MYCOLOGICAL RESEARCH 112 (2008) 316–330

British Mycological Society promoting fungal science

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

Secondary structure of ITS2 rRNA provides taxonomic characters for systematic studies — a case in *Lycoperdaceae* (*Basidiomycota*)

Dirk KRÜGER[†,*], *Andrea GARGAS*[‡]

The Botany Department, University of Wisconsin, 430 Lincoln Drive, Madison, WI 53706, USA

ARTICLE INFO

Article history:
Received 6 April 2006
Received in revised form
24 September 2007
Accepted 30 October 2007
Corresponding Editor: Michael Weiss

Keywords:
Consensus structure
Internal transcribed spacer
ITS2 rDNA
Nomenclature

ABSTRACT

The secondary structure of the ITS2 rDNA transcript (pre-rRNA) could provide information for identifying homologous nucleotide characters useful for cladistic inference of relationships. Such structure data could become taxonomic characters. This work compares the effect of several modern nucleotide alignment strategies, including those making use of structure data, on phylogenetic inference. From both the phylogenetic analyses and comparative secondary structure, implications for taxonomy and evolution of puffball fungi are discussed. *Lycoperdaceae* remain insufficiently resolved with present taxon and data sampling. Neither alignment allows statistically robust phylogenetic hypotheses under any current optimality criterion. The secondary structure data at this time are best used as accessory taxonomic characters as their phylogenetic resolving power and confidence in validity is limited compared with underlying nucleotide characters. We introduce a preliminary nomenclature convention to describe secondary structure for defining consensus features. These consensus structures are illustrated for the clades /Calvatia, /Handkea–Echinatum, /Vascellum, /Morganella, and /Plumbea–Paludosa (*Bovista*).

Introduction

rRNA and secondary structure

The rRNA genes have long been seen as the ultimate tracer of evolutionary relationships. Historically, targeting these regions of the genomes has been eased by primer annealing site conservation and the redundancy of the array, offering a higher number of templates for PCR. Although protein-coding genes are transcribed into mRNA ultimately processed by excision of introns and translation into proteins, RNA is the ultimate phenotypic expression of rDNA. In rRNA, spacers are excised, and the RNA itself is the molecule for selection to act upon. RNA folds up in stretches that are, in fact, double-stranded, while others remain single-stranded.

rRNA genes have been widely used in systematic studies in fungi and beyond, and are common targets for identifying and quantifying phylotypes in medical and environmental samples. The 'coding' (we use coding as coding for RNA) SSU rRNA and LSU rRNA genes are highly conserved. Evidence suggests that secondary structures of the initial transcript play

* Corresponding author. Tel.: +49 (345) 558 5425.
E-mail address: dirk.krueger@ufz.de
† Present address: Helmholtz-Zentrum für Umweltforschung GmbH – UFZ, Department of Soil Ecology, Theodor-Lieser-Str. 4, D-06120 Halle/Saale, Germany.
‡ Present address: Symbiology LLC, Middleton, WI 53562-1230, USA.

0953-7562/$ – see front matter
doi:10.1016/j.mycres.2007.10.019

important roles in ribosome assembly (Lalev & Nazar 1999; Lalev *et al.* 2000; Lalev & Nazar 2001), and putative secondary structures have long been recognized and archived for the coding regions of rDNA (van de Peer *et al.* 2000; Cannone *et al.* 2002), and recently for the internal spacers as well (Wolf *et al.* 2005). Bioinformatics programs, such as Mfold (Zuker 2003), Pfold (Knudsen & Hein 2003), and RNAviz (de Rijk *et al.* 2003), compute and depict putative secondary structure of any nucleic acid molecule, and a universal XML-based syntax was proposed for RNA structure bioinformatics (Waugh *et al.* 2002).

Characters from nuclear ribosomal gene sequences have been used for hypotheses of phylogenetic relationships among even distantly related organisms. Molecular phylogenetic analyses require the alignment of homologous sequence characters, and guidance from secondary structure information may aid in the alignment of homologous regions for phylogenetic analysis among plant and animal species (Jobes & Thien 1997; Goertzen *et al.* 2003; Xia *et al.* 2003), even between genomes as evolutionarily distant as of eukaryotic nuclei, prokaryotes, and eukaryotic organelles (Cedergren *et al.* 1988). Improved alignments are thus an alternative to excluding or recoding ambiguous indels (Lutzoni *et al.* 2000; Lee 2001; Young & Healy 2001; Creer *et al.* 2006). A further application of structure information to phylogenetics is recoding structure into new characters states (Caetano-Annolés 2002; Smith *et al.* 2004).

ITS regions have been used for phylogenetic analyses at the species to generic level, yet their primary nucleotide sequence often contains insertions and deletions (indels) making alignment difficult much beyond infraspecific levels. This ITS variability led to the assumption that non-coding ITS1 and ITS2 regions were mere 'junk DNA', whose evolution resulted from the accumulation of chance mutations unfettered by any functional constraints. However, research on plants and green algae suggested that ITS rDNA sequences provide evidence at a super-generic level (Baldwin *et al.* 1995; Hershkovitz & Zimmer 1996; Mai & Coleman 1997) and contain diagnostic characters for deeper divergences (Hershkovitz & Lewis 1996).

Sections of ITS2 transcripts are consistently predicted to form conserved stem-loop structures. Complementary base changes and the recovery of secondary structure motifs independent of primary sequence renders secondary structure highly appropriate for sequence alignment and taxonomy. Wolf *et al.* (2005) compiled a database of predicted secondary structures of ITS2. Incorporating some of these recent advances, we are now able to scrutinize the non-coding spacers and introns for secondary structure features and better employ them for biological systematics.

We have previously shown that the compression of phylogenetic information content inherent in the recoding of structural data *versus* underlying primary DNA sequence data precipitates a decrease of phylogenetic resolution (Krüger & Gargas 2004; Krüger *et al.* 2006). Refining models of rDNA sequence evolution and experimental verification of predicted structures will increase the reliability of rDNA as a phylogenetic marker.

After using secondary structure in polypore fungi (Krüger & Gargas 2004; Krüger *et al.* 2006), in the present study we (1) deduce and compare putative secondary structure features from ITS2 sequences from puffballs (*Lycoperdaceae*); (2) explore the effects of secondary structure incorporation into alignment and, subsequently, on the outcomes of phylogenetic analyses; and (3) define consensus structures as taxonomic characters for clades on phylogenetic trees. Finally, we discuss parsimonious explanations of DNA sequence changes responsible for the predicted secondary structures in *Lycoperdaceae* within a phylogenetic context and compared with previously used morphotaxonomic characters delimiting taxa.

Mutational events include compensatory base changes, one-sided changes (mutations leading to introduction or loss of new features), resizing changes (mutations increasing or decreasing feature sizes) or silent changes (mutations not altering the secondary structure). These trends cannot yet be described as directional, and the timing of putative events relative to clade separation is unknown (an event might be suggested by a homoplasious nucleotide character or we might miss data to correctly infer an event's occurrence). This precludes two useful future uses of pointing out such hypothetical events: (1) possible confirmation with independent data; and (2) possible experimentation on the feasibility of such event to allow formation of a functional ribosome.

Lycoperdaceae

Puffballs (gasteromycetes) are common inhabitants of grasslands, although some are found in forests, e.g. on wood. One example is the giant puffball, *Calvatia gigantea*, which produces millions of spores in its fruit body cavity (Ingold 1971); puffball spores may be potently allergenic (Levetin *et al.* 1992; Horner *et al.* 1993). Within the main family, *Lycoperdaceae*, the value of capillitia (hyphal threads) and spores — two traditional micromorphological characters for delimitation of puffball genera — remain ambiguous. The genera *Bovista* and *Lycoperdon* share common morphological types of capillitia inside their spore-producing cavities (the *Bovista* type, the intermediate type, and the *Lycoperdon* type of capillitium; all occur among *Bovista* species; Krüger *et al.* 2001; Bates 2004), calling into question the usefulness of this character and/or the traditional generic circumscriptions. Currently, available molecular sequence data do not satisfactorily delimit genera *Bovista* and *Lycoperdon* (Krüger *et al.* 2001; Bates 2004). The genus *Handkea* (Kreisel 1989) may be separated from *Calvatia* in micromorphology based on the presence of a distinct capillitium type featuring slit-like depressions in the cell walls, rather than pores.

Puffballs of the order *Lycoperdales*, as currently circumscribed, are believed to be a monophyletic group, although its former enclosing taxon, class *Gasteromycetes*, is now considered an artificial group (Hawksworth *et al.* 1995; Hibbett *et al.* 1997; Moncalvo *et al.* 2002). Morphological characters supporting the DNA-confirmed segregation of *Geastrales* (earthstars) from the order *Lycoperdales* (required to establish monophyly of the *Lycoperdales*) have been highlighted by Iosifidou & Agerer (2002). *Lycoperdaceae* are closely related to agarics (mushrooms such as *Lepiota* spp. and *Agaricus* spp.) (Vellinga 2004) and are perhaps best included in the *Agaricaceae*.

Material and methods

PCR and sequencing

Voucher information, origin of sequence data, and DNA extraction methods are provided in Krüger *et al.* (2001) and Krüger & Kreisel (2003). New sequence data (geographic origin and voucher information included in Table 1) was generated after amplification from extracted DNA using primers NS7UTK (Krüger 2002; samples CFRO, DA51, DA98, HERB), ITS4B (Gardes & Bruns 1993; HERB) and ITS4D (5′-TGT CAA AGA CGG GAT TCT C, Tm: 52.7 °C; DA51) and sequenced using ITS1 (White *et al.* 1990; CFRO, DA30, HERB), ITS4 (White *et al.* 1990; all five new sequences), ITS4D (DA51) or NS7UTK (DA51). We used PCR RedMix™ Plus 2X Master Mix Kit (GeneChoice, Frederick, MD), supplemented with 1 M betaine monohydrate (Acros Organics, Geel) as enhancer (Rees *et al.* 1993), and cleaned amplicons with ExoSAP-IT® (USB, Cleveland, OH). PCR products were cycle-sequenced with dye-dideoxy-terminators (BigDye v. 3.1, Applied Biosystems, Foster City, CA), purified using the MagDTR™ Magnetic Resin (Edge BioSystems, Gaithersburg, MD) and electrophoresed on automated capillary sequencing systems at the UW Madison Center for DNA Sequencing. Electropherogram files were checked in Chromas v. 1.45 (Technelysium, Tewantin, Queensland), then sequence data were imported and assembled in BioEdit v. 5.0.9 (Hall 1999).

