Molecular Phylogenetics and Evolution 66 (2013) 1089–1094

Contents lists available at SciVerse ScienceDirect

Molecular Phylogenetics and Evolution

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

Short Communication

Organellar genome, nuclear ribosomal DNA repeat unit, and microsatellites isolated from a small-scale of 454 GS FLX sequencing on two mosses

Yang Liu [a], Laura L. Forrest [a,b], Jillian D. Bainard [c], Jessica M. Budke [a], Bernard Goffinet [a,*]

[a] Department of Ecology and Evolutionary Biology, University of Connecticut, 75 North Eagleville Road, Storrs, CT 06269-3043, USA
[b] Royal Botanic Garden Edinburgh, 20a Inverleith Row, Edinburgh, UK
[c] Department of Integrative Biology, University of Guelph, Guelph, Ontario, N1G 2W1 Canada

ARTICLE INFO

Article history:
Received 3 July 2012
Revised 30 November 2012
Accepted 6 December 2012
Available online 20 December 2012

Keywords:
Entosthodon obtusus
Funaria hygrometrica
High-throughput sequencing
Microsatellite
Organellar genomes
nrDNA repeat unit

ABSTRACT

Recent innovations in high-throughput DNA sequencing methodology (next generation sequencing technologies [NGS]) allow for the generation of large amounts of high quality data that may be particularly critical for resolving ambiguous relationships such as those resulting from rapid radiations. Application of NGS technology to bryology is limited to assembling entire nuclear or organellar genomes of selected exemplars of major lineages (e.g., classes). Here we outline how organellar genomes and the entire nuclear ribosomal DNA repeat can be obtained from minimal amounts of moss tissue via small-scale 454 GS FLX sequencing. We sampled two Funariaceae species, *Funaria hygrometrica* and *Entosthodon obtusus*, and assembled nearly complete organellar genomes and the whole nuclear ribosomal DNA repeat unit (18S-ITS1-5.8S-ITS2-26S-IGS1-5S-IGS2) for both taxa. Sequence data from these species were compared to sequences from another Funariaceae species, *Physcomitrella patens*, revealing low overall degrees of divergence of the organellar genomes and nrDNA genes with substitutions spread rather evenly across their length, and high divergence within the external spacers of the nrDNA repeat. Furthermore, we detected numerous microsatellites among the 454 assemblies. This study demonstrates that NGS methodology can be applied to mosses to target large genomic regions and identify microsatellites.

1. Introduction

The moss family Funariaceae includes *Physcomitrella patens* (Hedw.) Bruch & Schimp. and *Funaria hygrometrica* Hedw., two model organisms for the study of hybridization, physiology, and genetics in mosses (Knight et al., 2009). The family comprises approximately 250–400 species distributed among three speciose genera (*Funaria* Hedw., *Entosthodon* Schwägr., and *Physcomitrium* [Brid.] Brid) and 13 monospecific or species poor genera (Crosby et al., 1999). Sequence variation within 10 loci sampled from all three genomic compartments provided robust signal for the earliest divergences within the family, and evidence for the polyphyly of *Entosthodon* and *Physcomitrium*, but was insufficient to resolve the relationship within the crown group, which accommodates an estimated 90% of the species, and is considered to have arisen from a rapid diversification (Liu et al., 2012a). At the DNA level, and assuming a conserved rate of substitution, a rapid radiation may be characterized by few nucleotide changes between two sister taxa at any given locus. Hence resolving the relationships among taxa emerging from a rapid succession of cladogenic events requires sequencing an extensive set of loci to uncover sufficient synapomorphic substitutions to identify and support unique common ancestries. Resolving, for example, the explosive radiation within the Saxifragales required 25–50 Kb of nucleotide data (Jian et al., 2008). In mosses, the problem of acquiring sufficient data for reconstructing a robust phylogeny from discrete loci may be further compounded by their significantly lower substitution rate (Stenøien, 2008) and high levels of homoplasy among more deeply rooted lineages. Hence broad genomic sampling of characters may be advantageous, if not critical, for inferring relationships among major lineages composing the backbone of the moss tree of life.

