ACADEMIC PRESS

Molecular Phylogenetics and Evolution 25 (2002) 10–26

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.academicpress.com

Use of the nuclear gene *glyceraldehyde 3-phosphate dehydrogenase* for phylogeny reconstruction of recently diverged lineages in *Mitthyridium* (Musci: Calymperaceae)

Dennis P. Wall*

Department of Integrative Biology, University and Jepson Herbaria, University of California, Berkeley, USA

Received 19 June 2001; received in revised form 18 January 2002

Abstract

A portion of the nuclear gene *glyceraldehyde 3-phosphate dehydrogenase* (*gpd*) was sequenced in 26 representatives of the paleotropical moss, *Mitthyridium*, and a group of 20 outgroup taxa to assess its utility for phylogenetic reconstruction compared with the better understood chloroplast markers, *rps*4 and *trn*L. Primers based on plant and fungal sequences were designed to amplify *gpd* in plants universally with the exclusion of fungal contaminants. The piece amplified spanned 4 introns and 3 of 9 exons, based on comparisons with complete sequence from *Arabidopsis*. Size variation in *gpd* ranged from 891 to 1007 bp, in part attributable to 6 indels of variable length found within the introns. Intron 6 contributed most of the length variation and contained a variable purine-repeat motif of possible use as a microsatellite. Phylogenetic analyses of the full *gpd* amplicon yielded well-resolved trees that were in nearly full accord with the trees derived from the cpDNA partitions for analyses of both the ingroup and ingroup + outgroup taxon sets. Pairwise nucleotide substitution rates of *gpd* were as much as 2.2 times higher than those in *rps*4 and 2.8 times higher than in *trn*L. Excision of the introns left suitable numbers of parsimony informative characters and demonstrated that the full *gpd* amplicon could be compartmentalized to provide resolution for both shallow and deep phylogenetic branches. Exons of *gpd* were found to behave in a clock-like fashion for the 26 ingroup taxa and select outgroups. In general, *gpd* was found to hold great promise not only for improving resolution of chloroplast-derived phylogenies, but also for phylogenetic reconstruction of recent, diversifying lineages.

1. Introduction

Few nuclear genes are currently available for molecular phylogenetic studies, especially ones that meet all of the criteria thought likely to allow reconstruction of historical relationships. Frequently used nuclear regions like the Internal and External Transcribed Spacer regions of nuclear ribosomal DNA, while remaining widely useful within the systematic community, can be problematical for phylogenetic reconstruction due to divergent paralogous evolution (Buckler et al., 1997) or insufficient variation. Thus, the demand for new nuclear genes remains high for several reasons, including the need for better phylogenetic resolution at shallow phylogenetic levels, the need to test results derived from other genomes and from morphology, and finally the need to identify biological phenomena such as reticulation and convergence. To date, few searches for phylogenetically useful nuclear genes have produced adequate rewards, despite the large size of the nuclear genome. Some notable exceptions include the small subunit of ribulose 1,5-bisphosphate carboxylase (*rbc*S), alcohol dehydrogenase (Adh), and chalcone synthase (Chs), but each of these exists as multi-gene families and thus also presents problems of paralogous evolution (Clegg et al., 1997). Moreover, these genes, especially Adh and Chs, may have undergone excessive recombination that could have clouded their actual history (Clegg et al., 1997). Now, the rapid accumulation of expressed sequence tag libraries (ESTs) and whole nuclear genome databases promises to greatly assist the search for new nuclear markers. Perhaps the most obvious uses of ESTs and genomic databases are gene identification and discovery. More extensive EST

* Present address. Department of Biological Sciences, Stanford University, Stanford, CA 94305, USA.
E-mail address: dpwall@stanford.edu.

1055-7903/02/$ - see front matter
PII: S1055-7903(02)00355-X

databases will reveal homologs from other plants at shallow or deep levels of history; such sequence identification will aid greatly in the development of universal primers. The development of an encyclopedic set of ESTs from organisms that span the nodes of the land plant phylogeny is imminent, despite a current bias on flowering plants. Mining those databases should reveal new genes useful for all levels of phylogenetic resolution and advance our understanding of evolution dramatically.

Of the 329 ESTs now available in GenBank (composing 8,114,353 gene sequences), two are mosses (and compose only 1010 sequences of the total database), which bodes well for the future of moss as a model system. Still, the moss system remains underexploited, not only as a model to understand plant evolution, but also as a model to comprehend molecular evolution generally, despite its biological importance and amenity to laboratory study. That said, genetic research on mosses is heading in very exciting directions (Cove, 2000; Panvisavas et al., 1999; Reski, 1998; Reski et al., 1998; Wood et al., 2000), especially since efficient gene targeting and disruption by homologous recombination has become routinely possible (Girke et al., 1998; Schaefer and Zyrd, 1997). It is likely that mosses will take a more premier role.

The understanding of moss phylogeny is growing, but is hindered by a lack of nuclear genetic input, especially for closely related taxa. So far, studies in moss phylogenetics have built large databases of primarily chloroplastic markers and smaller databases of traditional nuclear markers, i.e., ITS and 18s rDNA. More nuclear input is needed. In the present paper, I demonstrate the phylogenetic utility of a nuclear gene, glyceraldehyde 3-phosphate dehydrogenase, for moss phylogenetics.

1.1. Glyceraldehyde 3-phosphate dehydrogenase

A paper by Strand et al. (1997) developed primer pairs for a number of useful low-copy (or potentially single-copy) nuclear genes. They highlighted the increasing utility of GenBank, EMBL, and DDBJ for gene targeting and primer design. To develop their primer sets, the researchers presumably used a group of angiosperms—at the time probably the only taxa with suitable sequence availability. Among their targets was a partial sequence of the gene, glyceraldehyde 3-phosphate dehydrogenase (their amplicon was referred to as *g3pdh*, but is hereafter called *gpd*). This gene amplicon was the most promising among the set of markers since it was successfully amplified in every taxon and did not produce multiple bands (Strand et al., 1997). Their *gpd* primers have been used in only one other subsequent plant study-on cultivars of the angiosperm species *Manihot esculenta* (Olsen and Schaal, 1999). To date, the utility of the *gpd* primers for other plant lineages besides angiosperms has not been demonstrated. The increasing availability of plant ESTs will make it possible to broaden the usability of the primers, either through modification or as a pointer to other portions of the full gene.

The full gene encodes *gpd* (GAPDH) a common catalytic enzyme responsible for the conversion of *glyceraldehyde 3-phosphate* into 3-phosphoglycerate and is centrally important to both glycolysis and the Calvin cycle in eukaryotes and eubacteria (Figge et al., 1999). The two genes in the GAPDH family of eukaryotes, GapC ($=$ the entire gene of which the partial sequence, *gpd*, is a part) and GapAB, are known to be nuclear encoded. GapAB encodes chloroplast Calvin cycle GAPDH in plants and is highly divergent ($>50\%$) from its gene family member, GapC. GapC encodes the cytosolic GAPDH of glycolytic–gluconeogenetic function (Figge et al., 1999). While GapC (*gpd*) has been used for phylogenetic studies of some organismal groups (Fagan et al., 1998; Henze et al., 1995; Viscogliosi and Mueller, 1998), it has not been widely studied in plants (Martin et al., 1993; Olsen and Schaal, 1999; Schaal et al., 1998; Schaal and Olsen, 2000). In fact, its variability in plants is known from only two empirical studies (Martin et al., 1993; Olsen and Schaal, 1999). Those studies report two very different nucleotide substitution rates between taxon pairs for GapC. One reported that the mutation rate of GapC parallels the relatively slow-evolving chloroplast gene *rbc*L (Martin et al., 1993); the other demonstrated sequence variability suitable for phylogeography within species (Olsen and Schaal, 1999). A possible reason for this discrepancy is that the first study observed variation from cDNAs in a limited taxonomic sample, while the second observed variation among multiple cultivars in the tropical crop plant cassava using a smaller piece of the full GapC that spanned 4 introns. The difference in reported substitution rates may indicate differences in mutation rates between exons and introns of the gene and wide applicability of *gpd* for phylogenetic studies at deep and shallow levels of resolution. These studies agree that *gpd* is a very promising single copy (or low-copy) nuclear gene for plant systematics (Olsen and Schaal, 1999; Schaal and Olsen, 2000; Strand et al., 1997). In the present study, I develop *gpd* for use in phylogeny reconstruction of mosses at various taxonomic levels, but especially within lineages of the poorly known moss, *Mitthyridium.*

1.2. The study organism, Mitthyridium

Mitthyridium belongs to the tropical moss family Calymperaceae and is monophyletic (La Farge et al., 2000; Wheeler et al., in press). The group is endemic to the paleotropics, an uncommonly restricted geographic

range for a moss genus. It is most diverse in the Malay Peninsula and may have spread across both the paleotropical Pacific and Indian Oceans from this center of diversity, although this hypothesis remains untested. While previous taxonomic treatments of *Mitthyridium* make it possible to recognize major morphological entities in the group (Eddy, 1988; Nowak, 1980) the genus is notorious for its difficult taxonomy as circumscriptions of species are weak, debated, and often changed (Reese, 1994; Reese et al., 1994; Reese et al., 1986). The difficult taxonomy of *Mitthyridium*, a reflection of the graded phenotypic variation, and the relatively limited distribution range of *Mitthyridium* have led previous authors to suggest that the group is recently derived and in the process of rapid diversification (Reese et al., 1986). If *Mitthyridium* is a young lineage, then variation at the molecular level should also be limited, especially in more slowly evolving genes. Wheeler et al. (in press) demonstrated a lack of variation in the chloroplast gene, *rbc*L, across 5 *Mitthyridium* species. Resolving relationships in this group evidently requires more rapidly evolving molecular markers than *rbc*L. Here, I compare levels of sequence variation and phylogenetic utility of the nuclear gene *gpd* with the better-known, chloroplast markers, *rps*4 and *trn*L.

2. Materials and methods

2.1. Taxon selection

2.1.1. Ingroup dataset

Twenty-six *Mitthyridium* exemplars were chosen to represent the range of morphological variations, after examination of 300 samples collected from across the geographical range of the genus (Table 1). Floristic keys, taxonomic treatments (Eddy, 1988; Nowak, 1980; Reese et al., 1986), and general specimen examination were used to guide the selection of specimens for analysis; specimens were stratified into distinct morphological groups from which samples were chosen randomly. This sampling strategy was used in an attempt to ensure that at least one example of each morphotype in *Mitthyridium* was used as an operational taxonomic unit for phylogenetic analyses.

