Received: 30 October 2020 | Revised: 2 July 2021 | Accepted: 19 July 2021

DOI: 10.1111/1755-0998.13478

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Transcriptome-wide SNPs for *Botrychium lunaria* ferns enable fine-grained analysis of ploidy and population structure

Vinciane Mossion[1] | Benjamin Dauphin[1,2] | Jason Grant[1] | Michael Kessler[3] | Niklaus Zemp[4] | Daniel Croll[1]

[1]Laboratory of Evolutionary Genetics, University of Neuchâtel, Neuchâtel, Switzerland

[2]Swiss Federal Research Institute WSL, Birmensdorf, Switzerland

[3]Department of Systematic and Evolutionary Botany, University of Zürich, Zurich, Switzerland

[4]Genetic Diversity Centre (GDC), ETH Zurich, Zurich, Switzerland

Correspondence
Daniel Croll, Laboratory of Evolutionary Genetics, University of Neuchâtel, Neuchâtel, Switzerland.
Email: daniel.croll@unine.ch

Funding information
Université de Neuchâtel

Abstract

Ferns are the second most diverse group of land plants after angiosperms. Extant species occupy a wide range of habitats and contribute significantly to ecosystem functioning. Despite the importance of ferns, most taxa are poorly covered by genomic resources and within-species studies based on high-resolution markers are entirely lacking. The genus *Botrychium* belongs to the family Ophioglossaceae, which includes species with very large genomes and chromosome numbers (e.g., *Ophioglossum reticulatum* $2n = 1520$). The genus has a cosmopolitan distribution with 35 species, half of which are polyploids. Here, we establish a transcriptome for *Botrychium lunaria* (L.) Sw., a diploid species with an extremely large genome of about ~19.0–23.7 Gb. We assembled 25,677 high-quality transcripts with an average length of 1,333 bp based on deep RNA-sequencing of a single individual. We sequenced 11 additional transcriptomes of individuals from two populations in Switzerland, including the population of the reference individual. Based on read mapping to reference transcript sequences, we identified 374,463 single nucleotide polymorphisms (SNPs) segregating among individuals for an average density of 14 SNPs per kilobase. We found that all 12 transcriptomes were most likely from diploid individuals. The transcriptome-wide markers provided unprecedented resolution of the population genetic structure, revealing substantial variation in heterozygosity among individuals. We also constructed a phylogenomic tree of 92 taxa representing all fern orders to ascertain the placement of the genus *Botrychium*. High-quality transcriptomic resources and SNP sets constitute powerful population genomic resources to investigate the ecology, and evolution of fern populations.

KEYWORDS
ferns, Ophioglossaceae, ploidy, population genomics, SNP development, transcriptome assembly

1 | INTRODUCTION

Ferns (Polypodiopsida) are a highly diverse group of cosmopolitan vascular plants and present in most climates. Around 85% of the species richness is found in the tropics (Page, 2002; Ranker & Haufler, 2008), where ferns have diversified into a multitude of habitats including deserts, grasslands, forest understory, mountainous regions, and aquatic environments (Mehltreter et al., 2010). Ferns play key roles in ecosystem functioning, including serving as a habitat for invertebrates (Ellwood & Foster, 2004), shaping plant recolonization of disturbed habitats (Walker, 1994) and influencing the composition of tree species communities (George & Bazzaz, 1999a, 1999b). Most ferns are characterized by two independent life stages (i.e., gametophytic and sporophytic; Pinson et al., 2016). Ferns have complex and idiosyncratic life cycles. For example, many fern species are capable of versatile reproductive modes (Sessa et al., 2012) including apomixis (Grusz, 2016), sporophytic and gametophytic selfing, and outcrossing (Barker & Wolf, 2010; Haufler et al., 2016). Flexibility in the reproductive mode plays a crucial role in the foundation and persistence of isolated populations from a few spores, and thus likely contributed to the evolution of fern lifestyles. In addition, the reproductive mode likely impacts the tempo of evolution and lineage diversification.

Phylogenetic analyses have resolved the position of ferns as the sister group to seed plants (Pryer et al., 2001) and as the second earliest diverging lineage of vascular land plants (Pryer et al., 2004; Raubeson & Jansen, 1992). The root age of ferns has been estimated to be c. 360–431 million years, underlining the deep divergence among ferns lineages (Des Marais et al., 2003; Lehtonen et al., 2017; Magallón et al., 2013; Pryer et al., 2004; Qi et al., 2018; Rothfels et al., 2015; Testo & Sundue, 2016; Wikström & Kenrick, 2001; Zhong et al., 2014). Fern phylogenies have been mostly established based on chloroplast markers (Grewe et al., 2013; Kuo et al., 2011; Lu et al., 2015; Rai & Graham, 2010; Schuettpelz & Pryer, 2007; Testo & Sundue, 2016), sometimes combined either with mitochondrial markers (Knie et al., 2015) or nuclear markers (Pryer et al., 2001, 2004) or both (Qiu et al., 2007). The heavy reliance on plastid and mitochondrial markers largely prevents analyses of reticulation events (i.e., hybridization and introgression). Yet, polyploidizations occur at one third of the fern speciation rate (Wood et al., 2009). In contrast, a series of recent phylogenomic studies have highlighted the power of transcriptome-based approaches (Leebens-Mack et al., 2019; Qi et al., 2018; Rothfels et al., 2013, 2015; Shen et al., 2018; Wickett et al., 2014). Expanding genome- or transcriptome-wide data sets are essential to overcome challenges imposed by high degrees of paralogy in many fern genomes.

The first insights into the structure of fern genomes were provided by the complete genome sequences of *Azolla filiculoides* and *Salvinia cucullata* (Li et al., 2018). These two genera are known for their small genome sizes (*A. filiculoides* 753 Mb and *S. cucullata* 255 Mb; Li et al., 2018). For example, the genome sizes of *A. microphylla* and *S. molesta* were estimated at 1C = 0.77 pg or ~0.75 Gband 1C = 2.25 pg, respectively (Clark et al., 2016; Obermayer et al., 2002). Conversely, many fern genomes are of enormous size (e.g., *Tmesipteris obliqua* 1C = 150.61 pg; Hidalgo et al., 2017). Major progress in the establishment of genomic resources was made with the sequencing of 73 fern transcriptomes (Carpenter et al., 2019; Leebens-Mack et al., 2019). Such data sets have been successfully used to develop single copy nuclear markers to resolve deep evolutionary relationships among ferns (Rothfels et al., 2013, 2015). Transcriptome assemblies are also an important tool to develop genotyping approaches and overcome challenges associated with extremely large fern genomes (Bennett & Leitch, 2001; Hanson & Leitch, 2002; Obermayer et al., 2002). These approaches typically reduce genome complexity but still provide sufficient polymorphic markers to conduct population genomics analyses (Seeb et al., 2011). Establishing transcriptomic data sets for understudied fern clades will bring new insights into fern diversification.

An important genus lacking transcriptomic resources is *Botrychium* belonging to subclass Ophioglossidae (PPG I, 2016). This subclass is characterized by a subterranean gametophytic stage (Field et al., 2015; Jeffrey, 1898; Winther & Friedman, 2007) and extremely large and complex genomes (e.g., *Ophioglossum petiolatum* 1C = 65.55 pg; Obermayer et al., 2002). *Botrychium* occurs in open habitats on nearly every continent across a broad temperate and boreal distribution. The genus is divided into three monophyletic clades as estimated by non-coding plastid markers (Simplex-Campestre, Lanceolatum and Lunaria; Dauphin et al., 2017), containing 35 recognized taxa (PPG I, 2016). The challenge of identifying *Botrychium* taxa based on morphology is underlined by claims of cryptic species (Clausen, 1938; Hauk, 1995). Ambiguous morphologies are sometimes caused by polyploidization which is a major driver of speciation as half of the known *Botrychium* species are allopolyploids (Dauphin et al., 2018). Nuclear markers have resolved the parental origins of these allopolyploid taxa and have provided insights into the genus radiation (Dauphin et al., 2018). Additionally, the reconstruction of maternal lineages of *Botrychium* has revealed high genetic diversity within the Lunaria clade, highlighting the uncertainty of taxonomic assignments (Dauphin et al., 2014, 2017; Maccagni et al., 2017). Previous population genetic studies based on isozymes have shown a lack of genetic differentiation among morphologically recognized types (Williams et al., 2016), and the low amount of genetic variation detected within *Botrychium* populations suggests pervasive self-fertilization (Farrar, 1998; Hauk & Haufler, 1999; Williams, 2021). Furthermore, genetic differentiation among populations and regions was found to be low, suggesting that gene flow may occur (Birkeland et al., 2017; Camacho & Liston, 2001; Swartz & Brunsfeld, 2002). These studies highlight the need for powerful, genome-wide marker systems to resolve population structures, life histories, and taxonomy of eusporangiate ferns.

In this study, we developed transcriptome-wide SNPs for the fern species *Botrychium lunaria* (L.) Sw., which has a large genome size of about 24.1 pg (2C; ~23.7 Gb; Veselý et al., 2012), establishing the first such resource for ferns. By analysing the transcriptome of 12 individuals mapped against the newly established reference transcriptome, we show the power of the discovered SNPs to

TABLE 1 Populations, accessions and voucher information. Geographic coordinates are given in the WGS84 reference system (decimal degrees)

Individual identifier	Population	Location	Latitude	Longitude	Altitude (m)	Date	Voucher number	Absolute genome size[a]	Deposit[b] institute
CHA_I_1	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	NE000101258	–	NEU
CHA_I_3	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	NE000101257	–	NEU
CHA_I_5	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	NE000101256	–	NEU
CHA_I_6	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	NE000101255	–	NEU
CHA_I_7	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	CHA_I_7	20.58	UNINE
CHA_I_8	Chasseral	Chasseral	47.12974	7.04934	1549.71	07.06.2017	NE000101254	–	NEU
IIT2_H2	Val d'Hérens	Forclaz	46.08741	7.54379	2346.15	2015	IIT2_H2	19.38	UNINE
IIT1_H5	Val d'Hérens	Forclaz	46.08851	7.53906	2346.15	2015	IIT1_H5	19.97	UNINE
IIT2_B2	Val d'Hérens	Forclaz	46.08741	7.54379	2406.10	2015	IIT2_B2	19.84	UNINE
IT1_A3	Val d'Hérens	Mase	46.20432	7.48350	2406.62	2015	IT1_A3	19.40	UNINE
IT1_H1	Val d'Hérens	Mase	46.20432	7.48350	2406.62	2015	IT1_H1	19.84	UNINE
IT2_D4	Val d'Hérens	Mase	46.19642	7.48502	2424.07	2015	IT2_D4	–	UNINE

[a]Expressed as 2C-values and pictograms.

[b]NEU: Herbarium of the University of Neuchâtel; UNINE: University of Neuchâtel, Institute of Biology Herbarium. The specimens deposited at UNINE are frozen samples stored at −80°C.

resolve population-level variation in heterozygosity and to test for evidence of ploidy variation. We further demonstrate the power of transcriptome-wide markers to resolve phylogenetic relationships at the genus level and among deeply divergent fern lineages.

2 | MATERIALS AND METHODS

2.1 | Sampling, library preparation and sequencing

Leaf material of *B. lunaria* was obtained from three locations in Switzerland: two in the Val d'Hérens in the Pennine Alps, Mase and Forclaz, within approximately 30 km, and one at the Chasseral in the Jura Mountains (Table 1). The two regions are separated by ~120 km and the Alpine population was sampled on meadows over an altitudinal range of 1500 to 2400 m (Table 1). Leaves of six individuals from Val d'Hérens and from Chasseral were collected in July 2015 and June 2017, respectively. Plant material was wrapped in aluminum foil and frozen immediately in liquid nitrogen. Total RNA was extracted from trophophores (i.e., the sterile part of the leaves) using the RNAeasy Plant Mini Kit (Qiagen) and DNA was eliminated using DNase I digestion. Total RNA was quantified using a Qubit fluorometer (Invitrogen, Thermo Fisher Scientific) with the RNA Broad-Range assay kit (Invitrogen, Thermo Fisher Scientific) and quality-checked using an Agilent 2200 Tape Station (Agilent Technologies, Inc.). Samples were diluted to 100 ng/μl in RNase free ultra-pure water before library preparation. The RNA-sequencing libraries were prepared following a TruSeq RNA library preparation protocol (Illumina, Inc.) enriching for polyadenylated RNAs. After quality assessment on an Agilent 2200 Tape Station, libraries were pooled and sequenced in 150 bp single-end mode on one lane of an Illumina HiSeq 4000 sequencer.

