RESEARCH PAPERS

Ancient reticulation, incomplete lineage sorting and the evolution of the pyrenoid at the dawn of hornwort diversification

Gabriel Peñaloza-Bojacá [1,*], Adaíses Maciel-Silva [1], D. Christine Cargill [2], David Bell [3], Emily B. Sessa [4], Fay-Wei Li [5], J. Gordon Burleigh [6], Stuart F. McDaniel [6], E. Christine Davis [6], Lorena Endara [7], N. Salazar Allen [8], Peter Schafran [9], Sahut Chantanaorrapint [10], Jeff Duckett [11], Silvia Pressel [11], Claudia Solís-Lemus [12], Karen Renzaglia [13] and Juan Carlos Villarreal A. [14,*]

[1] Departamento de Botânica, Universidade Federal de Minas Gerais, Brazil; adaises.maciel@gmail.com
[2] Australian National Herbarium Canberra, Australia; Chris.Cargill@dcceew.gov.au
[3] Department of Botany, University of British Columbia, Canada; DBell@rbge.org.uk
[4] New York Botanical Garden, USA; esessa@nybg.org
[5] Plant Biology Section, Cornell University, USA; fl329@cornell.edu
[6] Department of Biology, University of Florida, USA; gburleigh@ufl.edu, stuartmcdaniel@ufl.edu, christine.davis@ufl.edu
[7] Department of Biological Sciences, Clemson University, USA; cendara@clemson.edu
[8] Independent Researcher, Panama; SalazarN@si.edu
[9] Boyce Thompson Institute, USA; ps997@cornell.edu
[10] Department of Biology, Faculty of Science, Prince of Songkla University, Thailand; sahut.c@psu.ac.th
[11] Natural History Museum, London, UK; j.g.duckett@qmul.ac.uk, s.pressel@nhm.ac.uk
[12] Wisconsin Institute for Discovery, University of Wisconsin-Madison, USA; solislemus@wisc.edu
[13] Department of Plant Biology, Southern Illinois University Carbondale, USA; renzaglia@siu.edu
[14] Département de Biologie, Université Laval, Canada

* Correspondence: gpenaloza.bojaca@gmail.com; jcvil9@ulaval.ca

Gabriel Peñaloza-Bojacá, **J. Gordon Burleigh**, **Karen Renzaglia**, and **Juan Carlos Villarreal A.** contributed equally to this article.

ABSTRACT

Background and Aims: The evolutionary history of hornworts has been significantly explored through phylogenomic analyses, revealing unexpected complexities in the placement of Leiosporoceros, a putative sister lineage to other hornworts. Our understanding of the evolutionary history of hornworts and the role of incomplete lineage sorting (ILS) or ancient reticulation poses challenges in resolving their exact position and comprehending the diversification and evolutionary processes of hornworts.

Methods: Using the GoFlag probe set, we sequenced 405 exons representing 234 nuclear genes, sampling 79 hornwort specimens, including representatives of all hornwort genera. We inferred the species phylogeny from gene tree analyses using concatenated and coalescence approaches, assessed ancient reticulation, ILS, and estimated the timing of divergences based on fossil calibrations.

Key Results: Our results revealed surprising insights into the placement of *Leiosporoceros*, the putative sister lineage to the other hornworts. Placement of *Leiosporoceros* is complicated due to extreme levels of gene tree incongruence, likely resulting from incomplete lineage sorting and ancient reticulation, the latter identified using a network approach to identify evidence of gene flow among hornwort lineages.

Conclusions: Our analyses showed that divergences within hornwort lineages, particularly during the mid-Cretaceous and Paleogene, correlating with the mass extinction near the Creta-ceous-Paleogene boundary. Similarly, our findings align with changes observed in other groups during this period (e.g., mosses, ferns, and gymnosperms) highlighting a parallel pattern of diversification. Additionally, we showed that hornworts likely originated in Carboniferous, coinciding with a significant decline in atmospheric CO2 levels during the Devonian. As suggested across nearly all photosynthetic organisms possessing pyrenoid, it's likely that the hornwort pyrenoid, potentially evolved in the common ancestor of the group in response to the substantial drop in atmospheric CO2 levels.

Keywords: Bryophytes; CO2 levels; Cretaceous-Paleogene boundary; Evolutionary history; Gene flow; Hybridization; Phylogenomic discordance; Rapid diversification.

64 INTRODUCTION

65 With nearly 230 species worldwide, hornworts are an enigmatic lineage that is sister to the other
66 bryophytes (Villarreal *et al.* 2010). The importance of hornworts lies on a combination of unique
67 features, such as indeterminant sporophyte growth governed by an intercalary meristem, an
68 undifferentiated gametophyte thallus, symbiotic associations with nitrogen-fixing cyanobacteria,
69 and the presence of a pyrenoid-based biophysical carbon-concentrating mechanism (Renzaglia *et al.*
70 2009; Villarreal and Renner 2012; Li *et al.* 2017). Pyrenoids are found throughout eukaryotic algae
71 (red, brown, green, chrysophytes, etc.) and have been repeatedly lost in all lineages (Atkinson *et al.*
72 2016; Meyer *et al.* 2020). The pyrenoid is a liquid-liquid phase aggregation of the enzyme Rubisco
73 and turbocharges CO_2 to the Rubisco active site enhancing photosynthesis in aquatic environments
74 where CO_2 diffuses slower than air (Heureux *et al.* 2017; Meyer *et al.* 2017, 2020). Most of the
75 knowledge of the pyrenoid assembly come from the model green alga *Chlamydomonas reinhardtii*
76 P.A.Dang. and includes constituent transport proteins and the protein-protein interaction that are
77 involved in pyrenoid formation (Uniacke and Zerges 2009; Zhan *et al.* 2018). Hornworts are the
78 only group of land plants possessing a pyrenoid, and the presence of this trait has intrigued
79 biologists for centuries (Renzaglia *et al.* 2009). The most recent reconstruction of pyrenoid
80 evolution inferred that the trait emerged five times during hornwort diversification (Villarreal and
81 Renner 2012), but this analysis was based on organellar data alone.

82 A recent phylogenomic study of bryophytes using 228 nuclear genes resolved many
83 previously ambiguous nodes in liverworts and mosses (Bechteler *et al.* 2023) but also highlighted
84 many cases of gene tree incongruence among bryophyte lineages. The availability of large number
85 of genes for phylogenetic reconstruction has illuminated the complex evolutionary processes
86 underlying the diversification of many clades. For example, conflicting gene tree topologies may
87 result from incomplete lineage sorting (ILS) and/or ancient or gene flow between distinct linages
88 (Knowles *et al.* 2018; Cooper *et al.* 2023). In mosses (Guan *et al.* 2018) and liverworts, ILS may be
89 associated with rapid radiations, and transcriptomic data suggests a history of ancient reticulation

events within liverworts (Dong *et al.* 2022). However, the phylogenomic discordance may also arise due to either systematic error in the analyses or stochastic error in tree reconstruction (Degnan and Rosenberg 2009). Thus, while these phylogenomic data provide new opportunities to elucidate the processes of evolution generating species diversity through time, they also present new analytical challenges that in many cases are relatively unexplored.

Molecular phylogenetic studies of hornworts have generally recovered a consistent topology that is also supported by morphology and serves as the basis for the current classification (Duff *et al.* 2007). *Leiosporoceros* Hässel has a unique suite of morphological characters that include smooth bean-shaped spores in isobilateral tetrads and *Nostoc* strands and pyrenoid less plastids (Villarreal and Renzaglia 2015). *Leiosporoceros* has been frequently recovered as the sister to all other hornworts based on analyses of organellar loci (Duff *et al.* 2007; Villarreal *et al.* 2015; Bell *et al.* 2020), and over 400 loci sequenced for a handful hornwort representatives (Leebens-Mack *et al.* 2019). Based on this topology, Villarreal & Renner (2012) recovered a pyrenoid-free ancestor in all hornworts, particularly due to the fact that *Leiosporoceros* lacks pyrenoids and no other outgroup possess the trait. However, a recent analysis of 12 hornwort species and 234 nuclear loci failed to resolve the node leading to *Leiosporoceros* (Bechteler *et al.* 2023), suggesting alternative topologies that may impact a reconstruction of the pyrenoid suggesting a secondary loss in *Leiosporoceros.*

Here we reevaluate hornwort relationships using data from numerous nuclear loci across all hornwort genera reassessment of hornwort phylogeny and divergence times and their implications for understanding the evolution of the pyrenoid. We also assess the potential role of ILS and reticulation throughout the history of hornworts. We inferred the phylogeny using both concatenation and coalescent-based method and assessed the effects of various data filtering strategies and analytical methods to infer the underlying causes of phylogenetic discordance. We finally address the evolution of the hornwort pyrenoid in the light of our phylogenomic hypotheses.

MATERIALS AND METHODS

Taxon sampling

We sampled a total of 79 hornwort specimens, including representatives of all five families and eleven hornwort genera (Söderström *et al.* 2016). In addition, we sampled six representatives of mosses (*Sphagnum magellanicum* Brid., *Ceratodon purpureus* (Hedw.) Brid., *Pseudanomodon attenuates* (Hedw.) Ignatov and Fedosov, *Bryum argenteum* Hedw.) and ten liverworts (*Riccardia incurvate* Lindb., *Riccardia latifrons* (Lindb.) Lindb., *Sphaerocarpos cristatus* M. Howe, *Sphaerocarpos drewei* Wigglesw*., Marchantia polymorpha* L., *Plagiochasma cordatum* Lehm. and Lindenb., *Ricciocarpos natans* (L.) Corda, *Riccia bicarinata* Lindb., *Riccia beyrichiana* Hampe ex Lehm., *Targionia hypophylla* L., *Monoclea gottschei* Lindb., *Monoclea forsteri* Hook.) as outgroup taxa.

DNA extraction

We extracted DNA using an E.Z.N.A. DNA (Omega, Bio-Tek, United States) extraction kit and Gel Extraction Kit (QIAquick) according to the manufacturer's protocol. In addition was used the modified cetyltrimethylammonium bromide (CTAB) extraction protocol (Doyle and Doyle 1987) described by (Breinholt *et al.* 2021; Bechteler *et al.* 2023) for most of the samples. This protocol includes lysing the cells by centrifuging them and washing with two rounds of 24:1 v/v chloroform–isoamyl alcohol, followed by cold isopropanol precipitation and a 70% v/v ethanol wash. Next was added 2 µL of 10 mg/ml RNase A (QIAGEN, Valencia, CA, USA) to remove RNA contamination between chloroform washes.

Target enrichment and sequencing assembly

We assembled a combination of published target enrichment data from Bechteler et al., 2023; Breinholt et al., 2021 and data that is new to this study. Data from the 16 samples (14 hornworts, 1 moss, and 1 liverwort; Supplementary Data Table S1) first published in Breinholt et al. (2021) were

generated using the GoFlag 451 flagellate land plant probe set, which covers 451 exons found in 248 single or low copy nuclear genes. These unlinked genes appear to be scattered across hornwort chromosomes from species in all clades, based on the genomes *Anthoceros agrestis* Paton, *Leiosporoceros dussii* (Steph.) Hässel, *Phaeoceros carolinianus* (Michx.) Prosk. and *Phaeomegaceros chiloensis* (Steph.) J.C. Villarreal (P. Schafran, pers. comm.). The data from the nine samples (5 hornworts, 2 mosses, and 2 liverworts; Supplementary Data Table S1) first published in Bechteler et al. (2023), as well as the 70 samples (61 hornworts, 3 mosses, and 6 liverworts; Supplementary Data Table S1) new to this study were generated using the GoFlag 408 flagellate land plant probe. The GoFlag 408 probe set is a subset of the GoFlag 451 probe set that covers 408 exons found in 229 genes. Library construction, target enrichment, and sequencing were done by RAPiD Genomics (Gainesville, FL, USA) using protocols described in Breinholt et al. (2021). The enriched, pooled libraries were sequenced on an Illumina HiSeq 3,000 platform (Illumina; 2 × 100 bp), and the paired end raw reads are available in the NCBI SRA database (Supplementary Data Table S1).

We assembled sequence alignments for the target regions covered by the probe sets (i.e., conserved exons) from the raw sequence data using the GoFlag pipeline described by Brienholt et al. (2021; see script 1, this and all scripts are in the online repository). In cases where there appears to be greater than simple allelic variation at a locus, the pipeline may retain more than one sequence from a sample. In these cases, to minimize the inclusion of possible paralogs, we removed all copies from these samples in the locus alignment. Next, we pruned columns in the alignments that had fewer than 10 nucleotides from the output PHYLIP files. We also concatenated the alignments from any exons (i.e., loci) that were found in the same gene (see "GFvs1kp.txt" in script 1, which maps GoFlag loci to their “single copy” 1KP gene). As a result, we obtained 234 alignments, each representing a single gene and containing between one and nine exons. These alignments were used to build “gene” trees. Combining loci from the same gene implicitly assumes that there is no

recombination within the gene; however, longer gene alignment may also reduce stochastic errors in gene-tree inference.

Gene tree and species tree estimation

We inferred maximum likelihood (ML) gene trees from each of 234 genes using RAxML v8.2.9 (Stamatakis 2014) with the GTR+Γ model with 100 random starting points. Statistical confidence of each gene tree was assessed by performing 100 bootstrap (BP) replicates (script 2). To estimate the species tree, we ran ASTRAL-III, which uses a coalescent-based approach to estimate the species tree (Zhang *et al.* 2018), using all 234 ML gene trees from the RAxML analyses as input. We assessed support for the ASTRAL topology based on the local posterior probabilities (LPP) and quartet support values (Sayyari and Mirarab 2016) for the main topology (q1), the first alternative topology (q2), and the second alternative topology (q3; script 3).

We also inferred an ML phylogenetic tree from a concatenated supermatrix of all gene alignments using IQ-TREE v.2.1.3 (Nguyen *et al.* 2015), with "-S option". We partitioned the supermatrix alignment by gene and used ModelFinder (Kalyaanamoorthy *et al.* 2017) to automatically select the best substitution model for each partition. Branch support was assessed using ultrafast bootstrapping (1000 replicates; script 4; Hoang *et al.* 2017).

