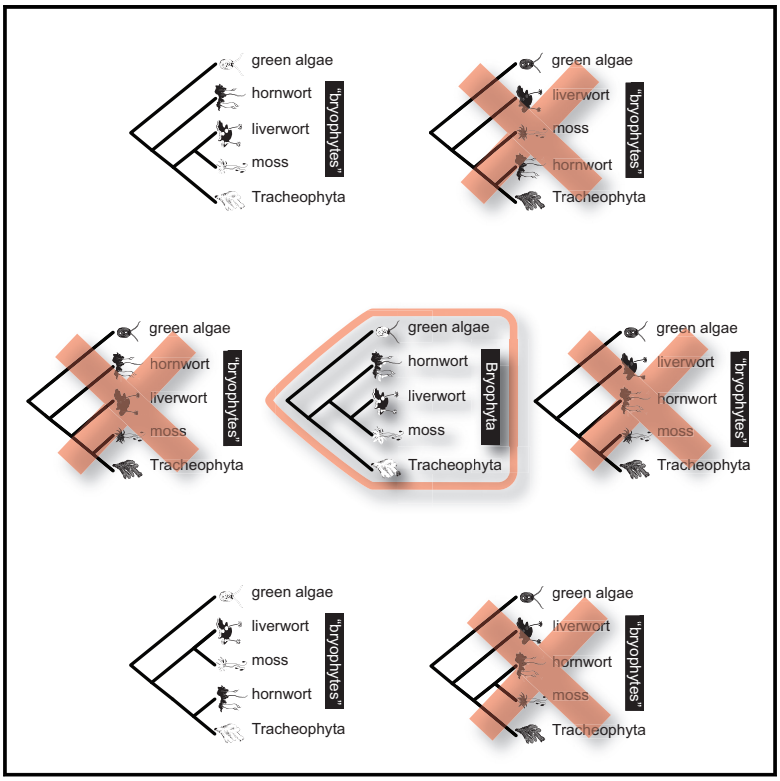
Article

Current Biology

The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte

Graphical Abstract



Highlights

- Early land plant relationships are extremely uncertain
- We resolve the "Setaphyta" clade of liverworts plus mosses
- The simple body plan of liverworts results from loss of ancestral characters
- The ancestral land plant was more complex

Authors

Mark N. Puttick, Jennifer L. Morris, Tom A. Williams, ..., Harald Schneider, Davide Pisani, Philip C.J. Donoghue

Correspondence

harald@xtbg.ac.cn (H.S.), davide.pisani@bristol.ac.uk (D.P.), phil.donoghue@bristol.ac.uk (P.C.J.D.)

In Brief

Puttick et al. resolve a "Setaphyta" clade uniting liverworts and mosses and support for bryophyte monophyly. Their results indicate that the ancestral land plant was more complex than has been envisaged based on phylogenies recognizing liverworts as the sister lineage to all other embryophytes.

Puttick et al., 2018, Current Biology *28*, 733–745
March 5, 2018
https://doi.org/10.1016/j.cub.2018.01.063

CellPress

Current Biology
Article

The Interrelationships of Land Plants and the Nature of the Ancestral Embryophyte

Mark N. Puttick,[1,2,3,9] Jennifer L. Morris,[1,4,9] Tom A. Williams,[2] Cymon J. Cox,[5] Dianne Edwards,[4] Paul Kenrick,[6] Silvia Pressel,[3] Charles H. Wellman,[7] Harald Schneider,[3,8,*] Davide Pisani,[1,3,*] and Philip C.J. Donoghue[1,10,*]
[1]School of Earth Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[2]School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
[3]Department of Life Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
[4]School of Earth and Ocean Sciences, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK
[5]Centro de Ciências do Mar, Universidade do Algarve, Gambelas, 8005-319 Faro, Portugal
[6]Department of Earth Sciences, The Natural History Museum, Cromwell Road, London SW7 5BD, UK
[7]Department of Animal and Plant Sciences, University of Sheffield, Alfred Denny Building, Western Bank, Sheffield S10 2TN, UK
[8]Center of Integrative Conservation, Xishuangbanna Tropical Botanical Garden, Chinese Academy of Sciences, Menglun, Yunnan, China
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: harald@xtbg.ac.cn (H.S.), davide.pisani@bristol.ac.uk (D.P.), phil.donoghue@bristol.ac.uk (P.C.J.D.)
https://doi.org/10.1016/j.cub.2018.01.063

SUMMARY

The evolutionary emergence of land plant body plans transformed the planet. However, our understanding of this formative episode is mired in the uncertainty associated with the phylogenetic relationships among bryophytes (hornworts, liverworts, and mosses) and tracheophytes (vascular plants). Here we attempt to clarify this problem by analyzing a large transcriptomic dataset with models that allow for compositional heterogeneity between sites. Zygnematophyceae is resolved as sister to land plants, but we obtain several distinct relationships between bryophytes and tracheophytes. Concatenated sequence analyses that can explicitly accommodate site-specific compositional heterogeneity give more support for a mosses-liverworts clade, "Setaphyta," as the sister to all other land plants, and weak support for hornworts as the sister to all other land plants. Bryophyte monophyly is supported by gene concatenation analyses using models explicitly accommodating lineage-specific compositional heterogeneity and analyses of gene trees. Both maximum-likelihood analyses that compare the fit of each gene tree to proposed species trees and Bayesian supertree estimation based on gene trees support bryophyte monophyly. Of the 15 distinct rooted relationships for embryophytes, we reject all but three hypotheses, which differ only in the position of hornworts. Our results imply that the ancestral embryophyte was more complex than has been envisaged based on topologies recognizing liverworts as the sister lineage to all other embryophytes. This requires many phenotypic character losses and transformations in the liverwort lineage, diminishes inconsistency between phylogeny and the fossil record, and prompts re-evaluation of the phylogenetic affinity of early land plant fossils, the majority of which are considered stem tracheophytes.

INTRODUCTION

The evolutionary emergence of land plant body plans is one of the most formative episodes in the evolution of our planet [1, 2]. Land plant innovations, including stomata, vascular and rooting systems, symbioses with fungi, and, eventually, leaves, expanded the sequestration of CO_2 through photosynthesis and silicate weathering [1]. These weathering effects resulted in changes in atmospheric CO_2 over long timescales [1, 3, 4], as well as key changes to the environment such as the development of soils [5]. Early land plants have even been invoked in shaping terrestrial landscapes by constraining sedimentological processes [6]. Unfortunately, a detailed understanding of this episode is obscured by uncertainty associated with the phylogenetic relationships among bryophytes (hornworts, liverworts, and mosses) and tracheophytes (vascular plants), for which almost every possible solution has been proposed (e.g., [7]; Figures 1A–1G). In the absence of phylogenetic resolution, it is not possible to establish the sequence in which embryophyte, bryophyte, and tracheophyte body plan characters were assembled. This is a prerequisite for determining their intrinsic molecular developmental causes and extrinsic environmental consequences, the phylogenetic interpretation of fossil embryophytes, and, consequently, establishing the timescale over which these characters evolved—facilitating tests of hypotheses on their role in transforming the planet.

Although the monophyly of Embryophyta (land plants) and Tracheophyta (vascular plants) is universally accepted, various hypotheses on the interrelationships of bryophytes (hornworts, liverworts, and mosses) and tracheophytes have gained support in the last 30 years. The fundamental distinction is between bryophyte monophyly (e.g., [7–11]; Figure 1A) and paraphyly (e.g., [12–17]; Figures 1B–1G). Indeed, at least seven alternative

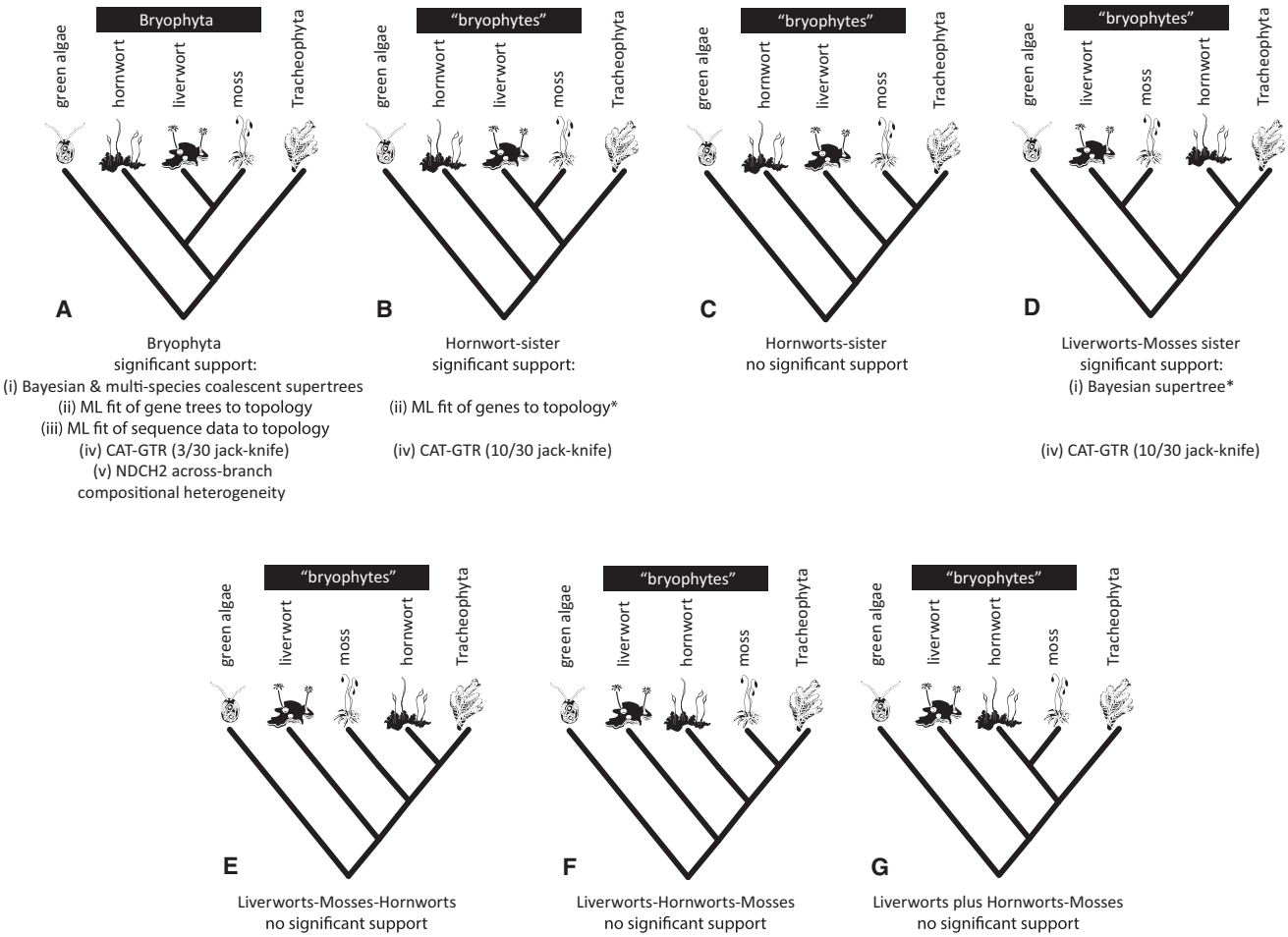


Figure 1. The Seven Competing Topologies Identified between the Principal Embryophyte Lineages and Tracheophytes, Including a Summary of the Support for the Three Main Topologies Identified in the Analyses

