



OPEN

Patterns and Rates of Plastid *rps12* Gene Evolution Inferred in a Phylogenetic Context using Plastomic Data of Ferns

Shanshan Liu¹, Zhen Wang¹, Hui Wang², Yingjuan Su^{1,3} ✉ & Ting Wang⁴ ✉

The *trans*-splicing *rps12* gene of fern plastomes (plastid genomes) exhibits a unique structure owing to its variations in intragenic exon location and intron content, and thus, it provides an excellent model system for examining the effect of plastid gene structure on rates and patterns of molecular evolution. In this study, 16 complete fern plastome sequences were newly generated via the Illumina HiSeq sequencing platform. We reconstructed the phylogeny of ferns and inferred the patterns and rates of plastid *rps12* gene evolution in a phylogenetic context by combining these plastome data with those of previously published fern species. We uncovered the diversity of fern plastome evolution by characterizing the structures of these genomes and obtained a highly supported phylogenetic framework for ferns. Furthermore, our results revealed molecular evolutionary patterns that were completely different from the patterns revealed in previous studies. There were significant differences in the patterns and rates of nucleotide substitutions in both intron-containing and intron-less *rps12* alleles. Rate heterogeneity between single-copy (SC) and inverted repeat (IR) exons was evident. Unexpectedly, however, IR exons exhibited significantly higher synonymous substitution rates (dS) than SC exons, a pattern that contrasts the regional effect responsible for decreased rates of nucleotide substitutions in IRs. Our results reveal that structural changes in plastid genes have important effects on evolutionary rates, and we propose possible mechanisms to explain the variations in the nucleotide substitution rates of this unusual gene.

Plastid ribosomes are ubiquitous organelles in plant cells and play a vital role in the biosynthesis of proteins. In higher plants, plastid ribosomes contain approximately 60 ribosomal proteins that are encoded in both the plastid and the nuclear genetic compartments¹. Among the plastid-encoded ribosomal protein gene structures, *rps12* is the most notable. The plastid ribosomal protein S12 encoded by the *rps12* gene is a highly conserved protein located in the functional center of the 30S subunit of the ribosome². In fern plastomes (plastid genomes), where *rps12* is a *trans*-splicing gene, this gene is split into three exons by two introns in most, but not all, ferns, and one intron (intron I) is discontinuous. The first exon of the *rps12* gene is generally located in the large single-copy (LSC) region, whereas the second and third exons reside in the inverted repeats (IRs); the two IR copies have identical sequences but opposite transcriptional directions. More importantly, the second intron is lacking from the *rps12* gene of all species belonging to three basal fern lineages: Psilotales, Ophioglossales, and Equisetales^{3,4}. Thus, two distinct *rps12* gene types were identified in ferns based on the presence or absence of intron II, and type I and II genes corresponded to intron-containing and intron-less genes, respectively. Plastid genomic structural alterations, such as inversions, duplications, and gene or intron loss, are often accompanied by an increase in the rate of plastome sequence evolution^{5,6}. Therefore, different copy numbers of exons and the presence or absence of introns have garnered substantial interest as an avenue to explore the evolutionary patterns of this unique gene.

Introns are highly stable components of land plant plastomes, and it is widely believed that a basic set of introns was established prior to the divergence between vascular and nonvascular plants, because this intron set is shared by different taxonomic groups (such as charophytes, bryophytes, and spermatophytes)⁷. The

¹School of Life Sciences, Sun Yat-sen University, Guangzhou, 510275, China. ²Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen, 518004, China. ³Research Institute of Sun Yat-sen University in Shenzhen, Shenzhen, 518057, China. ⁴College of Life Sciences, South China Agricultural University, Guangzhou, 510642, China. ✉e-mail: suyj@mail.sysu.edu.cn; tingwang@scau.edu.cn

Species	Raw data (G)	Clean data (G)	Raw Reads	Clean reads	Coverage (×)	Accession number
<i>L. microphyllum</i>	2.43	2.08	8,083,733	6,927,768	198.45	MN623356
<i>P. bifurcatum</i>	6.99	6.36	23,309,670	21,215,174	552.65	MN623367
<i>L. hederaceum</i>	2.28	2.21	7,585,298	7,066,889	260.67	MN623364
<i>S. yakushimensis</i>	2.94	2.71	9,815,752	9,033,527	284.70	MN623352
<i>T. decurrens</i>	2.17	2.05	7,228,697	6,826,579	24.29	MN623363
<i>N. cordifolia</i>	2.04	1.81	6,794,240	6,026,844	85.16	MN623365
<i>D. sieboldii</i>	3.26	3.00	10,851,235	10,014,316	221.45	MN623354
<i>B. subcordata</i>	2.48	2.28	8,276,022	7,604,656	127.23	MN623358
<i>P. triphyllum</i>	5.79	5.37	19,289,987	17,902,643	356.00	MN623361
<i>P. decursive-pinnata</i>	3.22	2.86	10,736,184	9,531,427	79.56	MN623353
<i>G. erubescens</i>	2.89	2.63	9,643,034	8,770,301	1196.58	MN623355
<i>O. gibba</i>	3.25	3.06	11,775,566	11,370,848	125.57	MN623360
<i>B. insignis</i>	3.86	3.58	12,880,011	11,926,722	178.90	MN623366
<i>D. maximum</i>	3.13	2.87	10,430,597	9,565,536	128.62	MN623359
<i>S. lepifera</i>	3.10	2.87	10,346,974	9,561,719	253.00	MN623357
<i>P. subadnata</i>	2.80	2.56	9,342,117	8,540,938	393.80	MN623362

Table 1. Summary of the sequencing data for 16 fern species.

intron contents of fern plastomes are also highly conserved, with no gains and few losses during fern evolution^{4,8}. The interspecies variation in the intron content of fern plastomes mainly reflects the complete loss of several intron-containing genes in specific lineages (e.g., intron-containing *rps16* genes are absent in Psilotales, Ophioglossales, and Equisetales⁴), or the presence of intron-less alleles in some lineages (e.g., intron-less *rps12* genes are present in Psilotales, Ophioglossales, and Equisetales; *Lygodium japonicum* lacks *rpoC1* introns that are commonly found in other ferns⁸). Intron losses have been associated with elevated substitution rates in plastid genes, as previously described^{5,9,10}, but there is an extreme bias in taxon sampling of these studies. The vast majority of these studies are based on angiosperms. Consequently, we cannot determine whether this pattern is a universal phenomenon in the plastome or an independent evolutionary event in certain lineages or genes. The plastid *rps12* gene in ferns provides an excellent model system to study the effect of intron losses on evolutionary rate and test the generality of this evolutionary pattern.

Another distinctive feature of the *rps12* gene in ferns is the variation in copy number among its exons. As mentioned previously, in most ferns, the first exon of *rps12* is located in the LSC, whereas the second and third exons reside in the IRs. It is well known that two completely identical IR regions are prominent structural features of the plastomes of nearly all land plants. The regional effect of the evolutionary pattern of plastid genes, in which the evolutionary rate of genes located in the IRs is lower than that of genes in the single-copy (SC) regions, has been documented with extensive research^{11–15}. The sequence identity of IRs can be maintained by copy-dependent DNA repair because when mutations are introduced into one IR copy, the other copy provides a template for error correction^{13,16}, thereby suppressing the substitution rate in the IRs. Biased gene conversion, as an efficient mutation-correcting mechanism, can result in a genome with different copy regions that have different mutation rates^{16,17}. There are up to 900 genomic copies in a single plastid¹⁸, whereas the duplicative property of the IRs provides an even greater number of copies; therefore, the frequency of gene conversion in the IRs should be higher than that in the SC regions, and this phenomenon might be responsible for a significantly lower evolutionary rate in the IRs than in the SC regions^{14,16}.

