at www.sciencemag.org/cgi/content/full/294/5550/2348/DC1.
16. D. M. Hillis, B. K. Mable, C. Moritz, Eds., *Molecular Systematics* (Sinauer, Sunderland, MA, 1996), pp. 515–543.
17. N. Goldman, J. P. Anderson, A. G. Rodrigo. *Syst. Biol.* **49**, 652 (2000).
18. B. Larget, D. Simon, *Mol. Biol. Evol.* **16**, 750 (1999).
19. J. P. Huelsenbeck, F. Ronquist. *Bioinformatics* **17**, 754 (2001).
20. D. M. Hillis, J. J. Bull, *Syst. Biol.* **42**, 182 (1993).
21. E. Eizirik, W. J. Murphy, S. J. O'Brien *J. Hered.* **92**, 212 (2001).
22. A. R. Wallace, *The Geographical Distribution of Animals* (Hafner, New York, 1962).
23. W. D. Matthew, *Ann. N.Y. Acad. Sci.* **24**, 171 (1915).
24. R. Rainger, *Agenda for Antiquity: Henry Fairfield Osborn and Vertebrate Paleontology at the American Museum of Natural History, 1890–1935* (Univ. of Alabama Press, Tuscaloosa, 1991).
25. A. Rambaut, L. Bromham, *Mol. Biol. Evol.* **15**, 442 (1998).
26. N. Takezaki, A. Rzhetsky, M. Nei, *Mol. Biol. Evol.* **12**, 823 (1995).
27. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
28. A. G. Smith, D. G. Smith, B. M. Funnell, *Atlas of Cenozoic and Mesozoic Coastlines*. (Cambridge Univ. Press, Cambridge, 1994).
29. W. W. Hay *et al.*, *Geol. Soc. Am. Spec. Pap. 332* (1999), pp. 1–48.
30. D. Penny, M. Hasegawa, P. J. Waddell, M. D. Hendy, *Syst. Biol.* **48**, 76 (1999).
31. M. C. McKenna, S. K. Bell, *Classification of Mammals Above the Species Level* (Columbia Univ. Press, New York, 1997).
32. M. Foote, J. P. Hunter, C. M. Janis, J. J. Sepkoski Jr., *Science* **283**, 1310 (1999).
33. J. D. Archibald, A. O. Averlanov, D. G. Ekdale, *Nature* **414**, 62 (2001).
34. M. S. Springer *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6241 (2001).
35. Supported by the NSF (M.S.S.), the Training and Mobility of Researchers (TMR) program of the European Commission (W.W.d.J. and M.J.S.), and the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil (E.E.).

17 October 2001; accepted 12 November 2001

The Closest Living Relatives of Land Plants

Kenneth G. Karol,[1]* Richard M. McCourt,[2] Matthew T. Cimino,[1] Charles F. Delwiche[1]

The embryophytes (land plants) have long been thought to be related to the green algal group Charophyta, though the nature of this relationship and the origin of the land plants have remained unresolved. A four-gene phylogenetic analysis was conducted to investigate these relationships. This analysis supports the hypothesis that the land plants are placed phylogenetically within the Charophyta, identifies the Charales (stoneworts) as the closest living relatives of plants, and shows the Coleochaetales as sister to this Charales/land plant assemblage. The results also support the unicellular flagellate *Mesostigma* as the earliest branch of the charophyte lineage. These findings provide insight into the nature of the ancestor of plants, and have broad implications for understanding the transition from aquatic green algae to terrestrial plants.

The evolutionary origin of the embryophytes (or land plants) from their green algal ancestor was a pivotal event in the history of life. This monophyletic group has altered the biosphere and now dominates the terrestrial environment, but uncertainty as to the identity of their closest living relatives has persisted in the literature after more than a century of scrutiny (*1–3*). Morphological and molecular studies have identified two distinct lineages within the green plants *sensu lato*, termed Charophyta and Chlorophyta. The Charophyta comprise the land plants and at least five lineages (orders) of fresh water green algae, and are sister to the Chlorophyta, which consist of essentially all other green algae. Previous molecular analyses have verified monophyly of most of the charophyte orders (*4–6*), but branching patterns among these lineages have been only weakly supported, with results that were sensitive to taxon selection and method of phylogenetic reconstruction. Similarly, analyses of morphological and genome structural data have clarified some relationships (*7–10*), but have been limited by the number of characters available, uncertain homology assessment, and a lack of character independence.