Secondary structure prediction

Initial alignments were compiled using ClustalX PC software (Thompson *et al.* 1997) with manual minimization of gaps using BioEdit. ITS2 rDNA regions were identified by comparison to published sequences for *Saccharomyces cerevisiae*; the 5.8S DNA region to Rubin (1973), the LSU rDNA region to Bayev *et al.* (1981), and the ITS2-proximal stem by comparison to structures in Côte & Peculis (2001). After delimitation of their ITS2-proximal stems, initial folding for each ITS2 rRNA sequence was based on models predicted by Mfold version 3.1 (Zuker 1989; Zuker 2003, http://www.bioinfo.rpi.edu/applications/mfold/rna/form1.cgi) with default conditions set to: linear RNA sequence, folding temperature of 37 °C, 20 % suboptimality, upper bound of 50 foldings, no limit to the maximum physical distance between paired bases, no constraint information, maximum number of nucleotides in a bulge or loop = 30, maximum asymmetry of an interior/bulge loop = 30). From the hypothesized foldings, those most closely resembling the four conserved pairing regions (stems) hypothesized for plants and green algae (Mai & Coleman 1997) and especially basidiomycete fungi (Krüger & Gargas 2004; Krüger *et al.* 2006) were carefully scrutinized. This procedure

Table 1 – Number of potential secondary structures predicted

Taxon	GenBank number	Number of Mfold structures for entire ITS2	Number of Mfold structures when using putative first two stem-loops	Reference
Agaricus bisporus	DQ404388	27	6	
Bovista paludosa	AJ237630	5	1	
B. plumbea	AJ237629	4	1	
B. polymorpha	AJ237613	1	1	
B. pusilla	AJ237631	6	1	
Calvatia fragilis collection CFRO	AJ617493	8 (3′ end cut off)	1	
C. gigantea collection DA30	AJ617492	2 (3′ end cut off)	1	Germany: Mecklenburg (herbarium CFRM FP-98552)
Disciseda candida	DQ112626	24	2	
Handkea excipuliformis collection HERB	AJ617491	9	1	Germany: Bavaria (herbarium REG s.n. 30.11.95 leg. Debus, det. Besl)
H. utriformis collection DA51	AJ617490	4	1	UK: Scotland (CFRM FP-98553)
Leucoagaricus americanus	AY176407	22	6	Vellinga (2004)
Lepiota cristata	AJ237628	8	2 (one incompatible with any other structure)	
Lycoperdon echinatum	AJ237622	3	1	
L. perlatum	AJ237627	7	1	
Macrolepiota konradii collection DA98	AJ617494	8	1	
Myrmicocrypta urichi symbiont T39	AF079720	20	4	Mueller *et al.* (1998)
Morganella fuliginea	AF485065	4	1	
M. pyriformis	AJ237620	3	1	
M. subincarnata	AJ237626	7	1	
Vascellum cf. pratense	AB067725	5	1	Terashima *et al.* (2002)
V. pratense	AJ237625	6	1	

Origin of specimens and references to herbarium specimens or literature citations are included for new sequence data, otherwise available from Krüger *et al.* (2001).

was iterative, with ongoing comparison and refinement for each taxon.

Alignment derivation and refinement

Alignments and their subsequent use in phylogenetic bioinformatics are summarized in Supplementary Material Table S2 and made available as supplemental data online. The ITS region sequence data were initially aligned online using a MAFFT web server (Katoh *et al.* 2005) and a MUSCLE web server (Edgar 2004). Locally, we used the POY (Wheeler *et al.* 2003; commands -likelihood -iafiles -noprealigned -impliedalignment) and QAlign2 Panta Rhei (Sammeth *et al.* 2006) software, the latter with the ClustalW (Chenna *et al.* 2003), Divide-and-Conquer (DCA, Stoye 1998), T-Coffee (Notredame *et al.* 2000), and DIALIGN (Morgenstern 1999) algorithms. Finally, after exclusion of the three most divergent sequences (*Agaricus*, *Leucoagaricus*, and *Disciseda*), Align-m (van Walle *et al.* 2004) was also utilized for alignment. Without this exclusion, Align-m alignments became too indel-rich and divergent from all other alignments. The 3′ of the *Disciseda* ITS2 sequence deposited at the sequence repositories is extremely divergent to that of the other puffballs. All alignments were performed with the default settings. Starting with pairwise visual comparison of the MAFFT and MUSCLE data using AltAVisT (Morgenstern *et al.* 2003) online, nucleotides participating in competing alignment segments were manually excluded. This AltAVisT comparison and subsequent exclusion went on as described in Supplementary Material Table S2, ultimately leading to the complete Alignment 1, representing the nucleotide fraction most confidently seen as homologously aligned.

This Alignment 1 was then trimmed to the boundaries corresponding to the P1 through P2 region (Krüger & Gargas 2004; Krüger *et al.* 2006), this being called Alignment 1A. Alignment 2 was created from the original MAFFT alignment, under exclusion of the four most diverse sequences, and trimmed to the P1/P2 ITS2 region (Krüger & Gargas 2004). Only understood by PAUP (Swofford 2002), GapCoder (Young & Healy 2001) and INAASE (Lutzoni *et al.* 2000) were used to recode ambiguities. Alignment 2 was modified to Alignment 3 by inserting characters delimiting boundaries of single-stranded RNA features (ssRNA). The character code ‘‘/’’ denoted the end of a helix, and thus start of ssRNA, while ‘‘\’’ denoted the end of ssRNA and start of a helical double-stranded RNA (dsRNA). A spaceholder ‘‘*’’ was inserted into the primary sequence data of taxa missing the above characters, a character ‘‘@’’ equated with indel within otherwise ssRNA. We arbitrarily placed a character weight of 5 on \, /, and * (*versus* 1 for standard nucleotides in helices) and used equal ratios (1 each) for interchanges between characters \, /, and *. These characters and weighting schemes were only implored in PAUP, but not in other programs. Alignment 4 follows the assumption that actual nucleotide sequence within bulges and loops is not important for the function of these, in all nucleotides within \ and / were replaced with a character ‘‘L’’. This character was equated with missing (code Z) and fully ambiguous sites (IUPAC code N) in PAUP.

Phylogenetic analyses

To first assess the total ITS region MP hypothesis, Alignment 1 was subjected to jackknife 37 % resampling (Efron & Gong 1983; Farris *et al.* 1996) in PAUP (Swofford 2002). Upon Modeltest (Posada & Crandall 1998)/MtGui (www.genedrift.org)/MrModeltest (Nylander 2004) estimating the model of sequence evolution (Supplementary Material Table S3), ML and Bayesian inference (MrBayes block written using http://130.132.86.77/viburnum/mrbayes.html) were used for further evaluation of Alignment 1. FastME (Desper & Gascuel 2002, 2004) was utilized for balanced minimum-evolution (BME) phylogenetic inference on K2P distance matrices from PHYLIP dnadist.exe (Felsenstein 2005). T-REX (Makarenkov 2001) calculated a reticulogram for showing conflict within Alignment 1 with a starting tree from BioNJ (Gascuel 1997) of a GTR least-square distance matrix from PAUP (Supplementary Material Table S3), and reticulation stopped upon parameter Q1 optimization.

Alignments 1A and 2 were subjected to the same analyses as Alignment 1. The ancestral sequence to all puffball taxa was estimated on Alignment 2, using Bayesian approaches, as well as parsimony (PAUP) and the pamp.exe program in PAML (Yang 1997).

The structure-enhanced Alignments 3 and 4 were subjected to jackknife parsimony analysis in PAUP to make use of the recoding understood by PAUP only. Finally, we compared symmetric tree distance (Robinson & Foulds 1981; Felsenstein 2005; treedist.exe in PHYLIP) calculated for the 21-taxa alignments 1 and 1A, and the 17-taxa alignments 2, 3 and 4 by using the UPGMA (Sneath & Sokal 1973) algorithm in PHYLIP’s neighbor.exe (Felsenstein 2005). This, as well as the parsimony scores for all MP consensus trees, and average BS, jackknife, and Bayesian PPs (all above 50 %) was used to estimate performance of each analysis. These branch support values are not readily comparable (Cummings *et al.* 2003; Alfaro *et al.* 2003).

Tree searches in PAUP involved nearest-neighbour interchange swapping, except for Alignment 1, which was also analysed using tree bisection–reconnection (TBR). Phylogenetic trees were usually visualized using PhyloDraw (Choi *et al.* 2000) and GhostScript (http://www.cs.wisc.edu/~ghost/). Consensus trees from FastME and PHYML were generated using the consense.exe program in PHYLIP.

Results

Parsimony and BME

Alignment 1 (TreeBASE, study accession number S1941; all alignments are available as supplemental data in truncated NEXUS format) contained 21 unique sequences with 565 nucleotide and gap characters. Alignment 1A then decreased to just 78 characters with 16 that were parsimony-informative. Alignments 2, 3, and 4, with 17 taxa each, contained 96, 160, and 160 characters each, with 35, 55, and 50 characters that were parsimony informative, respectively.

