Complete mitochondrial genome sequence of the "copper moss" *Mielichhoferia elongata* reveals independent *nad7* gene functionality loss

Denis V. Goruynov[1], Svetlana V. Goryunova[2], Oxana I. Kuznetsova[3], Maria D. Logacheva[1], Irina A. Milyutina[1], Alina V. Fedorova[1], Michael S. Ignatov[3] and Aleksey V. Troitsky[1]

[1] Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russia
[2] Institute of General Genetics Russian Academy of Science, Moscow, Russia
[3] Tsitsin Main Botanical Garden Russian Academy of Science, Moscow, Russia

ABSTRACT

The mitochondrial genome of moss *Mielichhoferia elongata* has been sequenced and assembled with Spades genome assembler. It consists of 100,342 base pairs and has practically the same gene set and order as in other known bryophyte chondriomes. The genome contains 66 genes including three rRNAs, 24 tRNAs, and 40 conserved mitochondrial proteins genes. Unlike the majority of previously sequenced bryophyte mitogenomes, it lacks the functional *nad7* gene. The phylogenetic reconstruction and scrutiny analysis of the primary structure of *nad7* gene carried out in this study suggest its independent pseudogenization in different bryophyte lineages. Evaluation of the microsatellite (simple sequence repeat) content of the *M. elongata* mitochondrial genome indicates that it could be used as a tool in further studies as a phylogenetic marker. The strongly supported phylogenetic tree presented here, derived from 33 protein coding sequences of 40 bryophyte species, is consistent with other reconstructions based on a number of different data sets.

INTRODUCTION

Bryophytes (mosses, liverworts, and hornworts) represent an ancient group of higher plant evolution that shows a dominance of the gametophyte stage in the life cycles. These nonvascular pioneers of land plants first acquired adaptations that enabled the transition from aquatic to terrestrial habitats. Mosses (Bryophyta), branched off from the stem of the Embryophyta phylogenetic tree after the Marchantiophyta and before the separation of the Anthocerotophyta (*Liu et al., 2014*; *Qiu et al., 2006*).

The mitogenomes (MGs) of mosses have recently become a target of sequencing efforts for phylogenetic reconstructions due to their compact size and a higher degree of synteny than is observed in vascular plants (*Liu, Medina & Goffinet, 2014*). The NCBI RefSeq

database (http://www.ncbi.nlm.nih.gov/refseq) currently contains 39 mitochondrial genomes for members of this group of plants. This quite limited data set includes only representatives of nine orders from the Bryopsida and three orders from other classes of mosses, so it clearly does not perfectly reflect bryophyte diversity. The aim of the present study is to extend bryophyte taxonomical coverage and expand the phylogenetic analysis to include MG data from the still unexplored order Bryales. For this purpose the complete MG of *Mielichhoferia elongata* (Hoppe & Hornsch.) Nees & Hornsch. was sequenced.

The plant list (http://www.theplantlist.org, Version 1.1 September 2013) contains 142 accepted species names of *Mielichhoferia* Nees & Hornsch. The taxonomical status of *Mielichhoferia* remains under debate. The genus has usually been treated as the subfamily Mielichhoferioideae within the Bryaceae, although some authors have attributed it to a separate family, the Mielichhoferiaceae (*Hill et al., 2006*; *Shaw, 2014*). The only molecular study of phylogenetic relationships of *Mielichhoferia* placed the Mielichhoferiaceae within the Mniaceae according to the *trnL-F* and *rps4* sequence data (*Guerra et al., 2011*).

Several moss and hepatic species are restricted to substrates enriched in heavy metals. These bryophytes that show an affinity for metalliferous substrates have been referred to as "copper mosses" (*Antonovics, Bradshaw & Turner, 1971*; *Chopra & Kumra, 1988*; *Persson, 1948*, *1956*; *Shaw, 1987*, *1989*). *M. elongata* Homsch. (*Shaw, 2000*) and the closely related *Mielichhoferia mielichhoferiana* (Funck.) Loeske are among the species that are highly tolerant and largely restricted to substrates enriched in copper. These species are widely distributed around the globe, but are always rare. They grow in habitats rich in copper (often associated with other metals) and inorganic sulfides, which results in a very low pH. These habitats represent areas damaged by mining (mine waste tailings) or metal-rich rocks. The heavy metal tolerance mechanisms are not well understood and apparently vary across species. Metals are adsorbed by the cell walls and are accumulated in cells (*Antonovics, Bradshaw & Turner, 1971*; *Antreich, Sassmann & Lang, 2016*; *Brown, 1982*; *Meharg, 2005*; *Tyler, 1990*).

Knowledge of the sequence of the mitochondrial genome of *M. elongata* will be useful both for finding an appropriate taxonomic treatment for the taxa and for population studies within the Mielichhoferia. The latter studies are particularly important in light of the disruptive character of the habitat area, the rarity of these species, and ongoing habitat damage.

MATERIALS AND METHODS

Sample collection and DNA isolation

The *M. elongata* samples were collected from July 12 to 17, 2011 in the area near Mus-Khaya Peak (62°31′–36′N, 140°56′–141°07′E), Republic of Sakha (Yakutia) and deposited in MHA, the Herbarium of the Main Botanical Garden Russian Academy of Science, Moscow (*Ignatova et al., 2011*). This moss was originally identified in a cited paper as *M. mielichhoferiana*. However, a subsequent analysis of the nuclear rDNA 5.8S-ITS 2 region attributed this plant to that of morphologically hardly distinguishable

M. elongata (Fig. S1). Rocks in the area are especially rich in MnS, with other heavy metals (Pb, Sn, As, Zn, Ag, etc., usually as sulfides) present in high concentrations. Consequently, many brooks have very acidic water and sulfur deposits along them. Siderite (iron carbonate) forms red outcrops ("Red rocks") rich in iron and is always enriched with other heavy metals. When the outcrops are dry, *M. elongata* is the only moss that grows on this substrate or at least it is the only particularly abundant one.

DNA was extracted from specimens in the herbarium collection that had been gathered with a minimal soil amount and dried using ordinary herbarium techniques (in a paper envelope, under a tent, in the shade for several days until dry), and then stored in the herbarium at room temperature. A Nucleospin Plant DNA Kit (Macherey Nagel, Düren, Germany) was used for total DNA extraction from whole shoots of plants according to the manufacturers' protocol. A yield of about 2 μg DNA was obtained according to measurements determined with a Qubit fluorometer (Invitrogen, Carlsbad, CA, USA).