Identifying the closest living relatives of land plants has been difficult. Roughly 470 million years of evolution since the colonization of the land, coupled with rapid radiation and numerous extinction events (*2, 3, 11*), has resulted in an inherently difficult phylogenetic problem, with much information from the early, common history of evolution obscured by subsequent evolution in the now independent lineages (*12*).

To investigate the evolutionary origin of land plants and identify the closest living relatives of this group, we analyzed DNA sequence data from four genes representing three plant genomes: *atpB* and *rbcL* (plastid), *nad5* (mitochondrial), and the small subunit (SSU) rRNA gene (nuclear). The data set used for phylogenetic analyses excludes introns and unalignable regions for a total length of 5147 base pairs [Appendix 1 (*13*)] (*14*). We sampled 34 representative charophytes, including eight land plants, and six outgroup taxa [Appendix 2 (*13*)]. The data were analyzed with Bayesian inference (BI), maximum likelihood (ML), maximum parsimony (MP), and minimum evolution with two distance measures [LogDet (ME-ld) and maximum likelihood (GTR+I+Γ; ME-ml) distances] [Appendix 3 (*13*)]. Both BI and ML are probabilistic methods that utilize explicit models of sequence evolution to test phylogenetic hypotheses. Advantages of BI are that it is relatively fast and provides probabilistic measures of tree strength that are more directly comparable with traditional statistical measures than those more commonly used in phylogenetic analyses (*15, 16*). To measure phylogenetic stability, posterior probabilities (PP) as inferred by BI were calculated and bootstrapping was performed for the ML, MP, and ME analyses.

Using BI and ML on the combined four-gene data set (Fig. 1), we found the order Charales sister to the land plants with strong statistical support (PP = 1.0, ML = 94) and a monophyletic Coleochaetales sister to the Charales/land plant clade (PP = 1.0, ML = 59). The MP and ME analyses [Appendix 4 (*13*)] also support the result that Charales have a closer relationship to land plants than do Coleochaetales (MP = 80, ME-ld = 97, ME-ml = 92). The overall structure of the best tree is consistent with previous work in that the classically recognized orders were also recovered (land plants, PP = 1.0, ML = 100, MP = 100, ME-ld = 100, ME-ml = 100; Charales, PP = 1.0, ML = 100, MP = 100, ME-ld = 100, ME-ml = 100; Coleochaetales, PP = 1.0, ML = 62, MP = <50, ME-ld = 75, ME-ml = <50; Zygnematales, PP = 1.0, ML = 99, MP = 93, ME-ld = 68, ME-ml = <50; and Klebsormidiales PP = 1.0, ML = 100, MP = 100, ME-ld = 100, ME-ml = 100). There was also support for placement of the enigmatic filamentous alga *Entransia* (*6*) with the Klebsormidiales (PP = 1.0, ML = 77, MP = 77, ME-ld = <50, ME-ml = 64). The rare, monotypic genus *Chlorokybus* was found sister to the remainder of the unambiguous charophytes, while all analyses strongly support the inclusion of *Mesostigma* within the Charophyta (PP = 1.0, ML = 97, MP = 100, ME-ld = 100, ME-ml = 100).

The phylogenetic placement of *Mesostigma*, a unicellular, scaly green flagellate

[1]Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA. [2]Department of Botany, Academy of Natural Sciences, 1900 Benjamin Franklin Parkway, Philadelphia, PA 19103, USA.

*To whom correspondence should be addressed. E-mail: karol@umail.umd.edu

has been controversial. Traditionally classified with like forms as a prasinophyte, it also has been allied with the Charophyta. The phylogenetic position of *Mesostigma* is critical to understanding the evolution of form and structure in the lineage that gave rise to land plants. Like the results presented here, analyses of actin sequences place *Mesostigma* at the base of the Charophyta (*17*), and analyses of SSU rRNA gene sequence data place it among them (albeit in close association with *Chaetosphaeridium*, a grouping not supported by other data) (*5, 18*). By contrast, maximum likelihood analyses of amino-acid data from both the plastid and mitochondrial genomes of *Mesostigma* find strong support for placement of this genus as sister to all green algae rather than as a basal charophyte lineage (*19, 20*). The latter analyses differ from those presented here in the number of taxa sampled (8 versus 40). When divergence times are large and internal branches short, limited taxon sampling can lead to inaccurate phylogenies (*12*). If taxon sampling explains this conflict, then one would predict convergence on the phylogeny presented here as additional organellar genomes become available.

