Pl. Syst. Evol. 254: 93–103 (2005)
DOI 10.1007/s00606-005-0337-1

Plant Systematics and Evolution
Printed in Austria

Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages

V. V. Goremykin and **F. H. Hellwig**

Institut für Spezielle Botanik, Universität Jena, Jena, Germany

Received September 8, 2004; accepted May 2, 2005
Published online: 27 June 2005

Abstract. The problem of relationships among the major basal living groups of land plants is long standing, yet the uncertainty as to the phylogenetic affinity of these lines persists in the literature. Molecular and modern cladistic studies of the phylogenetic relationships of the above groups resulted in a large number of conflicting topologies. However, with the exception of the cladistic analyses of spermatogenesis, suggesting monophyly of extant bryophytes, these studies agree the paraphyletic bryophyte grade is basal within the embryophyte tree. Here we would like to present analyses on the basis of the concatenated datasets of nucleotide and amino-acid sequences of 57 protein-coding genes common to 17 chloroplast genomes of land plants and a charophyte alga *Chaetosphaeridium globosum*. Character-wise, these are the largest datasets currently available to address the problem of basal relationships within embryophytes. Main lineages of bryophytes, i.e liverworts, hornworts and mosses are represented in our alignments with a single taxon, whereas 14 taxa represent the tracheophytes. With our data, phylogeny with liverwort basal appears to be and artifact related to high and unequal A + T contents among the sequences analysed. Reducing this compositional bias and applying methods developed to counter it, we recovered an alternative, strongly supported topology wherein both bryophytes and tracheophytes are monophyletic. Within bryophytes, hornworts are basal and liverworts are sister to mosses.

Key words: Bryophytes, chloroplast genomes, molecular evolution.

Introduction

Origin of embryophytes changed the life on Earth in most dramatic way, creating new environments and paving the way for the rise and evolution of terrestrial faunas (see Kenrick and Crane 1997a). Latest common ancestor of all land plants existed roughly 475 million years ago (Wellman et al. 2003) and was, of all living forms, most closely related to modern Charales (Karol et. al. 2001). First green plants colonized and formed terrestrial environments somewhere in the mid-Ordovician age (Strother et al. 1996). Fossil record documents wide dispersal of land plant spores in the Llarvirnian sediments (~475 million years ago) (Wellman et al. 2003). Much remains unknown about the paths the early diversification of land plants took. Abundance of the tetrads among those spores prompted a suggestion, that the first terrestrial floras were dominated by the liverwort-like plants (Gray 1993). However, absence of the associated macrofossils precludes any strong conclusions

on that matter to be drawn. Dispersal of the spores typical for land plants predate the appearance of the first macrofossils of land plants by ~50 millon years (Kenrick and Crane 1997a). These late Silurian macrofossil assemblages reveal primitive land plants to be by the most part small and morphologically simple forms with isoform life cycles, i.e similar gametophyte and sporophyte generations (Schweitzer 1990). These plants (protracheophytes) have been assigned a place within the vascular plant line of evolution (Kenrick and Crane 1997b). Their morphologically relatively complex gametophyte generation stands, however, in sharp contrast to strongly reduced gametophytes typical of the extant vasular plants and is reminiscent of free-living gametophytes of the extant bryophytes. Presence of columella in some of these early Silurian fossils (*Horneophyton*) only underlines this resemblance. However, fossil record of bryophytes can be traced back unequivocally only as far as late Devonian (~370 million years ago) (Kenrick and Crane 1997b). Such large gaps in fossil record hinder investigation of the life cycle evolution of early land plants. Another obstacle to this research is that the relationships among basal embryophyte clades (polysporangiophytes, liverworts, hornworts and mosses) remain unclear (see Duff and Nickrent 1999 and Nickrent et al. 2000 for the review).

Analyses of morphological characters from the above four lineages yielded topologies, pointing to absolutely different scenarious of early land plant evolution. Mishler and Curchill in 1984 first applied the cladistic approach in the study emphasizing bryophytes. They suggested that the tracheophyte clade is sister to the mosses, which implies paraphyly of bryophytes. Another important implication of this topology is that the gametophyte-dominated life cycle of the first land plants was succeeded by the sporophyte-dominated life cycle in the tracheophyte lineage. The tracheophyte affinity of the mosses also found support in the later investigation by Kenrick and Crane (1997b). Study of spermatogenesis by Garbary et al. (1993) pointed to the monophyly of bryophytes, basal position of hornworts within bryophytes and sister group relationships between liverworts and mosses. This topology was later advocated by Maden with co-workers in 1997 in their investigation of sperm ultrastructure. Within this evolutionary scenario gameto- and sporophyte-dominated life cycles must have evolved independently from one another. In 1998 Garbary and Renzaglia, however, added morphological data unrelated to the male gametogenesis to their matrix and published a topology which supported paraphyly of bryophytes. On their tree vascular plants are sister not to the mosses, but to the clade uniting mosses with liverworts. Similar result was received by Mishler et al. in 1994. In a later work by Renzaglia and colleagues (2000), however, the authors suggested that the earlier perception of leptoids and hydroids of bryophytes as being homologous to the sieve elements and tracheids, is wrong and, again, advocated bryophyte monophyly. Qiu et al. (1998) interpreted the absence of the three mitochondrial introns in liverworts and green algae in favour of the liverwort-basal topology (i.e as synapomorphous indels). However, their results indicate that some of these introns were also missing in the mtDNA from the non-liverwort land plants, which makes the synapomorphous status of these characters (or of their absence) subject to interpretation.

