The Mitochondrial Genome of the Moss *Physcomitrella patens* Sheds New Light on Mitochondrial Evolution in Land Plants

Kimihiro Terasawa, Masaki Odahara,† Yukihiro Kabeya,*‡ Tatsuhiko Kikugawa,* Yasuhiko Sekine,† Makoto Fujiwara,* and Naoki Sato**

*Department of Life Sciences, Graduate School of Arts and Sciences, University of Tokyo, Komaba, Meguro-ku, Tokyo, Japan; †Department of Life Science, College of Science, Rikkyo University, Ikebukuro, Toshima-ku, Tokyo, Japan; and ‡Center for Gene Research, Nagoya University, Furocho, Chikusa, Nagoya, Japan

The phylogenetic positions of bryophytes and charophytes, together with their genome features, are important for understanding early land plant evolution. Here we report the complete nucleotide sequence (105,340 bp) of the circular-mapping mitochondrial DNA of the moss *Physcomitrella patens*. Available evidence suggests that the multipartite structure of the mitochondrial genome in flowering plants does not occur in *Physcomitrella*. It contains genes for 3 rRNAs (*rnl*, *rns*, and *rrn5*), 24 tRNAs, and 42 conserved mitochondrial proteins (14 ribosomal proteins, 4 *ccm* proteins, 9 nicotinamide adenine dinucleotide dehydrogenase subunits, 5 ATPase subunits, 2 succinate dehydrogenase subunits, apocytochrome *b*, 3 cytochrome oxidase subunits, and 4 other proteins). We estimate that 5 tRNA genes are missing that might be encoded by the nuclear genome. The overall mitochondrial genome structure is similar in *Physcomitrella*, *Chara vulgaris*, *Chaetosphaeridium globosum*, and *Marchantia polymorpha*, with easily identifiable inversions and translocations. Significant synteny with angiosperm and chlorophyte mitochondrial genomes was not detected. Phylogenetic analysis of 18 conserved proteins suggests that the moss–liverwort clade is sister to angiosperms, which is consistent with a previous analysis of chloroplast genes but is not consistent with some analyses using mitochondrial sequences. In *Physcomitrella*, 27 introns are present within 16 genes. Nine of its intron positions are shared with angiosperms and 4 with *Marchantia*, which in turn shares only one intron position with angiosperms. The phylogenetic analysis as well as the syntenic structure suggest that the mitochondrial genomes of *Physcomitrella* and *Marchantia* retain prototype features among land plant mitochondrial genomes.

Introduction

Mitochondria are thought to be descendants of an endosymbiont related to α-Proteobacteria, (Gray et al. 1999; Fitzpatrick et al. 2006), but they are highly diversified in different phyla of photosynthetic organisms (Knoop 2004). Although basic functions of mitochondria have been characterized in yeast and mammals, mitochondria in photosynthetic organisms play various additional important roles in concert with other organelles, such as photorespiration (Mackenzie and McIntosh 1999). The size of the mitochondrial genome increased dramatically in most flowering plants: the mitochondrial genomes of chlorophycean green algae are about 15–96 kbp in size, whereas the mitochondrial genomes (master circles) of flowering plants are about 300–600 kbp (supplementary table 1, Supplementary Material online; for a review, see Bullerwell and Gray 2004). In addition, a reassociation kinetic study suggested that the mitochondrial genomes of some cucurbits are greater than 1 and even 2 Mbp (Ward et al. 1981). Gene content of the mitochondrial genome has been reduced by gene loss and gene transfer to the nucleus during the evolution of angiosperms (Adams et al. 2002). In addition, the mitochondrial genomes of flowering plants are heterogeneous or multipartite because of intramolecular recombination events. This heterogeneity leads to a certain population of mitochondria having imperfect genomes. To ascertain the transmission of correct mitochondrial genetic information, a master circle is thought to be inherited (for reviews, see Backert et al. 1997; Ogihara et al. 2005). However, the phylogenetic origin and lower boundary of such multipartite mitochondrial genomes are still uncertain.

One important problem in comparing mitochondrial genomes in plants is the lack of sequence data of mitochondria in bryophytes, pteridophytes, and gymnosperms. The mitochondrial genomes of various green algae as well as many angiosperms have been sequenced, but *Marchantia polymorpha* (liverwort; Oda et al. 1992) is the only bryophyte in which the mitochondrial genome sequence has been reported. In pteridophytes and gymnosperms, only gene fragments have been used for phylogenetic analysis. To shed light on the evolution of the mitochondrial genome in lower land plants, we sequenced the mitochondrial genome of a moss, *Physcomitrella patens*, for which the chloroplast genome (Sugiura et al. 2003) and preliminary nuclear genomic sequences are available (PHYSCObase: Nishiyama et al. 2003 and http://moss.nibb.ac.jp/).

Introns are common in mitochondria of plants and algae, and detailed comparative analyses have been performed (Burger et al. 1999; Turmel et al. 2003). Generally speaking, angiosperm mitochondria have more introns than do the mitochondria of algae. Both Group 1 and Group 2 introns are found in plant mitochondria. In some studies, introns have been used as phylogenetic markers (Qiu et al. 1998; Pruchner et al. 2002). In the *nad7* gene, we found that the first 2 introns are conserved in *Physcomitrella* and angiosperms, whereas *Marchantia* lacks these introns (Hashimoto and Sato 2001). We still do not know if insertion of introns is a good phylogenetic marker because introns are also lost during evolution, such as the loss of the third intron of the *nad7* gene in tobacco.

Physcomitrella patens is a model organism of lower land plants, in which gene manipulation, notably targeted gene disruption, is feasible. Genomic data are a prerequisite for a detailed expression analysis of mitochondrial genes.

Key words: bryophyte evolution, genome rearrangement, introns, mitochondrial genome, plant phylogeny.

E-mail: naokisat@bio.c.u-tokyo.ac.jp.

Mol. Biol. Evol. 24(3):699–709. 2007
doi:10.1093/molbev/msl198
Advance Access publication December 14, 2006

© The Author 2006. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

However, only a few gene sequences of *Physcomitrella* are known, including *nad7* (Hashimoto and Sato 2001), *nad2* (Beckert et al. 2001), *nad5* (Beckert et al. 1999), and *cox3* (Marienfeld et al. 1991). The mitochondrial genome sequence of *Physcomitrella* reported here gives insight into the phylogeny of land plants and the origin of multipartite mitochondrial genomes in plants. It also provides a basis for studies on the regulation of mitochondrial gene expression.

Materials and Methods

Culture Conditions and DNA Purification

Physcomitrella patens was grown on agar plates (Hashimoto and Sato 2001). Total genomic DNA including organellar DNA was isolated by the Cetyltrimethylammonium bromide (CTAB) method (Draper and Scott 1988). Protoplasts were prepared from protonemata that had been grown for 4 days after inoculation, and mitochondria were isolated from them as described previously (Kabeya et al. 2002; Kabeya and Sato 2005). DNA was isolated from the mitochondria by CsCl density gradient centrifugation essentially according to the procedure used to prepare chloroplast DNA (Sato et al. 1993).

Polymerase Chain Reaction Amplification of Large Fragments of Mitochondrial DNA

In the initial stage of the study, large genomic fragments were amplified by polymerase chain reaction (PCR) and sequenced. These data were used to verify the genomic DNA assembly obtained in shotgun sequencing. First, large fragments of mitochondrial DNA were amplified using the LA PCR kit (Takara, Kyoto, Japan), using the total genomic DNA as a template. Various combinations of primers, developed for known mitochondrial sequences of *Physcomitrella* (*nad7*, *nad2*, *nad5*, and *coxIII*), were used to test if amplification products were produced. Additional primers were synthesized based on highly conserved regions of known mitochondrial genes. The primers are listed in supplementary table 2 (Supplementary Material online). The temperature cycle of LA PCR consisted of initial denaturation at 94 °C for 3 min, 30 cycles of amplification shuttle (98 °C for 20 s and 68 °C for 15 min), and final extension at 72 °C for 5 min. Amplified large DNA fragments were cloned into pBluescript II SK+. After restriction mapping, smaller fragments were subcloned for sequencing. By this method, 3 large contigs (about 50 kbp in total) were sequenced.

