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.

© 2015 Elsevier Inc. All rights reserved.

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/© 2015 Elsevier Inc. All rights reserved.

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

Download English Version:

https://daneshyari.com/en/article/2833755

Download Persian Version:

https://daneshyari.com/article/2833755

Daneshyari.com