



The chloroplast genome of the desiccation-tolerant moss *Pseudocrossidium replicatum* (Taylor) R.H. Zander

Miguel A. Cevallos¹ , Gabriela Guerrero¹ , Selma Ríos² , Analilia Arroyo² , Miguel Angel Villalobos²  and Helena Porta³ 

¹*Centro de Ciencias Genómicas, Programa de Genómica Evolutiva, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico.*

²*Centro de Investigación en Biotecnología Aplicada, Laboratorio de Biología Molecular y Biotecnología de Plantas, Instituto Politécnico Nacional. Tepetitla de Lardizabal, Tlaxcala, Mexico.*

³*Instituto de Biotecnología, Departamento de Biología Molecular de Plantas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico.*

Abstract

Mosses in conjunction with hornworts and liverworts are collectively referred to as bryophytes. These seedless, nonvascular plants are the closest extant relatives of early terrestrial plants and their study is essential to understand the evolutionary first steps of land plants. Here we report the complete chloroplast (cp) genome sequence of *Pseudocrossidium replicatum*, a moss belonging to the Pottiaceae family that is common in the central highlands of Mexico, in South America, in southern USA, and in Kenya. The cp genome (plastome) of *P. replicatum* is 123,512 bp in size, comprising inverted repeats of 9,886 bp and single-copy regions of 85,146 bp (LSC) and 18,594 bp (SSC). The plastome encodes 82 different proteins, 31 different tRNAs, and 4 different rRNAs. Phylogenetic analysis using 16 cp protein-coding genes demonstrated that *P. replicatum* is closely related to *Syntrichia ruralis*, and the most basal mosses are *Takakia lepidozioides* followed by *Sphagnum palustre*. Our analysis indicates that during the evolution of the mosses' plastome, eight genes were lost. The complete plastome sequence reported here can be useful in evolutionary and population genetics.

Keywords: Bryophytes, Pottiaceae, chloroplast, next-generation sequencing.

Received: June 26, 2018; Accepted: November 14, 2018.

The chloroplast (cp) is not only the organelle where photosynthesis takes place, but also the site where a wide variety of functions essential for plant metabolism occur, like nitrate and sulfate assimilation, synthesis of almost all amino acids and fatty acids, chlorophylls, and carotenoids (Jensen and Leister, 2014). The cp genome (plastome) consists of a single circular double stranded DNA molecule, and for most of the land plants its structure is well conserved. The plastome can be divided into four regions: two copies of an inverted repeat (IR) region, a large-single-copy (LSC) region, and a small-single-copy (SSC) region. This pattern is maintained through a flip-flop homologous recombination mechanism (Palmer *et al.*, 1983). Gene number and content of plastomes is highly conserved, however, the sequence of the non-coding intergenic spacer regions varies to a large extent. In several plant lineages, comparative genomics studies have shown structural changes that

include gene or intron losses, loss of an IR region, inversions, and other kinds of genome rearrangements (Yurina *et al.*, 2017). Considering these features, together with the cp uniparental mode of inheritance and that their genomes are almost recombination free, the study of cp genome sequences is an invaluable tool to resolve taxonomic issues and to understand the origin and evolution of the plant kingdom.

Bryophytes are a paraphyletic group of land plants that comprises hornworts, liverworts, and mosses. A common characteristic of this heterogeneous group is that its sporophyte (diploid phase) is unbranched and contains a single sporangium that is attached to the gametophyte, the photosynthetic and haploid phase of the bryophyte life cycle. Bryophytes are the most primitive terrestrial group of plants that have recently emerged as models for plant biology studies. Particularly mosses have been receiving more attention because their gametophyte is haploid, they display low anatomical complexity, and have a rapid cell cycle (Cove, 2005). Mosses are a highly diverse clade containing more than 12,800 species (Delgadillo-Moya, 2014). The

Send correspondence to Helena Porta. Instituto de Biotecnología, Departamento de Biología Molecular de Plantas, Universidad Nacional Autónoma de México, Apartado postal 510-3, C.P. 62250, Cuernavaca, Morelos, México. E-mail: helena@ibt.unam.mx.

growth and sexual reproduction of these land plants is limited to the sites where water is available, however, they possess mechanisms that allow their survival during the dry season. Today, around 1750 complete plastomes of land plants have been determined (NCBI, 2016): 15 correspond to Bryophyte cp, and of these only nine belong to mosses (Rensing *et al.*, 2009). To have better understanding about the origin and evolution of the first land plants many more cp sequences are needed.