Measuring concatenation and coalescent-based topological signal

Since our analyses showed discordant topologies in the branch leading to *Leiosporoceros dussii*, we applied a method to assess and compare inconsistent genes between concatenation-based IQ-TREE (T1) and Quartet-based ASTRAL (T2; Shen *et al.* 2021). For incongruent internal bipartition(s) between T1 and T2, we defined a concatenation-based gene-wise phylogenetic signal as the difference in gene-wise log-likelihood score (ΔGLS) for T1 versus T2 and a quartet-based gene-wise phylogenetic signal as the difference in gene-wise quartet score (ΔGQS) for T1 versus T2 (see supplementary information; (Shen *et al.* 2021)). ΔGLS and ΔGQS values can be positive, negative, or zero. We assessed the consistency of gene-wise phylogenetic signal by calculating the following

two measures for every gene: -**consistent**: genes whose ΔGLS >0 and ΔGQS >0 (from T1) and genes whose ΔGLS <0 and ΔGQS <0 (from T2); -**inconsistent**: genes whose ΔGLS ≥0 and ΔGQS ≤0 or vice versa (Shen *et al.* 2021). We also examined the topologies using two different filtered datasets. First, we reduced the dataset by filtering out genes that did not include a sequence from *Leiosporoceros*. This filter datasets dataset (Only_Leios) included 195 genes. We also filtered datasets to include only consistent and only inconsistent genes (Supplementary Data Table S2; script 5).

Branch support and concordance analyses

Branch support was measured using bootstrap and local posterior probabilities (LPP) in IQ-TREE and ASTRAL, respectively. Additionally, we performed concordance analyses in the following datasets: 1) main data set (Full_data with 234 genes and Only_Leios with 195 genes); 2) subset of data (Full_data consistent with 165 genes and Only_Leios consistent with 133 genes); and 3) Divergent topologies (concatenation-based IQ-TREE, T1 and Quartet-based ASTRAL, T2). We calculated gene (gCF) and site concordance factors (sCF) to investigate topological conflict around each branch of the species tree in IQ-TREE with "–gcf and –scf " options (Minh *et al.* 2020). In addition, IQ-TREE estimates gDFP, gene discordant factor due to paraphyly or the gene discordance factor due to lack of information in the genes of the quartet (script 6). These analyses were conducted on every branch of the species tree. The gCF and sCF represents the percentage of decisive gene trees and sites. Gene concordance factor and site concordance factors (gCF/sCF) were categorized as follows: weak < 33%, moderate 33%–50%, strong > 50% (following (Minh *et al.* 2020; Cooper *et al.* 2023; Bechteler *et al.* 2023). We performed exploratory analyses to assess whether gene tree patterns are consistent with the neutral incomplete lineage sorting (ILS) model using IQ-TREE. To do this, a $\chi 2$-test was used to determine whether the frequency of gene trees (gCF) and sites (sCF) supporting the alternate topologies was significantly different (Chan *et al.* 2020). Under the assumption of ILS, the discordant topologies should be supported by an approximately equal number of gene trees or sites, which would result in a non-significant $\chi 2$ (Minh

et al. 2020). Thus, rejecting the χ2 suggests that processes other than ILS, including gene tree error, may be contributing to the discordance. For this test we followed the approach by (Chan *et al.* 2020) such that χ2-test was performed in R (Supplementary Data Table S3; Lanfear *et al.* 2018; Chan *et al.* 2020; R Core Team 2020).

Alternative topologies, especially the position of *Leiosporoceros*, were tested using tree topology tests (with -wpl to print partition-wise log likelihoods for both trees) and testing constrained tree in IQ-TREE (script 7). We used several tree topologies tests in IQ-TREE using the RELL approximation (Kishino *et al.* 1990) including bootstrap proportion (BP), Kishino-Hasegawa test (Kishino and Hasegawa 1989), Shimodaira-Hasegawa test (Shimodaira and Hasegawa 1999), expected likelihood weights (Strimmer and Rambaut 2002) and approximately unbiased (AU) test (Shimodaira 2002). We then identified the most influential genes using partition-wise log likelihoods (-wpl) to identify which genes were contributing most phylogenetic signal towards one tree but not the other. IQ-TREE looks at the gene-wise log-likelihood (logL) differences between the two given trees T1 and T2. Genes that have the largest logL(T1)-logL(T2) will favor T1, whereas genes showing the largest logL(T2)-logL(T1) will favor T2 (Supplementary Data Table S4; Minh *et al.* 2018, 2019).

Network inferences

We used the Species Networks applying the Quartets (SNaQ) method within the PhyloNetworks (Solís-Lemus and Ané 2016; Solís-Lemus *et al.* 2017) to examine the contribution of ILS and reticulation to the phylogenetic history of hornworts. This package uses maximum pseudolikelihood to fit a network while also accounting for ILS. SNaQ uses concordance factors (CFs), which are the frequencies of the three possible unrooted topologies of each set of four taxa (i.e., quartets) in a sample of gene trees (Solís-Lemus *et al.* 2017), as input. Because these analyses are computationally demanding on large datasets, we randomly subsampled at the generic level (55 species from all genera) of hornworts. We then ran SNaQ allowing the maximum number of

reticulations to vary from $h = 0$ to $h = 4$ and selected the best h using a slope heuristic suggested by Solís-Lemus and Ané (2016). As input, we used the 234 gene trees, and the T1 (concatenation-based IQ-TREE) and T2 (Coalescence-based ASTRAL) to limit noise introduced by missing data. Each SNaQ run was performed with 10 optimization iterations (script 8). In addition, we used Hybrid Detector (HyDe) to detect the phylogenetic invariants using a coalescent model with hybridization to infer probability of hybridization of three ingroup taxa relative to an outgroup taxon in hornworts (Blischak *et al.* 2018). The parameter γ represents the probability that gene trees with a hybrid population sister to parent X would arise under the parental population trees, whereas 1-γ is the probability of a hybrid population being sister to parent Y (script 9).

Divergence time estimation

Divergence time estimation was performed using treePL (Smith and O'Meara 2012). The ML tree was the input for treePL based on the Full_data dataset (234 genes) and gene trees previously generated with RAxML (see above) for concatenation-based IQ-TREE (T1) and Coalescence-based ASTRAL (T2). All trees were rooted in the outgroup taxa using the program pxrr in phyx (Brown *et al.* 2017) before being used in the dating analyses (script 10). We constrained three hornworts nodes using fossil evidence (see below) and implemented a maximum age constraint of the root of 515 Ma (Morris *et al.* 2018; Feldberg *et al.* 2021; Ignatov and Maslova 2021; Bechteler *et al.* 2023). Priming and cross-validation analysis was performed using the best-ML tree and all four calibrations. Best optimization parameters for concatenation-based analyses and quartet-based analyses were as follows: for IQ-TREE (T1), "opt=2, moredetail; optad=2, moredetailad; optcvad=1, moredetailcvad"; and for ASTRAL (T2), "opt=5; optad=5; optcvad=3, moredetailcvad". Cross-validation was conducted five times using these parameters and indicated stable values of a smoothing parameter = 10. To obtain confidence intervals on the dated tree, the treePL analysis was run with the bootstrap replicates using the same calibration, optimization and cross-validation values as outlined above. Trees were visualized in FigTree v.1.4.3 (Rambaut 2017).

The fossil record in hornworts is sparse, and earliest reliable fossils are only from the Upper Cretaceous (Villarreal and Renner 2012). We constrained the root to 515 (± 10) Ma, using a normal distribution based on previous studies (Feldberg *et al.* 2021; Ignatov and Maslova 2021; Bechteler *et al.* 2023). The age of *Anthoceros* L. spore type A from the Baqueró Formation, Argentina (118.56 ± 3.7 Ma; Archangelsky and Villar de Seone 1996) was used to constrain the stem node of *Anthoceros* and the rest of the hornworts. The spore exine is ornamented with warts on the proximal face resembling two extant Neotropical species, *Anthoceros tuberculatus* Lehm. et Lindenb. and *Anthoceros tristanianus* J.C. Villarreal, J.J. Engel et Váňa (Peñaloza-Bojacá *et al.* 2020). We used the age of *Notothylites nirulai* from the Deccan Intertrappean beds of Mohgaonka, India (Maastrichtian, 65–70 Ma; Chitaley and Yawale 1980) to constrain the age of the stem node of *Notothylas* Sull. ex A. Gray and *Phaeoceros* Prosk. (including *Paraphymatoceros* Hässel). The petrified fossil of an entire plant has similar thallus size, sporophyte size, lack of stomata in sporophyte, and elater shape as extant *Notothylas*. The third calibration was based on a fossil assigned to *Phaeoceros* sp. from the Uscari formation, Costa Rica (Lower Miocene 15–23 Ma; Graham 1987). This spore fossil has six depressions in its distal face and closely resembles extant *Phaeomegaceros fimbriatus* (Gottsche) R.J. Duff, J.C. Villarreal, Cargill et Renzaglia in size and ornamentation (Villarreal and Renner 2012). For the liverwort outgroup, we incorporated the fossil *Ricciopsis ferganica* dating 227 Ma (Moisan *et al.* 2012). Additionally, we used two constraints 448 Ma for liverworts and 417 Ma for mosses, following the results by Bechteler et al. (2023). We report mean ages and the 95% high posterior density (HPD) values.

RESULTS

Phylogenomic inferences

The ML concatenated supermatrix and the coalescence-based analyses from ASTRAL-III using sequences from 234 genes, containing 80,855 total sites, yielded mostly congruent topologies. Four

of the five hornwort families had similar topologies between analyses with BS support = 100% (Supplementary Data Fig. S1) and LPP = >0.9 (Supplementary Data Fig. S2). The families recovered are: Dendrocerotaceae J.Haseg. (Including *Dendroceros* Nees, *Megaceros* Campb., *Nothoceros* (R.M. Schust.) J. Haseg. and *Phaeomegaceros* R.J. Duff)*;* Phymatocerotaceae R.J. Duff (including only *Phymatoceros* Stotler); Notothyladaceae Müll. Frib. ex Prosk. (including *Notothylas*, *Paraphymatoceros*, and *Phaeoceros*)*;* and Anthocerotaceae Dumort (including *Anthoceros,* and *Folioceros* D.C.Bharadwaj). However, the position of the monotypic genus *Leiosporoceros* differs in the two analyses. ML analysis places *Leiosporoceros* as sister to the rest of the hornworts (BS support = 100%; Fig. 1, T1). ASTRAL-III analyses place *Leiosporoceros* as sister to the family Anthocerotaceae (LPP = 0.7; Fig. 1, T2). Another alternative topology was observed with the coalescence and concatenation analyses (T3), with the Anthocerotaceae family identified as the sister group to the clade consisting of *Leiosporoceros* and the other hornwort families (Fig. 1; T3). This topology exhibited lower support values (LPP = <0.3; BS <70%) at the node corresponding to *Leiosporoceros,* and tree topology tests did also not yield significant values supporting this topology (see below). Therefore, our analyses were centered on making comparisons between T1 and T2.

Genes inconsistency behavior in concatenation and coalescent approaches

To assess the origin of the inconsistency in the placement of *Leiosporoceros*, we quantified the distribution of concatenation-based phylogenetic signal (T1: *Leiosporoceros* sister to all hornworts) with quartet- based phylogenetic signal (T2: *Leiosporoceros* sister to Anthocerotaceae) for every data set (see Table 1). The phylogenomic matrices with ΔGLS and ΔGQS values for every gene are given in Supplementary Data Table S2. We found that when analyzing our full data (234 genes) using concatenation-based methods, the proportion of genes (55.9%) recovering topology T1 (Supplementary Data Fig. S1) is higher than that recovering topology T2 (44.1%). However, with quartet-based analysis, the genes recovered T2 more frequently (52.6% of the 234 genes) than T1 (47.4%). Similar results were obtained when excluding genes that did not in which *Leiosporoceros*

316 was not present (Only Leios: 195 genes; Supplementary Data Table S2). The concatenation
317 approach supported T1 with 107 genes (54.5%; Supplementary Data Fig. S3), while the coalescence
318 approach favored T2 with 98 genes (51.8%; Supplementary Data Fig. S4).

319 The phylogenetic gene signals between concatenation-based (T1) and quartet-based (T2)
320 approaches showed similar results. For the full dataset, 161/234 (68.8%) of the genes were
321 consistent, that is, their ΔGLS and ΔGQS values had the same signs. On the other hand, we found
322 that 73/234 (31.2%) of the genes were not consistent (ΔGLS and ΔGQS values had opposite signs).
323 For the Only_Leios dataset, 133/195 (68.2%) of the genes were consistent, and 62/195 (31.8%) of
324 the genes were not consistent.

325 *Phylogenetic discordance*

326 T1 (*Leiosporoceros* sister to all hornworts) was inferred with the full dataset only with
327 concatenation-based in IQ-TREE and presented high values of support (BS = 100%) and
328 congruence (gCF > 90 % and sCF > 50%; Supplementary Data Fig. S5). IQ-TREE analyses with
329 only the consistent genes produced the same result (Supplementary Data Fig. S6). In contrast, data
330 sets with T2 (*Leiosporoceros* sister to Anthocerotaceae) as the main topology, inferred with quartet-
331 based in ASTRAL (Supplementary Data Fig. S7) and base species tree inferred with coalescence-
332 based method (ASTRAL-III) vs Base tree genes inferred with concatenation method (IQ-TREE;
333 Supplementary Data Fig. S8), presented low levels of support (LPP < 0.7) and high levels of
334 discordance (gCF and sCF < 37%; Figs. 1; S7), even with consistent genes (Supplementary Data
335 Figs. S9, S10). Similar results were obtained from ASTRAL analyses with the Only_Leios genes
336 (Supplementary Data Figs. S4, S11) and their consistent genes (Supplementary Data Figs. S12,
337 S13).

