WILEY-BLACKWELL

Cladistics (2019) 1–10

Cladistics

10.1111/cla.12396

Chloroplast phylogenomics of liverworts: a reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales

Ying Yu[a], Jun-Bo Yang[b], Wen-Zhang Ma[c], Silvia Pressel[d], Hong-Mei Liu[e], Yu-Huan Wu[a,*] and Harald Schneider[f,*]

[a]*College of Life and Environmental Sciences, Hangzhou Normal University, Hangzhou 311121, China;* [b]*CAS Plant Germplasm and Genomics Center, Germplasm Bank of Wild Species, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China;* [c]*CAS Key Laboratory for Plant Biodiversity and Biogeography of East Asia, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming 650201, China;* [d]*Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK;* [e]*Key Laboratory of Tropical Plant Resources and Sustainable Use, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China;* [f]*Center of Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan 666303, China*

Accepted 30 May 2019

Abstract

As one of the four main lineages diverging from the early diversification of land plants, the phylogeny of liverworts holds the information about nearly 500 Myr of independent adaptation to changing environments. Thus, resolving the phylogenetic history of liverworts will provide unique insights into the successful diversification of early land plants in terrestrial ecosystems. However, the deep diverging events of this group remain incompletely resolved, such as the definite position of Ptilidiales. Here, we aimed to reconstruct the backbone relationships of liverworts using 84 protein-coding chloroplast genes, a dataset comprising 35 representatives from all major lineages of liverworts, and three phylogenetic analyses, namely maximum parsimony, maximum likelihood and Bayesian inference. To test the impact of composition biases, the phylogenetic analyses were carried out using three alignments representing the same dataset either as: (i) nucleotides, (ii) amino acids, or (iii) recoded nucleotides applying ambiguity base code. Chloroplast genome data consistently supported the monophyletic origin of three major lineages in liverworts, as well as the majority of backbone relationships. Ptilidiales were found to be sister to Jungermanniales. The rapid accumulation of G/C tracks as a consequence of increased GC content is an important cause for the long branches inferred in this group. Our study not only provides empirical evidence to support the significance of plastid genome sequencing to reconstruct the phylogeny of this important plant lineage, but also suggests that the GC content has played a critical role in the evolutionary dynamics of plastid genomes in land plants.

Introduction

Reconstruction of phylogenetic history requires resolution of incongruences caused by the complexity of the evolutionary processes shaping the tree-of-life. Several hypotheses have been proposed to explain these incongruences, including long-branch attraction, incomplete lineage sorting, absence of phylogenetic signals, gene heterogeneity, small sample size and inappropriate models employed in the analytical methods (e.g. Felsenstein, 1978; Philippe et al., 2005; Degnan and Rosenberg, 2009; Philippe and Roure, 2011). Even when sister relationships appear to be strongly supported by bootstrap or posterior confidence values, these may, in fact, be artifacts resulting from biases, such as nucleotide composition biases (Cox et al., 2014; de Sousa et al., 2018) and/or gene heterogeneity (Philippe et al., 2011). Recently, numerous studies have sought to resolve phylogenetic inconsistency by analyzing large DNA matrices (e.g. genome-wide

*Corresponding authors:
E-mail addresses: yuhuanwu@hznu.edu.cn (Y-HW); harald@xtbg.ac.cn (HS)

datasets) combined with complex substitution rate models and/or including more taxa (e.g. Moore et al., 2007; Giribet, 2016; Puttick et al., 2018). However, these methods have been widely used to resolve phylogenetic conflicts in animals and seed plants (e.g. Moore et al., 2007; Dunn et al., 2008; Giribet, 2016), and were recently applied to clarify the relationships among four major land plant lineages (mosses, liverworts, hornworts and vascular plants, Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2018), but are rarely used to resolve the backbone phylogeny of liverworts, even though more than one species of this group had been included in previous phylogenomic studies (Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2018).

In the past two decades, numerous phylogenetic studies have greatly improved our understanding of the origin of liverworts (Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2018) as well as the relationships among major lineages in this group (Heinrichs et al., 2005, 2007; Forrest et al., 2006; He-Nygrén et al., 2006; Villarreal et al., 2016). These remarkable improvements include that liverworts were strongly supported as sister to mosses in recent phylogenomic studies (Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2018) than the alternative hypothesis of this group as the earliest lineage of land plants (Qiu et al., 1998, 2006). These findings, consistent with those of the first chloroplast phylogenomic analysis (Nishiyama et al., 2004), not only provide new insights into the deepest diverging events of land plants, but also enable us to reappraise hypotheses on the evolution of key innovative morphological characters (e.g. stomata) and evolutionary events, based—hitherto—on early land plant fossils (Morris et al., 2018; Puttick et al., 2018). In turn, these novel results reignite discussions on the ecological–evolutionary history of liverworts (see, for example, Cooper et al., 2012; Feldberg et al., 2014; Laenen et al., 2016a, b). Furthermore, instead of the traditional division of liverworts into two lineages based on the body-plan of the gametophyte generation—thus, thalloid and leafy liverworts—the currently accepted classification scheme recognizes three major lineages. The Haplomitropsida, comprising two morphologically distinct but species poor orders, Calobryales and Treubiales, were found to be sister to the core liverworts (Heinrichs et al., 2005). The latter group comprises the Marchantiopsida (five orders, including all complex thalloid taxa, Villarreal et al., 2016) and the Jungermanniopsida (eight orders, including most simple thalloid and nearly all leafy species; a simplified phylogeny of liverworts is shown in Figure S1). Despite such remarkable progress, the backbone phylogeny of liverworts remains incompletely resolved, such as the placement of Ptilidiales. This possibly is due to insufficient genetic information provided by single gene or a few loci, or biases caused by sequence heterogeneity. Here, we used a large-scale genomic dataset to test this hypothesis with particular focus on the position of the order Ptilidiales. Resolving the position of this order is arguably crucial in order to elucidate the divergence of two major species-rich clades of leafy liverworts, namely Jungermanniales and Porellales.

As currently recognized, the order Ptilidiales comprises three small but well-defined families (Davis, 2004; Heinrichs et al., 2005; Forrest et al., 2006; Hendry et al., 2007; Liu et al., 2008), namely Herzogianthaceae (2 spp.), Neotrichocoleaceae (2 spp.) and Ptilidiaceae (3 spp.). Species in this order are united by several morphological characters, such as rhizoids in fascicles originating from the underleaf-base, *Frullania*-type branches, and ovoid to ellipsoidal capsules that are 4- to 7-stratose (Crandall-Stotler et al., 2009). Nevertheless, the placement of Ptilidiales remains controversial. In some studies, this order was placed as sister to Porellales, whereas in others it was resolved as sister to Jungermanniales (e.g. Davis, 2004; Forrest et al., 2006; Cooper et al., 2012).

