


Phylotranscriptomics of liverworts: revisiting the backbone phylogeny and ancestral gene duplications

Shanshan Dong¹, Jin Yu^{1,2}, Li Zhang¹, Bernard Goffinet³ and Yang Liu^{1,*} 

¹Laboratory of Southern Subtropical Plant Diversity, Fairy Lake Botanical Garden, Shenzhen & Chinese Academy of Sciences, Shenzhen 518004, Guangdong, China, ²State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen 518083, Guangdong, China and ³Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269-3043, USA

*For correspondence. E-mail yang.liu0508@gmail.com

Received: 24 May 2022 Returned for revision: 26 August 2022 Editorial decision: 6 September 2022 Accepted: 8 September 2022
Electronically published: 8 September 2022

- **Background and Aims** With some 7300 extant species, liverworts (Marchantiophyta) represent one of the major land plant lineages. The backbone relationships, such as the phylogenetic position of Ptilidiales, and the occurrence and timing of whole-genome duplications, are still contentious.
- **Methods** Based on analyses of the newly generated transcriptome data for 38 liverworts and complemented with those publicly available, we reconstructed the evolutionary history of liverworts and inferred gene duplication events along the 55 taxon liverwort species tree.
- **Key Results** Our phylogenomic study provided an ordinal-level liverwort nuclear phylogeny and identified extensive gene tree conflicts and cyto-nuclear incongruences. Gene duplication analyses based on integrated phylogenomics and *Ks* distributions indicated no evidence of whole-genome duplication events along the backbone phylogeny of liverworts.
- **Conclusions** With a broadened sampling of liverwort transcriptomes, we re-evaluated the backbone phylogeny of liverworts, and provided evidence for ancient hybridizations followed by incomplete lineage sorting that shaped the deep evolutionary history of liverworts. The lack of whole-genome duplication during the deep evolution of liverworts indicates that liverworts might represent one of the few major embryophyte lineages whose evolution was not driven by whole-genome duplications.

Key words: Liverworts, Marchantiophyta, evolution, transcriptome, phylogeny, gene duplication.

INTRODUCTION

With about 7300 species recognized in some 386 genera and 87 families (Söderström *et al.*, 2016), liverworts represent one of the most speciose major lineages of land plants. The phylogenetic position of liverworts has been contentious for decades, with various hypotheses proposed, i.e. liverworts sister to the rest of land plants, sister to mosses, sister to tracheophytes or sister to a clade including mosses and tracheophytes (as reviewed in Puttick *et al.*, 2018 and Goffinet, 2000). High-throughput sequencing technology has raised phylogenomic approaches to new heights. Recent nuclear phylogenomic analyses have converged on a sister relationship of liverworts and mosses, which, along with hornworts, form the bryophyte clade that is sister to extant tracheophytes (Puttick *et al.*, 2018; One Thousand Plant Transcriptomes Initiative, 2019; Harris *et al.*, 2020; Su *et al.*, 2021). Liverworts exhibit morphological heterogeneity, with three major morphotypes (i.e. complex thalloid, simple thalloid and leafy), a broad ecological distribution with terrestrial, aquatic, epiphytic and epiphyllous ecotypes, and various types of fungal associations (Read *et al.*, 2000), and are characterized most notably by an ancestral loss of stomata (Harris *et al.*, 2020).

A sufficiently resolved phylogenetic tree would provide a basis for studying character evolution within this group. However, phylogenomic reconstructions based on a large set of nuclear sequences with a broad sampling have not been carried out focusing on this group, and the backbone relationships of liverworts are still subject to controversy. The following relationships, in particular, remain ambiguous: relationships of Ptilidiales to Jungermanniales and Porellales (Liu *et al.*, 2008; Yu *et al.*, 2019), of *Pellia* to the rest of the Pelliidae (Forrest *et al.*, 2006) or Jungermanniopsida (Crandall-Stotler *et al.*, 2005; Haberle *et al.*, 2008), among the complex thalloids, i.e. Marchantiales (Forrest *et al.*, 2006; Villarreal *et al.*, 2016; Flores *et al.*, 2017, 2020), and also the deep relationships within the two most speciose orders, Jungermanniales and Porellales (Heinrichs *et al.*, 2007; Yu *et al.*, 2019).

Recent phylogenomic studies of liverworts based on plastid and mitochondrial genomes are not conclusive, and the aforementioned systematic ambiguities persisted. Plastid phylogenomic analyses (Yu *et al.*, 2019) based on 35 genera are largely congruent with previous analyses based on discrete genes (Forrest *et al.*, 2006), but incongruent with the recent plastid phylogenomic analyses by Dong *et al.* (2021b), with regard to the placements of Ptilidiales. The order is weakly

supported as a sister group to Jungermanniales in Yu *et al.* (2019) but strongly supported as the sister to Porellales in Dong *et al.* (2021b). Based on the mitochondrial genes, Dong *et al.* (2021a) recovered a moderately supported hypothesis whereby the Ptilidiales are sister to Jungermanniales. These apparent phylogenetic conflicts between mitochondrial and plastid data necessitate further nuclear phylogenomic studies incorporating more genes and involving extended taxon sampling.

With the development of high-throughput sequencing methods, several techniques have been utilized for generating genome-scale data in plant phylogenetic studies, such as transcriptome sequencing (e.g. One Thousand Plant Transcriptomes, 2019), hybrid capture (e.g. Liu *et al.*, 2019) and genome skimming methods (e.g. Dong *et al.*, 2022). The transcriptome is the readout of a genome, and transcriptome sequencing is an efficient and cost-effective way for recovering thousands of genes. Sequencing and analysing transcriptomes do not rely on genomic references, which is, however, a necessity when using a hybrid capture method (Lemmon *et al.*, 2012). The size of liverwort genomes ranges widely from approx. 200 Mb in *Marchantia* to approx. 20 Gb in *Phyllohallia* (<https://cvalues.science.kew.org>), which makes the genome skimming method not applicable when sampling liverworts at a broad phylogenetic spectrum. In addition, sequencing transcriptomes recovers unselected, and richer genomic information than that of the hybrid capture methods, which can be further mined for the analyses of gene family evolution. The phylotranscriptomic method has been widely used to reconstruct the relationships and explore genomic features in plant groups at both high and low taxonomic ranks (One Thousand Plant Transcriptomes Initiative, 2019). Sampling the vegetative transcriptome of ordinal exemplars may similarly yield a reliable phylogeny for liverworts and provide a robust basis for inferring the evolution of morphological traits, as well as of genomic characters such as whole-genome duplication (WGD) events.

Whole-genome duplications occurred throughout the eukaryote tree of life, especially in the plant kingdom, and are considered a driving force of species diversification and evolutionary innovations (Van de Peer *et al.*, 2017; Ren *et al.*, 2018; Wu *et al.*, 2019). A WGD event leads to genetic redundancies and is usually followed by lineage-specific loss of the duplicated genes, but may also provide opportunities for sub-/neo-functionalizations (Birchler and Yang, 2022). WGD contributes to adaptation to new niches, survival of severe environmental stress and subsequent rapid accumulations of species diversity (Schrantz *et al.*, 2012; Ren *et al.*, 2018). Liverworts represent a highly diversified lineage, and the number of liverwort species exceeds that of hornworts, lycophytes and gymnosperms. However, few signatures of ancient WGDs have been observed in liverworts (Bowman *et al.*, 2017), in contrast to mosses, whose diversification is characterized by several well-identified WGD events (Devos *et al.*, 2016; Lang *et al.*, 2018; Gao *et al.*, 2020; Carey *et al.*, 2021; Silva *et al.*, 2021). Liverworts might represent one of the most WGD-depauperate lineages among land plants, with only one putative WGD inferred at the most recent common ancestor (MRCA) of Jungermanniopsida as reported by the OneKP project (One Thousand Plant Transcriptomes Initiative, 2019). The *Marchantia* genome (Bowman *et al.*,

2017) bears no signatures of past WGDs, suggesting that the lineage of complex thalloid liverworts may not have undergone an ancient WGD.

Based on fossil evidence and molecular dating studies, the stem groups of extant liverworts diverged deep in time (Morris *et al.*, 2018), so that *Ks* saturation might have blurred molecular evolution patterns, hampering the tracing of WGD events (Wu *et al.*, 2019). Large-scale gene tree-based phylogenomic studies have been widely used to detect ancestral gene duplications and estimate their phylogenetic timing (Jiao *et al.*, 2011, 2012; Yang *et al.*, 2018). Gao *et al.* (2020) inferred gene duplication events in mosses with a transcriptome dataset and proposed six WGD events distributed across the backbone phylogeny of mosses. Whether WGDs similarly occurred during the evolutionary history of liverworts and shaped their diversification remains to be tested.

In this study, we newly generated transcriptomes for 38 liverwort accessions, representing 36 species, 11 genera and six orders of liverworts. With the additional 17 liverwort transcriptomes published by the OneKP project, our current sampling of 55 liverwort accessions broadened phylogenetic sampling of liverworts to 14 out of 15 orders (Fig. 1). We aimed to reconstruct the backbone phylogeny of liverworts, explore cyto-nuclear incongruences and infer possible ancestral gene duplications during the evolution of liverworts.

MATERIALS AND METHODS

Taxa sampling and DNA extraction

Fresh samples of 38 liverwort accessions were collected from China, Madagascar, New Zealand the USA and Vietnam, with voucher specimens deposited in SZG (Herbarium of Shenzhen FairyLake Botanical Garden, Shenzhen, China) and CONN (George Safford Torrey Herbarium at the University of Connecticut) (Supplementary data Table S1). We also included 18 liverwort transcriptomes from the OneKP project, and a liverwort (i.e. *Marchantia polymorpha*) genome representative in our analyses. Our exemplars span the liverwort phylogeny, representing all but one order (i.e. Neohodgsoniales) of extant liverworts (Crandall-Stotler *et al.*, 2009b). Each newly collected liverwort sample was cleaned with distilled water, dried, examined under the dissecting microscope for potential contaminants and processed for RNA extraction using the RNeasy Plant Mini Kit (Qiagen, Valencia, CA, USA). Each sample's RNA quality and quantity were examined using 1 % agarose gel electrophoresis, a Qubit fluorometer (Invitrogen) and a NanoDrop 2000 spectrophotometer.

