Cladistics

The International Journal of the Willi Hennig Society

WILEY

wileyonlinelibrary.com/journal/cla

New insights into the phylogeny of the complex thalloid liverworts (Marchantiopsida) based on chloroplast genomes

You-Liang Xiang[a,†], Xin-Jie Jin[a,b,†], Chao Shen[a] (iD), Xia-Fang Cheng[a], Lei Shu*[a,c] (iD) and Rui-Liang Zhu*[a,c,d] (iD)

[a]*Bryology Laboratory, School of Life Sciences, East China Normal University, Shanghai, 200241, China;* [b]*College of Life and Environmental Sciences, Wenzhou University, Wenzhou, 325035, China;* [c]*Shanghai Institute of Eco-Chongming (SIEC), Shanghai, 200062, China;* [d]*Tiantong National Station of Forest Ecosystem, Shanghai Key Laboratory for Urban Ecological Processes and Eco-Restoration, East China Normal University, Shanghai, 200241, China*

Abstract

Marchantiopsida (complex thalloid liverworts) are one of the earliest lineages of embryophytes (land plants), and well-known for their air pores and chambers, pegged rhizoids, and absence of organellular RNA editing sites. Despite their importance to an understanding of early embryophyte evolution, many key nodes within this class remain poorly resolved, owing to the paucity of genetic loci previously available for phylogenetic analyses. Here, we sequenced 54 plastomes, representing 28 genera, nearly all families, and all orders of Marchantiopsida. Based on these plastomes, we present a hypothesis of deep relationships within the class, and make the first investigations of gene contents and synteny. Overall, the Marchantiopsida plastomes were well-conserved, with the exception of the genus *Cyathodium* that has plastomes with higher GC content, fewer single sequence repeats (SSRs), and more structural variations, implying that this genus might possess RNA editing sites. Abundant repetitive elements and six highly divergent regions were identified as suitable for future infrafamilial taxonomic studies. The phylogenetic topology of Sphaerocarpales, Neohodgsoniales and Blasiales within Marchantiopsida was essentially congruent with previous studies but generally we obtained higher support values. Based on molecular evidence and previous morphological studies, we include Lunulariales in Marchantiales and suggest the retention of narrowed delimitation of monotypic families. The phylogenetic relationships within Marchantiales were better resolved, and 13 monophyletic families were recovered. Our analyses confirmed that the loss of intron 2 of *ycf3* is a synapomorphy of Marchantiidae. Finally, we propose a new genus, *Asterellopsis* (Aytoniaceae), and present an updated classification of Marchantiopsida. The highly supported phylogenetic backbone provided here establishes a framework for future comparative and evolutionary studies of the complex thalloid liverworts.
© 2022 The Willi Hennig Society..

Key words: classification, comparative genomics, cp genomes, Marchantiophyta, phylogenomics

Introduction

The liverworts (Marchantiophyta) are the second-largest division of bryophytes comprising *c.* 7300 species (Söderström et al., 2016). Marchantiopsida, usually known as the complex thalloid liverworts, are the second-largest class of Marchantiophyta, with 330–340 species (Long et al., 2016; Söderström et al., 2016; Villarreal et al., 2016). Marchantiopsida are recognized by several striking characters, including a stalked carpocephalum, pored air chambers, pegged rhizoids (Crandall-Stotler et al., 2009; Duckett et al., 2014) and a lack of RNA editing sites in organelles (Dong et al., 2019; Rüdinger et al., 2012; Yu et al., 2019).

*Corresponding author:
E-mail address:*rlzhu@bio.ecnu.edu.cn; lshu@bio.ecnu.edu.cn
[†]These authors contributed equally to this work.

These interesting characters identify Marchantiopsida as a terrestrial plant group of evolutionary and phylogenetic importance (Bowman et al., 2017).

Although Marchantiopsida comprise only about 34 genera in 19 families (Akiyama and Odrzykoski, 2020; Long et al., 2016; Söderström et al., 2016; Villarreal et al., 2016), its classification has been historically difficult and unstable, possibly as a consequence of the reduction and re-evolution of morphological characters in this class (Villarreal et al., 2016), and the classification of Marchantiopsida relies mainly on the sporophyte characters, which are not easy to obtain (Crandall-Stotler et al., 2009; Schuster, 1992). Indeed, several critical revisions have been proposed based on molecular analyses of a single or a few genetic markers (Flores et al., 2018, 2021; Long et al., 2016; Söderström et al., 2016; Villarreal et al., 2016). The most important changes to the accepted classification dictated by these molecular analyses include: (1) inclusion of Blasiales with simple thalli in Marchantiopsida, transfer of *Neohodgsonia* Perss. from the Marchantiaceae to its own family and order, recognition of a monogeneric Lunulariales, incorporation of Monocleales and Ricciales into Marchantiales, transfer of *Peltolepis* Lindb. from Monosoleniaceae to Cleveaceae, and recognition of a monogeneric Dumortieraceae (Long, 2006a); (2) transfer of *Monocarpus* D.J. Carr from Marchantiales to Sphaerocarpales (Forrest et al., 2015); and (3) treatment of *Bucegia* Radian and *Preissia* Corda as synonyms of *Marchantia* L. and transfer of Exormothecaceae to Corsiniaceae (Long et al., 2016). In recent comprehensive analyses of the Marchantiopsida based on 11 genes, 128 morphological characters and four Marchantiopsida fossils, Flores et al. (2018, 2021) proposed the reduction of Lunulariales and introduction of several broadly defined families, including Corsiniaceae, Monocleaceae, Ricciaceae and Targioniaceae. Although molecular analyses based on genetic markers have successfully delimited subclasses and orders within Marchantiopsida (Flores et al., 2018, 2021; Villarreal et al., 2016), most deep nodes within Marchantiales remain ambiguous or, at best, have low support values. Additionally, the relationships among some key lineages have yet to be resolved, specifically in Marchantiales. For instance, the genus *Aitchisoniella* Kashyap was placed in the Cleveaceae by Villarreal et al. (2016) based on an analysis of 11 gene fragments, but in Corsiniaceae by Flores et al. (2018, 2021) using the same 11 gene fragments in conjunction with 128 morphological characters and four fossils. In Aytoniaceae, *Asterella* P. Beauv. was monophyletic based on morphological characters (Long, 2006b), but polyphyletic based on genetic data (Borovichev et al., 2015; Long et al., 2000). Furthermore, the phylogenetic hypotheses presented in all previous studies showed low nodal support values and short internal branches for most early splits within Marchantiopsida. The deep relationships of this group seem to be very difficult to resolve using a few genetic loci even when supplemented with morphological data. Therefore, we performed analyses with more DNA markers in order to better resolve lineages of Marchantiopsida.

Complete chloroplast genome sequences possess much valuable information for the phylogenetic analyses of all plant lineages. In liverworts, chloroplast genomes have helped to clarify the backbone phylogeny (Dong et al., 2021; Yu et al., 2019). Several plastomes have been obtained for complex thalloid liverworts (Dong et al., 2021; Sawicki et al., 2020; Yu et al., 2019). However, owing to limited taxon sampling, many relationships within Marchantiopsida remain untested. Compared with the vascular plants, phylogenetic studies of liverworts based on complete chloroplast genomes are few. Here, we sequenced 54 new complete chloroplast genomes, representing all orders and the main lineages of the Marchantiopsida. The aims of our study were: (1) to perform phylogenetic analysis of Marchantiopsida based on plastomes and infer intergeneric relationships; (2) to investigate the structure and content of the plastid genomes in Marchantiopsida; and (3) to explore phylogenetic implications and provide a basis for future infrafamilial taxonomic studies of the complex thalloid liverworts.

Materials and methods

Taxon sampling, sample preparation, and sequencing

The chloroplast genomes of 77 samples of Marchantiopsida were included in this study, representing all five orders, 90% of all families and 82% of all genera. The chloroplast genomes of 54 species were newly sequenced for this study, and 23 more genomes were downloaded from GenBank (Table S1). Two species in class Haplomitriopsida, *Apotreubia nana* (S. Hatt. & Inoue) S. Hatt. & Mizut. and *Haplomitrium mnioides* (Lindb.) R.M. Schust., were selected as outgroup terminals following previous studies (Flores et al., 2021; Villarreal et al., 2016); The chloroplast genomes for these species also were newly sequenced in this study (Table S1).

All plant samples sequenced in this study were obtained from the Herbarium of East China Normal University, China (HSNU). Before DNA extraction, specimens were thoroughly cleaned to remove the algae, fungi and soil debris. Genomic DNA was extracted from *c.* 20 mg of silica-dried gametophytic tissue using DNA Plantzol Reagent (Hangzhou Lifefeng Biotechnology Co., Ltd, Hangzhou, China). Genomic DNA samples were sheared into fragments (< 800 bp) using Illumina Nextera XT DNA library preparation, and the quality of the fragments was checked using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). Short-insert (350-bp) paired-end libraries were prepared and sequenced by the Beijing Genomics Institute (Shenzhen, China). Indexed samples were pooled and run in a single lane of an Illumina HiSeq 2500 with a read length of 150 bp.

Chloroplast genome assembly and annotation

For the newly sequenced liverwort species, we used the GetOrganelle pipeline (Jin et al., 2020) for the *de novo* assembly of the complete circular chloroplast genome. The genomes were annotated automatically using CPGAVAS2 (Shi et al., 2019) and then adjusted using Geneious v.11.0.3 (Kearse et al., 2012) based on the *Pellia endiviifolia* (Dicks.) Dumort. reference plastome. Circular plastome maps were drawn using OrganellarGenome DRAW (Lohse et al., 2013). Previously published genomes were re-annotated using the same method.