In order to test these two alternative hypotheses of Ptilidiales, we newly generated chloroplast genome sequences for 32 bryophyte taxa, including 31 liverwort species and one hornwort taxon. By combining these data with four previously published genomes, we reconstructed a phylogeny for 35 liverwort species representing three classes and 13 of 15 orders using 84 protein-coding chloroplast genes and three methods of analysis, namely maximum parsimony (MP), maximum likelihood (ML) and Bayesian inference (BI). To test the potential biases caused by sequence composition (e.g. Cox et al., 2014; Cox, 2018; de Sousa et al., 2018), the analyses were carried out using three alignments representing the same dataset scored as: (i) nucleotides, (ii) amino acids, and (iii) recoded nucleotides applying ambiguity base code (Liu and Bundschuh, 2005). Our goal was to not only provide a robust backbone phylogeny of liverworts by elucidating the position of Ptilidiales, but also to shed novel lights into the pattern of chloroplast genome evolution in liverworts.

Materials and methods

Taxon sampling, DNA sequencing and alignments

The in-group consisted of 35 liverwort species (*c*. 0.48% of liverwort species diversity), representing 33 families (42% of family diversity) and 13 orders (87% of order diversity). The chloroplast genome data for 31 liverwort species were newly generated as part of our project on the diversity of liverwort chloroplast

genomes (see, also, Yu et al., unpublished) and four were extracted from previously published studies: *Aneura pinguis* (Myszczyński et al., 2017), *Marchantia polymorpha* (Ohyama et al., 1986), *Apopellia endiviifolia* (Grosche et al., 2012) and *Ptilidium pulcherrimum* (Forrest et al., 2011). Specimen information and accession numbers were listed in Table S1. Special attention was given to the sampling of Ptilidiales (2 spp.), Porellales (7 spp.) and Jungermanniales (11 spp.). As for the outgroup, seven species were sampled to represent the other two lineages of bryophytes, mosses (5 spp.) and hornworts (2 spp.). The outgroup sampling reflects the current hypothesis of the diversification of early land plants, in which liverworts and mosses form a monophylum, and hornworts are likely the sister to this clade (Puttick et al., 2018; de Sousa et al., 2018). The chloroplast genome data for six out-group taxa were downloaded from GenBank, whereas the chloroplast genome sequence of one hornwort species *Megaceros flagellaris* was generated in this study (Table S1).

Total genomic DNA was extracted from 10 to 50 μg of dry gametophyte material using Tiangen DNAsecure Plant Kit (DP320). A genome library was constructed using Illumina Nextera XT DNA library preparation based on *c*. 500-bp-long DNA-fragments obtained by shearing the genomic DNA following the manufacturer's manual (Illumina, San Diego, CA, USA). By generating 90-bp-long paired-end sequences using an Illumina HiSeq 2000 at BGI-Shenzhen, about 200 Mb of sequences were generated for each sample. These analyses were conducted at the Germplasm Bank of Wild Species, Kunming Institute of Botany (KIB, Kunming, China).

Sequences were assembled into scaffolds employing CLC Genomic workbench v.12 (https://www.qiagenbioinformatics.com/products/clc-genomics-workbench/) involving quality control of the raw sequences with the NGS QC Tool Kit (http://nipgr.res.in/ngsqctoolkit.html) with cut-off values for read length and PHRED quality scores set as recommended in Yang et al. (2014). The target 84 protein-coding chloroplast genes were extracted from scaffolds using Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) utilizing previously published liverwort genomes as reference. Then, a nucleotide sequence alignment concatenating 84 protein-coding genes (NSA) with a length of 58 731 bp were generated using Geneious R11 (Kearse et al., 2012). To test the effects of composition heterogeneity, we rescored the nucleotide dataset as an amino acid alignment (ASA, a length of 19 577 amino acids) and a recoded nucleotide sequence alignment using ambiguity base code (RSA, a length of 58 731 bp). The codon-degeneracy coding method follows Regier et al. (2010) and Cox et al. (2014). The final alignments and best data partition schemes for all analyses were deposited in the Figshare Digital Repository (https://doi.org/10.6084/m9.figshare.8206865.v1).

Phylogenetic analyses and frequency of poly(G/C)

All three alignments were analyzed using three phylogenetic methods: MP, ML and BI. First, we used PartitionFinder 2 (Lanfear et al., 2017) to determine the optimal data partition scheme and nucleotide substitution models for the three alignments, resulting in 207 partitions for NSA, 10 for ASA and 151 for RSA. Secondly, we conducted MP analyses using Mega 6 (Tamura et al., 2013) with 100 bootstrap replicates (MP-BS) and set all other parameters in the default settings, including max-mini branch-and-bound search and the number of initial trees. All nucleotide changes were set to be weighted equally and the codon position selected. The ML analyses were conducted using RaxML-HPC2 8.2.10 (Stamatakis, 2014) with 100 bootstrap replicates (ML-BS). The optimum model GTR + G + I estimated using PartitionFinder was applied. We performed all BI analyses using MrBayes 3.2.6 (Ronquist et al., 2012) with default two runs, four chains, and unlinked rates for five billion generations with sampling every 2500th generations. The convergence of runs and burn-in phase was assessed using Tracer v.1.4.1 (http://beast.bio.ed.ac.uk/Tracer). Bayesian posterior probabilities (BPP) were calculated for the maximum clade credibility tree of all samples after discarding 20% trees as burn-in. In the two model-based approaches, ML and BI, we set on the number of parameters, whereas the parameter values were estimated empirically synchronous to the tree search to enable a more extensive search of the parameter space.

In primary ML and BI analyses, we found that some branches were relatively longer than others, coincident with their greater number of homopolymeric poly-guanine/poly-cytosine tracts (G/C tracts). The G/C tracts have been assumed to be a consequence of DNA polymerase slippage, the latter, in turn, is largely affected by the level of GC content (Boyer et al., 2002; Kiktev et al., 2018). Therefore, positive correlations between branch length, diversity of G/C tracts and GC content were expected. To test it, we first compared the branch length and G/C tracts for 11 pairs of sister lineages (sister lineages selected using the phylogeny generated; Table S2) to assess whether longer branches are underpinned by a greater diversity of G/C tracts. Secondly, we optimized two variants, diversity of G/C tracts (number of G/C tracts per kb) and GC content (Table S3), for the samples in the phylogeny using the parsimony ancestral character reconstruction approach, and further tested their correlation using Felsenstein independent contrast (Felsenstein,

1985) as implemented in MESQUITE 3.51 (Maddison and Maddison, 2018).

