The complete plastid genome sequence of *Trichomanes trollii* (Hymenophyllaceae)

Samuli Lehtonen

S. Lehtonen (http://orcid.org/0000-0001-6235-2026) (samile@utu.fi) Biodiversity Unit, University of Turku FI-20014 Turku, Finland.

Abstract

Plastid genomes have become an important source of information for clarifying phylogenetic relationships especially where traditional molecular systematic studies have failed to produce strongly supported hypotheses. The exact phylogenetic position of an early diverging fern order Hymenophyllales has remained uncertain due to poor support in published phylogenetic studies. High-throughput sequencing of the first complete plastid genome of the genus *Trichomanes* revealed genome structure similar to the closest relatives and the phylogenetic analysis resulted in equally poorly supported topology as the previous studies that were based on only a few molecular markers. It seems that, at least with the current poor taxonomic sampling of complete plastid genomes, the exact relationships between Hymenophyllales, Gleicheniales, and other nonosmundalean leptosporangiate ferns cannot be firmly established.

Introduction

The phylogenetic relationships of extant ferns are now understood well enough so that higher-level classifications are converging into an apparently natural taxonomy (Smith et al. 2006, Christenhusz et al. 2011, Christenhusz and Chase 2014, PPG I 2016). Taxonomically broad molecular systematic investigations in ferns have mostly applied only a relatively few plastid markers (e.g. Schuettpelz and Pryer 2007, Lehtonen 2011, Testo and Sundue 2016), but complete plastid genomes (plastomes) have been increasingly used in recent years to verify the backbone of fern phylogeny (e.g. Gao et al. 2013, Grewe et al. 2013, Lu et al. 2015, Labiak and Karol 2017, Wei et al. 2017). Together with other high-throughput sequencing approaches (Rothfels et al. 2015, Wolf et al. 2018, Qi et al. 2018, Shen et al. 2018) these more character-rich studies have resolved and confirmed several previously uncertain nodes in the fern tree of life.

However, the exact phylogenetic relationships of some major fern lineages still remain controversial. For example, the phylogenetic position of horsetails (*Equisetum*) among the early fern lineages has remained elusive ever since they were first resolved as ferns (e.g. Pryer et al. 2001, Wikström and Pryer 2005, Karol et al. 2010, Knie et al. 2015, Qi et al. 2018, Shen et al. 2018). Filmy ferns (Hymenophyllales) represent another deep fern lineage with unclear relationships. They clearly belong to the early diverging leptosporangiate ferns, but some studies have resolved them as sister to Gleicheniales (e.g. Pryer et al. 2004, Lehtonen et al. 2017) while most studies seem to resolve them as sister to all other nonosmundalean leptosporangiate ferns (e.g. Pryer et al. 2001, Schuettpelz and Pryer 2007, Lehtonen 2011, Knie et al. 2015, Rai and Graham 2015, Rothfels et al. 2015, Wolf et al. 2018). Yet, some analyses have found Gleicheniales diverging before Hymenophyllales (Kuo et al. 2011, Knie et al. 2015), or Gleicheniales as a paraphyletic grade in respect to Hymenophyllales (Qi et al. 2018, Shen et al. 2018).

The order Hymenophyllales consists of a single family (Hymenophyllaceae) with nine genera and more than 400 species (PPG I 2016). They are either epiphytic or terrestrial and predominantly tropical (Kramer and Green 1990). A typical feature giving the common name "filmy fern" is the membranaceous lamina which lack stomata and has only one cell-layer between the veins (Kramer and Green 1990). These characteristics make the fossil record of filmy ferns very poor, with *Hymenophyllum iwatsukii* Herrera et al. from the Early Cretaceous and *Hopetedia praetermissa* Axsmith et al. from the Late Triassic being exceptionally well preserved examples (Axsmith et al. 2001, Herrera et al. 2017). Based on fossil-calibrated molecular dating the origin of filmy fern lineage has been dated at Permian or Carboniferous (Schuettpelz and Pryer 2006, 2009, Testo and Sundue 2016, Lehtonen et al. 2017).

Recently, Kuo et al. (2018) published plastomes of two Hymenophyllales genera, *Hymenophyllum* and *Callistopteris*. In this paper the first complete plastome of genus *Trichomanes* is presented with its structure and gene content described within a phylogenetic context.