The seven principal competing hypotheses of bryophyte and tracheophyte relationships: (A) bryophyte monophyly; (B) hornworts sister to a clade of mosses plus liverworts, itself sister to tracheophytes; (C) mosses, liverworts, and hornworts are successive sister lineages of tracheophytes; (D) a clade of liverworts and mosses as a sister lineage of hornworts plus tracheophytes; (E) hornworts, mosses, and liverworts are successive sister lineages of tracheophytes; (F) mosses, hornworts, and liverworts are successive sister lineages of tracheophytes; and (G) hornworts plus mosses comprise a clade, sister to tracheophytes, and liverworts are an outgroup to all three. There is significant support for monophyletic bryophytes from analyses employing Bayesian supertree estimation from gene trees (i), significant tests of the maximum-likelihood fit of gene trees (ii) and sequence data (iii) to the topologies, and CAT-GTR analyses in Phylobayes (iv). However, both Bayesian supertree estimation (i) and maximum-likelihood fit of gene trees (ii) significantly reject hornworts sister and mosses-liverworts sister, but both of these topologies are found more consistently in jack-knife CAT-GTR analyses than monophyly (iv).

topologies have been proposed from morphological and molecular analyses (Figures 1A–1G; Table 1), leading to the current consensus of a polytomy between mosses, liverworts, hornworts, and tracheophytes [18]. The identity of the embryophyte outgroup remains equally uncertain, with older studies supporting the morphologically complex stoneworts (Charales) as the immediate sister clade and more recent studies supporting the "pond scum" Zygnematophyceae as the land plant sister group (e.g., [7, 19, 20]).

The failure of different studies to reach congruence on the fundamental relationships among embryophytes is most likely due to inadequate phylogenetic models rather than insufficient data [11, 21, 22]. For example, the lack of consensus at the root of embryophytes has been attributed to directional evolution in nucleotide sequences leading to compositional biases [11] and the use of overly simplistic models with assumptions—stationarity, reversibility, and homogeneity—that are violated in real sequence data [22]. Therefore, any attempt to understand these early relationships requires substitution models and analytic methods that can incorporate these complexities.

Here we analyze a large transcriptomic amino acid alignment from 103 species of algae (Chlorophyta and Streptophyta) and Embryophyta (mosses, hornworts, liverworts, and tracheophytes) from Wickett et al. [7] using a model (CAT-GTR+G [23]) that accounts for compositional heterogeneity across sites [23] and a data-recoding strategy (Dayhoff-6) that reduces lineage-specific compositional heterogeneity; the same recoded datasets were also analyzed using the node-discrete composition heterogeneity (NCDH) model [24] that explicitly accounts for

Table 1. Summary of the Topologies Identified between the Principal Embryophyte Lineages, as well as the Algal Outgroup to Embryophyta and the Data Type Analyzed

Embryophyte Sister Group[a]	Data	Reference
Bryophyte Monophyly		
Charales	nuclear rDNA[b]	Hori et al. [56]
Charales	nuclear rDNA[b]	Steele et al. [57]
Coleochaetales/Charales	sperm ultrastructure	Garbary et al. [8]
Coleochaetales	sperm ultrastructure	Mishler et al. [14]
Charales	sperm ultrastructure	Renzaglia et al. [58]
charophytes/chlorophytes	chloroplast amino acids	Nishiyama et al. [9]
Chlorophyta/Coleochaetales	chloroplast nucleotides[b]	Quandt et al. [59]
Coleochaetales	chloroplast nucleotides[c]	Goremykin and Hellwig [10]
Coleochaetales	chloroplast amino acids[b]	Rodríguez-Ezpeleta et al. [60]
Zygnematales	chloroplast amino acids and nucleotides[b]	Turmel et al. [61]
Zygnematales	chloroplast amino acids	Lemieux et al. [62]
Zygnematales	chloroplast amino acids	Cox et al. [11]
Zygnematales	chloroplast amino acids	Civan et al. [63]
Zygnematales	nuclear amino acids	Wickett et al. [7]
(Hornwort, ((Liverwort, Moss), Tracheophyta))		
Coleochaetales	nuclear rDNA	Hedderson et al. [38]
Coleochatales/Klebsormidiales	nuclear rDNA	Hedderson et al. [64]
Coleochaetales	sperm ultrastructure	Garbary and Renzaglia [40]
Chlorophytes	mitochondrial rDNA	Duff and Nickrent [65]
Chlorophyta/Coleochaetales	nuclear and chloroplast nucleotides[c]	Nishiyama and Kato [66]
Charales/Zygnematales	nuclear, mitochondrial, chloroplast nucleotides and nuclear rDNA[c]	Nickrent et al. [67]
Coleochaetales	morphology, nuclear and mitochondrial nucleotides	Renzaglia et al. [58]
Zygnematales	nuclear nucleotides[c]	Wickett et al. [7]
(Hornwort, (Liverwort, (Moss, Tracheophyta)))		
Charales	morphology	Bremer et al. [68]
Coleochaetales	morphology, nuclear rDNA	Mishler et al. [14]
((Liverwort, Moss), (Hornwort, Tracheophyta))		
Coleochaetales	chloroplast nucleotides[d]	Nishiyama et al. [9]
Zygnematales	chloroplast nucleotides	Turmel et al. [61]
Zygnematales	chloroplast nucleotides	Lemieux et al. [62]
Zygnematales	chloroplast nucleotides[c]	Qiu et al. [17]
Zygnematales	chloroplast nucleotides[c]	Gao et al. [20]
Zygnematales	chloroplast nucleotides	Karol et al. [69]
Zygnematales	chloroplast amino acids	Ruhfel et al. [19]
Zygnematales	chloroplast amino acids	Lemieux et al. [70]
(Liverwort, (Moss, (Hornwort, Tracheophyta)))		
"Charophyceae"	morphology	Parenti [71]
Coleochaetes/Charales	chloroplast nucleotides	Lewis et al. [72]
Charales	chloroplast nucleotides	Delwiche et al. [73]
Coleochaetales	chloroplast genome structure	Kelch et al. [74]
Coleochaetales	chloroplast nucleotides[c]	Nishiyama et al. [9]
Coleochaetales	chloroplast nucleotides	Wolf et al. [75]
Charales/Coleochaetales	mitochondrial nucleotides	Groth-Malonek et al. [76]
Charales or Zygnematales	chloroplast nuclecotide and rDNA, and mitochondrial and nuclear rDNA	Qiu et al. [17]
Charales	chloroplast and mitochondrial nucleotides, and nuclear, chloroplast, and mitochondrial rDNA	Qiu et al. [77]

(*Continued on next page*)

Table 1. *Continued*

Embryophyte Sister Group[a]	Data	Reference
Charales	chloroplast nucleotides	Smith et al. [27]
Zygnematales	chloroplast nucleotides	Gao et al. [20]
Zygnematales	chloroplast nucleotides	Chang and Graham [28]
Chlorophytes/charophytes	chloroplast nucleotides and nuclear rDNA	Fiz-Palacios et al. [78]
Coleochaetales	chloroplast nucleotides	Magallón et al. [79]
Charales or Zygnematales	mitochondrial nucleotides	Turmel et al. [80]
Charales/Coleochaetales	chloroplast nucleotides	Kim et al. [81]
Charales	mitochondrial nucleotides and amino acids	Liu et al. [21]
Zygnematales	chloroplast nucleotides	Ruhfel et al. [19]
Zygnematales	chloroplast nucleotides	Zhong et al. [29]
(Liverwort (Hornwort (Moss, Tracheophyta)))		
"charophytes"	morphology	Mishler and Churchill [13]
Charales/Coleochaetales	morphology	Bremer [31]
Coleochaetales	morphology	Bremer et al. [68]
Coleochaetales	morphology and chloroplast and nuclear rRNA	Mishler et al. [14]
Charales/Coleochaetales	morphology	Kenrick and Crane [15]
Charales	chloroplast and mitochondrial nucleotides, nuclear rDNA	Karol et al. [45]
(Liverwort ((Moss, Hornwort), Tracheophyta))		
Klebsormidiales/Coeleochaetales	nuclear rRNA	Waters et al. [12]
Klebsormidiales	chloroplast and nuclear rDNA	Mishler et al. [14]
Chlorophytes	mitochondrial rDNA	Duff and Nickrent [65]
(Moss, (Liverwort, (Hornwort, Tracheophyta)))		
Charales	mitochondrial nucleotides	Liu et al. [21]
Coleochaetales	chloroplast amino acids	Lemieux et al. [82]
(Moss, (Hornwort, (Liverwort, Tracheophyta)))		
Charales	nuclear amino acids	Floyd et al. [83]

[a]Either the designated outgroups or, if identified, the most closely related taxon to land plants.
[b]Note that although bryophytes were monophyletic, they were embedded in a paraphyletic tracheophyte.
[c]Protein coding-genes 1+2 codon positions only [7, 9, 10, 20, 67].
[d]Excluding the nucleotide sites coding leucine and third codon positions and fourfold degenerate sites.

lineage-specific compositional heterogeneity. Finally, we explore the relative fit among sequence data and gene trees to seven proposed hypotheses for the relationships of early land plants (Figures 1A–1G). Analyses under the CAT-GTR model are not conclusive, supporting several topologies. However, bryophyte monophyly is significantly favored over all alternative topologies (1) in superalignment analyses using the NDCH2 model, (2) when comparing the likelihood fit of gene trees and sequences to different hypotheses, and (3) through Bayesian supertree analyses of gene trees.