Genomic Libraries

Mitochondrial DNA was physically fragmented by sonication and fractionated by preparative electrophoresis in 0.7% agarose. DNA fragments from 1.6 to 3.0 kbp in size were recovered using a QIAQuick Gel Extraction Kit (Qiagen, Hilden, Germany), blunted with T4 DNA polymerase, and cloned into the *Hinc*II site of pUC118. Three thousand white colonies were picked, grown in 100 μl Luria–Bertani (LB) medium in 96-well microplates, and then stored frozen in the presence of 20% glycerol. Sequencing was performed by Hitachi Technologies (Tokyo, Japan).

Assembly and Annotation

Sequences were assembled by the AutoAssembler program (Applied Biosystems, Foster City, CA). Genes for proteins and rRNAs were estimated by similarity searches using various mitochondrial genes as queries. Small open reading frames (ORFs) were then extracted, and their homology to known sequences was analyzed. Base composition and GC/AT skew were calculated using SISEQ (Sato 2000) version 1.30. A circular map was drawn with GenoMap (Sato and Ehira 2003). Genes for tRNA were searched using tRNA-SCAN (http://nsato4.c.u-tokyo.ac.jp/old/GenoMap/GenoMap.html; Lowe and Eddy 1997).

Phylogenetic Analysis

All protein sequences in the 23 mitochondrial genomes of plants and algae in GenBank entries (supplementary table 1, Supplementary Material online) were clustered using Gclust (Sato et al. 2005; for the clusters, see the Mt23 data set at http://gclust.c.u-tokyo.ac.jp/). Both amino acid and coding DNA data sets were prepared using SISEQ as follows. First, an amino acid alignment was prepared for each protein cluster using Clustal X version 1.83 (Thompson et al. 1994). Only sites having less than 20% gaps were selected by "getclu" (a SISEQ command). A DNA alignment exactly corresponding to each amino acid alignment was prepared by "nucaln." The first and the second codon positions were then extracted to get the final alignment ("splcod" and "seqcat"). The DNA alignment (18 genes, 9,918 sites, 15 operational taxonomic units [OTUs]) was used to infer the phylogeny by Bayesian inference (BI) using MrBayes version 3.1 (Ronquist and Huelsenbeck 2003). Before the BI calculation, suitable substitution models were evaluated by Modeltest version 3.7 (Posada and Crandall 1998) with PAUP* version 4 beta 10 (Swofford 1998). The results suggested that the best model was GTR + I + G (general time reversible, invariable sites, gamma distribution). The BI calculation was performed using the following parameters: nst = 6 (general time reversible), rates = adgamma (autocorrelated rates across sites with a gamma distribution; this option is not included in the test sets for Modeltest but is selected after comparison of various different models; see table 1), nucmodel = 4×4, ngen = 1,000,000, samplefreq = 200, burnin = 2,000. Several "rate" parameters were used (e.g., adgamma, invgamma, propinv, gamma, and equal) for the alignment using the first and second codons (table 1). All parameters were independently set for each gene. Another data set including 4 genes and 22 OTUs (excluding *Pseudendoclonium*) was processed in a similar way. As controls, calculations were also done for the nucleic acid alignment having all codons as well as for the amino acid alignment. A Neighbor-Joining (NJ) tree, with the minimal evolution (ME) optimization, was also calculated using identical data in MEGA version 3.1 (Kumar et al. 2004) on a Windows PC. For DNA sequences, a Jukes–Cantor (JC) model with gamma = 2.0 was used. For protein sequences, a Jones–Taylor–Thornton (JTT) model was used with gamma = 2.0. Maximum parsimony (MP) calculation was performed with PAUP* (see above) on an Apple PowerMac G5. A heuristic search was done with

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

Table 1
Evaluation of Various Parameters in the Phylogenetic Analysis

Model	Parameter	Generations	Run	Average standard deviation of split frequencies	Mean marginal likelihood	Topology	Support for (PPT, MPO)
DNA sequences with the first and second codons							
Nst = 1	Adgamma	1,000,000	1	0.009179	−92,137.41	1B	0.067
Nst = 2	Adgamma	1,000,000	1	0.008550	−91,699.29	1B	0.372
Nst = 6	Invgamma	1,000,000	1	0.005359	−91,848.46	2B	0.85
			2	0.010920	−91,853.28	2B	0.89
			3	0.022339	−91,838.10	2B	0.86
Nst = 6	Adgamma	1,000,000	1	0.001175	−91,315.64	2B	0.69
			2	0.005123	**−91,308.68**	2B	0.75
			3	0.001001	−91,309.86	2B	0.73
Nst = 6	Equal	1,000,000	1	0.051558	−95,950.44	31A	1.00
Nst = 6	Propinv	1,000,000	1	0.029722	−92,924.17	2A	0.99
Nst = 6	Gamma	1,000,000	1	0.007603	−91,894.24	2B	0.91
NJ analysis (JC model and gamma = 2.0)						33A	1.00
MP analysis						32B	0.91
DNA sequences with all 3 codons							
Nst = 6	Invgamma	1,000,000	1	0.009571	−167303.49	1A	0.00
			2	0.001854	−167300.25	1A	0.00
			3	0.009600	**−167293.92**	1A	0.00
Nst = 6	Adgamma	1,000,000	1	0.005887	−167362.18	1A	0.00
NJ analysis (JC model and gamma = 2.0)						33A	0.99
MP analysis						4	0.89
Amino acid sequences							
Whelan and Goldman	Invgamma	1,000,000	1	0.000031	−90491.18	2A	1.00
			2	0.000298	**−90478.65**	2A	1.00
			3	0.000079	−90480.38	2A	1.00
NJ analysis (JTT model and gamma = 2.0)						32A	1.00
MP analysis						34A	0.94

NOTE.—The data in figure 3*A* were analyzed with various parameters. Results obtained by the BI method are shown with additional results from NJ and MP. The results obtained using DNA sequences with the first 2 codons or the 3 codons, as well as the results with protein sequences are presented. Other combinations of parameters such as nst = 2 with invgamma for the 3 codons, which gave an inferior value of lnL, were calculated but not included in this table. The best record for each analysis is shown in bold face. The calculation was done using MrBayes version 3.1.2 compiled with the Message-Passing Interface (MPI) option, which was run using 8 CPUs at a time on supercomputers (Sun Fire 15, 1.0, or 1.2 GHz, 96 CPUs, 288 GB memory) in the Human Genome Center, the University of Tokyo. The support values for NJ and MP analyses are the bootstrap confidence level obtained from 1,000 replicates. A graphical representation of tree topology types is shown in supplementary figure 1 (Supplementary Material online). Topologies are as follows:

1A = ((((((((Angio,MPO),PPT),CHR),CHA),MST),(NEP,PWI)),((Red,RSA),PLI)),REC);
1B = ((((((((Angio,MPO),PPT),CHR),CHA),MST),(NEP,PWI)),(Red,RSA)),(PLI,REC));
2A = (((((((Angio,(PPT,MPO)),CHR),CHA),MST),(NEP,PWI)),((Red,RSA),PLI)),REC);
2B = (((((((Angio,(PPT,MPO)),CHR),CHA),MST),(NEP,PWI)),(Red,RSA)),(PLI,REC));
31A =(((((((Angio,(PPT,MPO)),CHR),CHA),NEP),(MST,PWI)),((Red,RSA),PLI)),REC);
32A = ((((((((Angio,(PPT,MPO)),CHR),CHA),NEP),PWI),MST),((Red,RSA),PLI)),REC);
32B = ((((((((Angio,(PPT,MPO)),CHR),CHA),NEP),PWI),MST),(Red,RSA)),(PLI,REC));
33A = ((((((((Angio,(PPT,MPO)),CHR),CHA),NEP),MST),PWI),((Red,RSA),PLI)),REC);
34A = ((((((((Angio,(PPT,MPO)),CHR),CHA),(NEP,PWI)),MST),((Red,RSA),PLI)),REC));
4 = (((((((Angio,(PPT,MPO)),CHR),CHA),(NEP,PWI),MST),(Red,(RSA,PLI))),REC));
Angio = (maize,*Arabidopsis*);
Red = (*Cyanidioschyzon*,(*Porphyra*,*Chondrus*));
PPT = *Physcomitrella;*
MPO = *Marchantia;*
CHR = *Chara;*
CHA = *Chaetosphaeridium;*
MST = *Mesostigma;*
NEP = *Nephroselmis;*
PWI = *Prototheca;*
RSA = *Rhodomonas;*
PLI = *Pylaiella;*
REC = *Reclinomonas;*

random stepwise addition. In both cases, 1,000 replicates were calculated for each bootstrap analysis.