Pseudocrossidium replicatum, belongs to the moss family Pottiaceae and can be found in the central highlands of Mexico, South America, southern USA, and Kenya. In the work presented here and with the aim to expand our understanding of these primitive land plants, we present the complete genome sequence of the *P. replicatum* plastome and the comparison of its structure with the other moss cp genomes.

Our *P. replicatum* specimen (sporophyte and gametophyte) was collected in Ixtacuixtla, Tlaxcala, Mexico (19-20-03.3 N, 98-21-59.9 W, 2159 msnm), in January 2008. An *in vitro* culture was started from a single spore and maintained since then through weekly subcultures in a growth chamber at 23 °C and 30% of relative humidity, with a photoperiod of 16/8 h light/dark. The genomic DNA used in this study was extracted from 7 day-old protonemata, cultivated *in vitro* in PpNH4 culture medium, using the ZR Plant/Seed DNA Microprep kit and following the manufacturer's instructions (ZymoResearch). A DNA pair-end library (PE 2 x 75) for Illumina NextSeq500 platform was prepared and sequenced at Unidad Universitaria de Secuenciación Masiva de DNA de la Universidad Nacional Autónoma de México. Reads were assembled with SPAdes-3.9.0 and all contigs that matched the cp genome of *Syntrichia ruralis* were used to orient *P. replicatum* contigs (Bankevich *et al.*, 2012). We selected a minimal contigs set that covered all the *S. ruralis* plastome. Overlapping regions between contigs and borders between the inverted repeat regions were confirmed by PCR and Sanger sequencing. The genome was annotated with the Dual Organellar GenoMe Annotator (DOGMA), BLAST homology search, and tRNAscan-SE (Wyman *et al.*, 2004; Lowe and Chan 2016). The circular genome map was obtained with the Organellar Genome DRAW (Lohse *et al.*, 2013). Protein alignments were done with Clustal_Omega, and a Neighbor joining phylogenetic tree was constructed with MEGA7, using the concatenated alignment of 16 protein sequences (Sievers and Higgins, 2014; Kumar *et al.*, 2016). The complete cp genome of *P. replicatum* with its gene annotation was deposited in NCBI GenBank with accession number MG132071.

The complete cp genome of *P. replicatum* has a total length of 123,512 bp. It possesses two inverted repeats (IRs), each one of 9,886 bp, a large single copy (LSC) region of 85,146 bp, and a small single copy (SSC) region of 18,594 bp (Figure 1). The cp genome of this organism pos-

ses 125 genes, with 82 encoding different proteins, 31 different tRNAs, and four different rRNAs. Among the protein encoding genes, most are involved in photosynthesis functions, including genes encoding the photosystems I and II, products involved in the assembly and stability of photosystem I, the large subunit of Rubisco, cytochrome b6/f complex, cytochrome c synthesis, and the eleven subunits of the NADH-plastoquinone oxidoreductase. Genes encoding the six subunits of the ATP synthase are also present. Additionally, this genome encodes proteins participating in transcription and translation, like the subunits of RNA polymerase, 21 ribosomal proteins, the translational initiation factor 1, and MatK, an intron maturase (Table 1).

The two inverted repeat regions contain the rRNA genes (23S, 16S, 5S and 4.5S) and five tRNA genes, each. Within the SSC region lay seven of the genes encoding the NADH-plastoquinone oxidoreductase along with three genes encoding ribosomal proteins and four genes involved in other functions (*chlL*, *chlN*, *ycf1* and *psaC*). This organization remains fairly constant in the moss plastomes compared here. The rest of the genes are located in the LSC region.

