Exploring the phylogeny of the marattialean ferns

Samuli Lehtonen, Péter Pocza, Gaurav Sablok, Jaakko Hyvönen, Dirk N. Karger, and Jorge Flores

Biodiversity Unit, University of Turku, FI-20014, Turku, Finland; Finnish Museum of Natural History (Botany), University of Helsinki, PO Box 7, FI-00014, Helsinki, Finland; OEB and ViPS, University of Helsinki, PO Box 65, FI-00014, Helsinki, Finland; Swiss Federal Research Institute, WSL, 8903, Birmensdorf, Switzerland

Accepted 16 April 2020

Abstract

The eusporangiate marattialean ferns represent an ancient radiation with a rich fossil record but limited modern diversity in the tropics. The long evolutionary history without close extant relatives has confounded studies of the phylogenetic origin, rooting and timing of marattialean ferns. Here we present new complete plastid genomes of six marattialean species and compiled a plastid genome dataset representing all of the currently accepted marattialean genera. We further supplemented this dataset by compiling a large dataset of mitochondrial genes and a phenotypic data matrix covering both extant and extinct representatives of the lineage. Our phylogenomic and total-evidence analyses corroborated the postulated position of marattialean ferns as the sister to leptosporangiate ferns, and the position of Danaea as the sister to the remaining extant marattialean genera. However, our results provide new evidence that Christensenia is sister to Marattia and that M. cicutifolia actually belongs to Eupodium. The apparently highly reduced rate of molecular evolution in marattialean ferns provides a challenge for dating the key phylogenetic events with molecular clock approaches. We instead applied a parsimony-based total-evidence dating approach, which suggested a Triassic age for the extant crown group. The modern distribution can best be explained as mainly resulting from vicariance following the breakup of Pangaea and Gondwana. We resolved the fossil genera Marattiopsis, Danaeopsis and Qasi-mia as members of the monophyletic family Marattiaceae, and the Carboniferous genera Sydneia and Radstockia as the monophyletic sister of all other marattialean ferns.

© 2020 The Authors. Cladistics published by John Wiley & Sons Ltd on behalf of Willi Hennig Society.

Introduction

The eusporangiate fern order Marattiales represents a Palaeozoic radiation, currently comprised of ≈100 species in six genera (Murdock, 2008a; PPG I, 2016). Although their current distribution is restricted to the wet tropical regions (Murdock, 2008a; PPG I, 2016), a diverse fossil record has been found from all continents, including Antarctica (Stidd, 1974; Escapa et al., 2014). Due to its ancient origin and abundant fossil record, the phylogeny of the Marattiales has received considerable interest from both palaeontological and neontological perspectives (Mamay, 1950; Hill and Camus, 1986a; Hill and Camus, 1986b; Liu et al., 2000a; Christenhusz, 2007; Li and Lu, 2007; Christenhusz et al., 2008; Murdock, 2008b; Senterre et al., 2014; Cleal, 2015; Rothwell et al., 2018). Although the phylogenetic relationships of the early diverging fern lineages have remained ambiguous for a long time (Pryer et al., 2001; Wikström and Pryer, 2005; Rothwell and Nixon, 2006; Karol et al., 2010; Knie et al., 2015), the most recent studies based on phylogenomic data of plastomes (Grewe et al., 2013; Kim et al., 2014; Lu et al., 2015; Labiak and Karol, 2017; Gitzendanner et al., 2018; Kuo et al., 2018; Lehtonen, 2018; Lehtonen and Cárdenas, 2019) and transcriptome data (Rothfels et al., 2015; Qi et al., 2018; Shen et al., 2018) generally support a sister relationship between the marattialean and leptosporangiate ferns (but see Wickett et al., 2014, and One Thousand Plant Transcriptomes Initiative,
2019, for alternative resolutions). Based on the fossil record and molecular dating this split had occurred by the early Carboniferous with a rapid diversification until marattialean ferns reached their maximal diversity during the Palaeozoic (DiMichele and Phillips, 2002; Lehtonen et al., 2017). This was followed by extinctions at the Permo-Triassic boundary and the rise of more modern-appearing forms in the Mesozoic (DiMichele and Phillips, 2002; Lehtonen et al., 2017) with apparently continually diminishing ecological importance, as suggested by the almost complete lack of Cenozoic fossils (Collinson, 2001).

Marattiales generally is divided into two families: the now extinct Psaroniaceae (often called Asterothecaeeae, see Cleal, 2015, for discussion about the correct name) and the extant Marattiaceae (Morgan, 1959). Psaroniaceae is a predominantly Palaeozoic family of mainly very large, arborescent ferns, which formed an important and sometimes dominant component of the coal swamp vegetation (Phillips et al., 1985; Millay, 1997; DiMichele and Phillips, 2002). Typical characteristics of Psaroniaceae include a massive, trunk-like stem with thick root mantle, highly divided large leaves with pecopteroid pinnule shape, and sporangia aggregated into radial synangia with limited fusion between the sporangia (Stidd, 1974; Millay, 1997). By contrast, Marattiaceae stems vary from slender and creeping, to large, fleshy and globular stems clothed with large stipules. Leaves are less divided with taeniopterid pinnules, and sporangia are generally fused into a bilaterally symmetrical synangia (Stidd, 1974; Hill and Camus, 1986a; Camus, 1990). Recently, Sydneideae, a new subfamily of Psaroniaceae was described for *Sydneia manleyi* Pšenička et al. with atypical sphenopteroid foliage and radially symmetrical synangia (Pšenička et al., 2014). The taxonomy of the fossil Marattiales has been studied extensively but is confounded by imperfect and variable preservation, making it difficult to reconstruct whole plants based on separately preserved stems, foliage fragments and reproductive structures (Lesnikowska, 1989), or to match permineralized material with adpresion fossils (Cleal, 2015).

The taxonomic concepts of the extant Marattiaceae have varied over time, and the species-level taxonomy still remains disputed especially in the larger genera *Danaea* Sm. and *Angiopteris* Hoffm. (Christenhusz, 2007; Murdock, 2008a). The genus-level classification, however, was revised based on molecular studies by Murdock (2008a, b) and has since been accepted by authors (Christenhusz and Chase, 2014; Senterre et al., 2014; Arana, 2016; PPG I, 2016; Tuomisto et al., 2018). When compared with the older concepts, the most dramatic change was the splitting of the paraphyletic genus *Marattia* Sw. into three segregate genera: *Marattia*, *Eupodium* J.Sm. and *Ptisana* Murdock. This classification is problematic from a palaeobotanical perspective, as numerous fossil species have been traditionally placed in the broadly defined *Marattia*, or *Marattioptis* Schimp., a name that was established for those fossils that have generally similar morphological characteristics compared to extant species (Schimper, 1869). The problem is not only that the diagnostic morphological characters of the newly split genera are difficult to discern in fossil material, but also that many fossils seem to show a mixed set of these characters (Bomfleur et al., 2013; Escapa et al., 2014; Kvaček, 2014).

Given the ancient origin of the lineage, its lack of close extant relatives with the somewhat diffuse nature of the extant genera, and the presence of fossil material mixing the putative synapomorphies of the extant genera, it is not surprising that rooting the extant crown group Marattiaceae has remained problematic (Murdock, 2008b; Rothwell et al., 2018). Accordingly, this results in an uncertain interpretation of the character evolution and the biogeographical history of Marattiales. Although the marattialean phylogeny has been investigated quite widely, previous studies have either ignored the fossil evidence (Li and Lu, 2007; Christenhusz et al., 2008; Murdock, 2008b), or used very few if any molecular data from the extant species (Hill and Camus, 1986a; Liu et al., 2000a; Rothwell et al., 2018). The latest analyses (Li and Lu, 2007; Christenhusz et al., 2008; Murdock, 2008b; Senterre et al., 2014; Rothwell et al., 2018) have incorporated a good sampling of the extant species, but unfortunately included only a few molecular markers; hence, the potential of genomic characters has thus far not been exploited.

Over the past few years the number of completely sequenced fern plastomes has increased rapidly. Together with the transcriptome-based nuclear data, plastomes have been increasingly used to resolve challenging nodes of the fern phylogeny (Grewe et al., 2013; Kim et al., 2014; Lu et al., 2015; Labiak and Karol, 2017; Kuo et al., 2018; Lehtonen, 2018; Lehtonen and Cárdenas, 2019). At the same time, understanding of the structural evolution of the fern plastome has advanced greatly, leading to the emergence of a better view on how the ancestral genome structure has dynamically evolved through inversions, inverted repeat border expansions and contractions, and apparent insertions and deletions of mobile open reading frames (ORFs) into the plastome (Wolf et al., 2010; Gao et al., 2011; Grewe et al., 2013; Li et al., 2016; Robison et al., 2018; Lehtonen and Cárdenas, 2019). Marattialean ferns are of special interest in this context due to their phylogenetic position and apparent lack of RNA editing, a feature very common among the more derived ferns (Roper et al., 2007; Kim et al., 2014; Li et al., 2018). Thus, wider sampling
of marattialean plastomes, along with resolving their phylogenetic position and timing of origin, are of great interest for aiding the understanding of the evolution of fern plastid genome organization.

In this study, we generated a dataset of complete plastomes that represent all extant genera of Marattiaaceae and compiled a phenotypic data matrix that represents all fairly well-known representatives of the marattialean ferns. Our data also allowed us to compile a large set of mitochondrial sequence data to be analyzed together with plastid and phenotypic characters. We then used these data to infer phylogenetic relationships and timing of diversification to investigate biogeographical hypotheses and explore the plastome genome evolution of this ancient fern order. We agree with Fitzhugh (2006) that there are no valid reasons to ignore any data that are potentially informative about phylogeny. As discussed by Wheeler et al. (2006), the commonly assumed distinction between genotypic and phenotypic data is artificial. We do acknowledge that there are real problems in combining all of the information in the same analyses, but this is not an excuse to ignore them, as the benefits obtained far outweigh the potential problems they cause. Studies that do not include known fossils implicitly treat them as separate from the phylogeny of extant organisms, which is an unrealistic stance.