Results

All analyses resulted in trees with highly similar topologies, except for the position of Ptilidiales and some family-level relationships in Jungermanniales and Marchantiales (Figs 1 and 2). Haplomitropsida were found to be sister to core liverworts including two lineages corresponding to Marchantiopsida and Jungermaniopsida (MP-BS > 95%, ML-BS = 100%, BPP $P = 1.0$; Figs 1 and 2). The Blasiales were supported as sister to the remaining Marchantiopsida, followed by Neohodgsoniales and Lunulariales (MP-BS ≥ 95%, ML-BS > 95%, BPP $P = 1.0$; Figs 1a,c and 2a–c). Within Marchantiales, the family-level relationships were inconsistent between MP and ML/BI methods as well as among the three treatments (Figs 1 and 2). In the former case, using the NSA treatment, Marchantiaceae were found to be sister to Cyathodiaceae in the MP analysis (MP-BS = 98%; Fig. 1a), and in turn sister to remaining Marchantiale families in both the ML and BI analyses (ML-BS = 100%, BPP $P = 1.0$; Fig. 2a). In the latter case, taking the BI analyses for an example, *Riccia* (Ricciaceae) was resolved as sister to *Conocephalum* (Conocephalaceae) using the NSA treatments (BPP $P = 1.0$; Fig. 2a), to the clade consisting of *Conocephalum* and *Cyathodium* (Cyathodiaceae) using the ASA treatments (BPP $P = 1.0$; Fig. 2b), and to *Cythodum* using the RAS treatments (BPP $P = 0.98$; Fig. 2c).

In the Jungermaniopsida, all analyses supported a division of simple thalloid liverworts plus a monophyletic leafy liverwort lineage Pleuroziaceae into two clades: Pelliidae and Metzgeriidae (MP-BS > 95%, ML-BS > 80% and BPP $P > 0.95$; Figs 1 and 2). These two subclasses form a grade leading to the core leafy liverworts Jungermanniidae, in which three subclades were recognized, namely Jungermanniales, Porellales and Ptilidiales (MP-BS > 95%, ML-BS = 100% and BPP $P = 1.0$; Figs 1 and 2). Our analyses recovered alternative positions for Ptilidiales, that is, this order was resolved as sister to Porellales using the NSA treatments (ML-BS = 94%, BPP $P = 1.0$; Fig. 2a), but as sister to Jungermanniales using the ASA and RSA treatments (MP-BS = 85%, ML-BS = 81%, BPP $P > 0.95$; Figs 1b and 2b,c). Within Porellales, Jubulineae was found to be the sister to a clade consisting of Radulineae and Porellineae. The relationships among three major subclades in Jungermanniales remain problematic. The suborder Jungermanniineae was resolved as sister to Cephaloziineae using the NSA treatments in the BI analyses (BPP $P = 1.0$; Fig. 2a), but as sister to Lophocoleineae using the RSA treatments in the MP and BI analyses (MP-BS = 81%, BPP $P = 1.0$; Figs 1c and 2c).

A positive correlation between branch length and diversity of G/C tracts was found in six of 11 lineage pairs (Table S2). A parallel link was found between the diversity of G/C tracts and GC content, supported in both the optimization analyses and Felsenstein independent contrast ($P < 0.01$). In addition, the diversity of G/C tracts and GC content show a pattern of phylogenetic clustering, in particular the values of both parameters were relatively low in the early diverging lineages and increased in derived groups. A remarkable exception is the genus *Haplomitrium* showing the highest GC content and greatest diversity of G/C tracts (Fig. 3, Table S2).

Discussion

Our results confirm the capability of the phylogenomic approach using whole chloroplast genome sequences to resolve phylogenetic relationships among bryophyte lineages, so long as the analyses take into account possible misleading processes such as amino acid biases (see Cox et al., 2014; Cox, 2018). By doing so, the results supported the hypothesis that Ptilidiales are sister to the Jungermanniales. This topology remains the most likely hypothesis to date, pending further analyses integrating more information such as incorporating several hundreds of nuclear genome-based loci.

The backbone phylogeny of liverworts

Our phylogenetic analyses using whole chloroplast genome data not only support the current classification of liverworts proposed by Söderström et al. (2016), but also provide new evidence for previously unresolved relationships. The inconsistent placement of Ptilidales among the three alignments indicates that the affinity of Ptilidiales and Porellales recovered in previous studies (Heinrichs et al., 2005, 2007; He-Nygrén et al., 2006) as well as in our NSA analyses (Fig. 1a) may be caused by composition base biases among synonymous substitutions. Nevertheless, the alternative hypothesis of Ptilidiales as sister to Jungermanniales did not obtain high bootstrap values in most of the MP and ML analyses (BS < 85%; Figs 1 and 2), although the posterior probability values were $P = 1.00$ in the BI analyses (Fig. 2b,c). The low bootstrap values could be interpreted as a consequence of either insufficient net phylogenetic signals as indicated by its relative short branch length connecting clades, or unsuitability of the MP and ML methods to correctly handle multiple substitutions based on the present sample size (Hendy and Penny, 1989; Philippe