To investigate the potential cyto-nuclear incongruences, we also downloaded the representative organellar genomes of liverworts (Supplementary data Table S2) from the NCBI organellar genome databases (<https://www.ncbi.nlm.nih.gov/genome/browse#!/organelles/>) with an ingroup taxon sampling consistent with that of the transcriptome dataset (Supplementary data Table S1). These mitochondrial and plastid genome data were used for phylogenetic reconstructions using the MP (maximum parsimony) and ML (maximum likelihood) methods.

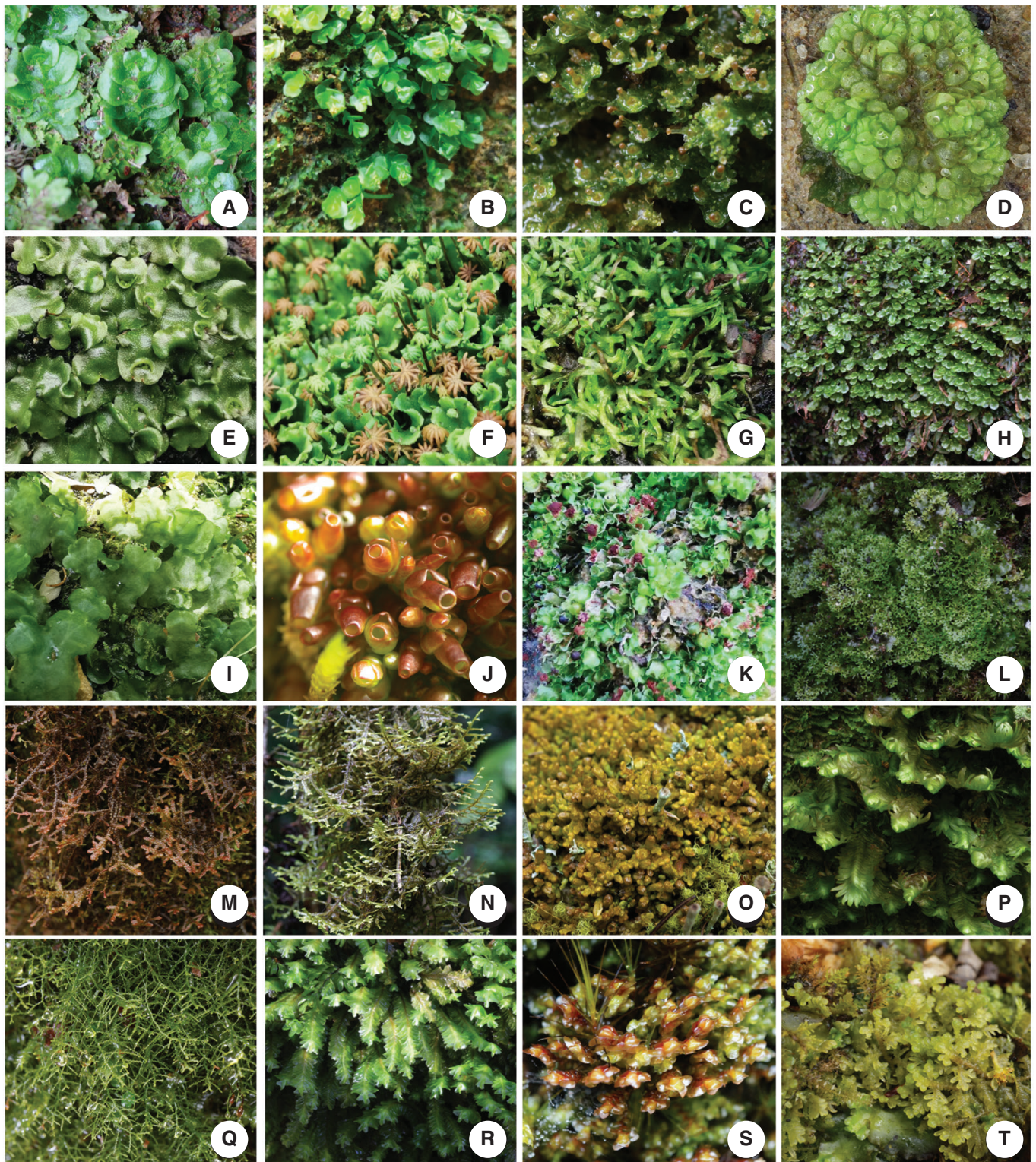


FIG. 1. Morphological diversity of liverworts. (A) *Treubia lacunosa*, Treubiales (photographed by J. G. Duckett, New Zealand); (B) *Haplomitrium mnioides*, Haplomitriales (photographed by S. Dong, Fujian, China); (C) *Blasia pusilla*, Blasiales (photographed by L. Zhang, Hunan, China); (D) *Sphaerocarpos texanus*, Sphaerocarpaceae (photographed by Blanka Aguero, North Carolina, USA); (E) *Lunularia cruciata*, Lunulariales (photographed by Y. Liu, lab culture, collected from North Rhine-Westphalia, Germany); (F) *Marchantia polymorpha*, Marchantiales (photographed by S. Dong, Fujian, China); (G) *Riccia fluitans*, Marchantiales (photographed by Y. Liu, Connecticut, USA); (H) *Pallavicinia ambigua*, Pallaviciniales (photographed by L. Zhang, Guizhou, China); (I) *Aneura maxima*, Metzgeriales (photographed by Y. Liu, Chongqing, China); (J) *Pleurozia acinose*, Pleuroziales (photographed by L. Zhang, Hainan, China); (K) *Fossombronina cristula*, Fossombroniales (photographed by Y. Liu, Shenzhen, China); (L) *Pellia endiviifolia*, Pelliiales (photographed by L. Zhang, Guizhou, China); (M) *Frullania* sp., Porellales (photographed by L. Zhang, Tibet, China); (N) *Ptychanthus striatus*, Porellales (photographed by L. Zhang, Guangdong, China); (O) *Ptilidium pulcherrimum*, Ptilidiales (photographed by L. Zhang, Xinjiang, China); (P) *Schistochila* sp., Schistochilales (photographed by L. Zhang, Gunung Jerai, Malaysia); (Q) *Lepidozia trichodes*, Jungermanniales (photographed by L. Zhang, Sichuan, China); (R) *Plagiochila* sp., Jungermanniales (photographed by L. Zhang, Tasmania, Australia); (S) *Scapania* sp., Jungermanniales (photographed by L. Zhang, Hunan, China); (T) *Trichocolea tomentella*, Jungermanniales (photographed by L. Zhang, Hunan, China).

Sequencing and genome assembly

Approximately 1 µg of high-quality RNA was used to generate paired-end sequencing libraries with the insert fragments of 200–300 bp of the corresponding cDNA. The libraries were sequenced on an Illumina HiSeq 2000 sequencing platform at the WuXiNextCode (Shanghai, China). Approximately 6 Gb of sequencing data were produced for each sample. The raw next-generation sequencing data were trimmed and filtered for adaptors, low-quality reads, undersized inserts and duplicate reads using Trimmomatic v0.36 (Bolger *et al.*, 2014). The cleaned reads for each species were *de novo* assembled using the TRINITY v2.13.1 (Grabherr *et al.*, 2013). The longest transcript for each gene was selected and used for subsequent analyses. Liverworts usually form symbiotic mycorrhiza and diverse fungal associations; hence, transcriptome data from the field-collected liverworts generally contain contaminant sequences from fungi, bacteria, viruses and even animals. To remove these before phylogenomic analyses, the transcript sequences were searched against the NCBI NT database with an e-value cut-off of 1e-6. Those transcripts with their top ten hits falling into the above-mentioned four categories were removed. The software TRANSDECODER v3.0.0 (<https://github.com/TransDecoder>) was used to predict the protein-coding regions for each clean dataset. CD-HIT v4.6.7 (Li and Godzik, 2006) was used to reduce the redundancy in the resultant gene set with -c 0.98 -aS 0.9 following Gao *et al.* (2020).

Phylogenomic analyses

For nuclear phylogenomic analyses of liverworts, we included 55 liverwort ingroup taxa and selected two species of mosses as outgroups. ORTHOFINDER v2.4.0 (Emms and Kelly, 2019) was used to cluster transcripts into homologue clusters with default settings (i.e. MCL Inflation = 1.5). The software KINFIN v1.0.3 (Laetsch and Blaxter, 2017) was used to select the most likely single-copy homologue clusters with the ORTHOFINDER output with default settings. Those clusters containing single-copy genes in >70 % of the taxa were selected for downstream phylogenomic analyses (Supplementary data Table S3). Each gene cluster was filtered to retain only single-copy genes and was aligned using a local version of TRANSLATORX (Abascal *et al.*, 2010), which first translates the nucleotide sequences into amino acid sequences using the standard genetic code, and then aligns those amino acid sequences using MAFFT v7.0 (Katoh *et al.*, 2005). The alignment is trimmed for ambiguous portions by GBLOCKS v0.91b (Talavera and Castresana, 2007) with -b3 (the least stringent settings with a maximum number of contiguous non-conserved positions) set to 8, -b4 (minimum length of a block) set to 5, -b5 (allowed gap positions) set to half and -b6 (use similarity matrices) set to yes. The cleaned amino acid alignment was then used as a guide to generate the nucleotide sequence alignment. The resultant individual codon alignments for each gene were concatenated into the combined nucleotide (NT) datasets with SEQKIT v0.3.1.1 (Shen *et al.*, 2016). As the third codon positions are prone to

substitution saturation and GC content heterogeneity (Liu *et al.*, 2014) (Supplementary data Table S4), for a lineage with a long evolutionary history such as liverworts, we also generated an NT12 dataset excluding the third codon positions. For each dataset, we estimated the ML trees with IQTREE2 v2.2.0 (Quang *et al.*, 2020) with 1000 ultra-fast bootstrap replicates (with the command: `iqtree2 -s alignment.phy -bb 1000 -m MF`) and inferred the coalescent tree with ASTRAL v5.7.3 (Mirarab *et al.*, 2014) with default settings. We generated the ML trees for the concatenated NT and NT12 datasets. MODELFINDER (Kalyaanamoorthy *et al.*, 2017) as implemented in IQTREE2 v2.2.0 (Quang *et al.*, 2020) was used to select the best-fit nucleotide substitution models, and ultra-fast 1000 bootstrap replicates were used. We also inferred the MP majority rule consensus tree for the concatenated datasets using the software MPBOOT v1.1.0 (Hoang *et al.*, 2018) with ultra-fast 1000 bootstrap replicates (with command: `-s alignment.phy -bb 1000`). MPBOOT searches for optimal trees using the ratchet based on branch swapping sub-tree pruning regrafting (Goloboff *et al.*, 2021).