Results and Discussion

Overall Features of the Mitochondrial Genome

The assembled mitochondrial genome sequence of *P. patens* has 105,340 bp (fig. 1). Forty-two protein-coding genes were identified, which are homologs of conserved mitochondrial protein genes (supplementary table 3, Supplementary Material online). Many additional putative ORFs were also detected by the computational survey, but they are not included in the current annotation because the significance of these reading frames is not clear. There was a single copy of each of the ribosomal RNA genes. Twenty-four tRNA genes were also found. Interestingly, many genes

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

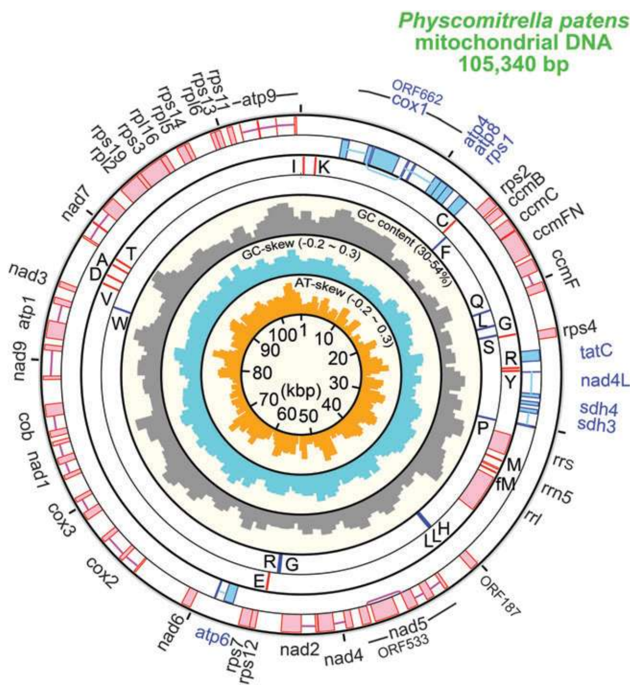

FIG. 1.—Circular map of the mitochondrial genome of *Physcomitrella patens*. The outermost 2 circles show the positions of protein-coding genes (exons are linked with lines). The next 2 circles show the positions of RNA genes. In both cases, genes encoded by the clockwise strand are colored red, whereas genes encoded by the counterclockwise strand are colored blue. The names of the protein-coding genes are indicated outside the circle. The names of tRNA genes are indicated by the amino acids. The central 3 circles show GC content (gray), GC skew (sky blue), and AT skew (orange), from the outer to the inner circle, respectively. These values were averaged over 1,000-bp regions without overlap and are shown in the ranges indicated. The calculation was done with SISEQ (Sato 2000), drawn by GenoMap (Sato and Ehira 2003), and embellished in Adobe Photoshop version CS2.

were encoded by the clockwise strand. The overall GC content was 40.6%, which was higher than that of the *P. patens* chloroplast genome (28.5%). There was no distinct bias in the GC content or in the GC or the AT skew (fig. 1).

Comparison of Global Genome Structure

Supplementary table 4 (Supplementary Material online) shows the gene content of the mitochondrial genomes of *Physcomitrella* and various green plants. The *ccm* genes encoding heme-biosynthesis enzymes are missing in *Mesostigma* and *Chaetosphaeridium*. The gene content of bryophytes and *Chara* is not very different, whereas some angiosperms have a notably smaller number of genes, as noted previously (Turmel et al. 2003). There is no evident relationship between the presence/absence of genes and the phylogeny (see below). For example, the *rpl14* gene was found in *Mesostigma* and *Chara*, but not in other taxa. The *rpl2* gene was missing in *Mesostigma* and maize. The *rps8* gene was found in only *Marchantia*. The *rps13* gene was missing in *Chara* and *Arabidopsis*. This comparison is consistent with the previous observation that gene transfer to the nucleus occurred many times independently (Adams et al. 2002).

A detailed comparison of synteny in the mitochondrial genomes of *M. polymorpha* (Oda et al. 1992), *P. patens*, *Chara vulgaris* (Turmel et al. 2003), and *Chaetosphaeridium globosum* (Turmel et al. 2002a) is shown in figure 2. These taxa are hereafter referred to by their genus names. Among these 4 species, synteny was found in many gene clusters, such as the ribosomal protein (RP) cluster, the *ccm* gene cluster, and 2 regions containing the *nad* and *cox* genes. The large difference in size of these 4 genomes (187, 105, 68, and 57 kb, respectively) is therefore due to the intergenic regions and introns. Gene arrangement was particularly similar between *Physcomitrella* and *Chara*. The large segment *cob-nad1-cox3-cox2-nad6-atp6-rps7-rps12-nad2-nad4-nad5* (the underlined section is transcribed in the reverse orientation) is conserved in this pair of genomes, although inverted in this representation (blue lines in fig. 2). In addition, *cob*, *nad1*, and *rps7-rps12* are inverted locally within this cluster. The gene clusters *nad9-atp1-nad3-nad7*-[RP genes]-*atp9* and *cox1* to c*cmF* are also conserved. The exact content of the RP cluster is compared below. The region from *rps4* to *rrl* in *Physcomitrella* is also conserved in *Chara*, with some rearrangements. This comparison clearly indicates that the *Chara* mitochondrial genome structure is highly related to that of *Physcomitrella*. It is also easy to trace changes between *Chaetosphaeridium* and *Chara* or *Physcomitrella*. In *Chaetosphaeridium*, no large inversion is found with respect to *Physcomitrella*, and the *ccm* genes are missing in *Chaetosphaeridium*. Comparison of *Marchantia* and *Physcomitrella* is also interesting. The 2 genomes are essentially the same, as indicated by 4 segments: *cox2* to *atp1*, *nad5* to *nad6*, *rps2* to *ccmF*, and *atp4* to *cox1*. Some other clusters, such as the RP gene cluster, are translocated with respect to these 4 segments. Despite these translocations, a good correspondence was found in the overall gene arrangement in these 2 genomes. In addition, uncharacterized putative protein genes (black) with no homology are inserted in the intergenic regions of *Marchantia*. In the light of good conservation of synteny in *Physcomitrella* and *Marchantia*, these unidentified *Marchantia* ORFs might be inserted in the liverwort lineage.

A significant degree of synteny is only found within these 4 species, not including angiosperm mitochondrial genomes that have retained only a small number of genes from the well-conserved RP clusters. In the prasinophyte *Mesostigma*, most of these genes are present (Turmel et al. 2002b), but they are dispersed throughout the mitochondrial genome. Other genes that were usually clustered in our 4-species comparison are also extensively translocated in *Mesostigma*. These findings demonstrate the close relatedness of liverworts, mosses, and charophytes.