It is interesting to note that in the cladistic analyses by Kenrick in Crane (1997b) on the trees just two and three steps longer than the most parsimonous one (which supported bryophyte paraphyly) bryophytes were monophyletic. On the trees two steps longer Silurian protracheophytes assumed a position at the base of the tracheophyte cluster. On the trees three steps longer they appeared at the very base of the land plant tree. Such tree instability shows that the morphological analyses of the vegetative forms may not be reliable for elucidating early land plant evolution due to scarcity of characters available for cladistic analyses. Cladistic investigation of the basal

relationships among major lines of land plants is objectively complicated by difficulty of establishing homologous characters between bryophyte and tracheophyte gametophytes and sporophytes, and by scarcity of these characters (Renzaglia et al. 2000, Kenrick 2000).

So far, application of molecular techniques was not able to clarify evolution of the basal embryophyte lineages. Topologies suggested by molecular studies are highly contradicting. Depending on the marker, either mosses (chloroplast 16S and 23S rRNAs - Mishler et al. 1992), or hornworts (*rbc*L - Lewis, et al. 1997, *rbc*L and three small-subunit rDNA genes - Nickrent et al. 2000), or liverworts (*cox*3 (mit.) - Malek et al. 1996), or liverworts and mosses taken together (18S rDNA (nuc.) - Hedderson, et al. 1996, *psb*A (chloroplast) - Crowe et al. 1997, 19S rDNA (mit.) - Duff and Nickrent 1999, five plastid genes - Nishiyama and Kato 1999) were suggested to share sister group relationship to the tracheophytes. In the recent study by Nishiyama et al. (2004), in which a combined dataset of amino acid sequences of 51 chloroplast genes was used, the authors found support for the sister group relationship between the tracheophytes and all bryophytes. Analyses of nucleotide sequences of these genes were not conclusive in the above study.

Notwithstanding the differences, these trees have something in common: all intermediate branches linking four major land plant lineages together are much shorter than the branches linking the common ancestor of each respective group to its distant extant posterity. This indicates that the particular intractability of reconstruction of the phylogenetic relationships of the four basal land plant lineages can be attributed to relatively short times of common history these lineages shared as compared to huge timespans in which they developed independently afterwards. Such old radiation would lead to the sparsity of characters bearing witness to those remote evolutionary events, which would explain contradictory topologies received with different markers as resulting from character sampling errors. It seems to us that the solution of this problem may lie in the accumulation of large molecular datasets, as it was done by Nishiyama with colleagues (2004). Now, as complete chloroplast genomes of all three major extant bryophyte lines have been published, it appears to be of interest to try to reconstruct the basal land plant topology using wealth of homologous data they provide, with a focus on the nucleotide sequences.

Materials and methods

Sequences of 57 protein-coding genes (see Table 1) shared by completely sequenced and annotated genomes of *Marchantia* (Ohyama et al. 1986), *Nicotiana* (Shinozaki et al. 1986), *Oryza* (Hiratsuka et al. 1989), *Pinus* (Wakasugi et al. 1994), *Zea* (Maier et al. 1995), *Arabidopsis* (Sato et al. 1999), *Oenothera* (Hupfer et al. 2000), *Lotus* (Kato et al. 2000), *Spinacia* (Schmitz-Linneweber et al. 2001), *Triticum* (Ogihara et al. 2002), *Psilotum* (Wakasugi et al. direct submission to Genbank), *Chaetosphaeridium* (Turmel et al. 2002), *Adiantum* (Wolf et al. 2003), *Anthoceros* (Kugita et al. 2003), *Physcomitrella* (Sugiura et al. 2003), *Calycanthus* (Goremykin et al. 2003a), *Amborella* (Goremykin et al. 2003b), *Nymphaea* (Goremykin et al. 2004) were extracted with our GBSEARCH script and then were converted into fasta format. They were sorted into 57 files each containing sequences of an individual gene and written in one directory. Based on those files we automatically produced alignments of all codon positions, of the first and the second codon positions, of the third codon positions and of the translated amino acid sequences for each individual gene with the help of our ALNMAKER script. These individual alignments of 57 genes of 18 OTUs (operational taxonomic units) were manually concatenated and edited to produce i) a 40578 positions long alignment containing original gene sequences, ii) a 13712 positions long alignment containing their third codon positions, iii) a 26436 positions long alignment containing their first and the second codon positions and iiii) a 13556 amino acid positions long alignment of their translated sequences, which were subsequently used for phylogeny reconstruction. We used PAUP* (Swofford 2002) and Phylip (Felsenstein 1989) packages for