Random or "shotgun" sequencing based on a total genomic extract yields a population of DNA sequences that can be assembled within and compared among samples. However, the overlap in homologous loci between samples, or the likelihood of randomly sampling homologous loci twice, is a function of the concentration of the locus in the extract and the scale of the sequencing effort. Loci present in single copies in the nuclear genomes will be sampled at a much lower frequency than multiple copy nuclear loci and organellar loci. A full picotiter plate on the 454 platform may generate as much as 0.4 Gb of data, which for a nuclear genome size of 1 Gb would lead to an average sequence depth of 0.4× (i.e., each nucleotide in the genome would be sequenced 0.4 times on average). On average, individual organellar genome sizes are less than 0.1% of the nuclear genome size, but estimates suggest

* Corresponding author.
E-mail address: bernard.goffinet@uconn.edu (B. Goffinet).

1055-7903/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.ympev.2012.12.006

a single plant leaf cell, since it contains numerous plastids and mitochondria, also contains many copies of each organellar genome. In angiosperms, for example, a leaf cell may hold 1000–50,000 plastid (pt) genomes (Shaver et al., 2006), and several hundreds to thousands of copies of the nuclear ribosomal DNA repeat (nrDNA) (Rogers and Bendich, 1987). Using a random sequencing approach, single copy nuclear loci are thus less likely to be captured and recovering homologous loci between two independent runs is highly unlikely. Consequently most of the nuclear data obtained from NGS may not be comparable. The imbalance in concentrations between genomic compartments has, however, an advantage, as loci present in high copies are more likely sequenced.

Comparison of publically available records indicates that organellar genomes are conserved in size within mosses. The plastid genome is composed of approximately 123 Kb (Oliver et al., 2010) and the mitochondrial (mt) genome comprises ca. 105 Kb (Liu et al., 2012b). In addition, the nuclear ribosomal repeat consists of an 11 Kb nrDNA repeat unit (Wicke et al., 2011). Together these loci comprise a total of ~230,000 nucleotides that could be acquired through NGS and provide a vast and hence promising source of phylogenetic markers.

The potential of these multicopy regions for resolving phylogenetic relationships is increasingly recognized (Kane et al., 2012; Nock et al., 2011; Steele et al., 2012). The relative ease with which they are sequenced and assembled even led to a proposal to use entire plastid genomes, or a combination of plastid and nuclear ribosomal sequences, as a DNA barcode (Nock et al., 2011), or ultra-barcode (Kane et al., 2012). The plastid genome is highly conserved among land plants, both in terms of gene content and gene order. Structural changes are typically rare and hence powerful phylogenetic markers (Mishler and Kelch, 2009), such as in mosses for the circumscription of the Funariales (Goffinet et al., 2007).

At present only six plastid genomes have been assembled for bryophytes (http://www.ncbi.nlm.nih.gov/genomes/GenomesGroup.cgi?taxid=2759&opt=plastid) and only Forrest et al. (2011) bioinformatically extracted the entire genome from 454 data as part of their effort to assemble organellar genomes to reconstruct the backbone of the liverwort tree of life. Within liverworts, the genomes are completely alignable and, with the exception of the non-photosynthetic *Aneura*, may differ only in the presence or absence of the *cys*A and *cys*T genes (Wickett et al., 2011). In mosses, gene content is identical between the distantly related *P. patens* and *Syntrichia ruralis* (Hedw.) F. Weber and D. Mohr (Oliver et al., 2010), although the former differs by a large inversion in the large single copy unit, a character that has been shown to be diagnostic of the Funariales (Goffinet et al., 2007). The only gene loss reported within mosses is that of the *rpo*A gene, which characterizes all arthrodontous mosses and *Tetraphis* Hedw. (Goffinet et al., 2005).

The mitochondrial genome has been vastly ignored as a source of phylogenomic makers, partially due to its great lability in gene order among vascular plants (Knoop, 2004). Within bryophytes, synteny of mitochondrial genomes is weak among the three lineages of bryophytes but seemingly high within each lineage (i.e., liverworts, mosses, and hornworts) (Liu et al., 2012b). For example, *P. patens* and *Anomodon rugelii* (Müll. Hal.) Keissl. (Liu et al., 2011) share the same mitochondrial gene content and gene order. Consequently, ambiguities in the alignment are few and the suite of homologous sites extremely high, making this genome an ideal source of potential phylogenetically informative characters.