Gene tree conflict was calculated and visualized with PHYPARTS v0.0.1 (Smith *et al.*, 2015) with default parameters (with commands: `java -jar phyparts-0.0.1-SNAPSHOT-jar-with-dependencies.jar -a 1 -v -d individual.trees -m astral.tre -o NT12_out && python phypartspiecharts.py astral.tre NT12_out 1480`). Quartet supports were characterized with ASTRAL v5.7.3 (Mirarab *et al.*, 2014) with the -t 8 option. For organellar phylogenomic analyses, each protein-coding gene was extracted in Geneious v10.0.1 (<https://www.geneious.com>), aligned with MAFFT v7.0 (Katoh *et al.*, 2005) and concatenated with SEQKIT v0.3.1.1 (Shen *et al.*, 2016). The programs IQTREE2 v2.2.0 and MPBOOT v1.1.0 were also used to calculate the organellar ML and MP trees, respectively, with parameter settings the same as those used for nuclear phylogenomic analyses. Concordance across the nuclear gene trees and the three cellular compartments was visualized using DISCOVISTA v1.0 (Sayyari *et al.*, 2018), which summarizes/visualizes the discordance among a set of phylogenetic trees. It compares the topology and support values among trees and visualizes the conflicts based on the threshold defined by the user. In the current study, strongly supported lineages are defined as those branches that received bootstrap support of ≥95 %, and weakly supported lineages are those that were recovered but received bootstrap support <95 %. Weakly rejected clades correspond to clades that are not present in the tree but are compatible if branches with low support (<85 %) collapse. Strongly rejected lineages are those that are not present in the tree and are not compatible even when branches with low support (<85 %) collapse. The phylogenetic trees were visualized and rooted in FigTree v1.4.1 (Rambaut, 2014). The ML estimates of branch-specific substitution rates were calculated with HyPhy v2.0 (Pond *et al.*, 2005) under the MG94W9 codon model (Muse and Gaut, 1994), allowing for independent estimation of d_N and d_S values for each branch (the local parameters option) following Richardson *et al.* (2013), using the concatenated nucleotide data matrix (excluding outgroups) and the corresponding ML tree inferred from the concatenated nucleotide dataset with IQTREE2 as mentioned above.

Examination of cyto-nuclear incongruences

We contrasted the topologies inferred from mitochondrial and plastid data against that obtained from nuclear loci and identified incongruences within the order of complex thal- loid, the Marchantiales, and among the early splits within the Jungermanniopsida, focusing on the position of *Pellia* and *Ptilidium*. These topological conflicts may reflect ancient in- complete lineage sorting (ILS) or hybridization. To test whether ILS alone can explain the incongruences observed, 400 000 gene trees were simulated under the assumption of the ILS scenar- io with the astral coalescent species tree. Then the topological frequencies of the simulated gene trees were summarized and contrasted with those of the observed gene trees (empirical gene trees from the phylogenomic analyses). Carefully selected representative taxa with good transcriptome data quality (as estimated by BUSCO assessments) will maximize the rela- tive gene tree space for the observed dataset. We build three additional phylogenomic datasets from ORTHOFINDER clus- tering results, i.e. (1) with 42 generic representatives for species network analyses to test the impact of hybridization, (2) with seven species (*Asterella wallichiana*, *Conocephalum concicum*, *Dumortiera hirsuta*, *Marchantia polymorpha*, *Monosolenium tenerum*, *Riccia berychiana* and *Wiesnerella denudata*) from Marchantiales for ILS simulation analyses and (3) with six representatives (*Marchantia polymorpha*, *Pellia endiviifolia*, *Pallavicinia lyellii*, *Ptilidium pulcherrimum*, *Porella navicularis* and *Nowellia curvifolia*) for ILS simulation analyses among the deep lineages of the Jungermanniopsida. We tested the hypoth- esis of ILS using the simulation methods as implemented in the R package PHYBASE v2.0 (Liu and Yu, 2010) and the hypoth- esis of hybridization using the InferNetwork_MPL option as implemented in PHYLONET v2.4 (Than et al., 2008) fol- lowing the methods used by Yang et al. (2020) and Wang et al. (2018). The frequencies of gene tree topologies were compared between the observed and simulated tree datasets, as well as the gene tree distances to species tree for the two datasets.

Gene duplication analyses

All 55 liverwort species, along with *Physcomitrium* (*Physcomitrella*) *patens* and *Selaginella moellendorffii* were used for gene family clustering with ORTHOFINDER v2.4.0 (Emms and Kelly, 2019) to infer gene duplication events. We filtered the homologue clusters and selected those that con- tained the outgroup *Selaginella* and at least 20 ingroup liver- wort species to build the gene family trees. These gene families were aligned with MAFFT v7.0 (Katoh et al., 2005), optimized with GBLOCKS v0.91b (Talavera and Castresana, 2007) with the least stringent settings with the aforementioned parameters, removing ambiguously aligned sequences with sites containing >50 % gaps. The resulting alignments were used as the input data for RAXML v8.0 (Stamatakis, 2006) for ML tree infer- ences with the PROTGAMMAAUTO model with 100 fast bootstrap replicates (with the command: raxml -f a -x 12345 -# 100 -m PROTGAMMAAUTO -s alignment.phy -n alignment -p 12345). The software finds the optimal tree under the popular ML criterion with the starting tree generated by parsimony inference. The resulting gene family trees were reconciled

and summarized against the species tree with the software TREE2GD v1.0.40 (<https://sourceforge.net/projects/tree2gd>) with the bootstrap threshold parameter set to 50. Although this software uses the bootstrap support thresholds of >50 % to increase its precision in identifying gene duplications, it uses only the gene trees with the outgroup taxon *Selaginella* included; thus, a majority of gene trees without *Selaginella* were discarded, which might result in biases due to a reduced sample size of gene family trees. Moreover, this software tends to have overestimated duplications for the deeper nodes be- cause it only requires that the two clades of paralogues share at least two species. For these reasons, we also inferred ancestral gene duplications with a novel, fast and scalable duplication- loss-coalescent (DLC) resolution algorithm as implemented in ORTHOFINDER v2.4.0 which applied a hybrid algorithm of the species-overlap method (Huerta-Cepas et al., 2007) and DLCpar analysis (Wu et al., 2014) with all gene trees in- cluded. Those nodes showing evidence of a gene duplication event through overlapping species sets with >50 % of the spe- cies overlapped for the two clades of paralogues were analysed using the DLC model to find the most parsimonious inter- pretation for each node, in terms of which genes diverged at the gene duplication event and which diverged at a speciation event (Emms and Kelly, 2019).

We calculated the K_s values for pairs of paralogues in repre- sentative liverwort species using the FASTKs pipeline ([https:// github.com/mrmckain/FASTKs](https://github.com/mrmckain/FASTKs)) that incorporates BLASTN (Camacho et al., 2009), MAFFT v7.0 (Katoh et al., 2005), PAL2NAL v14 (Suyama et al., 2006) and PAML v.4.4c (Yang, 2007). We drew the K_s plot with package GGLOT2 (Ginestet, 2011) as implemented in R 3.6.3 (<https://www.r-project.org>). We also inferred K_s -based age distributions for paralogous genes using WGD v1.1.2 (Zwaenepoel and Van de Peer, 2018). An all vs. all blast on protein sequences was conducted using BLASTP with an E-value cut-off of $1e-10$, and then the MCL package (Enright et al., 2002) was used for the reconstruc- tion of gene families, and MAFFT v7.0 (Katoh et al., 2005) for alignment within each family. Finally, gene families (with n members) of $n \times (n - 1)/2 > \text{'max_pairwise'}$ were removed, and the phylogenetic tree was built for each gene family using FASTTREE v1.6.0 (Price et al., 2009). The K_s values of each pairwise comparison were obtained using ML in the CODEML program of PAML v.4.4c (Yang, 2007), and the weighted K_s values were used for K_s distribution construction. Finally, we performed a mixture modelling for all possible WGD infer- ences using the BGMM method.

