Exploring non-coding mitochondrial DNA sequences in bryophyte molecular evolution

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Ute Volkmar

aus

Halle/Saale

Bonn 2010

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter: Prof. Dr. Volker Knoop
2. Gutachter: Prof. Dr. Jan-Peter Frahm

Tag der Promotion: 16. April 2010
Erscheinungsjahr: 2010

Contents

1 Summary

The thesis presented here focussed on the molecular evolution of non-coding mitochondrial DNA in liverworts and mosses.

To address remaining questions in moss phylogeny, three mitochondrial gene regions were investigated and established as novel molecular markers: the *nad*5-*nad*4 intergenic spacer region (Wahrmund *et al.* 2009) and the two group I introns in the *cob* gene and *cox*1 gene (cobi420, Wahrmund *et al.* 2010 and cox1i624, Volkmar and Knoop subm). Phylogenetic trees based on the single loci and concatenated data sets identified a placement of *Catoscopium*, *Drummondia* and *Timmiella* in the basal Dicranidae, proposed the exclusion of Gigaspermaceae from the Funariidae and a potential sister relationship of the nematodontous moss classes Tetraphidopsida and Polytrichopsida (Wahrmund *et al.* 2009, 2010). A different resolution of the basal-most moss taxa by chloroplast and mitochondrial markers was observed (Volkmar and Knoop subm) and its implications on molecular and morphological evolution in mosses discussed.

The hitherto assumed slow evolution of the mitochondrial genome in liverworts was contrasted with the discovery of recombinational activity in the intergenic region of the *trn*A-*trn*T-*nad*7 cluster, an ancient gene arrangement that is also present in algae and mosses. During liverwort evolution, an inversion and at least three independent losses of the *trn*T and adjacent spacer regions resulted in independent size decreases (Wahrmund *et al.* 2008). The *nad*7 gene, part of this gene cluster, is a pseudogene in all jungermanniid and marchantiid liverworts investigated. An ancient gene transfer to the nuclear genome occurred probably in the common ancestor of both classes, more than 350 million years ago (Groth-Malonek *et al.* 2007b). The exceptionally long retention of the pseudogene indicates an underlying but yet unknown function. In the three haplomitriid liverwort genera *Haplomitrium*, *Apotreubia* and *Treubia*, however, the mitochondrial *nad*7 gene is intact and functional and experiences an extremely varying degree of C-to-U RNA editing, a modification of mRNAs in plant organelles to reconstitute conserved codon identities.

2 Introduction

2.1 Peculiarities of plant mitochondrial genomes

Despite being derived from the same α-proteobacterial-like ancestor, there are many differences between animal and plant mitochondrial genomes – manifested in many aspects such as (i) complexities of the DNA molecules, (ii) sizes, (iii) numbers and orders of genes encoded, (iv) intron presences and their conservation, (v) affinity for interorganellar gene transfer and uptake of foreign DNA and (vi) the ability for and extent of RNA editing.

Regarding these traits, animal mitochondrial DNA is a small and compact molecule (with appr. 16 kb) and typically contains 37 intronless genes (13 genes for proteins necessary in the respiratory chain, 22 tRNAs and two rRNAs). Rarely, introns have been found in mitochondrial genes of enigmatic taxa representing basal lineages of the metazoa phylogeny e.g. the placozoon *Trichoplax adherens* (Burger *et al.* 2009), the bilaterian *Nephtys spec*. (Valles *et al.* 2008), some cnidaria like *Metridium senile* (Beagley *et al.* 1998) and few porifera like *Tetilla spec*. (Rot *et al.* 2006). In animals, the highly conserved gene order with only small spacer regions is in stark contrast to the high substitution rate found in the coding sequences.

In plants (Fig. 1.1), in contrast, exactly the opposite is found: In mitochondrial genomes of strikingly varying sizes (appr. 58 to more than 4000 kb) flexible gene order and gene content are closely linked with highly conserved coding sequences. There are only few examples of gene arrangements that are conserved across different land plant clades and these are usually found in early land plants (see below). In maize (*Zea mays*) it was shown that the gene order can vary greatly even between two cytotypes due to homologous recombination and the formation of smaller subcircles (Fauron *et al.* 1995). Such 'multipartite' structures are now generally recognized as a feature of angiosperm mitochondrial genomes (Sugiyama *et al.* 2005). Likewise 'early' tracheophytes such as the quillwort *Isoetes engelmannii* feature mtDNA reflecting frequent genomic rearrangements (Grewe *et al.* 2009). It is however not associated with a large genome size, as *Isoetes engelmannii* owns the smallest yet sequenced mitochondrial genome (appr. 58 kb).

The publication of several completely sequenced mitochondrial (mt) genomes has weakened the generalization that the evolution of land plants (Fig. 1.1) is accompanied by the enlargement of mi-

tochondrial DNA. Best examples are the moss *Physcomitrella patens* (105 kb, Terasawa *et al.* 2007) and the lycophyte *Isoetes engelmannii* (58 kb, Grewe *et al.* 2009) whose mt genomes are considerably smaller than that of the liverwort *Marchantia polymorpha* (187 kb, Oda *et al.* 1992), a member of the earliest land plant group. In addition, the mt genome of the gymnosperm *Cycas taitungensis* (415 kb, Chaw *et al.* 2008) is larger than that of the angiosperm *Arabidopsis thaliana* (Unseld *et al.* 1997). Even within closely related species and genera, the mt genome size can differ enormously as shown in *Cucumis melo* (muskmelon, appr. 1600 kb) and *Cucumis sativus* (cucumber, appr. 1000 kb, Ward *et al.* 1981). The largest plant mt genome known so far does not belong to a member of the quite derived Cucurbitaceae (angiosperms, eurosids I) but rather to the basal-most angiosperm *Amborella trichopoda* (appr. 4000 kb, J. Palmer, pers. comm.).

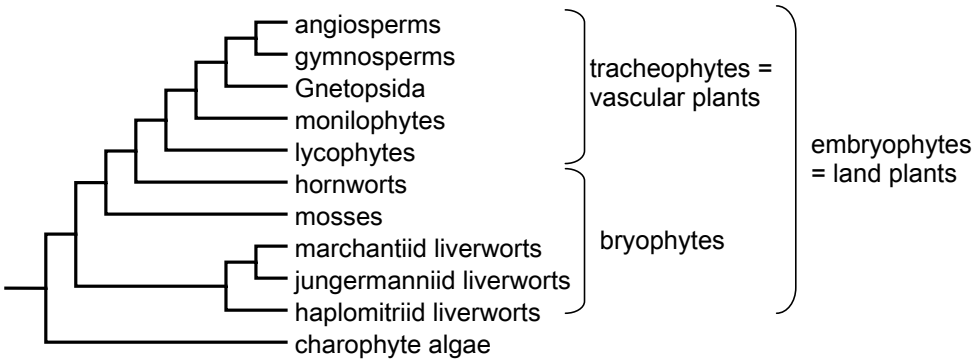


Figure 2.1: Overview of land plant evolution, rooted with charophyte algae as sister clade to all land plants, after Qiu *et al.* (2006)

An increase in genome size in general is not caused by a higher number of genes or more introns encoded. It is mainly the consequence of larger spacer regions and in part due to the uptake of 'foreign' DNA. Looking at one of only few spacer regions that are conserved between the three bryophyte divisions, the *nad*5-*nad*4 spacer (Groth-Malonek *et al.* 2007a), its size enlarges to more than 3 kb in hornworts before the gene continuity is disrupted in the early tracheophytes. In addition, the size of introns can increase. Mitochondrial introns (like the ones in the chloroplast genome) can be distinguished by their three-dimensional structure into group I and group II introns. The folding into loop and pairing regions allows close proximity of the two exons and facilitates splicing. The extension of introns usually occurs in loop 5 and 8 of group I introns and in domain IV of group II introns and can even result in breaking of genes into parts due to an interrupted transcription. Then, trans-splicing of several pre-mRNAs is necessary to assemble one continuous intron-less transcript. In angiosperms only group II introns are affected (*nad*1, *nad*2, *nad*5, Malek and Knoop 1998) and the same trans-splicing

introns are also present in gymnosperms (Chaw *et al.* 2008). The first trans-splicing group I intron was found recently in the chondriome of the lycophyte *Isoetes engelmannii* (cox1i1305, Grewe *et al.* 2009).

The number of protein-coding genes encoded by the mitochondrial genomes slightly varies in the course of land plant evolution. The highest number is encoded by the liverwort *Marchantia polymorpha* (41 genes, Oda *et al.* 1992), followed by 39 in the moss *Physcomitrella patens* (Terasawa *et al.* 2007), while most of the angiosperm mt genomes contain 29-37 protein-coding genes (*Brassica napus*, *Oryza sativa*, Handa 2003, Notsu *et al.* 2002). The smallest number of genes is found in *Megaceros aenigmaticus* with only 21 functional protein-coding genes, surprisingly eleven additional genes are retained as pseudogenes (Li *et al.* 2009). The transferred genes mostly encode matrix-soluble ribosomal proteins or tRNAs (Knoop 2004) that are then found in the nuclear genome. In angiosperms, far most of the gene transfer to the nuclear genome affect genes for subunits of the ribosomal proteins (*rps* and *rpl*, Adams *et al.* 2000, 2002). In some taxa all or most of those genes have been transferred to the nucleus (e.g. *Lachnocaulon*, *Allium*, *Phlox*, Adams *et al.* 2002), even more than once independently during angiosperm evolution (e.g. *rps*10, Adams *et al.* 2000). So far, only few genes encoding parts of the respiratory chain experienced a gene transfer. Two subunits of the succinate dehydrogenase (*sdh*3 and *sdh*4) were transferred several times independently in angiosperms (Adams *et al.* 2001), while *cox*1 (a subunit of cytochrome oxidase) encountered multiple silencing of the nuclear and mitochondrial copy in angiosperm after a single gene transfer (Adams *et al.* 1999). To retain their function, the transferred gene must acquire an N-terminal signal sequence allowing translated proteins to be imported into the mitochondrion, only known exceptions are *rps*10 nuclear copies in spinach, maize and *Oxalis* that do not possess upstream mitochondrial target sequences (Adams *et al.* 2000).

During the evolution of land plants, the amount of introns in the mitochondrial genome decreases slightly, notably with respect to group I introns (Chaw *et al.* 2008). Especially noteworthy is the complete loss of group I introns in the hornworts *Megaceros aenigmaticus* (Li *et al.* 2009) and *Phaeoceros laevis* (Xue *et al.* 2010), the gymnosperm *Cycas taitungensis* (Chaw *et al.* 2008) and angiosperms (Knoop 2004) after a decrease early in land plant evolution (seven group I introns in the liverwort *Marchantia polymorpha*, three each in the moss *Physcomitrella patens* and the lycophyte *Isoetes engelmannii*). An exception is a sporadically occurring group I intron in the *cox*1 gene of angiosperms that has been acquired independently via horizontal gene transfer (Cho *et al.* 1998, Sanchez-Puerta *et al.* 2008).

Several introns are only found in certain land plant clades. Mosses and liverworts, for instance, share only two group I and three group II introns. In very rare cases, introns seem to be gained or lost exclusively in single taxa, eg. nad5i753 is lacking so far only in the moss *Takakia ceratophylla* (this work and accession DQ268963), while the presence of nad5i391 is so far limited to the lycophyte

Huperzia selago (Vangerow *et al.* 1999). Noteworthy are also independent losses of nad5i1242 in the monilophytes *Ophioglossum* and *Equisetum* (Vangerow *et al.* 1999) and *Anemia* (Fischer 2009).

For the insertion of foreign DNA, two types can be distinguished, an intergenome gene transfer between the three genomes within the cells of one organism or the uptake of DNA fragments from other species. DNA acquired via both types is found in land plant mitochondrial genomes, although the ability to do so arose at different periods of land plant evolution.

As insertion of foreign DNA into the mitochondrial genome has not been shown for the three bryophyte divisions, it seems to be a derived feature of tracheophytes.The lycophyte *Isoetes engelmannii* for instance has incorporated two nuclear gene fragments and a 1.2 kb fragment of the chloroplast genome (Grewe *et al.* 2009). Moreover, the mt genome of the gymnosperm *Cycas taitungensis* contains a total of about 18 kb chloroplast DNA including 14 genes, that are supposed to be non-funtional pseudogenes (Chaw *et al.* 2008, Wang *et al.* 2007). In addition, a comparable amount of chloroplast (cp)-derived DNA is also found in the chondriome of angiosperms (Wang *et al.* 2007).

Some examples of horizontal gene transfer (i.e. crossing species-borders) into the mitochondrial genome have been so far identified exclusively in tracheophytes, although bryophytes can serve as donor species (Keeling and Palmer 2008, Richardson and Palmer 2007). Some species of the gymnosperm genus *Gnetum* contain a *nad*1 gene harbouring an angiosperm group II intron additional to their native *nad*1 copy (Won and Renner 2003). The occurrence of complete mitochondrial genes acquired by horizontal gene transfer was shown for e.g. *rps*2 and *rps*11 (Bergthorsson *et al.* 2003) and extended to *atp*1, which is retained as a pseudogene (Mower *et al.* 2004, Nickrent *et al.* 2004, Schönenberger *et al.* 2005). While the recipient is usually a gymnosperm or angiosperm, one example of gene transfer to the mitochondrial genome of a fern (*Botrychium*) is known (Davis *et al.* 2005). The integration of foreign DNA can also lead to the formation of chimeric mitochondrial genes. During the invasion of the *cox*1 group I intron, parts of the adjacent exon regions are also transferred (Sanchez-Puerta *et al.* 2008). In some angiosperms, the mitochondrial copy of *atp*1 contains parts of the chloroplast *atp*A gene (Hao and Palmer 2009). The most prominent example of mitochondrial-targeted horizontal gene transfer is the basal-most angiosperm *Amborella trichopoda*, whose mt genome has been sequenced lately revealing that a quarter of its large genome size contains DNA sequences from other organisms (J. Palmer, pers. comm.).

Finally, one more feature that is strikingly different between animal and plant mitochondrial genomes is the presence and extent of RNA editing. RNA editing, a post-transcriptional modification of pre-mRNA to re-create conserved amino acid codons and of other RNAs to restore structural base pairings (e.g. tRNAs), is rare in Metazoa and mostly found in the nuclear genome. An example in humans is e.g. the formation of a stop codon by cytidine to uridine exchange in the pre-mRNA of

apolipoprotein B protein in intestine, resulting in a shorter transcript (Chen *et al.* 1987, Powell *et al.* 1987). Besides C-to-U RNA editing, A-to-I (G) RNA editing occurs and is for instance necessary to regulate correct functioning of glutamate receptors and other neurotransmitter receptors in the human brain (Maas *et al.* 2009, Seeburg *et al.* 1998). In plants, RNA editing only affects plastid or mitochondrial transcripts. The C-to-U RNA editing is found in all land plants except for the Marchantiopsida (Steinhauser *et al.* 1999). 30 C-to-U RNA editing sites are present in the plastid (Tillich *et al.* 2005) and 441 in mitochondrial transcripts (Giegé and Brennicke 1999) of the angiosperm *Arabidopsis thaliana*. In hornworts, ferns and lycophytes frequent U-to-C RNA editing can be observed, sometimes exceeding the 'conventional' RNA editing (e.g. in the *nad*5 mRNA of the schizoid ferns *Anemia phyllitis* and *Lygodium flexuosum*, Fischer 2009). In addition, hornworts, ferns and lycophytes harbour the highest RNA editing frequencies (Steinhauser *et al.* 1999, Vangerow *et al.* 1999). In the lycophyte *Isoetes engelmannii* for example, the *atp*1 transcript requires RNA editing of 128 nucleotides, changing 115 out of 515 codons and thereby also removing 14 encoded stop codons (Grewe *et al.* 2009).

Recent findings suggest that proteins responsible for RNA editing in chloroplasts (Kotera *et al.* 2005) and in mitochondria (Zehrmann *et al.* 2009) all belong to a large family of so-called pentatricopeptide repeat (PPR) proteins (Lurin *et al.* 2004, Shikanai 2006) and that each protein recognizes a specific RNA editing site. Rüdinger and colleagues could show that the occurrence and diversity of a PPR-subfamily of DYW-proteins in land plant clades capable of RNA editing correlates well with their RNA editing extent (Rüdinger *et al.* 2008, Salone *et al.* 2007).

2.2 Mitochondrial genomes in liverworts and mosses

Three completely sequenced mitochondrial genomes of liverworts and mosses are currently available, from the marchantiid liverwort *Marchantia polymorpha* (Oda *et al.* 1992), the jungermanniid liverwort *Pleurozia purpurea* (Wang *et al.* 2010) and from the moss *Physcomitrella patens* (Terasawa *et al.* 2007). Only by comparing these genomes, hot spots with high recombinational activity and highly conserved regions can be detected. The two liverwort mt genomes are highly conserved, with identical gene order, high sequence similarity and less than a third of the spacer regions showing length differences of more than 200 nucleotides. The genome sizes of 169 kb and 186 kb are very similar and only two genes and one intron are missing in *Pleurozia* compared to *Marchantia* (Wang *et al.* 2010).

Comparing liverwort and moss mitochondrial genomes revealed the conserved presence of most of the genes, except for *rps*8 that is only present in the liverworts; *nad*7, a pseudogene in the two liverworts but functional in the moss; *rps*10, *trn*N(guu) and *trn*S(gcu), that are not found in the moss, but in the liverworts. With only some inversions existing (e.g. *rrn18-rrn*5*-rrn26*, *sdh*4*-nad*4L), the order of most genes are identical in both divisions (Terasawa *et al.* 2007). The intron content, however, is

remarkably different. Of 27 introns in *Physcomitrella* and 31 and 32 introns in *Pleurozia* and *Marchantia*, respectively, only five are shared between the three genomes (cox1i624g1, nad5i753g1, atp9i87g2, cox1i511g2, nad4Li283g2).

The frequency of C-to-U RNA editing is strikingly different in all three taxa. In *Marchantia*, no mitochondrial transcript requires RNA editing. *Pleurozia*, however, belonging to the Jungermanniopsida, has the ability to edit its transcripts and is, e.g. with six potential editing sites in *nad*5 and *nad*4 each, one of the lower RNA editing taxa in its clade (unpublished data). In contrast, a total of eleven RNA editing sites in nine mitochondrial genes were found in the complete chondriome of *Physcomitrella patens* (Rüdinger *et al.* 2009).

The mitochondrial genome of mosses and liverworts has a simple structure and is supposedly found as one large DNA molecule (Oda *et al.* 1992, Terasawa *et al.* 2007). Although the genome of the bacterial ancestor and the mitochondrial genomes of green algae are circular, a linear configuration for the mitochondrial genome of early land plants is possible. In *Marchantia* only 0 to 5 % of the total mitochondrial DNA is circularly arranged (Oldenburg and Bendich 1998), most of it are circularly permuted linear molecules (i.e. circular molecules that are linearised at different positions of the circle) or head-to-tail concatemers (Oldenburg and Bendich 2001). In *Physcomitrella*, linear molecules were found, but thought to be artefacts of DNA extraction (Terasawa *et al.* 2007). However, it is likely that linearized chondriomes are also dominating in the mitochondria of *Physcomitrella* and mosses in general. Nevertheless, the complex multi-partite arrangement commonly known from tracheophytes, has not been found in liverworts and mosses.

2.3 Genes, introns and spacer regions: different loci for phylogenetic analyses

The nuclear, the chloroplast and the mitochondrial genome are hosted within a plant cell and all three can be used for phylogenetic analyses, albeit for different hierarchy levels and different tasks as they differ by their nucleotide substitution rate. The nuclear genome has the highest substitution rate in plants, followed by the plastome (appr. half of the nuclear rate, Ks=1-3 x 10^{-9} substitutions per site per year, Wolfe *et al.* 1987). The substitution rate of the mitochondrial genome is a third to a fourth compared to the plastome (Ks=0.2-1 x 10^{-9}, Palmer and Herbon 1988, Wolfe *et al.* 1987). However, there are lineages with higher mitochondrial substitution rates, e.g. within the genus *Plantago* (Cho *et al.* 2004) and *Pelargonium* (Parkinson *et al.* 2005), where mitochondrial substitution rates even exceed nuclear rates. The synonymous substitution rates presented here refer to coding sequences and are much higher in non-coding regions, i.e. introns and intergenic spacers.

Within each genome, genes, introns and/or intergenic spacer regions can be explored for their phylogenetic potential. This section aims at giving only a short introduction on their properties and application of selected loci in plant phylogeny. Frequent gene duplications and losses within the nuclear genome often harbour the risk of comparing paralogous gene copies rather than orthologues (defined by descent). Therefore, only single copy genes, such as *gap*C (cytosolic glycerine aldehyde dehydrogenase, Martin *et al.* 1993) or *adh* (alcohol dehydrogenase, Small *et al.* 1998), are used to resolve genera or orders. The problem of paralogues can be avoided with high copy genes, like histone genes or genes encoding ribosomal RNAs as long as they are prone to concerted evolution resulting in maintaining identical copies (Baldwin *et al.* 1995). The nuclear large (26S) and small (18S) ribosomal RNA genes have been sequenced for large clade phylogeny (e.g. in angiosperms, Nickrent and Soltis 1995), because they are highly conserved and slowly evolving. The intergenic spacer between these genes (e.g. ITS, the internal transcribed spacer, or ETS, the external transcribed spacer, of the ribosomal DNA) yield more variability and hence sufficient information for genus or family phylogeny (Baldwin *et al.* 1995, Garcia-Jacas *et al.* 2001, Wissemann and Ritz 2005, and others). However, the analyses of more sequences revealed that concerted evolution is not a general rule in all taxa (e.g. Harpke and Peterson 2006) especially when interspecific hybridization occurs (e.g. Peterson *et al.* 2009). Very few studies have focussed on nuclear intron regions (Howarth and Baum 2002), e.g. due to their high sequence and length variabilities which impede the assignment of homologous alignment positions.

Chloroplast sequences have a long history in plant phylogenetic studies. *rbc*L (the large subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase, RuBisCO) was the first gene to be sequenced for angiosperm phylogeny (Chase *et al.* 1993). *ndh*F, *atp*B, *mat*K and *rps*4 are other examples used for inferring large clade relationships (Hilu *et al.* 2003, Olmstead *et al.* 2000, Qiu *et al.* 2005). As the chloroplast sequences are slowly evolving, changes can only accumulate over a long period of time. Hence, at lower phylogenetic level (e.g. genera or species), robust phylogenetic tree can hardly be obtained with these genes. Intergenic spacer regions, like *trn*T-L, *trn*L-F (Taberlet *et al.* 1991), *atp*B-*rbc*L (Shaw *et al.* 2005, Small *et al.* 1998) or introns, e.g. within *trn*L (Taberlet *et al.* 1991), were therefore explored with varying success.

Only few mitochondrial genes have found their way into phylogenetic analyses of seed plants. Especially in angiosperms, database entries are rare and limited to genes encoding proteins of the respiratory chain (*cox*1, *cox*2, *cob*, *atp*9). The frequent rearrangements in the mitochondrial genome of tracheophytes make the use of mitochondrial spacer regions difficult, as the gene continuity can be disrupted even between closely related genera (Ogihara *et al.* 2005, Sugiyama *et al.* 2005). In contrast, the high degree of sequence conservation allows to resolve ancient speciation events and deep-level phylogeny. As a long time is needed to establish sequence differences, relationships between land plant

divisions can be obtained with statistical confidence. In addition, the occurrence and conservation of introns within the genome add support to the phylogenetic trees. So far, mitochondrial gene and intron sequences have only been employed with larger taxon sampling in early land plants (i.e. the three bryophyte divisions, Beckert *et al.* 1999, 2001, Dombrovska and Qiu 2004, Pruchner *et al.* 2001, 2002, Qiu *et al.* 1998) and early tracheophytes (ferns and lycophytes, Groth-Malonek *et al.* 2005, Nickrent *et al.* 2000, Vangerow *et al.* 1999). Genes encoding subunits of the respiratory chain, like *nad*5, *nad*4, *nad*2 and *cox*1, are favoured especially because of their introns. As organellar introns have to maintain a specific secondary structure to enable correct splicing, conserved pairing regions alternate with variable loop regions. Hence, a variety of informative characters in variable loop regions increases the support for recent splits while conserved and slower evolving pairing regions strengthen older nodes. However, phylogenies based on mitochondrial spacer regions are still rare (Groth-Malonek and Knoop 2005, Groth-Malonek *et al.* 2007a). The first study to apply the *nad*5-*nad*4 spacer region for moss phylogeny will be presented here.

2.4 Current understanding of moss phylogeny and remaining questions

Mosses belong to the earliest land plant clades. While liverworts are the first braching division, mosses are sister to hornworts and tracheophytes (Dombrovska and Qiu 2004, Forrest *et al.* 2006, Groth-Malonek *et al.* 2005, Knoop 2004, Qiu *et al.* 2006, Samigullin *et al.* 1998). The haploid gametophyte is the major life form and divided into stemlet, leaflets and rhizoids. The classification of subdivisions and classes is based mainly on the sporophyte morphology (see Fig. 2.2A). The sporophyte itself contains the haploid spores in the spore capsule sitting on a seta (sporophyte stalk), usually protected by a calyptra (cap). The spore capsule releases the spores either by splitting into one or more vertical fissures or in the more derived mosses by dropping the lid (operculum). In operculate mosses, the open spore capsule is surrounded by one or two rings of teeth-like structures, called the peristome. The peristome is made of hygroscopic tissue that allows movements to inhibit or facilitate spore release. Its morphology can be applied to distinguish major moss lineages. As the hierarchical levels change through time and studies, not necessarily based on novel data, a recent systematic treatment (Goffinet *et al.* 2008) is used as reference here and figure 2.3 summarizes the currently acknowledged moss phylogeny.

Mosses of the monogeneric subdivisions Sphagnophytina, Takakiophytina, Andreaeophytina and Andreaeobryophytina do not possess a peristome. Except for the genus *Sphagnum*, whose spore capsule opens by removing a lid, the spores are released through vertical fissures. Either Sphagnophytina or Takakiophytina or both forming a sister clade are the earliest branching moss subdivisions. Andreaeo-

phytina and Andreaeobryophytina are assembled in a monophyletic group and split off next. The fifth subdivision, Bryophytina, contains most of the moss variety and is further divided into three classes. Polytrichopsida and Tetraphidopsida belong to the so-called nematodontous mosses. Their peristome ring is built by layers of whole cells in contrast to the peristome of arthrodontous mosses constructed out of thickened remnants of cell walls. Depending on further features of the peristome, the arthrodontous Bryopsida are traditionally separated into five (or more) subclasses: Buxbaumiidae and Diphysciidae (both monogeneric), Dicranidae, Bryidae (sometimes splitted into a paraphyletic Bryidae and monophyletic Hypnidae) and Funariidae. The peristome of the Dicranidae consists of only one ring of teeth, with two teethlets on the tip of each haplolepideous tooth (Fig. 2.2B). Bryidae and Funariidae both have two peristome rings. In the Bryidae, the diplolepideous teeth (Fig. 2.2C) of the exostome alternate with the endostome teeth (diplolepideous alternate peristome), while they face each other in the Funariidae (diplolepideous opposite peristome).

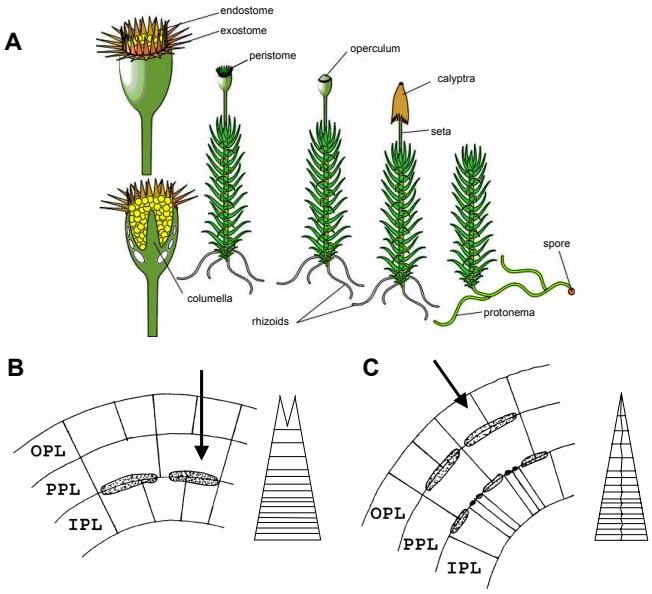


Figure 2.2: A) Overview of moss life cycle showing a diplolepideous, diplostomous (two rings) peristome with opposite teeth with morphological characters indicated, modified after www.biologie.uni-hamburg.de, B) and C) schematic view onto arthrodontous peristome, OPL = outer peristomal layer, PPL = primary peristomal, IPL = inner peristomal layer, arrow indicate view axis, modifed after Magombo (2003), B) peristome with thickened cell walls between PPL and IPL and view onto outer surface of a haplolepideous (i.e. one-scale) tooth, C) diplolepideous alternate peristome with thickened cell walls between OPL and PPL (forming the exostome) and PPL and IPL (forming the endostome) with view onto outer surface of an diplolepideous (two-scale) exostome tooth.

To explore the relationships between the five subdivisions and within the Bryophytina, several molecular loci have been employed. Large clade phylogenies based on the nuclear 18S rDNA, plastid

*rbc*L and *rps*4 genes and the mitochondrial *nad*5 and *nad*2, single or in combination have been reconstructed (e.g. Beckert *et al.* 1999, 2001, Bell *et al.* 2007, Buck *et al.* 2000, Cox *et al.* 2004, Goffinet *et al.* 2001, Hedderson *et al.* 1996, Magombo 2003, Newton *et al.* 2000).

In recent years a consensus appears to emerge from the phylogenetic studies. Most of the relationships within the Dicranidae, Bryidae and Funariidae, as well as within the Polytrichopsida receive very good statistical support with different molecular markers and are thus commonly acknowledged (e.g. Bell and Newton 2005, Goffinet and Cox 2000, Hedderson *et al.* 2004, Hyvönen *et al.* 2004). The position of some enigmatic taxa is still debatable, as it is either varying depending on the molecular markers used or receives only low statistical support.

Catoscopium, for instance, belonging to the monogeneric family of Catoscopiaceae is traditionally placed within the Bryales, because of its diplolepideous peristome. Chloroplast data supported different placements either excluded from the Bryales, with a closer, although unsupported, affinity to the outgroup of the study (Virtanen 2003), within a clade containing Funariidae and Dicranidae (Goffinet *et al.* 2001) or a basal position in the Dicranidae (Hedderson *et al.* 2004).

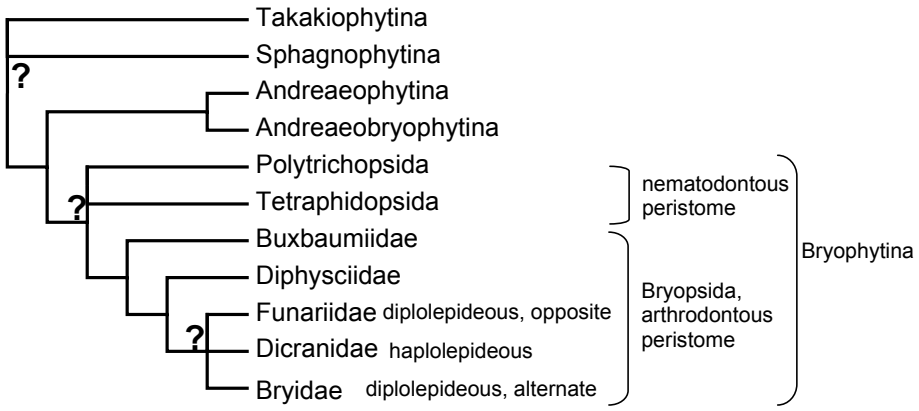


Figure 2.3: Cladogram reflecting the current understanding of moss phylogeny, unclear relationships are shown as polytomies and indicated with a question mark (modified after Goffinet *et al.* 2008).

Timmiella is still placed into the Pottiales (Goffinet *et al.* 2008), although molecular data especially from the chloroplast revealed a position at the base of the Dicranidae, next to *Catoscopium* (Hedderson *et al.* 2004).

The peristome of members of the genus *Timmia*, contains two rings of teeth. Hence, this genus was traditionally classified as a Bryidae. Molecular data, however, suggested a phylogenetic position among the basal-most arthrodontous mosses (Cox *et al.* 2004, Goffinet *et al.* 2001, Newton *et al.* 2000). A subsequent morphological investigation of the *Timmia* peristome revealed its unique nature (Budke *et al.* 2007). The symmetry of divisions forming the endostome is similar to the Funariidae, however

the number of endostome teeth suggest an affinity to the Bryidae (Budke *et al.* 2007). This intermediate morphology also reflects the still unsupported assignment within the Bryopsida.

Mosses with diplolepideous opposite or lacking peristomes are usually put into the subclass Funariidae. That this clade (containing Gigaspermaceae, Funariaceae, Disceliaceae, Encalyptaceae and Timmiaceae) is not monophyletic, was shown by several studies (Beckert *et al.* 1999, 2001, Cox *et al.* 2004, Goffinet *et al.* 2001, Newton *et al.* 2000). Even when *Timmia* is excluded, the question of monophyly remains. The support for a closer relationship of Disceliaceae, Funariaceae and Encalyptaceae emerged in recent years (Cox *et al.* 2004, Goffinet and Cox 2000) and is supported by a 71 kb inversion in the chloroplast genome of all members (Goffinet *et al.* 2007). A sister relationship with the Gigaspermaceae is suggested but not statistically supported (Cox *et al.* 2004, Goffinet and Cox 2000). In addition, the chloroplast inversion is absent in the Gigaspermaceae (Goffinet et al., 2007).

The relationship of the three species-rich Bryopsida subclasses Bryidae, Funariidae and Dicranidae is far from being resolved. Several studies attempting to address this issue with varying taxon samplings resulted in different, weakly supported topologies with very short branches at the backbone of the phylogenetic trees (Beckert *et al.* 1999, 2001, Cox *et al.* 2004, Goffinet *et al.* 2001, Magombo 2003, Newton *et al.* 2000, Stech and Frey 2008).

A putative sister relationship between the two nematodontous classes Tetraphidopsida and Polytrichopsida is sometimes suggested using mitochondrial loci but with low branch reliability (Beckert *et al.* 2001). Chloroplast markers resolve a likewise unsupported grouping of *Tetraphis* and *Buxbaumia* (Goffinet *et al.* 2001, Magombo 2003, Newton *et al.* 2000). A third hypothesis places *Tetraphis* as sister to all other Bryopsida (Cox *et al.* 2004).

The basal-most moss genus still remains to be found. It requires the choice of an outgroup from either liverworts or hornworts, the two adjacent land plant clades. Therefore, the intron or spacer region investigated has to be conserved in position and has to have retained enough similarity for an unambiguous sequence alignment. A monophyletic clade of *Takakia* and *Sphagnum* is resolved at the base of the mosses with nuclear and chloroplast loci analysed (Cox *et al.* 2004, Newton *et al.* 2000, Qiu *et al.* 2006): a grouping, that is not seen when only mitochondrial markers are used (Dombrovska and Qiu 2004). In addition, morphological and biochemical studies hint to a separation of *Takakia* from all other mosses due to its unique combination of moss and liverwort characters (Smith and Davison 1993) and the use of different secondary metabolites (R. Mues, pers. comm., Asakawa 2004, Markham and Porter 1979, Markham and Given 1988).

Novel mitochondrial loci were established to specifically address these open questions. A variable spacer region (*nad*5-*nad*4 intergenic spacer) and one group I intron (cobi420), conserved in all mosses, have been investigated to clarify the placement of several taxa within the class of Bryopsida. Another

group I intron, cox1i624, conserved in mosses and liverworts, was explored with focus on the position of early branching moss and liverwort genera. The molecular evolution of group I intron structures and the *nad*5-*nad*4 spacer region within the mosses and between mosses and liverworts was also investigated to further support monophyletic taxa through the occurrence of synapomorphic indel or duplication events.

2.5 Molecular evolution in liverwort chondriomes

A comparison of the two completely sequenced mitochondrial genomes of the jungermanniid liverwort *Pleurozia purpurea* (Wang *et al.* 2010) and the marchantiid liverwort *Marchantia polymorpha* (Oda *et al.* 1992) revealed a low degree of sequence variation and slowly evolving genome structures. The deficient amount of sequence differences especially within the Marchantiopsida is also reflected by short branches, low resolution and weak support in liverwort phylogenies.

In several multi-taxa and multi-locus studies a consensus phylogeny with three or four classes emerged (see Fig. 2.1, Crandall-Stotler *et al.* 2005, 2009, Davis 2004, Forrest *et al.* 2006, Heinrichs *et al.* 2005). The basal-most clade, the class of Haplomitriopsida, contains the three genera *Haplomitrium*, *Treubia* and *Apotreubia*, found on long, isolated branches (sometimes further divided into Treubiopsida and Haplomitriopsida). This clade is sister to a joint group of Marchantiopsida and Jungermanniopsida. The Marchantiopsida embrace all complex-thalloid liverworts, with *Blasia* being the first-branching genus. The Jungermanniopsida unite two simple-thalloid (Fossombroniales and Metzgeriales or a paraphyletic Metzgeriidae) and two leafy orders (Porellales and Jungermanniales or Jungermanniidae).

An assumed slow evolution of the mitochondrial genome in liverworts is indicated by the loss of only two tRNA genes and one intron when comparing the two completely sequenced chondriomes of *Marchantia* and *Pleurozia* (Wang *et al.* 2010). Further support for this hypothesis comes from short branches in mitochondrial loci based phylogenies, especially affecting the Marchantiopsida (e.g. Beckert *et al.* 1999, Forrest *et al.* 2006). However, the loss of *trn*T(ggu) in *Pleurozia* is accompanied by a smaller intergenic spacer region between the adjacent *trn*A and Ψ-*nad*7 genes (Wang *et al.* 2010). In addition, a large deletion at the 3' end of the *nad*7 pseudogene shortens its size, while *Marchantia* retained the full length (Wang *et al.* 2010).

The transfer of the functional *nad*7 gene into the nuclear genome is an exceptionally rare event, especially as a pseudogene copy is retained in the chondriome during most of the liverwort evolution (Groth-Malonek *et al.* 2007b). Although genes for ribosomal proteins are frequently transferred in angiosperms, only few genes encoding proteins of the respiratory chain were affected (Adams *et al.* 2002). Subunits of the succinate dehydrogenase have been transferred to the nuclear genome in the gym-

nosperm *Cycas taitungensis* (*sdh*4, Chaw *et al.* 2008) and several times independently in angiosperms (*sdh*3 and *sdh*4, Adams *et al.* 2001) and in legumes, a gene transfer of *cox*2 (a subunit of the cytochrome oxidase) has been discovered (Adams *et al.* 1999). However, when the pattern of mitochondrial gene loss in angiosperms was assessed, *nad*7 was found to reside in the mitochondrial genome in all taxa investigated so far (Adams *et al.* 2002). The only other known loss of *nad*7 occurred in hornworts. In *Megaceros aenigmaticus* and *Phaeoceros laevis nad*7 is missing from the mitochondrial genome (except for 137 bp of the first exon), while eleven other pseudogenes (nine of them shared between both species) are still present (Li *et al.* 2009, Xue *et al.* 2010). To date, the liverworts are the only known land plant group that retained the *nad*7 pseudogene with its two introns after a functional gene transfer to the nuclear genome (Groth-Malonek *et al.* 2007b, Kobayashi *et al.* 1997). Surprisingly, the pseudogene is still transcribed in *Marchantia* but the introns are not correctly spliced (Takemura et al., 1995). While a rapid degradation is the assumed fate of pseudogenes in the mitochondrial genome, the question of an underlying cause of its long-term retention in liverworts arises. However, considering the inferred extremely slow evolution in mitochondrial genomes of liverworts in conjunction with a similar retention of pseudogenes in hornworts, even several hundred million years might not have been enough time to dispose of the *nad*7 pseudogene.

In the light of RNA editing, the mitochondrial genome of liverworts is far from being conservative in its evolution (Malek *et al.* 1996, Steinhauser *et al.* 1999). In the marchantiid liverworts, no RNA editing was found to be necessary as conserved amino acid codons are already present. In the Jungermanniopsida, RNA editing is required and varies between species and especially between the two clades of leafy and simple-thalloid liverworts. While more RNA editing occurs in the simple thalloids, less is observed in the leafy liverworts (Groth-Malonek *et al.* 2005, Steinhauser *et al.* 1999). In the basal-most Haplomitriopsida, a surprisingly high amount of RNA editing can be found in *Haplomitrium*, exceeding known RNA editing frequencies in liverworts and mosses (Groth-Malonek *et al.* 2005, Rüdinger *et al.* 2008). Hence, the ability for RNA editing, acquired by the ancestor of all land plant clades, is lost at least once in the Marchantiopsida (Groth-Malonek *et al.* 2007b, Rüdinger *et al.* 2008).

This study focusses on an ancient gene arrangement of *trn*A-*trn*T-*nad*7, conserved in algae, liverworts, mosses and hornworts, that displays high recombinational activity. The *nad*7 gene, part of this gene cluster, is a pseudogene in most liverworts, remaining in the chondriome after a successful gene transfer to the nuclear genome. The fate of the pseudogene and its functional mitochondrial counterpart will be examined. The finding presented here will challenge the assumed slow molecular evolution of liverwort chondriomes.

3 Results

3.1 Novel mitochondrial markers for moss phylogeny

3.1.1 Ute Wahrmund, Theresia Rein, Kai F. Müller, Milena Groth-Malonek and Volker Knoop (2009): Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. *Plant Systematics and Evolution* 282 (3): 241-255

A novel mitochondrial locus was established and its phylogenetic potential assessed. For the first time a plant mitochondrial spacer region was used for phylogenetic analyses in an exemplary study of mosses. The *nad*5-*nad*4 gene arrangement is conserved in liverworts, mosses and hornworts. In mosses, its size ranged between 900 and 1100 bp, with the longest spacer regions found in the earlier branching moss lineages and an observed size decrease during moss evolution. The phylogenetic tree based on the *nad*5-*nad*4 spacer alone showed a good resolution, but statistical support was reached only for subclades, mostly within the Bryidae (see Fig. 1 in the article). The combined data set containing mitochondrial *nad*5 (incl. nad5i753 group I intron), *nad*2 (incl. nad2i156 group II intron) and plastid *rbc*L and *rps*4, however, resulted in a well supported phylogenetic tree (Fig. 3). Especially noteworthy is the placement of *Timmiella* as basal-most genus of the Dicranidae and a possible, yet unsupported monophyly of the nematodontous mosses *Tetraphis* and Polytrichopsida. In addition, monophyly of the Funariales, possibly including *Timma*, was suggested albeit in the absence of any member of the Gigaspermales. A polytomy prevented further insights into the relationships between Funariidae, Bryidae and Dicranidae. A comparison of resolution and robustness (in terms of statistical support) of each of the five single data set trees revealed that *nad*5 including nad5i753 group I intron exhibited a resolution comparable to the fusion tree with statistical support values only slightly decreased.

Contribution: I extended a given *nad*5-*nad*4 spacer data set by 28 taxa, also adding missing sequences of the other four loci. I prepared the alignments, calculated and interpreted the phylogenetic trees and helped writing the manuscript.

Plant Syst Evol
DOI 10.1007/s00606-008-0043-x

ORIGINAL ARTICLE

Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci

Ute Wahrmund · Theresia Rein · Kai F. Müller · Milena Groth-Malonek · Volker Knoop

Received: 2 September 2007 / Accepted: 28 March 2008

Abstract Given the frequent genomic recombinations in plant mitochondrial DNA, intergenic regions of this organelle genome had so far not been considered as loci of potential phylogenetic information. Based on the recent evidence that an evolutionary ancient mitochondrial *nad*5-*nad*4 gene continuum is conserved in bryophytes we have compiled a dataset for a phylogenetically wide sampling of 50 mosses covering this intergenic region. The length of the intergenic region was generally in the range of 585 (*Diphyscium*) to 646 bp (*Tomentypnum*) with rare exceptions, for example all Polytrichales taxa showing a 200 bp deletion as an apparent synapomorphy of this order. Phylogenetic information in the novel marker sequence was compared with that of a *nad*5 gene region containing a group I intron and a *nad*2 gene region containing a group II intron as well as with two widely sampled chloroplast data sets, *rbc*L and *rps*4. Indel evolution in the three types of non-coding mitochondrial sequences is obviously more taxon-dependent than locus-dependent, indicating lineage-specific insertion/deletion rates. For example, larger sequence deletions are a general feature in *Schistostega* and *Tetraphis*. Although confidence for particular nodes in the phylogeny was found to vary among data sets, gene trees were essentially without conflict with respect to well-supported ones and add up in information towards a reasonably well-resolved moss phylogeny. However, while a consensus on the latter is clearly emerging, sufficient confidence is still lacking for the first dichotomies among the arthrodontous mosses leading into subclasses Bryidae, Dicranidae and Funariidae and the relative placement of nematodontous mosses (Polytrichales and Tetraphidales) on the backbone of early moss phylogeny.

Keywords Bryophytes · Group I intron · Group II intron · Indels · Intergenic region · Mitochondrial DNA · Mosses · Phylogeny · RNA editing

Abbreviations

bp	Base pairs
mtDNA	Mitochondrial DNA

Introduction

An overwhelming majority of phylogenetic studies among plants above family level has relied on studies of loci from the chloroplast genome. As in other plant groups,

U. Wahrmund · T. Rein · M. Groth-Malonek · V. Knoop (✉)
IZMB, Institut für Zelluläre und Molekulare Botanik,
Abt. Molekulare Evolution, Universität Bonn, Kirschallee 1,
53115 Bonn, Germany
e-mail: volker.knoop@uni-bonn.de

U. Wahrmund
e-mail: ute.wahrmund@uni-bonn.de

T. Rein
e-mail: theresia.rein@gmx.de

M. Groth-Malonek
e-mail: groth-malonek@uni-bonn.de

K. F. Müller
Nees-Institut für Biodiversität der Pflanzen, Universität Bonn,
Meckenheimer Allee 170, 53115 Bonn, Germany
e-mail: kaimueller@uni-bonn.de

the results from molecular phylogenetics and systematics have provided many novel insights for a natural and cladistically informative taxonomy for mosses (Goffinet and Buck 2004), one of the classes of bryophytes. For the analysis of bryophytes in general and mosses in particular we have explored the phylogenetic utility of mitochondrial DNA to complement the widely used chloroplast data sets (Beckert et al. 1999, 2001). We found that the mitochondrial *nad*2 gene (harbouring nad2i156, a group II intron universally conserved among mosses) and the *nad*5 gene (featuring nad7i753, a group I intron conserved between mosses and liverworts), which encode subunits of complex I, the NADH-ubiquinone-oxidoreductase, deliver congruent and complementary phylogenetic information (Beckert et al. 2001). The *nad*5 group I intron nad5i753 in particular has subsequently been appreciated as a locus contributing phylogenetic information also in several multigene studies of early land plant phylogeny (e.g. Karol et al. 2001; Cox et al. 2004; Davis 2004; Hyvönen et al. 2004; Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005; Bell et al. 2007; Quandt et al. 2007).

In contrast to chloroplast DNA, the plant mitochondrial DNA, at least in angiosperms, is frequently rearranged and consequently gene linkages are only rarely conserved (Knoop 2004). Therefore, intergenic regions in plant mitochondrial DNA have so far not been considered as phylogenetically informative regions. Recently, however, we found that the gene order of three genes, *nad*5-*nad*4-*nad*2 is universally conserved among early land plants across all three bryophyte lineages—hornworts, liverworts and mosses—and that sizes of around 600 bp for the *nad*5-*nad*4 intergenic region in mosses may make this locus an interesting candidate as a first plant mitochondrial intergenic region for phylogenetic analyses (Groth-Malonek et al. 2007). To explore its phylogenetic utility we have now compiled a *nad*5-*nad*4 spacer data set for 50 mosses with a wide taxonomic sampling. For comparison with the previously compiled data sets of the *nad*5 and *nad*2 genes with their different types of structured organellar introns (group I and group II) and with the widely sampled chloroplast coding regions of *rbc*L and *rps*4, we have filled taxonomic gaps with the aim to compare phylogenetic information.

Phylogenetic utility of indels (for a review on the topic see e.g. Müller 2006; Simmons et al. 2007) has been the focus of various studies utilizing nuclear or chloroplast (cp) phylogenetic markers, but not for plant mitochondrial (mt) DNA. The compilation of three parallel noncoding mt datasets in this study provides the opportunity not only to address phylogenetic information in indels from these markers, but also to compare length mutational patterns across noncoding mtDNA regions.

Methods

Molecular work

Total nucleic acids were extracted either using the Plant DNeasy Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol or in CTAB buffer (Cetyl-trimethyl-ammonniumbromide) followed by phenol/chloroform extraction and isopropanol precipitation. PCR amplifications were performed using primer pair n5rb2F (5′-ggtgctattgaaatcttgggtcc-3′) and n4lb2R (5′-acaaagaataamgagataccatctata cc-3′) for of the mitochondrial *nad*5-*nad*4 intergenic spacer, primer pair K (5′-atatgtctgaggatccgcatag-3′) and L (5′-aactttggccaaggatcctacaaa-3′) for the mitochondrial *nad*5 gene region containing group I intron nad5i753, primer pair nad2up (5′-ggagttgtntttagtacctctaa-3′) and nad2do (5′-agtagtaacgayttntcacgatccat-3′ for the mitochondrial *nad*2 gene region containing group II intron nad2i156 and primer pair rbclF (5′-gcagcatttcgratgactc-3′) and rbclR (5′-ctcattacgggcttgtacac-3′) for the chloroplast *rbc*L gene region. Where already available for a given genus sequences of these loci were taken from the NCBI database as indicated in Table 1. Primers for the *nad*5-*nad*4 intergenic region were designed to cover 226 and 254 bp of the flanking *nad*5 and *nad*4 genes, respectively. The PCR amplification assay usually contained 2.5 µl 10 × PCR buffer, 250 µM of each dNTP, 1 µM of each primer, 1 U DNA polymerase and double-distilled water added up to 25 µl. Either Genaxxon (Biberach, Germany) Taq polymerase E or the TripleMaster PCR System (Eppendorf, Hilden, Germany) was used with their respective buffer as supplied. PCR was performed with an initial denaturation step of 3 min 94°C, followed by 10 touch down cycles with a stepwise decrease of hybridization temperature from 50 to 42°C held for 30 s and synthesis steps of 72°C for 3 min, continued for 30 further cycles at the 42°C hybridization temperature and finally by a terminal elongation at 72°C for 7 min. Gel purified PCR products were cloned into the pGEM-T Easy Vector (Promega, Mannheim, Germany) and clones were commercially sequenced at Macrogen, Inc. (Seoul, Korea).