Table 1. Genes used in our analyses, with indication of numbers of different position classes in their alignments (third codon position excluded)

Gene name	Alignment length	constant	number of sites:	
			non-informative	informative
*atp*A	1030	723	117	190
*atp*B	1006	744	128	134
*atp*E	286	116	56	114
*atp*F	390	152	93	145
*atp*H	164	124	24	16
*atp*I	504	327	73	104
*cem*A	1041	318	398	325
*clp*P	502	246	117	139
*lhb*A	126	80	18	28
*pet*A	650	385	132	133
*pet*B	470	384	53	33
*pet*D	362	256	72	34
*pet*G	78	52	14	12
*pet*L	170	98	47	25
*pet*N	62	37	16	9
*psa*A	1510	1214	152	144
*psa*B	1472	1172	160	140
*psa*C	164	132	19	13
*psa*I	106	67	13	26
*psa*J	90	53	17	20
*psb*A	708	646	30	32
*psb*B	1026	778	132	116
*psb*C	977	818	77	82
*psb*D	708	611	48	49
*psb*E	168	133	17	18
*psb*F	80	59	9	12
*psb*H	164	97	22	45
*psb*I	138	110	22	6
*psb*J	86	55	17	14
*psb*K	126	63	24	39
*psb*L	78	56	14	8
*psb*M	76	42	16	18
*psb*N	88	63	15	10
*psb*T	88	64	5	19
*rbc*L	962	771	82	109
*rpl*14	250	138	55	57
*rpl*16	288	178	43	67
*rpl*2	558	272	117	169
*rpl*20	270	95	57	118
*rpl*32	156	62	35	59
*rpl*33	146	59	33	54
*rpl*36	76	46	16	14
*rpo*B	2233	1019	482	732
*rpo*C1	1474	709	270	495
*rpo*C2	3405	1233	781	1391
*rps*11	294	150	53	91
*rps*12	274	199	38	37
*rps*14	208	89	53	66
*rps*18	344	183	82	79
*rps*19	190	99	29	62
*rps*2	482	204	107	171
*rps*3	516	199	126	191
*rps*4	431	183	110	138
*rps*7	314	158	68	88
*rps*8	278	99	66	113
*ycf*3	368	245	70	53
*ycf*4	508	272	102	134

maximum parsimony analyses andTree-Puzzle program (Strimmer and von Haeseler 1996) for maximum likelihood analyses. Analyses utilising LogDet distance method were performed with PAUP* package and LDDist (Thollesson 2004) program. 5% chi-square tests of compositional homogeneity were performed with the Tree-Puzzle program.

Results

Analysis of the nucleotide alignments. First, we analysed the 42000 positions long alignment of 18 OTUs (operational taxonomic units), containing all codon positions. Maximum parsimony heuristic searches as implemented in Paup* and Phylip packages yielded each a single tree. The topologies of these trees were congruent, wherein the branch separating *Chaetosphaeridium* with *Marchantia* from the rest of the species, the one separating the last two species and *Physcomitrella* from the rest of the species and the one separating algae and all three bryophytes under analysis from the rest of the land plants were recovered with 100% bootstrap proportion support (BP).

The affinity of *Marchantia* to *Chaetosphaeridium* in the above alignment in which all codon positions are included is supported by 320 positions in which these two species share a base to the exclusion of other species. These

320 positions appear to be of crucial importance for the stability of the topology with liverwort basal, since their deletion changes the MP inference of the bryophyte relationships to *Chaetosphaeridium* (*Physcomitrella* (*Marchantia* (*Anthoceros* (vascular plants))), wherein the branch bearing all plants under analysis with the exception of *Physcomitrella* and *Chaetosphaeridium* is supported by 100% BP. Of 320 positions supporting affinity of liverwort with charophyte alga only 31 (9,7%) contain G or C in the sequences of these species. Removal of 289 positions in which liverwort and charophyte sequences share A or T to the exclusion of other species also leads to the above change in tree topology, supported by 100/100 BP. Given compositional bias of these positions and their strong influence on the phylogeny inference, a possibility that the topology is affected by a systematic error cannot be rejected.

Having found that, we wished to check compositional homogeneity in the alignment. In the 5% chi-square test performed with Tree-Puzzle (Strimmer and von Haeseler 1996), only two sequences (*Arabidopsis* and *Lotus*) passed the set threshold. This urged us to examine the compositional biases existing in our data in detail. Since we previously noticed that the third codon positions on *Nymphaea alba* cpDNA have significantly lower G + C content than the first and second codon positions (Goremykin et al. 2004), we wished to check if this is the general case. The results of our survey of the G + C content of all codon positions for the concatenated sequences of the 57 genes from all species under analysis are presented in Table 2. We determined that *Marchantia*, *Physcomitrella* and *Chaetosphaeridium* have the lowest overall G + C-content among the species analysed. This bias can be observed in all codon positions. However, it is the third codon positions which i) exhibit the strongest absolute bias and ii) show greatest irregularities in that bias among the species analysed. On average, the third codon positions exhibit ~10% less G + C content versus the second codon positions and ~20% less G + C content than the first codon positions in angiosperms, whereas in the sequences of *Marchantia*, *Physcomitrella* and *Chaetosphaeridium* that difference is even more pronounced, reaching ~15% (second vs. third position) and ~30% (first vs. third position). The G + C content of the *Anthoceros* genes is higher and comes close to the values found for the tracheophytes. Differences in G + C content, not exceeding 5,5% in the first and the second positions in the species under analysis reach as high as 24.8% in the third codon position.

