A peer-reviewed version of this preprint was published in PeerJ on 2 February 2018.

View the peer-reviewed version (peerj.com/articles/4350), which is the preferred citable publication unless you specifically need to cite this preprint.

Goryunov DV, Goryunova SV, Kuznetsova OI, Logacheva MD, Milyutina IA, Fedorova AV, Ignatov MS, Troitsky AV. 2018. Complete mitochondrial genome sequence of the “copper moss” *Mielichhoferia elongata* reveals independent *nad7* gene functionality loss. PeerJ 6:e4350
https://doi.org/10.7717/peerj.4350

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

Denis V Goruynov [Corresp., 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], **Aleksey V Troitsky** [Corresp. 1]

[1] Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation

[2] Institute of General Genetics Russian Academy of Science, Moscow, Russian Federation

[3] Tsitsin Main Botanical Garden Russian Academy of Science, Moscow, Russian Federation

Corresponding Authors: Denis V Goruynov, Aleksey V Troitsky
Email address: denis.goryunov@mail.ru, bobr@belozersky.msu.ru

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 its 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 *Mielichhoferia 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.

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

Denis V. Goryunov[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], Aleksey V. Troitsky[1]

[1]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow, Russian Federation

[2]Institute of General Genetics Russian Academy of Science, Moscow, Russian Federation

[3]Tsitsin Main Botanical Garden Russian Academy of Science, Moscow, Russian Federation

Corresponding authors:

Aleksey V. Troitsky[1]

e-mail bobr@belozersky.msu.ru

Denis V. Goryunov[1]

e-mail: denis.goryunov@mail.ru

ABSTRACT

The mitochondrial genome of the moss *Mielichhoferia elongata* has been sequenced and assembled with the Spades genome assembler. It consists of 100,342 base pairs and has practically the same gene set and order as reported for other known bryophyte chondriomes. It is the smallest known mitogenome among bryophytes. The genome contains 66 genes including three rRNAs, 24 tRNAs, and 39 conserved mitochondrial proteins genes. Unlike the majority of previously sequenced bryophyte mitogenomes, this mitogenome lacks a functional *nad7* gene. The phylogenetic reconstruction and scrutiny analysis of the primary structure of the *nad7* gene carried out in this study suggest an independent pseudogenization in different bryophyte lineages. Evaluation of the microsatellite (simple sequence repeat) content of the*Mielichhoferia 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 9 orders from the Bryopsida and 3 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, 2009). 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, 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 et al., 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 et al., 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 12-17, 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 *Mielichhoferia 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 sulfurr 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, 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, USA).

Library preparation and sequencing

A 500 ng sample of genomic DNA was fragmented using a Covaris S220 sonicator (Covaris, USA) and a library was prepared using TruSeq DNA sample preparation kit (Illumina, USA). The concentration of the prepared library was measured with the Qubit fluorometer (Invitrogen, USA) and qPCR and fragment length distribution was determined with Bioanalyzer 2100 (Agilent). The library was diluted to 10 pM and used for cluster generation on a cBot instrument with TruSeq PE Cluster Kit v3 reagents (Illumina, USA). Sequencing was performed on a HiSeq2000 sequencer with read length of 101 from both ends of the fragments. About 6 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 100X 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 invesigation. 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 and the liverwort *Treubia lacunose* available in GenBank (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 4 rate categories was used. Posterior probabilities (PP) for trees and parameters were saved every 1000 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 (Figure 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 3 rRNAs (*rrn18*, *rrn26*, and *rrn5*), 24 tRNAs, and 39 conserved mitochondrial proteins (15 ribosomal proteins, 4 ccm proteins, 8 nicotinamide adenine dinucleotide dehydrogenase subunits, 5 ATPase subunits, 2 succinate dehydrogenase subunits, 1 apocytochrome b, 3 cytochrome oxidase subunits, and 1 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; 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 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 exon 2 and exon 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 Figure 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 (Figure 2).

SSR analysis of the *M. elongata* mitochondrial genome

Following more stringent criteria (Zhao et al., 2016) of perfect SSR locus identification (minimal number of repeating units ≥10 for mononucleotides, ≥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 and Figure 3). Most microsatellites refer to mono- and dinucleotides

186 classes (35 and 28 loci, respectively). Trinucleotides are the least frequent SSRs group in the
187 genome (one locus). No hexanucleotide microsatellite repeats occur in the genome. Among all
188 the SSRs, 87.67% are composed only of A/T bases. The total length of the SSR loci is 852 bp,
189 which comprises approximately 0.85% of the genome length.

