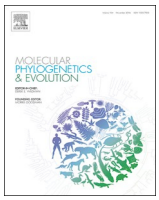
Molecular Phylogenetics and Evolution 154 (2021) 106965

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev



High gene space divergence contrasts with frozen vegetative architecture in the moss family Funariaceae

Nasim Rahmatpour[a], Neranjan V. Perera[b], Vijender Singh[b], Jill L. Wegrzyn[a,b], Bernard Goffinet[a,*]

[a] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
[b] Institute for Systems Genomics, Computational Biology Core, University of Connecticut, Storrs, CT, USA

ARTICLE INFO

Keywords:
Mosses
Genome evolution
Genome divergence
Transcriptome
Funaria hygrometrica
Physcomitrium (Physcomitrella) patens
Physcomitrium pyriforme
Funariaceae

ABSTRACT

A new paradigm has slowly emerged regarding the diversification of bryophytes, with inferences from molecular data highlighting a dynamic evolution of their genome. However, comparative studies of expressed genes among closely related taxa is so far missing. Here we contrast the dimensions of the vegetative transcriptome of *Funaria hygrometrica* and *Physcomitrium pyriforme* against the genome of their relative, *Physcomitrium (Physcomitrella) patens*. These three species of Funariaceae share highly conserved vegetative bodies, and are partially sympatric, growing on mineral soil in mostly temperate regions. We analyzed the vegetative gametophytic transcriptome of *F. hygrometrica* and *P. pyriforme* and mapped short reads, transcripts, and proteins to the genome and gene space of *P. patens*. Only about half of the transcripts of *F. hygrometrica* map to their ortholog in *P. patens*, whereas at least 90% of those of *P. pyriforme* align to loci in *P. patens*. Such divergence is unexpected given the high morphological similarity of the gametophyte but reflects the estimated times of divergence of *F. hygrometrica* and *P. pyriforme* from *P. patens*, namely 55 and 20 mya, respectively. The newly sampled transcriptomes bear signatures of at least one, rather ancient, whole genome duplication (WGD), which may be shared with one reported for *P. patens*. The transcriptomes of *F. hygrometrica* and *P. pyriforme* reveal significant contractions or expansions of different gene families. While transcriptomes offer only an incomplete estimate of the gene space, the high number of transcripts obtained suggest a significant divergence in gene sequences, and gene number among the three species, indicative of a rather strong, dynamic genome evolution, shaped in part by whole, partial or localized genome duplication. The gene ontology of their specific and rapidly-evolving protein families, suggests that the evolution of the Funariaceae may have been driven by the diversification of metabolic genes that may optimize the adaptations to environmental conditions, a hypothesis well in line with ecological patterns in the genetic diversity and structure in seed plants.

1. Introduction

Bryophytes, i.e., liverworts, hornworts and mosses, are the only extant land plants with a life cycle dominated by the haploid generation, the gametophyte. Although comprising as many as 20,000 species today (Vanderpoorten and Goffinet, 2009), they have been regarded as evolutionary dead-ends (Crum, 1972) owing to their haploid vegetative stage, considered highly vulnerable to deleterious mutations, their propensity to maintaining morphological cohesiveness over broad geographic ranges and their putative old age. The first inferences based on molecular data quickly revealed, however, a high level of genetic diversity (Cummins and Wyatt, 1981; Wyatt et al., 1989; Såstad et al., 1999; Stenøien and Såstad, 1999; Shaw, 2001). Glime (1990) suggested that the diversification of bryophytes was perhaps driven by ecophysiological adaptations, in particular those pertaining to optimizations of the parameters (Stark, 2017) shaping their generalized poikilohydry and dehydration or desiccation tolerance (Proctor et al., 2007). Bryophytes lack an effective water absorbing rooting system and mosses also lack compensatory mycorrhizal associations that could supply water from the substrate (Field et al., 2015). Thus, to regulate their internal water content bryophytes in general, and mosses in particular, depend on atmospheric moisture uptake through the surface of their vegetative body, which is covered by a thin, water permeable cuticle (Jeffree, 2006). Their ability to withstand periodic dehydration or desiccation relies on repair mechanisms (Wood, 2007; Fisher, 2008) whose modifications may be shaping speciation in bryophytes.

In the last decade, bryophytes have emerged as lineages that are anything but genetically depauperate or evolutionary static: (a) the

* Corresponding author.
E-mail address: bernard.goffinet@uconn.edu (B. Goffinet).

https://doi.org/10.1016/j.ympev.2020.106965
Received 10 January 2020; Received in revised form 13 September 2020; Accepted 14 September 2020
Available online 19 September 2020
1055-7903/

genome of *Physcomitrella* (now and hereafter *Physcomitrium patens* fide Medina et al., 2019) comprises 32,926 genes, and thus more genes than the flowering plant *Arabidopsis thaliana* (Litt, 2013); (b) bryophytes have undergone rapid radiations following changes in their biotic environment (e.g., epiphytic liverworts [Feldberg et al., 2014] or large climatic shifts (e.g., *Sphagnum*, Shaw et al., 2010); (c) transcriptomic or genomic data for any bryophyte species studied so far harbor signatures of large or whole nuclear genome duplications (Rensing et al., 2007; Shaw et al., 2008; Beike et al., 2014; Szövényi et al., 2015; Devos et al., 2016; Johnson et al., 2016; Lang et al., 2018) that may provide new templates for innovations such that the gene space is highly dynamic (Johnson et al., 2016); (d) patterns of moss or liverwort diversity through time reflect repeated burst of diversifications (Laenen et al., 2014); (e) hybridization, once thought to play no role in the diversification of mosses (Vitt, 1971) is now emerging as a widespread (Natcheva and Cronberg, 2004) and hence potentially important mechanism of speciation (Shaw, 2009); and finally, (f) the rate of nuclear non-synonymous substitutions relative to those in the organellar genomes in much higher in mosses than in lineages of vascular plants (Liu et al., 2019). Thus, the relatively young age of bryophyte species, their high diversity distributed across a broad ecological spectrum suggest a rather highly dynamic evolutionary history. Szövényi et al. (2011), as part of their selection of genes for the comparison in expression level between *F. hygrometrica* Hedw. gametophytes and sporophytes, reported that only about 30% of the 454 reads could be mapped unambiguously against the *P. patens* gene models or genome. Although they did consider the two species closely related they did not discuss the unexpected degree of divergence further. To gain deeper insight into the genomic divergence between moss species from a single family, here the Funariaceae, we contrasted the vegetative transcriptome of *F. hygrometrica* and *P. pyriforme* to the gene space inferred from the entire genome of *P. patens*.

The Funariaceae have long served as model systems in evolutionary bryology, starting with the pioneering studies by von Wettstein (1924) in reproductive isolation, and trait inheritance, to the emergence of *P. patens* as a model taxon in evolutionary developmental biology of land plants (Cove et al., 2009; Prigge & Bezanilla, 2010; Mallett et al., 2019; Rensing et al., 2020). The cosmopolitan family comprises perhaps 200 species, all of which are annuals growing on soil, in seasonally wet or moist regions. Recent estimates of divergence times within the moss family Funariaceae suggest that the family originated approximately 90 mya, with the Funarioideae, which comprise *F. hygrometrica* and *P. pyriforme*, diverging about 58 mya (Medina et al., 2018). During this long period of diversification, the architecture of the vegetative generation remained virtually unchanged, with the gametophyte composed of a single, short, typically unbranched axis, with cells weakly differentiated across the leaf lamina, except for those composing the median midrib. The gametophyte is so strongly conserved that vegetative plants of genera that diverged about 60 mya are virtually identical (Fig. 1). By contrast, the spore bearing generation, in particular the length of the stalk subtending the sporangium, the mode of dehiscence of the sporangium and the development of teeth lining the sporangial mouth following dehiscence and regulating spore release have undergone dramatic transformations (Fife, 1985). The spectrum of morphological complexity of the sporophyte, shaped by differential expression of orthologous genes and some species-specific genes (Kirbis et al., 2020), spans from the most complex in *F. hygrometrica* to the most reduced in *P. patens*, with *P. pyriforme* exhibiting an intermediate architecture.

Here we contrast the vegetative transcriptome of *F. hygrometrica* to the gene space of *P. patens*. Given the high divergence observed, and the possibility that it may be an artifact of a transcriptome to genome comparison, we validated our observations by confirming the identity of the *F. hygrometrica* through mapping against its draft genome, by generating a vegetative transcriptome for *P. pyriforme*, which we then compared to the transcriptome of *F. hygrometrica* and the genome of *P. patens*.

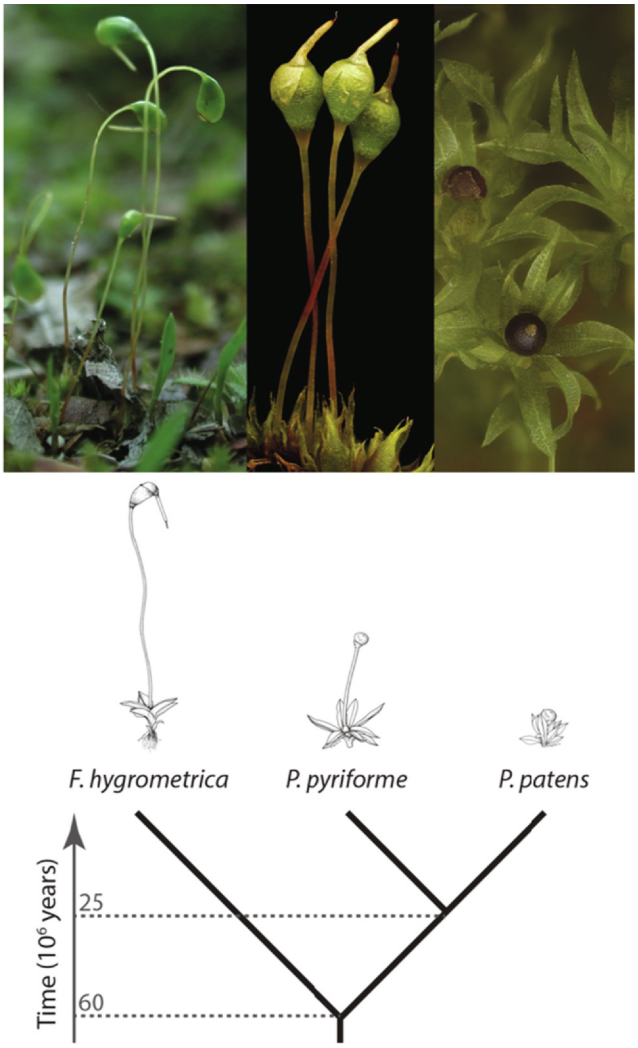


Fig. 1. The three species of Funariaceae included in the study (i.e., from left to right, *Funaria hygrometrica*, *Physcomitrium pyriforme* and *Physcomitrium* [*Physcomitrella*] *patens*), their relationship and estimated time of divergence fide Medina et al. (2018).

