Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 31 (2004) 505–516
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Evolution of multiple paralogous adenosine kinase genes in the moss genus *Hygroamblystegium*: phylogenetic implications

Alain Vanderpoorten,* A. Jonathan Shaw, and Cymon J. Cox

Department of Biology, Duke University, Durham, NC 27708, USA

Received 7 March 2003; revised 17 September 2003

Abstract

Maximum likelihood analyses of DNA sequences from two chloroplast regions, *trn*L–*trn*F and *atp*B–*rbc*L, and the internal transcribed spacers of 18S–5.8S–26S rRNA gene array, were performed to resolve species relationships within the moss genus *Hygroamblystegium*. Constraining morphospecies to monophyly resulted in significantly less likely trees for *H. tenax*, but not for the other species. The lack of support for most clades and the partial incongruence among topologies necessitated the use of another independent, more variable region, namely the adenosine kinase gene (*adk*). Sequences for *adk* were polymorphic but were present as multiple copies within individuals, making parology a problem for phylogenetic analyses. *Adk* evolution was reconstructed using a reconciled gene tree approach in which duplications and losses were minimized in the context of an estimate of the species tree derived from the analysis of the cp and nrDNA sequence data. Additional resolution of the species tree was then obtained by searching for reconstructions that further reduced *adk* duplications and losses. All the traditionally recognized morphospecies appeared to be polyphyletic in the resulting tree. Together with previous data from different molecular markers, the results support the interpretation that *Hygroamblystegium* represents a recent radiation in which molecular and morphological evolution have been uncoupled.

Keywords: Gene trees; Adenosine kinase gene; Gene duplication; Paralogy; Polyploids; Moss; Hygroamblystegium

1. Introduction

At the species level, incongruence among gene trees caused by deep coalescence, hybridization, or a combination of both, often necessitates the use of multiple genes to arrive at a reliable estimate of the species tree (Maddison, 1997; and references therein). Consequently, species phylogenies combining independent DNA sequences from the chloroplast, mitochondrial, and nuclear genomes have become more common (see Soltis et al., 1998 for review). Although nuclear genes potentially contain a wealth of information about organismal phylogeny, they are often present in multiple, paralogous copies (Page, 2000; Page and Charleston, 1997). Multiplicity of sequences from the same taxon, combined with uneven taxonomic sampling, often obfuscates the identification of orthologous sequences and greatly complicates the relationships between gene and species trees (D'Erchia et al., 1996). In contrast, mitochondrial (mt) and chloroplast (cp) genomes have the virtue of comprising single-copy genes. Nevertheless, because cytoplasmic DNA is inherited as a single unit, phylogenies derived from different cytoplasmic genes are not independent estimates of organismal phylogeny (Page, 2000). In addition, cytoplasmic genomes are considerably smaller in size than nuclear genomes and, although mtDNA has proven to be a suitable tool for phylogenetic analyses at low taxonomic levels in animals (Avise, 2000; but see, e.g., Hugall et al., 1999), plant cytoplasmic DNA is by far less variable at similar taxonomic levels (Soltis et al., 1998).

The low level of variability in the chloroplast genome is a limiting factor for inferring the phylogeny of some plant taxa. In pleurocarpous mosses, which, with acrocarpous and cladocarpous taxa, constitute the class

* Corresponding author. Present address: Department of Life Sciences, University of Liège, B-22, Sart Tilman, B-4000 Liège, Belgium. Fax: +3210473490.
E-mail address: alain@duke.edu (A. Vanderpoorten).

1055-7903/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2003.09.020

Bryopsida (Division Bryophyta), cpDNA has such a low level of sequence diversity that non-coding regions such as the *trn*L intron and the *trn*L–*trn*F spacer have been used to infer ordinal relationships (Buck et al., 2000; De Luna et al., 2000). By contrast, these regions were variable enough at the population level for addressing phylogeographic issues in some acrocarps (McDaniel and Shaw, 2003). Within the mosses, nuclear DNA sequence data for phylogenetic reconstruction at low taxonomic level has been limited to the internal transcribed spacers of the 18S–5.8S–26S rRNA genes (e.g., Quandt et al., 2000; Shaw and Allen, 2000; Vanderpoorten et al., 2001) and, most recently, to the glyceraldehyde 3-phosphate dehydrogenase gene (Wall, 2002).

The primary goal of this research was to seek additional nuclear DNA regions that would help resolve the phylogeny of *Hygroamblystegium*, one of the most taxonomically problematic genera of pleurocarpous mosses. Morphologically, the typical expression of each *Hygroamblystegium* species is recognizable, but the existence of a continuum of intermediates makes it almost impossible to define species boundaries (Crum and Anderson, 1981). In previous molecular phylogenetic analyses, cp- and nrDNA sequence data resolved a monophyletic, strongly supported *Hygroamblystegium* genus (Vanderpoorten et al., 2003) but failed to resolve species relationships within it. Moreover, the conflicting placement of certain species in chloroplast and nuclear topologies suggested either that lineage sorting was incomplete or that these species had hybridized (Vanderpoorten et al., 2001, 2002a).

A screening of primers for several nuclear regions suggested that the gene coding for the adenosine kinase (*adk*), an enzyme catalyzing the phosphorylation of adenosine to adenosine monophosphate (Schomberg and Stephan, 1997), could be amplified in *Hygroamblystegium*. The gene includes 11 introns (Moffatt et al., 2000) that may prove to be phylogenetically useful at low taxonomic levels. However, the phylogenetic analysis of the *adk* gene sequences was complicated by infragenomic polymorphism.

In the continuing quest for informative genes for use in molecular systematics, phylogenetic analysis of paralogous gene copies is an increasingly important subject (e.g., Hedin and Maddison, 2001). New methodologies, including uninode coding (Simmons et al., 2000) and reconciled gene trees (Page, 2000; Page and Charleston, 1997), have been developed to address these problems. In this paper, we describe phylogenetic analyses of *adk* sequences, interpreted in the context of organismal relationships inferred from three other gene regions. We show that reconciliation of relationships among *adk* sequences with organismal relationships inferred from the other genes requires multiple gains and losses of *adk* gene copies. Finally, we analyze *adk* variation in combination with data from other genes, using the reconciled gene tree method (Page and Charleston, 1997), to clarify relationships among *Hygroamblystegium* species.

2. Materials and methods

2.1. Taxon sampling

The existence of *Hygroamblystegium* as a segregate genus from *Amblystegium* s. str. (*A. serpens*) was recently confirmed by analyses of nucleotide sequence data (Vanderpoorten et al., 2003). Within *Hygroamblystegium,* morphologically ill-defined genus with a tendency to occur in wet or aquatic habitats, no less than 46 species have been included (Missouri Botanical Garden nomenclatural data base; http://mobot.mobot.org/W3T/ Search/most.html). Recent phylogenetic analyses however indicated that a number of taxa were erroneously included within the genus due to morphological convergence. For example, *H. irriguum* var. *spinifolium* and *H. curvicaule* were both shown to belong to the genus, *Cratoneuron* (Vanderpoorten et al., 2002b). *Hygroamblystegium macroneuron*, a puzzling species only known from a few localities in Texas, has even been reinterpreted as a member of another family, the Brachytheciaceae (Vanderpoorten et al., 2001, 2002b), within which its taxonomic position has been recently established (Ignatov and Huttunen, 2002). In addition, extensive reductions to synonymy within the genus were recently published (e.g., Ochyra and Streimann, 2001). As a consequence, no less than 13 synonyms of *H. varium* were reported in a revision only concerning the Amblystegiaceae of southern Chile (Ochyra and Matteri, 2001). A few species are known only from the type locality and/or old herbarium collections (e.g., *H. calcareum*) and therefore precluded their inclusion in this study, which included five widely recognized species: *H. fluviatile*, *H. humile*, *H. noterophyllum*, *H. tenax*, and *H. varium*.