2.1.2. Outgroup dataset

The outgroup taxa were chosen from within the newly clarified moss family Calymperaceae, to which *Mitthyridium* belongs (Wheeler et al., in press). Exemplars of 13 paleotropical taxa and one neotropical taxon within the traditionally recognized but

Table 1
Mitthyridium exemplars used in the molecular analyses

Genomic I.D.	Taxon	Geographic locality	Voucher information	GB *rps*4	GB *trn*L	GB *gpd*
M218	*junquilianum*	New Caledonia	DPWall 1086	AY046977	AY047014	AF400898
M114	*obtusifolium*	Moorea	DPWall	AF226733	AF231142	AF400899
M221	*undulatum*	Fiji	DPWall 2439	AY046978	AY047015	AF400900
M230	*constrictum*	Samoa	DPWall 2978	AY046979	AY047016	AF400902
M264	*subluteum*	Gabon	Orban 1995	AY046980	AY047017	AF400903
M292	*fasciculatum*	Australia	Mishler 7/21/1998	AY046981	AY047018	AF400916
M246	*fasciculatum*	Australia	Mishler 7/23/1 998	AY046982	AY047019	AF400919
M342	*perundulatum*	Papua New Guinea	Streimann 40876	AY046983	AY047020	AF400924
M364	*junquilianum*	New Caledonia	DPWall 1043	AY046984	AY047021	AF400928
M367	*perundulatum*	Borneo	DPWall 3673	AY046985	AY047022	AF400930
M368	*perundulatum*	Borneo	DPWall 3653	AY046986	AY047023	AF400931
M369	*fasciculatum*	Borneo	DPWall 3652	AY046987	AY047024	AF400932
M373	*luteum*	Borneo	DPWall 3685	AY046988	AY047025	AF400933
M377	*luteum*	Borneo	DPWall 3637	AY046989	AY047026	AF400936
M391	*luteum*	Peninsular Malaysia	DPWall 3885	AY046990	AY047027	AF400947
M395	*undulatum*	Peninsular Malaysia	DPWall 3864	AY046991	AY047028	AF400950
M396	*repens*	Borneo	DPWall 3784	AY046992	AY047029	AF400951
M401	*repens*	Peninsular Malaysia	DPWall 3848	AY046993	AY047030	AF400953
M404	*luteum*	Peninsular Malaysia	DPWall 3830	AY046994	AY047031	AF400955
M405	*microundulatum*	Peninsular Malaysia	DPWall 3859	AY046995	AY047032	AF400956
M422	*obtusifolium*	Fiji	DPWall 2620	AY046996	AY047033	AF400959
M423	*undulatum*	Fiji	DPWall 2546	AY046997	AY047034	AF400960
M425	*undulatum*	Fiji	DPWall 2541	AY046998	AY047035	AF400961
M433	*microundulatum*	Seychelles	Orban 1995	AY046999	AY047036	AF400967
M803	*constrictum*	Vanuatu	Streimann 63086	AY047000	AY047037	AF400968
M809	*constrictum*	Society Islands	Wall UC herbarium	AY047001	AY047038	AF400972

Exemplars were chosen a priori on the basis of morphological distinctness after examination of 300 specimens collected from across the geographic range occupied by *Mitthyridium*. No attempt was made to name semaphoronts; only genomic accession number was used in phylogenetic analyses. GB, GenBank accession numbers for the three data partitions, *rps*4, *trn*L, and *gpd*.

polyphyletic genus *Syrrhopodon* were chosen as the outgroup. Wheeler et al. (in press) identified certain *Syrrhopodon* species as the closest outgroup to *Mitthyridium*, but the taxa used in that study (*Syrrhopodon fimbriatulus* and *S. gardneri*) were phylogenetically distant. Inclusion of a large number of possible outgroups was warranted because *Syrrhopodon* is the largest genus in the family Calymperaceae, lacks taxonomic clarity, and is polyphyletic (Wheeler et al., in press) (Table 2).

To further ensure the identification of the sistergroup to *Mitthyridium*, 6 taxa more distantly related to *Mitthyridium* than the 13 *Syrrhopodon* were selected on the basis of their position near *Mitthyridium* in the phylogenetic results reported by Wheeler et al. (in press) (Table 1). *Calymperes* was considered too phylogenetically distant to *Mitthyridium* and was not included. A dataset was created that included all in- and outgroup taxa.

2.2. DNA isolation

Total genomic DNA was extracted from field-collected and herbarium specimens using DNeasy Plant Mini Kits (Qiagen, Chatsworth, CA, USA) following manufacturer's protocol. For each specimen, a sample of young, apical meristematic tissue was carefully examined under a dissecting scope to detect and remove any attached foreign tissues. Only single ramets were used. Vouchers are deposited in the University Herbarium at the University of California, Berkeley (UC) (Tables 1 and 2).

2.3. Gene selection and primer design

Primers uniC and uniF were used to amplify the chloroplast region *trn*L (Taberlet et al., 1991)—a region that spans one *trn*A and an intergenic spacer. Forward primer *rps*5′ and reverse primer *trn*S were used to amplify chloroplast gene *rps*4, which encodes a small ribosomal protein (Souza-Chies et al., 1997). Finally the nuclear gene glyceraldehyde 3-phosphate dehydrogenase was identified as an ideal nuclear gene candidate. Although other authors have designed primers for glyceraldehyde 3-phosphate dehydrogenase (*gpd*) (Olsen and Schaal, 1999; Strand et al., 1997), these primers preferentially amplified bacterial contaminants in the mosses studied here (Wall, personal observation). Therefore, new primers were designed.

Two primers (forward primer—*GPD 1790F* **GTC TTC ACY GAC AAR GAC AAG GCT** (24 bases); reverse primer—*GPD 3050R* **CTG TAA CCC CAR TCR TTG TCY TAC CA** (26 bases)) were designed to amplify a portion of *gpd* using the transcript for the moss *Physchomitrella patens* (X72381) together with sequences of *gpd* from *Selaginella lepidophylla* (U96623), *Arabidopsis thaliana* (M64119), *Nicotiana tabacum* (M14419), *Zea mays* (U45855), as well as the fungi *Glomerella cingulata* (M93427), *Neurospora crassa* (U56397), and *Aspergillus nidulans* (M19694) to prevent aberrant amplification of fungal sequence. Because of this broad phylogenetic spectrum, the primers are expected to be widely applicable to all major clades of land plants. Boundaries of the exons and introns were determined by comparison with the full GapC sequence of

Table 2
Outgroup taxa used in the molecular analyses

Genomic	Taxon	Geographic locality	Voucher information	GB *rps*4	GB *trn*L	GB *gpd*
G215	*Syrropodon loreus*	Australia	Tangney RT-01-B	AY046965	AY047002	AF400977
G200	*Syrropodon apertus*	New Caledonia	DPWall 1628	AY046973	AY047010	AF400980
G440	*Syrropodon croceus*	Australia	Streimann 64497	AY046970	AY047007	AF400979
G319	*Syrropodon mahensis*	Seychelles	Orban 1995	AY046966	AY047003	AF400978
G136	*Leucophanes albescens*	Moorea	DPWall 6/26/1997	AF226751	AF231158	AY135222
G318	*Leucophanes seychellarum*	Seychelles	Orban 1995	AY046974	AY047011	AY135223
G241	*Leucophanes glaucum*	Australia	Mishler 7/23/1998	AF226752	AF231159	AY135224
G240	*Syrropodon confertus*	Australia	Mishler 7/23/1998	AF226743	AF231151	AY135225
G261	*Exodictyon incrassatum*	Fiji	DPWall 2527	AF226776	AF231189	AY135226
G243	*Exostratum blumei*	Australia	Mishler 7/24/1998	AF226753	AF231160	AY135227
G244	*Arthrocormus*	Australia	Mishler 7/24/1998	AF226750	AF231157	AY135228
G437	*Syrropodon perarmatus*	Vanuatu	Streimann 62600	AY046967	AY047004	AY135229
G441	*Syrropodon muelleri*	Australia	Streimann 64404	AY046976	AY047013	AY135230
G439	*Syrropodon trachyphyllus*	Australia	Streimann 64418	AY046969	AY047006	AY135231
G116	*Syrropodon banksii*	Moorea	Wall UC herbarium	AF231148	AF226740	AY135232
G826	*Syrropodon prolifer*	Xalapa	DPWall 5/14/1999	AY046975	AY047012	AY135233
G446	*Syrropodon albovaginatus*	Samoa	DPWall 3098	AY046971	AY047008	AY135234
G447	*Syrropodon ciliatus*	Fiji	DPWall 2935	AY046972	AY047009	AY135235
G438	*Syrropodon armatus*	Australia	Streimann 64094	AY046968	AY047005	AF400982
G247	*Syrropodon fimbriatulus*	Australia	Mishler 8/29/1998	AF226742	AF231150	AF400981

The taxa were chosen in part on the basis of previous phylogenetic results (La Farge et al., 2000; Wheeler et al., in press). GB, GenBank accession numbers for the three data partitions, *rps*4, *trn*L, and *gpd*.

Arabidopsis thaliana and the cDNA sequence for *Physcomitrella patens.*

2.4. PCR and sequencing strategies

PCR reaction mixtures each contained 0.5 units of AmpliTaq Gold Polymerase (PE Applied Biosystems), 5 μl of the supplied 10X Buffer II, 0.1 mM each dNTP, 1.25 mM $MgCl_2$, and 1.25 mM of each primer.

MJ Research DNA Engine Thermal Cycler (MJ Research) was programmed to run the following PCR cycle: an initial hot start at 95 °C for 12 min; then, 45 cycles of 95 °C for 1 min, 58.5 °C for 1 min, and 72 °C for 1 min 30 s. A 7-min 72 °C extension step terminated the run. Reactions were stored at 4 °C. Products were visualized with ethidium bromide on 1% agarose gel. Amplicons were purified with kits (Qiagen, Chatsworth, CA) and then processed by cycle sequencing using BigDye-Terminator chemistry (PE Applied Biosystems) on an ABI model 377 automated fluorescent sequencer in the Molecular Phylogenetics Laboratory at the University of California, Berkeley.

2.5. Sequence manipulation and database assembly

The initial sequences from each amplicon were compared to GenBank, EMBL, and DDBJ databases using BLAST for early detection of mistakenly amplified sequences. Sequence files were aligned by eye using the program Sequence Navigator (PE Applied Biosystems) or directly in NEXUS format; two NEXUS files were created, one for the ingroup (26 *Mitthyridium* taxa—Table 1) and another for the inclusive compartment (*Mitthyridium* taxa (Table 1) plus outgroup taxa (Table 2)). Coding regions (i.e., *rps*4 and exons of *gpd*) were translated into amino acid sequences as an internal check on the accuracy of each edited nucleotide sequence. Alignments of both cpDNA regions, *trn*L and *rps*4, were unambiguous and the same for both the ingroup and inclusive compartments. However, because of the variability found in *gpd*, it was necessary to align *gpd* differently among the ingroup and inclusive compartments. Also, in the inclusive NEXUS file a character set that divided *gpd* into exons and introns was created to check for effects of intron variability on phylogenetic results.

