Molecular Phylogenetics and Evolution 96 (2016) 195–199

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Infraspecific variation within and across complete organellar genomes and nuclear ribosomal repeats in a moss ☆

Lily R. Lewis[a,c,*], Yang Liu[a], Ricardo Rozzi[b,c], Bernard Goffinet[a]

[a] Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., Storrs, CT 06269, USA
[b] Department of Philosophy, University of North Texas, 1704 West Mulberry, Denton, TX 76201, USA
[c] Omora Ethnobotanical Park, Institute of Ecology and Biodiversity, and Universidad de Magallanes, Puerto Williams, Antarctic Province, Chile

ARTICLE INFO

Article history:
Received 21 July 2015
Revised 3 December 2015
Accepted 9 December 2015
Available online 24 December 2015

Keywords:
Genomics
Infraspecific polymorphism
Early land plant
Bryophyte
Tetraplodon

ABSTRACT

Bryophytes (mosses, liverworts, and hornworts) are diverse and ecologically and evolutionarily significant yet genome scale data sets and analyses remain extremely sparse relative to other groups of plants, and are completely lacking at the infraspecific level. By sequencing the complete organellar genomes and nuclear ribosomal repeat from seven patches of a South American sub-Antarctic neo-endemic non-model moss, we present the first characterization of infraspecific polymorphism within and across the three genomic compartments for a bryophyte. Diversity within patches is accounted for by both intraindividual and interindividual variation for the nuclear ribosomal repeat and plastid genome, respectively. This represents the most extensive infraspecific genomic dataset generated for an early land plant lineage thus far and provides insight into relative rates of substitution between organellar genomes, including high rates of nonsynonymous to synonymous substitutions.

1. Introduction

Early land plants, collectively known as bryophytes (including mosses, liverworts, and hornworts) comprise approximately 20,000 species, occur on all continents and in all biomes (Vanderpoorten and Goffinet, 2009) and compose a lineage whose origin predates that of vascular plants (Clarke et al., 2011; Wickett et al., 2014). Despite their diversity, and their ecological and evolutionary significance, genomic resources for bryophytes are extremely limited relative to other groups of plants such as angiosperms (Wu et al., 2015). Methodological advances have simplified the sequencing of complete organellar genomes and the nuclear ribosomal repeat of non-model organisms (Liu et al., 2013), but the model organism *Physcomitrella patens* remains the only bryophyte species for which all three complete and annotated sequences have been reported (Sugiura et al., 2003; Terasawa et al., 2007; Rensing et al., 2008). Genome scale comparisons among bryophytes thus also remain extremely scarce (e.g., Liu et al., 2013, 2014; Sawicki et al., 2015). No studies to date have explored infraspecific genomic variation and relative rates of substitution across all three genomic compartments in a bryophyte based on complete organellar and nuclear ribosomal repeat sequences.

We sequenced the complete organellar genomes and nuclear ribosomal repeat for seven patches of the South American sub-Antarctic neo-endemic dung moss *Tetraplodon fuegianus* Besch. This monoecious species originated from a single amphitropical dispersal event during the Miocene to early Pliocene (8.63 Ma [95% highest posterior density 3.07–10.11 Ma]; Lewis et al., 2014), and forms spatially discrete patches on dung or carrion. Unlike most mosses, which disperse their spores by wind, dung mosses recruit flies to carry spores to their specific substrates. The flies are attracted to the moss through both visual and chemical cues that mimic dung and or carrion and inadvertently pick-up the sticky spores, which may fall off when flies land on the natural substrate. This fly mediated spore dispersal syndrome provides an efficient means of local to regional dispersal (Koponen, 1990; Marino et al., 2009).

In order to obtain sufficient DNA for next generation sequencing, multiple stems from the same patch of moss were pooled into a single DNA extract. Thus, each DNA extract represents a single patch, composed of an unknown number of individuals, as spores yield multiple individual stems and discrete patches may arise from multiple spores (Mägdefrau, 1982). *Tetraplodon fuegianus* is the second bryophyte, after the model species *Physcomitrella patens*, for which the complete chloroplast (cp) and MT genomes and nuclear ribosomal repeat (nrr) have been sequenced

☆ This paper was edited by the Associate Editor Elizabeth Zimmer.