Phylogenetic analyses

Sequence handling, alignment and distance-based phylogenetic analyses were done with the MEGA 3.1 software (Kumar et al. 2004). Sequences were prealigned using the CLUSTAL algorithm implemented in MEGA and then adjusted manually for proper indel placement to place indels as oligonucleotide motif duplication or deletions where obvious. Bootstrapping was done with 10,000 replicates to determine node reliabilities. Distance-based phylogenies were calculated using MEGA 3.1 with the

Table 1 List of moss species and the respective accession numbers of the five loci used for molecular phylogenetic reconstructions

Taxonomy superorder clade order family	Species	*nad*5-*nad*4 spacer (size in bp)	*nad*5	*nad*2	*rbc*L	*rps*4
Sphagnales	*Sphagnum fallax*	DQ098673 (638)	AJ001225	AJ299524	AB013673	AY309730 *S. squarrosum*
Takakiales	*Takakia lepidozioides*	EU095268 (631)	AJ291553	AJ299525	AF244565	AF306950
Andreaeales	*Andreaea nivalis*	DQ098672 (611)	AJ299526	AJ299526	AF478198 *A. nitida*	AJ617675 *A. rupestris*
Polytrichales	*Atrichum undulatum*	EU095269 (416)	AJ001229	AJ299527	AY118236	AY137681
	Dawsonia spec.	EU095270 (411)	AY908804 *D. superba* (1111)	EU095309	AF208410 *D. papuana*	AF208419 *D. papuana*
	Oligotrichum hercynicum	EU095271 (416)	AY908805 *O. parallelum* (1112)	EU095310	AY118242 *O. austroaligerum*	AY137688
	Pogonatum urnigerum	EU095272 (413)	AJ291554	AJ299528	AF478206 *P. perichaetiale*	AF478258 *P. perichaetiale*
Tetraphidales	*Tetraphis pellucida*	EU095273 (622)	AJ224855	AJ299529	AF478203	AF478251
Buxbaumiales	*Buxbaumia aphylla*	EU095274 (628)	AJ291555	AJ299531	AF478212	AF231897
Diphysciales	*Diphyscium sessile*	EU095275 (585)	Z98972	AJ299530	AF478219 *D. fasciculatum*	AF478270 *D. fasciculatum*
Funariidae						
Timmiales	*Timmia bavarica*	EU095276 (621)	Z98963	AJ299532	AF478242 *T. megapolitana*	AF222902 *T. megapolitana*
Encalyptales	*Bryobrittonia longipes*	EU095277 (618)	AY908790 (1097)	EU095311	AJ275168	AF023778
	Encalypta streptocarpa	EU095278 (627)	AJ291556	AJ299533	AF478239	AF478282
Funariales	*Funaria hygrometrica*	EU095279 (618)	Z98959	AJ299534	AF005513	AF023776
	Physcomitrella patens	DQ098674 (618)	Z98960	AJ299535	AP005672	NC_005087
	Physcomitrium pyriforme	EU095280 (617)	AY908933 *P. lorentzii* (948)	EU095312	EU095319	AF223045
Disceliaceae	*Discelium nudum*	EU095281 (623)	AY908956 (1088)	EU095313	EU095320	AF223063

Table 1 continued

Taxonomy superorder clade order family	Species	*nad*5-*nad*4 spacer (size in bp)	*nad*5	*nad*2	*rbc*L	*rps*4
Dicranidae						
Scouleriales Drummondiaceae	*Drummondia prorepens*	EU095282 (622)	AY908926 *D. obtusifolia* (1275)	n.a.	AF232697 *D. obtusifolia*	AF306977
Grimmiales	*Coscinodon cribrosus*	EU095283 (623)	AY908918 *C. calyptratus* (1088)	EU095314	AB125575	AJ553978
	Grimmia donniana	EU095284 (622)	AY908919 *G. plagiopodia* (1097)	EU095315	AF231305 *G. pulvinata*	AF222900 *G. pulvinata*
	Racomitrium lanuginosum	EU095285 (621)	AJ291561	AJ299542	AB125582 *R. japonicum*	AJ553982
Seligeriaceae	*Blindia acuta*	EU095286 (625)	AY908928 (1131)	EU095316	AF478232 *B. magellanica*	AF478278 *B. magellanica*
Dicranales	*Orthodicranum montanum*	EU095287 (619)	AJ291558	AJ299537	AF231311 *O. fulvum*	AF231288 *O. fulvum*
Fissidentaceae	*Fissidens cristatus*	DQ098675 (631)	Z98954	AJ299541	AF226810 *F. mooreae*	AF223056 *F. subbasilaris*
Ditrichaceae	*Ceratodon purpureus*	EU095288 (619)	Z98955	AJ299538	EU095321	AJ554004
	Ditrichum cylindricum	EU095289 (621)	AJ291559	AJ299539	AF231080 *D. ambiguum*	AJ554009 *D. pusillum*
Schistostegaceae	*Schistostega pennata*	EU095290 (532)	AJ224856	AJ299546	AY631206	AF265359
Leucobryaceae	*Leucobryum glaucum*	EU095291 (612)	AJ291560	AJ299540	AB124788	AJ554003
Pottiales	*Pottia truncata*	EU095292 (620)	Z98957	AJ299543	AB125592 *P. intermedia*	AF480987 *P. pallida*
	Timmiella spec.	EU095293 (621)	AY908958 *T. anomala* (1096)	EU095317	AF478236 *T. crassinervis*	AY908163 *T. anomala*
	Tortula latifolia	EU095294 (620)	AJ291562	AJ299544	AF226823 *T. obtusissima*	AF481041 *T. muralis*
Cinclidotaceae	*Cinclidotus riparius*	EU095295 (621)	AJ291563	AJ299545	AF231079 *C. mucronatus*	AF480975 *C. fontinaloides*
Bryidae						
Splachnales	*Splachnum ampullaceum*	EU095296 (619)	EU095308	EU095318	AF231071	AJ251308

Table 1 continued

Taxonomy superorder clade order family	Species	*nad*5-*nad*4 spacer (size in bp)	*nad*5	*nad*2	*rbc*L	*rps*4
Orthotrichales	*Ulota crispa*	EU095297 (617)	AJ291568	AJ299553	AY631208	AY618370 *U. hutchinsiae*
Hedwigiales	*Hedwigia ciliata*	EU095298 (622)	Z98966	AJ299554	AF005517	AF478289
Rhacocarpaceae	*Rhacocarpus purpurascens*	EU095299 (620)	Z98967	AJ299555	AJ275171	AF023815
Bryales Bartramicaceae	*Bartramia halleriana*	EU095300 (623)	Z98961	AJ299547	AF231090	AF265358
	Plagiopus oederi	EU095301 (622)	Z98962	AJ299548	DQ481540	AF023833
Mniaceae	*Mnium hornum*	EU095302 (621)	AJ291567	AJ299552	AF226820	AF023796
	Pohlia nutans	EU095303 (624)	AJ291565	AJ299550	AJ275175 *P. cruda*	AF023795 *P.cruda*
Aulacomniaceae	*Aulacomnium androgynum*	EU095304 (621)	AJ291564	AJ299549	AJ275180 *A. turgidum*	AF023809 *A. turgidum*
Orthodontiaceae	*Orthodontium lineare*	EU095305 (620)	AJ291566	AJ299551	AJ275174	AF023800
Hypnales	*Herzogiella seligeri*	DQ098681 (643)	AJ291573	AJ299561	EU095322	AF469815 *H. striatella*
Fontinalaceae	*Fontinalis antipyretica*	EU095306 (643)	AJ291570	AJ299558	AB050949	AF023817
Amblystegiaceae	*Hygrohypnum ochraceum*	DQ098679 (643)	AJ291574	AJ299562	EU095323	AY908620 *H. smithii*
	Scorpidium scorpioides	DQ098680 (643)	AJ291575	AJ299563	EU095324	AY908584
Brachytheciaceae	*Tomentypnum nitens*	DQ098677 (646)	AJ291572	AJ299560	AB024676	AY908567 *T. falcifolium*
Leucodontaceae	*Pterogonium gracile*	EU095307 (393)	Z98968	AJ299556	AY631194	AY907970

Table 1 continued

Taxonomy superorder clade order family	Species	*nad5-nad4* spacer (size in bp)	*nad5*	*nad2*	*rbc*L	*rps*4
Neckeraceae	*Homalia trichomanoides*	DQ098683 (644)	AJ291569	AJ299557	EU095325	AY908276
	Thamnobryum alopecurum	DQ098678 (643)	AJ291571	AJ299559	AY532392	AF023834

Systematic designations are according to a recent comprehensive classification (Goffinet and Buck 2004). Most entries for the *nad5-nad4* intergenic region (EU095268-EU095307) have been determined in the course of this study, some (DQ098672-DQ098683) had been determined before (Groth-Malonek et al. 2007). To fill in taxonomic gaps for the other loci, one new *nad5* sequence (accession EU095308), ten new *nad2* sequences (entries EU095309-EU095318) and seven new *rbc*L sequences (EU095319-EU095325) were newly determined and deposited in the database. We were unable to obtain the *nad2* amplicon (n.a.) for *Drummondia*. All other accessions were retrieved from the databases, occasionally for alternative species of the same genus, as indicated

Tamura 3-parameter nucleotide substitution model and pair wise deletion of gaps. Phylogenetic analyses with maximum parsimony were done with Paup 4.0b10 (Swofford 2003) using stepwise taxon addition, 10 random sequence additions, tree bisection and reconnection (TBR). Bayesian phylogenetic analyses were conducted with MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) using the general time reversible model of sequence evolution with gamma distributed rates and a proportion of invariant sites (GTR + Γ + I) for 10,000,000 generations, every 10th tree sampled in two parallel runs. The burnin was manually determined generously as 10,000 trees by which stationarity in the distribution of the log likelihoods of both runs had been reached. Phylogenetic trees were determined for each of the five data sets separately and for the fused five-loci-alignment. The Bayesian tree of the fused data matrix was calculated with the same model as above but divided into five separate partitions (chloroplast coding, mitochondrial coding, intergenic region, group I intron and group II intron, respectively) with the rate priors variable for all partitions and the following parameters unlinked: stationary nucleotide frequencies (statefreq), substitution rates of the GTR rate matrix (revmat), shape parameter of gamma distribution (shape) and proportion of invariable sites (pinvar). Alignment sizes were 1,222 positions for the *nad5*-*nad4* region, 2,126 positions for *nad5* including nad5i753, 2,356 positions for *nad2* including nad2i156, 1,293 positions for *rbc*L and 552 positions for *rps*4, respectively. Data matrices and trees have been submitted to TreeBASE under accession number S1977.

Indel coding, reconstruction of ancestral indel states, and statistical analysis

Indels were coded using SeqState v 1.33 (Müller 2005) according to simple indel coding (SIC, Simmons and Ochoterena 2000) and modified complex indel coding (MCIC, Müller 2006). The contribution to the overall phylogenetic signal was tested by the effect of including indel information on Bremer support. Since each of the three datasets alone contributes only a limited number of non-homoplastic informative characters, the many equally parsimonious trees accumulated by PAUP slowed down conventional tree searches during Bremer support calculation. Therefore, the parsimony ratchet approach for decay analysis as implemented in PRAP (Müller 2004) was used. Based on the topology from Fig. 3 (combined markers), ancestral indel character states were output for each node with help of Mesquite (Maddison and Maddison 2003), and parsed with a Perl script written for this study (available from K.M.). The script was used to determine unambiguous state transformations for each indel character on each branch (i.e. between two nodes), to determine which of

these were insertions rather than deletions and how many nucleotides were involved, and to output lists with frequency and lengths of insertions and deletion per branch, ready for subsequent statistical analysis. As a nonparametric and non-linear measure of correlation, Spearman's Rank Correlation Coefficient was calculated and significance assessed via Spearman's *t* test.

Results and discussion

The *nad*5-*nad*4 spacer as a novel mitochondrial marker locus for phylogenetic analysis

The gene continuity between *nad*5 and *nad*4 initially found to be conserved in the mitochondrial DNA across the bryophyte lineages (Groth-Malonek et al. 2007) turned out to be conserved in all 50 mosses of our taxon sampling (Table 1). Whereas the *nad*5-*nad*4 spacer sequence is highly divergent and unalignable between mosses and liverworts, the moss sequences of our taxonomically wide sampling could be unambiguously aligned over 742 positions, flanked by indel-free *nad*5 and *nad*4 coding regions of 226 and 254 bp, respectively, in the amplicon (excluding oligonucleotide primers). A phylogenetic tree based on the alignment of the *nad*5-*nad*4 amplicon alone is shown in Fig. 1. The overall topology of the tree is in accordance with current insights about moss phylogeny, which we will discuss below in detail in the context of fused data sets (Fig. 3; Table 3).

In the *nad*5-*nad*4 spacer, indels have the potential to contribute phylogenetic signal. Many indels in the *nad*5-*nad*4 intergenic region, however, so far appear as autapomorphies in our current taxon sampling. Most frequently observed are insertions of short oligonucleotide duplications (Figs. 1, 2). Some indels, however, are clearly synapomorphies supporting certain clades. In the case of our sampling, adding indels resolves additional nodes, enhances Bremer support by one step in other nodes, and in a total of five nodes adds more than two steps (up to nine) to the Bremer support observed without indels. With few exceptions, the effects of SIC (simple indel coding) and MCIC (modified complex indel coding) were identical, in agreement with the observation that both indel coding methods seem to perform about the same in the absence of frequent highly complex gap patterns (Simmons et al. 2007). The most notable of the synapomorphic indels is a large deletion of 200 bp which occurs in all four Polytrichales species of our sampling. Monophyly of the Polytrichales is further supported by a 3-bp-deletion and the 19-bp-deletion shown in Fig. 2. A pentanucleotide duplication (TAAAG) is common to all Polytrichales except *Dawsonia*, in which a subsequent deletion occurred (comprising the pentanucleotide motif plus adjacent nucleotides), according to reconstruction of ancestral MCIC character states. Another significant synapomorphy is a 22 bp insertion (likewise a direct tandem sequence duplication) at alignment position 668 shared by all Hypnales taxa (Fig. 1). On smaller taxonomic scales a YTT to GGCA change at position 497 supports the sister grouping of the Neckeraceae *Homalia* and *Thamnobryum* (Fig. 1). An independent large sequence deletion of 233 bp, fully encompassing the one in Polytrichales but starting 5 bp upstream occurs as an autapomorphy in *Pterogonium*. *Schistostega* features two independent, non-overlapping deletions of 25 and 64 bp, respectively. A trinucleotide insertion (NAC) at position 363 of the alignment (accompanied by an independent 3-bp-deletion in *Sphagnum* and a 2-bp-insertion in *Tetraphis*) is shared among all derived mosses including *Diphyscium*. An extended clade, further including *Buxbaumia*, is supported by a deletion of 7 bp at alignment position 574 (located within the region of the large Polytrichales deletion).

Obvious hotspots of variability in the intergenic *nad*5-*nad*4 region are three purine-rich stretches, each of which varies in length between 6 and 12 nt after positions 380, 453 and 612, respectively and a variable T-stretch of 2–9 nt at position 624. In one of the cases, a TA_9 motif is common to the Grimmiaceae (*Coscinidon-Grimmia-Racomitrium*) clade (but not in *Blindia*) and also occurs, obviously independently, in the Pottiales (*Cinclidotus-Pottia-Tortula*) clade (not in *Timmiella*). Two striking apparent homoplasies are a 4-bp-deletion (ATAG) in position 876 shared by *Bryobrittonia* and *Cinclidotus* and a 3-bp-insertion (TGA) in *Bryobrittonia, Buxbaumia, Cinclidotus* and *Leucobryum*. A more detailed analysis of the relative homoplasy in indels will benefit from progress in resolving parts of the tree with rather low confidence such as the Bryidae-Dicranidae-Funariidae subtree. In summary, in the light of high sequence conservation even in an intergenic region of plant mtDNA, much of the phylogenetic signal stems from nucleotide exchanges rather than indels. However, some of the indels which are currently diagnosed as autapomorphies may prove informative at yet denser taxon sampling. An indel of this type which could help to resolve the Bryidae-Dicranidae-Funariidae topology (see below) is, however, obviously lacking.

Comparing indel evolution and phylogenetic signal in different organelle genome regions

To compare the phylogenetic information contained in different loci we strived for taxonomically congruent data sets for the selection of 50 moss genera for the mitochondrial *nad*5 group I intron and the *nad*2 group II intron and the uninterrupted reading frames of chloroplast genes

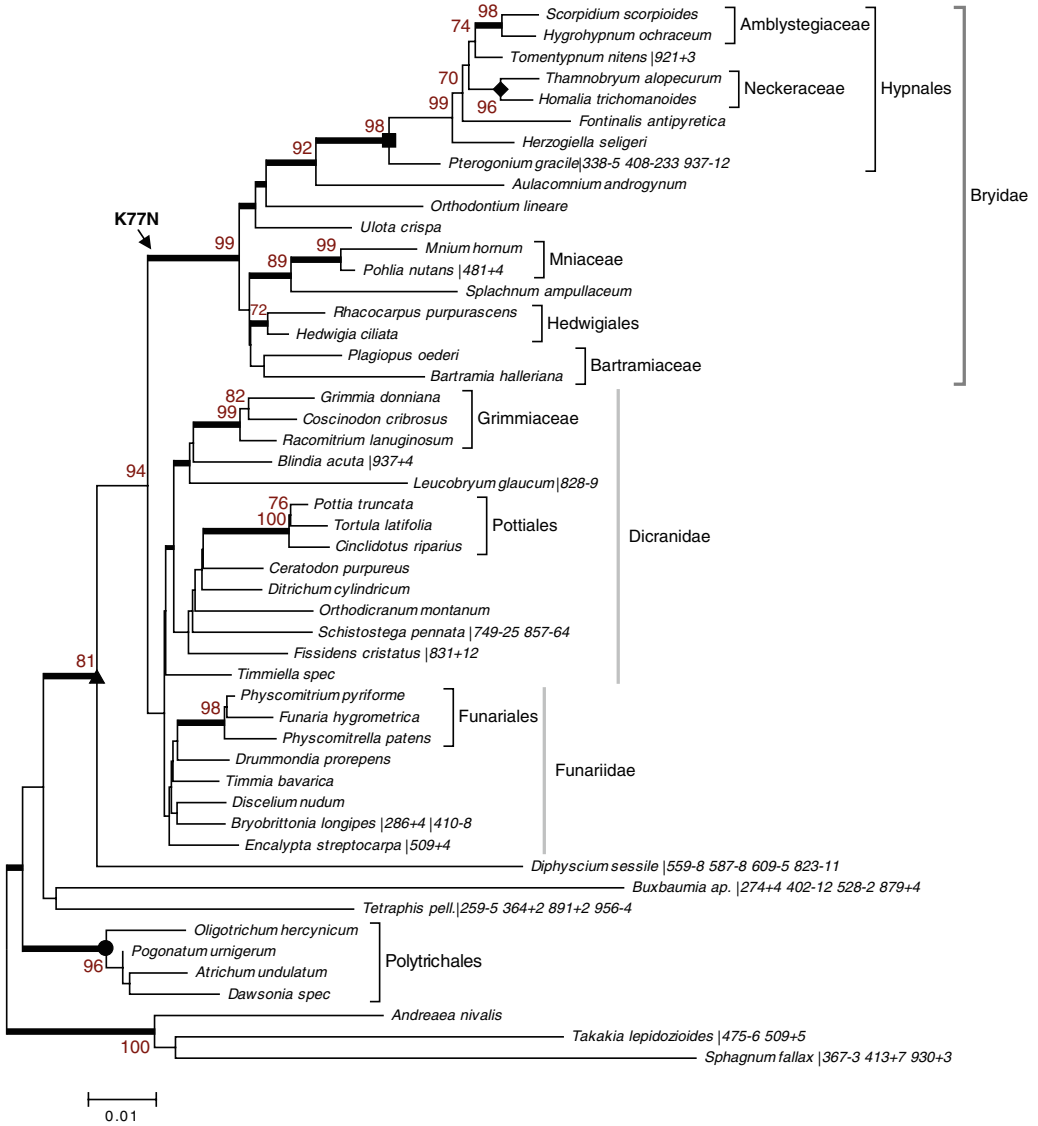


Fig. 1 Phylogenetic tree based on the *nad*5-*nad*4 sequence alignment for the 50 moss species under investigation (NJ tree based on T3P distances, pairwise gap deletion, with bootstrap support from 10,000 replicates indicated only where at least 70%). *Thickened internode lines* of increasing width indicate supports of 0.99 or 1.00, respectively, for a given clade from the Bayesian analysis run in parallel. Recognizable families, orders and higher level clades under debate are indicated with *brackets*. Subclasses Dicranidae and Funariidae (*grey lines*) emerge (except for the placement of *Drummondia)* but remain without significant support. *Symbols* indicate independent node support based on synapomorphic indels in the alignment as described in the text for Polytrichales (*filled circle*), Hypnales (*square*), Neckeraceae (*diamond*) or the arthrodontous mosses including *Diphyscium* (*triangle*), respectively. The *arrow* points to a lysine (K) to asparagine (N) amino acid change in *nad*5 as an obvious synapomorphy of the Bryidae. Autapomorphic indels larger than 1 bp are annotated with the species names after the vertical line with alignment position, ± for insertions and deletions, and the indel size, respectively

*rbc*L and *rps*4 which were already widely sampled and represented in the database (Table 1). To this end we filled all the remaining taxonomic gaps (with the sole exception of *nad*2 in *Drummondia*, which we could not retrieve).

The availability of data sets for all three different types of non-coding mtDNA—an intergenic region (*nad*5-*nad*4), a group I intron (nad5i753) and a group II intron (nad2i156)—allows for the first time a comparison of their

modes of evolution (Table 2). One might expect a higher frequency of indels in the intergenic spacer as there should be less evolutionary constraint than in group I or II introns with their pronounced secondary and tertiary structures, but this does not seem to be the case.

Autapomorphic single or oligonucleotide insertions seem to occur in the nad5i753 group I intron and the nad2i156 group II intron in very similar frequencies compared with the *nad*5-*nad*4 spacer region without any obvious increase in frequency in a particular taxon (Table 2). Similarly, stretches of polypurine sequences of variable length are likewise observed in both types of introns. As in the *nad*5-*nad*4 spacer region much of the phylogenetic information in the introns resides in nucleotide exchanges rather than in shared indels under current taxon sampling density. As in the spacer region, the indel evolution is not entirely free of homoplasies in the introns either: A 4-bp-deletion in the *nad*2 group II intron is shared by the distant taxa *Ceratodon* and *Plagiopus*. An 8-bp-deletion and an independent 6-bp-deletion in the *nad*2 group II intron on the other hand are obvious synapomorphies of the Pottiaceae and the Polytrichales, respectively, in the data set. Other, larger sequence deletions are not observed in this group II intron. The *nad*5 group I intron data set in contrast shows larger deletions of 4, 5 and 136 bp, respectively, as obvious synapomorphies of the Funariaceae. The larger of these deletions has a 5′ end which is identical to an independent, yet much larger, deletion of 304 bp in *Tetraphis* and fully encompasses a somewhat smaller deletion of 55 bp in *Hedwigia* (Table 2). These and another large non-overlapping deletion in *Schistostega* of 80 bp are located in the highly variable loop L8 of group I intron nad5i753 (Beckert et al. 1999). This loop is clearly the most variable region of sequence variation in the non-coding mtDNAs compared here for the moss taxon sampling. Some cases suggest that modes of indel evolution are more taxon-dependent than locus-dependent (Table 2). Most striking are the deletions of larger sequence stretches in both the intergenic region and the group I intron of *Schistostega* and the multiple independent smaller sequence deletions in *Buxbaumia* and *Diphyscium*, although the latter may simply be a result of the long isolated branches that these taxa occupy (see below and Fig. 3).

Visual examination of alignments implied that indels were frequent and/or long in some taxa, regardless of the marker, while other taxa had only few and/or shorter indels across all markers. To test for such lineage effects in indel evolution across the three mt regions compared, correlation analysis based on approximations of relative indel rates on branches were performed. Based on Spearman's Rank Correlation Coefficient, indel evolution (in terms of number of insertion/deletion events per branch) is significantly correlated ($p < 0.001$) among markers (correlation coefficients of 0.638 group II intron vs. spacer, 0.513 group II intron vs. group I intron and 0.388 group I intron vs. spacer, respectively), thereby supporting lineage effects that dominate locus-specific length mutational patterns.

Protein coding sequences and RNA editing in *nad*5/*nad*4

The *nad*5 and *nad*4 coding regions flanking the spacer show the usual requirements of C-to-U RNA editing to reconstitute conserved codons in plant organelle mRNAs. Like the indels in the intergenic region, many of the editing sites are also autapomorphies in the present sampling. RNA editing in mosses is comparatively rare but *Takakia* confirms elevated editing requirement in this genus: 5 out of 75 *nad*5 codons (including generation of the *nad*5 stop codon) and 6 out of 84 *nad*4 codons (including generation of the *nad*4 start codon). Apparent synapomorphies are a requirement in *nad*5 for a proline-to-leucine editing (P50L, codon numbering for *nad*5 beginning with the first codon in the amplicon) in Polytrichales (except *Oligotrichum*) and independently in six Hypnales taxa, but not in *Fontinalis* and *Herzogiella*, corroborating their phylogenetic basal position in the order as observed in the fused data set phylogeny (Fig. 3). A notable change in an amino acid position unrelated to RNA editing is K77N in *nad*5 which appears as a synapomorphy of the well supported Bryidae clade (Fig. 1). In *nad*4 a H9Y exchange is a synapomorphy of Hedwigiales. Some RNA editing requirements are apparent homoplasies in the data set: an arginine-to-tryptophane change (R57W) in *Takakia* and *Orthodontium* and a proline-to-leucine (P71L) change in *Aulacomnium* and *Ditrichum* in *nad*5 and P17L and R56W RNA editings in *nad*4 occurring in several unrelated species.

Phylogenetic information in five different loci

Phylogenetic trees were calculated for all data sets separately and for the fused data matrix which contained 7,549 characters (1,222 for the *nad*5-*nad*4-spacer, 2,126 for *nad*5, 2,356 for *nad*2, 1,293 for *rbc*L and 552 for *rps*4). The phylogenetic tree based on the complete data set is shown in Fig. 3. For clarity we have summarized node support from the different contributing data sets in Table 3.

The earliest dichotomy in the phylogeny of extant mosses is as yet unclear but likely separates *Takakia*, *Sphagnum* or both genera from the remainder of mosses and we have used the joint clade to root the trees. In any case, there is support for setting *Sphagnum* and *Takakia* apart from the other taxa in the unrooted trees and *Andreaea* was placed sister to all remaining taxa by all loci except *rbc*L (Table 3). The monophyly of Polytrichales is

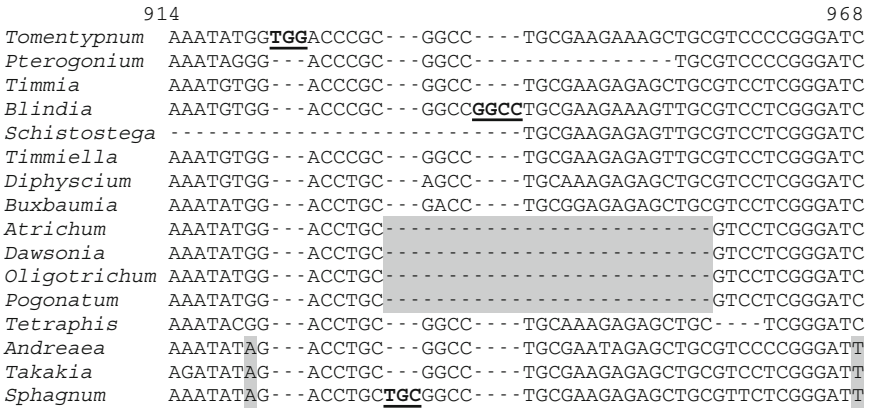


Fig. 2 A small section of the *nad*5-*nad*4 spacer alignment to illustrate indel evolution (full alignment positions 914 to 968, located 24 bp upstream of the *nad*4 start codon). Autapomorphic insertions in this selection in *Tomentypnum, Sphagnum* and *Blindia* are obviously derived from trinucleotide or tetranucletide sequence repeats, respectively (*bold*, *underlined*). Autapomorphic deletions of 12, 67 or 4 bp have occurred in *Pterogonium, Schistostega* (not fully shown) and *Tetraphis*, respectively. A 19 bp deletion is a synapomorphy in the Polytrichales and two base transitions in this region separate *Andreaea, Takakia* and *Sphagnum* from the other mosses (*grey shadings*)

confirmed by all loci (Table 3) but the branching order of the nematodontous mosses, Polytrichales and *Tetraphis*, on the backbone of moss phylogeny or a possible sister group relationship of Polytrichales and Tetraphidales so far remains without significant support. Clear support for the placement of mosses with pleated endostomes, *Buxbaumia* and *Diphyscium*, as successive sister clades basal to the arthrodontous mosses (Fig. 3) is mostly derived from phylogenetic signal in the mitochondrial *nad*5 and *nad*2 data sets (Table 3), as had been observed earlier at lower taxon sampling (Beckert et al. 2001). An alternative placement of *Buxbaumia* as sister taxon to *Tetraphis* is suggested by *rps*4 with the given taxon sampling (Table 3) but as in the case of the *nad*5-*nad*4 spacer (Fig. 1) remains unsupported. Among the derived, arthrodontous mosses three clades have been designated as subclasses Bryidae, Dicranidae and Funariidae (Goffinet and Buck 2004), respectively, generally with decreasing support from molecular data in this order (Fig. 3).

The Bryidae (the clade of mosses with diplolepidous-alternate peristomes supposed to contain orders Splachnales, Orthotrichales, Hedwigiales, Bryales, Rhizogoniales, Ptychomniales, Hookeriales and Hypnales) are unequivocally supported by molecular data (Beckert et al. 1999, 2001; Cox et al. 2000, 2004; Goffinet and Buck 2004) including the novel *nad*5-*nad*4 spacer region (Fig. 1) except for the distance based analysis of *rbc*L at our taxon sampling (Table 3). Among the Bryidae the Hypnales but not the Bryales find support as a monophyletic clade from the molecular data. In one clade of the Bryidae, the Hypnales are successively joined by *Aulacomnium, Orthodontium* and *Ulota* (Figs. 1, 3; Table 3) and this is in full accord with recent studies (Quandt et al. 2007; Bell et al. 2007) who have raised these families to the rank of orders (Aulacomniales, Orthodontiales and Orthotrichales) to secure a monophyly of the Bryales. A taxonomically wide concept of Bryales is rendered paraphyletic in any case, in full accord with previous molecular analyses (Cox et al. 2000). The placement of two other families of the Bryales, the Bartramiaceae and Mniaceae, still remains unsettled, although sister relationships of Bartramiaceae with Hedwigiales and Mniaceae with Splachnum, respectively, may be possible (Fig. 3; Table 3). Recent findings place Mniaceae within the Bryales and the Bartramiaceae in their own order (Quandt et al. 2007; Bell et al. 2007), but the position of the Bartramiales is yet unclear and a sister relationship with Hedwigiales still possible. With regard to Splachnales, both studies did not resolve their placement unambiguously, however a sister relationships of Mniaceae and *Splachnum* appears very unlikely.

Strong support for the monophyly of the Dicranidae, the clade of mosses with haplolepidous peristomes, comes mostly from *rbc*L in our taxon sampling (Table 3). Previous molecular studies have investigated the phylogeny of the Dicranidae with wide taxonomic sampling using chloroplast loci in single gene *rps*4 (Hedderson et al. 2004) or multigene (*rbc*L, *rps*4, *trn*L-F) studies (La Farge et al. 2000, 2002). Overall, the recent modern molecular plant phylogenies based on reasonable markers and sufficient taxon sampling seem to converge on the same tree topologies and the remaining open points indeed appear to reflect short internode distances indicating quick succession of lineage branchings for which informative synapomorphies have yet to be identified. *Timmiella*,

Fifty mosses on five trees

Table 2 Indels larger than 1 bp which are clearly identified as 2, 3, 4, 5 bp or larger tandem sequence duplications and of sequence deletions, respectively, in the three non-coding mtDNA regions are listed for the respective taxa in alphabetic order

Loci	nad5-nad4 intergenic region						nad2i156 group II intron						nad5i753 group I intron					
	Number of tandem duplications					Deletions	Number of tandem duplications					Deletions	Number of tandem duplications					Deletions
Taxa	2 bp	3 bp	4 bp	5 bp	Larger	No. (sizes)	2 bp	3 bp	4 bp	5 bp	Larger	No. (sizes)	2 bp	3 bp	4 bp	5 bp	Larger	No. (sizes)
Andreaea						1 (17)						2 (9, 11)						
Atrichum																1		
Bartramia											1 (9)							
Blindia			1															
Buxbaumia			2			2 (2, 12)						4 (3, 9 12, 20)	1					3 (2, 4, 21)
Bryobrittonia			1			2 (4, 8)												
Ceratodon												2 (4, 5)						
Cinclidotus						1 (4)												
Diphyscium						4 (5, 8, 8, 11)	1					1 (10)			1			5 (4, 4, 5, 10, 17)
Encalypta			1															
Fissidens					1 (12)					1		1 (2)						1 (4)
Funariaceae																		3 (4, 5, 136)
Hedwigia																		1 (55)
Homalia							1											
Hypnales					1 (22)													
Leucobryum					1 (9)					2	1 (6)							
Orthodicranum									1			1 (4)						
Plagiopus												1 (4)						
Pohlia			1															
Polytrichales				1		3 (3, 19, 200)						1 (6)						
Pottia														1				
Pottiaceae												1 (8)			1			
Pterogonium						3 (5, 12, 233)												
Racomitrium												1 (6)						1 (6)
Schistostega						2 (25, 64)										1		3 (4, 11, 80)
Sphagnum		1			1 (7)											1		
Takakia				1					1									2 (3, 6)
Tetraphis	2					2 (4, 5)		1			1 (6)		1	1				1 (304)
Tomentypnum		1																
Ulota															1			1 (7)

Regions of ambiguous alignment and polypurine-stretches of variable length were not considered

conventionally included among Pottiaceae (Goffinet and Buck 2004) clearly has to be placed in a basal grade of "proto-haplolepidous" taxa (Hedderson et al. 2004) branching before *Drummondia* according to the molecular data (Fig. 3) whereas (other) Pottiacae/-ales, including *Cinclidotus*, find perfect support as a clade (Figs. 1, 3). Likewise, the monophyly of the Grimmiaceae, a sister group of *Grimmia* and *Coscinodon*, and of Grimmiales including *Blindia* (Seligeriaceae), find significant support in the molecular phylogenies (Fig. 3; Table 3). On the other hand neither the taxonomically wide concepts of Dicranales nor the Ditrichaceae are supported. The grade of *Ceratodon* and *Ditrichum* (Ditrichaceae) followed by *Schistostega* as successive sister groups to Pottiales and ultimately joined by a *Fissidens-Orthodicranum* sister group as well as the placement of *Leucobryum* (Fig. 3) are

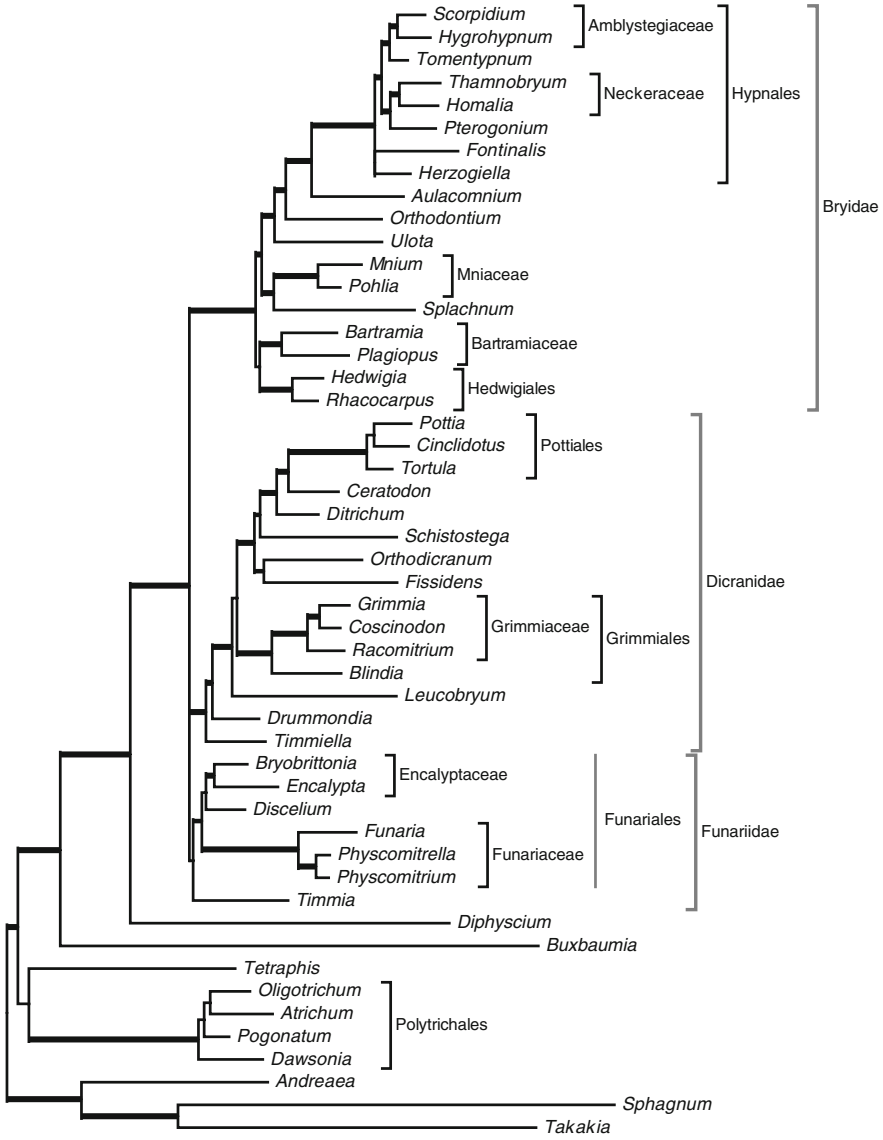


Fig. 3 Phylogenetic tree for the fused *nad*5-*nad*4 + *nad*5 + *nad*2 + *rps*4 + *rbc*L data set. The phylogeny shown is the Bayesian consensus tree topology. *Thickened internode lines* of increasing width indicate Bayesian support of 0.99 or 1.00, respectively

entirely compatible with an *rps*4 single gene study at higher taxon sampling (Hedderson et al. 2004), where the respective nodes lacked support, however.

The Funariidae (the clade of mosses with diplolepidous-opposite peristomes, now supposed to include *Timmia*, *Discelium*, the Encalyptales and the Funariales) generally find the weakest support from molecular data as one of the three subclasses of arthrodontous mosses. In the tree based on the fused data set (Fig. 3) the Funariidae appear as a clade but without significant support. Whether the Funariidae are the sister group to the Bryidae, to the Dicranidae or to the two joint subclasses remains unsettled. One key taxon in this problematic region is *Timmia*, given the now obvious exclusion of Timmiaceae from the Bryales according to molecular data, which make a reconsideration of morphological and reinvestigation of developmental (largely peristome) characters necessary (Budke and Goffinet 2006; Budke et al. 2007). The Funariaceae are well supported, and among these, a sister group of *Physcomitrium* and *Physcomitrella* (Fig. 3; Table 3). The sister relationship of *Bryobrittonia* and *Encalypta* (Encalyptales) emerges with *nad*5 and *rps*4 (Table 3). A very important observation concerning the monophyly of a subclade in the Funariidae cluster is a 71 kb inversion of the chloroplast genome shared by Encalyptales, Funariaceae and *Discelium* to the exclusion of

Fifty mosses on five trees

Table 3 Clades indicated in the left column received node confidences from the five individual data sets and the fused data in Neighbour-Joining Bootstrap Analysis (10,000 replicates, numbers as percentages before the slash) and Bayesian posterior node probabilities (decimal numbers, after the slash) as indicated

Clades	*nad*5-*nad*4	*nad*5	*nad*2	*rbc*L	*rps*4	Fused
(*Sphagnum, Takakia*)	60/0.92	99/1.00	66/1.00	99/1.00	96/1.00	100/1.00
(*Sphagnum, Takakia, Andreaea*)	100/1.00	90/1.00	100/1.00	0/0.00	87/1.00	100/1.00
(*Atrichum, Dawsonia, Oligotrichum, Pogonatum*), i.e. Polytrichales	96/1.00	100/1.00	100/1.00	100/1.00	98/1.00	100/1.00
(*Sphagnum, Takakia, Andreaea,* Polytrichales, *Tetraphis*)	0/0.00	86/0.99	100/1.00	0/0.00	0/0.00	97/1.00
(*Sphagnum, Takakia, Andreaea,* Polytrichales, *Tetraphis, Buxbaumia*)	81/1.00	93/1.00	84/1.00	0/0.70	0/1.00	95/1.00
(*Sphagnum, Takakia, Andreaea,* Polytrichales, *Tetraphis, Buxbaumia, Diphyscium*)	94/1.00	99/1.00	99/0.81	100/1.00	90/1.00	100/1.00
Funariidae (i.e. "DEFT" = *Discelium*-Encalyptales-Funariales-*Timmia* clade)	0/0.00	0/0.00	0/0.00	0/0.00	0/0.00	83/0.65
(*Physcomitrella, Physcomitrium*)	0/0.00	84/0.00	96/0.99	100/1.00	99/0.93	100/1.00
(*Physcomitrella, Physcomitrium, Funaria*), i.e. Funariales	98/1.00	94/1.00	100/1.00	100/1.00	99/1.00	100/1.00
(*Bryobrittonia, Encalypta*), i.e. Encalyptales	0/0.00	0/0.98	0/0.00	0/0.00	75/0.99	0/1.00
(Encalyptales, *Discelium*)	0/0.00	0/0.00	0/0.00	94/0.89	81/0.00	99/0.78
(Encalyptales, *Discelium*, Funariales)	0/0.00	0/0.00	0/0.00	0/0.00	68/1.00	69/0.99
Dicranidae	0/0.60	49/1.00	0/0.00	83/1.00	0/0.90	93/1.00
(*Pottia, Tortula*), i.e. Pottiaceae	76/0.00	78/0.92	0/0.00	0/0.00	83/1.00	65/0.00
(*Pottia, Tortula, Cinclidotus*), i.e. Pottiales excluding *Timmiella*	100/1.00	100/1.00	100/1.00	100/1.00	99/1.00	100/1.00
(Pottiales, *Ceratodon*)	37/0.94	47/0.56	0/0.00	47/1.00	0/0.00	83/1.00
(Pottiales, *Ceratodon, Ditrichum*)	54/0.96	79/1.00	99/1.00	0/0.00	0/0.00	99/1.00
(Pottiales, *Ceratodon, Ditrichum, Schistostega*)	56/0.00	67/0.51	0/0.00	0/0.62	0/0.00	89/0.99
(*Fissidens, Orthodicranum*)	0/0.00	90/0.99	0/0.00	0/0.00	0/0.00	87/0.99
(*Fissidens, Orthodicranum,* Ditrichaceae, Pottiales, *Schistostega*)	50/0.97	80/1.00	88/1.00	0/0.00	0/0.00	93/1.00
(*Grimmia, Coscinodon*)	82/0.71	100/1.00	51/0.00	90/0.97	89/0.98	100/1.00
(*Grimmia, Coscinodon, Racomitrium*), i.e. Grimmiaceae	99/1.00	91/1.00	100/1.00	96/1.00	99/1.00	100/1.00
(Grimmiaceae, *Blindia*), i.e. Grimmiales	44/0.00	59/0.00	100/1.00	81/1.00	51/0.94	100/1.00
(*Fissidens, Orthodicranum,* Ditrichaceae, Pottiales, *Schistostega*, Grimmiales, *Leucobryum*)	46/0.98	24/1.00	0/0.00	32/1.00	43/1.00	0/1.00
(*Fissidens, Orthodicranum*, Ditrichaceae, Pottiales, *Schistostega*, Grimmiales, *Leucobryum, Drummondia*)	0/0.00	28/0.97	n.a.	0/0.00	0/0.82	100/1.00
Bryidae	96/1.00	99/1.00	99/1.00	0/1.00	90/1.00	100/1.00
(*Homalia, Thamnobryum*), i.e. Neckeraceae	96/0.84	0/0.00	56/0.91	0/0.00	89/1.00	98/1.00
(Neckeraceae, *Pterogonium*)	0/0.00	0/0.82	83/0.00	0/0.00	0/0.00	0/1.00
(*Hygrohypnum, Scorpidium*), i.e. Amblystegiaceae	100/1.00	0/0.69	70/1.00	0/0.00	0/0.0	95/1.00
(Amblystegiaceae, *Tomentypnum*)	87/0.85	0/0.95	87/1.00	0/0.00	0/0.0	100/1.00
(Neckeraceae, *Pterogonium*, Amblystegiaceae, *Tomentypnum*)	0/0.00	0/0.84	0/0.00	0/0.00	41/0.71	0/1.00
Hypnales	98/1.00	98/1.00	100/1.00	87/1.00	92/1.00	100/1.00
(Hypnales, *Aulacomnium*)	92/1.00	85/1.00	55/0.98	0/0.00	0/0.00	95/1.00
(Hypnales, *Aulacomnium, Orthodontium*)	58/1.00	60/0.80	0/0.00	0/0.00	0/0.00	95/1.00
(Hypnales, *Aulacomnium, Orthodontium, Ulota*)	64/1.00	86/1.00	79/0.99	0/0.00	0/0.00	95/1.00
(*Mnium, Pohlia*), i.e. Miniaceae	99/1.00	100/1.00	96/1.00	99/1.00	99/1.00	100/1.00
(Mniaceae, *Splachnum*)	89/1.00	70/0.87	0/0.00	0/0.00	0/0.00	58/1.00
(Hypnales, *Aulacomnium, Orthodontium, Ulota,* Mniaceae, *Splachnum*)	0/1.00	0/0.00	0/0.00	0/0.00	0/0.00	0/1.00
(*Hedwigia, Rhacocarpus*), i.e. Hedwigiales	72/1.00	99/1.00	99/1.00	100/1.00	90/0.99	100/1.00
(*Bartramia, Plagiopus*), i.e. Bartramiaceae	66/0.84	82/1.00	94/1.00	99/1.00	89/0.98	100/1.00
(Bartramiaceae, Hedwigiales)	42/0.00	45/0.00	62/0.00	0/0.00	34/0.00	91/0.73

Values of zero indicate absence of a given clade in the respective analysis

Timmia (Goffinet et al. 2007), excellently supporting the basal placement of *Timmia* if included in the Funariidae as observed here (Fig. 3).

Conclusions and outlook

Bryophyte molecular phylogenetics has reached a state of maturity where different data sets at sufficient taxon sampling lead to very similar (at least not to contradictory, where node supports are significant) conclusions. Both the postulate of sufficient taxon sampling as well as the need for exploring the most informative of loci are now generally appreciated and extending the one while neglecting the other is a useless endeavour. Even branch lengths in moss phylograms based on different molecular loci seem to tell similar stories. Although some genera at highly crucial positions in the moss phylogeny—maybe most notably e.g. the taxa with pleated endostomes at the base of arthrodontous mosses, *Buxbaumia* and *Diphyscium*—reside on long terminal branches, which very likely also in the future cannot be broken by further sampling of extant species, their phylogenetic positions on the backbone of moss phylogeny seem to be resolved with reasonable confidence. In other cases, however, internodes only slightly deeper in the moss phylogeny such as those separating the nematodontous orders Polytrichales and Tetraphidales are congruently so short in molecular phylogenies that a reliably resolved phylogeny is lacking. We could not find convincing support for the assumed monophyly of Buxbaumiaceae with nematodontous mosses (Goffinet et al. 2001) or a sister group relationship of Tetraphidales with the arthrodontous mosses (Magombo 2003). Among the arthrodontous mosses several genera (*Bryobartramia, Drummondia, Timmia, Timmiella*, etc.) have received new systematic assignments (Goffinet and Buck 2004), several other concepts will need further revision in the future (e.g. the Dicranales). Some of the problematic taxa are key to the most notable remaining phylogenetic problem at the base of arthrodonts: better convincing support for the Funariids and Dicranids including resolution of their basal lineages and the resolution of the two clades, when fully confirmed, relative to the Bryidae.

Taxon sampling may certainly be crucial for several regions of unresolved branchings in the moss phylogeny. The addition of *Tetrodontium* as the second genus in the Tetraphidales and of the isolated genus *Oedipodium*, presumably branching close to these, is a case in point; the inclusion of taxa from the unique family Gigaspermaceae another. On the other hand, the very densely taxon-sampled *rps*4 gene alone including these taxa can likewise not resolve the moss backbone topology in single gene analyses here (not shown), showing that increase in taxon sampling and exploring informative novel molecular loci must go hand in hand.

The phylogenetic value of indel evolution has received increased attention also for issues of deep phylogeny, for example in the early angiosperm diversifications (e.g. Löhne and Borsch 2005; Worberg et al. 2007). Here we observed that indel evolution in the mitochondrial DNA of mosses is clearly more a matter of lineage than of locus functionality (group I vs. group II intron or an intergenic region). Synapomorphic indels in the three types of non-coding mitochondrial regions (as well as in the *rps*4 coding region), however, mainly lend further support only to those clades that had been determined with confidence before. The loss of a codon in *rps*4 of the Funariaceae is a typical example (Goffinet and Cox 2000). In other words, indel accumulation has been found to be largely proportional to nucleotide exchange when considering sufficiently long branches.

Where molecular phylogenies may be hampered by the effects of long-branch attraction between isolated and/or ancient linages, other molecular changes on the genomic level are a welcome alternative. The loss of the *rpo*A gene from the chloroplast genome of mosses (and its presumed foregone functional transfer to the nucleus) is an example (Sugiura et al. 2003; Sugita et al. 2004). However, the *rpo*A gene turned out to be lost twice independently: in the lineage of arthrodontous mosses including *Diphyscium* but excluding *Buxbaumia* and in the Tetraphidales (Goffinet et al. 2005) and hence is again not phylogenetically informative with respect to the placement of the two nematodontous orders. In contrast, the 71 kbp chloroplast inversion mentioned above (Goffinet et al. 2007) is a perfect example of welcome external support for a clade in the Funariidae to the exclusion of *Timmia*.

Acknowledgments The authors are very grateful to Drs. Jan-Peter Frahm, Bonn; Bernard Goffinet, Storrs CT; Hermann Muhle, Ulm; Yin-Long Qiu, Ann Arbor MI and Dietmar Quandt, Dresden, who made biological material or DNA samples, respectively, available. Financial support through the Deutsche Forschungsgemeinschaft (DFG) is gratefully acknowledged. The last author is grateful to the Deutsche Bahn for continuous, reliable and comfortable ICE transportation commuting between Bonn and Ulm on a weekly basis also in preparation of the present manuscript.