Both Charales and Coleochaetales have long been considered to be close relatives of the land plants (*1, 21–23*). Key morphological characters uniting these three lineages include branched filamentous growth, oogamous sexual reproduction, and phragmoplastic cell division, along with a suite of ultrastructural and biochemical features (*2*). In light of similar morphological traits (i.e., parenchyma-like tissue, placental transfer cell wall ingrowths, and zygote retention), the genus *Coleochaete* and, in some instances, a single species, *C. orbicularis*, has been discussed as a possible sister taxon to land plants (*8, 24*). Our results indicate that the Coleochaetales are monophyletic and less closely related to the land plants than the Charales. Both Bayesian inference and bootstrap analyses permit evaluation of alternative hypotheses; we were unable to identify any alternative hypothesis with nontrivial support (*25*).

The Charales also share numerous characteristics with land plants, some of which are not found in the Coleochaetales. These include gross sperm morphology and ultrastructure (*26*), numerous discoidal chloroplasts per cell, protonemal filaments, complete absence of zoospores (sperm are the only flagellate cells), and encasement of the egg by sterile jacket cells (cortication) prior to fertilization (*10, 21*). Our data suggest that many of the similarities between Charales and land plants reflect homology rather than convergent evolution. Cortication of the zygote reminiscent of that in Charales is found in some species of *Coleochaete*, but occurs only after fertilization of the egg, and zygote cortication is not thought to occur in *Chaetosphaeridium* (*10*). In addition, primary plasmodesmata have been confirmed in the Charales, a character shared with land plants (*27*). Although plasmodesmata have been described in *Coleochaete*, it is unknown whether their development is primary or secondary in nature.

Identification of the Charales as the sister taxon to land plants with the Coleochaetales as sister to the Charales/land plant clade suggests that the common ancestor of land plants was a branched, filamentous organism with a haplontic life cycle and oogamous reproduction. The early stages of development in the Charales involve formation of protonemal filaments reminiscent of those found in some mosses and other land plants, which suggests that a similar heteromorphic development might have occurred in the common ancestor. Other characteristics of this ancestor, including both developmental and biochemical features, may explain not only how their descendants came to survive on land, but also how they ultimately came to dominate terrestrial ecosystems. Moreover, the charophytes have important applications in a wide range of disciplines (Charales in cell biology, Coleochaetales in ultrastructure, and Zygnematales in physiology) (*10*). Consequently, a robust phylogeny relating these taxa to land plants can place this work in an evolutionary context and lead to the identification and development of appropriate model systems for future studies.

Although it is tempting to envision the origin of land plants as having been from amorphous pond scum, these data indicate that the common ancestor of land plants and their closest algal relatives was a relatively complex organism. The extant Charales are the remnants of a once diverse, but now largely extinct, group which includes some of the oldest known plant fossils [roughly 420 million years ago (Ma) from the late Ordovician] (*11, 28*). While the fossil record for the other charophyte orders is fragmentary at best (*29*), the molecular phylogenetic data presented here (Fig. 1) suggest that these lineages diversified more than 470 Ma. While not species-rich, these algae hold a key position in the tree of life and, consequently, represent an important part of eukaryotic diversity.

