

# Frequent allopolyploidy with distant progenitors in the moss genera *Physcomitrium* and *Entosthodon* (Funariaceae) identified via subgenome phasing of targeted nuclear genes

Nikisha Patel<sup>1</sup>, Rafael Medina<sup>2</sup>, Lindsay D. Williams<sup>3</sup>, Olivia Lemieux<sup>4</sup>, Bernard Goffinet<sup>4</sup>, Matthew G. Johnson<sup>3</sup>

<sup>1</sup>Biology Department, Trinity College, Hartford, CT, United States

<sup>2</sup>Biodiversidad, Ecología y Evolución, Universidad Complutense de Madrid, Madrid, Spain

<sup>3</sup>Department of Biological Sciences, Texas Tech University, Lubbock, TX, United States

<sup>4</sup>Department of Ecology and Evolutionary Biology, University of Connecticut, Mansfield, CT, United States

Corresponding author: Department of Biological Sciences, Texas Tech University 2901 Main Street, Lubbock, TX 79404, United States. Email: [matt.johnson@ttu.edu](mailto:matt.johnson@ttu.edu)

## Abstract

Allopolyploids represent a new frontier in species discovery among embryophytes. Within mosses, allopolyploid discovery is challenged by low morphological complexity. The rapid expansion of sequencing approaches in addition to computational developments to identifying genome merger and whole-genome duplication using variation among nuclear loci representing homeologs has allowed for increased allopolyploid discovery among mosses. Here, we test a novel approach to phasing homeologs within loci and phasing loci across subgenomes, or subgenome assignment, called Homologizer, in the family Funariaceae. We confirm the intergeneric hybrid nature of *Entosthodon hungaricus*, and the allopolyploid origin of *Physcomitrium eurystomum* and one population of *Physcomitrium collenchymatum*. We also reveal that hybridization gave rise to *Physcomitrium immersum*, as well as to yet unrecognized lineages sharing the phenotype of *Physcomitrium pyriforme* and *Physcomitrium sphaericum*. Our findings demonstrate the utility of our approach when working with polyploid genomes, and its value in identifying progenitor species using target capture data.

**Keywords:** bryophytes, polyploidy, hybridization, target capture, phylogenomics, allopolyploidy

## Introduction

The accurate quantification of species diversity is central to evolutionary biology. To this end, a molecular phylogenetic approach is indispensable, though it is increasingly recognized that, in plants, lineages of hybrid origin are frequently excluded from such analyses (Barker et al., 2016; Patel et al., 2021; Soltis et al., 2007). Allopolyploidy is a form of polyploidy in which a genome comprises multiple subgenomes inherited from two or more distinct progenitor species, and such merger may serve as a potential mechanism of speciation (Lohaus & Van de Peer, 2016; Soltis et al., 2015). The accurate identification of allopolyploids and their inclusion in phylogenetic analyses is, however, impeded by their tendency to be morphologically cryptic, as well as by challenges in phasing homeologous sequences within loci and accurately assigning the variants of all loci to parental subgenomes (Rothfels, 2021). The phylogenetic placement of allopolyploid subgenomes is critical for the identification of parental lineages and therefore a fuller understanding of trait evolution, ecology, and biogeography of groups including allopolyploids and their subgenomes (Oxelman et al., 2017; Rothfels, 2021; Rothfels et al., 2017). The use of DNA

data in the identification of allopolyploid parents has been widely accomplished in crop plants including *Gossypium* and *Triticum* (Chen et al., 2020; Grover et al., 2015; Ozkan et al., 2001), but methodological challenges remain in accomplishing this on a large scale for non-model plant species.

With a molecular approach, allopolyploid detection relies on the identification of variation within homeologous loci. We refer to homeologs, the copies from each of the progenitors of an allopolyploid and which are assumed to have no recombination, in contrast to allelic variation that segregates at a locus. In addition, here, heterozygosity refers to single nucleotide polymorphisms (SNPs) distinguishing homeologs from each subgenome in an allopolyploid genome. Historically, studies of polyploid species complexes have detected allopolyploids and identified homeologs by assessing nuclear markers for evidence of heterozygosity in the form of mixed nucleotide signals (for Sanger sequencing) or multiple allele sizes (for microsatellites; Clark & Schreier, 2017; Jorgensen & Barrington, 2017; Patel et al., 2018; Tate et al., 2006).

Phylogenetic inference is also used in identifying allopolyploid lineages. Historically, incongruence in phylogenetic position between nuclear and plastid trees was the basis for

Received February 3, 2022; revisions received August 30, 2023; accepted September 21, 2023

Associate Editor: Andrea Betancourt; Handling Editor: Tracey Chapman

© The Author(s) 2023. Published by Oxford University Press on behalf of The Society for the Study of Evolution (SSE).

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

identifying potential hybrids. However, an organellar genome represents only the maternal heritage and is considered a single gene by virtue of lacking recombination. Hence, phylogenetic inference increasingly relies on nuclear genes (Doyle, 2021). Although nuclear genetic markers are biparentally inherited, if phylogenetic inference is based on them without regard for within-sample variation (i.e., assembling reads to produce consensus sequences) sequences could be chimeric and not correctly resolve allopolyploid lineages (Oxelman et al., 2017; Šlenker et al., 2021). Single-molecule methods representing individual alleles or homeologs generated via vector cloning or two-step polymerase chain reaction (PCR) have been used to manually identify homeologs for a few discrete nuclear genes (Beike et al., 2014; Jorgensen & Barrington, 2017; Tate et al., 2006), but such an approach is not applicable to larger sets of loci, which may be needed to enhance accuracy in parental identification, particularly in cases of hybridization following rapid radiation or incomplete lineage sorting.

High-throughput sequencing methods have substantially improved phylogenetic resolution for non-polyploid species, but the inclusion of polyploids in phylogenetic analysis has been more challenging. Genome reduction approaches such as Hyb-Seq have allowed for rapid sequencing of hundreds of low-copy nuclear genes (Hale et al., 2020; Weitemier et al., 2014), and species trees are typically estimated under the coalescent model from unrooted gene trees. This approach provides data for robust phylogenetic inference and, by using pipelines such as HybPiper (Johnson et al., 2016) and HybPhaser (Nauheimer et al., 2021), allows for the discovery of likely allopolyploids through the detection of multiple copies and high heterozygosity (SNPs distinguishing homeologs) within biparentally inherited nuclear markers. However, the identification of progenitor lineages (subgenome phasing) is not trivial and remains a significant challenge, crucial to elucidating evolutionary history. Some studies have used parallel amplicon sequencing to generate a few nuclear loci and identified homeologs via manual comparison of gene trees. These approaches avoid problems associated with read assembly by utilizing long-reads sequencing such as PacBio, but scaling this analytical approach to the number of loci typical to a target capture approach like Hyb-Seq is not feasible (Rothfels et al., 2017).

To incorporate sequences representing subgenomes in a phylogeny, the homeolog sequences must be *in phase*, meaning that homeologs are assigned to the correct subgenome across loci. Within loci, tools such as ReadBackedPhasing in the Genome Analysis Toolkit can associate neighboring sequence variants using the sequence reads within a locus (Francioli et al., 2017). Across loci, additional evidence is needed to assign single-locus phased homeologs to each subgenome. Recently developed computational approaches allow for the subgenome phasing of dozens to hundreds of loci and hence the identification of hybrid progenitors. HybPhaser (Nauheimer et al., 2021) uses a two-step phylogenetic approach that identifies putative hybrids and re-assembles sequences using clade affinities, therefore, requiring a priori identification of appropriate references for each subgenome. The AlleleSorting pipeline (Šlenker et al., 2021) uses a distance-based approach to separate homeologs and assign them to subgenomes.

Here, we infer the subgenome phase of sequences using the Homologizer method (Freyman et al., 2023), implemented in RevBayes (Höhna et al., 2016). Homologizer differs from

the other recently developed methods by inferring the subgenome phase simultaneously with a phylogeny that includes the phased subgenomes of any allopolyploids as separate tips. Accordingly, an a priori reference sequence is not required.

Among mosses, shifts in ploidy arise from allopolyploidy or genome merger (Natcheva & Cronberg, 2004; Swangproh & Cronberg, 2021) or from autopolyploidy, that is intraspecific genome doubling (Patel et al., 2021). Historically, allopolyploids have been considered both rare and insignificant to moss evolution (Smith, 1978; Vitt, 1971). However, molecular phylogenetic studies in families such as Sphagnaceae (Ricca & Shaw, 2010), Mniaceae (Wyatt & Odrzykoski, 1998), and Funariaceae (Beike et al., 2014; McDaniel et al., 2010) find that allopolyploids play an important role in moss evolution. Allopolyploid mosses may, however, not be morphologically distinct from, or intermediate to, their progenitors, and the identification of their hybrid origin would thus rely on exploring molecular data for extensive fixed heterozygosity indicative of allopolyploid genomes. Accordingly, the exploration of molecular data toward the discovery of novel polyploid lineages has the potential to reveal significant morphologically cryptic diversity in mosses.

Our implementation of the Homologizer method to discover polyploids and their progenitors focuses on the Funariaceae, a cosmopolitan family of approximately 300 annual moss species (McIntosh, 2007). Most species in this family reside in the genera *Funaria*, *Entosthodon*, and *Physcomitrium* (Fife, 1985). The Funariaceae, like other moss families, include several complexes of putative, and likely unrecognized, polyploid species (Patel et al., 2021). Like other bryophytes, the life cycle of Funariaceae comprises a free-living vegetative haploid generation, the gametophyte, with, following sexual reproduction, a permanently attached diploid generation, the sporophyte (Figure 1). Note that when speaking of the ploidy level of mosses, it is customary to use the gametophyte generation as the standard—thus a “diploid” moss is a polyploid and it has a tetraploid sporophyte. The Funariaceae exhibit narrow phenotypic diversity in the gametophyte, in contrast to extensive trait variation in the sporophyte such as seta length, peristome architecture, or sporangial complexity and symmetry (Fife, 1985).

In this study we aim (1) first to confirm the allopolyploidy of three species of the Funariaceae using 50 nuclear genes: *Entosthodon hungaricus*, *Physcomitrium eurystomum*, and *Physcomitrium collenchymatum*, and investigate new evidence of allopolyploidy in *Physcomitrium immersum*, *Physcomitrium sphaericum*, and *Physcomitrium pyriforme*. All species except *P. collenchymatum* were sampled for target capture sequencing in Medina et al. (2019) but excluded from phylogenetic analysis as they showed heterozygosity congruent with a potential allopolyploid origin. The ploidy and evolutionary origins of *P. collenchymatum* are additionally considered here to address previously conflicting reports of allopolyploidy associated with the *P. collenchymatum* phenotype. McDaniel et al. (2010), Beike et al. (2014), and Ostendorf et al. (2021), each using the same nuclear dataset, report, based on a single identical population, that *P. collenchymatum* is an allopolyploid whereas Medina et al. (2018, 2019) treated other populations with this phenotype as haploid. Our second aim (2) is to infer evolutionary origins for component subgenomes and in doing so (3) highlight the utility and drawbacks of Homologizer as an approach to subgenome phasing and in identifying maternal and paternal



**Figure 1.** The gametophytic (haploid 1N) and sporophytic (diploid 2N) generations of *Physcomitrium pyriforme* s. lato.

progenitors of hybrid species. The development and implementation of a new approach to the discovery and characterization of these lineages will allow us to better understand the contribution of allopolyploidy to potential hidden moss diversity.

