Molecular Phylogenetics and Evolution 120 (2018) 342–353

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Rapid allopolyploid radiation of moonwort ferns (*Botrychium*; Ophioglossaceae) revealed by PacBio sequencing of homologous and homeologous nuclear regions

Benjamin Dauphin[a,b,*], Jason R. Grant[a], Donald R. Farrar[c], Carl J. Rothfels[b]

[a] *Laboratoire de Génétique Évolutive, Université de Neuchâtel, rue Emile-Argand 11, 2000 Neuchâtel, Switzerland*
[b] *Department of Integrative Biology, University of California at Berkeley, 3040 Valley Life Sciences Building, Berkeley, CA 94720-3140, USA*
[c] *Department of Ecology, Evolution, and Organismal Biology, Iowa State University, 353 Bessey Hall, Ames, IA 50011-1020, USA*

ARTICLE INFO

Keywords:
Ferns
Low-copy markers
PacBio
Polyploid network
PURC
Reticulate evolution

ABSTRACT

Polyploidy is a major speciation process in vascular plants, and is postulated to be particularly important in shaping the diversity of extant ferns. However, limitations in the availability of bi-parental markers for ferns have greatly limited phylogenetic investigation of polyploidy in this group. With a large number of allopolyploid species, the genus *Botrychium* is a classic example in ferns where recurrent polyploidy is postulated to have driven frequent speciation events. Here, we use PacBio sequencing and the PURC bioinformatics pipeline to capture all homeologous or allelic copies of four long (~1 kb) low-copy nuclear regions from a sample of 45 specimens (25 diploids and 20 polyploids) representing 37 *Botrychium* taxa, and three outgroups. This sample includes most currently recognized *Botrychium* species in Europe and North America, and the majority of our specimens were genotyped with co-dominant nuclear allozymes to ensure species identification. We analyzed the sequence data using maximum likelihood (ML) and Bayesian inference (BI) concatenated-data ("gene tree") approaches to explore the relationships among *Botrychium* species. Finally, we estimated divergence times among *Botrychium* lineages and inferred the multi-labeled polyploid species tree showing the origins of the polyploid taxa, and their relationships to each other and to their diploid progenitors. We found strong support for the monophyly of the major lineages within *Botrychium* and identified most of the parental donors of the polyploids; these results largely corroborate earlier morphological and allozyme-based investigations. Each polyploid had at least two distinct homeologs, indicating that all sampled polyploids are likely allopolyploids (rather than autopolyploids). Our divergence-time analyses revealed that these allopolyploid lineages originated recently—within the last two million years—and thus that the genus has undergone a recent radiation, correlated with multiple independent allopolyploidizations across the phylogeny. Also, we found strong parental biases in the formation of allopolyploids, with individual diploid species participating multiple times as either the maternal or paternal donor (but not both). Finally, we discuss the role of polyploidy in the evolutionary history of *Botrychium* and the interspecific reproductive barriers possibly involved in these parental biases.

1. Introduction

Polyploidy is a major speciation process in plants (Stebbins, 1950; Soltis and Soltis, 2009; Rothfels and Otto, 2016), particularly within ferns, which have rates of polyploid speciation approximately twice those of angiosperms (Otto and Whitton, 2000; Wood et al., 2009). The fern genus *Botrychium* (Ophioglossaceae) is a classic example where recurrent inter-specific hybridization between diploid progenitors has been postulated to have formed an allopolyploid complex (Gilman et al., 2015; Williams et al., 2016; for other examples in ferns see, e.g. *Myriopteris*, Grusz et al. 2009; *Dryopteris*, Sessa et al., 2012; *Cystopteris*, Rothfels et al., 2014). *Botrychium* is the largest genus of the Botrychioideae subfamily, containing approximately 48 taxa (24 diploids and 24 polyploids), of which 35 are currently accepted and 13 remain under investigation (Pteridophyte Phylogeny Group 2016; Dauphin et al., 2017). *Botrychium* species are small perennial plants, colloquially known as moonworts, and have a relatively simple form constituted by a common stalk divided into a trophophore and a sporophore (Clausen, 1938). Bisexual gametophytes of *Botrychium* are underground (Wagner et al., 1985), non photosynthetic, and colonized by endomycorrhizal species of Glomeromycota (Winther and Friedman, 2007). The genus has a broad circumtemperate to circumboreal distribution, primarily in

* Corresponding author at: Laboratoire de Génétique Évolutive, Université de Neuchâtel, rue Emile-Argand 11, 2000 Neuchâtel, Switzerland.
E-mail address: benjamin.dauphin@unine.ch (B. Dauphin).

https://doi.org/10.1016/j.ympev.2017.11.025
Received 5 September 2017; Received in revised form 12 November 2017; Accepted 30 November 2017
Available online 11 December 2017
1055-7903/

Table 1
Allele counts and coverage in parentheses per taxon.

Taxa	Accession	Ploidy	Locus			
			ApPEFP_C	*CRY2cA*	*CRY2cB*	*transducin*
B. alaskense var. *alaskense*	BD1204709	4	2 (48, 34)	2 (50, 42)	1 (70)	2 (40, 19)
B. alaskense var. *salchaketense*	BD1206801	4	2 (21, 17)	2 (26, 24)	1 (40)	2 (49, 45)
B. angustisegmentum	DF19039	2	1 (40)	1 (43)	1 (85)	1 (49)
B. ascendens	DF17467	4	2 (15, 6)	2 (22, 17)	1 (22)	2 (19, 8)
B. boreale	BD1203001	4	2 (24, 24)	2 (30, 25)	1 (47)	2 (14, 9)
B. campestre var. *campestre*	DF954	2	1 (23)	1 (57)	1 (44)	1 (31)
B. campestre var. *lineare*	DF16696	2	1 (54)	1 (10)	1 (5)	1 (17)
B. crenulatum	DF19611	2	1 (75)	1 (24)	NA	1 (73)
B. dusenii	DF17085	4	2 (53, 47)	2 (49, 39)	1 (5)	2 (36, 27)
B. echo	DF17698	4	2 (49, 42)	2 (60, 30)	2 (53, 43)	2 (100, 56)
B. farrarii	DF18455	2	1 (58)	1 (13)	1 (13)	1 (7)
B. furculatum	DF19259	4	1 (56)	1 (106)	1 (157)	2 (48, 36)
B. gallicomontanum	DF14083	4	1 (6)	1 (5)	1 (5)	2 (63, 53)
B. hesperium	DF18178	4	2 (80, 53)	2 (37, 34)	2 (73, 46)	2 (66, 52)
B. lanceolatum “green”	BD1202311	2	1 (30)	1 (25)	1 (52)	1 (25)
B. lanceolatum “green”	DF19171	2	1 (71)	1 (85)	1 (137)	1 (49)
B. lanceolatum “green”	DF19337	2	1 (48)	1 (43)	1 (62)	1 (52)
B. lanceolatum “red”	DF19326	2	1 (30)	1 (31)	1 (44)	1 (10)
B. lanceolatum “red” × *B. pallidum*	DF17272	4	2 (48, 34)	2 (31, 30)	2 (28, 25)	2 (26, 14)
B. lunaria 2	AM11I	2	1 (113)	1 (93)	1 (16)	1 (37)
B. lunaria var. *lunaria*	BD1204116	2	1 (21)	1 (10)	NA	NA
B. lunaria var. *melzeri*	DF19145	2	1 (29)	1 (35)	NA	1 (9)
B. matricariifolium	DF19017	4	2 (30, 29)	2 (8, 7)	2 (26, 7)	NA
B. matricariifolium	DF19067	4	2 (48, 25)	2 (28, 28)	2 (45, 41)	2 (20, 13)
B. michiganense	DF18145	4	2 (45, 33)	2 (42, 32)	2 (33, 33)	2 (27, 10)
B. minganense	BD1205713	4	2 (53, 41)	2 (33, 23)	1 (57)	2 (51, 42)
B. minganense	DF17538	4	2 (65, 50)	2 (38, 27)	1 (38)	2 (21, 11)
B. montanum	DF18528	2	1 (22)	1 (27)	1 (24)	1 (28)
B. mormo	DF890	2	1 (30)	1 (35)	1 (55)	1 (32)
B. neolunaria	DF18299	2	1 (23)	1 (16)	NA	1 (5)
B. neolunaria	DF19201	2	1 (58)	1 (85)	1 (14)	1 (86)
B. nordicum	DF17819	2	1 (30)	1 (25)	NA	1 (13)
B. pallidum	DF17260	2	1 (23)	1 (11)	1 (24)	1 (19)
B. pallidum	DF18729	2	1 (17)	1 (25)	1 (31)	1 (13)
B. paradoxum	DF16672	4	2 (39, 12)	1 (56)	1 (74)	2 (29, 20)
B. pedunculosum	DF5413	4	2 (80, 56)	2 (41, 34)	2 (37, 24)	2 (33, 15)
B. pinnatum	BD1206619	4	2 (78, 22)	2 (43, 23)	1 (56)	2 (58, 50)
B. pseudopinnatum	DF18314	6	3 (21, 18, 15)	3 (33, 32, 20)	2 (46, 41)	3 (57, 49, 41)
B. pumicola	DF17915	2	1 (110)	1 (22)	1 (36)	1 (60)
B. simplex 2	DF17299	2	1 (27)	1 (16)	1 (17)	1 (8)
B. simplex var. *compositum*	DF18799	2	1 (65)	1 (18)	1 (11)	1 (40)
B. spathulatum	DF18803	4	2 (43, 35)	1 (81)	1 (8)	2 (58, 32)
B. tenebrosum	AM1502	2	1 (66)	1 (73)	1 (44)	1 (57)
B. tenebrosum	DF17416	2	1 (60)	1 (60)	1 (44)	1 (30)
B. tunux	AM23A	2	1 (22)	1 (55)	1 (13)	1 (97)
Botrypus virginianus	BD1206330	4	1 (52)	1 (49)	2 (101, 48)	1 (9)
Sceptridium multifidum	AM1503	2	1 (5)	1 (5)	NA	1 (80)
Sceptridium multifidum	BD1205311	2	1 (52)	1 (40)	1 (61)	1 (49)

NA, no data available since PCR amplification failed multiple times. Both *B. lunaria* 2 and *B. simplex* 2 are under investigation to determine whether or not they warrant recognition as a new species or variety.

the northern hemisphere, where species occur in various specific habitats, for example in alpine meadows or along older roadsides and utility corridors.

