



OPEN The organellar genomes of Pellidae (Marchantiophyta): the evidence of cryptic speciation, conflicting phylogenies and extraordinary reduction of mitogenomes in simple thalloid liverwort lineage

Łukasz Pauksztó^{1✉}, Piotr Górski², Katarzyna Krawczyk¹, Mateusz Maździarz¹, Monika Szczecińska³, Monika Ślipiko¹ & Jakub Sawicki¹

Organellar genomes of liverworts are considered as one of the most stable among plants, with rare events of gene loss and structural rearrangements. However, not all lineages of liverworts are equally explored in the field of organellar genomics, and subclass Pellidae is one of the less known. Hybrid assembly, using both short- and long-read technologies enabled the assembly of repeat-rich mitogenomes of *Pellia* and *Apopellia* revealing extraordinary reduction of length in the latter which impacts only intergenic spacers. The mitogenomes of *Apopellia* were revealed to be the smallest among all known liverworts—109 k bp, despite retaining all introns. The study also showed the loss of one *tRNA* gene in *Apopellia* mitogenome, although it had no impact on the codon usage pattern of mitochondrial protein coding genes. Moreover, it was revealed that *Apopellia* and *Pellia* differ in codon usage by plastome CDSs, despite identical *tRNA* gene content. Molecular identification of species is especially important where traditional taxonomic methods fail, especially within Pellidae where cryptic speciation is well recognized. The simple morphology of these species and a tendency towards environmental plasticity make them complicated in identification. Application of super-barcodes, based on complete mitochondrial or plastid genomes sequences enable identification of all cryptic lineages within *Apopellia* and *Pellia* genera, however in some particular cases, mitogenomes were more efficient in species delimitation than plastomes.

The species fundamental unit in systematic biology is not strictly defined and several species concepts function in parallel in science. No less than 24 definitions of species exist¹ but the most common and popular definition of species is based on morphological differences², however not all species can be defined and described that way. Often in the taxa with lower morphological complexity, the accumulation of genetic and ecological differences is not correlated with the accumulation of morphological variation, this situation leads to the rise of cryptic species. Cryptic species are characterised by different ecological preferences, pattern of geographic distribution, distinctive genetic differences, and the absence or minor morphological variation³.

The cryptic speciation of liverworts is a well studied phenomenon, and known from all evolutionary lineages of these organisms^{4–7}. In recent years integrative taxonomy approach led to description of several species of liverworts, previously considered as cryptic lineages distinguishable only on the basis of molecular markers^{8,9}. However, most of successfully morphologically characterised and described former cryptic taxa belong to lineages with more complex morphology, whereas the lower structural complexity of simple thalloid liverworts makes macro- and microscopic delimitation more challenging. Based on morphology, liverworts have been

¹Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn, Plac Łódzki 1, 10-727 Olsztyn, Poland. ²Department of Botany, Poznań University of Life Sciences, ul. Wojska Polskiego 71C, 60-625 Poznań, Poland. ³Department of Ecology and Environmental Protection, University of Warmia and Mazury in Olsztyn, Plac Łódzki 3, 10-727 Olsztyn, Poland. ✉email: lukasz.paukszt@uwm.edu.pl

subdivided into an early divergent Haplomitriopsida, the Marchantiopsida (complex thalloids), and the Jungermanniopsida, which comprises two morphological subgroups: simple thalloids and leafy hepatics^{10,11}. The first mentioned class, Haplomitriopsida, includes leafy plants, radially symmetric, with polystratose, identical leaves, arranged in three rows. Plants classified as Marchantiopsida develop thick thallus differentiated into dorsal zones (with air-chambers and air-pores) and ventral ones with storage tissue. The third group (Jungermanniopsida) comprises both leafy and thallose plants differing in distinctive features from those mentioned above. Leafy Jungermanniopsida (Jungermanniidae subclass) have unistatose and morphologically diverse (lateral and ventral) leaves, bilateral symmetry and differ in these respects from Haplomitriopsida. Simple thalloids (Pelliidae and Metzgeriidae subclasses) develop thin thallus (mainly unistatose) without air-chambers and pores. Both of these groups share many features. The most striking is polystratose thallus without air-chambers and presence of a central strand of dead, water conducting cells (known also as midrib). An important difference is the location of the gametangia. Gynoecia in Pelliidae are usually anacrogynous, on dorsal surface of midrib or stem (here are also antheridia), whereas gynoecia in Metzgeriidae are acrogynous, on abbreviated lateral or ventral branches (with antheridia). The group of Pelliidae is more taxonomically diverse (three orders and 21 genera), while the Metzgeriidae include two orders (and eight genera), but on a global scale, more species are described within the last group¹².

Subclass Pelliidae includes both thalloid and leafy forms. The order Pelliiales, which according to recent molecular phylogenies is the earliest diverging lineage within simple thalloid liverworts^{13,14}, comprises two families: Noterocladaceae and Pelliaceae. The first family is monospecific, with leafy and South American *Noteroclada confluens* Taylor^{15,16}. The second one, Pelliaceae, comprises two genera, *Apopellia* and *Pellia*, with fleshy thalli and an obscure defined midrib, with seven species restricted to cool and temperate region of Holarctic^{12,15}. The order Fossombroniales includes plants with obliquely inserted succubous leaves in two ranks (Fossombroniaceae; the most numerous group within the order), and thallose with well- or ill-defined midrib (Allisoniaceae, Calyculariaceae, Petalophyllaceae, and Makinoaceae)^{12,15,17}. The last order within Pelliidae (Pallaviciniales) comprises plants with prostrate thallus and well-defined midrib, with 1–2(4) central strands of elongate conducting cells. These thallose forms are found within four families that belong to this order: *Pallaviciniaceae* (with eight genera: *Pallavicinia*, *Jensenia*, *Podomitrium*, *Greeneothallus*, *Seppeltia*, *Symphyogyna*, *Symphyogynopsis*, *Xenothallus*), *Hymenophytaceae* (with one genus *Hymenophyton*), *Moerckiacae* (*Hattorianthus*, *Moerckia*), and *Sandeothallaceae* (with one genus *Sandeothallus*)¹². A complete exception within the Pallaviciniales is the Phyllothalliaceae family, which has leafy forms with the leaves opposite, distant to contiguous, and well defined internodes¹¹. This family includes one genus (*Phyllothallus*) with two species occurring in the Southern Hemisphere^{12,18}. The phylogenetic studies divide Pelliidae into three orders: Pelliiales, Fossombroniales and Pallaviciniales, but monophyly of subfamily and evolutionary relationships among them are still a subject of discussion¹⁹. The order Pelliiales, which according to recent molecular phylogenies is the earliest diverging lineage within simple thalloid liverworts^{13,14}, including genera like *Apopellia* and *Pellia*, which are becoming model taxa for studying of the speciation mechanisms. The study of cryptic speciation of genus *Pellia* s.l. has a long history, starting with the description of polyploid *P. borealis* at the beginning of the twentieth century²⁰. The development of the isoenzyme electrophoresis methods enabled discovery of cryptic evolutionary lineages within *Pellia* sensu lato^{21,22}, which were later confirmed by various PCR based markers^{23–25}. Recently, an integrative approach based on morphology and molecular dataset has enabled the split of the genus *Pellia* into *Apopellia* and *Pellia* sensu stricto²⁶. The former comprises three taxonomic species, *A. apicola*, *A. megaspora* and *A. endiviifolia*, with different patterns of distribution. The European population of *A. endiviifolia* is diverging into two evolutionary lineages named A (typical form) and B (water form), which, besides molecular characters, differ in types of microhabitats^{24–26}.

Despite recent advances in liverwort organellar genomics, the simply thalloid Pelliidae are poorly explored in that field. Chloroplast genome of *Apopellia endiviifolia* was among the first sequenced liverwort plastomes²⁷, but since then only plastomes of *Fossombronia cristula*, *Makinoa crispata*, *Pallavicinia lyellii*¹⁹ and water form of *A. endiviifolia*²⁸ have been sequenced. The mitogenomic resources of Pelliidae were limited to three species of Fossombroniales: *F. cristula*, *F. foveolata* and *M. crispata*^{29,30} and mitogenomes of orders Pelliiales and Pallaviciniales remain unexplored.

Liverwort plastomes, as well as other plant plastomes, exhibit two copies of inverted repeats (IRa and IRb) separated by two single copy regions: the large single copy (LSC) and the small single copy (SSC). Unlike vascular plants, where there is a wide range of variation in plastome structure at the family or even genus level^{31,32}, the gene content and gene order in the plastomes of liverworts are highly conserved across all major liverwort lineages, including complex thalloid, simple thalloid, and leafy species^{19,33–35}. This consistency has been maintained despite millions of years of evolution separating these lineages³⁶. The liverwort mitogenomes, unlike in tracheophytes, are characterised by stable structure and low number of repeats³⁷ which play main role in mitogenomic rearrangements^{38,39}.

In this study we sequenced, assembled and annotated 44 organellar genomes of Pelliidae, with special focus on Pelliiales and their cryptic species complexes. Both organellar genomes were evaluated towards their superbarcoding potential.

Material and methods

Plant material. Plant material consisted of 22 fresh samples of *Pellia* s.l. collected from different geographical regions and different types of habitats and two herbarium specimens of *Fossombronia foveolata* and *Moerckia flotoviana*. The specimen and identification researchers details are given in Supplementary Table S1.

The cryptic species of the *Pellia epiphylla* complex were identified using peroxidase isozyme staining according to Zielinski methodology²². To confirm morphological identification of *P. borealis*, the number of chromosomes were counted for each specimen of *P. epiphylla* / *P. borealis* complex as described in earlier study²². Specimens

with nine chromosomes were delimited on either N or S based on presence different alleles in peroxidase locus. The samples with 18 chromosomes and peroxidase band pattern specific for *Pellia epiphylla* N were classified as *P. borealis*. Specimens with poor band staining or without confirmed chromosome numbers were excluded from analysis. For delimitation of *Apopellia* cryptic species intron–exon splice markers (ISJ) were used²⁴ according to previously published protocol⁴⁰. Experimental research and field studies on plants (either cultivated or wild), including the collection of plant material, complied with relevant institutional, national, and international guidelines and legislation.