2. Material and methods

2.1. Culturing

Cultures were established from spores sampled from one population of *F. hygrometrica* (*Budke* 145, CONN) and *P. pyriforme* from Connecticut (*Goffinet* 9276, CONN). Operculate capsules were sterilized in a 1% bleach (NaClO) solution for two minutes and rinsed in distilled water. Spores from individual capsules were sown on Knop's medium (Collier and Hughes, 1982) and allowed to germinate in sealed petri dishes. Buds developing on filamentous protonema were isolated and transferred to separate plates to provide genetically identical populations. The gametophytes from one plate were split among several plates to propagate and generate sufficient isogenic replicates. Cultures were kept in a growth chamber for approximately two months at 20–25 °C with 14–16 h of light (50–70 μmol s^{-1} m^{-2}). Gametophytes were transferred to a rich sandy loam soil mix in PlantCon containers (MP Biomedicals, Solon, OH, USA) and allowed to grow for three months under the same light and temperature regimes.

2.2. RNA extraction and library preparation

Three replicate cultures of *F. hygrometrica* and *P. pyriforme* were

selected (i.e., G1, G2 & G3) (stems, leaves and rhizoids, but no gametangia) were harvested for RNA extraction. The tissue was homogenized in liquid nitrogen, and RNA extracted following the RNAzol® RT protocol (Chomczynski et al., 2010). The six libraries were prepared with the TruSeq Stranded mRNA kit and sequenced on an Illumina NextSeq 500 platform on a single lane (100 bp PE mid-output) at the Center for Genome Innovation (CGI) at University of Connecticut. The Illumina short reads are submitted to NCBI (SRA accession: PRJNA421369).

2.3. Quality control, assembly, and annotation

The quality of the Illumina paired-end reads was assessed by FastQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Each quality-controlled library was assembled with the *de novo* transcriptome assembler, TRINITY v.2.0.6 (minimum contig length set to 300) (Grabherr et al., 2011). RSEM v.1.3.0 (Li and Dewey, 2011) was used to quantify transcriptome assemblies. Only assembled contigs with FPKM (fragments per kilobase of transcripts per million mapped reads) $>$ 0.5 were retained. All three filtered assemblies were clustered with USEARCH v8.1.1861 (Edgar, 2010) at 90% to construct a single non-redundant reference. TRANSDECODER v.3.0.0 (https://transdecoder.github.io/) was used to identify and trim predicted Open Reading Frames (ORFs). The frame-selected genes were provided along with translated protein sequences. The process of frame selection predicts the most optimal reading frame across the assembled transcripts, removes UTR, and translates the sequences for the longest open reading frame. In the event that an open reading frame of minimum length cannot be observed, the transcript was removed. Functional annotations of the *de novo* assembled transcripts were obtained using homology searches via EnTAP v.0.8.0 pipeline (Hart et al., 2019). The internal transcripts (transcripts without stop and start codon) were removed from the final reference and the remaining transcripts were compared to the National Center for Biotechnology Information's non-redundant protein database (E-value: 1.0E-5) Query and target coverage were set to 80% and 60%, respectively. Alignment to amoeba, bacteria, fungi and insects were applied as contaminant filters. Gene ontology (GO) terms were integrated based upon gene family assignment via EggNOG mapper (Blake et al., 2015; Huerta-Cepas et al., 2015). To assess the transcriptome assembly and annotation completeness, the reference transcriptome was compared with the BUSCO (Benchmarking Universal Single-Copy Orthologs) viridiplantae_odb10 database of 430 ortholog proteins, which includes *P. patens* as the only bryophyte (Simão et al., 2015).

After internal and contaminant transcript removal, the remaining transcripts of *F. hygrometrica* was mapped to the draft genome (unpublished data) to validate the transcriptome quality.

2.4. Alignments to P. patens genome

Funaria hygrometrica and *P. pyriforme* transcriptomic short reads were aligned to the *P. patens* genome (v3.3 accessed from Phytozome 12) (https://phytozome.jgi.doe.gov/pz/portal.html) using the STAR v.2.5.1b (splice-aware aligner) (Dobin et al., 2013). In addition, the translated and filtered transcriptome was aligned to the same genome via the protein to genome aligner EXONERATE v.2.4.0 (alignment score = 500, percentage = 70, minimum intron length = 10) (Slater and Birney, 2005).

2.5. Analysis of whole genome duplication

The likelihood of whole genome duplication was estimated based on the distribution of Ks or synonymous substitution per synonymous site for pairs of paralogs. Paralogous genes in each three species were identified via USEARCH (Edgar, 2010). To identify paralogous genes, the minimum identity threshold was set to 80% at the nucleotide level. An in-house Python script parsed the two longest nucleotide sequences from each paralog set and aligned them via MUSCLE (Edgar, 2004) (The scripts associated with this paper are available at https://gitlab.com/NasimR/funaria_gametophyte_transcriptome). The quality of sequence alignments was investigated using BELVU (Barson and Griffiths, 2016) and passing alignments (score $>$ 3.5) were selected for calculating Ks value with CODEML program in PAML (Yang, 1997). In order to estimate the relative timing of whole genome duplications relative to the divergence of *F. hygrometrica* and *P. patens*, the distribution of Ks values of paralogs pairs in each species was compared to the Ks value of orthologs pairs shared by *Funaria-Physcomitrium*. Orthologous pairs were identified by clustering the translated sequences via ORTHOFINDER v.1.1.4 (Li et al., 2003). The longest single-copy ortholog (SCO) and best paired alignments were selected from each orthogroup and the sequences were aligned via MUSCLE (Edgar, 2004). The quality of sequence alignments was investigated using BELVU (Barson and Griffiths, 2016) and passing alignments (score $>$ 3.5) were selected for the Ka/Ks analysis with CODEML program in PAML (Yang, 1997).

2.6. Gene families

To assess the distribution of *F. hygrometrica* transcripts among families of orthologous genes present in *P. pyriforme* and *P. patens*, first we compared only these three Funariaceae proteins; *F. hygrometrica*, *P. pyriforme* (internal transcripts and contaminants excluded) and *P. patens* (v3.3) using ORTHOFINDER v.1.1.4 (Li et al., 2003). A secondary analysis included other, most distant mosses. The moss proteins available for this comparison are these three Funariaceae proteins, *Ceratodon purpureus* (Szövényi et al., 2015), 48 taxa spanning the Bryopsida (Johnson et al., 2016; and Johnson *et al.* unpublished), and three *Sphagnum* species downloaded from the 1KP project (Matasci et al., 2014). All transcriptomes were frame selected and processed as described above. Species trees were generated within ORTHOFINDER and analyzed to confirm results of the analysis.

2.7. Rate of changes in gene family size

The branch lengths of the species tree from ORTHOFINDER (included 52 mosses) were multiplied by 1000, then the tree was converted to ultrametric via the ETE3 package in Python for use in CAFE analysis. The species tree and gene family table generated via ORTHOFINDER were used to estimate the rate of changes in gene family size via CAFE 4.1 (Han et al., 2013). CAFE can calculate the number of gene families with expansion, contraction and no change along each branch and also for each node in the tree. A single λ (the birth and death parameter) was calculated for the whole tree. The default 0.01 *P*-value cutoff was used to define the rapidly evolving gene families. The functional annotations of rapidly-evolving gene families and GO level comparisons was extracted from the EnTAP results and the R package GoSeq, respectively (Young et al., 2010).

3. Results

3.1. The vegetative transcriptome of F. hygrometrica

Following quality assessment, *de novo* assembly of short reads via TRINITY generated 28,350 (G1), 39,498 (G2) and 42,167 (G3) transcripts for these replicates of *F. hygrometrica* (Table 1). Filtering based on expression (alignment of short reads) reduced the independent assemblies to 26,920 (G1), 37,095 (G2) and 39,028 (G3) transcripts. The unique transcripts represented in the clustered reference (G1 + G2 + G3) totaled 63,967 (Table 1). In the process of frame selection, UTRs and sequences not matching the minimum length criterion were removed, which reduced the number of transcripts to 53,283 (Table 1) with an average length and N50 of 829 and 1062, respectively. The retained transcripts were classified into complete, partial and internal based on identified 5′ and 3′ positions (23,619

Table 1
Statistics of three replicate transcriptomes (G1, G2, G3) of *F. hygrometrica* and *P. pyriforme* gametophytes.

