Abstract

For about a century, spanning the eras of early genetics to state-of-the-art biotechnology, the moss *Physcomitrella patens* has been a popular object of biological research. Meanwhile it has become an established model organism in plant evolutionary and developmental biology, mainly due to a combination of two factors: its phylogenetic key position in the plant tree of life and the sum of its favorable biological features. As a member of an early diverging land plant lineage – the bryophytes – *Physcomitrella* fills the gap between other models of the green lineage such as aquatic algae and flowering plants. The advantages of small stature and short generation cycles, accompanied by established and reliable cultivation techniques provide researchers with a robust, relatively fast, and easy cultivation for experiments in a laboratory environment. Precise genome engineering is enabled by the moss’s haploid-dominant lifestyle and its specifically high rate of homologous recombination during DNA repair, that is routinely utilized through an extensive molecular toolkit for efficient gene targeting since 1998. *Physcomitrella*’s genome was sequenced about a decade ago, making it the first bryophyte and even one of the first plants to be chosen for such a whole-genome shotgun sequencing approach. Ever since, the annotation of this “flagship genome” has been subject to constant improvement by an active community through the internet resource cosmoss.org which provides a central platform for knowledge exchange as well as bioinformatics data and tools.
Key Concepts

The moss *Physcomitrella patens* is an established experimental model in plant evolutionary and developmental biology: important because of its basal position in the plant tree of life.

The dominant haploid generation and the relative ease of cultivation have strongly contributed to its popularity in modern biology.

Large-scale mutant collections are available and together with the genetic map allow forward genetics approaches.

Efficient gene targeting and reverse genetics in the moss are enabled by its high rate of homologous recombination allowing precise genome engineering.

The *Physcomitrella patens* genome was the first sequenced bryophyte genome and reveals insights into land plant evolution.
Introduction

Historically, because of its haploid-dominant generation, few tissue types traceable to the fate of a single cell, as well as its small stature and straightforward cultural needs, *Physcomitrella patens* (Hedw.) Bruch & Schimp. has been the focus of genetic studies for about 100 years. Over this time, it has developed into a plant experimental model in evolutionary developmental plant biology: important because it is one of the only models that resides in a basal position in the phylogeny of land plants and as such is more closely derived from an early land plant than present experimental models, bridging the gap between green algae and flowering plants. An established laboratory strain, complemented by standardized culture protocols and an extensive molecular toolkit for genome engineering further expanded the impact of this moss on plant biology. To date, the sequences of the nuclear, plastid, and mitochondrial genomes have revealed striking insights into plant evolution, enlightening the necessary genetic adaptations to life on land.

This chapter aims at introducing *Physcomitrella patens* to a wider audience, while focussing on its genetic and genomic features and providing an entry point into further, more detailed literature. Therefore, it provides an overview of the moss’s development as the subject of early genetic inquiries and the progress toward its significance as a flagship plant experimental model in state-of-the-art plant research. When applicable, example studies and reviews are being addressed to offer a proper introduction into the practical scientific work on *Physcomitrella*.

The Nature of *Physcomitrella patens*

In comparison to other prominent plant experimental models, the small moss species *Physcomitrella patens* might appear relatively unimposing. It is its size, short life cycle, and, not at least, its phylogenetic position that makes it not only a good experimental model but also a highly interesting species in the context of plant evolution.

Life Cycle

The life cycle of *Physcomitrella patens* exhibits a heteromorphic alternation of generations with a dominant haploid gametophyte and a morphologically distinct, reduced diploid sporophyte, typical of all bryophytes. A germinating meiospore gives rise to filamentous protonemata by apical tip growth producing two distinct cell types representing the juvenile gametophyte: the chloroplast-enriched chloronema cells with cell walls perpendicular to the growth axis and caulonema cells with fewer chloroplasts and oblique cell walls. Caulonemal cells produce meristematic buds with three-faced apical cells that develop into the adult gametophyte (gametophore) with basal rhizoids and an erect, leafy stem that is capable of generating the sex organs (gametangia). As *Physcomitrella patens* is monoecious,
the male antheridium and the female archegonium are located on the same gametophore, and self-fertilization is common. In the presence of water, motile male spermatozoids are able to reach and fertilize the egg cell inside the archegonium, which initiates the development of the diploid sporophytic embryo. The diploid sporophyte develops a short seta with an apical spore capsule bearing the characteristic calyptra. Inside the capsule, haploid spores are produced through meiosis and are released subsequent to the ripening and break-open of the capsule.

The male antheridium and the female archegonium are located on the same gametophore, and self-fertilization is common. In the presence of water, motile male spermatozoids are able to reach and fertilize the egg cell inside the archegonium, which initiates the development of the diploid sporophytic embryo. The sporophyte develops a short seta with an apical spore capsule bearing the characteristic calyptra and loses its initial ability to photosynthesize during its maturation. During the entire process, the sporophyte is strongly dependent on nutritional supplies from the gametophyte. Inside the spore capsule, a few thousands of haploid spores are produced through meiosis and are released subsequent to the ripening and opening of the capsule (Fig. 1).

**Phylogenetic Position**

One of the most crucial events in plant evolution is marked by the conquest of land and the rise of embryophytes, demanding substantial morphological and physiological adaptations to the new environment. The last common ancestor of the two land plant lineages of tracheophytes (vascular plants) and bryophytes (liverworts, mosses, hornworts) lived about 500 million years ago (Lang et al. 2010). This early divergence describes two radically different approaches to life on land. At the same time, it anchors the bryophyte *Physcomitrella patens*, as a member of the Funariaceae...
family inside the class Bryopsida, at an important basal position in the context of land plant evolution. The species represents those plants that bridge the gap between aquatic green algae, such as *Chlamydomonas reinhardtii*, and the tracheophytes, for example, *Selaginella moellendorfii* or even *Arabidopsis thaliana*. As such a “basal” land plant species, *Physcomitrella patens* provides an excellent reference to study the necessary adaptations to early life on land. The stronger effects of gravity (as the plant grew in stature), exposure to higher levels of UV irradiation, varying temperatures, and last but not least a fluctuating availability of water all defined the adaptations to the terrestrial lifestyle. In addition, the basal position of *Physcomitrella patens* in the land plant phylogeny enables comparative analyses with more complex species within the plant kingdom that can assist the identification of genes and gene families that are either conserved or emerged throughout the individual course of evolution, e.g., for specific adaptation to an environmental selective force.

The Funariaceae Family and *Physcomitrella patens* Ecotypes

*Physcomitrella patens* (Hedw.) Bruch & Schimp. belongs to the cosmopolitan Funariaceae family within the order Funariales of the class Bryopsida. The Funariaceae comprise approximately 250–400 species accommodated in a minimum of 15 genera of small, terricolous, monoicous, annual, and biennial mosses. However, the taxonomic classification of single genera and species is topic of constant debate in the bryophyte scientific community and will probably profit from future genomic insights. While the family members generally share a rather uniform vegetative, gametophytic body, they differ substantially in terms of sporophyte morphology and ecology. The sporophyte of *Physcomitrella patens* is characterized by a reduced seta and the lack of a peristome on the indehiscent spore capsule, a putative adaption to spore dispersal by water birds. In contrast, the Funariaceae species *Funaria hygrometrica* exhibits an elongated seta and dehiscent capsules with a complex peristome adapted to humidity- and wind-dependent spore dispersal. Interestingly, the sporophyte morphology of *Physcomitrella* is assumed to represent the evolutionary derived state that followed a secondary reduction of the precedent complex structure.

