MOLECULAR PHYLOGENETICS OF MOSSES AND RELATIVES

by

Ying Chang

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies

(Botany)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

July 2011

ABSTRACT

Substantial ambiguities still remain concerning the broad backbone of moss phylogeny. I surveyed 17 slowly evolving plastid genes from representative taxa to reconstruct phylogenetic relationships among the major lineages of mosses in the overall context of land-plant phylogeny. I first designed 78 bryophyte-specific primers and demonstrated that they permit straightforward amplification and sequencing of 14 core genes across a broad range of bryophytes (three of the 17 genes required more effort). In combination, these genes can generate sturdy and well-resolved phylogenetic inferences of higher-order moss phylogeny, with little evidence of conflict among different data partitions or analyses. Liverworts are strongly supported as the sister group of the remaining land plants, and hornworts as sister to vascular plants. Within mosses, besides confirming some previously published findings based on other markers, my results substantially improve support for major branching patterns that were ambiguous before. The monogeneric classes Takakiopsida and Sphagnopsida likely represent the first and second split within moss phylogeny, respectively. However, this result is shown to be sensitive to the strategy used to estimate DNA substitution model parameter values and to different data partitioning methods. Regarding the placement of remaining nonperistomate lineages, the [[[Andreaeobryopsida, Andreaeopsida], Oedipodiopsida], peristomate mosses] arrangement receives moderate to strong support. Among peristomate mosses, relationships among Polytrichopsida, Tetraphidopsida and Bryopsida remain unclear, as do the earliest splits within sublcass Bryidae. A [Funariidae, [Timmiidae, [Dicranidae, Bryidae]]] arrangement is strongly supported, as are major relationships within subclasses Funariidae and Dicranidae. I also reconstructed the phylogeny of

the nonperistomate moss family Andreaeaceae, with a focus on costate taxa, using two complementary sets of plastid markers and taxa. The major subgenera (*Andreaea* and *Chasmocalyx*) and sections of *Andreaea* (*Andreaea* and *Nerviae*) are rejected as monophyletic. Well-supported lineages include clades comprising: (1) *Andreaea nivalis* and *A. rigida* (northern hemisphere members of subgenus *Chasmocalyx*) and *A. blyttii* (section *Nerviae*); (2) most of the remainder of *Nerviae*; (3) a mixture of costate and ecostate species from *Chasmocalyx, Nerviae*, all sampled members of section *Andreaea*, and subgenus *Acroschisma*. Relationships among the major lineages, including the root of the family, are all well supported.

PREFACE

A version of Chapter 2 has been published: Chang, Y., and S. W. Graham (2011), Bryophyte-specific primers for retrieving plastid genes suitable for phylogenetic inference. American Journal of Botany 98: e109-e113. I carried out the laboratory work, designed the primers, and wrote the manuscript. Sean W. Graham provided valuable guidance with respect to primer design and writing.

A version of Chapter 3 has been published: Chang, Y., and S. W. Graham (2011), Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. American Journal of Botany 98: 839-849. I carried out the laboratory work and data analyses, and wrote the manuscript. Sean W. Graham provided insights into taxonomic sampling, data analyses and contributed to the writing.

A version of Chapter 4 will be submitted for publication: Chang, Y., and S. W. Graham, Backbone phylogeny of peristomate and nonperistomate mosses based on multiple plastid markers. I carried out the laboratory work and data analyses, and wrote the manuscript. Sean W. Graham provided insights into taxonomic sampling, data analyses and contributed to the writing.

A version of Chapter 5 will be submitted for publication: Chang, Y., and S. W. Graham, Molecular phylogenetics of Andreaeaceae. I carried out the laboratory work and data analyses, and wrote the manuscript. Sean W. Graham provided insights into data analyses and contributed to the writing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

First and foremost I want to thank my Ph.D supervisor, Dr. Sean W. Graham, for his all his contributions of time, ideas, and funding. His guidance through my Ph.D. period has been extremely valuable to me. I also thank other members of my supervisory committee (Drs. Wilf Schofield, Mary Berbee, and Wayne Maddison) for their help to my thesis work.

I want to express my gratitude to Dr. Barbara Murray. Not only did she provide me with valuable specimens, she has also generously shared her understanding of the morphology and evolution of Andreaeaceae with me.

I would like to thank former and current members from Graham lab, especially Vivienne Lam and Will Iles, for all the discussions, suggestions and help during the course of my degree. I owe my thanks to Shona Ellis, whose love and enthusiasm for bryophytes are contagious. It would not have been possible to write this doctoral thesis without the help and support of many other colleagues and friends. Thank you!

Finally, I would like to thank my parents for all their love and encouragement, and my husband Jun, who has always been loving, patient, and encouraging.

Dedication

Dr. Wilfred B. Schofield

1 INTRODUCTION

1.1 Overview of early land-plant phylogeny

Land plants (embryophytes) comprise four major extant lineages: liverworts (Marchantiophyta), mosses (Bryophyta), hornworts (Anthoceratophyta), and vascular plants (Tracheophyta). The former three are collectively referred as bryophytes. All land plants have a sporic life cycle, i.e., an alternation of generations. Unlike vascular plants, bryophytes are characterized by a gametophyte-dominant life history: the gametophyte generation is the major photosynthesizing phase, while the sporophyte generation is at least partially dependent on the female gametophyte for nutrients. The sporophytes of all extant bryophytes are unbranched and bear a single sporangium, in contrast to the branched (polysporangiate) sporophytes found in vascular plants. These attributes were considered to be evidence of the unique common ancestry shared by all bryophytes (Vanderpoorten and Goffinet, 2009). However, since the middle 1900s, various authors have challenged the monophyly of bryophytes and proposed that they represent independent evolutionary lines (e.g., Bold, 1956; Steere 1969; Crandall-Stotler, 1980). Early cladistic studies on morphological and biochemical characters by Mishler and Churchill (1984, 1985) demonstrated the paraphyly of bryophytes. Since then, an increasing number of studies has formed a consensus that each bryophyte lineage is monophyletic and that they form a grade subtending vascular plants (e.g., Kenrick & Crane, 1997; Garbary and Renzaglia, 1998; Qiu et al., 1998; however, Nishiyama, 2004; Goremykin and Hellwig, 2005). In contrast, the branching order of the three bryophyte lineages, in relation to each other and the vascular plants, still remains controversial (Qiu, 2008). Various hypotheses have been proposed over the years based

on multiple lines of data, but rarely with convincing bootstrap support (e.g., Lewis et al., 1997; Nickrent et al., 2000; Kelch et al., 2004; see also Fig. 1.1). An unambiguous reconstruction of early land-plant phylogeny is essential for understanding plant evolution, as it would help in reconstructions of the early morphological innovations that developed in land plants after the invasion of the terrestrial environment, and also provide a comparative framework for studying the genome evolution of early land plants.

1.2 Overview of backbone relationships of bryophytes

Among the three bryophyte lineages, hornworts are the smallest group, with approximately 150 species (Renzaglia et al., 2009). Hornworts are so named because of their horn-like sporophytes. Unlike the sporophytes of liverworts and mosses, the hornwort sporophyte possesses a basal intercalary meristem and so has indeterminate growth (in most species). The gametophyte is always thalloid and does not differentiate into stems and leaves. Hornworts are also known for their symbiotic relationship with nitrogen-fixing cyanobacteria and endomycorrizhal fungi (Vanderpoorten and Goffinet, 2009). Compared to the other two bryophyte groups, hornworts remain relatively unexplored. The number of species and genera found in hornworts has only been crudely estimated (Duff et al., 2007), and due to different opinions on the interpretation of some morphological characters, hornwort classification schemes have been highly variable (reviewed in Duff et al., 2004). In the past decade, molecular markers have been used to elucidate the phylogenetic relationships of hornworts (e.g., Stech et al., 2003; Duff et al., 2007). A broad degree of consensus has now been reached on this (Fig. 1.2a), corroborated by subsequent ultrastructural and developmental studies (Renzaglia et al., 2007). The monotypic *Leiosporoceros* is found to be highly divergent compared to the other hornworts

(Duff et al., 2004) and is now recognized as a class, Leiosporocerotopsida (Duff et al., 2004; Renzaglia et al., 2009). The remaining taxa are included in the other hornwort class, Anthocerotopsida. Three well-segregated clades are found within Anthocerotopsida (Duff et al., 2004 and 2007), each treated as a subclass (Anthocerotidae, Notothylatidae and Dendrocerotidae).

There are about 5,000 described liverwort species in total. Unlike the uniform thalloid gametophyte found in hornworts, the gametophytes of liverworts demonstrate a remarkable morphological diversity. They can be leafy in form with different leaf insertion patterns, and different numbers of lobes, or thalloid, with or without much internal differentiation. The sporophytes of liverworts, on the other hand, are more or less uniform among different liverwort species. Two major evolutionary lines were traditionally recognized in liverworts, based mainly on gametophytic characters (Schofield, 1985). The liverworts have been divided into two groups: marchantioids (= complex thalloid liverworts), and jungermannioids (= leafy and simple thalloid liverworts) (Crandall-Stotler et al., 2009). A large number of molecular studies have been carried out on the higher-order phylogeny of liverworts. These studies have brought new insights into the backbone relationships of liverworts (e.g., Davis, 2004; Forrest et al., 2006; He-Nygén, et al., 2006; Wilson et al., 2007). One of the major findings in these studies is the recovery of a clade comprising *Haplomitrium* and *Treubia,* two leafy taxa, as the sister-group of all other liverworts. *Blasia* and *Cavicularis,* two simple thalloid liverworts, then comprise a clade (Blasiidae) that is the sister group of the complex thalloid liverworts (Marchantiidae), and the clade consisting of Marchantiidae and Blasiidae (now collectively recognized as Marchantiopsida) is the sister group of the-Jungermannioids (Jungermanniopsida). Within Jungermanniopsida, Pelliidae and

Metzgeriidae form a grade of simple thalloid liverworts subtending the leafy taxa (Jungermanniidae) (Fig. 1.2b).

With about 13,000 species, mosses are the largest bryophyte group and the second largest major plant group after flowering plants (Cox et al., 2010). Unlike liverworts and hornworts, the gametophyte of extant mosses is always leafy, with leaves usually spirally arranged around the stem. A typical moss sporophyte is characterized by having a robust seta and a ring or two of peristome teeth aligned with the opening of the capsule (sporangium), although multiple nonperistomate lineages (which have capsules opening without teeth) represent the earliest splits in moss phylogeny (Fig. 1.2). The nonperistomate mosses are mostly species-poor (one species each in Oedipodiopsida and Andreaeobryopsida, two species in Takakiopsida, ~50 species in Andreaeopsida, and ~200 species in Sphagnopsida), but they represent ancient and morphologically distinctive evolutionary lines whose relationships to each other are still not clear (Goffinet and Buck, 2004).

Our understanding of the overall classification and relationships of the peristomate mosses has traditionally relied heavily on characters associated with the architecture and development of peristome teeth, due to their (presumed) conservative nature in moss evolution (e.g., Fleischer, 1904-1923; Dixon, 1932; Vitt, 1984). There are two major types of peristome teeth: nematodontous and arthrodontous teeth. The former type is made up of whole cells, while the latter is composed of cell-wall fragments. Nematodontous taxa are further arranged into two classes, Tetraphidopsida and Polytrichopsida. Both classes are species-poor, with only five species in Tetraphidopsida and ~200 species in Polytrichopsida (Bell and Hyvonen, 2010). In contrast, the only arthrodontous moss class, Bryopsida, harbours about 90% of all moss species

(~12,000 species) and shows great diversity in peristome morphology (Goffinet and Buck, 2004). There are three major types arthrodontous teeth: diplolepideous-opposite, diplolepideous-alternate and haplolepideous. These major types of peristome teeth respectively characterize three subclasses: Funariidae, Bryidae and Dicranidae. However, not all types of peristome teeth fall nicely into one of the three categories. For example, the teeth found in *Buxbaumia* (Buxbaumiidae) appear transitional between nematodontous and arthrodontous types, and *Timmia* (Timmiidae) possesses teeth with a unique architecture that is hard to interpret (Budke et al., 2007).

Despite extensive studies on the development and structure of peristome teeth, their evolutionary transitions are still poorly understood (Goffinet et al., 2001). Moreover, the morphologies of peristome teeth are generally too conserved to provide information about relationships within major groups (Vitt et al., 1998; Magombo, 2003). Many gametophytic characters, on the other hand, are known to be highly homoplastic and plagued with reversals and convergences, making it potentially problematic to use them in the inference of moss backbone relationships (Renzaglia et al., 2007; Goffinet et al., 2009)

In the past two decades, DNA sequence data, or a combination of DNA sequence and morphological data, have been used in a considerable number of phylogenetic studies to investigate higher-order relationships among major moss groups (peristomate and nonperistomate) (e.g., Newton et al., 2000; Beckert et al., 2001; Goffinet et al., 2001; Magombo 2003; Cox et al., 2004; Wahrmund et al., 2010). Despite multiple examples of concordance among different phylogenetic studies, there are also multiple points along the backbone of moss phylogeny that do not have satisfactory resolution–i.e., those that are either polytomous, poorly

supported, contradictory among studies, or unsatisfactorily corroborated by morphological synapomorphies. Higher-order moss phylogeny is therefore in need of substantial additional investigation (Goffinet et al., 2009). This is the major focus of my thesis (Chapters 2-4).

Among nonperistomate classes that represent the early splits in moss evolution, only two comprise more than two species: Andreaeopsida and Sphagnopsida. Unlike Sphagnopsida, which has been well studied (e.g., Ligrone and Duckett, 1998; Vitt, 2006; Shaw et al., 2010), the systematics of the monofamililial Andreaeopsida is mostly understudied. There has been no treatment of the whole family since the early work of Roth (1903-1904, 1910-1911), and phylogenetic relationships within the family have not been addressed. A robust phylogeny of the family would not only help improve our understanding of evolutionary transitions in the family, but it would also help to evaluate the naturalness of its current classification, and provide evidence for future taxonomic revisions of the family, if needed. Furthermore, it would help to place the family more firmly in early moss phylogeny and improve our understanding of the deepest splits in moss evolution. This is the focus of the research described in Chapter 5.

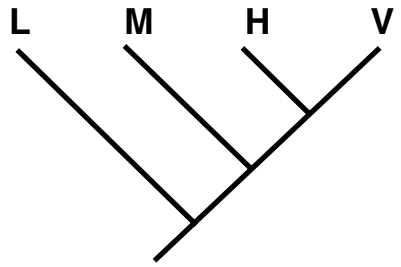
1.3 Objectives of the thesis

The overall goal of Chapters 2-4 is to provide a more resolved and better supported picture of the higher-order phylogeny of mosses and relatives, and in particular to improve our current understanding of the early branches of moss phylogeny. The inference of the higher-order relationships of bryophytes has relied on a limited number of molecular markers to date (Stech and Quandt, 2010). In this thesis, I develop 78 new bryophyte-specific primers to retrieve a suite of plastid markers (Chapter 2). The targeted plastid regions include four single genes (*atp*B, *ndh*F, *rbc*L and *rpl*2), and four multiple gene clusters (*psb*E-*psb*F-*psb*L-*psb*J, *psb*D-*psb*C,

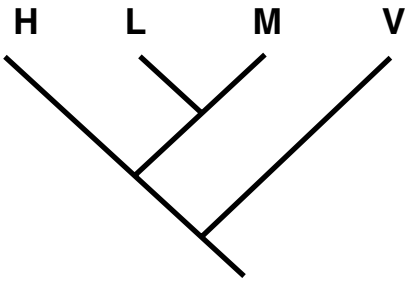
*psb*B-*psb*T-*psb*N-*psb*H, *rps*12-*rps*7-*ndh*B). These regions have proved useful for elucidating higher-order relationships among various vascular plant groups (Graham and Olmstead, 2000) and should be suitable for reconstructing bryophyte backbone phylogenies. I assess the utility of these primers in retrieving these regions using a test set of exemplar species that represent a broad range of bryophyte taxa.

I demonstrate the performance of the targeted DNA markers for inferring the broad phylogenetic backbone of mosses and their placement in land-plant phylogeny by first focusing on the nonperistomate moss lineages (Chapter 3). In this chapter I also examine whether identification and removal of the most rapidly evolving nucleotides in the plastid markers used here may have an impact on the phylogenetic inference using parsimony (as compared to maximum likelihood). I then substantially expand taxon sampling in peristomate mosses to characterize the overall phylogenetic relationships within this major moss clade (Chapter 4). In this chapter I also investigate the effect of DNA substitution model parameter settings on maximum likelihood (ML) based inference (Chapter 4). Finally, in Chapter 5 I present two complementary sets of plastid markers used at different densities of taxon sampling to characterize the broad phylogenetic structure of the granite moss family, Andreaeaceae, which represents the second largest group of nonperistomate mosses. I assess the monophyly of current infrageneric taxonomy (subgenera, sections and species) and summarize initial evidence on the evolution of some morphological characters and their potential utility in future classification schemes for the family (Chapter 5).

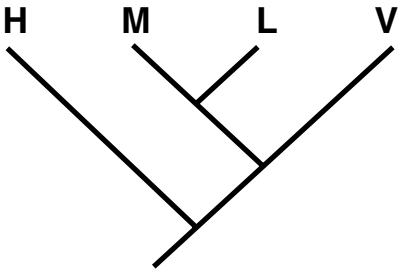
Fig. 1.1. Selection of alternative hypotheses regarding land-plant relationships. H: hornworts; L: liverworts; M: mosses; V: vascular plants



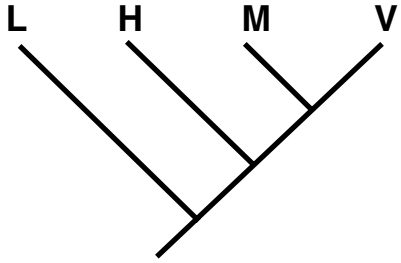
e.g., Qiu et al., 2006



e.g., Nishiyama et al., 2004



e.g., Renzaglia et al., 2000



e.g., Mishler and Churchill, 1984

Fig. 1.2a. Summary of higher-order relationships of hornworts (Anthocerotophyta). The classification follows Renzaglia et al. (2009).

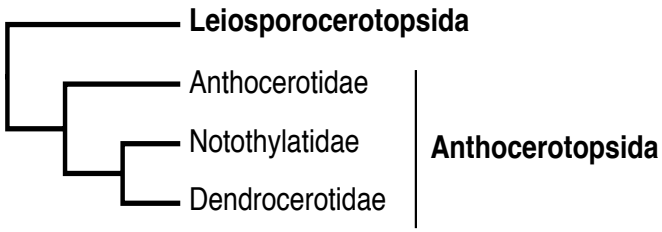


Fig. 1.2b. Summary of higher-order relationships of liverworts (Marchantiophyta). The classification follows Crandall-Stotler et al. (2009).

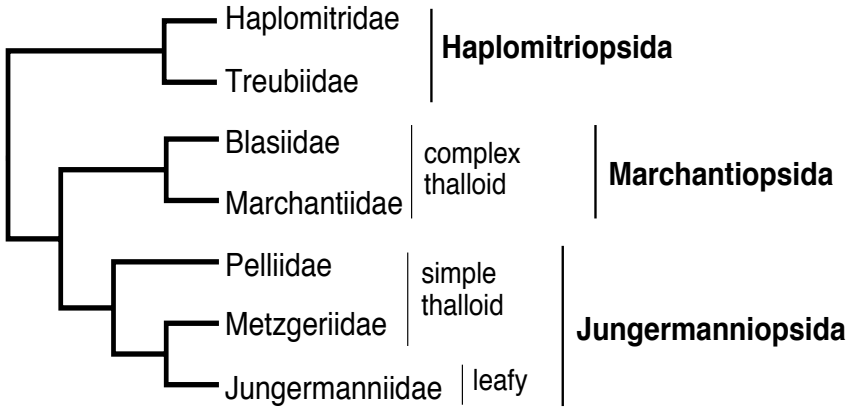
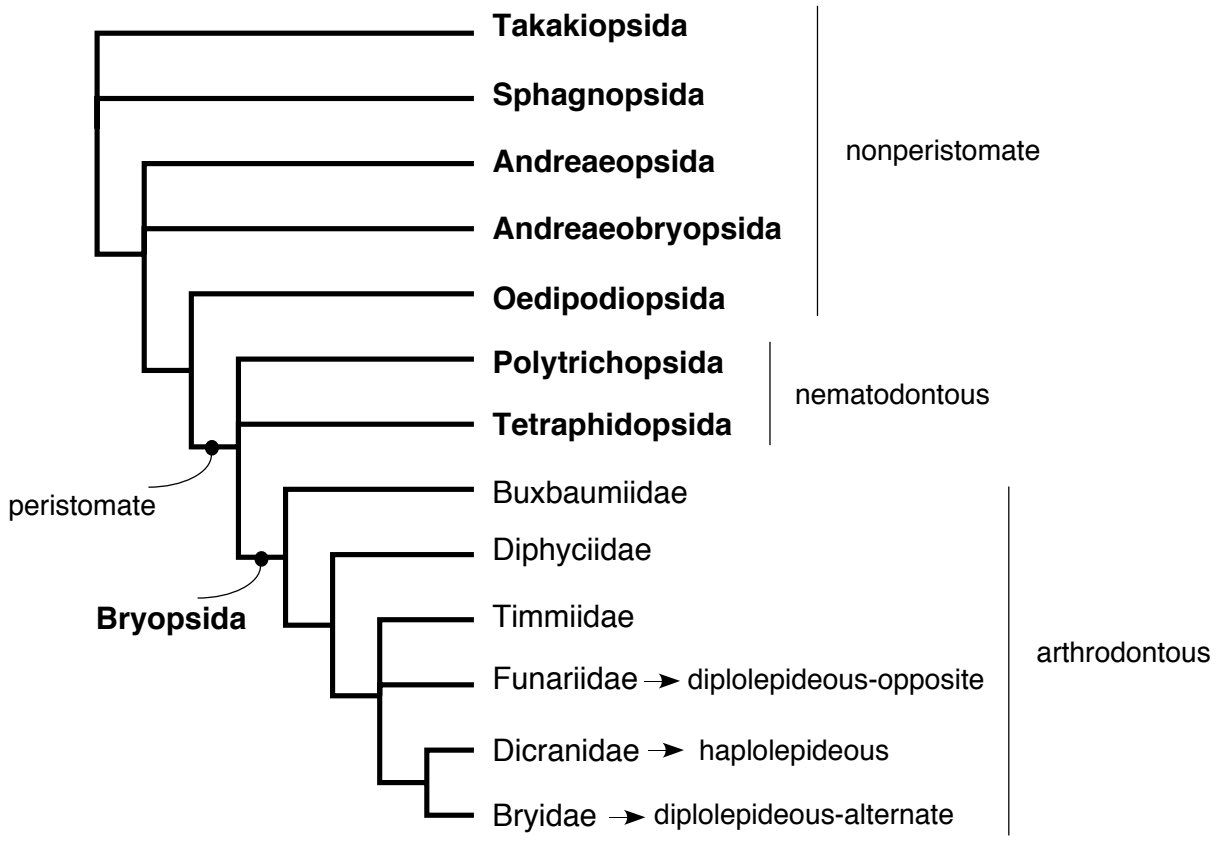


Fig. 1.3. Summary of higher-order relationships of mosses. The classification follows Goffinet et al. (2009)



2 BRYOPHYTE-SPECIFIC PRIMERS FOR RETRIEVING PLASTID GENES SUITABLE FOR PHYLOGENETIC INFERENCE[1]

2.1 Brief synopsis

I present here a total of 78 new bryophyte-specific primers that permit retrieval of 17 slowly evolving plastid genes and their associated introns and intergenic spacers. These regions were chosen to facilitate accurate phylogenetic inference across a broad range of mosses and other bryophytes. The primers were designed using an initial sampling of exemplar bryophytes and other plants. I assessed the ability of the new primers to amplify and sequence these regions using a test set of 11 exemplar bryophytes. The newly designed primers allowed ready retrieval of 14 of the 17 targeted genes from a broad range of bryophyte taxa, and should prove useful for future studies of bryophyte phylogeny.

2.2 Introduction

The higher-order relationships of mosses (Bryophyta) are arguably understudied compared to other major embryophyte clades, and studies of higher-order moss phylogeny have focused on relatively few plastid genome regions to date. Among the most widely used plastid markers are the *rbc*L and *rps4* loci and the *trn*L-*trn*F region (e.g., Tsubota et al., 2003; Cox et al., 2004). I have developed a series of new primers to amplify and sequence multiple additional

[1]A version of this chapter has been published: CHANG, Y., AND S. W. GRAHAM. 2011. Bryophyte-specific primers for retrieving plastid genes suitable for phylogenetic inference. *Am. J. Bot.* 98:e109-e113

conserved regions in the plastid genomes of bryophytes. The regions that can be surveyed with these primers are those used successfully in the inference of higher-order relationships at multiple levels of angiosperm, seed-plant and vascular-plant phylogeny (e.g., Graham and Olmstead, 2000). They comprise 10 photosystem II (*psb*) genes located in three major clusters (*psb*D-*psb*C, *psb*E-*psb*F-*psb*L-*psb*J, and *psb*B-*psb*T-*psb*N-*psb*H), two NADH dehydrogenase subunit genes (*ndh*F and *ndh*B), three ribosomal protein genes (3'-*rps*12, *rps*7 and *rpl*2), an ATP synthase subunit gene (*atp*B), and the gene coding for the large subunit of Rubisco (*rbc*L). Among the 17 loci that I surveyed, only *rbc*L has been widely used in other studies of bryophyte phylogeny (e.g., Tsubota et al., 2003; Chang et al., 2006), although some of them have been sampled sporadically using (in part) primers previously designed for tracheophytes (e.g., the *psb*T-*psb*N-*psb*H region, Quandt et al., 2003). Here I document the utility of these primers in complementing the previously published primers for retrieving the targeted regions, based on a preliminary sampling of "exemplar" (representative) mosses and other bryophytes (hornworts and liverworts; Anthocerotophyta and Marchantiophyta).

2.3 Methods and results

2.3.1 "Universal" primer design and testing

I chose candidate priming sites in plastid regions used previously to infer higher-order vascular-plant phylogeny, and designed oligonucleotides for amplification and sequencing of bryophytes using criteria described in Graham and Olmstead (2000). This design process involved several iterations of DNA sequence alignments comprising new and previously published plastid sequences (Appendix 1). I employed Oligo 4.0-s (National Biosciences Inc.

Plymouth, Minnesota, USA) to assess GC content and melting temperature of oligos, and to test for problematic hairpin structures or primer dimers (see Graham and Olmstead, 2000). Twelve taxa (Appendix 1) were newly sequenced during primer development. I also tested how well the new primers work in amplification and sequencing using an additional batch of "test" taxa (i.e., seven additional mosses, two liverworts and two hornworts, Appendix 1).

2.3.2 DNA extraction, amplification and sequencing

I extracted total genomic DNA from exemplar taxa (Appendix 1), following the method of Doyle and Doyle (1987) or using a QIAGEN DNeasy plant mini kit (QIAGEN, Valencia, California, USA). The PCR profile includes an initial denaturation at 94^{o}C for 5 min, followed by 35-40 cycles of 94^{o}C for 1 min, 45^{o}, 43^{o} or 41^{o}C for 1 min, 72^{o}C for 2 min, with a final extension at 72^{o}C for 10 min. I cleaned PCR products with the QIAquick PCR purification kit (QIAGEN, Valencia, California, USA) and performed direct sequencing using an ABI BigDye Terminator kit version 3.1 (Applied Biosystems, Foster City, California, USA). The reactions were run on 377 or 3700 automated sequencers (Applied Biosystems, Foster City, California, USA).

2.3.3 Sequence editing and alignment

I initiated this study by sequencing portions of the regions of interest using primers previously designed for vascular plants (Graham and Olmstead, 2000; Rai and Graham, 2010). I compiled chromatograms and called bases using Sequencher ver. 4.1 (Gene Codes Corp., Ann Arbor, Michigan, USA). I aligned the new sequences from each major region with a selection of previously published sequences from a phylogenetically diverse range of vascular plants and

charophycean algae with Se-Al ver. 2.0 (Rambaut, 2002), using criteria laid out in Graham et al. (2000). I included sequences only from the coding regions of the vascular plants and algae to minimize alignment ambiguities, but aligned all noncoding regions across bryophytes. I used these alignments to design more primers, and repeated this design process until I obtained satisfactory primer coverage for the targeted regions.

In total I designed 77 new primers for the targeted regions (Figs. 2.1-2.7). For the region spanning *psb*E, *psb*F, *psb*L and *psb*J, previously published primers (see Fig. 4 in Graham and Olmstead, 2000) generally worked well across the bryophytes that I surveyed. For *ndh*F (Fig. 2.3), the portion of the gene that I examined is the more conservative 5'-end (Olmstead and Sweere, 1994). After extensive primer design (Fig. 2.7) and testing I was unable to develop a core set of primers that work well across bryophytes for the *rps*12-*rps*7-*ndh*B region, and so I abandoned further primer design for it.

I amplified and generated sequencing products for the remaining 14 core targeted regions for a further seven mosses, two liverworts and two hornworts (Appendix 1). Focusing on these taxa, I found amplifications typically worked with only one or two trials per amplification product. I could not amplify the *ndh*F region for the liverworts *Haplomitrium* and *Treubia*, the *rpl*2 region for the hornworts *Leiosporoceros* and *Phaeoceros,* or the *psb*E-*psb*F-*psb*L-*psb*J region for the moss *Tetrodontium,* the liverwort *Haplomitrium*, and the hornwort *Phaeoceros*. The primers worked for sequencing with fewer trials on average for mosses, followed by liverworts and then hornworts (Table 2.1; this table is not a systematic exploration of all the new primers, but rather is intended to document my experience based on regular lab flow when working with each region. As a guide the most commonly used sequencing primers are noted in

Figs. 2.1-2.6 using an asterisk). This is not surprising, since much of my primer development work was based on alignments that included a high proportion of mosses. I demonstrated the ability of these conserved regions to infer splits along the backbone of moss phylogeny by using a subset of taxa I sampled in this study, together with new and previously published sequences (Chapter 3). These genes yield sturdy and well-resolved phylogenetic inferences using maximum likelihood (ML) analysis (Chapters 3, 4) that are largely congruent with other studies, showing that the targeted regions are suitable for inferring higher-order relationships among bryophytes.

2.4 Conclusions

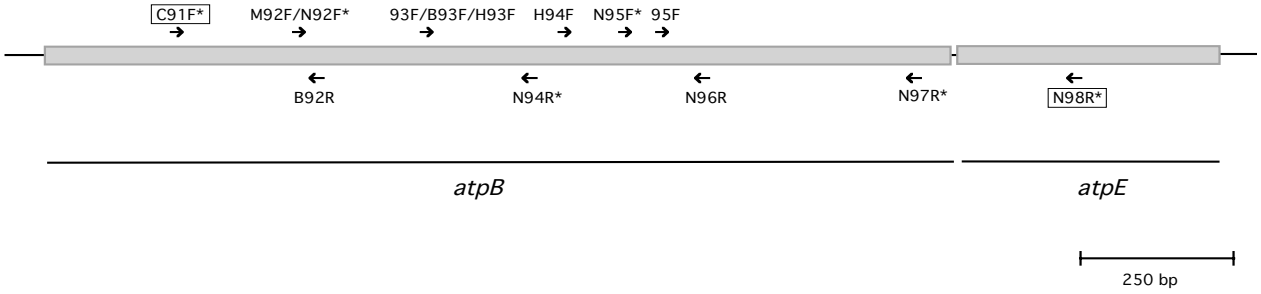
I find that the new primers designed in this study (Figs. 2.1-2.6) work well across a broad range of bryophytes, permitting straightforward amplification and sequencing of the targeted regions, except for the 3'–*rps*12-*rps*7-*ndh*B region (Fig. 2.7), which is more difficult to retrieve. These primers are good complements to those already described for vascular plants (e.g., Graham and Olmstead, 2000; Rai and Graham, 2010), and I expect that they will be suitable for retrieval of DNA sequence data from most other bryophytes for use in studies addressing bryophyte phylogeny.

Table 2.1. Details of the amplification and sequencing products for 14 core plastid genes from the 11 taxa used in primer testing (*psb*DC = *psb*D-*psb*C; *psb*BTNH = *psb*B-*psb*T-*psb*N-*psb*H; *psb*EFLJ = *psb*E-*psb*F-*psb*L-*psb*J). The average number of successful sequencing reactions is based on the number of sequences included in final contigs for 11 "test" taxa. Asterisk: data from a single species; n.a.: "not applicable" due to failed amplification.

Region	Length (kb)	[1]Amplification primer pairs for test taxa	Average # of successful/failed sequencing reactions		
			Mosses	Liverworts	Hornworts
*atp*B	1.4	C91F/N98R	6/1	5/1.5	8/0.5
*rbc*L	1.4	80F/N86R	6/1.5	6/1	7/1.5
*ndh*F	1.3	N30F/36R or N31F/N34R	6/1	n.a	6/2 *
*psb*DC	2.2	N40F or C40F/B47R, 44F/51R	8/1.2	10/0	13/1 *
*psb*BTNH	2.2	60F/66R, 65F or N65F/B71R	10.5/1.5	9/3	11/1
*psb*EFLJ	0.7	55F or B55F/58R	4/0.5	3/0 *	4/1
*rpl*2	1.2	N20F or F20F/B25R or N21F/N24R	5/1	8/3 *	n.a.

[1]Exhaustive list: the most commonly used amplification primers are noted in the figures.

Fig. 2.1. Primer map of *atp*B region, and sequences of primers used to amplify and sequence this region. All new primers were designed by Y. Chang, except for B92R (designed by H.S. Rai, University of British Columbia). Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for all bryophytes, 'M' for mosses, and 'H' for liverworts, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. The scale (bp) is relative to *Physcomitrella patens* (primers not to scale).