Figure 2 and supplementary table 5 (Supplementary Material online) show the arrangement of RP genes. The major RP cluster consists of so-called bacterial “S10, *spc*, and alpha operons.” However, all members of these operons in the mitochondrial genomes are arranged into a large genomic segment. Whether these genes are transcribed as a single transcriptional unit remains to be determined. In addition, nonribosomal genes such as *tufA*, *secY*, and *rpoA* are included in these operons in bacteria and chloroplasts, which is not the case in mitochondria except in jakobid flagellates (Lang et al. 1997). Gene arrangement in *Rickettsia prowazekii*, one of the α-Proteobacteria

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

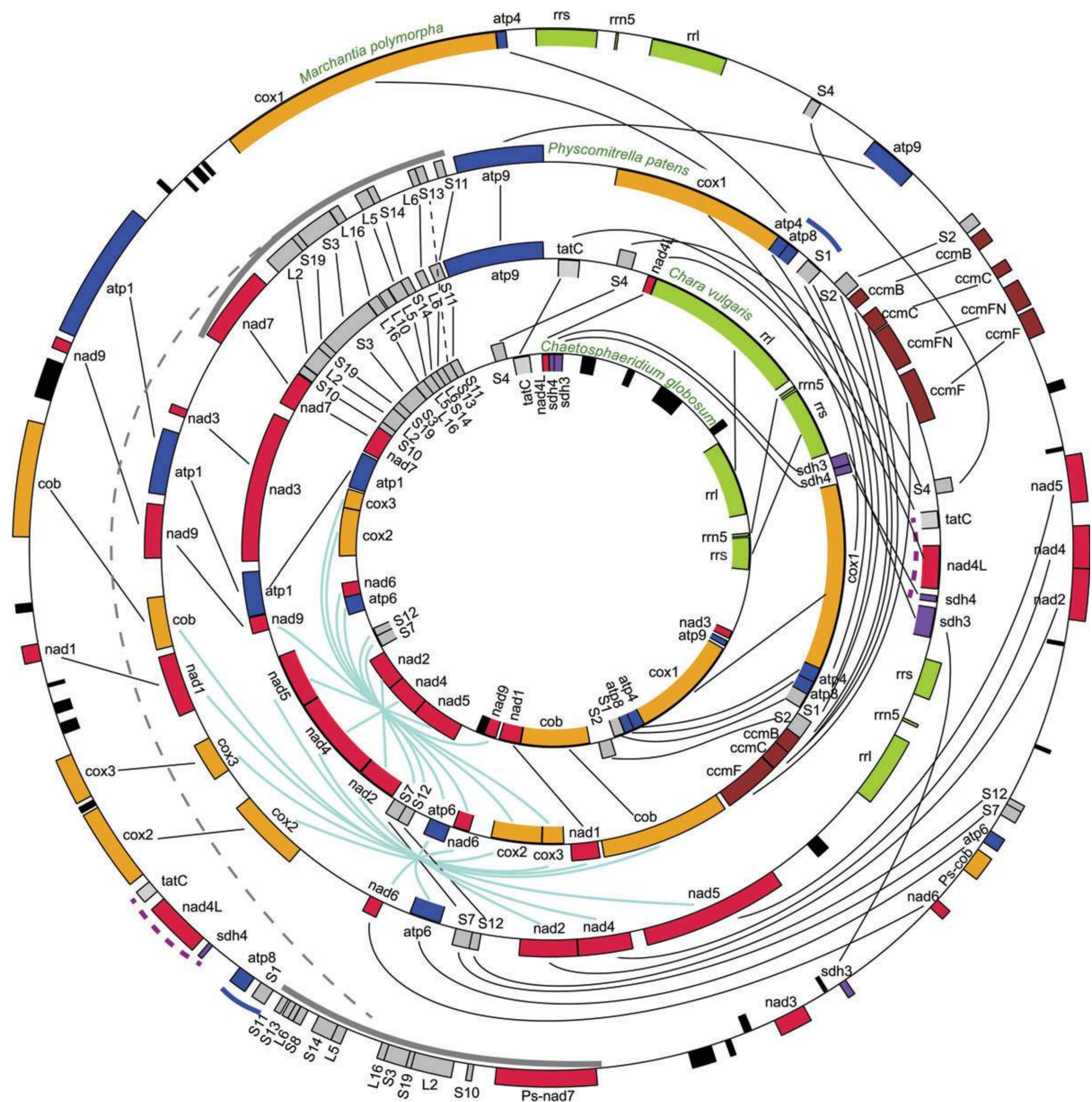

FIG. 2.—Comparison of synteny in *Marchantia*, *Physcomitrella, Chara*, and *Chaetosphaeridium* mitochondrial genomes. Protein-coding genes were illustrated using GenoMap. The 4 mitochondrial genomes are shown from outside to inside, respectively. Genes are color coded as follows: RP genes and *tatC* gene, gray; ribosomal RNAs, green; *nad* genes, red; *sdh* genes, purple; *cox* and *cob* genes, orange; *atp* genes, blue; *ccm* genes, brown; unidentified protein genes, black. The tRNA genes are not shown. The direction is reversed in *Chara* and *Chaetosphaeridium* with respect to the database entries to show better synteny. Major syntenic regions are shown by curved lines, and major genes included in the regions are indicated. The *rrn* operon, the *sdh4* cluster and the *nad7*-ribosomal cluster were translocated during the diversification of the 2 bryophytes. The blue lines crossing at a point between *Physcomitrella* and *Chara* or between *Chara* and *Chaetosphaeridium* indicate a large inversion. Except this large inversion and translocation around the *rrn* operon, synteny in *Physcomitrella* and *Chara* is nearly perfect.

thought to be most related to the mitochondrial ancestor, is also shown at the bottom of supplementary table 5 (Supplementary Material online). Essentially identical gene arrangement is conserved in *Physcomitrella*, *Marchantia*, *Chara*, *Chaetosphaeridium*, and *Nephroselmis*, as well as the unicellular red alga *Cyanidioschyzon merolae* and the primitive protist *Reclinomonas americana*. All sequenced mitochondrial genomes of flowering plants lack most of the RP genes, and only a trace of the S10 cluster remains. However, this may not be the general rule in angiosperms because Adams et al. (2002) showed by Southern blotting that some angiosperms retain as many as 40 protein genes, including a more complete set of RP genes. The 2 green algae, *Mesostigma* and *Prototheca*, have highly translocated RP genes. Another RP cluster, the S12 or *str* operon in bacteria, is reduced in mitochondrial genomes and present as a S12-S7 cluster. However, this cluster is conserved in many organisms, including *Physcomitrella*. Other orphan RP genes encoding S4, S2, and S1 are also conserved in bryophytes and charophytes.

These results suggest that a large RP cluster was conserved at the base of land plants. Although we still need data on the mitochondrial genomes of gymnosperms and pteridophytes, a reasonable hypothesis is that the RP cluster was

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

conserved even at the base of the angiosperms, although many RP genes were independently lost thereafter in many lineages of angiosperms. In green algae (except charophytes), the selection pressure of genome reduction led to smaller genomes with many rearrangements. Because of these circumstances, the mitochondrial genomes of charophytes and bryophytes retain the most primitive form of mitochondrial genome in the green lineage.

Codon Usage and tRNA Genes

All 61 codons are used in the *Physcomitrella* mitochondria (supplementary table 6, Supplementary Material online). However, the tRNA genes identified in the mitochondrial genome are insufficient to decode the codons; at least the tRNA genes corresponding to the ACA/G(Thr), CGA/G(Arg), AUU/C(Ile), AAU/C(Asn), and AGU/C(Ser) codons were missing (supplementary table 6, Supplementary Material online). In the mitochondrial genomes of various plant and algal species, several tRNA genes are missing (supplementary table 4, Supplementary Material online), and some of them are thought to be encoded by the nuclear genome and imported into mitochondria. The orthologs of tRNA-Thr(GGU), which is encoded by the mitochondrial genome of *Physcomitrella*, are known to recognize the ACU and ACC codons, but not ACA and ACG. This indicates mitochondrial import of at least one additional threonine isoacceptor tRNA, such as tRNA-Thr(AGU), which is imported into mitochondria in *Marchantia* (Akashi et al. 1997). The first nucleotide residue A of the anticodon of this tRNA may be modified to inosine (I), thereby enabling all the CAN codons to be read. The tRNA-Arg(ACG) is the only tRNA that corresponds to the CGN codons in the *Physcomitrella* mitochondria and is likely to decode the CGU and CGC codons. However, this tRNA might read all CGN codons, as reported in yeast mitochondria (Sibler et al. 1986). Because the mitochondrial genomes of *Mesostigma* and *Marchantia* encode tRNA-Arg(UCG) (supplementary table 4, Supplementary Material online), an ortholog encoded by the *Physcomitrella* nuclear genome may be imported into mitochondria to read the CGA and CGG codons.