The parsimony jackknife consensus (Fig 1) scored with average jackknife partition supports of *ca* 69 % (72 % using TBR

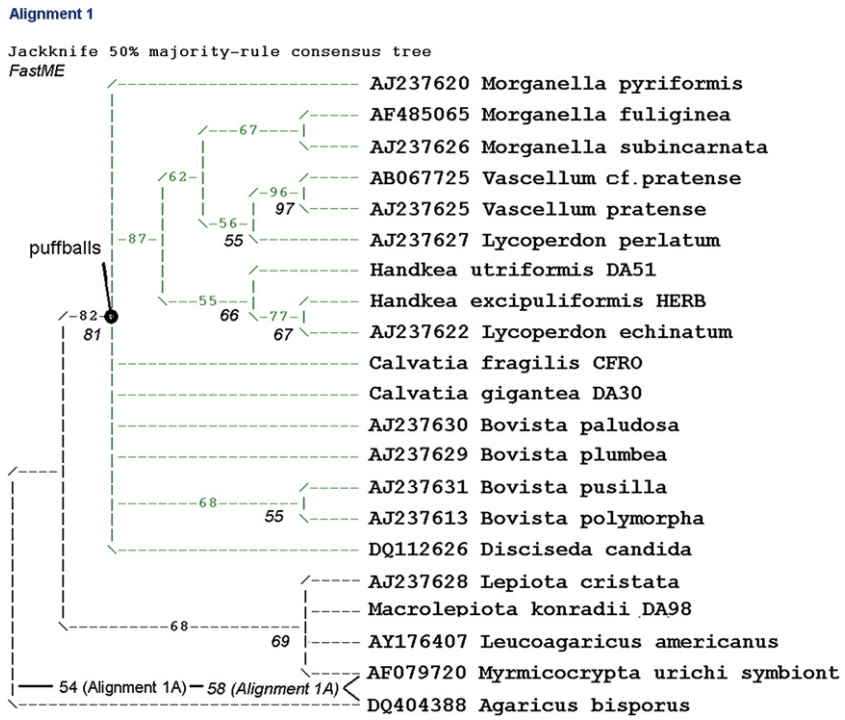

Fig 1 – Alignments 1, 1A. Outgroup *Agaricus*. Topology and support values on branches from 50 % majority rule consensus of 100 parsimony jackknife replicates performed on Alignment 1. Branch supports in italics from 100 BME (FastME) BS replicates. Only partitions supported over 50 % in Alignment 1A (*Myrmicocrypta* and *Agaricus versus* all other taxa) given at the according location.

swapping). The shortening of the data to only the P1–P2 region of ITS2 resulted in a decrease of the average jackknife values to 53 %. Alignment 2 yielded 66 % in the jackknife.

Bayesian inference behaved differently in that the decrease of characters from Alignment 1 to 1A actually increased the PPs (see Supplementary Material Table S1; Bayesian topology not shown). Alignment 2 then decreased in PPs again as compared with Alignment 1A. BS values for ML analyses and BME declined from Alignment 1 through 1A to 2.

All MP and BME analyses retained the monophyly of the puffball taxa with Alignment 1 and 1A (Fig 1). However, the consensus trees with Alignment 1A are almost unresolved (only one branch support over 50 %), but already with Alignment 1, resolution within the puffball taxa themselves is low.

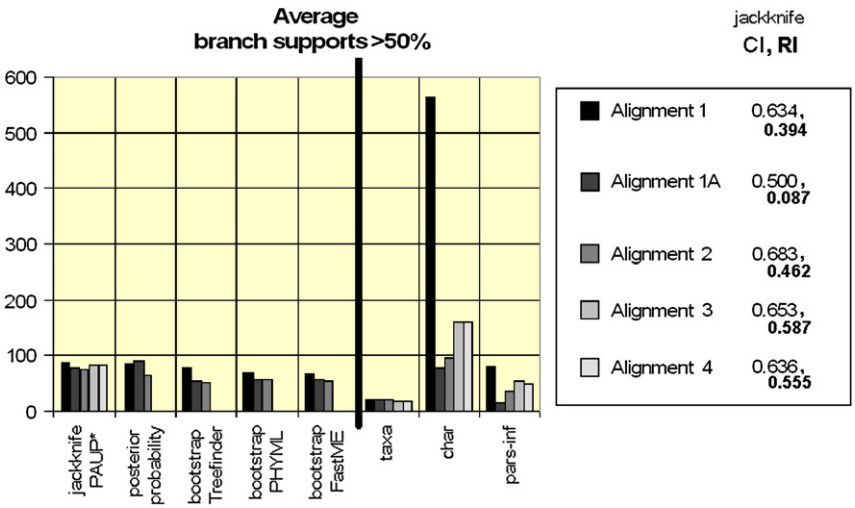

Fig 2 – Plot of average branch support values for Alignments 1, 1A, 2 in parsimony jackknife, Bayesian PPs, Treefinder ML BS, PHYML ML BS, BME BS. To the right follow number of taxa, characters, and parsimony-informative characters in Alignments 1, 1A, 2, 3, 4. The legend includes the CI of the jackknife consensus for each Alignment.

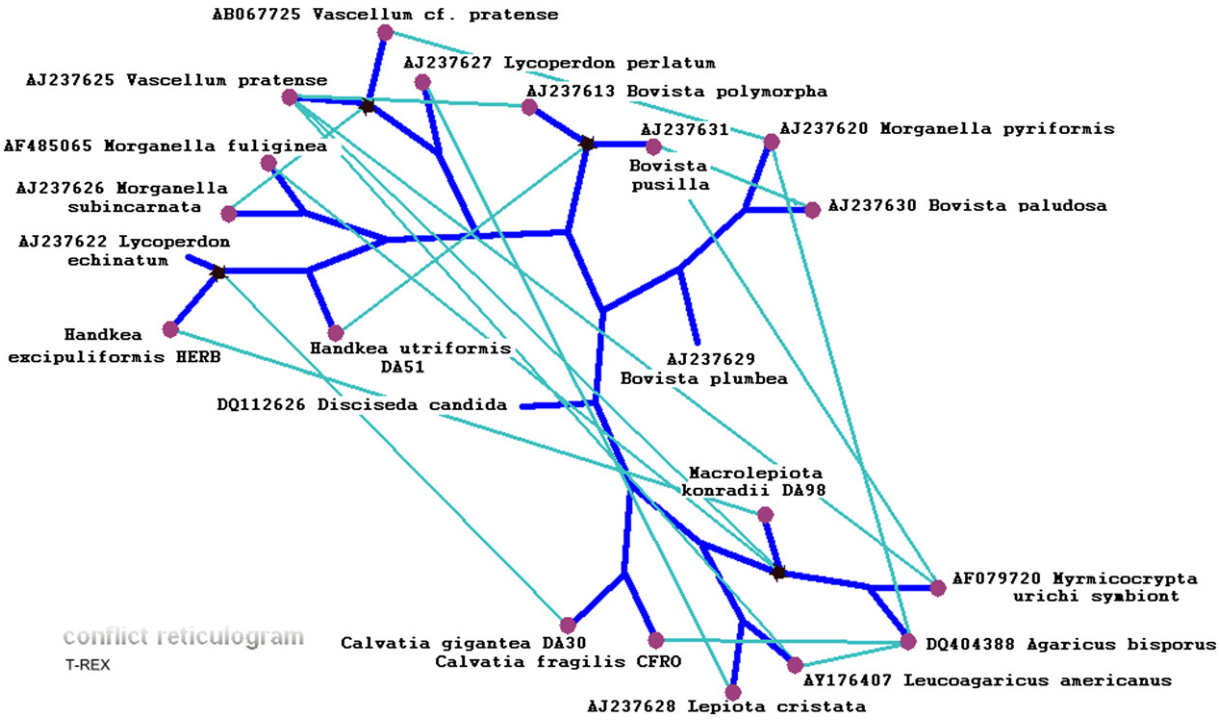

Fig 3 – T-REX reticulogram of analysis of conflict within Alignment 1. Node markers and tip markers indicate all positions on the topology implicated in conflict. The reticulations can be seen as strings pulling the topology to alternative explanations of phylogenetic relationships.

ML and Bayesian inference

Likelihood-based phylogenetic estimates are summarized in Supplementary Material Fig S1. In it, Alignment 1 displays a moderate support of puffball taxa monophyly, which is not anymore the case with Alignments 1A and 2. BS in PHYML and Treefinder are decreasing from Alignment 1 through 1A to 2, while the peak in MrBayes, as PP, is with Alignment 1A.

Distance-based reticulogram

The T-REX reticulogram within Alignment 1 (Fig 3) shows the high level of conflict apparent. Numerous reticulations affect 18 of the 21 tips directly. AJ237625 *Vascellum pratense* shows the highest number of reticulations connecting it with four tips and nodes in the reticulogram, while it is otherwise well anchored in all analyses on Alignment 1 in a monophyletic branch with the other *V. cf. pratense*. Puffball monophyly is supported in all Alignment 1 analyses, including the backbone topology of the reticulogram.

Although the shortening of the data to Alignment 1A and 2 results in massive loss of resolution, Alignment 1 always shows monophyly of *Lycoperdon echinatum* + *Handkea*. Generally, *L. perlatum* switches between a concladic position with *Morganella subincarnata* and *M. fuliginea*, and one with *Vascellum* (FastME topology not shown). Monophyly of *Bovista, Morganella,* and *Calvatia* is generally not supported with over

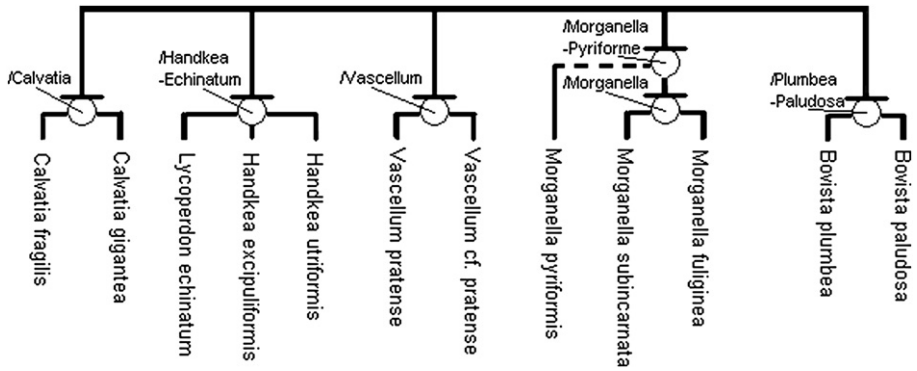

Fig 4 – Outtake from Alignment 1 Treefinder ML phylogeny (Supplementary Material Fig. S1). The labelled clades are discussed in subsequent figures in regards to secondary structure.