Analysed alone, the third codon positions provide strong support for the topology with liverwort basal. The parts of phylogenetic trees depicting relationships of the bryophytes inferred with all MP and distance methods implemented in PAUP* and two ML models implemented in the Tree-Puzzle program (Tamura-Nei and Hasegawa, Kishino and Yano, both applied with default settings) on the basis of the above alignment (32 methods in total, including LogDet) were congruent and have had the following topology: *Chaetosphaeridium* (*Marchantia* (*Physcomitrella* (*Anthoceros* (vascular plants)))). All above branches received the highest support in these analyses. In the alignment of the third codon positions only two species (*Arabidopsis* and *Lotus*) passed the 5% chi-square test implemented in the Tree-Puzzle program. The distances from *Chaetosphaeridium* to *Marchantia*, *Chaetosphaeridium* to *Physcomitrella* and *Chaetosphaeridium* to *Anthoceros* calculated with the help of the above program using Tamura-Nei model of substitutions on the basis of the alignment of the third codon positions are, correspondingly, 0.95, 0.97 and 1.14 substitutions per site.

Removal of the third codon positions from the total alignment significantly decreases distances from *Chaetosphaeridium* to bryophytes (in the above order, 0.17, 0.17, 0.20 subst./pos.). Yet it does not help to completely overcome compositional bias in the data, since 10 species, including *Chaetosphaeridium*, *Marchantia* and *Physcomitrella* (but not *Anthoceros*) still do not pass the 5%

chi-square test. MP analysis with the first and the second codon position alignment recovers the above mentioned tree with liverwort basal. Among 77 positions in which *Marchantia* and the charophyte alga share a character to the exclusion of other OTUs in the above alignment, only 16 positions have G or C in the sequences of these two species. Interesting is, that when MP search is performed only with positions from the above partition, which are not affected by a G+C bias for this algorithm (those are positions in which all OTUs share either G and C characters or A and T characters, including positions with one, or two exceptions if they are different), another topology of the basal relationships among the embryophytes was recovered, albeit with low BP support (PAUP*, bootstrap majority-rule consensus tree). On that tree, the most basal split in land plants separates tracheophytes from bryophytes. *Anthoceros* appears basal among bryophytes and is a sister to the branch bearing *Marchantia* and *Physcomitrella*. The trees built using LogDet method as implemented in PAUP* and LDDist (Thollesson et al. 2004) programs from the alignment with the removed third codon positions, shared the same topology for bryophytes (Fig. 1). Monophyly of the bryophytes and tracheophytes and the branch bearing *Marchantia* with *Physcomitrella* received 100% BP support. Excluding invariant positions with capture-recapture method of Steel et al. (2000) and applying rate heterogeneity as implemented in LDDist results in somewhat lower support for the above branches (bryophyte monophyly - 94%, tracheophyte monophyly - 92%, sister group relationship between *Physcomitrella* and *Marchantia* - 99%).

Analysis of the translated sequences. Another way to alleviate the problems due to compositional bias in the sequences is to translate them into amino-acids. It has been proposed that, since protein sequences are under greater pressure to maintain their structure, their composition can be expected to be more homogeneous (Loomis and Smith 1990; Lockhart et al. 1992). Indeed, translation increases number of sequences which pass 5% chi-square test by 9 (11 species in total). This time all spermatophytes and *Psilotum* pass this test. Still, the problem persists for bryophytes. Fortunately, a program, which was specifically developed to correct for compositional biases in protein data has been recently published (Thollesson 2004). This program (LDDist) utilised the paralinear method suggested by Lake for protein sequences in 1994. The tree built from 13556 positions long amino acid alignment of concateneted sequences of 57 genes using the above program contains the basal tracheophyte/bryophyte split that we detected with nucleotide 1.+2. pos. data. Within the bryophytes, *Anthoceros* splits off first and *Marchantia* is a sister to *Physcomitrella*. The bryophyte topology is supported with maximum BP values and the monophyly of vascular plants receives 96/100 BP support when the program is run with default parameters. Applying rate heterogeneity and excluding invariant positions with the method of Steel et al (2000), we received 99% BP for the branch bearing all bryophytes, 95% BP for the branch bearing *Marchantia* and *Physcomitrella*, and 86% for the monophyly of tracheophytes.