	F. hygrometrica			*P. pyriforme*		
Statistics	(G1)	(G2)	(G3)	(G1)	(G2)	(G3)
Illumina 100 bp PE reads	49,324,732	34,002,818	37,100,201	12,911798	13,547,068	12,406,247
TRINITY assembly	28,350	39,498	42,167	46,386	43,275	48,784
RSEM filtering	26,920	37,095	39,028	45,301	42,492	47,599
USEARCH clustering	63,967			83,430		
TRANSDECODER	53,283			70,040		
	23,619 complete, 12,152 internal, 17,512 partial			41,659 complete, 9411 internal, 18,970 partial		
EnTAP annotation	23,745 unique hits, 17,386 no hits, 258 contaminants			41,627 unique hits, 19,002 no hits, 220 contaminants		
BUSCO	358/430 = 83.25%			383/430 = 89.06%		

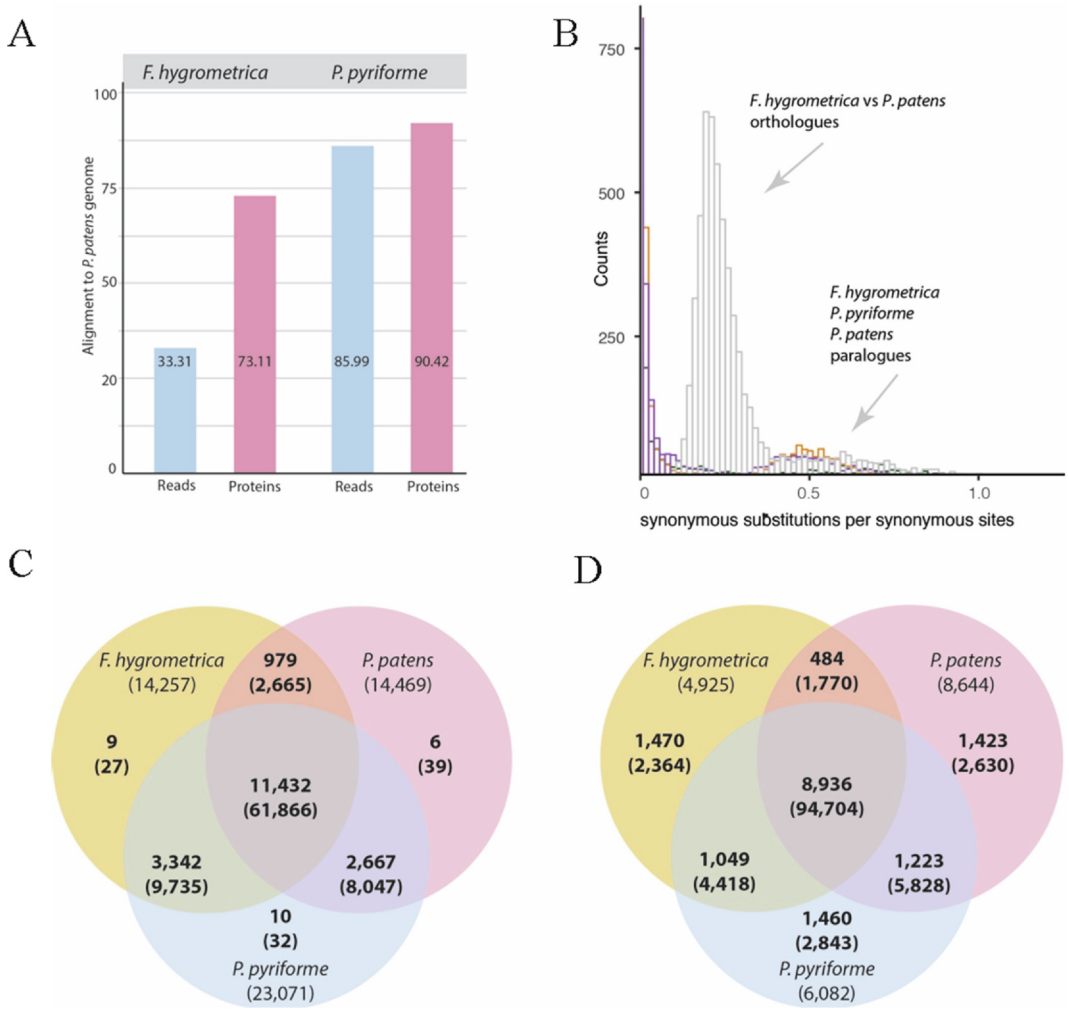


Fig. 2. A. Alignment of *Funaria hygrometrica* and *Physcomitrium pyriforme* vs. *P. patens* genome based on reads (via STAR) and translated transcriptome (proteins) (500 alignment score and 70 percent via EXONERATE). *F. hygrometrica* contains 40,873 and *P. pyriforme* contains 60,409 number of proteins after internal transcripts and contamination removal. B. Synonymous substitution per synonymous site (Ks) frequency plot for paralogous gene pairs of three species. The Ks frequency plots are drawn from 1563, 2180 and 2369 paralogs for the three replicates of *F. hygrometrica*, 3257, 2893 and 3471 from the three replicates of *P. pyriforme* and 6411 of *P. patens*. B. Ks distribution of paralogous gene pairs superimposed with *F. hygrometrica-P. patens* orthologous gene pairs. C. Protein family overlap among three Funariaceae species based on ORTHOFINDER run on these three species. D. Protein family overlap among three Funariaceae species extracted from ORTHFINDER run on 52 mosses. *(numbers in parentheses are the number of proteins in protein families, the numbers below the names refer to unassigned proteins).

complete, 12,152 internal, and 17,512 partial; Tables 1 & S1). Following the removal of the internal transcripts, the complete reference contained 41,131 sequences (Table S2). Functional annotation of the translated reference transcriptome revealed that nearly 57% (23,487) had unique hits to the non-redundant protein database. Of the remaining, 258 (1.09%) were recognized as contaminants, with 116 of these belonging to fungi, 110 to bacteria, 15 to amoeba and 17 to insects. Assigning proteins to gene families using EggNOG, provided an annotation for 9237 transcripts previously not aligned to any protein, resulting in a final annotation rate of 80% (i.e., 32,982 of 41,131)

(Table S2). Of the annotated transcripts, 16,045 aligned to proteins of *P. patens*. Of the 32,982 annotated transcripts, 19,449 (59%) were associated with at least one Gene Ontology (GO) term. The present vegetative transcriptome of *F. hygrometrica* of after internal and contamination removal contains 40,873 transcripts comprises an estimated 83.25% (complete) to 90.46% (including fragmented) of the green plant universal single copy orthologs (Tables 1 & S3).

3.2. The P. pyriforme vegetative transcriptome

The assembly of *P. pyriforme* generated 46,386, 43,275 and 48,784 transcripts for G1, G2 and G3, respectively (Table 1). Sequencing depth for *P. pyriforme* samples were lower (2–4 fold less) than *F. hygrometrica*'s since they were sequenced on a single lane with its sporophyte tissues (Table 1). RSEM filtering reduced number of the transcripts to 45,301 (G1), 42,492 (G2) and 47,599 (G3; Table 1). The number of unique transcripts after clustering was 83,430 (Table 1). This number reduced to 70,040 after frame selection (Table 1) with an average length and N50 of 955 and 1260, respectively. The final single reference was classified into 41,659 complete, 9411 internal and 18,970 partial transcripts (Tables 1 & S1). Following the removal of internal transcripts, the complete reference contained 60,629 sequences (Table S2). Nearly 68.65% (41,627) had unique hits and 19,002 had no hit to the non-redundant protein database. A small fraction of transcripts (i.e., 0.53%, or 220 transcripts) were recognized as contaminants with 61 of these attributed to fungi, 146 to bacteria, three to amoeba and 10 to insects. EggNOG provided annotations for an additional 9720 transcripts, resulting in 51,347 annotated transcripts (84%). Of the total annotated transcripts, 34,272 aligned to the proteins of *P. patens* (Table S2), and 30,046 could be characterized with at least one GO term. The remaining 9282 were not characterized in current databases. The vegetative transcriptome of *P. pyriforme* after contamination and internal removal (60,409 transcripts) comprises an estimated 89.06% (complete) to 94.65% (including fragmented) of the green plant universal single copy orthologs (Tables 1 & S2).

3.3. Mapping of transcripts against genomic references

3.3.1. F. hygrometrica mapped against itself and P. patens

Following filtering of the transcriptome, 33.31% of the original Illumina short reads of *F. hygrometrica* mapped to the *P. patens* genome. The rate of protein alignment was 73.11%, and thus 10,988 proteins did not align to the genome (Fig. 2A). Finally, 94.79% of *F. hygrometrica* transcripts aligned to the draft *F. hygrometrica* genome.

To examine species-specific proteins in *F. hygrometrica*, we combined the protein alignments and an independent gene family assessment. About 43% (10,988) of the translated transcripts or proteins failed to align (via EXONERATE). Of the 10,988, gene family assessment indicated that 7107 do not form protein families with *P. patens* (Fig. 3A). To verify that those are encoded from real genes and not the result of assembly artifacts, we aligned and identified 6398 in the draft genome of *F. hygrometrica* (using 98% identity and 95% coverage via GMAP). Among these *F. hygrometrica* specific proteins, 5692 could not be characterized through sequence similarly search against any species in Genbank (NCBI nr database). The remaining 706 annotated either through similarity search or gene family assignment primarily were derived from *P. patens* (51.22%) (Fig. 3A).

3.3.2. P. pyriforme mapped against P. patens

On average, 85.99% of the short reads of the three libraries mapped to the *P. patens* genome**.** Contrasting the proteins of *P. pyriforme* to the genome of *P. patens* reveals a much higher fraction of protein mapping (90.42%) than observed for *F. hygrometrica* (Fig. 2A), Of the 5787 proteins that did not align to the *P. patens* genome, 3936 of those were not able to form a gene family with any of *P. patens* proteins. Of those, 3454 proteins were not characterized via sequence similarity search to any species in Genbank (NCBI nr database). The remaining 482 annotated either through similarity search or gene family assignment primarily were derived from *P. patens* (26.39%) (Fig. 3B).