Library preparation and sequencing

A 500 ng sample of genomic DNA was fragmented using a Covaris S220 sonicator (Covaris, Woburn, MA, USA) and a library was prepared using TruSeq DNA sample preparation kit (Illumina, Mountain View, CA, USA). The concentration of the prepared library was measured with the Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) and qPCR and fragment length distribution was determined with Bioanalyzer 2100 (Agilent, Santa Clara, CA, USA). The library was diluted to 10 pM and used for cluster generation on a cBot instrument with TruSeq PE Cluster Kit v3 reagents (Illumina, Mountain View, CA, USA). Sequencing was performed on a HiSeq2000 sequencer with read length of 101 from both ends of the fragments. About six million read pairs were obtained.

Mitogenome assembly and annotation

Raw sequencing reads were preprocessed with Trimmomatic software (*Bolger, Lohse & Usadel, 2014*) to remove adapters and low-quality data from further analysis. The whole genome assembly was then accomplished using the Spades assembler (*Bankevich et al., 2012*). A Blast database was generated from the assembled contigs, and a Blast search was performed against the *Physcomitrella patens* MG sequence (*Terasawa et al., 2007*) using the standalone NCBI BLAST-2.2.29+ (*Altschul et al., 1990*). The longest hit was the *M. elongata* complete MG. Iterative mapping was carried out using Geneious R10 software (https://www.geneious.com; *Kearse et al., 2012*) to verify the assembled genome. The resulting sequence had almost 100× coverage depth. The correctness of the genome boundaries was verified by PCR amplification followed by Sanger sequencing. Initial reads mapping to the genome sequence with Bowtie 2 (*Lingmead et al., 2009*) was applied as an additional genome structure verification step.

Genome annotation based on sequence similarity was performed using Geneious software. The MG sequence of *Bartramia pomiformis* which gave a maximum score in a BLAST search against a *M. elongata* MG query was applied as a reference. The annotated genome sequence was submitted to GenBank (accession number: MF417767). A circular

genome map was drawn using the CGView Server (*Grant & Stothard, 2008*; http://stothard.afns.ualberta.ca/cgview_server).

SSR analysis

Simple sequence repeats (SSRs) were detected and located in the MG of *M. elongata* using GMATo v1.2 software (*Wang, Lu & Luo, 2013*).

Phylogenomic analysis

Phylogenetic reconstruction was conducted by selecting only functional protein-coding sequences (CDS) present in MGs of all bryophytes under investigation. A total of 33 of these CDS are known, including *atp1, atp4, atp6, atp8, atp9, cob, cox1, cox2, cox3, nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad9, rpL2, rpL5, rpL6, rpL16, rps1, rps2, rps4, rps7, rps11, rps12, rps13, rps14, rps19, sdh3, sdh4, and tatC*. These were extracted from the MG sequences of 39 mosses, liverwort *Treubia lacunose* available in GenBank (http://www.ncbi.nlm.nih.gov), and the *M. elongata* sequenced in this work. The GenBank files were imported into Geneious R10 and merged to export a fasta dataset file. All sequences from this dataset were aligned using the default option implemented in MAFFT (*Katoh & Standley, 2013*). The final alignment was adjusted manually in BioEdit 7.2.5 (*Hall, 1999*).

Phylogenetic reconstruction was performed using the Bayesian method with the program MrBayes v3.2.6 (*Ronquist et al., 2012*). For Bayesian analyses, we used a parallel MPI version of MrBayes (*Altekar et al., 2004*). Two simultaneous runs of Metropolis Coupled Markov Chain Monte Carlo (MC3), both with one cold and seven heated chains were performed for 10 million generations. Two starting trees were chosen randomly. The general time reversible evolutionary model (GTR+I+G) with four rate categories was used. Posterior probabilities (PP) for trees and parameters were saved every 1,000 generations and parameters for each data partition were sampled independently from each other; the first 25% of the trees was discarded in each run. Bayesian PPs were used as branch support values.

RESULTS

Structure of the *M. elongata* mitogenome

The MG of *M. elongata* is 100,342 bp in length and has a typical circular structure (Fig. 1). The nucleotide composition of this genome has a GC content of 39.8%. The MG of *M. elongata* contains 66 genes including genes for three rRNAs (*rrn18, rrn26,* and *rrn5*), 24 tRNAs, and 39 conserved mitochondrial proteins (15 ribosomal proteins, four ccm proteins, eight nicotinamide adenine dinucleotide dehydrogenase subunits, five ATPase subunits, two succinate dehydrogenase subunits, one apocytochrome b, three cytochrome oxidase subunits, and one twin-arginine translocation complex subunit). Besides the functional genes, a single pseudogene, *nad7*, resides in the genome (Table 1).