**Materials and methods**

**Taxon sampling**

We aimed to investigate the phylogeny of marattialean ferns at two different levels. First, we explored their rooting and position within the overall fern phylogeny by analyzing a set of complete plastomes. Second, we further explored the marattialean rooting, phylogenetic resolution, and its timing by coding a taxonomically broad matrix of phenotypic characters from the representatives of extant and extinct marattialeans.

The extant marattialeans are currently classified into six genera (Murdock, 2008a; PPG I, 2016). We sampled the extant taxa at this level. Two complete marattialean plastomes were available in GenBank (Roper et al., 2007; Zhu et al., 2015), both of them representing the genus *Angiopteris*. Thus, we extracted DNA from samples from the remaining four genera. The highest-quality extractions from each genus were selected for Illumina sequencing (see Molecular data). The extant marattialeans are currently classified into six genera (Murdock, 2008a; PPG I, 2016). We sampled the extant taxa at this level. Two complete marattialean plastomes were available in GenBank (Roper et al., 2007; Zhu et al., 2015), both of them representing the genus *Angiopteris*. Thus, we extracted DNA from samples from the remaining four genera. The highest-quality extractions from each genus were selected for Illumina sequencing (see Molecular data).

**Plastome coverage was evaluated in SAMTOOLS (Li et al., 2009) and BEDTOOLS (Quinlan and Hall, 2010) under the *genomewide* function by remapping the sequenced reads to the respective plastomes. To assess the synteny of the newly sequenced plastomes with the published ones we aligned the genomes with GETORGANELLE (Jin et al., 2018), NOVOPLASTY v.2.6.3 (Dierckxsens et al., 2016), VELVET v.1.2.08 (Zerbino and Birney, 2008), and GENEIOUS v.9.1.8 (www.geneious.com). The contigs thus obtained were then de novo assembled in GENEIOUS and the reads were mapped to the assemblies for verification and manual correction. Some gaps remained in the plastomes of *Ptisana novoguineensis* (Rosenst.) Murdock, *Eupodium kauflussii* (J.Sm.) J.Sm. and *Marattia laxa* Kunze; and these were filled with Sanger sequencing using custom-designed primers (Table S1). The complete plastomes were annotated in GENIOUS by comparing the initial annotations obtained in Dual Organeller GenoME Annotator (DOGMA; Wyman et al., 2004) with annotations on published sequences and ORFs.

Plastome coverage was evaluated in SAMTOOLS (Li et al., 2009) and BEDTOOLS (Quinlan and Hall, 2010) under the *genomewide* function by remapping the sequenced reads to the respective plastomes. To assess the synteny of the newly sequenced plastomes with the published ones we aligned the genomes with LASTZ (Harris, 2007) and MUTTER v.3.1 (Kurtz et al., 2004). Plastome coverage, gene tracks and synteny alignments were visualized with CIRCOS (Krzywinski et al., 2009) and GGBIO (Yin et al., 2012). A custom database was set up from the annotated plastid genomes in GENIOUS, which was used to search previously described Mobile Open Reading Frames in *Fern Organelles* (MORFFO; Robison et al., 2018) using the default settings in tBLASTX (Altschul et al., 1990).

We performed a comparative analysis of simple sequence repeats using MISA by evaluating both simple as well as compound repeats (Thiel et al., 2003; Beier et al., 2017). MISA was used to analyze the perfect microsatellites often abbreviated as simple sequence repeats (SSRs) with a defined length of *n* = 10 in the case of mono-, *n* = 6 in the case of di-, and *n* = 3 in the case of tri-, tetra-, penta- and hexa-nucleotide repeats. For the compound repeats, two defined SSRs should be interrupted by 100 bp. Microstructural events such as inversions and single nucleotide polymorphism (SNP) (including deletions and substitutions) were identified using the pairwise alignments in LASTZ and MUMMER. Following the alignments, the show-snps feature in MUMMER along with the mummer plot was used for the identification of the plastome-wide and gene-wise plastomic variations.

We also produced a mitochondrial DNA (mtDNA) matrix by first mapping the reads of *Danaea sellowiana* C.Presl, the sample that had
the highest coverage of cpDNA, to the complete mitochondrion of *Ophioglossum californicum* Prantl (Guo et al., 2016) in GENEIOUS.

Analyses of the phylogenetic position of Marattiales

For comparative phylogenetic analysis, a matrix of both the fern plastomes and seed plant outgroups (Table S2) was constructed by extracting the genes accD, atpA, atpB, atpE, atpH, atpI, ccsA, ccmA, chlB, chlF, chlN, infA, marK, ndhC, ndhD, ndhE, ndhF, ndtG, ndtH, ndtJ, ndtK, petA, petG, petL, petN, psaA, psaB, psbA, psbC, psbI, psa1, psbB, psbD, psbE, psbF, psbH, psbI, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL, rpl14, rpl20, rpl21, rpl22, rpl23, rpL36, rpO4A, rpO4B, rpO4C2, rps2, rps3, rps4, rps5, rps11, rps14, rps15, rps18, rps19, ycf4 and ycf12 (67 genes in total), and performing alignments of coding regions with MACSE (Ranwez et al., 2018). The alignments subsequently were masked for the internal stop codons following the frameshift alignment algorithm correction as implemented in MACSE, and trimmed using trimAl (Capella-Gutierrez et al., 2009). Finally, before the construction of the super-matrix, the presence of terminal stop codons was checked; if identified they were subsequently removed from the trimmed alignments. The final matrix was constructed using **SequenceMatrix v1.1.8** (Vaidya et al., 2011). The phylogeny was inferred using two different maximum-likelihood (ML) methods. First, we used **IQTREE** (Nguyen et al., 2013) with the approximate likelihood-ratio test (aLRT), Shi–Hasegawa likelihood (sh-LRT), Akaike information criterion (AIC) and Bayesian information criterion (BIC) using **MODELFINDER** (Kalyaanamoorthy et al., 2017). **UFBoot** (Hoang et al., 2018) was used for ultrafast bootstrap (BS) calculation. The key idea behind UFBoot is to keep trees encountered during the ML-tree search for the original sequence alignment and to use them to evaluate the tree likelihoods for the BS sequence alignment. UFBoot provides relatively unbiased BS estimates under mild model

**Table 1**
The plastome data used in this study

<table>
<thead>
<tr>
<th>Taxon</th>
<th>Accession code</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Adiantum capillus-veneris</em> L.</td>
<td>NC_004766</td>
<td>Wolf et al. (2003)</td>
</tr>
<tr>
<td><em>Alostaphila spinulosa</em> (Hook.) R.M.Tryon</td>
<td>NC_012818</td>
<td>Gao et al. (2009)</td>
</tr>
<tr>
<td><em>Angiopteris angustifolia</em> C.Presl</td>
<td>NC_026300</td>
<td>Zhu et al. (2015)</td>
</tr>
<tr>
<td><em>Angiopteris evecta</em> (Forst.) Hoffm.</td>
<td>NC_008829</td>
<td>Roper et al. (2007)</td>
</tr>
<tr>
<td><em>Christensenia aesculifolia</em> (Blume) Maxon</td>
<td>MN412587</td>
<td>This study</td>
</tr>
<tr>
<td><em>Cycas panzhihuaensis</em> L.Zhou &amp; S.Y.Yang</td>
<td>MN_031413</td>
<td>Han et al. (2017)</td>
</tr>
<tr>
<td><em>Cycas revoluta</em> Bedd.</td>
<td>NC_020319</td>
<td>Li et al., unpublished</td>
</tr>
<tr>
<td><em>Cycas taitungensis</em> C.F.Shen et al.</td>
<td>NC_009618</td>
<td>Wu et al. (2007)</td>
</tr>
<tr>
<td><em>Cytromium devexiculacae</em> (Koidz.) Ching</td>
<td>NC_028542</td>
<td>Lu et al. (2015)</td>
</tr>
<tr>
<td><em>Cytromium falcatum</em> (L.I.) C.Presl</td>
<td>NC_028705</td>
<td>Choi and Park, unpublished</td>
</tr>
<tr>
<td><em>Danaea sellowiana</em> Presl</td>
<td>MN412588</td>
<td>This study</td>
</tr>
<tr>
<td><em>Diploterigium glaucum</em> (Thunb. ex. Houtt.) Nakai</td>
<td>NC_024158</td>
<td>Kim et al. (2014)</td>
</tr>
<tr>
<td><em>Eupodium cicatfolium</em> (Kaufl.) Lehtonen</td>
<td>NC_021590</td>
<td>This study</td>
</tr>
<tr>
<td><em>Eupodium kaufussi</em> (J.Sm.) J.Sm.</td>
<td>NC_021589</td>
<td>This study</td>
</tr>
<tr>
<td><em>Equisetum arvense</em> L.</td>
<td>NC_014699</td>
<td>Karol et al. (2010)</td>
</tr>
<tr>
<td><em>Equisetum hyemale</em> L.</td>
<td>NC_021589</td>
<td>Grewe et al. (2013)</td>
</tr>
<tr>
<td><em>Eupodium cicutifolium</em> (Kaulf.) Lehtonen</td>
<td>NC_021590</td>
<td>This study</td>
</tr>
<tr>
<td><em>Ginkgo biloba</em> L.</td>
<td>NC_016986</td>
<td>Li et al., unpublished</td>
</tr>
<tr>
<td><em>Lygodium japonicum</em> (Thunb.) Sw.</td>
<td>NC_022136</td>
<td>Gao et al. (2013)</td>
</tr>
<tr>
<td><em>Marattia laxa</em> Kunze</td>
<td>MN412591</td>
<td>This study</td>
</tr>
<tr>
<td><em>Marsilea crenata</em> C.Presl</td>
<td>NC_022137</td>
<td>Gao et al. (2013)</td>
</tr>
<tr>
<td><em>Myriopteris lindeheimeri</em> J.Sm.</td>
<td>NC_014592</td>
<td>Wolf et al. (2011)</td>
</tr>
<tr>
<td><em>Ophioglossum californicum</em> Prantl</td>
<td>NC_020147</td>
<td>Grewe et al. (2013)</td>
</tr>
<tr>
<td><em>Osmunda cinnamomea</em> (L.) C.Presl</td>
<td>NC_021579</td>
<td>Kim et al. (2014)</td>
</tr>
<tr>
<td><em>Pilottum nudum</em> (L.) P.Beauv.</td>
<td>NC_003386</td>
<td>Wakasugi et al., unpublished</td>
</tr>
<tr>
<td><em>Pteridium aquilinum</em> (L.) Kuhn</td>
<td>NC_014348</td>
<td>Der (2010)</td>
</tr>
<tr>
<td><em>Pteris aquilina</em> (L.) C.Chr.</td>
<td>MN412592</td>
<td>This study</td>
</tr>
<tr>
<td><em>Pitunia novoguineensis</em> (Rosenst.) Murdock</td>
<td>NC_028543</td>
<td>Lu et al. (2015)</td>
</tr>
</tbody>
</table>

Seventy-seven phenotypic characters were obtained largely from the literature and scored for the studied taxa (Appendix 1). Scoring was based mainly on literature review but was confirmed for most of the extant taxa by studying the collections deposited in the herbarium of the University of Turku (TUR; see Appendix 1). A total of 18 characters were parsimony-uninformative in the current data matrix but are still listed here as they would be informative for broader taxonomic sampling of Marattiaceae and Osmundaceae. The list of characters and their states can be found in the Appendix 2, and the scored data matrix in the Appendix 3. All of the data matrices and resulting trees are available at TreeBASE (S25298), and the phenotypic data at MorphoBank (P3655).
misseffects and reduces computing time while achieving more unbiased branch supports than standard BS (Hoang et al., 2018). Second, we ran RAxML v.8.2.9 under the GTR-GAMMA and CAT models (Stamatakis, 2014).