## Materials and methods

### Accessions and taxonomy

Each specimen is discussed and presented in phylogenetic analysis using its DNA accession number to support future taxonomic work (Supplementary Table S1). This four-digit identifier is consistent among vouchers, DNA extractions, and sequence data submitted to NCBI's GenBank. Medina et al. (2019) did not include in their nuclear target capture phylogeny 14 samples present in the organellar plastome phylogeny

(Medina et al., 2018) though the data were published (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA674709>). These 14 samples were initially flagged as allopolyploid due to (i) high numbers of paralog warnings during sequence assembly with HybPiper 1.3.1 (Johnson et al., 2016), and (ii) high mean heterozygosity across all loci (>10 bp/locus) in gametophyte tissue estimated by mapping reads to a *Physcomitrium patens* reference sequence. Gametophytes of non-polyploid mosses are haploid and thus are expected to have no or only negligible (i.e., less than 5% of sites, then attributed to sequencing error) detected heterozygosity. Thirteen of the 14 putatively allopolyploid samples are included here in a novel phylogenetic analysis to identify progenitor lineages of these hybrids. One putative allopolyploid is excluded as subsequent analysis revealed poor-quality data. In addition, we introduced one accession of *P. collenchymatum*, to address conflicting reports of hybrid status in previous publications. As a result, the final sampling includes 60 accessions comprising 46 haploids from Medina et al. (2019), 13 putative allopolyploids, and one additional sample of *P. collenchymatum*. We identified all samples to their closest morphological species concept using the most appropriate regional floras (e.g., Frey et al., 2006; Goffinet, 2007; McIntosh, 2007).

The newly included sample, labeled *P. collenchymatum* (5274), is collected from a propagated culture (Culture accession number 40061) from the International Moss Stock Culture (IMSC, Freiburg, Germany) that served as the exemplar for this species in prior studies (Beike et al., 2014; McDaniel et al., 2010; Ostendorf et al., 2021). Procedures for targeted sequencing of 806 nuclear loci for this sample followed Medina et al. (2019) with the following differences: we extracted genomic DNA using a modified CTAB (cetyltrimethylammonium bromide)/chloroform method, sheared it with the Fragmentase enzyme (New England Biolabs, Ipswich, MA, USA), and prepared shotgun sequencing libraries with the NEBNext Ultra-II DNA kit modified to use half-volumes compared with manufacturer recommendations. We pooled the new sample with other mosses not included in the present study for target enrichment using the MyBaits v3 protocol (Arbor Biosciences, Ann Arbor, MI, USA) and sequenced 2 × 300 paired-end reads on an Illumina MiSeq at the Texas Tech University Center for Biotechnology and Genomics.

### Sequence assembly

We recovered targeted sequences from the 13 allopolyploid samples omitted by Medina et al. (2019) in their nuclear phylogenetic analyses, using HybPiper v 1.3.1 (Johnson et al., 2016), as well as from the new sample of *P. collenchymatum*. For each sample, we extracted supercontig sequences, representing exons and flanking non-coding regions for each locus. We called variants within each sample using the GATK (version 4.1.0.6) Variant Detection Workflow (Van der Auwera & O'Connor, 2020), which began by mapping reads from a sample to the recovered supercontigs from the same sample. We used MarkDuplicates (Picard Toolkit, 2019) to identify and remove duplicate reads via common mapping location prior to variant calling with HaplotypeCaller. We saved only the SNPs and used a hard filter to retain only SNPs with high mapping quality, read depth, and base call quality (QD < 5.0 && FS > 60.0 && MQ < 40.0 && MQRankSum < -12.5 && ReadPosRankSum < -8.0).

For samples with an F-statistic greater than 0.5, indicating higher-than-expected heterozygosity, we phased homeologs

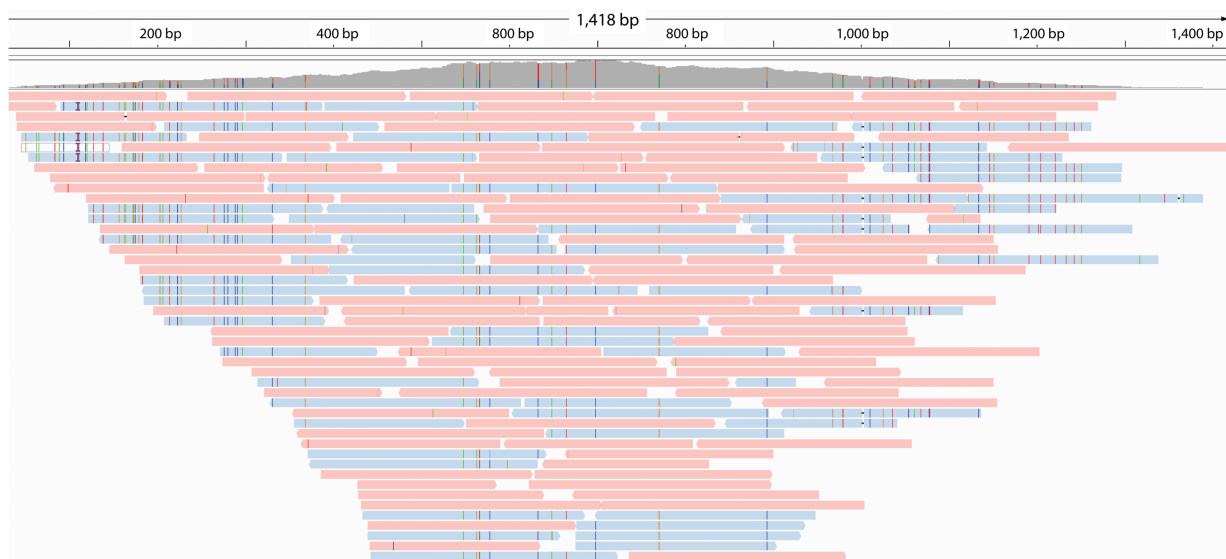
within each locus using a method similar to that of Kates et al. (2018). We used WhatsHap (version 1.1, Martin et al., 2016) to phase variants and extracted phased sequences at each locus using the Python script `haplonerate.py` ([github.com/mossmatters/phyloscripts](https://github.com/mossmatters/phyloscripts)). The latter script is necessary to sort sequences within the largest phaseable block (Figure 2), that is, the largest stretch of sequence within a locus where variants can be phased using the sequence reads. While the method of Kates et al. (2018) retains variants outside the largest block as ambiguous sites, we chose the more conservative approach of deleting all sequences that fell outside the largest block. This is consistent with the approach taken by Tiley et al. (2021).

To determine the likely ploidy for our samples with a high F-statistic, we used nQuire (accessed from GitHub 2022-01-01, Weiß et al., 2018) to fit Gaussian mixture models to within-sample allele frequencies. In nQuire documentation, sequence variants associated with a locus are referred to as “alleles” though in the present study, when multiple “alleles” are detected, these would be homeologs. This approach estimates the frequency of SNPs with multiple allelic variants. Various allelic ratios (e.g., 0.5/0.5 for diploids) are generated for each ploidy. Since we have sampled and sequenced gametophytic tissue, we accordingly expect that polyploids will have a ploidy of two or higher. Because nQuire does not have a model for haploid samples, we were only able to analyze our 14 putatively allopolyploid samples to choose among diploid, triploid, and tetraploid models via maximum likelihood. For all samples, we implemented denoising within nQuire to eliminate low-quality reads with noise generated from repetitive sequences.

### Subgenome phasing

We completed the GATK + WhatsHap workflow for the 648 genes used in Medina et al. (2019) for the 14 polyploid samples, resulting in reads assembled into homeologs for each locus (Figure 2). Any sequence outside the largest phase block within a locus was deleted from the polyploid samples. As a full Bayesian inference method, computational limitations are

a concern, so we chose to phase subgenomes using 50 nuclear loci. We first sampled two subsets of these genes (25 genes each) and analyzed them independently using Homologizer. One accession (*P. pyriforme*—3798) has missing data for one of the 50 genes selected, gene number 8583 (Supplementary Table S2). Each locus yielded consistent topologies and placement of each allopolyploid subgenome, suggesting that a subset of single-copy nuclear loci likely yields a phylogeny representative of the full dataset. In a Homologizer analysis that we will refer to as *iteration one*, samples with ploidies higher than diploid are still phased as diploids. Since attempts to phase three subgenomes for a given sample in iterations two and three result in greater genic conflict and lack of convergence, respectively, iteration one is presented to show subgenome phasing for allodiploids. For iteration one, we added two phased homeologous sequences per polyploid sample (28 alleles) to the haploid nuclear marker alignments (46 sequences) from Medina et al. (2019) resulting in 74 sequences for each gene. We added homeolog sequences from our 14 putative allopolyploid samples to the trimmed supercontig alignments from Medina et al. (2019) using MAFFT v 7.407 with the `addsequence` and `keplength` options (Katoh & Frith, 2012). Homologizer (Freyman et al., 2023) is a function in the software RevBayes (version 1.1, Höhna et al., 2016) that jointly infers the subgenome phase across loci along with the phylogeny from a matrix of DNA sequences. We followed the methods available from [github.com/wf8/homeolog\\_phasing](https://github.com/wf8/homeolog_phasing) to prepare a RevBayes control file containing prior associations between homeolog sequences and subgenomes for each sample, proposal frequencies for swapping homeologs between subgenomes, and standard priors for a GTR (general time-reversible) substitution model. We ran RevBayes for 10,000 generations (representing over 5.32 million Metropolis–Hastings moves, Höhna et al., 2016) on four independent runs and checked for chain stationarity in each run via a stable posterior probability and an effective sample size (ESS) of > 200 (a convention for Bayesian inference, Drummond & Bouckaert, 2015) for all parameters using Tracer v1.7 (Rambaut et al., 2018).



**Figure 2.** Reads associated with nuclear marker 7379 from *P. immersum*—3176 indicating single nucleotide polymorphisms (SNPs) in a subset of reads representing two distinct subgenomes in the allodiploids. This is representative of the read-backed phasing approach.

In order to attempt to represent polyploids greater than diploids with more than two tips, as we did in iteration one, we implemented *iteration two*, or the “fixed-diploid” approach, as follows: First, we tested the capacity of the Homologizer to make subgenome assignments when dealing with a ploidy higher than two, so we chose to impose triploidy on samples *P. pyriforme*—3410 and 3798. For populations with a ploidy greater than two, as estimated by nquire, variants were inferred with GATK changing the—ploidy flag where appropriate. We used the “polyphase” function in WhatsHap to generate phased haplotypes within each locus. Because earlier runs struggled with convergence when including higher-ploidy samples, we took a tiered approach to phased subgenomes of the higher-ploidy samples: first, we used the output of iteration one to set (fix) the subgenome phasing for diploids across loci in a second run of Homologizer, where we only inferred phase for the samples with higher ploidy (*P. pyriforme*—3410 and 3798). Finally, we implemented a third approach, *iteration three*, or the “dummy-tip” analysis, for the incorporation of higher ploidies represented by more than two tips. The “dummy tip” method has previously been employed for higher ploidy sequences (Freyman et al., 2023), where a blank sequence was inserted for a high-ploidy sample as a way to accommodate unsampled or lost homeologs. We ran ten independent runs of Homologizer for 5,000 generations for both the “fixed diploid” and “dummy tip” methods.

For all analyses, in addition to the Bayesian phylogeny inferred using RevBayes, we inferred a maximum quartet species tree using ASTRAL-III (Zhang et al., 2017). Using the joint posterior probability (the posterior probability determined jointly among loci) for the subgenome phase from Homologizer, we modified the sequence labels for all 14 polyploid samples to indicate assigned subgenomes. We inferred gene trees for each locus, with the subgenomes treated as their own tips, using IQTREE (version 2.0.3, Minh et al., 2020), with substitution models selected using ModelFinderPlus (Kalaanamorthy et al., 2017) and assessed support based on 1000 UltraFast Bootstrap (Hoang et al., 2018) replicates. We collapsed gene tree branches with less than 30% support using the `nw_ed` function in `newick_utils` (accessed from GitHub 2020-06-01, Junier & Zdobnov, 2010). After inferring a species tree from the collapsed gene trees with ASTRAL-III (version 5.6.3, Zhang et al., 2017), we used `phyparts` (version 0.0.1, Smith et al., 2015) to assess bipartitions among the 50 genes and visualized concordance and conflict with `phypartspiecharts.py` and `minorityreport.py` (github.com/mossmatters/phyloscripts). To determine the extent to which conflict among gene trees could be attributed to the inclusion of allopolyploid subgenome sequences, a second ASTRAL tree including only haploids in the dataset was generated.