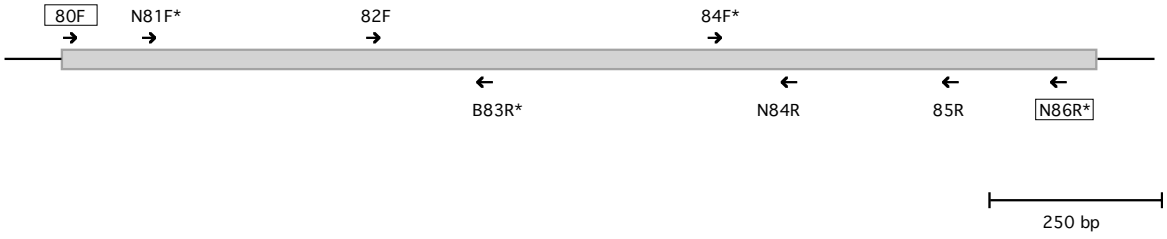
Forward primers

M92F: GGTGGTGTATCTGTATTTGGTGG
N92F: CAACTTTYCCKATTCATAGATCTGCTCC
H93F: AATGATCTTTACATGGAAATGAAAGAATC
H94F: GAAGTTTCTGCTTTATTAGGWAGAATGCC
N95F: CAATTCAAGCAGTTYAYGTACCTGC

Reverse primers

B92R: TCCACYACTTTAATTCCTGTTTC
N94R: AAAAGTACRTCTTGYTTATTAACATC
N96R: GTAGAAGTTGAATCTAAAGGATCYAC
N97R: TCTAATTCTCCASAAAGAATCATTTG
N98R: TACCATAGCAAAACCACCCATTA

Fig. 2.2. Primer map of *rbc*L region, and sequences of primers used to amplify and sequence this region. All new primers were designed by Y. Chang. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for bryophytes, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. The scale (bp) is relative to *Physcomitrella patens* (primers not to scale).



Forward primers

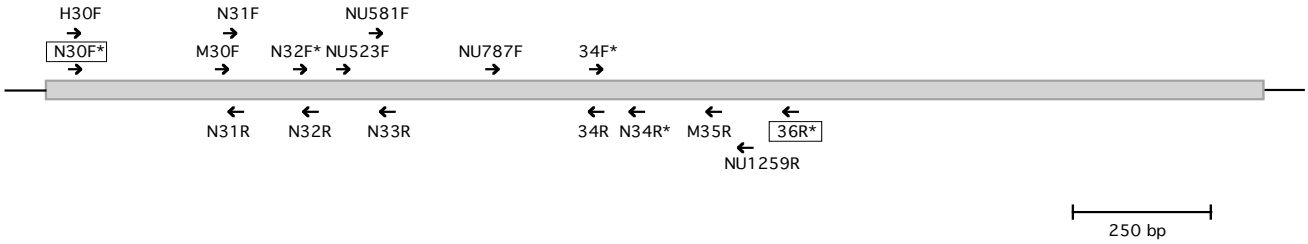
N81F: GCAGCATTTCGTATGACTCC

Reverse primers

N84R: CCTTCAAGTTTACCTACTACAGTAC

N86R: ATTTCTTTCCAWACTTCACAAGCAG

Fig. 2.3. Primer map of *ndh*F region, and sequences of primers used to amplify and sequence this region. Note that only the first ~1300 bp of *ndh*F is amplified. All new primers were designed by Y. Chang. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for bryophytes, 'M' for mosses, 'H' for liverworts and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. Scales (bp) are relative to *Physcomitrella patens* (primers not to scale).



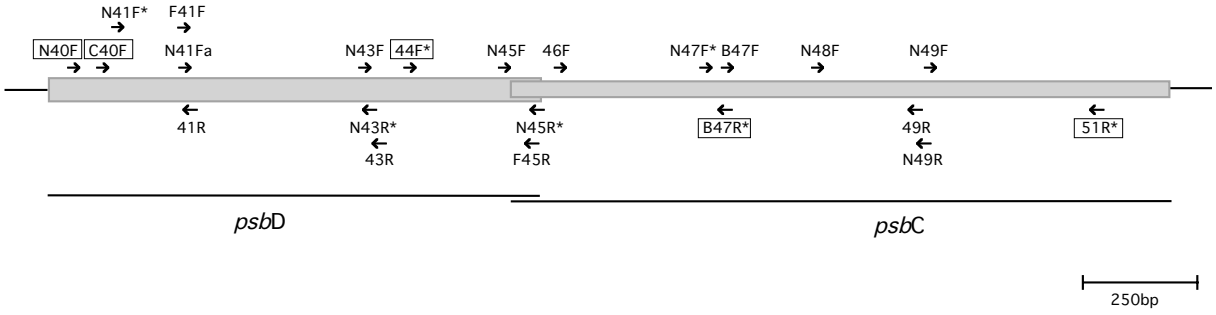
Forward primers

N30F: CCATTTATAGCTTCTATTTTAATAGG
H30F: ATGGAACTTATATTTCAAAATGTTTG
M30F: TAGGAGTTCTTGTTATGATTYATAG
N31F: GTTATGATTTATAGCGATAGTTAYATGTC
N32F: TTTTGGGAATTAGTAGGAATGTG

Reverse primers

N31R: GACATATAACTATCRCTATAAATCATAAC
N32R: TAAGAACACATTCCTACTAATTCCC
N33R: TCAAATTCAAAACTACCTGTYATCC
N34R: AGCTTTTGAATAAGCATGTG
M35R: AAGAAAAGTTRTTCCTGTAATTGG

Fig. 2.4. Primer map of *psb*D-*psb*C region, and sequences of primers used to amplify and sequence this region. This region is recovered using overlapping amplifications: *psbD-psb*C. All new primers were designed by Y. Chang. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for bryophytes, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. Scales (bp) are relative to *Physcomitrella patens* (primers not to scale).



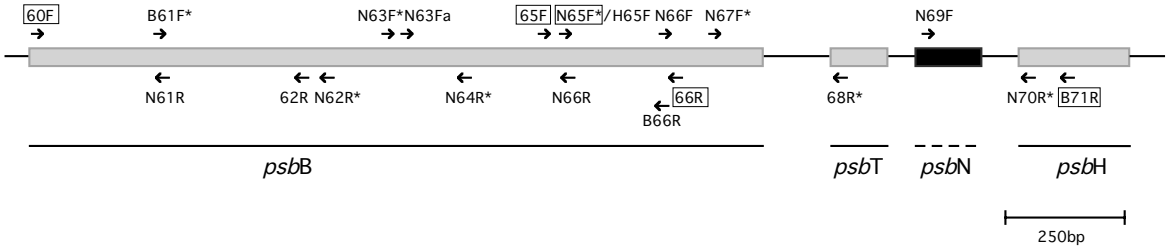
Forward primers

N40F: ATTTGATAGTATGGATGACTGGCTAAG
N41F: TAGGTGGRTGGTTTACAGGTAC
N41Fa: GAAGCACAAGGAGATTTTACTCGTTGGTG
N43F: GATGGTGATGGTGCAAAYACATTCCGTGC
N45F: GCTTGGATGGCAGCYCAAGATCAGCC
N47F: CCAGAAACTTTAGAAGAATCTTTCCC
N48F: TTGGTGGMGARGGATGGATTGTTAG
N49F: GTGAATTTTATGGTCCTACTGG

Reverse primers

N43R: CACGGAATGTATTTGCRCCATCACCATC
N45R: GGTTCCATTAAAGAGCGTTTCCAC
N49R: CCAGTAGGACCATAAAAYTCAC

Fig. 2.5. Primer map of *psb*B-*psb*T-*psb*N-*psb*H region, and sequences of primers used to amplify and sequence this region. All new primers were designed by Y. Chang. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for all bryophytes, 'H' for liverworts, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Note that *psb*N is on the complementary strand to *psbB*, *psb*T and *psb*H. Boxed primers are those used most frequently in amplification of the test taxa. Scales (bp) are relative to *Physcomitrella patens* (primers not to scale).



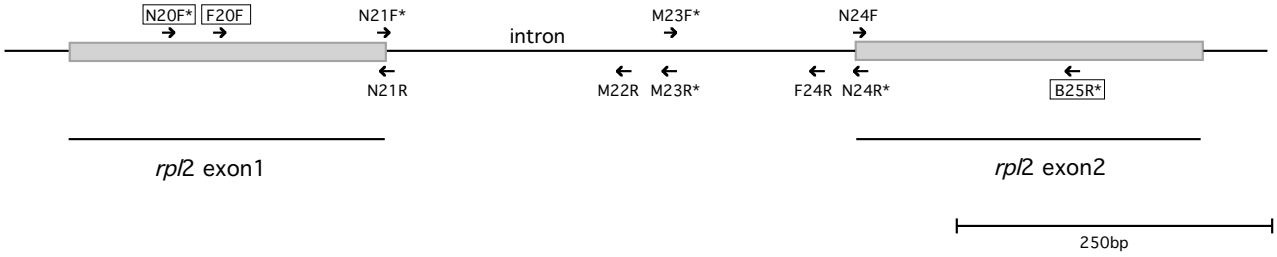
Forward primers

N63F: GCTTTTGTTGTTGCTGGAACTATGTG
N63Fa: GCAACTCCRGTAGAATTRTTTGG
N65F: AGTAGATGAAGAAGGAATTGTTAG
H65F: AGTAGATGAAGAAGGAATTGTTMGAGC
N66F: CAGATGGTGTTTTCCGTAGTAGTCC
N67F: GAGACGTTTTTGCTGGTATTGATCC
N69F: TGCTCTTCAAAAGGATCYCTAAG

Reverse primers

N61R: ACACCTTCATAACTCCAAAKACCTG
N62R: CCTGCTARTATACCTAAAATACCKGC
N64R: CWGGAATTTTAGACCAAGC
N66R: GCTCTAACAATTCCTTSTTCATCTAC
N70R: GGAGTATCATCAATAATTTGTGTAGC

Fig. 2.6. Primer map of *rpl*2 region, and sequences of primers used to amplify and sequence this region. All new primers were designed by Y. Chang. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for all bryophytes, 'M' for mosses, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. Scales (bp) are relative to *Physcomitrella patens* (primers not to scale).



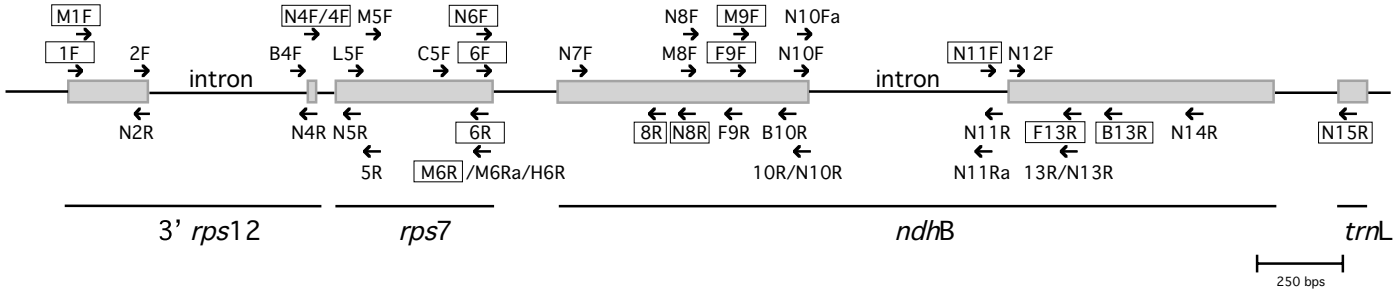
Forward primers

N20F: AAACAAGGTCGTAATAATCGAGG
N21F: GCTTTACCTTTGAGTGCGGTTTG
M23F: GCCTCCTAAGTTATTAGTAAAC
N24F: CTTCAACCAATATGCCTTTAGG

Reverse primers

N21R: AATAAATAATTCAAACCGCACTCAAAG
M22R: TGCCTTTGACCATCAATTYGAATG
M23R: TTGTTTACTAATAACTTAGGAGGC
N24R: CCTAAAGGCATATTGGTTGAAG

Fig. 2.7. Primer map for 3'-*rps*12-*rps7*-*ndh*B region, and sequences of primers used to amplify and sequence this region. This region is recovered using overlapping amplifications. Asterisks indicate the most commonly used sequencing primers for the first six regions. Primers starting with 'N' were designed to work for all bryophytes, 'M' for mosses, and 'H' for liverworts, and the remainder for vascular plants (e.g., Graham and Olmstead, 2000). Boxed primers are those used most frequently in amplification of the test taxa. Scales (bp) are relative to *Physcomitrella patens* (primers not to scale).



Forward primers

M1F: CGTAAAGTAGCAAGAGTAAGAYTAAC
N4F: AATCCACCCTACAATAYGGAG
M5F: AAAAATGGAAAAAAATCATTAGC
N6F: CTCATAGAATGGCTGAAGCTAATAGAGC
N7F: ACTATTTTGCCYGAATGTRTTCTTA
N8F: GGAGGAATGTTTTTATGYGGTGC
M8F: GCAACTATCGGAGGAATGTTYTTATG
M9F: GGAACAAGTTCATCTATTTTAGYTTATGG
N10F: AGTACCTTTTCATCAATGGACTCC
N10Fa: CAATGGACTCCYGATGTMTATGAAGG
N11F: ATTACTTAGGAGCTGTGTG
N12F: CCTACYCCAGTCGTTGCTTTTCTTTC

Reverse primers

N2R: TACAACGCACTAGAACGCCCTTGTTGAC
N4R: CTTTTTYACTCCATATTGTRGGG
N5R: CGATTACGATAAATYGGATC
M6R: GCTCTATTAGCTTCAGCCATTYTATGAG
M6Ra: TGTGCGAAAGCTCTATTGGCTTCTG
N8R: GCACCRCATAAAAACATTCCTC
N10R: CCTTCATAWACATCRGGAGTCCATTG
N11R: TATTTTCTCTCATTCAAAACYGTGC
N11Ra: AAACTTTYATTTCACACAGCTCC
N13R: CGTTTCATACTTGTYTGAGTAATAG
N14R: TAAATAATAATATATAGARATAACGCTTGT
N15R: CGTGTCTACCATTTCACCACCAAGGC

3. INFERRING THE HIGHER-ORDER PHYLOGENY OF MOSSES (BRYOPHYTA) AND RELATIVES USING A LARGE, MULTIGENE PLASTID DATA SET[1]

3.1 Brief synopsis

I surveyed 14–17 plastid genes from a broadly representative taxonomic sampling of the major bryophyte lineages, including all major lines of nonperistomate mosses. I inferred major clades with at least as strong support as other studies that used more taxa, and corroborated current views of overall embryophyte relationships—[liverworts, [mosses, [hornworts, tracheophytes]]]—with strong maximum likelihood (ML) bootstrap support. I also placed Zygnematales as the sister group of embryophytes with moderate ML bootstrap support. Within mosses, I confirmed Oedipodiaceae as the sister group of the large clade of peristomate taxa. Likelihood analysis firmly placed Takakiaceae as the sister group of all other mosses, a strong conflict with parsimony results, flagging a possible instance of long-branch attraction. Parsimony converged on the *Takakia*-sister result when rapidly evolving characters were removed, depending on the tree used to classify site rates. In general, however, my findings broadly support the utility of this 14-gene set from the plastome for future, more densely sampled phylogenetic studies of mosses and relatives.

---

[1]A version of this chapter has been published: CHANG, Y., AND S. W. GRAHAM. 2011. Inferring the higher-order phylogeny of mosses (Bryophyta) and relatives using a large, multigene plastid data set. *Am. J. Bot.* 98: 839-849.

3.2 Introduction

The mosses (Bryophyta) comprise a significant portion of plant diversity (~13,000 extant species, Shaw, 2009). They are the direct descendents of the earliest divisions in land-plant phylogeny, and are functionally important in many terrestrial ecosystems (e.g., Rydin, 2009). The precise placement of mosses in plant phylogeny—i.e., their position relative to the two other bryophyte clades (liverworts, hornworts) and to the vascular plants—has been controversial (reviewed in Qiu, 2008). In addition, despite substantial progress in understanding the overall pattern of moss phylogeny, considerable room for improvement clearly remains (Goffinet et al., 2009). Thus, new tools for inferring phylogenetic relationships in mosses and other bryophytes would be valuable for more confidently resolving their higher-order relationships, and this in turn would facilitate more precise reconstructions of their evolution, and improve the stability and functionality of current classification schemes.

Relationships among the multiple ancient and species-poor moss lineages that lack peristomes on their sporangia will be particularly critical to resolve, as they define the deepest splits in moss phylogeny. For example, while many molecular analyses recover a clade comprising the nonperistomate moss genera *Sphagnum* and *Takakia*, with this clade then as the sister group of all other mosses (e.g., Newton et al., 2000; Cox et al., 2004; Qiu et al., 2006, 2007; Shaw et al., 2010a, b), no morphological characters support this result (Newton et al., 2000; Frey and Stech, 2009), and there has been speculation that it represents a long-branch artefact (Shaw and Renzaglia, 2004). Subsequent splits in early moss phylogeny involve three distinctive nonperistomate genera: *Andreaea*, *Andreaeobryum* and *Oedipodium* (see Shaw and Renzaglia, 2004). *Oedipodium* may be the sister group of peristomate mosses (e.g., Cox et al.,

2004) but the precise relationship of this clade to *Andreaea* and *Andreaeobryum*, and of the latter genera to each other, are not clear (reviewed in Shaw and Renzaglia, 2004). Finally, major uncertainties in the backbone of moss phylogeny are by no means limited to the nonperistomate taxa. Within the peristomate clade, the precise relationships among the two nematodontous clades (Polytrichopsida and Tetraphidopsida) and the arthrodontous clade (Bryopsida), for example, have also not been resolved satisfactorily, as noted by Goffinet et al. (2009).

Loci from the nuclear and mitochondrial genomes have contributed to our understanding of plant phylogeny, but plastid (chloroplast) genes have generally played a more central role to date. There are several reasons for this (Olmstead and Palmer, 1994). For example, the coding portions of plastid genes are generally highly conserved, and as plastid genes are single-copy or effectively so (the two copies of genes present in the plastid inverted repeat region are identical), the orthology of plastid genes from distantly or closely related taxa is uncontroversial, an advantage over most nuclear genes (Kellogg and Bennetzen, 2004). The plastid genome also does not suffer from lateral gene transfer or the considerable rate heterogeneity sometimes observed in mitochondrial genes (e.g., Palmer et al., 2000; Bergthorsson et al., 2004; Qiu et al., 2010). In addition, the generally highly conserved gene content and gene order of plastid genomes makes it relatively straightforward to design primers to retrieve orthologous plastid genes from distantly related taxa (e.g., Graham and Olmstead, 2000a). Studies of higher-order moss phylogeny have focused on relatively few plastid genome regions (e.g., Table 3.1; see also Stech and Quandt, 2010, for a recent review). I therefore developed a series of new primers to amplify and sequence 17 plastid conserved genes and associated regions (Chapter 2) that have been used successfully in the inference of higher-order relationships at multiple levels of

angiosperm, seed-plant and vascular-plant phylogeny (Graham and Olmstead, 2000a; Rai et al., 2003, 2008; Graham et al., 2006; Saarela et al., 2007; Zgurski et al., 2008; Rai and Graham, 2010).

I focus here on using these plastid regions to infer the earliest splits in moss phylogeny and their position in embryophyte phylogeny, by using an exemplar-based approach that samples representatives of the major groups of mosses, liverworts and hornworts, including all lineages of nonperistomate mosses. I demonstrate that this sampling approach recovers relationships that are both well supported and broadly congruent with other studies, at least using maximum likelihood. Finally, I investigate whether identifying and removing the most rapidly evolving sites may help to ameliorate possible instances of mis-inference using parsimony. I restrict my inferences to maximum likelihood and parsimony analysis, as Bayesian inference can yield inflated or skewed branch support values in situations where ML or MP inference may be more conservative, such as the reconstruction of ancient phylogenetic splits (e.g., Suzuki et al., 2002; Cummings et al., 2003; Simmons et al., 2004; Kolaczkowski and Thornton, 2007; Susko, 2008).

3.3 Materials and methods

3.3.1 Taxonomic sampling strategy

I generated new data from 19 bryophyte taxa (five liverworts, two hornworts, 12 mosses) for phylogenetic analysis in combination with previously published sequences from one liverwort, one hornwort, two mosses, 15 vascular plants and five charophycean algae (Appendix 2). Carefully focused exemplar-based sampling allows the broad outline of phylogenetic relationships to be inferred, even where the species surveyed represent a small fraction of species

numbers, as is the case in all current studies that survey the broad backbone of bryophyte phylogeny (Table 3.1). I carefully chose exemplar species that split the broad backbone of bryophyte phylogeny at major internal nodes of interest. In general, my relatively sparse sampling permits exact comparisons to other studies that may have different sampling densities in descendant clades. For example, I sampled all nonperistomate moss lineages by including at least one exemplar species from the major genera, and included multiple representative species that straddle the root nodes of major clades (such as liverworts, hornworts and peristomate mosses), so far as these are known.

3.3.2 Sequence recovery and alignment

With a few minor exceptions, I recovered data from 14 plastid genes and associated non-coding regions from all taxa (Appendix 2). For most taxa I also generated sequence data for a region comprising three additional plastid genes (3'*rps*12, *rps*7 and *ndh*B) that proved significantly harder to retrieve (Chapter 2). This region was not recovered or attempted for a subset of taxa (i.e., *Buxbaumia aphylla*, *Bryum capillare*, *Treubia lacunose, Haplomitrium hookeri, Metzgeria conjugata, Pleurozia purpurea, Leiosporoceros dussii* and *Phaeoceros carolinianus*). Methods for DNA extraction, amplification, sequencing and base-calling follow Graham and Olmstead (2000a) and Chapter 2. The targeted markers were amplified and sequenced as eight distinct regions (four multi-gene clusters, *psb*D-*psb*C*, psb*E-*psb*F-*psb*L-*psb*J*,* *psb*B-*psb*T-*psb*N-*psb*H, and 3'-*rps*12-*rps*7-*ndh*B, and four single-gene regions, *ndh*F, *rpl*2, *rbc*L, and *atp*B, respectively) using primers and protocols described in Graham and Olmstead (2000a), Rai and Graham (2010) and Chapter 2. I recovered sequences from replicate amplification products generated from independent DNA extracts to detect pipetting errors or cross-

contamination. I detected none, although I observed several clear instances of polymorphisms from extracts involving multiple individuals from the same species, including *Andreaea rupestris* (for *psb*D), *Dicranum scoparium* (for *ndh*B, *ndh*F, *psb*F and *rbc*L) and *Orthotrichum lyellii* (for the *ndh*B, *rps*7, *rps*12 introns and the *rps*12-*rps*7 spacer region). I recorded these within-population polymorphisms using ambiguity codes in the consensus sequences. All regions were sequenced at least twice to generate contigs that are largely bidirectional.

I used the published sequences of *Physcomitrella patens* (Sugiura et al., 2003; Miyata and Sugita, 2004) to define gene and exon boundaries, and excluded noncoding regions from the vascular plants and algae to minimize alignment ambiguities. Sequences were aligned for each major region using Se-Al ver. 2.0 (Rambaut, 2002), according to criteria set out in Graham et al. (2000); see also Kelchner (2000) and Simmons and Ochoterena (2000). I offset hard-to-align noncoding regions that were frequently limited to single taxa in a staggered manner (Graham et al., 2006), although some of these regions include aligned blocks with multiple taxa, potentially providing phylogenetic information. The final alignment for all regions is 25,064 bp in length. This approach contributes significantly to the length of the alignment (Table 3.1), which is derived from ~10-13 kb of unaligned data for the taxa from which I attempted to sequence all 17-gene regions, and from ~8-10 kb of unaligned data where I attempted to sequence a subset of 14-gene regions (i.e., excluding 3'-*rps*12-*rps*7-*ndh*B region).

3.3.3 Phylogenetic analyses—Maximum likelihood and parsimony analyses

I performed maximum likelihood (ML) and parsimony (MP) analyses using Garli 1.0 (Zwickl, 2006) and PAUP* 4.0 b10 (Swofford, 2002) respectively, on concatenated matrices or subsets of them representing individual data partitions. The MP analyses used default settings

with TBR branch-swapping, but with 100 random addition replicates. I used the GTR + Γ + I model in ML analyses, with all model parameters estimated and two independent search replicates per analysis. I assessed branch support in ML and MP analyses using non-parametric bootstrapping (Felsenstein, 1985) with 300 bootstrap replicates, using simple taxon addition for MP, and otherwise using the search conditions described above. In the following discussion, I consider bootstrap support less than 70% as weak, between 70% and 90% as moderate, and higher than 90% as strong (e.g., Graham et al., 1998). I also considered several alternative tree arrangements (see Results) by constraining the monophyly of clades of interest that were not seen in the optimal trees, searching for the best ML tree satisfying a given constraint, and evaluating the significance of the difference in shortest unconstrained vs. constrained trees using the Approximately Unbiased (AU) test (Shimodaira, 2002) as implemented in CONSEL (Shimodaira and Hasegawa, 2001). All tree comparisons for a given data set were made simultaneously (see Goldman et al., 2000) and considered the full data set.

I ran individual ML analyses on eight subpartitions of the data comprising individual genes (*atp*B*, rbc*L*, ndh*F, *rpl*2) or local clusters of genes that included intergenic spacer regions (*psb*D-*psb*C*, psb*E-*psb*F-*psb*L-*psb*J*, psb*B-*psb*T-*psb*N-*psb*H*,* 3'*rps*12-*rps*7-*ndh*B). In all but one instance (*ndh*B in 3'*rps*12-*rps*7-*ndh*B) the multigene clusters comprise cotranscribed genes with related function (i.e., photosystem II or ribosomal protein genes). Three genes include introns (one each in *rpl*2, 3'*rps*12 and *ndh*B). I found two cases of possible strong conflict among regions, considering ML bootstrap support values. These concerned the relative arrangement of *Atrichum* and *Buxbaumia* for the *psb*D-*psb*C region (listed as clades p vs. p2 in Table 3.2), and an arrangement of *Chara* and *Zygnema* relative to land plants for the *psb*E-*psb*F-*psb*L-*psb*J

region (clades a and a2 in Table 3.2). These conflicts were not significant according to the AU test ($P > 0.05$), and so I concatenated all plastid regions into a combined data set using four different permutations: all 17 genes combined (with or without associated noncoding regions for bryophytes), and 14 genes combined, either with or without the associated noncoding regions for bryophytes. The 14- vs. 17-gene comparisons were included to examine whether the absence of the three-gene region that was difficult to obtain in some taxa (the *rps*12-*rps*7-*ndh*B region) had a substantial effect on bootstrap support; comparisons with and without noncoding regions were performed to assess the effect of including these regions. For MP analysis I considered only the case with all 17 genes combined, including noncoding regions. For this combination 6,280 aligned sites are potentially parsimony informative, and 1,856 sites are variable but parsimony uninformative (for comparison, for the 14-gene combination including noncoding data, 4,708 aligned sites are potentially parsimony informative, and 1,093 sites are variable but parsimony uninformative).

3.3.4 Effect of rapidly evolving sites on MP inference

I observed several moderate to strong conflicts between the ML and MP analysis of the concatenated data (see Results). I therefore investigated the effects of very rapidly evolving sites on MP-based inference by using HyPhy (Pond et al., 2005) to classify nucleotide sites (for the full 17-gene version, with coding and noncoding regions) into nine discrete ML rate classes, which I then set up as character sets. I repeated the site-rate assignments using the best ML and MP trees, respectively, to assess whether the tree considered for these rate classifications affected subsequent phylogenetic inference (Graham and Iles, 2009). The rate classes (RC) range from RC0 to RC8 (i.e., slowest to fastest; Burleigh and Mathews, 2004). I ran MP analysis on

three pooled subsets of rate classes: RC0-RC5, RC0-RC6, and RC78 (i.e., two slow subsets and one fast subset of the data).

3.4 Results

3.4.1 Phylogenetic inference—ML analysis

The results for vascular plants have been reported elsewhere (e.g., Graham and Olmstead, 2000; Rai et al., 2003), and so I focus here on phylogenetic results for bryophytes and algal outgroups. The best ML trees inferred from the various concatenated 14- and 17-gene matrices are identical to the topology shown in Fig. 3.1. Not surprisingly, the trees inferred from these combined analyses are better supported than in the analyses of individual data subpartitions (Table 3.2), which are nonetheless almost entirely congruent with each other (two exceptions are noted in the Materials and Methods). ML bootstrap support for most clades from the concatenated 14- and 17-gene data sets is generally highly comparable across analyses (Table 3.2), and almost all multi-family clades that involve bryophytes receive 80-100% ML bootstrap support from all four ML analyses (i.e., clades b-k, l, n-p, q, s, t, v, w; Fig. 1, Table 3.2); these clades represent almost all of the early backbone of bryophyte phylogeny. A clade supporting *Zygnema* (representing Zygnematales) as the sister group of land plants (clade a, Fig. 3.1) is moderately well supported by ML analyses based on the 17-gene data sets, with or without noncoding regions (Table 3.2). Only one clade is weakly supported in all ML analyses (clade u, representing a branch within Bryidae). Inclusion vs. exclusion of the hard-to-retrieve 3'*rps*12-*rps*7-*ndh*B region seems to have little effect at this taxon sampling. A minor exception is clade r, which is slightly weaker when this region is excluded. Excluding the noncoding regions also has

a relatively minor effect, with only two clades differing noticeably. One of these, clade m (Andreaeaceae-Andreaeobryaceae), is 16-21% less well supported with the noncoding regions included; the other, clade r (corresponding to the most-recent common ancestor of Dicranidae and Bryidae), is 30-34% better supported with the noncoding regions included.

3.4.2 Phylogenetic inference—MP analysis

Most bryophyte clades are identical in shortest trees inferred by the MP and ML analyses (Fig. 3.2), with three major exceptions. For MP, the first split in moss phylogeny is defined by a clade comprising *Sphagnum* and *Takakia*. This arrangement strongly contradicts the ML results (i.e., clades k2 vs. k, respectively, in Table 3.2). In addition, the liverworts are not monophyletic on the best MP tree, although their lack of monophyly is not well supported by MP bootstrap analysis (Fig. 3.2). A moderately well supported clade comprising *Zygnema* (Zygnematales) and *Chara* (Charales) contradicts the ML results (i.e., clades a2 and a, both have 72% bootstrap support, see Figs. 3.1, 3.2; Table 3.2). The MP results therefore weakly contradict the ML analysis concerning the sister group of land plants, as the clade comprising embryophytes, *Chara* and *Zygnema* has only 64% MP bootstrap support (Fig. 3.2). Seven multifamily clades have substantially reduced bootstrap support from MP compared to their generally strong support in ML bootstrap analysis (i.e., clades e, i, l-p), and only one clade is substantially stronger (clade u, which is still only moderately supported by MP bootstrap analysis, Table 3.2).

3.4.3 Exploration of conflicts between ML and MP analysis

The AU test rejects the MP arrangement for the deepest split in moss phylogeny (Fig. 2): when *Sphagnum* and *Takakia* are constrained as a clade, the resulting ML tree is significantly

longer than the best ML arrangement, in which *Takakia* is the sister group of all other mosses (Table 3.3). A third possible arrangement, with *Sphagnum* sister to all other mosses, is also rejected by the AU test (Table 3.3). An ML tree resulting from a search in which *Zygnema* and *Chara* are constrained as a clade (as in the best MP tree, Fig. 3.2) is not significantly longer than the moderately well supported arrangement found in the unconstrained ML tree in Fig. 1 (Table 3.3); note that several other possible outgroup arrangements assessed here cannot be rejected.

The effect of rate filtering on MP inference of the first split in moss phylogeny is sensitive to the tree used to make site-rate classifications. MP analysis of rate-filtered data recovered the ML arrangement, with *Takakia* sister to mosses for RC0-RC5, but only when the ML tree was used to generate rate classifications, and then with only moderate support (Table 3.4). In all other rate-filtered MP analyses, the *Sphagnum*-*Takakia* clade was recovered with moderate to strong MP bootstrap support. MP analysis recovers the ML arrangement for the sister-group of land plants when the most rapidly evolving sites are removed from consideration, with weak to strong bootstrap support. This depended in part on the precise set of characters considered (see RC0-RC5 and RC0-RC6 results concerning *Zygnema* and *Chara* in Table 3.4). For the RC0-RC5 results, this did not depend on whether the tree used for rate classifications was the ML or MP tree.

3.5 Discussion

3.5.1 A well supported backbone of bryophyte phylogeny

The ML relationships I recovered for the three major clades of bryophytes, with liverworts strongly supported as the sister group of other land plants, and hornworts as the sister group of vascular plants (Fig. 3.1, Table 3.5), agree with other recent studies (reviewed in Qiu, 2008). These relationships have been controversial until relatively recently, but although multiple lines of evidence now support them, they have not always been well supported by bootstrap analysis (e.g., a maximum of 87% ML or MP bootstrap support among studies summarized in Table 3.5, compared to 94-100% here for ML; clades c and d in Table 3.2). Other major relationships reported here within liverworts, hornworts and peristomate mosses are also consistent with recent studies that include broader samples of taxa for fewer genes, generally with at least as good ML bootstrap support here as reported elsewhere (Table 3.5).

My exemplar-based taxon-sampling typically permits me to make exact comparisons of the support for major clades to other recent studies (see Materials and Methods, and see footnote 2 in Table 3.5). Twenty of the 23 multi-family clades labeled in Fig. 3.1 have at least 90% ML bootstrap support (based on the 17-gene matrix that includes noncoding data in bryophytes; Tables 3.2, 3.5). The plastid genes employed here have also been used successfully in inferring the deep phylogeny of various vascular plant groups (e.g., Graham and Olmstead, 2000a; Graham et al., 2006; Saarela et al., 2007; Rai et al., 2008; Zgurski et al., 2008; Rai and Graham, 2010). A region comprising three of the 17 genes was difficult to retrieve for some bryophytes (Chapter 2). However, excluding these three genes (and associated noncoding regions) generally

had minimal effect on phylogenetic inference: the topology recovered using ML was the same for various 14- and 17-gene combinations in ML analyses, and support values were generally comparable (Table 3.2). This suggests that the 14-gene combination will generally provide sufficient information to infer major splits in bryophyte deep phylogeny with strong support. The gene set surveyed here is the largest attempted to date for this taxonomic breadth of bryophytes (Table 3.1) and our results indicate that they should be useful in future studies of higher-order relationships in the major clades of liverworts, hornworts and peristomate mosses.