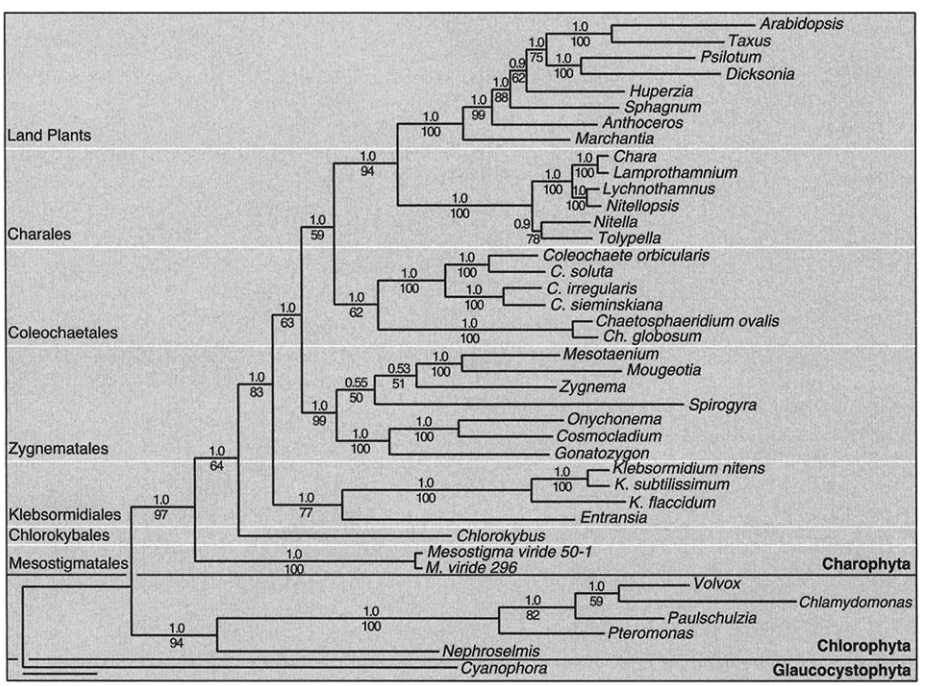

Fig. 1. Phylogenetic relationships for Charophyta determined by Bayesian inference from the combined four-gene data set. The maximum likelihood tree (–ln = 64499.87863) was of identical topology. Posterior probabilities are noted above branches and maximum likelihood bootstrap values are below branches. The topology is drawn with *Cyanophora* rooting the tree. Branch lengths are mean values and are proportional to the number of substitutions per site (bar, 0.05 substitutions/site). Taxonomy is modified from (*23*).

References and Notes

1. F. O. Bower, *The Origin of Land Flora. A Theory Based upon the Facts of Alternation* (Macmillan, London, 1908).
2. L. E. Graham, *The Origin of Land Plants* (Wiley, New York, 1993).
3. P. Kenrick, P. R. Crane, *The Origin and Early Diversification of Land Plants*, Smithsonian Series in Comparative Evolutionary Biology (Smithsonian Institution Press, Washington, DC, 1997).
4. R. L. Chapman *et al.*, in *Systematics of Plants II*, D. E. Soltis, P. S. Soltis, J. J. Doyle, Eds. (Kluwer Academic, Norwell, MA, 1998), pp. 508–540.
5. B. Marin, M. Melkonian, *Protist* **150**, 399 (1999).
6. R. M. McCourt *et al.*, *J. Phycol.* **36**, 747 (2000).
7. H. J. Sluiman, *Plant Syst. Evol.* **149**, 217 (1985).