We time-calibrated the plastome phylogeny with BEAST 1.10.1 (Suchard et al., 2018). We set a Yule tree prior to keep the model as simple as possible, even if the two-parameter birth-death model might have fitted the data better (Gernhard, 2008). Based on simulation studies it appears that the choice of tree prior has relatively little impact as long as the sequence data are informative (Serav et al., 2019) and we use a high number of calibration priors. We set calibration priors for 11 nodes, including the root node, using exponential prior distributions with a hard minimum and a soft maximum (with 5% prior probability distribution exceeding the constraint) ages for the nodes other than the root node, for which we applied a uniform prior (Table 2). The exponential prior favors ages close to the hard minimum age, therefore unrealistically assuming that sampled fossils represent the earliest occurrences of their lineages. We alleviated this problem by using relatively old soft maximum age constraints to flatten the prior distribution. As minimum ages, we used the minimum age boundaries of the geological formations from which the relevant fossils were found by applying the timescale of Walker et al. (2018). For most of the nodes we followed the soft maximum age constraints justified by Lehtonen et al. (2017). For seed plants, we applied the same soft maximum age as for the ferns and euphyllophytes (root node); this was set as middle Silurian, following Hao and Xue (2013). The molecular rates vary greatly between the fern lineages (Korall et al., 2010; Rothfels and Schuettpelz, 2013) and it has been shown that in such cases the random local clock (RLC) model (Drummond and Suchard, 2010) outperforms other clock models (Crisp et al., 2014). We therefore applied the RLC model and ran 14 chains of $30 \times 10^7$ generations sampling every 5000 generations, using the computing facilities of the CSC - IT Center for Science Ltd (csc.fi).

For the time-calibration, the molecular data were divided by genes and a greedy search was performed in PartitionFinder v.2.1.1 (Guindon et al., 2010; Lanfear et al., 2012, 2017) to determine the best partition strategy with associated substitution models under the Bayesian Information Criteria. This resulted in four character sets, three of which were assigned GTR + I + G and one with TVM + G model of evolution. Convergence of the runs and effective sample sizes were checked in Tracer v.1.7.1 (Rambaut et al., 2018) before combining them in LogCombiner v.1.10.1 (Rambaut and Drummond, 2002–2018a) with a 25% burn-in. A maximum clade credibility tree was reconstructed using TreeAnnotator v.1.10.1 (Rambaut and Drummond, 2002–2018b) and visualized using FigTree v.1.4.4 (Rambaut, 2006–2018). To investigate the possible deviation of specified and effective calibration priors due to prior interactions (Heled and Drummond, 2011), we re-ran the analysis by sampling from the prior only.

### Analyses of the phylogenetic relationships within Marattiales

The more detailed analyses on the marattialean relationships were performed for various data combinations to examine how the signal varies between pheno- and genotypic data, or if the inclusion of fossil terminals has a significant role in determining the topology. Hence, we separately analyzed cpDNA and mtDNA datasets and a phenotypic dataset including and excluding the fossils. We finally combined the datasets of all pheno- and genotypic characters including and excluding the fossils.

Tree searches under parsimony as optimality criterion were performed with TNT v.1.5 (Goloboff and Catalano, 2016) using 10 initial replicates per hit. For each replicate, 20 iterations of Tree Fusing subsequently were conducted to search for optimal trees (Goloboff, 1999). Searches were terminated after the best score was hit seven times. Support values were estimated with Symmetric Resampling, employing the difference between the most frequent groups and their most frequent contradictory group (GC; Goloboff, 2003).

The phylogenetic hypotheses were assessed for their sensitivity to the variation in the analysis parameters. In addition to equal weighting, the data were analyzed under extended implied weighting by using four different concavity values as reference ($k = 5$, $k = 10$, $k = 15$ and $k = 20$; Goloboff, 2014). In this set of analyses, individual characters were weighted by considering the proportion of missing entries, where each missing entry was assumed to have half of the homoplas of the observed entries ($P = 0.5$). Although the entire range of weighting values were employed to evaluate sensitivity, $k =$ 15 was arbitrarily selected as an intermediate concavity value (regarding $k = 5$ and equal weighting as the strongest and weakest weighting schemes, respectively) to further estimate nodal support, divergence times, and to be used as a reference for investigating the sensitivity.

In addition to the parsimony analyses, we analyzed the same datasets using MrBayes v.3.2.6 (Ronquist et al., 2012b) in CIPRES (Miller et al., 2010). The data partitioning and model selection were performed with PartitionFinder (Lanfear et al., 2017) as indicated above. For the phenotypic dataset, we applied the Markov $k$ model

---

**Table 2**

Node calibrations applied in the BEAST analysis

<table>
<thead>
<tr>
<th>Clade</th>
<th>Stem/ crown</th>
<th>Fossil</th>
<th>Hard min age</th>
<th>Soft max age</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Euphyllyphya</td>
<td>Stem</td>
<td>Euphyllyphyon</td>
<td>407.6 Ma*</td>
<td>427.4 Ma*</td>
<td>Hao and Xue (2013)</td>
</tr>
<tr>
<td>Seed plants</td>
<td>Crown</td>
<td>Cordaitales</td>
<td>315.2 Ma</td>
<td>407.6 Ma</td>
<td>Falcon-Lang (2005)</td>
</tr>
<tr>
<td>Ferns</td>
<td>Stem</td>
<td>Ibyka</td>
<td>382.7 Ma</td>
<td>407.6 Ma</td>
<td>Skog and Banks (1973)</td>
</tr>
<tr>
<td>Equisetales</td>
<td>Stem</td>
<td>Archaeocalamites</td>
<td>358.9 Ma</td>
<td>407.6 Ma</td>
<td>Stewart and Rothwell (1993)</td>
</tr>
<tr>
<td>Marattiales</td>
<td>Stem</td>
<td>Psaronius</td>
<td>323.2 Ma</td>
<td>407.6 Ma</td>
<td>Gerrienne et al. (1999)</td>
</tr>
<tr>
<td>Osmundales</td>
<td>Stem</td>
<td>Grammatopteris</td>
<td>272.95 Ma</td>
<td>358.9 Ma</td>
<td>Röllner and Galtier (2002)</td>
</tr>
<tr>
<td>Gleicheniales</td>
<td>Stem</td>
<td>Chansitheca</td>
<td>272.95 Ma</td>
<td>358.9 Ma</td>
<td>He et al. (2016)</td>
</tr>
<tr>
<td>Schizaeales</td>
<td>Stem</td>
<td>Stachypteris</td>
<td>168.3 Ma</td>
<td>298.9 Ma</td>
<td>Wikström et al. (2002)</td>
</tr>
<tr>
<td>Marsileaceae</td>
<td>Stem</td>
<td>Marsileaceaphylum</td>
<td>139.8 Ma</td>
<td>201.3 Ma</td>
<td>Hu et al. (2008)</td>
</tr>
<tr>
<td>Pteridaceae</td>
<td>Stem</td>
<td>Pteris</td>
<td>93.9 Ma</td>
<td>201.3 Ma</td>
<td>Krassilov and Bacchia (2000)</td>
</tr>
<tr>
<td>Woodwardia</td>
<td>Stem</td>
<td>Woodwardia</td>
<td>56.0 Ma</td>
<td>145.0 Ma</td>
<td>Wang et al. (2006)</td>
</tr>
</tbody>
</table>

*Uniform prior.
†Monophyly enforced.
with only variable characters (Mkv) (Lewis, 2001). Both the cpDNA and mtDNA datasets were divided in four partitions when analyzed separately; in the combined analysis these data were divided in six partitions. Details about the data partitioning and model settings can be found in Appendix S1. For each data combination analyzed, two independent MrBayes runs, each comprising four chains, were conducted with sampling every 2000 generations. In total 2 x 10^7 generations were run for each analysis and the convergence was assessed in Tracer v.1.7.1 (Rambaut et al., 2018). All of the effective sample sizes were >500 after discarding the first 25% of the sample as burn-in.

Although tip-dating methods that include fossil data are available (Marjanović and Laurin, 2007; Ronquist et al., 2012a; Sterli et al., 2013; Grimm et al., 2015), they are relatively rarely used. Furthermore, the phylogenetic placement of fossils, which often is unstable, is seldom considered explicitly (Pyron, 2011). Our initial trials to tip-date the phylogeny with fossils in MrBayes were unsatisfactory due to poor convergence and great sensitivity to changes in prior settings. To provide an approximation of the divergence times within Marattiales, a parsimony-based method using the stratigraphic data associated with fossils (Sterli et al., 2013) was employed instead. This approach relies on both optimizing minimum ages of nodes and incorporating the phylogenetic uncertainty in placement of fossils by means of bootstrapping trees (Sterli et al., 2013). The optimization approach of the age character is derived from the metrics of phylogenetic stratigraphic fit (Siddall, 1996, 1998; Pol and Norell, 2001), whereby the transformation costs among character states are given by a symmetrical step matrix based on the absolute time difference between each state. Following Pol and Norell (2001), the transformation costs were set to be irreversible to an older age. Hence, allowing gaps in the stratigraphic data ("ghost lineages") to be taken into account (Pol et al., 2004). Divergence times then are approximated from the branch lengths after the optimization of this age character (Sterli et al., 2013).