Absolute genome sizes were estimated by flow cytometry to assess the ploidy level of the six specimens for which frozen material was available, following the one-step methodology of Doležel et al. (1998). Frozen material was silica-dried before being processed.

2.2 | De novo assembly, filtering and quality assessment

Sequencing reads were quality-checked using FastQC v. 0.11.7 (Andrews, 2010) and trimmed using Trimmomatic v. 0.38 (Bolger et al., 2014). Reads were retained if the leading and trailing base qualities were >5, had a 4-bp sliding window >15, and a minimum read length of 36 bp. Trimmed reads from a single reference individual were then de novo assembled using Trinity v. 2.8.3 (Haas et al., 2013). The other individual transcriptomes were used for SNP calling (see below). The reference individual choice was made based on the highest trimmed transcript sequence number. We used the pseudo-alignment percentage calculated by Kallisto v. 0.45.0 (Bray et al., 2016) to assess the representativeness of the raw assembly across the twelve sequenced individuals in total. Candidate coding regions

were identified using TransDecoder v. 5.3.0 (Haas et al., 2013). Only transcripts with an open reading frame (ORF) of at least 100 amino acids were kept. We also retained only the longest isoform per transcript using the Trinity v. 2.8.3 toolkit. We used Diamond v. 0.9.24 (Buchfink et al., 2015) to screen the transcript assembly against the NCBI nonredundant protein (nr) and UniVec databases to identify potential foreign RNA contaminants. The best hit for each transcript was assigned at the phylum-level using the R package taxise v. 0.9.7 (Chamberlain & Szöcs, 2013) in RStudio v. 1.2.1335 (R Development Core Team, 2020; RStudio Team, 2015). We excluded all transcripts with a best hit outside of the plant kingdom (i.e., Viridiplantae). We did not filter for hits within the plant kingdom due to the scarce representation of seedless plant sequences in the NCBI nr and UniVec databases. Transcript redundancies were detected using CD-HIT-EST v.4.8.1 (Weizhong & Godzik, 2006) with a global sequence identity of 0.95% and a word size of 10. Redundant transcripts were pruned from the final assembly retaining the longest transcript per cluster. The completeness of the transcriptome assemblies was assessed using BUSCO v. 5.0.0 with the viridiplantae_odb10 database (Simão et al., 2015). Data were visualized using the R package ggplot2 v. 3.3.2 (Wickham, 2016).

2.3 | Variant calling

We generated alignment BAM files for each individual against the transcriptome using the short read aligner Bowtie2 v. 2.4.2 (Langmead, 2010) and SAMtools v. 1.11 (Li et al., 2009). Depth coverage of the reference individual was extracted using SAMtools idxstats. Alignments were processed with HaplotypeCaller implemented in the Genome Analysis Toolkit (GATK) v. 4.1.8.1 (Van der Auwera et al., 2013; DePristo et al., 2011; McKenna et al., 2010) for single nucleotide polymorphism (SNP) calling. The resulting gvcf files were combined and genotyped using the GATK CombineGVCF and GenotypeGVCF tools, respectively. We excluded monomorphic sites from further analysis. We filtered SNPs for the number of genotyped chromosomes (AN ≥ 20) out of a maximum of 24 (12 diploid individuals). Quality criteria QUAL > 100, QualByDepth > 5.0, RMSMappingQuality > 20.0, MappingQualityRankSumTest retained values >−2.0 and <2.0, and ReadPosRankSumTest and BaseQualityRankSumTest retained values >−2.0 and <2.0 were defined following the best practices and were applied to flag low-quality loci (Figure S1). We removed SNPs failing the above filters using VCFtools v. 0.1.16 (Danecek et al., 2011) and added a filter to retain only biallelic SNPs. Analyses were performed using the R packages vcfR v. 1.12.0 (Knaus & Grünwald, 2017) and the SNP statistics among transcripts were visualized using ggplot2 v. 3.3.2.

2.4 | Population genetics analyses

Intra-individual allele frequencies were calculated for each individual and SNP locus using the mapped read depth per allele (AD). The frequency distributions were plotted per individual. We subsampled the number of SNPs by selecting one SNP every 1,000 bp of transcriptomic sequence using VCFtools v. 0.1.16. The subsampling was performed to obtain a representative set of markers across all polymorphic transcripts. Furthermore, the subsampling reduced linkage disequilibrium among markers to reduce biases in the representation of the population genetic structure. We performed principal component analyses (PCA) and calculated the pairwise Nei's F_{ST} and the mean heterozygosity (H_e) per location and per individual (Nei, 1987). These analyses were performed using the R packages vcfR v. 1.12.0, adegenet v. 2.1.3 (Jombart & Ahmed, 2011) and hierfstat v. 0.5–7 (Goudet, 2005), and data were visualized using ggplot2 v. 3.3.2.

2.5 | Functional annotation

We functionally characterized encoded protein sequences based on gene ontology (GO) terms. We summarized GO terms by selecting the least redundant annotations among the 30 most frequent terms per ontology (biological process BP, cellular component CC, and molecular function MF). Analyses were performed using the Bioconductor packages AnnotationDbi v. 1.46.0 (Pagès et al., 2020), GO.db v. 3.8.2 (Carlson, 2020), GSEABase v. 1.34.0 (Morgan et al., 2020), annotate v. 1.62.0 (Gentleman, 2020), and data were visualized using the R package ggplot2 v. 3.3.2.

2.6 | Genus-level phylogenetic analyses

To estimate the phylogenetic placement of the 12 individuals included in our study, we retrieved sequences of four previously analysed nuclear regions (i.e., *ApPEFP_C*, *CRY2cA*, *CRY2cB*, and *transducin*) of both diploid and polyploid *Botrychium* taxa (Dauphin et al., 2018). We searched homologous sequences in the transcriptome assembly using BLAST v. 2.9.0 (Altschul et al., 1990). If the associated transcript was found in the assembly, we used BCFtools v. 1.9 (Li, 2011) to retrieve the corresponding transcript from the 11 remaining individuals using the VCF file information. Sequence alignments were performed with MAFFT v. 7.470 under the G-INS-i strategy and default parameters (Katoh et al., 2015; Katoh & Standley, 2013). Multiple alignments were visually inspected and manually adjusted using Geneious v. 8.1.9 (Kearse et al., 2012). Phylogenetic trees were inferred using maximum likelihood (ML) in RAxML-NG v. 0.9.0 (Kozlov et al., 2019). We ran tree inferences with a fixed random seed of 2 under the HKY+GAMMA model based on model settings by Dauphin et al. (2018) to ensure reproducibility. The tree search was performed using 25 random- and 25 parsimony-based starting trees. The branch support was estimated using 1,000 bootstrap replicates and calculated according to the transfer bootstrap expectation matrix (Lemoine et al., 2018). The support values were depicted on the best-scoring ML tree. The tree was visualized using the R packages ape v. 5.4–1 (Paradis & Schliep, 2019) and ggtree v.2.4.0 (Yu et al., 2017).

2.7 | Phylogenomic analyses

We performed phylogenomic analyses across ferns by including the newly established *B. lunaria* transcriptome for a total of 95 transcriptomes including 86 fern species (18 eusporangiates and 68 leptosporangiates), six Spermatophyta and two Lycopodiopsida (Table S1; Leebens-Mack et al., 2019; Qi et al., 2018; Shen et al., 2018). The Spermatophyta and Lycopodiopsida species represent the outgroup in the analysis. We first performed an orthologue search using OrthoFinder v. 2.3.12 (Emms & Kelly, 2015) including the newly established *B. lunaria* transcriptome and protein sequences of other transcriptomes. We retained orthogroups shared by every member of the taxa set regardless of the gene copy number. We used ASTRAL-Pro v. 1.1.5, which was specifically developed to summarize multi-copy gene trees into a species tree (Zhang et al., 2020). To reduce the computational load, we processed only orthogroups for which at least half of the species had a single-copy gene and excluded two orthogroups with very high paralogue counts ($n \geq 700$). Sequences of the orthogroups subset were subsequently aligned with MAFFT v. 7.475 under the G-INS-i strategy and default parameters. The optimal substitution model was assessed for each orthogroup alignment using Modeltest-NG v. 0.1.6 (Darriba et al., 2020). Finally, unrooted gene trees were built using maximum likelihood (ML) in RAxML-NG v. 1.0.1. We ran tree inferences under the best model according to the Akaike Information Criterion (AICc) with a fixed random seed of 12,345. The tree search was performed on 25 random and 25 parsimony-based starting trees. The inferred gene trees (i.e., best ML trees) were used to estimate a species tree with ASTRAL-Pro. Branch support was calculated using local posterior probabilities (Sayyari & Mirarab, 2016). In order to display internal branch lengths on our final trees, we first added arbitrary terminal branch lengths using the python script (add-bl.py) provided with ASTRAL. Species trees were edited using the R packages treeio v.1.14.0 (Wang et al., 2020) and ape v. 5.4–1. The arbitrary terminal branch lengths were removed for the final tree visualizations.

3 | RESULTS

3.1 | Sample collection and transcriptome assembly

We successfully analyzed transcriptomes of 12 *B. lunaria* individuals. The transcriptome sequencing produced 14.6–50.1 million reads per individual. After quality trimming, we retained 97.0%–99.2% of the reads (Figure 1a, Table S2). The highest number of high-quality reads (49.5 million) was obtained for the Chasseral individual CHA_I_1. We selected this individual to assemble a reference transcriptome for the species. The raw assembly for CHA_I_1 contained 167,306 transcripts for a total of 87,537 candidate genes before redundancy filtering. Mapping reads from all 12 individuals to the raw transcriptome assembly showed a pseudoalignment rate (i.e., percentage of mapped reads) between 74.2%–82.5% regardless of the population of origin (Figure 1a, Table S2). We analyzed all assembled transcripts for the presence of high-confidence open reading frames (ORF; ≥100 amino acids). We retained 69,280 transcripts (41.4%) covering 26,139 predicted genes (Figure 1b). Next, we selected the longest transcript for each gene (Figure 1d). We screened each gene against the complete nonredundant protein and the UniVec database of NCBI and found evidence for contamination in 438 transcripts. Most contaminant sequences were associated with viruses, fungi, or bacteria (Figure 1c). We performed an additional screen for redundant transcripts using CD-HIT-EST. We identified 16 clusters and retained only a unique representative transcript per cluster for further analyses. The final assembly consisted of 25,677 unique transcripts spanning a total of 3,423 Mb. The average and median transcript lengths were 1,333 and 967 bp, respectively (Figure 1d). The N50 of the final transcriptome was 1,995 bp with an average GC content of 44.3% (Table 2). GO terms were assigned to 11,137 transcripts (43.4%; Figure 2d, Table S3).

3.2 | A highly complete *B. lunaria* transcriptome

The completeness of the assembled *B. lunaria* transcriptome assessed using BUSCO matched the completeness of complete genome assemblies of other ferns. Importantly, the absence of fern species among the 30 species constituting the BUSCO viridiplantae_odb10 database might lead to an underestimation of the assembly completeness. We found 90.1% complete single-copy, 0.9% complete duplicates, 5.9% fragmented and 3.1% missing genes for the *B. lunaria* transcriptome (Figure 1e, Table 3). This is comparable to the only two complete genome assemblies of ferns: *A. filiculoides* with 79.3% and *S. cucullata* with 90.4% complete single-copy genes (Figure 1e, Table 3). The two Salviniaceae genomes and the *B. lunaria* transcriptome exhibited a comparable number of missing BUSCO genes (2.8%–1.6% and 3.1%, respectively; Figure 1e, Table 3) and of fragmented BUSCO genes (9.9%–4.7% and 5.9% respectively; Figure 1e, Table 3). The mapped reads coverage depth of the reference individual to the assembled transcripts is on average 1649× with a range of 4 to 514,622×. Transcript coverage is strongly skewed towards low coverage, whereas ~7% show coverage >4000 reads (Figure 2a). The read coverage shows no obvious association with transcript length (Figure 2b).