190 **Phylogenetic analysis**

191 The alignment of 33 mitochondrial protein CDS of 40 moss taxa and hepatic *Treubia*
192 *lacunosa* (Haplomitriopsida, Treubiidae, Treubiales, Treubiaceae) consists of 24,827 positions.
193 The Bayesian phylogenetic tree inferred from this data with the hepatic *T. lacunosa* as an
194 outgroup is shown in Figure 4. Most nodes of the tree have very high PP supports. Two
195 exceptions are two nodes among the Orthotrichaceae.
196

197 ## DISCUSSION

198 We performed sequencing and analysis of the MG of *M. elongata,* a rare "copper moss" with
199 an ambiguous taxonomic status. The-MG size significantly varies even among closely related
200 flowering plants (Allen et al., 2007; Alverson et al., 2010; Cho et al., 2004; Sloan et al., 2010;
201 2012), but it is extremely stable in bryophytes (Liu, Medina & Goffinet, 2014). The MG of *M.*
202 *elongata* is 383 bp smaller than the genome of *B. aphylla* (Liu, Medina & Goffinet, 2014), which
203 to date is the smallest MG among bryophytes. However, the MG of *M. elongata* contains the
204 same set of genes and a similar genome structure to that of other mosses. The only difference is a
205 pseudogenization of the *nad7* gene.

206 This locus encodes subunit 7 of NADH dehydrogenase (NDH-1 or complex I of the
207 mitochondrial electron transfer chain) is located on the inner mitochondrial membrane and plays
208 an important role in oxidative phosphorylation process (Bonen et al., 1994). NDH-1 is a quite
209 complicated protein complex, consisting of approximately 30-40 subunits (Kerscher et al., 2008).
210 The majority of the subunits are encoded in nuclear genome, but several proteins of the complex

211 are specified by mitochondrial genes (Bonen et al., 1994).

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

220 Intact *nad7* genes were found in the MGs of different angiosperms clades (Adams & Palmer,
221 2003) and in representatives of hornworts, lycophytes, ferns and gymnosperms (Guo et al., 2017;
222 Li et al., 2009; Xue et al., 2010). However several exceptions were noted in different
223 evolutionary lineages. Therefore, the absence of a functional *nad7* gene was noted in the MG of
224 *Nicotiana sylvestris* cytoplasmic male sterile (CMS) mutants (Pla et al., 1995) and in the
225 lycophyte *Huperzia squarrosa* (Liu et al., 2012). In the liverwort *Marchantia polymorpha,* a
226 functional *nad7* gene was transferred from the MG to nucleus, but the pseudogene was preserved
227 in the MG (Kobayashi et al., 1997). Pseudogenization of *nad7* was observed in 11 other
228 liverwort groups, whereas the intact gene was found in *Haplomitrium mnioides* MG (Groth-
229 Malonek et al., 2007). This discovery suggested a basal placement of the taxon among liverworts.
230 Overall, pseudogenization of *nad7* may have occurred independently in different unrelated
231 lineages of embryophytes.

232 A total of 73 simple sequence repeats (SSRs, microsatellites) loci were identified in the MG
233 of *M. elongata*. SSRs are common in plant and animal genomes and could play an important role
234 in gene functioning (Li et al., 2004). Besides the occurrence of the SSR loci in nuclear genomes,
235 microsatellite repeats are present in plastids and MGs as well (Kumar, Kapil, & Shanker, 2014;
236 Sablok, 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.

M. 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 Figure 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, Sean & Graham (2013), and based on 41 concatenated mitochondrial protein-coding genes from19 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

We thank reviewers and the editors for their insightful suggestions and comments on the paper. We are grateful to Paul Stothard for assistance in genome map preparation.

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*

291 *Ecological Research* **7**:1-85 DOI https://doi.org/10.1016/S0065-2504(08)60202-0.
292 Antreich S, Sassmann S, Lang I. 2016. Limited accumulation of copper in heavy metal adapted
293 mosses. *Plant Physiology and Biochemistry* **101**:141-148. DOI
294 http://dx.doi.org/10.1016/j.plaphy.2016.02.005.
295 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko
296 SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA,
297 Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-
298 cell sequencing. *Journal of Computational Biology* **19**(5):455-477 DOI
299 10.1089/cmb.2012.0021.
300 Bell NE, Boore JL, Mishler BD, Hyvönen J. 2014. Organellar genomes of the four-toothed moss,
301 Tetraphis pellucida. *BMC Genomics* **15**(1):383 DOI 10.1186/1471-2164-15-383.
302 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
303 data. *Bioinformatics* **30**(15):2114-2120 DOI 10.1093/bioinformatics/btu170.
304 Bonen L, Williams K, Bird S, Wood C. 1994. The NADH dehydrogenase subunit 7 gene is
305 interrupted by four group II introns in the wheat mitochondrial genome. *Molecular and*
306 *General Genetics MGG* **244**:81-89 DOI https://doi.org/10.1007/BF00280190
307 Brown DH. 1982. Mineral nutrition. In: Smith AJE. (Ed.), *Bryophyte Ecology* 383-444. London -
308 N.-Y, Chapman and Hall pp.511. DOI 10.1007/978-94-009-5891-3.
309 Chang Y, Sean W. Graham SW. 2013. Patterns of clade support across the major lineages of
310 moss phylogeny. *Cladistics* **30**(6):590–606 DOI 10.1111/cla.12066.
311 Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily
312 elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of*
313 *Sciences of the United States of America* **101**(51):17741–17746 DOI
314 10.1073/pnas.0408302101.
315 Chopra RN, Kumra PK. 1988. *Biology of Bryophytes*. New York: Wiley & Sons.
316 Grant JR, Stothard P. 2008. The CGView Server: a comparative genomics tool for circular
317 genomes. *Nucleic Acids Research* **36**:W181-W18 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 https://doi.org/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**: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. *The New Phytologist* **213:**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**:198–267 DOI http://dx.doi.org/10.1179/174328206X119998.