In seed plants, the nuclear ribosomal DNA repeat unit includes the three genes coding for rRNA subunits (18S, 5.8S, and 26S), the internal transcribed spacers (ITS1 and ITS2), and the intergenic spacer (IGS). In bryophytes, this region further includes the 5S nrDNA subunit, located between consecutive 26S and 18S genes, separating the IGS region into an IGS1 and IGS2 (Wicke et al., 2011), in the repeat unit. The non-coding regions of the nrDNA repeat are well known as sources of variable characters (Vanderpoorten et al., 2006), and the IGS regions in particular carry abundant substitutions, although the latter region remains poorly explored due to amplification difficulties (Poczai and Hyvönen, 2010). In bryophytes the entire repeat has only been sequenced for *F. hygrometrica* and *Marchantia polymorpha* L. (Vanderpoorten et al., 2006), but is being assembled for *P. patens* (Lang pers. Com. on March, 2012) as part of the *Physcomitrella* genome project (Rensing et al., 2008).

These three regions of the cellular genome, present in numerous copies in each cell, provide optimal candidates for efficient sampling of homologous loci from all genomic compartments using NGS methods. Here we test the yield of a partial NGS run on a 454 platform to recover these regions, and compare their variation among three members of the Funariaceae. We performed a multiplexed NGS with *E. obtusus* (Hedw.) Lindb. and *F. hygrometrica* and aligned the regions to the available organellar genomes of *P. patens* and the nrDNA repeat of *F. hygrometrica*. Besides the organellar and nrDNA sequences, we also isolated microsatellites, which can potentially be used for phylogeographic and population studies. Overall we confirm the results of Forrest et al. (2011) whereby organellar genomes and nrDNA repeats can be successfully obtained for bryophytes through 454 shotgun sequencing, by contributing new, nearly complete plastid and mitochondrial genomes and nrDNA repeat for two mosses.

2. Materials and methods

2.1. Plant material and genome size determination

Spores were collected from herbarium specimens of *F. hygrometrica* (Goffinet 5576; Chile) and *E. obtusus* (Holyoak 04-87; Ireland), and grown on soil for about two months. Vouchers for both species have been deposited in George Safford Torrey Herbarium at the University of Connecticut (CONN). Genome size was determined by flow cytometry according to Bainard et al. (2010). Fresh cultured gametophytes of *E. obtusus* were co-chopped with the standard *Zea mays* L. 'CE-777' (2C DNA content = 5.43 pg) and *F. hygrometrica* with *Raphanus sativus* L. 'Saxa' (2C DNA content = 1.11 pg) in LB01 buffer with 150 ug/ml propidium iodide and 0.5 ug/ml RNase A (Sigma, St. Louis). Each sample was replicated three times and analyzed on a Partec CyFlow SL (Partec, Munster, Germany). The moss 1C DNA content was determined using the ratio between the fluorescence of the standard and sample nuclei.

2.2. DNA extraction and next generation sequencing

Approximately 0.3 g and 0.4 g of fresh gametophytes were harvested from *F. hygrometrica* and *E. obtusus*, respectively. The plant material was ground in liquid nitrogen, and used for DNA extraction by a modified CTAB method (Forrest et al., 2011). DNA quantity and quality were examined by Qubit fluorometer system with a Quant-iT™ ds-DNA BR Assay (Invitrogen, San Diego, CA, USA) and NanoDrop Spectrophotometer ND-1000 (NanoDrop Technologies, Wilmington, DE, USA), respectively. In total, 1.5 μg and 2.0 μg of genomic DNA were collected for *F. hygrometrica* and *E. obtusus*, respectively, and were sent to the IGSP Sequencing Core Facility (Duke University) for GS FLX 454 shotgun library preparation using the Titanium system (Roche 454 Life Science, Branford, CT, USA). The two species and an unrelated moss were multiplexed using MID-tags, and put in an equimolar pool for amplification. Finally, the samples were run on half a picotiter plate on a Genome

Sequencer FLX Instrument (Roche 454 Life Science, Branford, CT, USA).