RESULTS

The results of our phylogenetic analyses overwhelmingly support the clade uniting liverworts and mosses that we name "Setaphyta." To a lesser degree, we find support for a sister relationship between Setaphyta and hornworts and, hence, the monophyly of the bryophytes (Figures 2, 3, and 4). Support for the monophyly of the three bryophyte groups is mainly from the use of hypothesis tests using prior topologies, but also supertree estimation from gene trees.

Compositional Heterogeneity among Sites

The position of hornworts is unresolved in a focal analysis of genes present in at least 95% of taxa (20,512 amino acids). Analysis of these data with the CAT-GTR model yields a phylogeny in which the relationships between the bryophyte phyla and the tracheophytes are unresolved (Figures 3 and 4). Using posterior predictive tests to estimate the number of Dayhoff-recoded amino acids at each site in the alignment, there is a significant difference between the empirical data and those estimated from the CAT-GTR model (*Z* score: 3.85), but this model fit is better than the prediction from the GTR model with no compositional heterogeneity and the empirical data (*Z* score: 27.24). Analyses without the hornworts did not change the relationships among the remaining bryophyte and tracheophyte clades.

Five fully resolved alternative topologies are supported by the jack-knife analyses of the full dataset (Figure 4). The joint most commonly sampled consensus tree (10/30 replicates) from the jack-knife analyses shows hornworts as sister to tracheophytes, with liverworts and mosses comprising a sister clade at the base of embryophytes (Figure 4B). The position of hornworts changes in the remaining phylogenies to be either at the base of

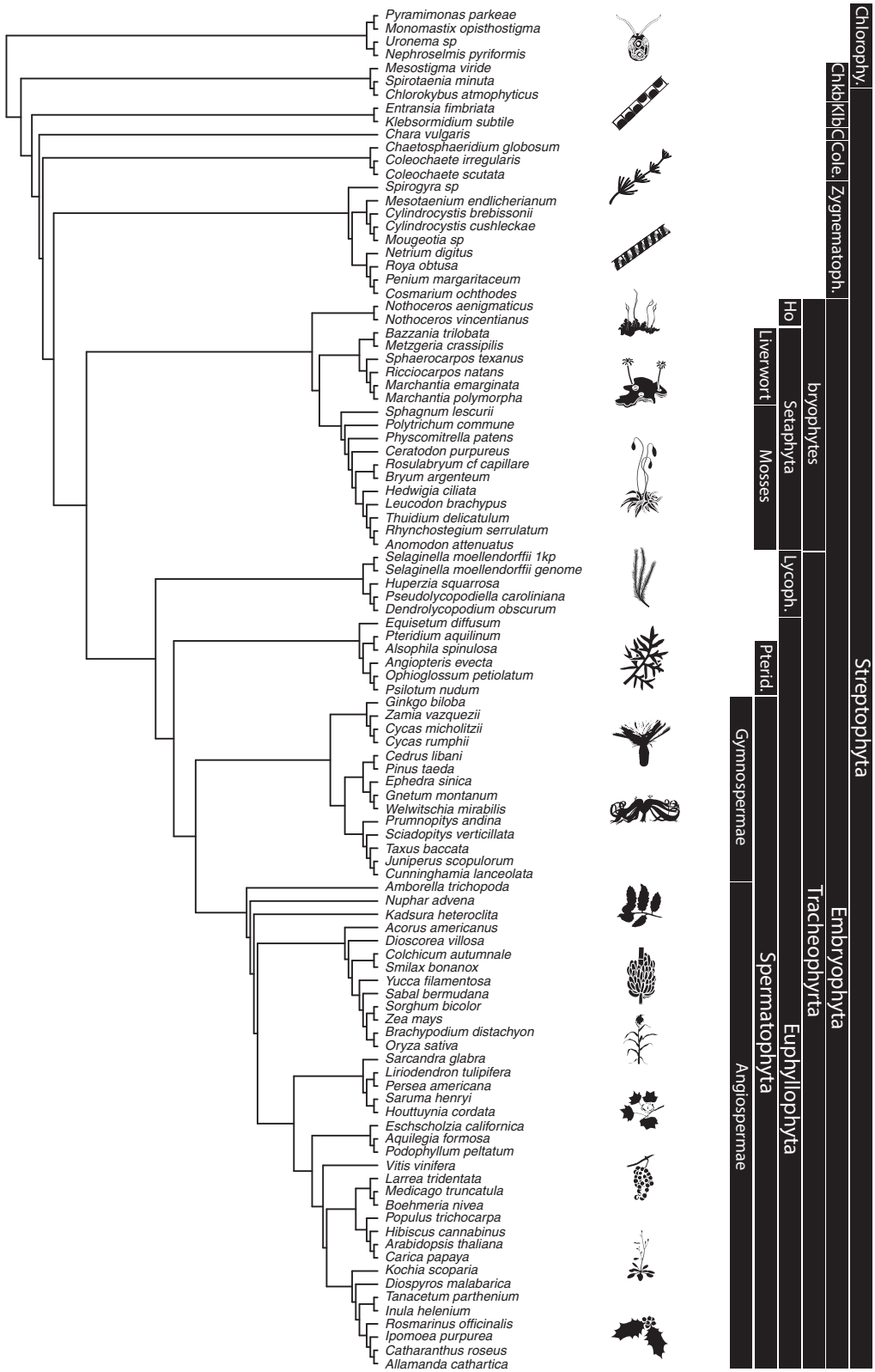


Figure 2. The Main Topology Supported by Analyses Recovers a Monophyletic Bryophyta

Abbreviations: Chloro, Chlorophyta; Chkb, Chlorokybophyceae; Klb, Klebsormidiophyceae; C, Charophyceae; Cole, Coleochaetophyceae; Zygnemato, Zygnematophyceae; Ho, hornworts; Lycophy, Lycophyta; Pterid., Pteridophyta. Images of *Klebsormidium*, *Chara*, and *Spirogyra* are from Phylopic.

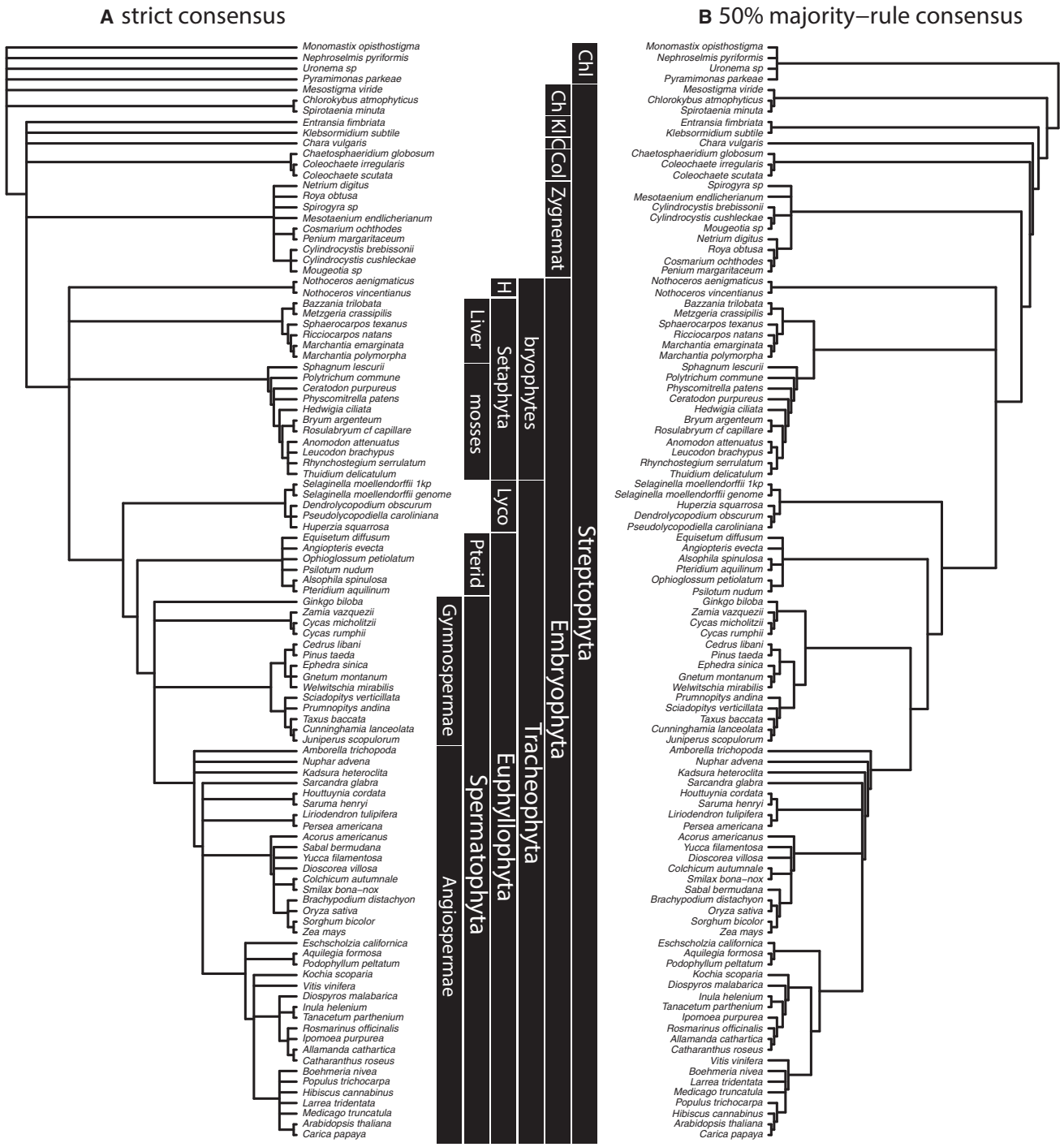


Figure 3. Result of Analysis of 30 Sampled 20,000 Amino Acid, Dayhoff-Recoded Datasets Analyzed under the CAT-GTR Model

Strict (A) and 50% majority-rule (B) consensus. Abbreviations: Chl, Chlorophyta; Ch, Chlorokybophyceae; Kl, Klebsormidiophyceae; C, Charophyceae; Col, Coleochaetophyceae; Zygnemat, Zygnematophyceae; H, hornworts; Liv, liverworts; Lyco, Lycophyta; Pterid., Pteridophyta.

embryophytes with the moss-liverwort clade sister to tracheophytes (10/30 replicates; Figure 4) or within monophyletic bryophytes at the base of Embryophyta (4/30 replicates; Figure 4C). The remaining topologies show variations of a polytomy containing bryophytes at the base of embryophytes. A strict consensus of the results from each jack-knife analysis indicates a polytomy between the bryophyte phyla and tracheophytes (Figure 3A) and the 50% majority-rule consensus shows a similar polytomy, but with mosses and liverworts forming a clade (Figure 3B).