3.4. Transcriptomic signatures of whole genome duplications in F. hygrometrica and P. pyriforme

Patterns in the distribution of Ks between intragenomic paralogs are congruent with both transcriptomes having undergone a whole genome duplication (Fig. 2B). Superimposing these distributions over the distribution of Ks between interspecific orthologs suggests that the divergence between paralogs is greater than between orthologs.

3.5. Gene family evolution

Protein family determination based on only three Funariaceae species identified 15,762, 17,551 and 15,084 gene families for *F. hygrometrica*, *P. pyriforme* and *P. patens*, respectively and 14,257, 23,071, 14,469 unassigned proteins for each respectively, which were not included in any determined protein families. (Fig. 2C). Furthermore six, ten and nine protein families were identified specifically in *F. hygrometrica*, *P. pyriforme* and *P. patens*, respectively (Fig. 2C).

Assessing shared ancestry among proteins and including more moss transcriptomes, provided more accurate protein family determination and decreased the number of unassigned proteins (Fig. 2D, Fig. S4). The gene family overlap between three Funariaceae species extracted based on this protein family determination, indicates *F. hygrometrica* and *P. patens* overlapped in 9420 orthogroups, *F. hygrometrica* and *P. pyriforme* overlapped in 9985 orthogroups, and *P. patens* and *P. pyriforme* overlapped in 10,159 orthogroups (Fig. 2D, Fig. S4).

3.6. Rate of changes in gene family size

The terminal branch of *F. hygrometrica*, is characterized by 5271 and 5626 expanded and contracted gene families, respectively. Among those, 24 of gene families are rapidly evolving (statistically significant), either through rapid expansion (21) or contraction (3). The terminal branch leading to *P. pyriforme* is characterized by 8131 and 2131 expanded and contracted gene families, respectively. Among those, 629 are rapidly evolving either through expansion (6 1 8) or contraction (11). The terminal branch leading to *P. patens* is characterized by 1005 and 8273 expanded and contracted gene families, respectively. Among those, 151 families underwent rapid evolution either through expansion (45) or contraction (106; Fig. S1).

4. Discussion

Mosses compose a diverse lineage that arose early in the diversification of land plants (Clarke et al, 2011; Cardona-Correa et al., 2016; Morris et al., 2018; Leebens-Mack et al., 2019) and that underwent repeated bursts of radiation that ultimately yielded an estimated diversity of 13,000 extant species (Laenen et al., 2014). Long considered evolutionary static, mosses have emerged as dynamic evolutionary entities (e.g., Medina et al., 2018). Although the genetic underpinning of moss development is increasingly being unraveled (e.g., Harrison et al., 2009; Xiao et al. 2011; Vidali and Bezanilla, 2012; Landberg et al., 2013; Chater et al., 2016; Ortiz-Ramírez et al., 2016; Kofuji et al., 2018; Coudert et al., 2019), virtually nothing is known in terms of comparative developmental genetics within mosses (Goffinet and Buck, 2013), except for the recent study by Kirbis et al. (2020). Phylogenetic inferences suggest that conspicuous morphological transformations may obfuscate a shared ancestry (Goffinet and Shaw, 2002) and that morphological similarity can be a poor predictor of uniquely shared evolutionary ancestry (e.g., Goffinet et al., 2007). The overall simplicity of the vegetative body is also a poor predictor of genomic architecture and size of the gene space, a discrepancy revealed by the higher gene

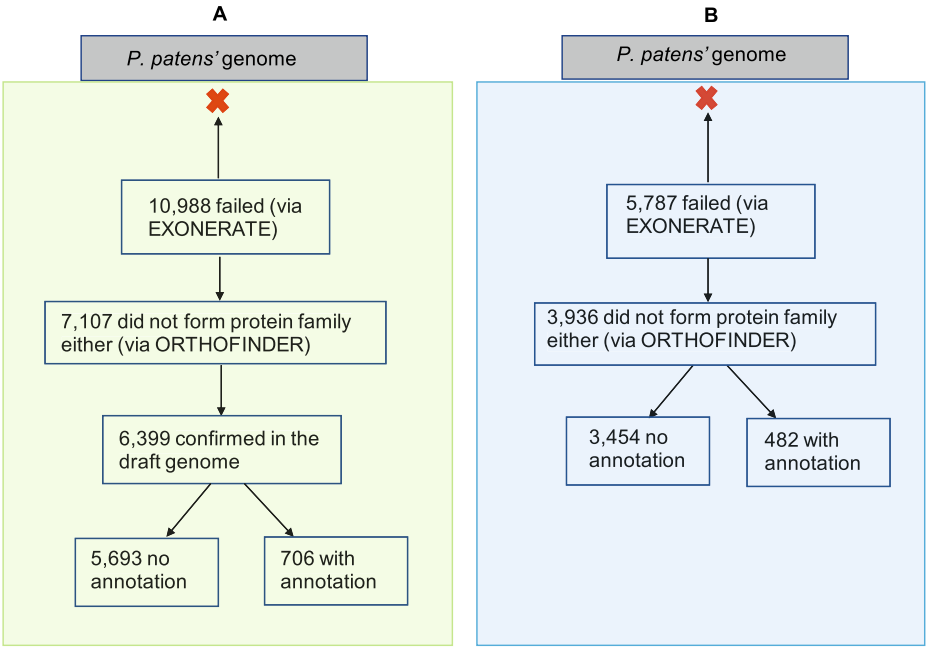


Fig. 3. Workflow for finding novel genes in *F, hygrometrica* and *P. pyriforme* transcriptomes by comparative analysis to *P. patens* genome and gene space through different tools. A. *F. hygrometrica* vs. *P. patens*. B. *P. pyriforme* vs. *P. patens*.

content of the moss *P. patens* (i.e., 32,926 coding genes) compared to the flowering plant *Arabidopsis thaliana* (i.e., 27,655 coding genes; https://phytozome.jgi.doe.gov/pz/portal.html). Considering the generalized poikilohydry of mosses (Proctor and Tuba, 2002; Wood, 2007; Gao et al., 2015; Gao et al., 2018), such putative diversity of the genomic tool box may allow mosses to modify and optimize their ecophysiology along abiotic ecological gradients (Glime, 1990), and diversify within a broader ecological space (e.g., Shaw, 1985). Szövényi et al. (2011) had previously reported that only about a third of the 454 reads generated from the haploid and diploid generations of *F. hygrometrica* could be mapped unambiguously against the *P. patens* genome. Following a validation of our transcriptome against the draft genome of *F. hygrometrica* we confirm this high level of differentiation against *P. patens*, as well as against its close relative *P. pyriforme*, and reveal significant variation in gene family sizes. Such degree of genomic divergence between mosses sharing virtually identical vegetative bodies strengthens the emerging paradigm that bryophytes are evolutionarily highly dynamic lineages, with much of the signatures of diversification hidden below the morphological surface.

4.1. Vegetative transcriptomes of F. hygrometrica and P. pyriforme

The vegetative (i.e., gametophytic) transcriptome of *F. hygrometrica* comprises 40,873 transcripts, with an estimated 83.25% (complete) to 90.46% (including fragmented) of the set of green plants universal single copy orthologs (Table 1 & S3). Compared to the total proteome space of *P. patens* (i.e., 86,669 protein-coding transcripts based on 32,926 genes), this likely offers a representative and significant capture of the transcriptome of *F. hygrometrica.* The transcriptome of the vegetative plants of *P. pyriforme* (i.e., 60,409 transcripts with at least 89.06% complete transcripts of the green plants universal single copy orthologs, Table 1 & S3) is equally of high quality. Although, transcriptomes typically represent a slightly inflated value of the unique gene space, statistics regarding the completeness, functional annotation, and assembly itself provide support for the high quality reference. Filters applied to reduce fragmentation, assembly artifacts, and contamination provided an improvement on the base assemblies.

4.2. Mapping rates of transcriptomes against the P. patens and draft F. hygrometrica genomes

While only 33.31% of the raw reads of *F. hygrometrica* mapped against the *P. patens* genome about 73.11% of the putative proteins (after excluding transcripts with lower expression [FPKM $<$ 0.5] and internal and high confidence contaminants) aligned against the *P. patens* genome (Fig. 2A). These rates are lower than those for reads and proteins of *P. pyriforme* (i.e., 85.99%, and 90.42%, respectively; Fig. 2A) and this pattern likely reflects the congeneric relationship of *P. pyriforme* and *P. patens* (Medina et al., 2019) and hence also the more recent divergence, with a most recent common ancestry estimated at 23 versus 58 mya with *F. hygrometrica* (Medina et al., 2018). Nearly 95% of proteins of *F. hygrometrica* aligned to the draft genome of the same species (Szövényi et al. unpublished data) based on a 98% identity and 95% coverage criteria, confirming that the original transcripts belong to *F. hygrometrica*. Of the 10,988 proteins of *F. hygrometrica* that did not align to the genome of *P. patens*, 3881 did share putative gene families with *P. patens* proteins, while 7107 did not. Yet, of these 6399 could be aligned to the *F. hygrometrica* draft genome, further confirming that the basis for the observed divergence is real.

Whether the 6399 and 3936 proteins to *F. hygrometrica* and *P. pyriforme* that seemingly lack a close ortholog in *P. patens* emerged as unique to these species due to significant sequence divergence beyond the mapping threshold of (alignment score = 500, percentage = 70, minimum intron length = 10) or more simply due to the loss of the orthologs in *P. patens*, coupled with high divergence between the *F. hygrometrica* and *P. pyriforme* orthologs, is not clear, and must await a comprehensive phylogenomic analysis of the Funariaceae. An alternative hypothesis would be a *de novo* origin of these loci from existing non-coding regions through overprinting or exonization for example, or from ancestral genes through fission, fusion or retroposition (Van Oss and Carvunis, 2019). While *de novo* gene birth has been demonstrated in plants, it does not appear to account for a significant number of new genes (i.e., 782) in *Arabidopsis* (Li et al., 2016) or n *Oryza* (i.e., 175, Zhang et al., 2019). *De novo* origin of genes in bryophytes has not yet been explored, due most likely to the lack of fully assembled and annotated genomes other that *P. patens* (Lang et al., 2018), and in particular of close or confamilial relatives of the latter (e.g., *F. hygrometrica*)

that would allow for synteny-based approaches to the identification of *de novo* genes from non-genic ancestors (McLysaght and Hurst, 2016). Testing the hypothesis of newly originating genes accounting for some of the observed divergence must await the final assembly and annotation of the *F. hygrometrica* genome.

