Dynamism in the fern plastome structure observed through the phylogeny

SAMULI LEHTONEN[1] and GLENDA G. CÁRDENAS [2]*

[1]Biodiversity Unit, 20014 University of Turku, Finland

[2]Department of Biology, 20014 University of Turku, Finland

*Corresponding author: glecar@utu.fi

Running title: Dynamism in the fern plastome structure

Abstract

The growing number of completely sequenced fern plastomes has revealed a structurally more dynamic genome than previously thought. Especially the boundaries of inverted repeats have undergone expansions and reductions across many fern lineages among other structural changes in plastid genome organization. In this study, we generated eight new complete fern plastomes, each representing a different family, in order to improve the sampling of plastomes in the fern order Polypodiales. We inferred phylogenetic trees using Bayesian, maximum likelihood and parsimony methods, and applying different data partitioning strategies. The trees produced under different analytical approaches were very similar, consolidating a robust hypothesis of fern phylogeny, with a few still persistent uncertainties. Structural changes in the genomes include the presence of recently found mobile open reading frames which were concentrated in inverted repeats and in large single copy *rpoB-psbZ* and *rps4-psaI* regions. The observed more common presence of mobile open reading frames in Polypodiineae in contrast to Aspleniineae may be a signal of distinct evolutionary pattern of plastome structure between these two major fern clades.

ADDITIONAL KEYWORDS: data partitioning – MORFFO – NGS – plastid genome – plastome evolution

INTRODUCTION

The introduction of molecular systematics has revolutionized our understanding on fern relationships (Christenhusz & Chase, 2014). From the early beginning, fern molecular systematics has relied on plastid genes (Hasebe et al., 1993; Pryer, Smith & Skog, 1995; Kranz & Huss, 1996). Only a few complete fern plastomes were published (eg. Wolf et al., 2003; Roper et al., 2007) until the generation of plastomes became easier by the advance of Next Generation Sequencing techniques (eg. Raman, Choi & Park, 2016; Labiak & Karol, 2017; Wei et al., 2017; Sun et al. 2017; Liu et al. 2018; Kuo et al., 2018). Currently, the number of available fern plastomes is rapidly growing, but sampling is still taxonomically very sparse in comparison to seed plants and many important fern lineages still remain completely unsampled.

As the number of published fern plastomes has increased, the previously held idea of the relative stability of plastome oganisation is now challenged (Mower & Vickrey, 2018). When the first sequence of a fern plastome was published for *Adiantum capillus-veneris* L., several rearrangements were detected relative to other vascular plants (Wolf et al., 2003). It is now clear that the inverted repeats (IRs) in ferns have experienced several changes in relation to their land plant common ancestors (Grewe et al., 2013), and that IRs in Polypodiales appear particularly dynamic (Logacheva et al., 2017). Labiak & Karol (2017) reported the loss of 31 genes from the plastome in Schizaeales, including all *ndh* genes, and together with changes in IR boundaries this has resulted in dramatically reduced small single copy (SSC) region in this fern group.

Another recently revealed dynamic pattern in fern plastome structure is the presence of mobile open reading frames (Logacheva et al., 2017; Robison et al., 2018). These Mobile Open Reading Frames in Fern Organelles (MORFFO) appear to be located adjacent to inferred sites of genomic inversions, intergenic expansions, and changes to the inverted repeats (Robison et al., 2018). It is unknown if the presence of MORFFO elements are driving the genomic restructuring, or if restructuring events are influencing the appearance of the MORFFO elements (Robison et al., 2018). It appears that MORFFO elements can move between plastid, mitochondrial, and nuclear genomes, and they may be major drivers of structural genome evolution in the fern plastomes (Robison et al., 2018).

The generation of complete plastomes is not only reveling changes in genome structure but also shedding more light on evolutionary processes of ferns (e.g. Gao et al., 2009; Gao et al., 2013; Labiak & Karol, 2017; Gitzendanner et al., 2018) and helping to infer phylogenetic relationships which have remained poorly resolved in earlier studies (Grewe et al., 2013; Wei et al., 2017; Kuo et al., 2018). However, genomic scale data are computationally challenging and the results may be compromised by analytical short cuts used to speed up the analyses. Indeed, genomic data may give false confidence and does not automatically lead to correct solution for difficult phylogenetic problems (Grewe et al., 2013; Kuo et al., 2018). One important aspect in phylogenetic inference is to incorporate evolutionary heterogeneity of the data into the analysis, through data partitioning (eg. Brown & Lemmon, 2007; Rota & Wahlberg, 2012; Frandsen et

al., 2015; Kainer & Lanfear, 2015). Alternative partitioning strategies have been used in fern plastome analyses, from analysing unpartitioned data to partitioning the data by genes and or codon positions. A novel strategy to better accommodate the rate heterogeneity was proposed by Rota et al. (2018). In this method, the sites in an alignment are grouped into partitions with uniform rate variation based on their relative evolutionary rates.

In this study, we present eight new complete fern plastomes and use them together with a set of the previously published plastomes to infer phylogenetic relationships under different analytical settings, and investigate the plastome structural evolution and presence of MORFFO elements throughout the fern phylogeny.

MATERIALS AND METHODS

DATA PRODUCTION

Eight species representing eight families of Polypodiales were selected for plastome sequencing. Total genomic DNA was extracted from silicagel dried leaves using NucleoSpin Plant II kit (Macherey-Nagel, Germany) or E.Z.N.A. SP Plant DNA Kit (Omega Bio-tek, Doraville, Georgia) following the protocol for dry samples. Success of DNA extraction was verified with NanoDrop and the DNA quality was measured using Qubit at the laboratory of Genetics at the University of Turku. DNA extractions were sent to Institute for Molecular Medicine Finland (FIMM), Helsinki, for library preparation and sequencing. DNA was fragmented

into an average size of 250-350 bp and sequenced with paired-end 101 bp reads using Illumina HiSeq 2500 platform.

Quality control of the raw reads was completed with a wrapper tool Trim Galore version 0.4.4, http://www.bioinformatics.babraham.ac.uk/projects/trim_galore, which used Cutadapt version 1.15 (Marti, 2011) to remove the poor quality portions (Phred score threshold = 20) as well as the adapters added during the library preparation from the reads.

Trimmed reads were assembled into contigs using various approaches. The first set of contigs was produced in Geneious 11.0.5 by mapping the reads to reference genomes obtained from GenBank (*Lepisorus clathratus* Ching, KY419704 and *Pseudophegopteris aurita* (Hook.) Ching, NC_035861). An additional set of contigs was built *de novo* using GetOrganelle (Jin et al., 2018) and NOVOPlasty version 2.6.7 (Dierckxsens, Mardulyn & Smits, 2016). Contigs thus obtained were assembled in Geneious using De Novo Assembly tool and manually edited when necessary. The plastome of *Tectaria panamensis* (Hook.) R.M.Tryon & A.F.Tryon could not be fully completed using the Illumina reads and the remaining gaps were filled by Sanger sequencing. The final assemblies were verified by mapping the original reads to the assemblies.

To facilitate and improve the accuracy of the assembly, we identified large single copy (LSC), small single copy (SSC) and inverted repeats (IR) using BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and Dual Organellar GenoMe Annotator (DOGMA) (Wyman, Jansen & Boore, 2004). Genomes were initially

annotated with DOGMA. Annotations were further edited by predicting RNA editing sites using PREPACT (Plant RNA Editing-Prediction & Analysis Computer Tool) v3.12.0 (Lenz et al., 2010), and by inspecting stop codons using ExPASy (Bioinformatics Resourse Portal) (Gasteiger et al., 2003). GenomeVx (Conant & Wolfe, 2008) was used to visualize the plastome of *Lomariopsis japurensis* (C.Martius) J.Sm. (Fig. 1). The complete plastome sequences are available in GenBank and the accession numbers are listed in Table 1.

