American Journal of Botany

RESEARCH ARTICL

Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution

David Bell[1,2,3], Qianshi Lin[1,2], Wesley K. Gerelle[1,2], Steve Joya[1], Ying Chang[1,4], Z. Nathan Taylor[5], Carl J. Rothfels[6], Anders Larsson[7], Juan Carlos Villarreal[8,9], Fay-Wei Li[10,11], Lisa Pokorny[12,13], Péter Szövényi[14], Barbara Crandall-Stotler[15], Lisa DeGironimo[16], Sandra K. Floyd[17], David J. Beerling[18], Michael K. Deyholos[19], Matt von Konrat[20], Shona Ellis[1], A. Jonathan Shaw[21], Tao Chen[22], Gane K.-S. Wong[23,24,25], Dennis W. Stevenson[26], Jeffrey D. Palmer[5], and Sean W. Graham[1,2,27]

Manuscript received 4 June 2019; revision accepted 4 November 2019.

[1] Department of Botany, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

[2] UBC Botanical Garden and Centre for Plant Research, University of British Columbia, 6804 Marine Drive SW, Vancouver, British Columbia V6T 1Z4, Canada

[3] Royal Botanic Garden, 20A Inverleith Row, Edinburgh EH3 5LR, UK

[4] Department of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon 97331, USA

[5] Department of Biology, Indiana University, Bloomington, Indiana 47405, USA

[6] University Herbarium and Department of Integrative Biology, University of California Berkeley, Berkeley, California 94702, USA

[7] Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden

[8] Department of Biology, Université Laval, Québec G1V 0A6, Canada

[9] Smithsonian Tropical Research Institute, Panama City, Panama

[10] Boyce Thompson Institute, Ithaca, New York 14853, USA

[11] Plant Biology Section, Cornell University, Ithaca, New York 14853, USA

[12] Royal Botanic Gardens, Kew, Richmond TW9 3DS, Surrey, UK

[13] Centre for Plant Biotechnology and Genomics (CBGP, UPM-INIA), 28223, Pozuelo de Alarcón (Madrid), Spain

[14] Department of Systematic and Evolutionary Botany, University of Zurich, Zollikerstrasse 107, 8008, Zurich, Switzerland

[15] School of Biological Sciences, Southern Illinois University, Carbondale, Illinois 62901, USA

[16] Department of Biology, College of Arts and Science, New York University, New York, New York 10003, USA

[17] School of Biological Sciences, Monash University, Melbourne, Victoria 3800, Australia

[18] Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK

[19] Department of Biology, University of British Columbia, Kelowna, British Columbia V1V 1V7, Canada

[20] Field Museum of Natural History, Chicago, Illinois 60605, USA

[21] Department of Biology, Duke University, Durham, North Carolina 27708, USA

PREMISE: Phylogenetic trees of bryophytes provide important evolutionary context for land plants. However, published inferences of overall embryophyte relationships vary considerably. We performed phylogenomic analyses of bryophytes and relatives using both mitochondrial and plastid gene sets, and investigated bryophyte plastome evolution.

METHODS: We employed diverse likelihood-based analyses to infer large-scale bryophyte phylogeny for mitochondrial and plastid data sets. We tested for changes in purifying selection in plastid genes of a mycoheterotrophic liverwort (*Aneura mirabilis*) and a putatively mycoheterotrophic moss (*Buxbaumia*), and compared 15 bryophyte plastomes for major structural rearrangements.

RESULTS: Overall land-plant relationships conflict across analyses, generally weakly. However, an underlying (unrooted) four-taxon tree is consistent across most analyses and published studies. Despite gene coverage patchiness, relationships within mosses, liverworts, and hornworts are largely congruent with previous studies, with plastid results generally better supported. Exclusion of RNA edit sites restores cases of unexpected non-monophyly to monophyly for *Takakia* and two hornwort genera. Relaxed purifying selection affects multiple plastid genes in mycoheterotrophic *Aneura* but not *Buxbaumia*. Plastid genome structure is nearly invariant across bryophytes, but the *tuf*A locus, presumed lost in embryophytes, is unexpectedly retained in several mosses.

CONCLUSIONS: A common unrooted tree underlies embryophyte phylogeny, [(liverworts, mosses), (hornworts, vascular plants)]; rooting inconsistency across studies likely reflects substantial distance to algal outgroups. Analyses combining genomic and transcriptomic data may be misled locally for heavily RNA-edited taxa. The *Buxbaumia* plastome lacks hallmarks of relaxed selection found in mycoheterotrophic *Aneura*. Autotrophic bryophyte plastomes, including *Buxbaumia*, hardly vary in overall structure.

KEY WORDS Anthocerotophyta (hornworts); Bryophyta (mosses); embryophyte relationships; long-branch outgroups; Marchantiophyta (liverworts); mycoheterotrophic bryophytes; organellar evolution; phylogenetic incongruence; RNA editing; tree rooting.

[22] Shenzhen Fairy Lake Botanical Garden, Chinese Academy of Sciences, Shenzhen, Guangdong 518004, China

[23] Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada

[24] Department of Medicine, University of Alberta, Edmonton, Alberta T6G 2E1, Canada

[25] BGI-Shenzhen, Beishan Industrial Zone, Yantian District, Shenzhen 518083, China

[26] New York Botanical Garden, Bronx, New York 10458, USA

[27] Author for correspondence (e-mail: swgraham@interchange.ubc.ca)

Citation: Bell, D., Q. Lin, W. K. Gerelle, S. Joya, Y. Chang, Z. N. Taylor, C. J. Rothfels, et al. 2019. Organellomic data sets confirm a cryptic consensus on (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution. *American Journal of Botany* 107(1): 1–25.

doi:10.1002/ajb2.1397

INTRODUCTION

Despite extensive investigation, phylogenetic relationships among the four major groups of land plants (embryophytes)—vascular plants, mosses, liverworts, and hornworts (the latter three comprising the bryophytes)—are still unsettled (e.g., Qiu et al., 2006, 2007; Cox et al., 2014; Wickett et al., 2014; Gitzendanner et al., 2018; de Sousa et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). The bryophytes are distinguishable from vascular plants in having a diploid sporophyte generation attached to (and dependent on) the persistent haploid gametophyte. Vascular plants, by contrast, have a dominant sporophyte generation, and reduced and often dependent gametophytes. The three major bryophyte lineages differ substantially from each other, for example in gametophyte and sporophyte structure, embryology, mechanisms of sporangium dehiscence, and methods of spore dispersal (e.g., Campbell, 1895; Schofield, 1985; Crum, 2001; Crandall-Stotler et al., 2009; Goffinet et al., 2009; Renzaglia et al., 2009; Vanderpoorten and Goffinet, 2009; Ligrone et al., 2012). Despite their considerable morphological diversity, these three lineages were long treated together at the division (phylum) rank (e.g., Campbell, 1895; Smith, 1955; Schofield, 1985), although even by the mid-nineteenth century authors had begun to challenge the naturalness (monophyly) of bryophytes (reviewed in Crandall-Stotler, 1980). A growing body of phylogenetic evidence from morphological and molecular studies now supports the view that extant bryophytes comprise three distinct branches of land plants, with extant vascular plants (tracheophytes) representing the fourth major land-plant lineage (Mishler and Churchill, 1984; Mishler et al., 1994; Hedderson et al., 1996; Kenrick and Crane, 1997a; Lewis et al., 1997; Qiu et al., 1998; Nishiyama et al., 2004; Forrest et al., 2006; Gao et al., 2010; Karol et al., 2010; Chang and Graham, 2011; Liu et al., 2014a; Ruhfel et al., 2014; Wickett et al., 2014; Puttick et al., 2018; for an alternative viewpoint, see de Sousa et al., 2019).

Early phylogenetic studies based on one or only a few loci reported conflicting results regarding relationships among the four extant land-plant lineages (reviewed in Qiu, 2008). Later studies that considered additional sequence data, focusing mainly on the plastid genome (plastome) and using more refined methods of phylogenetic inference, appeared to be converging on an overall topology that placed liverworts as the sister group of all other land plants, and hornworts sister to vascular plants (e.g., Forrest et al., 2006; Qiu et al., 2006, 2007; Chang and Graham, 2011; Magallón et al., 2013). However, additional studies using larger data sets from plastid or mitochondrial genomes (e.g., for plastids: Gao et al., 2010; Karol et al., 2010; Civáň et al., 2014; Cox et al., 2014; Ruhfel et al., 2014; Gitzendanner et al., 2018; for mitochondria: Turmel et al., 2013; Liu et al., 2014a), or from the nuclear genome (Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019), recovered alternative and conflicting topologies, both among and within studies. Multiple studies have supported the monophyly of extant bryophytes in at least some analyses (Civáň et al., 2014; Cox et al., 2014; Wickett et al., 2014; Gitzendanner et al., 2018; Puttick et al., 2018; de Sousa et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019).

Despite these contradictory inferences, for the vast majority of molecular phylogenetic and phylogenomic studies the inferred relationships among the four major groups of land plants have an underlying and largely unremarked-upon consistency: they correspond to a single unrooted four-taxon tree, [(liverworts, mosses),(hornworts, vascular plants)] (Tree 1 in Fig. 1; study details in Table 1; see also Cox, 2018: fig. 1), with relatively few studies recovering either of the other two possible alternative unrooted relationships (Trees 2 and 3 in Fig. 1). If this is the correct unrooted tree, it is likely that most of the ambiguity about the relationships among the four extant land-plant lineages is due to the substantial evolutionary distance (long branches) between land plants and their closest streptophyte algal relatives (e.g., Zygnematophyceae [Wodniok et al., 2011; Wickett et al., 2014]; Coleochaetales [Finet et al., 2010]; for examples of other long-branch situations in plant phylogeny, see Graham et al. 2002, Murdock 2008, Graham and Iles 2009, and Rothfels et al. 2012). Long-branch attraction (Felsenstein, 1978; Hendy and Penny, 1989) and other types of systematic bias (e.g., compositional biases; Philippe et al., 2011) can be hard to deal with, even by using dense taxon sampling (e.g., Hedtke et al., 2006; Heath et al., 2008) or complex substitution models (e.g., Liu et al., 2014a; Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2019). Increasing the amount of data per taxon helps address the effects of stochastic error on short branches (e.g., Graham et al., 1998; Rothfels et al., 2015), but genome-scale samplings may also make systematic errors related to inadequate models more apparent (Philippe et al., 2011). Despite paying careful attention to substitutional dynamics (e.g., Cox et al., 2014; de Sousa et al., 2019), it is possible that no current nucleotide or amino-acid substitution model captures DNA or protein evolution

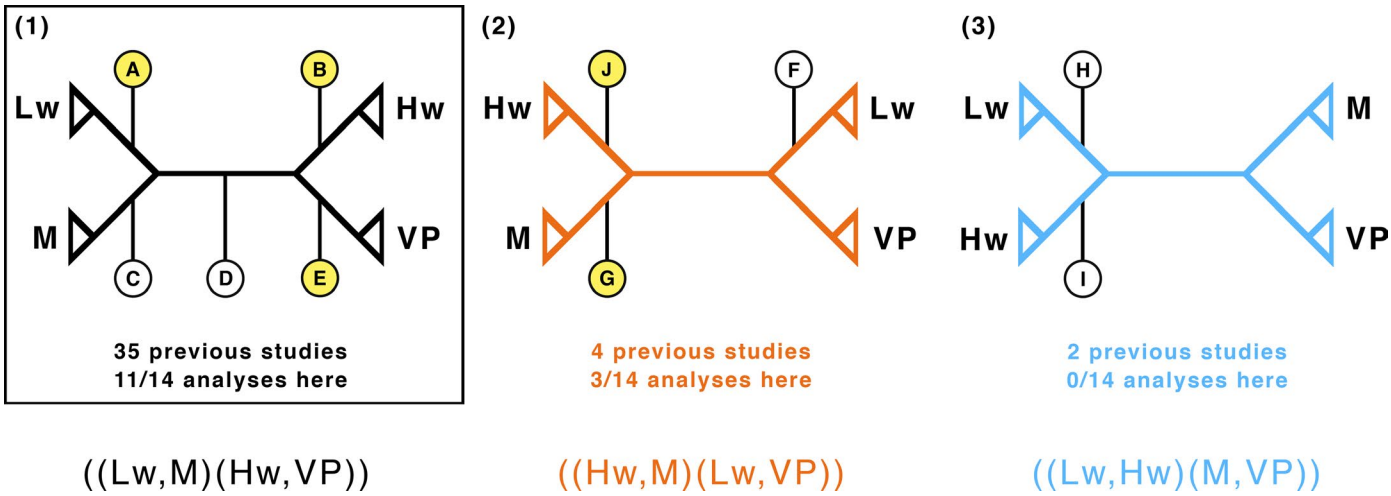

FIGURE 1. Four-taxon trees showing the three possible unrooted relationships of the four major land-plant lineages (Lw: liverworts; M: mosses; Hw: hornworts; VP: vascular plants). Tree 1 (black) shows the most frequently recovered four-taxon tree in major analyses from previously published studies and this study. Trees 2 (orange) and 3 (blue) show the two possible alternative arrangements of land-plant relationships. A–I represent the rooting points obtained in major analyses from previous studies (see Table 1 for details). Rooting points shaded yellow (A, B, E, G, J) represent those inferred here from likelihood analyses of plastid (pt) or mitochondrial (mt) nucleotide data with all positions included (DNA), nucleotide data with third codon positions excluded (1+2), and amino-acid data (AA); A = pt DNA, B = pt 1+2 and mt AA, E = pt AA, G = mt DNA, and J = mt 1+2. The number of published studies recovering each unrooted topology is noted below each tree, along with the number of times each was recovered in the 14 analyses in this study.

sufficiently well to fully address this inference challenge (e.g., see the strongly supported conflicts summarized in Wickett et al., 2014: fig. 4). The scope of genomic data sets is now rapidly expanding (e.g., Wickett et al., 2014; Cheng et al., 2018), but because different data sets and analytical approaches conflict in what they infer, and because it is not clear that any current substitution models fully account for the complexity of gene evolution across 470–515 Ma of land-plant evolution (Morris et al., 2018), and beyond to algal relatives, the relationships among the four major land plant groups may still be considered unclear (cf. de Sousa et al., 2019).

We examine overall bryophyte relationships among and within major lineages, using plastid and mitochondrial data, including new genome-skim-based data sets for 25 taxa, organellar genomic data retrieved from GenBank (e.g., Liu et al., 2014a, b; Shaw et al., 2016; Park et al., 2018), transcriptome-based data from the One Thousand Plants project (1KP; One Thousand Plant Transcriptomes Initiative, 2019), and collections of Sanger sequences from key lineages currently lacking substantial genomic data (e.g., Cox et al., 2004; Forrest et al., 2006; Duff et al., 2007; Chang and Graham, 2011). Our taxon sampling includes a phylogenetically broader representation of major bryophyte lineages than most recent studies, except for Gitzendanner et al. (2018), and advances on the latter study by the inclusion of a parallel comparison of mitochondrial genes retrieved from a largely parallel set of taxa. We include an examination of the possible effect of different DNA and amino-acid substitutional schemes on phylogenetic inference, such as partitioned likelihood models (e.g., Lanfear et al., 2012, 2016) that take account of different substitutional dynamics among functionally or evolutionarily distinctive subsets of data.

We also test for possible changes in selection pressure in the plastomes of a known heterotrophic liverwort (*Aneura mirabilis*; Wickett et al., 2008) and a purportedly heterotrophic moss (*Buxbaumia aphylla*; Leake, 1994). Such changes are a major hallmark of heterotrophy (e.g., Lam et al., 2015), so their presence or absence may help support or refute the heterotrophic status of *Buxbaumia*. Heterotrophic taxa can have highly unusual DNA substitutional dynamics and substantial plastome reduction through gene loss, which may influence their suitability for phylogenetic inference. However, Lam et al. (2015, 2016, 2018) demonstrated that data from even highly reduced and rapidly evolving plastid genes or plastomes can be used to place heterotrophic plants with confidence. We were particularly interested in characterizing the *Buxbaumia* plastome in light of its purported heterotrophic status, given the placement of this small genus as sister to all other members of Bryopsida, which is by far the most species-rich class of mosses (e.g., Qiu et al., 2006, 2007; Chang and Graham, 2011, 2014; Liu et al., 2014a, 2019; Gitzendanner et al., 2018).

Because gene recovery is sometimes patchy across taxa (e.g., due to incomplete gene recovery from transcriptomes), our study effectively also incorporates a practical examination of how well data sets with incomplete gene occupancy perform. One expectation is that including incompletely sequenced taxa would tend to increase the accuracy of phylogenetic inference when taxon sampling is otherwise incomplete (e.g., Wiens, 2005, 2006; Burleigh et al., 2009; Wiens and Tiu, 2012; Jiang et al., 2014). However, we also know that combining patchily sampled matrices with long branches can be problematic (Wiens, 2005, 2006). Nonetheless, recent analyses considering very rapidly evolving but highly reduced plastomes suggest that model-based analyses of organellar phylogenomic inference can cope well in situations in which some taxa have exceptionally long branches coupled with limited or variable gene sampling (Lam et al., 2018).

The availability of multiple completed plastomes across bryophytes, including for two taxa newly completed here (*Buxbaumia aphylla* and *Diphyscium foliosum*), also permits comparison of genome structural evolution across bryophytes. This allows us to considerably extend the observation that plastomes of bryophytes

TABLE 1. Summary of underlying four-taxon land-plant trees and land-plant root points implied in previously published molecular phylogenetic studies, including details of the genome (plastid, "pt"; mitochondrial, "mt"; nuclear, "nr"), data type (full nucleotide data, "DNA"; selected codon positions only, "1+2", "1+3", "2+3"; RY coded, "RY"; amino acid, "AA"), type of analysis (parsimony, "MP"; maximum likelihood, "ML"; Bayesian, "BI"; neighbor-joining, "NJ"; coalescent, "ASTRAL"), number of loci analyzed, number of bryophytes/total number of taxa included, and the level of support (weak, moderate, or strong) for each. Support values are generally bootstrap (BS) for ML and MP, and posterior probability (PP) for BI (weak: <70% BS, <0.95 PP; strong: >95% BS, >0.99 PP). Studies are organized according to implied unrooted trees in Fig. 1. Lw: liverworts; M: mosses; Hw: hornworts; VP: vascular plants.

