Review

Exploring plant biodiversity: the *Physcomitrella* genome and beyond

Daniel Lang[1], Andreas D. Zimmer[1], Stefan A. Rensing[2] and Ralf Reski[1,2,3]

[1] Plant Biotechnology, Faculty of Biology, University of Freiburg, Schaenzlestr. 1, D-79104 Freiburg, Germany
[2] Freiburg Initiative for Systems Biology (FRISYS), Faculty of Biology, University of Freiburg, Schaenzlestr. 1, D-79104 Freiburg, Germany
[3] Centre for Biological Signalling Studies (bioss), University of Freiburg, Schaenzlestr. 1, D-79104 Freiburg, Germany

For decades, plant molecular biology has focused on only a few angiosperm species. Recently, the ~500 mega base pairs (Mb) of the haploid *Physcomitrella patens* genome were sequenced and annotated. Mosses such as *P. patens* occupy a key evolutionary position halfway between green algae and flowering plants. This draft genome, in comparison to existing genome data from other plants, allows evolutionary insights into the conquest of land by plants and the molecular biodiversity that land plants exhibit. As a model organism, *P. patens* provides a well-developed molecular toolbox, including efficient gene targeting in combination with the morphologically simple moss tissues. We describe current as well as future tools for *P. patens* research and the prospects they offer for plant research in general.

Biology of the evo-devo model plant *Physcomitrella patens*

The moss *Physcomitrella patens* (Hedw.) Bruch and Schimp. (*P. patens*) has been developed as a model organism over the last two decades, providing a well-developed molecular toolbox including efficient gene targeting in combination with the morphologically simple moss tissues. Both organellar genomes, the mitochondrial [1] and the chloroplast [2] genomes, are fully sequenced and have already revealed valuable insights into the evolution of Plantae (Glaucophyta, Rhodophyta and Viridiplantae; Figure 1; example applications of the organellar genome are given in Refs [3–5]). Recently, the draft of the *P. patens* nuclear genome was reported [6]. Here, we review the principal findings, provide an overview of the available resources in *P. patens* research and show possible future avenues of plant research.

P. patens (synonym: *Aphanorrhegma patens* (Hedw.) Lindb.) belongs to the family *Funariaceae* within the order *Funariales* as part of the class *Bryopsida*. *Funariaceae* are a family of short-lived, minute to medium-sized, light to yellow-green and annual to biennial plants that grow gregarious to open tufts. The worldwide family consists of ~16 genera containing an estimated 300 species (see the Global Biodiversity Information Facility's *Funariaceae* entry: http://data.gbif.org/species/browse/taxon/13145815). *P. patens* is a monoecious, self-fertile, annual opportunist growing in late summer to autumn in open, unshaded, limy, loamy, moist and disturbed but nutrient-rich habitats, often close to the waterline. The heterophasic lifecycle was reported to be completed in the wild in about four weeks (J.P. Frahm, personal communication), but it takes six to eight weeks in the laboratory.

First version of the *P. patens* nuclear genome

The *P. patens* genome was sequenced as part of the U.S. Department of Energy's community sequencing program by a whole-genome shotgun approach at the Joint Genome Institute (JGI) in 2005. The draft sequence was published in early 2008 [6]. Three libraries with different insert sizes were sequenced to 8.6x clone coverage. Based on microscopy and flow cytometry, the haploid genome was estimated to consist of 27 chromosomes with a total length of ~510 mega base pairs (Mb) [7]. The overall scaffold length of the current v1.1 genome assembly (~480 Mb, see Table 1) matches this estimate. However, the sequences

Glossary

Contig: contiguous consensus sequences assembled from multiple sequence reads.
Gypsy and Copia-like retrotransposons: retrotransposons propagate via an RNA intermediate, which is reverse-transcribed and packed into virus-like particles. The major structural differences between the Gypsy- and Copia-like groups is the order of the reverse transcriptase (RT) and integrase (INT) coding regions in their *pol* genes (the gag and env domains can also be included). Gypsy-like: gag-RT-INT-env; Copia-like: gag-INT-RT-env [66].
Helitron: class II/DNA transposons that amplify via a rolling-circle mechanism.
L50: genome-assembly-specific terminology – the length of the scaffold that separates the top half (N50) of the assembled genome from the remainder of smaller scaffolds, if the sequences are ordered by size.
Long-terminal repeat (LTR): a directly repeated sequence at each end of a retrovirus or retrotransposon that is necessary for reverse transcription, integration and transcription.
Mate pair: end sequences derived from the same physical clone, representing opposite ends of the clone.
Monoecious: both sexual organs are present on a single individual.
N50: genome-assembly-specific terminology – the number of scaffolds that represent the top half of the assembled genome, if the sequences are ordered by size.
Peristome: in mosses, the peristome is a specialized structure in the sporangium that allows for gradual spore discharge, instead of releasing them all at once. Most mosses produce a capsule with a lid (the operculum), which falls off once the spores inside are mature and thus ready to be dispersed. The opening thus revealed is called the stoma (meaning 'mouth') and is surrounded by one or two peristomes. Each peristome is a ring of triangular 'teeth' formed by the remnants of thickened cell walls (arthrodontous) or whole cells (nematodontous).
Scaffold: scaffolds are sets of ordered, oriented contigs positioned relative to each other by mate pairs whose reads are present in adjacent contigs.
Seta: part of the sporophytic tissue (diploid). A stalk supporting the capsule of a moss or liverwort and supplying it with nutrients.
Synonymous site: silent codon positions that can mutate without changing the resulting amino acid.

Corresponding author: Reski, R. (ralf.reski@biologie.uni-freiburg.de).