### Maternal and paternal subgenomes

Each of the allopolyploid samples in the present study, except the newly sampled *P. collenchymatum*—5274, was included in the organellar phylogenomic reconstruction by Medina et al. (2018) and hence should have one nuclear subgenome in a phylogenetic position identical to that in the plastid exome phylogeny reconstructed from the same samples (Medina et al., 2018). These relationships are therefore used here to infer the maternally inherited subgenome in each allopolyploid.

For one sample, *P. pyriforme*—3410, substantial gene conflict in the positioning of its subgenomes indicated a need to rule out contamination and ensure that *P. pyriforme*—3410 in

the present study was the sample for which plastid DNA was extracted and sequenced by Medina et al. (2018). Three plastid genes were recovered from target capture data generated for *P. pyriforme*—3410 generated by Medina et al. (2019) using HybPiper v 1.3.1 (Johnson et al., 2016). Consensus sequences were aligned with single gene alignments from Medina et al. (2018) to assess sequence similarity with *P. pyriforme*—3410.

### Data availability

All nucleotide alignments, RevBayes control files, gene trees, and species trees are available via the Dryad Digital repository (10.5061/dryad.0p2ngf23f). For target capture baits used, see Medina et al. (2019) and Dryad Digital repository (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.0p2ngf23f>).

### *Physcomitrium collenchymatum* sequencing

We tested whether conflicting reports of allopolyploidy in *P. collenchymatum* were attributable to culture contamination, by comparing a barcoding locus sequenced from DNA extracted from the culture propagated in our laboratory (5274), the original IMSC culture (40061), and the herbarium voucher from which the later was established. In addition, two plastid markers are sequenced for *P. collenchymatum*—5274 and used as the basis of maternal subgenome identity.

### Sanger sequencing

For target-capture sequencing of the nuclear loci 4780 and 7379, both homeologs were consistently recovered in allopolyploids. Accordingly, Sanger sequencing was used to generate reads for nuclear markers 4780 and 7379 (Supplementary Table S3) for the IMSC culture (40061) and our associated culture (5274) of *P. collenchymatum* to confirm their identity (Supplementary Table S1). Nuclear sequences associated with the two cultures were compared with the sequences generated from the original dry voucher specimen (Homberg 1155 MO). In addition, two plastid markers, *psbA-trnH* and *rps4-trnS*, were sequenced for the culture *P. collenchymatum*—5274 to discern the maternal subgenome based on phylogenetic relatedness with haploid samples (Supplementary Table S3).

### DNA extraction and PCR amplification

For live cultures, total genomic DNA was extracted using the NucleoSpin Plant II Midi kit (Macherey-Nagel, Düren, Germany) following the manufacturer’s instructions. Nuclear markers 7379 and 4780 as well as plastid markers *psbA-trnH* and *rps4-trnS* were amplified using primers detailed in Supplementary Table S3. For the nuclear markers, PCRs were performed in a final volume of 25  $\mu$ l with 12.5  $\mu$ l of GoTaq Green Mastermix (Promega, Madison WI, USA), 9.5  $\mu$ l of water, 1  $\mu$ l of the forward and reverse primer (10  $\mu$ l), and 1  $\mu$ l of genomic DNA extract. PCR consisted of a 3 min hot start at 94°C, followed by 40 cycles of denaturation (1 min, 94°C), annealing (1 min, 50°C), and extension (1 min, 70°C), ended by a final extension step of 10 min. For plastid markers *psbA-trnH* and *rps4-trnS*, a nested PCR was used. Both PCRs consisted of a 3 min hot start at 94°C, followed by 40 cycles of denaturation (1 min at 94°C), annealing (1 min at 50°C), and extension (1 min at 70°C), ended by a final extension step of 10 min. PCR products were cleaned using the ExoSAP-IT protocol (USB-Affymetrix, Cleveland OH, USA) and sequenced at Eurofins Genomics on an ABI prism sequencing platform (Louisville, KY).

## Results

### Within-locus phased assembly

Haploid samples are evidenced by a low proportion of heterozygous sites identified by GATK (Supplementary Table S1) and are represented in phylogenetic analyses with only one sequence per locus. Among the allopolyploid samples previously excluded from the Medina et al. (2019) nuclear phylogeny, the number of variant sites per locus within a sample ranged from 42.86 (*E. hungaricus*—3177) to 133.74 (*P. pyriforme*—3410) sites per gene, compared to a range of 0.00–7.62 for the haploid samples. The samples with high numbers of within-sample variants include four samples of *P. eurystomum*, four of *P. sphaericum*, two of *P. pyriforme*, one of *P. immersum*, two of *E. hungaricus*, and one of *P. collenchymatum* (Supplementary Table S1).

### nQuire

Diploidy is the best-fit model for 12 hypothesized allopolyploids, whereas triploidy is the best-fit model for *P. pyriforme*—3410 and tetraploidy is the best-fit model for *P. pyriforme*—3798. (Supplementary Table S4; ploidy assessments are based on the gametophyte, hence the nQuire results suggest that *P. pyriforme*—3798 has octoploid sporophytes, see Figure 1). It should be remembered that while nQuire provides a hypothesis for ploidy, it is designed for polyploid detection in angiosperms where the dominant generation is diploid, rather than the haploid gametophyte in mosses that are sampled here.

### Subgenome phasing and phylogenetic relationships

The apparent triploid and tetraploid, *P. pyriforme*—3410 and 3798, are treated as diploids in iteration one of our analyses and as higher-level ploidies (three subgenomes) in iteration two (the “fixed diploid” approach) to test Homologizer’s capacity to phase higher-level ploidies and determine future approaches to improve its robustness. To evaluate the effect of treating *P. pyriforme*—3410 and 3798 as diploids, in iteration one, on inferred phylogenetic relationships among the allodiploid subgenomes, a RevBayes tree excluding *P. pyriforme*—3410 and 3798 is included (Supplementary Figure S1). The topology is largely consistent with that of the ASTRAL tree including subgenomes representing all polyploids (Supplementary Figure S2). The exception is the circumscription of subgenomes associated with *E. hungaricus*—3177 and 3838, since the maternal subgenome of each is sister to the maternal subgenome of triploid *P. pyriforme*—3410 in the iteration one analysis (Supplementary Figure S2). As a result, the RevBayes tree excluding the higher-level ploidies resolves all subgenomes of *E. hungaricus*—3177 and 3838 in one clade on a short branch (Supplementary Figure S1). The remainder of the allopolyploid samples are apparent allodiploids (Supplementary Table S4). The total length of the trimmed alignments for 50 phased nuclear supercontigs was 187,988 bp (per gene average = 3,759 bp, range 1,668–12,446 bp). The iteration two analysis testing the incorporation of greater than two subgenomes per sample represents *P. pyriforme*—3410 and 3798 with three tips, or three subgenomes. The topology resulting from the ASTRAL tree for the second iteration is largely consistent with the ASTRAL tree from the first iteration representing these populations with two subgenomes. The topologies differ in the position of the

likely paternal subgenome of *P. pyriforme*—3798, and paternal and maternal subgenomes of *P. pyriforme*—3410. When treated as triploids, *P. pyriforme*—3410 includes two subgenomes in a clade with a subgenome of *P. pyriforme*—3798 sister to the *P. pyriforme* species complex, and one subgenome sister to the *Physcomitrium japonicum* and *P. pyriforme* species complexes. The remaining two subgenomes of *P. pyriforme*—3798 are in the *P. pyriforme* species complex (Supplementary Figure S4).

Homologizer yielded two main outputs: (i) a Bayesian phylogenetic inference (including the haploid samples and the subgenomes for each allopolyploid sample) and (ii) posterior probabilities for assigning homeologs of each gene to a subgenome. After discarding the first 25% of the run as burn-in, we calculated a maximum credibility tree that shows full support at nearly all nodes for iteration one (Supplementary Figure S6) and iteration two (Supplementary Figure S7). Convergence and stationarity of Homologizer run with all polyploid samples coded as diploid (iteration one) was confirmed using ESS > 200 for all parameters and a stable posterior probability after burn-in. By contrast, Homologizer runs struggled to reach convergence when the two higher-ploidy samples (*P. pyriforme*—3410 and 3798) were represented by three subgenomes and phased simultaneously with allodiploids (iteration two). For the “fixed diploid” analysis, which fixed the subgenome phase for diploid samples while we attempted to phase the remaining two samples as triploid, only seven of 10 runs converged within 5,000 generations. Finally, for iteration three, the “dummy tip” analysis where a fourth blank sequence was added for the two higher ploidy samples, each of the 10 runs converged on a different posterior probability (Supplementary Figure S8).

Phylogenetic inferences in iterations one and two resolve a tree topology largely consistent with Medina et al. (2019) (Figure 3; Supplementary Figure S7). As before, *P. pyriforme* is resolved as paraphyletic. In iteration one, only two nodes are not resolved with a 100% posterior probability, and the posterior probability of subgenome assignment was maximal (PP = 1.0) for all samples and loci, which allowed assigning each homeolog to a subgenome for further analysis (Supplementary Figure S6).

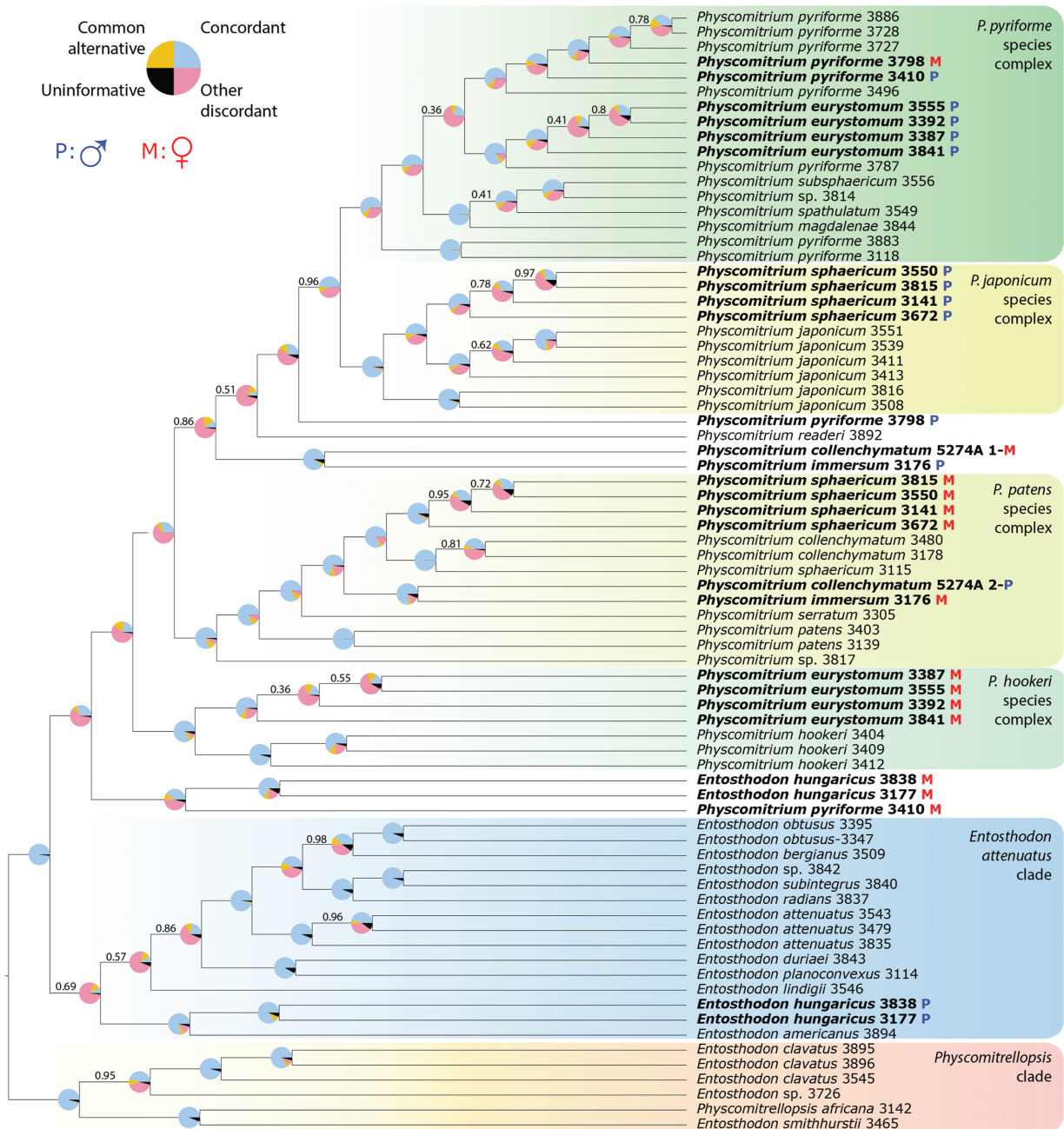
The subgenomes of *E. hungaricus*—3177 and 3838 are the two most phylogenetically distant subgenomes in the sampling: the maternal subgenome belongs to a lineage sister to all *Physcomitrium* species and the paternal subgenome originated from the *Entosthodon attenuatus* clade (Figures 3 and 4). The paternal subgenomes of *E. hungaricus*—3177 and 3828 are resolved as part of the *Entosthodon* clade with minimal genic conflict (39/50 genes are concordant). However, gene conflict (only 20/50 genes are concordant) is substantial at the ancestral node of the clade including the maternal subgenomes of *E. hungaricus*—3177 and 3828 (Supplementary Figure S2). In iteration two, there is little genic conflict in the resolution of the paternal genomes and substantially more conflict in the resolution of the maternal genomes (Supplementary Figure S4).