Ignatova EA, Ivanova EI, Ivanov OV, Ignatov MS. 2011. Mosses of the Mus-Khaya mountain (Yakutia, Asiatic Russia). *Arctoa* **20**: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,

345 Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious
346 Basic: an integrated and extendable desktop software platform for the organization and
347 analysis of sequence data. *Bioinformatics* **28**(12):1647-1649 DOI
348 10.1093/bioinformatics/bts199.
349 Kerscher S, Dröse S, Zickermann V, Brandt U. 2008. The three families of respiratory
350 NADHdehydrogenases. In: Schäfer G., Penefsky HS. (eds) Bioenergetics. Results and
351 Problems in Cell Differentiation **45**:185-222.
352 Kimura M. 1980 A simple method for estimating evolutionary rates of base substitutions through
353 comparative studies of nucleotide sequences. *Journal of Molecular Evolution*. **16**:111-120.
354 Kobayashi Y1, Knoop V, Fukuzawa H, Brennicke A, Ohyama K. 1997. Interorganellar gene
355 transfer in bryophytes: the functional nad7 gene is nuclear encoded in *Marchantia*
356 *polymorpha*. *Molecular and General Genetics MGG* **265**(5):589-592.
357
358 Kumar M, Kapil A, Shanker A. 2014. MitoSatPlant: Mitochondrial microsatellites database of
359 viridiplantae. Mitochondrion Part B November 334-337 DOI
360 https://doi.org/10.1016/j.mito.2014.02.002
361 Li Y-C, Korol AB, Fahima T, Nevo E. 2004. Microsatellites within genes: structure, function,
362 and evolution. *Molecular Biology and Evolution* **21**(6):991-1007 DOI
363 https://doi.org/10.1093/molbev/msh073.
364 Li L, Wang B, Liu Y, Qiu Y-L. 2009. The complete mitochondrial genome sequence of the
365 hornwort Megaceros aenigmaticus shows a mixed mode of conservative yet dynamic
366 evolution in early land plant mitochondrial genomes. *Journal of Molecular Evolution* **68**:665–
367 678 DOI 10.1007/s00239-009-9240-7.
368 Lingmead B, Trapnell C, Pop M, Salzberg SL. 2009.Ultrafast and memory-efficient alignment of
369 short DNA sequences to the human genome. *Genome Biology* **10**(3):R25 DOI 10.1186/gb-
370 2009-10-3-r25.
371 Liu Y, Cox CJ, Wang W, Goffinet B. 2014. Mitochondrial phylogenomics of early land plants:

372 mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias.
373 *Systematic Biology* **63**(6):862–878 DOI 10.1093/sysbio/syu049.
374 Liu Y, Medina R, Goffinet B. 2014. 350 million years of mitochondrial genome stasis in mosses,
375 an early land plant lineage. *Molecular Biology and Evolution* **31**(10):2586–2591 DOI
376 10.1093/molbev/msu199.
377 Liu Y , Wang B, Cui P, Li L, Xue J-Y, Yu J, Qiu Y-L. 2012. The Mitochondrial Genome of the
378 Lycophyte *Huperzia squarrosa*: The Most Archaic Form in Vascular Plants. *PLos One* 7(4)
379 DOI https://doi.org/10.1371/journal.pone.0035168.
380 Meharg AA. 2005 Mechanisms of plant resistance to metal and metalloid ions and potential
381 biotechnological applications. *Plant and Soil* **274**:163–174 DOI 10.1007/s11104-004-0262-z.
382 Nishikawa T, Vaughan DA, Kadowaki K. 2005. Phylogenetic analysis of Oryza species, based
383 on simple sequence repeats and their flanking nucleotide sequences from the mitochondrial
384 and chloroplast genomes. *Theoretical and Applied Genetics* **110**(4):696-705 DOI
385 10.1007/s00122-004-1895-2.
386 Palmer JD, Herbon LA. 1988. Plant mitochondrial DNA evolved rapidly in structure, but slowly
387 in sequence. *Journal of Molecular Evolution* **28**(1-2):87–97. DOI
388 http://dx.doi.org/10.1007/BF02143500.
389 Persson H. 1948. On the discovery of Merceya ligulata in the Azores with a discussion of the so-
390 called "copper mosses." *Revue Bryologique et Lichenologique*. **17**:75-88.
391 Persson H.1956. Studies in "copper mosses." *The Journal of the Hattori Botanical Laboratory*
392 **17**:1-18.
393 Pla M, Mathieu C, De Paepe R, Chétrit P, Vedel F. 1995. Deletion of the last two exons of the
394 mitochondrial nad7 gene results in lack of the NAD7 polypeptide in a *Nicotiana sylvestris*
395 CMS mutant. *Molecular and General Genetics MGG* 248(1):79-88.
396 Qiu, YL, Li L, Wang B, Chen Z, Knoop V, Groth Malonek M, Dombrovska O, Lee J, Kent L,
397 Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall Stotler B,
398 Duff RJ, Stech M, Frey F, Quandt D, Davis CC. 2006. The deepest divergences in land