8. L. E. Graham, C. F. Delwiche, B. D. Mishler, *Adv. Bryol.* **4**, 213 (1991).
9. B. D. Mishler, S. P. Churchill, *Brittonia* **36**, 406 (1984).
10. L. E. Graham, L. W. Wilcox, *Algae* (Prentice-Hall, Upper Saddle River, NJ, 2000).
11. M. Feist, N. Grambast-Fessard, in *Calcareous Algae and Stromatolites*, R. Riding, Ed. (Springer-Verlag, Berlin, 1991), pp. 189–203.
12. J. Felsenstein, *Syst. Zool.* **27**, 401 (1978).
13. Supplementary material is available on *Science* Online at www.sciencemag.org/cgi/content/full/294/5550/2351/DC1.
14. Polymerase chain reaction (PCR) and sequencing: Total cellular DNA was isolated by the CTAB method [J. J. Doyle, J. L. Doyle, *Phytochem. Bull.* **19**, 11 (1987)], UNSET method (a high-urea, SDS extraction buffer) or using the Nucleon Phytopure Plant DNA extraction kit (Amersham Pharmacia Biotech) following the manufacturer's protocol from fresh thalli growing in uni-algal condition. The genes were amplified by PCR with gene specific primers (*atpB* upstream: 5′-TGTTACTTGTGAAGTTCAACA-3′; *atpB* downstream: 5′-CTAAATAAAATGCTTGTTCAGG-3′; *rbcL* upstream: 5′-ATGTCACCACAAACAGAAACTAAAGC-3′; *rbcL* downstream: 5′-AATTCAAATTTATTTCTTTCC-3′; *nad5* upstream: 5′-GTAGGTGATTTTGGATTAGC-3′: *nad5* downstream: 5′-GTACCTAAACCAATCATCATATC-3′; SSU upstream: 5′-GTAGTCATATGCTTGTCTC-3′: SSU downstream: 5′-CTTGTTACGACTTCTCCT-3′) and sequenced using either an ABI-PRISM 377 or 3100 DNA sequencer (PE Applied Biosystems) according to the manufacturer's protocols. The resulting sequence chromatograms were edited and compiled into a single alignment using Sequencher 3.1.1 (Gene Codes Corp.) and exported in NEXUS format for phylogenetic analyses. Many published SSU rRNA gene sequences were difficult to align to published secondary structure models. Small subunit sequences that could not be matched to such structure models were resequenced for this study (*13*). A single intron was found in the *Coleochaete orbicularis nad5* sequence and the distribution of introns in *nad5* was examined in the taxa within our study. No introns were found in any other species of *Coleochaete* or other algal charophyte *nad5* sequence sampled. Introns with the same insertion point as that of *C. orbicularis* were only found in *Sphagnum* (a moss) and *Marchantia* (a liverwort) which share a sequence identity of 69.39%, compared with only 37.82% and 37.81% to *C. orbicularis*, respectively. *Anthoceros* (a hornwort) has an apparently unrelated intron inserted 128 base pairs downstream with 37.35% identity with that of *Sphagnum*, 35.99% identity to *Marchantia*, and 39.46% to *C. orbicularis*. For comparison, pairs of random sequences with similar base composition and length as the natural sequences had an average of 37.78% sequence identity. These data suggest that the *C. orbicularis nad5* intron was acquired independently from that shared by *Sphagnum* and *Marchantia*.
15. J. P. Huelsenbeck, J. P. Bollback, in *Handbook of Statistical Genetics*, M. Bishop, Ed. (Wiley, London, 2001).
16. J. P. Huelsenbeck, F. Ronquist, R. Nielsen, J. P. Bollback, *Science* **294**, 2310 (2001).
17. D. Bhattacharya, K. Weber, S. S. An, W. Berning-Koch, *J. Mol. Evol.* **47**, 544 (1998).
18. H. J. Sluiman, C. Guihal, *J. Phycol.* **35**, 395 (1999).
19. C. Lemieux, C. Otis, M. Turmel, *Nature* **403**, 649 (2000).
20. C. Lemieux, C. Otis, M. Turmel, in press.
21. F. E. Fritsch, *The Structure and the Reproduction of the Algae* (Cambridge Univ. Press, London, 1935), vol. I.
22. J. D. Pickett-Heaps, H. J. Marchant, *Cytobios* **6**, 255 (1972).
23. K. R. Mattox, K. D. Stewart, in *The Systematics of the Green Algae*, D. E. G. Irvine, D. M. John, Eds. (Academic Press, London, 1984), pp. 29–72.
24. B. D. Mishler, S. P. Churchill, *Cladistics* **1**, 305 (1985).
25. Alternative hypotheses that were explored include: *Coleochaete orbicularis* sister to land plants, PP = 0.0, ML = 0.0%; *Coleochaete* sister to land plants, PP = 0.0, ML = 0.0%; Coleochaetales sister to land plants, PP = 0.0, ML = 0.0%; Coleochaetales sister to Charales, PP = 0.0, ML = 0.4%.
26. T. M. Duncan, K. S. Renzaglia, D. J. Garbary, *Pl. Syst. Evol.* **204**, 125 (1997).
27. M. E. Cook, L. E. Graham, C. E. J. Botha, C. A. Lavin, *Am. J. Bot.* **84**, 1169 (1997).
28. M. Feist, R. Feist, *Nature* **385**, 401 (1997).
29. H. Tappan, *The Paleobiology of Plant Protists* (Freeman, New York, 1980).
30. We thank T. Bachvaroff, T. Cooke, G. French, M. Hibbs, J. Lewandowski, T. Marushak, and E. Zimmer for critical comments; C. Drummond, S. Snyder, and A. Zeccardi for technical assistance; J. Bollback and J. Huelsenbeck for important discussions and assistance with Bayesian analyses; M. Casanova, M. Feist, and V. Proctor for material; F. Lang *et al.*, C. Lemieux, C. Otis, and M. Turmel for unpublished sequence data; and S. Fritz, A. Kaspar, R. Sudman, K. Sytsma, and the GPPRGC ("Deep Green"; USDA) for help with development of this project. This work was supported by NSF grant DEB-9978117 and is dedicated to the memory of C. C. Delwiche.