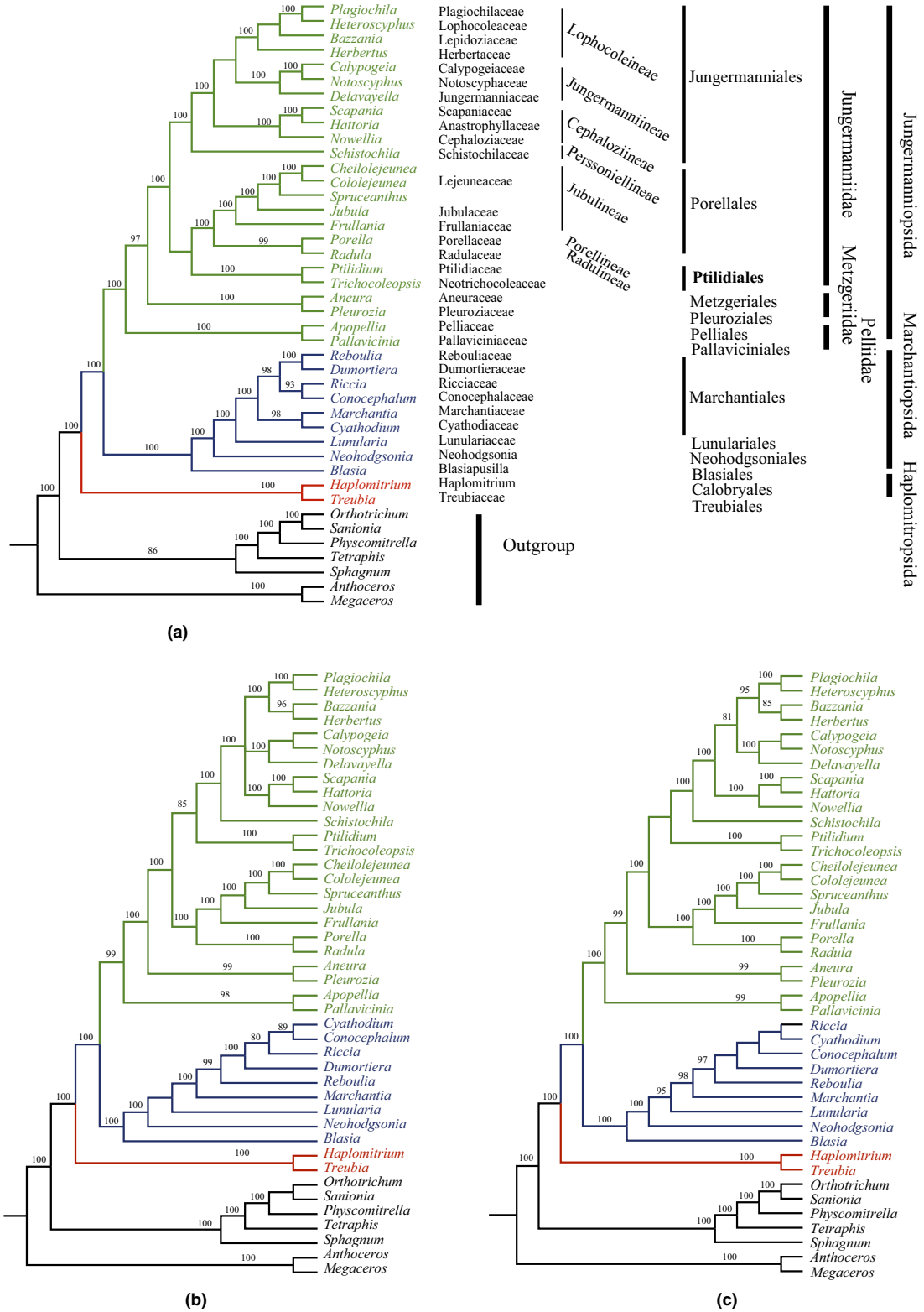

Fig. 1. The three hypotheses of the liverwort phylogeny using maximum parsimony (MP) analyses based on a sampling of 42 bryophyte taxa and three alignments comprising of 84 chloroplast protein-coding genes scored as (a) nucleotides, (b) amino acids and (c) recoded nucleotides applying the ambiguity codon code (Cox et al., 2014). Only the genera of representatives are shown. The ingroup includes 35 liverwort taxa covering 13 of 15 orders and divides into three major lineages: Haplomitropsida (red), Marchantiopsida (blue) and Jungermanniopsida (green). The numbers above branches are MP Bootstrap values (MP-BS) > 80%.