Data type	Analysis	Number of loci	Number of taxa	Support	Reference
A: Outgroups,(Lw(M(Hw,VP))) = unrooted Tree 1					
pt 1+2	MP	51	3/20	Weak	Nishiyama et al., 2004
pt DNA, 1+2, AA	ML, MP	73	3/20	Weak–strong	Wolf et al., 2005
pt DNA, 1+3, 2+3	ML, MP	67	3/36	Strong	Qiu et al., 2006
pt DNA	ML	79	5/42	Moderate	Gao et al., 2010
pt DNA	ML, BI	49	4/43	Weak	Karol et al., 2010
mt AA	ML, BI	40	7/25	Weak	Turmel et al., 2013
pt DNA	ML	72	5/30	Weak	Zhong et al., 2013
pt DNA	ML, BI	49	4/43	Weak–strong	Cox et al., 2014
pt DNA	ML, MP, BI	89	5/35	Weak–strong	Kim et al., 2014
mt AA	ML, BI	41	20/60	Moderate–strong	Liu et al., 2014a
mt DNA	BI	41	20/60	Weak	Liu et al., 2014a
pt DNA, RY	ML	78	6/360	Weak–moderate	Ruhfel et al., 2014
pt DNA	ML	1	30/40	Weak	Lewis et al., 1997
pt DNA	ML	1	5/39	Weak	Delwiche et al., 2002
pt, mt, nr DNA	MP, BI	5	204/210	Weak	Forrest et al., 2006
pt, mt, nr DNA	ML, MP	6	96/193	Moderate–strong	Qiu et al., 2006
pt, mt, nr DNA	ML	7	95/192	Moderate	Qiu et al., 2007
pt DNA	ML	1	1088/13,533	N/A	Smith et al., 2009
pt DNA	ML	17	23/43	Strong	Chang and Graham, 2011
pt DNA	ML	5	8/81	Strong	Magallón et al., 2013
pt DNA	ML	6	96/193	Moderate	Cox et al., 2014
B: Outgroups,(Hw(VP(Lw,M))) = unrooted Tree 1					
nr 1+2	ML	674	19/103	Strong	Wickett et al., 2014
nr AA	ML, BI	148–852	19/103	Significant	Puttick et al., 2018
nr DNA, 1+2, AA	ML	620	19/103	Moderate–strong	de Sousa et al., 2019
nr DNA	ML, BI	100	10/26	Weak–strong	de Sousa et al., 2019
nr rDNA	MP	1	18/27	Weak [a]	Hedderson et al., 1996
mt rDNA	ML, MP	1	9/20	Weak	Duff and Nickrent, 1999
pt AA	ML	5	5/12	Moderate	Nishiyama and Kato, 1999
pt, mt, nr DNA	ML, MP	4	7/30	Weak–strong	Nickrent et al., 2000
nr, mt rDNA	MP	2	11/26	Weak	Renzaglia et al., 2000
nr 1+2, AA	ML, ASTRAL	410	74/1178	Weak–strong	One Thousand Plant Transcriptomes Initiative, 2019
C: Outgroups,(M(Lw(Hw,VP))) = unrooted Tree 1					
pt DNA	ML	72	5/30	Moderate	Zhong et al., 2013
pt AA	ML, BI	71	3/47	Weak	Lemieux et al., 2014
mt DNA	ML, BI	41	20/60	Moderate–strong	Liu et al., 2014a
nr DNA	BI	100	10/26	Strong	de Sousa et al., 2019
D: Outgroups,((Lw,M)(Hw,VP)) = unrooted Tree 1					
pt 1+2	ML, MP	67	3/36	Weak–strong	Qiu et al., 2006
pt 1+2	ML	56	3/33	Moderate	Turmel et al., 2006
pt DNA	ML, MP	45	3/45	Weak	Lemieux et al., 2007
pt 1+2	ML	79	5/42	Strong	Gao et al., 2010
pt DNA	ML, BI	49	4/43	Weak–strong	Karol et al., 2010
pt DNA	BI	72	5/30	Strong	Zhong et al., 2013
pt DNA	BI	49	4/43	Weak–strong	Cox et al., 2014
pt 1+2, AA	ML	78	6/360	Weak–moderate	Ruhfel et al., 2014
pt 1+2, AA	ML, BI	88	4/28	Weak–strong	Lemieux et al., 2016
nr AA	BI	148–852	19/103	Significant	Puttick et al., 2018
E: Outgroups,(VP(Hw(Lw,M))) = unrooted Tree 1					
pt AA	ML	51	3/20	Strong	Nishiyama et al., 2004
pt 1+2	NJ	57	3/18	Strong	Goremykin and Hellwig, 2005
pt AA	ML, MP	45	3/45	Weak–moderate	Lemieux et al., 2007
pt AA	ML, BI	49	4/43	Moderate–strong	Karol et al., 2010
pt AA	ML	43	5/16	Moderate–strong	Shanker et al., 2011
pt AA	BI	83	6/30	Weak	Civáň et al., 2014

(Continued)

TABLE 1. (Continued)

Data type	Analysis	Number of loci	Number of taxa	Support	Reference
pt AA	BI	49	4/43	Weak–strong	Cox et al., 2014
nr AA	ML	674	19/103	Moderate	Wickett et al., 2014
nr 1+2, AA	ASTRAL	424	19/103	Strong	Wickett et al., 2014
pt AA	ML	78	85/1879	Strong	Gitzendanner et al., 2018
nr AA	ML, BI	148–852	19/103	Significant	Puttick et al., 2018
nr 1+2, AA	ML	620	19/103	Weak–moderate	de Sousa et al., 2019
nr DNA, AA	BI	100	10/26	Strong	de Sousa et al., 2019
nr AA	ASTRAL	410	74/1178	Weak–strong	One Thousand Plant Transcriptomes Initiative, 2019
pt AA	ML	78	74/1178	Weak	One Thousand Plant Transcriptomes Initiative, 2019
F: Outgroups,(Lw(VP(Hw,M))) = unrooted Tree 2					
nr rDNA	MP	2	8/15	Weak [a]	Waters et al., 1992
mt rDNA	MP	1	9/27	Weak	Duff and Nickrent, 1999
pt, nr DNA	ML	5	186/699	Weak–moderate	Fiz-Palacios et al., 2011
G: Outgroups,(M(Hw(Lw,VP))) = unrooted Tree 2					
nr AA	BI	5	3/14	Strong	Floyd et al., 2006
H: Outgroups,(Lw(Hw(M,VP))) = unrooted Tree 3					
nr rDNA [b]	MP	2	9/59	Weak [a]	Mishler et al., 1994
nr rDNA [b]	MP	1	6/9	Weak [a]	Mishler et al., 1994
pt, mt, nr DNA	ML, BI	4	3/40	Moderate–strong	Karol et al., 2001
I: Outgroups,(Hw(Lw(M,VP))) = unrooted Tree 3					
nr rDNA [b]	MP	1	6/9	Weak [a]	Mishler et al., 1994

[a]Decay indices, <4 considered weak here.
[b]Molecular data combined with morphological data.

are nearly identical in gene order, gene content, and intron content (e.g., Mower and Vickrey, 2018).

Relationships within major bryophyte lineages

Diversity within each of the three major bryophyte lineages is well characterized morphologically, but the application of molecular phylogenetic methods has revolutionized traditional classifications based primarily on morphological traits (e.g., Crandall-Stotler et al., 2009; Goffinet et al., 2009; Renzaglia et al., 2009). In contrast to the conflicting relationships often recovered among the four major land-plant lineages, molecular phylogenetic results have been generally congruent with respect to relationships recovered within each bryophyte lineage, with a largely consistent "backbone" inferred in sufficiently well-sampled molecular analyses of the liverworts (e.g., Crandall-Stotler et al., 2005; Heinrichs et al., 2005; Forrest et al., 2006; He-Nygrén et al., 2006; Chang and Graham, 2011; Gitzendanner et al., 2018) and hornworts (e.g., Stech et al., 2003; Duff et al., 2004, 2007; Villarreal et al., 2010).

In the mosses, molecular analyses have also resolved many of the deep relationships, but there are still several critical nodes that remain contentious (e.g., Newton et al., 2000; Cox et al., 2004; Qiu et al., 2006; Chang and Graham, 2011, 2014; Liu et al., 2019). Of these, the relationships among Takakiopsida, Sphagnopsida, and the rest of the mosses (i.e., whether Takakiopsida is sister to all other mosses or forms a clade with Sphagnopsida) are perhaps the most intriguing, as conflicting topologies are often recovered with strong support (Qiu et al., 2006; Chang and Graham, 2011, 2014; Liu et al., 2014a, 2019; One Thousand Plant Transcriptomes Initiative, 2019). The source of this conflict is uncertain, but Chang and Graham (2014) observed that relationships were sensitive to different model or method assumptions. In addition, conflicts in relationships among Oedipodiopsida, Tetraphidopsida, Polytrichopsida, and Bryopsida are consistently recovered with only weak to moderate support (Newton et al., 2000; Goffinet et al., 2001; Magombo, 2003; Cox et al., 2004, 2010; Qiu et al., 2006, 2007; Volkmar and Knoop, 2010; Chang and Graham, 2014; Gitzendanner et al., 2018; Liu et al., 2019). Liu et al. (2019) suggest that rapid successive origin of the lineages leading to Oedipodiopsida, Tetraphidopsida, and Polytrichopsida relative to the Bryopsida explains the often weak signals and incongruence concerning relationships among them, which has also been argued to explain poor resolution of inferred relationships among the major lineages of pleurocarpous mosses (Hypnanae; Shaw et al., 2003; Huttunen et al., 2012).

Study aims

We present phylogenomic and molecular evolutionary analyses of data sets derived from the plastid and mitochondrial genomes of a broad sample of bryophytes, including substantial new data. The aims of our study are (1) to reconstruct overall land-plant phylogeny using complete and partially complete plastid and mitochondrial gene sets; (2) to characterize the evolution of plastid genome structure across a representative sampling of bryophyte species; and (3) to test for changes in selective regime in the plastomes of a mycoheterotrophic liverwort (*Aneura mirabilis*) and a putatively mycoheterotrophic moss (*Buxbaumia aphylla*) relative to other bryophytes.

MATERIALS AND METHODS

Taxon sampling

We generated genomic data from shotgun sequencing of 19 moss and 6 liverwort species, and recovered plastid and mitochondrial gene sets from two additional sources: we retrieved plastid and mitochondrial transcriptomic data for 43 taxa from the 1KP project

(Matasci et al., 2014); and full organellar genomes or sets of genes from GenBank (Appendix S1). Our sampling focuses on the taxonomic diversity of major bryophyte clades, but also includes a broad sampling of vascular plants and algal outgroup taxa to enable reconstruction of deep phylogenetic relationships within land plants. In total we included 159 samples representing 138 species for plastid phylogenetic analyses (78 mosses, 40 liverworts, 12 hornworts, 24 vascular plants, and 5 algae) and 157 samples representing 139 species for mitochondrial phylogenetic analyses (78 mosses, 36 liverworts, 13 hornworts, 25 vascular plants, and 5 algae). There are 89 bryophyte taxa in common between the two data sets (52 mosses, 29 liverworts, and 8 hornworts).

DNA extraction, library preparation, de novo contig assembly and gene annotation

For shotgun sequencing of organellar genomes, we extracted total DNA using DNeasy Plant Mini Kits (Qiagen, Valencia, California, USA) following the manufacturer's protocol, but with an extended incubation at 65°C. We prepared whole-genome shotgun sequencing libraries and sequenced them as multiplexed paired-end reads on a HiSeq 2000 machine (Illumina, San Diego, California, USA), following Lam et al. (2018).

We processed and (de novo-) assembled genome-skim reads following Lam et al. (2018). Plastid contigs were retrieved with BLASTn (Altschul et al., 1990), using several bryophytes as reference taxa (i.e., *Marchantia polymorpha*, GenBank accession NC_001319.1; *Physcomitrella patens*, GenBank accession NC_005087.1; *Ptilidium pulcherrimum*, GenBank accession NC_015402.1; *Syntrichia ruralis*, GenBank accession NC_012052.1). For *Buxbaumia aphylla* and *Diphyscium foliosum* we assembled fully circularized plastomes, connecting contigs across gaps using polymerase chain reaction (PCR) amplification of the original DNAs, followed by Sanger sequencing, using custom-designed primers (see Lam et al., 2015).

Matrix construction and sequence alignment

We constructed a matrix comprising 85 plastid genes (81 protein genes and 4 rRNA genes) with 159 taxon terminals and a separate matrix comprising 39 mitochondrial protein genes (mitochondrial rRNA genes not included) with 157 taxon terminals. We excluded two rapidly evolving genes due to alignment difficulties: *ycf*2 (for all taxa) and *ycf*1 (for non-bryophytes). We extracted genes from 1KP assemblies (generated by the 1KP consortium using SOAPdenovo-Trans; Xie et al., 2014) using BLASTn (Altschul et al., 1990) with an e-value cutoff of 1e-20 and all other settings left as default (Camacho et al., 2009). Published organellar genomes of ten bryophyte species were used as query sequences for the BLAST searches (*Nothoceros aenigmaticus*, *Anthoceros angustus*, *Marchantia paleacea*, *Ptilidium pulcherrimum*, *Pellia endiviifolia*, *Physcomitrella patens*, *Sanionia uncinata*, *Sphagnum palustre*, *Syntrichia ruralis*, and *Tetraphis pellucida*). Occasionally, the assembled scaffolds recovered for a gene had short overlaps but were not pieced together by SOAPdenovo-Trans; in these cases, we joined them manually using IUPAC (International Union of Pure and Applied Chemistry) ambiguity codes for occasional mismatches. In a few cases where multiple scaffolds for individual genes overlapped, we used the variant closest to a known related species (based on published species relationships). We coded missing genes (which may represent only unrecovered genes in the case of transcriptome data and for taxa represented by Sanger data; Appendices S2, S3) or pseudogenized genes (apart from those in the mycoheterotrophic liverwort *Aneura mirabilis*; Wickett et al., 2008) as gaps, and treated them as missing data in analyses. Individual gene files were initially aligned using Muscle (Edgar, 2004) in AliView version 1.18 (Larsson, 2014), with a preliminary test tree run for each gene using RAxML BlackBox (Stamatakis et al., 2008) to identify possible contaminants (none were observed). Alignments were then refined manually (Graham et al., 2000) by identifying and staggering regions that were difficult to align globally, and a second round of test trees was run. Such offsetting avoids dubious alignment of rapidly evolving regions (e.g., "overalignment"; Swain, 2018), minimizing errors due to misalignment but potentially retaining sequence variation within the locally aligned offset blocks. Such variation may be phylogenetically informative within these blocks and would otherwise be lost by removing or masking such regions. This allows fuller use of the available sequence data (e.g., Steane, 1999; Graham et al., 2006; Swain, 2018). Alignments for protein genes were maintained in-frame and the individual gene alignments were concatenated into a 79,785 bp plastid matrix and a 42,963 bp mitochondrial matrix. Transcriptomic and genomic sequences from the same taxon were maintained as separate terminals. Translated versions of both matrices were generated using AliView. The DNA and amino-acid alignments for both genomes are publicly available from TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S25209).

Phylogenetic inference

We performed five separate maximum-likelihood (ML) analyses per genome using RAxML version 8.2.10 (Stamatakis, 2014) as implemented on the CIPRES Portal (Miller et al., 2010), with separate unpartitioned and partitioned analyses of both the nucleotide and amino-acid matrices, and with an additional unpartitioned analysis of the nucleotide data with third codon positions excluded to avoid possible effects of saturation or strong compositional bias in this codon position (e.g., Ruhfel et al., 2014; Wickett et al., 2014). The partitioned analyses accommodate different substitution models (or model parameters) for different subsets of the data (see below). For all ML analyses, we conducted 20 independent searches for the best tree and estimated branch support using 500 bootstrap replicates (Felsenstein, 1985). We considered strongly supported branches to have bootstrap support of ≥95%, and poorly supported branches to have <70% bootstrap support (e.g., Soltis and Soltis, 2003). For the partitioned nucleotide analyses, we partitioned the matrices by codon position for each protein gene (a gene-by-codon-based scheme; e.g., Lam et al., 2015; Rothfels et al., 2015), with rRNA genes considered as additional data partitions, resulting in 247 initial data partitions for the plastid matrix and 117 for the mitochondrial matrix. We used PartitionFinder2 (Lanfear et al., 2016) to regroup partitions with similar substitution models and model parameter values under the Bayesian information criterion, resulting in 41 final partitions for the plastid DNA matrix and 25 for the mitochondrial DNA matrix, with GTR+G or GTR+I+G DNA substitution models recovered as the best fit for individual partitions (Appendix S4). The optimal model for the unpartitioned plastid and mitochondrial nucleotide data sets was GTR+I+G. We initially partitioned the amino-acid matrices by gene and used PartitionFinder2 to find the best partitioning scheme, resulting in 77 final partitions for the plastid matrix and nine for the mitochondrial matrix. The optimal model for the unpartitioned

amino-acid matrices was STMTREV (Liu et al., 2014a) for both genomes. To test the possible distorting effects of long outgroup branches on ingroup topology, we also ran each unpartitioned analysis with the algal outgroup taxa removed, using the same models and model parameters as the corresponding unpartitioned analyses with outgroup taxa included. Because we uncovered substantial incongruence between the two organellar genomes concerning overall relationships among the four major land-plant clades and some well-supported incongruence within clades (see below), we did not combine data from the two genomes here (e.g., Huelsenbeck et al., 1996).

RNA editing

RNA editing is thought to have originated early in land-plant evolution, apparently absent in algae but present in all major land-plant lineages except the complex thalloid liverworts (e.g., Freyer et al., 1997; Rüdinger et al., 2012). In our study, genomic and transcriptomic data were available for the same species for 14 and 16 plastomes and mitogenomes, respectively. We directly compared DNA and cDNA sequences in these cases to find evidence of RNA editing, rather than relying on inferential methods (e.g., Mower, 2009; Lenz and Knoop, 2013; Robison and Wolf, 2019). In most cases the differences recovered likely reflect a mixture of DNA substitutional variation within species and RNA edits (C-to-U or U-to-C edit types are assumed to be the only kind of RNA editing in the plastome and mitogenome; Takenaka et al., 2013; Appendix S5). A limitation of this approach is that in most cases we did not have genomic and transcriptomic data from the same individuals or populations (Appendix S5). However, several taxa had a high proportion of C-to-U or U-to-C differences between genomic and transcriptomic data (collectively representing >75% of genome/transcriptome differences for genes in plastomes of *Timmia*, *Buxbaumia*, *Takakia*, and *Nothoceros*, and mitogenomes of *Timmia*, *Pallavicinia*, *Takakia*, and *Nothoceros*; Appendix S5), likely indicative of the existence of RNA editing in these taxa. We focused our attention on such cases for the organellar genomes of the moss *Takakia* and the hornwort *Nothoceros*, because both had by far the highest incidence in both organellar genomes that we observed (collectively representing about 95–100% of all differences for both species and organelles; Appendix S5); we assumed that these differences nearly all represent RNA editing events. To test for the effect of RNA editing on local phylogenetic inference for *Takakia* and *Nothoceros*, we reran the unpartitioned plastid and mitochondrial nucleotide analyses, excluding all positions (i.e., entire columns in each matrix) that contained these C-to-U or U-to-C sites for *Takakia* and/or *Nothoceros*.

Plastome structure comparison

We used Mauve version 2.4.0 (Darling et al., 2004, 2010) to compare plastome gene order across a taxonomically diverse selection of bryophyte species (two hornworts, five liverworts, and nine mosses, including two mosses newly completed here, *Buxbaumia aphylla* and *Diphyscium foliosum*; Appendices S6, S7). This version of Mauve (progressiveMauve) uses an algorithm based on a sum-of-pairs approach to align regions of homology between two or more sequences in an alignment (locally colinear blocks [LCBs]). The LCBs are positioned using progressive alignment based on the approach of CLUSTALW (Thompson et al., 1994; Darling et al., 2004). We used a seed weight of 19 and otherwise used default settings, with one copy of the inverted repeat region (IR) excluded for the structural comparisons.