2.3. De novo assembly and mapping with reference genomes

The raw 454 reads were trimmed to remove adaptor and low-quality sequences, and de novo assembled into contigs using GS Assembler (Roche 454 Life Science, Branford, CT, USA) using the default parameters (minimum overlap between reads 40 bases, minimum overlap identity 90%, reward for a match 2, penalty for a mismatch -3). This approach may not yield a single contig matching the entire genome, and hence the 454 reads were also directly mapped against the reference plastid (NC_005087) and mitochondrial (NC_007945) genomes of *P. patens* using GS Mapper (Roche 454 Life Science, Branford, CT, USA) with the default settings (minimum overlap between reads 40 bases, minimum overlap identity 90%, reward for a match 2, penalty for a mismatch −3). When performing the mapping analyses, one copy of the Inverted Repeat was removed from the plastid reference. For the assembly of the nrDNA repeat for our two samples, the publically available sequence for *F. hygrometrica* was (GenBank accession: X80212) used as a reference.

2.4. Sequence alignment and comparisons

The mapped, or de novo assembled contigs were checked for their quality, and imported into Geneious 5.4.4 (http://www.geneious.com/; Biomatters Ltd., Auckland, New Zealand). The contigs were aligned against each reference sequence and preliminarily annotated in Geneious. One copy of the Inverted Repeat was removed from the plastid reference during the assembly. All gaps among contigs were replaced by equivalent number of "N"s. The assembled plastid and mitochondrial genomes and nuclear ribosomal DNA units of *E. obtusus* and *F. hygrometrica,* and the relevant *P. patens* sequence were aligned using the genome alignment software Mauve 2.3.1 (Darling et al., 2004). As the *P. patens* nrDNA unit is unavailable, we compared the *F. hygrometrica* nrDNA unit to the previously sequenced one (X80212). Each aligned dataset was then imported into MEGA 5.05 (Tamura et al., 2011), which was used to calculate the proportion of variable sites. In addition, all three genomic regions were compared using the online program VISTA (http://genome.lbl.gov/vista/) to visualize the distribution of variations.

2.5. Sequence accession numbers

The assembled plastid (accessions JQ753069 and JQ753071), mitochondrial genome scaffolds (JQ753070, JQ713171), and nrDNA repeat unit sequences (JQ736824, JQ736823) for *E. obtusus* and *F. hygrometrica* are deposited in the GenBank database (Table 1).

2.6. Detecting microsatellite markers

All 454 assemblies were screened for potential microsatellite markers for both species. All de novo assembled contigs (FASTA format) were run through msatCommander 0.8.2 (Faircloth, 2008), accepting mononucleotide repeats of $\geqslant 10$, dinucleotide repeats of $\geqslant 6$, trinucleotide repeats of $\geqslant 4$, tetranucleotide repeats of $\geqslant 4$, pentanucleotide repeats of $\geqslant 4$, and hexanucleotide repeats of $\geqslant 4$. Primers were designed through Primer3 nested inside the msatCommander program.

3. Results and discussion

Organellar genomes and the entire nuclear ribosomal DNA repeat are currently being targeted for phylogenetic reconstructions in plants, either individually (e.g., plastid genome (Wu and Ge, 2012)), or in combination (plastid and nrDNA sequences (Kane et al., 2012), or plastid, nrDNA sequences, and partial mitochondrial coding genes (Steele et al., 2012)). Thus far, the NGS methodology has not been used to mine these genomic regions for inferring phylogenetic histories among bryophytes. Here we demonstrate that nearly complete organellar genomes and entire nuclear ribosomal DNA repeats can be assembled based on a small-scale 454 sequencing of a total genomic extract from less than half a gram of fresh moss tissue. These regions can be extensively if not completely aligned among the three mosses, thereby extending the source of phylogenetic characters to the non-coding regions of these genomic regions.

3.1. Genomic sequence assembly

In total, 135,697 reads (47 Mb in size with an average length of 342 bases) were generated for *E. obtusus.* Although only a small fraction could be mapped to the *P. patens* pt and mt genomes (1.96% and 1.40%, respectively; Table 1), these reads could be assembled to cover 92.8% of the plastid and 97.9% of the mitochondrial reference genomes. For *F. hygrometrica* we obtained 403,252 reads (144 Mb in size with an average length of 359 bases), of which 5.36% and 1.14% mapped to *P. patens* pt and mt genomes, respectively. The 454 contigs covered 90.7% and 98.7% of each reference *P. patens* organellar genome. The average sequencing depth of *E. obtusus* pt and mt genomes are 8× and 6×, and *F. hygrometrica* 75× and 11×, respectively (Table 1).