While the Funariaceae exhibit an overall worldwide distribution, *Physcomitrella* species occur mutually exclusively to certain regions. Ecotypes of the two other species in the *Physcomitrella* genus, *Physcomitrella readeri* and *Physcomitrella magdalenae*, were collected in Japan, Australia, and Africa, respectively. However, isolates of *Physcomitrella patens* ecotypes exclusively cover the northern hemisphere, more precisely the USA and Europe, including France, Germany, Norway, Sweden, the United Kingdom, and Ukraine. The most prominent of these ecotypes is represented by the established laboratory strain “Gransden,” that is, by now, the de facto standard in laboratories worldwide. Originally isolated in 1962 by H.L.K. Whitehouse from a single spore in Gransden Wood, Huntingdonshire (UK), the vast majority of modern genetic and molecular biological experiments
that involve *Physcomitrella patens* employ this strain. Highlighting this role, the “Gransden” ecotype was the chosen strain for the *Physcomitrella* genome project. Though reports on chromosome counts differed, cytogenetic analyses indicated the currently accepted chromosome number of \( n = 27 \) (Reski et al. 1994), and flow cytometric analyses determined its size to be about 511 megabasepairs (Schween et al. 2003). A collection of about 40 Funariaceae isolates, including the *Physcomitrella* ecotypes, is cryo-conserved in a long-term storage facility – the International Moss Stock Center (IMSC, http://www.moss-stock-center.org/) at the University of Freiburg, Germany. The IMSC is a community-oriented, nonprofit resource that provides moss researchers with a facility to safely store individual strains by cryopreservation. In addition, moss material can also be ordered by the community to ensure the use of standardized stock material in worldwide scientific applications. In this regard, the IMSC also maintains about 60 transgenic lines that comprise a collection of knockout mutants and reporter lines.

**Features and Genetics**

*Physcomitrella patens* combines a variety of favorable properties, which accounts for its success as a desirable plant model for genetic and developmental studies. As a target for genetic manipulation, it profits from its short generation time, fast growth rate, small stature, and reduced morphological complexity with traceable cell lines. It also has the added advantage of being relatively simple to transform with high efficiency.

**Lab Culture**

In the laboratory environment, protonemal tissue of *Physcomitrella patens* is grown using simple microbiological culture techniques. Under controlled conditions, the moss plants complete their entire life cycle in approximately 4 months. Standard Petri dishes with solidified agar medium containing only inorganic salts can be used to produce dense tissue from a protonemal “solution” within a few days, made possible by the ability of the moss to regenerate a whole plant from almost each single cell. Larger amounts of protonemal tissues can also be grown in liquid medium using small to medium size flasks or with the use of larger bioreactors. Moss bioreactors are typically glass containers with capacities of 5–30 liters with stirring and ventilation mechanisms to ensure an even exposure to light and an optimal air mix (Fig. 2). In high-throughput production protocols, for example, in pharmaceutical applications (Reski et al. 2015), “wave” bioreactors are often employed. Typically, these consist of large plastic bags containing up to 600 liters of medium placed on a shaking device to ensure constant wave-like movement of the medium.
Classical Genetics and Mutagenesis

Bryophyte species in general and mosses in particular have been subjects of genetic research since the early twentieth century, when Mendelian genetics were rediscovered and put into the context of Darwinian evolution. The natural

Fig. 2 Habitus and laboratory culture. (A) Natural *Physcomitrella patens* population covering a clump on a fallow acre near Freiburg, Germany. (B) Close-up of the sporophyte, comprising the short seta that bears the spore capsule with its calyptra. Subfigures C–E illustrate different laboratory culture techniques: gametophytes growing on solid medium in a Petri dish (C), filamentous protonemata in liquid culture D, and protonemata in the moss bioreactor (E).
occurrence of variants and hybrids in the bryophytes attracted the attention of early geneticists, and experiments were facilitated by their short generation time and undemanding culture. Examples of scientific breakthroughs from early bryophyte genetic studies include the discovery of sex chromosomes in plants (Allen 1917), the description of heterochromatin and the continuity of chromosomes during the mitotic cell cycle (Heitz 1928), and the development of X-ray mutagenesis (Knapp 1936).

The haploid nature of the moss gametophyte was determined from the early observation of non-Mendelian inheritance of genes (von Wettstein 1928). The crossing of two haploid strains results in the formation of a diploid sporophyte which in turn generates haploid spores. Examining the progeny of such a cross, two alleles of one gene segregate 1:1 in the F1 generation, and mutant alleles of two unlinked genes segregate in a 1:1:1:1 ratio, corresponding to a 3:1 ratio of mutant to wild type. While these segregation ratios provided the first evidence for the dominance of the haploid genotype in mosses, they also highlighted the clear advantage of mosses for the isolation of mutant strains for genetic investigations: in contrast to crosses in diploid species, mutants can be identified directly without the need for additional inbreeding and the establishment of an F2 generation. Clearly, this only applies to nonlethal phenotypes. However, the study of recessive lethal mutants is still possible by the use of somatic hybridization, i.e., the fusion of two protoplasts and regeneration of diploid gametophytes. Although these hybrids of Physcomitrella patens are self-fertile, they develop a reduced number of sex organs, particularly archegonia, as well as exhibiting “sluggish” spermatozoid movement, which putatively accounts for an overall slower completion of the life cycle. Starting in the 1960s with the work of Engel, a plethora of auxotrophic mutants and developmentally abnormal Physcomitrella patens strains have been generated, e.g., by exposure to X-rays and alkylating agents (Engel 1968) and the subsequent mass screening of spores. The ability of protoplasts (single cells) to regenerate whole plants has led to the recent replacement of mutagenized spores as the starting material for mutant strains. Protoplasts, either directly treated with mutagenic agents or derived from treated protonemal tissue by digestion, are now the preferred source of mutants.

**Transformation**

Over the last three decades, an extensive toolkit has been established for direct transformation, i.e., the delivery of exogenous DNA into cells of Physcomitrella patens. Depending on the retention and inheritance of the transformed DNA and the corresponding phenotype, two basic types of transformation can be distinguished: stable and transient. Transformation with a plasmid containing an antibiotic resistance reporter cassette demonstrates these two possibilities. Under continuous selection pressure, i.e., the growth on antibiotic media, transformed plants express the reporter genes and thus show antibiotic resistance. Removing the antibiotic from the growth medium for a period of time, thus lowering the selection pressure for resistance, generates the two classes of transformed plants. Stable transformants
with genome integration of the transgene inherit the resistance meiotically and produce offspring that encode and express the resistance to the antibiotic. In contrast, plants that lose or do not inherit the resistant phenotype have not undergone genome integration of the transgene, thus are only transiently transformed. Researchers suggested that in the case of lost resistance, the transgene may have been replicated and expressed extrachromosomally when the plant was under selection and the plasmids lost after a period of relaxed conditions (Ashton et al. 2000). The research team transformed *Physcomitrella patens* protoplasts using a previously unused plasmid, pBI426, carrying an antibiotic resistance marker and a β-glucuronidase (GUS) reporter gene. After regeneration of gametophytes and initial selection, they repeatedly transferred the plants between selective and nonselective media for defined periods. A minority of plants (~0.2%) showed stable transformation even after several rounds of relaxed and increased selection pressure; however, the majority lost their ability to grow on selective medium. GUS staining was possible for all stable transformants and initially for tissue of transient transformants in direct contact with the selective medium. In accordance with the observable growth deficiency, transient transformants also lost GUS activity after several rounds of relaxed selection. Southern blot analyses lead to the assumption that these transient transformants temporarily contained 3–40 extrachromosomal copies of the plasmid. This was the first report of a transgenic photosynthetic plant to be generated by the expression of genes from an autonomously replicating extrachromosomal element. Based on these findings and the fact that stable genome integration of transgenes is the desired goal of most experiments, the discrimination of stable transformants from transient ones is mandatory. Typically, this is accomplished by the before mentioned method, i.e., the removal of the selective condition and a subsequent renewed selection period(s).

Transformation methods for *P. patens* have been constantly improved, and to date, a variety of methods are available to researchers (Strotbek et al. 2013). *Physcomitrella patens* was considered not to be transformable via *Agrobacterium tumefaciens* for a long period of time; however, the recent discovery of new virulent strains enabled the application of this method (Cove et al. 2009). *A. tumefaciens* directs the integration of foreign DNA into a plant’s genome via the T-DNA region of its Ti plasmid and is usually a method of choice for transformation of vascular plants. Although the genomic integration of the introduced DNA is nondirected and random via this method, it does exclusively lead to the generation of stable mutants, at high transformation rates, using *Physcomitrella* protoplasts. In addition to this true classical method of plant transformation, it is also possible to deliver foreign DNA directly into cells mechanically using microinjection and biolistics. With a microinjection needle, DNA can be injected into filament cells and even directly into nuclei of *Physcomitrella patens* cells. However, this method demands the special cultivation of moss filaments in absolute darkness for a defined period, to allow cell wall penetration. Additionally, the demanding injection technique has to be applied at a precise region basal to the nucleus to prevent the bursting of cells. So far, only transient mutants have been analyzed, thus reliable evidence of stable transformation is lacking. The biolistic method relies on a particle gun to bombard cells with
plasmid-coated gold or tungsten particles. There are no specific restrictions to the kind of targeted plant tissue, and efficiency is comparable to that seen for vascular plants. Unfortunately, as for microinjection, there have been no reports of stable mutants to date.

