Origin of land plants: Do conjugating green algae hold the key?

Sabina Wodniok1, Henner Brinkmann2†, Gernot Glöckner3, Andrew J Heidel4, Hervé Philippe2, Michael Melkonian1 and Burkhard Becker1*

Abstract

Background: The terrestrial habitat was colonized by the ancestors of modern land plants about 500 to 470 million years ago. Today it is widely accepted that land plants (embryophytes) evolved from streptophyte algae, also referred to as charophycean algae. The streptophyte algae are a paraphyletic group of green algae, ranging from unicellular flagellates to morphologically complex forms such as the stoneworts (Charales). For a better understanding of the evolution of land plants, it is of prime importance to identify the streptophyte algae that are the sister-group to the embryophytes. The Charales, the Coleochoetales or more recently the Zygnematales have been considered to be the sister group of the embryophytes. However, despite many years of phylogenetic studies, this question has not been resolved and remains controversial.

Results: Here, we use a large data set of nuclear-encoded genes (129 proteins) from 40 green plant taxa (Viridiplantae) including 21 embryophytes and six streptophyte algae, representing all major streptophyte algal lineages, to investigate the phylogenetic relationships of streptophyte algae and embryophytes. Our phylogenetic analyses indicate that either the Zygnematales or a clade consisting of the Zygnematales and the Coleochaetales are the sister group to embryophytes.

Conclusions: Our analyses support the notion that the Charales are not the closest living relatives of embryophytes. Instead, the Zygnematales or a clade consisting of Zygnematales and Coleochaetales are most likely the sister group of embryophytes. Although this result is in agreement with a previously published phylogenetic study of chloroplast genomes, additional data are needed to confirm this conclusion. A Zygnematales/embryophyte sister group relationship has important implications for early land plant evolution. If substantiated, it should allow us to address important questions regarding the primary adaptations of viridiplants during the conquest of land. Clearly, the biology of the Zygnematales will receive renewed interest in the future.

Background

The ancestors of modern land plants (embryophytes) colonized the terrestrial habitat about 500 to 470 million years ago (Ordovician period [1-3]). This event was undoubtedly one of the most important steps in the evolution of life on earth [4-6], thereby establishing the path to our current terrestrial ecosystems [7] and significantly changing the atmospheric oxygen concentration [8,9]. Since this time three major groups of land plants evolved: bryophytes (liverworts, hornworts and mosses), pteridophytes (lycophytes and monilophytes) and spermatophytes with the latter dominating most habitats today.

It is widely accepted that embryophytes evolved from green algae, or more specifically, from a small but diverse group of green algae known as the streptophyte algae (charophycean algae). Streptophyte algae and embryophytes together constitute the division Streptophyta, which likely split from the Chlorophyta (all other green algae) about 725-1200 MY ago [10-12]. Streptophyta and Chlorophyta comprise the Viridiplantae, one of the three evolutionary lineages derived from the single primary endosymbiosis of a cyanobacterium and a eukaryotic host cell [13].

The Streptophyta are characterized by several morphological (e.g., structure of flagellate reproductive cells,
if present [14]), and physiological characters (e.g., occurrence of glyceraldehyde-3-phosphate dehydrogenase isoform B, GAPDH B [15], leaf peroxisome type of photorespiration [16,17]). Furthermore, several typical embryophyte traits have evolved within the streptophyte algae (e.g., cell division using a phragmoplast, structure of the cellulose synthase complex [4]). However, the streptophyte algae differ greatly in cellular organization and reproduction. Molecular phylogenies indicate that the Mesostigmatales and Chlorokybales form a clade that is a sister-group to all other streptophytes, currently containing only two genera: the biflagellate Mesostigma and the saccinoid (non-motile cells occurring in packages of four) Chlorokybus [18,19]. The Klebsormidiales, which is comprised of filamentous algae [14], is the sister group to the remaining streptophyte algae and the embryophytes. The phylogenetic position of the other three groups of streptophyte algae is currently controversial. The conjugating green algae (Zygnematales) today represent the most species-rich group of streptophyte algae and are characterized by their unique mode of sexual reproduction. They have completely lost flagellate cells, using instead conjugation for sexual reproduction [20]. The conjugating green algae include both filamentous and unicellular forms. The last two groups of streptophyte algae, the Coleochaetales and Charales, are filamentous with apical growth and an oogamous mode of sexual reproduction. Based on morphological complexity, either of the latter two groups have been suggested to be the sister group of the embryophytes [4]. In many illustrations referring to the evolution of streptophyte algae and embryophytes in textbooks [e.g. [21]] or review articles [14,22,23], the Charales (stoneworts) are depicted as the sister group of the embryophytes. The strongest support for a sister group relationship between Charales and embryophytes was obtained in all analyses. Bootstrap support values were computed for both methods, using the chlorophytes as outgroup to root the trees.