Allopolyploidy in *Botrychium* was first hypothesized based on taxa with morphological characters intermediate between their putative parents (Wagner and Wagner, 1981, 1983, 1986). In contrast, several polyploids are superficially indistinguishable from diploid taxa, and were first identified by karyological studies that determined their polyploid status and chromosome number (Wagner and Lord, 1956; Wagner, 1993). The hypotheses of polyploidy (and specifically allopolyploidy) were then refined and greatly elaborated by 15 years of investigative work by D.R. Farrar (see Farrar, 2011), who characterized the genetic profiles of nearly all *Botrychium* taxa in North America (see also Hauk and Haufler, 1999) based on 22 co-dominant enzyme loci. These data enabled the reconstruction of the bi-parental origin of allopolyploids and the proposal of new hybridization hypotheses (Stensvold et al., 2002; Zika and Farrar, 2009; Gilman et al., 2015). This genetic fingerprinting has since been complemented by other molecular tools, such as amplified fragment length polymorphism (AFLP) or microsatellites, to provide new insights on the genetic diversity within specific allopolyploid complexes (Williams and Waller, 2012; Ellis, 2014; Williams et al., 2016).

Previous molecular phylogenies have mostly focused on the relationships of maternal lineages among *Botrychium* taxa (Vogel et al., 1998; Hauk et al., 2012; Williams and Waller, 2012; Dauphin et al., 2014; Dauphin et al., 2017). The genus comprises an estimated 48 taxa divided among three major clades: Lanceolatum (3 diploids, 14 polyploids), Lunaria (10 diploids, 1 polyloid), and Simplex-Campestre (11 diploids, 9 polyploids; Dauphin et al., 2017). Previous studies additionally revealed various morphologically cryptic taxa (including both diploids and polyploids) and provided a phylogenetic foundation that helped identify diagnostic morphological characters (Stensvold and Farrar, 2017; Williams et al., 2016; Meza-Torres et al., 2017; Farrar and Gilman, 2017). However, despite the recent advances in our

understanding of the maternal phylogenetic relationships among *Botrychium* taxa, phylogenetic information about the paternal lines through nuclear sequencing is lacking.

Low-copy nuclear regions, which provide information on both maternal and paternal evolutionary histories, are crucial for phylogenetic studies of polyploid groups. Unfortunately, there are no universal primer sets targeting a single-copy nuclear gene available for ferns, probably due to limited genetic resources, and the deep divergences within ferns; for example, Ophioglossidae, the subclass to which *Botrychium* belongs (PPG I 2016), diverged from its extant sister group approximately 350 mya (Rothfels et al., 2015). However, with recent advances in sequencing technologies, pteridologists have new opportunities to target variable nuclear regions in their group of interest (Rothfels et al., 2013). In particular, it is now possible to investigate reticulate evolution of polyploid groups by simultaneously sequencing all homeologs present within individual accessions without cloning, and thus to capture and characterize the bi-parental copies of a targeted gene (Rothfels et al., 2017).

In this study, we used the PacBio sequencing approach, in conjunction with the PURC analysis pipeline (Rothfels et al., 2017), to investigate the nuclear phylogenetic relationships of *Botrychium* diploid species and the reticulate histories of allopolyploids, infer their origins, and characterize the temporal scale of allopolyploidy events. Moreover, we investigated the combination of diploids in the formation of allopolyploids and their tendencies to be either maternal or paternal progenitors. Finally, we discuss our results in a broader context about the role of polyploidy in generating new species and the possible reproductive barriers involved in the composition of parental contributors of allopolyploids.

2. Materials and methods

2.1. Plant material and DNA extraction

We analyzed 45 specimens (25 diploids and 20 polyploids) of *Botrychium* and three outgroup accessions from the related genera *Botrypus* and *Sceptridium*; in total we sampled 37 distinct named *Botrychium* taxa (Table 1). Our sampling includes representatives of each of the three major clades of *Botrychium* (Lanceolatum, Lunaria, and Simplex-Campestre; Hauk et al., 2012), and encompasses the phylogenetic diversity found in each of them, as inferred from plastid loci (Dauphin et al., 2014). Importantly, 35 of the 45 *Botrychium* specimens included were additionally assayed for 22 allozyme loci to ensure species identification and to connect our sequence-based results with the existing wealth of allozyme data. All vouchers were deposited at either the herbarium of the University of Neuchâtel (NEU) or at the Ada Hayden Herbarium (ISC) at Iowa State University.

Genomic DNA was extracted from herbarium specimens or silica-dried material using the DNeasy Plant Mini Kit according to the manufacturer's protocol (Qiagen, Hilden, Düsseldorf, Germany). Quantity and quality of total DNA was inspected with spectrophotometry (NanoDrop 2000, Thermo Scientifc, Wilmington, USA), fluorescence (Qubit Fluorometric Quantification, Thermo Scientifc, Wilmington, USA), and on agarose gels (1%).

2.2. Choice of candidate genes, PCR amplification tests and optimization

Primer sets were designed from transcriptome data from *Botrypus* and *Sceptridium*, targeting intron-containing regions approximately 1 kb in length for 10 single-copy nuclear markers (*ApPEFP_C*, *CRY2*, *CRY4*, *DET1*, *gapCpSh*, *IBR3*, *pgiC*, *SQD1*, *TPLATE*, and *transducin*; Rothfels et al., 2013) using the program Primer3 (Koressaar and Remm 2007), as implemented in Geneious v.6.1 (Kearse et al., 2012; http://www.geneious.com). PCR amplification was carried out in 25 µl reactions using 12.5 µl of GoTaq G2 colorless master mix (Promega, Madison, Wisconsin, USA), 1 mM of $MgCl_2$, 10 µM of each primer, and 5 ng of DNA. For these amplification tests, material included representatives of each major clade of *Botrychium*, and two outgroups (*Botrypus virginianus* and *Sceptridium dissectum*; Hauk et al., 2003). Approximately 4–10 primer combinations were tested on gradient PCR for each locus with standard PCR conditions: initial denaturation at 94 °C for 180 sec, 35 cycles of 94 °C for 60 sec, 48–65 °C for 45 sec, and 72 °C for 90 sec, with a final extension at 72 °C for 600 sec. The highest annealing temperature in the PCR gradient was adjusted for each primer set, taking the lower T_M of the two primers, subtracting 5 °C, and decreasing up to T_M − 14 °C. Amplified products were visualized on 1% agarose gel at 100 V for 90 min to confirm amplification success.

2.3. Sanger sequencing and variability tests

To assess phylogenetic informativeness at the species level for each marker, we sequenced a subset of diploid species using an ABI3130 XL Automated Sequencer by Macrogen Europe (Amsterdam, Netherlands), using the same amplification primers. We roughly estimated the variability of each locus although several chromatograms were ambiguous with multiple peaks. For the *SQD1* region, the largest band (~800 bp) was extracted using the QIAquick Gel Extraction kit and manufacturer's guidelines (Qiagen, Hilden, Düsseldorf, Germany), and purified before sequencing.

2.4. Pool preparation and the PacBio sequencing of homologous and homeologous copies

Four regions were both consistently amplifiable and informative at the species level in *Botrychium*: *ApPEFP_C*, *CRY2cA*, *CRY2cB*, and *transducin* (Table 2). For each of these regions we designed barcoded forward primers using the first 48 primers provided by Pacific Biosciences and followed the data-generation protocol of Rothfels et al. (2017). Briefly, each sample was individually barcoded with a unique 16-bp sequence linked to the 5′ end of the forward primer (reverse primers were not barcoded). PCR products were inspected on a 1% agarose gel and band intensity was scored by eye according to six categories: blank, very weak, weak, medium, strong, and very strong. These band strengths were taken to correspond approximately to concentrations of 5, 10, 15, 20, 30, and 50 ng/µl respectively (Rothfels et al., 2017). Amplicons were pooled in approximately equal quantities based on these concentration estimates (doubled for tetraploid taxa), keeping a constant targeted sequencing coverage of about 36X for the target sequences, for a total of 414 potential targeted sequences. The pooled amplicons were cleaned with an AMPure XP bead purification as recommended by Pacific Biosciences (2016).

Table 2
Primer sets and PCR conditions.