Tests of changes in selective regime in plastid genes

To test for shifts in selective regime, specifically relaxation or strengthening of purifying selection, in heterotrophic and putatively heterotrophic bryophytes (*Aneura mirabilis* and *Buxbaumia aphylla*, respectively), we examined the d_N/d_S, or ω values (ratios of nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site), for plastid genes in *B. aphylla* compared to other moss taxa, and in *A. mirabilis* compared to other liverworts. We used our original unpartitioned plastid nucleotide tree (see below) and individual plastid gene alignments, pruning all taxa from the alignments and tree except for the target species (*B. aphylla* and *A. mirabilis*) and a selection of relatives (for *B. aphylla*, the mosses *Takakia lepidozioides*, *Sphagnum palustre*, *Andreaea nivalis*, *Tetraphis pellucida*, *Polytrichum juniperinum*, *Diphyscium foliosum*, *Timmia austriaca*, *Physcomitrella patens*, *Racomitrium muticum*, *Dicranum scottianum*, *Syntrichia ruralis*, *Bryum argenteum*, *Pulvigera lyellii*, *Hookeria lucens*, *Fontinalis antipyretica*, and *Sanionia uncinata*; and for *A. mirabilis*, the liverworts *Marchantia paleacea*, *Conocephalum conicum*, *Dumortiera hirsuta*, *Pellia endiviifolia*, *Pallavicinia lyellii*, *Ptilidium pulcherrimum*, *Porella cordaeana*, *Lejeunea patens*, *Herbertus stramineus*, and *Mylia taylorii*). In each case, this yielded a reduced tree and reduced matrices for individual genes. To overcome possible issues associated with insufficient data in shorter genes, we also repeated the analyses using concatenated sets of loci representing multi-subunit gene complexes (i.e., for *atp*, *chl*, *ndh*, *pet*, *psa*, *psb*, *rpl*, *rpo*, and *rps*). We performed tests of change in selective regime using both the branch test (Yang, 1998) and the RELAX test (Wertheim et al., 2015), following Joyce et al. (2018). We conducted branch tests using the codeml module of PAML version 4.9h (Yang, 2007), testing for changes in selective regime of each gene or gene complex (including 15 pseudogenized photosynthetic genes in *A. mirabilis*). The branch test detects changes in ω values in each gene/gene complex in the taxon of interest, compared to the rest of the tree. Two likelihood models are compared: the first estimates a single ω value by assuming that all branches in the tree evolve under the same selective regime (model M_0); the second allows a different selective regime (model M_1) for the test branch (the foreground ω) compared to the rest of the tree (the background ω). The separate sets of branch tests for *B. aphylla* and *A. mirabilis* considered mosses or liverworts as background branches, respectively. We evaluated the significance of the branch tests using a likelihood ratio test, considering a chi-square distribution with one degree of freedom. A significant result indicates there is a difference between the foreground and background ω (i.e., that the test taxon is evolving under a different selective regime to the rest of the tree). We also used RELAX (Wertheim et al., 2015) to test for evidence of relaxed or intensified selection. RELAX assigns every site to a rate class (ω_1, ω_2, and ω_3, representing purifying selection, neutral evolution, and positive selection, respectively) for the test taxon compared to the reference taxa (here the other moss or liverwort species). A null model constrains the test and reference taxa to behave the same for each class; an alternative model allows for relaxation or intensification of selection in the test taxon compared to the null case. Cases of relaxation are inferred when values in each rate class approach 1.0 (neutral evolution); intensification represents a strengthening of selection away from 1.0, in either direction. In this test, rate changes are summarized as a relaxation coefficient (*k*), where

$k < 1$ indicates relaxation, and $k > 1$ indicates intensification of selection. We corrected tests of significance for multiple tests (i.e., genes/gene complexes), adjusting the alpha value to allow for a false discovery rate of 0.05 (Benjamini and Hochberg, 1995).

RESULTS

Gene recovery

Gene occupancy in the final alignments is summarized visually in Appendices S2 and S3, including for a few taxa represented by only a handful of genes from published Sanger-sequencing based studies. For most newly extracted samples (19 moss and 6 liverwort species; Appendix S1), genome skimming recovered a high proportion of organellar genes, with 98.6% mean recovery of plastid genes, except for *Andreaeobryum macrosporum* (75.3%) and *Hedwigia stellata* (76.5%), and 98.5% average recovery of mitochondrial genes, except for *H. stellata* (74.4%) (Appendices S2, S3). Gene recovery for 1KP samples was lower, with 65.2% mean recovery of plastid genes (73.9% in mosses, 57.1% in liverworts, and 50% in hornworts), and 78.2% of mitochondrial genes (86.7% in mosses, 83.4% in liverworts, and 24.5% in hornworts, although this low value for hornwort mitochondrial genes is a consequence of extensive gene losses and pseudogenizations within this lineage; Li et al., 2009; Xue et al., 2010; Villarreal et al., 2018).

Relationships among the four major land-plant clades

All rooted plastid and mitochondrial analyses recovered land plants as monophyletic with 100% likelihood bootstrap support (BS), but relationships among the four major land-plant lineages (mosses, liverworts, hornworts, and vascular plants) differed depending on the genome sequenced and whether nucleotide (DNA) or amino-acid (AA) data were analyzed (Appendices S8–S15; summarized for major lineages in Fig. 2). For both the plastid and mitochondrial genomes the DNA-based analyses resolved bryophytes as paraphyletic with strong (plastid; Fig. 2A; Appendices S8, S9) to moderate bootstrap support (mitochondrial; Fig. 2B; Appendices S12, S13). Only the plastid DNA results had strong support for all major lineages and relationships, recovering a strongly supported tree with liverworts as the sister group of the rest of the land plants, and hornworts sister to vascular plants, in both partitioned and unpartitioned bootstrap analyses (Fig. 2A; Appendices S8, S9). Analyses of the plastid AA data resolved bryophytes as monophyletic, although this was not well supported by partitioned and unpartitioned analyses (BS <50%; Fig. 2A; Appendices S10, S11), and there was only moderate support for a clade comprising liverworts and mosses. The mitochondrial DNA analyses resolved mosses as sister to all other land plants, and hornworts sister to a clade of liverworts and vascular plants, arrangements that had, respectively, moderate and weak support from partitioned and unpartitioned analyses (Fig. 2B; Appendices S12, S13). The mitochondrial AA analyses recovered hornworts as the sister group of other land plants (vascular plants, liverworts and mosses, with the latter two recovered as sister groups), but with weak support in both the partitioned and unpartitioned analyses (Fig. 2B; Appendices S14, S15).

Considering the underlying (de-rooted) relationships among ingroup lineages only from rooted analyses, the plastid analyses (both DNA and AA) and mitochondrial AA analyses recovered the same underlying unrooted four-taxon tree. These analyses effectively differ only in the point of attachment of the algal outgroup (Tree 1 in Fig. 1). The rooted mitochondrial DNA analyses correspond to an alternative unrooted four-taxon arrangement, considering ingroup lineages only (Tree 2 in Fig. 1), but the corresponding bipartition among the four land-plant clades in that tree was poorly supported (66–69%; Fig. 2B).

Analyses with the third codon position excluded for plastid or mitochondrial data recovered different rooted topologies compared to both the DNA and AA analyses, although these relationships were also poorly supported (Fig. 2; Appendices S16, S17). However, these analyses imply the same underlying unrooted four-taxon trees as their respective full-data analyses (Trees 1 and 2 in Fig. 1).

With only ingroup taxa considered in unpartitioned likelihood analysis (algal outgroups not included during tree inference), plastid DNA, plastid AA, mitochondrial DNA, and mitochondrial AA all recovered the same four-taxon relationships for land plants as all the rooted plastid analyses and the mitochondrial AA analyses (i.e., Tree 1 in Fig. 1, consistent with most other published studies; Table 1; Appendices S18–S21; results summarized in Fig. 3). This four-taxon tree was strongly supported by the unrooted plastid analyses and weakly supported by unrooted mitochondrial analyses (Fig. 3; Appendices S18–S21).

Relationships within bryophyte lineages

In general, the plastid results recovered more well-supported clades than the mitochondrial analyses, with mostly longer branches within the land plants, particularly in the liverworts and mosses (Appendices S22–S24). Within the major bryophyte lineages, the only strongly supported conflicts concerned relationships among Takakiopsida, Sphagnopsida, and the rest of the mosses, and the deepest splits of the Bryidae (Figs. 4 and 5; Appendices S8–S17). Plastid analyses based on the full nucleotide data set and a version with the third codon positions excluded both resolved a clade of Takakiopsida and Sphagnopsida (with 100% bootstrap support, BS, in the unpartitioned DNA analysis) sister to the rest of the mosses (Fig. 4; Appendices S8, S9, S16). By contrast, the plastid AA and all mitochondrial analyses found strong support for a grade in which Takakiopsida is the sister group of the rest of the mosses (BS 96–100%; Figs. 4 and 5; Appendices S10–S15, S17). In Bryidae, the plastid DNA analyses resolved a clade of *Philonotis* and *Hedwigia* as defining the deepest split in that subclass (BS 100/88/55% for unpartitioned DNA, partitioned DNA, and third-codon-position-excluded analyses, respectively; Fig. 4; Appendices S8, S9, S16), while all mitochondrial analyses resolved *Philonotis* alone as sister to the rest of Bryidae, with strong support for the latter arrangement in the full mitochondrial DNA analyses and the analysis with third codon positions excluded (BS 98/100/97% for unpartitioned DNA, partitioned DNA and third-codon-position-excluded analyses, respectively; Fig. 5; Appendices S12, S13, S17).

Other noteworthy conflicts among the different analyses were not strongly supported in one or both sets of the conflicting analyses. All plastid analyses recovered Tetraphidopsida and Polytrichopsida as a grade relative to Bryopsida, with moderate support in the DNA analyses (BS 79/90%), while the mitochondrial DNA analyses recovered these two taxa as a clade, also with moderate support (BS 88/89%; Figs. 4 and 5; Appendices S8–S17). These respective relationships were poorly supported in both sets of AA analyses and DNA analyses with third codon positions excluded. In the liverworts all plastid analyses

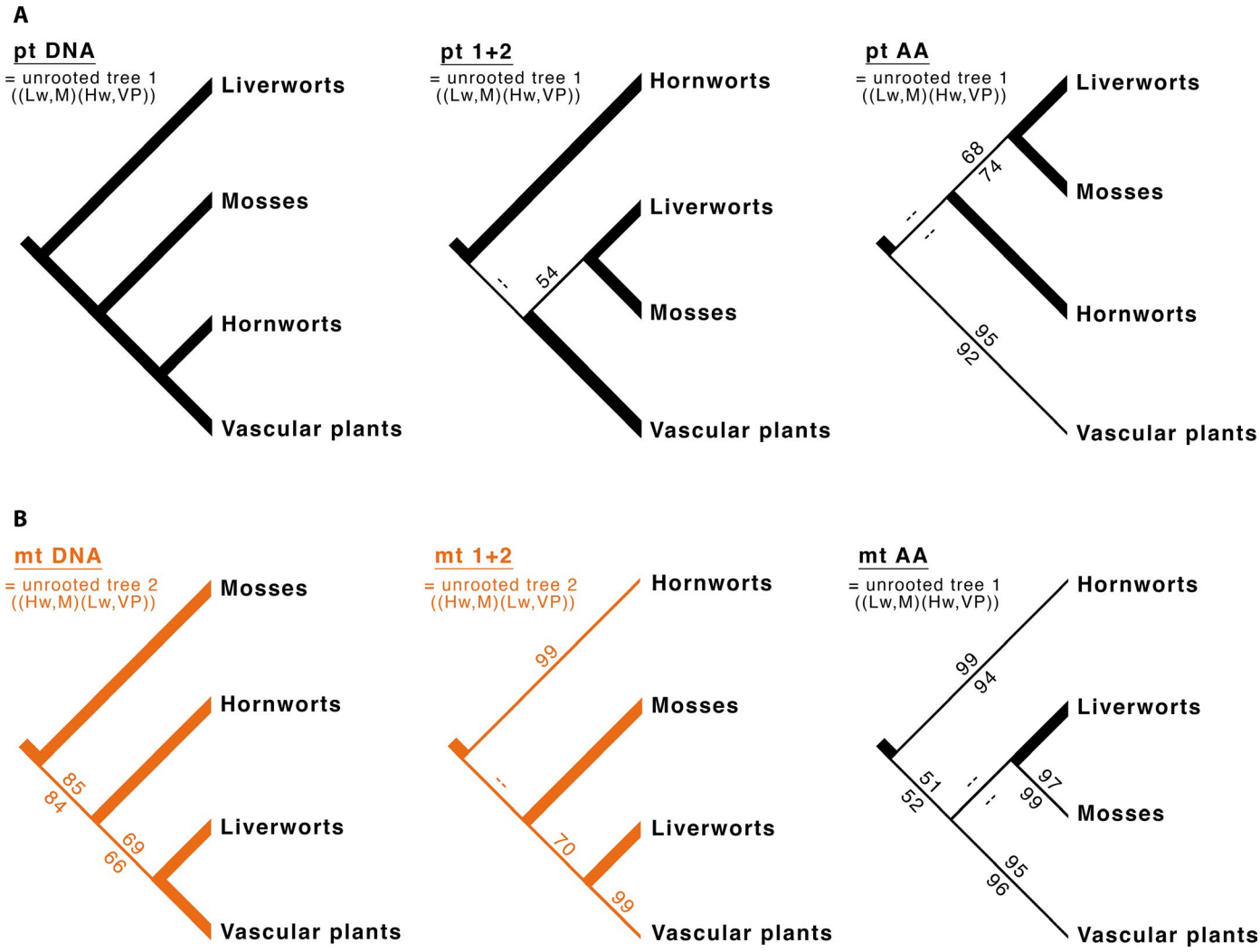

FIGURE 2. Summary of land-plant relationships inferred from likelihood analyses of (A) plastid (pt) and (B) mitochondrial (mt) DNA with all positions included (DNA), DNA with third codon positions excluded (1+2), and amino-acid data (AA). Outgroup algal taxa and relationships within the four major land-plant clades are removed for clarity. Bootstrap support (BS) >50% is indicated (unpartitioned and partitioned values are above and below branches, respectively, except for the 1+2 analyses where only unpartitioned data were analyzed; thickened branches represent 100% BS from both partitioned and unpartitioned analyses, "–" indicates BS <50%). Unrooted land-plant relationships (corresponding to unrooted trees 1–3 in Fig. 1) are annotated beside each figure.

resolved Pelliales and Pallaviciniales as a clade with strong support in the DNA analyses (BS 100%), whereas the mitochondrial analyses recovered these two taxa as a weakly supported grade with respect to the remaining Jungermanniopsida (Fig. 6; Appendices S8–S17). The position of Ptilidiales also varied between analyses, but only the unpartitioned plastid DNA and the plastid analyses with third codon positions excluded resolved its position with strong support (both with BS 99%, as sister to the Porellales). Relationships among hornwort lineages were generally resolved with strong support in the plastid analyses, but with considerably lower support in the mitochondrial analyses. The latter also resolved both *Nothoceros* and *Phaeoceros* as non-monophyletic, with moderate support (Fig. 7A, C; Appendices S8–S17).

RNA editing and phylogenetic inference

Two hornwort genera (*Nothoceros* and *Phaeoceros*) and one moss species (*Takakia lepidozioides*), all widely considered to be monophyletic, were recovered as non-monophyletic with moderate to strong support in the mitochondrial analyses (Figs. 5 and 7C; Appendices S12–S15), and in the case of *Takakia*, in the plastid and mitochondrial analyses (Figs. 4, 5, and 8A, B; Appendices S8–S15). RNA editing is common in the organellar genomes of *Takakia* and hornworts (Sper-Whitis et al., 1996; Steinhauser et al., 1999; Pruchner et al., 2001; Kugita et al., 2003a, b; Duff and Moore, 2005; Duff, 2006; Sugita et al., 2006; Yura et al., 2008; Rüdinger et al., 2012), and all three non-monophyletic groups comprise a mixture of genomic and transcriptomic sequences. Therefore, we reran the unpartitioned DNA-based analyses after excluding sites of RNA editing in *T. lepidozioides* and *Nothoceros aenigmaticus*, the only members of these groups for which both genomic and transcriptomic data were available. Direct comparison of DNA and cDNA sequences identified 764 and 568 C-to-U differences between the genomes and transcriptomes of the plastid and mitochondrion, respectively,

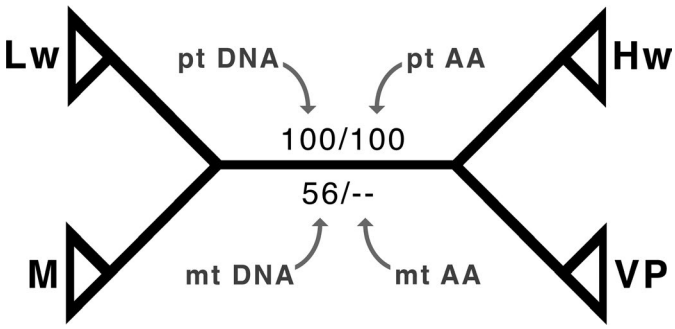

FIGURE 3. Summary of land-plant relationships inferred from each of four likelihood analyses where the outgroup algal taxa were excluded, presented as an unrooted four-taxon tree (= Tree 1 in Fig. 1; see also Appendices S18–S21); bootstrap values >50% represent unpartitioned analyses of plastid DNA and amino-acid data (above central branch), and mitochondrial DNA and amino-acid data (below branch). "–" indicates bootstrap support <50%.

of *T. lepidozioides*, and 99 and 38 for *N. aenigmaticus*. For the latter species, we also identified 19 plastid and one mitochondrial U-to-C differences, consistent with evidence that reverse, U-to-C, editing occurs in hornworts (but not mosses; Kugita et al., 2003a, b; Duff and Moore, 2005; Duff, 2006; Schallenberg-Rüdinger and Knoop, 2016).

Upon exclusion of these RNA edit sites, *T. lepidozioides* was recovered as monophyletic with strong support (BS 100%) in both the plastid and mitochondrial analyses (Fig. 8c, d; Appendices S25, S26), whereas with edit sites included, both analyses recovered this species as paraphyletic, also with strong support, with *T. ceratophylla* nested within *T. lepidozioides* (Figs. 4, 5, and 8A, C; Appendices S8, S12). With edit sites excluded from the mitochondrial analyses, the hornworts *Nothoceros* and *Phaeoceros* were also both recovered as monophyletic, albeit with poor support in the case of *Phaeoceros* (Fig. 7D; Appendix S26). By contrast, with edit sites included, the mitochondrial analyses recovered *Phaeoceros* and *Nothoceros* as non-monophyletic with strong or moderate support, respectively (Fig. 7C; Appendix S12). Note that while the plastid analyses recovered the same set of *Nothoceros* taxa as monophyletic regardless of whether or not edit sites were excluded—and indeed recovered the same hornwort topology overall—they are irrelevant to the question of *Phaeoceros* monophyly because only one species of this genus was sampled in the plastid data set, represented by two transcriptomic samples (Fig. 7A, B; Appendices S8, S25).