7 August 2001; accepted 9 November 2001

Water Permeation Across Biological Membranes: Mechanism and Dynamics of Aquaporin-1 and GlpF

Bert L. de Groot and Helmut Grubmüller*

"Real time" molecular dynamics simulations of water permeation through human aquaporin-1 (AQP1) and the bacterial glycerol facilitator GlpF are presented. We obtained time-resolved, atomic-resolution models of the permeation mechanism across these highly selective membrane channels. Both proteins act as two-stage filters: Conserved fingerprint [asparagine-proline-alanine (NPA)] motifs form a selectivity-determining region; a second (aromatic/arginine) region is proposed to function as a proton filter. Hydrophobic regions near the NPA motifs are rate-limiting water barriers. In AQP1, a fine-tuned water dipole rotation during passage is essential for water selectivity. In GlpF, a glycerol-mediated "induced fit" gating motion is proposed to generate selectivity for glycerol over water.

Aquaglyceroporins constitute a large family of integral membrane proteins that facilitate highly efficient and specific passive permeation of water and other small uncharged solutes across biological membranes (*1*, *2*). Osmotic water regulation is essential for all life forms, and aquaglyceroporins are found throughout nature, with nearly 300 proteins identified and sequenced so far. In humans, more than 10 different aquaporins with specialized functionality are expressed in tissues as diverse as kidney, red blood cells, and brain. Malfunctions of these proteins cause a wide range of diseases, including nephrogenic diabetes insipidus, congenital cataract, and impaired hearing (*1*, *3*, *4*).

The human water channel aquaporin-1 (AQP1) (Fig. 1) (*5*) permeates water molecules across the membrane at a rate of 3×10^9 s^{-1} per channel (*6–8*), with an activation energy nearly as low as the one associated with the self-diffusion rate in bulk water (*8*). The homologous bacterial glycerol facilitator GlpF is selective for glycerol and other linear alcohols (*9*, *10*) and shows lower water permeability (*10*, *11*) despite a wider pore. The low activation energies allow one to study entire water-permeation events through both proteins by molecular dynamics (MD) simulations in "real time," without the need to accelerate the process by additional driving forces.

The structural models of human AQP1 (*12–14*) and the atomic structure of GlpF from *Escherichia coli* (*15*) have confirmed and extended the early sequence-based "hourglass" model (*16*): The walls of the pore are formed by six transmembrane helices, 1 through 6, connected by five loops, A through E; the pore center is formed by the two highly conserved fingerprint asparagine-proline-alanine (NPA) motifs contained in the B and E loops, which fold back into the protein. The C-terminal halves of these two loops form two short helices that together form a seventh, kinked transmembrane helix. Despite a wealth of experimental data, major issues need to be resolved at the atomic level: How is this extremely high rate achieved while maintaining strict selectivity? How are ions, and particularly protons, excluded, even though they are known to be conducted well by hydrogen-bonded water chains (*17*, *18*)? What is the exact pathway of water molecules through the channel? How are the known structural differences between AQP1 and GlpF reflected in the permeation mechanism? An especially intriguing question is how GlpF facilitates permeation of (larger) glycerol mole-

Theoretical Molecular Biophysics Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, 37077 Göttingen, Germany.

*To whom correspondence should be addressed. E-mail: hgrubmu@gwdg.de