338 Despite the difference between T1 and T2, and the discrepancy values between the data sets,
339 these were not significantly different ($\chi 2 > 0.05$) in the nodes leading to *Leiosporoceros*
340 (Supplementary Data Table S3). For T1, gDF1 and gDF2 were not significantly different

341 ($\chi 2 > 0.05$), and we observed the similar result to sDF1 and sDF2. This is consistent with high levels
342 of incomplete lineage sorting (ILS) in the gene trees. Furthermore, when exploring the
343 *Leiosporoceros* node for T2 across the different data sets, we observed that in most cases, the gDF1
344 and gDF2 were not significantly different ($\chi 2 > 0.05$). However, in the coalescence and consistent
345 gene analyses, gDF1 and gDF2 were significantly different, we cannot reject the null hypothesis of
346 equal frequencies of the alternate topologies ($\chi 2 > 0.05$), suggesting evidence of reticulation. For
347 sDF1 and sDF2 were not significantly different in all data sets for T2. This confirms the presence of
348 ILS and reticulation in the gene trees.

349 The tree topology tests showed a slight tendency to support T1 and to reject T2 and T3
350 (Table 2). In addition, we found that the most influential genes have a similar result for T1 and T2
351 topologies. Where our full data (234 genes) showed T1= 90 genes, T2= 78 genes, and T3= 62
352 genes; Full data of the genes were consistent (161 genes) showed T1= 61 genes, T2= 63 genes, and
353 T3= 36; Only Leios (195 genes) showed T1= 79 genes, T2= 63 genes, and T3= 52; and Only Leios
354 of the genes were consistent (133 genes) showed T1= 51 genes, T2= 55 genes, and T3= 27.
355 Therefore, the contribution of genes may be similar in the phylogenetic signal at T1 and T2.
356 Another node revealing some incongruence was the nodes leading to *Notothylas*,
357 *Paraphymatoceros*, and *Phaeoceros*. The node resolved the same topology using concatenation and
358 coalescent-based methods, but with gDFP values close to 30%, which indicates rates of paraphyly
359 or lack of information in genes (nodes G, H, I in Fig. 1, respectively; Supplementary Data Figs. S3,
360 S5-S13).

361 *Reticulation events in hornworts*

362 We explored phylogenetic networks with different numbers of hybridization events (*h*= 0 to 3;
363 Supplementary Data Figs. S14, S15). The best scores were two reticulations events for T1 (-Ploglik
364 = 255.42; Supplementary Data Fig. S16) and three for T2 (-Ploglik = 254.74; Supplementary Data
365 Fig. S17), suggesting reticulation events in the evolutionary history of hornworts (Fig. 2).

Two hybridization events were recovered based on T1 (Fig. 2A; Supplementary Data Table S5). The first in *Nothoceros* with $\gamma = 0.115$ (gamma value), showing a low hybridization signal, and the stem branch leading to *Leiosporoceros*, *Anthoceros* and *Folioceros* with $\gamma = 0.885$, showing a high hybridization signal. The second event occurred between *Leiosporoceros* with $\gamma = 0.0292$ and *Folioceros* with $\gamma = 0.971$, suggesting a possible hybridization event in *Folioceros*, having *Leiosporoceros* and *Anthoceros* as parental groups. On the other hand, when observing the Quartet CF (observed vs. expected) derived from the gene trees of the SNaQ hybridization analysis, we identified potential discrepancies in the phylogenetic relationships of quartets within the gene trees compared to the phylogenetic network, particularly concerning taxa interactions involving *Leiosporoceros* (Fig. 2C). The results suggest discrepancies or possible hybridization events involving this genus in the gene trees compared to the phylogenetic network.

Likewise, three reticulation events were detected in the analyses that included T2 (Fig. 2B; Supplementary Data Table S5). The first among *Notothylas* with a value of $\gamma = 0.069$ and *Phaeoceros* with a value of $\gamma = 0.931$, suggesting a possible hybridization event in *Phaeoceros*, with *Notothylas* and *Paraphymatoceros* as parental groups. The second event occurred between the *Dendroceros* and *Megaceros* clade with a value of $\gamma = 0.0929$ and the *Leiosporoceros*, *Anthoceros* and *Folioceros* clade with a value of $\gamma = 0.907$. The third event was recorded in *Anthoceros* with a value of $\gamma = 0.972$ and a possible extinct or unsampled taxon with a value of $\gamma = 0.0278$. *Anthoceros* and this species possibly presented a case of reticulated evolution with *Leiosporoceros* and an extinct taxon. Additionally, Quartet CF (observed vs expected) showed reasonable agreement between the gene trees and the phylogenetic network (Fig. 2C), suggesting a strong correlation between these two variables. Despite limited conflict in network analyzes and quartet scoring, gene flow analysis using HyDe showed extensive gene flow among hornwort genera, and there were only a few pairs of taxa between which no gene flow was identified (Supplementary Data Table S6).

Hornworts divergence times

Two topologies (T1 and T2) were dated to explore the alternative evolutionary timelines of hornworts implied by the different trees. The crown age of hornworts ranges from the 337,6 Ma (HPD 330,7-348,5 Ma) for T1 and 322,5 Ma (HPD 316,8-328,2 Ma) for T2 (Supplementary Data Figs. S18, S19 for HPD values for all clades) in the Carboniferous. Both chronograms showed similar divergence ages for most taxonomic groups (Fig. 1). For the orders we estimated an origin (stem age) in the Carboniferous and Permian times for Leiosporocerotales Hässel T1: 337,6 Ma (HPD: 330,7-348,6 Ma) and T2: 288,3 Ma (HPD: 316,8-328,2 Ma); Anthocerotales Limpr. T1: 299,4 Ma (HPD: 291,9-307,5 Ma) and T2: 288,3 Ma (HPD: 280,3-295,5 Ma) originated in the Permian. Notothyladales Hyvönen et Piippo and Dendrocerotales Hässel originated in the Jurassic T1: 182,1 Ma (HPD: 170,7-193,5 Ma) and T2: 190,5 Ma (HPD: 182,9-198,6 Ma), and Phymatocerotales R.J. Duff T1: 129,3 Ma (HPD: 120,1-139,5 Ma) and T2: 129,2 Ma (HPD: 119,4-138,0 Ma) in the Cretaceous times. Finally, most genera diversified (crown age) in the Paleogene (Fig. 1; Supplementary Data Figs. S18, S19), from HPD 24-58 Ma in T1 and HPD 25-59 Ma in T2. *Paraphymatoceros* has an origin between the Cretaceous and Paleogene T1: 73,6 Ma (HPD: 57,4-93,2 Ma) and T2: 73,9 Ma (HPD: 57,0-91,5 Ma). *Phymatoceros* has a Neogene origin T1: 17,8 Ma (HPD:14,3-23,2 Ma) and T2: 17,9 Ma (HPD: 13,5-23,8 Ma).

DISCUSSION

Prior to the advent of molecular phylogenetics, the systematics and classification of hornworts were widely debated and subject to conflicting interpretations (Hasegawa 1988; Hässel de Menéndez 1988). Following the first comprehensive molecular analysis across hornwort diversity (Duff *et al.* 2007), a consensus emerged that challenged existing concepts of hornwort relationships. The widely recognized 5 or 6 genera were segregated into 10 (11) genetically distinct taxa. *Leiosporoceros* surfaced as the sister taxon to the other hornworts, and the distinct morphological features of the

genus provided support for this position (Villarreal and Renzaglia 2006; Renzaglia *et al.* 2007). Our analysis of sequences from 234 nuclear genes from 60 hornwort species, which includes representatives of all genera, is consistent with previous studies using organellar loci with one important exception (Duff *et al.* 2007; Villarreal *et al.* 2015): our analyses reveal incongruence with respect to the placement of the branch leading to *Leiosporoceros.* We have shown that the position of *Leiosporoceros*, at the dawn of hornwort diversification may be influenced by incomplete lineage sorting and ancient reticulation events.

Our analyses also support a Carboniferous origin of hornworts, coinciding with the significant decrease in CO_2 levels beginning in the Devonian period (Fig. 1). We propose that the hornwort pyrenoid may have emerged as a response to the substantial reduction in atmospheric CO_2, triggering a carbon concentration mechanism in the most recent common ancestor of the group with subsequent losses. This hypothesis has been put forward for all photosynthetic organisms possessing pyrenoids, ranging for chlorarachniophytes to green algae as a case of convergent evolution of this complex trait (Villarreal and Renner 2012; He *et al.* 2023; Ruaud *et al.* 2024). All those organisms have a very ancient (even Precambrian) origin while hornworts as a clade evolved in the Carboniferous. Another equally plausible hypothesis is that the pyrenoid existed in the ancestor of all land plants and was subsequently lost in setaphytes (mosses and liverworts) and tracheophytes (Robison *et al.* 2024). Within hornworts, the trait has been lost several times with some genera entirely lacking pyrenoids such *Leiosporoceros*, and *Megaceros*. In contrast, most taxa of all rich genera such *Anthoceros* (67 spp.) *Phaeoceros* (35) and *Dendroceros* (39) possess pyrenoids (Fig. 3). Additionally, *Nothoceros vincentianus* (Lehm. et Lindenb.) J.C. Villarreal has individuals specimens both with and without plastids containing pyrenoids (Villarreal and Renner 2012). The function, genetic underpinnings, and physiology of the hornwort pyrenoid remain poorly known.

Gene tree/species tree conflict and ILS in hornworts

Most previous studies based on a few organellar loci with high levels of hornwort sampling (Duff *et al.* 2007; Renzaglia *et al.* 2009; Villarreal and Renner 2013; Villarreal *et al.* 2015; Villarreal and Renzaglia 2015) or low levels of hornwort sampling with a large number of markers (Leebens-Mack *et al.* 2019; Breinholt *et al.* 2021) recovered *Leiosporoceros* as the sister to all hornworts. The quite distinct gametophyte and sporophyte morphology of *Leiosporoceros* supported this phylogenetic placement. However, Duff et al. (2007, their Fig. 11) removed 193 edited sites from *rbc*L and *nad*5 and recovered *Leiosporoceros* as sister to Anthocerotaceae, suggesting that the widely accepted position of the taxon may be due to the low rate of RNA editing in *Leiosporoceros* (Villarreal *et al.* 2018). RNA edited sites significantly contribute to the overall number of phylogenetically informative positions, thereby influencing the reconstruction of phylogeny. A recent nuclear gene phylogeny by Bechteler et al. (2023) that included 12 of species of hornworts also questioned the placement of *Leiosporoceros*, suggesting high levels of phylogenomic incongruence at the *Leiosporoceros* node.

We recovered two distinct topologies, depending on the method of analysis. ML concatenation analysis strongly supports *Leiosporoceros* sister to all hornworts (T1). Concatenation implicitly assumes that all genes share the same history, and therefore, it does not account for ILS or reticulation. This can introduce biases, which may lead to inconsistent or erroneous results with high confidence (Song *et al.* 2012; Xi *et al.* 2014; Cai *et al.* 2021). In contrast, our ASTRAL-III analyses, which are based on a multispecies coalescent (MSC) model, tend to recover *Leiosporoceros* sister to Anthocerotaceae (T2), although with low support (Song *et al.* 2012; Xi *et al.* 2014; Reddy *et al.* 2017; Zhang *et al.* 2018; Cai *et al.* 2021). However, these analyses may be biased by error in the input gene trees.

The ASTRAL results revealed gCFs and sCFs values were similar between T1 and T2 in the various datasets analyzed, with high levels of discordance in the placement of *Leiosporoceros* (gCF and sCF < 37%; see Fig. 1). Two topologies have nearly the same proportion of trees and sites supporting them, and the lack of resolution might be attributed to presence of ILS. Incomplete

lineage sorting is a prominent biological factor responsible for generating higher levels of inconsistency among topologies and has been reported in nearly all plant groups (Pérez-Escobar *et al.* 2021). In liverworts, Marchantiales and Jungermanniopsida exhibit cytonuclear inconsistencies in their topologies due to the presence of ILS (Dong *et al.* 2022), which also appears to occur in the moss *Sphagnum* (Meleshko *et al.* 2021).

Evidence of hybridization in hornworts

Our hybridization analyses confirm the present of ancestral reticulation and introgressions among the major ancestral lineages of hornworts (Fig. 2), suggesting complex patterns of gene flow among genera during their diversification. This contributes to the discordance observed in our two genetic tree topologies (T1 and T2, Figs. 1, 2) as well as the phylogenetic position of *Leiosporoceros*. Hybridization has proven to be a significant source of variation in the evolutionary history of organisms (Anderson 1949; Taylor and Larson 2019) because of the potential to generate novel characteristics more rapidly than other processes such as mutation (Soltis 2013; Suarez-Gonzalez *et al.* 2018; Stull *et al.* 2023).

Although hybridization has been suggested in hornworts (Proskauer 1957, 1969), it has been considered to be rare or of limited evolutionary importance in the group. In contrast, hybrids have been reported between species in mosses and liverworts (Natcheva and Cronberg 2004) and ancient reticulation has been proposed between major liverwort classes (Dong *et al.* 2022). For example, in mosses, there is evidence of hybridization between ancestral lineages of *Sphagnum* L. (Natcheva and Cronberg 2007). Consequently, recent gene flow between species within the genus is limited, despite the interspecific hybridization documented in the group (Meleshko *et al.* 2021). Similarly, in liverworts, signals of ancient hybridization and ILS were reported in the backbone phylogeny involving the ancestors of simple thalloids, leafy liverworts, and complex thalloids (Dong *et al.* 2022). Introgression and hybridization events have been identified in groups across the tree of life, including cyanobacteria (Pardo-De la Hoz *et al.* 2023), lichens (Keuler *et al.* 2022), ferns (Huang *et*

al. 2020; Chen *et al.* 2023), angiosperms (Sun *et al.* 2020), Rosaceae (Hodel *et al.* 2022; Debray *et al.* 2022), gymnosperms (Liu *et al.* 2022), dragonflies (Suvorov *et al.* 2022), snakes (Schöneberg *et al.* 2023), and primates (Vanderpool *et al.* 2020). The widespread occurrence of these events underscores the fundamental role of hybridization in shaping the evolutionary trajectory and diversification of organisms.