* Corresponding author at: Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Rd., Storrs, CT 06269, USA.

E-mail addresses: LilyRLewis@gmail.com (L.R. Lewis), yang.liu@uconn.edu (Y. Liu), ricardo.rozzi@unt.edu (R. Rozzi), bernard.goffinet@uconn.edu (B. Goffinet).

http://dx.doi.org/10.1016/j.ympev.2015.12.005
1055-7903/

assembled and annotated, and the first to have all three sequences reported at once.

2. Materials and methods

DNA was extracted from eight patches of *T. fuegianus* collected in the Antarctic and Magallanes provinces in the Magallanes region of Chile (Supplementary Table S1). Multiple gametophytes and/or sporophytes from each patch of moss were pooled to yield sufficient DNA for TruSeq library preparation and sequenced on the Illumina HiSeq2000 platform (samples 1–7) following Liu et al. (2014) or the 454 platform (sample 8) following Liu et al. (2013).

Filtered paired-end reads were *de novo* assembled in CLC genomics workbench 6.5, with quality scores retained and reads mapped back to contigs (80% similarity across at least 50% of reads, mismatch cost of 2, insertion and deletion costs of 3). *De novo* contigs were blasted (blastn – somewhat similar setting) against a custom database including the cp from *Syntrichia ruralis* (NC_012052) and *Physcomitrella patens* (NC_005087), mt from *Anomodon attenuatus* (NC_021931) and *Physcomitrella patens* (NC_007945), and the complete *Funaria hygrometrica* nrr sequence (X80212). Blast identified mt and nrr contigs for all patches were combined and *de novo* assembled in Geneious 7.0.4 and aligned to the *Anomodon* mt and *Funaria* nrr sequences, respectively. *De novo* assemblies for each sample did not generate significant cp genome coverage, so *de novo* assembly was accomplished by pooling reads across 4 samples (sample numbers 1, 2, 7, and 8). Contigs were blast identified as above using the *Physcomitrella patens* and *Syntrichia ruralis* cp genomes and *de novo* assembled in Geneious 7.0.4. Junctions between the large single copy (LSC), inverted repeats A and B (IRA; IRB), and small single copy (SSC) were sequenced on an ABI3100 Genetic Analyzer (Applied Biosystems, Grand Island, NY, USA). The consensus pre-draft mt, nrr and cp were annotated based on the references, and manually edited using ExPASy translate tool for coding regions (Gasteiger et al., 2003) and tRNAscan SE 1.21 webserver for tRNAs (Lowe and Eddy, 1997). Sample 8, sequenced on the 454 platform, was not used in further analyses due to poor coverage.

A reiterative reads mapping process (similarity fraction of 0.95 over at least 0.90 of the read length with equal mismatch, insertion and deletion costs) was used for each sample to create a draft consensus from pre-drafts, and then a final read mapping to the draft was screened for variants. A final consensus was generated with ambiguity codes denoting variant sites with a 100× minimum depth of quality filtered reads. Variants with a frequency lower than 0.30 were filtered out as noise and only those represented by a minimum nucleotide count of 30 were retained. Final consensus sequences were aligned in Geneious 7.0.4, using the progressive Mauve algorithm (Darling et al., 2004; SI Alignment). In order to confirm select variants, distinguishing between sequencing error (Wu et al., 2015) versus interindividual and intraindividual variation (i.e., paralogy), DNA was re-extracted from individual gametophyte stem tips sampled randomly across a patch. Sanger sequencing following Lewis et al. (2014) was used to sequence selected variants. Variant positions were checked for double chromatogram peaks and variation across individuals from the same patch of moss.