Structure of *nad7* gene in bryophytes

The lack of a functional gene copy of the *nad7* gene has been reported previously in the MG of hornworts and the majority of liverworts (*Groth-Malonek et al., 2007*;
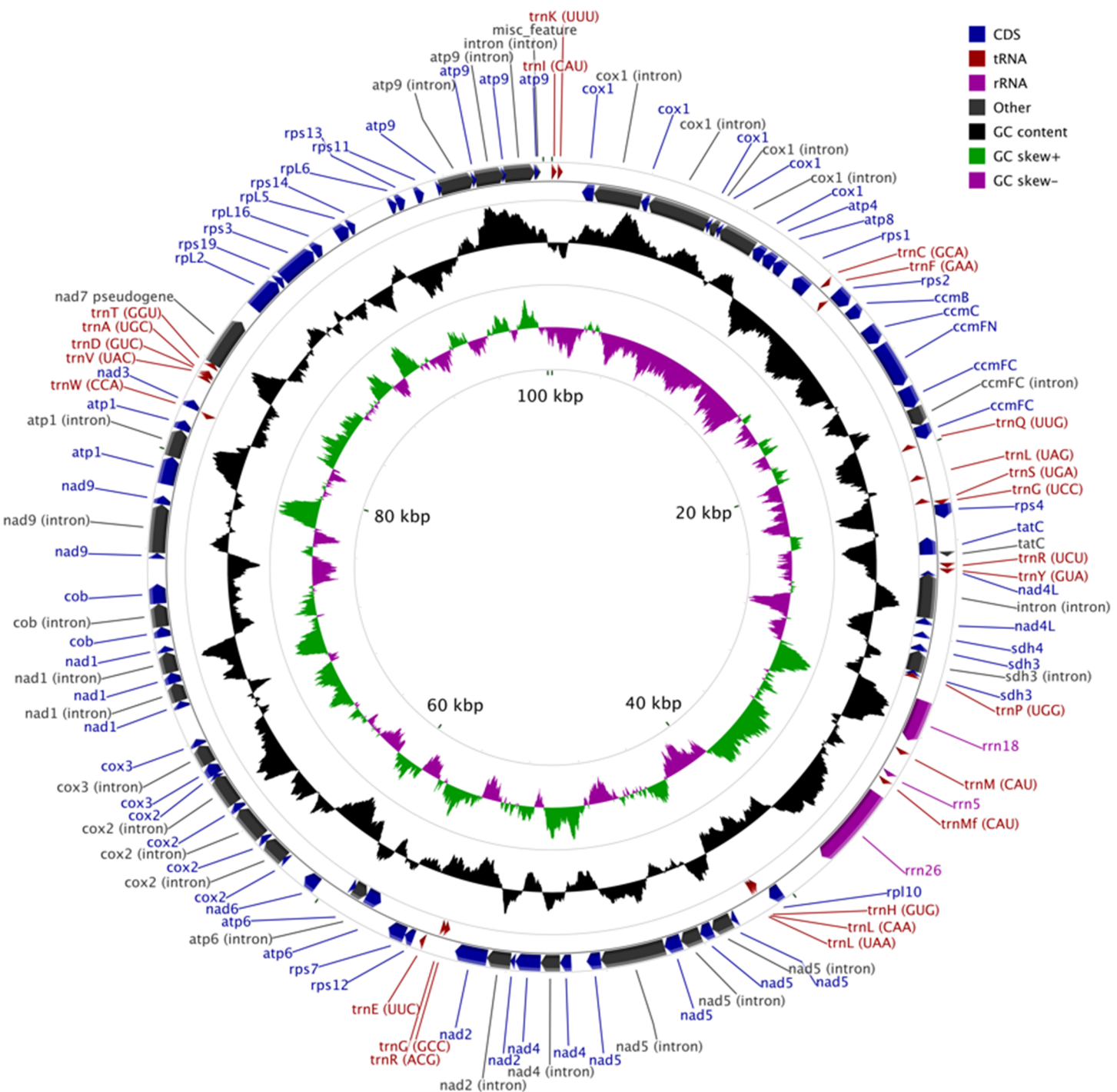
Figure 1 Mitogenome map of *Mielichhoferia elongata* (MF417767) consisting of 100,342 base pairs.
Full-size ☑ DOI: 10.7717/peerj.4350/fig-1

Li et al., 2009; *Xue et al., 2010*). Evolution and losses of the functionality of the gene copies within mosses also deserve special attention and scrutiny. Pseudogenization of the *nad7* gene is currently described for *Tetraphis pellucida* and *Buxbaumia aphylla* (*Bell et al., 2014*; *Liu, Medina & Goffinet, 2014*), whereas all other sequenced bryophyte MGs have a

Table 1 Gene contents in *Mielichhoferia elongata* mitogenome (66 genes, one pseudogene).

Category	Group of genes	Genes	Number of genes
RNA genes	rRNAs	*rrn18, rrn26, rrn5*	3
	tRNAs	*trnA (UGC), trnC (GCA), trnD (GUC), trnE (UUC), trnF (GAA), trnG (GCC), trnG (UCC), trnH (GUG), trnI (CAU), trnK (UUU), trnL (CAA), trnL (UAA), trnL (UAG), trnM (CAU), trnMf (CAU), trnP (UGG), trnQ (UUG), trnR (ACG), trnR (UCU), trnS (UGA), trnT (GGU), trnV (UAC), trnW (CCA), trnY (GUA)*	24
Conserved mitochondrial proteins	Large ribosomal subunits	*rpl10, rpL16, rpL2, rpL5, rpL6*	5
	Small ribosomal subunits	*rps1, rps11, rps12, rps13, rps14, rps19, rps2, rps3, rps4, rps7*	10
	Cytochrome c maturation proteins	*ccmB, ccmC, ccmFC, ccmFN*	4
	Nicotinamide adenine dinucleotide dehydrogenase subunits	*nad1, nad2, nad3, nad4, nad4L, nad5, nad6, nad9*	8
	ATPase subunits	*atp1, atp4, atp6, atp8, atp9*	5
	Succinate dehydrogenase subunits	*sdh3, sdh4*	2
	Apocytochrome b	*cob*	1
	Cytochrome oxidase subunits	*cox1, cox2, cox3*	3
	Twin arginine translocation complex subunit	*tatC*	1
Pseudogenes		*nad7pseudo*	1

functional gene, that consists of three exons separated by two introns. The only known exception is the *nad7* locus structure in MG of *Hypnum imponens* (NC 024516), its functional gene consists of only two exons and one intron sequences. The intron 2 of the gene was lost and exons 2 and 3 were merged together in one exon sequence. The low conservation of the pseudogene sequences has created difficulties in constructing a reliable nucleotide alignment and unambiguously judging whether exons 2 and 3 are completely deleted in either these chondriomes or whether some exon remnants are still preserved. We performed a Tblastn search of these exons amino acid sequence of the *nad7* gene in *T. pellucida* and *B. aphylla* and confirmed the absence of exon 2 in *B. aphylla* and exon 3 in both species. The same finding is evident from Fig. S2 with the alignment of *nad7* from *B. aphylla*, *T. pellucida*, *M. elongata*, and six other moss species. It agrees with the earlier data provided by *Bell et al. (2014)* on the structure of the *T. pellucida* MG. In addition, *B. aphylla* and *T. pellucida* pseudogenes have deletions in the sequences of the first gene exon, although at different locations. The main difference in the *nad7* pseudogene primary structure in these bryophytes is two deletions in the sequence of exon 2 in *B. aphylla*, whereas *T. pellucida* has an intact exon 2 sequence. By contrast, the *nad7* pseudogene of *M. elongata* completely lacks the second exon and has intact exon 1 sequence and exon 3 with frame shift mutation as a result of 2 bp insertion located at 190 bp from 5′ end of the exon (Fig. 2).

