Chloroplast Phylogeny Indicates that Bryophytes Are Monophyletic

Tomoaki Nishiyama,* Paul G. Wolf,† Masanori Kugita,‡¹ Robert B. Sinclair,*§¹ Mamoru Sugita,||¹ Chika Sugiru,†¹ Tatsuya Wakahagi,‡¹ Kyoji Yamada,‡¹ Koichi Yoshinaga,‡¹ Kazuo Yamaguchi,‡¹ Kunihiko Ueda,** and Mitsuyasu Hasebe*††

*National Institute for Basic Biology, Okazaki, Japan; †Department of Biology, Utah State University; ‡Faculty of Science, Shizuoka University, Shizuoka, Japan; §The Jackson Laboratory, Bar Harbour, Maine; ||Center for Gene Research, Nagoya University, Nagoya, Japan; †¹Department of Biology, Faculty of Science, Toyama University, Toyama, Japan; **Division of Functional Genomics, Advanced Science Research Center, Kanazawa University, Kanazawa, Japan; ††Department of Molecular Biomechanics, Sokendai, Okazaki, Japan

Opinions on the basal relationship of land plants vary considerably and no phylogenetic tree with significant statistical support has been obtained. Here, we report phylogenetic analyses using 51 genes from the entire chloroplast genome sequences of 20 representative green plant species. The analyses, using translated amino acid sequences, indicated that extant bryophytes (mosses, liverworts, and hornworts) form a monophyletic group with high statistical confidence and that extant bryophytes are likely sisters to extant vascular plants, although the support for monophyletic vascular plants was not strong. Analyzes at the nucleotide level could not resolve the basal relationship with statistical confidence. Bryophyte monophyly inferred using amino acid sequences has a good statistical foundation and is not rejected statistically by other data sets. We propose bryophyte monophyly as the currently best hypothesis.

Introduction

More than 300,000 green plant species now cover the Earth’s surface, and these plants are descended from a common ancestor that lived during the Ordovician, more than 470 MYA (Kenrick and Crane 1997a, 1997b; Kenrick 2000). Recent morphological and molecular phylogenetic studies have revealed that the closest relative of land plants is a group of freshwater green algae (reviewed in Graham and Wilcox [2000] and Chapman and Waters [2002]). However, the earliest branches of the land plant phylogeny remain unclear.

The oldest fossil records of land plants are spore tetrads from the mid-Ordovician (Gray 1993) that are thought to be sporopollenin-bearing meiospores and are suggested to be similar to spores of liverworts because of lamellae in the spore walls. Fossil plant fragments containing similar spores, recently found in Ordovician deposits from Oman, appear to be from embryophytes (Wellman, Osterloff, and Mohiuddin 2003). However, the affinity of these spores to extant liverworts is uncertain because no associated mega-fossils have been found, and similar tetrads are also produced by mosses (Gray 1993). A cladistic analysis of several key morphological characters supports the basal branching of liverworts (Mishler and Churchill 1984), whereas other cladistic analyses of ultrastructural, biochemical, and developmental characters (Garbary and Renzaglia 1998; Renzaglia et al. 2000) support a basal position of hornworts rather than liverworts. Data from antheridial development and the complex process of spermatoogenesis (Garbary, Renzaglia, and Duckett 1993; Renzaglia et al. 2000; Renzaglia and Garbary 2001) support a monophyly of extant bryophytes (mosses, liverworts, and hornworts). Results using molecular data are also controversial. Analyses using DNA sequences of multiple genes suggest a basal position for hornworts (Nishiyama and Kato 1999), although the branching patterns are only weakly supported. Absence of group II introns in three mitochondrial genes in both liverworts and green algae supports the liverwort basal hypothesis (Qiu et al. 1998), although repeated intron losses have been observed in land plant lineages (Qiu et al. 1998) and the possibility of parallel evolution cannot be excluded. The consensus of these conflicting results yields an unresolved polytomy at the base of land plants. Determining these relationships would not only resolve a critical area of the land plant phylogeny, but it would also be important for understanding the characteristics of the earliest land plants and their subsequent evolution.