However, with a growing quantity of plastome data available for comparative genomics, this hypothesis of the regional effect of the IRs has been refuted by several studies^{19–21}. In addition to the impact of IRs on the evolutionary rate, the non-IR localization, locus- and lineage-specific also have significant effects on evolutionary rate heterogeneity in plastomes^{19,21–23}. Therefore, we suggest that each of these factors alone cannot sufficiently explain the rates and patterns of molecular evolution in plastid genes. For this reason, we turned our research focus to the split *rps12* gene because of its location in both the SC and IR regions. The variation in the exon location and intron content of the *rps12* gene in fern plastomes provides a unique opportunity to explore the effect of gene structure on sequence evolution. To thoroughly understand the pattern of *rps12* gene evolution in ferns, in the current study, we employed greatly expanded taxon sampling, including 91 fern species, and undertook a broad survey to investigate the impact of structural variation on plastid gene evolution.

Results and Discussion

Organization and dynamic structural evolution of plastid genomes in ferns. Whole-genome sequencing using an Illumina HiSeq platform generated 6,794,240–23,309,670 raw reads for 16 samples. We obtained 6,026,844–21,215,174 clean reads by removing adaptors and low-quality read pairs (Table 1). Following *de novo* and reference-guided assembly strategies, 16 newly sequenced fern plastomes were each assembled into a single circular molecule. Both genome size and GC content were relatively conserved among all species (Table 2). The size of the 16 plastome sequences ranged from 148,928 bp in *Dryopteris sieboldii* to 164,857 bp in *Selliguea yakushimensis*, and all plastomes displayed a typical quadripartite structure consisting of a large-single copy region (LSC, 79,002–92,033 bp), a small single-copy region (SSC, 19,484–27,733 bp), and a pair of inverted

Species	Genome size (bp)	LSC (bp)	IR (bp)	SSC (bp)	GC %	Gene	<i>trnR</i> -UCG	<i>trnV</i> -UAC	<i>trnT</i> -UGU	<i>trnN</i> -GUU
<i>L. microphyllum</i>	158,029	81,244	27,494	21,797	41.83	132	+	+	+	+
<i>P. bifurcatum</i>	156,985	79,002	28,249	21,485	39.91	130	-	-	+	+
<i>L. hederaceum</i>	152,337	81,395	24,593	21,756	42.65	132	+	+	+	+
<i>S. yakushimensis</i>	164,857	80,975	32,017	19,848	40.80	135	+	+	+	+
<i>T. decurrens</i>	151,258	82,963	23,256	21,783	41.98	129	-	+	+	-
<i>N. cordifolia</i>	149,152	82,020	22,850	21,432	39.36	131	-	+	+	+
<i>D. sieboldii</i>	148,928	82,251	22,528	21,621	43.08	131	+	-	+	+
<i>B. subcordata</i>	153,428	83,087	24,476	21,389	42.39	132	+	+	+	+
<i>P. triphyllum</i>	151,908	82,777	23,617	21,897	42.82	132	+	+	+	+
<i>P. decursive-pinnata</i>	150,995	82,344	23,530	21,591	42.37	132	+	+	+	+
<i>G. erubescens</i>	156,961	82,715	26,229	21,788	43.15	132	+	+	+	+
<i>O. gibba</i>	159,641	92,033	23,018	21,572	43.53	132	+	+	+	+
<i>B. insignis</i>	149,734	81,453	23,387	21,508	41.40	131	+	-	+	+
<i>D. maximum</i>	150,984	82,293	23,462	21,767	43.90	132	+	+	+	+
<i>S. lepifera</i>	162,216	86,349	24,067	27,733	40.80	132	+	+	ψ	+
<i>P. subadnata</i>	159,998	89,960	24,307	21,424	42.94	132	+	+	ψ	+

Table 2. Plastome features of the sequenced species. Plus and minus signs denote genes that are present and absent, respectively, in the corresponding species. ψ represents pseudogenes.

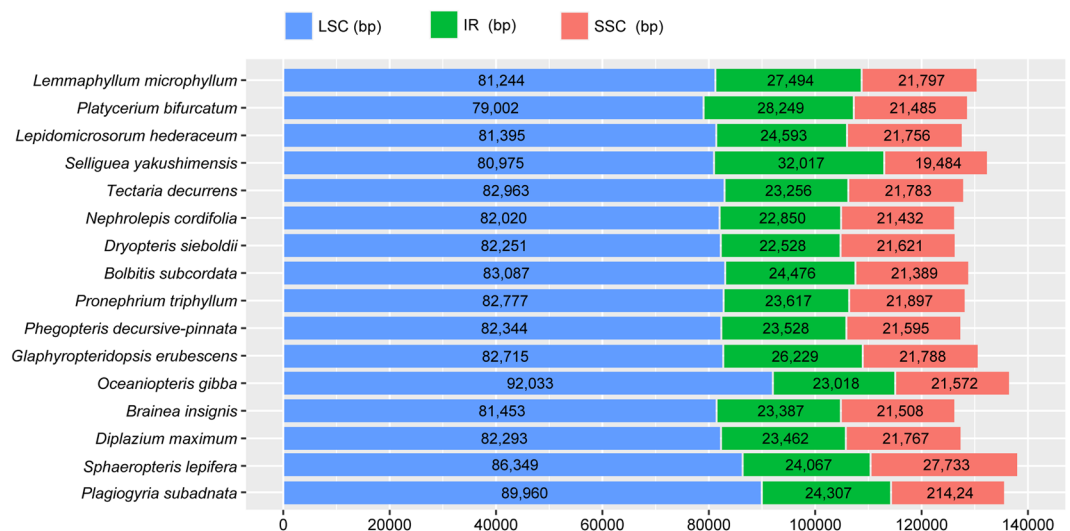


Figure 1. Sizes of each part of 16 fern complete plastome sequences.

repeats (IRs, 22,528–32,017 bp) (Fig. 1 and Table 2). Across all sequenced ferns, there were 84–86 protein coding genes, 27–29 tRNA genes and 4 rRNA genes. The polypods and tree ferns had similar coding gene contents, with a few notable distinctions. Compared with the 84 protein coding genes inferred to be present in the ancestral plastomes of polypods, tree ferns had an additional *psaM* gene that was shared with most other non-polypods. The majority of the sequenced samples displayed typical fern intron contents. The *rps12* gene in all the sequenced species was classified as type I, and the first exon was located in the LSC, far away from the second and third exons of *rps12*, which were present in two copies in the IRs.

Although the plastome structure remained relatively stable, some species showed exceptional variability in tRNA gene content (Table 2). Several tRNA gene losses from multiple independent lineages were detected, including the loss of *trnR*-UCG in *Platynerium bifurcatum*, *Tectaria decurrens*, and *Nephrolepis cordifolia*; the loss of *trnV*-UAC in *P. bifurcatum*, *D. sieboldii*, and *Brainea insignis*; the loss of *trnN*-GUU in *T. decurrens*; and the pseudogenization of *trnT*-UGU in *Sphaeropteris lepifera* and *Plagiogyria subadnata* (Table 2). All of these genes have also been parallelly lost in other polypods plastomes with the exception of the *trnN*. The *trnN* is one of the core set genes contained in IRs, generally adjacent to either *ndhF* or *chlL* at the IR/SSC borders. In ferns, IRs are generally considered to be the most stable part of the plastome because genomic rearrangement rarely occurs in these regions. However, in contrast, a recent study indicated that the IR sequences and gene contents were highly variable in polypods²⁴. Our results are in accordance with the latter findings and provide additional evidence for the dynamic evolution of IR regions among closely related polypod plastomes.