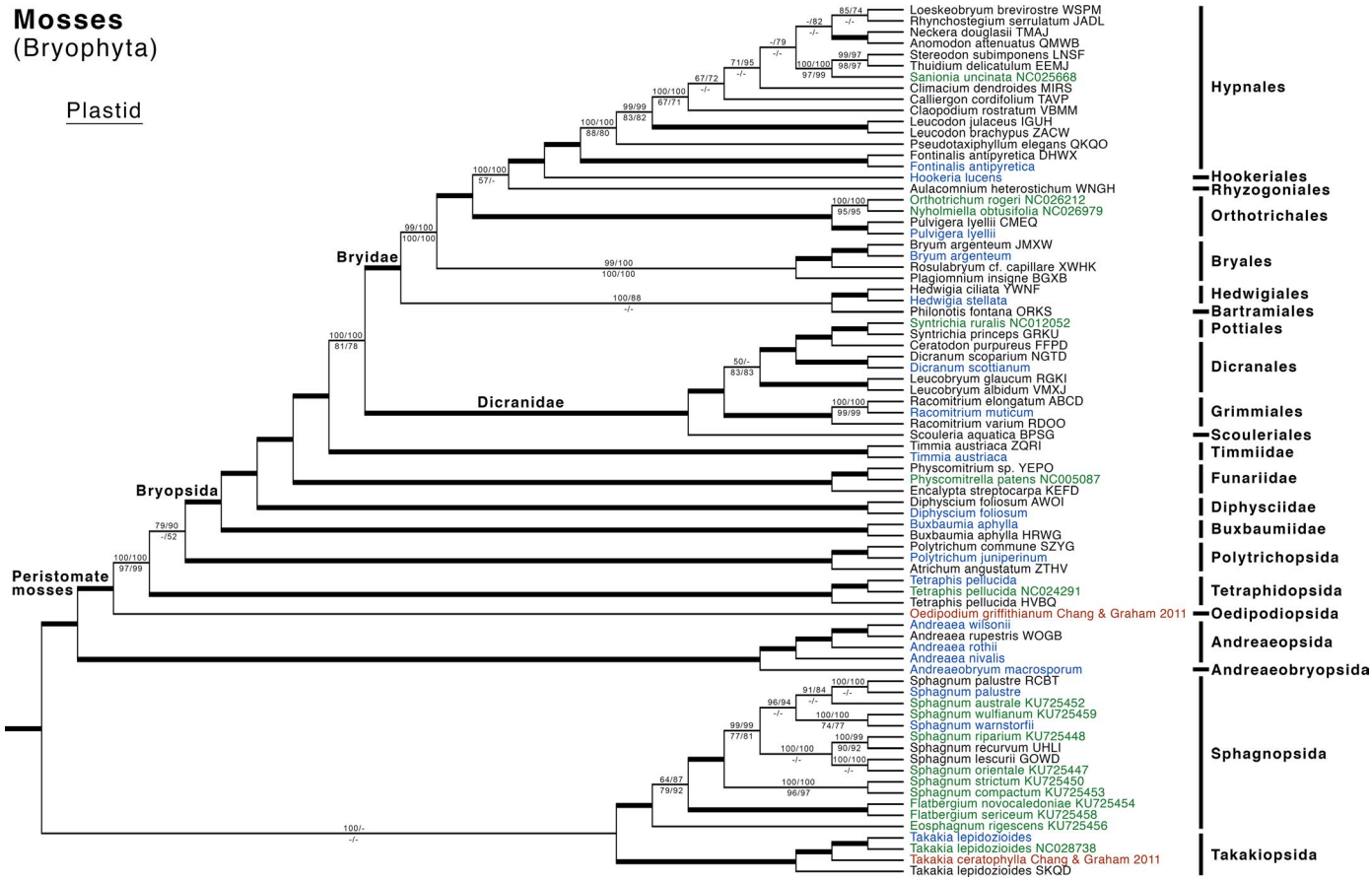

FIGURE 4. Cladogram of relationships in the mosses (Bryophyta) taken from the plastid unpartitioned nucleotide analysis, with likelihood bootstrap (BS) support values >50% from the different analyses indicated (unpartitioned and partitioned DNA above branches; unpartitioned and partitioned amino-acid support values below branches). Thickened branches represent 100% BS in all four analyses. Colors reflect data source; blue = newly sequenced taxa (unannotated); black = transcriptomic gene sets retrieved from the 1KP project (annotated with 1KP four-letter codes); green = genomic gene sets retrieved from GenBank (annotated with GenBank accession numbers); red = Sanger sequencing sets of genes from GenBank (annotated with original publication reference).

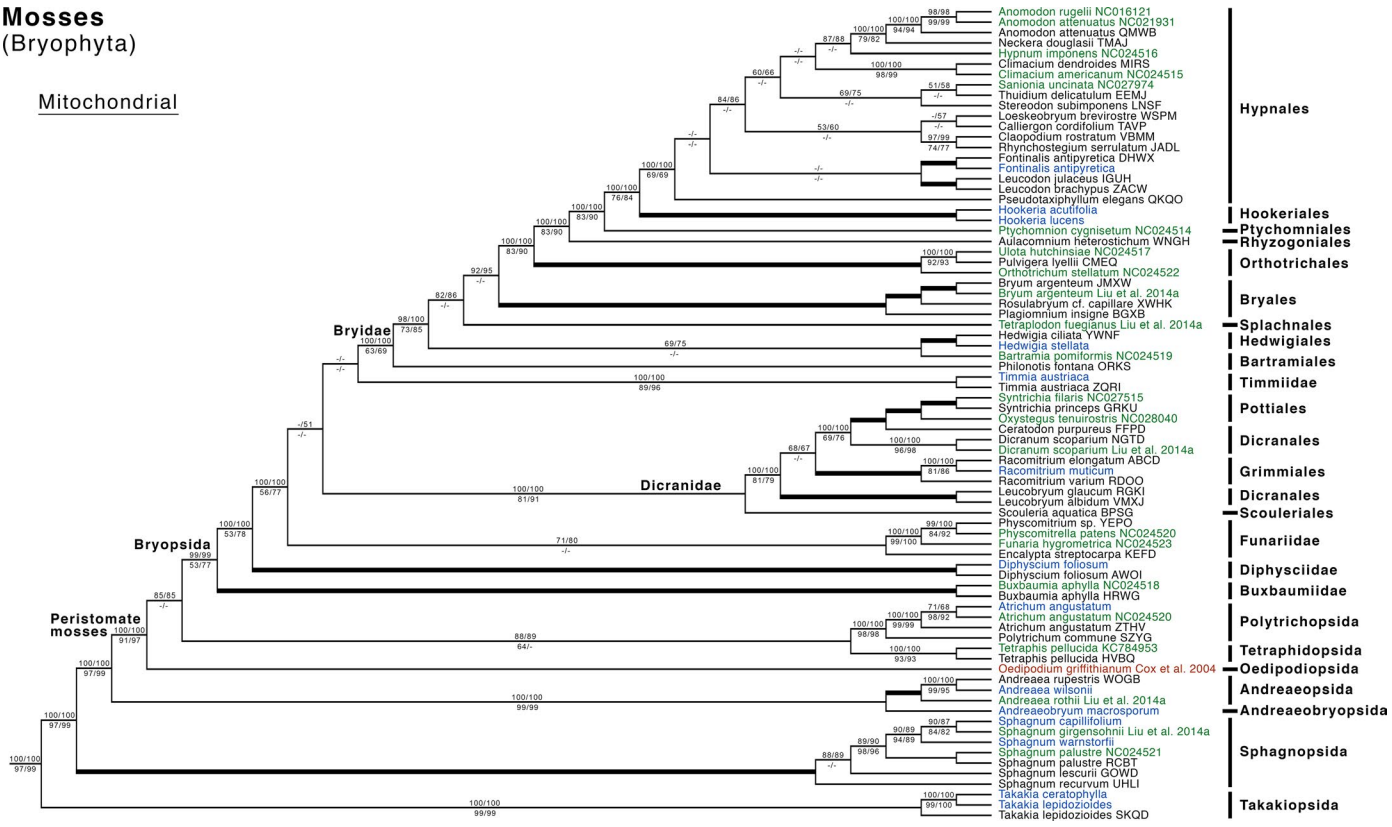

FIGURE 5. Cladogram of relationships in the mosses (Bryophyta) taken from the mitochondrial unpartitioned nucleotide analysis, with likelihood bootstrap support (BS) values >50% from the different analyses indicated (unpartitioned and partitioned DNA analyses above branches; unpartitioned and partitioned amino-acid support values below branches). Thickened branches represent 100% BS in all four analyses. Colors reflect data source; blue = newly sequenced taxa (unannotated); black = transcriptomic gene sets retrieved from the 1KP project (annotated with 1KP four-letter codes); green = genomic gene sets retrieved from GenBank (annotated with GenBank accession numbers); red = Sanger sequencing sets of genes from GenBank (annotated with original publication reference).

Plastome structural evolution

The Mauve comparative analysis of genome synteny (Fig. 9) for 15 fully sequenced plastomes demonstrates no additional changes in gene order across the diverse bryophytes sampled (two hornworts, four liverworts, and nine mosses) beyond three previously recorded changes in gene order. There are no changes specific to any of the three major lineages. A giant (71 kb) inversion is present in the large single copy (LSC) region of Funariidae (represented here by the moss *Physcomitrella patens*), as reported by Goffinet et al. (2007). In the liverworts, *Aneura mirabilis* experienced substantial gene loss and a 1 kb inversion of *psb*E-*pet*L (Wickett et al., 2008), and in the hornworts *Anthoceros angustus* is the only bryophyte to show any structural change in the plastid IR, due to a 4 kb expansion of its IR at the LSC border that includes *ndh*B, *rps*7, and the 3′ exon of *rps*12 (Kugita et al., 2003b).

Plastome sizes in bryophytes vary from >150 kb in the hornworts to ~120 kb in the liverworts (except for the reduced genome of non-photosynthetic *A. mirabilis*, at 108 kb), with the mosses exhibiting variation in size from almost 150 kb in *Takakia lepidozioides* to ~123 kb in the Bryopsida (Appendix S27; note that one copy of the IR is excluded in this figure). Most of this variation is related to expansion or contraction of intergenic spacer sizes, ranging from 53,481 bp combined in *Anthoceros angustus* to 20,370 bp in *Pellia endiviifolia* (Appendix S27).

Plastid gene content is highly conserved across bryophytes (Appendices S27, S28), with only the non-photosynthetic *A. mirabilis* exhibiting many gene losses (Wickett et al., 2008). The purportedly mycoheterotrophic moss *Buxbaumia aphylla* exhibits no gene loss compared to other mosses. Excluding *A. mirabilis*, only 12 gene losses (6 genes lost once, and 3 twice) have occurred for the 88 protein genes clearly present in the common ancestor of the 14 bryophyte plastomes compared (Appendix S28). In addition to these 88 protein genes, there is the enigmatic case of the *tuf*A gene, present only in the moss genus *Sphagnum palustre* and (as a pseudogene) in *Takakia lepidozioides* among all land plants. Note that the *Sphagnum* and *Takakia tufA* sequences are probably orthologs, given that they have the same genomic location and similar base compositions and are substantially more similar to each other than to any other *tuf*A sequences (their closest relatives are from green algal plastomes; data not shown).

Intron content is also highly conserved across bryophytes. Twenty-two plastome introns were present in the common ancestor of land plants (Lemieux et al., 2016). There is only a single loss per se of these 22 introns across the 15 examined bryophytes (of one of two *ycf*3 introns in *Marchantia*; Appendix S28), with all other differences in the intron counts (see Appendix S27) being the result of full or partial gene loss. There is also only a single case of intron gain in bryophyte plastomes, of a single intron in the *rrn*23 gene of *Anthoceros angustus* (Kugita et al., 2003b; Appendix S28).

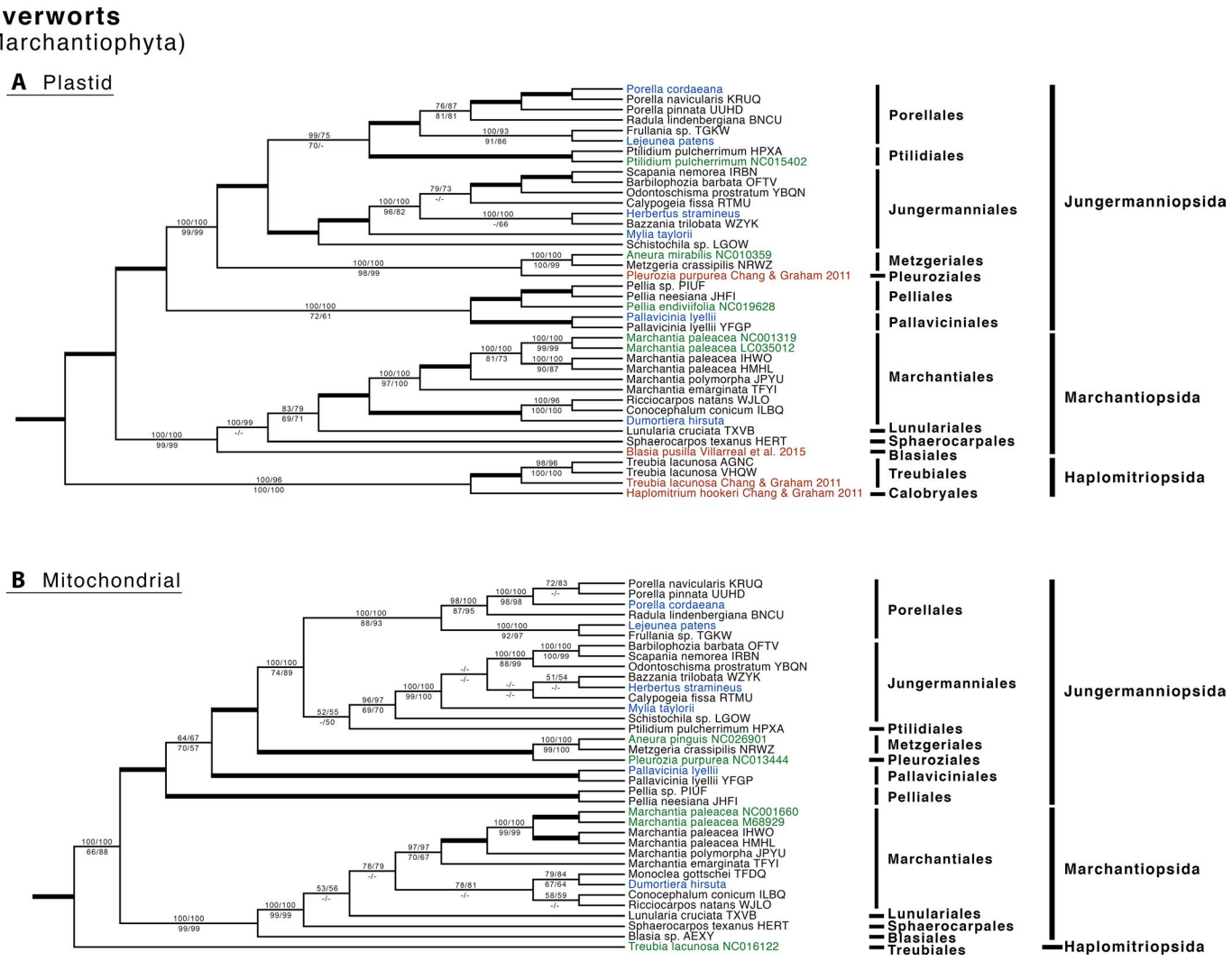

FIGURE 6. Cladogram of relationships in the liverworts (Marchantiophyta) for (A) the plastid results and (B) the mitochondrial results, taken from the unpartitioned nucleotide analyses, with likelihood bootstrap support (BS) values >50% from the different analyses indicated (unpartitioned and partitioned DNA analyses above branches; unpartitioned and partitioned amino-acid support values below branches). Thickened branches represent 100% BS in all four analyses. Values above branches indicate BS support values for the unpartitioned and partitioned nucleotide analyses and below branches indicate BS support from the unpartitioned and partitioned amino acid analyses, respectively. Colors reflect data source; blue = newly sequenced taxa (unannotated); black = transcriptomic gene sets retrieved from the 1KP project (annotated with 1KP four-letter codes); green = genomic gene sets retrieved from GenBank (annotated with GenBank accession numbers); red = Sanger sequencing sets of genes from GenBank (annotated with original publication reference).

Tests of change in selective regime in plastid genes

The branch tests for changes in selective regime indicate that several genes and functionally related groups of genes in *Aneura mirabilis* (*acc*D, *atp*, *cem*A, *ndh*, *pet*, *psb*, *rpl*, *rpo*, and *rps*) have a significantly detectable increase in ω ("d_N/d_S") compared to other liverworts, consistent with relaxation of purifying selection for them, while *rbc*L shows a significant decrease in ω, consistent with a strengthening of purifying selection compared to other sampled liverworts (Fig. 10). *Buxbaumia aphylla* shows only minor changes in ω in comparison to other sampled mosses, with no significant changes in the majority of genes and gene groups, except for *rbc*L and the *psb* genes, in which a strengthening of purifying selection is indicated.

The RELAX tests for *B. aphylla* mirrored the results of the branch tests, with only *rbc*L and the *psb* complex showing significant results, and *k*-value >1, consistent with intensification of selection (Appendix S29). All genes and gene groups showed significant relaxation of selection ($k < 1$) in *A. mirabilis* except for *inf*A, *rbc*L, *ycf*3, and *ycf*12, which had no significant change (for *rbc*L this is in contrast to the branch-test analysis, which showed a significant strengthening of purifying selection; see above).

DISCUSSION

Inference of land-plant rooted and unrooted relationships

Resolving the branching order of the four major land-plant lineages has proved to be a major challenge for plant systematics, and

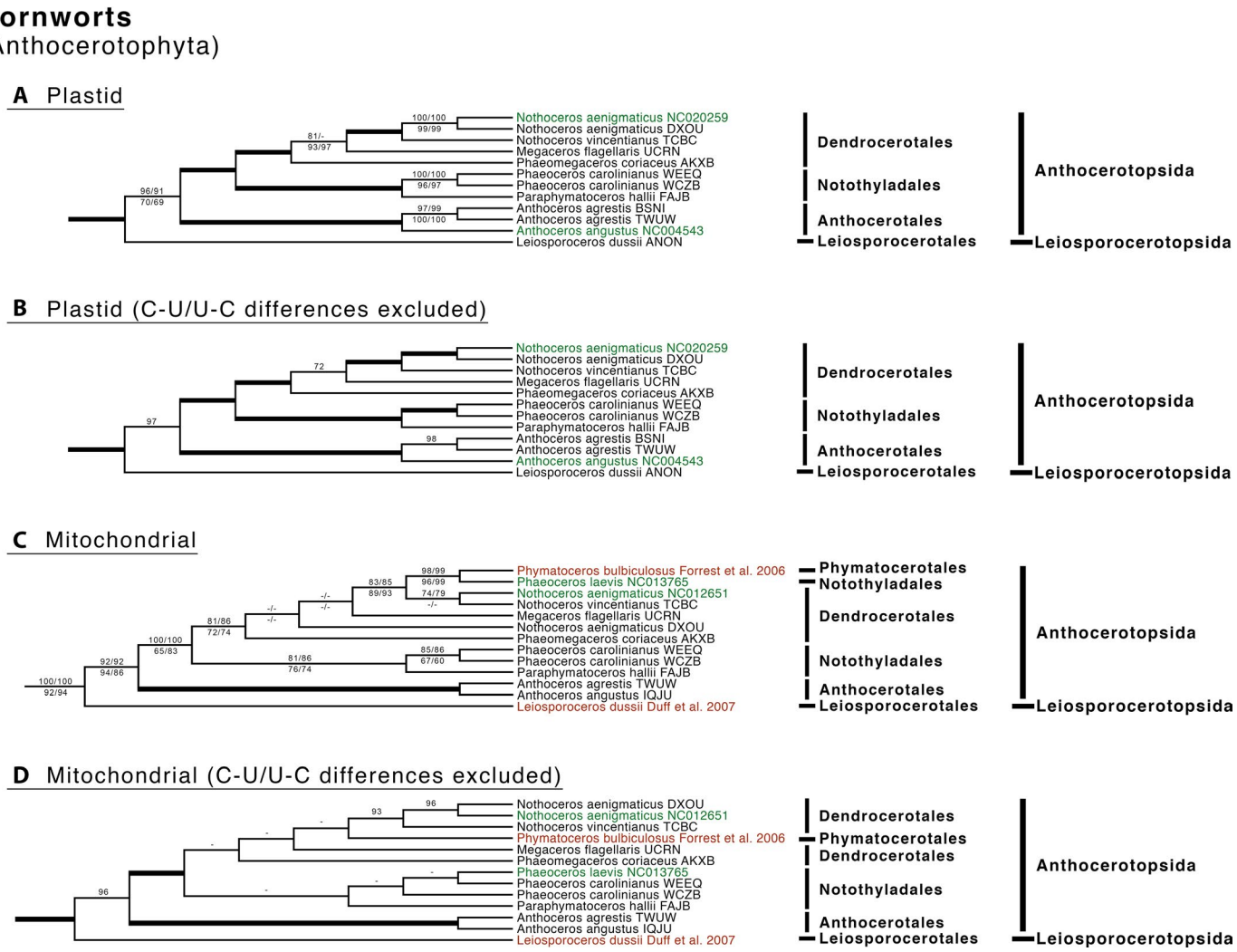

FIGURE 7. Cladograms of relationships in the hornworts (Anthocerotophyta) for (A) the plastid unpartitioned and partitioned DNA and AA results, (B) the plastid unpartitioned DNA results with C-to-U and U-to-C RNA edit sites from *Takakia lepidozioides* and *Nothoceros aenigmaticus* excluded, (C) the mitochondrial unpartitioned and partitioned DNA and AA results, and (D) the mitochondrial unpartitioned DNA results with suspected C-to-U and U-to-C RNA edit sites from *T. lepidozioides* and *N. aenigmaticus* excluded. Likelihood bootstrap support (BS) values >50% from the different analyses are indicated (unpartitioned and partitioned DNA analyses above branches; unpartitioned and partitioned amino-acid support values below branches). Thickened branches represent 100% BS support in all four analyses in A and C and in the single analysis in B and D. Colors reflect data source; black = transcriptomic gene sets retrieved from the 1KP project (annotated with 1KP four-letter codes); green = genomic gene sets retrieved from GenBank (annotated with GenBank accession numbers); red = Sanger sequencing genes from GenBank (annotated with original publication reference).