Insertions/deletions (indels) of *gpd* were coded into binary characters to ensure their contribution to the phylogenetic outcome; the indel sequences themselves were also used in the analyses; their alignment across taxa was unambiguous. All sequences were submitted to GenBank (Tables 1 and 2).

2.6. Phylogenetic analyses

PAUP* 4.0 (Swofford, 2000) was used for all parsimony, likelihood, and decay analyses of the data partitions separately and in combination. Gaps were treated as missing data. In all heuristic searches using parsimony, starting trees were obtained via random addition and branch swapping was performed using tree–bisection–reconnection (TBR). At least 100 replicate searches were conducted for all analyses.

Whenever small enough (that is <7000), the set of most parsimonious trees was sorted on the basis of likelihood score using the nucleotide substitution parameters defined in the HKY-85 model with Γ-distributed rate variation. The trees with the single highest likelihood score were chosen for display.

2.6.1. Incongruence

To test for incongruence among the 3 data partitions, *rps*4, *trn*L, and *gpd*, the partition homogeneity was implemented (Farris et al., 1995; Kellogg et al., 1996; Mason-Gamer and Kellogg, 1996). One thousand replicates were used for each partition to generate the null distributions. All partition homogeneity tests were performed using PAUP* 4.0 (Swofford, 2000).

2.6.2. Combined analyses

Topological incongruence between trees based on different genes can be a reflection of either different history (Maddison, 1997) or some kind of systematic error (Swofford et al., 1996). Either problem may cause a rejection of the null hypothesis in a partition homogeneity test. Therefore, phylogenetic analyses were conducted on all combinations of the three genes, regardless of the results obtained from the homogeneity test. In some cases, a series of strict consensus analyses was conducted on trees derived from the separate gene partitions to look for regions of incongruence.

2.6.3. Character support

Decay indices (also known as Bremer support values) implemented in TreeRot.v2 (Sorenson, 1999) were performed to provide measures of support for each node. Values of zero were not illustrated.

Genes that share the same organismal history but differ greatly in rate of mutation may appear parts of separate process partitions. To assess whether mutation rate accounts for heterogeneity between data sets it is instructive to examine what characters support particular nodes on a phylogeny. Therefore, *gpd* exons, *gpd* introns, *rps*4, and *trn*L were separately optimized onto the *inclusive* total evidence phylogeny using PAUP* 4.0. The branch length data were gathered for each data partition excluding uninformative characters. Those branch length data were then sorted by node order from the tips to the base and placed into node classes according to this sorting. Class "0" represented the branch length from the terminal taxa to the first coalescent event, class 1 represented the branch length from the first to the second node, and so on to class "4." A

histogram was used to show the relative character support per data partition within each node class.

2.7. Tests for recombination and molecular selection in glyceraldehyde 3-phosphate dehydrogenase

PLATO (Grassly and Rambaut, 1998) was used to detect anomalously evolving regions within complete *gpd* sequences for the ingroup taxa. This program uses a sliding window of varying sizes to find regions of an alignment that reject a global phylogenetic hypothesis calculated across all sites given a tree (Fig. 4d) and a model of sequence evolution (in this case HKY-85 + Γ rate heterogeneity). PAML (Yang, 2000) was used to determine rates of nonsynonymous (dn) and synonymous (ds) substitutions and their ratio (ω). This estimation of ω was by the method of Yang and Nielsen (2000) (equal weighting of pathways).

2.8. Test for evolutionary rate constancy in glyceraldehyde 3-phosphate dehydrogenase

The likelihood-ratio tests for rate constancy of molecular evolution were conducted on the set of equally parsimonious trees found using total evidence (i.e., using all data partitions). Trees derived from total evidence were judged to be the most robust and possibly the most accurate representation of the organismal history and thus best for conducting tests for evolutionary rate constancy in *gpd*. However, a set of tests was conducted in which the trees used to evaluate rate constancy were built by the exact same data partition (either the full *gpd* or *gpd* excluding introns) whose likelihood was being assessed. In no instance was the level of significance either for or against the null model affected by such topological differences between trees based on different data partitions. At most, use of the total evidence phylogeny biased the results against favoring the null model of rate constancy, making the tests more conservative.

A test for rate constancy was conducted on the ingroup dataset and a taxon compartment that contained the 26 *Mitthyridium* exemplars as well as 4 other taxa (G215, G200, G319, and G440) that were found to be the closest relatives to *Mitthyridium* based on the phylogenetic analyses presented. Additional outgroup taxa were added singly and the likelihood-ratio tests were conducted in the same fashion as described below. The addition of taxa, guided by the phylogenetic results of the inclusive data compartment, started with G247 and G438 (Table 2) and proceeded until the null of rate constancy was rejected. A HKY-85 model with Γ-distributed rate variation was used as the model of sequence evolution in all tests. This model was found to best explain the data after performing a series of likelihood ratio tests on different models of sequence evolution.

Variation in rates across lineages was examined by using a tree-wide likelihood ratio test to compare rate-constant and rate variable models of molecular evolution (Felsenstein, 1988; Huelsenbeck and Rannala, 1997). Formulae for determining degrees of freedom for the test of rate constancy across lineages assume a fully dichotomous tree (Felsenstein, 1988). Degrees of freedom for the test of rate constancy across lineages are equal to the difference between the number of parameters in the rate constant and rate variable models. In the rate constant model for the ingroup data compartment there were 22 internal node ages and one rate parameter (23 parameters); in the rate variable model there was one parameter for each branch length on the unrooted topology (50 parameters), leaving 27 degrees of freedom. The degrees of freedom for the 30 taxon data compartment (ingroup + 4 outgroups) were calculated in the same way—29 parameters in the constrained model, 58 in the unconstrained—and totaled 29. The degrees of freedom were adjusted accordingly with the addition of other outgroups (starting with G247). The same likelihood ratio tests were conducted on *gpd* data with and without the introns removed.

3. Results

3.1. gpd Structure, size, and composition

3.1.1. General

Fig. 1 is a diagram of the region of the *gpd* gene used in the present analysis. Although the whole gene in *Arabidopsis* consists of approximately 2705 bp spanning eight introns and nine exons, the region shown in Fig. 1

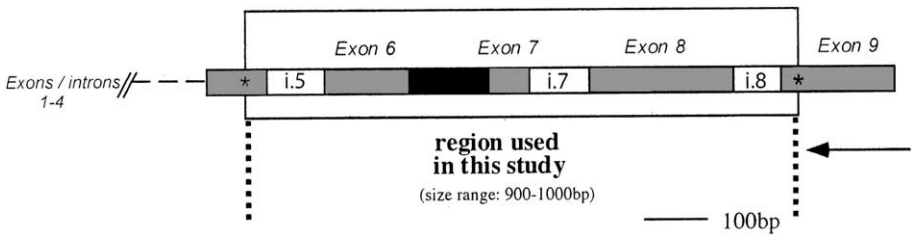

Fig. 1. Schematic of *gpd*. Asterisks indicate the position of the forward and reverse primers. The dark region in the intron 6 indicates a purine-repeat motif that varied in number among the taxa sequenced. Intron 6 also contains 3 of the 6 indels found throughout the sequences studied.

was chosen for its size (~1000 bp) and consequent ease of sequencing using only two primers (See Section 2). In no instance did the primers amplify fungal contaminants.

The length of the section sequenced for the present analysis varied from 891 to 1007 bp. The bulk of the length variation was found in intron 7. Only small portions of the exons 5 and 9 were sequenced; exons 6–8 were sequenced entirely and together consist of 183 amino acids (totaling 549 bp; Fig. 1). The exons, though a rich source of nucleotide variation, were invariant in length among the taxa studied here.

3.1.2. Insertion/deletions

The introns of the gene *gpd* were rich in indels (insertions/deletions). Among the ingroup taxa, six informative indels were found, three in the first 100 bp of amplified sequence (intron 5), and three in the region from 390 to 436 bp (intron 6). The indels ranged in length from six to 16 bp, but did not vary in length across taxa. In the exons, two amino acid indels were found, one at the start of exon 6 and one at the end, just before the start of intron 6. Alignment of the indels was unambiguous among the ingroup taxa.

The alignment of *gpd* across the ingroup and inclusive datasets differed slightly especially with regard to indel characteristics. Of the six indels identified in the ingroup dataset only the third was found among the 20 outgroups. Specifically, taxa G318, G240, G241, G244, G261, G143, and G136 (Table 2) possessed sequence at the third indel, whereas the other 13 outgroup taxa lacked sequence at this indel.

3.1.3. Base frequencies

The base frequencies were largely identical across all taxa included in both the ingroup and inclusive datasets (Tables 1 and 2). The average base frequencies were nearly equal for each base at 0.23579 (A), 0.24977 (C), 0.27180 (G), and 0.24264 (T) (exons and introns combined). The introns were rich in G and C, with average frequencies at 0.24356 (A), 0.13965 (C), 0.30705 (G), and 0.30974 (T). Intron 6 contained a large purine-repeat region, consisting of variable repeats of an AGG motif, perhaps useful as a microsatellite. Also, intron 6 was found to be the largest of the four introns sequenced (Fig. 1) and the intron responsible for a large percentage of the length variation found in *gpd* among the 46 taxa in this study.