Since the early 1990s, Polyethylene glycol (PEG)-mediated transformation of *Physcomitrella patens* protoplasts has been developed into the method of choice (Schaefer et al. 1991). It enables the use of relatively large DNA constructs and at the same time requires no sophisticated technology or instruments. PEG stabilizes the protoplasts and allows for the uptake of negatively charged DNA through the cell membrane. After transformation, protoplasts are transferred to regeneration medium and undergo subsequent selection. In addition to the undemanding requirements, PEG transformation provides both, high transformation efficiency and a low rate of multiple integration of the transgene into the genome. Remarkably, the method can also be employed for the transformation of chloroplasts in *Physcomitrella patens* without any additional requirements (Sugiura and Sugita 2004). Integration of DNA into the plastid genome relies exclusively on homologous recombination and thus offers nearly the same possibilities that are available for nuclear genome editing (see subsequent sections). Of course, this has been facilitated by the availability of the chloroplast genome sequence, sequenced prior to the nuclear genome of *Physcomitrella*. Undoubtedly, the ability to directly target genes encoded in the plastid genome offers an invaluable tool for studies of genes related to photosynthesis, especially in the context of plant evolution.

**Homologous Recombination and Gene Targeting**

One of the most distinctive features of *Physcomitrella patens* is the high degree of homologous recombination (HR) in somatic cells. In all living cells, HR represents a repair mechanism for DNA double-strand breaks that is highly conserved throughout the plant kingdom. Just like other genomic disruptions, double-strand breaks occur naturally in cellular processes such as DNA duplication and are induced by chemical and physical agents. Yet, by undermining chromatin stability, they pose a huge threat to genome integrity and can cause severe mutations or even cell death. While angiosperms rely mainly on the alternative repair mechanism of nonhomologous end joining (NHEJ), which illegitimately recombines DNA strands, *Physcomitrella patens* predominantly employs HR (Schaefer and Zryd 1997). Through this precise mechanism, two double-stranded DNA molecules that share homologous sequences perform a temporary crossover, in which whole stretches of DNA are exchanged, under guidance of specific enzymes, between the strands without the loss of a single nucleotide. As HR relies on a high degree of sequence identity between the two recombining strands, it produces a legitimate recombination of DNA (Fig. 3).

In practice, HR provides a powerful tool for directed delivery of DNA to a predetermined locus in the genome and the study of gene function in the moss through gene targeting (Strepp et al. 1998). The term “gene targeting” describes a technique for modification of a locus using defined sequences homologous to the
target region of a gene or gene element by exploitation of the HR mechanism. As HR is especially efficient in *Physcomitrella patens*, exhibiting efficiencies only known from the microbiological field, gene targeting via HR is a distinctive feature of the moss in comparison to other plant models. Thus, *Physcomitrella* has often been described as the “green yeast.” The integration efficiency for practical applications depends on certain properties of the transgenic construct. In general, DNA is delivered into the protoplast in linear form, after PCR duplication or enzymatic digestion from its original cloning vector, and consists of a gene cassette plus a selection marker. The construct carries two homologous flanking sequences corresponding to the 5' and 3' ends of the target region in the genome: typically the minimal length is 400–500 bp for each, and longer stretches generally increase integration rates. In addition, a uniform length of the two flanking sequences seems to have a positive effect. Successful integration via HR results in the replacement of the target region by the gene cassette. In this context, two different strategies can be distinguished: targeted knockout and knockin of a gene. The targeted replacement of a gene generally leads to the loss of its function, i.e., a knockout of the respective gene. In contrast, the same method is used to replace genomic regions with homologous sequences from a similar gene from another species or to fuse additional sequences, often regulatory, to the target locus. In the former case, the HR-directed gene replacement enables gene complementation studies, and in the latter, it offers an elegant way of fusing sequences encoding reporter proteins, such as the green

**Fig. 3** Simplified schema of homologous recombination for gene targeting. A double-strand DNA (dsDNA) construct with specifically designed flanking regions is introduced into the nucleus where its ends are recognized as double-strand breaks (DSB), and it is introduced to the DSB repair mechanism of the cell. Consequently, the construct is guided to the genomic locus harboring sequence stretches identical to the flanks (a). Single strands at both ends of the construct invade the identical genomic stretches and anneal to the complementary strands (b). After resolution of the complex structure, the transfected sequence is integrated into the genome (c).
fluorescence protein (GFP) or β-glucuronidase (GUS), to the coding region of a gene of interest. The measurement of the activity of the reporter proteins in-turn enables localization and interaction studies of endogenous gene products. The efficacy of such an approach was shown in a recent study of genes responsible for mitotic cell division in *Physcomitrella patens* (Nakaoka et al. 2012). Different candidate genes, e.g., coding for augmin and others, previously described as tubulin associated, were independently knocked down via RNAi (see section “RNAi”). In order to test for the phenotypic consequences, the coding sequences of the two reporter proteins GFP and RFP (red fluorescent protein) were fused to the genes of tubulin and histone, respectively. Confocal microscopic time-lapse photography of the fluorescent reporters in the different knocked-down lines revealed a dominant role for augmin in the generation of microtubules.

**Forward Genetics and the Genetic Map**

For a long time, forward genetics has represented the standard method for linking a function with a specific gene. The approach relies primarily on untargeted mutagenesis via chemicals such as alkylating agents or physical factors such as X-ray and UV irradiation. Subsequent screening of mutant plants reveals interesting phenotypes that are subject to further investigation to identify the responsible genes that have been mutated. For *Physcomitrella patens*, a huge effort was made to create several mutant collections for the genetics community, and from the 1990s, the large-scale mutagenesis of *Physcomitrella* via transformation with tagged constructs involving genomic DNA (gDNA) or complementary DNA (cDNA) was presented as an alternative to chemical and/or physical mutation.

In one of the early efforts, the moss was transformed using a gDNA library in which DNA fragments were tagged via so-called shuttle mutagenesis using mini-transposons (Nishiyama et al. 2000). The researchers extracted genomic DNA, generated 3–6 kb fragments, and then used an established *Escherichia coli* system to integrate mini-transposons that contained an antibiotic resistance gene as well as a gene encoding for GUS into the generated gDNAs. Upon PEG-mediated protoplast transformation with these constructs, plants where selected for antibiotic resistance and phenotypically classified. Of the 5264 stable transformants, 203 mutants had developmental and morphological abnormalities in comparison to the wild type, and 129 mutants had GUS activity, e.g., in mucilage hairs (multicellular hairs that form at the base of a leaf), in young leaves, in the apical cell of very young buds and in protonema including chloronema and caulonema, respectively.

Subsequently, a similar approach yielded a total of 16,203 transformants (Egener et al. 2002). In this instance, wild-type mRNA was extracted from protonemata and reverse transcribed to create a cDNA library that was then target to advanced shuttle mutagenesis including an antibiotic resistance selection marker using a different transposon-derived system. The use of cDNA was assumed to enable highly specific HR-based integration into actively transcribed genes rather than genome-wide integration by use of gDNA. The transformation of *Physcomitrella* resulted in
2636 mutant plants that differed from the wild type in terms of developmental and morphological as well as physiological characteristics such as nutritional requirements.

By far the largest tagged DNA mutagenesis project led to the production of 73,329 mutant plants, classified in 16 different categories, consisting of plant structure, color, gametophore development, cell shape, etc. (Schween et al. 2005). The transformation was again accomplished using cDNA libraries to generate mutants; however, in this case, the libraries each were created using a variety of different starting tissues, such as protonema, gametophyte, and sporophyte tissue, as well as from moss treated with various agents, such as the phytohormones cytokinin and abscisic acid (ABA). Interestingly, the sporophyte and gametophore libraries were subtracted cDNA libraries, i.e., mRNA from one tissue was used to capture complementary cDNA from the other tissue in order to obtain specific cDNA exclusive to the latter tissue. In addition to phenotypic descriptions, the libraries were normalized, and extensive sequencing yielded over 110,000 expressed sequence tags (ESTs).