50 %. Fig 4 shows an outtake of Alignment 1 ML analysis, with *Morganella* rather a non-monophyletic group (/- Morganella). For each of the shown puffball groupings, we will later discuss secondary structure evolution.

Secondary structure and ancestral structure prediction

A summary of ITS2 rRNA structures predicted by Mfold appears in Table 1. Most sequences yielded only one structure after excluding primary sequence data before P1 and after P2 (Krüger & Gargas 2004). These structures were compatible with at least one of the structures predicted when using the entire available sequence (not so with the *Calvatia* sequences if the available sequence data at 3′ was not arbitrarily shortened), and compatible with each other across the alignment.

Fig 5 shows the structure predictions as to position of the P1 and P2 stem-loops and the joiner between them. For this example, *Morganella subincarnata*, we visually compare the prediction downloaded from the ITS2 online database (Wolf *et al.* 2005); our similar prediction based on the same entire ITS2, and just using P1 through P2 in Mfold. The difference within P1 through P2 between the database and our folding mainly concerns the extent and position of ssRNA within P1. Krüger & Gargas (2004) suggested that structure prediction beyond P2 is much more difficult, as can be seen by the differences between the database and our folding prediction.

Predicted ancestral sequences for all puffballs are shown in Supplementary Material Fig S2. They differ somewhat, partly because PAUP included ambiguities in the reconstructed sequence. Without these ambiguities, the predicted secondary structures would be more similar to each other. Overall, the PAUP reconstruction shows more dsRNA in agreement with all puffball taxa than does the MrBayes prediction. PAMP ancestral reconstruction yielded a much shorter sequence and structure.

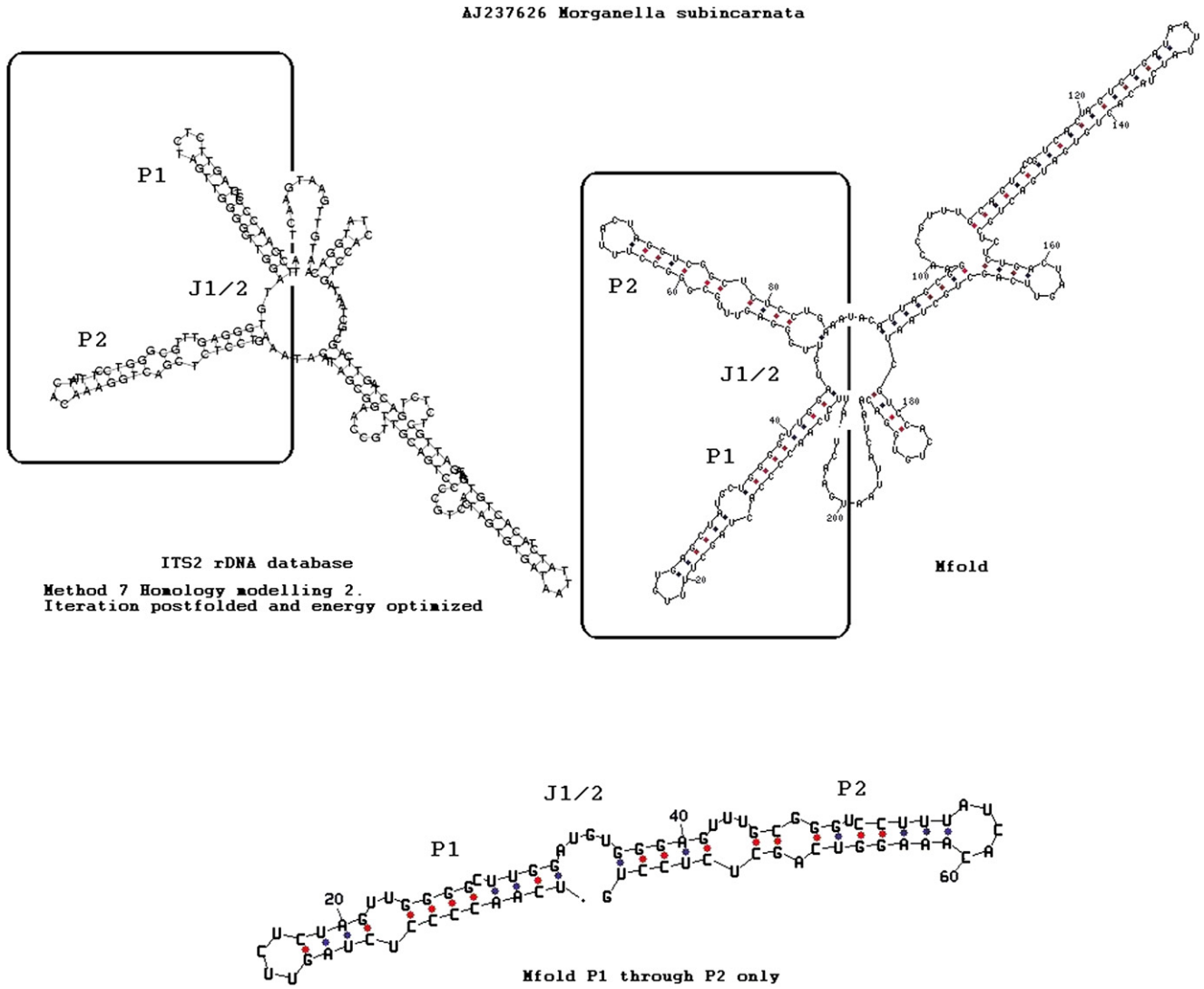

Fig 5 – Example secondary structure of *Morganella pyriformis*. The top two include the entire ITS2 spacer, with the P1, P2, and J1/2 areas marked. Top left = from the ITS2 database, top right = our Mfold prediction. Below, Mfold prediction parsing only P1, J1/2, P2 region.

When secondary structure recoding was used, parsimony analysis (Figs 2 and 6) was again not well resolved, and *Lepiota* was also located within the puffballs. The resolution is even further decreased compared with Alignment 2; the CI is slightly further decreased.

Supplementary Material Figs S3 and S4 offer a comprehensive view of tree distances of the major analyses. Within the analyses on the 21-taxa alignments, Alignment 1 and Alignment 1A tend to separate out in tree distances (Supplementary Material Fig S3). The same separation by alignment holds true for 17-taxa alignments (Supplementary Material Fig S4). This shows that the alignment overrides effects of differences in phylogenetic optimality criteria and algorithms.

Given that Alignment 1 is the richest in terms of characters, most conservative alignment with the best phylogenetic resolution, it shall serve as the guideline to discuss secondary structure evolution in the following putatively 'robust' clades visible in the outtake of phylogenetic trees depicted as Fig 4.

The following illustrations capture the consensus structures delimited from ITS2 rDNA sequences within each of the clades contained in Fig 4: /Calvatia (Fig 7), /Handkea–Echinatum (Fig 8), /Vascellum (Fig 9), /Morganella (Fig 10), and /Plumbea–Paludosa (Fig 11). /Morganella–Pyriforme is apparently not monophyletic (≠). As a new, not yet formally published nomenclature is applied, we shortly introduce this annotation tool of structure descriptors here. Hypothetical events during sequence evolution that gave rise to distinct structures are discussed in the following section of the paper.

A descriptor is a short annotation of putative or real secondary structure information allowing the nucleotide-independent drawing and the formulation of consensus structures at different taxonomic levels, as well as allowing implementation in future computer programs that formulate structures and recode it for phylogenetic analysis. Inclusion of helix characters belonging to dsRNA is optional, as these are not needed for nucleotide-independent drawing of structure. Helix characters might be appended at the end of the descriptor. A shorthand descriptor lacks any mentioning of primary sequence nucleotide characters. The shorthand version is intended for straightforward drawing of the overall structure. A longhand descriptor would also include details of primary sequence in brackets. Sequence is written 5′ to 3′ using upper case letters, and follows IUPAC conventions for ambiguous nucleotides. Descriptors may focus on a limited portion of the ITS2, as do our consensus structures. For consensus structure descriptors only the conserved regions are included for the appropriate taxonomic level. Longhand descriptors may optionally specify the start (in nucleotide position number) counted from the ITS2 5′ end. Helices or structural subfeatures are numbered sequentially within each pairing region (or insertional, side pairing region) starting at the basal end of the P region (i.e. h1, h2 etc). Position numbers are derived from the number of paired bases in the helices of each pairing region. The most basal base pair carries number 1. Counting proceeds from the base to the terminus of each pairing region. An internal bulge loop is considered as being left or right when it would appear to the left or right of the terminal hairpin loop on a flat drawing of the structure.

