A mitochondrial perspective on early land plants: new loci in evolving chondriomes

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der Mathematisch-Naturwissenschaftlichen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Milena Groth-Malonek

aus

Halle / Saale

Bonn, d. 23. Januar 2006

1. Gutachter: Prof. Dr. V. Knoop

2. Gutachter: Prof. Dr. W. Barthlott

Tag der Prüfung: 25. April 2006

Diese Dissertation ist auf dem Hochschulschriftenserver der ULB Bonn
http://hss.ulb.uni-bonn.de/diss_online elektronisch publiziert

Erscheinungsjahr: 2006

Table of Contents

1 General Introduction: mitochondrial DNA and why we want to know

The earliest diversifications of land plants occurred up to 500 Mio years ago and are still unresolved. So far it is commonly accepted that bryophyte-like organisms were the first land plants (Kenrick and Crane 1997; Wellman et al. 2003). The green algae order Charales comprises the closest living relatives to these land plants (Malek et al. 1996; Karol et al. 2001; Turmel et al. 2003), but the relations between the three morphologically and genetically very different bryophyte groups - liverworts, mosses, and hornworts - and all tracheophytes are still widely debated. Recent studies based on molecular data support the relationship of liverworts as sisters to all other land plants (Qiu et al. 1998), and the rare hornworts as sisters to all tracheophytes (Knoop 2004; Groth-Malonek et al. 2005), therefore resolving bryophytes as a paraphyletic lineage (see also Mishler 1986; Mishler et al. 1994). However, another recent study proposed a monophyly of bryophytes, based on chloroplast protein sequences (Nishiyama et al. 2004).

Why is it so difficult to obtain a clear morphological or genetic picture of these relationships? Very likely the diversification of liverworts, mosses and hornworts into separate classes took place in only a short period of geological time, so that the backbone nodes of phylogenetic trees including all three classes are lying close together on relatively short internode branches and timescales. After their establishment, the three groups developed independently of each other with possibly fast radiations, and probably very different grades of differentiation and extinction rates. The worst problem to resolve these very old phylogenetic backbone events is the lack of informative, conserved characters which developed some 400 to 500 Mio years ago. Morphological features seem to provide controversial conclusions for that matter. Palaeobotany on the other hand provides abundant tracheophyte fossils (Forey et al. 2004), but lacks ancient bryophyte macrofossils. Only few mesofossils are discussed as potential liverworts, mosses, of hornworts (Edwards 2000; Kenrick 2003; Wellman et al. 2003). Consequently, the reconstruction of evolution through analyses of molecular

data is not only necessary but very likely the only way to resolve the problem. Every plant carries three genomes: the nuclear genome, the chloroplast genome (plastome), and the mitochondrial genome (chondriome). For the understanding of land plant evolution all three genomes are subject to close interrogations.

To obtain a reasonable theory for the phylogeny of all land plants, a major problem that has to be addressed are the different ages of the plant clades. A group aged as much as 450 mio years has to be compared to recently developed groups like the angiosperms (120 mio y). Direct comparison is only possible, if homologous regions can be analysed that have very slow rates of evolution, so that enough original information is still conserved. In this thesis the focus lies on the extension of data sets and development of new markers for the analysis of lower land plant relationships, with emphasis on "bryophytes", mostly liverworts. This group is very likely the oldest group of all land plants (Qiu et al. 1998), and therefore vital for the understanding of land plant evolution.

The is chondriome known as the most slowly evolving plant genome (Wolfe et al. 1987; Palmer and Herbon 1988) and therefore a promising candidate for the understanding of early land plant evolution. The chondriome of the liverwort *Marchantia polymorpha* was the first fully sequenced land plant chondriome (Oda et al. 1992a), followed by the angiosperm thale-cress *Arabidopsis thaliana* (Unseld et al. 1997). At present, only six further chondriomes are completely sequenced, all of them from angiosperms: sugar beet *Beta vulgaris* (Kubo et al. 2000), rapeseed *Brassica napus* (Handa 2003), rice *Oryza sativa* (Notsu et al. 2002), maize *Zea mays* (Clifton et al. 2004), wheat *Triticum aestivum* (Ogihara et al. 2005), and tobacco *Nicotiana tabacum* (Sugiyama et al. 2005). In addition, chondriome sequences of a few related algae are available, notably from the Charales alga *Chara vulgaris*, the probably closest living relative of the land plant lineage (Turmel et al. 2003), and *Chaetosphaeridium globosum*, a Coleochaetales alga (Turmel et al. 2002), candidate for a sister group to the Charales-land plant clade (Karol et al. 2001). The chondriome of *Chara* is 68 kb in size with 68 genes and 27 introns; *Chaetosphaeridium* encodes 67 genes and 11 introns within its 56 kb of mitochondrial DNA. In contrast to that, the *Arabidopsis* chondriome comprises only 57 genes including 23 introns in 367 kb (Unseld et al. 1997) , therefore increasing the amount of

"junk" DNA, but decreasing the number of genes (still) encoded on the mitochondrial genome. *Brassica napus*, like *Arabidopsis* a member of the Brassicaceae, has a chondriome of "only" 222 kb, therefore making it hard to find conclusive phylogenetic patterns in the reorganisation events of the plant mitochondrial DNA, which are probably mainly a result of frequent genomic recombination. The chondriomes of angiosperms in general are very variable in size and are known to be as large as 2.400 kb in the Cucurbitaceae (Ward et al. 1981).

Land plant chloroplast genomes (plastomes) are circular DNAs consisting of a Large Single Copy Region (LSC) and a Small Single Copy Region (SSC) which are separated by two Inverted Repeats (IR). The succession of the genes in LSC and SSC is highly conserved and structural changes occur usually as a varying expansion of the IR regions and therefore the duplication of some genes more or less, i.e. in the hornwort *Anthoceros formosae* (Kugita et al. 2003a). The transfer of genes to the nucleus, a frequent phenomenon in angiosperm mitochondria, is rarely observed for the plastome, the *rpoA* gene in *Physcomitrella patens* is an interesting exception (Sugiura et al. 2003). The overall sizes of the 21 fully sequenced land plant plastomes varies only slightly in the size range of 117 to 163 kb (Kim and Lee 2004), with the exception of the parasitic non-photosynthetic Orobanchaceae *Epifagus virginiana*, which contains a very reduced chloroplast genome of only 70 kb (Wolfe et al. 1992). Interestingly, organisation and gene content of algal plastomes can vary to a much higher degree, and reaches from 89 kb in *Codium fragile* (Manhart et al. 1989) to up to 1500 kb in *Acetabularia* (Simpson and Stern 2002).

The main differences of the two organelle genome types are the organisation of the genes and the size of the intergenic sequences. In fact, the "slow" evolution of the chondriomes is restricted to the very low mutation rate of protein coding exon regions. Its structural changes like disruptions of gene continuities (Palmer and Herbon 1988), and much higher mutation rates in non-coding DNA like introns or spacers between genes, could provide valuable tools for phylogenetic analyses on class and order level, as will be shown partially in this thesis.

Chapter 2 of this study refers to a newly established mitochondrial locus for phylogenetic analyses, the *nad4* gene. This gene was tested for its potential to resolve issues of basal land plant phylogeny on the example of the liverworts, probably the earliest diverging group of bryophytes. The secondary structure of the *nad4* group II intron conserved in liverworts is presented here with its folding pattern for the first time. A smaller part of the chapter is assigned to the study of the *nad4* homologues in mosses, hornworts and tracheophytes. The intron content and conservation pattern of *nad4* over all land plants could lead to further insights into the evolution and relationships of the major land plant groups.

Chapter 3 describes an extended study of the *nad5* gene that was already established as a phylogenetic marker locus in mosses and ferns, and became an effective phylogenetic tool in all lower land plants, including liverworts and hornworts. This study includes the sequencing of several newly analysed liverworts and the revision of the folding pattern of the group I intron included in this gene, which is a frequently sequenced locus shared by liverworts and mosses, and has been found to exhibit some unusual features in the liverwort genus *Pellia*, where it is singificantly smaller than in other liverworts. It also includes a correlation of the phylogenetic topology with a combined dataset derived from *nad5*, *nad4*, the cloroplast *rps4* and *rbcL* genes, partially derived from sequences obtained from public databases.

Chapter 4 is a study of the evolution of non-coding regions of the mitochondrial genome. This is an approach that has been used frequently on chloroplast or nuclear DNA but is completely new for plant mitochondrial DNA, which is known to present very few stable gene continuities in angiosperms. This aspect had not been investigated in non-tracheophytes so far. Two different gene clusters were analysed – the *nad5-nad4-nad2* and *trnA-trnT-nad7* regions – and revealed several interesting features shown for the first time in lower land plants, including the loss and possible regain of a *trnT* gene in one of the spacer regions.

Chapter 5 focuses on the evolution of *nad7*, a pseudogene restricted to liverworts and so far the only known land plant case of a highly conserved mitochondrial *nad* gene (of the

typical mitochondrially encoded group of *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9*) that underwent a transfer of the functional copy from the mitochondrial genome (chondriome) into the nucleus. The study of pseudogene development gives new insights into the mechanisms of unconstrained DNA evolution in different groups of liverworts and their relationship to each other.

All chapters are independent of each other and can principally be read in any order. They are accompanied by a general introduction into the special aspects of mitochondrial DNA and the phylogeny of lower land plants (Chapter 1), and followed by a synopsis (Chapter 6). Literature references for all parts can be found at the end of the thesis.

2 The mitochondrial *nad4* gene

2.1 Introduction

The mitochondrial *nad4* gene encodes subunit 4 of the NADH ubiquinone oxidoreductase, which is also known as complex I of the mitochondrial respiratory chain. All subunits are highly conserved in their amino acid sequence because of the vital importance of this protein complex. Most of the at least 34 subunits are encoded in the nuclear genome, but the subunit genes *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, *nad7*, and *nad9* are encoded on the chondriome of all land plants (Knoop 2004). The only exception known so far is the *nad7* gene of the complex thalloid liverwort *Marchantia polymorpha* (Oda et al. 1992a) (see chapter 5).

Some of these genes have been used for phylogenetic analyses, especially *nad5* (see Chapter 3) and *nad2* (Beckert et al. 2001). Both genes are disrupted through trans-spliced introns in angiosperms (Knoop 2004 and therein). In contrast to that, all of the three group II introns that occur in the *nad4* gene of angiosperms are cis-arranged, the two downstream introns are found to be occasionally and independently lost (Fig. 2-1). Only the generally conserved 5'-intron is universally present in mosses (Pruchner et al. 2001). In the absence of the angiosperm-type introns, an alternative group II intron is exclusively conserved in liverworts.

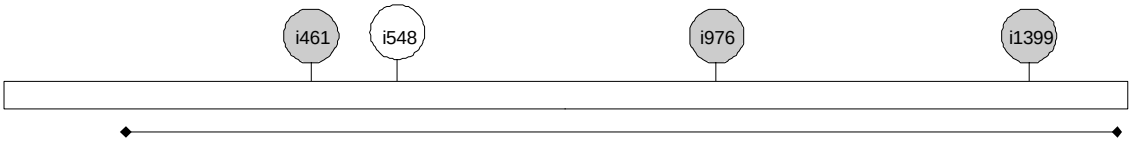

Fig. 2-1. Graphical overview of the *nad4* gene in land plants, adapted from Prucher et al. 2001. Shaded circles indicate angiosperm group II introns, the open circle designates the group II intron that is conserved in liverworts. The line beneath the graphic delineates the analysed region of the gene.

The usefulness of *nad4* as a phylogenetic marker was tested on the bryophyte group of liverworts, because liverworts were so far rather poorly sampled for phylogenetic studies based on mitochondrial data (Beckert et al. 1999). Other studies on liverworts included only data from chloroplast and nuclear genomes (e.g. Samigullin et al. 1998). During the last two years one other mitochondrial gene (*nad5*, see chapter 3) gained attention for phylogenetic studies and was included in the first large scale liverwort

phylogenies (Davis 2004; Forrest and Crandall-Stotler 2004; Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005). This gene includes a group I intron in liverworts, large parts of which were often excluded from the analyses.

2.2 Material and Methods

2.2.1 General methods and strategies

Fresh bryophyte material was collected by S. R. Gradstein, H. Groth, M. Groth-Malonek, J. Heinrichs, M. Lindner, Y.-L. Qiu, M. Schwertfeger, and M. Shimamura. Sterile culture specimens were obtained by courtesy of Prof. H. Becker, Saarbrücken. Fern DNAs were prepared from living plants from the Botanical Garden Bonn. Vouchers are deposited in the herbarium of the Dept. of Molecular Evolution, IZMB, University Bonn and/or in the Herbarium Goettingen (GOET) (Table 2-3). Additional DNA was prepared in the former Knoop laboratory in Ulm. Additional sequences that were included for the analyses are listed with their respective accession number from Genbank (NCBI), outgroup data for the mosses are given in chapter 2.3.2, table 2-3. Total nucleic acids were extracted from green plant material in the presence of cetyl-trimethyl-ammonium-bromide (CTAB) (modified after Doyle and Doyle 1990). PCR amplification assays contained 1 µl template DNA or cDNA (approximately 10 ng – 0.5 µg), 1 unit Taq-DNA-Polymerase (Genaxxon) or Silverstar-Taq (Eurogentec), 5 µl corresponding 10x PCR buffer, 2-3 mM $MgCl_2$, 200 µM dNTPs each, 0.2 mM of each primer, 2-4 % DMSO, 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: 92 °C 1 min, 57 °C to 50 °C for 1 min, 72 °C for 2 min, followed by 30 cycles of 92 °C 1 min, 50 °C for 1 min, 72 °C for 2 min – 2.5 min, and a final step of synthesis for 15 min at 72 °C. Primers used for the DNA assays are given in table 2-2. PCR-fragments were sequenced directly on an ABI 3100 capillary sequencer using the BigDye ™ Terminator Cycle Sequencing v2.0 kit (PE Biosystems), or cloned into the pGEM-T Easy vector (Promega) and sequenced on an ALF Express II (Amersham Biosciences) using the Sequenase Cy5 dye Terminator kit or the Flourescent Labelled Primer Cycle Sequencing kit (Amersham Biosciences), or were commercially sequenced (Macrogen Inc., Korea).

Sequences were aligned with BioEdit 7.0.1 (Hall 1999) and MEGA3 (Kumar et al. 2004), using the implemented Clustal algorithm, and manually adjusted. Graphics were designed with OpenOffice 2.0 (www.openoffice.org, Sun Microsystems Inc.) and MEGA3.

To obtain a phylogenetic tree, at first the best fitting evolution model for the data set was estimated by Modeltest 3.7 (Posada and Crandall 1998), with the preference of the implemented AIC (Akaike Information Criterion) over the also included hLRTs (hierarchical Likelihood Ratio Test) for the final choice of the most appropriate model, following the recommendation of Posada and Buckley (Posada and Buckley 2004).

Phylogenetic analyses were carried out by Bayesian Inference approach: implementation in MrBayes 3.1 (Huelsenbeck and Ronquist 2001) with the following parameters: all partitions unlinked, two independent runs with: four parallel chains, one heated chain, 1.000.000 generations, every 100th tree sampled, burnin set to 6.000 trees, which was estimated after the conversion of the two independent runs, model parameters see table 2-1.

Table 2-1: Models implemented in MrBayes for liverwort phylogeny based on the *nad4* gene

character set (partition)	model selected by AIC (modeltest 3.1)	parameters implemented in MrBayes
exons *nad4*	GTR+I+G	revmatpr = fixed(2.0757, 3.7931, 0.3069, 0.9919, 11.3371, 1.0000) statefreqpr = fixed(0.2585, 0.2132, 0.1959, 0.3324), shapepr= exponential(0.8181) pinvarpr = fixed(0.2145), ratepr= variable, nst = 6, rates = gamma
intron *nad4*	TVM+G	revmatpr = fixed(0.9545, 2.4145, 0.3375, 0.8789, 2.4145, 1.0000) statefreqpr = fixed(0.2455, 0.2434, 0.3026, 0.2085), shapepr= exponential(0.7358) pinvarpr = fixed(0), ratepr= variable, nst = 6, rates = gamma

2.2.2 The phylogenetic study on *nad4* in liverworts: Outgroup selection

The selection of an outgroup for the phylogenetic study on liverworts was tested with Charophyte algae, the closest ancestors to land plants: *Chara vulgaris* in combination or without *Chaetosphaeridium globosum*. The *nad4* gene of *Chaeotsphaeridium* is intronless, whereas *Chara* carries one group II intron that is a positional homologue to the angiosperm intron nad4i976, which is not present in liverworts. This intron sequence was excluded from the analysis, but, apart from the good statistical separation of the algal outgroup, several ingroup clusters were only weakly supported. This effect was even more pronounced when both algae were included in the study. On the other hand, mosses are considered to be an early diverging land plant group like liverworts, and have been used as alternative outgroup in other phylogenetic liverwort studies, although these studies analyse different loci for their phylogenies (Davis 2004; Forrest and Crandall-Stotler 2004). Mosses carry a single group II intron in the *nad4* gene that is a homologue to the angiosperm intron nad4i461, which is absent in liverworts. These intron sequences were excluded from the dataset prior to phylogenetic analysis, resulting in the selection of four intronless mosses as outgroups for the presented study.
Moss sequences represent a suitable outgroup for the exon analysis, but a large part of the liverwort dataset is comprised of the group II intron nad4i548. No possible outgroup for the liverwort intron could be identified, as this intron is unique for liverworts in all land plants and Charophytes analysed, and the most similar introns from the fully sequenced chondriome of the liverwort *Marchantia*, nad7i336 and rpl2i28, were not well alignable and gave no suitable results as an outgroup, and could also add an artificial bias towards the Marchantiopsida as the group closest to the selected outgroup intron.

Table 2-2: Primers for PCR assays of the land plant mitochondrial *nad4* gene

primer designation and location	primer sequence	location on the *Marchantia polymorpha* chondriome
nad4		from 6469 to 8855
nad4upliver	agg aag cct tat tat ttt ggt gat cc	6531…6556
n4up	aca gcc aaa ttt car ttt gtg gaa	6655…6678
n4upv2	aaa ttt car ttt gtg gaa ann ntt cga tgg ctt cc	6661…6695
n4uphw1	aay atc aat ttt tat wtr ggt ata gay gg	6703…6731
n4MOSSi1up	ctt tca tga ttg ctg trt tty gc	6851…6873
n4i+	att att ata ggn gtw tgg ggt tcy	6931…6958
n4i-	gtc trg aac ccc awa cnc cta taa	6931…6958
n4MOSSi1do	ata ttt gta rat cag tgg ttc ctg	7951…7974
n4+hw	tta tta acc aca gaa ttt agt gag	7974…8000
n4-hw	ccr ctc act aaa ttc tgt ggt taa	7974…8000
n4i23-.cy5	aat att tgg cgc cgc tca cta aat tct g	7984…8012
n4+1	gts aaa gtg cct atg gta cca gt	8043…8065
n4+1v2	gts aaa gtg cct atg gta cca gtt cat att tgg	8043…8075
n4-1	gtc gct tca gga aac atg gg	8175…8194
n4-1v2	aac ata cca ata gtc acn nna ttc ata tga gct ac	8304…8338
n4-1v2kurz	aca tac caa tag tca caa aat tca aat gg	8309…8337
n4+2hw	aac ata cag gga att gra ggt agc at	8346…8371
n4uphw2	agc agc ttt atc ggg gaa ttt cty	8550…8575
n4-2	tam gcs gcg cct aaa atc atc cc	8628…8650
n4dohw1	cca aaa acc aca cga tta tat arc c	8659…8683
n4dov2	tcc atg ttg cac taa gtt act tac gga ngt atg cat	8808…8843
n4do	tya ats aaa ttt tcc atg ttg cac	8832…8855
intron n4i461		between 6929 and 6930
n4i1+	ggg tag tct tgt gtg taa gca tag	approx. 120 bp from 5'-end of intron
n4i1-	ctg tag gta ccc act ccc ttc tc	approx. 70 bp from 3'-end of intron
intron n4i548		between 7014 and 7916
nad4i2+	gca tgg ggt gtt cta tgt aaa gc	7067…7089
n4mittei2do	ccc tta gca gaa tca tgt ccg t	7514…7535
nad4i2-	aac ctc aac tac cca ata aaa cc	7876…7898
intron n4i976		between 8343 and 8344
n4i976up	gca gca cgg ctc tac gga g	approx. 290 bp from 5'-end of intron
n4i976dov1 (A)	ccc ata ttc tga aac gaa ggc a	approx. 360 bp from 3'-end of intron
n4i976dov2 (D)	cga ata gga ttg tgc cgt caa tgg	approx. 340 bp from 3'-end of intron
intron n4i1399	no internal primers available	between 8766 and 8767

Table 2-3: Taxa used for the phylogenetic analyses of the mitochondrial *nad4* gene in liverworts

Phylogeny	Species	collection number	Accession number and sequence length
Blasiopsida	*Blasia pusilla* L.	J. Heinrichs 2291	2254 bp
Haplomitriopsida	*Haplomitrium mnioides* (Lindb.) Schust.	M. Shimamura s.n.	2208 bp
Jungermanniopsida	*Frullania tamarisci* (L.) Dumort.	J. Heinrichs 4382	2133 bp
leafy I / Porellales	*Lejeunea cavifolia* (Ehrh.) Lindb.	Ulm-collection s.n.	2247 bp
	Lepidolaena hodgsoniae Grolle	MGM031218-02SC	2265 bp
	Porella platyphylla (L.) Pfeiff.	J. Heinrichs 4383	2315 bp
	Ptilidium pulcherrimum (G.Web) Vainio	Heinrichs & Gradstein 4395	2231 bp
	Radula complanata (L.) Dum.	MGM031218-14SC	2301 bp
leafy II / Jungermanniales	*Anthelia julacea* (L.) Dumort.	J. Heinrichs s.n.	2263 bp
	Bazzania trilobata (L.) Gray	Ulm-collection s.n.	AJ310800, 2157 bp
	Calypogeia muelleriana (Schiffner) K. Müller	J. Heinrichs 4375	2243 bp
	Diplophyllum albicans (L.) Dumort.	J. Heinrichs 4371	2174 bp
	Gymnomitrion concinnatum (Lightf.) Corda	J. Heinrichs 4394	2322 bp
	Harpanthus flotovianus (Nees) Nees	J. Heinrichs 4390	2276 bp
	Herbertus sendtneri (Nees) Lindb.	J. Heinrichs 4377	2202 bp
	Jamesoniella autumnalis (DC.) Steph.	Ulm-collection s.n.	2192 bp
	Lophocolea cuspidata	MGM	2323 bp
	Mylia taylorii (Hook.) Gray	J. Heinrichs 4387	2146 bp
	Nardia scalaris Gray	J. Heinrichs 4389	2225 bp
	Plagiochila asplenioides (L.) Dumort.	J. Heinrichs & H. Groth 4369	2134 bp
	Scapania nemorea (L.) Grolle	J. Heinrichs 4372	2206 bp
	Trichocolea tomentella (Ehrh.) Dumort.	MGM031218-03SC	2231 bp
	Tritomaria quinquedentata (Huds.) H.Buch	J. Heinrichs 4381	2188 bp
simple thalloids I / Fossombroniales	*Fossombronia alaskana* Steere & Inoue	MGM031218-07SC	2253 bp
	Fossombronia pusilla (L.) Nees	Ulm-collection s.n.	2232 bp
	Noteroclada confluens Taylor ex Hook. & Wilson	live culture Goettingen	2240 bp
	Pellia endiviifolia (Dicks.) Dum.	MGM031218-12SC	1730 bp
	Symphyogyna brasiliensis	live culture Goettingen	2240 bp
	Symphyogyna brogniartii	SC	2321 bp
simple thalloids II / Metzgeriales	*Aneura pinguis* (L.) Dumort.	MGM031218-01SC	2182 bp
	Apometzgeria spec.	Ulm-collection s.n.	2308 bp
	Metzgeria furcata (L.) Dumort.	J. Heinrichs 4384	2304 bp
Marchantiopsida / complex thalloids	*Asterella blumeana* (Nees) Pandé Srivastava et Khan.	MGM031218-06SC	2265 bp
	Bucegia romanica Radian	Ulm-collection s.n.	2313 bp
	Conocephalum conicum (L.) Underw.	Groth & Schwertfeger s.n.	2203 bp
	Corsinia coriandrina (Spreng.) Lindb.	Ulm-collection s.n.	AJ310801, 2151 bp
	Lunularia cruciata (L.) Dum. ex Lindb.	Groth & Schwertfeger s.n.	AJ310803, 2151 bp
	Marchantia polymorpha L.	--	NC 001660, 2387 bp
	Monoclea gottschei Lindb.	live culture Goettingen	2223 bp
	Monosolenium tenerum Griff./Sunita Kapila & SS Kumar	live culture Goettingen	2258 bp
	Oxymitra incrassata (Brotero) Sérgio & Sim-Sim	MGM031218-11SC	2295 bp
	Reboulia hemisphaerica (L.) Raddi	MGM031218-04SC	2172 bp
	Riccia breidleri Steph.	ML-030826	2268 bp
	Riccia fluitans L.	Ulm-collection s.n.	AJ310802, 2155 bp
	Ricciocarpos natans (L.) Corda	MGM031218-05SC	2265 bp
	Riella spec.	Ulm-collection s.n.	2086 bp
	Sphaerocarpos donnellii Aust.	Ulm-collection s.n.	2210 bp
	Targionia hypophylla L.	Ulm-collection s.n.	2124 bp
Treubiopsida	*Apotreubia nana* (S. Hatt. & Inoue) S. Hatt. & Mizut.	LF198/ Long 30451	2263 bp
	Treubia lacunosa (Colenso) Prosk. lenta Taylor ex Prosk.	LF28/Stotler&Crandall-Stotler 4561 (ABSH)	2318 bp
	Treubia pygmea	LF30/Stotler&Crandall-Stotler 4582 (ABSH)	2290 bp

s.n. = sine numero (lat.), “without number”, meaning here without explicit collection or voucher number

2.3 Results and Discussion

2.3.1 The *nad4* gene in liverworts

The *nad4* gene in all liverworts investigated was found to carry only one universally conserved intron (Fig. 2-2). This intron, nad4i548, has already been described as conserved in liverworts, but absent in mosses (Pruchner et al. 2001). Its folding pattern (Fig. 2-3) reveals the structure of a group II intron (Michel et al. 1989).

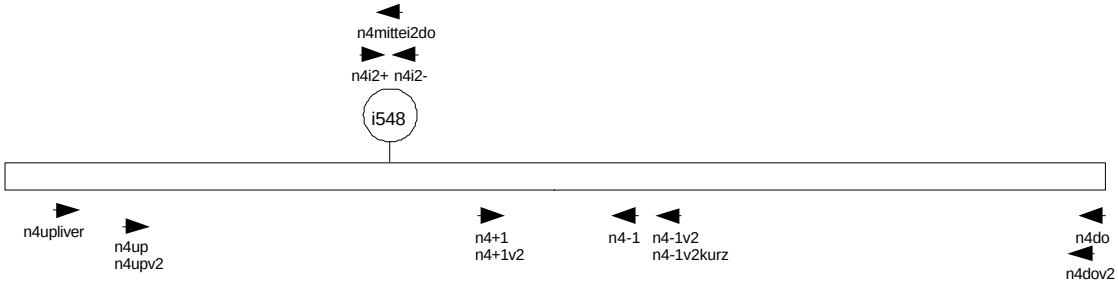

Fig. 2-2: Overview of the mitochondrial *nad4* gene in liverworts. The group II intron is indicated by a pinhead. Intron designation relies on the nucleotide position in the *nad4* gene of the reference liverwort *Marchantia polymorpha* after which the intron is inserted. Primer locations and directions are symbolized by arrows. Primers underneath the bar are located in exon regions, primers above the bar are located in the intron.

2.3.1.1 Structure and conservation of the group II intron nad4i548

Group II introns are catalytic RNAs which are capable of excising themselves out of pre-mRNA (splicing). Several of them encode the ORF (open reading frame) of a Reverse Transcriptase (RT) which catalyzes the splicing process (maturase activity). The RT is also necessary for the retroelement activity of several introns, the ability to reinsert the spliced intron into another region of the genome.

Organellar group II introns in land plants are usually located in highly conserved genes like *nad5* or *cox1* in mitochondria, or *petD* and *trnT* in chloroplasts. They tend to lose the RT encoding ORFs, an effect that is not exclusive to angiosperms but was already noted in mosses (Dombrovska and Qiu 2004). In the single completely sequenced chondriome of an early land plant, the liverwort *Marchantia polymorpha*, only 9 out of 24 group II introns carry an ORF, and all of them are restricted in their appearance to liverworts only (Turmel et al. 2003 and therein). In contrast to that, only one intact ORF remains in angiosperms (mat-R), in the intron nad1i725. Interestingly, only one of all mitochondrial group II introns is shared between *Marchantia* and angiosperms

(nad2i718), and only one group II intron, nad3i152, is shared by *Marchantia* and the green alga *Chara*, which still carries an intact ORF in this alga, but not in the liverwort.

The present study confirms the occurrence of the intron nad4i548 (Fig. 2-3) as restricted to liverworts. It is composed of all six domains typical for group II introns (e. g. Michel and Dujon 1983; Michel et al. 1989; Robart and Zimmerly 2005) including the hairpin structure of domain V that is highly conserved and used for intron identification (Knoop et al. 1994). It does not carry an ORF in its domain IV.

The presented folding pattern has been extrapolated from the alignment of 50 liverworts (see table 2-3 for the species names). Structural features of domain I assign it to the subgroup A1, which is the usual type for plant mitochondrial introns (Toor et al. 2001).

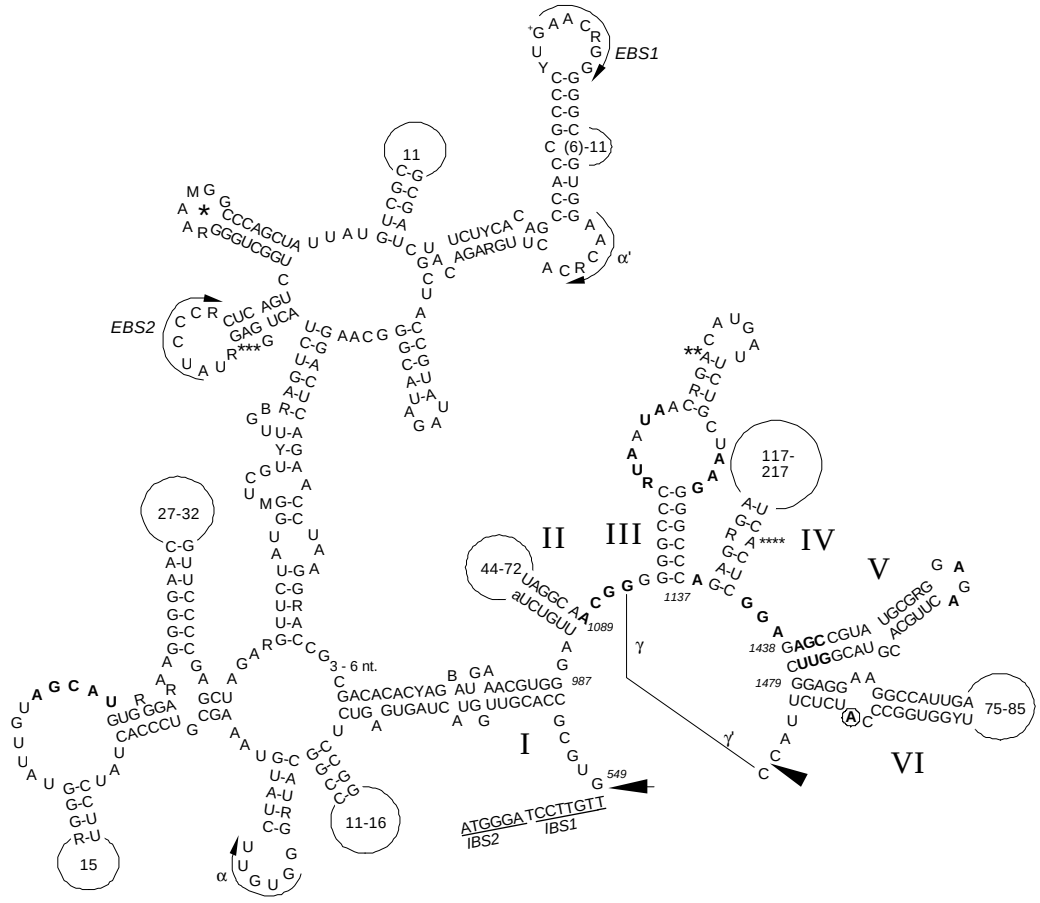

Fig. 2-3. Folding of group II intron (subgroup IIA1) nad4i548, unique for liverworts. Arrows indicate beginning and end of the intron. Roman numbers indicate the six intron domains. * loop is inverted in *Monoclea* and *Oxymitra*: CCTTTT; ** insertion of ACGGA in *Treubia*, *** insertion of GAG in *Oxymitra*, **** insertion of ACC in *Frullania* and *Lejeunea*. EBS: exon binding site, IBS: intron binding site; γ–γ' indicates a tertiary single-base-pair interaction, $^{+}$G near EBS1 indicates δ–δ' pairing with the first nucleotide of the 3'-exon.