In order to identify the unique characters of the Marchantiopsida plastomes, we used Geneious v.11.0.3 (Kearse et al., 2012) to compare GC content, genome size, and noncoding content among the 77 plastomes of Marchantiopsida, two plastomes of Haplomitriopsida and 54 plastomes of Jungermanniopsida that were downloaded from GenBank (Table S1).

Identification of molecular markers

Simple sequence repeats (SSRs) across the Marchantiopsida and outgroup plastomes were detected using the MISA-web application (Beier et al., 2017). Thresholds for the minimum number of repeat units for each type of SSR were set as follows: ten for mononucleotides, eight for di-nucleotides, and five each for tri-, tetra-, penta- and hexa-nucleotides. To explore sequence divergence and hotspot regions in Marchantiopsida plastomes, nucleotide variability values (Pi) were calculated using a sliding window analysis in DnaSP v.6.11.01 (Rozas et al., 2017). In this analysis, the step size was set to 200 bp, with a 600-bp window.

Phylogenetic analyses

We investigated phylogenetic relationships among the 77 Marchantiopsida samples, with the two Haplomitriopsida species as outgroup terminals (Table S1), based on chloroplast genomes. We initially performed phylogenetic analyses using three different datasets of 79 sequences each: the entire chloroplast genome minus the second inverted repeat region (IRa), to avoid considering the same information twice; concatenated noncoding sequences, including intergenic spacers and introns; and concatenated amino acid sequences from the exons of protein-coding genes. MAFFT v.7.149b (Katoh and Standley, 2013) was used to align sequences: nucleotide and amino acid sequences were aligned in batches with "–auto" strategy, and ambiguously aligned regions were trimmed manually. The final alignments of the three datasets were deposited in the Figshare Digital Repository (https://figshare.com/articles/dataset/Marchantiopsida_data/19203359). We then performed maximum-likelihood (ML) analyses using a concatenation approach for the three datasets separately by IQ-Tree v.2.0.6 (Nguyen et al., 2014), using the defaulted settings with the implementation of hill-climbing and stochastic nearest neighbour interchange (NNI) operations for tree search, 1000 ultra-fast bootstrap replicates (ML-BS), and the best-fit model selected by ModelFinder (Kalyaanamoorthy et al., 2017). We compared topologies among the three ML trees and found no conflict among the highly supported branches of the three trees except for *Monosolenium tenerum* Griff. (Figs S1–S2). The dataset of concatenated amino acid sequences not only avoids synonymous substitutions in nucleotide datasets but also possesses the minimum amount of data, which could save computation. Therefore, we reconstructed trees based on this dataset using both Bayesian inference (BI) and maximum parsimony (MP). BI was performed using MrBayes v.3.2.7 (Ronquist et al., 2012) on the CIPRES Science Gateway website (Miller et al., 2010). The best-fit nucleotide substitution model (Cprev) was selected based on three measures [the Akaike Information Criterion (AIC), the corrected Akaike Information Criterion (AICc), and the Bayesian Information Criterion (BIC)] in PartitionFinder2 (Lanfear et al., 2017). The Markov chain Monte Carlo (MCMC) algorithm was run for 5 000 000 generations, with two parallel runs and default setting of runs, a random starting tree and tree sampling every 1000 generations. We assumed that the MCMC chains had converged when the average standard deviation of the split frequencies (ASDFs) was ≤0.01. Bayesian posterior probabilities (BPPs) were calculated for the maximum clade credibility tree of all samples after discarding the first 20% of the trees as burn-in. MP analysis was performed using PAUP* v.4.0b10 (Swofford, 2002), with 1000 replicates of random stepwise additions and tree bisection-reconnection (TBR) branch swapping for heuristic searches. We measured support for the reconstructed clades using 1000 bootstrap replicates (MP-BS), with the starting tree generated by stepwise addition and TBR branch swapping.

In order to provide the best estimate of the phylogeny of Marchantiopsida, we also undertook phylogenetic analysis using a coalescent method which considered variation in the phylogenetic signal across plastid genes and had been verified to offer a high level of accuracy with plastid data for determining phylogeny (Gonçalves et al., 2019). We first inferred individual ML gene trees (based on the 79 amino acid sequence alignment) using RAxML v.8.2.10 (Stamatakis, 2014) with a PROTGAMMA model and 1000 replicates to assess clade support. We then submitted these individual-gene trees to ASTRAL-III v.5.6.3 (Zhang et al., 2018) to infer the coalescent species tree. Node support was measured based on local posterior probabilities (LPP; Sayyari and Mirarab, 2016).

Results

Chloroplast genome features

The plastomes of 77 Marchantiopsida samples (including the 54 that were sequenced for this study) ranged from 116 912 bp (*Cyathodium smaragdinum* Schiffn.) to 123 481 bp (*Exormotheca bischlerae* Furuki & Higuchi), with an average size of 121 746 bp and a 1.14-fold chloroplast genome size range (Fig. 1). As in the plastomes of most embryophytes, the plastomes of Marchantiopsida exhibit a typical quadripartite structure, consisting of a pair of inverted repeated regions (IRs: 9165 – 11 420 bp) separated by a large single-copy region (LSC: 78 716 – 83 380 bp) and a small single-copy region (SSC: 19 479 – 20 254 bp). All Marchantiopsida plastomes possess 126–136 unique genes (82–90 protein-coding genes, 36–38 tRNA genes and eight rRNA genes) arranged in the same order. Of these, 12 protein-coding genes and six tRNAs contained at least one intron (Fig. S3). Across all liverwort chloroplast genomes examined, the size of the LSC correlated directly with overall genome size, being largest in *Haplomitrium mnioides* (Lindb.) R.M. Schust. and smallest in *Aneura mirabilis* (Malmb.) Wickett & Goffinet, and the size of the SSC and IR generally remained similar across all genomes, irrespective of total genome size. Of three species, *Conocephalum salebrosum* Szweyk., Buczk. & Odrzyk., *Dumortiera hirsuta* (Sw.) Nees and *Haplomitrium*

mnioides, the IR regions expanded and were unusually large. This expansion was driven by inclusion of the genes *rps7* (*Conocephalum salebrosum*), *ndhB* and *rps12* (*Dumortiera hirsuta*), and *ndhF* (*Haplomitrium mnioides*) (Fig. S4). In addition, the SSC was unusually small in *Apotreubia nana*, *Aneura mirabilis* (Malmb.) Wickett & Goffinet, and *Cololejeunea lanciloba* Steph., owing to gene losses: eight genes belonging to the *ndh* gene family (*ndhA*, *ndhB*, *ndhD*, *ndhE*, *ndhF*, *ndhG*, *ndhH*, *ndhI*) in *Apotreubia nana*, seven genes (*ndhA*, *ndhG*, *ndhH*, *ndhI*, *ndhK*, *psaA*, *ycf3*) in *Aneura mirabilis*, and six genes (*ccsA*, *cysA*, *cyst*, *ndhF*, *rpl21*, *rpl32*) in *Cololejeunea lanciloba* (Fig. S4).

Compared with Jungermanniopsida, GC content and plastome size seem to be more conserved in Marchantiopsida aside from the genus *Cyathodium*, which has a smaller plastome size and a higher GC content than other genera of Marchantiopsida (Fig. 1). In addition, the loss of intron 2 of *ycf3*, which was reported previously (Dong et al., 2021), is a synapomorphy of Marchantiidae. In Haplomitriopsida, and most species of Jungermanniopsida, *cysA-T* is lost or pseudogenized, but this gene has been retained and is functional in all species of Marchantiopsida. Despite the rather static genomic structures across Marchantiopsida, we identified one intron loss (*rps12*) in *Monoclea forsteri* Hook. and loss of six genes (*petG*, *petL*, *psbE*, *psbF*, *psbJ*, *psbL*) in *Neohodgsonia mirabilis* (Perss.) Perss. (Fig. S4). Also notable is *Asterella grollei* D.G. Long, which has a degraded *ycf2* with only two short reading frames.