Results
Zygnematales alone or together with Coleochaetales as the sister group of embryophytes

New ESTs were sequenced from the streptophyte algae Klebsormium subtile, Coleochaete scutata, and Chara vulgaris and the chlorophyte alga Pyramimonas parkeae (see Material and Methods for details). We assembled a data set of 129 expressed genes (30,270 unambiguously aligned amino acid positions) for 40 viridiplant taxa including six streptophyte algae (Mesostigma, Klebsormium, Chara, Coleochaete, Closterium and Spirogyra) using the chlorophytes as outgroup to root the trees.

The data set was analyzed by maximum likelihood (ML) and Bayesian inference (BI) methods using several evolutionary models. We first evaluated the fits of the models to our data set using cross validation (Table 1). The site-heterogeneous CATGTR model is the best of the four models under study. The site-heterogeneous CAT model, which assumes uniform exchangeability rates among amino acids, has a much better fit than the site-homogeneous LG+F and GTR models, and is just slightly worse than the CATGTR model. Interestingly, the data set is sufficiently large to accurately estimate the amino acid exchangeability rates, since the GTR model has a better fit to the data than the LG+F model, where these parameters were learned from numerous alignments [28]. The simplifying assumption of equal rates of the CAT model, albeit biologically unsound and rejected by cross validation (in favor of the CATGTR model), has the advantage of allowing a significant increase in computational speed [29], and was therefore used for bootstrap analysis.

Despite very different model fits, the same tree topology (Figure 1) was obtained in all analyses. Bootstrap support values were computed for both methods, using the site homogeneous GTR+Γ4 model (ML) and the site heterogeneous CAT+Γ4 model (BI). The posterior probabilities of all nodes for both the CAT+Γ4 and CATGTR +Γ4 models were 1 except for three nodes (0.99 each, indicated with an asterisk in Figure 1). The molecular phylogeny of embryophytes and chlorophyte algae (outgroup) is in agreement with other recently published phylogenies [14,23,30-32] and supports the monophyly of liverworts and mosses which is however still a matter of debate [33]. The phylogeny of the streptophyte algae is

Table 1 Cross-validation results for the data set of 40 viridiplant species and 30,270 position (a positive score indicates a better fit)

| Model compared | Likelihood difference (±SD) |
|----------------|-----------------------------|
| GTR+Γ4 vs LG+Γ4 | 377.99 ± 28.51              |
| CAT+Γ4 vs LG+Γ4 | 1187.27 ± 78.75             |
| CATGTR+Γ4 vs LG+Γ4 | 1547.08 ± 63.35             |
asymmetrical. *Mesostigma* is sister to all the remaining streptophytes, as in other studies without *Chlorokybus* [18,19,24]. There is a long, highly supported, branch at the base of the clade uniting the other streptophyte algae and embryophytes, likely indicating the elapse of a substantial amount of time. In contrast, the phylogeny of the remaining streptophyte algae resembles an adaptive radiation, with the five major lineages, including the embryophytes, appearing serial, but with relatively short internal branches. Within this clade, *Klebsormidium* is sister to all the other species. The Charales are sister to a clade comprising Coleochaetales, Zygnematales and
embryophytes and are therefore an unlikely candidate as the sister-group of the embryophytes (only BP of 5% and 2% with CAT+Γ4 and GTR+Γ4 models, respectively). The latter clade is moderately well supported (BP of 88% and 94%) and its support is lower than for the clade Klebsormidium+Chara+Coleochaetales+Zygnematales+embryophytes (100%). However, the sister-group relationship of embryophytes and Zygnematales (BP 83% and 54%) is less well supported, especially in the analysis under the site-homogeneous GTR+Γ4 model. A fraction of the bootstrap replicates supports the alternative topology that unites Coleochaete with the Zygnematales (BP 12% and 35%, respectively). The data indicate that Mesostigma, Chara, Klebsormidium and Coleochaete are evolving at a comparable and moderate rate, with embryophytes and Zygnematales evolving faster. For instance, the Zygnematales appear to have evolved twice as fast as Coleochaete.