Several specimens putatively attributed to these species were sampled in order to cover as completely as possible each species' range of morphological variation and hence test morphological species concepts. Furthermore, one species of each of the closely related genera, *Cratoneuron*, *Leptodictyum*, and *Vittia*, were used as outgroups, as were four specimens of the sister genus, *Amblystegium* (Table 1).

2.2. DNA extractions, PCR conditions, cloning, and sequencing procedures

DNA was extracted, amplified and sequenced for the ITS, *trn*L–*trn*F, and *atp*B–*rbc*L regions according to the protocols described in Vanderpoorten et al. (2002a). About 900 bp of the adk gene were amplified in a

Table 1
Taxon sampling, vouchers, and GenBank Accession Numbers

Putative species	Voucher	GenBank accession number[a]
Outgroups		
Amblystegium serpens (Hedw.) B., S. & G. #1	Schofield 106313	AY009827; AF322326; AF168152; AF465121
#prim	Vanderpoorten 4158	AF465006; AF464963; AF464987; AF465116; AF465099
#sec	Vanderpoorten 4630	AF465007; AF464964; AF464987; AF465133; AF465117; AF465057AF465115; AF465114
#Hprim	Vanderpoorten s.n.	AF465001; AF464957; AF464980; AF465138; AF465137; AF465123; AF465024; AF465122
Cratoneuron filicinum (Hedw.) Spruce	Lewis 87262	AY009817; AF322332; AY009812; AF465101; AF465091; AF465096; AF465100; AF465119
Leptodictyum riparium (Schimp.) Warnst.	Bowers and Haynes 15869	AY009830; AF322325; AF16816; AF465059; AF465076; AF465093; AF465094; AF465095
Vittia elimbata Hedenäs, Vanderpoorten & Goffinet	Lewis 87826	AF465017; AF464976; AF46499; AF465136; AF465033; AF465135
Ingroups		
H. fluviatile (Hedw.) Loeske #1	Allen 16372	AY009822; AF322324; AF168154; AF465037; AF465039; AF465141
#2	Vanderpoorten 4913	AY009875; AF464954; AF46497; AF465154; AF465153; AF465051; AF465090; AF465052
#3	Anderson and Mishler 25804	AY009876; AF464955; AF464978; AF465109; AF465143; AF465142; AF465064; AF465053
#10	Vanderpoorten 3739	AF465000; AF464956; AF464979; AF465084; AF465140; AF465085; AF465097; AF465056
H. humile (P. Beauv.) Vanderpoorten, Hedenäs Goffinet #2	Buck 15943	AY009823; AF322359; AF168165; AF465132; AF465110; AF465020; AF465148; AF465022; AF465130
#sec	Vanderpoorten 4163	AF465002; AF464958; AF464981AF465025; AF465026; AF465120; AF465055; AF465105
#3	Nelson and Moore 7347	AY009874; AF464960; AF464983; AF465035; AF465023; AF465112AF465036
#3prim	Vanderpoorten s.n.	AF465003; AF464959; AF464982; AF465126; AF465061; AF465134; AF465018; AF465019
#4	Vanderpoorten s.n.	AF465004; AF464961; AF464984; AF465048; AF465080; AF465107; AF465078; AF465079; AF465151
#5	Vanderpoorten 4198	AF465005; AF464962; AF464985; AF465106; AF465060; AF465021; AF465054
H. noterophyllum (Sull.) Warnst. #1	Ley 669	AF465014; AF464973; AF464996; AF465040; AF465139; AF465038; AF465152; AF465062
#2	Buck 15914	AF465015; AF464974; AF464997; AF465074; AF465075
#3	Allen 6323	AF465016; AF464975; AF464998; AF465065; AF465086; AF465087
H. tenax (Hedw.) Jenn. #1	Schofield and Belland 94941	AY980024; AF322360; AF168164; AF465066; AF465070; AF465147; AF465073
#2	Vanderpoorten 4195	AY009820; AF322361; AF168156; AF465029; AF465108; AF465030; AF465102; AF465046; AF465003
#3	Vanderpoorten 4263	AY009821; AF322327; AF168157; AF465041; AF465044; AF465058; AF465042
#4	Bowers 15612	AY009871; AF464965; AF464988; AF465063; AF465069; AF465131
#7	Vanderpoorten s.n.	AF465008; AF464966; AF464989; AF465150; AF465071; AF465072; AF465077; AF465149
#8	Vanderpoorten s.n.	AF465009; AF464967; AF464990; AF465129; AF465045; AF465034; AF465043
#9	Vanderpoorten 4181	AF465010; AF464968; AF464991; AF465098; AF465088; AF465089; AF465027
H. varium (Hedw.) Mönk. #1	Anderson 27620	AY009825; AF322328; AF168159; AF465032; AF465031; AF465092; AF465113
#2	Risk 10825	AY009872; AF464969; AF464992; AF465049; AF465050; AF465081; AF465083; AF465082
#8	Vanderpoorten 4165	AF465011; AF464970; AF464993; AF465118; AF465144; AF465145; AF465127; AF465125
#9	Vanderpoorten 4200	AF465012; AF464971; AF464994; AF465028; AF465047; AF465111; AF465128; AF465124
#sec	Vanderpoorten 896	AF465013; AF464972; AF464995; AF465067; AF465146; AF465068

Note. All vouchers are kept at DUKE. The first three accession numbers correspond to the *trn*L–*trn*F, *atp*B–*rbc*L, and ITS sequence data, respectively, the last ones to the *adk* sequence data.

Table 2
Adk primers used for PCR and sequencing

F[a]	GAAGAAGCCAGAAAACTGGGC
R[a]	GTCACCCCATCTTCAGCAAC
1F[b]	AAGCTTTTCCCGTAAGT
2R[b]	ACTTACGGGAAAAGCTT
3R[b]	GGTCCCCTGGGTAATAAC
4F[b]	TTTCATCCCATCGGTGG

[a] Primers designed by John Wheeler (UC Berkeley).
[b] Specific primers designed for the purpose of this study.

preliminary survey of five accessions representing different species using the forward (F) and reverse (R) primers shown in Table 2. The amplified region started at about 195 bp downstream of the 155th codon and ended at the 252th codon of the *adk* cDNA isolated from the moss, *Physcomitrella patens* (von Schwartzenberg et al., 1998). Coding parts of the amplified region thus included 292 of the 1175 nucleotides of the entire *Physcomitrella adk* cDNA. PCRs were performed using the following cycle conditions: 30 cycles of 1 min at 97 °C, 1 min at 50 °C and 3 min at 72 °C, preceded by an initial melting step at 97 °C and followed by a final extension period of 7 min at 72 °C. The band of highest molecular weight obtained by this procedure, constant in all five samples, was subsequently gel-cut and sequenced. In these sequences, roughly equal peak height of two or more nucleotide states suggested superimposition of several gene copies, making it necessary to use cloning techniques prior to sequencing. PCR products were cloned using the TOPO-TA cloning kit and sequenced using the F and R primers (Table 2). From the sequences obtained, two specific pairs of primers, an external one, 4F/3R, and an internal one, 1F/2R, were designed (Table 2). The pair of primers F/3R, which gave a unique band of high molecular weight in all the investigated specimens, was used for PCR amplification. The PCR products were cloned as described above. Products from the second PCR were sequenced using the primers F and 1F, and re-sequenced with 2R and 3R.