Combined analyses of DNA sequences from multiple loci have proven useful in inferring deep phylogenetic relationships (Qiu et al. 1999; Pryer et al. 2001), as predicted by simulation studies (Cummings, Otto, and Wakeley 1995). With the accumulation of genome information, the entire chloroplast genome sequences of 15 seed plants, one liverwort (Marchantia polymorpha), four green algae, three red algae, and one cryptophyte have been reported in the GenBank RefSeq division. Because these published data do not cover all major lineages of land plants, we determined the entire chloroplast genome sequences of two pteridophytes, Psilotum nudum (accession number: NC_003386) and Adiantum capillus-veneris (accession number: AY178864; Wolf et al. 2003), a hornwort Anthoceros angustus (accession number: AB086179; Kugita et al. 2003), and a moss, Physcomitrella patens (accession number: AP005672; Sugiu et al. 2003). We analyzed 51 genes that are found in every chloroplast genome of the sampled land plants and chlorophytes (26,937 bp / 8,979 amino acid sites in total), to infer the phylogeny of land plants and to examine the utility of the amino acid versus nucleotide sequence analyses.

¹ These authors contributed equally to this work.

Key words: land plants, bryophytes, codon usage, nucleotide composition, LogDet.

E-mail: mhasebe@nibb.ac.jp.

doi:10.1093/molbev/msi203
Advance Access publication July 7, 2004

Molecular Biology and Evolution vol. 21 no. 10 © Society for Molecular Biology and Evolution 2004; all rights reserved.
Materials and Methods

Coding sequences were extracted from the annotated sequences using bioruby (http://bioruby.org/). Chloroplast genome sequences were determined for a moss (Physcomitrella patens), a hornwort (Anthoceros angustus), and two pteridophytes (Adiantum capillus-veneris and Psilotum nudum). In addition to these four taxa, 16 green plant chloroplast DNA genome sequences in the DNA database (table 1) were used for phylogenetic analyses. The nucleotide sequences of 51 genes that are present in all the chloroplast genomes of green plants were individually aligned using the program fftnsi in MAFFT (Katoh et al. 2003). In addition to these four taxa, 16 green plant chloroplast DNA genome sequences in the DNA database (table 1) were used for phylogenetic analyses. The nucleotide sequences of 51 genes that are present in all the chloroplast genomes of green plants were individually aligned using the program fftnsi in MAFFT (Katoh et al. 2002) version 3.85. The 51 genes are atpA, atpB, atpE, atpF, atpH, atpL, petA, petB, petD, petG, psaA, psaB, psaC, psaD, psbA, psbB, psbC, psbD, psbE, psbF, psbH, psbI, psbK, psbL, psbM, psbN, psbT, psbZ, rbcL, rpl2, rpl14, rpl16, rpl20, rpoB, rpoC1, rpoC2, rps2, rps3, rps4, rps7, rps8, rps11, rps12, rps14, rps18, rps19, ycf3, and ycf4. Unambiguously aligned regions were selected for further analyses and concatenated. The overlapping region of psbD and psbC was excluded. The data matrix is available online as Supplementary Material.

The selected regions included 26,937 nucleotide sites. Because RNA editing is abundant in Anthoceros angustus (Kugita et al. 2003), we used cDNA sequence of A. capillus-veneris. The amino acid matrix comprised 20 taxa and 8,979 amino acid sites. Pairwise maximum-likelihood (ML) distances were calculated using ProtML (Adachi and Hasegawa 1996) and a Neighbor-Joining (NJ) tree was obtained with Njdist (Adachi and Hasegawa 1996). A local rearrangement search was performed with the NJ tree as the starting tree. Local bootstrap probabilities were calculated with resampling using the estimated log-likelihood method (Adachi and Hasegawa 1996). A user tree was prepared for each of the 105 possible bifurcating topologies among six groups of taxa: outgroups (O), mosses (M), liverworts (L), hornworts (H), monilophytes (F), and seed plants (S). The monilophytes include ferns, psilotophytes (whisk ferns), and equisetophytes (horsetails) (Kenrick and Crane 1997b, p. 228; Pryer et al. 2001; Judd et al. 2002, pp. 162–168), and in our study they are represented by Adiantum capillus-veneris and Psilotum nudum. A. capillus-veneris and P. nudum are forced to form a clade, and the topologies within seed plants and within the outgroup were constrained as the ML tree found in the local rearrangement search (fig. 1). The estimated log-likelihood under the JTT model (Jones, Taylor, and Thornton 1992), adjusted with the empirical amino acid frequency (JTT-F model) of each site under the 105 topologies, was calculated using ProtML (Adachi and Hasegawa 1996). The log-likelihood under the JTT model (Jones, Taylor, and Thornton 1992) was calculated for each gene to take account of rate differences among genes.