The specimens *P. sphaericum*—3141, 3550, 3672, and 3815 are from China and their paternal subgenome is from the *P. japonicum* species complex, whereas the maternal subgenome is from the *P. patens* species complex related to (haploid) *P. collenchymatum*. The respective subgenomes of *P. immersum*—3176 and *P. collenchymatum*—5274 are sister to each other, with one parental subgenome sister to

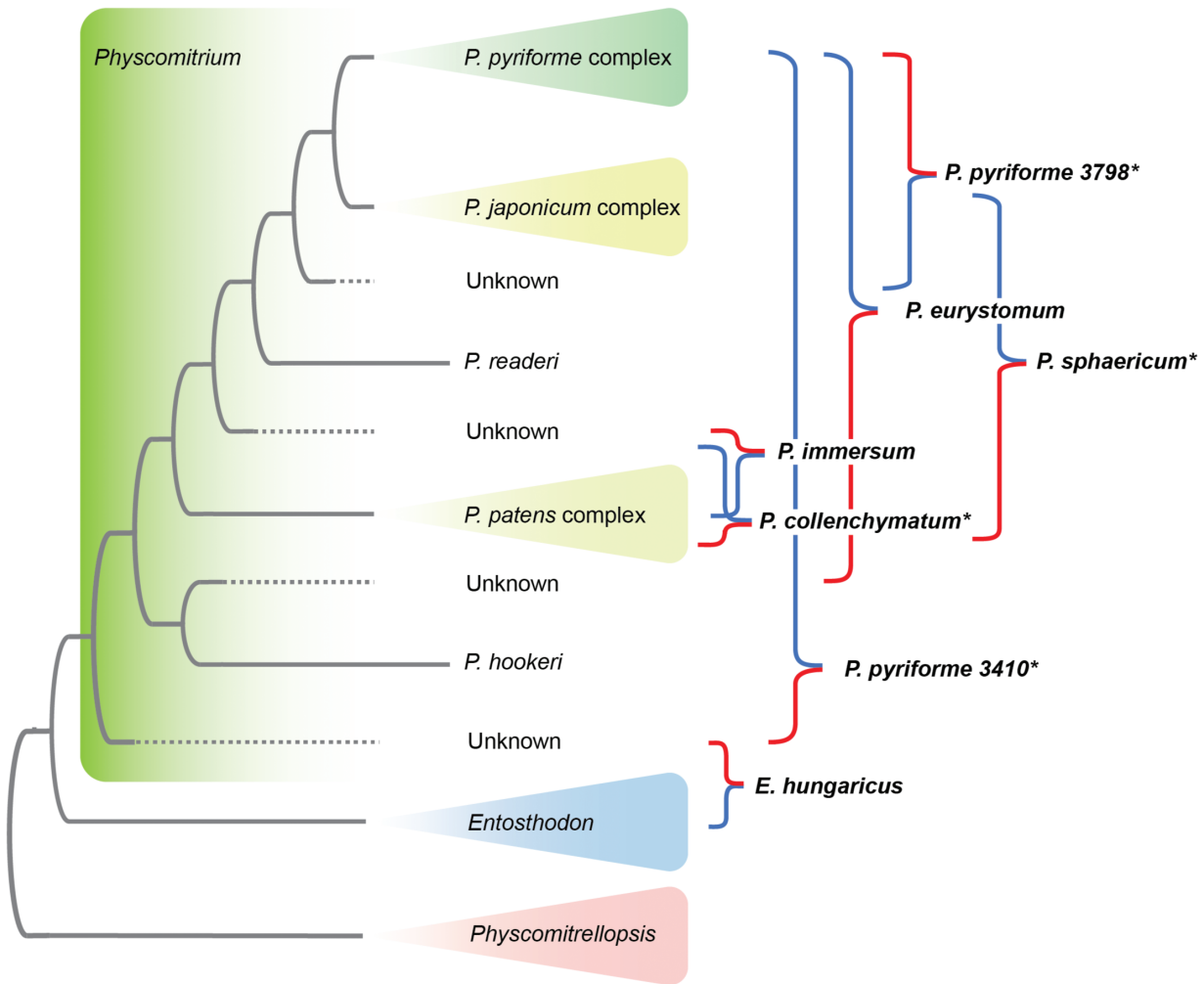
the clade comprising *Physcomitrium readeri*, the progenitor to the *P. japonicum* and *P. pyriforme* species complexes and the second sister to a clade including *P. collenchymatum* and *Physcomitrium serratum* (Figures 3 and 4; Supplementary Figure S7). However, the maternal and paternal contributions of the two proposed progenitors to these two allopolyploid lineages are reversed with the maternal subgenome of *P. immersum*—3176 in the *P. patens* species complex and the maternal subgenome of *P. collenchymatum*—5274 sister to the *P. pyriforme* species complex (Figure 3). *P. eurystomum*—3355, 3392, 3387, and 3841 comprise two subgenomes, with the

paternal parent closely related to *P. pyriforme*—3787 and the maternal one sister to *Physcomitrium hookeri* (Figure 3; Supplementary Figure S7).

In the iteration one analysis, our ASTRAL analysis of gene trees inferred from phased subgenome sequences revealed a topology largely congruent with the concatenated topology from RevBayes (Figure 3; Supplementary Figures S5 and S6). The only partially supported incongruence pertains to the placement of the paternal subgenome of *P. pyriforme*—3798: in the RevBayes tree it is nested in a clade with the maternal genome of *P. pyriforme*—3410, whereas in the ASTRAL



**Figure 3.** Phylogenetic reconstruction of haploid samples and phased subgenomes of allopolyploids using ASTRAL from 50 gene trees. Pie charts on each branch indicate the proportion of gene trees concordant with that bipartition (blue), proportion of gene trees with the most common discordant bipartition for that bipartition (orange), proportion of gene trees with all other discordant bipartitions for that bipartition (red), and proportion of gene trees with no support due to missing data for that bipartition (black). Support values are local posterior probability from ASTRAL; only values less than 1.0 are shown. Boxes indicate named clades as discussed in text. Taxa in bold are allopolyploids represented by their maternal (M) and paternal (P) subgenomes.



**Figure 4.** Phylogenetic relationships among major clades represented in Figure 3, inferred using ASTRAL based on 50 genes. Allopolyploid accessions are indicated with brackets adjacent to the phylogeny. Red and blue lines indicate maternal and paternal heritage, respectively. Taxon names currently applied to allopolyploid and haploid accessions are marked with an asterisk.

tree its placement is unsupported along the backbone of *Physcomitrium*. Other disagreements, including the placement of *P. readeri*, are unsupported in the ASTRAL analysis. Short internodes in part indicate a high level of conflict among the 50 genes, particularly for the monophyly of the clade comprising the *P. pyriforme* and *P. japonicum* species complexes and for the placement of *P. readeri* and subgenomes of *P. immersum* and of the polyploid *P. collenchymatum* (Figure 3; Supplementary Figure S5).

Bipartition analysis using phyparts (Smith et al., 2015) reveals some incongruence among genes supporting relationships of allopolyploid subgenomes to haploid samples. For example, in the iteration one analysis, which treats *P. pyriforme*—3410 as a diploid, the species tree placement of the maternal subgenome is resolved with the *E. hungaricus* maternal subgenome by 20 of 50 genes. The placement of the paternal subgenome of *P. pyriforme*—3410 with haploid samples of *P. pyriforme* from North Carolina in the *P. pyriforme* species complex is supported by 22 genes. In both cases, conflict arises from substantial minority bipartitions: nine genes reconstruct the two homeologs from sample *P. pyriforme*—3410 as sister to each other, while the homeologs from the two *E. hungaricus* samples form a clade in 15 gene

trees (Supplementary Figure S2). Similarly, in the iteration two analysis, i.e., “fixed diploid” approach including three subgenome tips for *P. pyriforme*—3410 and 3798, the nodes supporting placement of each subgenome is supported by at most 15 genes (Supplementary Figure S4).

By contrast, the placement of *P. immersum* subgenomes sister to those of the allodiploid sample of *P. collenchymatum*—5274 in iteration one is consistent for the paternal (45 gene trees) and maternal subgenomes (39 gene trees) with no dominant minority bipartition (Supplementary Figure S2). Similarly, in iteration two, each *P. immersum* subgenome is sister to a *P. collenchymatum* subgenome, supported by a majority of genes (47 gene trees; 41 gene trees; Supplementary Figure S4). For the *P. sphaericum* samples from China, homeologs from one subgenome consistently form a clade sister to haploid *P. collenchymatum* (iteration one: 41 gene trees; iteration two: 40 gene trees). There is also support among the majority of gene trees for the other subgenome of *P. sphaericum* as part of the *P. japonicum* species complex (iteration one: 47 gene trees; iteration two: 46 gene trees). The genic support for the affinity of subgenomes of *P. eurystomum* is more variable (Supplementary Figures S2 and S4).

## Sanger sequencing results

The original culture of *P. collenchymatum* from IMSC (40061), our own culture (5274), and the original voucher specimen (Homberg 1155, MO) associated with these cultures, share identical sequences for two nuclear markers, each with the same signatures of heterozygosity in each sample (Supplementary Figure S3). In addition, phylogenetic affinities of the plastid sequences *psbA-trnH* and *rps4-trnS* generated from both cultures of *P. collenchymatum* indicate the maternal subgenome is sister to the *P. pyriforme* species complex (Supplementary Figure S9). Assembled chromatograms are available on Dryad.