The size variation of the intron is very small in complex thalloid liverworts, with a median size of 899 bp, ranging from 894 to 903 bp. The largest liverwort group, the leafy liverworts, have the same median size, but a deviation from 873 to 912 bp, with the exception of *Frullania*, which has a singular deletion in the loop of domain IV that reduces the intron size to 818 bp. The most variable size is visible in simple thalloid liverworts, with a size range from 845 bp to 925 bp around a median of 892 bp. One exception from this is *Symphyogyna brasiliensis* with only 826 bp, mostly due to two deletions in the domain IV. Both deletions are not shared by its sister species *Symphyogyna brogniartii*.

Of the 51 analysed liverworts only *Pellia endiviifolia* exhibits a series of major differences in the folding pattern of the intron structure. This species is a member of the simple thalloid liverworts, the most variable group; an assignment that is not contested by morphological or molecular studies. Its intron is only 660 bp long, which is more than 200 base pairs shorter than the average size (Fig. 2-4). This is mostly due to reduced loop sizes of domains II, IV and VI. In addition to that, some parts of the remnants of these loops are different to the sequences of other liverworts and therefore difficult to align. There are three regions in domain I where nucleotides are missing from non-loop structure elements (shaded gaps in figure 2-4). Two nucleotides are missing from domain VI, only six bases upstream of the A nucleotide necessary for the lariat formation of the intron during the splicing process. Six of these deletions are located in the forth subdomain of domain I, which carries the exon binding sites EBS1 and EBS2. One of the deletions is located in the main stem of the subdomain, leading to an alternative folding pattern for this stem as shown in figure 2-4. Combined with another deletion and the insertion of three bases in the core circle of the subdomain these findings suggest a different stereometry of the affected subdomain, mostly in the angles of the hairpin structures around the core circle.

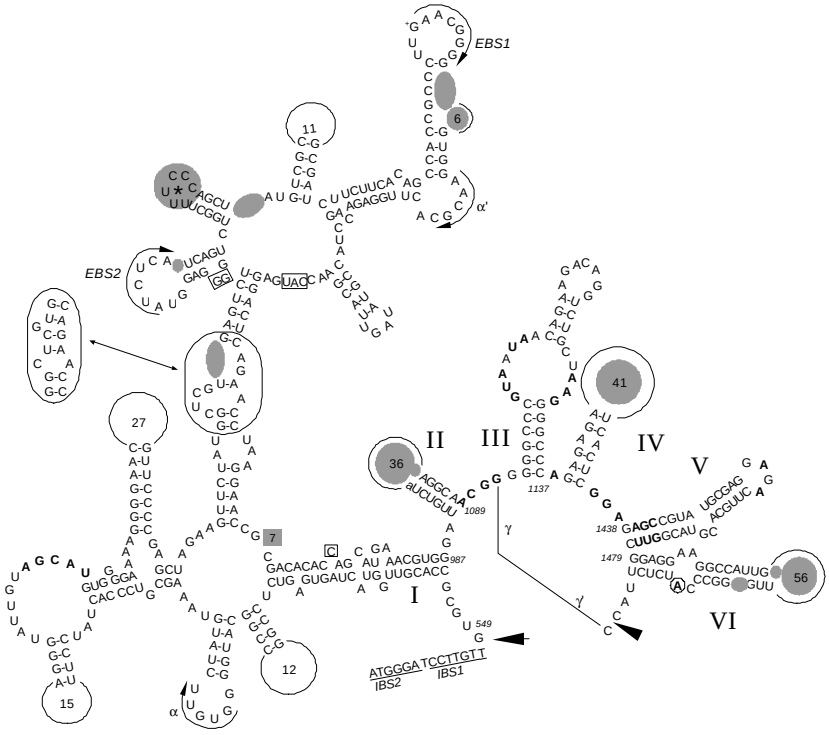

Fig. 2-4: Folding of group II intron nad4i548 for *Pellia endiviifolia*. Shaded regions depict deletions of the *Pellia* intron relative to the liverwort intron nad4i548 consensus pattern in figure 2-2. Boxed bases in domain I are: C → mismatch, putative editing site, GG → mismatch, UAC → insertion. Asterisk: loss of the original loop, replacement by mutated flanking bases.

The impact of these differences on *Pellia* is not known, because no cDNA sequences for direct comparison are available. As no ORF is included in domain IV of this intron (like in all liverworts), the loss of large parts of this domain should be rather indifferent. Provided that the gene analysed here is indeed the only copy of *nad4* encoded on the genomes of *Pellia*, it is likely that the intron is spliced normally, because *nad4* exon sequences seem not affected. It could also be possible, that a second copy of the gene, including a "normal" intron, is encoded either on the chondriome or in the nuclear genome. The only known land plant case of a second copy of a *nad*-gene is the example of *nad7* in the liverwort *Marchantia polymorpha*, which is a pseudogene on the chondriome, but has a functional copy in the nuclear genome (see also chapter 5). In that case, no mitochondrial intron but a nuclear intron was identified in the nuclear copy of *nad7*, and the nucleotide sequence of the two gene versions were clearly distinguishable. Even though no evidence for multiple gene-versions or pseudogenes has been found in *Pellia*, this possibility can not be ruled out.

Interestingly, this reduced size of introns is also prominent in *Pellia* sequences of the *nad5* intron nad5i753, which is a group I intron (see chapter 3), and probably also in the group II intron of the mitochondrial *trnS* which has been only partially sequenced (Davis 2004). Possibly, this is a pattern that extends to all mitochondrial introns in *Pellia*. However, it definitely does not extend to the next closely related genus *Noteroclada* that was also sequenced in this study. A study of the chloroplast *trnK* intron including *Pellia* (Hausner et al. 2006) noted only two mispaired stem nucleotides in domain I, no large indels are mentioned. As no other chloroplast intron sequences are available from Genbank, no general point can be made about *Pellia* introns here, but it should be analysed whether the "reduced size effect" is restricted to mitochondrial introns. To gain knowledge about the extension of the phenomenon, more intron-containing genes and more species should be analysed including different taxonomic ranks. As only two genera are described from the Pelliaceae, *Pellia* and *Noteroclada*, other species of the genus *Pellia* should be investigated for further studies.

2.3.1.2 RNA editing in liverworts: studies on *nad4*

DNA sequences do not always mirror the amino acid composition of the encoded protein. Differences are due to the splicing of introns and another very important mechanism, the RNA editing. The kind of editing that is typical for land plant organelles is the change of C to U (see Fig. 2-5), probably established through a deamination reaction. It occurs in all major land plant groups in differing degrees, with the exception of one group of liverworts, the complex thalloids (Steinhauser et al. 1999).

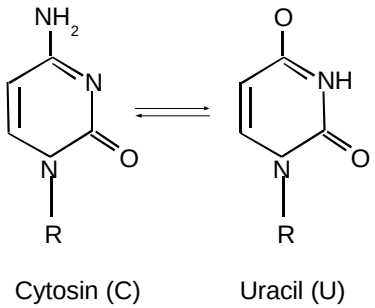

Fig. 2-5: Chemical structure of the nucleotide bases Cytosin and Uracil. RNA editing in land plants is based on the exchange of these two bases.

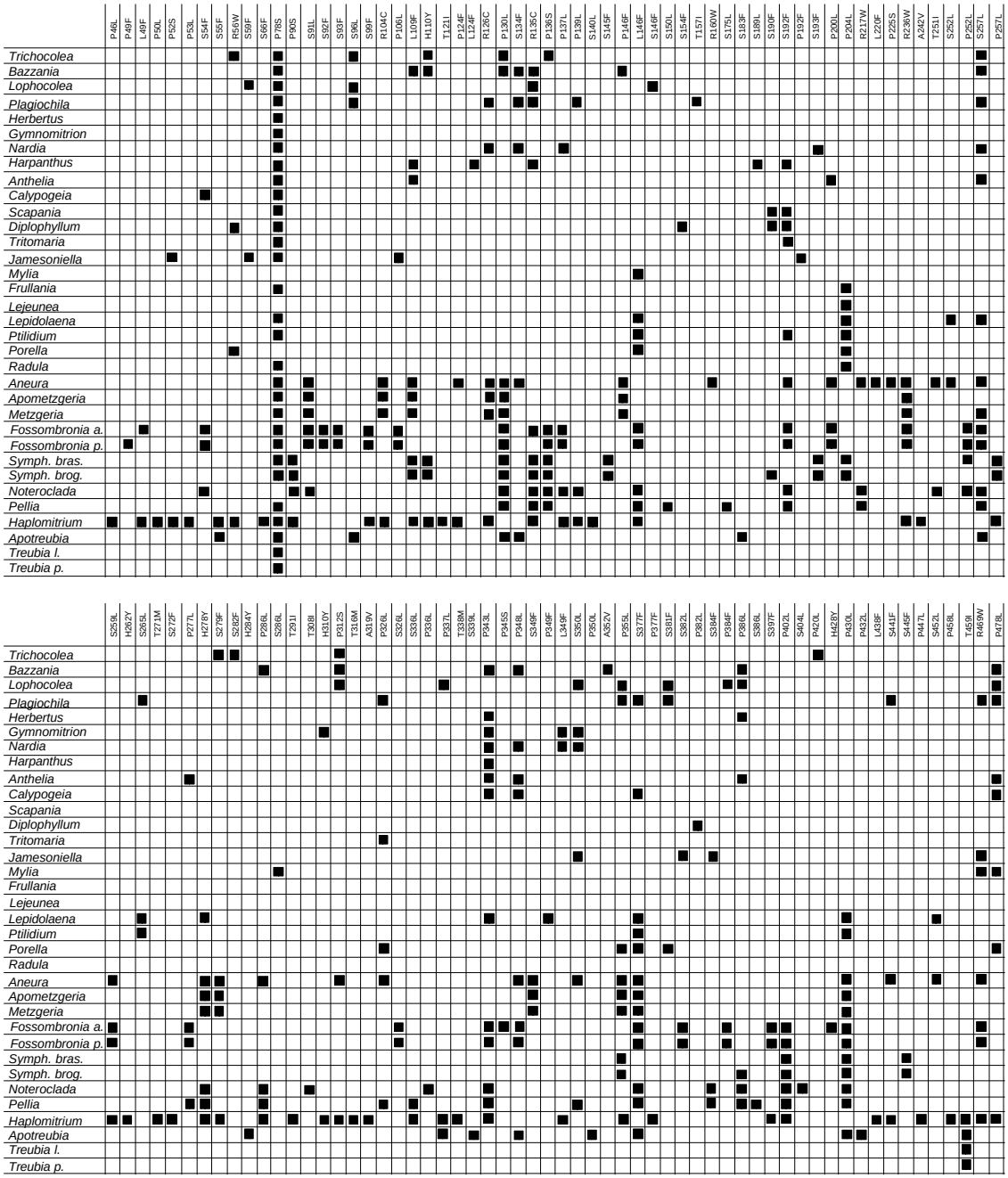

Fig. 2-6: RNA editing in the mitochondrial *nad4* gene in liverworts. Columns denote the putative editing sites: amino acid translated from the DNA sequence, number of the amino acid counted in relation to *Marchantia*, residue that would result from the edtiting

Another, but much rarer type of RNA editing is the “reverse” editing, the exchange from U to C, which is essential to remove stop codons that disrupt the CDS. This type has been found in hornworts, where it is necessary to correct stop codons in more than half of the genes in the case of the chloroplast genome (Yoshinaga et al. 1996; Yoshinaga et al. 1997; Kugita et al. 2003b), and is rather frequent in ferns (Vangerow et

al. 1999; Wolf et al. 2004), but very rare in gymnosperms and angiosperms (Hiesel et al. 1994; Freyer et al. 1997). Nevertheless, it has been found in both mitochondrial and chloroplast genomes, although the amount is much higher in mitochondria.

In this study we analysed the largest scale liverwort data set regarding the existence of RNA editing sites of a mitochondrial gene. Still, no evidence could be found for any RNA editing in complex thalloid liverworts, as was already proposed earlier. Putative Editing sites in simple thalloid and leafy liverworts were identified by comparison with the *Marchantia polymorpha* chondriome and its related taxa. All findings are presented in figure 2-6.

The full size protein as deduced from the *Marchantia polymorpha* chondriome sequence is 496 amino acids (aa) long. The average number of amino acids that were checked for editing sites is 438 aa per taxon, which is almost 90 % of the proposed gene length. In this region a total of 120 putatively edited nucleotide sites were identified, 47 of which are unique to a single species. They result in 103 changed amino acids, meaning that more than every fifth aa of *nad4* is target for RNA editing in at least one liverwort species. All proposed nucleotide changes are C to U editings.
The highest number of putative editing sites was identified in *Haplomitrium* with 53 edited aa, comprising almost every 8th aa of the analysed region (Fig. 2-6). This is an extraordinarily large amount, with an even slightly higher ratio than the findings from the *nad5* gene, where an average of every 9th aa is edited in this genus (Groth-Malonek et al. 2005), rendering it the most strongly editing species in that study, which also included hornwort, fern, and angiosperm sequences in comparison.
However, the confirmation of the proposed editing sites by the analysis of cDNA sequences from the same taxa should be added in further experiments.

2.3.1.3 A phylogenetic study in liverworts: the usefulness of *nad4* as a novel marker gene

The phylogeny of all liverworts based on molecular data has come into focus only recently (Davis 2004; He-Nygren et al. 2004; Forrest and Crandall-Stotler 2005). Mostly chloroplast and nuclear loci were used for these approaches, and only one single

mitochondrial locus, the *nad5* gene, has been included in some studies, where large parts of its intron are usually excluded from the study because of missing parts or alignment problems .

This phylogenetic approach presented here (Fig. 2-7) includes for the first time a liverwort taxon sampling over all major subgroups that is solely based on a mitochondrial gene, and that includes one mitochondrial group II intron (see chapter 2-1).

The study is meant to establish whether the *nad4* gene is an adequate locus for phylogenetic analyses in liverworts, based on the assumption that this very ancient land plant group could be well understood by the use of highly conserved and slowly evolving DNA sequences like genes from the chondriome (Wolfe et al. 1987).

OVERALL TOPOLOGY

The basal-most branching event (Fig. 2-7) supports the bifurcation of the morphologically simple thalloid *Blasia* and the rest of all liverworts. *Blasia* has been proposed to be placed in basal position of the complex thalloid clade by recent genetic approaches (Davis 2004; Forrest and Crandall-Stotler 2004), or treated as a simple thalloid liverwort of the order Metzgeriales by morphological analyses (Renzaglia 1982; Schuster 1992) or *rbcL* analysis (Wheeler 2000). Other studies placed *Blasia* in a basal position to the rest of the liverworts (Stech and Frey 2001; He-Nygren et al. 2004), classified as a separate class Blasiopsida. In this study *Blasia* is placed in an unranked relationship with three other groups. One of them are the complex thalloid liverworts (Marchantiopsida), that are considered a well defined monophyletic group in most studies, and usually placed as sister to the rest of the liverworts. Another clade is formed by the combined taxa of the Haplomitriopsida and Treubiopsida. All species of this group are very rare, and tend to have a high genetical distance from the rest of the liverworts, resulting in very long branches and potential misplacing due to long branch attraction. The third unranked clade is comprised of all Jungermanniopsida, including the two growth forms of the simple thalloids and the leafy liverworts.

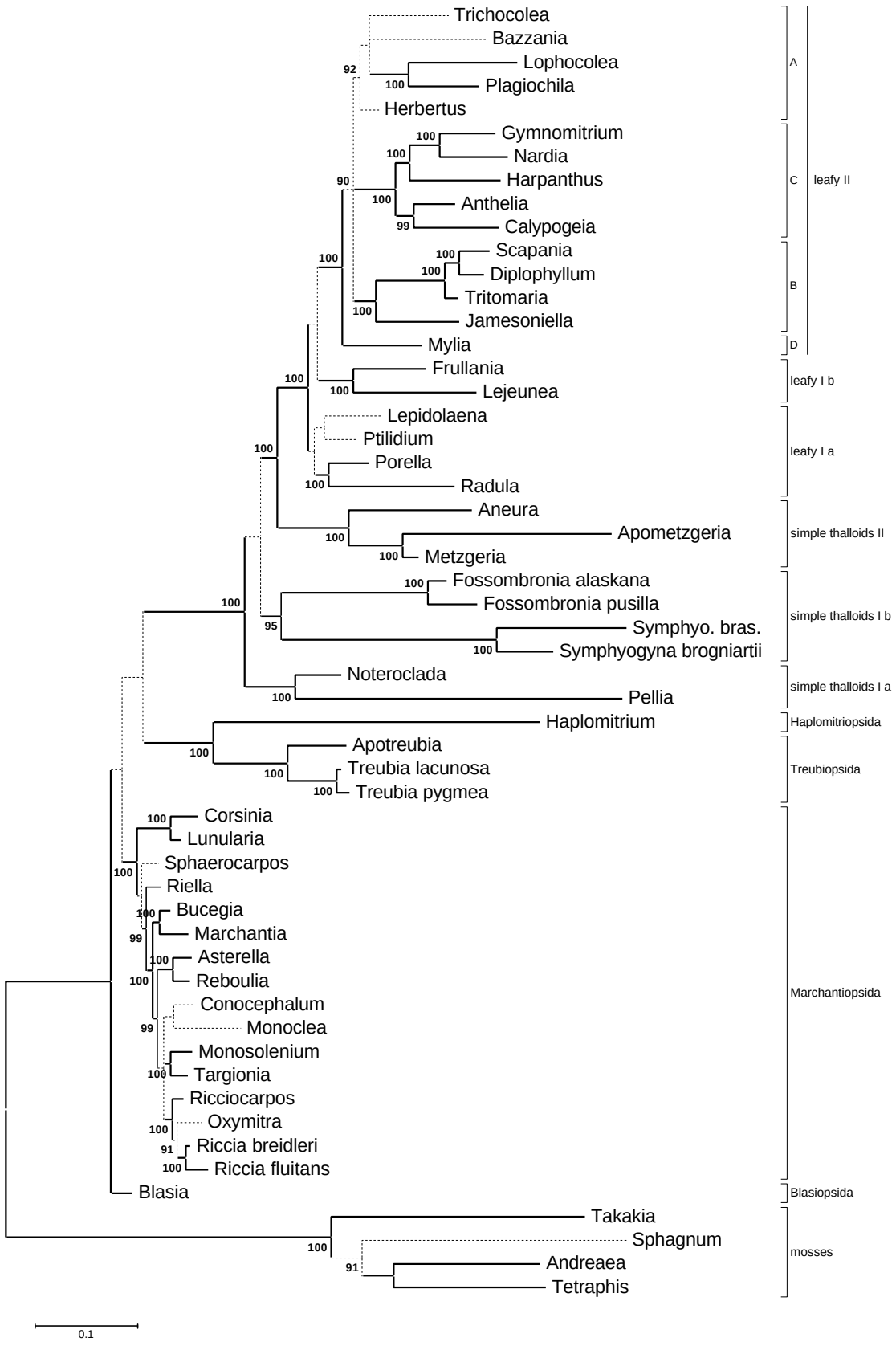

Fig. 2-7: Phylogeny of liverworts based on the mitochondrial *nad4* gene. Dotted lines are branches without statistical support >90 % through Bayesian Posterior Probabilities. Thin lines are weakly supported nodes (95-99 %), strong lines have maximum support (100 %).

MARCHANTIOPSIDA

The basal-most group, which is not supported in an unequivocal position, but strongly supported as a clade, is comprised of the Marchantiopsida (complex thalloid liverworts), including 16 taxa from 15 genera (Fig. 2-8). This group represents only about 5 percent of all liverwort species, and is genetically very homogenous. Other molecular approaches to obtain a phylogeny of the complex thalloid liverworts included nuclear (LSU) rDNA genes (Boisselier-Dubayle et al. 1997; Boisselier-Dubayle et al. 2002), nuclear *18S rDNA* (Bopp and Capesius 1996; Capesius and Bopp 1997), the combination of *nucLSU* and chloroplast *trnL-trnF*-spacer (Wheeler 2000), or chloroplast *rbcL* (Lewis et al. 1997). All studies exhibited low support for most subgroups and / or conflicting results, with the exception of one recent all-liverwort approach involving 8 loci (five chloroplast, one mitochondrial, two nuclear) that included a relatively large set of 12 taxa plus the reconsidered former simple thalloid liverworts of the Blasiaceae, *Blasia* and *Cavicularia* (Forrest and Crandall-Stotler 2005).

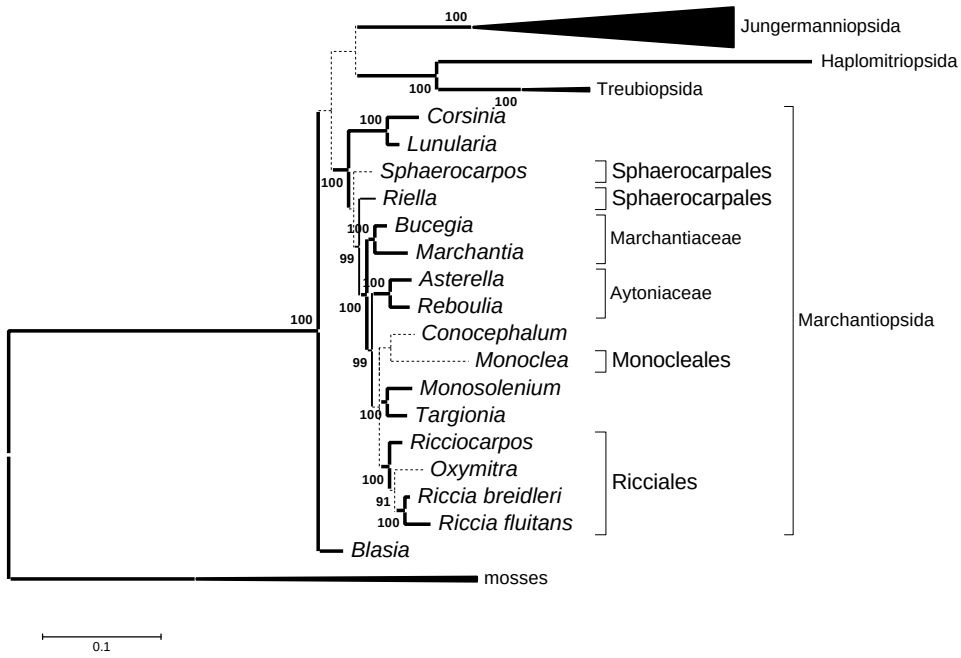

Fig. 2-8: Backbone phylogeny of liverworts based on the mitochondrial *nad4* gene. Some clades are collapsed for enhanced visuality. Subtree includes additional taxonomic descriptions. Dotted lines are branches without statistical support >90 % through Bayesian Posterior Probabilities. Thin lines are weakly supported nodes (95-99 %), strong lines have maximum support (100 %). Complete tree see figure 2-7.

In this study (Fig. 2-8) the clade exhibits strong support for several branches within the Marchantiopsida clade that could give further insights into the phylogeny of this liverwort group. The basalmost branch is comprised of *Corsinia* and *Lunularia* in a strongly supported relationship alongside with *Sphaerocarpos*. This placement of *Lunularia* and *Sphaerocarpos* was already proposed several times (mentioned above) and seems to be a valid assessment of the earliest-diverging complex thalloids. *Corsinia* has formerly been placed in the crown group (Boisselier-Dubayle et al. 2002), but exhibited the same strong relationship with *Lunularia* in an approach using *nad5* data (Beckert et al. 1999). As those *nad5* sequences were obtained using the same DNA samples as in this *nad4* study, alternative DNA samples should be tested to confirm this placement. *Riella* is taxonomically placed in the subclass Sphaerocarpidae, together with *Sphaerocarpos* (Frey and Stech 2005). However, no unambiguous monophyletic Sphaerocarpidae-clade could be identified here (Fig. 2-8), but there is also no support for a paraphyletic or polyphyletic relationship. *Riella* has in fact been shown in a monophyletic group with *Sphaerocarpos* by a phylogenetic approach using only nuclear sequences (Wheeler 2000). The next diverging branch is composed of two members of the Marchantiaceae, *Marchantia* and *Bucegia*. The placement of this family in a rather basal position of the complex thalloids has also been proposed before, albeit with only moderate support (Wheeler 2000; Boisselier-Dubayle et al. 2002). One part of the crown group consists of the monophyletic family Aytoniaceae (*Asterella* and *Reboulia*). The genus *Asterella* has previously been discussed as a polyphyletic group in this family (Long et al. 2000), but the placement of the family as a monophyletic clade was strongly supported in that study, as it is here. Another part of the crown group is a well supported clade formed by the four members of the order Ricciales. This order is comprised of two families, the Oximitriaceae and the Ricciaceae, including the only two genera *Riccia* and *Ricciocarpos*. The monophyly of the family Ricciaceae is not fully supported here, as *Oxymitra* is placed in an unsupported position between the two genera, suggesting a possibly polyphyletic family. Nevertheless, the order Ricciales is strongly supported as a monophyletic group. A new taxon that has not been included in in the previous analyses is the genus *Monosolenium* (Monosoleniacae), which is taxonomically placed in the suborder Marchantiineae of the order Marchantiales. *Monosolenium* clusters together with *Targionia*, the single member of the suborder

Targioniineae sampled in this survey. This suborder consists of a single family with the single genus *Targionia*, and should probably be included in the suborder Marchantiineae as no clear placement in a sistergroup relationship to the other subclasses can be postulated, at least on genetical basis. Two other unsupported crown group taxa are *Conocephalum*, a Marchantiineae taxon, and *Monoclea*, which is considered to be a member of the order Monocleales, but is clearly nested here within the Marchantiales clade (Fig. 2-8). Monocleales have been morphologically described by the absence of ventral scales in the gametophyte and the occurrence of a massive seta (Crandall-Stotler and Stotler 2000), and especially the rather simple thalloid appearance. As these are autapomorphic characters for a Marchantiopsiid liverwort, no connection between the orders Marchantiales, Ricciales, or Sphaerocarpiidae/ Sphaerocarpales can been made. In this molecular approach no sistergroup relationship to any of them was supported. As many morphogical features, especially in the complex thalloid liverworts, show reduced states that are not easily identified as homoplastic (Boisselier-Dubayle et al. 1997; Boisselier-Dubayle et al. 2002), genetic data could shed some light on the matter, as there are many more character states available from DNA than from morphological approaches. The data of this study do not support a separate placement of *Monoclea* in an order Monocleales, but rather suggest a placement within the Marchantiales. The order Marchantiales itself is clearly paraphyletic if the order Ricciales is regarded as a valid taxonomic unit. To improve the knowledge of these unclear relationships more data, especially from further loci, should be analysed.

HAPLOMITRIOPSIDA / TREUBIOPSIDA

This clade is comprised of three taxa of the Treubiopsida and one of the Haplomitriopsida (Fig. 2-7). They tend to be placed on the basis of all liverworts, either Treubiopsida alone (Stech and Frey 2001) or in combination together (Crandall-Stotler et al. 2005) or have a different placement with Haplomitriopsida connected to the simple thalloid liverworts (Stech and Frey 2001). Morphologically, the placement is still debated, and they have been placed with *Blasia* as separate orders Haplomitriales, Treubiales and Blasiales at the basis of the class Jungermanniopsida, subclass Metzgeriidae (Crandall-Stotler and Stotler 2000). The latest molecular and taxonomic studies positioned *Blasia* at the basis of the complex thalloid clade, and the

Haplomitriopsida/Treubiopsida group in a separate cluster as the earliest diverging group of all liverworts (e. g. Forrest and Crandall-Stotler 2004; Frey and Stech 2005). Both groups clearly exhibit an extensive genetic distance to the rest of the liverworts, and *Haplomitrium* is also known for its high degree of RNA editing (Groth-Malonek et al. 2005), a feature that is either lost or has not yet evolved in Marchantiopsida (Steinhauser et al. 1999). In this study no placement of this group can be proposed, but a joint placement of both classes is supported so far.

JUNGERMANNIOPSIDA (SIMPLE THALLOID & LEAFY LIVERWORTS)

This class can be divided into two morphologically distinct groups, the simple thalloid and the leafy liverworts (Metzgeriidae vs. Jungermanniidae, respectively), the latter comprising approximately 95 % of all extant liverwort taxa (Fig. 2-7). Recent studies revealed that some morphological features, especially the simple thalloid state, have evolved several times during evolution, as for instance the occurrence of simple thalloid vegetative phases in some leafy liverworts, and are therefore possible homoplastic characters. The simple thalloid taxa of this study are separated in three well supported clades: simple thalloids Ia, Ib, and II (Fig. 2-7). The placements of clade II as a sister group to the subclass Jungermanniidae renders the simple thalloid taxa paraphyletic (Crandall-Stotler and Stotler 2000). This agrees with recent genetic studies (Davis 2004; Forrest and Crandall-Stotler 2004; Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005) and led to the proposal of a distinct separation of one superclass for the simple thalloid clade I and another superclass for clade II in connection with the leafy liverworts (Frey and Stech 2005), accompanied by a superclass for the Marchantiopsida including *Blasia* and another one for the Treubiopsida/Haplomitriopsida clade. It should be noted, that the simple thalloids clade Ia is composed of the strongly supported clade of the two genera *Noteroclada* and *Pellia*, which are morpho-taxonomically placed in the family Pelliaceae and were resolved together in several molecular analyses (Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005), but were recently regarded as an unclear relationship based on very different versions of *rbcL* sequences available from public databases (Heinrichs et al. 2005; Frey and Stech 2005). The *nad4* data of this study concur with the placement of *Noteroclada* and *Pellia* in a close relationship (Fig. 2-7).

JUNGERMANNIIDAE (LEAFY LIVERWORTS)

This group is the most taxon-rich group of all liverworts, and is comprised of several well supported subgroups (Fig. 2-7). Basalmost are two subclades referred to as leafy Ia and leafy Ib, that are taxonomically circumscribed as the order Porellales, which was recently expanded to include the former Radulales and also *Ptilidium* and *Lepidolaena* based on a *rbcL* phylogeny (Heinrichs et al. 2005). In that study no statistical support for this grouping was evident, but the circumscription as Porellales originated rather from the separation of the suggestively combined but unsupported clades of the *rbcL* phylogeny from a well supported sister clade Jungermanniales. The present study based on *nad4* data cannot contradict this placement, as there is no apparent difference to the unsupported placement of *Ptilidium* or *Lepidolaena* in this study (Fig. 2-7). The strong support of a *Lejeunea-Frullania* clade, however, has been proposed before by morphological and genetic studies (Crandall-Stotler and Stotler 2000 and therein). The high support for the clade composed of *Radula* and *Porella* mirrors their placement in the former order Porellales sensu Schljakov (1972) as equally ranked suborders, rather than the placement of *Radula* in a separate order Radulales parallel to Porellales (*Frullania, Lejeunea, Porella*) (Crandall-Stotler and Stotler 2000, later referred to as CSS).

The crown group of the phylogeny is composed of the group leafy II that is subdivided in three clades A, B, C, and the basal placed genus *Mylia* (Jungermanniaceae s.str.) (Fig. 2-7). This taxon has not been included in recent large scale phylogenetic approaches of liverwort phylogeny, but, relating to two phylogenetic approaches analysing the Lophoziaceae and related families using chloroplast data (Schill et al. 2004; Yatsentyuk et al. 2004), the placement of *Mylia* should be in clade C. However, *Mylia* has been designated as clade D, because no affiliation to any of the clades A, B, or C is supported in this study (Fig. 2-7). Clade B consists of *Jamesoniella, Scapania, Tritomaria*, and *Diplophyllum* in a strongly supported relationship. *Jamesoniella* and *Tritomaria* are genera of the family Lophoziaceae sensu Schill (this family is included in the Jungermanniaceae sensu CSS), *Scapania* and *Diplophyllum* are genera of the family Scapaniaceae sensu Schill, or divided into the two families Scapaniaceae and Diplophyllaceae sensu Potemkin (1999). The close relationship of the latter two taxa is supported in this study (Fig. 2-7). The relationship of *Jamesoniella* (subfamily

Jamesonielloideae) and *Tritomaria* (subfamily Lophozioideae) is paraphyletic, with the placement of the Scapaniaceae as a crown group, which is identical to the findings of Schill et al. (2004). In that study the inclusion of the Scapaniaceae in the Lophoziaceae was considered premature. The results of these *nad4* data strongly support at least a close relationship of all four genera.