This difference in evolutionary rate suggests that the grouping of embryophytes and Zygnematales could be due to a long-branch attraction (LBA) artifact [34]. To explore this possibility, we analyzed two reduced taxon samples, where the 15 fastest-evolving land plants have been discarded (Figure 2A) and the 13 long-branched chlorophytes and Mesostigma were not used as an outgroup (Figure 2B). The impact of LBA should be reduced in both cases. Again, the four models and the two methods (ML and BI) lead to identical topologies for both data sets. Interestingly, in both cases, the topology differs from the one of the complete data set (40 species, Figure 1) by the appearance of a sister-group relationship between Coleochaete and Zygnematales. However, this grouping receives non-significant support (PP between 0.51 and 0.90 and BP of 36 and 72%). In both cases, the Zygnematales plus Coleochaete clade is the closest relative to embryophytes (however, with limited support when the outgroups were removed [BP of 74%, Figure 2B]). It is noteworthy, that in none of the analyses was Chara recovered as a sister clade to the embryophytes (low BP: 2% in Figure 2A and 19% in Figure 2B).

Another cause of systematic errors in phylogenetic inference is the compositional heterogeneity across taxa [35,36]. A principal component analysis of the amino acid composition (Figure 3) demonstrates that Coleochaete, Chara and Spirogyra and most embryophytes have a similar composition. Other organisms (e.g. Prototheca, Chlorella, Closterium and Klebsormidium) show much larger compositional differences, but are correctly placed in a phylogenetic tree due to the presence of a strong phylogenetic signal, as the strongly supported monophyly of Closterium+Spirogyra and Selaginella+Huperzia illustrate. We explored the potential impact of compositional heterogeneity on our inference by using the Dayhoff recoding, an approach known to be efficient [37-39]. Interestingly, the three models (GTR, CAT and CATGTR) and the three taxon samples in Figures 1 and 2A, B all lead to the same topology as in Figure 1. Since the Dayhoff recoding reduces not only compositional heterogeneity but also saturation, the sister-group relationship between the Zygnematales and embryophytes as observed in Figure 1 is less likely to result from systematic error.

Discussion

Previous studies of the phylogeny of Streptophyta were restricted mainly to ribosomal RNA or sequences of organellar origin [24-27]. We now, for the first time, large data sets of nuclear-encoded proteins for phylogenetic studies in this important evolutionary lineage. Our phylogenetic analyses are in agreement with both phylogenies obtained using a data set of concatenated plastid proteins or ribosomal RNAs [26,27], but are in conflict with the 4 gene tree mentioned above [24]. In contrast, Coleochaete was found to be sister to embryophytes [40] in a recent analysis based on 77 nuclear encoded ribosomal proteins (12,459 amino acid positions). However, as this study failed to recover the monophyly of the Coleochaetales (placing Chaetosphaeridium within the Zygnematales) the conclusions from this study should be treated with caution. In the 4 gene analysis the topology (Zygnematales, (Coleochaetales, (Charales, embryophytes))) was observed. This analysis suggested that the streptophyte algae regularly (without reversions) evolved towards increasing morphological complexity (resulting in a larger number of shared morphological characters with embryophytes). In contrast, our results and the results obtained using chloroplast data [25,26] suggest that most likely the morphologically simpler Zygnematales (or a clade consisting of Zygnematales and Coleochaetales) is the sister group of embryophytes, rather than the Charales. It seems plausible that the simpler morphology of extant Zygnematales represents a secondary simplification, similar to the loss of flagellate cells in this group, which may actually represent an adaptation to ensure sexual reproduction in the absence of free water [41]. Alternatively, the morphological complexity of the Charales and Coleochaetales might have evolved independently after the three evolutionary lines (Coleochaetales, Charales, and Zygnematales) diverged. This kind of scenario was already proposed by Stebbins and Hill [41]: They suggested that the early evolution of streptophyte algae took place in a moist terrestrial habitat and involved rather simple unicellular types. They considered the extant Coleochaetales, Charales, Klebsormidiales and Zygnematales to be derived forms with a secondary aquatic life style. A fast initial radiation at the time of colonization of the terrestrial habitat by the ancestors of
Figure 2 Consensus Tree inferred by PhyloBayes using reduced data sets of 25 taxa (A. Spermatophytes and Selaginella eliminated) or 26 taxa (B. Chlorophytes and Mesostigma eliminated). The same methods and models were used as in Figure 1, with the only exception that no bootstrap analyses were performed for the PhyloBayes analyses, for which only the posterior probabilities are given. The alternative taxon samplings were aimed at either eliminating the fast evolving embryophytes (all spermatophytes and Selaginella) (A) or the distantly-related outgroup sequences eliminating chlorophytes and Mesostigma (B).
modern streptophyte algae as proposed by Stebbins and Hill [41] may also explain the relative difficulty to infer
the phylogenetic relationships among the four groups of streptophyte algae, enhanced by the accelerated evolu-
tionary rate of the Zygnematales. However, in contrast to
Stebbins and Hill [41], we argue that at least some of the
morphological complexity had evolved prior to the early
radiation of the streptophyte algae for the following rea-
sons: (1) Based on the available fossil record, the Charales
already had a morphology similar to that of extant forms
in the Silurian period [42], (2) the available EST data
indicate that Zygnematales, Coleochaetales and Charales
possess homologues of a number of proteins that are
involved in the development of morphologically complex
structures in embryophytes, such as GNOM, Wuschel,
Meristemlayer 1, MIKc-type MADS-box protein ([43,44]
and unpublished results). Taken together, these data
lend support to the idea that the extant morphological
complexity of the Charales and Coleochaetales is an
ancient trait that may have been secondarily lost in the
Zygnematales.