The observed divergence in terms of low mapping rate of *F. hygrometrica* transcripts to *P. patens* is somewhat unexpected given the high overall similarity in vegetative body, in life strategies (i.e., short-lived annual life cycle) and in general ecology (i.e., terricolous, on mineral soil). It is, however, concordant with the hypothesis that the diversification of mosses, and in this case of the Funariaceae in particular, may be driven by metabolic optimization along ecological gradients (Glime, 1990), within a morphologically stable, and thus highly selected for, architecture of the leafy stem.

4.3. Evidence for whole genome duplication in F. hygrometrica and P. pyriforme

Ancient whole or large-scale genome duplications may characterize the evolutionary history of all mosses for which transcriptomes (e.g., Devos et al., 2016; Johnson et al., 2016; Pederson, 2019) or fully assembled genomes (Lang et al., 2018) are available. The transcriptomes of both *F. hygrometrica* and *P. pyriforme* bear signatures of an ancient WGD (Fig. 2B), and based on a comparison of substitutions between intragenomic paralogs and interspecific orthologs, the WGD likely preceded the split of *F. hygrometrica* and the *Physcomitrium* lineage (Fig. 2B). Lang et al. (2018) proposed that the genome of *P. patens* underwent one and likely two WGD, about 40–48 mya and 27–35 mya. The former estimate is within the range inferred based on plastid data by Medina et al. (2018) (40–75 mya), for the divergence of the Funarioideae suggesting its ancestor may have undergone a WGD. Although one more recent WGD may have occurred in *P. patens* (Lang et al., 2018; Gao et al., 2019), the evidence for a second genome duplication in *F. hygrometrica* is currently lacking and may await analysis of the full genome. Indeed, inferences of WGD based on paralog divergences estimated from transcriptomic data alone, can be biased especially if gene retention following duplication is low or if paralogs generated via WGD are more expressed than other duplicated paralogs, and hence multiple lines of evidence integrating genomic and phylogenetic data should be sought to provide robust support for proposed WGD (Tiley et al., 2018).

Although WGDs may have catalyzed genic innovation and perhaps gene family expansion and contraction, the number of genes duplicated via WGD tend to decrease with time since divergence (Qiao et al., 2019), and hence the impact of the WGD shared by the Funarioideae may be minimal. By contrast, more recent events may account for many of the duplicated genes in either or both *P. pyriforme* and *P. patens*. Furthermore, tandem and proximal duplication may be more effective in promoting gene innovation (Qiao et al., 2019). The transcriptomic data offer little or no insight into the extend and distribution of the duplication, and thus the contribution of whole versus tandem duplications to the evolution of new functional genes must await the analysis of the chromosome-level assembly of the *F. hygrometrica* genome.

4.4. Lineage specific shifts in gene family size in Funariaceae

The three Funariaceae sampled here exhibit significant differences in gene family size due to rapid evolution either through expansion or contraction (Fig. S2). The terminal branch of *F. hygrometrica*, is characterized by 24 gene families, which are rapidly evolving, either through rapid expansion (21) or contraction (3). The low rate of mapping of *F. hygrometrica* reads and proteins to *P. patens* results from of divergence between orthologs, but since the time of their divergence not many changes occurred in *F. hygrometrica* gene families size (24 rapidly evolving, Fig. S2). On the other hand, *P. pyriforme* shows more genomic overlap with *P. patens* but its terminal branch is characterized by significant differences in gene family size compared to *P. patens* (629 rapidly evolving, Fig. S2). The disparity in the size of their gene families may result from independent whole or partial duplications or contraction/loss in the other species, as well as tandem or proximal duplication. It should also be noted that transcriptomes provide only a partial estimate of this phenomenon. A summary of interesting trends in species-specific and rapidly evolving gene families are presented here.

Families with functions described as copper chaperone, copper ion binding, calcium-dependent protein kinase (CDPK), universal stress protein, callose synthase, allen oxide synthase, heat stress transcription factor and riboflavin biosynthesis protein were identified among *F. hygrometrica*-specific gene families (Table S4). Copper chaperones are requisite to distribute cooper to cellular compartments or copper binding proteins in yeast and mammals, also with similar homologs in *Arabidopsis* (Chu et al., 2005). One destination includes superoxide dismutase proteins (SODs), which can bind to copper/zinc (CuZnSOD). CuZnSODs are essential to protect the cells against Reactive Oxygen Species (ROS) through catalyzing dismutation of superoxide radicals (Beyer et al., 1991). Possession of copper chaperone and copper ion binding proteins in four *F. hygrometrica* specific gene families may explain why different *F. hygrometrica* populations have shown a wide range of tolerance to copper and zinc metal in their substrate (Shaw, 1987; Shaw et al., 1987; Shaw, 1988).

Calcium-dependent protein kinase gene family (CDPK) are associated with a variety of physiological activities and signaling pathways through calcium-stimulated protein phosphorylation, as studied in plants (Cheng et al., 2002). Their activity includes hormone responses (Bethke et al., 1995), many abiotic and biotic stress pathways: cold, wound, salinity, drought, and pathogen defense response (Knight and Knight, 2001).

Universal stress proteins, known by the heat shock protein domain (HSP domain) are common in prokaryotes, also are identified in angiosperms. Their function is studied well in prokaryotes. They aid in cell survival after toxic chemical and osmotic stress exposure (Nystrom and Neidhardt, 1992; Nystrom and Neidhardt 1993; Kerk et al., 2003). In angiosperms, increased temperature or heat shock leads to expression of heat shock proteins which are a type of molecular chaperone to bind to denatured proteins to prevent irreversible protein inactivation and aggregation (Waters et al., 1996). Additionally, heat shock proteins are also synthesized in many angiosperms in response to other stress treatments such as heavy metal and salt stress (Czarnecka et al., 1988; Harrington and Alm, 1988).

Riboflavin biosynthesis and heat stress transcription factor were also recognized among the *F. hygrometrica* specific gene families. Riboflavin is a fundamental cofactor in a variety of mainstream metabolic enzymes such as citric acid cycle, fatty acid oxidation, mitochondrial electron transport and photosynthesis (Jordan et al., 1999). Riboflavin biosynthesis protein can provoke signal transduction pathways and activation of pathogenesis-related genes in plants and eventually lead to disease resistance (Dong and Beer, 2000). Heat stress transcription factors, a final element of the signal transduction chain in *Arabidopsis*, mediates gene response to heat and chemical stresses (Nover et al., 2001).

The function of specific proteins may relate to the ability of *F. hygrometrica* to occupy disturbed areas and exhibit, perhaps, a much broader geographic and ecological distribution. Magdy et al. (2016) described a high genetic diversity and structure within and among *F. hygrometrica* ‘s populations along ecological gradients, arguing for its potential microecological adaptations. Furthermore, the functional characterization of the rapidly-expanding gene families in *F. hygrometrica* shows that these are overrepresented in GOs, including: sulfur compound metabolic process, organic compound metabolic process, cofactor metabolic (inorganic such as zinc, iron and copper cofactors involving in some chemical reaction) and some defensive mechanism such as viral response and response to osmotic stress (Tables S5 & S6; Fig. S2).

Only a few of *P. pyriforme*-specific proteins could be functionally annotated, such as those of the blue light photoreceptors gene families (Table S4). Blue light receptors control different developmental processes such as flowering time, phototropism and stomatal opening (Briggs and Huala, 1999), and may be involved in controlling gametangial genesis and sporophyte development, which may be controlled by temperature and light thresholds. In vitro gametangial development in *P. pyriforme* is typically induced by temperatures around 7 °C (Nakosteen and Hughes, 1978) as in *F. hygrometrica* (Monroe, 1965), versus 17°C in *P. patens* (Engel, 1968). However, we have observed development of gametangia populations of *P. pyriforme* other than the one sampled here in the absence of cold treatment, and hence under conditions approximating those under which *P. patens* forms sex organs. The functional characterization of rapidly-expanding gene families in *P. pyriforme* and *P. patens* revealed gene families related to reproductive processes, light and hormone response respectively (Tables S5 &S6).

The GO enrichment of *P. pyriforme* rapidly-expanding gene families exhibits many terminologies related to the sexual life cycle and reproductive system development (Fig. S3). The expression of such loci is unexpected given that cultures were phenotypically sterile, although we cannot exclude that gametangial development had been initiated. Overrepresented GOs related to reproductive terminologies in rapidly-expanding gene families of *P. pyriforme* may be linked to the accelerated gametophyte development in this species compared to *F. hygrometrica*. *Physcomitrium pyriforme* and *P. patens* may undergo a somewhat accelerated development, not just in their reduced sporophytic generation but also in their gametophyte. *Funaria hygrometrica* completes its life cycle in four months, whereas *P. pyriforme* and *P. patens* do so in three and two months, respectively (Nakosteen and Hughes, 1978).

5. Conclusions

The comparison of high-quality vegetative transcriptomes of Funariaceae belonging to two genera, *Funaria* and *Physcomitrium*, and their contrast against the genome of *P. patens* reveal significant divergence in gene space, reflective of the long evolutionary divergence between the genera and also of a highly dynamic evolution of the genomes, which is in sharp contrast with the rather static architecture of the vegetative body over 60 million years. While some genic innovations may be related to further specialization in life history traits, most of the functional annotations point at innovations in metabolic genes. The latter pattern may support the hypothesis formulated by Glime (1990) that the evolution of mosses, which are poikilohydric, and mycorrhizae devoid organisms, is driven, if not primarily, at least significantly by physiological adaptations along environmental gradients.

