Chloroplast Genome Analysis to Detect Transition Link between Bryophytes and Pteridophytes: A Bioinformatics Approach

Asheesh Shanker* and Vinay Sharma

Department of Bioscience and Biotechnology, Banasthali University, Rajasthan, India

*Corresponding author email: ashomics@gmail.com

Abstract: Bryophytes (mosses, liverworts and hornworts) are the most basal lineages among extant land plants. However, there is controversy regarding their monophyletic or paraphyletic origin. Moreover, the most basal group of bryophytes also remains questionable. Due to this it is a matter of debate which lineage of bryophytes acts as transition link between bryophytes and pteridophytes. In order to resolve these issues we conducted a bioinformatics analysis on available chloroplast genomes of 1 green alga, 5 bryophytes, 4 pteridophytes, and 6 gymnosperms. Our study is based on multiple genome analysis through whole proteome comparison. The higher number of orthologous identified in *Anthoceros* (a hornwort), in comparison to other bryophytes considered, against all pteridophytes strongly support hornworts as transition link between bryophytes and pteridophytes.

Introduction

Among land plants, bryophytes (mosses, liverworts and hornworts) appear to be the most basal lineages among extant land plants (Kenrick and Crane, 1997). However, the monophyly of major

groups of bryophytes as well as most basal group of bryophytes remains questionable (Qiu *et al,* 1998; Nishiyama *et al,* 2004; Qiu *et al,* 2006; Shanker *et al,* 2011). Most of these studies show fundamentally different tree topologies and are unable to resolve the dilemma of bryophyte's common ancestry and the lineage which primarily inhabited the land. Two theories have been proposed for earliest diverging bryophyte lineage, (i) liverworts basal topology (LBT; Mishler *et al,* 1994; Qiu *et al,* 1998, 2006) and (ii) hornworts basal topology (HBT; Renzaglia and Duckett, 1991; Garbary and Renzaglia, 1998; Nickrent *et al,* 2000; Nishiyama *et al,* 2004). Despite all these efforts the relationship of bryophytes continues to be unresolved therefore new approaches are required to get new insights on bryophytes relationship.

Chloroplasts are cytoplasmic organelles present in green plants which contain their own autonomously replicating DNA genome and encode a number of components for the process of photosynthesis (Palmer, 1985; Sugiura, 1989). Since the determination of first chloroplast genome sequences of *Marchantia polymorpha* (Ohyama *et al,* 1986) and tobacco, *Nicotiana tabacum* var. Bright Yellow 4 (Shinozaki *et al,* 1986) a large number of chloroplast genomes have been sequenced which belong to various algae, bryophytes, pteridophytes, gymnosperms and angiosperms (NCBI Eukaryotae Organelles List). The adequate number of available complete chloroplast genome sequences makes it feasible to compare them and to reliably infer the phylogenetic relationship and possible course of plant evolution.

Multiple genome analysis through whole proteome comparison provides a means to detect conservation between genomes. In doing so, sequence similarity search tools with specific search criterion were used to identify the number of homologous sequences between organisms, and based on this the degree of similarity between genomes was calculated. The similarity provides insight on conservedness among various organisms utilizing the treasure of evolutionary information present in complete genome sequences by identifying orthologous sequences. In the present study a data set of chloroplast proteomes from 16 plants was used to identify transition link among bryophytes and pteridophytes.

Materials and Methods

Chloroplast Genome Sequences

Various chloroplast genomes belonging to different lineages and supposedly representing the hierarchy of plant evolution were used in the study. The chloroplast genome sequences of all these organisms are available at National Center for Biotechnology Information (NCBI) organelles database. These organisms included 1 green alga, *Chara vulgaris,* 5 bryophytes, *Aneura mirabilis, Anthoceros formosae, Marchantia polymorpha, Physcomitrella patens, Syntrichia ruralis,* 4 pteridophytes, *Adiantum capillus-veneris, Psilotum nudum, Huperzia lucidula, Angiopteris evecta,* and 6 gymnosperms, *Cycas taitungensis, Pinus thunbergii, Ephedra equisetina, Welwitschia mirabilis, Gnetum parvifolium, Keteleeria davidiana.* The protein sequences of these chloroplast genomes were downloaded from NCBI. The accession number of chloroplast genomes along with number of proteins is shown in table 1.

Table 1. Genome information of organisms along with number of protein sequences.