References

Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad*5 gene. Pl Syst Evol 218:179–192

Beckert S, Muhle H, Pruchner D, Knoop V (2001) The mitochondrial *nad*2 gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Molec Phylogenet Evol 18:117–126

Bell NE, Quandt D, O'Brien TJ, Newton AE (2007) Taxonomy and phylogeny in the earliest diverging pleurocarps: square holes and bifurcating pegs. Bryologist 110:533–560

Budke JM, Goffinet B (2006) Phylogenetic analyses of Timmiaceae (Bryophyta: Musci) based on nuclear and chloroplast sequence data. Syst Bot 31:633–641

Budke JM, Jones CS, Goffinet B (2007) Development of the enigmatic peristome of Timmia megapolitana (Timmiaceae; Bryophyta). Amer J Bot 94:460–467

Cox CJ, Goffinet B, Newton AE, Shaw AJ, Hedderson TAJ (2000) Phylogenetic relationships among the diplolepideous-alternate mosses (Bryidae) inferred from nuclear and chloroplast DNA sequences. Bryologist 103:224–241

Cox CJ, Goffinet B, Shaw AJ, Boles SB (2004) Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. Syst Bot 29:234–250

Crandall-Stotler B, Forrest LL, Stotler RE (2005) Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). Taxon 54:299–316

Davis EC (2004) A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). In: Goffinet B, Hollowell V, Magill R (eds) Molecular systematics of bryophytes. Missouri Botanical Garden Press, St Louis, pp 61–86

Forrest LL, Crandall-Stotler BJ (2005) Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. J Hattori Bot Lab 97:127–159

Goffinet B, Buck WR (2004) Systematics of the bryophyta (mosses): from molecules to a revised classification. In: Goffinet B, Hollowell V, Magill R (eds) Molecular systematics of bryophytes. Missouri Botanical Garden Press, St Louis, pp 205–239

Goffinet B, Cox CJ (2000) Phylogenetic relationships among basalmost arthrodontous mosses with special emphasis on the evolutionary significance of the Funariineae. Bryologist 103:212–223

Goffinet B, Cox CJ, Shaw AJ, Hedderson TAJ (2001) The bryophyta (mosses): systematic and evolutionary inferences from an rps4 gene (cpDNA) phylogeny. Ann Bot 87:191–208

Goffinet B, Wickett NJ, Shaw AJ, Cox CJ (2005) Phylogenetic significance of the rpoA loss in the chloroplast genome of mosses. Taxon 54:353–360

Goffinet B, Wickett NJ, Werner O, Ros RM, Shaw AJ, Cox CJ (2007) Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). Ann Bot (Lond) 99:747–753

Groth-Malonek M, Rein T, Wilson R, Groth H, Heinrichs J, Knoop V (2007) Different fates of two mitochondrial gene spacers in early land plant evolution. Int J Pl Sci 168:709–717

Hedderson TA, Murray DJ, Cox CJ, Nowell TL (2004) Phylogenetic relationships of haplolepideous mosses (Dicranidae) inferred from *rps*4 gene sequences. Syst Bot 29:29–41

Hyvönen J, Koskinen S, Smith Merrill GL, Hedderson TA, Stenroos S (2004) Phylogeny of the Polytrichales (Bryophyta) based on simultaneous analysis of molecular and morphological data. Molec Phylogenet Evol 31:915–928

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

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46:123–139

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinf 5:150–163

La Farge C, Mishler BD, Wheeler JA, Wall DP, Johannes K, Schaffer S, Shaw AJ (2000) Phylogenetic relationships within the haplolepideous mosses. Bryologist 103:257–276

La Farge C, Shaw AJ, Vitt DH (2002) The circumscription of the Dicranaceae (Bryopsida) based on the chloroplast regions *trn*L-*trn*F and *rps*4. Syst Bot 27:435–452

Löhne C, Borsch T (2005) Molecular evolution and phylogenetic utility of the petD group II intron: a case study in basal angiosperms. Molec Biol Evol 22:317–332

Maddison WP, Maddison DR (2003) Mesquite: a modular system for evolutionary analysis. [1.0]

Magombo ZLK (2003) The phylogeny of basal peristomate mosses: evidence from cpDNA, and implications for peristome evolution. Syst Bot 28:24–38

Müller K (2004) PRAP-computation of Bremer support for large data sets. Molec Phylogenet Evol 31:780–782

Müller K (2005) SeqState: primer design and sequence statistics for phylogenetic DNA datasets. Appl Bioinformatics 4:65–69

Müller K (2006) Incorporating information from length-mutational events into phylogenetic analysis. Molec Phylogenet Evol 38:667–676

Quandt D, Bell N, Stech M (2007) Unravelling the knot: the Pulchrinodaceae fam. nov (Bryales). Nova Hedwigia 21–39

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinf 19:1572–1574

Simmons MP, Ochoterena H (2000) Gaps as characters in sequence-based phylogenetic analyses. Syst Biol 49:369–381

Simmons MP, Müller K, Norton AP (2007) The relative performance of indel-coding methods in simulations. Molec Phylogenet Evol 44:724–740

Sugita M, Sugiura C, Arikawa T, Higuchi M (2004) Molecular evidence of an *rpo*A gene in the basal moss chloroplast genomes: *rpo*A is a useful molecular marker for phylogenetic analysis of mosses. Hikobia 14:171–175

Sugiura C, Kobayashi Y, Aoki S, Sugita C, Sugita M (2003) Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of rpoA from the chloroplast to the nucleus. Nucleic Acids Res 31:5324–5331

Swofford DL (2003) PAUP* Phylogenetic analysis using parsimony (* and Other Methods). Version 4. Sinauer Assoc., Sunderland, MA. [4]. Sunderland, MA, Sinauer Associates

Worberg A, Quandt D, Barniske A-M, Löhne C, Hilu KW, Borsch T (2007) Phylogeny of basal eudicots: insights from non-coding and rapidly evolving DNA. Organisms Diversity & Evolution 7:55–77

3.1.2 Ute Wahrmund, Dietmar Quandt and Volker Knoop (2010): The phylogeny of mosses – addressing open issues with a new mitochondrial locus: group I intron cobi420. *Molecular Phylogenetics and Evolution* 54 (2): 417-426

After nad5i753, a second group I intron in the mitochondrial genome of mosses was investigated and evaluated phylogenetically. Group I intron cobi420 was found to be conserved in all mosses analysed and a size reduction during moss evolution was observed. Moreover, three taxa of the Gigaspermaceae and *Catoscopium* were now included in the taxon sampling. The phylogenetic tree based on cobi420 alone, exhibited good resolution and supported the monophyly of all major lineages except the Funariidae. A well supported and resolved phylogram was the result of analysing the fused data set (cobi420, *nad*5, *nad*2, *nad*5-*nad*4 spacer, *rbc*L and *rps*4). To rule out artefacts of long-branch attraction, we successively excluded taxa with long branches, referred to as a 'taxon quenching' approach. The topology, however, was never affected except for the respective excluded taxon. Again, a monophyly of nematodontous mosses (*Tetraphis* and Polytrichopsida) was shown, although weakly supported (see Fig. 3 in the publication). The core Funariidae formed a monophyletic clade to the exclusion of the Gigaspermaceae and *Timmia*. *Timmiella* and *Catoscopium* were clearly placed as the basal-most genera of the Dicranidae, while *Timmia*, the Gigaspermaceae and the Bryidae grouped together unsupported. Except for the placement of *Buxbaumia* and *Diphyscium*, the backbone within the Bryopsida consisted of short nodes, not resolving the relationships in the major lineages with support. Thus, a rapid radiation within a short time appeared to have given rise to the ancestors of the three major subclasses, the genus *Timmia* and the Gigaspermaceae.

Contribution: I obtained and sequenced the cobi420 amplificates, built the alignment, calculated and interpreted the phylogenetic trees and I wrote large parts of the manuscript.

Molecular Phylogenetics and Evolution 54 (2010) 417–426

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The phylogeny of mosses – Addressing open issues with a new mitochondrial locus: Group I intron cobi420

Ute Wahrmund [a,*], Dietmar Quandt [b], Volker Knoop [a]

[a] IZMB – Institut für Zelluläre und Molekulare Botanik, Abt. Molekulare Evolution, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany
[b] NEES – Institut für Biodiversität der Pflanzen, Abt. Phylogenie und Evolution der Pflanzen, Universität Bonn, Meckenheimer Allee 170, 53115 Bonn, Germany

ARTICLE INFO

Article history:
Received 25 April 2009
Revised 12 September 2009
Accepted 22 September 2009
Available online 21 October 2009

Keywords:
Inversions
Mitochondrial DNA
Group I intron
Bryophytes

ABSTRACT

Mosses are an ancient clade of land plants, set apart from the two other monophyletic groups of bryophytes, the liverworts and the hornworts. Different molecular data sets seem to converge towards a consensus backbone phylogeny of mosses. Nevertheless, for some crucial nodes open questions remain, which obviously require additional phylogenetic information. We here report that a group I intron in the mitochondrial *cob* gene (cobi420) is universally conserved in the mosses, including basal genera such as *Sphagnum* and *Takakia*. Well resolved phylogenetic trees were obtained for 56 mosses of wide phylogenetic sampling, e.g. supporting a placement of *Drummondia* branching after the genera *Timmiella* and *Catoscopium* basal in the Dicranidae. Taxon gaps have been filled in the previously established data sets nad5i753, nad2i156, the *nad*5-*nad*4 intergenic region and chloroplast *rbc*L and *rps*4 genes. A concatenated 6-loci analysis suggests reconsideration of a Funariidae concept which includes Gigaspermaceae but leaves the isolated genus *Timmia* at the radiation of arthrodontous mosses unresolved.

1. Introduction

As for many other groups of life, molecular data have had a major impact on our understanding of the evolution and phylogeny of mosses and consequently on modern systematics and taxonomy of this clade (Goffinet et al., 2008; Stech and Frey, 2008). Mosses are a particularly interesting example given the general scarcity of reliable morphological characters, further complicated by frequent independent reductions of morphological complexity (e.g. Huttunen et al., 2004; Olsson et al., 2009). Beside features of the enduring gametophyte state of bryophytes, the classical systematics of mosses has for the last 200 years largely relied on the position of the sporophyte and its peristome structures, especially the rings of teeth lining the opening of the mature spore capsule. The arthrodontous mosses feature elaborate peristomes that are composed of periclinal cell wall plates rather than whole cells. Two main categories are recognized for the architecture of the arthrodontous peristomes: taxa with haplolepideous (dicranid) and diplolepideous peristomes. In the latter, the outer surface of each tooth is typically composed of two cells, rather than one cell as in the former. However, peristome teeth are not homologous between the two types: The diplolepideous peristome typically comprises two concentric rings whereas only one ring is found in the haplolepideous peristome. The teeth of the concentric rings of the diplolepidous peristome can either face (funariid) or alternate (bryid) with one another. Varying degrees of secondary reductions complicate morphological assignments in many cases.

The prime importance of molecular data for elucidating a phylogeny of mosses has clearly been stated more than 10 years ago (Goffinet et al., 1998). By and large, the affiliations of most taxa with the three corresponding moss clades Dicranidae, Funariidae and Bryidae, respectively, have been recognized in molecular systematics. However, in some cases, molecular phylogenies have strongly suggested reconsiderations, a prominent example being the genus *Timmia* (Budke et al., 2007; Budke and Goffinet, 2006), whose classical placement in the diplolepideous alternate Bryales has never been found with molecular data. Similarly, several other isolated moss genera (e.g. *Bryobartramia*, *Catoscopium*, *Diphyscium*, *Drummondia*, *Splachnobryum*, *Timmiella*) have also been shown to occupy unexpected positions in molecular phylogenies which contradict their classical systematic assignments.

Like in other plant groups, a few nuclear and several chloroplast loci (most notably *rbc*L, *rps*4, *trn*L/*trn*F) have been used most widely in molecular phylogenetic analyses (see references cited in Goffinet et al. (2004) and Quandt and Stech (2003)). Alternatively, mitochondrial loci have also been suggested for phylogenetic studies, both including a conserved group I (nad5i753, Beckert et al., 1999) as well as a group II intron (nad2i156, Beckert et al., 2001) and recently also a first example for an intergenic mitochondrial region (*nad*5-*nad*4, Groth-Malonek et al., 2007; Wahrmund et al., 2009). Subsequently, the group I intron residing in *nad*5 has frequently been included in various multigene studies

* Corresponding author. Fax: +49 (0)228 73 6467.
E-mail addresses: ute.wahrmund@uni-bonn.de (U. Wahrmund), quandt@uni-bonn.de (D. Quandt), volker.knoop@uni-bonn.de (V. Knoop).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.09.031

(Bell et al., 2007; Bell and Newton, 2005; Hyvönen et al., 2004; Newton et al., 2000; Quandt et al., 2007).

From such comparative and/or multigene studies (Cox et al., 2004; Newton et al., 2000; Wahrmund et al., 2009) the general insight appears to emerge that the phylogenetic trees retrieved from different loci are largely without conflict once data sets of reasonably large taxon sampling have been established. However, at the same time many phylogenetic issues and open questions remain to be clarified, such as the origin of the arthrodontous peristome. The arthrodontous mosses with their morphologically distinct peristomes have clearly been identified as a monophyletic group and raised to the rank of class Bryopsida by many systematists. Nevertheless, the relationships between the three major clades Bryidae, Dicranidae and Funariidae and the placement of key taxa such as the Gigaspermaceae or the enigmatic genus *Timmia* are unresolved.

With this study we present a detailed description of the cobi420 group I intron, its conservation among mosses, the molecular evolution of its secondary structure and its variability and usefulness in revealing relationships within bryophytes. To address the controversial issues of the moss backbone phylogeny outlined above we have included several critical and isolated taxa such as *Timmiella*, *Drummondia*, *Catoscopium*, three genera of the Gigaspermaceae and three species of the genus *Timmia* in our taxon sampling. Moreover, remaining taxonomic gaps in the data matrices for five other loci previously analyzed have been filled and the novel cobi420 locus has been integrated into a 6-loci molecular phylogenetic analysis of 56 moss taxa. Finally, to address potential problems of long branches in the phylogeny, where no recent taxa can be added in the future, we have used a "taxon quenching" approach by systematically removing long branching taxa to determine the robustness of phylogenetic positioning and the impact on the respective node support values.

2. Materials and methods

Extraction of total nucleic acids either followed a CTAB (cetyltrimethyl ammonium bromide) method or used the NucleoSpin plant DNA kit (Macherey-Nagel, Düren, Germany). PCR amplification was performed using either cobi420up (5′-atg att rta aca gct ttt ata gga tac g-3′) and cobi420do (5′-att gcr cta gcy aag ctt gta atg act g-3′) or cobi420up2 (5′-gaa tta gtt tgg tgt ctt gga g-3′) and cobi420do2 (5′-agc caa gtt act ata gtg tct cc-3′) or both combinations in a nested approach. Amplifications of other regions previously established for phylogenetic analyses used primers as reported before (Wahrmund et al., 2009). Each PCR reaction contained 1× PCR buffer, 200 μM of each dNTP, 0.3 μM of each primer, 0.5 U DNA polymerase, 1 μl DNA (≈10–30 ng) and double-distilled water added up to 25 μl. We used GoTaq Polymerase (Promega, Mannheim, Germany) or the PCR extension system (5prime, Hamburg, Germany) with their supplied buffers. The temperature profile of a typical PCR exhibited a 3 min initial denaturation at 95 °C, followed by 35 cycles with 30 s denaturation at 95 °C, 30 s annealing at 50 °C, 3 min elongation at 72 °C each and a final elongation step for 7 min at 72 °C. PCR fragments were ligated into pGEM-T Easy vector (Promega) and positive plasmid clones were commercially sequenced at Macrogen Ltd. (Seoul, Korea). Sequences were assembled and aligned in Mega 4 (Tamura et al., 2007) or PhyDE (Müller et al., 2005), respectively, and manual alignment editing followed criteria laid out by Kelchner (2000). All newly generated sequences are deposited in the primary sequence databases and their accession numbers, as well as those obtained in previous studies, are listed in Table 1. The first 40 bp of *rps*4 sequence entries were not taken into consideration, as the downloaded sequences usually contained the primer information and were of poor quality. We detected five hairpin-associated inversions in the combined data matrix, two in the *cob* intron and three in the *nad*2 intron (Table 2). Inversions were treated as reverse complement in the phylogenetic analyses, in this way substitutions that occurred prior to the inversion event are recovered (see Quandt et al. (2003) for a detailed discussion). Regions of ambiguous alignment (homonucleotide stretches of highly variable length) were excluded from the alignment for phylogenetic analyses (Table 2) but this likewise remained without influence on statistically supported nodes.

Phylogenetic trees were calculated with MrBayes 3.1.2 (Ronquist and Huelsenbeck, 2003) using the GTR + Γ + I model (General Time Reversible model with Gamma distributed substitution rates and a proportion of Invariable sites) of nucleotide substitution, 1,000,000 generations with every 100th tree sampled, four runs in parallel with four chains each (3,000,000 generations, every 1000th tree sampled, 10 parallel runs for the concatenated data set). The consensus tree and posterior probability (PP) of clades were calculated based upon the trees sampled after the chains converged. Stationary was reached well before 250,000 generations (set as burn-in point). PPs of 0.90 and higher were plotted on the maximum likelihood tree inferred with Treefinder (Jobb et al., 2004) using the GTR + Γ + I model. Likewise, Treefinder was used to determine ML bootstrap node support from 100 replicates. For the concatenated data set partitions were defined as chloroplast coding (*rbc*L and *rps*4), mitochondrial coding (coding region of *nad*5, *nad*2, *nad*4, *cob*), mitochondrial group I introns (nad5i753 and cobi420), mitochondrial group II intron (nad2i156) and mitochondrial intergenic spacer region (*nad*5-*nad*4 IGS) and their parameters treated as unlinked.

Secondary structure representation of the *cob* intron was guided by published consensus ribozyme structures (Michel and Westhof, 1990) using the Cech et al. (1994) model as scaffold. Due to the large sequence and length variation of the peripheral stem-loop regions P5 and P9 these were calculated separately using *RNAstructure* 4.6 available at http://rna.urmc.rochester.edu/rnastructure.html (Mathews et al., 2004). Categorization was done using the Group I Intron Sequence and Structure Database (GISSD) at http://www.rna.whu.edu.cn/gissd/ (Zhou et al., 2008).

3. Results

3.1. The new locus: cobi420

Only two additional group I introns other than nad5i753 were identified in the completely sequenced *Physcomitrella patens* (Funariales) mitochondrial genome (Terasawa et al., 2006): one in the *cox*1 and one in the *cob* gene. Given the exceptionally small size of the former intron (cox1i624) of only 395 bp in *Physcomitrella*, we decided to investigate the latter, cobi420 of 954 bp with respect to its conservation in other mosses.

The intron is present in all mosses analyzed here (Table 1), adding to the previous observations on universal conservation of mitochondrial introns in mosses (Pruchner et al., 2001, 2002). We successfully amplified, cloned and sequenced cobi420 and small parts of its flanking exons (14 bp of exon 1 and 22 bp of exon 2) for 56 moss taxa, representing 54 genera. We observed a narrow intron size range of 930–960 bp in most mosses, with unique expansions up to 1017 bp in the Encalyptaceae (*Bryobrittonia*) and 1068 bp in *Sphagnum*. Conversely, three independent decreases in intron sequence lengths were observed: down to 781 bp in *Buxbaumia*, 529 bp in *Schistostega* and only 424 bp in *Leucobryum*. To understand intron conservation and size variations in terms of molecular evolution we established a cobi420 secondary structure (Fig. 1) using a domain by domain folding approach

Table 1
List of moss taxa investigated in this study with systematic assignments according to a recent classification (Goffinet et al., 2008; Goffinet, in press). Taxa with phylogenetic placements differing from the systematic assignments (see Figs. 2 and 3) are underlined. Lengths of the respective cobi420 intron sequences (bold) are indicated. Accessions numbers are given for the five other organelle genome loci analyzed previously. In 18 cases (bold), sequence lengths are given for those 5 loci, where novel sequences were retrieved in the course of this study to fill taxon gaps. We were unable to obtain the *nad*2 amplicon (n.a.) for *Drummondia*. Species names are indicated where another species of the respective genus has been used for the 6-loci phylogenetic analysis. Numbers in brackets indicate the length of a given sequence where only an amplicon of smaller size has been retrieved for the respective species. The sequence of an extremely stable hairpin structure (~20 bp) in cobi420 of *Tetraphis* (~970 bp) could not be fully resolved and remained encoded as Ns.

Taxonomy **Subdivision** **Class** Subclass Superorder Order Family	Species	cobi420	*nad*5-*nad*4 spacer	*nad*5 with nad5i753	*nad*2 with nad2i156	*rbc*L	*rps*4
Sphagnophytina	*Sphagnum fallax*	**FJ870693** **1068 bp** ***S. capillifolium***	DQ098673	AJ001225	AJ299524	AB013673	AY309730 *S. squarrosum*
Takakiophytina	*Takakia lepidozioides*	**FJ870694** **995 bp**	EU095268	AJ291553	AJ299525	AF244565	AF306950
Andreaeophytina	*Andreaea nivalis*	**FJ870695** **1030 bp**	DQ098672	AJ299526	AJ299526	AF478198 *A. nitida*	AJ617675 *A. rupestris*
Bryophytina **Polytrichopsida**	*Atrichum undulatum*	**FJ870696** **1003 bp**	EU095269	AJ001229	AJ299527	AY118236	AY137681
	Dawsonia spec.	**FJ870697** **1003 bp**	EU095270	AY908804 *D. superba* (1111)	EU095309	AF208410 *D. papuana*	AF208419 *D. papuana*
	Oligotrichum hercynicum	**FJ870698** **1003 bp**	EU095271	AY908805 *O. parallelum* (1112)	EU095310	AY118242 *O. austroaligerum*	AY137688
	Pogonatum urnigerum	**FJ870699** **997 bp**	EU095272	AJ291554	AJ299528	AF478206 *P. perichaetiale*	AF478258 *P. perichaetiale*
Tetraphidopsida	*Tetraphis pellucida*	**FJ870700** **~970 bp**	EU095273	AJ224855	AJ299529	AF478203	AF306954
Bryopsida Buxbaumiidae	*Buxbaumia aphylla*	**FJ870701** **781 bp**	EU095274	AJ291555	AJ299531	AF478212	AF231897
Diphysciidae	*Diphyscium sessile*	**FJ870702** **947 bp**	EU095275	Z98972	AJ299530	AF478219 *D. fasciculatum*	AF478270 *D. fasciculatum*
Timmiidae							
Timmiales	*Timmia austriaca*	**FJ870703** **919 bp**	**FJ870748** **1110 bp**	AY312890 *T. megapolitana 1*	**FJ870755** **2215 bp**	AJ275185	AF223035
	Timmia bavarica	**FJ870704** **923 bp**	EU095276	AJ622820	AJ299532	AF478242 *T. sibirica*	AF222902 *T. sibirica*
	Timmia norvegica	**FJ870705** **919 bp**	**FJ870749** **1110 bp**	AY908982 *T. megapolitana 2*	**FJ870756** **2217 bp**	AJ275166 *T. megapolitana*	AF023775 *T. megapolitana*
Funariidae							
Gigaspermales	*Chamaebryum pottioides*	**FJ870706** **936 bp**	**FJ870750** **1105 bp**	AY908983	**FJ870757** **2182 bp**	**FJ870761** **1149 bp**	AF223051
	Gigaspermum repens	**FJ870707** **937 bp**	**FJ870751** **1102 bp**	AY908974	**FJ870758** **2133 bp**	**FJ870762** **1149 bp**	AF231064
	Oedipodiella australis	**FJ870708** **937 bp**	**FJ870752** **1102 bp**	**FJ870754** **1576 bp**	**FJ870759** **2216 bp**	**FJ870763** **1149 bp**	**FJ870765** **651 bp**
Encalyptales	*Bryobrittonia longipes*	**FJ870709** **1017 bp**	EU095277	AY908790 (1097)	EU095311	AJ275168	AF023778
	Encalypta streptocarpa	**FJ870710** **990 bp**	EU095278	AJ291556	AJ299533	AF478239	AF478282
Funariales	*Funaria hygrometrica*	**FJ870711** **954 bp**	EU095279	Z98959	AJ299534	AF005513	AF023776
	Physcomitrella patens	**NC_007945** **954 bp**	DQ098674	Z98960	AJ299535	AP005672	NC_005087
	Physcomitrium pyriforme	**FJ870712** **954 bp**	EU095280	AY908933 *P. lorentzii* (948)	EU095312	EU095319	AF223045
Disceliaceae	*Discelium nudum*	**FJ870713** **932 bp**	EU095281	AY908956 (1088)	EU095313	EU095320	AF223063
Dicranidae							
Scouleriales Drummondiaceae	*Drummondia prorepens*	**FJ870714** **946 bp**	EU095282	AY908926 *D. obtusifolia* (1275)	n.a.	AF232697 *D. obtusifolia*	AF306977
Grimmiales	*Coscinodon cribrosus*	**FJ870715** **946 bp**	EU095283	AY908918 *C. calyptratus* (1088)	EU095314	AB125575	AJ553978
	Grimmia donniana	**FJ870716** **946 bp**	EU095284	AY908919 *G. plagiopodia* (1097)	EU095315	AF231305 *G. pulvinata*	AF222900 *G. pulvinata*
	Racomitrium lanuginosum	**FJ870717** **946 bp**	EU095285	AJ291561	AJ299542	AB125582 *R. japonicum*	AJ553982
Seligeriaceae	*Blindia acuta*	**FJ870718** **947 bp**	EU095286	AY908928 (1131)	EU095316	AF478232 *B. magellanica*	AF478278 *B. magellanica*
Dicranales	*Orthodicranum montanum*	**FJ870719** **944 bp**	EU095287	AJ291558	AJ299537	AF231311 *O. fulvum*	AF231288 *O. fulvum*
Fissidentaceae	*Fissidens cristatus*	**FJ870720** **953 bp**	DQ098675	Z98954	AJ299541	AF226810 *F. mooreae*	AF223056 *F. subbasilaris*

(*continued on next page*)

Table 1 (*continued*)

Taxonomy **Subdivision** **Class** Subclass Superorder Order Family	Species	cobi420	*nad*5-*nad*4 spacer	*nad*5 with nad5i753	*nad*2 with nad2i156	*rbc*L	*rps*4
Ditrichaceae	*Ceratodon purpureus*	**FJ870721** **942 bp**	EU095288	Z98955	AJ299538	EU095321	AJ554004
	Ditrichum cylindricum	**FJ870722** **946 bp**	EU095289	AJ291559	AJ299539	AF231080 *D. ambiguum*	AJ554009 *D. pusillum*
Schistostegaceae	*Schistostega pennata*	**FJ870723** **529 bp**	EU095290	AJ224856	AJ299546	AY631206	AF265359
Leucobryaceae	*Leucobryum glaucum*	**FJ870724** **424 bp**	EU095291	AJ291560	AJ299540	AB124788	AJ554003
Pottiales	*Pottia truncata*	**FJ870725** **942 bp**	EU095292	Z98957	AJ299543	AB125592 *P. intermedia*	AF480987 *P. pallida*
	Timmiella spec.	**FJ870726** **938 bp**	EU095293	AY908958 *T. anomala* (1096)	EU095317	AF478236 *T. crassinervis*	AY908163 *T. anomala*
	Tortula latifolia	**FJ870727** **937 bp**	EU095294	AJ291562	AJ299544	AF226823 *T. obtusissima*	AF481041 *T. muralis*
Cinclidotaceae	*Cinclidotus riparius*	**FJ870728** **945 bp**	EU095295	AJ291563	AJ299545	AF231079 *C. mucronatus*	AF480975 *C. fontinaloides*
Bryidae							
Bryanae Splachnales	*Splachnum ampullaceum*	**FJ870729** **944 bp**	EU095296	EU095308	EU095318	AF231071	AJ251308
Orthotrichales	*Ulota crispa*	**FJ870730** **941 bp**	EU095297	AJ291568	AJ299553	AY631208	AY618370 *U. hutchinsiae*
Hedwigiales	*Hedwigia ciliata*	**FJ870731** **944 bp**	EU095298	Z98966	AJ299554	AF005517	AF478289
Rhacocarpaceae	*Rhacocarpus purpurascens*	**FJ870732** **945 bp**	EU095299	Z98967	AJ299555	AJ275171	AF023815
Bartramiales	*Bartramia halleriana*	**FJ870733** **944 bp**	EU095300	Z98961	AJ299547	AF231090	AF265358
	Plagiopus oederi	**FJ870734** **944 bp**	EU095301	Z98962	AJ299548	DQ481540	AF023833
Bryales	*Catoscopium nigritum*	**FJ870735** **936 bp**	**FJ870753** **1110 bp**	AY908927	**FJ870760** **2218 bp**	**FJ870764** **1149 bp**	AF307001
	Mnium hornum	**FJ870736** **943 bp**	EU095302	AJ291567	AJ299552	AF226820	AF023796
	Pohlia nutans	**FJ870737** **943 bp**	EU095303	AJ291565	AJ299550	AJ275175 *P. cruda*	AF023795 *P. cruda*
Rhizogoniales Aulacomniaceae	*Aulacomnium androgynum*	**FJ870738** **946 bp**	EU095304	AJ291564	AJ299549	AJ275180 *A. turgidum*	AF023809 *A. turgidum*
Orthodontiaceae	*Orthodontium lineare*	**FJ870739** **943 bp**	EU095305	AJ291566	AJ299551	AJ275174	AF023800
Hypnanae	*Herzogiella seligeri*	**FJ870740** **943 bp**	DQ098681	AJ291573	AJ299561	EU095322	AF469815 *H. striatella*
Fontinalaceae	*Fontinalis antipyretica*	**FJ870741** **943 bp**	EU095306	AJ291570	AJ299558	AB050949	AF023817
Amblystegiaceae	*Hygrohypnum ochraceum*	**FJ870742** **944 bp**	DQ098679	AJ291574	AJ299562	EU095323	AY908620 *H. smithii*
	Scorpidium scorpioides	**FJ870743** **943 bp**	DQ098680	AJ291575	AJ299563	EU095324	AY908584
Brachytheciaceae	*Tomentypnum nitens*	**FJ870744** **943 bp**	DQ098677	AJ291572	AJ299560	AB024676	AY908567 *T. falcifolium*
Leucodontaceae	*Pterogonium gracile*	**FJ870745** **943 bp**	EU095307	Z98968	AJ299556	AY631194	AY907970
Neckeraceae	*Homalia trichomanoides*	**FJ870746** **939 bp**	DQ098683	AJ291569	AJ299557	EU095325	AY908276
	Thamnobryum alopecurum	**FJ870747** **943 bp**	DQ098678	AJ291571	AJ299559	AY532392	AF023834

and plotting the data on the Cech et al. (1994) scaffold. In accordance with GISSD the intron can be classified as a IB intron, IB4 receives the highest score. Size variation mainly occurred in the variable stem-loop regions P5 and P9 without any apparent effect on the formation of the core ribozyme structure (Fig. 1). The most variable region is located in the stem loop P9, demonstrated not only by its tremendous 8.5-fold size variation (87 bp in *Leucobryum* vs. 740 bp in *Sphagnum*) but also by substitutions, duplications and the numerous occurrences of indels.

A detailed analysis of the cobi420 intron sequence alignment immediately suggested structural synapomorphies such as an 11 base motif (CTATAATTTTA) that is found as a repetitive element in the terminal part of P5 existing in five or eight tandem repeats exclusively in the two Encalyptaceae of our taxon sampling, *Encalypta* and *Bryobrittonia*. A four base duplicated sequence motif (TGTC) is only present in P5.2b of the three Funariaceae of our taxon sampling.

The stem-loop region P9 (Fig. 1) in contrast contains three large deletions independently acquired by *Buxbaumia*, *Schistostega* and *Leucobryum* as obvious molecular autapomorphies in the given taxon sampling. This is an interesting finding regarding the isolated position of the two former genera. If deletions in this cobi420 re-

Table 2
Locations (i.e., absolute alignment positions in the combined data set) of regions of extreme variability and ambiguous alignment (V) and the detected hairpin-associated inversions (I) that were left out (V) or inverted (I) for phylogenetic analyses.

No.	Position	Region
V1	969–991	cobi420
V2	1139–1170	cobi420
V3	2576–2588	nad2i156
I1	536–541	cobi420
I2	789–794	cobi420
I3	1633–1650	nad2i156
I4	1920–1925	nad2i156
I5	2255–2262	nad2i156

gion will turn out to accumulate in a step-wise manner among taxa closely related to *Leucobryum*, it may be of use in future phylogenetic studies of the species-rich Leucobryaceae. In P9, a four base deletion (ATTA) is found as a synapomorphy of the Gigaspermaceae, similarly a five base pair deletion defines the Encalyptaceae. The most striking synapomorphy, however, is a pentanucleotide (CGGCT) duplication which unites *Drummondia* with the derived Dicranidae (see below).

The cob group I intron cobi420 is unique in occurring in mosses but being absent from all other land plants known so far. It has a homologue in the alga *Chaetosphaeridium globosum* at the same position, however, there are no obvious sequence similarities, suggesting an independent intron gain rather than an orthologue.

3.2. Phylogenetic analyses with cobi420

We aimed for a moss taxon sampling that allows on the one hand for comparison and fusion with previously available data sets and at the same time to address some of the open phylogenetic questions outlined above. To this end we have included *Catoscopi-*

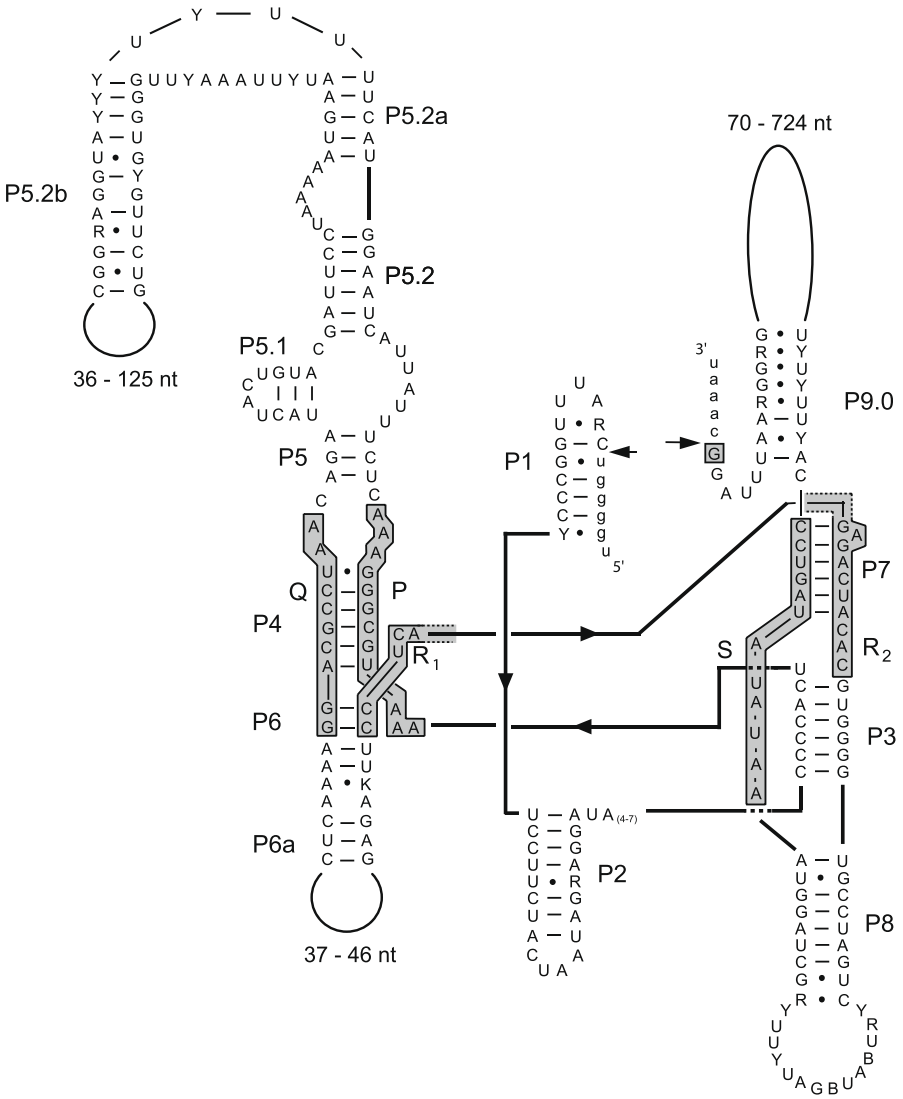


Fig. 1. Secondary structure consensus for the *cob* group I intron cobi420 derived from sequences of 56 moss taxa. All positions with more than five substitutions in the 56 moss sequences (i.e. 10%) are shown in the IUPAC ambiguity code. The conserved sequence motifs P, Q, R, S and the G preceding the *cob* 3′exon are depicted in grey boxes. Length variability is indicated for the variable stem loops P5, P6 and P9, respectively. Splice sites are indicated by arrows. Paired regions are numbered according to international convention (Michel and Westhof, 1990).

um, three species of the Gigaspermaceae, and three *Timmia* species in our studies and in addition to determining the cobi420 also filled the corresponding holes in the *rbc*L/*rps*4/*nad*5/*nad*2/*nad*5-*nad*4 spacer data matrix (Table 1).

The alignment of all 56 moss cobi420 sequences (including 36 bp of the flanking exons) had a length of 1391 characters mainly due to the incorporation of indels in loop regions P5 and P9. A resulting Bayesian phylogenetic tree with node supports independently determined by Maximum Likelihood bootstrapping is shown in Fig. 2. Since all potential outgroup taxa lack the cobi420 intron, the tree is rooted with *Sphagnum* and *Takakia*. The phylogenetic tree based on cobi420 alone is in perfect accordance with previous insights on moss phylogeny (Goffinet and Buck, 2004) and of a phylogenetic resolution comparable to the *nad*5 group I intron when the extended taxon sampling is taken into account (Beckert et al., 1999; Wahrmund et al., 2009). For example, *Buxbaumia* and *Diphyscium* are clearly confirmed in their position as gradual sister taxa to the core arthrodontous mosses. The Gigaspermaceae/-ales are clearly supported as monophyletic. Within the arthrodontous mosses, a basal polytomy of five major clades clearly emerges: the well-supported Bryidae, the Dicranidae (including *Catoscopium*), the Funariidae, the Gigaspermales and the genus *Timmia*.

3.3. Phylogenetic analyses with a 6-loci data set

Remaining taxon gaps were filled for the previously established loci *rbc*L, *rps*4, *nad*5, *nad*2 and the *nad*5-*nad*4 spacer with respect to the 56 moss species sampling (Table 1). A well resolved phylogenetic tree was obtained (Fig. 3) based on a concatenated data set containing 8921 characters (five partitions: chloroplast coding and mitochondrial coding sequences, mitochondrial group I introns, mitochondrial group II intron and mitochondrial spacer region). Individual support for several crucial nodes from analyses of the single-locus data sets is summarized in Table 3. The phylogenetic backbone of moss phylogeny confirms the earlier insights including the placement of crucial genera *Buxbaumia* and *Diphyscium*. The placement of *Tetraphis* (Tetraphidales) as potential sister to the Polytrichopsida among the nematodontous mosses receives no convincing support and has to be further considered unresolved. Again, within the arthrodontous mosses, five clades (Bryidae, Gigaspermaceae, *Timmia*, Funariidae and Dicranidae) become apparent but with much stronger support than in the single-locus analyses (Table 3). Among the Dicranidae, *Drummondia* is now clearly linked into to the Dicranids to the exclusion of *Catoscopium* and *Timmiella*, well supported with the synapomorphic pentanucleotide duplication in cobi240. The exclusion of these genera from their classical assignments to Bryales or Pottiaceae, respectively, is obvious in any case.

There is no resolution for the relationships of the three major clades of arthrodontous mosses, the Dicranidae, Funariidae and Bryidae. An intriguing outcome of the new multilocus phylogenetic analysis, however, is the placement of Gigaspermaceae separate from the Funariidae but rather linked as a potential sister group to the Bryidae. Likewise there is no support for an inclusion of the genus *Timmia* to the Funariidae which is placed with weak support basal to a Gigaspermales–Bryidae dichotomy.

3.4. Addressing remaining problems and potential artefacts

In order to test the robustness of the now derived moss phylogeny, we excluded species sitting on long branches (*Diphyscium* and *Buxbaumia*, alone or combined, *Gigaspermum* or *Catoscopium*) individually from independent phylogenetic analyses. Except for the deleted taxa, the topologies of the resulting trees remained the very same and resulted in only minor changes of the statistical support values suggesting that the molecular phylogenies are unaffected by major artefacts and reflect true phylogenetic cladogenesis. Most notably, the polytomy of the five clades of arthrodontous mosses (not considering the insufficiently supported sister grouping of Gigaspermaceae and Bryidae, Fig. 3) was unaffected by any of these taxon deletions.

4. Discussion

With adequate taxon sampling, recent molecular studies using different genetic loci appear to steer towards a backbone phylogeny of mosses with an ever increasing number of well-supported nodes (Goffinet and Buck, 2004; Newton et al., 2000; Wahrmund et al., 2009). In fact, most of the recent molecular studies have not disagreed on phylogenetic issues but rather identified the very same open questions regarding only weakly supported nodes and/or have identified crucial taxa that were phylogenetically misplaced on the basis of morphological characters (or a lack thereof). The here presented cobi420 locus is a case in point but given the remaining open issues even in multi-loci analyses, yet other molecular markers are obviously needed to address the positioning of still problematic taxa.

We have here presented cobi420 as a second mitochondrial group I intron for phylogenetic analyses in mosses after nad5i753 had been introduced for this purpose 10 years ago (Beckert et al., 1999). As was observed recently in comparison of non-coding mitochondrial loci (Wahrmund et al., 2009), an overall general trend towards size reduction is observed when the derived arthrodontous mosses are compared to basal genera such as *Sphagnum* and *Takakia*. The smallest cobi420 introns with independently acquired large deletions were found in *Schistostega* and *Leucobryum* (Table 1). *Schistostega* also features a smaller *nad*5-*nad*4 spacer region (532 bp vs. mostly over 620 bp) and a smaller nad5i753 intron. These observations could make *Schistostega* in its isolated position among the dicranids an interesting object of study for the wide-scale analysis of its mitochondrial DNA.

Small inversion as reported here (Table 2) that are often associated with hairpins and well known from non-coding plastid regions, where numerous examples can be found in the literature (e.g. Kelchner and Wendel, 1996; Kim and Lee, 2005; Quandt et al., 2003), had as yet not been reported from mitochondrial regions. The now observed frequencies in cobi420 of mitochondrial small-scale sequence inversions, the association with hairpins and their sizes are similar to reports for plastid regions (Quandt et al., 2003; Quandt and Stech, 2004) and point to similar mechanisms creating sequence inversions in the organellar DNAs.

The phylogenetic information contained in cobi420 apparently resembles the one in the *nad*5 locus with nad5i753. Yet, at the same time a very clear point is made that different loci are able to support certain branches that remain weakly supported with other data sets (Table 3). The obvious synapomorphic pentanucleotide duplication in cobi420 strongly supporting the placement of *Drummondia* among the Dicranids (Fig. 3) to the exclusion of *Timmiella* and *Catoscopium* is such a case in point. The basal placement of *Drummondia* in the Dicranids sharing this synapomorphy, however, could not be resolved by cobi420 alone but had to rely on the concatenated data set.

The unresolved polytomy at the base of the arthrodontous mosses that remains unresolved, and is apparently also unaffected by long-branch taxa, suggests that the five groups emerging here may indeed have evolved during a rapid radiation at nearly the same geological time. We suggest that for the time being *Timmia*, the Gigaspermaceae and the (core) Funariidae should not be lumped together into a subclass Funariidae or Timmiidae *sensu lato*. Except for the Bryidae with diplolepideous opposite peristomes, the short branches, notably of the Dicranidae and Funariidae,

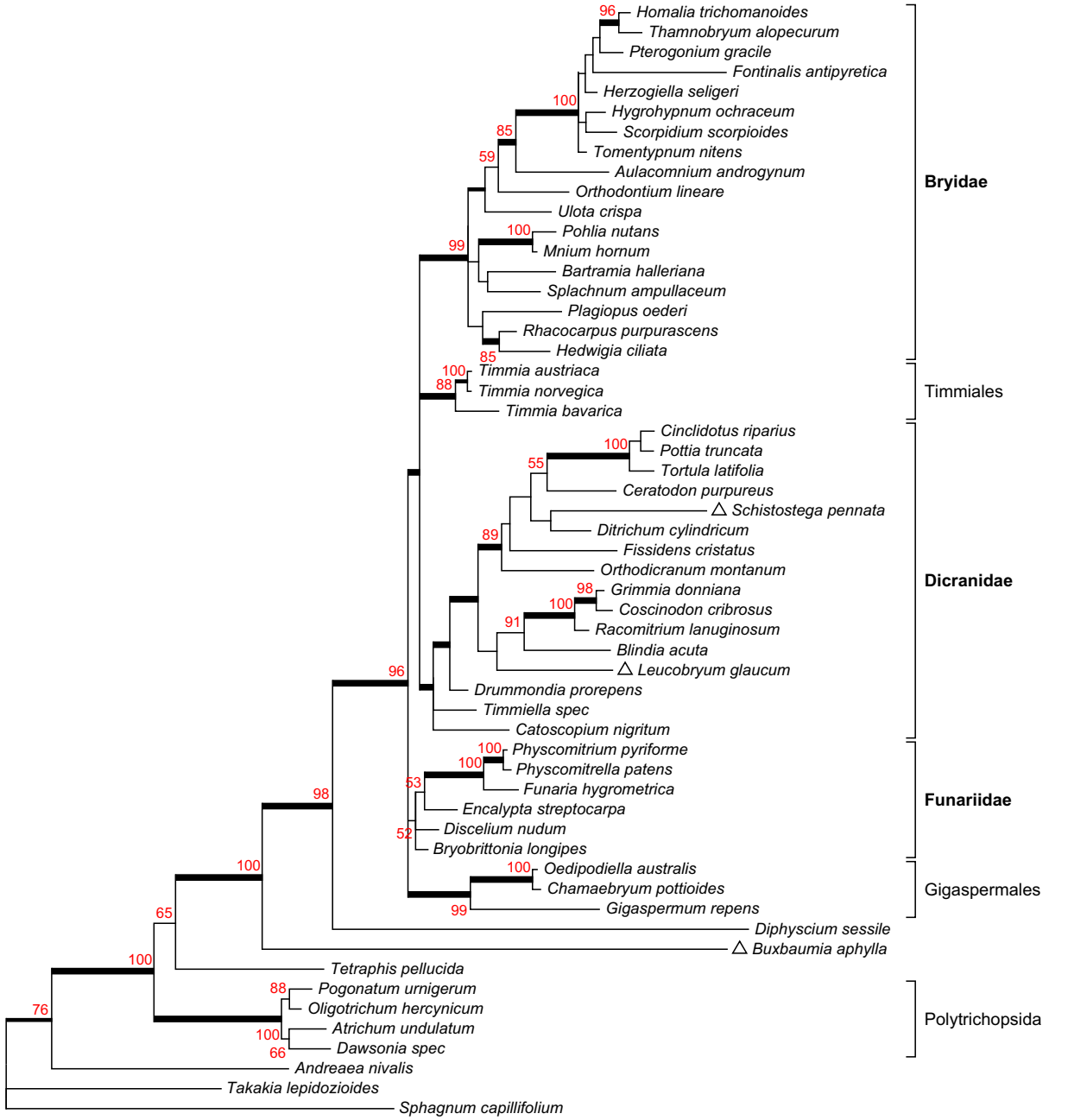


Fig. 2. Bayesian phylogenetic tree based on the cobi420 intron and adjacent exon sequences calculated with the GTR + Γ + I substitution model (1 mio. generations, 4 runs with 4 chains each, every 100th tree sampled). Nodes supported with medium (96–98) or high (99–100) posterior probabilities are indicated by increasingly thickened internode branches. Maximum likelihood bootstrap support values of 100 replicates are shown above the branches where >50%. Open triangles indicate independent large deletions in L9 of the cobi420 intron structures.

could possibly reflect a relatively short period of time in which the morphological novelties in peristome architecture emerged and led to the different radiation events during the subsequent evolution. That period may indeed have been so short that synapomorphies in non-coding regions, that could characterize the relationships between those five clades, either could not accumulate or have been drowned in homoplasies since then. Similarly, the relationships of nematodontous groups along the backbone of early moss phylogeny with a potential sister group relationship of Tetraphidopsida and Polytrichopsida remain an open issue.

The Gigaspermaceae, as represented by the three here included species of the family, are monophyletic in all single data set trees (Table 3) as well as in the fusion tree. The Funariidae *sensu strictu* excluding the Gigaspermaceae are characterized by a large 71 kb inversion in the chloroplast genome that was found in the Funariaceae, Disceliaceae and Encalyptaceae but not in the Gigasperma-

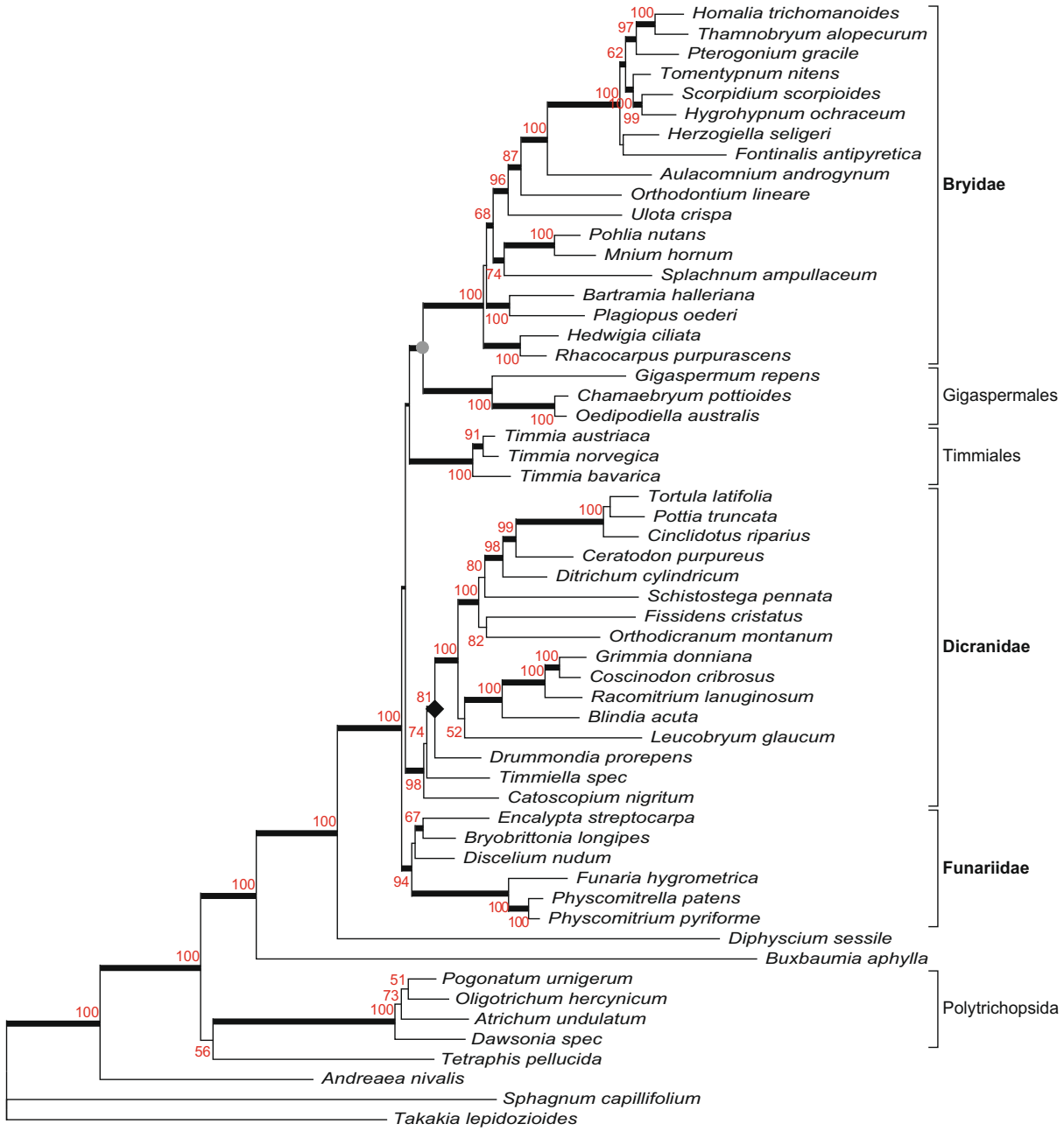


Fig. 3. Bayesian phylogenetic tree based on the concatenated 6-loci data set (cobi420, *nad*5, *nad*2, *nad*5-*nad*4 spacer, *rbc*L, *rps*4; 8921 characters). The data were partitioned into chloroplast and mitochondrial coding sequences, mitochondrial group I introns, mitochondrial group II intron and mitochondrial spacer region and the GTR + Γ + I substitution model was used (3 mio. generations, 10 parallel runs with 4 chains each, every 1000th tree sampled). Nodes supported with medium (96–98) or high (99–100) posterior probabilities are indicated by increasingly thickened internode branches. Maximum Likelihood bootstrap support values are shown where >50%. The clade of *Drummondia* and the core Dicranidae supported by a pentanucleotide duplication in cobi420 is denoted with a filled diamond and the potential sister group relationship linking Gigaspermales and the Bryidae is denoted with a grey circle, respectively.

ceae (Goffinet et al., 2007). Hence, while we see good support to include the isolated Disceliaceae into the Funariidae on the basis of molecular data (Fig. 3 and the cpDNA inversion), we would suggest, in accordance with Stech and Frey (2008), treating Gigaspermaceae/-ales separately at present.