Bacterial tRNA-Ile(LAU), which is derived from tRNA-Ile(CAU) by C to L (lysidine) modification, recognizes the AUA codon but not the AUG codon (Muramatsu et al. 1988). The "L" modification seems to exist in potato mitochondria (Weber et al. 1990), so the AUA codon may be read by modified tRNA-Ile(LAU) in *Physcomitrella*. The tRNA-Ile(GAU), which decodes the AUU and AUC codons, as well as tRNA-Asn(GUU) and tRNA-Ser(GCU), is encoded by the mitochondrial genomes of *Mesostigma*, *Chaetosphaeridium*, and *Chara* (supplementary table 4, Supplementary Material online) but not by *Physcomitrella*. Thus, these tRNAs could be imported from the nucleus in *Physcomitrella*.

Phylogenetic Analysis of Mitochondrial Protein Genes

The phylogeny of the mitochondrial genomes was estimated by BI. The protein-coding genes that were conserved in 16 genomes, namely, *cob*, *cox1*, *cox2*, *cox3*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *atp6*, *atp8*, *atp9*, *rpl16*, *rps12*, *rps3*, and *tatC* genes, were selected. The first 2 codon positions were used for the BI calculation. However, the Markov Chain Monte Carlo (MCMC) runs during the BI analysis did not converge around *Pseudendoclonium*, which was finally excluded from the analysis, giving a tree of 15 species in total.

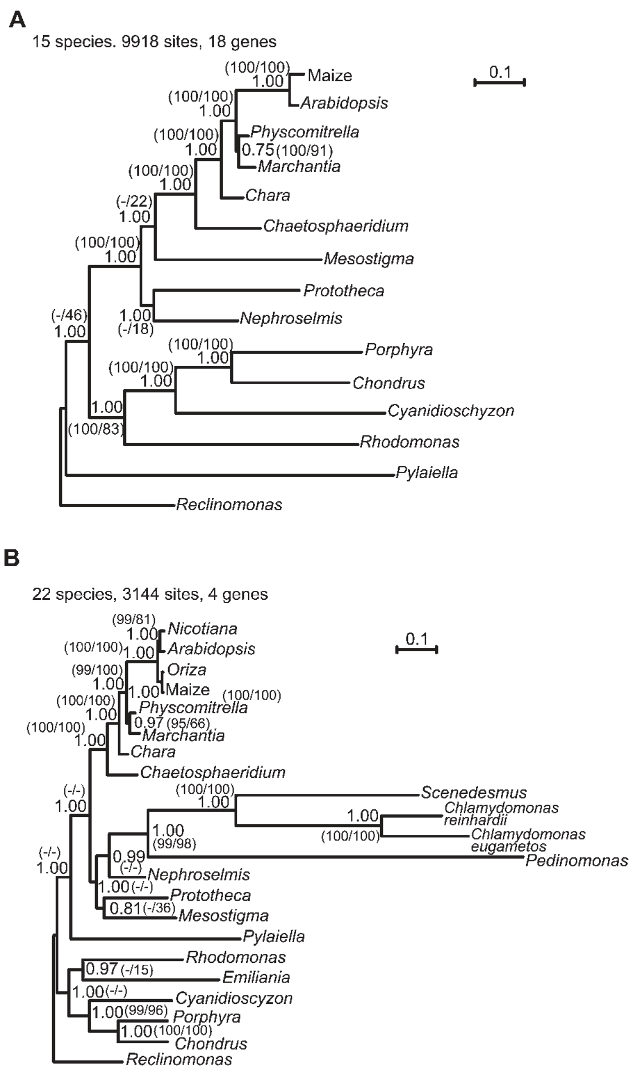

FIG. 3.—BI phylogenetic tree of photosynthetic organisms based on protein genes encoded by the mitochondrial genomes. (*A*) Consensus tree with 15 species, including 18 genes and 9,918 sites. (*B*). Consensus tree with 22 species, including 4 genes and 3,144 sites. Methodological details for BI calculation are given in the text. The posterior probability of each branch is shown with the bootstrap confidence levels (percentage based on 1,000 replicates) obtained by the NJ/MP methods (in parentheses). The NJ calculation was performed with MEGA version 3.1 (Kumar et al. 2004) for Windows with the ME option, JC model, and a gamma distribution (gamma = 2.0). The MP calculation was performed with PAUP* version 4 beta 10 for MacOS 9.2 (Altivec) on an Apple PowerMac G5 with the heuristic search option. The details of genomes used are as listed in supplementary table 1 (Supplementary Material online).

The resulting phylogenetic tree is shown in figure 3*A*. There were only 2 trees that remained after the long MCMC run (1,000,000 generations). Reproducibility of MCMC optimization was confirmed by 3 different runs (table 1). As a control, another data set of 4 genes (*cob*, *cox1*, *nad4*, and *nad5*) having 3,144 sites in 22 species was also

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

examined (fig. 3*B*). The global pattern was conserved but somewhat deformed due to the long distance of some green algae (*Scenedesmus*, 2 species of *Chlamydomonas*, and *Pedinomonas*). We finally selected the first data set and then analyzed it in more detail by changing the parameters. The parameter set nst = 6 (general time reversible) and rates = adgamma (autocorrelated rates across sites, with a gamma distribution) was the best (table 1, upper half). The results showed that *Physcomitrella* and *Marchantia* were monophyletic, and this group was sister to the angiosperms. To further evaluate the results, BI calculations using the 3 codon positions and protein sequences were also done (table 1, lower half). An identical data set was used to calculate NJ and MP trees, with 1,000 bootstrap replicates (fig. 3*A*). The monophyly of *Physcomitrella* and *Marchantia* is supported by most analyses, except BI with 3 codon positions. This was also supported by analyses of the chloroplast genome (Nishiyama et al. 2004; Sato 2006). The reliability of the topologies was tested by the approximately unbiased (AU) test (Shimodaira and Hasegawa 2001). Among the 10 topologies supported by one of the analyses (table 1), topologies 1A and 1B, in which *Marchantia* and *Physcomitrella* are paraphyletic, were not rejected at $P = 0.05$ (supplementary table 8, Supplementary Materials online). This is also consistent with the results of BI analyses showing that support for bryophyte monophyly is less than 0.9. These results suggest that we need additional data, most notably additional taxa (pteridophytes, gymnosperms, and hornworts), to obtain a more solid support for the monophyly of bryophytes.

Chara is the closest relative to the land plants (fig. 3*A*). In a previous report on the *Chara* chloroplast genome (Turmel et al. 2006), *Chara* was shown to be at the base of all land plants and charophytes. Therefore, we found a different branching order of *Chara* and *Chaetosphaeridium*. However, the synteny analysis of mitochondrial genomes (fig. 2) suggested a closer similarity of *Chara* than *Chaetosphaeridium* to the 2 bryophytes. The current findings are also different from a previous intron analysis suggesting that liverworts are the earliest land plants (Qiu et al. 1998). The AU test, however, clearly showed that the topologies with *Chara* and *Chaetosphaeridium* reversed are not supported at $P = 0.01$ (supplementary table 8, Supplementary Material online).

Various tree topologies including bryophytes have been published recently. Renzaglia et al. (2000) summarized morphological data and suggested the monophyly of bryophytes, placing the hornworts at the base of bryophytes. Nishiyama et al. (2004) tested various methods by using chloroplast protein sequences and concluded that bryophytes are monophyletic. In their analysis, however, the same data set with DNA sequences gave noticeably different results. Many phylogenetic analyses using various protein sequences, such as guanosine adenine dinucleotide phosphate (Petersen et al. 2003), *nad1* (Dombrovska and Qiu 2004), *nad5* (Groth-Malonek et al. 2005), and chloroplast sequences including a lycopod sequence (Wolf et al. 2005), indicate serial sister group relationships, with liverworts branching first. In most cases, hornworts are closest to vascular plants. In genome-based analyses such as presented here, the number of sequences is maximal, whereas the number of taxa is limited. Notably, we have no genome data for hornworts and pteridophytes. We have only one mitochondrial genome for the liverworts and mosses. These are limitations to the reliable inference of phylogenetic relationships among land plants.