3.2. gpd Sequence divergence

3.2.1. Ingroup

The pairwise distances found for each of the three genes differed markedly (Fig. 2). The pairwise distances of *gpd* varied from 0.036 to 0.067 with an average pairwise distance of 0.047 (Fig. 2). Conversely, *rps*4

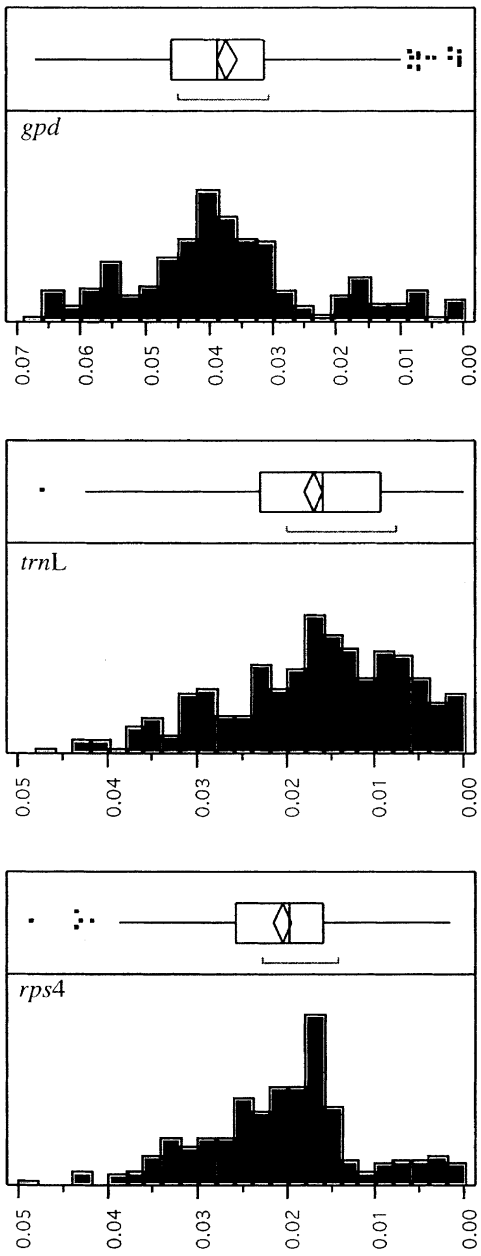

Fig. 2. Histogram of pairwise distances between all pairs within the exclusive, 26 taxon dataset (see Table 1). A boxplot is provided to the right of each histogram and shows the mean, quartiles, and outliers of each set of pairwise distances.

pairwise distances varied from a minimum of 0.0 to a maximum of 0.0489 with an average of 0.0235. *trn*L pairwise distances varied from 0.00 to 0.0425, mean = 0.0194 (Fig. 2).

3.2.2. Inclusive

In the inclusive dataset, *gpd* pairwise distances ranged from 0.0332 to 0.201 (mean = 0.106). The average pairwise differences of *gpd* with and without the introns removed changed significantly from 0.11 to 0.066, respectively. In *rps*4, the pairwise divergences ranged

from 0.00 to 0.103 (mean = 0.048) and from 0.00 to 0.088 (mean = 0.036) in *trn*L. Fig. 3 displays all pairwise divergences from the inclusive dataset across all three genes. Each histogram demonstrates a bimodal distribution; the higher peak corresponds to the comparisons among more distantly related taxa and the lower peak corresponds to the more closely related taxa (i.e., *Mitthyridium* exemplars). The average pairwise distances among outgroup taxa were 0.108 in *gpd*, 0.056 in *rps*4, and 0.049 in *trn*L.

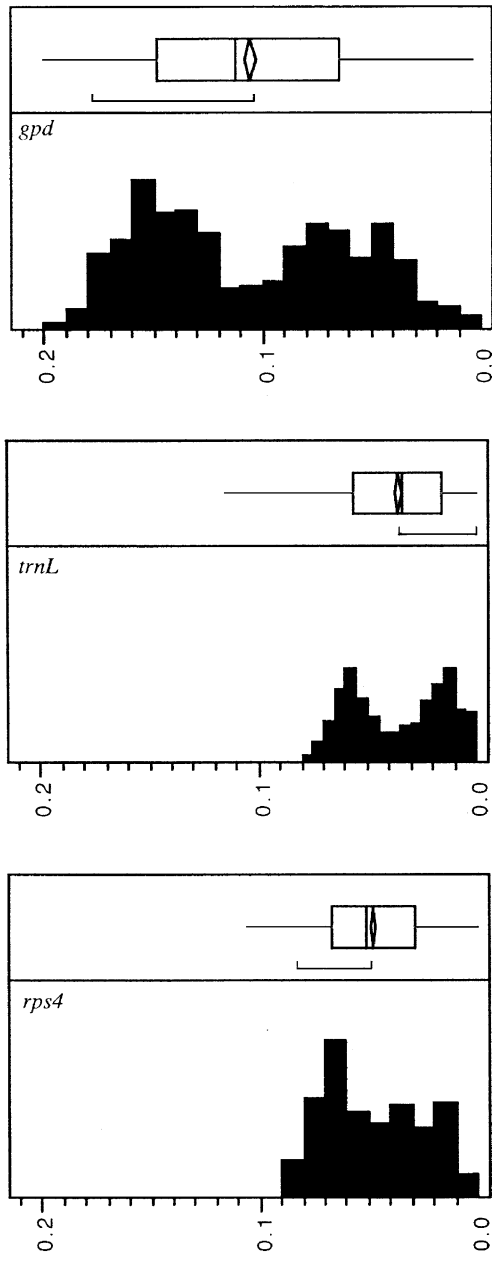

Fig. 3. Histogram of pairwise distances between all pairs within the inclusive, 46 taxon dataset (see Table 2). A boxplot is provided to the right of each histogram and shows the mean, quartiles, and outliers of each set of pairwise distances.

3.3. Phylogenetic analysis

For the purpose of comparison between the phylogenies of the ingroup and inclusive data compartments, four clades were identified as A, B, C, and D. The composition of those clades is shown in Fig. 4a.

3.3.1. Ingroup dataset phylogenetic results

All ingroup trees were rooted using clade A (M230, M803, and M809) after resolving their position at the base of the *Mitthyridium* clade in the inclusive analyses. Maximum parsimony analyses of the *gpd*, *rps*4, and *trn*L data partitions produced three different topologies (Figs. 4a–c) whose main differences were: the position of clade B, the presence of clade C, and the positions of taxa M218, M364, and M342 (Fig. 4). The *trn*L tree lacked much resolution but remained largely congruent with the *rps*4 and *gpd* trees (Fig. 4c). The major toplogical incongruencies in the separate gene phylogenies were among branches in clades C and D and at the node clarifying the relationships among clades B, C, and D. Fig. 5 juxtaposes the *rps*4 and *gpd* topologies to indicate the main differences after branches with decay values of zero were collapsed. The minor differences between trees based on the two partitions were the positions of taxa M218, M364, M342, and M264 (Fig. 5).

A partition homogeneity test demonstrated that *rps*4 and *gpd* were compatible ($p = 0.12$). Combining *rps*4 and *gpd* produced 8 equally parsimonious trees of length 290, (consistency index (CI) = 0.8414), which retained components of both data partitions such as the position of clade B found using *gpd* alone and the position of taxa M218, M364, and M342 found with *rps*4 alone. Many of the relationships were strongly supported, with an exception being the node distinguishing clades C and D as sister to clade B (similar to the result shown in Fig. 4a). The partition homogeneity test indicated that *trn*L was incongruent with both *rps*4 and *gpd*, respectively, and that all three genes when combined were incongruous ($p < 0.01$).

A combined analysis of the three data partitions reconstructed 4 equally parsimonious trees of length 367 (Fig. 4d). This combined tree contained elements of both the nuclear- and chloroplast-derived phylogenies. Specifically, this total evidence tree differed from the *gpd* topology again in the placement of the M218, M364, M264, and M342, the result discovered when the *gpd* topology was examined against *rps*4 (Fig. 5). Clade B differed in position from that found in the combined *rps*4/*gpd* tree and the tree derived from the *gpd* data alone (Fig. 4a), but was identical to its position in the *rps*4 tree (Fig. 4b). The position of clade B, however, was relatively poorly supported (decay = 1; Fig. 4d).

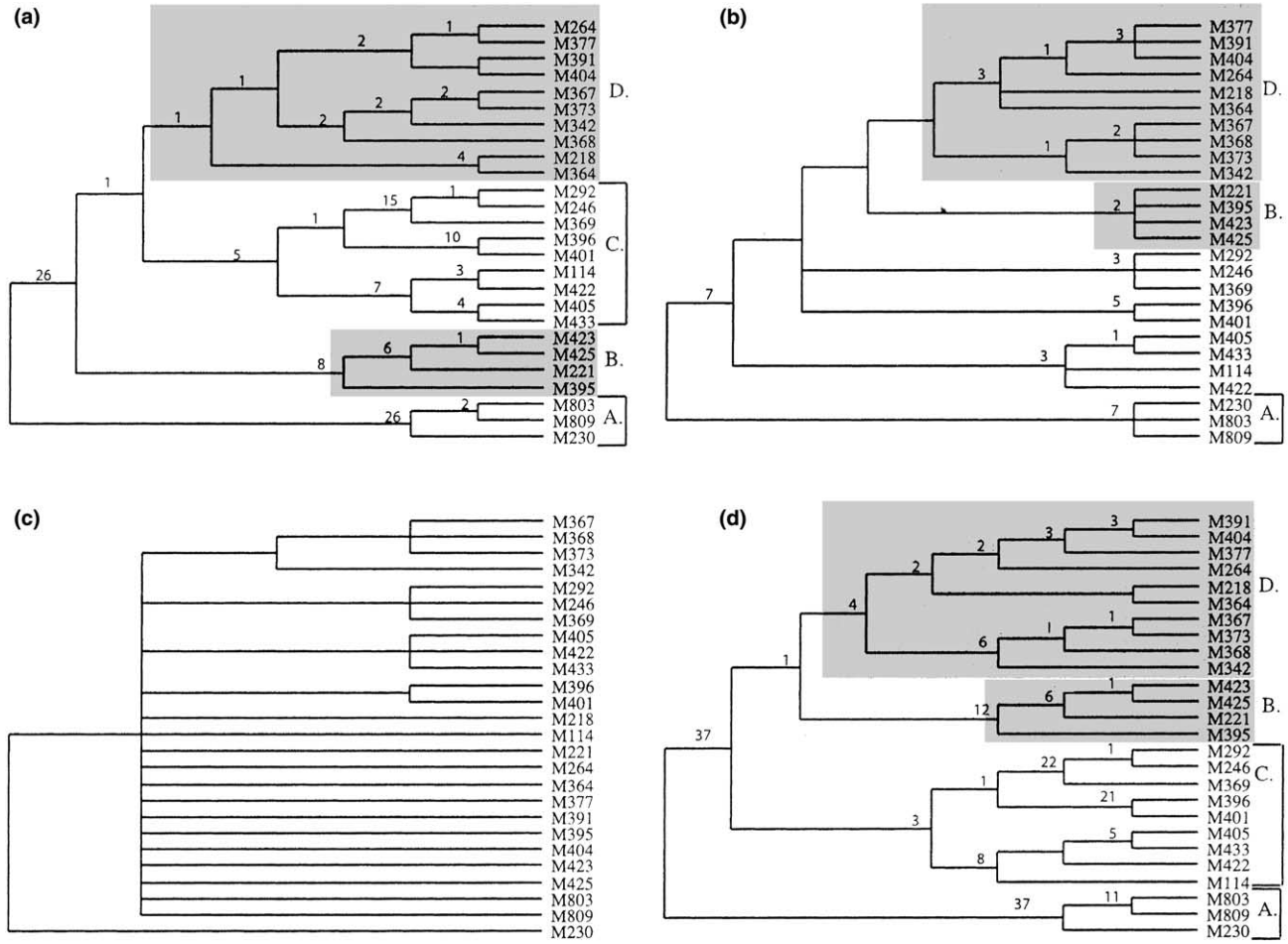

Fig. 4. Maximum parsimony trees for each of the three data partitions and a total evidence analysis. Common clades are designated by a letter code, A–D. Those letters serve as the basis for comparison in later analyses. Trees are rooted with clade A. (a) Tree based on the nuclear gene *gpd*. Tree length = 202; CI = 0.8168. (b) Strict consensus of 234 equally parsimonious trees based on the chloroplast gene, *rps*4; tree length = 82; CI = 0.9634. (c) Strict consensus of 1390 equally parsimonious trees based on the chloroplast gene, *trn*L; tree length = 54, CI = 0.8148. (d) 1 tree with the highest likelihood score of 4 maximally parsimonious trees (length = 367; CI = 0.7847) from combined analysis of the 3 data partitions.