Clade C consists of *Gymnomitrion* (suborder Jungermanniineae, Gymnomitriaceae), *Nardia* (suborder Jungermanniineae, Jungermanniaceae), *Harpanthus* (suborder Lophocoleineae, Geocalycaceae), *Anthelia* (suborder Antheliineae, Antheliaceae), and *Calypogeia* (suborder Lepidoziineae, Calypogeiaceae). Clade A includes *Plagiochila* (suborder Lophocoleineae, Plagiochilaceae) and *Lophocolea* (Lophocoleineae, Geocalycaceae sensu CSS (incl. Lophocoleaceae)) in a strongly supported clade, accompanied by *Bazzania* (suborder Lepidoziineae, Lepidoziaceae), *Herbertus* (suborder Herbertineae, Herbertaceae), and *Trichocolea* (order Lepicoleales sensu CSS, suborder Lepicolaeninae, fam. Trichocoleaceae) (Fig. 2-7).

The strong support of the Plagiochilaceae (*Plagiochila*) and Geocalycaceae (*Lophocolea*) is mirrored with *rbcL* data (Heinrichs et al. 2005), but that study does not include the genus *Harpanthus* that is placed in clade C here (Fig. 2-7), in accordance with the data from Davis (2004). The family concept of the Geocalycaceae is clearly in need of a taxonomic revision, as it appears to be a polyphyletic group in independent studies. Further polyphyletic groups identified here are the suborder Lepidoziineae (*Calypogeia* and *Bazzania*), and the order Lepicoleales sensu CSS (*Trichocolea*, leafy II, *Lepidolaena* leafy Ia). These findings are congruent to the results of Davis (2004) and Heinrichs et al. (2005), and support them with data from an independent new locus on the mitochondrial genome.

CONCLUSION

In conclusion it can be said, that the *nad4* gene is indeed useful to establish phylogenetic relationships in liverworts, but the topology of the resulting tree is lacking a well supported backbone. The main groups were identified with rather high statistical support. In comparison with other studies concerning liverworts, the complex thalloids exhibited a better statistical support for the topology when based on *nad4* data. This could be very useful to establish a well resolved phylogeny for this group with a larger

taxon sampling, probably in connection with other molecular markers. This approach could also be used to gain more support for the backbone phylogeny.

2.3.2 The *nad4* gene in mosses

In a previous study concerning the intron distribution of bryophyte mitochondria compared to flowering plants (Pruchner et al. 2001) *nad4* amplicons of two mosses were sequenced for the first time: *Timmia* and *Takakia*. Both taxa exhibited a single intron: nad4i461 (Fig. 2-8), which is also present in angiosperms. These findings supported the taxonomic position of the enigmatic genus *Takakia* as a moss. The genus

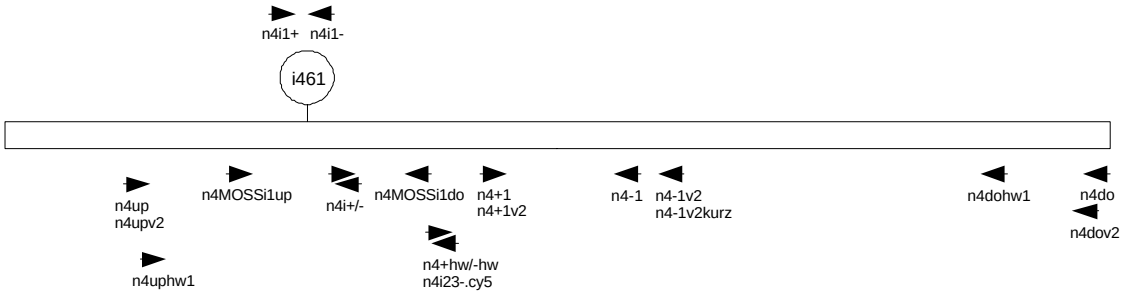

Fig. 2-8: Overview of the mitochondrial *nad4* gene in mosses. The group II intron is indicated by a pinhead. Intron designation relies on the nucleotide position in the *nad4* gene of the reference liverwort *Marchantia polymorpha* after which the intron is inserted. Primer locations and directions are symbolized by arrows. Primers underneath the bar are located in exon regions, primers above the bar are located in the intron.

has been placed as a liverwort in earlier studies (e.g. Schuster 1984), until moss-like sporophytes were discovered (Smith and Davison 1993). Molecular studies supported the latter placement (Hedderson et al. 1998; Pruchner et al. 2001; Beckert et al. 2001).

An extension of the *nad4* data set for mosses in this study (Tab. 2-4) verifies the conservation pattern of nad4i461, and includes a first analysis of the phylogenetic potential of the *nad4* locus in mosses. The intron distribution of 13 *nad4* sequences of mosses is shown in figure 2-7. In all species only the intron nad4i461 has been found, including *Takakia*. This supports the taxonomic placement of this genus within the mosses, as has been shown by phylogenetic approaches using *nad2* and *nad5* data (Beckert et al. 2001).

Table 2-4: Taxa used for the phylogenetic analyses of the *nad4* gene in mosses

Phylogeny	Species	collection number	Accession number and sequence length *nad4*	Accession number and sequence length *rbcL*	Accession number and sequence length *rps4*
Andreaeopsida	*Andreaea nivalis* Hook.	Muhle140897-3	2052 bp	AF478199 1354 bp (*rupestris*)	AJ617675 568 bp (*rupestris*)
Sphagnopsida	*Sphagnum fallax* Klinggr.	Muhle180597-2	2022 bp	AB013673 1305 bp	AF307004 569 bp (*tenerum*)
Takakiopsida	*Takakia lepidozioides* Hatt. et H. Inoue	Qiu97126	AJ409092 1990 bp	AF244565 1200 bp	AJ269687 609 bp
Tetraphidopsida	*Tetraphis pellucida* Hedw.	Ulm-collection s.n.	2036 bp	AF478203 1347 bp	AF231896 561 bp
Bryopsida	*Bartramia halleriana* Hedw.	Muhle140897-6	2032 bp (VK)	AF491009 1285 bp	AF265358 602 bp
	Buxbaumia aphylla Hedw.	Muhle070398-1	1938 bp	AY118230 1297 bp	AY137677 584 bp
	Diphyscium sessile Lindb.	Muhle191097-2	1999 bp (VK)	AY312928 1315 bp (*foliosum*)	AJ251065 609 bp (*foliosum*)
	Fissidens cristatus Wilson & Mitt.	Muhle200497-3	2034 bp	AF226810 1329 bp (*mooreae*)	AF223056 588 bp (*subbasilaris*)
	Isothecium alopecurum (Hedw.) Spruce	Muhle291197-6	2035 bp	AB029385 1428 bp (*Platyhypnidium riparioides*)	AY306933 570 bp (*myosuroides*)
	Mnium hornum Hedw.	Muhle090897-2	2036 bp	AF226820 1347 bp	AF023796 601 bp
	Physcomitrella patens (Hedw.) Bruch & Schimp.	Ulm-collection s.n.	2036 bp	AP005672 1428 bp	AF223044 586 bp
	Timmia bavarica Hessl.	Muhle1611197-1	AJ409093 2029 bp	AJ275185 1334 bp (*austriaca*)	AF223035 588 bp (*austriaca*)
	Ulota crispa (Hedw.) Brid.	Muhle200497-6	2029 bp	AY631208 1346 bp	AF306972 570 bp

VK stands for sequences obtained earlier in the laboratory of V. Knoop. Alternative species for *rbcL* and *rps4* analyses are given in italics.

A recent approach for the classification of mosses has been published by Goffinet and Buck (2004), where several rearrangements especially regarding the class level were made, as well as the establishments of five superclasses. The taxonomic placement of the genus *Takakia* in a separate superclass, or the arrangement of the classes Sphagnopsida and Andreaeopsida in superclasses that include only one class, respectively, mirrors the still ongoing search for the relationship between the earliest diverging mosses, based on the extremely different morphology of these plants. Usually, it is agreed upon the placement of *Sphagnum*, *Takakia*, often combined with *Andreaea*, as the basal-most branches of the phylogenetic tree (e. g. Beckert et al. 2001; Cox et al. 2004). An identical topology was obtained by midpoint rooting (see also Figs. 2-9, 2-10).

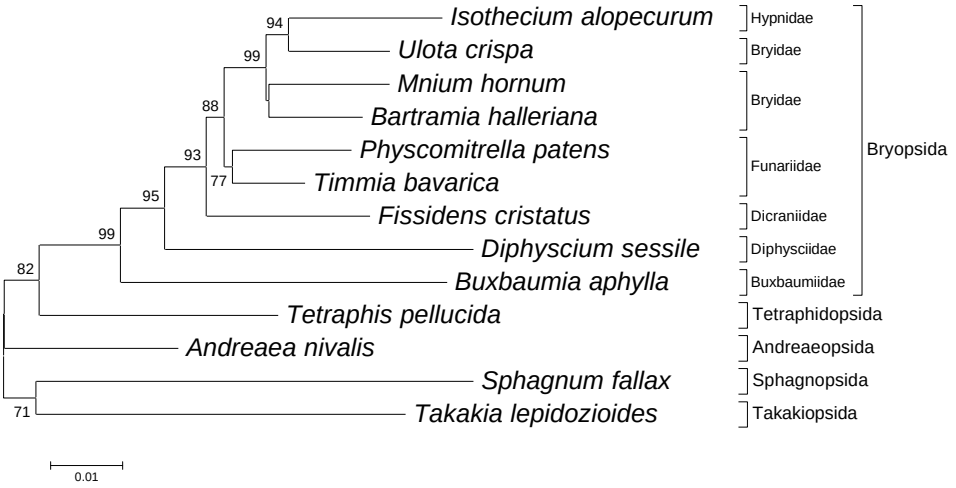

Fig. 2-9: Phylogenetic tree based on the mitochondrial *nad4* gene in mosses. The tree was obtained by Neighbor-joining using Kimura-2-Parameter distances, pairwise gap deletions, uniform rates among sites, and 10.000 bootstrap replicates; bootstrap values above 70 are given next to the respective nodes. Systematic classification is adopted from Goffinet and Buck 2004.

The relationships of the early diverging classes Tetraphidopsida, Andreaeopsida, Sphagnopsida, and Takakiopsida are not resolved in this tree, although there is a well supported distinction between them and the class Bryopsida, which comprises the majority of the taxa in this tree and of all mosses. In this class the topology exhibits a well supported backbone. No separation of the subclass Bryidae is supported, but this group is known to be paraphyletic (Goffinet and Buck 2004 and therein). The overall arrangement of the taxa is in congruence with previous studies on mosses (Beckert et al. 2001; Cox et al. 2004), and could therefore very well constitute a novel marker for moss phylogeny. The statistical support for the topology is definitely competitive to other markers with the same taxon sampling, like *rbcL* or *rps4* (Fig. 2-10).

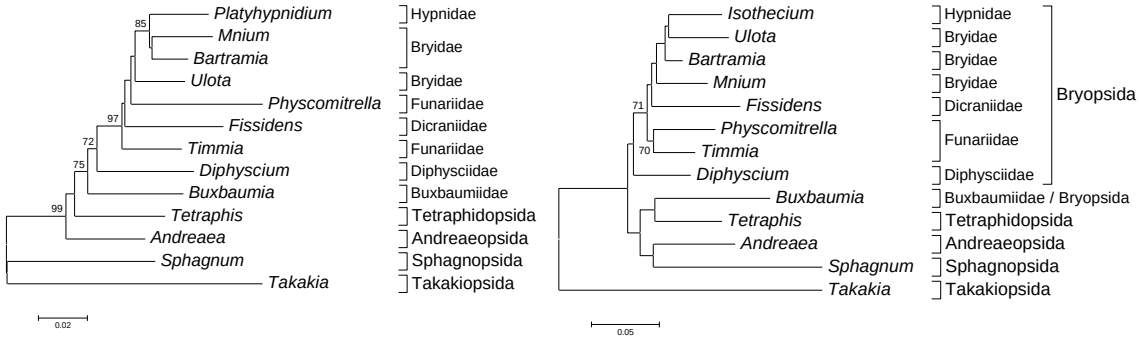

Fig. 2-10: Phylogenetic tree based on the mitochondrial *rbcL* in mosses (left) and *rps4* (right). Taxon sampling is based on the data for the *nad4* gene (Fig. 2-9). The tree was obtained by Neighbor-joining using Kimura-2-Parameter distances, bootstrap values above 70 are given next to the respective nodes. Systematic classification is adopted from Goffinet and Buck 2004.

It should be noticed, however, that the spacer region located upstream of the *nad4* gene, between *nad5* and *nad4*, is less than half the size of the gene studied here, and exhibits very good statistical support as well (see chapter 4, also in: Rein, Groth-Malonek, Knoop: "Mitochondrial gene spacers as novel phylogenetic markers: a case study in mosses", under revision).

2.3.3 The *nad4* gene in hornworts

Hornworts are the smallest and rarest group of bryophytes, with only about 100 taxa worldwide. Nevertheless, they comprise a key link to the understanding of the phylogeny of all land plants because of their widely debated placement, e.g. at the basis of all land plants (Mishler et al. 1994; Garbary and Renzaglia 1998) versus the placement as the sister group to all tracheophytes (Lewis et al. 1997; Groth-Malonek et al. 2005). A large part of this discussion originates from the morphological aspect that hornworts have no stem and leaves but a thallus, which at a first glance reminds of a complex thalloid liverwort. As the acquisition of hornwort DNA is also rather difficult due to the scarcity of these plants, only few taxa, if any, were included in recent molecular studies, with only few exceptions (Cargill et al. 2005 and therein; Duff and Moore 2005; Duff 2006).

The sequences of hornwort DNA studied here (table 2-5), are restricted to few fragments of the gene, and span almost the whole exon region, but do not include complete sequences of the identified introns. Nevertheless, the overall gene structure as known so far can be depicted as in figure 2-11.

Table 2-5. Overview of *nad4* hornwort sequences used for intron identification.

Species	Intron	Accession number and sequence length
Anthoceros agrestis Paton	nad4i461, present	DQ267609, 278 bp, intron: bases 190-278
Anthoceros agrestis Paton	nad4i548, absent	AJ409090, 317 bp
Anthoceros agrestis Paton	nad4i976, present	AJ409091, 506 bp, intron: bases 1-191
Megaceros spec.	nad4i1399, absent	276 bp

Two group II introns were identified in the *nad4* gene of hornworts. One of them, the intron nad4i461, is present in mosses and angiosperms in the same location (Fig. 2-11).

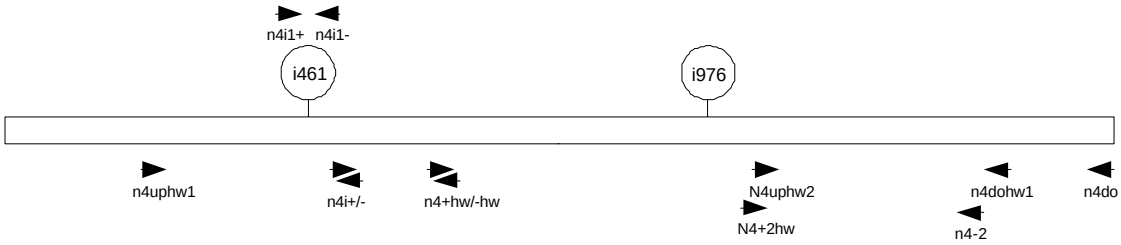

Fig. 2-11: Overview of the mitochondrial *nad4* gene in hornworts. Group II introns are indicated by pinheads. Intron designation relays to the nucleotide position in the *nad4* gene of the reference liverwort *Marchantia polymorpha* after which the intron is inserted. Primer locations and directions are symbolized by arrows. Primers underneath the bar are located in exon regions, primers above the bar are located in introns.

Even without the full sequence available, it can reasonably be proposed that they are vertically transferred homologues, as this pattern is known from other introns in *nad5* (nad5i230 and nad5i1455, respectively). The second intron found in *nad4* of hornworts is nad4i976. An intron in this position of the reading frame is known from angiosperms, where it can occasionally be lost, as has been shown in lettuce (Geiss et al. 1994). Hence, this intron, which is absent in mosses and liverworts, may be seen as an additional synapomorphy of a hornwort-tracheophyte clade. However, this is obscured by the fact that a positional homologue is present in the alga *Chara*. This resembles the case of the intron rps3i74, which is also known from the *Chara* chondriome, missing in the liverwort *Marchantia*, and conserved in some angiosperms, e.g. *Arabidopsis* (Turmel et al. 2003). The possibility that the positional homologues in *Chara* are of rather xenolog origin, and are not related to the introns identified in higher plants or hornworts, respectively, cannot be ruled out here, but could possibly be concluded by detailed comparisons of complete intron sequences in future studies.

2.3.4 The *nad4* gene in lycophytes and ferns

This study shows for the first time the intron distribution of the mitochondrial *nad4* gene in two polypod ferns (Fig. 2-12, table 2-5) and in fragments of lycophyte sequences (table 2-5 and V. Knoop, pers. comm., figure 2-14).

Table 2-5. Overview of *nad4* lycophyte and monilophyte sequences used for intron identification.

Species	Voucher number	concerning intron	Accession number and length
Asplenium scolopendrium	GF050421	i461 present, i548 absent, i976 present, i1399 absent	DQ267606, 4873 bp
Dryopteris uniformis	BGBN 22023	i461 present, i548 absent, i976 present, i1399 absent	4741 bp
Isoetes velata	MP040402	i976 absent	DQ304074, 562 bp
Huperzia selago	GS991028	i976 absent	DQ304072, 562 bp

Both species, *Asplenium* and *Dryopteris*, carry two group II introns in their *nad4* gene (Fig. 2-12), which are placed in the reading frame in identical positions as two angiosperm introns. One of them, nad4i461, is also conserved in mosses (chapter 2.3.2),

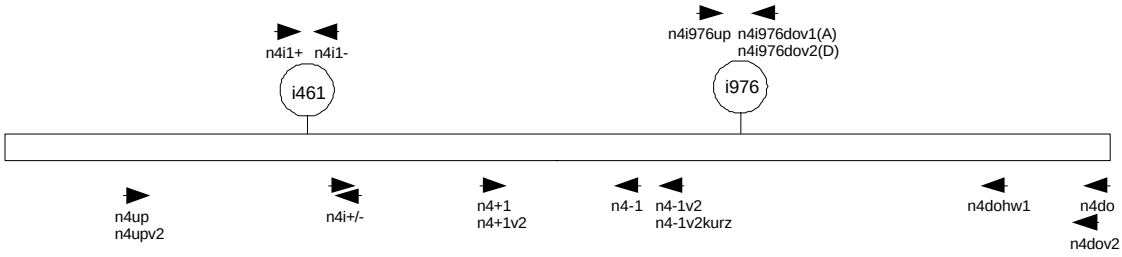

Fig. 2-12: Overview of the mitochondrial *nad4* gene in *Asplenium* and *Dryopteris*. Group II introns are indicated by circles. Intron designation relays to the nucleotide position in the *nad4* gene of the reference liverwort *Marchantia polymorpha* after which the intron is inserted. Primer locations and directions are symbolized by arrows. Primers underneath the bar are located in exon regions, primers above the bar are located in introns.

hornworts (2.3.3), and lycophytes (V. Knoop, unpublished data); nad4i976 has been found in hornworts, but not in lycophytes. In the light of the considerations above, this may reflect a secondary loss in the lycophyte clade. The alignment of the amino acid translation of the fern nucleotide sequences reveals 9 stop codons in *Asplenium* and 8 stop codons in *Dryopteris* (Fig. 2-13). Compared to the protein sequence of the liverwort *Marchantia polymorpha*, all of them are likely to be removed by reverse RNA editing, the change from U to C. This would mean a rescue of TGA (stop) through the change to CGA (R, arginine), and of TAA and TAG (stops) to CAA or CAG (Q, glutamine), respectively.

```
A. scolo. TI*WLPHSNINFIIGIDXTLLSLAVLTMSSILIRILVGWSCIEGFNKEYIIAFSIGESFMIAVSRMPDTSPLHVFFESVLILMS
D. uni.   ..R..........X..G.......S.............S....E.............T..........L....X.....S..
M.poly.   ..R...Y.....Y....GIS.FFVI..TFLTP.C....FYSVKSYK...M...F.C...L...FCSL.LLIFY........P.F

A. scolo. ITIGVRGSR*RKIRAAH*FLSHTPPGSVFMLLAISLIFF*TGTTDLHISLTTESSEWR*ILPRLAFLASFPXXVPMVXVHIRXP
D. uni.   .........*.......*.....................*...I.............*.....................L.
M.poly.   .I...W...Q...K..YQ.FLY.LM..LL.....LF...Q......Q.L....F..R.Q..LWI..F...SVK....P...WL.

A. scolo. EAHVEXLTAGFVILAGIPXKLGTYGFLRFSVPMFXXATLYFTPFVYTLSVIAITYTLPTTIR*IDLKKIIAYSSVAYMNFVTTG
D. uni.   ...X.V...........LS.....................................................*.....................
M.poly.   .....AP...S......LL...........I...PE........I........I..SL....Q.............H.....I.

A. scolo. MSSSNIQGIEGSILPTSSHGLVLLALSSCVGALYD*YKT*PVRYYGGPASTMPNFPTIPLLSTPANMSLPGTSSSIGEFPILVG
D. uni.   ..................................*...*.....................................................
M.poly.   .F.L..........LML.....SS..FL........RH..RI.K....LV....I.S..F.FF.L..........F....L....

A. scolo. AF*KNSLVATSAALGMILGAAYSPWLYNRVVFGNFKPNFILKFSDSNRREVLIFLPFIAGVIWMGVYPEVFLECMHTSVSNS
D. uni.   ..*.....................................................................................
M.poly.   ..QR......L............L.....................L............V.....................L
```

Fig. 2-13: Alignment of the *nad4* gene protein sequence of the liverwort *Marchantia polymorpha* with the amino acid translation of the nucleotide sequences of *Dryopteris uniformis* and *Asplenium scolopendrium*. Amino acids identical to the uppermost line are indicated by dots, potential editing sites are given in bold letters, stop codons are marked by asterisks. X stands for not translatable positions due to ambiguities in the nucleotide sequence.

The occurrence of stop codons in the amino acid translation is typical for mitochondrial fern sequences, as can also be seen in the fern *nad5* sequences (Vangerow et al. 1999). Reverse editing is known from hornworts, rather common in ferns, and much rarer again in angiosperms. Usually it is much less frequent than the "normal" land plant editing, the change from C to U (see also chapter 2.2).

2.3.5 Overview of the intron distribution in *nad4* for all land plants

The findings of this study on the mitochondrial *nad4* gene in land plants are combined in figure 2-14, as an overview of all identified introns.

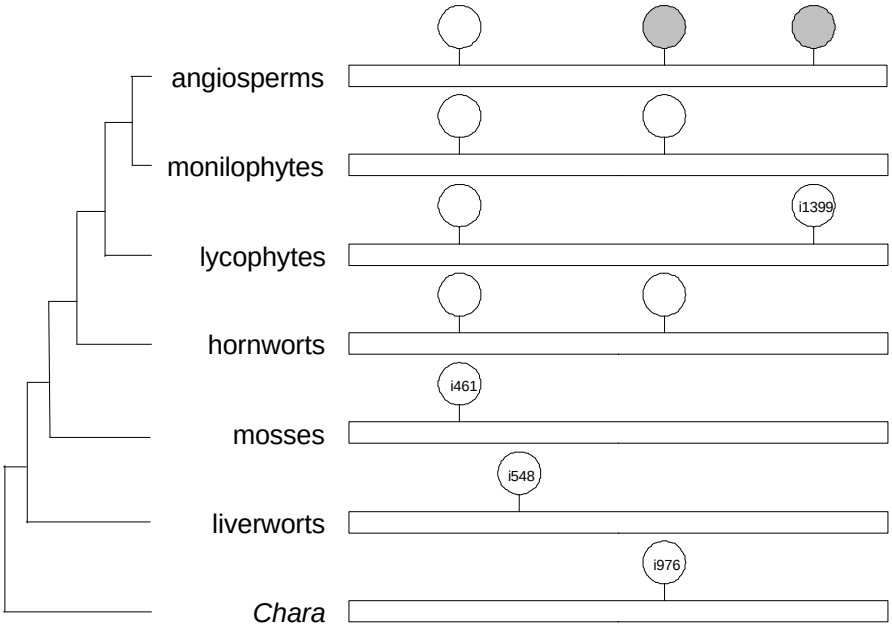

Fig. 2-14: Overview of intron distribution in the mitochondrial *nad4* gene in land plants. Circles indicate group II introns, grey circles stand for introns that can occasionally be lost in angiosperms. Intron designation relays to the nucleotide position of the CDS after which the intron is placed, in reference to the *Marchantia polymorpha* Genbank entry NC_001660, adopted after Dombrovska and Qiu (2004). The phylogenetic tree is extrapolated from recent studies about land plant phylogeny.

Four group II introns are known to interrupt the coding region of the mitochondrial *nad4* gene in land plants (fig. 2-14). One of these introns, **nad4i548**, is unique for liverworts, and is the first mitochondrial group II intron studied extensively for this bryophyte group.

Another intron, **nad4i976**, is already conserved in *Chara* (Fig. 2-14), yet not in *Chaetosphaeridium*. It has not been detected in any liverwort or moss sequence, but homologues have been identified in hornworts (V.Knoop, unpublished), monilophytes (this study), and angiosperms (e.g. Unseld et al. 1997). No evidence was found for its occurrence in *Isoetes*, a lycophyte, but as this group is very diverse in its conservation of introns despite its low number of species, the data should be regarded here with special caution before generalizing the data for the whole lycophyte group. The intron can also be occasionally lost from angiosperm mitochondria (Gass et al. 1992).

The intron **nad4i461** has been observed in all mosses, hornworts, lycophytes, monilophytes, and angiosperms studied (Fig. 2-14). This conservation pattern is identical to the *nad5* introns nad5i230 and nad5i1455 (Groth-Malonek et al. 2005), even though nad5i1455 evolved into a trans-splicing intron in angiosperms. An almost identical pattern was also shown for nad1i728 (Dombrovska and Qiu 2004), which is reported to be absent from *nad1* of the gymnosperm *Ephedra* (southern blot analyses in Qiu et al. 1998) and whose conservation state is unknown for *Isoetes* (Dombrovska and Qiu 2004). This tendency to conserve and vertically inherit introns that occur as early in the land plant evolution as in mosses is an indication for a phylogeny that is based on the early divergence of liverworts from the rest of the land plants, and mosses and hornworts in a basal position to tracheophytes. Hornworts are proposed to be even more closely connected to tracheophytes than mosses (Lewis et al. 1997; Samigullin et al. 2003; Groth-Malonek et al. 2005), one reason for that is the sharing of the intron nad5i1477 (which is absent in monilophytes, trans-splicing in angiosperms). Other argumentations are based on their morphological affinity to the potentially intermediate fossils between bryophytes and tracheophytes like *Hornea* (Campbell 1924), and chemical similarities regarding cell-wall xylans (Carafa et al. 2005).

The fourth intron identified in *nad4* is **nad4i1399**, a strictly tracheophyte intron that has been shown to occur in lycophytes (V. Knoop, pers. comm.) and angiosperms (Fig. 2-14), and has probably been secondarily lost in monilophytes. It is also occasionally absent in some angiosperms (Gass et al. 1992; Geiss et al. 1994).

The overall pattern of the intron distribution in the *nad4* gene (Fig. 2-14) can be applied for the following reconstruction of the land plant phylogeny: the closest ancestor to land

plants is the Charales alga *Chara*, as the angiosperm intron nad4i976 is already conserved here. The earliest bifurcation of the phylogenetic tree occurs between liverworts (loss of nad4i976, autapomorphic gain of nad4i548) and the rest of the land plants. Mosses comprise the earliest diverging branch of the non-liverwort lineage, and acquire the intron nad4i461, but lose nad4i976. Hornworts follow in a paraphyletic relationship, as the sistergroup to all tracheophytes. They do possess the Charales intron nad4i461. The tracheophytes lineage places lycophytes as the basal-most group, which gains nad4i1399, but loses nad4i461 again. Monilophytes as the paraphyletic next group and the sister group to spermatophytes, share their two introns with angiosperms, but do not carry nad4i1399. Actually, the intron distribution is identical to the hornworts.

Obviously, both angiosperm introns nad4i976 and nad4i1399 that are not always conserved in angiosperms seem to be rather less continuously conserved in older plant groups as well (Fig. 2-14). As only two complete monilophyte sequences and few parts of lycophytes were analysed in this study, it is possible that the apparent loss of any of the introns is a phenomenon restricted to some taxa of these plant groups, as it is the case in angiosperms. Further analysis of additional species of this gene could shed light on this problem. This of course applies to hornworts as well. In the case of liverworts and mosses the taxon sampling is considered to be widespread enough to refuse the option of the inheritance of these two introns at an earlier time than proposed here.

3 The *nad5* gene revisited: Six years after Beckert et al. (mosses) and Vangerow et al. (ferns): new insights into a now highly popular marker for lower land plant phylogeny

3.1 Introduction

Analyses of the mitochondrial *nad5* gene in land plants did not begin with phylogenetic approaches, but with the discovery of two trans-splicing group II introns in angiosperms (Knoop et al. 1991; Pereira et al. 1991), which are separated by a very small exon of only 22 nucleotides (Fig. 3-1). They are accompanied by two "normaly" splicing cis-arranged introns in angiosperms, **nad5i230** and **nad5i1872**.

Following this line, the ancestors of the trans-splicing introns were identified in lower land plants (Malek et al. 1997). One of them, **nad5i1455**, was found in the fern *Asplenium*, where it is not trans-splicing, but cis-arranged. The other one, **nad5i1477**, is absent in this species, but present in the lycophyte *Isoetes*, where it is extraordinary small with only 434 bp, and in the bryophyte *Anthoceros* (hornwort) with a size of 2391 bp, where it is cis-arranged in both cases (Malek and Knoop 1998). The complete sequencing of a liverwort chondriome (*Marchantia polymorpha*, Oda et al. 1992a) lead to the finding of a unique group I intron, nad**5i753**, which is absent in vascular plants.

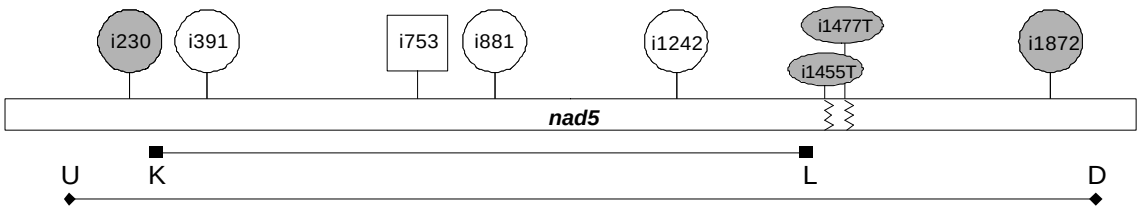

Fig. 3-1: Overview of the *nad5* gene structure in land plants. Group I intron depicted as square, group II introns as circles. Distruption of the gene in angiosperms through trans-splicing introns is marked by zig-zag-lines. Grey shaded introns are conserved in angiosperms. Nad5i391 is present only in *Huperzia selago*, nad5i753 is conserved in liverworts and mosses, nad5i881 is present in some hornworts, nad5i1242 is carried by most ferns (monilophytes).