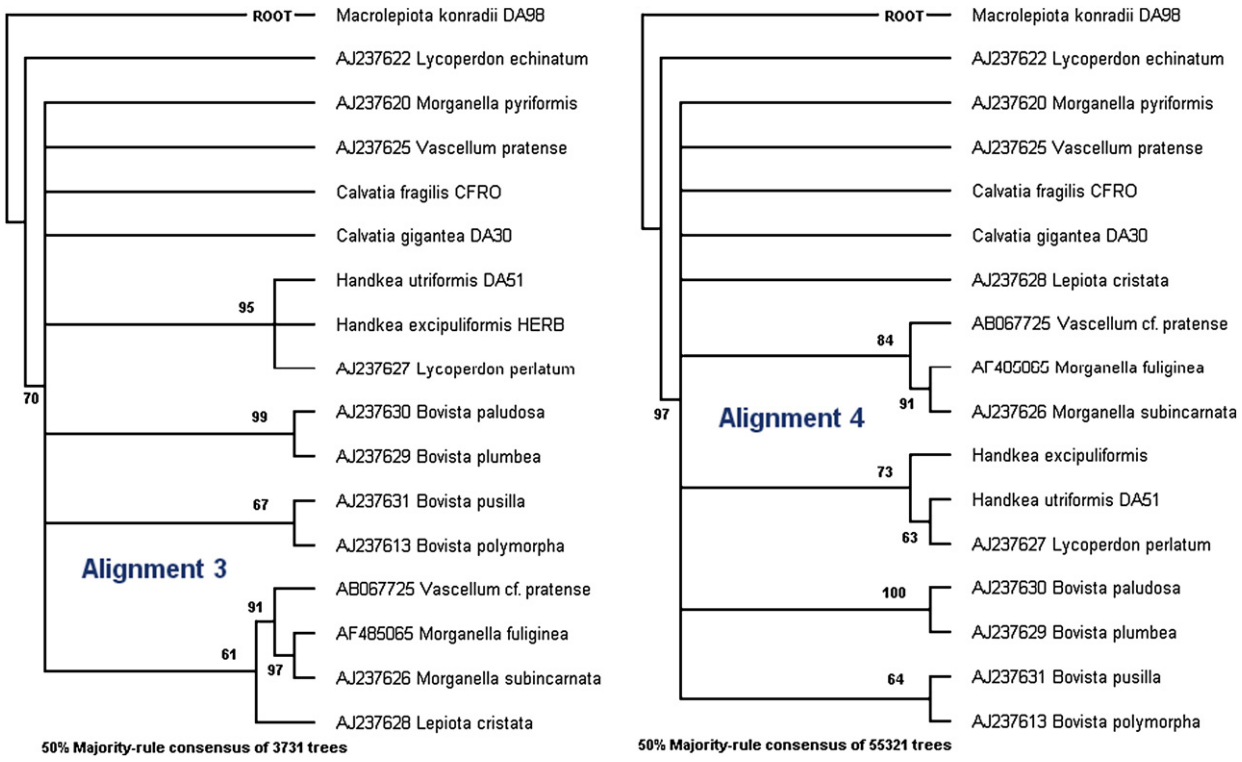

Fig 6 – Alignments 3, 4, 50 % majority rule of jackknife parsimony analysis.

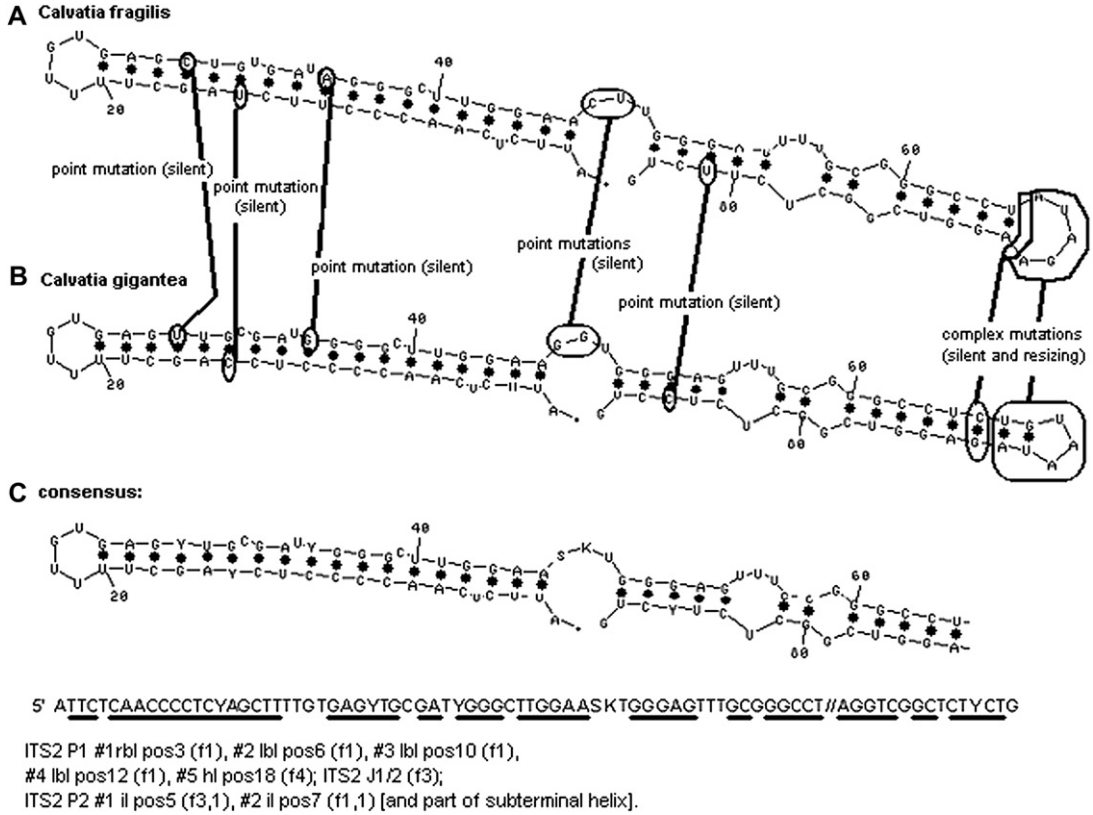

Fig 7 – (A) ITS2 P1 P2 secondary structure for *Calvatia fragilis*. (B) ITS2 P1 P2 secondary structure for *Calvatia gigantean*. (C) Consensus of secondary structure motifs within /Calvatia. Underlined nucleotides in DNA sequence = dsDNA, // = interruption in consensus.

In other words, a bulge is a left bulge loop if one encounters it before the terminal hairpin loop when reading along the primary sequence data from 5′ to 3′. In the descriptors, the following acronyms are used: J1/2, joiner between P1 and P2; pos, position number; hl, terminal hairpin loop; rbl, right bulge loop; lbl, left bulge loop; il, interior loop on both sides; f, number of free bases.

The large *Calvatia* puffballs included in our survey apparently have few differences in their structure (Fig 7), mainly involving silent, one-sided point mutations. A series of mutations may have altered the terminal hairpin loops of P2. The *Handkea* secondary structure (Fig 8) is remarkably stable among included taxa, with the major difference (terminal loop of P2) caused only by inclusion of an ambiguity code in the sequence parsed into Mfold. A number of complementary base changes caused resizing when *Handkea* and *Lycoperdon echinatum* are compared. For example, a resizing deletion of two nucleotides in the P2 stem led to the removal of all three interior loops as inferred for *Handkea*, resulting in a single, one-sided bulge loop in *L. echinatum*. The effect of another single point mutation, T–C in P2, on the structure can not be judged and the relative order of mutational events is unknown.

Both compared *Vascellum* sequences (Fig 9) display only one silent mutation in the P2 stem. However, *Vascellum cf. pratense* (AB067725; perhaps *Vascellum curtisii*, see Terashima *et al.* 2002; Krüger & Kreisel 2003; sequence was not generated by us) appears to have a remarkably shortened P1 length, apparently caused by deletion of 10 nucleotides. This may have led to the formation of a new terminal loop as well as a subterminal interior loop. The pairings otherwise compatible with AJ237625 are still discernible in the folding of AB067725 of Terashima *et al.* (2002) and indicated in Fig 9 as dashed lines. Between *Morganella fuliginea* and *M. subincarnata* (Fig 10) another deletion of many nucleotides, as shown in Fig 9 for *Vascellum*, may have occurred, leading to shorter P1 in *M. subincarnata*. The only other resizing mutations affected the terminal loop of P2. The only structure-altering mutations within /Bovista (Fig 11) here affected the size of the J1/2, the formation of the P2 terminal loop, and the size of the first (basal) interior loop of P2 (double deletion). *Bovista paludosa* and *B. plumbea* have virtually identical ITS2 rRNA structures (Fig 11), only one silent mutation is inferable.

Discussion

Practising cladists are aware that the phylogenies inferred from the limited signal of single genes may not be accurate to describe the overall evolutionary history of the organism (Rokas & Carroll 2005), e.g. due to the reticulate and