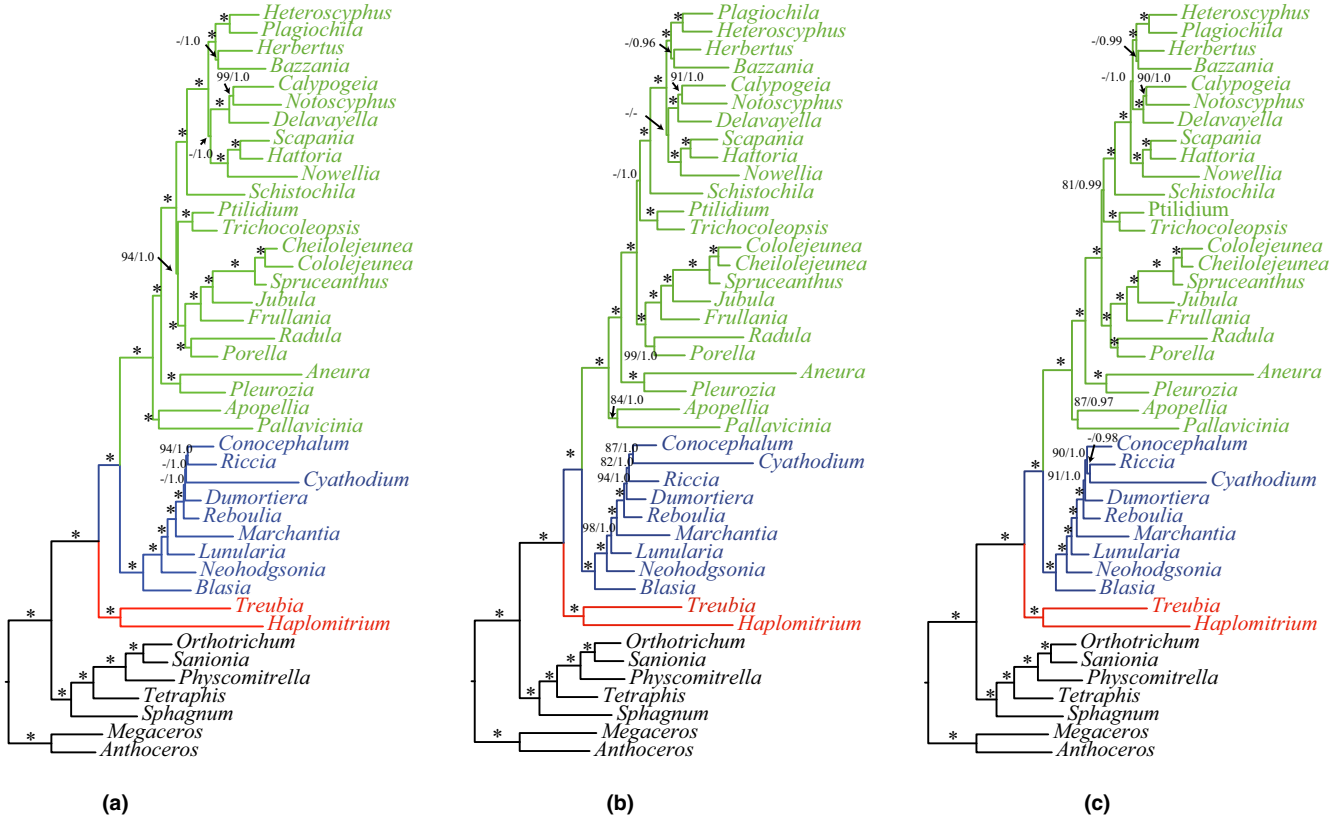

Fig. 2. The three hypotheses of the liverwort phylogeny using maximum-likelihood (ML) analyses based on a sampling of 42 bryophyte taxa and three alignments comprising of 84 chloroplast protein-coding genes scored as (a) nucleotides, (b) amino acids, and (c) recoded nucleotides applying the ambiguity codon code (Cox et al., 2014). Only the genera of representatives are shown. The ingroup includes 35 liverwort taxa covering 13 of 15 orders and divides into three major lineages: Haplomitropsida (red), Marchantiopsida (blue) and Jungermanniopsida (green). The numbers above branches are MP Bootstrap values (ML-BS)/Bayesian Posterior Probability (BPP). "*"and "–" indicate ML-BS = 100%/BPP = 1.0 and ML-BS < 85%/BPP < 0.95, respectively.

et al., 2011). Composition base biases also may explain, in part, topological conflicts recovered among the phylogenetic hypotheses of other lineages obtained using three alignments, with or without taking this bias into account. Examples included the affinities among Marchantiales families, and those among Lophocoleineae, Jungermanniineae and Cephaloziineae (Figs 1 and 2). However, we cannot reject alternative hypotheses, such as amino acid biases related to sequence heterogeneity (e.g. the GC-rich level affecting the content of amino acids; Foster et al., 1997) given the conflicts in the resolution of Jungermanniineae between the ASA and RSA treatments (Fig. 2b,c), and long-branch attractions (Philippe et al., 2011) given the variations of branch length among *Conocephalum*, *Cyathodium* and *Riccia* (Fig. 2).

A division of Porellales into the three subclades Radulineae, Jubulineae and Porellineae, reported here using genome-scale datasets, is in agreement with previous studies based on genomic regions (Forrest et al., 2006; Cooper et al., 2012). However, affinities among these three subclades remain problematic, as Radulineae are supported as sister to Porellineae reported herein (Figs 1 and 2), whilst in the multi-locus analyses the former is well resolved as sister to Jubulineae (Forrest et al., 2006).

Branch length, diversity of G/C tracts and GC content

Variations in estimates of branch length are caused by differences in mutation rates or related mechanisms (Lanfear et al., 2010; Bromham et al., 2015). In this study, the branches leading to *Aneura* (Jungermanniopsida), *Cyathodium* (Marchantiopsida) and *Haplomitrium* (Haplomitriopsida) are relatively longer than those leading to their closest relatives, indicating that these groups have accumulated more genetic changes since diverging from the common ancestor. However, we know little about the processes shaping this evolutionary pattern of liverwort chloroplast genomes, as rather few phylogenetic/phylogenomic studies have been conducted to explore the effects of intrinsic (e.g. body design and GC content) and/or extrinsic factors (e.g. biological productivity and environmental

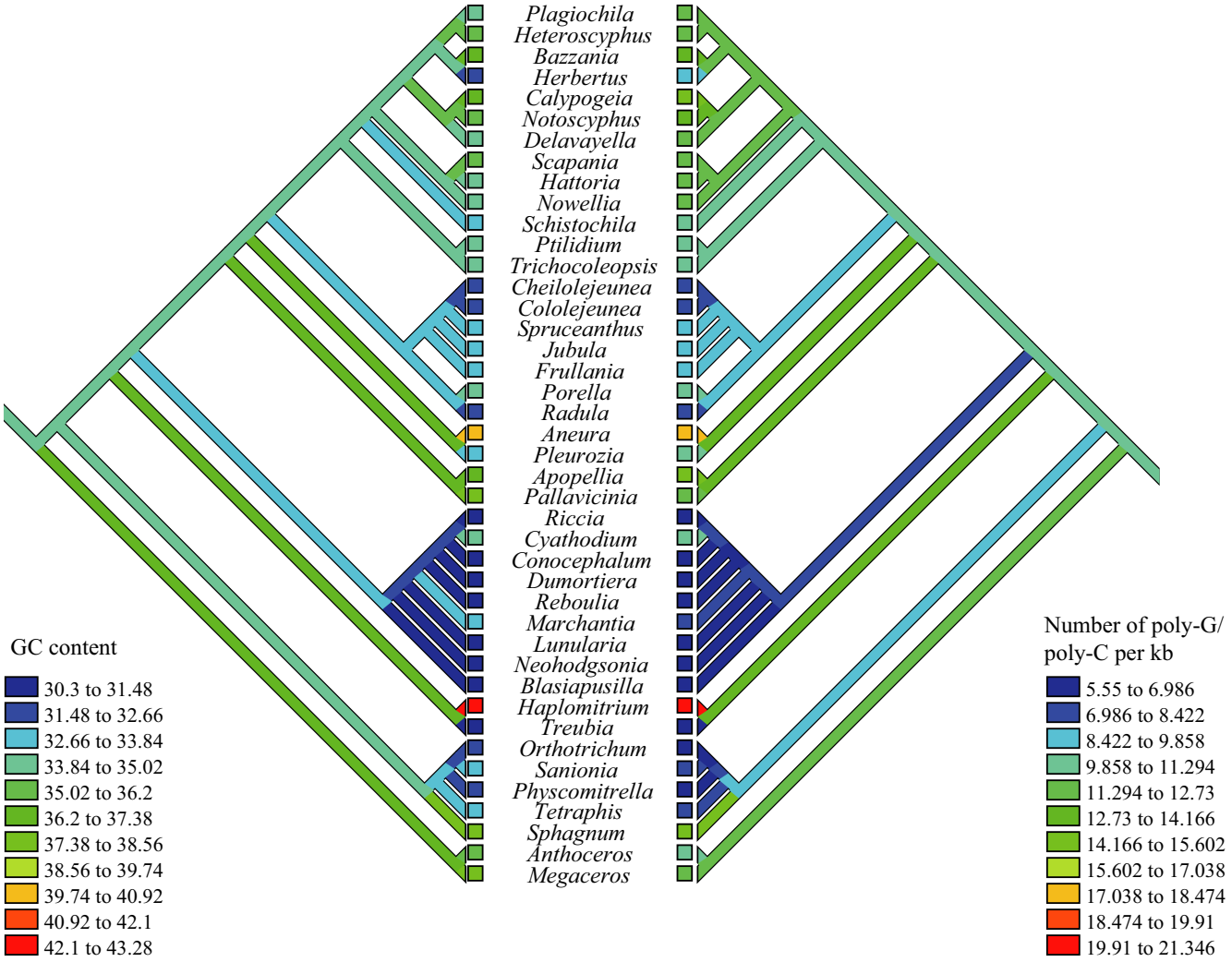

Fig. 3. Optimization of the diversity of G/C tracts and GC content on the phylogeny of liverworts obtained using the recoded nucleotide sequence dataset and maximum parsimony (MP analysis. Ancestral states of these two characters were reconstructed using the MP method.

conditions) on the rates of molecular evolution in liverworts, although considerable efforts have been made to reveal the heterogeneity of substitution rates in this group (Quandt and Stech, 2005; Villarreal et al., 2016).