S. No.	Organism	Abr*	Accession Number	Number of Proteins
Algae				
1.	*Chara vulgaris*	Cv	NC_008097	105
Bryophytes				
2.	*Aneura mirabilis*	Am	NC_010359	62
3.	*Anthoceros formosae*	Af	NC_004543	90
4.	*Marchantia polymorpha*	Mp	NC_001319	89
5.	*Physcomitrella patens*	Pp	NC_005087	85
6.	*Syntrichia ruralis*	Sr	NC_012052	81
Pteridophytes				
7.	*Adiantum capillus-veneris*	Ac	NC_004766	87
8.	*Psilotum nudum*	Pn	NC_003386	101
9.	*Huperzia lucidula*	Hl	NC_006861	87
10.	*Angiopteris evecta*	Ae	NC_008829	88
Gymnosperms				
11.	*Cycas taitungensis*	Ct	NC_009618	122
12.	*Pinus thunbergii*	Pt	NC_001631	155
13.	*Ephedra equisetina*	Ee	NC_011954	72
14.	*Welwitschia mirabilis*	Wm	NC_010654	70
15.	*Gnetum parvifolium*	Gp	NC_011942	66
16.	*Keteleeria davidiana*	Kd	NC_011930	75

*Abbreviations of name of organism.**

Orthologs Identification between Chloroplast Proteomes

The presence of homolog(s) defines the degree of conservation between genomes. Genes/proteins that are separated by speciation are called orthologous and they are often detected as the best reciprocal match. Therefore, the bidirectional best hit method (BBH; Overbeek *et al,* 1999) was used for the identification of orthologous proteins among all organisms. BBH method was implemented using PERL scripts that could automatically build database of all proteomes except the query proteome. Thus all the surveyed proteomes once become query proteome whereas at the same time other proteomes remain in the database. Organism specific comparisons of each protein sequence from query proteome against proteome database of every other organism were performed using protein-protein BLAST (Basic Local Alignment Search Tool; Altschul *et al,* 1997). BLAST results were parsed using stringent measure of both sequence coverage and percent sequence similarity (positivity).

The sequence positivity in BLAST result represents similar characters along with identical characters, while identity considers only exactly matched characters. Therefore, in a similarity search it is quite possible that protein sequences from different organisms have higher positivity percentage than identity percentage. This is simply due to conservative substitution which is not considered while calculating identity. Therefore a fixed match criterion of ≥50% positivity was used to parse all BLAST results. Parsed hits were used to identify putative reciprocal orthologous sequences. Due to variability in the size of genomes their comparison requires an appropriate methodology and parameters to detect an unbiased degree of conservation between them. Therefore, newly defined parameters, cumulative sum of homologs (CSH) and actual proportion of homologs (APH), developed by Shanker et al. (2009) were applied to orthologs identified and unbiased conservation was calculated between chloroplast genomes.

Results and Discussion

The results of identified orthologs along with their APH values among representatives of bryophytes and pteridophytes are presented in table 2. Considering specified criteria of sequence positivity the proteome comparison of organism A with organism B identify putative orthologs. The comparison of total m organisms leads to an m × m matrix. The cells of this matrix contain total number of putative orthologs identified between organism A and organism B. The higher APH values identified in *Anthoceros* (a hornwort; table 2), in comparison to other bryophytes considered, against all pteridophytes strongly support hornworts as the sister to vascular plants. These results are consistent with several recent findings (Qiu *et al,* 2006; Pena *et al,* 2008). Contrary to this a number of studies (Renzaglia and Duckett, 1991; Garbary and Renzaglia, 1998; Renzaglia *et al,* 2000; Nishiyama *et al,* 2004) show hornworts as basal most land plant. Overall, higher number of orthologs identified and corresponding higher APH values suggest hornworts as transition link between bryophytes and pteridophytes.

In our earlier study to detect unbiased conservation between genomes, *Anthoceros* was found to show higher similarity with *Psilotum,* a pteridophyte. Representatives of pteridophytes, *Psilotum* and *Adiantum* also showed higher similarity with *Anthoceros.* Moreover A comparison of bryophytes with pteridophytes, gymnosperms and angiosperms shows unequivocal higher similarity with *Anthoceros* than *Marchantia* (Shanker *et al,* 2009). This also supports hornworts as transition link between bryophytes and pteridophytes and is in agreement with a previous work

based on morphological characters, which showed hornworts similar to pteridophytes (Bierhorst, 1971; Qiu *et al,* 2006).

Our analysis based on sequence comparison of chloroplast protein sequences shows conservation between genomes. As more distant organisms show less sequence similarity in comparison to closely related organisms. We take advantage to this fact and identified hornworts as transition link between bryophytes and pteridophytes. A comprehensive study with a larger number of representatives is needed to further resolve this relationship.

Acknowledgements

References

ALTSCHUL, S. F., MADDEN, T. L., SCHÄFFER, A. A., ZHANG, J., ZHANG, Z., MILLER, W. AND LIPMAN, D. J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res., 25: 3389-3402.