The introduction of the mitochondrial *nad5* gene as a tool for phylogenetic analysis started with an investigation of bryophytes, including all three groups: liverworts, mosses, and hornworts (Beckert et al. 1999). In that study the intron nad5i753 was also found in mosses, rendering it the only known mitochondrial intron shared by liverworts

and mosses so far. No complete moss chondriome sequence is as yet available for comparison. Additionally, a novel group II intron **nad5i881** (Fig. 3-1) was identified in one of two hornworts included in that study (present in *Anthoceros husnotii*, missing in *Phaeroceros* spec.). This partial conservation pattern in hornworts was also confirmed by recent data (Duff 2006), where it is only found in the supposedly earliest-diverging hornwort *Leiosporoceros* and its closest related genera *Anthoceros*, *Folioceros*, and *Sphaerosporoceros*, but not in the crown group composed of the genus *Phaeroceros*, which would indicate a secondary loss of the intron nad5i881 in hornworts. The study by Beckert et al. (1999) also established *nad5* for the first time as a valuable phylogenetic marker gene for bryophytes, as the analyses established a clear distinction of all three major bryophytes groups and some good statistical support for new insights into the subclass and order relationships of mosses.

Following this promising beginning, another phylogenetic study was published in the same year (Vangerow et al. 1999) concerning a large scale phylogenetic approach on ferns and fern allies. In that study, another group II intron (**nad5i1242**, figure 3-1) was newly identified that is conserved only in ferns and lycophytes, with the exception of the genera *Ophioglossum* and *Equisetum*. The *nad5* gene of the lycophyte *Huperzia selago* contains an additional group II intron, **nad5i391**, that is very similar to nad5i1242, and has only been found in this species.

The usefulness of the very slowly evolving sequences of the mitochondrial DNA (Wolfe et al. 1987; Palmer and Herbon 1988) for phylogenetic studies was shown before by a approach using *coxIII* (Malek et al. 1996). Soon other studies started to utilise the novel locus *nad5* for the analysis of phylogenetic relationships of other plant groups, like the identification of Charales instead of Coleochaetales as the extant sister group to all land plants (Karol et al. 2001), or a study on the Pinaceae (Wang et al. 2000). In recent years, bryophytes, which were rather poorly analysed before by molecular methods, became the focus of several new studies, some of them implementing the *nad5* gene. This applies to mosses (Cox et al. 2004; Goffinet et al. 2004; Bell and Newton 2004), and, in a lesser extend, to liverworts (Davis 2004; Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005), the latter two focussing on one group of liverworts, the simple thalloids. The latest study of the intron distribution in *nad5* over all major land plant groups identified two of the angiosperm

introns, the cis-arranged nad5i230 and the trans-spliced nad5i1455, as cis-arranged and already conserved as early as in mosses.

This study analyses the present data available in public databases and reviews the folding pattern of the group I intron nad5i753 published earlier by Beckert et al. (1999). It combines already published data sets from liverworts with sequences from new taxa, and correlates the resulting phylogeny to a multi-gene approach based on *nad5*, *nad4*, and the chloroplast genes *rbcL* and *rps4* (see also chapter 4 for detailed analyses on a multi-gene approach on complex thalloid liverworts).

3.2 The *nad5* gene in liverworts: combined data from several labs

3.2.1 Material and Methods

DNA extraction and PCR assay strategies see chapter 2.2.1 “General methods and strategies”. Taxa used for PCR assays and additional sequences that were used for correlation and alignments are shown with their respective accession number from Genbank (NCBI), and/or their respective sequence length (table 3-2). Primers used for the DNA assays are published in (Beckert et al. 1999).

Parameters for the phylogenetic analyses implemented in MrBayes see chapter 2.2.1, exceptions are: 1.000.000 generations, every 100th tree sampled, burnin set to 8.000 trees, model parameters see table 3-1.

Table 3-1: Models implemented in MrBayes for *nad5* phylogeny in liverworts

character set (partition)	model selected by AIC (modeltest 3.1)	parameters implemented in MrBayes
exons *nad5*	GTR+I+G	revmatpr = fixed(2.2952, 4.9628, 0.4582, 1.0251, 15.4857, 1.0000) statefreqpr=fixed(0.2344, 0.2105, 0.2030, 0.3521), shapepr= exponential(0.7328) pinvarpr = fixed(0.1915), ratepr= variable, nst = 6, rates = gamma
intron *nad5*	GTR+G	revmatpr = fixed(0.6775, 2.9705, 0.2405, 0.5448, 1.7513, 1.0000) statefreqpr= fixed(0.2934, 0.2297, 0.2157, 0.2612), shapepr= exponential(0.8643) pinvarpr = fixed(0), ratepr = variable, nst = 6, rates = gamma
exons *nad4*	GTR+I+G	revmatpr = fixed(2.0757, 3.7931, 0.3069, 0.9919, 11.3371, 1.0000) statefreqpr= fixed(0.2585, 0.2132, 0.1959, 0.3324), shapepr= exponential(0.8181) pinvarpr = fixed(0.2145), ratepr= variable, nst = 6, rates = gamma
intron *nad4*	TVM+G	revmatpr = fixed(0.9545, 2.4145, 0.3375, 0.8789, 2.4145, 1.0000) statefreqpr= fixed(0.2455, 0.2434, 0.3026, 0.2085), shapepr= exponential(0.7358) pinvarpr = fixed(0), ratepr= variable, nst = 6, rates = gamma
rbcL	GTR+I+G	revmatpr = fixed(2.4998, 7.3054, 0.7947, 2.3309, 13.5614, 1.0000) statefreqpr= fixed(0.2921, 0.1562, 0.1933, 0.3584), shapepr = exponential(1.1534) pinvarpr = fixed(0.4775), ratepr= variable, nst = 6, rates = gamma
rps4	TVM+I+G	revmatpr = fixed(1.5391, 7.2094, 0.2991, 1.6475, 7.2094, 1.0000) statefreqpr= fixed(0.3892, 0.1275, 0.1696, 0.3137), shapepr = exponential(1.1591) pinvarpr = fixed(0.2227), ratepr= variable, nst = 6, rates = gamma

Table 3-2: Taxa used for the phylogeny of liverworts including the *nad5* gene.

Taxonomy	Taxon	*nad5*	*nad4*	*rbcL*	*rps4*
Jungermanniopsida	*Frullania moniliata* (Reinw., Blume & Nees) Mont.	AY688752 1800 bp	---	AY507401 1364 bp	---
Porellales	*Frullania tamarisci* (L.) Dumort.	2491 bp	2133 bp	AY302453 1042 bp	AY462349 576 bp
(leafy I)	*Lejeunea cavifolia* (Ehrh.) Lindb.	2493 bp	2247 bp	AY548102 1309 bp	AY462363 567 bp (*catanduana*)
	Lepidolaena hodgsoniae Grolle	2502 bp	2265 bp	AY462310 1038 bp (*taylorii*)	AY462368 576 bp (*taylorii*)
	Porella navicularis (Lehm. et Lindenb.) Lindb.	AY688767 1209 bp	2315 bp (*platyphylla*)	AY507420 1474 bp	AY462387 576 bp (*platyphylla*)
	Ptilidium pulcherrimum (G.Web) Vainio	2498 bp	2231 bp	AY302460 1042 bp	AY462388 576 bp
	Radula complanata (L.) Dum.		2301 bp	AY302461 1042 bp	AY608104 589 bp
Jungermanniales	*Anastrophyllum michauxii* (F. Weber) H. Buch ex A. Evans	AY688743 1800 bp	---	AY507390 1473 bp	AY507433 603 bp
(leafy II)	*Anthelia julacea* (L.) Dumort	---	2263 bp	---	AY608044 589 bp
	Bazzania trilobata (L.) Gray	AJ622815 2503 bp	AJ310800 2157 bp	L11056 1624 bp	AY608048 589 bp (*B. spec.*)
	Calypogeia muelleriana (Schiffner) K. Müller	2490 bp	2243 bp	U87065 1347 bp	AY608052 589 bp
	Diplophyllum albicans (L.) Dumort.	2489 bp	2174 bp	AY507397 1416 bp (*obtusifolium*)	AY608060 589 bp
	Gymnomitrion concinnatum (Lightf.) Corda	2517 bp	2322 bp	---	AY608065 589 bp
	Harpanthus flotovianus (Nees) Nees	2486 bp	2276 bp	---	AY608069 589 bp (*scutatus*)
	Herbertus sendtneri (Nees) Lindb.		2202 bp	AY507404 1467 bp (*alpinus*)	AY462353 569 bp
	Jamesoniella autumnalis (DC.) Steph.	AJ000700 1774 bp	2192 bp	AY462303 1038 bp	AJ251066 744 bp
	Lophocolea cuspidata	---	2323 bp	AY149854 1052 bp	AF231889 640 bp (*heterophylla*)
	Mylia taylorii (Hook.) Gray	---	2146 bp	---	---
	Nardia scalaris Gray	---	2225 bp	AY462316 1038 bp (*assamica*)	AY608092 589 bp
	Plagiochila asplenioides (L.) Dumort.	AJ000704 1799 bp	2134 bp	AY699996 1380 bp	AY547693 609 bp
	Scapania nemorea (L.) Grolle	AJ000706 1780 bp	2206 bp	AY507423 1484 bp	AY507464 603 bp
	Schistochila appendiculata (Hook.) Dum. ex Trev.	AY688770 1801 bp	---	AY507424 701 bp	AY507465 587 bp
	Trichocolea tomentella (Ehrh.) Dumort.	2499 bp	2231 bp	AY608040 1090 bp (*tomentosa*)	AY608118 589 bp
	Tritomaria quinquedentata (Huds.) H.Buch	2479 bp	2188 bp	AY700003 1380 bp	AY608119 589 bp
Fossombroniales	*Calycularia crispula* Mitt.	AY688747 1822 bp	---	AY507395 1469 bp	AY507437 603 bp
(simple thalloids I)	*Fossombronia angulosa* (Dicks.) Raddi.	AY688750 1313 bp	2253 bp (*alaskana*)	AY507398 1499 bp	AY507440 603 bp
	Fossombronia pusilla (L.) Nees	AJ000699 1747 bp	2232 bp	AF536231 1347 bp	AY608062 587 bp
	Hymenophyton flabellatum (Labill.) Dumort.	AY688755 1712 bp	---	AY507406 1499 bp	AY462357 575 bp
	Jensenia connivens (Colenso) Grolle	AY734748 1794 bp	---	AY688782 1496 bp	AY507450 602 bp
	Jensenia spinosa (Lindbg. & Gott.) Grolle	AY734747 1790 bp	---	AY734689 1515 bp	AY734698 728 bp
	Moerckia flotoviana (Nees) Schiffn.	AJ223717 1800 bp	---	AY507412 1468 bp (*blyttii*)	AY507454 603 bp (*blyttii*)
	Noteroclada confluens Taylor ex Hook. & Wilson	AJ622816 2489 bp	2240 bp	AY688784 1493 bp	AY688797 655 bp
	Pallavicinia rubristipa Schiffn.	AY734753 1790 bp	---	AY734693 1437 bp	AY734702 724 bp
	Pallavicinia xiphoides (Hook. f. & Tayl.) Trev.	AY734752 1758 bp	---	AY734692 1499 bp	AY734700 754 bp
	Pellia endiviifolia (Dicks.) Dum.	2476 bp	1730 bp	AY688786 701 bp	AY688800 659 bp
	Symphyogyna brasiliensis Nees & Mont.	1618 bp	2240 bp	AY734694 1504 bp (*marginata*)	AY734703 601 bp (*marginata*)
	Symphyogyna brogniartii	AY734754 1501 bp (*marginata*)	2321 bp	AY688789 1469 bp	AY608112 589 bp

Taxonomy	Taxon	*nad5*	*nad4*	*rbcL*	*rps4*
Metzgeriales (simple thalloids II)	*Aneura pinguis* (L.) Dumort.	AY688744 1785 bp	2182 bp	AY507391 1461 bp	AY608043 579 bp
	Apometzgeria frontipilis		2308 bp (*spec.*)	---	AY608045 589 bp
	Metzgeria conjugata Lindb.	AJ000703 1769 bp	2304 bp (*furcata*)	AY507411 1471 bp	AY507453 603 bp
	Riccardia capillacea (Steph.) Meenks & DeJong	AY688768 1785 bp	---	---	---
Blasiopsida	*Blasia pusilla* L.	2444 bp	2254 bp	AF536232 1347 bp	AY507436 603 bp
Marchantiopsida (complex thalloids)	*Asterella blumeana* (Nees) Pande Srivastava et Khan.	---	2265 bp	U87064 1347 bp (tenella)	---
	Bucegia romanica Radian	AJ001031 1794 bp	2313 bp	---	---
	Conocephalum conicum (L.) Underw.	2495 bp	2203 bp	AY688778 1353 bp	AY688791 725 bp
	Corsinia coriandrina (Spreng.) Lindb.	AJ622813 2492 bp	AJ310801 2151 bp	---	---
	Lunularia cruciata (L.) Dum. ex Lindb.	AJ001002 1792 bp	AJ310803 2151 bp	U87077 1347 bp	AY688795 985 bp
	Marchantia polymorpha L.	NC_001660 2682 bp	NC_001660 2387 bp	NC_001319 1428 bp	NC_001319 609 bp
	Monoclea gottschei Lindb.	AJ622814 2474 bp	2223 bp	AY507414 1343 bp	AY507455 549 bp
	Monosolenium tenerum Griff./Sunita Kapila & SS Kumar	2499 bp	2258 bp	---	---
	Oxymitra incrassata (Brotero) Sergio & Sim-Sim	---	2295 bp	---	---
	Reboulia hemisphaerica (L.) Raddi	2490 bp	2172 bp	AY462326 1038 bp	AY688801 741 bp
	Riccia fluitans L.	---	AJ310802 2155 bp	---	AY608107 567 bp
	Riccia breidleri Steph.	---	2268 bp	AY507422 1068 bp (*huebeneriana*)	AY507463 549 bp (*huebeneriana*)
	Ricciocarpos natans (L.) Corda	AJ001032 1789 bp	2265 bp	U87089 1347 bp	AJ251062 815 bp
	Riella spec.	---	2086 bp	---	---
	Sphaerocarpos donnellii Aust.	AJ001033 1797 bp	2210 bp	AY507425 1482 bp (*texanus*)	AY608110 580 bp (*texanus*)
	Targionia hypophylla L.	AJ001001 1793 bp	2124 bp	AY507427 1353 bp	AY688805 1153 bp
Haplomitriopsida	*Haplomitrium mnioides* (Lindb.) Schust.	AJ409111 2479 bp	2208 bp	AB013678 1317 bp	AY507444 600 bp
	Haplomitrium gibbsiae (Steph.) Schust.	AY688753 1793 bp	---	AY688781 1307 bp	AY688793 790 bp
Treubiopsida	*Apotreubia nana* (S. Hatt. & Inoue) S. Hatt. & Mizut.	2451 bp	2263 bp	---	---
	Treubia lacunosa (Colenso) Prosk. lenta Taylor ex Prosk.	2465 bp	2318 bp	AY507428 1403 bp	AY507468 603 bp
	Treubia pygmaea Schust.	2465 bp	2290 bp	AY507429 1280 bp	AY507469 603 bp
Bryophyta	*Andreaea rupestris* Hedw.	AJ001227 1948 bp	2052 bp (*nivalis*)	AY312925 1295 bp (*wilsonii*)	AJ617675 840 bp
	Sphagnum fallax Klinggr.	AJ622817 2083 bp	2022 bp	AF231887 1482 bp (*palustre*)	AF231893 639 bp (*cuspidatum*)
	Takakia lepidozioides Hatt. et H. Inoue	AJ291553 1952 bp	AJ409092 1990 bp	AY312936 1312 bp	AF306950 570 bp
	Tetraphis pellucida Hedw.	AJ224855 1637 bp	2036 bp	---	AF231896 636 bp

3.2.2 Structure and conservation of the group I intron nad5i753

The gene structure of the mitochondrial *nad5* gene is rather simple in liverworts: only one intron, nad5i753, has been identified, and is carried unanimously by all liverworts (and mosses). The analysis of its consensus sequence over 51 liverwort species (table 3-2) reveals the folding pattern of a group I intron, as was previously published by Beckert et al. (1999). This earlier study presented a secondary structure deduced from the alignment of 15 liverworts and 30 mosses, drawn from the example of the moss *Brachythecium rutabulum*.

As more data from liverworts are available today, a reassessment of the folding pattern has been made to fully incorporate all information into a structural estimation of nad5i753 exclusively for liverworts (Fig. 3-2).

The splicing of group I introns requires two highly conserved nucleotides: a U preceding the 5' splice site, and a G preceding the 3' site (Burke 1988). Both nucleotides are present in nad5i753 (marked in bold letters, figure 3-2).

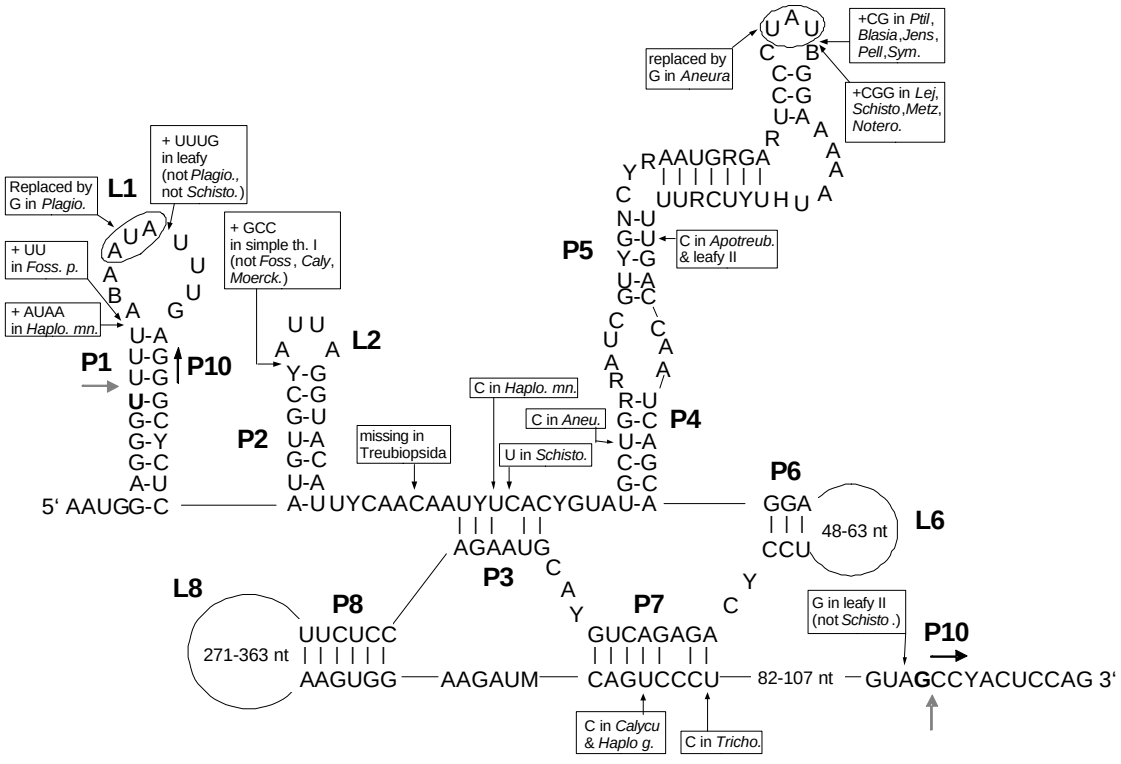

Fig. 3-2: Secondary structure of group I intron nad5i753 in liverworts. Grey arrows indicate splice sites. Intron sequence is based on a consensus sequence of 51 liverworts, folding pattern adapted from Beckert et al. (1999). Length variations, mismatches restricted to few taxa, and other notable exceptions are included in boxes, with the respective taxa designations shortened from species names in table 2-1.

The regions designated as L6 (L = loop) and L8 are very variable in length and folding structure, and were therefore not included in the consensus figure. Correlating this updated folding pattern with the one published in Beckert et al. (1999), most of the core nucleotide sequences are identical. Nevertheless, novel indels were now identified in L1, L2, and P5 (P = pair, Fig. 3-2). The single nucleotide indel between P2 and P3 is unique to the three Treubiopsida species (Fig. 3-2). Interestingly, this taxonomically difficult liverwort group is considered to be closely related to another unusual liverwort class, the Haplomitriopsida, which is also included in this study with two taxa, but does not show this particular indel. The occurrence of multiple indels is frequent in group I introns, and usually does not collide with the correct splicing process (Burke 1988; Quandt and Stech 2005 and therein).

P3 and P7 each exhibit one highly conserved C/A mismatch, which could indicate potential editing sites. The C would in that case be edited to a U, and reconstitute a U-A base pairing. However, these mismatches are conserved in all liverworts including the complex thalloids, which are incapable of RNA editing and are not able to repair these particular mismatches.

One genus in this comprehensive study exhibited a distinctly different sequence: the simple thalloid genus *Pellia*. The sequence from *Pellia endiviifolia* was the only one included in the taxa selection for the phylogenetic approach, but sequences from three species were available for the reconstruction of the intron folding pattern: in addition to *P. endiviifolia*, *P. appalachiana* (1396 bp, Genbank Accession: AY688762) included a complete intron sequence, albeit with several ambiguities towards the 3' end, and one *P. epiphylla* sequence (1611 bp, AY688764) was lacking only very few nucleotides of the intron, also on the 3' side.

The reconstruction of the secondary structure reveals a nucleotide sequence for the first domain P1/L1 (Fig. 3-3) that differs from the liverwort consensus sequence (Fig. 3-2) in all three *Pellia* species, comprising a different, larger loop L1 in *Pellia appalachiana*, and an additional hairpin structure in *P. epiphylla* and *P. endiviifolia*. The second domain consisting of P2 and L2 exhibits a shape rather similar to the consensus, but the actual nucleotide sequence is almost completely different, and these differences are conserved in all *Pellia* species (Fig. 3-3). In addition to that, the folding structure and sequence of P5 displays a succession of variants (Fig. 3-3), with a different sequence

but rather consensus-like structure in *Pellia appalachiana*, a more pronounced terminal stem in *P. epiphylla*, and the longest terminal stem in *P. endiviifolia*.

The regions that are variably sized in the consensus structure, L6 and L8, are peculiar in *Pellia* as well, as L6 is almost doubled in size, whereas L8 is comprised of roughly 30-40 % of the average number of nucleotides from to other liverworts.

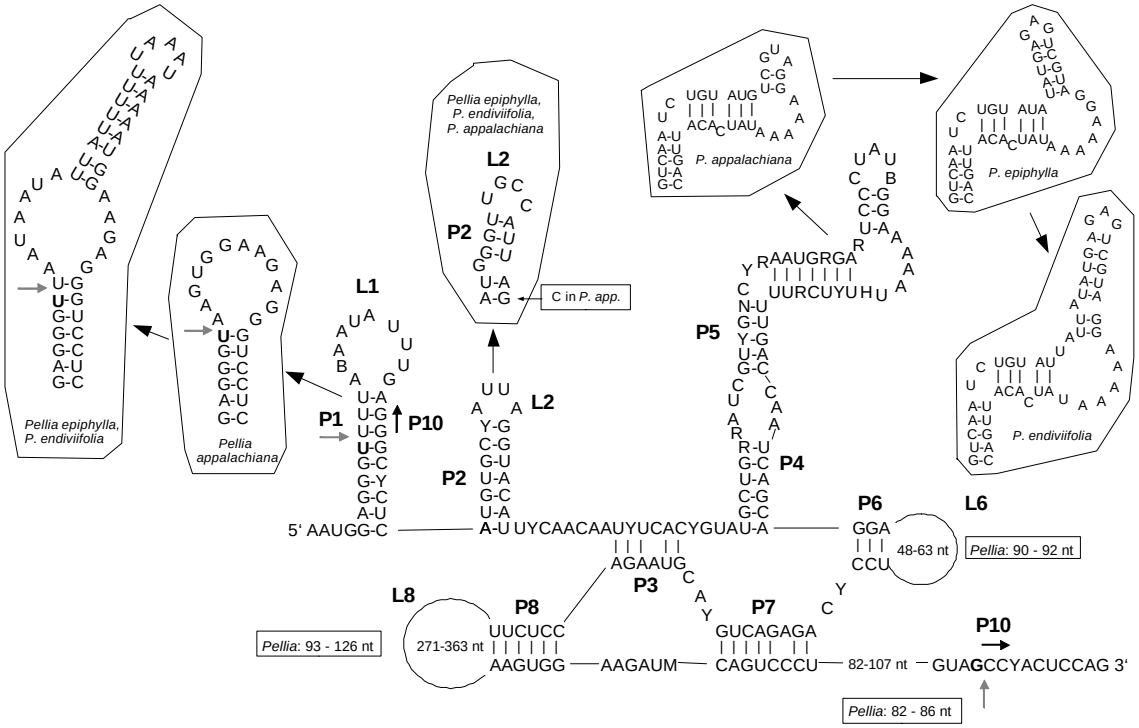

Fig. 3-3: Secondary structure of group I intro n nad5i753 in the liverwort genus *Pellia*. Grey arrows indicate splice sites. Intron sequence is based on a consensus sequence of 51 liverworts, folding pattern adapted from Beckert et al. (1999). *Pellia*-specific structure differences to the liverwort consensus are added for P1/L1, P2/L2, and P5. Nucleotide numbers of the three size-variable regions are given in boxes.

As much as the identified structural peculiarities suggest a rather basal position of *P. appalachiana* in a potential phylogeny of the genus *Pellia*, which then gains additional indels in further taxa, this topology is not retrieved in a study focussing on simple thalloid liverworts (Forrest and Crandall-Stotler 2005). A phylogenetic study of *Pellia* based on isozyme and nuclear *trnL*-spacers is available, including unfortunately only European taxa (Fiedorow et al. 2001). Additional sequencing, e.g. of the *nad5* gene intron from further *Pellia* species, could improve the knowledge about intron evolution in this unusual genus.

Interestingly, the effect of very reduced sizes of variable regions is not restricted to the one group I intron nad5i753 in *Pellia*. Unique patterns of evolution have also been

observed in the group II intron nad4i548 conserved in the mitochondrial *nad4* gene (see chapter 2), where the intron is reduced to 660 bp, compared with a median size of approximately 900 bp in other liverworts. A possible third example for *Pellia* is the group II intron which is exclusively conserved in the mitochondrial *trnS* of liverworts, a locus that has been partially sequenced by Davis (2004) for phylogenetic studies.

The second and only other genus of the family Pelliaceae is *Noteroclada*, which has been included in this study. It sequence shows none of the patterns that were observed in *Pellia*.

To confirm the correct splicing of the intron nad5i753 in *Pellia* despite of the different secondary structure (Fig. 3-3), cDNA analyses are necessary. It would also be possible to postulate the occurrence of a second gene copy containing an intron similar to the other liverworts that exists in parallel to the observed version. However, no evidence has been found to support this latter hypothesis.

3.2.3 The *nad5* gene as a phylogenetic marker in liverworts

3.2.3.1 Material and methods: Sequence selection

The *nad5* gene is the longest protein coding region of the land plant chondriome, with approximately 670 encoded amino acids. The potential of the mitochondrial *nad5* gene as a locus for phylogenetic analyses in bryophytes has been shown for the first time by Beckert et al. (1999). These analyses have been based on sequences that were composed of approximately 1100 bp reading frame plus the intron nad5i753, which has a size of an additional 700 bp, the so called K-L region (Fig. 3-4). This region was chosen due to the fact that two trans-splicing introns disrupt the continuity of the gene in angiosperms on the 3’ side of the selected portion (Fig. 3-4). During the last two years the locus was additionally exploited for liverworts by Davis (2004) and Forrest (Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005). The sequences obtained by Forrest et al. were focused on a broad taxon sampling in a taxonomically difficult subgroup of the liverworts, the simple thalloids, as mentioned before. The data produced by Davis were more widely sampled in a taxonomical sense, but in all cases the maximum extension of the sequences equalled the K-L region. It was even often restricted to sequences of the intron, as in many cases of the data obtained by Davis. In a recent study (Groth-

Malonek et al. 2005) the previously analysed part of the gene was extended to the U-D region (Fig. 3-4), because no trans-splicing was detected in liverworts, as well as none of the introns that were conserved in angiosperms. This led to an addition of approx. 700 bp to the former sequence length, increasing the amount of data available for phylogenetic studies.

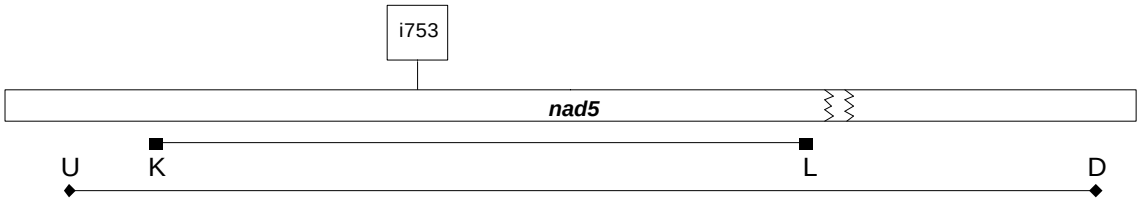

Fig. 3-4: **Graphical overview of the *nad5* gene in liverworts.** The group I intron is depicted as a square. Disruption of the gene in angiosperms through trans-splicing introns is marked by zig-zag-lines. K-L spans the region sequenced in previous studies (approx. 1800 bp), U-D spans the region analysed in Groth-Malonek et al. (2005) and this study (approx. 2500 bp). See figure 3-1 for comparison.

The analysis of all available data from the public database GenBank (NCBI) displayed significant differences in sequence length, ranging from 370 bp to 1800 bp. Preliminary trials to obtain a large scale taxon sampling for a comprehensive liverwort study on *nad5* revealed additional problems due to differences in the phylogenetic placement of sequences obtained by different labs from the same taxon, as for instance in the cases of *Monoclea*, *Ptilidium*, or *Porella*. Additionally, sequences for some taxa were available in parallel from different sources, but of differing quality, as they included for example unlikely frame shifts in the coding region (e.g. *Conocephalum*) that were not identified in other sequences of the same taxon, or clusters of ambiguities, which reduces the confidence in the correctness of the flanking sequence. Finally, empirical trials with different data sets showed that the length of the sequences is strongly correlated to the statistical support obtained for the backbone of the resulting topology.

Therefore a selection was made out of the 134 sequences available (combined count from published data deposited by the respective authors in Genbank, and data available from my studies). Duplicate sequences from the same taxon were reduced to one sequence, which was selected using the following criteria: 1) the smallest amount of ambiguous nucleotides and frame shifts, 2) the longest sequence, 3) the most likely placement in the phylogenetic tree compared to other molecular liverwort studies, if ambiguous. All sequences shorter than 1200 bp were excluded from the study. This

particular restriction was chosen by empirical weighting of sequence length versus taxon sampling. The maximum amount of taxa from the same genus was reduced to two, to prevent a weighting of the tree towards a single subgroup. This selection led to a data set of 51 liverworts including all major subgroups with a reasonable number of taxa (table 3-2).

For the selection of a suitable outgroup several combinations were evaluated. The primary testing of mosses as an outgroup was an obvious choice, as they also share the intron nad5i753 and could therefore provide equal sized sequences (see chapter 2 for similar reflections about the *nad4* phylogeny). The four selected sequences from *Takakia*, *Sphagnum*, *Andreaea*, and *Tetraphis* were obtained from the K-L region of *nad5* (Fig. 3-4), and comprise taxa that are placed on basal branches in moss topologies. They were also used successfully as outgroup in other studies (e.g. Forrest and Crandall-Stotler 2004). Other potential taxa for the outgroup were the algae *Chara* and *Chaetosphaeridium*, whose chondriome sequences are completely available in Genbank (Turmel et al. 2002; Turmel et al. 2003). Including them in the data set resulted in a reduction of the coherence of the overall topology. This lead to a pull-down of *Pellia* to the basis of the liverworts, when exon and intron sequences were used for the phylogenetic approach, or alternatively the loss of any support for the Marchantiopsida cluster (complex thalloids), when intron-less sequences were analysed. This occurred independently of the inclusion or exclusion of the selected mosses, and was accompanied by an increase of the number of supported nodes when no algae were included in the data set and/or intron and exon sequences were analysed, respectively.

Therefore the above mentioned four mosses were chosen as the best fitting outgroup for this phylogenetic approach on liverworts, based on the mitochondrial *nad5* gene data selected before.

3.2.3.2 Results and discussion of the *nad5* liverwort topology

The selected data set was phylogenetically analysed using a Baysian approach (see Mat. & Meth.), which resulted in the topology presented in figure 3-5 (see also chapter 2 for a similar approach on *nad4* data). The backbone of this tree is very strongly supported, presenting a clear division between the mosses and the liverworts. The basal-most clade of the liverworts (Figs. 3-5) is represented by a combined cluster of the Haplomitriopsida (two taxa) and the Treubiopsida (three taxa).