Recent comparative studies have put forward several hypotheses to explain long branches as a consequence of enhanced mutation rates caused by such factors as short generation time, small population size, fast metabolic rates and relaxed natural selections (Bromham et al., 2015). In a recent study, Villarreal et al. (2016) suggested that in *Cyathodium* a postulated higher mutation rate was related to its short generation time, which may simply increase the replication mutations (Eyre-Walker and Gaut, 1997), despite a lack of empirical data. However, changes in the generation time alone are seemingly insufficient to explain all cases of mutation rate heterogeneity observed so far, and other factors such as effective population size, growth rate, changes in molecular mechanisms or environmental conditions also should be contemplated. For example, theoretically well-grounded arguments predict a dramatic reduction of effective population size in haploids with the onset of self-fertilization (Szövényi et al., 2017), and in the present study not all liverworts that are considered to be short-lived, such as epiphyllous taxa (e.g. *Cololejeunea*; Schuster, 1988; Zartman and Nascimento, 2006), have long branches. This hypothesis requires further investigation using an increased sampling of plastid genomes of bryophytes.

Furthermore, after careful examination of substitution sites in our samples, with particular focus on sister lineages showing considerable variations in estimated branch lengths, we found that poly-G/poly-C tracts (G/C tracts) accumulated rapidly in the plastome of taxa nested in clades with longer branches,

such as *Haplomitrium*, *Cyathodium* and *Aneura*, although this correlation was not always robust (Fig. 3, Table S3). The G/C tracts were assumed to be a consequence of DNA replication slippage, a process that could be affected by multiple extrinsic (e.g. selection, gene drift, mating preference and generation time) and/or intrinsic (e.g. genome structure) factors (Aguilera and García-Muse, 2013). Unravelling the mechanisms that lead to the accumulation of long branches calls for multi-comparative studies with non-target variants fixed and for the application of a larger dataset, both in terms of samples and sequences. Conversely, some branches in Marchantiales and Jungermanniales were notably short, coincident with a small number of G/C tracts and low level of GC content (Fig. 3, Table S3). These results may be interpreted from the perspective of low mutation rates (e.g. Marchantiales; Villarreal et al., 2016) or rapid diversification (e.g. Jungermanniales; Feldberg et al., 2014).

Although correlations between branch length and diversity of G/C tracts remain inconsistent, possibly due to other complex factors such as phylogenetic conservation, the link between diversity of G/C tracts and GC content is supported in both the optimization analyses and Felsenstein independent contrast (Fig. 3). These findings are in agreement with previous molecular studies (Northam et al., 2006; Kiktev et al., 2018) exploring the complex impacts of GC content on the DNA polymerase slippage, which often lead to rapid accumulation of homopolymeric poly-(dN), and subsequently to accelerating mutation rates. In addition, the diversity of G/C tracts and GC content shows informative taxonomic values, especially at the order level. For example, Marchantiopsida are well characterized by GC contents <32.70% and GC tracts <8.4 per kb, whereas Jungermanniopsida are characterized by GC contents >32.70% and GC tracts >8.4 per kb. The highest GC contents and GC tracts were found in the Haplomitropsida—the sister lineage to all other extant liverworts. Denser taxon samplings are needed to confirm all of these trends and taxonomic information about the GC content and GC tracts.

Accessing the effects of sequence length, codon heterogeneity, sample size and reconstructed methods on phylogenetic inference

By using chloroplast genome data, we obtained high posterior confidence values for all suprafamily relationships left unresolved by multi-gene analyses (e.g. the position of Ptilidiales in leafy liverworts, the early branching order within Marchantiales, and the relationships among and within three major subclades in Jungermanniales; Figs 1 and 2), although some of them remained weakly supported in the MP and ML analyses. These results are consistent with the views of Philippe et al. (2011) and Philippe and Roure (2011) that adding sequences is not sufficient on its own to resolve phylogenetic incongruences. Compared to sequence length and methods of analysis (MP, ML and BI), codon heterogeneity appears to have greater effects on the backbone phylogeny of liverworts, as evidenced by inconsistencies between phylogenetic hypotheses obtained using nucleotide and amino acid/recoded nucleotide alignments. Similar conclusions also were reached by previous studies (Cox et al., 2014; Li et al., 2014; de Sousa et al., 2018). Thus, it is essential to re-evaluate phylogenetic affinities using protein-coding genes while taking composition biases in account. Nevertheless, codon heterogeneity is difficult to quantify, as it varies across sequences, genes, genomes and taxa in terms of the uneven translation selections (Sharp et al., 2005; Rispe et al., 2007; Koonin and Wolf, 2010).

It is well known that adding more taxa can increase the ratio of phylogenetic to nonphylogenetic signals, allowing for better detection of multiple substitutions (Hendy and Penny, 1989; Baurain et al., 2007; Philippe et al., 2011). This is why recent phylogenetic studies have obtained more robust and rigorous phylogenies by using large-scale taxon-character datasets (Sims et al., 2009; Ruhfel et al., 2014; Gitzendanner et al., 2018). Therefore, remaining problematic relationships in liverworts, especially at low rank, such as the generic level relationships in the largest leafy family Lejeuneaceae (Wilson et al., 2007) may be resolved by increasing taxon sampling, together with appropriate analyzed methods and inclusion of more sequences.

Acknowledgements

We are grateful to the herbaria of Kunming Institute of Botany (KUN) and Natural History Museum London (BM) for providing DNA materials. This work is financially supported by National Natural Science Foundation of China (31500182 and 41571049), Natural Environment Research Council (NE/N002067/1) and Large-scale scientific facilities of the Chinese Academy of Sciences (2017-LSF-GBOWS-02).