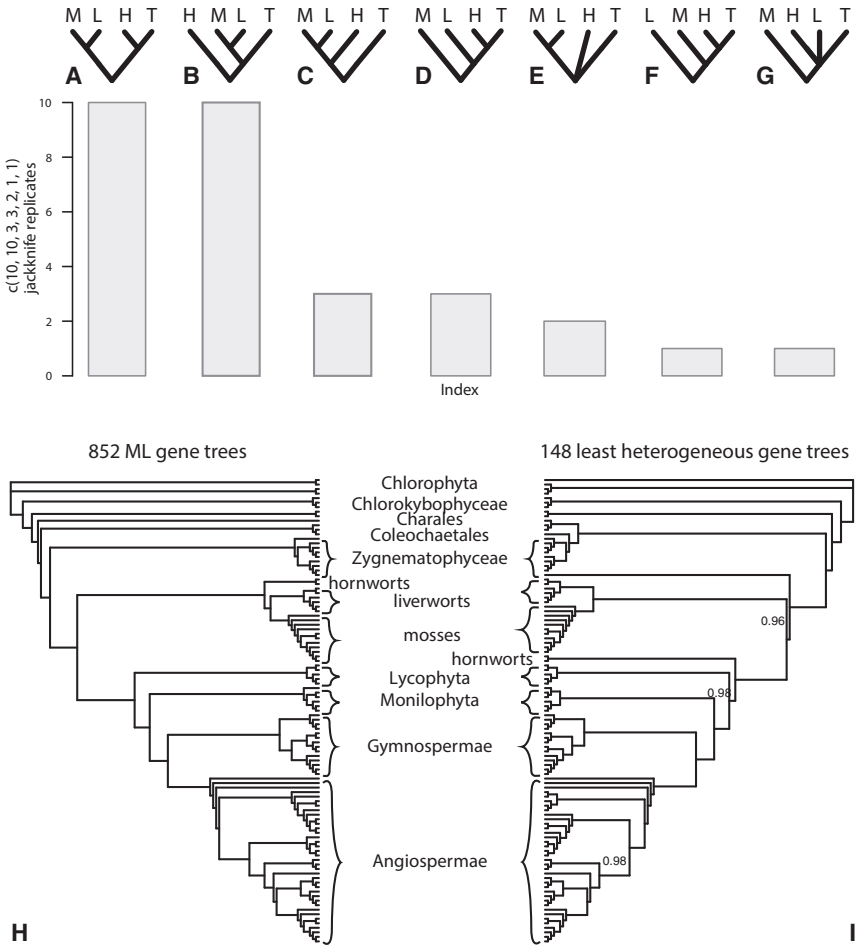


Figure 4. Assessing Support for Competing Topologies

The number of times seven topologies were produced from the 30 jack-knife analyses of Dayhoff-recoded data under the Bayesian CAT-GTR model (A–G) and the topology from Bayesian supertree estimation using the method of Steel and Rodrigo [25] based on 852 gene trees (H) and on the 148 least compositionally heterogeneous genes (I). The figure summarizes the consensus trees produced from each of the jack-knife runs independently: (A) a clade of liverworts and mosses as a sister lineage of hornworts plus tracheophytes; (B) hornworts sister to a clade of mosses plus liverworts, itself sister to tracheophytes; (C) bryophyte monophyly; (D) mosses, liverworts, and hornworts are successive sister lineages of tracheophytes; (E) mosses plus liverworts comprise a clade sister to tracheophytes, the position of hornworts unresolved; (F) hornworts, mosses, and liverworts are successive sister lineages of tracheophytes; and (G) the position of hornworts and liverworts unresolved, sister to tracheophytes, with mosses branching earlier. The Bayesian supertree analysis of 852 genes recovers a bryophyte clade, including the setaphyte clade of liverworts and mosses (H); however, bryophytes are paraphyletic in the tree of 148 genes, with hornworts sister to tracheophytes (I). Abbreviations: H, hornwort; M, moss; L, liverwort; T, tracheophyta.

Across-Branch Compositional Heterogeneity

The NDCH2 analysis of the Dayhoff-recoded 148 least heterogeneous genes recovered monophyletic bryophytes with maximal (posterior probability [PP] = 1.0) posterior probability. CAT-GTR analysis of these data supports a liverwort plus moss clade sister to a hornworts plus tracheophytes clade, with maximal posterior probability (PP = 1.0).

Fit of Sequence Alignments to Proposed Topologies

Analyses of the sequence data rejected all topologies except monophyletic bryophytes (Table 2). None of the five other alternative topologies was sampled during the bootstrap analyses (monophyletic bryophytes proportion = 1). Bryophyte monophyly received the highest support of summed likelihoods and is significantly supported compared to all alternative topologies using approximately unbiased (AU), Kishino-Hasegawa (KH), and Shimodaira-Hasegawa (SH) tests (Table 2).

Supertree Estimation using Gene Trees

Results from analyses of gene trees using Bayesian supertree inference [25] strongly support monophyletic bryophytes. The posterior probability of the split supporting monophyletic bryophytes was 1 across two independent chains (Figure 4H). When the 148 least compositionally heterogeneous genes are analyzed alone (Figure 4I), a liverwort-moss sister clade to the remaining embryophytes is recovered (posterior probability = 0.98). Further, the 148 least heterogeneous genes indicate that *Chara* is sister to embryophytes, whereas all other estimated supertrees support a Zygnematophyceae-Embryophyta relationship (Figure 4I). The effective sample size (ESS) of all parameters was >200, and the median estimate of the free beta parameter was 2.06.

We also estimated multi-species coalescent supertrees in Astral [26]. These trees support bryophyte monophyly with low support (52.5) and also support a sister-group relationship between *Chara* and the embryophytes.

Fit of Gene Trees to Proposed Topologies

Comparisons of the likelihood of gene trees and seven proposed embryophyte relationships support bryophyte monophyly (Table 3). All alternative hypotheses can be rejected using the AU and KH tests, although the more conservative SH test cannot reject the topology in which liverworts and mosses are recovered as the sister clade to the remaining embryophytes (Table 3, top). However, when using the least heterogeneous 148 gene dataset, it is not possible to reject liverworts-mosses sister, monophyletic bryophytes, or liverworts plus hornworts-mosses topologies using the AU test (Table 3, bottom).

DISCUSSION

The phylogenetic relationship among early land plants has been one of the most recalcitrant problems in phylogenetics. We find support for three topologies of bryophytes and tracheophytes

Table 2. The Fit of the Weighted Alignment to the Seven Hypotheses of Early Land Plant Topology Using RELL Bootstrapping

Topology	Likelihood	Δ Likelihood	RELL Bootstrap Proportion	AU	SH	KH
Monophyletic bryophytes	−19,802,012.115	0.000	1	1	1	1
Hornworts sister	−19,802,243.57	231.459	0	0	0.0013	0
Liverworts-mosses sister	−19,802,248.226	236.111	0	0	0.0004	0
Liverworts-mosses-hornworts	−19,803,217.59	1,205.477	0	0.0002	0	0
Liverworts plus hornworts-mosses	−19,803,340.12	1,328.01	0	0.0019	0	0
Hornworts-liverworts-mosses	−19,803,399.22	1,387.107	0	0.0057	0	0
Liverworts-hornworts-mosses	−19,803,472.14	1,460.029	0	0	0	0

The summed likelihood and bootstrap proportion of alignments supports the monophyletic bryophytes topology and significantly rejects all alternative topologies. RELL, resampling of estimated log likelihoods; AU, approximately unbiased; KH, Kishino-Hasegawa; SH, Shimodaira-Hasegawa.

from analyses using models that accommodate sequence heterogeneity: (1) monophyly of the bryophytes, (2) hornworts alone earliest branching, and (3) liverworts plus mosses earliest branching. Bryophyte monophyly is consistently supported by analyses that compare the fit of sequence data and gene trees to hypothesized relationships, as well as by analyses performed using ASTRAL. Support for bryophyte monophyly is also found in the results of analyses accommodating across-branch compositional heterogeneity. Overall, these results suggest bryophyte monophyly, and support for alternative topologies might be a consequence of incomplete lineage sorting and lineage specific compositional heterogeneity in the data. Bryophyte monophyly is congruent with recent studies that accommodate across-branch compositional heterogeneity [11] but conflict with results from a large-scale concatenation analysis using simpler models [7] (Figure 2).