Unfortunately, most collected mutants still miss precise association of a phenotype to a specific genetic locus. It is only recently that researchers have begun to elucidate some of these connections. With recent technological advances, a study from 2010 revealed the genetic basis for long-described Physcomitrella patens mutants that exhibit abnormal responses to auxin treatments (Prigge et al. 2010). These authors used information derived from homologous genes of Arabidopsis thaliana to identify genes in Physcomitrella patens putatively involved in the response to auxin that delivers an accelerated and early transition from chloronema to caulonema. Using mutants that had been generated in 1979, described as being arrested at the primary chloronemal stage or delayed in the chloronema-to-caulonema transition, they sequenced the putative Physcomitrella patens genes. Comparisons between the wild-type genes and the mutant sequences revealed lesions in a specific motif of three moss genes that result in a single amino acid substitution in the corresponding protein. Thus, the genetic basis for the defect was identified successfully 30 years after the description of the phenotypic mutants.

For the vast majority of mutants, such links between a phenotype and a specific genetic change could not be assigned as yet. One major reason for this predicament is the lack of sufficient molecular markers at the time of each individual mutagenesis study. Molecular markers denote sequence polymorphisms between individuals of the same species or different species and allow the mapping of genetic loci in map-based cloning experiments. Co-segregation analyses of the mutant allele and such markers enable the determination of a mutation’s genetic background, based on the assumption that physically close loci on a chromosome are less frequently recombined during meiosis than distant ones. A first step in the establishment of such markers for Physcomitrella patens was accomplished in 2006, in the form of 110 polymorphic short sequence repeats (SSR), identified via a computer-based search against an expressed sequence tag (EST) database of the “Gransden” strain (von Stackelberg et al. 2006). Upon comparing the SSR sequences with different Physcomitrella strains, the genetically distant French strain “Villersexel” was used to
identify polymorphic SSRs. SSRs represent DNA sequences of mono- up to hexa-nucleotide repeats with genome-wide distribution yet accumulate in non-repetitive DNA and untranslated 3' and 5' regions of genes. The ESTs had been sequenced previously as part of the moss transcriptome assembly, and as such (Rensing et al. 2002a), they offered absolute linkage of the markers to genes with putative or known function.

In 2008, the publication of a reference genetic linkage map for *Physcomitrella patens* marked a milestone for moss forward genetics (Kamisugi et al. 2008). As in the previously described project, researchers used mapping populations of “Gransden” and “Villersexel” generated in Freiburg, Germany, and Leeds, England. However, in addition to 42 SSRs, another 1378 amplified fragment length polymorphisms (AFLP) formed the basis for the map. AFLPs represent a different type of molecular marker and are produced upon processing of genomic DNA by digestion with two defined restriction enzymes and subsequent PCR amplification of the fragments. High-resolution electrophoresis is used to visualize the pattern and lengths of these AFLPs. As this pattern is specific to a species or an individual organism, it enables straightforward genotyping, although the AFLP loci are anonymous, i.e., their sequence context is primarily unknown. However, analyses of the AFLP mapping population enabled the computational construction of a genetic map that was subsequently integrated with SSR-based linkage information to anchor it to the genome sequence. The resulting map comprised of 1420 markers and suggested high genome coverage, as 99.8% of the *Physcomitrella patens* genome was predicted to be within 10 centimorgan (cM) of a mapped marker.

**Reverse Genetics**

Complementing the classic methodology of forward genetics, established gene targeting methods together with bioinformatics enable reverse genetics on a far larger scale. The combination of transcriptomic and genomic data provide a basis for an expression-based characterization of genomic regions. In addition, searches against databases covering annotated sequences from different species enable functional inference based on homology with candidate genes. With the advent of next-generation sequencing technology, the volume of available data is growing continuously, facilitating large-scale bioinformatics-based gene identification. Candidate genes of interest are then targeted for gene disruption and generation of loss-of-function phenotypes. Reverse genetics rely on the assumption that such knockouts reveal the respective function of a gene. In *Physcomitrella patens*, highly efficient gene targeting through homologous recombination presents an excellent tool for the study of genes of unknown function. The comment from 1998, years before *Physcomitrella*’s entry into the genomic age, “until recently, this approach [reverse genetics via gene targeting] was not available to plant molecular biologists” highlighted the potential of the moss for reverse genetics and emphasized its obvious experimental amenity and efficient HR compared to *Arabidopsis* (Reski 1998). By targeted gene disruption of the *Physcomitrella* homologue of the bacterial cell
division protein FtsZ and microscopic characterization of the knockout phenotype revealing the formation of macrochloroplasts, Strepp et al. (1998) were able to report the first organelar division protein to be identified in eukaryotes (Fig. 4).

Since the late 1990s, reverse genetics have developed into a standard technique for the identification of gene function in Physcomitrella patens, as evidenced by an ever-increasing number of studies. An exemplar of such investigations is the functional characterization of FtsZ proteins performed via multiple targeted gene knockouts (Martin et al. 2009). The FtsZ is a self-assembling GTPase that forms a contractile ring at the division site during plastid division in plants and traces back to an ancestral homologue in cyanobacteria. In contrast to bacteria, which encode only a single FtsZ protein, plants harbor at least three different nuclear-encoded proteins in two families, FtsZ1 and FtsZ2. The Physcomitrella genome encodes for five FtsZ proteins in total, comprising two members of each family FtsZ1 and FtsZ2 as well as a single member of a third family, FtsZ3. To elucidate their putative functions in the moss, Martin et al. (2009) performed targeted knockouts of the five FtsZ genes. Phenotypic analysis of the resulting knockout plants revealed mainly three different patterns, correlated with Physcomitrella’s three FtsZ families: FtsZ1 proteins were characterized as interacting partners to fulfill functions related to the maintenance of chloroplast shape, the FtsZ2 family members appear responsible for plastid division, and FtsZ3 was characterized as a molecular link between cell and chloroplast division in Physcomitrella patens based on its co-localization to the chloroplast as well as to the cytoplasm.

**Fig. 4** Two principles of gene targeting by homologous recombination (inspired by Fig. 5 in Strotbek et al. 2013). (a) Targeted knockout of a gene by disruption of its coding sequence, e.g., via integration of a different gene for functional studies. (b) Targeted knockin, e.g., via fusion of a reporter gene to a gene of interest, for localization or interaction studies.
RNA Interference and Gene Silencing

The inhibition of gene expression by RNA interference (RNAi) is a highly conserved regulatory pathway expressed in all eukaryotes. The conservation of the pathway is widely attributed to be a consequence of its role as an early defense system against viruses or other forms of foreign genetic material and, especially in plants, as a control system for self-propagating transposable elements. A detailed review on the role and regulatory mechanisms involving RNAi in Physcomitrella patens was provided by Arif et al. (2013). In general, introduced double-stranded RNA (dsRNA), either derived from a virus or delivered experimentally via a transcribable genetic element, is processed by Dicer-like (DCL) proteins to release small interfering RNAs (siRNAs). Each siRNA then guides an RNA-induced silencing complex (RISC) to a specific target messenger RNA (mRNA) that has a matching antisense sequence. An Argonaute protein (AGO) within the RISC cleaves the target mRNA which is subsequently degraded. In this way, RNAi prevents the translation of an mRNA, i.e., silences a gene at the posttranscriptional level. In Physcomitrella patens, this mechanism is utilized for targeted silencing of single genes or whole gene families. Members of gene families typically share at least short segments of common sequence allowing all family members to be targeted by a single RNAi. The cDNA of a target gene region (containing the common sequence) is cloned into an expression cassette as an inverted repeat separated by a spacer region. After successful transformation of the construct, its expressed mRNA, containing the self-complementary inverted repeats, folds back on itself to form a dsRNA. The dsRNA is processed through the RNAi pathway and gives rise to siRNAs that sufficiently guide the RISC in posttranscriptional silencing of multiple members or the entire gene family.

However, the dispersal of a diverse set of siRNAs from one dsRNA might lead to unintended side effects such as off-target gene silencing and therefore is not always the method of choice.