Molecular markers

We detected numerous SSR microsatellites in the plastomes of 77 Marchantiopsida samples and two outgroup species (Fig. S5). Most SSRs were located in LSC and SSC regions, with a few in IR regions. Mono-nucleotide repeats were the most abundant, followed by di-nucleotide repeats. Fewer total SSRs and SSR types were identified in the outgroup plastomes. In total, eight types of SSRs were found across the Marchantiopsida plastomes. Although five SSR types were identified in most plastomes, the plastomes of two species (*Riccia junghuhniana* Nees & Lindenb. and *Sauteria* sp. 1) possessed seven SSR types, the maximum observed in this study (Fig. S5). The total number of SSRs in the Marchantiopsida plastomes ranged
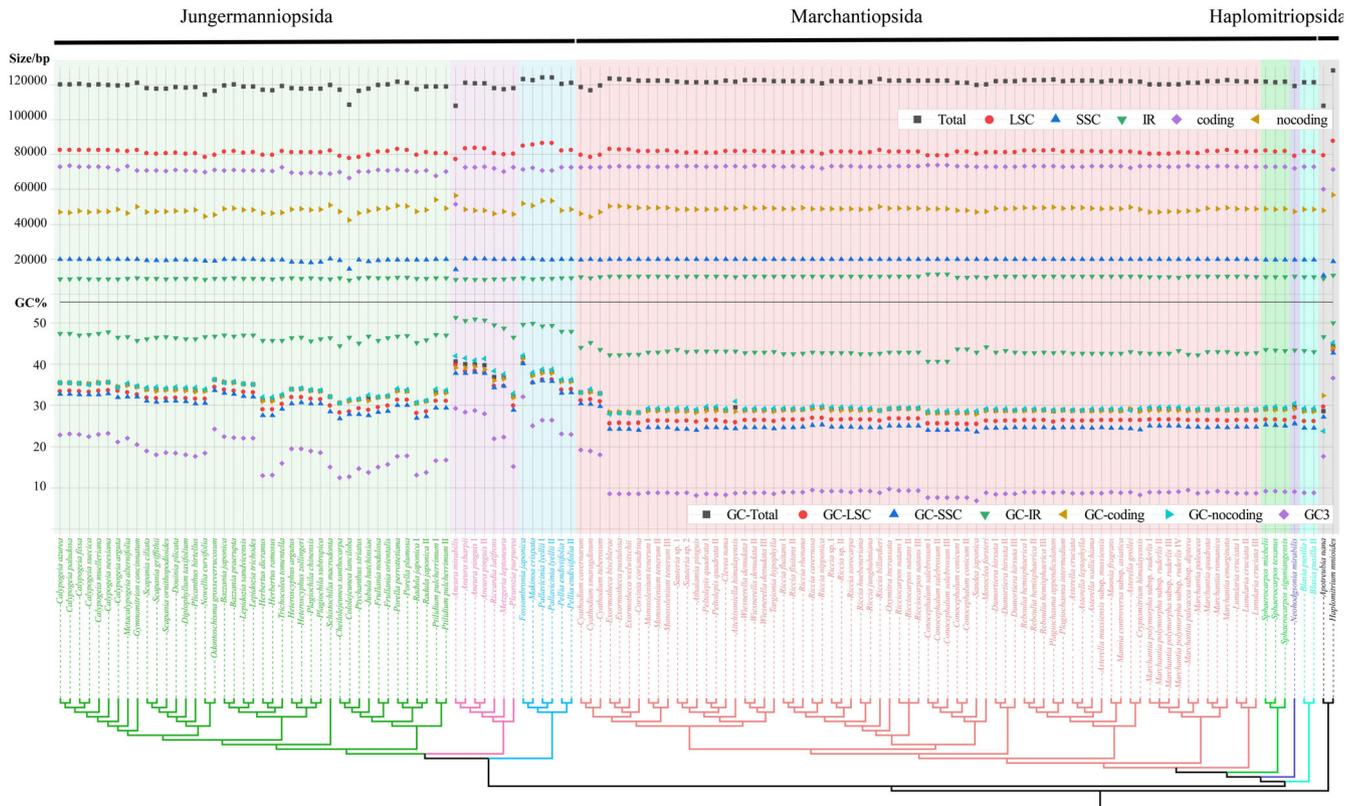

Fig. 1. Comparison of size and GC content among chloroplast genomes of three classes of liverworts (Haplomitriopsida, Jungermanniopsida and Marchantiopsida), based on the whole genome, the LSC region, the SSC region and the IR regions. Taxa along the *x*-axis are sorted and coloured based on the ML phylogeny. The following orders are represented: Blasiales (light blue), Neohodgsoniales (plum), Sphaerocarpales (aquamarine) and Marchantiales (lavender).

from 59 (*Cyathodium smaragdinum*) to 208 (*Exormotheca welwitschii* Steph.), and SSR type counts varied among samples, even within a single species (see for example *Dumortiera hirsuta, Ricciocarpos natans* (L.) Corda and *Reboulia hemisphaerica* (L.) Raddi; Fig. S5).

Plastomes of Marchantiopsida include the following 141 regions: two rRNA genes, five tRNA genes, 64 intergenic spacers and 70 coding genes (Fig. S6). The variation range of nucleotide variability across sites was widest in the SSC regions (Pi = 0.04022–0.22469, mean 0.08831), followed by the LSC regions (Pi = 0.06989–0.20039, mean 0.08705), and the IR regions (Pi = 0.00456–0.08130, mean 0.04652) (Fig. S6). Regions with the greatest nucleotide variability were intergenic: *trnX6-trnL* (Pi = 0.22469) and *ycf2-trnD* (Pi = 0.20039). Another 13 intergenic regions (*ndhB-trnL, ndhH-rps15, petA-psbJ, psaJ-rpl33, rbcL-trnR, rpl14-rpl16, rpl23-trnM, rps7-ndhB, trnD-trnY, trnF-ndhJ, trnT-trnL, trnX2-psbK, trnX1-ycf12*) had relatively high nucleotide diversity (Pi > 0.15000). The most variable protein-coding regions were the genes related to the *ycf* family: *ycf2* (Pi = 0.15105), *ycf1-1* (Pi = 0.09956) and *ycf1-2* (Pi = 0.10694). Other highly variable protein-coding regions (Pi > 0.09000) were *rpoC2, ndhF, ndhB* and *ndhG*. The genetic variability of the tRNA and rRNA sites was somewhat lower, with the greatest variability observed in *matK* (Pi = 0.09891), followed by *23rrn* (Pi = 0.00945).

Phylogenomic analyses

The concatenated amino acid matrix consisted of 25 624 sites, among which 13 527 were constant and 8950 were parsimony-informative. The topologies of the MP, ML and BI trees were largely congruent with each other (Fig. 2), as well as with the ASTRAL coalescent analysis (Fig. S7). The five previously identified orders of Marchantiopsida (Blasiales, Lunulariales, Marchantiales, Neohodgsoniales and Sphaerocarpales) all were found as monophyletic groups with high support values across all analyses (MP-BS = 100, ML-BS = 100, BPP = 1; Fig. 2). The order Blasiales are sister to the rest of Marchantiopsida, followed by Neohodgsoniales, Sphaerocarpales, Lunulariales and Marchantiales.

Marchantiales, the largest order in the Marchantiopsida, was divided into five major clades (clades A–E; Fig. 2). Clade A (MP-BS = 60, ML-BS = 100, BPP = 1) is composed of six families: Cleveaceae, Corsiniaceae, Cyathodiaceae, Monosoleniaceae, Targioniaceae and Wiesnerellaceae. Two sister relationships were recovered in all analyses with maximum support (Corsiniaceae and Cyathodiaceae, Targioniaceae and Wiesnerellaceae). Cleveaceae was resolved

to be sister to the subclade of Corsiniaceae and Cyathodiaceae with high support values in ML and BI estimation but not supported in MP estimation (MP-BS = –, ML-BS = 83, BPP = 1). The position of Monosoleniaceae was dubious, referring to conflict topologies among the analyses of three datasets: (((Targioniaceae, Wiesnerellaceae), Monosoleniaceae), ((Corsiniaceae and Cyathodiaceae), Cleveaceae)) in the amino acid dataset, ((Targioniaceae, Wiesnerellaceae), (((Corsiniaceae and Cyathodiaceae), Cleveaceae), Monosoleniaceae)) in the concatenated noncoding dataset, and (Monosoleniaceae, ((Targioniaceae, Wiesnerellaceae), ((Corsiniaceae and Cyathodiaceae), Cleveaceae))) in the entire chloroplast genome dataset. Clade B was well-supported as sister to clade A (MP-BS = 93, ML-BS = 100, BPP = 1), and included four families: Conocephalaceae, Monocleaceae, Oxymitraceae and Ricciaceae. The latter was not found to be monophyletic because Oxymitraceae was nested within this family, as a sister to *Riccia* L. (MP-BS = 100, ML-BS = 100, BPP = 1; Fig. 2). Within the Conocephalaceae, *Sandea* Lindb. was recovered sister to *Conocephalum* Hill with high support values (MP-BS = 100, ML-BS = 100, BPP = 1; Fig. 2). Clades C, D and E all had maximum support values (MP-BS =100, ML-BS = 100, BPP = 1) and contained only one family each (Aytoniaceae, Dumortieraceae and Marchantiaceae, respectively; Fig. 2). The family Aytoniaceae includes five genera: *Asterella, Cryptomitrium* Austin ex Underw., *Mannia* Corda*, Plagiochasma* Lehm. and *Reboulia* Raddi. However, *Asterella* was polyphyletic because *A. grollei* was recovered in a sister relationship with *Cryptomitrium himalayense* Kashyap. (MP-BS = 100, ML-BS = 100, BPP = 1; Fig. 2).