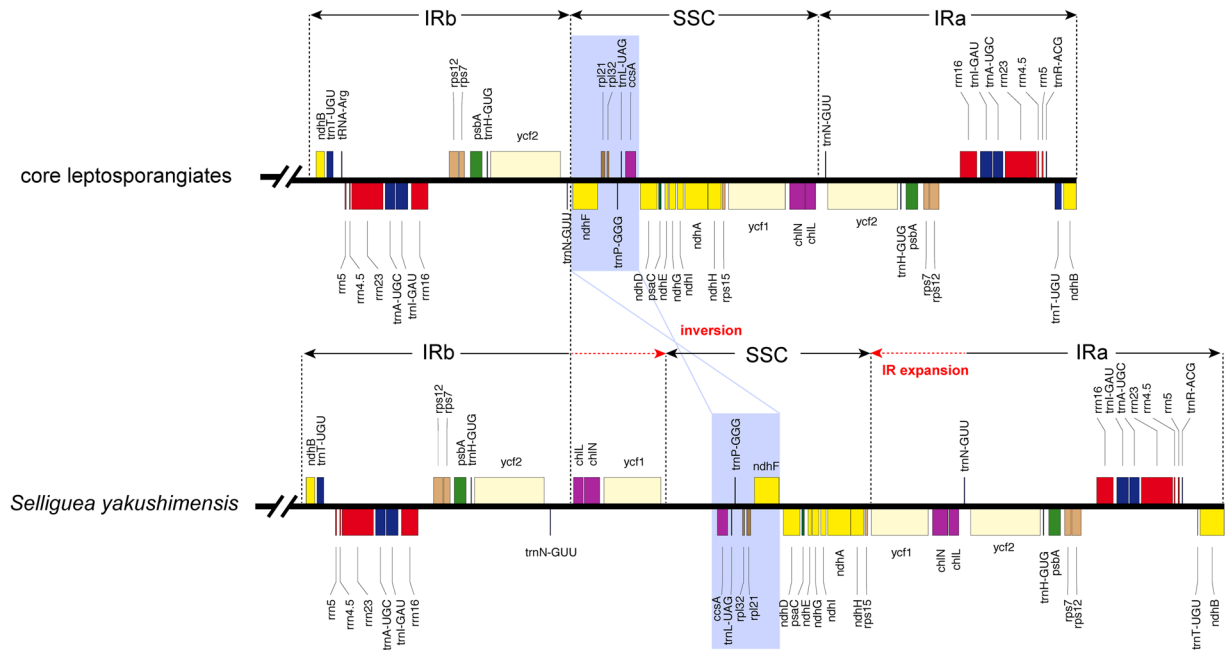


Figure 2. Unique structural changes in the plasmome of *Selligaea yakushimensis* compared to those of other core leptosporangiates. Genes are represented by colored boxes above or below of the black chromosome bar according to the direction of transcription. The novel inversion that occurs in the *S. yakushimensis* is shown in a purple box. The red dashed arrow indicates the range of the IR expansion in the *S. yakushimensis* plasmome.

Polypods are the lineage of most derived ferns that diversified in the Cretaceous period, displaying an ecologically opportunistic response to the diversification of angiosperms²⁵. The plastomes of polypods have undergone multiple complex genomic reconfigurations during fern evolution, and thus, their plastomes differ substantially from the plastomes of basal ferns (Psilotales, Ophioglossales, Marattiales, and Equisetales). Plastome evolution among polypods is considered relatively static compared with that in lineages other than polypods²⁶. Surprisingly, distinct genome organizations were identified in *S. yakushimensis* based on inversions and IR boundary variation. A major variation in the *S. yakushimensis* plastome relative to the genomes of other core leptosporangiates is the altered location from the *ndhF-ccsA* (Fig. 2). In addition, IR expansion into the SSC resulted in a duplication of the *ycf1*, *chL*, and *chLN* genes in the *S. yakushimensis* plastome. The extent of IR expansion in the *S. yakushimensis* plastomes is unprecedented among ferns. The IRs in *S. yakushimensis* are up to 32 kb in length (Fig. 1 and Table 2), whereas the longest IR found previously was 29 kb in *Cibotium barometz*²⁷. These rare genome structure variations in polypods were detected for the first time.

Phylogenetic analyses. Both the 93- and 84-taxon datasets based on 50 protein-coding genes showed consistent phylogenetic framework, differing only in the support values for some nodes (Fig. 3 and Supplementary Fig. S1). Along the backbone, Equisetales was uncovered as a sister clade to the remaining ferns with strong support, followed by a highly supported joint Ophioglossales + Psilotales clade, itself sister to a clade with Marattiales (Fig. 3). The phylogenetic relationships among these four basal fern orders are the most debated topics in fern phylogeny. Most previous studies using nuclear genes^{28–30}, a combination of mitochondrial and plastid sequences³¹, several plastid genes^{32,33}, and whole plastome sequences also obtained this topology^{34,35}. In contrast, other phyloplastomic studies tend to support grouping Equisetales and Ophioglossales + Psilotales together as a monophyletic group and sister to the remaining ferns^{36,37}. In addition, the recent phylotranscriptomic analyses have also revealed a distinct topology of relationships among the basal fern orders, showing that Equisetales is the sister group to all other ferns, whereas Marattiales and Psilotales + Ophioglossales form a monophyletic group³⁸. Given that plastid genes generally evolve more slowly than the nuclear genes, these topological differences may be due to different numbers of phylogenetically informative sites contained within the diverse molecular data³⁹.

The diversification of leptosporangiates occurred after that of euposporangiate ferns, within which Osmundales was the earliest-diverging lineage, and then, Gleicheniales and Hymenophyllales formed a monophyletic group sister to the remaining non-Osmundales ferns (Fig. 3). The phylogenetic position of Hymenophyllales remains debated. Most phylogenetic analyses have depicted two alternative relationships: Hymenophyllales as a sister clade to Gleicheniales + the rest of non-Osmundales leptosporangiates^{25,32,33,40–43}, or Hymenophyllales and Gleicheniales together form a clade that is sister to the remaining non-Osmundales leptosporangiates^{29,38,44}. Interestingly, both topologies have been found in more recent plastid phylogeny reconstructions using different data types and partition schemes³⁷. Kuo *et al.*³⁷ showed that the use of plastome organization features also fails to provide additional support for either of these two topologies. This phylogenetic uncertainty is probably due to sparse taxon sampling, data types, model selection, and tree inference methods. Although the relative positions