Depending on the completeness of the sampled fossil record, branches might be assigned zero lengths in this type of calibration methods (Wang and Lloyd, 2016). To avoid these artificial zero-length branches, they were "smoothed" by using a constant minimum ages of nodes and incorporating the phylogenetic uncertainty in placement of fossils by means of bootstrapping trees (Sterli et al., 2013). The optimization approach of the age character is derived from the metrics of phylogenetic stratigraphic fit (Siddall, 1996, 1998; Pol and Norell, 2001), whereby the transformation costs among character states are given by a symmetrical step matrix based on the absolute time difference between each state. Following Pol and Norell (2001), the transformation costs were set to be irreversible to an older age. Hence, allowing gaps in the stratigraphic data ("ghost lineages") to be taken into account (Pol et al., 2004). Divergence times then are approximated from the branch lengths after the optimization of this age character (Sterli et al., 2013).

Despite the fact that many species of Pitsana are local endemics, some are quite widespread and are distributed almost throughout the range of the genus (Murdock, 2008a). Marattia laxa, however, covers the range of Marattia as here defined with the exception of Hawaii, where a very closely related M. douglasii is present (Murdock, 2008a). Furthermore, the natural range of the most widespread species in our analysis, Angiopteris evecta, covers multiple biogeographical regions and the whole range of its genus. Thus, in the absence of complete taxonomic sampling it seems reasonable to apply genus-level ranges for these taxa. For the fossil species we only coded the positive occurrences, and coded question marks for the regions from where they were not observed.

We constructed a time-stratified palaeogeographical model taking into account changes in continental plate positions for seven time intervals: 0–30, 30–50, 50–70, 70–90, 90–110, 110–180 and 180–350 Ma, following the plate tectonic model in Scotese (2016). We applied a different connectivity matrix for each time interval (Appendix S2) and set Antarctica as an area not allowed for the latest time interval when the continent has been glaciated (Carter et al., 2017). The area connectivity matrices were constructed so that a dispersal probability of 1.0 was assigned to all areas directly connected and a probability of 0.5 to areas that were either connected through another area(s) or were separated by a relatively narrow sea. A dispersal probability of 0.01 was assigned for areas separated by wide oceanic barriers and for Northern and Southern Hemisphere continents that were not directly connected in the Pangaea-configuration (e.g. dispersal probability of 1.0 from directly connected Australia to Antarctica, and 0.5 from Australia to indirectly connected South America, but 0.01 from Southern Hemisphere Australia to Northern Hemisphere Eurasia at 180–350 Ma).

**Results**

**Plastome structure**

We assembled and annotated six plastid genomes representing the Marattiales. High-throughput sequencing ranged from 15 922 in the case of *Pitsana novoguineensis* to 347 362 plastid genome reads for *Danaea sellowiana* (Table 3). Mapping of the reads to the *de novo* plastid genomes indicated that a mean coverage ranged from ×11 (P. novoguineensis) to ×240 (D. sellowiana) (see Fig. S1). Of the assembled plastomes, that of *D. sellowiana* was the smallest in size (145 892 bp), whereas *Eupodium kaulfussii* was the largest (151 986 bp). The genome structure of marattialean ferns showed similar quadripartite structure as in the seed plants, comprising a large-single copy (LSC), small-single copy (SSC) and two inverted repeat (IR) regions. Across Marattiales, gene content seems to be stable, with 86–87 protein-coding genes, 37 tRNA genes and four rRNA genes annotated in the genomes (Fig. 1). The distribution of these genes was similar to seed plants or other fern species, with 18 genes located in the SSC, 15 genes in the IRs and

**Biogeography**

We analyzed the biogeographical history of Marattiales based on the parsimony-dated total-evidence phylogeny and by applying the dispersal-extinction-cladogenesis (DEC; Ree and Smith, 2008) model as implemented in R/BIOGEOBARS v.0.2.1 (Matzke, 2013a, b). We considered the following nine biogeographical regions in our analyses: Eurasia, Africa, Madagascar, India, Australia, North America, South America, Oceania and Antarctica. For *Marattia laxa* and *Pitsana novoguineensis*, we coded the geographical ranges of their respective genera in order to reconstruct the genus-level distribution patterns; in other extant taxa the species-level distributions covered the genus-level distributions. The application of genus ranges to species could be seen as a violation of the biogeographical model, but we consider this approach justified in our case. This is because despite the fact that many species of Pitsana are local endemics, some are quite widespread and are distributed almost throughout the range of the genus (Murdock, 2008a). Marattia laxa, however, covers the range of Marattia as here defined with the exception of Hawaii, where a very closely related *M. douglasii* is present (Murdock, 2008a). Furthermore, the natural range of the most widespread species in our analysis, *Angiopteris evecta*, covers multiple biogeographical regions and the whole range of its genus. Thus, in the absence of complete taxonomic sampling it seems reasonable to apply genus-level ranges for these taxa. For the fossil species we only coded the positive occurrences, and coded question marks for the regions from where they were not observed.
90 genes in the LSC regions (Fig. S2). The overall GC
content of the plastomes varied between 32.8 and
40.8%, whereas 29.5–31.2% of the whole genome was
noncoding, which is congruent with previous reports
(Robison et al., 2018). However, a lower GC content
was observed for *E. cicutifolium* (32.8) and *E. kaul-
fussii* (32.8), reflecting different mutation/conversion
biases in these genomes.

There were 18 intron-containing genes among
Marattiacae plastomes. Among these, 16 genes had a
single intron (ten protein-coding and 6 tRNA), and
two (*ycf3, clpP*) had two introns, whereas the *rpoC1*
intron has been lost from *D. sellowiana*. The loss of
the same intron has been reported previously in Japa-
nese climbing fern (*Lygodium japonicum*; Gao et al.,
2013). The largest intron was the group II mitochon-
drial ORF-like intron of the *trnK-UUU* gene (2387–
2427 bp) encompassing the variable *matK*. The large
gene *rps12* appeared to be trans-spliced, with one exon
located in the LSC and two exons in the IRs. We con-
irmed the presence of the *ycf66* gene across Maratti-
aceae. This highly unstable gene has been lost
independently at least four times, and has been pseu-
dogenized in ferns (Gao et al., 2011). The adjacent
*chL* and *chlN* genes were located in the SSC region,
creating an operon in the complete nucleotide
sequences of marattialean ferns, whereas the *chlB* gene
was disjunct in the LSC region. These three genes
encode the subunits of the light-independent enzyme
protochlorophyllide oxidoreductase required for
chlorophyll formation in the dark. Their presence has
been confirmed in the plastid genomes of at least some
conifers, green algae and photosynthetic bacteria, but
they are absent from major lineages of Poaceae and
Solanaceae (Nazir and Khan, 2012). The *psbC/pshD*
genes were the only overlapping genes among the plas-
tid genomes. The partial overlap of these genes encod-
ing the D2 and CP43 proteins of the photosystem II
complex could be attributed to cotranscription. In
*ccsA, rpoB* and *rps15* genes, we observed the use of
the alternative start codon ACG instead of the com-
mon AUG in *Marattia laxa*. Genes with such excep-
tional start codons are RNA edited in Asteraceae
(Sablok et al., 2019) and Solanaceae (Amiryousefi
et al., 2018). The diversity of RNA editing in ferns is
somewhat unclear, as the abundant U-to-C edits are
lacking in major lineages. There is evidence, howev-
er, for abundant C-to-U and U-to-C back edits in early
diverging (*Equisetum L.*, *Psilotum Sw.*) and in derived
leptosporangiate ferns (*Adiantum L.*, *Pteridium Gled.
ex Scop*) (Guo et al., 2015; Li et al., 2018). RNA edit-
ing might be present in *M. laxa* but confirming this
would require further transcriptomic data.

We searched for and characterized mobile elements
collectively termed as MORFFO described from fern
plastomes (Logacheva et al., 2017; Kim and Kim,
Our analysis indicated the absence of MORFFO elements in Marattiaceae except for the following cases: fractions of MORFFO1 were found between \textit{trnT} and \textit{trnFM} in \textit{Eupodium cicutifolium} and \textit{E. kaulfussii}, MORFFO2 between \textit{trnL} and \textit{ndhB} in \textit{Angiopteris evecta}, and MORFFO3 element between \textit{trnL} and \textit{ndhB} in \textit{Christensenia aesculifolia}, respectively. The MORFFO1 hit in \textit{Eupodium} over-lapped with a 936-bp-long ORF. Likewise, we found an uninterrupted ORF of >1300 bp located between \textit{trnL} and \textit{ndhB} in \textit{Christensenia} and in both \textit{Angiopteris} species. Accordingly, our BLAST search matched the MORFFO3 at the 5' end of this ORF in \textit{C. aesculifolia} and MORFFO2 at the 3' end of the ORF in the same genomic position in \textit{A. evecta}. No match was

Fig. 1. Syntenic visualization of the Marattiaceae plastomes showing a generally conserved structure. Whole-genome plastome alignment using LASTZ and visualization using Circos
found to the ORF at this position in _A. angustifolia_, and in other species no uninterrupted ORFs of significant length were found at this position. We further queried the translated amino acid sequences of these ORFs against the NCBI protein database using BLASTP (Altschul et al., 1990). The >1300-bp-long ORF in _C. aesculifolia_, _A. evecta_ and _A. angustifolia_ all matched with the annotated ORF531 in the _Mankyua chejuensis_ B.Y.Sun et al. plastome (KP205433; Kim and Kim, 2018). The 936-bp-long ORF in _Eupodium_ matched with the annotated ORF295 in the _Mankyua_ B.Y.Sun et al. plastome.