Twelve protein-encoding genes and six different tRNA genes have introns. The largest one, with 2,259 bp, belongs to the *trnK-UUU* genes and is large enough to embrace the complete *matK* gene. Additionally, two genes, *atpF* and *rpoB*, are possibly transcribed and modified by RNA editing.

Until now and including the plastome reported here, there are 10 sequenced moss cp genomes: *Nyholmiella obtusifolia* (NCBI: NC_026979.1), *Orthotrichum rogeri* (NCBI: NC_026212.1), *Syntrichia ruralis* (Oliver *et al.*, 2010), *Physcomitrella patens* (Sugiura *et al.*, 2003), *Sanionia uncinata* (Park *et al.*, 2018), *Sphagnum palustre* (Shaw *et al.*, 2016), *Takakia lepidozoides* (NC_028738) *Tetraphis pellucida* (Bell *et al.*, 2014), and *Tetraplodon fuegianus* (Lewis *et al.*, 2016). These 10 species are representatives of eight different moss families. These plastomes vary in size, gene content and, to a lesser extent, in the location of the genes in the LSC/IRB/SSC/IRA boundary regions.

The length of the *P. replicatum* cp genome (123,512 bp) falls within the range reported for the other moss plastomes, considering that *S. ruralis* has the smallest plastome (122,630 bp) and *T. lepidozoides* the largest (149,016 bp). The total GC content of the *P. replicatum* cp genome is 28.2%, which is similar to that of its closest relative (28.38%). In general, we found that the moss cp genomes have a very similar repertoire of genes, though we were able to identify a few differences in some of them. The largest genomes contain more genes, especially those encoding for proteins: *T. lepidozoides* encodes 91 different genes and *S. palustre* (plastome size of 140,040 bp) encodes 85 distinct proteins. Congruently, the number of protein encoding genes of the *P. replicatum* plastome lies in the aver-