3. Results and discussion

3.1. Structure of organellar genomes and nuclear ribosomal repeat is conserved

The cp genome of *Tetraplodon fuegianus* has a quadripartite architecture and is 123,670–123,672 bp long (Table 1), with length variation due to an AT indel in the *ycf*4-*psa*L intergenic region (Table 2). The gene order and content is conserved relative to *Syntrichia ruralis* (Oliver et al., 2010) and *Tetraphis pellucida* (Bell et al., 2014). The GC content of 28.7% is similar to that of *Physcomitrella patens* (28.5%; Sugiura et al., 2003) and *Tetraphis pellucida* (29.4%; Bell et al., 2014), as well as the liverwort *Marchantia polymorpha* (28.8%, Ohyama et al., 1986), but lower than that reported for the liverwort *Ptilidium* (33.2%; Forrest et al., 2011) and the hornworts (32.9%; *Anthoceros formosae*; Kugita, 2003; 35% *Nothoceros aenigmaticus*; Villarreal et al., 2013). The mt genome is 104,741 bp long, with a conserved gene order relative to all other surveyed mosses spanning the phylogenetic history of mosses (Liu et al., 2014). The mt of *T. fuegianus* provides additional support, from a previously un-sampled moss family, for the structurally static nature of the moss mt relative to the highly labile angiosperm mt (Liu et al., 2014). The nrr varies in length between 10,394 and 10,398 bp due to indels in the IGS2, and comprises four ribosomal RNA genes arranged in the characteristic L-type organization of mosses and streptophyte algae, i.e. with the 18S, 5.8S, 26S and 5S genes arranged in tandem repeats (Wicke et al., 2011).

3.2. Infraspecific polymorphism in organellar genomes and nuclear ribosomal repeat

Screening the complete organellar genomes and the nrr within and across patches of *Tetraplodon fuegianus* revealed infraspecific polymorphisms in each of the genomic compartments (Fig. 1; Table 2). The cp genome comprised 16 polymorphic sites (i.e., 0.013%) scattered across the chromosome. Five single nucleotide polymorphisms, including two transitions and two transversions were located in coding sequences, resulting in four nonsynonymous substitutions and one synonymous substitution. A single polymorphic site (i.e., a C/T transition) was recovered within the mt genome within patches 2 and 3, resulting in an amino acid change in the second exon of the *nad*1 gene (position 72,549). Twenty-one polymorphisms were detected in the nrr, representing 0.21% of the sequence.

To distinguish intra- from interindividual variation in the cp and nrr, select polymorphic sites were resequenced from DNA extracts representing single gametophyte stems rather than pooled stems from a given patch. Sanger sequencing in forward and reverse directions of the A/G variant in the *rps*12-*trn*V (GAC) intergenic region (aligned position 84,936 supplementary alignment B; Table 2) of the chloroplast across eight individual stems resampled from a single patch (sample 2) recovered the A state in two and the G state in six individual stems (Fig. 1). Confirmation of this polymorphic site using individual stem DNA extracts rules out sequencing error (Wu et al., 2015) and heteroplasmy (Wolfe and Randle, 2004), providing strong support for interindividual cp variation within discrete patches of *T. fuegianus*. Sequencing of the A/G polymorphism in the ITS2 (nrr) for 2 individual gametophyte stems from two distinct patches (samples 1 and 2) recovered signals for both A and G character states (i.e., double peaks) at the variant site for each individual stem (Fig. 1). Thus, the polymorphism recovered in the nrr is at least in part a result of intraindividual variation, as is the case for *Tortula muralis* for which it was suggested that weakly divergent repeat copies likely arose from mutation rates exceeding the pace of concerted evolution while highly divergent copies were a result of ancestral gene flow (Košnar et al., 2012).

3.3. Rates of polymorphism vary across genomic compartments

The relative rates of substitution between genomic compartments of *Tetraplodon fuegianus* are variable. The mt has the slowest substitution rate, with the cp being 13.7 times faster. Wolfe et al. (1987) inferred a cp rate of substitution 3 times faster than the

Table 1
Size, composition, and GC content of organellar genomes and nuclear ribosomal repeat (Genbank accession numbers in parentheses) for *Tetraplodon fuegianus*, including number of genes/total base pairs/percent of total size of each sequence for protein coding sequences and tRNAs.