Discussion

The third codon positions of the chloroplast protein-coding genes under analysis share high and, what is potentially more dangerous, more irregular compositional biases in comparison with the first and the second codon positions (see Table 1). With G+C- content reaching as low as ~14% in the third codon positions of chloroplast genes of *Marchantia* and *Physcomitrella*, these biases can pose a particularly serious problem for phylogeny reconstruction in bryophytes. It has been shown, that log determinant/paralinear analysis (Steel 1994, Lockhart et al. 1994, Lake 1994), the method of phylogeny reconstruction which had been developed to overcome

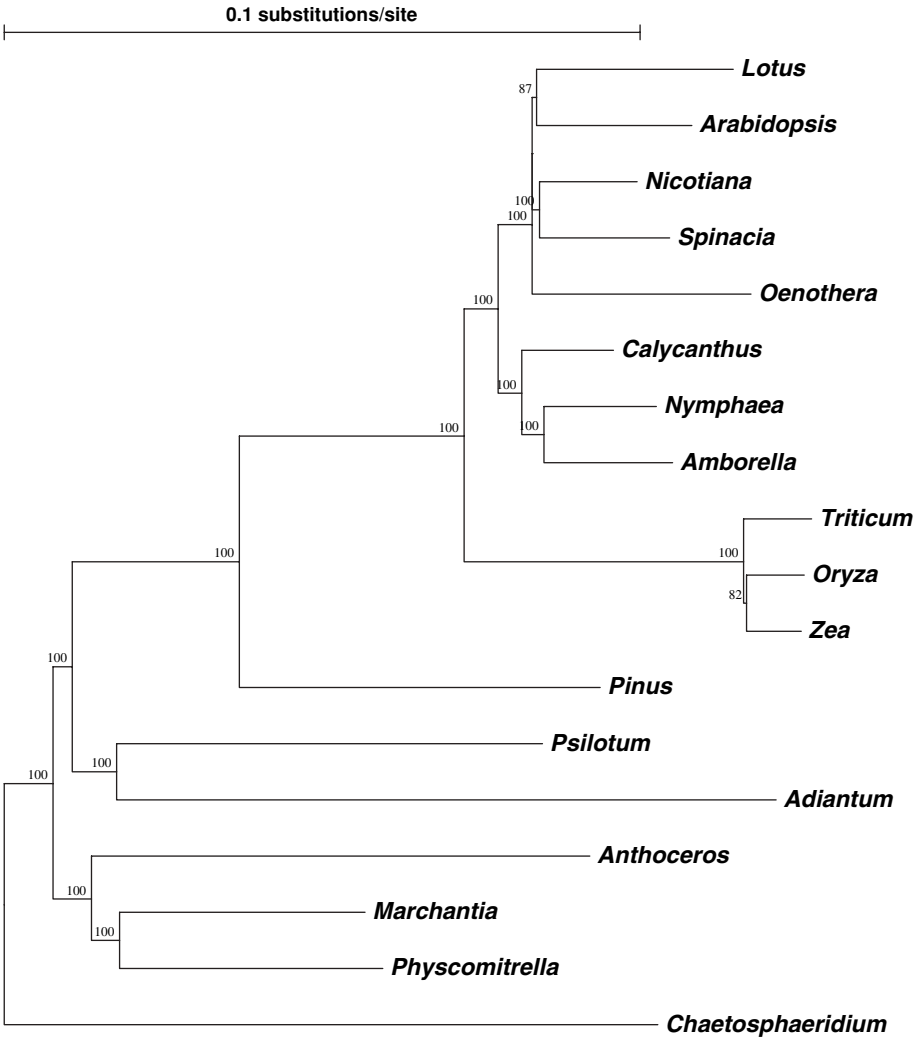

Fig. 1. Neighbor-joining tree built from LogDet/paralinear distances on the basis of the 26436 pos. long alignment of the first and the second codon positions sampled from 57 proteins common to completely sequenced chloroplast genomes and a charophyte alga *Chaetosphaeridium globosum*. The numbers to the right of the tree nodes indicate BP support

compositional biases, can fail to elucidate the correct topology when bias in the data is especially severe (Foster and Hickey 1999). Previously we have reported that the high rates of substitutions in the synonymous codon positions in chloroplast genes can pose problems for phylogeny reconstruction (Goremykin et al. 1996 and references therein). These high rates lead to distances close to subst. per site in pairwise comparisons between *Chaetosphaeridium* and bryophytes when only the third codon positions of 57 genes land plants and the algae plastomes share, are analysed.

Taking above considerations into account, one can conclude, that the result obtained with the alignment of the third codon positions, namely, the topology *Chaetosphaeridium* (*Marchantia* (*Physcomitrella* (*Anthoceros* (vascular plants)))) may be not reliable and that the use of the third position data for elucidating land plant phylogeny should be avoided.