Alternatively, as was found by the comparison of Volvox,
a multicellular system, with its close relative Chlamydomo-
nas reinhardtii, the evolution of multi-cellular structures
seems to rely mainly on the reorganization and differential
regulation of already existing genes [45]. From this point
of view, the complex morphologies of the Charales and
Coleochaetales could have evolved completely independ-
ently by using the toolbox already present in the com-
mon ancestor of these streptophyte algae. Although our
results, based on the complete data set, favor the Zygnem-
atales as the sister group of the embryophytes, the
results from the two alternative taxon sampling tests, in
which we tried to reduce as much as possible potential
disturbing influences of the LBA artifacts, seem to point
rather to a sister group relationship between Coleochaete
and the Zygnematales. In contrast, Dayhoff coding favors
the grouping of the Zygnematales with the embryophytes,
whatever the taxon sampling. Since this recoding is
expected to reduce several sources of systematic errors,
this topology is more likely. However, the support in this
part of the tree remains limited, and large-scale genomic
data from more streptophyte algae (especially Coleochaete,
Charales and Klebsormidiales) are needed to resolve
this question. Whatever the relative position of Coleo-
chaete and the Zygnematales, our analysis supports the
scenario of a secondary loss of morphological complexity
in the Zygnematales.

The first land plants encountered a more extreme
environment compared to a freshwater habitat, with
large fluctuations in water content (wetting and desicca-
tion), radiation intensity (visible light and UV) and
nutrient supply. Potentially, the last common ancestor
of the Zygnematales and embryophytes was better
adapted to these types of environmental stressors than other streptophyte algae. The more variable environmental conditions might have also favored the evolution of more complex signaling pathways [46]. Rensing et al. [46] discussed several proteins likely to be important for the adaptation of embryophytes to their terrestrial habitat. Preliminary analysis of the available ESTs from streptophyte algae indicate that expressed genes similar to most of the proteins listed by Rensing et al. [46] can be found in various streptophyte algae (Table 2). For example, proteins similar to major light harvesting complex II proteins (lhcb1-3), which were considered to be missing from green algae [46-48], are clearly found in ESTs from streptophyte algae except Mesostigma (Table 2). Expressed genes similar to late embryo abundant (LEA) proteins known to protect spermatophyte seeds from desiccation [49] are found in several streptophyte algae (Table 2).

The probable fast radiation of the derived lineages of streptophyte algae (see above) in conjunction with the secondary morphological simplification of the Zygnematales makes it difficult to find any synapomorphies for the possible sister group relationship of the Zygnematales and embryophytes. We note two complex traits that might potentially support this relationship. Firstly, components of the “auxin signaling machinery” are highly conserved in embryophytes [46,50], but appear to be absent in streptophyte algae, except for the auxin binding protein (abp1), which can be found in various green algae including chlorophytes [50]. However, as also noted by De Smet et al. [51] the recently published ESTs from Spirogyra [52] include expressed genes similar to components (ARF, PIN, PINOID, Table 2) of the embryophyte-specific “auxin signaling machinery”. Secondly, embryophytes generally show chloroplast movements in response to high (avoidance response) or low light (accumulation response), which has been shown to be of ecological importance [53].

Conclusions
Knowledge of the phylogenetic relationships within streptophyte algae is of crucial importance for developing a realistic scenario for the colonization of the terrestrial habitat and the origin and early evolution of embryophytes. Phylogenomic analyses of nuclear and chloroplast data now indicate that the Charales are most likely not the closest living extant relatives of the embryophytes despite their morphological complexity. Instead, the analyses favor either the Zygnematales or, less likely, a clade consisting of the Zygnematales and Coleochaetales as the sister group of embryophytes. An extended taxon sampling and/or analyses of larger data sets such as complete genomes/transcriptomes will likely be necessary to shed further light on the elusive sister group of the embryophyte plants.