a consensus answer has arguably yet to be reached for the rooted versions of land-plant phylogeny (e.g., Qiu et al., 2006, 2007; Chang and Graham, 2011; Liu et al., 2014a; Ruhfel et al., 2014; Wickett et al., 2014; Gitzendanner et al., 2018; Puttick et al., 2018; de Sousa et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). Analyses of deep land-plant relationships may be sensitive to details of taxon sampling (e.g., Wickett et al., 2014) and substitution model assumptions (e.g., Chang and Graham, 2014; Liu et al., 2014a; Cox, 2018; de Sousa et al., 2019). Our results continue with this picture of incongruence: for example, different arrangements among the major land-plant lineages are obtained depending on the genome analyzed (plastid vs. mitochondrion) and whether DNA or AA data are used. Thus, no consistent results are found across any of the four data sets that included all codon positions or AA data (Appendices S8–S15; summarized in Fig. 2), although the plastid DNA analysis that excluded third codon positions had the same weakly supported arrangement as the mitochondrial AA data (Fig. 2; Appendices S14–S16). Different subsets of the data have substantially different DNA substitutional dynamics (Appendix S4), which can potentially be accommodated by using data partitioning schemes that group similarly evolving subsets of the data, allowing different models or model parameters to be applied to these subsets during phylogenetic inference (e.g., Lanfear et al., 2012, 2016). However, including this major refinement in model complexity had little effect on inferred relationships for a given data type here, as partitioned and unpartitioned versions of analysis yielded similar results in each case (Figs. 2 and 4–7; Appendices S8–S15). We also examined the effect of excluding third codon positions, which typically evolve substantially faster than the other two codon positions; their elevated rate has been hypothesized to lead to saturation

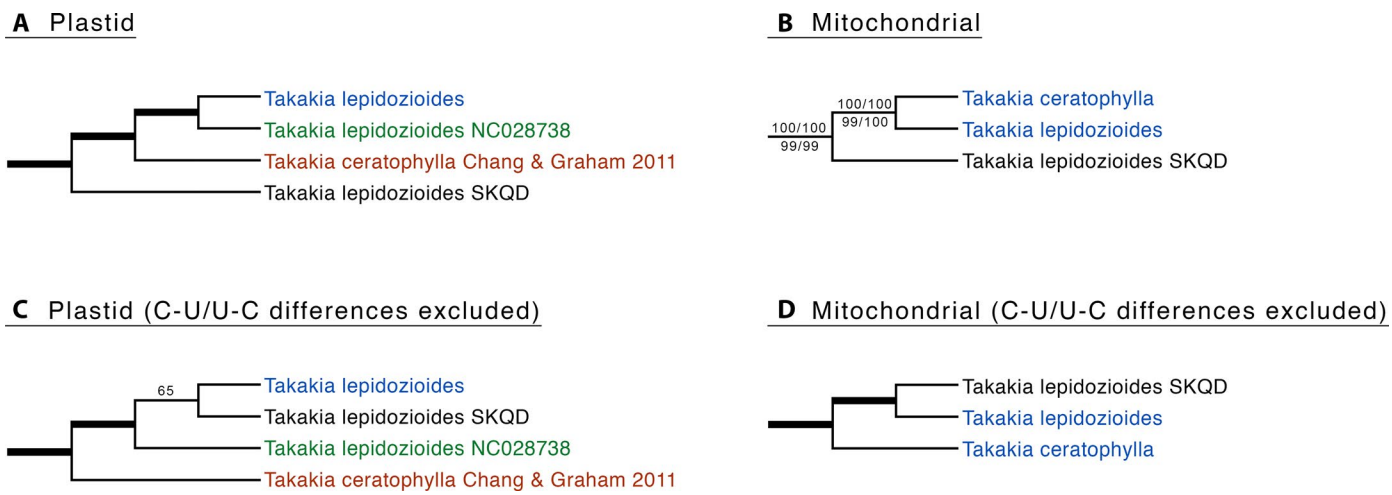

FIGURE 8. Cladograms of relationships in *Takakia* for (A) the plastid unpartitioned and partitioned DNA and AA data, with all sites included, (B) the mitochondrial unpartitioned and partitioned DNA and AA data, with all sites included, (C) the plastid data with C-to-U and U-to-C RNA edit sites from *Takakia lepidozioides* and *Nothoceros aenigmaticus* excluded, and (D) the mitochondrial data with C-to-U and U-to-C RNA edit sites from *T. lepidozioides* and *N. aenigmaticus* excluded. Likelihood bootstrap support (BS) values from the different analyses are indicated, with unpartitioned and partitioned DNA analyses above branches, and unpartitioned and partitioned amino-acid support values below branches for A and B, and unpartitioned DNA analyses above branches for C and D. Thickened branches represent 100% BS support in the four analyses in A and B and in the single analysis in C and D. Colors reflect data source; blue = newly sequenced taxa (unannotated); black = transcriptomic gene sets retrieved from the 1KP project (annotated with 1KP four-letter codes); green = genomic gene sets retrieved from GenBank (annotated with GenBank accession numbers); red = Sanger sequencing sets of genes from GenBank (annotated with original publication reference).

issues (e.g., in the nuclear genome, and when taxon sampling is limited; Wickett et al., 2014; de Sousa et al., 2019). Our analyses that excluded the third codon position led to different land-plant arrangements for both plastid and mitochondrial data sets (Fig. 2; Appendices S16, S17). However, these differences were poorly supported (BS < 70%) in the analyses where the third codon position was excluded. Inferred relationships within mosses, liverworts and hornworts were also generally consistent with corresponding analyses with the third codon positions excluded vs. included, although often with poorer support in the former case (Appendices S16, S17). This reduced branch support is not surprising, given that fewer, more slowly evolving sites were included in these analyses.

Our mitochondrial results differed from the plastid results in having generally shorter branches across the tree (Appendices S22–S24), consistent with the slower rate of sequence evolution of this genome in plants (e.g., Wolfe et al., 1987; Palmer and Herbon, 1988). This relative lack of sequence variation in this organellar genome likely accounts for the shorter branches and, coupled with a smaller total amount of sequence data, the generally lower support values in our mitochondrial trees (Appendices S22–S24). Compared to the plastid DNA trees, the conflicting topologies inferred here from both plastid AA and mitochondrial DNA and AA data have comparatively weak support (Fig. 2), likely due in part to the relative volume of information in each data set (around half the number of genes in the mitochondrial matrix and one-third the number of sites in each AA matrix, although the latter have more character states per site).

In strong contrast to this diversity of rooted trees, a majority of previously published results and most of our results are consistent with the same unrooted four-taxon tree (Tree 1 in Fig. 1; Table 1). Rooting uncertainty among the four major lineages of land plants likely reflects the extremely long branches connecting land plants and their closest sampled algal relatives (e.g., Appendix S22). In our study, only the analyses of the full mitochondrial DNA data set, and the same data set with third codon positions excluded, failed to recover this unrooted tree (instead recovering Tree 2 in Fig. 1, with weak to moderate support; Appendices S10, S11, S15). This result contrasts with that of Liu et al. (2014a), who recovered the commonly inferred unrooted topology (= Tree 1 in Fig. 1), with different root points, for both their mitochondrial DNA and AA analyses (Fig. 1; Table 1). However, when we reran our original unpartitioned analyses with the algal outgroup taxa removed, all four analyses (plastid DNA, plastid AA, mitochondrial DNA, mitochondrial AA; DNA analyses consider all nucleotides) recovered the same basic unrooted relationship (the same unrooted four-taxon tree, = Tree 1 in Fig. 1) for the four major clades, including the mitochondrial DNA analysis (Fig. 3; Appendices S18–S21).

This mostly consistent recovery of the same underlying four-taxon tree in our rooted and unrooted analyses (with and without algal outgroups), and in most published studies, implies that the large distance between land plants and their algal ancestors is a major contributor to the difficulty in resolving deep relationships in land plants (e.g., Chang and Graham, 2011). It also suggests that the unrooted four-taxon relationship among land-plant lineages (Tree 1 in Fig. 1; Fig. 3) may be satisfactorily resolved, or at least that a strong consensus has now been reached. Nevertheless, understanding how these lineages actually relate to each other on the true rooted tree is crucial to our understanding of evolution across land plants; uncertainty surrounding the root point restricts our ability to infer ancestral states for traits of interest in early land plants (e.g., Doyle, 2017; Puttick et al., 2018). The only analyses in our study to recover a consistently strongly supported backbone for overall rooted land-plant relationships were those considering the full plastid DNA data. Both the unpartitioned and partitioned analyses of this data set resolved the bryophytes as a grade, with liverworts sister to the rest of the land plants (the latter

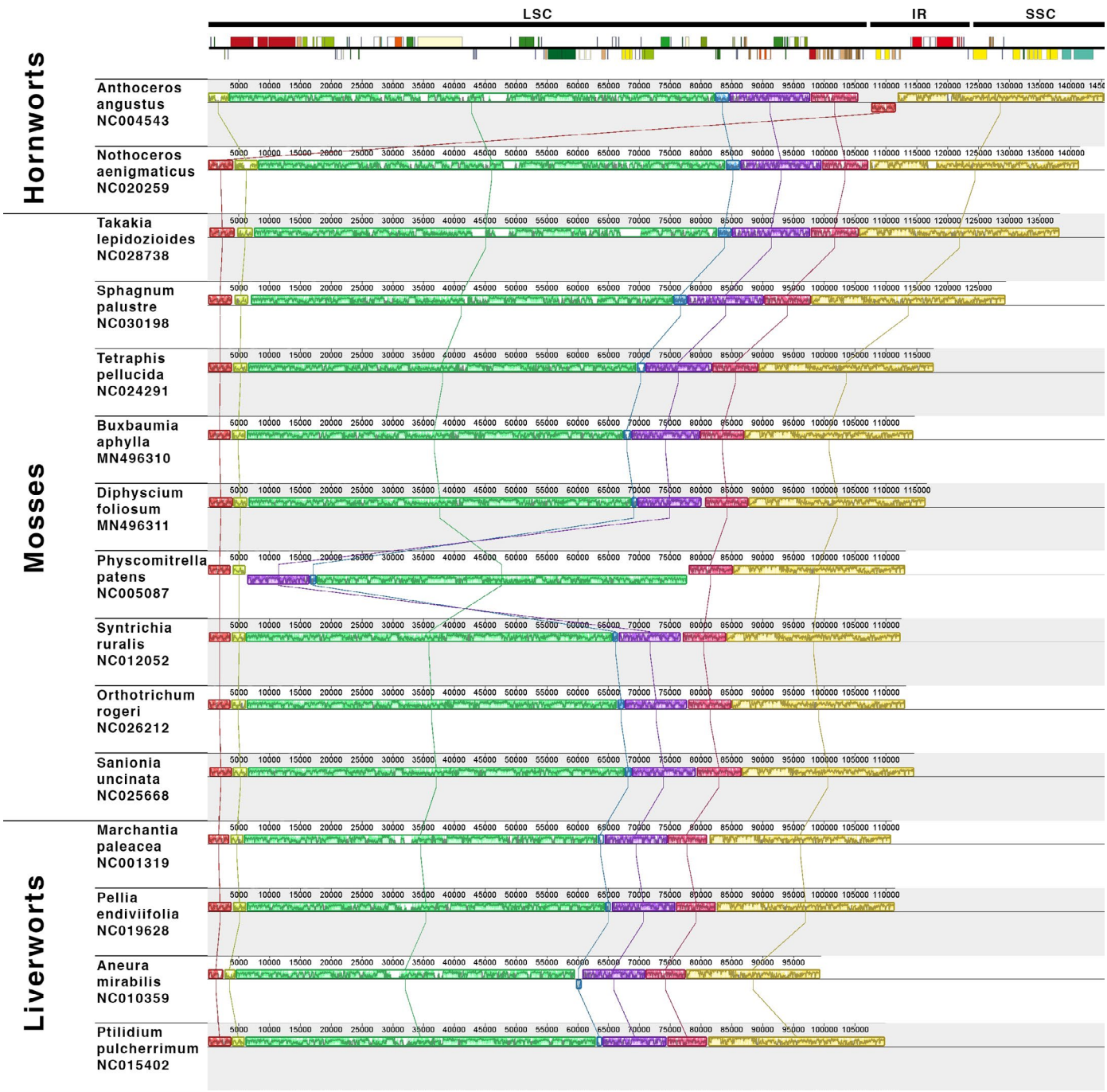

FIGURE 9. Mauve-based alignment comparing plastome structure of 15 bryophyte species, with one copy of the inverted repeat region removed. A linear map of *Anthoceros angustus* appears first for reference. Colored "locally colinear blocks" (LCBs) have shared gene order between plastomes. LCBs appearing above the central line are colinear and in the same orientation and those below align in reverse complement. Colored lines link LCBs shared between taxa.

clade of mosses, hornworts, and vascular plants referred to as stomatophytes, stomata-bearing plants; Kenrick and Crane, 1997a), and hornworts sister to vascular plants (Fig. 2A; Appendices S8, S9). This is also the most frequently recovered topology in published molecular studies (Table 1), mostly inferred using plastid DNA data (e.g., Nishiyama et al., 2004; Forrest et al., 2006; Qiu et al., 2006, 2007; Gao et al., 2010; Karol et al., 2010; Chang and Graham, 2011; Ruhfel et al., 2014) but also in several mitochondrial AA- and DNA-based studies (Turmel et al., 2013; Liu et al., 2014a). By contrast, recent attempts to resolve land-plant relationships using nuclear transcriptomic data have not recovered this topology, but instead have consistently recovered a clade comprising liverworts and mosses (referred to as setaphytes, in which the sporangium is elevated on a stalk known as a seta; Renzaglia and Garbary, 2001), with various relationships then observed among setaphytes, hornworts, and vascular plants, including monophyletic bryophytes in

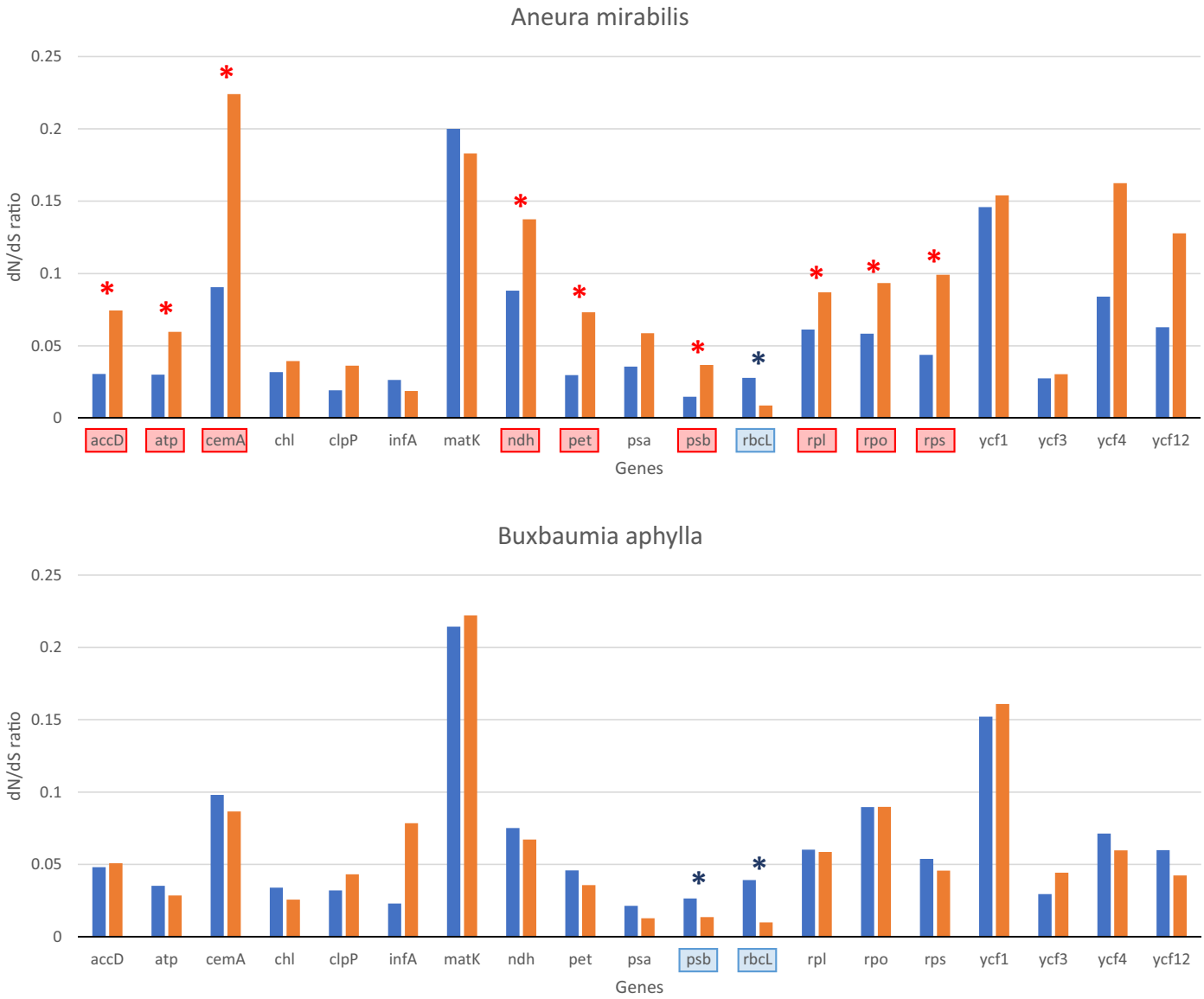

FIGURE 10. Summary of branch test results for plastid genes and gene complexes in *Aneura mirabilis* and *Buxbaumia aphylla*. d_N/d_S (ω) measures the proportion of nonsynonymous substitutions per nonsynonymous site (d_N) to synonymous substitutions per synonymous site (d_S). ω was assessed using a maximum likelihood approach. fg-ω is the "foreground" d_N/d_S ratio for *A. mirabilis* and *B. aphylla* (orange), and bg-ω is the "background" d_N/d_S ratio for the other taxa in the tree (blue). Genes and gene complexes with significant results are marked with an asterisk above and a colored box around the gene name (red for significant relaxation of selection and blue for significant intensification of selection). Note that *ycf*66 is absent from the plastome of *A. mirabilis* and the results for *B. aphylla* are excluded here for consistency (see Appendices S29, S30).

some analyses (Wickett et al., 2014; Puttick et al., 2018; de Sousa et al., 2019; One Thousand Plant Transcriptomes Initiative, 2019). Analyses using plastid AA data also frequently recover bryophytes as monophyletic, with hornworts sister to setaphytes (Nishiyama et al., 2004; Turmel et al., 2006; Lemieux et al., 2007; Shanker et al., 2011; Civáň et al., 2014; Cox et al., 2014; Gitzendanner et al., 2018; One Thousand Plant Transcriptomes Initiative, 2019). In our plastid AA analyses, bryophytes formed a poorly supported clade (BS 38/39%), in contrast to the strong support (BS 95%) found by Gitzendanner et al. (2018) for monophyly of bryophytes, despite the considerable sampling overlap, with both our studies having independently retrieved gene sets from 1KP source data (Matasci et al., 2014). These differences may partly reflect different matrix assembly strategies between our studies, given that Gitzendanner et al. (2018) deleted difficult-to-align regions from their 78-plastid-gene, 1879-taxon matrix resulting in 18,328 AA positions, compared to our inclusion of staggered alignment blocks for difficult-to-align regions in our 85-plastid-gene, 159-taxon matrix of 26,595 AA positions. The difference in alignment length and rooted tree inference may also reflect differences in taxon sampling (we included more bryophytes but far fewer vascular plants and algae, although we represented the major lineages) and in the total number of genes and choice of genes included in each study (only 74 genes in common between studies). Additionally, the One Thousand Plant Transcriptomes Initiative (2019) recovered bryophytes as monophyletic with weak support (BS 64%), based

on an analysis of plastid AA data from a subset of 1178 of the same samples included in the Gitzendanner et al. (2018) analyses.