Contigs were constructed from single-stranded forward and reverse sequences using Sequencher 3.0 (Gencodes Corp.). Sequences were aligned manually using Se.Al. 2.0 (http://evolve.zoo.ox.ac.uk/software/SeAl/main.html). Positions where different nucleotide states were in conflict were treated as missing characters and gaps were inserted where necessary to preserve positional homology. Coding and non-coding regions were identified by comparison with those of other moss sequences available on GenBank.

2.3. Phylogenetic analyses

The substitution model best fitting each genomic-region data set was chosen by calculating the likelihood of a neighbor-joining tree under 56 different nested likelihood models in PAUP 4.0b8 (Swofford, 2001) and assessing the models by hierarchical likelihood ratio tests as implemented by Modeltest 3.0 (Posada and Crandall, 1998). The parameters of the selected models were then fixed and used to analyze the ITS and cpDNA data sets under maximum likelihood using heuristic searches with 300 random addition replicates and TBR branch swapping. Bootstrap analyses under ML with 300 replicates and simple taxon addition were performed to assess support for clades. Congruence between data partitions was tested by inspection of the topologies and bootstrap scores above 70% resulting from the separate ML analyses. If bootstrap analyses on two partitions provided support >70% for two different phylogenetic relationships for the same set of accessions, this would be considered as a potential incongruence among partitions (Mason-Gamer and Kellogg, 1996; Miadlikowska and Lutzoni, 2000), in which case accessions causing the incongruence were removed from the analysis.

In the ITS plus cpDNA combined data set, constraint analyses were run to test species monophyly. Constraint trees containing two clades, one including all the accessions identified (morphologically) as a particular species and the other including accessions of all remaining taxa, were constructed for each species. Optimal topologies containing the constraint were sought under likelihood (with the same model and fixed parameters as described above). Significant departure from the unconstrained optimal topology was tested using the Shimodaira–Hasegawa test (Shimodaira and Hasegawa, 1999) with 1000 replicates of full optimization.

Heuristic searches under ML with 100 random addition replicates were performed on the *adk* data matrix. Support for branches was assessed using a Bayesian inference procedure implemented in MrBayes 2.01 (Huelsenbeck et al., 2001). Analyses were conducted after removal of all identical clones belonging to the same accession. Six individual analyses were conducted under a general time-reversible model of substitution with among-site rate heterogeneity modeled by a gamma distribution. The substitution matrix, shape of the gamma distribution, and base frequencies were all estimated during the Markov chain Monte Carlo (mcmc) procedure. Each analysis was run for 1,000,000 generations with sampling of parameters performed every 10 generations. The number of generations needed to reach stationarity (i.e., the "burnin") in the mcmc algorithm was estimated by visual inspection of the plot of ML score at each sampling point using GNUPlot-3.7.p1 (Williams et al., 1999). The trees of the "burnin" for each run were excluded from the tree set, and the trees from each run were combined to form the full sample of trees assumed to be representative of the posterior probability distribution. Posterior probabilities for branches were calculated by constructing a 50% majority-rule consensus tree in PAUP.

Initial analyses of the *adk* data set indicated that copies of the gene from a single individual were often quite divergent. Sequences from different individuals were frequently more similar than were sequences from an individual. In order to investigate evolutionary patterns in the *adk* gene, the ITS plus cpDNA consensus tree was taken as a starting hypothesis of species relationships in an analysis of the *adk* data set using the "reconciled tree" approach as implemented in GeneTree 1.0 (Page, 1998). For this purpose, the *adk* 50% majority-rule consensus tree from the Bayesian analysis, whose polytomies were randomly resolved, was compared to the ITS plus cpDNA consensus tree. The minimum number of duplication and losses required to reconcile the *adk* tree with the ITS plus cpDNA consensus tree was used as the optimality criterion for calculating the "cost" of the reconciled tree. Subsequently, 100 heuristic searches were performed by perturbing the initial species tree in search of the species tree that yielded the lowest cost reconciled tree. The space of all possible species trees was limited by constraining the outgroup taxa to form a sister clade to the ingroup taxa, and ingroup topology was constrained to include all clades supported by $\geqslant 70\%$ bootstrap support in the original analysis of combined cpDNA plus ITS. The alternate "nearest neighbor interchanges" and "subtree pruning and regrafting" strategy were chosen for tree perturbation. When a perturbation produced an improvement, the search was immediately restarted using the better tree (i.e., steepest ascent was not enforced).

Table 3
Alignment length and percent of variable sites in each partition

Partition	Alignment length (in ntds)	% of variable sites
ITS	918	11.0 (8.3)
3′ end of the 18S gene, 5.8S gene, and 5′ end of the 26S gene	160	0.6 (0.6)
ITS1	383	11.2 (9.7)
ITS2	375	15.2 (10.1)
*trn*L–*trn*F	461	5.2 (1.7)
*trn*L gene and 5′ end of the *trn*F gene	397	4.3 (0.5)
*trn*L–*trn*F spacer	64	10.9 (9.4)
*atp*B–*rbc*L spacer, including 3′ end of the *rbc*L gene and 5′ end of the *atp*B gene	675	6.5 (2.7)
Adk	909	47.5 (38.1)
Exons	291	29.2 (22.0)
Introns	618	56.1 (45.6)

Note. Percent values are calculated by dividing the number of variable characters by the total number of characters of the region. Percent with the outgroup species excluded are in parentheses.

3. Results

For each data partition, the DNA sequence characteristics are presented in Table 3 and the substitution model parameters maximizing likelihoods of the data are presented in Table 4. The two chloroplast regions, *trn*L–*trn*F and *atp*B–*rbc*L, yielded 5.2% (1.7% after exclusion of the outgroup taxa) and 6.5% (2.7%) variable sites, respectively. Analyses of the cpDNA data matrix resulted in two most likely trees ($-\ln L = 1811.7094$).

Table 4
Substitution models selected for the ML analyses

		ITS	cpDNA	ITS + cpDNA[a]	*adk*[a]
Model selected		HKY[b] + G[c]	GTR[d] + I[e]	GTR + G	GTR + G
$-\ln L$		1665.7615	1812.3741	3570.2852	4661.7639
Base frequencies		A = 0.1724 C = 0.3271 G = 0.2982 T = 0.2023	A = 0.3969 C = 0.1028 G = 0.1367 T = 0.3636	A = 0.3102 C = 0.1884 G = 0.2006 T = 0.3008	A = 0.2223 C = 0.2250 G = 0.2274 T = 0.3253
Substitution model		Ti/Tv ratio = 0.6794	A–C = 1.7315 A–G = 2.3881 A–T = 0.2494 C–G = 1.7732 C–T = 2.3881	A–C = 1.0846 A–G = 1.5680 A–T = 0.3015 C–G = 2.2564 C–T = 1.5680	A–C = 0.8099 A–G = 3.0901 A–T = 1.1780 C–G = 0.3920 C–T = 3.0901
Among-site rate variation	Proportion of invariable sites	0	0.7535	0	0
	Gamma distribution shape parameter	0.3930	Equal rates for all sites	0.2008	2.9822

[a] Analyses conducted after deletion of *H. varium* #2, *H. tenax* #2, and *H. humile* #sec.
[b] Hasegawa et al. (1985) substitution model.
[c] Shape parameter of the gamma distribution.
[d] Rodriguez et al. (1990) substitution model.
[e] Proportion of invariable sites.