Material and methods

Genome sequencing, assembly and annotation

Total genomic DNA was extracted from silica dried material of *Trichomanes trollii* Bergdolt (DNA voucher: Brazil: Amazonas, *Tuomisto et al. 15461* TUR) using E.Z.N.A. SP Plant DNA Kit (Omega Bio-tek, Doraville, GA) and paired-end 101 bp reads with insert size ~300 bp were sequenced by Illumina HiSeq 2500 Sequencing System at FIMM Technology Center, Finland. Raw sequence data were cleaned and trimmed using cutadapt (Martin 2011) and assembled de novo in Geneious 9.1.8 (Biomatters, New Zealand) with an average 61 $\times$ coverage.

The complete plastome was initially annotated in Dual Organeller GenoMe Annotator (DOGMA; Wyman et al. 2004) with annotations manually complemented and corrected in Geneious by comparison with homologous genes in published plastomes. Putative RNA editing was predicted with Prepact 2.0 (Lenz and Knoop 2013) and by comparing homologous genes in related species. The annotated plastome was visualized with GenomeVx (Conant and Wolfe 2008) and deposited at GenBank with accession number MH348274.

Phylogenetic analyses

In order to resolve the phylogenetic placement of *Trichomanes trollii* a set of available plastomes representing the main fern lineages were downloaded from GenBank. Plastomes of one angiosperm, two gymnosperms, and one lycopod were used as outgroups. A list of GenBank accession codes is given in Appendix 1.

All the protein coding sequences were extracted from the annotated plastomes using Geneious and aligned with default parameters in mafft v.7.125 (Katoh and Standley 2013), resulting in a dataset of 85 genes. In order to remove regions of uncertain alignment Gblocks 0.91 (Castresana 2000) was run allowing gap positions within the final blocks, otherwise the default parameters were used. The cleaned sequence alignments were concatenated using SequenceMatrix (Vaidya et al. 2011) resulting in a matrix of 67,801 characters and analysed under maximum-likelihood (ML) criterion. The ML analyses were performed with default parameters in Garli 2.0 (Zwickl 2006) by

running 100 independent runs. A general time reversible model of sequence evolution with gamma distributed rate variation among the sites (GTR+G) was applied across the concatenated data since this is the most general model of DNA evolution. 100 bootstrap replicates were performed in Garli with the same parameter settings but running only five independent ML runs per replicate. Results of the bootstrap replicates were summarised with SumTrees program (Sukumaran and Holder 2010a) in DendroPy package (Sukumaran and Holder 2010b). The final alignments as well as the resulting ML tree are available at TreeBASE (S22949).

Results

Genome organization

The complete plastome of *Trichomanes trollii* is 150,965 base pairs (bp) with a large single copy (LSC) region of 90,979 bp and a small single copy (SSC) region of 20,998 bp separated from each other by two inverted repeats (IRs), each of 19,494 bp (Figure 1). The genome contains 118 genes, when the IR is considered only once, and has an overall GC content of 38.1%. The SSC and IRs are bounded by *ndhF* and *chlL* genes and *rps12*, *rps7*, and *ndhB* genes are incorporated into the IRs. Comparison of predicted coding regions and RNA editing sites suggested that at least 38 sites in 23 genes required RNA editing for protein translation. Three of these sites were C-to-U edits and 35 U-to-C edits. The plastome organization is basically the same as in closely related *Callistopteris*, with inverted repeat including *ndhB*, *rps7*, *rps12*, and part of the *ycf2* (Kuo et al. 2018).

Phylogenetic analysis

The resulting ML tree agreed with the accepted consensus view of fern relationships (PPG I 2016) with the exception that *Equisetum* was placed as sister to Ophioglossidae and not to all other ferns (Figure 2). *Trichomanes trollii* was resolved as the sister of all the remaining nonosmundalean leptosporangiate ferns but with only 67% bootstrap support. The remaining 33% of the bootstrap replicates resolved *Trichomanes* as the sister of *Diplopterygium* (Gleicheniales). All the other nodes received ≥99% bootstrap support.