### *Physcomitrium pyriforme*—3410 identity

Three plastid genes were recovered from target capture data generated for *P. pyriforme*—3410 by Medina et al. (2019) using HybPiper v 1.3.1 (Johnson et al., 2016). Consensus sequences were aligned with single gene alignments from Medina et al. (2018) to assess similarity with *P. pyriforme*—3410 via target capture of organellar exons for that study. The sequences recovered from target capture data associated with *P. pyriforme*—3410 used in the present study were consistent with SNPs of plastid sequences published for *P. pyriforme*—3410 in Medina et al. (2018), confirming that the same sample is analyzed in the present study (Supplementary File S1).

## Discussion

The present study contributes to a growing body of research on polyploidy and hybridization contributing to the evolution and diversification of the Funariaceae and provides evidence for overlooked diversity. We find evidence of allopolyploidy across multiple populations of several species in the *Entosthodon*–*Physcomitrium* complex, suggesting independent hybridization events in the evolutionary history of the group. We confirm the hybrid origins of *P. eurystomum*, *P. collenchymatum*, and *E. hungaricus* (Beike et al., 2014; McDaniel et al., 2010; Ostendorf et al., 2021) and reveal the allopolyploid nature of other accessions corresponding to the morphologically distinct species *P. immersum*, *P. pyriforme*, and *P. sphaericum*. This analysis allowed us to explore the limitations of the subgenome phasing approach, Homologizer, relative to other similar methods. For instance, *P. pyriforme*—3410 and 3798 are the first proposed triploid and tetraploid accessions identified in the *P. pyriforme* species complex using a molecular approach, but the accurate circumscription of their subgenomes remains a challenge, particularly given the need to accurately phase subgenomes. Furthermore, our approaches to homeolog and subgenome phasing are novel in that Homologizer is the first tree-based method that phases gene copies using all loci simultaneously. The application of a subgenome phasing approach using target capture data is novel for mosses, which is of particular importance since the haploid phase is sampled. As a result, this study lays the foundation for rapidly developing computational approaches to elucidating reticulate species complexes rather than excluding polyploid taxa when inferring phylogenies.

### Evolutionary origins of known allopolyploids in *Physcomitrium* and *Entosthodon*

The prior evidence for an allopolyploid origin of *E. hungaricus*, *P. eurystomum*, and *P. collenchymatum* (Beike et

al., 2014; McDaniel et al., 2010; Ostendorf et al., 2021) is here significantly strengthened when extended to 50 loci. Furthermore, the ability to phase the subgenomes and estimate their phylogenetic affinities within a broader taxonomic sampling (Medina et al., 2018, 2019) revealed novel hypotheses regarding their progenitor species.

Difficulties in discerning the evolutionary origins of the European endemic *E. hungaricus* suggest that further taxon sampling may be needed to identify progenitors. The two *E. hungaricus* populations sampled here are from Austria and Hungary. The paternal subgenome of both samples is sister to the North American *Entosthodon americanus* in the *E. attenuatus* clade, whereas its maternal subgenome is sister to a clade of North American lineages (Figures 3 and 4; Supplementary Figures S2 and S4). The phylogenetic placement of each paternal subgenome is well resolved with little conflict among component gene tree topologies (Figure 3; Supplementary Figure S4). Given that the maternal subgenome of *E. hungaricus* is sister to a subgenome of an allopolyploid population of *P. pyriforme* in iteration one (Figure 3), the two allopolyploids may share one progenitor not sampled here or that is extinct (Karbstein et al., 2022). However, there is evidence that *P. pyriforme*—3410 subgenomes are incorrectly phased contributing to misinference of *E. hungaricus* subgenome positions. For instance, the clade including the maternal subgenomes of allopolyploid *E. hungaricus* is resolved as sister to the *Physcomitrium* clade with substantial genic conflict in iteration one (Supplementary Figure S2) and ASTRAL phylogenetic inference excluding *P. pyriforme*—3410 subgenomes results in the resolution of the maternal subgenomes in the *Entosthodon* clade (Supplementary Figure S1). In addition, in iteration two *P. pyriforme*—3410 groups with three subgenomes and hence three tips, none of these tips remain sister to a subgenome of *E. hungaricus*—3177 and 3838 (Supplementary Figure S4).

The two subgenomes of *P. eurystomum* are resolved sister to *P. hookeri* and one population (3787) of *P. pyriforme*, respectively. Neither parental subgenome is closely related to *P. sphaericum* as previously suggested (Beike et al., 2014; McDaniel et al., 2010; Ostendorf et al., 2021) although our sampling does not include European populations of the latter species. This morphospecies includes haploid and hybrid populations and neither subgenome of the latter forms a clade with the haploid sample (Figures 3 and 4; Supplementary Figure S4).

Populations morphologically matching the North American endemic *P. collenchymatum* are either haploid (Medina et al., 2018, 2019) or allopolyploid (Beike et al., 2014; McDaniel et al., 2010; Ostendorf et al., 2021). Sampling numerous nuclear loci for the IMSC stock culture (accession number 40061) of *P. collenchymatum*—5274, which was used in the latter studies, confirms extensive heterozygosity and hence allopolyploidy (Supplementary Table S1). Furthermore, heterozygosity in the nuclear barcoding loci 4780 and 7379 (Supplementary Figure S3) confirms that the original collection serves as a voucher of the IMSC culture (i.e., Homberg 1155), and which is morphologically consistent with *P. collenchymatum*, is an allopolyploid. Consequently, the *P. collenchymatum* morphotype is represented by two distinct lineages, one haploid and one allopolyploid, the latter resulting from a hybridization event involving a haploid species of the *P. patens* species complex and a yet unidentified haploid lineage sister to the clade comprising *P. readeri*, the progenitor to the *P. japonicum* and

*P. pyriforme* species complexes (Figures 3 and 4). These parental lineages also gave rise via hybridization to *P. immersum* (see below), although with reversed progenitor roles. Whether the name *P. collenchymatum* should apply to the haploid or the allopolyploid lineages must await the characterization on the type specimen of *P. collenchymatum* collected in Missouri (Gier, 1955), as both ploidies are herein reported from this state.

### Newly identified allopolyploid lineages

*P. immersum* is here newly identified as another allodiploid species of *Physcomitrium*. Its two subgenomes are highly similar to those of the hybrid accession of *P. collenchymatum*—5274. However, phylogenetic inferences from plastid data reveal opposite progenitor roles, with one parental lineage in the *P. patens* complex serving as the maternal parent of *P. immersum* and the paternal parent of *P. collenchymatum* (Figure 3; Supplementary Figure S9). Distinct hybrid origins of these species, despite the sister relationships of their respective subgenomes, is consistent with their conspicuously distinct sporophytes (e.g., capsule immersed in *P. immersum* and exserted in *P. collenchymatum*). Repeated hybridization resulting in multiple morphologically distinct allopolyploid lineages is known in ferns and angiosperms (Patel et al., 2018; Sessa et al., 2018; Soltis et al., 2014) and may often represent independent hybridization with inverted parental roles (Sessa et al., 2018).

*P. sphaericum* is currently regarded as a haploid species, based on a genotyped population from Germany (McDaniel et al., 2010). In Asia, populations matching this species' morphotype are either haploid (3115) or allopolyploid (3141, 3550, 3672, and 3815; Figure 3). For the latter, the paternal subgenome is in the *P. japonicum* species complex, whereas the maternal parent may be closely related to the haploid *P. collenchymatum* in the *P. patens* species complex (Figures 3 and 4; Supplementary Figure S4). The haploid and allopolyploid Asian populations of *P. sphaericum* are morphologically indistinguishable. In addition, given strong morphological similarity between *P. sphaericum* and *P. japonicum* (Xing-jiang et al., 2003), if indeed the allopolyploid *P. sphaericum* results from hybridization between these two haploid lineages, one would expect it to be perhaps morphologically indistinguishable from either progenitor.

*P. pyriforme* has a long-documented history of polyploidy, with Patel et al. (2021) noting karyotypic evidence for at least six cytotypes ( $n = 9, 18, 27, 36, 45,$  and  $54$ ; see Fritsch, 1991). In the present study, three of the nine *P. pyriforme* populations sampled are allopolyploid, and *P. pyriforme*—3410 and 3798 appear to be triploid and tetraploid, respectively (Supplementary Table S4). However, we find substantial challenges to inferring progenitors for these higher-level (above diploid) polyploids, evident in the implementation of iterations two and three. While some studies of reticulate evolution phase subgenomes from individuals with ploidies greater than two and incorporate associated subgenomes into phylogenetic analysis, they largely benefit from the use of long-read sequencing approaches such as PacBio, which obviates the necessity of phased sequence assembly (Dauphin et al., 2018; Rothfels et al., 2017). With short-reads-based target capture sequencing employed here and increasingly in phylogenetics (Hale et al., 2020), misassembly and uneven sequencing depth across genes must be contended with and present a challenge to allele phasing in GATK and hence phylogenetic resolution

(Eaton et al., 2017; Karbstein et al., 2022). Often, short reads can lead to multiple phased blocks, complicating our ability to phase variants in relationship to variants in other discontinuous blocks (Kates et al., 2018; Šlenker et al., 2021). Recent studies leveraging short-read target-capture data in subgenome phasing and parsing allopolyploid origins among plants manage these complications with various approaches including only retaining the largest phase block (Šlenker et al., 2021; Tiley et al., 2021). We have dealt with the matter of linking variants by using reads-backed phasing and retention of only the largest phased block (Figure 2). Nonetheless, without sufficient read depth, accurate phasing of each subgenome becomes increasingly difficult as ploidy increases (Šlenker et al., 2021). Consequently, the incorporation of more than two genomes for the triploid and tetraploid samples in our phylogenetic analysis (Supplementary Figure S4), may contribute to the substantial gene conflict at nodes representing divergence of the subgenomes of *P. pyriforme*—3798 and *P. pyriforme*—3410, respectively. For instance, in iteration two only 11 genes support the relationship of one subgenome of *P. pyriforme*—3410 to the *P. pyriforme* clade (Supplementary Figure S4). Subgenome phasing could be improved with additional target capture data to increase read-depth. Phylogenetic resolution of subgenomes for ploidies higher than two is a critical area of improvement for both read-backed phasing within loci and similar target capture-based approaches to subgenomic assignment and phylogenetic analysis.

Although the phylogenetic position of each subgenome of *P. pyriforme*—3410 and 3798 varies depending on the number of subgenomes phased, and with substantial gene conflict, we can reliably infer the identity and correct position of the maternal subgenome using the organellar phylogeny presented in Medina et al. (2018). Plastid phylogenetic inference from Medina et al. (2018) resolves *P. pyriforme*—3410 as sister to *E. hungaricus*—3177 and 3838, which is in conflict with all *P. pyriforme*—3410 subgenome positions when represented by three tips in the iteration two (“fixed diploid”) analysis (Supplementary Figure S4), but consistent with the phylogenetic position of the inferred maternal subgenome of 3410 when it is represented by two subgenomes in the iteration one analysis (Figure 3). Still, both analyses indicate substantial gene conflict at nodes ancestral to each 3410 subgenome, suggesting that these conflicts are potentially a result of incomplete lineage sorting or, more problematically, recombination among subgenomes (Oxelman et al., 2017).

### Implications of allopolyploidy for moss speciation and diversification

With the inclusion of allopolyploid lineages in taxonomically broad, multi-locus analyses, it is possible to address several core questions in moss evolution. Here, we explore the effect of polyploidization on diversification and trait evolution among polyploid complexes, as well as the extent of hybridization among long-diverged lineages.