DNA extraction. DNA from the fresh samples was isolated by the modified CTAB procedure. The liquid nitrogen-ground thalii were thoroughly mixed with 3 ml preheated CTAB isolation buffer (2% CTAB, 100 mM Tris–HCl, pH 8.0, 20 mM EDTA, 1.4 M NaCl and 2% β -mercaptoethanol) and incubated at 55 °C for 1 h. After three chloroform extractions, the DNA was precipitated and dissolved in sterile, deionized H₂O. Total genomic DNA from herbarium specimens was extracted using ZR Plant/Seed DNA MiniPrep™ kit (Zymo Research Corp., Irvine, CA, USA) according to the manufacturer's protocol. The purity of DNA samples was assessed spectrophotometrically using Cary 60 spectrophotometer (Agilent). DNA quantity was estimated using the Qubit fluorometer and Qubit™ dsDNA BR Assay Kit (Invitrogen, Carsbad, NM, USA). DNA quality was checked by electrophoresis in 0.5% agarose gel stained with Euryx Simple Safe (Eurx, Gdańsk, Poland). The extracted DNA prior to long-read sequencing was carefully examined using TapeStation Genomic DNA kit (Agilent) and additionally cleaned using Genomic DNA Clean and Concentrator kit (Zymo, Irvine, USA).

Library preparation, sequencing and assembly. The genomic libraries were constructed with DNA TrueSeq (Illumina) and were sequenced using HiSeqX (Illumina) to generate 150 bp paired-end reads at Macrogen Inc. (Seoul, Korea) with 350 bp insert in size between paired-ends. Afterwards, sequencing reads were cleaned by removing the adaptor sequences and low quality reads with Trimmomatic v0.36⁴¹. The long-read libraries were constructed using Ligation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies) and NEBNext™ Companion Module for Oxford Nanopore Technologies™ Ligation Sequencing (New England Biolabs) according to manufacturer's protocol and sequenced using MinION MK1B portable device (ONT) and R.9.4.1 Flow Cell (ONT). The Flow Cell was prepared for sequencing with Flow Cell Priming Kit EXP-FLP002 (ONT). Sequence reads were basecalled using high-accuracy (HAC) mode of Guppy 6.0.7 basecaller.

The previously sequenced Pellidae genomes, including plastome of *Apopellia endiviifolia* (NC_019628) and mitogenomes of *Fossombronia cristula* (MK645818) and *Makinoa crispata* (MK230958) couldn't be used as direct references due to low sequence similarity, especially considering *Pellia* sensu stricto genus. Therefore to assemble *Pellia* reference genomes we applied previously published approaches⁴². To assemble the *Fossombronia crispula* plastome SSR8246023 reads archive was used. The specimen details including GenBank accession numbers of newly assembled genomes are given in Supplementary Table S1.

The reference mitochondrial genomes for *Pellia* were assembled using Flye 2.8⁴³ with default settings to overcome long regions of short tandem repeats. The remaining *Pellia* organellar genomes were assembled using NOVOplasty 4.3.1⁴⁴. The GeSeq software was used for annotation of organellar genomes⁴⁵ and genome maps were drawn using OGDRAW web server⁴⁶. The junctions between chloroplast inverted repeats and single copy regions were visualised using IRscope software⁴⁷.

Complete chloroplast and mitogenome genome sequences of 75 specimens, including 17 from Pellidae were used for phylogenetic analysis. Organellar genomes were initially aligned using MAFFT and ambiguously aligned regions were trimmed by Gblocks 0.91⁴⁸. However, due to significant evolutionary divergence, reliable alignment wasn't possible for most plastid intergenic and intronic regions, therefore for the main phylogenetic analyses only protein-coding genes (PCG) sequences were selected. The mitogenome dataset was also reduced to PCG, due to the loss of most intergenic spacers in genus *Apopellia* and short tandem repeat richness of *Pellia* intergenic spacers.

Phylogenetic analysis was carried out using the maximum likelihood (ML). The optimal model for the plastome dataset was identified as GTR + F + I + G4 by ModelFinder based on the Bayesian information criterion (BIC). ML analysis was performed using IQ-tree⁴⁹. Mitochondrion and plastome trees were compared using cophylo function of phytools 1.0-3 R package⁵⁰.

In order to explore plastome and mitogenome gene-tree conflict across liverworts phylogeny the single gene trees were calculated for each plastid and mitochondrial PCG.

The consistency among single genes was evaluated by the maximum-likelihood (ML) analysis for every locus that contained parsimony informative characters using the RAxML v7.2.3 plugin⁵¹ for Geneious Prime 2022 with the GTR + I + G model and bootstraps estimated using 1000 replicates. Potential conflicts among genes were visualised by constructing a supernetwork using SplitsTree4⁵² according to the approach given by Liu et al.³⁷, except the number of replicates was increased to 1000 and the software used for creating consensus trees was Geneious Prime 2022. The supernetworks were created separately for chloroplast and mitochondrial datasets.

Genetic variation, barcoding and molecular species delimitation. Comparative analysis of chloroplast genomes was carried out in the Spider 1.1-5 program⁵³ based on inter- and intraspecific distances that were calculated using the Kimura 2-parameter model (K2P) of nucleotide substitution. Due to reasons mentioned above *Apopellia* and *Pellia* datasets were analysed separately. Barcoding analyses of entire *Pellia*/*Apopellia* organellar genomes and their 500 bp-long fragments generated by sliding window (step by 100 bp) were made in Spider. The discrete molecular diagnostic characters (MDCs) for each species were calculated according to the Jörger and Schrödl⁵⁴ approach using FASTACHAR software⁵⁵. Nucleotide diversity analyses were carried out using the PopGenome package⁵⁶ with the window size and steps as in Spider analyses. Genetic similarity

among *Apopellia* and *Pellia* species was calculated using *simplot* function of *ggmsa* package⁵⁷, with sliding windows parameters as in previous analyses and accession numbers OL654073 (cpDNA), OQ236455 (mtDNA) and OQ280821 (cpDNA), OQ236469 (mtDNA) as references for *Apopellia* and *Pellia* respectively. Next to species delimitation analysis in *Spider*, *ASAP*: assemble species by automatic partitioning⁵⁸ was applied. This method generates ten partitions, computing for each partition the *asap*-score, which is the average of two components: probability of panmixia (p-value) and barcoding gap width. The lower the score, the better the partition for the dataset. As input data alignments of plastomes and mitogenomes were used separately in each analysis.

Codon usage analysis. The relative synonymous codon usage (RSCU) indicator was calculated based on the number of times a particular codon is observed, relative to the number of times that the codon would be observed assuming a uniform synonymous codon usage⁵⁹. The RSCU values were calculated using the *seqinr* 4.2 R package⁶⁰ and visualised using *ggplot2*⁶¹.

To calculate difference in codon usage bias between *Apopellia* and *Pellia* protein-coding genes we applied the Measure Independent of Length and Composition (MILC) with hard filtering option to remove sequences shorter than 80 AA, as implemented in *coRdon* 1.15.0 package^{62,63}. Differences in codon usage (CU) between genera were visualised using *intraBplot* function of *coRdon* package⁶³.

Results

Characteristics of Pelliidae chloroplast genomes. Newly sequenced plastid genomes of Pelliidae are circular molecules consisting of regions typical for land plants. The total length of assembled plastomes ranged from 115,878 bp in *Pellia neesiana* (Fig. 1) to 124,598 bp in *Moerckia flotoviana*. The GC content ranged from 35.8% in water form of *A. endiviifolia* to 41.6% in *F. foveolata*.

One hundred twenty-two unique genes (taking into account only one copy of inverted repeat regions) were identified in the plastomes of *Apopellia* and *Moerckia*: 81 protein-coding genes, four ribosomal RNAs, 31 transfer RNAs and six *ycf* genes of an indeterminate function. The plastomes of *Fossombronia* and *Pellia* genera lack *cysA* and *cysT* genes, additionally in the *Fossombronia*, clusters of *trnS*(GCU)-*psbI*-*trnS*(GCA) genes are missing. The gene order seems to be conserved in Pelliidae, with exception of translocation of *rps12-rps7-ndhB* to the minus strand of LSC/IR boundary.

Plastid genome variation. The nucleotide diversity (Π) values of 500 bp long frames across plastomes varies from 0 in rRNA genes of IR regions to 0.034 in *ycf12-trnQ* and *trnV-atpE* intergenic spacers (igs) (Fig. 2).

The values within IR regions were almost five times lower than within single copy LSC and SSC parts of the plastome. In the both genera the single copy regions near LSC-IRb and IRb-SSC regions were also highly variable, especially in the case of SSC, where *ndhF* gene is located. The second peak of nucleotide diversity within the SSC region was within the *ycf1* gene. However, the gene near the SSC-IRA border, *chlL*, was among less variable regions of *Pellia* and *Apopellia* plastomes.

The patterns of nucleotide diversity within *Apopellia* and *Pellia* genera usually overlap with few additional hotspots of the former found in the *rpoC2* gene and igs between *atpI-atpH*, *psbM-trnS*, *psbA-trnH* and *trnV-atpE* (Fig. 2). Lower variation in *Apopellia* than *Pellia* was revealed at 5' end of *ycf2*, across whole *accD*, *rpl22* and *chlL* genes.

The analysis of sequence similarity within the same 500 bp window revealed some regions of intraspecific variation, which was generally higher in *Pellia* than *Apopellia* (Fig. 2). In the case of the latter slight variation at cryptic species level was found within the *ndhF* gene and the *trnG*(UCC)-*ycf12* spacer of the typical form. The mentioned spacer was also variable within *P. epiphylla* N and S. These two species shared intraspecific variation within few other plastome regions including *ndhB-ycf66*, *atpH-atpE*, *ycf2-cysA*, *ndhC-trnV* and *ndhH-ycf1b*. Additionally, plastomes of *P. epiphylla* S revealed variation in the *ycf2* gene, *ycf3-rps4* region and between *rrn12* and *rrn23* genes of IR.

Characteristics of Pelliidae mitochondrial genome. The newly assembled mitochondrial genomes have almost identical gene content and order, as previously published (MK230958, MK749462 and MK230936) with exception of the *trnR*-UCG loss in the *Apopellia* genus (Fig. 3a). The mitogenomes of Pelliidae ranged from 108,928 bp in typical form of *Apopellia endiviifolia* to 179,862 bp in *Moerckia flotoviana* (Table S1), but the length of the molecule doesn't correlate with intron content (Fig. 3b).