Protein region	Primer names (Forward, Reverse)	Sequence (5′–3′) (Forward, Reverse)	PCR Program[a]	Expected length (bp)
ApPEFP_C	76AP_F2, 559AP_R2	TCATTGGGTGGGTGCAGGACTG, CTGCTGGAAGTGCAGTTATTCT	659035	900
CRY2	1749C2_F5, 4483C2_R4	TGGGAAGTCAATGATGATCAAGG, GCCKTACATACTCTCCATGAGG	589035	1150
transducin	1904TR_F4, 1826TR_R6	GAAAGAGGTAGAAGCTCATTCTGG, TGAGCAAACAARGACCGGTGGC	609035	1300

[a] First two digits is the annealing temperature (°C), the following two is elongation time (in seconds), and the two last is the number of cycles.

Table 3
Dataset characteristics.

Locus	Aligned length in bp	No. specimens	Missing data[a]	No. Alleles	No. Parsimony information sites	No. Variable sites	No. SNPs	Best model	lnL	BIC scores	Implemented model in RaxML	Implemented model in MrBayes	Implemented model in BEAST2
ApPEPC	905 (915)	45 (48)	0%	21 (23)	30 (66)	41 (116)	11 (50)	HKY + G	−2048.2941	5044.4186	HKY + G	HKY + G	HKY + G
CRY2cA	1037 (1115)	45 (48)	0%	30 (32)	38 (93)	51 (138)	13 (45)	TIM3 + G	−2481.5967	5270.7011	HKY + G	GTR + G	TN93 + G
CRY2cB	1066 (1084)	40 (42)	13%	23 (26)	31 (65)	48 (132)	17 (65)	HKY + I	−2498.3041	5758.3452	HKY + G	HKY + I	HKY + I
transducin	1221 (1764)	43 (46)	4%	30 (33)	77 (117)	96 (259)	19 (142)	HKY + G	−4026.9818	9063.1344	HKY + G	HKY + G	HKY + G
All loci	4229 (4878)	45 (48)	4%	104 (114)	176 (341)	236 (645)	60 (302)				Unlinked	Unlinked	Unlinked
Overall in %					4.16 (6.99)	5.58 (13.22)	1.41 (6.19)						

Note: outgroups are considered in parentheses.
[a] For each locus, a specimen was counted as missing data when no allele or homeolog was successfully sequenced.

We submitted 150 µl of the cleaned pool for library preparation using the P6-C4 chemistry and the SMRTbell™ Template Prep Kit, which includes a repair-ends and ligation reaction, and a purification step (Pacific Biosciences, 2016). Sequencing was conducted on a single PacBio SMRT cell, based on Circular Consensus Sequencing (CCS) technology (Travers et al., 2010) using the PacBio RS II platform (Pacific Biosciences, 2015). Library preparation and sequencing were performed at the Sequencing and Genomic Technologies Core Facility of the Duke University Center for Genomic and Computational Biology (Durham, North Carolina, USA).

2.5. Bioinformatics pipeline

We processed the raw CCS reads with the Pipeline for Untangling Reticulate Complexes (PURC; Rothfels et al., 2017), which relies on four other dependencies: BLAST+ v.2.4 (Camacho et al., 2009); CUTADAPT v.1.3 (Martin, 2011); MUSCLE v.3.8.31 (Edgar, 2004); and USEARCH v.8.1 (Edgar, 2010). Thus, we converted the fastq file to fasta and removed all raw CCS reads that were < 600 bases long or had more than five expected errors using USEARCH's fastq_filter command (Edgar, 2010). Then, PURC was run on the filtered CCS to de-multiplex the reads, remove primer and barcode sequences, annotate the reads with locus and accession names, remove chimeras, correct sequencing and PCR errors through an iterative clustering approach, and infer final alignments for each locus (Rothfels et al., 2017).

To explore the effectiveness of PURC in inferring the true biological sequences from the PacBio CCS reads, we tested ten different clustering regimes (Supplemental Table 1). The first five regimes (a, b, c, d, and e) followed the default UCHIME settings (Edgar et al., 2011), required a minimum cluster size of five reads for a cluster to be retained, and differed in their sequence-similarity cutoffs for two sequences to be clustered together (Supplemental Table 1). The second five regimes (f, g, h, i, and j) used the same clustering parameters as for the first five, but with more stringent chimera detection parameters (abundance_skew, minh, xn, and dn; Edgar et al., 2011; Supplemental Table 1). For each locus, the alignments from all regimes were merged and a phylogeny inferred from that master alignment using the program AliView v.1.18 (Larsson, 2014). The inspection of branchlength (similarity of sequences) and topology (polytomies) of these phylogenies revealed clear cross-regime consensuses as to what the true biological sequences were; results from those regimes that differed from this consensus were clearly due to, e.g., failures of that regime to capture all the chimeras present or to erroneous lumping together of distinct sequences (see Rothfels et al., 2017). The final sequences were submitted to GenBank (accession numbers MG710551-MG710617, MG710618-MG710682, MG710683-MG710733, and MG710734-MG710799) and final alignments of each locus were deposited on Dryad (https://doi.org/10.5061/dryad.6q8d6).

2.6. Combining homeologs of nuclear loci

All 48 specimens included in this study were previously analyzed in a global plastid phylogeny of *Botrychium* (Dauphin et al., 2017). Because the chloroplast genome is maternally inherited in ferns (Gastony and Yatskievych, 1992; Vogel et al., 1998; Guillon and Raquin, 2000), we deduced for each polyploid taxon the maternal homeolog (labeled "A") as the one that matched with its plastid phylogenetic placement. The second homoelog copy (labeled "B") is therefore considered to be paternal. For the concatenated data analyses the labeled homeologs (A and B) were treated as independent accessions.

2.7. Nuclear gene tree inference

We analyzed our four nuclear loci with maximum likelihood (ML) and Bayesian inference (BI) using RAxML v.8 (Stamatakis, 2014) and MrBayes v.3.2 (Ronquist et al., 2012), respectively, to infer individual

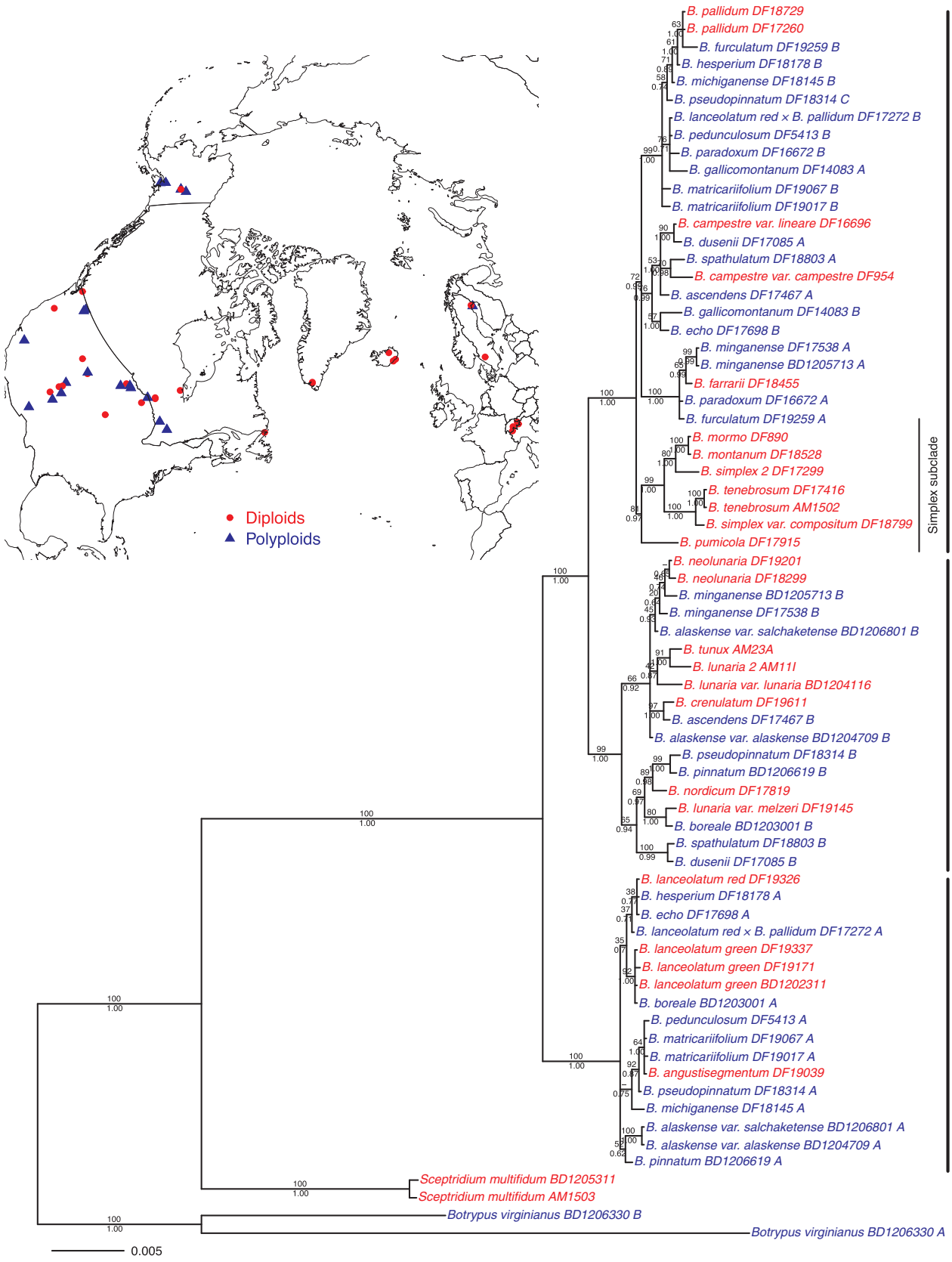

Fig. 1. Bayesian phylogram of the multi-labeled gene tree from the concatenated nuclear data. Maximum likelihood bootstrap support values are depicted above branches, and posterior probabilities are below branches. Diploid species are shown in red and the polyploids in blue. For the polyploids, each subgenome (that from the maternal progenitor and that from the paternal progenitor) appears on the tree individually: "A" for the maternal subgenome, "B" for the paternal subgenome, and "C" for the additional subgenome of the single hexaploid. The map shows the collection sites of the sampled accessions, with red circles for diploids and blue triangles for polyploids.