Discussion

Plastome variation in the class Marchantiopsida

The sizes of the studied plastomes (77 of Marchantiopsida plastomes with 54 newly sequenced and 23 previously published, two of Haplomitriopsida plastomes, and 54 of Jungermanniopsida plastomes) seem to show quite a small variation around 120 000 bp, in contrast to the hypothesis that plastome size would tend to decrease in liverworts over evolutionary time (Dong et al., 2021). Although all plastomes generally are more conserved in the IR and SSC regions than in the LSC region, plastome structure and synteny are more variable in Haplomitriopsida and Jungermanniopsida as compared to Marchantiopsida. Within Marchantiopsida, the lengths of various sequence regions and sequence types (IRs, LSC, SSC, coding sequences and noncoding sequences) varied little.
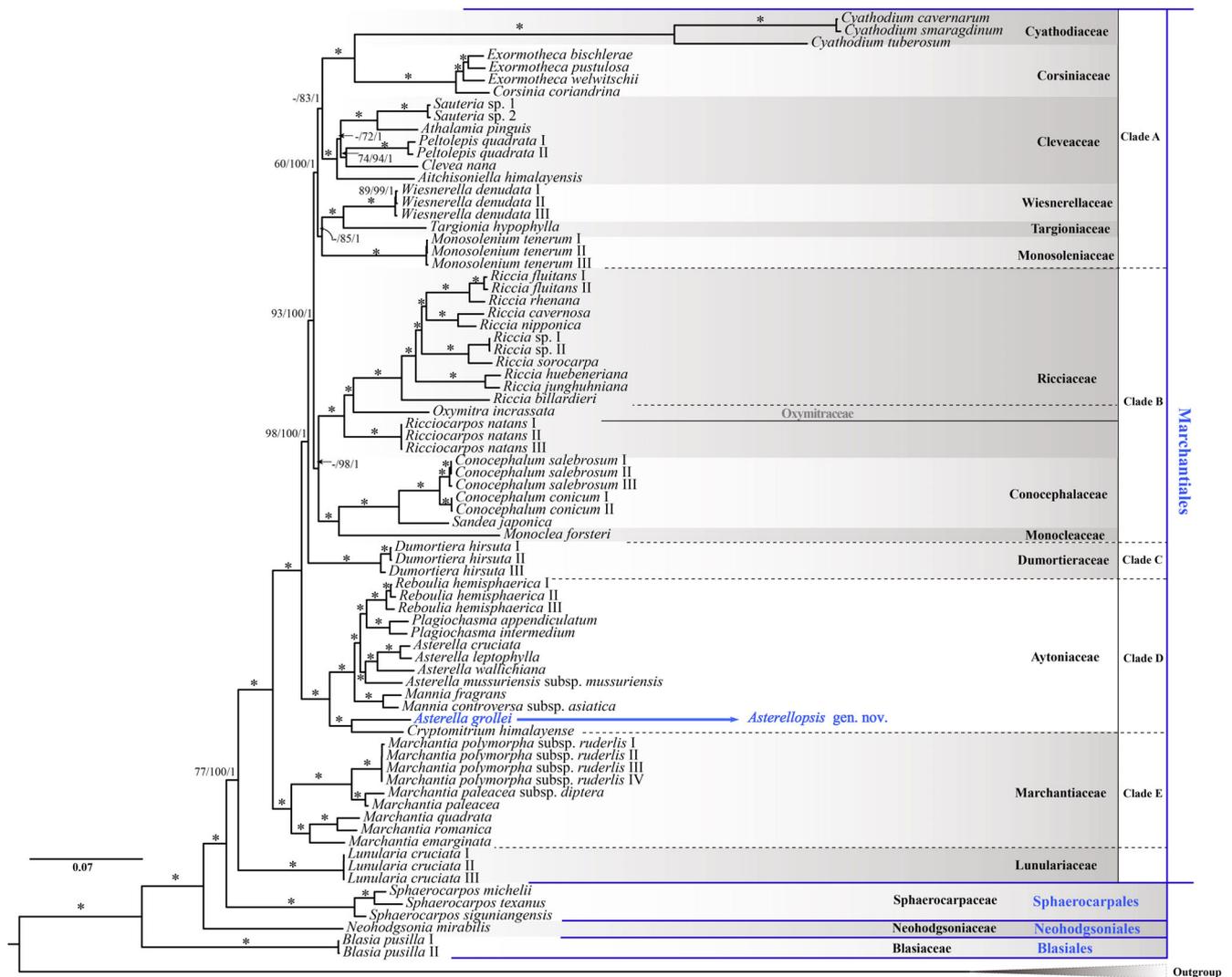
Fig. 2. Bayesian inference tree of Marchantiopsida, inferred from the concatenated amino acid sequences. Numbers at nodes correspond to MP-BS values, ML-BS values and BPP values. Asterisks are shown to indicate MP-BS = 100, ML-BS = 100 and BPP = 1. A dash "–" indicates that this relationship was not supported in the corresponding analysis. Clades A–E are the five major clades in Marchantiidae.

Likewise, GC contents across the Marchantiopsida plastomes range only from 28.1% to 33.4%, a distinctly smaller range than that recovered among other species of liverworts (Dong et al., 2021; Yu et al., 2019), indicating substantial conservation. GC content in the noncoding intergenic regions of Marchantiopsida plastomes is much lower than that in the coding regions, as has been observed in most embryophytes (Bock, 2007). The high plastome conservation across Marchantiopsida coincides with the low rate of molecular evolution in this class (Villarreal et al., 2016).

The conserved GC content of Marchantiopsida is consistent with the hypothesis that RNA editing does not occur in the plastid genomes of this class because the number of RNA editing sites in the liverwort plastomes is positively correlated with the GC contents (Dong et al., 2021; Rüdinger et al., 2008). The one exception is the genus *Cyathodium*: the plastome sequences of species in this genus, specifically at the third codon position, have a much higher GC content than that in other genera of Marchantiopsida (Fig. 1). Examination of the annotation of the previously published plastome of *Cyathodium smaragdinum* (GenBank no. MW429509; Frangedakis et al., 2021) showed that *rbcL* had two RNA editing sites, whereas *ndhA*, *ndhI*, *psbB*, *rpl12*, *rps18* and *ycf4* had one RNA editing site each. We identified identical nucleotide bases in the RNA editing positions in two of the *Cyathodium* plastomes sequenced herein. Thus, RNA editing of organellar genomes might have been absent in Marchantiopsida except for *Cyathodium*, which should

be taken into consideration in future studies involving dense taxon sampling and transcriptome data of this genus. Additionally, repetitive DNA can induce rearrangements and promote genomic instability (Blazier et al., 2016; Sinn et al., 2018; Wu et al., 2021), whereas the abundance of repetitive sequences of plastomes in *Cyathodium* is slightly lower than in other genera of Marchantiopsida. Specifically, although intron II of the chloroplast gene *ycf3*, which has two introns, is missing from all plastomes of Marchantiidae, intron I also is absent in *C. cavernarum* Kunze and *C. smaragdinum*. The *ycf3* gene is essential for photosystem I assembly (Boudreau et al., 1997; Wang et al., 2020), and the loss of even one intron affects *ycf3* gene expression (Petersen et al., 2011). Thus, based on our phylogenetic hypothesis, intron I might have been lost recently, whereas intron II was more likely lost in the common ancestor of Marchantiidae. In combination, its distinctive genomic structure, ephemeral life cycle (Allen and Korpelainen, 2006) and unusually rapid rate of organellar-marker evolution (Villarreal et al., 2016) indicate that *Cyathodium* stands out from the rest of the complex thalloid liverworts and might have experienced a distinct evolution pattern, which should be carefully explored in future studies on character evolution and biogeography with greater taxon sampling.

The other plastome variations that we observed, such as IR expansion, gene loss and degradation, seem to have occurred independently during the evolutionary history of Marchantiopsida, and thus may not provide relevant phylogenetic information. The plastomes of Marchantiopsida included more SSRs than the plastomes of other species of liverworts (Sawicki et al., 2020). Owing to their high rates of polymorphism, SSRs are an important source of molecular markers and have been studied extensively in mosses (Kamisugi et al., 2008; Sawicki et al., 2012; Beike et al., 2014), but rarely in liverworts. Across Marchantiopsida, we identified six coding regions (*matK*, *ndhB*, *ndhF*, *ndhG*, *rpoC2*, *ycf2*) and 15 intergeneric regions with high nucleotide diversity. Most of the identified intergeneric regions are too short to be recommended as barcodes, but these six highly variable coding regions may be suitable for future infrafamilial phylogenetic analysis. Indeed, the abovementioned six coding regions previously have been identified as highly variable regions in many plant species (Fan et al., 2018; Machado et al., 2020; Wu et al., 2018; Ye et al., 2018; Zhang et al., 2020), as well as in the liverwort genus *Calypogeia* Arnell (Ślipiko et al., 2020).

Backbone phylogenetics of Marchantiopsida

An appropriate combination of data is one of the most important determinants of accurate phylogenetic estimation (Heath et al., 2008). Previous studies have used nuclear ribosomal subunit 26S, organellular genes and morphological characters to infer the phylogeny of complex thalloid liverworts (Flores et al., 2018; Flores et al., 2021; Forrest et al., 2006; Villarreal et al., 2016). However, phylogenetic relationships among and within the complex thalloid liverwort families are challenging to resolve and have been revised frequently based on different datasets (e.g. Flores et al., 2018; Flores et al., 2021; Forrest et al., 2006; Villarreal et al., 2016). This persistent topological incongruence between different phylogenetic hypotheses has hindered studies of adaptive evolution. Here, our phylogenetic analyses based on complete plastid genomes, noncoding sequences and concatenated amino acids provided strong support for the monophyly of three groups Blasiales, Neohodgsoniales and Sphaerocarpales, in agreement with previous molecular and morphological evidence (Flores et al., 2018; Flores et al., 2021; Villarreal et al., 2016), and improved the support values for phylogenetic relationships within the Marchantiales.