Mapping reads against the *F. hygrometrica* nrDNA repeat, revealed that 2.79% and 3.51% of the raw sequences of *E. obtusus* and *F. hygrometrica* were homologous, and these could be assembled into complete nuclear rDNA repeat regions. The average sequencing depth of nrDNA region is 104× for *E. obtusus* and 461× for *F. hygrometrica* (Table 1).

In summary, only a small part (ca. 6% in *E. obtusus* and 10% in *F. hygrometrica*) of the 454 reads matches the organellar genomes and high copy nrDNA sequences. Almost all the other data,

Table 1
Data characteristics of plastid, mitochondrial genomes and the nuclear ribosomal DNA repeat unit generated from *Entosthodon obtusus* and *Funaria hygrometrica* 454 assemblies.

Taxon	Sequence[a]	Mapped reads	Mapped bases	No. contigs	Assembly length	Average coverage[c]	Average depth (x)	GenBank accession
Entosthodon	pt genome	2664 (1.96%)	860,563 (1.86%)	51	104,766	92.8%	8	JQ753069
	mt genome	1905 (1.40%)	579,169 (1.25%)	21	102,971	97.9%	6	JQ753070
	nrDNA unit[b]	3792 (2.79%)	1,225,728 (2.64%)	1	11,792	100%	104	JQ736824
Funaria	pt genome	21,614 (5.36%)	7,648,937 (5.30%)	51	102,295	90.7%	75	JQ753071
	mt genome	4588 (1.14%)	1,136,348 (0.79%)	13	103,830	98.7%	11	JQ713171
	nrDNA unit	14,140 (3.51%)	5,106,111 (3.54%)	1	11,077	100%	461	JQ736823

[a] The sequences were assembled by GS Mapper (v2.5.3).
[b] The *Entosthodon* nrDNA sequence was assembled by GS assembler (v2.5.3).
[c] Estimated by comparing to *Physcomitrella patens* sequences.

Table 2
Comparisons of organellar and nuclear rDNA repeat unit sequences between *Physcomitrella patens* and the newly targeted *Funaria hygrometrica* and *Entosthodon obtusus*.

Sequence	Comparison	Whole genome (bp)		Exons only (bp)	
		Length	Mutations (%)	Length	Mutations (%)
pt genome	Among 3 taxa	100,237	5818 (5.8%)	72,100	3382 (4.7%)
	Funaria vs. *Entosthodon*	100,237	5077 (5.1%)	72,100	3012 (4.2%)
	Funaria vs. *Physcomitrella*	100,146	4666 (4.7%)	72,100	2778 (3.9%)
	Physcomitrella vs. *Entosthodon*	100,006	1815 (1.8%)	72,100	1012 (1.4%)
mt genome	Among 3 taxa	102,260	1573 (1.5%)	32,325	244 (0.8%)
	Funaria vs. *Entosthodon*	102,257	1282 (1.3%)	32,325	216 (0.7%)
	Funaria vs. *Physcomitrella*	102,033	1146 (1.1%)	32,325	145 (0.5%)
	Physcomitrella vs. *Entosthodon*	101,997	696 (0.7%)	32,325	128 (0.4%)
nrDNA unit	Among 3 samples	12,642	1225 (9.7%)	5538	55 (1.0%)
	Funaria vs. *Entosthodon*	12,532	1002 (8.0%)	5537	47 (0.8%)
	Funaria (X80212) vs. *Entosthodon*	12,523	1001 (8.0%)	5537	54 (1.0%)
	Funaria (X80212) vs. *Funaria*	11,246	176 (1.6%)	5538	8 (0.1%)

accounting for 44 Mb in *E. obtusus* and 130 Mb in *F. hygrometrica*, are single-copy, low-copy nuclear sequences, or other multiple copy nuclear regions (such as tandem sequences and mobile elements) that are not comparable and hence not homologous between samples. Based on flow cytometry, the size of the nuclear genomes of *E. obtusus* and *F. hygrometrica* are estimated at 1139 Mb and 921 Mb, respectively. Hence, the average sequencing depth of any site in the nuclear genomes of *E. obtusus* and *F. hygrometrica* would be 0.04× and 0.14×, which is clearly two orders of magnitude less than the high copy loci.