PHYLOGENETIC ANALYSES

For phylogenetic analyses 41 additional fern, three seed plant, and one lycophyte plastomes were downloaded from GenBank (Table 1). Protein coding sequences were extracted in Geneious and sequences were aligned using MAFFT v7.149 (Katoh & Standley, 2013). Poorly aligned positions were removed using Gblocks v0.91b, using default settings with the exception of Allowed gap position that was set to "with half" (Castresana, 2000; Talavera & Castresana, 2007). Cleaned alignments were concatenated using SequenceMatrix v 1.7.8. (Vaidya, Lohman & Meier, 2011). The final data matrix contained 67,017 characters and is available from TreeBase (TB2:S23597).

Phylogenetic trees were inferred using parsimony, maximum likelihood (ML) and Bayesian approaches. The parsimony tree was inferred using TNT v1.5 (Goloboff & Catalano, 2016) with the following search parameters: xmult = hits 10 nocss replications 5 fuse 2 norss ratchet 10 hold 5 keepall. Nodal support

support was measured by running 1,000 jackknife replicates with the search parameters: mult=ratchet replic 10 hold 5.

The ML and Bayesian analyses were run for unpartitioned and partitioned data matrices. Data was partitioned either by sites or by genes. The partitioning by sites was based on first estimating the rate of evolution for each site using Tree-Independent Generation of Evolutionary Rates (TIGER) (Cummins & McInerney, 2011), and then applying a python script RatePartitions to group the sites to partitions according to their relative evolutionary rates (Rota et al., 2018). We used RatePartitions to create three different partition schemes by varying the division factor *d* (*d*=2.5, *d*=3.5 and *d*=4.5), resulting in the data partitioned into fewer (lower values of *d*) of larger (higher values of *d*) number of partitions. Partitionfinder v1.1.0 (Lanfear et al., 2012) was used to compare the partitioning strategies and to choose optimal evolutionary models. The optimal evolutionary model for the non-partitioned data was selected by jModelTest 2.1.10 (Guindon & Gascuel, 2003; Darriba et al., 2012). The selection for the best scheme among partitioned and unpartitioned matrices was based on Bayesian Information Criterion (BIC). Computationally heavy analyses were run either at CSC (IT Center for Science) or Cipres Science Gateway (Miller, Pfeiffer & Schwartz, 2010).

STRUCTURAL EVOLUTION OF THE FERN PLASTOME

We investigated the structural evolution of fern plastomes by mapping the inversions and shifts of the inverted repeat boundaries on the inferred phylogeny (Fig. 2). Furthermore, we conducted in Geneious a local BLASTN

(Altschul et al., 1990) search for the plastomes included in the present study to locate the three MORFFO sequences recognised by Robison et al. (2018). We used the *morffo1*, *morffo2*, and *morffo3* consensus sequences published by Robison et al. (2018) as the query sequences to locate the putatively homologous open reading frames.

RESULTS

NEW PLASTOMES

We generated and annotated eight new complete circular fern plastomes of the following species: *Davallia fejeensis* Hook., *Lindsaea line*aris Sw., *Lomariopsis japurensis* (C.Martius) J.Sm., *Nephrolepis biserrata* (Sw.) Schott, *Oleandra articulata* (Sw.) C.Presl, *Pecluma dulce* (Poir.) F.C. Assis & Salino, *Saccoloma inaequale* (Kunze) Mett., and *Tectaria panamensis* (Hook.) R.M.Tryon & A.F.Tryon. Plastomes have the typical main structure and are composed of a Large Single Copy (LSC), a Small Single Copy (SSC) and two Inverted Repeats (IR_A and IR_B) located between LSC and SSC (Fig. 1).

All the eight species have the same gene content and order with the following exceptions. *Pecluma dulce* and *Lindsaea linearis* are missing *rps16*, and *trnT-UGU* has been lost from *Lindsaea linearis*. There is only one copy of *trnN-GUU* located in the SSC in *Davallia*, in contrast to the other newly produced plastomes, where this gene is located in IRs (Figs. 1 and 2B).

The generated genomes contained 82–84 genes that encode for proteins, 28–29 that encode for tRNAs and four that encode for rRNAs. The total length of the plastomes spanned from 149,256 bp (*Lindsaea linearis*) to 172,311 bp (*Saccoloma inaequale*), the latter being the largest fern plastome published so far. Table 2 shows main features of the new plastomes.

PHYLOGENETIC ANALYSES

The best data partitioning strategy according to BIC score was TIGER+RatePartition with *d*=3.5 (in bold in Table 3) and it was used as preferred partitioning strategy in the ML and Bayesian analyses. Unpartitioned data gave the worst BIC score, followed by partitioning the data by genes. The preferred partitioning scheme splitted the data into 20 partitions for which three different evolutionary models were applied (GTR+I for one partition, SYM for 7 partitions and GTR+G for 12 partitions).

The following five phylogenetic trees were produced: Bayesian tree using the preferred data partitioning, Bayesian tree using unpartitioned data, ML tree using the preferred data partitioning, ML tree using unpartitioned data and parsimony tree. We run the model-based analyses using the best and worst data partitioning strategies to investigate the sensitivity of the results to different data partitioning strategies.

The resulting phylogenetic trees had largely similar topologies with only a few differences to be detailed below. Most of the nodes received high support

throughout the different analyses, yet, nodes conflicting among the trees had generally low support (Fig. 3A).

Bayesian inference and ML produced identical topologies using the partitioned data. As well, the trees based on the unpartitioned data were similar between these methods. In the following paragraphs, we will refer to the tree produced by Bayesian inference and ML with partitioned data as "partitioned" and "unpartitioned" refers to the tree produced by Bayesian inference and ML with unpartitioned data. There were only two topological differences between partitioned and unpartitioned trees. In the partitioned tree, *Trichomanes trollii* was resolved in a clade with *Diplopterygium glaucum*, and this clade was resolved as sisters to the remaining nonosmundaceous leptosporangiate ferns with a posterior probability (PP) of 0.98 and 60% bootstrap support (BS). In the unpartitioned tree, *Trichomanes trollii* and *Diplopterygium glaucum* were resolved as successively diverging lineages with PP=1 and BS=75. The only other difference between partitioned and unpartitioned trees was in the order in which Pteridaceae and Dennstaedtiaceae diverged. Partitioned tree supported earlier divergence of Pteridaceae (PP=1, BS=74), while in the unpartitioned tree, Dennstaedtiaceae diverged earlier (PP=0.85, BS=53) (Fig. 3A and 3B).

The parsimony tree (not shown) differed from the model based trees by a few nodes. Parsimony analysis resolved *Equisetum* as sister to all the other ferns with a 99% jackknife support whereas all the model based analyses resolved *Equisetum* as sister to a clade formed by Ophioglossaceae and Psilotaceae. Parsimony tree agrees with unpartitioned tree in resolving *Trichomanes* and

Diplopterygium as successive lineages and with partitioned tree in the relative divergence order of Pteridaceae and Dennstaedtiaceae (Figs. 3A and 3B). Within Polypodiineae, parsimony resolved *Lomariopsis* and *Nephrolepis* as sisters with 100% jackknife support in contrast to model based analyses, where they are successive lineages leading to Tectariaceae, Oleandraceae, Davalliaceae and Polypodiaceae. Within Aspleniineae, parsimony tree differs from the model based trees in the positions of *Deparia* (Athyriaceae) and *Woodsia* (Woodsiaceae). In the model based analyses Woodsia is sister to Athyriaceae + Blechnaceae + Onocleaceae clade and *Deparia* is a part of Athyriaceae, whereas in the parsimony tree *Deparia* and *Woodsia* form a clade.