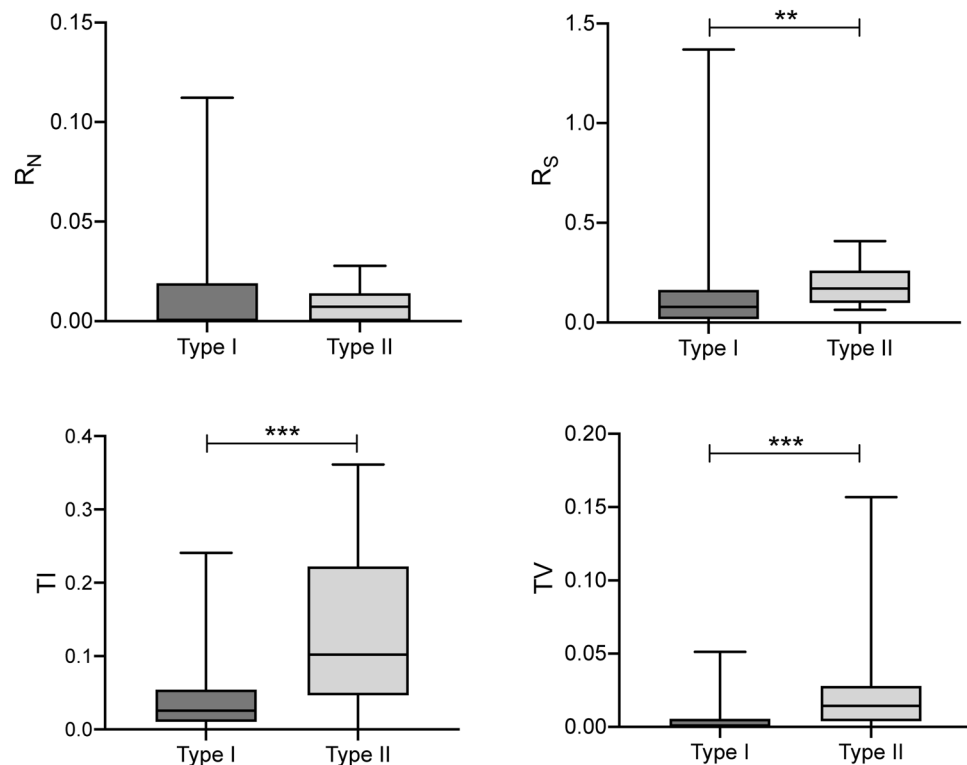


Figure 4. Comparison of R_N , R_S , TI, and TV rates between intron-containing (type I) and intron-less (type II) *rps12* genes in ferns. Asterisks indicate (**) $P < 0.01$ and (***) $P < 0.001$.

Impact of intron loss on the *rps12* evolutionary rate. The occurrence of *rps12* intron loss in three basal fern lineages is considered to be an important evolutionary event in ferns. If the phylogenetic relationships among ferns are indeed consistent with our analysis, which recovered Equisetales as a sister clade to the remaining ferns, then the intron of the *rps12* gene would have been independently lost at least twice during fern evolution⁴. Research suggests that the rate of plastome sequence evolution is generally affected by structural changes^{5,6}; thus, we sought to investigate the impact of intron loss on the evolutionary rate of the *rps12* genes.

Branch length in a phylogenetic tree represents an estimate of the amount of sequence divergence in the corresponding lineage, which is equal to the product of the absolute substitution rate and time⁴⁶. For this reason, we cannot ignore the timescale over which molecular rates change in comparisons of substitution rates among taxa. Therefore, we tested for an intron effect on substitution rate changes by comparing the absolute substitution rates of the type I and type II genes. Our results showed that the *trans*-splicing *rps12* genes in all examined species were highly conserved with a size of 372 bp, encoding a total of 123 amino acids. The differences in the rates between the two types of genes were mainly reflected in the rates of synonymous substitution (R_S). Wilcoxon rank sum tests showed that the values of R_S for type II genes were significantly higher than those for type I genes ($P < 0.01$), whereas the values of the rates of nonsynonymous substitution (R_N) were not significantly different ($P = 0.8019$) (Fig. 4; Supplementary Fig. S2 and Table S1). Moreover, in order to detect whether the selection pressures acting on the two types of genes were significantly different, we compared dN/dS ratios in a phylogenetic context via a model-based approach. A likelihood ratio test (LRT) was used to compare the fits of two models: the null model, where values of dN/dS were not significantly different between type I and type II, and the alternative model, where type I had different dN/dS ratios relative to type II. Overall, the alternative model was significantly different from the null model ($P < 0.05$).

Currently, the most widely accepted mechanism of intron loss is a reverse transcriptase (RT)-mediated model (namely, retroprocessing), which is a process of integrating intron-less cDNA generated by reverse transcription of the corresponding mRNA into the genome by homologous recombination^{47–49}. Another two possible mechanisms of intron loss include genomic deletion⁵⁰ and exonization of intronic sequences⁵¹. In these cases, the intron is often removed imprecisely, changing the intron/exon borders. Consequently, the *rps12* intron loss that occurred in the plastomes of ferns is indicative of the first mechanism because the exon boundaries of the intron-less genes have been shown to be perfectly matched to that of the intron-containing genes. Previous studies have reported that the accelerated evolution of *clpP1* in many seed plant lineages is associated with intron loss^{9,10,52}. A possible explanation for this acceleration in *clpP1* evolution is that reverse transcriptases and/or RNA polymerases have higher error rates during retroprocessing⁵³. Although the rapid evolution of the *clpP1* gene could also relate to a hybrid effect (e.g., duplications, indels, and pseudogenization), we found a significantly higher R_S for type II genes than for type I genes in the fairly conserved gene *rps12*, indicating that accelerated gene evolutionary rates are correlated with loss of introns. Therefore, this phenomenon may be a prevalent genome-wide pattern.

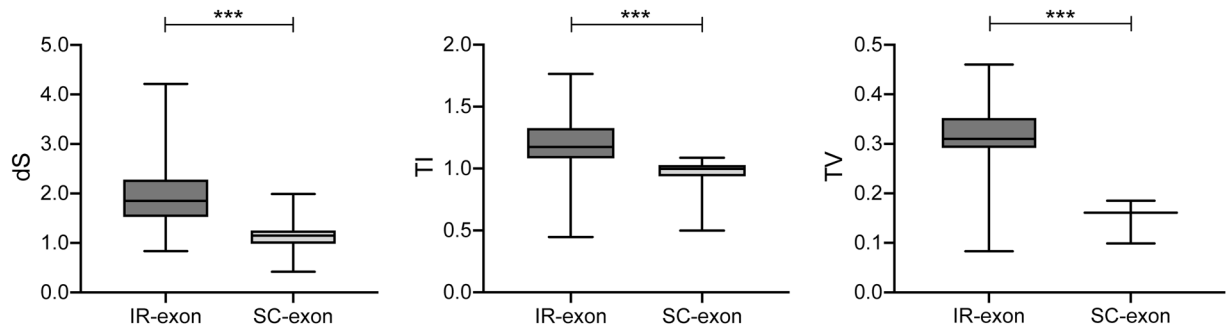


Figure 5. Comparison of dS, TI, and TV values between IR and SC exons of *rps12* genes in ferns. Asterisks indicate (***) $P < 0.001$.

Furthermore, we analyzed the two well-known types of genetic mutations, transitions (TIs) and transversions (TVs), to test whether there were differences in mutation pattern between type I and type II genes. Our results showed that TIs occurred more frequently than TVs in both type I and type II genes (Fig. 4; Supplementary Fig. S2 and Table S2). From the perspective of natural selection, this “transition bias” phenomenon is considered to indicate that selection disfavors transversions, as transversions are more likely to alter the amino acid sequence of proteins than transitions⁵⁴. Comparison between the TI and TV values showed that both values were significantly higher in type II than in type I genes ($P < 0.001$) (Fig. 4; Supplementary Fig. S2 and Table S2). An LRT was also implemented to identify rate shifts between the type I and type II branches, and the results indicated that TI/TV was significantly different in the type I branches than in the type II branches ($P < 0.001$). That is, transition and transversion events that occurred in type II genes resulted in more synonymous changes than those in type I genes. This finding may explain the significant change in R_S between the type I and type II genes, whereas R_N did not exhibit significant changes.