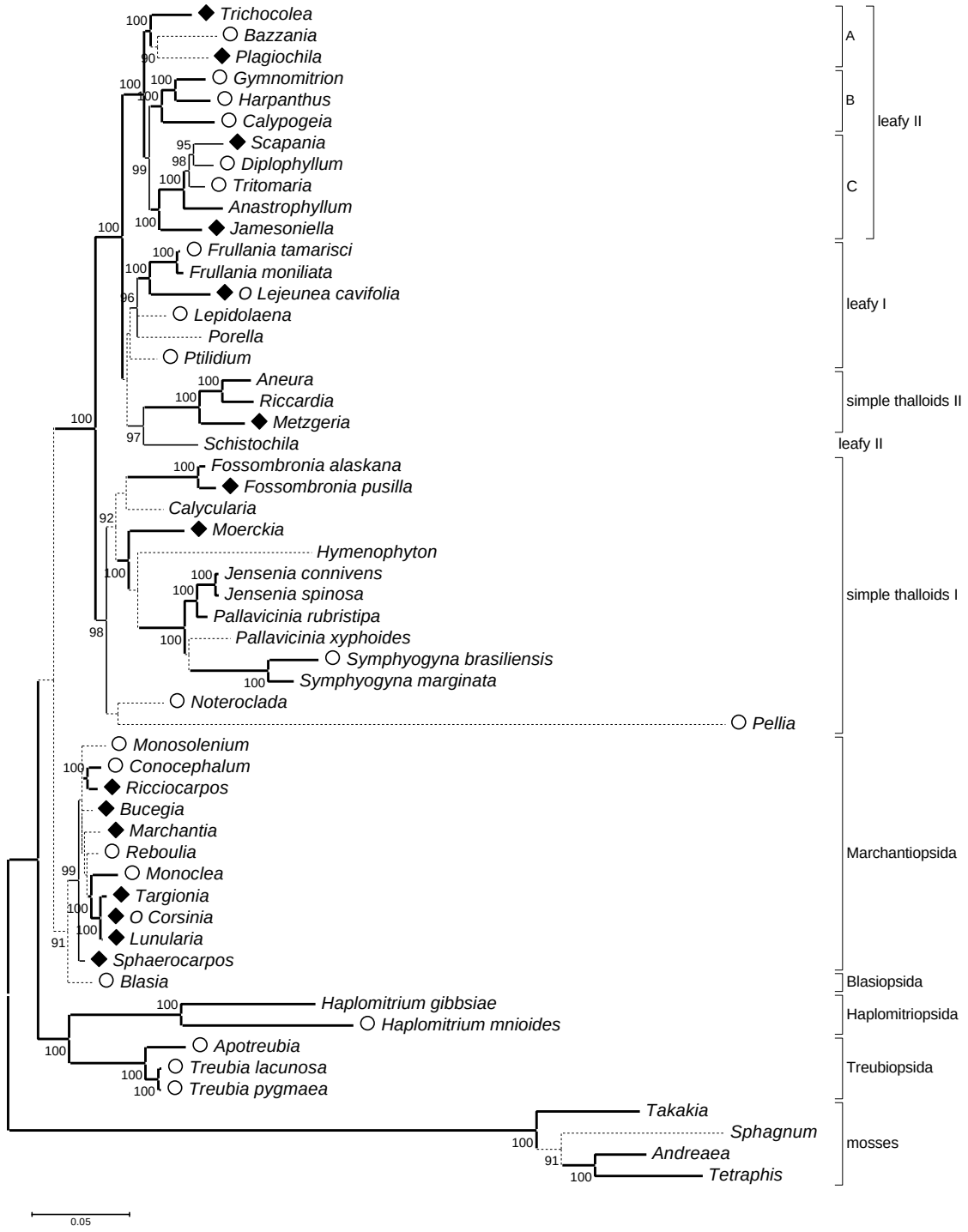

Fig. 3-5: Phylogenetic tree based on the mitochondrial *nad5* gene in liverworts. Baysian Posterior Probabilities are given on the respective nodes if exceeding 90 %. Thick lines indicate strong statistical support of the topology (PP=100), thin lines denote moderate support (95<PP>99), dotted lines circumscribe nodes that are not statistically supported. The same topology was obtained by Neighbour joining method, albeit with sometimes weaker support of the nodes by Bootstrap values. Taxonomic designations follow Davis et al. (2004) and Frey and Stech (2005). Black diamonds accompany taxa that were present in Beckert et al. (1999), open circles indicate sequences that span the U-D region, unmarked taxa were taken from Genbank.

The combined placement of these liverwort groups on the basis of all liverworts has been proposed before based on chloroplast data (Forrest and Crandall-Stotler 2004) and with a three-genome-approach (Forrest and Crandall-Stotler 2005). The placement of *Haplomitrium* without Treubiopsida varied in different studies, with a placement on the basis of all liverworts (morphology, Schuster 2000) or on the basis of the simple thalloids (review, Crandall-Stotler and Stotler 2000), or inside the simple thalloid I clade (chloroplast, He-Nygren et al. 2004). It is not unequivocal whether the combined placement of the two classes on the basis of the topology could result from long-branch attraction of the Haplomitriopsida towards the Treubiopsida. However, all three taxa from the Treubiopsida clade present rather moderate branch lengths (Fig. 3-5). A 6 bp indel in the variable region between P7 and P10 of the intron nad5i753, which is exclusively conserved in Haplomitriopsida and Treubiopsida, additionally supports the notion of a genuine sistergroup relationship, but is not unequivocal in its origin (Fig. 3-6).

complex thalloids:	A----------TA
or alternatively:	AA----------TA
Blasia:	AAA----------TA
leafy I: only *Porella* and *Lejeunea*	----------TA
leafy II without *Trito.*, *James.*, *Schisto*	----------TATA
leafy II: only *Bazz.*, *Gymno.*	---------TATATA
Haplomitrium gibbsiae	---TATACA------
Haplomitrium mnioides	---TCTACG------
Apotreubia, *Treubia lac.*, *Treubia pyg.*	---TAGACA------
all other taxa, incl. mosses:	---------------

Fig. 3-6: Section of the alignment of the intron nad5i753 in liverworts. Dashes stand for gaps. Taxa see table 3-2.

The second basal clade of the *nad5* phylogeny (Fig. 3-5) is composed of *Blasia*, the single member of the “simple-thalloid-like” Blasiopsida, and the complex thalloid liverworts (Marchantiopsida). Their molecular connection has been recently proposed by different studies (e.g. Wheeler 2000; Davis 2004), but obtained no statistical support for its suggestive placement in this topology (Fig. 3-5). The complex thalloid liverworts are well supported as a separate clade, but internal nodes are mostly lacking.

The upper part of the tree is composed of the Jungermanniopsida (Fig. 3-5), which include the morphologically rather different groups of the simple thalloid (Metzgeriidae) and the leafy liverworts (Jungermanniidae). The distiction of these subgroups is sometimes problematic, as the morphological traits include simple thalloid vegetative phases in some Jungermanniidae, and leafy, or nodal, morphology in some Metzgeriidae (Crandall-Stotler et al. 2005 and therein). Molecular studies, which are usually based on loci independent from morphological characters, revealed a partitioning of the simple thalloid liverworts into a paraphyletic lineage composed of two groups, the so called simple thalloids I and II (Davis 2004). In the phylogeny derived from the mitochondrial *nad5* gene (Fig. 3-7), members of both groups were analysed. Jungermanniopsida as a whole were identified as a strongly supported monophyletic clade, which is divided into two subclades, placing the simple thalloids I as sistergroup to the rest of the Jungermanniopsida. The simple thalloid I clade includes both genera of the Pelliaceae, *Noteroclada* and *Pellia*, albeit with no statistical support for a definite placement as sister taxa (Fig. 3-7). *Pellia* is placed on a very long branch, which is likely due to the singular sequence variety of this species concerning the intron nad5i753 (see chapter 3.2.2). A comparative phylogenetic approach based solely on exon sequences led to a decrease of the overall statistical support of the otherwise similar topology (and is therefore not shown here), but exhibited a well supported combined placement of *Pellia* and *Noteroclada* due to strong exon similarities.

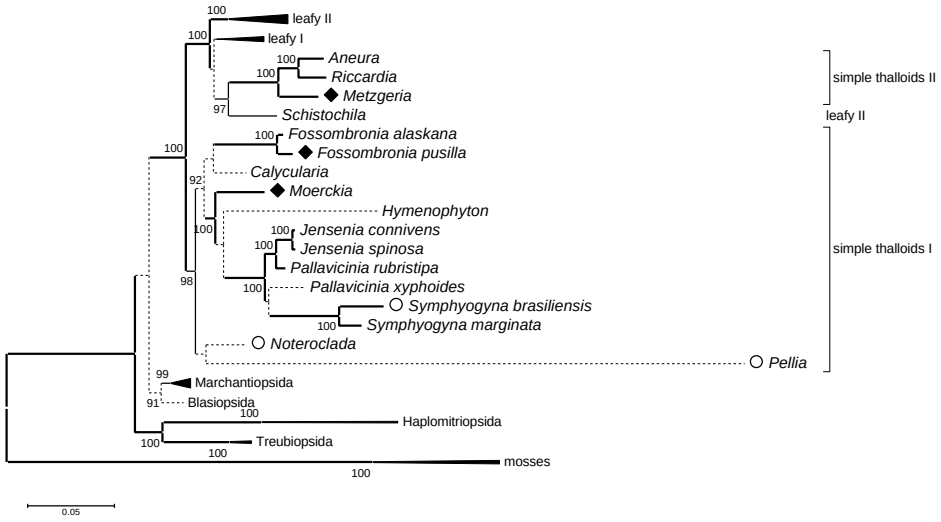

Fig. 3-7: Liverwort phylogeny derived from the mitochondrial *nad5* gene, with leafy and complex liverworts, Haplomitriopsida, Treubiopsida, and mosses as collapsed clades for enhanced visibility, full tree presented in figure 3-5.

The second group of the simple thalloids is nested with strong support in the leafy liverworts clade. This tripartite cluster is composed of the leafy clades I and II and the simple thalloid II, with unclear relation to each other. *Schistochila* is placed with moderate support on the basis of the simple thalloid clade (Fig. 3-5), although it is of leafy appearance and has been shown to reside on the basis of the leafy II clade (or its equivalent) in several studies (Davis 2004; Forrest and Crandall-Stotler 2004; He-Nygren et al. 2004). *Ptilidium*, the proposedly basal-most taxon of the leafy I clade, is also not supported in its placement (Fig. 3-5). The phylogenetic analysis based on *nad5* data is probably not high enough to resolve these nodes. Another study restricted to *rbcL* data exhibits the same lack of statistical support (Heinrichs et al. 2005).

The leafy clade I is composed of three clades, annotated as A, B, C (Fig. 3-5) following the proposal of Davis (2004). These subclades are highly supported and confirm earlier findings of these clusters, as has also been shown by *nad4* data (for detailed analysis see chapter 2: The *nad4* gene in liverworts).

3.2.3.3 Extended taxon sampling and a multi-gene approach

Rather few taxa were sampled from some clades, e.g. leafy IIA or the simple thalloids II. Additionally, the placement of taxa like *Schistochila* and *Ptilidium* is no unequivocally resolved (Fig. 3-5). As the addition of taxa to the *nad5* data set resulted in the reduction of statistical support due to the inclusion of rather short sequences, an enhanced taxon sampling was combined with the addition of three further molecular loci: the mitochondrial *nad4* gene (chapter 2), the chloroplast gene *rbcL* (ribulose-bisphosphat-carboxylase, large subunit), which was tested independently for liverwort phylogeny (Heinrichs et al. 2005), and the chloroplast *rps4* gene (small ribosomal subunit protein 4), which also proved to be useful for this approach (Knoop, in revision).

The extended taxon sampling included 63 liverwort species that were phylogenetically analysed, using a Bayesian approach implementing independent modelling of six different partitions of the data set (Tabs. 3-1, 3-2). This resulted in a highly supported backbone that supports most of the previously proposed liverwort groups, including the paraphyletic simple thalloids I and II (Fig. 3-8). Only the leafy I group is based on a weak node.

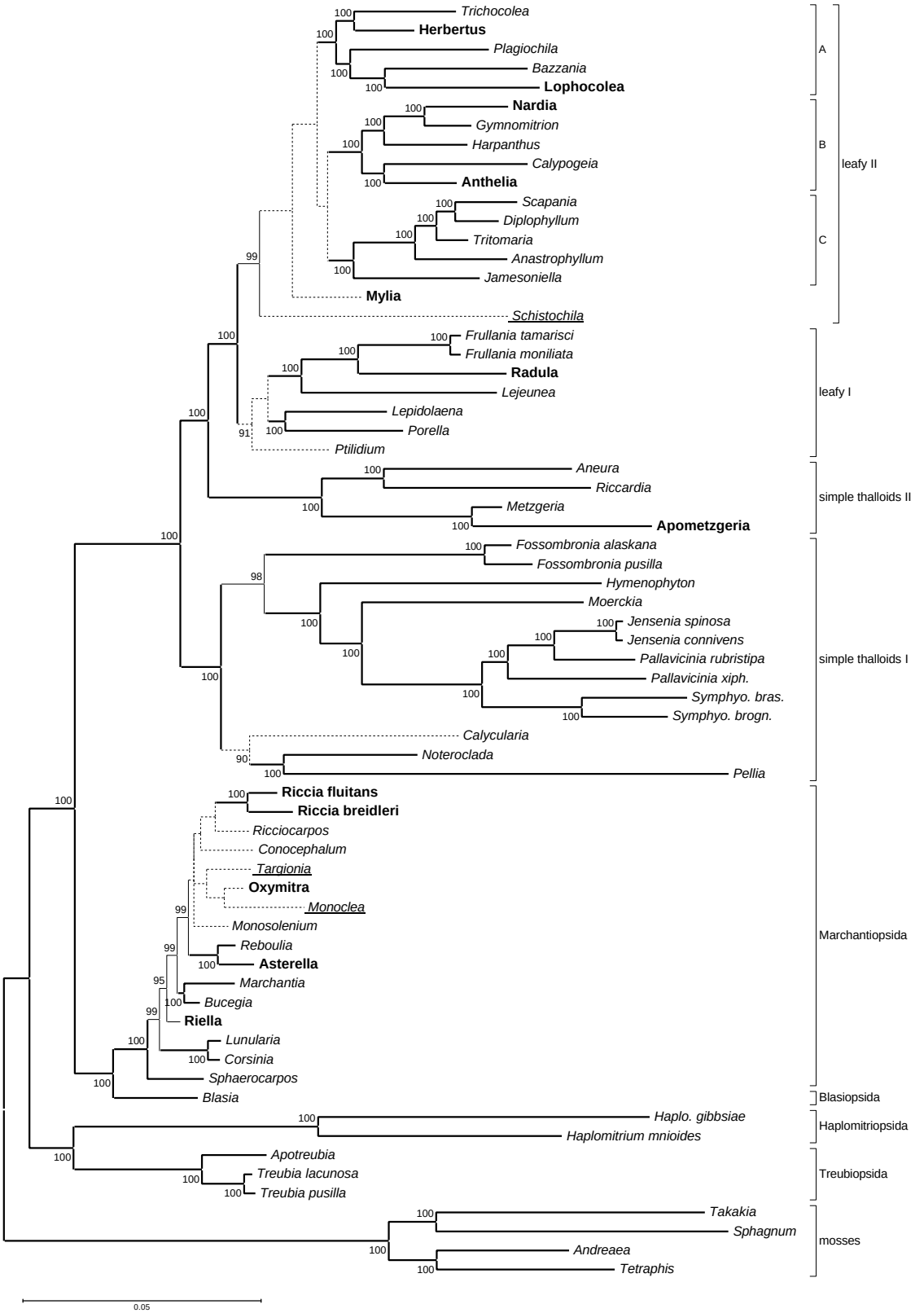

Fig. 3-8: Phylogenetic tree based on *nad5*, *nad4*, chloroplast *rps4* and *rbcL* in liverworts. Baysian Posterior Probabilities are given on the respective nodes if exceeding 90 %. Thick lines indicate strong statistical support of the topology (PP=100), thin lines denote moderate support (95<PP>99), dotted lines represent branches that are not statistically supported. Bold straight taxa are added to the taxon sampling of figure 3-5, underlined species are supported in a different position compared to figure 3-5.

Additional support was also gained in the Marchantiopsida clade (Fig. 3-8), where the addition of five species and, more importantly, at least 2000 bp additional data resulted in an internal backbone that has not been present in the *nad5* phylogeny. This is most likely based primarily on the *nad4* data which were shown to enhance support for this group (see chapter 2).

Better support was also gained for the simple thalloid I clade, where *Calycularia* is the only taxon without at least moderate support. Two species from the genus *Pallavicinia* are clearly paraphyletic (Fig. 3-8). This concurs with a study previously presented by Crandall-Stotler et al. (2005) that included three species of this genus, which were resolved in a polyphyly. The mentioned study focuses on simple thalloid liverworts in both morphological and molecular analyses, but does only suggest a broader taxon sampling for this problematic situation. This genus should probably be subject of a morphological review, as molecular data do not support it as a monophyletic group. A second possibility could also be a mix-up of DNA or sequences prior to the analyses, as they were not produced independently but from the same laboratory. Independent analyses, including different plant vouchers, should therefore be added.

The topology of the rest of the Juntermanniopsida is very well resolved with few exceptions (Fig. 3-8). One of them is the still unclear position of *Ptilidium* to the clades leafy I and leafy II, with a slight tendency towards the leafy I clade (Fig. 3-8). The other one is the now different placement of *Schistochila*, which has been located close to the simple thalloid clade II in the *nad5*-only phylogeny (Fig. 3-5). This taxon is now placed in a sistergroup relationship with the tripartite crown group of the leafy II clade, which reflects the placement of this species previously proposed by other studies (Davis 2004; He-Nygren et al. 2004), albeit complemented with the intermediately placed taxon *Mylia*, that is as yet only represented by *nad4* data in this analysis. Its position was also not resolved in an phylogenetic approach restricted to *nad4* data (chapter 2).

3.3 Discussion of *nad5* as a useful marker for land plant phylogeny

The here presented study was aimed firstly as a review of the so far available publications based on the mitochondrial *nad5* gene in all land plants (chapter 3.1), and

secondly as a closer look into its development as a standard liverwort marker, where it represents the singular mitochondrial locus investigated on a larger scale so far except for the *nad4* gene introduced with the studies in this work (chapter 2). The revision of the secondary structure of the included intron nad5i753 could assist in future studies of the gene, to prevent alignment problems of such unusual examples like the observed *Pellia* sequences.

Apart from additional *Pellia* species, which would be of interest due to their peculiar intron features, some interesting taxa are still missing from the data set, whose placements in liverwort phylogenies are still unresolved. One of these cases is the genus *Pleurozia*, which has been morphologically placed as a highly isolated lineage of the Jungermanniidae (Crandall-Stotler and Stotler 2000), but was surprisingly resolved as sister to the simple thalloid II clade in recent molecular studies (Davis 2004; He-Nygren et al. 2004). Few morphological characters, like the lenticular apical cell, do indeed link *Pleurozia* to this clade, but most features are rather similar to Jungermanniid characters (Crandall-Stotler et al. 2005 and therein). For *Pleurozia* there is already a *nad5* sequence available, but its inclusion into the data set resulted in no conclusive placement, probably due to its restricted size of only 852 bp.

A very interesting taxon would also be *Cavicularia*. It constitutes a sister genus to the isolated *Blasia*, which is placed on the basis of all Marchantiopsida (complex thalloids), although its appearance resembles a simple thalloid liverwort (see also chapters 2, 4, and 5). The highly supported, but isolated placement of *Blasia* could be confirmed with an additional supporting branch, provided its sister taxon is included in the phylogeny.

Leafy liverworts are a very speciose group, comprising at least 5500 species worldwide (Frey and Stech 2005); many of them are rare and endemic. Therefore it would be impossible to sample all leafy taxa that could potentially be of interest for liverwort phylogenies. Nevertheless, a couple of isolated genera were sampled that are as yet unsatisfactorily resolved by the presented phylogenies. These taxa are usually placed basal to well supported subclades, like *Ptilidium*, *Schistochila*, and *Mylia* in the here presented phylogeny based on *nad5* data. Other genera of unresolved relation would be *Adelanthus*, *Syzygiella*, and *Jamesoniella*, which tend to cluster together basal to the leafy II B clade, although they are morphologically positioned in different families.

From leafy I (Porellales) there is also a very rare taxon missing, *Jubula*, that should be resolved as sister to the Lejeuneaceae, according to morphological studies (Weis 2001).
Even if used as the single marker gene for a phylogenetic approach, the *nad5* gene can provide enough information to obtain a strong backbone phylogeny, and also support several internal nodes, in the probably earliest diverging land plants, the liverworts.
The obtained topology is highly congruent to comparable molecular liverwort phylogenies (Davis 2004; He-Nygren et al. 2004; Forrest and Crandall-Stotler 2005), including the well supported paraphyly of the Metzgeriidae (simple thalloid) taxa. This is especially true when not only the intron, but also major parts of the exons are included in the data set. In a combination with other loci, it provided a very well resolved topology of liverworts, and should probably be included in all future phylogenetic studies of large scale approaches concerning this plant group.

4 Evolution patterns of mitochondrial DNA: highly divergent development of intergenic regions (spacers) in bryophytes

4.1 Introduction

Plant mitochondrial DNA is known for its highly conserved exon sequences but a much less conserved structure regarding the arrangement of its genes (Palmer and Herbon 1988). These observations were mostly derived from the comparison of angiosperm chondriome sequences or algal counterparts, but only one complete sequence of a non-angiosperm land plant is known, the chondriome of the liverwort *Marchantia polymorpha* (Oda et al. 1992a). This study aims at the understanding of chondriome evolution in lower land plants, and concentrates on two gene clusters: first, the combination of two tRNAs, potentially flexible and small genes that are located upstream of an important and highly conserved *nad* gene; second, the physical linkage of three *nad* genes that are proposed to be conserved on the mitochondrial DNA of all land plants (Knoop 2004 and therein), in contrast to tRNAs that can be lost in favour of nuclear encoded genes or import of chloroplast alternatives. Both gene clusters are known to be disrupted in several angiosperms (e.g. Unseld et al. 1997).

4.2 The gene cluster *trnA-trnT-nad7*: conservation vs. loss of a tRNA in bryophytes

The mitochondrial gene cluster *trnA*(UGC)-*trnT*(GGU)-*nad7* is conserved in the Charophyte algae *Chara vulgaris* (Turmel et al. 2003) and *Chaetosphaeridium globosum* (Turmel et al. 2002). *TrnA* encodes the tRNA for alanine (anti-codon UGC), *trnT* the one for threonine (anti-codon GGU). *Nad7* encodes subunit 7 of the NADH ubiquinone oxidoreductase, complex I of the mitochondrial respiratory chain. The same gene continuity is found in the moss *Physcomitrella patens* (Fig. 4-1).

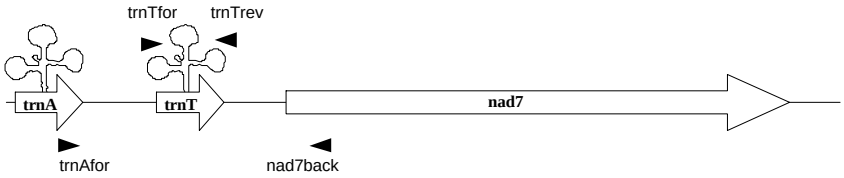

Fig. 4-1: Gene arrangement of the mitochondrial gene cluster *trnA-trnT-nad7* in green algae and *Physcomitrella*. White arrows indicate the orientation of the respective genes on the chondriome; black arrows indicate the location of the primers used for PCR amplification assays.

An identical gene order was identified in the liverwort *Marchantia polymorpha* (Oda et al. 1992a), with the extraordinary exception that the *trnT* gene is arranged in the opposite orientation. Nevertheless, it is indeed the same gene encoding for the same *trnT*(GGU), as can be seen by alignment of the sequences and comparison of the anticodon (fig. 4-2). This gene cluster is not existent in angiosperms, because none of the fully sequenced chondriomes of other land plants contains a homologue of either of both tRNAs (Unseld et al. 1997; Kubo et al. 2000; Notsu et al. 2002; Handa 2003; Sugiyama et al. 2005; Ogihara et al. 2005).

4.2.1 Material and Methods

DNA extraction and PCR assay strategies see chapter 2.2.1 “General methods and strategies”. Taxa used for PCR assays and additional sequences that were used for correlation and alignments are shown with their respective accession number from Genbank (NCBI), and/or their respective sequence length (tables 4-1 and 4-2). Primers used for the DNA assays are published in (Beckert et al. 1999).

Taxonomic classification follows recent literature for mosses (Goffinet and Buck 2004) and for liverworts (Crandall-Stotler and Stotler 2000; Frey and Stech 2005). DNA and RNA were differentially precipitated in the presence of 3 M lithium acetate. OmniScript^{TM} Reverse Transcriptase (Qiagen) was used for cDNA synthesis.

Primer sequences used for PCR of the gene cluster *trnA-trnT-nad7* were trnAfor (5’- tcg gtt caa vtc cga tcg tt cca - 3’), nad7back (5’- acc atg agc agc wgg rtg ttg agg - 3’), trnTfor (5’- cat ggt aag gga aag gtc tcc -3’), and trnTrev (5’- gga ggc ctt tcc ctt acc atg - 3’).

Table 4-1: Taxa analysed for the mitochondrial *trnA-trnT-nad7* gene cluster

Taxonomy	Species	Voucher number or DNA signature	Accession number
green algae	*Chara vulgaris* L.	--	NC_005255
LIVERWORTS	*Herbertus sendtneri* (Nees) Lindb.	J. Heinrichs 4377	870 bp
Jungermanniopsida / **leafy**	*Anthelia julacea* (L.) Dumort.	J. Heinrichs s.n.	897 bp
	Porella platyphylla (L.) Pfeiff.	J. Heinrichs 4383	581 bp
Jungermanniopsida / **simple thalloid I**	*Fossombronia pusilla* (L.) Nees	Ulm-collection s.n.	872 bp
	Noteroclada confluens Taylor ex Hook. & Wilson	live culture Goettingen	911 bp
simple thalloid II	*Metzgeria furcata* (L.) Dumort.	J. Heinrichs 4384	891 bp
Marchantiopsida / **complex thalloid**	*Marchantia polymorpha* L.	--- / Grewe s.n.	NC_001660/ confirmation-seq.: 1911 bp
	Bucegia romanica Radian	Ulm-collection s.n.	1926 bp
	Lunularia cruciata (L.) Dum. ex Lindb.	Groth & Schwertfeger s.n.	650 bp (only 5' seq.)
	Corsinia coriandrina (Spreng.) Lindb.	Ulm-collection s.n.	1307 bp (only 3' seq.)
	Asterella blumeana (Nees) Pandé Srivastava et Khan.	MGM031218-06SC	699 bp
	Conocephalum conicum (L.) Underw.	Groth & Schwertfeger s.n.	707 bp
	Monoclea gottschei Lindb.	live culture Goettingen	700 bp
	Oxymitra incrassata (Brotero) Sérgio & Sim-Sim	MGM031218-11SC	710 bp
	Riccia breidleri Steph.	ML-030826	709 bp
	Targionia hypophylla L.	Ulm-collection s.n.	709 bp
	Monoselenium tenerum Griff./Sunita Kapila & SS Kumar	live culture Goettingen	704 bp
Haplomitriopsida/ Treubiopsida	*Treubia lacunosa* (Colenso) Prosk. lenta Taylor ex Prosk.	LF28 / Stotler&Crandall-Stotler 4561 (ABSH)	411 bp
MOSSES	*Dawsonia* spec.	Pruchner s.n.	632 bp
Polytrichopsida	*Pogonatum urnigerum* (Hedw.) P. Beauv.	Muhle170997-15	627 bp
Takakiopsida	*Takakia lepidozioides* Hatt. et H. Inoue	Qiu97126	553 bp
Bryopsida Funariidae	*Physcomitrella patens* (Hedw.) Bruch & Schimp.	--- / Ulm-collection s.n.	AB035479 / confirmation-PCR: 579 bp
Dicraniidae	*Ditrichum cylindricum* (Hedw.) Grout	Muhle281097-2	579 bp
Bryidae	*Mnium hornum* Hedw.	Muhle090897-2	443 bp
	Pohlia nutans (Hedw.) Lindb.	Muhle090897-4	430 bp
	Homalia trichomanoides (Hedw.) Schimp.	Muhle291197-3	572 bp
Hypnidae	*Hygrohypnum ochraceum* (Turner ex Wilson) Loeske	Muhle191197-4	572 bp
	Leskea polycarpa Hedw.	Muhle231197-1	336 bp

s.n. = sine numero (lat.), "without number", meaning here without explicit collection or voucher number

Phylogenetic analyses of complex thalloid liverworts were carried out by Bayesian Inference approach: estimation of appropriate models for each partition of the data set by Modeltest 3.7 (Posada and Crandall 1998) using the Akaike Information Criterion (AIC), and implementation in MrBayes 3.1 with the following parameters: all partitions unlinked, two independent runs with: four parallel chains, one heated chain, 500.000 generations, every 100th tree sampled, burnin set to 4.000 trees, taxa used for the study see table 4-2, model parameters see table 4-3.

Table 4-2:

Taxa and accession numbers for Marchantiopsida (complex thalloid liverworts) phylogeny

Taxonomy	Species	nad4	nad5	rbcL	rps4
Blasiopsida (outgroup)	*Blasia pusilla*	2254 bp	2444 bp	AF536232 1347 bp	AY507436 603 bp
Marchantiidae	*Marchantia polymorpha*	NC_001660 2387 bp	NC_001660 2682 bp	X04465 1428 bp	X04465 609 bp
Marchantiales	*Bucegia romanica*	2313 bp	AJ001031 1794 bp	n.a.	n.a.
Marchantiineae	*Lunularia cruciata*	AJ310803 2151 bp	AJ001002 1792 bp	U87077 1347 bp	AY688795 985 bp
	Asterella blumeana	2265 bp	n.a.	U87064 1347 bp (*A. tenella*)	n.a.
	Conocephalum conicum	2203 bp	2495 bp	AY688778 1353 bp	AY688791 725 bp
	Monoselenium tenerum	2258 bp	2499 bp	n.a.	n.a.
	Reboulia haemisphaerica	2172 bp	2490 bp	AY462326 1038 bp	AY688801 741 bp
Corsiniineae	*Corsinia coriandrina*	AJ310801 2151 bp	2492 bp	n.a.	n.a.
Targioniineae	*Targionia hypophylla*	2124 bp	AJ001001 1793 bp	AY507427 1353 bp	AY688805 1153 bp
Monocleales	*Monoclea gottschei*	2223 bp	AJ622814 2474 bp	AY507414 1343 bp	AY507455 549 bp
Ricciales	*Riccia breidleri*	2268 bp	n.a.	AY507422 1068 bp (*R. huebeneriana*)	AY507463 549 bp (*R. huebeneriana*)
	Riccia fluitans	AJ310802 2155 bp	n.a.	n.a.	AY608107 567 bp
	Ricciocarpos natans	2265 bp	AJ001032 1789 bp	U87089 1347 bp	AJ251062 815 bp
	Oxymitra incrassata	2295 bp	n.a.	n.a.	n.a.
Sphaerocarpidae	*Riella spec.*	2086 bp	n.a.	n.a.	n.a.
Sphaerocarpales	*Sphaerocarpos donnellii*	2210 bp	AJ001033 1797 bp	AY507425 1482 bp (*S. texanus*)	AY608110 580 bp (*S. texanus*)

n.a.: Sequence unknown, no data available from Genbank

Table 4-3: Models implemented in MrBayes for Marchantiopsida phylogeny

character set (partition)	model selected by AIC (modeltest 3.1)	parameters implemented in MrBayes
exons nad4	TVM+I	revmatpr = fixed(1.6572, 1.7718, 0.0486, 0.7804, 1.7718, 1.0000) statefreqpr = fixed(0.2540, 0.1755, 0.1775, 0.3930) pinvarpr = fixed(0.8515), ratepr = variable, nst = 6, rates = equal
intron nad4	K81uf+I	revmatpr = fixed(1.0000, 1.1822, 0.3570, 0.3570, 1.1822, 1.0000) statefreqpr = fixed(0.2777, 0.2206, 0.2885, 0.2132) pinvarpr = fixed(0.8042), ratepr = variable, nst = 6, rates = equal
exons nad5	TVM+I	revmatpr = fixed(3.8621, 3.6031, 0.2463, 1.6726, 3.6031, 1.0000) statefreqpr = fixed(0.2394, 0.1737, 0.1940, 0.3929) pinvarpr = fixed(0.7948), ratepr = variable, nst = 6, rates = equal
intron nad5	K81uf+I	revmatpr = fixed(1.0000, 1.0653, 0.4322, 0.4322, 1.0653, 1.0000) statefreqpr = fixed(0.3547, 0.2017, 0.2158, 0.2278) pinvarpr = fixed(0.7771), ratepr = variable, nst = 6, rates = equal
rbcL	GTR+I+G	revmatpr = fixed(2.1881, 3.2360, 1.0473, 2.5230, 8.4993, 1.0000) statefreqpr = fixed(0.2907, 0.1590, 0.2153, 0.3350), pinvarpr = fixed(0.4018) shapepr = exponential(0.3241), ratepr = variable, nst = 6, rates = gamma
rps4	GTR+I+G	revmatpr = fixed(1.7185, 3.5430, 0.3349, 0.9751, 2.2087, 1.0000) statefreqpr = fixed(0.4160 0.1250 0.1198, 0.3392), pinvarpr = fixed(0.4557) shapepr = exponential(0.7078), ratepr = variable, nst = 6, rates = gamma

4.2.2 Results for the *trnA-trnT-nad7* intergenic region

PCR assays of the region between *trnA* and *nad7* and the respective sequencing of the resulting PCR products identified *trnT* sequences in ten mosses and four liverworts, that are conform to the consensus cloverleaf folding pattern (Fig. 4-2). In all cases the same anticodon was conserved as GGU. Only few differences were found between the species, most notably a mismatch in the acceptor stem that occurs in some basal mosses and the liverwort *Marchantia*. The more common type of the base pairing observed in this acceptor stem is a combination of U and A, the respective mismatch a C and A combination, occurring in basal mosses and the liverworts that carry a *trnT* gene.