SSR analysis of the *M. elongata* mitochondrial genome

Following more stringent criteria (*Zhao, Zhu & Liu, 2016*) of perfect SSR locus identification (minimal number of repeating units ≥10 for mononucleotides,
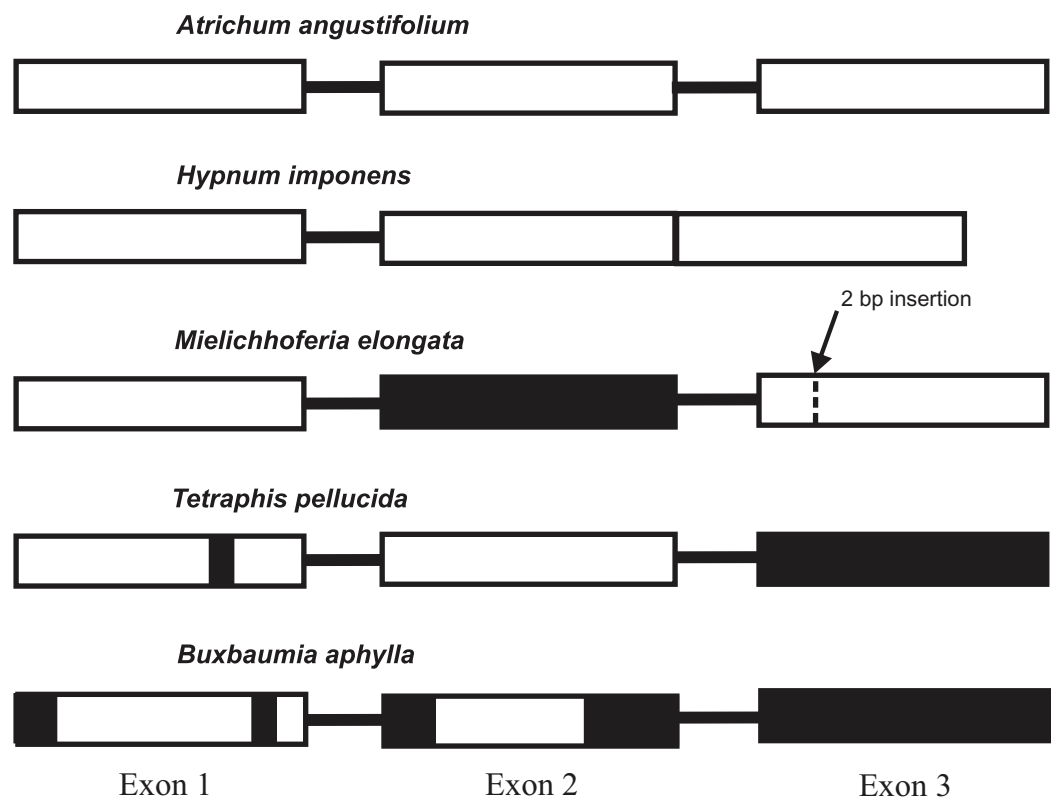
Figure 2 The diversity of the mitochondrial *nad7* gene exon structure in mosses. The majority of the sequenced moss mitogenomes have the same type of locus structure found in *Atrichum angustifolium*. The black filled sections indicate absent exons (or parts of them).

Full-size ☑ DOI: 10.7717/peerj.4350/fig-2

≥5 for dinucleotides, ≥4 for trinucleotides, and ≥3 for tetra-, penta- and hexanucleotides) 73 SSR loci were identified in the MG of *M. elongata* (Table 2; Fig. 3). Most microsatellites refer to mono- and dinucleotides classes (35 and 28 loci, respectively). Trinucleotides are the least frequent SSRs group in the genome (one locus). No hexanucleotide microsatellite repeats occur in the genome. Among all the SSRs, 87.67% are composed only of A/T bases. The total length of the SSR loci is 852 bp, which comprises approximately 0.85% of the genome length.

Phylogenetic analysis

The alignment of 33 mitochondrial protein CDS of 40 moss taxa and hepatic *Treubia lacunosa* (Haplomitriopsida, Treubiidae, Treubiales, Treubiaceae) consists of 24,827 positions. The Bayesian phylogenetic tree inferred from this data with the hepatic *T. lacunosa* as an outgroup is shown in Fig. 4. Most nodes of the tree have very high PP supports. Two exceptions are two nodes among the Orthotrichaceae.

DISCUSSION

We performed sequencing and analysis of the MG of *M. elongata,* a rare "copper moss" with an ambiguous taxonomic status. The MG size significantly varies even among closely

Table 2 SSR-loci of *Mielichhoferia elongata* mitogenome.

Type of repeat unit	Motif	Repetitions	StartPos	EndPos
Mono-	A	10	269	278
Mono-	A	10	13,526	13,535
Mono-	A	10	22,179	22,188
Mono-	A	10	25,861	25,870
Mono-	A	10	40,307	40,316
Mono-	A	10	46,592	46,601
Mono-	A	10	49,092	49,101
Mono-	A	10	52,217	52,226
Mono-	A	10	54,565	54,574
Mono-	A	10	62,618	62,627
Mono-	A	10	88,341	88,350
Mono-	A	10	91,128	91,137
Mono-	A	10	93,879	93,888
Mono-	A	11	39,390	39,400
Mono-	A	12	16,182	16,193
Mono-	G	10	98,368	98,377
Mono-	G	12	52,784	52,795
Mono-	G	12	57,418	57,429
Mono-	T	10	29,873	29,882
Mono-	T	10	46,703	46,712
Mono-	T	10	47,865	47,874
Mono-	T	10	56,400	56,409
Mono-	T	10	57,552	57,561
Mono-	T	10	86,976	86,985
Mono-	T	10	94,469	94,478
Mono-	T	10	99,146	99,155
Mono-	T	11	16,200	16,210
Mono-	T	11	25,885	25,895
Mono-	T	11	40,958	40,968
Mono-	T	11	50,459	50,469
Mono-	T	11	58,416	58,426
Mono-	T	11	95,793	95,803
Mono-	T	12	17,608	17,619
Mono-	T	12	100,200	100,211
Mono-	T	15	11,233	11,247
Di-	AT	5	32,938	32,947
Di-	AT	5	54,921	54,930
Di-	AT	6	14,278	14,289
Di-	AT	6	14,298	14,309
Di-	AT	6	59,230	59,241
Di-	AT	7	70,407	70,420

Table 2 (*continued*).