gene trees and the tree from the concatenated data. For each locus we implemented the optimal substitution model (Table 3), based on BIC scores (Posada and Buckley, 2004) calculated with the program jModelTest v.2 (Darriba et al., 2012). We ran ML tree searches using a randomized stepwise addition maximum parsimony starting tree and gamma-distributed site rates, and we performed 1000 bootstrap replicates on the best-scoring ML tree with a bootstrap random number seed. For the Bayesian inference, parameters were unlinked among partitions (loci) and each locus was permitted its own average rate. We implemented the same substitution models as for the ML analyses, except we used a GTR instead of a HKY for the *CRY2cA* locus (Table 3). We kept all other priors at their default values and we ran two independent runs of 10 million MCMC generations, each including three heated and one cold chain, a chain temp of 0.1, sampling trees every 1000 generations, and a burn-in of 25%. A phylogram summarizing the 7500 remaining trees was produced, on which PP were reported (Fig. 1).

2.8. Divergence time estimates

To estimate divergence times within *Botrychium*, we analyzed the concatenated nuclear data (partitioned by locus, substitution models as above) under an uncorrelated lognormal relaxed clock model using BEAST 2 v.2.4 (Bouckaert et al., 2014). Site models and clock models were unlinked among partitions, and a birth-death tree prior was applied. We set broad priors on the clock rate with an exponential distribution (mean of 10.0) for *ucldMean.c* and a gamma distribution ($\alpha = 0.5396$, $ß = 0.3819$) for *ucldStdev.c*. We constrained the age of three well-supported nodes, using two macrofossils, one attributable to *Botrypus* (Rothwell and Stockey, 1989) and the other to *Sceptridium* (Bozukov et al., 2010), and one secondarily derived age estimate for the divergence of the Lanceolatum and Lunaria clades (Stensvold, 2008). Using the fossils as minimum age of the clades, we applied a uniform prior distribution on the root (divergence *Botrypus-Sceptridium* + *Botrychium*) with a lower bound of 57.0 million years (myr), and a uniform prior for the divergence *Sceptridium-Botrychium* with a lower bound of 23.0 myr. We set a normal distribution for the divergence time of the Lanceolatum-Lunaria clades, with a mean of 5 myr and a sigma of 2.0 to keep broad priors, and we constrained the calibrated clades as monophyletic to facilitate the convergence of the analysis.

We ran four independent analyses of this model for 100 million generations, with parameters sampled every 1000 generations. We inspected convergence in Tracer v.1.6 (Drummond and Rambaut, 2007), and we excluded the first 10 million generations from each of the runs as burn-in; the effective sample size (ESS) of each parameter was greater than 300. We then summarized our post burn-in samples with TreeAnnotator v.2.4 (Bouckaert et al., 2014) to generate the maximum clade credibility (MCC) chronogram depicting the divergence time estimates with 95% highest posterior density (HPD) intervals.

2.9. Multi-labeled species tree inference

To investigate more deeply the polyploidization events, we inferred a multi-labeled species tree (or "genome tree") using AlloPPNet (Jones et al., 2013). Because AlloPPNet is limited to diploid and tetraploid taxa, for this analysis we removed accessions of the hexaploid *B. pseudopinnatum* (DF18314) and kept the remaining 47 specimens; 25 diploid, 19 tetraploid, and three outgroups for a total of 36 taxa, with one or two specimens for each taxon.

We used the R (R Core Team 2015) script and AlloPPNet R codes (AlloppDT_5beastxml_toplevel.r and AlloppDT_6beastxml_bits.r) available online (Jones, 2014) to generate a XML file implementable in BEAST v.1.8.2 (Drummond and Rambaut 2007). To facilitate convergence of the analysis, we manually added a starting tree (obtained from our previous ML phylogenetic analysis excluding the *B. pseudopinnatum* accession) and constrained the monophyly of the three major clades (Lanceolatum, Lunaria, and Simplex-Campestre). We partitioned the nuclear data by locus and kept the default HKY substitution model and strict molecular clock for each gene.

We ran two independent analyses of this model for 500 million generations, sampling parameters every 10,000 generations. We inspected consistency of parameters among the two runs using Tracer v.1.6 (Drummond and Rambaut 2007) to ensure convergence and adequate ESS values. Then, we generated a MCC tree with a burn-in of 10% using TreeAnnotator v.2.4 (Bouckaert et al., 2014). All phylogenetic analyses were run on the Cyberinfrastructure for Phylogenetic Research portal (CIPRES; Miller et al., 2010).

3. Results

3.1. Nuclear loci and PURC regimes comparison

Of the over 100 primer sets tested, only three consistently amplified long nuclear fragments in single copy (*ApPEFP_C* and *transducin*) or had similarly-sized copies that were indistinguishable when viewed on an agarose gel (*CRY2cA* and *CRY2cB*; Table 2). From the PacBio run, we obtained 60,151 CCS reads, of which 33,417 were > 600 bp long and had fewer than five expected errors (Supplementary Fig. 1). Cleaning the dataset, PURC detected and removed 180 interlocus concatemers, 10,316 sequences with failed barcodes, 970 sequences with more than two barcodes, and 85 sequences that were unclassifiable (contaminants that did not match any of the reference sequences).

In our PURC regime comparisons, we recovered similar numbers of inferred alleles between stringent (f, g, h, i, and j) and less stringent regimes (a, b, c, d, and e) for *ApPEFP_C* and *transducin*, but different numbers for *CRY2cA* and *CRY2cB*. Manual inspection of alignments showed that the less stringent regimes more consistently recovered the expected number of alleles, especially in cases where the CCS coverage was low (e.g., between 5 and 10X). Therefore, we retained the inferences from regime d as the most appropriate for *ApPEFP_C*, *CRY2cA*, and *transducin*, and regime c for *CRY2cB*, although some exceptions were applied for low CCS coverage sequences, based on manual inspections of the alignments.

Because intra-gametophytic selfing and homozygous populations are common in *Botrychium* (Hauk and Haufler, 1999; Stensvold and Farrar, 2017), we expected most diploids to have a single sequence for each locus, the tetraploids to have two sequences (fixed heterozygosity via allopolyploidy), and the single hexaploid to have three sequences. In general, our results matched expectations for the diploids (Table 1), although the tetraploids frequently had fewer sequences than expected (Table 1). In total, we obtained 23 different alleles for *ApPEFP_C*, 32 for *CRY2cA*, 26 for *CRY2cB*, and 33 for *transducin* (Table 3).

3.2. Nuclear phylogeny of the concatenated data

Our nuclear dataset contained four loci with alignment lengths ranging from 915 to 1764 base pairs (bp), for a total of 4878 aligned bp for 48 accessions (including outgroups). The *transducin* locus had the most variable sites and *ApPEFP_C* was the most conserved region (Table 3).

Topologies of the ML and BI trees inferred from the concatenated nuclear data were congruent, with 56% of branches strongly supported (BV > 70 and PP > 0.95; Fig. 1). We found strong support for the three major clades (Lanceolatum: 100/1.00; Lunaria: 99/1.00; Simplex-Campestre: 100/1.00; Fig. 1).

3.3. Divergence time estimates

All parameters from the divergence-time analyses had post-burnin pooled ESSs > 3200. The mean estimate for the *Botrypus – Sceptridium* divergence time was 60.20 myr (HPD, 57.00–71.3 myr), that for the *Sceptridium – Botrychium* divergence was 36.19 myr (HPD,