The species of *Fossombroniales* order lack both introns of the *atp1* gene, and in the genus *Fossombronia* lost two introns of *cox1*, being the second and third largest mitogenomes of Pelliidae (179,676 bp and 175,222 bp respectively). In contrast, over 70 kbp smaller mitogenomes of *Apopellia* contain complete introns set. The reduction of *Apopellia* mitogenome in comparison to the closest related *Pellia* is due to the loss of nucleotides in intergenic regions (Fig. 4). The intergenic spacers like *atp4-trnMf*, *trnV-trnD*, *sdh4-nad4L*, *nad4L-tatC* were over 80% shorter, while in the most of igs reduction fall within 40–60% range. Among the larger igs the smallest reduction, by 28.4% were in the *atp1-cox1*. Seven very short intergenic spacers were longer in *Apopellia* than in *Pellia* and in the one longer between *cob-nad9* genes with the difference lower than 5%. Despite extraordinary reduction of igs, the identity between genes remains high, mostly above 90% (Fig. 4).

Mitochondrial genome variation. The nucleotide diversity (Π) values of 500 bp-long frames across mitogenomes ranged from 0 in frames located along protein-coding genes (like *nad4* and *nad5*) to 0.11 in *Apopellia trnL*(CAA)-*trnL*(UAA) igs (Fig. 5). The Π values were generally lower in *Pellia*, with the highest peak (0.048) revealed within *ccmB-ccmC* igs. Beside this spacer, only five regions of *Pellia* mitogenome had values exceeding 0.04 including the following igs: *nad2-trnR*(UCU), *trnY*(GUA)-*trnR*(ACG), *trnN*(GUU)-*trnP*(UGG),



Figure 2. Comparison of diagnostic nucleotides, sequences similarity and nucleotide diversity between plastome of *Apopellia* and *Peltia* genera together with intragenetic specificity of *Peltia* (spring, water and typical) and *Apopellia* (N, S, and neesiana) specimen.

genera of Pallaviciniales. The order Pelliiales further splits into two clades corresponding to genera *Apopellia* and *Peltia*. Within the former, all cryptic lineages including typical, water and spring forms were resolved as monophyletic clades with maximum bootstrap support. The water habitat lineages (water and spring) form a common clade sister to the typical form. The second clade of Pelliiales were formed by species of the genus *Peltia*, revealing *P. neesiana* as early divergent within this group. Cryptic species of *P. epiphylla* formed monophyletic, maximally supported groups with a common clade. Allopolyploid species *P. borealis* was included in *P. epiphylla* N clade.

The concentrated mitochondrial gene dataset consists of 45,852 nucleotides including 18,257 variable and 13,051 parsimony informative. The mitochondrial dataset generated a tree with similar, but not identical topology as the plastid dataset. Analysis of mitochondrial genes didn't support monophyly of Pelliidae and resolved Pelliiales as earliest divergent among Jungermanniopsida, however support for common Fossombroniales/Pallaviciniales—Jungermanniidae clade wasn't maximal (BS 91%), while the plastid dataset supported monophyly

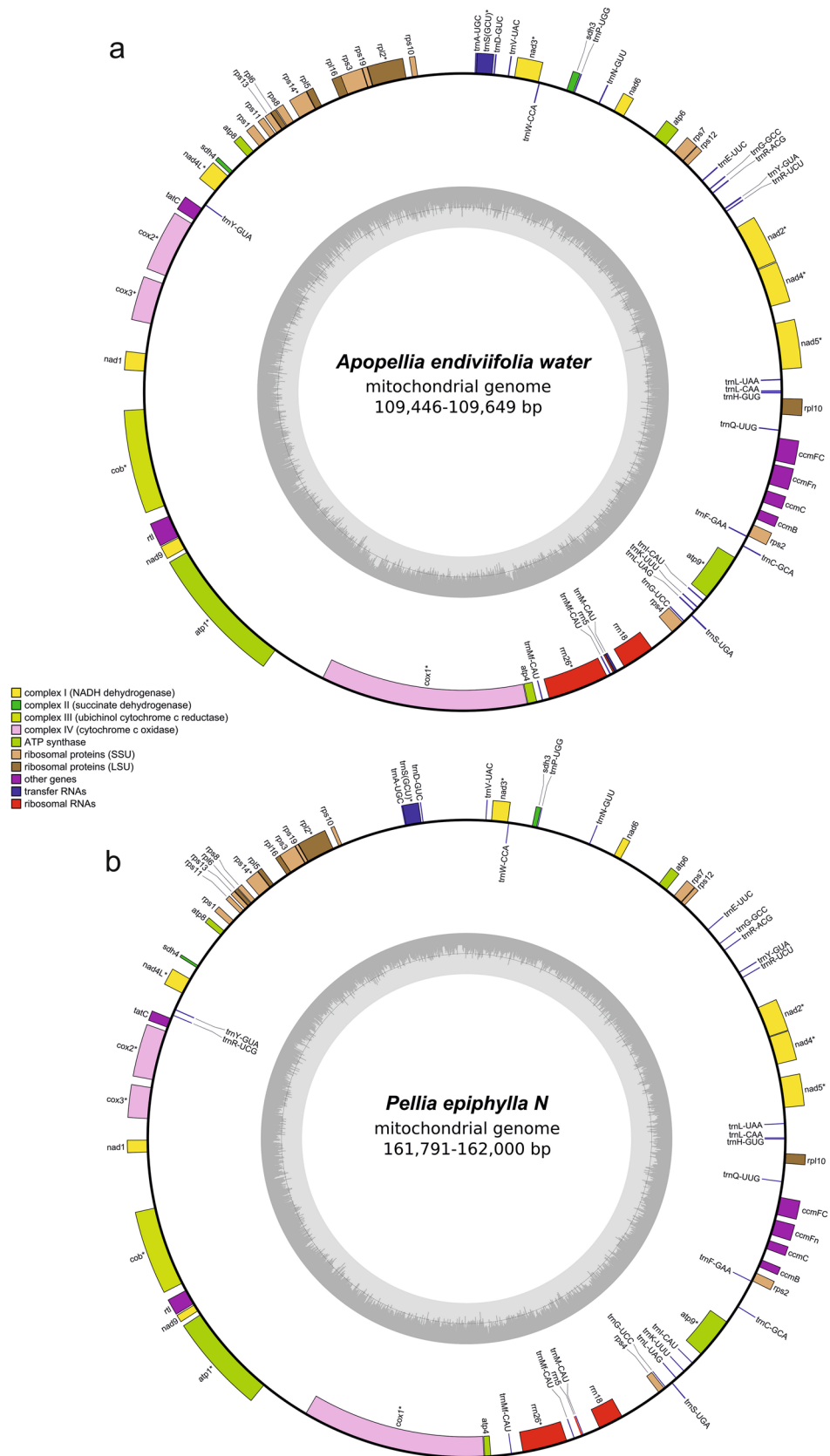


Figure 3. Mitochondrial genomes of (a) *Apopellia endiviifolia* water form and (b) *Pellia epiphylla* N. Genes inside and outside the outer circle are transcribed in counterclockwise and clockwise direction, respectively. The genes are colour-coded based on their function. The inner circle visualises the GC content. * indicates genes containing introns.

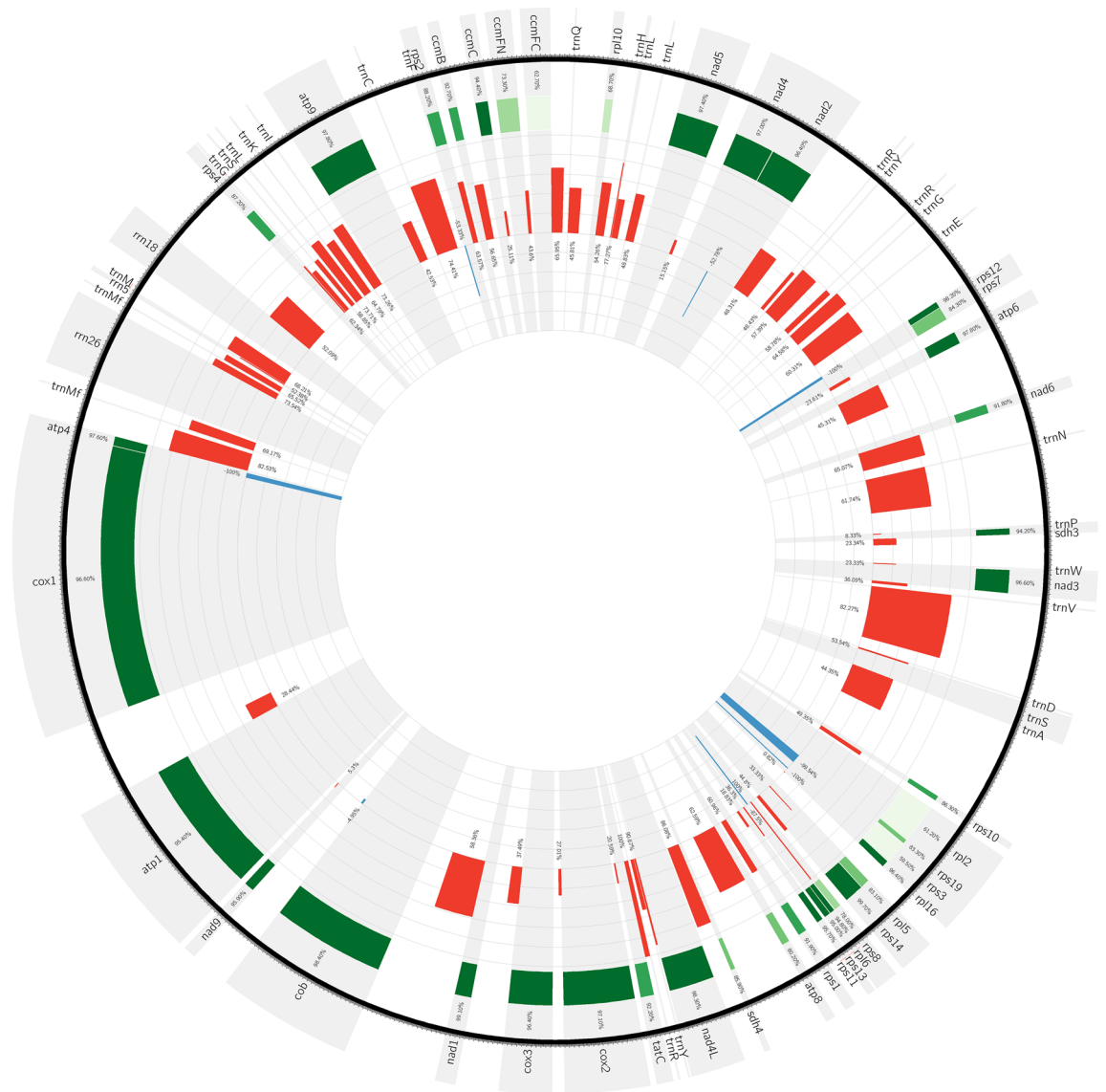


Figure 4. Circular visualisation of difference in spacers length between *Pellia* and *Apopellia*. Red and blue histograms represent percentage differences in intergenic spacers sequences. Height of red bars show increasing ratio length for *Pellia* versus *Apopellia* and width indicate direct dissimilarity between sequences length of both species. The blue bars show *Apopellia* advantage in length. The green heatmap depicts pairwise alignment of each coding region between both species.

of Pellidae with maximum value. Another significant difference was revealed within suborder Lophocoleineae where species of *Herbertus* were placed as the earliest divergent. (Fig. 5). Minor differences comprise relationships within complex thaloid liverworts: (1) the mitochondrial dataset placed *Asterella* and *Dumortiera* within common clade (but poorly supported—BS 69%); (2) relationships at intraspecific level within *Pellia neesiana* and *P. epiphylla* N.

