Communication

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

Jakub Sawicki [1,]* [ID]**, Alina Bączkiewicz** [2]**, Katarzyna Buczkowska** [2]**, Piotr Górski** [3]**, Katarzyna Krawczyk** [1]**, Patryk Mizia** [1]**, Kamil Myszczyński** [1] [ID]**, Monika Ślipiko** [1] [ID] **and Monika Szczecińska** [1]

[1] Department of Botany and Nature Protection, University of Warmia and Mazury in Olsztyn; 10-719 Olsztyn, Poland; katarzyna.krawczyk@uwm.edu.pl (K.K.); patryk.mizia@gmail.com (P.M.); kamil.myszczynski@gmail.com (K.M.); monika.slipiko@uwm.edu.pl (M.Ś.); monika.szczecinska@uwm.edu.pl (M.S.)

[2] Department of Genetics, Faculty of Biology, Adam Mickiewicz University in Poznań, 61-614 Poznań, Poland; alinbacz@amu.edu.pl (A.B.); androsac@amu.edu.pl (K.B.)

[3] Department of Botany, Poznań University of Life Sciences, 60-625 Poznań, Poland; piotr.gorski@up.poznan.pl

[*] Correspondence: jakub.sawicki@uwm.edu.pl; Tel.: +48-89-523-3494; Fax: +48-89-523 3546

check for
updates

Abstract: The chloroplast genomes of liverworts, an early land plant lineage, exhibit stable structure and gene content, however the known resources are very limited. The newly sequenced plastomes of *Conocephalum, Riccia* and *Sphaerocarpos* species revealed an increase of simple sequence repeats during the diversification of complex thalloid liverwort lineage. The presence of long TA motifs forced applying the long-read nanopore sequencing method for proper and dependable plastome assembly, since the length of dinucleotide repeats overcome the length of Illumina short reads. The accumulation of SSRs (simple sequence repeats) enabled the expansion of inverted repeats by the incorporation of *rps*12 and *rps*7 genes, which were part of large single copy (LSC) regions in the previously sequenced plastomes. The expansion of inverted repeat (IR) at the genus level is reported for the first time for non-flowering plants. Moreover, comparative analyses with remaining liverwort lineages revealed that the presence of SSR in plastomes is specific for simple thalloid species. Phylogenomic analysis resulted in trees confirming monophyly of Marchantiidae and partially congruent with previous studies, due to dataset-dependent results of *Dumortiera-Reboulia* relationships. Despite the lower evolutionary rate of Marchantiales plastomes, significant barcoding gap was detected, even for recently divergent holarctic *Conocephalum* species. The sliding window analyses revealed the presence of 18 optimal (500 bp long) barcodes that enable the molecular identification of all studied species.

Keywords: chloroplast genome; inverted repeats expansion; liverworts; cpSSR; nanopore sequencing; super-barcoding; phylogeny

1. Introduction

The plastid genomes are the most frequently used molecules in evolutionary studies on plants, including phylogenetics, phylogeography, and population genetics [1–3]. They also serve as barcoding resources for most of the taxonomic groups [4], with only a few exceptions when plastid-derived

barcodes fail [5,6]. In opposition to vascular plants, plastid genomes of early land plants—bryophytes are yet poorly explored.

Bryophyte plastomes, as well as well-studied angiosperm 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). In contrast to vascular plants, where plastome structure variation is observed at the family or even at genus level [7,8], in bryophytes the gene content and order seem to be very stable across evolutionary lineages. Gene content and gene order in the plastome is conserved in all main liverwort lineages, including complex thalloid, simple thalloid, and leafy species [9–11], despite it being millions of years since their divergence [12]. The mosses, phylogenetically sister to liverworts, are known from at least two different conformations [13] of chloroplast genome due to LSC inversion in Funariaceae [14], however the gene content of IRs, LSC, and SSC regions remained unchanged. In vascular plants of structural rearrangements that were observed within plastome is a shift in the IR boundary. The IRs of land plants fluctuate in size, either increasing the content of genes and non-coding DNA segments that otherwise would be single-copy, or losing some duplications via the transfer to single copy regions [15].

The species of genus *Conocephalum* belongs to an early divergent liverwort group—Marchantiidae characterized by thick thallus differentiated into dorsal zone (with air-chambers and air-pores) and ventral zone with storage tissue. *Conocephalum* species is one of the most common complex thallose liverwort in Holarctic, occupying wet, shaded sites in lowlands, highlands, and mountains. *Conocephalum* became the object of interest of plant science researchers, including taxonomists, geneticist, and evolutionists, due to large thalli and widespread occurrence. Biochemical followed by isoenzymatic analyses revealed the presence of cryptic speciation within *C. conicum* s.l. [16,17]. The detailed morpho-anatomical analyses resulted in the description of numerous diagnostics features enabling distinguishing *C. salebrosum* from *C. conicum* s.s. [18].

The recent phylogenetic studies revealed the lower genetic diversity of Marchantiidae plastomes in comparison to evolutionarily younger Jungermannidae [9,12], but the reasons of slower evolution of the formers are not known. The lack or low level of RNA editing in comparison to other liverworts is another characteristic feature of Marchantiidae [19,20].

In this study, we sequenced, assembled, annotated, and analyzed the plastomes of two holarctic *Conocephalum* species, which revealed unique features, including the first exception from the stable structure of liverwort plastomes. The assembly of SSR rich plastomes required the application of long-read sequencing technology, which was used for the first time in study on liverworts. Additionally, we sequenced two plastomes of early (*Sphaerocarpos texanus*) and late divergent (*Riccia fluitans*) taxa of Marchantiidae to verify whether the features that were observed in *Conocephalum* fit into any evolutionarily-wide scenario. Comparative analyses revealed species-specific SNP that are useful in molecular delimitation of *Conocephalum* species.

2. Results and Discussion

2.1. Characteristics of the Newly Sequenced Chloroplast Genomes

The short reads sequencing of prepared libraries resulted in the output data in 2–7Gb range (Table S1). The long-read nanopore sequencing was performed to provide reliable assembly over long TA-repeats regions of *Conocephalum* plastomes, resulting in over 183,000 reads (within the range 1000–83,000 bp, N50 = 2200 bp), which enabled the validation of ambiguous short-reads assembly.