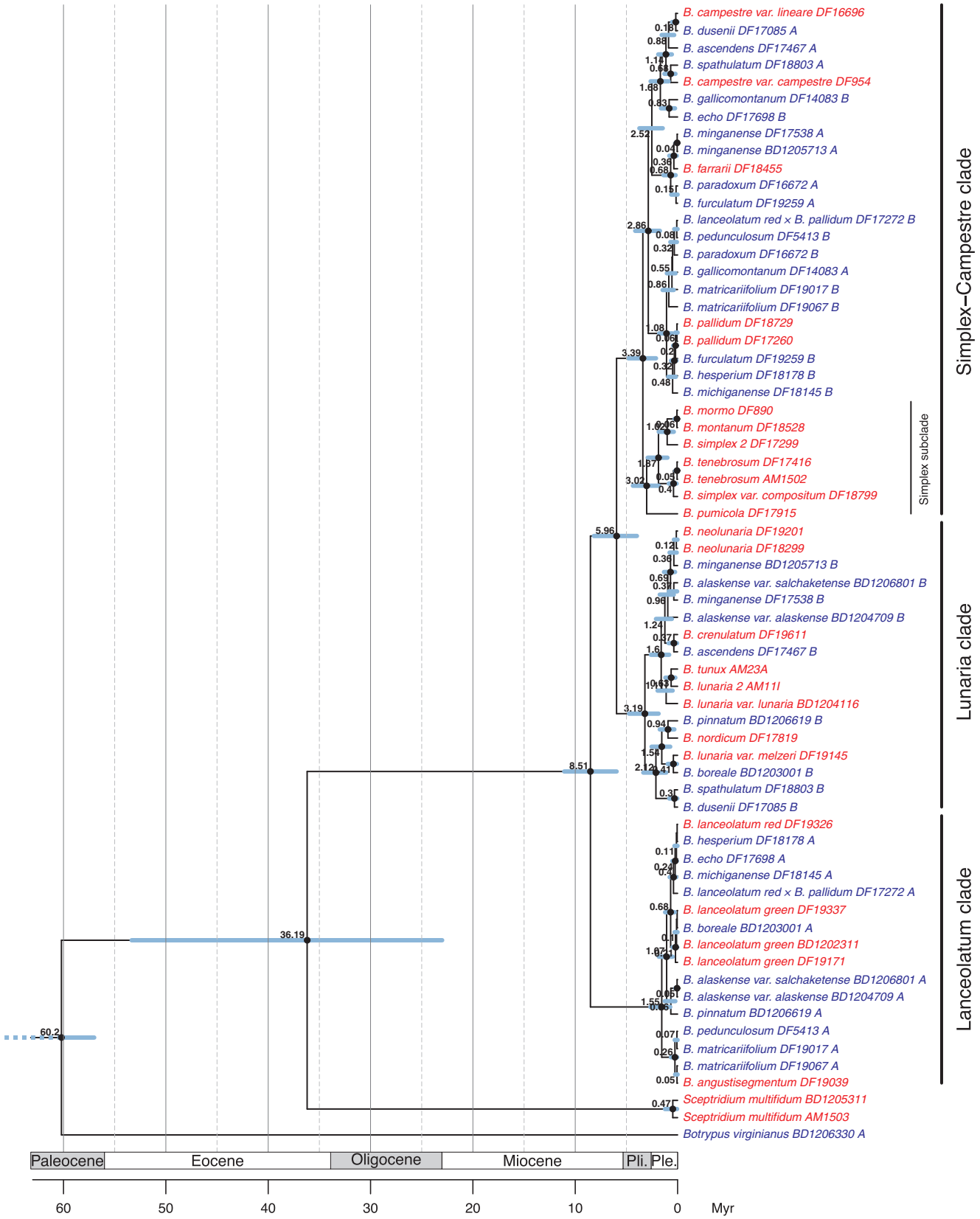

Fig. 2. Time-calibrated *Botrychium* phylogeny from the concatenated nuclear data. Black dots indicate a posterior probability > .95 and node ages are specified on the left side of nodes. The color scheme and subgenome designations are as in Fig. 1 and the geological time scale follows the International Commission on Stratigraphy (2012; http://www.stratigraphy.org/index.php/ics-chart-timescale).

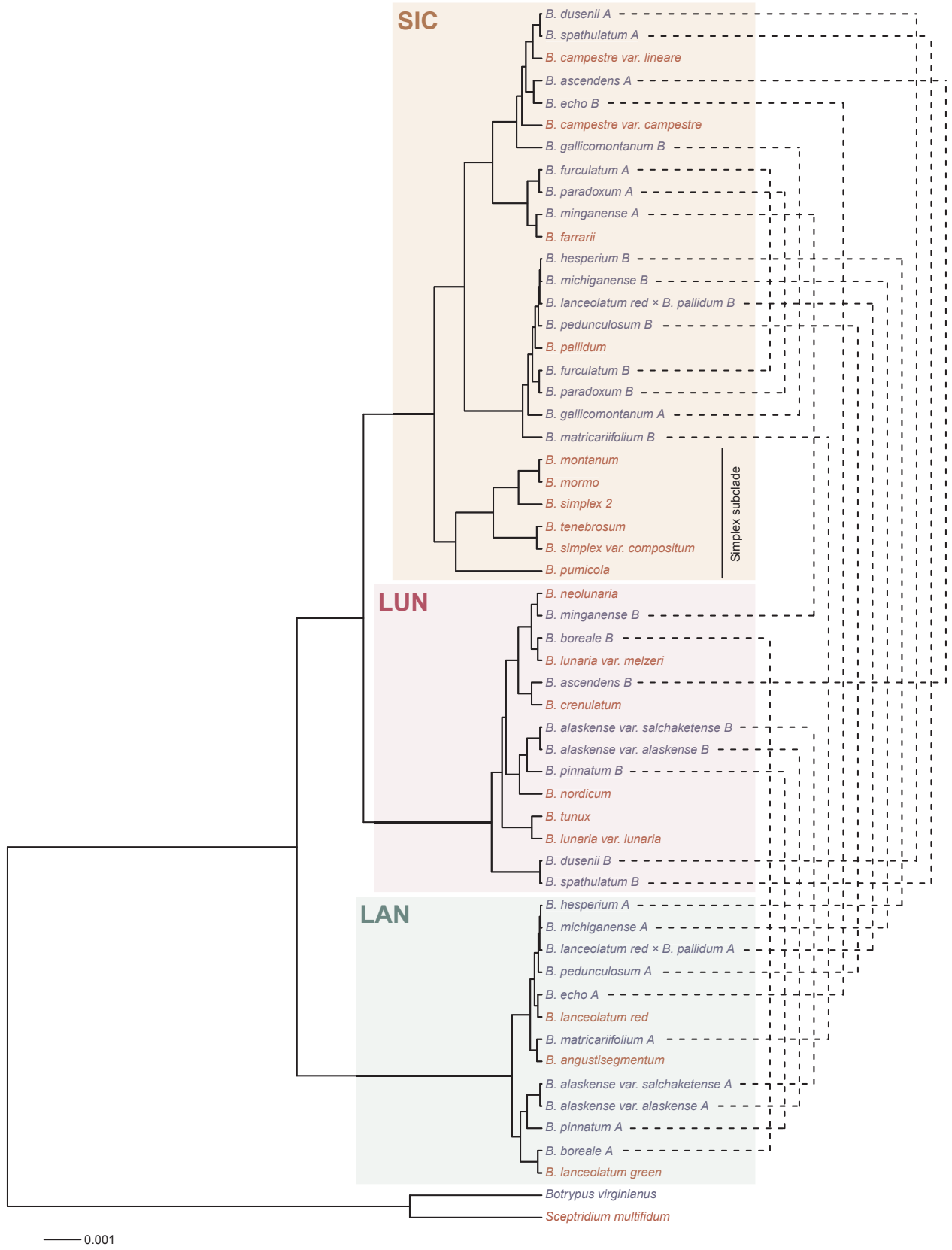

Fig. 3. The multi-labeled maximum clade credibility species tree inferred by AlloPPNet from our nuclear data. Dashed lines on the right-hand side connect subgenomes present in individual polyploid taxa. The color scheme and subgenome designations are as in Fig. 1.

23–53.34 myr), and that between the Lanceolatum and Lunaria clades was 8.51 myr (HPD, 5.96–11.08 myr; Fig. 2).

3.4. Multi-labeled species tree

In our allopolyploid network inference, the majority of parameters were estimated with ESS values > 300 for both combined runs, with some exceptions for the *apspnetwork* statistics. We infered a mean of 11 allopolyploidy events (HPD, 10–12). In total, 45,000 trees were retained and used to generate the MCC tree (highest log clade credibility = −94.707745; Fig. 3).

4. Discussion

4.1. Generating low-copy nuclear data for polyploid complexes

PacBio sequencing combined with the bioinformatics pipeline PURC (Pipeline for Untangling Reticulate Complexes; Rothfels et al., 2017) provides new opportunities to study groups with polyploid taxa and limited genomic resources. As a first result, we provided primer sets targeting four long nuclear loci for *Botrychium*, *Botrypus*, and *Sceptridium*, which now enables investigations of the evolutionary history of their nuclear genomes. Our success for generating low-copy nuclear data was primarily due to the availability of transcriptomes of species closely related to the group of interest (Rothfels et al., 2013), which made possible primer design. Based on a single PacBio run, we sequenced 45 specimens of *Botrychium* (25 diploids, 19 allotetraploids, and one hexaploid) and three outgroups, in a much faster and more economical way than is possible using conventional cloning and Sanger sequencing. The use of PURC greatly improved the detection of PCR-mediated chimeras and simultaneously retrieved with success the homologs of diploids and homeologs of allopolyploids.

4.2. Corroborating plastid-based inferences of relationship among diploid species

A first relevant outcome is the concordance of our nuclear phylogeny with earlier inferences based on plastid data for the relationships among the three major clades of *Botrychium*, with the first divergence separating the Lanceolatum (LAN) clade from the two other major lineages (the Lunaria [LUN] and Simplex-Campestre clade [SIC]; Hauk et al., 2012; Williams and Waller, 2012; Dauphin et al., 2014; Dauphin et al., 2017). The deep phylogenetic divergences among the genera *Botrychium*, *Botrypus*, and *Sceptridium*, are also consistent with those inferred in the plastid phylogeny of Ophioglossaceae (Hauk et al., 2003).