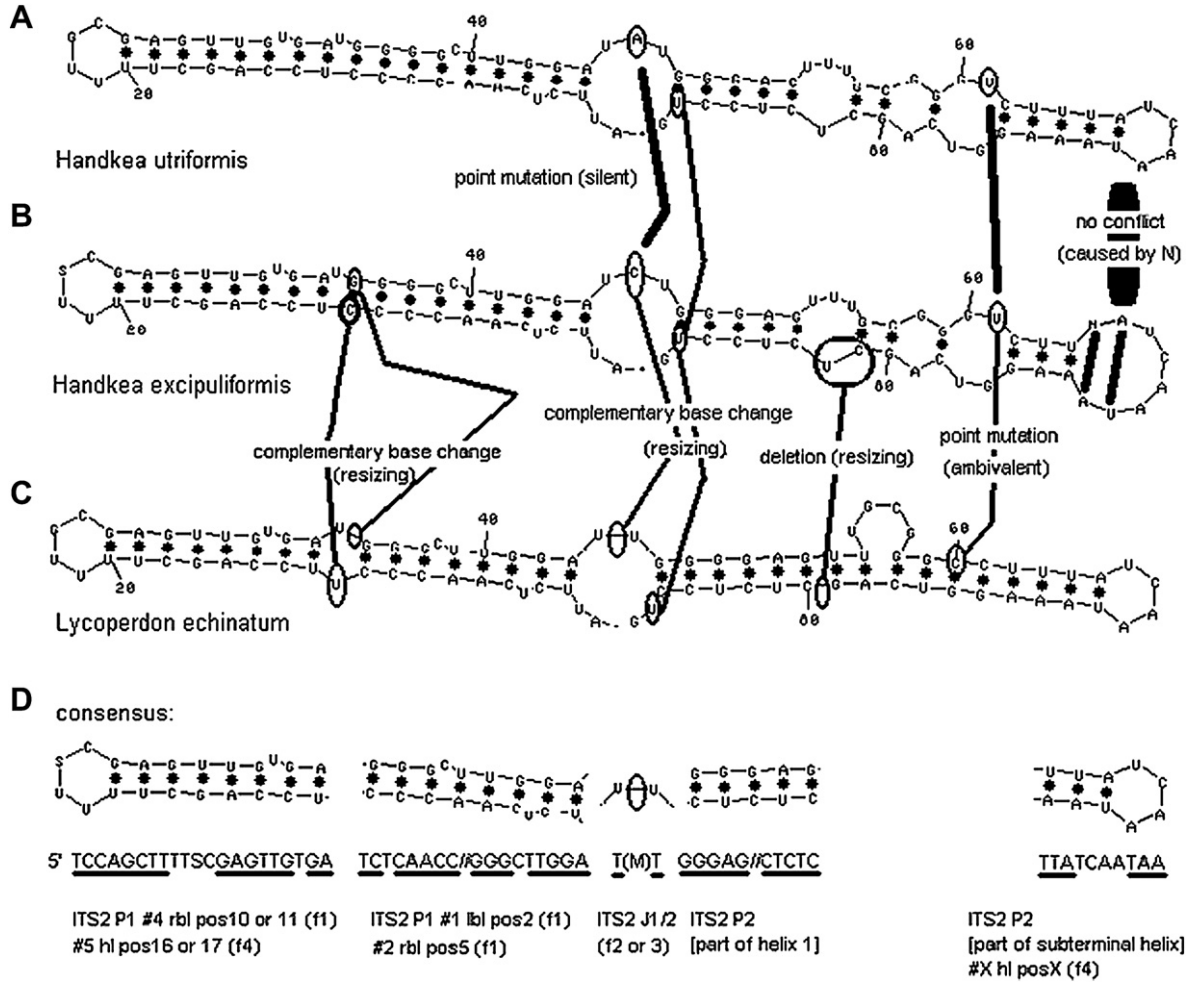

Fig 8 – (A) ITS2 P1 P2 secondary structure for *Handkea utriformis*. (B) ITS2 P1 P2 secondary structure for *Handkea excipuliformis*. (C) ITS2 P1 P2 secondary structure for *Lycoperdon echinatum*. (D) Consensus of secondary structure motifs within /Handkea–Echinatum.

heterochronous nature of gene phylogenies. In our exploration, we consider a tree topology suggested as consensus of most-parsimonious and other trees from the most conservatively retained alignment as the optimally deduced phylogeny. The phylogenetic clades identified are used to compare secondary structure within the clade's ITS2 rRNA structure, suggest evolutionary trends (e.g. a change in structure position and extent) arising from past mutational events, and define potential consensus structures for identifying each organism.

Phylogenetic inference (Alignment 1) and traditional taxonomy

Fig 4 is a compilation of identifiable, presumably robust clades from parsimony and likelihood analyses, with newly introduced cladonyms. The topology, based on parsimony, was also corroborated using likelihood approaches, including Bayesian inference.

Among traditional morphotaxonomic characters, the difference of slits *versus* pores in capillitia appears consistent with *Calvatia* and *Handkea* belonging to different clades. Albeit the clade of *Handkea* may include taxa with pored capillitia as does *Calvatia* (*Lycoperdon echinatum* or *L. perlatum*). The inclusion of *M. pyriformis* in *Morganella* (Krüger & Kreisel 2003) is not supported with any of the datasets and analyses presented here. ML phylogenetic analyses on ITS performed by Bates (2004) indicate that *Disciseda* may actually be closely related to *Morganella pyriformis*. Bates (2004) used both likelihood and parsimony analyses on ITS region sequences of a different set of taxa than that of Krüger *et al.* (2001) and Krüger & Kreisel (2003). The monophyly of *Bovista* here is also not statistically supported by any alignment and any analysis, whereas the two *Bovista* puffballs with the *Bovista*-type capillitium (*B. plumbea* and *B. paludosa*) are usually supported as monophyletic. Bates' phylogenetic analyses also generally show that ITS data still leave most puffballs unresolved.

Effect of different alignment strategies

Alignment programs are unlikely to be efficient enough to recover a true alignment and phylogeny. The direct

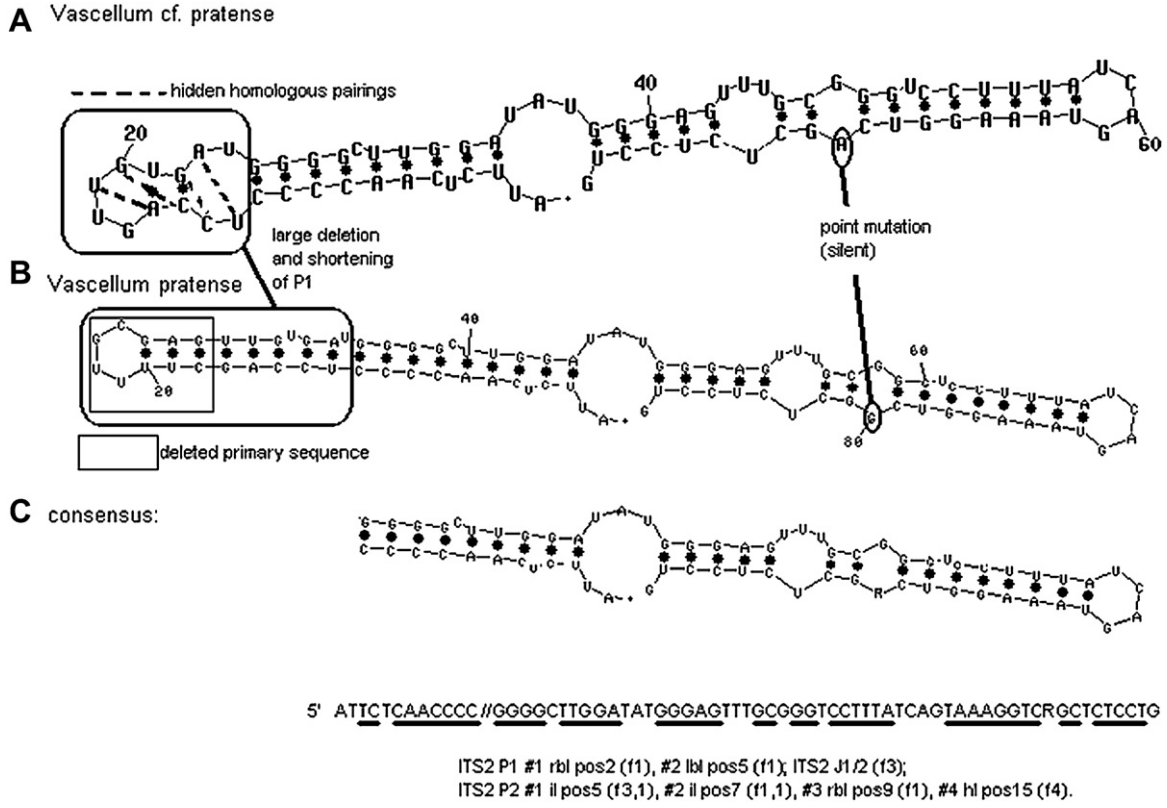

Fig 9 – (A) ITS2 P1 P2 secondary structure for *Vascellum cf. pratense*. (B) ITS2 P1 P2 secondary structure for *Vascellum pratense*. (C) Consensus of secondary structure motifs within /Vascellum.

optimization in POY has very recently been found to be unreliable; (Kjer *et al.* 2007; Ogden & Rosenberg 2007) yet at the time this alignment method is still used in the process of delimiting unambiguous alignment parts. In our study, neither compared recoding treatment of ITS2 ambiguously-alignable parts, nor secondary-structure-informed parts yielded robust phylogenetic estimates consistent with analyses on the conservatively selected nucleotides in Alignment 1, nor those in

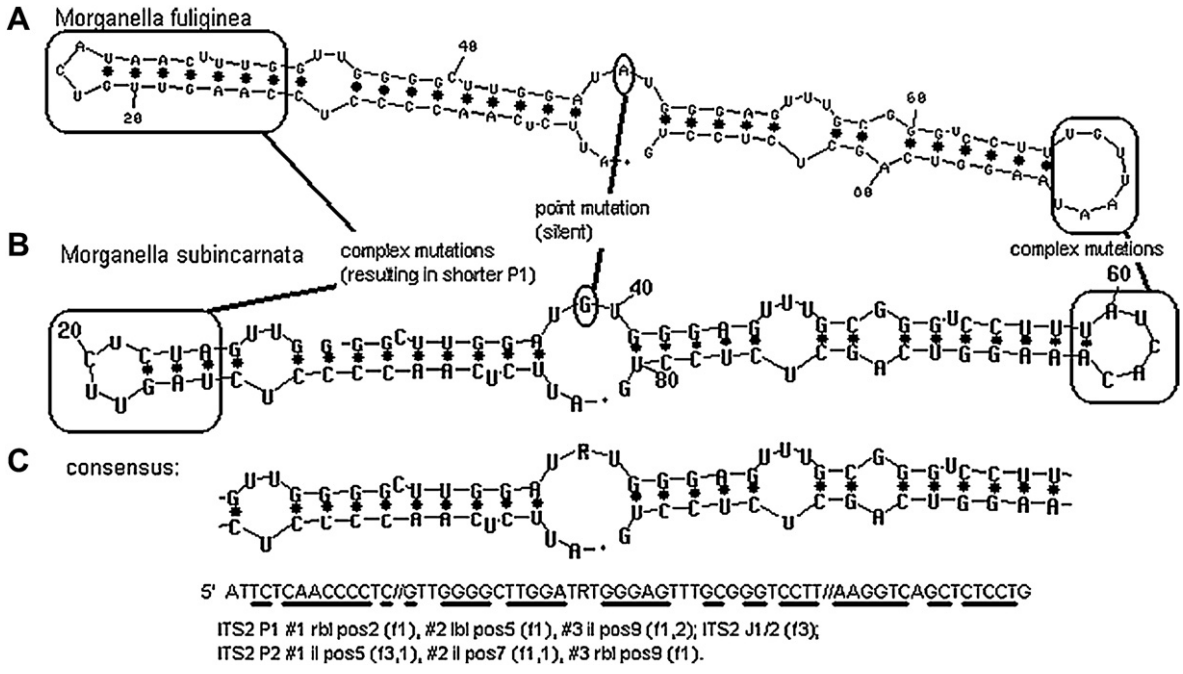

Fig 10 – (A) ITS2 P1 P2 secondary structure for *Morganella fuliginea*. (B) ITS2 P1 P2 secondary structure for *Morganella subincarnata*. (C) Consensus of secondary structure motifs within /Morganella.