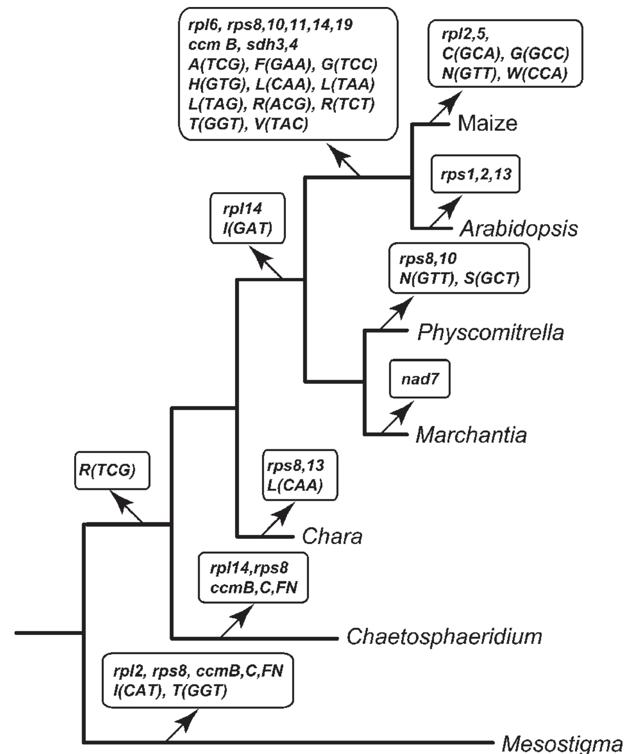

FIG. 4.—Loss of genes during the evolution of mitochondrial genomes. Gene losses are mapped on the phylogenetic tree.

We estimated the formal number of gene-loss events based on the obtained phylogenetic tree (fig. 4). *Rps8* is only conserved in *Marchantia*; therefore, we must assume loss of *rps8* in many lineages. A number of genes seem to have been lost during the evolution of angiosperms, but as noted above, some angiosperms do retain many genes. The loss of many genes may have occurred independently in various lineages during the speciation of angiosperms.

Conservation of Intron Positions

By homology analysis, we found 2 Group 1 introns (i8 of *cox1* and i3 of *nad5*) and 25 Group 2 introns in 16 genes of the *Physcomitrella* mitochondrial genome. The positions of introns were used as a marker of phylogeny (Qiu et al. 1998; Hashimoto and Sato 2001; Pruchner et al. 2002), and this seems true in some cases, such as the *nad7* gene (fig. 5). Angiosperms (except some plants such as tobacco) have 4 introns (i1, i2, i4, and i5) in their *nad7* gene (Hashimoto and Sato 2001). The same is true for *Ginkgo biloba* (gymnosperm). Pteridophytes contain the first 3 introns (i1, i2, and i4). The first 2 introns (i1 and i2) of the *nad7* gene are conserved in *Physcomitrella* and other mosses (Nakamura et al. 2002). However, the history of intron insertion/removal is complicated because introns are mobile. The *nad7* gene in *Marchantia* was translocated to the nuclear genome, and the mitochondrial pseudogene, which retains the

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

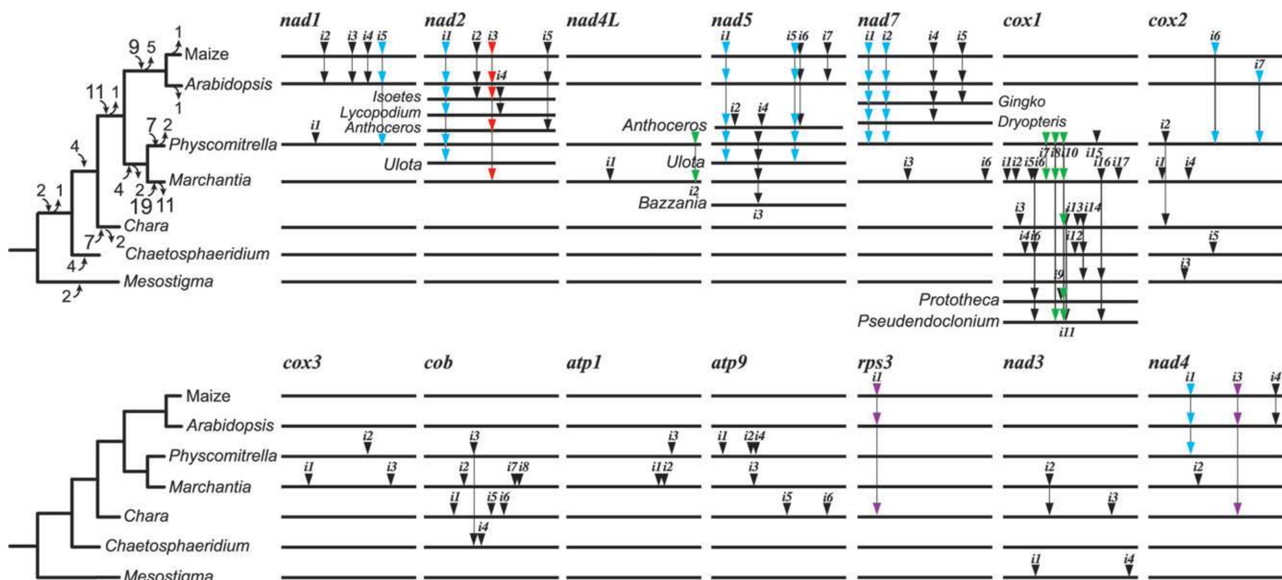

FIG. 5.—Comparison of mitochondrial introns in various plants. Each arrowhead represents the position of an intron insertion. Introns are numbered according to their position. Data sources: *Anthoceros agrestis* (hornwort: AJ409117, AJ627605, AJ627606, AJ627607); *Bazzania trilobata* (liverwort: AJ622815); *Dryopteris crassirhizoma* (fern: accession numbers AB259580–AB259585); *Equisetum arvense* (horsetail: AB259586–AB259589); *Ginkgo biloba* (gymnosperm: AB259572–AB259579); *Isoetes dureui* (lycopod: AJ409119); *Lycopodium annotinum* (lycopod: AJ409120); *Ulota crispa* (moss: AJ299553, AJ622819). Other data sources are described in supplementary table 1 (Supplementary Material online). The phylogenetic tree on the left indicates the number of additions and removals of introns. The sequence alignments underlying these data are available as supplementary figure 2 (Supplementary Material online) (a large file).

complete coding region but is inactivated by 6 termination codons, contains a totally different set of introns, i3 and i6 (introns X and Y in Hashimoto and Sato 2001). Intron i3 was found to be conserved in some liverworts (Nakamura et al. 2002) related to *Marchantia*, suggesting that this intron was inserted during the evolution of the Marchantiales.

Results of a survey of the presence/absence of introns are shown in figure 5 and supplementary table 4 (Supplementary Material online). A complete set of alignments of gene sequences containing introns introns is also available as supplementary figure 2 (Supplementary Material online). Here we clearly distinguish the absence of a gene and absence of an intron, which is sometimes confused in a previous table of introns (Turmel et al. 2003). Figure 5 summarizes the position of introns within the genes that contain introns. In the *nad2* gene, the first intron i1 is conserved in *Physcomitrella* and angiosperms (fig. 5 and Pruchner et al. 2002). Intron i3 in the *nad2* gene is the only one that is shared by *Marchantia* and angiosperms. Conserved introns in *Physcomitrella* and angiosperms are also found in *nad1*, *nad4*, *nad5*, and *cox2* (blue arrowheads in fig. 5 and Turmel et al. 2003; Groth-Malonek et al. 2005). The only intron i1 in the *rps3* gene is conserved in angiosperms and *Chara*. Intron i2 in the *nad4L* gene is conserved in *Physcomitrella* and *Marchantia*. The introns in the *cob*, *atp1*, and *atp9* genes are variable in bryophytes and charophytes, and these introns are not found in angiosperms. The situation in the *cox1* gene is complicated, but some introns such as i7, i8, and i10 are conserved in *Physcomitrella* and *Marchantia*. In summary, *Physcomitrella* contains 27 introns in 16 mitochondrial genes, and 9 of them are shared with angiosperms. Four other introns are shared with *Marchantia*, whereas only a single intron is shared between *Marchantia* and angiosperms. Based on the phylogeny in figure 3*A*, these results are explained by assuming that intron insertion occurred continuously during the evolution of spermatophytes and that many introns, once introduced in the common ancestor of land plants, were lost and new introns were introduced during the evolution of liverworts related to *Marchantia* (phylogenetic tree with intron flow in fig. 5). Introns might sometimes serve as markers of phylogeny, but all intron data can be interpreted by adequately assuming insertion and removal according to the phylogenetic relationship that was established by a more sophisticated computational method.