RESULTS

Transcriptome assemblies and homologue clusters

A total of 57 species were included in the phylogenomic analyses (Fig. 2) (Supplementary data Table S1). The non-redundant pro- tein sequences for each species ranged from 9908 (*Sphaerocarpos texanus*) to 73 862 (*Nowellia curvifolia*), with an average 26 566 uni-genes. BUSCO assessments recovered a complete BUSCO ratio ranging from 45.3 % (*Lunularia cruciata*) to 96.7 % (*Marchantia polymorpha*), with an average ratio of 66.2 % (Supplementary data Table S1). Peptide sequences from 57 species

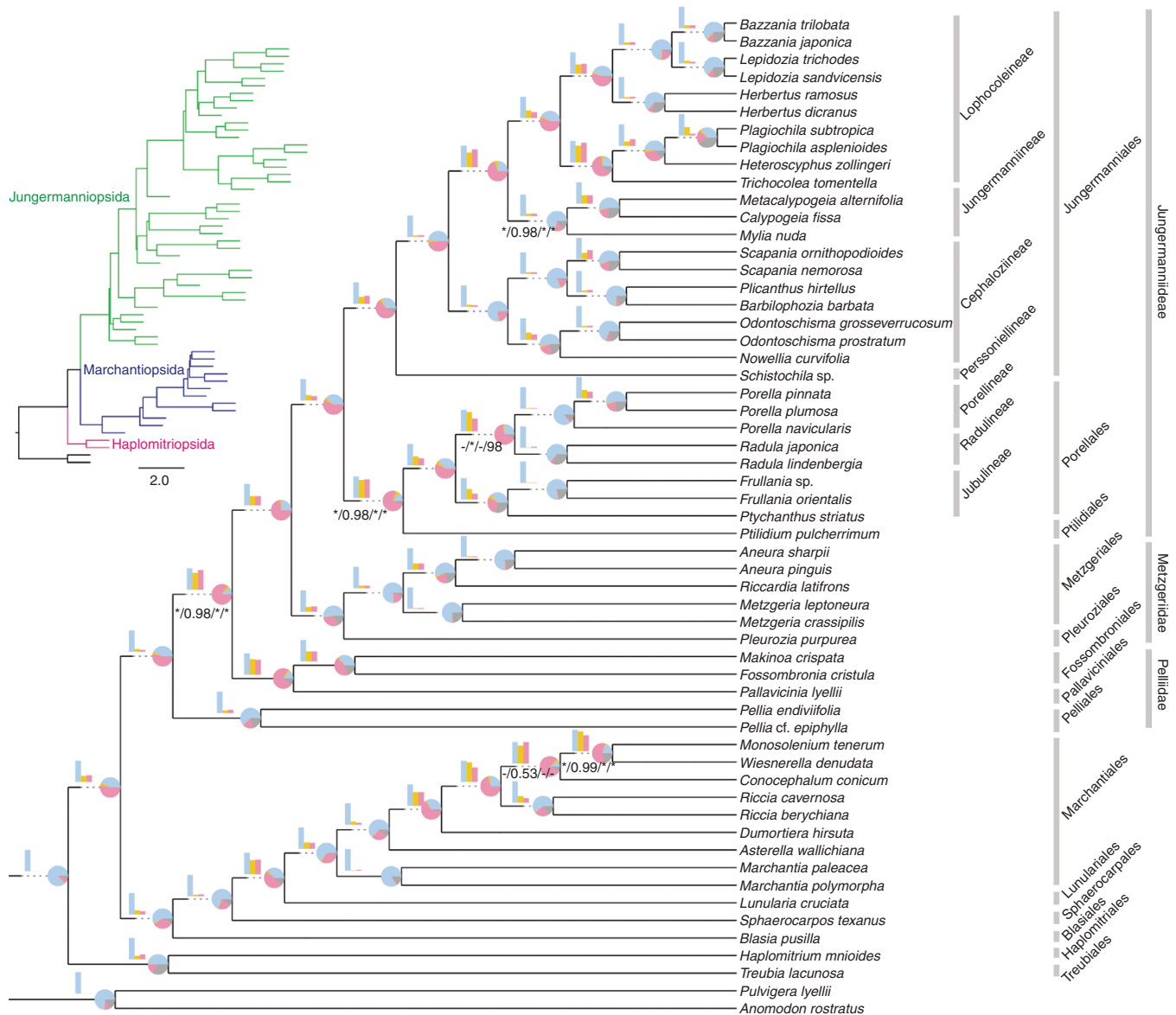


FIG. 2. The 57 species liverwort coalescent tree based on 1480 low-copy nuclear genes with the third codon positions excluded. Individual gene tree incongruences are shown as pie charts at the nodes and quartet supports as bar charts above the branches. For the pie chart, blue, orange, pink and black areas represent the percentages of gene trees showing concordance, top alternative topology, other conflicting topologies and no signal, respectively. For the bar chart, blue, orange and pink bar lengths represent quartet supports for the main topology (q1), the first alternative topology (q2) and the second alternative topology (q3), respectively. Astral PP for codon dataset (PP NT) and codon 1st + 2st (PP NT12) dataset, IQTREE BPP for codon dataset (BPP NT) and codon 1st + 2st dataset (BPP NT12), MPBOOT BPP for codon dataset (MBPP NT) and codon 1st + 2st dataset (MBPP NT12) are shown below the branches as PP NT/PP NT12/BPP NT/BPP NT12/MBPP NT/MBPP NT12. Branches are maximally supported unless otherwise indicated. Asterisks represent those maximally supported by either analysis.

were sorted into 84 468 homologue clusters, which were reduced to 1480 by retaining only those containing single-copy genes for at least 70 % of the taxa (Supplementary data Table S3), for nuclear phylogenomic reconstruction purposes. For gene duplication analyses with TREE2GD, the homologous gene clusters were reduced to 4929 by retaining those containing the outgroup *Selaginella* and at least 20 of the ingroup liverwort taxa. For duplication analyses with ORTHOFINDER, a total of 28 321 gene families were used.

Phylogenomic reconstructions

Alignments of 1480 mostly single-copy gene families were concatenated to generate an NT supermatrix that contained

1 244 946 sites, with 801 094 (64.4 %) sites potentially parsimony informative. We also produced an NT12 supermatrix of 829 964 sites (of which 399 616 sites are potentially parsimony informative) with third codon positions excluded. The contributions of each species to the nuclear single-copy orthogroups are presented in Supplementary data Table S1 (in the column 'Loci Recovered'). GTR + I + R5 and GTR + I + R6 models were selected as the best-fit models for phylogenomic reconstruction for NT and NT12 supermatrices, respectively. Our nuclear phylogenomic reconstructions (NT coalescent, NT12 coalescent trees, NT concat ML, NT12 concat ML, NT concat MP and NT12 concat MP) are mostly congruent with each other, with the exception of the deep relationships within Porellales

and the affinities of *Conocephalum* in Marchantiales. *Radula* formed a sister group relationship with *Porella* in the NT12 trees or Jubulineae in the NT trees. *Conocephalum* formed a well-supported sister group relationship with *Riccia* in all trees but not in the NT12 coalescent tree (Supplementary data Fig. S1).

Overall, the present phylogenetic reconstructions are generally consistent with the current liverwort classification (Fig. 1; Supplementary data Fig. S2), with Haplomitriopsida consisting of *Haplomitrium* and *Treubia* emerging from the first split, followed by successive divergences of complex thalloids (Marchantiopsida), Pelliidae, Metzgeriidae and Jungermanniidae, which comprise Porellales and Jungermanniales. *Pellia* was weakly to strongly supported as sister to the rest of Jungermanniopsida in the nuclear and mitochondrial trees, resulting in a paraphyletic, rather than a monophyletic Pelliidae, as supported in the plastid trees (Fig. 3; Supplementary data Figs S1 and S2). The quartet supports and the gene tree conflicts at the focal node further corroborated the unstable position of *Pellia* (Fig. 2). The enigmatic *Ptilidium* is consistently and strongly supported as sister to Porellales, and Jungermanniineae are supported as sister to Lophocoleineae (Fig. 2). The three major liverwort lineages show similar nucleotide substitution rates (Supplementary data Fig. S3), in contrast to the exceptionally slow substitution rates in organellar loci in complex thalloid liverworts (Villarreal et al., 2016).

Gene tree conflicts and cyto-nuclear incongruences

Nodes showing intense gene tree conflicts also tended to show strong conflicts in quartet supports (Fig. 2). The affected

nodes are mainly those previously identified as problematic across analyses (Fig. 3), namely the relationships among the three lineages of Jungermanniales or the three lineages of Porellales, of the Pelliidae or *Ptilidium*, and within the Marchantiales (Fig. 3). The most significant conflict involves *Ptilidium*, which was supported as sister to Porellales in the nuclear and plastid trees but sister to Jungermanniales in the mitochondrial tree. Nuclear and mitochondrial analyses resolve *Pellia* as sister to the rest of Jungermanniopsida, whereas the plastid data support a unique ancestry with the rest of Pelliidae. In the complex thalloid clade, *Asterella* and *Dumortiera* constitute sister taxa in the mitochondrial tree only, and *Conocephalum* and *Riccia* are sister lineages in organellar trees and all nuclear trees, except in the NT12 coalescent tree where such a relationship was not recovered (Fig. 3; Supplementary data Fig. S3). The alternative placements of deep relationships within Porellales might be related to the substitutional saturation in the third codon positions of nuclear genes (Supplementary data Fig. S4), since the nuclear NT12 dataset consistently recovered the sister relationship of *Porella* and *Radula*, which is congruent with the organellar trees.

