisited. *Mol. Biol. Evol.* **2013**, *30*, 177–190. [[CrossRef](#)]
35. Van de Peer, Y. Computational approaches to unveiling ancient genome duplications. *Nat. Rev. Genet.* **2004**, *5*, 752. [[CrossRef](#)] [[PubMed](#)]
36. Vanneste, K.; Baele, G.; Maere, S.; Van de Peer, Y. Analysis of 41 plant genomes supports a wave of successful genome duplications in association with the Cretaceous-Paleogene boundary. *Genome Res.* **2014**, *24*, 1334–1347. [[CrossRef](#)]
37. Tang, H.; Wang, X.; Bowers, J.E.; Ming, R.; Alam, M.; Paterson, A.H. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* **2008**, *18*, 1944–1954. [[CrossRef](#)] [[PubMed](#)]
38. Hickok, L.G.; Warne, T.R.; Fribourg, R.S. The biology of the fern *Ceratopteris* and its use as a model system. *Int. J. Plant Sci.* **1955**, *156*, 332–345. [[CrossRef](#)]
39. Tewksbury, J.J.; Huey, R.B.; Deutsch, C.A. Putting the heat on tropical animals. *Science* **2008**, *320*, 1296–1297. [[CrossRef](#)]
40. Zachos, J.C.; Dickens, G.R.; Zeebe, R.E. An early Cenozoic perspective on greenhouse warming and carbon-cycle dynamics. *Nature* **2008**, *451*, 279–283. [[CrossRef](#)]
41. Wendel, J.F. Genome evolution in polyploids. *Plant Mol. Evol.* **2000**, *42*, 225–249.
42. Ramsey, J. Polyploidy and ecological adaptation in wild yarrow. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7096–7101. [[CrossRef](#)]
43. Chao, D.Y.; Dilkes, B.P.; Luo, H.; Douglas, A.; Yakubova, E.; Lahner, B.; Salt, D.E. Polyploids exhibit higher potassium uptake and salinity tolerance in *Arabidopsis*. *Science* **2013**, *341*, 658–659. [[CrossRef](#)]
44. Diallo, A.M.; Nielsen, L.R.; Kjaer, E.D.; Patersen, K.K.; Raebild, A. Polyploidy can confer superiority to West African *Acacia senegal* (L.) Willd. trees. *Front. Plant Sci.* **2016**, *7*, 821. [[CrossRef](#)]
45. Warne, T.R.; Lloyd, R.M. The role of spore germination and gametophyte development in habitat selection: Temperature responses in certain temperate and tropical ferns. *Bull. Torrey Bot. Club* **1980**, *107*, 57–64. [[CrossRef](#)]
46. Muenster, T.; Pahnke, J.; Di Rosa, A.; Kim, J.T.; Martin, W.; Saedler, H.; Theissen, G. Floral homeotic genes were recruited from homologous MADS-box genes preexisting in the common ancestor of ferns and seed plants. *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2415–2420. [[CrossRef](#)]
47. Hasebe, M.; Wen, C.-K.; Kato, M.; Banks, J.A. Characterization of MADS homeotic genes in the fern *Ceratopteris richardii*. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6222–6227. [[CrossRef](#)]
48. Theissen, G.; Becker, A.; Di Rosa, A.; Kanno, A.; Kim, J.T.; Münster, T.; Winter, K.U.; Saedler, H. A short history of MADS-box genes in plants. *Plant Mol. Evol.* **2000**, *42*, 115–149.