It was proposed that the family Lunulariaceae, composed of a single monospecific genus *Lunularia* Adans., should be recognized as an independent order based on two morphological characters: the crescent-shaped gemma receptacles and the elongate massive seta exserted from a tubular involucre (Crandall-Stotler et al., 2009; Long, 2006a). Previous phylogeny studies recovered a sister relationship between Lunulariales and Marchantiales (Flores et al., 2018, 2021; Villarreal et al., 2016), which also is supported by our result. Marchantiales previously defined were usually characterized by the differentiated thallus, dimorphic rhizoids, and epidermis with either simple or compound air pores (Crandall-Stotler and Stotler, 2000). With the inclusion of Monocleales and Ricciales in Machantiales (Long, 2006a) and the exclusion of *Monocarpus* from Machantiales (Forrest et al., 2015; Villarreal et al., 2016), Flores et al. (2018) pointed out that previously defined Marchantiales lack diagnostic morphological characters and suggested that Lunulariales be included in Marchantiales because the two groups share 15 morphological synapomorphies, such as hexagonal/pentagonal scale cells, lanceolate shaped scales and thin pegged rhizoids. In addition, a further three shared characters are observed in the two groups: (1) capsule dehiscence by a lid, (2) perigonial chambers aggregated in terminal cushions on the thallus and (3) massive seta, elongating before spore dispersal. Therefore, the present study accepts Flores et al.'s (2018) proposal for inclusion of Lunulariales in Marchantiales.

Previous studies have agreed as to the phylogenetic placement of the family Marchantiaceae (Boisselier-Dubayle et al., 2002; Dong et al., 2021; Flores

et al., 2018, 2021; Villarreal et al., 2016; Yu et al., 2020). However, owing to fragmentary sampling in the plastome analyses (Dong et al., 2021; Yu et al., 2020) and incongruent topologies between multiple-locus and morphological data (Boisselier-Dubayle et al., 2002; Villarreal et al., 2016; Flores et al., 2018, 2021), the phylogenic positions of other families in Marchantiales had remained ambiguous. Here, based on sampling that covered 96% of the genera in Marchantiales, our phylogeny recovered Marchantiaceae as sister to the rest of Marchantiidae. Although Dumortieraceae previously were revealed as the second branching node within Marchantiidae with moderate to high support values (Villarreal et al., 2016; Flores et al., 2018, 2021), our result strongly supported Aytoniaceae to be the second one and Dumortieraceae the third. Conocephalaceae were recovered sister to Monocleaceae with high support value, and the Conocephalaceae + Monocleaceae clade was sister to Ricciaceae. Flores et al. (2021) suggested that the monogeneric Conocephalaceae should be synonymized with the Monocleaceae based on a shared synapomorphy (the loss of germ tubes). However, Conocephalaceae are markedly different from the Monocleaceae in having differentiated thalli with air chambers, air pores, ventral scales, and sporophytes on stalked receptacles (Schuster, 1992; Crandall-Stotler et al., 2009). Given the notable loss of the *rps12* intron in the Monocleaceae plastomes and the distinctive morphological characters of the Conocephalaceae, we propose the retention of both families, consistent with previous studies (Jovet-Ast, 2005; Forrest et al., 2006; Crandall-Stotler et al., 2009; Villarreal et al., 2016). Likewise, as the sporophytes of Wiesnerellaceae are on stalked receptacles whereas those of Targioniaceae on the ventral surface of the thallus apex, we object to the inclusion of Wiesnerellaceae in Targioniaceae, as proposed by Flores et al. (2021). In Corsiniaceae, the sporophytes are on the female receptacles or on dorsal surface of the thallus. In Cyathodiaceae, the sporophytes are on the ventral surface of the thallus apex. The ventral scale appendages are absent in Cyathodiaceae but present in Corsiniaceae. Finally, the plastome characters of Cyathodiaceae differ distinctly from those of Corsiniaceae. We thus propose the retention of both Cyathodiaceae and Corsiniaceae instead of reducing Cyathodiaceae to Corsiniaceae as in Flores et al. (2021). The monotypic Oxymitraceae were tentatively synonymized with Ricciaceae (Flores et al., 2021) based on morphological similarities and the sister relationship between families in multiple topologies (Villarreal et al., 2016; Flores et al., 2018, 2021). Our results were consistent with this suggestion: we recovered *Oxymitra* Bisch. ex Lindenb. within Ricciaceae and sister to *Ricca* L. with high support values. Several important synapomorphies for Oxymitraceae

and Ricciaceae have been reported in previous studies, such as perigonia embedded in receptacles, cleistocarpous capsules and the lack of elaters (Schuster, 1992; Flores et al., 2018). The phylogenetic position of *Aitchisoniella* has been controversial because of inconsistent topologies between molecular and total-evidence data (Long et al., 2016; Villarreal et al., 2016; Flores et al., 2018; Flores et al., 2021). Our analysis, which included one sample of *Aitchisoniella,* strongly supported the inclusion of *Aitchisoniella* in Cleveaceae, which was consistent with the molecular analysis of Villarreal et al. (2016). As there are only partial sequences of one gene (*atpB*) for *Aitchisoniella* in previous studies, the incongruent topology in Flores et al. (2018, 2021), which recovered *Aitchisoniella* as sister to *Stephensoniella* Kashyap and nested within Corsiniaceae, might result from the weighting strategy for missing data applied in MP analysis.

Discussion of the proposed new genus, Asterellopsis (Fig. 3)

Asterella initially was defined to include a large group of chambered liverworts with *c.* 80 species that were characterized by the presence of pseudoperianths, but the precise definition of this genus has been controversial since its establishment (Borovichev et al., 2015; Grolle, 1983; Long et al., 2000; Long, 2005; Long, 2006b). Schill et al. (2010) suggested that *Asterella* should be distinguished not only by pseudoperianths, but also by the lid of sporophyte (capsule) and spore ornamentations. Owing to continuous changes of identified characters for this genus, the number of species in *Asterella* has been reduced to 34 (Söderström et al., 2016). However, recent phylogenies had suggested that even this reduced genus may be polyphyletic (Long et al., 2000; Borovichev et al., 2015; Villarreal et al., 2016). Consistent with this, we recovered *A. grollei* D.G. Long in a well-supported sister relationship with *Cryptomitrium himalayense*. The genus *Cryptomitrium*, however, differs markedly from *A. grollei* in having bivalved involucres and flattened carpocephalum, and lacking pseudoperianths and germinal apertures on the spore surface (Long, 1999; Long, 2006b). Careful examination of morphological characters showed that *A. grollei* was distinguished from other species of *Asterella* by germinal apertures on the spore surface and an undivided spherical carpocephalum (Fig. 3). Based on this morphological and molecular evidence, we propose the erection of a new genus, *Asterellopsis* (Aytoniaceae), to accommodate this interesting species. *Asterella palmeri* (Austin) Underw. shares some morphological characteristics with *Asterellopsis grollei* and may be a member of *Asterellopsis,* yet it was not included in our
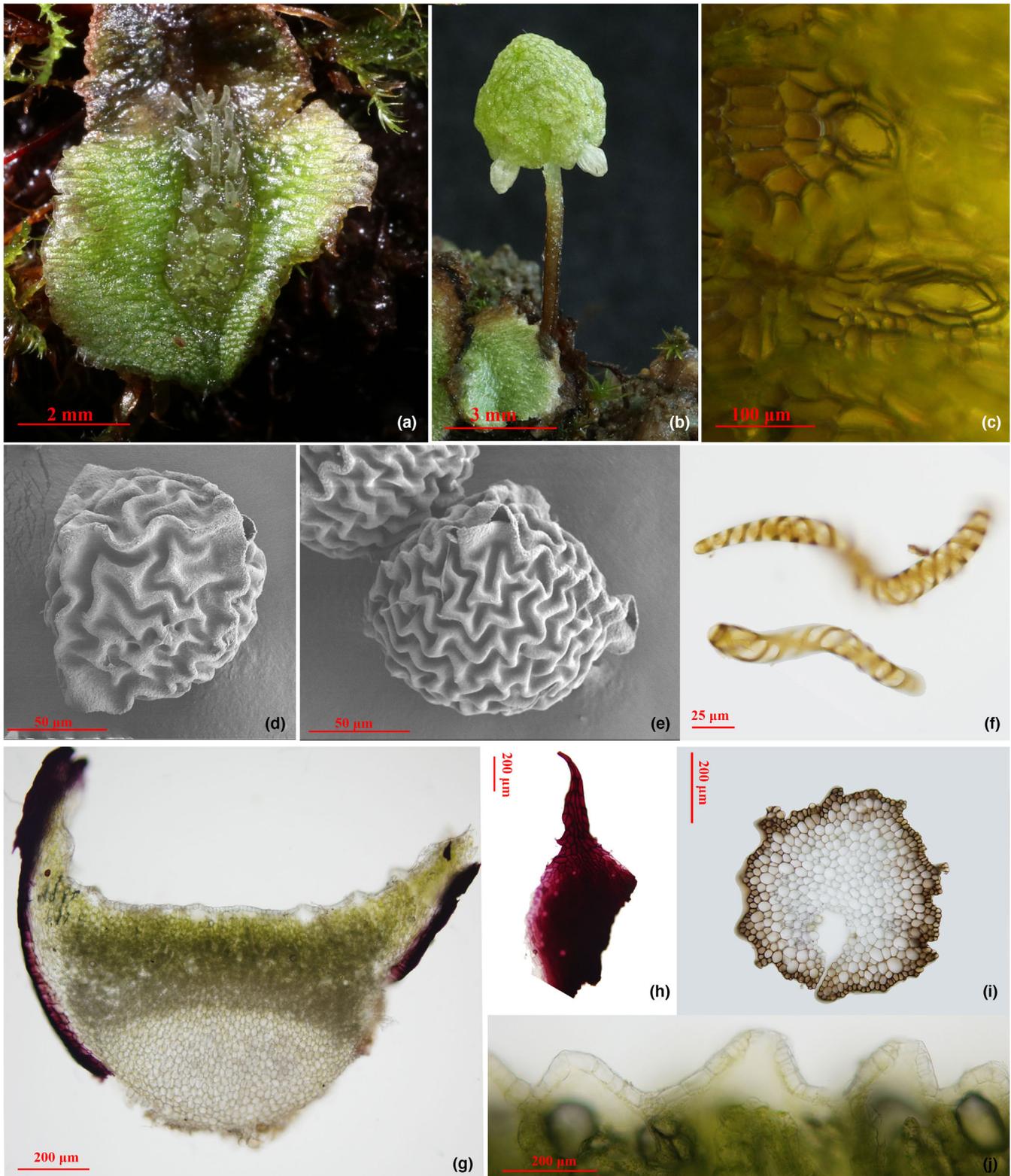
Fig. 3. *Asterellopsis grollei* (D.G. Long) R.L. Zhu & You L. Xiang: (a) thallus with antheridia; (b) thallus with female receptacle; (c) air pore on thallus; (d–e) spores with germinal apertures; (f) elaters; (g) transverse section of thallus; (h) ventral scale with appendage; (i) transverse section of female stalk; and (j) transverse section of air pores. All images of voucher specimen R.L. Zhu 20160815-18A (HSNU).