3.2. Phylogenomic signals

Among the three species, 5.8% of the aligned plastid sequences are variable (Table 2), which is nearly four times more than the mitochondrial genome (1.5%). Although such degrees of variation may seem low, considering the size of these genomic regions, the absolute number of variable sites is high, 5818 bp and 1573 bp, respectively. *P. patens* and *E. obtusus* are part of the crown group of Funariaceae, which is sister to *Funaria* (Liu et al., 2012a). Not surprisingly then, the degree of divergence in the present genomic regions is lower between the former two species than between either one and *F. hygrometrica* (Table 2). However, relative divergence between the two species differs, since the plastid genomes exhibit more variation than the mitochondrial genomes (Table 2). For example *P. patens* and *E. obtusus* differ in only 696 sites of the mt genome compared to 1815 pt genome sites. For both plastid and mitochondrial genomes, the variable sites are scattered along the genomes and no obvious variation hotspots were observed in the alignments (Fig. 1). For nuclear rDNA repeat unit, most variable sites locate within the internal transcribed spacers and the intergenic spacers (Table 2 and Fig. 2), and the IGS itself contains 176 variable sites between the here and previously sequenced *F. hygrometrica* samples (Table 2).

3.3. Microsatellites isolation

Traditional approaches for identifying microsatellite markers are time consuming, while NGS outputs have proven to provide a rich sampling of microsatellites (Zalapa et al., 2012). Within this study, the screening revealed that 107 contigs, representing 5.19% of the *E. obtusus* assemblies, contained 142 microsatellites; 285 contigs, representing 5.46% of *Funaria* assemblies, contained 403 microsatellites. Meanwhile, 47 and 170 primer pairs flanking microsatellites were successfully generated for *E. obtusus* and *F. hygrometrica*, respectively (Table S1). Based on the results, NGS

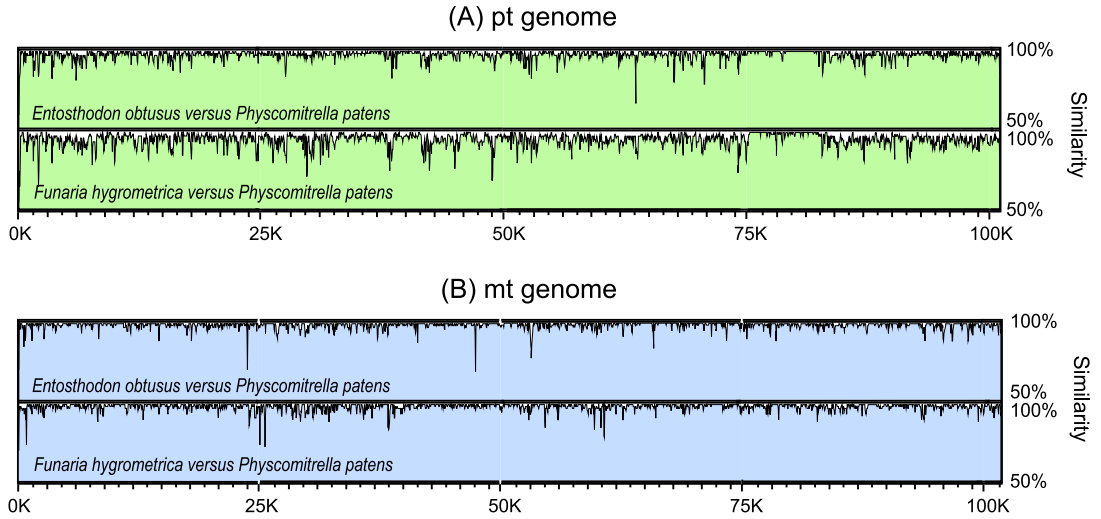

Fig. 1. Pairwise comparison of plastid (A) and mitochondrial (B) genomes of *Funaria hygrometrica* and *Entosthodon obtusus* against those of *Physcomitrella patens*. Sequences are compared in blocks of 100 nucleotides, and the average similarity across the block is plotted.