Hybridization and incomplete lineage sorting

The strong cyto-nuclear incongruences in liverwort phylogeny were further investigated for potential evidence of ILS or hybridization (Fig. 4A). Several putative gene flow events were detected on the generic level phylogeny with 42 liverwort representatives. The species tree network of $n = 6$ (Fig. 4A)

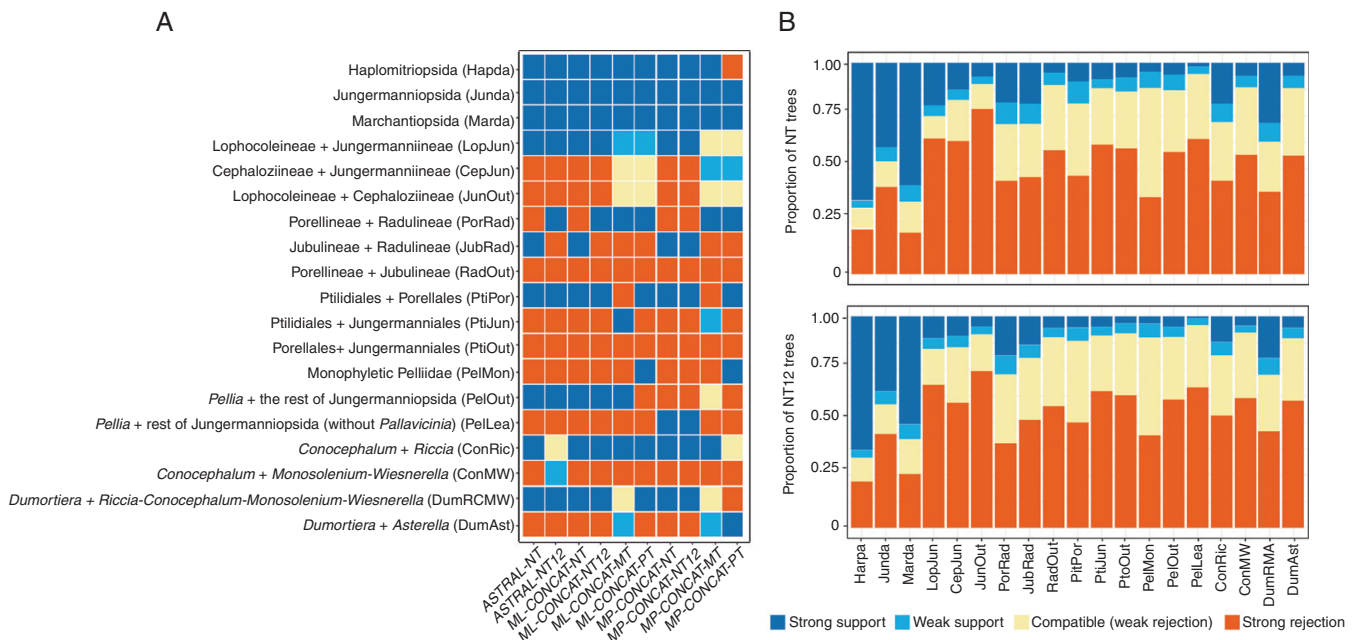


FIG. 3. Statistics of species trees and gene trees. (A) Species tree summarization using DiscoVista. Rows correspond to focal splits, and columns correspond to the phylogenetic results of different methods reported in nuclear, mitochondrial and plastid datasets. (B) Gene tree compatibility. The portion of gene trees for which focal splits are highly (or weakly) supported (or rejected). MT, mitochondrial genes; PT, plastid genes; NT, nucleotide sequences of nuclear genes; NT12, codon 1st + 2nd positions of nuclear genes. ASTRAL, coalescent tree inference method using Astral software; CONCAT, concatenated datasets; ML TREE, maximum likelihood tree inferred with IQTREE; MP TREE, maximum parsimony tree inferred with MPBOOT.

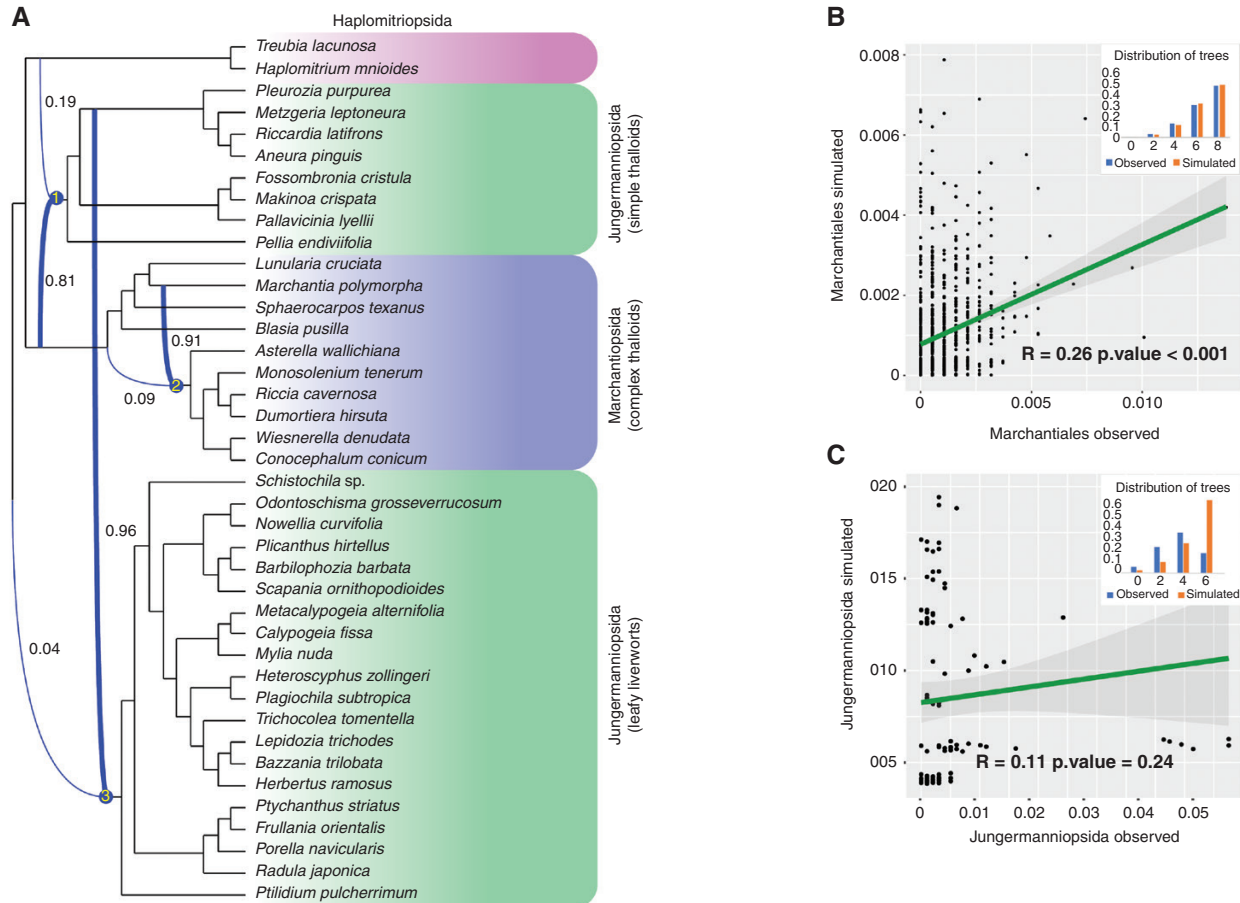


FIG. 4. Hybridization and ILS simulation results. (A) The best species tree network was inferred with PHYLONET (reticulations = 6). Blue lines and dots denote reticulation edges and reticulation nodes, respectively. Those edges consistent with conventional phylogenetic tree topology are highlighted with thick lines. The inheritance probabilities are shown on the reticulation edges. Gene tree frequencies for the observed and simulated datasets for Marchantiales (B) and Jungermanniopsida (C). The upper right bar charts indicate the species tree distance for the observed and simulated datasets.

had a higher probability in likelihood than any other reticulations (Supplementary data Fig. S5). The backbone relationships of the liverwort network tree were largely consistent with the currently accepted liverwort phylogeny, albeit with three reticulation nodes (Fig. 4A). Reticulation node 1 denoted that about 81 % of the loci in the ancestor of Jungermanniopsida were inherited from the ancestor of Marchantiopsida while 19 % of the loci in the ancestor of Jungermanniopsida were inherited from the ancestor of Haplomitriopsida. Reticulation node 2 denoted a scenario where about 91 % of the loci in the ancestor of Ricciales–Marchantiales (without *M. polymorpha*) shared an MRCA with loci from *M. polymorpha*, and 9 % of the loci in the ancestor of Ricciales–Marchantiales shared an MRCA with loci from the ancestor of Marchantiopsida. Reticulation node 3 denoted that roughly 96 % of the loci in the ancestor of leafy liverworts were inherited from the ancestor of Metzgeriidae and 4 % of the loci in the ancestor of leafy liverworts were inherited from the ancestor of liverworts (Fig. 4A).

The ILS simulation results for the backbone relationships of Jungermanniopsida, represented by six taxa, and Marchantiales, represented by seven taxa, might point to different evolutionary processes for the species involved in the

two groups. The Marchantiales dataset showed similar frequencies of different tree topologies, as well as similar distributions of gene tree distance (compared with the nuclear species tree), comparing the simulated and the empirical datasets, respectively (Fig. 4B), indicating that incomplete lineage sorting alone could explain the cyto-nuclear gene tree incongruences in Marchantiales. The Jungermanniopsida datasets showed distinct tree topology frequencies and gene tree distance distributions (Fig. 4C), and with no significant correlations detected between the simulated and the observed datasets, indicating that ILS alone could not explain the cyto-nuclear tree incongruences observed on the backbone relationships of Jungermanniopsida, and that instead hybridization might have played an important role in shaping their evolutionary history.

Ancestral gene duplications

The K_s distribution plots were generated for 52 liverwort species involved in the current study (Supplementary data Fig. S6). Considering that most liverwort species form symbiotic associations with fungi and bacteria (Read *et al.*, 2000), our

field-collected specimens inevitably contained contaminations from the symbionts and other micro-organisms. Those transcripts having their top ten hits falling into the categories of bacteria, fungi, viruses and animals were removed, and the rest, including those (the bulk) that did not yield hits from the NCBI NT database, were kept, acknowledging that these might also contain unidentified contaminants. Therefore, the *Ks* plot of liverworts should be interpreted with prudence. Although most of the liverwort species show no evidence of WGDs, some liverwort species show a *Ks* peak at around 0.5, which might suggest ancestral duplication at the shallow nodes of liverworts rather than WGDs at the backbone liverwort phylogeny, considering their deep split time (Morris et al., 2018).

Large-scale gene tree-based analyses based on 4929 multicopy homologue groups using the program TREE2GD identified negligible numbers of gene duplications along the backbone phylogeny of liverworts (Fig. 5). The number of gene duplication events at the MRCA of liverworts was only 203 (approx. 5.18 %), and at the MRCA of Marchantiopsida plus Jungermanniopsida was 129 (approx. 2.23 %) events. Our study did not identify the previously reported putative WGD event at the MRCA of Jungermanniopsida (One Thousand Plant Transcriptomes Initiative, 2019), and only 34 (approx. 1.09 %) gene duplication events were identified in our analysis. Gene duplication results inferred with ORTHOFINDER with 28 321 gene trees (Supplementary data Fig. S7) were consistent with the TREE2GD result, with very small numbers of gene duplication events identified along the liverwort backbone phylogeny. The two lineages of Haplomitriopsida did not show significant gene duplication events (Fig. 5), indicating that the *Ks* peaks observed in *Treubia* and *Haplomitrium* were not the signals of shared gene duplications. The same is also true regarding other liverwort species that show *Ks* peaks (Supplementary data Fig. S6), such as *Blasia pusilla*, *Pellia endiviifolia*, *Aneura pinguis*, *Pleurozia purpurea*, *Nowellia curvifolia* and the two species of *Lepidozia*.