BIERHORST, D. W. 1971. Morphology of vascular plants. New York, MacMillan, pp. 560.

GARBARY, D. J. AND RENZAGLIA, K. S. 1998. Bryophyte phylogeny and the evolution of land plants: Evidence from development and ultrastructure. In: Bryology for the twenty-first century (eds. J. W. BATES, N. W. ASHTON and J. G. DUCKETT), Maney Publishing and The British Bryological Society, Leeds, England, pp. 45-63.

KENRICK, P. AND CRANE, P. R. 1997. The origin and early evolution of plants on land. Nature, 389: 33-39.

MISHLER, B. D., LEWIS, L. A., BUCHHEIM, M. A., RENZAGLIA, K. S., GARBARY, D. J., DELWICHE, C. F., ZECHMAN, F. W., KANTZ, T. S. AND CHAPMAN, R. L. 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. AND DUFF, R. J. 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., SINCLAIR, R. B., SUGITA, M., SUGIURA, C., WAKASUGI, T., YAMADA, K., YOSHINAGA, K., YAMAGUCHI, K., UEDA, K. AND HASEBE, M. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. Mol. Biol. Evol., 21: 1813-1819.

OHYAMA, K., FUKUZAWA, H., KOHCHI, T., SHIRAI, H., SANO, T., SANO, S., UMESONO, K., SHIKI, Y., TAKEUCHI, M., CHANG, Z., AOTA, S., INOKUCHI, H. AND OZEKI, H. 1986. Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature, 322: 572-574.

OVERBEEK, R., FONSTEIN, M., D'SOUZA, M., PUSCH, G. D. AND MALTSEV, N. 1999. The use of gene clusters to infer functional coupling. Proc. Natl. Acad. Sci. USA, 96: 2896-2901.

PALMER, J. D. 1985. Comparative organization of chloroplast genomes. Ann. Rev. Genet., 19: 325-354.

PENA, M. J., DARVILL, A. G., EBERHARD, S., YORK, W. S. AND O'NEILL, M. A. 2008. Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants. Glycobiology, 18: 891-904.

QIU, Y. L., CHO, Y., COX, J. C. AND PALMER, J. D. 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., CHEN, Z., KNOOP, V., GROTH-MALONEK, M., DOMBROVSKA, O., LEE, J., KENT, L., REST, J., ESTABROOK, G. F., HENDRY, T. A., TAYLOR, D. W., TESTA, C. M., AMBROS, M., CRANDALL-STOTLER, B., DUFF, R. J., STECH, M., FREY, W., QUANDT, D. AND DAVIS, C. C. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc. Natl. Acad. Sci. USA, 103: 15511-15516.

RENZAGLIA, K. S. AND DUCKETT, J. G. 1991. Towards an understanding of the differences between the blepharoplast of mosses and liverworts, comparisons with hornworts, biflagellated lycopods and charophytes: A numerical analysis. New Phytol., 117: 187-208.

RENZAGLIA, K. S., DUFF, R. J. T., NICKRENT, D. L. AND GARBARY, D. J. 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., 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.

SHANKER, A., SHARMA, V. AND DANIELL, H. 2009. A novel index to identify unbiased conservation between proteomes. IJIB, 7: 32-38.

SHINOZAKI, K., OHME, M., TANAKA, M., WAKASUGI, T., HAYASHIDA, N., MATSUBAYASHI, T., ZAITA, N., CHUNWONGSE, J., OBOKATA, J., YAMAGUCHI-SHINOZAKI, K., OHTO, C., TORAZAWA, K., MENG, B. Y., SUGITA, M., DENO, H., KAMOGASHIRA, T., YAMADA, K., KUSUDA, J., TAKAIWA, F., KATO, A., TOHDOH, N., SHIMADA, H. AND SUGIURA M. 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO, 5: 2043-2049.

SUGIURA, M. 1989. The chloroplast chromosomes in land plants. Annu. Rev. Cell Biol., 5: 51-70.

online January 17, 2012

Table 2. Putative reciprocal orthologs identified among organisms along with APH values in parenthesis (see table 1 for organisms' abbreviations).