phylogenetic analysis owing to the lack of fresh material.

Taxonomic treatment

For nomenclatural purposes, we include the following formal treatment. Our revised classification of the Marchantiopsida (Table 1) was modified from Söderström et al. (2016) and Flores et al. (2021) based on the results reported herein. A total of 35 genera in 18 families belonging to four orders in two subclasses were included.

1. **Marchantiales** Limpr. in Cohn, Krypt.-Fl. Schlesien 1: 239. 336. 1877.

= Lunulariales D.G. Long, Edinburgh J. Bot. 63: 259. 2006.

Descriptions: Thallus usually differentiated, with simple or compound air pores (rarely absent); air-chambers absent or present; ventral scales present, sometimes absent, with or without appendages; rhizoids usually dimorphic, sometimes only smooth; idioblastic oil cells present or not; monoicous or dioicous; sporophyte on stalked receptacles, ventral at the thallus apex, along the thallus margin, in a dorsal depression of the thallus or embedded singly in the thallus; involucres bivalved, cup-shaped, scale- or flap-like, or tubular, sometimes absent; pseudoperianths absent or present; seta massive, short or absent; elaters present or absent; capsule dehiscence by longitudinal valves, longitudinal slit or lid, sometimes cleistocarpous; gemmae cups present or absent, crescent-shaped or cup-shaped; gemmae present in a few taxa.

2. **Ricciaceae** Rchb., Bot. Damen: 255. 1828.

= Oxymitraceae Müll. Frib. ex Grolle, J. Bryol. 7: 215. 1972.

Descriptions: Thallus dichotomously branched, differentiated, with air pores and air chambers, or with air pores but air chambers absent or vestigial to narrow gaps in some species; ventral scales present or

Table 1
The updated classification of Marchantiopsida.

Subclass	Order	Family	Genus
Blasiidae	Blasiales	Blasiaceae	*Blasia* L.
			Cavicularia Steph.
Marchantiidae	Marchantiales	Aytoniaceae	*Asterella* P. Beauv.
			Asterellopsis R.L. Zhu & You L. Xiang
			Cryptomitrium Austin ex Underw.
			Mannia Corda
			Plagiochasma Lehm.
			Reboulia Raddi
		Cleveaceae	*Aitchisoniella* Kashyap
			Athalamia Falc.
			Clevea Lindb.
			Peltolepis Lindb.
			Sauteria Nees
		Conocephalaceae	*Conocephalum* Hill
			Sandea Lindb.
		Corsiniaceae	*Corsinia* Raddi
			Cronisia Berk.
			Exormotheca Mitt.
		Cyathodiaceae	*Cyathodium* Kunze ex Lehm.
		Dumortieraceae	*Dumortiera* Nees
		Lunulariaceae	*Lunularia* Adans.
		Marchantiaceae	*Marchantia* L.
		Monocleaceae	*Monoclea* Hook.
		Monosoleniaceae	*Monosolenium* Griff.
		Ricciaceae	*Oxymitra* Bisch. ex Lindenb.
			Riccia L.
			Ricciocarpos Corda
		Targioniaceae	*Targionia* L.
		Wiesnerellaceae	*Wiesnerella* Schiffn.
	Neohodgsoniales	Neohodgsoniaceae	*Neohodgsonia* Perss.
	Sphaerocarpales	Monocarpaceae	*Monocarpus* D.J. Carr
		Riellaceae	*Austroriella* Cargill & J. Milne
			Riella Mont.
		Sphaerocarpaceae	*Geothallus* Campb.
			Sphaerocarpos Boehm.

absent. Antheridium embedded in the thallus, scattered. Sporophytes without seta and foot, elaters absent; capsules cleistocarpous.

With the synonymization of the Oxymitraceae, the cosmopolitan Ricciaceae is now represented by three genera (*Oxymitra*, *Riccia* and *Ricciocarpos*) (Table 1).

3. **Asterellopsis** R.L. Zhu & YouL. Xiang, **gen.n.** (Fig. 3)

Type species: *Asterellopsis grollei* (D.G. Long) R.L. Zhu & YouL. Xiang (≡ *Asterella grollei* D.G. Long, Bryologist 102(2): 169. 1999).

Descriptions: Thallus dichotomously branched, differentiated, with air pores and air-chambers, air pores simple; ventral scales in two rows; gemma cups absent; rhizoids dimorphic; female receptacles spherical, not dehiscent; the stalk of female receptacles with one rhizoid furrow; involucres flap-like, with pseudoperianths, spores with germinal apertures.

Etymology: The generic name is formed by the combination of "*Asterella*" and "*opsis*," meaning similar to the genus *Asterella*.

Species and distribution: monospecific, China and the Himalaya.

4. **Asterellopsis grollei** (D.G. Long) R.L. Zhu & YouL. Xiang, **comb.n.**

≡ *Asterella grollei* D.G. Long, Bryologist 102(2): 169. 1999. Type: Nepal, Sankhuwasabha District, Nehe Kharka, S. side of Barun Khola, 27°45′N, 87°10′E, 3075 m (above sea level), steep eroding calcareous slope of gully below cliffs, on friable soil, 29.IX.1991, Long 20777 (holotype, E; isotype, JE).

Representative specimens examined: China. Qinghai, Huangzhong Co., 36°17′55.77″ N, 101°41′53.13″ E, 2850 m, on soil, R.L. Zhu et al. 20200806–25 (HSNU); Huzhutuzu Autonomous Co., 36°45′13.48″N, 102°32′52.36″E, 2725 m, on soil, L. Shu & C. Promma 20180805–11 (HSNU); Huzhutuzu Autonomous Co., 36°44′56.58″N, 102°32′46.71″E, 2763 m asl, on soil, L. Shu & C. Promma 20180805–24 (HSNU); Xinghai Co., 34°58′54.05″N, 100°8′4.96″E, 3353 m, on soil, R.L. Zhu et al. 20 200809–5 (HSNU); Sichuan, Ruoergai Co., 34°04′29.22″N, 102°37′04.94″E, 3519 m, on soil, C. Shen et al. 20200827–32 (HSNU); Xiaojin Co., 31°11.133′N, 102°45.268′E, 3574 m, on soil, R.L. Zhu 20160815-18A (HSNU); Xiaojin Co., 31°12′19.42″N, 102°45′43.27″E, 3685 m, on soil, Y.L. Xiang & C. Shen 20190905–21 (HSNU). Yunnan, Xianggelila Co., 28°00′49.21″N, 100°02′18.32″E, 3881 m, on soil, Y.L. Xiang 20160919–41 (HSNU); Xianggelila Co., 27°37′01.35″N, 99°53′25.30″E, 4134 m, on soil, Y.L. Xiang 20160919–111 (HSNU); Y.L. Xiang 20160919–119 (HSNU).

Distribution and habitat: China and Nepal; on soil at altitudes of 2710–3881 m in coniferous forests.

Conclusions

Here, we used plastome data to investigate deep phylogenetic relationships within Marchantiopsida. We demonstrated that plastome synteny and length were highly conserved across Marchantiopsida, but identified several instances of species-specific dynamic evolutionary events. Comparative chloroplast genome analysis within this class indicates that some chloroplast genes and introns have been completely, or partially lost. Most of these losses are independent, with the exception of the loss of intron 2 of *ycf3*, which is synapomorphic for Marchantiidae. The distinct plastome structure and higher GC content in the *Cyathodium* plastomes indicate that RNA editing sites might be present in this genus, which should be noted in future studies, as it has been assumed previously that the complex thalloid liverworts completely lack RNA editing sites. The result of analyses of identification of molecular markers suggests that SSRs and six highly divergent genetic regions are suitable for species identification in Marchantiopsida. Phylogenetic analyses based on plastome sequences help to resolve phylogenetic relationships within Marchantiopsida, with high support values for both deep and shallow nodes. Based on our analyses, we herein propose an updated classification of Marchantiopsida. Our study provides a framework for future studies, which should involve characters and fossils, employ more extensive data from representative groups, such as *Cavicularia* Steph., *Cronisia* Berk., Riellaceae and Monocarpaceae, and aim at understanding the patterns and processes that generated the species diversity of Marchantiopsida.