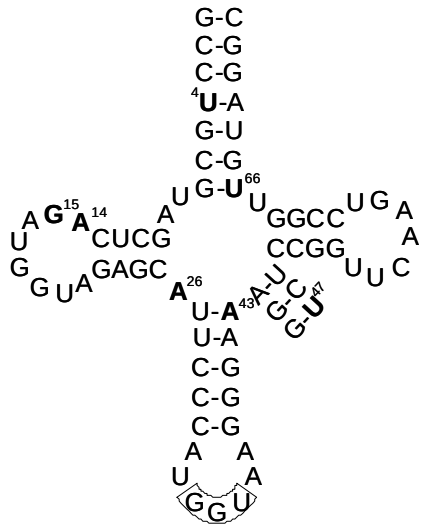

Fig. 4-2. The trnT(GGU) RNA structure. The anticodon GGU is boxed. The tRNA sequence as shown is identical in *Physcomitrella*, *Ditrichum*, *Leskea*, *Mnium*, *Pohlia*, and *Homalia*, differences in other species are indicated by nucleotide positions: nucleotid number 4 is exchanged from U to C in *Takakia*, *Marchantia*, *Bucegia*, *Corsinia*, *Lunularia*, nucleotid 14: A to G in *Hygrohypnum*, nucleotid 15: G to A in *Pogonatum*, and *Dawsonia*, nucleotid 26: A to G in *Takakia*, nucleotid 43: A to G in *Dawsonia*, nucleotid 47: U to C in *Marchantia*, *Bucegia*, *Corsinia*, and *Lunularia*, nucleotid 66: U to C in *Marchantia*, *Bucegia*, *Corsinia*, *Lunularia*, *Pogonatum*, *Takakia*, and *Dawsonia*.

The PCR assays on the *trnA-nad7* region and sequencing of the resulting PCR products revealed a loss of the *trnT* gene in all major groups of liverworts, with the exception of four taxa of the complex thalloid liverworts: *Marchantia* itself, accompanied by an additional taxon of the Marchantiaceae, *Bucegia*, and the two less closely related taxa *Corsinia* and *Lunularia* (Fig. 4-3).

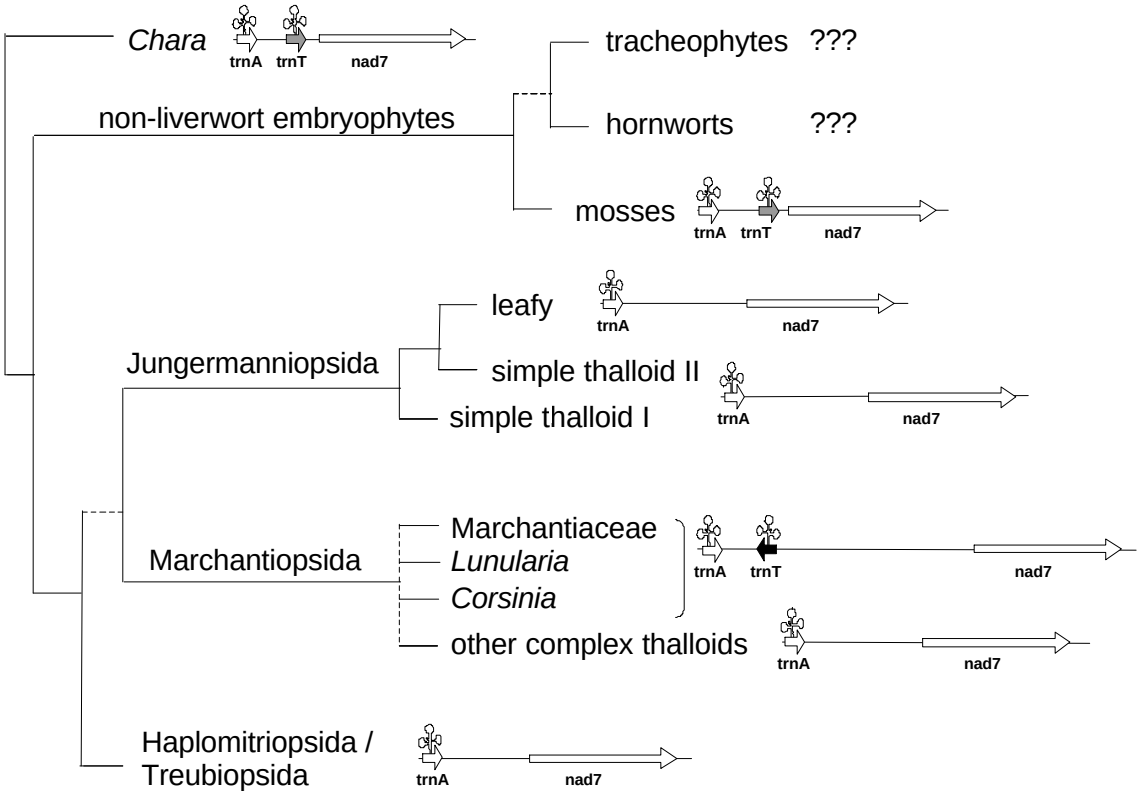

Fig. 4-3: Graphical overview of the gene cluster *trnA-trnT-nad7* in the alga *Chara* and in the investigated bryophytes, plotted on a potential phylogenetic tree derived from recent independent molecular analyses of liverwort relationships. Dotted lines indicate not clearly resolved branching events. Details for the analysed taxa are given in table 4-1.

As only a few taxa of one liverwort group (complex thalloids) carry the inverted *trnT* gene, a phylogenetic method was used to establish the relationship of those species with each other (fig. 4-4). Two mitochondrial (*nad5, nad4*) and two plastid genes (*rps4, rbcL*) were combined in a dataset and analyzed using a 6-model combination with a Bayesian Posterior Probabilities approach (see table 4-3 and methods).

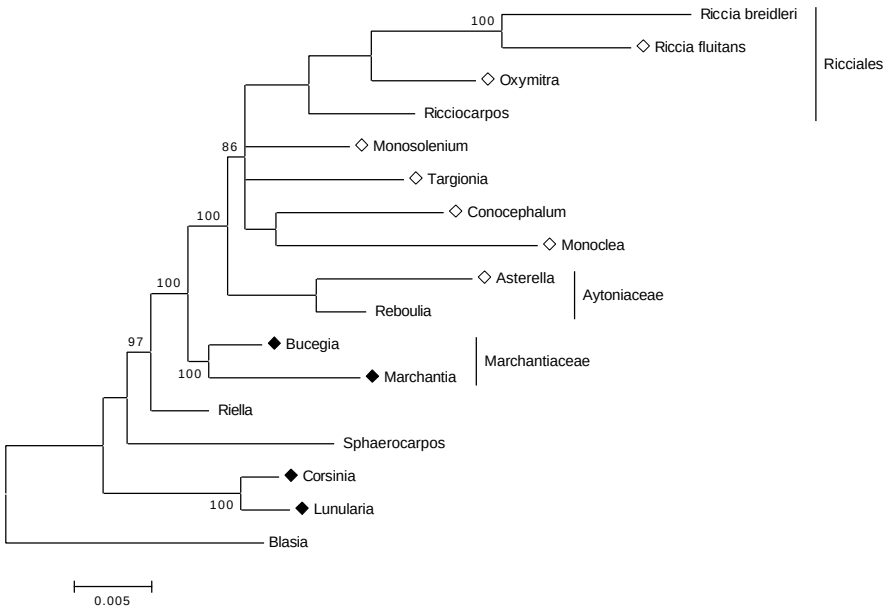

Fig. 4-4: Phylogeny of Marchantiopsida, inferred from the combined data set of *nad4*, *nad5*, *rbcL*, and *rps4*. Bayesian Posterior Probability values > 85% are indicated on the respective nodes, black diamonds indicate the presence of the inverted *trnT* gene, white diamonds refer to sequences without *trnT*. Taxa without diamonds were not sampled in this spacer study.

The Marchantiopsida group is shown here with *Blasia* as the basalmost taxon, as was established by other recent liverwort phylogenetic analyses based on molecular data (Wheeler 2000; Davis 2004; Forrest and Crandall-Stotler 2004). The complex thalloid group is known to be more conserved in its DNA sequences than other liverwort groups, and therefore its phylogenetic resolution in larger liverwort phylogenies is rather poor. This is the reason for the strictly complex thalloid taxon sampling for figure 4-4, as the inclusion of non-complex taxa reduces the resolution of the statistical support of the phylogeny.

No PCR products for the gene cluster *trnA-trnT-nad7* could be obtained from hornworts, a bryophyte group that is very distinct from liverworts and mosses, nor from any tracheophytes.

4.2.3 Discussion of the evolution of the gene region *trnA-trnT-nad7*

The removal of *trnT* from the gene cluster *trnA-trnT-nad7* seems to be restricted to liverworts. In fact, it could also be argued that mosses show a very close affinity to green algae in this case and therefore strengthen a theory of mosses as the earliest land

plants. Although the relationships of all bryophyte groups and tracheophytes are widely debated, this particular possibility has been clearly refused by several studies favoring either a monophyly of all bryophytes (e. g. Garbary et al. 1993; Nishiyama et al. 2004), or different constructs of a polyphyly with hornworts as the basalmost group (e. g. Malek et al. 1996; Nickrent et al. 2000; Renzaglia et al. 2000), or the most widely believed hypothesis of liverworts as the earliest land plants (Qiu et al. 1998; Kugita et al. 2003a; Groth-Malonek et al. 2005).

As no PCR products could be obtained from hornworts, the third bryophyte group, it is possible that the gene continuity is already disrupted here. On the other hand, hornwort DNA sequences are not easily deduced from homologous genes of other plants, because they exhibit an extraordinary amount of editing sites including the “reverse” editing from U to C and the more frequent plant organelle editing from C to U. For that reason, predictions of primer sequences from known DNA sequences or conserved amino acids are very unreliable for hornworts and often lead to a difficulty in amplifying PCR products. Missing data could therefore reflect a PCR assay problem instead of a disruption of the gene cluster. This could be verified by Southern blot analysis, where, in the case of a conserved gene continuity, probes for all three genes would then hybridize to the same restriction fragment on the blotting membrane. As tRNA genes provide only very short probes for hybridization and the amount of RNA editing in all three genes is not known, this experiment could be rather difficult to establish. A much more promising approach is the attempt to sequence a complete hornwort chondriome, as is already in progress in other laboratories.

But if the previously assumed theory is indeed correct and hornworts don not share the gene continuity of *trnA-trnT-nad7* with mosses and green algae, then this would be an indication for the following scenario: the common ancestor of the first land plants derived from Charales algae like *Chara* and inherited the whole intact cluster (Fig. 4-3). After the divergence of liverworts from the rest of the land plants (the non-liverwort lineage) the common ancestor of all liverworts experienced a rearrangement of this locus and transferred the *trnT* gene to a hitherto unknown location. The earliest diverging group of the non-liverwort lineage would then be the mosses, which share the intact gene cluster with *Chara* and the presumed ancestor of all land plants. After the divergence of the ancestor of all hornworts from the mosses, the gene cluster is

disrupted and completely lost (Fig. 4-3), also in the ancestor of the tracheophyte lineage that stands presumably in a sistergroup relationship with hornworts (Groth-Malonek et al. 2005).

Interestingly enough, both tRNAs are missing from all known angiosperm chondriomes. Their removal could have taken place as early as in hornworts, or sometime during the early evolution of tracheophytes like lycophytes or ferns. As none of the angiosperms are from an early lineage, as for instance *Amborella* or Nympheales, it could also be a loss restricted to a few angiosperms.

The insertion or possible inversion of the inversely oriented *trnT* gene, however, is presumably at least one independent gain in several complex thalloid liverworts. The phylogenetic relationships of complex thalloid liverworts are not clear due to a relatively small number of taxa and morphological characters, although several molecular approaches have been tried, including nuclear (*LSU*) rDNA genes (Boisselier-Dubayle et al. 1997; Boisselier-Dubayle et al. 2002), nuclear *18S rDNA* (Bopp and Capesius 1996; Capesius and Bopp 1997), the combination of *nucLSU* and chloroplast *trnL-trnF*-spacer (Wheeler 2000), or chloroplast *rbcL* (Lewis et al. 1997). The results were mostly weakly supported or contradicting. A recent approach involving 8 loci (five chloroplast, one mitochondrial, two nuclear located) combined previously used loci with a set of 12 taxa plus the reconsidered former simple thalloid liverworts *Blasia* and *Cavicularia* (Forrest and Crandall-Stotler 2005).

In conclusion of these studies a few assumptions were made: The two monospecific genera *Blasia* and *Cavicularia* form the basalmost group of the complex thalloid liverworts and should be used as an outgroup for further analyses. They are followed presumably by *Sphaerocarpos*, *Neohodgsonia*, and *Lunularia*, in unresolved relations. The next group is formed by the Marchantiaceae in a sister group relationship to an unresolved crown group that includes the Ricciales.

A study of the new locus *nad4* in liverworts (see chapter 2) indicated a relatively high amount of information in the complex thalloid taxa. Based on the available *nad4* taxon sampling, an additional dataset was obtained from three further genes (*nad5*, and the chloroplast genes *rbcL* and *rps4*). The combined data were then phylogenetically analysed (see material and methods). As shown in figure 4-4, the basalmost complex thalloid liverworts after the divergence of *Blasia* are indeed *Sphaerocarpos* and

Lunularia (*Neohodgsonia* was not sampled here), and *Corsinia*, the latter two in a highly supported sister relationship. *Sphaerocarpos* is a member of the subclass Sphaerocarpidae, which is sister to the subclass Marchantiidae. *Riella* is sampled here as a second member of the former group, and is placed on the next following branch with moderate support. This suggests a potential paraphyly in the Sphaerocarpidae, but the data are not conclusive here, especially as only *nad4* sequences were obtained from *Riella* (table 4-2). The only taxa that were sampled for the *trnA-trnT-nad7* study from this basal group of complex thalloids so far are *Corsinia* and *Lunularia*, which share the occurrence of the inverted *trnT* gene. Also no *Blasia* sequence could be obtained. Therefore it is not possible to ascertain whether the *trnT* occurrence is restricted to the subclass Marchantiidae, or whether it extends to the subclass Sphaerocarpidae or to the supposedly earliest diverging Blasiopsida.

The next diverging group of the complex liverworts are the two members of the Marchantiaceae, *Marchantia* and *Bucegia*. Both carry the *trnT* gene between *trnA* and *nad7*. The following crown group is composed of the two members *Asterella* and *Reboulia* of the family Aytoniaceae, which is strongly supported as a monophyletic group, and an unresolved combination of three members of the Marchantiales, four members of the Ricciales, and one member of the monogeneric order Monocleales. Seven of these ten taxa were sampled for this study, and none of them carries the *trnT* gene in its *trnA-nad7* spacer.

The overall topology of the phylogenetic tree in figure 4-4 is in congruence with former studies (mentioned above), although this particular and rather extended taxon sampling did not occur in any of these analyses. In conclusion of the results three scenarios are possible: 1) The re-insertion of *trnT* into the *trnA-nad7* spacer occurred at an early stage of the evolution of complex thalloid liverworts, possible after the divergence of *Blasia* or after the divergence of the Marchantiidae subclass, followed by a secondary loss later in the crown group. 2) If the phylogenetic studies were lacking sufficient taxon sampling to obtain a true version of the complex thalloid liverwort evolution, the group of the four *trnT* carrying taxa could in fact be monophyletic, and the gain of *trnT* would be a singular event in a distinct basal marchantiid subgroup. 3) The reinsertion of the *trnT* gene could be a frequent event in complex thalloid liverworts, which has not been discovered because of the low taxon sampling of 11 species out of the complex thalloid

liverwort group that comprises approximately 17 families, 32 genera and 350 species (Frey and Stech 2005).

At the moment the first scenario seems the most likely one as the topology of the phylogenetic study is assumed to be correct. However, the possibility of a frequent gain and loss of the *trnT* can not be dismissed, as the mechanism of the process itself is not clear, and the circumstances that led to the event could occur several times in the evolution of liverworts, or any plants for that matter. The likeliness of such an event is especially high in this spacer region, because the mitochondrial *nad7* gene that is flanking this region is a pseudogene in *Marchantia* (Oda et al. 1992a) and presumably also in other liverworts (see chapter 6). As the functional version of the gene was transferred to the nucleus (Kobayashi et al. 1997), the 5'- region of the mitochondrial gene has not necessarily to be this conserved any more, as it does not have to encode promotor sequences. Interestingly, the mitochondrial pseudogene is still transcribed at least in the case of *Marchantia* (Takemura et al. 1995) despite the proposed reinsertion of the *trnT* gene.

The *trnT* gene itself is also a very interesting case. It is clearly a *trnT*(GGU) gene that is very similar to the moss and algae mitochondrial *trnT*(GGU) genes, because not only the anticodon but almost the complete sequence is identical in these plants. It is therefore assumed here that the inverse *trnT* found in some liverworts is indeed the same gene or a copy of the original *trnT* that was inherited vertically from the common ancestor of all land plants.

One notable exception of the sequence homologies is the occurrence of a mismatch in the acceptor stem in liverworts and some mosses. The nucleotide C of the C-A mismatch could be exchanged with a U by RNA editing and rescue the mismatch in this way. This is likely in the case of the mosses, but no editing ability was detected in one subgroup of liverworts, the complex thalloid liverworts (Marchantiopsida) (Steinhauser et al. 1999). All liverworts that were shown to carry a *trnT* gene in the spacer between *trnA* and *nad7* are members of that group. Therefore a rescue of the mismatch is supposedly impossible. The occurrence of mismatches in the stem regions of tRNA structures seems in fact to be rather regular for *Marchantia*, as 18 of the 27 species of mitochondrially encoded tRNAs have mismatches in one or more stem parts (compare

Oda et al. 1992b), although in most cases this is restricted to the proximal pair of the anticodon stem or the distal pair of the D-loop arm, positions with rather low influence on the overall folding structure.

It should be noticed that the *trnT*(GGU) gene detected in the spacer between *trnA* and *nad7* is the only *trnT* gene encoded on the *Marchantia* chondriome. No tRNA species for the codon recognition of ACR (also threonine) was identified (Oda et al. 1992b). It is likely that the missing tRNAs are encoded on the nuclear genome and imported from the cytoplasm into the mitochondrion. Alternatively, another possibility to compensate for the missing codon recognition of ACR could be a "two out of three" mechanism, that could lead to the recognition of ACR via the single available *trnT*(GGU) (Lagerkvist 1978).

4.3 The gene cluster *nad5-nad4-nad2*

The genes *nad5*, *nad4*, and *nad2* are highly conserved mitochondrial genes coding for subunits of complex I of the respiratory chain. These three genes are identically arranged in a gene cluster in the chondriomes of the streptophyte algae *Chara vulgaris* and *Chaetosphaeridium globosum* as well as in the liverwort *Marchantia polymorpha* (Fig. 4-5).

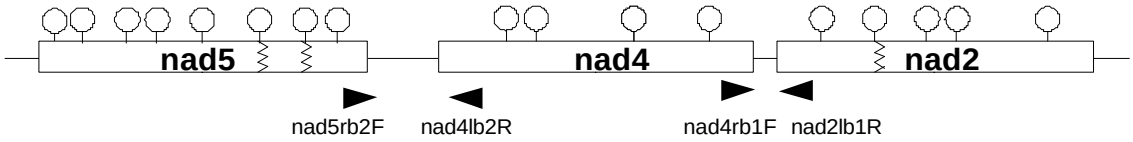

Fig. 4-5: Graphical overview of the *nad5-nad4-nad2* gene cluster. Circles stand for introns identified in the genes in different land plants, zig-zag-lines illustrate the disruption of the genes in angiosperms through trans-splicing introns. Arrows indicate the location of primers used in PCR assays.

They are even co-transcribed in the latter (Nozato et al. 1993). The gene arrangement is not only broken up in angiosperms, even more so, *nad5* and *nad2* are individually disrupted in their gene continuities by trans-splicing introns. As can be judged from the complete mitochondrial sequence of *Arabidopsis thaliana* and other angiosperm mitochondria, all three genes and/or their respective parts are located in different

sequence environments on the chondriomes. Tracing the break-up of this gene cluster could give interesting insights into the evolution of lower land plant chondriomes.

4.3.1 Material and Methods

Methods used for the study of this gene cluster are identical to the ones used for the analysis of the gene cluster *trnA-trnT-nad7* (see chapter 4.2 and chapter 2.2). Primer sequences for the *nad5-nad4-nad2* gene cluster were nad5rb2F (5'- ggt gct att gaa atc ttg ggt cc - 3'), nad4lb2R (5'- aca aag aat aam gag ata cca tct ata cc - 3'), nad4rb1F (5' - gga gtt att tgg atg ggt gtt tac - 3'), nad2lb1R (5' - aac act aag taa acc aag cca acy cac - 3'). Additional internal primers were used to obtain complete sequences.
Taxa sampled for this study are given in table 4-4.

4.3.2 Results for the *nad5-nad4-nad2* gene cluster

The spacer between *nad5* and *nad4* is only 57 base pairs long in *Chara*, the probably closest extant ancestor of land plants (Karol et al. 2001). In *Chaetosphaeridium*, another charophyte alga, which is rather less closely related to land plants (Turmel et al. 2002), this spacer is also conserved and even smaller with only 15 bp. The complete chondriome sequence of the complex thalloid liverwort *Marchantia polymorpha* (Oda et al. 1992a) exhibits the same gene continuity but a much larger intergenic region with 1311 bp in size.
This study extended the taxon sampling with four other liverworts, and included for the first time sequences of three mosses, two hornworts and the quillwort *Isoetes* (Tab. 4-4), to establish an overview for the evolution of the conserved *nad5-nad4* spacer over all three bryophyte classes and an early tracheophyte.
PCR assays of the gene cluster resulted in very different sizes of the intergenic region in the major bryophyte groups (fig. 4-6, tab. 4-4). The analysed region is much larger in land plants than in algae, and is comprised of 1000 to 1300 bp in liverworts, as was known already from *Marchantia*, with the longest sequences found in complex thalloids and significantly shorter spacers in the leafy liverwort *Frullania* and the simple thalloid *Metzgeria*. This liverwort spacer is disrupted by a sequence that is homologous to intron

i783 of the *cob* gene. This gene encodes apocytochrome B, a highly conserved mitochondrial gene that is present in *Marchantia* as an intronless pseudogene version and a functional copy that carries three introns.

It should be noticed that the taxon *Blasia* is controversially discussed in its taxonomic placement. This genus has been treated as a simple thalloid liverwort in morphological studies and has been reconsidered as closely connected to complex thalloid liverworts in recent molecular studies. The sequence of *Blasia*'s *nad5-nad4* spacer is even longer than the two other sampled complex thalloid liverworts, with an extension of an additional 14 bp compared to *Marchantia*, and 2 bp compared to *Riccia*, respectively. This supports other molecular analyses proposing a close relationship of this genus with complex thalloids.

Table 4-4. Taxa sampled for the *nad5-nad4-nad2* gene cluster study and spacer length inferred from the obtained sequences.

Taxonomy	Species	length & accession number nad5-nad4 spacer	length & accession number nad4-nad2 spacer
Charophyta	*Chara vulgaris*	57 bp NC_005255	44 bp NC_005255
	Chaetosphaeridium globosum	15 bp NC_004118	18 bp NC_004118
liverworts Marchantiopsida complex thalloids	*Marchantia polymorpha*	1311 bp NC_001660	26 bp NC_001660
	Riccia fluitans	1323 bp DQ098657	26 bp DQ098663
Blasiales	*Blasia pusilla*	1325 bp	26 bp
Jungermanniopsida simple thalloids leafy	*Metzgeria furcata*	1047 bp DQ098658	26 bp DQ098664
	Frullania brasiliensis	1052 bp DQ098659	26 bp DQ098665
mosses	*Sphagnum fallax*	637 bp DQ098673	26 bp DQ098666
	Physcomitrella patens	617 bp DQ098674	26 bp DQ098667
	Mnium hornum	621 bp DQ098676	26 bp DQ098668
hornworts	*Anthoceros agrestis*	3328 bp DQ098661	26 bp DQ098669
	Megaceros spec.	2885 bp DQ098662	26 bp DQ098670
Lycopodiopsida	*Isoetes lacustris*	no PCR product	17 bp DQ098671

The *nad5-nad4* spacer is significantly smaller in mosses than in liverworts, with a size of approx. 600 bp (fig. 4-6, tab. 4-4). No significant similarities to other sequences were identified in this spacer region. An extended taxon sampling (Rein, Groth-Malonek, Knoop: "Mitochondrial gene spacers as novel phylogenetic markers: a case study in mosses", in preparation) shows an easily alignable region that contains a high amount of

variable sites, and which could be a putatively useful novel marker for phylogenetic approaches in mosses.

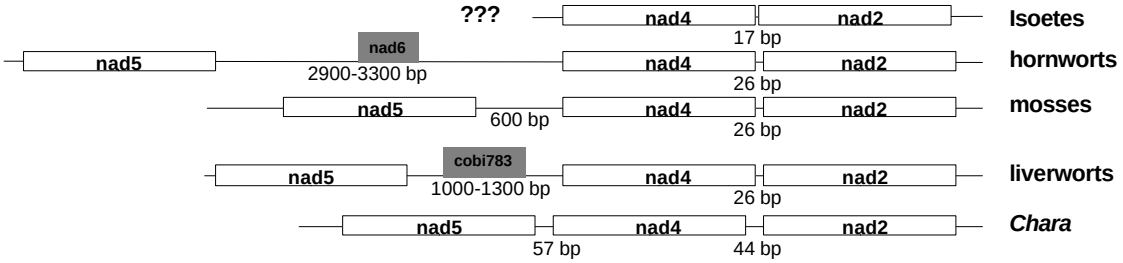

Fig. 4-6. Graphical overview of the mitochondrial gene cluster *nad5-nad4-nad2*. Grey boxes indicate homologies of spacer fragments with other identified elements of mitochondrial DNA.

The *nad5-nad4* spacer in hornworts is even larger with up to 3300 bp in size (fig. 4-6). Part of the spacer is a region with a high similarity to the *nad6* gene. This region is disrupted by frame shifts and stop codons, rendering it a pseudogene (not shown here). This is the first known case of a *nad6* pseudogene in land plants, where it is usually one of the highly conserved mitochondrial genes that are not transferred to the nucleus but functionally encoded on the chondriome (see Knoop 2004 and therein).
No PCR product could be obtained from the intergenic region between *nad5* and *nad4* in the lycophyte *Isoetes*, or other tracheophytes.

The region between *nad4* and *nad2* is comprised of just 44 bp in *Chara* and 18 bp in *Chaetosphaeridium*. This study revealed a spacer of a highly constant size in all three bryophyte groups with exactly 26 bp, which is not only co-transcribed in complex thalloid liverworts like *Marchantia* (Nozato et al. 1993), but also in the leafy liverwort *Frullania tamarisci*, as could be shown by PCR on cDNA.
This spacer is even smaller in the early diverging tracheophyte *Isoetes* with only 17 base pairs (fig. 4-6). No other tracheophyte PCR product could be obtained.

4.3.3 Discussion of the evolution of the gene region *nad5-nad4-nad2*

The situation in the gene cluster *nad5-nad4-nad2* is very different to that in the *trnA-trnT-nad7* cluster as all three genes are highly typical mitochondrial genes and are

present and functional on the chondriome in all land plants, in contrast to the early transfer of mitochondrial tRNAs to the nucleus, and the special case of the decay of the *nad7* pseudogene in liverworts. Although these three *nad* genes are functionally connected and even co-transcribed in different liverworts, their gene continuity is lost in higher plants. The results of my study show that even two spacers as closely connected as in this gene cluster can undergo very different styles of evolution.

The smaller spacer between *nad4* and *nad2* is conserved in size in all three main groups of bryophytes: liverworts, mosses, and hornworts (Fig. 4-6). It is even smaller than its counterpart in green algae, which is a clearly surprising discovery.
Seen in direct comparison, the *Chara* chondriome has a size of 68 kb, whereon 68 genes are encoded (Turmel et al. 2003), and *Marchantia* with a very similar gene content exhibits a chondriome size of 186.6 kb (Oda et al. 1992a), almost three times as large. This increased size is partly based upon the aquisition of several introns, but mainly on the size increase of non-coding spacer regions between genes. Therefore a reduction of a spacer and its high conservation over different bryophyte groups, as seen in the case of the nad4-nad2 spacer, is in contrast to the typical picture of mitochondrial DNA evolution.
Interestingly, the further size reduction in the tracheophyte *Isoetes* could be expected, as this genus generally exhibits a very unusual pattern of its chondriome evolution. It is known to preferentially carry very small introns (Malek and Knoop 1998; Pruchner et al. 2002), that are still homologous in position and structure to their counterparts in other plant groups, but lack large parts of variable regions like domain IV in group II introns. *Isoetes* seems to employ mechanisms to reduce the size of its chondriome through the reduction of non-coding features, or otherwise lacks mechanisms that enlarge these parts. This rule possibly applies to its intergenic regions as well, and therefore leads to this very small spacer between *nad4* and *nad2* (Fig. 4-6). As no PCR products could be obtained from other lycophytes like *Huperzia* or *Lycopodium*, which do not exhibit this kind of evolution (at least as far as is known from *nad5* and *nad2* studies), no general point can be made about the conservation of this gene continuity in early tracheophytes. It is assumed, however, that the gene arrangement is broken up sometime before the evolution of angiosperms, as no further PCR products could be

obtained from ferns (e.g. *Equisetum, Psilotum, Asplenium, Dicksonia*), or gymnosperms (e.g. *Gnetum, Welwitschia, Ginkgo, Pinus, Abies*), either.

The spacer between *nad5* and *nad4* is clearly a different case, as it is much expanded in all land plant groups compared to *Chara*. Liverworts, which are set apart from other land plants by the occurrence of several unique introns in their chondriome, exhibit a spacer region that is about 20 times longer than the algal counterpart. Much of this spacer is composed of a sequence that is very similar to the second intron of *Marchantia*'s *cob* gene, an intron that is so far known from *Marchantia polymorpha* only (Oda et al. 1992a). Similiarities to this intron could be identified in all obtained liverwort sequences, albeit with a significant part of the intron missing in the Jungermanniopsid taxa *Frullania* and *Metzgeria*. This indicates that the original *cob* intron (one of three group II introns restricted to *Marchantia* in this gene) was gained by the common ancestor of all liverworts, as is similarly assumed for instance in the cases of nad4i548 (chapter 2), nad7i336 or nad7i1113 (chapter 5), and that a copy of this intron not only inserted itself into *cob*, but also into the spacer between *nad5* and *nad4* in an early stage of the evolution of liverworts, predating the diversification of Jungermanniopsida (simple thalloid and leafy liverworts) and Marchantiopsida (complex thalloid liverworts).