3.3.2. Inclusive dataset phylogenetic results

gpd Alone. Maximum parsimony analysis of the 46 taxon dataset revealed 2 equally parsimonious trees of length 805 (CI = 0.6733). The most likely of those two trees is shown in Fig. 6a, although the two trees differ only in the position of taxon M292. Given the large difference between pairwise distances with and without introns (described above), a maximum parsimony analysis of *gpd* with no introns was conducted to compare with the trees derived from *gpd* with introns. This no-intron analysis produced 16 trees of length 476 (CI = 0.6912) (Fig. 6b). Those trees were sorted using maximum likelihood; the tree with the highest likelihood score differed minimally from the tree based on the full *gpd* data partition. The only difference found was in the placement of taxon M264. Clades A–D all were reconstructed using both partitions and the topology of *Mitthyridium* was found identical to that shown in the ingroup total evidence tree described above (Fig. 4d). No topological differences were found among the outgroup taxa. These trees and subsequent inclusive analyses were rooted using the clade containing G439, G116, G826, G446, and G447.

rps4 Only. An analysis of *rps*4 data alone yielded 6281 equally parsimonious trees of length 375 (CI = 0.6453). These trees were sorted using likelihood (HKY-85 with Γ-distributed rate variation), finding 161 equally likely from the set of 6281 most parsimonious. The consensus of those 161 trees is shown in Fig. 7a. The position of clade B differed from that found in previous analyses, this time embedded within clade C (itself no longer monophyletic to the exclusion of clade B). However, this relationship dissolved in the strict consensus of all 6281 most parsimonious (Fig. 7b), as there is general lack of resolution of the shallow splits.

*gpd vs. rps*4. A consensus of the 161 *rps*4 trees (the most likely of the 6281 most parsimonious) and the trees derived from the full *gpd* data (introns included) produced a topology identical to the full consensus of the 6281 *rps*4 trees alone (Fig. 7b). The topologies of the

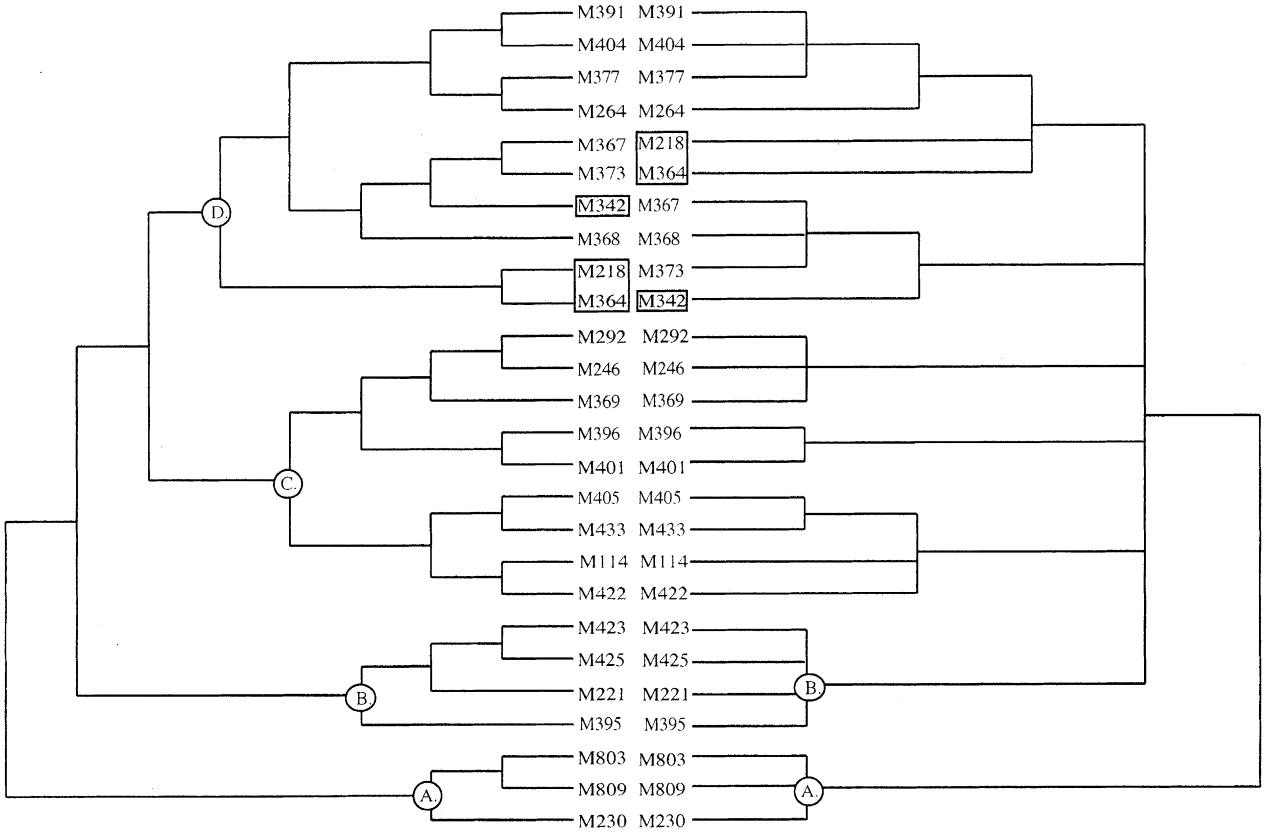

Fig. 5. The *gpd* (left) and *rps*4 (right) phylogenies juxtaposed. Topological differences are in boxes. All branches with decay indices of zero were collapsed. Fig. 4 shows decay values on all nonzero branches.

ingroup portions of the *rps*4-derived and *gpd*-derived trees differed primarily at the deeper splits, such as in the placement of clade B and the presence of clade C. The topologies did differ in minor ways at the shallower splits, such as with the placement of taxon M342 (Fig. 7b). However, the topology of the outgroup taxa differed only among the branches within the clade containing G439, G116, G446, G447, and G826 (Fig. 7b).

*trn*L *Alone.* A maximum parsimony analysis yielded 156,947 trees of length 243. *trn*L again failed to provide adequate variation for branches in the ingroup portion of this larger analysis. However, the topology was not in conflict with *rps*4 or *gpd*, especially among the deeper splits of the outgroup taxa. As with *rps*4, clades compatible with the *gpd* phylogeny were found among the ingroup taxa of this inclusive *trn*L consensus.

Phylogenetic congruence. A test for congruence among the three different data partitions revealed that only *rps*4 and *gpd* are combinable, as in the ingroup dataset described above. Although not statistically congruent with *rps*4 and *gpd* via the partition homogeneity test, *trn*L was fairly well resolved and concordant with the other two data partitions among the outgroup taxa.

*Combined analyses. gpd with rps*4—Maximum parsimony analysis of the combined *rps*4 and full *gpd* datasets yielded 8 trees of equal length (1205, CI = 0.6506). These were sorted using maximum likelihood to identify a single most likely tree (Fig. 8a). Clade B settled in a position topologically identical to that found in the ingroup total evidence phylogeny (Fig. 4d). *Total evidence*—Maximum parsimony analysis of the 46 taxon dataset revealed 12 equally parsimonious trees of length 1483 (CI = 0.6601). The most likely of those is shown in Fig. 8b. The total evidence and the *gpd*/*rps*4 maximum likelihood trees differed only in the position of taxa M342, M395, and M114; otherwise the trees were congruent. *cpDNA*—Maximum parsimony analysis of the 46 taxon dataset revealed 13,307 equally parsimonious trees of length 646 (CI = 0.6765). These trees differed minimally at the shallow splits from the combined *rps*4 and *gpd* results and presented no novel relationships.

3.4. Character support across the three data partitions

The number of parsimony informative characters differed considerably among the three data partitions and for both taxon compartments. For the ingroup alignment, *rps*4 had 41 parsimony informative (of 639 chars); *trn*L had 28 parsimony informative (of 540 chars), and *gpd* had 112 parsimony informative (of 867 chars). In the inclusive compartment (ingroup + out-