Methods
Preparation of cDNA libraries and EST sequencing
The preparation of cDNA libraries for Pyramimonas parkeae, Klebsormidium subtile and Coleochaete scutata, EST sequencing and processing of the primary reads have all been described by Wodniok et al. [56]. Chara vulgaris zygotes were collected from the botanical garden of the Universität zu Köln. Zygotes were surface-sterilized using the following protocol (modified after [57,58]): after

Table 2 Proteins proposed to be important in the adaptation to the terrestrial habitat [46] are present in streptophyte algae

| Gene name                  | TAIR number | Klebsormidium subtile | Chara vulgaris | Coleochaete scutata | Coleochaete orbicularis | Closterium spec. | Spirogyra pratense |
|----------------------------|-------------|-----------------------|----------------|--------------------|-------------------------|-----------------|-------------------|
| Photosynthesis             | Lhcb3       | At5g4270              | 8 (e-119)      | 9 (e-38)           | 3 (e-117)               | 78 (e-102)      | (e-126)           | 37 (e-126)        |
| Dessication tolerance      | Lea1        | At3g1810              | 1 (e-20)       | 1 (e-33)           | 1 (e-19)                | 1 (e-189)       | n.d.              | ?                 |
| Ethylene signaling         | EIN         | At5g03280             | n.d.           | n.d.               | 1 (e-22)                | n.d.            | n.d.              | n.d.              |
|                            | ETR         | At1g66340             | n.d.           | n.d.               | n.d.                    | n.d.            | 1 (e-105)         | n.d.              |
|                            | ACS         | At5g65800             | n.d.           | n.d.               | 1 (e-53)                | n.d.            | 1 (e-49)          | n.d.              |
| Auxin signaling            | ARF         | At5g62010             | n.d.           | n.d.               | n.d.                    | n.d.            | 1 (e-44)          | n.d.              |
|                            | PIN         | At1g73590             | n.d.           | n.d.               | n.d.                    | n.d.            | 1 (e-33)          | n.d.              |
|                            | PINOID      | At2g34650             | n.d.           | n.d.               | n.d.                    | n.d.            | 1 (e-72)          | n.d.              |

BLAST analyses were performed to identify putative homologues in the indicated streptophyte algae. The number of contigs and the best e-value obtained are given.

n.d. not detected, ? no clear result. 1) for most species ESTs showing similarity to several different Lea proteins were found (e.g., in Klebsormidium ESTs we detected additional sequences showing similarity to Lea14, Lea76 and Lea D29).
washed with distilled water, zygotes were rinsed with 
EtOH (70%, 1-2 min), followed by sodium hypochlorite 
(7-12%, 20-25 min). In some experiments the EtOH step 
was omitted. Surface-sterilized zygotes were rinsed repeat-
edly with sterile distilled water to remove hypochlorite 
etanol. For germination, single zygotes were each 
placed in a well of a microtiter plate and incubated in 
Chara-medium ([57,58]) at 24°C and a 14/10 h light dark 
cycle at 20 - 40 μEm-2s-1. Zygo
tes germinated after 
4-6 weeks. After germination, young plants were trans-
ferred into 100 ml Erlenmeyer flasks containing 10 ml 
agar overlaid with 40 ml Chara-medium. Cultures often 
underwent sexual reproduction within one year. Prepara-
tion of cDNA libraries and Sanger sequencing was done as 
described earlier [56,59].

For 454 sequencing Chara total RNA was isolated 
using Trizol (Invitrogen) following the manufacturer’s 
instructions. The cDNA library was made from the 
RNA with the Mint cDNA synthesis kit (Evrogen) using 
21 cycles in the PCR amplification step. The cDNA sub-
sequently was converted to a Roche/454 sequencing 
library (rapid) according to the manufacturer’s protocols.
Sequencing of this library yielded 740,341 raw reads 
with 245 Mb raw sequence data. Assembly resulted in 
13,615 contigs spanning 6.5 Mb.

Phylogenetic analyses

The data set assembly and the detection of possible non-
orthologous sequences were performed as described else-
where in detail [60,61]. Briefly, the latter approach is based on the assumption that the tree obtained in the phylo-
genetic analysis of the concatenated data set (super-matrix) is a good proxy of the “true” tree. All single gene data sets 
were analyzed separately with RAxML using the LG+Γ4 
model including 100 bootstrap pseudo-replicates [62]. 
Nodes of the single gene trees that are in conflict to the 
reference tree (super-matrix) and that are supported by a 
bootstrap value ≥ 70% are considered to be incongruent. 
Most of these conflicts are usually due to problems of the phylogenetic reconstruction (stochastic, but also systema-
tic errors), most often the conflict can be resolved by a 
nearest-neighbor-interchange (NNI). However, occasion-
ally there are also true conflicts related to the presence of 
paralogous or xenologous sequences and the genes were 
therefore discarded from the super-matrix.