399 plants inferred from phylogenomic evidence. *Proceedings of the National Academy of*
400 *Sciences of the United States of America* **103**:15511-15516 DOI 10.1073/pnas.0603335103.
401 Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L,.
402 Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: Efficient Bayesian phylogenetic inference
403 and model choice across a large model space. *Systematic Biology* **61**(3): 539–542. DOI
404 10.1093/sysbio/sys029.
405 Sablok G, Padma Raju GV, Mudunuri SB, Prabha R, Singh DP,Baev V, Yahubyan G,Ralph PJ,
406 La Porta N. 2015. ChloroMitoSSRDB 2.00: more genomes, more repeats, unifying SSRs
407 search patterns and on-the-fly repeat detection. *Database* 1-10 DOI 10.1093/database/bav084.
408 Saitou N, Nei M. 1987. The neighbor-joining method: A new method for reconstructing
409 phylogenetic trees. *Molecular Biology and Evolution* **4**:406-425.
410 Shaw AJ. 2000. Molecular phylogeography and cryptic speciation in the mosses, Mielichhoferia
411 elongata and M. mielichhoferiana (Bryaceae). *Molecular Ecology* **9**:595–608 DOI
412 10.1046/j.1365-294x.2000.00907.x.
413 Shaw AJ. 2009. Mielichhoferiaceae. In: *Flora of North America* **28:**190 DOI
414 http://www.efloras.org/florataxon.aspx?flora_id=1&taxon_id=20921.
415 Shaw J. 1987. Evolution of heavy metal tolerance in bryophytes II. An ecological and
416 experimental investigation of the "copper moss", *Scopelophila cataractae* (Pottiaceae).
417 *American Journal of Botany* **74**(6):813-821.
418 Shaw J. 1989. *Heavy Metal Tolerance in Plants: Evolutionary Aspects*. Boca Raton, Florida:
419 CRC Press. pp.268.
420 Sloan DB, Alverson AJ, Chuckalovcak JP, Wu M, McCauley DE, Palmer JD, Taylor DR. 2012.
421 Rapid evolution of enormous, multichromosomal genomes in flowering plant mitochondria
422 with exceptionally high mutation rates. *PLOS Biology* **10**(1):e1001241
423 doi:10.1371/journal.pbio.1001241
424 Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Sato N. 2007. The mitochondrial
425 genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land

plants. *Molecular Biology and Evolution* **24**:699-709 DOI 10.1093/molbev/msl198.

Tyler G. 1990. Bryophytes and heavy metals: a literature review. *Botanical Journal of the Linnean Society* **104**:231-253 DOI https://doi.org/10.1111/j.1095-8339.1990.tb02220.x.

Van de Peer Y, De Wachter R. 1994. TREECON for Windows: a software package for the construction and drawing of evolutionary trees for the Microsoft Windows environment. *Bioinformatics* **10(**5**):** 569-570. DOI https://doi.org/10.1093/bioinformatics/10.5.569.

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 https://doi.org/10.3109/19401736.2014.880889.

FIGURE LEGENDS

Figure 1. Mitogenome map of *M. elongata* (MF417767) consisting of 100,342 base pairs.

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).

453 Figure 3. Simple sequence repeat unit length distribution in *M. elongata* mitogenome. n - the
454 numbers of base pairs (n=1, 2, 3, 4, and 5) in different microsatellite classes. N - the number of
455 loci in each SSR category.
456 Figure 4. Bayesian phylogenetic tree of 40 Bryophyta species constructed for 33 mitochondrial
457 protein coding sequences. The hepatic *Treubia lacunosa* was used as an outgroup. All nodes,
458 except those indicated on the tree, have maximal posterior probability values equal to 1,0.
459 Asterisks indicate taxa with pseudogenization of *nad7*.
460 Figure S1. The *Mielichhoferia mielichhoferiana/M. elongata* evolutionary tree.
461 The phylogenetic tree based on nuclear rDNA region (5.8S rDNA-ITS 2-5‘-end of 26S rDNA).
462 The alignment consists of 440 positions. The tree reconstruction was conducted in TREECON
463 software (Van de Peer & De Wachter, 1994) using the Neighbor-Joining method (Saitou & Nei
464 1987) with 500 bootstrap replications. Bootstrap support values >50% are shown next to the
465 branches. The evolutionary distances were computed using the Kimura method (1980) with gaps
466 taken into account as it implemented in the TREECON package.
467 Figure S2. The alignment of *nad7* genes from nine moss species.
468 The alignment was the map was created by the MAFFT program with the subsequent manual
469 checking. Only generic names of species are indicated for brevity. The full species names are
470 shown in the Figure 4. Yellow columns mark exon/intron boundaries.
471

Figure 1

Mitogenome map of *Mielichhoferia elongata* (MF417767) consisting of 100,342 base pairs.