1360-1385/$ – see front matter © 2008 Published by Elsevier Ltd. doi:10.1016/j.tplants.2008.07.002 Available online 30 August 2008

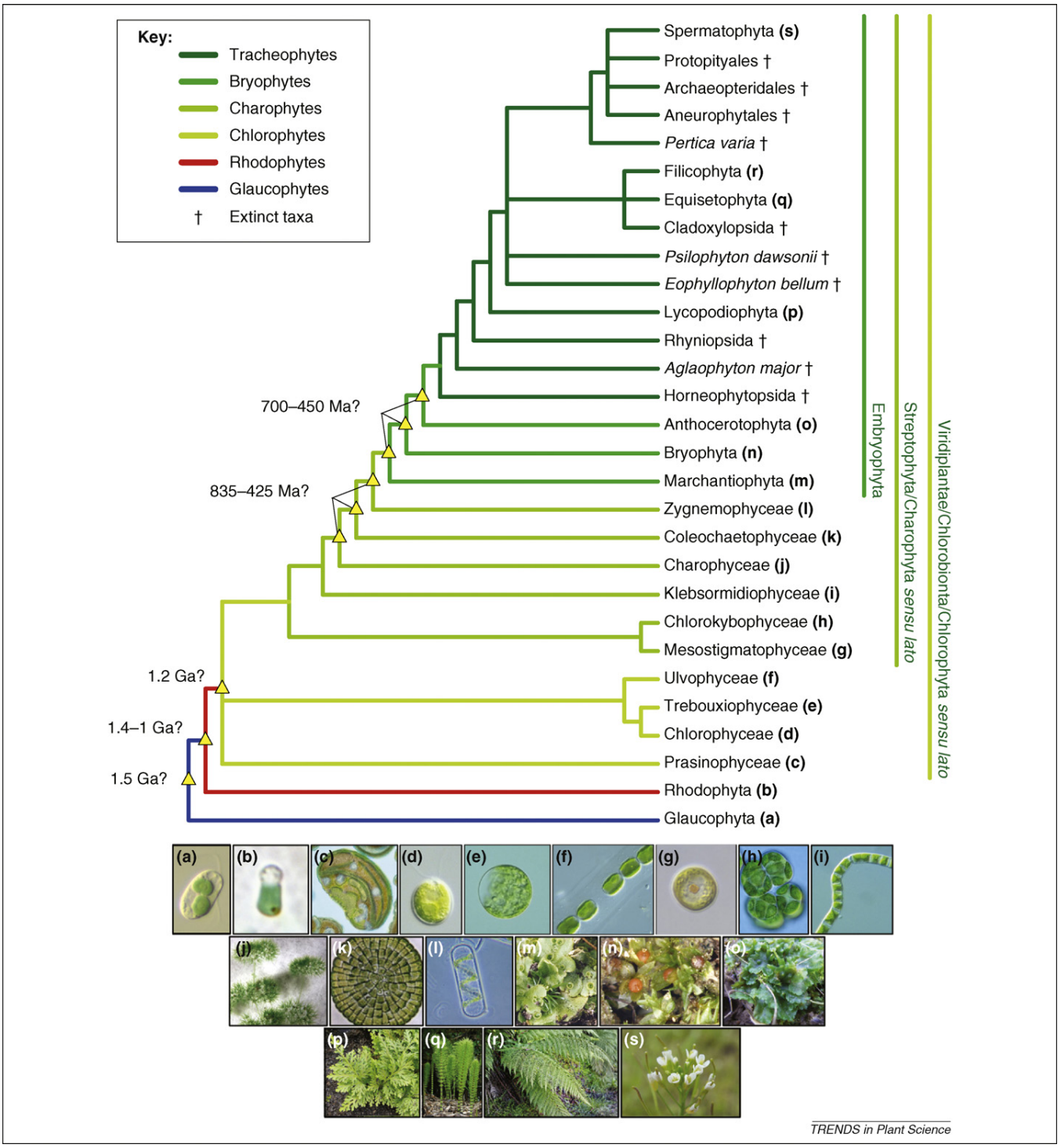

Figure 1. Phylogenetic relationships among Plantae. Cladogram depicting the putative evolutionary relationships among the three lineages of the Plantae supergroup [67]. Divergence times for the nodes discussed in the text are annotated and depicted as yellow triangles (derived from molecular clock estimates found in [18,19] and the fossil record [22,24,68]). The overall topology is based on the cladograms 'Green Plants' (http://www.tolweb.org/Green_plants) and 'Embryophytes' (http://www.tolweb.org/Embryophytes/20582) from the Tree of Life Web Project. Several relationships were subsequently revised according to literature (glaucophyte/red/green [69], Mesostigmatales/Chlorokybales [70], charophytes [5], bryophytes [3]). To better illustrate the organismal diversity of the green linage and to introduce the model organisms used in the comparative genomics and phylogenetic analyses presented here, photographs of representative species from the extant lineages are shown at the bottom: **(a)** *Cyanophora paradoxa*, **(b)** *Cyanidioschyzon merolae*, **(c)** *Ostreococcus sp.*, **(d)** *Chlamydomonas reinhardtii*, **(e)** *Trebouxia sp.*, **(f)** *Geminella mutabilis*, **(g)** *Mesostigma viride*, **(h)** *Chlorokybus sp.*, **(i)** *Klebsormidium accidum*, **(j)** *Nitella tenuissima*, **(k)** *Coleochaete sp.*, **(l)** *Spirogyra sp.*, **(m)** *Marchantia polymorpha*, **(n)** *Physcomitrella patens*, **(o)** *Anthoceros levis*, **(p)** *Selaginella moellendorfii*, **(q)** *Equisetum telmateia*, **(r)** *Dicksonia antarctica* and **(s)** *Arabidopsis thaliana*. Sources for images are provided in supplementary Table S1. Ga = billion years ago.

are still scattered across more than 2000 genomic scaffolds, revealing gaps in genomic sequences.