The plastomes of *Conocephalum* species were 120,963 bp and 122,488 bp long in *C. conicum* and *C. salebrosum*, respectively. The structure is typical for most plants, including a pair of IR regions (each of 9612 bp in case of *C. conicum* and 11,400 bp in the case of *C. salebrosum*) separated by LSC (81,795 bp in case of *C. conicum* and 79,825 bp in case of *C. salebrosum*) and SSC (19,946 bp in the case of *C. conicum* and 19,866 bp in case of *C. salebrosum*) regions (Figure 1).
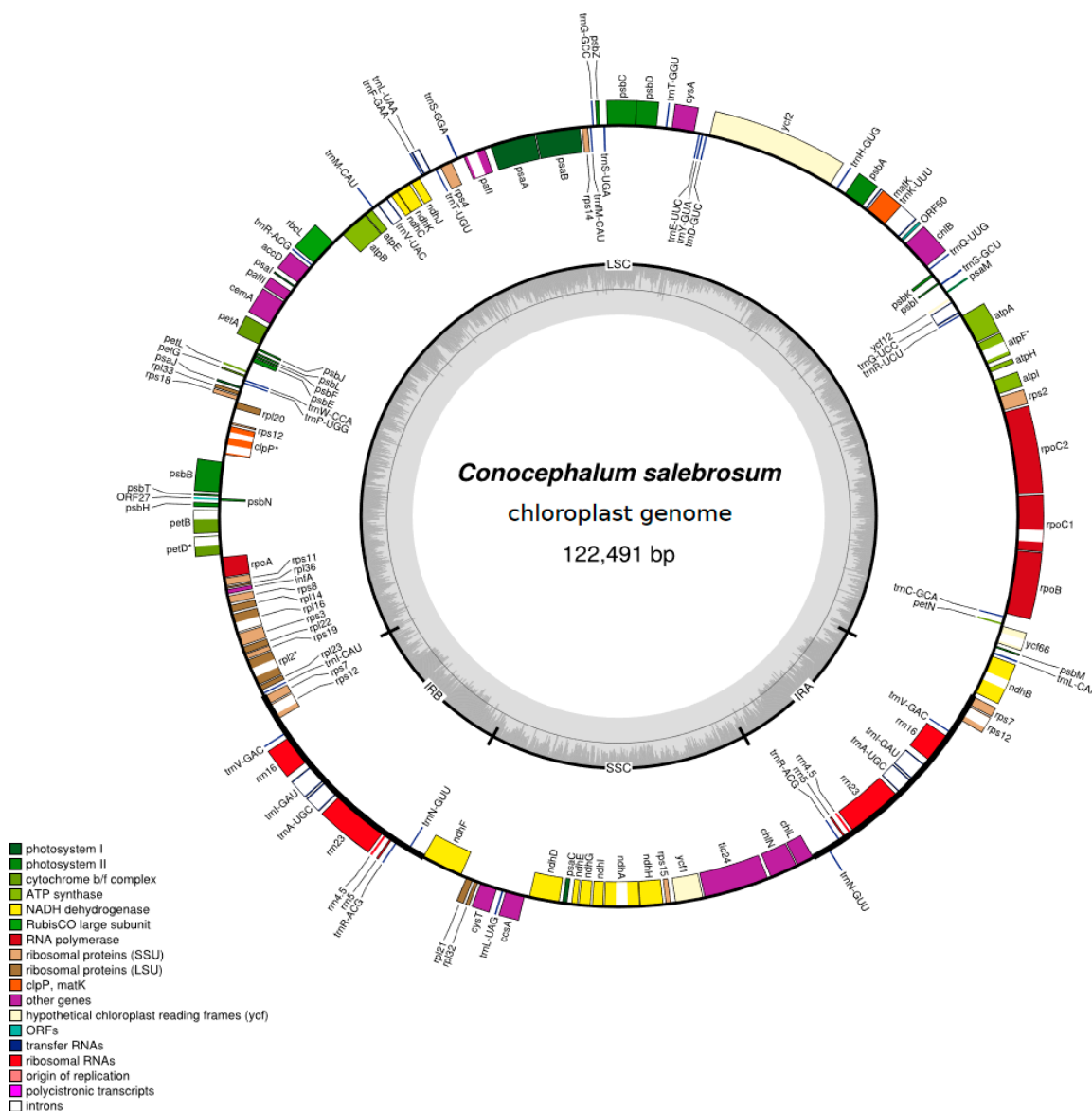
Figure 1. The chloroplast genome of *Conocephalum salebrosum*. Genes inside and outside the outer circle are transcribed in counterclockwise and clockwise directions, respectively. The genes are color-coded based on their function. The dashed area in the inner circle visualizes the G/C content.

The newly sequenced plastomes of *Riccia fluitans* and *Sphaerocarpos texanus* fall within the range with 121,999 and 121,377 bp, respectively.

The interspecific variation in plastome length of *Conocephalum* species is higher than in the previously analysed genera with the 1525 bp difference. Comparative analysis of the chloroplast genomes of six *Aneura pinguis* cryptic species revealed that the length of the chloroplast genomes ranged from 120,698 to 121,140 bp [21], resulting in the 442 bp difference. The first sequenced liverwort plastome of *Marchantia paleacea* [22] and the later sequenced plastome of the same species differ in length by 56 bp, while the three sequenced plastomes of *M. polymorpha* differ in length by 86 bp, so the interspecific difference could be up to 860 bp, depending on the specimen. The sequenced plastomes of water-living liverwort *Riccia fluitans* differ in length by 317 bp, which is currently the biggest gap at the species level among known Marchantiales chloroplast genomes. In contrast to *Marchantia* and *Riccia* the *Conocephalum* plastomes do not reveal any intraspecific variation in length.

The enlargement of *C. salebrosum* plastome is mostly connected with the expansion of IR regions by the inclusion of two genes: *rps*12 and *rps*7. The expansion of IRs was validated with nanopore long-read sequencing and by increasing Illumina short-read coverage over duplicated *rps*12 and *rps*7 regions (Supplementary Figure S1).

This is the first exception from the conservative structure of liverwort plastomes, since all of the previously sequenced ones shared the same IR, LSC, and SSR gene content. In the typical structure of liverwort plastome, *rps*12 and *rps*7 genes are located at 5' end of LSC region between *trn*V—the last gene of IRa, and *ndh*B genes [22]. However, in the case of *C. salebrosum*, the additional copies of *rps*12 and *rps*7 are located between *trn*V-GAC (IRa) and *trn*L-CAU, which in all known liverwort plastomes is the last LSC gene at the 3' end (Figure. 2).

The IR expansion seems to have evolved recently, since the plastome of the sister species *C. conicum* revealed a structure typical for liverworts with the IR junction presenting between *trn*V-*rpl*23 and *trn*V-*rps*12, respectively, for the IRa and IRb regions (Figure 2.). The analysed *Conocephalum* species diverged ca 15 My ago [12], therefore the IR expansion seems to be more recent than the species divergence.
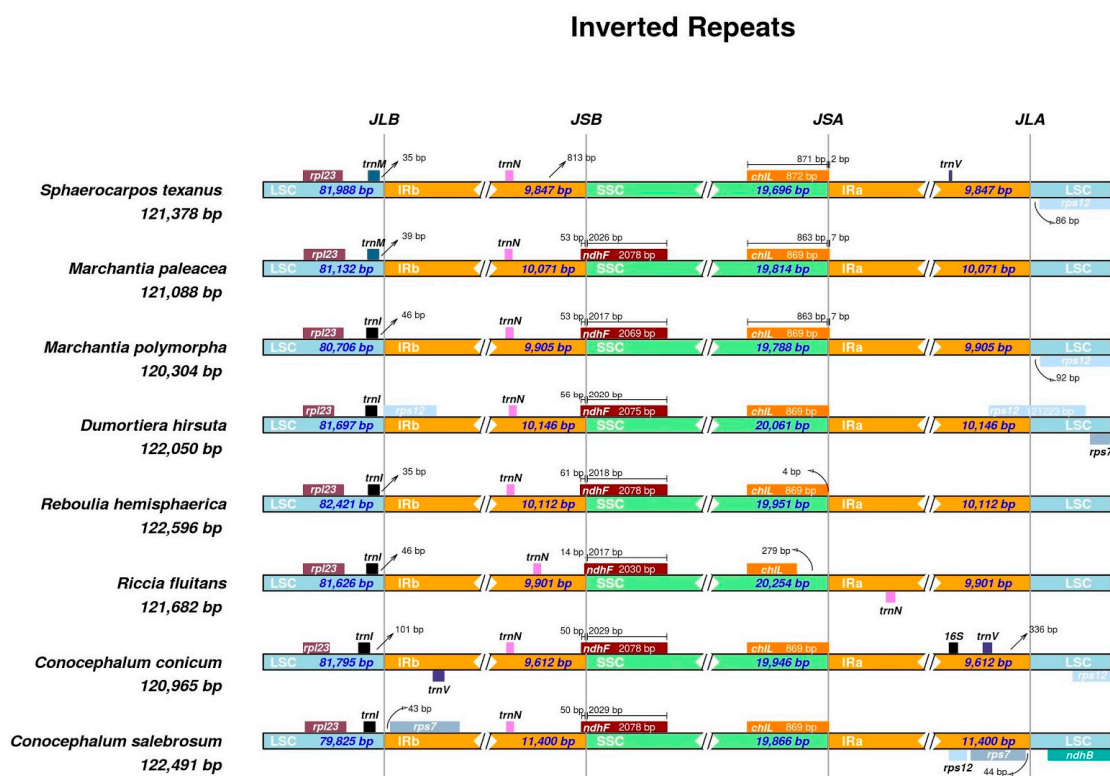