Rejection of Proposed Relationships

In the last 30 years, at least seven topologies have been proposed for early land plants (Figures 1A–1G; Table 1). Our analyses allow us to narrow this topological uncertainty down solely to the position of hornworts: as the sister to a clade of mosses and liverworts in monophyletic bryophytes (Figure 1A), as the sister lineage to other embryophytes (Figure 1B), or as the sister lineage of tracheophytes (Figure 1D). With all methods, we can reject previous hypotheses that do not find a moss-liverwort clade (i.e., Table 1; Figures 1D and 1E–1G), such as the successive branching of hornworts, liverworts, and mosses sister to the tracheophytes [13, 19–21, 27–31]. The three remaining topologies represent a fundamental split in the topology with either monophyly or paraphyly of bryophytes (hornworts or liverworts-mosses as the sister lineage to all remaining embryophytes). Wickett et al. [7] found support for both bryophyte paraphyly (hornworts sister—based on concatenation analyses of nucleotide data) and bryophyte monophyly (based upon a coalescent analysis of a reduced dataset of amino acids). Here we find equivocal support for these hypotheses through estimation of phylogenies based on CAT-GTR-model-based analyses of amino acid data (Figures 3 and 4).

Difficulties in the Resolution of Early Land Plant Phylogeny

Despite three decades of research, the phylogeny of early land plants remains unresolved with large-scale analyses based upon morphology (e.g., [15]) and transcriptomes (e.g., [7]). Our analyses suggest that this continuing controversy is due to a combination of factors: embryophyte sequence data exhibit significant compositional heterogeneity that is difficult to model even with the best current approaches [22]; there is a paucity of sequence data for key hornwort lineages; and biological effects, such as incomplete lineage sorting, may be masking early-branching relationships.

In our analyses, the choice of model did impact the inferred tree topology, and models accounting for across-site or across-tree heterogeneity did not agree with each other. Site specific heterogeneity is very high in this dataset and cannot be accounted for in full also when applying Dayhoff recoding ($Z = |6|$). In addition, it is evident from the results of the CAT-GTR+G analyses, which resolve different tree topologies, that the problem may also be further complicated by the presence of incomplete lineage sorting in the data. However, both the use of ASTRAL coalescent methods and the use of models that accommodate lineage-specific compositional heterogeneity support bryophyte monophyly. Thus, from a modeling perspective, resolving the deepest relationships among land plants is a challenging problem. The application of site-heterogeneous models has resolved phylogenetic controversy elsewhere in the tree of life (e.g., [32, 33]), but, at present, it is unable to definitively resolve early land plant relationships (Figure 4). Furthermore, modeling of compositional heterogeneity between sites may not be sufficiently complex to resolve land plant relationships, as it will not capture compositional heterogeneity among lineages [24] or non-reversibility in sequence evolution [34]. An additional factor is that the genome-scale amino acid datasets that are now available for many embryophyte lineages are too large to be easily modeled using the best-fitting models available. In particular, achieving convergence with the CAT-GTR model in Phylobayes or the NDCH2 model of P4 becomes challenging beyond datasets of around 20,000 aligned positions, even when using parallelization.

A second important factor relates to the sampling of key lineages. Despite the modeling difficulties outlined above, our analyses enabled us to resolve robustly all of the main branches of the land plant tree with the exception of the hornworts—the most poorly sampled major lineage in our dataset. Hornworts are currently represented by just two transcriptomes from congeneric species (*Nothoceros*), leaving four of the five orders unrepresented. Thus, improved genomic or transcriptomic

Table 3. Comparison of the Sum of Approximate Likelihoods of Gene Trees, When Fit to Seven Alternative Topologies of Early Plant Relationships

Topology	Approximate Likelihood	Δ Likelihood	Multiscale Bootstrap Proportion	AU	KH	SH
852 Maximum-Likelihood Estimate Trees						
Monophyletic bryophytes	−56,488.38	0	0.994	0.995	0.995	1
Liverworts-mosses sister	−56,545.555	57.2	0.006	0.006	0.005	0.050
Hornwort sister	−56,563.42	75	9e−5	7e−5	5e−5	0.023
Liverworts plus hornworts-mosses	−56,697.43	209.1	3e−6	1e−5	0	0
Hornworts-liverworts-mosses	−56,749.25	260.9	7e−20	2e−56	0	0
Liverworts-mosses-hornworts	−56,772.48	284.1	3e−25	4e−82	0	0
Liverworts-hornworts-mosses	−56,784.99	296.6	3e−6	2e−5	0	0
148 Least Compositional Heterogeneous Maximum-Likelihood Estimate Trees						
Liverworts-mosses sister	−9,816.57	0	0.823	0.969	0.937	0.996
Monophyletic bryophytes	−9,830.48	13.9	0.067	0.091	0.063	0.3
Liverworts plus hornworts-mosses	−9,833.95	17.4	0.128	0.079	0.077	0.179
Hornwort sister	−9,833.95	17.4	0.020	0.031	0.026	0.199
Liverworts-mosses-hornworts	−9,842.65	26.1	0.008	0.017	0.018	0.063
Hornworts-liverworts-mosses	−9,844.38	27.8	0.008	0.02	0.007	0.023
Liverworts-hornworts-mosses	−9,844.38	27.8	0.001	0.004	0.008	0.027

Datasets were analyzed using the method of Steel and Rodrigo [25]. Top: for the 852 genes, the beta value was set to 2.05 (the median value from the P4 analyses). Over all gene trees, the summed likelihood is highest for the monophyletic bryophytes topology. There is significant support for monophyletic bryophytes over all other potential land plant relationships using the AU and KH. Bottom: for the 148 gene trees with the least compositional heterogeneity, the beta value was set to 2. Analyses were run with the maximum-likelihood estimate topologies. The liverworts-mosses sister tree was the topology with the highest support, but both monophyletic bryophytes and liverworts plus hornworts mosses could not be rejected.

sampling of hornworts may help to place this final recalcitrant branch in the land plant tree. Liverwort diversity is also under represented, particularly for early-branching lineages. Thus, the quality of the dataset could be improved significantly through addition of transcriptome data for the hornworts *Leiosporoceros* and/or *Anthoceros*, as well as the earliest-diverging lineages of haplomitropsid liverworts, *Haplomitrium* and/or *Treubia*.

Finally, the current intractability of land plant relationships must be, at least in part, biological. Part of the reason may be horizontal gene transfer, which has been documented between hornworts and pteridophytes [35], as well as mosses and angiosperms [36], potentially biasing the estimation of phylogenies using concatenation. The short branches separating the four main lineages of land plants may indicate a rapid adaptive radiation resulting in conflicts among gene trees as a consequence of incomplete lineage sorting of gene polymorphisms. Future studies may aim to identify sets of gene trees supporting alternative topologies as a consequence of these biological processes.

Support for Monophyletic Bryophytes and the Liverwort-Moss Clade "Setaphyta"

Bayesian supertree estimation [25], gene tree coalescent analysis [26], and maximum-likelihood analyses [37] of gene trees all support monophyletic bryophytes [7–11, 38–40]. Wickett et al. [7] also found monophyletic bryophytes based on gene tree coalescence analyses of amino acids, and other studies have suggested that support for relationships other than bryophyte monophyly is artifactual in chloroplast data [11]. We find the support for bryophyte monophyly based on analysis of gene trees, but there is additional support from analyses of the fit of sequence data to competing phylogenetic hypotheses (Tables 2 and 3). For example, significant tests of the relative fit of sequence data to topologies reject hornworts sister, and the highest support and bootstrap proportion is for monophyletic bryophytes (Table 2). Our results are compatible with concatenation precluding resolution of early land plant phylogeny, possibly due to incomplete lineage sorting (e.g., [41]). Furthermore, the approach we apply here [37] is similar to the gene genealogy interrogation (GGI) methodology of Arcila et al. [42], who found that testing between the fit of gene trees to hypothesized species trees could be used to resolve difficult areas of the tree of life. It is possible the gene tree approach here can overcome known difficulties in sequence evolution of early land plants [11, 22] to resolve these difficult relationships when concatenation-based methods struggle.

While support for a monophyletic Bryophyta is not unequivocal, we find overwhelming support for a clade of mosses and liverworts that we name "Setaphyta" after the sporophyte seta, a stalk supporting the capsule in both liverworts and mosses. However, we define Setaphyta phylogenetically, not based on an apomorphy, as the clade comprised of living mosses and liverworts, their last common ancestor and all of its descendants.

Implications for Early Land Plant Evolution

Most recent analyses (e.g., [19, 21, 28, 29]) have supported liverworts as the sister group to all other embryophytes. This result has been compelling because liverworts are missing a number of characters that are otherwise shared between mosses and vascular plants, e.g., stomata. As such, liverworts have been envisaged as an evolutionary halfway stage between primitively

aquatic algae and full-blown terrestrial stomatophytes and, hence, the widespread adoption of the liverwort *Marchantia polymorpha* as a developmental [43] and genome [44] model. However, regardless of the phylogenetic position of hornworts, the results of our analyses indicate that liverwort morphology resulted from loss of traits, rather than reflecting a primitive absence of embryophyte characters. Mosses have frequently been considered the closest bryophyte relative of the tracheophytes [13–16, 30, 31, 45]; our results indicate that this does not reflect close kinship, but, rather, the retention of shared primitive embryophyte characters that have been lost principally in liverworts.

Our results question the relevance of liverworts like *Marchantia* as a developmental [43], genomic [44], and biogeochemical [46] model for the ancestral embryophyte in evolutionary studies. Indeed, attempts to infer the developmental biology of ancestral embryophytes requires phylogenetically constrained insights from both bryophyte model systems, *Marchantia* and *Physcomitrella*, in comparison to tracheophyte models like *Selaginella*. This is because no one lineage can be considered a developmental, genetic, or physiological surrogate for the ancestral embryophyte (cf. [47]); all are a mélange of shared primitive and unshared derived characters specific to their respective lineages. Only through phylogenetically constrained comparative developmental biology is it possible to disambiguate the sequence in which their shared and derived characters were assembled during the evolution of land plant body plans. Nevertheless, our results indicate that for those organ systems that, in liverworts, are secondarily simplified or absent through loss as a consequence of loss of function or unique adaptations to life on land (e.g., gas exchange, spore wall structure, sporangium development and structure, rhizoid structure, and thalloid gametophyte form), *Marchantia* and other liverworts might not serve as an appropriate model for anything other than liverworts themselves. In particular, recognition of a moss-liverwort setaphyte clade impacts significantly on contemporary efforts to uncover the molecular developmental basis of the gametophyte-sporophyte phenotypic differentiation in extant land plant lineages [48, 49]. Paraphyletic bryophytes, with liverworts or hornworts sister, is consistent with the hypothesis of an ancestral dimorphic life cycle with gametophyte dominance extending to the level of life-long nutritional dependence of sporophyte on gametophyte. However, bryophyte monophyly is compatible with more varied life-cycle states, *viz*. dimorphic with gametophyte dominance, dimorphic with sporophyte dominance, dimorphic with co-dominance of both generations, and nearly monomorphic generations. The setaphyte clade also has significant implications for the origin of sporophytic characters in land plants, including the evolution of placental tissue, the loss of stomata in liverworts and some lineages of mosses, and the origin of their sporophyte developmental programs [50, 51]. This derived phylogenetic position for the stomata-free liverworts indicates that stomata are ancestral to land plants (though it does not resolve questions concerning the ancestral function of these structures [52]).