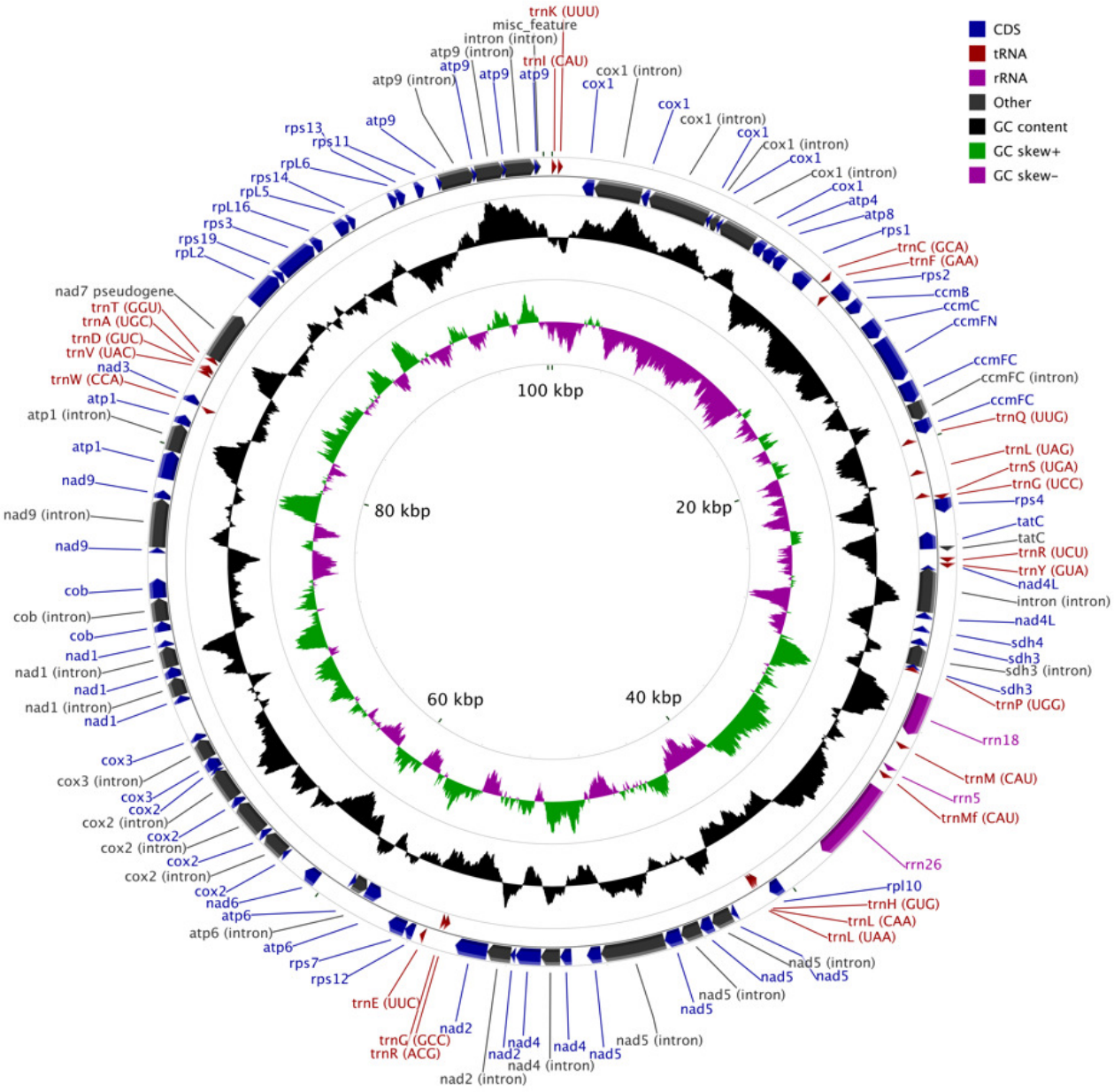

Figure 2 (on next page)

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).