Broad-scale phylogeny and node calibration

The plastomes of Marattiaceae and other ferns were analyzed under ML and node-calibrated with Bayesian inference. All of these analyses resulted in the same tree topology (Fig. 2). _Equisetum_ was resolved as the sister to Ophioglossidae with a low support value (PP = 0.62, BS = 72), and Marattiaceae as the sister to leptosporangiate ferns (PP = 0.80, BS = 100). The family Dennstaedtiaceae, represented by _Pteridium_ in our analyses, was resolved as diverging earlier than Pteridaceae, represented by _Adiantum_ and _Myriopteris_ Fée, albeit with a low posterior probability (PP = 0.95) and no BS value. The relative phylogenetic positions of Dennstaedtiaceae and Pteridaceae have remained notoriously difficult to resolve (e.g. Schuettpelz and Pryer, 2007; Lehtonen, 2011; Rothfels et al., 2015; Qi et al., 2018; Shen et al., 2018; Lehtonen and Cárdenas, 2019). Within Marattiaceae, _Danacea_ was resolved as the sister to the remaining taxa in all the analyses, but not with maximum support value (PP = 0.97, BS = 94). _Maratta cicuitifolia_ was sister to _Eupodium kaulfussii_ with maximum support value, whereas _M. laxa_ was resolved in a distinct position as sister to _Christensenia_, but without maximum support value (PP = 0.89, BS = 94). The _Christensenia-M. laxa_ clade was sister to _Angiopteris_, and these together formed a sister clade to the _Pisana-Eupodium-M. cicuitifolia_ clade.

Divergence time analysis using BEAST did not converge well, as can be seen from the trace plot (Fig. 2). The effective sample size (ESS) values were >145 for each parameter. In the maximum clade credibility tree, ferns diverged from the seed plants at 420 Ma (95% CI: 409–427) and the fern crown group diverged at 388 Ma (95% CI: 359–420). Marattiales originated at 330 Ma (95% CI: 311–347). In most cases the specified priors, as they were set up, closely matched the effective prior distributions as observed through running the analysis without data (Fig. 2). This was not the case with the seed plant prior, and even less so with the Marattiales prior, where the exponential prior was set up with an offset at 323.3 Ma, but the effective prior peaked at 410 Ma.

Marattialean phylogeny and parsimony dating

The phylogenetic relationships within Marattiales were analyzed in more detail by expanding the plastome data with mtDNA and morphology; the latter data also coded for selected fossil terminals. Analyses of these data under Bayesian inference and parsimony resulted in largely congruent topologies (Fig. 3). Implied weighting in parsimony analyses did not change the topologies except for the analyses including fossils, in which case the resolution—and to some degree the topology—varied depending on the _k_-value. The parsimony analysis of phenotypic data including only extant taxa resolved _Danacea_ as the first diverging lineage within Marattiales, followed by _Christensenia_. The remaining species formed a clade with the two species of _Angiopteris_ resolved as sisters, as were _E. kaulfussii_ and _M. cicuitifolia_. _M. laxa_ and _P. novoguineensis_ remained unresolved within this clade. The topology remained the same across all the _k_-values. The Bayesian analysis resulted in basically the same topology, but with _P. novoguineensis_ and _M. laxa_ resolved as successively diverging lineages on the branch leading to _Angiopteris_.

The topological arrangements within the extant species slightly changed when the extant taxa were included in the analysis. In parsimony analyses, the positions of _M. laxa_ and _P. novoguineensis_ somewhat varied under equal weighting and _k_ = 5, but with higher _k_-values they were constantly resolved as successively diverging lineages on the branch leading to _Angiopteris_. In the Bayesian analysis the inclusion of fossil taxa made _Danacea_ and _Christensenia_ sisters and collapsed _P. novoguineensis_ and _M. laxa_ to the same position as in parsimony analysis of extant taxa only. Beyond the extant species, the limited resolution present in Bayesian and equally weighted parsimony trees was largely congruent. In both trees, _Qasimia_ Hill et al. (with two species) was sister to the clade that included all the extant Marattiales. The Bayesian analysis also resolved _Daneaopsis_ Heer ex Schimp. (with three species, one of them missing a nomenclatural combination in _Daneaopsis_) within this clade. The fossil taxon _Marattiopsis vodrakae_ Kváček was quite consistently resolved as sister to the extant _Eupodium kaulfussii_, except only in parsimony analysis with _k_ = 5. The remaining species of _Marattiopsis_ were somewhat unstable, although constantly placed within the same clade with the extant
Marattiaceae. *Angiopteris blackii* Van Cittert, another fossil taxon, constantly formed a clade with the extant *Angiopteris*. The Bayesian analyses and all of the parsimony analyses, except the equally weighted analysis, resolved *Sydneia* Pšenička et al. and *Radstockia* Kidston as a clade sister to the remaining Marattiaceae. Psaroniaceae mostly remained unresolved in the Bayesian and equally weighted parsimony analyses, but under implied weighting this extinct family was almost constantly resolved as sister to Marattiaceae (with *Sydneia* and *Radstockia* as a sister clade to the Psaroniaceae-Marattiaceae clade). The relationships within Psaroniaceae varied depending on the k-value, but *Danaeites* Goeppert and *Millaya* Mapes & Schabilion formed a sister clade to the remaining family almost constantly.

Fig. 2. Plastome phylogeny of ferns and seed plant outgroup. (a) Node-calibrated Bayesian maximum clade credibility tree with 95% HPD interval for node ages presented in horizontal bars. Asterisks denote calibrated nodes, specified calibration priors are shown with dotted lines, density distributions sampled from prior only (effective priors) are shown for these nodes in green, and posterior density distributions in blue. Nodal support values (Bayesian PP/ML bootstrap) are shown if not maximally supported. For the marattialean ferns, the tree dated with parsimony bootstrapping (see Fig. 4) is shown in the background with blue dotted lines. (b) The ML tree for the same plastome data. (c) Post-burn-in likelihood traces of the 14 chains are combined to show the sampled joint probability and the poor convergence of the chains. A horizontal line is provided for illustrative purposes.
The plastome data resulted in somewhat different topologies depending on the optimality criterion. Parsimony analysis resolved Danaea and Christensenia as successive lineages leading to the remaining Marattiaceae with maximum support value, Ptisana with maximum support value as sister to the E. kaulfussii-M. cicutifolia clade and M. laxa as sister to Angiopteris, albeit without support value. By contrast, the Bayesian analysis resolved Christensenia as sister to M. laxa (PP = 0.99), and these together as sister to Angiopteris; this clade was sister to a clade in which Ptisana was sister to the E. kaulfussii-M. cicutifolia clade (PP = 1).

The mtDNA analyses resulted in congruent topologies with high support values across all the analyses. Danaea was resolved as the sister to the remaining Marattiaceae, Christensenia as sister to M. laxa, and these together as sister to Angiopteris; this clade was sister to a clade in which Ptisana was sister to the E. kaulfussii-M. cicutifolia clade.

The parsimony analysis of phenotypic and DNA data of extant taxa only resulted in the same topology clade where Ptisana was sister (PP = 1) to the E. kaulfussii-M. cicutifolia clade (PP = 1).
as mtDNA but with decreased support value for the sister relationships between the Christensenia-M. laxa clade and the sister relationships of this clade to Angiopteris. Bayesian analysis resulted in the same topology as the plastome data with Danaea as sister to the Ptisana-Eupodium clade (PP = 1). In the total-evidence analyses (fossils included), the extinct Angiopteris blackii was resolved as sister to the extant Angiopteris, and this clade was sister to the Christensenia-M. laxa clade. This resolution was recovered across the parsimony analyses of varying k-values and the Bayesian analysis. Although M. cicutifolia and E. kaulfussii were sisters and together formed a sister clade to M. vodrazkae in the Bayesian analysis, in the parsimony analyses M. vodrazkae was constantly the sister of E. kaulfussii and these formed the sister clade of M. cicutifolia. These three terminals formed the sister group of Ptisana in all of the analyses. In the Bayesian analysis, the relationships of these two clades with Danaea and the remaining species of Marattioptis remained unresolved. In the parsimony analyses, Danaea was constantly sister to Marattioptis and the remaining extant Marattiaceae. Within Marattioptis, M. anglica Thomas, M. patagonica Escapa et al. and M. asiatica Kawasai formed a clade under the implied weights but remained unresolved under the equal weights. In the Bayesian analysis and in parsimony analyses with k = 5 and k = 10, the fossil genus Danaeopsis was resolved as the sister to the clade of the extant Marattiaceae and fossil Marattioptis; these together formed a sister clade to the fossil genus Qasimia. With the value k = 15 (and k = 20) the relative positions of Qasimia and Danaeopsis were interchanged so that Qasimia was the sister of the most exclusive clade containing all the extant Marattiaceae. In all analyses, Qasimia, Danaeopsis, Marattioptis and the extant Marattiaceae formed a clade that excluded all the remaining species in the analyses and are here considered as members of the family Marattiaceae.

The remaining species, that is Psaroniaceae, either remained largely unresolved (Bayesian and equally weighted parsimony) or had good but contradicting resolutions (implied weighted parsimony). However, Sydneia manleyi and Radstockia kidstonii Taylor were constantly resolved as sisters of each other, and they almost constantly formed the sister lineage to the remaining Marattiaceae (implied weighted parsimony) or were resolved outside the remaining Marattiaceae (Bayesian). Furthermore, Danaeites rigida Gu & Zhi and Millaya tularosana Mapes & Schabiliion constantly formed a clade, either in an unresolved position (Bayes, equally weighted parsimony) or they were a sister of Marattiaceae (k = 5, k = 10), or sister of a large clade including Marattiaceae and the remaining Psaroniaceae (k = 15, k = 20). Araiangium pygmaeum (Graham) Millay (k = 15, k = 20) and Gemellitheca saudica Wagner et al. (k = 5, k = 10) also were invariably resolved outside of the Psaroniaceae, but otherwise the family remained a monophyletic sister of Marattiaceae in the implied weighted parsimony analyses.

The parsimony dating of the total-evidence phylogeny resulted in an age estimate ranging from 307 Ma (BUR_{min}) to 320 Ma (BUR_{max}) for the Marattiaceae. The split between Danaea and the rest of the extant Marattiaceae was estimated to have occurred in the Late Triassic (201–236 Ma). Angiopteris was estimated to have separated during the Late Triassic–early Jurassic at the same time with the splitting of Christensenia-M. laxa and Ptisana-Eupodium clades. The estimated age range for the Christensenia-M. laxa split was extremely large due to the lack of any internal fossils and phylogenetic uncertainty in bootstrapping. However, the weighted mean age computed over the BS replicates placed this split at the late Jurassic. The split between Ptisana and Eupodium was placed at the Late Cretaceous. In this case, the fossil Marattioptis vodrazkae provided an internal calibration point as sister of E. kaulfussii, but because the phylogenetic position of Ptisana was stable across the BS replicates, it was fixed in the time tree just below the Eupodium.