Figure 2. The differences in inverted repeat (IR) borders among Marchantiidae.

Several mechanisms, such as gene conversion, double-stand breaks (DBS), and genomic deletion, are considered to be involved in the IR expansion process [23,24]. The IR expansion might be possible due to the presence of "TA" tandem repeats within the regions flanking duplicated genes, which are not present or are more reduced in other known liverwort plastomes.

The plastomes of *Conocephalum* species contained 30% more repeats than chloroplast genomes of their relatives from genera *Marchantia*, *Dumortiera*, *Reboulia*, *Riccia*, and *Sphaerocarpos*. In comparison to the seed plants, bryophyte plastomes are richer in SSRs, with the exception of mononucleotide repeats [23], however the numbers of repeats do not seem to have an impact on structure stability by enabling the recombination. The positive correlation between the presence of SSRs and the frequency of recombination was found in human nuclear genome [24], but in the plant organellar genomes this

correlation was only found in the plastomes of Geraniaceae [25]. The plastome of *C. conicum* possesses a typical structure, despite the presence of longer TA repeats.

The expansion of IRs were found in several taxa of green algae [26] and vascular plants [15], but this process characterizes rather higher taxonomic units—orders and families. Among closely related species, the gain or loss of genes by IRs is known among several genera of flowering plants [27,28], however, for non-flowering plants, the differences in IR gene content at the genus level are reported for the first time.

2.2. Phylogenetics Relationships Based on Plastomes Datasets

The reliable alignment of complete genomes is difficult or impossible to obtain due to the high divergence of intergenic and intronic regions of liverwort plastomes, therefore the main phylogenetic analyses were split into two datasets: the first one included coding regions (CDS and ribosomal genes) of all liverwort lineages, while the second one, based on complete genomes, only comprised complex thalloid species.

The analysed dataset that was based on CDS and ribosomal genes delivered results that were mostly congruent with previous phylogenetic studies [12,19,20,29] and divided liverworts into five main lineages: Haplomitriidae, represented here by *Haplomitrium blumei*, which served as an outgroup, early divergent complex thalloids, followed by two lineages of simple thalloid and leafy liverworts (Figure 3)
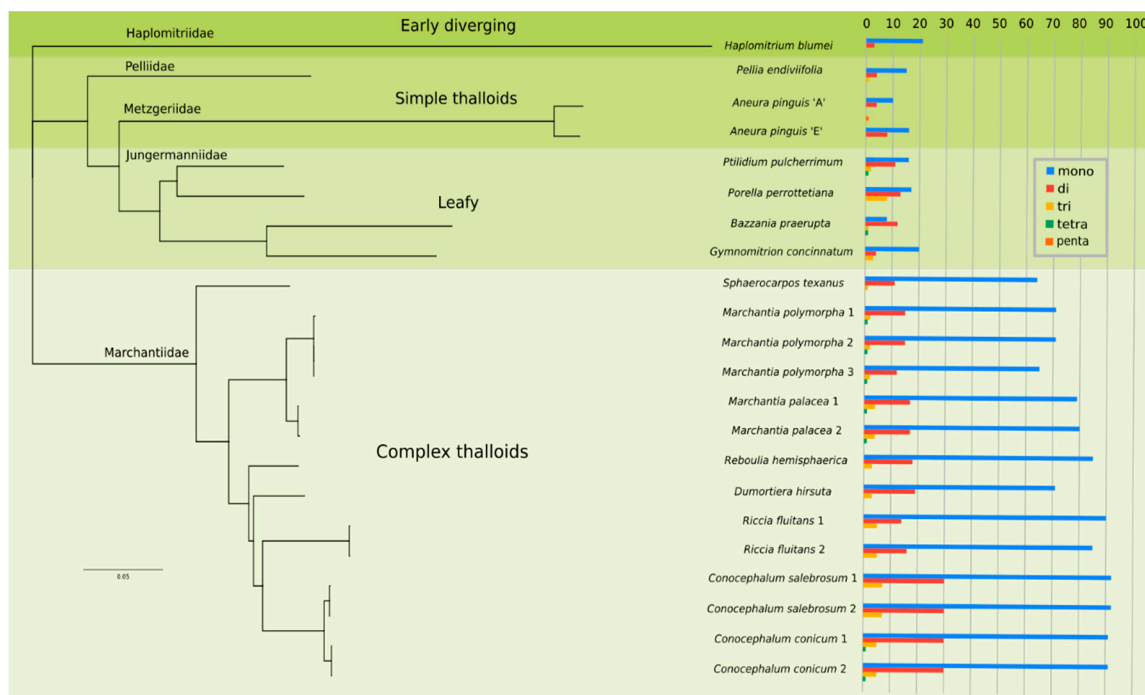

Figure 3. Phylogram based on coding regions of liverworts plastomes derived from a Bayesian analysis. The posterior probabilities values were equal to 1 for each of the clade. Right side of the figure provides data on SSR abundance in liverworts plastomes.

Bayesian and Maximum Likelihood analyses both provided the trees with identical topology and clade support. All of the groups mentioned above formed distinct clades with maximum bootstrap support and Bayesian posterior probabilities. Phylogeny that is based on the complete set of plastome coding regions is congruent with previous studies [12,29], however phylogenetic analysis of Marchantiales based on complete chloroplast genomes revealed incongruence with previously published data due to the position of *Dumortiera hirsuta* and *Reboulia hemisphaerica*. In the latest, most