Data availability statement

Specimen information and accession numbers in GenBank for the molecular data used in this paper were listed in Table S1.

References

Allen, N.S. and Korpelainen, H., 2006. Notes on Neotropical *Cyathodium*. Cryptogam. Bryol. 27, 85–96.

Akiyama, H. and Odrzykoski, I.J., 2020. Phylogenetic re-examination of the genus *Conocephalum* Hill. (Marchantiales: Conocephalaceae). Bryophyt. Divers. Evol. 42, 1–18. https://doi.org/10.11646/bde.42.1.1.

Beier, S., Thiel, T., Münch, T., Scholz, U. and Mascher, M., 2017. MISA-web: A web server for microsatellite prediction. Bioinformatics 33, 2583–2585. https://doi.org/10.1093/bioinformatics/btx198.

Beike, A.K., von Stackelberg, M., Schallenberg-Rűdinger, M., Hanke, S.T., Follo, M., Quandt, D., McDaniel, S.F., Reski, R., Tan, B.C. and Rensing, S.A., 2014. Molecular evidence for convergent evolution and allopolyploid speciation within the *Physcomitrium-Physcomitrella* species complex. BMC Evol. Biol. 14(158), 10.1186/1471-2148-14-158.

Blazier, J.C., Jansen, R.K., Mower, J.P., Govindu, M., Zhang, J., Weng, M. and Ruhlman, T.A., 2016. Variable presence of the inverted repeat and plastome stability in *Erodium*. Ann. Bot. 117, 1209–1220. https://doi.org/10.1093/aob/mcw065.

Bock, R., 2007. Structure, function, and inheritance of plastid genomes. In: Bock, R. (Ed.), Cell and Molecular Biology of Plastids. Springer, Berlin, pp. 29–63.

Boisselier-Dubayle, M.C., Lambourdière, J. and Bischler, H., 2002. Molecular phylogenies support multiple morphological reductions in the liverwort subclass Marchantiidae (Bryophyta). Mol. Phylogenet. Evol. 24, 66–377. https://doi.org/10.1016/s1055-7903(02)00201-4.

Borovichev, E.A., Bakalin, V.A. and Vilnet, A.A., 2015. Revision of the Russian Marchantiales. II. A review of the genus *Asterella* P. Beauv. Arctoa 24, 294–313. https://doi.org/10.15298/arctoa.24.26.

Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M. and Rochaix, J.D., 1997. The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. EMBO J. 16, 6095–6104. https://doi.org/10.1093/emboj/16.20.6095.

Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F. et al., 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell 171, 287–304. https://doi.org/10.1016/j.cell.2017.09.030.

Crandall-Stotler, B., Stotler, R., 2000. Morphology and classification of the Marchantiophyta. In: Shaw, A. and Goffinet, B. (Eds.) Bryophyte Biology. Cambridge: Cambridge University Press, pp. 21–70.

Crandall-Stotler, B., Stotler, R.E. and Long, D.G., 2009. Phylogeny and classification of the Marchantiophyta. Edinb. J. Bot. 66, 155–198. https://doi.org/10.1017/S0960428609005393.

Dong, S., Zhang, S., Zhang, L., Wu, H., Goffinet, B. and Liu, Y., 2021. Plastid genomes and phylogenomics of liverworts (Marchantiophyta): Conserved genome structure but highest relative plastid substitution rate in land plants. Mol. Phylogenet. Evol. 161, 107171. https://doi.org/10.1016/j.ympev.2021.107171.

Dong, S., Zhao, C., Zhang, S., Wu, H., Mu, W., Wei, T., Li, N., Wan, T., Liu, H., Cui, J. et al., 2019. The amount of RNA editing sites in liverwort organellar genes is correlated with GC content and nuclear PPR protein diversity. Genome Biol. Evol. 11, 3233–3239. https://doi.org/10.1093/gbe/evz232.

Duckett, J.G., Ligrone, R., Renzaglia, K.S. and Pressel, S., 2014. Pegged and smooth rhizoids in complex thalloid liverworts (Marchantiopsida): Structure, function and evolution. Bot. J. Linn. Soc. 174, 68–92. https://doi.org/10.1111/boj.12121.

Fan, W.B., Wu, Y., Yang, J., Shahzad, K. and Li, Z.H., 2018. Comparative chloroplast genomics of Dipsacales species: Insights into sequence variation, adaptive evolution, and phylogenetic relationships. Front. Plant Sci. 9, 689. https://doi.org/10.3389/fpls.2018.00689.

Flores, J.R., Bippus, A.C., Suárez, G.M. and Hyvönen, J., 2021. Defying death: Incorporating fossils into the phylogeny of the complex thalloid liverworts (Marchantiidae, Marchantiophyta) confirms high order clades but reveals discrepancies in family-level relationships. Cladistics 37, 231–247. https://doi.org/10.1111/cla.12442.

Flores, J.R., Catalano, S.A., Muñoz, J. and Suárez, G.M., 2018. Combined phylogenetic analysis of the subclass Marchantiidae (Marchantiophyta): Towards a robustly diagnosed classification. Cladistics 34, 517–541. https://doi.org/10.1111/cla.12225.

Forrest, L.L., Davis, E.C., Long, D.G., Crandall-Stotler, B.J., Clark, A. and Hollingsworth, M.L., 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): Multiple taxa, genomes and analyses. Bryologist 109, 303–334. https://doi.org/10.1639/0007-2745(2006)109[303:utehot]2.0.co;2.

Forrest, L.L., Long, D.G., Cargill, D.C., Hart, M.L., Milne, J., Schill, D.B., Seppelt, R.D. and Villarreal, J.C., 2015. On *Monocarpus* (Monocarpaceae, Marchantiopsida), an isolated salt-pan complex thalloid liverwort. Aust. Syst. Bot. 28, 137–144. https://doi.org/10.1071/SB15012.

Frangedakis, E., Guzman-Chavez, F., Rebmann, M., Markel, K., Yu, Y., Perraki, A., Tse, S.W., Liu, Y., Rever, J., Sauret-Gueto, S. et al., 2021. Construction of DNA tools for hyperexpression in *Marchantia* chloroplasts. ACS Synth. Biol. 10, 1651–1666. https://doi.org/10.1021/acssynbio.0c00637.

Grolle, R., 1983. Nomina generica Hepaticarum; references, types and synonymies. Acta Bot. Fenn. 121, 1–63.

Gonçalves, D.J.P., Simpson, B.B., Ortiz, E.M., Shimizu, G.H. and Jansen, R.K., 2019. Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. Mol. Phylogenet. Evol. 138, 219–232. https://doi.org/10.1016/j.ympev.2019.05.022.

Heath, T.A., Hedtke, S.M. and Hillis, D.M., 2008. Taxon sampling and the accuracy of phylogenetic analyses. J. Syst. Evol. 46, 239–257.

Jin, J.J., Yu, W.B., Yang, J.B., Song, Y., Depamphilis, C.W., Yi, T.S. and Li, D.Z., 2020. GetOrganelle: A fast and versatile toolkit for accurate de novo assembly of organelle genomes. Genome Biol. 21, 241. https://doi.org/10.1186/s13059-020-02154-5.

Jovet-Ast, S., 2005. Riccia. In: Bischler-Causse, H., Gradstein, S.R., Jovet-Ast, S., Long, D.G. and Salazar, A.N. (Eds.), Marchantiidae, Flora Neotropica Monograph. NY Botanical Garden Press, New York, NY, pp. 25–123.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A. and Jermiin, L.S., 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Kamisugi, Y., von Stackelberg, M., Lang, D., Care, M., Reski, R., Rensing, S.A. and Cuming, A.C., 2008. A sequence-anchored genetic linkage map for the moss, *Physcomitrella patens*. Plant J. 56, 855–866. https://doi.org/10.1111/j.1365-313x.2008.03637.x.

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. https://doi.org/10.1093/molbev/mst010.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C. et al., 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649. https://doi.org/10.1093/bioinformatics/bts199.

Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T. and Calcott, B., 2017. PartitionFinder 2: New methods for selecting

partitioned models of evolution for molecular and morphological phylogenetic analyses. Mol. Biol. Evol. 34, 772–773. https://doi.org/10.1093/molbev/msw260.

Lohse, M., Drechsel, O., Kahlau, S. and Bock, R., 2013. OrganellarGenomeDRAW—A suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. Nucleic Acids Res. 41, W575–W581. https://doi.org/10.1093/nar/gkt289.

Long, D.G., 1999. Studies on the genus *Asterella*. IV. *Asterella grollei* sp. nov., a new species from eastern Asia related to the American *A. palmeri*. Bryologist 102, 169–178.

Long, D., 2005. Studies on the genus *Asterella* (Aytoniaceae) VI: Infrageneric classification in *Asterella*. J. Hattori. Bot. Lab. 97, 249–261.

Long, D.G., 2006a. New higher taxa of complex thalloid liverworts (Marchantiophyta-Marchantiopsida). Edinb. J. Bot. 63, 257–262. https://doi.org/10.1017/S0960428606000606.

Long, D.G., 2006b. Revision of the genus *Asterella* P. Beauv. in Eurasia. Bryophyt. Biblioth. 63, 1–299.

Long, D.G., Forrest, L.L., Villarreal, J.C. and Crandall-Stotler, B.J., 2016. Taxonomic changes in Marchantiaceae, Corsiniaceae and Cleveaceae (Marchantiidae, Marchantiophyta). Phytotaxa 252, 77–80. https://doi.org/10.11646/phytotaxa.252.1.9.