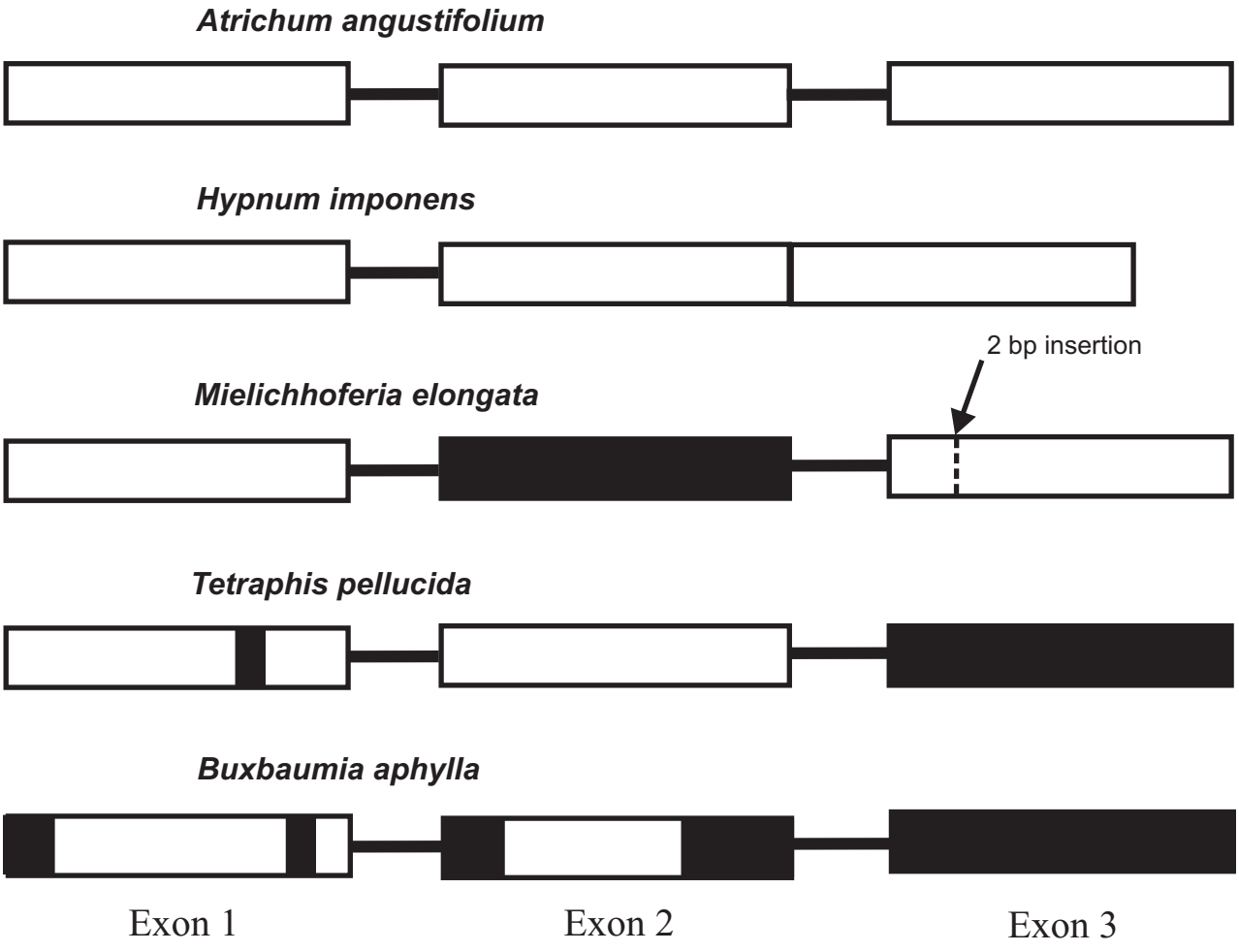
Atrichum angustifolium
Hypnum imponens
2 bp insertion
Mielichhoferia elongata
Tetraphis pellucida
Buxbaumia aphylla
Exon 1
Exon 2
Exon 3

Figure 3(on next page)

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.