Discussion

The exact phylogenetic position of Hymenophyllales remains uncertain. Although resolved as the sister of the nonosmundalean leptosporangiate ferns, the support remained low despite of analysing a data matrix of 85 plastid genes. The poor support is likely related to the combination of short internal and long terminal branches in this part of the tree. It is noteworthy that this node remained equally poorly supported in an analysis of 25 nuclear loci with limited taxonomic sampling (Wolf et al. 2018). In taxonomically better sampled analyses of much larger number of nuclear loci Hymenophyllales and Gleicheniales formed a clade, but Gleicheniales was paraphyletic with respect to Hymenophyllales (Qi et al. 2018, Shen et al. 2018). In contrast, the plastome analysis by Kuo et al. (2018) supported, under most analyses, the monophyly of Gleicheniales, but the relationship between Hymenophyllales and Gleicheniales remained ambiguous. It seems clear that the monophyly of Gleicheniales and its relationship with Hymenophyllales needs to be investigated further also by taking into account morphological data and fossil evidence.

Other nodes in the tree were fully supported. This includes the early diverging Marattiales (represented by *Angiopteris*) and *Equisetum*, lineages whose positions have remained uncertain for a long time. Recently, the placement of Marattiales seems to be converging as sister group to leptosporangiate ferns (Gao et al. 2013, Zhong et al. 2014, Knie et al. 2015, Lu et al. 2015, Rothfels et al. 2015, Labiak and Karol 2017, Kuo et al. 2018, Qi et al. 2018) and this position is further supported here. Likewise, several recent studies (Karol et al. 2010, Grewe et al. 2013, Ruhfel et al. 2014, Zhong et al. 2014, Lu et al. 2015, Gitzendanner et al. 2018) place *Equisetum* as sister of Ophioglossidae in agreement with the present results and this position appears better supported than the alternative arrangement of *Equisetum* as sister to all other ferns (e.g. Kuo et al. 2011, Kim et al. 2014, Knie et al. 2015, Rothfels et al. 2015, Labiak and Karol 2017). However, the recent large-scale analyses using massive transcriptome data unequivocally support *Equisetum* as the sister of all other ferns (Qi et al. 2018, Shen et al. 2018).

The number of genes with internal stop codons, putatively requiring RNA editing, are high in the leptosporangiate ferns excluding *Osmundastrum* (Kim et al. 2014). The basal fern lineages in general have very few, if any, RNA edit sites (Kim et al. 2014). The number of genes interrupted by stop codons in *Trichomanes* are comparable with the numbers in other nonosmundalean leptosporangiates (Kim et al. 2014), suggesting that high levels of RNA editing is a common pattern in the plastomes of nonosmundalean leptosporangiate ferns. In contrast to previous reports in ferns, U-to-C edits associated with internal stop codons seem to be more common in *Trichomanes* than C-to-U edits associated with the start codons (Wolf et al. 2004, Guo et al. 2015, Labiak and Karol 2017). However, the number of editing sites estimated here is only a minimum estimate required to repair stop or start codons while the majority of confirmed editing sites change amino acids for another (Wolf et al. 2004). Complete plastid transcriptomes would be needed to confirm the putative RNA editing patterns.

Acknowledgements – Material used in this study was collected by H. Tuomisto, G. Zuquim, J. Prado, K. Ruokolainen and J. de Souza under the permit CNPq EXC n° 030/07-CR.

Funding – This work was financially supported by the Finnish Cultural Foundation.

References

Axsmith, B. J. et al. 2001. A filmy fern from the Upper Triassic of North Carolina (USA). – Am. J. Bot. 88: 1558–1567.

Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. – Mol. Biol. Evol. 17: 540–552.

Christenhusz, M. J. M. and Chase, M. W. 2014. Trends and concepts in fern classification. – Ann. Bot. 113: 571–594.

Christenhusz, M. J. M. et al. 2011. A linear sequence of extant families and genera of lycophytes and ferns. – Phytotaxa 19: 7–54.

Conant, G. C. and Wolfe, K. H. 2008. GenomeVx: simple web-based creation of editable circular chromosome maps. – Bioinformatics 24: 861–862.

Der, J.P. 2010. Genomic perspectives on evolution in bracken fern. – Ph.D. Dissertation. Utah State University, Utah, USA.