Long, D.G., Möller, M. and Preston, J., 2000. Phylogenetic relationships of *Asterella* (Aytoniaceae, Marchantiopsida) inferred from chloroplast DNA sequences. Bryologist 103, 625–644. https://doi.org/10.1639/0007-2745(2000)103[0625:PROAAM]2.0.CO;2.

Machado, L.O., Vieira, L.D.N., Stefenon, V.M., Faoro, H., Pedrosa, F.O., Guerra, M.P. and Nodari, R.O., 2020. Molecular relationships of *Campomanesia xanthocarpa* within Myrtaceae based on the complete plastome sequence and on the plastid *ycf2* gene. Genet. Mol. Biol. 10, e20180377. https://doi.org/10.1590/1678-4685-GMB-2018-0377.

Miller, M.A., Pfeiffer, W. and Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. In: Gateway Computing Environments Workshop (GCE). IEEE, New Orleans, LA. pp. 1–8.

Nguyen, L., Schmidt, H.A., von Haeseler, A. and Minh, B.Q., 2014. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Petersen, K., Schöttler, M.A., Karcher, D., Thiele, W. and Bock, R., 2011. Elimination of a group II intron from a plastid gene causes a mutant phenotype. Nucleic Acids Res. 39, 5181–5192. https://doi.org/10.1093/nar/gkr105.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A. and Huelsenbeck, J.P., 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. https://doi.org/10.1093/sysbio/sys029.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J.C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S.E. and Sánchez-Gracia, A., 2017. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol. Biol. Evol. 34, 3299–3302. https://doi.org/10.1093/molbev/msx248.

Rüdinger, M., Polsakiewicz, M. and Knoop, V., 2008. Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. Mol. Biol. Evol. 25, 1405–1414. https://doi.org/10.1093/molbev/msn084.

Rüdinger, M., Volkmar, U., Lenz, H., Groth-Malonek, M. and Knoop, V., 2012. Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. J. Mol. Evol. 74, 37–51. https://doi.org/10.1007/s00239-012-9486-3.

Sawicki, J., Bączkiewicz, A., Buczkowska, K., Górski, P., Krawczyk, K., Mizia, P., Myszczyński, K., Ślipikoipiko, M. and Szczecińska, M., 2020. The increase of simple sequence repeats during diversification of Marchantiidae, an early land plant lineage, leads to the first known expansion of inverted repeats in the evolutionarily-stable structure of liverwort plastomes. Genes 11, 299. https://doi.org/10.3390/genes11030299.

Sawicki, J., Kwaśniewski, M., Szczecińska, M., Chwiałkowska, K., Milewicz, M. and Plášek, V., 2012. Isolation and characterizationof Simple Sequence Repeats (SSR) Markers from the moss genus Orthotrichum using a small throughput pyrosequencing machine. Int. J. Mol. Sci 13, 7586–7593. https://doi.org/10.3390/ijms13067586

Sayyari, E. and Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.

Schill, D.B., Long, D.G. and Forrest, L.L., 2010. A molecular phylogenetic study of *Mannia* (Marchantiophyta, Aytoniaceae) using chloroplast and nuclear markers. Bryologist 113, 164–179. https://doi.org/10.1639/0007-2745-113.1.164.

Schuster, R.M., 1992. Studies on Marchantiales, I–III. J. Hattori Bot. Lab. 71, 267–287.

Shi, L., Chen, H., Jiang, M., Wang, L., Wu, X., Huang, L. and Liu, C., 2019. CPGAVAS2, an integrated plastome sequence annotator and analyzer. Nucleic Acids Res. 47, W65–W73. https://doi.org/10.1093/nar/gkz345.

Sinn, B.T., Sedmak, D.D., Kelly, L.M. and Freudenstein, J.V., 2018. Total duplication of the small single copy region in the angiosperm plastome: Rearrangement and inverted repeat instability in *Asarum*. Am. J. Bot. 105, 71–84. https://doi.org/10.1002/ajb2.1001.

Ślipiko, M., Myszczyński, K., Buczkowska, K., Bączkiewicz, A., Szczecińska, M. and Sawicki, J., 2020. Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. BMC Plant Biol. 20, 243. https://doi.org/10.1186/s12870-020-02435-y.

Söderström, L., Hagborg, A., von Konrat, M., Bartholomew-Began, S., Bell, D., Briscoe, L., Brown, E., Cargill, C.D., Pinheiro da Costa, D., Crandall-Stotler, B.J. et al., 2016. World checklist of hornworts and liverworts. PhytoKeys 59, 1–828. https://doi.org/10.3897/phytokeys.59.6261.

Stamatakis, A., 2014. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Swofford, D.L., 2002. PAUP*: Phylogenetic analysis using parsimony (and other methods), version 4.0 Beta. Sinauer associates, Sunderland, Massachusetts. https://paup.phylosolutions.com/.

Villarreal, A.J.C., Crandall-Stotler, B.J., Hart, M.L., Long, D.G. and Forrest, L.L., 2016. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. New Phytol. 209, 1734–1746. https://doi.org/10.1111/nph.13716.

Wang, X., Yang, Z., Zhang, Y., Zhou, W., Zhang, A. and Lu, C., 2020. Pentatricopeptide repeat protein PHOTOSYSTEM I BIOGENESIS FACTOR2 is required for splicing of *ycf3*. J. Integr. Plant Biol. 62, 1741–1761. https://doi.org/10.1111/jipb.12936.

Wu, S., Chen, J., Li, Y., Liu, A., Li, A., Yin, M., Shrestha, N., Liu, J. and Ren, G., 2021. Extensive genomic rearrangements mediated by repetitive sequences in plastomes of *Medicago* and its relatives. BMC Plant Biol. 21, 421. https://doi.org/10.1186/s12870-021-03202-3.

Wu, Y., Liu, F., Yang, D.G., Li, W., Zhou, X.J. and Li, Z.H., 2018. Comparative chloroplast genomics of *Gossypium* species: Insights into repeat sequence variations and phylogeny. Front. Plant Sci. 9, 376. https://doi.org/10.3389/fpls.2018.00376.

Ye, W.Q., Yap, Z.Y., Li, P., Comes, H. and Qiu, Y.X., 2018. Plastome organization, genome-based phylogeny and evolution of plastid genes in Podophylloideae (Berberidaceae). Mol. Phylogenet. Evol. 127, 978–987. https://doi.org/10.1016/j.ympev.2018.07.001.

Yu, Y., Liu, H.M., Yang, J.B., Ma, W.Z., Pressel, S., Wu, Y.H. and Schneider, H., 2019. Exploring the plastid genome disparity of liverworts. J. Syst. Evol. 57, 382–394. https://doi.org/10.1111/jse.12515.

Yu, Y., Yang, J.B., Ma, W.Z., Pressel, S., Liu, H.M., Wu, Y.H. and Schneider, H., 2020. Chloroplast phylogenomics of liverworts: A reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. Cladistics 36, 184–193. https://doi.org/10.1111/cla.12396.

Zhang, C., Rabiee, M., Sayyari, E. and Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang, X., Sun, Y., Landis, J.B., Lv, Z., Shen, J., Zhang, H., Lin, N., Li, L., Sun, J., Deng, T. et al., 2020. Plastome phylogenomic study of Gentianeae (Gentianaceae): Widespread gene tree discordance and its association with evolutionary rate heterogeneity of plastid genes. BMC Plant Biol. 20, 340. https://doi.org/10.1186/s12870-020-02518-w.

Supporting Information

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

Table S1 Voucher and sequence information for the 133 liverwort plastomes included in this study. Accession numbers in bold correspond to sequences generated herein.

Fig. S1. ML tree of the Marchantiopsida, inferred from the entire chloroplast genome minus the second inverted repeat region (IRa). Numbers at nodes correspond to ML bootstrap values (ML-BS).

Fig. S2. ML tree of the Marchantiopsida, inferred from the concatenated noncoding sequences. Numbers at nodes correspond to ML bootstrap values (ML-BS).

Fig. S3. Gene map of a representative Marchantiopsida chloroplast genome generated in the present study (*Marchantia quadrata* Scop.). Genes shown outside the outer circle are transcribed clockwise and those inside are transcribed counterclockwise. Genes are coloured by functional group. Histogram in the shaded area of the inner circle corresponds to GC content.

Fig. S4. Gene arrangements (on the *x*-axis) in plastomes of Marchantiopsida species and the two outgroups (on the *y*-axis) which are mapped on the ML phylogeny shown in Fig. 2, showing gene loss (red), pseudogenes (yellow), duplicated loci (dark blue), intron loss (green) and gene inversion (blue).

Fig. S5. Types and counts of SSRs in the chloroplast genomes of species of the Marchantiopsida, mapped on the phylogenetic tree shown in Fig. 2. SSRs are coloured by type; bar lengths correspond to the number (on the *x*-axis) of SSRs of each type in a given sample. Numbers to the right of each bar indicate the total number of SSR types identified in the sample.

Fig. S6. Sliding window analysis of nucleotide diversity (Pi) along the plastid genome among Marchantiopsida species. The plastome regions LSC (large single-copy), SSC (small single-copy) and IR (inverted repeated regions) are shown along the *x*-axis.

Fig. S7. Species tree of Marchantiopsida, inferred from the amino acid data using a coalescent approach in ASTRAL. Numbers above branches indicate node support (LPP)