The final data set was assembled using Scafos [63] and 
consisted of 46 taxa, including six distant outgroup 
two glaucophytes and four red algae), with a total of 
30,270 amino acid positions coming from 129 genes. To 
reduce the amount of missing data three chimerical 
sequences were made within the streptophytes, two at the 
genus and one at the family level, as well as a higher 
order (class) chimera invoking the Florideophyceae. By 
allowing no more than 16 missing taxa for any given 
gene, the final amount of missing data in the superma-
trix was 30%. The concatenated data set and three sub-
sets were made within the streptophytes, two at 
the family level, as well as a higher 
order (class) chimera invoking the Florideophyceae. By 

Phylogenetic analysis of the three Dayhoff-recoded data 
sets was performed using PhyloBayes with the GTR+Γ4 
and CATGTR+Γ4 models [29]. Two independent chains 
were run per analysis for 10,000 cycles (with each 10th 
cycle sampled) and their bipartitions were compared 
after the elimination of the “burn-in” in order to test the 
quality of the convergence. The maximal difference 
observed between bipartition frequencies of two inde-
pendent runs was always lower than 0.1. Furthermore, a 
bootstrap analysis with 100 pseudo-replicates was per-
formed under the CAT+Γ4 model, a chain per dataset 
was run for the same length and sampled as above, the 
“burn-in” was fixed to 1,000 cycles. Each of the 100 
resulting consensus trees was then used as an input for 
the program Consense of the Phylip package to generate 
the bootstrap consensus tree [65]. We performed cross 
validation tests to evaluate the fit of the four models 
used (LG, GTR, CAT and CATGTR). The analysis was 
performed in PhyloBayes, using ten randomly generated 
replicates, in which the original data set was divided 
into training data sets (9/10 of the positions) to estimate 
the parameters of the given model and into the test data 
sets (1/10 of the positions) to calculate with these para-
meters the likelihood scores.

The amino-acid composition of the 46 species data set 
was visualized by assembling a 20 × 46 matrix containing 
the frequency of each amino acid per species using the 
program NET from the MUST package [66]. This matrix 
was then displayed as a two-dimensional plot in a prin-
cipal component analysis, as implemented in the R pack-
age. To counteract sequence bias, we recoded the 20 
amino acids into six groups as previously proposed [39]. 
Phylogenetic analysis of the three Dayhoff-recoded data 
sets was performed using PhyloBayes with the GTR+Γ4, 
CAT+Γ4 and CATGTR+Γ4 models.

Data access

EST reads (Sanger) were deposited in Genbank under the 
following accession numbers: Klebsormidium subtile 
(LIBEST_027068), Coleochaete scutata (LIBEST_027067). 
The 454 sequence of Chara vulgaris can be found in the 
Sequence Read Archive (SRP005673). The alignment has 
been deposited to Treebase (S11199).
Acknowledgements
This work was supported by the DFG (Be1779/7-2 and 7-4) and a DAAD fellowship to SW. HP is funded by CRC and NSERC and RQHP for computational resources.

Author details
1Bozeninstitut Koln, Botanik, Universitat zu Koln, Zülpicher Straße 47b, 50674 Köln, Germany. 2Centre Robert-Cedergren, Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, QC H3C3J7, Canada. 3Leibniz-Institute of Freshwater Ecology and Inland Fisheries, IGB Muggelseeedamm 310, D-12587 Berlin, Germany. 4Fritz-Lipmann-Institut, Köln, Germany. 2Centre Robert-Cedergren, Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, Qc H3C3J7, Canada.

References
1. Sanderson MJ, Thorne JL, Wikstrom N, Bremer K: Early evolution of land plants: Phylogeny, physiology, and plant divergence times. Annu Rev Ecol Syst 2005, 36:729-748.

Received: 10 December 2010 Accepted: 18 April 2011 Published: 18 April 2011

References
1. Sanderson MJ, Therne JL, Wikstrom N, Bremer K: Molecular evidence on plant divergence times. Am J Bot 2004, 91(10):1656-1665.

2. Rubenstein CV, Gemmerne P, de la Puente GS, Astin RA, Steemans P: Early Middle Ordovician evidence for land plants in Argentina (eastern Gondwana). New Phytol 2010, 188(2):365-369.