Gao, L. et al. 2009. Complete chloroplast genome sequence of a tree fern *Alsophila spinulosa*: insights into evolutionary changes in fern chloroplast genomes. – BMC Evol. Biol. 9: 130–14.

Gao, L. et al. 2013. 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.
Gitzendanner, M. A. et al. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. – Am. J. Bot. 105: 291–301.
Goremykin, V. V. 2003. Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. – Mol. Biol. Evol. 20: 1499–1505.
Grewe, F. et al. 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 Evol. Biol. 13: 8.
Guo, W. et al. 2015. Variable frequency of plastid RNA editing among ferns and repeated loss of uridine-to-cytidine editing from vascular plants. – PLoS ONE 10: e0117075.
Herrera, F. et al. 2017. An exquisitely preserved filmy fern (Hymenophyllaceae) from the Early Cretaceous of Mongolia. – Am. J. Bot. 104: 1370–1381.
Karol, K. G. et al. 2010. 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.
Katoh, K. and Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. – Mol. Biol. Evol. 30: 772–780.
Kim, H. T. et al. 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. – Mol. Cells 37: 372–382.
Knie, N. et al. 2015. Horsetails are the sister group to all other monilophytes and Marattiales are sister to leptosporangiate ferns. – Mol. Phylogenet. Evol. 90: 140–149.
Kramer, K.U. and Green, P.S. 1990. Pteridophytes and gymnosperms. – In: Kubitzki, K. (ed.), The families and genera of vascular plants. Berlin, Springer-Verlag, pp. 1–404.
Kuo, L.-Y. et al. 2011. First insights into fern matK phylogeny. – Mol. Phylogenet. Evol. 59: 556–566.
Kuo, L.-Y. et al. 2018. Order-level fern plastome phylogenomics: new insights from Hymenophyllales. Am. J. Bot. 105: 1545–1555.
Labiak, P. H. and Karol, K. G. 2017. Plastome sequences of an ancient fern lineage reveal remarkable changes in gene content and architecture. – Am. J. Bot. 104: 1008–1018.
Lehtonen, S. 2011. Towards resolving the complete fern tree of life. – PLoS ONE 6: e24851.
Lehtonen, S. et al. 2017. Environmentally driven extinction and opportunistic origination explain fern diversification patterns. – Sci. Rep. 7: 4831.
Lu, J.-M. et al. 2015. Chloroplast phylogenomics resolves key relationships in ferns. – J. Syst. Evol. 53: 448–457.
Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal, 17: 10–12.
PPG I. 2016. A community-derived classification for extant lycophytes and ferns. – J. Syst. Evol. 54: 1–41.
Pryer, K. M. et al. 2001. Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. – Nature 409: 618–622.
Pryer, K. M. et al. 2004. Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. – Am. J. Bot. 91: 1582–1598.
Qi, X. et al. 2018. A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. – Mol. Phylogenet. Evol. 127: 961–977.