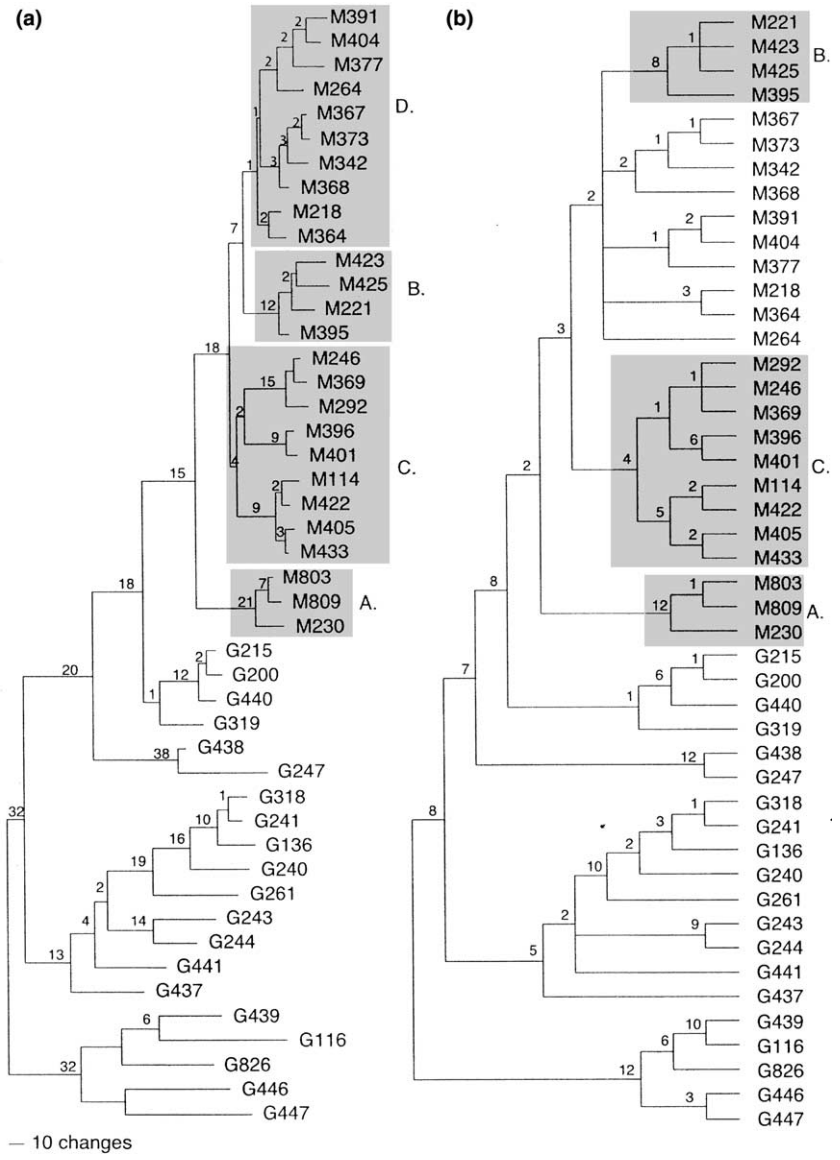

Fig. 6. Maximum parsimony trees based on the nuclear gene, *gpd* with and without introns, respectively. Decay values are indicated above branches, except when 0. (a) One of two maximum parsimony trees found using both exons and introns of *gpd* (TL = 805, CI = 0.6733). The tree presented here has the highest likelihood score, based on a HKY-85 substitution model with Γ-distributed rate heterogeneity. (b) Strict consensus of 16 trees of length 476 (CI = 0.6912) reconstructed using *gpd* with its introns removed.

group), *rps*4 had 129, *trn*L had 71, *gpd* (introns + exons) had 268, and *gpd* introns alone had 150 parsimony informative characters.

The fact that there were more informative characters in the introns of *gpd* than in the entire *rps*4 gene and far more than in *trn*L across the 46 taxa sampled suggested that each data set provided signal at potentially very different levels in the phylogenetic hierarchy. Despite the test for incongruence indicating the incompatibility of all three genes for combined analysis (owing only to the *trn*L data, as *gpd* and *rps*4 were found to correspond to the same process partition), characters were optimized separately onto the total evidence phylogeny and the branch lengths were separately charted. This demonstrated the various character contributions at progressively deeper nodes in the phylogeny (Fig. 9). *gpd* and especially *gpd* intronic characters were important at the shallowest levels of the phylogeny, while the slower evolving chloroplast genes proved largely invariant at the shallowest splits. *rps*4 was more variable than *trn*L at intermediate and deep splits, while *trn*L was most variable at the deepest splits (Fig. 9).

3.5. Tests for molecular selection and recombination in *gpd*

Fourteen anomalously evolving regions were discovered in the set of 26 ingroup gpd sequences (Table 3). In

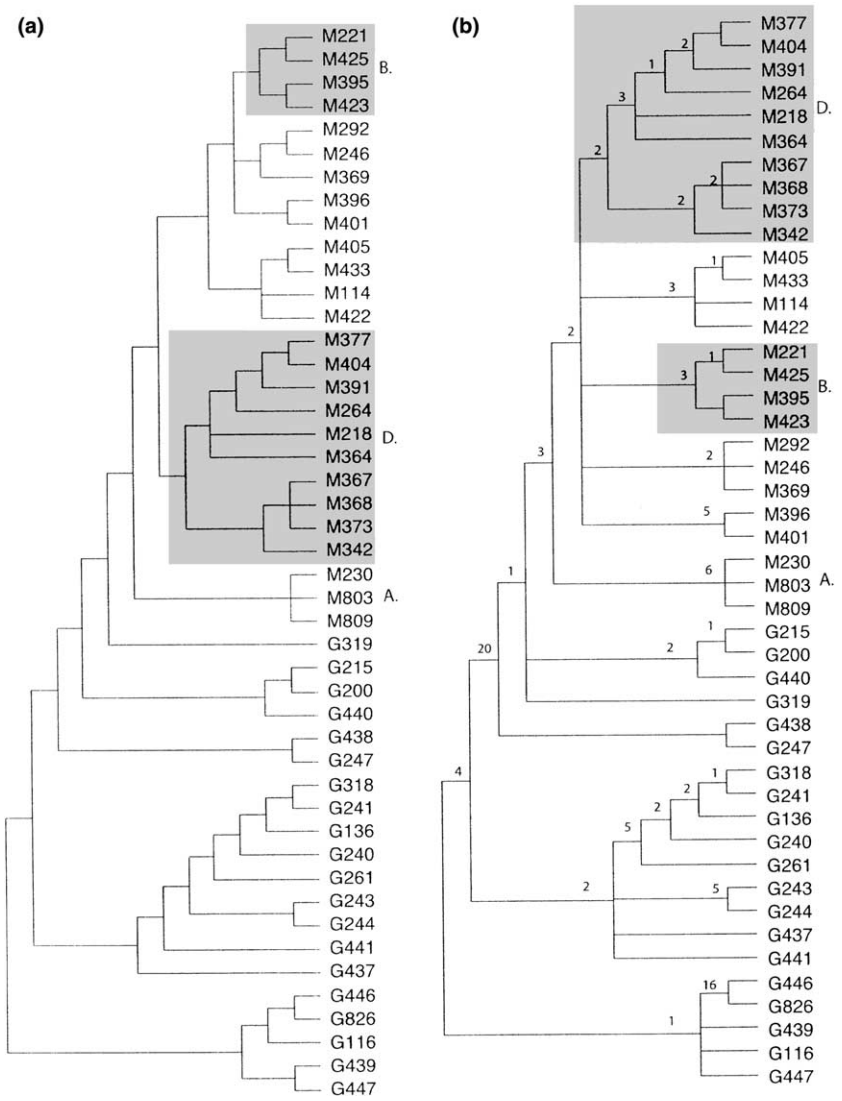

Fig. 7. Consensus of maximum parsimony trees derived from the analysis of the chloroplast gene, *rps*4; the trees are rooted using with the G446, G826, G447, G439, and G116. (a) Strict consensus of 161 most likely trees from a set of 6281 most parsimonious trees found by analysis of the *rps*4 data partition (TL = 375, CI = 0.6453). (b) Strict consensus of the 6281 most parsimonious trees found using *rps*4. Decay values are given above branches on the strict consensus of all most parsimonious trees. This tree shown is identical to the strict consensus of the 161 *rps*4 trees with the 8 maximally parsimonious trees found using *gpd* (introns included).

total, these anomalous regions composed approximately 80% of the full *gpd* sequence. Still the average degree of homoplasy in the anomalous regions (CI = 0.78) did not differ greatly from the average CI yielded by the non-anomalous regions (CI = 0.70) (Table 3).

The average d_N/d_S (ω) ratio in *gpd* for the 26 taxa was 0.437. However, some lineages were discovered to have ω ratios $\geqslant 1$. These were taxa M395, M405, and M433. The chloroplast gene *rps*4 was also found to have low ω, averaging 0.14 and no evidence of positive selection.

3.6. Tests of evolutionary rate constancy in gpd

The tests for rate constancy in the full *gpd* data partition were all significant ($p < 0.05$), indicating a lack of clock-like evolution for the ingroup data compartment and the ingroup + outgroup data compartment. However, with the introns removed the hypothesis of rate constancy within *gpd* could not be rejected in any instance at $\alpha = 0.05$. The four total evidence phylogenies found in the parsimony search described above for the 26 *Mitthyridium* exemplars were tested in turn. The $-2\ln$ likelihood ratio (LR) values were 36.1 and 39.6. The $-2\ln$ LRs for the 30 taxon compartment were smaller and ranged from 30.8 to 31.4 ($p > 0.10$). The tests in which taxon G247 was added (for which there were 12 MP trees described above) produced $-2\ln$ LR values that ranged from 42.8 to 50.4. Only 4 of the 12 MP trees did not allow rejection of the null of rate constancy. The addition of any other outgroup taxa