3.3 | Phylogenetic assignment of all analyzed transcriptomes

The analysis of nuclear barcoding loci confirmed the 12 individuals included in the transcriptome analyses as *B. lunaria*. Among the four nuclear loci previously sequenced in a broad sample of *Botrychium* species, three loci displayed sequence variation nearly exclusively in intronic sequences (Dauphin et al., 2018). Therefore, no comparison with our transcriptomic sequences was possible. We focused on the locus CRY2cA carrying sufficient informative sites in the coding regions to produce a well-supported phylogeny. The combined data set for CRY2cA included 67 individuals, representing 38 *Botrychium* taxa

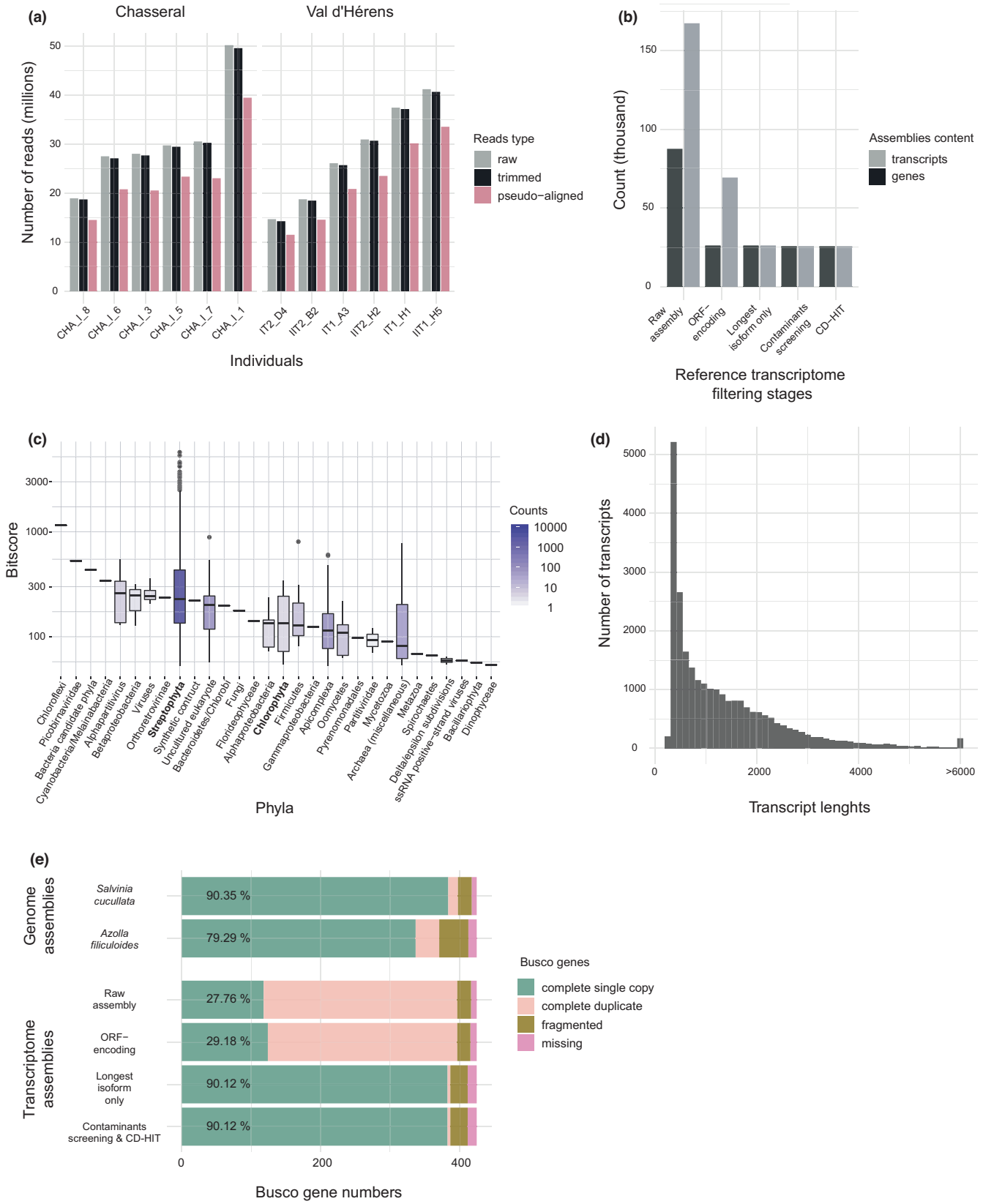


FIGURE 1 De novo assembly of the *Botrychium lunaria* transcriptome. (a) Distribution of the raw, trimmed, and pseudoaligned read numbers per individual. (b) Transcripts and gene assembly content at each filtering stage of the transcriptome. (c) Contaminant transcript sequences detected by phylum. Retained phyla are indicated in bold. See methods for further details. (d) Distribution of the assembled transcript lengths given in base pairs (bp). (e) BUSCO genes detected at each filtering stage of the *B. lunaria* transcriptome and for two genome assemblies of ferns (*Azolla filliculoides* and *Salvinia cucullata*; Li et al., 2018)

TABLE 2 Overview of assembly statistics over the different transcript filtering stages

Filtering stage	Assembled bases	Transcripts	Genes	N50-longest isoform	N50-all	GC%
Raw assembly	56,273,802	167,306	87,537	1,689	1,089	43.68
ORF-encoding	34,588,465	69,280	26,139	2,152	1,988	44.00
Longest isoform only	34,588,465	26,139	26,139	1,988	1,988	44.31
Contaminant screening	34,245,455	25,701	25,701	1,995	1,995	44.30
CD-HIT	34,230,067	25,677	25,677	1,995	1,995	44.30

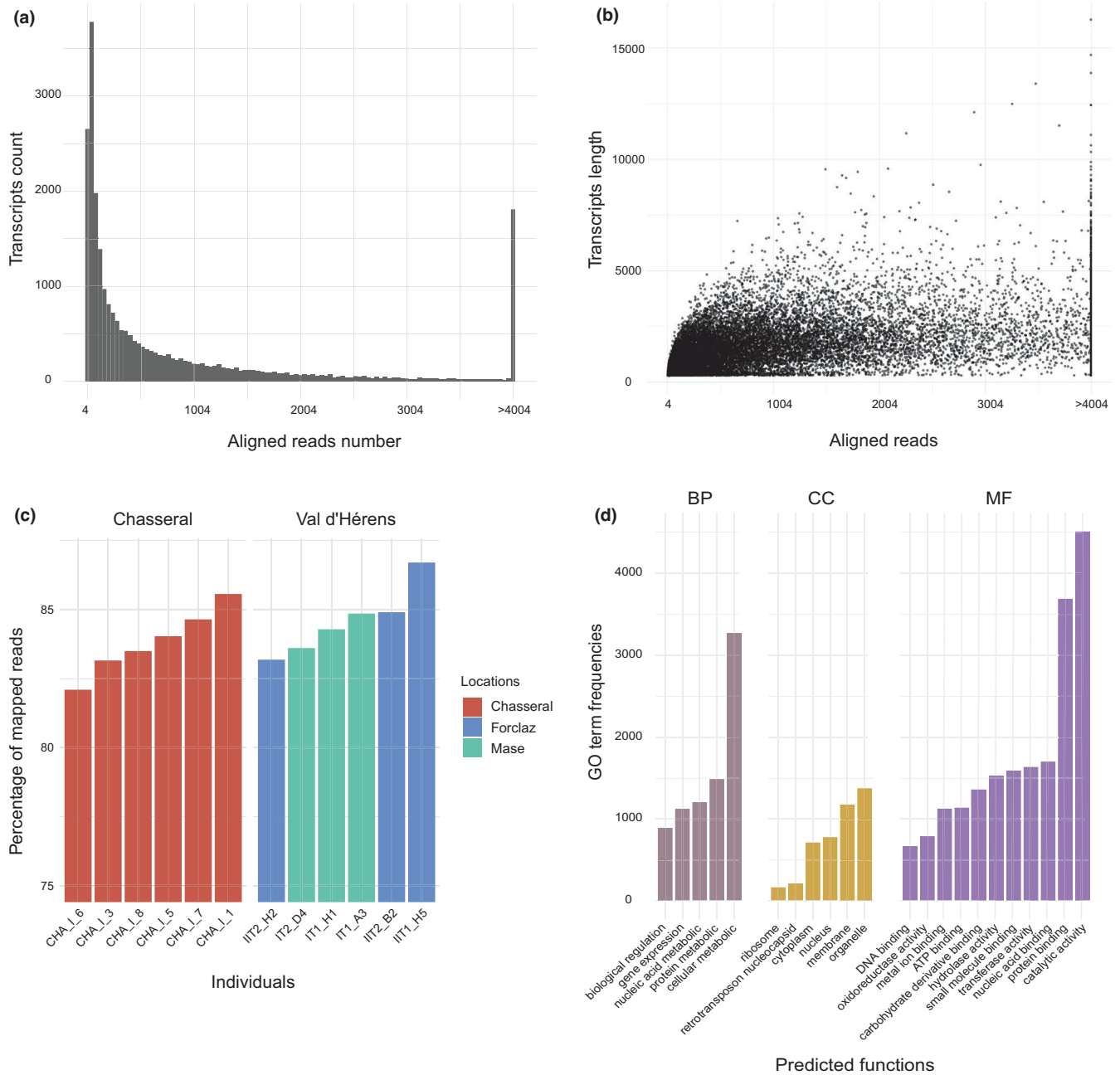


FIGURE 2 Analyses of the assembly coverage. (a) Number of aligned reads per assembled transcript for the reference individual. (b) Aligned reads of the reference individual according to the assembled transcript length. (c) Mapping rate of all 12 individuals from the Chasseral and Val d'Hérens populations, including the subpopulations Forclaz and Mase. The reference individual used to assemble the transcriptome was CHA_I_1. (d) Characterizations of predicted functions encoded by the transcriptome. Gene ontology (GO) term annotations are shown for the 30 most frequent terms per ontology (BP, biological process; CC, cellular component; MF, molecular function). GO terms with highly similar functions are excluded from the representation

TABLE 3 Analyses of assembly completeness using BUSCO genes. The *B. lunaria* transcriptome is compared to two genome assemblies of ferns (*Azolla filliculoides* and *Salvinia cucullata*)

Species name	Filtering stage	C[a]	CS[b]	CD[c]	F[d]	M[e]	N[f]	Data set
Azolla filliculoides	–	371	337	34	42	12	425	viridiplantae_odb10
Salvinia cucullata	–	398	384	14	20	7	425	viridiplantae_odb10
B. lunaria	Raw assembly	397	118	279	20	8	425	viridiplantae_odb10
B. lunaria	ORF-encoding	397	124	273	19	9	425	viridiplantae_odb10
B. lunaria	Longest isoform only	387	383	4	25	13	425	viridiplantae_odb10
B. lunaria	Contaminant screening	387	383	4	25	13	425	viridiplantae_odb10
B. lunaria	CD-HIT	387	383	4	25	13	425	viridiplantae_odb10

Notes: [a]Complete genes, [b]Complete and single copy genes, [c]Complete and duplicated genes, [d]Fragmented genes, [e]Missing genes, [f]Total number of BUSCO genes in the data set.

and an outgroup constituted by *Sceptridium multifidum* and *Botrypus virginianum* (Dauphin et al., 2018; Table S4). The multiple sequence alignment contained a total of 3,579 sites and 153 patterns. The main clades Lanceolatum, Lunaria, and Simplex-Campestre were resolved as monophyletic (Figure S2). Lanceolatum was resolved as a sister group to the Simplex-Campestre and Lunaria clade. All individuals from the Chasseral and Val d'Hérens grouped with *B. lunaria* var. *lunaria* and formed a well-supported clade.