However, within the three clades themselves, there are several cases of plastid-nuclear incongruence among the diploids (see for comparison Dauphin et al., 2017), and some inconsistencies between our concatenated-based analysis and the species tree. In the LAN clade, the two diploids *B. angustisegmentum* and *B. lanceolatum* “green” are supported in two distinct clades, while *B. lanceolatum* “red” is embedded in a poorly supported group (Fig. 1). In contrast, in our species tree (Fig. 3), *B. angustisegmentum* and *B. lanceolatum* “red” form a monophyletic group, which is inconsistent with allozyme-based genetic distance that supports *B. lanceolatum* “red” as sister to the remaining LAN clade members (Stensvold and Farrar, 2017). Because of weak differentiation among the LAN diploids, presumably due to recent speciation events, these phylogenetic relationships remain unclear (Figs. 1–3).

Within the LUN clade, all diploid species are differentiated in our nuclear phylogeny, including recently published cryptic taxa. Compared to earlier plastid phylogenies, we found a greater than expected divergence between the varieties of *B. lunaria* (var. *lunaria* and *melzeri*), and between *B. lunaria* and *B. nordicum* (Fig. 3), which is congruent with their genetic distance based on the allozyme allele frequencies and supports the species status of *B. nordicum* (Stensvold and Farrar, 2017). However, the phylogenetic placements of the *B. lunaria* varieties and *B. tunux* were inconsistent between our concatenation-based analysis and the species tree.

Plastid and nuclear genomes generate a concordant phylogenetic signal for the SIC clade, with an exception for the placement of *B. pallidum*. In the plastid phylogenetic reconstruction, *B. pallidum* is the sister group to the rest of the Simplex subclade (ML = 95; PP = 1.00; see Fig. 4 in Dauphin et al., 2017), whereas that species is embedded within the Campestre clade in our nuclear phylogeny. Both the concatenated-data tree and the species tree support the clade formed by the four diploids *B. campestre* var. *campestre*, *B. campestre* var. *lineare*, *B. farrarii*, and *B. pallidum* (Figs. 1–3). This result is in agreement with morphological data, where *B. pallidum* displays a tendency to form two lobes on the basal pinna, typical of the Campestre clade (Wagner and Wagner, 1990).

4.3. Simplex subclade free of polypoids

Within the Simplex-Campestre clade, the Simplex subclade comprises six taxa (*B. montanum*, *B. mormo*, *B. pumicola*, *B. simplex* var. *compositum*, *B. simplex 2*, and *B. tenebrosum*), none of which form any allopolyploids (Figs. 1–3). Yet, the Simplex subclade is widely distributed in North America, and also occurs in several regions of Europe and its species are found with other diploid taxa in natural populations, which should allow inter-specific hybridization (Abbott, 2017). At least two pre-zygotic barriers may explain in part the apparent lack of polyploids involving members of this clade. First, the bisexual gametophytes of *Botrychium* taxa may have different phenological cycles so that non-synchrony of gametangial maturity in members of the Simplex subclade could possibly be involved, and needs further investigation. The second scenario is that cytogenetic incompatibilities prevent allopolyploidy between diploids of the Simplex subclade and the other members of *Botrychium*. Supporting this idea is the significantly smaller genome size of *B. simplex* in comparison with all other species of the genus (except *B. pallidum*; Williams and Waller, 2012), suggesting that differences in their chromosomal structure may prevent pairing in meiosis. Interestingly, several sterile hybrids (*B. lunaria* × *B. simplex*) were reported between diploid taxa of SIC and LUN, but these cytogenetic incompatibilities appear permeable at the intra-clade level and the diploid taxa of the Simplex subclade do show the capability to produce viable homoploid hybrids (between *B. simplex* var. *compositum* and *B. simplex 2*) within natural populations (Wagner, 1980, 1991; Wagner and Beitel, 1985; Wagner and Wagner, 1988). An analogous case of homoploid hybrids is found in the LUN clade where there are several introgressed genotypes between *B. lunaria* and *B. neolunaria* (Stensvold, 2008), but surprinsingly, these two closely related species have also formed an allopolyploid, *B. yaaxudakeit* (Stensvold et al., 2002; Stensvold and Farrar, 2017). Further works are critical to elucidate the mechanisms underlying their reproductive barriers at conspecific level.

4.4. Confirming parentage of allopolyploids

Reticulation is extensive in *Botrychium* and involves diploid progenitors among all three major clades. On the basis of our phylogenetic results, we identified most of the parental donors of polyploids, which largely corroborate earlier morphology and allozyme-based investigations (Figs. 3 and 4). Over half (12 of 20) of the diploid taxa contributed to at least one polyploid, and most show a clear bias in favor of being either a maternal or paternal progenitor. Most of the diploid donors contributed to two or three allotetraploids but some are much more prolific, such as *B. pallidum*, which is the progenitor of nine allotetraploids. *Botrychium pallidum’s* prolific polyploid formation may in part be explained by its widespread ancestral occurrence in North America where it likely occurred sympatrically with both eastern and western diploid species of *Botrychium*. Among the allotetraploids, our results

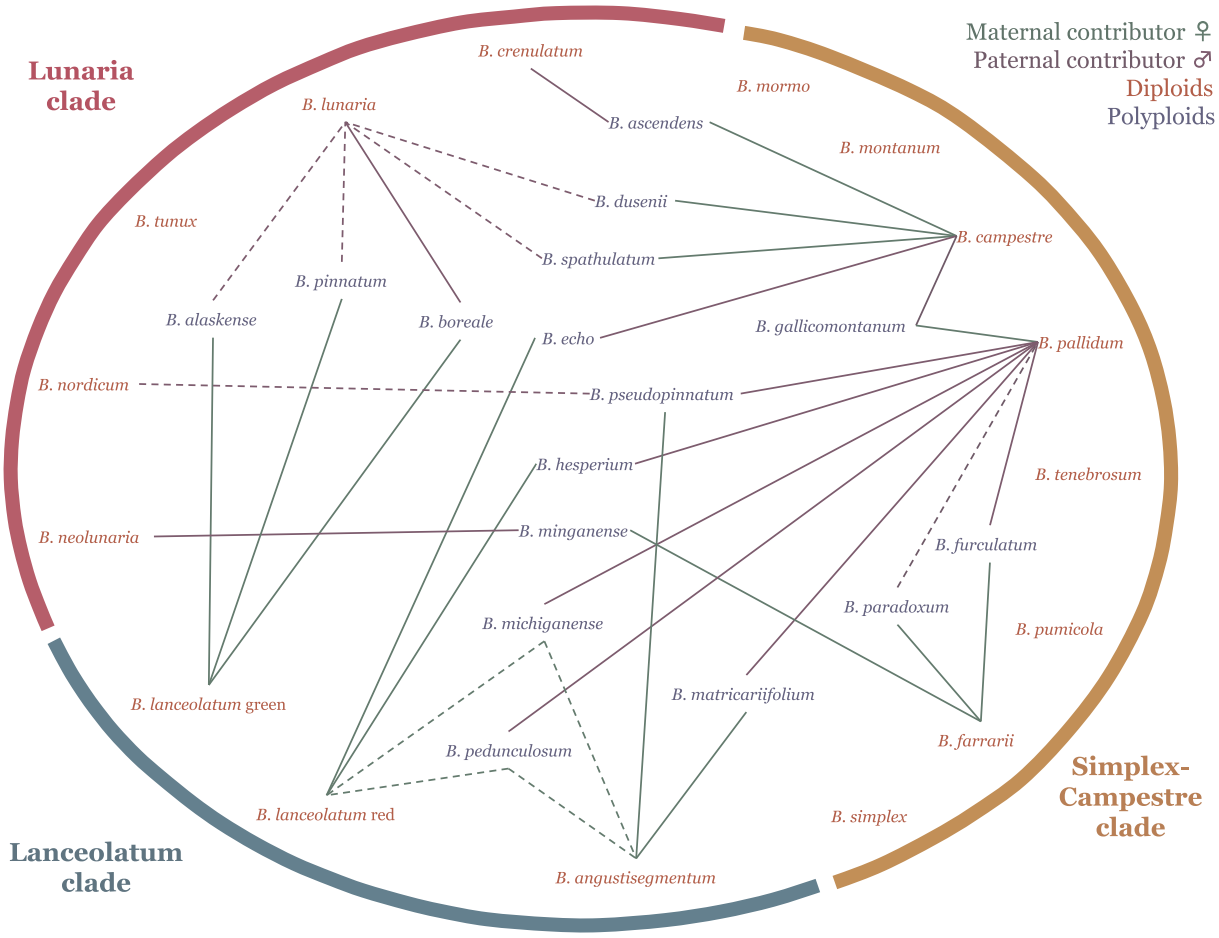

Fig. 4. Reticulation diagram, showing the origins of the sampled tetraploids and the single hexaploid (*B. pseudopinnatum*). Maternal contributors of polyploids are shown as green lines and their paternal contributors as violet lines. Uncertainties of parental progenitors—cases where our results either conflict with earlier hypotheses (Farrar, 2011; Gilman et al., 2015; Meza-Torres et al., 2017), or are incongruent between Figs. 1 and 3, or even could be erroneous due to unsampled diploids (Stensvold and Farrar, 2017)—are represented with dashed lines. Diploid progenitors are represented under the species rank.

support the parentage hypotheses of *B. ascendens* (Zika and Farrar, 2009), *B. matricariifolium* (Williams et al., 2016), *B. michiganense* (Gilman et al., 2015), and the majority of the remaining unpublished taxa, although some discrepancies (symbolized as dashed lines in Fig. 4) between our results and allozyme-based genetic profiles (Farrar, 2011) exist, as described below.