Indeed, our unequivocal resolution of Setaphyta requires that the last common ancestor of liverworts plus mosses was much more complex—more stomatophyte like, or even tracheophyte like—than has been perceived hitherto. This may explain, for example, differences in the conducting tissues of bryophytes compared to those of vascular plants [53], perhaps inherited from a more tracheophyte-like ancestral embryophyte. As such, the results of our phylogenetic analyses have implications for interpreting the affinity and, ultimately, the evolutionary significance of the earliest fossil plants. The majority are currently interpreted as stem tracheophytes, but it remains possible that some are stem-lineage representatives of a more complex embryophyte or bryophyte crown ancestor [54].

Conclusions

Our results highlight the hornworts as a key lineage for improved genomic or transcriptomic sampling to definitively resolve land plant phylogeny. Our finding of monophyletic bryophytes could have major impacts on understanding the macroevolution of embryophytes. Specifically, the resolution of bryophytes as monophyletic could greatly affect the estimation of the ancestral state of life history for land plants [55] and the divergence times of lineages. Early bryophytes are unknown from the fossil record [1], and vascular plants are known from the Silurian [2], so the resolution of monophyletic bryophytes belies a cryptic history of bryophyte evolution intrinsic to the hypothesis of paraphyletic bryophytes. Further confirmation of these results could be provided via total-evidence approaches that directly incorporate information from the fossil record and improved sampling of hornworts.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- METHOD DETAILS
  - Dataset selection
  - Phylogenetic models
  - Compositional heterogeneity among sites
  - Focal analysis
  - Across-branch compositional heterogeneity
  - Tests of proposed embryophyte relationships and sequence alignments
  - Tests of proposed embryophyte relationships and inferred gene trees

ACKNOWLEDGMENTS

We thank the University of Bristol Advanced Computing Research Centre and University College London for providing access to supercomputing infrastructure. This work was supported by the NERC (NE/N002067/1 to P.C.J.D., D.E., H.S., P.K., S.P., and C.W.), the Royal Society (Wolfson Merit Award to P.C.J.D.; University Research Fellowship to T.W.), and Fundação para a Ciência e a Technologia (FCT), Portugal (PTDC/BIA-EVF/1499/2014 to C.J.C.). We also thank the reviewers and other members of the Bristol Palaeobiology Research Group for their formative feedback on the development of this study.

AUTHOR CONTRIBUTIONS

Conceptualization, P.C.J.D., D.E., H.S., P.K., S.P., and C.H.W.; Methodology, C.J.C., P.C.J.D., M.N.P., J.M., D.P., H.S., and T.W.; Investigation, M.N.P., J.M., and T.W.; Interpretation, all authors; Writing – Original Draft, P.C.J.D.,

M.N.P., J.M., D.P., H.S., and T.W.; Writing – Review & Editing, all authors; Supervision, P.C.J.D., D.P., and H.S.; Funding Acquisition, C.J.C., P.C.J.D., D.E., H.S., P.K., S.P., and C.H.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: November 30, 2017
Revised: January 15, 2018
Accepted: January 22, 2018
Published: February 15, 2018

REFERENCES

1. Kenrick, P., Wellman, C.H., Schneider, H., and Edgecombe, G.D. (2012). A timeline for terrestrialization: consequences for the carbon cycle in the Palaeozoic. Philos. Trans. R. Soc. Lond. B Biol. Sci. *367*, 519–536.
2. Gensel, P.G. (2008). The earliest land plants. Annu. Rev. Ecol. Evol. Syst. *39*, 459–477.
3. Berner, R.A. (1997). The rise of plants and their effect on weathering and atmospheric CO_2. Science *276*, 544–546.
4. Kenrick, P., and Crane, P.R. (1997). The origin and early evolution of plants on land. Nature *389*, 33–39.
5. Retallack, G.J. (1997). Early forest soils and their role in Devonian global change. Science *276*, 583–585.
6. Gibling, M.R., and Davies, N.S. (2012). Palaeozoic landscapes shaped by plant evolution. Nat. Geosci. *5*, 99–105.
7. Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA *111*, E4859–E4868.
8. Garbary, D.J., Renzaglia, K.S., and Duckett, J.G. (1993). The phylogeny of land plants: a cladistic analysis based on male gametogenesis. Plant Syst. Evol. *188*, 237–269.
9. Nishiyama, T., Wolf, P.G., Kugita, M., Sinclair, R.B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K., et al. (2004). Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol. Biol. Evol. *21*, 1813–1819.
10. Goremykin, V.V., and Hellwig, F.H. (2005). Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. Plant Syst. Evol. *254*, 93–103.
11. Cox, C.J., Li, B., Foster, P.G., Embley, T.M., and Civán, P. (2014). Conflicting phylogenies for early land plants are caused by composition biases among synonymous substitutions. Syst. Biol. *63*, 272–279.
12. Waters, D.A., Buchheim, M.A., Dewey, R.A., and Chapman, R.L. (1992). Preliminary inferences of the phylogeny of bryophytes from nuclear-encoded ribosomal RNA sequences. Am. J. Bot. *79*, 459–466.
13. Mishler, B.D., and Churchill, S.P. (1984). A cladistic approach to the phylogeny of the 'bryophytes'. Brittonia *36*, 406–424.
14. Mishler, B.D., Lewis, L.A., Buchheim, M.A., Renzaglia, K.S., Garbary, D.J., Delwiche, C.F., Zechman, F.W., Kantz, T.S., and Chapman, R.L. (1994). Phylogenetic relantionships of the 'green algae' and 'bryophytes'. Ann. Mo. Bot. Gard. *81*, 451–483.
15. Kenrick, P., and Crane, P.R. (1997). The Origin and Early Diversification of Land Plants: A Cladistic Study (Smithsonian Institution Press).
16. Qiu, Y.L., Cho, Y., Cox, J.C., and Palmer, J.D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature *394*, 671–674.
17. Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., et al. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA *103*, 15511–15516.
18. Gerrienne, P., Servais, T., and Vecoli, M. (2016). Plant evolution and terrestrialization during Palaeozoic times—The phylogenetic context. Rev. Palaeobot. Palynol. *227*, 4–18.
19. Ruhfel, B.R., Gitzendanner, M.A., Soltis, P.S., Soltis, D.E., and Burleigh, J.G. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. BMC Evol. Biol. *14*, 23.
20. Gao, L., Su, Y.-J., and Wang, T. (2010). Plastid genome sequencing, comparative genomics, and phylogenomics: current status and prospects. J. Syst. Evol. *48*, 77–93.
21. Liu, Y., Cox, C.J., Wang, W., and Goffinet, B. (2014). Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. Syst. Biol. *63*, 862–878.
22. Cooper, E.D. (2014). Overly simplistic substitution models obscure green plant phylogeny. Trends Plant Sci. *19*, 576–582.
23. Lartillot, N., and Philippe, H. (2004). A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol. Biol. Evol. *21*, 1095–1109.
24. Foster, P.G. (2004). Modeling compositional heterogeneity. Syst. Biol. *53*, 485–495.
25. Steel, M., and Rodrigo, A. (2008). Maximum likelihood supertrees. Syst. Biol. *57*, 243–250.
26. Mirarab, S., Bayzid, M.S., Boussau, B., and Warnow, T. (2014). Statistical binning enables an accurate coalescent-based estimation of the avian tree. Science *346*, 1250463.
27. Smith, S.A., Beaulieu, J.M., and Donoghue, M.J. (2009). Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches. BMC Evol. Biol. *9*, 37.
28. Chang, Y., and Graham, S.W. (2011). Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. Am. J. Bot. *98*, 839–849.
29. Zhong, B., Xi, Z., Goremykin, V.V., Fong, R., McLenachan, P.A., Novis, P.M., Davis, C.C., and Penny, D. (2014). Streptophyte algae and the origin of land plants revisited using heterogeneous models with three new algal chloroplast genomes. Mol. Biol. Evol. *31*, 177–183.
30. Mishler, B.D., Thrall, P.H., Hopple, J., John, S., De Luna, E., and Vilgalys, R. (1992). A molecular approach to the phylogeny of bryophytes: cladistic analysis of chloroplast-encoded 16S and 23S ribosomal RNA genes. Bryologist *95*, 172–180.
31. Bremer, K. (1985). Summary of green plant phylogeny and classification. Cladistics *1*, 369–385.
32. Delsuc, F., Brinkmann, H., Chourrout, D., and Philippe, H. (2006). Tunicates and not cephalochordates are the closest living relatives of vertebrates. Nature *439*, 965–968.
33. Pisani, D., Pett, W., Dohrmann, M., Feuda, R., Rota-Stabelli, O., Philippe, H., Lartillot, N., and Wörheide, G. (2015). Genomic data do not support comb jellies as the sister group to all other animals. Proc. Natl. Acad. Sci. USA *112*, 15402–15407.
34. Williams, T.A., Heaps, S.E., Cherlin, S., Nye, T.M., Boys, R.J., and Embley, T.M. (2015). New substitution models for rooting phylogenetic trees. Philos. Trans. R. Soc. Lond. B Biol. Sci. *370*, 20140336.
35. Li, F.W., Villarreal, J.C., Kelly, S., Rothfels, C.J., Melkonian, M., Frangedakis, E., Ruhsam, M., Sigel, E.M., Der, J.P., Pittermann, J., et al. (2014). Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. Proc. Natl. Acad. Sci. USA *111*, 6672–6677.
36. Soucy, S.M., Huang, J., and Gogarten, J.P. (2015). Horizontal gene transfer: building the web of life. Nat. Rev. Genet. *16*, 472–482.
37. Akanni, W.A., Creevey, C.J., Wilkinson, M., and Pisani, D. (2014). L.U.St: a tool for approximated maximum likelihood supertree reconstruction. BMC Bioinformatics *15*, 183.