Roper, J. M. et al. 2007. The complete plastid genome sequence of *Angiopteris evecta* (G. Forst.) Hoffm. (Marattiaceae). – Am. Fern J. 97: 95–106.
Rothfels, C. J. et al. 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. – Am. J. Bot. 102: 1089–1107.
Ruhfel, B. R. et al. 2014. From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. – BMC Evol. Biol. 14: 23.
Ruixiang, X, et al. 2018. The first complete chloroplast genome of a traditional Chinese medicinal herb *Odontosoria chinensis* (Lindsaeaceae). – Mitochondrial DNA B Resour. 3: 292–293.
Schuettpelz, E. and Pryer, K.M. 2006. Reconciling extreme branch length differences: decoupling time and rate through the evolutionary history of filmy ferns. – Syst. Biol. 55: 485–502.
Schuettpelz, E. and Pryer, K. M. 2007. Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. – Taxon 56: 1037–1050.
Schuettpelz, E. and Pryer, K.M. 2009. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. – Proc. Natl. Acad. Sci. USA 106: 11200–11205.
Shen, H. et al. 2018. Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. – GigaScience, 7: 1–11.
Smith, A. R. et al. 2006. A classification for extant ferns. – Taxon 55: 705–731.
Sukumaran, J. and Holder, M. T. 2010a. SumTrees: phylogenetic tree summarizatio. 4.4.0. Available at https://github.com/jeetsukumaran/DendroPy.
Sukumaran, J. and Holder, M. T. 2010b. DendroPy: a Python library for phylogenetic computing. – Bioinformatics 26: 1569–1571.
Testo, W. and Sundue, M. 2016. A 4000-species dataset provides new insight into the evolution of ferns. – Mol. Phylogenet. Evol. 105: 200–211.
Vaidya, G. et al. 2011. SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. – Cladistics 27: 171–180.
Wei, R. et al. 2017. Plastid phylogenomics resolve deep relationships among eupolypod II ferns with rapid radiation and rate heterogeneity. – Genome Biol. Evol. 9: 1646–1657.
Wikström, N., and Pryer, K. M. 2005. Incongruence between primary sequence data and the distribution of a mitochondrial atp1 group II intron among ferns and horsetails. – Mol. Phylogenet. Evol. 36: 484–493.
Wolf, P. G. et al. 2018. Target sequence capture of nuclear-encoded genes for phylogenetic analysis in ferns. – Appl. Plant Sci. 120: e01148–8.
Wolf, P. G. et al. 2004. High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. – Gene 339: 89–97.
Wolf, P. G. et al. 2003. Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. – DNA Res. 10: 59–65.
Wu, C.-S. et al. 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. – Mol. Biol. Evol. 24: 1366–1379.
Wyman, S. K. et al. 2004. Automatic annotation of organellar genomes with DOGMA. – Bioinformatics 20: 3252–3255.
Zhong, B. et al. 2014. Two new fern chloroplasts and decelerated evolution linked to the long generation time in tree ferns. – Genome Biol. Evol. 6: 1166–1173.
Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation. The University of Texas at Austin, Texas, USA.

APPENDIX 1. List of species analysed with the respective GenBank accession numbers and references.

Actinostachys pennula (Sw.) Hook. KU764518 (Labiak and Karol 2017), *Adiantum capillus-veneris* L. NC_004766 (Wolf et al. 2003), *Alsophila spinulosa* (Wall. ex Hook.) R.M.Tryon NC_012818 (Gao et al. 2009), *Amborella trichopoda* Baill. NC_005086 (Goremykin et al. 2003), *Angiopteris evecta* (G.Forst) Hoffm. DQ821119 (Roper et al. 2007), *Ceratopteris richardii* Brongn. KM052729 (Marchant et al. unpublished), *Cycas taitungensis* C.F.Shen et al. NC_009618 (Wu et al. 2007), *Cyrtomium devexiscapulae* (Koidz.) Ching KT599100 (Lu et al. 2015), *Cystopteris chinensis* (Ching) X.C.Zhang & R.Wei KY427337 (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), *Ginkgo biloba* L. NC_016986 (Li et al. unpublished), *Isoetes flaccida* Shuttlew. NC_014675 (Karol et al. 2010), *Lepisorus clathratus* Ching NC_035739 (Wei et al. 2017), *Lygodium japonicum* (Thunb.) Sw. KC536645 (Gao et al. 2013), *Mankyua chejuensis* B.Y.Sun et al. KP205433 (Kim and Kim unpublished), *Marsilea crenata* C.Presl NC_022137 (Gao et al. 2013), *Odontosoria chinensis* (L.) J.Sm. MG913608 (Ruixiang et al. 2018), *Ophioglossum californicum* Prantl NC_020147 (Grewe et al. 2013), *Osmundastrum cinnamomeum* (L.) C.Presl NC_024157 (Kim et al. 2014), *Plagiogyria glauca* (Blume) Mett. KP136831 (Marchant et al. unpublished), *Psilotum nudum* L. NC_003386 (Wakasugi et al. unpublished), *Pteridium aquilinum* (L.) Kuhn NC_014348 (Der 2010), *Schizaea elegans* (Vahl) Sw. KX258660 (Labiak and Karol 2017), *Pseudophegopteris aurita* (Hook.) Ching NC_035861 (Wei et al. 2017), *Tmesipteris elongata* P.A.Dang KJ569699 (Zhong et al. 2014), *Trichomanes trollii* Bergdolt MH348274 (this study).