Alternatively, a more specific silencing mechanism can be achieved by using micro RNAs (miRNAs). An miRNA primary transcript is encoded by a discrete miRNA gene and upon expression forms a hairpin-like dsRNA as it contains internal self-complementary sequences. In a multistep reaction guided by a member of the DCL family, among other associated proteins, the primary transcript is processed into a miRNA precursor that gives rise to a single miRNA that is complementary to its target gene. This miRNA then guides the RISC to a target transcript in a highly specific mode. In a manner similar to the RNAi via siRNA mechanism, subsequent AGO- and DCL-mediated mRNA cleavage results in the degradation of the target mRNA: a form of posttranscriptional gene silencing. In Physcomitrella, it was shown that the miRNA pathway can also lead to transcriptional inhibition by guiding an RNA-induced transcriptional silencing complex (RITS) that alters DNA methylation patterns depending on the ratio of the miRNA to the target mRNA (Khraiwesh et al. 2010). In the PpDCL1b mutant, target mRNA cleavage did not occur even though the target mRNA transcript levels were reduced suggesting an epigenetic silencing of the target mRNA-encoding gene. Indeed, sequencing of target mRNA
genes in this mutant revealed increased methylation compared to wild type. The proposed pathway implies the reimport of the RITS into the nucleus and subsequent targeting of the genomic locus complementary to the miRNA/target duplex. Local DNA methylation then leads to the transcriptional inactivation of the specific gene. The pathway complements the classical model of miRNA function and, at the same time, constitutes a novel mechanism of eukaryotic gene regulation: targeted epigenetic silencing via specific RNA/DNA interaction. Interestingly, miRNA-mediated epigenetic gene silencing has become the center of attention in current medical research, since the investigation of human cancer-associated miRNAs has indicated the presence of mature miRNAs in the nucleus (Hwang et al. 2007). While miRNAs are known to regulate components of the epigenetic machinery, they might also directly control DNA methylation and histone modifications in mammals (Lujambio and Lowe 2012), using the pathway that was first described in Physcomitrella patens (Fig. 5).

The exploitation of the miRNA pathway for gene knockdown was demonstrated in Physcomitrella patens with the transformation of an expression cassette containing a conserved miRNA gene from Arabidopsis thaliana. The miRNA-generating region of the miRNA gene was replaced with a specifically designed sequence, leading to the generation of an artificial miRNA (amiRNA) that targeted the FtsZ2–1 transcript (Khraiwesh et al. 2008). Expression of the primary transcript, as well as the correct formation of the required hairpin-like structure were achieved which resulted in the successful processing and release of the amiRNA. The target gene was posttranscriptionally silenced as demonstrated by the formation of macrochloroplasts – the same phenotype of a previously described FtsZ2–1 knockout mutant.

**Genomics**

With the publication of the nuclear genome (Rensing et al. 2008), Physcomitrella patens entered the genomic age. Based on its phylogenetic position, bridging the gap between green algae and flowering plants, the genome has enabled (comparative) studies in the context of plant evolution. For the first time, it provided a resource to gain scientific insight into the conquest of land by plants. In the course of ever-growing data and constant improvement of genomic annotation, Physcomitrella’s role as an informative genomic reference is undisputed within the plant community.

**History of the Nuclear Genome**

The origins of genomic studies using the attributes of Physcomitrella patens date back to the 1990s (Reski et al. 1994). In these first efforts, long before the genome was at hand, transcriptome sequencing of isolated complementary DNA (cDNA) by subtractive hybridization was performed on a large scale. A comparison of sequence similarity between these cDNAs and annotated genes of flowering plants was used to
identify homologous as well as putative species-specific genes. Subsequent gene discovery projects were initiated in Freiburg (Germany), Okazaki (Japan), Leeds (England), and St. Louis (USA) utilizing large-scale expressed sequence tag (EST) sequencing. These efforts yielded the first virtual *Physcomitrella patens*...
transcriptome, an extensive data resource of both high-quality sequences and a low degree of redundancy (Rensing et al. 2002b). Although the number of protein-coding genes was estimated to be ~25,000 – a number comparable to a similar estimate for *Arabidopsis thaliana* – nearly half shared no sequence homology with sequences from any other organism. In addition, the majority of the longer virtual transcripts were only fragments of whole genes. Based on the growing general interest in *Physcomitrella* as a plant experimental model, driven by the analysis of the transcriptome, and the limited validity of the transcriptomic data for addressing important genomic questions, the common will to sequence *Physcomitrella patens*’ genome emerged in the moss community. Consequently, in 2004, the International Moss Genome Consortium dedicated to pursue this task was established in Freiburg at the annual moss conference. As part of the Community Sequencing Program of the US Department of Energy’s Joint Genome Institute (JGI), a whole-genome shotgun sequencing approach was executed, with the “Gransden” laboratory strain as a reference strain from which the DNA was extracted. This program succeeded in establishing the genome of *Physcomitrella patens* as the first bryophyte genome (and one of the first plant genomes) to be sequenced. The JGI justified its decision to take on the task by stating that “[t]he moss *Physcomitrella patens* is becoming widely recognized as an experimental organism of choice not only for basic molecular, cytological, and developmental questions in plant biology, but also as a key link in understanding plant evolutionary questions, especially those related to genome evolution. *Physcomitrella* is well placed phylogenetically to provide important comparisons with the flowering plants. In terms of evolutionary distance, *Physcomitrella* is to the flowering plants what *Drosophila* is to humans. Having the full *Physcomitrella* genome available will greatly inform bioinformatic comparisons and functional genomics in plants, just as the mouse, *Fugu*, and *Drosophila* genomes have informed animal biology.” After assembly and initial structural and functional annotation, the genome version 1.1 was released in 2007, and its features were published in 2008 (Rensing et al. 2008). The draft genome had a size of 480 Mbp and contained 35,938 protein-coding genes and as expected, it revealed important insights into the conquest of land by plants. Around the same time, the *Physcomitrella patens* genome resource “cosmoss.org” was launched to provide the community with a central web portal for exploration and annotation, in the tradition of comparable model organism databases like “TAIR” for *Arabidopsis thaliana* or “FlyBase” for *Drosophila melanogaster*. Since that time, the community has maintained a continuous effort to further improve genome annotation, underscored by the decision by JGI to extend the *Physcomitrella* resource as one of its “flagship” plant genomes. The development of the genetic linkage map based on amplified fragment length polymorphisms (AFLP) and simple sequence repeat (SSR) markers (Kamisugi et al. 2008) is one prime example of these efforts, and it provides an invaluable information resource for targeted identification of genes connected to mutant phenotypes in forward genetic studies. In addition, an ambitious single nucleotide polymorphism (SNP) genotyping analysis was performed, which resulted in the identification of approximately 600,000 high-confidence SNPs that are distinguishable between the *Physcomitrella patens* laboratory strain “Gransden” and
the genetically divergent “Villersexel” strain from France. These mapping tools enabled the assembly and annotation of a yet unpublished chromosome-based genome release. The latest official genome version 1.6 (V1.6) was published in 2013 as a complete re-annotation of the initial release (Zimmer et al. 2013). Along with the incorporation of increased transcriptomic data representing a multitude of developmental stages and tissue types, V1.6 introduced updated protein-coding gene annotations with improved EST support, untranslated region (UTR) annotation, and alternative splicing isoforms. The complete set is comprised of 32,275 protein-coding genes with accompanying extended information describing nonprotein-coding loci.

**Expression Profiling**

For many years, gene expression analyses were limited to either a single or only a few genes using such techniques as mRNA quantification via Northern blotting. The development of biological chip technologies in the mid-1990s introduced the first platform for parallel, large-scale quantitative gene expression analysis using what became known as the DNA microarray. In general, microarrays consisted of a solid surface coated with thousands of individual evenly spaced “spots” each containing identical single-stranded DNA probes bound to the surface matrix in the nano- to picomole concentration range. The probes typically consisted of cDNAs complementary to each mRNA representing each member of a gene set of interest, e.g., all known protein coding sequences of an organism. Upon loading the array and hybridizing a fluorescently labeled sample RNA to the probes, subsequent removal of non-hybridized material, the probe:mRNA hybrid is stably bound to the solid matrix of the array. The level of hybridization of sample RNA to the probe DNA is measured with a scanning device that detects the fluorescent signal for each spot. Depending upon the respective conditional context of sampled RNA, e.g., tissue type or treatment, the specific response to the tissue or condition in terms of gene expression is quantifiably measured.