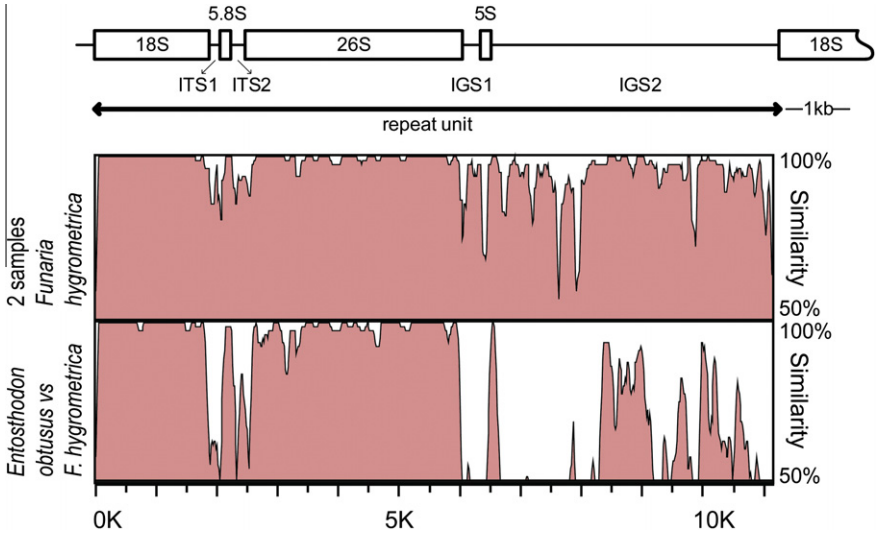

Fig. 2. Pairwise comparison of the nuclear ribosomal DNA repeats (structure shown above) of *Funaria hygrometrica* and *Entosthodon obtusus* against the publically available sequence of another *F. hygrometrica* sample (GenBank Acc. # X80212). Sequences are compared in blocks of 100 nucleotides, and the average similarity across the block is plotted.

technology appears to be an efficient way of discovering microsatellite markers in mosses. Except for the in silico analyses, no further PCR confirmation or polymorphism detection was performed on these microsatellites. The microsatellite primers and 454 assemblies are provided in supplementary Table S2 and S3.

4. Conclusion and future prospects

In the current study, we generated nearly completed plastid and mitochondrial genomes, and the entire nuclear rDNA repeat unit for two species of the moss family Funariaceae, *E. obtusus* and *F. hygrometrica*. By comparing these to the publicly available organellar genomes for *P. patens* and the nrDNA repeat of anther sample of *F. hygrometrica*, we demonstrated that these regions carry numerous variable and potentially phylogenetically informative sites. Shotgun sequencing of total genomic extracts followed by bioinformatic extraction of high copy templates such as organellar genomes and nrDNA repeats, provides a critical tool for sampling large sets of characters needed to resolve deep divergences among moss lineages and also recent radiations, an evolutionary pattern characterizing families such as the Funariaceae (Liu et al., 2012a) and Sphagnaceae (Shaw et al., 2010).

Acknowledgments

This research was supported by the National Science Foundations (DEB-0919284 and EF-0531557) to BG. The authors thank Paul Kron for assistance with the flow cytometry experiment. The UConn Bioinformatics Facility provided computing resources for 454 sequence assembly.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2012.12.006.

References

Bainard, J.D., Fazekas, A.J., Newmaster, S.G., 2010. Methodology significantly affects genome size estimates: quantitative evidence using bryophytes. Cytometry Part A 77, 725–732.

Crosby, M.R., Magill, R.E., Allen, B., He, S., 1999. A Checklist of the Mosses. Missouri Botanical Garden, St. Louis.

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

Faircloth, B.C., 2008. MSATCOMMANDER: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8, 92–94.

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 Ecology and Evolution 144, 29–43.

Goffinet, B., Wickett, N.J., Shaw, A.J., Cox, C.J., 2005. Phylogenetic significance of the *rpo*A loss in the chloroplast genome of mosses. Taxon 54, 353–360.

Goffinet, B., Wickett, N.J., Werner, O., Ros, R.M., Shaw, A.J., Cox, C.J., 2007. Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). Annals of Botany 99, 747–753.

Jian, S., Soltis, P.S., Gitzendanner, M.A., Moore, M.J., Li, R., Hendry, T.A., Qiu, Y.L., Dhingra, A., Bell, C.D., Soltis, D.E., 2008. Resolving an ancient, rapid radiation in Saxifragales. Systematic Biology 57, 38–57.

Kane, N., Sveinsson, S., Dempewolf, H., Yang, J.Y., Zhang, D., Engels, J.M.M., Cronk, Q., 2012. Ultra-barcoding in cacao (*Theobroma* spp.; Malvaceae) using whole chloroplast genomes and nuclear ribosomal DNA. American Journal of Botany 99, 320–329.