49. Veron, A.S.; Kaufmann, K.; Bornberg-Bauer, E. Evidence of interaction network evolution by wholegenome duplications: A case study in MADS-box proteins. *Mol. Biol. Evol.* **2007**, *24*, 670–678. [[CrossRef](#)] [[PubMed](#)]
50. Airoidi, C.A.; Davies, B. Gene duplication and the evolution of plant MADS-box transcription factors. *J. Genet. Genomics* **2012**, *39*, 157–165. [[CrossRef](#)] [[PubMed](#)]

51. Thangavel, G.; Nayar, S. A survey of MIKC type MADS-box genes in non-seed plants: Algae, Bryophytes, Lycophytes and Ferns. *Front. Plant Sci.* **2018**, *9*, 510. [[CrossRef](#)] [[PubMed](#)]
52. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
53. Simao, F.A.; Waterhouse, R.M.; Ioannidi, P.; Kriventseva, E.V.; Zdobnov, E.M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **2015**, *31*, 3210–3212. [[CrossRef](#)] [[PubMed](#)]
54. Sollars, E.S.A.; Harper, A.L.; Kelly, L.J.; Sambles, C.M.; Ramirez-Gonzalez, R.H.; Swarbreck, D.; Kaithakottil, G.; Cooper, E.D.; Uauy, C.; Havlickova, L.; et al. Genome sequence and genetic diversity of European ash trees. *Nature* **2017**, *541*, 212–216. [[CrossRef](#)] [[PubMed](#)]
55. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [[CrossRef](#)] [[PubMed](#)]
56. Yang, Z. PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **2007**, *24*, 1586–1591. [[CrossRef](#)] [[PubMed](#)]
57. Fraley, C.; Raftery, A.E. *MCLUST Version 3: An R Package for Normal Mixture Modeling and Model-based Clustering*; Department of Statistic, University of Washington: Seattle, WA, USA, 2006.
58. Zar, J.H. *Biostatistical Analysis*, 4th ed.; Prentice Hall International: Englewood Cliffs, NJ, USA, 1999.
59. Moreno-Hagelsieb, G.; Latimer, K. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics* **2007**, *24*, 319–324. [[CrossRef](#)]
60. Schuettpelz, E.; Schneider, H.; Smith, A.R.; Hovenkamp, P.; Prado, J.; Rouhan, G.; Salino, A.; Sundue, M.; Almeida, T.E.; Parris, B.; et al. A community-derived classification for extant lycophytes and ferns. *J. Syst. Evol.* **2016**, *54*, 563–603.
61. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **2012**, *29*, 1969–1973. [[CrossRef](#)]
62. Krassilov, V.; Bacchia, F. Cenomanian florule of Nammoura, Lebanon. *Cretac. Res.* **2000**, *21*, 785–799. [[CrossRef](#)]
63. Collinson, M.E. Cainozoic ferns and their distribution. *Brittonia* **2001**, *53*, 172–235. [[CrossRef](#)]
64. Bonde, S.D.; Kumaran, K.P.N. The oldest macrofossil record of the mangrove fern *Acrostichum* L. from the Late Cretaceous Deccan Intertrappean beds of India. *Cretac. Res.* **2002**, *23*, 149–152. [[CrossRef](#)]
65. De Smet, R.; Van de Peer, Y. Redundancy and rewiring of genetic networks following genome-wide duplication events. *Curr. Opin. Plant Biol.* **2012**, *15*, 168–176. [[CrossRef](#)]
66. Wang, Y.; Wang, X.; Paterson, A.H. Genome and gene duplications and gene expression divergence: A view from plants. *Ann. N. Y. Acad. Sci.* **2012**, *1256*, 1–14. [[CrossRef](#)] [[PubMed](#)]
67. Tian, T.; Liu, Y.; Yan, H.; You, Q.; Yi, X.; Du, Z.; Xu, W.; Su, Z. agriGO v2. 0: A GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Res.* **2017**, *45*, W122–W129. [[CrossRef](#)] [[PubMed](#)]
68. Ronquist, F.; Huelsenbeck, J.P. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **2003**, *19*, 1572–1574. [[CrossRef](#)] [[PubMed](#)]
69. Rambaut, A. *FigTree v1.4.2, a Graphical Viewer of Phylogenetic Trees*; University of Edinburgh: Edinburgh, UK, 2014. Available online: <http://tree.bio.ed.ac.uk/software/figtree> (accessed on 8 October 2018).