The supernetworks (Supplementary Fig. 1) based on single-gene trees of the mitochondrial and plastid dataset didn't reveal significant phylogenetic conflicts among genes, especially during diversification of the major liverwort lineages (Fig. 6). The competing splits within Pellidae appeared only in the *Pellia* clade, where some minor incongruences in the mitochondrial dataset were detected among *Pellia epiphylla* N and *P. neesiana* lineages.

Codon usage differentiation of *Apopellia* and *Pellia*. Comparative codon usage analysis revealed a difference between *Apopellia* and *Pellia* plastome genes, despite sharing by these genera identical sets of plastid *trnS* (Fig. 7). However, in the case of mitochondrial protein-coding genes, the CU patterns of both genera were similar and the loss of one *trnR*(*ucg*) gene by *Apopellia* mitogenomes didn't impact CU. The differences in codon usage among *Apopellia* and *Pellia* aren't equally distributed among amino acids (Supplementary Fig. 2, Fig. 7). In the *Apopellia* plastomes most common stop codon 'taa' is preferred over 'tag', while in *Pellia* all stop codons have similar frequency. Patterns of CU for amino acids like tyrosine, glutamine, glutamate, asparagine, aspartate and



Figure 5. Comparison of diagnostic nucleotides, sequences similarity and nucleotide diversity between mitogenome of *Apopellia* and *Peltia* genera together with intrageneric specificity of *Peltia* (spring, water and typical) and *Apopellia* (N, S, and nessiana) specimen. Sliding window analyses of *Apopellia* and *Peltia* mitogenomes is equal to 500 bp.

cysteine were similar in both genera. Where amino acids were coded by more than two codons, the differences in CU usually comprised two of them, like in the case of alanine, where ‘gct’ codon is more frequent than ‘gcc’ in *Apopellia*, while the opposite situation was revealed in *Peltia*.

The general pattern of CU in the mitogenome and plastome was similar (Fig. 7) and as mentioned above, without clear differences between studied genera. Within *Apopellia* differences were found in ‘water’ form, in which serotonine ‘tcg’ and ‘tct’ codon frequencies are different than in remaining species. This kind of variation

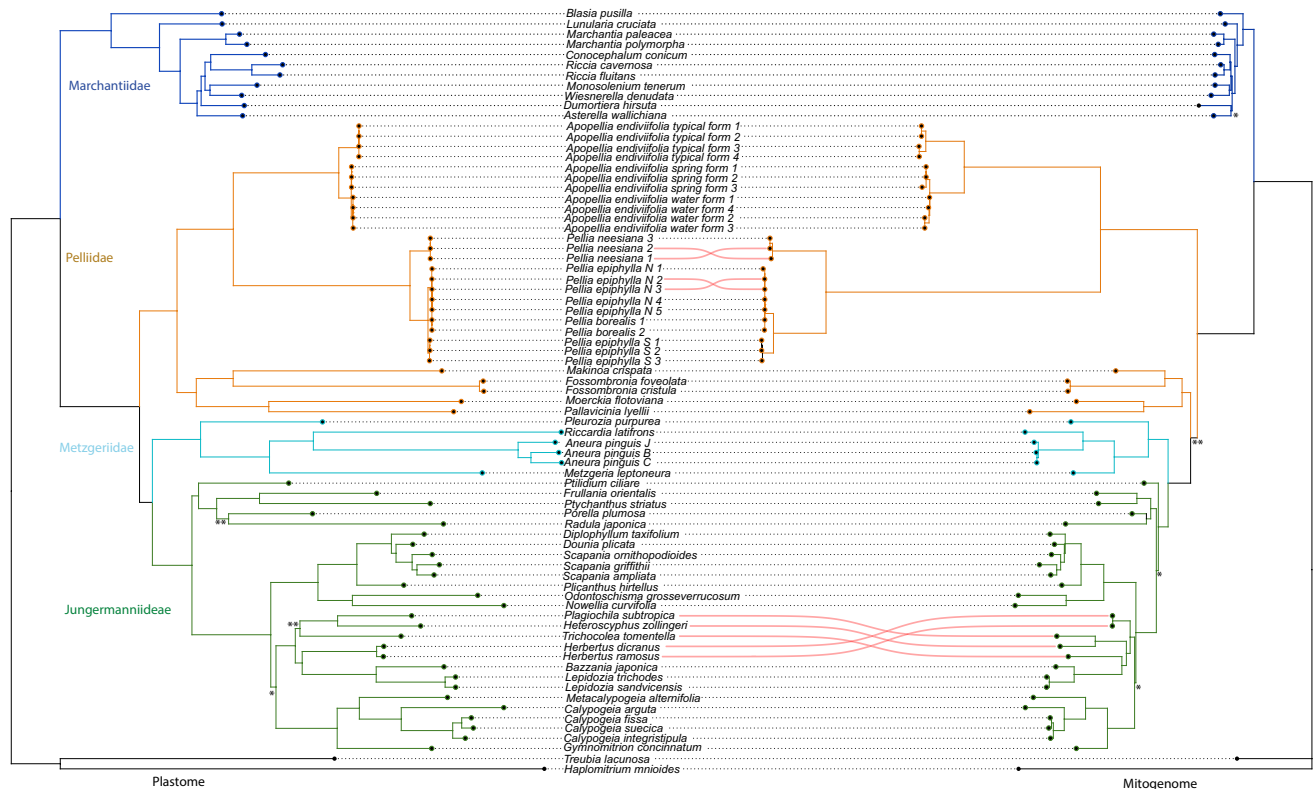


Figure 6. Phylogenetic topology inferred from concentrated plastid (left) and mitochondrial dataset (right) using ML method. All clades are maximally supported unless indicated otherwise with double asterisk (for BS values within 85–99 range) or single asterisk (for BS values below 85). Clade positions were optimised using cophylo function in the phylotools R package.

was also found in CU of leucine and arginine amino acids, where ‘water’ form of *A. endiviifolia* revealed different CU than *Pellia* and *A. endiviifolia* typical and spring forms.

Barcoding and MDCs. The number of detected plastid MDCs in *Apopellia* and *Pellia* genera varied from 0 to 4717 (Supplementary Table 2). No MDCs were identified for allopolyploid *Pellia borealis*, as two analysed specimens fell within variation of *Pellia epiphylla* N. Therefore plastomes of *P. borealis* were excluded from barcoding and molecular delimitation analyses. Among other species, the lowest number of MDCs (136) was detected for the spring form of *A. endiviifolia* followed by *P. epiphylla* S with and *P. epiphylla* N with 235 and 393 MDCs respectively. The highest numbers of MDCs were revealed for *P. neesiana* (4717) and for typical form of *A. endiviifolia* (3161). The impact of multiple specimen sampling on the number of detected MDCs, with the exception of the *P. epiphylla* complex, was minimal, below 1% decrease in MDCs. In the case of *P. epiphylla* S the number of MDCs decreased by 31% (mean for individuals 342 vs 235 at species level) and 15% for *P. epiphylla* N (mean for individuals 463 vs 393 at species level).

The number of detected mitochondrial MDCs in *Apopellia* and *Pellia* genera varied from four to 30,538 (Supplementary Table 2). The lowest number of MDCs was revealed for *P. borealis*, while the highest for *P. neesiana*. The cryptic species of the *P. epiphylla* complex have 2979 and 3307 MDCs for N and S respectively. Among *Apopellia endiviifolia* cryptic species the biggest number of MDCs was identified for the typical form (20,260), while the water and spring form revealed 309 and 264 MDCs respectively.

Comparison of infra- and interspecific distances based on the plastome confirmed the presence of a significant barcoding gap among all species, including cryptic ones (Supplementary Fig. 3). In the case of mitochondrial genomes, the barcoding gaps are smaller but still significant. The share of nucleotide diagnostic characters is not equally distributed among analysed species (Fig. 2).

ASAP analysis performed for data based on the chloroplast genome generated four identical partitions with low values of asap-score ranging from two to three (Fig. 8a). These four proposed divisions indicated six species among testing liverwort groups. These results fully reflect the species split among the cryptic complexes in the genera *Pellia* and *Apopellia*. Within the *Pellia* genus, cladogram indicated that *P. epiphylla* N and *P. epiphylla* S are definitely separated species (black circle, Fig. 8a). As different species, cladogram also clearly showed two forms of *Apopellia endiviifolia*: water and spring (black circle, Fig. 8a).