Type of repeat unit	Motif	Repetitions	StartPos	EndPos
Di-	TA	5	195	204
Di-	TA	5	279	288
Di-	TA	5	466	475
Di-	TA	5	27,730	27,739
Di-	TA	5	41,770	41,779
Di-	TA	5	44,628	44,637
Di-	TA	5	62,826	62,835
Di-	TA	5	68,954	68,963
Di-	TA	5	69,190	69,199
Di-	TA	6	12,557	12,568
Di-	TA	6	86,244	86,255
Di-	TA	6	94,457	94,468
Di-	TA	7	10,767	10,780
Di-	TA	7	19,813	19,826
Di-	TA	7	25,871	25,884
Di-	TA	7	28,304	28,317
Di-	TA	7	29,340	29,353
Di-	TA	7	41,786	41,799
Di-	TA	8	57,533	57,548
Di-	TA	8	69,397	69,412
Di-	TA	10	100,045	100,064
Di-	TA	11	72,289	72,310
Tri-	TTA	4	70,696	70,707
Tetra-	AATA	3	54,140	54,151
Tetra-	ATAA	3	25,162	25,173
Tetra-	ATAG	3	10,865	10,876
Tetra-	ATTT	3	69,685	69,696
Tetra-	CATA	3	25,129	25,140
Tetra-	TACC	3	76,426	76,437
Tetra-	TAGA	3	85,926	85,937
Penta-	AACAA	3	54,704	54,718
Penta-	AAGAA	3	75,527	75,541

related flowering plants (*Allen et al., 2007*; *Alverson et al., 2010*; *Cho et al., 2004*; *Sloan et al., 2012*), but it is extremely stable in bryophytes (*Liu, Medina & Goffinet, 2014*). The MG of *M. elongata* is 383 bp smaller than the genome of *B. aphylla* (*Liu, Medina & Goffinet, 2014*), which to date is the smallest MG among bryophytes. However, the MG of *M. elongata* contains the same set of genes and a similar genome structure to that of other mosses. The only difference is a pseudogenization of the *nad7* gene.

This locus encodes subunit 7 of NADH dehydrogenase (NDH-1 or complex I of the mitochondrial electron transfer chain) is located on the inner mitochondrial membrane and plays an important role in oxidative phosphorylation process (*Bonen et al., 1994*).
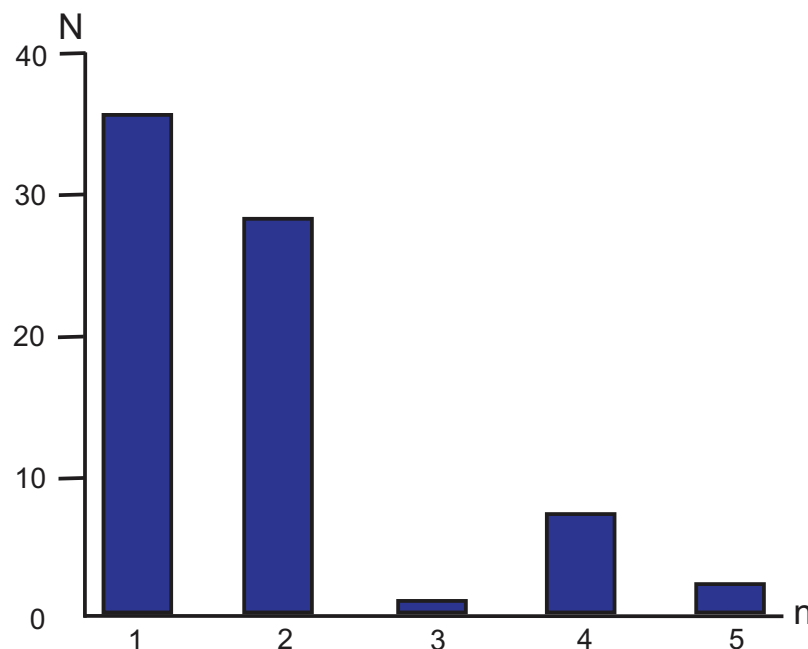
Figure 3 Simple sequence repeat unit length distribution in *M. elongata* mitogenome. *n*—the numbers of base pairs ($n = 1, 2, 3, 4,$ and 5) in different microsatellite classes. *N*—the number of loci in each SSR category. Full-size ☒ DOI: 10.7717/peerj.4350/fig-3

NDH-1 is a quite complicated protein complex, consisting of approximately 30–40 subunits (*Kerscher et al., 2008*). The majority of the subunits are encoded in nuclear genome, but several proteins of the complex are specified by mitochondrial genes (*Bonen et al., 1994*).

Although the MGs of the Bryophyta are highly stable in terms of their gene content, there are two other mosses, *B. aphylla* and *T. pellucida* that lack the intact open reading frame (ORF) of the *nad7* gene in their MGs (*Bell et al., 2014*; *Liu, Medina & Goffinet, 2014*). In our study, we found that the exon structure of *nad7* pseudogene of *M. elongata* differs substantially when compared with that of the MGs of *B. aphylla* and *T. pellucida*. Taking into account the close location of the later on the constructed a phylogenetic tree (Fig. 4) and the extremely distant position of *M. elongata* relative to them, the loss of the functionality of the *nad7* gene can be concluded to have occurred at least twice during the evolutionary history of the mosses.