38. Hedderson, T.A.J., Chapman, R.L., and Rootes, W.L. (1996). Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Plant Syst. Evol. *200*, 213–224.

39. Malek, O., Lättig, K., Hiesel, R., Brennicke, A., and Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J. *15*, 1403–1411.

40. Garbary, D.J., and Renzaglia, K.S. (1988). Bryophyte phylogeny and the evolution of land plants: evidence from development and ultrastructure. In Bryology for the Twenty-First Century, J.W. Bates, N.W. Ashton, and J.G. Duckett, eds. (Maney and the British Bryological Society), pp. 45–63.

41. Edwards, S.V., Liu, L., and Pearl, D.K. (2007). High-resolution species trees without concatenation. Proc. Natl. Acad. Sci. USA *104*, 5936–5941.

42. Arcila, D., Ortí, G., Vari, R., Armbruster, J.W., Stiassny, M.L.J., Ko, K.D., Sabaj, M.H., Lundberg, J., Revell, L.J., and Betancur-R, R. (2017). Genome-wide interrogation advances resolution of recalcitrant groups in the tree of life. Nat Ecol Evol *1*, 20.

43. Shimamura, M. (2016). *Marchantia polymorpha*: taxonomy, phylogeny and morphology of a model system. Plant Cell Physiol. *57*, 230–256.

44. Bowman, J.L., Kohchi, T., Yamato, K.T., Jenkins, J., Shu, S., Ishizaki, K., Yamaoka, S., Nishihama, R., Nakamura, Y., Berger, F., et al. (2017). Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. Cell *171*, 287–304.e15.

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

46. Fletcher, B.J., Brentnall, S.J., Quick, W.P., and Beerling, D.J. (2006). BRYOCARB: a process-based model of thallose liverwort carbon isotope fractionation in response to CO2, O2, light and temperature. Geochim. Cosmochim. Acta *70*, 5676–5691.

47. Conway-Morris, S. (2003). The Cambrian "explosion" of metazoans and molecular biology: would Darwin be satisfied? Int. J. Dev. Biol. *47*, 505–515.

48. Niklas, K.J., and Kutschera, U. (2010). The evolution of the land plant life cycle. New Phytol. *185*, 27–41.

49. Bowman, J.L., Sakakibara, K., Furumizu, C., and Dierschke, T. (2016). Evolution in the cycles of Life. Annu. Rev. Genet. *50*, 133–154.

50. Ligrone, R., Duckett, J.G., and Renzaglia, K.S. (2012). Major transitions in the evolution of early land plants: a bryological perspective. Ann. Bot. (Lond.) *109*, 851–871.

51. Ligrone, R., Duckett, J.G., and Renzaglia, K.S. (2012). The origin of the sporophyte shoot in land plants: a bryological perspective. Ann. Bot. (Lond.) *110*, 935–941.

52. Field, K.J., Duckett, J.G., Cameron, D.D., and Pressel, S. (2015). Stomatal density and aperture in non-vascular land plants are non-responsive to above-ambient atmospheric CO_2 concentrations. Ann. Bot. (Lond.) *115*, 915–922.

53. Hebant, C. (1977). The Conducting Tissues of Bryophytes (J. Cramer).

54. Edwards, D., Morris, J.L., Richardson, J.B., and Kenrick, P. (2014). Cryptospores and cryptophytes reveal hidden diversity in early land floras. New Phytol. *202*, 50–78.

55. Donoghue, M.J. (2005). Key innovations, convergence, and success: macroevolutionary lessons from plant phylogeny. Paleobiology *31*, 77–93.

56. Hori, H., Lim, B.-L., and Osawa, S. (1985). Evolution of green plants as deduced from 5S rRNA sequences. Proc. Natl. Acad. Sci. USA *82*, 820–823.

57. Steele, K.P., Holsinger, K.E., and Jansen, R.K. (1991). Assessing the reliability of 5S rRNA sequence data for phylogenetic analysis in green plants. Mol. Biol. Evol. *8*, 240–248.

58. Renzaglia, K.S., Nickrent, D.L., and Garbary, D.J.; Duff RJT (2000). Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos. Trans. R. Soc. Lond. B Biol. Sci. *355*, 769–793.

59. Quandt, D., Müller, K., Stech, M., Frahm, J.-P., Frey, W., Hilu, K.W., and Borsch, T. (2004). Molecular evolution of the chloroplast trnL-F region in land plants. In Molecular Systematics of Bryophytes, B. Goffinet, V. Hollowell, and R. Magill, eds. (Missouri Botanical Garden Press), pp. 13–37.

60. Rodríguez-Ezpeleta, N., Philippe, H., Brinkmann, H., Becker, B., and Melkonian, M. (2007). Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of *Mesostigma* in the Streptophyta. Mol. Biol. Evol. *24*, 723–731.

61. Turmel, M., Otis, C., and Lemieux, C. (2006). The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. Mol. Biol. Evol. *23*, 1324–1338.

62. Lemieux, C., Otis, C., and Turmel, M. (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 Biol. *5*, 2.

63. Civaň, P., Foster, P.G., Embley, M.T., Séneca, A., and Cox, C.J. (2014). Analyses of charophyte chloroplast genomes help characterize the ancestral chloroplast genome of land plants. Genome Biol. Evol. *6*, 897–911.

64. Hedderson, T.A., Chapman, R., and Cox, C.J. (1998). Bryophytes and the origins and diversi¢cation of land plants: new evidence from molecules. In Bryology for the Twenty-First Century, J.W. Bates, N.W. Ashton, and J.G. Duckett, eds. (The British Bryological Society and Maney Publishing), pp. 65–77.

65. Duff, R.J., and Nickrent, D.L. (1999). Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Am. J. Bot. *86*, 372–386.

66. Nishiyama, T., and Kato, M. (1999). Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. Mol. Biol. Evol. *16*, 1027–1036.

67. Nickrent, D.L., Parkinson, C.L., Palmer, J.D., and Duff, R.J. (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. Mol. Biol. Evol. *17*, 1885–1895.

68. Bremer, K., Humphries, C.J., Mishler, B.D., and Churchill, S.P. (1987). On cladistic relationships in green plants. Taxon *36*, 339–349.

69. Karol, K.G., Arumuganathan, K., Boore, J.L., Duffy, A.M., Everett, K.D.E., Hall, J.D., Hansen, S.K., Kuehl, J.V., Mandoli, D.F., Mishler, B.D., et al. (2010). Complete plastome sequences of *Equisetum arvense* and *Isoetes flaccida*: implications for phylogeny and plastid genome evolution of early land plant lineages. BMC Evol. Biol. *10*, 321.

70. Lemieux, C., Otis, C., and Turmel, M. (2016). Comparative Chloroplast Genome Analyses of Streptophyte Green Algae Uncover Major Structural Alterations in the Klebsormidiophyceae, Coleochaetophyceae and Zygnematophyceae. Front. Plant Sci. *7*, 697.

71. Parenti, L.R. (1980). A phylogenetic analysis of the land plants. Biol. J. Linn. Soc. Lond. *13*, 225–242.

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

73. Delwiche, C.F., Karol, K.G., Cimino, M.T., and Sytsma, K.J. (2002). Phylogeny of the genus *Coleochaete* (Coleochaetales, Charophyta) and related taxa inferred by analysis of the chloroplast gene rbcL1. J. Phycol. *38*, 394–403.

74. Kelch, D.G., Driskell, A., and Mishler, B.D. (2004). Inferring phylogeny using genomic characters: a case study using land plant plastomes. Monographs in Systematic Botany from the Missouri Botanical Garden *98*, 3–12.

75. Wolf, P.G., Karol, K.G., Mandoli, D.F., Kuehl, J., Arumuganathan, K., Ellis, M.W., Mishler, B.D., Kelch, D.G., Olmstead, R.G., and Boore, J.L. (2005). The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). Gene *350*, 117–128.