complex study of Marchantiales using 11 loci from three genomes (one nuclear, three mitochondrial, and seven plastid), *Dumortiera* was resolved as an earlier divergent than *Reboulia* [12], while the results that were based on complete plastid genomes placed these species in a common clade (Figure S2), however weakly (58% of ML bootstrap support) or unsupported (in case of Bayesian analysis). The analysis of complete genomes, including spacers and introns, did not provide satisfactory resolution of *Reboulia-Dumortiera* relationships (Figure S2). The decrease of phylogenetic resolution due to the expansion of the dataset by spacers and introns is rather uncommon, since non-coding regions are usually more variable. However, the barcoding analyses revealed that phylogenetic incongruence is accumulated in the IR region of plastome, therefore additional analysis of single copy plastome regions were performed. The exclusion of IR region did not change the topology of trees (Figure 4), but provided significant (0.99 of Bayesian inference) and reasonable (70% ML BS) support for the *Reboulia-Dumortiera* clade. The monophyly of this clade was also reported in recent studies providing results that were based on mitogenomic data [19].
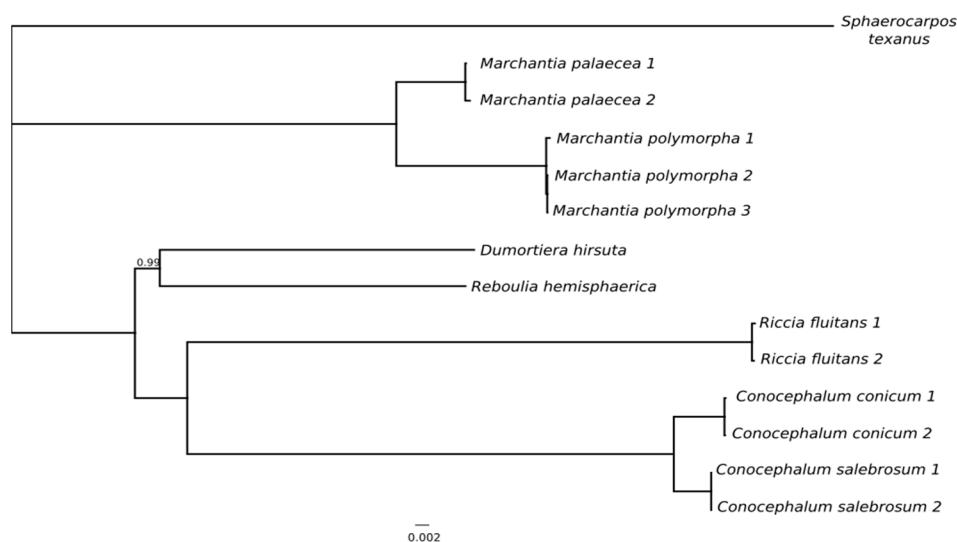

Figure 4. Phylogram of Marchantiidae derived from a Bayesian analysis of complete large single copy (LSC) and small single copy (SSC) regions. The posterior probabilities values lower than 1 are shown. The IR region was excluded from analysis due to incongruent phylogenetic signal.

2.3. Mining and Characterisation of SSRs in Liverwort Plastomes

The liverwort plastomes were previously considered as SSR rich, especially in comparison with tracheophytes, however this conclusion had been made based on very limited data, including only *Marchantia* [23]. Analysis of newly sequenced and available plastomes of liverworts performed in this study revealed a wide range of SSRs with different accumulation patterns (Figure 3), which only confirmed the previous observations in the case of Marchantiidae [23]. The species belonging to Marchantiidae revealed a high number of repeats, especially mono- and dinucleotides, however longer motifs, including tri-, tetra-, and pentanucleotide repeats, were too rare to draw any conclusion. The number and length of mononucleotide repeats present in the analysed plastomes of complex thalloids was rather stable. The number of repeats ranged from 64 in the early divergent *Sphaerocarpos* to 92 in *C. salebrosum* (Figure 3, Table S1) and their length (Table S2) fell into the range from 13 (*C. salebrosum*) to 17 (*Conocephalum conicum*, *Marchantia paleacea*, *Sphaerocapos texanus*).

The chloroplast genomes of *Conocephalum* species revealed a stable number of SSRs, the highest among liverworts (91–92 mono-, 30–32 dinucleotide), but they were very diverse in terms of their length. The longest dinucleotide motif that was observed in *C. conicum* was almost two times longer than in *C. salebrosum* (63 vs 33 AT repeats). Therefore, the reliable assembly of *C. conicum* plastome had to be supported with nanopore long-read sequencing.

Among the *Marchantia* species the interspecific differences were noticed in the case of mono-, di-, and trinucleotide repeats and, in all cases, the greatest number of SSRs was found in *M. paleacea* (Figure 3, Table S1). The tri- and tetranucleotide repeats were 75% longer in *M. paleacea,* albeit the length of mono- and dinucleotide SSRs was similar among the analysed species. The higher intrageneric variation of SSR present in *Marchantia,* could be explained by the more than twice longer regions in comparison to relatively younger species of *Conocephalum* [12]. The three available plastomes of *Riccia fluitans* revealed the highest variation in the numbers of mononucleotide repeats that fall within the range of 85 to 93.

The plastome of *Dumortriera hirsuta*, sister to *Marchantia* genus, contain a similar number of SSRs, while the basal for Marchantiidae, *Sphaerocarpos texanus* were characterized by the lowest number of identified repeats. The trend of SSR cumulation during complex thalloid diversification is clearly visible, despite the limited data (Figure 3).

The plastomes of Jungermanniales revealed a smaller number of repeats (Figure 3, Table S1), similar to those that were found in vascular plants [23], so the previously published conclusion should be narrowed only to complex thalloid liverworts. The SSR-gaining scenario is also supported by the low numbers of SSRs in the early divergent liverwort *H. blumei*, where only 24 SSRs were found including 21 mononucleotide and three dinucleotide repeats (Figure 3).

2.4. Plastome Diversity and Molecular Delimitations of Marchantiales

The dataset containing 14 complete sequences of chloroplast genomes revealed relatively low level of genetic diversity. Similar to other taxonomic groups [30–32], the lowest mean nucleotide diversity (0.0182) was detected in the IR region, while the single copy regions were more diverse, with Pi values of 0.0415 and 0.0453 for LSC and SSC, respectively (Figure 5).
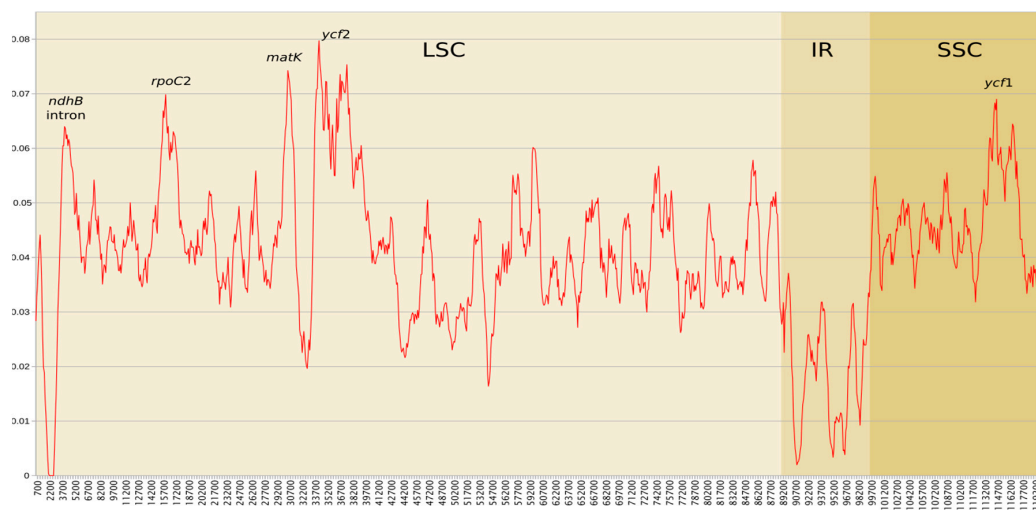

Figure 5. Nucleotide diversity (Pi) by sliding window analysis in the aligned whole plastomes (excluding second IR) of 14 Marchantiidae specimens. Window length 1000 bp, step size 100bp.