**Biogeographical reconstructions**

Our biogeographical reconstruction based on the DEC model and parsimony-dated phylogeny supports the North American origin of Marattiaceae during the Pennsylvanian (Fig. 4). The branches leading to Marattiaceae and the most recent common ancestor (MRCA) of the extant Marattiaceae were reconstructed as having a Eurasian origin. According to the reconstructed history, the genus Danaea reached its current South American distribution from the ancestral Eurasian range through North America while these continents were still connected. Marattioptis patagonica was reconstructed to have spread via the same route independently. The Ptisana-Eupodium clade shows dispersal from the ancestral Eurasian range to Africa followed by expansion throughout the Gondwana that was still intact at that time, with a more recent split between the palaeotropical Ptisana and Antarctic-Neotropical Eupodium (including M. vodrazkae). The origin of the Christensenia-M. laxa clade was linked with dispersal from the ancestral Eurasian range to the directly connected Oceania, from where the Marattia was modelled to have reached its current Central American–Hawaiian range through Eurasia. Angiopteris originated within the ancestral Eurasian range, from where it was estimated to have reached its current range in Madagascar–India–Oceania–Australia.
by dispersal through Africa, where the genus is currently not present, to Oceania.

Within the fossil Psaroniaceae, several range expansions from the ancestral range in North America to Eurasia, South America, and Africa were inferred. These continents were connected at the time of these range expansions. The sole survivor of *Scolecopteris* Zenker into the Mesozoic, *S. antarctica* Delevoryas et al. (Delevoryas et al., 1992), was modelled to have dispersed from North America to Antarctica through Africa; all of these continents were connected during that time.

**Discussion**

**Marattialean phylogenetics**

Our results agree with the now widely supported view that Marattiales is the sister lineage of the leptosporangiate ferns (Knie et al., 2015; Lu et al., 2015; Rothfels et al., 2015; Qi et al., 2018; Lehtonen and Cardenas, 2019). The persisting questions include the phylogenetic relationships between the extant genera—especially their rooting (Murdock, 2008b)—as well as how the extinct *Marattiopsis* is related to the extant
taxa (Escapa et al., 2014; Kvaček, 2014), and how the Palaeozoic forms are related to the morphologically modern forms (Mamay, 1950; Stidd, 1974; Hill et al., 1985).

Murdock (2008b) resolved the root of extant Marattiaceae on the branch connecting Danaea with the remaining genera but noted that the root position was sensitive to the dataset, outgroup choice and optimality criteria used. We also found the root on this branch in most analyses, but despite much larger datasets the alternative phylogenetic resolutions and root positions persisted depending on the data and optimality criterion applied. Murdock (2008b) noted that the branch leading to Danaea is the longest internal branch and therefore this rooting could represent a case of long-branch attraction (Bergsten, 2005). We recovered this root position in all of the analyses except the Bayesian analyses of plastome and total-evidence data. The same root position also was recovered in the ML analysis of 146 nuclear genes by Qi et al. (2018), who sampled four genera of Marattiaceae (Marattia and Eupodium were not sampled). Our total-evidence analyses also effectively broke down the long branch leading to Danaea, yet recovered the same root position under the parsimony analysis. However, in the Bayesian total-evidence tree the relationships between Danaea, Ptisana-Eupodium and Angiopteris-Marattia-Christensenia clades remain uncertain, and with extant taxa only, the root position is displaced from the Danaea branch. It has been noted that the Mk model may not be suitable for morphological datasets (Goloboff et al., 2019). However, as we recovered the root along the Danaea branch in the Bayesian analyses for phenotypic data (albeit with the topologies slightly different from the parsimony topologies), this may not be the problem in our case. Instead, it appears that the root position in the Bayesian analyses of plastome data depends on the outgroup; so that the analyses with Osmundastrum as the sole outgroup misplace the root as compared to the analysis incorporating a wider sampling of fern and seed plant plastomes.

Christensenia is a distinctive, and in many respects, highly autapomorphic genus (Hill and Camus, 1986a; Murdock, 2008b; Liu et al., 2019). This genus remained somewhat unstable in our analyses. Unlike Murdock (2008b) or Rothwell et al. (2018), we found that the most plausible phylogenetic position for Christensenia is a sister relationship with Marattia. An alternative position, supported by the plastome data under parsimony and phenotypic evidence, resolved Christensenia as a distinct lineage splitting off after Danaea had been separated from the rest of the Marattiaceae. Our analyses also recovered Marattia cicutifolia, a species previously not included in phylogenetic analyses, as more closely related to E. kaufmannii than M. laxa (M. laxa is the sister of M. alata Sw., the type of the genus; Murdock, 2008a). Hence, we transfer M. cicutifolia into Eupodium (see Taxonomy below).

The systematic significance of the name Marattiosis has recently been re-evaluated by palaeobotanists (Bomfleur et al., 2013; Escapa et al., 2014; Kvaček, 2014). After noting that several morphological features defining the genera now segregated from Marattia s.l. can be found mixed in the fossil taxa, they concluded that a broadly defined (i.e. paraphyletic) Marattiosis is needed to accommodate the fragmentary fossil material (Bomfleur et al., 2013; Escapa et al., 2014; Kvaček, 2014). However, Escapa et al. (2014) considered most of the fossil species to probably be closest to Ptisana. We coded five Marattiosis species for our analyses and found that M. vodrazkae from the Campanian of Antarctica was resolved in Eupodium, whereas the position of M. aganzhenensis (Yang et al.) Escapa et al. remained ambiguous and the remaining three species formed a clade in our preferred tree ($k = 15$, the same topology also was recovered under $k = 5$ and $k = 20$). It may therefore be possible to assign at least some fossils into the modern segregate genera, most notably M. vodrazkae and another unnamed but similarly stalked marattialean synangia from Antarctica (Vera and Césari, 2016) into Eupodium. It should be noted, however, that Rothwell et al. (2018) resolved these taxa differently.

The appearance of apparently modern Marattiaceae in the Mesozoic with distinctive morphologies compared to the Palaeozoic forms has puzzled palaeobotanists (Mamay, 1950; Stidd, 1974; Delevoryas et al., 1992; Zhifeng and Thomas, 1993). Our results suggest that the Permian Qasimia and Triassic Danaeopsis, two genera with taeniopteroid foliage, are among the oldest Marattiaceae. The sporangia of Qasimia are bilaterally fused into a synangia closely resembling some of the extant forms (Hill et al., 1985). However, in Danaeopsis the sporangia are arranged in two rows covering the entire lower lamina and are not fused into a synangia as in extant forms (Kustatscher et al., 2012). This topology is somewhat comparable with the results of Rothwell et al. (2018), who resolved these genera nested within the Marattiaceae. Apparently Qasimia never survived the Permian mass extinction, but Danaeopsis became an important component of the Triassic floras (Kustatscher et al., 2012). Some of our analyses ($k = 5$, $k = 20$, Bayesian) resolved these genera in reversed positions, which better fits the stratigraphy.

Considering Psaroniaceae, we found support for recognizing the subfamily Sydneideae extended to include not only Sydheia but also Radstockia, another genus with sphenopteroid foliage. We thus resolved this subfamily outside of the family, as sister to the remaining Marattiaceae. Rothwell et al. (2018) also found Radstockia as sister to the rest of Marattiaceae, although
their analysis placed *Sydneia* in a distant position deeply embedded in Psaroniaceae. The importance of *Radstockia* in understanding marattialean evolution has been recognized earlier and our finding may help to understand the origin and still unknown relationships of the marattialeans with other Palaeozoic ferns (Mamay, 1950; Stidd, 1974; Delevoryas et al., 1992). However, further resolution requires a taxonomically broader total-evidence analysis that also would include fossils related to Osmundales (Wang et al., 2014; Grimm et al., 2015) and other Palaeozoic ferns (Rothwell and Nixon, 2006). The close relationship between *Millaya* and *Danaeites* recognized here has been suggested before (Mapes and Schabilion, 1979b), but our preferred tree excludes them from the Psaroniaceae. Because the resolution within Psaroniaceae remained highly unstable, the relationships within this family are not discussed further. We note, however, that *Gemelolitheca* Wagner et al. and *Buritiranopteris* Tavares et al. generally were resolved as early radiations and often as sisters, thus supporting similar resolutions for these taxa by Rothwell et al. (2018), and that we generally recovered monophyletic *Scolecopterus*.

**Time-calibration and historical biogeography of Marattiales**

Tip-dating phylogenies by coding fossil taxa as such into a total-evidence analyses has been considered as an alternative to the more widely used indirect dating approach of calibrating nodes with minimum age estimates taken from the fossil record (Ronquist et al., 2012a; Grimm et al., 2015). The inherent problem of the latter approach is the uncertainty in the phylogenetic placement of fossil calibration points (Ronquist et al., 2012a). Another problem is that multiple calibration priors may interact with each other and with other priors and thus influence the effective priors in an unpredictable way (Grimm et al., 2015). We included multiple calibration priors into our random local clock analysis of taxonomically broad plastome data to better model the presumed rate shifts. However, the result was that the effective priors in some cases strongly deviated from the specified priors and despite running a large number of long chains, the runs never properly converged. Our attempts to tip-date the marattialean tree using a Bayesian total-evidence approach likewise failed due to convergence issues. Hence, we relied on a model-free parsimony dating approach (Sterli et al., 2013) to infer the phylogenetic position of the fossil taxa and the uncertainty ranges for the node ages.