The early-branching position of the isolated genus *Timmia* within the arthrodontous mosses diverging after the split from *Buxbaumia* and *Diphyscium* continues to be unclear also after analysis of the novel cobi420 locus which here shows the same shortcomings as the previously analyzed molecular loci. The same can be said for the first branching lineages of the Bryidae (Hedwigiales, Bartramiales, Splachnales), however, this may well be an issue of present taxon under-sampling, as the branching order among early diverging diplolepideous-alternate taxa was not in the focus of the pres-

Table 3
Comparison of node supports (Bayesian posterior probabilities (PP)/ML bootstrap values (BS)) from the six individual data sets and the concatenated 6-loci data set. Values are indicated for all nodes where at least one data set has a BS exceeding 70 or PP exceeding 0.96. Nodes supports that are affected by a lack of Drummondia nad2 sequence are marked with an asterisk.

Clades	6 loci	cobi420	*nad*5	*nad*2	*nad*5-*nad*4 spacer	*rbc*L	*rps*4
(**Bryophytina**, *Andreaea*)	100/1.00	76/0.96	99/1.00	99/1.00	74/–	99/1.00	100/1.00
Bryophytina	100/1.00	100/1.00	95/1.00	100/1.00	100/1.00	–/–	98/1.00
Polytrichopsida	100/1.00	100/1.00	100/1.00	100/1.00	97/1.00	100/1.00	100/1.00
Bryopsida	100/1.00	100/1.00	91/1.00	100/1.00	–/–	–/–	–/–
Bryopsida excl. *Buxbaumia*	100/1.00	98/1.00	100/1.00	100/1.00	98/1.00	–/–	83/1.00
Bryopsida excl. *Buxbaumia* and *Diphyscium*	100/1.00	96/1.00	96/1.00	–/–	–/–	100/1.00	92/1.00
Timmiales	100/1.00	88/1.00	100/1.00	100/1.00	97/1.00	100/1.00	100/1.00
Funariidae excl. Gigaspermales	94/1.00	–/–	–/–	–/–	–/–	–/–	74/1.00
Funariaceae	100/1.00	100/1.00	93/1.00	100/1.00	100/1.00	100/1.00	100/1.00
Encalyptales	–/1.00	–/–	–/–	–/–	–/–	–/–	89/1.00
Gigaspermales	100/1.00	99/1.00	92/1.00	100/1.00	95/1.00	92/1.00	96/1.00
(Gigaspermaceae, Bryidae)	–/1.00	–/–	–/–	76/1.00	98/1.00	–/–	–/–
(Dicranidae, *Timmiella*, *Catoscopium*)	98/1.00	–/0.99	–/1.00	–/–	–/–	84/1.00	–/–
Dicranidae excl. *Catoscopium*	74/–	–/–	–/–	–/–*	–/–	–/–	–/–
Dicranidae excl. *Catoscopium* and *Timmiella* (Scouleriales, Grimmiales, Dicranales, Pottiales)	81/1.00	–/0.99	–/–	92*/1.00*	–/–	–/–	–/–
(*Ceratodon*, Pottiales)	99/1.00	–/–	–/–	–/–	–/–	76/1.00	–/–
(*Ditrichum*, *Ceratodon*, Pottiales)	98/1.00	–/–	96/1.00	100/1.00	–/–	–/–	–/–
(*Schistostega*, *Ditrichum*, *Ceratodon*, Pottiales)	80/–	–/–	–/–	–/–	–/–	–/–	–/–
(*Orthodicranum*, *Fissidens*)	82/–	–/–	100/0.98	–/–	–/–	–/–	–/–
(Dicranales + Pottiales excl. *Leucobryum*)	100/1.00	89/1.00	91/1.00	95/1.00	–/–	–/–	–/0.97
Grimmiales	100/1.00	100/1.00	96/1.00	100/1.00	100/1.00	94/1.00	99/1.00
(*Coscinodon*, *Grimmia*)	100/1.00	98/1.00	100/1.00	–/–	–/–	90/0.96	88/0.97
(*Blindia*, Grimmiales)	100/1.00	91/–	–/–	100/1.00	–/–	87/1.00	–/–
(*Blindia*, Grimmiales, *Leucobryum*)	–/–	–/–	–/–	–/–	79/1.00	–/–	–/–
(*Fissidens*, *Leucobryum*)	–/–	–/–	–/–	–/–	–/–	80/1.00	–/–
Bryidae	100/1.00	99/1.00	98/1.00	95/1.00	98/1.00	–/1.00	96/1.00
Bartramiales	100/1.00	–/–	77/1.00	85/1.00	–/–	92/1.00	88/0.97
Hedwigiales	100/1.00	85/0.99	100/1.00	100/1.00	79/1.00	100/1.00	76/1.00
Bryales	100/1.00	100/1.00	100/1.00	93/1.00	100/1.00	97/1.00	100/100
(*Splachnum*, Bryales)	74/1.00	–/–	–/–	–/–	90/1.00	–/–	–/–
(*Splachnum*, Bryales, *Ulota*, Rhizogoniales, Hypnanae)	–/1.00	–/–	–/–	–/–	71/0.99	–/–	–/–
(*Ulota*, Rhizogoniales, Hypnanae)	96/1.00	–/0.98	80/1.00	75/0.97	89/1.00	–/–	–/–
(Rhizogoniales, Hypnanae)	87/1.00	–/–	73/–	–/–	71/1.00	–/–	–/–
(Hypnanae, *Aulacomnium*)	100/1.00	85/0.99	92/1.00	–/–	95/1.00	–/–	–/–
Hypnanae	100/1.00	100/1.00	100/1.00	100/1.00	100/1.00	93/1.00	86/1.00
(*Thamnobryum*, *Homalia*)	100/1.00	96/1.00	–/–	–/–	91/–	100/1.00	85/1.00
(*Pterogonium*, *Thamnobryum*, *Homalia*)	97/1.00	–/–	–/–	–/–	–/–	96/1.00	–/–
(*Hygrohypnum*, *Scorpidium*)	99/1.00	–/–	–/–	71/1.00	97/1.00	87/0.99	–/–
(*Tomentypnum*, *Hygrohypnum*, *Scorpidium*)	100/1.00	–/–	–/–	82/1.00	–/–	–/–	–/–
(Amblystegiaceae, Brachytheciaceae, Leucodontaceae, Neckeraceae)	–/1.00	–/–	–/–	–/–	–/–	–/–	–/–
(*Fontinalis*, *Herzogiella*)	–/–	–/–	–/–	–/0.97	–/–	–/–	–/–
(*Fontinalis*, *Herzogiella*, *Tomentypnum*, *Hygrohypnum*, *Scorpidium*)	–/–	–/–	–/–	–/99	–/–	–/–	–/–

ent study. However, where a larger taxon sampling for molecular studies is limited due to scarcity of extant, living taxa, further insights can in any case only be expected to be gained from phylogenetic information hidden in rare synapomorphies of novel molecular loci. Beside such synapomorphies in sequence alignments other genomic features such as the chloroplast genome inversion in the Funariidae mentioned above are obvious examples. In other cases though, such genomic features, similar to indels, somewhat disappointingly rather just add to information on already well supported branches. An example is the chloroplast *rpo*A gene translocation into the nucleus (Goffinet et al., 2005), which rather confirms the linkage of *Diphyscium* into the core Bryopsida to the exclusion of *Buxbaumia* (Beckert et al., 2001).

In some cases, a yet denser taxon sampling of informative loci such as the here established cobi420 may hold promise for future studies. An example along those lines could be phylogenetic studies within the basal-most lineages of the Bryidae, e.g. the Hedwigiales and the Bartramiales, for which the branching patterns are unclear as yet but which may be improved once further genera are sampled. Which additional taxa will ultimately lead to an improvement of phylogenetic trees through extended taxon sampling is not easy to predict. The inclusion of isolated genera or those with doubtful systematic assignment (e.g. *Brachydontium*, *Bryoxiphium*, *Oedipodium*, *Scouleria*, *Tetrodontium*, *Wardia*) in future phylogenetic analyses will be highly important in any case. Somewhat contrastingly, the quest for yet further informative molecular loci is certainly guided by educated guesses, but the discovery of novel molecular synapomorphies elucidating the remaining open issues in a given phylogeny may simply be a matter of luck.

Acknowledgments

The authors are grateful to Drs. Jan-Peter Frahm (Bonn, Germany), Bernard Goffinet (Storrs, CT, USA), Hermann Muhle (Ulm, Germany) for providing plant material and to the Deutsche Forschungsgemeinschaft DFG for funding (Kn411/6). We thank two anonymous reviewers for many helpful suggestions to improve the manuscript.

References

Beckert, S., Muhle, H., Pruchner, D., Knoop, V., 2001. The mitochondrial nad2 gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol. Phylogenet. Evol. 18, 117–126.

Beckert, S., Steinhauser, S., Muhle, H., Knoop, V., 1999. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst. Evol. 218, 179–192.
Bell, N.E., Newton, A.E., 2005. The paraphyly of *Hypnodendron* and the phylogeny of related non-hypnanaean pleurocarpous mosses inferred from chloroplast and mitochondrial sequence data. Syst. Bot. 30, 34–51.
Bell, N.E., Quandt, D., O'Brien, T.J., Newton, A.E., 2007. Taxonomy and phylogeny in the earliest diverging pleurocarps: square holes and bifurcating pegs. Bryologist 110, 533–560.
Budke, J.M., Goffinet, B., 2006. Phylogenetic analyses of Timmiaceae (Bryophyta: Musci) based on nuclear and chloroplast sequence data. Syst. Bot. 31, 633–641.
Budke, J.M., Jones, C.S., Goffinet, B., 2007. Development of the enigmatic peristome of *Timmia megapolitana* (Timmiaceae; Bryophyta). Am. J. Bot. 94, 460–467.
Cech, T.R., Damberger, S.H., Gutell, R.R., 1994. Representation of the secondary and tertiary structure of group I introns. Nat. Struct. Biol. 1, 273–280.
Cox, C.J., Goffinet, B., Shaw, A.J., Boles, S.B., 2004. Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. Syst. Bot. 29, 234–250.
Goffinet, B., in press. Addenda to the classification of mosses. I. Andreaeophytina stat. nov. and Andreaeobryophytina stat. nov. The Bryologist 112.
Goffinet, B., Bayer, R.J., Vitt, D.H., 1998. Circumscription and phylogeny of the Orthotrichales (Bryopsida) inferred from rbcL sequence analyses. Am. J. Bot. 85, 1324–1337.
Goffinet, B., Buck, W.R., 2004. Systematics of the bryophyta (mosses): from molecules to a revised classification. In: Goffinet, B., Hollowell, V., Magill, R. (Eds.), Molecular Systematics of Bryophytes. Missouri Botanical Garden Press, St. Louis, pp. 205–239.
Goffinet, B., Buck, W.R., Shaw, A.J., 2008. Morphology, anatomy, and classification of the Bryophyta. In: Goffinet, B., Shaw, A.J. (Eds.), Bryophyte Biology. Cambridge University Press, Cambridge, pp. 55–138.
Goffinet, B., Shaw, A.J., Cox, C.J., 2004. Phylogenetic inferences in the dung-moss family Splachnaceae from analyses of cpDNA sequence data and implications for the evolution of entomophily. Am. J. Bot. 91, 748–759.
Goffinet, B., Wickett, N.J., Shaw, A.J., Cox, C.J., 2005. Phylogenetic significance of the *rpoA* loss in the chloroplast genome of mosses. Taxon 54, 353–360.
Goffinet, B., Wickett, N.J., Werner, O., Ros, R.M., Shaw, A.J., Cox, C.J., 2007. Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). Ann. Bot. (Lond.) 99, 747–753.
Groth-Malonek, M., Rein, T., Wilson, R., Groth, H., Heinrichs, J., Knoop, V., 2007. Different fates of two mitochondrial gene spacers in early land plant evolution. Int. J. Plant Sci. 168, 709–717.
Huttunen, S., Ignatov, M.S., Müller, K., Quandt, D., 2004. The phylogeny and evolution of epiphytism in three moss families Meteoriaceae, Brachytheciaceae and Lembophyllaceae. Monogr. Syst. Bot. Gard. 98, 328–361.
Hyvönen, J., Koskinen, S., Smith Merrill, G.L., Hedderson, T.A., Stenroos, S., 2004. Phylogeny of the Polytrichales (Bryophyta) based on simultaneous analysis of molecular and morphological data. Mol. Phylogenet. Evol. 31, 915–928.
Jobb, G., von Haeseler, A., Strimmer, K., 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol. Biol. 4, 18.
Kelchner, S.A., 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. Ann. MO Bot. Gard. 87, 482–498.
Kelchner, S.A., Wendel, J.F., 1996. Hairpins create minute inversions in non-coding regions of chloroplast DNA. Curr. Genet. 30, 259–262.
Kim, K.J., Lee, H.L., 2005. Widespread occurrence of small inversions in the chloroplast genomes of land plants. Mol. Cell 19, 104–113.
Mathews, D.H., Disney, M.D., Childs, J.L., Schroeder, S.J., Zuker, M., Turner, D.H., 2004. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. Proc. Natl. Acad. Sci. USA 101, 7287–7292.
Michel, F., Westhof, E., 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. J. Mol. Biol. 216, 585–610.
Müller, K.F., Quandt, D., Müller, J., Neinhuis, C., 2005. PhyDE 0.995: Phylogenetic Data Editor, www.phyde.de.
Newton, A.E., Cox, C.J., Duckett, J.G., Wheeler, J.A., Goffinet, B., Hedderson, T.A.J., Mishler, B.D., 2000. Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. Bryologist 103, 187–211.
Olsson, S., Buchbender, V., Huttunen, S., Enroth, J., Hedenas, L., Quandt, D., 2009. Evolution of the Neckeraceae (Bryophyta): resolving the backbone phylogeny and identifying ancestral character states. Syst. Biodivers. 7.
Pruchner, D., Beckert, S., Muhle, H., Knoop, V., 2002. Divergent intron conservation in the mitochondrial *nad2* gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. J. Mol. Evol. 55, 265–271.
Pruchner, D., Nassal, B., Schindler, M., Knoop, V., 2001. Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol. Genet. Genomics 266, 608–613.
Quandt, D., Bell, N., Stech, M., 2007. Unravelling the knot: the Pulchrinodaceae fam. nov (Bryales). Nova Hedwigia, 21–39.
Quandt, D., Müller, K., Huttunen, S., 2003. Characterisation of the chloroplast DNA psbT-H region and the influence of dyad symmetrical elements on phylogenetic reconstructions. Plant Biol. 5, 400–410.
Quandt, D., Stech, M., 2003. Molecular systematics of bryophytes in context of land plant evolution. In: Sharma, A.K., Sharma, A. (Eds.), Plant Genome: Biodiversity and Evolution. Science Publishers, Enfield, NH, pp. 267–295.
Quandt, D., Stech, M., 2004. Molecular evolution of the trnT(UGU)–trnF(GAA) region in bryophytes. Plant Biol. 6, 545–554.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Stech, M., Frey, W., 2008. A morpho-molecular classification of the mosses (Bryophyta). Nova Hedwigia 86, 1–21.
Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24, 1596–1599.
Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., Sato, N., 2006. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol. Biol. Evol. 24, 699–709.
Wahrmund, U., Rein, T., Müller, K.F., Groth-Malonek, M., Knoop, V., 2009. Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. Plant Syst. Evol. 282, 241.
Zhou, Y., Lu, C., Wu, Q.J., Wang, Y., Sun, Z.T., Deng, J.C., Zhang, Y., 2008. GISSD: Group I Intron Sequence and Structure Database. Nucleic Acids Res. 36, D31–D37.

3.1.3 Ute Wahrmund and Volker Knoop: The lacking roots of mosses: discrepancies in chloroplast and mitochondrial data, including the novel intron locus cox1i624. *Journal of Molecular Evolution*, submitted

Determining the root of mosses was addressed by group I intron cox1i624, which was discovered to be universally conserved between mosses and liverworts, similar to nad5i753. The taxon sampling included representatives of the known early branching moss genera such as *Takakia*, *Sphagnum*, *Andreaea*, *Andreaeobryum*, *Tetraphis*, *Tetrodontium*, some genera of the Polytrichopsida, *Buxbaumia*, *Diphyscium* and several other Bryopsida taxa. In addition, jungermanniid and marchantiid liverworts as well as *Apotreubia* and *Treubia* were incorporated. The intron is present in all mosses and liverworts analysed here. However, while the size of the intron decreased during moss evolution, an increase was observed during liverwort evolution. The phylogenetic trees based on mitochondrial and chloroplast loci were in concordance except for the basal-most moss genus. While mitochondrial data unanimously recognized *Takakia* as first branching moss genus, plastid data suggested a clade of *Sphagnum* and *Takakia* (plus *Andreaea* and *Andreaeobryum* with *rps*4) as sister to all other mosses. Analysing the fused data set resulted in a moderately supported *Sphagnum-Takakia* clade at the base of the mosses, followed by a monophyletic group of *Andreaea* and *Andreaeobryum* (see Fig. 4 in the manuscript). A clade of nematodontous mosses Polytrichopsida and Tetraphidopsida (including *Tetrodontium*) received good statistical support for the first time. Within the Bryopsida, the taxon sampling was clearly too limited to clarify relationships between and within Dicranidae, Bryidae and Funariidae. Nevertheless, the backbone again contained short, unsupported branches. In the liverworts, the resolved topology is congruent with the current understanding of liverwort evolution. *Treubia* and *Apotreubia* form a well supported monophyletic group, sister to an equally well supported joint Marchantiopsida-Jungermanniopsida clade.

Contribution: I sequenced the locus and additionally closed gaps in the other data sets, aligned the sequences, calculated and interpreted the phylogenetic tree, inferred the intron folding and contributed to manuscript preparation.

The lacking roots of mosses: discrepancies in chloroplast and mitochondrial data, including the novel intron locus cox1i624

Ute Volkmar and Volker Knoop

IZMB- Institut für Zelluläre und Molekulare Botanik

Abt. Molekulare Evolution, Universität Bonn, Kirschallee 1, 53115 Bonn

Key Words: bryophytes, group I intron, plant mitochondrial DNA

E-mails: ute.volkmar@uni-bonn.de, volker.knoop@uni-bonn.de

Corresponding author:
Volker Knoop
volker.knoop@uni-bonn.de
+49-(0)228-73-6466
FAX: +49-(0)228-73-6467

Abstract

Liverworts are well supported as the sister group to all other land plants (embryophytes) by molecular data. Observations strongly supporting this earliest dichotomy in embryophyte evolution are the strikingly different introns occurring in the mitochondrial DNAs of liverworts versus non-liverwort embryophytes (NLE), including the mosses. A final conclusion on the most basal lineages of mosses, for which genera such as *Sphagnum* and *Takakia* are the most likely candidates, is lacking. We have now investigated cox1i624, a mitochondrial group I intron conserved between the moss *Physcomitrella patens* and the liverwort *Marchantia polymorpha.* Focussing on a sampling of liverwort and moss genera, which had previously been identified as early branching taxa in their respective clades, we find that group I intron cox1i624 is universally conserved in all 33 mosses and eleven liverworts investigated. The group I intron core secondary structure is well conserved between the two ancient land plant clades. However, whereas dramatic size reductions are seen in the moss phylogeny, the opposite is observed for liverworts. The new cox1i624g1 locus was used for phylogenetic tree reconstruction also in combination with data sets of nad5i753g1 as well as chloroplast loci *rbc*L and *rps*4. The phylogenetic analyses revealed (i) very good support for the Treubiopsida as sister clade to all other liverworts, (ii) a sister group relationship of the nematodontous Tetraphidopsida and Polytrichopsida and (iii) two rivalling hypotheses about the basal-most moss genus with mitochondrial loci favouring an isolated *Takakia* as sister to all other mosses and chloroplast loci prefer a *Takakia-Sphagnum* clade.

Introduction

Correctly identifying the basal lineages of taxa in the phylogeny of a given clade is frequently a matter of suitable outgroup taxa. Even when the closest extant relatives to the studied ingroup are well known, a long time period of separation since the split from the last common ancestor frequently hinders phylogenetic analyses due to a risk for artefacts such as long-branch attraction. The phylogeny of the earliest land plants is a case in point: Although well supported as the closest relatives of land plants (embryophytes), and consequently the most appropriate outgroups for embryophyte phylogeny, charophyte algae such as the genera *Chara* or *Chaetosphaeridium* have accumulated numerous independent character changes after their split from the land plant lineage more that 500 mio. years ago. This is well exemplified by mitochondrial genomic features such as intron occurrence. Many plant mitochondrial introns have apparently been gained with the emergence of the two earliest embryophyte lineages, the liverworts and all other embryophytes, provisionally labelled NLE: non-liverwort embryopyhtes (Pruchner et al. 2001; Qiu et al. 1998). Independent intron gains and losses in the NLE lineage had furthermore proposed a sister group relationship of hornworts and vascular plants (tracheophytes) with mosses as the sister group to the joint clade (Groth-Malonek et al. 2005). This model of early land plant phylogeny which postulates liverworts and mosses as the earliest emerging monophyletic clades has been supported by multi-gene phylogenies at high-density taxon sampling (Qiu et al. 2006).

The majority of mitochondrial introns are independent autapomorphies for mosses or liverworts. However, some of them are conserved between mosses and liverworts and may hence reflect very early intron gains with establishment of the embryophytes. The mitochondrial introns shared between the fully sequenced mitochondrial DNAs of the liverwort *Marchantia polymorpha* (Oda et al. 1992) and the moss *Physcomitrella patens* (Terasawa et al. 2007) are two group I introns (nad5i753 and cox1i624) and two group II introns (atp9i87 and cox1i512). Group I intron nad5i753 was established as a useful locus for

phylogenetic analysis long ago (Beckert et al. 1999) and has subsequently been integrated in several multigene studies of early land plant evolution (Bell and Newton 2005; Cox et al. 2004; Forrest et al. 2006; Hyvönen et al. 2004; Karol et al. 2001). We have recently described the phylogenetic potential of another group I intron in mosses, cobi420, which, however, is neither present in liverworts nor in hornworts, the two other clades of bryophytes (Wahrmund et al. 2009a).

Whereas the unique class of Treubiopsida (Haplomitriopsida) including the genera *Haplomitrium, Apotreubia* and *Treubia* has repeatedly been found as the sister group to all other liverworts (Forrest et al. 2006; Forrest and Crandall-Stotler 2005; Groth-Malonek et al. 2007; Stech and Frey 2008), the earliest branching lineages of mosses is still debatable. Depending on taxon sampling, outgroup choice and the molecular markers that have been used, the resulting trees either had weak or no support for a branching of the basal moss lineages (Cox et al. 2004; Dombrovska and Qiu 2004) or weakly supported a joint *Takakia-Sphagnum* clade as sister to all other mosses (Newton et al. 2000). Similar results were also obtained in phylogenetic studies focussing on liverworts with mosses used as an outgroup at low taxon sampling (Crandall-Stotler et al. 2005; He-Nygrén et al. 2006). Certainly, the inclusion of representatives of other basal moss lineages such as genera of Andreaeophytina, Andreaeobryophytina, Tetraphidopsida or Polytrichopsida can influence the outcome of those analyses.

The present study now investigates the second group I intron shared between moss and liverwort mitochondrial genomes, cox1i624. Phylogenetic reconstructions based on cox1i624 alone and from combined data sets of with *nad*5 including nad5i753 and the intron-less chloroplast loci *rbc*L and *rps*4 were performed. We focused on a wide taxon sampling with representatives of basal moss lineages (*Takakia*, *Sphagnum*, *Andreaea*, *Andreaeobryum*, *Tetraphis* and *Tetrodontium*) in addition to representatives from all major recognized liverwort clades.

Material and Methods

Total nucleic acids were extracted using protocols employing CTAB (Cetyl-trimethyl-ammoniumbromide) as a detergent for cell lysis (Doyle and Doyle 1990) or the NucleoSpin Plant kit (Macherey-Nagel, Düren, Germany). The cox1i624 intron with short flanking *cox*1 gene exon regions was PCR-amplified with oligonucleotide primers cox1i624up (5'-cag cat tcc tac ttt tat tat cyc ttc c-3') and cox1i624do (5'-taa agt tcc tat cag tta ata aca tgg-3') or alternatively with cox1i624up2 (5'-atg cgt ggg cca gga atg-3') and cox1i624do2 (5'-cag atg ctg gta taa aat tgg-3') or both primer combinations in a nested PCR approach. The PCR reactions of 25 µl contained 2.5 mM $MgCl_2$, 200 µM of each dNTP, 0.3 µM of each primer, 0.5 U GoTaq (Promega, Mannheim, Germany) or PCR extender system (*Taq-Pfu* mixture, 5prime, Hamburg, Germany) and 1 µl DNA (appr. 10-30 ng) in the respective buffer supplied by the manufacturer. A typical temperature profile for amplification started with 3 min at 94 °C for denaturation followed by 35 cycles of 30 sec at 94 °C for denaturation, 30 sec of 50 °C for primer annealing and 3 min at 72 °C for elongation, which were followed by a final elongation step at 72 °C for 7 min. The PCR samples were analysed and purified using agarose gel electrophoresis. PCR products of appropriate size were excised, purified with the NucleoSpin Extract kit (Macherey-Nagel) and cloned into the pGEM-T Easy vector (Promega). DNA sequencing of at least three independent clones was done commercially by Macrogen Inc. (Seoul, Korea). Sequences were edited and aligned manually in MEGA 4.0.2 (Tamura et al. 2007). Maximum likelihood phylogenetic trees were calculated with Treefinder Version Oct. 2008 (Jobb et al. 2004) using the GTR+Γ+I substitution model. Bootstrap values were determined from 1000 replicates. Bayesian phylogenetic analyses were done with MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003), likewises using GTR+Γ+I substitution model and analyses runing for 1 million generations with every 100th tree sampled and a burn-in phase of discarded trees until stationarity of log likelihoods was reached. For the combined

data set, a partitioning for mitochondrial coding, mitochondrial intron and chloroplast coding sequences was used with all parameters unlinked.

Results

Conservation and structure of the cox1i624 group I intron

We started our investigations from the observation that group I intron cox1i624 in the *cox*1 gene encoding subunit one of the mitochondrial cytochrome oxidase is conserved in position between the mitochondrial genomes of the liverwort *Marchantia polymorpha* and the moss *Physcomitrella patens* (Fig. 1). With primers targeting the flanking *cox*1 exons in close proximity, we found cox1i624 universally conserved in all 44 additional taxa (33 mosses and eleven liverworts) investigated here (Table 1). Our taxon sampling was aimed to include most early branching moss lineages representing all major superclasses including rare basal taxa such as *Andreaeobryum macrosporum* and *Tetrodontium brownianum*. While the focus of this study was clearly set on identifying the earliest branching lineages of mosses, we nevertheless sampled liverwort taxa of wide phylogenetic distance to include representatives of all major liverwort clades.

The cox1i624 intron sizes ranged from 366 bp (in *Leucobryum*) to 513 bp (in *Takakia*) among mosses, but up to 1167 bp (in *Blasia*) in the liverworts. The secondary structure folding of cox1i624 shown in fig. 2 follows the model proposed for the *Marchantia polymorpha* cox1i624 intron by Ohta and colleagues (1993). The secondary structure model is well supported with our novel sequences determined for diverse liverwort and moss taxa given that most important pairing regions and primary sequence motifs were found conserved. The larger intron sizes in liverworts vs. mosses result mainly from an increase in the size of the unstructured loop L8 sequence and a different intron structure in the 3' part. In mosses, P9 can be folded parsimoniously with only a small loop region of 39 nucleotides (Fig. 2), whereas in liverworts two pairing regions, P9 and P9.1, with larger loop sizes can be formed.

In our present taxon sampling, duplications or deletions are mostly found as autapomorphies of single taxa or synapomorphies for either all mosses or all liverworts. The reduction of cox1i624 intron size is obviously a feature evolving during moss evolution: The basal taxa have longer intron sequences, mainly due to larger loops L5 and L8, while in more derived mosses these regions are reduced in size. A particularly noteworthy observation is made for *Leucobryum glaucum* with the smallest cox1i624 intron size of only 366 nucleotides. Strikingly, such significant autapomorphic size reductions have also been observed for other non-coding mitochondrial loci previously investigated in *Leucobryum* (Wahrmund et al. 2009b; Wahrmund et al. 2009a).

Surprisingly, liverworts show exactly the opposite picture with intron cox1i624 increasing instead of decreasing in size during phylogenetic diversification of the clade. The earliest branching taxa of the Treubiopsida (*Treubia* and *Apotreubia)* show significantly shorter intron sequences than the more derived marchantiid or jungermanniid liverworts (~915 nucleotides vs. ~1167 nucleotides).

Phylogenetic analysis based on cox1i624

A maximum likelihood phylogram based on the cox1i624 alignment of 44 taxa (eleven liverworts, 33 moss taxa) is shown in figure 3. The tree is displayed as midpoint-rooted on the long branch separating the liverworts from the mosses. The liverwort phylogeny shown is supported by the current understanding of Treubiopsida (*Treubia* and *Apotreubia*) as sister clade to a well-supported dichotomy of jungermanniid and marchantiid liverworts (Crandall-Stotler et al. 2005; Forrest et al. 2006). In the marchantiid clade *Blasia* is branching first, followed subsequently by *Sphaerocarpus* and *Lunularia*, whereas *Marchantia*, *Oxymitra* and *Conocephalum* form the crown group, likewise in full accord with phylogenetic insights from other loci (Forrest et al. 2006; Wahrmund et al. 2008).

Among the mosses, bootstrap and posterior probability values (see Table 2) support *Takakia* as the most basal genus, followed by *Sphagnum*. *Andreaea* and *Andreaeobryum* are placed as sister genera on the following branch. *Tetrodontium* and *Tetraphis* are expectedly joint (Tetraphidopsida) and weakly supported as sister group to the Polytrichopsida. Within the Bryopsida resolution and statistical support are low due to the scarcity of characters available and hence, most relationships are lost in polytomies.

Phylogenetic tree based on four loci

The phylogenetic tree based on a combined data set additionally including the mitochondrial *nad*5 locus (including group I intron nad5i753) and the chloroplast loci *rbc*L and *rps*4 strongly resembles the one based on cox1i624 alone (Fig. 4) except for a joint group of *Sphagnum* and *Takakia* as sister to all mosses. Statistical supports for individual nodes from the four individual loci and the concatenated data set are shown in table 2. In the single data set trees, this topology is only found in *rbc*L. In the tree based on *rps*4, *Takakia* is placed as sister to *Sphagnum* in a clade also comprising *Andreaea* and *Andreaeobryum*.

Among the liverworts, the phylogenetic topology derived from the concatenated data is exactly the same as from cox1i624 alone, but with generally increased statistical node supports. The single exception here is interestingly a decreased node support from the concatenated data vs. cox1i624 for the *Aneura-Radula* sister grouping (Fig. 4), which is now well supported by multigene analyses of liverworts at high taxon density (Forrest et al. 2006). Interestingly, this linkage is not recognized by the three other data sets (Tab. 2), suggesting that the here presented novel cox1i624 locus – in the light of its extended size in liverworts – may actually be well-suited alternative novel phylogenetic marker locus for further studies in the liverwort clade.

Within the moss clade, *Tetrodontium* and *Tetraphis* again form a monophyletic clade, now very strongly confirmed as sister group to the Polytrichopsida. Within the Bryopsida,

Buxbaumia is now well supported as the most basal taxon and *Diphyscium* branches off next, a branching order that is well confirmed (Beckert et al. 2001). The unresolved pentachotomy among the derived arthrodontous mosses using cox1i624 alone, however, is only marginally better resolved with the concatenated data set, with short, yet unsupported branches. The trees are fully congruent with the model that the five monophyletic clades emerging here – Timmiidae, Gigaspermales, Funariidae, Bryidae and Dicranidae – have indeed originated near-simultaneously in a very short period of geological time as has recently been observed also with other studies at higher moss taxon sampling (Wahrmund et al. 2009b; Wahrmund et al. 2009a).

Testing for potential artifacts

We tested the robustness of the phylogenetic tree based on the four concatenated data sets by repeating the phylogenetic analyses after successively excluding basal taxa sitting on long branches (*Andreaea*, *Andreaeobryum*, *Buxbaumia*, *Diphyscium*, the Tetraphiodpsida and Polytrichopsida), an approach we had recently referred to as taxon-quenching (Wahrmund et al. 2009a). We generally only observed changes in bootstrap support for nodes close to the excluded (quenched) taxon but found node support (and toplogy, of course) of the more distant nodes unaffected. The most dramatic changes seen was an increase in the bootstrap value for the *Takakia-Sphagnum* sister relationship from 79% to 89% when *Andreaeobryum* was eliminated and similarly an increased support of a Polytrichopsida-Tetraphidopsida clade (from 83 to 96%) when *Buxbaumia* was excluded.

Discussion

The mitochondrial group I intron cox1i624 presented here was intended as a novel, phylogenetically informative locus among bryophytes, mainly to elucidate the evolutionary history of the basal-most branches among mosses. Similar to nad5i753 – the only other

group I intron shared between mosses and liverworts that has previously been established as phylogenetic marker locus (Beckert et al. 1999) – we found coxi624 universally conserved in all taxa investigated. Interestingly, similar to nad5i753 it is also absent in hornworts (Li et al. 2009).

We have included representatives of all basal moss classes, i.e. of the monogeneric clades Takakiophytina, Sphagnophytina, Andreaeophytina and Andreaeobryophytina, as well as the first branching genera of the Bryophytina, the Tetraphidopsida and Polytrichopsida in our taxon sampling. Not explicitly addressing the issue, previous studies also resolved the root of mosses with varying outcome. Newton and colleagues (2000) used three chloroplast and one nuclear marker, but only in combination with morphological characters a monophyly of *Takakia* and *Sphagnum* sister to all other mosses received low MP bootstrap support. Cox and colleagues (2004) analysed eight different loci from all three genomes, which resulted in a higher, yet still only moderate, ML bootstrap value for this topology. Certainly, aside from different loci investigated, different taxon samplings also affect the outcome of molecular phylogenies and only comparisons with identical taxon samplings such as the one shown in table 2 are truly informative on that issue.

For this study we have strived to fill remaining gaps in basal lineages and included key genera such as *Andreaea*, *Andreaeobryum, Tetrodontium* as well as both described species of *Takakia* in our sampling to minimize the risk of long-branch artefacts. Our phylogenetic reconstruction has identified *Andreaea* and *Andreaeobryum* as sister genera with good statistical support, a relationship seen before without or with low statistical support (Cox et al. 2004; Goffinet et al. 2001; Newton et al. 2000). Morphologically, the two genera are reasonably similar which may justify their joint treatment in a joint taxon. Similarly, *Tetrodontium*, a presumed sister genus of *Tetraphis* in the Tetraphidopsida was indeed resolved at that position in the molecular phylogenies. We wished to resolve the hitherto doubtful sister relationship of Tetraphidopsida and Polytrichopsida, which emerged with some

molecular data sets but never received significant support. The sister relationship of Tetraphidopsida and Polytrichopsida was identified with good support in the concatenated data set, mostly with phylogentic signal from cox1i624 (Tab. 2). Morphologically, both classes share a nematodontous peristome – the teeth of the peristome are build up by whole cells rather than cell wall remnants as in the class of Bryopsida, the arthrodontous mosses. Our taxon quenching analysis shows an increased branch reliability when both genera of the Tetraphidopsida – *Tetraphis* and *Tetrodontium* – are included, highlighting again the importance of a careful and extensive taxon sampling to reduce the risks of long-branch attraction.

No increased resolution or support on the other hand was found for the five monophyletic clades in the core Bryophytina: Timmiidae, Gigaspermales, Funariidae, Dicranidae and Bryidae. Given that different and mostly weakly supported relationships of these five clades have been identified (see Wahrmund et al. 2009a for discussion and references) a simultaneous formation of all five clades during a very short period of time is indeed a likely explanation. None of the coding sequences, introns or intergenic spacer regions investigated so far had been able to resolve the relationship between them. Given that this polytomy can hardly be addressed by yet further taxon sampling, an appropriate molecular marker retaining phylogenetic information on the closely space cladogenic events remains to be found. In the light of its small size among mosses the cox1i624 locus is, however, very unlikely to contribute to remaining questions of moss phylogeny such as this unresolved pentachotomy of derived arthorodontrous mosses. In contrast, in the light of its significant size extension among the liverworts, cox1i624 may prove a useful tool to elucidate remaining open question in the phylogeny of that division of bryophytes, the presumed oldest clade of land plants.

In conclusion, the novel cox1i624 locus for phylogenetic studies among bryophytes in combination with extended taxon sampling for all early branching lineages resulted in a well-

supported backbone phylogeny of mosses resolving hitherto unclear issues: a robust resolution of the sister group relationships of *Andreaea* and *Andreaeobryum* and of Tetraphidopsida and Polytrichopsida. On the other hand, whereas both mitochondrial group I introns, cox1i624 and nad5i753, suggest the placement of the genus *Takakia* as sister to all other mosses, this topology is overruled by the two chloroplast genes, *rbc*L and *rps*4, finding *Takakia* together with *Sphagnum* in a well-supported sister grouping (notably with *rps*4 suggesting a joint clade of *Takakia, Sphagnum*, Andreaeophytina and Andreaeobryophytina, see Table 2).

A systematic assignment of *Takakia* on morphological grounds alone was difficult for a long time when only the gametophyte was known. When antheridia and sporophytes were discovered, *Takakia* was placed into the mosses and assumed to be close to the Andreaeophytina (Renzaglia et al. 1997; Smith and Davison 1993). This is by now well supported by molecular data including the observations that *Takakia* universally shares (other) mitochondrial introns with mosses, not with liverworts (Beckert et al. 2001; Dombrovska and Qiu 2004; Pruchner et al. 2001). One unique feature of *Takakia* development, found nowhere else in mosses, is the opening of the spore capsule with a single vertical fissure – even *Andreaea* and *Andreaeobryum* have multiple fissures (Smith and Davison 1993). In *Sphagnum* the tip of the spore capsule falls off, a mechanism that has independently evolved also in the Bryophytina (Goffinet et al. 2008). Thus, the modes of spore capsule opening seem to vary independently throughout the most basal moss clades. *Takakia* could then have either retained the ancestral plesiomorphic state or developed a unique mode of its own during its long time of independent evolution separate from the other mosses.

A well resolved root of moss phylogeny could reliably trace the directions of molecular and morphological character changes during moss evolution. On the molecular level, significant sequence divergence distinguishes members of the Bryopsida from the basal lineages, usually not only displayed by substitutions but also by length variations in non-

coding regions. The longest mitochondrial intron and spacer regions have been found in *Sphagnum* and *Takakia*, whereas a reduction in size in addition to a high conservation of structural features (e.g. paired intron regions) occurred during moss evolution (Beckert et al. 1999; Beckert et al. 2001; Wahrmund et al. 2009b; Wahrmund et al. 2009a). In chloroplasts, a similar trend is discernible in intron and spacer regions (Quandt and Stech 2004). Hence, we may assume that the ancestral mosses may contain somewhat larger chloroplast and mitochondrial genomes due to extended intron and intergenic sequences, which were gradually streamlined during diversification of the more derived lineages.

Acknowledgements

We wish to thank Prof. Jan-Peter Frahm, Bonn, Dr. Hermann Muhle, Ulm, Prof. Rüdiger Mues, Saarbrücken and Prof. Yin-Long Qiu, Ann Arbor, for their friendly help in providing biological materials. We are grateful to the Deutsche Forschungsgemeinschaft DFG for support of our work via grant Kn411/6 and wish to thank Milena Groth-Malonek for providing *nad*5 liverwort sequences and the students of our lab courses PMEP and WP20 for their concentrated work and successful contributions to this project. Preparing this manuscript was finished on the 20th anniversary of the fall of the Berlin wall (on November, 9th in 1989), experienced by the senior author of this paper in what was West-Berlin at this time. We wish to express our deepest gratitude to all people who participated in initiating the peaceful revolution as well as those developing it a historic success and in this way also paving the way for the GDR-born first author to freely choose her career-path.

Figure legends

Fig. 1. Overview of the mitochondrial *cox*1 gene with group I introns (boxes) and group II introns (circles), present in *Marchantia polymorpha* (dark grey shading) or *Physcomitrella patens* (light grey shading) or in both (hatched). Intron nomenclature is according to the preceding exon nucleotide of the mature coding sequence as previously suggested (Dombrovska and Qiu 2004). The here analysed intron group I cox1i624 is highlighted.

Fig. 2. Secondary structure consensus model of mitochondrial group I intron cox1i624 derived from sequences of 33 mosses and eleven liverworts. IUPAC nucleotide ambiguity codes are shown when more than 10 % (i.e. four taxa) differ in a given sequence position. Exon regions are shown in italics, the splice sites are marked with arrows. Conserved base pairings are numbered P1 through P9, the size ranges of loops L6, L8, L9 and L9.1 are indicated. Mosses feature a simpler secondary structure at the cox1i624 end with only P9 (dotted line) as compared to the two paired region (P9 and P9.1) in liverworts.

Fig. 3. Maximum-Likelihood phylogram based on cox1i624 and adjacent exon regions calculated with GTR+Γ+I substitution model by Treefinder. ML bootstrap values >50 % (1000 bootstrap replicates) are shown. Bayesian posterior probability supports (GTR+Γ+I substitution model, 4 runs with 4 chains each, 1 million generations, every 100th tree sampled, burnin=400 trees) are reflected by increased width of respective branches (0.96-0.98: 3 pt, 0.99-1.00: 5 pt).

Fig. 4. Maximum-Likelihood phylogram based on the fused data set including mitochondrial loci cox1i624 (and adjacent exon regions), *nad*5 (including nad5i753) and chloroplast loci *rbc*L and *rps*4. Phylogeny was determined using the GTR+Γ+I substitution model and data partitioning into mitochondrial coding, mitochondrial intro and chloroplast sequences with

Treefinder. ML bootstrap values >50 % (1000 replicates) are shown. Bayesian posterior probabilities (GTR+Γ+I substitution model and data partitioning [mitochondrial coding, chloroplast coding and mitochondrial intron regions], 4 runs with 4 chains each, 1 million generations, every 100th tree sampled, burnin=400 trees) are reflected by thickness of branches (0.99-1.00 by 5 pt, 0.96-0.99 by 3 pt).

Figure 1

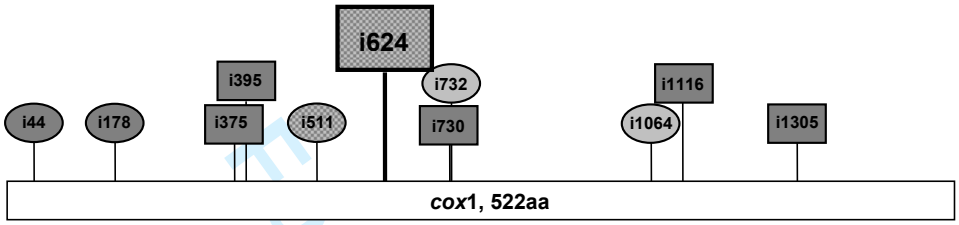


Figure 2

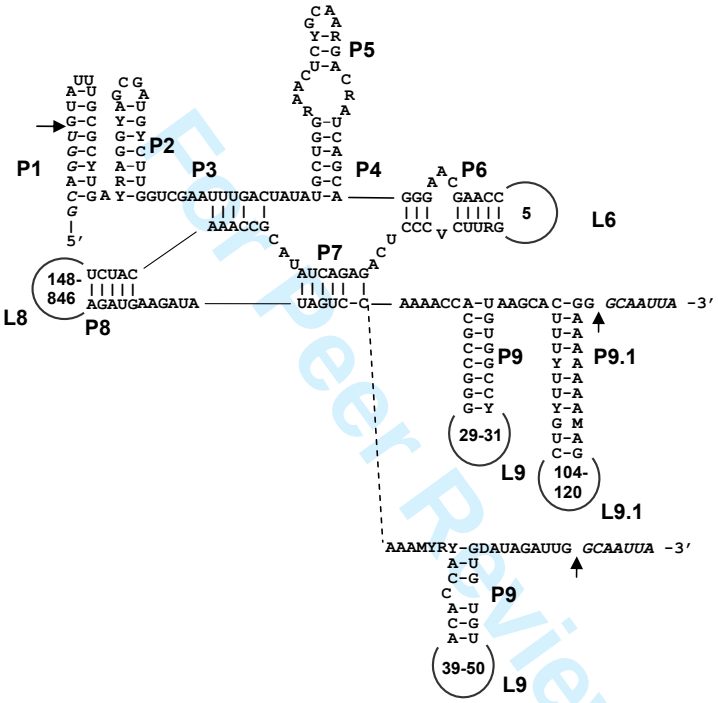


Figure 3

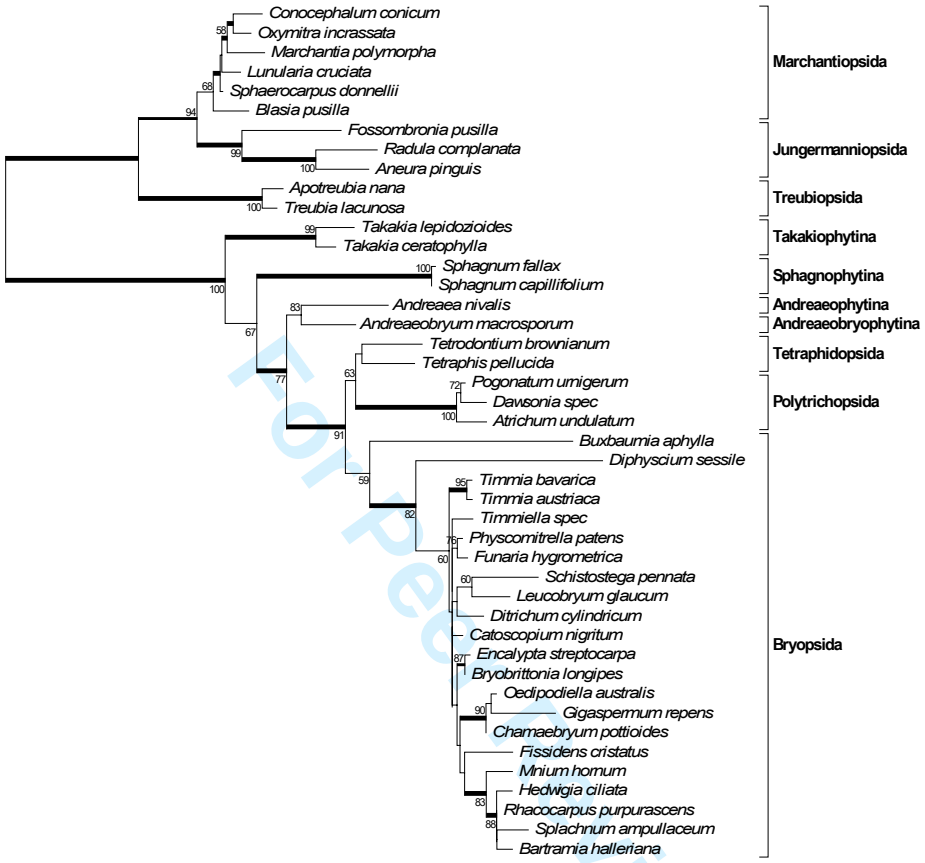


Figure 4

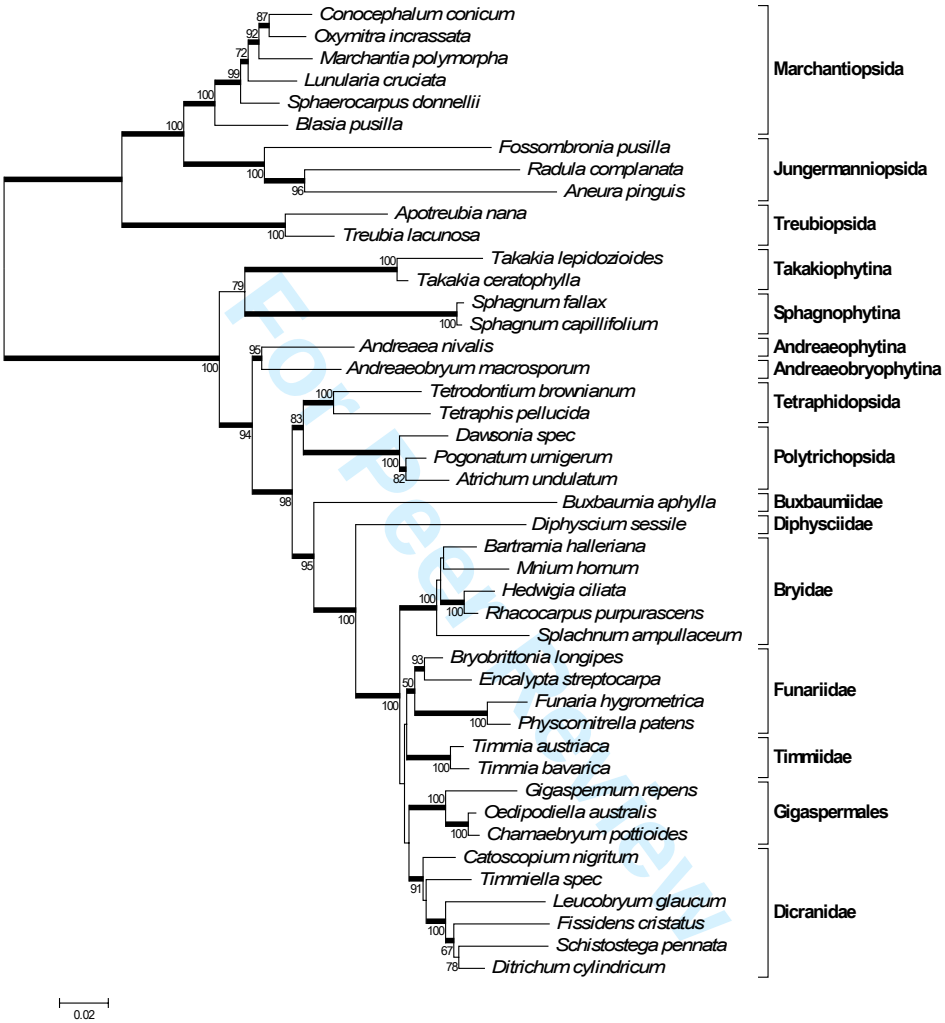


Table 1. List of moss and liverwort taxa and the accession numbers for the loci used in phylogenetic analyses investigated during this study. Respective lengths of cox1i624 are indicated. All cox1i624 sequences except for *Marchantia* and *Physcomitrella* and the *nad*5 sequences of *Andreaeobryum* and *Treubia* were determined in the course of this study. Epithets are given where sequences from other species of the same genus were used in the concatenated data set.