DISCUSSION

Gene tree conflicts and cyto-nuclear incongruences characterized the backbone phylogeny of liverworts

Our nuclear phylogenomic reconstructions yielded a well-supported liverwort phylogeny that is mostly consistent with recent phylogenetic studies (Forrest et al., 2006; Yu et al., 2019; Dong et al., 2021b), with all classes, orders and families resolved as monophyletic except for Pelliidae, which are recovered as paraphyletic. Despite maximum support for most nodes in nuclear coalescent ML and MP trees, we observed strong gene tree incongruences as evidenced by quartet supports and gene tree mapping results on the NT12 coalescent tree, such as the node subtending the majority of Pelliidae (without *Pellia*), and the node of most Jungermanniopsida (without *Pellia*), the node involving *Ptilidium*, the node involving *Radula* and *Porella*, the node involving Lophocoleineae and Jungermanniineae, and the nodes involving several species in Marchantiales. The relationships among Porellales recovered from the nuclear NT and NT12 datasets were incongruent, with *Radula* sister to Jubulineae based on the former, vs. sister to Porellineae in the latter, a result consistent with inferences from organellar

data. This incongruence may arise from extensive homoplasy due to substitutional saturation in the third codon positions introducing phylogenetic artefacts (Liu et al., 2014).

Cyto-nuclear incongruences characterize nodes across the phylogeny of land plants (Meng et al., 2021) including, as shown here, of liverworts. *Pellia* was supported as sister to the remaining Jungermanniopsida in the nuclear (Fig. 2) and mitochondrial trees but sister to the remaining Pelliidae in the plastid tree (Supplementary data Fig. S2). The sister relationship of *Ptilidium* with Porellales is congruent with the plastid tree but incongruent with the mitochondrial tree wherein *Ptilidium* is resolved as sister to Jungermanniales. Within complex thalloids, Marchantiales are characterized by inconsistent internal relationships across different datasets, i.e. *Asterella* and *Dumortiera* constitute sister taxa in the mitochondrial tree only, and *Conocephalum* and *Riccia* are consistently strongly supported as sister taxa except in the nuclear NT12 coalescent tree, with the focal nodes showing similar quartet support values, suggesting gene tree discordances (Sayyari and Mirarab, 2016).

These phylogenetic incongruences among gene trees and different genomic compartments might suggest ILS (Yang et al., 2020; Wakeley, 2009), hybridizations (Morales-Briones et al., 2018) and lateral gene transfers (Tofigh et al., 2011), gene losses and duplications (Goodman et al., 1979; Page, 1994). The most significant of these nodes coincided with the long-standing systematic ambiguities, indicating that these areas of the tree that are poorly resolved may reflect a diversity of speciation processes, some of which may be incompatible with a bifurcating tree. As has been observed in angiosperms (Vriesendorp and Bakker, 2005), the phylogeny of liverworts might be better represented by species networks.

Hybridization and ILS imprint the evolution of liverworts

Cyto-nuclear incongruences mainly persisted within Marchantiales and regarding the relationships of *Pellia* and *Ptilidium* to the major lineages of Jungermanniopsida. The species network recovered here (Fig. 4) might be a good representation of liverwort morphological classifications, as the network identified four major clades, corresponding to Haplomitriopsida, complex thalloids, simple thalloids and leafy liverworts. As shown in Fig. 3, the reticulation edges of thick lines connecting these four lineages reflect a good ordinal divergence pattern consistent with current molecular phylogenomic studies (Forrest et al., 2006; Yu et al., 2019; Dong et al., 2021b), and the reticulation edges of thin lines represent gene flows from other lineages. For example, all simple thalloid species clustered in one clade that shows signatures of gene flow from Haplomitriopsida in its MRCA (Fig. 4A). Although we did not detect gene flows directly involving the two problematic lineages, *Pellia* and *Ptilidium*, the immediate ancestor of each lineage shows evidence of gene flow, suggesting that ancient hybridization played essential roles in the evolution of liverworts.

Incomplete lineage sorting is another critical factor that may account for gene tree and species tree discordances (Sayyari and Mirarab, 2016). The ILS simulation analyses (Yang et al., 2020) identified a significant correlation between the observed topological frequencies and the simulated frequencies, indicating that ILS alone could explain the cyto-nuclear

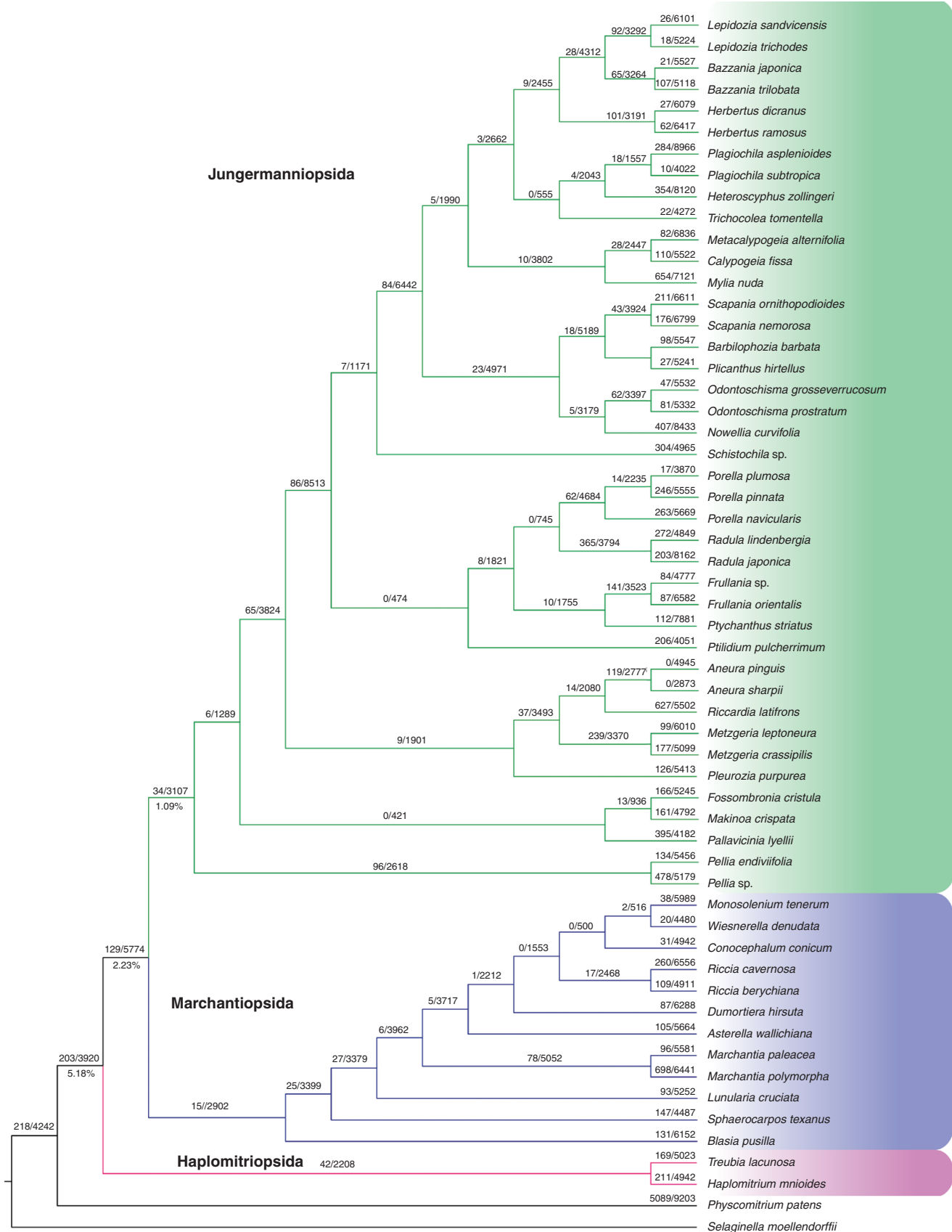


FIG. 5. Gene duplications along liverwort phylogeny inferred with the software TREE2GD based on large-scale phylogenomic analyses. The number of duplication nodes with bootstrap values >50 % and all the nodes detected in the trees were counted and labelled at the corresponding ancestral nodes on the species tree.

incongruences observed in Marchantiales. The ILS simulations in Jungermanniopsida failed to identify such a strong correlation (Fig. 4), suggesting that other processes, such as hybridization, gene duplications and losses, might have also shaped the gene tree incongruences involving *Pellia* and *Ptilidium*.

Liverworts can reproduce clonally via specialized or undifferentiated diaspores (Crandall-Stotler *et al.*, 2009a) or sexually via motile sperm cells. Hybridization may have contributed to the diversification of liverworts, but it was expected to be rare (Laenen *et al.*, 2016). However, phenotypic diagnosis of genome mergers is challenged by the relatively simple and microscopic character space-defining species, such that evidence for hybridization emerged primarily from molecular evidence (Boisselier-Dubayle *et al.*, 1998; Buczkowska *et al.*, 2012). Linde *et al.* (2020) revealed reticulation during the diversification of the *M. polymorpha* species complex, suggesting that hybridization and introgression may be more frequent than previously considered.