Single Molecular Species of the Mitochondrial Genome

Because the mitochondrial genomes of angiosperms are multipartite and consist of various different molecular species that are produced by complex DNA recombination (Bullerwell and Gray 2004), we analyzed molecular species of *Physcomitrella* mitochondrial genomes by field-inversion electrophoresis (fig. 6). The results showed that the major molecular species was about 100 kbp (representing a probably linearized species produced during the manipulation). A putative circular molecule was detected at a position equivalent to about 300 kbp linear DNA (asterisk). The latter could also represent linear multimers as found in the *Marchantia* mitochondrial genome (Oldenburg and Bendich 1998, 2001). This result suggests that the mitochondrial genome of *Physcomitrella* does not consist of a multipartite structure as in angiosperms (Klein et al. 1994; Fauron et al. 1995; Ogihara et al. 2005; Sugiyama et al. 2005). Also, the mitochondrial genome of *Marchantia* is known to be a mixture of genome-sized and multigenomic molecules in linear and branched form (Oldenburg and

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

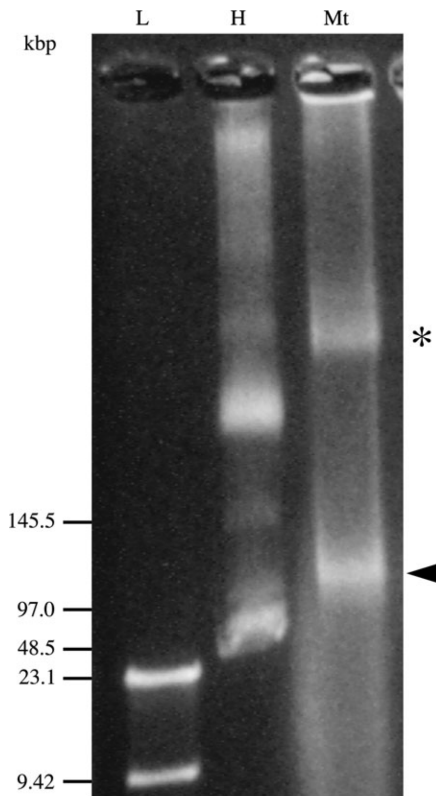

FIG. 6.—Field-inversion electrophoresis of mitochondrial genome. Mitochondrial DNA was electrophoresed in a 1.0% agarose gel in 0.5× TBE (44.5 mM Tris-borate, 1 mM EDTA) at 6 °C for 20 h with a program in which switch times ramped from 3 to 17 s and the switch time ratio was kept at 2:1. The electric field was 6.0 V/cm. This figure shows the fluorescence of DNA after staining with ethidium bromide. Markers: L, *Hin*dIII-cut lambda DNA (23, 9.2 kbp as well as smaller fragments); H, lambda DNA ladder (multiples of 48.5 kbp). Mt, *Physcomitrella patens* mitochondria; asterisk, circular or concatenated linear molecule; arrowhead, linear molecule.

Bendich 1998, 2001). Multipartite structures caused by multiple recombination as seen in angiosperms may have started during the evolution of pteridophytes or seed plants. This is also supported by the fact that gene order is largely conserved in the 2 bryophytes and the 2 charophytes (fig. 2), whereas gene synteny has been lost in the mitochondria of flowering plants. Additionally, repeat sequences that might be involved in genome recombination were searched for in the *Physcomitrella* mitochondrial genome (supplementary fig. 3, Supplementary Material online). No significant repeats that could lead to genome rearrangements as found in angiosperm mitochondrial genomes were detected.

However, this observation is a clue to the evolution of multipartite or linear multimer structures. During the evolution of mitochondrial genomes, intron insertions accumulated. The introns are likely to have originated from transposons, and some introns likely originated from identical or related transposons. In this case, homologous recombination may have occurred between the homologous introns. Normally, this is not allowed because such recombination will lead to truncated, nonfunctional essential genes. However, in *Marchantia*, for example, the *nad7* gene has been transferred to the nucleus and has become a mitochondrial pseudogene. The *nad7* pseudogene contains 2 introns (fig. 5). Recombination involving these introns might lead to rearranged molecules without loss of essential genes. In higher plants, insertion of introns and gene transfer to the nucleus occur very frequently, and this situation allows some genome recombination leading to large rearrangements or multipartite structures. This is just a possible scenario, but the frequent occurrence of repeat sequences within the introns may be a clue to the evolution of mitochondrial genomes.

Other Features

The mitochondrial genomes of angiosperms have several features besides multipartite organization and large size, such as RNA editing and trans-splicing. First, we searched for the presence of chloroplast sequences. A BlastP analysis using the whole sequence of the *Physcomitrella* mitochondrial genome as a query against a collection of 44 chloroplast genome sequences detected significant homology of *rnl*, *rnl*, and the first exon of *atp1*. All these are attributed to homology between mitochondrial and chloroplast orthologs and are not due to promiscuous sequences as found in angiosperms. Second, we found at least 7 sites of putative C to T RNA editing (supplementary table 7, Supplementary Materials online). Third, we did not detect any evidence of *trans*-spliced genes. These results suggest that some of the features of angiosperm mitochondria were already acquired during the evolution of land plants.

Concluding Remarks

In summary, the mitochondrial genome of *Physcomitrella*, which is the smallest among the sequenced mitochondrial genomes in land plants, sheds new light on the evolution of mitochondrial genomes, namely, the close relatedness of bryophytes and charophytes as well as the prototype structure of the *Physcomitrella* mitochondrial genome among its land plant counterparts. Sequence-based phylogenetic analysis clearly supports the conclusion that bryophytes are a sister group to flowering plants. The complex rearrangements and recombinations found in the mitochondrial genomes of flowering plants may have started during the evolution of vascular plants or flowering plants. In this regard, more sequence information is needed on pteridophyte mitochondrial genomes. Finally, the insertion/removal of introns was found to be a less informative marker of evolution of mitochondrial genomes of land plants because liverworts have many introns at positions quite different from those in mosses and vascular plants. In addition to evolutionary insights, the mitochondrial genome sequence of *Physcomitrella* will be useful in analyzing mitochondrial gene expression because the presence of 2 phage-like mitochondrial RNA polymerases have been unequivocally identified in this moss (Kabeya et al. 2002; Kabeya and Sato 2005). *Physcomitrella* is an ideal organism to study plant mitochondria because the mitochondria are large and easily identified under a microscope in protonemal or leaf cells (see micrographs in Kabeya and Sato 2005), a situation that is very different in the green leaves of higher plants. This will allow detailed analysis of gene expression, most notably of promoter and transcription factors.

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

Supplementary Material

Supplementary tables 1–8 and figures 1–3 and sequence data for *P. patens* mitochondrial genome deposited in the GenBank/EMBL/DNA Data Bank of Japan database under accession number AB251495 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Dr S. Tabata, Kazusa DNA Research Institute, for help in constructing the genomic library. One of the authors, T.K., left the laboratory in the spring of 2005 following graduation. This work was supported in part by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (nos 15370017, 17018010, 18017005, and 16GS0304).

Literature Cited

Adams KL, Qiu Y-L, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 99: 9905–9912.

Akashi K, Hirayama J, Takenaka M, Yamaoka S, Suyama Y, Fukuzawa H, Ohyama K. 1997. Accumulation of nuclear-encoded tRNA(Thr) (AGU) in mitochondria of the liverwort *Marchantia polymorpha*. Biochim Biophys Acta. 1350: 262–266.

Backert S, Nielsen BL, Börner T. 1997. The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci. 2:477–483.