STRUCTURAL EVOLUTION AND MORFFO ELEMENTS

BLASTN search returned matches to the MORFFO consensus sequences throughout the fern phylogeny (Fig. 3C). The MORFFO elements appeared particularly common in Polypodiineae and, in comparison, remarkably rare in the sister group, Aspleniineae. MORFFO elements also appear to be rarely present in the plastomes of ferns other than Polypodiales. Most of the MORFFO elements were located in the IRs and some are associated with the *rpoB-psbZ* region of LSC (Fig. 3C and 3D), both of these genomic regions having dynamic history. Additionally, we found MORFFO elements located in the *rps4-psaI* region of LSC in five species (Figs. 3C and 3D). In this region the only genomic rearrangements have been gene losses in some lineages. This region also appears to be generally rather conserved in length (~11,700 bp), except that in the species where MORFFO elements are inserted the region can be up to 5,000

bp longer. Generally, but not always, we found *morffo1*, *morffo2*, and *morffo3* to be located adjacent to one another, but in *Odontosoria chinensis* and *Hymenasplenium unilaterale* they were found at different genomic positions within the plastome.

DISCUSSION

Our nearly identical trees between different analytical approaches and data partitioning schemes imply robustness of the complete plastome data, probably due to the large number of characters. However, certain nodes in the fern phylogeny have remained notoriously difficult to resolve with confidence, and remain sensitive to data partitioning and analytical method in our analyses.

Our parsimony analysis resolved *Equisetum* as sister of all the other ferns, albeit with poor support. The same topology has been supported in recent analyses of large nuclear datasets (Rothfels et al., 2015; Qi et al., 2018; Shen et al., 2018) and a combined analysis of limited number of mitochondrial (*atp1*, *nad5*) and plastid (*atpA*, *atpB*, *matK*, *rbcL*, *rps4*) loci (Knie et al., 2015). Some of the fern plastome analyses have also supported this topology (Kim, Chung & Kim, 2014; Labiak & Karol, 2017). The position of *Equisetum* in our model-based analyses, irrespective of data partitioning, is fully supported as sister to the clade Ophioglossales + Psilotales and together they form a sister clade to the remaining ferns. This is in agreement with a growing number of plastome studies (Karol et al., 2010; Grewe et al., 2013; Kim et al., 2014; Ruhfel et al., 2014; Zhong et al., 2014; Lu et al., 2015; Gitzendanner et al., 2018; Kuo et al., 2018;

Lehtonen 2018), leaving the apparent incongruence with the nuclear DNA phylogenies puzzling.

The phylogenetic position of Hymenophyllales has remained unstable since the first molecular fern phylogenies (Hasebe et al., 1994) and despite the use of genomic data the problem still remains (Kuo et al., 2018; Lehtonen, 2018). Both our parsimony and unpartitioned analyses resolved Hymenophyllales as sister to all other nonosmundalean leptosporangiate ferns. This resolution agrees with Lehtonen (2018), whose plastome data was unpartitioned and analysed with ML using GTR+G model, and Kuo et al. (2018), who partitioned their data by genes and codon position and inferred the tree with ML and best fit models for each partition. As well, Gitzendanner et al. (2018) revealed the same topology, although without any support. Their massive dataset necessitated serious analytical shortcuts which may have compromised some of the results. In contrast, our partitioned analyses resolved Hymenophyllales forming a clade with Gleicheniales, but without strong support. This topology is in agreement with Lehtonen et al. (2017), who used four plastid markers to infer Bayesian time calibrated phylogeny. In addition, the clade formed by Hymenophyllales and Gleicheniales is also supported by large nuclear data sets (Qi et al., 2018; Shen et al., 2018), although in these studies Gleicheniales was not resolved monophyletic.

Among the more derived ferns the position of *Saccoloma* (Lehtonen, Wahlberg & Christenhusz, 2012) and the branching order of Pteridaceae and Dennstaedtiaceae (Schuettpelz & Pryer, 2007; Wolf et al., 2015) have remained a

source of debate. All our analyses resolved *Saccoloma* forming a clade with Lindsaeaceae and this position was also fully supported in the best sampled nuclear phylogeny so far (Qi et al., 2018). It thus seems that the phylogenetic position of *Saccoloma* can now be considered resolved. Our parsimony and partitioned trees agree in Pteridaceae diverging before Dennstaedtiaceae, a topology that is now generally accepted (PPG I, 2016) and also fully supported in the analyses of nuclear data (Qi et al., 2018; Shen et al., 2018). Alternative resolutions for these families have been supported by some earlier studies (Schuettpelz & Pryer, 2007) and by our unpartitioned trees.

Within Polypodiineae the relationship between Lomariopsidaceae and Nephrolepidaceae has remained controversial. While some analyses have supported a sister relationship (e.g. Tsutsumi & Masahiro, 2006; Schuettpelz & Pryer, 2007; Li et al., 2009; Hennequin et al., 2010; Zhang et al., 2016), including our parsimony analysis, others have resolved these families as successively branching lineages with Lomariopsidaceae diverging earlier (e.g. Lehtonen, 2011; Christenhusz, Jones & Lehtonen, 2013; Liu et al., 2013). This latter topology is supported by nuclear genomic data (Qi et al., 2018; Shen et al., 2018) and by our model-based analyses of plastome data. Additional taxon sampling of genomic data is desired, since the analyses supporting the sister relationship generally have a broader sampling of taxa, whereas the alternative topology is generally supported by studies sampling fewer taxa but much more character data.

Previous studies in fern plastome phylogenetics (Gao et al., 2009; Labiak & Karol, 2017; Wei et al., 2017; Kuo et al., 2018; Lehtonen et al., 2018) have generally partitioned their data by genes and sometimes codon positions, or analysed the data unpartitioned. We found unpartitioning and partitioning by genes to be the clearly suboptimal data partitioning strategies based on BIC scores, thus supporting the similar conclusions by Rota et al. (2018). However, we did not evaluate partitioning our data by codon positions, which Rota et al. (2018) found also to be a suboptimal strategy. In our study, the TIGER+RatePartition strategy with *d*=3.5 was found to be the most suitable partitioning scheme for our data, among the ones tried. Apparently, partitioning according to the relative evolutionary rate of individual sites better represents the evolutionary variation of the data than partitioning by genes, which may result in a highly heterogeneous information within partitions since evolutionary rate may greatly differ within a gene.

We did not find a phylogenetic pattern regarding to genes missing from the newly generated plastomes but these genes have repeatedly been lost in other fern lineages. Thus, *rps16* lost in *Pecluma* and *Lindsaea* was also reported to be missing from *Ophioglossum*, *Psilotum* and *Equisetum* plastomes (Grewe et al., 2013). The *trnT-UGU* gene absent from *Lindsaea* has been lost in various fern lineages including *Hymenophyllum*, *Callistopteris*, *Dipteris* (Kuo et al., 2018), and from the vittarioid ferns (Robison et al., 2018). We found *trnR-UCG* gene located between *rbcL* and *accD* genes in all the newly generated genomes of Polypodiales, although this gene has thought to be exclusive to tree ferns (Gao et al., 2009; Raman et al., 2016), non-core leptosporangiates, and basal ferns

(Raman et al., 2016). The position of *trnN-GUU* gene in the SSC in *Davallia*, instead in IRs, is a novel arrangement within the Polypodiales and has previously been reported only in *Schizaea* (Labiak & Karol, 2017).

The ancestral fern plastome probably had the same structure with the collinear plastomes of Equisetales, Ophioglossales and Osmundales (Mower & Vickrey, 2018). Within the early diverging ferns this structure has been modified in Psilotales by expansions of IR_A to LSC and IR_B to SSC (Grewe et al., 2013, Fig. 2A). Apparently independent expansion of IR_A to LSC followed by an expansion of IR_B to LSC have taken place in Marattiales (Roper et al., 2007, Fig. 2A). The IR_A-LSC border appears to be especially labile in fern plastomes, as the border has obviously shifted various times in the lineages leading to nonosmundalean leptosporangiates. Shift in the border location is related to an inversion affecting *ndhB-rps12* in Gleicheniales (Kim et al., 2014, Fig. 2A). A major reorganisation of the border region has taken place in the lineage ancestral to Schizaeales and their sister group, involving the long-recognised double inversion with apparent expansions-contractions of IRs to explain the inclusion of *psbA-ycf2* into IRs and reverted orientation of *ndhB* (Fig. 2B).