	Chloroplast (KU095851)				Mitochondrial (KT373818)	NRR (KU095852)
	LSC	SSC	IR[a]	Total		
Size (bp)	84,946	18,692	10,016	123,670	104,741	10,401
Protein coding	68/53,833/63.4	14/15,336/82.1	0	82/65,169/52.7	40/32,143/30.7	0
tRNA	26/1987/2.3	2/154/0.8	5/363/3.62	33/2867/2.3	24/1801/1.7	0
rRNA	0	0	4/4522/45.1	4/9044/7.3	3/5232/5.0	4/5522/53.1
G/C (%)	26	25.3	43.3	28.7	40	54

Note: Abbreviations include: large single copy (LSC); small single copy (SSC); inverted repeat (IR); transfer RNA (tRNA); ribosomal RNA (rRNA); nuclear ribosomal repeat (nrr).
[a] Inverted repeat (IR) genes and sequence length are only counted once.

Table 2
Polymorphic positions in the organellar genomes and nuclear ribosomal repeat of *Tetraplodon fuegianus*, based on within and across patch screening.

Aligned position	Location	Length (bp)	Variant	Amino acid	Distribution within/between patches
Chloroplast genome 123,670–123,672 bp					
1784	*rpl*2 (LSC)	1	R (A or G) TS	Q	3
19,649	*psb*J-*pet*A (LSC)	1	W (A or T) TV	n/a	All
19,654		1	Y (C or T) TS	n/a	All
19,660		1	W (A or T) TV	n/a	1, 2
24,035	*ycf*4-*psa*L (LSC)	2	AT indel	n/a	4
46,280	*trn*T (GGU)-*trn*G(UUC) (LSC)	1	Y (C or T) TS	n/a	7
47,662	*ycf*2	1	G/T TV	H/N	Inter-patch
54,850	*psb*A-*trn*K(UUU) (LSC)	1	R (A or G) TS	n/a	3
55,639	*mat*K (LSC)	1	R (A or G) TS	G/S	6
56,686		1	Y (C or T) TS	H/Y	6
60,827	*psb*I-*trn*S(GCU) (LSC)	1	R (A or G) TS	n/a	1
61,851	*ycf*12-*trn*G(UCC) (LSC)	1	A/G TS	n/a	Inter-patch
69,665	*rpo*C2 (LSC)	1	A/T TV	N/I	Inter-patch
83,334	*ndh*B-*rps*7 (LSC)	1	W (A or T) TV	n/a	1, 6
84,936	*rps*12-*trn*V(GAC) (IRA)*	1	R (A or G) TS	n/a	2*, 3
108,033	*psa*C-*ndh*D (SSC)	4	WWWW (A or T) TV	n/a	1, 6
Mitochondrial genome 104,741 bp					
72,549	*nad*1 exon 2	1	Y (C or T) TS	P/L	3, 2
Nuclear ribosomal repeat 10,394–10,398 bp					
2289	ITS2	1	R (A or G) TS	n/a	1*, 2*, 5, 6, 7
6202	IGS1	1	R (A or G) TS	n/a	2, 3, 5, 6, 7
6408		1	M (A or C) TV	n/a	1, 2, 3, 5, 6, 7
6410		1	R (A or G) TS	n/a	1, 2, 3, 5, 6, 7
7602	IGS2	1	W (A or T) TV	n/a	6, 7
7604		1	T indel	n/a	Inter-patch
7606		2	AT indel	n/a	Inter-patch
7612		1	G indel	n/a	Inter-patch
7720		1	M (A or C) TV	n/a	All
7728		1	M (A or C) TV	n/a	All
7763		1	Y (C or T) TS	n/a	All
7770		1	R (A or G) TS	n/a	All
7865		1	Y (C or T) TS	n/a	All
7871		1	Y (C or T) TS	n/a	All
8101		1	M (A or C) TV	n/a	4, 5, 6
8120		1	Y (C or T) TS	n/a	1, 2, 3, 5, 6, 7
8174		1	K (G or T) TV	n/a	2, 4, 5, 6, 7
8181		1	R indel	n/a	4
8840		1	T indel	n/a	Inter-patch
9640		1	C indel	n/a	Inter-patch
9853		1	R (A or G) TS	n/a	4, 6, 7