When the third codon positions are removed from our data, support of the above

topology with MP decreases, but not disappears. As one can see from Table 2 the compositional bias in the first and the second codon positions still can favour the branch separating *Marchantia*, *Physcomitrella* and *Chaetosphaeridium* from the rest of the species. Among 78 positions supporting liverwort affinity to the charophyte alga, only 16 have G or C bases in the sequences of these species. Removing part of phylogenetic signal affected by bias in the alignment of the first and the second codon positions results in loss of support for the topology with liverwort basal. The tree assumes another topology, with the first split among the land plants separating bryophytes from tracheophytes. Within bryophytes, *Anthoceros* is basal and *Marchantia* is a sister to *Physcomitrella*. BP support for this topology is low with MP, which is probably attributable to the massive loss of informative positions, resulting from removal of the part of the alignment affected by the bias. However, when we applied the LogDet/paralinear method to the original first and second pos. data retaining all phylogenetic signal, this topology received strong support.

Translation to amino-acids has long been suggested to decrease the compositional bias. With our data, while not removing compositional bias completely, it alleviates it for many species. The tree topology favoured by the protein LogDet method as implemented in LDDist program (Thollesson 2004) built on the basis of the aminoacid alignment is the same as the one we recovered with nucleotide LogDet on the basis of the first and the second codon positions. Relationships among bryophytes are strongly supported.

Given strong compositional biases in the data, we believe the topology received with protein and nucleotide LogDet methods is more likely to reflect the underlying true tree. These analyses confirm the monophyly of bryophytes suggested in several morphological studies (Garbary et al. 1993, Maden et al. 1997, Renzaglia et al. 2000) and by the analyses of amino acid sequences of chloroplast genes (Nishiyama et al. 2004). We were able to demonstrate that the nucleotide sequences of

Table 2. G + C content across the codon positions of 57 chloroplast protein-coding genes for species under analysis

codon positions:	1 + 2	1st	2nd	3rd
Oryza	45.1	49.7	40.6	31.2
Triticum	45.0	49.4	40.6	30.5
Zea	45.0	49.5	40.5	31.0
Arabidopsis	44.9	49.2	40.6	28.1
Spinacia	45.3	49.8	40.8	28.6
Oenothera	45.8	50.2	41.4	32.6
Lotus	44.5	48.6	40.4	28.1
Nicotiana	45.6	50.2	41.0	29.6
Calycanthus	46.0	50.4	41.7	31.8
Amborella	45.9	50.3	41.5	31.1
Nymphaea	46.0	50.5	41.4	32.2
Pinus	45.6	50.2	41.1	29.9
Adiantum	46.7	50.9	42.6	37.5
Psilotum	44.1	48.5	39.6	26.1
Anthoceros	43.9	47.4	40.4	22.8
Physcomitrella	42.4	46.6	38.2	14.2
Marchantia	41.5	45.4	37.7	12.7
Chaetosphaeridium	41.6	45.0	38.3	15.4

the genes encoded by cpDNA also support the sister group relationship between the bryophytes and the tracheophytes and that the alternative topology with liverworts most basal among the land plants is a result of a compositional bias characteristic of these genes.

An importance of this result for the ongoing discussion on the early embryophyte diversification lies in its implication that the vascular plants are not derived bryophytes. This is in sharp contradiction to the currently widely accepted view, as for example; expressed in (Kenrick 2000): ''The weight of evidence... implies that vascular plants *are* bryophytes with a highly modified life history''. The suggestion of ''highly modified life history'' stems from profound differences in the life cycles of tracheophytes and bryophytes. A necessary assumption of this ''modified history'' is that, while in the first phase of land colonisation by plants, the selection consistently supported complex photosynthetic gametophytes, something happened later, which made such developed gametophyte a burden. To our knowledge, no theory was put forward to elaborate as to the nature of this sudden change. Furthermore, the fact that no macrofossils of bryophytes are known from the time predating Silurian protracheophytes, seems to contradict the suggestions of earliest bryophyte-like flora (Gray 1993). Our topology makes very plausible another scenario of evolution, advocated by Remy et al. (1993), in which plants with dominant sporophyte and gametophyte generations independently developed from the protracheophytes with isomorphic life cycles. Once established, the respective dominant generation in both lines became ever increasingly complex, while the other generation underwent reduction. In bryophytes development went from photosynthetic sporophyte of hornworts to the reduced and parasitic sporophytes of liverworts and mosses. Presence of columella in protracheophyte *Horneophyton* (Kenrick and Crane 1997b) is in compliance with the basal position of protracheophytes to all of the land plants. The columella, accordingly was lost once in tracheophytes, and one other time in mosses.

Our choice of methods for phylogeny reconstruction was strongly limited due to compositional bias. Therefore it is desirable to check the results we report on a basis of some extensive, non-biased data employing substitution models which assume the stationary evolution process. We also would like to point out that our results must be perceived as preliminary, since the choice of taxa in this study was limited by the low number of complete chloroplast genomes available.

References

Crowe C. T., Pike L. M., Cross K. S., Renzaglia K. S. (1997) The *psb*A gene sequence can be used to infer phylogenetic relationships among the major lineages of bryophytes. Amer. J. Bot. 84: 14–15.

Duff R. J., Nickrent D. L. (1999) Phylogenetic relationships of land plants using mitochondrial small-subunit rDNA sequences. Amer. J. Bot. 86: 372–386.