The likelihood under a discrete gamma model of rate heterogeneity among sites (Yang 1994) was calculated with PAML version 3.13d (Yang 1997). The likelihood under the JTT-F model of each site under the 105 topologies, was calculated using ProtML (Adachi and Hasegawa 1996). The log-likelihood under the JTT model (Jones, Taylor, and Thornton 1992) was calculated for each gene to take account of rate differences among genes. The likelihood under a discrete gamma model of rate heterogeneity among sites (Yang 1994) was calculated with PAML version 3.13d (Yang 1997). The likelihood under the JTT-F model of each site under the 105 topologies, was calculated using ProtML (Adachi and Hasegawa 1996). The log-likelihood under the JTT model (Jones, Taylor, and Thornton 1992) was calculated for each gene to take account of rate differences among genes. The likelihood under a discrete gamma model of rate heterogeneity among sites (Yang 1994) was calculated with PAML version 3.13d (Yang 1997). The likelihood under the JTT-F model of each site under the 105 topologies, was calculated using ProtML (Adachi and Hasegawa 1996). The log-likelihood under the JTT model (Jones, Taylor, and Thornton 1992) was calculated for each gene to take account of rate differences among genes.

Results

Heterogeneity in base composition at all codon positions is recognized as a problem for phylogenetic analyses of chloroplast genes (Steel, Lockhart, and Penny, 1993;
We observed such compositional differences across entire chloroplast genomes; GC composition ranged from 12.9% in the third codon of *Marchantia polymorpha* to 38.3% in *Adiantum capillus-veneris*. We also found differences in codon usage for the first codon positions of leucine (table 1). Fifty-five percent of the leucine codons start with U in *Nephroselmis olivacea* compared to 26% in *Chaetosphaeridium globosum*. Most conventional methods of phylogenetic analysis are statistically consistent (i.e., reach the true topology with infinite data) only when the sequences evolved under a stationary Markov process (Gu and Li 1996). The large differences in the base composition indicate that the evolutionary process is not stationary. Therefore, conventional methods may give positively misleading results with high bootstrap probabilities (Lockhart et al. 1994). To overcome this problem, we applied three approaches: (1) we eliminated codon usage differences by analyzing amino acids sequences; (2) we removed the sites exhibiting compositional heterogeneity, i.e., all the leucine codons and third codon positions; and (3) we used LogDet/paralinear distances (Lake 1994; Lockhart et al. 1994; Steel 1994; Gu and Li 1996), which are consistent under nonstationary processes.

**Analysis of Deduced Amino Acid Sequence Data**

A local rearrangement search starting with the NJ tree (Saitou and Nei 1987) under the JTT-F model (Jones, Taylor, and Thornton 1992; Adachi and Hasegawa 1996) found an ML tree (ln likelihood = −89478.09; fig. 1). The three bryophytes, *Marchantia polymorpha*, *Physcomitrella patens*, and *Anthoceros angustus*, formed a monophyletic group with high local bootstrap probability (99%). Previously well-recognized monophyletic groups, the eudicots, monocots, angiosperms, seed plants, and land plants (reviewed in Judd et al. 2002, pp. 153–183), each formed a monophyletic group with high local bootstrap probability. Vascular plants formed a monophyletic group, though the bootstrap value was low (55%). This contrasts previous studies with high bootstrap support for vascular plants (Hedderson, Chapman, and Cox 1998; Nickrent et al. 2000). To test whether other possible trees have a significantly less likelihood than the ML tree, the bootstrap test and AU test (Shimodaira 2002) were used for all 105 possible topologies among seed plants, monilophytes, hornworts, liverworts, mosses, and outgroups. We constrained the topology within seed plants and within outgroup taxa. In the bootstrap test, topology 1 (fig. 2), which had the same topology as the ML tree, had 52.8% boot-
Table 2