3. Lang D, Weich B, Timmerhaus G, Richards S, Riano-Pachon DM, Corea LGG, Reski R, Mueller-Roeber B, Rensing SA: Genome-Wide Phylogenetic Comparative Analysis of Plant Transcriptional Regulation: A Timeline of Loss, Gain, Expansion, and Correlation with Complexity. Genome Biol Evol 2010, 2:468-503.

4. Graham LE: Origin of land plants. New York: John Wiley & Sons, Inc.; 1993.

5. Kenrick P, Crane PR: The origin and early diversification of land plants. Washington, London: Smithsonian Institution Press; 1997.

6. Bateman RM, Crane PR, DImichelle WA, Kenrick PR, Rowe NP, Speck T, Stein WE: Early evolution of land plants: Phylogeny, physiology, and ecology of the primary terrestrial radiation. Annu Rev Ecol Syst 1998, 29:261-292.

7. Waters ER: Molecular adaptation and the origin of land plants. Molecular Phylogenetics and Evolution 2003, 29(4):456-463.

8. Beman RA: Atmospheric oxygen over Phanerozoic time. Proc Natl Acad Sci USA 1999, 96(20):10955-10957.

9. Scott AC, Glasspool IJ: Phylogeny and evolution of chlorophyte algae and land plants. Journal of Systematics and Evolution 2008, 46(3):287-306.

10. Douglas RM, Delwiche CF, Karel KC: Chlorophyte algae and land plant origins. Trends Ecol Evol 2004, 19(12):661-666.

11. Qiu YL: Phylogeny and evolution of chlorophyte algae and land plants. Journal of Systematics and Evolution 2008, 46(3):287-306.

12. Karol KG, McCourt RM, Cimino MT, Delwiche CF: The closest living relatives of land plants. Science 2001, 294(5550):2351-2353.

13. Turner M, Otis C, Lemieux C: The chloroplast genome sequence of Chara vulgaris sheds new light into the closest green algal relatives of land plants. Mol Biol Evol 2006, 23(9):1324-1339.

14. Turmel M, Pombert JF, Chatebois P, Otis C, Lemieux C: The green algal ancestry of land plants as revealed by the chloroplast genome. Int J Plant Sci 2007, 168(5):679-689.

15. Turmel M, Ehara M, Otis C, Lemieux C: Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences. J Phycol 2002, 38(2):364-375.

16. Le SQ, Gascoquel O: An improved general amino acid substitution matrix. Mol Biol Evol 2008, 25(7):1307-1320.

17. Lartillot N, Philippe H: A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. Mol Biol Evol 2004, 21(6):1095-1109.

18. Mann B, Melkonian M: Molecular Phylogeny and Classification of the Mamiellaphyceae class. nov (Chlorophyta) based on Sequence Comparisons of the Nuclear- and Plastid-encoded rRNA Operons. Protist 2010, 161(2):304-336.

19. Turmel M, Gagnon MC, O’Keele CJ, Otis C, Lemieux C: The chloroplast genomes of the green algae Pyramimonas, Monomastix, and Pyrococcus mediterraneus carry within the origin of the secondary chloroplasts of euplastids. Mol Biol Evol 2009, 26(3):631-648.

20. Gogolnikin W, Hellwig FH: Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. Plant Syst Evol 2005, 254(1-2):93-103.

21. Qiu Y, Li LB, Wang B, Chen ZD, Knoop V, Groth-Malonek M, Kelly CJ, Otis C, Lemieux C: The chloroplast genome of Chara vulgaris sheds new light into the closest green algal relatives of land plants. Mol Biol Evol 2006, 23(9):1324-1339.

22. Phillips MJ, Delwiche F, Penny D: Genome-scale phylogeny and the detection of systematic biases. Mol Biol Evol 2004, 21(7):1455-1458.

23. Cox CJ, Foster PG, Hirt RP, Harris SR, Embley TM: The archaebacterial origin of eukaryotes. Proc Natl Acad Sci USA 2008, 105(51):20356-20361.

24. Rodriguez-Ezpeleta N, Brinkmann H, Broue B, Lartillot N, Lang BF, Philippe H: Detecting and overcoming systematic errors in genome-scale phylogenies. Systematic Biology 2007, 56(3):389-399.

25. Hidy J, Hirt RP, Dolezal P, Bardonova L, Foster PG, Tachezy J, Martin Embley T: Trichomonas hydrogenosomes contain the NADH dehydrogenase module of mitochondrial complex i. Nature 2004, 432(7017):618-622.