Complicated rate variation in *rps12* exons. Due to a series of genomic inversions across the fern phylogeny, 84 of 93 sampled species showed the same pattern of exon distribution, which was divided into IR and SC exons. The consensus sequences of the SC and IR exons of the fern *rps12* genes were 114 bp (encoding the 1st to 38th amino acids) and 232 bp in length (encoding the 39th to 123rd amino acids), respectively. Since nonsynonymous mutations are more strongly affected by selective constraints, while synonymous mutations are largely invisible to natural selection, a comparison of the synonymous substitution rates provides a better understanding of DNA sequence evolution⁵⁵. In order to compare with previous research more intuitively, we used the same pairwise comparison method as the previous studies when investigating the difference between the rates of IR and SC exons^{12,21}. Our results revealed a more complicated rate variation in the *rps12* exons than expected. The dS values for the IR exon were significantly higher than those for the SC exon ($P < 0.001$), a pattern that contrasts the regional effect responsible for decreased rates of nucleotide substitutions in the IRs (Fig. 5; Supplementary Table S3). The results of an LRT between the one-rate and two-rate models showed significant rate changes relative to the position of an exon in either the IR or SC region ($P < 0.01$).

A previous study indicated that the total substitution rates decreased after gene translocation into the IRs by using a model-based approach across the fern phylogeny¹¹. Moreover, Zhu *et al.*¹² compared the dS values of IR and SC genes across land plants using the pairwise estimation method, including some representative fern species. Based on very limited taxon sampling, they observed that former IR genes that moved into the SC region due to genome rearrangements underwent corresponding dS accelerations. However, we discovered that the IR exons of a given gene in fern species had, in fact, significantly higher dS rates on average than the corresponding SC exons. The most substantial distinction between our study and previous studies is that we compared the IR exons of *rps12* with the SC exons of the same gene rather than other SC genes. Furthermore, we employed vastly expanded taxon sampling that included not only the major fern lineages but also additional species of derived lineages. Comparing the rate changes in exons located in different regions of a given gene can effectively avoid rate heterogeneity caused by lineage effects⁸. The intragenic location changes in the *rps12* exons present a unique opportunity to test the effect of IRs on evolutionary rate variation. Thus, the pattern of substitution rate inhibition in the IRs may not be applicable to the fern *rps12* gene. Indeed, several studies have shown that a pattern of decreased IR substitution rates is not universally suitable among vascular plants, as increased IR substitution rates were previously observed in some genes of *Pelargonium*^{19,21}, *Plantago*, and *Silene*¹². Several possible mechanisms have previously been proposed to explain substitution rate increases in plastid genes. These abnormally evolving genes could result from local hypermutation potentially induced by a high level of error-prone double-strand break repair¹². A subsequent study by Weng *et al.*²¹ expanded taxon sampling to characterize rates and patterns of evolution in *Pelargonium*, and the results showed that the anomalous rate acceleration observed in *Pelargonium* plastomes could be explained by a mixture of locus-specific, lineage-specific, and IR-dependent effects.

Thus, the mechanisms responsible for generating substitution rate variation may be different in each case. Considering that homologous recombination could enhance copy-correction activity in the IRs via gene conversion, the most direct explanation is the reduction in recombinant activity in IRs. Furthermore, a model of recombination repair^{16,56} could serve as another possible explanation for their observations of increased dS. Recombination repair is also known as post-replication repair⁵⁶, which is a process of repairing impaired

Family	Genus	Species	Sampling site
Polypodiaceae	<i>Lemmaphyllum</i>	<i>L. microphyllum</i>	FLBG
	<i>Platyserium</i>	<i>P. bifurcatum</i>	FLBG
	<i>Lepidomicrosorium</i>	<i>L. hederaceum</i>	WBG
	<i>Selliguea</i>	<i>S. yakushimensis</i>	WBG
Thelypteridaceae	<i>Pronephrium</i>	<i>P. triphyllum</i>	FLBG
	<i>Phegopteris</i>	<i>P. decursive-pinnata</i>	FLBG
	<i>Glaphyopteridopsis</i>	<i>G. erubescens</i>	WBG
Dryopteridaceae	<i>Bolbitis</i>	<i>B. subcordata</i>	FLBG
	<i>Dryopteris</i>	<i>D. sieboldii</i>	FLBG
Blechnaceae	<i>Oceaniopteris</i>	<i>O. gibba</i>	FLBG
	<i>Brainea</i>	<i>B. insignis</i>	FLBG
Tectariaceae	<i>Tectaria</i>	<i>T. decurrens</i>	FLBG
Nephrolepidaceae	<i>Nephrolepis</i>	<i>N. cordifolia</i>	FLBG
Athyriaceae	<i>Diplazium</i>	<i>D. maximum</i>	FLBG
Cyatheaceae	<i>Sphaeropteris</i>	<i>S. lepifera</i>	FLBG
Plagiogyriaceae	<i>Plagiogyria</i>	<i>P. subadnata</i>	WBG

Table 3. Taxa sampled in this study. FLBG, Fairy Lake Botanical Garden, Chinese Academy of Sciences (CSA); WBG, Wuhan Botanical Garden, CAS.

molecules by using undamaged molecules as donors, and if this repair is error-prone, single base pair substitutions will be generated because the repair process involves gap-filling DNA synthesis¹⁶. Additional copies of IR exons in a cell or organelle can act as donors to ensure the efficiency of recombination repair. Consequently, we speculate that recombinant activity and recombination repair may have larger effects than the IR on the substitution rate variation in the *rps12* gene.

Furthermore, values of TI and TV for SC and IR exons were also estimated. The results showed that both the TI and TV values in the SC exons were significantly higher than those in the IR exons (Fig. 5; Supplementary Table S3), and an LRT rejected the one-rate model in favor of the two-rate model ($P < 0.001$). Likewise, comparisons of the DNA sequences of the *rps12* gene showed that the base composition in the two exons was not uniform and an excess of transitional over transversional substitutions was present. One reason for the transition bias observed in these exons could be a mutational bias due to the intrinsic properties of DNA, as purines and pyrimidines have different conformational sizes.

Materials and Methods

Taxon sampling and DNA extraction. Ferns are a species-rich lineage of vascular plants, occupying a high diversity of ecological niches, and are a major component of the earth's land flora⁵⁷. However, only 132 complete plastomes are available from GenBank. All 132 complete fern plastome sequences in the NCBI RefSeq collection as of 10 Feb 2019 were downloaded from GenBank, and the annotated coding sequences and the number of exons in *rps12* were extracted from these sequences. Because the *rps12* gene is highly conserved in fern plastomes and no substantial variation in the *rps12* sequence was observed among congeners, only one sample was chosen from each genus (except for two samples that were selected from *Equisetum*) among the previously published fern plastomes to reduce redundancy in the dataset. Then, additional plastome sequences from 16 fern species were sequenced to ensure the coverage of additional fern clades. The increased sampling was undertaken to better understand the evolutionary patterns of *rps12* gene in ferns. Sequenced species mainly included tree ferns (Cyatheales) and polypod ferns (Polypodiales), which contain most of the extant fern diversity. This sampling strategy resulted in 93 samples representing all 11 extant fern orders and 32 families and two plastomes of lycophytes (outgroup) (Supplementary Table S4).

In this study, fresh leaves of 16 newly sequenced fern species were sampled from Wuhan Botanical Garden, Chinese Academy of Sciences (CSA), and Fairy Lake Botanical Garden (CSA), respectively (Table 3). Samples were taken from young leaves of each plant and were either flash-frozen in liquid nitrogen or placed in paper envelopes and dried with silica gel. Genomic DNA was extracted from the silica-dried or lyophilized leaf tissue of each sample using the Plant Genomic DNA Kit (Tiangen Biotech., Beijing, China). The DNA concentration and purity assessments were performed using agarose gel electrophoresis and a NanoDrop spectrophotometer (Thermo Scientific, Carlsbad, CA, USA). Isolations with concentrations ≥ 150 ng/ μ l were chosen for Illumina sequencing.