Here, we detect signals of ancient hybridization in the backbone phylogeny of liverworts involving the ancestor of simple thalloids and the ancestor of Haplomitriopsida; the ancestor of leafy liverworts and the ancestor of liverworts; and the ancestor of complex thalloids and the core Marchantiales (excluding *M. polymorpha*). We also provide evidence for ILS in a complex thalloid clade. Hence, the current study appears to be consistent with a recent study in peat moss (Meleshko *et al.*, 2021) that supported ancient hybridizations among ancestral lineages followed by ILS rather than recent gene flow as shaping cytonuclear incongruences.

Genomic evolution of liverworts

Despite a few allopolyploid cases (Boisselier-Dubayle *et al.*, 1998; Buczkowska *et al.*, 2012), the haploid chromosome number of liverworts is generally eight, nine or ten (Fritsch, 1991), suggesting that the karyotype of liverworts is relatively stable [compared with, for example, mosses (Patel *et al.*, 2021)] and its evolution overall is rarely shaped by WGDs. The *Marchantia* genome (Bowman *et al.*, 2017) holds no signatures of WGD, lending support to the hypothesis that liverworts might not have undergone genome doubling events. Analyses of transcriptome data (One Thousand Plant Transcriptomes Initiative, 2019), however, suggested that liverworts might have experienced at least one WGD, namely in the stem clade of Jungermanniopsida. *Ks* distribution analyses failed to recover the WGD in the stem of Jungermanniopsida, but the WGD signal in the *Ks* plot might be lost in lineages with an ancient divergence time, such as the liverwort deep clades (Morris *et al.*, 2018). Our phylogenomic analyses with extended liverwort transcriptome sampling also rejected the hypothesis of WGD along the liverwort backbone phylogeny, although some liverwort species might have experienced their lineage-specific WGDs. While the crown leafy liverworts (i.e. Jungermanniales and Porellales) also show species radiation as in mosses (Laenen *et al.*, 2016), liverworts are indeed one of the most WGD-depauperate embryophyte lineages, like hornworts (One Thousand Plant Transcriptomes Initiative, 2019; Zhang *et al.*, 2020). Whether genetic mechanisms, or the architecture of liverworts, especially their short-lived sporophyte exhibiting

perhaps less totipotency of its cells, and hence less potential for aposporous propagation yielding polyploid gametophyte, could account for this peculiar phenomenon remains to be explored.

SUPPLEMENTARY DATA

Supplementary data are available online at <https://academic.oup.com/aob> and consist of the following. Figure S1: nuclear liverwort phylogeny based on Astral coalescent analyses, concatenated maximum likelihood analyses and maximum parsimony analyses for codon datasets and those without third positions. Figure S2: liverwort phylogeny based on concatenated protein-coding genes for the mitochondrial and plastid data. Figure S3: nucleotide substitution rates in liverworts. Figure S4: distribution of the GC content across all single-copy genes for each species. Figure S5: phylonet analyses of 42 liverwort representatives on genus-level sampling with 1388 rooted gene trees investigated. Figure S6: *Ks* distributions for liverwort species involved in the current study. Figure S7: detection of focal nodes with a high concentration of ancestral paralogous gene duplication events. Table S1: summary of taxa sampling and sequencing.

ACKNOWLEDGEMENTS

We are extremely grateful to Dr Qin Zuo and Hui Dong (Fairy Lake Botanical Garden) for assistance in acquiring liverwort material of *Fossombronina cristula*, Dr David Glenny (Landcare Research, New Zealand) for sharing material of *Treubia lacunosa*, and Dr J. G. Duckett and Dr Blanka Aguero for providing the photos of *Treubia lacunosa* and *Sphaerocarpos texanus*, respectively. Dr Zhen Li from Ghent University, Belgium, and Dr Jiayu Xue from Nanjing Agricultural University assisted in the *Ks* analyses. We gratefully acknowledge the lab assistance by Na Li and Jiachen Chen at the Fairy Lake Botanical Garden. Y.L. conceived and designed the study. Y.L., S.D. and B.G. collected the fresh liverwort samples. S.D. carried out the lab experiments and analyses, and drafted the article. Y.L. and B.G. revised the article, and all the authors approved the article. The authors alone are responsible for the content and writing of the article.

The authors report no conflicts of interest.

DATA DEPOSITION

The transcriptome data have been submitted to the NCBI under the SRA accession numbers SRR8202174–SRR8202215 as listed in Supplementary data Table S1. The transcriptome assemblies, annotations and dataset used for phylogenomic analyses have been deposited in the Figshare data repository under the doi: 10.6084/m9.figshare.14452662. All relevant data are available from the authors

FUNDING

This project was supported by the Scientific Foundation of the Urban Management Bureau of Shenzhen (grant nos 202106 to S.D. and 440300221271500200125 to Y.L.).