Table 2 Statistical Test for the 105 Topologies Under Various Models of Rate Heterogeneity

<table>
<thead>
<tr>
<th>Topology</th>
<th>JTT-F AU</th>
<th>Independent Gene AU</th>
<th>Gamma 8 Category AU</th>
<th>JTT-F BP</th>
<th>Independent Gene BP</th>
<th>Gamma 8 Category BP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.707</td>
<td>0.528</td>
<td>0.981</td>
<td>0.847</td>
<td>0.645</td>
<td>0.836</td>
</tr>
<tr>
<td>2</td>
<td>0.532</td>
<td>0.443</td>
<td>0.151</td>
<td>0.107</td>
<td>0.321</td>
<td>0.306</td>
</tr>
<tr>
<td>3</td>
<td>0.019</td>
<td>0.005</td>
<td>0.018</td>
<td>0.006</td>
<td>0.005</td>
<td>0.003</td>
</tr>
</tbody>
</table>

* Topologies are shown in figure 2.
* AU and BP are P values for the approximately unbiased test and bootstrap probability for the bootstrap test, respectively.

To consider rate heterogeneity among genes and sites, we performed separate gene analyses, in which each gene is assumed to have evolved independently (“independent gene” in table 2). This was performed under the JTT model using ProtML. We also performed a discrete gamma approximation with eight categories (“gamma 8 category” in table 2) using PAML (Yang 1997). The log-likelihood of every tree was calculated. These analyses also supported topology 1 as the ML tree (table 2; table S1 in the online Supplementary Material).

To consider rate heterogeneity among genes and sites, we performed separate gene analyses, in which each gene is assumed to have evolved independently (“independent gene” in table 2). This was performed under the JTT model using ProtML. We also performed a discrete gamma approximation with eight categories (“gamma 8 category” in table 2) using PAML (Yang 1997). The log-likelihood of every tree was calculated. These analyses also supported topology 1 as the ML tree (table 2; table S1 in the online Supplementary Material).

Fig. 3.—The most parsimonious tree using nucleotides of the first and second codon positions. 4,228 of 17,958 sites were parsimony-informative. Tree length = 14,487. Consistency index excluding uninformative characters = 0.52. Retention index = 0.62. Bootstrap probabilities are indicated on branches. Branch lengths are proportional to the number of substitutions on the branches by ACCTRAN optimization.