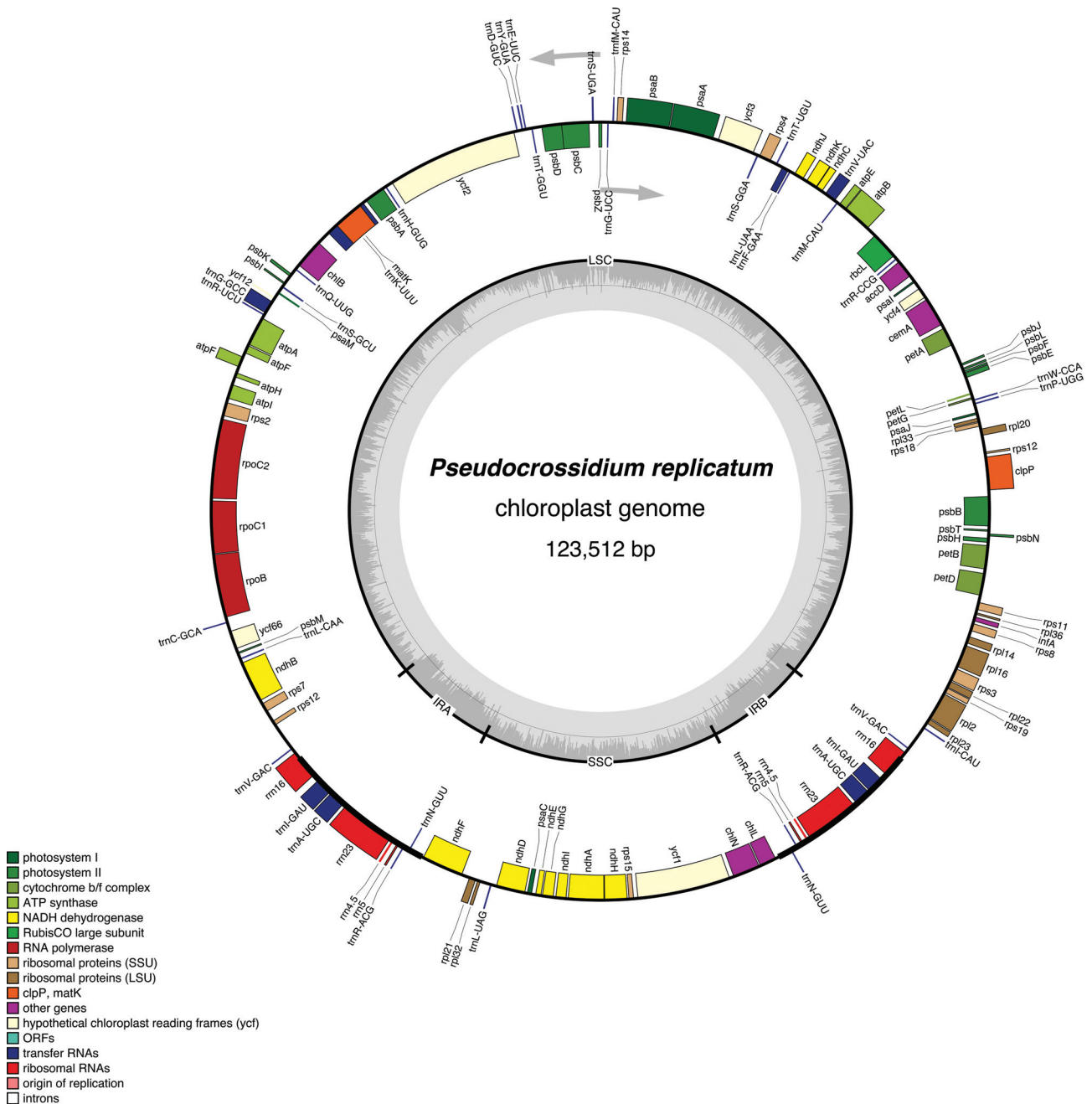


Figure 1 - Gene map of the moss *P. replicatum*. Genes in the inside of the map are transcribed clockwise, and genes in the outside of the map are transcribed counterclockwise. Exons are indicated as closed boxes. Genes with related function are shown in the same color.

age: 82. The number of rRNA genes is the same in all moss plastomes (8 genes). With only one exception, *N. obtusifolia* that contains 32 different tRNAs, the rest encode 31.

To determine the relationships between the moss genomes analyzed here, we constructed a Neighbor joining phylogenetic tree using a concatenated alignment of 16 proteins and including the fern *Osmundastrum cinnamomeum* as outgroup (Figure 2). This figure shows that the most basal chloroplast moss is *T. lepidozoides*, which has the largest plastome and encodes the largest gene

set. The next basal cp genome corresponds to *S. palustre* that possesses the second largest plastome and also encodes the second largest gene set. The *T. lepidozoides* plastome possess eight genes that are absent in the other moss cp genomes (*cysA*, *petN*, *rps12*, *rps16*, *cysT*, *rpoA* *ccsA* and *tufA*), and these probably were transferred to the nucleus in very ancient events. Surprisingly, the *petN* gene is present in *T. lepidozoides* and *P. patens*, suggesting that this gene was lost early in moss evolution, but was regained later by an unknown mechanism in the ancestors of *P. patens*.

Table 1 - Genes encoded in *Pseudocrossidium replicatum* cpDNA.