Although these methods are not commonly used in phylogenetic analyses of extant taxa, model-free calibration approaches have been applied before to datasets dominated by fossils (e.g. Laurin, 2004; Marjanović and Laurin, 2007; Wang and Lloyd, 2016; Lloyd et al., 2016). In this latter case, where missing data are abundant, model assumptions regarding character change rates—and subsequent branch lengths—are not always fulfilled (Lloyd et al., 2016). Consequently, approaches assigning minimum node ages on the basis of stratigraphic data might be deemed as more conservative under those circumstances (Wang and Lloyd, 2016; Lloyd et al., 2016). However, it should be noted that these calibration methods are still prone to produce zero-length branches and subsequently misestimate divergence times when the fossil record is poorly sampled (Wang and Lloyd, 2016). To avoid this problem, authors have commonly employed a fixed minimum branch length (Marjanović and Laurin, 2007; Lloyd et al., 2016). Nevertheless, setting a minimum branch length is hardly straightforward because it is dependent on the taxonomic sampling, tree topology and the quality of the fossil record, hence arbitrarily fixed most of the time (Marjanović and Laurin, 2007; Sterli et al., 2013; Wang and Lloyd, 2016). In our analyses, the selection of a minimum branch length followed Sterli et al. (2013).

The parsimony dating resulted in a minimum age estimate of 201–236 Ma (Late Triassic) for the most recent common ancestor (MRCA) of the extant Marattiaceae. This closely corresponds with the Bayesian estimations of 185–224 Ma for this node by Lehtonen et al. (2017), who considered *Marattiopsis asiatica* as a member of *Ptisana* and hence constrained the origin of *Ptisana* at 176 Ma. They also estimated a similar age (214–242 Ma) for the origin of Marattiaceae directly from the fossil evidence (Lehtonen et al., 2017). Smith et al. (2010) also dated the MRCA of extant Marattiaceae at the Triassic–Jurassic boundary by using an internal calibration of *Marattia-Angiopteris* split at 166.1 Ma. By contrast, Testo and Sundue (2016) did not assign any calibration points inside Marattiaceae and they obtained a much younger estimate of 154–165 Ma for the MRCA of Marattiaceae. Even more dramatically different was our poorly converged BEAST estimate of just 76–06 Ma for this node (Fig. 2), also inferred without internal calibration points within Marattiaceae. Marattialean ferns are known to have an anomalously slow rate of molecular evolution (Solits et al., 2002), perhaps reflecting their long generation time (Sharpe, 1993), as suggested previously for the tree ferns with a similarly slow rate of molecular evolution (Korall et al., 2010). Consequently, correctly modelling the evolutionary rate of Marattiales in the absence of close relatives may be extremely difficult and we consider our parsimony approach better justified in this case.

Our DEC model suggested a North American origin for Marattiales as a whole, as well as for Psaroniaceae, and a Eurasian origin for the Marattiaceae. According to this scenario, *Danaea* dispersed to its current
neotropical range through North America and M. patagonica independently dispersed the same route. Danaea fossils have been reported from North America, although they have been later considered misidentified (Collinson, 2001). The distribution of other extant genera likewise follow the patterns of continental breakup during the Mesozoic. The sister genera Pitisana and Eupodium are biogeographically disjunct; the former has a Paleotropical range and the latter Neotropical. According to our model, this pattern is a result of dispersal from Eurasia through Africa to southern Pangea, while all of these continents were still connected, and then splitting between the Neotropical-Antarctic Eupodium lineage and Paleotropical Pitisana. In our tree (Fig. 4), this split is dated by the Campanian Marattia-Cristensenia vodrazi cae (Kvaček, 2014). However, if the similarly stalked synangia from the Lower Cretaceous of Antarctica (Vera and Césari, 2016) also belong to Eupodium, the split between Pitisana and Eupodium would perfectly match the breakup of Africa from the Antarctica (Jokat et al., 2003). Murdock (2008b) briefly commented on the biogeography of Marattia by considering the following two alternative scenarios: the current distribution in the American tropics and Hawaii was achieved either by dispersal from the American continent to Hawaii, or vice versa. Our model suggests that the ancestor of the Marattia-Christensenia lineage dispersed from Eurasia to Southeast Asia and thus favours the Oriental origin of Marattia. Furthermore, our model suggests a Eurasian origin for Angiopteris. Thus, each of the extant main clades seem to have originated in different parts of the Pangea as it was breaking up. This separation was probably promoted by palaeoclimatic patterns with Marattia-Christensenia restricted to far east, Angiopteris to Eurasia, Danaea to North America, and Pitisana-Eupodium to the Southern Hemisphere, where moist climates were available (see Escapa et al., 2014: fig. 8).

Plastome evolution

Comparative genome analysis across embryophytes has greatly improved our understanding about plastid genome evolution. The first plastid genome studies revealed previously unreported features of marattialean plastid genome evolution. The matK gene is located within the trnK intron in all embryophytes except in non-Osmundales leptosporangiate ferns, where this gene has lost its flanking regions (Wolf et al., 2010). The presence of the trnK intron has previously been reported from Osmundales (Duffy et al., 2009) and from Angiopteris evecta (Roper et al., 2007). Our study confirms that the loss of trnK appeared in non-Osmundales leptosporangiate ferns by confirming the presence of trnK in all extant marattialean genera. Previous studies reported the pseudogenization of the highly divergent gene coding for the hypothetical protein ycf1 in A. evecta (Roper et al., 2007). This gene is interspersed by an 817-bp direct repeat, which is missing from the other marattialean fern plastid genomes. Thus, the pseudogenization of this gene seems to be autapomorphic in A. evecta, because ycf1 genes remain functional in other Marattiaceae. The divergence of this specific gene is not surprising because many other insertions and deletions have been reported from other embryophytes. For example, the degradation of this gene is almost exclusively accounted for by the plastome size reduction in the graminid clade of Poaceae (Pozczi and Hyvönen, 2017).

The gene order in Marattiaceae showed high similarity to those observed in seed plants as members of the family showed only the trnG-trnT inversion characteristic to all ferns. The plastid genome structure evolution in Marattiaceae has remained constant, because the sequenced plastomes showed high synteny among all extant genera (Fig. 1). This could be correlated with evolutionary trends in the family implicating a correlation between slower genome structure and morphological evolution, also suggested by Roper et al. (2007). A recent study highlighted the importance of MORFFO-type mobile elements adjacent to inversion sites in shaping the plastid genome evolution of ferns (Robison et al., 2018). The origin and function of these elements are unknown, but they appear to be linked with the structural genome evolution. We identified and characterized these elements in Marattiaceae and found that MORFFO elements are absent from Danaea, Marattia and Pitisana but they are partially present in intergenic-spacer regions (IGS) of Angiopteris, Christensenia and Eupodium. The lack of relative abundance of such elements could partially explain the absence of dynamic genomic re-arrangements in Marattiaceae. However, the observed MORFFO2 and MORFFO3 elements in Angiopteris and Christensenia matched a single ORF (ORF531) in the plastome of Mankuya (Ophioglossales), suggesting that these elements may share a common origin in early diverging ferns, even if they are often found in distinct positions within the plastomes of the leptosporangiate ferns (Lehtonen and Cárdenas, 2019). Kim and Kim (2018) suggested that this ORF may be of
bacterial origin. We further matched the MORFFO1 element found in the two *Eupodium* species with the ORF295 in *Mankuya*. This ORF was found to be similar to the protein found in green alga (Kim and Kim, 2018). In *Eupodium* this ORF is associated with tandem repeats, and is located within a 1378-bp-long insertion. It was speculated that such plastome ORF insertions could originate from intercellular transfer from the mitochondrial genome (Logacheva et al., 2017). Some of these ORFs encode functional proteins, whereas others contain conserved domains or have become pseudogenes. Such structural changes in the plastome show greater complexity as compared to simple nucleotide substitutions (e.g. Poczai and Hyvönen, 2013), and such changes could be good phylogenetic markers.

**Taxonomy**


The genus *Eupodium* has been considered to contain two (Murdock, 2008a) or three (Christenhusz, 2010) species. The genus is diagnosed based on the presence of usually only a single leaf at a time, stalked synangia, and awns on the midrib of the pinnae. Murdock (2008b) did not sample *M. cicutifolia*, but nevertheless placed the species in *Marattia* s.s. because it has morphological characters more typical of that genus, despite of having shortly stalked synangia in some cases (Murdock, 2008a). Another morphological character common to *M. cicutifolia* and *Eupodium* is the spinulose exosporangiate ornamentation (del Carmen Lavalle et al., 2011), which is quite similar to that found in *M. cicutifolia* (Camus, 1990; Murdock, 2008a). The absence of awns, the presence of multiple leaves simultaneously, and general synangial and foliar characteristics in *M. cicutifolia* are more similar to *Marattia* and *Pitsana*. The phylogenetic position of *M. cicutifolia* in this study necessitates its transfer to *Eupodium*, but at the same time partly negates the morphological circumscription of the genus. Under the current circumscription, only the stalked synangia is diagnostic for *Eupodium*, but this character is not always clearly developed in *E. cicutifolium*. Biogeographically *Eupodium* remains a South American genus with a few occurrences in Central America and Hispaniola (Christenhusz, 2010), whereas *Marattia* is now restricted to Central America, Hawaii, Jamaica and Cuba (Murdock, 2008a).

**Acknowledgements**

This work was financially supported by the Finnish Cultural Foundation grant (SL) and Academy of Finland grant no. 295595 (JH), which supported Gaurav Sablok. DNK received funds from the Swiss National Funds (SNF) grant no. 148691. We acknowledge the Willi Hennig Society for making TNT freely available.

**References**


and Marattiales are sister to leptosporangiate ferns. Mol. Phylogenet. Evol. 90, 140–149.


Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1. Plastome coverage and gene track visualization of the sequenced plastomes. Plastome coverage was evaluated by remapping the corresponding reads to the plastomes, and gene tracks were derived from the corresponding GFF files.

Figure S2. Maps of the newly generated plastomes.

Table S1. Gene-wise alignment statistics and informative sites per gene alignment in the plastome analysis of the ferns with seed plant outgroups.

Appendix S1. Data partitioning and evolutionary models in the Bayesian analyses.

Appendix S2. Biogeographical connectivity matrices through time.

Appendix 1

Taxa investigated with literature references and herbarium specimens studied for the character coding.

Figure S1. Plastome coverage and gene track visualization of the sequenced plastomes. Plastome coverage was evaluated by remapping the corresponding reads to the plastomes, and gene tracks were derived from the corresponding GFF files.

Figure S2. Maps of the newly generated plastomes.