In some cases, polyploidy in angiosperms is associated with a higher rate of speciation (Jiao et al., 2014; Landis et al., 2018; Van de Peer et al., 2017). Others advocate the dead-end hypothesis (Mayrose et al., 2011; 2015). Within the Funariaceae, no two morphologically distinct allopolyploid species share a unique pair of common ancestors with the same progenitor roles; Thus, whereas allopolyploidization may be a frequent speciation mechanism in the Funariaceae, even when involving identical progenitor lineages as in the

case of *P. immersum* and *P. collenchymatum*, genome merger does not appear to be followed by subsequent speciation events. However, the lack of diversification may be attributable, in part, to the recency of allopolyploid divergence or a need for more dense taxon sampling (Dauphin et al., 2018; Escudero et al., 2014; Mayrose et al., 2015). Relying on maternal node age estimates in Medina et al. (2018) from plastid phylogenetic analysis of populations sampled in the present study, allopolyploid lineages likely arose within the last 5 my (*E. hungaricus*) or even within only the last 1 my (e.g., *P. sphaericum*, *P. immersum*, and *P. eurystomum*). Given the relatively recent divergence of allopolyploids in the Funariaceae, it is likely that all progenitors are extant. Furthermore, the taxon sampling here is narrow with respect to the 300 species described in the Funariaceae and the population sampling is limited to one or two populations per species. Progenitor hypotheses would be refined by an expanded taxon and population sampling.

Allopolyploid genomes, by combining homeologs from divergent lineages, are historically thought to present physiological challenges to survival and reproduction (Buggs et al., 2008; Comai, 2005). Accordingly, it is expected that hybridization is less common between more divergent lineages because of the accumulation of phenotypic divergence and hence potential reproductive barriers (Coyne & Orr, 2004; Moyle et al., 2004; Presgraves, 2002). Nonetheless, intergeneric hybrids among lineages that diverged up to 50 mya are known in ferns and angiosperms (Forster & Dale, 1983; Lehtonen, 2018; Rothfels et al., 2015). In our inferences, the parental genomes of each allopolyploid vary in their date of divergence. The most extreme is *E. hungaricus*, a fertile taxon that combines genomes of parental lineages that diverged approximately 30 mya (Medina et al., 2018). The proposed progenitor genomes of *P. eurystomum*, *P. sphaericum*, and *P. immersum* diverged less than 15 mya, though here we have not estimated the date of the hybridization event itself.

Since we identified samples based on the morphological species concept, our inferences reveal that at least three morphospecies (i.e., *P. collenchymatum*, *P. pyriforme*, and *P. sphaericum*) comprise both haploid and allopolyploid genotypes, challenging diversity estimates, taxonomic delimitation and nomenclatural assignment. The extent to which morphological features can be used to reliably identify species in the *Physcomitrium–Entosthodon* complex requires further study. The size of spores is correlated with genome size in both ferns and mosses and hence may be valuable in identifying polyploids in a narrow taxonomic context (Barrington et al., 2020; Košnar et al., 2012). The taxonomic treatment in the complex may need to follow the example of *Sphagnum*, in which hybridization has been frequently demonstrated (Cronberg & Natcheva, 2002; Karlin & Smouse, 2017; Ricca & Shaw, 2010; Ricca et al., 2011). For example, the *Sphagnum subsecundum* complex includes haploid and allodiploid species as well as morphospecies that include multiple cytotypes (Shaw et al., 2012). In the *Physcomitrium–Entosthodon* complex, detailed surveys of morphological characters are needed to make similar progress. Because the type specimens of the taxa in the present study are 60–120 years old and sometimes consist of a few tiny plants (and hence, unavailable for destructive sampling), estimating ploidy, even based on heterozygosity of barcodes, would be difficult or impossible. A critical revision of morphological characters may diagnose lineages of distinct ploidy and hence could allow assigning a morphotype

to a type specimen and applying the correct name. However, these studies are still lacking and, as currently circumscribed, morphospecies comprise populations of haploid and hybrid populations.

### Challenges to allopolyploid discovery, homeolog phasing, and homeolog sorting

The inclusion of allopolyploid lineages along with haploid taxa in a phylogeny required two phasing steps: phasing of reads within loci and of homeologs within samples. We initially identified potential allopolyploids via an abundance of “paralog warnings” in HybPiper. We then characterized the degree of heterozygosity after mapping reads to the HybPiper output and identifying within-sample variants with GATK. Our results suggest the paralog warning method was not perfect; several samples, including *E. hungaricus*—3177, did not have many paralog warnings but were nevertheless highly heterozygous when put through the GATK workflow. It is likely that low sequencing depth can confound the ability of HybPiper version 1.3 to identify multiple copies, because it only does so when multiple full-length contigs are recovered for a gene. Notably, version 2 of HybPiper purports to alleviate this shortcoming and will produce warnings if any part of the assembled sequence has multiple representative contigs (see [github.com/HybPiper/wiki](https://github.com/HybPiper/wiki)). An alternative approach to allopolyploid discovery with target-capture data may be to use HybPhaser (Nauheimer et al., 2021) for a large-scale review of target-capture data to identify homeologs and to discover hidden paralogs and contaminants.

Our read-backed phasing workflow using GATK and WhatsHap proved generally effective in phasing subgenomes, though it presented some challenges when phasing subgenomes of higher-level ploidies. The workflow often resulted in fully phased sequences across thousands of bases, suggesting that our nuclear target capture genes have many short introns that our relatively long (300 bp) paired-end Illumina reads could sequence through. The within-locus phasing (assembly) workflow takes advantage of programs written for phasing heterozygosity in diploid-dominant organisms. For triploids, in our iteration two analysis (fixed-diploid), we’ve modified our read-backed phasing workflow to incorporate the “polyphase” function in WhatsHap allowing for the phasing of haplotypes at ploidies greater than two (Schrinner et al., 2020). Whereas three distinct haplotypes were phased for *P. pyriforme*—3410 and 3798, substantial gene conflict is evident in the phylogenetic position of each subgenome (Supplementary Figure S4). Detecting autopolyploids and unbalanced allopolyploids such as triploids using this approach may be much more difficult by requiring divergence between subgenomes substantial enough for detection as unique homeologs among multiple genes as well as sufficient read depth across variants (Karlin & Smouse, 2017). This challenge was reflected in the inconsistency in posterior probability across Homologizer runs when our two higher ploidy samples were coded as triploids in the iteration two (fixed diploid) analysis. It is possible that recombination among subgenomes, as has been previously reported in triploid *Sphagnum* mosses (Karlin & Smouse, 2017), may have occurred following genome doubling in the polyploid *P. pyriforme* lineages. The phenomenon of “subgenome dropout” may also contribute to the observed gene conflict by preventing recovery of all loci for each subgenome (Adams & Wendel, 2005; Grigoryan et al., 2021; Leitch & Bennett,

2004). This could contribute to the lack of convergence in the Homologizer and the high levels of subgenome discordance observed in the species tree reconstructed from the iteration two Homologizer results (Supplementary Figure S4).

Newly added features in programs including WhatsHap “polyphase” (Schrunner et al., 2020) and H-POP-G (Xie et al., 2016) are promising, but more studies using these features to analyze polyploid target capture data are needed to explore the utility of these tools in confidently calling variants (homeologs in the present study) with lower sequencing depth. Our target capture data were generated from the enrichment of genomic DNA extracted from gametophytic tissue exclusively. This simplifies analysis by necessitating only the assignment of alleles to monoploid genomes. Analyzing data derived from sporophyte allodiploids, as we did for those of gametophytes, would be more complicated as it requires allele phasing and subgenome assignment for up to six variants at each locus. This is a key issue in that one major benefit of target capture sequencing is that it can be an effective method for generating subgenomic data even for fragmented DNA derived from herbarium specimens (Viruel et al., 2016, 2019).

Our results rely on subgenome phasing in the RevBayes plugin Homologizer to phase among-loci, and we highlight its utility for the integration of polyploids in phylogenetic systematic studies. However, the scaling potential of the Homologizer is limited without future improvements. For instance, with 50 genes (representing about 188 Kb) and 74 tips, RevBayes took over 6 days to complete 10,000 generations on a high-performance computer.

In this study, we observe much more discordance among gene trees than Medina et al. (2019). Although phylogenetic inference in Medina et al. (2019) is based on 648 genes compared to 50 in the present study, much of the discordance we observed is likely attributable to the inclusion of phased allopolyploid subgenomes in gene tree construction. This is evident when comparing the phylogeny excluding allopolyploids (Supplementary Figure S10) with the phylogeny including all samples (Figure 3; Supplementary Figure S4), both based on the same 50 genes. Many of the discordant gene trees result from phased subgenome sequences resolved as sister to each other, which suggests that some genes may have more complicated inheritance, i.e., the loss or duplication of one homeolog after polyploidization. Investigating the mode of polyploid inheritance and homeologous exchange would require phasing many more genes than can currently be practically supported by Homologizer. One approach could be to conduct iterative phasing by phasing multiple sets of genes in parallel (e.g., 50 sets of 50 genes each), and then using a final phasing of one gene per set to connect all homeologs into subgenomes for final analysis. We also limited our search to genes with complete data for all samples and increasing gene sampling will need to accommodate missing data.

Finally, although target capture sequencing of low and single-copy nuclear markers is critical to the identification and phylogenetic analysis of allopolyploids and allopolyploid subgenomes, our study also highlights the importance of highly supported organellar phylogenies to the identification of allopolyploid progenitors. For each polyploid hybrid species, we resolved one subgenome in a position consistent with that of the organellar genome proposed by Medina et al. (2018), except for *P. collenchymatum*—5274, which was not sampled in the previous study. Although the plastid exome likely represents just one coalescent history, it is valuable for

identifying the maternal lineage in hybrids and determining whether polyploids have recurrent formation.

## Conclusions

The diversification of the Funariaceae is characterized by significant reticulation among multiple species complexes, and even between genera. The Funariaceae exemplifies growing evidence for the significance of polyploidy in moss speciation and evolution (Sawangproh & Cronberg, 2021). Numerous studies (Barker et al., 2016; Schmickl et al., 2017; Soltis et al., 2009) in plants suggest that polyploids constitute a substantial proportion of total plant species diversity (Wood et al., 2009) and hence accurate circumscription of polyploid subgenomes facilitates the testing of questions in evolutionary biology pertaining to diversification. Here, we demonstrate the importance of phasing and sorting allopolyploid subgenomes for phylogenomic analysis. The implementation of Homologizer as a tool for the incorporation of allopolyploids in phylogenetic analysis is a critical step toward elucidating the role of allopolyploidy in the evolution and diversification of plants.

## Supplementary material

Supplementary material is available online at *Evolution*.

## Data availability

All nucleotide alignments, RevBayes control files, gene trees, and species trees are available via the Dryad Digital repository (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.0p2ngf23f>). For target capture baits used, see Medina et al. (2019) and Dryad Digital repository (<https://datadryad.org/stash/dataset/doi:10.5061/dryad.8rq9465>).

## Author contributions

N.P.: Target capture data collection for some samples and Sanger sequencing data collection. Phylogenetic analyses, interpretation, and manuscript composition. R.M.: Target capture data collection for some samples, specimen identification, and advising on analysis and interpretation. L.D.W.: Support in Homologizer analysis and documentation of computational protocols. O.L.: DNA extraction and Sanger sequencing of *Physcomitrium collenchymatum* cultures and vouchers. Analysis of sequences to confirm identity. B.G.: Manuscript composition and advising on analysis and interpretation. Figure generation. Specimen identification. M.G.J.: Data analysis using HybPiper and Homologizer. Generation of sequence statistics. Manuscript composition. Figure generation.

## Funding

This research was supported by the National Science Foundation with grants to B.G. (NSF-DEB 1753811), R.M. (NSF-DEB 1753673), and M.G.J. (NSF-DEB 1753800).

*Conflict of interest:* The authors have no conflict of interest to report.

## Acknowledgments

We thank the IMSC for providing cultures of *Physcomitrium collenchymatum* and John Atwood and Bruce Allen (MO)

for the loan of the original vouchers of these cultures. We thank C. Rothfels and W. Freyman for assistance running Homologizer.