| Gene categories | Group of genes | Gene names | |
|-------------------------------|-------------------------------------|--|--|
| Photosynthesis | Rubisco | <i>rbcL</i> | |
| | Photosystem I | <i>psaA, psbA, psbC, psal, psaj, psam</i> | |
| | Assembly/stability of photosystem I | <i>ycf1, ycf2, <u>ycf3</u>, ycf4, ycf12, <u>ycf66</u></i> | |
| | Photosystem II | <i>psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbJ, psbK, psbL, psbM, psbN, psbT, psbZ</i> | |
| | ATP Synthase | <i>atpA, atpB, atpE, atpF, atpH, atpI</i> | |
| | Cytochrome b/f complex | <i>petA, <u>petB</u>, <u>petD</u>, petG, petL</i> | |
| | Chlorophyll biosynthesis | <i>chlB, chlL, chlN</i> | |
| | NADPH dehydrogenase | <i><u>ndhA</u>, <u>ndhB</u>, ndhC, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI, ndhJ, ndhK</i> | |
| | Transcription and translation | Transcription | <i>rpoB, <u>rpoC1</u>, rpoC2</i> |
| | | Ribosomal proteins | <i><u>rpl2</u>, rpl14, <u>rpl16</u>, rpl20, rpl21, rpl22, rpl23, rpl32, rpl33, rpl36, rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps15, rps18, rps19</i> |
| Translation initiation factor | | <i>infA</i> | |
| RNA | | Ribosomal RNA | <i>rrn5, rrn16, rrn4.5, rrn23S,</i> |
| | Transfer RNA | <i><u>trnA-UGC</u>, trnC-GCA, trnD-GUC, trnE-UUC, trnF-GAA, <u>trnG-GCC</u>, trnG-UCC, trnH-GUG, trnI-CAU, <u>trnL-GAU</u>, trnK-UUU, trnL-CAA, <u>trnL-UAA</u>, trnL-UAG, trnM-CAU, trnN-GUU, trnP-UGG, trnQ-UUG, trnR-ACG, trnR-CCG, trnR-UCU, trnS-GCU, trnS-GGA, trnS-UGA, trnT-GGU, trnT-UGU, trnV-GAC, <u>trnV-UAC</u>, trnW-CCA, trnY-GUA, trnM-CAU</i> | |
| Other | RNA processing | <i>matK</i> | |
| | Carbon metabolism | <i>cemA</i> | |
| | Proteolysis | <i><u>clpP</u></i> | |
| | Other proteins | <i>accD</i> | |

Underlined genes contain introns.

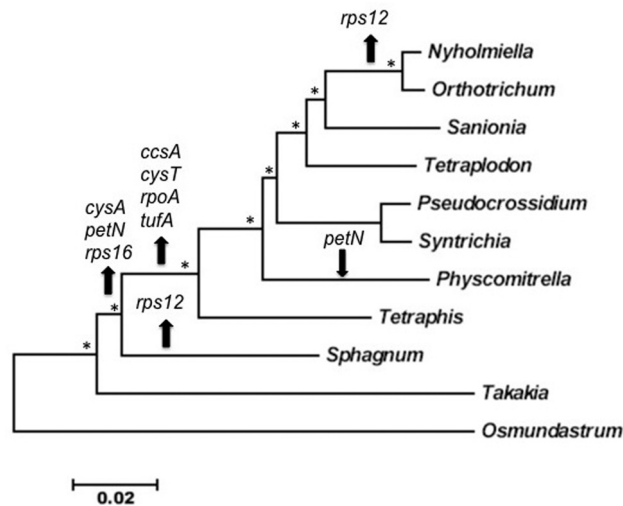


Figure 2 - Phylogenetic tree of chloroplast genomes of mosses. A Neighbor joining tree was constructed using a concatenated alignment of 16 proteins: *accD, atpA, atpB, cemA, chlB, chlL, clpP, ndhB, ndhF, ndhH, ndhK, petA, psaA, psbC, rbcL*, and *rps2*. The chloroplast sequence of the fern *Osmundastrum cinnamomeum* was used as outgroup. Asterisks indicate boot-strap values > 85%. Arrows indicate gene insertion (down) or deletion (up). The scale represents number of amino acid substitution per site.