Intact *nad7* genes were found in the MGs of different angiosperms clades (*Adams & Palmer, 2003*) and in representatives of hornworts, lycophytes, ferns and gymnosperms (*Guo et al., 2017*; *Li et al., 2009*; *Xue et al., 2010*). However several exceptions were noted in different evolutionary lineages. Therefore, the absence of a functional *nad7* gene was noted in the MG of *Nicotiana sylvestris* cytoplasmic male sterile (CMS) mutants (*Pla et al., 1995*) and in the lycophyte *Huperzia squarrosa* (*Liu et al., 2012*). In the liverwort *Marchantia polymorpha*, a functional *nad7* gene was transferred from the MG to nucleus, but the pseudogene was preserved in the MG (*Kobayashi et al., 1997*).
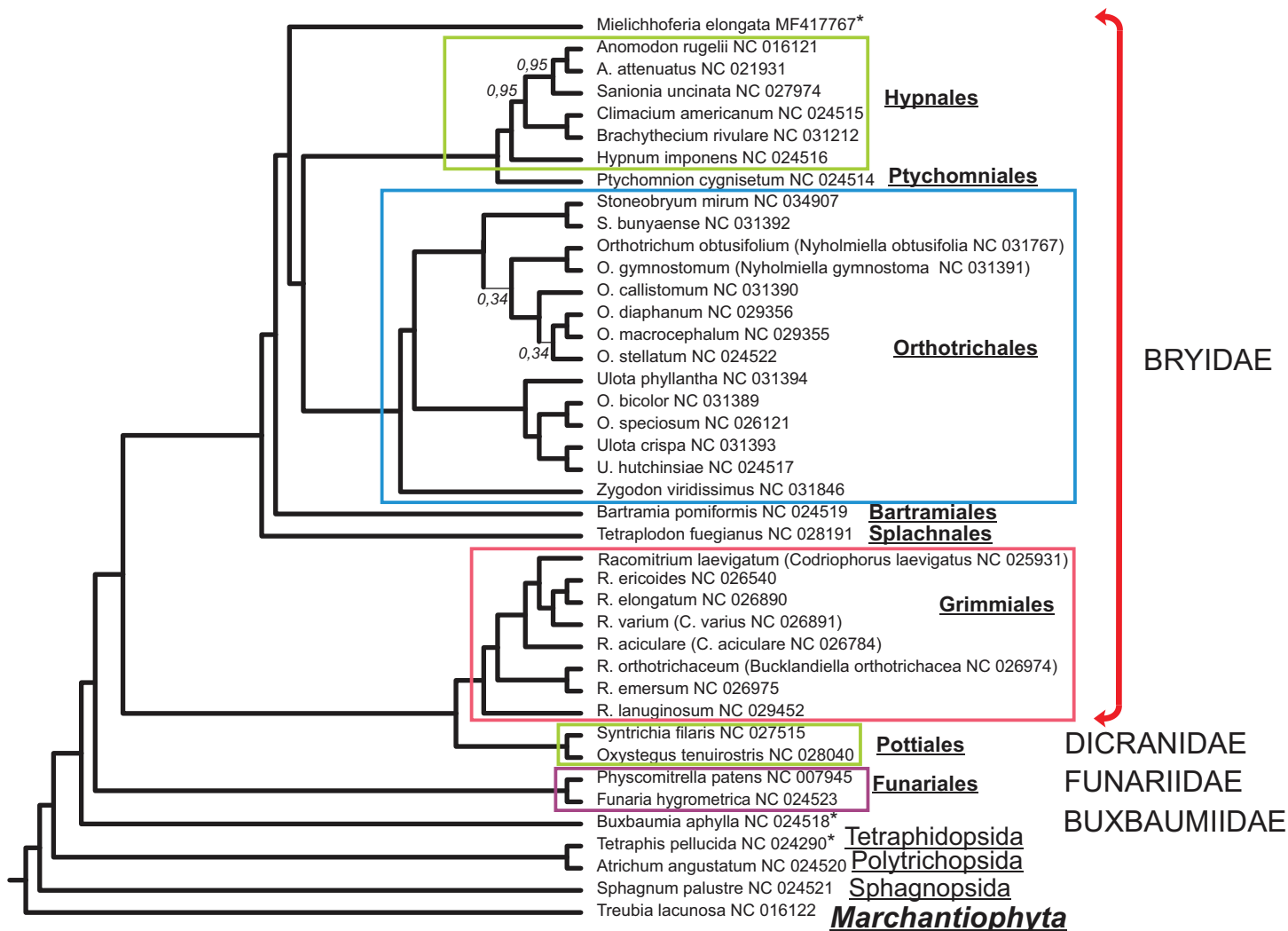
Figure 4 **Bayesian phylogenetic tree of 40 Bryophyta species constructed for 33 mitochondrial protein coding sequences.** The hepatic *Treubia lacunosa* was used as an outgroup. All nodes, except where indicated on the tree, have maximal posterior probability values equal to 1.0. Asterisks indicate taxa with pseudogenization of *nad7*. Full-size 🖼 DOI: 10.7717/peerj.4350/fig-4

Pseudogenization of *nad7* was observed in 11 other liverwort groups, whereas the intact gene was found in *Haplomitrium mnioides* MG (*Groth-Malonek et al., 2007*). This discovery suggested a basal placement of the taxon among liverworts. Overall, pseudogenization of *nad7* may have occurred independently in different unrelated lineages of embryophytes.

A total of 73 SSRs (microsatellites) loci were identified in the MG of *M. elongata*. SSRs are common in plant and animal genomes and could play an important role in gene functioning (*Li et al., 2004*). Besides the occurrence of the SSR loci in nuclear genomes, microsatellite repeats are present in plastids and MGs as well (*Kumar, Kapil & Shanker, 2014*; *Sablok et al., 2015*). However, to date, it is much less known about distribution and functions of microsatellites in bryophyte genomes. SSR loci are usually characterized by

high mutation rate, and therefore actively used as molecular markers in population genetics surveys (*Zalapa et al., 2012*). Molecular markers based on organellar microsatellites have been used successfully for phylogeny reconstruction at the genus taxonomic level and for intraspecific variation analysis (*Ishii, Mori & Ogihara, 2001*; *Nishikawa, Vaughan & Kadowaki, 2005*). The SSR loci revealed in the MG of *M. elongata* could therefore be further investigated to obtain informative markers for using in monitoring programs for *Mielichhoferia* species. That is especially important due to the disruptive character of the habitat area, the rarity of the species, and ongoing habitat damage.

Mielichhoferia elongata represents a separate branch on a phylogenetic tree within the Bryidae and is closest to the Hypnales/Ptychomniales/Orthotrichales group. However, the absence of a MG sequence for the Mniaceae and Bryaceae representatives preclude clarification of the taxonomic position of *Mielichhoferia*. The phylogenetic tree depicted in Fig. 4 inferred from 33 mitochondrial CDSes of 40 mosses species with liverwort as an outgroup, is consistent with other reconstructions based on 14–17 plastid genes from 43 moss species representing the major lineages summarized by *Chang & Graham (2013)*, and based on 41 concatenated mitochondrial protein-coding genes from 19 Bryophyta species (*Liu et al., 2014*). Although plant mitochondrial sequences evolve slowly (*Palmer & Herbon, 1988*), phylogenomic analyses can be effective for bryophytes taxa of both lower and higher ranks. Of course, the remarks of *Liu et al. (2014)* and other earlier authors should be kept in minds; namely, that even high support does not guarantee that an inferred phylogeny is approaching the true evolutionary history.