Felsenstein J. (1989) PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics 5: 164–166.

Foster P. G., Hickey D. A. (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Molec. Evol. 48: 284–290.

Garbary D. J., Renzaglia K. S., Duckett J. G. (1993) The phylogeny of land plants - a cladistic analysis based on male gametogenesis. Pl. Syst. Evol. 188: 237–269.

Garbary D. J., Renzaglia K. S. (1998) Bryophyte phylogeny and the evolution of land plants: evidence from development and ultrastructure. In: Bates J. W., Ashton N. W., Duckett J. G. (eds.) Bryology for the twenty-first century. Maney Publishing and British Bryological Society, Leeds, U.K., pp. 45–63.

Goremykin V. V., Hirsch-Ernst K. I., Wolfl S., Hellwig F. H. (2003a) Analysis of the *Amborella trichopoda* chloroplast genome sequence suggests that *Amborella* is not a basal angiosperm. Molec. Biol. Evol. 20: 1499–14505.

Goremykin V. V., Hirsch-Ernst K. I., Wolfl S., Hellwig F. H. (2003b) The chloroplast genome of the ''basal'' angiosperm *Calycanthus fertilis* - structural and phylogenetic analyses. Pl. Syst. Evol. 242: 119–135.

Goremykin V. V., Hirsch-Ernst K., Wolfl S., Hellwig F. H. (2004) The chloroplast genome of *Nymphala alba*: whole genome analyses and the problem of identifying the most basal angiosperm. Molec. Biol. Evol. 21: 1445–1454.

Gray J. (1993) Major Paleozoic land plant evolutionary bio-events. Palaeogeogr. Palaeoclimatol. Palaeoecol. 104: 153–169.

Hedderson T. A., Chapman R. L., Rootes W. L. (1996) Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. Pl. Syst. Evol. 200: 213–224.

Hiratsuka J., Shimada H., R. Whittier R. et al. (1989) The complete sequence of the rice (*Oryza sativa*) chloroplast genome: intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. Mol. Gen. Genet. 217: 185–194.

Hupfer H., Swiatek M., Hornung S., Hermann R.G., Maier R.M., Chiu W.-L., Sears B. (2000) Complete nucleotide sequence of the *Oenothera elata* plastid chromosome, representing plastome I of the five distinguishable *Euoenothera* plastomes. Mol. Gen. Genet. 263: 581–585.

Karol K. G., McCourt R. M., Cimino M. T., Delwiche C. F. (2001) The closest living relatives of land plants. Science 294: 2351–2353.

Kato T., Kaneko T., Sato S., Nakamura Y., Tabata S. (2000) Complete structure of the chloroplast genome of a legume, *Lotus japonicus*. DNA Res. 7: 323–330.

Kenrick P., Crane P. R. (1997a) The origin and early evolution of plants on land. Nature 389: 33–39.

Kenrick P., Crane P. R. (1997b) The origin and early diversification of land plants: a cladistic study. Smithsonian Institution Press, Washington.

Kenrick P. (2000) The relationships of vascular plants. Philos. Trans., Ser. B. 355: 847–855.

Kugita M., Kaneko A., Yamamoto Y., Takeya Y., Matsumoto T., Yoshinaga K. (2003) The complete nucleotide sequence of the hornwort (*Anthoceros formosae*) chloroplast genome: insight into the earliest land plants. Nucl. Acids Res. 31: 716–721.

Lake J. A. (1994) Reconstructing evolutionary trees from DNA and protein sequences: paralinear distances. Proc. Natl. Acad. Sci. USA 91: 1455–1459.

Lewis L. A., Mishler B. D., Vilgalys R. (1997) Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbc*L. Molec. Phylogenet. Evol. 7: 377–393.

Lockhart P. J., Howe C. J., Bryant D. A., Beanland T. J., Larkum A.W. (1992) Substitutional bias confounds inference of cyanelle origins from sequence data. J. Molec. Evol. 34: 153–162.

Lockhart P. J., Steel M. A., Hendy M. D., Penny D. (1994) Recovering evolutionary trees under a more realistic model of sequence evolution. Molec. Biol. Evol. 11: 605–612.

Loomis W.F., Smith D.W. (1990) Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. Proc. Natl. Acad. Sci. USA 87: 9093–9097.

Maden A. R., Whittier D. P., Garbary D. J., Renzaglia K. S. (1997) Ultrastructure of the spermatozoid of *Lycopodiella lateralis* (Lycopodiaceae). Canad. J. Bot. 75: 1728–1738.

Maier R. M., Neckermann K., Igloi G. L., Kossel H. (1995) Complete sequence of the maize chloroplast genome: gene content, hotspots of divergence and fine tuning of genetic information by transcript editing. J. Molec. Biol. 251: 614–628.

Malek O., Laettig K., Hiesel R., Brennicke A., Knoop V. (1996) RNA editing in bryophytes and a molecular phylogeny of land plants. EMBO J. 15: 1403–1411.