The IR that were located genes and spacers are known from their lower substitution rates in comparison to those located in single copy parts of plastomes [6,7,13,33]. Genes that are transferred from SC regions to IRs tend to decrease substitution rates due to copy-dependent repair activity [13,34]. Lower nucleotide diversity of IR regions is also observed in liverworts and mosses, but available data are still very limited [21,31]. The comparative analysis of six *Aneura* plastomes revealed 4.6-fold decrease of IR nucleotide diversity in comparison to single copy regions [21]. Similar values were also obtained from plastomes of Orthotrichaceae moss family [31]. The plastomes of Marchantiidae that were sequenced and analysed in this study revealed lower nucleotide diversity than among species of *Aneura* genus, however the Marchantiidae are known from low evolutionary rates of their organellar

genomes [12]. In case of genes *rps*7 and *rps*12 that are transferred to IR of *C. salebrosum* plastome it is too early to conclude whether this transfer influences their substitution ratio, since the plastome sequences of only four individuals are known. In the single copy regions, the highest value was found for *ycf*2 gene (pi = 0.008), which is often considered as a promising barcode in many other taxa [7,35]. The second (*matK*) and third (*rpoC*2) most variable regions of Marchantiidae plastome also prove their potential as barcodes in many plant groups and are among most commonly used regions in plant barcoding [36–39]. The remaining regions with pi > 0.006, the *ndh*B intron in LSC and *ycf*1 gene in SSC were also previously considered as potential barcodes in plants [40,41].

In the case of five species from three genera (*Conocephalum*, *Marchantia* and *Riccia*), at least two individuals were sequenced, thus enabling barcoding analyses, which revealed the presence of significant barcoding gap for all the species (Figure. S3).

The highest interspecific distances were detected in the case of *Riccia fluitans* (K2P distance 9.7 percent) with *Marchantia* in the middle (K2P distance ranged from 2.55 to 2.58 percent) and the lowest in the case of *Conocephalum* species (K2P distance 1.05 percent). These results correspond well with the estimated divergence times of *Marchantia* and *Conocephalum* species, which are 40 Mya and 15 Mya, respectively [12]. The sliding window analyses across plastomes revealed the lowest K2P distances and the lowest numbers of diagnostic nucleotides across the IR region (Figure 6A,E). The IR region was also characterised by the greatest number of zero cell in K2P distance matrix and the highest proportion of zero non-conspecific K2P distances (Figure 6B,C). The above parameters are also correlated with the disturbance of phylogenetic signal at the IR regions, which is the main source of the incongruent NJ trees (Figure 6D). None of 18 detected 500 bp frames that enable molecular delimitation of all analysed species were found in IR, three of them were located in SSC, while also remaining in LSC regions (Figure 6F).
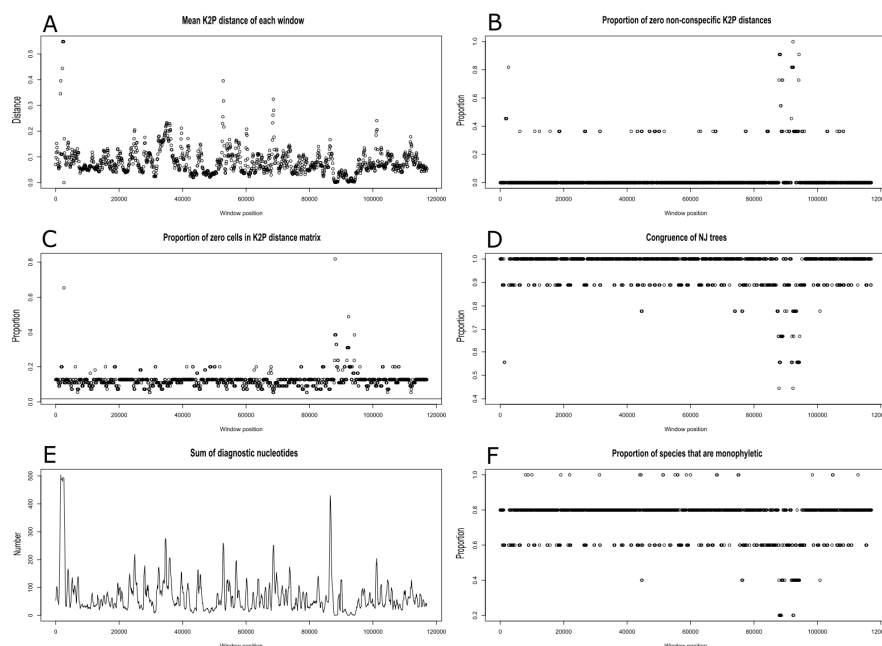

Figure 6. Results of several analyses across the plastome sequences of five Marchantiidae species with at least two sequenced specimens using the sliding window method. (**A**)—the plot of the mean Kimura 2-parameter distance matrix for each 500 bp-long window. (**B**)—the proportion of zero non-conspecific distances. (**C**)—the proportion of zero cells in the distance matrix. The unbroken horizontal line crossing the y-axis at 0 is the proportion of zero cells in the distance matrix created from the full dataset. (**D**)—the proportion of clades that are identical between the windows and the full dataset. (**E**)—the sum of diagnostic nucleotide positions for all species. (**F**)—the proportion of species that are monophyletic.

The genetic diversity of complex thalloid liverwort plastomes seems to be lower in comparison to simply thalloid and leafy taxa [12]; however, available genomic resources are still very limited. The complete plastome sequences are known from the four genera of complex thalloids, including *Marchantia* [42], *Dumortiera* [43], *Reboulia*, *Riccia* [29], and four species sequenced in this study. The sequenced genomes of simple thallioids are limited to genera *Pellia* [44] and *Aneura* [21,45], while the sequenced complete leafy liverwort plastomes comprise 14 species [9,10,29].

3. Materials and Methods

3.1. DNA Extraction

The total genomic DNA of liverworts were extracted from fresh tissue stored in−20 °C. The DNA was extracted using the Qiagen Plant MiniSpin (Qiagen, Germany). Ca 1 cm^2 of cleaned thallus was ground with silica beads in a MiniBead-Beater homogenizer for 50 s and subsequently processed according to the manufacturer protocol. The DNA of *Riccia fluitans* and *Sphaerocarpos texanus* were extracted from 1 cm long thalli from herbarium specimens.

The DNA quantity was estimated using Qubit fluorometer and Qubit™ dsDNA BR Assay Kit (Invitrogen, Carsbad, NM, USA). DNA quality was checked by electrophoresis in 0.5% agarose gel that was stained with Euryx Simple Save (Eurx, Gdańsk, Poland). Table S1 provides sample specimen details.

The extracted DNA of *Conocephalum* species prior to nanopore sequencing was carefully examined and additionally cleaned while using Genomic DNA Clean and Concentrator kit (Zymo, Irvine, CA, USA).

3.2. Library Construction and Sequencing

The libraries for short-read sequencing were prepared using Qiagen FX library preparation kit according to the manufacturer protocol and sequenced while using HiSeq X platform (Illumina) by Macrogen Inc. (Seoul, Korea) to generate 150 bp paired-end reads with a 350 bp insert between paired reads.

The long-read libraries were constructed using Rapid sequencing kit RAD-004 (Oxford Nanopore, UK), according to manufacturer's protocol and sequenced using the MinION MK1b portable device (Oxford Nanopore) with default settings and Guppy basecaller for improved read accuracy.