Table S1. Gene-wise alignment statistics and informative sites per gene alignment in the plastome analysis of the ferns with seed plant outgroups.

Appendix S1. Data partitioning and evolutionary models in the Bayesian analyses.

Appendix S2. Biogeographical connectivity matrices through time.

Appendix 2

Phenotypic character descriptions and argumentation.

1. Root pith sclerified: (0) no; (1) yes. Similar to Hill and Camus (1986a), char. 12; Pryer et al. (1995), char. 21; Rothwell (1999), char. 48; Murdock (2008b), char. 1; Rothwell et al. (2018), char. 6. Uninformative in the current matrix.

1. Continuous band of sclerenchymatous fibres in cortex of root: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 13; Murdock (2008b), char. 2; Rothwell et al. (2018), char. 7.

2. Root hairs: (0) absent; (1) present. Similar to Stevenson and Loconte (1996), char. 3; Jud et al. (2008), char. 1267; Schneider et al. (2009), char. 41.

3. Root hair structure: (0) unicellular; (1) multicellular. Similar to Hill and Camus (1986a), char. 15; Stevenson and Loconte (1996), char. 3; Schneider et al. (2009), char. 42; Rothwell et al. (2018), char. 9. Uninformative in the current matrix.

4. Root anatomy: (0) diarch; (1) polyarch. Similar to Pryer et al. (1999), char. 16; Pryer et al. (1995), char. 35; Schneider et al. (2009), char. 45; Rothwell et al. (2018), char. 8. Uninformative in the current matrix.

5. Outer root cortex: (0) sclerenchymatous; (1) parenchymatous. Similar to Pryer et al. (1995), char. 94; Schneider et al. (2009), char. 48. Uninformative in the current matrix.
1. **Root mantle**: (0) absent; (1) present. Similar to Rothwell et al. (2018), char. 4.

2. **Stem symmetry**: (0) radial; (1) dorsiventral. Similar to Hill and Camus (1986a), chars. 17 and 19; Pryer et al. (1995), char. 26; Stevenson and Loconte (1996), char. 8; Rothwell (1999), char. 5; Murdock (2008b), char. 3; Jud et al. (2008), char. 128; Schneider et al. (2009), char. 25; Rothwell et al. (2018), chars. 1 and 12.

3. **Stem shape**: (0) elongate; (1) squat. Similar to Hill and Camus (1986a), char. 18; Rothwell et al. (2018), char. 2.

4. **Stem sclerenchyma**: (0) absent; (1) present. Similar to Hill and Camus (1986a), chars. 11 and 14; Schneider et al. (2009), char. 32; Wang et al. (2014), char. 15; Rothwell et al. (2018), char. 18.

5. **Mature shoot stele type**: (0) euctaphic; (1) siphonosteole; (2) dictyosteole. Similar to Hill and Camus (1986a) chars. 1–4; Pryer et al. (1995), char. 27; Stevenson and Loconte (1996), char. 19; Rothwell (1999), chars. 27–39; Jud et al. (2008), char. 1232–1233; Schneider et al. (2009), char. 26; Wang et al. (2014), char. 3 and 10; Rothwell et al. (2018), char. 14.

6. **Vascular stele cycles**: (0) monocyclic; (1) polycyclic. Similar to Hill and Camus (1986a), char. 2; Pryer et al. (1995), char. 28; Rothwell (1999), char. 34; Schneider et al. (2009), char. 27; Rothwell et al. (2018), char. 16.

7. **Pneumathodes**: (0) absent; (1) scattered all around petiole. Similar to Schneider et al. (2009), char. 13. Uninformative in the current matrix.

8. **Lamina margins of pinnae or pinnules excluding apex**: (0) reflexed; (1) not reflexed. Similar to Hill and Camus (1986a), char. 44; Liu et al. (2000a), char. 5; Rothwell et al. (2018), char. 61.

9. **Fertile pinnules**: (0) not reflexed; (1) reflexed to protect the sori. Similar to Hill and Camus (1986a), char. 38; Rothwell et al. (2018), char. 59.

10. **False veins**: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 44; Liu et al. (2000a), char. 5; Rothwell et al. (2018), char. 61.

11. **Forking of veins in fertile frond**: (0) mainly simple; (1) normally forking near the midrib; (2) forking in the lamina; (3) irregularly compound with cuneate segments; (4) incised. Similar to Hill and Camus (1986a), char. 37; Liu et al. (2000a), char. 7; Rothwell et al. (2018), char. 60.

12. **Pinnule base**: (0) rounded; (1) auriculate cordate acute. Similar to Hill and Camus (1986a), chars. 45; Murdock (2008a). The extant members of *Ptisana* typically have sheathing structures called stipules at the bases of rachises (Hill and Camus, 1986a). Osmundalean ferns often have sheathing structures called stipules at their rachis bases, but these structures are not considered homologous with the marattialean stipules.

13. **Reflexed margin completely envelopes the sporangia**: (0) no; (1) yes. Similar to Rothwell et al. (2018), char. 62.

14. **Vein density**: (0) ≤ 7 per cm; (1) 9–16 per cm; (2) > 30 per cm. Similar to Hill and Camus (1986a), chars. 6 and 26–28; Pryer et al. (1995), char. 15; Stevenson and Loconte (1996), char. 32; Murdock (2008b), chars. 4–7; Schneider et al. (2009), char. 12; Rothwell et al. (2018), chars. 109–113.

15. **Soft spines on veins**: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 39; Murdock (2008b), char. 33; Rothwell et al. (2018), char. 51.

16. **Sutures at pinnule bases**: (0) absent; (1) present. Based on *Pitsana* typically have sutures at pinnule bases (Murdock, 2008a) and, because some of the fossil species are always preserved as detached pinnules, it has been suggested that they also had sutures (Escapa et al., 2014; Kvaček, 2014). However, in the lack of clear evidence these fossils were here coded as having unknown state. Uninformative in the current matrix.

17. **Ultimate pinnule size**: (0) ≤ 1.5 cm; (1) 1.5–4 cm; (2) > 4 cm. Similar to Wang et al. (2014), char. 25; Rothwell et al. (2018), char. 30.
42. Stomatal type: (0) anomocytic; (1) cyclocytic. Similar to Hill and Camus (1986a), char. 31; Pryer et al. (1995), char. 12; Stevenson and Loconte (1996), char. 35; Liu et al. (2000a), char. 9; Schneider et al. (2009), char. 56; Rothwell et al. (2018), char. 114.

43. Stomatal shape: (0) elliptical; (1) circular. Similar to Murdock (2008b), char. 19; Rothwell et al. (2018), char. 116. Uninformative in the current matrix.

44. Stomatal guard cells: (0) functional; (1) permanently open. Similar to Hill and Camus (1986a), char. 30; Murdock (2008b), char. 18; Rothwell et al. (2018), char. 115. Uninformative in the current matrix.

45. Stomatal complex: (0) surficial or slightly sunken; (1) raised. Similar to Murdock (2008b), char. 20; Rothwell et al. (2018), char. 117.

46. Idioblasts in abaxial epidermis: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 36; Murdock (2008b), char. 23; Rothwell et al. (2018), char. 35.

47. Laminar idioblast groupings: (0) solitary or small clusters; (1) dense areas of idioblasts. Similar to Murdock (2008b), char. 24; Rothwell et al. (2018), char. 36.

48. Laminar tissue expanded between sori/sporangia: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 50; Liu et al. (2000a), char. 15; Murdock (2008b), char. 10; Rothwell et al. (2018), char. 55.

49. Sporangial grouping: (0) absent; (1) present. Similar to Pryer et al. (1995), char. 47; Rothwell (1999), char. 65; Schneider et al. (2009), char. 59.

50. Sporangial stalk width: (0) four to six cell rows wide; (1) more than six cell rows. Similar to Pryer et al. (1995), char. 45; Stevenson and Loconte (1996), char. 63; Rothwell (1999), char. 71; Schneider et al. (2009), char. 74. Uninformative in the current matrix.

51. Sporangial development: (0) eusporangiate; (1) lepto sporangiate. Similar to Pryer et al. (1995), char. 42; Stevenson and Loconte (1996), char. 48; Rothwell (1999), chars. 73 and 79; Schneider et al. (2009), char. 72. Uninformative in the current matrix.

52. Annullus: (0) absent; (1) present. Similar to Pryer et al. (1995), char. 56; Stevenson and Loconte (1996), char. 64; Rothwell (1999), char. 74; Liu et al. (2000a), char. 21; Schneider et al. (2009), char. 76; Rothwell et al. (2018), char. 63. Uninformative in the current matrix.

53. Shape of sporangia: (0) elongate; (1) ovoid; (2) obovoid; (3) globular. Similar to Cleal (2015), char. 4.

54. Sporangial aperture development: (0) apertures develop longitudinally; (1) apertures develop apically; (2) apertures develop vertically. Similar to Hill and Camus (1986a), char. 69; Liu et al. (2000a), char. 19; Murdock (2008b), char. 45; Rothwell et al. (2018), char. 88.

55. Sporangial aperture shape: (0) slit; (1) pore. Similar to Hill and Camus (1986a), char. 69; Liu et al. (2000a), char. 19; Murdock (2008b), char. 44; Rothwell et al. (2018), char. 87. Uninformative in the current matrix.

56. Sporangial tips: (0) absent; (1) present. Similar to Liu et al. (2000a), char. 23; Murdock (2008b), char. 49; Rothwell et al. (2018), char. 83.

57. Sporangial tip length: (0) short; (1) long. Similar to Rothwell et al. (2018), char. 84.

58. Size of sporangia: (0) <1 mm; (1) 1-10 mm. Similar to Rothwell et al. (2018), char. 80.

59. Sporangia dorsal wall thickness: (0) one to two cells; (1) three or more cells. Similar to Liu et al. (2000a), char. 22; Rothwell et al. (2018), char. 82.

60. Indument borne on sporangia: (0) absent; (1) present. Similar to Hill and Camus (1986a), char. 49; Murdock (2008b), char. 29; Rothwell et al. (2018), char. 43.
### Appendix 3

Phenotypic character data matrix. Polymorphisms are marked by letters as follows: a = [0 1], b = [1 2], c = [0 1 2]. Dashes indicate inapplicable states, question marks indicate unknown states.