The mitogenome-based dataset gave slightly weaker results. Only one partition (asap-score = 7.5) proposed by ASAP was congruent with real division of *Pellia* and *Apopellia*, indicating a total of six species within these genera (Fig. 8b). On the other hand, one partition with quite low asap-score with a value of 3.5 grouped together only water and spring form of *Apopellia*. Although two ASAP-partitions put forms of *Pellia epiphylla* together,

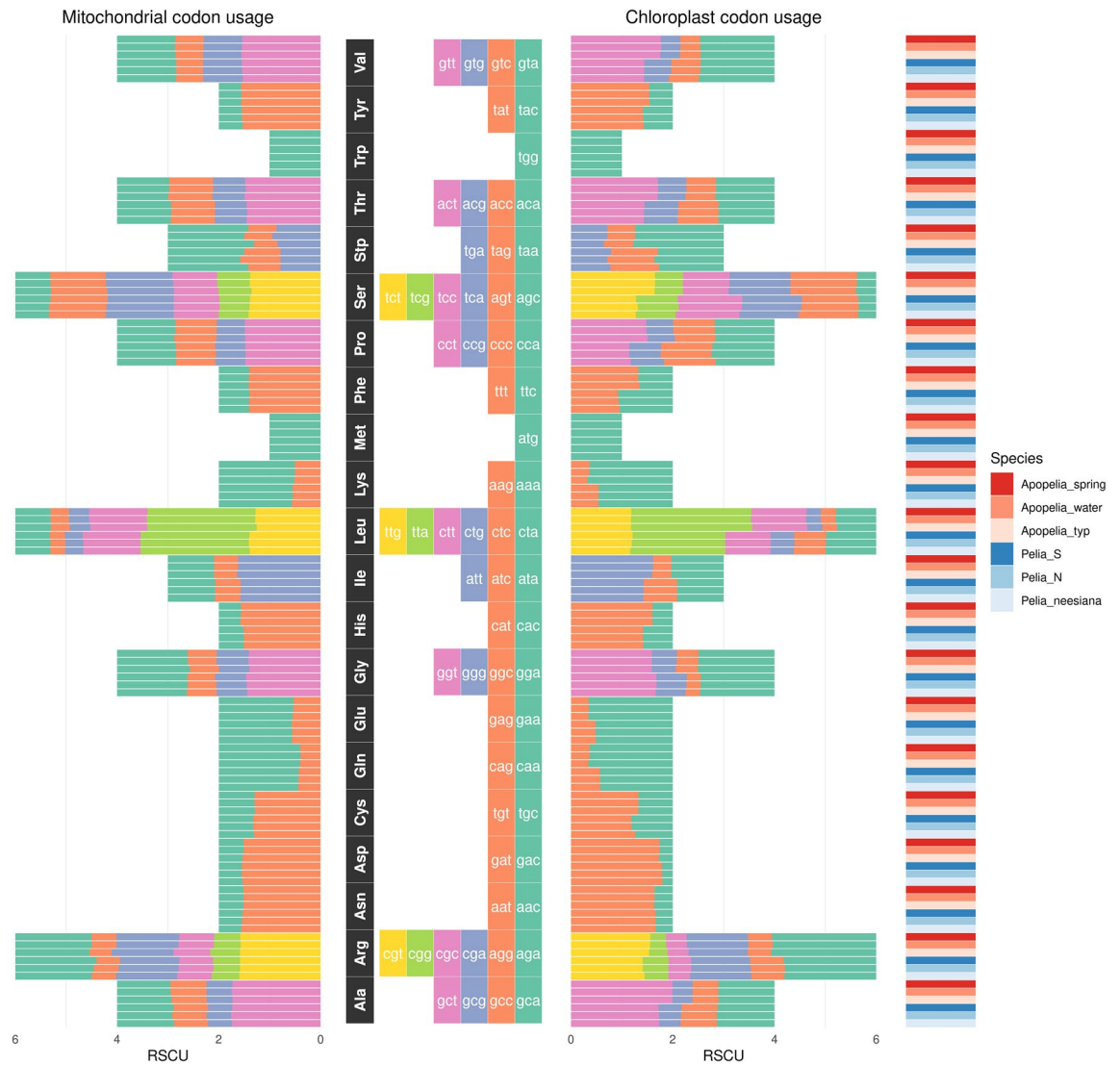


Figure 7. Codon usage of mitochondrial and chloroplast genomes for six selected species of *Pelliidae*. The main histograms (left—mitochondrial, right—chloroplast) depict RSCU values for each amino-acid's coding triplets based on MILC method.

the cladogram next to the bar chart suggests that *Pellia epiphylla* is not a homogeneous group and the division into N and S types is justified (red circle, Fig. 8b).

Discussion

Organellar genomes of *Pelliidae*. The liverwort organellar genomes are known from its conserved structure and gene content and results of *Pelliidae* analysis confirm this observation. In contrast to the chloroplast genomes of angiosperms, the structural heteroplasmy connected with SSC subunit orientation wasn't detected in *Pelliidae*, as well as in previous studies on leafy liverwort *Scapania undulata*⁶⁴. Long read sequencing identified only one plastome haplotype also in gymnosperms and pteridophytes, suggesting that alternative haplotype is specific for angiosperms⁶⁵. The gene order of *Pelliales* and *Fossombroniales* is typical for most of liverworts, but plastome structure of *Moerckia* confirms that the translocation of *rps12-rps7-ndhB* cluster seems to be characteristic for *Pallaviciniales*¹⁹.

Plastomes of the *Pellia* genus lost two genes (*cysA* and *cysT*), but gene order remains stable. These genes are present in the sister genera *Apopellia*, *Makinoa* and *Moerckia*^{19,28}, but in the remaining genera of *Pelliidae* (*Fossombronia*, *Pallavicinia*, *Pellia*) are pseudogenised, as well as in many other leafy liverworts¹⁹. Further gene losses include *trnS(GCU)-psbI-trnS(GCA)* in *Makinoa* and *Fossombronia*. However, plastome length in *Pelliidae* doesn't correlate with gene losses, while the second largest plastome of *Pallavicinia* (124,103 bp) and the smallest of *Pellia neesiana* (115,875 bp) have identical gene content, whereas plastomes of *Apopellia* with complete gene sets (120,537–120,947 bp), are smaller than *Pallavicinia* without *cysA* and *cysT* genes. Independently of gene loss, plastome reduction was also found in leafy Cephaloziineae, which plastomes ranged between 118,571 and 114,423 bp, despite identical gene sets^{28,64}. Plastid gene losses are scattered along different evolutionary lineages

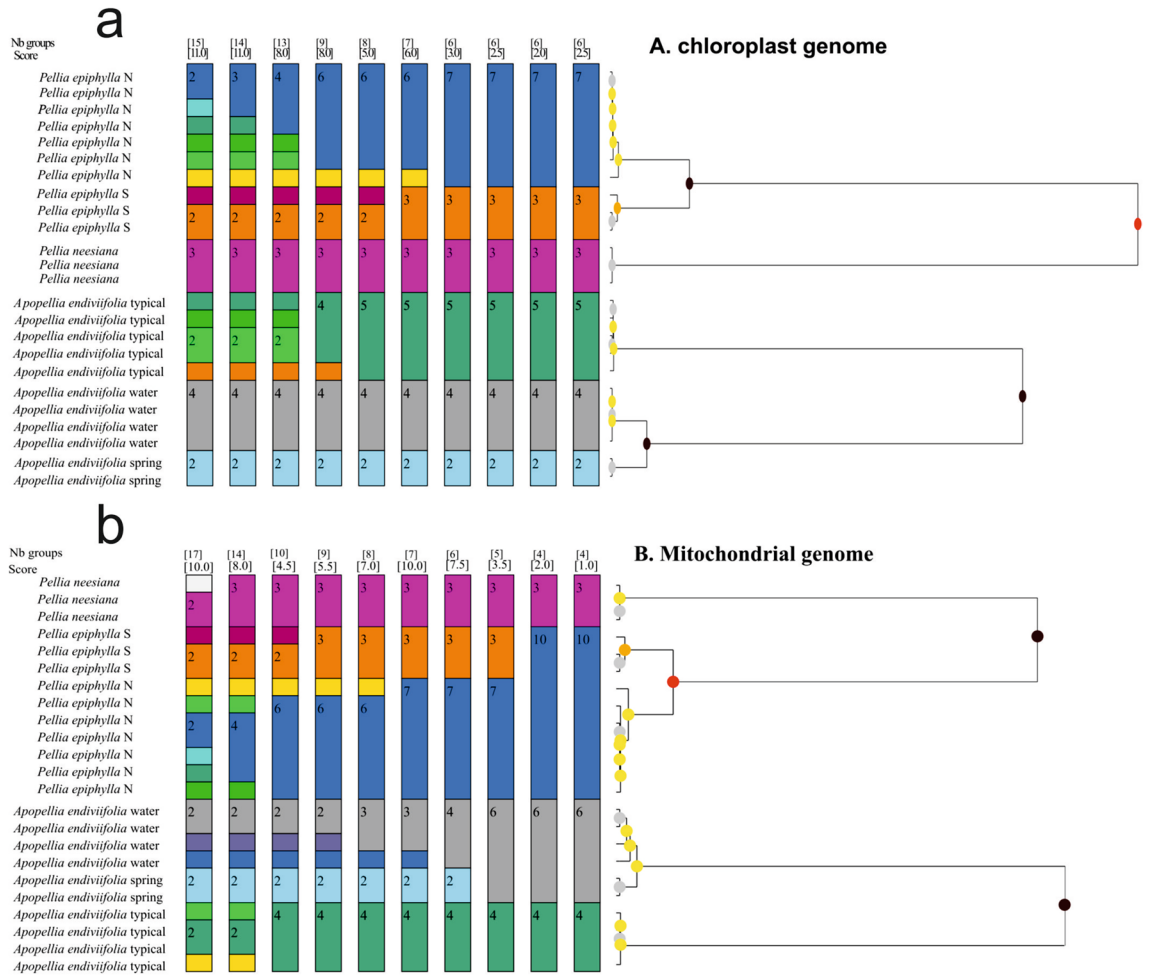


Figure 8. Species delimitation by ASAP analysis for chloroplast (a) and mitochondrial (b) genomes. The colorful fields indicate the groups of species. Every field contains the number of individuals. Above the coroful bars the coefficient asap-score (the lower value) and number of species (the upper value) recognized for the whole dataset are presented. The cladogram supports the species delimitation presented on the bar charts.

including loss of *cysA* and *cysT* genes in *Ptilidium pulcherrimum* and *Cheilelejeunea xanthocarpa*^{33,35} and additional four genes (*ndhF*, *rpl21*, *rps32* and *ccsA*) in *Cololejeunea lanciloba*⁶⁶. Single genes are also missing in *Aneura pinguis* (*ycf66*), *Haplomitrium blumei* (*psaM*) and *Schistochila macrodonta* (*ndhF*)^{66,67}. Heterotrophic *Aneura mirabilis* has a reduced plastome due to loss of six genes (*ndhA-E-G-K-I* and *ycf66*) or pseudogenization (19 genes) of genes involved in the process of photosynthesis. Two liverwort species increase the number of plastid genes by transferring them to IRs: *Haplomitrium blumei* transferred *ndhF* gene from SSC⁶⁶, while *Conocephalum salebrosum* transferred *rps7* and *rps12* from LSC⁶⁸. Minor changes in genomes are related to loss of introns: complex thalloid liverworts, with the exception of Blasiales, lost introns of *ycf3* and *rps12* genes, which phenomenon was also reported in species from distantly related genera like *Herbertus*, *Ptychantus*, *Radula* and *Metzgeria*^{19,35,68}. Single loss of introns of *ndhA* and *ndhB* genes were found in *Radula japonica* and *Treubia lacunosa*¹⁹.