LITERATURE CITED

- Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research* **38**: W7–W13. doi:10.1093/nar/gkq291.
- Birchler JA, Yang H. 2022. The multiple fates of gene duplications: deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *The Plant Cell* **34**: 2466–2474. doi:10.1093/plcell/koac076.
- Boisselier-Dubayle MC, Lambourdiere J, Bischler H. 1998. The leafy liverwort *Porella baueri* (Porellaceae) is an allopolyploid. *Plant Systematics and Evolution* **210**: 175–197. doi:10.1007/bf00985667.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120. doi:10.1093/bioinformatics/btu170.
- Bowman J, Kohchi T, Yamato KT, et al. 2017. Insights into land plant evolution garnered from the *Marchantia polymorpha* genome. *Cell* **171**: 287–304.
- Buczowska K, Sawicki J, Szczecińska M, Klama H, Bczkiewicz A. 2012. Allopolyploid speciation of *Calypogeia sphagnicola* (Jungermanniopsida, Calypogeiaceae) based on isozyme and DNA markers. *Plant Systematics and Evolution* **298**: 549–560.
- Camacho C, Coulouris G, Avagyan V, et al. 2009. Blast+: architecture and applications. *BMC Bioinformatics* **10**: 421.
- Carey SB, Jenkins JW, Lovell JT, et al. 2021. Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. *Science Advances* **7**: eabn2488.
- Crandall-Stotler BJ, Forrest LL, Stotler RE. 2005. Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). *Taxon* **54**: 299–316. doi:10.2307/25065359.
- Crandall-Stotler B, Stotler RE, Long DG. 2009a. Morphology and classification of the Marchantiophyta. In: Goffine B, Shaw AJ, eds. *Bryophyte biology*. Cambridge: Cambridge University Press, 1–54.
- Crandall-Stotler B, Stotler RE, Long DG. 2009b. Phylogeny and classification of the Marchantiophyta. *Edinburgh Journal of Botany* **66**: 155–198. doi:10.1017/s0960428609005393.
- Devos N, Szovenyi P, Weston DJ, Rothfels CJ, Johnson MG, Shaw AJ. 2016. Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). *New Phytologist* **211**: 300–318. doi:10.1111/nph.13887.
- Dong S, Li H, Goffinet B, Liu Y. 2021a. Exploring the impact of RNA editing on mitochondrial phylogenetic analyses in liverworts, an early land plant lineage. *Journal of Systematics and Evolution* **60**: 16–22. doi:10.1111/jse.12706.
- Dong S, Zhang S, Zhang L, Wu H, Liu Y. 2021b. Plastid genomes and phylogenomics of liverworts (Marchantiophyta): conserved genome structure but highest relative plastid substitution rate in land plants. *Molecular Phylogenetics and Evolution* **161**: 107171.
- Dong SS, Wang YL, Xia NH, et al. 2022. Plastid and nuclear phylogenomic incongruences and biogeographic implications of *Magnolia sl* (Magnoliaceae). *Journal of Systematics and Evolution* **60**: 1–15.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biology* **20**: 238.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research* **30**: 1575–1584. doi:10.1093/nar/30.7.1575.
- Flores JR, Catalano SA, Munoz J, Suarez GM. 2017. Combined phylogenetic analysis of the subclass Marchantiidae (Marchantiophyta): towards a robustly diagnosed classification. *Cladistics* **34**: 5171–5541. doi:10.1111/cla.12225.
- Flores JR, Bippus AC, Suárez GM, Hyvönen J. 2020. Defying death: incorporating fossils into the phylogeny of the complex thalloid liverworts (Marchantiidae, Marchantiophyta) confirms high order clades but reveals discrepancies in family-level relationships. *Cladistics* **37**: 231–247. doi:10.1111/cla.12442.
- Forrest LL, Davis EC, Long DG, Crandall-Stotler BJ, Clark A, Hollingsworth ML. 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. *Bryologist* **109**: 303–334.
- Fritsch R. 1991. Index to bryophyte chromosome counts. *Bryophytorum Bibliotheca* **40**: 1–352.
- Gao B, Chen MX, Li XS, et al. 2020. Ancestral gene duplications in mosses characterized by integrated phylogenomic analyses. *Journal of Systematics and Evolution* **60**: 144–159.
- Ginestet C. 2011. GGPlot2: Elegant graphics for data analysis. *Journal of the Royal Statistical Society Series A: Statistics in Society* **174**: 245–246.
- Goffinet B. 2000. Origin and phylogenetic relationships of the Bryophyta. In: Goffine B, Shaw AJ, eds. *Bryophyte biology*. Cambridge: Cambridge University Press, 124–149.
- Goloboff PA, Catalano SA, Torres A. 2021. Parsimony analysis of phylogenomic datasets (ii): evaluation of paup*, mega and mpboot. *Cladistics* **38**: 126–146. doi:10.1111/cla.12476.
- Goodman M, Czelusniak J, Moore G, Romero-Herrera A, Matsuda G. 1979. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. *Systematic Zoology* **28**: 132–163.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Amit I. 2013. Trinity: reconstructing a full-length transcriptome without a genome from RNA-seq data. *Nature Biotechnology* **29**: 644.
- Haberle RC, Fourcade HM, Boore JL, Jansen RK. 2008. Extensive rearrangements in the chloroplast genome of *Trachelium caeruleum* are associated with repeats and tRNA genes. *Journal of Molecular Evolution* **66**: 350–361. doi:10.1007/s00239-008-9086-4.
- Harris BJ, Harrison CJ, Hetherington A, Williams T. 2020. Phylogenomic evidence for the monophyly of bryophytes and the reductive evolution of stomata. *Current Biology* **30**: 2001–2012.
- Heinrichs J, Hentschel J, Wilson R, Feldberg K, Schneider H. 2007. Evolution of leafy liverworts (Jungermanniidae, Marchantiophyta), estimating divergence times from chloroplast DNA sequences using penalized likelihood with integrated fossil evidence. *Taxon* **56**: 31–44.
- Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ. 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evolutionary Biology* **18**: 11.
- Huerta-Cepas J, Dopazo H, Dopazo J, Gabaldon T. 2007. The human phylome. *Genome Biology* **8**: R109. doi:10.1186/gb-2007-8-6-r109.
- Jiao Y, Wickett NJ, Ayyampalayam S, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97–100.
- Jiao YN, Leebens-Mack J, Ayyampalayam S, et al. 2012. A genome triplcation associated with early diversification of the core eudicots. *Genome Biology* **12**: R3.
- Kalyanamoorthy S, Minh BQ, Wong TKF, Haeseler AV, Jeremiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods* **14**: 587–589.
- Katoh K, Kuma K, Toh H, Miyata T. 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511–518. doi:10.1093/nar/gki198.
- Laenen B, Machac A, Gradstein SR, et al. 2016. Increased diversification rates follow shifts to bisexuality in liverworts. *New Phytologist* **210**: 1121–1129. doi:10.1111/nph.13835.
- Laetsch D, Blaxter M. 2017. KinFin: software for taxon-aware analysis of clustered protein sequences. *G3: Genes, Genomes, Genetics* **7**: 3349–3357.
- Lang D, Ullrich K, Murat F, et al. 2018. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. *The Plant Journal* **93**: 515–533.
- Lemmon AR, Emme SA, Lemmon EM. 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* **61**: 727–744. doi:10.1093/sysbio/sys049.
- Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658–1659. doi:10.1093/bioinformatics/btl158.
- Linde AM, Sawangproh W, Cronberg N, Szovenyi P, Lagercrantz U. 2020. Evolutionary history of the *Marchantia polymorpha* complex. *Frontiers in Plant Science* **11**: 829. doi:10.3389/fpls.2020.00829.
- Liu L, Yu L. 2010. Phybase: an R package for species tree analysis. *Bioinformatics* **26**: 962–963. doi:10.1093/bioinformatics/btq062.
- Liu Y, Jia Y, Wang W, Chen ZD, Davis EC, Qiu YJ. 2008. Phylogenetic relationships of two endemic genera from East Asia, *Trichocoleopsis* and *Neotrichocolea* (Hepaticae). *Annals of the Missouri Botanical Garden* **95**: 459–470.
- Liu Y, Cox CJ, Wang W, Bernard G. 2014. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* **6**: 862–878.
- Liu Y, Johnson MG, Cox CJ, et al. 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. *Nature Communication* **10**: 1485.
- Meleshko O, Martin MD, Korneliusen 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. *Molecular Biology and Evolution* **38**: 2750–2766.
- Meng K, Chen S, Xu K, et al. 2021. Phylogenomic analyses based on genome-skimming data reveal cyto-nuclear discordance in the evolutionary history of *Cotoneaster* (Rosaceae). *Molecular Phylogenetics and Evolution* **158**: 107083. doi:10.1016/j.ympev.2021.107083.
- Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. 2014. Astral: genome-scale coalescent-based species tree estimation. *Bioinformatics* **30**: i541–i548. doi:10.1093/bioinformatics/btu462.
- Morales-Briones DF, Liston A, Tank DC. 2018. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the neotropical genus *Lachemilla* (Rosaceae). *New Phytologist* **218**: 1668–1684.
- Morris JL, Puttick MN, Clark JW, et al. 2018. The timescale of early land plant evolution. *Proceedings of the National Academy of Sciences, USA* **115**: E2274–E2283.
- Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* **11**: 715–724. doi:10.1093/oxfordjournals.molbev.a040152.
- One Thousand Plant Transcriptomes Initiative. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* **574**: 679–685.
- Page R. 1994. Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. *Systematic Biology* **43**: 58–77.
- Patel N, Medina R, Johnson M, Goffinet B. 2021. Karyotypic diversity and cryptic speciation: have we vastly underestimated moss species diversity? *Bryophyte Diversity and Evolution* **43**: 150–163.
- Pond SL, Kosakovsky F, Simon DW, Muse SV. 2005. Hyphy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.
- Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Molecular Biology and Evolution* **26**: 1641–1650. doi:10.1093/molbev/msp077.
- Puttick MN, Morris JL, Williams TA, et al. 2018. The interrelationships of land plants and the nature of the ancestral embryophyte. *Current Biology* **28**: 733–745.
- Quang MB, Schmidt HA, Olga C, et al. 2020. IQTREE2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**: 1530–1534.
- Rambaut A. 2014. FigTree 1.4.2. <http://tree.bio.ed.ac.uk/>.
- Read DJ, Duckett JG, Francis R, Ligron R, Russell A. 2000. Symbiotic fungal associations in ‘lower’ land plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **355**: 815–830. doi:10.1098/rstb.2000.0617.
- Ren R, Wang H, Guo C, et al. 2018. Widespread whole genome duplications contribute to genome complexity and species diversity in angiosperms. *Molecular Plant* **11**: 414–428. doi:10.1016/j.molp.2018.01.002.
- Richardson AO, Rice DW, Young GJ, Alverson AJ, Palmer JD. 2013. The ‘fossilized’ mitochondrial genome of *Liriodendron tulipifera*: ancestral gene content and order, ancestral editing sites, and extraordinarily low mutation rate. *BMC Biology* **11**: 29.
- Sayyari E, Mirarab S. 2016. Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution* **33**: 1654–1668. doi:10.1093/molbev/msw079.
- Sayyari E, Whitfield JB, Mirarab S. 2018. DiscoVista: interpretable visualizations of gene tree discordance. *Molecular Phylogenetics and Evolution* **122**: 110–115. doi:10.1016/j.ympev.2018.01.019.
- Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. *Current Opinion in Plant Biology* **15**: 147–153. doi:10.1016/j.pbi.2012.03.011.
- Shen W, Le S, Li Y, Hu F, Quan Z. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* **11**: e0163.
- Silva AT, Gao B, Fisher KM, et al. 2021. To dry perchance to live: insights from the genome of the desiccation-tolerant biocrust moss *Syntrichia caninervis*. *The Plant Journal* **105**: 1339–1356.
- Smith S, Moore M, Brown JW, Yang Y. 2015. Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology* **15**: 150.
- Söderström L, Hagborg A, von Konrat M, et al. 2016. World checklist of hornworts and liverworts. *Phytokeys* **59**: 1–828.
- Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688–2690. doi:10.1093/bioinformatics/btl446.
- Su D, Yang L, Shi X, et al. 2021. Large-scale phylogenomic analyses reveal the monophyly of bryophytes and Neoproterozoic origin of land plants. *Molecular Biology and Evolution* **38**: 3332–3344. doi:10.1093/molbev/msab106.
- Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**: W609–W612. doi:10.1093/nar/gkl315.
- Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* **56**: 564–577. doi:10.1080/10635150701472164.
- Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinformatics* **9**: 322. doi:10.1186/1471-2105-9-322.
- Tofigh A, Hallett M, Lagergren J. 2011. Simultaneous identification of duplications and lateral gene transfers. *IEEE/ACM Transactions on Computational Biology and Bioinformatics* **8**: 517–535. doi:10.1109/TCBB.2010.14.
- Van de Peer Y, Mizrahi E, Marchal K. 2017. The evolutionary significance of polyploidy. *Nature Reviews. Genetics* **18**: 411–424. doi:10.1038/nrg.2017.26.
- Villarreal AJ, Crandall-Stotler BJ, Hart ML, Long DG, Forrest LL. 2016. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytologist* **209**: 1734–1746.
- Vriesendorp B, Bakker FT. 2005. Reconstructing patterns of reticulate evolution in angiosperms: what can we do? *Taxon* **54**: 593–604. doi:10.2307/25065417.
- Wakeley J. 2009. Coalescent theory: an introduction. *Systematic Biology* **58**: 162–165.
- Wang K, Lenstra JA, Liang L, Hu Q, Liu J. 2018. Incomplete lineage sorting rather than hybridization explains the inconsistent phylogeny of the wisent. *Communications Biology* **1**: 169.
- Wu S, Han B, Jiao Y. 2019. Genetic contribution of paleopolyploidy to adaptive evolution in angiosperms. *Molecular Plant* **13**: 59–71.
- Wu YC, Rasmussen MD, Bansal MS, Kellis M. 2014. Most parsimonious reconciliation in the presence of gene duplication, loss, and deep coalescence using labeled coalescent trees. *Genome Research* **24**: 475–486. doi:10.1101/gr.161968.113.
- Yang Y, Moore M, Brockington S, et al. 2018. Improved transcriptome sampling pinpoints 26 ancient and more recent polyploidy events in Caryophyllales, including two allopolyploidy events. *New Phytologist* **217**: 855–870.
- Yang Y, Sun P, Lv L, et al. 2020. Prickly waterlily and rigid hornwort genomes shed light on early angiosperm evolution. *Nature Plants* **6**: 215–222. doi:10.1038/s41477-020-0594-6.
- Yang Z. 2007. PAML4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution* **24**: 1586–1591. doi:10.1093/molbev/msm088.
- Yu Y, Yang J, Ma W, et al. 2019. Chloroplast phylogenomics of liverworts: a reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. *Cladistics* **36**: 184–193. doi:10.1111/cla.12396.
- Zhang J, Fu XX, Li RQ, et al. 2020. The hornwort genome and early land plant evolution. *Nature Plants* **6**: 107–118. doi:10.1038/s41477-019-0588-4.
- Zwaenepoel A, Van de Peer Y. 2018. WGD – simple command line tools for the analysis of ancient whole genome duplications. *Bioinformatics* **35**: 2153–2155. doi:10.1093/bioinformatics/bty915.