We found three different patterns in the LSC/IRB/SSC/IRA boundary regions in the moss cp genomes (Figure 3). The first pattern is shared between *T. pellucida*, *N. obtusifolia*, *O. rogeri*, *S. palustre*, *T.*

fuegianus, and *P. replicatum*. The second is represented in the *T. lepidozoides*, *P. patens*, and *S. uncinata* plastomes, and the third pattern pertains only to *S. ruralis*. These patterns are not congruent with the phylogenetic tree, indicat-

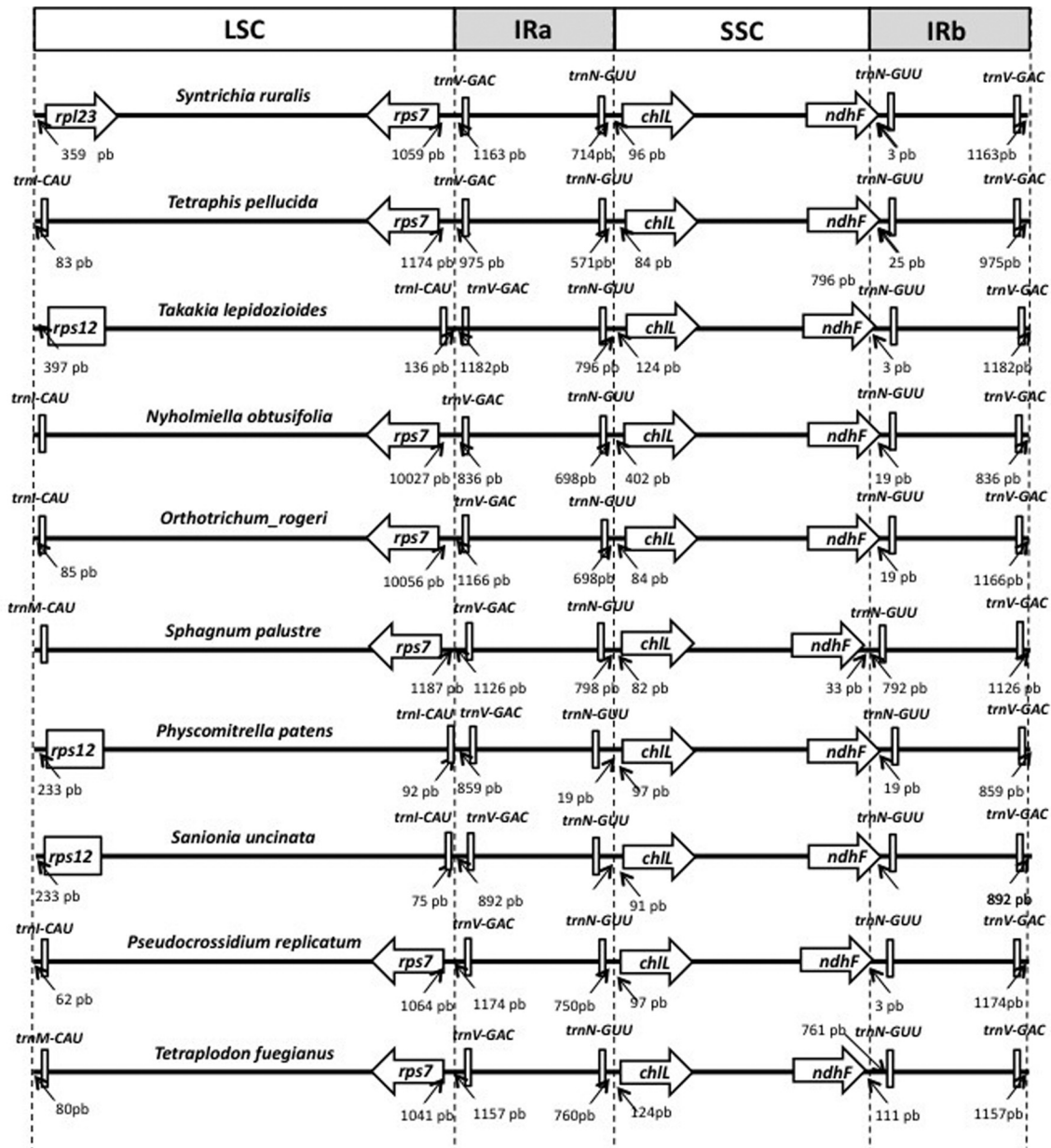


Figure 3 - Comparison of chloroplast genome structure between mosses. Shown are the genome distances and gene position in the vicinity of LSC, IRB, SSC, and IRA in the genome structure of mosses cp.

ing that the LSC/IRB/SSC/IRA boundary regions are constantly changing as a consequence of the flip-flop recombination described earlier.