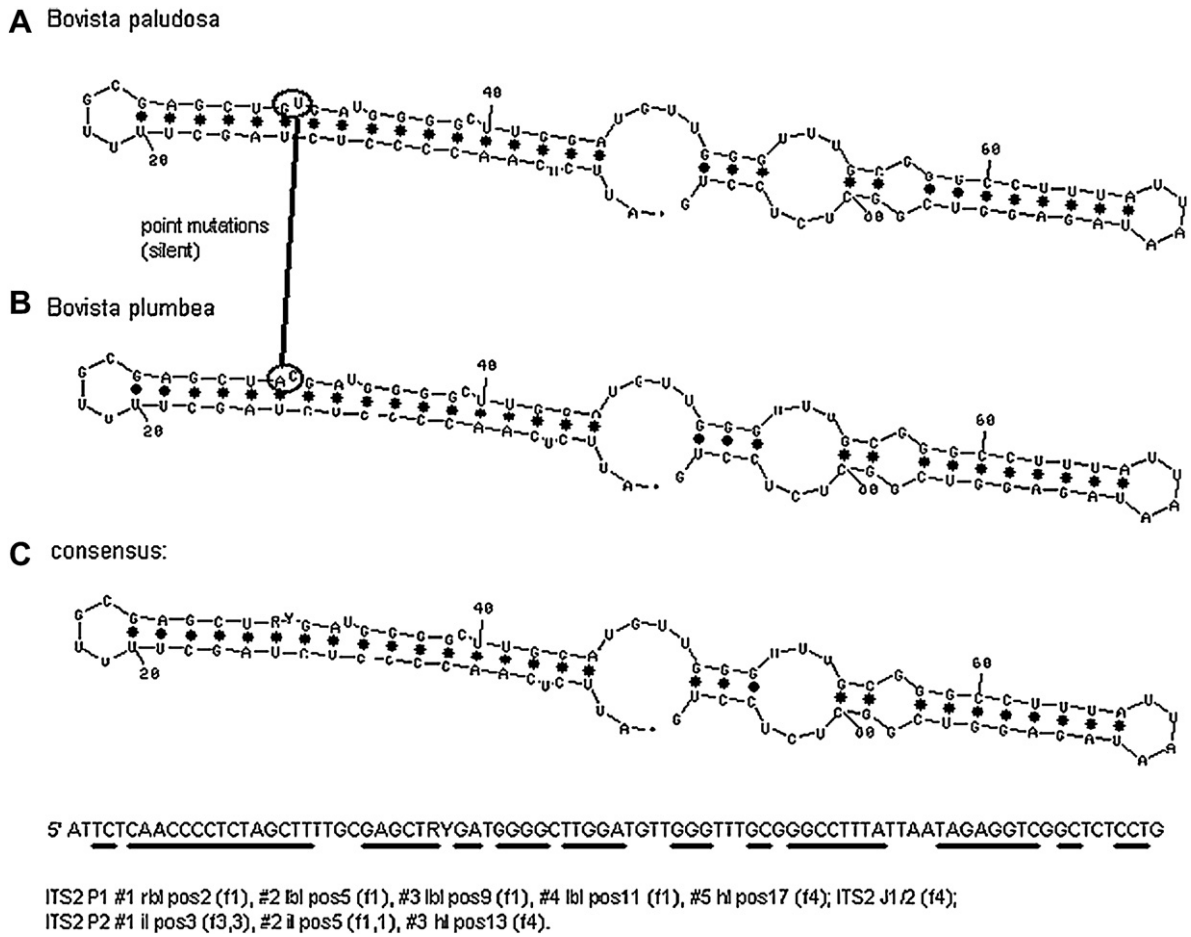

Fig 11 – (A) ITS2 P1 P2 secondary structure for *Bovista paludosa*. (B) ITS2 P1 P2 secondary structure for *Bovista plumbea*. (C) Consensus of secondary structure motifs within within /Plumbea–Paludosa.

previously published works. The applied characters and weighting schemes for secondary structure information are necessarily arbitrary as there are no empirically defined models of evolution considering secondary structure. In Alignments 3 and 4 such weighting causes phylogenies to present with little resolution (Fig 6). This is most pronounced in the placement of *Lepiota cristata* within the puffballs. Trees generated with full Alignment 1 are more statistically robust, with some topological differences between likelihood and parsimony strategies with implications for taxonomically emending genus *Morganella* (Krüger & Kreisel 2003).

The recoding of secondary structure and subsequent problems of positional homology and exclusion/inclusion in analysis are tightly interwoven with alternative treatments of indels, as recently explored by Creer *et al.* (2006) for introns in ribosomal and protein-coding genes. In both cases, models of molecular evolution are unavailable, recoding may result in loss of available informative characters, positional homology might be violated, but a simple recoding may outperform omission of data or more complex schemes of data incorporation. Further, increasing potential character non-independence must be avoided. In our present work, we have not simultaneously used characters describing structure and the underlying nucleotides other than by introduction of characters for the presence and absence of beginning and ending of structural motifs.

Secondary structure used as semi-independent taxonomic character

We suggest continuing the utilization of secondary structure as taxonomic characters for comparison with independently inferred phylogenies (Krüger & Gargas 2004). Future likelihood models of secondary structure evolution should be formulated, making inclusion of structure characters in phylogenetic analysis more feasible.

We here define, with Figs 7–11, taxonomic characters coinciding with phylogenetically defined cladonyms that are supported in either Fig 1 or Supplementary Material Fig. S1. We suggest that independent confirmation of the existence and viability of deduced secondary structure requires experimentally mutating primary sequences or interfering with ssRNA features. For use in phylogenetic analysis and molecular detection or debilitation, a wider survey of these novel molecular taxonomic characters is suggested and would require language implementing a fine-tuned structure nomenclature.

Acknowledgements

A.G. thanks UW Botany Department and D.K. the UW Graduate School for financial support. D.K. also thanks Hanns Kreisel for introduction to gasteromycete taxonomy and access to gasteromycete specimens. Both authors thank Mirna Santana, Marie T. Trest, and Matthew P. Nelsen for helpful comments.

Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.mycres.2007.10.019

REFERENCES

Alfaro ME, Zoller S, Lutzoni F, 2003. Bayes or bootstrap? A simulation study comparing the performance of Bayesian Markov chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. *Molecular Biology and Evolution* **20**: 255–266.

Baldwin BG, Sanderson MJ, Porter JM, Wojiechowski MF, Campbell CS, Donoghue MJ, 1995. The ITS region of nuclear ribosomal DNA: a valuable source of evidence on angiosperm phylogeny. *Annals of the Missouri Botanical Garden* **82**: 247–277.

Bates ST, 2004. *Arizona members of the Geastraceae and Lycoperdaceae (Basidiomycota, Fungi)*. MSc thesis, Arizona State University, Tempe.

Bayev A, Georgiev OI, Hadjiolov AA, Nikolaev N, Skryabin KG, Zakharyev VM, 1981. The structure of the yeast ribosomal RNA genes. 3. Precise mapping of the 18 S and 25 S rRNA genes and structure of the adjacent regions. *Nucleic Acids Research* **9**: 789–799.

Caetano-Annolés G, 2002. Tracing the evolution of RNA structure in ribosomes. *Nucleic Acids Research* **11**: 2575–2587.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR, 2002. The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* **3**: 15.

Cedergren R, Gray MW, Abel Y, Sankoff D, 1988. The evolutionary relationships among known life forms. *Molecular Evolution* **28**: 98–112.

Côte CA, Peculis BA, 2001. Role of the ITS2-proximal stem and evidence for indirect recognition of processing sites in pre-rRNA processing in yeast. *Nucleic Acids Research* **29**: 2106–2116.

Chenna R, Sugawara H, Koike T, Lopez R, Gibson TJ, Higgins DG, Thompson JD, 2003. Multiple sequence alignment with the Clustal series of programs. *Bioinformatics* **31**: 3497–3500.

Choi JH, Jung HY, Kim HS, Cho HG, 2000. PhyloDraw: a phylogenetic tree drawing system. *Bioinformatics* **16**: 1056–1058.

Creer S, Pook CE, Malhotra A, Thorpe RS, 2006. Optimal intron analyses in the trimeresurus radiation of Asian pitvipers. *Systematic Biology* **55**: 57–72.

Cummings MP, Handley SA, Myers DS, Reed DL, Rokas A, Winka K, 2003. Comparing bootstrap and posterior probability values in the four-taxon case. *Systematic Biology* **52**: 477–487.

Desper R, Gascuel O, 2002. Fast and accurate phylogeny reconstruction algorithms based on the minimum-evolution principle. *Journal of Computational Biology* **9**: 687–705.

Desper R, Gascuel O, 2004. Theoretical foundation of the balanced minimum evolution method of phylogenetic inference and its relationship to weighted least-squares tree fitting. *Molecular Biology and Evolution* **21**: 587–598.

Edgar RC, 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**: 113.

Efron B, Gong G, 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. *American Statistician* **37**: 36–48.

Farris JS, Albert VA, Källersjö M, Lipscomb D, Kluge AG, 1996. Parsimony jackknifing outperforms neighbor-joining. *Cladistics* **12**: 99–124.

Felsenstein J, 2005. *PHYLIP (Phylogeny Inference Package)*. Version 3.6. Department of Genome Sciences, University of Washington, Seattle (distributed by the author).

Gardes M, Bruns TD, 1993. ITS primers with enhanced specificity for basidiomycetes—application to the identification of mycorrhizae and rusts. *Molecular Ecology* **2**: 113–118.

Gascuel O, 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution* **14**: 685–695.