Taxonomy Subdivision Class Subclass Order Family	Species	cox1i624	*nad*5 with nad5i753	*rbc*L	*rps*4
MOSSES **Sphagnophytina**	*Sphagnum capillifolium*	502 bp XYZ	AY309560 *S. cuspidatum*	AY309694 *S. cuspidatum*	AF478246 *S. cuspidatum*
	Sphagnum fallax	507 bp XYZ	AJ001225	AB013673	AY309730 *S. squarrosum*
Takakiophytina	*Takakia ceratophylla*	526 bp XYZ	DQ268963	DQ646000	DQ268993
	Takakia lepidozioides	513 bp XYZ	AJ291553	AF244565	AF306950
Andreaeophytina	*Andreaea nivalis*	483 bp XYZ	AJ299526	AF478198 *A. nitida*	AJ617675 *A. rupestris*
Andreaeobryophytina	*Andreaeobryum macrosporum*	473 bp XYZ	XYZ	AF231059	AF306953
Bryophytina **Polytrichopsida**	*Atrichum undulatum*	479 bp XYZ	AJ001229	AY118236	AY137681
	Dawsonia spec.	479 bp XYZ	AY908804 *D. superba* (1111)	AF208410 *D. papuana*	AF208419 *D. papuana*
	Pogonatum urnigerum	480 bp XYZ	AJ291554	AF478206 *P. perichaetiale*	AF478258 *P. perichaetiale*
Tetraphidopsida	*Tetrodontium brownianum*	485 bp XYZ	AY908809	AF478205	AY908022
	Tetraphis pellucida	487 bp XYZ	AJ224855	AF478203	AF306954
Bryopsida Buxbaumiidae	*Buxbaumia aphylla*	477 bp XYZ	AJ291555	AF478212	AF231897
Diphysciidae	*Diphyscium sessile*	380 bp XYZ	Z98972	AF478219 *D. fasciculatum*	AF478270 *D. fasciculatum*
Timmiidae Timmiales	*Timmia austriaca*	391 bp XYZ	AY312890 *T. megapolitana*	AJ275185	AF223035
	Timmia bavarica	391 bp XYZ	AJ622820	AF478242 *T. megapolitana*	AF222902 *T. sibirica*
Funariidae Encalyptales	*Bryobrittonia longipes*	391 bp XYZ	AY908790	AJ275168	AF023778
	Encalypta streptocarpa	390 bp XYZ	AJ291556	AF478239	AF478282
Funariales	*Funaria hygrometrica*	395 bp XYZ	Z98959	AF005513	AF023776
	Physcomitrella patens	395 bp XYZ	Z98960	AP005672	NC_005087
Gigaspermales	*Chamaebryum pottioides*	391 bp XYZ	AY908983	FJ870761	AF223051
	Gigaspermum repens	395 bp XYZ	AY908974	FJ870762	AF223050
	Oedipodiella australis	391 bp XYZ	FJ870754	FJ870763	FJ870765
Dicranidae Dicranales					
Fissidentaceae	*Fissidens cristatus*	395 bp XYZ	Z98954	AF226810 *F. mooreae*	AF223056 *F. subbasilaris*
Ditrichaceae	*Ditrichum cylindricum*	391 bp XYZ	AJ291559	AF231080 *D. ambiguum*	AJ554009 *D. pusillum*
Leucobryaceae	*Leucobryum glaucum*	366 bp XYZ	AJ291560	AB124788	AJ554003
Schistostegaceae	*Schistostega pennata*	388 bp XYZ	AJ224856	AY631206	AF265359
Pottiales	*Timmiella spec.*	391 bp XYZ	AY908958 *T. anomala*	AF478236 *T. crassinervis*	AY908163 *T. anomala*
Bryidae					
Splachnales	*Splachnum ampullaceum*	390 bp XYZ	EU095308	AF231071	AJ251308
Hedwigiales	*Hedwigia ciliata*	390 bp XYZ	Z98966	AF005517	AF478289
Rhacocarpaceae	*Rhacocarpus purpurascens*	390 bp XYZ	Z98967	AJ275171	AF023815
Bartramiales	*Bartramia halleriana*	390 bp XYZ	Z98961	AF231090	AF265358
Bryales	*Catoscopium nigritum*	386 bp XYZ	AY908927	FJ870764 1149 bp	AF307001

	Mnium hornum	390 bp XYZ	AJ291567	AF226820	AF023796
LIVERWORTS Haplomitriopsida/ Treubiopsida	*Apotreubia nana*	915 bp XYZ	EU519192	AY877389	DQ268983
	Treubia lacunosa	919 bp XYZ	XYZ	AY507428	AY507468
Marchantiopsida	*Blasia pusilla*	1167 bp XYZ	EU519187	DQ645982	AY507436
	Conocephalum conicum	1133 bp XYZ	EU519188	AY688778	DQ220678
	Lunularia cruciata	1130 bp XYZ	AJ001002	U87077	AY688795
	Marchantia polymorpha	1148 bp XYZ	NC001660	NC001660	X04465
	Oxymitra incrassata	1124 bp XYZ	EU519190	EU519195	EU519197
	Sphaerocarpus donnellii	1135 bp XYZ	AJ001033	AY507425 *S. texanus*	AY608110 *S. texanus*
Jungermanniopsida					
Porellales	*Radula complanata*	1073 bp XYZ	AY688767	AY302461	AM398295
Fossombroniales					
	Fossombronia pusilla	1127 bp XYZ	AJ000699	AF536231	AY507440
Metzgeriales	*Aneura pinguis*	1039 bp XYZ	AY688744	AY507391	DQ983852

Table 2. Branch support for the concatenated 4-loci and the individual data sets. All clades receiving minimally 75% bootstrap support from at least one of the data sets are shown. Bayesian posterior probabilities are given after the slash. Dashes denote absence of the given clade in the respective phylogeny.

Clades	4 loci	cox1i624	*nad*5	*rbc*L	*rps*4
Liverworts*	100/1.00	100/1.00	100/1.00	99/1.00	100/1.00
Treubiopsida	100/1.00	100/1.00	100/1.00	100/1.00	100/1.00
(Jungermanniopsida, Marchantiopsida)	100/1.00	94/0.98	77/0.95	96/1.00	94/1.00
Jungermanniopsida	100/1.00	99/1.00	93/1.00	94/1.00	98/1.00
(*Aneura*, *Radula*)	96/1.00	100/1.00	-/-	-/-	-/-
Marchantiopsida	100/1.00	68/<0.90	66/0.95	100/1.00	64/<0.90
Marchantiopsida excl. *Blasia*	99/1.00	<50/0.99	92/1.00	77/1.00	73/1.00
Marchantiopsida excl. *Blasia*, *Sphaerocarpus*	72/0.99	<50/<0.90	<50/<0.90	62/0.97	-/-
(*Conocephalum*, *Oxymitra*, *Marchantia*)	92/1.00	58/1.00	-/-	75/1.00	<50/<0.90
(*Conocephalum*, *Oxymitra*)	87/1.00	<50/1.00	65/0.99	87/1.00	-/-
Mosses*	100/1.00	100/1.00	100/1.00	99/1.00	100/1.00
Mosses excl. *Takakia*	-/-	67/<0.90	85/0.99	-/-	-/-
Mosses excl. *Takakia*, *Sphagnum*	94/1.00	77/1.00	90/1.00	58/1.00	-/-
(*Sphagnum*, *Takakia*)	79/<0.90	-/-	-/-	98/1.00	95/0.98
(*Sphagnum*, *Takakia*, *Andreaeobryum*, *Andreaea*)	-/-	-/-	-/-	-/-	79/<0.90
(*Andreaea*, *Andreaeobryum*)	95/1.00	83/0.93	76/0.92	<50/0.95	-/-
Mosses excl. *Takakia*, *Sphagnum*, *Andreaea*, *Andreaeobryum*	98/1.00	91/1.00	90/1.00	<50/0.99	-/-
(*Tetraphis*, *Tetrodontium*)	100/1.00	<50/-	73/0.98	92/1.00	93/1.00
(Tetraphidopsida, Polytrichopsida)	83/0.99	63/0.91	<50/<0.90	<50/0.96	-/-
Bryopsida	95/1.00	59/0.91	88/1.00	<50/0.98	-/-
Bryopsida excl. *Buxbaumia*	100/1.00	82/1.00	99/1.00	55/0.99	60/<0.90
Bryopsida excl. *Buxbaumia*, *Diphyscium*	100/1.00	60/<0.90	95/1.00	100/1.00	73/<0.90
Timmiidae	100/1.00	95/1.00	100/1.00	100/1.00	99/1.00
Gigaspermales	100/1.00	90/1.00	88/1.00	83/0.98	100/1.00
(*Oedipodiella*, *Chamaebryum*)	100/1.00	-/-	100/1.00	98/1.00	81/0.99
(*Bryobrittonia*, *Encalypta*)	93/1.00	87/0.96	-/0.92	-/-	95/1.00
(*Physcomitrella*, *Funaria*)	100/1.00	76/0.95	100/1.00	100/1.00	100/1.00
Dicranidae	91/1.00	-/-	<50/1.00	83/1.00	<50/0.97
Dicranidae excl. *Catoscopium*, *Timmiella*	100/1.00	-/-	92/1.00	62/1.00	73/1.00
(*Ditrichum*, *Schistostega*, *Fissidens*)	67/1.00	-/-	94/1.00	-/-	54/<0.90
(*Ditrichum*, *Schistostega*)	78/0.94	-/-	<50/-	<50/<0.90	<50/-
(*Fissidens*, *Leucobryum*)	-/-	-/-	-/-	85/1.00	-/-
Bryidae	100/1.00	83/1.00	100/1.00	<50/1.00	99/1.00
Bryidae excl. *Mnium*	-/-	88/0.98	<50/-	-/-	-/-
Bryidae excl. *Splachnum*	<50/-	-/-	-/-	75/<90	-/-
(*Hedwigia*, *Rhacocarpus*)	100/1.00	-/-	100/1.00	100/1.00	81/1.00

*support of branch separating mosses and liverworts

References

Beckert S, Muhle H, Pruchner D, Knoop V (2001) The mitochondrial nad2 gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. Mol Phylogenet Evol 18:117-126

Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst Evol 218:179-192

Bell NE and Newton AE (2005) The paraphyly of *Hypnodendron* and the phylogeny of related non-hypnanaean pleurocarpous mosses inferred from chloroplast and mitochondrial sequence data. Syst Bot 30:34-51

Cox CJ, Goffinet B, Shaw AJ, Boles SB (2004) Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. Syst Bot 29:234-250

Crandall-Stotler B, Forrest LL, Stotler RE (2005) Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). Taxon 54:299-316

Dombrovska E and Qiu YL (2004) Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol 32:246-263

Doyle JJ and Doyle JL (1990) Isolation of plant DNA from fresh tissue. Focus 12:13-15

Forrest LL and Crandall-Stotler BJ (2005) Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. J Hattori Bot Lab 97:127-159

Forrest LL, Davis EC, Long DG, Crandall-Stotler BJ, Clark A, Hollingsworth ML (2006) Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. Bryologist 109:303-334

Goffinet B, Buck WR, Shaw AJ (2008) Morphology, anatomy, and classification of the Bryophyta. In: Goffinet B, Shaw AJ. (eds) Bryophyte Biology. Cambridge University Press, Cambridge, pp 55-138

Goffinet B, Cox CJ, Shaw AJ, Hedderson TAJ (2001) The bryophyta (mosses): Systematic and evolutionary inferences from an *rps4* gene (cpDNA) phylogeny. Annals of Botany 87:191-208

Groth-Malonek M, Pruchner D, Grewe F, Knoop V (2005) Ancestors of *trans*-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. Mol Biol Evol 22:117-125

Groth-Malonek M, Wahrmund U, Polsakiewicz M, Knoop V (2007) Evolution of a pseudogene: Exclusive survival of a functional mitochondrial *nad7* gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. Mol Biol Evol 24:1068-1074

He-Nygrén X, Juslén A, Ahonen I, Glenny D, Piippo S (2006) Illuminating the evolutionary history of liverworts (Marchantiophyta) - towards a natural classification. Cladistics 22:1-31

Hyvönen J, Koskinen S, Smith Merrill GL, Hedderson TA, Stenroos S (2004) Phylogeny of the Polytrichales (Bryophyta) based on simultaneous analysis of molecular and morphological data. Mol Phylogenet Evol 31:915-928

Jobb G, von Haeseler A, Strimmer K (2004) TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evol Biol 4:18

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

Li L, Wang B, Liu Y, Qiu YL (2009) The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. J Mol Evol 68:665-678

Newton AE, Cox CJ, Duckett JG, Wheeler JA, Goffinet B, Hedderson TAJ, Mishler BD (2000) Evolution of the major moss lineages: Phylogenetic analyses based on multiple gene sequences and morphology. Bryologist 103:187-211

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992) Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1-7

Ohta E, Oda K, Yamato K, Nakamura Y, Takemura M, Nozato N, Akashi K, Ohyama K, Michel F (1993) Group I introns in the liverwort mitochondrial genome: the gene coding for subunit 1 of cytochrome oxidase shares five intron positions with its fungal counterparts. Nucleic Acids Res 21:1297-1305

Pruchner D, Nassal B, Schindler M, Knoop V (2001) Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol Genet Genom 266:608-613

Qiu YL, Cho YR, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671-674

Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci U S A 103:15511-15516

Quandt D and Stech M (2004) Molecular evolution of the trnT(UGU)-trnF(GAA) region in bryophytes. Plant Biol 6:545-554

Renzaglia KS, McFarland KD, Smith DK (1997) Anatomy and ultrastructure of the sporophyte of *Takakia ceratophylla* (Bryophyta). Am J Botan 84:1337-1350

Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinf 19:1572-1574

Smith DK and Davison PG (1993) Antheridia and sporophytes in *Takakia ceratophylla* (Mitt.) Grolle: Evidence for reclassification among the mosses. J Hattori Bot Lab 73:263-271

Stech M and Frey W (2008) A morpho-molecular classification of the mosses (Bryophyta). Nova Hedwigia 86:1-21

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596-1599

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N (2007) The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. Mol Biol Evol 24:699-709

Wahrmund U, Groth-Malonek M, Knoop V (2008) Tracing Plant Mitochondrial DNA Evolution: Rearrangements of the Ancient Mitochondrial Gene Cluster *trn*A-*trn*T-*nad*7 in Liverwort Phylogeny. J Mol Evol 66:621-629

Wahrmund U, Quandt D, Knoop V (2009a) The phylogeny of mosses - addressing open issues with a new mitochondrial locus: group I intron cobi420. Mol Phylogenet Evol in press

Wahrmund U, Rein T, Müller KF, Groth-Malonek M, Knoop V (2009b) Fifty mosses on five trees: Comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. Plant Syst Evol 282:241-255

3.2 Mitochondrial molecular evolution in liverworts

3.2.1 Ute Wahrmund, Milena Groth-Malonek and Volker Knoop (2008): Tracing plant mitochondrial DNA evolution: Rearrangements of the ancient mitochondrial gene cluster *trn*A-*trn*T-*nad*7 in liverwort phylogeny. *Journal of Molecular Evolution* 66: 621-629

The *trn*A-*trn*T-*nad*7 gene arrangement is known to be conserved in the completely sequenced mitochondrial genomes of the alga *Chara vulgaris*, the moss *Physcomitrella patens* and the liverwort *Marchantia polymorpha*. While the size of the *trn*A-*nad*7 spacer region remained rather small in the alga and in the moss, it increased in *Marchantia* and the *trn*T is found in inverted direction. A survey across liverwort clades revealed frequent changes of spacer size and structure due to assumed recombinational events. In *Blasia*, the basal-most genus of the Marchantiopsida, the *trn*T is present in the same direction as in mosses and *Chara*. Sequence comparison unmasked the inversion of large part of the spacer region flanked by an imperfect inverted repeat motif. Only four other investigated marchantiid liverworts showed the same setting as *Marchantia*. All other analysed Marchantiopsida have lost the *trn*T gene and adjacent spacer regions. In Jungermanniopsida, a different part of the spacer including the *trn*T gene was independently deleted. Most of the spacer region of *Apotreubia*, a member of the basal-most liverwort clade, was not alignable with any other liverwort sequence determined during this study, again the *trn*T is lacking. Thus, this highly conserved region experienced frequent recombinational activity, i.e. an inversion and at least three independent losses of the *trn*T during liverwort evolution.

Contribution: I obtained half of the sequences, built the alignment, calculated the phylogenetic tree. I also discovered the patterns of frequent inversions, deletions and insertions and wrote part of the manuscript.

J Mol Evol (2008) 66:621–629
DOI 10.1007/s00239-008-9114-4

Tracing Plant Mitochondrial DNA Evolution: Rearrangements of the Ancient Mitochondrial Gene Cluster *trn*A-*trn*T-*nad*7 in Liverwort Phylogeny

Ute Wahrmund · Milena Groth-Malonek · Volker Knoop

Received: 18 December 2007 / Accepted: 22 April 2008 / Published online: 17 May 2008

Abstract Whereas frequent recombination characterizes flowering plant mitochondrial genomes, some mitochondrial gene arrangements may, in contrast, be conserved between streptophyte algae and early land plant clades (bryophytes). Here we explore the evolutionary fate of the mitochondrial gene arrangement *trn*A-*trn*T-*nad*7, which is conserved among the alga *Chara*, the moss *Physcomitrella*, and the liverwort *Marchantia*, although *trn*T is inverted in orientation in the latter. Surprisingly, we now find that the *Chara*-type gene arrangement is generally conserved in mosses, but that *trn*T is lacking between *trn*A and *nad*7 in all simple-thalloid and leafy (jungermanniid) liverworts. The ancient gene continuity *trn*A-*trn*T-*nad*7 is, however, conserved in *Blasia*, representing the sister lineage to all other complex-thalloid (marchantiid) liverworts. The recombinogenic insertion of short sequence stretches, including *nad*5 and *rps*7 pseudogene fragments copied from elsewhere in the liverwort mtDNA, likely mediated a subsequent inversion of *trn*T and flanking sequences in a basal grade of marchantiid liverworts, which was then followed by an independent secondary loss of *trn*T in derived marchantiid taxa later in evolution. In contrast to the previously observed extreme degree of coding sequence conservation and the assumed absence of active recombination in *Marchantia* mtDNA, this now reveals a surprisingly dynamic evolution of marchantiid liverwort mitochondrial genomes.

Keywords Liverworts · Phylogeny · Mitochondrial DNA · Intergenic region · Recombination

Abbreviations

bp	Base pair
mtDNA	Mitochondrial DNA

Introduction

Plant mitochondrial genomes are generally known for their highly conserved coding sequences but rapidly changing gene orders and coexisting genomic arrangements (Knoop 2004; Mackenzie et al. 1994; Mackenzie and McIntosh 1999; Ogihara et al. 2005; Sugiyama et al. 2005). Plant mitochondrial DNA (mtDNA) may even vary in structure between isolates of angiosperm species, for example, among the ecotypes of the model plant *Arabidopsis thaliana* (Ullrich et al. 1997) or among isolates of common beans or soybeans (Arrieta-Montiel et al. 2001; Moeykens et al. 1995). However, such an ongoing, highly frequent recombination of mtDNA may be an evolutionary gain only after the rise of vascular plants (tracheophytes). Groth-Malonek and colleagues have recently reported that an ancient gene cluster, the *nad*5-*nad*4-*nad*2 gene arrangement, found in the alga *Chara*, is universally conserved among liverworts, mosses, and hornworts (Groth-Malonek et al. 2007a). While the *nad*5-*nad*4 intergenic region showed size increases to varying degrees in the three bryophyte divisions (e.g., ~500 base pairs [bp] in mosses and up to more than ~3,000 bp in hornworts), the *nad*4-*nad*2 spacer was strikingly conserved with its tiny size of only 26 bp across all bryophyte clades.

U. Wahrmund · M. Groth-Malonek · V. Knoop (✉)
IZMB—Institut für Zelluläre und Molekulare Botanik, Abt. Molekulare Evolution, Universität Bonn, Kirschallee 1, 53115 Bonn, Germany
e-mail: volker.knoop@uni-bonn.de

U. Wahrmund
e-mail: ute.wahrmund@uni-bonn.de

M. Groth-Malonek
e-mail: groth-malonek@uni-bonn.de

We have investigated another gene cluster that was found conserved when the completely sequenced chondriomes of the alga *Chara vulgaris* (Turmel et al. 2003), the liverwort *Marchantia polymorpha* (Oda et al. 1992b), and the moss *Physcomitrella patens* (Terasawa et al. 2006) were compared: the *trn*A-*trn*T-*nad*7 region. Besides the significantly larger size of the intergenic region between *trn*A and *nad*7 in *Marchantia* (1866 bp, compared to 528 bp in *Physcomitrella* and 124 bp in *Chara*), the *trn*T gene is present in inverted orientation in the liverwort. Moreover, the functional *nad*7 gene resides in the *Marchantia* nuclear genome (Kobayashi et al. 1997), while six stop codons render the mitochondrial copy a transcribed pseudogene (Takemura et al. 1995). Groth-Malonek and colleagues have recently shown that *nad*7 is retained as a pseudogene in all liverwort lineages but is an apparently intact and expressed gene in *Haplomitrium* (Haplomitriopsida), which is a representative of the sister lineage (Haplomitriopsida/Treubiopsida) to all other liverworts (Forrest et al. 2006; Groth-Malonek et al. 2007b).

The conservation of *nad*7 as a pseudogene and the enigmatic upstream inversion of *trn*T in *Marchantia* relative to the alga and the moss prompted us to investigate the molecular evolution of this region in related taxa. Whereas the *trn*A-*trn*T-*nad*7 region is conserved in this order and direction of transcription in diverse mosses, we find that it has experienced major changes among the liverworts, which include insertions of pseudogene fragments and noncoding sequence copies, subsequent inversion of *trn*T among marchantiid liverworts, and, finally, independent losses of *trn*T in the two major liverwort clades. These events of molecular evolution can be conveniently mapped onto the consensus phylogeny of liverworts as the sister clade to all other land plants (embryophytes) which is now clearly emerging after a series of recent molecular phylogenetic studies (Crandall-Stotler et al. 2005; Davis 2004; Forrest et al. 2006; Forrest and Crandall-Stotler 2004, 2005; Heinrichs et al. 2005; Qiu et al. 2006). Our new findings document significant genomic plasticity in the mtDNA of marchantiid liverworts and, as such, stand in contrast to the hitherto observed strong conservation of mitochondrial coding sequences (Beckert et al. 1999; Dombrovska and Qiu 2004; Forrest et al. 2006; Forrest and Crandall-Stotler 2005; Pruchner et al. 2002), the apparent and exclusive absence of RNA editing in this early land plant clade (Steinhauser et al. 1999), and, most importantly, the absence of active, ongoing genomic mtDNA recombination in *Marchantia* (Oda et al. 1992a, b; Oldenburg and Bendich 1998).

Materials and Methods

Total nucleic acids were extracted using the CTAB (cetyltrimethyl-ammonium bromide) method or a plant DNA extraction kit (Macherey-Nagel, Düren, Germany). To PCR amplify the *trn*A-*nad*7 region, the primers trnAfor (5′-tcggttcaavtccgatcgtctcca-3′) and nad7back (5′-accatgagcagc wggrtgttgagg-3′) were used. PCRs usually contained 2.5 µl 10× PCR buffer, a 250 µM concentration of each dNTP, a 1 µM concentration of each primer, 1 U of DNA polymerase, 1 µl of DNA, and double-distilled water added up to 25 µl. Either Genaxxon *Taq* DNA polymerase S (Biberach, Germany) or the Triple Master PCR System (Eppendorf, Hamburg, Germany) and the respective buffers supplied by the manufacturers were used. Typical amplifications were performed with 95°C for 3 min as the denaturation step, followed by 35 cycles at 95°C for 30 s, 50°C for 30 s, 72°C for 3 min, and a final elongation step at 72°C for 7 min. Gel-purified PCR fragments were cloned into pGEM-Teasy (Promega, Mannheim, Germany) and sequenced commercially by Macrogen, Inc. (Seoul, Korea). Nucleotide sequences were edited and aligned manually using MEGA 4.0 (Kumar et al. 2004). Phylogenetic trees were calculated either with the neighbor-joining algorithm (Tamura three-parameter, pairwise deletion, and bootstrap with 10,000 replicates) with MEGA or by Bayesian phylogenetic analyses using MrBayes v3.1.2win (Ronquist and Huelsenbeck 2003) with the GTR + G + I model, partitions unlinked, for 10 million generations, with every 100th tree sampled (burn-in = 1000). Burn-in was determined as stationarity in the log likelihood plots based on the summarizing parameters of the MrBayes output. Folding of tRNAs was done with RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi) and subsequent manual editing. The BLAST search against the complete mitochondrial genome of *Marchantia polymorpha* (NC_001660 [Oda et al. 1992]) was conducted using the Local BLAST tool implemented in BioEdit v7.0.5.3 (Hall 1999).

Results

We first wished to check whether the conservation of gene orders *trn*A-*trn*T-*nad*7 in this direction of transcription in the alga *Chara vulgaris* and the model moss *Physcomitrella patens* was just a coincidental exception (Fig. 1). To this end we designed primers in *trn*A and *nad*7 (Fig. 1) and used them in PCR amplifications over a taxonomically wide sampling of moss DNAs (Table 1). The PCR products we retrieved successfully showed only a moderate variation in size (not shown), mostly similar to what was to be expected for *Physcomitrella* (534 bp), with only exceptional size reductions in *Mnium, Pohlia, Takakia*, and *Leskea* (Table 1). PCR amplification products were cloned and their nature was verified by sequencing. In all cases, *trn*T was found to be present in the same direction of transcription as in *Chara* and *Physcomitrella*, initially

J Mol Evol (2008) 66:621–629

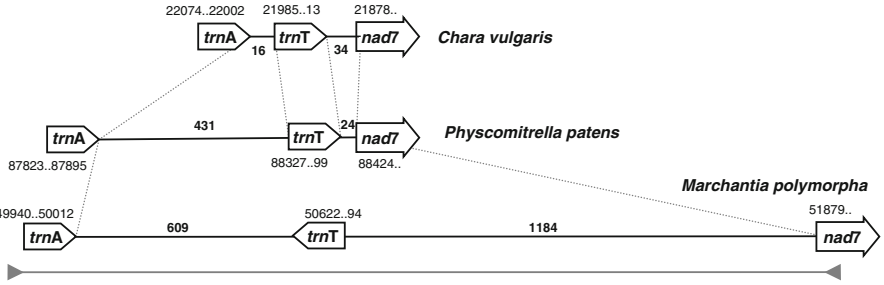


Fig. 1 Conservation of the mitochondrial *trn*A-*trn*T-*nad*7 region in the alga *Chara vulgaris*, the moss *Physcomitrella patens*, and the liverwort *Marchantia polymorpha* (with *trn*T in inverted orientation) allowed for the design of oligonucleotide primers (*arrowheads*) bordering a PCR amplicon (*bottom*) to probe conservation of the gene arrangements in other bryophyte taxa via PCR amplification. The indicated sequence numbering follows the complete mtDNA database entries for *Chara vulgaris* (NC_005255), *Physcomitrella patens* (AB251495), and *Marchantia polymorpha* (NC_001660), respectively. Drawing is not to scale

suggesting that this gene arrangement is likely conserved among early land plants as similarly observed before for the mitochondrial *nad*5-*nad*4-*nad*2 gene continuity (Groth-Malonek et al. 2007a).

Subsequently, we amplified the *trn*A-*nad*7 region for 11 jungermanniid and 15 marchantiid liverworts. In contrast to mosses, the analysis of PCR products immediately suggested significantly greater size variation in liverworts, ranging from ~700 bp in some marchantiid taxa, to ~900 bp for most jungermanniid taxa, up to nearly 2,000 bp in other marchantiids, a size fitting the expectation for *Marchantia* (Fig. 2). The nature of the amplification products was again verified through cloning and sequencing, which revealed intergenic distances between *trn*A and *nad*7 ranging from 536 bp in *Porella* up to 1,868 bp in *Lunularia* (Table 1). Among the jungermanniid liverworts the *trn*A-*nad*7 intergenic region reached up to 903 bp in *Noteroclada* (Table 1), but none of the *trn*A-*nad*7 intergenic regions in 11 jungermanniid species contained a *trn*T gene (Fig. 3).

Among the marchantiid species significantly different results were obtained. In *Blasia*, the presumed sister genus to a clade of all other marchantiid taxa (Forrest et al. 2006), a *trn*T gene was identified in the same orientation as *trn*A and *nad*7, i.e., as in the mosses and in the alga *Chara* (Fig. 3). The *trn*A-*trn*T spacer had a length of 926 bp, the *trn*T-*nad*7 spacer of 850 bp (Table 1), thus revealing a significant size increase of the intergenic regions compared to *Chara* and the mosses.

In a further four of the marchantiid taxa investigated here (*Bucegia, Lunularia, Riella*, and *Sphaerocarpos*), *trn*T was identified between *trn*A and *nad*7, but in all these cases its coding sequence was inverted compared to *Blasia* (Table 1), thus reflecting the *Marchantia* situation (Fig. 3; Marchantiopsida 2). Similar to all jungermanniid taxa, *trn*T and surrounding sequences are lacking altogether from the other 10 marchantiid species in our taxon sampling (Table 1), obviously due to a major deletion in the *trn*A-*nad*7 intergenic region (Fig. 3; Marchantiopsida 3). Finally, we strived to include a representative taxon from the Haplomitriopsida as well, given that this class is now well supported as an ancient lineage of liverworts, sister to the dichotomy of the marchantiid and jungermanniid clades. We were able to obtain a PCR product for the *trn*A-*nad*7 region from *Apotreubia nana* as a representative taxon for this group. Upon cloning and sequencing we found that the large intergenic spacer in *Apotreubia* carries a unique sequence insertion without similarity to any other sequences in the database but that *trn*T is absent between *trn*A and *nad*7 (Fig. 3).

Detailed sequence analyses revealed that several stretches in the liverwort *trn*A-*nad*7 spacers share significant similarities with other regions of the *Marchantia* mtDNA. Most notable is a highly conserved sequence stretch of ~100 bp copied from the central coding region of the *rps*7 gene encoding protein 7 of the small ribosomal subunit (with an internal deletion of about 90 bp) located 15 bp upstream of *nad*7 in all liverworts (Fig. 3). Interestingly, another intergenic region in the *Marchantia* mtDNA (*rps*1-*nad*4L) carries a corresponding *rps*7 pseudogene fragment completely including this *rps*7 homology, but without the internal deletion observed upstream of *nad*7. A yet larger pseudogene fragment of ~200 bp derived from the 5′ part of the mitochondrial *nad*5 gene is inserted upstream of *trn*T in *Blasia* and took part in the major sequence inversion including *trn*T observed in some marchantiid species (Fig. 3; Marchantiopsida 2). Other regions carry sequence elements of 20–70 bp repeated elsewhere in noncoding regions (introns and spacers) of the *Marchantia* mtDNA, suggesting ancient recombination events on evolutionary timescales. One such sequence element repeated upstream of tRNA genes *trn*E and *trn*Q in the *Marchantia* mtDNA (REPeq in Fig. 3) also participated in the major sequence inversion in Marchantiopsida 2.

The large sequence inversion in Marchantiopsida 2 is precisely bordered by a perfect 21-bp inverted repeat motif in

Table 1 List of sequence accessions for the taxa under investigation

Taxon	*trn*A-*nad7*	*nad5*	*rbc*L	*rps*4	*trn*T orientation	Spacer size, *trn*A-*trn*T	Spacer size, *trn*T-*nad7*
Mosses							
Physcomitrella patens	AB251495	AB251495	NC_005087	NC_005087	Forward	431	24
Dawsonia sp.	EU519151	AY908804	AY118237	AY908016	Forward	490	24
Ditrichum cylindricum	EU519152	AJ291559	AF231080	AJ554009	Forward	436	25
Homalia trichomanoides	EU519153	AJ291569	EU095325	AY908276	Forward	431	24
Hygrohypnum ochraceum	EU519154	AJ291574	EU095323	AY908620	Forward	431	24
Leskea polycarpa	EU519155	AJ291576	EU519193	AF143042	Forward	194	24
Mnium hornum	EU519156	AJ291567	AF226820	AF023796	Forward	288	24
Pogonatum urnigerum	EU519157	AJ291554	DQ120779	AF208426	Forward	486	24
Pohlia nutans	EU519158	AJ291565	AJ275175	AY907983	Forward	288	24
Takakia lepidozioides	EU519159	AJ291553	AY312936	AF306950	Forward	294	23
Liverworts							
Marchantiopsida							
Asterella blumeana	EU519160	DQ268909	U87064	DQ268983	—	654	
Blasia pusilla	EU519161	EU519187	AF536232	AY507436	Forward	926	850
Bucegia romanica	EU519162	AJ001031	EU519194	—	Reverse	601	1183
Conocephalum conicum	EU519163	EU51918	AY688778	DQ220678	—	662	
Lunularia cruciata	EU519164	AJ001002	U87077	AY688795	Reverse	607	1187
Marchantia polymorpha	NC001660	NC001660	X04465	X04465	Reverse	609	1184
Monoclea gottschei	EU519165	AJ622814	AY507414	AY507455	—	655	
Monosolenium tenerum	EU519166	DQ268944	DQ286017	DQ220691	—	659	
Oxymitra incrassata	EU519167	EU519190	EU519195	EU519197	—	665	
Reboulia hemisphaerica	EU519168	EU519189	AY462326	AY688801	—	656	
Riccia breidleri	EU519169	DQ268957	AY507422	*AY507463*	—	658	
Riccia fluitans	EU519170	DQ268956	DQ286023	AY608107	—	664	
Ricciocarpos natans	EU519171	AJ001032	U87089	AJ251062	—	661	
Riella sp.	EU519172	DQ268959	DQ645959	EU519196	Reverse	605	1188
Sphaerocarpos donnellii	EU519173	AJ001033	AY507425	AY608110	Reverse	609	1179
Targionia hypophylla	EU519174	AJ001001	AY507427	AY688805	—	664	
Jungermanniopsida							
Anthelia julacea	EU519175	—	DQ026581	AY608044	—	852	
Blepharostoma trichophyllum	EU519176	—	DQ645964	AY462343	—	871	
Calypogeia muelleriana	EU519177	EU519191	U87065	AY608052	—	827	
Fossombronia pusilla	EU519178	AJ000699	AF536231	AY608062	—	827	
Herbertus sendtneri	EU519179	DQ268927	AY507404	AY462353	—	825	
Lophocolea cuspidata	EU519180	DQ268932	AY149845	AF231889	—	870	
Metzgeria furcata	EU519181	AJ000703	AY507411	AY507453	—	846	
Noteroclada confluens	EU519182	AJ622816	AY688784	AY688797	—	903	
Plagiochila asplenioides	EU519183	AJ000704	AY699996	AY547693	—	828	
Porella platyphylla	EU519184	AY688767	AY507420	AY462387	—	536	
Trichocolea tomentella	EU519185	AJ000707	AY608040	AY608118	—	828	
Haplomitriopsida							
Apotreubia nana	EU519186	EU519192	AY877389	DQ268983	—	2106	
Alga							
Chara vulgaris	NC_005255	NC_005255	NC_008097	NC_008097	Forward	16	34

Note: Sizes of the mitochondrial intergenic regions and the direction of transcription of *trn*T, when present, are indicated. The respective PCR amplicons additionally included 45 bp of the *nad7* amino terminus (excluding primers). All sequence accessions of the mitochondrial *trn*A-*nad7* regions were determined in the course of the present study (36: EU519151–EU519186) except for those of *Chara vulgaris*, *Marchantia polymorpha*, and *Physcomitrella patens*, which were previously available in the database. Furthermore, a total of 11 novel sequences (EU519187–EU519197) were obtained to fill in holes in the three-gene matrix used to derive the phylogeny shown in Fig. 4. Available sequence accessions from different species were used in the three-gene matrix to determine the phylogeny shown in Fig. 4 for some genera (*Asterella, Riccia, Riella, Shaerocarpos, Herbertus, Lophocolea, Porella*, and *Trichocolea*)

J Mol Evol (2008) 66:621–629

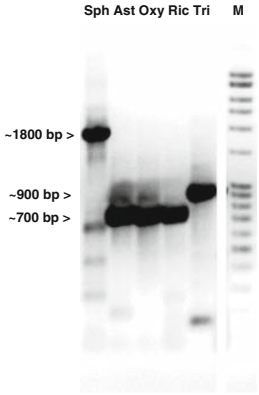


Fig. 2 Exemplary PCR amplification assays of the *trn*A-*nad*7 spacer region with DNA from *Sphaerocarpos* (Sph), *Asterella* (Ast), *Oxymitra* (Oxy), *Riccia* (Ric), and *Trichocolea* (Tri). The identities of the respective major amplicon products (sizes indicated) were subsequently verified through cloning and sequencing. Occasional, minor accompanying bands were gel-excised, cloned, and sequenced as well, but turned out to be products of nonspecific mispriming in all cases

Blasia. Other sequence stretches composed of homologous sequence stretches repeated multiple times in noncoding regions of the *Marchantia* mtDNA (boxes labeled REPm in Fig. 3) may have functionally contributed to this recombination event given their locations flanking the large sequence inversion, possibly by creating the necessary homologous stretches as substrates for recombination.

Interestingly, the large sequence insertion found in *Apotreubia* is flanked by a 22-bp motif (Fig. 3), in this case present as an imperfect direct sequence repeat. Similarly extended sequence repeats are not identified as flanking the large deletion event in Marchantiopsida 3, which completely removed most parts of the inverted region, resulting simply in a short run of guanidine nucleotides (Fig. 3).

The *trn*A-*nad*7 regions of all marchantiid liverworts except *Blasia* can easily be aligned, leaving a large gap of more than 1200 bp in the taxa without *trn*T. Likewise, most of the spacer sequence in the jungermanniid liverworts can be aligned with the marchantiid species except for a unique region of about 130 bp that is not found in the latter.

A reasonably well-resolved liverwort phylogeny on the basis of molecular data is available (e.g., Forrest et al. 2006), which can now be used to trace the series of molecular rearrangements in the evolution of the *trn*A-*nad*7 region in this early land plant clade. We have here used the available sequence data for the mitochondrial

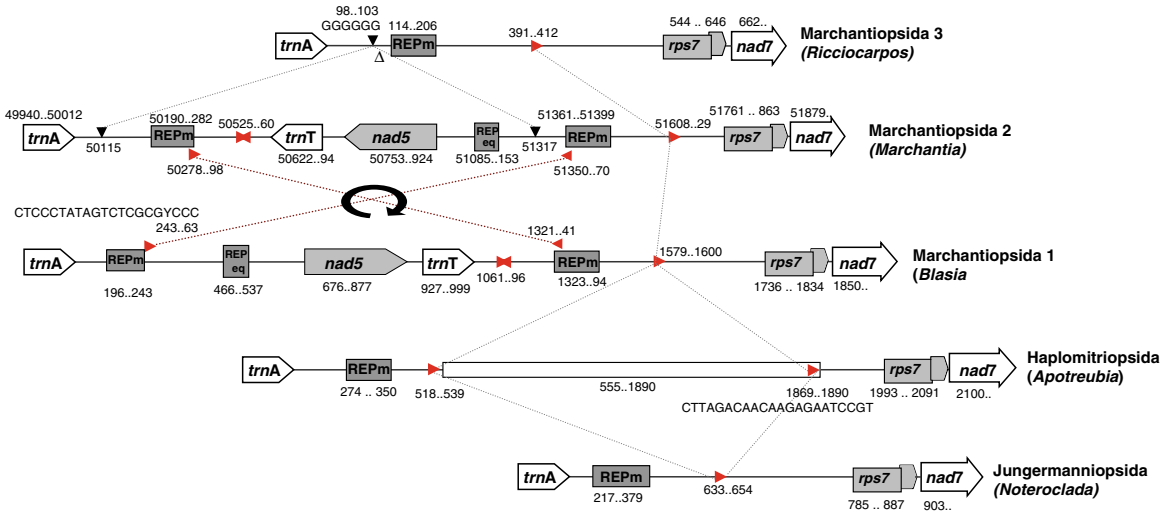


Fig. 3 The order of *trn*A, *trn*T, and *nad*7 in this direction of transcription in *Chara* and *Physcomitrella* turned out to be conserved in all mosses now investigated (not shown) but only in *Blasia* among the liverworts (Marchantiopsida 1). Sequence numbering starts with the first nucleotide following *trn*A for the selected exemplar taxa as indicated except for *Marchantia*, which is as in Fig. 1. The major part of the *trn*A-*nad*7 intergenic region including *trn*T is inversed in orientation in one group of marchantiid liverworts (Marchantiopsida 2). The gene for *trn*T is absent between *trn*A and *nad*7 in all Jungermanniopsida investigated, in a subclade of marchantiid liverworts (Marchantiopsida 3), and in *Apotreubia* (Haplomitriopsida). Two larger sequence stretches (*light gray*) are pseudogene fragment copies of *nad*5 and *rps*7. Several other regions (dark-gray boxes labeled REPm) are composed of sequence fragments with significant similarity to other noncoding regions repeated numerous times elsewhere in the *Marchantia* mtDNA, indicating recombination events (e.g., in introns cobi372, cox1i511, and nad4Li100 and in intergenic regions *atp*9-*trn*C, *rps*11-*rps*1, *trn*S-*trn*L, *nad*3-*trn*V, and *cob*-*nad*9, respectively). REPeq indicates a sequence motif repeated upstream of *trn*E and of *trn*Q, respectively. A large sequence insertion in *Apotreubia* (*white box*) without significant similarity to any sequence currently in the database and the inverted sequence in Marchantiopsida are flanked by imperfect direct or inverted sequence repeats (*arrowhead*), respectively, as indicated. The double-arrowhead indicates a 36-bp imperfect palindrome sequence downstream of *trn*T in Marchantiopsida (AAAGCRAGTG TTTTTTTMKAAAAAARCACTYGCTTT). Drawing is not to scale

*nad*5 gene and the chloroplast *rbc*L and *rps*4 genes (with some gaps filled in the course of this study) to reconstruct the phylogeny for a taxonomically congruent data set (Fig. 4). Most likely, the ancient *trn*A-*trn*T-*nad*7 arrangement existed when liverworts emerged but has only survived in *Blasia*, the sister lineage to a clade of all other marchantiid liverworts. The loss of *trn*T does not characterize a monophyletic group but has likely occurred three times independently: in the lineage of jungermanniid liverworts, in a monophyletic clade of derived marchantiids, and in the *Apotreubia* lineage (Fig. 4). The five taxa of marchantiid liverworts with an inverted *trn*T represent a basal, paraphyletic grade.

To explore the potential utility of the now investigated *trn*A-*nad*7 intergenic region as a candidate phylogenetically informative locus, we constructed phylogenetic trees based on the spacer sequence data set. In an approach using the complete alignment encompassing all taxa, the five basal marchantiids with inverted *trn*T were, somewhat expectedly, artificially retrieved as monophyletic due to the large number of shared characters in the inverted sequence region (not shown). On the other hand, well-resolved phylogenetic trees were retrieved for the taxonomic subsets with the same arrangements of the intergenic region, i.e., the jungermanniid taxa (Fig. 5A), and the marchantiid taxa with the large inverted *trn*T region (Fig. 5B). Phylogenetic resolution, however, was significantly lower for the derived marchantiid taxa (Fig. 5C), featuring a much smaller intergenic region lacking the (inverted) *trn*T sequence stretch, in line with the extreme primary sequence conservation that had been observed for marchantiids before. Compared to the now reasonably well-supported liverwort phylogenies (or see Fig. 4), it can be stated that despite the small taxon sampling, the *trn*A-*nad*7 region may be a welcome addition in phylogenetic analyses for the remaining unanswered questions among jungermanniid or basal marchantiid taxa.

The secondary structure of *trn*T(ggu), which was found to be subject to genomic recombination, remains largely unaffected in the majority of taxa (Fig. 6). An extra 3-bp hairpin arm between the anticodon and the pseudouridine arm is conserved in all species, and only minor nucleotide exchanges of nonpaired regions were observed in positions

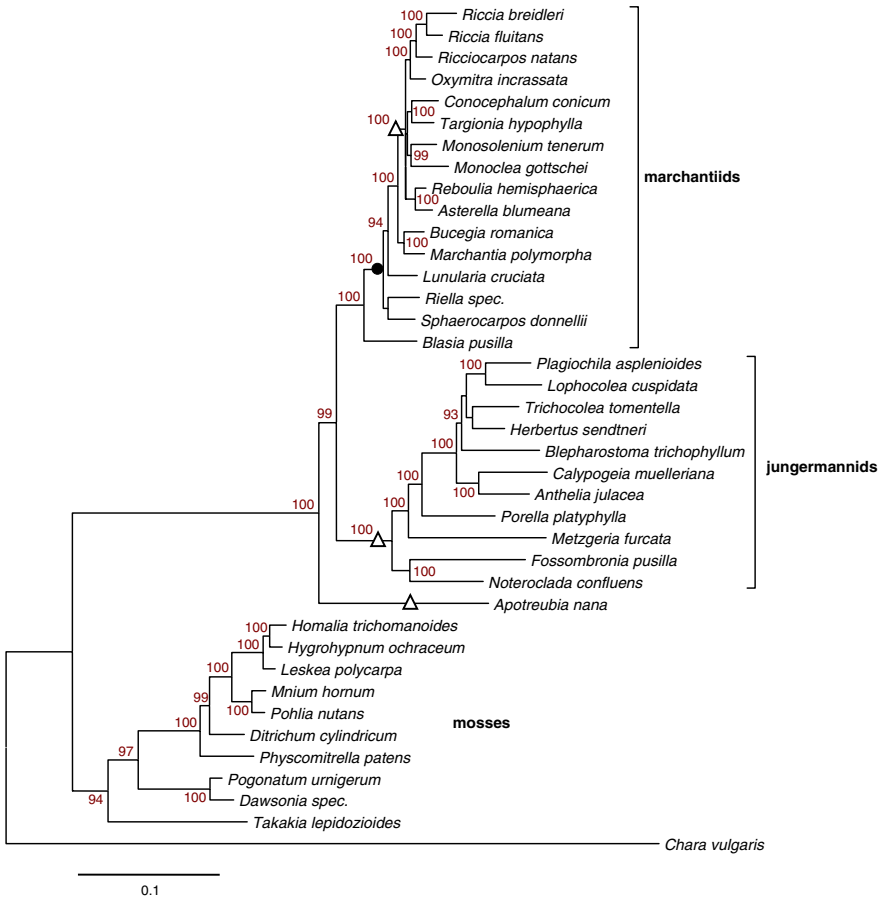


Fig. 4 A Bayesian phylogenetic tree of the liverworts under investigation based on fused *nad*5, *rbc*L, and *rps*4 gene data sets (GTR + G + I substitution model, two parallel runs for 10 million generations each, every 100th tree sampled, 1,000 trees discarded as burn-in). Posterior probabilities are indicated as percentage node supports in the phylogram where at least 90. Branch lengths are means of the branch length posterior probability distribution of all sampled trees. The filled circle indicates inversion of a major part of the *trn*A-*nad*7 spacer including *trn*T in the marchantiid liverworts. Open triangles indicate independent secondary losses of *trn*T and flanking sequences in *Apotreubia*, in the jungermanniid, and in the derived marchantiid liverworts, respectively

J Mol Evol (2008) 66:621–629

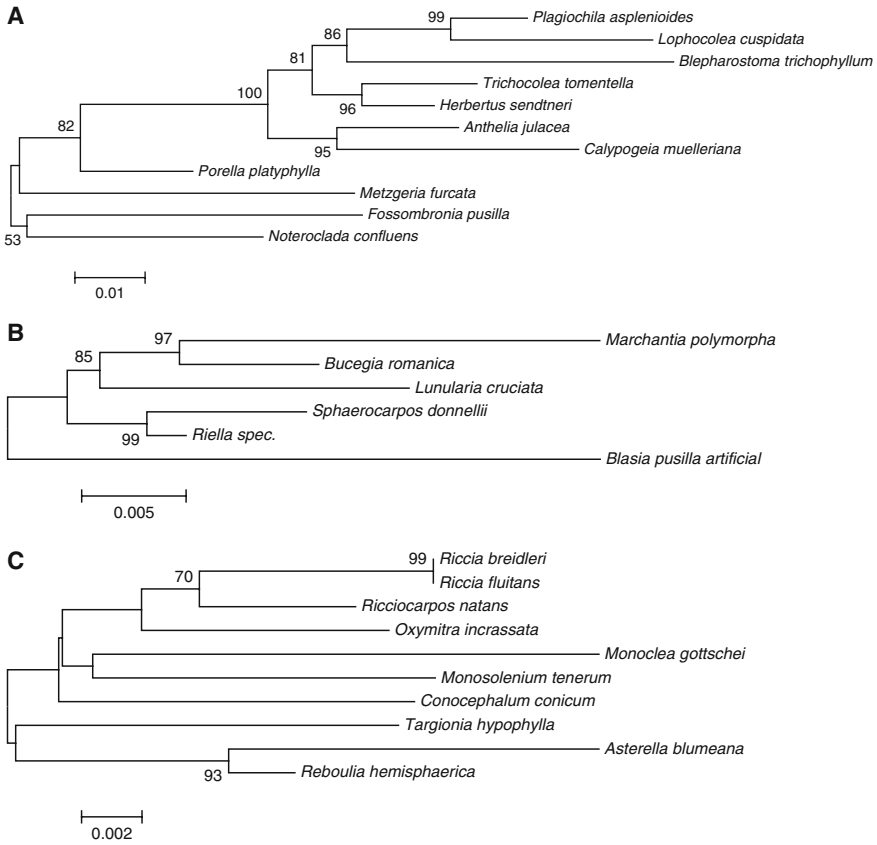


Fig. 5 Phylogenetic trees based on alignments of the *trn*A-*nad*7 regions of jungermanniid liverworts (**A**), of basal marchantiid liverworts with inverted *trn*T (**B**), and derived marchantiid liverworts without *trn*T in the intergenic region (**C**). *Blasia* was included in the alignment to derive the phylogeny under B with an artificially inverted *nad*5-*trn*T fragment. Shown are neighbor-joining trees using Tamura three-parameter corrected distances (pairwise deletion) with bootstrap support from 1,000 replicates as a conservative measure of node reliability

15, 16, 26, and 32 in some mosses and *Chara*, and base-paired positions 47 and 66 are alternatively part of G-C or G-U base pairs, respectively (Fig. 6). Somewhat more conspicuous is the U-to-C transition in position 4 in *Chara vulgaris* (shared by *Takakia lepidozioides* and all liverworts), leading to a C-A mismatch in the acceptor stem, and the U-to-G transversion in the acceptor stem of *Sphaerocarpos*, which potentially impede the function of *trn*T to functionally decode ACY threonine codons.