Within Schizaeales, *Lygodium* appears to have an unmodified plastome organization in relation to the extremely modified plastomes of *Actinostachys* and *Schizaea* (Labiak & Karol, 2017, Fig. 2B). The latter plastomes have lost a large number of genes and IR expansion-contractions have affected the whole SSC and the borders of LSC and IRs. Salviniales, Cyatheales and Polypodiales share a double inversion affecting the *rpoB-psbZ* region (Gao et al., 2011, Fig. 2B).

These orders share an generally identical plastome, but in *Marsilea* the complete *ndhB* gene is located in LSC whereas in Cyatheales and Polypodiales it is partially located in the IRs (Gao et al., 2013, Fig. 2B). Furthermore, in *Plagiogyria* an inversion within the *rpoB-psbZ* region has resulted in a gene order similar to the putative intermediate of the earlier double inversion of the region (Gao et al., 2011, Fig. 2B). We observed a novel structural change in the plastome of *Davallia*, where *trnN* gene has been excluded from the IRs and is located in SSC instead (Fig. 2B).

Similar to what Robison et al. (2018) found in Pteridaceae, we found that MORFFO elements are often located in the regions of most dynamic plastome structure, close to the sites of inversions or IR borders. The association of MORFFOs with these dynamic regions implies their evolutionary significance in the plastome structure evolution. However, while MORFFO elements were always located near to each other in Pteridaceae (Robison et al., 2018), we found them in a far away positions in *Odontosoria* and *Hymenasplenium*. In *Odontosoria*, *morffo1* and *morffo2* were located in the LSC between *psbM-petN* (within the *rpoB-psbZ* region), but *morffo3* was found in IRs. In *Hymenasplenium morffo1* was located in IRs but *morffo3* between *rps4* and *trnL* (within *rps4-psaI* region) in the LSC. This finding suggest that MORFFOs can move independently from each other, but the reason why they are mostly found together remains unclear and a matter of further investigation. A novel finding in this study was the presence of MORFFO elements within the *rps4-psaI* region. Previously, this region has not been recognised as a hot spot for genomic rearrangements, but we found that it

has experienced some gene losses and has been greatly expanded with the insertion of MORFFO elements in five plastomes (Fig. 3B).

The size variation of fern plastomes seems to be closely associated to the presence of MORFFO elements. The largest fern plastome so far recorded, *Saccoloma inaequale* (172,311 bp), has all three MORFFO elements in IR, thus doubling the size effect. Moreover, the *atpA-trnR-UCU* and *petA-psbJ* intergenic regions were ~1,000 bp longer than in close relatives. These regions contained several open reading frames which may turn out to be MORFFO elements. In contrast, the closely related *Lindsaea linearis* had no MORFFO elements and it has one of the smallest fern plastomes (149,256 bp). Similarly, within the Polypodiineae, species lacking MORFFOs have smaller plastomes (149,643–151,684 bp) than the species with MORFFOs (152,479–163,068 bp).

The presence of MORFFO elements in more than half of the sampled Polypodiineae plastomes, in contrast to only two out of 17 sampled in Aspleniineae, is suggestive of distintive evolutionary pattern between these two clades. Alternatively, it may also be the artifact of the still very limited sampling of complete plastomes. Could the rarity of MORFFO elements in Aspleniineae predict higher plastome stability in contrast to other Polypodiales with an apparently more frequent presence of MORFFO elements? Clarifying this and related questions require the generation of taxonomically much more densely sampled plastome data.

ACKNOWLEDGMENTS

We thank Willi Hennig Society for making TNT freely available. The authors wish to acknowledge CSC – IT Center for Science, Finland, for generous computational resources. This study was funded by grants from Kone Foundation, Oskar Öflunds Foundation, Department of Biology of the University of Turku to GC and a Finnish Cultural Foundation grant to SL.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology* **215**: 403–410.

Brown JM, Lemmon AR. 2007. The importance of data partitioning and the utility of bayes factors in Bayesian phylogenetics. *Systematic Biology* **56**: 643–655.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**: 540–552.

Christenhusz MJM, Jones M, Lehtonen S. 2013. Phylogenetic placement of the enigmatic fern genus *Dracoglossum*. *American Fern Journal* **103**: 131–138.

Christenhusz MJM, Chase MW. 2014. Trends and concepts in fern classification. *Annals of Botany* **0**: 1–24.

Conant GC, Wolfe KH. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. *Bioinformatics* **24**: 861–862.

Cummins CA, McInerney JO. 2011. A method for inferring the rate of evolution of homologous characters that can potentially improve phylogenetic

inference, resolve deep divergence and correct systematic biases. *Systematic Biology* **60**: 833–844.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods* **9**: 772.

Der JP. 2010. Genomic perspectives on evolution in bracken fern. Dotoral thesis, Utah: Utah State University.

Dierckxsens N, Mardulyn P, Smits G. 2016. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Research* **45**: e 18.

Frandsen PB, Calcott B, Mayer C, Lanfear R. 2015. Automatic selection of partitioning schemes for phylogenetic analyses using iterative k-means clustering of site rates. *BMC Evolutionary Biology* **15**: 13.

Gao L, Yi X, Yang Y-X, Su Y-J, Wang T. 2009. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. *BMC Evolutionary Biology* **9**: 130.

Gao L, Zhou Y, Wang Z-W, Su Y-J, Wang T. 2011. Evolution of the *rpoB-psbZ* region in fern plastid genomes: notable structural rearrangements and highly variable intergenic spacers. *BMC Plant Biology* **11**: 64.

Gao L, Wang B, Wang Z-W, Zhou Y, Su Y-J, Wang T. 2013. Plastome sequences of *Lygodium japonicum* and *Marsilea crenata* reveal the genome organization transformation from basal ferns to core Leptosporangiates. *Genome Biology and Evolution* **5**: 1403–1407.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research* **31**: 3784–3788.

Gitzendanner MA, Soltis PS, Wong GK-S, Ruhfel BR, Soltis DE. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* **0**: 1–11.

Goloboff P, Catalano S. 2016. TNT, version 1.5, with a full implementation of phylogenetic morphometrics. *Cladistics* **32**: 221–238.

Goremykin VV, Hirsch-Ernst KI, Wölf S, Hellwig FH. 2003. Analysis of the Amborella trichopoda chloroplast genome sequence suggests that Amborella is not a basal angiosperm. *Molecular Biology and Evolution* **20**: 1499–1505.

Grewe F, Guo W, Gubbels EA, Hansen AK, Mower JP. 2013. 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 Evolutionary Biology* **13**: 8.

Guindon S, Gascuel O. 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. *Systematic Biology* **52**: 696–704.

Hasebe M, Ito M, Kofuji R, Ueda K, Iwatsuki K. 1993. Phylogenetic relationships of ferns deduced from *rbcL* gene sequence. *Journal of Molecular Evolution* **37**: 476–482.

Hasebe M, Omori T, Nakazawa M, Sano T, Kato M, Iwatsuki K. 1994. *rbcL* gene sequences provide evidence for the evolutionary lineages of leptosporangiate ferns. *Proceedings of the National Academy of Sciences* **91**: 5730–5734.

Hennequin S, Hovenkamp P, Christenhusz MJM, Schneider H. 2010. Phylogenetics and biogeography of *Nephrolepis* – a tale of old settlers and young tramps. *Botanical Journal of the Linnean Society* **164**: 113–127.

Jin J-J, Yu W-B, Yang J-B, Song Y, Yi T-S, Li D-Z. 2018. GetOrganelle: a simple and fast pipeline for de novo assembly of a complete circular chloroplast genome using genome skimming data. *BioRxiv* doi:10.1101/256479.

Kainer D, Lanfear R. 2015. The effects of partitioning on phylogenetic inference. *Molecular Biology and Evolution* **32**: 1611-1627.