The respective intergenic region in mosses is comprised of only approx. 600 bp that include no similarities to other introns, but is nevertheless around ten times as long as the algae spacer. This locus could be a valuable new tool for phylogenetic analyses in the moss lineage as it is short but yet very variable. Its main disadvantage would be the fact that no close outgroup is available, as this spacer region developed very differently in all bryophyte groups, and can not be aligned over any two groups out of liverworts, mosses, or hornworts.

The most expanded spacer region between *nad5* and *nad4* that could be amplified with a PCR assay has been detected in two hornworts, comprising a spacer of 2900 bp and 3300 bp, respectively. These sequences already carry evidence for the interruption of the gene continuity, as a degenerated version of *nad6* has been identified between *nad5* and *nad4*. *Nad6* encodes for a rather small subunit of complex I of only approx. 200 amino acids, with almost the full-length reading frame still being

discernible in both hornworts, albeit disrupted by frame shifts and stop codons. It is frequently found in different positions on the chondriome relative to other genes, as for instance in a co-transcribed position downstream of the gene encoding subunit 6 of ATP-synthase (*atp6*) in maize (Haouazine-Takvorian et al. 1997) or downstream of mitochondrial *rps4* in tobacco (Sugiyama et al. 2005). A 3'-severed sequence of *nad6* encoding 100 amino acids has also been identified in a secondary copy on the *Marchantia polymorpha* chondriome (Yamato et al. 1993). It is not clear whether the identified copy in the hornwort spacer was functional at any point, or where the actually functional *nad6* gene is located on the hornwort chondriome, as no complete sequence is available from any hornwort taxon. Nevertheless it can be postulated that the insertion of *nad6* in the *nad5-nad4* spacer did at least lead to a further size increase, that could eventually lead to a disruption of the gene continuity. No PCR products could be obtained from any tracheophyte taxon, not even from *Isoetes*, which exhibits a conserved *nad4-nad2* gene linkage (Fig. 4-6). If the theory that hornworts are ancestors to the tracheophytes is correct, as postulated recently with molecular approaches (Lewis et al. 1997; Groth-Malonek et al. 2005), based on chemical analyses (Carafa et al. 2005), or derived from the comparison with fossil data of *Rhynia* and *Hornea* as early as 1924 (Campbell 1924), we would see here the slow development of a gene cluster disruption on the way from Charophyte algae to tracheophytes, frozen in different stages of its development, presented by several taxa of every major group of bryophytes.

4.3.4 Conclusion

The selected gene clusters *trnA-trnT-nad7* and *nad5-nad4-nad2* analysed in this study show very different aspects of the evolution of mitochondrial DNA. The rearrangement of genes, like the insertion of a *nad6* pseudogene in a hornwort spacer (Fig. 4-6), the loss of tRNAs, shown here as the loss of *trnT* in some liverworts (Fig. 4-3), and insights into the origin of parts of intergenic regions, like the similarity of large parts of the *nad5-nad4* spacer with an *cob*-intron (Fig. 4-6), are all typical mitochondrial developments that were formerly only known from angiosperms, as no comparable study was available for lower land plants.

This study includes data from all major bryophyte groups and an early tracheophyte. The evolutionary step from bryophytes to tracheophytes seems to be accompanied by an increase in structural changes of the chondriome, as no conservation of any of the analysed spacers could be found. The single exception is the *nad4-nad2* spacer, but this finding was restricted to one small group of lycophytes, the quillworts (*Isoetes*), and only one species therein (Fig. 4-3).

It is also noteworthy that the evolutionary pattern is very distinct and different in all three major bryophyte groups: liverworts, mosses, and hornworts. Liverworts, especially complex thalloids, tend to expand their spacers much more than mosses, and hornworts even more so. In the case of the *nad5-nad4* spacer, the latter seem to be already on the way to disrupt the spacer through the insertion of another gene, and the destiny of the whole gene cluster *trnA-trnT-nad7* is unknown here.

In the cases where the spacer is still conserved in differrent bryophyte groups, an alignment of these spacers between them is not possible, not just due to the different size of the spacers but to the very differing nucleotide sequences. In contrast to that, the alignment of sequences from a single group is rather straightforward, especially in mosses.

Further studies could continue here and obtain data from non-angiosperm tracheophytes to complement the knowledge about these mitochondrial regions with data from a broader range of all land plants.

5 Evolution of a pseudogene: the mitochondrial *nad7* gene in liverworts

5.1 Introduction

Plant mitochondrial DNA underwent frequent structural changes. These interesting processes are only recently exploited as possibilities to gain further understanding of the evolution of lower land plants (Groth-Malonek and Knoop 2005). Former analyses of angiosperm mitochondria concentrated mainly on gene transfer from the mitochondrial genome (chondriome) to the nucleus. Such transfers of single genes from the chondriome to the nuclear genome are known to occur frequently and independently in angiosperms (Adams et al. 1999; Adams et al. 2000; Palmer et al. 2000; Adams and Palmer 2003). Mostly, genes for proteins of the small (*rps*) or large (*rpl*) ribosomal subunits are affected, only rarely components of the respiratory chain, such as *cox2* (Nugent and Palmer 1991). However, several genes are known to be universally conserved in the mitochondrial genome of embryophytes, from green algae to angiosperms (Knoop 2004). Notably, this includes nine *nad* genes coding for subunits of the NADH ubiquinone oxidoreductase, complex I of the respiratory chain.

One noteworthy exception is the *nad7* gene, which is a mitochondrial pseudogene in *Marchantia polymorpha*, a complex thalloid liverwort (Oda et al. 1992a). In this case, a nuclear copy is functional and the mitochondrial version has become a pseudogene due to the mutational introduction of six stop codons (Takemura et al. 1995; Kobayashi et al. 1997). It could be proposed that these stop codons are rescued by RNA editing, as is common for instance in hornworts. Most liverwort groups are known to be capable of RNA editing (Steinhauser et al. 1999), but this does not include the complex thalloids like *Marchantia*.

The mitochondrial *nad7* gene in land plants carries three to four group II introns in angiosperms (Fig. 5-1), ancestral homologues of two of them are conserved in mosses (Pruchner et al. 2001). Two unrelated group II introns are present in the mitochondrial nad7 pseudogene of *Marchantia polymorpha* (Oda et al. 1992a). The functional nuclear copy of *nad7* in *Marchantia* is 5'-extended to provide an appropriate target sequence for

organellar import, and this aminoterminal extension of the reading frame is interrupted by a typical spliceosomal nuclear intron (Kobayashi et al. 1997). No further introns were identified, expectedly none of the mitochondrial group II introns. As it must be assumed that the nuclear copy originated from the mitochondrial *nad7* gene, this could be an indication that the transfer to the nucleus was mediated via a mature intron-less mRNA. An alternative would be a very ancient transfer event that took place when no organellar introns where yet acquired. Although the mitochondrial *nad7* gene is actively transcribed in *Marchantia* as was shown through northern blot analysis, no splicing of the two introns was detectable (Takemura et al. 1995). The secondary structures of these introns are mostly conform with the group II intron consensus, but in both cases intron and exon binding sites are not completely compatible (Takemura et al. 1995). 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. Their insertions could possibly even have mediated the process of pseudogene degeneration, as a non-spliceable intron would prevent the translation of a correct amino acid sequence, and so decided the fate of the mitochondrial copy that remained in coexistence after the transfer of a *nad7* copy to the nucleus. Alternatively, splicing function may have been lost after pseudogene degeneration or, as a third alternative, both introns could have been inserted into *Marchantia*'s mitochondrial *nad7* gene only after its evolvement into a pseudogene.

The focus of this study was to extend our knowledge about *nad7* in liverworts with special emphasis on the aspect of the evolution of a pseudogene.

Several molecular data support the assumption of liverworts as the earliest land plants (Qiu et al. 1998; Groth-Malonek et al. 2005). They share a serial sister group relationship with mosses, hornworts and tracheophytes (Groth-Malonek et al. 2005). Earlier analyses show that *nad7* of mosses do not possess any stop codons in their reading frame and therefore very likely represent a functional mitochondrial *nad7* (Pruchner et al. 2001), which also generally holds true for angiosperms. The closest ancestors to land plants are Charophyte algae, of which the chondriomes of *Chara vulgaris* (Turmel et al. 2003) and *Chaetosphaeridium globosum* (Turmel et al. 2002) are fully sequenced. Both species carry functional *nad7* genes in their mitochondria as well.

Obviously the mitochondrial *nad7* pseudogene evolved after the diversification of the liverwort group, and is restricted to them.

Liverworts (Marchantiophyta) are divided into up to five clades based on morphological and recently published genetic data (Davis 2004; Crandall-Stotler et al. 2005; Forrest and Crandall-Stotler 2005; Frey and Stech 2005). The earliest diverging group is supposedly formed by the enigmatic groups of Haplomitriaceae and/or Treubiaceae, albeit with generally weak support in the phylogenetic analyses. The complex thalloids such as *Marchantia* are a second clade with a sister group status to the rest of the liverworts, which are divided into two clades of simple thalloids, and finally the monophyletic leafy liverworts (Fig. 5-2). Alternatively the *Haplomitrium*/*Treubia* group may be an early lineage of the simple thalloids.

An insight into the evolution of the *nad7* pseudogene could also enlighten our knowledge about liverwort phylogeny in general.

5.2 Material and Methods

DNA extraction and PCR assay strategies see chapter 2.2.1 “General methods and strategies”. Taxa used for PCR assays and additional sequences that were used for correlation and alignments are shown with their respective accession number from Genbank (NCBI), and/or their respective sequence length (table 5-1). DNA and RNA were differentially precipitated in the presence of 3 M lithium acetate. OmniScript ™ Reverse Transcriptase (Qiagen) was used for cDNA synthesis. 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’), primers for the cDNA assay were nad7up.2 (5’- gga cct caa cay cct gct gct cat gg -3’) and nad7do2 (5’- tct atc tac ctc tcc aaa cac aat -3’). Five of the six mitochondrial *nad7* stop codons of *Marchantia* are located between the two liverwort-specific introns nad7i336 and nad7i1113. This exon is therefore an attractive region for sequence studies, and could show evidence for pseudogenes in other liverwort taxa. Assuming that the two introns would be conserved, two intron-based primers (nad7i336up and nad7i1113do) were designed to ensure amplification of the mitochondrial copy in the analysed liverworts rather than a potential nuclear version, and to establish whether both introns are conserved in liverwort groups other

than complex thalloids. In addition to that, primers at the beginning and the end of the mitochondrial reading frame were used to amplify cDNA of *Haplomitrium*.

Table 5-1: Taxa and sequences included in this study

Taxonomy	Species	Voucher	Accession number
Haplomitriopsida	*Haplomitrium mnioides* (Lindb.) Schust.	M. Shimamura s.n.	1321 bp cDNA: 891 bp
Marchantiopsida / Blasiopsida	*Blasia pusilla* L.	J. Heinrichs 2291	1081 bp
Marchantiopsida s.str.	*Marchantia polymorpha* L. chondriome	---	NC 001660 5668 bp
	Marchantia polymorpha L. nuclear *nad7*	---	1942 bp Kobayashi et al. (1997)
	Bucegia romanica Radian	Ulm-collection s.n.	1325 bp
	Conocephalum conicum (L.) Underw.	Groth & Schwertfeger s.n.	1316 bp
	Monosolenium tenerum Griff./Sunita Kapila & SS Kumar	live culture Goettingen	1322 bp
	Lunularia cruciata (L.) Dum. ex Lindb.	Groth & Schwertfeger s.n.	1282 bp
Jungermanniopsida simple thalloids	*Aneura pinguis* (L.) Dumort.	MGM031218-01SC	1198 bp
Jungermanniopsida leafy liverworts	*Lepidolaena hodgsoniae* Grolle	MGM031218-02SC	1352 bp
	Calypogeia muelleriana (Schiffner) K. Müller	J. Heinrichs 4375	1232 bp
	Frullania tamarisci (L.) Dumort.	J. Heinrichs 4382	1233 bp
	Harpanthus flotovianus (Nees) Nees	J. Heinrichs 4390	1230 bp
	Scapania nemorea (L.) Grolle	J. Heinrichs 4372	1233 bp
Charophyta	*Chara vulgaris* L.	---	1182 bp NC_xxx
Spermatophyta	*Arabidopsis thaliana* L.	---	NC_000932 6083 bp

5.3 Results

The amplification and sequencing of the *nad7* region between the two liverwort-specific introns nad7i336 and nad7i1113 (Fig. 5-1) was successful for several species, including members of all main groups of liverworts.

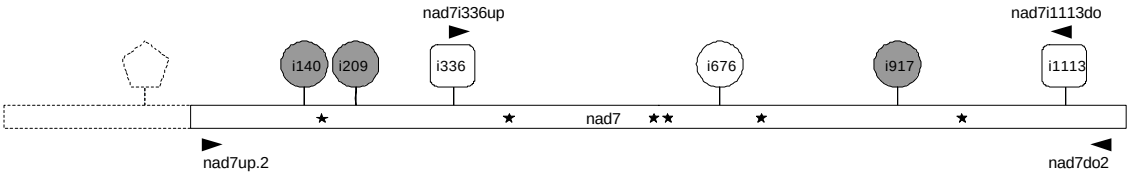

Fig. 5-1. Graphical overview of the *nad7* gene in land plants, adapted from Prucher et al. 2001. Dotted lines delineate the 5'-extension of the nuclear *nad7* copy of *Marchantia polymorpha*, including one nuclear type intron (pentagon). Shaded circles indicate group II introns that are conserved in angiosperm chondriomes, the open circle designates an intron that is lost in tobacco. Squares represent group II introns identified in *Marchantia*, here shown to be generally present in liverworts. Asterisks indicate the position of six stop codons in the mitochondrial *nad7* gene of *Marchantia polymorpha*, which render it a pseudogene. Two primer binding sites located in exons are indicated by arrows below the gene, two intron-based binding sites are shown above the respective introns.

The exon 2 sequence of *Haplomitrium* is intact (Fig. 5-2). The alignment of its putative amino acid sequence with the respective sequences of *Marchantia*, the thale-cress *Arabidopsis* and the moss *Physcomitrella* show no stop codons (Fig. 5-2), but several RNA editing sites that were confirmed by cDNA sequence comparison in the case of *Haplomitrium*.

```
Arabid. mt    HSLALTTHAMDVGALTPFLWAFEEREKLLEFYERVSGARMHASFIRPGGVAQDLPLGLCRDIDSFTQQFASRIDELEEMSTGNRIWK
Physco. mt    .L.....................................Y.........M....SE..FL...........I...L.N.....
Haplo.  mt    .L...........P.....................P.......Y.............SE..SP.............L.N.....
March.  nuc   .L.S.....................MM.........L..AY.........M....SE..FL...........I...L.N.....
March.  mt    .L.......I.................*.............Y......G.......SE..FL.............K*.N...*.
Chara   mt    .L.....................................Y.............SE.LFL...........V...L.N.....

Arabid. mt    QRLVDIGTVTAQQAKDWGFSGVMLRGPGVCWDSRRAAPYDVHDQSDLDVPVGTRGDRYDRYCIRIEEMRQSLRII-VQCLNQMPSGM
Physco. mt    .............L...........S.....L.KS.....YN.LSF.........C..............I...-M..........
Haplo.  mt    .............V........S..S.....L.KS.....Y.RL.F.........C..............I...-M.........V
March.  nuc   .......V..GEE.M......P....S.IK..L...S...CY.KLEF.I.......C....LV........I...-A....D..N..
March.  mt    .............V...........S....NL.K*.L...Y.RL.FE.---...R.C....Y.........I...IM..........
Chara   mt    .............M...........S..S..L.K......Y.KVEF.........C.....V...........-...........

Arabid. mt    IKADDRKLCPPSRCRMKLSMESSIHHFELYTEGFSVPASSTYTAVEAPKGEFGVFLVSNGSNRPYRRKIRAPGSAHSQGLDSMSKHH
Physco. mt    .............SQ..Q....L....K.............................T.....C......F..L....F.....
Haplo.  mt    .......PR....SQ..Q.T..L...LKPH...V....P..H.................T.....C.T....F..L..........
March.  nuc   .......IT....TQ..Q....L....K.....YH...GA.............Y.....T.....C......FG.L....F.....
March.  mt    ........G.TA.S...Q....L....K....SV..R.............*.........T.....C..T...F..L....F.....
Chara   mt    .............SQ..Q....L....K.....V.........C..........Y.....T.....C......F..L....F.....
```

Fig. 5-2. Amino acid sequence alignment of the mitochondrial *nad7* gene exon 2 of *Haplomitrium* and selected taxa (angiosperm *Arabidopsis thaliana*, moss *Physcomitrella patens*, liverwort *Marchantia polymorpha*, alga *Chara vulgaris*). Stop codons are indicated by asterisks, editing sites in *Haplomitrium* identified by DNA-cDNA-comparison are marked with bold letters. Amino acids identical to the uppermost line are indicated by dots, dashes represent alignment gaps. Mitochondrial *nad7* genes are given as "mt", a nuclear copy as "nuc".

The basalmost genus of the group of the complex thalloids liverworts is *Blasia*, an enigmatic species that was widely considered to be a simple thalloid liverwort and was recently linked to the complex thalloids by morphological and genetic studies. In *Blasia* a large part of the analysed exon sequence is missing (Fig. 5-3), and this loss is accompanied by a shift of the reading frame. Therefore, *Blasia* has indeed a *nad7* pseudogene, but not due to stop codon introduction through point mutations.

Nad7 fragments from other marchantiopsid species show closer similarity to the *Marchantia* sequence (Fig. 5-3). All members of the complex thalloid group, including *Marchantia*, carry stop codons in their *nad7* reading frames. Intriguingly, one of them is shared by all members of the group and the basal-most Jungermanniopsiid liverwort *Aneura* (S 3728), two further stops are shared by all Marchantiopsida except for *Lunularia* (S3483 and S 4064). Nine other stop codons are unique to single species and distributed over the whole second exon of the *nad7* gene (Fig. 5-3).

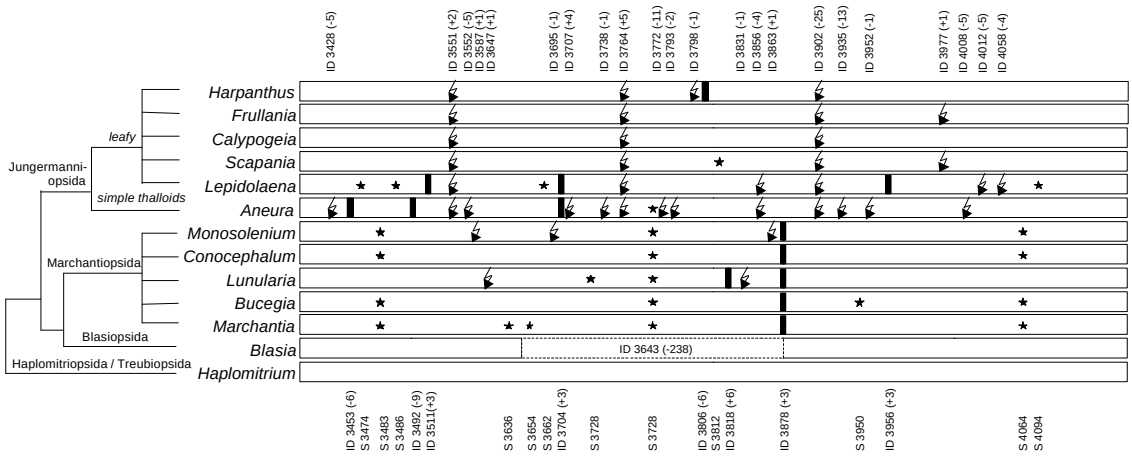

Fig. 5-3: Graphic overview of the *nad7* exon region between the introns nad7i336 and nad7i1113 for *Marchantia polymorpha* and 12 other liverworts. Asterisks represent stop codons (S), arrows indicate indels (ID) that result in frame shifts, black boxes depict indels (ID) that do not represent frameshifts but codon losses or gains. Their respective positions are indicated by the *Marchantia polymorpha* nucleotide position of the *nad7* gene on the 5' side of the stop codons and indels. Dotted lines circumscribe the loss of a major part of the exon in *Blasia*. Phylogenetic relationships of the taxa are shown to the left, summarizing findings from several recent studies.

The reading frames of taxa from the groups of simple thalloid and leafy liverworts are disrupted by several indels, as well as the reading frames of the complex thalloid liverworts *Monosolenium* and *Lunularia* (Fig. 5-3). Three indels are shared by all Jungermanniopsiid taxa: ID 3551 (+2), ID 3764 (+5), ID 3902 (-25), all constituting frame shifts, respectively. One other frame shifting indel is shared between the simple thalloid *Aneura* and the basal leafy liverwort *Lepidolaena* (ID 3856 (+4)). Indels that do not result in frame shifts are much rarer (black boxes in Fig. 5-3). One of them, ID 3704 (+3), is again shared between *Aneura* and *Lepidolaena*, another one is present in all Marchantiopsida (ID 3878 (+3)).

5.4 Discussion of the evolution of *nad7* in liverworts

The *nad7* gene is one of the most conserved genes of the plant mitochondrial genome, and therefore its development into a pseudogene and the existence of a functional nuclear copy is a rare case with land plants. In fact, this is the only known event of functional gene transfer from the mitochondrion in non-angiosperm land plants, and the only one that involves one of the nine mitochondrially encoded *nad*-genes. The transfer

event is apparently restricted to liverworts, because a presumably functional mitochondrial version is known from both *Chara* (Turmel et al. 2003) and mosses (Pruchner et al. 2001).

As mentioned before, the *Chara nad7* gene is intron-free, and all analyzed mosses share the group II introns nad7i140 and nad7i209 with flowering plants (Fig. 5-1). *Marchantia* shares none of these introns, but carries two different group II introns. Although the mitochondrial pseudogene of *Marchantia* is transcribed, both introns are not spliced (Takemura et al. 1995).

The situation is quite different in the case of the rare liverwort *Haplomitrium mnioides*, which has recently been proposed as a member of the earliest diverging group of all liverworts (Forrest and Crandall-Stotler 2004; Forrest and Crandall-Stotler 2005). Its cDNA was obtained using primers at the end of the coding sequence (CDS), the respective DNA sequence for exon 2 was obtained with primers anchoring in both liverwort introns. The PCR and the RT-PCR products document both the presence of the two liverwort-type introns and their functional splicing. The alignment of all obtained liverwort sequences clearly documents the occurrence of both introns in all liverwort taxa. Several base exchanges of C to T were found in *Haplomitrium* when DNA and cDNA sequences were compared, indicating an exchange of C to U on mRNA level (Fig. 5-2). This kind of RNA editing is common in all land plants, with the exception of complex thalloid liverworts. *Haplomitrium* is known to have an unusually extensive amount of editing sites in *nad5*, another mitochondrially encoded nad-gene (Groth-Malonek et al. 2005). In the case of *nad7*, there is evidence for a high amount of RNA editing as could be expected from a functional mitochondrial gene of this taxon. Therefore, it is assumed here that *nad7* is a functional mitochondrial gene in the liverwort *Haplomitrium*.

This is not the case for *Blasia*, which morphologically resembles a simple thalloid liverwort, but is on the molecular and ultrastructural level more closely related to complex thalloid liverworts like *Marchantia* (Renzaglia and Duckett 1987; Garbary and Renzaglia 1998). Recent studies of liverwort phylogeny place *Blasia* as the basalmost branch of the complex thalloid (Marchantiopsida) clade (Davis 2004; Forrest and Crandall-Stotler 2004; Forrest and Crandall-Stotler 2005), where it is occasionally referred to as the separate class Blasiopsida (Stech and Frey 2001). In the case of the

mitochondrial *nad7* gene a unique feature was found in *Blasia*: it was the only sequence lacking a large portion of exon 2, resulting in a frame shift and ultimately in a putative pseudogene (Fig. 5-3). No stop codons were detected, and no further frame shifts could be found. All other analysed complex thalloid taxa share a different mode of pseudogene evolution: they carry stop codons. *Marchantia* is the most stop-rich taxon sequenced here, with 5 stop codons distributed over exon 2 (Figs. 5-2 and 5-3). *Bucegia*, a second member of the same family as *Marchantia* (Marchantiaceae), shares three of these stops (S 3483, S 3728, S 4064), and features a single stop that is unique for *Bucegia* (S 3728). The less closely related taxa *Monosolenium* and *Conocephalum* share all three conserved stop codons without the occurrence of new ones, but in the case of *Monosolenium* an additional three indels were detected. All three indels constitute frame shifts, as they are the size of only one single nucleotide. The even further distant *Lunularia* shares only one stop codon (Fig. 5-3, S 3728), has a unique stop codon in addition to that, and includes also two single-nucleotide indels. Therefore, only one of the stop codons, S 3728, is conserved in all surveyed complex thalloid liverworts. This stop actually occurs in *Aneura* as well, which is a member of the earliest diverging branch of the Jungermanniopsida, the sister group of the Marchantiopsida, and could therefore be the earliest occurring stop codon in the pseudogene evolution of *nad7*. Interestingly, 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. If *Blasia*'s placement as the earliest diverging taxon of the complex thalloid clade is correct, then it can be postulated that the transfer of a functional *nad7* copy to the nucleus must have occurred before the evolution of the whole complex thalloid liverwort clade including *Blasia*. Only the lifting of any evolutionary constraint from the mitochondrial *nad7* due to the occurrence of a functional version in the nucleus can allow for a pseudogene in the mitochondrial genome.

Aneura, which is, as a simple thalloid liverwort, a member of supposedly early diverging groups of the Jungermanniopsida, shows the extraordinary high number of 16 indels in the *nad7* exon 2. All further taxa show fewer indels, from three (*Calypogeia*) up to nine (*Lepidolaena*), and occasionally stop codons. All Jungermanniopsid liverwort sequences share three frame shifts in the analysed exon

(Fig. 5-3), and all of them are constituting frame shifts. An interesting difference between the frame shifts found in Marchantiopsida and Jungermanniopsida is their size. All four Marchantiopsid indels are comprised of only one nucleotide. In the case of the 23 identified Jungermanniopsida indels, only two of the indels in the sequence of *Aneura* are this small, and also another indel of *Harpanthus*, a leafy liverwort, has this size. Single-nucleotide indels are very rare, and they can often be linked to sequencing errors, especially in coding regions. This can not be totally ruled out here, as it was not possible to obtain and sequence a multitude of clones from some species (e.g. *Monosolenium*), which should be obtained to prove the correctness of the sequences. However, if any combination of these particular indels (Fig. 5-3, ID 3587, ID 3647, ID 3695, ID 3738, ID 3798, ID 3831, ID 3863, ID 3952, ID 3977) would be excluded from the overall picture, all sequences involved would still represent pseudogenes due to different indels or stop codons. Nevertheless, the occurrence of positionally identical stop codons and indels that are present in taxonomically linked species points furthermore to a rather correct estimation of the situation in the mitochondrial *nad7* gene in liverworts.

CONCLUSION

All liverwort *nad7* sequences investigated here were shown to be pseudogenes with the notable exception of *Haplomitrium*, the taxon that has been proposed to be the earliest diverging branch of the liverwort phylogeny (Forrest and Crandall-Stotler 2005). The *nad7* gene of this taxon undergoes typical organellar RNA editing (Fig. 5-2) and group II intron splicing, and is proposed to represent the functional mitochondrial gene. Necessarily, the complete reading frame needs to be fully sequenced for ultimate clarity, since mutations in the flanking regions cannot be completely ruled out.

The observations support the model that both introns were present in a functional mitochondrial *nad7* gene of an extinct liverwort ancestor and were correctly spliced at that time. Their appearance obviously predates the occurrence of stop codons in the *nad7* gene, if the proposed liverwort phylogeny is correct. Three reported mismatches of exon binding sites (EBS) 1 & 2 of nad7i1113 are proposed to prevent its splicing in *Marchantia* (Takemura et al. 1995). The flanking exon site including intron binding sites (IBS) 1 & 2 was aligned with several angiosperm and moss data (not shown).

IBS1 & 2 of *Marchantia* are identical to the homologous exon region in several species, including *Arabidopsis*, tobacco, the moss *Takakia*, and *Haplomitrium*. Therefore the proposed mismatches, if correctly predicted, are derived from mutations of the exon binding sites in the intron. In contrast to that, the single mismatch that was reported for nad7i336 IBS1 is very likely an exon located mutation in *Marchantia*, because the corresponding nucleotide in the EBS1 was proposed to be a U, and all moss and angiosperm sequences share a corresponding A, whereas the *Marchantia* sequence displays a mismatching C.

Taken the knowledge that the introns were obtained before a pseudogene developed, they could still have been inserted into the gene before or after the nuclear transfer. It is also not clear on which point in early plant evolution that transfer occurred, because no nuclear PCR product could be obtained from any taxon, although primers were designed from *Marchantia*'s nuclear *nad7* sequence, and all analysed taxa were tested, including DNA of *Marchantia polymorpha*. Also no contradiction could be found for the model that the transfer process was mediated by mature mRNA, as was proposed by Kobayashi (1997). Indeed that process was frequently proposed, e.g. for the transfer of *cox2* to the nucleus in legumes (Adams et al. 1999) and for *rps10* in angiosperms (Adams et al. 2000), which was independently transferred several times. In the cases of angiosperms the nuclear version was always found to resemble the mature mRNA, because the nuclear sequence contained edited versions of the putative editing sites discovered on the mitochondrial DNA. This comparison is not useful for *Marchantia*, because the marchantiid liverworts do not edit their mitochondrial RNA, and therefore no distinction between DNA and mature mRNA sequence is possible except for the occurrence of introns. To obtain more information about the history of nuclear *nad7* in liverworts, it is necessary to analyze the nuclear versions of simple thalloid or leafy taxa that are known to be capable of RNA editing.

It seems rather likely that the pseudogenisation of the mitochondrial version took place after the divergence of the basal *Haplomitrium*-clade, supporting the recent theory of the *Haplomitrium/Treubia* clade as the earliest diverged group of all liverworts. Still, an independent loss of the mitochondrial *nad7* function in the two large liverwort groups Marchantiopsida (complex thalloids) and Jungermanniopsida (simple thalloids and leafy liverworts) can not be ruled out, although it is not the most parsimonious explanation.

The loss of a large part of exon 2 in *Blasia* versus the development of stop codons and frame shifts in the rest of the liverworts are very likely independent events, that were only possible because an alternative gene version was already in place in the nucleus and, most importantly, functional. The question remains, whether *Haplomitrium* or even *Chara* already carry a *nad7* copy in their nuclear genome, and whether it is already fused with a targeting sequence for the mitochondrion, which would be obligatory for a functional replacement, and was found for *cox2* in soybean (Nugent and Palmer 1991) and other legumes (Adams et al. 1999; Palmer et al. 2000). In the case of *nad7*, a nuclear type intron was indeed detected in the functional gene copy, accompanied by a 5'-extension, which probably comprises the targeting sequence for the transport of the resulting protein into the mitochondrion (Kobayashi et al. 1997).

6 Synopsis

Mitochondrial DNA in land plants: very slowly evolving gene sequences vs. rampant gene relocation and loss - what do we learn from the studies presented here?

This study concentrated (mostly) on the presumably earliest land plant groups, the bryophytes. Few molecular data have previously been accumulated from these organisms compared for instance to flowering plants, and most of them were obtained for pure phylogenetic studies, leaving particular modes of organelle genome evolution anappreciated. The evolution of such peculiar aspects as RNA editing, which is known from chloroplast and mitochondrial DNA of flowering plants, began already in bryophytes. As the editing frequency is much higher in mitochondrial than in chloroplast RNA, the study of selected additional mitochondrial genes in the earliest land plants promised further insights.