Two clades with low bootstrap support, one including all the accessions of *A. serpens* and an accession of *H. humile* (#sec), and the other including all the other species, were resolved in the strict consensus. Within the large *Hygroamblystegium* clade, relationships were poorly supported, with the exception of a clade with 85% bootstrap support (b.s.) including two accessions of *H. fluviatile* (65% b.s.) and four accessions of *H. tenax* (95% b.s.).

Analyses of the ITS data matrix, which included 9.7 and 10.1% variable sites after removal of the outgroup species for ITS1 and ITS2, respectively (Table 3),

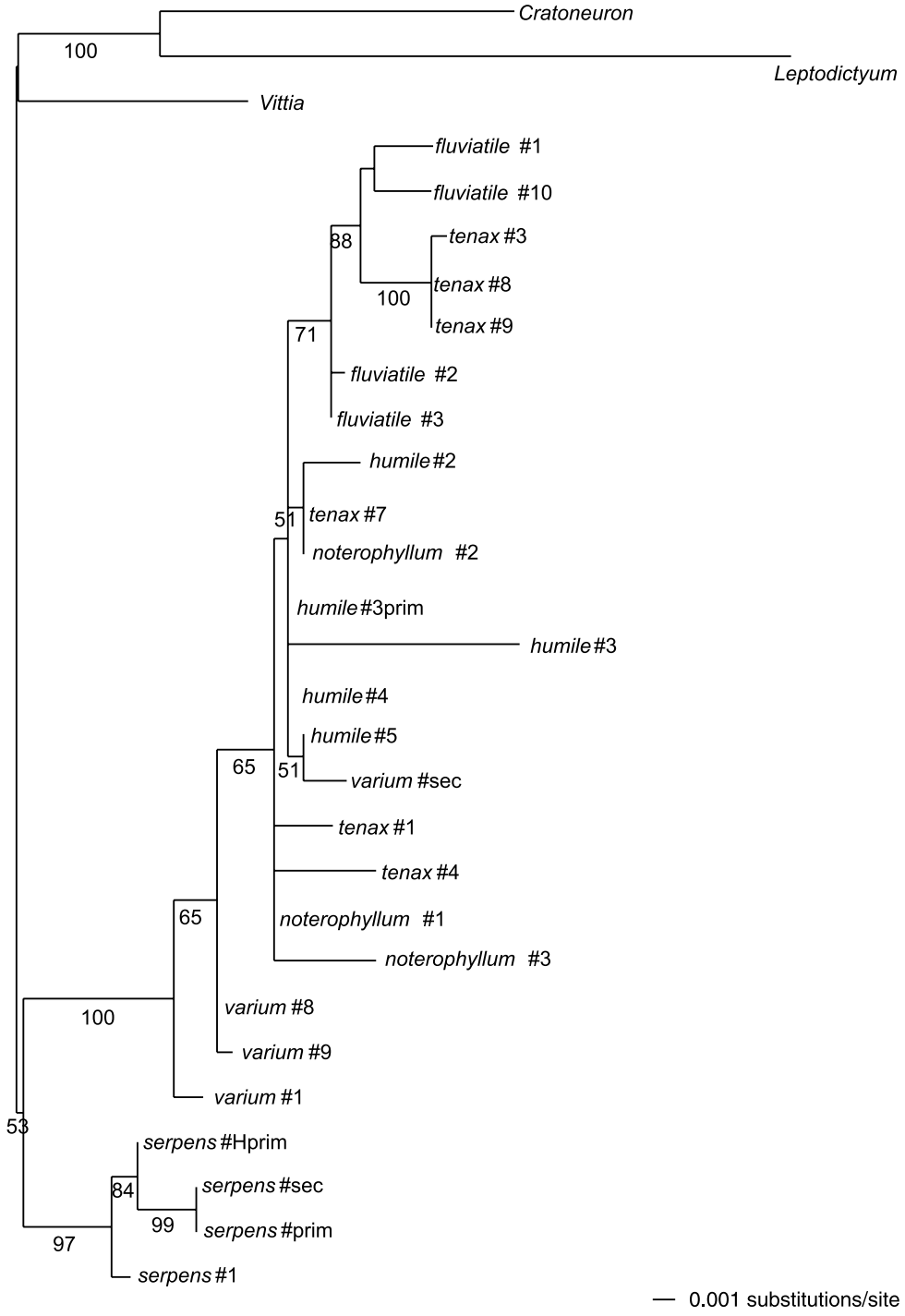

Fig. 1. Phylogram of the unique most likely tree from the ML analysis of the combined ITS plus cpDNA data matrix after deletion of the accessions causing incongruence and using the substitution model and fixed parameters as described in Table 4. Bootstrap values higher than 50% are indicated below the branches.

resulted in a single most likely tree ($-\ln L = 1677.7859$). An accession of *H. varium* (#2) was nested with *Leptodictyum* (100% b.s.) and an accession of *H. tenax* (#2) was nested with *Cratoneuron* (88% b.s.). A clade consisting of four *A. serpens* collections was well-supported and sister to a larger clade containing the *Hygroamblystegium* species. Relationships within the *Hygroamblystegium* clade, which was supported at 100%, were poorly resolved.

The positions of three accessions were incongruent between the cpDNA and ITS phylogenies. In the cpDNA phylogeny, *H. varium* #2 and *H. tenax* #2 were nested within the *Hygroamblystegium* clade, whereas they were sister to *Leptodictyum* and *Cratoneuron*, respectively, in the ITS phylogeny. *H. humile* #sec was nested within the *A. serpens* clade in the cpDNA topology but within the main clade including all the other ingroup taxa in the ITS phylogeny. These three accessions were removed before combining the ITS and cpDNA data sets. The combined analyses resulted in a single most likely tree ($-\ln L = 3552.9610$) (Fig. 1). The four accessions of *A. serpens* formed a strongly supported clade sister to *Hygroamblystegium*. Within the *Hygroamblystegium* clade, the only well supported clades were two accessions of *H. fluviatile* (60% b.s), three accessions of *H. tenax* (100% b.s.), and the clade joining these two lineages (71% b.s.). The other accessions of *H. fluviatile* and *H. tenax* fell outside the latter clade. Constraining species to monophyly resulted in significantly less likely trees for *H. tenax* (Table 5), but not the other species.

After removal of the outgroups, 38.1% of sites were variable in the *adk* data set (Table 3). The data set included four introns that shared identical nucleotides at 54.4% of the sites. In the four exons, 78.0% of sites had identical nucleotides. Multiple *adk* copies were cloned from several individuals. Average infragenomic variability reached 2.6% (standard deviation of 1.4%) in the exons and 8.7% (standard deviation of 6.4%) in the introns. The highest divergence levels were reached in *H. tenax* #2, *H. humile* #sec, and *H. varium* #2, with more than 4.8% variability in the exons and 18% variability in the introns.