Genome sequencing, assembly, and annotation. Genomic DNA was sheared using a Covaris M220 focused-ultrasonication device (Covaris Inc., MS, USA) to a mean fragment size of 300 bp. Paired-end libraries were prepared using the NEBNext Ultra DNA Library Prep Kit (New England Biolabs, Ipswich, MA) for sequencing on an Illumina HiSeq 4000 platform (Illumina Inc., San Diego, CA, USA). Subsequently, 150 bp paired-end reads were produced with an insert size of ~ 300 bp. Following enrichment, we obtained the raw data for 16 species, ranging from 2.04–6.99 Gb (Table 1). Raw data were then filtered, and adaptor sequences were removed with Trimmomatic v0.33⁵⁸. Trimming was performed from both ends of each read, removing all bases with a quality lower than Q20, and keeping only reads of 50 nt or longer. The quality-filtered reads were subjected to *de novo*

assembly using Velvet v1.2.08⁵⁹. Each assembled contig was BLAST searched against previously reported fern plastomes to identify plastid contigs. Gaps that remained in the assembled draft sequence were filled by polymerase chain reaction (PCR) using specific primers that were designed from the regions flanking the gaps.

We annotated plastomes in DOGMA⁶⁰. Plastid genes were corrected in each draft genome by performing homology searches with BLASTX or BLASTN against the previously published fern plastomes. Genes were annotated as pseudogenes if they showed disruptions in reading frames or frameshifts caused by nontriplet insertions or deletions. Transfer RNA (tRNA) genes were verified using two online programs, ARAGORN⁶¹ and tRNAscan-SE⁶². Plastome maps for all species were drawn using OGDRAW⁶³. All fully annotated plastome sequences were deposited in NCBI GenBank under accession numbers MN623352–MN623367 (Table 1).

Phylogenetic analyses. Phylogenetic analyses were performed on two separate protein-coding gene matrices. The first included 50 shared genes from all the sampled taxa. The second dataset included 50 gene sequences from 84 taxa. This matrix included the same genes as the first dataset except that those taxa with a single copy of *rps12* were removed (Supplementary Table S4). All sequence alignments for each gene were conducted using MAFFT v7.310⁶⁴, and trimmed to exclude poorly aligned positions using Gblocks v0.91b⁶⁵ with default settings. The trimmed alignments were concatenated in SequenceMatrix v1.0⁶⁶. Phylogenetic trees from the concatenated dataset were estimated using the maximum likelihood (ML) method in RAxML v8.2.4⁶⁷ with the standard GTRGAMMA model by 1,000 rapid bootstrap replicates.

Estimation of evolutionary rates. Nucleotide sequences for each part exon of the *rps12* gene from 91 fern species and two lycophyte outgroups, *Huperzia lucidula* and *Isoetes flaccida*, were extracted from plastomes and spliced into a complete coding region. For this step, the exon/intron boundaries of each sample must be carefully examined, and incorrectly annotated genes were manually adjusted if necessary because misalignment would result in the erroneous inference of substitutions rates. The coding sequences of the *rps12* gene were aligned at the protein level using the Align Codon option of ClustalW as implemented in MEGA7⁶⁸.

Absolute substitution rates for *rps12* were calculated according to the methods that have been described previously^{69,70}. In brief, divergence times for all nodes within ferns were estimated using the penalized likelihood approach in the program r8s⁷¹, with a time constraint of 354 million years for the crown group age of ferns²⁵. The ML tree constructed from the dataset of all sampled taxa was used as the starting tree. We performed 100 bootstrap replicates to calculate standard errors for the divergence time of each node. The branch lengths of dS (number of substitutions per synonymous site) and dN (number of substitutions per nonsynonymous site) trees were computed using the codeml module in PAML v4.8⁷². The standard errors of dN and dS branches were calculated from the standard errors of their corresponding total branch lengths as reported by PAML. Absolute rates of synonymous (R_s) and nonsynonymous (R_n) substitution for each branch were calculated by dividing the synonymous/nonsynonymous branch length by the length of time corresponding to that branch.

To compare the nucleotide mutation patterns of the two gene types, the TI and TV rates for *rps12* were also estimated in the phylogenetic context of ferns using the HKY85 model (the local parameters option) as implemented in HyPhy⁷³. The significance of the differences in R_s , R_n , TI, and TV between the type I and type II branches were evaluated using Wilcoxon rank sum tests. We used a model-based method to test whether the dN/dS changes differed between the type I and type II branches in a phylogenetic context. The phylogenetic tree of all 93 species generated using ML was used as a constraint tree. The null model, where one dN/dS ratio was fixed across ferns, was compared with an alternative model, where the type I and type II clades were allowed to have different dN/dS ratios. LRTs were performed in PAML v4.8⁷² to compare the goodness of fit of the two models. We also tested for possible rate heterogeneity in the TI/TV ratios across different branches in a phylogenetic context. The LRT scheme for detecting a significant difference in TI/TV between type I and type II genes was conducted using the HKY85 model in HyPhy⁷³, in which the null and alternative models specified were the same as those used when detecting dN/dS.

Furthermore, to investigate whether the evolutionary rate variations in the *rps12* exon corresponded to its location, that is, the IR or SC regions, we performed a parallel analysis of 84 species to determine the effects of IR location on the substitution rate of a given gene. Nine samples were excluded from this analysis because their *rps12* exon sequences were all single copies (Supplementary Table S4). Based on their locations, the *rps12* exon sequences were extracted from the plastomes and aligned at the protein level, and the alignments were concatenated separately. The pairwise dS rates for the IR and SC exons between the two lycophytes and each fern species were calculated using the codeml module in PAML v4.8⁷² (seqtype = 1, runmode = -2, and CodonFreq = 2 in the codeml.ctl files). The ML tree inferred from 50 plastid genes of 82 taxa was used as a constraint tree. Likewise, we calculated the pairwise TI and TV of the two exons by using the HKY85 model implemented in HyPhy to compare the nucleotide mutation patterns. Wilcoxon rank sum tests were used to test the rate differences between the SC and IR exons. To investigate whether the dN/dS changes in exons corresponded to their location, on the phylogeny, a model-based rate analysis was used. LRTs were performed using the MG94xHKY85 codon model in HyPhy⁷³, in which the one-rate model specified a shared dN/dS across the whole gene, whereas the two-rate model allowed the IR and SC exons to have different dN/dS values. The same setting was applied to the LRTs for TI/TV, except that a HKY85 substitution model was applied.

Data availability

The complete plastome sequences generated in current study are available in GenBank, <https://www.ncbi.nlm.nih.gov/genbank/> (accession numbers are described in the text).