## References

- Adams, K. L., & Wendel, J. F. (2005). Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, 8(2), 135–141. <https://doi.org/10.1016/j.pbi.2005.01.001>
- Barker, M. S., Arrigo, N., Baniaga, A. E., Li, Z., & Levin, D. A. (2016). On the relative abundance of autopolyploids and allopolyploids. *The New Phytologist*, 210(2), 391–398. <https://doi.org/10.1111/nph.13698>
- Barrington, D. S., Patel, N. R., & Southgate, M. W. (2020). Inferring the impacts of evolutionary history and ecological constraints on spore size and shape in the ferns. *Applications in Plant Sciences*, 8(4), e11339. <https://doi.org/10.1002/aps3.11339>
- Beike, A. K., von Stackelberg, M., 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 Evolutionary Biology*, 14(1), 1–20.
- Buggs, R. J., Soltis, P. S., Mavrodiev, E. V., Symonds, V. V., & Soltis, D. E. (2008). Does phylogenetic distance between parental genomes govern the success of polyploids. *Castanea*, 73(2), 74–93.
- Chen, Z., Sreedasyam, A., Ando, A., Song, Q., Santiago, L., Hulse-Kemp, A., Ding, M., Ye, W., Kirkbride, R., Jenkins, J., Plott, C., Lovell, J., Lin, Y., Vaughn, R., Liu, B., Simpson, S., Scheffler, B., Wen, L., Sasaki, C., ... Schmutz, J. (2020). Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nature Genetics*, 52(5), 525–533.
- Clark, L. V., & Schreier, A. D. (2017). Resolving microsatellite genotype ambiguity in populations of allopolyploid and diploidized autopolyploid organisms using negative correlations between allelic variables. *Molecular Ecology Resources*, 17(5), 1090–1103. <https://doi.org/10.1111/1755-0998.12639>
- Comai, L. (2005). The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6(11), 836–846. <https://doi.org/10.1038/nrg1711>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation* (Vol. 37). Sinauer Associates.
- Cronberg, N., & Natcheva, R. (2002). Hybridization between the peat mosses, *Sphagnum capillifolium* and *S. quinquefarium* (Sphagnaceae, Bryophyta) as inferred by morphological characters and isozyme markers. *Plant Systematics and Evolution*, 234(1), 53–70.
- Dauphin, B., Grant, J. R., Farrar, D. R., & Rothfels, C. J. (2018). Rapid allopolyploid radiation of moonwort ferns (*Botrychium*; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions. *Molecular Phylogenetics and Evolution*, 120, 342–353. <https://doi.org/10.1016/j.ympev.2017.11.025>
- Doyle, J. J. (2021). Defining coalescent genes: theory meets practice in organelle phylogenomics. *Systematic Biology*, 71(2), 476–489. <https://doi.org/10.1093/sysbio/syab053>
- Drummond, A. J., & Bouckaert, R. R. (2015). *Bayesian evolutionary analysis with BEAST*. Cambridge University Press.
- Eaton, D. A. R., Spriggs, E. L., Park, B., & Donoghue, M. J. (2017). Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. *Systematic Biology*, 66(3), 399–412. <https://doi.org/10.1093/sysbio/syw092>
- Escudero, M., Martín-Bravo, S., Mayrose, I., Fernández-Mazuecos, M., Fiz-Palacios, O., Hipp, A. L., Pimentel, M., Jiménez-Mejías, P., Valcárcel, V., Vargas, P., & Luceño, M. (2014). Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. *PLoS One*, 9(1), e85266.
- Fife, A. J. (1985). A generic revision of the Funariaceae (Bryophyta: Musci) Part I. *Journal of the Hattori Botanical Laboratory*, 58, 149–196.
- Forster, B. P., & Dale, J. E. (1983). Effects of parental embryo and endosperm mitotic cycle times on development of hybrids between barley and rye. *Annals of Botany*, 52(5), 613–620. <https://doi.org/10.1093/oxfordjournals.aob.a086617>
- Francioli, L. C., Cretu-Stancu, M., Garimella, K. V., Fromer, M., Kloosterman, W. P., Samocha, K. E., Neale, B. M., Daly, J., Bankes, E., DePristo, M. A., & de Bakker, P. I. (2017). A framework for the detection of de novo mutations in family-based sequencing data. *European Journal of Human Genetics*, 25(2), 227–233.
- Frey, W., Frahm, J. P., Fischer, E., & Lobin, W. (2006). In T. L. Blockeel (Ed.), *The liverworts, mosses and ferns of Europe* (English edition). Harley Books.
- Freyman, W. A., Johnson, M. G., & Rothfels, C. J. (2023). Homologizer: Phylogenetic phasing of gene copies into polyploid subgenomes. *Methods in Ecology and Evolution*, 14(5), 1230–1244. <https://doi.org/10.1111/2041-210x.14072>
- Fritsch, R. (1991). Index to bryophyte chromosome counts. *Bryophytum Bibliotheca*, 40, 1–352.
- Gier, L. G. (1955). *Physcomitrium collenchymatum*. *Transactions of the Kansas Academy of Science (1903-)*, 58(3), 330–333. <https://doi.org/10.2307/3626099>
- Goffinet B. (2007). *Aphanorrhagma* Sull. *Physcomitrella* Bruch & Schimper. Pages: 181–182, respectively. In Flora of North America Editorial Committee (Eds.), *Flora of North America: North of Mexico* (Vol. 27, pp. 194–195). Flora of North America Editorial Committee.
- Grigoryan, M. Y., Bobrov, A. A., Brunton, D. F., Volkova, P. A., Logacheva, M. D., & Neretina, T. V. (2021). Next generation DNA sequencing reveals allopolyploid origin of decaploid *Isoetes lacustris* (Isoëtaceae). *Aquatic Botany*, 170, 103326. <https://doi.org/10.1016/j.aquabot.2020.103326>
- Grover, C. E., Gallagher, J. P., Jareczek, J. J., Page, J. T., Udall, J. A., Gore, M. A., & Wendel, J. F. (2015). Re-evaluating the phylogeny of allopolyploid *Gossypium* L. *Molecular Phylogenetics and Evolution*, 92, 45–52. <https://doi.org/10.1016/j.ympev.2015.05.023>
- Hale, H., Gardner, E. M., Viruel, J., Pokorny, L., & Johnson, M. G. (2020). Strategies for reducing per-sample costs in target capture sequencing for phylogenomics and population genomics in plants. *Applications in Plant Sciences*, 8(4), e11337. <https://doi.org/10.1002/aps3.11337>
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35, 518–522.
- Höhna, S., Landis, M. J., Heath, T. A., Boussau, B., Lartillot, N., Moore, B. R., Huelsenbeck, J. P., & Ronquist, F. (2016). RevBayes: Bayesian phylogenetic inference using graphical models and an interactive model-specification language. *Systematic Biology*, 65(4), 726–736. <https://doi.org/10.1093/sysbio/syw021>
- Jiao, Y., Li, J., Tang, H., & Paterson, A. H. (2014). Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. *Plant Cell*, 26(7), 2792–2802. <https://doi.org/10.1105/tpc.114.127597>
- Johnson, M. G., Gardner, E. M., Liu, Y., Medina, R., Goffinet, B., Shaw, A. J., Zeregs, N. J. C., & Wickett, N. J. (2016). HybPiper: Extracting coding sequence and introns for phylogenetics from high throughput sequencing reads using target enrichment. *Applications in Plant Sciences*, 4(7), 1600016.
- Jorgensen, S. A., & Barrington, D. S. (2017). Two Beringian origins for the allotetraploid fern *Polystichum braunii* (Dryopteridaceae). *Systematic Botany*, 42(1), 6–16. <https://doi.org/10.1600/036364417x694557>
- Junier, T., & Zdobnov, E. M. (2010). The Newick utilities: High-throughput phylogenetic tree processing in the UNIX shell. *Bioinformatics*, 26(13), 1669–1670. <https://doi.org/10.1093/bioinformatics/btq243>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermini, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, 14(6), 587–589. <https://doi.org/10.1038/nmeth.4285>
- Karbstein, K., Tomasello, S., Hodač, L., Wagner, N., Marinček, P., Barke, B. H., Paetzold, C., & Hörandl, E. (2022). Untying Gordian knots: Unraveling reticulate polyploid plant evolution by genomic