A major part of this thesis, chapter 2, is dedicated to the analysis of a novel marker gene, *nad4*, following its evolution from algae to angiosperms. Many parallels can be drawn between this gene and the already well understood *nad5* gene, which developed from a gene of interest for intron-splicing studies in angiosperms into a renowned phylogenetic tool for bryophytes and non-flowering plants. Both genes, *nad4* and *nad5*, have been shown to represent useful loci for phylogenetic analyses in liverworts (chapters 2 and 3), in the case of *nad4* here shown for the first time. The *nad5* gene exhibits an extraordinaryly high degree of RNA editing in such unique liverworts as *Haplomitrium*, and a significant number of reverse editing events in hornworts and ferns, and the same holds true for *nad4* (chapter 2). This supports the notion that a species, which reveals an elevated number of editing sites in one gene, is also prone to show the same picture in other genes, rendering this phenomenon lineage specific rather than an effect unique for single mitochondrial loci. The detailed analysis of *nad4* editing sites in 51 liverworts (chapter 2) also supports the previously noted hypothesis that no RNA editing occurs at all in the complex thalloid liverworts, and again no evidence for reverse U-to-C-editing has been found in any liverwort. Combining all data regarding this subject in early land plants with the phylogenetic studies incorporated in chapters 2 to 5 of this thesis, it can be concluded that the C-to-U-RNA editing developed in the common ancestor of all land plants, and is present to a very high degree in a member of

the potentially earliest diverging liverwort lineage: *Haplomitrium*. The missing evidence for editing events in complex thalloid liverworts reflects a potential secondary loss in this distinct and clearly monophyletic group. As the probably earliest diverging extant member of mosses, the genus *Takakia*, likewise exhibits a very high RNA editing rate, it could be postulated that the common ancestor of all land plants, which would also be the common ancestor of liverworts and mosses in our proposed land plant phylogeny, would similiarily exhibit much RNA editing. The general mechanism of land plant RNA editing has not been revealed so far, but current theories converge on the point that a number of nucleotide-sequence-specific RNA-binding proteins must be involved, because the observed editing events are extremely site specific in the analysed angiosperms. Assuming that the basal placement of *Haplomitrium* and *Takakia* is reflecting the real evolution of land plants, both could have inherited the ancestral RNA-binding protein(s). No reverse U-to-C-editing was found in either liverworts or mosses. This phenomenon appears for the first time in hornworts. Interestingly enough, *Leiosporoceros*, which is considered the possibly earliest diverging hornwort lineage, displays a very small degree of RNA editing compared with the average levels in hornworts.

Another form of “editing” RNA naturally is the splicing of introns. Organellar introns are distinguished by their secondary structure into the very differing groups I and II. Group I introns constitute the predominant type in green algae, and are present in parallel with group II introns in early land plants, whereas angiosperms carry only group II introns in their chondriomes, with a unique exception in *Peperomia*. A member of each group has been analysed in detail in the presented study, based on liverworts in both cases for enhanced comparability. The simple thalloid genus *Pellia* exhibited surprising differences in both the group I intron in *nad5* (chapter 3) and the group II intron in *nad4* (chapter 2), considerably reducing the size of the two introns. This is the first case of such an occurrence in non-tracheophytes, this effect has only been noticed before in the lycophyte *Isoetes* on several occasions, including a group II intron in *nad5*. Therefore special attention was given to these unique features (chapter 2 and 3).

Apart from their structure, the analysed introns were also part of phylogenetic studies obtained to gain insights into the phylogeny of liverworts (chapter 2 and 3), and, in the case of *nad4*, also with a small taxon sampling for mosses (chapter 2). These data were

compared to results from studies about another aspect of molecular evolution that can be studied most favourably on mitochondrial DNA: the rearrangement of genes and the evolution of the spacers that separate them (chapter 4). This study includes a first approach towards mitochondrial spacer regions in non-tracheophytes, starting out from a comparison of shared gene clusters between green algae and liverworts, and leading to interesting insights into very different developments of closely located spacers in the gene cluster *nad5-nad4-nad2* (chapter 4). This includes a study on the varying occurence of a tRNA gene in complex liverworts in the gene cluster *trnA-trnT-nad7* (chapter 4), a plant group that is otherwise known to have extremely conserved gene sequences.

Last, but not least, another peculiarity has been closely observed which is also so far a unique study in lower land plants: the evolution of the *nad7* pseudogene (chapter 5), providing an interesting view on the evolutionary mechanisms within plant chondriomes. It becomes clear that some liverwort groups, like the simple thalloids and the leafy liverworts (Jungermanniopsida), undergo a rather faster deconstruction of a pseudogene, involving the multiple occurence of frameshifts as is known from angiosperms, opposed to the very slow disintegration of the *nad7* gene in complex thalloid liverworts, which merely includes the introduction of stop codons, leaving the gene sequence otherwise intact.

Combining all data gained from many different angles about the evolution of early land plants, it becomes obvious that bryophytes are by no means one land plant group separate from tracheophytes. Indeed all details proved that “bryophytes” are clearly divided into three very different groups, liverworts, mosses, and hornworts, each of them exhibiting their own typical style of evolution, maybe seen best in the spacer study reported in chapter 4.

Outlook

Linking the presented findings on bryophytes to the most basal tracheophytes, the lycophytes and ferns, would be the next step in the evolutionary order. In that regard hornworts are of particular importance, as the analyses presented here helped to accumulate even more indications supporting them as the most closely related bryophyte group to tracheophytes.

Another very important step would be the sequencing and detailed analysis of a data set that is comprised of the complete chondriomes of several land plants, including such unusual liverworts like *Pellia* and *Haplomitrium*, and of course the first chondriome sequences of mosses, hornworts, lycophytes, ferns, and gymnosperms. This latter group has also not been included in the studies presented here, and sequences from these plants should be added to the data sets in further studies.

Clearly, this study has shown the way for a couple of interesting projects in the future, answering a few questions, but also never ceasing to produce new ones, always on the way to a more accurate picture of the true evolution of early land plants,

or, to say it with Charles Darwin's words:
„There will come a time, though I shall not live to see it, when we will have a fair representation of the genealogy of all living organisms".

7 Summary

Mitochondrial DNA in land plants is characterised by very slowly evolving gene sequences in contrast to a very variable genome structure displaying frequent gene relocations and transfers. Studies in this thesis address different aspects of chondriome evolution with a focus on early land plants. The mitochondrial *nad4* gene was established as a novel phylogenetic marker for liverworts, including an analysis of the secondary structure of the group II intron nad4i548 conserved in these plants. A review of the already well known mitochondrial *nad5* gene is accompanied by a revision of the folding structure of the included group I intron nad5i753, revealing peculiarities unique for the liverwort genus *Pellia*. A phylogenetic study on liverworts is reported, which combines the novel *nad4* sequences with *nad5* data from several labs, and the chloroplast genes *rbcL* and *rps4*, resulting in a well supported topology. The gene continuities of the gene clusters *nad5-nad4-nad2* and *trnA-trnT-nad7* were of special interest regarding the variability of the mitochondrial genome in the earliest diverging land plant lineages, and revealed very different patterns of evolution in the analysed spacers and among different plant groups. Loss and potential regain of the *trnT* gene was found in liverworts. Finally, studies on the degeneration of *nad7* into a pseudogene in liverworts identified different modes of sequence degeneration in the major liverwort subclades.

Zusammenfassung

Die mitochondriale DNA der Landpflanzen ist charakterisiert durch sehr langsam evolvierende Gensequenzen, die im Gegensatz zu einer durch häufige Umordnungen und Gen-Transfers bedingt sehr variablen Genomstruktur stehen. In dieser Dissertation werden verschiedene Aspekte der Chondriom-Evolution angesprochen, der Schwerpunkt liegt bei den frühen Landpflanzen. Das mitochondriale Gen *nad4* wird als neuer phylogenetischer Marker für die Lebermoose etabliert, was eine Analyse der Sekudärstruktur des in dieser Pflanzengruppe konservierten Gruppe II Introns nad4i548 einschließt. Ein Überblick über das bereits gut bekannte mitochondriale Gen *nad5* wird begleitet von einer Revision der Faltungsstruktur des darin enthaltenen Gruppe I Introns nad5i753, bei der ungewöhnliche Besonderheiten der Lebermoosgattung *Pellia* deutlich wurden. Eine phylogenetische Studie der Lebermoose, die sowohl die neuen *nad4* Sequenzen als auch *nad5* Daten verschiedener Arbeitsgruppen und die chloroplastidären Gene *rbcL* und *rps4* vereinigt, führt zu einer gut gestützten Topologie. Die Reihenfolge der Gene in den Gen-Gruppierungen *nad5-nad4-nad2* und *trnA-trnT-nad7* sind von besonderem Interesse für Untersuchungen der Variabilität des mitochondrialen Genoms in den am frühesten divergierenden Gruppen der Landpflanzen. Ihre Untersuchung führt zu der Präsentation sehr unterschiedlicher Evolutionsmodelle in den analysierten Spacern und verschiedenen Pflanzengruppen. Hierbei wird der Verlust und potentielle Wiedergewinn des *trnT* Gens in Lebermoosen gezeigt. Desweiteren führen Studien der Degeneration des *nad7* Gens zu einem Pseudogen in Lebermoosen zu der Identifizierung von unterschiedlichen Arten der Sequenzdegeneration in den Großgruppen der Lebermoose.

8 General Appendix

This appendix constitutes a short overview over the media and buffer used as well as their respective receipts (table 9-1).

Chemicals and plastics were obtained from Applichem, Roth, Labomedic, Sarstedt, Sigma, and Merck. Primers were synthesised by Invitrogen, Qiagen Operon, biomers.net, or Metabion.

Table 9-1: Media and buffer receipts

Medium / Buffer	Component	Final concentration
1 x TE-Buffer (pH 8.0)	EDTA-Na_4	1 mM
	Tris	10 mM
10 x TBE-Buffer (pH 8.0)	Tris	0.9 M
	Boric acid	0.2 M
	EDTA-Na_4	20 mM
3 x Loading Dye for agarose gels	cresol red	0.04 % (w/v)
	Ficoll	13.0 % (w/v)
LB medium (Luria broth):	NaCl	1 % (w/v)
complete mix, composed of:	tryptone	1 % (w/v)
	yeast extract	0.5 % (w/v)
for solid Petri dishes add:	agar agar	1.5 % (w/v)
for selective media add:	Ampicillin	100 µg/ml
SOC medium (pH 7.0)	KCl	2.5 mM
	NaCl	10 mM
	$MgCl_2$	10 mM
	$MgSO_4$	10 mM
	glucose	20 mM
	tryptone	2 % (w/v)
	yeast extract	0.5 (w/v)
CTAB buffer	CTAB	2 % (w/v)
	NaCl	1.4 M
	EDTA-Na_4	20 mM
	Tris-HCl	100 mM
	PVP 40	1 % (w/v)
add immediately before use:	β–Mercaptoethanol	100 mM
Buffer 1 for manual plasmid preparation (Bibdo 1)	Tris (pH 8.0)	25 mM
	EDTA- Na_4	10 mM
	glucose	50 mM
	lysozyme	4 mg / ml
Bibdo 2	NaOH	0.2 N
	SDS	1 % (w/v)
Bibdo 3	Potassium acetate	3 M
	acetic acid	11.5 % (v/v)

Additional kits used for plasmid preparation were obtained from Macherey-Nagel, Qiagen, or Eppendorf.

Procedures followed the provided instructions given in the manuals of all kits.

9 Acknowledgements

First of all I want to thank Prof. Volker Knoop for giving me the chance to study for my PhD under his highly skilled supervision and with his generous support. I have learned a lot from him and count myself lucky to have been a part of his research team.
My sincere thanks go also to Prof. Barthlott, who agreed to co-review my thesis on rather short notice, and to Prof. Wägele and Prof. Litt for their agreement to participate in my PhD committee.
I also want to extend my thanks to the rest of the team, with special thanks to Monika Polsakiewicz for being a caring and courageous chef of the lab, Katrin Weyand for her friendliness, excellent proof reading skills and a never ceasing supply with light American entertainment, Karolin Eifler, Felix Grexe, Michael Gebert, Theresia Rein, Julia Lee and the rest of the group for a nice working environment, interesting discussions and lots of fun, and Katja Kirsch for her never ending patience with too many dirty cuvettes and minipreps.
I want to thank the students of the four advanced block courses I co-led in Bonn, who taught me a lot about patience, self-control, and leadership qualities, and still found the time to produce some useful data.
I am grateful and indebted to several people who provided me with DNA or plant samples: Jochen Heinrichs, Prof. Hans Becker, Prof. Rob Gradstein, Melanie Lindner, Henk Groth, Michael Schwertfeger, Prof. Yin-Long Qiu, Laura Lowe Forrest, Prof. Sofia Szweykowska-Kulinska, Michael Stech, the Botanical Garden of Bonn, and the members of the former Knoop laboratory in Ulm.
Very special thanks go to my husband Henk, who always believed in me and supported me in every possible way. I want to extend my heartfelt thanks to my parents and my parents-in-law for their continuous support and trust.

10 References

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: 354-357.

Adams, K.L. and Palmer, J.D. 2003. Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Molecular Phylogenetics and Evolution* 29: 380-395.

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 *cox2* genes in legumes. *Proceedings of the National Academy of Sciences of the United States of America* 96: 13863-13868.

Beckert, S., Muhle, H., Pruchner, D., and Knoop, V. 2001. The mitochondrial *nad2* gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. *Molecular Phylogenetics and Evolution* 18: 117-126.

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

Bell, N.E. and Newton, A.E. 2004. Systematic studies of non-hypnanaean pleurocarps: Establishing a phylogenetic framework for investigating the origins of pleurocarpy. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 290-319. Missouri Botanical Garden Press, St. Louis.

Boisselier-Dubayle, M.C., Lambourdiere, J., and Bischler, H. 2002. Molecular phylogenies support multiple morphological reductions in the liverwort subclass Marchantiidae (Bryophyta). *Molecular Phylogenetics and Evolution* 24: 66-77.

Boisselier-Dubayle, M.C., Lambourdiere, J., Leclerc, M.C., and Bischler, H. 1997. Phylogenetic relationships in the Marchantiales (Hepaticae). Apparent incongruence between morphological and molecular data. *Evolution* 320: 1013-1020.

Bopp, M. and Capesius, I. 1996. New aspects of bryophyte taxonomy provided by a molecular approach. *Botanica Acta* 109: 368-372.

Burke, J.M. 1988. Molecular genetics of group I introns: RNA structures and protein factors required for splicing - A review. *Gene* 73: 259-271.

Campbell, D.H. 1924. A remarkable development of the sporophyte in *Anthoceros fusiformis*, Aust. *Annals of Botany* 38: 473-483.

Capesius, I. and Bopp, M. 1997. New classification of liverworts based on molecular and morphological data. *Plant Systematics and Evolution* 207: 87-97.

Carafa, A., Duckett, J.G., Knox, J.P., and Ligrone, R. 2005. Distribution of cell-wall xylans in bryophytes and tracheophytes: new insights into basal interrelationships of landplants. *New Phytologist* 168: 231-240.

Cargill, C.D., Renzaglia, K.S., Villareal, J.C., and Duff, R.J. 2005. Generic concepts within hornworts: historical review, contemporary insights and future directions. *Australian Systematic Botany* 18: 7-16.

Clifton, S.W., Minx, P., Fauron, C.M.R., Gibson, M., Allen, J.O., Sun, H., Thompson, M., Barbazuk, W.B., Kanuganti, S., Tayloe, C., Meyer, L., Wilson, R.K., and Newton, K.J. 2004. Sequence and Comparative Analysis of the Maize NB Mitochondrial Genome. *Plant Physiology* 136: 3486-3503.

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: 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: 299-316.

Crandall-Stotler, B.J. and Stotler, R.E. 2000. Morphology and classification of the Marchantiophyta. *In* Shaw, A.J. and Goffinet, B. [eds.]: *Bryophyte Biology*: 21-70. Cambridge University Press, Cambridge.

Davis, C.E. 2004. A molecular phylogeny of leafy liverworts (Jungermanniidae: Marchantiophyta). *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 61-86. Missouri Botanical Garden Press.

Dombrovska, O. and Qiu, Y.L. 2004. Distribution of introns in the mitochondrial gene *nad1* in land plants: phylogenetic and molecular evolutionary implications. *Molecular Phylogenetics and Evolution* 32: 246-263.

Doyle, J.J. and Doyle, J.L. 1990. Isolation of plant DNA from fresh tissue. *Focus* 12: 13-15.

Duff, R.J. 2006. Divergent RNA editing frequencies in hornwort mitochondrial *nad5* sequences. *Gene* In Press.

Duff, R.J. and Moore, F.B. 2005. Pervasive RNA editing among hornwort *rbcL* transcripts except *Leiosporoceros*. *Journal of Molecular Evolution* 61: 571-578.

Edwards, D. 2000. The role of mid-palaeozoic mesofossils in the detection of early bryophytes. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 355: 733-754.

Fiedorow, P., Odrzykoski, I., Szweykowski, J., and Szweykowska-Kulinska, Z. 2001. Phylogeny of the European species of the genus *Pellia* (Hepaticae; Metzgeriales) based on the molecular data from nuclear *tRNA*(*Leu*)(CAA) intergenic sequences. *Gene* 262: 309-315.

Forey, P.L., Fortey, R.A., Kenrick, P., and Smith, A.B. 2004. Taxonomy and fossils: a critical appraisal. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 359: 639-653.

Forrest, L.L. and Crandall-Stotler, B.J. 2004. A phylogeny of the simple thalloid liverworts (Jungermanniopsida, Metzgeriidae) as inferred from five chloroplast genes. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany from the Missouri Botanical Garden* 98: 119-140. Missouri Botanical Garden Press, St. Louis.

Forrest, L.L. and Crandall-Stotler, B.J. 2005. Progress towards a robust phylogeny for the liverworts, with particular focus on the simple thalloids. *Journal of the Hattori Botanical Laboratory* 97: 127-159.

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

Freyer, R., KieferMeyer, M.C., and Kossel, H. 1997. Occurrence of plastid RNA editing in all major lineages of land plants. *Proceedings of the National Academy of Sciences of the United States of America* 94: 6285-6290.

Garbary, D.J. and Renzaglia, K.S. 1998. Bryophyte phylogeny and the evolution of land plants: Evidence from development and ultrastructure. *In* Bates, J.W., Ashton, N.W., and Duckett, J.G. [eds.]: *Bryology for the Twenty-first century*: 45-63. Maney, Leeds.

Garbary, D.J., Renzaglia, K.S., and Duckett, J.G. 1993. The phylogeny of land plants - a cladistic analysis based on male gametogenesis. *Plant Systematics and Evolution* 188: 237-269.

Gass, D.A., Makaroff, C.A., and Palmer, J.D. 1992. Variable intron content of the NADH dehydrogenase subunit 4 gene of plant mitochondria. *Curent Genetics* 21: 423-430.

Geiss, K.T., Abbas, G.M., and Makaroff, C.A. 1994. Intron loss from the NADH dehydrogenase subunit 4 gene of lettuce mitochondrial DNA: evidence for homologous recombination of a cDNA intermediate. *Molecular and General Genetics* 243: 97-105.

Goffinet, B., Shaw, A.J., Cox, C.J., Wickett, N.J., and Boles, S.B. 2004. Phylogenetic inferences in the Orthotrichoideae (Orthotrichaceae, Bryophyta) based on variation in four loci from all genomes. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 270-289. Missouri Botanical Garden Press, St. Louis.

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. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany from the Missouri Botanical Garden* 98: 205-239. Missouri Botanical Garden Press, St. Louis.

Groth-Malonek, M. and Knoop, V. 2005. Bryophytes and other basal land plants: the mitochondrial perspective. *Taxon* 54: 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: 117-125.

Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

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: 5907-5916.

Haouazine-Takvorian, N., Takvorian, A., Jubier, M.F., and Lejeune, B. 1997. Genes encoding subunit 6 of NADH dehydrogenase and subunit 6 of ATP synthase are co-transcribed in maize mitochondria. *Current Genetics* 31: 63-69.

Hausner, G., Olson, R., Simon, D., Johnson, I., Sanders, E.R., Karol, K.G., McCourt, R.M., and Zimmerly, S. 2006. Origin and evolution of the chloroplast *trnK* (*matK*) intron: a model for evolution of group II intron RNA structures. *Molecular Biology and Evolution* 23: 380-391.

He-Nygren, X., Ahonen, I., Juslen, A., Glenny, D., and Piippo, S. 2004. Phylogeny of liverworts - Beyond a leaf and a thallus. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 87-118. Missouri Botanical Garden Press, St. Louis.

Hedderson, T.A., Chapman, R.L., and Cox, C.J. 1998. Bryophytes and the origins and diversification of land plants: New evidence from molecules. *In* Bates, J.W., Ashton, N.W., and Duckett, J.G. [eds.]: *Bryology for the twenty-first century* Maney and Br. Bryol. Soc., Leeds.

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

Hiesel, R., Combettes, B., and Brennicke, A. 1994. Evidence for RNA editing in mitochondria of all major groups of land plants except the Bryophyta. *Proceedings of the National Academy of Sciences of the United States of America* 91: 629-633.

Huelsenbeck, J.P. and Ronquist, F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

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

Kenrick, P. 2003. Fishing for the first plants. *Nature* 425: 248-249.

Kenrick, P. and Crane, P.R. 1997. The origin and early evolution of plants on land. *Nature* 389: 33-39.

Kim, K.J. and Lee, H.L. 2004. Complete chloroplast genome sequences from Korean ginseng (*Panax schinseng* Nees) and comparative analysis of sequence evolution among 17 vascular plants. *DNA Research* 11: 247-261.

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

Knoop, V., Kloska, S., and Brennicke, A. 1994. On the identification of group II introns in nucleotide sequence data. *Journal of Molecular Biology* 242: 389-396.

Knoop, V., Schuster, W., Wissinger, B., and Brennicke, A. 1991. Trans splicing integrates an exon of 22 nucleotides into the *nad5* mRNA in higher plant mitochondria. *EMBO Journal* 10: 3483-3493.

Kobayashi, Y., Knoop, V., Fukuzawa, H., Brennicke, A., and Ohyama, K. 1997. Interorganellar gene transfer in bryophytes: the functional *nad7* gene is nuclear encoded in *Marchantia polymorpha*. *Molecular and General Genetics* 256: 589-592.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris* L.) reveals a novel gene for tRNA(Cys)(GCA). *Nucleic Acids Research* 28: 2571-2576.

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

Kugita, M., Yamamoto, Y., Fujikawa, T., Matsumoto, T., and Yoshinaga, K. 2003b. RNA editing in hornwort chloroplasts makes more than half the genes functional. *Nucleic Acids Research* 31: 2417-2423.

Kumar, S., Tamura, K., and Nei, M. 2004. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150-163.

Lagerkvist, U. 1978. "Two out of Three": An alternative method for codon reading. *Proceedings of the National Academy of Sciences of the United States of America* 75: 1759-1762.

Lewis, L.A., Mishler, B.D., and Vilgalys, R. 1997. Phylogenetic relationships of the liverworts (Hepaticae), a basal embryophyte lineage, inferred from nucleotide sequence data of the chloroplast gene *rbcL*. *Molecular Phylogenetics and Evolution* 7: 377-393.

Long, D.G., Möller, M., and Preston, J. 2000. Phylogenetic relationships of *Asterella* (Aytoniaceae, Marchantiopsida) inferred from chloroplast DNA sequences. *The Bryologist* 103: 625-644.

Malek, O., Brennicke, A., and Knoop, V. 1997. Evolution of trans-splicing plant mitochondrial introns in pre-Permian times. *Proceedings of the National Academy of Sciences of the United States of America* 94: 553-558.

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: 1599-1609.

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: 1403-1411.

Manhart, J.R., Kelly, K., Dudock, B.S., and Palmer, J.D. 1989. Unusual characteristics of *Codium fragile* chloroplast DNA revealed by physical and gene mapping. *Molecular Genetics and Genomics* 216: 417-421.

Michel, F. and Dujon, B. 1983. Conservation of RNA secondary structures in two intron families including mitochondrial-, chloroplast- and nuclear-encoded members. *EMBO Journal* 2: 33-38.

Michel, F., Umesono, K., and Ozeki, H. 1989. Comparative and functional anatomy of group II catalytic introns - a review. *Gene* 82: 5-30.

Mishler, B.D. 1986. A Hennigian approach to bryophyte phylogeny. *Journal of Bryology* 14: 71-81.

Mishler, B.D., Lewis, L.A., Buchheim, M.A., Renzaglia, K.S., Garbary, D.J., Delwiche, C.F., Zechman, F.W., Kantz, T.S., and Chapman, R.L. 1994. Phylogenetic relationships of the green algae and bryophytes. *Annals of the Missouri Botanical Garden* 81: 451-483.

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: 1885-1895.

Nishiyama, T., Wolf, P.G., Kugita, M., Sinclair, R.B., Sugita, M., Sugiura, C., Wakasugi, T., Yamada, K., Yoshinaga, K., Yamaguchi, K., Ueda, K., and Hasebe, M. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Molecular Biology and Evolution* 21: 1813-1819.

Notsu, Y., Masood, S., Nishikawa, T., Kubo, N., Akiduki, G., Nakazono, M., Hirai, A., and Kadowaki, K. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. *Molecular Genetics and Genomics* 268: 434-445.

Nozato, N., Oda, K., Yamato, K., Ohta, E., Takemura, M., Akashi, K., Fukuzawa, H., and Ohyama, K. 1993. Cotranscriptional expression of mitochondrial genes for subunits of NADH dehydrogenase, *nad5*, *nad4*, *nad2*, in *Marchantia polymorpha*. *Molecular and General Genetics* 237: 343-350.

Nugent, J.M. and Palmer, J.D. 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., and Kohchi, T. 1992a. 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: 1-7.

Oda, K., Yamato, K., Ohta, E., Nakamura, Y., Takemura, M., Nozato, N., Akashi, K., and Ohyama, K. 1992b. Transfer RNA genes in the mitochondrial genome from a liverwort, *Marchantia polymorpha*: the absence of chloroplast-like tRNAs. *Nucleic Acids Research* 20: 3773-3777.

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: 6235-6250.

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y.L., and Song, K. 2000. Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences of the United States of America* 97: 6960-6966.

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

Pereira, d.S., Jubier, M.F., Delcher, E., Lancelin, D., and Lejeune, B. 1991. A trans-splicing model for the expression of the tripartite *nad5* gene in wheat and maize mitochondria. *The Plant Cell* 3: 1363-1378.

Posada, D. and Buckley, T.R. 2004. Model selection and model averaging in phylogenetics: advantages of akaike information criterion and bayesian approaches over likelihood ratio tests. *Systematic Biology* 53: 793-808.

Posada, D. and Crandall, K.A. 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.

Potemkin, A.D. 1999. Circumscription of the family Scapaniaceae, with segregation of the new family Diplophyllaceae (Hepaticae). *Annales Botanici Fennici* 36: 271-283.

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

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: 608-613.

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: 671-674.

Quandt, D. and Stech, M. 2005. Molecular evolution of the *trnL*UAA intron in bryophytes. *Molecular Phylogenetics and Evolution* 36: 429-443.

Renzaglia, K.S. 1982. A comparative developmental investigation of the gametophyte generation in the Metzgeriales (Hepatophyta). *Bryophytorum Bibliotheca* 24: 1-253.

Renzaglia, K.S. and Duckett, J.G. 1987. Comparative ultrastructural studies of spermatogenesis in the Metzgeriales (Hepatophyta) II. The blepharoplast of *Blasia pusilla*. *American Journal of Botany* 74: 53-64.

Renzaglia, K.S., Duff, R.J.T., Nickrent, D.L., and Garbary, D.J. 2000. Vegetative and reproductive innovations of early land plants: implications for a unified phylogeny. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* 355: 769-793.

Robart, A.R. and Zimmerly, S. 2005. Group II intron retroelements: function and diversity. *Cytogenetic and Genome Research* 110: 589-597.

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: 47-51.

Samigullin, T.H., Yacentyuk, S.P., Degtyaryeva, G.V., Valiejo-Roman, K.M., Bobrova, V.K., Capesius, I., Martin, W.F., Troitsky, A.V., Filin, V.R., and Antonov, A.S. 2003. Paraphyly of bryophytes and close relationship of hornworts and vascular plants inferred from analysis of chloroplast rDNA ITS (cpITS) spacer sequences. *Arctoa* 11: 31-43.

Schill, D.B., Long, D.G., Moeller, M., and Squirrell, J. 2004. Phylogentic relationships between Lophoziaceae and Scapaniaceae based on chloroplast sequences. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 141-149. Missouri Botanical Garden Press, St. Louis.

Schljakov, R.N. 1972. On the higher taxa of liverworts - class Hepaticae s. str. (in Russian). *Botanitscheskii Journal* 57: 496-508.

Schuster, R.M. 1984. Evolution, phylogeny and classification of the Hepaticae. *In* Schuster, R.M. [ed.]:*New Manual of Bryology* 2: 892-1070. Hattori Botanical Laboratory, Nichinan.

Schuster, R.M. 1992. *Hepaticae and Anthocerotae of North America V.* Field Museum of Natural History, Chicago.

Schuster, R.M. 2000. Austral Hepaticae, I. *Nova Hedwigia* Beih. 118: 1-524.

Simpson, C.L. and Stern, D.B. 2002. The treasure trove of algal chloroplast genomes. Surprises in architecture and gene content, and their functional implications. *Plant Physiology* 129: 957-966.

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

Stech, M. and Frey, W. 2001. CpDNA-relationship and classification of the liverworts (Hepaticophytina, Bryophyta). *Nova Hedwigia* 72: 45-58.

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: 303-312.

Sugiura, C., Kobayashi, Y., Aoki, S., Sugita, C., and 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 Research* 31: 5324-5331.

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: 603-615.

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: 565-570.

Toor, N., Hausner, G., and Zimmerly, S. 2001. Coevolution of group II intron RNA structures with their intron-encoded reverse transcriptases. *RNA* 7: 1142-1152.

Turmel, M., Otis, C., and 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. *Proceedings of the National Academy of Sciences of the United States of America* 99: 11275-11280.

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: 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: 57-61.

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

Wang, X.Q., Tank, D.C., and Sang, T. 2000. Phylogeny and divergence times in Pinaceae: evidence from three genomes. *Molecular Biology and Evolution* 17: 773-781.

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: 793-803.

Weis, G. 2001. Morphologische und anatomische Untersuchungen der Sporophyten bei den Jubulaceae Klinggr. und Lejeuneaceae Caesares-Gil (Hepaticae) und deren systematische Bedeutung. *Bryophytorum Bibliotheca* 57: 1-307.

Wellman, C.H., Osterloff, P.L., and Mohiuddin, U. 2003. Fragments of the earliest land plants. *Nature* 425: 282-285.

Wheeler, J.A. 2000. Molecular phylogenetic reconstructions of the marchantioid liverwort radiation. *The Bryologist* 103: 314-333.

Wolf, P.G., Rowe, C.A., and Hasebe, M. 2004. High levels of RNA editing in a vascular plant chloroplast genome: analysis of transcripts from the fern *Adiantum capillus-veneris*. *Gene* 339: 89-97.

Wolfe, K.H., Morden, C.W., and Palmer, J.D. 1992. Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proceedings of the National Academy of Sciences of the United States of America* 89: 10648-10652.

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.

Yamato, K., Nozato, N., Oda, K., Ohta, E., Takemura, M., Akashi, K., and Ohyama, K. 1993. Occurrence and transcription of genes for *nad1*, *nad3*, *nad4L*, and *nad6*, coding for NADH dehydrogenase subunits 1, 3, 4L, and 6, in liverwort mitochondria. *Current Genetics* 23: 526-531.

Yatsentyuk, S.P., Konstantinova, N.A., Ignatov, M.S., Hyvönen, J., and Troitsky, A.V. 2004. On phylogeny of Lophoziaceae and related families (Hepaticae, Jungermanniales) based on *trnL-trnF* intron-spacer sequences of chloroplast DNA. *In* Goffinet, B., Hollowell, V., and Magill, R. [eds.]: Molecular Systematics of Bryophytes. *Monographs in Systematic Botany* 98: 150-165. Missouri Botanical Garden Press, St. Louis.

Yoshinaga, K., Iinuma, H., Masuzawa, T., and Uedal, K. 1996. Extensive RNA editing of U to C in addition to C to U substitution in the *rbc*L transcripts of hornwort chloroplasts and the origin of RNA editing in green plants. *Nucleic Acids Research* 24: 1008-1014.

Yoshinaga, K., Kakehi, T., Shima, Y., Iinuma, H., Masuzawa, T., and Ueno, M. 1997. Extensive RNA editing and possible double-stranded structures determining editing sites in the *atp*B transcripts of hornwort chloroplasts. *Nucleic Acids Research* 25: 4830-4834.