

---

## Paraphyly of bryophytes inferred using chloroplast sequences

Asheesh Shanker

Department of Bioscience and Biotechnology, Banasthali University,  
Rajasthan, India, Email: ashomics@gmail.com

### **Abstract**

The availability of organelle genome sequences in public databases facilitates the reconstruction of land plants phylogeny. Therefore the present phylogenetic analysis was conducted using chloroplast genome sequences of plants to study phylogenetic relationship of bryophytes. The representative taxa were selected considering the availability of organelle genome sequences in database. Inferred tree topologies showed paraphyly of bryophytes with sister relationship between hornworts and vascular plants irrespective of the method (Maximum Likelihood and Bayesian Inference) and type of sequences (Nucleotide and Protein) used. Among bryophytes liverworts are the first to diverge therefore appeared as sister to land plants. Moreover land plants formed a monophyletic group.

### **Introduction**

Bryophytes are the earliest land plants sharing ancestry with charophycean algae (Kenrick and Crane 1997, Renzaglia et al. 2000, Qiu et al. 2006, Qiu 2008, Finet et al. 2010, Gao et al. 2010). The three extant bryophyte lineages (liverworts, mosses and hornworts) are supposed to diverge before the lineage ancestral to existing tracheophytes (Ligrone et al. 2012). Several molecular phylogenetic studies based on sequences from chloroplast, mitochondrial and nuclear genomes were conducted to resolve the relationship among bryophytes lineages but they give conflicting results (Mishler et al. 1994, Lewis et al. 1997, Nickrent et al. 2000, Nishiyama et al. 2004, Groth-Malonek et al. 2005, Qiu et al. 2006, Shanker et al. 2011, Chang and Graham 2011, Shanker and Sharma 2012). As more and more chloroplast genomes are added in public databases their sequences can be used to generate organelle genome based phylogenies of land plants. Therefore the present study was designed to test the established views of bryophytes relationship based on chloroplast genome sequences.

## Materials and methods

A total of 18 organisms were selected with sequenced chloroplast genomes (Table 1). Coding sequences (CDS) and corresponding protein sequences of these organelle genomes were downloaded from National Center for Biotechnology Information (NCBI). A data set of chloroplast orthologous protein sequences was created as given in Shanker et al (2009). Each set of orthologous protein sequences was aligned using MUSCLE (Edgar 2004), to get the multiple sequence alignment (MSA). These orthologous protein sequences were combined to generate concatenated data set. The MSA of corresponding CDS was generated with the help of PAL2NAL (Suyama et al. 2006), which uses protein MSA to guide the nucleotide alignment. CDS alignments were also assembled into concatenated data set, thus generating combined 54 genes alignment.

Table 1. Information of chloroplast genomes.