Timing of diversification in hornworts and the evolution of the pyrenoid

Since the placement of *Leiosporoceros* remains elusive, we explored the timing of hornworts diversification in two distinct scenarios (T1 and T2). Our estimates place the origin of the crown age of hornworts in the Carboniferous period, in line with previous studies (Villarreal and Renner 2012; Morris *et al.* 2018; Harris *et al.* 2022; Bechteler *et al.* 2023). The Carboniferous and Cretaceous-Paleogene origin of hornworts and their diversity may have been influenced by the environmental characteristics of the time. The atmospheric drop of CO_2 levels started in the Devonian (>275 ppm) and went down to nearly current levels in the Carboniferous (<190 ppm; Igamberdiev and Lea 2006), potentially triggering the evolution of carbon concentrating mechanism in algae (Badger *et al.* 2000). A physical carbon concentration mechanism such as a pyrenoid is found scattered across many algal lineages (Meyer *et al.* 2020; Mukherjee 2020; Barrett *et al.* 2021). The polyphyletic pyrenoid evolved in aquatic environments to overcome the low diffusion of CO_2 in water, and it seems to be an ancient structure present in the common ancestors of all photosynthetic eukaryotes. The evolution of the hornwort pyrenoid has remained enigmatic because of the terrestrial nature of these plants.

The topology in which *Leisporoceros* is sister to the rest of the hornworts (T1), and the absence of pyrenoids in *Leiosporoceros* and outgroup taxa (other land plants) weighed heavily on previous ancestral trait reconstructions that identified a de novo origin of the hornwort pyrenoid (Villarreal & Renner 2012). It is possible that pyrenoids evolved in response to the dramatic drop of CO2 in the Carboniferous. The structurally distinct and complex pyrenoid in *Dendroceros* may be

related to the microenvironment of this epiphytic genus (Vaughn *et al.* 1992; Schuette and Renzaglia 2010; Villarreal and Renner 2012). The alternative topology presented here (T2) suggests that the ancestor of the group may have possessed the trait. Since the pyrenoid has evolved in many ancient algal lineages, we cannot rule out the possibility that the pyrenoid was present in the ancestor of all land plants and subsequently lost in setaphytes and tracheophytes.

Our findings also highlight a period of heightened diversification in hornworts during the mid-Cretaceous and Paleogene with nearly synchronous emergence of all genera, similar to the cycad radiation (Nagalingum *et al.* 2011; Condamine *et al.* 2015). This period of increased diversification aligns with the aftermath of the Cretaceous–Paleogene (K–Pg) mass extinction (Springer *et al.* 2003; Renne *et al.* 2013; Vajda and Bercovici 2014). Changes in pollen and spore assemblages across the K-Pg boundary have provided crucial insights into vegetation responses during a global environmental crisis triggered by the asteroid impact (Vajda and Bercovici 2014). The Cretaceous period is also characterized by significant shifts in terrestrial plant and animal communities, such as the explosion of flowering plants (Davis *et al.* 2005; Lutzoni *et al.* 2018; Donoghue *et al.* 2021) and mammal diversification (Vajda and Bercovici 2014; Ramírez-Barahona *et al.* 2020).

Bryophytes represent another group of plants that underwent significant diversification during the Cretaceous-Paleogene period. For example, species-rich lineages of liverworts and mosses, such as Hypnales W.R. Buck & Vitt and Lejeuneaceae Cavers, diversified nearly in parallel with angiosperms during this time (Feldberg *et al.* 2014; Laenen *et al.* 2014; Bechteler *et al.* 2023). Additionally, there is evidence of profound divergences in complex thalloid liverwort species near the K/T global extinction event. These species exhibited a xerophytic lifestyle with desiccation tolerance strategies (e.g., *Corsinia* Raddi, *Exormotheca* Mitt., *Plagiochasma* Lehm., *Riccia* L., *Targionia* L. (Bischler-causse *et al.* 2012; Villarreal *et al.* 2016). Concurrently, other organisms experienced bursts of diversification, such as leptosporangiate ferns (Schuettpelz and Pryer 2009),

conifers (Taxodiaceae, Cupressaceae, Araucariaceae; Axsmith and Jacobs 2005), cycads, and ginkgoes (Johnson 2002; Vajda and Bercovici 2014).

SUPPLEMENTARY DATA

Supplementary data are available at Annals of Botany online and consist of the following.

Figure S1: Hornworts phylogeny inferred with concatenation method (IQ-TREE) from nucleotide data showing bootstrap support with 234 genes.

Figure S2: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data including the local posterior probability (Lpp) values with 234 genes.

Figure S3: Hornworts phylogeny inferred with concatenation method (IQ-TREE) from nucleotide data with 195 genes (only genes with *Leiosporoceros dussii*), including bootstrap support and concordance values for each node outlined in Suppl. Table S4.

Figure S4: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data with 195 genes (only genes with *Leiosporoceros dussii*), including local posterior probability and concordance values for each node outlined in Suppl. Table S4.

Figure S5: Hornworts phylogeny inferred with concatenation method (IQ-TREE) from nucleotide data with 234 genes including bootstrap support and concordance values for each node outlined in Suppl. Table S4.

Figure S6: Hornworts phylogeny inferred with concatenation method (IQ-TREE) from nucleotide data with 161consistent genes including bootstrap support and concordance values for each node outlined in Suppl. Table S4.

Figure S7: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data with 234 genes including local posterior probability values and concordance values for each node outlined in Suppl. Table S4.

Figure S8: Hornworts phylogeny inferred from nucleotide data with 234 genes including local posterior probability values and concordance values for each node outlined in Suppl. Table S4. Base species tree inferred with coalescence-based method (ASTRAL-III) vs Base tree genes inferred with concatenation method (IQ-TREE).

Figure S9: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data with 161 consistent genes including local posterior probability values and concordance values for each node outlined in Suppl. Table S4.

Figure S10: Hornworts phylogeny inferred from nucleotide data with 161 consistent genes including local posterior probability values and concordance values for each node outlined in Suppl. Table S4. Base species tree inferred with coalescence-based method (ASTRAL-III) vs Base tree genes inferred with concatenation method (IQ-TREE).

Figure S11: Hornworts phylogeny inferred from nucleotide data with 195 genes (only genes with *Leiosporoceros dussii*), including local posterior probability values and concordance values for each node outlined in Suppl. Table S4. Base species tree inferred with coalescence-based method (ASTRAL-III) vs Base tree genes inferred with concatenation method (IQ-TREE).

Figure S12: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data with 133 consistent genes (only genes with *Leiosporoceros dussii*), including local posterior probability values and concordance values for each node outlined in Suppl. Table S4.

Figure S13: Hornworts phylogeny inferred from nucleotide data with 133 consistent genes (only genes with *Leiosporoceros dussii*), including local posterior probability values and concordance

values for each node outlined in Suppl. Table S4. Base species tree inferred with coalescence-based method (ASTRAL-III) vs Base tree genes inferred with concatenation method (IQ-TREE).

Figure S14: Estimates of ancient reticulate evolution (SNaQ) from all hornwort genera using 234 genes and hmax = 3 (number of hybridizations). Estimated network based on T1 topology (*Leiosporoceros* sister to all hornworts) with best scores across hybrid values (-Ploglik).

Figure S15: Estimates of ancient reticulate evolution (SNaQ) from all hornwort genera using 234 genes and hmax = 3 (number of hybridizations). Estimated network based on T2 topology (*Leiosporoceros* sister to Anthocerotaceae) with best scores across hybrid values (-Ploglik).

Figure S16: Plot scores between hybrid values. Hmax are the number of hybridization events (0-3). |It was based on T1 topology (Leiosporoceros sister to all hornworts).

Figure S17: Plot scores between hybrid values. Hmax are the number of hybridization events (0-3). It was based on T2 topology (*Leiosporoceros* sister to Anthocerotaceae).

Figure S18: Divergence time estimates for hornwort genera using the concatenation method (IQ-TREE) based on 234 nuclear genes. Node heights represent mean ages, and bars indicate the 95% highest posterior density intervals.

Figure S19: Divergence time estimates for hornwort genera using the coalescence-based method (ASTRAL-III) based on 234 nuclear genes. Node heights represent mean ages, and bars indicate the 95% highest posterior density intervals.

Figure S20: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data including the quartet values with 234 genes, for the three possible topologies, presented as percentages near the node (100/0/0).

Figure S21: Hornworts phylogeny inferred with coalescence-based method (ASTRAL-III) from nucleotide data including the local posterior probability (lpp) values with 234 genes, for the three possible topologies, presented as percentages near the node (100/0/0).

610 **Table S1:** Taxon sampling Taxon sampling indicating their accession number, DNA sample ID and
611 its respective voucher information (i.e., Collection number, Locality and Herbarium). Species are
612 listed in alphabetical order showing their respective taxonomic ranks.

613 **Table S2:** Consistent and inconsistent genes between concatenation-based IQ-TREE (T1) and
614 quartet-based ASTRAL (T2) in hornworts.

615 **Table S3:** Concordance factors for the nucleotide data stemming from IQ-TREE and ASTRAL
616 analyses from hornworts.

617 **Table S4:** Test alternative topologies in IQ-TREE, identifying most influential genes with partition-
618 wise log likelihoods. FD: Full-Data; FD_Co: Full-Data consistent genes, ON: Only genes whit
619 *Leiosporoceros*, OLCo: Only genes consistent whit *Leiosporoceros*.

620 **Table S5:** Analyses with SNaQ to explore possible reticulation events in the evolutionary history of
621 hornworts.

622 **Table S6:** Analysis in HyDe with gene flow for hornworts. Database with 234 genes and all
623 hornworts genera.

624

625 AVAILABILITY OF DATA

626 The raw sequence reads for all samples have been deposited in the NCBI SRA database (and all
627 SRA accession numbers can be found in Supplementary Data Table S1). All scripts, phylogenetic
628 nucleotide alignments, and resulting gene trees and species trees are available at Github:
629 https://github.com/gpenalozabojaca/Hornwort-diversification-.git.

630

631

632

AUTHOR CONTRIBUTIONS

The concept and methodology of the study was developed by G.F.B.P., J.G.B., K.R., and J.C.V.A.; specimens were collected by G.F.B.P., J.G.B., K.R., A.S.M.S., D.C.C, D.B., E.B.S., F.W.L., S.F.M., E.C.D., L.E., N.S.A, P.S., S.C., J.D., S.P., C.S.L., and J.C.V.A.; DNA was extracted and data were generated by G.F.B.P., J.G.B., L.E, and J.C.V.A.; analyses were performed by G.F.B.P. and J.C.V.A.; writing—original draft preparation by G.F.B.P., K.R., and J.C.V.A.; writing—review and editing by G.F.B.P., J.G.B., K.R., A.S.M.S., D.C.C, D.B., E.B.S., F.W.L., S.F.M., E.C.D., L.E., N.S.A, P.S., S.C., J.D., S.P., C.S.L., and J.C.V.A.; funding for the study was obtained by G.F.B.P., J.G.B., E.C.D., E.B.S., S.F.M., P.S., and J.C.V.A. All authors have read and agreed to the published version of the manuscript.

FUNDING INFORMATION

This work was supported by the NSF collaborative project "Building a Comprehensive Evolutionary History of Flagellate Plants" [DEB 1541506 to J.G.B., E.C.D., S.F.M., and E.B.S.]. P.S. was funded by the NSF IOS [2109789]. G.F.P.B. was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nivel Superior, Brasil [CAPES- Finance Code 001], the Latin American and the Caribbean Macro Universities, the postgraduate Mobility Program 02-2017, the International Association for Plant Taxonomy [IAPT—Research Grant 2019], the Emerging Leaders in the Americas Program – ELAP [2018-2019]; and the Rufford Foundation [Small Grants 2018-2019]. J.C.V.A. was funded by the Canada Research Chair [950-232698] the CRNSG- RGPIN [05967–2016] and the Canadian Foundation for Innovation [projects 36781, 39135].

ACKNOWLEDGMENTS

We thank all the herbaria that provided specimens and the anonymous reviewers and the Associate Editor for their constructive comments on previous versions of the manuscript. G.F.P.B. and J.C.V.A. thank Adriel Sierra and the Louis-Marie Herbarium (UL) for the support.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

678 LITERATURE CITED

679 **Anderson E**. **1949**. *Introgressive hybridization*. New York: J. Wiley.

680 **Archangelsky S, Villar de Seone L**. **1996**. Estudios palinógicos de la formación Baqueró
681 (Cretácico), provincia de Santa Cruz, Argentina. *Ameghiniana* **33**: 7–19.

682 **Atkinson N, Feike D, Mackinder L, *et al.* 2016**. Introducing an algal carbon-concentrating
683 mechanism into higher plants: location and incorporation of key components. *Plant Biotechnology*
684 *Journal* **14**: 1302–1315.

685 **Axsmith BJ, Jacobs BF**. **2005**. The conifer *Frenelopsis ramosissima* (Cheirolepidiaceae) in the
686 lower Cretaceous of texas: systematic, biogeographical, and paleoecological implications.
687 *International Journal of Plant Sciences* **166**: 327–337.

688 **Badger MR, von Caemmerer S, Ruuska S, Nakano H**. **2000**. Electron flow to oxygen in higher
689 plants and algae: rates and control of direct photoreduction (Mehler reaction) and rubisco
690 oxygenase (C.B. Osmond, C.H. Foyer, and G. Bock, Eds.). *Philosophical Transactions of the Royal*
691 *Society of London. Series B: Biological Sciences* **355**: 1433–1446.

692 **Badger MR, Hanson D, Price GD**. **2002**. Evolution and diversity of CO2 concentrating
693 mechanisms in cyanobacteria. *Functional Plant Biology* **29**: 161.

694 **Barrett J, Girr P, Mackinder LCM**. **2021**. Pyrenoids: CO2-fixing phase separated liquid
695 organelles. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1868**: 118949.

696 **Bechteler J, Peñaloza-Bojacá G, Bell D, *et al.* 2023**. Comprehensive phylogenomic time tree of
697 bryophytes reveals deep relationships and uncovers gene incongruences in the last 500 million years
698 of diversification. *American Journal of Botany* **110**: 1–20.

699 **Bell D, Lin Q, Gerelle WK, *et al.* 2020**. Organellomic data sets confirm a cryptic consensus on
700 (unrooted) land-plant relationships and provide new insights into bryophyte molecular evolution.
701 *American Journal of Botany* **107**: 91–115.

Berner RA. **2001**. GEOCARB III: A revised model of atmospheric CO2 over Phanerozoic time. *American Journal of Science* **301**: 182–204.

Bischler-causse H, Gradstein R, Jovet-ast S, Long D, Allen NS. **2012**. Marchantiidae. *Flora Neotropica Monographs* **97**: 1–267.