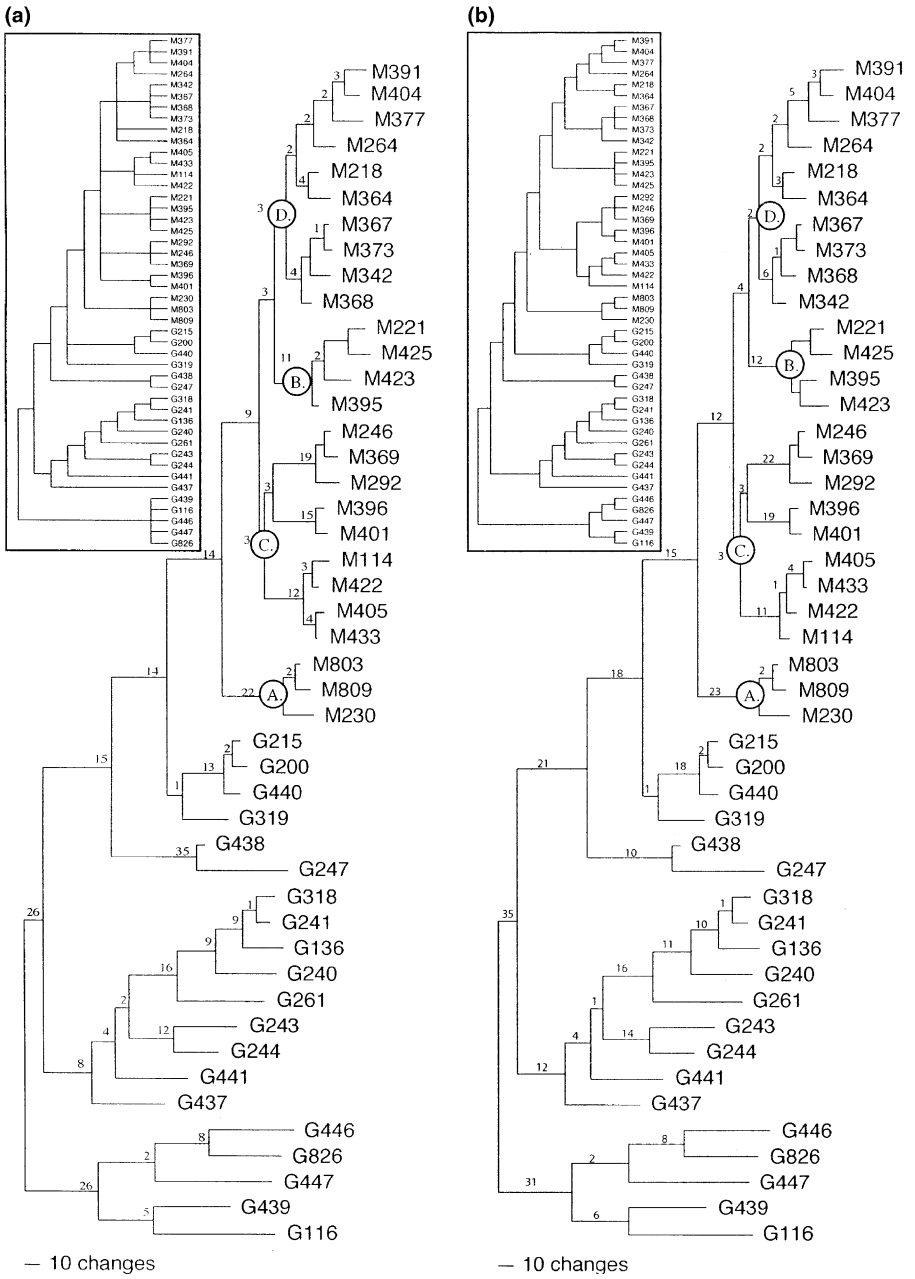

Fig. 8. Maximally parsimonious trees from analyses of combined data partitions; the trees are rooted using with the clade G446, G826, G447, G439, and G116. (a) One of the eight maximum parsimony trees found from analysis of the combined-gene data set—the nuclear gene, *gpd* and the chloroplast gene, *rps*4 (TL = 1205, CI = 0.6506). Inset to the left is the strict consensus of those 8 trees. (b) One of the 12 total-evidence phylogenies derived from simultaneous analysis of *gpd*, *rps*4 and the chloroplast region *trn*L (TL = 1483; CI = 0.6601). Inset to the left is the strict consensus of those 12 most parsimonious trees. The two phylograms shown here were found to have the highest likelihood score after sorting their respective sets of maximally parsimony trees using the model HKY-85 with Γ-distributed rate variation.

caused a dramatic rejection of the hypothesis of rate constancy, with the smallest $-2\ln$ LR = 94.3 (after having added G438).

4. Discussion

This study examined the utility of the nuclear gene *gpd* for phylogeny reconstruction in plants, with special emphasis the moss group *Mitthyridium*. While the gene is well known biochemically and has been used for phylogeny reconstruction in other organisms, it has not been widely used for plant systematics and has not been used before for phylogenetic studies within mosses. Two new primers were presented that amplify a portion of *gpd* small enough to be sequenced in two reads. The primers were designed to amplify *gpd* across all major lineages of green plants and to avoid errant sequencing

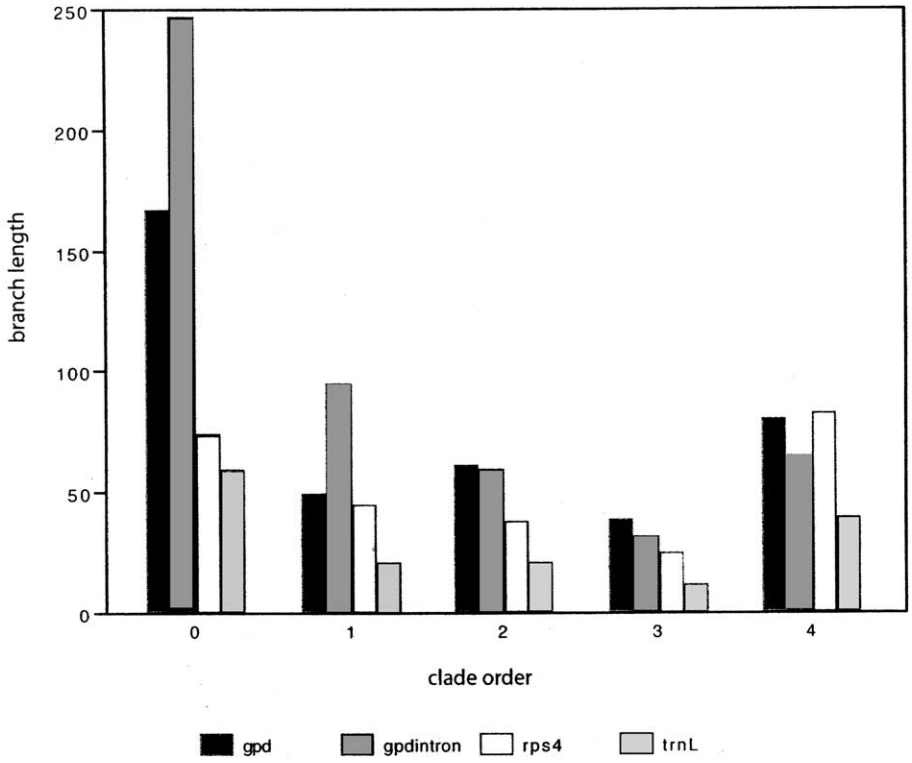

Fig. 9. The sum of branch lengths per node class after separate optimization of *gpd* (exons and introns), *gpd* introns alone, and the chloroplast markers *rps*4 and *trn*L characters to the branches of the total evidence *inclusive* phylogeny. Uninformative characters were excluded prior to optimization. The nodes were sorted from deep to shallow and placed into one of the 4 classes to indicate the hierarchical level of character support. Class "0" represents the branch length from the terminal taxa to the first coalescent event, class 1 represents the branch length from the first to the second node, and so on to class "4."

Table 3
Anomalously evolving regions in *gpd*

Coordinates	*Z*-value	CI	gpd-Region
31–35	28.569901	0.67	intron_1
52–56	52.330233	0.50	intron_1
63–69	59.512408	1.00	intron_1
79–99	106.889824	1.00	intron_1
102–110	104.499634	1.00	intron_1
115–146	138.378016	0.83	intron_1
149–155	80.844706	1.00	exon_6
160–167	85.892196	1.00	exon_6
171–371	331.689845	0.73	exon_6
379–386	97.327255	0.38	intron_6
411–425	97.428034	0.67	intron_6
436–489	163.126998	0.75	intron_6
522–558	166.859676	0.64	exon_7
566–943	366.628570	0.88	exon_7

Z-values were provided by PLATO using a HKY-85 + Γ rate heterogeneity model of sequence evolution. Regions in *gpd* are illustrated in Fig. 1. CI, consistency index.

of fungal contaminants, especially those that appeared to be close symbionts with mosses. The portion amplified spanned 4 introns and three complete exons that corresponded to the exon 6, 7, and 8 of 9 total exons (based on comparison with the completely sequenced *gpd* of *Arabidopsis*). The sequence variation in *gpd* among the ingroup was significantly higher than that found in the chloroplast DNA regions, *rps*4 and *trn*L. Consistency indices were similarly high across all 3 genes for the ingroup taxa compartment, averaging 0.83; the consistency indices for the inclusive compartment were lower, as expected with the increase in number of taxa, but again were nearly identical across the 3 genes. These results indicated that homoplasy was generally minimal in the data and that no sequence differed considerably from any other with regard to homoplasy. Still, the rate of sequence variation differed significantly between the exons and introns of *gpd*. Regardless, a qualitative assessment of saturation and homoplasy (Reed and Sperling, 1999; Zamudio et al., 1997) in which uncorrected pairwise distances are plotted against corrected divergences (Tamura and Nei, 1993) demonstrated that *gpd* did not become saturated even at pairwise divergences greater than 19%. The excision of the introns leaves a suitable number of parsimony informative characters for accurate phylogeny reconstruction, suggesting that if *gpd* did reach a saturation point when reconstructing history among a set of even more distantly related taxa than the set examined here, excision of introns may increase the homologous signal.

The portion of *gpd* sequenced here allowed for an accurate phylogenetic reconstruction of *Mitthyridium*

and its closest outgroups. This is the first appearance of a *Mitthyridium* phylogeny beyond a brief treatment that helped strengthen the hypothesis that *Mitthyridium* is monophyletic (La Farge et al., 2000). Although previous studies have found polyphyletic "Syrrhopodon"—a vast and poorly studied member of the Calymperaceae—to be the most likely candidate for the close outgroup to *Mitthyridium* (La Farge et al., 2000; Wheeler et al., in press), until now the specific outgroup taxa within "Syrrhopodon" had not been identified. This study has served to more securely identify the outgroups closest to *Mitthyridium* as *Syhrropodon apertus*, *S. croceus*, *S. mahensis*, and *S. loreus.* The present study is largely in accord with previous taxonomic concepts of *Mitthyridium* (Nowak, 1980; Reese et al., 1986). However, because the intent of the present paper was to examine the utility of *gpd* rather than study the phylogenetic taxonomy of *Mitthyridium*, the sampling was inadequate to suggest possible new taxonomies or test previous concepts (e.g., Reese, 1994). A more thorough treatment of *Mitthyridium*, including detailed descriptions of morphological and molecular evolution, is currently in preparation for publication and may be viewed online—http://ucjeps.herb.berkeley.edu/bryolab/students/dpwall/mono—as a phylogenetic monograph, the first of its kind.

The present study has further served to demonstrate the utility of beyond previous studies (e.g., Olsen and Schaal, 1999) in finding that *gpd* data are reasonably congruent with cpDNA. Not only does this congruence help corroborate that the genes share a common history presumably reflective of the organismal history, it shows that *gpd* may also be of use at deep taxonomic levels. Thus, *gpd* has the intronic variation suitable to resolve recent divergence events within *Mitthyridium*, where *rps*4 and especially *trn*L lack sufficient variation, as well as the exonic variation to reinforce topologies already supported by both *rps*4 and *trn*L. Furthermore, the congruence among the three data partitions lends additional evidence that *gpd* is an effective marker for reconstructing shallower splits, where it is often necessary to rely on a single gene (since so few genes with appropriate variation are known at present).