Discussion

Liverworts are now supported unequivocally as a monophyletic clade by molecular data and, as such, are also reasonably well supported as the sister group to all other land plants (Qiu et al. 1998, 2006). *Marchantia polymorpha* is widely considered as the prototype liverwort. However, it should be noted that as a complex-thalloid species, it represents only one of several major clades of liverworts. With the present study it becomes clear that *Marchantia* may even represent more of an exception than a rule in mtDNA organization, given that *trn*T and most of the *trn*A-*nad*7 intergenic spacer are lacking in most marchantiid and all jungermanniid liverworts as well as in *Apotreubia* representing the Haplomitriopsida.

This is the first report of significant divergence in mtDNA organization within the liverwort clade. Here we have traced the recombination events in the *trn*A-*nad*7 region, which now allows us to plot parsimoniously a series of events in its molecular evolution onto a phylogeny of this early land plant clade. Notably, in contrast to the frequently recombining mtDNAs of flowering plants, the mtDNA of *Marchantia polymorpha* had been not only mapped as but also physically identified as a single circular molecule (Oda et al. 1992a). No subgenomic circles typical for angiosperm mtDNAs have been identified, however, concatemers of the circular genome and linear forms, possibly representing replication intermediates, seem to be present in significant amounts in *Marchantia* mtDNA preparations (Oldenburg and Bendich 1998, 2001). The now identified series of recombination events in the marchantiid liverwort lineage clearly documents ancient recombinational processes that were active on evolutionary timescales. The extensive rearrangements in the intergenic region studied here starkly contrast with the striking

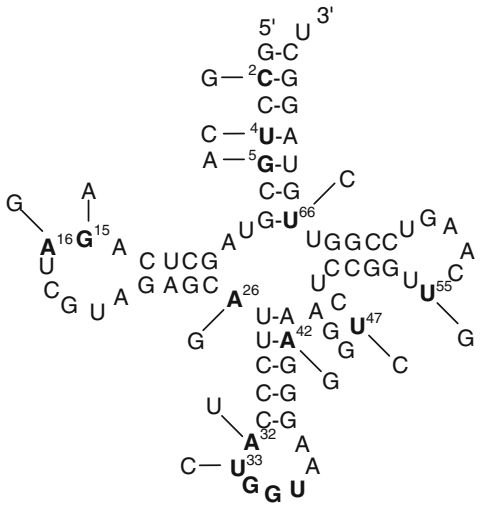


Fig. 6 Cloverleaf secondary structure of *trn*T(GGU) obtained using the RNAfold WWW service (http://rna.tbi.univie.ac.at/cgi-bin/RNAfold.cgi), with nucleotide positions showing exchanges in certain taxa highlighted in boldface. 2-*Sphaerocarpos*—G; 4-*Chara*, *Takakia*, and all liverworts—C; 5-*Chara*—A; 15-*Dawsonia* and *Pogonatum*—A; 16-*Chara*—G; 26-*Takakia*—G; 32-*Chara*—T; 33-*Blasia*—C; 42-*Dawsonia*—G; 47-*Marchantia, Sphaerocarpos, Riella, Lunularia*, and *Blasia*—C; 55-*Sphaerocarpos* and *Riella*—G; 66-*Dawsonia*, *Pogonatum*, *Takakia*, and liverworts—C

conservation, even across all three bryophyte divisions, of two other intergenic regions, *nad*5-*nad*4 and *nad*4-*nad*2 (Groth-Malonek et al. 2007a).

When rearrangements of organellar genomes occur as rare, one-time events in evolutionary history, they are of particular use for phylogenetic studies by defining monophyletic clades. A chloroplast DNA inversion, which clearly sets lycopodiophytes, but not the whisk fern *Psilotum*, apart from other tracheophytes (Raubeson and Jansen 1992b), is a typical example of an early phylogenetic insight now well documented through molecular studies confirming the identification of euphyllophytes as a monophyletic group (Pryer et al. 2001). Along similar lines a 22-kb inversion of cpDNA in Asteraceae except Barnadesioidae and the absence of an inverted repeat sequence copy as a synapomorphy of conifers are other examples (Jansen and Palmer 1987; Raubeson and Jansen 1992a). Similarly, the loss of *rpo*A in arthrodontous mosses (Goffinet et al. 2005) or the large inversion of 71 kb in the plastome in funariid mosses except the Gigaspermaceae described more recently (Goffinet et al. 2007) are significant phylogenetically informative events. A similar case in point is the degeneration of *nad*7 into a pseudogene in all liverworts except the Haplomitriopsida (Groth-Malonek et al. 2007b), which independently confirms a placement of the latter as sister to all other liverworts. In a similar manner, the new data presented here now independently confirm the unequivocal placement of *Blasia* among, and basal to all other, marchantiid liverworts. Rare molecular apomorphies like these identified on genomic scales are welcome support for phylogenies including such isolated basal lineages which may be subject to artificial long-branch attraction in sequence-based molecular phylogenies.

Moreover, this study has shown that the loss of *trn*T from the *trn*A-*nad*7 region has occurred several times independently—once in the jungermanniid lineage, once (after a major sequence inversion) in a lineage of derived marchantiid liverworts (and in *Apotreubia*)—and those events are in full accord with current insights on liverwort phylogeny. The now investigated mtDNA region holds promise as a phylogenetically informative locus, largely for the jungermanniid liverworts (due to faster sequence evolution) and possibly also for the basal marchantiid taxa (mainly due to the longer intergenic region). In contrast, it cannot offer phylogenetic resolution for the derived marchantiids given the large deletion within the intergenic region and in the light of the high degree of sequence conservation.

Acknowledgments We are very grateful for financial support through the Deutsche Forschungsgemeinschaft DFG and to colleagues Jochen Heinrichs (Göttingen, Germany), Jan-Peter Frahm (Bonn, Germany), Hermann Muhle (Ulm, Germany), Yin-Long Qiu (Ann Arbor, MI, USA), Barbara Crandall-Stotler (Carbondale, IL, USA), and Laura Lowe-Forrest (Edinburgh, UK) for help in obtaining certain plant materials and/or DNA samples, respectively. U.W. would like to thank Michael Volkmar for constant support and ad hoc advice.

References

Arrieta-Montiel M, Lyznik A, Woloszynska M, Janska H, Tohme J, Mackenzie S (2001) Tracing evolutionary and developmental implications of mitochondrial stoichiometric shifting in the common bean. Genetics 158:851–864

Beckert S, Steinhauser S, Muhle H, Knoop V (1999) A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad5* gene. Plant Syst Evol 218:179–192

Crandall-Stotler B, Forrest LL, Stotler RE (2005) Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). Taxon 54:299–316

Davis EC (2004) A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). In: Goffinet B, Hollowell V, Magill R (eds) Molecular systematics of bryophytes. Missouri Botanical Garden Press, St. Louis, pp 61–86

Dombrovska E, Qiu YL (2004) Distribution of introns in the mitochondrial gene *nad*1 in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol 32:246–263

Forrest LL, Crandall-Stotler BJ (2004) A phylogeny of the simple thalloid liverworts (Jungermanniopsida, Metzgeriidae) as inferred from five chloroplast genes. In: Goffinet B, Hollowell V, Magill R (eds) Molecular systematics of bryophytes. Missouri Botanical Garden Press, St. Louis, pp 119–140

Forrest LL, Crandall-Stotler BJ (2005) Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. J Hattori Bot Lab 97:127–159

Forrest LL, Davis EC, Long DG, Crandall-Stotler BJ, Clark A, Hollingsworth ML (2006) Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. Bryologist 109:303–334

Goffinet B, Wickett NJ, Shaw AJ, Cox CJ (2005) Phylogenetic significance of the rpoA loss in the chloroplast genome of mosses. Taxon 54:353–360

Goffinet B, Wickett NJ, Werner O, Ros RM, Shaw AJ, Cox CJ (2007) Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). Ann Bot (Lond) 99:747–753

Groth-Malonek M, Rein T, Wilson R, Groth H, Heinrichs J, Knoop V (2007a) Different fates of two mitochondrial gene spacers in early land plant evolution. Int J Plant Sci 168:709–717

Groth-Malonek M, Wahrmund U, Polsakiewicz M, Knoop V (2007b) Evolution of a pseudogene: exclusive survival of a functional mitochondrial nad7 gene supports Haplomitrium as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. Mol Biol Evol 24:1068–1074

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Heinrichs J, Gradstein SR, Wilson R, Schneider H (2005) Towards a natural classification of liverworts (Marchantiophyta) based on the chloroplast gene rbcL. Crypto Bryol 26:215–233

Jansen RK, Palmer JD (1987) A chloroplast DNA inversion marks an ancient evolutionary split in the sunflower family (Asteraceae). Proc Natl Acad Sci USA 84:5818–5822

Knoop V (2004) The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet 46:123–139

Kobayashi Y, Knoop V, Fukuzawa H, Brennicke A, Ohyama K (1997) Interorganellar gene transfer in bryophytes: the functional nad7 gene is nuclear encoded in Marchantia polymorpha. Mol Gen Genet 256:589–592

Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinf 5:150–163

Mackenzie S, McIntosh L (1999) Higher plant mitochondria. Plant Cell 11:571–585

Mackenzie S, He S, Lyznik A (1994) The elusive plant mitochondrion as a genetic system. Plant Physiol 105:775–780

Moeykens CA, Mackenzie SA, Shoemaker RC (1995) Mitochondrial genome diversity in soybean—repeats and rearrangements. Plant Mol Biol 29:245–254

Oda K, Kohchi T, Ohyama K (1992a) Mitochondrial DNA of Marchantia polymorpha as a single circular form with no incorporation of foreign DNA. Biosci Biotechnol Biochem 56:132–135

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, Ohyama K (1992b) Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol 223:1–7

Ogihara Y, Yamazaki Y, Murai K, Kanno A, Terachi T, Shiina T, Miyashita N, Nasuda S, Nakamura C, Mori N, Takumi S, Murata M, Futo S, Tsunewaki K (2005) Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Res 33:6235–6250

Oldenburg DJ, Bendich AJ (1998) The structure of mitochondrial DNA from the liverwort, Marchantia polymorpha. J Mol Biol 276:745–758

Oldenburg DJ, Bendich AJ (2001) Mitochondrial DNA from the liverwort Marchantia polymorpha: circularly permuted linear molecules, head-to-tail concatemers, and a 5′ protein. J Mol Biol 310:549–562

Pruchner D, Beckert S, Muhle H, Knoop V (2002) Divergent intron conservation in the mitochondrial *nad2* gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. J Mol Evol 55:265–271

Pryer KM, Schneider H, Smith AR, Cranfill R, Wolf PG, Hunt JS, Sipes SD (2001) Horsetails and ferns are a monophyletic group and the closest living relatives to seed plants. Nature 409:618–622

Qiu YL, Cho YR, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674

Qiu YL, Li L, Wang B, Chen Z, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, Davis CC (2006) The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA 103:15511–15516

Raubeson LA, Jansen RK (1992a) A rare chloroplast DNA structural mutation is shared by all conifers. Biochem Syst Ecol 20:17–24

Raubeson LA, Jansen RK (1992b) Chloroplast DNA evidence on the ancient evolutionary split in vascular land plants. Science 255:1697–1699

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V (1999) Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? J Mol Evol 48:303–312

Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genom 272:603–615

Takemura M, Nozato N, Oda K, Kobayashi Y, Fukuzawa H, Ohyama K (1995) Active transcription of the pseudogene for subunit-7 of the Nadh dehydrogenase in Marchantia-Polymorpha mitochondria. Mol Gen Genet 247:565–570

Terasawa K, Odahara M, Kabeya Y, Kikugawa T, Sekine Y, Fujiwara M, Sato N (2006) The mitochondrial genome of the moss Physcomitrella patens sheds new light on mitochondrial evolution in land plants. Mol Biol Evol 24:699–709

Turmel M, Otis C, Lemieux C (2003) The mitochondrial genome of Chara vulgaris: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell 15:1888–1903

Ullrich H, Lättig K, Brennicke A, Knoop V (1997) Mitochondrial DNA variations and nuclear RFLPs reflect different genetic similarities among 23 *Arabidopsis thaliana* ecotypes. Plant Mol Biol 33:37–45

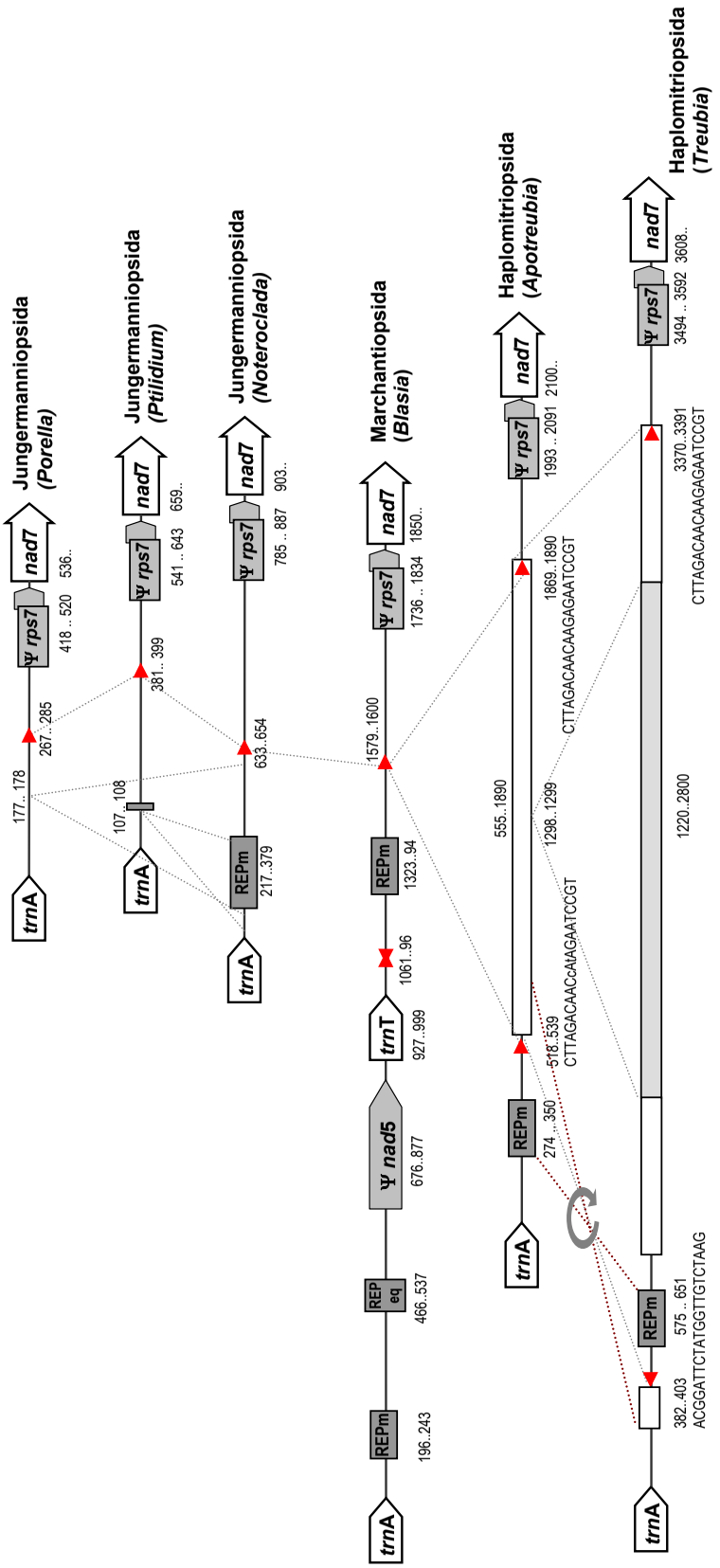
Further unpublished results

Additional *trn*A-*nad*7 spacer sequences were subsequently obtained for four other jungermanniid taxa (*Gymnomitrion concinnatum*, *Lepidogyna hodgsoniae*, *Nardia scalaris* and *Ptilidium pulcherrimum*) as well as for *Treubia lacunosa*, another species of the Haplomitriopsida. All four jungermanniids share the lack of *trn*T and the small spacer size, extending and fully confirming the published conclusion. In the light of the novel sequences, further independently acquired deletions resulting in a yet smaller *trn*A-*nad*7 spacer region of 659 bp and 536 bp were discovered in *Ptilidium* and the already published sequence of *Porella* (see Fig. 3.1).

The *trn*A-*nad*7 spacer sequence in *Treubia* is as peculiar as in its sister genus *Apotreubia*. The size increased to nearly 3700 bp with yet another 1600 bp of unknown sequence inserted, without similarity to any of other sequence in the databases. While the 5' and 3' region were perfectly homologous to Apotreubia and the other liverworts, an inversion affected the already known repetition motif (REPm), the 5' repeat region and a small part of the insert only found in Apotreubia and Treubia (Fig. 3.1).

Figure 3.1: Extended version of Fig. 3 in Wahrmund *et al.* (2008), now also including the *Porella*, *Ptilidium* and *Treubia trn*A-*nad*7 spacer regions. Numbers refer to position of elements, starting at the first nucleotide after the *trn*A. White larger arrows indicate the genes *trn*A, *trn*T, *nad*7, grey arrows represent pseudogene fragments of *nad*5 and *rps*7. Repetitive motifs also found elsewhere in the mitochondrial genome of *Marchantia* are shown in grey boxes labelled REP. Smaller arrows depict assumed recombination points for insertions, deletions and inversions. Two independently acquired deletions are found in *Porella* and *Ptilidium*. An inversion affected a smaller part of the spacer in *Treubia* and *Apotreubia*. The grey bar indicates unknown sequence inserted or deleted in *Apotreubia* after nucleotide 1298, the white bar displays sequences only found in *Apotreubia* and *Treubia*. ⇒

Figure 3.1:


Jungermanniopsida (Porella)
Jungermanniopsida (Ptilidium)
Jungermanniopsida (Noteroclada)
Marchantiopsida (Blasia)
Haplomitriopsida (Apotreubia)
Haplomitriopsida (Treubia)
trnA
Ψ rps7
nad7
REPm
REP eq
trnT
Ψ nad5
177..178
267..285
418..520
536..
107..108
381..399
541..643
659..
633..654
785..887
903..
1579..1600
1736..1834
1850..
196..243
217..379
1061..96
1323..94
466..537
676..877
927..999
555..1890
518..539
274..350
1869..1890
1993..2091
2100..
1298..1299
1220..2800
382..403
575..651
3370..3391
3494..3592
3608..
CTTAGACAACAAGAGAATCCGT
CTTAGACAACAAGAGAATCCGT
CTTAGACAACAAGAGAATCCGT
ACGGATTCTATGGTTGTCTAAG

3.2.2 Milena Groth-Malonek, Ute Wahrmund, Monika Polsakiewicz and Volker Knoop: Evolution of a pseudogene: Exclusive survival of a functional mitochondrial *nad*7 gene supports *Haplomitrium* as the earliest liverwort lineage and proposes a secondary loss of RNA editing in Marchantiidae. *Molecular Biology and Evolution* 24 (4): 1068-1074

In *Marchantia* a *nad*7 pseudogene, degenerated by six stop codons, resides in the chondriome harbouring two group II introns, while an intron-less copy transferred to the nuclear genome provides the essential gene product. To assess whether this gene transfer and subsequent pseudogenization occurred recently, mitochondrial *nad*7 sequences of representatives of all three major liverwort clades were investigated. All Marchantiopsida except *Blasia*, the basal-most genus, possess pseudogenes with stop codons and minor deletions or inversions. Exceptionally, in *Blasia* a large deletion occurs in the mitochondrial *nad*7 locus. Within the Jungermanniopsida a different mode of degeneration introduced frequent insertions and deletions thereby disrupting the reading frame. Only *Haplomitrium*, belonging to the basal-most clade, showed no indication of a reading frame degenerated by stop codons or frame shifts, also confirmed on cDNA level. The cDNA analyses revealed correctly spliced sequences, identical to the mitochondrial copy except for numerous RNA editing changes. A comparison of nuclear derived *nad*7 sequences of *Marchantia* and *Harpanthus* displayed nucleotide exchanges compared to functional mitochondrial *nad*7 sequences of the alga *Chara vulgaris* and the moss *Physcomitrella patens*, suggesting a one-time nuclear gene transfer in the ancestor of marchantiid and jungermanniid liverworts. Thus, the pseudogene was retained for about 400 million years.

Contribution: I obtained sequences for the 5' mitochondrial *nad*7 copy in *Haplomitrium* and sequenced additional *Haplomitrium* cDNA clones to confirm correct splicing and RNA editing.

Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial *nad7* Gene Supports *Haplomitrium* as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae

Milena Groth-Malonek, Ute Wahrmund, Monika Polsakiewicz, and Volker Knoop

Institut für Zelluläre und Molekulare Botanik, Abteilung Molekulare Evolution, Kirschallee 1, D-53115 Bonn, Germany

Gene transfer from the mitochondrion into the nucleus is a corollary of the endosymbiont hypothesis. The frequent and independent transfer of genes for mitochondrial ribosomal proteins is well documented with many examples in angiosperms, whereas transfer of genes for components of the respiratory chain is a rarity. A notable exception is the *nad7* gene, encoding subunit 7 of complex I, in the liverwort *Marchantia polymorpha*, which resides as a full-length, intron-carrying and transcribed, but nonspliced pseudogene in the chondriome, whereas its functional counterpart is nuclear encoded. To elucidate the patterns of pseudogene degeneration, we have investigated the mitochondrial *nad7* locus in 12 other liverworts of broad phylogenetic distribution. We find that the mitochondrial *nad7* gene is nonfunctional in 11 of them. However, the modes of pseudogene degeneration vary: whereas point mutations, accompanied by single-nucleotide indels, predominantly introduce stop codons into the reading frame in marchantiid liverworts, larger indels introduce frameshifts in the simple thalloid and leafy jungermanniid taxa. Most notably, however, the mitochondrial *nad7* reading frame appears to be intact in the isolated liverwort genus *Haplomitrium*. Its functional expression is shown by cDNA analysis identifying typical RNA-editing events to reconstitute conserved codon identities and also confirming functional splicing of the 2 liverwort-specific group II introns. We interpret our results 1) to indicate the presence of a functional mitochondrial *nad7* gene in the earliest land plants and strongly supporting a basal placement of *Haplomitrium* among the liverworts, 2) to indicate different modes of pseudogene degeneration and chondriome evolution in the later branching liverwort clades, 3) to suggest a surprisingly long maintenance of a nonfunctional gene in the presumed oldest group of land plants, and 4) to support the model of a secondary loss of RNA-editing activity in marchantiid liverworts.

Introduction

In contrast to animals, the mitochondrial DNA of plants is characterized by larger genomic complexity and significant evolutionary plasticity. Additional genes, the presence of introns, the incorporation of foreign DNA from the nucleus and the chloroplast, frequent genomic recombination, and an ongoing functional gene transfer from the mitochondrion to the nucleus are typical features of plant mitochondrial genomes (Knoop 2004). Functional transfer of individual genes from the mitochondrial to the nuclear genome is known to occur frequently and independently in flowering plants (Adams et al. 2001, 2002; Adams and Palmer 2003). In most instances, genes for proteins of the small (*rps* genes) or large (*rpl* genes) subunits of mitochondrial ribosomes or tRNA genes are subject to gene transfer, sometimes also subunits of complex II of the respiratory chain (*sdh* genes). The core set of typical mitochondrial genes encoding subunits of the respiratory chain protein complexes I (*nad* genes), III (*cox* genes), and IV (*cob*) and of the ATPase (*atp*) is generally found to be universally conserved in the mitochondrial genomes of land plants (embryophytes) and in green algae. A noteworthy exception is the *cox2* gene, for which the establishment of a functional gene copy in the nucleus and the following inactivation and disintegration of the original mitochondrial copy can been traced in leguminous plants (Nugent and Palmer 1991; Covello and Gray 1992; Adams et al. 1999). The quick degeneration of a formerly functional mitochondrial gene into a pseudogene after establishment of a functional gene copy in the nucleus is typical of angiosperms. Indeed, a complete loss of the nonfunctional mitochondrial gene barely leaving traces are often the first conclusive hint for a functional gene transfer in the history of a respective taxon (Adams et al. 2000).

A second interesting example of a core respiratory subunit gene transfer event is the case of *nad7* in the liverwort *Marchantia polymorpha*, in fact also the only clearly documented example of functional gene transfer in a non-angiosperm land plant (Kobayashi et al. 1997). The functional nuclear copy of *nad7* in *Marchantia* has a typical 5′ reading frame extension encoding the appropriate target sequence for organellar import. Somewhat in contrast to what is generally observed in angiosperms, the mitochondrial *nad7* copy in *Marchantia* is turned into a nonfunctional pseudogene through the introduction of 6 stop codons but otherwise remains intact and complete over the full extension of the reading frame from start to stop (Oda et al. 1992). This may indicate an evolutionary recent event of gene transfer and could suggest the presence of functional *nad7* genes in mitochondria of related liverworts.

In flowering plants (angiosperms), the functional mitochondrial *nad7* gene carries 3 or 4 group II introns (Bonen et al. 1994), and the 2 upstream introns are also conserved in mosses (Pruchner et al. 2001), but the liverwort gene contains none of the angiosperm-like introns. Two unrelated group II introns are present in different positions in the *nad7* pseudogene of *Marchantia*. The presence of unrelated introns in mitochondrial genes of liverworts in comparison with other land plants is a general observation (Pruchner et al. 2001; Knoop 2004). Although the mitochondrial *nad7* pseudogene was shown to be transcribed in *Marchantia*, no splicing of the 2 introns was detectable (Takemura et al. 1995), and this raises the possibility that the lack of splicing functionality was involved in pseudogene degeneration. The secondary structures of the 2 *Marchantia* introns mostly conform with the group II intron consensus, but in both cases, intron and exon binding sites are not

Key words: livercorts, mitochondrial DNA, pseudogene evolution, RNA editing, phylogeny.

E-mail: volker.knoop@uni-bonn.de.

Mol. Biol. Evol. 24(4):1068–1074. 2007
doi:10.1093/molbev/msm026
Advance Access publication February 5, 2007

Table 1
List of Liverworts Investigated in This Study

Taxonomy	Species	Voucher	Size/accession number
Haplomitriopsida	*Haplomitrium mnioides* (Lindb.) Schust.	M. Shimamura s.n.	DNA: 3570 bp/EF010864 cDNA: 916 bp/EF010865
Marchantiopsida/Blasiopsida	*Blasia pusilla* L.	J. Heinrichs 2291	1032 bp/EF010866
Marchantiopsida, sensu stricto	*Marchantia polymorpha* L. chondriome	—	5668 bp/NC 001660
	Marchantia polymorpha L., nuclear *nad7*	—	1942 bp, Kobayashi et al. (1997)
	Bucegia romanica Radian	Ulm-collection s.n.	1276 bp/EF010867
	Conocephalum conicum (L.) Underw.	Groth and Schwertfeger s.n.	1267 bp/EF010868
	Monosolenium tenerum Griff./Sunita Kapila and SS Kumar	live culture Goettingen	1276 bp/EF010869
	Lunularia cruciata (L.) Dum. ex Lindb.	Groth and Schwertfeger s.n.	1283 bp/EF010870
Jungermanniopsida simple thalloids	*Aneura pinguis* (L.) Dumort.	MGM031218-01SC	1142 bp/EF010871
Jungermanniopsida leafy liverworts	*Lepidogyna hodgsoniae* Grolle	MGM031218-02SC	1198 bp/EF010872
	Calypogeia muelleriana (Schiffner) K. Müller	J. Heinrichs 4375	1232 bp/EF010873
	Frullania tamarisci (L.) Dumort.	J. Heinrichs 4382	1233 bp/EF010874
	Harpanthus flotovianus (Nees) Nees	J. Heinrichs 4390	mitochondria: 1230 bp/EF010875, nucleus: 680 bp/EF010876
	Scapania nemorea (L.) Grolle	J. Heinrichs 4372	1233 bp/EF010877

NOTE.—Sequences of the amplicons are given with their respective lengths and have been deposited in the database under the novel accession numbers indicated. Sequences from *Chara vulgaris* (NC_005255), *Marchantia polymorpha*, and *Physcomitrella patens* (1713035479) included for comparison were taken from the database. s.n., sine numero

completely compatible. As no other plant group with a functional mitochondrial *nad7* gene shares these particular introns, it is as yet unclear whether they were correctly spliced at any time in evolution.

Liverworts represent an evolutionary old land plant clade and are in fact possibly the phylogenetic sister group to a clade comprising all other land plants including mosses, hornworts, lycophytes, monilophytes (ferns, horsetails, and whisk ferns), and seed plants (angiosperms and gymnosperms) (Qiu et al. 2006). Mitochondrial gene sequences including the positionally stable group I and group II introns have contributed to current models of land plants phylogeny (Qiu et al. 1998; Groth-Malonek and Knoop 2005; Groth-Malonek et al. 2005). Algae closely related to the land plants such as *Chara vulgaris* (Turmel et al. 2003) and *Chaetosphaeridium globosum* (Turmel et al. 2002) carry functional *nad7* genes in their mitochondria, but these genes do not contain any introns. Mosses have functional mitochondrial nad7 genes and carry 2 of the angiosperm-type group II introns (Hashimoto and Sato 2001; Pruchner et al. 2001). Hence, the data indicate presence of a functional *nad7* gene in the ancestor of all embryophytes and a subsequent degeneration into a pseudogene in *Marchantia* and possibly in related liverwort taxa.

An improved understanding of liverwort phylogeny is currently emerging from multigene studies (Davis 2004; Forrest and Crandall-Stotler 2004, 2005; He-Nygrén et al. 2004; Frey and Stech 2005; Heinrichs et al. 2005; Knoop V, unpublished data). The classical morphological distinction of complex thalloid taxa (marchantiid) versus simple thalloid and leafy (metzgeriid/jungermanniid) taxa is well corroborated by the molecular analyses, but several novel insights have emerged. We wished to trace the evolutionary history of the *nad7* pseudogene in liverworts more and less closely related to *Marchantia*. The objective of our study was 2-fold: 1) to address the modes of mitochondrial gene disintegration in a plant group so far not in the focus of gene transfer studies and 2) to obtain additional molecular data useful for phylogeny reconstruction in this ancient land plant group.

Materials and Methods

Plant taxa under study are listed in table 1. Total nucleic acids were extracted from green plant material in the presence of cetyl-trimethyl-ammonium-bromide. DNA and RNA were differentially precipitated in the presence of 3 M lithium acetate. OmniScript Reverse Transcriptase (Qiagen, Hiden, Germany) was used for cDNA synthesis. Polymerase chain reaction (PCR) amplification assays contained 1 μl template DNA or cDNA (approximately 10 ng–0.5 μg), 1 unit *Taq* DNA polymerase (Genaxxon, Biberach, Germany) or Silverstar *Taq* (Eurogentec, Seraing, Belgium), 5 μl corresponding 10× PCR buffer, 2–3 mM $MgCl_2$, 200 μM dNTPs each, 0.2 mM of each primer, 2–4% DMSO (Dimethyne sulfonide), and double-distilled water added up to 50 μl. A typical amplification assay included initial denaturation at 92 °C for 1 min, followed by 10 cycles of 92 °C for 1 min, 57–50 °C for 1 min, 72 °C for 2 min, followed by 30 cycles of 92 °C for 1 min, 50 °C for 1 min, 72 °C for 2–2.5 min, and a final step of synthesis for 15 min at 72 °C. Primers used for the DNA assays were n7i336up (5′-ggt agg act ctc gta att gga ttg c-3′) and n7i1113do (5′-gtt gta ttc acc cag aca ata acc-3′), for the nuclear gene assay 7E1+ (5′-caa cac cct gca gct cat ggt g-3′) and 7E3− (5′-cca aac aca ata tct cga gta cc-3′), and for the cDNA assay were nad7up2 (5′-atg atg gcn car gaa cay gc-3′) and nad7do2 (5′-tct atc tac ctc tcc aaa cac aat-3′). Primers to amplify the 5′ and 3′ termini of the mitochondrial *nad7* open reading frame (ORF) in *Haplomitrium* were nad7upv23 (5′-ccg tag ata ttt atg cta tgg ttt gaa tgg-3′) and nad7i336do (5′-ggg tct cga agt tyc gca tcg agg aac g-3′) and nad7i56up (5′-tac mgc agt aga agc rcc taa agg-3′) and nad7downstream (5′-gtc cta cct gtg caa tta gta gtc-3′), respectively. PCR

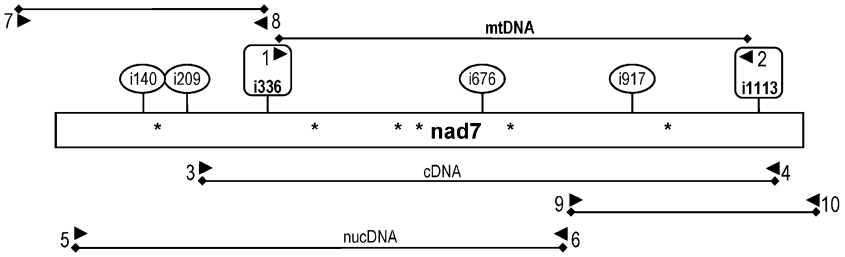


FIG. 1.—Graphical overview of the mitochondrial *nad7* gene in land plants. Asterisks indicate the positions of 6 stop codons in the mitochondrial *nad7* gene of the liverwort *Marchantia polymorpha*, which turn it into a pseudogene. Rounded squares represent group II introns known in *Marchantia*, ellipses indicate introns known in angiosperms and mosses. Intron nomenclature numbers indicate the position with respect to the upstream nucleotides in the continuous reading frame. Primers anchoring in nad7i336 and nad7i1113 (arrowheads 1: nad7i336up and 2: nad7i1113do) were used to amplify the large central *nad7* region, demonstrating the presence of both introns in liverworts in general. Primers binding in flanking exons (arrowheads 3: nad7up2, 4: nad7do, 5: 7E1+, and 6: 7E3−) were used for amplification from cDNA in *Haplomitrium* and to retrieve a nuclear gene copy in *Harpanthus*. Primers 7: nad7upv23, 8: nad7i336do, 9: nad7i56, and 10: nad7downstream were finally successfully used to amplify and determine the complete *nad7* coding sequence in *Haplomitrium*.

fragments were sequenced directly on an ABI 3100 capillary sequencer using the BigDye Terminator Cycle Sequencing v2.0 kit (PE Biosystems, Foster City, CA), or cloned into the pGEM-T Easy vector (Promega, Manhein, Germany) and sequenced on an ALF Express II (Amersham Biosciences) using the Sequenase Cy5 Dye Terminator kit or the fluorescent labeled Primer Cycle Sequencing kit (GE Healthcare, Munich, Germany), or were commercially sequenced (Macrogen Inc., Korea). Sequences were aligned with MEGA3 (Kumar et al. 2004) using the implemented Clustal algorithm and manually adjusted.

Results

Five of the 6 stop codons in the mitochondrial *nad7* pseudogene are located between the 2 large group II introns nad7i336 and nad7i1113 in the liverwort *M. polymorpha* (fig. 1, intron designations follow Dombrovska and Qiu 2004). Hence, the large central exon between them was an attractive region to address pseudogene evolution assuming that the 2 introns would be conserved in other liverwort taxa. Design of primers targeting the mitochondrial introns could ensure amplification of the mitochondrial copy in liverworts rather than a potential nuclear version. Moreover, this approach circumvented the risk of PCR failure using primers targeting terminal exon regions given the large group II intron sizes or after potential disintegration of gene termini through recombination events.

The amplification and sequencing of the *nad7* region with primers anchoring in the 2 liverwort-specific introns nad7i336 and nad7i1113 (fig. 1) was indeed successful for a taxonomically diverse spectrum of liverworts. PCR products of expected sizes around 1300 bp were obtained with the exception of *Blasia* where a product of about 1000 bp was retrieved (table 1). All PCR products were sequenced to clarify the status of the mitochondrial *nad7* sequences. We first checked other taxa of the marchantiid group of complex thalloid liverworts closely related to *Marchantia*. As in *Marchantia*, the *nad7* reading frames were also found to carry stop codons in *Bucegia, Conocephalum, Lunularia*, and *Monosolenium* (fig. 2). Surprisingly, one of the *Marchantia* stop codons (s3765) is shared among all taxa (except *Blasia*), and 2 other ones (s3483 and s4064) are shared with all taxa except *Lunularia* (and *Blasia*). The 2 remaining stops are unique to *Marchantia*, and one novel stop codon each is identified in *Lunularia* and *Bucegia*. As in *Marchantia*, there are no reading frameshifts in the closely related taxon *Bucegia* and only 1 or 2 single-nucleotide frameshifts, respectively, in the other taxa. These very minor differences and the otherwise high degree of sequence similarities even in a pseudogene correlate well with the extreme degree of sequence conservation of functional mitochondrial genes in the marchantiids that had been observed before (Beckert et al. 1999).

The genus *Blasia* is of particular interest to the phylogeny of liverworts because recent molecular data have strongly suggested its inclusion among the marchantiids in a basal position instead of its classic assignment to the simple leafy, jungermanniid taxa. Unusually, in *Blasia*, we observed an internal deletion of 238 bp removing a large portion of the *nad7* reading frame (fig. 2). Interestingly, this large deletion is immediately upstream of a codon insertion (in3878+3) universally present in the marchantiid taxa. Neither frameshifts nor stop codons were observed in the regions flanking the large deletion in *Blasia*, possibly suggesting that those observed in the derived marchantiids were gained after separation from the *Blasia* lineage and so possibly support its basal placement.

In the next step, we included phylogenetically more distant taxa of the Jungermanniopsida (simple thalloid and leafy taxa) into our survey. Again, our taxonomic spectrum was wide so as to include representatives from the well-established clades as based on recent insights on liverwort phylogeny. As in the marchantiids, the *nad7* reading frame was found destroyed in all cases. However, whereas single-nucleotide frameshifts or stop codons render *nad7* a pseudogene in the marchantiids, oligonucleotide insertions and deletions of up to 25 bases are characteristic of *nad7* pseudogenes in the Jungermanniopsida (fig. 2). Three frameshifting indels are shared by all Jungermanniopsiid taxa: in3551+2, in3764+5, and in3902−25. Other indels occur independently and at particularly high frequency in *Aneura*, a simple thalloid liverwort and member of supposedly early diverging groups of the Jungermanniopsida, which shows 16 indels in total

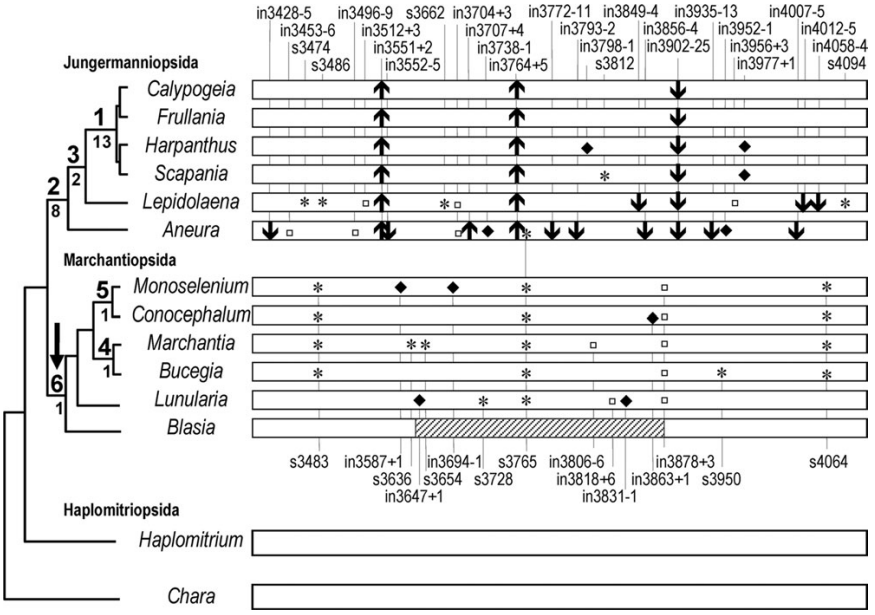


FIG. 2.—Graphical overview of the *nad7* exon region between the introns nad7i336 and nad7i1113 for *Marchantia polymorpha* and the 12 liverworts investigated in this study. Asterisks represent stop codons (s), rhombs represent single-nucleotide frameshifts, up and down arrows indicate insertions or deletions that result in frameshifts (in), the open squares indicate codon indels that do not disrupt the reading frame. The respective positions are indicated with the *M. polymorpha* nucleotide position of the *nad7* gene (including introns) followed by the number of inserted (+) or deleted (−) nucleotides, respectively. The hatched box represents the deletion of a major part of the central *nad7* exon upstream of in3878+3 in *Blasia*. Presumed phylogenetic relationships of the taxa are shown to the left, summarizing findings from several recent studies (Davis 2004; Forrest and Crandall-Stotler 2004; He-Nygrén et al. 2004). Cladistic support for some of the nodes (node identifier above branch) also comes from several synapomorphic codon changes (number below branch) in the *nad7* pseudogene amplicon (for amino acid positions of the *Chara* sequence refer to fig. 3. Node 1: D123G, L176P, Y183H, G206S, S219G, G227E, L244S, I262T, Q272R, S304D, H/R310Q, L314P, L340S; node 2: M122T, R151G, S183Y, M191V, M275T, F340L, N344D, G364E; node 3: E135G, G161E; node 4: A163G; node 5: R196C; and node 6: E336S). Loss of RNA editing has presumably occurred at node 6 (arrow).

in the central *nad7* exon. All other taxa show fewer indels, up to 9 in *Lepidogyna*.

The clear distinction of jungermanniid and marchantiid taxa is also clearly supported by 8 codon changes, which appear as synapomorphies of the Jungermanniopsida and one as synapomorphy for the clade of marchantiids including *Blasia* (fig. 2). A further 13 such changes confirm a clade of 4 derived jungermanniid taxa.

Although these data so far suggested that *nad7* is generally a pseudogene in liverwort mitochondria, we finally included the genus *Haplomitrium* in our taxon sampling. As in the case of *Blasia*, molecular data had recently suggested a new taxonomic placement, possibly at the base of liverworts as a whole (Crandall-Stotler et al. 2005). The complete *nad7* coding region between the 2 liverwort-type introns was found intact in *Haplomitrium*. Neither frameshifts nor stop codons or frame-conserving codon indels were found. Moreover, no significant mutations affecting conserved amino acids were observed other than some codon exchanges that could potentially be corrected by the C-to-U type of RNA editing typical for plant organellar genes to reconstitute codon identities, which were previously found in abundance in this genus in a study of the *nad5* gene (Groth-Malonek et al. 2005). To address whether the *Haplomitrium* sequence actually represents a functional sequence, we used RT–PCR. Because functional splicing of the liverwort intron sequences would be a fundamental prerequisite for expression, we extended the amplicon with primers anchoring in the upstream and downstream flanking exons (fig. 1). An RT (reverse transcription)–PCR product of a size expected for correct splicing was obtained and sequenced. Comparison of cDNA and DNA sequences indeed confirmed correct and precise splicing at the expected sites and showed complete sequence identity between cDNA and DNA except for all 15 expected RNA-editing positions to reconstitute conserved codon identities as now confirmed by the cDNA (fig. 3). A remote possibility would be that the cDNA could have been derived from a transcript of a *nad7* gene copy very recently transposed into the nucleus in *Haplomitrium* that has not accumulated any nucleotide exchanges. To address this possibility, we wished to clone the 5′ and 3′ terminal sequences of the *nad7* ORF in *Haplomitrium*. This proved difficult given that flanking intergenic sequences are generally only rarely conserved in plant mitochondrial DNA and required several attempts with different primers. However, in the end, we succeeded to amplify and clone the 5′ end of *nad7* with a downstream primer anchoring in nad7i336 and a 5′ primer anchoring 39 bp upstream of the *nad7* start codon. The 5′ end of the *Haplomitrium nad7* gene is colinear with the mitochondrial homologues in other taxa and identical with the cDNA sequence overlapping in the exon upstream of i336 except for 3 further codons found to be edited as could be expected (fig. 3). The methionine start codon in the *Haplomitrium nad7* ORF is located at the same position as the start codon in other mitochondrial nad7 genes. In addition, a stop—only 12 codons upstream

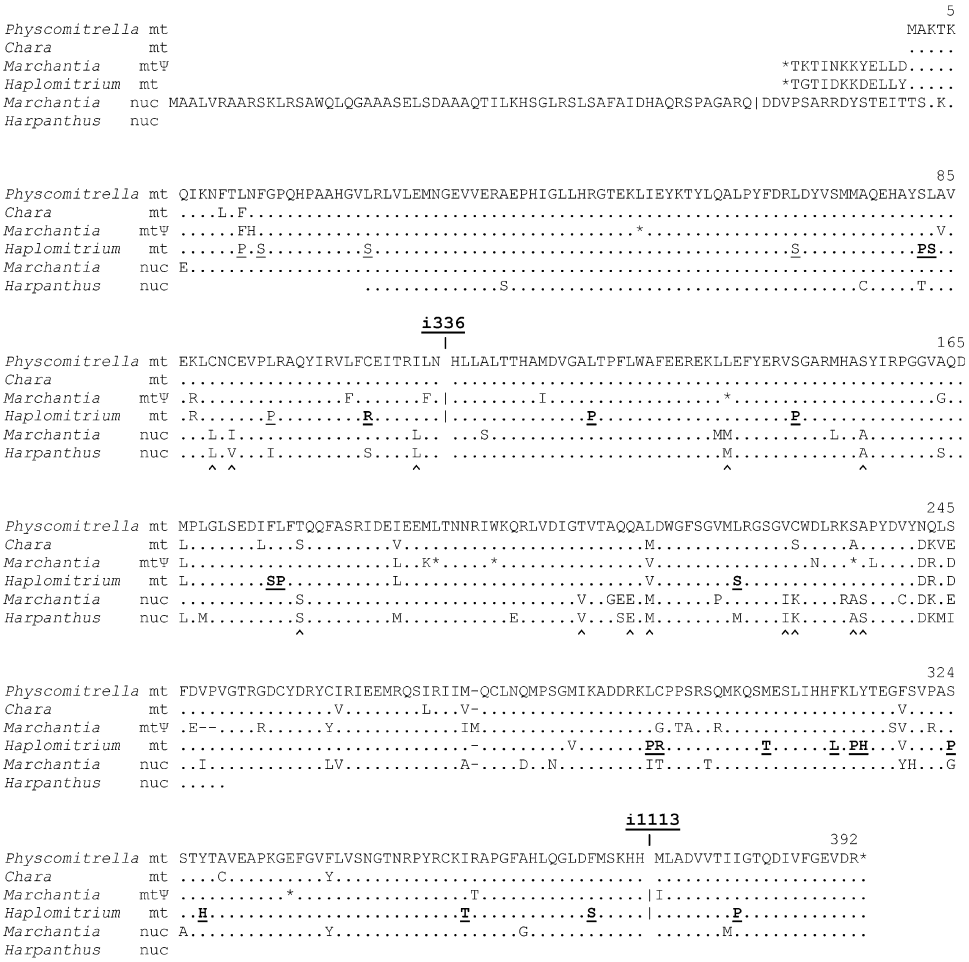


FIG. 3.—Amino acid sequence alignment of the *nad7* gene of *Haplomitrium mnioides* with the corresponding regions in the mtDNAs of the moss *Physcomitrella patens*, the alga *Chara vulgaris*, and both the mitochondrial pseudogene sequence and the functional nuclear gene copy in the liverwort *Marchantia polymorpha*. Intron positions of the liverwort mitochondrial introns i336 and i1113 in the *Marchantia* and *Haplomitrium* mitochondrial sequences and of the spliceosomal intron in the amino terminal extension of the *Marchantia* nuclear copy are indicated by vertical lines. The protein translation of the likely nuclear and functional *nad7* sequence isolated from *Harpanthus flotovianus* is included. Codon synapomorphies of the 2 nuclear sequences in *Harpanthus* and *Marchantia* are indicated by carets below the sequences. Amino acids are numbered for the *Chara* nad7 protein sequence as a reference. Amino acids identical to the *Physcomitrella* sequence are indicated by dots, dashes represent alignment gaps, and stop codons in the mitochondrial pseudogene of *Marchantia* are indicated by asterisks. The amino acid positions requiring RNA editing in *Haplomitrium* to reconstitute the conserved canonical codon identities in *Chara* or *Physcomitrella*, respectively, are indicated by underlining and those confirmed by cDNA analysis are shown in bold. There is an in-frame stop codon 36 nucleotides upstream of the presumed *nad7* start codon in the *Marchantia* and *Haplomitrium* mitochondrial sequences.

(coincidentally conserved at this position in the *Marchantia* mitochondrial sequence)—clearly documents absence of a potential targeting sequence extension that could be indicative of a nuclear gene copy. Efforts to amplify the 3′ end of the *Haplomitrium nad7* gene in the end were successful with a 5′ primer anchoring in the coding region upstream of i1113 and a 3′ primer anchoring immediately downstream of the stop codon in *Haplomitrium* to determine the conserved 3′ end of *nad7*. Again, the sequence is perfectly colinear with its mitochondrial homologues in other taxa (fig. 3), including the highly conserved carboxy terminal amino acid motif GEVDR and the stop codon position and is identical with the cDNA sequence except for the confirmation of an expectedly edited proline codon (fig. 3).