Mishler B. D., Churchill S. P. (1984) A cladistic approach to the phylogeny of the "bryophytes". Brittonia 36: 406–424.

Mishler B. D., Thrall P. H., Hopple J. J. S., DeLuna E., Vilgalys R. (1992) A molecular approach to the phylogeny of bryophytes: cladistic analysis of chloroplast-encoded 16S and 23S ribosomal RNA genes. Bryologist 95: 172–180.

Mishler B. D., Lewis L. A., Buchheim M. A., Renzaglia K. S., Garbary D. J., Delwiche C. F., Zechman F. W., Kantz T. S., Chapman R. L. (1994) Phylogenetic relationships of the "green algae" and "bryophytes". Ann. Missouri Bot. Gard. 81: 451–483.

Nickrent D. L., Parkinson C. L., Palmer J. D., Duff R. J. (2000) Multigene phylogeny of land plants with special reference to Bryophytes and the earliest land plants. Molec. Biol. Evol. 17: 1885–1895.

Nishiyama T., Kato M. (1999) Molecular phylogenetic analysis among bryophytes and tracheophytes based on combined data of plastid coded genes and the 18S rRNA gene. Molec. Biol. Evol. 16: 1027–1036.

Nishiyama T., Wolf P. G., Kugita M., Sinclair R.B., Sugita M., Sugiura C., Wakasugi T., Yamada K., Yoshinaga K., Yamaguchi K., Ueda K., Hasebe M. (2004) Chloroplast phylogeny indicates that bryophytes are monophyletic. Molec. Biol. Evol. 21: 1813–1819.

OgiharaY., Isono K., Kojim T. et al. (2002) Structural features of a wheat plastome as revealed by complete sequencing of chloroplast DNA. Molec. Genet. Gen. 266: 740–746.

Ohyama K., Fukuzawa H., Kohchi T. et al. (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322: 572–574.

Qiu Y. L., Cho Y., Cox J. C., Palmer J. D. (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394: 671–674.

Remy W., Gensel P.G., Hass H. (1993) The gametophyte generation of some early devonian land plants. Int. J. Plant Sci. 154: 35–58.

Renzaglia K. S., Duff R. J. T., Nickrent D. L., Garbary D. J. (2000) Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. Philos. Trans. Ser. B. 355: 769–793.

Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S. (1999) Complete structure of the chloroplast genome of *Arabidopsis thaliana*. DNA Res. 6: 283–290.

Schmitz-Linneweber C., Maier R. M., Alcaraz J. P., Cottet A., Herrmann R. G., Mache R. (2001) The plastid chromosome of spinach (*Spinacia oleracea*): complete nucleotide sequence and gene organization. Pl. Molec. Biol. 45: 307–315.

Schweitzer H-J. (1990) Der Generationswechsel rheinischer Psilophyten. Bonner Paläobotanische Mitteilungen 8: 1–19.

Shinozaki K., Ohme M., Tanaka M. et al. (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J. 5: 2043–2049.

Steel M.A. (1994) Recovering a tree from the leaf colourations it generates under a Markov model. Appl. Math. Lett. 7: 19–23.

Steel M., Huson D., Lockhart P. J. (2000) Invariable sites models and their use in phylogeny reconstruction. Syst. Biol. 49: 225–232.

Strimmer K., von Haeseler A. (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Molec. Biol. Evol. 13: 964–969.

Strother P. K., Al-Hajiri S., Traverse A. (1996) New evidence for land plants from the lower middle Ordovician of Saudi Arabia. Geology 24: 55–58.

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 *rpo*A from the chloroplast to the nucleus. Nucl. Acids Res. 31: 5324–5331.

Swofford D. L. PAUP* (2002) Phylogenetic Analysis Using Parsimony (* and other Methods). Version 4. Sinauer associates, Sunderland, Massachusetts.

Thollesson M. (2004) LDDist: a Perl module for calculating LogDet pair-wise distances for protein and nucleotide sequences. Bioinformatics. 20: 416–418.

Turmel M., Otis C., Lemieux C. (2002) 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.

Wakasugi T., Tsudzuki J., Ito S., Nakashima K., Tsudzuki T., Sugiura M. (1994) Loss of all ndh genes as determined by sequencing the entire chloroplast genome of the black pine *Pinus thunbergii*. Proc. Natl. Acad. Sci. USA 91:9794–9798.

Wellman C. H., Osterloff P. L., Mohiuddin U. (2003) Fragments of the earliest land plants. Nature 425: 282–285.

Wolf P. G., Rowe C. A., Sinclair R. B., Hasebe M. (2003) Complete nucleotide sequence of the chloroplast genome from a leptosporangiate fern, *Adiantum capillus-veneris* L. DNA Res. 10: 59–65.

Address of the authors: Vadim Goremykin, Frank H. Hellwig, Institut für Spezielle Botanik, Friedrich-Schiller-Universität Jena, Philosophenweg 16, 07743 Jena, Germany. (e-mail: Vadim.Goremykin@uni-jena.de)