A threat for large-scale sequencing projects is accidental contamination by sequences from other organisms. Although not overly discussed, contaminations introduced by genomic DNA or plate switches are common phenomena in published genomes. They present a serious complication, especially to comparative genomics approaches.

Table 1. Scaffold and gene model statistics of the *P. patens* genome v1.1

Scaffold statistics	
Scaffolds total	2106
Scaffold sequence total	480 Mb (5.4% gaps)
Scaffold N50	111 scaffolds
Scaffold L50	1.32 Mb
Sequence depth derived from the assembly	8.63x ± 0.10
Largest scaffold	5.39 Mb
Model statistics	**Average**
Gene length (bp)	2389.42
Transcript length (bp)	1195.77
Protein length (aa)	362.84
Exons per gene	4.87
Exon length (bp)	245.62
Intron length (bp)	310.57
Genes per Mb	74.9

The initial *P. patens* genome assembly suffered from both, plate-switch-derived scaffolds and sequences from the genus *Bacillus* for at least 1.3% of the scaffolds. Combining multivariate data mining and wet laboratory analysis, we were able to remove most of the contaminations for the published version 1.1 of the *P. patens* genome [6]. Subsequently, we have removed additional plate-switch contaminants, resulting in v1.2 of the moss genome.

Protein-encoding genes

From v1.1 of the *P. patens* genome ~36 000 protein-encoding genes were predicted [6], whereas clustered EST (expressed sequence tag) data suggested only ~25 000 protein-encoding moss genes [8]. More detailed analysis prompted us to exclude gene models overlapping with transposable elements and tRNA and microRNA (miRNA) genes before release of v1.2 of the *P. patens* genome. From this, we can predict 27 949 protein-encoding *P. patens* genes. Hence, the complexity in terms of raw number of protein-encoding genes in *P. patens* (~30 000 [6]) is in the same range as those of morphologically more complex organisms, such as *Homo sapiens* or *Arabidopsis thaliana* [9]. Table 2 provides a more detailed comparison between sequenced genomes of Plantae (Figure 1).

A comparison of gene structure revealed introns to be larger in *P. patens* than in *A. thaliana*, yet position and amount of introns are approximately the same (Table 1). In *P. patens* the average G/C content of exons is 50%, whereas it is 44% in *A. thaliana* [10]. The average codon bias is the same in both plants, but the distribution pattern is different, with 15% unbiased moss genes [10]. In *P. patens*, highly expressed genes are more compact with shorter introns, reminiscent of the situation in animal genomes and not yet described for plant genomes [11].

Small RNAs and tRNAs

The *P. patens* genome harbours at least 220 genes coding for miRNAs from 99 families [12,13], a complexity also found in seed plants (Table 2). In addition, the entire complement of small RNAs (sRNAs) has been targeted by high-throughput sequencing of samples from three developmental stages, yielding a total of 561 102 sRNA reads representing 214 996 unique sequences. From those, 127 135 (59%) could be mapped to the draft genome sequence [14]. Very often, sRNA genes map to annotations for retrotransposons, a fact that might indicate massive post-transcriptional gene silencing (PTGS) of transposable elements in *P. patens*.

In addition, we found 432 putative genes encoding tRNAs in *P. patens*. From these, 417 tRNAs (97%) code for one of the 20 standard amino acids. In all cases, at least one of the anti-codons per amino acid was found, including a gene for a selenocysteine tRNA, but not all of the possible 57 anti-codon tRNAs seem to be used.

Transposable elements

About half of the ~500 Mb *P. patens* genome consists of long-terminal-repeat retrotransposons (LTR-Rs), and 4795 of them are predicted to be full length. From these, 46% are Gypsy-like and 2% are Copia-like LTR-Rs. Nested regions are very common, with 14% of the LTR elements inserted into other LTR elements [6]. The activity of these transposable elements might be controlled by PTGS via sRNAs (see above). Further analysis of the spliced alignments of ~340 000 ESTs covering the complete live cycle and various treatments revealed that up to 3% of a given library originate from regions annotated as LTR-Rs. In contrast to all other eukaryotic genomes analysed to date, *P. patens* contains only one family of Helitron rolling-circle DNA transposons [6]. Further class II DNA transposable elements have not been analysed so far.

With the current draft we have an initial overview of the genomic structure and a good understanding of the comp-

Table 2. Overview of published photosynthetic eukaryote genomes

	Nuclear genome								Plastid genome		Mitochondrial genome	
	Genome size (1C) (Mb)	Chromosomes (1C)	Ploidy (x)	GC (%)	Protein-coding genes	miRNAs[a]	miRNA families[a]	Repetitive/ transposable elements[b]	Genome size (kb)	Encoded proteins	Genome size (kb)	Encoded proteins
Populus trichocarpa	485	19	2	36.7	45 555	215	33	37%	157	101	803	52
Arabidopsis thaliana	157	5	2	36.0	26 819	184	109	10%	155	87	367	117
Oryza sativa	490	12	2	43.6	~32 000	243	62	30%	135	64	492	54
Physcomitrella patens	511	27	1	38.7	~30 000	220	99	48%	123	85	105	42
Chlamydomonas reinhardtii	120	17	1	64.0	15 143	49	47	11%	204	69	15	8
Ostreococcus tauri	12	20	1	58.0	7618	N.A.	N.A.	N.A.	72	86	44	65
Phaeodactylum tricornutum	26	33	2	48.5	10 010	N.A.	N.A.	N.A.	117	130	N.A.	N.A.
Thalassiosira pseudonana	32	24	2	47.0	11 390	N.A.	N.A.	2%	129	127	44	40
Cyanidioschyzon merolae	16	20	1	55.0	5014	N.A.	N.A.	6%	150	207	32	34

Abbreviations: GC, GC content; kb, kilo base pair; N.A., not available.
[a]Data from the miRBase registry, release 10.0, http://microrna.sanger.ac.uk/.
[b]Percentage of the genome.

lement of protein-coding and non-protein-coding regions of a haploid moss genome. The internet resource http://www.cosmoss.org provides access to the past, current (v1.2) and future versions of the *P. patens* genome and annotation using BLAST, sequence retrieval and an integrated genome browser. The assembly has been ridded of further contaminants and false predictions originating from non-protein-coding regions: for example, miRNA precursors, tRNAs and transposable elements have been isolated from the set of protein coding genes.