The nanopore reads were polished using the hybrid read error correction method in order to use sequencing reads of the best quality. First, the Burrows–Wheeler Transform (BWT) of the short-read Illumina dataset was constructed while using ropeBWT2 [46]. Next, FMLRC [47] was used to build the FM-index and correct errors occurring within nanopore reads.

3.3. Genomes Assembling, Annotation and Comparative Analyses

Sequencing reads were cleaned by removing the adaptor sequences and low-quality reads with Trimmomatic v0.36 [48] and assembled while using SPAdes 3.12.0 [49]. Reference plastome sequences of *Marchantia polymorpha* (NC035977.1) were used to identify the organellar genomes of Marchantiidae among the generated contigs. The detailed workflow was similar to that previously published [50].

The initial assemblies of *Conocephalum* plastomes revealed several long TA-rich regions, which made reliable assembly using only short reads difficult, therefore the nanopore sequencing was applied to confirm the assumed linkage between contigs flanking by "TA" motifs. The polished nanopore reads were mapped on Illumina reads based contig flanked by microsatellites motifs using Geneious Prime 2019 build-in mapper with high-sensitivity settings and minimum overlap set to 2000 bp. The consensus sequenced with nanopore overlaps longer than flanking SSR enabled the proper orientation of input contigs. The obtained circular assemblies were revalidated by mapping short-reads libraries in Geneious Prime 2019.

Annotated organellar genomes GenBank files were used to draw gene maps while using the OrganellarGenome DRAW 1.31 tool [51], and the maps were examined for further comparisons of gene order and content.

The four junctions between single-copy segments and inverted repeats in *Riccia fluitans* and *Sphaerocapros texanus* were confirmed using PCR-based product sequencing of the assembled genomes. Supplementary File 1 provides the amplifications conditions, used reagents, and primers. Purified PCR products were sequenced in both directions using the ABI BigDye 3.1 Terminator Cycle Kit (Applied Biosystems, Foster City, CA, USA) and visualized with the ABI Prism 3130 Automated DNA Sequencer (Applied Biosystems). The sequences that were obtained with the Sanger method were aligned with the assembled genomes using the Geneious Prime 2019 assembly software to check for any differences.

The chloroplast microsatellite markers were identified using the msatcommander application [52] with default settings.

3.4. Phylogenomics Analyses

Phylogenetic analyses were performed using the chloroplast genomes of 32 species, including all known and newly sequenced plastomes, complex thalloids, and selected species of remaining liverworts lineages. MAFFT software [53] was used to create an alignment of LSC, IR, and SSR chloroplast regions. Next, with the use of PartitionFinder2 [54], the best partitioning schemes and corresponding substitution models of the alignment were estimated. Afterwards, Bayesian and ML analyses were conducted using MrBayes 3.2.6 [55] and RaxML [56] software, respectively, based on the alignment and obtained models. The MCMC algorithm was run for 5,000,000 generations (sampling every 500) with four incrementally heated chains (starting from random trees). The Tracer 1.7.1 [57] software was used to determine the number of generations needed to reach stationarity, which occurred at approximately 300,000 generations. Therefore, the first 600 trees were discarded as burn-in, and the remaining trees were used to develop a Bayesian consensus tree. The ML analysis was performed with the partitioning scheme and nucleotide substitution blocks that were generated by PartitionFinder2, using default parameters. Bootstrap analyses were performed with 2000 replicates to assess the nodal support. The phylogenetics analyses were conducted while using three datasets: the coding regions of plastomes for all of the analysed liverworts, complete plastome sequences of Marchantiidae (since its unable to provide reliable alignment of intergenic regions among phylogenetically distant liverworts lineages), and the dataset containing complete single copy regions of Marchantiidae. The latter was performed due to the detected incongruence in phylogenetic signal among single copy regions and inverted repeats.

3.5. Diversity Analyses

Available and newly sequenced Marchantiidae chloroplast genomes were aligned using the MAFFT v5 genome aligner [53]. Visually checked alignments were imported in the R environment and analysed using the PopGenome package [58]. The nucleotide diversity (Pi) was calculated for each of the plastome region (LSC, IR and SSC) and using a sliding window approach with 1000 bp window length and 100 bp step. Barcoding analyzes of *Conocephalum*, *Marchantia* and *Riccia* plastomes were performed in Spider software [59] using 500 bp-long sliding window with 100 bp step.

4. Conclusions

The result of our study revealed the increase of SSR during the diversification of complex thalloid liverworts, which, in the case of SSRs located near IR-LSC boundaries, leads to the expansion of IRs via homologous recombination. The growing length and number of dinucleotide repeats, which often override the length of commonly used Illumina 100–150 bp long reads, is pushing towards the new long-read sequencing technologies, like nanopore sequencing, which enable reliable structural assembly of repeat-rich molecules.

The newly sequenced genomes confirmed the stable structure of liverwort plastomes, with one exception for *Conocephalum salebrosum*, where the significant expansion of IRs was detected for the first time. This expansion might be the result of SSR accumulation during the diversification of Marchantiidae. Despite being SSR rich, the plastomes of complex thalloids are less diverse than other liverwort lineages. The low genetic diversity of Marchantiales chloroplast genomes does not exclude their use as a good source of barcodes. Screening for potential barcodes revealed 18 optimal, 500 bp long frames. Each of them enables the molecular identification of all the analysed taxa. The presented study also points out another important aspect of phylogenetics studies—the larger datasets based on complete genomes do not always resolve evolutionary relationships better. In this particular case, an incongruent phylogenetic signal came out from slowly evolving inverted repeat regions.

Supplementary Materials: The following are available at http://www.mdpi.com/2073-4425/11/3/299/s1. Table S1: The list of specimens used in this study including sequencing results and numbers of SSRs. Table S2: The number and length of SSRs detected in analyzed plastomes, Supplementary File 1: The conditions of PCR reaction used during study, Figure S1: Read coverage LSC-IR border of *Conocephalum salebrosum* chloroplast genome. Figure S2: Phylogram of Marchantiidae derived from a Maximum Likelihood analysis of complete chloroplast sequences. The posterior probabilities values lower than 1 and bootstrap values lower than 100% are shown. The Bayesian analysis resulted in the tree with identical topology but unsupported (0.72 posterior probability) *Dumortiera-Reboulia* clade. Figure S3: Line plot of the barcode gap for the five species of Marchantiales. For each individual in the dataset, grey lines represent the furthest intraspecific distance (bottom of line value), and the closest interspecific distance (top of line value).