Note: For multi-nucleotide variants, the starting position is listed. Positions and patches marked with an "*" were re-sequenced with Sanger sequencing. Abbreviations include: large single copy (LSC); small single copy (SSC); inverted repeat (IR); transition (TS); transversion (TV); not applicable (n/a).

mt based on synonymous substitutions from discrete organellar loci for a sampling of land plants. Screening of complete organellar genomes for intravarietal polymorphisms in rice also recovered faster rates of substitution in cp versus mt genomes (Zhang et al., 2012a,b). Not all land plants, however, share this pattern of relative organellar substitution rates. Extensive variation has been reported from the mt genome of the resurrection fern, *Boea hygrometrica*, relative to a complete absence of polymorphic sites across the cp genome (Zhang et al., 2012a,b). Similarly, much higher rates of polymorphism were recovered in the mt compared to the cp genomes for various cultivars of date palm (Yang et al., 2010; Fang et al., 2012). Mosses have notably conserved mt genomes in terms of gene order across the macroevolutionary tree of life (Liu et al., 2014). Here we provide evidence supporting the highly conserved nucleotide sequence of the mt genome at the microevolutionary scale. While it is tempting to say that the nrr has a substitution rate 16.2 times faster than the cp, evidence of intraindividual variation (i.e., polymorphic copies) in the nrr prevents us from distinguishing whether the polymorphism arose after divergence from ancestral populations, or whether the ancestor that was dispersed to