3.4 | Transcriptome-wide phylogenomic trees of ferns

Phylogenomic analyses based on the transcriptomes of 91 fern species (Leebens-Mack et al., 2019; Qi et al., 2018; Shen et al., 2018) including the new transcriptome for *B. lunaria* were congruent with previous studies. We identified 41,186 orthogroups in total (Table S5) of which we retained 761 orthogroups and 92 species to construct a phylogenomic tree (Table S6). The species tree was well-resolved and branch support values from the local posterior probability (LPP) were high (>0.95; Figure 5). The species tree topology was consistent with the most recent fern phylogenies (Knie et al., 2015; Kuo et al., 2011; Leebens-Mack et al., 2019; Lu et al., 2015; Qi et al., 2018; Rai & Graham, 2010; Rothfels et al., 2015; Shen et al., 2018; Testo & Sundue, 2016) and with the current consensus classification (PPG I, 2016). The *B. lunaria* transcriptome clustered with the sister genus *Sceptridium* and the closely related genus *Botrypus*. Among the earliest divergent ferns (i.e., eusporangiate and early leptosporangiate), we identified Equisetales as the sister clade to all other ferns, and Marattiales as the sister clade to all leptosporangiates. Consistent with recent work, the Gleicheniales were recovered as paraphyletic and the deep eupolypod relationships remained largely unresolved in our phylogeny (Figure S3).

3.5 | Ploidy assessments and analyses of within-species transcriptome-wide polymorphisms

The mapping rate of reads from each individual against the assembled transcriptome sequences varied between 82.1%–86.7% (Figure 2c). The highest mapping rate was found for the individual IIT1_H5 (86.7%), which was slightly higher than the mapping rate of the reference individual used to establish the transcriptome (CHA_I_1, 85.5%; Table S4). We found no meaningful difference in mapping rates among populations. Based on reads aligned against reference transcripts, we assessed intra-individual allele frequency distributions based on read counts. Diploid individuals should show a singular, dominant peak at 0.5 corresponding to heterozygous SNPs (Chen et al., 2018). We indeed found a major peak around a frequency of 0.5 without any secondary peaks at 0.25 or 0.75, suggesting that all individuals are likely diploid (Figure 3a). However, very recent autopolypoids would also share this signature. The broad peak shape is likely due to stochasticity in read coverage at each SNP among individuals. Low read coverage can produce individual allele frequency estimates far from 0.5. For independent evidence of ploidy levels, we used flow cytometry. The estimated absolute genome sizes varied between 19.38–20.58 pg among the six analyzed individuals (2C-values; Table 1). Hence, there is no indication for ploidy variation and the genome size estimates match previous reports for the diploid genome of *B. lunaria* (Veselý et al., 2012). Given this evidence for diploidy, we proceeded with calling SNPs assuming a diploid state in all individuals. We recovered a total of 376,463 high-quality biallelic SNPs after filtering. The average number of SNPs per transcript was 17 and the maximum number was 257 (Figure 3b). The SNP density per transcript had a mean of 14, a median of 10, and a maximum of 153 SNPs per kb (Figure 3c). The median SNP density decreased slightly with transcript length (Figure 3d).

3.6 | Population structure and heterozygosity

The SNP genotyping data from the 12 individuals revealed a clear population differentiation between the two main sampling locations (Figure 4a). The first principal component (PC1, 17% of total variance explained) of the PCA identified a divergent genotype in the Chasseral population (CHA_I_7, Figure 4b). The second principal component (PC2, 12%) separated the two populations Chasseral and Val d'Hérens (Figure 4b). We performed a second PCA excluding the CHA_I_7 individual and found the Chasseral population to be more

diverse than Val d'Hérens (Figure 4c). We found no apparent differentiation between the two locations sampled in Val d'Hérens, but the sampling coverage was limited ($n = 3$ per location). The pairwise F_{ST} between populations was low (0.040). Mean heterozygosity was slightly higher in Val d'Hérens ($H_e = 0.20$) than in the Chasseral population (0.17; Figure 4d). We found similar levels of variation in individual heterozygosity among populations ranging from 0.16 to 0.21, except for CHA_I_7, which was an outlier in the PCA (Figure 4e). CHA_I_7 showed less than half the heterozygosity ($H_e = 0.05$) compared to other members of the same population.

4 | DISCUSSION

We established a high-quality transcriptome for the genus *Botrychium* filling an important gap in the coverage of eusporangiate ferns. The completeness of the transcriptomic gene space was comparable to well-assembled fern genomes. Using 12 individuals of the same species sampled in two regions, we were able to generate the first intraspecies transcriptome-wide SNP data set for ferns in general. Our analyses revealed a clear genetic distinction between the two *B. lunaria* populations and among the twelve individuals challenging the general assumption of gametophytic selfing being the dominant reproductive mode in *Botrychium* populations. A phylogenomic tree based on 761 orthologous genes confirmed the phylogenetic position of the genus among other fern lineages.

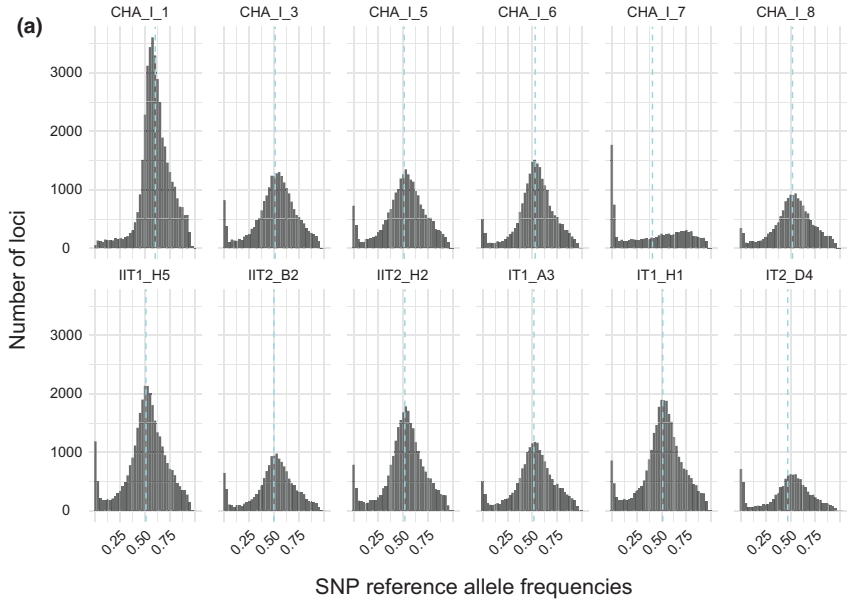
4.1 | Establishment of a transcriptome for *B. lunaria*

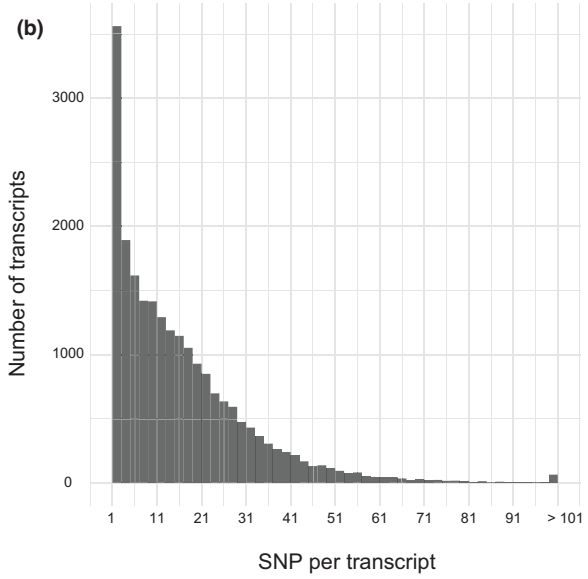
Generating a representative transcriptome assembly is challenging because not all genes are expressed in all tissues and life cycle stages. Across the life cycle of ferns, gene expression patterns are largely overlapping (Sigel et al., 2018), but the covered gene space is usually increased by including multiple target tissues. For *Botrychium*, only the trophophore and the sporophore were adequate tissues for the extraction of RNA since underground tissues are colonized by arbuscular mycorrhizal fungi (AMF; Winther & Friedman, 2007) leading to numerous contaminants. Because we only included trophophore tissue, the assembled transcriptome potentially underrepresents sporogenesis- and root-specific genes. Despite these challenges, our *B. lunaria* transcriptome has a fairly complete gene space in comparison to a wide range of assembled transcriptomes (Der et al., 2011; Leebens-Mack et al., 2019; Qi et al., 2018; Shen et al., 2018; Figure 1e). It is important to note that database-dependent tools such as BUSCO consistently underestimate transcriptome completeness if the database was compiled without closely related species. The challenge in using BUSCO is exemplified by the absence of ferns species in the viridiplantae data set. The gene space of assembled fern genomes tends to show less fragmented BUSCO genes compared to the *B. lunaria* transcriptome (Li et al., 2018). However, the *B. lunaria* transcriptome is consistent with other high-quality fern transcriptomes (Leebens-Mack et al., 2019; Qi et al., 2018; Shen et al., 2018). Missing gene segments in assembled transcriptomes are often caused by uneven read depth among genes or alternative splicing complicating gene recovery. The completeness of the *B. lunaria* transcriptome compared to other fern genomes and transcriptomes provides a powerful tool for phylogenomic and population genetic analyses.

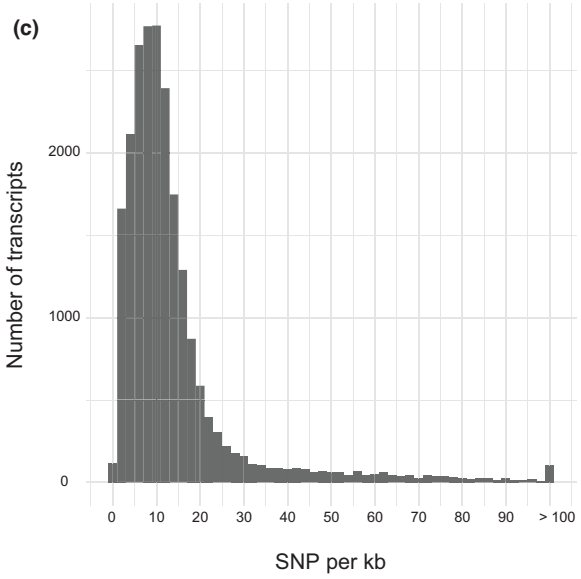
The *B. lunaria* transcriptome enables strong phylogenomic inference at different taxonomic levels, overcoming challenges associated with the small number of nuclear and chloroplast markers available for most ferns. *Botrychium* taxa have subtle morphological characteristics, hence taxonomy relies largely on molecular data (Dauphin et al., 2014, 2017; Maccagni et al., 2017; Stensvold & Farrar, 2017; Stensvold et al., 2002). We have placed the individual *B. lunaria* transcriptomes among other closely related taxa by retrieving orthologous genes, which were previously used for phylogenetic analyses. The newly established transcriptome will enable investigations of the gene content of *Botrychium*. The tens of thousands of markers across the coding sequences allow powerful genome-wide studies of population and species differentiation. Highly dense marker sets are critical for ecological genomics investigations to for example, identify loci underlying climatic adaptation, recent gene flow events and consequences of variation in the reproductive mode. Beyond this, refined analyses of transcriptomic markers can help to retrace the evolution of the extensive ploidy variation among *Botrychium*.