References

Aguilera, A. and García-Muse, T., 2013. Causes of genome instability. Annu. Rev. Genet. 47, 1–32.

Baurain, D., Brinkmann, H. and Philippe, H., 2007. Lack of resolution in the animal phylogeny: closely spaced cladogeneses or undetected systematic errors? Mol. Biol. Evol. 24, 6–9.

Boyer, J.C., Yamada, N.A., Roques, C.N., Hatch, S.B., Riess, K. and Farber, R.A., 2002. Sequence dependent instability of mononucleotide microsatellites in cultured mismatch repair proficient and deficient mammalian cells. Hum. Mol. Genet. 11, 707–713.

Bromham, L., Hua, X., Lanfear, R. and Cowman, P.F., 2015. Exploring the relationships between mutation rates, life history, genome size, environment, and species richness in flowering plants. Am. Nat. 185, 507–524.

Cooper, E.D., Henwood, M.J. and Brown, E.A., 2012. Are the liverworts really that old?' Cretaceous origins and Cenozoic diversifications in Lepidoziaceae reflect a recurrent theme in liverwort evolution. Biol. J. Linn. Soc. 107, 425–441.

Cox, C.J., 2018. Land plant molecular phylogenetics: a review with comments on evaluating incongruence among phylogenies. Crit. Rev. Plant Sci. 24, 1–15.

Cox, C.J., Li, B., Foster, P.G., Embley, T.M. and Civan, P., 2014. Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. 63, 272–279.

Crandall-Stotler, B., Stotler, R. and Long, D.G., 2009. Phylogeny and classification of the Marchantiophyta. Edinb. J. Bot. 66, 155–198.

Davis, C., 2004. A molecular phylogeny of leafy liverworts (Jungermanniidae, Marchantiophyta). In: Goffinet, B., Hollowell, V., Magill, R. (Eds.), Molecular Systematics of Bryophytes. Monographs in Systematic Botany 98. Missouri Botanical Garden, St Louis, MO, pp. 61–86.

Degnan, J.H. and Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol. Evol. 24, 332–340.

Dunn, C.W., Hejnol, A., Matus, D.Q., Pang, K., Browne, W.E., Smith, S.A., Seaver, E., Rouse, G.W., Obst, M., Edgecombe, G.D. *et al*., 2008. Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452, 745.

Eyre-Walker, A. and Gaut, B.S., 1997. Correlated rates of synonymous site evolution across plant genomes. Mol. Biol. Evol. 14, 455–460.

Feldberg, K., Schneider, H., Stadler, T., Schaefer-Verwimp, A., Schmidt, A.R. and Heinrichs, J., 2014. Epiphytic leafy liverworts diversified in angiosperm-dominated forests. Sci. Rep. 4, 5974.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

Felsenstein, J., 1985. Phylogenies and the comparative Method. Am. Nat. 125, 1–15.

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.

Forrest, L.L., Wickett, N.J., Cox, C.J. and Goffinet, B., 2011. Deep sequencing of *Ptilidium* (Ptilidiaceae) suggests evolutionary stasis in liverwort plastid genome structure. Plant Ecol. Evol. 144, 29–43.

Foster, P.G., Jermiin, L.S. and Hickey, D.A., 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. 44, 282–288.

Giribet, G., 2016. Genomics and the animal tree of life: conflicts and future prospects. Zool. Scr. 45, 14–21.

Gitzendanner, M.A., Soltis, P.S., Wong, G.K.S., Ruhfel, B.R. and Soltis, D.E., 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. Am. J. Bot. 105, 291–301.

Grosche, C., Funk, H.T., Maier, U.G. and Zauner, S., 2012. The chloroplast genome of *Pellia endiviifolia*: gene content, RNA-editing pattern, and the origin of chloroplast editing. Genome Biol. Evol. 4, 1349–1357.

Heinrichs, J., Gradstein, S., Wilson, R. and Schneider, H., 2005. Towards a natural classification of liverworts (Marchantiopyta) based on the chloroplast gene *rbc*L. Cryptogam. Bryol. 26, 131–150.

Heinrichs, J., Hentschel, J., Wilson, R., Feldberg, K. and Schneider, H., 2007. Evolution of leafy liverworts (Jungermanniidae, Marchantiophyta): estimating divergence times from chloroplast DNA sequences using penalized likelihood with integrated fossil evidence. Taxon 56, 31–44.

Hendry, T.A., Wang, B., Yang, Y., Davis, E.C., Braggins, J.E., Schuster, R.M. and Qiu, Y.L., 2007. Evaluating phylogenetic positions of four liverworts from New Zealand, *Neogrollea notabilis*, *Jackiella curvata*, *Goebelobryum unguiculatum* and *Herzogianthus vaginatus*, using three chloroplast genes. Bryologist 110, 738–751.

Hendy, M.D. and Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Biol. 38, 297–309.

He-Nygrén, X., Juslén, A., Ahonen, I., Glenny, D. and Piippo, S., 2006. Illuminating the evolutionary history of liverworts (Marchantiophyta) – Towards a natural classification. Cladistics 22, 1–31.

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.

Kiktev, D.A., Sheng, Z.W., Lobachev, K.S. and Petes, T.D., 2018. GC content elevates mutation and recombination rates in the yeast *Saccharomyces cerevisiae*. Proc. Natl Acad. Sci. USA 115, E7109–E7118.

Koonin, E.V. and Wolf, Y.I., 2010. Constraints and plasticity in genome and molecular-phenome evolution. Nat. Rev. Genet. 11, 487–498.

Laenen, B., Machac, A., Gradstein, S.R., Shaw, B., Patiño, J., Désamoré, A., Goffinet, B., Cox, C.J., Shaw, A.J. and Vanderpoorten, A., 2016a. Increased diversification rates follow shifts to bisexuality in liverworts. New Phytol. 210, 1121–1129.

Laenen, B., Machac, A., Gradstein, S.R., Shaw, B., Patiño, J., Desamore, A., Goffinet, B., Cox, C.J., Shaw, A.J. and Vanderpoorten, A., 2016b. Geographical range in liverworts: does sex really matter? J. Biogeogr. 43, 627–635.

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.

Lanfear, R., Welch, J.J. and Bromham, L., 2010. Watching the clock: studying variation in rates of molecular evolution between species. Trends Ecol. Evol. 25, 495–503.

Li, B., Lopes, J.S., Foster, P.G., Embley, T.M. and Cox, C.J., 2014. Compositional biases among synonymous substitutions cause conflict between gene and protein trees for plastid origins. Mol. Biol. Evol. 31, 1697–1709.