In contrast to plastomes, evolutionary losses of mitochondrial introns seem to be common in liverworts, but mainly concern leafy liverwort lineages, while the intron content of thallose liverworts is stable^{29,30,69}. Out of ten intron losses only the loss of *trnL* could happen once and remaining losses were independent across different evolutionary lineages of leafy liverworts⁷⁰. Unlike previously sequenced genomes of Pelliidae (*Fossombronia cristula* and *Makinia crispata*), the mitogenomes of *Pellia* and *Apopellia* have introns in the *atp1* gene and in contrast to *Fossombronia*, keep complete *cox1* intron set^{30,70}. The gene content of liverwort mitogenomes is more stable than that of the plastomes, beside the *nad7* loss, which is present only in early divergent lineages of Haplomitriales and Treubiales and the loss of *ccmC* and *ccmFn* in *Treubia lacunosa*⁷¹. The variation of tRNA genes among liverworts comprises the loss of *trnT*-GGU after divergence Marchantiopsida and Jungermanniopsida and *trnR*-UGC after divergence of Pelliidae. The *trnR*-UGC seems to be independently lost in *Apopellia*, even considering incongruent results on monophyly or paraphyly of Pelliace^{14,66,70,72}, which doesn't question monophyly of *Apopellia*/*Pellia* clade^{14,26}. The loss of the *trnR*-UGC gene didn't impact codon usage pattern of mitochondrial protein coding genes—both genera revealed similar CU pattern (Figs. 7, 8). On the other hand, *Apopellia* and *Pellia* differ in CU by plastome CDS, despite identical tRNA gene content. The lack of correlation between codon usage and the

number of tRNA was previously reported in bacterial genomes^{73,74}. Prokaryotes seem to respond to gains or losses of tRNA genes by modulating the concentration of tRNAs rather than modifying its frequency on codon usage⁷³.

Despite the presence of some repeats, which are considered as a main factor of mitochondrial structure dynamics in vascular plants^{75,76}, the structure of liverwort mitogenomes is stable³⁷. The exceptions were found for the mitogenome of *Gymnomitrium concinnatum* belonging to Jungermanniales, where two recombination events were detected³⁴ and in *Dumortiera hirsuta* (Marchantiales) where one 96 kb inversion was found⁷⁷. The possibility of recombination was also suggested for 12 other liverworts, however most frequent variants support typical mitogenome structure²⁹.

Despite the nearly identical gene content (loss of *trnR*-UCG in *Apopellia*) and order the mitogenomes of *Pellia* and *Apopellia* differ in size. The newly sequenced *Apopellia* mitogenomes were smallest among known liverworts, over 49 kbp smaller than those found in *Pellia* and falling within 109,022–109,828 bp in range. The size of *Apopellia* mitogenomes is closer to the values found in mosses^{37,78,79}, than to those known in liverwort mitogenomes ranging from 142,510 bp in *Tritomaria quinquedentata* to 187,628 bp in *Monosolenium tenerum*^{29,69}. The size reduction of *Apopellia* mitogenome is not caused by gene or intron losses, but by the reduction of intergenic spacers, which affects all regions with exception of only few spacers like *cob-nad9*, *cox3-nad9* and usually long *atp1-cox1*, which was reduced only by 31%. The reduction or increase of mitogenome length in leafy liverworts is often caused by indels between genes *nad3* and *rpl10*. This region contains the pseudogenized (with exception of *Haplomitrium* and *Treubia* where *nad7* is functional) *nad7* gene and three tRNAs: *trnV*-UAC, *trnD*-GUC, *trnS*-GCU and is a main source of mitogenome size variation among *Nowellia* and *Scapania* species²⁸. Comparative analyses of Pelliidae mitogenomes also confirmed this observation revealing over 11 kbp difference in *nad3-rpl10* region between *Makinoa* and *Fossombronia* species, with differ by 8 kbp in total mitogenome length (*Makinoa* contains two more *cox1* introns). This region is also twofold smaller in *Apopellia* (5.3 kbp) than in *Pellia* (11.4 kbp), but it comprises only 11% of total mitogenome reduction in the former.

Phylogenetic placement of Pelliidae based on plastid sequence is mostly congruent with previous studies^{28,35,70}. One minor difference is the position of *Conocephalum conicum*, which in the work of Dong et al. was resolved as an unusual clade, probably due to plastome misassembly (GB accession number MK645816). The latest study on complex thaloid phylogeny confirms the position of *Conocephalum* as sister to *Riccia*⁸⁰. Chloroplast dataset supports monophyly of Pelliidae, however the mitogenome dataset resolved Fossombroniales and Pallaviciniiales in common clade with remaining Jungermanniopsida, while Pelliiales was resolved as the earliest divergent from that class. Cyto-nuclear incongruences are common across phylogeny of vascular plants⁸¹ and were also confirmed within Pelliidae, where the application of transcriptomic data revealed that mito-phylogenetic results support nuclear gene topology¹⁹. Incongruences among nuclear, mitochondrial and plastid dataset are often explained by incomplete lineage sorting⁸², horizontal gene transfers⁸³ and hybridization events⁸⁴, however in the case of Pelliidae aforementioned mechanisms are unlikely. The organellar genomes of Pelliidae seem to evolve much faster than any other subclass of liverworts.

The topology of the obtained trees is congruent with previous studies on Pelliiales and fully supports splitting of *Pellia* sensu lato by placing *Pellia endiviifolia* in the separate genus *Apopellia*²⁶. The splitting of the *Pellia* genus is not only supported by typical sequence phylogenetic divergence but also by differences in the structures and genome gene content as discussed above. Moreover, the mitochondrial dataset revealed a paraphyly of Pelliaceae and resolved *Apopellia* as a sister group to Fossombroniales.

Super-barcoding and molecular diagnostic characters. Application of organellar genomes as super-barcodes enabled identification of all analysed evolutionary lineages of *Pellia* s.l. with the exception of allopolyploid *P. borealis* which inherited organellar genomes from *P. epiphylla* N⁸⁵.

The number of plastid MDCs characterising *Apopellia* species decreased in comparison to previous study²⁸, due to expanding the dataset by a spring form, which reveals some intermediate characters between water and terrestrial forms, but it still, like *P. epiphylla* S and N, falls within the range known from *Calypogeia*—a taxonomically good liverwort species^{86,87}.

Based on our data it is hard to estimate the effect of multiple specimen sampling on presence of diagnostic nucleotides or MDCs. In four out of six species (including cryptic) the impact of sample number within 2–6 range on plastid MDCs was below 1% (species of *Apopellia* and *P. neesiana*). The decrease in plastid MDCs by 31% was detected in *P. epiphylla* S (N = 3) and by 15% in *P. epiphylla* N (N = 6), but it was not correlated with sample size. The mitogenomes reveal contrasting patterns, but again, not correlated with sample size. The decrease in MDCs in the case of *P. epiphylla* cryptic species was 7% and 9% for S and N respectively, but in the case of *A. endiviifolia* water form decrease in MDCs was 35%.

The mitogenomes deliver more MDCs than the plastome, however two factors should be discussed. The number of detected MDCs depends on alignment size, which in the case of mitogenomes was as much as 71% longer, however it doesn't explain the over 5.6-fold higher number of MDCs. In this particular case, the main reason for the outnumber of MDCs in mitogenome is a bigger number of large indels in mitochondrial dataset (1385 indels with mean length of 146.2 bp) in comparison to plastid dataset (1629 indels of 9.1 bp mean length). Thus the mitogenomes were more efficient in delimitation of cryptic species within the *P. epiphylla* complex, while plastomes performed better in the case of *A. endiviifolia* complex. These results can be explained by differences in genome sizes, since the mitogenome of *Apopellia* is smaller than plastome, which is quite unusual in plants, with exception of mosses^{37,78,88}.

Comparative organellar genomics support previous hypotheses about allopolyploid origin of *Pellia borealis*, which pointed out *Pellia epiphylla* N as a maternal species⁸⁵. Both organellar genomes found in *P. borealis* were almost identical to those found in *P. epiphylla* N. The lack of plastome MDCs and only four MDCs in mitogenomes suggest the late Pleistocene origin of *P. borealis*, when ranges of the N and S lineages could form a natural

hybridization zone. The allopolyploid *Calypogeia muelleriana*, which carries the maternally inherited plastome of *C. integristupula*, revealed specific SNPs enabling molecular delimitation of this species based on the plastome and mitochondrial datasets^{86,87}.

The distribution of *Pellia epiphylla* cryptic species can't be described as northern (N) and southern (S) lineages and their ranges should be revisited. Previous studies revealed geographical vicariance despite similar ecological requirements^{22,89}, but nowadays we can't confirm this pattern, since N lineage was also found in Carpathians. Coexistence of cryptic species lineages is common in other known complexes as *Aneura pinguis*³ or *Apopellia endiviifolia*²².

Our results confirm the presence of the third cryptic lineage named “C” within *Apopellia endiviifolia*, described previously by Polok et al.²⁴ as “well-headed” form. Phylogenetic analyses resolved species C as a sister to A which corresponds to similar ecological preferences of these forms. However, based on current knowledge on distribution of species C, its habitat preferences are much narrower than in the case of species A. All known specimens of species C occurred directly in the springs, they cover the bottoms of forming watercourses. The earlier studies suggested that species A is known mainly from the mountains in the direct neighbourhood of running water, with very restricted occurrence in the lowlands, which corresponds to preferred microhabitat distribution⁹⁰. The sites with clear and fast running shallow water, which this lineage depends on, are more common in the mountains. The geographic distribution of species B is concentrated in lowlands with rare occurrence in the mountains and the species prefers denuded soil and rock detritus⁹⁰. The species C is currently known from five sites, including three lowland and two mountain locations²⁴ and its scattered distribution can be explained by presence of spring areas. Morphological studies on *A. endiviifolia* revealed significant differences between cryptic species in quantitative traits like size of basal and marginal cells, even after six months of greenhouse culture²⁴. The species C had the smallest basal and marginal cells within *A. endiviifolia* complex. However, further ecological and morphological studies are needed to better characterise cryptic species within *A. endiviifolia* which could lead to formal taxonomic recognition. A species called *Apopellia megaspora* was identified in the early 1980s as one of the cryptic North American species of *A. endiviifolia*⁹¹. In addition to differences in spore diameters, it is distinguished from other *Apopellia* species by the absence of thallus proliferation in autumn. Another cryptic species of simple thalloid liverwort, *Aneura pinguis*, has been more thoroughly studied, but despite the presence of various lineage-specific DNA markers^{3,67,92} and many biochemical compounds⁹³, it lacks significant morpho-anatomical differentiation. Furthermore, the cryptic lineages of *A. endiviifolia* and some of *A. pinguis* differ in their preferred microhabitats, which can be characterised not only by general ecological descriptions, but also by specific parameters such as pH and concentrations of calcium, magnesium, potassium, and sodium⁹⁴.