**Discussion**

The monophyly of extant bryophytes was inferred by our analyses using translated amino acid sequence data of currently available chloroplast genomes. Analyses using nucleotide sequences did not resolve phylogenetic relationships of the three bryophytes with statistical confidence. Amino acid and nucleotide data each have their own merits and drawbacks. Amino acid sequences have frequently been used to infer deep phylogeny because they avoid problems with saturation of silent substitution and differential GC content (Russo, Takezaki, and Nei 1996; Nishiyama and Kato 1999; Simmons, Ochoterena, and Freudenstein 2002 and references therein). But, Simmons, Ochoterena, and Freudenstein (2002) criticized the use of amino acid translation, showing that nucleotide data outperform deduced amino acid data in a phylogenetic analysis of angiosperms, which diverged during the last 200 Myr (Judd et al. 2002, pp. 175–176). However, analyses using nucleotide sequences may not be robust when nucleotide compositions are apparently different among taxa, because phylogenetic analyses usually assume a stationary Markov process (Lockhart et al. 1994). In an attempt to remove characters with different nucleotide compositions, we excluded third codon positions (fig. 3) or the nucleotide sites coding leucine and third codon positions (fig. 4). We also compensated for differences in nucleotide composition by LogDet transformation (fig. 5). Neither the removal of characters nor the LogDet transformation resolved the relationships of extant bryophytes with high statistical support, whereas resolution with high statistical support was achieved with amino acid translation. These results might be explained if a uniform rate of change does not describe evolution of the fourfold degener-
erate sites. These results may also suggest that nucleotide composition bias can increase in proportion to divergence time among taxa. Nucleotide sequences including the third codon position data are more likely to be useful for inferring phylogeny among taxa that diverged relatively recently, such as angiosperms, which diverged within ~200 Myr (Judd et al. 2002, pp. 175–176). If compositional heterogeneity is not extreme and the spatial pattern of substitution is simple, it would follow that nucleotide data will outperform deduced amino acid data because the former has more phylogenetic information than the latter. However, when sequences become highly diverged, the spatial substitution pattern may become more complex and problematic for distance-based correction formulae (Steel, Huson, and Lockhart 2000). The main lineages of land plants probably diverged as many as 470 MYA (Kenrick and Crane 1997; Kenrick 2000). While our hypothesis was supported with high statistical confidence with presently available data, the bryophyte monophyly hypothesis should be tested with more data and further analysis. Current analyses include only one sample per group from mosses, liverworts, and bryophytes, which are supported by molecular sequence data (Nickrent et al. 2000; Pryer et al. 2001). If our inference of monophyletic bryophytes reflects the true relationships, this would suggest that spermatogenesis characters have retained more phylogenetic signal than other morphological characters. Phylogenetic studies of land plant relationships using morphological characters are also incongruent. Most studies suggest that extant bryophytes are paraphyletic (Mishler and Churchill 1984, 1985; Renzaglia et al. 2000), whereas monophyly of extant bryophytes and a sister relationship of mosses and liverworts were inferred using characters of spermatogenesis (Garbary, Renzaglia, and Duckett 1993; Renzaglia et al. 2000; Renzaglia and Garbary 2001). Analyses of spermatogenesis characters also supported monilophytes involving ferns, psilotophytes (whisk ferns), and equisetophytes (horsetails), which are supported by DNA sequence data (Nickrent et al. 2000; Pryer et al. 2001). If our inference of monophyletic bryophytes reflects the true relationships, this would suggest that spermatogenesis characters have retained more phylogenetic signal than other morphological characters.

Phylogenetic analyses using DNA sequences suggest that extant bryophytes are paraphyletic (Lewis, Mishler, and Vilgalys 1997; Duff and Nickrent 1999; Nishiyama and Kato 1999; Nickrent et al. 2000). However, only the MP-3Ti analysis by Nickrent et al. (2000) provided evidence of strong statistical support. Moreover, monophyly of bryophytes was not rejected in the Kishino-Hasegawa (1989) test by Nickrent et al. (2000). The sister relationship of mosses and liverworts received relatively strong support in recent analyses using multiple genes (Nishiyama and Kato 1999; Nickrent et al. 2000). Marchantia appears on a basally diverging branch in land plants with high bootstrap probability in a combined analysis of atpB, rbcL, nad5, and the small subunit rRNA gene (Karol et al. 2001). However, that study used DNA sequences of chloroplast genes without attempting to compensate for the nucleotide composition bias, which means that the bootstrap probabilities may not be a good estimator of the reliability of branches and tree topologies (Lockhart et al. 1994; Swofford et al. 2001; Felsenstein 2003, pp. 272–274). Therefore, the extant bryophyte monophyletic hypothesis inferred in this study stands as the best one statistically supported by molecular sequence data.