Blischak PD, Chifman J, Wolfe AD, Kubatko LS. **2018**. HyDe: A Python package for genome-scale hybridization detection (D Posada, Ed.). *Systematic Biology* **67**: 821–829.

Breinholt JW, Carey SB, Tiley GP, *et al.* 2021. A target enrichment probe set for resolving the flagellate land plant tree of life. *Applications in Plant Sciences* **9**: 1–28.

Brown J, Walker J, Smith S. **2017**. Phyx: phylogenetic tools for unix (J Kelso, Ed.). *Bioinformatics* **33**: 1886–1888.

Cai L, Xi Z, Lemmon EM, *et al.* 2021. The perfect storm: gene tree estimation error, incomplete lineage sorting, and ancient gene flow explain the most recalcitrant ancient angiosperm clade, Malpighiales (M Fishbein, Ed.). *Systematic Biology* **70**: 491–507.

Chan KO, Hutter CR, Wood PL, Grismer LL, Brown RM. **2020**. Target-capture phylogenomics provide insights on gene and species tree discordances in Old World treefrogs (Anura: Rhacophoridae). *Proceedings. Biological sciences* **287**: 20202102.

Chen H, Fang Y, Zwaenepoel A, Huang S, Van de Peer Y, Li Z. **2023**. Revisiting ancient polyploidy in Leptosporangiate ferns. *New Phytologist* **237**: 1405–1417.

Chitaley S, Yawale N. **1980**. On *Notothylites nirulai* gen. et sp. nov. a petrified sporogonium from the Deccan-Intertrappean beds of Mohgaonkalan M. P. (India). *Botanique* **9**: 111–118.

Condamine FL, Nagalingum NS, Marshall CR, Morlon H. **2015**. Origin and diversification of living cycads: a cautionary tale on the impact of the branching process prior in bayesian molecular dating. *BMC Evolutionary Biology* **15**: 65.

Cooper BJ, Moore MJ, Douglas NA, *et al.* 2023. Target enrichment and extensive population sampling help untangle the recent, rapid radiation of *Oenothera* sect. *Calylophus* (J Beaulieu, Ed.). *Systematic Biology* **72**: 249–263.

Davis CC, Webb CO, Wurdack KJ, Jaramillo CA, Donoghue MJ. **2005**. Explosive radiation of Malpighiales supports a mid-cretaceous origin of modern tropical rain forests. *The American Naturalist* **165**: E36–E65.

Debray K, Le Paslier M-C, Bérard A, *et al.* 2022. Unveiling the patterns of reticulated evolutionary processes with phylogenomics: hybridization and polyploidy in the genus *Rosa* (J Mandel, Ed.). *Systematic Biology* **71**: 547–569.

Degnan JH, Rosenberg NA. **2009**. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* **24**: 332–340.

Dong S, Yu J, Zhang L, Goffinet B, Liu Y. **2022**. Phylotranscriptomics of liverworts: revisiting the backbone phylogeny and ancestral gene duplications. *Annals of Botany*: 1–13.

Donoghue P, Harrison C, Paps J, Schneider H. **2021**. The evolutionary emergence of land plants. *Current Biology* **31**: R1281–R1298.

Doyle Jeff, Doyle Jane. **1987**. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bulletin* **19**: 11–15.

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

Feldberg K, Gradstein SR, Gröhn C, *et al.* 2021. Checklist of fossil liverworts suitable for calibrating phylogenetic reconstructions. *Bryophyte Diversity and Evolution* **43**: 14–71.

Feldberg K, Schneider H, Stadler T, Schäfer-Verwimp A, Schmidt AR, Heinrichs J. **2014**. Epiphytic leafy liverworts diversified in angiosperm-dominated forests. *Scientific Reports* **4**: 5974.

748 **Graham A**. **1987**. Miocene communities and paleoenvironments of southern Costa Rica. *American*
749 *Journal of Botany* **74**: 1501–1518.

750 **Guan Y, Liu L, Wang Q, *et al.* 2018**. Gene refashioning through innovative shifting of reading
751 frames in mosses. *Nature Communications* **9**: 1555.

752 **Harris BJ, Clark JW, Schrempf D, *et al.* 2022**. Divergent evolutionary trajectories of bryophytes
753 and tracheophytes from a complex common ancestor of land plants. *Nature Ecology & Evolution* **6**:
754 1634–1643.

755 **Hasegawa J**. **1988**. A proposal for a new system of the Anthocerotae, with a revision of the genera.
756 *Journal of the Hattori Botanical Laboratory* **64**: 87–95.

757 **Hässel de Menéndez G**. **1988**. A proposal for a new classification of the genera within the
758 Anthocerotophyta. *Journal of the Hattori Botanical Laboratory* **64**: 71–86.

759 **He S, Crans VL, Jonikas MC**. **2023**. The pyrenoid: the eukaryotic CO2-concentrating organelle.
760 *The Plant Cell* **35**: 3236–3259.

761 **Heureux A, Young J, Whitney S, *et al.* 2017**. The role of Rubisco kinetics and pyrenoid
762 morphology in shaping the CCM of haptophyte microalgae. *Journal of Experimental Botany* **68**:
763 3959–3969.

764 **Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh L**. **2017**. UFBoot2: Improving the
765 ultrafast bootstrap approximation. *Molecular Biology and Evolution* **35**: 518–522.

766 **Hodel RGJ, Zimmer EA, Liu B-B, Wen J**. **2022**. Synthesis of nuclear and chloroplast data
767 combined with network analyses sup-ports the polyploid origin of the apple tribe and the hybrid
768 origin of the Male-ae—Gillenieae clade. *Frontiers in Plant Science* **12**: 1–15.

769 **Huang C, Qi X, Chen D, Qi J, Ma H**. **2020**. Recurrent genome duplication events likely
770 contributed to both the ancient and recent rise of ferns. *Journal of Integrative Plant Biology* **62**:
771 433–455.

Igamberdiev AU, Lea PJ. **2006**. Land plants equilibrate O2 and CO2 concentrations in the atmosphere. *Photosynthesis Research* **87**: 177–194.

Ignatov MS, Maslova E V. 2021. Fossil mosses: What do they tell us about moss evolution? *Bryophyte Diversity and Evolution* **43**: 72–97.

Johnson KR. **2002**. Megaflora of the hell creek and lower fort union formations in the Western Dakotas: vegetational response to climate change, the cretaceous-tertiary boundary event, and rapid marine transgression In: *The Hell Creek Formation and the Cretaceous-Tertiary boundary in the northern Great Plains: An Integrated continental record of the end of the Cretaceous*. Geological Society of America, 329–391.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. **2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.

Keuler R, Jensen J, Barcena-Peña A, *et al.* 2022. Interpreting phylogenetic conflict: hybridization in the most speciose genus of lichen-forming fungi. *Molecular Phylogenetics and Evolution* **174**: 107543.

Kishino H, Hasegawa M. **1989**. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *Journal of Molecular Evolution* **29**: 170–179.

Kishino H, Miyata T, Hasegawa M. **1990**. Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *Journal of Molecular Evolution* **31**: 151–160.

Knowles LL, Huang H, Sukumaran J, Smith SA. **2018**. A matter of phylogenetic scale: Distinguishing incomplete lineage sorting from lateral gene transfer as the cause of gene tree discord in recent versus deep diversification histories. *American Journal of Botany* **105**: 376–384.

Laenen B, Shaw B, Schneider H, *et al.* 2014. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. *Nature Communications* **5**: 5134.

Lanfear R, Minh B, Hahn M. **2018**. *Calculating and interpreting gene- and site-concordance factors in phylogenomics*. http://www.robertlanfear.com/blog/files/concordance_factors.html. 16 May 2022.

Leebens-Mack J, Barker M, Carpenter E, *et al.* 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.

Li F-W, Villarreal JC, Szövényi P. **2017**. Hornworts: an overlooked window into carbon-concentrating mechanisms. *Trends in Plant Science* **22**: 275–277.

Liu X-Q, Xia X-M, Chen L, Wang X-Q. **2022**. Phylogeny and evolution of Cupressaceae: updates on intergeneric relationships and new insights on ancient intergeneric hybridization. *Molecular Phylogenetics and Evolution* **177**: 107606.

Lutzoni F, Nowak MD, Alfaro ME, *et al.* 2018. Contemporaneous radiations of fungi and plants linked to symbiosis. *Nature Communications* **9**: 5451.

Meleshko O, Martin MD, Korneliussen TS, *et al.* 2021. Extensive genome-wide phylogenetic discordance is due to incomplete lineage sorting and not ongoing introgression in a rapidly radiated bryophyte genus (J de Meaux, Ed.). *Molecular Biology and Evolution* **38**: 2750–2766.

Meyer M, Goudet MMM, Griffiths H. **2020**. The algal pyrenoid In: Larkum A, Grossmann A, Raven J, eds. *Photosynthesis in Algae: Biochemical and Physiological Mechanisms Volume 45*. Switzerland AG: Springer Nature, 179–203.

Meyer M, Whittaker C, Griffiths H. **2017**. The algal pyrenoid: key unanswered questions. *Journal of Experimental Botany* **68**: 3739–3749.

Minh BQ, Hahn MW, Lanfear R. **2020**. New methods to calculate concordance factors for phylogenomic datasets (M Rosenberg, Ed.). *Molecular Biology and Evolution* **37**: 2727–2733.

Minh B, Trifinopoulos J, Schrempf D, Schmidt HA. **2018**. IQ-TREE version 2.2.0: Tutorials and manual phylogenomic software by maximum likelihood. : 1–156.

820 **Minh BQ, Trifinopoulos J, Schrempf D, Schmidt HA**. **2019**. IQ-TREE version 2.0: Tutorials and
821 manual phylogenomic software by maximum likelihood. *Manual/tutorial*.

822 **Moisan P, Voigt S, Schneider JW, Kerp H**. **2012**. New fossil bryophytes from the Triassic
823 Madygen Lagerstätte (SW Kyrgyzstan). *Review of Palaeobotany and Palynology* **187**: 29–37.

824 **Morris JL, Puttick MN, Clark JW, *et al.* 2018**. The timescale of early land plant evolution.
825 *Proceedings of the National Academy of Sciences* **115**: E2274–E2283.

826 **Mukherjee A**. **2020**. CO2 concentration in *Chlamydomonas reinhardtii:* effect of the pyrenoid
827 starch sheath. *Plant Physiology* **182**: 1796–1797.

828 **Nagalingum NS, Marshall CR, Quental TB, Rai HS, Little DP, Mathews S**. **2011**. Recent
829 synchronous radiation of a living fossil. *Science* **334**: 796–799.

830 **Natcheva R, Cronberg N**. **2004**. What do we know about hybridization among bryophytes in
831 nature? *Canadian Journal of Botany* **82**: 1687–1704.

832 **Natcheva R, Cronberg N**. **2007**. Maternal transmission of cytoplasmic DNA in interspecific
833 hybrids of peat mosses, *Sphagnum* (Bryophyta). *Journal of Evolutionary Biology* **20**: 1613–1616.

834 **Nguyen L-T, Schmidt H, von Haeseler A, Minh BQ**. **2015**. IQ-TREE: A fast and effective
835 stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and*
836 *Evolution* **32**: 268–274.

837 **Pardo-De la Hoz CJ, Magain N, Piatkowski B, *et al.* 2023**. Ancient rapid radiation explains most
838 conflicts among gene trees and well-supported phylogenomic trees of Nostocalean Cyanobacteria
839 (C Lemus-Solis, Ed.). *Systematic Biology* **72**: 694–712.

840 **Peñaloza-Bojacá GF, de Oliveira MF, Maciel-Silva AS**. **2020**. *Anthoceros tristanianus* J.C.
841 Villarreal, J.J. Engel & Váňa in Váňa & Engel (Anthocerotaceae, Anthocerotophyta) a new record
842 for South America from the Colombian bryophyte flora. *Cryptogamie, Bryologie* **41**: 105.

Pérez-Escobar OA, Dodsworth S, Bogarín D, *et al.* 2021. Hundreds of nuclear and plastid loci yield novel insights into orchid relationships. *American Journal of Botany* **108**: ajb2.1702.

Proskauer J. **1957**. Studies on Anthocerotales V. *Phytomorphology* **7**: 113–135.

Proskauer J. **1969**. Studies on Anthocerotales. VIII. *Phytomorphology* **19**: 52–66.

R Core Team R. **2020**. *R: A language and environment for statistical computing*. https://www.r-project.org/. 4 May 2024.

Rambaut A. **2017**. *FigTree-version 1.4. 3, a graphical viewer of phylogenetic trees.* http://tree.bio.ed.ac.uk/software/figtree/. 20 Aug. 2023.

Ramírez-Barahona S, Sauquet H, Magallón S. **2020**. The delayed and geographically heterogeneous diversification of flowering plant families. *Nature Ecology & Evolution* **4**: 1232–1238.

Reddy S, Kimball RT, Pandey A, *et al.* 2017. Why do phylogenomic data sets yield conflicting trees? Data type influences the avian tree of life more than taxon sampling. *Systematic Biology* **66**: 857–879.

Renne P, Deino A, Hilgen F, *et al.* 2013. Time scales of critical events around the cretaceous-paleogene boundary. *Science* **339**: 684–687.

Renzaglia K, Schuette S, Duff J, *et al.* 2007. Bryophyte phylogeny: Advancing the molecular and morphological frontiers. *The Bryologist* **110**: 179–213.

Renzaglia K, Villarreal JC, Duff RJ, Goffinet B. **2009**. New insights into morphology, anatomy, and systematics of hornworts In: Shaw AJ, ed. *Bryophyte Biology*. Cambridge: Cambridge University Press, 139–172.

Robison TA, Oh ZG, Lafferty D, *et al.* 2024. Hornworts reveal a spatial model for pyrenoid-based CO2-concentrating mechanisms in land plants. *bioRxiv*: 1–20.

Ruaud S, Nötzold SI, Waller M, *et al.* 2024. Molecular underpinnings of hornwort carbon concentrating mechanisms: subcellular localization of putative key molecular components in the model hornwort Anthoceros agrestis. *bioRxiv*: 1–36.

Sayyari E, Mirarab S. **2016**. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33**: 1654–1668.