Finally, we wished to verify that a functional *nad7* gene copy resides in the nucleus of jungermanniid species. A PCR amplification with the exon-based primers initially employed to retrieve the functional nuclear counterpart in *Marchantia* (Kobayashi et al. 1997) was successful in amplifying the corresponding sequence (from amino acid positions 24 through 253) in *Harpanthus*. We were unable to amplify this region from other taxa in our sampling, a result likely due to a combination of lower sequence conservation in the more divergent nuclear gene copies and genome complexities. In the case of *Haplomitrium*, however, the failure to detect any *nad7* copy other than the obviously mitochondrial one with this and any of the other primer combinations used could certainly also reflect a phylogenetic status before nuclear

gene transfer, which may have occurred only in the stem lineage of the remaining liverworts after split from *Haplomitrium* (fig. 2). The (presumably nuclear) *Harpanthus* sequence is, as expected, free of group II (and other) introns or frameshifts and shares most sequence similarity with the *Marchantia* nuclear *nad7* gene (fig. 3). Several codon positions in the *Harpanthus nad7* sequence are shared with the *Marchantia* nuclear *nad7* sequence and are not shared with any of the mitochondrial sequences included in these analyses (C89L, C91I/V, I110L, L141M, S155A, T207V, Q212E, V228I, C229K, A236S). This suggests an ancient gene transfer establishing the functional nuclear *nad7* gene before the split of jungermanniid and marchantiid taxa.

Discussion

Numerous examples of frequent and independent gene transfer from the mitochondrion to the nucleus have been reported, however, mainly for genes encoding ribosomal proteins among angiosperms. No example for a functional gene transfer of any one of the *nad* genes among flowering plants had been reported (Adams et al. 2002). Hence, the case of *nad7* gene transfer in a liverwort (Kobayashi et al. 1997) is noteworthy in 2 respects: it is the only known functional gene transfer event from the mitochondrion in nonangiosperm land plants and the only one that involves one of the 9 mitochondrially encoded *nad* genes.

We report that the mitochondrial *nad7* pseudogene sequences are generally present in liverworts across a wide taxonomic range. These sequences carry both liverwort-type group II introns known from *M. polymorpha*, nad7i336 and nad7i1113. Although the mitochondrial pseudogene of *Marchantia* is transcribed, both introns are not spliced. In contrast, mosses apparently have functional *nad7* genes in their mitochondria, which carry 2 upstream angiosperm-type group II introns nad7i140 and nad7i209 (Hashimoto and Sato 2001; Pruchner et al. 2001). This finding in itself is not surprising given that plant mitochondrial introns are generally stable in position within a particular phylogenetic clade (Beckert et al. 1999; Dombrovska and Qiu 2004; e.g., Vangerow et al. 1999; Pruchner et al. 2002; Qiu et al. 2006). Notably, the very different occurrence of mitochondrial introns in liverworts and other land plants is strong evidence for the deepest dichotomy of embryophyte phylogeny (Qiu et al. 1998, 2006).

The persistence of the mitochondrial *nad7* pseudogene in liverworts is a puzzling observation. If liverworts indeed are the sister group to all other land plants and in so far a phylogenetically very old clade, presumably exceeding an age of 400 Myr, why is a nonfunctional pseudogene conserved for so long in their mitochondrial genomes with only minor changes—notably very minor ones in the marchantiid subgroup?

Nad7 sequence analysis indicates that the mode of pseudogene degeneration differs in the 2 liverwort subclades. Whereas the *nad7* gene is rendered nonfunctional through introduction of stop codons and single-base indels in the marchantiids, indels of several bases dominate in the jungermanniid taxa. The fact that all surveyed members within each taxon share similar classes of mutations suggests that different modes of pseudogenization arose early in the 2 extant clades. One of the stop codons (s3765) is conserved in all surveyed complex thalloid liverworts and also occurs in *Aneura*, the earliest diverging branch of the Jungermanniopsida in our sampling. The large gap that was found in *Blasia* is, when aligned with other liverwort sequences, placed in a region that includes the position of this highly conserved stop codon. Hence, this could possibly be the one stop codon that initiated the loss of functionality of the mitochondrial *nad7* gene at least in marchantiids, if not the entire clade of non-*Haplomitrium* liverworts when assuming a later reversal in the jungermanniids. Currently, the most parsimonious explanation for the observations is the establishment of a functional *nad7* copy in the nucleus at least before the split of jungermanniid and marchantiid taxa. This is well supported by our finding of a functional copy of *nad7* in the jungermanniid taxon *Harpanthus*, which exhibits a high degree of similarity to the nuclear *nad7* sequence of *Marchantia*.

The presence of a functionally spliced and edited mitochondrial *nad7* copy in *Haplomitrium* is particularly noteworthy in other respects. First, it serves as an independent genomic character that supports the basal phylogenetic placement of this genus among liverworts as a whole. Furthermore, it clearly supports the idea that both liverwort-type group II introns, nad7i336 and nad7i1113, were present in a functional mitochondrial *nad7* gene of the primordial liverwort ancestor and were correctly spliced at that time. Finally, the finding also lends strong support to a phylogenetically plausible scenario of secondary loss of RNA-editing activity (fig. 2), which is absent in the marchantiid liverworts (Steinhauser et al. 1999), instead of independent gain of editing activity in *Haplomitrium*, the Jungermanniopsida, and the clade of all other land plants.

Disintegration of the mitochondrial gene copy after functional transfer to the nucleus is frequently so fast in angiosperms that absence of a mitochondrial gene copy had been taken as an indicator of gene transfer in the first place (Adams and Palmer 2003). A notable exception has only very recently been reported for the survival of rps14 pseudogenes among the grasses for maybe some 80 Myr (Ong and Palmer 2006). The dominating point mutations and small-scale indels observed in liverworts stand in contrast to the otherwise (quick) recombinational disruption of pseudogenes known in angiosperms. The high recombinational activity typical for the mitochondrial DNAs of angiosperms may be a later evolutionary gain in the tracheophyte lineage, not present to that extent in liverworts. The new observations are in line with the absence of disrupted (*trans*-splicing) introns in bryophytes, which instead are conventionally *cis*-arranged there (Malek and Knoop 1998; Dombrovska and Qiu 2004; Groth-Malonek et al. 2005).

Acknowledgments

We are very grateful to colleagues Hans Becker, Saarbrücken, S. Robbert Gradstein, Jochen Heinrichs, Henk Groth, Michael Schwertfeger, Göttingen, Yin-Long Qiu, Ann Arbor, and Masaki Shimamura, Hiroshima for providing biological materials and to Felix Grewe for comments on the manuscript. Research in the authors' laboratory is supported by the Deutsche Forschungsgemeischaft.

Literature Cited

Adams KL, Daley DO, Qiu YL, Whelan J, Palmer JD. 2000. Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. Nature. 408:354–357.

Adams KL, Palmer JD. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. Mol Phylogenet Evol. 29:380–395.

Adams KL, Qiu YL, Stoutemyer M, Palmer JD. 2002. Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. Proc Natl Acad Sci USA. 99:9905–9912.

Adams KL, Rosenblueth M, Qiu YL, Palmer JD. 2001. Multiple losses and transfers to the nucleus of two mitochondrial succinate dehydrogenase genes during angiosperm evolution. Genetics. 158:1289–1300.

Adams KL, Song KM, Roessler PG, Nugent JM, Doyle JL, Doyle JJ, Palmer JD. 1999. Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial cox2 genes in legumes. Proc Natl Acad Sci USA. 96:13863–13868.

Beckert S, Steinhauser S, Muhle H, Knoop V. 1999. A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial nad5 gene. Plant Syst Evol. 218:179–192.

Bonen L, Williams K, Bird S, Wood C. 1994. The NADH dehydrogenase subunit-7 gene is interrupted by 4 group-II introns in the wheat mitochondrial genome. Mol Gen Genet. 244:81–89.

Covello PS, Gray MW. 1992. Silent mitochondrial and active nuclear genes for subunit 2 of cytochrome c oxidase (cox2) in soybean: evidence for RNA-mediated gene transfer. EMBO J. 11:3815–3820.

Crandall-Stotler B, Forrest LL, Stotler RE. 2005. Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). Taxon. 54:299–316.

Davis EC. 2004. A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). In: Goffinet B, Hollowell V, Magill R, editors. Molecular systematics of bryophytes. St Louis (MO): Missouri Botanical Garden Press. p. 61–86.

Dombrovska E, Qiu YL. 2004. Distribution of introns in the mitochondrial gene nad1 in land plants: phylogenetic and molecular evolutionary implications. Mol Phylogenet Evol. 32:246–263.

Forrest LL, Crandall-Stotler BJ. 2004. A phylogeny of the simple thalloid liverworts (Jungermanniopsida, Metzgeriidae) as inferred from five chloroplast genes. In: Goffinet B, Hollowell V, Magill R, editors. Molecular systematics of bryophytes. St Louis (MO): Missouri Botanical Garden Press. p. 119–140.

Forrest LL, Crandall-Stotler BJ. 2005. Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. J Hattori Bot Lab. 97:127–159.

Frey W, Stech M. 2005. A morpho-molecular classification of the liverworts (Hepaticophytina, Bryophyta). Nova Hedwigia. 81:55–78.

Groth-Malonek M, Knoop V. 2005. Bryophytes and other basal land plants: the mitochondrial perspective. Taxon. 54:293–297.

Groth-Malonek M, Pruchner D, Grewe F, Knoop V. 2005. Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses to vascular plants. Mol Biol Evol. 22:117–125.

Hashimoto K, Sato N. 2001. Characterization of the mitochondrial nad7 gene in Physcomitrella patens: similarity with angiosperm nad7 genes. Plant Sci. 160:807–815.

Heinrichs J, Gradstein SR, Wilson R, Schneider H. 2005. Towards a natural classification of liverworts (Marchantiophyta) based on the chloroplast gene rbcL. Cryptogam Bryol. 26:215–233.

He-Nygrén X, Ahonen I, Juslén A, Glenny DPS. 2004. Phylogeny of liverworts—beyond a leaf and a thallus. In: Goffinet B, Hollowell V, Magill R, editors. Molecular systematics of bryophytes. St Louis (MO): Missouri Botanical Garden Press. p. 87–118.

Knoop V. 2004. The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. Curr Genet. 46:123–139.

Kobayashi Y, Knoop V, Fukuzawa H, Brennicke A, Ohyama K. 1997. Interorganellar gene transfer in bryophytes: the functional nad7 gene is nuclear encoded in Marchantia polymorpha. Mol Gen Genet. 256:589–592.

Kumar S, Tamura K, Nei M. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. 5:150–163.

Malek O, Knoop V. 1998. Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. RNA. 4:1599–1609.

Nugent JM, Palmer JD. 1991. RNA-mediated transfer of the gene coxII from the mitochondrion to the nucleus during flowering plant evolution. Cell. 66:473–481.

Oda K, Yamato K, Ohta E, Nakamura Y, Takemura M, Nozato N, Akashi K, Kanegae T, Ogura Y, Kohchi T, et al. (11 co-authors). 1992. Gene organization deduced from the complete sequence of liverwort Marchantia polymorpha mitochondrial DNA. A primitive form of plant mitochondrial genome. J Mol Biol. 223:1–7.

Ong HC, Palmer JD. 2006. Pervasive survival of expressed mitochondrial rps14 pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. BMC Evol Biol. 6:55.

Pruchner D, Beckert S, Muhle H, Knoop V. 2002. Divergent intron conservation in the mitochondrial nad2 gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. J Mol Evol. 55:265–271.

Pruchner D, Nassal B, Schindler M, Knoop V. 2001. Mosses share mitochondrial group II introns with flowering plants, not with liverworts. Mol Genet Genom. 266:608–613.

Qiu YL, Cho YR, Cox JC, Palmer JD. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature. 394:671–674.

Qiu YL, Li L, Wang B, et al. (21 co-authors). 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proc Natl Acad Sci USA. 103:15511–15516.

Steinhauser S, Beckert S, Capesius I, Malek O, Knoop V. 1999. Plant mitochondrial RNA editing: extreme in hornworts and dividing the liverworts? J Mol Evol. 48:303–312.

Takemura M, Nozato N, Oda K, Kobayashi Y, Fukuzawa H, Ohyama K. 1995. Active transcription of the pseudogene for subunit-7 of the NADH dehydrogenase in marchantia-polymorpha mitochondria. Mol Gen Genet. 247:565–570.

Turmel M, Otis C, Lemieux C. 2002. The chloroplast and mitochondrial genome sequences of the charophyte Chaetosphaeridium globosum: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. Proc Natl Acad Sci USA. 99:11275–11280.

Turmel M, Otis C, Lemieux C. 2003. The mitochondrial genome of Chara vulgaris: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell. 15:1888–1903.

Vangerow S, Teerkorn T, Knoop V. 1999. Phylogenetic information in the mitochondrial nad5 gene of pteridophytes: RNA editing and intron sequences. Plant Biol. 1:235–243.

Charles Delwiche, Associate Editor

Accepted January 31, 2007

Further unpublished results

Further sequences for the second exon and small parts of the adjacent intron regions of the mitochondrial *nad*7 copy were obtained for the jungermanniid taxa *Metzgeria furcata*, *Noteroclada confluens* and *Fossombronia pusilla* confirming the degeneration by frequent insertions or deletions (Fig. 3.2). In addition, the survey was extended to the complete pseudogene region in *Blasia*, the basal-most marchantiid liverwort, which showed a smaller insertion as well as a stop codon in the first pseudogene exon. The completely sequenced mitochondrial genome of *Pleurozia purpurea* (Wang *et al.* 2010) allowed the inclusion of its Ψ-*nad*7. Surprisingly, it showed a larger sequence deletion – similar, but unequal to *Blasia*.

Besides smaller indels and a stop codon, also conserved in its position in *Marchantia*, two large deletions are found in *Pleurozia*. One affects the second exon and one causes the loss of the third exon and large parts of the group II intron i1113, leaving only domains I and II. In contrast, nearly the complete first exon except for 18 nucleotides 5' of the group II intron i336 is missing in *Aneura*. Amplification was only successful using *trn*A anchored primers yielding a fragment of an exceptionally small *trn*A-*nad*7 spacer (170 bp), 18 nucleotides of the first pseudogene exon and 1754 bp of the nad7i336 intron. The last exon could not be acquired by PCR, questioning its presence also in *Aneura*. In all taxa investigated so far, group II intron i336 has not been deleted. One of the primers for amplifying the second pseudogene exon in the eleven published taxa and three novel ones binds at the 3' end of loop IV in nad7i336, comparable to a region that is already deleted in the second intron i1113 in *Pleurozia*. Hence, at least the last appr. 300 nucleotides, covering domains V and VI completely, were present in those taxa. Nearly the complete intron i336 could be sequenced in *Aneura* and *Calycularia*, with only domain IV missing in *Aneura* and domain IV to VI in *Calycularia*. The approach to sequence the missing intron regions (and the second exon in *Calycularia*) failed. However, the complete sequence of the first and third exon and both introns was successfully obtained for *Blasia* in addition to the already published sequence of the second exon. An investigation of the acquired nad7i336 sequences, surprisingly, did not reveal any larger deletions, not even in *Pleurozia* (Wang *et al.* 2010) and *Blasia* (and domain I, II, III, V, VI in *Aneura*), where larger deletions were already found in other regions of the pseudogene. Due to mismatches in vital binding regions in the pseudogene intron, it is not supposed to be spliced out correctly. However, as the cause for the long retention of the pseudogene is still unclear, an unknown function resting within the intron region might prevent the complete degradation of the pseudogene and loss from the mitochondrial genome.

As *Haplomitrium* still harbours a functional *nad*7 copy in its chondriome, *nad*7 sequences of *Apotreubia* and *Treubia* were further obtained and investigated. Here again, neither mutations leading to stop codons nor frame shifts induced by insertions or deletions are found (Fig. 3.2). From sequence

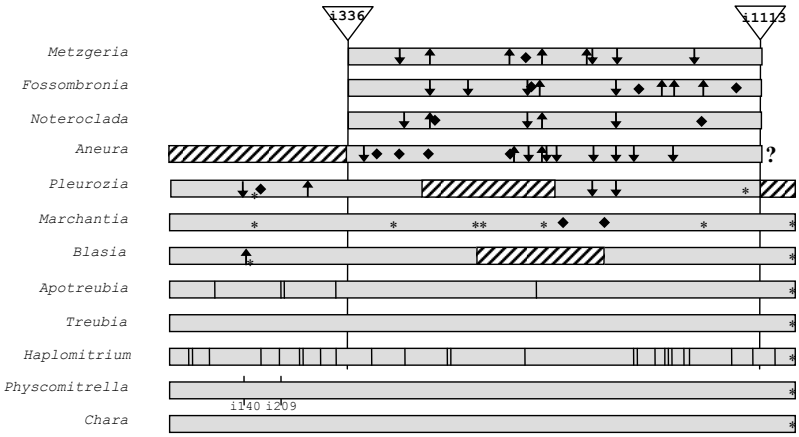


Figure 3.2: Graphical overview of the complete coding region of the functional mitochondrial *nad*7 for the three haplomitriid liverworts *Haplomitrium*, *Treubia* and *Apotreubia*, the moss *Physcomitrella patens* and the alga *Chara vulgaris* in combination with the pseudogenes of seven other liverworts, two marchantiid: *Marchantia* and *Blasia* and five jungermanniid: *Aneura*, *Fossombronia*, *Metzgeria*, *Noteroclada* and *Pleurozia* (this study and Wang *et al.* 2010). Asterisks indicate stop codons, arrows insertions (up) or deletions (down) that result in frameshifts, filled rhombs indels that do not disrupt the reading frame. Hatched boxes represent larger deletions. The position of the two liverwort group II introns i336 and i1113 as well as two introns in *Physcomitrella* are shown. The *Pleurozia* Ψ-nad7i1113 intron sequence homology is breaking off after intron domain II, also lacking the γ interacting nucleotide (see Fig. 3.3B). The amino acid positions requiring RNA editing to reconstitute the conserved codon identities in *Haplomitrium* (22/25 confirmed by cDNA analysis) and *Apotreubia* are indicated by vertical lines.

comparison a secondary structure model of both introns was proposed (Fig. 3.3). Necessary base pairing is provided, thus allowing folding and correct splicing. Unfortunately, RNA of these rare and isolated taxa could not be obtained, restraining confirmation of transcription and splicing by cDNA analyses. Consequently, RNA editing positions could likewise only be suggested.

Nevertheless, the potential RNA editing frequency differs strikingly from what is known in *Haplomitrium* (22 confirmed plus three proposed editing sites). While only five editing events are proposed for *Apotreubia*, none is necessary in *Treubia* (Fig. 3.2), showing a surprising reduction of RNA editing in closely related taxa.

The sequence of the complete functional nuclear *nad*7 copy of *Marchantia* have been obtained before (Kobayashi *et al.* 1997). The authors also identified the presumed mitochondrial target sequence (i.e. upstream protein sequence extension prior to the actual *nad*7 coding region) encoding an amphiphilic helix that could facilitate import of the *nad*7 protein into the mitochondrion. If the gene transfer into the nucleus occurred in the ancestor of marchantiid and jungermanniid liverworts and was

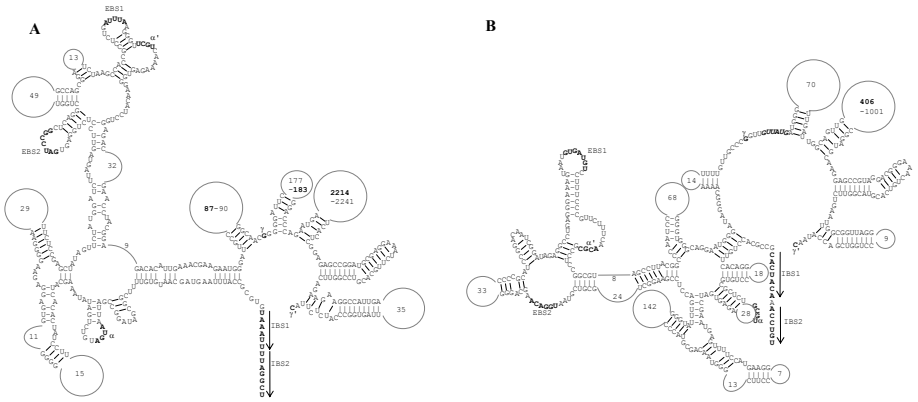


Figure 3.3: Proposed secondary structure of a) *nad*7 intron i336 and b) *nad*7 intron i1113 in haplomitriid liverworts (*Haplomitrium mnioides*, *Apotreubia nana* and *Treubia lacunosa*). Domaines numbered according to structural conventions (Michel *et al.* 1989). Interacting nucleotides are shown in bold and named accordingly. Nucleotides in intron i1113 region between domain II and III not present in *Haplomitrium* are indicated in bold italics.

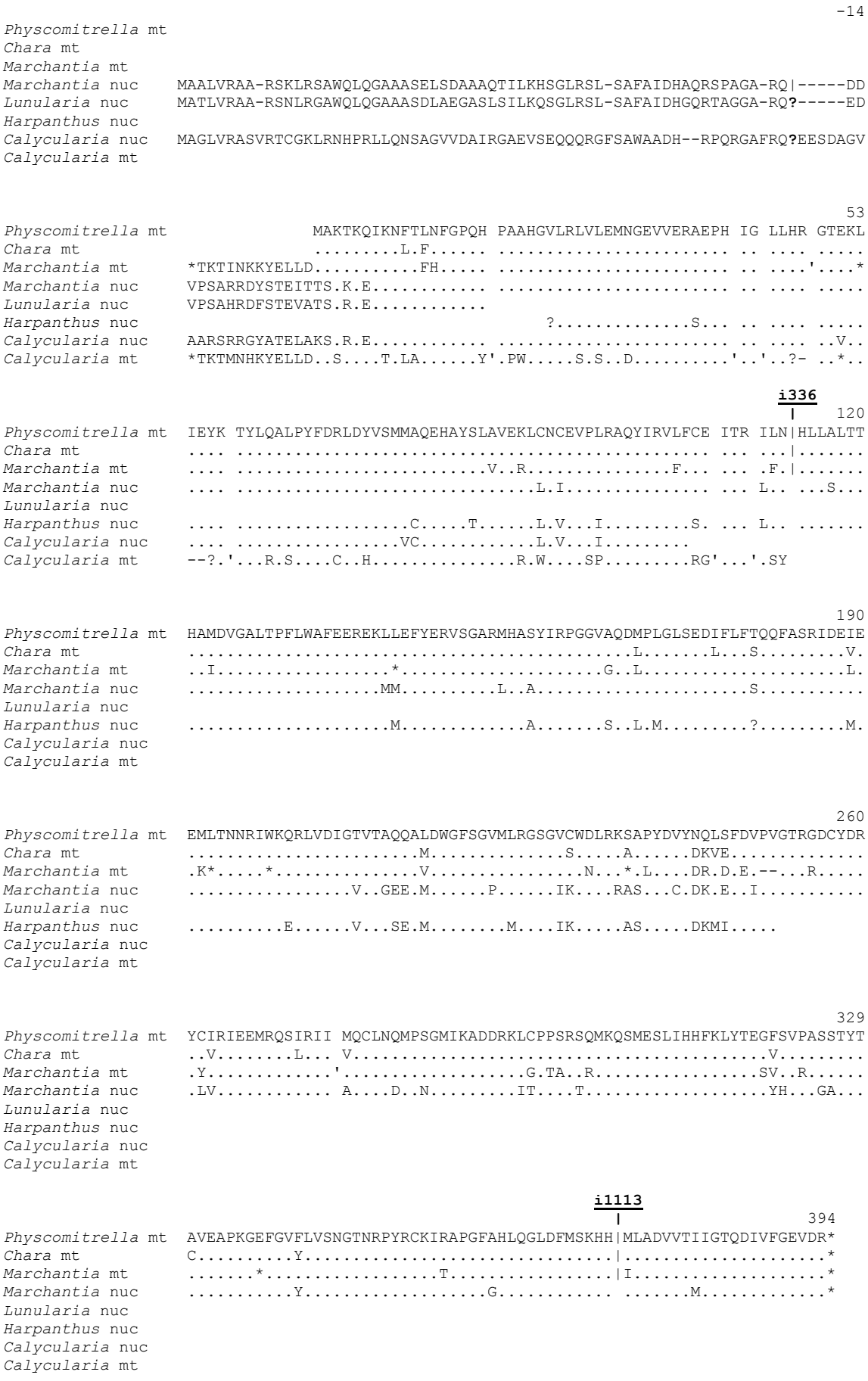
thus a one-time event, a similar mitochondrial target peptide should be observed. Therefore, the nuclear *nad*7 copy of the jungermanniid liverwort *Calycularia crispula* and the marchantiid *Lunularia cruciata* was targeted by 5'RACE. For *Calycularia*, a region containing 312 nucleotides of the coding region and 294 nucleotides of the according 5' UTR plus target signal (including a stop codon prior to the presumed start codon, Fig. 3.4) was successfully amplified. The *Lunularia* sequence included 99 nucleotides of the coding region and 403 upstream nucleotides with two stop codons prior to the presumed start. *nad*7 coding sequences of the marchantiid *Marchantia* and *Lunularia* and the jungermanniid *Harpanthus* and *Calycularia* are nearly identical, demonstrating a high selective pressure to retain its functionality. Only minor changes can be observed, especially noteworthy is the replacement of A75C, C91V and L95I in both jungermanniid taxa, reflecting their common ancestry. In comparison with the functional mitochondrial *nad*7 sequence of *Physcomitrella patens* and *Chara vulgaris* (Fig. 3.4) and the three haplomitriid liverworts (not shown), changes at amino acid position T4KR and Q6E were found in the nuclear *nad*7 copies of *Marchantia*, *Lunularia* and *Calycularia* and C89L in *Marchantia*, *Calycularia* and *Harpanthus*.

The deduced amino acid sequence of the *Calycularia* 5'RACE sequence beginning with the presumed start codon was predicted to contain a mitochondrial localisation peptide by TargetP (82.1 %, Emanuelsson et al., 2000), WoLFPSORT (7/13, Horton et al., 2007) and PA-SUB (100 %, Lu et al., 2004). In contrast to the coding region, the mitochondrial target sequences of *Marchantia* and *Lunularia* compared to *Calycularia* are variable and share only few short peptide motifs (e.g. SAFAIDH in *Marchantia* and *Lunularia* vs. SAWAADH in *Calycularia*). These differences could be the result

of the evolutionary distance between both classes in addition to high substitution rates in the nuclear genome. Even the two closer related marchantiid liverworts *Marchantia* and *Lunularia* show sequence deviations. The intron in the target peptide region of *Marchantia* would clearly indicate a common origin of the nuclear copy, when also found in a jungermanniid taxon. An approach to obtain this region with genomic DNA of *Calycularia* and *Lunularia* failed, while the intron sequence of *Marchantia* could be amplified and sequenced without difficulties.

Figure 3.4: Amino acid sequence alignment of the nuclear *nad*7 gene of *Marchantia polymorpha*, *Lunularia cruciata*, *Harpanthus flotovianus* and *Calycularia crispula* with the corresponding regions in the mtDNAs of all three species in addition to the moss *Physcomitrella patens* and the alga *Chara vulgaris*. Intron positions of the liverwort mitochondrial introns i336 and i1113 and the spliceosomal intron in the N-terminal extension of the *Marchantia* nuclear copy are indicated by vertical lines. Amino acids are numbered with the *Physcomitrella nad*7 protein sequence as reference. Amino acids identical to the *Physcomitrella* sequence are indicated by dots, dashes represent alignment gaps. Stop codons in the mitochondrial pseudogene of *Marchantia* are indicated by asterisks, indels in the mitochondrial pseudogene fragment of *Calycularia* by apostrophes. ⇒

Figure 3.4:


-14
Physcomitrella mt
Chara mt
Marchantia mt
Marchantia nuc MAALVRAA-RSKLRSAWQLQGAAASELSDAAAQTILKHSGLRSL-SAFAIDHAQRSPAGA-RQ|-----DD
Lunularia nuc MATLVRAA-RSNLRGAWQLQGAAASDLAEGASLSILKQSGLRSL-SAFAIDHGQRTAGGA-RQ?-----ED
Harpanthus nuc
Calycularia nuc MAGLVRASVRTCGKLRNHPRLLQNSAGVVDAIRGAEVSEQQQRGFSAWAADH--RPQRGAFRQ?EESDAGV
Calycularia mt
53
Physcomitrella mt MAKTKQIKNFTLNFGPQH PAAHGVLRLVLEMNGEVVERAEPH IG LLHR GTEKL
Chara mt .....L.F..... ....
Marchantia mt *TKTINKKYELLD.....FH..... ....'....*
Marchantia nuc VPSARRDYSTEITTS.K.E..... ....
Lunularia nuc VPSAHRDFSTEVATS.R.E.....
Harpanthus nuc ?.....S... .. ..
Calycularia nuc AARSRRGYATELAKS.R.E..... ....V..
Calycularia mt *TKTMNHKYELLD..S....T.LA.....Y'.PW.....S.S..D.....'..'..?- ..*..
i336
120
Physcomitrella mt IEYK TYLQALPYFDRLDYVSMMAQEHAYSLAVEKLCNCEVPLRAQYIRVLFCE ITR ILN|HLLALTT
Chara mt ....|.....
Marchantia mt .....V..R.....F... ... .F.|.....
Marchantia nuc .....L.I..... ... L.. ...S...
Lunularia nuc
Harpanthus nuc .....C.....T.....L.V...I.....S. ... L.. .....
Calycularia nuc .....VC.....L.V...I.....
Calycularia mt --?.'...R.S....C..H.....R.W....SP.....RG'...'.SY
190
Physcomitrella mt HAMDVGALTPFLWAFEEREKLLEFYERVSGARMHASYIRPGGVAQDMPLGLSEDIFLFTQQFASRIDEIE
Chara mt .....L.....L...S.....V.
Marchantia mt ..I.....*.....G..L.....L.
Marchantia nuc .....MM.....L..A.....S.....
Lunularia nuc
Harpanthus nuc .....M.....A.....S..L.M.....?.....M.
Calycularia nuc
Calycularia mt
260
Physcomitrella mt EMLTNNRIWKQRLVDIGTVTAQQALDWGFSGVMLRGSGVCWDLRKSAPYDVYNQLSFDVPVGTRGDCYDR
Chara mt .....M.....S.....A.....DKVE.....
Marchantia mt .K*.....*.....V.....N...*.L....DR.D.E.--...R.....
Marchantia nuc .....V..GEE.M.....P.....IK....RAS...C.DK.E..I.....
Lunularia nuc
Harpanthus nuc .....E.....V...SE.M.....M....IK.....AS.....DKMI.....
Calycularia nuc
Calycularia mt
329
Physcomitrella mt YCIRIEEMRQSIRII MQCLNQMPSGMIKADDRKLCPPSRSQMKQSMESLIHHFKLYTEGFSVPASSTYT
Chara mt ..V.....L... V.....V.....
Marchantia mt .Y.....'.....G.TA..R.....SV..R.....
Marchantia nuc .LV..... A....D..N.....IT....T.....YH...GA...
Lunularia nuc
Harpanthus nuc
Calycularia nuc
Calycularia mt
i1113
394
Physcomitrella mt AVEAPKGEFGVFLVSNGTNRPYRCKIRAPGFAHLQGLDFMSKHH|MLADVVTIIGTQDIVFGEVDR*
Chara mt C.....Y.....|.....*
Marchantia mt .....*.....T.....|I.....*
Marchantia nuc .....Y.....G..... .....M.....*
Lunularia nuc
Harpanthus nuc
Calycularia nuc
Calycularia mt

4 Discussion

4.1 The impact of novel mitochondrial markers on moss phylogeny

For decades, detailed morphological studies built the base for a moss classification (e.g. Brotherus 1924, 1925, Buck 1991, Dixon 1932, Vitt 1984). In particular, anatomy and morphology of peristome types in the sporophyte generation played a major role. Many systematic assignments were corroborated by molecular markers. Several taxa, however, were found, according to molecular phylogenies, to belong to a different clade than previously classified, e.g. due to a parallel evolution of morphological features. With the introduction of novel mitochondrial markers presented here, the phylogenetic placement of some of those enigmatic taxa has been addressed. Moreover, a clear consensus on the backbone phylogeny of the basal moss lineages is emerging, an issue that was hardly considered with classic systematic treatments.

Placement of some enigmatic taxa

Catoscopium and *Timmiella*, assigned to different subclasses based on morphological characters, are now confidently placed as gradual sister species at the base of the Dicranidae (Volkmar and Knoop subm, Wahrmund *et al.* 2010), a phylogenetic placement that was only weakly supported before (Goffinet *et al.* 2001, Hedderson *et al.* 2004). In the case of *Catoscopium*, the presence of two rings of peristome teeth led to an assignment to the Bryidae, although the endostome is reduced to a mere hyaline membrane (Goffinet and Buck 2004). In *Timmiella*, the single peristome ring consists of 16 rudimentary teeth with two filaments at the tip of each (Zander 1994), resembling the haplolepideous peristome of the Pottiaceae. The teeth of the single peristome ring of the Dicranidae are at positions comparable to the endostome teeth of the diplolepideous opposite peristome of the Funariidae. Considering that all Dicranids are nested within the diplolepideous mosses, the ancestor of all Dicranids could well have possessed two peristome rings. Thus, the phylogenetic placement of *Catoscopium* as basal-most dicranid might reflect the ancestral diplostome (two rings) peristome state. While the exostome was reduced in all other Dicranidae, in *Catoscopium* instead, the endostome could have been altered. Did the modification occur only in *Catoscopium* or are other basal dicranids affected? As *Timmiella* already possesses the typical haplolepideous peristome of dicranid mosses, the change of peristome architec-

ture might have happened fast. Nevertheless, some taxa morphologically placed in one of the other subclasses could well belong to the basal Dicranidae, but have not been investigated with molecular markers so far. A straightforward approach to discard the exostome during the critical phase of the alteration of peristome architecture could have given way in favour of the exploration of different peristome structures before the 'typical' dicranid peristome succeeded. Including more taxa with morphologically ambiguous peristomes into molecular analyses and detailed morphological studies of different stages of peristome development can then yield further insight into peristome evolution and the transition from diplolepideous to haplolepideous peristomes.

The phylogenetic position of *Timmia*, in contrast, remained unclear also with the novel mitochondrial loci analysed. Depending on the markers and taxon sampling used, its affinities to other clades varied, but have never received any significant statistical support. *Timmia* was either included in the Funariidae (Cox *et al.* 2004, Goffinet *et al.* 2001), in the basal arthrodontous mosses (Goffinet and Cox 2000, Newton *et al.* 2000) or as sister to a Funariidae-Dicranidae clade (Beckert *et al.* 1999, 2001). The novel mitochondrial markers introduced in this study for the time being suggest an independent placement as a separate subclass (Volkmar and Knoop subm, Wahrmund *et al.* 2009, 2010). *Timmia* clearly belongs to the core Bryopsida (excluding *Buxbaumia* and *Diphyscium*). However, due to the long period of isolated evolution, long branch attraction might influence its positioning in phylogenetic trees. Adding more *Timmia* species resulted in a clear exclusion from the other clades. Still, the backbone of the core Bryopsida is poorly supported with short branches leading to the monophyletic groups (Volkmar and Knoop subm, Wahrmund *et al.* 2010). Considering only well-supported branches, a polytomy of five clades (*Timmia*, Gigaspermaceae, Bryidae, Funariidae and Dicranidae) remains. The unique peristome only present in *Timmia* does not clarify its position and might have evolved independently from the other peristome types. While the exostome contains 16 teeth, the endostome is a basal membrane with 64 equal-sized filaments on top (Budke *et al.* 2007). Some members of the Bryidae also possess 64 endostome filaments. Their size, however, varies due to asymmetric cell divisions in the inner peristomal layer (Magombo 2003). Symmetric cell divisions of the inner peristomal layer is also found in Funaria (Budke *et al.* 2007), but there resulting in only 16 endostome teeth. Thus, neither morphological peristome features nor molecular data can clearly place *Timmia* into one of the three subclasses. Considering a phylogenetic placement of *Timmia* separate from all other groups and a peristome that combines characters of two subclasses (Funariidae and Bryidae), it might resemble the ancestral peristome of the core Bryopsida that gave rise to the other peristome types.

Including members of the Gigaspermaceae in the phylogenetic analyses questioned a Funariidae concept incorporating this family (Goffinet *et al.* 2007). Similar to the 71 kb inversion in the chloroplast present in all other Funariidae families (Disceliaceae, Funariaceae, Encalyptaceae) it suggests the ex-

clusion of the Gigaspermaceae from a Funariidae sensu strictu (Volkmar and Knoop subm, Wahrmund *et al.* 2010). With the novel molecular data, even a weakly supported sister relationship of the Gigaspermaceae with the Bryidae can be observed (Wahrmund *et al.* 2010). The Gigaspermaceae were assigned to the Funariidae because of their leaf morphology and the peristome which lacks peristome teeth (gymnostomous), that is also found in *Discelium* and several Funariaceae taxa (Goffinet and Buck 2004). Hence, if an exclusion of the Gigaspermaceae from the Funariidae will be verified in the future, the gymnostomous peristome must have evolved independently in both groups. Given the unresolved polytomy from which the three major subclasses, *Timmia* and the Gigaspermaceae emerge (Wahrmund *et al.* 2010), many morphological changes that led to the formation of these five clades – e.g. in peristome architecture – have supposedly occurred within a short period of time. Currently, only few synapomorphic characters can be found in the so far analysed molecular data and these are obscured by background noise when phylogenetic trees are inferred. Unfortunately, no unique indel or inversion event in the plastid or mitochondrial genomes have yet been identified which could support a close relationship between any two members of these five clades. Moreover, events of gene transfer to the nucleus rather add strength to already known relationships instead of resolving the polytomy in the core Bryopsida. The only known example so far is the loss of *rpo*A in all Bryopsida except *Buxbaumia*, that also occurred in *Tetraphis* (Goffinet *et al.* 2005). Thus, two independent losses, one in *Tetraphis* and one in the ancestor of *Diphyscium* and all core Bryopsida, have to be assumed, nevertheless strengthen the support for a common ancestor of the latter relationship.

The novel mitochondrial loci clearly support a joint clade of the Tetraphidopsida and Polytrichopsida (Volkmar and Knoop subm, Wahrmund *et al.* 2009, 2010). Both groups share a nematodontous peristome built by whole cells. However, the presence of four teeth in Tetraphidopsida was supposed to be a plesiomorphic character, especially because more teeth, found in the Polytrichopsida (16-64), and the sequence of cell divisions similar to arthrodontous peristome development suggested a closer relationship of Polytrichopsida and Bryopsida (Magombo 2003). A sister relationship of Polytrichopsida and Bryopsida has been observed only once with the chloroplast marker *rps*4 (Magombo 2003). An extended taxon sampling with the same marker resulted in a phylogeny that included *Buxbaumia* into a Tetraphidopsida-Polytrichopsida-clade (Goffinet *et al.* 2001). Most other molecular markers however, displayed an unsupported monophyletic nematodontous moss clade (Beckert *et al.* 1999, 2001, Wahrmund *et al.* 2009, 2010). With the inclusion of *Tetrodontium* and the exploration of the novel mitochondrial group I intron cox1i624, good statistical support validates this clade well (Volkmar and Knoop subm). In addition, chemical analyses of secondary metabolites also identified a similar phenolic pattern mainly consisting of tri- and tetrahydroxycoumarin derivatives in members of Tetraphidopsida and Polytrichopsida (Jung *et al.* 1994, 1995).

Andreaea and *Andreaeobryum*, although placed in separate subdivisions are clearly recognized as being a well supported monophyletic entity with all markers investigated (Volkmar and Knoop subm) and could as well be united into a single subdivision. This close relationship between both taxa was seen before albeit without or with low statistical support (Cox *et al.* 2004, Goffinet *et al.* 2001, Newton *et al.* 2000).

The basal-most moss genus and features of the moss ancestor

Although all mitochondrial loci analysed so far place *Takakia* at the base of all mosses, the basal-most moss genus is still not determined with convincing statistical support. Chloroplast loci and concatenated data sets with loci from both organelle genomes more or less support a clade of *Sphagnum* and *Takakia* as sister to all other mosses. This basal-most grouping is in so far debatable as it is only seen inferred with *rbc*L. A phylogenetic tree based on *rps*4 alone identified *Sphagnum*, *Takakia*, *Andreaea* and *Andreaeobryum* as the basal-most clade. The apparent contradiction between the chloroplast data sets certainly contrasts the agreement of all mitochondrial markers. In addition, the possession of unique morphological characters likewise sets *Takakia* aside from all other mosses. Its whole gametophyte habit does not resemble any of the extant moss gametophytes. In fact, *Takakia* was earlier placed into the liverwort order Calobryales as sister to *Haplomitrium* (e.g. Schuster 1966). Only after a moss-like sporophyte was discovered, it was recognized as a moss. As the opening of the spore capsule with one vertical fissure is comparable to the four vertical fissures that open the spore capsule of *Andreaeobryum*, both genera were united in the subclass Takakiidae within the class Andreaeopsida (Smith and Davison 1993). However, the unique combination of sporophytic moss and gametophytic liverwort morphological characters (Smith and Davison 1993) has no resemblance to any other moss. The chloroplast favoured *Takakia-Sphagnum* clade is likewise not undermined by morphological characters. With respect to its secondary metabolites, the oil composition of *Takakia* is more similar to liverworts than to mosses, underlining a basal position within the mosses, when interpreted as a plesiomorphy (Anhut *et al.* 1984, Asakawa 2004, Markham and Given 1988, Saritas *et al.* 2001). In addition, *Takakia* contains a variety of unique flavone glucosides (Markham and Porter 1979). These features, however, could be interpreted as either ancestral (plesiomorphic) to all mosses and lost or modified later during moss evolution or these peculiarities are uniquely acquired in *Takakia* as autapomorphies after it evolved from the ancestor shared with *Sphagnum*.

A molecular character state unique to *Takakia* is the extraordinarily high amount of C-to-U RNA editing compared to all other mosses (Sugita *et al.* 2006, Yura *et al.* 2008). Assuming that *Takakia* resembles the moss ancestor in its RNA editing frequency, a very high RNA editing rate in early branching mosses gradually decreases during moss evolution (unpublished observation), possibly as

a consequence of a better adaptation to living on land and more efficient protection against mutagenic UV light. In the case of an assumed basal-most *Sphagnum-Takakia* clade, the common ancestor of all mosses not necessarily had to have high RNA editing frequencies, as *Sphagnum* has considerably less RNA editing sites than *Takakia* but still more than any other moss (Steinhauser *et al.* 1999). Thus, the high amount of RNA editing in *Takakia* could also have increased in its lineage independently, for instance, as a necessity to compensate for a higher mutation rate.

A comparison of mitochondrial spacer and intron regions surprisingly showed a pattern of size decreases of these sequences during moss evolution. Noticeably larger spacer and introns were found in the early branching mosses, especially in *Takakia* and *Sphagnum*. Here again, a basal position of both taxa is corroborated without support, however, for either their sister relationship or for a placement of one of them as basal-most moss genus.

Novel molecular markers and the future of moss phylogeny

A critical assessment of the novel mitochondrial loci revealed that phylogenetic trees based on single data sets alone could not fully resolve all relationships. Each marker yielded sufficient information to support other monophyletic groups. The *nad*5-*nad*4 intergenic spacer, for instance, reliably resolved relationships in the Bryidae, but lacked resolution within the Dicranidae (Wahrmund *et al.* 2009). With cobi420, the backbone of mosses and especially the five clades in the core Bryopsida received good support (Wahrmund *et al.* 2010). Both loci could not address the question of the basal-most moss genus as the spacer region is too divergent between mosses and liverworts, hornworts or the charophyte alga *Chara vulgaris*, the closest living relative to land plants (Groth-Malonek *et al.* 2007a). The group I intron cobi420 is absent in liverworts and both introns found at the same position in the alga *Chara vulgaris* (Turmel *et al.* 2003) and the hornwort *Phaeoceros laevis* (Xue *et al.* 2010) are not homologous to the moss group I intron. Thus, only introns conserved in mosses and a neighbouring divisions allow to resolve the earliest branches by rooting with either the alga *Chara*, liverworts or hornworts as sister to all mosses. Besides nad5i753, cox1i624 was introduced as the only other group I intron found in both mosses and liverworts to reach a consensus on the root of mosses (Volkmar and Knoop subm).

The contradiction between chloroplast and mitochondrial data, however, left the issue of *Takakia*, *Sphagnum* or a joint clade of the two genera as the basal-most moss taxon unanswered (Volkmar and Knoop subm). The origin of the different signals might lie in ancient hybridization events between *Takakia* and *Sphagnum* resulting in (chimeric) chloroplast genomes that resolve the hybridization event but not the true phylogeny. It should also be kept in mind, that the mitochondrial marker consists of variable introns or spacers while the chloroplast loci investigated here contain only coding regions. Thus, more phylogenetic information should reside in the mitochondrial markers. In addition, intron

and spacer regions are not much affected by the high RNA editing frequency found in *Takakia* (Sugita *et al.* 2006, Yura *et al.* 2008). The basal placement of *Takakia*, inferred by mitochondrial loci, should therefore be considered as being more reliable.

The taxon sampling included members of all early branching moss lineages except for *Oedipodium* that was placed sister to all other Bryopsida by several studies (Cox *et al.* 2004, Goffinet *et al.* 2001, Newton *et al.* 2000). It is lacking a peristome either as retaining a plesiomorphic state in its intermediate position between aperistomate and peristomate mosses or as a secondary loss after it split off (Goffinet *et al.* 2001, Magombo 2003). Whether its inclusion in the phylogenetic analyses would solve the matter of the root of mosses is hard to predict.

Will the exploration of more novel molecular markers finally resolve the moss phylogeny completely? It can be doubted. Single molecular loci do not recognize all clades and certainly experience the occurrence of homoplasy. Phylogenetic trees based on concatenated data sets will face more background noise the more data sets are included. The effort invested into establishing new molecular markers and screening them for large amounts of taxa should therefore be used for carefully selecting an appropriate taxon sampling with loci already tested. A novel molecular marker that contains synapomorphies to support a certain clade does not necessarily resolve and support this clade when included in data set of several loci. Approaches to successively exclude isolated taxa from the data set diminished the risk of long branch attraction and tested the reliability of clades (Volkmar and Knoop subm, Wahrmund *et al.* 2010). Moreover, increase in statistical support by adding critical taxa can sometimes be observed, e.g. the inclusion of *Tetrodontium* strengthened a monophyly of nematodontous mosses (Volkmar and Knoop subm), supported only weakly before. Insight into the last remaining open questions will more likely come from larger-scale molecular changes, like the 71 kb inversion in the chloroplast genome of Funariidae (Goffinet *et al.* 2007). Comparison of structural changes of completely sequenced organelle genomes of at least one other moss with the already sequenced genomes of *Physcomitrella patens* will help to identify them. Until then, clues on phylogenetic mysteries have to rely more or less on lucky guesses.

Conclusion

Comparing the genomic features derived from studying several mitochondrial markers gives insight into mitochondrial DNA evolution that accompanied moss phylogeny. The larger size of intron and spacer region observable in all early branching moss genera is most likely an ancestral character. During moss evolution spacers and introns were then streamlined to smaller sizes with only vital features preserved allowing e.g. faster replication and transcription. The independent, significant size reductions found e.g. in *Leucobryum* and *Schistostega* (Volkmar and Knoop subm, Wahrmund *et al.* 2009, 2010) indi-

cate that a further decrease is still possible. Investigating C-to-U RNA editing frequencies, a decline in editing rate occurred during moss evolution, with highest amount of editing sites found in *Takakia*. A similar pattern of RNA editing reduction is observed also in liverworts. In *Haplomitrium*, belonging to the basal-most liverwort class of Haplomitriopsida, the RNA editing frequency greatly exceeds the rate of *Takakia* (Groth-Malonek *et al.* 2005, Rüdinger *et al.* 2008) and is considerably decreased in the other two haplomitriid genera. A tendency to diminish the amount of RNA editing during Jungermanniopsida evolution can also be observed as well as a complete loss of RNA editing in Marchantiopsida (Steinhauser *et al.* 1999). With the transition from living in water to living on land, the use of RNA editing could have been a way to cope with higher UV radiation before DNA repair mechanisms in the organelles adapted to the environment thereby restricting the need for it.

4.2 New insights into mitochondrial evolution of liverworts: promiscuous spacer regions and long retained pseudogenes

Within the liverworts, completely sequenced mitochondrial genomes are currently known from two species, the jungermanniid *Pleurozia purpurea* (Wang *et al.* 2010) and the marchantiid *Marchantia polymorpha* (Oda *et al.* 1992). The two species belong to different classes that split from a common ancestor approximately 370 million years ago (Heinrichs *et al.* 2007). During the long time of separate evolution not much seemed to have happened within the chondriomes. The gene order is still the same and high sequence similarity can be observed (Wang *et al.* 2010). Only two genes (*trn*R[ucg] and *trn*T[ggu]) and one intron rrn18i1065g2 are missing in *Pleurozia* compared to *Marchantia* (Wang *et al.* 2010). In addition, a large part of the group II intron nad7i1113 and the third exon of the *nad*7 pseudogene are lost in *Pleurozia* (Wang *et al.* 2010). Thus, it appears that only minor structural changes affected the mitochondrial genome during most of the liverwort evolution. In contrast to tracheophyte chondriomes with their frequent recombination resulting in different gene orders and subcircles with varying gene content, the liverwort mitochondrial genomes give the impression of slow evolution and high degree of conservatism. Similarly, a comparison of the sequenced mitochondrial genomes of two hornworts, *Megaceros aenigmaticus* (Li *et al.* 2009) and *Phaeoceros laevis* (Xue *et al.* 2010), did not exhibit much differences and showed a largely conserved mode of evolution except for a large number of degenerated pseudogenes that are retained in hornwort chondriomes (Xue *et al.* 2010). Are the bryophyte chondriomes simply not capable of recombination and have frequent rearrangements in e.g. gene order not been invented? Are they not "embarked on the journey of volatile evolution, in particular showing reshuffled gene order" (Li *et al.* 2009) as the mitochondrial genomes of tracheophytes are said to be?

The *trn*A-*trn*T-*nad*7 spacer: a highy variable region in a conserved environment

A closer look at one of the gene regions that vary between *Pleurozia* and *Marchantia*, the *trn*A-*trn*T-*nad*7 spacer, reveals high variability instead of conserved evolution. Here, frequent recombination shaped the spacer region independently in all three liverwort classes. The *trn*T and adjacent spacer regions were inverted once and lost at least three times independently during the liverwort evolution (Wahrmund *et al.* 2008).

In the ancestor of all liverworts, the *trn*T was most likely present between *trn*A and *nad*7 in the same direction as still found nowadays in the alga *Chara* and in mosses. As they are found in all liverworts analysed, some of the repetitive motifs and the small part of *rps*7 could have been introduced in the common liverwort ancestor (see 3.2.1). In contrast, the large size of the spacer in the haplomitriid liverworts *Apotreubia* and *Treubia* is probably acquired later by insertion of large fragments of non-coding DNA. Here, at least one inversion event can be detected (see Fig. 3.1). In the marchantiid liverworts, the basal-most position of *Blasia* is corroborated by possessing the gene arrangement preceding the inversion. After *Blasia* split off, the inversion of the *trn*T with neighbouring parts of the spacer occurred in the ancestor of all other marchantiid liverworts. Later, probably in the ancestor of all Marchantiales except *Marchantia* and *Bucegia*, a large part of the spacer was deleted. The presence of the larger spacer region including the inverted *trn*T also supports a basal placement of *Marchantia* and *Bucegia* within their order (Forrest *et al.* 2006). In the ancestor of all jungermanniids, a different segment of the spacer including the *trn*T was deleted, shortening the spacer size. In addition, independent deletions in the spacer region were found in *Ptilidium* and *Porella*, further decreasing its length (see Fig. 3.1).