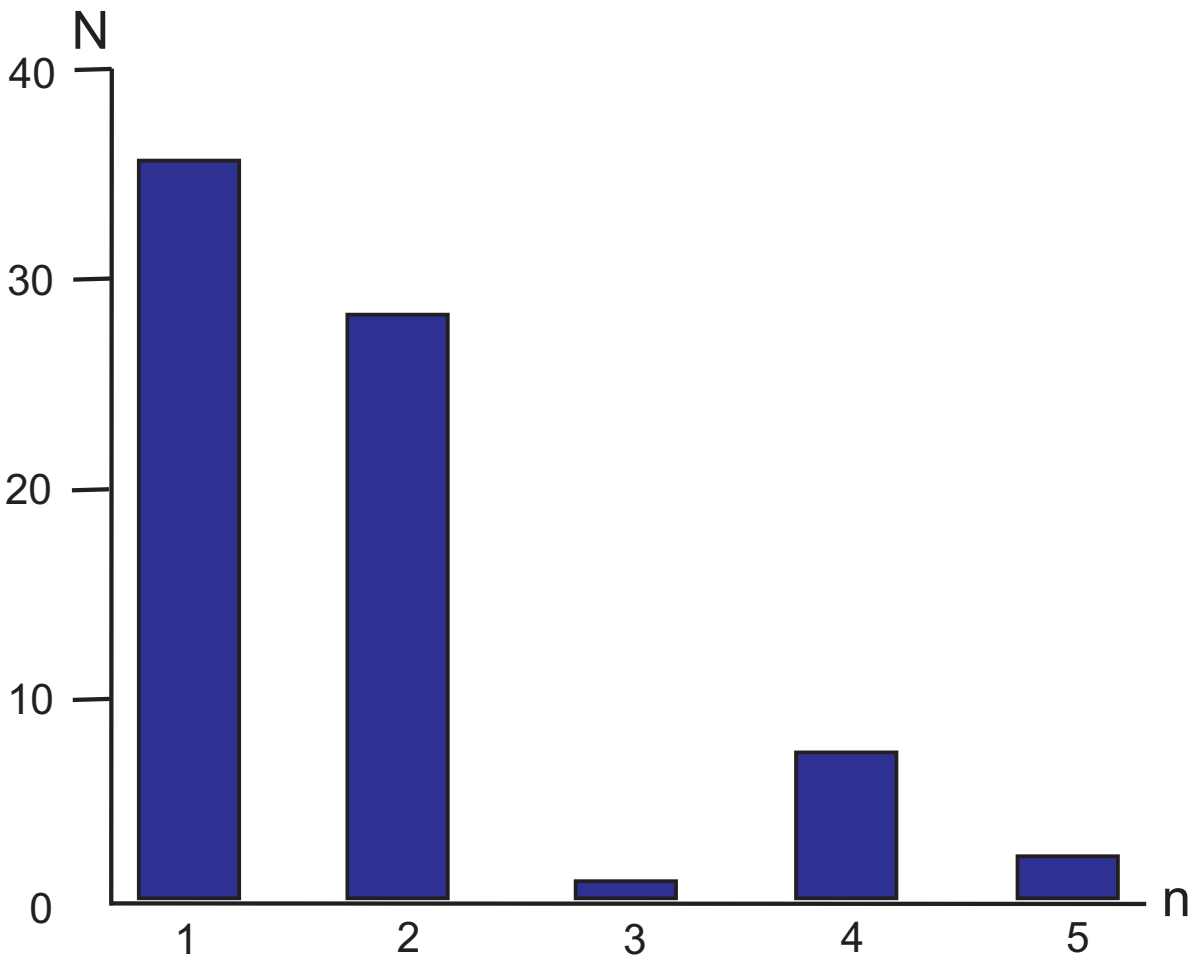
N
40
30
20
10
0
1
2
3
4
5
n

Figure 4 (on next page)

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 indicated on the tree, have maximal posterior probability values equal to 1,0. Asterisks indicate taxa with pseudogenization of *nad7*.

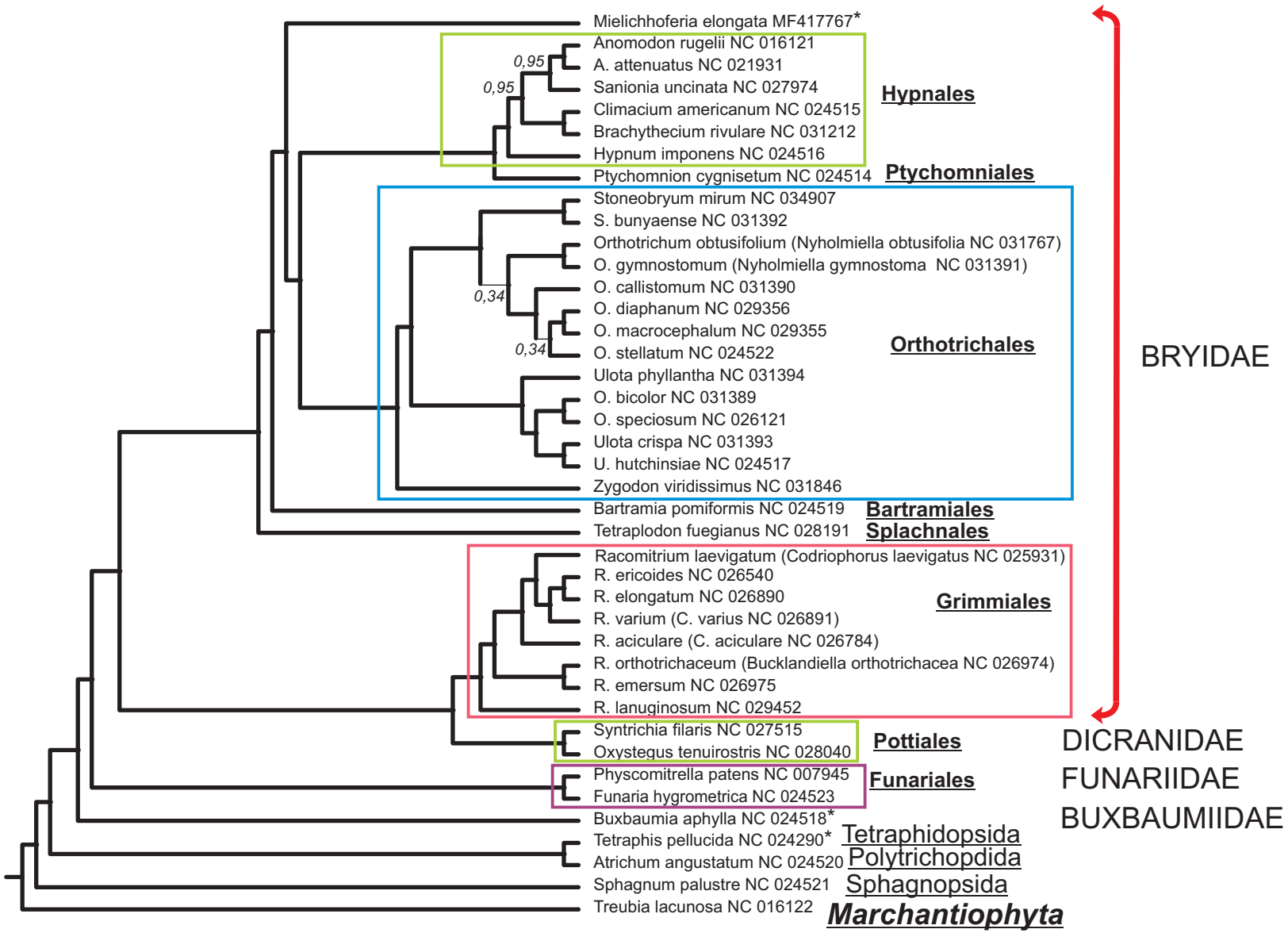
Mielichhoferia elongata MF417767*
Anomodon rugelii NC 016121
A. attenuatus NC 021931
Sanionia uncinata NC 027974
Climacium americanum NC 024515
Brachythecium rivulare NC 031212
Hypnum imponens NC 024516
0,95
0,95
Hypnales
Ptychomnion cygnisetum NC 024514
Ptychomniales
Stoneobryum mirum NC 034907
S. bunyaense NC 031392
Orthotrichum obtusifolium (Nyholmiella obtusifolia NC 031767)
O. gymnostomum (Nyholmiella gymnostoma NC 031391)
O. callistomum NC 031390
O. diaphanum NC 029356
O. macrocephalum NC 029355
O. stellatum NC 024522
0,34
0,34
Orthotrichales
Ulota phyllantha NC 031394
O. bicolor NC 031389
O. speciosum NC 026121
Ulota crispa NC 031393
U. hutchinsiae NC 024517
Zygodon viridissimus NC 031846
Bartramia pomiformis NC 024519
Bartramiales
Tetraplodon fuegianus NC 028191
Splachnales
BRYIDAE
Racomitrium laevigatum (Codriophorus laevigatus NC 025931)
R. ericoides NC 026540
R. elongatum NC 026890
R. varium (C. varius NC 026891)
R. aciculare (C. aciculare NC 026784)
R. orthotrichaceum (Bucklandiella orthotrichacea NC 026974)
R. emersum NC 026975
R. lanuginosum NC 029452
Grimmiales
Syntrichia filaris NC 027515
Oxystegus tenuirostris NC 028040
Pottiales
DICRANIDAE
Physcomitrella patens NC 007945
Funaria hygrometrica NC 024523
Funariales
FUNARIIDAE
Buxbaumia aphylla NC 024518*
BUXBAUMIIDAE
Tetraphis pellucida NC 024290*
Tetraphidopsida
Atrichum angustatum NC 024520
Polytrichopdida
Sphagnum palustre NC 024521
Sphagnopsida
Treubia lacunosa NC 016122
Marchantiophyta