Schöneberg Y, Winter S, Arribas O, *et al.* 2023. Genomics reveals broad hybridization in deeply divergent Palearctic grass and water snakes (Natrix spp.). *Molecular Phylogenetics and Evolution* **184**.

Schuette S, Renzaglia K. **2010**. Development of multicellular spores in the hornwort genus Dendroceros (Dendrocerotaceae, Anthocerotophyta) and the occurrence of endospory in Bryophytes. *Nova Hedwigia* **91**: 301–316.

Schuettpelz E, Pryer KM. **2009**. Evidence for a Cenozoic radiation of ferns in an angiosperm-dominated canopy. *Proceedings of the National Academy of Sciences* **106**: 11200–11205.

Shen X-X, Steenwyk JL, Rokas A. **2021**. Dissecting incongruence between concatenation- and quartet-based approaches in phylogenomic data (J Brown, Ed.). *Systematic Biology* **0**: 1–18.

Shimodaira H. **2002**. An approximately unbiased test of phylogenetic tree selection (N Goldman, Ed.). *Systematic Biology* **51**: 492–508.

Shimodaira H, Hasegawa M. **1999**. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Molecular Biology and Evolution* **16**: 1114–1116.

Smith S, O'Meara B. **2012**. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics* **28**: 2689–2690.

Söderström L, Hagborg A, von Konrat M, *et al.* 2016. World checklist of hornworts and liverworts. *PhytoKeys* **59**: 1–828.

889 **Solís-Lemus C, Ané C**. **2016**. Inferring phylogenetic networks with maximum pseudolikelihood
890 under incomplete lineage sorting (S Edwards, Ed.). *PLOS Genetics* **12**: e1005896.

891 **Solís-Lemus C, Bastide P, Ané C**. **2017**. PhyloNetworks: A package for phylogenetic networks.
892 *Molecular Biology and Evolution* **34**: 3292–3298.

893 **Soltis PS**. **2013**. Hybridization, speciation and novelty. *Journal of Evolutionary Biology* **26**: 291–
894 293.

895 **Song S, Liu L, Edwards S V., Wu S**. **2012**. Resolving conflict in eutherian mammal phylogeny
896 using phylogenomics and the multispecies coalescent model. *Proceedings of the National Academy*
897 *of Sciences* **109**: 14942–14947.

898 **Springer MS, Murphy WJ, Eizirik E, O'Brien SJ**. **2003**. Placental mammal diversification and
899 the Cretaceous–Tertiary boundary. *Proceedings of the National Academy of Sciences* **100**: 1056–
900 1061.

901 **Stamatakis A**. **2014**. RAxML v8.2.X Manual. *Manual/tutorial*: 1–5.

902 **Steinthorsdottir M, Vajda V**. **2015**. Early Jurassic (late Pliensbachian) CO2 concentrations based
903 on stomatal analysis of fossil conifer leaves from eastern Australia. *Gondwana Research* **27**: 932–
904 939.

905 **Strimmer K, Rambaut A**. **2002**. Inferring confidence sets of possibly misspecified gene trees.
906 *Proceedings of the Royal Society of London. Series B: Biological Sciences* **269**: 137–142.

907 **Stull GW, Pham KK, Soltis PS, Soltis DE**. **2023**. Deep reticulation: the long legacy of
908 hybridization in vascular plant evolution. *The Plant Journal* **114**: 743–766.

909 **Suarez-Gonzalez A, Lexer C, Cronk QCB**. **2018**. Adaptive introgression: a plant perspective.
910 *Biology Letters* **14**: 20170688.

911 **Sun M, Folk R, Gitzendanner M, *et al.* 2020**. Recent accelerated diversification in rosids occurred

912 outside the tropics. *Nature Communications* **11**: 3333.

913 **Suvorov A, Scornavacca C, Fujimoto MS, *et al.* 2022**. Deep ancestral introgression shapes
914 evolutionary history of Dragonflies and Damselflies. *Systematic Biology* **71**: 526–546.

915 **Taylor SA, Larson EL**. **2019**. Insights from genomes into the evolutionary importance and
916 prevalence of hybridization in nature. *Nature Ecology & Evolution* **3**: 170–177.

917 **Uniacke J, Zerges W**. **2009**. Chloroplast protein targeting involves localized translation in
918 *Chlamydomonas*. *Proceedings of the National Academy of Sciences of the United States of America*
919 **106**: 1439–1444.

920 **Vajda V, Bercovici A**. **2014**. The global vegetation pattern across the Cretaceous–Paleogene mass
921 extinction interval: A template for other extinction events. *Global and Planetary Change* **122**: 29–
922 49.

923 **Vanderpool D, Minh BQ, Lanfear R, *et al.* 2020**. Primate phylogenomics uncovers multiple rapid
924 radiations and ancient interspecific introgression (CD Jiggins, Ed.). *PLOS Biology* **18**: e3000954.

925 **Vaughn K, Ligrone R, Owen H, *et al.* 1992**. The anthocerote chloroplast: a review. *New*
926 *Phytologist* **120**: 169–190.

927 **Villarreal JC, Crandall-Stotler B, Hart M, Long D, Forrest L**. **2016**. Divergence times and the
928 evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow
929 molecular rate. *New Phytologist* **209**: 1734–1746.

930 **Villarreal JC, Cusimano N, Renner S**. **2015**. Biogeography and diversification rates in hornworts:
931 The limitations of diversification modeling. *Taxon* **64**: 229–238.

932 **Villarreal JC, Frey W, Cargill C**. **2010**. *Bryophyte Biology, Phylogeography, Systematics and*
933 *Evolution in the Southern Hemisphere. A Volume in Honour of Dr. Gabriela Hässel de Menéndez†*.
934 Cramer in der Gebr. Borntraeger Verlagsbuchhandlung, Stuttgart: Nova Hedwigia.

Villarreal JC, Renner S. **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* **109**: 18873–18878.

Villarreal JC, Renner S. **2013**. Correlates of monoicy and dioicy in hornworts, the apparent sister group to vascular plants. *BMC Evolutionary Biology* **13**: 1–8.

Villarreal JC, Renzaglia K. **2006**. Structure and development of *Nostoc* strands in *Leiosporoceros dussii* (Anthocerotophyta): a novel symbiosis in land plants. *American Journal of Botany* **93**: 693–705.

Villarreal JC, Renzaglia K. **2015**. The hornworts: important advancements in early land plant evolution. *Journal of Bryology* **37**: 157–170.

Villarreal JC, Turmel M, Bourgouin-Couture M, *et al.* 2018. Genome-wide organellar analyses from the hornwort *Leiosporoceros dussii* show low frequency of RNA editing (F Martinez-Abarca, Ed.). *PLoS ONE* **13**: e0200491.

Xi Z, Liu L, Rest JS, Davis CC. **2014**. Coalescent versus concatenation methods and the placement of *Amborella* as sister to water lilies. *Systematic Biology* **63**: 919–932.

Zhan Y, Marchand CH, Maes A, *et al.* 2018. Pyrenoid functions revealed by proteomics in *Chlamydomonas reinhardtii*. *PLoS ONE* **13**: 1–20.

Zhang C, Rabiee M, Sayyari E, Mirarab S. **2018**. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**: 153.

958 TABLES

959 **Table 1.** Main hornwort topologies. T1: *Leiosporoceros dussii* (Steph.) Hässel sister to the other
960 families of hornworts (ML concatenation method); T2: *Leiosporoceros dussii* is sister group of
961 Anthocerotaceae (ASTRAL coalescent method); T3: Anthocerotaceae sister to the clade of
962 *Leiosporoceros dussii* and other families of hornworts (alternative topology in coalescence and
963 concatenation method). The T3 topology does not have support at the *L. dussii* node. The full
964 dataset includes the 234 genes. In contrast, *Leiosporoceros* lacks 39 genes (P. Schafran pers.
965 comm.) and we created the Only_leios dataset and performed similar analyses.

966

Data set	**Number of genes**	**Genes supporting each topology**		
		T1	**T2**	**T3**
Full_data	234 (100%)	90 (38,5%)	78 (33,4%)	62 (26,5%)
Full_data consistent genes	161 (100%)	61 (37,8%)	63 (39,1%)	36 (22,4%)
Only_leios	195 (100%)	79 (40,5%)	63 (32,3%)	52 (26,7%)
Only_leios consistent genes	133 (100%)	51 (38,4%)	55 (41,3%)	27 (20,3%)

967

968

969

970

971

972

973

974 **Table 2.** Tree topology tests (T1: *Leiosporoceros dussii* is sister to the other families of hornworts;
975 T2: *Leiosporoceros dussii* is sister group of Anthocerotaceae; T3: Anthocerotaceae sister to the
976 clade of *Leiosporoceros dussii* and other families of hornworts. DeltaL: logL difference from the
977 maximal logl in the set; bp-RELL: bootstrap proportion using RELL method; p-KH: p-value of one-
978 sided Kishino-Hasegawa test; p-SH: p-value of Shimodaira-Hasegawa test; c-ELW: Expected
979 Likelihood Weight; p-AU: p-value of approximately unbiased (AU) test). Plus, signs denote the
980 95% confidence sets and minus signs denote significant exclusion. All tests performed 10000
981 resampling using the RELL method in IQ-Tree.

982

	Data set	Topology	logL	deltaL	bp-RELL		p-KH		p-SH		c-ELW		p-AU	
Full_data		T1	-1202258,58	0	1	+	1	+	1	+	1	+	1	+
Full_data		T2	-1202466,911	208,33	0	-	0,0001	-	0,0001	-	1,68e-07	-	1,46e-05	-
Full_data		T3	-1202511,488	252,91	0	-	0	-	0	-	1,19e-37	-	4,81e-64	-
Full_data	Consistent genes	T1	-848377,7561	0	0,998	+	0,997	+	1	+	0,998	+	0,997	+
Full_data	Consistent genes	T2	-848511,2003	133,44	0,002	-	0,0027	-	0,0031	-	0,00213	-	0,00308	-
Full_data	Consistent genes	T3	-848604,8049	227,05	0	-	0	-	0	-	1,16e-25	-	5,78e-48	-
Only_Leios		T1	-1015587,001	0	0,953	+	0,947	+	1	+	0,952	+	0,96	+
Only_Leios		T2	-1015632,14	45,14	0,047	-	0,053	+	0,063	+	0,0479	-	0,0429	-
Only_Leios		T3	-1015673,389	86,39	0	-	0	-	0	-	3,43e-09	-	6,79e-13	-
Only_Leios	Consistent genes	T1	-743956,0191	0	0,999	+	0,999	+	1	+	0,999	+	0,999	+
Only_Leios	Consistent genes	T2	-744114,4576	158,44	0,0006	-	0,0006	-	0,0006	-	0,000635	-	0,000994	-
Only_Leios	Consistent genes	T3	-744189,0778	233,06	0	-	0	-	0	-	5,51e-33	-	0,000309	-

983

984

985 FIGURES
986

987 **Fig. 1:** Divergence time estimates of hornworts genera based on an ASTRAL analysis of 234
988 nuclear genes. The nodes of interest (highlighted by major clade color) have been labeled (A-N) to
989 illustrate quartet values from gene concordance factor (gcf) values (left), site concordance factor
990 (sCF) values (center). gCF and sCF represent the percentage of decisive genes and sites at each
991 branch, respectively. Pies for quartet values from concordance factor values: CF (topology shown)
992 and alternative options (DF1, DF2, DFP) and site concordance factor (sCF, sDF1, sDF2); see Figure
993 S9, Table S4 for the concordance factors for all nodes. Quartet values from ASTRAL analyses
994 presented as pie-charts (right). ASTRAL pies are divided into q1 or topology shown (purple), q2
995 (blue, 2nd alternative hypothesis) and q3 (orange, 3rd alternative hypothesis) with the percentage for
996 q1 included in the pie diagram; see Appendix 5, figures S21, S22 for quartet values for all other
997 nodes and local posterior probabilities. The detailed chronogram with node heights represents mean
998 ages and bars the 95% highest posterior density intervals reported in Appendix 5, figure S19.
999 Numbers represent the calibrations (Material and Methods): 1: Mosses divergence, 2: Liverworts
1000 divergence, 3: *Ricciopsis ferganica* fossil, 4: *Anthoceros* spore type A; 5: *Notothylites nirulai* fossil;
1001 and 6: fossil assigned to *Phaeomegaceros* sp. (Inset). Average values of atmospheric CO2 (ppm)
1002 levels during the last 500 Ma (Berner 2001; Badger *et al.* 2002; Renne *et al.* 2013; Steinthorsdottir
1003 and Vajda 2015).

1004

1005 **Fig. 2:** Estimates of ancient reticulate evolution (SNaQ) from all hornwort genera using 234 genes
1006 and number of hybridizations (hmax = 3). A) estimated network based on T1 topology
1007 (*Leiosporoceros* sister to all hornworts) with best scores across hybrid values -Ploglik = 255.42. B)
1008 estimated network based on T2 topology (*Leiosporoceros* sister to Anthocerotaceae) with best
1009 scores across hybrid values -Ploglik =254.74. On the left corner, we present the recovered topology
1010 and to the right the correlation graphic with the observed concordance factors versus the expected
1011 concordance factors. Light blue quartets include *Leiosporoceros*. C) Best estimated network with
1012 possible reticulation events based on T1 and T2 topologies. The red and blue lines indicate hybrid
1013 edges. The red and blue numbers indicate estimated inheritance probabilities from major and minor
1014 parental species, respectively. See Supplementary Data figures S14-S15 and Table S5.

1015

1016 **Fig. 3**: Hornwort habit and plastid diversity. A) Gametophyte and sporophyte of *Phaeoceros*
1017 *minutus* (Mitt.) S.W. Arnell (South Africa); B) Chloroplast without pyrenoid in *Nothoceros*

1018 *minarum* (Nees) J.C. Villarreal (Brazil); C) Chloroplast with pyrenoid in *Dendroceros crispatus*
1019 (Hook.) Nees (Colombia); D) TEM image of the pyrenoid of *Phaeoceros carolinianus* (Michx.)
1020 Prosk. (USA). Scales A: 2 mm; B-D: 10 μm. Credits: A) Des Callaghan; B-C: GFPB; and D: KR.