26. Finet C, Timme RE, Delwiche CF, Marletaz F, Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants. Curr Biol 20(24):2217-2222.

27. Stebbins GL, Hill GJC: Did multicellular plants invade the land. Am Nat 1980, 115(3):542-553.
42. Martin-Closas C: The fossil record and evolution of freshwater plants: A review. Geol Acta 2003, 1(4):315-338.
43. Timme R, Delwiche C: Uncovering the evolutionary origin of plant molecular processes: comparison of Coleochaete (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes. BMC Plant Biology 2010, 10(1):96.
44. Tanabe Y, Hasebe M, Sekimoto H, Nishiyama T, Kitami M, Henschel K, Munster T, Theissen G, Nozaki H, Ito M: Characterization of MADS-box genes in charophycean green algae and its implication for the evolution of MADS-box genes. Proc Natl Acad Sci USA 2005, 102(7):2436-2441.
45. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishi I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, et al: Genomic analysis of organismal complexity in the multicellular green alga Volvox carteri. Science 2010, 329(5988):223-226.
46. Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud PF, Lindquist EA, Kamisugi Y, et al: The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science 2008, 319(5859):64-69.
47. Kozol AG, Borza T, Ishida KI, Keeling P, Lee RW, Durnford DG: Structural and functional diversification of the light-harvesting antennae in chlorophyll a/b-containing organisms. Plant Physiol 2007, 143(4):1802-1816.
48. Nellison J, Durnford D: Structural and functional diversification of the light-harvesting complexes in photosynthetic eukaryotes. Photosynth Res 2010, 106(1):7-11.
49. Hundertmark M, Hinch DA: LEA (Late Embryogenesis Abundant) proteins and their encoding genes in Arabidopsis thaliana. BMC Genomics 2008, 9(1):118.
50. Lau S, Shao N, Bock R, Jurgens G, De Smet I: Auxin signaling in algal lineages: fact or myth? Trends Plant Sci 2009, 14(4):162-168.
51. De Smet I, Vois L, Lau S, Wilson M, Shao N, Timme R, Swarup R, Kerr L, Hodgman C, Bock R, et al: Unraveling the evolution of auxin signaling. Plant Physiol 2010, pp10.168161.
52. Timme RE, Delwiche CF: Uncovering the evolutionary origin of plant molecular processes: comparison of Coleochaete (Coleochaetales) and Spirogyra (Zygnematales) transcriptomes. BMC Plant Biology 2010, 10(1):96.
53. Wada M, Kagawa T, Sato Y: Chloroplast movement. Annu Rev Plant Biol 2003, 54:455-468.
54. Suetsugu N, Wada M: Chloroplast photorelocation movement mediated by phototropin family proteins in green plants. Biochim Biophys Acta 2007, 1778(9):927-935.
55. Suetsugu N, Wada M: Phytochrome-dependent photomovement responses mediated by phototropin family proteins in cryptogam plants. Photosynthetica 2007, 83(1):87-93.
56. Wodniok S, Simon A, Glockner G, Becker B: Gain and loss of polyadenylation signals during evolution of green algae. BMC Evolutionary Biology 2007, 7:55.
57. Forsberg C: Nutritional Studies of Chara in Axenic Cultures. Physiol Plant 1965, 18(2):275.
58. Forsberg C: Sterile Germination of Oospores of Chara and Seeds of Najas Marina. Physiol Plant 1965, 18(1):128.
59. Becker B, Feja N, Melkonian M: Analysis of expressed sequence tags (ESTs) from the scaly green flagellate Scherffelia dubia Pascher emend. Melkonian et Preisig, Protot 2001, 152(2):139-147.
60. Pick KS, Philippe H, Schreiber F, Erpenbeck D, Jackson DJ, Wrede P, Wiens M, Alle A, Morgenstern B, Manuel M, et al: Improved phylogenomic taxon sampling noticeably affects nonbilaterian relationships. Mol Biol Evol 2010, 27(9):1983-1987.
61. Philippe H, Dereelle R, Lopez P, Pick K, Borchelliini C, Bouy-Easnault N, Vacelet J, Renard E, Houliston E, Queinnec E, et al: Phylogenomics revises traditional views on deep animal relationships. Curr Biol 2009, 19(8):706-712.
62. Stamatakis A: RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 2006, 22(21):2686-2690.
63. Roure B, Rodrigues-Espelena N, Philippe H: SCafos: a tool for selection, concatenation and fusion of sequences for phylogenomics. BMC Plant Biology 2007, 7:12.
64. Lartillot N, Lepage T, Blanquart S: PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. Bioinformatics 2009, 25(17):2286-2288.