Beckert S, Muhle H, Pruchner D, Knoop V. 2001. The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol Phylogenet Evol. 18:117–126.

Beckert S, Steinhauser S, Muhle H, Brennicke A, Knoop V. 1999. Mitochondrial nucleotide sequences yield a high resolution molecular phylogeny of bryophytes. Plant Syst Evol. 218:179–192.

Bullerwell CE, Gray MW. 2004. Evolution of the mitochondrial genome: protist connections to animals, fungi and plants. Curr Opin Microbiol. 7:528–534.

Burger G, Saint-Louis D, Gray MW, Lang BF. 1999. Complete sequence of the mitochondrial DNA of the red alga *Porphyra purpurea*: Cyanobacterial introns and shared ancestry of red and green algae. Plant Cell. 11:1675–1694.

Dombrovska O, Qiu Y-L. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol. 32: 246–263.

Draper J, Scott R. 1988. The isolation of plant nucleic acids. In: Draper J, Scott R, Armitage P, Walden R, editors. Plant genetic transformation and gene expression: a laboratory manual. Oxford: Blackwell Scientific Publications. p. 199–236.

Fauron C, Casper M, Gao Y, Moore B. 1995. The maize mitochondrial genome: dynamic, yet functional. Trends Genet. 11:228–235.

Fitzpatrick DA, Creevey CJ, McInerney JO. 2006. Genome phylogenies indicate a meaningful alpha-proteobacterial phylogeny and support a grouping of the mitochondria with the Rickettsiales. Mol Biol Evol. 23:74–85.

Gray MW, Burger G, Lang BF. 1999. Mitochondrial evolution. Science. 283:1476–1481.

Groth-Malonek M, Pruchner D, Grewe F, Knoop V. 2005. Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol. 22:117–125.

Hashimoto K, Sato N. 2001. Characterization of the mitochondrial *nad7* gene in *Physcomitrella patens*: similarity with angiosperm *nad7* genes. Plant Sci. 160:807–815.

Kabeya Y, Hashimoto K, Sato N. 2002. Identification and characterization of two phage-type RNA polymerase cDNAs in the moss *Physcomitrella patens*: implication of recent evolution of nuclear-encoded RNA polymerase of plastids in plants. Plant Cell Physiol. 43:245–255.

Kabeya Y, Sato N. 2005. Unique translation initiation at the second AUG codon determines mitochondrial localization of the phage-type RNA polymerases in the moss *Physcomitrella patens*. Plant Physiol. 138:369–382.

Klein M, Eckert-Ossenkopp U, Schmiedeberg I, Brandt P, Unseld M, Brennicke A, Schuster W. 1994. Physical mapping of the mitochondrial genome of *Arabidopsis thaliana* by cosmid and YAC clones. Plant J. 6:447–455.

Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 46:123–139.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Lang BF, Burger G, O'Kelly CJ, Cedergren R, Golding GB, Lemieux C, Sankoff D, Turmel M, Gray MW. 1997. An ancestral mitochondrial DNA resembling a eubacterial genome in miniature. Nature. 387:483–487.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25:955–964.

Mackenzie S, McIntosh L. 1999. Higher plant mitochondria. Plant Cell. 11:571–585.

Marienfeld JR, Reski R, Abel WO. 1991. The first analysed archegoniate mitochondrial gene (*COX3*) exhibits extraordinary features. Curr Genet. 20:319–329.

Muramatsu T, Nishikawa K, Nemoto F, Kuchino Y, Nishimura S, Miyazawa T, Yokoyama S. 1988. Codon and amino-acid specificities of a transfer RNA are both converted by a single post-transcriptional modification. Nature. 336:179–181.

Nakamura M, Kabeya Y, Hashimoto K, Sato N. 2002. Phylogenetic analysis of land plants using introns within the *nad7* gene. Plant Cell Physiol. 43(Suppl):371.

Nishiyama T, Fujita T, Shin-I T, et al. (12 co-authors). 2003. Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. Proc Natl Acad Sci USA. 100:8007–8012.

Nishiyama T, Wolf PG, Kugita M, et al. (12 co-authors). 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol Biol Evol. 21:1813–1819.

Oda K, Yamato K, Ohta E, et al. (11 co-authors). 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol. 223:1–7.

Ogihara Y, Yamazaki Y, Murai K, et al. (14 co-authors). 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res. 33:6235–6250.

Oldenburg DJ, Bendich AJ. 1998. The structure of mitochondrial DNA from the liverwort, *Marchantia polymorpha*. J Mol Biol. 276:745–758.

Oldenburg DJ, Bendich AJ. 2001. Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022

molecules, head-to-tail concatemers, and a 5′ protein. J Mol Biol. 310:549–562.
Petersen J, Brinkmann H, Cerff R. 2003. Origin, evolution, and metabolic role of a novel glycolytic GAPDH enzyme recruited by land plant plastids. J Mol Evol. 57:16–26.
Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. Bioinformatics. 14:817–818.
Pruchner D, Beckert S, Muhle H, Knoop V. 2002. Divergent intron conservation in the mitochondrial *nad2* gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. J Mol Evol. 55:265–271.
Qiu YL, Cho Y, Cox JC, Palmer JD. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature. 394:671–674.
Renzaglia KS, Duff RJ, Nickrent DL, Garbary DJ. 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos Trans R Soc Lond B Biol Sci. 355:769–793.
Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19:1572–1574.
Sato N. 2000. SISEQ: manipulation of multiple sequence and large database files for common platforms. Bioinformatics. 16:180–181.
Sato N. 2006. Origin and evolution of plastids: genomic view on the unification and diversity of plastids. In: Wise RR, Hoober JK, editors. The structure and function of plastids. Dordrecht, the Netherlands: Springer. p. 75–102, the Netherlands.
Sato N, Albrieux C, Joyard J, Douce R, Kuroiwa T. 1993. Detection and characterization of a plastid envelope DNA-binding protein which may anchor plastid nucleoids. EMBO J. 12:555–561.
Sato N, Ehira S. 2003. GenoMap, a circular genome data viewer. Bioinformatics. 19:1583–1584.
Sato N, Ishikawa M, Fujiwara M, Sonoike K. 2005. Mass identification of chloroplast proteins of endosymbiont origin by phylogenetic profiling based on organism-optimized homologous protein groups. Genome Inform. 16:56–68.
Shimodaira H, Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics. 17:1246–1247.
Sibler AP, Dirheimer G, Martin RP. 1986. Codon reading patterns in *Saccharomyces cerevisiae* mitochondria based on sequences of mitochondrial tRNAs. FEBS Lett. 194:131–138.
Sugiura C, Kobayashi Y, Aoki S, Sugita C, 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.
Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochindrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics. 272:603–615.
Swofford DL. 1998. PAUP Phylogenetic analysis using parsimony and other methods. Sunderland (MA): Sinauer Associates.
Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.
Turmel M, Otis C, Lemieux C. 2002a. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci USA. 99:11275–11280.
Turmel M, Otis C, Lemieux C. 2002b. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. Mol Biol Evol. 19:24–38.
Turmel M, Otis C, Lemieux C. 2003. The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell. 15:1888–1903.
Turmel M, Otis C, Lemieux C. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algae relatives of land plants. Mol Biol Evol. 23:1324–1338.
Ward BL, Anderson RS, Bendich AJ. 1981. The mitochondrial genome is large and variable in a family of plants (cucurbitaceae). Cell. 25:793–803.
Weber F, Dietrich A, Weil JH, Marechal-Drouard L. 1990. A potato mitochondrial isoleucine tRNA is coded for by a mitochondrial gene possessing a methionine anticodon. Nucleic Acids Res. 18:5027–5030.
Wolf PG, Karol KG, Mandoli DF, Kuehl J, Arumuganathan K, Ellis MW, Mishler BD, Kelch DG, Olmstead RG, Boore JL. 2005. The first complete chloroplast genome sequence of a lycophyte, Huperzia lucidula (Lycopodiaceae). Gene. 350:117–128.

Franz Lang, Associate Editor

Accepted December 11, 2006

Downloaded from https://academic.oup.com/mbe/article/24/3/699/1238269 by guest on 20 August 2022