CRediT authorship contribution statement

Nasim Rahmatpour: Conceptualization, Writing - original draft. **Neranjan V. Perera:** . **Vijender Singh:** . **Jill L. Wegrzyn:** Conceptualization, Funding acquisition, Writing - review & editing. **Bernard Goffinet:** Conceptualization, Funding acquisition, Writing - review & editing.

Acknowledgements

This work was partially funded through grants from the National Science Foundation (DEB-1146295 & DEB-1753811 to BG) and from the Vice President for Research (Research Excellence Program) at the University of Connecticut to BG and JW. We also acknowledge the support of the Institute for Systems Genomics, including the Center for Genome Innovation and the Computational Biology Core at the University of Connecticut. Furthermore, we are grateful for support from T. Gerlach to the BG lab, and for Dr. Patel's (BG lab) careful reading of the final version of the manuscript.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2020.106965.

References

Barson, G., Griffiths, E., 2016. SeqTools: visual tools for manual analysis of sequence alignments. BMC Res. Notes 9, 39–48. https://doi.org/10.1186/s13104-016-1847-3.
Beike, A.K., Stackelberg, M.V., Schallenberg-Rüdinger, M., Hanke, S.T., Follo, M., Quandt, D., McDaniel, S.F., Reski, R., Tan, B.C., Rensing, S.A., 2014. Molecular evidence for convergent evolution and allopolyploid speciation within the *Physcomitrium-Physcomitrella* species complex. BMC Evol. Biol. 14, 158–177. https://doi.org/10.1186/1471-2148-14-158.
Bethke, P.C., Gilroy, S., Jones, R.L., 1995. Calcium and plant hormone action. In: Davies, P.J. (Ed.), Plant Hormones. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 298–317.
Beyer, W., Imlay, J., Fridovich, I., 1991. Superoxide dismutases. Prog. Nucleic Acid Res. Mol. Biol. 40, 221–253. https://doi.org/10.1016/S0079-6603(08)60843-0.
Blake, J.A., Christie, K.R., Dolan, M.E., et al., 2015. Gene ontology consortium: going forward. Nucleic Acids Res. 43, D1049–D1056. https://doi.org/10.1093/nar/gku1179.
Briggs, W.R., Huala, E., 1999. Blue-light photoreceptors in higher plants. Annu. Rev. Cell Dev. Biol. 15, 33–62. https://doi.org/10.1146/annurev.cellbio.15.1.33.
Cardona-Correa, C., Piotrowski, M.J., Knack, J.J., Kodner, R.E., Geary, D.H., Graham, L.E., 2016. Peat moss–like vegetative remains from Ordovician carbonates. Int. J. Plant Sci. 177, 523–538. https://doi.org/10.1086/686242.
Chater, C.C., Caine, R.S., Tomek, M., Wallace, S., Kamisugi, Y., Cuming, A.C., Lang, D., MacAlister, C.A., Casson, S., Bergmann, D.C., Decker, E.L., 2016. Origin and function of stomata in the moss *Physcomitrella patens*. Nat. Plants 2, 1–7. https://doi.org/10.1038/nplants.2016.179.
Cheng, S., Willmann, M.R., Chen, H., Sheen, J., 2002. Calcium signaling through protein kinases. The *Arabidopsis* Calcium-dependent protein kinase gene family. Plant Physiol. 129, 469–485. https://doi.org/10.1104/pp.005645.
Chomczynski, P., Wilfinger, W., Kennedy, A., Rymaszewski, M., Mackey, K., 2010. RNAzol® RT: a new single-step method for isolation of RNA. an4–an5. Nature Methods 7. https://doi.org/10.1038/nmeth.f.315.
Chu, C., Lee, W., Guo, W., Pan, S., Chen, L., Li, H., Jinn, T., 2005. A copper chaperone for superoxide dismutase that confers three types of copper/zinc superoxide dismutase activity in *Arabidopsis*. Plant Physiol. 52, 425–436. https://doi.org/10.1104/pp.105.065284.
Clarke, J.T., Warnock, R.C.M., Donoghue, P.C.J., 2011. Establishing a time-scale for plant evolution. New Phytol. 192, 266–301. https://doi.org/10.1111/j.1469-8137.2011.03794.x.
Collier, P.A., Hughes, K.W., 1982. Life cycle of the moss *Physcomitrella patens*, in culture. J. Tissue Cult. Methods 7, 19–22. https://doi.org/10.1007/BF01666875.
Coudert, Y., Novák, O., Harrison, C.J., 2019. A KNOX-cytokinin regulatory module predates the origin of indeterminate vascular plants. Curr. Biol. 29, 2743–2750. https://doi.org/10.1016/j.cub.2019.06.083.
Cove, D.J., Perroud, P.-F., Charron, A.J., McDaniel, S.F., Khandelwal, A., Quatrano, R.S., 2009. The moss Physcomitrella patens: a novel model system for plant development and genetic studies. Harbor Protocols 20 (2) pdb.emo115.
Crum, H., 1972. The geographic origins of the mosses of North America's eastern deciduous forest. J. Hattori Bot. Lab. 35, 269–298.
Cummins, H., Wyatt, R., 1981. Genetic variability in natural populations of the moss *Atrichum angustatum*. Bryologist 84, 30–38. https://doi.org/10.2307/3242975.
Czarnecka, E., Nagao, R.T., Key, J.L., Gurley, W.B., 1988. Characterization of Gmhsp26-A, a stress gene encoding a divergent heat shock protein of soybean: heavy-metal-induced inhibition of intron processing. Mol. Cell. Biol. 8, 1113–1122. https://doi.org/10.1128/MCB.8.3.1113.
Devos, N., Szövényi, P., Weston, D.J., Rothfels, C.J., Johnson, M.G., Shaw, A.J., 2016. Analyses of transcriptome sequences reveal multiple ancient large-scale duplication events in the ancestor of Sphagnopsida (Bryophyta). New Phytol. 211, 300–318. https://doi.org/10.1111/nph.13887.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., Gingera, T.R., 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.
Dong, H., Beer, S.V., 2000. Riboflavin induces disease resistance in plants by activating a novel signal transduction pathway. Phytopathology 90, 801–811. https://doi.org/10.1094/phyto.2000.90.8.801.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. https://doi.org/10.1093/bioinformatics/btq461.
Engel, P.P., 1968. The induction of biomedical and morphological mutants in the moss *Physcomitrella patens*. Am. J. Bot. 55, 438–446. https://doi.org/10.1002/j.1537-2197.1968.tb07397.x.
Feldberg, K., Schneider, H., Stadler, T., Schäfer-Verwimp, A., Schmidt, A.R., Heinrichs, J., 2014. Epiphytic leafy liverworts diversified in angiosperm-dominated forests. Sci. Rep. 4, 5974. https://doi.org/10.1038/srep05974.
Field, K.J., Pressel, S., Duckett, J.G., Rimington, W.R., Bidartondo, M.I., 2015. Symbiotic options for the conquest of land. Trends Ecol. Evol. 30, 477–486. https://doi.org/10.1016/j.tree.2015.05.007.