Karol KG, Arumuganathan K, Boore JL, Duffy AM, Everett KDE, Hall JD, Hansen SK, Kuehl JV, Mandoli DF, Mishler BD, Olmstead RG, Renzaglia KS, Wolf PG. 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evolutionary Biology* **10**: 321.

Katoh K, Standley D M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30**: 772–780.

Kim HT, Chung MG, Kim KJ. 2014. Chloroplast genome evolution in early diverged Leptosporangiate ferns. *Molecules and cells* **37**: 372–382.

Kim HT, Kim K-J. 2018. Evolution of six novel ORFs in the plastome of *Mankyua chejuense* and phylogeny of eusporangiate ferns. *Scientific Reports* **8**: 16466.

Knie N, Fischer S, Grewe F, Polsakiewicz M, Knoop V. 2015. Horsetails are the sister group to all other monilophytes and Marattiales are sister to

leptosporangiate ferns. *Molecular Phylogenetics and Evolution* **90**: 140–149.

Kranz HD, Huss VAR. 1996. Molecular evolution of pteridophytes and their relationship to seed plants: evidence from complete 18S rRNA gene sequences. *Plant Systematics and Evolution* **202**: 1–11.

Kuo L-Y, Qi X, Ma H, Li F-W. 2018. Order- level fern plastome phylogenomics: new insights from Hymenophyllales. *American Journal of Botany* **105**: 1–11.

Labiak PH, Karol KG. 2017. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. *American Journal of Botany* **104**: 1008–1018.

Lanfear R, Calcott B, Ho SYW, Guindon S. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Molecular Biology and Evolution* **29**: 1695–1701.

Lehtonen S. 2011. Towards resolving the complete fern tree of life. *PLoS ONE* **6**: 1–5.

Lehtonen S, Wahlberg N, Christenhusz MJM. 2012. Diversification of lindsaeoid ferns and phylogenetic uncertainty of early polypod relationships. *Botanical Journal of the Linnean Society* **170**: 489–503.

Lehtonen S, Silvestro D, Karger DN, Scotese Ch, Tuomisto H, Kessler M, Peña C, Wahlberg N, Antonelli A. 2017. Environmentally driven extinction and opportunistic origination explain fern diversification patterns. *Scientific reports* **7**: 4831.

Lehtonen S. 2018. The complete plastid genome sequence of *Trichomanes trollii* (Hymenophyllaceae). *Nordic Journal of Botany* **36**: e02072.

Lenz H, Rüdinger M, Volkmar U, Fischeret S, Herres S, Grewe F, Knoop V. 2010. Introducing the plant RNA editing prediction and analysis computer tool PREPACT and an update on RNA editing site nomenclature. *Current Genetics* **56**: 189–201.

Li F-W, Tan BC, Buchbender V, Moran RC, Rouhan G, Wang Ch-N, Quandt D. 2009. Identifying a mysterious aquatic fern gametophyte. *Plant Systematics and Evolution* **281**: 77–86.

Liu H-M, Jiang R-H, Guo J, Hovenkamp P, Perrie LR, Shepherd L, Hennequin S, Schneider H. 2013. Towards a phylogenetic classification of the climbing fern genus *Arthropteris*. *Taxon* **62**: 688–700.

Liu S, Ping J, Wang Z, Wang T, Su Y. 2018. Complete chloroplast genome of the tree fern *Alsophila podophylla* (Cyatheaceae), *Mitochondrial DNA Part B* **3**: 48–49.

Logacheva MD, Krinitsina AA, Belenikin MS, Khafizov K, Konorov EA, Kuptsov SV, Speranskaya AS. 2017. Comparative analysis of inverted repeats of polypod fern (Polypodiales) plastomes reveals two hypervariable regions. *BMC Plant Biology* **17**(Suppl 2): 255.

Lu JM, Zhang N, Du XY, Wen J, Li DZ. 2015. Chloroplast phylogenomics resolves key relationships in ferns. *Journal of Systematics and Evolution* **53**: 448–457.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal* **17**: 10–12.

Miller MA, Pfeiffer W, Schwartz T. 2010. "Creating the CIPRES Science Gateway for inference of large phylogenetic trees" in Proceedings of the

Gateway Computing Environments Workshop (GCE), New Orleans, LA pp 1–8.

Mower JP, Vickrey TL. 2018. Chapter nine - Structural diversity among plastid genomes of land plants. In: Chaw S-M, Jansen RK, eds. *Advances in Botanical Research*. United Kingdom: Academic Press, 263–292.

PPG I. 2016. A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution* **54**: 563–603.

Pryer KM, Smith AR, Skog JE. 1995. Phylogenetic relationship of extant ferns based on evidence from morphology and *rbcL* sequences. *American Fern Journal* **85**: 205–282.

Qi X, Kuo L-Y, Guo Ch, Li H, Li Z, Qi J, Wang L, Hu Y, Xiang J, Zhang C, Guo J, Huang Ch-H, Ma H. 2018. A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Molecular Phylogenetics and Evolution* **127**: 961–977.

Raman G, Choi KS, Park S. 2016. Phylogenetic relationships of the fern *Cyrtomium falcatum* (Dryopteridaceae) from Dokdo island based on chloroplast genome sequencing. *Genes* **7**: 115.

Robison TA, Grusz AL, Wolf PG, Mower JP, Fauskee BD, Sosa K, Schuettpelz E. 2018. Mobile elements shape plastome evolution in ferns. *Genome Biology and Evolution* **10**: 2558–2571.

Roper JM, Hansen SK, Wolf PG, Karol KG, Mandoli DF, Everett KDE, Kuehl J, Boore JL. 2007. The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffin. (Marattiaceae). *American Fern Journal* **97**: 95–106.

Rota J, Wahlberg N. 2012. **Exploration** of data partitioning in an eight-gene data set: phylogeny of metalmark moths (Lepidoptera, Choreutidae). *Zoologica Scripta* **41**: 536–546.

Rota J, Malm T, Chazot N, Peña C, Wahlberg N. 2018. A simple method for data partitioning based on relative evolutionary rates. *PeerJ* **6**: e5498.

Rothfels CJ, Li F-W, Sigel EM, Huiet L, Larsson A, Burge DO, Ruhsam M, Deyholos M, Soltis DE, Stewart JrCN, Shaw ShW, Pokorny L, Chen T, dePamphilis C, DeGironimo L, Chen L, Wei X, Sun X, Korall P, Stevenson DW, Graham SW, Wong GK-S, Pryer KM. 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *American Journal of Botany* **102**: 1089–1107.

Ruhfel BR, Gitzendanner MA, Soltis PS, Soltis DE, Burleigh JG. 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* **14**: 23.

Schuettpelz E, Pryer KM. 2007. Fern phylogeny inferred from 400 leptosporangiate species and three. *Taxon* **56**: 1037–1050.

Shen H, Jin D, Shu J-P, Zhou X-L, Lei M, Wei R, Shang H, Wei H-J, Zhang R, Liu L, Gu Y-F, Zhang X-Ch, Yan Y-H. 2018. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience* **7**: 1–11.

Sun M, Li J, Li D, Shi L. 2017. Complete chloroplast genome sequence of the medical fern *Drynaria roosii* and its phylogenetic analysis. *Mitochondrial DNA Part B* 2: 7–8.

Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564–577.

Tsutsumi C, Kato M. 2006. Evolution of epiphytes in Davalliaceae and related ferns. *Botanical Journal of the Linnean Society* **151**: 495–510.

Vaidya G, Lohman DJ, Meier R. 2011. SequenceMatrix: concatenation software for the fast assembly of multigene datasets with character set and codon information. *Cladistics* **27**: 171–180.

Wei R, Yan Y-H, Harris AJ, Kang J-S, Shen H, Xiang Q-P, Zhang X-C. 2017. Plastid phylogenomics resolve deep relationships among eupolypod II ferns with rapid radiation and rate heterogeneity. *Genome Biology and Evolution* **9**: 1646–1657.