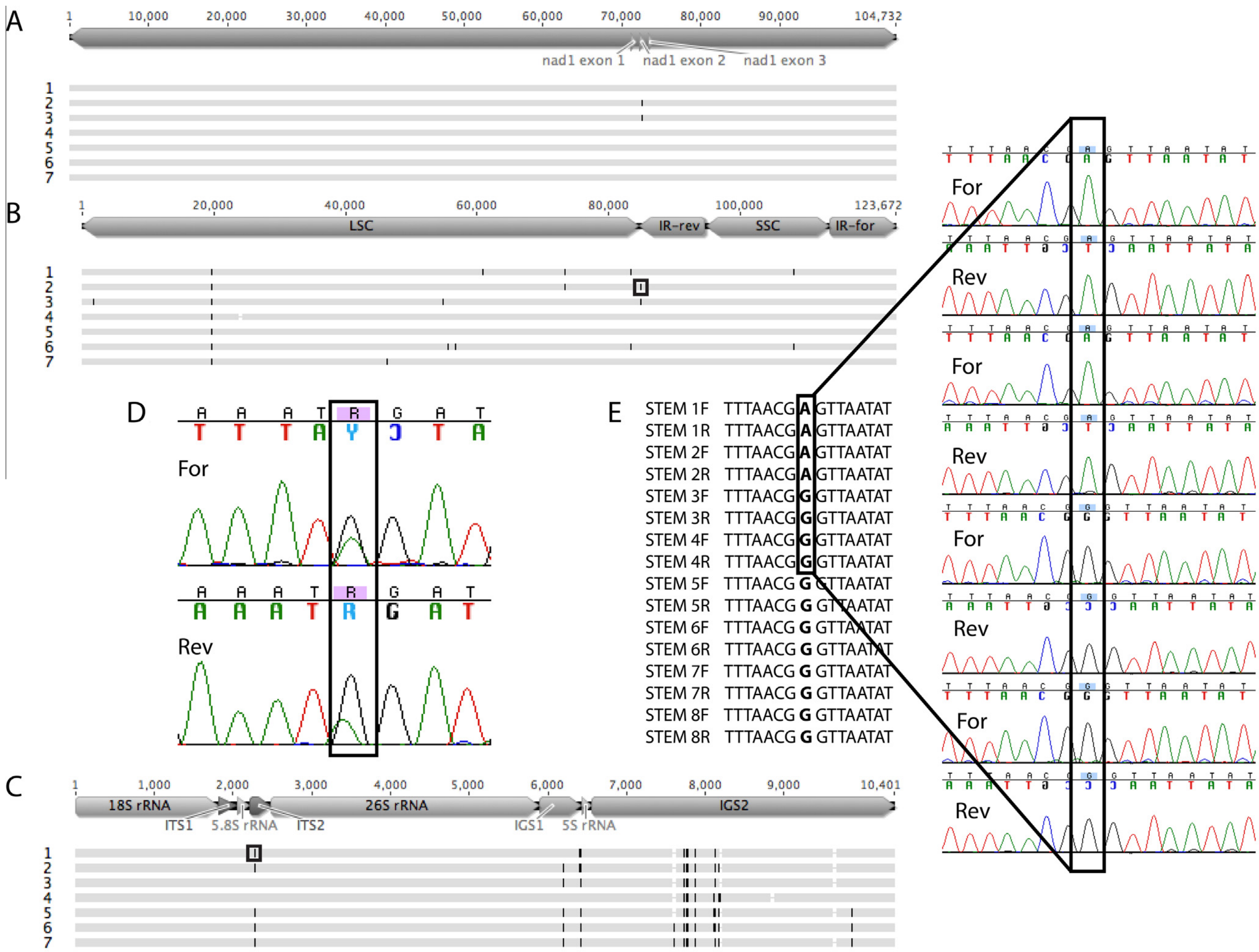

Fig. 1. Complete mitochondrial (A) and chloroplast (B) genomes and nuclear ribosomal repeat sequences (C) aligned for seven patches of *Tetraplodon fuegianus*. Polymorphic positions within patches (1–7) are denoted with a black mark for substitutions and a white space for indels. Sanger sequencing of select variants in the cp genome and nrr from single stem DNA extracts was done to differentiate between interindividual and intraindividual variation as the source of patch polymorphism. Re-sequenced variants are denoted with a box in the cp (B) and nrr (C) alignments. Sequencing results and chromatograms for both forward and reverse directions are shown for the nrr (D) and cp (E).

southernmost Chile approximately 8.63 Ma (Lewis et al., 2014) harbored polymorphic nrr copies.

3.4. High ratio of nonsynonymous to synonymous substitutions

The polymorphisms recovered in the organellar genomes of *Tetraplodon fuegianus* represent a high number of nonsynonymous to synonymous substitutions, with 4:1 ratio in the cp and 1:0 in the mt. Screening of the complete cp genome of the angiosperm *Arabis alpina* revealed a ratio of 24:13 nonsynonymous to synonymous infraspecific substitutions (Melodelima and Lobréaux, 2013), and 6:5 for the nearly complete cp genome of the moss *Syntrichia ruralis* (Oliver et al., 2010). In protein coding genes directly related to photosynthesis, only synonymous substitutions were detected for *A. alpina*, suggesting higher selective pressure on photosynthesis genes (Melodelima and Lobréaux, 2013). Two of the four nonsynonymous substitutions discovered in the cp of *T. fuegianus* are in the gene *mat*K, which is known to be highly variable at both the nucleotide and amino acid levels (Johnson and Soltis, 1994; Liu et al., 2010), despite evidence of its role in chloroplast function (Barthet and Hilu, 2007). A single nonsynonymous substitution was also found in the *mat*K of *S. ruralis* (Oliver et al., 2010). All other genes associated with photosynthesis lack polymorphisms in the *T. fuegianus* cp genome. The remaining two nonsynonymous substitutions in the cp of *T. fuegianus* occur in the unknown protein-coding gene *ycf*2 and RNA polymerase gene *rpo*C2. Two nonsynonymous substitutions have also been found in *ycf*2 and three in the *rpo*C2 of *A. alpina* (Melodelima and Lobréaux, 2013) and three in *ycf*2 of *S. ruralis* (Oliver et al., 2010). Whether the occurrence of nonsynonymous substitutions in these genes signifies relaxed or positive selective pressures warrants further study. If positive selection is at play, adaptation to local conditions (Magdy et al., in press) since colonization of the sub-Antarctic may account for the accumulation of nonsynonymous substitutions in the plastid genome of *T. fuegianus*.