Analysis of the *adk* data matrix under ML after exclusion of the three accessions whose positions were incongruent between ITS and cpDNA data partitions yielded a single tree ($-\ln L = 4661.7639$) (Fig. 2). Almost all accessions contained gene copies in at least two different *adk* gene tree clades. Examination of branch lengths on the phylogram (Fig. 2) indicates that substitutions accumulate along long, well-supported branches leading to clades including accessions of different species with little molecular divergence. A minimum of 41 duplication events and 256 losses were necessary to reconcile the *adk* gene tree associated with the three species trees (Fig. 3). Among the 41 duplication events, 19 were based on physical evidence (i.e., the existence of different gene copies within an organism) whereas 22 were inferred from incongruence between gene and species trees. Only six of the inferred duplications occurred on well-supported nodes (Fig. 3). The strict consensus of the six rival species trees is presented in Fig. 4. This tree represents our best estimate of *Hygroamblystegium* phylogeny. Well-supported relationships inferred from cpDNA plus ITS (Fig. 1) are preserved in this reconstruction (by enforcing them as topological constraints), and additional resolution is provided by *adk*. This refined species tree involves the lowest number of *adk* duplications/losses within the context of the topological constraints imposed. All the *Hygroamblystegium* species appeared to be polyphyletic (Fig. 4).

Table 5
Results of the Shimodaira–Hasegawa tests of species monophyly for the combined cpDNA and ITS data set

Species with enforced monophyly	$-\ln L$	Difference in $-\ln L$	p[a]
None (best tree)	3552.9610		
tenax	3589.2536	36.2926	0.007*
Fluviatile	3561.1675	8.2065	0.107
Humile	3564.4172	11.4562	0.085
Varium	3569.1561	16.1951	0.098
Noterophyllum	3574.6900	21.729	0087

[a] The p value represents the probability of accepting the null hypothesis of no difference between the optimal trees. An asterisk indicates a significant difference.

4. Discussion

A remarkable feature of *Hygroamblystegium* genomes is the presence, within individuals, of multiple copies of the *adk* gene. This situation contrasts with that observed in other mosses for which the gene has been sequenced, e.g., *Physcomitrella* (von Schwartzenberg et al., 1998) and *Ceratodon* (Shaw et al., 2002). The infra-genomic polymorphism found in *Hygroamblystegium* strongly suggests duplications of *adk* itself, or the entire genome. Both coding and noncoding regions of *adk* in *Hygroamblystegium* were substantially less divergent among paralogs than the 11% divergence found between the coding sequences of the two *adk* copies recently isolated in *Arabidopsis* (Moffatt et al., 2000). Nevertheless, many gene copies from a single *Hygroamblystegium* plant showed less similarity with one another than with copies from other accessions and/or other species. Reports of extensive, within-organism polymorphism is usually associated with divergent evolution of gene arrays, hybridization, or formation of pseudogenes (see Campbell et al., 1997; Hugall et al., 1999 and references therein).

There is a growing body of evidence that allopolyploidization, one of the main pathways of hybrid

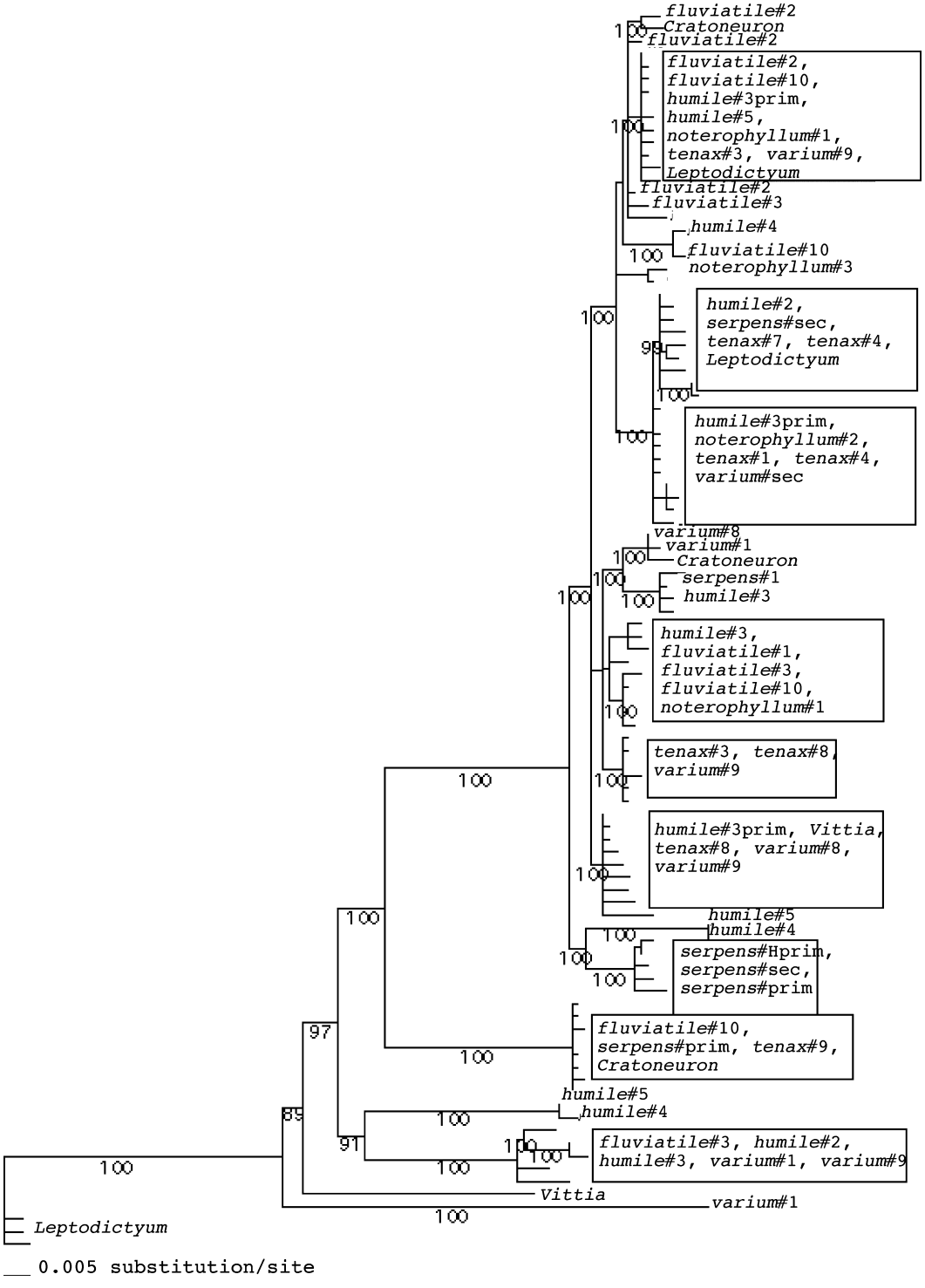

Fig. 2. Phylogram of the unique most likely tree from the ML analysis of the *adk* data matrix using the substitution model and fixed parameters as described in Table 4. Numbers below the branches represent the posterior probabilities of the Bayesian analysis, i.e., the percentage of trees in which a branch is resolved in the 50% majority rule consensus tree from the sample of 355,000 trees assumed to be representative of the posterior probability distribution.