Table 1(on next page)

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

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

Table 2 (on next page)

SSR-loci of *Mielichhoferia elongata* mitogenome.

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

Type of repeat unit	Motif	Repetitions	StartPos	EndPos
mono-	A	10	269	278
mono-	A	10	13526	13535
mono-	A	10	22179	22188
mono-	A	10	25861	25870
mono-	A	10	40307	40316
mono-	A	10	46592	46601
mono-	A	10	49092	49101
mono-	A	10	52217	52226
mono-	A	10	54565	54574
mono-	A	10	62618	62627
mono-	A	10	88341	88350
mono-	A	10	91128	91137
mono-	A	10	93879	93888
mono-	A	11	39390	39400
mono-	A	12	16182	16193
mono-	G	10	98368	98377
mono-	G	12	52784	52795
mono-	G	12	57418	57429
mono-	T	10	29873	29882
mono-	T	10	46703	46712
mono-	T	10	47865	47874
mono-	T	10	56400	56409
mono-	T	10	57552	57561
mono-	T	10	86976	86985
mono-	T	10	94469	94478
mono-	T	10	99146	99155
mono-	T	11	16200	16210
mono-	T	11	25885	25895
mono-	T	11	40958	40968
mono-	T	11	50459	50469
mono-	T	11	58416	58426
mono-	T	11	95793	95803
mono-	T	12	17608	17619

mono-	T	12	100200	100211
mono-	T	15	11233	11247
di-	AT	5	32938	32947
di-	AT	5	54921	54930
di-	AT	6	14278	14289
di-	AT	6	14298	14309
di-	AT	6	59230	59241
di-	AT	7	70407	70420
di-	TA	5	195	204
di-	TA	5	279	288
di-	TA	5	466	475
di-	TA	5	27730	27739
di-	TA	5	41770	41779
di-	TA	5	44628	44637
di-	TA	5	62826	62835
di-	TA	5	68954	68963
di-	TA	5	69190	69199
di-	TA	6	12557	12568
di-	TA	6	86244	86255
di-	TA	6	94457	94468
di-	TA	7	10767	10780
di-	TA	7	19813	19826
di-	TA	7	25871	25884
di-	TA	7	28304	28317
di-	TA	7	29340	29353
di-	TA	7	41786	41799
di-	TA	8	57533	57548
di-	TA	8	69397	69412
di-	TA	10	100045	100064
di-	TA	11	72289	72310
tri-	TTA	4	70696	70707
tetra-	AATA	3	54140	54151
tetra-	ATAA	3	25162	25173
tetra-	ATAG	3	10865	10876
tetra-	ATTT	3	69685	69696
tetra-	CATA	3	25129	25140
tetra-	TACC	3	76426	76437
tetra-	TAGA	3	85926	85937
penta-	AACAA	3	54704	54718

penta-	AAGAA	3	75527	75541