Relationships within the bryophyte clades are generally consistent among analyses, with the only well-supported conflicts (the relationships between Takakiopsida, Sphagnopsida, and other mosses, and between *Philonotis*, *Hedwigia*, and the rest of the Bryidae) and a handful of poorly supported conflicts (including the relative branching orders of Tetraphidopsida and Polytrichopsida with respect to Bryopsida, and the branching order of various taxa in Bryidae) being consistently problematic in studies of bryophyte phylogenetics (e.g., Newton et al., 2000; Cox et al., 2004; Qiu et al., 2006; Chang and Graham, 2014; Liu et al., 2019). Chang and Graham (2014) noted that relationships among Takakiopsida, Sphagnopsida, and other mosses are sensitive to different modeling parameters, and it is possible that this effect is exacerbated by the long evolutionary distances between these groups. By contrast, the consistently poorly resolved incongruent relationships among Tetraphidopsida, Polytrichopsida, and Bryopsida, and among various lineages in the Bryidae, are likely due to poor phylogenetic signal caused by rapid divergences at these nodes (e.g., Shaw et al., 2003; Liu et al., 2019).

Limited effect of missing data

Missing data in the form of absent genes and/or taxa can be a concern for phylogenetic reconstruction (e.g., the sparsely sampled supermatrices examined in Simmons and Goloboff, 2014). However, including incompletely sampled taxa can also improve phylogenetic accuracy by acting to break up problematic long branches (e.g., Burleigh et al., 2009; Wiens and Tiu, 2012; Jiang et al., 2014). Gene recovery for several species included in our analyses was limited, with a few taxa being represented by only a handful of genes (summarized in Appendices S2, S3). However, this appears to have had little overall effect on phylogenetic inference for the affected taxa, as inferred topologies within the three bryophyte clades were largely congruent with other published studies with regard to these species (e.g., Forrest et al., 2006; Duff et al., 2007; Chang and Graham, 2011, 2014; Villarreal and Renner, 2012; Liu et al., 2019). Specifically, we recovered strong support for the placement of taxa with very limited gene sampling that occupy critical positions in bryophyte phylogeny (e.g., *Haplomitrium hookeri*, *Pleurozia purpurea*, and *Oedipodium griffithianum*).

Simmons and Goloboff (2014) noted that RAxML bootstrap values in patchily sampled supermatrices may be inflated compared to those inferred from more thorough tree searches, which may be a caveat for our findings concerning these taxa. However, several lines of evidence suggest that the recovered relationships were not positively misled by limited data available for them. The most sparsely sampled taxa here (those labeled "Sanger" in Appendices S2 and S3) place in the same or consistent positions in other studies with different sets of taxa and genes (e.g., *Oedipodium*; see Newton et al., 2000; Liu et al., 2019), where the number of sampled genes for these taxa in the original study was comparable to other taxa (e.g., *Treubia* and *Haplomitrium* in Chang and Graham, 2011). These poorly sampled taxa also place in congruent positions using the two organellar data sets here (e.g., the moss *Oedipodium*, based on only 3 mitochondrial genes or 14 plastid genes; Fig. 4 vs. Fig. 5; Appendices S2, S3). The most sparsely sampled major clade here was the mitochondrial data set for hornworts, reflecting gene loss and pseudogenization in the mitogenomes of this lineage (e.g., Li et al., 2009; Xue et al., 2010; Villarreal et al., 2018; Appendix S3; pseudogenized genes in these taxa are excluded from analysis here). However, there were no conflicts between our mitochondrial and plastid inferences of hornworts that were well supported (Fig. 7), suggesting that the missing data did not strongly mislead our analyses. Lam et al. (2018) also examined the phylogenetic placement of multiple taxa in which a subset had substantial missing plastome phylogenomic data due to effects of heterotrophy on photosynthetic genes; their data set had the additional complication that the retained genes tend to be rapidly evolving. Their likelihood analyses yielded positions for these often highly reduced and rapidly evolving taxa that were consistent with those of other studies that used different sets of genes or taxa, further supporting the inclusion of such taxa in analysis of organellar-scale matrices. These considerations suggest that our analyses were not unduly affected by missing data, although further study of this possibility would be useful in future large-scale genomic data sets of land plants.

Organellar RNA editing and localized phylogenetic mis-inference

We found high levels of C-to-U editing in both organelles of the moss *Takakia lepidozioides*. In the case of the plastome, this finding is not unexpected but nonetheless extends the high level of C-to-U editing (302 edits) reported from comparison of a number of plastid gene and cDNA sequences (Sugita et al., 2006; Yura et al., 2008) to 764 C-to-U edits across the set of 72 plastid genes recovered here from both DNA and cDNA sources (Appendix S2). Our data provide the first direct, DNA/cDNA evidence of extensive mitochondrial RNA editing in *T. lepidozioides*, confirming and extending to the genome scale previous predictions of high levels of RNA editing for four *T. lepidozioides* mitochondrial genes (Pruchner et al., 2001; Rüdinger et al., 2012), with 568 edits across the 36 mitochondrial genes recovered from both DNA and cDNA sources (Appendix S3).

Several authors have voiced concern over the effects of RNA editing on inference of hornwort phylogeny (e.g., Duff et al., 2007; Villarreal et al., 2013). Bowe and dePamphilis (1996) noted that RNA editing may be problematic when genomic (DNA-derived) and transcriptomic (RNA-derived) data are used in the same analysis (as is the case in our study), because the same forms of the gene in different taxa may tend to group together as a result of independent RNA edit sites being mistakenly interpreted as homologies. The C-to-U and U-to-C RNA editing sites inferred here for *Nothoceros* may explain contrasting inferences of hornwort phylogeny in mitochondrial ML analyses when these sites are included vs. excluded (Fig. 7C, D). However, some of these relationships are weakly supported in both cases, except for moderate and strong support for non-monophyly of *Nothoceros* and *Phaeoceros*, respectively, when all sites are included, and strong support for monophyly of *Nothoceros* when edit sites are excluded (Fig. 7C, D; Appendices S12–S15, S26). Our comparisons of transcriptomic and genomic data for plastid genes of *Nothoceros*, together with published data for *Anthoceros* (Kugita et al., 2003a), suggest a greater volume of C-to-U differences in the plastid data matrix than in the mitochondrial data. However, the plastid subtree for hornworts is consistent when analyses either included or excluded edit sites (Fig. 7A, B; Appendices S8–S11, S25). This suggests that if RNA editing has an effect on phylogenetic inference using plastid data for this clade, the effect is overridden by strong signal in the remaining nucleotides.

Excluding C-to-U and U-to-C sites had a very strong effect on species relationships within the moss *Takakia*. With all sites included, *T. ceratophylla* was nested within *T. lepidozioides* with 100% BS support from both plastid and mitochondrial analyses (Figs. 4 and 5; Appendices S8–S15), whereas upon edit-site exclusion *T. lepidozioides* was monophyletic, also with 100% BS support (Fig. 8; Appendices S25, S26). Thus, these sites likely misled phylogenetic inference within *Takakia*, but this appears to be correctable by their exclusion from analysis. Several other taxa also had indications of some RNA editing based on a relatively high fraction of C-to-U or U-to-C differences between their organellar genomes and corresponding transcriptome data (Appendix S5; and see Methods). However, for all these taxa, the pair of genomic and transcriptomic sequences representing a species always fell out as sister taxa in our main analyses (Figs. 4–7), without any need for the sort of manipulations we employed for *Takakia* or *Nothoceros* here.

Plastome structural stability in bryophytes

The plastome of most photosynthetic land plants is relatively conserved in size, gene order, and gene and intron content, such that inversions and other structural changes are usually straightforward to infer in comparisons of even distantly related lineages (e.g., Wicke et al., 2011). Aside from short indels and inversions, the most frequent class of structural mutation in land-plant plastomes is change in the extent of the plastid IR region due to its ebb and flow (expansion and contraction) with respect to the flanking large and small single-copy regions (LSC and SSC), which is responsible for much of the observed range of plastome sizes in land plants (Goulding et al., 1996; Mower and Vickrey, 2018).

The number of complete bryophyte plastomes (15) compared here more than doubles the number compared in any previous study (e.g., Mower and Vickery, 2018; Park et al., 2018). Yet these 15 genomes differ by only two inversions, both previously reported (Goffinet et al., 2007; Wickett et al., 2008), and only the hornwort *Anthoceros angustus* displays a significant expansion or contraction of the IR, also previously reported (Kugita et al., 2003b). Plastid gene and intron content are also highly conserved among the 14 photosynthetic bryophytes, with the liverwort *Aneura mirabilis* having lost many genes (and their accompanying introns) in keeping with its non-photosynthetic, fully mycoheterotrophic condition (e.g., Wickett et al., 2008; Graham et al., 2017). Our findings considerably extend and thereby strengthen the conclusions of previous studies (Forrest et al., 2011; Mower and Vickrey, 2018; Park et al., 2018) that plastome structure, including gene and intron content, is highly conserved in bryophytes.

Variation in total plastome length (particularly in the mosses) is probably the most striking global difference across the bryophytes. With the exception of *Aneura mirabilis* (see preceding paragraph), little of this variation can be accounted for by gene and intron loss (Appendix S28) or, excepting *Anthoceros angustus* (Kugita et al., 2003b), by IR expansion or contraction. Rather, the major contributor to plastome size variation in bryophytes is variation in the size of intergenic spacers (Appendix S27).

One surprise from the plastome comparisons concerns the *tuf*A gene, which encodes protein-synthesis elongation factor EF-Tu. In the first documented case of relatively modern-day functional transfer of an organellar gene to the nucleus, this essential gene was inferred to have been lost from the plastome and transferred to the nucleus in a green algal ancestor of land plants (Baldauf and Palmer, 1990). Subsequent studies indicated that this transfer probably occurred in the common ancestor of land plants and the Zygnematophyceae, the probable green-algal sister group to land plants (e.g., Chang and Graham, 2011; Timme et al., 2012; Wickett et al., 2014) and showed that *tuf*A is absent from all >2000 sequenced plastomes of land plants. It thus came as a surprise to find an intact *tuf*A gene in the plastome of the moss *Sphagnum palustre* and a likely orthologous *tuf*A pseudogene in the *T. lepidozioides* plastome (note that intact *tuf*A genes were annotated in plastomes of 38 *Sphagnum* species, but their presence went unnoted in the relevant manuscript; Shaw et al., 2016).

The phylogenetically puzzling presence of *tuf*A in these moss plastomes can be explained by at least four scenarios that are considerably complex and/or involve highly unlikely events. One scenario postulates a much earlier functional transfer of *tuf*A, in a land plant/Zygnematophyceae common ancestor, than its subsequent loss from the plastome, with retention of functional *tuf*A copies in both the plastid and nucleus in multiple lineages for many millions of years. This scenario would invoke three to five losses of plastid *tuf*A following its transfer, once on the branch leading to Zygnematophyceae, once in the common ancestor of mosses exclusive of Takakiopsida and Sphagnopsida, and one to three losses deep in land plant phylogeny (once according to the "mt DNA" tree in Fig. 2 and two or three times under the other five trees in Fig. 2). The flip side of this scenario is one that postulates three to five independent *tuf*A transfers (and phylogenetically coincident losses), these occurring in the same lineages as in the above scenario. In their simplest form, the other two scenarios postulate a single loss of *tuf*A from the plastome, phylogenetically coincident with its functional transfer to the nucleus in a land plant/Zygnematophyceae common ancestor, followed by plastome reacquisition of *tuf*A in the *Takakia*/*Sphagnum* clade recovered in Fig. 4 and Appendices S8, S9, and S16. These two scenarios become more complicated, and thus less likely, requiring either two reacquisition events or an additional transfer to the nucleus, under the topology with *Takakia* sister to other mosses, recovered in Figs. 4 and 5 and Appendices S10–S15 and S17. One reacquisition scenario invokes reverse transfer of *tuf*A from the nucleus to the plastid, while the other invokes horizontal gene transfer, perhaps from a green algal lineage (see Results). There are more-or-less apt precedents for all four scenarios in green-plant organellar evolution (Adams et al., 1999, 2002; Qiu et al., 2014; Atluri et al., 2015; Leliaert and Lopez-Bautista, 2015; Wu et al., 2017), and so each should be considered a viable possibility. We hope that our unexpected discovery of *tuf*A in certain moss plastomes will stimulate studies designed to achieve resolution among the several complex scenarios that could have given rise to this intriguing evolutionary puzzle.

The *Buxbaumia aphylla* plastome lacks signals of relaxed purifying selection evident in *Aneura mirabilis*

Heterotrophic angiosperms often display highly elevated rates of plastid nucleotide substitution, coupled with plastid gene loss (e.g., Graham et al., 2017; Lam et al., 2018) and sometimes extensive or moderate genome rearrangements (e.g., Logacheva et al., 2014; Lam et al., 2015). Only one bryophyte lineage is known to be non-photosynthetic (Merckx et al., 2013a, b), the

fully mycoheterotrophic liverwort *Aneura mirabilis*. This species derives its nutrition entirely from fungal partners (Bidartondo et al., 2003), and its genome shows unmistakable signs of loss of photosynthetic function (Wickett et al., 2008). In addition, the moss genus *Buxbaumia* has been described as a saprophyte (Campbell, 1895; Schofield, 1985), and as a putative partial mycoheterotroph (i.e., capable of photosynthesis but also partnering with soil fungi as a source of supplementary food supply; Leake, 1994; Bidartondo, 2005). This genus of ~12 species (Schofield, 2007) is characterized by an extremely reduced gametophyte, most noticeably in the leafy gametophore (Duckett et al., 2004). However, its status as a putative mycoheterotroph has been questioned by Duckett et al. (2004), who found no association between the rhizoids of *Buxbaumia* and nearby fungal hyphae, supporting the idea that it is not even partially mycoheterotrophic. In retrospect, this is not unexpected, given that few mosses are thought to have interactions with soil fungi in general, in contrast to liverworts and hornworts, which frequently contain endophytic fungi in their gametophytes (Felix, 1988; Read et al., 2000). Mycoheterotrophs typically undergo a sequence of plastid gene pseudogenization and loss beginning with the NADH dehydrogenase (*ndh*) complex followed by most genes relating to photosynthetic functions (e.g., Graham et al., 2017). The complete plastome of *B. aphylla* presented here (Appendix S6) reveals that all plastid genes are retained in *Buxbaumia* as fully open reading frames, and show no evidence of *ndh* gene loss, unlike other partially heterotrophic plants (summarized in Graham et al., 2017).

However, we speculated that retained *ndh* genes, or other plastid photosynthesis-related genes in *Buxbaumia*, might still be experiencing relaxation of selection if the genus were in the early stages of mycoheterotrophy (in turn a possible concern for phylogenetic inference). We addressed the possibility of altered selection by examining the d_N/d_S values in the species of interest compared to known autotrophic relatives. Several tests have been used to look for changes in selective regime consistent with transitions toward partial or full heterotrophy in parasitic plants (e.g., Petersen et al., 2015; Cusimano and Wicke, 2016; Wicke et al., 2016) and mycoheterotrophic plants (e.g., Barrett et al., 2014; Lam et al., 2015; Logacheva et al., 2016; Braukmann et al., 2017; Joyce et al., 2018). We performed these tests separately for both *B. aphylla*, compared to other mosses, and *A. mirabilis*, compared to other liverworts, including the latter as a positive control to demonstrate the effectiveness of the test (*Aneura* is a known fully mycoheterotrophic plant). The results of these branch and RELAX tests (Fig. 10; Appendix S29) suggest considerable relaxation of purifying selection in many of the plastid genes of *A. mirabilis*. Together with the loss or pseudogenization of all *ndh* and many photosynthesis genes (Wickett et al., 2008; Appendix S28), this is consistent with it having transitioned to a fully mycoheterotrophic lifestyle (Graham et al., 2017). One quirk for *Aneura* is that the branch test suggested stronger purifying selection in its *rbc*L gene (although the RELAX test failed to detect significant change in the selective regime for it). Whether under similar or intensified selection compared to autotrophs, this suggests that *rbc*L is still functional in *Aneura*, which is consistent with the secondary role of Rubisco in lipid biosynthesis (e.g., Schwender et al., 2004). Of the 15 pseudogenes identified in *A. mirabilis* by Wickett et al. (2008) and included in the branch tests of individual genes here, only six showed significant relaxation of selection (Appendix S30). This may be an artifact of the relatively short length of several of these pseudogenes or the relative recency of loss of function.

In contrast to the results for the non-photosynthetic *A. mirabilis*, the only significant shift in selective regime detected in plastid genes for the putatively mycoheterotrophic *B. aphylla* was an intensification of purifying selection in the *psb* and *rbc*L genes, detected by the branch and RELAX tests (Fig. 10; Appendix S29). The reason for this intensification is unknown, although it may be related to the reduced size of the gametophyte generation in this plant. The absence of relaxation or release of selection in any genes in *Buxbaumia* is a contrast with other partial heterotrophs (e.g., Barrett et al., 2014; Petersen et al., 2015; Cusimano and Wicke, 2016; Wicke et al., 2016). The lack of evidence of change in selective regime in photosynthetic genes may mean that it is in the very early stages of partial heterotrophy. However, our findings are also consistent with the morphological observations of Duckett et al. (2004), who found no physical interactions between this moss and soil fungi. Taken together, these two lines of evidence do not support mycoheterotrophy in *B. aphylla*.

CONCLUSIONS AND FUTURE DIRECTIONS

Despite increasing access to large genomic data sets, relationships among the four major land-plant lineages are still contradictory and inconsistent across and even within studies, including this one, which features the largest taxon sampling for bryophyte organellar data sets to date. However, the basic unrooted topology inferred among the four major land-plant lineages (liverworts, mosses, hornworts, and vascular plants) is congruent across most published studies and most of our analyses (Figs. 1 and 3; Table 1). This supports the idea that the very long branch connecting land plants to their algal outgroups, coupled with relatively short branches connecting the major land-plant clades, is the major source of conflict among studies, and that we now understand the unrooted tree uniting these lineages. The best-supported inferences in our study are those based on partitioned and unpartitioned analyses of plastid DNA data (Fig. 2A; Appendices S8, S9). These analyses include third codon positions, which are more likely to experience saturation effects. However, removal of these sites or replacement with degeneracy codes (Cox et al., 2014; Wickett et al., 2014) may not be necessary for plastid genes, which evolve more slowly than nuclear genes (Wolfe et al., 1987; Drouin et al., 2008; Rothfels and Schuettpelz, 2014), and corrections for saturation may in general be less important with well-sampled phylogenies (e.g., Pollock et al., 2002; Hedtke et al., 2006).