3.5.2 Nonperistomate mosses and the earliest splits in moss phylogeny

Our ML analyses strongly support the monotypic *Oedipodium* as the sister group of peristomate mosses, congruent with the results of Cox et al. (2004), and consistent with the suggestions of Newton et al. (2000) and Goffinet et al. (2001). Support for a sister-group relationship between *Andreaea* and the monotypic *Andreaeobryum* is consistently found in our ML analysis (Table 3.2), and also in other studies that include both taxa (summarized in Cox et al., 2004; see also Table 3.5). This clade is moderately to strongly supported here in the 14- and 17-gene analyses that included coding regions only, but oddly, including noncoding regions reduced support by 16-21%. Volkmar and Knoop (2010) recently recovered this relationship with strong support, based on combined evidence from five plastid and mitochondrial genes. In the future I will sample *Andreaea* more densely for the current wholly plastid-based sampling. Sampling more densely within the clade comprising *Sphagnum* and its generic segregates would also be useful, and may help break up this branch substantially (Shaw et al., 2010a, b).

The most striking result here is the finding in ML analysis that *Takakia* is the sister group of all other mosses, with the next split being between *Sphagnum* and the remainder (Fig. 3.1). The position of *Takakia* in bryophyte phylogeny is considered problematic, and indeed before a sporophyte generation was known it was placed with liverworts based on its simple gametophyte morphology (reviewed in Kenrick and Crane, 1997; Cox et al., 2004). Published studies disagree on the position of *Takakia*, with some recovering the *Takakia*-sister arrangement (Yatsenyuk, 2001; Forrest et al., 2006; Volkmar and Knoop, 2010). Forrest et al. obtained strong support for this placement using Bayesian inference (see Table 3.5); Volkmar and Knoop recovered this from analysis of mitochondrial but not plastid data. Other researchers have obtained a contrasting arrangement with *Takakia* sister to *Sphagnum* (e.g., Cox et al., 2004; Qiu et al., 2007). I observed a strong conflict between my ML and MP analyses, the former strongly preferring *Takakia* as the sister group of all other mosses, the latter placing it strongly with *Sphagnum*, with these two then weakly supported as the sister group of other mosses (Fig. 3.2). The contrasting placements of *Takakia* may therefore be indicative of a long-branch attraction problem; ML analyses are known to be less prone than MP to systematic error (e.g., Chang, 1996; Huelsenbeck, 1997, 1998; Sullivan and Swofford, 2001; Swofford et al., 2001). However, the *Takakia*-*Sphagnum* arrangement has also been recovered with moderately strong support from ML bootstrap analysis in other published studies (e.g., Table 3.5).

I examined this question further by repeating the MP analysis after removing nucleotides that evolve rapidly according to an ML-based rate classification, since these characters may be more prone to saturation effects (e.g., Felsenstein, 1983). While I recovered a moderately well-supported *Takakia*-sister result using this data transformation, this result was sensitive to the

reference tree used for rate classifications, which suggests that this method of rate filtering may not be suitable for ameliorating long-branch problems in MP analysis of bryophyte phylogeny (see also Graham and Iles, 2009, who came to a similar conclusion for angiosperms). The strong conflict reported here concerning the first split in moss phylogeny underlines that this question warrants further careful study. However, the majority of the long terminal branches among nonperistomate mosses can not be broken up by denser taxon sampling, as I sampled all known extant species for three of five of these lineages (i.e., *Andreaeobryum*, *Oedipodium*, *Takakia*).

3.5.3 The sister-group of land plants

The closest algal relative to land plants also remains controversial. Charales, Coleochaetales, Zygnematales, or various combinations of these taxa are posited as the sister group of land plants (Becker and Marin, 2009). Recent studies based on DNA and protein-based analyses of whole plastid genomes (Turmel et al., 2006, 2007) recovered Zygnematales as the sister group of embryophytes, generally with strong ML support. In contrast, Finet et al. (2010) analyzed a large number of nuclear ribosomal protein genes and found strong ML support for *Coleochaete* (Coleochaetales) as the sister group of land plants, with *Chaetosphaeridium*, another member of Coleochaetales, nested in Zygnematales. However, an expanded nuclear gene analysis supports Zygnematales as the sister group of embryophytes (Wodniok et al., 2011). Although my sampling of outgroups to land plants is relatively limited, I included representatives of all major clades here, and my ML analysis moderately supports the Zygnematales-sister hypothesis (Fig. 3.1); note the moderate conflict with my MP results (cf. Figs. 3.1, 3.2). When I filtered out the highest rate sites based on the best ML or MP tree, considering the slowest sites (i.e., RC0-5), MP recovered the arrangement with Zygnematales as the sister group of land plants

with moderate to strong bootstrap support (96% and 79% respectively, depending on whether the best ML and MP tree was used for site rate assignment; Table 3.4). However, the AU test could not reject three of four alternative hypotheses for the sister group of land plants here using the full data set (Table 3.3; only Charales as sister is rejected at $P < 0.05$). A Zygnematales-sister arrangement has been questioned as a long-branch artefact by Qiu (2008). This may be true, and the existence of long-branch artefacts may be supported by the moderate ML vs. MP conflict observed here (clades a vs. a2 in Table 3.5). An expanded taxon sampling would certainly be useful in all current phylogenetic studies that examine the transition from charophytes to bryophytes.

3.5 Conclusions and future work

Phylogenetic congruence (the recovery of the same or similar tree topologies using different genes and/or taxon samplings) is a key criterion in the systematist's toolkit for assessing the accuracy of recovered relationships, as it supports the idea that the inferred branching relationships reflect evolutionary history rather than random error (e.g., Penny et al., 1982; Chase et al., 1993; Hillis, 1995; Graham et al., 1998). In contrast, topological discordance among studies or between likelihood and parsimony analysis may be indicative of systematic error, such as long-branch attraction (Felsenstein, 1978). I have shown that a concatenated subset of 14 plastid regions, that are straightforward to amplify and sequence (Chapter 2), collectively permit inference of the broad backbone of bryophyte phylogeny in the broader context of early land-plant evolution. These loci corroborate corresponding clades in other studies, generally with strong bootstrap support here, and in several cases provide the highest support values observed among these studies. I therefore conclude that these plastid regions will be valuable for

addressing further surveys of deep and recent phylogenetic relationships in bryophytes, and I do so in Chapter 4 for the most species-rich group of bryophytes, the peristomate mosses. Finally, it should be straightforward to use amplification products from my primers in multiplexed next-generation sequencing reads, as has been done for some whole plastome studies (e.g., Cronn et al., 2008). This could facilitate more rapid retrieval of this gene set from large numbers of taxa, including those represented by small or degraded samples (many of my current samples came from limited or herbarium samples). The regions explored here are straightforward to align and annotate, and my current study also suggests that the gene set examined here may be sufficient to recover many deep splits with confidence without recourse to recovering whole plastomes for all taxa. Our approach should therefore complement ongoing efforts to collect whole-plastid genome data sets for mosses and relatives.

Table 3.1. Selected recent multigene studies of higher-level phylogenetic relationships in various bryophyte lineages. ML = maximum likelihood; MP = parsimony; BI = Bayesian inference; NJ = Neighbor joining; nu = nuclear; pt = plastid; mt= mitochondrial.

Authors	Primary focus group (no. of exemplar species)	Other groups (no. of exemplar species)	No. of genes	Aligned nucleotides	Genes	Method(s) of analysis
This study	**Bryophytes (23)**	**Green algae (5)** **Vascular plants (15)**	**14-17**	**25,064** [a]	**All pt: The 14-gene set comprises *atp*B, *ndh*F, *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*J, *psb*L, *psb*N, *psb*H, *psb*T, *rbc*L, *rpl*2. The 17-gene set adds *ndh*B, *rps*7, 3'-*rps*12**	**ML, MP**
Cox et al. (2004)	Mosses (30)	Liverworts (4)	8	11,071	mt: *nad*5, *nad*7 nu: 18S rDNA, 26S rDNA pt: *trn*L-*trn*F, *psb*A, *rbc*L, *rps*4	ML, MP, BI
Forrest et al. (2006)	Liverworts (173)	Algae (4) Mosses (10) Hornworts (5) Vascular plants (2)	5	8,245	mt: *nad*5 nu: 26S rDNA pt: *psb*A, *rbc*L, *rps*4	MP, BI
Qiu et al. (2006)	Land plants (184)	Green algae (9)	6	13,631	mt: LSU nu: 18S rDNA pt: *atp*B, LSU, SSU, *rbc*L	ML, MP
Qiu et al. (2007)	Land plants (181)	Green algae (9)	7	14,553	mt: *atp*1, LSU nu: 18S rDNA pt: *atp*B, LSU, SSU, *rbc*L	ML, MP
Volkmar and Knoop (2010)	Mosses (33)	Liverwort (11)	4	5,453	mt: *cox*1 intron, *nad*5 pt: *rbc*L, *rps*4	ML, BI
Wahrmund et al. (2009)	Mosses (50)	--	5	7,549	mt: *nad*2, *nad*5 pt: *rbc*L, *rps*4	MP, BI, NJ
Wahrmund et al. (2010)	Mosses (56)	--	6	8,921	mt: *cob* intron, *nad*2, *nad*5 pt: *rbc*L, *rps*4	ML, BI

[a] Aligned length for the taxon subset considered in the present study, including offsets (staggered regions typically limited to unalignable portions for single taxa). The corresponding unaligned length ranges from ~8-13 kb per taxon in bryophytes, depending in part on whether we attempted to retrieve 14 or 17 genes (see Appendix 1 for list of genes retrieved by taxon). Noncoding regions were excluded for algae and vascular plants.

Table 3.2. Summary of bootstrap support for multi-family clades of bryophytes and relatives (excluding vascular plants) based on different data combinations and sub-partitions. Clade labels correspond to branches (taxon bipartitions) in Figs. 3.1 and 3.2, except branch p2 (which defines a clade comprising *Atrichum* and *Buxbaumia*). All listed clades have at least 70% support from at least one analysis; bootstrap values <50% are noted as "--". Some taxa have missing data for individual regions (see Results); the corresponding branch support is labeled as "na." Abbreviations: *psb*DC = *psb*D-*psb*C; *psb*BTNH = *psb*B-*psb*T-*psb*N-*psb*H; *psb*EFLJ = *psb*E-*psb*F-*psb*L-*psb*J; individual regions include coding and noncoding regions where relevant.

	Combined data sets					Individual regions							
	14 genes (coding only)	14 genes + noncoding regions	17 genes (coding only)	17 genes + noncoding regions	17 genes + noncoding regions	*atp*B	*rbc*L	*ndh*F	*psb*DC	*psb*EFLJ	*psb*BTNH	*rpl*2	*3'rps*12-*rps*7-*ndh*B
	ML	ML	ML	ML	MP	ML	ML	ML	ML	ML	ML	ML	ML
a	--	--	76	72	--	--	--	--	--	--	--	--	93
a2	--	--	--	--	72	--	--	--	--	72	--	--	--
b	100	100	100	100	100	78	82	100	95	--	96	--	100
c	99	99	100	100	93	74	76	--	--	--	--	--	75
d	96	94	97	97	90	61	67	--	63	--	78	--	70
e	100	100	100	100	57	--	--	na	75	57	96	55	na
f	100	98	96	94	89	57	63	na	78	na	98	--	na
g	100	100	100	100	96	55	60	99	95	96	90	--	100
h	100	100	100	100	97	99	99	98	90	62	100	--	na
i	91	93	86	93	75	--	52	na	77	82	58	94	na
j	88	93	100	100	85	--	--	96	--	--	--	--	100
k	81	89	97	99	--	--	--	82	--	--	--	na	94
k2	--	--	--	--	99	--	--	--	--	64	--	na	--
l	100	100	100	100	--	--	--	94	--	--	--	--	100
m	95	79	85	64	--	--	--	63	--	--	79	--	--
n	100	100	100	100	62	63	58	97	--	--	87	--	na
o	100	100	100	100	78	--	50	79	--	--	97	--	100
p	95	95	94	93	81	68	66	92	--	53	94	--	na
p2	--	--	--	--	--	--	--	--	98	--	--	--	--
q	100	100	100	100	100	91	93	100	83	58	100	61	100
r	53	87	65	95	100	--	--	--	55	90	--	58	77
s	100	100	100	100	100	57	60	--	95	94	75	85	100
t	100	100	100	100	100	--	--	100	87	81	82	74	na
u	--	--	--	54	80	--	60	--	--	51	72	69	100
v	100	100	100	100	100	100	100	100	100	95	100	na	na
w	100	100	100	100	99	92	90	93	64	na	100	na	na

Table 3.3. Approximately Unbiased (AU) tests of whether suboptimal arrangements concerning early moss phylogeny and the sister group of land plants are significantly worse than the optimal one. In the best likelihood tree, *Takakia* is the sister group of all other mosses, and *Zygnema* the sister group of embryophytes. ΔL = increase in –*ln*L compared to the best tree. The analyses consider all data combined (i.e., all 17 genes and conservative noncoding regions). The charophyte orders represented here are: Charales (*Chara*); Coleochaetales (*Chaetosphaeridium*); Zygnematales (*Zygnema*).

Arrangement of interest		ΔL	*p* value
First split in mosses	(*Takakia*, (all other mosses))	best ML tree	
	(*Sphagnum*, (all other mosses))	24.3441	0*
	((*Sphagnum*, *Takakia*), (all other mosses))	19.9754	0.0289*
Sister group of land plants	*Zygnema*	best ML tree	
	Chaetosphaeridium	12.9220	0.1529
	Chaetosphaeridium-Zygnema	15.151	0.1656
	Chara-Zygnema	14.0910	0.3137
	Chara	19.9806	0.030*

Notes: * $P < 0.05$: significantly worse.

Table 3.4. Bootstrap support for contrasting arrangements at the base of mosses, and concerning the sister group of land plants. The analyses considered all data combined (i.e., all 17 genes and associated noncoding regions) or subsets involving different rate classes (RC; see text) as determined using different reference trees. Bootstrap values with <50% support are indicated as '--'.

Bootstrap support determined by:	ML (MP) (unfiltered)	MP only (using rate-filtered data)					
Tree for rate classification [1]	n/a	(1) Best ML tree			(2) Best MP tree		
Rate classes [2]	n/a	RC0-RC5	RC0-RC6	RC78	RC0-RC5	RC0-RC6	RC78
Arrangement of interest:							
Takakia sister to other mosses	100 (--)	81	--	--	--	--	--
Takakia sister to *Sphagnum*	(--) 99	--	84	100	73	96	98
Zygnema sister to land plants	72 (--)	96	85	--	79	--	--
Zygnema sister to *Chara* [3]	-- (72)	--	--	78	--	51	52

[1] Site rate classifications determined using the best ML or MP tree, respectively.

[2] Subsets of sites used in MP inference, according to the rate classification. RC0-RC5 includes the slowest six classes. RC78 includes the two fastest rate classes.

[3] [*Zygnema* + *Chara*] then sister to land plants, except for RC78, where *Chaeotosphaeridium* (Coleochaetales) is sister to land plants.

Table 3.5. Support for multi-family clades of bryophytes and relatives, excluding vascular plants, across selected studies, as determined by likelihood bootstrap (ML), parsimony bootstrap (MP) or posterior probability (PP, expressed as a percentage here). Support values for each study are for the largest combinations of genes noted in Tables 3.1 and 3.2 (e.g., 17 genes and noncoding regions for the current study). Labels in first column correspond to branches in Figs. 3.1 and 3.2 (except a3). Author abbreviations: C04, Cox et al. 2004; F06, Forrest et al. (2006); Q07, Qiu et al. 2007; V10, Volkmar and Knoop (2010); W10, Wahrmund et al. (2010). Clade composition comparisons are relative to Fig. 1 in Qiu et al. (2007) for clades b-d, Fig. 1 in Forrest et al. (2006) for clades e-i, Fig. 3 in Cox et al. (2004) for clades j and k2-n, Fig. 3 in Wahrmund et al. (2010) for clades o-u, and Fig. 2 in Duff et al. (2007) for clades v, w. MRCA = most recent common ancestor.

	Taxon bipartition (clade on rooted tree)	This study ML (MP)	C04 ML (MP, PP[1])	F06 MP (PP[1])	Q07 ML	V10 ML (PP)	W10 ML (PP)
a	[Zygnematales + land plants]	72 (<50)	--	--	--	--	--
a2	[Zygnematales + Charales]	<50 (72)	--	--	--	--	--
a3	[Charales + land plants]	<50 (<50)	--	100 (--)	93	--	--
b	Land plants	100 (100)	--	99 (--)	100	--	--
c	Landplants excl. liverworts	100 (93)	--	59 (--)	87	--	--
d	[Hornworts + vascular plants]	97 (90)	--	61 (100–100)	87	--	--
e	Liverworts	100 (57)	--	100 (100–100)	100	100 (100)	--
f	Haplomitriopsida	94 (89)	--	99 (100–100)	83	100 (100) [2]	--
g	Liverworts excl. Haplomitriopsida	100 (96)	--	98 (100–100)	92	100 (100)	--
h	Jungermanniopsida	100 (97) [2]	95 (75, 100–100) [2]	100 (100–100)	100	100 (100)	--
i	Metzgeriidae	93 (75)	--	99 (100–100)	100 [2]	--	--
j	Mosses	100 (85)	100 (100, 100–100)	98 (100–100)	100	100 (100)	--
k	Mosses excl. Takakiaceae	99 (<50)	--	-- (58–97)	--	--	--
k2	[Takakiaceae + Sphagnaceae]	<50 (99)	86 (100, 99–100)	--	86	79 (--)	--
l	Mosses excl. Takakiaceae & Sphagnaceae	100 (<50)	67 (--, 99–100)	96 (100–100)	100	94 (100)	100 (100)
m	[Andreaeaceae + Andreaeobryaceae]	64 (<50)	-- (--, 86–90)	--	--	95 (100)	--
n	[Oedipodiaceae + peristomate mosses]	100 (62)	91 (--, 99–100)	--	--	--	--
o	Peristomate mosses	100 (78)	72 (--, 99–100)	94 (100–100)	100	98 (100)	100 (100)
p	Bryopsida	93 (81)	83 (--, 99–100)	(= t)	100	95 (100)	100 (100)
q	Bryopsida excl. Buxbaumiaceae	100 (100) [2]	100 (77, 100–100)	(= t)	100	100 (100)	100 (100)
r	MRCA of Dicranidae & Bryidae	95 (100)	78 (79, 100–100)	--	100	100 (100)	<50 (--)
s	Dicranidae	100 (100) [2]	100 (100, 100–100) [2]	--	100 [2]	91 (100) [2]	98 (100)
t	Bryidae	100 (100) [2]	100 (100, 100–100)	100 (100–100) [2]	100	100 (100)	100 (100)
u	MRCA of Orthotrichaceae & Brachytheciaceae	54 (80)	-- (--, 53–85)	--	<50	--	96 (100)
v	Hornworts	100 (100)	--	100 (100–100)	100	--	--
w	Anthocerotopsida	100 (99)	--	75 (91-100)	85	--	--

--: Precise support not noted or branch not applicable given the taxa sampled.
[1] Range of posterior possibilities noted when multiple Bayesian analyses were done.
[2] MRCA of exemplar taxa is less inclusive than in corresponding reference studies.

Fig. 3.1. Phylogram of best ML tree based on combined analysis of 17 genes and associated noncoding regions (-*ln*L = 181836.287). ML bootstrap values indicated as numbers beside branches; those with <50% support indicated as '--'. Letter labels indicating major clades of bryophytes and early land plants referred to in Tables 3.2 and 3.5 and the text.

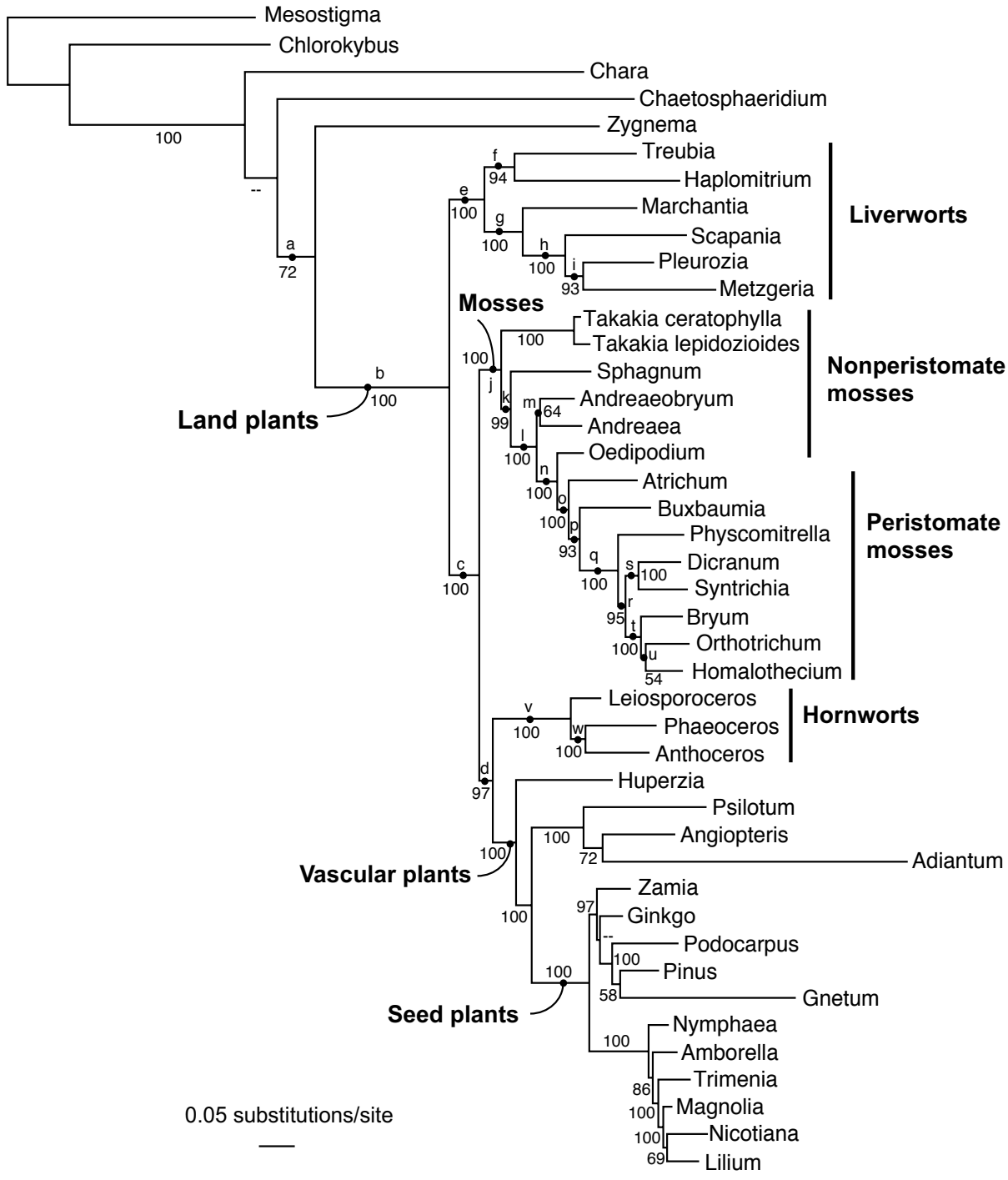
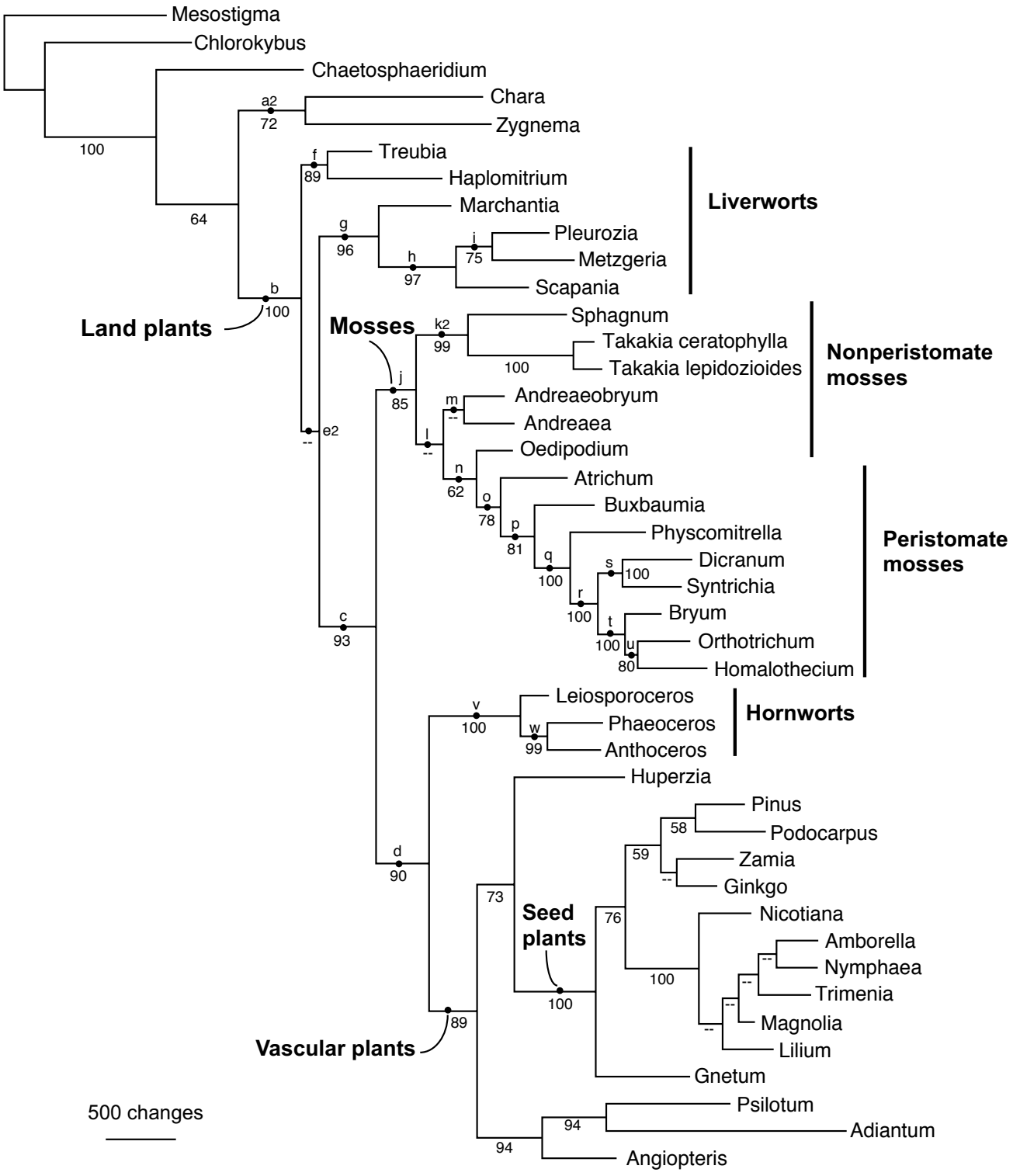

Mesostigma
Chlorokybus
Chara
Chaetosphaeridium
Zygnema
Treubia
Haplomitrium
Marchantia
Scapania
Pleurozia
Metzgeria
Liverworts
Mosses
Takakia ceratophylla
Takakia lepidozioides
Sphagnum
Andreaeobryum
Andreaea
Oedipodium
Nonperistomate mosses
Land plants
Atrichum
Buxbaumia
Physcomitrella
Dicranum
Syntrichia
Bryum
Orthotrichum
Homalothecium
Peristomate mosses
Leiosporoceros
Phaeoceros
Anthoceros
Hornworts
Huperzia
Psilotum
Angiopteris
Adiantum
Vascular plants
Zamia
Ginkgo
Podocarpus
Pinus
Gnetum
Seed plants
Nymphaea
Amborella
Trimenia
Magnolia
Nicotiana
Lilium
0.05 substitutions/site

Fig. 3.2. Phylogram of the single best MP tree based on combined analysis of 17 genes and associated noncoding regions (tree length = 36527 steps). MP bootstrap values indicated as numbers beside branches; those with <50% support indicated as '--'. Letter labels indicating major clades of bryophytes and early land plants referred to in Tables 3.2 and 3.5 and the text.


Mesostigma
Chlorokybus
Chaetosphaeridium
Chara
Zygnema
Treubia
Haplomitrium
Marchantia
Pleurozia
Metzgeria
Scapania
Liverworts
Land plants
Mosses
Sphagnum
Takakia ceratophylla
Takakia lepidozioides
Nonperistomate mosses
Andreaeobryum
Andreaea
Oedipodium
Atrichum
Buxbaumia
Physcomitrella
Dicranum
Syntrichia
Bryum
Orthotrichum
Homalothecium
Peristomate mosses
Leiosporoceros
Phaeoceros
Anthoceros
Hornworts
Huperzia
Pinus
Podocarpus
Zamia
Ginkgo
Seed plants
Nicotiana
Amborella
Nymphaea
Trimenia
Magnolia
Lilium
Gnetum
Vascular plants
Psilotum
Adiantum
Angiopteris
500 changes

4. BACKBONE PHYLOGENY OF PERISTOMATE AND NONPERISTOMATE MOSSES BASED ON MULTIPLE PLASTID MARKERS

4.1 Brief synopsis

To further investigate relationships among major groups of peristomate mosses, I retrieved 14-17 genes from a total of 35 peristomate species, and additional representatives from the nonperistomate lineages. The phylogenetic relationships recovered here generally were consistent with other studies on peristomate mosses, and I recovered improved support for multiple clades, including several that have not been satisfactorily resolved to date. For example, within Bryopsida, Timmiidae was found to be the sister group of a clade comprising Dicranidae and Bryidae, with Funariidae then the sister group of Bryidae-Dicranidae-Timmiidae, all with strong maximum likelihood (ML) bootstrap support. Higher-order relationships within Funariidae and Dicranidae were also generally well resolved and supported, while relationships within Bryidae were less well supported, possibly reflecting a rapidly early radiation in this major clade of mosses. Conflicting arrangements of Takakiaceae and Sphagnaceae at the base of moss phylogeny were obtained with ML inference, depending on the method of data partitioning and how base frequencies were parameterized in the DNA substitution model.

4.2 Introduction

The relatively simple morphology of mosses results in a paucity of characters suitable for the inference of higher-order phylogenetic relationships. While sporophyte characters like those

related to the morphology and development of peristome teeth are useful for defining major taxonomic groups, due to their presumed evolutionary stability, they are not numerous or ubiquitous enough to permit confident inference of all aspects of the broad backbone of moss phylogeny (Goffinet et al., 2001; Cox et al., 2004). Gametophytic characters, on the other hand, appear to suffer from frequent secondary loss and convergence, which may lead to incorrect inference of higher-order relationships (Goffinet et al., 2009). For the past two decades, studies based on molecular data have therefore provided important new evidence for improving our understanding of phylogenetic relationships among and within major moss groups (e.g., Goffinet et al., 2001, La Farge et al., 2002; also Fig. 4.1). While a considerable degree of consensus has been reached, substantial ambiguities still remain, including multiple relationships along the major backbones of moss phylogeny (Renzaglia et al., 2007; Goffinet et al., 2009).

The earliest splits in moss phylogeny are defined by nonperistomate mosses. In one plausible scenario, *Takakia* (Takakiopsida: Takakiaceae) is the sister group of the remaining mosses, with *Sphagnum* and relatives (Sphagnopsida) the next successive sister group (e.g., Forrest et al., 2006; Chapter 3); alternatively a clade comprising Takakiopsida and Sphagnopsida may be the sister group to all other mosses (e.g., Cox et al., 2004, Qiu et al., 2007; Chapter 3). A clade comprising Andreaeopsida and Andreaeobryopsida may then represent the next successive sister group (Frey and Stech, 2009; Chapter 3), with the monotypic Oedipodiopsida the sister group of all peristomate mosses (Cox et al., 2004; Wahrmund et al., 2009; Chapter 3). Within the peristomate moss clade, two lineages (Polytrichopsida and Tetraphidopsida) have solely nematodontous peristome teeth, and one has arthrodontous peristome teeth (Bryopsida). The relationships among these three lineages remain uncertain (e.g., Qiu et al., 2006, 2007; Volkmar

and Knoop, 2010), although evidence from peristome development may support a close relationship between Tetraphidopsida and Bryopsida (Shaw and Anderson, 1988). The arthrodontous taxa (Bryopsida), which comprise ~90% of all mosses (Shaw et al., 2003), include several lineages that are somewhat intermediate in peristome form and development between arthrodontous and nematodontous types, such as *Buxbaumia* (Buxbaumiidae). Most of the remaining "true" arthrodontous taxa can be further defined by their possession of diplolepideous-opposite, diplolepideous-alternate or haplolepideous peristomes. The diplolepideous-alternate mosses are classified in subclass Bryidae, and haplolepideous mosses are placed in Dicranidae. These two subclasses may be sister groups to each other (Newton et al., 2000; Cox et al., 2004; Qiu et al., 2006). The interpretation of the peristome in *Timmia* (Timmiidae) and *Encalypta* (Funariidae) remains controversial (Shaw and Renzaglia, 2004); the phylogenetic position of these taxa relative to the diplolepideous-opposite mosses (Funariales), and Bryidae-Dicranidae is also unclear (e.g., Volkmar and Knoop, 2010; Wahrmund et al., 2010).