Received: 6 November 2019; Accepted: 14 May 2020;

Published online: 10 June 2020

References

- Eneas-Filho, J., Hartley, M. R. & Mache, R. Pea chloroplast ribosomal proteins: characterization and site of synthesis. *Mol. Gen. Genet.* **184**, 484–488, <https://doi.org/10.1007/BF00352527> (1981).
- Yamaguchi, K. & Subramanian, A. R. Proteomic identification of all plastid-specific ribosomal proteins in higher plant chloroplast 30S ribosomal subunit. *Eur. J. Biochem.* **270**, 190–205, <https://doi.org/10.1046/j.1432-1033.2003.03359.x> (2003).
- Karol, K. G. *et al.* Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evol. Biol.* **10**, 321, <https://doi.org/10.1186/1471-2148-10-321> (2010).
- Grewe, F., Guo, W., Gubbels, E. A., Hansen, A. K. & Mower, J. P. Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evol. Biol.* **13**, 8, <https://doi.org/10.1186/1471-2148-13-8> (2013).
- Jansen, R. K. *et al.* Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proc. Natl. Acad. Sci. USA* **104**, 19369–19374, <https://doi.org/10.1073/pnas.0709121104> (2007).
- Weng, M. L., Blazier, J. C., Govindu, M. & Jansen, R. K. Reconstruction of the ancestral plastid genome in Geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **31**, 645–659, <https://doi.org/10.1093/molbev/mst257> (2013).
- Barkan, A. Intron splicing in plant organelles in *Molecular Biology and Biotechnology of Plant Organelles* (eds Daniell, H. & Chase, C.) 295–322, https://doi.org/10.1007/978-1-4020-3166-3_11 (Springer, Dordrecht, 2004).
- Gao, L. *et al.* Plastome sequences of *Lygodium japonicum* and *Marsilea crenata* reveal the genome organization transformation from basal ferns to core leptosporangiates. *Genome Biol. Evol.* **5**, 1403–1407, <https://doi.org/10.1093/gbe/evt099> (2013).
- Erixon, P. & Oxelman, B. Whole-gene positive selection, elevated synonymous substitution rates, duplication, and indel evolution of the chloroplast *clpP1* gene. *PLoS One* **3**, e1386, <https://doi.org/10.1371/journal.pone.0001386> (2008).
- Park, S. *et al.* Contrasting patterns of nucleotide substitution rates provide insight into dynamic evolution of plastid and mitochondrial genomes of *Geranium*. *Genome Biol. Evol.* **9**, 1766–1780, <https://doi.org/10.1093/gbe/evx124> (2017).
- Li, F. W., Kuo, L. Y., Pryer, K. M. & Rothfels, C. J. Genes translocated into the plastid inverted repeat show decelerated substitution rates and elevated GC content. *Genome Biol. Evol.* **8**, 2452–2458, [10.1093/gbe/evw167](https://doi.org/10.1093/gbe/evw167) (2016).
- Zhu, A., Guo, W., Gupta, S., Fan, W. & Mower, J. P. Evolutionary dynamics of the plastid inverted repeat: the effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **209**, 1747–1756, <https://doi.org/10.1111/nph.13743> (2016).
- Wolfe, K. H., Li, W. H. & Sharp, P. M. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proc. Natl. Acad. Sci. USA* **84**, 9054–9058, <https://doi.org/10.1073/pnas.84.24.9054> (1987).
- Perry, A. S. & Wolfe, K. H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J. Mol. Evol.* **55**, 501–508, <https://doi.org/10.1007/PL00020998> (2002).
- Yamane, K., Yano, K. & Kawahara, T. Pattern and rate of indel evolution inferred from whole chloroplast intergenic regions in sugarcane, maize and rice. *DNA Res.* **13**, 197–204, <https://doi.org/10.1093/dnares/dsl012> (2006).
- Birky, C. & Walsh, J. Biased gene conversion, copy number, and apparent mutation rate differences within chloroplast and bacterial genomes. *Genetics* **130**, 677–683 (1992).
- Khakhlova, O. & Bock, R. Elimination of deleterious mutations in plastid genomes by gene conversion. *Plant J.* **46**, 85–94, <https://doi.org/10.1111/j.1365-3113X.2006.02673.x> (2006).
- Bendich, A. J. Why do chloroplasts and mitochondria contain so many copies of their genome? *BioEssays* **6**, 279–282, <https://doi.org/10.1002/bies.950060608> (1987).
- Guisinger, M. M., Kuehl, J. V., Boore, J. L. & Jansen, R. K. Genome-wide analyses of Geraniaceae plastid DNA reveal unprecedented patterns of increased nucleotide substitutions. *Proc. Natl. Acad. Sci. USA* **105**, 18424–18429, <https://doi.org/10.1073/pnas.0806759105> (2008).
- Blazier, J. C. *et al.* Variable presence of the inverted repeat and plastome stability in *Erodium*. *Ann. Bot.* **117**, 1209–1220, <https://doi.org/10.1093/aob/mcw065> (2016).
- Weng, M. L., Ruhlman, T. A. & Jansen, R. K. Expansion of inverted repeat does not decrease substitution rates in *Pelargonium* plastid genomes. *New Phytol.* **214**, 842–851, <https://doi.org/10.1111/nph.14375> (2017).
- Sloan, D. B., Alverson, A. J., Wu, M., Palmer, J. D. & Taylor, D. R. Recent acceleration of plastid sequence and structural evolution coincides with extreme mitochondrial divergence in the angiosperm genus *Silene*. *Genome Biol. Evol.* **4**, 294–306, <https://doi.org/10.1093/gbe/evs006> (2012).
- Sloan, D. B., Triant, D. A., Wu, M. & Taylor, D. R. Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol. Biol. Evol.* **31**, 673–682, <https://doi.org/10.1093/molbev/mst259> (2013).
- Logacheva, M. D. *et al.* Comparative analysis of inverted repeats of polypod fern (Polypodiales) plastomes reveals two hypervariable regions. *BMC Plant Biol.* **17**, 255, <https://doi.org/10.1186/s12870-017-1195-z> (2017).
- Schneider, H. *et al.* Ferns diversified in the shadow of angiosperms. *Nature* **428**, 553–557, <https://doi.org/10.1038/nature02361> (2004).
- Wolf, P. G., Roper, J. M. & Duffy, A. M. The evolution of chloroplast genome structure in ferns. *Genome* **53**, 731–738, <https://doi.org/10.1139/G10-061> (2010).
- Liu, S. S., Wang, Z., Wang, T. & Su, Y. J. The complete chloroplast genome of *Cibotium barometz* (Cibotiaceae), an endangered CITES medicinal fern. *Mitochondrial DNA Part B* **3**, 464–465, <https://doi.org/10.1080/23802359.2018.1462128> (2018).
- Rothfels, C. J. *et al.* The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *Am. J. Bot.* **102**, 1089–1107, <https://doi.org/10.3732/ajb.1500089> (2015).
- Qi, X. *et al.* A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Mol. Phylogenet. Evol.* **127**, 961–977, <https://doi.org/10.1016/j.ympev.2018.06.043> (2018).
- Wolf, P. G. *et al.* Target sequence capture of nuclear-encoded genes for phylogenetic analysis in ferns. *Appl. Plant Sci.* **6**, e01148, <https://doi.org/10.1002/aps3.1148> (2018).
- Knier, N., Fischer, S., Grewe, F., Polsakiewicz, M. & Knoop, V. Horsetails are the sister group to all other monilophytes and Marattiales are sister to leptosporangiate ferns. *Mol. Phylogenet. Evol.* **90**, 140–149, <https://doi.org/10.1016/j.ympev.2015.05.008> (2015).
- Rai, H. S. & Graham, S. W. Utility of a large, multigene plastid data set in inferring higher-order relationships in ferns and relatives (monilophytes). *Am. J. Bot.* **97**, 1444–1456, <https://doi.org/10.3732/ajb.0900305> (2010).
- Testo, W. & Sundue, M. A 4000-species dataset provides new insight into the evolution of ferns. *Mol. Phylogenet. Evol.* **105**, 200–211, <https://doi.org/10.1016/j.ympev.2016.09.003> (2016).
- Kim, H. T., Chung, M. G. & Kim, K. J. Chloroplast genome evolution in early diverged leptosporangiate ferns. *Mol. Cells* **37**, 372–382, <https://doi.org/10.14348/molcells.2014.2296> (2014).
- Labiak, P. H. & Karol, K. G. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. *Am. J. Bot.* **104**, 1008–1018, <https://doi.org/10.3732/ajb.1700135> (2017).
- Lu, J. M., Zhang, N., Du, X. Y., Wen, J. & Li, D. Z. Chloroplast phylogenomics resolves key relationships in ferns. *J. Syst. Evol.* **53**, 448–457, <https://doi.org/10.1111/jse.12180> (2015).
- Kuo, L. Y., Qi, X., Ma, H. & Li, F. W. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. *Am. J. Bot.* **105**, 1545–1555, <https://doi.org/10.1002/ajb2.1152> (2018).
- Shen, H. *et al.* Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* **7**, 1–11, <https://doi.org/10.1093/gigascience/gix116> (2018).