Knight, C.D., Perroud, P.F., Cove, D.J., 2009. The moss *Physcomitrella patens*. Annual Plant Reviews 36, 1–349.

Knoop, V., 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Current Genetics 46, 123–139.

Liu, Y., Xue, J.Y., Wang, B., Li, L., Qiu, Y.L., 2011. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: dynamic and conservative evolution. PLoS ONE 6, e25836.

Liu, Y., Budke, J.M., Goffinet, B., 2012a. Phylogenetic inference rejects sporophyte based classification of the Funariaceae (Bryophyta): rapid radiation suggests rampant homoplasy in sporophyte evolution. Molecular Phylogenetics and Evolution 62, 130–145.

Liu, Y., Wang, B., Li, L., Qiu, Y.-L., Xue, J., 2012b. Conservative and dynamic evolution of mitochondrial genomes in early land plants. In: Bock, R., Knoop, V. (Eds.), Genomics of Chloroplasts and Mitochondria. Springer, Netherlands, pp. 159–174.

Mishler, B., Kelch, D., 2009. Phylogenomics and early land plant evolution. In: Goffinet, B., Shaw, A.J. (Eds.), Bryophyte Biology, second ed. Cambridge University Press, Cambridge, United Kingdom, pp. 173–197.

Nock, C.J., Waters, D.L.E., Edwards, M.A., Bowen, S.G., Rice, N., Cordeiro, G.M., Henry, R.J., 2011. Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnology Journal 9, 328–333.

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.

Poczai, P., Hyvönen, J., 2010. Nuclear ribosomal spacer regions in plant phylogenetics: problems and prospects. Molecular Biology Reports 37, 1897–1912.

Rensing, S.A. et al., 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319, 64–69.

Rogers, S.O., Bendich, A.J., 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. Plant Molecular Biology 9, 509–520.
Shaver, J.M., Oldenburg, D.J., Bendich, A.J., 2006. Changes in chloroplast DNA during development in tobacco, *Medicago truncatula*, pea, and maize. Planta 224, 72–82.
Shaw, A.J., Devos, N., Cox, C.J., Boles, S.B., Shaw, B., Buchanan, A.M., Cave, L., Seppelt, R., 2010. Peatmoss (*Sphagnum*) diversification associated with Miocene Northern Hemisphere climatic cooling? Molecular Phylogenetics and Evolution 55, 1139–1145.
Steele, P.R., Hertweck, K.L., Mayfield, D., McKain, M.R., Leebens-Mack, J., Pires, J.C., 2012. Quality and quantity of data recovered from massively parallel sequencing: examples in Asparagales and Poaceae. American Journal of Botany 99, 330–348.
Stenøien, H., 2008. Slow molecular evolution in 18S rDNA, *rbc*L and *nad*5 genes of mosses compared with higher plants. Journal of Evolutionary Biology 21, 566–571.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., Kumar, S., 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Molecular Biology and Evolution 28, 2731–2739.
Vanderpoorten, A., Quandt, D., Goffinet, B., 2006. Utility of the internal transcribed spacers of the 18S-5.8S-26S nuclear ribosomal DNA in land plant systematics with special emphasis on Bryophytes. Plant genome: biodiversity and evolution—vol. 2, Part B, pp. 385–407.
Wicke, S., Costa, A., Muṇoz, J., Quandt, D., 2011. Restless 5S: the re-arrangement (s) and evolution of the nuclear ribosomal DNA in land plants. Molecular Phylogenetics and Evolution 61, 321–332.
Wickett, N.J., Forrest, L.L., Budke, J.M., Shaw, B., Goffinet, B., 2011. Frequent pseudogenization and loss of the plastid-encoded sulfate-transport gene cysA throughout the evolution of liverworts. American Journal of Botany 98, 1263–1275.
Wu, Z.Q., Ge, S., 2012. The phylogeny of the BEP clade in grasses revisited: evidence from the whole-genome sequences of chloroplasts. Molecular Phylogenetics and Evolution 62, 573–578.
Zalapa, J.E., Cuevas, H., Zhu, H., Steffan, S., Senalik, D., Zeldin, E., McCown, B., Harbut, R., Simon, P., 2012. Using next-generation sequencing approaches to isolate simple sequence repeat (SSR) loci in the plant sciences. American Journal of Botany 99, 193–208.