Acknowledgments

We are grateful to Dr. Michael Dunn for helping us edit our paper and to Dr. Luis Lozano for his advice on bioinformatics. AAB and MAVL are grateful for the financial support of SIP-IPN (20121738 and 20131591) and CONACYT Ciencia Básica 62174.

Conflict of Interest

The authors declare no conflict of interest.

Author contributions

MAC and HP conceived, designed the study and drafted the manuscript, MAC, MAV, AA, SR conducted the experiments and analyzed the data. All authors read and approved the final version.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski

- AD *et al.* (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455-477.
- Bell NE, Boore JL, Mishler BD and Hyvonen J (2014) Organellar genomes of the four-toothed moss, *Tetraphis pellucida*. *BMC Genomics* 15:383.
- Cove D (2005) The moss *Physcomitrella patens*. *Annu Rev Genet* 39:339-358.
- Delgadillo-Moya C (2014) Biodiversity of Bryophyta (mosses) in Mexico. *Rev Mex Biodiv* 85:100-105.
- Jensen PE and Leister D (2014) Chloroplast evolution, structure and functions. *F1000Prime Rep* 6:40.
- Kumar S, Stecher G and Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870-1874.
- Lewis LR, Liu Y, Rozzi R and Goffinet B (2016) Intraspecific variation within and across complete organellar genomes and nuclear ribosomal repeats in a moss. *Mol Phylogenet Evol* 96:195-199.
- Lohse M, Drechsel O, Kahlau S and Bock R (2013) OrganellarGenomeDRAW - a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res* 41:W575-W581.
- Lowe TM and Chan PP (2016) tRNAscan-SE On-line: Integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54-W57.
- Oliver MJ, Murdock AG, Mishler BD, Kuehl JV, Boore JL, Mandoli DF, Everett KD, Wolf PG, Duffy AM and Karol KG (2010) Chloroplast genome sequence of the moss *Tortula ruralis*: Gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. *BMC Genomics* 11:143.
- Palmer JD, Shields CR, Cohen DB and Orton TJ (1983) Chloroplast DNA evolution and the origin of amphidiploid *Brassica* species. *Theor Appl Genet* 65:181-189.
- Park M, Park H, Lee H, Lee BH and Lee J (2018) The complete plastome sequence of an Antarctic bryophyte *Sanionia uncinata* (Hedw.) Loeske. *Int J Mol Sci* 19:709
- Rensing S, Lang D and Zimmer AD (2009) Comparative genomics. *Annu Plant Rev* 36:42-75.
- Shaw JA, Devos N, Liu Y, Cox CJ, Goffinet B, Flatberg KI and Shaw B (2016) Organellar phylogenomics of an emerging model system: *Sphagnum* (peatmoss). *Ann Bot* 118:185-196.
- Sievers F and Higgins DG (2014) Clustal Omega, accurate alignment of very large numbers of sequences. *Methods Mol Biol* 1079:105-116.
- Sugiura C, Kobayashi Y, Aoki S, Sugita C and Sugita M (2003) 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* 31:5324-5331.
- Wyman SK, Jansen RK and Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20:3252-3255.
- Yurina NP, Sharapova LS and Odintsova MS (2017) Structure of plastid genomes of photosynthetic eukaryotes. *Biochemistry (Mosc)* 82:678-691.

Associate Editor: Hong Luo

License information: This is an open-access article distributed under the terms of the Creative Commons Attribution License (type CC-BY), which permits unrestricted use, distribution and reproduction in any medium, provided the original article is properly cited.