4. Conclusion

The DNA sampling and sequencing approach used here facilitated the novel identification of infraspecific polymorphisms both within and across discrete patches of moss for all three genomic compartments, and thus the comparison of relative rates of polymorphisms between these compartments and the assessment of mutation types and frequencies across coding and non-coding regions. Genome scale sequencing is becoming increasingly cost effective, making it feasible to generate extensive data sets for non-model organisms. Sequencing across multiple infraspecific populations provides valuable insight into the frequency and patterns of polymorphism at a fine evolutionary scale, which facilitates identification of markers of interest to molecular evolutionary biologists.

Acknowledgments

This work was supported by the National Science Foundation (DEB-1311405 to L.R.L.; DEB-1212505 to B.G.). Fieldwork in the Cape Horn Archipelago, Chile, was made possible by Iniciativa Científica Milenio and Comisión Nacional de Investigación Científica y Tecnológica (ICM P05-002 and Basal-CONICYT PFB-23, respectively, to Institute of Ecology and Biodiversity, IEB-Chile, R.R.), and the National Science Foundation (DEB-0948380 to William R. Buck). For comments on earlier drafts we thank Rafael Medina, Jessica Budke, Paul Lewis, and Kent Holsinger.

Appendix A. Supplementary material

Supplementary alignments: Positions of polymorphisms listed in Table 2 and discussed in the text refer to aligned positions. A: Mitochondrial genomes aligned, B: Chloroplast genomes aligned, C: Nuclear ribosomal repeat sequences aligned. Supplementary Table S1 and alignments associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.12.005.

References

Barthet, M.M., Hilu, K.W., 2007. Expression of matK, functional and evolutionary implications. Am. J. Bot. 94, 1402–1412.

Bell, N.E., Boore, J.L., Mishler, B.D., Hyvönen, J., 2014. Organellar genomes of the four-toothed moss, *Tetraphis pellucida*. BMC Genomics 15, 383.

Clarke, J.T., Warnock, R.C.M., Donoghue, P.C.J., 2011. Establishing a time-scale for plant evolution. New Phytol. 192, 266–301.

Darling, A.C.E., Mau, B., Blattner, F.R., Perna, N.T., 2004. Multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 14, 1394–1403.

Fang, Y., Wu, H., Zhang, T., Yang, M., Yin, Y., Pan, L., 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. PLoS One 7, e37164.

Forrest, L.L., Wickett, N.J., Cox, C.J., Goffinet, B., 2011. Deep sequencing of *Ptilidium* (Ptilidiaceae) suggests evolutionary stasis in liverwort plastid genome structure. Plant Ecol. Evol. 144, 29–43.

Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., Bairoch, A., 2003. ExPASy: the proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. 31, 3784–3788.

Johnson, L.A., Soltis, D.E., 1994. *mat*K DNA sequences and phylogenetic reconstruction in Saxifragaceae s. str. Syst. Bot. 19, 143–156.

Koponen, A., 1990. Entomophily in the Splachnaceae. J. Linn. Soc., Bot. 104, 115–127.

Košnar, J., Herbstová, M., Kolář, F., Koutecký, P., Kučera, J., 2012. A case study of intragenomic ITS variation in bryophytes: assessment of gene flow and role of polyploidy in the origin of European taxa of the *Tortula muralis* (Musci: Pottiaceae) complex. Taxon 61, 709–720.

Kugita, M., 2003. The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucleic Acids Res. 31, 716–721.

Lewis, L.R., Rozzi, R., Goffinet, B., 2014. Direct long-distance dispersal shapes a New World amphitropical disjunction in the dispersal-limited dung moss *Tetraplodon* (Bryopsida: Splachnaceae). J. Biogeogr. 41, 2385–2395.

Liu, Y., Forrest, L.L., Bainard, J.D., Budke, J.M., Goffinet, B., 2013. Organellar genome, nuclear ribosomal DNA repeat unit, and microsatellites isolated from a small-scale of 454 GS FLX sequencing on two mosses. Mol. Phylogenet. Evol. 66, 1089–1094.

Liu, Y., Medina, R., Goffinet, B., 2014. 350 million years of mitochondrial genome stasis in mosses, an early land plant lineage. Mol. Biol. Evol. 31, 8–13.