Results presented in this study provide novel insights on evolution of liverwort organellar genomes, revealing high differentiation of mitogenomes among closely related taxa. Both plastid and mitochondrial genomes proved their usefulness in molecular delimitation of cryptic species of *Apopellia* and *Pellia*. Surprisingly, effectiveness of mitogenomes in species discrimination within Pellidae was on a similar level as plastome.

Data availability

The datasets used and/or analysed during the current study available from the corresponding author on reasonable request.

Received: 27 January 2023; Accepted: 15 May 2023

Published online: 23 May 2023

References

- Mayden, R. L. A hierarchy of species concepts: the denouement in the saga of the species problem. *Species. The units of biodiversity* (1997).
- Wiens, J. J. Species delimitation: New approaches for discovering diversity. *Syst. Biol.* **56**, 875–878 (2007).
- Bączkiewicz, A., Szczecińska, M., Sawicki, J., Stebel, A. & Buczkowska, K. DNA barcoding, ecology and geography of the cryptic species of *Aneura pinguis* and their relationships with *Aneura maxima* and *Aneura mirabilis* (Metzgeriales, Marchantiophyta). *PLoS One* **12**, e0188837 (2017).
- Heinrichs, J. et al. Formalizing morphologically cryptic biological entities: New insights from DNA taxonomy, hybridization, and biogeography in the leafy liverwort *Porella platyphylla* (Jungermanniopsida Porellales). *Am. J. Bot.* **98**, 1252–1262 (2011).
- Wachowiak, W., Bączkiewicz, A., Chudzińska, E. & Buczkowska, K. Cryptic speciation in liverworts—A case study in the *Aneura pinguis* complex. *Bot. J. Linn. Soc.* **155**, 273–282 (2007).
- Therrien, J. P., Crandall-Stotler, B. J. & Stotler, R. E. Morphological and genetic variation in *Porella platyphylla* and *P. platyphylloidea* and their systematic implications. *Bryologist* **101**, 1–19 (1998).
- Bakalin, V. A., Vilnet, A. A., Choi, S. S. & Nguyen, V. S. *Blepharostoma trichophyllum* s.l. (Marchantiophyta): The complex of sibling species and hybrids. *Plants* **9**, 1–28 (2020).
- Szweykowski, J., Buczkowska, K. & Odrzykoski, I. J. *Conocephalum salebrosum* (Marchantiopsida, Conocephalaceae)—A new Holarctic liverwort species. *Plant Syst. Evol.* **253**, 133–158 (2005).
- Buczkowska, K. et al. Does *Calypogeia azurea* (Calypogeiaceae, Marchantiophyta) occur outside Europe? Molecular and morphological evidence. *PLoS One* **13**, e0204561 (2018).
- Shaw, J. & Renzaglia, K. Phylogeny and diversification of bryophytes. *Am. J. Bot.* **91**, 1557–1581 (2004).
- Crandall-Stotler, B., Stotler, R. E. & Long, D. G. Phylogeny and classification of the Marchantiophyta. *Edinb. J. Bot.* **66**, 155–198 (2009).
- Söderström, L. et al. World checklist of hornworts and liverworts. *PhytoKeys* **59**, 1–828 (2016).
- Forrest, L. L. et al. Unraveling the evolutionary history of the liverworts (Marchantiophyta): Multiple taxa, genomes and analyses. *Bryologist* **109**, 303–334 (2006).
- He-Nygrén, X., Juslén, A., Ahonen, I., Glenney, D. & Piippo, S. Illuminating the evolutionary history of liverworts (Marchantiophyta)—Towards a natural classification. *Cladistics* **22**, 1–31 (2006).
- Frey, W. & Stech, M. A morpho-molecular classification of the liverworts (Hepaticophytina, Bryophyta). *Nova Hedwigia* **81**, 55–78 (2005).

16. Crandall-Stotler, B., Stotler, R. E., Zhang, L. & Forrest, L. L. On the morphology, systematics and phylogeny of Noteroclada (Noterocladaceae, Marchantiophyta). *Nova Hedwigia* **91**, 421–450 (2010).
17. Damsholt, K. *Illustrated Flora of Nordic Liverworts and Hornworts*, 2nd ed. <https://www.worldcat.org/formats-editions/756403484> (Oikos Editorial Office, 2009).
18. Schuster, R. M. Studies on antipodal Hepaticae; IX. Phyllohalliaceae. *Trans. Br. Bryol. Soc.* **5**, 283–288 (1967).
19. Dong, S. *et al.* Plastid genomes and phylogenomics of liverworts (Marchantiophyta): Conserved genome structure but highest relative plastid substitution rate in land plants. *Mol. Phylogenet. Evol.* **161**, 107171 (2021).
20. Lorbeer, G. Die Zytologie Der Lebermoose Mit Besonderer Berücksichtigung Allgemeiner Chromosomenfragen: Tl 1. *J. Bot.* **80**, 567–818 (1934).
21. Krzakowa, M. Evolution and speciation in *Pellia*, with special reference to the *Pellia megaspora-endiviifolia* complex (Metzgeriales) IV. Isozyme investigations. *J. Bryol.* **11**, 447–450 (1981).
22. Zieliński, R. Interpretation of electrophoretic patterns in population genetics of bryophytes: VI. Genetic variation and evolution of the liverwort genus *Pellia* with special reference to Central European Territory. *Lindbergia* **12**, 87–96 (1987).
23. Szwejkowska-Kulińska, Z., Pacak, A. & Jankowiak, K. New DNA markers for discrimination between closely-related species and for the reconstruction of historical events; an example using liverworts. *Cell. Mol. Biol. Lett.* **7**, 403–416 (2002).
24. Polok, K. *et al.* Evolutionary divergence within *Pellia endiviifolia* [Dicks.] Dum. from Poland. *Seria Biologiczna. Uniwersytet im. Adama Mickiewicza w Poznaniu* **72** (2005).
25. Pacak, A., Fiedorow, R., Dabert, J. & Szwejkowska-Kulińska, Z. RAPD technique for taxonomic studies of *Pellia epiphylla*-complex (Hepaticae, Metzgeriales). *Genetica* **104**, 179–187 (1998).
26. Schütz, N., Quandt, D. & Nebel, M. The position of the genus *Apopellia* stat. nov. within the Pelliales (Marchantiophytina: Jungermanniopsida). *Taxon* **65**, 221–234 (2016).
27. Grosche, C., Funk, H. T., Maier, U. G. & Zauner, S. The chloroplast genome of *Pellia endiviifolia*: Gene content, RNA-editing pattern, and the origin of chloroplast editing. *Genome Biol. Evol.* **4**, 1349–1357 (2012).
28. Sawicki, J., Krawczyk, K., Ślipiko, M., Szandar, K. & Szczecińska, M. Comparative analysis of *Apopellia endiviifolia* plastomes reveals a strikingly high level of differentiation between its terrestrial and water form. *Diversity (Basel)* **13**, 674 (2021).
29. Dong, S. *et al.* Mitochondrial genomes of the early land plant lineage liverworts (Marchantiophyta): Conserved genome structure, and ongoing low frequency recombination. *BMC Genom.* **20**, 1–14 (2019).
30. Myszczynski, K., Ślipiko, M. & Sawicki, J. Potential of transcript editing across mitogenomes of early land plants shows novel and familiar trends. *Int. J. Mol. Sci.* **20**, 2963 (2019).
31. Szczecińska, M. & Sawicki, J. Genomic resources of three pulsatilla species reveal evolutionary hotspots, species-specific sites and variable plastid structure in the family Ranunculaceae. *Int. J. Mol. Sci.* **16**, 22258–22279 (2015).
32. Nevill, P. G. *et al.* Plastome-wide rearrangements and gene losses in carnivorous Droseraceae. *Genome Biol. Evol.* **11**, 472–485 (2019).
33. Forrest, L. L., Wickett, N. J., Cox, C. J. & Goffinet, B. Deep sequencing of *Ptilidium* (Ptilidiaceae) suggests evolutionary stasis in liverwort plastid genome structure. *Plant Ecol. Evol.* **144**, 29–43 (2011).
34. Myszczynski, K., Górski, P., Ślipiko, M. & Sawicki, J. Sequencing of organellar genomes of *Gymnomitrium concinnatum* (Jungermanniales) revealed the first exception in the structure and gene order of evolutionary stable liverworts mitogenomes. *BMC Plant Biol.* **18**, 1–12 (2018).
35. Yu, Y. *et al.* Chloroplast phylogenomics of liverworts: A reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. *Cladistics* **36**, 184–193 (2020).
36. Villarreal, A. *et al.* Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytol.* **209**, 1734–1746 (2016).
37. Liu, Y., Cox, C. J., Wang, W. & Goffinet, B. Mitochondrial phylogenomics of early land plants: Mitigating the effects of saturation, compositional heterogeneity, and codon-usage bias. *Syst. Biol.* **63**, 862–878 (2014).
38. Gualberto, J. M. & Newton, K. J. Plant mitochondrial genomes: Dynamics and mechanisms of mutation. *Annu. Rev. Plant Biol.* **68**, 225–252 (2017).
39. Gualberto, J. M. *et al.* The plant mitochondrial genome: Dynamics and maintenance. *Biochimie* **100**, 107–120. <https://doi.org/10.1016/j.biochi.2013.09.016> (2014).
40. Sawicki, J. & Szczecińska, M. A comparison of PCR-based markers for the molecular identification of *Sphagnum* species of the section Acutifolia. *Acta Soc. Bot. Pol.* **80**, 185–192 (2011).
41. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
42. Szczecińska, M., Gomolińska, A., Szkudlarz, P. & Sawicki, J. Plastid and nuclear genomic resources of a relict and endangered plant species: *Chamaedaphne calyculata* (L.) Moench (Ericaceae). *Turk. J. Bot.* **38**, 1229–1238 (2014).
43. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
44. Dierckxens, N., Mardulyn, P. & Smits, G. NOVOPlasty: De novo assembly of organelle genomes from whole genome data. *Nucleic Acids Res.* **45**, e18 (2017).
45. Tillich, M. *et al.* GeSeq—Versatile and accurate annotation of organelle genomes. *Nucleic Acids Res.* **45**, W6–W11 (2017).
46. Greiner, S., Lehwark, P. & Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **47**, 59–64 (2019).
47. Amiroussi, A., Hyvönen, J. & Poczai, P. IRscope: An online program to visualize the junction sites of chloroplast genomes. *Bioinformatics* **34**, 3030–3031 (2018).
48. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
49. Minh, B. Q. *et al.* IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
50. Revell, L. J. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012).
51. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
52. Huson, D. H. SplitsTree: Analyzing and visualizing evolutionary data. *Bioinformatics* **14**, 68–73 (1998).
53. Brown, S. D. J. *et al.* Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol. Ecol. Resour.* **12**, 562–565 (2012).
54. Jörger, K. M. & Schrödl, M. How to describe a cryptic species? Practical challenges of molecular taxonomy. *Front. Zool.* **10**, 1–27 (2013).
55. Merkelbach, L. M. & Borges, L. M. S. Make every species count: Fastachar software for rapid determination of molecular diagnostic characters to describe species. *Mol. Ecol. Resour.* **20**, 1761–1768 (2020).
56. Pfeifer, B., Wittelsbürger, U., Ramos-Onsins, S. E. & Lercher, M. J. PopGenome: An efficient swiss army knife for population genomic analyses in R. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
57. Zhou, L. *et al.* ggmsa: A visual exploration tool for multiple sequence alignment and associated data. *Brief Bioinform.* **23**, bbac222 (2022).