Author Contributions: Conceptualization, J.S.; Methodology, J.S., K.K. and P.M; Software, K.M. and M.S; Validation, J.S., and K.K.; Formal analysis, J.S, M.S.; Investigation, J.S, K.K, M.Ś. And P.M.; resources, A.B., K.B, P.G.; Data curation, K.M.; Writing—original draft preparation, J.S.; Writing—review and editing, K.B., P.M., M.Ś. and K.M.; Visualization, P.M, J.S.; Supervision, J.S.; Project administration, J.S.; Funding acquisition, J.S., A.B. and K.B. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

LSC	Large Singe Copy Region
SSC	Small Single Copy Region
IR	Inverted Repeats Regions
Bp	Base Pair
BS	Bootstrap Support
SSR	Simple Sequence Repeats
ML	Maximum Likelihood
MCMC	Monte Carlo Markov Chains
NJ	Neighbour-Joining

References

1. Vitelli, M.; Vessella, F.; Cardoni, S.; Pollegioni, P.; Denk, T.; Grimm, G.W.; Simeone, M.C. Phylogeographic structuring of plastome diversity in Mediterranean oaks (Quercus Group Ilex, Fagaceae). *Tree Genet. Genomes* **2017**, *13*. [CrossRef]
2. Huang, B.; Ruess, H.; Liang, Q.; Colleoni, C.; Spooner, D.M. Analyses of 202 plastid genomes elucidate the phylogeny of Solanum section Petota. *Sci. Rep.* **2019**, *9*, 1–7. [CrossRef]
3. Magdy, M.; Ou, L.; Yu, H.; Chen, R.; Zhou, Y.; Hassan, H.; Feng, B.; Taitano, N.; van der Knaap, E.; Zou, X.; et al. Pan-plastome approach empowers the assessment of genetic variation in cultivated Capsicum species. *Sci. Rep.* **2019**, *8*, 665–677. [CrossRef] [PubMed]
4. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biol. Rev.* **2015**, *90*, 157–166. [CrossRef]

5.   Szczecińska, M.; Łazarski, G.; Bilska, K.; Sawicki, J. The complete plastid genome and nuclear genome markers provide molecular evidence for the hybrid origin of pulsatilla × hackelii Pohl. *Turk. J. Bot.* **2017**, *41*, 329–337. [CrossRef]

6.   Krawczyk, K.; Nobis, M.; Myszczyński, K.; Klichowska, E.; Sawicki, J. Plastid super-barcodes as a tool for species discrimination in feather grasses (Poaceae: Stipa). *Sci. Rep.* **2018**, *8*, 1–10. [CrossRef]

7.   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.* **2015**, *16*, 22258–22279. [CrossRef]

8.   Nevill, P.G.; Howell, K.A.; Cross, A.T.; Williams, A.V.; Zhong, X.; Tonti-Filippini, J.; Boykin, L.M.; Dixon, K.W.; Small, I. Plastome-wide rearrangements and gene losses in carnivorous droseraceae. *Genome Biol. Evol.* **2019**, *11*, 472–485. [CrossRef]

9.   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.* **2011**, *144*, 29–43. [CrossRef]

10. Myszczyński, K.; Górski, P.; Ślipiko, M.; Sawicki, J. Sequencing of organellar genomes of Gymnomitrion concinnatum (Jungermanniales) revealed the first exception in the structure and gene order of evolutionary stable liverworts mitogenomes. *BMC Plant Biol.* **2018**, *18*, 1–12. [CrossRef]

11. Yu, Y.; Liu, H.M.; Yang, J.B.; Ma, W.Z.; Pressel, S.; Wu, Y.H.; Schneider, H. Exploring the plastid genome disparity of liverworts. *J. Syst. Evol.* **2019**, *57*, 382–394. [CrossRef]

12. Villarreal A., J.C.; Crandall-Stotler, B.J.; Hart, M.L.; Long, D.G.; Forrest, L.L. Divergence times and the evolution of morphological complexity in an early land plant lineage (Marchantiopsida) with a slow molecular rate. *New Phytol.* **2016**, *209*, 1734–1746. [CrossRef] [PubMed]

13. Park, M.; Park, H.; Lee, H.; Lee, B.H.; Lee, J. The complete plastome sequence of an antarctic bryophyte Sanionia uncinata (Hedw.) loeske. *Int. J. Mol. Sci.* **2018**, *19*, 709. [CrossRef] [PubMed]

14. Sugiura, C.; Kobayashi, Y.; Aoki, S.; Sugita, C.; Sugita, M. Complete chloroplast DNA sequence of the moss Physcomitrella patens: Evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. *Nucleic Acids Res.* **2003**, *31*, 5324–5331. [CrossRef]

15. Plunkett, G.M.; Downie, S.R. Expansion and Contraction of the Chloroplast Inverted Repeat in Apiaceae Subfamily Apioideae. *Syst. Bot.* **2000**, *25*, 648. [CrossRef]

16. Markham, K.R.; Porter, L.J.; Mues, R.; Zinsmeister, H.D.; Brehm, B.G. Flavonoid variation in the liverwort Conocephalum conicum: Evidence for geographic races. *Phytochemistry* **1976**, *15*, 147–150. [CrossRef]

17. Szweykowski, J.; Krzakowa, M. Variation of four enzyme systems in Polish populations of Conocephalum conicum (L) Dumort (Hepaticae Marchantiales). *Bull. Acad. Pol. Sci., Ser. Sci. Biol.* **1979**, *27*, 37–41.

18. Szweykowski, J.; Buczkowska, K.; Odrzykoski, I.J. Conocephalum salebrosum (Marchantiopsida, Conocephalaceae)—A new Holarctic liverwort species. *Plant Syst. Evol.* **2005**, *253*, 133–158. [CrossRef]

19. Dong, S.; Zhao, C.; Zhang, S.; Wu, H.; Mu, W.; Wei, T.; Li, N.; Wan, T.; Liu, H.; Cui, J.; et al. The Amount of RNA Editing Sites in Liverwort Organellar Genes Is Correlated with GC Content and Nuclear PPR Protein Diversity. *Genome Biol. Evol.* **2019**, *11*, 3233–3239. [CrossRef]

20. Myszczyński, K.; Ślipiko, M.; Sawicki, J. Potential of transcript editing across mitogenomes of early land plants shows novel and familiar trends. *Int. J. Mol. Sci.* **2019**, *20*, 2963. [CrossRef]

21. Myszczyński, K.; Bączkiewicz, A.; Buczkowska, K.; Ślipiko, M.; Szczecińska, M.; Sawicki, J. The extraordinary variation of the organellar genomes of the Aneura pinguis revealed advanced cryptic speciation of the early land plants. *Sci. Rep.* **2017**, *7*, 1–12. [CrossRef] [PubMed]

22. Ohyama, K.; Fukuzawa, H.; Kohchi, T.; Shirai, H.; Sano, T.; Sano, S.; Umesono, K.; Shiki, Y.; Takeuchi, M.; Chang, Z.; et al. Chloroplast gene organization deduced from complete sequence of liverwort marchantia polymorpha chloroplast DNA. *Nature* **1986**, *322*, 572–574. [CrossRef]

23. George, B.; Bhatt, B.S.; Awasthi, M.; George, B.; Singh, A.K. Comparative analysis of microsatellites in chloroplast genomes of lower and higher plants. *Curr. Genet.* **2015**, *61*, 665–677. [CrossRef] [PubMed]

24. Brandström, M.; Bagshaw, A.T.; Gemmell, N.J.; Ellegren, H. The relationship between microsatellite polymorphism and recombination hot spots in the human genome. *Mol. Biol. Evol.* **2008**, *25*, 2579–2587. [CrossRef] [PubMed]

25. Weng, M.L.; Blazier, J.C.; Govindu, M.; Jansen, R.K. Reconstruction of the ancestral plastid genome in geraniaceae reveals a correlation between genome rearrangements, repeats, and nucleotide substitution rates. *Mol. Biol. Evol.* **2014**, *31*, 645–659. [CrossRef] [PubMed]

26. Sabir, J.S.M.; Yu, M.; Ashworth, M.P.; Baeshen, N.A.; Baeshen, M.N.; Bahieldin, A.; Theriot, E.C.; Jansen, R.K. Conserved gene order and expanded inverted repeats characterize plastid genomes of Thalassiosirales. *PLoS ONE* **2014**, *9*, e0107854. [CrossRef]

27. Weng, M.L.; Ruhlman, T.A.; Jansen, R.K. Expansion of inverted repeat does not decrease substitution rates in Pelargonium plastid genomes. *New Phytol.* **2017**, *214*, 842–851. [CrossRef]