With the generation of genomic data, the first microarray for Physcomitrella patens was produced in 2007 based on the “Agilent” platform and contained approximately 20,000 features based on 80,000 ESTs from Leeds, England, and Okazaki, Japan. This array enabled researchers to study the transcriptional responses of various tissues of Physcomitrella patens to multiple individual treatments, such as treatment with exogenous abscisic acid (ABA) or osmotic, salt, and dehydration stress treatments (Cuming et al. 2007). Expression profiling using this technology revealed 130 genes induced by dehydration, 56 by ABA, 10 by osmotic stress, and eight genes by salt stress, with 51 genes showing induction in more than one treatment. An extended homology-based search against genes of other plant species yielded ABA- and drought-responsive Physcomitrella homologues of angiosperm genes expressed during drought stress and seed development. These findings enabled the posing of hypotheses concerning the connection between conserved stress regulatory transcription factors exclusive to the seed developmental pathway
Accompanying the initial and subsequent genome sequence releases, several microarrays based on assembled transcripts were made available using the “CombiMatrix” (V1.1 and V1.2) and “NimbleGen” (V1.6) platforms. Since the availability of these chips enabled large-scale expression profiling for the “state-of-the-art” gene set, a variety of studies have been published focussing primarily on the transcriptional responses of the moss to biotic and abiotic stresses. In a similar fashion as previously described, one study characterized the transcriptional response of protonemal cells to salt stress and ABA treatments (Richardt et al. 2010). By focusing on genes encoding transcription-associated proteins, the team was able to demonstrate that responses to these treatments were regulated at the transcriptional level. In addition, based on the overlapping expression profiles, they suggested that ABA mediates the salt stress responses in *P. patens*. In the same year, another study revealed both distinct and conserved pathways of the UV-B response of *Physcomitrella patens* compared to that of *Arabidopsis thaliana* (Wolf et al. 2010). *Physcomitrella patens* prefers to grow in sun-exposed and open areas. Concordantly, *P. patens* was described as being capable of surviving more severe UV-B stress than *Arabidopsis*, and as such harbors conserved, yet widely expanded gene families primarily responding to UV-B. This has evolutionary connotations in that during the time when plants invaded the land, UV radiation posed a major challenge to survival, as the levels were reportedly much higher than in modern times. Thus, it is likely that *Physcomitrella* has retained the genetic structure present in the early land plants, and this would involve the conserved and widely expanded UVB-responsive gene families.

In a large-scale study focused on the transcriptome of *Physcomitrella patens* combining a time series of microarray measurements, between 1 and 24 h, with quantitative real-time PCR data, Beike et al. (2015) were able to characterize the cold response of the moss. The differentially expressed genes defined in this study were clustered based on their gene ontology annotation, i.e., information on the subcellular localization, molecular function, and involvement in biological processes of their respective products, if available. In this way, the team was able to identify the transcriptional changes and associated processes related to the early cold stress response and the emerging cold acclimation. Surprisingly, the early cold response is dominated by a number of species-specific genes, which putatively renders the response process species or lineage specific. This is difficult to explain as many plants exhibit cold stress responses. However, *Physcomitrella patens* is poikilohydric, i.e., it is not able to actively regulate cell or tissue water content unlike more complex plants, and as such, it exhibits levels of dehydration tolerance greater than the majority of angiosperms. Cold stress, especially for poikilohydric tissues, has a major dehydration component associated with the drop in temperature, and thus the combination of poikilohydry and increased dehydration tolerance might explain the observed highly species-specific response of *Physcomitrella patens*.

In order to establish a community resource specifically aimed at the distribution of expression analyses data for *Physcomitrella patens*, several large datasets were
made available online (Hiss et al. 2014). The microarray data originates from experiments utilizing a multitude of tissue types such as spores, protoplasts, protanemata, or gametophores, developmental stages such as germinating spores or developing gametangia, as well as various treatments and stresses such as different light intensities and qualities, dehydration, or different growth media. To ensure data integrity and future accessibility for the community, sample expression profiles were linked to Physcomitrella gene IDs and the tissue types; developmental stages and experimental conditions were semantically connected to respective ontology terms. As a front end for data access, the online meta-analysis tool “Genevestigator” was chosen, as it has successfully provided similar data for several other model plant species and it also features a gene expression search engine that is focused on integrating all contents and comparisons between experiments. The platform thus provides an invaluable information resource for moss researchers that are interested in the conditional expression of a Physcomitrella gene or wish to conduct cross-species comparisons of gene expression patterns. In a similar way, but relying on high-throughput sequencing of RNA (RNA-seq), the US Department of Energy’s Joint Genome Institute initiated its Plant Gene Atlas Project. The program aims at large-scale, coordinated development of reference transcriptomes for the institute’s flagship species: the alga Chlamydomonas reinhardtii, the grass model Setaria italica, soybean, poplar, and not least, Physcomitrella patens. At the same time, the project provides a test bed for new experimental technologies associated with dissection, capturing, and amplification of plant RNA.

### Genes and Gene Structures

For the gene models of the Physcomitrella patens V1.6 genome annotation (Zimmer et al. 2013), only transcripts with EST or full-length cDNA evidence were chosen, complemented by mapped sequences that had been experimentally confirmed and published by the community, independent of the genome project. In total, 38,357 transcripts that represent 32,275 protein-coding genes, with a mean transcript and gene length of 1389 bp and 2369 bp, respectively, have been described. A subset of 15,757 transcripts possess both 5' and 3' UTR annotation and another 21,464 transcripts having at least either one of the UTRs delineated, based on full support by experimental evidence. Alternative splicing (AS) occurs in 10.8% of loci with an average of 2.52 and a maximum of 11 transcripts per locus. The predominantly observed form of AS is defined by alternative intron acceptor sites, i.e., the use of a different 3' splice junction that results in an alteration of the 5' downstream exon boundary. The prevalent consequence of this mechanism in Physcomitrella patens is the retention of introns. In addition, the coding sequence is altered in 51% of transcripts that are target of AS, resulting in 2380 distinct proteins. Also 56% of AS transcripts exhibit an alteration of the 5' UTR. The effect of 5' UTR introns on gene expression, regulation, translation, and nonsense-mediated mRNA decay have been described independently. These findings, together with the unique length of
Physcomitrella 5' UTR introns, suggest that the moss makes frequent use of this type of gene regulation.

The moss genome contains a higher fraction of genes with a single exon (23.4%) than the genomes of green algae (<10%) and other land plants (~19%). While intron loss is common in the evolution of land plants and outnumbers intron gain in vascular plants, it may specifically occur in Physcomitrella as a result of retrocopying: the reverse transcription of mRNA followed by partial replacement of genomic DNA by the intron-less cDNA. In the case of Physcomitrella patens, this mechanism might be facilitated by gene conversion via homologous recombination. Another explanation, based on the observation that single-exon genes in the moss generally lack evidence for expression, is that these gene models might represent fragmentary predictions, nonprotein-coding genes, or pseudogenes.

A striking characteristic of the Physcomitrella patens genome is the presence of tandemly arrayed genes (TAG). In general, TAGs are paralogous genes, i.e., originate from a species-specific duplication event, residing closely together on a chromosome. Although the occurrence of TAGs was described in Arabidopsis thaliana and Oryza sativa (and others) and was observed to exist in an even higher proportion of their genomes, the TAGs of Physcomitrella patens deviate significantly in orientation and conservation. TAGs in Physcomitrella patens occur predominantly in an inverse orientation and feature highly conserved sequences even in the noncoding regions (Lang et al. 2008). These specific differences together with Physcomitrella patens' prevalent and efficient HR capabilities suggest a concerted evolution of these genes through gene conversion, that is, the replacement of one gene with its homologue to retain and conserve sequences. This might be a mechanism for the haploid dominant moss to maintain “pseudoalleles” of highly expressed genes. As described in the V1.6 release, TAGs contribute to the majority of orthologous gene clusters with only two member genes. With 27% of all gene clusters, their share is substantially higher in the moss compared to other species of plants (Table 1).