A major potential limitation of our data is that many taxa are represented by incomplete gene sets, because they were derived from available GenBank accessions, including Sanger sequencing data sets, and from transcriptomes that vary in gene recovery (Appendices S2, S3). Despite this, inferred relationships within each major bryophyte clade are largely congruent among our analyses and with other published phylogenies, supporting the view that data sets with patchy recovery of genes can perform well in phylogenetic inference (e.g., Burleigh et al., 2009; Wiens and Tiu, 2012; Jiang et al., 2014; Lam et al., 2018). The few relationships within bryophyte lineages that do remain contentious are likely due to substantial evolutionary distances among lineages in some instances (e.g., *Takakia*, *Sphagnum*, and other mosses), or due to rapid diversification among the relevant lineages in other cases (e.g., Tetraphidopsida, Polytrichopsida, and Bryopsida). We noted the presence of RNA edit sites in the moss *T. lepidozioides* and the

hornwort *Nothoceros aenigmaticus*, and demonstrated for these two bryophytes that these edits likely affect inference of local phylogenetic relationships.

Our data have additional implications for the molecular evolution of plastid genomes. As is the case for most groups of vascular plants, bryophyte plastomes are highly conserved in gene order, gene content, intron content, and, to a lesser extent, genome size. Although the number of complete bryophyte plastomes that have been sequenced to date is still small, they broadly represent the major bryophyte lineages. Nonetheless, it would be useful to expand on this sampling to survey for lineage-specific effects, which are strikingly evident in certain vascular plant lineages. As well, it should be interesting to explore the complex alternative scenarios outlined above to account for the one striking novelty in bryophyte plastomes uncovered in this study: the unexpected presence of the *tufA* gene in *Sphagnum* and *Takakia* against a backdrop of the gene's absence from *all* other examined land-plant plastomes. *Buxbaumia aphylla* conforms to the general pattern of plastome stasis and shows no evidence of gene loss or shifts in selective regime that might indicate that it is in the early stages of photosynthesis loss, as might be expected if it were mycoheterotrophic. We also found no evidence of relaxation of selection in any of its plastid genes, in contrast to the known mycoheterotrophic liverwort *A. mirabilis*. Our data are therefore not consistent with the existence of mycoheterotrophy in *B. aphylla*.

What hope is there for resolving the land-plant root satisfactorily (which of the rooting points of Tree 1 in Fig. 1 are correct)? The serendipitous discovery of cryptic algal lineages with a close relationship to land plants would help (in effect, a discovery parallel to the recent identification of a close non-photosynthetic sister group to red algae; Gawryluk et al., 2019). In the absence of such good luck, improved substitutional models may help, as would the discovery of rare genomic or morphological changes that unambiguously support one or another resolution of early land-plant relationships. It may also be useful to explore gene duplication events that may intersect the long branch connecting land plants and other streptophytes, closer to land plants than their algal outgroups, allowing confident duplicate gene rooting (Mathews and Donoghue, 1999; Simmons et al., 2000; Emms and Kelly, 2017). The increasing availability of large genomic data sets may facilitate the latter approach (e.g., Wickett et al., 2014; One Thousand Plant Transcriptomes Initiative, 2019). However, full understanding of early land-plant evolution will also require an improved early fossil record, and better integration of living and extinct lineages in phylogenetic inference (e.g., Kenrick and Crane, 1997b; Gensel, 2008; Rothwell et al., 2009).

ACKNOWLEDGMENTS

The authors thank P. Hollingsworth for advice and two anonymous reviewers for comments that helped to improve the text. This work was supported by an NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant (to S.W.G.), a University of British Columbia Four-Year Fellowship (to D.B.), and a U.S. National Science Foundation grant (to Z.N.T. and J.D.P.). P.S. was supported by the Swiss National Science Foundation (160004 and 131726) and by URPP Evolution in Action.

DATA AVAILABILITY

DNA and amino-acid alignments for the plastid and mitochondrial gene sets used in the study are available from TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S25209).

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Accession information for organellar genomes or gene sets used in the analyses, including GenBank accession numbers, four-letter codes for 1KP samples, and voucher information (collection number and herbarium) for newly sequenced samples.

APPENDIX S2. Heatmap summarizing gene coverage in final alignments for the plastid analyses, organized by major source of data.

APPENDIX S3. Heatmap summarizing gene coverage in final alignments for the mitochondrial analyses, organized by major source of data.

APPENDIX S4. Data partitioning schemes used in the partitioned analyses.

APPENDIX S5. Nucleotide variation for species represented by both genomic and transcriptomic data, with the frequency of each type of nucleotide difference noted.

APPENDIX S6. Circular plastome map of *Buxbaumia aphylla*.

APPENDIX S7. Circular plastome map of *Diphyscium foliosum*.

APPENDIX S8. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned plastid nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S9. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the gene-by-codon partitioned plastid nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S10. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned plastid amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S11. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the gene-partitioned plastid amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S12. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned mitochondrial nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S13. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the gene-by-codon partitioned mitochondrial nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S14. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned mitochondrial amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S15. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the gene-partitioned mitochondrial amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S16. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned plastid nucleotide analysis with the third codon position excluded, with likelihood bootstrap support values indicated above branches.

APPENDIX S17. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned mitochondrial nucleotide analysis with the third codon position excluded, with likelihood bootstrap support values indicated above branches.

APPENDIX S18. Cladogram of unrooted land-plant relationships inferred from the unpartitioned plastid nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S19. Cladogram of unrooted land-plant relationships inferred from the unpartitioned plastid amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S20. Cladogram of unrooted land-plant relationships inferred from the unpartitioned mitochondrial nucleotide analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S21. Cladogram of unrooted land-plant relationships inferred from the unpartitioned mitochondrial amino-acid analysis, with likelihood bootstrap support values indicated above branches.

APPENDIX S22. Phylograms of land-plant relationships with five algal outgroup taxa, from (a) plastid and (b) mitochondrial unpartitioned nucleotide analyses.

APPENDIX S23. Phylogram of land-plant relationships taken from the unpartitioned plastid nucleotide analysis, with outgroup algal taxa removed for clarity.

APPENDIX S24. Phylogram of land-plant relationships taken from the unpartitioned mitochondrial nucleotide analysis, with outgroup algal taxa removed for clarity.

APPENDIX S25. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned plastid nucleotide analysis, with C-to-U and U-to-C RNA edit sites from *Takakia lepidozioides* and *Nothoceros aenigmaticus* excluded.

APPENDIX S26. Cladogram of land-plant relationships relative to the algal outgroup, inferred from the unpartitioned mitochondrial nucleotide analysis, with C-to-U and U-to-C RNA edit sites from *Takakia lepidozioides* and *Nothoceros aenigmaticus* excluded.

APPENDIX S27. Bar chart quantifying the volumes of genic (in blue), intergenic spacer (grey) and intron (orange) sequence in each plastome compared in Fig. 9, with one copy of the inverted repeat region removed.

APPENDIX S28. Plastid gene content and size for taxa included in Fig. 9.

APPENDIX S29. Results for the PAML branch tests and the RELAX tests for *Aneura mirabilis* relative to other liverworts, and *Buxbaumia aphylla* relative to other mosses for each gene/gene complex.

APPENDIX S30. Branch test results for individual plastid genes in *Aneura mirabilis* and *Buxbaumia aphylla*.

LITERATURE CITED

Adams, K. L., K. Song, P. G. Roessler, J. M. Nugent, J. L. Doyle, J. J. Doyle, and J. D. Palmer. 1999. Intracellular gene transfer in action: dual transcription and multiple silencing of nuclear and mitochondrial *cox2* genes in legumes. *Proceedings of the National Academy of Sciences, USA* 96: 13863–13868.

Adams, K. L., Y.-L. Qiu, M. Stoutemyer, and J. D. Palmer. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences, USA* 99: 9905–9912.

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

Atluri, S., S. N. Rampersad, and L. Bonen. 2015. Retention of functional genes for S19 ribosomal protein in both the mitochondrion and nucleus for over 60 million years. *Molecular Genetics and Genomics* 290: 2325–2333.

Baldauf, S. L., and J. D. Palmer. 1990. Evolutionary transfer of the chloroplast *tuf*A gene to the nucleus. *Nature* 344: 262–265.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires, and C. Santos. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* 31: 3095–3112.

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistics Society, B* 57: 289–300.

Bidartondo, M. I. 2005. The evolutionary ecology of myco-heterotrophy. *New Phytologist* 167: 335–352.

Bidartondo, M.I., T.D. Bruns, M. Weiss, C. Sérgio, and D.J. Read. 2003. Specialized cheating of the ectomycorrhizal symbiosis by an epiparasitic liverwort. *Proceedings of the Royal Society of London, B, Biological Sciences* 270: 835–842.

Bowe, L. M., and C. W. dePamphilis. 1996. Effects of RNA editing and gene processing on phylogenetic reconstruction. *Molecular Biology and Evolution* 13: 1159–1166.

Braukmann, T. W. A., M. B. Broe, S. Stefanović, and J. V. Freudenstein. 2017. On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytologist* 216: 254–266.

Burleigh, J. G., K. W. Hilu, and D. E. Soltis. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evolutionary Biology* 9: 61.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, and T. L. Madden. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10: 421.

Campbell, D. H. 1895. The structure and development of mosses and ferns (Archegoniatae). Macmillan and Co., London, UK.

Chang, Y., and S. W. Graham. 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *American Journal of Botany* 98: 839–849.

Chang, Y., and S. W. Graham. 2014. Patterns of clade support across the major lineages of moss phylogeny. *Cladistics* 30: 590–606.

Cheng, S., M. Melkonian, S. A. Smith, S. Brockington, J. M. Archibald, P.-M. Delaux, F.-W. Li, et al. 2018. 10KP: A phylodiverse genome sequencing plan. *GigaScience* 7: giy013.

Civáň, P., P. G. Foster, M. T. Embley, A. Séneca, and C. J. Cox. 2014. Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. *Genome Biology and Evolution* 6: 897–911.

Cox, C. J. 2018. Land plant molecular phylogenetics: A review with comments on evaluating incongruence among phylogenies. *Critical Reviews in Plant Sciences*. 37: 113–127.

Cox, C. J., B. Goffinet, A. J. Shaw, and S. B. Boles. 2004. Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. *Systematic Botany* 29: 234–250.

Cox, C. J., B. Goffinet, N. J. Wickett, S. B. Boles, and A. J. Shaw. 2010. Moss diversity: a molecular phylogenetic analysis of genera. *Phytotaxa* 9: 175–195.

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

Crandall-Stotler, B. J. 1980. Morphogenetic designs and a theory of bryophyte origins and divergence. *BioScience* 30: 580–585.

Crandall-Stotler, B. J., L. L. Forrest, and R. E. Stotler. 2005. Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). *Taxon* 54: 299–316.

Crandall-Stotler, B. J., R. E. Stotler, and D. G. Long. 2009. Morphology and classification of the Marchantiophyta. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology. 2nd ed. 1–54. Cambridge University Press, Cambridge UK.

Crum, H. A. 2001. Structural Diversity of Bryophytes. University of Michigan Herbarium, Ann Arbor, Michigan, USA.

Cusimano, N., and S. Wicke. 2016. Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytologist* 210: 680–693.

Darling, A. E., B. Mau, F. R. Blattner, and N. T. Perna. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Research* 14: 1394–1403.

Darling, A. E., B. Mau, and N. T. Perna. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE* 5: e11147.

de Sousa, F., P. G. Foster, P. C. J. Donoghue, H. Schneider, and C. J. Cox. 2019. Nuclear protein phylogenies support the monophyly of the three bryophyte groups (Bryophyta Schimp.). *New Phytologist* 222: 565–575.

Delwiche, C. F., K. G. Karol, M. T. Cimino, and K. J. Sytsma. 2002. Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene *rbc*L. *Journal of Phycology* 38: 394–403.

Doyle, J. A. 2017. Phylogenetic analysis and morphological innovations in land plants. *Annual Plant Reviews* 45: 1–50.

Drouin, G., H. Daoud, and J. Xia. 2008. Relative rates of synonymous substitutions in the mitochondrial, chloroplast and nuclear genomes of seed plants. *Molecular Phylogenetics and Evolution* 49: 827–831.

Duckett, J. G., J. Burch, P. W. Fletcher, H. W. Matcham, D. J. Read, A. J. Russell, and S. Pressel. 2004. In vitro cultivation of bryophytes: a review of practicalities, problems, progress and promise. *Journal of Bryology* 26: 3–20.

Duff, R. J. 2006. Divergent RNA editing frequencies in hornwort mitochondrial *nad*5 sequences. *Gene* 366: 285–291.

Duff, R. J., and F. B.-G. Moore. 2005. Pervasive RNA editing among hornwort *rbc*L transcripts except *Leiosporoceros*. *Journal of Molecular Evolution* 61: 571–578.

Duff, R. J., and D. L. Nickrent. 1999. Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. *American Journal of Botany* 86: 372–386.

Duff, R. J., D. C. Cargill, J. C. Villarreal, and K. S. Renzaglia. 2004. Phylogenetic relationships of the hornworts based on *rbc*L sequence data: novel relationships and new insights. *Monographs in Systematic Botany from the Missouri Botanical Garden* 98: 41–58.

Duff, R. J., J. C. Villarreal, D. C. Cargill, and K. S. Renzaglia. 2007. Progress and challenges toward developing a phylogeny and classification of the hornworts. *The Bryologist* 110: 214–243.

Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Emms, D. M., and S. Kelly. 2017. STRIDE: species tree root inference from gene duplication events. *Molecular Biology and Evolution* 34: 3267–3278.

Felix, H. 1988. Fungi on bryophytes. *Botanica Helvetica* 98: 239–269.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

Finet, C., R. E. Timme, C. F. Delwiche, and F. Marlétaz. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology* 20: 2217–2222.

Fiz-Palacios, O., H. Schneider, J. Heinrichs, and V. Savolainen. 2011. Diversification of land plants: insights from a family-level phylogenetic analysis. *BMC Evolutionary Biology* 11: 341.

Floyd, S. K., C. S. Zalewski, and J. L. Bowman. 2006. Evolution of class III homeodomain-leucine zipper genes in streptophytes. *Genetics* 173: 373–388.

Forrest, L. L., E. C. Davis, D. G. Long, B. J. Crandall-Stotler, A. Clark, and M. L. Hollingsworth. 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): Multiple taxa, genomes and analysis. *The Bryologist* 109: 303–334.

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

Freyer, R., M.-C. Kiefer-Meyer, and H. Kössel. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proceedings of the National Academy of Sciences, USA* 94: 6285–6290.

Gao, L., Y. J. Su, and T. Wang. 2010. Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. *Journal of Systematics and Evolution* 48: 77–93.

Gawryluk, R. M. R., D. V. Tikhonenkov, E. Hehenbeger, F. Husnik, A. P. Mylnikov, and P. J. Keeling. 2019. Non-photosynthetic predators are sister to red algae. *Nature* 572: 240–243.

Gensel, P. G. 2008. The earliest land plants. *Annual Review of Ecology, Evolution, and Systematics* 39: 459–477.

Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany* 105: 291–301.

Goffinet, B., C. J. Cox, A. J. Shaw, and T. A. J. Hedderson. 2001. The Bryophyta (mosses): systematic and evolutionary inferences from an *rps*4 gene (cpDNA) phylogeny. *Annals of Botany* 87: 191–208.

Goffinet, B., N. J. Wickett, O. Werner, R. M. Ros, A. J. Shaw, and C. J. Cox. 2007. Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). *Annals of Botany* 99: 747–753.

Goffinet, B., W. R. Buck, and A. J. Shaw. 2009. Morphology, anatomy, and classification of the Bryophyta. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology. 2nd ed. 55–138. Cambridge University Press, Cambridge, UK.

Goremykin, V. V., and F. H. Hellwig. 2005. Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. *Plant Systematics and Evolution* 254: 93–103.

Goulding, S. E., R. G. Olmstead, C. W. Morden, and K. H. Wolfe. 1996. Ebb and flow of the chloroplast inverted repeat. *Molecular and General Genetics* 252: 195–206.

Graham, S. W., and W. J. D. Iles. 2009. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. *American Journal of Botany* 96: 216–227.

Graham, S. W., J. R. Kohn, B. R. Morton, J. E. Eckenwalder, and S. C. H. Barrett. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Systematic Biology* 47: 545–567.

Graham, S. W., P. A. Reeves, A. C. E. Burns, and R. G. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83–S96.

Graham, S. W., R. G. Olmstead, and S. C. H. Barrett. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution* 19: 1769–1781.

Graham, S. W., J. M. Zgurski, M. A. McPherson, D. M. Cherniawsky, J. M. Saarela, E. F. C. Horne, S. Y. Smith, et al. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: Comparative biology and evolution (excluding Poales), 3–21. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

Graham, S. W., V. K. Y. Lam, and V. S. F. T. Merckx. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist* 214: 48–55.

Heath, T. A., S. M. Hedtke, and D. M. Hillis. 2008. Taxon sampling and the accuracy of phylogenetic analyses. *Journal of Systematics and Evolution* 46: 239–257.

Hedderson, T. A., R. L. Chapman, and W. L. Rootes. 1996. Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Systematics and Evolution* 200: 213–224.

Hedtke, S. M., T. M. Townsend, and D. M. Hillis. 2006. Resolution of phylogenetic conflict in large data sets by increased taxon sampling. *Systematic Biology* 55: 522–529.

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

Hendy, M. D., and D. Penny. 1989. A framework for quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

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

Huelsenbeck, J. P., J. J. Bull, and C. W. Cunningham. 1996. Combining data in phylogenetic analysis. *Trends in Ecology & Evolution* 11: 152–158.

Huttunen, S., N. Bell, V. K. Bobrova, V. Buchbender, W. R. Buck, C. J. Cox, B. Goffinet, et al. 2012. Disentangling knots of rapid evolution: origin and diversification of the moss order Hypnales. *Journal of Bryology* 34: 187–211.

Jiang, W., S.-Y. Chen, H. Wang, D.-Z. Li, and J. J. Wiens. 2014. Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution* 80: 308–318.

Joyce, E. M., D. M. Crayn, V. K. Y. Lam, W. K. Gerelle, S. W. Graham, and L. Nauheimer. 2018. Evolution of *Geosiris* (Iridaceae): historical biogeography and plastid-genome evolution in a genus of non-photosynthetic tropical rainforest herbs disjunct across the Indian Ocean. *Australian Systematic Botany* 31: 504–523.

Karol, K. G., R. M. McCourt, M. T. Cimino, and C. F. Delwiche. 2001. The closest living relatives of land plants. *Science* 294: 2351–2353.

Karol, K. G., K. Arumuganathan, J. L. Boore, A. M. Duffy, K. D. E. Everett, J. D. Hall, S. K. Hansen, et al. 2010. Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. *BMC Evololutionary Biology* 10: 321.

Kenrick, P., and P. R. Crane. 1997a. The origin and early diversification of land plants. A cladistic study. Smithsonian Institution Press, Washington, DC, USA.

Kenrick, P., and P. R. Crane. 1997b. The origin and early evolution of plants on land. *Nature* 389: 33–39.