CONCLUSION

This study provides the complete MG sequence of the "copper moss" *Mielichhoferia elongata* consisting of 100,342 base pairs. It is the smallest known mitochondrial genome among bryophytes and non-parasitic tracheophytes. *M. elongata* is a moss with very specific requirements regarding environmental conditions; in particular, it is mostly confined to heavy metals enriched substrates. Although the MG has the same gene set as that found within previously studied mosses and does not demonstrate any special features associated with high heavy metal tolerance, it lacks a functional *nad7* gene. Based on the phylogeny reconstruction data and exon structure analysis of the gene, it has been deduced, that *nad7* pseudogenization took place independently not once in moss evolution. The phylogenetic tree presented in this study, inferred from the 33 mitochondrial CDS of 41 bryophyte species is consistent with the reconstructions made in earlier studies.

ACKNOWLEDGEMENTS

ADDITIONAL INFORMATION AND DECLARATIONS

Competing Interests
The authors declare that they have no competing interests.

Author Contributions
- Denis V. Goruynov analyzed the data, wrote the paper, prepared figures and/or tables.
- Svetlana V. Goryunova analyzed the data, wrote the paper, reviewed drafts of the paper.
- Oxana I. Kuznetsova performed the experiments.
- Maria D. Logacheva conceived and designed the experiments, performed the experiments, contributed reagents/materials/analysis tools.
- Irina A. Milyutina performed the experiments, analyzed the data.
- Alina V. Fedorova performed the experiments.
- Michael S. Ignatov reviewed drafts of the paper, plants collection and identification.
- Aleksey V. Troitsky conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, reviewed drafts of the paper.

DNA Deposition
The following information was supplied regarding the deposition of DNA sequences:
The *Mielichhoferia elongata* mitochondrial genome sequence described here is accessible via GenBank accession number MF417767.

Data Availability
The following information was supplied regarding data availability:
The raw data has been uploaded as Supplemental Dataset Files.

Supplemental Information
Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.4350#supplemental-information.

REFERENCES
Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* **29**(3):380–395 DOI 10.1016/S1055-7903(03)00194-5.

Allen JO, Fauron CM, Minx P, Roark L, Oddiraju S, Lin GN, Meyer L, Sun H, Kim K, Wang C, Du F, Xu D, Gibson M, Cifrese J, Clifton SW, Newton KJ. 2007. Comparisons among two fertile and three male-sterile mitochondrial genomes of maize. *Genetics* **177**(2):1173–1192 DOI 10.1534/genetics.107.073312.

Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. 2004. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20**(3):407–415 DOI 10.1093/bioinformatics/btg427.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* **215**(3):403–410 DOI 10.1016/S0022-2836(05)80360-2.

Alverson AJ, Wei XX, Rice DW, Stern DB, Barry K, Palmer JD. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* **27**(6):1436–1448 DOI 10.1093/molbev/msq029.

Antonovics J, Bradshaw AD, Turner RG. 1971. Heavy metal tolerance in plants. *Advances in Ecological Research* **7**:1–85 DOI 10.1016/S0065-2504(08)60202-0.

Antreich S, Sassmann S, Lang I. 2016. Limited accumulation of copper in heavy metal adapted mosses. *Plant Physiology and Biochemistry* **101**:141–148 DOI 10.1016/j.plaphy.2016.02.005.

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of Computational Biology* **19**(5):455–477 DOI 10.1089/cmb.2012.0021.

Bell NE, Boore JL, Mishler BD, Hyvönen J. 2014. Organellar genomes of the four-toothed moss, *Tetraphis pellucida*. *BMC Genomics* **15**(1):383 DOI 10.1186/1471-2164-15-383.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**(15):2114–2120 DOI 10.1093/bioinformatics/btu170.

Bonen L, Williams K, Bird S, Wood C. 1994. The NADH dehydrogenase subunit 7 gene is interrupted by four group II introns in the wheat mitochondrial genome. *Molecular and General Genetics MGG* **244**(1):81–89 DOI 10.1007/BF00280190.

Brown DH. 1982. Mineral nutrition. In: Smith AJE, ed. *Bryophyte Ecology.* London: Chapman and Hall, 383–444, 511.

Chang Y, Graham SW. 2013. Patterns of clade support across the major lineages of moss phylogeny. *Cladistics* **30**(6):590–606 DOI 10.1111/cla.12066.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America* **101**(51):17741–17746 DOI 10.1073/pnas.0408302101.

Chopra RN, Kumra PK. 1988. *Biology of Bryophytes.* New York: Wiley & Sons.

Grant JR, Stothard P. 2008. The CGView server: a comparative genomics tool for circular genomes. *Nucleic Acids Research* **36**:W181–W184 DOI 10.1093/nar/gkn179.

Groth-Malonek M, Wahrmund U, Polsakiewicz M, Knoop V. 2007. Evolution of a pseudogene: exclusive survival of a functional mitochondrial nad7 gene supports Haplomitrium as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Molecular Biology and Evolution* **24**(4):1068–1074 DOI 10.1093/molbev/msm026.

Guerra J, Jiménez-Martínez JF, Cano MJ, Jiménez-Fernández JA. 2011. A contribution to the phylogenetic study of Mielichhoferiaceae–Mniaceae (Bryophyta) based on molecular sequence data. *Nova Hedwigia* **93**(1):47–56 DOI 10.1127/0029-5035/2011/0093-0047.

Guo W, Zhu A, Fan W, Mower JP. 2017. Complete mitochondrial genomes from the ferns *Ophioglossum californicum* and *Psilotum nudum* are highly repetitive with the largest organellar introns. *New Phytologist* **213**(**1**):391–403 DOI 10.1111/nph.14135.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.

Hill MO, Bell N, Bruggeman-Nannenga MA, Brugués M, Cano MJ, Enroth J, Flatberg KI, Frahm J-P, Gallego MT, Garilleti R, Guerra J, Hedenäs L, Holyoak DT, Hyvönen J, Ignatov MS, Lara F, Mazimpaka V, Muñoz J, Söderström L. 2006. An annotated checklist of the mosses of Europe and Macaronesia. *Journal of Bryology* **28**(**3**):198–267 DOI 10.1179/174328206X119998.

Ignatova EA, Ivanova EI, Ivanov OV, Ignatov MS. 2011. Mosses of the Mus-Khaya mountain (Yakutia, Asiatic Russia). *Arctoa* **20**(**1**):211–226 DOI 10.15298/arctoa.20.17.