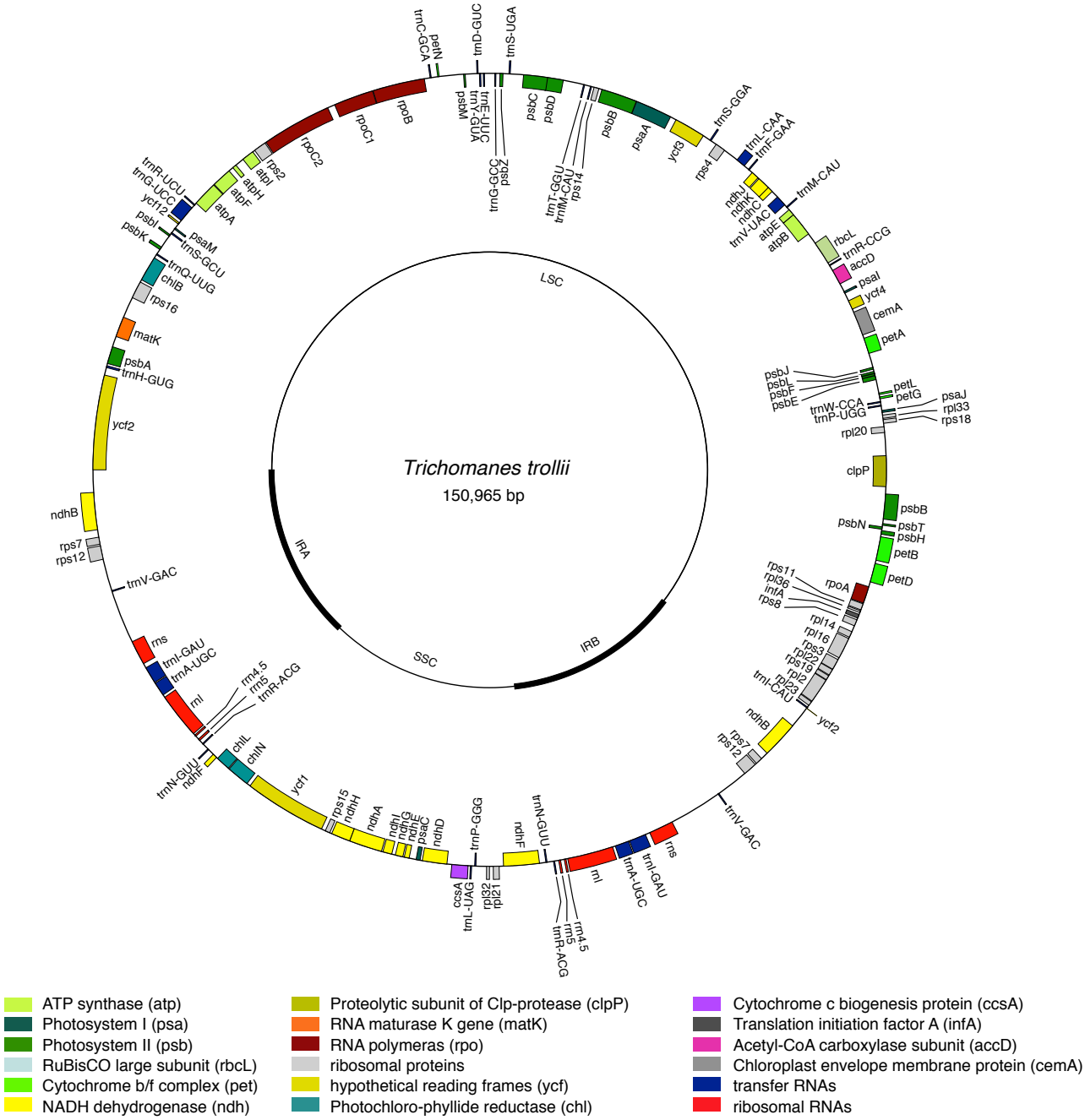

Figure 1. Plastome map of the *Trichomanes trollii*. Genes on the inside and outside of the outer circle are transcribed counterclockwise and clockwise, respectively. Thick black lines on the inner circle represent the inverted repeats (IRA and IRB) which separate the genome into the large (LSC) and small (SSC) single copy regions. Genes are coded by functional groups as shown in the key.