Liu, T.L. and Bundschuh, R., 2005. A model for codon position bias in RNA editing. Phys. Rev. Lett. arXiv:physics/0505001v1.

Liu, Y., Lia, Y., Wang, W., Chen, Z.D., Davis, E.C. and Qiu, Y.L., 2008. Phylogenetic relationships of two endemic genera from East Asia: *Trichocoleopsis* and *Neotrichocolea* (Hepaticae). Ann. Missouri Bot. Gard. 95, 459–470.

Maddison, W.P. and Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis. Ver. 3.51. http://www.mesquite project.org.

Moore, M.J., Bell, C.D., Soltis, P.S. and Soltis, D.E., 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proc. Natl Acad. Sci. USA 104, 19363–19368.

Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z.H., Schneider, H. and Donoghue, P.C.J., 2018. The timescale of early land plant evolution. Proc. Natl Acad. Sci. USA 115, E2274–E2283.

Myszczyński, K., Bączkiewicz, A., Buczkowska, K., Ślipiko, M., Szczecińska, M. and Sawicki, J., 2017. The extraordinary variation of the organellar genomes of the *Aneura pinguis* revealed advanced cryptic speciation of the early land plants. Sci. Rep. 7, 904.

Nishiyama, T., Wolf, P.G., Kugita, M., Sinclair, R.B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K. *et al*., 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol. Biol. Evol. 21, 1813–1819.

Northam, M.R., Garg, P., Baitin, D.M., Burgers, P.M.J. and Shcherbakova, P.V., 2006. A novel function of DNA polymerase zeta regulated by PCNA. EMBO J. 25, 4316–4325.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, Y., Sano, S., Umesono, K., Siki, Y., Takeuchi, M., Chang, Z. *et al*., 1986.

Chloroplast gene organization decided from complete sequence of liverwort *Marchantia polymorphs* chloroplast DNA. Nature 322, 572–574.
Philippe, H. and Roure, B., 2011. Difficult phylogenetic questions: more data, maybe; better methods, certainly. BMC Biol. 9, 91.
Philippe, H., Delsuc, F., Brinkmann, H. and Lartillot, N., 2005. Phylogenomics. Annu. Rev. Ecol. Evol. Syst. 36, 541–562.
Philippe, H., Brinkmann, H., Lavrov, D.V., Littlewood, D.T.J., Manuel, M., Worheide, G. and Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS Biol. 9, e1000602.
Puttick, M.N., Morris, J.L., Williams, T.A., Cox, C.J., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Schneider, H., Pisani, D. *et al.* , 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. Curr. Biol. 28, 733–745.
Qiu, Y.L., Cho, Y., Cox, J.C. and Palmer, J.D., 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394, 671–674.
Qiu, Y.L., Li, L.B., Wang, B., Chen, Z.D., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J. *et al.* , 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl Acad. Sci. USA 103, 15511–15516.
Quandt, D. and Stech, M., 2005. Molecular evolution of the *trn*L (UAA) intron in bryophytes. Mol. Phylogenet. Evol. 36, 429–443.
Regier, J.C., Shultz, J.W., Zwick, A., Hussey, A., Ball, B., Wetzer, R., Martin, J.W. and Cunningham, C.W., 2010. Arthropod relationships revealed by phylogenomic analysis of nuclear protein-coding sequences. Nature 463, 1079–1083.
Rispe, C., Legeai, F., Gauthier, J.P. and Tagu, D., 2007. Strong heterogeneity in nucleotidic composition and codon bias in the pea aphid (*Acyrthosiphon pisum*) shown by EST-based coding genome reconstruction. J. Mol. Evol. 65, 413–424.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, 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.
Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E. and Burleigh, J.G., 2014. From algae to angiosperms–inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. 14, 23.
Schuster, R.M., 1988. Ecology, reproductive biology and dispersal of Hepaticae in the tropics. J. Hattori Bot. Lab. 64, 237–269.
Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F. and Sockett, R.E., 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res. 33, 1141–1153.
Sims, G.E., Jun, S.-R., Wu, G.A. and Kim, S.-H., 2009. Whole-genome phylogeny of mammals: evolutionary information in genic and nongenetic regions. Proc. Natl Acad. Sci. USA 106, 17077–17082.
Söderström, L., Hagborg, A., von Konrat, M., Bartholomew-Began, S., Bell, D., Briscoe, L., Brown, E., Cargill, D.C., Costa, D.P., Crandall-Stotler, B.J. *et al.*, 2016. World checklist of hornworts and liverworts. Phytokeys 59, 1–828.
de Sousa, F., Foster, P.G., Donoghue, P.C.J., Schneider, H. and Cox, C.J., 2018. Nuclear protein phylogenetics support the monopoly of the three bryophyte groups (Bryophyta Schimp.). New Phytol. 222, 565–575.
Szövényi, P., Ullrich, K.K., Rensing, S.A., Lang, D., van Gessel, N., Stenoien, H.K., Conti, E. and Reski, R., 2017. Selfing in haploid plant and efficacy of selection: codon usage bias in the model moss *Physcomitrella patens*. Genome Biol. Evol. 9, 1528–1546.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729.
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.
Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A. *et al.* , 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl Acad. Sci. USA 111, E4859–E4868.
Wilson, R., Gradstein, S.R., Schneider, H. and Heinrichs, J., 2007. Unravelling the phylogeny of Lejeuneaceae (Jungermanniopsida): evidence for four main lineages. Mol. Phylogenet. Evol. 43, 270–282.
Yang, J.-B., Li, D.-Z. and Li, H.-T., 2014. Highly effective sequencing whole chloroplast genomes of angiosperms by nine novel universal primer pairs. Mol. Ecol. Resour. 14, 1024–1031.
Zartman, C.E. and Nascimento, H.E.M., 2006. Are habitat-tracking metacommunities dispersal limited? Inferences from abundance-occupancy patterns of epiphylls in Amazonian forest fragments Biol. Conserv. 127, 46–54.

Supporting Information

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

Fig. S1. A simplified phylogeny of liverworts generated for 15 orders following a classification proposed by Söderström et al. (2016).

Table S1 Forty-two bryophyte samples used in the phylogenetic analyses. The accession numbers for previously published genomes and voucher information of newly generated ones in the present study are listed.

Table S2 The branch length and diversity of G/C tracts for eleven lineage pairs selected based on phylogenies of liverworts obtained using NSA, ASA and RSA, respectively.

Table S3 Information of the GC content and diversity of G/C tracts of all samples used in the phylogeny.