speciation (Levin, 2000; Otto and Whitton, 2000), is common and widespread in mosses (see Derda and Wyatt, 2000 and references therein). A comparative analysis of the ITS, cpDNA, and *adk* data sets suggests that hybridization may have occurred in *Hygroamblystegium*. Indeed, the three accessions causing incongruence between the cpDNA and ITS topologies also possessed the highest levels of infragenomic polymorphism in *adk*

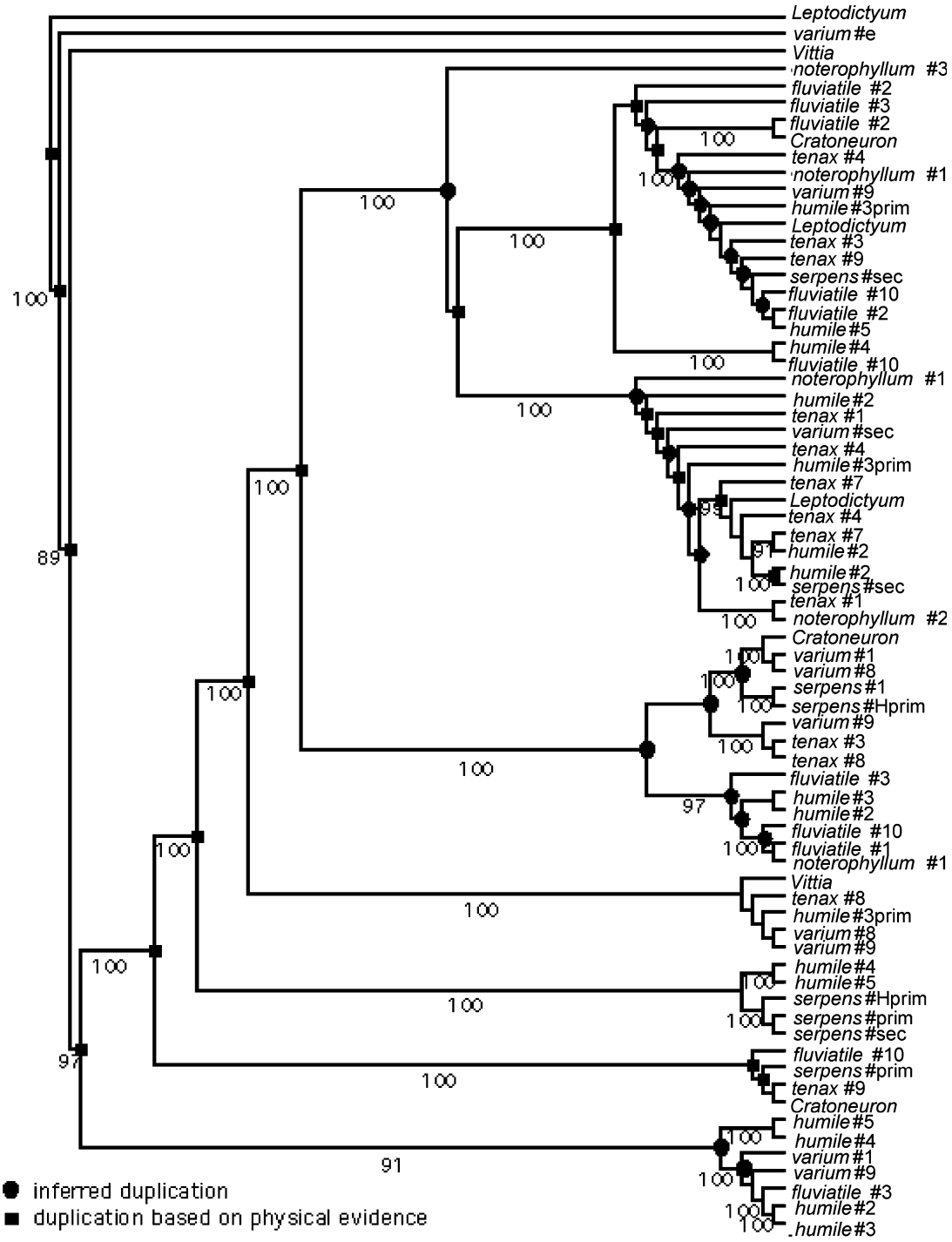

Fig. 3. Reconciled *adk* gene tree. Duplications that are necessary to reconcile the gene tree with the associated species trees, whose strict consensus is given in Fig. 4, are marked with a plain square (duplications based on physical evidence) or a plain circle (inferred duplication). Numbers below the branches represent the posterior probabilities of the Bayesian analysis.

gene sequences. Hybridization may account for both incongruence among ITS and cpDNA topologies and high divergence of gene copies within plants and within species. Most studies reporting allopolyploidization in mosses have, however, involved genera that reproduce sexually (e.g., Derda and Wyatt, 2000). By contrast, *Hygroamblystegium* species are rarely fertile (Crum and Anderson, 1981). Hence, repeated episodes of horizontal transfer are unlikely to fully account for the divergent paralogous *adk* sequences identified from within individuals of *Hygroamblystegium.*

High polyploid states, as those encountered in *Hygroamblystegium* and sister genera, such as *Amblystegium*, may provide an opportunity for different arrays of