Liu, Y., Yan, H.F., Cao, T., Ge, X.J., 2010. Evaluation of 10 plant barcodes in Bryophyta (Mosses). J. Syst. Evol. 48, 36–46.

Lowe, T.M., Eddy, S.R., 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25, 955–964.

Mägdefrau, K., 1982. Life-forms of Bryophytes. In: Smith, A.J.E. (Ed.), Bryophyte Ecology. Springer, Netherlands.

Magdy, M., Werner, O., McDaniel, S.F., Goffinet, B., Ros, R.M., 2015. Genomic scanning using AFLP to detect loci under selection in the moss *Funaria hygrometrica* along a climate gradient in the Sierra Nevada Mountains, Spain. Plant Biol. doi: http://dx.doi.org/10.1111/plb.12381 (in press).

Marino, P., Raguso, R., Goffinet, B., 2009. The ecology and evolution of fly-dispersed dung mosses (Splachnaceae): manipulating insect behavior through odour and visual cues. Symbiosis 47, 61–76.

Melodelima, C., Lobréaux, S., 2013. Complete *Arabis alpina* chloroplast genome sequence and insight into its polymorphism. Meta Gene 1, 65–75.

Ohyama, K., Fukuzawa, H., Kohchi, T., Shirai, H., Sano, T., Sano, S., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., et al., 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 332, 572–574.

Oliver, M.J., Murdock, A.G., Mishler, B.D., Kuehl, J.V., Boore, J.L., Mandoli, D.F., Everett, K.D.E., Wolf, P.G., Duffy, A.M., Karol, K.G., 2010. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. BMC Genomics 11, 143.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H., Nishiyama, T., Perroud, P.F., Lindquist, E.A., Kamisugi, Y., et al., 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69.

Sawicki, J., Szczecińska, M., Bednarek-Ochyra, H., Ochyra, R., 2015. Mitochondrial phylogenomics supports splitting the traditionally conceived genus *Racomitrium* (Bryophyta: Grimmiaceae). Nova Hedwigia 100, 293–371.

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., Sugita, M., 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpo*A from the chloroplast to the nucleus. Nucleic Acids Res. 31, 5324–5331.

Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., Sato, N., 2007. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol. Biol. Evol. 24, 699–709.

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

Villarreal, J.C., Forrest, L.L., Wickett, N., Goffinet, B., 2013. The plastid genome of the hornwort *Nothoceros aenigmaticus* (Dendrocerotaceae): phylogenetic signal in inverted repeat expansion, pseudogenization, and intron gain. Am. J. Bot. 100, 1–11.

Wicke, S., Costa, A., Muñoz, J., Quandt, D., 2011. Restless 5S: the re-arrangement(s) and evolution of the nuclear ribosomal DNA in land plants. Mol. Phylogenet. Evol. 61, 321–332.

Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al., 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. USA 111, E4859–E4868.

Wolfe, A.D., Randle, C.P., 2004. Recombination, heteroplasmy, haplotype polymorphism, and paralogy in plastid genes: implications for plant molecular systematics plastid genes. Syst. Botany 29, 1011–1020.

Wolfe, K.H., Li, W.H., Sharp, P.M., 1987. Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. Proc. Natl. Acad. Sci. USA 84, 9054–9058.

Wu, Z., Tembrock, L.R., Ge, S., 2015. Are differences in genomic data sets due to true biological variants or errors in genome assembly: an example from two chloroplast genomes. PLoS One 10, e0118019.

Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S., Yu, J., 2010. The complete chloroplast genome sequence of date palm (*Phoenix dactylifera* L.). PLoS One 5, e12762.

Zhang, T., Fang, Y., Wang, X., Deng, X., Zhang, X., Hu, S., Yu, J., 2012a. The complete chloroplast and mitochondrial genome sequences of *Boea hygrometrica*: insights into the evolution of plant organellar genomes. PLoS One 7, e30531.

Zhang, T., Hu, S., Zhang, G., Pan, L., Zhang, X., Al, I.S., 2012b. The organelle genomes of Hassawi rice (*Oryza sativa* L.) and its hybrid in Saudi Arabia: genome variation, rearrangement, and origins. PLoS One 7, e42041.