The earliest splits in moss evolution likely happened more than 300 million years ago (Newton et al., 2009) during relatively rapid early radiations that left correspondingly short internodes between major groups (Goffinet et al., 2009). Many lineages that arose at this time (e.g., nonperisomate mosses) are also quite species poor. As a consequence, even densely sampled phylogenies have substantially long branches in early moss phylogeny that are connected by short internodes. This combination may be particularly challenging for phylogenetic inference (e.g., Felsenstein, 1978, Penny and Hendy, 1989). Here I sample a relatively large number of conservative DNA markers from the plastid genome to reconstruct the backbone relationships of mosses. Slowly evolving markers may be less susceptible to any long

branch problems (Chapters 2, 3, and see Graham and Olmstead 2000 for a more detailed discussion).

The plastid markers I sample here are *atp*B, *ndh*F, *rbc*L and *rpl*2, and four clusters of genes with associated intergenic spacers (i.e., *psb*D-*psb*C*, psb*E-*psb*F-*psb*L-*psb*J*, psb*B-*psb*T-*psb*N-*psb*H, and 3'-*rps*12-*rps*7-*ndh*B). I developed bryophyte-specific primers for these genes and demonstrated their utility in inferring higher-order relationships of mosses based on a preliminary taxonomic sampling (Chapters 2, 3). Here I use these genes to infer the deep phylogenetic relationships of mosses using a substantially expanded taxa sampling (especially of peristomate mosses). I use different phylogenetic criteria for tree inference: maximum parsimony (MP) and several implementations of maximum likelihood (ML). I also show that at least one major relationship in moss phylogeny is sensitive to the stategy used to estimate model parameters for likelihood analysis.

4.3 Materials and methods

4.3.1 Taxonomic and genomic sampling

I sampled 75 taxa in total from various land-plant groups and algal relatives (Appendix 3). Within mosses, I sampled at least one species from each order, except for Ambuchananiales and Gigaspermales (the classification used here is that of Goffinet et al., 2009). A total of 43 exemplar mosses is included, among which 29 taxa are newly sequenced for this study. In addition, I included six liverworts, three hornworts and 15 vascular plants to represent the other major land-plant clades, and six charophycean and two chlorophycean algae as outgroups (e.g., Turmel et al., 2007).

Among all the targeted markers, the 3'*rps*12-*rps*7-*ndh*B region proved the most difficult to retrieve (Chapter 2). I therefore focussed on a subset of 14 genes and associated noncoding regions from all the taxa (Chapters 2, 3). However, I was able to add 3'*rps*12-*rps*7-*ndh*B for several taxa that may occupy potentially critical positions in the phylogeny (i.e., *Sphagnum squarrosum, Alophosia, Timmia, Funaria, Bartramia, Rhytidiadelphus* and *Hookeria*) based on results from previous studies (e.g., Newton et al., 2000; Cox et al., 2004; Qiu et al., 2007; Wahrmund et al., 2009 and 2010), or preliminary analyses based on the 14-gene matrix in this study.

4.3.2 Sequence recovery and alignment

Methods for DNA extraction, amplification, sequencing and base-calling follow Graham and Olmstead (2000) and Chapters 2, 3. I amplified and sequenced the targeted markers in eight distinct regions (comprising respectively four multi-gene clusters, *psb*D-*psb*C*, psb*E-*psb*F-*psb*L-*psb*J*, psb*B-*psb*T-*psb*N-*psb*H, and 3'-*rps*12-*rps*7-*ndh*B, and four single-gene regions, *ndh*F, *rpl*2, *rbc*L and *atp*B) using primers and protocols described in Graham and Olmstead (2000), Rai and Graham (2010) and Chapter 2. The newly obtained data were aligned to a previously generated alignment (Chapter 3) with Se-Al (Rambaut, 2002), following alignment criteria outlined in Graham et al. (2000). The alignments include several noncoding regions and introns that I aligned across mosses. I excluded these regions from other green plants (Chapter 3). The final alignment is 29,266 bp in length, derived from ~10-13 kb of unaligned data for the taxa from which I attempted to sequence all 17-gene regions, and from ~8-10 kb of unaligned data where I attempted to sequence a subset of 14-gene regions (i.e., excluding 3'-*rps*12-*rps*7-*ndh*B region). Across all taxa included in this study, 7,014 aligned sites are potentially parsimony informative,

and 1,824 are variable but parsimony uninformative; across mosses, 4,228 aligned sites are potentially parsimony informative, and 1,855 variable but parsimony uninformative.

4.3.3 Phylogenetic analysis

I first conducted phylogenetic analysis based on the eight regions described above, using maximum likelihood (ML). I searched for the best ML tree in each case using Garli 1.0 (Zwickl, 2006), with the GTR + Γ + I DNA substitution model, using ML estimates of the model parameters (i.e., with these optimized as part of tree inference). I assessed branch support using non-parametric bootstrapping (Felsenstein, 1985) with 500 bootstrap replicates for each region, with all model parameters estimated.

I compared ML bootstrap support values from individual regions and found no strong conflicts (i.e., no conflicting results with bootstrap values both higher than 70%; data not shown). In the previous chapter (Chapter 3), there were two moderate to strong conflicts regarding the relationship between *Atrichum* (Polytrichopsida, Polytrichaceae) and *Buxbaumia* (Bryopsida, Buxbaumiaceae), and the relative arrangement of *Zygnema*, other charophytes and the land plants. I no longer saw the conflicts among regions reported in Chapter 2 with the expanded taxon sampling used here. I concatenated all of the regions for a combined ML analysis that considered two alternative data partitionings for the analysis [i.e., either by individual region (the eight regions described above, including noncoding regions with relevant regions, where present), or by "codon position" (i.e., first, second and third, with noncoding regions included as a fourth partition)]. I also performed a combined ML analysis with no subpartitions, and analyzed the combined matrix using maximum parsimony (MP).

I used PAUP* 4.0 b10 (Swofford, 2002) for MP heuristic searches, using TBR branch-swapping and 100 random addition replicates to search for the shortest MP trees. For ML heuristic searches, I used Garli 1.0 (for nonpartitioned analyses), Garli-Part-0.97 (for partitioned analyses) and RAxML-7.2.6 (Stamatakis, 2006; Stamatakis et al., 2008). Ten independent replicates were carried out for each ML search. The RAxML analyses were executed using the CIPRES portals (Miller et al., 2009). In preliminary analyses, I noticed an apparent conflict among the earliest branches in moss phylogeny in analyses using these two programs (see Results). As base frequencies can be estimated in Garli using ML optimization or empirically from terminal taxa only, and RAxML allows only the latter, I explored the possible impact of these two estimation methods on phylogenetic inference by re-running Garli using empirical etimates of base frequencies.

In summary, for each matrix and data partitioning (unpartitioned, partitioned by codons or regions), I ran ML analyses once using RAxML (with empirical base frequencies) and twice using Garli (once each with ML-estimated vs. empirical base frequencies) (Table 4.1). I used bootstrap analysis with 500 pseudo-replicates for MP and the various ML analyses to assess branch support. For the bootstrap analyses, I used simple taxon addition for MP analysis, but otherwise used the search conditions described above.

I also considered three alternative tree arrangements concerning the earliest split in moss phylogeny: the optimal arrangement in the unpartitioned ML analysis, and two suboptimal arrangements (see Results). In the latter two cases I used Garli to perform heuristic searches with topological constraints that reflect each suboptimal arrangement (considering ML estimates of base frequencies and with no data partitions). I then used CONSEL (Shimodaira and Hasegawa,

2001) to assess whether suboptimal trees satisfying the constraints were significantly less optimal then the optimal tree, using the Approximately Unbiased (AU) test (Shimodaira, 2002). To do this test I determined individual site likelihoods using PAUP*, with base frequencies either estimated using ML or determined empirically, and again with no data partitions.

4.4 Results

4.4.1 Phylogenetic relationships inferred by different analyses

I discussed relationships among the four major land-plant lineages in Chapter 3. My results here for these lineages are consistent with what I reported there, and so here I focus on presenting multi-family relationships within the mosses. I first discuss the results from the Garli analyses that used ML estimates of base frequencies (Figs. 4.1, 4.2, 4.3; column 3-5 in Table 4.1), followed by a comparison of results from the other ML analyses (i.e., analyses using empirical estimates of base frequency in Garli and RaxML; columns 4-9 in Table 4.1), and the MP analysis (column 10 in Table 4.1). In the following discussion, I consider bootstrap support less than 70% as weak, between 70% and 90% as moderate, and higher than 90% as strong (e.g., Graham et al., 1998; Chapter 3). For the interpretation of the results below, note that the likelihood scores estimated by different programs should not be compared directly, due to the different methods of ML optimization used by different programs (Sundberg et al., 2008),

4.4.1.1 ML analyses using ML-estimated base frequencies

The results from the three Garli analyses using ML estimates of base frequencies are generally highly consistent with each other (see the first three Garli analyses in Table 4.1; one unpartitioned ML analysis; the two partitioned ML analyses). The best ML trees have essentially

the same topology (one is shown in Figs. 4.1-4.3), with only minor differences regarding the relative placements of the "early-diverging" Bryidae taxa such as *Tetraplodon* (Splachnales), *Bartramia* (Bartramiales) and *Hedwigia* (Hedwigiales). For comparison, among the 41 clades of interest recovered in the optimal tree based on the unpartitioned ML analysis, 33 clades receive moderate or strong in all three analyses; five clades are weakly supported in all analyses; three clades are moderately supported by analysis based on the matrix partitioned by codon-position but are weakly supported by the other two analyses (Table 4.1; Figs. 4.1, 4.2).

The monophyly of land plants as a whole, and of each of the four land-plant lineages (liverworts, mosses, hornworts, vascular plants), and the relationships among these lineages (Table 4.1a; clades b-h), are among the branches that are strongly supported in all three ML analyses. The monophyly of all bryophyte classes and subclasses, and the majority of relationships among these major clades (e.g., clades j, l, m, o, p, q, r, s, t, u, and cc) are also recovered with moderate or strong support from all three analyses. Other branching patterns receiving moderate or strong support in all three analyses include the relationships among some moss orders and families (Table 4.1b; Fig. 4.2).

The relationships recovered for nonperistomate mosses are consistent with those seen in Chapter 3, again with mostly strong support from all three analyses; noteably *Takakia* is recovered as the sister group of all other mosses. Within the perisomate mosses, the relationships among Tetraphidopsida and Bryopsida (clade n), the relationships among Grimmiales, Pottiales, and Dicranales s.s. (clade z), and the sister relationship between Hypopterygiaceae and Hypnales (clade mm) receive moderate support from the ML analysis with partitioning by codon position, but only weak support from the other two ML analyses. The clades with weak support from all

analyses mainly concern relationships within Bryidae, especially the placements of "early diverging" taxa (Splachnales, Hedwigiales, Bartramiales vs. the remaining Bridae; clades dd and ee), and the local placement of Orthotrichales (clade gg) and Ptychomniales (clade jj). The closest algal relative of land plants (clade a) also remains uncertain (Table 4.1) but is recovered as Zygnematales in the best ML trees in all three analyses.

4.4.1.2 ML analyses using empirical estimates of base frequencies

The results from the other ML analyses (i.e., ML analyses using Garli and RaxML in which base frequencies are estimated empirically rather than using ML) are generally highly consistent with the analyses in which all parameters are ML estimates (Table 4.1), except for the relative placements of *Takakia* (Takakiopsida) and *Sphagnum* (Sphagnopsida). In the unpartitioned analysis and the analysis partitioned by the eight major regions, *Takakia* is recovered as the sister group of *Sphagnum* (the *Takakia*-*Sphagnum* arrangement) using either Garli or RAxML (with 69-97% support, Table 4.1, see also Table 4.2). However, in the analysis partitioned by codon position, the *Takakia*-sister arrangement is recovered in the optimal tree, with 100% bootstrap support in both Garli and RAxML analyses. In all ML analyses, the optimal trees inferred when the data are partioned by codon position have better likelihood scores than the optimal trees inferred from unpartitioned data, or when data partitioning is by major region (Table 4.1).

4.4.1.3 MP analysis

Overall the results from the MP analysis are consistent with the results from ML analyses using ML estimates of base frequencies, except for the placements of *Takakia* and *Sphagnum*

(Table 4.1). As in Chapter 3, the MP analysis supports the *Takakia*-*Sphagnum* arrangement (here clade i2) with strong support (100%). For the other clades of interest, the MP bootstrap support is generally comparable to but lower than that observed in the ML analyses. Only 21 of the 33 clades with moderate or strong ML support also receive moderate or strong bootstrap support from MP analysis. The remaining 12 clades, most of which concern branching order within Dicranidae and Bryidae (e.g., clades ff and kk), in addition to the monophyly of liverworts (clade f) and the earliest splits mosses (e.g., clades j and k), receive only weak support in MP analyses. For the clades with weak ML support, the MP support is also generally weak, except for clade ee2 (Bartramiales-Hedwigiales; Table 4.1), which is strongly supported by MP analysis, but receives only weak to moderate support from the various ML analysis.

4.4.2 Exploration of the conflicting placements of *Takakia* and *Sphagnum*

The optimal trees found in ML analyses (with ML estimates of base frequencies) depict *Takakia* as the sister group of other mosses. I examined whether two alternative arrangements of *Takakia* and *Sphagnum* were significantly less optimal than the best tree by comparing the optimal arrangement to trees satisfying these alternative arrangements (i.e., either with *Takakia* sister to *Sphagnum*; or *Sphagnum* as the sister group of other mosses). The test used site likelihoods estimated with no data partitions for the whole matrix, but used two different ways of estimating base frequencies (ML vs. empirical estimates).

For the AU tests in which I used ML estimates of base frequencies the two suboptimal arrangements were strongly rejected (for *Takakia*-*Sphagnum*, and *Sphagnum*-sister, $\Delta ln\text{L} = 28.476$ and 31.621, respectively; $P < 0.001$; Table 4.3). When I used empirical estimates of base frequencies, the *Takakia*-*Sphagnum* arrangement was the best scoring of the three considered,

but the AU test could not reject the *Takakia*-sister arrangement (ΔlnL = 6.802; P = 0.998; Table 4.3). However, the *Sphagnum*-sister arrangement was still rejected (ΔlnL = 26.341; $P < 0.001$; Table 4.3).

The observed base frequencies differ substantially among codon positions. For example, the GC content of the first and second codon positions is more than twice of that of the third codon position and the noncoding regions (Table 4.4). Further comparison of the values of other nucleotide substitution model parameters reveals substantial differences among different codon positions (Table 4.4). I ran ML analyses using Garli with only the first and second codon positions included, or with only the third codon position and noncoding sites included, with either ML or empirical estimates of base frequencies. I recovered the *Takakia*-sister arrangement as the optimal tree in all cases, with moderate to strong bootstrap support (Table 4.2).

4.5 Discussion

4.5.1 Length variation of the noncoding regions in mosses

The noncoding plastid regions examined here for mosses are not obviously more conserved in length than in the seed plants (cf. Table 2 in Rai et al. 2003 and Table 4.5 here), a crown clade of comparable age. A few studies have reported that the "early diverging" mosses have longer intergenic spacers (IGS) or introns in various mitochondrial and plastid regions (e.g., Quandt and Stech, 2004; Wahrmund et al., 2009 and 2010). This has been considered indicative of the presence of larger plastid and mitochondrial genomes in ancestral mosses and of a subsequent streamlining in "derived" (peristomate) lineages (Volkmar and Knoop, 2010). My data, however, suggest that noncoding regions are not evolving consistently in terms of their

length across the major moss lineages and plastid regions examined (Table 4.5). Comparing mean lengths for the intergenic spacer regions and introns examined here, four of six spacer regions (between *psb* genes) are shorter in the nonperistomate mosses than the peristomate mosses, whereas five other noncoding regions are longer, and there is substantial variation in the lengths of most of the noncoding regions (see ranges and standard deviations in Table 4.5). In general there is little generality to patterns of change in the lengths of these regions between nonperistomate and peristomate mosses (the *ndh*B intron is a possible exception; it is substantially shorter in the peristomate mosses considered here).

4.5.2 Towards a robust backbone phylogeny of peristomate mosses

4.5.2.1 Nematodontous mosses

One of the major unresolved questions in moss evolution concerns the placement of two nematodontous lineages, the monofamilial classes Polytrichopsida and Tetraphidopsida. Both lineages are characterized by peristome teeth made from whole cells. However, the structure and development of the teeth are different between the two classes (Shaw and Renzaglia, 2004). In addition, Polytrichopsida and Tetraphidopsida differ substantially in many gametophytic features so it has been proposed that they are not closely related (e.g., Crum and Anderson, 1981; Shaw and Anderson, 1988). In contrast, the early stages of the development of peristome teeth in Tetraphidopsida is "virtually indistinguishable" from the patterns found in arthrodontous mosses (Bryopsida), which may be indicative of a close affinity between them (Shaw and Anderson, 1988). To date, however, a sister-group relationship between Tetraphidopsida and Bryopsida has not received strong support from any phylogenetic study (e.g., only weak support in Cox et al.,

2004; Fig. 4.3), and although I recovered this relationship in most of the ML analyses here (data not shown), it does not have strong support in any analysis (clade n; Table 4.1, see also Fig 4.2). This is despite a broadly representative taxon sampling that includes exemplar species from the deepest splits of these major clades: I sampled species representing the core polytrichids and the two earliest splits in Polytrichopsida (Bell and Hyvonen 2008) and included both genera of Tetraphidopsida. A confident resolution of this issue may therefore require whole plastid genome data or genomic structural characters.

4.5.2.2 Diplolepideous-opposite taxa and relatives

While taxa with typically diplolepideous-opposite teeth are included in Funariidae, species in Encalyptales (Funariidae) have diverse peristome teeth morphologies whose interpretation has been controversial, as has been the order's phylogenetic placement (Vitt, 1984; Newton et al., 2000; Magombo, 2003). Here I recovered a well-supported clade comprising Encalypatales (represented by *Encalypta*) and Funariales (represented by *Funaria* and *Physcomitrella*). This result is consistent with recent studies based on multi-gene analyses (e.g., Qiu et al., 2007; Wahrmund et al., 2010; Fig. 4.3). The recent discovery of a 71-kb inversion in the large single copy region of plastid genome in Encalyptales and Funariales (excluding Gigaspermaceae), which is absent in the genomes from other mosses (Goffinet, 2007), also supports this hypothesis.

The interpretation of the peristome teeth of *Timmia* (the sole genus of Timmiidae) is also controversial (e.g., Vitt, 1984; Budke et al., 2007). *Timmia* possesses two rings of arthrodontous teeth, with the exostome being diplolepideous. Unlike most of the other diplolepideous taxa, the endostome of Timmiidae has no regular segment (instead it has 64 cilia). *Timmia* has been

classified with the diplolepideous-alternate taxa (Bryidae), because the latter sometimes also possess cilia (e.g., Vitt, 1984). However, recent phylogenetic studies indicate that *Timmia* may be placed outside Bryidae, although generally without strong support (e.g., Newton et al., 2000; Cox et al., 2004; Volkmar and Knoop, 2010). Here I find strong support for a sister-group relationship between *Timmia* and a clade comprising Bryidae and Dicranidae, with Funariidae well supported as the sister-group of Timmiidae-Bryidae-Dicranidae (clades q, s and t in Fig. 4.2). These clades are all consistently well-supported in the different ML analyses (Table 4.1). This key relationship between Timmiidae and Bryidae-Dicranidae has not been recovered with strong support before (Fig. 4.3).

A recent study on the peristome development of *Timmia megapolitana* shows symmetric division of its eight inner peristomial layer (IPL) cells (Budke et al., 2007), similar to Funariales. Asymmetric cell division in the IPL is found in other "early diverging" peristomate taxa (e.g., *Diphyscium*, *Tetraphis*), and in the haplolepideous (Dicranidae) and diplolepideous-alternate taxa (Bryidae) (Budke et al., 2007). The placement of *Timmia* here therefore implies that symmetric division of IPL cells was gained independently in Funariales and Timmiales, or that it was gained before the divergence of Funariales, and then lost before the origin of the diplolepidous-alternate and haplolepideous mosses.

4.5.2.3 Diplolepideous-alternate and haplolepideous taxa (Bryidae and Dicranidae)

The existence of a common uniform cell division pattern in peristome teeth development of the haplolepideous mosses (subclass Dicranidae) supports the idea that this taxon represents a natural group (Edwards, 1979; Shaw, 1989). I found strong support for the monophyly of Dicranidae here. This is consistent with previous studies (e.g., La Farge et al., 2000; Tsubota et

al., 2003; Hedderson, 2004; Wahrmund et al., 2009; and see clade u in Fig. 4.3). The earliest phylogenetic splits within Dicranidae are represented here by *Scouleria* (Scouleriales) and *Bryoxiphium* (Bryoxiphiales). The relationships recovered here for these taxa (clades v and w; Fig. 4.1) are well suppported across most analyses (Table 4.1) and have been found with good support in only a few other studies (Fig. 4.3). The monophyly of the species-rich order Dicranales is strongly rejected here (Fig. 4.2), as *Leucobryum* (Dicranales) is strongly and consistently supported as the sister-group of *Archidium* (Archidales) (clade x; Table 4.1, Fig. 4.2) and two other members of Dicranales (*Dicranum* and *Schistostega*) form a grade subtending *Syntrichia* (Pottiales) with moderate to strong support from the different ML analyses (clade bb; Table 4.1, Fig. 4.2). Similar relationships have also been recovered in other studies (e.g., Qiu et al., 2006 and 2007; Wahrmund et al., 2009 and 2010, see also Fig. 4.3), supporting the need for a revised classification of Dicranales as a whole (La Farge et al., 2000; Hedderson et al., 2004).

The monophyly of Bryidae (the diplolepideous-alternate group) has been well established by various studies (e.g., Goffinet et al., 2001; Qiu et al., 2006; see also Fig. 4.3). Splachnales (*Tetraplodon*), Bartramiales (*Bartramia*), and Hedwigiales (*Hedwigia*) represent the first splits in Bryidae, although the exact branching order among them is poorly supported (Fig. 4.2, Table 4.1) as it is in other studies (Fig. 4.3). The branching order of the remaining deep splits in Bryidae finds improved support here in some cases (e.g., clades ff, hh, ii; Fig. 4.3), and the core pleurocarpous mosses (as defined in Bell et al., 2007; clade kk here) are moderately well supported as monophyletic. Substantial ambiguities still remain regarding the relationships of the Bryidae taxa that form the grade subtending the core pleurocarps, as various branching orders have been recovered in previous studies, generally with poor support (e.g., Goffinet et al., 2001;

Bell et al., 2007; O'Brien, 2007). Rhizogoniales and the core pleurocarps have been referred to collectively as "pleurocarpids" (Bell at el., 2007), but the monophyly of Rhizogoniales (and therefore pleurocarpids) is weakly rejected here and elsewhere, as *Rhizogonium* (Rhizogoniales) is weakly supported as the sister group of *Orthotrichium* (Orthotrichales) (see clade gg in Figs. 4.2, 4.3).

Robust inference of relationships among the "early-diverging" Bryidae lineages has proved to be difficult. In the best trees recovered here, there are short internodes defining the first splits, which in turn lead to long terminal lineages. Inference of these relationships may be especially susceptible to long-branch attraction. In addition, as there is only a limited number of synapomorphies that can be mapped onto those short internodes, sampling error may lead to the generally low observed branch support for these relationships (clades ee and dd). These relationships may become more clearly resolved by substantial expansion of taxon sampling (to break up the relatively long terminal branches) and addition of more data. Including additional genes with higher substitution rates may help, but this should be done cautiously, given the observed imbalance between short internal nodes and longer terminal branches (Fig. 4.2), and the possibility that this might contribute to long-branch attraction if faster genes are sampled. Improved taxon sampling may help particularly in the core pleurocarps, which represent about half of all moss species (Newton et al., 2007).

4.5.3 The earliest splits in moss phylogeny

The early splits in mosses are represented by the nonperistomate lineages. The branching order among these early-diverging taxa has long been of interest to bryologists. Two major unanswered questions concern relationships between Andreaeobryopsida and Andreaeopsida,

and between Takakiopsida and Sphagnopsida, in relation to other mosses. Here I sampled species representing every genus from the five early-diverging lineages, except for the extremely rare *Ambuchanania* (Sphagnopsida). Consistent with many other studies, *Takakia* (Takakiopsida) and *Sphagnum* (Sphagnopsida) are inferred here to represent the first split(s) among mosses. However, two conflicting placements of *Takakia* and *Sphagnum* are inferred by different ML analyses, as well as in various other studies (e.g., Yatsenyuk, 2001; Cox et al., 2004; Volkmar and Knoop, 2010; Chapter 3). The placements of *Takakia* and *Sphagnum* are heavily affected by different base frequency settings here (Table 4.1; see discussion below). In contrast, the placement of *Andreaeobryum* (Andreaeobryopsida) and *Andreaea* (Andreaeopsida) is recovered consistently among different ML analyses (Table 4.1): they form a clade that defines the next phylogenetic split in moss phylogeny. The sister-group relationship between *Andreaea* and *Andreaeobryum* receives consistently moderate to strong ML bootstrap support here (78% - 94%; Table 4.1). The long branch leading to each genus indicates an early split between these two taxa. The branch connecting *Andreaea* and *Andreaeobryum* to other mosses (branch j; Fig. 4.1), on the other hand, is quite short. This combination of long terminal branches and a short internode may explain the general lack of support for this relationship in most other studies to date (Fig. 4.3).

4.5.4 Placement of *Takakia* and *Sphagnum* in moss phylogeny

The backbone phylogenetic relationships inferred here for peristomate and nonperistomate mosses here are generally consistent among anlayses performed using different phylogenetic criteria, except for for the placements of *Takakia* and *Sphagnum*. I found that the base frequency settings can have striking effects on how the deepest split in moss phylogeny is

resolved (Table 4.1). Equilibrium base frequencies are usually estimated by optimizing them during inference of the best tree, or are calculated based on the observed base frequencies of terminal taxa. Here, application of ML-estimated base frequencies seems to outperform the application of empirical base frequencies, at least in terms of the overall consistency of results. When ML-estimated base frequencies are applied, the same arrangement (*Takakia*-sister) is recovered with moderate to strong support (>85%) using different data partitionings and partitions (Tables 4.1, 4.2). In contrast, strongly conflicting results are found in analyses of different data partitioning when empirical estimates of base frequencies are used.

The GC content of the genes examined here varies greatly among different codon positions and the noncoding regions (Table 4.4), but it does not differ substantially among different major lineages or among the different regions sampled here (data not shown). The GC content of the first and second codon positions is about twice that observed for third codon position and the noncoding sites (Table 4.4). Base composition heterogeneity among different sites may be only one aspect of different tempos and modes of nucleotide substitution in these plastid regions. For example, the values of the among-site rate variation (Γ) and invariant site (I) parameters also vary substantially among data partitions (Table 4.4), as do site-site substitution rate parameters (data not shown). Applying one set of parameters to all sites (or even within major partitions) may over-simplify the pattern of DNA substitution sufficiently to misdirect subsequent phylogenetic inference, at least concerning the very first split in moss phylogeny. The observation that ML vs. empirical estimates of base frequencies can have an effect on anaysis here may be only one symptom of a more complex inference problem, as ML analyses that examine each of the major codon positions individually (i.e., first and second, vs. third and

noncoding combinations), recover the *Takakia*-sister arrangement with moderately strong to strong support, regardless of how base frequencies are estimated (Table 4.2: columns 4, 5, 6), and despite the very different model parameters inferred here (Table 4.4). ML inference may therefore be performing adequately on these distinct partitions alone, whereas a strong interaction among data partitions and their estimated parameter values may be responsible for strongly conflicting inferences made using the full data set, in some cases.

Resolving the relationships among *Takakia*, *Sphagnum* and the remaining mosses is an important outstanding issue in moss phylogeny. I have demonstrated conflicting results here based on different MP and ML analyses, and earlier demonstrated that rate filtering based on ML rate classes is apparently not effective in ameliorating long-branch problems in MP analysis (Chapter 3). In this chapter I showed that ML analyses can be sensitive to parameter settings and data partition methods, in a way that can lead different phylogenetic programs to lead to strongly conflicting inferences. The results of the subsequent AU test, a common method to choose among competing hypotheses, are also conditional on different parameter settings. This suggests that extra caution should be taken when investigating and interpreting the *Takakia*-*Sphagnum* problem. However, my results suggest that *Takakia* is more likely to represent the deepest split in moss phylogeny, since this is the result that has the best ML value (when data are partitioned by codons, when subsets of these partitions are analyzed individually, or when ML-estimates of base frequencies are used). Nonetheless, it is clear that further investigation of this problem is needed.

Table 4.1a. Bootstrap support for relationships among the four major land plant lineages and algal relatives. Labels in the first column correspond to individual branches (taxon bipartitions) in Fig. 4.1. ML analyses were performed using Garli and RAxML, with likelihood-based or empirical estimates of base frequencies. ML analysis abbreviations: Un-P, unpartitioned; P-codon: partitioned by codon position; P-region: partitioned by major region. The likelihood values of the optimal trees were calculated using different programs (i.e., Garli or RAxML) and so are not directly comparable (Sundberg et al., 2008). Bootstrap support less than 50% is represented by --.

		ML									**MP**
		Garli, estimated base frequencies			Garli, empirical base frequencies			RAxML, empirical base frequencies			
		Un-P	P-codon	P-region	Un-P	P-codon	P-region	Un-P	P-codon	P-region	
	Log likelihood of optimal tree	247052	239161	244660	248329	239534	246678	249033	242990	248127	n/a
	Taxon bipartition (= clade on rooted tree)										
a	[Zygnematales + land plants]	--	65	--	--	63	--	62	77	--	--
b	Land plants	100	100	100	100	100	100	100	100	100	100
c	Landplants excl. liverworts	100	100	100	100	100	100	100	100	100	94
d	[Hornworts + vascular plants]	98	100	98	100	100	100	100	92	100	90
e	Liverworts	100	100	100	99	100	100	100	100	100	50
f	Mosses	100	100	100	99	100	99	100	100	100	100
g	Hornworts	100	100	100	100	100	100	100	100	100	100
h	Vascular plants	96	97	93	96	97	98	100	85	100	75

Table 4.1b. Bootstrap support for the multi-family clades of mosses. Labels in first column correspond to branches (taxon bipartitions) in Figs. 4.2 and 4.3 (except i2 and ee2). Clade composition comparisons are relative to Fig. 3 of Cox et al. (2004) for clades i, j-q, Fig. 3 of Wahrmund et al. (2010) for clades r-bb, and Fig. 2 of Bell et al. (2007) for clades cc-oo. MRCA = most recent common ancestor. ML analyses were performed using Garli and RAxML, with likelihood-based or empirical estimates base frequencies. ML analysis abbreviations: Un-P, unpartitioned; P-codon: partitioned by codon position; P-region: partitioned by region. The likelihood values of the optimal trees were calculated using different programs (i.e., Garli or RAxML) and so are not directly comparable (Sundberg et al., 2008). Bootstrap support less than 50% is represented by --.

		ML									**MP**
		Garli, estimated base frequencies			Garli, empirical base frequencies			RAxMl, empirical base frequencies			
		Un-P	P-codon	P-region	Un-P	P-codon	P-region	Un-P	P-codon	P-region	
	Log likelihood of optimal tree	247052	239161	244660	248329	239534	246678	249033	242990	248127	
	Taxon bipartition (= clade on rooted tree)										
i	Mosses excl. Takakiopsida	100	100	99	--	100	--	--	100	--	--
i2	Takakiopsida + Sphagnopsida	--	--	--	70	--	72	69	--	97	100
j	Andreaeopsida + Andreaeobryopsida	81	85	78	92	83	93	94	92	78	--
k	Mosses excl. Takakiopsida and Sphagnopsida	100	100	100	100	100	100	100	100	100	--
l	[Oedipodopsida + peristomate mosses]	100	100	100	100	100	100	100	100	100	97
m	Peristomate mosses	100	100	100	100	100	100	100	100	100	68
n	[Tetraphidopsida + Bryopsida]	56	72	--	63	73	59	--	85	68	--
o	Bryopsida	100	100	100	100	100	100	100	100	100	99
p	Bryopsida excl. Buxbaumiidae	100	100	100	100	100	100	100	100	100	100
q	Bryopsida excl. Buxbaumiidae & Diphysciidae	100	100	100	100	100	100	100	100	100	100
r	Funariidae[1]	99	99	99	99	99	99	84	100	100	89[1]
s	[Timmiidae + Dicranidae + Bryidae]	99	98	98	99	98	98	100	100	100	76
t	[Dicranidae + Bryidae]	94	89	96	93	90	95	86	100	92	89
u	Dicranidae[2]	100	100	100	100	100	100	100	100	100	100
v	Core Dicranidae[3] + Bryoxiphiales	91	96	89	82	93	81	76	100	60	--
w	Core Dicranidae[3]	99	99	98	96	98	93	100	100	87	--
x	[*Leucobryum* + Archidales]	100	100	100	100	100	100	100	100	100	100
y	Grimmiales	100	100	100	100	100	100	100	100	100	100
z	[Grimmiales + Pottiales + Dicranales s.s.[4]]	63	74	66	--	74	57	69	85	54	--
aa	Pottiales + Dicranales s.s.[4]	93	87	92	92	89	92	100	100	94	59
bb	MRCA of *Schistostega* & Pottiales	79	76	84	89	76	88	92	92	78	71
cc	Bryidae	100	100	100	100	100	100	100	100	100	100
dd	[Bartramiales + Splachnales]	--	--	--	--	--	--	--	--	--	--
ee	Bryidae excl. Hedwigiales	--	--	--	--	--	--	--	--	--	--
ee	[Bartramiales + Hedwigiales]	--	--	--	74	--	73	76	62	86	98
ff	[Bryales + Orthotrichales + pleurocarpids]	84	83	90	80	80	80	76	--	73	67
gg	[Orthotrichales + *Rhizogonium*]	--	53	--	--	--	--	--	--	--	67
hh	[Orthotrichales + pleurocarpids[5]]	73	83	80	69	76	73	52	91	68	--
ii	[*Aulacomnium* + core pleurocarps[6]]	91	87	91	84	86	79	84	68	92	67
jj	[Ptychomniales + Hypnodendrales]	--	55	--	--	--	--	--	--	--	72
kk	Core pleurocarps[6]	80	85	87	68	85	68	72	69	92	--
ll	[Hookeriales + Hypnales]	100	100	100	100	100	100	100	100	100	--
m	[Hypopterygiaceae + Hypnales]	--	70	51	59	73	57	60	77	54	--
nn	Hypnales	100	100	95	100	100	100	100	100	100	90
oo	MRCA of Brachytheciaceae and Hypnaceae	94	92	95	95	90	96	100	85	84	92

Table 4.2. The placement of *Takakia* and *Sphagnum* in various ML analyses (bootstrap support in parentheses), as determined using Garli, considering either likelihood or empirical estimates of base frequencies. 'T-sister': *Takakia*-sister to other mosses (including *Sphagnum*); 'T-S clade': clade comprising *Takakia* and *Sphagnum*. Note that several bootstrap support values (columns 2-4) are also cited in Table 4.1.