Wolf PG, Rowe CA, Sinclair RB, Hasebe M. 2003. Complete nucleotide sequence of the chloroplast genome from a Leptosporangiate fern, *Adiantum capillus-veneris* L. *DNA Research* **10**: 59–65.

Wolf PG, Sessa EB, Marchant DB, Li F-W, Rothfels CJ, Sigel EM, Gitzendanner MA, Visger CJ, Banks JA, Soltis DE, Soltis PS, Pryer KM, Der JP. 2015. An exploration into fern genome space. *Genome Biology and Evolution* **7**: 2533–2544.

Wolf PG, Der JP, Duffy AM, Davidson JB, Grusz AL, Pryer KM. 2011. The evolution of chloroplast genes and genomes in ferns. *Plant Molecular Biology* **76**: 251–261.

Wu CS, Wang YN, Liu SM, Chaw SM. 2007. Chloroplast genome (cpDNA) of *Cycas taitungensis* and 56 cp protein-coding genes of *Gnetum parvifolium*:

Insights into cpDNA evolution and phylogeny of extant seed plants. *Molecular Biology and Evolution* **24**: 1366–1379.

Wyman SK, Jansen RK, Boore JL. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* **20**: 3252–3255.

Xu R, Liu S, Wang Z, Wang T, Su Y. 2018. The first complete chloroplast genome of a traditional Chinese medicinal herb *Odontosoria chinensis* (Lindsaeaceae). *Mitochondrial DNA Part B* **3**: 292–293.

Zhang L, Schuettpelz E, Rothfels CJ, Zhou X-M, Gao X-F, Zhang L-B. 2016. Circumscription and phylogeny of the fern family Tectariaceae based on plastid and nuclear markers, with the description of two new genera: *Draconopteris* and *Malaifilix* (Tectariaceae). *Taxon* **65**: 723–738.

Zhong B, Fong R, Collins LJ, McLenachan PA, Penny D. 2014. Two new fern chloroplasts and decelerated evolution linked to the long generation time in tree ferns. *Genome Biology and Evolution* **6**: 1166–1173.

Figures and tables legends

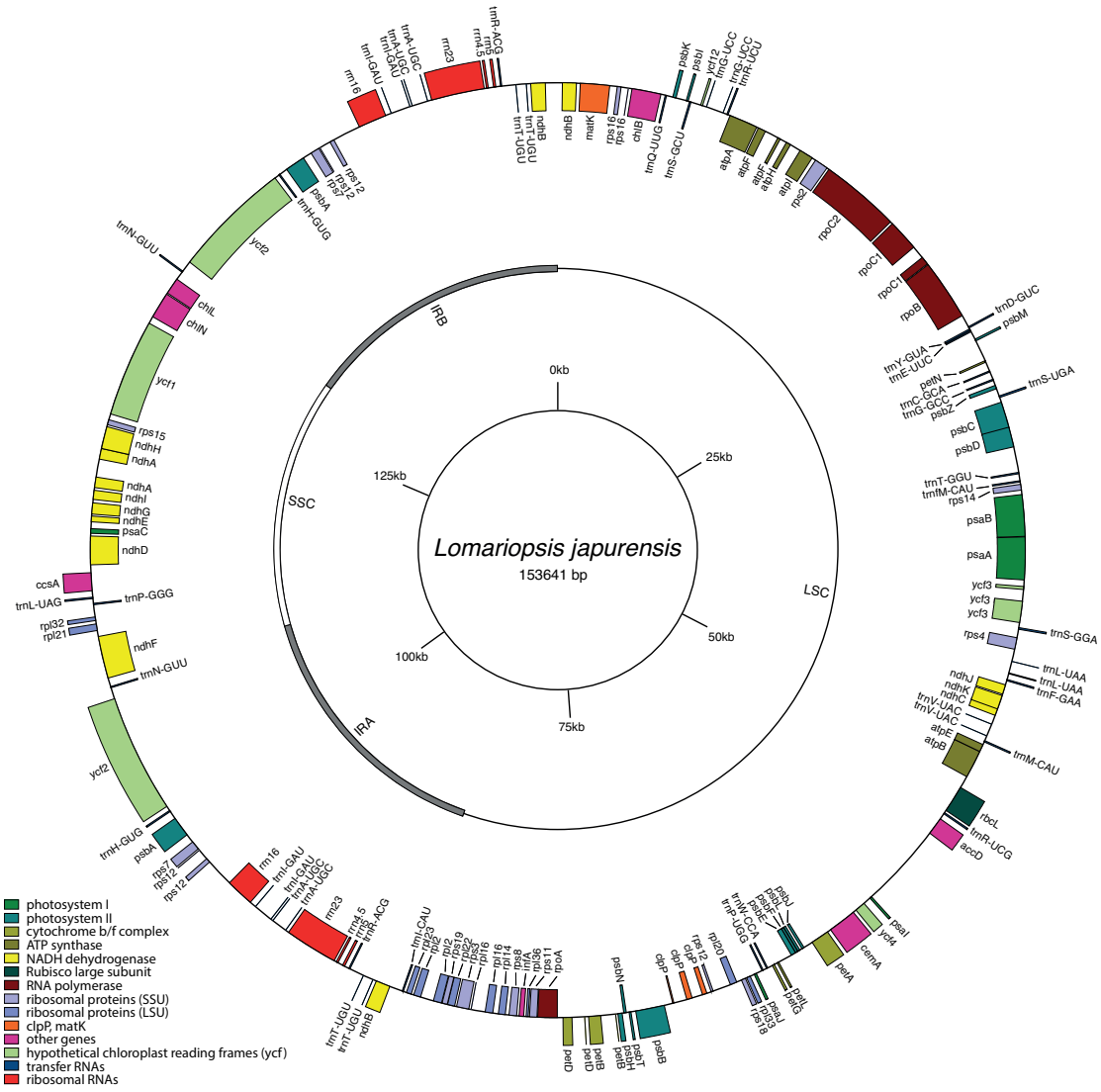

Figure 1. Visualisation of *Lomariopsis japurensis* plastome. The inner circle provides a scale, the middle circle indicates the locations of Large Single Copy (LSC), Small Single Copy (SSC), and Inverted Repeats (IR_A and IR_B), and the outer circle indicates the genes coded by functional groups as shown in the key. Genes on the inside are transcribed counterclockwise and genes on the outside are transcribed clockwise