58. Puillandre, N., Brouillet, S. & Achaz, G. ASAP: Assemble species by automatic partitioning. *Mol. Ecol. Resour.* **21**, 609–620 (2021).
59. Sharp, P. M., Tuohy, T. M. F. & Mosurski, K. R. Codon usage in yeast: Cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Res.* **14**, 5125–5143 (1986).
60. Charif, D. & Lobry, J. R. SeqinR 1.0-2: A contributed package to the r project for statistical computing devoted to biological sequences retrieval and analysis 207–232. https://doi.org/10.1007/978-3-540-35306-5_10 (2007).
61. Wickham, H. *Elegant Graphics for Data Analysis: ggplot2*. *Applied Spatial Data Analysis with R* (2008).
62. Supek, F. & Vlahoviček, K. Comparison of codon usage measures and their applicability in prediction of microbial gene expressivity. *BMC Bioinform.* **6**, 1–15 (2005).
63. Elek, A., Kuzman, M. & Vlahoviček, K. Package ‘coRdon’ type package title codon usage analysis and prediction of gene expressivity. *Master’s thesis* (2022).
64. Ciborowski, K., Skierkowski, B., Żukowska, K., Krawczyk, K. & Sawicki, J. Nanopore sequencing of chloroplast genome of *Scapania undulata* (L.) Dumort., 1835 (Scapaniaceae, Jungermanniales). *Mitochondrial DNA B Resour.* **7**, 1424–1426 (2022).
65. Wang, W. & Lanfear, R. Long-reads reveal that the chloroplast genome exists in two distinct versions in most plants. *Genome Biol. Evol.* **11**, 3372–3381. <https://doi.org/10.1093/gbe/evz256> (2019).
66. Yu, Y. *et al.* Exploring the plastid genome disparity of liverworts. *J. Syst. Evol.* **57**, 382–394 (2019).
67. Myszczynski, K. *et al.* The extraordinary variation of the organellar genomes of the *Aneura pinguis* revealed advanced cryptic speciation of the early land plants. *Sci. Rep.* **7**, 1–12 (2017).
68. Sawicki, J. *et al.* The increase of simple sequence repeats during diversification of Marchantiidae, an early land plant lineage, leads to the first known expansion of inverted repeats in the evolutionarily-stable structure of liverwort plastomes. *Genes (Basel)* **11**, 299 (2020).
69. Ślipiko, M. *et al.* Comparative analysis of four *Calypogeia* species revealed unexpected change in evolutionarily-stable liverwort mitogenomes. *Genes (Basel)* **8**, 395 (2017).
70. Dong, S. & Liu, Y. The mitochondrial genomes of bryophytes. *Bryophyt. Divers. Evol.* **43**, 112–126 (2021).
71. Liu, Y., Xue, J. Y., Wang, B., Li, L. & Qiu, Y. L. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: Dynamic and conservative evolution. *PLoS One* **6**, 25836 (2011).
72. Crandall-Stotler, B. J., Forrest, L. L. & Stotler, R. E. Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). *Taxon* **54**, 299–316 (2005).
73. Rojas, J. *et al.* Codon usage revisited: Lack of correlation between codon usage and the number of tRNA genes in enterobacteria. *Biochem. Biophys. Res. Commun.* **502**, 450–455 (2018).
74. Wei, Y., Silke, J. R. & Xia, X. An improved estimation of tRNA expression to better elucidate the coevolution between tRNA abundance and codon usage in bacteria. *Sci. Rep.* **9**, 3184 (2019).
75. André, C., Levy, A. & Walbot, V. Small repeated sequences and the structure of plant mitochondrial genomes. *Trends Genet.* **8**, 128–132 (1992).
76. Maréchal, A. & Brisson, N. Recombination and the maintenance of plant organelle genome stability. *New Phytol.* **186**, 299–317 (2010).
77. Kwon, W., Kim, Y. & Park, J. The complete chloroplast genome sequence of *Dumortiera hirsuta* (Sw.) Nees (Marchantiophyta, Dumortieraceae). *Mitochondrial DNA B Resour.* **4**, 318–319 (2019).
78. Sawicki, J. *et al.* Mitogenomic analyses support the recent division of the genus *Orthotrichum* (Orthotrichaceae, Bryophyta). *Sci. Rep.* **7**, 4408 (2017).
79. Sawicki, J., Szczecińska, M., Kulik, T., Gomolińska, A. M. & Plášek, V. The complete mitochondrial genome of the epiphytic moss *Orthotrichum speciosum*. *Mitochondrial DNA Part A* **27**, 1709–1710. <https://doi.org/10.3109/19401736.2014.961133> (2014).
80. Xiang, Y.-L. *et al.* The Macaronesian liverwort *Riccia boumanii* Dirkse, Losada and M. Stech (Marchantiophyta: Ricciaceae) confirmed new to Asia by morphological and molecular evidence. *Cryptogam Bryol.* **43**, 201–209 (2022).
81. Meng, K.-K. *et al.* Phylogenomic analyses based on genome-skimming data reveal cyto-nuclear discordance in the evolutionary history of *Cotoneaster* (Rosaceae). *Mol. Phylogenet. Evol.* **158**, 107083 (2021).
82. Wakeley, J. *Coalescent Theory: An Introduction* (Roberts and Company Publishers, 2009).
83. Tofigh, A., Hallett, M. & Lagergren, J. Simultaneous identification of duplications, losses, and lateral gene transfers. *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8**, 517–535 (2011).
84. Morales-Briones, D. F., Liston, A. & Tank, D. C. Phylogenomic analyses reveal a deep history of hybridization and polyploidy in the Neotropical genus *Lachemilla* (Rosaceae). *New Phytol.* **218**, 1668–1684 (2018).
85. Pacak, A. & Szwejkowska-Kulińska, Z. Organellar inheritance in liverworts: An example of *Pellia borealis*. *J. Mol. Evol.* **56**, 11–17 (2003).
86. Ślipiko, M., Myszczynski, K., Buczkowska, K., Bączkiewicz, A. & Sawicki, J. Super-mitobarcoding in plant species identification? It can work! The case of leafy liverworts belonging to the genus *Calypogeia*. *Int. J. Mol. Sci.* **23**, 15570 (2022).
87. Ślipiko, M. *et al.* Molecular delimitation of European leafy liverworts of the genus *Calypogeia* based on plastid super-barcodes. *BMC Plant Biol.* **20**, 243 (2020).
88. Mizia, P. *et al.* Comparative plastomes analysis reveals the first infrageneric evolutionary hotspots of *Orthotrichum* s.l. (Orthotrichaceae, Bryophyta). *Turk. J. Bot.* **43**, 444–457 (2019).
89. Szwejkowski, J. & Odrzykoski, I. J. Chemical differentiation of *Aneura pinguis* (L.) Dum. (Hepaticae, Aneuraceae) in Poland and some comments on application of enzymatic markers in bryology. *Bryophytes, Their Chemistry and Chemical Taxonomy* 437–448. <https://doi.org/10.1007/BF00937961> (1990).
90. Szwejkowski, J. An annotated checklist of Polish liverworts and hornworts: Krytyczna lista wątrobowców i glikiwów Polski. *Biodivers. Poland* **4**, 18–20 (2006).
91. Schuster, R. M. Evolution and speciation in *Pellia*, with special reference to the *Pellia megaspora-endiviifolia* complex (Metzgeriales) I. Taxonomy and distribution. *J. Bryol.* **11**, 411–431 (1981).
92. Bączkiewicz, A., Szczecińska, M., Sawicki, J., Stebel, A. & Buczkowska, K. DNA barcoding, ecology and geography of the cryptic species of *Aneura pinguis* and their relationships with *Aneura maxima* and *Aneura mirabilis* (Metzgeriales, Marchantiophyta). *PLoS One* **12**, e0188837 (2017).
93. Wawrzyniak, R., Wasiak, W., Bączkiewicz, A. & Buczkowska, K. Volatile compounds in cryptic species of the *Aneura pinguis* complex and *Aneura maxima* (Marchantiophyta, Metzgeriidae). *Phytochemistry* **105**, 115–122 (2014).
94. Bączkiewicz, A. *et al.* Geochemical alkalinity and acidity as preferential site-specific for three lineages liverwort of *Aneura pinguis* cryptic species A. *Sci. Rep.* **11**, 4334 (2021).

Author contributions

L.P., Writing—original draft preparation, Conceptualization, Methodology, Formal analysis, Visualization; K.K., Formal analysis, Methodology; P.G., Resources, Writing—original draft preparation; M.Sz., Resources, Formal analysis; M.M., Formal analysis, Visualization; M.Ś., Formal analysis, Writing—original draft preparation; J.S., Conceptualization, Formal analysis, Funding acquisition, Visualization, Supervision.

Funding

The sequencing of organellar genomes was financially supported by The National Science Center Kraków, Poland: Grant No. 2020/39/B/NZ8/02504.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-35269-3>.

Correspondence and requests for materials should be addressed to Ł.P.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023