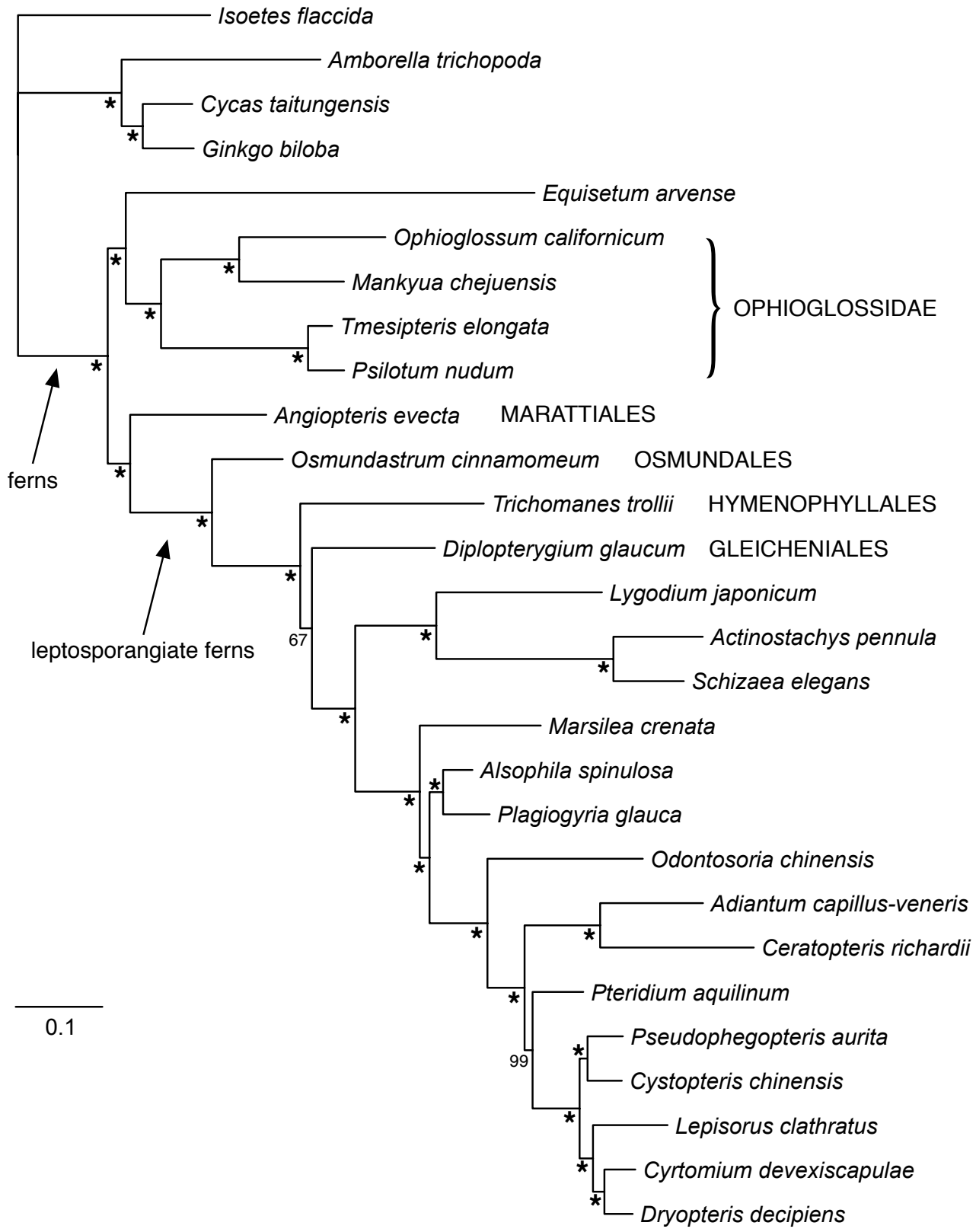

Figure 2. Plastome phylogeny based on simultaneous ML analysis of 85 coding regions. Bootstrap support is shown at the nodes; asterisk denotes 100% support.