Transposable Elements

In plant genomes, repetitive sequences account for a large fraction of the total DNA sequence. These repetitive elements are typically comprised of tandem repeats and simple sequence repeats (SSR), transposable elements (TE), and duplicated genomic fragments. Repeats in general, and TEs in particular, are of special evolutionary interest as they enhance the diversity of plant genomes. In some cases, TEs alone contribute to 80% of a plant genome (Feschotte et al. 2002). Such elements are capable of transposition, i.e., they possess the ability to change their position in the genome, and are enclosed by specific repetitive sequences. Depending on the nature of these repetitive sequences and the intermediate form of TEs during transposition, TEs can be divided into two classes. Class 1 elements transpose via an RNA intermediate and either have long terminal repeats (LTR) similar to LTR retrotransposons or are terminated by a SSR (usually polyA) similar to non-LTR retrotransposons. Class 1 elements also include long and short interspersed nuclear
elements, LINEs and SINEs, respectively. Class 2 elements transpose via a DNA intermediate. They possess terminal inverted repeats and can be subdivided according to whether they are autonomous or not.

In the current *Physcomitrella patens* genome annotation, V1.6, full-length LTR retrotransposons and related fragments have been described to cover about half of the entire genome (Zimmer et al. 2013). In comparison, *Physcomitrella* harbors approximately three times more LTR retrotransposons than *Arabidopsis thaliana* but three times fewer than rice.

The class 2 elements, Helitrons, nonautonomously transpose via a rolling-circle mechanism, in which a single DNA strand is transferred from the original genomic locus to the target site. The cell’s DNA repair machinery then uses this single strand as a template for synthesis of the second strand, which results in duplication of the Helitron. The genome of *Physcomitrella patens* contains only a single Helitron family comprised of 19 members. High sequence similarity between these Helitrons (96%) indicates their activity within the past 3 million years (Rensing et al. 2008).

**Noncoding RNAs**

The involvement of small RNAs (sRNA) in plant gene regulation is a hot topic in current research, and the extent of the regulatory role that sRNAs play is still only partly understood. However, their role in fundamental processes like development, morphogenesis, and genome defense is obvious. In *Physcomitrella patens*, the two

---

**Table 1** Evolution of the *Physcomitrella patens* genome annotation (modified from Zimmer et al. 2013)

<table>
<thead>
<tr>
<th></th>
<th>V1.1 Rensing et al. (2008)</th>
<th>V1.2 Lang et al. (2008)</th>
<th>V1.6 Zimmer et al. (2013)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genome size (Mb)</td>
<td>480</td>
<td>480</td>
<td>480</td>
</tr>
<tr>
<td>Scaffolds</td>
<td>2106</td>
<td>1995</td>
<td>1985</td>
</tr>
<tr>
<td>Protein-coding genes</td>
<td>35,938</td>
<td>27,966 (−7972)</td>
<td>32,275 (+4309)</td>
</tr>
<tr>
<td>Protein-coding genes with EST support</td>
<td>12,593</td>
<td>19,119 (+6523)</td>
<td>26,722 (+7603)</td>
</tr>
<tr>
<td>Protein-coding transcripts</td>
<td>35,938</td>
<td>27,966 (−7972)</td>
<td>38,357a (+10,391)</td>
</tr>
<tr>
<td>Annotated as alternatively spliced</td>
<td>–</td>
<td>–</td>
<td>3500</td>
</tr>
<tr>
<td>Genes with UTRs</td>
<td>4517</td>
<td>4515 (−2)</td>
<td>15,757a (+11,242)</td>
</tr>
<tr>
<td>Either UTR Genes</td>
<td>8418</td>
<td>8381 (−37)</td>
<td>16,010 (+7629)</td>
</tr>
<tr>
<td>Transcripts</td>
<td></td>
<td></td>
<td>21,464a</td>
</tr>
<tr>
<td>Gene density (kb per gene)</td>
<td>13.4</td>
<td>17.2</td>
<td>14.9a</td>
</tr>
<tr>
<td>Exons/gene</td>
<td>4.9</td>
<td>5.4</td>
<td>5.0a</td>
</tr>
<tr>
<td>Mean exon length (bp)</td>
<td>246</td>
<td>234</td>
<td>275a</td>
</tr>
<tr>
<td>Mean intron length (bp)</td>
<td>311</td>
<td>277</td>
<td>278a</td>
</tr>
</tbody>
</table>

*a* including splice variants and data from the miRBase registry, release 18.
main types of sRNAs, siRNAs, and miRNAs have been studied, with a focus on their biogenesis and mode of action. Axtell presented an exemplary and detailed review of the moss’s diverse noncoding RNA species (Axtell 2009). Small RNA annotation has greatly profited from next-generation RNA sequencing (RNAseq) that enabled high-throughput sRNA identification, complemented by an ever-growing set of bioinformatic tools. The V1.6 genome annotation includes 229 miRNAs from 108 miRNA families, sharing 47 families with Chlamydomonas reinhardtii and 187 families with Arabidopsis thaliana (Zimmer et al. 2013). Loci that either code for siRNAs or miRNAs seem to possess opposed properties (Coruh et al. 2015). MiRNAs typically occur in genic regions that have an overall lower DNA methylation state, i.e., enhanced accessibility to the transcription machinery. SiRNAs, on the other hand, contribute to the vast landscape of noncoding RNA in intergenic regions that contain a high ratio of repetitive sequences and dense methylation. Unfortunately, a large proportion of the latter still lacks proper annotation.

Besides small regulatory RNAs, other elementary noncoding RNA species are represented in the V1.6 genome annotation. A total of 432 tRNA loci have been identified (Rensing et al. 2009), with 417 of the encoded tRNAs encompassing all of the 20 standard amino acids, with at least one of the possible anticodons per amino acid. Thus, not all of the 57 possible tRNAs seem to be used in Physcomitrella patens. A gene for a selenocysteine tRNA was also identified in the set. In addition, there are 798 rDNA loci encoding ribosomal RNAs that were annotated in V1.6, as well as 213 small nucleolar RNA (snRNA) and 6 signal recognition particle (SRP) loci. SRPs occur in the spliceosome where snRNAs are associated with specific proteins in small nuclear ribonucleoprotein particles (snRNP) and catalyze the splicing of pre-mRNA. SRPs are also associated with ribonucleoproteins but contribute to the targeting of proteins to the endoplasmic reticulum of eukaryotes.

Organellar Genomes

The genome sequences and genome features for both the chloroplast and mitochondrion were published in 2003 (Sugiura et al. 2003) and 2007 (Terasawa et al. 2007), respectively. According to the endosymbiotic theory, chloroplasts originated via engulfment of a photosynthetic cyanobacterial ancestor, and over time, endosymbiotic gene transfer to the host cell nuclear genome led to a reduced plastid genome (plastome) – in land plants, the length typically ranges between 120 and 160 kbp. The specific features of chloroplast DNA (cpDNA) were reviewed as part of a general chapter on chloroplasts in Physcomitrella patens by Sugita and Aoki (2009). With 122,890 bp, the plastome of Physcomitrella patens lies between the liverwort Marchantia polymorpha (~121 kbp) and the hornwort Anthoceros formosae (~160 kbp) in size, closely to the plastome size of another bryophyte, Syntrichia ruralis, with 122,530 bp (Oliver et al. 2010). In terms of GC content, it ranges closer to the two bryophytes, with 28.5% compared to ~29% and ~33%, respectively, than to vascular plants (~38–39%). It comprises 83 protein-encoding genes, 31 tRNA and 4 rRNA genes, as well as a single pseudo-tRNA gene. Most
strikingly, comparative analysis revealed fundamental insights into cpDNA evolution in land plants. Firstly, a large inversion of 71 kbp was described to be present throughout Funariaceae, Disceliaceae, and Encalyptales – the largest inversion documented in plants so far. Thus, it is assumed that this inversion was already present in the last common ancestor of these three moss families. Secondly, the absence of the RNA polymerase alpha chain gene (rpoA) from cpDNA and its transfer to the nuclear genome were first described in *Physcomitrella patens*. The gene is present in the plastid genomes of other mosses and was lost twice in the evolutionary history of bryophytes.

The *Physcomitrella patens* mitochondrial genome (chondriome) was the second published bryophyte chondriome, and it was the smallest of reported land plant chondriomes at the time of sequencing. The circular-mapping chondriome DNA of *Physcomitrella patens* harbors 41 protein-encoding genes, 24 tRNA and 3 rRNA genes. In comparison to other species, neither significant synteny with chlorophytes and flowering plants could be found nor the multipartite structure reported to be present in the chondriomes of flowering plants. Yet, the mitochondrial genome has retained the prototypic features of land plant chondriomes. Phylogenetic comparisons supported the close relationship between *Physcomitrella patens*, the Charophyta and liverworts as well as the assumption that the clade of mosses and liverworts are a sister group to the flowering plants.