Phylogenetic analyses using DNA sequences suggest that extant bryophytes are paraphyletic (Lewis, Mishler, and Vilgalys 1997; Duff and Nickrent 1999; Nishiyama and Kato 1999; Nickrent et al. 2000). However, only the MP-3Ti analysis by Nickrent et al. (2000) provided evidence of strong statistical support. Moreover, monophyly of bryophytes was not rejected in the Kishino-Hasegawa (1989) test by Nickrent et al. (2000). The sister relationship of mosses and liverworts received relatively strong support in recent analyses using multiple genes (Nishiyama and Kato 1999; Nickrent et al. 2000). Marchantia appears on a basally diverging branch in land plants with high bootstrap probability in a combined analysis of atpB, rbcL, nad5, and the small subunit rRNA gene (Karol et al. 2001). However, that study used DNA sequences of chloroplast genes without attempting to compensate for the nucleotide composition bias, which means that the bootstrap probabilities may not be a good estimator of the reliability of branches and tree topologies (Lockhart et al. 1994; Swofford et al. 2001; Felsenstein 2003, pp. 272–274). Therefore, the extant bryophyte monophyletic hypothesis inferred in this study stands as the best one statistically supported by molecular sequence data.

**FIG. 5.**—Neighbor-Joining tree using LogDet distance based on 2,695 fourfold degenerate sites. The number on each branch indicates a bootstrap probability (%) estimated by 10,000 replicates. Branch lengths are proportional to the LogDet distance.
h hornworts and lack a major lineage of land plants, lycophytes. Some major lineages of bryophytes and lycophytes should be included based on previous phylogenetic analyses (Lewis, Mishler, and Vilgalys 1997; Garbary and Renzaglia 1998; Hedderson, Chapman, and Cox 1998; Nickrent et al. 2000; Renzaglia et al. 2000). Sphagnum, Takakia, Andreaea., and Andraeobryum from mosses and Treubia, Haplomitrium, Blasia, one simple thalloid, and one leafy liverwort from liverworts are candidate representatives. In addition to one homosporous and one heterosporous lycophyte, equisetophytes should be targeted for vascular plants. Furthermore, data from Charales may be a better outgroup to land plants than Coleochaetales (Turnel et al. 2002). Mitochondrial and nuclear genome data are also candidate sources of phylogenetic information that could provide an independent test of results based on chloroplast genome data.

The fossil record of plants older than rhyniophytes includes cuticles (Edwards, Duckett, and Richardson 1995) and spor e tetrad s (Gray 1993). These fossils have been considered to be related to extant liverworts (Edwards, Duckett, and Richardson 1995; Taylor 1995). Our results are not in line with this interpretation because we infer that liverworts do not descend from a basal node in bryophytes (fig. 2). Our results, however, are consistent with a "bryophyte-like ancestor" hypothesis (Graham and Gray 2001), whereby the above fossils represent a common ancestor of the three extant groups of bryophytes.

Supplementary Material

The complete table for statistical tests among 105 topologies is available online as a supplementary table. The nucleotide data matrix is available online in nexus format.

Acknowledgments

We thank two anonymous reviewers, Brent Mishler, Hiroyuki Akiyama, Hironori Deguchi, and members of the Hasebe laboratory for discussion and comments on the manuscript. Work on Anthoceros angustus has been published under the name of A. formosae. Thanks also to Carol Rowe for help sequencing the Adiantum chloroplast genome. The computations were partly done in the Computer Lab of the National Institute for Basic Biology. The research was supported by National Science Foundation grant #DEB-0228432 to P.G.W. and grants by MEXT and JSPS. Adiantum capillus-veneris, Anthoceros angustus, Physcomitrella patens, and Psilotum nudum were sequenced by P.G.W., R.B.S, and M.H.; M.K. and K.Y.; M.S. and C.S.; and T.Y. and K.Y, respectively. Correspondence in general should be addressed to M.H. (e-mail: mhasebe@nibb.ac.jp). Correspondence and requests for materials for chloroplast DNA s should be addressed to P.G.W. (Adiantum, e-mail: wolf@biology.usu.edu), K.Y. (Anthoceros, e-mail: sckyos@ipc.shizuoka.ac.jp), M.S. (Physcomitrella, e-mail: sugita@gene.nagoya-u.ac.jp), and K.Y. (Psilotum, e-mail: yamada@sci.toyama-u.ac.jp).

Literature Cited


Chloroplast Phylogeny Indicates Monophyletic Bryophytes 1819


Peter Lockhart, Associate Editor

Accepted June 28, 2004