Despite the obvious congruencies between the chloroplast DNA and *gpd*, a number of topological inconsistencies were found. The most glaring were the multiple branch shifts within clade D and the lack of resolution at the node distinguishing clades C and D. There are several possible reasons for toplogical incongruence across phylogenetic analyses using different genes. Among these are different histories and different rates of evolution. A partition homogeneity test suggested that *rps*4 and *gpd* do not have different histories. However, the number of anomalous regions found in *gpd* indicates that recombination may have occurred. Still, the levels of homology within the anomalous regions was high, higher in fact than in those regions found not to evolve anomalously. Furthermore, the general biology of *Mitthyrdium*, including diocy (males and females on separate plants), rarity of sexual reproduction (Reese et al., 1986), and the sparse distribution of populations (Wall, personal observation) make hybridization seem unlikely. Nevertheless, ancient hybridization cannot be ruled out and further tests should be attempted to address this potential problem.

Rates of evolution may differ between genes for numerous reasons including variable selection intensities and rapid diversification. In the present study, positive selection was detected in only 3 of the 26 *Mitthyridium* taxa; the others were found to have a low average d_N/d_S ratio. Similarly, positive selection was not discovered in *rps*4, suggesting that differential selection intensities may not play a role in shaping the differences between the two gene phylogenies. However, the rate of pairwise divergence among all 3 genes differed considerably. Thus, differences in phylogeney reconstruction (e.g., in clade D) and level of resolution, especially between *gpd* and *trn*L may be an artifact of the different rates of nucleotide substitutions across taxa rather than an indication of separate history.

Rapid radiation and consequently the existence of hard polytomies (Jackman et al., 1999) could be a further cause of the lack of topological resolution at the node between clades B, C, and D as well as among branches within clades C and D. Likelihood-ratio tests failed to reject the null hypothesis of a molecular clock in the exons of *gpd* for the 26 ingroup *Mitthyridium* taxa. Unfortunately, the age of the most recent common ancestor of *Mitthyridium* remains a mystery. Provided that a reasonable calibration can be found, the existence of a clock should allow an adequate calculation of the age and rate of diversification of *Mitthyridium*, hopefully shedding light on the rate and pattern of branching. Such a study is currently in preparation by the author. While the generality of this molecular clock has yet to be established, there is some promise that *gpd* may be of utility to plant systematists hoping to explore ages and rates of diversification.

The high rate of sequence evolution found in the introns of *gpd* holds promise for future phylogeographic studies in mosses, although more research on this gene at the moss population level is needed. Phylogeographic studies of plants lag significantly behind similar studies in animals (Olsen and Schaal, 1999; Schaal et al., 1998). Consequently, much of the field of phylogeography and its important theoretic developments, such as the application of coalescence theory, rely heavily on the empirical results of animals (Avise, 2000). Given the differences in their biology, plant studies will add empirical data to this developing field; new nuclear markers like *gpd* are likely to become more plentiful and should be exploited.

Although no more than half of the functional *gpd* gene was sequenced for the present analysis, designing primers to capture the entire gene could be done easily using the available sequences in GenBank. Future research should accomplish this; a common problem in molecular systematics is the tendency to sequence portions of genes. While this partial gene sequence may service the needs of the systematic community in the short term, it leaves little opportunity for collaboration with other fields like biochemistry. Furthermore, a growing set of empirical studies has shown that nucleotides do not vary independently, especially in coding regions or regions of known function. As the knowledge gap between gene sequence, protein form, and function narrows, the need for full genes will become increasingly paramount for improving the models used to reconstruct both organismal and molecular evolution.

Acknowledgments

This manuscript is a portion of my Ph.D. thesis completed at the University of California, Berkeley under the supervision of Dr. Brent D. Mishler and with the support of the National Science Foundation (PEET; DEB-9712347). Many thanks to Dr. John Wheeler, Dr. Bruce Baldwin, and Dr. Rosemary Gillespie for support and advice. Also, I extend my sincere gratitude to Rosalyn Sayman and Danica Harbaugh for technical assistance. The present manuscript was much improved by comments of two anonymous reviewers.

References

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

Buckler, E.S.I., Ippolito, A., Holtsford, T.P., 1997. The evolution of plant ribosomal DNA: divergent paralogues, pseudogenes and phylogenetic implications. Am. J. Bot. 84, 1.

Clegg, M.T., Cummings, M.P., Durbin, M.L., 1997. The evolution of plant nuclear genes. Proc. Natl. Acad. Sci. USA 94, 7791–7798.

Cove, D., 2000. The moss, *Physcomitrella patens*. J. Plant Growth Regul. 19, 275–283.

Eddy, A., 1988. A Handbook of Malesian Mosses. British Museum (Natural History), London.

Fagan, T., Hastings, J.W., Morse, D., 1998. The phylogeny of glyceraldehyde-3-phosphate dehydrogenase indicates lateral gene transfer from cryptomonads to dinoflagellates. J. Mol. Evol. 47, 633–639.

Farris, J.S., Kallersjo, M., Kluge, A.G., Bult, C., 1995. Constructing a significance test for incongruence. Syst. Biol. 44, 570–572.

Felsenstein, J., 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22, 521–565.

Figge, R.M., Schubert, M., Brinkmann, H., Cerff, R., 1999. Glyceraldehyde-3-phosphate dehydrogenase gene diversity in eubacteria and eukaryotes: evidence for intra- and inter-kingdom gene transfer. Mol. Biol. Evol. 16, 429–440.

Girke, T., Schmidt, H., Zaehringer, U., Reski, R., Heinz, E., 1998. Identification of a novel DELTA6-acyl-group desaturase by targeted gene disruption in Physcomitrella patens. Plant J. 15, 39–48.

Grassly, N.C., Rambaut, A., 1998. PLATO—Partial Likelihoods Assessed Through Optimisation, ver. 2.0. University of Oxford.

Henze, K., Badr, A., Wettern, M., Cerff, R., Martin, W., 1995. A nuclear gene of eubacterial origin in Euglena gracilis reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA 92, 9122–9126.

Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science (Washington, DC) 276, 227–232.

Jackman, T.R., Larson, K., De Queiroz, K., Losos, J.B., 1999. A. Phylogenetic relationships and tempo of early diversification in Anolis lizards. Syst. Biol. 48, 254–285.

Kellogg, E.A., Appels, R., Mason-Gamer, R.J., 1996. When genes tell different stories: the diploid genera of Triticeae (Gramineae). Syst. Bot. 21, 321–347.

La Farge, C., Mishler, B.D., Wheeler, J.A., Wall, D.P., Johannes, K., Schaffer, S., Shaw, A.J., 2000. Phylogenetic relationships within the haplolepideous mosses. Bryologist 103, 257–276.

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

Martin, W., Lydiate, D., Brinkmann, H., Forkmann, G., Saedler, H., Cerff, R., 1993. Molecular phylogenies in angiosperm evolution. Mol. Biol. Evol. 10, 140–162.

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

Nowak, H., 1980. Revision der Laubmoosgattung *Mitthyridium* (Mitten) Robinson fur Ozeanien (Calymperaceae). J. Cramer, Hirschberg.

Olsen, K.M., Schaal, B.A., 1999. Evidence on the origin of cassava: phylogeography of *Manihot esculenta*. Proc. Natl. Acad. Sci. USA 96, 5586–5591.

Panvisavas, N., Lu, G., Quatrano, R., Cove, D.J., Cuming, A.C., Knight, C.D., 1999. Gene targeting in the moss *Physcomitrella patens* using EST homologues of higher plant genes. J. Exp. Bot. 50, 18.

Reed, R.D., Sperling, F.A.H., 1999. Interaction of process partitions in phylogenetic analysis: an example from the swallowtail butterfly genus *Papilio*. Mol. Biol. Evol. 16, 286–297.

Reese, W.D., 1994. The subgenera of *Mitthyridium* (Musci). J. Hattori Bot. Lab. 0, 41–44.

Reese, W.D., Magill, R.E., Pocs, T., 1994. *Mitthyridium* microundulatum subsp. comoroensis subsp. nov. from Mayotte. Bryologist 97, 430–431.

Reese, W.D., Mohamed, H., Mohamed, A.D., 1986. A synopsis of *Mitthyridium* (Musci: Calymperaceae) in Malaysia and adjacent regions. Bryologist 89, 49–58.

Reski, R., 1998. Development, genetics and molecular biology of mosses. Bot. Acta 111, 1–15.

Reski, R., Reynolds, S., Wehe, M., Kleber-Janke, T., Kruse, S., 1998. Moss (*Physcomitrella patens*) expressed sequence tags include several sequences which are novel for plants. Bot. Acta 111, 143–149.

Schaal, B.A., Hayworth, D.A., Olsen, K.M., Rauscher, J.T., Smith, W.A., 1998. Phylogeographic studies in plants: problems and prospects. Mol. Ecol. 7, 465–474.

Schaal, B.A., Olsen, K.M., 2000. Gene genealogies and population variation in plants. Proc. Natl. Acad. Sci. USA 97, 7024–7029.

Schaefer, D.G., Zyrd, J.-P., 1997. Efficient gene targeting in the moss *Physcomitrella patens*. Plant J. 11, 1195–1206.

Sorenson, M.D., 1999. TreeRot, ver. 2. Boston University, Boston.

Souza-Chies, T.T., Bittar, G., Nadot, S., Carter, L., Besin, E., Lejeune, B., 1997. Phylogenetic analysis of Iridaceae with parsimony and distance methods using the plastid gene *rps*4. Plant Syst. Evol. 204, 109–123.

Strand, A.E., Leebens-Mack, J., Milligan, B.G., 1997. Nuclear DNA-based markers for plant evolutionary biology. Mol. Ecol. 6, 113–118.

Swofford, D.L., 2000. PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M.M., Mable, C., B, K. (Eds.), Molecular Systematics, second ed. Sinauer Associates, Inc., Sunderland, MA, USA, pp. 407–514.

Taberlet, P., Geilly, L., Pautou, G., Bouvet, J., 1991. Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mol. Biol. 17, 1105–1109.

Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.

Viscogliosi, E., Mueller, M., 1998. Phylogenetic relationships of the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase, from parabasalid flagellates. J. Mol. Evol. 47, 190–199.

Wheeler, J., Wall, D.P., Johannes, K., Mishler, B.D., in press. Congruence and convergence in the moss family Calymperaceae: phylogenetic analysis of two chloroplast genes (*rbc*L and *rps*4) and morphology. Syst. Biol.

Wood, A.J., Oliver, M.J., Cove, D.J., 2000. Bryophytes as model systems. Bryologist 103, 128–133.

Yang, Z., Nielsen, R., 2000. Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17, 32–43.

Yang, Z., 2000. Phylogenetic Analysis by Maximum Likelihood (PAML), ver. 3.0. University College London, London.

Zamudio, K.R., Jones, K.B., Ward, R.H., 1997. Molecular systematics of short-horned lizards: biogeography and taxonomy of a widespread species complex. Syst. Biol. 46, 284–305.