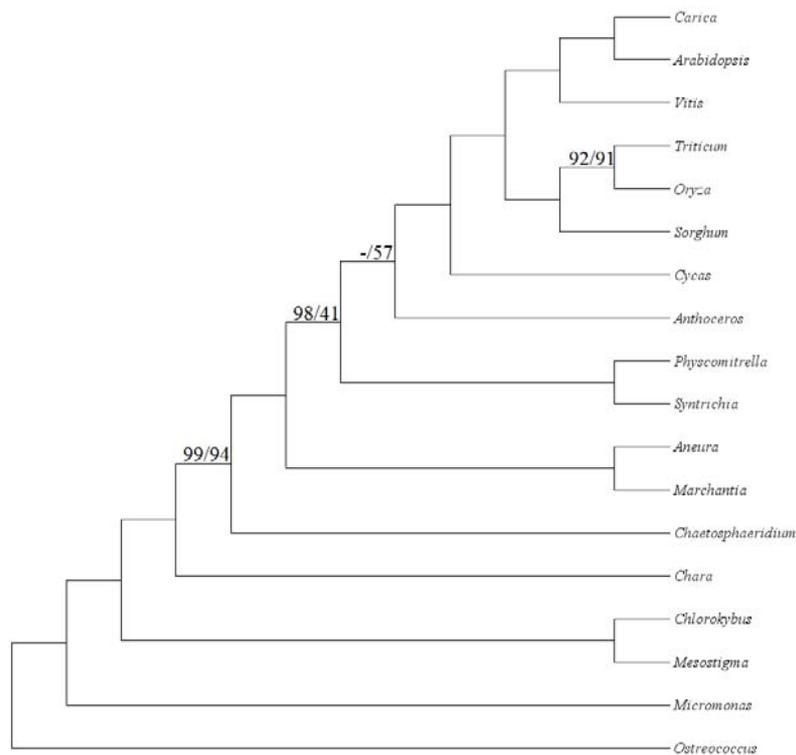
S. No.	Organisms	Chloroplast genome	
		Accession No.	No. of Protein Sequences
1.	<i>Chaetosphaeridium globosum</i>	NC_004115	98
2.	<i>Chara vulgaris</i>	NC_008097	105
3.	<i>Chlorokybus atmophyticus</i>	NC_008822	114
4.	<i>Mesostigma viride</i>	NC_002186	105
5.	<i>Micromonas</i> sp. RCC299	NC_012575	57
6.	<i>Ostreococcus tauri</i>	NC_008289	61
7.	<i>Anthoceros formosae</i>	NC_004543	90
8.	<i>Marchantia polymorpha</i>	NC_001319	89
9.	<i>Aneura mirabilis</i>	NC_010359	62
10.	<i>Syntrichia ruralis</i>	NC_012052	81
11.	<i>Physcomitrella patens</i> subsp. patens	NC_005087	85
12.	<i>Cycas taitungensis</i>	NC_009618	122
13.	<i>Arabidopsis thaliana</i>	NC_000932	85
14.	<i>Carica papaya</i>	NC_010323	84
15.	<i>Triticum aestivum</i>	NC_002762	83
16.	<i>Sorghum bicolor</i>	NC_008602	84
17.	<i>Oryza sativa</i> Indica Group	NC_008155	64
18.	<i>Vitis vinifera</i>	NC_007957	84

Mraic.pl (Nylander 2004), a PERL script, was used to determine the best fitting model of nucleotide substitution on data set. Considering Akai Information Criterion (AIC) nucleotide data sets showed GTRIG as the best model. To construct the phylogenetic trees maximum likelihood (ML) analysis was implemented using PHYML (Guindon and Gascuel 2003) with 500 bootstrap replicates. Moreover Bayesian Inference (BI) analysis was implemented using MrBayes (Huelsenbeck and Ronquist 2001). For BI analysis 1 run with four markov chains running for ten million generations were performed. The heated markov chain was sampled every 1000<sup>th</sup> generation. Convergence was detected using tracer and 25% samples were discarded as burn-in. CpRev (Adachi et al. 2000) model estimated for chloroplast encoded proteins were used to construct trees of chloroplast protein sequences data set. PHYML, with 500 bootstrap replicates, was used to construct phylogenetic trees from protein sequences. *Ostreococcus tauri*, member of chlorophyta, was used to root all the trees.

## Results and discussion

Earlier Nishiyama et al. (2004) showed that phylogenetic analysis using nucleotide sequences from chloroplast genomes did not resolve relationship with statistical confidence. Therefore in this study phylogenetic analysis was performed using nucleotide and corresponding protein sequences. The trees inferred showed identical tree topologies in which liverworts are the first to diverge with 100% bootstrap (BS) and 1 posterior probability (PP) support values therefore appeared as sister to all land plants (Fig. 1). *Anthoceros formosae*, a hornwort, showed sister relationship with tracheophytes with 100% BS/1 PP support values in nucleotide based trees while 57% BS support in protein sequences based tree.

Irrespective of taxon set, type of sequences and method of analysis used the paraphyly of bryophytes inferred in this analysis is in strong agreement with several earlier studies. In a study involving 350 diverse land plants and several red and green algae, absence of three mitochondrial group II introns in algae and liverworts was shown to support the bryophytes paraphyly hypothesis (Qiu et al. 1998). This view was also supported by analyses of morphological data sets (Kenrick and Crane 1997, Renzaglia et al. 2000). Moreover an analysis of three different data sets, a six gene data set from 193 green algae and land plants, a mitochondrial group II intron insertion site matrix and a chloroplast genome sequence matrix inferred paraphyly of bryophytes (Qiu et al. 2006).