Shanker & Sharma

Organism	Cv	Af	Am	Mp	Pp	Sr	Ac	Pn	Hl	Ae	Ct	Pt	Ee	Wm	Gp	Kd
Cv	-	81 (7.53)	56 (5.20)	77 (7.16)	79 (7.34)	75 (6.97)	79 (7.34)	77 (7.16)	80 (7.43)	83 (7.71)	79 (7.34)	64 (5.95)	62 (5.76)	58 (5.39)	61 (5.67)	65 (6.04)
Af	81 (7.03)	-	59 (5.12)	86 (7.47)	83 (7.20)	80 (6.94)	86 (7.47)	88 (7.64)	86 (7.47)	90 (7.81)	88 (7.64)	67 (5.82)	66 (5.73)	62 (5.38)	63 (5.47)	67 (5.82)
Am	56 (6.68)	59 (7.04)	-	58 (6.92)	59 (7.04)	57 (6.80)	58 (6.92)	58 (6.92)	58 (6.92)	62 (7.40)	56 (6.68)	53 (6.32)	54 (6.44)	48 (5.73)	48 (5.73)	54 (6.44)
Mp	77 (7.06)	86 (7.88)	58 (5.32)	-	79 (7.24)	77 (7.06)	78 (7.15)	83 (7.61)	80 (7.33)	84 (7.70)	80 (7.33)	65 (5.96)	61 (5.59)	58 (5.32)	61 (5.59)	64 (5.87)
Pp	79 (7.09)	83 (7.45)	59 (5.30)	79 (7.09)	-	81 (7.27)	79 (7.09)	84 (7.54)	83 (7.45)	86 (7.72)	83 (7.45)	68 (6.10)	64 (5.75)	58 (5.21)	60 (5.39)	68 (6.10)
Sr	75 (6.98)	80 (7.44)	57 (5.30)	77 (7.16)	81 (7.53)	-	77 (7.16)	81 (7.53)	80 (7.44)	83 (7.72)	78 (7.26)	66 (6.14)	60 (5.58)	56 (5.21)	57 (5.30)	67 (6.23)
Ac	79 (7.02)	86 (7.64)	58 (5.15)	78 (6.93)	79 (7.02)	77 (6.84)	-	86 (7.64)	84 (7.46)	88 (7.82)	84 (7.46)	68 (6.04)	65 (5.77)	62 (5.51)	63 (5.60)	69 (6.13)
Pn	77 (6.71)	88 (7.67)	58 (5.05)	83 (7.23)	84 (7.32)	81 (7.06)	86 (7.49)	-	87 (7.58)	93 (8.10)	87 (7.58)	67 (5.84)	61 (5.31)	64 (5.57)	64 (5.57)	68 (5.92)
Hl	80 (7.02)	86 (7.55)	58 (5.09)	80 (7.02)	83 (7.29)	80 (7.02)	84 (7.37)	87 (7.64)	-	89 (7.81)	85 (7.46)	69 (6.06)	65 (5.71)	61 (5.36)	63 (5.53)	69 (6.06)
Ae	83 (6.98)	90 (7.57)	62 (5.21)	84 (7.06)	86 (7.23)	83 (6.98)	88 (7.40)	93 (7.82)	89 (7.49)	-	92 (7.74)	70 (5.89)	69 (5.80)	64 (5.38)	66 (5.55)	70 (5.89)
Ct	79 (6.81)	88 (7.59)	56 (4.83)	80 (6.90)	83 (7.16)	78 (6.72)	84 (7.24)	87 (7.50)	85 (7.33)	92 (7.93)	-	74 (6.38)	69 (5.95)	65 (5.60)	66 (5.69)	74 (6.38)
Pt	64 (6.4)	67 (6.7)	53 (5.3)	65 (6.5)	68 (6.8)	66 (6.6)	68 (6.8)	67 (6.7)	69 (6.9)	70 (7)	74 (7.4)	-	68 (6.8)	61 (6.1)	62 (6.2)	78 (7.8)
Ee	62 (6.37)	66 (6.78)	54 (5.55)	61 (6.27)	64 (6.58)	60 (6.17)	65 (6.68)	61 (6.27)	65 (6.68)	69 (7.09)	69 (7.09)	68 (6.99)	-	72 (7.40)	70 (7.19)	67 (6.89)
Wm	58 (6.29)	62 (6.72)	48 (5.21)	58 (6.29)	58 (6.29)	56 (6.07)	62 (6.72)	64 (6.94)	61 (6.62)	64 (6.94)	65 (7.05)	61 (6.62)	72 (7.81)	-	71 (7.70)	62 (6.72)
Gp	61 (6.50)	63 (6.72)	48 (5.12)	61 (6.50)	60 (6.40)	57 (6.08)	63 (6.72)	64 (6.82)	63 (6.72)	66 (7.04)	66 (7.04)	62 (6.61)	70 (7.46)	71 (7.57)	-	63 (6.72)
Kd	65 (6.47)	67 (6.67)	54 (5.37)	64 (6.37)	68 (6.77)	67 (6.67)	69 (6.87)	68 (6.77)	69 (6.87)	70 (6.97)	74 (7.36)	78 (7.76)	67 (6.67)	62 (6.17)	63 (6.27)	-