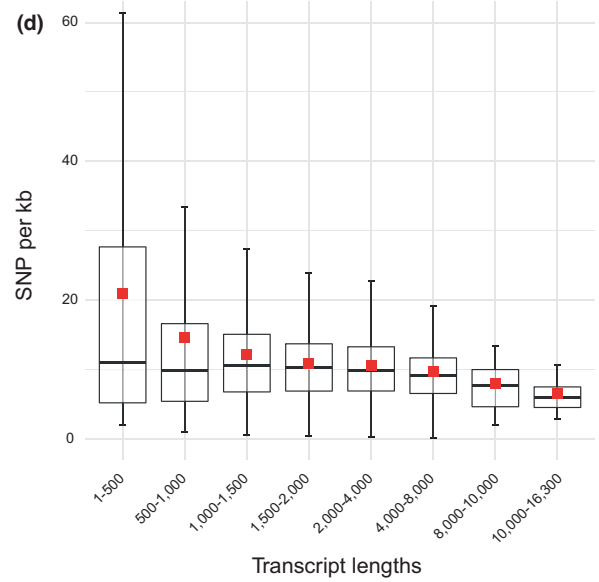
As an expansion of the phylogenetic analyses within *Botrychium*, we analysed orthologous genes across all ferns. The genus *Botrychium* was placed within the Ophioglossales with strong support (Figure 5). Furthermore, the phylogenetic placements of most fern clades were congruent with previous studies and highlighted unresolved discordances. For example, the paraphyly observed for the Gleicheniales in our species tree (Figure 5, Figure S3) corroborates recent findings (Qi et al., 2018; Shen et al., 2018). Sparse sampling can strongly influence tree topologies. For instance, the Matoniaceae, constituting one of the three Gleicheniales families (PPG I, 2016), are not represented in phylogenomic studies. Phylogenies based only on few barcoding loci, but with a more representative sampling, identified Gleicheniales as being a monophyletic clade (Pryer et al., 2004; Schuettpelz et al., 2006; Schuettpelz & Pryer, 2007). Phylogenomic studies including the Matoniaceae will be needed to ascertain the

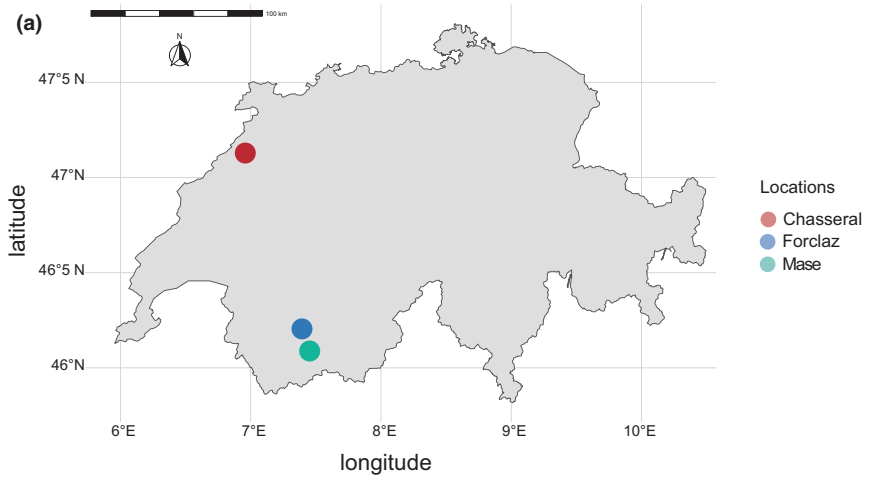
FIGURE 3 Analyses of population-level transcriptomic polymorphism. (a) Distribution of the transcriptome-wide SNP reference allele frequencies per individual estimated from mapped reads. The light-blue dashed lines show the mean reference allele frequency. Homozygous positions (frequencies 0 and 1) were excluded. (b) Number of SNPs per transcript up to 257. (c) Density of SNPs per transcript (i.e., number of SNPs per kb) up to 153. (d) SNP density according to bins of transcript length given in bp. The mean density is shown by a red rectangle and the median by the solid black line inside the box. The box outline indicates the first and third quartiles. The whiskers show the minimum (bottom one) and maximum (top one) number of SNPs per kb for each bin

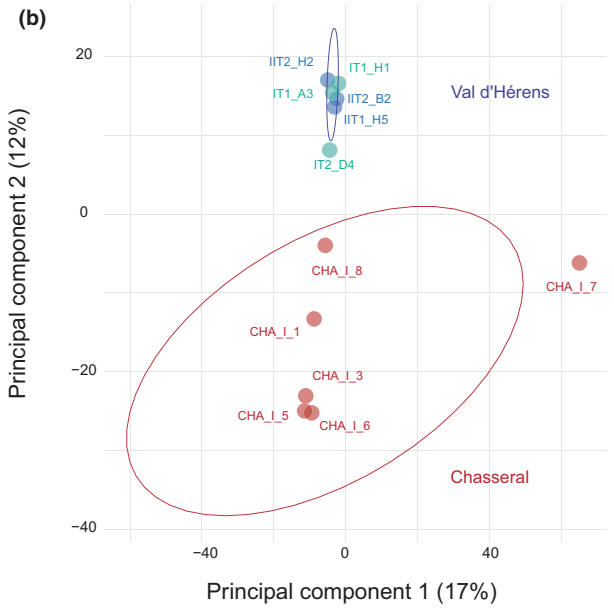


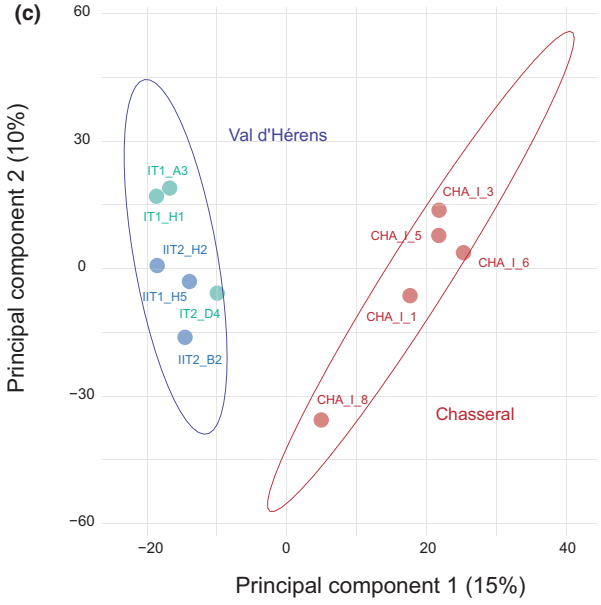


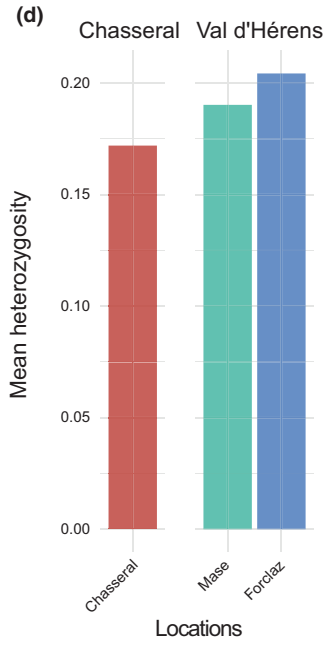












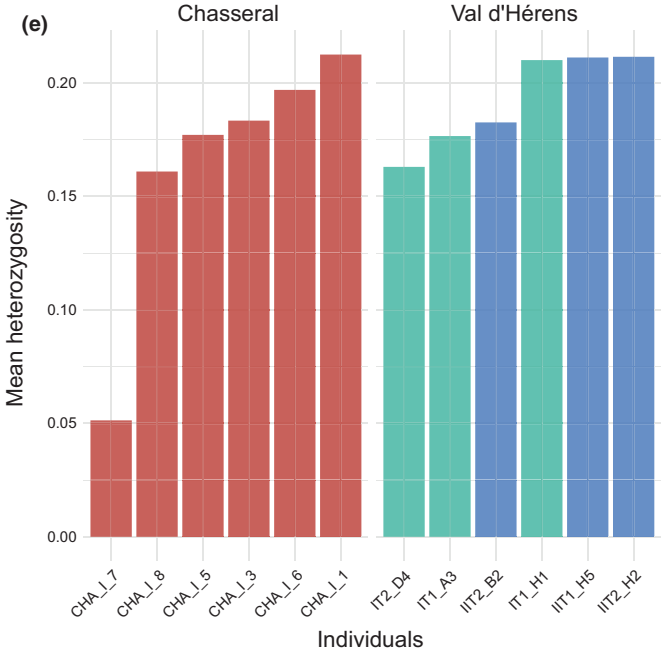


FIGURE 4 Population genetic structure and observed heterozygosity. (a) Sampling locations of *Botrychium lunaria* populations analysed in this study. (b) Principal component analysis (PCA) of the populations Chasseral and Val d'Hérens (sites Mase and Forclaz). (c) PCA of both populations excluding the CHA_I_7 outlier. PCA were analysed using a reduced SNP data set of a maximum of 1 SNP per kb of transcript. (d) Mean observed heterozygosity per location grouped by population. (e) Mean observed heterozygosity per individual grouped by population

placement of the Dipteridaceae outside of the order Gleicheniales. The resolution of our fern-wide phylogenomic analyses revealed some limitations in resolution at the within-genus level. *Equisetum hyemale* and *E. arvense* were recovered as sister species distant from *E. diffusum* even though *E. diffusum* and *E. arvense* belong to the same subgenus *Equisetum* which does not include *E. hyemale* (subgenus. *Hippochaete*; Christenhusz et al., 2021). The reduced power to resolve some phylogenetic relationships with genera might be explained by the scarce species representation within some genera and the degree of conservation of the included orthologues.

4.2 | Fine-grained resolution of population structure

The transcriptome-wide SNPs revealed clear geographical structuring between the two *B. lunaria* populations sampled from Switzerland (Figure 4b, c). The differentiation was apparent even when subsampling SNPs containing a maximum of 1 SNP per kb to avoid biases by highly polymorphic transcripts and high linkage disequilibrium. The full set of SNPs will provide a powerful tool for ecological genomics investigations to unravel loci contribution to adaptation. It has been generally assumed *Botrychium* species show no meaningful genetic differentiation within populations (Farrar, 1998; Hauk & Haufler, 1999; Williams, 2021) or low genetic differentiation among populations (Birkeland et al., 2017; Camacho & Liston, 2001; Dauphin et al., 2020; Swartz & Brunsfeld, 2002). However, the absence of genetic differentiation reported by previous studies may well stem from low marker resolution. The transcriptome-wide markers used here showed every individual was clearly distinct, and populations showed marked differentiation. The Chasseral and Val d'Hérens populations were collected in the Jura Mountains and Pennine Alps, respectively. The two sites are 120 km apart and separated by habitats unsuitable for *B. lunaria*. Therefore, reduced gene flow and genetic differentiation among populations is expected. We found no indication of genetic substructure among the two locations Mase and Forclaz within the Val d'Hérens valley. This suggests sufficient gene flow at the local scale or recent recolonization at the upper front of the valley, which is consistent with restriction-site associated DNA sequencing-based analyses of the same field sites (Dauphin, 2017). We found no evidence for higher levels of ploidy based on mapped read depths per individual (Figure 3a). The intraindividual allele frequency distributions showed some spread around the singular peak at frequency 0.5 (expected for diploids). The spread in the estimated frequencies likely reflects noise caused by low coverage SNP positions and was observed in other systems as well (Chen et al., 2018). The individual CHA_1_7 has a less pronounced and a shifted major peak. The individual does not cluster as closely as the other genotypes of the same population and is less heterozygous. We find no difference in sequencing quality or quantity that could explain the observed peak pattern. Unequal mapping of reads matching the two alleles may be a possible reason. Such patterns may be expected for introgressed sequences of higher divergence. Absolute genome size measurements corroborate the absence of higher level ploidy. This is in accordance with a broader sampling of genome sizes of 57 individuals across the *B. lunaria* group, which likewise recovered only diploid conditions (V. Mossion & M. Kessler, unpublished data), except for the allotetraploid species *B. yaaxudakeit*, which is only known from northwestern North America (Stensvold et al., 2002). Also, consistent with our findings, a recent study of Swiss populations based on allozyme markers found no evidence for fixed heterozygosity, a typical indicator of polyploidy (Dauphin et al., 2020). However, we cannot rule out that very recent polyploidization or autopolyploidization events may remain undetected.