Base frequency estimates in Garli	Matrix used in Garli				
	All (unpartitioned)	All (partitioned by major regions)	All (partitioned by three codon positions and noncoding regions)	First and second codon positions only	Third codon position and noncoding regions
ML estimates	T-sister (100%)	T-sister (99%)	T-sister (100%)	T-sister (86%)	T-sister (100%)
Empirical	T-S clade (70%)	T-S clade (72%)	T-sister (100%)	T-sister (82%)	T-sister (100%)

Table 4.3. Approximately Unbiased (AU) tests of whether suboptimal arrangements of *Takakia* and *Sphagnum* are significantly worse than the optimal one. Tree searches (with or without constraint) were based on the unpartitioned matrix using Garli, with likelihood-estimates of all DNA substitution model parameters. Likelihood values used in AU tests perfirormed in CONSEL were calculated in PAUP* (tree scores not comparable with Garli estimates in Table 4.1).

	Likelihood estimates of base frequencies (PAUP*)		Empirical estimates of base frequencies (PAUP*)	
	-*ln*L (PAUP*)	AU test result	-*ln*L (PAUP*)	AU test result
Takakia sister to other mosses	245901.643	Best tree	248346.914	$P = 0.999$
Takakia-Sphagnum	245930.119	$P < 0.001$	248340.113	Best tree
Sphagnum sister to other mosses	245933.264	$P < 0.001$	248366.454	$P < 0.001$

Table 4.4. DNA substitution model parameters estimated for different data partitions. All parameter values are likelihood estimates, but the Γ and I values were estimated given empirical estimates of base frequencies. Noncoding regions from green algae and vascular plants were not included in the current matrix. The base frequencies of non-coding regions are the average over bryophyte taxa sampled in this study.

	Observed base frequencies			ML estimate values	
	Pi (C)	Pi (G)	Pi (G+C)	Γ	I
All sites	0.1624	0.2025	0.3649	0.7249	0.2844
1st codons	0.1715	0.3320	0.5025	0.6941	0.3365
2nd codons	0.2337	0.1970	0.4307	0.5578	0.3991
3rd codons	0.1022	0.1078	0.2100	2.0568	0.0338
Noncoding	0.1049	0.1178	0.2227	0.9554	0.1026

Table 4.5. Variation in lengths of noncoding regions in nonperistomate vs. peristomate mosses (lengths rounded to nearest bp).

Region	Mean length	Standard deviation	Range
	nonperistomate / peristomate	nonperistomate / peristomate	nonperistomate / peristomate
psb*E-*psb*F-*psb*L-p*sbJ			
*psb*E-*psb*F	9/9	0/0	9/9-10
*psb*F-*psb*L	26/35	7/1	22-37/34-36
*psb*L-*psb*J	112/117	4/2	108-115/113-122
psbB-psbT-psbN*-p*sbH			
*psb*B-*psb*T	135/155	3/23	132-140/100-210
*psb*T-*psb*N	108/83	8/10	94-114/64-104
*psb*N-*psb*H	84/94	2/9	83-88/88-120
***rpl*2** intron	621/618	6/22	616-628/566-666
3'*rps*12-*rps*7-*ndh*B*			
*rps*12 intron	571/528	10/53	559-582/474-588
*rps*12-*rps*7	46 /44	14/15	25-53/20-55
*rps*7-*ndhB*	281/204	83/13	204-369/190-216
*ndh*B intron	765/641	58/22	722-855/614-671

* Data for peristomate mosses based on eight exemplar species (*Alophosia azorica, Atrichum selwynii, Tetraphis pellucida, Funaria hygrometrica, Timmia austriaca*, *Dicranum scoparium, Orthotrichum lyellii* and *Rhytidiadelphus loreus*).

Fig. 4.1. The relationships of land plants and algal relatives, inferred from ML analysis of an unpartitioned matrix with ML-estimates of base frequencies, using Garli (-*ln*L = 245901.624). The numbers by the branches are bootstrap support values. Support values lower than 50% are represented by a dash (--). Species names of land plants, bootstrap support for relationships within each land-plant lineage have been removed, and branches condensed vertically for simplicity. Clade labels are those used in Table 4.1.

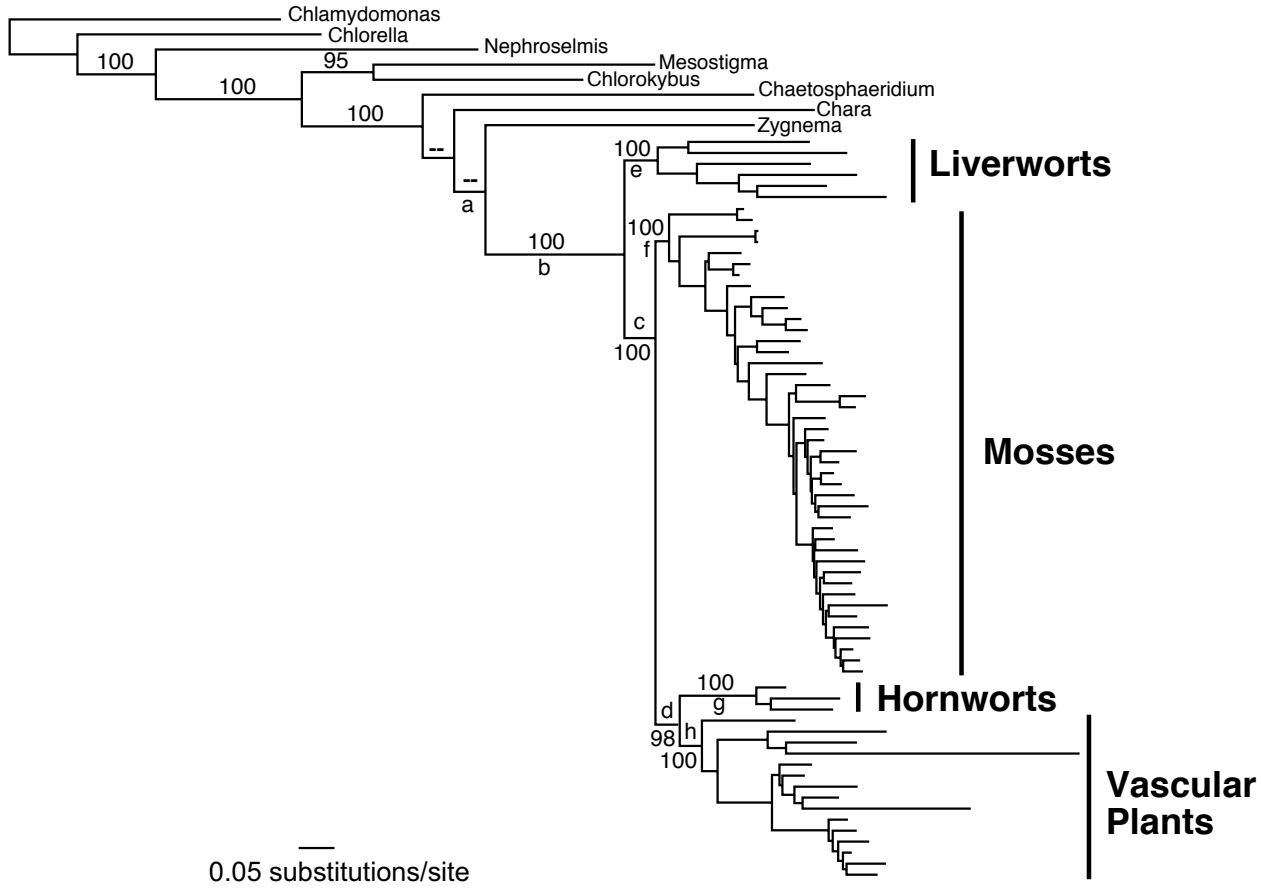


Fig. 4.2. Phylogram of moss phylogeny inferred from ML analysis of an unpartitioned data matrix with ML-estimates of base frequencies, using Garli (this phylogram corresponds to the portion labelled as mosses in Fig. 4.1; the stem lineage has been shortened). The classification follows Goffinet et al. (2009); orders are indicated for exemplar taxa in Bryidae, Dicranidae and Funariidae. Major peristome types: nematodontous – Tetraphidopsida and Polytrichopsida; arthrodontous – Bryopsida; diplolepideous-opposite – Funariidae; diplolepideous-alternate – Bryidae; haplolepideous – Dicranidae. Other major clades: core Dicranidae – clade w; pleurocarpids – clade hh (if it excluded *Orthotrichum*); core pleurocarps – clade kk (see Table 4.1 for clade definitions and support in other analyses). Thickened lines represent strong ML bootstrap support (>90%). ML Support values between 50% and 90% are labeled above or beside the branches; those with <50% support are indicated with a dash (--).

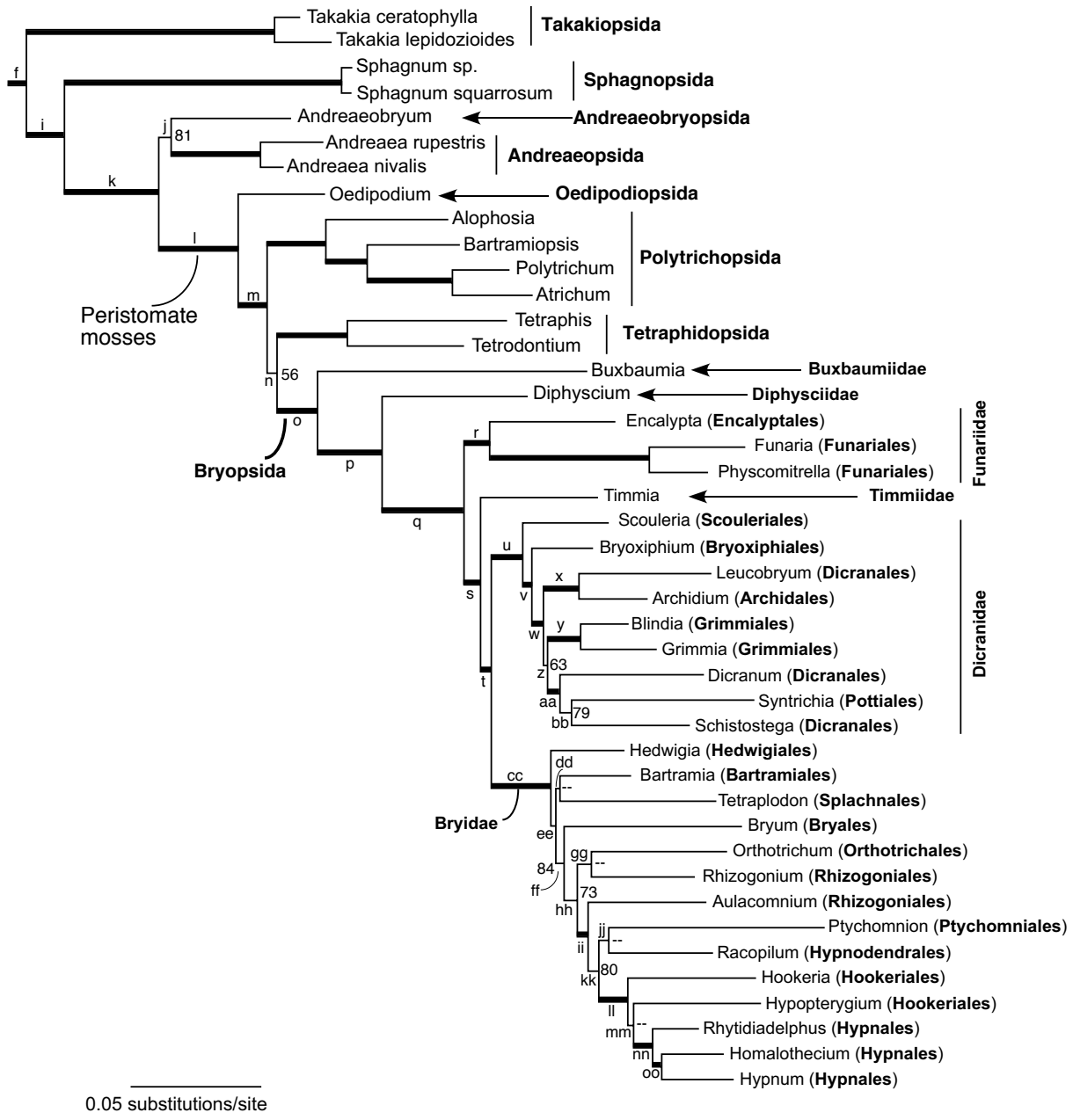
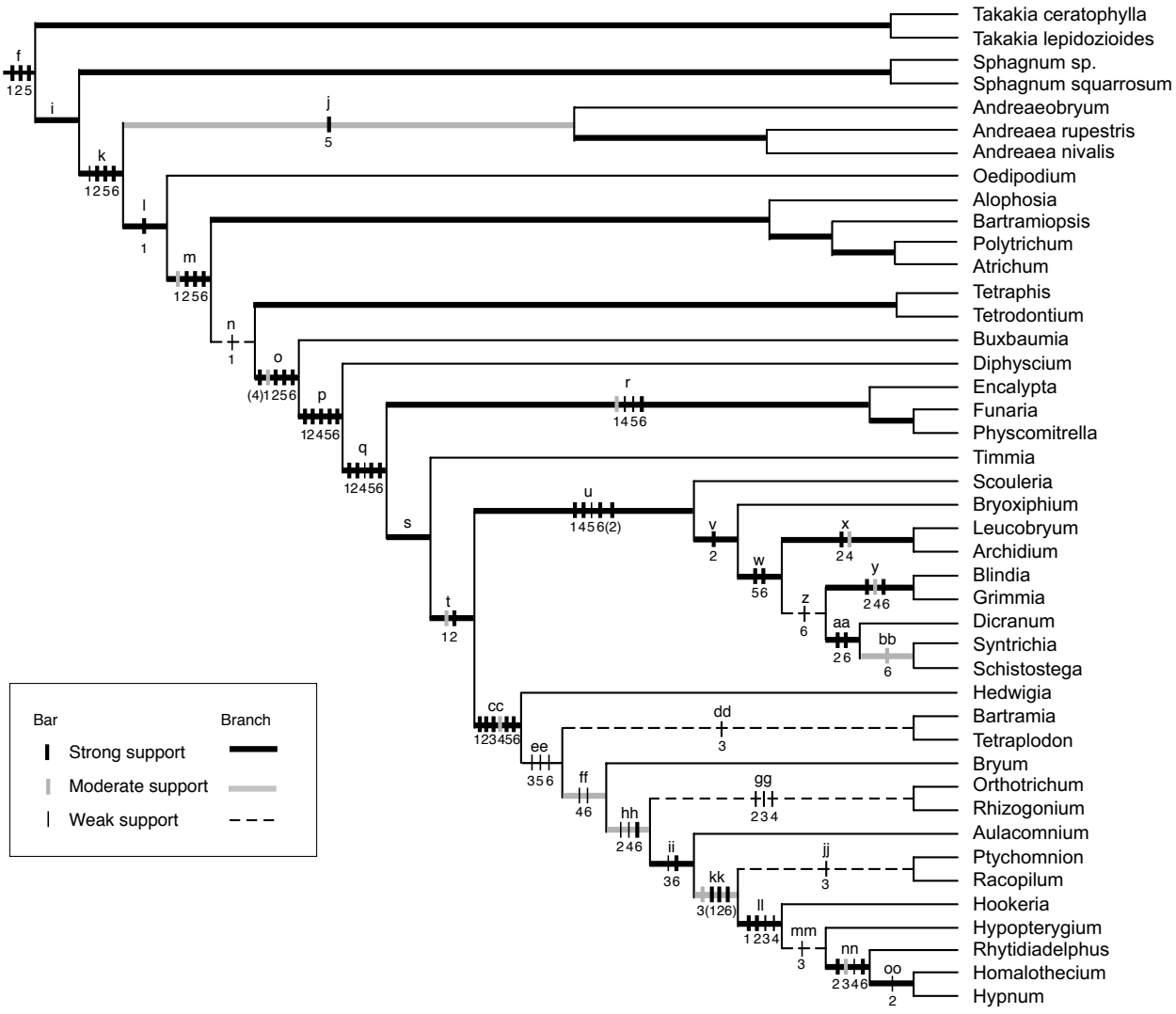

Takakia ceratophylla
Takakia lepidozioides
Takakiopsida
Sphagnum sp.
Sphagnum squarrosum
Sphagnopsida
Andreaeobryum
Andreaeobryopsida
Andreaea rupestris
Andreaea nivalis
Andreaeopsida
Oedipodium
Oedipodiopsida
Alophosia
Bartramiopsis
Polytrichum
Atrichum
Polytrichopsida
Peristomate mosses
Tetraphis
Tetrodontium
Tetraphidopsida
Buxbaumia
Buxbaumiidae
Diphyscium
Diphysciidae
Bryopsida
Encalypta (Encalyptales)
Funaria (Funariales)
Physcomitrella (Funariales)
Funariidae
Timmia
Timmiidae
Scouleria (Scouleriales)
Bryoxiphium (Bryoxiphiales)
Leucobryum (Dicranales)
Archidium (Archidales)
Blindia (Grimmiales)
Grimmia (Grimmiales)
Dicranum (Dicranales)
Syntrichia (Pottiales)
Schistostega (Dicranales)
Dicranidae
Bryidae
Hedwigia (Hedwigiales)
Bartramia (Bartramiales)
Tetraplodon (Splachnales)
Bryum (Bryales)
Orthotrichum (Orthotrichales)
Rhizogonium (Rhizogoniales)
Aulacomnium (Rhizogoniales)
Ptychomnion (Ptychomniales)
Racopilum (Hypnodendrales)
Hookeria (Hookeriales)
Hypopterygium (Hookeriales)
Rhytidiadelphus (Hypnales)
Homalothecium (Hypnales)
Hypnum (Hypnales)
0.05 substitutions/site

Fig. 4.3. Support for interfamilial relationships among mosses from this study and selected studies. The cladogram topology is from the ML analysis of an unpartitioned data matrix with ML-estimates of base frequencies, using Garli (Fig. 4.1, 4.2; see Table 4.1 for clade definitions and support in other analyses here). In several studies, the MRCA of the exemplar taxa is less inclusive than the relevant reference phylogeny. The least inclusive clades are shown; more inclusive interpretations are indicated in parentheses; in several cases, my exemplar-based sampling is also less inclusive than the reference (indicated with an asterisk). References: 1 -- Cox et al., 2004 (ML); 2 -- Bell et al., 2007 (MP); 3 -- Qiu et al., 2007 (ML); 4 -- Stech and Frey, 2008 (MP); 5 -- Volkmar and Knoop, 2010 (ML); 6 -- Wahrmund et al., 2010 (ML).


Takakia ceratophylla
Takakia lepidozioides
Sphagnum sp.
Sphagnum squarrosum
Andreaeobryum
Andreaea rupestris
Andreaea nivalis
Oedipodium
Alophosia
Bartramiopsis
Polytrichum
Atrichum
Tetraphis
Tetrodontium
Buxbaumia
Diphyscium
Encalypta
Funaria
Physcomitrella
Timmia
Scouleria
Bryoxiphium
Leucobryum
Archidium
Blindia
Grimmia
Dicranum
Syntrichia
Schistostega
Hedwigia
Bartramia
Tetraplodon
Bryum
Orthotrichum
Rhizogonium
Aulacomnium
Ptychomnion
Racopilum
Hookeria
Hypopterygium
Rhytidiadelphus
Homalothecium
Hypnum
f 125
i
j 5
k 1256
l 1
m 1256
n 1
o (4)1256
p 12456
q 12456
r 1456
s
u 1456(2)
v 2
x 24
w 56
y 246
z 6
aa 26
bb 6
t 12
cc 123456
dd 3
ee 356
ff 46
gg 234
hh 246
ii 36
jj 3
kk 3(126)
ll 1234
mm 3
nn 2346
oo 2
Bar
Branch
Strong support
Moderate support
Weak support

5 MOLECULAR PHYLOGENETICS OF ANDREAEACEAE

5.1 Brief synopsis

The monogeneric family Andreaeaceae is one of the major nonperistomate moss lineages and represents a very early split in moss phylogeny. To examine its phylogeny and test the current infrageneric taxonomy (subgenera and sections), I sampled six plastid DNA markers from a total of 39 specimens (representing 22 different species), with an emphasis on the costate taxa. I identified multiple well-supported major clades within the genus, but none of the currently recognized infrageneric taxa was supported. Relationships among the clades were also generally strongly supported by maximum likelihood (ML) bootstrap analysis. The lack of monophyly (or substantial intraspecific molecular divergences) of several species indicates possible cryptic speciation for these species. I sub-sampled five species that represent the major lineages of the family for the 14-gene set employed in earlier chapters. ML analysis of this expanded gene-set confirmed the overall relationships in Andreaeaceae found with the six-gene data set, and the sister-group relationship of Andreaeaceae with Andreaeobryaceae.

5.2 Introduction

Andreaea Hedw., the only genus of the granite moss family, Andreaeaceae (Bryophyta), comprises 50 ~ 75 species. *Andreaea* is mainly distributed in high latitude and high altitude regions of both hemispheres, with its greatest diversity in cool-temperate regions (Schofield, 1985; Murray, 2006). *Andreaea* is commonly found growing on granitic or other acidic rocks. In contrast to most mosses, *Andreaea* lacks peristome teeth to aid in spore release; instead, the sporangium of *Andreaea* opens up along four to eight longitudinal lines that extend from the

sporangium base up to its apex. When dry, the dehisced mature sporangium resembles a Chinese lantern, which gives *Andreaea* its other name, the Chinese lantern moss. The sporophyte lacks a seta and the sporangium is instead held up by a pseudopodium derived from gametophytic tissue. Collectively, these unusual morphological features support the view that *Andreaea* represents an isolated lineage in moss phylogeny (Mishler and Churchill, 1984; Murray 1988a). Recent phylogenetic studies that considered evidence from morphological and various molecular markers support an early divergence of the lineage leading to *Andreaea* in the evolution of mosses, as one of several nonperistomate lineages that arose before the origin of the large peristomate clade of mosses (e.g., Mishler and Chruchill 1984; Cox et al. 2004). A proposed sister-group relationship between *Andreaea* and the monotypic family Andreaeobryaceae has been confirmed by various phylogenetic studies, although generally with weak support until the recent studies by Volkmar and Knoop (2010) and Chapters 3, 4.

Despite advances in understanding its placement in moss phylogeny, the taxonomy of *Andreaea* needs further work. The most comprehensive recent treatments of the genus were performed by Schultze-Motel (1970) and Murray (1988b). The former treated only the costate species, which comprise approximately one third of species in the genus. The latter provided a detailed treatment of the *Andreaea* species of the British Isles, which has a similar species composition to North America (~11-12 species). The greatest diversity of *Andreaea* is found in the Southern Hemisphere (Murray, 1986), where it is understudied, except for the Australian taxa. *Andreaea* has generally been divided into two to three subgroups that have been treated at various ranks (e.g., Braithwaite, 1880-1887; Sainsbury, 1955; Schultze-Motel, 1970). The taxonomy used here largely follows Murray (1988a), with three subgenera recognized: subgenus

Acroschisma Hook. f. & Wilson (1 species), subgenus *Chasmocalyx* (Braithw.) Limpr. (~ 5 species), and subgenus *Andreaea* (~ 45 species). The only member of subgenus *Acroschisma*, *A. wilsonii* Hook. f., stands out from other *Andreaea* species. Due to its unique sporangium morphology, some authors treated *A. wilsonii* as its own genus (e.g., Ochyra et al., 2003). Subgenera *Andreaea* and *Chasmocalyx* are each defined based mainly on characteristics of their perichaetial leaves. *Chasmocalyx* has costate perichaetial leaves that are similar in shape to vegetative leaves and are generally not convoluted; in contrast, the perichaetial leaves of subgenus *Andreaea* are strongly differentiated from vegetative leaves and are convoluted. Subgenus *Andreaea* is further divided into two sections: sections *Nerviae* Card. ex Broth. (9-10 species) and *Andreaea* (~ 35 or more species), based mainly on the presence vs. absence of costa in the vegetative leaf, respectively. The ecostate section *Andreaea* was believed to be a heterogeneous group comprising several independent evolutionary lines, and requires further taxonomic work (Murray, 1988b).

Phylogenetic relationships among different *Andreaea* species have also not been assessed, except for an unpublished phylogenetic study based on two plastid regions (Hedderson, pers. comm.). A robust phylogeny would be useful for future taxonomic revision of the genus and permit reconstructions of ancestral character states, dates and biogeography. In this study, I attempted to reconstruct an overall framework for *Andreaea* phylogeny using multiple plastid markers. The major goals of this study are to 1) infer phylogenetic relationships using an exemplar-based sampling and; 2) test the monophyly of intra-generic taxa. The taxonomy of the costate species has been better studied; there is no worldwide treatment of the ecostate species. I focused my sampling efforts on the costate taxa, but sampled a broad selection of ecostate

species for which material was available, choosing ones that are relatively well defined morphologically.

5.3 Materials and methods

I used two different sets of plastid genes to examine phylogenetic relationships within *Andreaea*. For the first of these, I recovered six plastid regions with a variety of rates to assess deep and recent phylogenetic splits: *atp*B, *atp*B-*rbc*L, *ndh*F, *rbc*L, *rpo*B and *atp*F-*atp*H. The *rpo*B locus and the *atp*F-*atp*H intergenic spacer region have been used or proposed for DNA barcoding studies (e.g., Chase et al. 2007; Wang et al., 2010), the *atp*B-*rbc*L intergenic region has been widely used in bryophyte phylogenetic studies (Stech and Quant, 2010), and the *atp*B, *ndh*F, and *rbc*L genes have been used extensively in phylogenetic studies of various plant groups, including those involving higher-order relationships (e.g., Tsubota et al., 2003; Saarela et al., 2007; Smith et al., 2010; Chapters 3, 4). I sampled a total of 39 specimens for these markers from four species in subgenus *Chasmocalyx,* nine in section *Nerviae*, seven in section *Andreaea,* one in section *Acroschisma*, and one species (*A. fuegiana* [Cardot] S. W. Greene) with an uncertain taxonomic placement (Table 5.1, Appendix 4). Only two costate species were not sampled here: *A. depressinervis* Cardot. (with uncertain subgeneric placement) and *A. pachyphylla* (Müll. Hal.) Broth. (subgenus *Chasmocalyx*). No recent material is available for either species, as they have very restricted distribution and have been rarely collected in recent years. I included multiple samples from different geographical ranges for most species, where material was available, especially for the costate species (among the 13 sampled costate species, the majority of species were sampled with two or more accessions from different populations, except for *A. schofieldiana, A. crassinervia* and *A. megistospora*; Table 5.1). I used

Andreaeobryum macroporum Steere & B. M. Murray (Andreaeobryaceae) as an outgroup for this gene set.

Relatively long branches connect *Andreaea* and its putative sister group, *Andreaeobryum* (Volkmar and Knoop 2010; Chapters 3, 4). Rooting of major clades may be problematic when outgroups are distantly related to the ingroup (Graham et al., 2002), and this appears to be the case here, as the deepest split in *Andreaea* is only moderately well supported by bootstrap analysis of the 22-species, 6-region data set (see Results). Long-branch problems associated with distant outgroups may potentially be ameliorated by adding additional taxa to break up long outgroup branches (e.g., Graham et al., 2002; Graham and Iles, 2009). I therefore sampled multiple outgroups and added additional slowly evolving plastid markers to confirm overall root placement in *Andreaea* and relationships among the major subclades of *Andreaea*. I focused on five exemplar species of *Andreaea* that represent these lineages according to the 22-species data set mentioned above, and sampled and aligned 14 plastid regions (*atp*B, *rbc*L, and *ndh*F and 11 additional plastid genes: *psb*B, *psb*C, *psb*D, *psb*E, *psb*F, *psb*H, *psb*J, *psb*L, *psb*N, *psb*T, and *rpl*2, and associated noncoding regions, Chapter 2) to an existing matrix consisting of sequences from other major moss groups and relatives (the 14-gene matrix used in Chapter 4); these effectively served as additional outgroups. I was also interested in assessing whether including multiple *Andreaea* species would have a substantial effect on the strength of the sister-group relationship between *Andreaea* and *Andreaeobryum*, as only one *Andreaea* species has been included in most previous studies (e.g., Cox et al., 2004; Volkmar and Knoop, 2010; Chapters 3, 4), and this relationship was often only moderately well supported for the gene set considered here (Chapter 3, 4).

Protocols for DNA extraction, amplification and sequencing of targeted regions and for base calling and sequence alignment followed those described in Chapters 2 and 3. Primers used to amplify and sequence the targeted regions were previously published (Table 5.2). I used jModelTest 0.1.1 (Guindon and Gascuel, 2003; Posada, 2008) to choose a DNA substitution model for each individual region using the standard Akaike Information Criterion (AIC; Akaike, 1974), the AIC corrected for small samples (AICc), and the Bayesian Information Criterion (BIC; Schwarz, 1978). When different models were selected according to different criteria, I chose a model that fell within the 95% confidence interval for all criteria.

I carried out both maximum parsimony (MP) and maximum likelihood (ML) analyses. I used PAUP* 4.0 b10 (Swofford, 2002) for MP analyses, with tree-bisection-reconnection (TBR) branch-swapping for 100 random addition replicates, and otherwise with default settings. I used Garli 1.0 and Garli-Part-0.97 (Zwickl, 2006) for unpartitioned and partitioned ML analyses, respectively. In ML analyses, I ran ten independent replicates to search for the best tree, with the model parameters estimated. To assess branch support in MP and ML analyses, I performed non-parametric bootstrapping (Felsenstein, 1985) with 500 bootstrap replicates, using simple taxon addition for MP, but otherwise using the search conditions described above.

For the six-region sampling, I first did maximum likelihood analysis using Garli 1.0 (Zwickl, 2006) on individual regions. The results based on different individual regions showed no strong conflicts, and so I concatenated all regions into one matrix. I then performed MP and unpartitioned and partitioned ML analyses. For the 14-gene matrix, I performed unpartitioned ML analysis on the combined dataset, with the associated noncoding regions included or excluded. I used GTR + Γ + I as the DNA substitution model for both analyses. Throughout the

following discussion, I consider bootstrap support less than 70% as weak, between 70% and 90% as moderate, and higher than 90% as strong (e.g., Graham et al., 1998).

5.4 Results

In this section, I first report results for the six-region dataset. I recovered the four coding regions (*atp*B, *ndh*F, *rbc*L, and *rpo*B; see Table 5.3) for most specimens. Due to difficulties with amplification, data were missing from the two intergenic spacer regions, *atp*B-*rbc*L and especially *atp*F-*atp*H, for some specimens (Table 5.3). However, the majority of specimens had all six regions sequenced or only one region missing. Only one specimen (one of the *Andreaea microvaginata* accessions) had three missing regions (*rbc*L, *atp*B-*rbc*L and *atp*F-*atp*H spacers). I found no strong conflicts among regions for ML analysis of individual regions. I therefore performed ML (unpartitioned and partitioned) and MP analyses on a combined matrix for all six regions. The combined six-locus matrix has 5,551 aligned sites of which 364 sites are parsimony-informative (Table 5.3). The unpartitioned and partitioned best ML trees are identical topologically and have comparable bootstrap support. ML trees inferred from individual regions are generally consistent with the trees based on the combined analyses but with poorer support (Table 5.3). Best MP trees (not shown) are very similar topologically to the best ML trees, with no strong or moderate conflicts. In the following discussion, I mainly focus on the results from the partitioned ML analysis and the MP analysis based on the combined dataset.

Twenty-six of 35 clades recovered in the optimal ML tree based on the partitioned six-region matrix receive moderate or strong support in the partitioned ML and MP analyses of the combined dataset (Fig. 5.1 and Table 5.3). Twenty-two of 35 clades have at least 90% bootstrap support. Seven of 35 clades are weakly supported in both analyses. One clade (clade G in Fig.

5.1) is strongly supported by MP analysis but only weakly supported by the partitioned ML analyses. The two subgenera (*Andreaea*, *Chasmocalyx*) and the two sections of subgenus *Andreaea* (*Andreaea* and *Nerviae*) are not recovered as monophyletic in the best ML tree. However, four notable major clades within the genus are defined by major portions of these taxa (clades A-D in Fig. 5.1). These clades all have strong support (93%-100% bootstrap support in partitioned ML and MP analyses). A sister-group relationship between clades C and D is strongly supported (clade F in Fig. 5.1; 100% bootstrap support in partitioned ML and MP analyses). However, because the sister-group relationship between clades A and B receives only moderate support (clade E in Fig. 5.1; 77% and 78% bootstrap support in partitioned ML and MP analyses), the root of *Andreaea* as a whole is only moderately well supported.