Several allopolyploidy hypotheses presented remain tentative (Fig. 4), pending inclusion of putative new diploid taxa that may prove to be more probable diploid parents than those currently indicated. Specifically, unrepresented in our study is a diploid North American "*B. lunaria* var. *lunaria*" that appears to be distinct from European *B. lunaria* (Dauphin et al., 2017; Maccagni et al., 2017; Stensvold and Farrar, 2017). Consequently, and in consideration of the currently restricted occurrence of *B. nordicum* to Iceland and Norway, the *B. nordicum* parentage of the North American *B. pinnatum*, *B. pseudopinnatum, B. spathulatum,* and the South American *B. dusenii,* is questionable. Additionally, a putative new diploid form of *B. paradoxum* that may prove parental to allotetraploid *B. paradoxum* has recently been detected. Parentages in question due to non-inclusion of these candidate diploid taxa are indicated (dashed lines in Fig. 4) and further study with key diploid additions can be expected to further clarify these ambiguous parentages of *Botrychium* allopolyploids.

4.5. Predominance of allopolyploidy

In general, an allopolyploid taxon is easier to identify in the field than an autopolyploid that might be expected to have no or few morphological characters distinct from its parent (Ramsey and Schemske, 1998). Thus, although we cannot refute the presence of undetected autopolyploidy, our nuclear data demonstrate that all our sampled polyploids are allopolyploids, usually with parents from different major clades (Lanceolatum, Lunaria, and Simplex-Campestre; Fig. 3). *Botrychium* species are known to maintain gametophytic selfing as the predominant mode of reproduction (Hauk and Haufler, 1999; Stensvold and Farrar, 2017), which produces strict homozygotes in a single generation, and drastically decreases the allele diversity at the individual level (Haufler et al., 2016; Sessa et al., 2016). With fixed heterozygosity, the allopolyploids may have an evolutionary advantage relative to their diploid progenitors, including greater resilience under ecological changes (Hegarty and Hiscock, 2007; Abbott et al., 2013).

4.6. Recent rapid radiation in an ancient lineage

Probably the major finding of our study is the discovery of the rapid and recent radiation of allopolyploid taxa in *Botrychium*. Our divergence time estimates provide the first chronology of hybridization events within *Botrychium*, in which all presently known allopolyploids have arisen in the last two million years (Fig. 2). This timescale could be even shorter, given the probability, with greater sampling, of detecting diploids that may be more closely related to allopolyploids than those inferred in our phylogenetic analysis (e.g., the paternal donor of *B. dusenii* and *B. spathulatum* in Fig. 3). In the more conservative case, with the oldest ages considered (Fig. 2), *Botrychium* allopolyploids are young in the context of Ophioglossaceae, which have been diverging for

~160 my from the other fern lineages (Pryer et al., 2004; Rothfels et al., 2015). Comprised of 50% polyploids, speciation via allopolyploidy plays an important role in the genus, as similarly reported in many other fern lineages (*Asplenium*, Perrie and Brownsey, 2005; *Pteris*, Chao et al., 2012; *Dryopteris*, Sessa et al., 2012; *Cystopteris*, Rothfels et al., 2014).

4.7. Parental biases in the formation of allopolyploids

The allopolyploids in our dataset are not randomly formed from the pool of diploid donors—instead some diploids are over-represented in the formation of polyploids. Among the diploid progenitors of polyploids, there is an additional tendency for individual species to be either the maternal or paternal progenitor, but not both. A striking example is the three diploids of the LAN clade, which have contributed to fourteen allopolyploid taxa, but always as the maternal donor (Dauphin et al., 2017). In contrast, the prolific *B. pallidum* has generated nine allopolyploids as a paternal donor and only one as the maternal contributor. Similar patterns of non-random parental contributions have been inferred in *Asplenium* (Perrie et al., 2010), and *Dryopteris* (Sessa et al., 2012; Testo et al., 2015), whereas more reciprocal patterns (individual diploids functioning as both the maternal and paternal parents of polyploids) were reported for the polyploid *Astrolepis integerrima* (Beck et al., 2012) and *Polypodium hesperium* (Sigel et al., 2014). Reproductive traits such as sperm motility and antheridiogen pheromones may be involved in mediating the hybridization between underground bisexual gametophytes, leading to non-random polyploid formation (Tanaka et al., 2014; Testo et al., 2015).

5. Conclusion

PacBio sequencing technology in conjunction with the use of the PURC bioinformatic pipeline provides new insight into phylogenetics of *Botrychium* and the major role played by allopolyploidy in the diversification of this genus. Our study presents a timescale of inter-specific hybridization events and recent and rapid radiation of allopolyploids in the last two million years. Our findings also confirm a predominance of allopolyploidy, with no evidence of autopolyploidy in our accessions. Distinctive features of allopolyploid formation in *Botrychium* include strong biases for individual diploid species to be either the maternal or paternal contributors to polyploids, a tendency for parental diploids to be from different major clades, and unequal participation among species and clades, with particular species in certain clades parenting most allopolyploids.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the Swiss National Science Foundation (grant: 31003A_156456) and its associated mobility grant, and by the University of California Berkeley. The authors thank Fay-Wei Li, Alessio Maccaggni, Kathleen Pryer, and Eric Schuettpelz for sharing herbarium specimens. We also thank Ingrid Jordon-Thaden and Lydia Smith for their assistance with laboratory work, and the staff at Duke University's Sequencing and Genomic Technologies Core Facility. We thank three anonymous reviewers for their helpful comments and constructive criticism.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2017.11.025.

References

Abbott, R.J., 2017. Plant speciation across environmental gradients and the occurrence and nature of hybrid zones. J. Syst. Evol. 55, 238–258.
Abbott, R.J., Albach, D., Ansell, S., Arntzen, J.W., Baird, S.J., Bierne, N., et al., 2013. Hybridization and speciation. J. Evol. Biol. 26, 229–246.
Beck, J.B., Allison, J.R., Pryer, K.M., Windham, M.D., 2012. Identifying multiple origins of polyploid taxa: a multilocus study of the hybrid cloak fern (*Astrolepis integerrima*; Pteridaceae). Am. J. Bot. 99, 1857–1865.
Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537.
Bozukov, V., Tsenov, B., Vatsev, M., 2010. A first find of *Botrychium* (Ophioglossaceae) in Bulgarian palaeomacroflora. C. R. Acad. Bulg. Sci. 63, 889–892.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 1–9.
Chao, Y.S., Dong, S.Y., Chiang, Y.C., Liu, H.Y., Chiou, W.L., 2012. Extreme multiple reticulate origins of the *Pteris cadieri* complex (Pteridaceae). Int. J. Mol. Sci. 13, 4523–4544.
Clausen, R.T., 1938. A monograph of the Ophioglossaceae. Mem. Torrey Bot. Club 19, 1–177.
Darriba, D., Taboada, G.L., Doallo, Posada, R.D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 772–772.
Dauphin, B., Vieu, J., Grant, J.R., 2014. Molecular phylogenetics supports widespread cryptic species in moonworts (*Botrychiu*m s.s., Ophioglossaceae). Am. J. Bot. 101, 128–140.
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). Syst. Bot. 42. http://dx.doi.org/10.1600/036364417X696483.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 1–8.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461.
Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., Knight, R., 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27, 2194–2200.
Ellis, B.K., 2014. Creating Microsatellite Primers for Botrychium. Dissertation, Lake Forest College, Lake Forest, Illinois, U.S.A.
Farrar, D.R., 2011 (last updated 15 Dec. 2011). Systematics and taxonomy of the genus Botrychium. < http://www.public.iastate.edu/~herbarium/botrychium/Moonwort-Systematics.pdf > (accessed 01 May 2017).
Farrar, D.R., Gilman, A.V., 2017. Relationships in the *Botrychium campestre* (Ophioglossaceae) complex. Brittonia 69, 265–275.
Gastony, G.J., Yatskievych, G., 1992. Maternal inheritance of the chloroplast and mitochondrial genomes in cheilanthoid ferns. Am. J. Bot. 79, 716–722.
Gilman, A.V., Farrar, D.R., Zika, P.F., 2015. *Botrychium michiganense* sp. nov. (Ophioglossaceae), a new North American moonwort. J. Bot. Res. Inst. Texas 9, 295–309.
Grusz, A.L., Windham, M.D., Pryer, K.M., 2009. Deciphering the origins of apomictic polyploids in the *Cheilanthes yavapensis* complex (Pteridaceae). Am. J. Bot. 96, 1636–1645.
Guillon, J.M., Raquin, C., 2000. Maternal inheritance of chloroplasts in the horsetail *Equisetum variegatum* (Schleich.). Curr. Genet. 37, 53–56.
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, 928–937.
Hauk, W.D., Haufler, C.H., 1999. Isozyme variability among cryptic species of *Botrychium* subgenus *Botrychium* (Ophioglossaceae). Am. J. Bot. 86, 614–633.
Hauk, W.D., Kennedy, L., Hawke, H.M., 2012. A phylogenetic investigation of *Botrychium* s.s. (Ophioglossaceae): Evidence from three plastid DNA sequence datasets. Syst. Bot. 37, 320–330.
Hauk, W.D., Parks, C.R., Chase, M.W., 2003. Phylogenetic studies of Ophioglossaceae: evidence from *rbcL* and *trnL-F* plastid DNA sequences and morphology. Mol. Phylogenet. Evol. 28, 131–151.
Hegarty, M., Hiscock, S., 2007. Polyploidy: doubling up for evolutionary success. Curr. Biol. 17, R927–R929.
Jones, G., 2014. Manual for using AlloppNET for diploids and allotetraploids. MANUAL, 1–10. < https://sites.google.com/site/touchingthedata/software/allopolyploids > (accessed 16.03.01).
Jones, G., Sagitov, S., Oxelman, B., 2013. Statistical inference of allopolyploid species networks in the presence of incomplete lineage sorting. Syst. Biol. 62, 467–478.
Koressaar, T., Remm, M., 2007. Enhancements and modifications of primer design program Primer3. Bioinformatics 23, 1289–1291.
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., Mentjies, P., Drummond, A.J., 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28, 1647–1649.
Larsson, A., 2014. AliView: a fast and lightweight alignment viewer and editor for large datasets. Bioinformatics 30, 3276–3278.
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 Bot. 127, 185–196.