1021

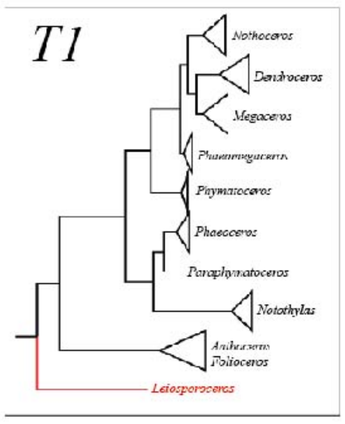
gDFP gCF gDF2 gDF1

sDF2 sCF sDF1

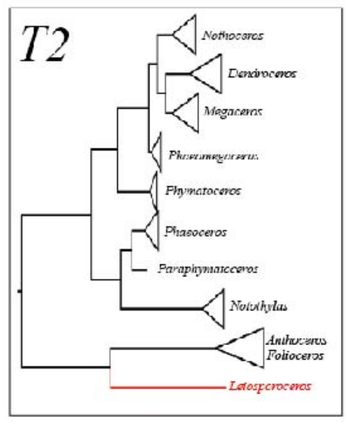
q3 q1 q2

A B C D E F G H I J K L M N

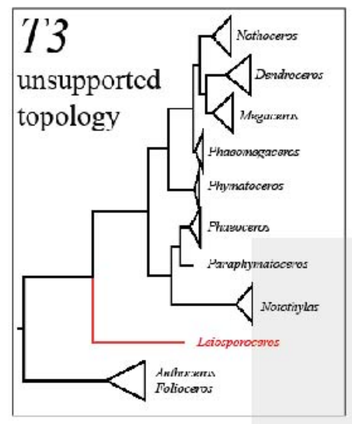
T1

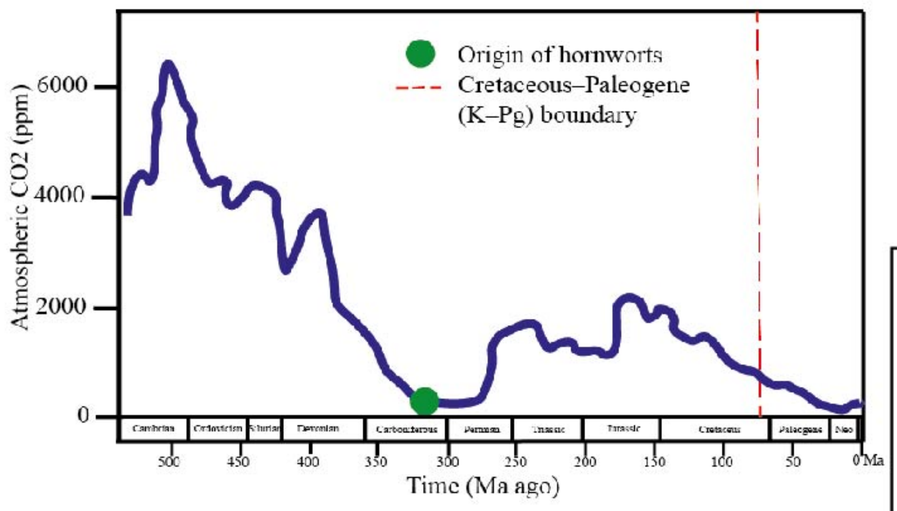


T2

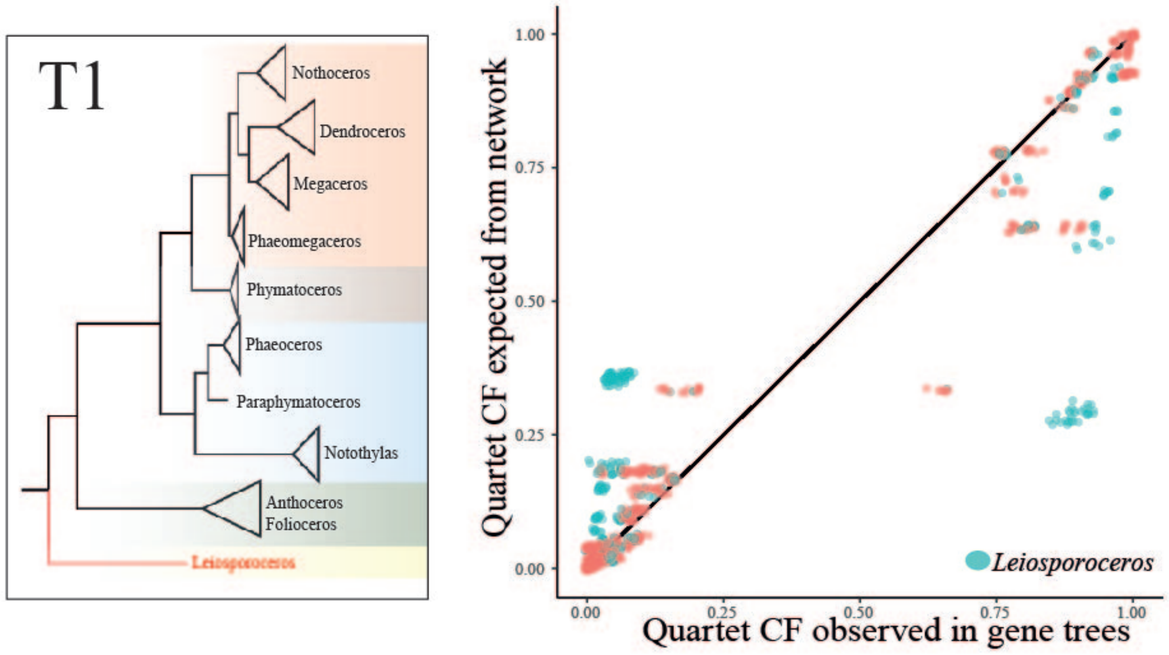


T3
unsupported topology



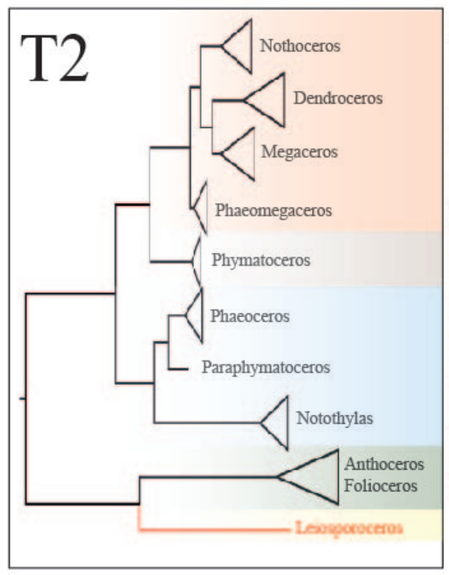


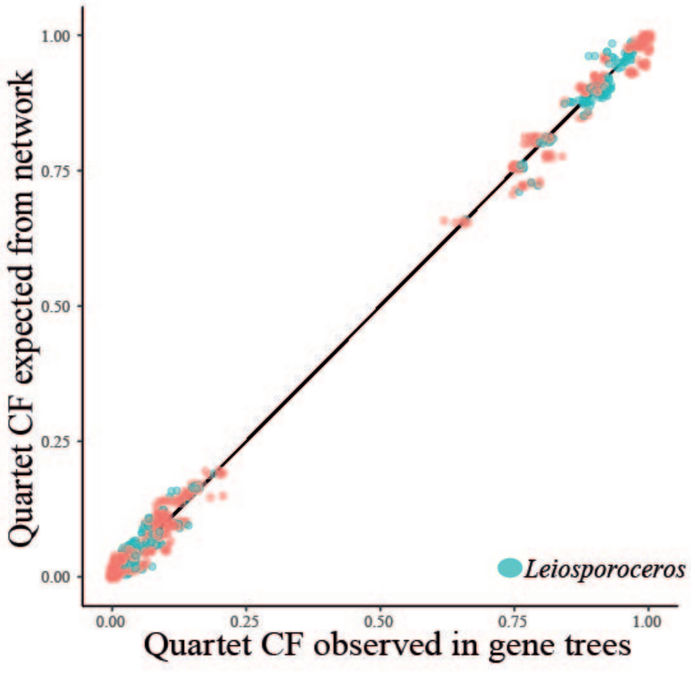
A


T1
Nothoceros
Dendroceros
Megaceros
Phaeomegaceros
Phymatoceros
Phaeoceros
Paraphymatoceros
Notothylas
Anthoceros
Folioceros
Leiosporoceros

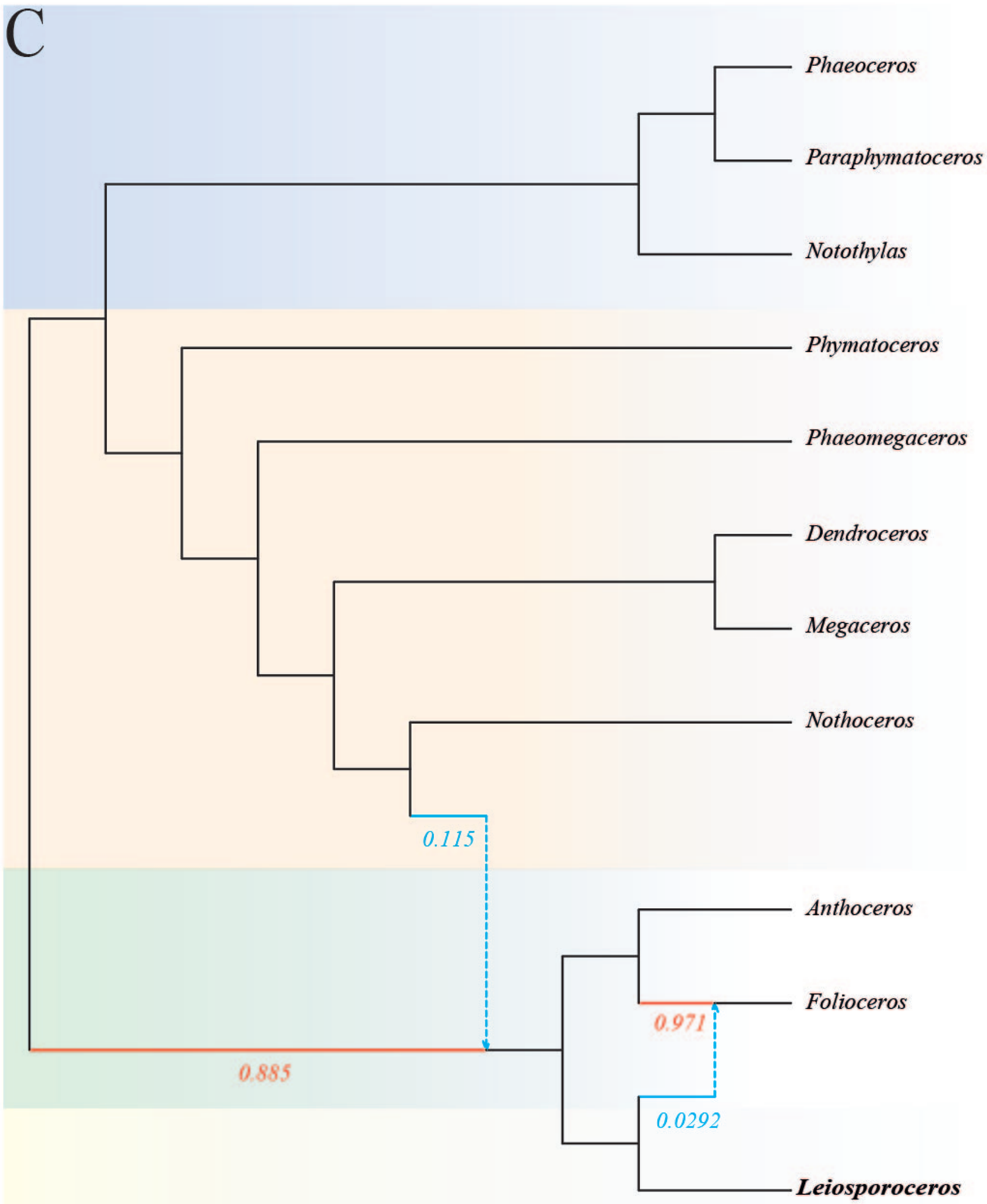
Quartet CF expected from network
Quartet CF observed in gene trees
0.00
0.25
0.50
0.75
1.00
Leiosporoceros