Towards the v2.0 genome

The JGI produced a wealth of sequence data by a whole-genome shotgun approach. In addition, the International *Physcomitrella* Genome Consortium aimed at producing additional data for high-quality genome annotation. For example, a total of ~250 000 ESTs covering the complete life cycle [6,15,16] aided the accurate prediction of nearly 13 000 gene models in v1.1. Currently, ~100 000 additional ESTs and, even more importantly, 20 000 full-length cDNAs are being used to generate improved training datasets that will enable even more accurate gene prediction. Furthermore, end sequences from *P. patens* genomic clones in bacterial artificial chromosomes (BACs) will be used to decrease the current high amount of scaffolds by super-scaffolding. In addition, a physically linked genetic map of *P. patens* is being produced using three different marker technologies, that is, single sequence repeats (SSRs, [17]), amplified fragment length polymorphisms (AFLPs) and ecotype-derived single-nucleotide polymorphisms (SNPs). Taken together, these efforts should result in an improved version of the *P. patens* genomic sequence where 27 linkage-group-derived pseudomolecules represent the golden path of the expected 27 chromosomes. Through expansion of the functional genomics toolbox of the model plant *P. patens* by a complete genetic map and assembled chromosomes, gene identification after forward genetics will become feasible and will increase *P. patens*' usefulness as an evolutionary model plant.

Early land plant evolution

The colonization of land by plants was an important step in the history of life. Approximately 1400 million years [18,19] of Viridiplantae (green algae and land plants, Figure 1) evolution and diversification shaped the biosphere and laid the foundation for all extant terrestrial ecosystems. Based on spore microfossils dated to the Mid-Ordovician (~475 megaannum [Ma]) to Early Silurian (~440 Ma) periods, the water-to-land transition of multicellular plant ancestors to land plants (embryophytes) took place at least 475 Ma [20–22]. These first land plants had to adapt to the harsh conditions of terrestrial existence, such as (UV) radiation, gravity, flooding and desiccation cycles, wind and other extremes in weather and nutrient supply. Adaptations encompassed substantial changes in morphology and cellular, physiological and regulatory processes, leading to enhanced osmoregulation and osmoprotection, desiccation and freezing tolerance, heat resistance, synthesis and accumulation of protective 'sunscreens' and enhanced DNA repair mechanisms [20,21,23].

As an evolutionary relict of their last common ancestor (LCA), all extant land plants exhibit an alternation of multicellular generations, that is, the sexual, haploid gametophyte and the asexual, diploid sporophyte. Whereas in Lower Devonian (~400 Ma; e.g. *Aglaophyton major* in Figure 1) land plant fossils the gametophytic and sporophytic generations share approximately equal morphological complexity [24], extant bryophytes (Figure 1, paraphyletic group comprising mosses, hornworts and liverworts) and vascular plants adopted different strategies for their diversification during their colonization of the land mass. The gametophyte was heavily reduced in vascular plants to gain independence from water for sexual reproduction [20,24], but sexual reproduction in bryophytes is dependent on water, which ensures the motility of their biflagellate sperm cells, and the sporophytes are at least in part nutritionally dependent on the gametophyte.

The affiliation of ultrastructures in spore-containing plant fragments from Late Ordovician (~450 Ma) rocks with liverwort spore walls [22] brought further evidence that some of the widely distributed microfossils in Mid-Ordovician sediments might represent liverwort spores. The earliest unequivocal liverwort megafossil is dated to the Middle Devonian (~380 Ma) [25]. There are also fossils from the Lower and Upper Carboniferous with some moss affinities, and the first unambiguous fossil record of mosses is dated to the Lower Permian (~270 Ma). Although the bryophytic fossil record of this pivotal period is scarce, it has been argued that the origin should date back farther than 450 Ma and that the diversification of bryophytes took place sometime between the Upper Ordovician and the Silurian periods [20]. Moreover, some molecular clock analyses date the separation of vascular plants and mosses to as long as ~700 Ma [26]. Therefore, it seems safe to say that the LCA of mosses and vascular plants was present at least 450 Ma and that the initial embryophyte divergences [3], as shown in Figure 1, separated the Marchantiophyta (liverworts), Bryophyta (mosses) and Anthocerotophyta (hornworts) from the remainder of the land plants, the vascular plants (Tracheophyta). However, the branching order of bryophytes and their sister relation to tracheophytes are still subjects of debate [20].

Acquisition of morphological complexity

Some genes are found in unicellular aquatic algae but not in land plants. They encode for proteins such as flagellar components necessary for motility of gametes and for cytoplasmic dynein with the regulatory dynactin complex, suggesting that the dynein-mediated transport system was lost during the conquest of land [6]. By contrast, a general expansion of gene families associated with the acquisition of multicellularity and morphological complexity on land can be observed [27]. These include, for example, signalling pathways for the phytohormones abscisic acid, auxin and cytokinin, as well as more elaborate transport capabilities [28,29]. *P. patens* relies on proteins acquired via endosymbiotic gene transfer for chloroplast maintenance and division [30,31] and has expanded their number to generate unique features such as an FtsZ (filamentous temperature-sensitive Z)-based plastoskeleton [32].