Population-level genetic diversity is indicative of the reproductive mode of *Botrychium* populations. Self-fertilization is considered common in homosporous ferns and includes sporophytic and gametophytic selfing (Gastony & Gottlieb, 1985; Klekowski & Baker, 1966; McCauley et al., 1985; Sessa et al., 2016; Soltis & Soltis, 1986, 1990, 1992). In sporophytic selfing, zygotes are produced by gametes from two distinct gametophytes that originate from a single sporophyte. In contrast, in gametophytic selfing, zygotes are produced from gametes of the same gametophyte. Gametophytic selfing is thought to be the main reproductive mode for the genus *Botrychium* (Hauk & Haufler, 1999) and does lead to completely homozygous plants within one generation (Klekowski & Lloyd, 1968). In a population undergoing largely gametophytic selfing, low genetic diversity would also be expected among individuals due to genetic drift. The genetic diversity of *B. lunaria* populations (Figure 4d, e) and the clear structure among sites (Figure 4b, c) suggest that sporophytic selfing or outcrossing are predominant. Only one individual in the Chasseral population showed signature of a recent gametophytic selfing event (Figure 4e). These findings contrast with the general assumption that gametophytic selfing is the dominant reproductive mode in the genus but corroborate evidence of outcrossing in Swiss populations (Dauphin et al., 2020).

Our study establishes a high-quality set of transcriptome-wide SNPs in a species of the early diverging fern genus *Botrychium*. With an estimated genome size of 19.4–24.6 (2C; ~19.0–23.7 Gb) for *B. lunaria*, the assembly of transcriptomes provides the only currently feasible approach to generate extensive genome-wide markers information. Our transcriptome-wide SNPs enable fine-grained demographic history analyses and investigate the consequences of the complex mating systems and putative polyploidization events. Furthermore, ecological genomics studies based on our transcriptomic markers will provide the first opportunities to dissect adaptation to the local environment within fern species.

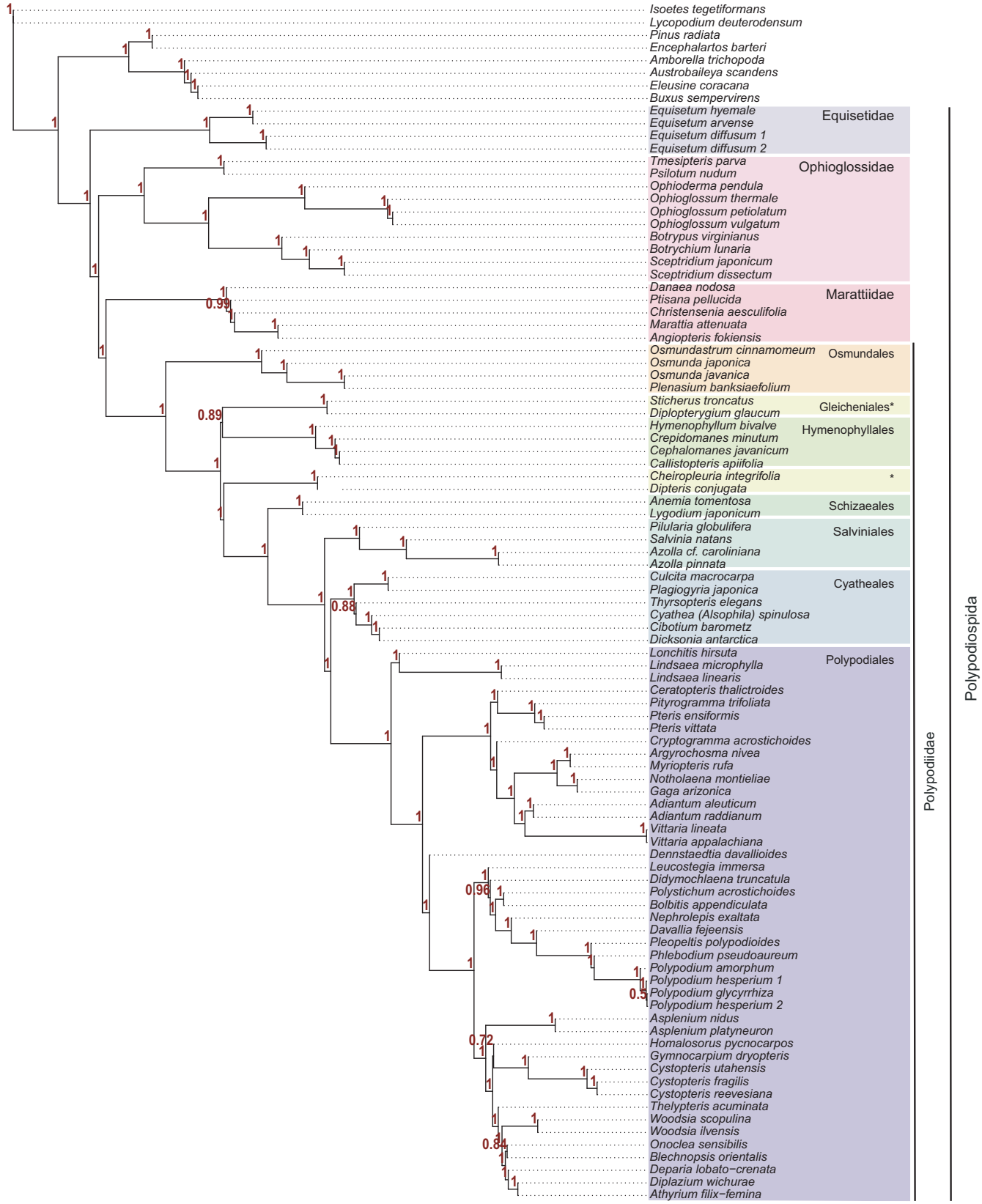


FIGURE 5 Phylogenomic relationships among ferns including *Botrychium lunaria*. A species tree of 761 orthologous genes including 92 taxa inferred by coalescence-based method implemented in ASTRAL-Pro. The branch support is indicated by local posterior probability (LPP) values. LPP values of the main topology are displayed in dark red above branches on left node sides. Fern subclasses (Equisetidae, Ophioglossidae, Marattiidae and Polypodiidae) are denoted by coloured rectangles or a vertical dark grey line on clade sides. Polypodiidae orders (Osmundales, Hymenophyllales, Gleicheniales, Schizaeales Salviniales, Cyatheales and Polypodiales) are designated by coloured rectangles. The internal branch lengths are in coalescent units. The tree with the full ASTRAL annotations is available as Supporting Information S7

ACKNOWLEDGEMENTS

We thank Frederic Sandoz for assistance in the fieldwork. We thank Aria Minder and Sylvia Kobel for input on laboratory methods. We acknowledge the 1KP sequencing consortium for advance access to transcriptomic sequences. Emilie Chanclud, Ursula Oggenfuss and Erik Koenen provided advice on data analysis. Ursula Oggenfuss, Leen N. Abraham, Simone Fouché, Pierre-Emmanuel Du Pasquier, Rosa-Angela Ston, Erik Koenen and Giacomo Zilio provided helpful comments on a previous manuscript version. Rosa-Angela Ston mounted herbarium vouchers. Elke Kessler conducted the flow cytometry analyses. Data produced and analysed in this paper were generated in collaboration with the Genetic Diversity Centre (GDC), ETH Zurich and the Functional Genomics Centre (FGC), Zurich. This work was supported by the Overhead fund of the University of Neuchâtel. Open Access Funding provided by Universite de Neuchatel.

AUTHOR CONTRIBUTIONS

V.M., B.D. and D.C. designed the study. V.M., M.K. and D.C. performed analyses, N.Z. contributed to analyses, B.D. and J.G. acquired funding. V.M. and D.C. wrote the manuscript with input from coauthors.

DATA AVAILABILITY STATEMENT

Raw sequencing reads and the assembled transcriptome have been made available at the NCBI Sequence Read Archive under the Bioproject accession PRJNA605155. The assembly was deposited under the accession number GIJZ00000000.1. Phylogenetic trees, alignments and protein sequences are available as Supporting Information S1–S7 on Zenodo (https://doi.org/10.5281/zenodo.5061267).

ORCID

Benjamin Dauphin https://orcid.org/0000-0003-0982-4252
Niklaus Zemp https://orcid.org/0000-0002-0991-5130
Daniel Croll https://orcid.org/0000-0002-2072-380X

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Andrews, S. (2010). *FastQC A Quality Control tool for High Throughput Sequence Data*. Retrieved August 30, 2019, from http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics*, *43*(1), 11.10.1-33. https://doi.org/10.1002/0471250953.bi1110s43

Barker, M. S., & Wolf, P. G. (2010). Unfurling fern biology in the genomics age. *BioScience*, *60*(3), 177–185. https://doi.org/10.1525/bio.2010.60.3.4

Bennett, M. D., & Leitch, I. J. (2001). Nuclear DNA amounts in pteridophytes. *Annals of Botany*, *87*(3), 335–345. https://doi.org/10.1006/anbo.2000.1339

Birkeland, S., Borgenkjetne, I., Brysting, A. K., Elven, R., & Alsos, I. G. (2017). Living on the edge: Conservation genetics of seven thermophilous plant species in a high arctic archipelago. *AoB PLANTS*, *9*(1), 1–21. https://doi.org/10.1093/aobpla/plx001

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. https://doi.org/10.1093/bioinformatics/btu170

Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, *34*(5), 525–527. https://doi.org/10.1038/nbt.3519

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Camacho, F. J., & Liston, A. (2001). Population structure and genetic diversity of *Botrychium pumicola* (Ophioglossaceae) based on inter-simple sequence repeats (ISSR). *American Journal of Botany*, *88*(6), 1065–1070. https://doi.org/10.2307/2657089

Carlson, M. (2020). GO.db: A set of annotation maps describing the entire Gene Ontology (Version 3.8.2). Bioconductor version: Release (3.10). https://doi.org/10.18129/B9.bioc.GO.db

Carpenter, E. J., Matasci, N., Ayyampalayam, S., Wu, S., Sun, J., Yu, J., Jimenez Vieira, F. R., Bowler, C., Dorrell, R. G., Gitzendanner, M. A., Li, L., Du, W., K. Ullrich, K., Wickett, N. J., Barkmann, T. J., Barker, M. S., Leebens-Mack, J. H., & Wong, G.-S. (2019). Access to RNA-sequencing data from 1,173 plant species: The 1000 Plant transcriptomes initiative (1KP). *GigaScience*, *8*(10), https://doi.org/10.1093/gigascience/giz126

Chamberlain, S. A., & Szöcs, E. (2013). taxize: Taxonomic search and retrieval in R. *F1000Research*, *2*, 191. https://doi.org/10.12688/f1000research.2-191.v2

Chen, E. C. H., Mathieu, S., Hoffrichter, A., Sedzielewska-Toro, K., Peart, M., Pelin, A., Ndikumana, S., Ropars, J., Dreissig, S., Fuchs, J., Brachmann, A., & Corradi, N. (2018). Single nucleus sequencing reveals evidence of inter-nucleus recombination in arbuscular mycorrhizal fungi. *eLife*, *7*, e39813. https://doi.org/10.7554/eLife.39813

Christenhusz, M. J. M., Chase, M. W., Fay, M. F., Hidalgo, O., Leitch, I. J., Pellicer, J., & Viruel, J. (2021). Biogeography and genome size evolution of the oldest extant vascular plant genus, *Equisetum* (Equisetaceae). *Annals of Botany*, *127*(5), 681–695. https://doi.org/10.1093/aob/mcab005

Clark, J., Hidalgo, O., Pellicer, J., Liu, H., Marquardt, J., Robert, Y., Christenhusz, M., Zhang, S., Gibby, M., Leitch, I. J., & Schneider, H. (2016). Genome evolution of ferns: Evidence for relative stasis of genome size across the fern phylogeny. *The New Phytologist*, *210*(3), 1072–1082. https://doi.org/10.1111/nph.13833

Clausen, R. T. (1938). A monograph of the Ophioglossaceae. *Memoirs of the Torrey Botanical Club*, *19*(2), 1–177.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Darriba, D., Posada, D., Kozlov, A. M., Stamatakis, A., Morel, B., & Flouri, T. (2020). ModelTest-NG: A new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*, *37*(1), 291–294. https://doi.org/10.1093/molbev/msz189

Dauphin, B. (2017). *Evolution of moonwort ferns* (Botrychium, *Ophioglossaceae) on local to global scales*. University of Neuchâtel.