**Figure 1.** Bryophytes phylogeny inferred from chloroplast nucleotide (PHYML and BI) and protein sequences (PHYML) data sets. All the nodes in nucleotide based tree inferred using Bayesian method have posterior probability 1. Bootstrap support values (<100%) for nucleotide and protein sequences based trees, respectively, are shown on the internal nodes.

Contrary to this the monophyly of bryophytes was inferred using spermatogenesis characters (Garbary et al. 1993, Renzaglia et al. 2000), analysis using amino acid sequences, translated from 51 chloroplast genes (Nishiyama et al. 2004). Moreover analysis based on 57 protein coding gene sequences common to 17 chloroplast genomes of land plants and a charophyte alga inferred monophyly of bryophytes (Goremykin and Hellwig 2005). Qiu (2008) suggested that the studies showing monophyly of bryophytes may suffer from limitation of taxon or character sampling. The present study successfully used available chloroplast genome sequences of plants and provide support to bryophytes paraphyly hypothesis. Addition of more bryophytes taxa will help to further strengthen these relationships.

## References

- Adachi, J., Waddell, P., Martin, W. et al. 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *J. Mol. Evol.* 50: 348-358.
- Chang, Y. and Graham, S.W. 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *Am. J. Bot.* 98: 839-849.
- Edgar, R.C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5: 113.
- Finet, C., Timme, R.E., Delwiche, C.F. et al. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology* 20: 2217-2222.
- Gao, L., Su, Y.J. and Wang, T. 2010. Plastid genome sequencing, comparative genomics and phylogenomics: current status and prospects. *J. Syst. Evol.* 48: 77-93.
- Garbary, D.J., Renzaglia, K.S. and Duckett, J.G. 1993. The phylogeny of land plants -a cladistic analysis based on male gametogenesis. *Plant Syst. Evol.* 188: 237-269.
- Goremykin, V.V. and Hellwig, F.H. 2005. Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. *Plant Syst. Evol.* 254: 93-103.
- Groth-Malonek, M., Pruchner, D., Grewe, F. et al. 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.
- Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52: 696-704.
- Huelsenbeck, J.P. and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.
- Kenrick, P. and Crane, P.R. 1997. The origin and early diversification of land plants: a cladistic study. Smithsonian Institution, Washington DC, pp. 441.
- Lewis, L.A., Mishler, B.D. and Vilgalys, R. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Mol. Phylogenet. Evol.* 7: 377-393.
- Ligrone, R., Duckett, J.G. and Renzaglia, K.S. 2012. Major transitions in the evolution of early land plants: a bryological perspective. *Ann. Bot.* 109: 851-871.
- Mishler, B.D., Lewis, L.A., Buchheim, M. A. et al. 1994. Phylogenetic relationships of the green algae and bryophytes. *Ann. Mo. Bot. Gard.* 81: 451-483.
- Nickrent, D.L., Parkinson, C.L., Palmer, J.D. et al. 2000. Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Mol. Biol. Evol.* 17: 1885-1895.
- Nishiyama, T., Wolf, P.G., Kugita, M. et al. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Mol. Biol. Evol.* 21: 1813-1819.
- Nylander, J.A.A. 2004. MrAIC.pl program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Qiu, Y. L. 2008. Phylogeny and evolution of charophytic algae and land plants. *J. Syst. Evol.* 46: 287-306.

- 
- Qiu, Y.L., Cho, Y., Cox, J. C. et al. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671-674.
- Qiu, Y.L., Li, L., Wang, B. et al. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. USA* 103: 15511-15516.
- Renzaglia, K.S., Duff, R.J.T., Nickrent, D.L. et al. 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.
- Shanker, A. and Sharma, V. 2012. Chloroplast genome analysis to detect transition link between bryophytes and pteridophytes: a bioinformatics approach. *Archive for Bryology* 121: 1-8.
- Shanker, A., Sharma, V. and Daniell, H. 2009. A novel index to identify unbiased conservation between proteomes. *International Journal of Integrative Biology* 7: 32-38.
- Shanker, A., Sharma, V. and Daniell, H. 2011. Phylogenomic evidence of bryophytes' monophyly using complete and incomplete data sets from chloroplast proteomes. *J. Plant Biochem. Biotechnol.* 20: 288-292.
- Suyama, M., Torrents, D. and Bork, P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res.* 34: W609-612.