Clades A and B are each composed of only one or two species. *Andreaea blyttii* (section *Nerviae*) is the only member of clade A. *Andreaea nivalis* and *A. rigida*, both from subgenus *Chasmocalyx*, comprise clade B. All six species resolved in clade C are from section *Nerviae*. There are three distinct lineages within clade C: *A.heinemannii* alone, a clade comprising *A. schofieldiana* and *A. frigida*, and a clade comprising *A. rothii, A. crassinervia* and *A. megistospora* (Fig. 5.1), which I refer to here as the *rothii* complex. Relationships among these three subclades of clade C are well supported: the *schofieldiana-frigida* clade is the sister group of the *rothii* complex. Clade D is heterogeneous, as it includes species from all four sub-generic groups. *Andreaea wilsonii* (section *Acroschisma*) is weakly supported as the sister group of the rest of clade D, with a small clade of two *Chasmocalyx* species, *A. nitida* and *A. australis*, the next successive sister group. The seven species sampled from section *Andreaea* do not form a clade, as a small clade of two section *Nerviae* species, *A. subulata* and *A. microvaginata*, is

nested in a grade of species in section *Andreaea* (including *A. fuegiana*). The relationships among these small clades within clade D are mostly weakly supported.

The monophyly of most of the 12 species with multiple accessions is strongly supported at the current taxon population samplings (i.e., *Andreaea alpina, A. australis, A. blyttii, A. heinemannii, A. microvaginata, A. rigida, A. subulata, A. frigida;* >98% ML and MP bootstrap support). However, *A. nivalis* and *A. rupestris* are found to be paraphyletic with moderate support. The monophyly of the remaining species and subspecies that are sampled for multiple accessions is only weakly supported (*A. rothii* spp. *falcata*) or rejected (*A. nitida, A. rothii* and *A. rothii* spp. *rothii*). The AU test cannot reject the monophyly of any of the species (or subspecies) found to be paraphyletic on the best ML topology (Table 5.4).

The best ML trees for the 14-gene matrix (with noncoding included or excluded) have an identical topology and comparable bootstrap support (Fig. 5.2). A clade of *Andreaeobryum* and *Andreaea* is recovered with moderate support (80% and 97% with and without the noncoding regions, respectively; Fig. 5.2), and is the sister group of a clade composed of *Oediopodium* and the peristomate mosses (100% bootstrap support; Fig. 5.2; see Chapters 3, 4). The monophyly of *Andreaea* is strongly supported (100% bootstrap support; Fig. 5.2). The root of *Andreaea* is consistent with the six-region analysis, but is strongly supported for this gene sampling (Fig. 5.2, branches E and F).

5.5 Discussion

Apart from an unpublished study by T. Hedderson (University of Cape Town, pers. comm.), this study provides the first formal phylogeny analysis of *Andreaea*, a highly distinctive nonperistomate moss genus, and the sole extant member of Andreaeaceae. Its monophyly is

strongly supported here for the two different gene and taxon samplings that I considered (Figs. 5.1, 5.2); more broadly, a large multigene analyses confirms the placement of *Andreaea* among the early diverging nonperistomate lineages mosses, and provides moderately strong to strong support for its sister-group relationship with *Andreaeobryum* (Andreaeobryaceae), consistent with previous studies (e.g., Cox et al., 2004; Chapters 3, 4). The deep phylogenetic split observed here between *Andreaea* and *Andreaeobryum* (Fig. 5.2) is consistent with a very early divergence between these lineages, although a formal estimate of the timing of this split remains to be done. Relationships among four major clades of *Andreaea* inferred here (i.e., clades A-D; Fig. 5.1) are mostly well-supported in the 6-region analyses (Fig. 5.1); the 14-gene analysis more firmly places the root node of *Andreaea* recovered by the 6-region analyses (cf. Figs. 5.1, 5.2).

For the 6-region analysis I sampled a broad selection of ecostate species (section *Andreaea* in subgenus *Andreaea*) and included nearly all of the costate species (i.e., species in subgenus *Chasmocalyx,* and section *Nerviae* of subgenus *Andreaea*). While the ecostate taxa need to be sampled more heavily in the future, it is clear that none of the currently recognized subgeneric taxa are monophyletic as currently circumscribed. Subgenus *Chasmocalyx* and section *Nerviae* are each divided into two distantly related clades that are separated by ecostate lineages, and the ecostate section *Andreaea* is part of a large, well-supported clade (clade D) that includes portions (subclades) of *Chasmocalyx* and *Nerviae* (Fig. 5.1). Assuming that the major phylogenetic results uncovered here using plastid data are an accurate reflection of species relationships, my study indicates the need for a substantial re-classification of the genus above the species level, in addition to a re-consideration of the homology of the characters that have traditionally been used to define current taxa at this level. I address the significance of the broad

phylogenetic backbone of *Andreaea* recovered here below, but defer formal changes to the classification of evolutionary reconstructions until a broader sampling of ecostate species can be included and the current results corroborated with data from additional genetic linkage groups (e.g., multiple nuclear genes).

5.5.1 Major lineages of *Andreaea*

I recover four major well-supported clades in this study (clades A-D; Fig. 5.1). The sole member of clade A, *Andreaea blyttii,* is currently included in section *Nerviae.* The isolated placement of *A. blyttii* from other *Nerviae* species here is consistent with earlier morphological observations that *A. blyttii* is distinct from other members of section *Nerviae* (Murray 1988b). The two members of clade B, *A. nivalis* and *A. rigida* (currently in subgenus *Chasmocalyx*), are also morphologically distinctive within *Andreaea*, having lanceolate leaves that are denticulate and papillose, and strong and well-defined costa that are percurrent or excurrent (Schultze-Motel, 1970).

Three distinctive lineages can be found in clade C, which comprises the majority of the species in section *Nerviae*. *Andreaea heinemannii* stands apart from other *Nerviae* because of its small size, its sporangium morphology, and its growth as a short turf on its usual habitats of basalt rocks (Murray, 2006). The European endemic *A. frigida* is sister to the North American endemic *A. schofieldiana*. Both species have lanceolate or gradually tapering leaves with strong costae (Murray 1987a). The three species within the *rothii*-complex are characterized by strongly costate leaves that have a more or less abrupt contraction from the leaf base to upper the part of the leaf (Murray 1987a).

Clade D is a large and heterogeneous group that includes species from all three sub-generic units, and includes all of the ecostate species sampled here. I can identify no obvious morphological synapomorphies for this group. However, all taxa in this clade are defined by a unique 8-base indel (inferred to be a deletion) in the *atp*F-*atp*H region. Based on our current understanding of *Andreaea* morphology, I predict that most, if not all, of the unsampled ecostate taxa belong to this clade.

5.5.2 Species monophyly and species delimitation

Molecular phylogenetic evidence is frequently used to aid in species circumscriptions (e.g., history-based phylogenetic species concepts, e.g., Baum and Donoghue, 1995), supplementing morphological definitions of species and in some cases revealing substantial non-monophyly of species in mosses (e.g., Vanderpoorten and Goffinet, 2006). High genetic variation (or lack of monophyly) accompanied by limited morphological differentiation can be indicative of cryptic speciation (Shaw et al., 2003), which appears to be common for bryophytes, especially those with wide distribution ranges (e.g., Fernandez et al., 2006). However, Vanderpoorten and Shaw (2010) caution that multiple linkage groups need to be considered when using molecular data to assess species boundaries, and remind us that species may often be paraphyletic, at least initially (consistent with the plesiospecies concept of Olmstead, 1995).

Nonetheless, the current phylogenetic data provide initial molecular information on the naturalness of individual species that can then be followed up with more intensive molecular and morphological study, and additional population sampling. Twelve species sampled here included multiple accessions; when I sampled multiple populations per species these were generally well separated geographically (see Fig. 5.1). At the current sampling intensity, eight of these were

recovered as monophyletic with strong support (>97% ML and MP bootstrap support; Fig. 5.1). Among these, *Andreaea alpina* and *A. heinemanii* have a broad distribution and occur in both hemispheres. *Andreaea alpina* is found widely in the Southern Hemisphere and in northwestern Europe and Greenland, while *A. heinemannii* has a disjunct distribution in the Northern Hemisphere, and has recently been recognized in Australia (Murray, 2006), a sample of which was included here. The genetic variation within each of these species, however, is quite different. The distance between the two *A. alpina* accessions is similar to the average intra-species genetic distance of the six other monophyletic species that have limited distributions (see branch lengths in Fig. 5.1). The average intra-species genetic distance of *A. heinemannii*, in contrast, is about five times higher than that of *A. alpina,* which suggests that cryptic speciation may have occurred. The remaining four species with multiple accessions are recovered as monophyletic with weak support, or paraphyletic with moderate support (none are polyphyletic at current samplings). However, the AU test cannot reject the monophyly of any of the four species or the two subspecies of *A. rothii* (Table 5.4); constraining the monophyly of each species does not result in trees that are significantly longer than the best ML tree. Sampling of additional populations from across the range of all of the species examined here will be required to further assess their monophyly (or lack). *Andreaea nitida* and *A. rupestris* would be good targets for initial study as both species are morphologically variable (Murray, 1988b and pers. comm).

5.5.3 Implications of *Andreaea* phylogeny for morphological evolution

I did not perform formal ancestral-state reconstructions here due to the relatively thin sampling of ecostate taxa in this study. Nonetheless, *Andreaea* phylogeny shows character-distribution patterns that are worth attention in light of the current phylogeny, including the

evolution of the sexual system. Our knowledge of sexuality in *Andreaea* is incomplete, but based on our current understanding, the majority of *Andreaea* appear to be monoicous (i.e., with sperm and eggs produced on the same gametophyte). The phylogenetic root of *Andreaea* defined here (Figs. 5.1, 5.2) divides the genus into a dioicous clade (clades A+B) and a primarily monoicous clade (clades C+D): the only known dioicous species are *A. blytii, A. nivalis,* and probably *A. australis* (Murray, 1988b and per comm.). As the sister group of *Andreaea* (*Andreaeobryum*) is dioicous (as is *Takakia* and some members of *Sphagnum*), I predict that dioicy is the ancestral sexual system in the genus, and that monoicy is a derived condition defining clade C and clade D (which contain the majority of *Andreaea* species).

All members of clades A, B, and C have leaves with costae. Clade D is more heterogeneous, as it includes both costate and ecostate taxa. *Andreaeobryum* is costate, and so I predict that leaves with costae were the ancestral condition in *Andreaea* as a whole, and that there were subsequently multiple transitions (gains and/or losses) within clade D. The presence or absence of a costa is clearly not a reliable marker considering current subgeneric taxonomy. It is not uncommon to observe that costa development and morphology vary within individual moss genera (e.g., Allen, 1987; Vanderpoorten, 2002), but few genera are comparable to *Andreaea* in terms of the diversity of morphology of their costa and costa-like structures. Within the genus, the costa may be prominent throughout the leaf (e.g., in *A. nivalis*), restricted to the lower part of the leaf (e.g., in *A. nitida*), or absent towards the leaf base (e.g., in *A. microvaginata*). In addition, the weak or weaker part of the costa can be two or three layers of cells, with little cell differentiation (Murray, 1988b; Frey and Stech, 2010). In some taxa considered to be ecostate, bi- or multi-stratose stripes can be found near the leaf base or along the

leaf apex that resemble a highly reduced costa (e.g., *A. wilsonii*) (pers. obs.). An extreme example is the multistratose stripes found along the leaf margin of *A. fuegiana*, which receives controversial interpretations from different authors (e.g., Schultz-Motel, 1970; Ochyra et al., 2003). The diversity of costa form clearly deserves more attention in a phylogenetic context.

Another intriguing pattern is the suite of characters that are shared by members of *Chasmocalyx*. Besides having large but undifferentiated and non-convolute perichaetial leaves, all species in this subgenus have a pseudopodium that develops from both archegonial and stem tissues. They also have auriculate leaf bases and stems with very small outer cells and a sharp transition to the larger inner cortex cells (Murray, 1988b). However, the four *Chasmocalyx* species sampled in this study are divided between two distantly related clades. This suggests that this suite of characters co-evolved in these two widely isolated lineages, although there is no obvious structural or developmental linkage among these characters.

5.5.4 The placement of *Andreaea wilsonii* and *A. fuegiana*

Among all the *Andreaea* species sampled in this study, *A. wilsonii* and *A. fuegiana* are perhaps the most distinctive in terms of their morphology. *Andreaea fuegiana* has two bi- or multi-stratose stripes near the leaf margin, starting from the base of the leaf and extending all the way to the apex. Whether these stripes are a form of costa remains debatable (e.g., Schultze-Motel 1970; Murray, 1988b; Ochyra et al. 2003). To emphasize the unique leaf morphology, Ochyra et al. (2003) proposed a new genus, *Bicosta,* to accommodate this species. *Andreaea wilsonii* stands out from other *Andreaea* species due to its unique capsule morphology. The sporangium is long-cylindric, similar to the typical bryopsid sporangium, and opens by eight divisions that are confined only to the apex of the sporangium. Based on these features, *A.*

wilsonii has been treated as its own section, subgenus or even as a separate genus (*Acroschisma*) by some authors (e.g., Wijk et al., 1959; Frey and Stech, 2009; Goffinet et al., 2009).

However, my phylogenetic results recover both species as deeply nested within the genus *Andreaea. Andreaea fuegiana* is well resolved as a member of clade D. Its closest relatives, *A. remotifolia* and *A. alpina* (Fig. 5.1) are typical ecostate taxa, and so its "bi-costa"-like stripes likely have an independent origin from other costa-like structures in the genus. The closest relatives of *A. wilsonii* remain uncertain. However, this species is clearly part of clade D, although it is also a relatively isolated lineage. This suggests that the unique capsule found in this species is a modification of the typical valved capsule of *Andreaea.* However, the relatively long branch leading to *A. wilsonii* (Fig. 5.1) may reflect in part an early divergence within clade D, and may be consistent with its relatively drastic change in capsule morphology.

5.6 Conclusions and future directions

A multigene analysis based on larger set of plastid markers confirms the early divergence of the lineage leading to *Andreaea* in moss evolution, and places it as the sister group of the monotypic *Andreaeobryum* with moderately strong support, consistent with previous studies (e.g., Cox et al., 2004; Volkmar and Knoop, 2010; Chapters 3, 4). It also confirms the monophyly of the genus *Andreaea* and supports relationships among and within four major subgeneric clades, permitting inference of the placement of the root node with strong support. I further investigated phylogenetic relationships within the genus using a set of six plastid markers. The phylogenetic relationships inferred from plastid data do not support the current subgeneric taxonomy, and hint at the paraphyly of several species, possibly indicative of cryptic speciation in some cases. Future elaborations of the phylogenetic backbone inferred here should

focus on sampling additional ecostate species and on adding data from the nuclear genome to corroborate or refute it. This will be essential for further improving our understanding of the evolution of this fascinating lineage of mosses.

Table 5.1. The taxonomy of the *Andreaea* species sampled in this study and the number of specimens included for each species. *: the taxonomic treatment of *A. fuegiana* and *A. wilsonii* is controversial (e.g., Vitt, 1984; Ochyra, 2003; Goffinet et al., 2009;).

Subgenus	Section	Species	# of specimens
Chasmocalyx costate; ~5 species		*A. australis* Mitt.	2
		A. nitida Hook. f. & Wilson	*3*
		A. nivalis Hook.	*2*
		A. rigida Wilson	*2*
Andreaea	***Nerviae*** costate; ~9 species	*A. blyttii* Schimp.	*3*
		A. crassinervia Bruch	*1*
		A. frigida Hüb.	*2*
		A. heinemannii Hampe & Müll. Hal.	*3*
		A. megistospora B. M. Murray	*1*
		A. microvaginata Müll. Hal.	*2*
		A. rothii Web. & Mohr.	*4*
		A. schofieldiana B. M. Murray	*1*
		A. subulata Harv.	*2*
	Andreaea ecostate; ~35 species	*A. acutifolia* Hook.f. & Wilson	*1*
		A. acuminata Mitt.	*1*
		A. alpestris (Thed.) Schimp.	*1*
		A. alpina Hedw.	*2*
		A. mutabilis Hook. f. & Wilson	*1*
		A. remotifolia Dusèn	*1*
		A. rupestris Hedw.	*2*
Acroschisma ecostate; 1 species		*A. wilsonii** Hook. f.	*1*
		*A. fuegiana** (Cardot) S. W. Greene	*1*

Table 5.2. Primers used to amplify and sequence the *rpo*B gene, and the intergenic spacers (IGS) between *atp*B and *rbc*L and between *atp*F and *atp*H. The 14-gene regions were amplified with the set of primers cited in Chapter 2 (Figs. 2.1 – 2.7). Citations for the primers: *atp*B-1 and *rbc*L-1 -- Chiang et al. (1998); *atp*F, *atp*H, *rpo*BajfF1, and *rpo*BajfR1 -- Fazekas et al. (2008); B92R -- Chapter 2.

	Forward primers	Reverse primers
*rpo*B	*rpo*BajfF1: TCTAATATGCAICGTCAAGC	*rpo*BajfR1: GAGGIGTTAITTIACCTAC
IGS *atp*B-*rbc*L	*atp*B-1: CCAACCGGAACACTTAARGGAGC B92R: TCCACYACTTTAATTCCTGTTTC	*rbc*L-1: GGTTGAGGAGTCATTCGAAATG
IGS *atp*F-*atp*H	*atp*F: ACTCGCACACACTCCCTTTCC	*atp*H: GCTTTTATGGAAGCTTTAACAAT

Table 5.3. Statistics for the various MP and ML analyses conducted on the six-region dataset.

Regions	Combined		*atp*B	*ndh*F	*rbc*L	*rpo*B	*atp*B-*rbc*L	*atp*F-*atp*H
Number of accessions included [a]	40		40	40	39	39	35	21
Aligned length	5551		1479	985	1369	562	596	559
Number of variable characters	622		124	133	114	74	39	138
Number of parsimony informative characters	364		65	79	67	55	24	74
Number of nodes with BS > 80% in MP analyses [b]	25 [c]		na	na	na	na	na	na
DNA substitution model used in ML analysis [d]	GTR+Γ+I	Partitioned [e]	TIM2+ Γ +I	TPM3uf + Γ +I	GTR+ Γ +I	TPM1uf+ I	TPM1uf+ I	TPM3uf
Likelihood score (best tree)	-12818.86	-12584.12	-3065.51	-2467.67	-2915.51	-1367.78	-1004.26	-1503.16
Number of nodes with BS > 80% in ML analysis [b]	24 [c]	23 [c]	11	10	8	5	5	4
Number of nodes with BS > 70% in ML analysis [b]	27	27	12	12	10	13	5	5

[a] 39 samples from *Andreaea* and one from *Andreaeobryum*.
[b] BS = bootstrap support.
[c] 22, 21 and 22 nodes with BS > 90% in the MP analysis, partitioned and unpartitioned ML analyses, respectively
[d] See Posada (2008) for specifics of DNA substitution models.
[e] In the partitioned ML analysis, individual regions were assigned the individual DNA substitution model shown in the rest of the table.

Table 5.4. Approximately Unbiased (AU) test of whether suboptimal arrangements constraining monophyly of several species are significantly worse than the paraphyletic arrangements in the optimal tree. ΔL = increase in $-ln$L compared to the best ML tree. The analyses consider the combined six-region dataset. All the analyses were based on the partitioned matrix.

Arrangement of interest	ΔlnL	*P* value*
Monophyly of *A. nivalis*	4.27	0.08
Monophyly of *A. rupestris*	6.95	0.21
Monophyly of *A. rothii*	4.30	0.20
Monophyly of *A. rothii* spp. *rothii*	8.86	0.08
Monophyly of *A. nitida*	0.00	0.85

Notes: * The monophyly is rejected when $P < 0.05$.

Fig. 5.1. Phylogeny of *Andreaea* inferred from the partitioned 6-region plastid matrix using maximum likelihood. The outgroup (*Andreaeobryum macrosporum*) is removed for clarity, with the stem lineage shortened. Numbers beside branches are bootstrap support values from partitioned ML analysis and MP analysis. The bootstrap support values from unpartitioned ML analysis (not shown) are largely comparable to those from the partitioned ML analysis. "--" represents support less than 50%. Labels by the dashed lines indicated current taxonomy: subgenus *Chasmocalyx,* sections *Nerviae* and *Andreaea* (in subgenus *Andreaea*).

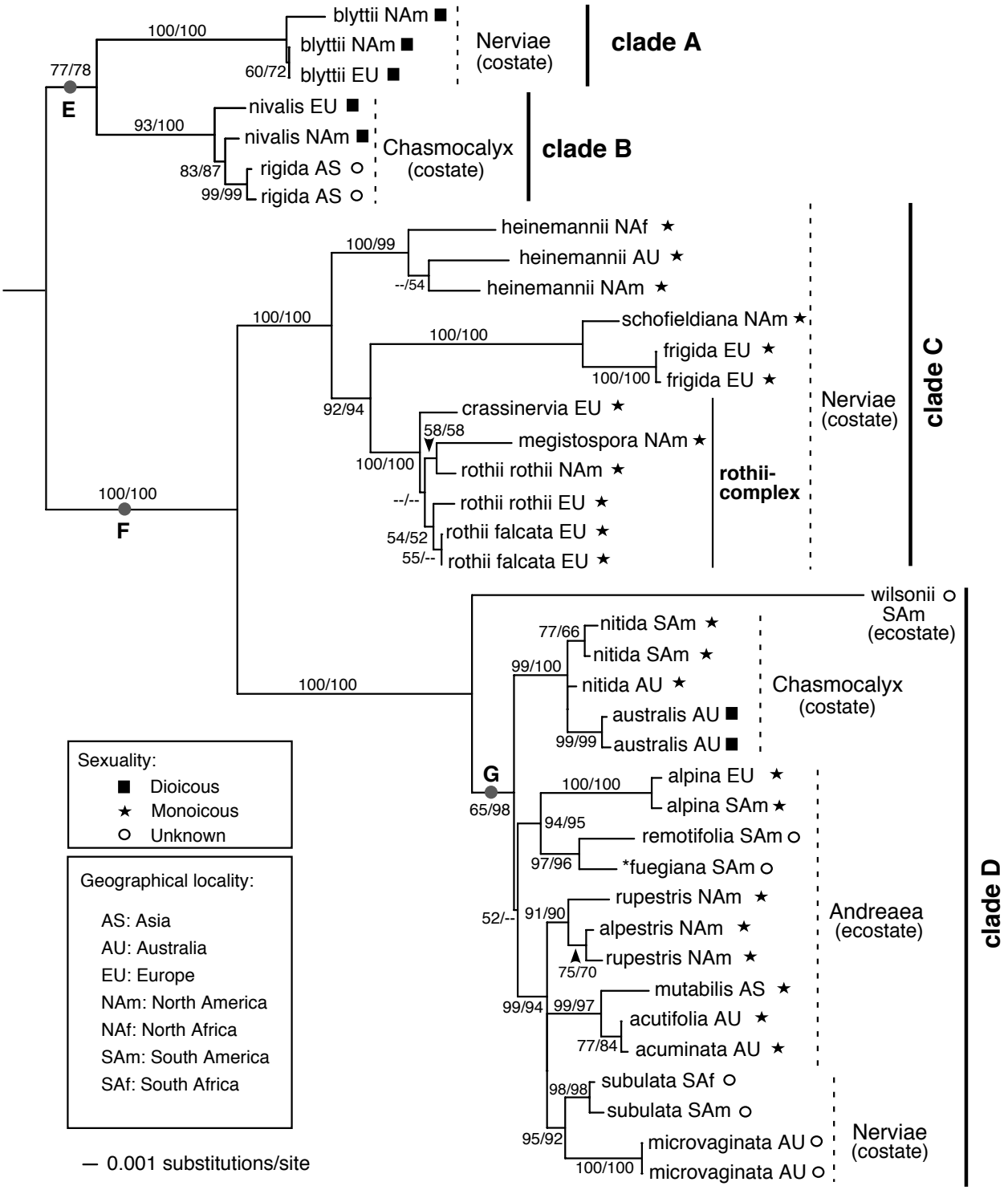
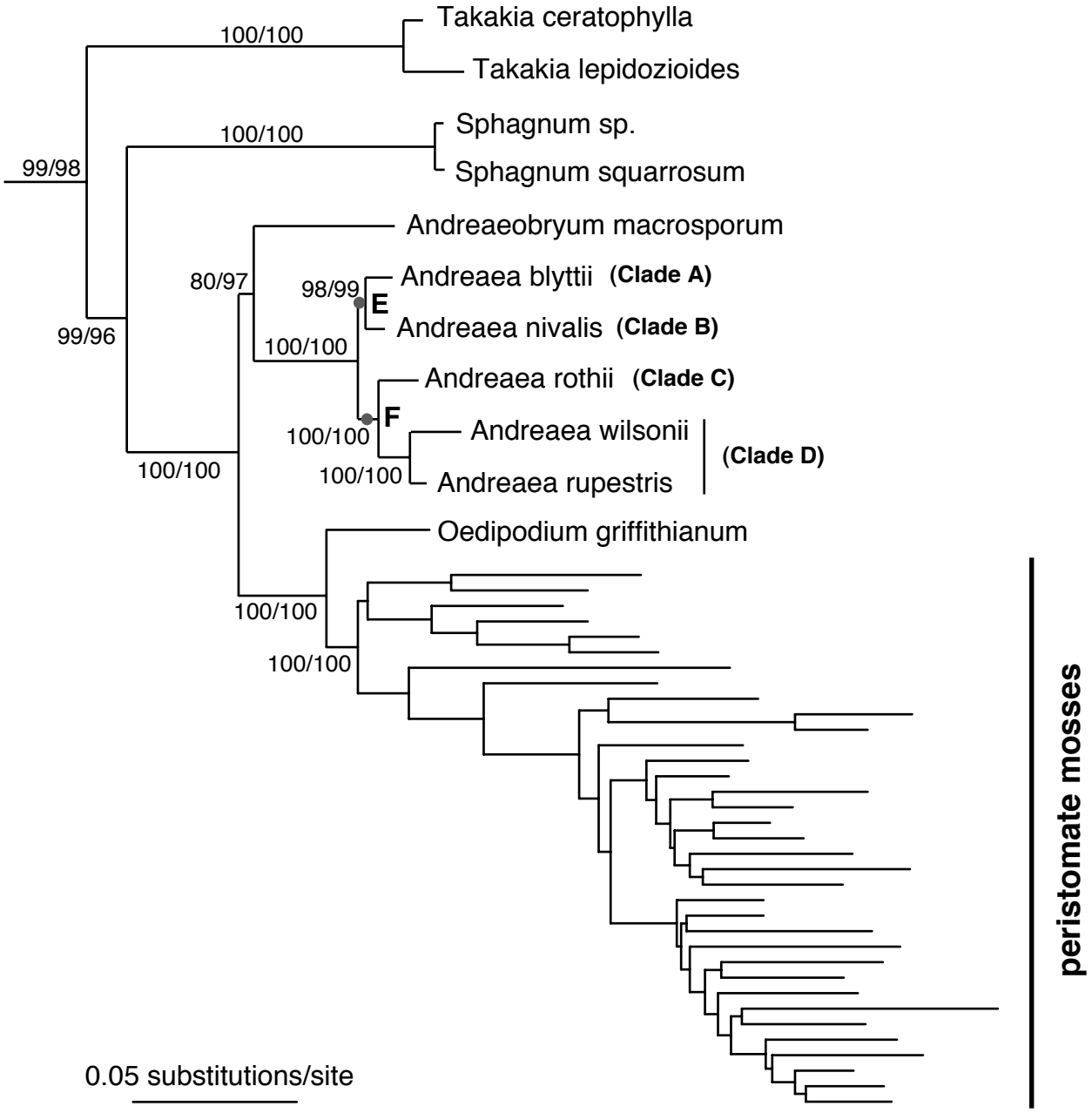

blyttii NAm ■
blyttii NAm ■
blyttii EU ■
100/100
60/72
77/78
E
Nerviae
(costate)
clade A
nivalis EU ■
nivalis NAm ■
93/100
rigida AS ○
rigida AS ○
83/87
99/99
Chasmocalyx
(costate)
clade B
heinemannii NAf ★
heinemannii AU ★
heinemannii NAm ★
100/99
--/54
100/100
schofieldiana NAm ★
frigida EU ★
frigida EU ★
100/100
100/100
92/94
crassinervia EU ★
58/58
megistospora NAm ★
rothii rothii NAm ★
100/100
rothii rothii EU ★
--/--
rothii falcata EU ★
54/52
rothii falcata EU ★
55/--
rothii-
complex
Nerviae
(costate)
clade C
100/100
F
wilsonii ○
SAm
(ecostate)
nitida SAm ★
77/66
nitida SAm ★
99/100
nitida AU ★
australis AU ■
99/99
australis AU ■
Chasmocalyx
(costate)
100/100
G
65/98
alpina EU ★
100/100
alpina SAm ★
94/95
remotifolia SAm ○
97/96
*fuegiana SAm ○
rupestris NAm ★
52/--
91/90
alpestris NAm ★
rupestris NAm ★
75/70
mutabilis AS ★
99/94
99/97
acutifolia AU ★
77/84
acuminata AU ★
Andreaea
(ecostate)
subulata SAf ○
98/98
subulata SAm ○
95/92
microvaginata AU ○
100/100
microvaginata AU ○
Nerviae
(costate)
clade D
Sexuality:
■ Dioicous
★ Monoicous
○ Unknown
Geographical locality:
AS: Asia
AU: Australia
EU: Europe
NAm: North America
NAf: North Africa
SAm: South America
SAf: South Africa
— 0.001 substitutions/site

Fig. 5.2. Phylogenetic placement of Andreaeaceae in moss evolution and the relationships of the major *Andreaea* clades based on ML analysis of the 14-gene matrix. Taxa from other plant groups (including algal outgroups) are removed for clarity, with the stem lineage shortened. Species names of peristomate mosses are removed and branches condensed vertically for simplicity. Only the branches representing the nonperistomate mosses are shown with bootstrap support: numbers beside branches are ML bootstrap values with and without the noncoding regions included, respectively. Clade labels correspond to those in Fig. 5.1.


Takakia ceratophylla
Takakia lepidozioides
100/100
Sphagnum sp.
Sphagnum squarrosum
100/100
99/98
Andreaeobryum macrosporum
80/97
98/99
Andreaea blyttii (Clade A)
E
Andreaea nivalis (Clade B)
99/96
100/100
Andreaea rothii (Clade C)
F
100/100
Andreaea wilsonii
(Clade D)
100/100
Andreaea rupestris
100/100
Oedipodium griffithianum
100/100
peristomate mosses
0.05 substitutions/site

6 CONCLUSIONS AND FUTURE DIRECTIONS

6.1 Major conclusions

A major goal of this thesis was to improve our understanding of the overall phylogenetic backbone of mosses in the context of early land-plant phylogeny. Chapter 2 focused on the development of tools (primers) for retrieving plastid genes from a broad range of bryophyte taxa, yielding a large set of loci suitable for higher-order phylogenetic inference. Subsequent chapters focused on different aspects of phylogenetic inference using these genes, including a demonstration of their utility for inferring very broad-scale relationships among the major embryophyte and moss lineages (Chapter 3), and an expanded taxon sampling focusing on the peristomate moss clade that includes most extant species (Chapter 4). Finally, I developed a well-supported phylogenetic backbone of one of the nonperistomate moss lineages, the granite-moss family, Andreaeaceae, using different sets of plastid genes and taxon samplings (Chapter 5). These phylogenetic backbones should provide frameworks for developing more natural classification schemes and to help shed light on one of the most intriguing sets of events during the evolution of life on earth, the origin and diversification of the early land plants. However, future inferences made about their morphological and genomic evolution in a phylogenetic framework will depend in part on the accuracy of tree inference. It is therefore important that the trees are as well corroborated and as strongly supported as possible. The inferences made in Chapters 3 and 4 largely corroborate those recovered in other studies using different taxon and gene/genomic samplings, and generally provide well-supported tree structure, in several cases with improved support here compared to other studies. The phylogenetic backbone of

Andreaeaceae, the subject of Chapter 5, has not been studied in detail before. These backbones should be useful for additional evolutionary and systematic research on the bryophyte branches of the tree of life.

The study of moss phylogeny has focused on extant taxa because of the paucity of records for fossil bryophytes (Shaw and Renzaglia, 2004). DNA sequence data have played a key role (Goffinet et al., 2009), but compared to vascular plants there has been a relatively limited number of plastid DNA markers used for studying backbone relationships (Stech and Quandt, 2010). In Chapter 2 I developed 78 new primers that permit ready retrieval of 14-17 plastid genes and associated noncoding regions (~8-13 kb unaligned bp) for phylogenetic inference of mosses and relatives. Most of the targeted regions have not been used in the inference of these backbone relationships before. I tested the utility of these primers for retrieving the targeted regions from a broad range of bryophyte taxa. One region comprising three genes (the 3'–*rps*12-*rps*7-*ndh*B region) proved more difficult to retrieve than the others, but in general the primers allowed straightforward amplification and sequencing of the targeted regions from a broad variety of bryophytes. These regions generally have a conservative rate of evolution (plastid DNA is considerably more slowly evolving than nuclear genes; Wolfe et al., 1987; Olmstead and Palmer, 1994) but span a range of evolutionary rates (e.g., coding vs. noncoding regions), and so should be useful at multiple levels of phylogenetic inference in moss and bryophyte phylogeny.

I demonstrated the utility of these regions in phylogenetic inference of the earliest splits in land-plant (embryophyte) phylogeny, even when using a relatively sparse taxon sampling of the major lineages of bryophytes (Chapter 3). The major land-plant groups were resolved as monophyletic with strong support from maximum likelihood (ML) analysis, as were the

relationships among them, corroborating studies based on other multi-gene analyses and plastome structural data (e.g., Kelch et al., 2004; Wolf et al., 2005; Qiu et al., 2006, 2007): in particular, I recovered the liverworts as the sister group of all other land plants, and the hornworts as the sister group of vascular plants, both with strong support. The deepest phylogenetic splits within liverworts and hornworts were also well supported here, and were consistent with current understanding of the backbone phylogeny in each group (e.g., Renzaglia et al., 2007). The charophycean algal lineage Zygnematales (represented by *Zygnema*) was recovered as the sister group of land plants with moderate ML bootstrap support, consistent with some other recent large-scale analyses of plastid and nuclear data (Turmel at el., 2006, 2007; Wodniok et al., 2011).