Dauphin, B., Farrar, D. R., Maccagni, A., & Grant, J. R. (2017). A worldwide molecular phylogeny provides new insight on cryptic diversity within the Moonworts (*Botrychium* s. S., Ophioglossaceae). *Systematic Botany*, *42*(4), https://doi.org/10.1600/036364417X696483

Dauphin, B., Grant, J. R., & Farrar, D. R. (2020). Outcrossing mating system of the early-divergent moonwort fern (*Botrychium lunaria*, Ophioglossaceae) revealed in the European Alps. *International Journal of Plant Sciences*, *181*(9), 926–936. https://doi.org/10.1086/710491

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

Dauphin, B., Vieu, J., & Grant, J. R. (2014). Molecular phylogenetics supports widespread cryptic species in moonworts (*Botrychium* s.s., Ophioglossaceae). *American Journal of Botany*, *101*(1), 128–140. https://doi.org/10.3732/ajb.1300154

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., & Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Der, J. P., Barker, M. S., Wickett, N. J., dePamphilis, C. W., & Wolf, P. G. (2011). De novo characterization of the gametophyte transcriptome in bracken fern, *Pteridium aquilinum*. *BMC Genomics*, *12*, 99. https://doi.org/10.1186/1471-2164-12-99

Des Marais, D. L., Smith, A. R., Britton, D. M., & Pryer, K. M. (2003). Phylogenetic relationships and evolution of extant horsetails, equisetum, based on chloroplast DNA sequence data (rbcL and trnL-F). *International Journal of Plant Sciences*, *164*(5), 737–751. https://doi.org/10.1086/376817

Doležel, J., Greilhuber, J., Lucretti, S., Meister, A., Lysák, M. A., Nardi, L., & Obermayer, R. (1998). Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Annals of Botany*, *82*, 17–26. https://doi.org/10.1006/anbo.1998.0730

Ellwood, M. D. F., & Foster, W. A. (2004). Doubling the estimate of invertebrate biomass in a rainforest canopy. *Nature*, *429*(6991), 549–551. https://doi.org/10.1038/nature02560

Emms, D. M., & Kelly, S. (2015). OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biology*, *16*(1), 157. https://doi.org/10.1186/s13059-015-0721-2

Farrar, D. R. (1998). Population genetics of moonwort Botrychium. In N. Berlin, P. Miller, J. Borovansky, U. S. Seal, & O. Byers (eds.), *Population and habitat viability assessment for the goblin fern* (Botrychium mormo) (pp. 109–113). The Conservation Breeding Specialist Group. Retrieved from The Conservation Breeding Specialist Group website: http://www.cpsg.org/content/goblin-fern-phva-1998

Field, K. J., Leake, J. R., Tille, S., Allinson, K. E., Rimington, W. R., Bidartondo, M. I., Beerling, D. J., & Cameron, D. D. (2015). From mycoheterotrophy to mutualism: Mycorrhizal specificity and functioning in *Ophioglossum vulgatum* sporophytes. *New Phytologist*, *205*(4), 1492–1502. https://doi.org/10.1111/nph.13263

Gastony, G. J., & Gottlieb, L. D. (1985). Genetic Variation in the Homosporous Fern *Pellaea andromedifolia*. *American Journal of Botany*, *72*(2), 257–267. https://doi.org/10.1002/j.1537-2197.1985.tb08290.x

Gentleman, R. (2020). annotate: Annotation for microarrays (Version 1.64.0). Bioconductor version: Release (3.10). https://doi.org/10.18129/B9.bioc.annotate

George, L. O., & Bazzaz, F. A. (1999a). The fern understory as an ecological filter: Emergence and establishment of canopy-tree seedlings. *Ecology*, *80*(3), 833–845. https://doi.org/10.1890/0012-9658(1999)080[0833:TFUAAE]2.0.CO;2

George, L. O., & Bazzaz, F. A. (1999b). The fern understory as an ecological filter: Growth and survival of canopy-tree seedlings. *Ecology*, *80*(3), 846–856. https://doi.org/10.1890/0012-9658(1999)080[0846:TFUAAE]2.0.CO;2

Goudet, J. (2005). Hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, *5*(1), 184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Grewe, F., Guo, W., Gubbels, E. A., Hansen, A. K., & Mower, J. P. (2013). Complete plastid genomes from *Ophioglossum californicum*, *Psilotum nudum*, and *Equisetum hyemale* reveal an ancestral land plant genome structure and resolve the position of Equisetales among monilophytes. *BMC Evolutionary Biology*, *13*, 8. https://doi.org/10.1186/1471-2148-13-8

Grusz, A. L. (2016). A current perspective on apomixis in ferns. *Journal of Systematics and Evolution*, *54*(6), 656–665. https://doi.org/10.1111/jse.12228

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., Couger, M. B., Eccles, D., Li, B. O., Lieber, M., MacManes, M. D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C. N., … Regev, A. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols*, *8*, 1494–1512. https://doi.org/10.1038/nprot.2013.084

Hanson, L., & Leitch, I. J. (2002). DNA amounts for five pteridophyte species fill phylogenetic gaps in C-value data. *Botanical Journal of the Linnean Society*, *140*(2), 169–173. https://doi.org/10.1046/j.1095-8339.2002.00083.x

Haufler, C. H., Pryer, K. M., Schuettpelz, E., Sessa, E. B., Farrar, D. R., Moran, R., Schneller, J. J., Watkins, J. E., & Windham, M. D. (2016). Sex and the single gametophyte: revising the homosporous vascular plant life cycle in light of contemporary research. *BioScience*, *66*(11), 928–937. https://doi.org/10.1093/biosci/biw108

Hauk, W. D. (1995). A Molecular assessment of relationships among cryptic species of *Botrychium* subgenus *Botrychium* (Ophioglossaceae). *American Fern Journal*, *85*(4), 375–394. https://doi.org/10.2307/1547816

Hauk, W. D., & Haufler, C. H. (1999). Isozyme variability among cryptic species of *Botrychium* subgenus *Botrychium* (Ophioglossaceae). *American Journal of Botany*, *86*(5), 614–633. https://doi.org/10.2307/2656570

Hidalgo, O., Pellicer, J., Christenhusz, M. J. M., Schneider, H., & Leitch, I. J. (2017). Genomic gigantism in the whisk-fern family (Psilotaceae): *Tmesipteris obliqua* challenges record holder *Paris japonica*. *Botanical Journal of the Linnean Society*, *183*(4), 509–514. https://doi.org/10.1093/botlinnean/box003

Jeffrey, E. C. (1898). The gametophyte of *Botrychium virginianum*. *Univ. of Toronto Studies. Biological Series.*, *1*, 3-32 pl.i-iv.

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2015). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, *30*(14), 3059–3066. https://doi.org/10.1093/nar/gkf436

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., & Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Klekowski, E. J., & Baker, H. G. (1966). Evolutionary significance of polyploidy in the pteridophyta. *Science*, *153*(3733), 305–307. https://doi.org/10.1126/science.153.3733.305

Klekowski, E. J., & Lloyd, R. M. (1968). Reproductive biology of the Pteridophyta: 1. General considerations and a study of *Onoclea sensibilis* L. *Botanical Journal of the Linnean Society*, *60*(383), 315–324. https://doi.org/10.1111/j.1095-8339.1968.tb00093.x

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: A package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44–53. https://doi.org/10.1111/1755-0998.12549

Knie, N., Fischer, S., Grewe, F., Polsakiewicz, M., & Knoop, V. (2015). Horsetails are the sister group to all other monilophytes and Marattiales are sister to leptosporangiate ferns. *Molecular Phylogenetics and Evolution*, *90*, 140–149. https://doi.org/10.1016/j.ympev.2015.05.008

Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: A fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, *1–3*, https://doi.org/10.1093/bioinformatics/btz305

Kuo, L.-Y., Li, F.-W., Chiou, W.-L., & Wang, C.-N. (2011). First insights into fern matK phylogeny. *Molecular Phylogenetics and Evolution*, *59*(3), 556–566. https://doi.org/10.1016/j.ympev.2011.03.010

Langmead, B. (2010). Aligning short sequencing reads with Bowtie. *Current Protocols in Bioinformatics*, *32*(1), 11.7.1-11.7.14. https://doi.org/10.1002/0471250953.bi1107s32

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., & Initiative, O. T. P. T. (2019). One thousand plant transcriptomes and the phylogenomics of green plants. *Nature*, *574*(7780), 679–685. https://doi.org/10.1038/s41586-019-1693-2

Lehtonen, S., Silvestro, D., Karger, D. N., Scotese, C., Tuomisto, H., Kessler, M., Peña, C., Wahlberg, N., & Antonelli, A. (2017). Environmentally driven extinction and opportunistic origination explain fern diversification patterns. *Scientific Reports*, *7*(1), 4831. https://doi.org/10.1038/s41598-017-05263-7

Lemoine, F., Entfellner, J.-B.-D., Wilkinson, E., Correia, D., Felipe, M. D., Oliveira, T. D., & Gascuel, O. (2018). Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*, *556*(7702), 452. https://doi.org/10.1038/s41586-018-0043-0

Li, F.-W., Brouwer, P., Carretero-Paulet, L., Cheng, S., de Vries, J., Delaux, P.-M., Eily, A., Koppers, N., Kuo, L.-Y., Li, Z., Simenc, M., Small, I., Wafula, E., Angarita, S., Barker, M. S., Bräutigam, A., dePamphilis, C., Gould, S., Hosmani, P. S., … Pryer, K. M. (2018). Fern genomes elucidate land plant evolution and cyanobacterial symbioses. *Nature Plants*, *4*(7), 460–472. https://doi.org/10.1038/s41477-018-0188-8

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987–2993. https://doi.org/10.1093/bioinformatics/btr509

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, & R.1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, *25*(16), 2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lu, J.-M., Zhang, N., Du, X.-Y., Wen, J., & Li, D.-Z. (2015). Chloroplast phylogenomics resolves key relationships in ferns. *Journal of Systematics and Evolution*, *53*(5), 448–457. https://doi.org/10.1111/jse.12180

Maccagni, A., Parisod, C., & Grant, J. R. (2017). Phylogeography of the moonwort fern *Botrychium lunaria* (Ophioglossaceae) based on chloroplast DNA in the Central-European Mountain System. *Alpine Botany*, *127*(2), 185–196. https://doi.org/10.1007/s00035-017-0188-8

Magallón, S., Hilu, K. W., & Quandt, D. (2013). Land plant evolutionary timeline: Gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *American Journal of Botany*, *100*(3), 556–573. https://doi.org/10.3732/ajb.1200416

McCauley, D. E., Whittier, D. P., & Reilly, L. M. (1985). Inbreeding and the rate of self-fertilization in a grape fern, *Botrychium dissectum*. *American Journal of Botany*, *72*(12), 1978–1981. https://doi.org/10.1002/j.1537-2197.1985.tb08471.x

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

Mehltreter, K., Walker, L. R., Sharpe, J. M., Kessler, M., Richardson, S. J., Hietz, P., & Sheffield, E. (2010). *Fern ecology*. Cambridge University Press. Retrieved from https://www.cambridge.org/ch/academic/subjects/life-sciences/plant-science/fern-ecology

Morgan, M., Falcon, S., & Gentleman, R. (2020). GSEABase: Gene set enrichment data structures and methods (Version 1.48.0). Bioconductor version: Release (3.10). https://doi.org/10.18129/B9.bioc.GSEABase

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Obermayer, R., Leitch, I. J., Hanson, L., & Bennett, M. D. (2002). Nuclear DNA C-values in 30 species double the familial representation in pteridophytes. *Annals of Botany*, *90*(2), 209–217. https://doi.org/10.1093/aob/mcf167

Page, C. N. (2002). Ecological strategies in fern evolution: A neopteridological overview. *Review of Palaeobotany and Palynology*, *119*(1), 1–33. https://doi.org/10.1016/S0034-6667(01)00127-0

Pagès, H., Carlson, M., Falcon, S., & Li, N. (2020). *AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor*. Bioconductor version: Release (3.10).