During colonization of the land mass, bryophytes and tracheophytes evolved and expanded with different strategies. While bryophytes developed a dominant sexual gametophyte that still combined sexual reproduction with the availability of free water, vascular plants, with their dominant sporophyte, became more independent of water in their sexual reproduction. The poikilohydric mosses (organisms without the ability to prevent desiccation) have significantly expanded two-component signal transduction systems (phosphorelays), efficient homology-based DNA repair and adaptation to shade/high light and de-/rehydration cycles [6]. Furthermore, their metabolism is uniquely complex due to redundancy and enzymatic pathways not found in seed plants [6,15,33]. By contrast, vascular plants lost vegetative dehydration tolerance and motile gametes but acquired signalling pathways for the phytohormones brassinosteriods, ethylene, gibberellic acid and jasmonic acid from pre-formed genetic modules [34–37]. Likewise, seed plants used already existing genes to develop flowers, including evolution of the floral regulator LEAFY [38] and of the MADS-box family of transcription factors [39–43]. Surprisingly, transcription factors that control root hair development in *A. thaliana* control rhizoid development in *P. patens*, revealing a basic regulatory toolbox in the LCA of bryophytes and angiosperms [44].

A phylogenomics approach on transcription-associated proteins (i.e. transcriptional regulators and transcription factors; TAPs) comparing *P. patens*, the diatom *Thalassiosira pseudonana*, the red alga *Cyanidioschyzon merolae*, the green alga *Chlamydomonas reinhardtii*, *Oryza sativa* and *A. thaliana* revealed that *P. patens* contains most of the TAPs found in seed plants. The transcription factor gene families that are absent in the moss are all of small size and have specialized functions in flower development and in lateral organ formation and thus might have emerged during the divergence of tracheophytes. On average, TAP gene families are two to three times larger in *P. patens* than in algae, and in angiosperms transcriptional regulators show an approximately fourfold increase and transcription factor families a ninefold increase as compared to algae [27], which correlates nicely with the increase in morphological complexity observed with the evolution from unicellular algae to multicellular land plants.

Duplication history and gene family evolution

Analyses of the *P. patens* virtual transcriptome [15], a huge collection of assembled ESTs, revealed that this moss, although a true haploid in genetic crosses, has undergone at least one whole-genome duplication in its evolutionary history [45]. This interpretation of *P. patens* being a paleopolyploid species was supported by analyses of the gene models predicted from v1.1 of the genome sequence [6]. In addition, the molecular analyses also indicate a duplication history that is significantly younger than 450 Ma, and thus might not be related to the conquest of land by the LCA of mosses and seed plants but to further diversification and speciation of the Funariaceae [45,46] (Figure 2).

Besides such large-scale duplications, smaller portions of the genome were duplicated in most plant genomes. One example of this is tandemly arrayed genes (TAGs), which are paralogous genes residing closely together on a chromosome. Such TAGs amount to more than 15% of all genes in the *A. thaliana* and *O. sativa* genomes [47]. By contrast, TAGs in *P. patens* amount to only 1% of the genome. Another unique feature is that TAGs in moss are

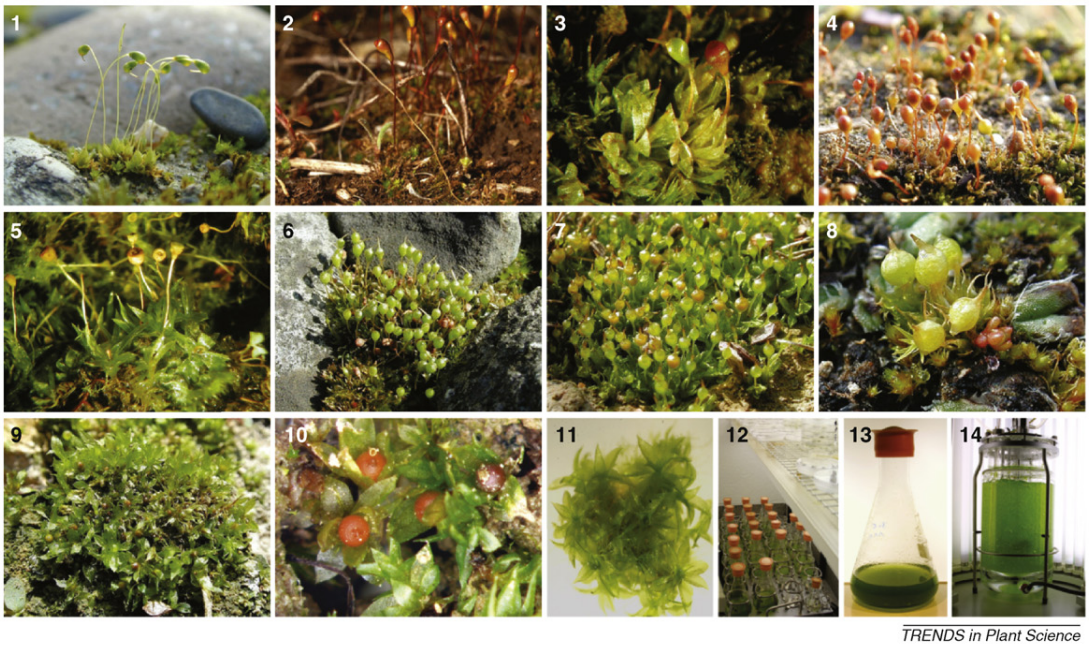

Figure 2. *Physcomitrella* and its taxonomic context within the *Funariaceae*. Nine selected *Funariaceae* found in Southern Germany are shown. Numbering left to right. Upper row: (1) *Funaria hygrometrica*, (2) *Funaria muhlenbergii*, (3) *Entosthodon fascicularis*, (4) *Entosthodon obtusus*. Middle row: (5) *Physcomitrium eurystomum*, (6) *Physcomitrium pyriforme*, (7) *Physcomitrium sphaericum*, (8) *Pyramidula tetragona*. Lower row: *Physcomitrella patens* in the wild (9–10) and in the laboratory (11–14; culture on plates, flasks and bioreactor [55]). Images: 1–9 Michael Lüth; 10 Mark von Stackelberg; 11–14 Anja Martin, Stefan Rensing and other members of Chair Plant Biotechnology.

highly conserved even in noncoding regions and that they are predominantly in inverse orientation [6]. This pattern suggests a concerted evolution by gene conversion and might be related to the exceptionally high reliance on sequence similarity for DNA repair observed in *P. patens* [48,49].