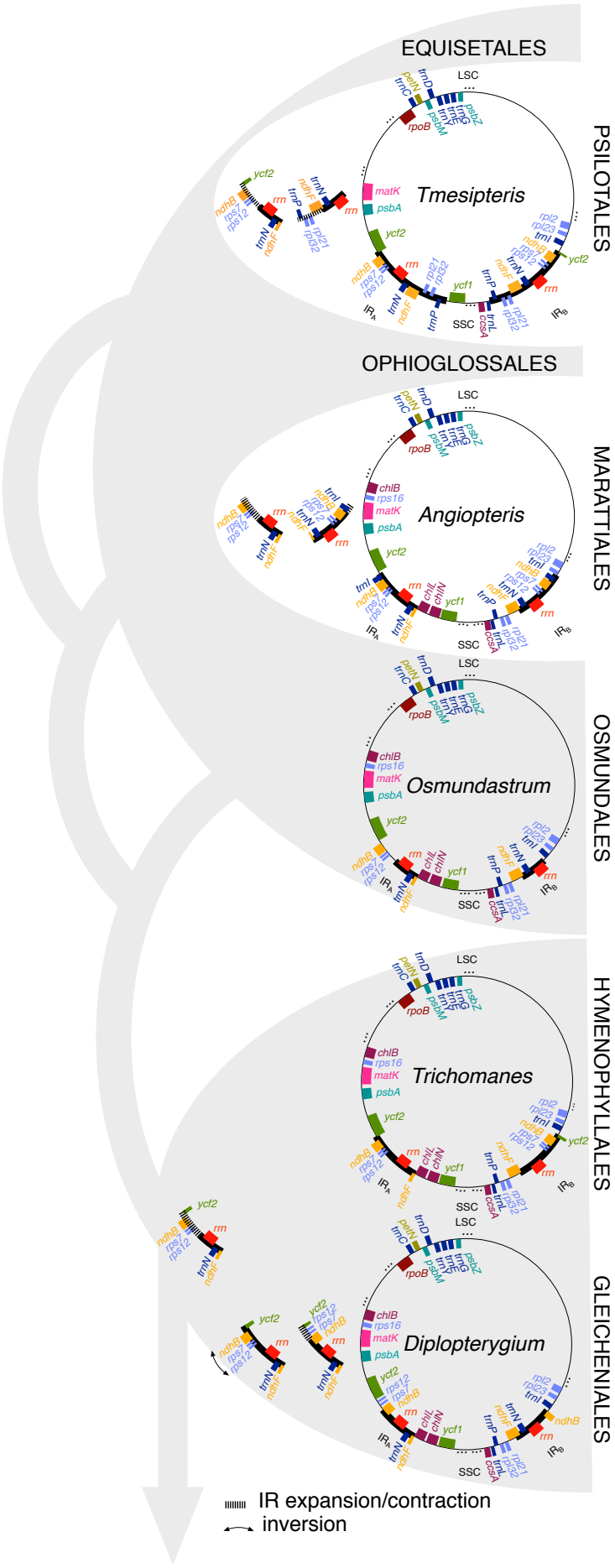
EQUISETALES
PSILOTALES
Tmesipteris
OPHIOGLOSSALES
MARATTIALES
Angiopteris
OSMUNDALES
Osmundastrum
HYMENOPHYLLALES
Trichomanes
GLEICHENIALES
Diplopterygium
LSC
SSC
IR expansion/contraction
inversion

Figure 2A. Schematic plastome maps displayed on the phylogeny indicating putative IR expansions/contractions and inversion events. Continues in figure 2B

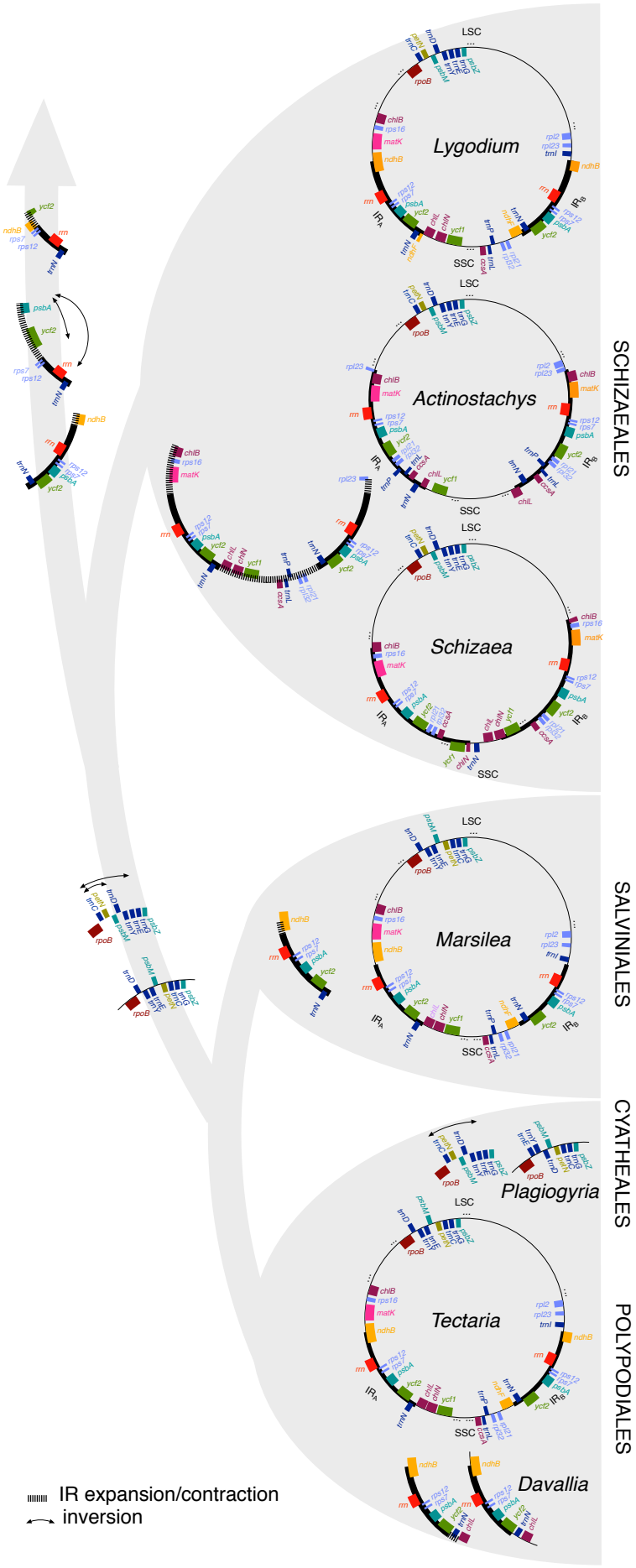
Lygodium
Actinostachys
Schizaea
Marsilea
Plagiogyria
Tectaria
Davallia
SCHIZAEALES
SALVINIALES
CYATHEALES
POLYPODIALES
IR expansion/contraction
inversion