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17, 1–10.
Meza-Torres, E.I., Stensvold, M.C., Farrar, D.R., Ferrucci, M.S., 2017. Circumscription of the South American moonwort *Botrychium* (Ophioglossaceae). Plant Biosyst. 151, 158–168.
Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE). IEEE, pp. 1–8.
Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401–437.
Pacific Biosciences, 2015. Revolutionize genomics with SMRT Sequencing, pp. 1–9. < http://pacificbiosciences.com/brochure > (accessed 16.03.01).
Pacific Biosciences, 2016. Procedure checklist – amplicon template preparation and sequencing. < http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Amplicon-Template-Preparation-and-Sequencing.pdf > (accessed 16.03.01).
Perrie, L.R., Brownsey, P.J., 2005. Insights into the biogeography and polyploid evolution of New Zealand *Asplenium* from chloroplast DNA sequence data. Am. Fern J. 95, 1–21.
Perrie, L.R., Shepherd, L.D., De Lange, P.J., Brownsey, P.J., 2010. Parallel polyploid speciation: distinct sympatric gene-pools of recurrently derived allo-octoploid *Asplenium* ferns. Mol. Ecol. 19, 2916–2932.
Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793–808.
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. Am. J. Bot. 91, 1582–1598.
Pteridophyte Phylogeny Group, 2016. A community-derived classification for extant lycopods and ferns. J. Syst. Evol. 54, 563–603.
R Development Core Team, 2015 (last updated 15.01.10). R: a language and environment for statistical computing. Version 3.2. R Foundation for Statistical Computing. Vienna. < https://www.rproject.org > (accessed 16.01.01).
Ramsey, J., Schemske, D.W., 1998. Pathways, mechanisms, and rates of polyploid formation in flowering plants. Annu. Rev. Ecol. Syst. 29, 467–501.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Rothfels, C.J., Johnson, A.K., Windham, M.D., Pryer, K.M., 2014. Low-copy nuclear data confirm rampant allopolyploidy in the Cystopteridaceae (Polypodiales). Taxon 63, 1026–1036.
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.K., Korall, P., 2013. Transcriptome-mining for single-copy nuclear markers in ferns. PLoS One 8, 1–18.
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., 2015. The evolutionary history of ferns inferred from 25 low-copy nuclear genes. Am. J. Bot. 102, 1089–1107.
Rothfels, C.J., Otto, S.P., 2016. Polyploid speciation. In: Kliman, R.M. (Ed.), Encyclopedia of Evolutionary Biology. Academic Press, Oxford, pp. 317–326.
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. New Phytol. 213, 413–429.
Rothwell, G.W., Stockey, R.A., 1989. Fossil Ophioglossales in the Paleocene of Western North America. Am. J. Bot. 76, 637–644.
Sessa, E.B., Testo, W.L., Watkins, J.E., 2016. On the widespread capacity for, and functional significance of, extreme inbreeding in ferns. New Phytol. 211, 1108–1119.
Sessa, E.B., Zimmer, E.A., Givnish, T.J., 2012. Reticulate evolution on a global scale: A nuclear phylogeny for New World *Dryopteris* (Dryopteridaceae). Mol. Phylogenet. Evol. 64, 563–581.
Sigel, E.M., Windham, M.D., Pryer, K.M., 2014. Evidence for reciprocal origins in *Polypodium hesperium* (Polypodiaceae): a fern model system for investigating how multiple origins shape allopolyploid genomes. Am. J. Bot 101, 1476–1485.
Soltis, P.S., Soltis, D.E., 2009. The role of hybridization in plant speciation. Annu. Rev. Plant Biol. 60, 561–588.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.
Stebbins, G.L., 1950. Variation Evolution in Plants. Columbia Univ. Press, New York.
Stensvold, M.C., 2008. A taxonomic phylogeographic study of the Botrychium lunaria complex. Dissertation, Iowa State University, Ames, Iowa, U.S.A.
Stensvold, M.C., Farrar, D.R., 2017. Genetic diversity in the worldwide *Botrychium lunaria* (Ophioglossaceae) complex, with new species and new combinations. Brittonia 69, 148–175.
Stensvold, M.C., Farrar, D.R., Johnson-Groh, C., 2002. Two new species of moonworts *(Botrychium* subg. *Botrychium*) from Alaska. Am. Fern J. 92, 150–160.
Tanaka, J., Yano, K., Aya, K., Hirano, K., Takehara, S., Koketsu, E., et al., 2014. Antheridiogen determines sex in ferns via a spatiotemporally split gibberellin synthesis pathway. Science 346, 469–473.
Testo, W.L., Watkins, J.E., Barrington, D.S., 2015. Dynamics of asymmetrical hybridization in North American wood ferns: reconciling patterns of inheritance with gametophyte reproductive biology. New Phytol. 206, 785–795.
Travers, K.J., Chin, C.S., Rank, D.R., Eid, J.S., Turner, S.W., 2010. A flexible and efficient template format for circular consensus sequencing and SNP detection. Nucleic Acids Res. 38, 1–8.
Vogel, J.C., Russell, S.J., Rumsey, F.J., Barrett, J.A., Gibby, M., 1998. Evidence for the maternal transmission of chloroplast DNA in the genus *Asplenium* (Aspleniaceae, Pteridophyta). Bot. Acta 111, 247–249.
Wagner, F.S., 1993. Chromosomes of North American grapeferns and moonworts (Ophioglossaceae: *Botrychium*). Contrib. Univ. Mich. Herb. 19, 83–92.
Wagner, W.H., 1980. A probable new hybrid grapefern *Botryhium matricariifolium* x *simplex*, from Central Michigan. Michigan Botanist 19, 31–36.
Wagner, W.H., 1991. New examples of the moonwort hybrid, *Botrychium matricariifolium* x *simplex* (Ophioglossaceae). Can. Field Nat. 105, 91–94.
Wagner, W.H., Beitel, J.M., 1985. Evidence for interspecific hybridization in pteridophytes with subterranean mycoparasitic gametophytes. P. Roy. Soc. Edinb. B 86, 273–281.
Wagner, W.H., Lord, L.P., 1956. The morphological and cytological distinctness of *Botrychium minganense* and *B. lunaria* in Michigan. B. Torrey Bot. Club 83, 261–280.
Wagner, W.H., Wagner, F.S., 1981. New species of moonworts, *Botrychium* subg. *Botrychium* (Ophioglossaceae), from North America. Am. Fern J. 71, 20–30.
Wagner, W.H., Wagner, F.S., 1983. Two moonworts of the Rocky Mountains; *Botrychium hesperium* and a new species formerly confused with it. Am. Fern J. 73, 53–62.
Wagner, W.H., Wagner, F.S., 1986. Three new species of moonworts *(Botrychium* subg. *Botrychium*) endemic in western North America. Am. Fern J. 76, 33–47.
Wagner, W.H., Wagner, F.S., 1988. Detecting *Botrychium* hybrids in the Lake Superior region. Michigan Bot. 27, 75–80.
Wagner, W.H., Wagner, F.S., 1990. Notes on the fan-leaflet group of moonworts in North America with descriptions of two new members. Am. Fern J. 80, 73–81.
Wagner, W.H., Wagner, F.S., Beital, J.M., 1985. Evidence for interspecific hybridization in pteridophytes with subterranean mycoparasitic gametophytes. Proc. Roy. Soc. Edin. 86B, 273–281.
Williams, E.W., Waller, D.M., 2012. Phylogenetic placement of species within the genus *Botrychium* s.s. (Ophioglossaceae) on the basis of plastid sequences, amplified fragment length polymorphisms, and flow cytometry. Int. J. Plant Sci. 173, 516–531.
Williams, E.W., Farrar, D.R., Henson, D., 2016. Cryptic speciation in allotetraploids: lessons from the *Botrychium matricariifolium* complex. Am. J. Bot. 103, 740–753.
Winther, J.L., Friedman, W.E., 2007. Arbuscular mycorrhizal symbionts in *Botrychium* (Ophioglossaceae). Am. J. Bot. 94, 1248–1255.
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. P. Natl. Acad. Sci. USA 106, 13875–13879.
Zika, P.F., Farrar, D.R., 2009. *Botrychium ascendens* WH Wagner (Ophioglossaceae) in Newfoundland and notes on its origin. Am. Fern J. 99, 249–259.