Prominent examples for this observation are the proteins of the light harvesting complex (LHC) superfamily, which are significantly enriched among TAGs. Likewise, early light-inducible proteins (ELIPs) are dramatically expanded in *P. patens* as compared to seed plants. These redundancies of highly abundant proteins might point to a unique robustness of the photosynthetic antenna and to efficient mechanisms to cope with photo-oxidative damage [6].

P. patens is a paleopolyploid, that is, the genome underwent at least one large-scale duplication event, which might be related to speciation. Tandem-duplicates found in the moss' genome deviate significantly from those found in di- and polyploid seed plants in terms of orientation and conservation. This pattern might be related to the moss' high rate of DNA repair by homologous recombination and be indicative of concerted evolution of these loci via gene conversion to allow the haploid organism to maintain 'pseudoalleles' of highly expressed genes.

Future avenues

Comparative genomics and phylogenomics analyses including algae, bryophytes, club mosses and seed plants will enhance our understanding of plant evolution [50]. With the *P. patens* genome at hand we can now learn how such a huge haploid genome retained its integrity over millions of years and which special molecular features are involved in its efficient homology-dependent DNA repair [48,51]. Transfer of such features to crop plants might have tremendous implications for plant biotechnology [52]. The simplicity and ease of handling of the tissue, combined with the fully sequenced genome, make this plant a prime candidate for systems biology approaches [53,54]. Because *P. patens* can be grown in bioreactors (Figure 2), it is already in use as a safe and cost-effective production system for complex biopharmaceuticals [55]. Furthermore, this moss is used in synthetic biology, a newly emerging field in life sciences that aims at understanding and verifying signalling pathways by genetic reconstruction in evolutionary distant organisms, as exemplified in Ref. [56]. Forward genetics in *P. patens* will now be targeted as in *A. thaliana*, making gene-identification from huge tagged mutant collections [57] less time-consuming. All these approaches will benefit from advanced microarray [58] and proteomic tools [59], which both heavily rely on a complete genome sequence [60]. In addition, an International Moss Stock Center (IMSC; http://www.moss-stock-center.org, currently under construction) is being established to store published moss mutants and to collect and to distribute *P. patens* ecotypes (http://www.cosmoss.org/ecomap.content). These will help to unravel new gene functions by analysis of genetic variation in natural populations, as has been applied successfully to *A. thaliana* [61,62]. With the *P. patens* genome as a reference, the application of ultra-high-throughput (re-)sequencing technologies (e.g. 454, Solexa, SOLiD) to different members of the Funariaceae (Figure 2) would allow investigation of, for example, the genetic factors determining the morphological diversity observed in the family's sporophytes, ranging from long setae and complex peristomes (*F. hygrometrica*) to reduced or complete absence of peristomes and setae (seen to varying extents in *Physcomitrium*, *Pyramidula tetragona* and *Physcomitrella patens*). Ecological diversity could be assessed by the application of the above-mentioned techniques to the available *Physcomitrella* ecotype collection.

Although *P. patens* is a phylogenetic link between algae and 'higher plants', conclusions should not be over-simplified. Extant bryophytes are not frozen in time. Instead, they have had a separate evolutionary history of at least 450 Ma since the early embryophyte divergence. Therefore, improved genomic taxon sampling is crucial for a better coverage of morphological and ecological diversity among the green tree of life (Figure 1). Two major steps in this direction are the full genome sequences of the club moss *Selaginella moellendorffii*, which will be published soon, and of the liverwort *Marchantia polymorpha*, which is scheduled for sequencing by the JGI in 2008. Both genomes will greatly improve taxonomic resolution of the embryophyte tree of life (Figure 1) for comparative inference and are ideally positioned to bridge the evolutionary gaps from *P. patens* to green algae (*M. polymorpha*) and seed plants (*S. moellendorffii*). Yet, although it is not as morphologically reduced as *P. patens, M. polymorpha* has many derived features when compared to the putative liverwort-like initial land plants [20]. Therefore, additional bryophyte genomes are necessary, and the community needs coverage of other parts of the green lineage as well, such as basal angiosperms, gymnosperms, ferns and charophyte algae [63]. However, because of the huge genome sizes of ferns and gymnosperms, sequencing further bryophyte genomes, where lower coverage [64,65] is feasible due to availability of the *P. patens* and *M. polymorpha* genomes, will have priority.

Acknowledgements

The authors are indebted to all members of the *Physcomitrella* genome consortium (http://www.mossgenome.org) for their support and cooperation within the community. We thank Michael Lüth, Mark von Stackelberg and Anja Martin for moss photographs and Erika Lang and Anne Katrin Prowse for proof-reading of the manuscript. Work in the laboratory is financed by Deutsche Forschungsgemeinschaft (RE 837/10), Bundesministerium für Bildung und Forschung (BioChancePlus-3: 0313852C; GABI-PRECISE: 0315057B; FRISYS: 0313921), Landesstiftung Baden-Württemberg (P-LS-RNS/40), German-Israeli Foundation (research grant: I-832–130.12/2004) and the Excellence Initiative of the German Federal and State Governments (EXC 294). We thank two anonymous reviewers for helpful comments and suggestions.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tplants.2008.07.002.