Figure 2B. Continued from figure 2A

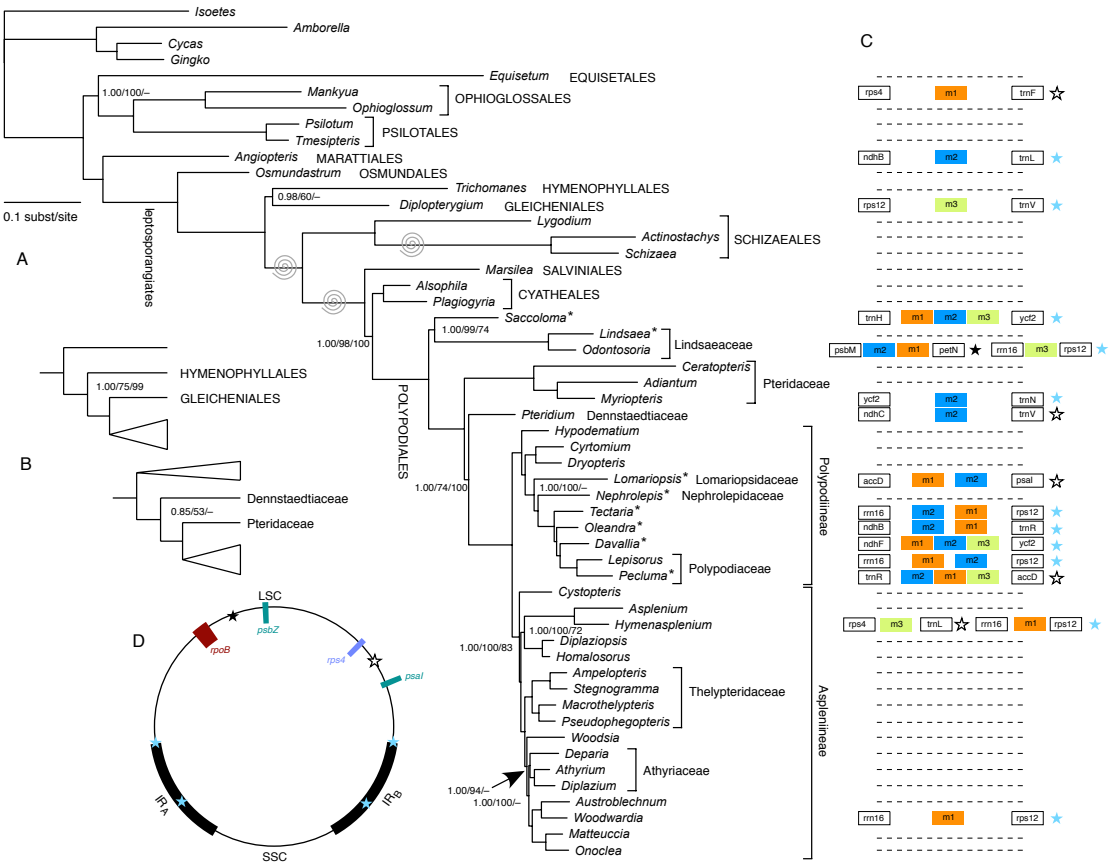

Figure 3. A. Phylogeny from the partitioned Maximum Likelihood analysis. Nodal support values (Bayesian PP in partitioned analysis/ML bootstrap in partitioned analysis/parsimony jackknife) are only given to nodes not fully supported. The whorls indicate branches along which major plastome reorganizations have taken place. Plastomes newly produced in this study are marked with an asterisk. B. Alternative topological arrangements obtained under unpartitioned Bayesian and ML analyses. Support values as in (A). C. Distribution and location of MORFFO elements in the fern phylogeny. Dashed line indicates that no MORFFO elements were found in the BLAST search. D. Generalized plastome map showing the approximate locations of MORFFO elements as indicated by the stars

Table 1. List of plastomes used in this article and GenBank accession numbers

Taxa	GenBank accession codes	Published
Ingroup		
Actinostachys pennula Hook.	KU764518	Labiak & Karol 2017
Adiantum capillus-veneris L.	NC_004766	Wolf et al. 2003
Alsophila spinulosa (Hook.) R.M.Tryon	NC_012818	Gao et al. 2009
Ampelopteris prolifera (Retz.) Copel.	NC_035835	Wei et al. 2017
Angiopteris evecta (G.Forst.) Hoffm.	DQ821119	Roper et al. 2007
Asplenium pekinense Hance	NC_035837	Wei et al. 2017
*Athyrium sinens*e Rupr.	NC_035839	Wei et al. 2017
Austroblechnum melanocaulon (Brack.) Gasper & V.A.O.Dittrich	NC_035840	Wei et al. 2017
Ceratopteris richardii Brongn.	KM052729	Marchant et al. unpublished
Cyrtomium devexicapulae (Koidz.) Ching	NC_028542	Lu et al. 2015
Cystopteris chinensis (Ching) X.C. Zhang & R. Wei	NC_035843	Wei et al. 2017
Davallia fejeensis Hook.	MK705750	This study
Deparia lancea (Thunb.) Fraser-Jenk.	NC_035844	Wei et al. 2017
Diplaziopsis javanica (Blume) C.Chr.	NC_035848	Wei et al. 2017
Diplazium bellum (C.B.Clarke) Bir in Mehra & Bir	NC_035849	Wei et al. 2017
Diplopterygium glaucum (Thunb. ex Houtt.) Nakai	NC_024158	Kim et al. 2014
Dryopteris decipiens (Hook.) Kuntze	NC_035854	Wei et al. 2017
Equisetum arvense L.	NC_014699	Karol et al. 2010
Homalosorus pycnocarpos (Spreng.) Pic.Serm.	NC_035855	Wei et al. 2017
Hymenasplenium unilaterale (Lam.) Hayata	NC_035856	Wei et al. 2017
Hypodematium crenatum Kuhn & Decken	NC_035857	Wei et al. 2017
Lepisorus clathratus Ching	NC_035739	Wei et al. 2017
Lindsaea linearis Sw.	MK705751	This study
Lomariopsis japurensis (C.Martius) J.Sm.	MK705752	This study
Lygodium japonicum (Thunb.) Sw.	KC536645	Gao et al. 2013
Macrothelypteris torresiana (Gaudich.) Ching	NC_035858	Wei et al. 2017
Mankyua chejuensis B.Y.Sun, M.H.Kim & C.H.Kim	KP205433	Kim & Kim 2018
Marsilea crenata C.Presl	NC_022137	Gao et al. 2013
Matteuccia struthiopteris (L.) Tod.	NC_035859	Wei et al. 2017
Myriopteris lindheimeri (Hook.) J. Sm.	NC_014592	Wolf et al. 2011
Nephrolepis biserrata (Sw.) Schott	MK705753	This study
Odontosoria chinensis (L.) J.Sm.	MG913608	Xu et al. 2018
Oleandra articulata (Sw.) C.Presl	MK705754	This study
Onoclea sensibilis L.	NC_035860	Wei et al. 2017
Ophioglossum californicum Prantl	NC_020147	Grewe et al. 2013
Osmundastrum cinnamomeum (L.) C.Presl	NC_024157	Kim et al. 2014
Pecluma dulce (Poir.) F.C. Assis & Salino	MK705755	This study
Plagiogyria glauca (Blume) Mett.	KP136831	Wolf et al. 2015

Pseudophegopteris aurita (Hook.) Ching	NC_035861	Wei et al. 2017
Psilotum nudum (L.) P.Beauv.	KC117179	Grewe et al. 2013
Pteridium aquilinum subsp. *aquilinum* (L.) Kuhn	NC_014348	Der 2010
Saccoloma inaequale (Kunze) Mett.	MK705756	This study
Schizaea elegans (Vahl) Sw.	NC_035807	Labiak & Karol 2017
Stegnogramma sagittifolia (Ching) L.J.He & X.C.Zhang	NC_035863	Wei et al. 2017
Tectaria panamensis (Hook.) R.M.Tryon & A.F.Tryon	MK705757	This study
Tmesipteris elongata P.A.Dang.	KJ569699	Zhong et al. 2014
Woodsia polystichoides D.C.Eaton	NC_035865	Wei et al. 2017
Woodwardia unigemmata (Makino) Nakai	NC_028543	Lu et al. 2015
Outgroup		
Amborella trichopoda Baill.	NC_005086	Goremykin et al. 2003
Cycas taitungensis C.F.Shen, K.D.Hill, C.H.Tsou & C.J.Chen	NC_009618	Wu et al. 2007
Gingko biloba L.	NC_016986	Li et al. unpublished
Isoetes flaccida Shuttlew.	NC_014675	Karol et al. 2010

Table 2. The main features of the newly produced plastomes

Taxa	Collector number (Herbarium)	LSC (bp)	SSC (bp)	IR (bp)	Genome size (bp)	Genes encode proteins	Genes encode rRNA	Genes encode tRNA	GC Freq	GC %	Coverage
Davallia fejeensis	HBG_0044-0736	81799	25704	22488	152479	84	4	29	62872	41.2	116
Lindsaea linearis	JB_589 (AAU)	81226	20592	23719	149256	82	4	28	59438	39.8	1472
Lomariopsis japurensis	SL_989 (TUR)	85152	21811	23339	153641	84	4	29	121	34.8	47
Nephrolepis biserrata	MJ_1189 (TUR)	82328	21433	22941	149643	84	4	29	126	35.3	242
Oleandra articulata	SL_950 (TUR)	82316	21594	29579	163068	84	4	28	67845	41.6	358
Pecluma dulce	MJ_919 (TUR)	86118	21453	23232	154035	83	4	29	62681	40.7	677
Saccoloma inaequale	MJ_1019 (TUR)	85158	21493	32830	172311	84	4	29	73158	42.5	108
Tectaria panamensis	MJ_1052 (TUR)	82698	21726	27116	158656	84	4	29	79	41.8	40

HBG = Helsinki Botanical Garden, JB = J.E. Braggins, MJ = M. Jones, SL = S. Lehtonen

Table 3. Partition strategies applied to the data and their corresponding BIC (Bayesian Information Criterion) score

Partition strategy	BIC score	Number of partitions
non-partition	2403619.06850	1
partition by site d=2.5	2287832.67120	14
partition by site d=3.5	**2287591.65526**	20
partition by site d=4.5	2287829.88921	27
partition by genes	2399336.51183	84