Ishii T, Mori N, Ogihara Y. 2001. Evaluation of allelic diversity at chloroplast microsatellite loci among common wheat and its ancestral species. *Theoretical and Applied Genetics* **103**(**6–7**):896–904 DOI 10.1007/s001220100715.

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

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

Kerscher S, Dröse S, Zickermann V, Brandt U. 2008. The three families of respiratory NADH dehydrogenases. In: Schäfer G, Penefsky HS, eds. *Bioenergetics. Results and Problems in Cell Differentiation.* Vol. 45. Berlin: Springer, 185–222.

Kobayashi Y, Knoop V, Fukuzawa H, Brennicke A, Ohyama K. 1997. Interorganellar gene transfer in bryophytes: the functional nad7 gene is nuclear encoded in *Marchantia polymorpha. Molecular and General Genetics MGG* **265**(**5**):589–592 DOI 10.1007/pl00008616.

Kumar M, Kapil A, Shanker A. 2014. MitoSatPlant: mitochondrial microsatellites database of viridiplantae. *Mitochondrion* **19**:334–337 DOI 10.1016/j.mito.2014.02.002.

Li Y-C, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: structure, function, and evolution. *Molecular Biology and Evolution* **21**(**6**):991–1007 DOI 10.1093/molbev/msh073.

Li L, Wang B, Liu Y, Qiu Y-L. 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. *Journal of Molecular Evolution* **68**(**6**):665–678 DOI 10.1007/s00239-009-9240-7.

Lingmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology* **10**(**3**):R25 DOI 10.1186/gb-2009-10-3-r25.

Liu Y, Cox CJ, Wang W, Goffinet B. 2014. Mitochondrial phylogenomics of early land plants: mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Systematic Biology* **63**(**6**):862–878 DOI 10.1093/sysbio/syu049.

Liu Y, Medina R, Goffinet B. 2014. 350 million years of mitochondrial genome stasis in mosses, an early land plant lineage. *Molecular Biology and Evolution* **31**(**10**):2586–2591 DOI 10.1093/molbev/msu199.

Liu Y, Wang B, Cui P, Li L, Xue J-Y, Yu J, Qiu Y-L. 2012. The mitochondrial genome of the lycophyte *Huperzia squarrosa*: the most archaic form in vascular plants. *PLOS ONE* **7**(**4**):e35168 DOI 10.1371/journal.pone.0035168.

Meharg AA. 2005. Mechanisms of plant resistance to metal and metalloid ions and potential biotechnological applications. *Plant and Soil* **274(1–2)**:163–174 DOI 10.1007/s11104-004-0262-z.

Nishikawa T, Vaughan DA, Kadowaki K. 2005. Phylogenetic analysis of Oryza species, based on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial and chloroplast genomes. *Theoretical and Applied Genetics* **110(4)**:696–705 DOI 10.1007/s00122-004-1895-2.

Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution* **28(1–2)**:87–97 DOI 10.1007/bf02143500.

Persson H. 1948. On the discovery of *Merceya ligulata* in the Azores with a discussion of the so-called "copper mosses". *Revue Bryologique et Lichenologique* **17**:75–88.

Persson H. 1956. Studies in "copper mosses". *Journal of the Hattori Botanical Laboratory* **17**:1–18.

Pla M, Mathieu C, De Paepe R, Chétrit P, Vedel F. 1995. Deletion of the last two exons of the mitochondrial nad7 gene results in lack of the NAD7 polypeptide in a *Nicotiana sylvestris* CMS mutant. *Molecular and General Genetics MGG* **248(1)**:79–88 DOI 10.1007/bf02456616.

Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall Stotler B, Duff RJ, Stech M, Frey F, Quandt D, Davis CC. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America* **103(42)**:15511–15516 DOI 10.1073/pnas.0603335103.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic Biology* **61(3)**:539–542 DOI 10.1093/sysbio/sys029.

Sablok G, Padma Raju GV, Mudunuri SB, Prabha R, Singh DP, Baev V, Yahubyan G, Ralph PJ, La Porta N. 2015. ChloroMitoSSRDB 2.00: more genomes, more repeats, unifying SSRs search patterns and on-the-fly repeat detection. *Database* **2015**:bav084 DOI 10.1093/database/bav084.

Shaw AJ. 1987. Evolution of heavy metal tolerance in bryophytes II. An ecological and experimental investigation of the "copper moss", *Scopelophila cataractae* (Pottiaceae). *American Journal of Botany* **74(6)**:813–821 DOI 10.2307/2443862.

Shaw AJ. 1989. *Heavy Metal Tolerance in Plants: Evolutionary Aspects.* Boca Raton: CRC Press, 268.

Shaw AJ. 2000. Molecular phylogeography and cryptic speciation in the mosses, *Mielichhoferia elongata* and *M. mielichhoferiana* (Bryaceae). *Molecular Ecology* **9(5)**:595–608 DOI 10.1046/j.1365-294x.2000.00907.x.

Shaw AJ. 2014. Mielichhoferiaceae. In: Flora of North America Editorial Committee, eds. Vol. 28. New York and Oxford, 190. *Available at http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=20921*.

Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012. Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria with exceptionally high mutation rates. *PLOS Biology* **10(1)**:e1001241 DOI 10.1371/journal.pbio.1001241.

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Sato N. 2007. The mitochondrial genome of the moss *Physcomitrella* patens sheds new light on mitochondrial evolution in land plants. *Molecular Biology and Evolution* **24(3)**:699–709 DOI 10.1093/molbev/msl198.

Goruynov et al. (2018), *PeerJ,* DOI 10.7717/peerj.4350

16/17

Tyler G. 1990. Bryophytes and heavy metals: a literature review. *Botanical Journal of the Linnean Society* **104(1–3)**:231–253 DOI 10.1111/j.1095-8339.1990.tb02220.x.

Wang X, Lu P, Luo Z. 2013. GMATo: a novel tool for the identification and analysis of microsatellites in large genomes. *Bioinformation* **9(10)**:541–544 DOI 10.6026/97320630009541.

Xue J-Y, Liu Y, Li L, Wang B, Qiu Y-L. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics* **56(1)**:53–61 DOI 10.1007/s00294-009-0279-1.

Zalapa JE, 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(2)**:193–208 DOI 10.3732/ajb.1100394.

Zhao C-X, Zhu R-L, Liu Y. 2016. Simple sequence repeats in bryophyte mitochondrial genomes. *Mitochondrial DNA Part A* **27(1)**:191–197 DOI 10.3109/19401736.2014.880889.