- data using the large *Ranunculus auricomus* species complex. *The New Phytologist*, 235(5), 2081–2098. <https://doi.org/10.1111/nph.18284>
- Karlin, E. F., & Smouse, P. E. (2017). Allo-allo-triploid *Sphagnum* × *falcatum*: Single individuals contain most of the Holantarctic diversity for ancestrally indicative markers. *Annals of Botany*, 120(2), 221–231. <https://doi.org/10.1093/aob/mcw269>
- Kates, H. R., Johnson, M. G., Gardner, E. M., Zerega, N. J., & Wickett, N. J. (2018). Allele phasing has minimal impact on phylogenetic reconstruction from targeted nuclear gene sequences in a case study of *Artocarpus*. *American Journal of Botany*, 105(3), 404–416.
- Katoh, K., & Frith, M. C. (2012). Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics*, 28(23), 3144–3146. <https://doi.org/10.1093/bioinformatics/bts578>
- Košnar, J., Herbstová, M., Kolář, F., Koutecký, P., & Kučera, J. (2012). A case study of intragenomic ITS variation in bryophytes: assessment of gene flow and role of polyploidy in the origin of European taxa of the *Tortula muralis* (Musci: Pottiaceae) complex. *Taxon*, 61(4), 709–720. <https://doi.org/10.1002/tax.614001>
- Landis, J. B., Soltis, D. E., Li, Z., Marx, H. E., Barker, M. S., Tank, D. C., & Soltis, P. S. (2018). Impact of whole-genome duplication events on diversification rates in angiosperms. *American Journal of Botany*, 105(3), 348–363. <https://doi.org/10.1002/ajb2.1060>
- Lehtonen, S. (2018). × *Lindsaeosoria flynnii* (Lindsaeaceae), another confirmed example of deep hybridization among the ferns. *American Fern Journal*, 108(1), 7–18. <https://doi.org/10.1640/0002-8444-108.1.7>
- Leitch, I. J., & Bennett, M. D. (2004). Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society*, 82(4), 651–663. <https://doi.org/10.1111/j.1095-8312.2004.00349.x>
- Lohaus, R., & Van de Peer, Y. (2016). Of dups and dinos: Evolution at the K/Pg boundary. *Current Opinion in Plant Biology*, 30, 62–69. <https://doi.org/10.1016/j.pbi.2016.01.006>
- Martin, M., Patterson, M., Garg, S., Fischer, S., Pisanti, N., Klau, G. W., Schönhuth, N., & Marschall, T. (2016). WhatsHap: Fast and accurate read-based phasing. *bioRxiv*. <https://doi.org/10.1101/085050>
- Mayrose, I., Zhan, S. H., Rothfels, C. J., Arrigo, N., Barker, M. S., Rieseberg, L. H., & Otto, S. P. (2015). Methods for studying polyploid diversification and the dead-end hypothesis: A reply to Soltis et al. (2014). *New Phytologist*, 206(1), 27–35.
- Mayrose, I., Zhan, S., Rothfels, C. J., Magnuson-Ford, K., Barker, M., Rieseberg, L. H., & Otto, S. (2011). Recently formed polyploid plants diversify at lower rates. *Science*, 333, 1257–1257.
- McDaniel, S. F., Von Stackelberg, M., Richardt, S., Quatrano, R. S., Reski, R., & Rensing, S. A. (2010). The speciation history of the *Physcomitrium*—*Physcomitrella* species complex. *Evolution*, 64(1), 217–231. <https://doi.org/10.1111/j.1558-5646.2009.00797.x>
- McIntosh, T. T. (2007). Funariaceae. Pages: 180–189. In Flora of North America Editorial Committee (Eds.), *Flora of North America: North of Mexico* (Vol. 27). Flora of North America Editorial Committee.
- 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 *Aphanorrhagma*. *Journal of Systematics and Evolution*, 57(4), 404–417. <https://doi.org/10.1111/jse.12516>
- Medina, R., Johnson, M., 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. *Molecular Phylogenetics and Evolution*, 120, 240–247. <https://doi.org/10.1016/j.ympev.2017.12.002>
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, 37(5), 1530–1534. <https://doi.org/10.1093/molbev/msaa015>
- Moyle, L. C., Olson, M. S., & Tiffin, P. (2004). Patterns of reproductive isolation in three angiosperm genera. *Evolution*, 58(6), 1195–1208. <https://doi.org/10.1111/j.0014-3820.2004.tb01700.x>
- Natcheva, R., & Cronberg, N. (2004). What do we know about hybridization among bryophytes in nature? *Canadian Journal of Botany*, 82(12), 1687–1704. <https://doi.org/10.1139/b04-139>
- Nauheimer, L., Weigner, N., Joyce, E., Crayn, D., Clarke, C., & Nargar, K. (2021). HybPhaser: A workflow for the detection and phasing of hybrids in target capture data sets. *Applications in Plant Sciences*, 9(7), e1144.
- Ostendorf, A. K., Van Gessel, N., Malkowsky, Y., Sabovljevic, M. S., Rensing, S. A., Roth-Nebelsick, A., & Reski, R. (2021). Polyploidization within the Funariaceae—A key principle behind speciation, sporophyte reduction and the high variance of spore diameters? *Bryophyte Diversity and Evolution*, 43(1), 164–179.
- Oxelman, B., Brysting, A. K., Jones, G. R., Marcussen, T., Oberprieler, C., & Pfeil, B. E. (2017). Phylogenetics of allopolyploids. *Annual Review of Ecology, Evolution, and Systematics*, 48(1), 543–557. <https://doi.org/10.1146/annurev-ecolsys-110316-022729>
- Ozkan, H., Levy, A. A., & Feldman, M. (2001). Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops*–*Triticum*) group. *Plant Cell*, 13(8), 1735–1747. <https://doi.org/10.1105/tpc.010082>
- Patel, N., Li, C. X., Zhang, L. B., & Barrington, D. S. (2018). Biodiversity and apomixis: Insights from the East-Asian holly ferns in *Polystichum* section *Xiphopolystichum*. *Molecular Phylogenetics and Evolution*, 127, 345–355. <https://doi.org/10.1016/j.ympev.2018.05.003>
- 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(1), 150–163.
- Picard Toolkit. (2019). Broad Institute, GitHub Repository. <http://broadinstitute.github.io/picard/>
- Presgraves, D. C. (2002). Patterns of postzygotic isolation in *Lepidoptera*. *Evolution*, 56(6), 1168–1183. <https://doi.org/10.1111/j.0014-3820.2002.tb01430.x>
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*, 67(5), 901–904. <https://doi.org/10.1093/sysbio/syy032>
- Ricca, M., & Shaw, A. J. (2010). Allopolyploidy and homoploid hybridization in the *Sphagnum subsecundum* complex (Sphagnaceae: Bryophyta). *Biological Journal of the Linnean Society*, 99(1), 135–151. <https://doi.org/10.1111/j.1095-8312.2009.01340.x>
- Ricca, M., Szövényi, P., Tensch, E. M., Johnson, M. G., & Shaw, A. J. (2011). Interploidal hybridization and mating patterns in the *Sphagnum subsecundum* complex. *Molecular Ecology*, 20(15), 3202–3218. <https://doi.org/10.1111/j.1365-294X.2011.05170.x>
- Rothfels, C. J. (2021). Polyploid phylogenetics. *The New Phytologist*, 230(1), 66–72. <https://doi.org/10.1111/nph.17105>
- Rothfels, C. J., Johnson, A. K., Hovenkamp, P. H., Swofford, D. L., Roskam, H. C., Fraser-Jenkins, C. R., Windham, M., & Pryer, K. M. (2015). Natural hybridization between genera that diverged from each other approximately 60 million years ago. *American Naturalist*, 185(3), 433–442.
- Rothfels, C. J., Pryer, K. M., & Li, F. W. (2017). Next-generation polyploid phylogenetics: Rapid resolution of hybrid polyploid complexes using PacBio single-molecule sequencing. *The New Phytologist*, 213(1), 413–429. <https://doi.org/10.1111/nph.14111>
- Sawangproh, W., & Cronberg, N. (2021). Evidence for interspecific hybridization in bryophytes during pre-molecular and molecular eras. *Bryophyte Diversity and Evolution*, 43(1), 180–205.
- Schmickl, R., Marburger, S., Bray, S., & Yant, L. (2017). Hybrids and horizontal transfer: Introgression allows adaptive allele discovery. *Journal of Experimental Botany*, 68(20), 5453–5470. <https://doi.org/10.1093/jxb/erx297>
- Schrinner, S. D., Mari, R. S., Ebler, J., Rautiainen, M., Seillier, L., Reimer, J. J., Usadel, B., Marschall, T., & Klau, G. W. (2020). Haplotype threading: Accurate polyploid phasing from long reads. *Genome Biology*, 21(1), 252. <https://doi.org/10.1186/s13059-020-02158-1>
- Sessa, E. B., Vicent, M., Chambers, S. M., & Gabriel y Galán, J. M. (2018). Evolution and reciprocal origins in Mediterranean ferns:

- The *Asplenium obovatum* and *A. adiantumnigrum* complexes. *Annals of the Missouri Botanical Garden*, 103(2), 175–187.
- Shaw, A. J., Shaw, B., Ricca, M., & Flatberg, K. I. (2012). A phylogenetic monograph of the *Sphagnum subsecundum* complex (Sphagnaceae) in eastern North America. *The Bryologist*, 115(1), 128–152. <https://doi.org/10.1639/0007-2745-115.1.128>
- Šlenker, M., Kantor, A., Marhold, K., Schmickl, R., Mandakova, T., Lysak, M. A., Perny, M., Cabonová, M., Slovák, M., & Zozomova-Lihova, J. (2021). Allele sorting as a novel approach to resolving the origin of allotetraploids using Hyb-Seq data: A case study of the Balkan mountain endemic *Cardamine barbaraeoides*. *Frontiers in Plant Sciences*, 12, 539.
- Smith, A. J. E. (1978). Cytogenetics, biosystematics and evolution in the Bryophyta. In H. W. Woolhouse (Ed.), *Advances in botanical research* (pp. 195–276). Academic Press.
- Smith, S. A., Moore, M. J., Brown, J. W., & Yang, Y. (2015). Analysis of phylogenomic datasets reveals conflict, concordance, and gene duplications with examples from animals and plants. *BMC Evolutionary Biology*, 15, 150. <https://doi.org/10.1186/s12862-015-0423-0>
- Soltis, D. E., Albert, V. A., Leebens-Mack, J., Bell, C. D., Paterson, A. H., Zheng, C., Sankoff, D., DePamphilis, C., Wall, P. K., & Soltis, P. S. (2009). Polyploidy and angiosperm diversification. *American Journal of Botany*, 96(1), 336–348.
- Soltis, D. E., Soltis, P. S., Schemske, D. W., Hancock, J. F., Thompson, J. N., Husband, B. C., & Judd, W. S. (2007). Autopolyploidy in angiosperms: Have we grossly underestimated the number of species? *Taxon*, 56(1), 13–30.
- Soltis, D. E., Visger, C. J., & Soltis, P. S. (2014). The polyploidy revolution then... and now: Stebbins revisited. *American Journal of Botany*, 101(7), 1057–1078. <https://doi.org/10.3732/ajb.1400178>
- Soltis, P. S., Marchant, D. B., Van de Peer, Y., & Soltis, D. E. (2015). Polyploidy and genome evolution in plants. *Current Opinion in Genetics & Development*, 35, 119–125. <https://doi.org/10.1016/j.gde.2015.11.003>
- Tate, J., Ni, Z., Scheen, A., Koh, J., Gilbert, C., Lefkowitz, D., Chen, Z., Soltis, P., & Soltis, D. E. (2006). Evolution and expression of homeologous loci in *Tragopogon miscellus* (Asteraceae), a recent and reciprocally formed allopolyploid. *Genetics*, 173, 1599–1611.
- Tiley, G. P., Crowl, A. A., Manos, P. S., Sessa, E. B., Solis-Lemus, C., Yoder, A. D., & Burleigh, J. G. (2021). Phasing alleles improves network inference with allopolyploids. *BioRxiv*, <https://doi.org/10.1101/2021.05.04.442457>
- Van de Peer, Y., Mizrachi, E., & Marchal, K. (2017). The evolutionary significance of polyploidy. *Nature Reviews Genetics*, 18(7), 411–424. <https://doi.org/10.1038/nrg.2017.26>
- Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the cloud: Using Docker, GATK, and WDL in Terra* (1st ed.). O'Reilly Media.
- Viruel, J., Conejero, M., Hidalgo, O., Pokorný, L., Powell, R. F., Forest, F., Kantar, M. B., Gomez, M. F., Graham, S. W., Gravendell, B., Wilkin, P., & Leitch, I. J. (2019). A target capture-based method to estimate ploidy from herbarium specimens. *Frontiers in Plant Sciences*, 10, 937.
- Viruel, J., Segarra-Moragues, J. G., Raz, L., Forest, F., Wilkin, P., Sanmartín, I., & Catalán, P. (2016). Late Cretaceous-early Eocene origin of yams (Dioscorea, Dioscoreaceae) in the Laurasian Palearctic and their subsequent Oligocene Miocene diversification. *Journal of Biogeography*, 43, 750–762.
- Vitt, D. H. (1971). The infrageneric evolution, phylogeny, and taxonomy of the genus *Orthotrichum* (Musci) in North America. *Nova Hedwigia*, 21, 683–711.
- Weiß, C. L., Pais, M., Cano, L. M., Kamoun, S., & Burbano, H. A. (2018). nQuire: A statistical framework for ploidy estimation using next generation sequencing. *BMC Bioinformatics*, 19(1), 122. <https://doi.org/10.1186/s12859-018-2128-z>
- Weitemier, K., Straub, S. C., Cronn, R. C., Fishbein, M., Schmickl, R., McDonnell, A., & Liston, A. (2014). Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics. *Applications in Plant Sciences*, 2(9), 1400042.
- Wood, T. E., Takebayashi, N., Barker, M. S., Mayrose, I., Greenspoon, P. B., & Rieseberg, L. H. (2009). The frequency of polyploid speciation in vascular plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106(33), 13875–13879. <https://doi.org/10.1073/pnas.0811575106>
- Wyatt, R., & Odrzykoski, I. J. (1998). On the origins of the allopolyploid moss *Plagiomnium cuspidatum*. *The Bryologist*, 101(2), 263–271.
- Xie, M., Wu, Q., Wang, J., & Jiang, T. (2016). H-PoP and H-PoPG: Heuristic partitioning algorithms for single individual haplotyping of polyploids. *Bioinformatics*, 32(24), 3735–3744. <https://doi.org/10.1093/bioinformatics/btw537>
- Xing-Jiang, L., He, S., & Da-cheng, Z. (2003). Funariaceae. In S. He (Ed.), *Moss flora of China. English Version. Vol. 3 Grimmiaceae-Tetraphidaceae*. Science Press & Missouri Botanical Garden.
- Zhang, C., Sayyari, E., & Mirarab, S. (2017). ASTRAL-III: Increased scalability and impacts of contracting low support branches. In J. Meidanis, & L. Nakhleh (Eds.), *Comparative genomics. RE-COMB-CG 2017. Lecture notes in Computer Science* (Vol. 10562, pp. 53–75). Springer International Publishing.