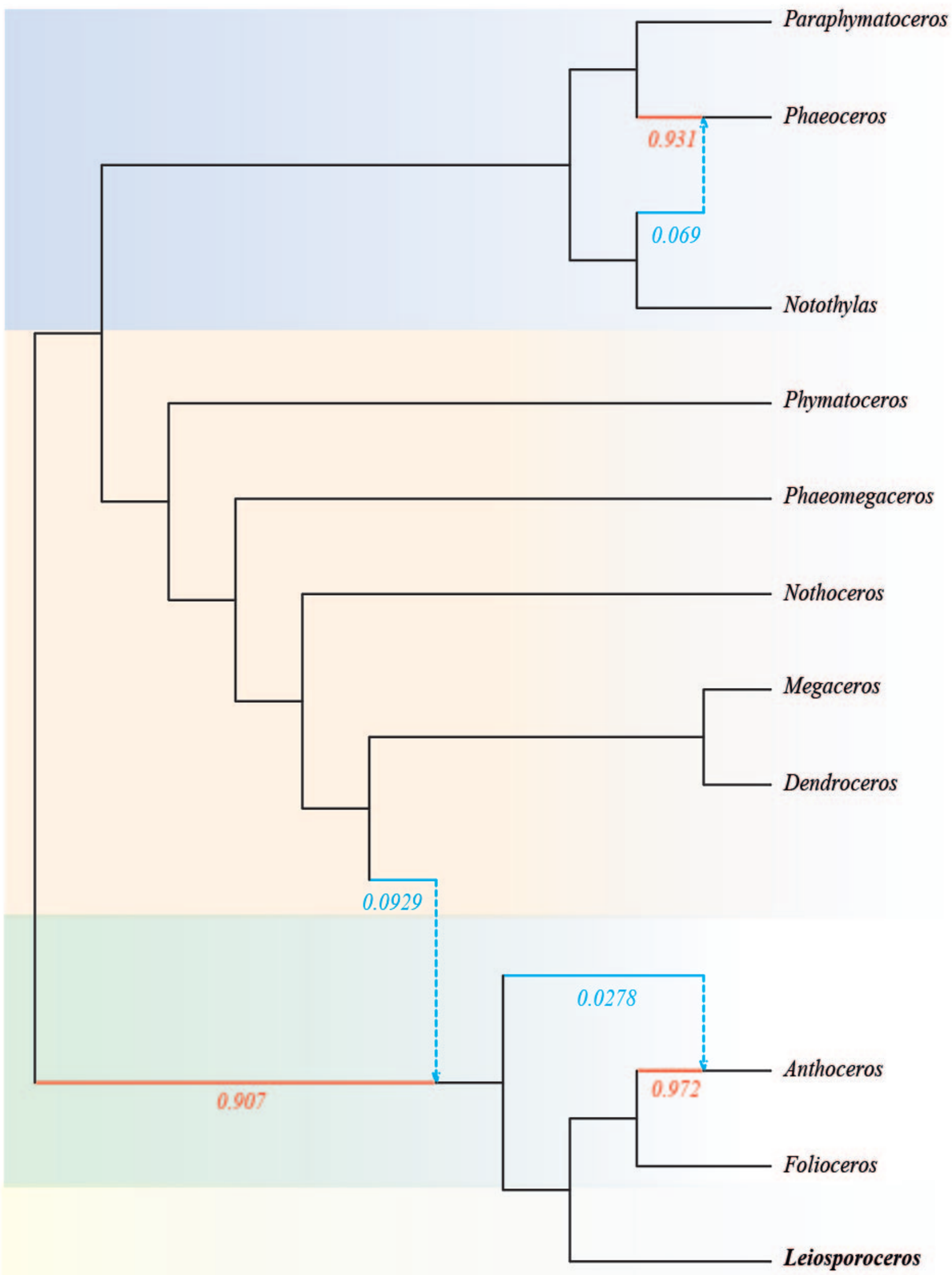
B


T2
Nothoceros
Dendroceros
Megaceros
Phaeomegaceros
Phymatoceros
Phaeoceros
Paraphymatoceros
Notothylas
Anthoceros
Folioceros
Leiosporoceros

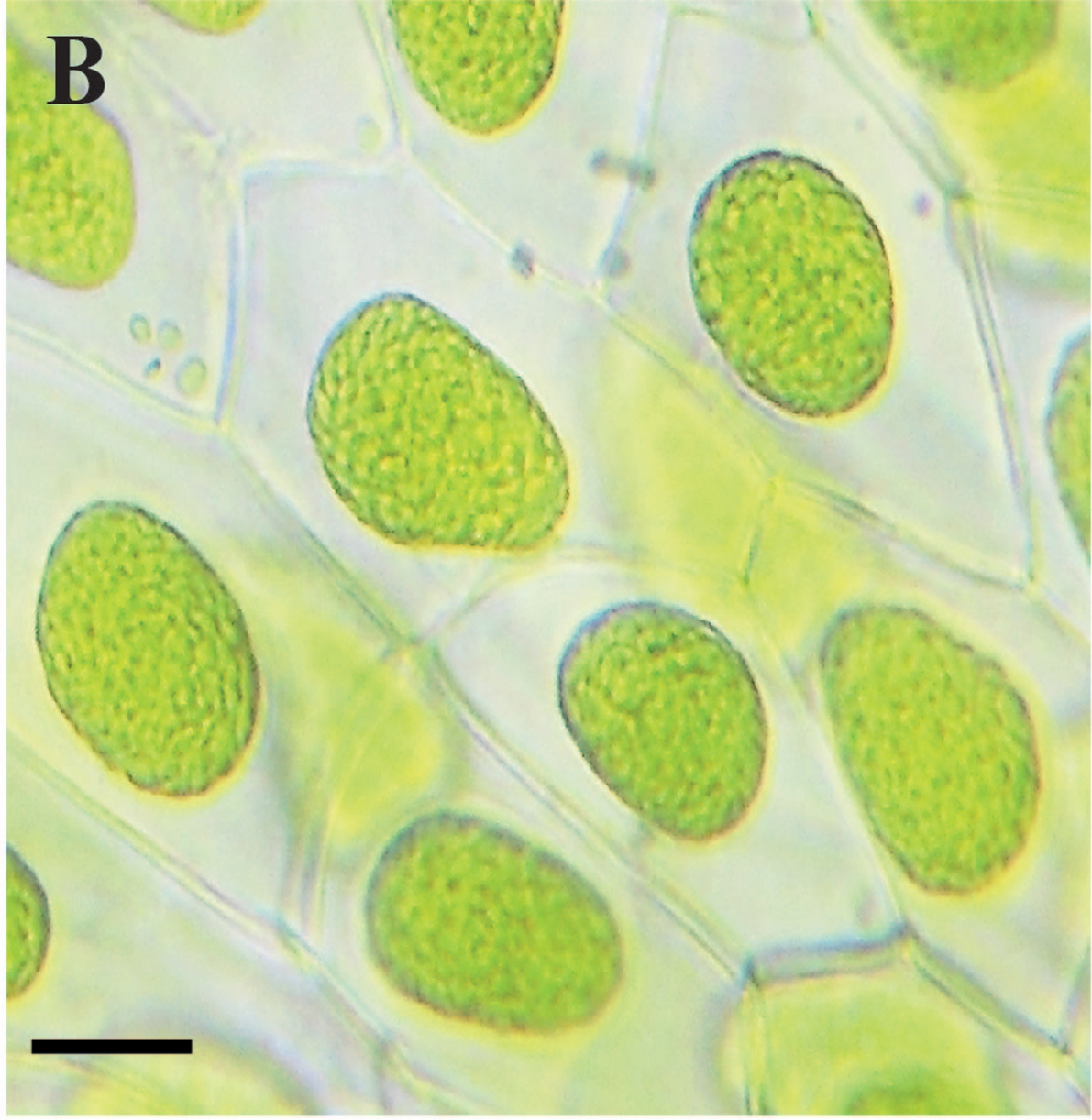

Quartet CF expected from network
Quartet CF observed in gene trees
0.00
0.25
0.50
0.75
1.00
Leiosporoceros

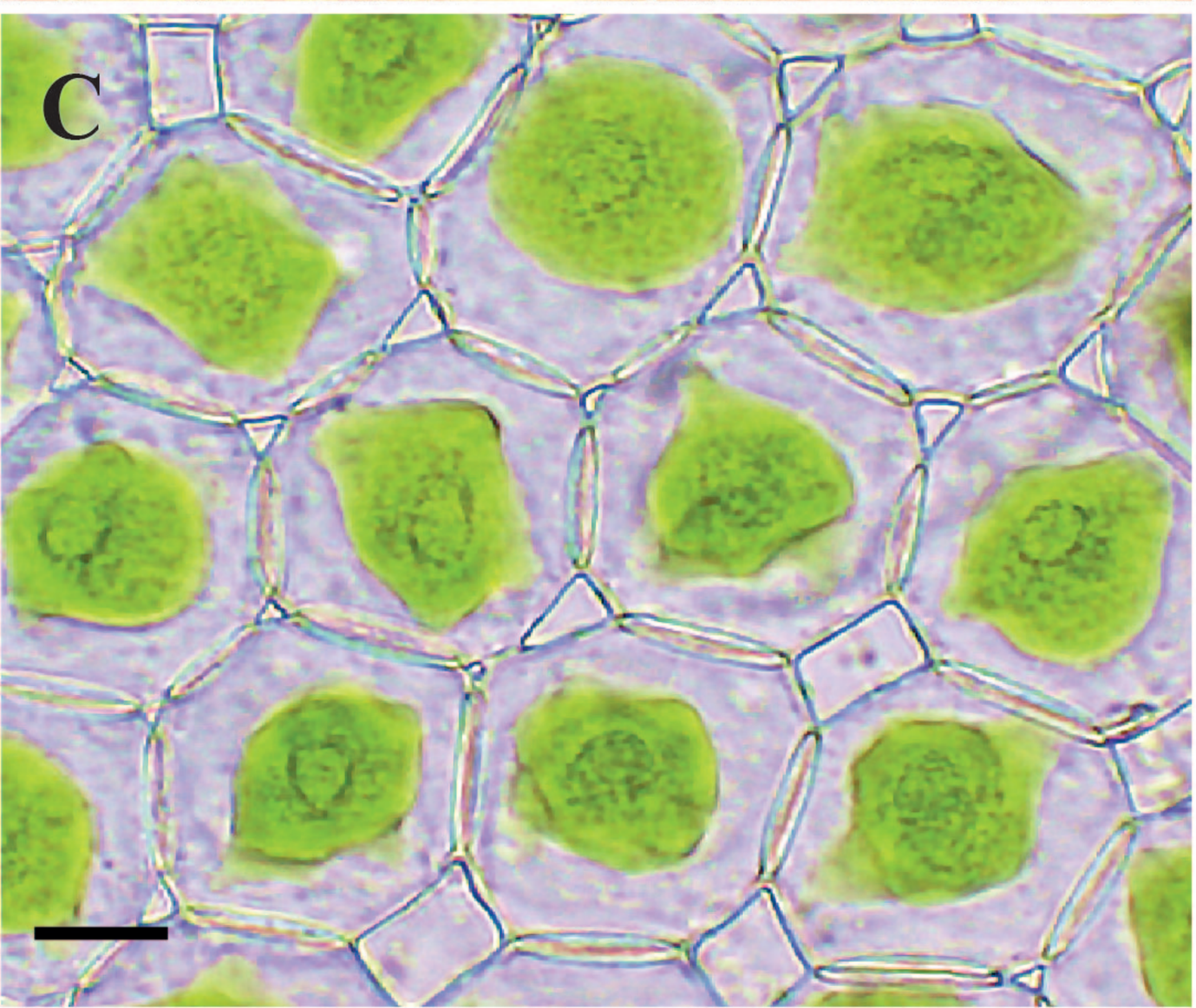
C

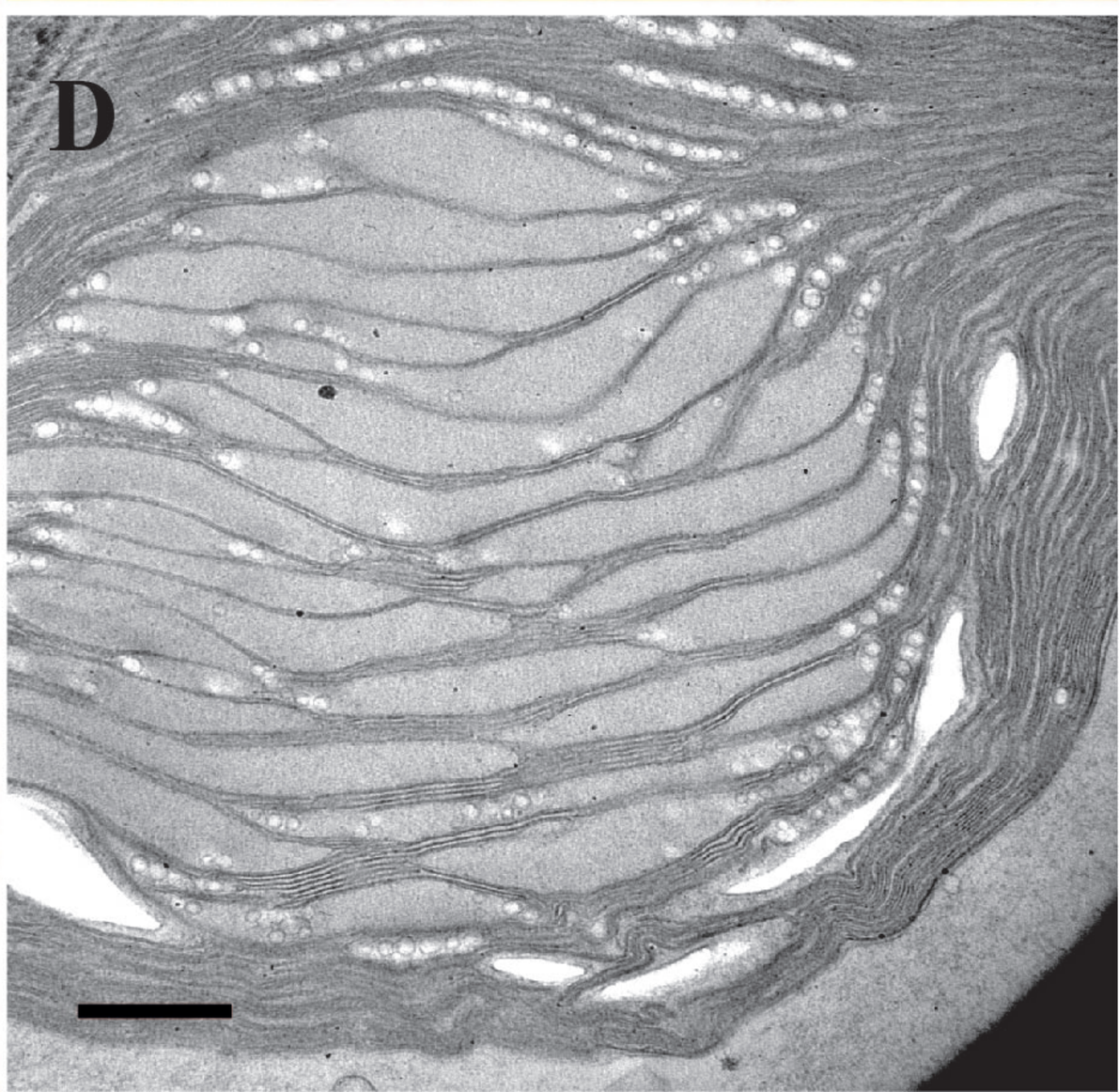

Phaeoceros
Paraphymatoceros
Notothylas
Phymatoceros
Phaeomegaceros
Dendroceros
Megaceros
Nothoceros
0.115
Anthoceros
Folioceros
0.971
0.885
0.0292
Leiosporoceros


Paraphymatoceros
Phaeoceros
0.931
0.069
Notothylas
Phymatoceros
Phaeomegaceros
Nothoceros
Megaceros
Dendroceros
0.0929
0.0278
Anthoceros
0.972
0.907
Folioceros
Leiosporoceros


A


B


C


D