Fife, A.J., 1985. A generic revision of the Funariaceae (Bryophyta: Musci). Part I. J. Hattori Bot. Lab. 58, 149–196.
Fisher, K.M., 2008. Bayesian reconstruction of ancestral expression of the LEA gene families reveals propagule-derived desiccation tolerance in resurrection plants. Am. J. Bot. 95, 506–515. https://doi.org/10.3732/ajb.95.4.506.
Gao, B., Chen, M., Li, X., Liang, Y., Zhang, D., Wood, A.J., Oliver, M.J., Zhang, J., 2019. Integrated phylogenomic analyses reveal recurrent ancestral large-scale duplication events in mosses. bioRxiv. https://doi.org/10.1101/603191.
Gao, B., Zhang, D., Li, X., Yang, H., Liang, Y., Chen, M., Zhang, Y., Zhang, J., 2018. Desiccation tolerance in bryophytes: the rehydration proteomes of *Bryum argenteum* provides insights into the resuscitation mechanism. J. Arid Land. 10, 152–167. https://doi.org/10.1007/s40333-017-0033-3.
Gao, B., Zhang, D., Li, X., Yang, H., Zhang, Y., Wood, A.J., 2015. De novo transcriptome characterization and gene expression profiling of the desiccation tolerant moss *Bryum argenteum* following rehydration. BMC Genomics 16, 1–14. https://doi.org/10.1186/s12864-015-1633-y.
Glime, J.M., 1990. Primitive or advanced? The Bryological Times 55, 5–7.
Goffinet, B., Buck, W.R., 2013. The evolution of body form in bryophytes. Ann. Plant Rev. 45, 51–90. https://doi.org/10.1002/9781118305881.ch2.
Goffinet, B., Shaw, A.J., 2002. Independent origins of cleistocarpy in the Splachnaceae: analyses of cpDNA sequences and polyphyly of the Voitioideae (Bryophyta). Syst. Bot. 27, 203–208. https://doi.org/10.1043/0363-6445-27.2.203.
Goffinet, B., Buck, W.R., Wall, M.A., 2007. *Orthotrichum freyanum* (Orthotrichaceae, Bryophyta), a new epiphytic species from Chile. Beiheft zur Nova Hedwigia 131, 1–11.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.
Han, M.V., Thomas, G.W.C., Lugo-Martinez, J., Hahn, M.W., 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Mol. Biol. Evol. 30, 1987–1997. https://doi.org/10.1093/molbev/mst100.
Harrington, H.M., Alm, D.M., 1988. Interaction of heat shock and salt shock in cultured tobacco cells. Plant Physiol. 88, 618–625. https://doi.org/10.1104/pp.88.3.618.
Harrison, C.J., Roeder, A.H.K., Meyerowitz, E.M., Langdale, J.A., 2009. Local cues and asymmetric cell divisions underpin body plan transitions in the moss *Physcomitrella patens*. Curr. Biol. 19, 461–471. https://doi.org/10.1016/j.cub.2009.02.050.
Hart, A.J., Ginzburg, S., Xu (Sam), M., Fisher, C.R., Rahmatpour, N., Mitton, J.B., Wegrzyn, J.L., 2019. EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. Mol. Ecol. Resour. 20, 591–604. https://doi.org/10.1111/1755-0998.13106.
Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., Rattei, T., Mende, D.R., Sunagawa, S., Kuhn, M., Jensen, L.J., Mering, C.V., Bork, P., 2015. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286–D293. https://doi.org/10.1093/nar/gkv1248.
Jeffree, C.E., 2006. The fine structure of the plant cuticle. Ann. Plant Rev. 23, 11–125.
Johnson, M.G., Malley, C., Goffinet, B., Shaw, A.J., Wickett, N.J., 2016. A phylotranscriptomic analysis of gene family expansion and evolution in the largest order of pleurocarpous mosses (Hypnales, Bryophyta). Mol. Phylogenet. Evol. 98, 29–40. https://doi.org/10.1016/j.ympev.2016.01.008.
Jordan, D.B., Bacot, K.O., Carlson, T.J., Kessel, M., Viitanene, P.V., 1999. Plant riboflavin biosynthesis. Cloning chloroplads localization, expression, purification, and partial characterization of spinach lumazine synthaze. J. Biol. Chem. 274, 22114–22121. https://doi.org/10.1074/jbc.274.31.22114.
Kerk, D., Bulgrien, J., Smith, D.W., Gribskov, M., 2003. *Arabidopsis* proteins containing similarity to the universal stress protein domain of Bacteria. Plant Physiol. 131, 1209–1219. https://doi.org/10.1104/pp.102.016006.
Kirbis, A., Waller, M., Ricca, M., Bont, Z., Neubauer, A., Goffinet, B., Szövényi, P., 2020. The transcriptomic landscape of differential sporophyte development in two mosses, *Physcomitrium (Physcomitrella) patens* and *Funaria hygrometrica*. Front. Plant Sci. 11, 747. https://doi.org/10.3389/fpls.2020.00747.
Knight, H., Knight, M.R., 2001. Abiotic stress signaling pathways: specificity and cross-talk. Trends Plant Sci. 6, 262–267. https://doi.org/10.1016/S1360-1385(01)01946-X.
Kofuji, R., Yagita, Y., Murata, T., Hasebe, M., 2018. Antheridial development in the moss *Physcomitrella patens*: implications for understanding stem cells in mosses. Philos. Trans. R. Soc. Lond. B Biol. Sci. 373, 20160494.
Laenen, B., Shaw, B., Schneider, H., Goffinet, B., Paradis, E., Désamoré, A., Heinrichs, J., Villarreal, J.C., Gradstein, S.R., McDaniel, S.F., Long, D.G., Forrest, L.L., Hollingsworth, M.L., Crandall-Stotler, B., Davis, E.C., Engel, J., Von Konrat, M., Cooper, E.D., Patiño, J., Cox, C.J., Vanderpoorten, A., Shaw, A.J., 2014. Extant diversity of bryophytes emerged from successive post-Mesozoic diversification bursts. Nat. Commun. 5, 1–6. https://doi.org/10.1038/ncomms6134.
Landberg, K., Pederson, E.R.A., Viaene, T., Bozorg, B., Friml, J., Jönsson, H., Thelander, M., Sundberg, E., 2013. The moss *Physcomitrella patens* reproductive organ development is highly organized, affected by the two SHI/STY genes and by the level of active auxin in the SHI/STY expression domain. Plant Physiol. 162, 1406–1419. https://doi.org/10.1104/pp.113.214023.
Lang, D., Ullrich, K.K., Murat, F., Fuchs, J., Jenkins, J., Haas, F.B., Piednoel, M., Gundlach, H., Van Bel, M., Meyberg, R., Vives, C., Morata, J., Symeonidi, A., Hiss, M., Muchero, W., Kamisugi, Y., Saleh, O., Blanc, G., Decker, E.L., van Gessel, N., Grimwood, J., Hayes, R.D., Graham, S.W., Gunter, L.E., McDaniel, S.F., Hoernstein, S.N.W., Larsson, A., Li, F.W., Perroud, P.F., Phillips, J., Ranjan, P., Rokshar, D.S., Rothfels, C.J., Schneider, L., Shu, S., Stevenson, D.W., Thümmler, F., Tillich, M., Villarreal Aguilar, J.C., Widiez, T., Wong, G.K.S., Wymore, A., Zhang, Y., Zimmer, A.D., Quatrano, R.S., Mayer, K.F.X., Goodstein, D., Casacuberta, J.M., Vandepoele, K., Reski, R., Cuming, A.C., Tuskan, G.A., Maumus, F., Salse, J., Schmutz, J., Rensing, S.A., 2018. The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. Plant J. 93, 515–533. https://doi.org/10.1111/tpj.13801.
Leebens-Mack, J.H., Barker, M.S., Carpenter, E.J., et al., 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature 574, 679–685. https://doi.org/10.1038/s41586-019-1693-2.
Li, B., Dewey, C.N., 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, 323–339. https://doi.org/10.1186/1471-2105-12-323.
Li, L., Stoeckert, C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189. https://doi.org/10.1101/gr.1224503.
Li, Z.-W., Chen, X., Wu, Q., Hagmann, J., Han, T.-S., Zou, Y.-P., Ge, S., Gu, Y.-L., 2016. On the origin of de novo genes in *Arabidopsis thaliana* populations. Genome Biol. Evol. 8, 2190–2202. https://doi.org/10.1093/gbe/evw164.
Litt, A., 2013. Comprative evoluionary genomics of land plants. Annu. Plant Rev. 45, 227–276. https://doi.org/10.1002/9781119312994.apr0493.
Liu, Y., Johnson, M.G., Cox, C.J., Medina, R., Devos, N., Vanderpoorten, A., Hedenäs, L., Bell, N.E., Shevock, J.R., Aguero, B., Quandt, D., Wickett, N.J., Shaw, A.J., Goffinet, B., 2019. Resolution of the ordinal phylogeny of mosses using targeted exons from organellar and nuclear genomes. Nat. Commun. 10, 1–11. https://doi.org/10.1038/s41467-019-09454-w.
Magdy, M., Werner, O., Mcdaniel, S.F., Goffinet, B., Ros, R.M., 2016. Genomic scanning using AFLP to detect loci under selection in the moss *Funaria hygrometrica* along a climate gradient in the Sierra Nevada Mountains, Spain. Plant Biol. 18, 280–288. https://doi.org/10.1111/plb.12381.
Mallett, D.R., Chang, M., Cheng, X., Bezanilla, M., 2019. Efficient and modular CRISPR-Cas9 vector system for *Physcomitrella patens*. Plant Direct 3, e00168–e215. https://doi.org/10.1002/pld3.168.
Matasci, N., Hung, L.-H., Yan, Z., Carpenter, E.J., Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Ayyampalayam, S., Braker, M., Burleigh, J.G., Gitzendanner, M.A., Wafula, E., Der, J.P., dePamphilis, C.W., Roure, B., Philippe, H., Ruhfel, B.R., Miles, N.W., Graham, S.W., Mathews, S., Surek, B., Melkonian, M., Soltis, D.E., Soltis, P.S., Rothfels, C., Pokorny, L., Shaw, A.J., DeGironimo, L., Stevenson, D.W., Villarreal, J.C., Chen, T., Kutchan, T.M., Rolf, M., Baucom, R.S., Deyholos, M.K., Samudrala, R., Tian, Z., Wu, X., Sun, X., Zhang, Y., Wang, J., Leebens-Mack, J., Wong, G.K.-S., 2014. Data access for the 1,000 Plants (1KP) project. GigaScience 3, 1–10. https://doi.org/10.1186/2047-217X-3-17.
Medina, R., Johnson, M.G., Liu, Y., Wickett, N.J., Shaw, A.J., Goffinet, B., 2019. Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorrhegma*. J. Syst. Evol. 57, 404–417. https://doi.org/10.1111/jse.12516.
Medina, R., Johnson, M.G., Liu, Y., Wilding, N., Hedderson, T.A., Wickett, N., Goffinet, B., 2018. Evolutionary dynamism in bryophytes: phylogenomic inferences confirm rapid radiation in the moss family Funariaceae. Mol. Phylogenet. Evol. 120, 240–247. https://doi.org/10.1016/j.ympev.2017.12.002.
McLysaght, A., Hurst, L.D., 2016. Open questions in the study of de novo genes: what, how and why. Nat. Rev. Genet. 17, 567–578. https://doi.org/10.1038/nrg.2016.78.
Monroe, J., 1965. Some factors evoking formation of sex organs in Funaria. Bryologist 68, 337–339. https://www.jstor.org/stable/3240607.
Morris, J.L., Puttick, M.N., Clark, J.W., Edwards, D., Kenrick, P., Pressel, S., Wellman, C.H., Yang, Z., Schneider, H., Donoghue, P.C.J., 2018. The timescale of early land plants evolution. PNAS 115, E2274–E2283. https://doi.org/10.1073/pnas.1719588115.
Nakosteen, P.C., Hughes, K.W., 1978. Sexual life cycle of three species of Funariaceae in culture. Bryologist 81, 307–314. https://doi.org/10.2307/324191.
Natcheva, R., Cronberg, N., 2004. What do we know about hybridization among bryophytes in nature? Can. J. Bot. 82, 1687–1704. https://doi.org/10.1139/b04-139.
Nover, L., Bharti, K., Döring, P., Mishra, S.K., Ganguli, A., Scharf, K.-D., 2001. *Arabidopsis* and the heat stress transcription factor world: how many heat stress transcription factors do we need? Cell Stress Chaperones 6, 177–189. https://doi.org/10.1379/1466-1268(2001)006<0177:aathst>2.0.co;2.
Nystrom, T., Neidhardt, F.C., 1992. Cloning, mapping and nucleotide sequencing of gene encoding universal stress protein in *Escherichia coli*. Mol. Microbiol. 6, 3187–3198. https://doi.org/10.1111/j.1365-2958.1992.tb01774.x.
Nystrom, T., Neidhardt, F.C., 1993. Isolation and properties of mutant of *Escherichia coli* with an insertional inactivation of the *uspA* gene, which encodes a universal stress protein. J. Bacteriol. 175, 3949–3956. https://doi.org/10.1128/jb.175.13.3949-3956.1993.
Ortiz-Ramírez, C., Hernandez-Coronado, M., Thamm, A., Catarino, B., Wang, M., Dolan, L., Feijó, J.A., Becker, J.D., 2016. A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. Mol. Plant 9, 205–220. https://doi.org/10.1016/j.molp.2015.12.002.
Pederson, E.R.A., 2019. Genome sequencing of *Pleurozium scherberi*: the assembled and annotated draft genome of pleurocarpous feather moss. G3-Genes Genom Genet. 9, 2791–2797. https://doi.org/10.1534/g3.119.400279.
Prigge, M.J., Bezanilla, M., 2010. Evolutionary crossroads in developmental biology: *Physcomitrella patens*. Development 137, 3535–3543. https://doi.org/10.1242/dev.049023.
Proctor, M.C.F., Tuba, Z., 2002. Poikilohydry and homoihydry: antithesis or spectrum of possibilities? New Phytol. 156, 327–349. https://doi.org/10.1046/j.1469-8137.2002.00526.x.