Relationships among the major lineages of nonperistomate mosses were also generally well supported. A sister-group relationship between Andreaeobryopsida (*Andreaeobryum*) and Andreaeopsida (*Andreaea*) received moderate to strong support from various ML analyses, and Oedipodiopsida (*Oedipodium*) was recovered the sister group of the peristomate mosses with strong support, consistent with previous studies (e.g., Cox et al., 2004). The very first evolutionary split in moss phylogeny has not been resolved satisfactorily to date. Although some studies have recovered a clade comprising *Sphagnum* and *Takakia* as the sister group of the remaining moss taxa (e.g., Cox et al., 2004; Qiu et al., 2006), others find *Takakia* as the sister group of all other mosses (e.g., Forrest et al., 2006; Volkmar and Knoop, their Fig. 3). No morphological synapomorphies satisfactorily support either hypothesis (e.g., Newton et al., 2001), and inference of this important feature of moss phylogeny may be prone to long-branch attraction (Shaw and Renzaglia, 2004). In Chapter 3 I recovered two conflicting hypotheses

regarding the earliest splits in moss phylogeny: *Takakia* as the sister group of *Sphagnum* (using maximum parsimony, MP), or *Takakia* as the sister group of all other mosses (using ML), both with strong bootstrap support. The latter '*Takakia*-sister' result may be more trustworthy, as ML is known to be less prone to long-branch attraction than MP, in general (Huelsenbeck, 1995). More rapidly evolving sites are expected to be more prone to saturation effects (Felsenstein, 1983). When I identified and removed these sites using a likelihood-based site-rate classification, I was able to recover the '*Takakia*-sister' result using MP analysis. However, this result was sensitive to the initial tree topology used to classify site rates, suggesting that ML-based rate filtering is not a general solution to tree mis-inference in MP analysis.

I explored the *Takakia* problem further in Chapter 4, a study that sampled additional exemplar taxa across the moss phylogeny. This study had a major focus on the peristomate mosses, but also included several additional nonperistomate taxa. I found that ML inferences concerning the position of *Takakia* can be highly sensitive to how base frequencies are handled in DNA substitution models (i.e., empirical vs. ML estimates of these parameters), and to the type of data partitioning used. I documented substantial differences in the tempo and mode of DNA substitution in different data partitions, but combined vs. separate analysis of these partitions did always not yield straightforward (predictable) differences concerning which arrangement of *Takakia* was favoured by ML analysis, suggesting that interactions among different data partitions may be at last partly responsible for misdirecting phylogenetic inference. ML analyses that partitioned the plastid data by codon partition (considering noncoding data as a fourth partition) yielded trees with higher likelihood scores than unpartitioned ML analyses or the ML analyses that partitioned the data by major plastid region, regardless of the program used

for ML inference. The codon-based partitioning recovered the *Takakia*-sister arrangement. Although the first split in moss phylogeny remains controversial, the strong conflict observed in ML analysis here for this relationship suggests that using ML estimates of base frequencies may be less misleading than using empirical estimates (the latter method is used as a short-cut in one widely used ML program, the current version of RAxML, Stamatakis, 2006). This conclusion assumes that the *Takakia*-sister arrangement is correct.

The major goal of Chapter 4 was to improve our understanding of relationships among the major lineages of peristomate mosses. The major clades recognized as classes and subclasses in current moss classification (Goffinet et al., 2009) were recovered as monophyletic at current taxon samplings with strong support, as were many of the relationships among them. For example, within Bryopsida, the sub-classes Buxbaumiidae and Diphysciidae were recovered as successive sister-groups to a major clade comprising Funariidae, Timmiidae, Bryidae and Dicranidae, with strong ML bootstrap support. The placement of Funariidae and Timmiidae with respect to the latter two clades has not previously been convincingly resolved. Here I recovered *Timmia* (Timmiidae, which has a unique peristome teeth morphology) with strong support as the sister group of the clade comprising Bryidae (diplolepideous-alternate peristomes) and Dicranidae (haplolepideous peristomes). The diplolepideous-opposite Funariidae was well supported as the sister group of the other three lineages. Relationships within Dicranidae and Funariidae were also generally resolved here, with improved support. However, within Bryidae, substantial ambiguities remain especially concerning the grade of lineages that subtends the core pleurocarps. Relationships between the two nematodontous lineages of mosses (classes

Polytrichopsida and Tetraphidopsida) relative to the arthrodontous Bryopsida, also eluded satisfactory resolution here.

The two most species-rich lineages of nonperistomate mosses are the peat mosses, Sphagnopsida, and the granite mosses, Andreaeaceae. Phylogenetic relationships within the former clade have been relatively well studied (e.g., Cronberg, 1996; Natcheva, 2007; Shaw et al., 2010a), but the latter are more poorly known. I characterized phylogenetic relationships within Andreaeaceae using two complementary taxon and plastid gene samplings (17 vs. 6 plastid loci). The 17-gene data set included a handful of exemplar taxa from the major lineages of Andreaeaceae (as defined by the 6-region analysis), and confirmed a sister-group relationship of the family to the monotypic family Andreaeobryaceae with moderate to strong ML bootstrap support. It also resolved the root split in the family with strong support. The more densely sampled 6-locus analysis included nearly all costate species and a broad selection of ecostate species. The current subgeneric taxonomy of Andreaeaceae, and the utility of the characters used to support them (e.g., the presence or absence of costa on leaves) were rejected, as I identified several major lineages that each includes species from different subgeneric units. The monophyly of each lineage and the relationships among them were strongly supported, indicating that a substantial revision of the genus above the species level is needed, and could be based on the clades identified here if these can be confirmed by other data sources. In addition, a subset of species were recovered as non-monophyletic according to the plastid data, with weak to moderate support (e.g., *Andreaea nivalis* and *A. rupestris*), implying that there may have been cryptic speciation in at least these lineages.

6.2 Future directions

The availability of large-scale genomic data sets promises to transform systematic and evolutionary research in bryology, permitting fairly ready retrieval of whole plastid genomes or transcriptome-based data sets, for example (Cronn et al. 2008; Wodniak et al. 2011). However, there is still a place for large-scale inferences of bryophyte phylogeny using more focused collections of plastid genes, not least to cleanly connect phylogenetic inference in these early branches of land-plant phylogeny to the extensive plastid-based inferences that have already been made in vascular plants (e.g., APG III 2009), and to permit retrieval of minimal DNA sequence data from fragmentary or degraded museum-based samples of poorly collected clades. It is not clear that sampling whole plastid genomes offers significant advantages to the smaller plastid gene set considered here, as the latter alone appear to permit inference of a well-supported phylogenetic backbone; nonetheless, some of the more poorly supported branches identified here would likely benefit from expansion of the plastid gene set used in phylogenetic inference.

The orthology of plastid-based genes is unambiguous (Olmstead and Palmer, 1994): the same cannot be said for nuclear genes, although reconciling gene tree conflicts among different linkage groups (plastid, mitochondrial and especially nuclear) can constitute an additional sources of phylogenetic information (e.g., Ness et al., in press). On the other hand, recovering the same clades using different linkage groups is a powerful tool for systematic inference, by permitting cross-validation (corroboration) of results obtained from individual genomes. It is therefore important to obtain a broadly sampled, well-supported plastid-based phylogenetic estimate as a starting point for comparison. This had not been done before for many of the

lineages investigated here. But clearly, more intensive taxon sampling would also be useful, particularly in species-rich clades like Bryidae that still have poor resolution of many early, key evolutionary splits. Continuing the taxon sampling started here in peristomate mosses for the regions surveyed here would be valuable, and should be relatively straightforward to do using 'next-generation' sequencing technology (e.g., by multiplexing sequencing products from different species for the gene set considered here).

The surprising result obtained here for *Takakia*, where different method of phylogenetic analysis (or even different ways of estimating basic parameters) led to conflicting phylogenetic estimates, underlines that we should remain vigilant for phylogenetic mis-inference. Given the great age of some the clades investigated here, it is perhaps surprising that more conflicts were not uncovered. However, it should also be borne in mind that lack of conflict in different analysis is no guarantee that any analysis is free from systematic bias.

Finally, clearly considerable additional work is needed for Andreaeaceae, one of a small handful of evolutionarily isolated lineages of nonperistomate mosses that collectively define the first branches of moss phylogeny. It is straightforward to identify what is needed next for molecular phylogenetic study of this family: more taxa/samples (particularly so for the ecostate species) and more genes (especially additional linkage groups). It will be important to link this work back to the growing body of work being developed using classical systematic methods that focus on morphology (e.g., Murray 1988b and 2006) to refine our understanding of the family's classification. But the deep divergences and lack of monophyly observed for some individual species underlines that future research in Andreaeaceae needs to blend the toolkits of classical

and molecular methods of moss systematics, and the insights that it is possible to gain from them.

REFERENCES

AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716-723.

ALLEN, B. H. 1987. A systematic account of *Pulchrinodus inflatus* (Musci: Pterobryaceae), genus novum. *New Zealand Journal of Botany* 25: 335-342.

ANGIOSPERM PHYLOGENY GROUP. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105-121.

BAUM, D. A., AND M. J. DONOGHUE. 1995. Choosing among alternative "phylogenetic" species concepts. *Systematic Botany* 20: 560-573.

BECKER, B., AND B. MARIN. 2009. Streptophyte algae and the origin of embryophytes. *Annals of Botany* 103: 999-1004.

BECKERT, S., H. MUHLE, D. PRUCHNER, AND V. KNOOP. 2001. The mitochondrial *nad*2 gene as a novel marker locus for phylogenetic analysis of early land plants: a comparative analysis in mosses. *Molecular Phylogenetics and Evolution* 18: 117-126.

BELL, N. E., AND J. HYVONEN. 2008. Rooting the Polytrichopsida: The phylogenetic position of Atrichopsis and the independent origin of the polytrichopsid peristome. *In* H. Mohamed, B. B. Baki, A. Nasrulhaq-Boyce, and P. K. Y. Lee [eds.], Bryology in the new millenium, 227 – 239. University of Malaya, Kuala Lumpur, Malaysia.

BELL, N. E., AND J. HYVONEN. 2010. A phylogenetic circumscription of Polytrichastrum (Polytrichaceae): Reassessment of sporophyte morphology supports molecular phylogeny. *American Journal of Botany* 97: 566-578.

BELL, N. E., D. QUANDT, T. J. O'BRIEN, AND A. E. NEWTON. 2007. Taxonomy and phylogeny in the earliest diverging pleurocarps: square holes and bifurcating pegs. *The Bryologist* 110: 533-560.

BERGTHORSSON, U., A. O. RICHARDSON, G. J. YOUNG, L. R. GOERTZEN, AND J. D. PALMER. 2004. Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. *Proceedings of the National Academy of Sciences USA* 101: 17747-17752.

BRAITHWAITE, R. 1880-1887. The British Moss Flora. Vol. I. London.

BUDKE, J. M., C. S. JONES, AND B. GOFFINET. 2007. Development of the enigmatic peristome of *Timmia megapolitana* (Timmiaceae; Bryophyta). *American Journal of Botany* 94: 460-467.

BURLEIGH, J. G., AND S. MATHEWS. 2004. Phylogenetic signal in nucleotide data from seed plants: implications for resolving the seed plant tree of life. *American Journal of Botany* 91: 1599-1613.

CHANG, J. T. 1996. Inconsistency of evolutionary tree topology reconstruction methods when substitution rates vary across characters. *Mathematical Biosciences* 134: 189-215.

CHANG, Y., C. W. KHIONG, S. W. GRAHAM, AND B. C. TAN. 2006. Molecular evidence for the systematic positions of two enigmatic mosses: *Pterogonidium pulchellum*

(Sematophyllaceae, Musci) and *Piloecium pseudorufescens* (Myuriaceae, Musci). *Canadian Journal of Botany* 84: 501-507.

CHASE, M. W., D. E. SOLTIS, R. G. OLMSTEAD, D. MORGAN, D. H. LES, B. D. MISHLER, M. R. DUVALL, R. A. PRICE, H. G. HILLS, Y.-L. QIU, K. A. KRON, J. H. RETTIG, E. CONTI, J. D. PALMER, J. R. MANHART, K. J. SYTSMA, H. J. MICHAELS, W. J. KRESS, K. G. KAROL, W. D. CLARK, M. HEDREN, S. G. BRANDON, R. K. JANSEN, K.-J. KIM, C. F. WIMPEE, J. F. SMITH, G. R. FURNIER, S. H. STRAUSS, Q.-Y. XIANG, G. M. PLUNKETT, P. S. SOLTIS, S. M. SWENSEN, S. E. WILLIAMS, P. A. GADEK, C. J. QUINN, L. E. EGUIARTE, E. GOLENBERG, G. H. LEARN, JR., S. W. GRAHAM, S. C. H. BARRETT, S. DAYANANDAN, AND V. A. ALBERT. 1993. Phylogenetics of seed plants: an analysis of nucleotide sequences from the plastid gene *rbc*L. *Annals of the Missouri Botanical Garden* 80: 528-580.

CHASE, M. W., R. S. COVWAN, P. M. HOLLINGSWORTH, C. VAN DE BERG, AND E. AL. 2007. A proposal for a standardised protocol to barcode all land plants. *Taxon* 56: 295-299.

COX, C. J., B. GOFFINET, A. J. SHAW, AND S. B. BOLES. 2004. Phylogenetic relationships among the mosses based on heterogeneous Bayesian analysis of multiple genes from multiple genomic compartments. *Systematic Botany* 29: 234-250.

COX, C. J., B. GOFFINET, N. WICKETT, S. B. BOLES, AND A. J. SHAW. 2010. Moss diversity: A molecular phylogenetic analysis of genera. *Phytotaxa* 9: 175-195.

CRANDALL-STOTLER, B. 1980. Morphogenetic designs and a theory of bryophyte origins and divergence. *BioScience* 30: 580-585.

CRANDALL-STOTLER, B., R. E. STOTLER, AND D. G. LONG. 2009. Morphology and classificaiton of the Marchantiophyta. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology, 1-54. Cambridge University Press, New York.

CRONBERG, N. 1996. Isozyme evidence of relationships withinSphagnum sect.Acutifolia (Sphagnaceae, Bryophyta). *Plant Systematics and Evolution* 203: 41-64.

CRONN, R, A. LISTON, M. PARKS, D. S. GERNANDT, R. SHEN, AND R MOCKLER. 2008. Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. *Nucleic Acids Research* 36: e122.

CRUM, H. A., AND L. E. ANDERSON. 1981. Mosses of Eastern North America. Columbia University Press, New York.

CUMMINGS, M. P., S. A. HANDLEY, D. S. MYERS, D. L. REED, A. ROKAS, AND K. WINKA. 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology* 52: 477-487.

DAVIS, C. 2004. A molecular phylogeny of leafy liverworts (Jungermanniidae, Marchantiophyta). *In* B. Goffinet, V. Hollowell, and R. Magill [eds.], Molecular Systematics of Bryophytes, 61-86. Missouri Botanical Garden Press, St. Louis.

DIXON, H. N. 1932. Classification of mosses. *In* F. Verdoorn [ed.], Manual of Bryology, 397-412. Martius Nijhoff, The Hague.

DOYLE, J. J., AND J. L DOYLE. 1987. A rapid isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

DUFF, R., D. C. CARGILL, J. C. VILLARREAL A., AND K. RENZAGLIA. 2004. Phylogenetic relationships of the hornworts based on *rbc*L sequence data: novel relationships and new

insights. *In* B. Goffinet, V. Hollowell, and R. Magill [eds.], Molecular Systematics of Bryophytes, 41-58. Missouri Botanical Garden Press, St. Louis.

DUFF, R. J., J. C. VILLARREAL, D. C. CARGILL, AND K. S. RENZAGLIA. 2007. Progress and challenges toward developing a phylogeny and classification of the hornworts. *The Bryologist* 110: 214-243.

EDWARDS, S. R. 1979. Taxonomic implications of cell patterns in haplolepidous moss peristomes. *In* G. C. S. Clarke and J. G. Duckett [eds.], Bryophyte Systematics. Academic Press, New York.

FAZEKAS, A. J., K. S. BURGESS, P. R. KESANAKURTI, S. W. GRAHAM, S. G. NEWMASTER, B. C. HUSBAND, D. M. PERCY, M. HAJIBABAEI, AND S. C. H. BARRETT. 2008. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* 3: e2802.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401-410.

_____. 1983. Parsimony in systematics: biological and statistical issues. *Annual Review of Ecology and Systematics* 14: 313-333.

_____. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783-791.

FERNANDEZ, C. C., J. R. SHEVOCK, A. N. GLAZER, AND J. N. THOMPSON. 2006. Cryptic species within the cosmopolitan desiccation-tolerant moss *Grimmia laevigata*. *Proceedings of the National Academy of Sciences of the United States of America* 103: 637-642.

FINET, C., R. E. TIMME, C. F. DELWICHE, AND F. MARLÉTAZ. 2010. Multigene phylogeny of the green lineage reveals the origin and diversification of land plants. *Current Biology* 20: 2217-2222.

FLEISCHER, M. 1904-1923. Die Musci der Flora von Buitenzorg (Zugleich Laubmoosflora von Java). Vols. 1-4. E. J. Brill, Leiden.

FORREST, L. L., E. C. DAVIS, D. G. LONG, B. J. CRANDALL-STOTLER, A. CLARK, AND M. L. HOLLINGSWORTH. 2006. Unraveling the evolutionary history of the liverworts (Marchantiophyta): multiple taxa, genomes and analyses. *The Bryologist* 109: 303-334.

FREY, W. AND M. STECH. 2009. Bryophyta (Musci, Mosses). *In* W. Frey (ed.). Syllabus of Plant Families A. Engler's Syllabus der Pflanzenfamilien. Part 3. Bryophytes and Seedless Vascular Plants. 13th Edition, 116-257. Gebr. Borntraeger Verlagsbuchhandlung, Stuttgart, Germany.

GARBARY, D. J., AND K. S. RENZAGLIA. 1998. Bryophyte phylogeny and the evolution of land plants: evidence from development and ultrastructure. *In* J. W. Bates, N. W. Ashton, and J. G. Duckett [eds.], Bryology for the twenty-first century, 45-63. The British Bryological Society and Maney Publishing, Leeds.

GOFFINET, B., AND W. R. BUCK. 2004. Systematics of the Bryophyta (mosses) from molecules to a revised classification. *In* B. Goffinet, V. Hollowell, and R. Magill [eds.], Molecular Systematics of Bryophytes, 204-239. Missouri Botanical Garden Press, St. Louis.

GOFFINET, B., C. J. COX, A. J. SHAW, AND T. A. J. HEDDERSON. 2001. The Bryophyta (Mosses): systematic and evolutionary inferences from an *rps*4 gene (cpDNA) phylogeny. *Annals of Botany* 87: 191-208.

GOFFINET, B., N. J. WICKETT, O. WERNER, R. M. ROS, A. J. SHAW, AND C. J. COX. 2007. Distribution and phylogenetic significance of the 71-kb inversion in the plastid genome in Funariidae (Bryophyta). *Annals of Botany* 99: 747-753.

GOFFINET, B., W. R. BUCK, AND A. J. SHAW. 2009. Morphology, anatomy, and classification of the Bryophyta. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology, 55-138. Cambridge University Press, New York.

GOREMYKIN, V. V., AND F. H. HELLWIG. 2005. Evidence for the most basal split in land plants dividing bryophyte and tracheophyte lineages. *Plant Systematics and Evolution* 254: 93-103.

GOLDMAN, N., J. P. ANDERSON, AND A. G. RODRIGO. 2000. Likelihood-based test of topologies in phylogenetics. *Systematic Biology* 49: 652-670.

GRAHAM, S. W., AND W. J. D. ILES. 2009. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. *American Journal of Botany* 96: 216-227.

GRAHAM, S. W., AND R. G. OLMSTEAD. 2000a. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87: 1712-1730.

GRAHAM, S. W., AND R. G. OLMSTEAD. 2000b. Evolutionary significance of an unusual chloroplast DNA inversion found in two basal angiosperm lineages. *Current Genetics* 37: 183–188.

GRAHAM, S. W., J. R. KOHN, B. R. MORTON, J. E. ECKENWALDER, AND S. C. H. BARRETT. 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Systematic Biology* 47: 545-567.

GRAHAM, S. W., P. A. REEVES, A. C. E. BURNS, AND R. G. OLMSTEAD. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution, and utility of indels and inversions in basal angiosperm phylogenetic inference. *International Journal of Plant Sciences* 161: S83-S96.

GRAHAM, S. W., R. G. OLMSTEAD, AND S. C. H. BARRETT. 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. *Molecular Biology and Evolution* 19: 1769-1781.

GRAHAM, S. W., J. M. ZGURSKI, M. A. MCPHERSON, D. M. CHERNIAWSKY, J. M. SAARELA, E. F. C. HORNE, S. Y. SMITH ET AL. 2006. Robust inference of monocot deep phylogeny using an expanded multigene plastid data set. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution (excluding Poales). Rancho Santa Ana Botanic Garden, Claremont, California, USA.

GUINDON, S. P., AND O. GASCUEL. 2003. A Simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* 52: 696-704.

HEDDERSON, T. A., D. J. MURRAY, C. J. COX, AND T. L. NOWELL. 2004. Phylogenetic relationships of haplolepideous mosses (Dicranidae) inferred from *rps*4 gene sequences. *Systematic Botany* 29: 29-41.

HENDY, M.D., AND D. PENNY. 1989. A framework for the quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

HE-NYGRÉN, X., A. JUSLÉN, I. AHONEN, D. GLENNY, AND SINIKKA PIIPPO. 2006. Illuminating the evolutionary history of liverworts (Marchantiophyta): towards a natural classification. *Cladistics* 22: 1-31.

HILLIS, D. M. 1995. Approaches for assessing phylogenetic accuracy. *Systematic Biology* 55: 44: 3-16.

HUELSENBECK, J. P. 1995. Performance of Phylogenetic Methods in Simulation. *Systematic Biology* 44: 17-48.

_____. 1997. Is the Felsenstein zone a fly trap? *Systematic Biology* 46: 69-74.

_____. 1998. Systematic bias in phylogenetic analysis: Is the Strepsiptera problem solved? *Systematic Biology* 47: 519-537.

KELCH, D. G., A. DRISKELL, AND B. D. MISHLER. 2004. Inferring phylogeny using genomic characters: a case study using land plant plastomes. *In* B. Goffinet, V. Hollowell, and R. Magill [eds.], Molecular systematics of bryophytes, 3-11. Missouri Botanical Garden Press, St. Louis.

KELCHNER, S. A. 2000. The evolution of non-coding chloroplast DNA and its application in plant systematics. *Annals of the Missouri Botanical Garden* 87: 482-498.

KELLOGG, E. A., AND J. L. BENNETZEN. 2004. The evolution of nuclear genome structure in seed plants. *American Journal of Botany.* 91: 1709-1725.

KENRICK, P., AND P. R. CRANE. 1997. The origin and early diversification of land plants: a cladistic study. Smithsonian Press, Washington, D.C., USA.

KOLACZKOWSKI, B., AND J. W. THORNTON. 2007. Effects of branch length uncertainty on Bayesian posterior probabilities for phylogenetic hypotheses. *Molecular Biology and Evolution* 24: 2108-2118.

LA FARGE, C., A. J. SHAW, AND D. H. VITT. 2002. The circumscription of the Dicranaceae (Bryopsida) based on the chloroplast regions *trn*L-*trn*F and *rps*4. *Systematic Botany* 27: 435-452.

LA FARGE, C., B. D. MISHLER, J. A. WHEELER, D. P. WALL, K. JOHANNES, S. SCHAFFER, AND A. J. SHAW. 2000. Phylogenetic relationships within the haplolepideous mosses. *The Bryologist* 103: 257-276.

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

LIGRONE, R., AND J. G. DUCKETT. 1998. Development of the leafy shoot in *Sphagnum* (Bryophyta) involves the activity of both apical and subapical meristems. *New Phytologist* 140: 581-595.

MAGOMBO, Z. L. K. 2003. The phylogeny of basal peristomate mosses: evidence from cpDNA, and implications for peristome evolution. *Systematic Botany* 28: 24-38.

MILLER, M. A., M. T. HOLDER, R. VOS, P. E. MIDFORD, T. LIEBOWITZ, L. CHAN, P. HOOVER, AND T. WARNOW 2009. The CIPRES Portals. URL: http://www.phylo.org/sub_sections/portal. (Archived by WebCite(r) at http://www.webcitation.org/5imQlJeQa)

MISHLER, B. D., AND S. P. CHURCHILL. 1984. A cladistic approach to the phylogeny of the "bryophytes". *Brittonia* 36: 406-424.

_____. 1985. Transistion to a land flora: phylogenetic relationships of the green algae and bryophytes *Cladistics* 1: 305-328.

MIYATA, Y., AND M. SUGITA. 2004. Tissue- and stage-specific RNA editing of *rps*14 transcripts in moss (*Physcomitrella patens*) chloroplasts. *Journal of Plant Physiology* 161: 113-115.

MURRAY, B. M. 1987a. *Andreaea schofieldiana* and *A. megistospora*, species novae, and taxonomic criteria for sect. Nerviae (Andreaeopsida). *The Bryologist* 90: 15-26.

_____. 1987b. Illustrated moss flora of arctic North America and Greenland. *Meddelelser om Groønland, Bioscience* 23: 15-26.

_____. 1988a. Systematics of the Andreaeopsida (Bryophyta): two orders with links to *Takakia*. *Beihefte zur Nova Hedwigia* 90: 289-336.

_____. 1988b. The genus *Andreaea* in Britain and Ireland. *Journal of Bryology* 15: 17-82.

_____. 2006. Andreaeaceae. *Flora of Australia* 51: 108-123, 390-391, 406-409.

NATCHEVA, R., AND N. CRONBERG. 2007. Recombination and introgression of nuclear and chloroplast genomes between the peat mosses, *Sphagnum capillifolium* and *Sphagnum quinquefarium*. *Molecular Ecology* 16: 811-818.

NESS, R. W., S. W. GRAHAM, AND S. C. H. BARRETT. In press. Reconciling gene and genome duplicaiton events: using multiple nuclear gene families to infer the phylogeny of the aquatic plant family Pontederiaceae. *Molecular Biology and Evolution*.

NEWTON, A. E., C. J. COX, J. G. DUCKETT, J. A. WHEELER, B. GOFFINET, T. A. J. HEDDERSON, AND B. D. MISHLER. 2000. Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. *The Bryologist* 103: 187-211.

NEWTON, A. E., C. J. COX, J. G. DUCKETT, J. A. WHEELER, B. GOFFINET, T. A. J. HEDDERSON, AND B. D. MISHLER. 2000. Evolution of the major moss lineages: phylogenetic analyses based on multiple gene sequences and morphology. *The Bryologist* 103: 187-211.

NEWTON, A. E., N. WILKSTRÖM, AND A. J. SHAW. 2009. Mosses (Bryophyta). *In* S. B. Hedges and S. Kumar [eds.], The Timetree of Life, 138-145. Oxford University Press, Oxford, UK.

NICKRENT, D. L., C. L. PARKINSON, J. D. PALMER, AND R. J. DUFF. 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., P. G. WOLF, M. KUGITA, R. B. SINCLAIR, M. SUGITA, C. SUGIURA, T. WAKASUGI, K. YAMADA, K. YOSHINAGA, K. YAMAGUCHI, K. UEDA, AND M. HASEBE. 2004. Chloroplast phylogeny indicates that bryophytes are monophyletic. *Molecular Biology and Evolution* 21: 1813-1819.

O'BRIEN, T. J. 2007. The phylogenetic distribution of pleurocarpous mosses: evidence from cpDNA sequences. *In* A. E. Newton and R. S. Tangney [eds.], Pleurocarpous mosses: systematics and evolution, 41–64. CRC Press, Boca Raton.

OCHYRA, R., J. ZARNOWIEC, AND H. BEDNAREK-OCHYRA. 2003. Census catalogue of Polish mosses. *Biodiversity of Poland* 3: 1-372.

OLMSTEAD, R. G. 1995. Species concepts and plesiomorphic species. *Systematic Botany* 20: 623-630.

OLMSTEAD, R. G., AND J. D. PALMER. 1994. Chloroplast DNA systematics: A review of methods and data analysis. *American Journal of Botany* 81: 1205–1224.

OLMSTEAD, R. G., AND J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Systematic Biology* 43: 467-481.

OLSSON, S., V. BUCHBENDER, J. ENROTH, L. HEDEN√§S, S. HUTTUNEN, AND D. QUANDT. 2009. Phylogenetic analyses reveal high levels of polyphyly among pleurocarpous lineages as well as novel clades. *The Bryologist* 112: 447-466.

PALMER, J. D., K. L. ADAMS, Y. CHO, C. L. PARKINSON, Y.-L. QIU, AND K. SONG. 2000. Dynamic evolution of plant mitochondrial genomes: Mobile genes and introns and highly variable mutation rates. *Proceedings of the National Academy of Sciences USA* 97: 6960-6966.

PENNY, D., L. R. FOULDS, AND M. D. HENDY. 1982. Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. *Nature* 297: 197-200.

POND, K. S. L., S. D. W. FROST, AND S. V. MUSE. 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21: 676-679.

POSADA, D. 2008. jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253-1256.

QIU, Y.-L. 2008. Phylogeny and evolution of charophytic algae and land plants. *Journal of Systematics and Evolution* 46: 287-306.

QIU, Y.-L., Y. CHO, J. C. COX, AND J. D. PALMER. 1998. The gain of three mitochondrial introns identifies liverworts as the earliest land plants. *Nature* 394: 671-674.

QIU, Y.-L., L. LI, B. WANG, Z. CHEN, V. KNOOP, M. GROTH-MALONEK, O. DOMBROVSKA, J. LEE, L. KENT, J. REST, G. F. ESTABROOK, T. A. HENDRY, D. W. TAYLOR, C. M. TESTA, M. AMBROS, B. CRANDALL-STOTLER, R. J. DUFF, M. STECH, W. FREY, D. QUANDT, AND

C. C. DAVIS. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. *Proceedings of the National Academy of Sciences* 103: 15511-15516.

QIU, Y.-L., L. LI, B. WANG, Z. CHEN, O. DOMBROVSKA, J. LEE, L. KENT, R. LI, R. W. JOBSON, T. A. HENDRY, D. W. TAYLOR, C. M. TESTA, AND M. AMBROS. 2007. A nonflowering land plant phylogeny inferred from nucleotide sequences of seven chloroplast, mitochondrial, and nuclear genes. *International Journal of Plant Sciences* 168: 691-708.

QIU, Y.-L., L. LI, B. WANG, J.-Y. XUE, T. A. HENDRY, R.-Q. LI, J. W. BROWN, Y. LIU, G. T. HUDSON, AND Z.-D. CHEN. 2010. Angiosperm phylogeny inferred from sequences of four mitochondrial genes. *Journal of Systematics and Evolution* 48: 391-425.

QUANDT, D., AND M. STECH. 2004. Molecular evolution of the *trn*TUGU-trnFGAA region in bryophytes. *Plant Biology* 6: 545-554.

QUANDT, D., K. MÜLLER, AND S. HUTTUNEN. 2003. Characterisation of the chloroplast DNA *psb*T-H region and the influence of dyad symmetrical elements on phylogenetic reconstructions. *Plant Biology* 5: 400-410.

RAI, H. S., AND S. W. GRAHAM. 2010. Utility of a large, multigene plastid data set in inferring higher-order relationships in ferns and relatives (monilophytes). *American Journal of Botany* 97: 1444-1456.

RAI, H. S., H. E. O'BRIEN, P. A. REEVES, R. G. OLMSTEAD, AND S. W. GRAHAM. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350-359.

RAI, H. S., P. A. REEVES, R. PEAKAL, R. G. OLMSTEAD, AND S. W. GRAHAM. 2008. Inference of higher-order conifer relationships from a multi-locus plastid data set. *Botany* 86: 658-669.

RAMBAUT, A. 2002. Se-Al sequence alignment editor, ver. 2.0a11. Oxford: University of Oxford.

RENZAGLIA, K. S., R. JOEL DUFF, D. L. NICKRENT, AND D. J. GARBARY. 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.

RENZAGLIA, K. S., S. SCHUETTE, R. J. DUFF, R. LIGRONE, A. J. SHAW, B. D. MISHLER, AND J. G. DUCKETT. 2007. Bryophyte phylogeny: Advancing the molecular and morphological frontiers. *The Bryologist* 110: 179-213.

RENZAGLIA, K., J. C. VILLARREAL A., AND R. DUFF. 2009. New insights into morphology, anatomy, and systematics of hornworts. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology, 139-171. Cambridge University Press, New York.

ROTH, G. 1903-1904. Die europäischen Laubmoose, Vol. 1. Leipzig.

ROTH, G. 1910-1911. Die aussereuropäischen Laubmoose, Vol. 1. Dresden.

RYDIN, H. 2009. Population and community ecology. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology, 393-444. Cambridge University Press, New York, New York, USA.

SAARELA, J. M., H. S. RAI, J. A. DOYLE, P. K. ENDRESS, S. MATHEWS, A. D. MARCHANT, B. G. BRIGGS, AND S. W. GRAHAM. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. *Nature* 446: 312-315.

SAINSBURY, G. O. K. 1955. A Handbook of the New Zealand Mosses. *The Royal Society of New Zealand Bulletin* 5: 1-490.

SCHOFIELD, W. B. 1985. Introduction to bryology. Macmillan, New York, USA.

SCHULTZE-MOTEL, W. 1970. Monographie der Laubmoosgattung *Andreaea*. 1. Die costaten Arten. *Willdenowia* 6: 25-110.

SCHWARZ, G. E. 1978. Estimating the dimension of a model. *Annals of Statistics* 6: 461-464.