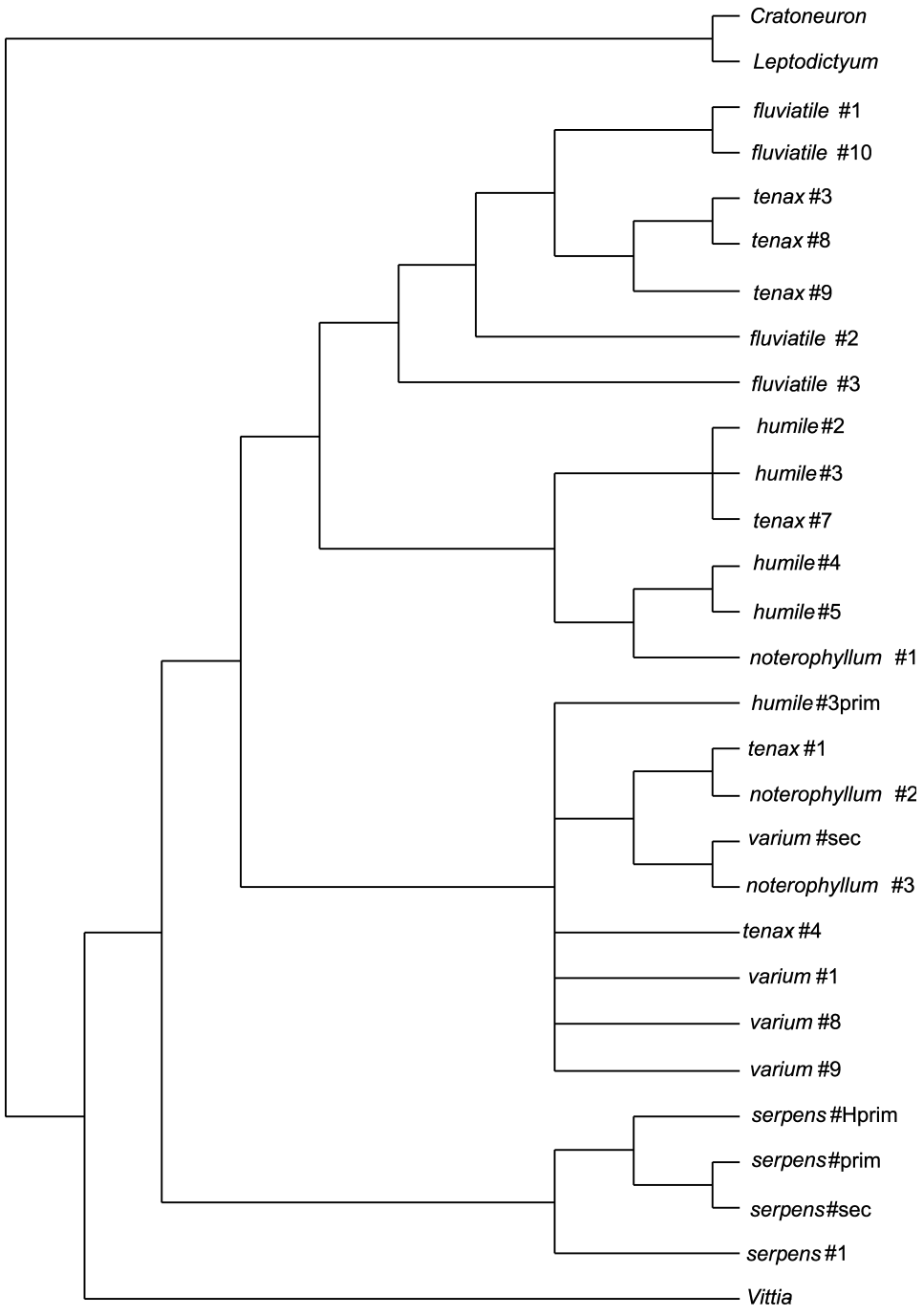

Fig. 4. Strict consensus of the three species trees associated with the reconciled *adk* gene tree of lowest cost.

nDNA to evolve independently (e.g., Suh et al., 1993; Wendel, 2000). As opposed, for example, to *Ceratodon*, a haploid species in which *adk* was found to be unilocus in a range of accessions (Shaw et al., 2002), *Hygroamblystegium* and *Amblystegium* species indeed are known to be variously polyploid. For example, chromosome numbers of $n = 12, 24, 48; 11, 19, 19 + 2m$ were reported in *A. serpens* (Fritsch, 1982).

Repeated events of gene duplication and losses may account for the observed polymorphism of *adk* in *Hygroamblystegium*. Using the "reconciled gene trees" approach (Page, 2000; Page and Charleston, 1997) (see Fig. 4 in Page and Charleston, 1997), the most parsimonious estimation of duplication events that are required to reconcile the *adk* gene tree with the species tree reaches the value of 41 across the genus. Sixteen of these duplications may be methodological artifacts. Indeed, these duplications were not based on physical evidence, i.e., existence of different gene copies in the same accession, but were instead required for reconciling the

species trees with the gene tree, and occurred at nodes that were randomly resolved for computing purpose and/or at poorly supported nodes. This means that the sequence data may support alternative gene trees that do not require these duplications. The 25 remaining duplication events were either based on physical evidence or occurred at well-supported nodes and are therefore more strongly supported.

Gene duplication must have happened at some point to account for the existence of divergent *adk* gene copies within a genome. The similarity among certain gene copies of some *Hygroamblystegium* and outgroup species, as well as the fairly long branches separating the different paralogous copies, suggest that duplication occurred at fairly deep phylogenetic level. Towards the terminal nodes, gene duplications may, however, be unlikely (Maddison, 1997). Alternatively, although all duplications discussed so far were interpreted as actual gene duplications, a number of them inferred near the terminal nodes, which were not based on physical evidence but were inferred to reconcile the gene tree with the species tree (see Figs. 1 and 3 in Page and Charleston, 1997), may also correspond to coalescence events of alleles, which occurred within the common ancestor of the current species (Page and Charleston, 1997). The interpretation, that *Hygroamblystegium* represents a recent radiation whose short timeframe did not allow the complete sorting of *adk* arrays, is supported by the low level of divergence of non-coding cpDNA regions and ITS.

Independent evolution of gene paralogs in ancestral species with subsequent incomplete lineage sorting of alleles and/or successive gene duplication and losses made it necessary to reconstruct the species tree compatible with the reconciled *adk* gene tree, with the number of evolutionary events, be they gene duplication and losses or lineage sorting events, minimized. The tree supported a monophyletic interpretation of *A. serpens* but each of the *Hygroamblystegium* morphospecies appeared polyphyletic. All the molecular markers that have been used to date, including AFLP, ITS, and cpDNA sequence data (Vanderpoorten et al., 2002b), also failed to resolve monophyletic *Hygroamblystegium* species. Accessions of one morphospecies, *H. tenax*, occurred in widely divergent clades defined by molecular data. As in agamosperm complexes (Gornall, 1999), *Hygroamblystegium* may be characterized by multiple processes that confound attempts at an unambiguous taxonomic treatment. These processes may include hybridization, polyploidization, phenotypic plasticity, and convergent evolution of morphologies commonly used to delineate species in the genus. In this context, the results presented here, coupled with a detailed morphological study on specimens grown under identical conditions, will serve as a basis for examining patterns of morphological evolution (Vanderpoorten and Jacquemart, 2003).

Acknowledgments

The senior author acknowledges financial support from the Belgian Funds for Scientific Research (FNRS). This research was also supported by NSF grant no. DEB-0089131 to A.J.S. We thank Sandra Boles for assistance in the laboratory, John Wheeler for providing information about *adk* primers, Stuart McDaniel and Isabelle Olivieri for constructive discussions, Rodrick Page for his advice concerning the use of his program, GeneTree, and Molly McMullen for curatorial assistance.

References

Avise, J.C., 2000. Phylogeography. Harvard University Press, Cambridge.

Buck, W.R., Goffinet, B., Shaw, A.J., 2000. Testing morphological concepts of orders of pleurocarpous mosses (Bryophyta) using phylogenetic reconstructions based on *trn*L–*trn*F and *rps*4 sequences. Mol. Phylogenet. Evol. 16, 180–198.

Campbell, C.S., Wojciechowski, M.F., Baldwin, B.G., Alice, L.A., Donoghue, M.J., 1997. Persistent nuclear ribosomal DNA sequence polymorphism in the *Amelanchier* agamic complex (Rosaceae). Mol. Biol. Evol. 14, 81–90.

Crum, H.A., Anderson, L.E., 1981. In: Mosses of Eastern North America, vol. 2. Columbia University Press, New York.

De Luna, E., Buck, W.R., Akiyama, H., Arikawa, T., Tsubota, H., Gonzalez, D., Newton, A.E., Shaw, A.J., 2000. Ordinal phylogeny within the *Hypnobryalean pleurocarpous* mosses inferred from cladistic analyses of three chloroplast DNA sequence data sets: *trn*L–F, *rps*4, and *rbc*L. Bryologist 103, 242–256.