Proctor, M.C., Oliver, M.J., Wood, A.J., Alpert, P., Stark, L.R., Cleavitt, N.L., Mishler, B.D., 2007. Desiccation-tolerance in bryophytes: a review. Bryologist 110, 595–621. https://doi.org/10.1639/0007-2745(2007)110[595:DIBAR]2.0.CO;2.

Qiao, X., Li, Q., Yin, H., Qi, K., Li, L., Wang, R., Zhang, S., Paterson, A.H., 2019. Gene duplication and evolution in recurring polyploidization–diploidization cycles in plants. Genome Biol. 20, 38–60. https://doi.org/10.1186/s13059-019-1650-2.

Rensing, S.A., Ick, J., Fawcett, J.A., Lang, D., Zimmer, A., Van de Peer, Y., Reski, R., 2007. An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. BMC Evol. Biol. 7, 130–139. https://doi.org/10.1186/1471-2148-7-130.

Rensing, S.A., Goffinet, B., Meyberg, R., Wu, S.-Z., Bezanilla, M., 2020. The moss *Physcomitrium (Physcomitrella) patens*: a model organism for non-seed plants. Plant Cell 32, 1361–1376. https://doi.org/10.1105/tpc.19.00828.

Såstad, S.M., Pedersen, B., Digre, K., 1999. Habitat-specific genetic effects on growth rate and morphology across pH and water-level gradients within a population of the moss *Sphagnum angustifolium* (Sphagnaceae). Am. J. Bot. 86, 1687–1698. https://doi.org/10.2307/2656667.

Shaw, A.J., 1985. The relevance of ecology to species concept in bryophytes. Bryologist 88, 199–206. https://www.jstor.org/stable/3243029.

Shaw, A.J., 1987. Effect of environmental pretreatment on tolerance to copper and zinc in the moss *Funaria hygrometrica*. Am. J. Bot. 74, 1466–1475. https://doi.org/10.1002/j.1537-2197.1987.tb12138.x.

Shaw, A.J., 1988. Genetic-variation for tolerance to Copper and Zinc within and among populations of the moss, *Funaria hygrometrica* Hedw. New Phytol. 109, 211–222. https://doi.org/10.1111/j.1469-8137.1988.tb03710.x.

Shaw, A.J., 2001. Biogeographic patterns and cryptic speciation in bryophytes. J. Biogeogr. 28, 161–253. https://doi.org/10.1046/j.1365-2699.2001.00530.x.

Shaw, A.J., 2009. Bryophyte species and speciation. In: Goffinet, B., Shaw, A.J. (Eds.), Bryophyte Bryology, second ed. Cambridge University Press, pp. 445–485.

Shaw, A.J., Antonovics, J., Anderson, L.E., 1987. Inter- and intraspecific variation of mosses in tolerance to Copper and Zinc. Evolution 41, 1312–1325. https://doi.org/10.2307/2409096.

Shaw, A.J., Devos, N., Cox, C.J., Boles, S.B., Shaw, B., Buchanan, A.M., Cave, L., Seppelt, R., 2010. Peatmoss (*Sphagnum*) diversification associated with Miocene Northern Hemisphere climatic cooling? Mol. Phylogenet. Evol. 55, 1139–1145. https://doi.org/10.1016/j.ympev.2010.01.020.

Shaw, A.J., Pokorny, L., Shaw, B., Ricca, M., Boles, S., Szövényi, P., 2008. Genetic structure and genealogy in the Sphagnum subsecundum complex (Sphagnaceae: Bryophyta). Mol. Phylogenetics. Evol. 49, 304–317. https://doi.org/10.1016/j.ympev.2008.06.009.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Slater, G.S., Birney, E., 2005. Automated generation of heuristics for biological sequence comparison. BMC Bioinf. 6, 3–41. https://doi.org/10.1186/1471-2105-6-31.

Stark, L.R., 2017. Ecology of desiccation tolerance in bryophytes: a conceptual framework and methodology. Bryologist 120, 130–165. https://doi.org/10.1639/0007-2745-120.2.130.

Stenøien, H.K., Såstad, S.M., 1999. Genetic structure in three haploid peat mosses (*Sphagnum*). Heredity 82, 391–400. https://doi.org/10.1038/sj.hdy.6884940.

Szövényi, P., Perroud, P.F., Symeonidi, A., Stevenson, S., Quatrano, R.S., Rensing, S.A., Cuming, A.C., Mcdaniel, S.F., 2015. De novo assembly and comparative analysis of the *Ceratodon purpureus* transcriptome. Mol. Ecol. Resour. 15, 203–215. https://doi.org/10.1111/1755-0998.12284.

Szövényi, P., Rensing, S.A., Lang, D., Wray, G.A., Shaw, A.J., 2011. Generation-biased gene expression in a bryophyte model system. Mol. Biol. Evol. 28, 803–812. https://doi.org/10.1093/molbev/msq254.

Tiley, G.P., Barker, M.S., Burleigh, J.G., 2018. Assessing the performance of Ks plots for detecting ancient whole genome duplications. Genome Biol. Evol. 10, 2882–2898. https://doi.org/10.1093/gbe/evy200.

Vanderpoorten, A., Goffinet, B., 2009. Introduction to Bryophyte Biology. Cambridge University Press, Cambridge, UK.

Van Oss, S.B., Carvunis, A.-R., 2019. De novo gene birth. PLoS Genet. 15, e1008160. https://doi.org/10.1371/journal.pgen.1008160.

Vidali, L., Bezanilla, M., 2012. *Physcomitrella patens*: a model for tip cell growth and differentiation. Curr. Opin. Plant Biol. 15, 625–631. https://doi.org/10.1016/j.pbi.2012.09.008.

Vitt, D.H., 1971. The infrageneric evolution, phylogeny, and taxonomy of the genus Orthotrichum (Musci) in North America. Nova Hedwigia 21, 683–711.

von Wettstein, F., 1924. Morphologie und Physiologie des Formwechsels der Moose auf genetischer Grundlage l. Zeitschrift für inductive Abstammungs- und Vererbunglehre 33, 1–236.

Waters, E.R., Lee, G.J., Vierling, E., 1996. Evolution, structure and function of the small heat shock proteins in plants. J. Exp. Bot. 47, 325–338. https://doi.org/10.1093/jxb/47.3.325.

Wood, A.J., 2007. The nature and distribution of vegetative desiccation-tolerance in hornworts, liverworts and mosses. Bryologist 110, 163–177. https://doi.org/10.1639/0007-2745(2007)110[163:IENFIB]2.0.CO;2.

Wyatt, R., Odrzykoski, I.J., Stoneburner, A., 1989. High levels of genetic variability in the haploid moss *Plagiomnium ciliare*. Evolution 43, 1085–1096. https://doi.org/10.1111/j.1558-5646.1989.tb02553.x.

Xiao, L., Wang, H., Wan, P., Kuang, T., He, Y., 2011. Genome-wide transcriptome analysis of gametophyte development in *Physcomitrella patens*. BMC Plant Biol. 11, 177. https://doi.org/10.1186/1471-2229-11-177.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Bioinformatics 13, 555–556. https://doi.org/10.1093/bioinformatics/13.5.555.

Young, M.D., Wakefield, M.J., Smyth, G.K., Oshlack, A., 2010. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 11, R14–R25. https://doi.org/10.1186/gb-2010-11-2-r14.

Zhang, L., Ren, Y., Yang, T., Li, G., Chen, J., Gschwend, A.R., Yu, Y., Hou, G., Zi, J., Xhou, R., Wen, B., Zang, J., Chougule, K., Wang, M., Copetti, D., Peng, Z., Zhang, C., Zhang, Y., Ouyang, Y., Wing, R.A., Liu, S., Long, M., 2019. Rapid evolution of protein diversity by de novo origination in *Oryza*. Nat. Ecol. Evol. 3, 679–690. https://doi.org/10.1038/s41559-019-0822-5.