SHAW, A. J. 2009. Bryophyte species and speciation. *In* B. Goffinet and A. J. Shaw [eds.], Bryophyte Biology, 445-485. Cambridge University Press, New York, New York, USA.

SHAW, J., AND L. E. ANDERSON. 1988. Peristome development in mosses in relation to systematics and evolution. II. *Tetraphis pellucida* (Tetraphidaceae). *American Journal of Botany* 75: 1019-1032.

SHAW, J., AND K. RENZAGLIA. 2004. Phylogeny and diversification of bryophytes. *American Journal of Botany* 91: 1557-1581.

SHAW, A. J., C. J. COX, B. GOFFINET, W. R. BUCK, AND S. B. BOLES. 2003. Phylogenetic evidence of a rapid radiation of pleurocarpous mosses (Bryophyta) *Evolution* 57: 2226-2241.

SHAW, A. J., N. DEVOS, C. J. COX, S. B. BOLES, B. SHAW, A. M. BUCHANAN, L. CAVE, AND R. SEPPELT. 2010a. Peatmoss (Sphagnum) diversification associated with Miocene Northern Hemisphere climatic cooling? *Molecular Phylogenetics and Evolution* 55: 1139-1145.

SHAW, A. J., C. J. COX, W. R. BUCK, N. DEVOS, A. M. BUCHANAN, L. CAVE, R. SEPPELT, ET AL. 2010b. Newly resolved relationships in an early land plant lineage: Bryophyta class Sphagnopsida (peat mosses). *American Journal of Botany* 97: 1511-1531.

SHIMODAIRA, H., AND M. HASEGAWA. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. *Bioinformatics* 17: 1246-1247.

SHIMODAIRA, H. 2002. An approximately unbiased test of phylogenetic tree selection. *Systematic Biology* 51: 492-508.

SIMMONS M. P., AND H. OCHOTERENA. 2000. Gaps as characters in sequence-based phylogenetic analyses. *Systematic Biology* 49: 369–381.

SIMMONS, M. P., K. M. PICKETT, AND M. MIYA. 2004. How meaningful are Bayesian support values? *Molecular Biology and Evolution* 21: 188-199.

SMITH, S. A., J. M. BEAULIEU, AND M. J. DONOGHUE. 2010. An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proceedings of the National Academy of Sciences* 107: 5897-5902.

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22: 2688-2690.

STAMATAKIS, A., P. HOOVER, AND J. ROUGEMONT. 2008. A fast bootstrapping algorithm for the RAxML web-servers. *Systematic Biology* 57: 758-771.

STECH, M., AND D. QUANDT. 2010. 20,000 species and five key markers: The status of molecular bryophyte phylogenetics. *Phytotaxa* 9: 196-228.

STECH, M., D. QUANDT, AND W. FREY. 2003. Molecular circumscription of the hornworts (Anthocerotophyta) based on the chloroplast DNA *trn*L–*trn*F region. *Journal of Plant Research* 116: 389-398.

STEERE, W. C. 1969. A new look at evolutionand phylogeny in bryophytes. *In* J. E. Gunkel [ed.], *Current Topics in Plant Science*, 134-143. Academic Press, New York, USA.

SUGIURA, C., Y. KOBAYASHI, S. AOKI, C. SUGITA, AND M. SUGITA. 2003. Complete chloroplast DNA sequence of the moss *Physcomitrella patens*: evidence for the loss and relocation of *rpo*A from the chloroplast to the nucleus. *Nucleic Acids Research* 31: 5324-5331.

SULLIVAN, J., AND D. L. SWOFFORD. 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated? *Systematic Biology* 50: 723-729.

SUNDBERG, K., H. CARROLL, Q. SNELL, AND M. CLEMENT. 2008. Incomparability of results between phylogenetic search programs. *International Conference on Bioinformatics and Computational Biology*, pp 81-84.

SUSKO, E. 2008. On the distributions of bootstrap support and posterior distributions for a star tree. *Systematic Biology* 57: 602-612.

SUZUKI, Y., G. V. GLAZKO, AND M. NEI. 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proceedings of the National Academy of Sciences USA* 99: 16138-16143.

SWOFFORD, D. L. 2002. PAUP*-Phylognetic Analyses Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland.

SWOFFORD, D. L., P. J. WADDELL, J. P. HUELSENBECK, P. G. FOSTER, P. O. LEWIS, AND J. S. ROGERS. 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. *Systematic Biology* 50: 525-539.

THIERS, B. 2010 [continuously updated]. Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden's Virtual Herbarium. http://sweetgum.nybg.org/ih/

TSUBOTA, H., Y. AGENO, B. ESTEBANEZ, T. YAMAGUCHI AND H. DEGUCHI. 2003. Molecular phylogeny of Grimmiales (Musci) based on chloroplast *rbc*L sequences. *Hikobia* 14: 55-70.

TURMEL, M., C. OTIS, AND C. LEMIEUX. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. *Molecular Biololgy and Evolution* 23: 1324-1338.

TURMEL, M., J.-F. POMBERT, P. CHARLEBOIS, C. OTIS, AND C. LEMIEUX. 2007. The Green algal ancestry of land plants as revealed by the chloroplast genome. *International Journal of Plant Sciences* 168: 679-689.

VANDERPOORTEN, A., AND B. GOFFINET. 2006. Mapping uncertainty and phylogenetic uncertainty in ancestral character state reconstruction: an example in the moss genus *Brachytheciastrum*. *Systematic Biology* 55: 957-971.

VANDERPOORTEN, A., AND B. GOFFINET. 2009. Introduction to Bryophytes. Cambridge University Press, Cambridge, UK.

VANDERPOORTEN, A., L. HEDENÅS, C. J. COX, AND A. J. SHAW. 2002. Phylogeny and morphological evolution of the Amblystegiaceae (Bryopsida). *Molecular Phylogenetics and Evolution* 23: 1-21.

VITT, D. H. 1980. A comparative study of *Andreaea acutifolia, A. mutabilis,* and *A. rupestris*. *New Zealand Journal of Botany* 18: 367-377.

_____. 1984. Classification of the Bryopsida. *In* R. M. Schuster [ed.], New Manual of Bryology, 696-759. Hattori Botanical Laboratory, Nichinan, Japan.

_____. 2006. Functional characteristics and indicators of boreal peatlands. *In* R. K. Wieder and D. H. Vitt [eds.], Boreal Peatland Ecology, 9-24. Springer-Verlag., Berlin.

VITT, D. H., B. GOFFINET, AND T. A. HEDDERSON. 1998. The ordinal classification of the mosses: Questions and answers for the 1990's. *In* N. J. W. Bates, W. Ashton, and J. G. Duckett [eds.], Bryology for the Twenty-first Century, 143-159. Maney Publishing and British Bryological Society, Leeds, U.K.

VOLKMAR, U., AND V. KNOOP. 2010. Introducing intron locus cox1i624 for phylogenetic analyses in bryophytes: on the issue of *Takakia* as sister genus to all other extant mosses. *Journal of Molecular Evolution* 70: 506-518.

WAHRMUND, U., T. REIN, K. MÜLLER, M. GROTH-MALONEK, AND V. KNOOP. 2009. Fifty mosses on five trees: comparing phylogenetic information in three types of non-coding mitochondrial DNA and two chloroplast loci. *Plant Systematics and Evolution* 282: 241-255.

WAHRMUND, U., D. QUANDT, AND V. KNOOP. 2010. The phylogeny of mosses - Addressing open issues with a new mitochondrial locus: Group I intron cobi420. *Molecular Phylogenetics and Evolution* 54: 417-426.

WANG, W., Y. WU, Y. YAN, M. ERMAKOVA, R. KERSTETTER, AND J. MESSING. 2010. DNA barcoding of the Lemnaceae, a family of aquatic monocots. *BMC Plant Biology* 10: 205.

WIJK, R. V. D., W. D. MARGADANT, AND P. A. FLORSCHÜTZ. 1959. Index Muscorum. 1 (A--C).

WILSON, R., S. ROBBERT GRADSTEIN, H. SCHNEIDER, AND J. HEINRICHS. 2007. Unravelling the phylogeny of Lejeuneaceae (Jungermanniopsida): Evidence for four main lineages. *Molecular Phylogenetics and Evolution* 43: 270-282.

WODNIOK, S., H. BRINKMANN, G. GLOCKNER, A. HEIDEL, H. PHILIPPE, M. MELKONIAN, AND B. BECKER. 2011. Origin of land plants: Do conjugating green algae hold the key? *BMC Evolutionary Biology* 11: 104.

WOLF, P. G., K. G. KAROL, D. F. MANDOLI, J. KUEHL, K. ARUMUGANATHAN, M. W. ELLIS, B. D. MISHLER, D. G. KELCH, R. G. OLMSTEAD, AND J. L. BOORE. 2005. The first complete chloroplast genome sequence of a lycophyte, *Huperzia lucidula* (Lycopodiaceae). *Gene* 350: 117-128.

WOLFE, K. H., W. H. LI, AND P. M. SHARP. 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.

YATSENTYUK, S. P. 2001. Molecular phylogeny of the Bryophyta and Lycophyta, according to results from some sequences of chloroplast DNA. Ph.D Dissertation, Moscow State University, Russia.

ZGURSKI, J. M., H. S. RAI, Q. M. FAI, D. J. BOGLER, J. FRANCISCO-ORTEGA, AND S. W. GRAHAM. 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. *Molecular Phylogenetics and Evolution* 47: 1232-1237.

ZWICKL, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. Dissertation, The University of Texas at Austin, Texas, USA.

APPENDICES

Appendix 1: Source information of bryophyte taxa used in primer design and testing in Chapter 2.

Herbarium abbreviations: ABSH, Hepatic Herbarium, Southern Illinois University; CONN, The George Safford Torrey Herbarium; E, Royal Botanic Garden, Edinburgh; UBC, University of British Columbia.

A: Previously published data used in primer development.

Adiantum capillus-veneris L., *Amborella trichopoda* Baill., *Angiopteris evecta* (G. Frost.) Hoffm., *Anthoceros formosae* Steph., *Chaetosphaeridium globosum* (Nordst.) Kleb., *Chara vulgaris* L., *Mesostigma viride* Lauterborn, *Chlorokybus atmophyticus* Geitler, *Ginkgo biloba* L., *Gnetum gnemon* L., *Huperzia lucidula* (Michx.) Trevisan, *Lilium superbum* L., *Magnolia stellata* Maxim., *Marchantia polymorpha* I., *Nicotiana tabacum* L., *Nymphaea odorata* Aiton, *Physcomitrella patens* (Hedw.) Bruch & Schimp, *Pinus koraiensis* Siebold & Zucc., *Podocarpus chinensis* Wall. ex. J. Forbes, *Psilotum nudum* (L.) Beauv., *Sphagnum* L. sp., *Trimenia moorei* (Oliv.) Philipson, *Zygnema circumcarinatum* Czurda.

B: Taxa sequenced during primer development.

Bryophyta: *Andreaea rupestris* Hedw. (Andreaeobryopsida, Andreaeaceae), Y. Chang 06-03, UBC. *Atrichum selwynii* Aust. (Polytrichopsida, Polytrichaceae), Y. Chang 06-01, UBC. *Bryum capillare* Hedw. (Bryopsida, Bryaceae), Y. Chang 07-01, UBC. *Buxbaumia aphylla* Hedw. (Bryopsida, Buxbaumiaceae), W. B. Schofield 126011A, UBC. *Dicranum scoparium* Hedw. (Bryopsida, Dicranaceae), D. Woods 5-04, UBC. *Homalothecium fulgescens* (Mitt. ex C. Muell.) Lawt. (Bryopsida, Brachytheciaceae), D. Woods 05-01, UBC. *Orthotrichum lyellii* Hook. & Tayl. (Bryopsida, Orthotrichaceae), D. Woods 05-03, UBC. *Syntrichia norvegica* F. Weber (Bryopsida, Pottiaceae), Y. Chang 06-08, UBC. *Takakia ceratophylla* (Mitt.) Grolle (Takakiopsida, Takakiaceae), W. B. Schofield 125387, UBC. *Takakia lepidozioides* Hattori & H. Inoue (Takakiopsida, Takakiaceae), Y. Chang 05-02 UBC.

Marchantiophyta: *Metzgeria conjugata* Lindb. (Jungermanniopsida, Metzgeriaceae), Y. Chang 07-02, UBC. *Scapania bolanderi* Aust. (Jungermanniopsida, Scapaniaceae), Y. Chang 06-05, UBC.

C. Taxa used in primer testing (see Table 2.1).

Bryophyta: *Aulacomnium palustre* (Hedw.) Schwaegr. (Bryopsida, Aulacomniaceae), W.B. Schofield 123402, UBC. *Bartramiopsis lescurii* (James) Kindb. (Bryopsida, Bartramiaceae), W.B. Schofield 125406, UBC. *Encalypta ciliata* Hedw. (Bryopsida, Encalyptaceae), W.B. Schofield 117021, UBC. *Oedipodium griffithianum* (Dicks.) Schwaegr. (Oedipodiopsida, Oedipodiaceae), W.B. Schofield 115800, UBC. *Ptychomnion cygnisetum* (C. Mull.) Kindb. (Bryopsida, Ptychomniaceae), Bruce Allen 26391, UBC. *Tetrodontium brownianum* (Dicks.) Schwaegr. (Tetraphidopsida, Tetraphidaceae), D.G. Long 37600, E. *Tetraplodon mnioides* (Hedw.) B.S.G. (Bryopsida, Splachnaceae), W.B. Schofield 121611, UBC.

Marchantiophyta: *Haplomitrium hookeri* (Sm.) Nees (Haplomitriopsida, Haplomitriaceae), D. G. Long 33741, E. *Treubia lacunosa* (Col.) Prosk. (Haplomitriopsida, Treubiaceae), Stotler & Crandall-Stotler 4561, ABSH.

Anthoceratophyta: *Leiosporoceros dussii* (Steph.) Hassel (Leiosporocerotopsida, Leiosporocerotaceae), Villarreal & Rodríguez 852A, CONN. *Phaeoceros carolinianus* (Michx.) Prosk. (Anthocerotopsida, Notothyladaceae), Bernard Goffinet 9652, UBC.

Appendix 2: Source information and GenBank numbers for taxa included in Chapter 3.

A: Data obtained in this study: For each taxa, the information is displayed in the following sequence: species (classification), voucher specimen, herbarium; GenBank accession: *atp*B, *ndh*F, *psb*B-T-N-H, *psb*D-C, *psb*E-F-L-J, *rbc*L, *rpl*2, *rps*12-*rps*7-*ndh*B. When there are two separate accessions for one region, the two accessions are connected by "&". "—" represents missing data. The classification follows Goffinet et al. (2009). Herbarium abbreviations (Thiers, 2010): ABSH: Hepatic Herbarium, Southern Illinois University; ALA, University of Alaska; CONN, University of Connecticut; UBC, University of British Columbia.

Bryophyta: *Andreaea rupestris* Hedw. (Andreaeobryopsida, Andreaeaceae), Y. Chang 06-03, UBC**;** GU295845, GU295930, GU295893, GU295869, GU295857, GU295869, GU295919, GU295905. *Andreaeobryum macrosporum* Steere & B.M.Murray (Andreaeobryopsida, Andreaeobryaceae), O.M. Afonina 153, ALA; HQ412998, HQ413022, HQ413017, HQ413028, HQ412991, HQ413005, HQ413012, HQ412994. *Atrichum selwynii* Aust. (Polytrichopsida, Polytrichaceae), Y. Chang 06-01, UBC; GU295846, GU295931**,** GU295894, GU295870, GU295858, GU295870, GU295920, GU295906 & GU295907. *Bryum capillare* Hedw. (Bryopsida, Bryaceae), Y. Chang 07-01, UBC; GU295850, GU295935, GU295898, GU295874, GU295862, GU295874, GU295924, —. *Buxbaumia aphylla* Hedw. (Bryopsida, Buxbaumiaceae), W. B. Schofield 126011A, UBC**;** GU295847, GU295932, GU295895, GU295871, GU295859, GU295871, GU295921, —. *Dicranum scoparium* Hedw. (Bryopsida, Dicranaceae), D. Woods 5-04, UBC**;** GU295848, GU295933, GU295896, GU295872, GU295860, GU295872, GU295922, GU295908 & GU295909. *Homalothecium fulgescens* (Mitt. ex C. Muell.) Lawt. (Bryopsida, Brachytheciaceae), D. Woods 05-01, UBC; GU295852, GU295937, GU295900, GU295876, GU295864, GU295876, GU295926, GU295914. *Oedipodium griffithianum* (Dicks.) Schwaegr. (Oedipodiopsida, Oedipodiaceae), W.B. Schofield 115800, UBC; HQ412999, HQ413023, HQ413018, HQ413029, HQ412992, HQ413006, HQ413013, —. *Orthotrichum lyellii* Hook. & Tayl. (Bryopsida, Orthotrichaceae), D. Woods 05-03, UBC; GU295851, GU295936, GU295899, GU295875, GU295863, GU295875, GU295925, GU295912 & GU295913. *Syntrichia norvegica* F. Weber (Bryopsida, Pottiaceae), Y. Chang 06-08, UBC; GU295849, GU295934, GU295897, GU295873, GU295861, GU295873, GU295923, GU295910 & GU295911. *Takakia ceratophylla* (Mitt.) Grolle (Takakiopsida, Takakiaceae), W. B. Schofield 125387, UBC; GU295843, GU295928, GU295891, GU295867, GU295855, GU295867, GU295917, GU295902 & GU295903. *Takakia lepidozioides* Hattori & H. Inoue (Takakiopsida, Takakiaceae), Y. Chang 05-02 UBC; GU295844, GU295929, GU295892, GU295880, GU295856, GU295868, GU295918, GU295904.

Marchantiophyta: *Haplomitrium hookeri* (Sm.) Nees (Haplomitriopsida, Haplomitriaceae), D. G. Long 33741, E; HQ412996, —, HQ413015, HQ413026, —, HQ413003, HQ413010, —. *Metzgeria conjugata* Lindb. (Jungermanniopsida, Metzgeriaceae), Y. Chang 07-02, UBC; GU295842, —, GU295890, GU295866, GU295854, GU295866, GU295916, —. *Pleurozia purpurea* Lindb. (Jungermanniopsida, Pleuroziaceae), W. B. Schofield 125173, UBC; HQ412997, HQ413021, HQ413016, HQ413027, HQ412990, HQ413004, HQ413011, —. *Scapania bolanderi* Aust. (Jungermanniopsida, Scapaniaceae), Y. Chang 06-05, UBC; GU295841, GU295927, GU295889, GU295865, GU295853, GU295865, GU295915, GU295901. *Treubia lacunosa* (Col.) Prosk. (Haplomitriopsida, Treubiaceae), Stotler & Crandall-Stotler 4561, ABSH; HQ412995, —, HQ413014, HQ413025, HQ412989, HQ413002, HQ413009, —.

Anthocerotophyta: *Leiosporoceros dussii* (Steph.) Hassel (Leiosporocerotopsida, Leiosporocerotaceae), Villarreal & Rodríguez 852A CONN; HQ413000, HQ413024, HQ413019, HQ413030, HQ412993, HQ413007, —, —. *Phaeoceros carolinianus* (Michx.) Prosk. (Anthocerotopsida, Notothyladaceae), B. Goffinet 9652, UBC; HQ413001, —, HQ413020, HQ413031, —, HQ413008, —, —.

B: Previously published data:

Plastid genomes: *Adiantum capillus*-veneris L. (NC_004766), *Angiopteris evecta* (G. Frost.) Hoffm. (NC_008829), *Anthoceros formosae* Steph. (NC_004543), *Chaetosphaeridium globosum* (Nordst.) Kleb. (NC_004115), *Chara vulgaris* L. (NC_008097), *Chlorokybus atmophyticus* Geitler (NC_008822), *Mesostigma viride* Lauterborn (NC_002186), *Huperzia lucidula* (Michx.) Trevisan (NC_006861), *Marchantia polymorpha* I. (NC_001319), *Psilotum nudum* (L.) Beauv. (NC_003386), *Physcomitrella patens* (Hedw.) Bruch & Schimp (NC_005087), *Pinus koraiensis* Siebold & Zucc. (NC_004677), *Nicotiana tabacum* L. (NC_006581), and *Zygnema circumcarinatum* Czurda (NC_008117).

Previously published sequences: *Amborella trichopoda* Baill. *Lilium superbum* L., *Ginkgo biloba* L., *Gnetum gnemon* L., *Magnolia stellata* Maxim., *Nymphaea odorata* Aiton, *Podocarpus chinensis* Wall. ex. J. Forbes., *Sphagnum* L. sp., and *Trimenia moorei* (Oliv.) Philipson. The source information and GenBank accession numbers are in Graham & Olmstead (2000a,b), Graham et al. (2002; 2006), Rai et al. (2003) and Rai & Graham (2010).

Appendix 3. Source information and GenBank numbers for taxa included in Chapter 4.

For each taxa newly sequenced, the information is displayed in the following sequence: species (classification), voucher specimen, herbarium; GenBank accession: *atp*B, *ndh*F, *psb*B-T-N-H, *psb*D-C, *psb*E-F-L-J, *rbc*L, *rpl*2, *rps*12-*rps*7-*ndh*B. When there are two separate accessions for one region, the two accessions are connected by "&". "—" represents missing data. The classification follows Goffinet et al. (2009). Herbarium abbreviations (Thiers, 2010): ABSH, Hepatic Herbarium, Southern Illinois University; CONN, University of Connecticut; MO, Missouri Botanical Garden; UBC, University of British Columbia. For data previously published, see Appendices 1, 2.

Sphagnopsida:

Sphagnum squarrosum Crome (Sphagnales, Sphagnaceae), W.B.Schofield 120039, UBC; JN162134, JN162175, JN162204, JN162233, JN162262, JN162289, JN162318, JN162163.

Andreaeopsida:

Andreaea nivalis Hook. (Andreaeales, Andreaeaceae), W.B.Schofield 102464, UBC; JN162135, JN162176, JN162205, JN162234, JN162263, JN162290, JN162319, —.

Polytrichopsida:

Alophosia azorica (Renauld & Cardot) Cardot (Polytrichales, Polytrichaceae), F.J.Rumsey 1998, UBC; JN162136, JN162177, JN162206, JN162235, JN162264, JN162291, JN162320, JN162164 & JN162165.

Bartramiopsis lescurii (James) Kindb. (Polytrichales, Polytrichaceae), W.B.Schofield & S.S.Talbot 125406, UBC; JN162137, JN162178, JN162207, JN162236, JN162265, JN162292, JN162321, —.

Polytrichum juniperinum Hedw. (Polytrichales, Polytrichaceae), Y.Chang 07-03, UBC; JN162138, JN162179, JN162208, JN162237, JN162266, JN162293, JN162322, —.

Tetraphidopsida:

Tetraphis pellucida Hedw. (Tetraphidales, Tetraphidaceaea), S. Ellis *s.n.*, UBC; JN162139, JN162180, JN162209, JN162238, JN162267, JN162294, JN162323, JN162166 & JN162167.

Tetrodontium brownianum (Dicks.) Schwägr. (Tetraphidales, Tetraphidaceaea), D.G.Long 37600, E; JN162140 JN162181 JN162210 JN162239, —, JN162295, JN162324, —.

Bryopsida:

Archidium hallii Austin (Archidales, Arcidaceae) S.P.Churchill & T. Florentin P20196, MO; JN162149, JN162190, JN162219, JN162248, JN162304, JN162333, —.

Aulacomnium palustre (Hedw.) Schwägr. (Rhizogoniales, Rhizogoniaceae), W.B.Schofield & L.Boldwin 123403, UBC; JN162155, JN162196, JN162225, JN162254, JN162281, JN162310, JN162339, —.

Bartramia pomiformis Hedw. (Bartramiales, Bartramiaceae), S. Ellis *s.n.*, UBC; JN162152, JN162193, JN162222, JN162251, JN162278, JN162307, JN162336, JN162171 & JN162172,

Blindia acuta (Hedw.) Bruch & Schimp. (Grimmiales, Seligeriaceae), S. Joya 08-19-07-1, UBC; JN162145, JN162186, JN162215, JN162244, JN162272, JN162300, JN162329, —.

Bryoxiphium norvegicum (Brid.) Mitt. (Bryoxiphiales, Bryoxiphiaceae), S.S.Talbot KIS529, UBC; JN162150, JN162191, JN162220, JN162249, JN162276, JN162305, JN162334, —.

Diphyscium foliosum(Hedw.) D. Mohr. (Diphysciales, Diphysciaceae), W.B.Scholfield 126011B, UBC; JN162141, JN162182, JN162211, JN162240, JN162268, JN162296, JN162325, —.

Encalypta ciliata Hedw. (Encalyptales, Encalyptaceae), W.B.Schofield, J.Harpel, & Forest Service Personnel 117021, UBC; JN162143, JN162184, JN162213, JN162242, JN162270, JN162298, JN162327, —.

Funaria hygrometrica Hedw. (Funariales, Funariaceae), Y.Chang 05-01, UBC; JN162144, JN162170, JN162185, JN162214, JN162243, JN162271, JN162299, JN162328.

Grimmia pulvinata (Hedw.) Sm. (Grimmiales, Grimmiaceae), Y.Chang 08-02, UBC; JN162146, JN162187, JN162216, JN162245, JN162273, JN162301, JN162330, —.

Hedwigia stellata Hedenås (Hedwigiales, Hedwigiaceae), S. Joya 03-26-07-1, UBC; JN162154, JN162195, JN162224, JN162253, JN162280, JN162309, JN162338, —.

Hookeria lucens (Hedw.) Sm. (Hookeriales, Hookeriaceae), D.Woods 05-31, UBC; JN162159, JN162200, JN162229, JN162258, JN162285, JN162314, JN162343, JN162173.

Hypnum circinale Hook. (Hypnales, Hypnaceae), Y.Chang 08-01, UBC; JN162162, JN162203, JN162232, JN162261, JN162288, JN162317, JN162346, —.

Hypopterygium fauriei Besch. (Hookeriales, Hypopterigiaceae), K. Dillman 2005-109, UBC; JN162160, JN162201, JN162230, JN162259, JN162286, JN162315, JN162344, —.

Leucobryum javense (Brid.) Mitt. (Dicranles, Leucobryaceae), Y.Chang 07-17, UBC; JN162147, JN162188, JN162217, JN162246, JN162274, JN162302, JN162331, —.

Ptychomnion cygnisetum (Müll. Hal.) Kindb. (Ptychomniales, Ptychomniaceae), B.Allen 26391, UBC; JN162156, JN162197, JN162226, JN162255, JN162282, JN162311, JN162340, —.

Racopilum cuspidigerum (Schwägr.) Ångström (Hypnodendrales, Racopilaceae), Y.Chang 07-20, UBC;

Rhizogonium graeffeanum (Müll. Hal.) A. Jaeger (Rhizogoniales, Rhizogoniaceae), Y.Chang 07-18, UBC; XXXX, XXXX, XXXX, XXXX, XXXX, XXXX.

Rhytidiadelphus loreus (Hedw.) Warnst. (Hypnales, Hylocomiaceae), D.Woods 05-05, UBC; JN162161, JN162202, JN162231, JN162260, JN162287, JN162316, JN162345, JN162174.

Schistostega pennata (Hedw.) F. Weber & D. Mohr (Dicranales, Schistostegaceae), Y.Chang 07-01, UBC; JN162148, JN162189, JN162218, JN162247, JN162275, JN162303, JN162332, —.

Scouleria aquatica Hook. (Scouleriales, Scouleriaceae), S. Joya 04-30-07-1, UBC; JN162151, JN162192, JN162221, JN162250, JN162277, JN162306, JN162335, —.

Tetraplodon mnioides (Sw. ex Hedw.) Bruch & Schimp. (Splachnales, Splachnaceae), W.B.Schofield & S.S.Talbot 121611, UBC; JN162153, JN162194, JN162223, JN162252, JN162279, JN162308, JN162337, —.

Timmia austriaca Hedw. (Timmiales, Timmiaceae), W.B.Schofield 123467, UBC; JN162142, JN162183, JN162212, JN162241, JN162269, JN162297, JN162326, JN162168 & JN162169.

Appendix 4: Source information and GenBank numbers for taxa included in Chapter 5.

For each taxa, the information is displayed in the following sequence: species, voucher specimen, herbarium; GenBank accession: *atp*B, IGS *atp*B-*rbc*L, IGS *atp*F-*atp*H, *ndh*F, *rbc*L, and *rpo*B. "—" represents missing data Herbarium abbreviations (Thiers, 2010): ALA, University of Alaska Museum of the North; E, Royal Botanic Garden, Edinburgh; University of Connecticut; NY, New York Botanical Garden; UBC, University of British Columbia.

Andreaea acuminata Mitt.

W.B.Schofield 127663, UBC; JN388728, JN388896, JN388861, JN388765, JN388801, JN388840

Andreaea acutifolia Hook. f. & Wilson.

B.M.Murray 93-10, ALA; JN388723, JN388893, —, JN388760, JN388796, JN388835.

Andreaea alpestris (Thed.) Schimp.

F.M.Boas 2001-313A, ALA; JN388722, JN388892, JN388860, JN388759, JN388795, JN388834.

Andreaea alpina Hedw.

D.G.Long 33219, E; JN388720, JN388889, —, JN388757, JN388793, JN388831.

W.R.Buck 46192, NY; JN388721, JN388890, JN388858, JN388758, JN388794, JN388832.

Andreaea australis F. Muell. ex Mitt.

Streimann 53547, NY; JN388717, JN388886, —, JN388754, JN388790, JN388828.

Streimann 53460, NY; JN388718, JN388887, JN388856, JN388755, JN388791, JN388829.

Andreaea blyttii Schimp.

W.B.Schofield 113953, UBC; JN388692, JN388862, JN388842, JN388729, JN388766, JN388802.

Heegaard 399a, ALA; JN388694, —, —, JN388731, JN388768, JN388804.

B.M.Murray 96-198, ALA; JN388693, JN388863, JN388843, JN388730, JN388767, JN388803.

Andreaea crassinervia Bruch.

B.M.Murray 98-493, ALA; JN388704, JN388874, —, JN388741, JN388778, JN388815

Andreaea frigida Huebener.

B.M.Murray 98-497, ALA; JN388701, JN388871, —, JN388738, JN388775, JN388812.

D.G.Long & G.P.Rothero 36961, E; JN388702, JN388872, JN388848, JN388739, JN388776, JN388813

Andreaea fuegiana (Cardot) S.W. Greene

W.R.Buck:47936, NY; JN388727 JN388895, —, JN388764, JN388800, JN388839.

Andreaea heinemannii Hampe & Müll. Hal.

B.M.Murray 92-293, ALA; JN388698, JN388868, —, JN388735, JN388772, JN388809.

B.M.Murray 92-29, ALA; JN388699, JN388869, —, JN388736, JN388773, JN388810.

O'Brien 3023, ALA; JN388700, JN388870, —, JN388737, JN388774, JN388811.

Andreaea megistospora B.M. Murray

W.B.Schofield 113960, UBC; JN388705, JN388875, —, JN388742, JN388779, JN388816.

Andreaea microvaginata Müll. Hal.

B.M.Murray 93-7, ALA; JN388712, JN388882 —, JN388749, JN388786, JN388823.

H.Streimann 15014, NY; JN388713, —, —, JN388750, —, JN388824.

Andreaea mutabilis Hook. f. & Wilson

J.Shevock 18076, ALA; JN388726, —, —, JN388763, JN388799, JN388838.

Andreaea nitida Hook. f. & Wilson

W.R.Buck 47938, NY; JN388716, JN388885, —, JN388753, JN388789, JN388827.

S.P.Churchill, Z.Magombo, M.Price19834, NY; JN388714, JN388883, JN388854, JN388751, JN388787, JN388825.

H.Streimann 51004, NY; JN388715, JN388884, JN388855, JN388752, JN388788, JN388826.

Andreaea nivalis Hook.

W.B.Schofield 120263, UBC; JN162135, JN388865, JN388845, JN162176, JN162290.

D.G.Long 13547, E; JN388695, JN388864, JN388844, JN388732, JN388769, JN388805.

Andreaea remotifolia Dusén

B.Allen 26668, NY; JN388725, —, —, JN388762, JN388798, JN388837.

Andreaea rigida Wilson

D.G.Long 26602, E; JN388697, JN388867, JN388847, JN388734, JN388771, JN388808.

D.G.Long & J.Shevock 37179, E; JN388696, JN388866, JN388846, JN388733, JN388770, JN388807.

Andreaea rothii spp. *falcata* (Schimp.) Lindb.

D.G.Long 27782, E; JN388707, JN388877, JN388851, JN388744, JN388781, JN388818.

B.M.Murray 98-480, ALA; JN388708, JN388878, —, JN388745, JN388782, JN388819.

Andreaea rothii spp. *rothii* F. Weber & D. Mohr.

B.M.Murray 98-485, ALA; JN388709, JN388879, —, JN388746, JN388783, JN388820.

W.B.Schofield 124669, UBC; JN388706, JN388876, JN388850, JN388743, JN388780, JN388817.

Andreaea rupestrisi Hedw.

Y.Chang 06-03, UBC**;** GU295930, JN388891, JN388859, GU295893, GU295869, JN388833.

B.M.Murray 10-01, JN388724, JN388894, —, JN388761, JN388797, JN388836

Andreaea schofieldiana B.M.Murray

W.B.Schofield 112012, UBC; JN388703, JN388873, JN388849, JN388740, JN388777, JN388814.

Andreaea subulata Harv.

T.A.Hedderson 13568, UBC; JN388710, JN388880, JN388852, JN388747, JN388784, JN388821.

W.R.Buck 45663, NY; JN388711, JN388881, JN388853, JN388748, JN388785, JN388822.

Andreaea wilsonii Hook. f.

B.Allen 26335, NY; JN388719, JN388888, JN388857, JN388756, JN388792,JN388830.

Andreaeobryum macrosporum Steere & B.M.Murray.

O.M. Afonina 153, ALA; HQ412998, —, JN388841, HQ413022, HQ413005, —.