**Evolution of Transcription-Associated Proteins**

Now that *Physcomitrella patens* is widely used as a model to study genetic adaptations during the evolution of land plants, a special emphasis has been put on the comparative analysis of transcription-associated proteins (TAPs), which largely account for modulating and regulating gene activity (Rensing et al. 2009). In eukaryotes, TAPs are involved in complex networks that maintain protein-encoding gene regulation. These networks center on specific transcription factors (TF) that bind to cis-active elements in order to activate or repress target gene expression. But they also include other transcriptional regulators (TR) such as coactivators and corepressors that bind and influence TFS, transcription initiation factors that recruit components of the transcription machinery by recognizing core promoter elements, and last but not least chromatin remodeling factors that mediate histone modifications and DNA methylation and thus affect the accessibility of DNA.

It was proposed that there is a positive correlation between the quantity of TAPs and the number of cell types in a respective organism: the comparative analysis of TAPs in *Physcomitrella patens* and diatom, red and green algae species as well as rice and Arabidopsis support this proposal. On average, the TAP gene families in *Physcomitrella* are two to three times higher than those found in the genomes of algae but only represent a fourth of total TAPs and putative TAPs seen in the genomes of vascular plants. The analysis also revealed that the moss contains representatives of most of the transcription-associated proteins found in seed plants. Only six novel angiosperm TF families emerged after the evolutionary split of
mosses and seed plants. These TFs are involved in flower development and lateral organ formation in the sporophyte in flowering plants. Interestingly, as unicellular algae do not contain most of the TF families present in land plants, the expansion of TAP families is most likely to have evolved during or after the switch from unicellularity to multicellularity. These hypotheses will greatly profit from the ongoing genomic sequencing and comparative genomic studies.

**Generative Polyploidization in Funariaceae**

Among land plants, generative polyploidy in the form of large-scale or whole-genome duplication (WGD) events is a widespread phenomenon. Rensing et al. (2013) highlighted the evolutionary importance of polyploidization with a focused study on the moss family, Funariaceae. Generally, WGD events are recognized as driving forces behind speciation and diversification of plant lineages. In addition, their crucial influence on the evolution of organismal complexity is widely hypothesized. Two forms of WGD events have been described: autopolyploidization, i.e., genome doubling within a species and without hybridization, and allo-polyploidization, that is, the doubling of the genome in association with the introduction of a genetically different chromosome sets resulting from the hybridization of two species. While autopolyploidization probably provides a short-term evolutionary advantage during periods of environmental change, allopolyploidization might increase the potential of genes to sub- and neo-functionalize and thus presents a possible driver of speciation. The fate of duplicated genomic regions is key to understanding the evolutionary importance of WGD events, as it is possible to draw conclusions from the diversifying effects of the post-WGD phase on paralogous genes.

In the context of haploid-dominant mosses, polyploidization might present a greater potential evolutionary advantage than for plants that have a dominant diploid phase. Additional copies of genes and chromosomes after a WGD event could introduce a buffer against somatic mutations that may eventually affect the germ line. A WGD event would also introduce the same potential for genetic sub- and neo-functionalization as such an event does for diploid-dominant organisms. WGD events also induce heterosis, i.e., the increased biological quality or capability of a hybrid. This might explain the observed increase in size of extant moss species compared to those found preserved in Eocene (45 million year old) amber. *Physcomitrella patens* has undergone at least one polyploidization between 30 and 60 million years ago and is thus referred to as a paleopolyploid (Rensing et al. 2007). This finding was based on the analysis of nearly 3000 paralogous genes and relied upon the calculation of synonymous substitution rates to estimate the divergence rates of the paralogues. This period, 30 and 60 million years ago, nicely coincides with the fossil record and falls in the time frame for the mass extinction events that occurred at the end of the Eocene (Stehlin 1909), ca. 33 million years ago. Phylogenetic analyses further implied that the ancient polyploidization observed in
Physcomitrella patens had affected the ancestral lineage that generated the Funariaceae.

For species in the Funariaceae, including Physcomitrella patens, natural hybrids appear frequently, presumably because of comparable habitats and growth in close proximity to one another (McDaniel et al. 2010; Beike et al. 2014). Experimental hybrids have been generated by artificial induction of autopolyploidization after transfection of P. patens protoplasts during transformation protocols. The polyploids, mostly diploid transformants, are indistinguishable from haploid plants based on morphological features alone.

As hybridization seems to represent a common phenomenon among Funariaceae, there is a great potential for natural polyploidization in this clade. Allopolyploidization as a result of hybridization occurs in three different ways: apospory, diplospory, or syndiplospory. Apospory is the direct regeneration of a diploid gametophyte from the tissue of a hybrid sporophyte without the generation of an actual spore, diplospory requires an unreduced diploid spore (failure to form haploid spores from a spore “mother” cell via meiosis), and syndiplospory is the fusion of two spore mother cells (sporocytes) prior to meiosis. Among Funariaceae, a variety of different chromosome counts have been reported, yet all of them more or less revert to a putative base number of seven chromosomes. In addition, the Funariaceae family exclusively consists of monoecious species that, as introduced earlier, most likely share a common WGD event. Considering these characteristics, the chromosome history of Physcomitrella patens might be reconstructed as follows: an ancient hybridization of two parental (male and female) chromosome sets of \( n = 7 \) followed by allopolyploidization and haploidization leading to a monoecious species with \( n = 14 \). An additional (auto- or allo-) polyploidization with subsequent haploidization and a putative aneuploidization, i.e., the loss of a single chromosome, could have produced the extant haploid moss with \( n = 27 \). Given that the first WGD event in this reconstruction corresponds to the allopolyploidization in the ancestry of all Funariaceae, this scenario can also explain the exclusive monoecy of the family species. In addition, speciation through hybridization might result from reproductive isolation due to allopolyploidization, providing an explanation for the great variation in chromosome numbers across this moss family.

Future Directions

Undoubtedly, the versatile natural and experimental features that account for the popularity of Physcomitrella patens as a model species were decisive for the genome-sequencing project. In addition, its phylogenetically important position has contributed to the moss’s status as a reference in evolutionary developmental plant biology. With the availability of transcriptomic and genomic data, the evolutionary adaptations and responses to the consequences of the conquest of land have been a major focus of recent studies. Gene function and interaction studies helped shape the importance of this moss to genetic studies at the molecular level and have enabled further targeted investigations. With the plastid genome sequence and an
established transformation technique at hand, future research into the intricacies of photosynthesis would profit highly from the studies utilizing the moss. Chloroplast division and directed protein import into the plastid are prime examples of questions that are currently under intense investigation and that, by using the moss, would further elucidate significant aspects of plant evolution, such as the endosymbiosis of chloroplasts and associated evolutionary processes of gene conversion and cellular adaptation.

As studies on the Funariaceae family are ongoing and sequence data for species other than Physcomitrella patens are being produced, additional key features of the mosses are open for investigation. Genome evolution, especially in haploid-dominant organisms, is one example of high interest. Hybridization among species is common, and traces of ancient polyploidization in the moss family have been described. Future studies tackling these findings present a unique chance to study the underlying mechanisms and evolutionary consequences surrounding genome evolution in haploid organisms.

Physcomitrella also offers the opportunity to uncover regulatory mechanisms underlying the alternation of generations, i.e., the switch between gametophytic and sporophytic development and vice versa (Horst et al. 2016). While both generations differ fundamentally at the morphological level, they share a common genome. Thus, genomic and transcriptomic studies in the context of Physcomitrella patens with its extensive annotation and short generation time have great potential to shed light into this fascinating topic.

All these different aspects will greatly benefit from the increasing growth in data and further experiments in the field. Special attention will be paid to population genomic approaches to study individuals from single moss populations, as such studies could provide valuable insight into the effects of selfing and the mechanisms related to genome evolution in the Funariaceae.

References


Oliver MJ, Murdock AG, Mishler BD, Kuehl JV, Boore JL, Mandoli DF, Everett KDE, Wolf PG, Duffy AM, Karol KG. Chloroplast genome sequence of the moss *Tortula ruralis*: gene content, polymorphism, and structural arrangement relative to other green plant chloroplast genomes. BMC Genomics. 2010;11:143.