39. Zhang, N., Zeng, L., Shan, H. & Ma, H. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* **195**, 923–937, <https://doi.org/10.1093/gigascience/gix116> (2012).
40. Pryer, K. M. *et al.* Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature* **409**, 618–622, <https://doi.org/10.1038/35054555> (2001).
41. Schuettpeiz, E., Korall, P. & Pryer, K. M. Plastid atpA data provide improved support for deep relationships among ferns. *Taxon* **55**, 897–906, <https://doi.org/10.2307/25065684> (2006).
42. Schuettpeiz, E. & Pryer, K. M. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon* **56**, 1037–1037, <https://doi.org/10.2307/25065903> (2007).
43. Qiu, Y. L. *et al.* The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. USA* **103**, 15511–15516, <https://doi.org/10.1073/pnas.0603335103> (2006).
44. Pryer, K. M. *et al.* Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *Am. J. Bot.* **91**, 1582–1598, <https://doi.org/10.3732/ajb.91.10.1582> (2004).
45. PPG I. A community-derived classification for extant lycophytes and ferns. *J. Syst. Evol.* **54**, 563–603, <https://doi.org/10.1111/jse.12229> (2016).
46. Mower, J. P., Touzet, P., Gummow, J. S., Delph, L. F. & Palmer, J. D. Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants. *BMC Evol. Biol.* **7**, 135, <https://doi.org/10.1186/1471-2148-7-135> (2007).
47. Derr, L. K. & Strathern, J. N. A role for reverse transcripts in gene conversion. *Nature* **361**, 170–173, <https://doi.org/10.1038/361170a0> (1993).
48. Mourier, T. & Jeffares, D. C. Eukaryotic intron loss. *Science* **300**, 1393–1393, <https://doi.org/10.1126/science.1080559> (2003).
49. Cohen, N. E., Shen, R. & Carmel, L. The role of reverse transcriptase in intron gain and loss mechanisms. *Mol. Biol. Evol.* **29**, 179–186, <https://doi.org/10.1093/molbev/msr192> (2011).
50. Llopart, A., Comeron, J. M., Brunet, F. G., Lachaise, D. & Long, M. Intron presence–absence polymorphism in *Drosophila* driven by positive Darwinian selection. *Proc. Natl. Acad. Sci. USA* **99**, 8121–8126, <https://doi.org/10.1073/pnas.122570299> (2002).
51. Parma, J., Christophe, D., Pohl, V. & Vassart, G. Structural organization of the 5' region of the thyroglobulin gene: Evidence for intron loss and “exonization” during evolution. *J. Mol. Biol.* **196**, 769–779, [https://doi.org/10.1016/0022-2836\(87\)90403-7](https://doi.org/10.1016/0022-2836(87)90403-7) (1987).
52. Williams, A. M., Friso, G., van Wijk, K. J. & Sloan, D. B. Extreme variation in rates of evolution in the plastid Clp protease complex. *Plant J.* **98**, 243–259, <https://doi.org/10.1111/tpj.14208> (2019).
53. Preston, B. D. Error-prone retrotransposition: rime of the ancient mutators. *Proc. Natl. Acad. Sci. USA* **93**, 7427–7431, <https://doi.org/10.1073/pnas.93.15.7427> (1996).
54. Vogel, F. & Kopun, M. Higher frequencies of transitions among point mutations. *J. Mol. Evol.* **9**, 159–180, <https://doi.org/10.1007/BF01732746> (1977).
55. Akashi, H. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* **139**, 1067–1076, <https://www.genetics.org/content/139/2/1067.short> (1995).
56. Sharp, P. M., Shields, D. C., Wolfe, K. H. & Li, W. H. Chromosomal location and evolutionary rate variation in enterobacterial genes. *Science* **246**, 808–810, <https://doi.org/10.1126/science.2683084> (1989).
57. Smith, A. R. *et al.* A classification for extant ferns. *Taxon* **55**, 705–731, <https://doi.org/10.2307/25065646> (2006).
58. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, <https://doi.org/10.1093/bioinformatics/btu170> (2014).
59. Zerbino, D. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829, <https://doi.org/10.1101/gr.074492.107> (2008).
60. Wyman, S. K., Jansen, R. K. & Boore, J. L. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**, 3252–3255, <https://doi.org/10.1093/bioinformatics/bth352> (2004).
61. Laslett, D. & Canback, B. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* **32**, 11–16, <https://doi.org/10.1093/nar/gkh152> (2004).
62. Lowe, T. M. & Chan, P. P. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* **44**, W54–W57, <https://doi.org/10.1093/nar/gkw413> (2016).
63. Lohse, M., Drechsel, O., Kahlau, S. & Bock, R. OrganellarGenomeDRAW—a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**, W575–W581, <https://doi.org/10.1093/nar/gkt289> (2013).
64. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780, <https://doi.org/10.1093/molbev/mst010> (2013).
65. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552, <https://doi.org/10.1093/oxfordjournals.molbev.a026334> (2000).
66. Vaidya, G., Lohman, D. J. & Meier, R. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* **27**, 171–180, <https://doi.org/10.1111/j.1096-0031.2010.00329.x> (2011).
67. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313, <https://doi.org/10.1093/bioinformatics/btu033> (2014).
68. Kumar, S., Stecher, G. & Tamura, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874, <https://doi.org/10.1093/molbev/msw054> (2016).
69. Cho, Y., Mower, J. P., Qiu, Y. L. & Palmer, J. D. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proc. Natl. Acad. Sci. USA* **101**, 17741–17746, <https://doi.org/10.1073/pnas.0408302101> (2004).
70. Parkinson, C. L. *et al.* Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evol. Biol.* **5**, 73, <https://doi.org/10.1186/1471-2148-5-73> (2005).
71. Sanderson, M. J. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**, 301–302, <https://doi.org/10.1093/bioinformatics/19.2.301> (2003).
72. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591, <https://doi.org/10.1093/molbev/msm088> (2007).
73. Pond, S. L. K. & Muse, S. V. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**, 676–679, <https://doi.org/10.1093/bioinformatics/bti079> (2005).

Acknowledgements

This work was supported by the National Natural Science Foundation of China [31370364, 31570652, 31670200, 31770587, and 31872670], the Natural Science Foundation of Guangdong Province, China [2016A030313320 and 2017A030313122], Science and Technology Planning Project of Guangdong Province, China [2017A030303007], Project of Department of Science and Technology of Shenzhen City, Guangdong, China [JCYJ20160425165447211, JCYJ20170413155402977, JCYJ20170818155249053, and JCYJ20190813172001780], and Science and Technology Planning Project of Guangzhou City, China [201804010389].

Author contributions

Ting Wang conceived and designed the experiments; Hui Wang, Shanshan Liu, and Yingjuan Su collected and contributed sample material; Zhen Wang performed the experiments; Shanshan Liu analyzed the data and wrote the manuscript; Ting Wang and Yingjuan Su reviewed the manuscript critically. All authors have read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-66219-y>.

Correspondence and requests for materials should be addressed to Y.S. or T.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020