28. Rabah, S.O.; Shrestha, B.; Hajrah, N.H.; Sabir, M.J.; Alharby, H.F.; Sabir, M.J.; Alhebshi, A.M.; Sabir, J.S.M.; Gilbert, L.E.; Ruhlman, T.A.; et al. Passiflora plastome sequencing reveals widespread genomic rearrangements. *J. Syst. Evol.* **2019**, *57*, 1–14. [CrossRef]

29. Yu, Y.; Yang, J.B.; Ma, W.Z.; Pressel, S.; Liu, H.M.; Wu, Y.H.; Schneider, H. Chloroplast phylogenomics of liverworts: A reappraisal of the backbone phylogeny of liverworts with emphasis on Ptilidiales. *Cladistics* **2019**, 1–10. [CrossRef]

30. Choi, S.J.; Kim, Y.; Choi, C. Chloroplast genome-based hypervariable markers for rapid authentication of six Korean pyropia species. *Diversity* **2019**, *11*, 220. [CrossRef]

31. Mizia, P.; Myszczyński, K.; Ślipiko, M.; Krawczyk, K.; Plášek, V.; Szczecińska, M.; Sawicki, J. Comparative plastomes analysis reveals the first infrageneric evolutionary hotspots of orthotrichum s.l. (Orthotrichaceae, Bryophyta). *Turk. J. Bot.* **2019**, *43*, 444–457. [CrossRef]

32. Somaratne, Y.; Guan, D.L.; Abbood, N.N.; Zhao, L.; Wang, W.Q.; Xu, S.Q. Comparison of the complete eragrostis pilosa chloroplast genome with its relatives in eragrostideae (Chloridoideae; poaceae). *Plants* **2019**, *8*, 485. [CrossRef] [PubMed]

33. Perry, A.S.; Wolfe, K.H. Nucleotide substitution rates in legume chloroplast DNA depend on the presence of the inverted repeat. *J. Mol. Evol.* **2002**, *55*, 501–508. [CrossRef] [PubMed]

34. Zhu, A.; Guo, W.; Gupta, S.; Fan, W.; Mower, J.P. Evolutionary dynamics of the plastid inverted repeat: The effects of expansion, contraction, and loss on substitution rates. *New Phytol.* **2016**, *209*, 1747–1756. [CrossRef]

35. Zhang, W.; Zhao, Y.; Yang, G.; Peng, J.; Chen, S.; Xu, Z. Determination of the evolutionary pressure on Camellia oleifera on Hainan Island using the complete chloroplast genome sequence. *PeerJ* **2019**, *2019*. [CrossRef]

36. Logacheva, M.D.; Penin, A.A.; Samigullin, T.H.; Vallejo-Roman, C.M.; Antonov, A.S. Phylogeny of flowering plants by the chloroplast genome sequences: In search of a "lucky gene". *Biochemistry* **2007**, *72*, 1324–1330. [CrossRef]

37. Ford, C.S.; Ayres, K.L.; Toomey, N.; Haider, N.; Van Alphen Stahl, J.; Kelly, L.J.; Wikström, N.; Hollingsworth, P.M.; Duff, R.J.; Hoot, S.B.; et al. Selection of candidate coding DNA barcoding regions for use on land plants. *Bot. J. Linn. Soc.* **2009**, *159*, 1–11. [CrossRef]

38. Kuang, D.Y.; Wu, H.; Wang, Y.L.; Gao, L.M.; Zhang, S.Z.; Lu, L. Complete chloroplast genome sequence of Magnolia kwangsiensis (Magnoliaceae): Implication for DNA barcoding and population genetics. *Genome* **2011**, *54*, 663–673. [CrossRef]

39. Krawczyk, K.; Szczecińska, M.; Sawicki, J. Evaluation of 11 single-locus and seven multilocus DNA barcodes in Lamium L. (Lamiaceae). *Mol. Ecol. Resour.* **2014**, *14*, 272–285. [CrossRef]

40. Dong, W.; Liu, J.; Yu, J.; Wang, L.; Zhou, S. Highly variable chloroplast markers for evaluating plant phylogeny at low taxonomic levels and for DNA barcoding. *PLoS ONE* **2012**, *7*, e0035071. [CrossRef]

41. Dong, W.; Xu, C.; Li, C.; Sun, J.; Zuo, Y.; Shi, S.; Cheng, T.; Guo, J.; Zhou, S. ycf1, the most promising plastid DNA barcode of land plants. *Sci. Rep.* **2015**, *5*, 8348. [CrossRef] [PubMed]

42. Kwon, W.; Kim, Y.; Park, J. The complete chloroplast genome of Korean Marchantia polymorpha subsp. ruderalis Bischl. & Boisselier: Low genetic diversity between Korea and Japan. *Mitochondrial DNA Part B* **2019**, *4*, 959–960. [CrossRef]

43. Kwon, W.; Kim, Y.; Park, J. The complete chloroplast genome sequence of Dumortiera hirsuta (Sw.) Nees (Marchantiophyta, Dumortieraceae). *Mitochondrial DNA Part B* **2019**, *4*, 318–319. [CrossRef]

44. 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.* **2012**, *4*, 1349–1357. [CrossRef]

45. Wickett, N.J.; Zhang, Y.; Hansen, S.K.; Roper, J.M.; Kuehl, J.V.; Plock, S.A.; Wolf, P.G.; Depamphilis, C.W.; Boore, J.L.; Goffinet, B. Functional gene losses occur with minimal size reduction in the plastid genome of the parasitic liverwort Aneura mirabilis. *Mol. Biol. Evol.* **2008**, *25*, 393–401. [CrossRef]

46. Li, H. Fast construction of FM-index for long sequence reads. *Bioinformatics* **2014**, *30*, 3274–3275. [CrossRef]

47. Wang, J.R.; Holt, J.; McMillan, L.; Jones, C.D. FMLRC: Hybrid long read error correction using an FM-index. *BMC Bioinform.* **2018**, *19*, 1–11. [CrossRef]

48. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

49. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]

50. 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.* **2014**, *38*, 1229–1238. [CrossRef]

51. Greiner, S.; Lehwark, P.; Bock, R. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Res.* **2019**, *47*, W59–W64. [CrossRef] [PubMed]

52. Faircloth, B.C. MSATCOMMANDER: Detection of microsatellite repeat arrays and automated, locus-specific primer design. *Mol. Ecol. Resour.* **2008**, *8*, 92–94. [CrossRef] [PubMed]

53. Katoh, K.; Kuma, K.I.; Toh, H.; Miyata, T. MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res.* **2005**, *33*, 511–518. [CrossRef] [PubMed]

54. Lanfear, R.; Frandsen, P.B.; Wright, A.M.; Senfeld, T.; Calcott, B. Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* **2017**, *34*, 772–773. [CrossRef] [PubMed]

55. Huelsenbeck, J.P.; Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **2001**, *17*, 754–755. [CrossRef] [PubMed]

56. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]

57. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **2018**, *67*, 901–904. [CrossRef]

58. 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.* **2014**, *31*, 1929–1936. [CrossRef]

59. Brown, S.D.J.; Collins, R.A.; Boyer, S.; Lefort, M.C.; Malumbres-Olarte, J.; Vink, C.J.; Cruickshank, R.H. Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Mol. Ecol. Resour.* **2012**, *12*, 562–565. [CrossRef]