Most of the supposed recombination points within the *trn*A-*trn*T-*nad*7 intergenic region are bordered by direct or inverted repeat sequences (Wahrmund *et al.* 2008), it can therefore be suggested that pairing of these repeats can result in inversions and deletions or even insertions of sequence stretches. In tracheophytes, small mitochondrial subcircles can be generated by intramolecular homologous recombination of repeat elements in the mitochondrial DNA (Fauron *et al.* 1995, Grewe *et al.* 2009, Sugiyama *et al.* 2005). The existence of those subcircles has not been shown for bryophytes (Oda *et al.* 1992, Terasawa *et al.* 2007). In liverworts, however, the recombinational activity found within the *trn*A-*trn*T-*nad*7 spacer during liverwort evolution indicates that some of the requirements for a dynamic structure of mitochondrial genomes are already met.

Except for the haplomitriids, the size of the *trn*A-*nad*7 region was diminished independently during the liverwort evolution. As a similar trend of reducing the size of non-coding sequences was also observed in moss evolution (see above), it is not necessarily limited to liverwort phylogeny. The majority of spacers are shorter in *Pleurozia* compared to *Marchantia* (Wang *et al.* 2010). Interestingly, the

intron lengths are for most introns similar in both species and do not account for the reduced overall chondriome length (Wang *et al.* 2010). An exception is cox1i624, whose intron size is increased in jungermanniids and marchantiids compared to haplomitriids (Volkmar and Knoop subm). Assuming that the tendency to shorten the genome length continued during liverwort evolution, the more recently evolved liverwort species within the Jungermanniopsida and Marchantiopsida might own even smaller mitochondrial genomes than known so far.

Why a particularly large *trn*A-*nad*7 spacer region is present in *Apotreubia* and *Treubia* and is not reduced as in the other two liverwort classes, is consequently hard to explain. As is known so far, both genera are very old, they have a rather limited and isolated distribution and might not have encountered many changes in their mitochondrial DNA. The inversion found within the *trn*A-*nad*7 spacer in both taxa indicates that the ability for recombination already existed but did not result in a decreased spacer length. Alternatively, the insertion of large fragments might have influenced the assembly of further repetitive elements and thereby the ability to delete parts of the spacer via recombination. In addition, the *nad*7 gene flanking this spacer is presumably still a functional gene in both genera (see below), while it is degenerated in marchantiid and jungermanniid liverworts rendering it a pseudogene there (Groth-Malonek *et al.* 2007b). The frequent recombination activity observed in Jungermanniopsida and Marchantiopsida might therefore be related to the presence of a pseudogene, as deletions or insertions near a functional gene can affect its transcription and thereby its function.

Whether the variations observed in the *trn*A-*nad*7 intergenic region mirror a general pattern, can only be assessed by investigating further spacer regions at wider taxon sampling. In the *nad*5-*nad*4-*nad*2 gene arrangement for instance, a high conservation of spacer size between several early land plant clades (*nad*4-*nad*2 spacer) as well as size reductions and expansions (*nad*5-*nad*4 spacer), even among liverworts, can be found without a pseudogene being involved (Groth-Malonek *et al.* 2007a). It is very likely that similar recombinations also occur at other spacers of the mitochondrial genome of liverworts. A simple comparison of only two species might not be sufficient to discover hot spots of recombination. Instead of sequencing yet another liverwort chondriome, spacers with considerably different sizes in *Pleurozia* and *Marchantia* should be surveyed across a selection of species covering all major liverwort lineages and assessed for their variability and for traces of recombination. Thus, the picture of extremely conserved liverwort chondriomes will most likely have to be revised. Recombinational activity is already present in liverworts even if its complexity does not reach levels found in tracheophytes (Fauron *et al.* 1995, Grewe *et al.* 2009, Sugiyama *et al.* 2005).

A long surviving *nad*7 pseudogene

The long retention of the *nad*7 pseudogene in the mitochondrial genome of marchantiid and jungermanniid liverworts is clearly arguing against fast-evolving liverwort chondriomes. After a successful gene transfer to the nucleus, the unnecessary mitochondrial copy is degraded, usually leaving no sign of its former existence. In fact, successful gene transfers are often assumed because an essential gene is lacking in the mitochondrial genome (Knoop 2004). There are only few other examples of pseudogenes having survived in plant mitochondrial genomes. In grasses (Poales), the mitochondrial *rps*14 gene was successfully transferred to the nucleus at least three times with a pseudogene copy residing in the chondriome of Poaceae and Cyperaceae for approximately 80 million years (Ong and Palmer 2006). The completely sequenced mitochondrial genomes of the hornworts *Megaceros aenigmaticus* and *Phaeoceros laevis* each contain eleven pseudogenes with nine being shared between them and these are supposedly retained for at least 100 million years (Li *et al.* 2009, Xue *et al.* 2010). In both hornworts, however, only a small fragment of the *nad*7 gene is left in the mitochondrial genome, with the gene degraded to a much larger extent than in liverworts possibly after a successful gene transfer to the nuclear genome. Given that the *nad*7 pseudogene was already present in the ancestor of marchantiid and jungermanniid liverworts, it obviously persisted for more than 370 million years (Groth-Malonek *et al.* 2007b, Heinrichs *et al.* 2007). There is still the probability that all the pseudogenes are on their way to be lost from the genome, the usual fate of a pseudogene. However, other genes have been transferred to the nucleus without leaving a trace of a pseudogene in the chondriome, e.g. *rpl*14, found in the charophyte algae *Mesostigma* and *Chara* but not in land plants (Turmel *et al.* 2002, 2003). In addition, the example of the *trn*A-*nad*7 spacer demonstrated that deletions have occurred independently in marchantiid and jungermanniid liverworts (Wahrmund *et al.* 2008) revealing an obvious tendency to reduce the size of the mitochondrial genome during evolution and to delete unnecessary features as quickly as possible. Thus, a long-term retention of pseudogenes may indicate an unknown function. In the case of *nad*7 in *Marchantia*, the mitochondrial pseudogene copy is still transcribed but not properly spliced (Takemura *et al.* 1995). The survey across the liverwort lineages revealed that different parts of the pseudogene are affected by deletions, e.g. the third exon and most part of the nad7i1113 group II intron in *Pleurozia*, a large part of the second exon in *Blasia* and most of the first exon in *Aneura*. Interestingly, the first intron, nad7i336g2, so far remained in the pseudogene without experiencing larger deletions (see Further results in 3.2.2). In addition, the sequence of this group II intron suggests folding of the RNA with minor mispairings (vital for the splicing process but not for the structure assembly) and contains remnants of a maturase or reverse transcriptase gene (Takemura *et al.* 1995). Assuming that the pseudogene is also transcribed in other liverworts, the intron can still be folded into its three-dimensional structure. A functioning of the maturase or reverse transcriptase is however prevented by

several frame shifts and stop codons (Takemura *et al.* 1995). From the nuclear genome it is known that small non-coding RNAs, derived from double-stranded RNA precursors, have specific functions in regulation of transcription, translation and in defense against viruses (Aravin *et al.* 2007, Bartel 2007, Filipowicz *et al.* 2008, He and Hannon 2004, Jones-Rhoades *et al.* 2006). Although it has not been shown for the organelle gene expression, a similar role might be appointed to small non-coding RNAs derived from paired (double-stranded) intron structures. By binding to target mRNAs (of, for instance, other genes of the respiratory chain) they could inhibit or elevate their translation. The occurrence of several types of small RNAs originated from intronic sequences is already known (Brown *et al.* 2008) albeit in the nuclear genome. Recently, small non-coding RNAs have been discovered in plant chloroplasts and mammal mitochondria (Lung *et al.* 2006), possibly also existing in plant mitochondria. The authors speculated about a possible function in gene regulation that would allow fast and efficient reactions to changing environmental conditions (Lung *et al.* 2006). A molecular machinery for recognition and processing would be necessary if the nuclear mechanism is adopted in the organelles and the nuclear encoded enzymes have to be imported into the mitochondria first. As it seems less likely, that in the mitochondria the presence of several proteins needed for processing non-coding RNAs has so long been overlooked, there is still the possibility that a completely different strategy of gene regulation involving RNA derived from the *nad*7 pseudogene transcript is present. The hypotheses presented here clearly rely on the ongoing transcription of the pseudogene in all liverworts, which has to be verified at first. Nevertheless, a pseudogene that is retained for such a long time when other transferred genes were deleted and spacer regions shortened during liverwort evolution indicates an essential function.

A comparison with other long retained pseudogenes, e.g. in Poales (Ong and Palmer 2006) and in hornworts (Li *et al.* 2009, Xue *et al.* 2010) might help to identify a possible function. The nearly intact mitochondrial *rps*14 pseudogene in Poales is transcribed and sometimes even subject to RNA editing but does not contain an intron (Ong and Palmer 2006). Here, the transcript itself might be able to build a hairpin structure, allowing to form a double-stranded RNA. Ong and Palmer (2006) suggested that the co-transcription with the adjacent *rpl*5 gene protects the pseudogene from degradation. A co-transcription of the *nad*7 pseudogene with neighbouring *rps*10 (Takemura *et al.* 1995) and *trn*A (own unpubl. observation) was also detected. If a deletion of the pseudogene resulted in a failure to transcribe and/or translate the essential *rps*10 gene, the retention of the Ψ-*nad*7 would be easily explained. Whether a smaller transcript size inhibits the transcription and translation of the latter gene is, however, questionable. In the hornwort *Megaceros aenigmaticus*, two of the mitochondrial pseudogenes, Ψ-*sdh*3 and Ψ-*ccm*Fc, contain group II introns (Li *et al.* 2009). Until now nothing is known about the transcription of the pseudogenes in the hornworts. Thus, it is still possible that the pseudogenes in Poales and hornworts could also be maintained for gene regulation at the level of RNA as it is here suggested for

the liverwort *nad*7 pseudogene.

The mode of degradation of the *nad*7 pseudogene is different in jungermanniid and marchantiid liverworts (Groth-Malonek *et al.* 2007b). While in the Marchantiopsida predominantly substitutions were found resulting in the formation of several stop codons, frequent insertion and deletion events led to the disruption of the reading frame in the Jungermanniopsida (see Fig. 3.2 Groth-Malonek *et al.* 2007b). The former coding region of the *nad*7 pseudogene in the Marchantiopsida is obviously almost unaffected by insertions and deletions and less affected by degeneration compared to the pseudogene in Jungermanniopsida. The observed pattern could be attributed to a generally lower mutation rate in the mitochondrial DNA in marchantiid liverworts, also evident from highly conserved coding sequences and shorter branches in phylogenetic trees (Beckert *et al.* 1999, Forrest *et al.* 2006, Wahrmund *et al.* 2008).

As *nad*7 is a pseudogene in all jungermanniid and marchantiid liverworts investigated, a parsimonious scenario would suggest that the successful gene transfer to the nucleus occurred in a common ancestor of both classes. Independent transfers cannot be ruled out, however because of the varying sequences obtained for the mitochondrial target signal in the marchantiid *Marchantia* (Kobayashi et al., 1997) and the jungermanniid *Calycularia* (see 3.2.2). The presently lacking information, whether the intron found in the upstream part of the *Marchantia* nuclear gene is also present in *Lunularia* and *Calycularia* likewise does not clarify the issue of a single gene transfer event. Although such a one-time event appears more likely, independent transfers of the same gene to the nuclear genome and coupling with mitochondrial target signal sequence have been observed in angiosperms also for genes encoding subunits of the respiratory chain (Adams *et al.* 2000, 2001). Nevertheless, the *cox*2 gene has been transferred only once to the nuclear genome with nuclear and mitochondrial copies being silenced several times independently during legume evolution (Adams *et al.* 1999). In addition, the sequence differences in mitochondrial target regions could be explained by a higher mutation rate in the nuclear genome compared to the chondriome (Wolfe et al., 1987). As long as import into the mitochondria and functionality of the protein is ensured, the nucleotide sequence can be altered.

A functional mitochondrial *nad*7 gene with varying RNA editing frequencies in the basal-most Haplomitriopsida

In *Haplomitrium*, a taxon of the most basal liverwort clade Haplomitriopsida, the mitochondrial *nad*7 copy is transcribed and properly spliced and thus still functional (Groth-Malonek et al., 2007b). For *Treubia* and *Apotreubia*, the only other haplomitriid genera, the mitochondrial *nad*7 sequences also suggest functionality as neither stop codons nor frame shift mutations can be observed. Within the haplomitriid liverworts, the *nad*7 sequences differ, mainly by their amount of RNA editing needed to

recreate codon identities. While *Haplomitrium* shows a high RNA editing frequency, considerably less RNA editing is proposed for *Apotreubia* and none at all for *Treubia*. The tendency to decrease the RNA editing frequency is also seen in the Jungermanniopsida, with more editing in the basal taxa such as Fossombroniales and Metzgeriales and less in the more derived taxa like Porellales and Jungermanniales (unpubl. observation). As the haplomitriids are clearly the basal-most class in the liverworts, the common ancestor of all liverworts was also capable of RNA editing. Thus, the lack of RNA editing in the Marchantiopsida is a secondary loss of this ability (Groth-Malonek *et al.* 2007b) as it is probably in *Treubia*, because so far no RNA editing is required there as well. As *Haplomitrium*, possessing the highest known RNA editing frequency in liverworts, is the basal-most genus in the basal-most class, it can be assumed that the ancestor of all liverworts also displayed a high amount of RNA editing. However, most of its RNA editing sites could have been acquired after *Haplomitrium* split from the other haplomitriid liverworts. In liverworts as in mosses the hypothesis of an ancestor with considerably less RNA editing than found in *Takakia* and *Haplomitrium* cannot be ruled out. In fact, DYW-PPR protein genes, involved in RNA editing (Kotera *et al.* 2005, Rüdinger *et al.* 2008, Salone *et al.* 2007, Takenaka *et al.* 2008, Zehrmann *et al.* 2009), are numerous in *Haplomitrium* but rarely cluster together with DYW-PPR protein genes from other liverworts. Most of them form a separate clade suggesting an independent expansion of the gene family in *Haplomitrium* (M. Rüdinger, pers. comm.) and therefore an accumulation of RNA editing sites not seen in the other liverworts.

Conclusion

The generalization of slow mitochondrial evolution in liverworts based on two completely sequenced mitochondrial genomes (Wang *et al.* 2010) should be considered with some caution. The long retained pseudogene, that could be interpreted in favour of highly conserved chondriomes, might as well harbour an unknown function. In addition, RNA editing frequencies display a remarkable variability during liverwort evolution, with at least one complete loss of the ability for RNA editing in Marchantiopsida (Groth-Malonek *et al.* 2007b) and independent decreases in Haplomitriopsida and Jungermanniopsida. Studies on spacer regions also suggest that size variation due to insertions and deletions, sometimes even affecting genes (Groth-Malonek *et al.* 2007a, Wahrmund *et al.* 2008), occur much more frequently than expected by the comparison of the two chondriomes (Wang *et al.* 2010). As traces of (direct or inverted) repeat elements are found at the border of the supposed recombination points, pairing of these regions and subsequent inversion, deletion or insertion can be assumed. In tracheophytes, direct repeat regions are responsible for recombinational activity that forms the mitochondrial subcircles (e.g. Fauron *et al.* 1995, Grewe *et al.* 2009, Sugiyama *et al.* 2005). Although the formation of smaller subcircles has not been shown in bryophytes (Oda *et al.* 1992, Terasawa *et al.* 2007), the ability to alter the mitochondrial

structure is already present albeit not to such an extent as in vascular plants. Hence, the mitochondrial genomes of the earliest land plants already possess first prerequisites for gene-shuffling found much more pronounced in the tracheophytes.

5 Acknowledgements

First of all I sincerely thank Prof. Volker Knoop for letting me work in his research group, for the inspiring topic, for the challenges I faced and the chances given during the almost four years of study under his supervision. I also want to thank Prof. Jan-Peter Frahm for agreeing to co-review my PhD thesis and Prof. Thomas Bartolomaeus and Prof. Jens Léon for participating in my PhD commission.

I am very grateful to Monika Polsakiewicz for supervising and maintaining a well-organized lab, for her help and support whenever she was needed.

I have been very happy to work with my colleague and friend Mareike. She always shared the ups and downs experienced in doing research and helped more often than she would guess. I also much appreciate Henning's help and his counsel, especially with, but not limited to, electronic devices.

Michael, Felix, Simon, Stefan, Julia H., Julia D., Johanna, Daniel, Milena, Karo and all other current and former members of the Knoop's lab have made the last years exiting and full of fun and action. I enjoyed it and thank you all.

I am also thankful to all members of Prof. Frahm's group for bryophyte material and interesting and funny field trips.

To Micha, Mama, my families and friends I am deeply indebted for their constant love, friendship and support.

Bibliography

Adams, K. L., Song, K., Roessler, P. G., Nugent, J. M., Doyle, J. L., Doyle, J. J., and Palmer, J. D. (1999). Intracellular gene transfer in action: dual transcription and multiple silencings of nuclear and mitochondrial *cox*2 genes in legumes. *Proceedings of the National Academy of Sciences of the United States of America*, 96(24):13863–13868.

Adams, K. L., Daley, D. O., Qiu, Y. L., Whelan, J., and Palmer, J. D. (2000). Repeated, recent and diverse transfers of a mitochondrial gene to the nucleus in flowering plants. *Nature*, 408(6810):354–357.

Adams, K. L., Ong, H. C., and Palmer, J. D. (2001). Mitochondrial gene transfer in pieces: Fission of the ribosomal *rpl*2 and partial or complete transfer to the nucleus. *Molecular Biology and Evolution*, 18(12):2289–2297.

Adams, K. L., Qiu, Y. L., Stoutemyer, M., and Palmer, J. D. (2002). Punctuated evolution of mitochondrial gene content: high and variable rates of mitochondrial gene loss and transfer to the nucleus during angiosperm evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 99(15):9905–9912.

Anhut, S., Zinsmeister, H. D., Mues, R., Barz, W., Mackenbrock, K., Köster, J., and Markham, K. R. (1984). First identification of isoflavones from a bryophyte. *Phytochemistry*, 23(5):1073–1075.

Aravin, A. A., Hannon, G. J., and Brennecke, J. (2007). The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science*, 318(5851):761–764.

Asakawa, Y. (2004). Chemosystematics of the Hepaticae. *Phytochemistry*, 65:623–669.

Baldwin, B. G., Sanderson, M. J., Porter, J. M., Wojciechowski, M. F., Campbell, C. S., and Donoghue, M. J. (1995). The ITS region of nuclear ribosomal DNA: A valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden*, 82(2):247–277.

Bartel, D. P. (2007). MicroRNAs: Genomics, biogenesis, mechanism, and function. *Cell*, 131(4):11–29.

Beagley, C. T., Okimoto, R., and Wolstenholme, D. R. (1998). The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics*, 148(3):1091–1108.

Beckert, S., Steinhauser, S., Muhle, H., and Knoop, V. (1999). A molecular phylogeny of bryophytes based on nucleotide sequences of the mitochondrial *nad*5 gene. *Plant Systematics and Evolution*, 218:179–192.

Beckert, S., Muhle, H., Pruchner, D., and Knoop, V. (2001). The mitochondrial *nad*2 gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. *Molecular Phylogenetics and Evolution*, 18(1):117–126.

Bell, N. E. and Newton, A. E. (2005). The Paraphyly of *Hypnodendron* and the Phylogeny of Related Non-Hypnanaean Pleurocarpous Mosses Inferred from Chloroplast and Mitochondrial Sequence Data. *Systematic Botany*, 31(1):34–51.

Bell, N. E., Quandt, D., O'Brien, T. J., and Newton, A. E. (2007). Taxonomy and phylogeny in the earliest diverging pleurocarps: square holes and bifurcating pegs. *The Bryologist*, 110(3):533–560.

Bergthorsson, U., Adams, K. L., Thomason, B., and Palmer, J. D. (2003). Widespread horizontal transfer of mitochondrial genes in flowering plants. *Nature*, 424(6945):197–201.

Brotherus, V. F. (1924). Musci. In Engler, A. and Prantl, K., editors, *Die natürlichen Pflanzenfamilien*, volume 10, pages 143–478. Ducker and Humblot, Berlin, 2 edition.

Brotherus, V. F. (1925). Musci. In Engler, A. and Prantl, K., editors, *Die natürlichen Pflanzenfamilien*, volume 11, pages 1–542. Duncker and Humblot, Berlin, 2 edition.

Brown, J. W. S., Marshall, D. F., and Echeverria, M. (2008). Intronic noncoding RNAs and splicing. *Trends in Plant Science*, 13(7):335–342.

Buck, W. R. (1991). The basis for familial classification of pleurocarpous mosses. *Advances in Bryology*, 4:169–185.

Buck, W. R., Goffinet, B., and Shaw, A. J. (2000). Testing Morphological Concepts of Orders of Pleurocarpous Mosses (Bryophyta) Using Phylogenetic Reconstructions Based on TRNL-TRNF and RPS4 Sequences. *Molecular Phylogenetics and Evolution*, 16(2):180–198.

Budke, J. M., Jones, C. S., and Goffinet, B. (2007). Development of the enigmatic peristome of *Timmia megapolitana* (Timmiaceae; Bryophyta). *American Journal of Botany*, 94(3):460–467.

Burger, G., Yan, Y. F., Javadi, P., and Lang, B. F. (2009). Group I-intron trans-splicing and mRNA editing in the mitochondria of placozoan animals. *Trends in Genetics*, 25(9):381–386.

Chase, M. W., Soltis, D. E., Olmstead, R. G., Morgan, D., Les, D. H., Mishler, B. D., Duvall, M. R., Price, R. A., Hills, H. G., Qiu, Y. L., Kron, K. A., Rettig, J. H., Conti, E., Palmer, J. D., Manhart, J. R., Sytsma, K. J., Michaels, H. J., Kress, W. J., Karol, K. G., Clark, W. D., Hedren, M., Gaut, B. S., Jansen, R. K., Kim, K. J., Wimpee, C. F., Smith, J. F., Furnier, G. R., Strauss, S. H., Xiang, Q. Y., Plunkett, G. M., Soltis, P. S., Swensen, S. M., Williams, S. E., Gadek, P. A., Quinn, C. J., Eguiarte, L. E., Golenberg, E., Learn, G. H., Graham, S. W., Barrett, S. C. H., Dayanandan, S., and Albert, V. A. (1993). Phylogenetics of seed plants - An analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden*, 80(3):528–580.

Chaw, S. M., Chun-Chieh Shih, A., Wang, D., Wu, Y. W., Liu, S. M., and Chou, T. Y. (2008). The Mitochondrial Genome of the Gymnosperm *Cycas taitungensis* Contains a Novel Family of Short Interspersed Elements, Bpu Sequences, and Abundant RNA Editing Sites. *Molecular Biology and Evolution*, 25(3):603–615.

Chen, S. H., Habib, G., Yang, C. Y., Gu, Z. W., Lee, B. R., Weng, S. A., Silberman, S. R., Cai, S. J., Deslypere, J. P., Rosseneu, M., Gotto, A. M., Li, W. H., and Chan, L. (1987). Apolipoprotein B-48 Is the Product of A Messenger-Rna with An Organ-Specific In-Frame Stop Codon. *Science*, 238(4825):363–366.

Cho, Y., Qiu, Y. L., Kuhlman, P., and Palmer, J. D. (1998). Explosive invasion of plant mitochondria by a group I intron. *Proceedings of the National Academy of Sciences of the United States of America*, 95(24):14244–14249.

Cho, Y., Mower, J. P., Qiu, Y. L., and Palmer, J. D. (2004). Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*, 101(51):17741–17746.

Cox, C. J., Goffinet, B., Shaw, A. J., and Boles, S. B. (2004). Phylogenetic relationships among the mosses based on heterogeneous bayesian analysis of multiple genes from multiple genomic compartments. *Systematic Botany*, 29(2):234–250.

Crandall-Stotler, B. J., Forrest, L. L., and Stotler, R. E. (2005). Evolutionary trends in the simple thalloid liverworts (Marchantiophyta, Jungermanniopsida subclass Metzgeriidae). *Taxon*, 54(2):299–316.

Crandall-Stotler, B. J., Stotler, R. E., and Long, D. G. (2009). Phylogeny and classification of the Marchantiophyta. *Edinburgh Journal of Botany*, 66(1):155–198.

Davis, C. C., Anderson, W. R., and Wurdack, K. J. (2005). Gene transfer from a parasitic flowering plant to a fern. *Proceedings of the Royal Society B*, 272:2237–2242.

Davis, C. E. (2004). A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). *Monographs in Systematic Botany of the Missouri Botanic Garden*, 98:61–86.

Dixon, H. N. (1932). Classification of mosses. In Verdoorn, F., editor, *Manual of Bryology*, pages 397–412. Nijhoff, Hague.

Dombrovska, O. and Qiu, Y. L. (2004). Distribution of introns in the mitochondrial gene *nad*1 in land plants: phylogenetic and molecular evolutionary implications. *Molecular Phylogenetics and Evolution*, 32(1):246–263.

Fauron, C., Casper, M., Gao, Y., and Moore, B. (1995). The maize mitochondrial genome: dynamic, yet functional. *Trends in Genetics*, 11(6):228–235.

Filipowicz, W., Bhattacharyya, S. N., and Sonenberg, N. (2008). Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nature Reviews Genetics*, 9(2):102–114.

Fischer, S. (2009). Diplomarbeit: Molekulare Phylogenie und Evolution der Moniliformopses: Untersuchung mitochondrialer Gene zur Aufklärung ungelöster Fragen im Stammbaum der Farne. *Universität Bonn*.

Forrest, L. L., Davis, E. C., Long, D. G., Crandall-Stotler, B. J., Clark, A., and Hollingsworth, M. L. (2006). Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. *The Bryologist*, 109(3):303–334.

Garcia-Jacas, N., Susanna, A., Garnatje, T., and Vilatersana, R. (2001). Generic delimitation and phylogeny of the subtribe Centaureinae (Asteraceae): A combined nuclear and chloroplast DNA analysis. *Annals of Botany*, 87(4):503–515.

Giegé, L. and Brennicke, A. (1999). RNA editing in *Arabidopsis* mitochondria effects 441 C to U changes in ORFs. *Proceedings of the National Academy of Sciences of the United States of America*, 96(26):15324–15329.

Goffinet, B. and Buck, W. R. (2004). Systematics of the Bryophyta (Mosses): from molecules to a revised classification. In Goffinet, B., Hollowell, V., and Magill, R., editors, *Molecular Systematics of Bryophytes*, Monographs in Systematic Botany from the Missouri Botanical Garden, pages 205–239. Missouri Botanical Garden Press, St. Louis.

Goffinet, B., Buck, W. R., and Shaw, A. J. (2008). Morphology, anatomy, and classification of the Bryophyta. In Goffinet, B. and Shaw, A. J., editors, *Bryophyte Biology*, pages 55–138. Cambridge University Press, Cambridge.

Goffinet, B. and Cox, C. J. (2000). Phylogenetic relationships among basal-most arthrodontous mosses with special emphasis on the evolutionary significance of the Funariinae. *The Bryologist*, 103(2):212–232.

Goffinet, B., Cox, C. J., Shaw, A. J., and Hedderson, T. A. J. (2001). The Bryophyta (Mosses): Systematic and Evolutionary Inferences from an *rps*4 Gene (cpDNA) Phylogeny. *Annals of Botany*, 87(2):191–208.

Goffinet, B., Wickett, N. J., Shaw, A. J., and Cox, C. J. (2005). Phylogenetic significance of the *rpo*A loss in the chloroplast genome of mosses. *Taxon*, 54(2):353–360.

Goffinet, B., Wickett, N. J., Werner, O., Ros, R. M., Shaw, A. J., and Cox, C. J. (2007). Distribution and Phylogenetic Significance of the 71-kb Inversion in the Plastid Genome in Funariidae (Bryophyta). *Annals of Botany*, 99(4):747–753.

Grewe, F., Viehoever, P., Weisshaar, B., and Knoop, V. (2009). A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Research*, 37(15):5093–5104.

Groth-Malonek, M. and Knoop, V. (2005). Bryophytes and other basal land plants: the mitochondrial perspective. *Taxon*, 54(2):293–297.

Groth-Malonek, M., Pruchner, D., Grewe, F., and Knoop, V. (2005). Ancestors of trans-splicing mitochondrial introns support serial sister group relationships of hornworts and mosses with vascular plants. *Molecular Biology and Evolution*, 22(1):117–125.

Groth-Malonek, M., Rein, T., Wilson, R., Groth, H., Heinrichs, J., and Knoop, V. (2007a). Different Fates of Two Mitochondrial Gene Spacers in Early Land Plant Evolution. *International Journal of Plant Sciences*, 168(5):709–717.

Groth-Malonek, M., Wahrmund, U., Polsakiewicz, M., and Knoop, V. (2007b). Evolution of a Pseudogene: Exclusive Survival of a Functional Mitochondrial *nad*7 Gene Supports *Haplomitrium* as the Earliest Liverwort Lineage and Proposes a Secondary Loss of RNA Editing in Marchantiidae. *Molecular Biology and Evolution*, 24(4):1068–1074.

Handa, H. (2003). The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Research*, 31(20):5907–5916.

Hao, W. and Palmer, J. D. (2009). Fine-scale mergers of chloroplast and mitochondrial genes create functional, transcompartmentally chimeric mitochondrial genes. *Proceedings of the National Academy of Sciences of the United States of America*, 106(39):16728–16733.

Harpke, D. and Peterson, A. (2006). Non-concerted ITS evolution in *Mammillaria* (Cactaceae). *Molecular Phylogenetics and Evolution*, 41(3):579–593.

He, L. and Hannon, G. J. (2004). MicroRNAs: Small RNAs with a big role in gene regulation. *Nature Reviews Genetics*, 5(7):522–531.

Hedderson, T. A., Chapman, R. L., and Rootes, W. L. (1996). Phylogenetic relationships of bryophytes inferred from nuclear-encoded rRNA gene sequences. *Plant Systematics and Evolution*, 200(3-4):213–224.

Hedderson, T. A., Murray, D. J., Cox, C. J., and Nowell, T. L. (2004). Phylogenetic relationships of haplolepideous mosses (Dicranidae) inferred from *rps*4 gene sequences. *Systematic Botany*, 29(1):29–41.

Heinrichs, J., Gradstein, S. R., Wilson, R., and Schneider, H. (2005). Towards a natural classification of liverworts (Marchantiophyta) based on the chloroplast gene *rbc*L. *Cryptogamie Bryologie*, 26(2):215–233.

Heinrichs, J., Hentschel, J., Wilson, R., Feldberg, K., and Schneider, H. (2007). Evolution of leafy liverworts (Jungermanniidae, Marchantiophyta): estimating divergence times from chloroplast DNA sequences using penalized likelihood with integrated fossil evidence. *Taxon*, 56(1):31–44.

Hilu, K. W., Borsch, T., Muller, K., Soltis, D. E., Soltis, P. S., Savolainen, V., Chase, M. W., Powell, M. P., Alice, L. A., Evans, R., Sauquet, H., Neinhuis, C., Slotta, T. A. B., Rohwer, J. G., Campbell, C. S., and Chatrou, L. W. (2003). Angiosperm phylogeny based on *mat*K sequence information. *American Journal of Botany*, 90(12):1758–1776.

Howarth, D. G. and Baum, D. A. (2002). Phylogenetic utility of a nuclear intron from nitrate reductase for the study of closely related plant species. *Molecular Phylogenetics and Evolution*, 23:525–528.

Hyvönen, J., Koskinen, S., Merrill, G. L., Hedderson, T. A., and Stenroos, S. (2004). Phylogeny of the Polytrichales (Bryophyta) based on simultaneous analysis of molecular and morphological data. *Molecular Phylogenetics and Evolution*, 31(3):915–928.

Jones-Rhoades, M. W., Bartel, D. P., and Bartel, B. (2006). MicroRNAs and their regulatory roles in plants. *Annual Review of Plant Biology*, 57:19–53.

Jung, M., Zinsmeister, H. D., and Geiger, H. (1994). New three and tetraoxygenated coumarin glucosides from the mosses *Atrichum undulatum* and *Polytrichum formosum*. *Zeitschrift der Naturforscher*, 49(c):859–868.

Jung, M., Geiger, H., and Zinsmeister, H. D. (1995). Tri- and Tetrahydroxycoumarin derivatives from *Tetraphis pellucida*. *Phytochemistry*, 39(2):379–381.

Keeling, P. J. and Palmer, J. D. (2008). Horizontal gene transfer in eukaryotic evolution. *Nature Reviews Genetics*, 9(8):605–618.

Knoop, V. (2004). The mitochondrial DNA of land plants: peculiarities in phylogenetic perspective. *Current Genetics*, 46(3):123–139.

Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A., and Ohyama, K. (1997). Interorganellar gene transfer in bryophytes: the functional *nad*7 gene is nuclear encoded in *Marchantia polymorpha*. *Molecular and General Genetics*, 256(5):589–592.

Kotera, E., Tasaka, M., and Shikanai, T. (2005). A pentatricopeptide repeat protein is essential for RNA editing in chloroplasts. *Nature*, 433(7023):326–330.

Li, L., Wang, B., Liu, Y., and Qiu, Y. L. (2009). The Complete Mitochondrial Genome Sequence of the Hornwort *Megaceros aenigmaticus* Shows a Mixed Mode of Conservative Yet Dynamic Evolution in Early Land Plant Mitochondrial Genomes. *Journal of Molecular Evolution*, 68:665–678.

Lung, B., Zemann, A., Madej, M. J., Schuelke, M., Techritz, S., Ruf, S., Bock, R., and Hüttenhofer, A. (2006). Identification of small non-coding RNAs from mitochondria and chloroplasts. *Nucleic Acid Research*, 34(14):3942–3852.

Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., Lecharny, A., Le Ret, M., Martin-Magniette, M. L., Mireau, H., Peeters, N., Renou, J. P., Szurek, B., Taconnat, L., and Small, I. (2004). Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. *The Plant Cell*, 16(8):2089–2103.

Maas, S., Kawahara, Y., Tamburro, K. M., and Nishikura, K. (2009). A-to-I RNA editing and human disease. *RNA Biology*, 3(1):1–9.

Magombo, Z. L. K. (2003). The Phylogeny of Basal Peristomate Mosses: Evidence from cpDNA, and Implications for Peristome Evolution. *Systematic Botany*, 28(1):24–38.

Malek, O., Lattig, K., Hiesel, R., Brennicke, A., and Knoop, V. (1996). RNA editing in bryophytes and a molecular phylogeny of land plants. *EMBO Journal*, 15(6):1403–1411.

Malek, O. and Knoop, V. (1998). Trans-splicing group II introns in plant mitochondria: the complete set of cis-arranged homologs in ferns, fern allies, and a hornwort. *RNA*, 4(12):1599–1609.

Markham, K. R. and Porter, L. J. (1979). Flavonoids of the primitive liverwort *Takakia* and their taxonomic and phylogenetic significance. *Phytochemistry*, 18:611–615.

Markham, K. R. and Given, D. R. (1988). The major flavonoids of an antarctic *Bryum*. *Phytochemistry*, 27(9):2843–2845.

Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H., and Cerff, R. (1993). Moleuclar Phylogenies in Angiosperm Evolution. *Molecular Biology and Evolution*, 10(1):140–162.

Michel, F., Kazuhiko, U., and Haruo, O. (1989). Comparative and functional anatomy of group II catalytic introns – a review. *Gene*, 82(1):5–30.

Mower, J. P., Stefanovic, S., and Palmer, J. D. (2004). Gene transfer from parasitic to host plants. *Nature*, 432:165–166.

Newton, A. E., Cox, C. J., Duckett, J. G., Wheeler, J. A., Goffinet, B., Hedderson, T. A. J., and Mishler, B. D. (2000). Evolution of the major moss lineages: Phylogenetic analyses based on multiple gene sequences and morphology. *The Bryologist*, 103(2):187–211.

Nickrent, D. L. and Soltis, D. E. (1995). A Comparison of Angiosperm Phylogenies from Nuclear 18S rDNA and *rbc*L Sequences. *Annals of the Missouri Botanic Garden*, 82(2):208–234.

Nickrent, D. L., Parkinson, C. L., Palmer, J. D., and Duff, R. J. (2000). Multigene phylogeny of land plants with special reference to bryophytes and the earliest land plants. *Molecular Biology and Evolution*, 17(12):1885–1895.

Nickrent, D. L., Blarer, A., Qiu, Y. L., Vidal-Russell, R., and Anderson, F. (2004). Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. *BMC Evolutionary Biology*, 4(1):40.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. (2002). The complete sequence of rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA

sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics*, 268(4):434–445.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., Kanegae, T., Ogura, Y., and Kohchi, T. (1992). Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. *Journal of Molecular Biology*, 223(1):1–7.

Ogihara, Y., Yamazaki, Y., Murai, K., Kanno, A., Terachi, T., Shiina, T., Miyashita, N., Nasuda, S., Nakamura, C., Mori, N., Takumi, S., Murata, M., Futo, S., and Tsunewaki, K. (2005). Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. *Nucleic Acids Research*, 33(19):6235–6250.

Oldenburg, D. J. and Bendich, A. J. (1998). The structure of mitochondrial DNA from the liverwort, *Marchantia polymorpha*. *Journal of Molecular Biology*, 276(4):745–758.

Oldenburg, D. J. and Bendich, A. J. (2001). Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. *Journal of Molecular Biology*, 310(3):549–562.

Olmstead, R. G., Kim, K. J., Jansen, R. K., and Wagstaff, S. J. (2000). The Phylogeny of the Asteridae sensu lato Based on Chloroplast *ndh*F Gene Sequences. *Molecular Phylogenetics and Evolution*, 16(1):96–112.

Ong, H. and Palmer, J. D. (2006). Pervasive survival of expressed mitochondrial *rps*14 pseudogenes in grasses and their relatives for 80 million years following three functional transfers to the nucleus. *BMC Evolutionary Biology*, 6(1):55.

Palmer, J. D. and Herbon, L. A. (1988). Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. *Journal of Molecular Evolution*, 28(1-2):87–97.

Parkinson, C., Mower, J., Qiu, Y. L., Shirk, A., Song, K., Young, N., dePamphilis, C., and Palmer, J. D. (2005). Multiple major increases and decreases in mitochondrial substitution rates in the plant family Geraniaceae. *BMC Evolutionary Biology*, 5(1):73.

Peterson, A., Harpke, D., Peruzzi, L., Levichev, I. G., Tison, J. M., and Peterson, J. (2009). Hybridization drives speciation in *Gagea* (Liliaceae). *Plant Systematics and Evolution*, 278(3-4):133–148.

Powell, L. M., Wallis, S. C., Pease, R. J., Edwards, Y. H., Knott, T. J., and Scott, J. (1987). A Novel Form of Tissue-Specific RNA Processing Produces Apolipoprotein-B48 in Intestine. *Cell*, 50(6):831–840.

Pruchner, D., Nassal, B., Schindler, M., and Knoop, V. (2001). Mosses share mitochondrial group II introns with flowering plants, not with liverworts. *Molecular Genetics and Genomics*, 266(4):608–613.

Pruchner, D., Beckert, S., Muhle, H., and Knoop, V. (2002). Divergent intron conservation in the mitochondrial *nad*2 gene: signatures for the three bryophyte classes (mosses, liverworts, and hornworts) and the lycophytes. *Journal of Molecular Evolution*, 55(3):265–271.

Qiu, Y. L., Cho, Y., Cox, J. C., and Palmer, J. D. (1998). The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature*, 394(6694):671–673.

Qiu, Y. L., Dombrovska, O., Lee, J., Li, L., Whitlock, B. A., Bernasconi-Quadroni, F., Rest, J. S., Davis, C. C., Borsch, T., Hilu, K., Renner, S. S., Soltis, D. E., Soltis, P. S., Zanis, M. J., Cannone, J. J., Gutell, R. R., Powell, M., Savolainen, V., Chatrou, L. W., and Chase, M. W. (2005). Phylogenetic analyses of basal angiosperms based on nine plastid, mitochondrial, and nuclear genes. *International Journal of Plant Sciences*, 166(5):815–842.

Qiu, Y. L., Li, L., Wang, B., Chen, Z., Knoop, V., Groth-Malonek, M., Dombrovska, O., Lee, J., Kent, L., Rest, J., Estabrook, G. F., Hendry, T. A., Taylor, D. W., Testa, C. M., Ambros, M., Crandall-Stotler, B. J., Duff, R. J., Stech, M., Frey, W., Quandt, D., and Davis, C. C. (2006). The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42):15511–15516.

Rüdinger, M., Polsakiewicz, M., and Knoop, V. (2008). Organellar RNA editing and plant-specific extensions of pentatricopeptide repeat proteins in jungermanniid but not in marchantiid liverworts. *Molecular Biology and Evolution*, 25(7):1405–1414.

Rüdinger, M., Funk, H. T., Rensing, S. A., Maier, U. G., and Knoop, V. (2009). RNA editing: only eleven sites are present in the *Physcomitrella patens* mitochondrial transcriptome and a universal nomenclature proposal. *Molecular Genetics and Genomics*, 281(5):473–481.

Richardson, A. O. and Palmer, J. D. (2007). Horizontal gene transfer in plants. *Journal of Experimental Botany*, 58:1–9.

Rot, C., Goldfarb, I., Ilan, M., and Huchon, D. (2006). Putative cross-kingdom horizontal gene transfer in sponge (Porifera) mitochondria. *BMC Evolutionary Biology*, 6:71.

Salone, V., Rüdinger, M., Polsakiewicz, M., Hoffmann, B., Groth-Malonek, M., Szurek, B., Small, I., Knoop, V., and Lurin, C. (2007). A hypothesis on the identification of the editing enzyme in plant organelles. *FEBS Letters*, 581(22):4132–4138.

Samigullin, T. H., Valiejo-Roman, K. M., Troitsky, A. V., Bobrova, V. K., Filin, V. R., Martin, W., and Antonov, A. S. (1998). Sequences of rDNA internal transcribed spacers from the chloroplast DNA of 26 bryophytes: properties and phylogenetic utility. *FEBS Letters*, 422(1):47–51.

Sanchez-Puerta, M. V., Cho, Y., Mower, J. P., Alverson, A. J., and Palmer, J. D. (2008). Frequent, phylogenetically local horizontal transfer of the *cox*1 group I intron in flowering plant mitochondria. *Molecular Biology and Evolution*, 25(8):1762–1777.

Saritas, Y., Sonwa, M. M., Iznaguen, H., König, W. A., Muhle, H., and Mues, R. (2001). Volatile constituents in mosses (Musci). *Phytochemistry*, 57:443–457.

Schönenberger, J., Anderberg, A. A., and Sytsma, K. J. (2005). Molecular phylogenetics and patterns of floral evolution in the Ericales. *International Journal of Plant Sciences*, 166(2):265–288.

Schuster, R. M. (1966). Studies on Hepaticae XV. Calobryales. *Nova Hedwigia*, 13:1–63.

Seeburg, P. H., Higuchi, M., and Sprengel, R. (1998). RNA editing of brain glutamate receptor channels: mechanism and physiology. *Brain Research Reviews*, 26(2-3):217–229.

Shaw, J., Lickey, E. B., Beck, J. T., Farmer, S. B., Liu, W., Miller, J., Siripun, K. C., Winder, C. T., Schilling, E. E., and Small, R. L. (2005). The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis. *American Journal of Botany*, 92(1):142–166.

Shikanai, T. (2006). RNA editing in plant organelles: machinery, physiological function and evolution. *Cellular and Molecular Life Sciences*, 63(6):698–708.

Small, R. L., Ryburn, J. A., Cronn, R. C., Seelanan, T., and Wendel, J. F. (1998). The tortoise and the hare: Choosing between noncoding plastome and nuclear ADH sequences for phylogeny reconstruction in a recently diverged plant group. *American Journal of Botany*, 85(9):1301–1315.

Smith, D. K. and Davison, P. G. (1993). Antheridia and Sporophytes in *Takakia ceratophylla* (Mitt.) Grolle: Evidence for reclassification among mosses. *Journal of Hattori Botanical Laboratory*, 73:263–271.

Stech, M. and Frey, W. (2008). A morpho-molecular classification of the mosses (Bryophyta). *Nova Hedwigia*, 86(1-2):1–21.

Steinhauser, S., Beckert, S., Capesius, I., Malek, O., and Knoop, V. (1999). Plant mitochondrial RNA editing - extreme in hornworts and dividing the liverworts? *Journal of Molecular Evolution*, 48(3):303–312.

Sugita, M., Miyata, Y., Maruyama, K., Sugiura, C., Arikawa, T., and Higuchi, M. (2006). Extensive RNA Editing in Transcripts from the *Psb*B Operon and *Rpo*A Gene of Plastids from the Enigmatic Moss *Takakia lepidozioides*. *Bioscience, Biotechnology, and Biochemistry*, 70(9):2268–2274.

Sugiyama, Y., Watase, Y., Nagase, M., Makita, N., Yagura, S., Hirai, A., and Sugiura, M. (2005). The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. *Molecular Genetics and Genomics*, 272(6):603–615.

Taberlet, P., Gielly, L., Pautou, G., and Bouvet, J. (1991). Universal primers for amplification of three non-coding regions of chloroplast DNA. *Plant Molecular Biology*, 17:1105–1109.

Takemura, M., Nozato, N., Oda, K., Kobayashi, Y., Fukuzawa, H., and Ohyama, K. (1995). Active transcription of the pseudogene for subunit 7 of the NADH dehydrogenase in *Marchantia polymorpha* mitochondria. *Molecular and General Genetics*, 247(5):565–570.

Takenaka, M., Verbitskiy, D., van der Merwe, J. A., Zehrmann, A., and Brennicke, A. (2008). The process of RNA editing in plant mitochondria. *Mitochondrion*, 8(1):35–46.

Terasawa, K., Odahara, M., Kabeya, Y., Kikugawa, T., Sekine, Y., Fujiwara, M., and Sato, N. (2007). The Mitochondrial Genome of the Moss *Physcomitrella patens* Sheds New Light on Mitochondrial Evolution in Land Plants. *Molecular Biology and Evolution*, 24(3):699–709.

Tillich, M., Funk, H. T., Schmitz-Linneweber, C., Poltnigg, P., Sabater, B., Martin, M., and Maier, R. M. (2005). Editing of plastid RNA in *Arabidopsis thaliana* ecotypes. *The Plant Journal*, 43:708–715.

Turmel, M., Otis, C., and Lemieux, C. (2002). The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Molecular Biology and Evolution*, 19(4):585–585.

Turmel, M., Otis, C., and Lemieux, C. (2003). The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *The Plant Cell*, 15(8):1888–1903.

Unseld, M., Marienfeld, J. R., Brandt, P., and Brennicke, A. (1997). The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924 nucleotides. *Nature Genetics*, 15(1):57–61.

Valles, Y., Halanych, K. M., and Boore, J. L. (2008). Group II Introns Break New Boundaries: Presence in a Bilaterian's Genome. *PLoS ONE*, 1:e1488.

Vangerow, S., Teerkorn, T., and Knoop, V. (1999). Phylogenetic information in the mitochondrial *nad*5 gene of pteridophytes: RNA editing and intron sequences. *Plant Biology*, 1(2):235–243.

Virtanen, V. (2003). Phylogeny of the Bartramiaceae (Bryopsida) based on morphology and on *rbc*L, *rps*4, and *trn*L-*trn*F sequence data. *The Bryologist*, 106(2):280–296.

Vitt, D. H. (1984). Classification of Bryopsida. In Schuster, R. M., editor, *New Manual of Bryology*, volume 2, pages 696–759. Hattori Botanical Laboratory, Nichinan.

Volkmar, U. and Knoop, V. (subm). The lacking roots of mosses: discrepancies in chloroplast and mitochondrial data, including the novel intron locus cox1i624. *Journal of Molecular Evolution*.

Wahrmund, U., Groth-Malonek, M., and Knoop, V. (2008). Tracing plant mitochondrial DNA evolution: Rearrangements of the ancient mitochondrial gene cluster *trn*A-*trn*T-*nad*7 in liverwort phylogeny. *Journal of Molecular Evolution*, 66(6):621–629.

Wahrmund, U., Rein, T., Müller, K. F., Groth-Malonek, M., and Knoop, V. (2009). Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. *Plant Systematics and Evolution*, 282(3-4):241–255.

Wahrmund, U., Quandt, D., and Knoop, V. (2010). The phylogeny of mosses - Addressing open issues with a new mitochondrial locus: Group I intron cobi420. *Molecular Phylogenetics and Evolution*, 54(2):417–426.

Wang, B., Xue, J., Li, L., and Qiu, Y. L. (2010). The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Current Genetics*, (in press).

Wang, D., Wu, Y. W., Shih, A. C. C., Wu, C. S., Wang, Y. N., and Chaw, S. M. (2007). Transfer of Chloroplast Genomic DNA to Mitochondrial Genome Occurred At Least 300 MYA. *Molecular Biology and Evolution*, 24(9):2040–2048.

Ward, B. L., Anderson, R. S., and Bendich, A. J. (1981). The mitochondrial genome is large and variable in a family of plants (Cucurbitaceae). *Cell*, 25(3):793–803.

Wissemann, V. and Ritz, C. M. (2005). The genus *Rosa* (Rosoideae, Rosaceae) revisited: molecular analysis of nrITS1 and *atp*B-*rbc*L intergenic spacer (IGS) versus conventional taxonomy. *Botanical Journal of the Linnean Society*, 147(3):275–290.

Wolfe, K. H., Li, W. H., and Sharp, P. M. (1987). Rates of nucleotide substitution vary greatly among plant mitochondrial, chloroplast, and nuclear DNAs. *Proceedings of the National Academy of Sciences of the United States of America*, 84:9054–9058.

Won, H. and Renner, S. S. (2003). Horizontal gene transfer from flowering plants to *Gnetum*. *Proceedings of the National Academy of Sciences of the United States of America*, 100(19):10824–10829.

Xue, J. Y., Liu, Y., Li, L., Wang, B., and Qiu, Y. L. (2010). The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics*, (in press).

Yura, K., Miyata, Y., Arikawa, T., Higuchi, M., and Sugita, M. (2008). Characteristics and Prediction of RNA Editing Sites in Transcripts of the Moss *Takakia lepidozioides* Chloroplast. *DNA Research*, 15(5):309–321.

Zander, R. (1994). *Timmiella*. In Sharp, A., editor, *The moss flora of Mexico*, volume 69, pages 243–246. New York Botanical Garden, New York.

Zehrmann, A., Verbitskiy, D., Merwe, J. A. v. d., Brennicke, A., and Takenaka, M. (2009). A DYW Domain-Containing Pentatricopeptide Repeat Protein Is Required for RNA Editing at Multiple Sites in Mitochondria of *Arabidopsis thaliana*. *Plant Cell*, 21(2):558–567.