References

1 Terasawa, K. *et al.* (2007) The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Mol. Biol. Evol.* 24, 699–709

2 Sugiura, C. *et al.* (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. *Nucleic Acids Res.* 31, 5324–5331
3 Qiu, Y.L. *et al.* (2006) The deepest divergences in land plants inferred from phylogenomic evidence. *Proc. Natl. Acad. Sci. U. S. A.* 103, 15511–15516
4 Turmel, M. *et al.* (2007) An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. *BMC Genomics* 8, 137
5 Turmel, M. *et al.* (2006) The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Mol. Biol. Evol.* 23, 1324–1338
6 Rensing, S.A. *et al.* (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. *Science* 319, 64–69
7 Schween, G. *et al.* (2003) Unique tissue-specific cell cycle in *Physcomitrella*. *Plant Biol.* 5, 50–58
8 Rensing, S.A. *et al.* (2002) Moss transcriptome and beyond. *Trends Plant Sci.* 7, 535–538
9 Sterck, L. *et al.* (2007) How many genes are there in plants (. . . and why are they there)? *Curr. Opin. Plant Biol.* 10, 199–203
10 Rensing, S.A. *et al.* (2005) Protein encoding genes in an ancient plant: analysis of codon usage, retained genes and splice sites in a moss, *Physcomitrella patens*. *BMC Genomics* 6, 43
11 Stenoien, H.K. (2007) Compact genes are highly expressed in the moss *Physcomitrella patens*. *J. Evol. Biol.* 20, 1223–1229
12 Axtell, M.J. *et al.* (2007) Common functions for diverse small RNAs of land plants. *Plant Cell* 19, 1750–1769
13 Fattash, I. *et al.* (2007) Evidence for the rapid expansion of microRNA-mediated regulation in early land plant evolution. *BMC Plant Biol.* 7, 13
14 Axtell, M.J. *et al.* (2006) A two-hit trigger for siRNA biogenesis in plants. *Cell* 127, 565–577
15 Lang, D. *et al.* (2005) Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism among mosses. *Plant Biol.* 7, 228–237
16 Nishiyama, T. *et al.* (2003) Comparative genomics of *Physcomitrella patens* gametophytic transcriptome and *Arabidopsis thaliana*: implication for land plant evolution. *Proc. Natl. Acad. Sci. U. S. A.* 100, 8007–8012
17 von Stackelberg, M. *et al.* (2006) Identification of genic moss SSR markers and a comparative analysis of twenty-four algal and plant gene indices reveal species-specific rather than group-specific characteristics of microsatellites. *BMC Plant Biol.* 6, 9
18 Yoon, H.S. *et al.* (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21, 809–818
19 Zimmer, A. *et al.* (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol. Genet. Genomics* 278, 393–402
20 Renzaglia, K.S. *et al.* (2007) Bryophyte phylogeny: advancing the molecular and morphological frontiers. *The Bryologist* 110, 179–213
21 Waters, E.R. (2003) Molecular adaptation and the origin of land plants. *Mol. Phylogenet. Evol.* 29, 456–463
22 Wellman, C.H. *et al.* (2003) Fragments of the earliest land plants. *Nature* 425, 282–285
23 Floyd, S.K. and Bowman, J.L. (2007) The ancestral developmental tool kit of land plants. *Int. J. Plant Sci.* 168, 1–35
24 Taylor, T.N. *et al.* (2005) Life history biology of early land plants: deciphering the gametophyte phase. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5892–5897
25 Hernick, L.V. *et al.* (2008) Earth's oldest liverworts – *Metzgeriothallus sharonae* sp. nov. from the Middle Devonian (Givetian) of eastern New York, USA. *Rev. Palaeobot. Palynol.* 148, 154–162
26 Hedges, S.B. *et al.* (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4, 2
27 Richardt, S. *et al.* (2007) PlanTAPDB: a phylogeny-based resource of plant transcription associated proteins. *Plant Physiol.* 143, 1452–1466
28 Fujita, T. *et al.* (2008) Convergent evolution of shoots in land plants: lack of auxin polar transport in moss shoots. *Evol. Dev.* 10, 176–186
29 Lienard, D. *et al.* (2008) Water transport by aquaporins in the extant plant *Physcomitrella patens*. *Plant Physiol.* 146, 1207–1218
30 Machida, M. *et al.* (2006) Genes for the peptidoglycan synthesis pathway are essential for chloroplast division in moss. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6753–6758
31 Suppanz, I. *et al.* (2007) An integrated physiological and genetic approach to the dynamics of FtsZ targeting and organisation in a moss, *Physcomitrella patens*. *Protoplasma* 232, 1–9
32 Gremillon, L. *et al.* (2007) Filamentous temperature-sensitive Z (FtsZ) isoforms specifically interact in the chloroplasts and in the cytosol of *Physcomitrella patens*. *New Phytol.* 176, 299–310
33 Kopriva, S. *et al.* (2007) The putative moss 3′-phosphoadenosine-5′-phosphosulfate reductase is a novel form of adenosine-5′-phosphosulfate reductase without an iron-sulfur cluster. *J. Biol. Chem.* 282, 22930–22938
34 Hirano, K. *et al.* (2007) The GID1-mediated gibberellin perception mechanism is conserved in the lycophyte *Selaginella moellendorffii* but not in the bryophyte *Physcomitrella patens*. *Plant Cell* 19, 3058–3079
35 Reski, R. (2006) Small molecules on the move: homeostasis, crosstalk, and molecular action of phytohormones. *Plant Biol. (Stuttg.)* 8, 277–280
36 Vandenbussche, F. *et al.* (2007) Evolutionary conservation of plant gibberellin signalling pathway components. *BMC Plant Biol.* 7, 65
37 Yasumura, Y. *et al.* (2007) Step-by-step acquisition of the gibberellin-DELLA growth-regulatory mechanism during land-plant evolution. *Curr. Biol.* 17, 1225–1230
38 Maizel, A. *et al.* (2005) The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260–263
39 Benlloch, R. *et al.* (2007) Floral initiation and inflorescence architecture: a comparative view. *Ann. Bot. (Lond.)* 100, 659–676
40 De Bodt, S. *et al.* (2003) And then there were many: MADS goes genomic. *Trends Plant Sci.* 8, 475–483
41 Riese, M. *et al.* (2005) Isolation and characterization of new MIKC*-type MADS-box genes from the moss *Physcomitrella patens*. *Plant Biol.* 7, 307–314
42 Singer, S.D. *et al.* (2007) Clues about the ancestral roles of plant MADS-box genes from a functional analysis of moss homologues. *Plant Cell Rep.* 26, 1155–1169
43 Verelst, W. *et al.* (2007) MIKC* MADS-protein complexes bind motifs enriched in the proximal region of late pollen-specific *Arabidopsis* promoters. *Plant Physiol.* 143, 447–460
44 Menand, B. *et al.* (2007) An ancient mechanism controls the development of cells with a rooting function in land plants. *Science* 316, 1477–1480
45 Rensing, S.A. *et al.* (2007) An ancient genome duplication contributed to the abundance of metabolic genes in the moss *Physcomitrella patens*. *BMC Evol. Biol.* 7, 130
46 Newton, A.E. *et al.* (2006) Dating the diversification of the pleurocarpous mosses. In *Pleurocarpous Mosses: Systematics and Evolution* (Tangney, N., ed.), Systematics Association
47 Rizzon, C. *et al.* (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. *PLOS Comput. Biol.* 2, e115
48 Kamisugi, Y. *et al.* (2006) The mechanism of gene targeting in *Physcomitrella patens*: homologous recombination, concatenation and multiple integration. *Nucleic Acids Res.* 34, 6205–6214
49 Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *J. Exp. Bot.* 56, 1–14
50 Nishiyama, T. (2007) Evolutionary developmental biology of nonflowering land plants. *Int. J. Plant Sci.* 168, 37–47
51 Markmann-Mulisch, U. *et al.* (2007) Differential requirements for RAD51 in *Physcomitrella patens* and *Arabidopsis thaliana* development and DNA damage repair. *Plant Cell* 19, 3080–3089
52 Hohe, A. and Reski, R. (2003) A tool for understanding homologous recombination in plants. *Plant Cell Rep.* 21, 1135–1142
53 Cove, D. *et al.* (2006) Mosses as model systems for the study of metabolism and development. *Annu. Rev. Plant Biol.* 57, 497–520
54 Decker, E.L. *et al.* (2006) Moss systems biology en route: phytohormones in *Physcomitrella* development. *Plant Biol.* 8, 397–405
55 Decker, E.L. and Reski, R. (2008) Current achievements in the production of complex biopharmaceuticals with moss bioreactors. *Bioprocess Biosyst. Eng.* 31, 3–9