76. Groth-Malonek, M., Pruchner, D., Grewe, F., and Knoop, V. (2005). Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol. Biol. Evol. *22*, 117–125.
77. Qiu, Y.L., Li, L.B., Wang, B., Chen, Z.D., Dombrovska, O., Lee, J., Kent, L., Li, R.Q., Jobson, R.W., Hendry, T.A., et al. (2007). A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. Int. J. Plant Sci. *168*, 691–708.
78. Fiz-Palacios, O., Schneider, H., Heinrichs, J., and Savolainen, V. (2011). Diversification of land plants: insights from a family-level phylogenetic analysis. BMC Evol. Biol. *11*, 341.
79. Magallón, S., Hilu, K.W., and Quandt, D. (2013). Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. Am. J. Bot. *100*, 556–573.
80. Turmel, M., Otis, C., and Lemieux, C. (2013). Tracing the evolution of streptophyte algae and their mitochondrial genome. Genome Biol. Evol. *5*, 1817–1835.
81. Kim, H.T., Chung, M.G., and Kim, K.J. (2014). Chloroplast genome evolution in early diverged leptosporangiate ferns. Mol. Cells *37*, 372–382.
82. Lemieux, C., Otis, C., and Turmel, M. (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.
83. Floyd, S.K., and Bowman, J.L. (2006). Distinct developmental mechanisms reflect the independent origins of leaves in vascular plants. Curr. Biol. *16*, 1911–1917.
84. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics *25*, 2286–2288.
85. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.
86. Shimodaira, H., and Hasegawa, M. (2001). CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics *17*, 1246–1247.
87. Team, R.C. (2016). R: a language and environment for statistical computing (R Foundation for Statistical Computing).
88. Plummer, M., Best, N., Cowles, K., and Vines, K. (2006). CODA: convergence diagnosis and output analysis for MCMC. R News *6*, 7–11.
89. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. Bioinformatics *20*, 289–290.
90. Bryant, D., and Steel, M. (2009). Computing the distribution of a tree metric. IEEE/ACM Trans. Comput. Biol. Bioinformatics *6*, 420–426.
91. Delsuc, F., Tsagkogeorga, G., Lartillot, N., and Philippe, H. (2008). Additional molecular support for the new chordate phylogeny. Genesis *46*, 592–604.
92. Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C.i. (1978). A model of evolutionary change in proteins. In Atlas of Protein Sequences and Structure, M.O. Dayhoff, ed. (National Biomedical Research Foundation), pp. 345–352.
93. Hrdy, I., Hirt, R.P., Dolezal, P., Bardonová, L., Foster, P.G., Tachezy, J., and Embley, T.M. (2004). Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex I. Nature *432*, 618–622.
94. Heaps, S.E., Nye, T.M., Boys, R.J., Williams, T.A., and Embley, T.M. (2014). Bayesian modelling of compositional heterogeneity in molecular phylogenetics. Stat. Appl. Genet. Mol. Biol. *13*, 589–609.
95. Quang, S., Gascuel, O., and Lartillot, N. (2008). Empirical profile mixture models for phylogenetic reconstruction. Bioinformatics *24*, 2317–2323.
96. Kishino, H., and Hasegawa, M. (1989). Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. J. Mol. Evol. *29*, 170–179.
97. Goldman, N., Anderson, J.P., and Rodrigo, A.G. (2000). Likelihood-based tests of topologies in phylogenetics. Syst. Biol. *49*, 652–670.
98. Yang, Z. (2014). Molecular Evolution: A Statistical Approach (Oxford University Press).
99. Shimodaira, H., and Hasegawa, M. (1999). Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. *16*, 1114–1116.
100. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.
101. Robinson, D.F., and Foulds, L.R. (1981). Comparison of phylogenetic trees. Math. Biosci. *53*, 131–147.
102. Akanni, W.A., Wilkinson, M., Creevey, C.J., Foster, P.G., and Pisani, D. (2015). Implementing and testing Bayesian and maximum-likelihood supertree methods in phylogenetics. R. Soc. Open Sci. *2*, 140436.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and analyzed data	This paper	https://doi.org/10.6084/m9.figshare.5633983
Software and Algorithms		
Phylobayes	Lartillot et al. [84]	https://github.com/bayesiancook/pbmpi
Lust	Akanni et al. [37]	https://afro-juju@bitbucket.org/afro-juju/l.u.st.git.
IQ-Tree	Nyugen et al. [85]	http://www.iqtree.org/
CONSEL	Shimodaira and Hasegawa [86]	http://stat.sys.i.kyoto-u.ac.jp/prog/consel/
ASTRAL	Mirarab et al. [26]	https://github.com/smirarab/ASTRAL
P4	Foster [24]	http://p4.nhm.ac.uk/
R	R Core Team [87]	https://cran.r-project.org/
CODA	Plummer et al. [88]	https://cran.r-project.org/web/packages/coda/coda.pdf
APE	Paradis et al. [89]	https://cran.r-project.org/web/packages/ape/ape.pdf

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Philip Donoghue (phil.donoghue@bristol.ac.uk).

METHOD DETAILS

Dataset selection

We used the amino acid transcriptome alignments from Wickett et al. [7]. The original authors presented and analyzed different variants of nucleotide and amino acid alignments of their data; their focal analyses were based upon a 674 gene dataset with the 3rd codon removed. We also employed their full dataset of 852 genes. We termed these datasets the 'reduced dataset' (674 genes) and 'full dataset' (852 genes) respectively. These datasets are available from figshare (https://doi.org/10.6084/m9.figshare.5633983).

Phylogenetic models

We employed a diversity of models to evaluate competing hypotheses on the phylogenetic relationships among the principal groups of land plants: (a) the CAT-GTR model in Phylobayes which models compositional heterogeneity among sites [84]; (b) testing the fit of gene trees and sequence data to proposed alternative topologies of land plant relationships [25, 37]; and (c) Bayesian supertree estimation of gene trees [24, 25, 90].

Compositional heterogeneity among sites

Phylobayes analyses are computationally expensive, limiting dataset size to alignments of around 20K amino acids. Our datasets are considerably larger and so we employed a jack-knifing approach to analyze the datasets in 20K amino acid samples [91].

Focal analysis

We sub-sampled the reduced 674 gene dataset, selecting genes that are present in at least 95% of taxa. This resulted in an alignment of 20,512 amino acids that we used for a 'focal analysis'. As this dataset only represents a portion of the original data, we used jack-knifing to sample 30 alignments of 20K amino acids with replacement, from the 852 gene dataset in our 'jack-knife analysis' [91]. Initial analyses indicated that it was not possible to convergence for the focal and jack-knifing datasets, despite running analyses for > 10K points. Therefore, to reduce heterogeneity and ease convergence, we re-coded the data into six-state Dayhoff groups for all subsequent analyses. Dayhoff recoding [92] groups the twenty amino acids into six "bins," each containing a set of biochemically similar amino acids. For all Phylobayes analyses we ran the CAT-GTR model in two independent chains until convergence was reached. We assessed convergence based on estimated effective sample size of parameters (target > 50) and the maximum number of bipartitions that differ between phylogenies (target < 0.3). All analyses were run using Phylobayes MPI 1.7 on the University College London Computer Science Cluster.

Across-branch compositional heterogeneity

The analyses described above made use of models that allow sequence composition to vary across the sites of the alignment, a pattern that may arise from the varying functional constraints experienced by different amino acids in a protein. But sequence composition can also vary among lineages, and this among-branch compositional heterogeneity has been shown to mislead phylogenetic inference by, for example, grouping lineages by shared compositional biases rather than common ancestry [11, 24].

To evaluate the degree of across-branch compositional heterogeneity in the Wickett et al. dataset, we applied a Chi-square test for compositional homogeneity [24] to the 852 single copy orthologs in the full dataset using P4 [7]. To account for phylogenetic structure, we generated a null distribution separately for each gene by simulating 1000 sequence alignments of the same length as the real data on the maximum likelihood single gene tree under a composition-homogeneous model LG + Γ, and compared the chi-square scores for the real and simulated data. If the chi-square statistic on the real data fell above the 95th percentile of the scores from the simulated data, we judged that gene to show significant across-branch compositional heterogeneity.

Of the 852 genes, only 148 displayed no evidence of significant across-branch compositional heterogeneity ($p >= 0.05$), indicating that this kind of compositional variation is a pervasive feature of the full dataset. To investigate whether failure to model across-branch variation in our other analyses might have led to phylogenetic error, we performed an additional set of phylogenetic analyses on the set of 148 'least heterogeneous' genes. First, we analyzed a concatenation of these genes using the CAT+GTR model, both on the "raw" amino acid alignment and on a Dayhoff-recoded data [93]. By only modeling substitutions between Dayhoff groups of amino-acids (bins), this recoding partially ameliorates both across-branch and across-site compositional heterogeneity, albeit at the cost of some loss of phylogenetic information.

We also inferred a phylogeny in P4 under the among-branch composition heterogeneous NDCH2 (with GTR + Γ) model [24]. The NDCH2 model is similar to the NDCH model (described in [24]), but where the composition is fully parameterised with a distinct composition vector for each node, including the root. The prior on the internal node composition vectors is the overall composition but the prior on leaf nodes is the model composition for that leaf (c.f. Prior A in [94]). For reasons of computational tractability, we only fit this model to the Dayhoff-recoded version of the composition homogeneous 148-gene alignment.

Tests of proposed embryophyte relationships and sequence alignments

We used significance test to compare the support for seven *a priori* hypotheses of early land plant relationships (Table 1; Figure 1) from the site likelihoods of 852 orthologous genes conserved across land plants analyzed under the C60 + Γ empirical mixture model [95]. The full alignment of 852 genes was tested using RELL bootstrapping by generating alignments from within gene partitions in the software IQ-Tree [85]. We then analyzed these site data using the Kishino and Hasegawa (KH) [96–98], Shimodaira and Hasegawa (SH) [99], and approximately unbiased (AU) tests [100]. The weighted KH and SH tests were employed in these analyses.

Tests of proposed embryophyte relationships and inferred gene trees

We estimated 852 gene trees using the C60 + Γ empirical mixture model [95]. Analyses based on these gene trees were done using method of Steel and Rodrigo [25]. In this approach the likelihood of a gene tree is compared to a given or inferred supertree using an exponential function based on the Robinson Foulds distances between the two trees [101]. We employed this approach to a) estimate a supertree given the gene trees in P4 [24] and b) test the fit of gene trees to supertrees representing hypotheses of early land plant relationships using the software Lust [37]. Finally, we also generated a supertree using these gene trees in ASTRAL [26].

Supertree estimation using gene trees

Supertree species trees can be estimated from input gene trees in a likelihood framework [25]. However, under a likelihood model this approach is unfeasible for datasets with large numbers of gene trees that contain many taxa [37]. Therefore, we estimated supertrees in an MCMC framework with the software P4 [24, 102] using the method described by Steel and Rodrigo [25]. We used a normalizing constant to ensure probabilities sum to one, and set the parameter beta (an estimate of data quality and quantity) to be estimated from the data (a method designated SR2008_rf_aZ_fb in P4) [90]. We ran two independent runs, and estimated convergence by checking the likelihood values effective sample size and stationarity in the R package *Coda* [87, 88]. For all methods, we estimated the consensus topology by using trees from each run after burn-in using the R package *APE* [89].

Gene trees fit to proposed topologies

We also used the software Lust [37] to estimate the approximate likelihood of each gene tree when fit to each of the seven hypothesized species trees respectively. We then used each of these approximate likelihood scores to test whether any of the species tree hypotheses could be rejected under three tests of significance (KH, SH, AU) in Consel [86]. For the analyses with 852 and 148 MLE trees, the beta parameter estimated from the P4 supertree was used in the analyses.