D'Erchia, H., Gissi, C., Pesole, G., Saccone, C., Arnason, U., 1996. The guinea-pig is not a rodent. Nature 381, 597–600.

Derda, G., Wyatt, R., 2000. Isozyme evidence regarding the origins of the allopolyploid species of *Polytrichastrum* (Polytrichaceae, Bryophyta). Plant Syst. Evol. 220, 37–53.

Fritsch, R., 1982. Index to chromosome numbers—Bryophyta. Junk, The Hague.

Gornall, R.J., 1999. Population genetic structure in agamospermous plants. In: Hollingsworth, P.M., Bateman, R.M., Gornall, R.J. (Eds.), Molecular Systematics and Plant Evolution. Taylor & Francis, London, pp. 118–138.

Hasegawa, M., Kishino, K., Yano, T., 1985. Dating the human–ape splitting by a molecular clock of mitochondrial DNH. J. Mol. Evol. 22, 160–174.

Hedin, M.C., Maddison, W.P., 2001. Phylogenetic utility and evidence for multiple copies of elongation factor-1 alpha in the spider genus *Habronattus* (Araneae : Salticidae). Mol. Biol. Evol. 18, 1512–1521.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollack, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Hugall, A., Stanton, J., Moritz, C., 1999. Reticulate evolution and the origins of ribosomal transcribed spacer diversity in apomictic Melaidogyne. Mol. Biol. Evol. 16, 157–164.

Ignatov, M.S., Huttunen, S., 2002. Brachytheciaceae (Bryophyta)—a family of sibling genera. Arctoa 11, 245–296.

Levin, D.H., 2000. The origin, expansion, and demise of plant species. Oxford Series in Ecology and Evolution.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Mason-Gamer, R., Kellogg, E., 1996. Testing for phylogenetic conflict among molecular data sets in the tribe Triticeae (Gramineae). Syst. Biol. 45, 524–545.

McDaniel, S., Shaw, A.J., 2003. Phylogeographic structure and cryptic speciation in the trans-antarctic moss *Pyrrhobryum minoides*. Evolution 57, 205–215.

Miadlikowska, J., Lutzoni, F., 2000. Phylogenetic revision of the genus *Peltigera* (lichen-forming Ascomycota) based on morphological, chemical, and large subunit nuclear ribosomal DNA. Int. J. Plant Sci. 161, 925–958.

Moffatt, B.A., Wang, L., Allen, M.S., Stevens, Y.Y., Qin, W., Snider, J., von Schwarzenberg, K., 2000. Adenosine kinase of *Arabidopsis*. Kinetic properties and gene expression. Plant Physiol. 124, 1775–1785.

Ochyra, R., Matteri, C.M., 2001. Bryophyta, Musci: Amblystegiaceae. Flora Cryptogamica de Tierra del Fuego 14, 1–95.

Ochyra, R., Streimann, H., 2001. Studies on austral Amblystegiaceae: I. Australian taxa. Nova Hedwigia 73, 453–460.

Otto, S.P., Whitton, J., 2000. Polyploid incidence and evolution. Annu. Rev. Genet. 34, 401–437.

Page, R.D.M., 1998. GeneTree: comparing gene and species trees using reconciled trees. Bioinformatics 14, 819–820.

Page, R.D.M., 2000. Extracting species trees from complex gene trees: reconciled trees and vertebrate phylogeny. Mol. Phylogenet. Evol. 14, 89–106.

Page, R.D.M., Charleston, M.A., 1997. From gene to organismal phylogeny: reconciled trees and the gene tree/species tree problem. Mol. Phylogenet. Evol. 7, 231–240.

Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.

Quandt, D., Tangney, R.S., Frahm, J.-P., Frey, W., 2000. A molecular contribution for understanding the Lembophyllaceae (Bryopsida) based on noncoding chloroplast regions (cpDNA) and ITS2 (nrDNA) sequence data—studies in austral temperate rain forest bryophytes 8. J. Hatt. Bot. Lab. 89, 71–92.

Rodriguez, F., Oliver, J.F., Marin, A., Medina, J.R., 1990. The general stochastic model of nucleotide substitution. J. Theor. Biol. 142, 485–501.

Schomberg, D., Stephan, D., 1997. Adenosine kinase. In: Schomberg, D., Stephan, D. (Eds.), Enzyme Handbook, vol. 13. Springer Verlag, Berlin, pp. 1–9.

Shaw, A.J., Allen, B., 2000. Phylogenetic relationships, morphological incongruence, and geographic speciation in the Fontinalaceae (Bryophyta). Mol. Phylogenet. Evol. 16, 225–237.

Shaw, A.J., McDaniel, S.F., Werner, O., Ros, R.M., 2002. New frontiers in bryology and lichenology—phylogeography and phylodemography. Bryologist 105, 373–383.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Simmons, M.P., Donovan Bailey, C., Nixon, K.C., 2000. Phylogeny reconstruction using duplicate genes. Mol. Biol. Evol. 17, 469–473.

Soltis, D.E., Soltis, P.S., Doyle, J.J. (Eds.), 1998. Molecular Systematics of Plants II: DNA Sequencing. Kluwer Academic Publishers, Boston.

Suh, Y., Thien, L.B., Reeve, H.E., Zimmer, E.H., 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. Am. J. Bot. 80, 1042–1055.

Swofford, D.L., 2001. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland.

Vanderpoorten, A., Jacquemart, A.-L., 2003. Evolutionary mode, tempo, and phylogenetic association of continuous morphological traits in the aquatic moss genus, *Amblystegium*. J. Evol. Biol., in press.

Vanderpoorten, A., Shaw, A.J., Goffinet, B., 2001. Testing controversial alignments in *Hygroamblystegium* and related genera (Amblystegiaceae: Bryopsida). Evidence from rDNA ITS sequences. Syst. Bot. 28, 470–479.

Vanderpoorten, A., Hedenäs, L., Cox, C.J., Shaw, A.J., 2002a. Phylogeny and morphological evolution of the Amblystegiaceae (Bryopsida). Mol. Phylogenet. Evol. 23, 1–21.

Vanderpoorten, A., Hedenäs, L., Cox, C.J., Shaw, A.J., 2002b. Circumscription, phylogenetic relationships and taxonomy of Amblystegiaceae inferred from nr and cpDNA sequence data and morphology. Taxon 51, 115–122.

Vanderpoorten, A., Goffinet, B., Hedenäs, L., Cox, C.J., Shaw, A.J., 2003. A taxonomic reassessment of the Vittiaceae (Hypnales, Bryopsida): evidence from phylogenetic analyses of combined chloroplast and nuclear sequence data. Plant Syst. Evol. 241, 1–12.

von Schwartzenberg, K., Kruse, S., Reski, R., Moffatt, B., Laloue, M., 1998. Cloning and characterization of an adenosine kinase from *Physcomitrella* involved in cytokinin metabolism. Plant J. 13, 249–257.

Wall, D.P., 2002. Use of the nuclear gene glyceraldehyde 3-phosphate dehydrogenase for phylogeny reconstruction of recently diverged lineages in *Mitthyridium* (Musci: Calymperaceae). Mol. Phylogenet. Evol. 25, 10–26.

Wendel, J.F., 2000. Genome evolution in polyploids. Plant Mol. Biol. 42, 225–249.

Williams, P., and multiple authors, 1999. Gnuplot 3.71. Distributed by the authors.