Kim, H. T., M. G. Chung, and K.-J. Kim. 2014. Chloroplast genome evolution in early diverged leptosporangiate ferns. *Molecules and Cells* 37: 372–382.

Kugita, M., Y. Yamamoto, T. Fujikawa, T. Matsumoto, and K. Yoshinaga. 2003a. RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Research* 31: 2417–2423.

Kugita, M., A. Kaneko, Y. Yamamoto, Y. Takeya, T. Matsumoto, and K. Yoshinaga. 2003b. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. *Nucleic Acids Research* 31: 716–721.

Lam, V. K. Y., M. Soto Gomez, and S. W. Graham. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lam, V. K. Y., V. S. F. T. Merckx, and S. W. Graham. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *American Journal of Botany* 103: 692–708.

Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. M. Neubig, J. R. Abbott, G. E. Beatty, J. Provan, M. Soto Gomez, and S. W. Graham. 2018. Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany* 105: 480–494.

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

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

Larsson, A. 2014. AliView: a fast and lightweight alignment viewer and editor for large data sets. *Bioinformatics* 30: 3276–3278.

Leake, J. R. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist* 127: 171–216.

Leliaert, F., and J. M. Lopez-Bautista. 2015. The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditionis* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *BMC Genomics* 16: 204.

Lemieux, C., C. Otis, and M. Turmel. 2007. A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biology* 5: 2.

Lemieux, C., C. Otis, and M. Turmel. 2014. Six newly sequenced chloroplast genomes from prasinophyte green algae provide insights into the relationships among prasinophyte lineages and the diversity of streamlined genome architecture in picoplanktonic species. *BMC Genomics* 15: 857.

Lemieux, C., C. Otis, and M. Turmel. 2016. Comparative chloroplast genome analyses of streptophyte green algae uncover major structural alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. *Frontiers in Plant Science* 7: 697.

Lenz, H., and V. Knoop. 2013. PREPACT 2.0: Predicting C-to-U and U-to-C RNA editing in organelle genome sequences with multiple references and curated RNA editing annotation. *Bioinformatics and Biology Insights* 7: 1–19.

Lewis, L., B. D. Mishler, and R. Vilgalys. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbc*L. *Molecular Phylogenetics and Evolution* 7: 377–393.

Li, L., B. Wang, Y. Liu, and Y.-L. Qiu. 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *Journal of Molecular Evolution* 68: 665–678.

Ligrone, R., J. G. Duckett, and K. S. Renzaglia. 2012. Major transitions in the evolution of early land plants: A bryological perspective. *Annals of Botany* 109: 851–871.

Liu, Y., C. J. Cox, W. Wang, and B. Goffinet. 2014a. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* 63: 862–878.

Liu, Y., R. Medina, and B. Goffinet. 2014b. 350 My of mitochondrial genome stasis in mosses, an early land plant lineage. *Molecular Biology and Evolution* 31: 2586–2591.

Liu, Y., M. G. Johnson, C. J. Cox, R. Medina, N. Devos, A. Vanderpoorten, L. Hedenäs, et al. 2019. Resolution of the backbone phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communications* 10: 1485.

Logacheva, M. D., M. I. Schelkunov, M. S. Nuraliev, T. H. Samigullin, and A. A. Penin. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.

Logacheva, M. D., M. I. Schelkunov, V. Y. Shtratnikova, M. V. Matveeva, and A. A. Penin. 2016. Comparative analysis of plastid genomes of non-photosynthetic Ericaceae and their photosynthetic relatives. *Scientific Reports* 6: 30042.

Magallón, S., K. W. Hilu, and D. Quandt. 2013. Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany* 100: 556–573.

Magombo, Z. L. K. 2003. The phylogeny of basal peristomate mosses: evidence from cpDNA, and implications for peristome evolution. *Systematic Botany* 28: 24–38.

Matasci, N., L.-H. Hung, Z. Yan, E. J. Carpenter, N. J. Wickett, S. Mirarab, N. Nguyen, et al. 2014. Data access for the 1,000 Plants (1KP) project. *GigaScience* 3: 17.

Mathews, S., and M. J. Donoghue. 1999. The root of angiosperm phylogeny inferred from duplicate phytochrome genes. *Science* 286: 947–950.

Merckx, V.S.F.T., J.V. Freudenstein, J. Kissling, M.J.M. Christenhusz, R.E. Stotler, B.J. Crandall-Stotler, N. Wickett, et al. 2013a. Taxonomy and classification.

In V.S.F.T. Merckx [ed.] Mycoheterotrophy: the biology of plants living on fungi. 19–101. Springer-Verlag, New York, USA.

Merckx, V.S.F.T., C.B. Mennes, K.G. Peay, and J. Geml. 2013b. Evolution and diversification. *In* V.S.F.T. Merckx [ed.] Mycoheterotrophy: the biology of plants living on fungi. 222–226. Springer-Verlag, New York, USA.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop. *GCE* 2010: 1–8.

Mishler, B. D., and S. P. Churchill. 1984. A cladistic approach to the phylogeny of the "bryophytes". *Brittonia* 36: 406–424.

Mishler, B. D., L. A. Lewis, M. A. Buchheim, K. S. Renzaglia, D. J. Garbary, C. F. Delwiche, F. W. Zechman, et al. 1994. Phylogenetic relationships of the "green algae" and "bryophytes". *Annals of the Missouri Botanical Garden* 81: 451–483.

Morris, J. L., M. N. Puttick, J. W. Clark, D. Edwards, P. Kenrick, S. Pressel, C. H. Wellman, et al. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* 115: E2274–E2283.

Mower, J. P. 2009. The PREP suite: predictive RNA editors for plant mitochondrial genes, chloroplast genes and user-defined alignments. *Nucleic Acids Research* 37: W253–W259.

Mower, J. P., and T. L. Vickrey. 2018. Structural diversity among plastid genomes of land plants. *In* S.-M. Chaw and R. K. Jansen [eds.], Advances in Botanical Research. Vol. 85, 263–292. Academic Press, Cambridge, MA, USA.

Murdock, A. G. 2008. Phylogeny of marattioid ferns (Marattiaceae): Inferring a root in the absence of a closely related outgroup. *American Journal of Botany* 95: 626–641.

Newton, A. E., C. J. Cox, J. G. Duckett, J. A. Wheeler, B. Goffinet, T. A. J. Hedderson, and B. D. Mishler. 2000. Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. *The Bryologist* 103: 187–211.

Nickrent, D. L., C. L. Parkinson, J. D. Palmer, and R. J. Duff. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution* 17: 1885–1895.

Nishiyama, T., and M. Kato. 1999. Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. *Molecular Biology and Evolution* 16: 1027–1036.

Nishiyama, T., P. G. Wolf, M. Kugita, R. B. Sinclair, M. Sugita, C. Sugiura, T. Wakasugi, et al. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Molecular Biology and Evolution* 21: 1813–1819.

One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and phylogenomics of green plants. *Nature* 574: 679–685.

Palmer, J. D., and L. A. Herbon. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* 28: 87–97.

Park, M., H. Park, H. Lee, B. Lee, and J. Lee. 2018. The complete plastome sequence of an Antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. *International Journal of Molecular Sciences* 19: 709.

Petersen, G., A. Cuenca, and O. Seberg. 2015. Plastome evolution in hemiparasitic Mistletoes. *Genome Biology and Evolution* 7: 2520–2532.

Philippe, H., H. Brinkmann, D. V. Lavrov, D. T. J. Littlewood, M. Manuel, G. Wörheide, and D. Baurain. 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. *PLoS Biology* 9: e1000602.

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Systematic Biology* 51: 664–671.

Pruchner, D., B. Nassal, M. Schindler, and V. Knoop. 2001. Mosses share mitochondrial group II introns with flowering plants, not with liverworts. *Molecular Genetics and Genomics* 266: 608–613.

Puttick, M. N., J. L. Morris, T. A. Williams, C. J. Cox, D. Edwards, P. Kenrick, S. Pressel, et al. 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology* 28: 733–745.

Qiu, Y.-L. 2008. Phylogeny and evolution of charophytic algae and land plants. *Journal of Systematics and Evolution* 46: 287–306.

Qiu, Y.-L., Y. R. Cho, J. C. Cox, and J. D. Palmer. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671–674.

Qiu, Y.-L., L. Li, B. Wang, Z. Chen, V. Knoop, M. Groth-Malonek, O. Dombrovska, et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences, USA* 103: 15511–15516.

Qiu, Y.-L., L. Li, B. Wang, Z. Chen, O. Dombrovska, J. Lee, L. Kent, et al. 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 168: 691–708.

Qiu, Y., S. J. Filipenko, A. Darracq, and K. L. Adams. 2014. Expression of a transferred nuclear gene in a mitochondrial genome. *Current Plant Biology* 1: 68–72.

Read, D. J., J. G. Duckett, R. Francis, R. Ligrone, and A. Russell. 2000. Symbiotic fungal associations in 'lower' land plants. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 355: 815–831.

Renzaglia, K. S., R. J. Duff, D. L. Nickrent, and D. J. Garbary. 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. *Philosophical Transactions of the Royal Society, B, Biological Sciences* 355: 769–793.

Renzaglia, K. S., and D. J. Garbary. 2001. Motile gametes of land plants: diversity, development, and evolution. *Critical Reviews in Plant Sciences* 20: 107–213.

Renzaglia, K. S., J. C. Villarreal, and R. J. Duff. 2009. New insights into morphology, anatomy, and systematics of hornworts. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology. 2nd ed. 139–171. Cambridge University Press, Cambridge, UK.

Robison, T. A., and P. G. Wolf. 2019. ReFernment: An R package for annotating RNA editing in plastid genomes. *Applications in Plant Sciences* 7: e1216.

Rothfels, C. J., and E. Schuettpelz. 2014. Accelerated rate of molecular evolution for vittarioid ferns is strong and not driven by selection. *Systematic Biology* 63: 31–54.

Rothfels, C. J., A. Larsson, L.-Y. Kuo, P. Korall, W.-L. Chiou, and K. M. Pryer. 2012. Overcoming deep roots, fast rates, and short internodes to resolve the ancient rapid radiation of eupolypod II ferns. *Systematic Biology* 61: 490–509.

Rothfels, C. J., F.-W. Li, E. M. Sigel, L. Huiet, A. Larsson, D. O. Burge, M. Ruhsam, et al. 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *American Journal of Botany* 102: 1089–1107.

Rothwell, G. R., W. L. Crepet, and R. A. Stockey. 2009. Is the anthophyte hypothesis alive and well? New evidence from the reproductive structures of Bennettitales. *American Journal of Botany* 96: 296–322.

Rüdinger, M., U. Volkmar, H. Lenz, M. Groth-Malonek, and V. Knoop. 2012. Nuclear DYW-type PPR gene families diversify with increasing RNA editing frequencies in liverwort and moss mitochondria. *Journal of Molecular Evolution* 74: 37–51.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms- inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Schallenberg-Rüdinger, M., and V. Knoop. 2016. Coevolution of organelle RNA editing and nuclear specificity factors in early land Plants. *In* S.A. Rensing [ed.] Advances in Botanical Research. Vol. 78. 37–93. Academic Press; Cambridge, MA, USA.

Schofield, W. B. 1985. Introduction to Bryology. Blackburn Press, New Jersey, USA.

Schofield, W.B. 2007. *Buxbaumia*. *In* Flora of North America Editorial Committee [eds.] Flora of North America North of Mexico. Vol. 27, 118–120. Oxford University Press; New York and Oxford.

Schwender, J., F. Goffman, J. B. Ohlrogge, and Y. Shachar-Hill. 2004. Rubisco without the Calvin cycle improves the carbon efficiency of developing green seeds. *Nature* 432: 779–782.

Shanker, A., V. Sharma, and H. Daniell. 2011. Phylogenomic evidence of bryophytes' monophyly using complete and incomplete data sets from chloroplast proteomes. *Journal of Plant Biochemistry and Biotechnology* 20: 288–292.

Shaw, A. J., C. J. Cox, B. Goffinet, W. R. Buck, and S. B. Boles. 2003. Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta). *Evolution* 57: 2226–2241.

Shaw, A. J., N. Devos, Y. Liu, C. J. Cox, B. Goffinet, K. I. Flatberg, and B. Shaw. 2016. Organellar phylogenomics of an emerging model system: *Sphagnum* (peatmoss). *Annals of Botany* 118: 185–196.

Simmons, M. P., and P. A. Goloboff. 2014. Dubious resolution and support from published sparse supermatrices: the importance of thorough tree searches. *Molecular Phylogenetics and Evolution* 78: 334–348.

Simmons, M. P., C. D. Bailey, and K. C. Nixon. 2000. Phylogeny reconstruction using duplicate genes. *Molecular Biology and Evolution* 17: 469–473.

Smith, G.M. 1955. Cryptogamic Botany, Vol. II, Bryophytes and Pteridophytes. 2nd ed. McGraw-Hill, New York, USA.

Smith, S. A., J. M. Beaulieu, and M. J. Donoghue. 2009. Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. *BMC Evolutionary Biology* 9: 37.

Soltis, P. S., and D. E. Soltis. 2003. Applying the bootstrap in phylogeny reconstruction. *Statistical Science* 18: 256–267.

Sper-Whitis, G. L., J. L. Moody, and J. C. Vaughn. 1996. Universality of mitochondrial RNA editing in cytochrome-*c* oxidase subunit I (*coxI*) among the land plants. *Biochimica et Biophysica Acta* 1307: 301–308.

Stamatakis, A. 2014. RAxML Version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web-servers. *Systematic Biology* 75: 758–771.

Steane, D. A., R. W. Scotland, D. J. Mabberley, and R. G. Olmstead. 1999. Molecular systematics of *Clerodendrum* (Lamiaceae): ITS sequences and total evidence. *American Journal of Botany* 86: 98–107.

Stech, M., D. Quandt, and W. Frey. 2003. Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA *trn*L-*trn*F region. *Journal of Plant Research* 116: 389–398.

Steinhauser, S., S. Beckert, I. Capesius, O. Malek, and V. Knoop. 1999. Plant mitochondrial RNA editing. *Journal of Molecular Evolution* 48: 303–312.

Sugita, M., Y. Miyata, K. Maruyama, C. Sugiura, T. Arikawa, and M. Higuchi. 2006. Extensive RNA editing in transcripts from the *psb*B operon and *rpo*A gene of plastids from the enigmatic moss *Takakia lepidozioides*. *Bioscience, Biotechnology, and Biochemistry* 70: 2268–2274.

Swain, T. D. 2018. Revisiting the phylogeny of Zanthidea (Cnidaria: Anthozoa): staggered alignment of hypervariable sequences improves species tree inference. *Molecular Phylogenetics and Evolution* 118: 1–12.

Takenaka, M., A. Zehrmann, D. Verbitsky, B. Härtel, and A. Brennicke. 2013. RNA editing in plants and its evolution. *Annual Review of Genetics* 47: 335–352.

Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22: 4673–4680.

Timme, R. E., T. R. Bachvaroff, and C. F. Delwiche. 2012. Broad phylogenomic sampling and the sister lineage of land plants. *PLoS ONE* 7: e29696.

Turmel, M., C. Otis, and C. Lemieux. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molecular Biology and Evolution* 23: 1324–1338.

Turmel, M., C. Otis, and C. Lemieux. 2013. Tracing the evolution of streptophyte algae and their mitochondrial genome. *Genome Biology and Evolution* 5: 1817–1835.

Vanderpoorten, A., and B. Goffinet. 2009. Introduction to Bryophytes. Cambridge University Press, Cambridge, UK.

Villarreal, J. C., and S. S. Renner. 2012. Hornwort pyrenoids, carbon-concentrating structures, evolved and were lost at least five times during the last 100 million years. *Proceedings of the National Academy of Sciences, USA* 109: 18873–18878.

Villarreal, J. C., B. Goffinet, R. J. Duff, and D. C. Cargill. 2010. Phylogenetic delineation of the genera *Nothoceros* and *Megaceros*. *The Bryologist* 113: 106–113.

Villarreal, J. C., L. L. Forrest, N. J. Wickett, and B. Goffinet. 2013. The plastid genome of the hornwort *Nothoceros aenigmaticus* (Dendrocerotaceae): Phylogenetic signal in inverted repeat expansion, pseudogenization, and intron gain. *American Journal of Botany* 103: 467–477.

Villarreal, J. C., M. Turmel, M. Bourgouin-Couture, J. Laroche, N. Salazar Allen, F.-W. Li, S. Cheng, et al. 2018. Genome-wide organellar analyses from the hornwort *Leiosporoceros dussii* show low frequency of RNA editing. *PLoS ONE* 13: e0200491.

Volkmar, U., and V. Knoop. 2010. Introducing intron locus cox1i624 for phylogenetic analyses in bryophytes: on the issue of *Takakia* as sister genus to all other extant mosses. *Journal of Molecular Evolution* 70: 506–518.

Waters, D. A., M. A. Buchheim, R. A. Dewey, and R. L. Chapman. 1992. Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. *American Journal of Botany* 79: 459–466.

Wertheim, J. O., B. Murrell, M. D. Smith, S. L. Kosakovsky Pond, and K. Scheffler. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Molecular Biology and Evolution* 32: 820–832.

Wicke, S., G. M. Schneeweiss, C. W. dePamphilis, K. F. Müller, and D. Quandt. 2011. The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Molecular Biology* 76: 273–297.

Wicke, S., K. F. Müller, C. W. dePamphilis, D. Quandt, S. Bellot, and G. Schneeweiss. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences, USA* 113: 9045–9050.

Wickett, N. J., Y. Zhang, S. K. Hansen, J. M. Roper, J. V. Kuehl, S. A. Plock, P. G. Wolf, et al. 2008. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort *Aneura mirabilis*. *Molecular Biology and Evolution* 25: 393–401.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.

Wiens, J. J. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? *Systematic Biology* 54: 731–742.

Wiens, J. J. 2006. Missing data and the design of phylogenetic analyses. *Journal of Biomedical Informatics* 39: 34–42.

Wiens, J. J., and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PloS ONE* 7: e42925.

Wodniok, S., H. Brinkmann, G. Glöckner, A. J. Heidel, H. Philippe, M. Melkonian, and B. Becker. 2011. Origin of land plants: Do conjugating green algae hold the key? *BMC Evolutionary Biology* 11: 104.

Wolf, P. G., K. G. Karol, D. F. Mandoli, J. Kuehl, K. Arumuganathan, M. W. Ellis, B. D. Mishler, et al. 2005. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene* 350: 117–128.

Wolfe, K. H., W. H. Li, and P. M. Sharp. 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences, USA* 84: 9054–9058.

Wu, Z., D. B. Sloan, C. W. Brown, M. Rosenblueth, J. D. Palmer, and H. C. Ong. 2017. Mitochondrial retroprocessing promoted functional transfers of *rpl5* to the nucleus in grasses. *Molecular Biology and Evolution* 34: 2340–2354.

Xie, Y., G. Wu, J. Tang, R. Luo, J. Patterson, S. Liu, W. Huang, et al. 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. *Bioinformatics* 30: 1660–1666.

Xue, Y., Y. Liu, L. Li, B. Wang, and Y.-L. Qiu. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics* 56: 53–61.

Yang, Z. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular Biology and Evolution* 15: 568–573.

Yang, Z. 2007. PAML4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* 24: 1586–1591.

Yura, K., Y. Miyata, T. Arikawa, M. Higuchi, and M. Sugita. 2008. Characteristics and prediction of RNA editing sites in transcripts of the moss *Takakia lepidozioides* chloroplast. *DNA Research* 15: 309–321.

Zhong, B., Z. Xi, V. V. Goremykin, R. Fong, P. A. McLenachan, P. M. Novis, C. C. Davis, and D. Penny. 2013. Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. *Molecular Biology and Evolution* 31: 177–183.