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Pinson, J. B., Chambers, S. M., Nitta, J. H., Kuo, L.-Y., & Sessa, E. B. (2016). The separation of generations: Biology and biogeography of long-lived sporophyteless fern gametophytes. *International Journal of Plant Sciences*, *178*(1), 1–18. https://doi.org/10.1086/688773

PPG I. (2016). A community-derived classification for extant lycophytes and ferns. *Journal of Systematics and Evolution*, *54*(6), 563–603. https://doi.org/10.1111/jse.12229

Pryer, K. M., Schneider, H., Smith, A. R., Cranfill, R., Wolf, P. G., Hunt, J. S., & Sipes, S. D. (2001). Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. *Nature*, *409*(6820), 618–622. https://doi.org/10.1038/35054555

Pryer, K. M., Schuettpelz, E., Wolf, P. G., Schneider, H., Smith, A. R., & Cranfill, R. (2004). Phylogeny and evolution of ferns (monilophytes) with a focus on the early leptosporangiate divergences. *American Journal of Botany*, *91*(10), 1582–1598. https://doi.org/10.3732/ajb.91.10.1582

Qi, X., Kuo, L.-Y., Guo, C., Li, H., Li, Z., Qi, J. I., Wang, L., Hu, Y. I., Xiang, J., Zhang, C., Guo, J., Huang, C.-H., & Ma, H. (2018). A well-resolved fern nuclear phylogeny reveals the evolution history of numerous transcription factor families. *Molecular Phylogenetics and Evolution*, *127*, 961–977. https://doi.org/10.1016/j.ympev.2018.06.043

Qiu, Y.-L., Li, L., Wang, B., Chen, Z., Dombrovska, O., Lee, J., Kent, L., Li, R., Jobson, R. W., Hendry, T. A., Taylor, D. W., Testa, C. M., & Ambros, M. (2007). A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences*, *168*(5), 691–708. https://doi.org/10.1086/513474

R Development Core Team (2020). *R: The R project for statistical computing*. Retrieved February 9, 2020, from https://www.r-project.org/

Rai, H. S., & Graham, S. W. (2010). Utility of a large, multigene plastid data set in inferring higher-order relationships in ferns and relatives

(monilophytes). *American Journal of Botany*, *97*(9), 1444–1456. https://doi.org/10.3732/ajb.0900305

Ranker, T. A., & Haufler, C. H. (2008). *Biology and evolution of ferns and lycophytes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511541827

Raubeson, L. A., & Jansen, R. K. (1992). Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. *Science (New York, N.Y.)*, *255*(5052), 1697–1699. https://doi.org/10.1126/science.255.5052.1697

Rothfels, C. J., Larsson, A., Li, F.-W., Sigel, E. M., Huiet, L., Burge, D. O., Ruhsam, M., Graham, S. W., Stevenson, D. W., Wong, G.-S., Korall, P., & Pryer, K. M. (2013). Transcriptome-mining for single-copy nuclear markers in ferns. *PLoS One*, *8*(10), e76957. https://doi.org/10.1371/journal.pone.0076957

Rothfels, C. J., Li, F.-W., Sigel, E. M., Huiet, L., Larsson, A., Burge, D. O., Ruhsam, M., Deyholos, M., Soltis, D. E., Stewart, C. N., Shaw, S. W., Pokorny, L., Chen, T., Pamphilis, C., DeGironimo, L., Chen, L. I., Wei, X., Sun, X., Korall, P., … Pryer, K. M. (2015). The evolutionary history of ferns inferred from 25 low-copy nuclear genes. *American Journal of Botany*, *102*(7), 1089–1107. https://doi.org/10.3732/ajb.1500089

RStudio Team. (2015). *RStudio: Integrated Development for R*. RStudio, Inc. Retrieved August 30, 2019, from https://www.rstudio.com/

Sayyari, E., & Mirarab, S. (2016). Fast coalescent-based computation of local branch support from quartet frequencies. *Molecular Biology and Evolution*, *33*(7), 1654–1668. https://doi.org/10.1093/molbev/msw079

Schuettpelz, E., Korall, P., & Pryer, K. M. (2006). Plastid atpA data provide improved support for deep relationships among ferns. *Taxon*, *55*(4), 897–906. https://doi.org/10.2307/25065684

Schuettpelz, E., & Pryer, K. M. (2007). Fern phylogeny inferred from 400 leptosporangiate species and three plastid genes. *Taxon*, *56*(4), 1037–1050. https://doi.org/10.2307/25065903

Seeb, J. E., Carvalho, G., Hauser, L., Naish, K., Roberts, S., & Seeb, L. W. (2011). Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in non-model organisms. *Molecular Ecology Resources*, *11*(s1), 1–8. https://doi.org/10.1111/j.1755-0998.2010.02979.x

Sessa, E. B., Testo, W. L., & Watkins, J. E. Jr (2016). On the widespread capacity for, and functional significance of, extreme inbreeding in ferns. *New Phytologist*, *211*(3), 1108–1119. https://doi.org/10.1111/nph.13985

Sessa, E. B., Zimmer, E. A., & Givnish, T. J. (2012). Unraveling reticulate evolution in North American *Dryopteris* (Dryopteridaceae). *BMC Evolutionary Biology*, *12*(1), 104. https://doi.org/10.1186/1471-2148-12-104

Shen, H., Jin, D., Shu, J.-P., Zhou, X.-L., Lei, M., Wei, R., Shang, H., Wei, H.-J., Zhang, R., Liu, L. I., Gu, Y.-F., Zhang, X.-C., & Yan, Y.-H. (2018). Large-scale phylogenomic analysis resolves a backbone phylogeny in ferns. *GigaScience*, *7*(2), 1–11. https://doi.org/10.1093/gigascience/gix116

Sigel, E. M., Schuettpelz, E., Pryer, K. M., & Der, J. P. (2018). Overlapping patterns of gene expression between gametophyte and sporophyte phases in the fern *Polypodium amorphum* (Polypodiales). *Frontiers in Plant Science*, *9*, 1–17. https://doi.org/10.3389/fpls.2018.01450

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*(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Soltis, D. E., & Soltis, P. S. (1986). Electrophoretic evidence for inbreeding in the fern *Botrychium virginianum* (ophioglossaceae). *American Journal of Botany*, *73*(4), 588–592. https://doi.org/10.1002/j.1537-2197.1986.tb12078.x

Soltis, D. E., & Soltis, P. S. (1992). The distribution of selfing rates in homosporous ferns. *American Journal of Botany*, *79*(1), 97–100. https://doi.org/10.1002/j.1537-2197.1992.tb12628.x

Soltis, P. S., & Soltis, D. E. (1990). Evolution of inbreeding and outcrossing in ferns and fern-allies. *Plant Species Biology*, *5*(1), 1–11. https://doi.org/10.1111/j.1442-1984.1990.tb00187.x

Stensvold, M. C., & Farrar, D. R. (2017). Genetic diversity in the worldwide *Botrychium lunaria* (Ophioglossaceae) complex, with new species and new combinations. *Brittonia*, *69*(2), 148–175. https://doi.org/10.1007/s12228-016-9455-z

Stensvold, M. C., Farrar, D. R., & Johnson-Groh, C. (2002). Two new species of moonworts (*Botrychium* subg. *Botrychium*) from Alaska. *American Fern Journal*, *92*(2), 150–160. https://doi.org/10.1640/0002-8444(2002)092[0150:TNSOMB]2.0.CO;2

Swartz, L. M., & Brunsfeld, S. J. (2002). The morphological and genetic distinctness of *Botrychium minganense* and *B. crenulatum* as assessed by morphometric analysis and RAPD markers. *American Fern Journal*, *92*(4), 249–269. https://doi.org/10.1640/0002-8444(2002)092[0249:TMAGDO]2.0.CO;2

Testo, W., & Sundue, M. (2016). A 4000-species dataset provides new insight into the evolution of ferns. *Molecular Phylogenetics and Evolution*, *105*, 200–211. https://doi.org/10.1016/j.ympev.2016.09.003

Veselý, P., Bureš, P., Šmarda, P., & Pavlíček, T. (2012). Genome size and DNA base composition of geophytes: The mirror of phenology and ecology? *Annals of Botany*, *109*(1), 65–75. https://doi.org/10.1093/aob/mcr267

Walker, L. R. (1994). Effects of fern thickets on woodland development on landslides in Puerto Rico. *Journal of Vegetation Science*, *5*(4), 525–532. https://doi.org/10.2307/3235979

Wang, L.-G., Lam, T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones, B. R., Bradley, T., Zhu, H., Guan, Y. I., Jiang, Y., & Yu, G. (2020). Treeio: An R package for phylogenetic tree input and output with richly annotated and associated data. *Molecular Biology and Evolution*, *37*(2), 599–603. https://doi.org/10.1093/molbev/msz240

Weizhong, L., & Godzik, A. (2006). Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics (Oxford, England)*, *22*(13), 1658–1659. https://doi.org/10.1093/bioinformatics/btl158

Wickett, N. J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M. S., Burleigh, J. G., Gitzendanner, M. A., Ruhfel, B. R., Wafula, E., Der, J. P., Graham, S. W., Mathews, S., Melkonian, M., Soltis, D. E., Soltis, P. S., Miles, N. W., … Leebens-Mack, J. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences of the USA*, *111*(45), E4859–E4868. https://doi.org/10.1073/pnas.1323926111

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*, 2nd ed. Springer International Publishing. https://doi.org/10.1007/978-3-319-24277-4

Wikström, N., & Kenrick, P. (2001). Evolution of Lycopodiaceae (Lycopsida): Estimating divergence times from rbcL gene sequences by use of nonparametric rate smoothing. *Molecular Phylogenetics and Evolution*, *19*(2), 177–186. https://doi.org/10.1006/mpev.2001.0936

Williams, E. W. (2021). Population genetics of species in the genera *Botrychium* and *Botrypus* (Ophioglossaceae). *American Fern Journal*, *111*(2), 129–146. https://doi.org/10.1640/0002-8444-111.2.129

Williams, E. W., Farrar, D. R., & Henson, D. (2016). Cryptic speciation in allotetraploids: Lessons from the *Botrychium matricariifolium* (Ophioglossaceae) complex. *American Journal of Botany*, *103*(4), 1–14.

Winther, J. L., & Friedman, W. E. (2007). Arbuscular mycorrhizal symbionts in *Botrychium* (Ophioglossaceae). *American Journal of Botany*, *94*(7), 1248–1255. https://doi.org/10.3732/ajb.94.7.1248

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 USA*, *106*(33), 13875–13879. https://doi.org/10.1073/pnas.0811575106

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T.-T.-Y. (2017). ggtree: An r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36. https://doi.org/10.1111/2041-210X.12628

Zhang, C., Scornavacca, C., Molloy, E. K., & Mirarab, S. (2020). ASTRAL-Pro: Quartet-based species-tree inference despite paralogy. *Molecular Biology and Evolution*, *37*(11), 3292–3307. https://doi.org/10.1093/molbev/msaa139

Zhong, B., Fong, R., Collins, L. J., McLenachan, P. A., & Penny, D. (2014). Two new fern chloroplasts and decelerated evolution linked to the long generation time in tree ferns. *Genome Biology and Evolution*, *6*(5), 1166–1173. https://doi.org/10.1093/gbe/evu087

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Mossion, V., Dauphin, B., Grant, J., Kessler, M., Zemp, N., & Croll, D. (2022). Transcriptome-wide SNPs for *Botrychium lunaria* ferns enable fine-grained analysis of ploidy and population structure. *Molecular Ecology Resources*, 22, 254–271. https://doi.org/10.1111/1755-0998.13478