56 Khandelwal, A. *et al.* (2007) Moonlighting activity of presenilin in plants is independent of γ-secretase and evolutionarily conserved. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13337
57 Schween, G. *et al.* (2005) Large-scale analysis of 73 329 *Physcomitrella* plants transformed with different gene disruption libraries: production parameters and mutant phenotypes. *Plant Biol.* 7, 228–237
58 Cuming, A.C. *et al.* (2007) Microarray analysis of transcriptional responses to abscisic acid and osmotic, salt, and drought stress in the moss, *Physcomitrella patens*. *New Phytol.* 176, 275–287
59 Heintz, D. *et al.* (2006) Rapid alteration of the phosphoproteome in the moss *Physcomitrella patens* after cytokinin treatment. *J. Proteome Res.* 5, 2283–2293
60 Quatrano, R.S. *et al.* (2007) *Physcomitrella patens*: mosses enter the genomic age. *Curr. Opin. Plant Biol.* 10, 182–189
61 Kim, S. *et al.* (2007) Recombination and linkage disequilibrium in *Arabidopsis thaliana*. *Nat. Genet.* 39, 1151–1155
62 Clark, R.M. *et al.* (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* 317, 338–342
63 Jackson, S. *et al.* (2006) Comparative sequencing of plant genomes: choices to make. *Plant Cell* 18, 1100–1104
64 Green, P. (2007) 2x genomes – does depth matter? *Genome Res.* 17, 1547–1549
65 Katari, M.S. *et al.* (2005) Comparing low coverage random shotgun sequence data from *Brassica oleracea* and *Oryza sativa* genome sequence for their ability to add to the annotation of *Arabidopsis thaliana*. *Genome Res.* 15, 496–504
66 Havecker, E.R. *et al.* (2004) The diversity of LTR retrotransposons. *Genome Biol.* 5, 225
67 Adl, S.M. *et al.* (2005) The new higher level classification of eukaryotes with emphasis on the taxonomy of protests. *J. Eukaryot. Microbiol.* 52, 399–451
68 Kelman, R. (2004) Charophyte algae from the Rhynie chert. *Trans. R. Soc. Edinburgh, Earth Sci.* 94, 445–455
69 Reyes-Prieto, A. and Bhattacharya, D. (2007) Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within Plantae. *Mol. Biol. Evol.* 24, 2358–2361
70 Lemieux, C. *et al.* (2007) A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biol.* 5, 2