Goertzen LR, Cannone JJ, Gutell RR, Jansen RK, 2003. ITS secondary structure derived from comparative analysis: implications for sequence alignment and phylogeny of the *Asteraceae*. *Molecular Phylogenetics and Evolution* **29**: 216–234.

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

Hawksworth DL, Kirk PM, Sutton BC, Pegler DN, 1995. *Ainsworth & Bisby's Dictionary of the Fungi*, 8th edn. CAB International, Wallingford.

Hershkovitz MA, Lewis LA, 1996. Deep-level diagnostic value of the rDNA-ITS region. *Molecular Biology and Evolution* **13**: 1276–1295.

Hershkovitz MA, Zimmer EA, 1996. Conservation patterns in angiosperm rDNA ITS2 sequences. *Nucleic Acids Research* **24**: 2857–2867.

Hibbett DS, Pine EM, Langer E, Langer G, Donoghue MJ, 1997. Evolution of gilled mushrooms and puffballs inferred from ribosomal DNA sequences. *Proceedings of the National Academy of Sciences, USA* **94**: 12002–12006.

Horner WE, Levetin E, Lehrer SB, 1993. Basidiospore allergen release: elution from intact spores. *Journal of Allergy and Clinical Immunology* **92**: 306–312.

Ingold CT, 1971. *Fungal Spores: their Liberation and Dispersal*. Clarendon Press, Oxford.

Iosifidou P, Agerer R, 2002. Die Rhizomorphen von *Gastrosporium simplex* und einige Gedanken zur systematischen Stellung der *Gastrosporiaceae* (*Hymenomycetes*, *Basidiomycota*). *Feddes Repertorium* **113**: 11–23.

Jobes DV, Thien LB, 1997. A conserved motif in the 5.8S ribosomal RNA (rRNA) gene is a useful diagnostic marker for plant internal transcribed spacer (ITS) sequences. *Plant Molecular Biology Reporter* **15**: 326–334.

Katoh K, Kuma K, Toh H, Miyata T, 2005. MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Research* **33**: 511–518.

Kjer KM, Gillespie JJ, Ober KA, 2007. Opinions on multiple sequence alignment, and an empirical comparison of

repeatability and accuracy between POY and structural alignment. *Systematic Biology* **56**: 133–146.
Knudsen B, Hein JJ, 2003. Pfold: RNA secondary structure prediction using stochastic context-free grammars. *Nucleic Acids Research* **31**: 3423–3428.
Kreisel H, 1989. Studies in the *Calvatia* complex (basidiomycetes). *Nova Hedwigia* **48**: 281–296.
Krüger D, Binder M, Fischer M, Kreisel H, 2001. The *Lycoperdales*. A molecular approach to the systematics of some gasteroid mushrooms. *Mycologia* **93**: 947–957.
Krüger D, 2002. *Monographic Studies in the Genus Polyporus* (*Basidiomycotina*). PhD thesis, University of Tennessee, Knoxville.
Krüger D, Kreisel H, 2003. Proposing *Morganella* subgen. *Apioperdon* subgen. nov. for the Puffball *Lycoperdon pyriforme*. *Mycotaxon* **86**: 169–177.
Krüger D, Gargas A, 2004. The basidiomycete genus *Polyporus*—an emendation based on phylogeny and putative secondary structure of ribosomal RNA molecules. *Feddes Repertorium* **115**: 530–546.
Krüger D, Petersen RH, Hughes KW, 2006. Molecular phylogenies and mating study data in *Polyporus* with special emphasis on group 'Melanopus' (*Basidiomycota*). *Mycological Progress* **5**: 185–206.
Lalev AI, Nazar RN, 1999. Structural equivalence in the transcribed spacers of pre-rRNA transcripts in *Schizosaccharomyces pombe*. *Nucleic Acids Research* **27**: 3071–3078.
Lalev AI, Nazar RN, 2001. A chaperone for ribosome maturation. *Journal of Biological Chemistry* **276**: 16655–16659.
Lalev AI, Abeyrathne PD, Nazar RN, 2000. A RAC protein-binding site in the internal transcribed spacer 2 of pre-rRNA transcripts from *Schizosaccharomyces pombe*. *Journal of Molecular Biology* **302**: 65–77.
Lee MSY, 2001. Unalignable sequences and molecular evolution. *Trends in Ecology and Evolution* **16**: 681–685.
Levetin E, Horner WE, Lehrer SB, 1992. Morphology and allergic properties of basidiospores from four *Calvatia* spp. *Mycologia* **84**: 759–767.
Lutzoni F, Wagner P, Reeb V, Zoller S, 2000. Integrating ambiguously aligned regions of DNA sequences in phylogenetic analyses without violating positional homology. *Systematic Biology* **49**: 628–651.
Mai JC, Coleman AW, 1997. The internal transcribed spacer 2 exhibits a common secondary structure in green algae and flowering plants. *Journal of Molecular Evolution* **44**: 258–271.
Makarenkov V, 2001. T-REX: reconstructing and visualizing phylogenetic trees and reticulation networks. *Bioinformatics* **17**: 664–668.
Moncalvo J-M, Vilgalys R, Redhead SA, Johnson JE, James TY, Aime MC, Hofstetter V, Verduin SJW, Larsson E, Baroni TJ, Thorn RG, Jacobsson S, Clémençon H, Miller Jr OK, 2002. One hundred and seventeen clades of euagarics. *Molecular Phylogenetics and Evolution* **23**: 357–400.
Morgenstern B, 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**: 211–218.
Morgenstern B, Goel S, Sczyrba A, Dress A, 2003. AltAVisT: comparing alternative multiple sequence alignments. *Bioinformatics* **19**: 425–426.
Mueller UG, Rehner SA, Schultz TR, 1998. The evolution of agriculture in ants. *Science* **281**: 2034–2038.
Notredame C, Higgins DG, Heringa J, 2000. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *Journal of Molecular Biology* **302**: 205–217.
Nylander JAA, 2004. *MrModeltest v.2.2*. Evolutionary Biology Centre, Uppsala University, Uppsala, Distributed by the author.
Ogden TH, Rosenberg MS, 2007. Alignment and topological accuracy of the direct optimization approach via POY and traditional phylogenetics via ClustalW + PAUP*. *Systematic Biology* **56**: 182–193.
Posada D, Crandall KA, 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
Rees WA, Yager TD, Korte J, von Hippel PH, 1993. Betaine can eliminate the base pair composition dependence of DNA melting. *Biochemistry* **32**: 137–144.
de Rijk P, Wuyts J, de Wachter R, 2003. RnaViz 2: an improved representation of RNA secondary structure. *Bioinformatics* **19**: 299–300.
Robinson DR, Foulds L, 1981. Comparison of phylogenetic trees. *Mathematical Biosciences* **53**: 131–147.
Rokas A, Carroll SB, 2005. More genes or more taxa? The relative contribution of gene number and taxon number to phylogenetic accuracy. *Molecular Biology and Evolution* **22**: 1337–1344.
Rubin GM, 1973. The nucleotide sequence of *S. cerevisiae* 5.8S rRNA. *Journal of Biological Chemistry* **248**: 3860–3875.
Sammeth M, Griebel T, Tille F, Stoye J, 2006. Panta rhei (QAlign2): an open graphical environment for sequence analysis. *Bioinformatics* **22**: 889–890.
Smith AD, Lui TW, Tillier ER, 2004. Empirical models for substitution in ribosomal RNA. *Moleclar Biology Evolution* **21**: 419–427.
Sneath PHA, Sokal RR, 1973. *Numerical Taxonomy—The Principles and Practice of Numerical Classification*. WH Freeman, San Francisco.
Stoye J, 1998. Multiple sequence alignment with the divide-and-conquer method. *Gene* **211**: GC45–GC56.
Swofford DL, 2002. *PAUP*—Phylogenetic analysis using parsimony (*and other methods)*, version 4.0. Sinauer Associates, Sunderland, MA.
Terashima Y, Ogiwara K, Kojima M, Kubo C, Seki A, Fujie A, 2002. Primers based on specific ITS sequences of rDNAs for PCR detection of two fairy ring fungi on turfgrass, *Vascellum pratense* and *Lycoperdon pusillum*. *Mycoscience* **43**: 261–265.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876–4882.
van de Peer Y, de Rijk P, Wuyts J, Winkelmans T, de Wachter R, 2000. The European small subunit ribosomal RNA database. *Nucleic Acids Research* **28**: 175–176.
van Walle I, Lasters I, Wyns L, 2004. Align-m—a new algorithm for multiple alignment of highly divergent sequences. *Bioinformatics* **20**: 1428–1435.
Vellinga EC, 2004. Genera in the family *Agaricaceae*: evidence from nrITS and nrLSU sequences. *Mycological Research* **108**: 354–377.
Waugh A, Gendron P, Altman R, Brown JW, Case D, Gautheret D, Harvey SC, Leontis N, Westbrook J, Westhof E, Zuker M, Major F, 2002. RNAML: a standard syntax for exchanging RNA information. *RNA* **8**: 707–717.
Wheeler WC, Gladstein DS, de Laet J, 2003. *POY*. Version 3.0. Available from: ftp.amnh.org/pub/molecular/poy
White TJ, Bruns T, Lee S, Taylor JW, 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (eds), *PCR Protocols: a guide to methods and applications*. Academic Press, San Diego, pp. 315–321.
Wolf M, Achtziger M, Schultz J, Dandekar T, Müller T, 2005. Homology modeling revealed more than 20,000 rRNA internal transcribed spacer 2 (ITS2) secondary structures. *RNA* **11**: 1616–1623.
Xia X, Xie Z, Kjer KM, 2003. 18S Ribosomal RNA and tetrapod phylogeny. *Systematic Biology* **52**: 283–295.

Yang Z, 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* **15**: 555–556.

Young ND, Healy J, 2001. *GapCoder: a computer program for including indels in phylogenetic analysis*. Available from authors.

Zuker M, 1989. On finding all suboptimal foldings of an RNA molecule. *Science* **244**: 48–52.

Zuker M, 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**: 3406–3415.