MycoKeys 4: 1–8 (2012)
doi: 10.3897/mycokeys.4.2619
www.pensoft.net/journals/mycokeys

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

Cylindrocladiella hahajimaensis, a new species of *Cylindrocladiella* transferred from *Verticillium*

Patrik Inderbitzin[1], Richard M. Bostock[1], Krishna V. Subbarao[1]

1 *Department of Plant Pathology, University of California Davis, One Shields Avenue, Davis, CA 95616, USA*

Corresponding author: *Krishna V. Subbarao* (kvsubbarao@ucdavis.edu)

Academic editor: *Andrew Miller* | Received 5 January 2011 | Accepted 6 June 2012 | Published 11 June 2012

Citation: Inderbitzin P, Bostock RM, Subbarao KV (2012) *Cylindrocladiella hahajimaensis*, a new species of *Cylindrocladiella* transferred from *Verticillium*. MycoKeys 4: 1–8. doi: 10.3897/mycokeys.4.2619

Abstract

We used phylogenetic analyses based on ribosomal internal transcribed spacer (ITS) and beta-tubulin (*TUB*) sequences and determined that the correct name of '*Verticillium hahajimaense*' was *Cylindrocladiella hahajimaensis*. A closest relative could not be determined, since *C. hahajimaensis* clustered within the *C. infestans* species complex, a poorly resolved group of taxa. *Cylindrocladiella hahajimaensis* differed from all other members of the *C. infestans* species complex by at least 18 substitutions at the two loci. Morphological characters supported the placement within *Cylindrocladiella*. In addition to the verticillate conidiophores mentioned in the type description, we found evidence for the presence of penicillate conidiophores. Other differences to the type description included the presence of yellow to brown-pigmented hyphae, and the hyaline instead of pigmented chlamydospores.

Keywords

Ascomycetes, Nectriaceae, taxonomy, phylogenetics

Introduction

The genus *Cylindrocladiella* comprises a small, monophyletic group of fungi in the Nectriaceae related to *Cylindrocladium* (Schoch et al. 2000). *Cylindrocladium* and *Cylindrocladiella* are morphologically similar, they form two different synanamorphs that resemble *Penicillium* and *Verticillium* in terms of branching pattern and phialide arrangement, but may bear long hyphal projections with swollen tips that are referred to

as filaments with terminal vesicles. Chlamydospores may also be present (Boesewinkel 1982). Differences between *Cylindrocladiella* and *Cylindrocladium* include aspects of conidiophore branching, filament morphology, appearance in culture, and conidia morphology and size (Boesewinkel 1982). The teleomorph of *Cylindrocladiella* where known, belongs to *Nectricladiella* (Schoch et al. 2000). *Cylindrocladiella* comprises ten species that have been isolated from soil as well as plant tissues where they may act as pathogens (van Coller et al. 2005; Zhang and Chi 1996).

Verticillate conidiophores that consist of narrow, elongate phialides arranged in whorls along a main axis, have evolved multiple times as determined by molecular data and phylogenetic analyses. Taxonomic conclusions from these studies have resulted in the traditional genus *Verticillium* being restricted to a small, monophyletic group of plant pathogens in the Plectosphaerellaceae (Gams et al. 2005; Zare et al. 2007) with the majority of former *Verticillium* species transferred to other genera, including *Lecanicillium* (Gams and Zare 2001; Zare and Gams 2001a; Zare and Gams 2008), *Pochonia* and *Haptocillium* (Zare and Gams 2001b; Zare et al. 2001), as well as *Gibellulopsis* and *Musicillium* (Zare et al. 2007).

During recent studies of *Verticillium* evolution and taxonomy (Inderbitzin et al. 2011a; Inderbitzin et al. 2011b), we received the ex-holotype culture of '*V. hahajimaense*' isolated from soil in Japan using cucumber seeds as bait (Watanabe et al. 2001). Phylogenetic analyses showed that '*V. hahajimaense*' belonged to *Cylindrocladiella* instead as documented and discussed below.

Methods

Taxon sampling, origins of fungal strains and growth conditions

An ex-holotype culture of '*V. hahajimaense*' (strain MAFF 238172, PD684) was obtained free of charge from the National Institute of Agrobiological Sciences Genebank (NIAS), Japan, single spore purified and maintained in a glycerol solution at -80°C (Inderbitzin et al. 2011a). For all 42 ingroup taxa used by van Coller et al. (2005), ribosomal internal transcribed spacer sequences (ITS) and beta-tubulin sequences (*TUB*) were downloaded from GenBank.

DNA extraction, PCR amplification, and DNA sequencing

For complete details see Inderbitzin et al. (2011b). The ITS region was PCR amplified using primers ITS1-F (Gardes and Bruns 1993) and ITS4, and sequenced with ITS5 and ITS4 (White et al. 1990), part of *TUB* was PCR amplified and sequenced with VTubf2/VTubR (Inderbitzin et al. 2011b).

Phylogenetic analyses

Three different datasets were analyzed, the single locus datasets ITS and *TUB* using parsimony as implemented in PAUP v.4.0b 10 (Swofford 2002), as well as a the combined ITS plus *TUB* dataset, using parsimony and likelihood implementations in PAUP, and a Bayesian approach using MRBAYES v3.0b4 (Ronquist and Huelsenbeck 2003). The likelihood and Bayesian analyses used an optimal model of DNA evolution determined using MODELTEST 3.7 (Posada and Crandall 1998). DNA sequences were assembled and aligned in GENEIOUS v4.8.5 (Drummond et al. 2010). For complete details of phylogenetic analyses, see Inderbitzin et al. (2011b).

Microscopy

Morphological observations were from cultures grown on PDA. Microscopy was performed using a Leica DM5000 B microscope (Leica Microsystems CMS GmbH, Wetzlar, Germany), with differential interference contrast (DIC) illumination of specimens mounted in water. Photographs were taken with a Leica DFC310 FX camera, using LEICA APPLICATION SUITE VERSION 3.6.0 software.

Results

DNA sequence data obtained and phylogenetic analyses

The ITS and *TUB* sequences of *Cylindrocladiella hahajimaensis* strain PD684 measured 462 and 480 bp in length, respectively (GenBank JN687561, JN687562), and were aligned with the *Cylindrocladiella* ingroup sequences by van Coller et al. (2005) resulting in 43 taxa datasets submitted to TreeBase (http://purl.org/phylo/treebase/phylows/study/TB2:S11935). ITS and *TUB*-derived most parsimonious trees were identical on a 70% bootstrap support level (data not shown), and were combined into one, 950 character dataset that was analyzed using parsimony, likelihood and Bayesian inference of phylogeny, the latter two methods implementing a HKY+I+G model of nucleotide substitution. The 22 most parsimonious trees were 227 steps in length and differed by branches that were supported by less then 70% of the bootstrap replicates (data not shown). One most parsimonious tree was midpoint rooted and is shown in Fig. 1. The most likely tree (-ln likelihood = 2580.47), and the Bayesian tree were congruent with the most parsimonious tree on 70% bootstrap and posterior probability levels, respectively (data not shown, but see Fig. 1 for support values). The tree topology obtained was as in van Coller (2005) for the taxa shared between the two studies, but with lower branch supports probably because our analyses included only two and not three loci. *Cylindrocladiella hahajimaensis* grouped in a well supported, but poorly resolved clade together with *C. lageniformis*, *C. infestans*, *Nectricladiella infestans*, *C. viticola* and a

Cylindrocladiella sp. (Fig. 1). This group, without *C. lageniformis*, was referred to as the '*C. infestans* species complex' by van Coller et al (2005). Among its closest relatives, *C. hahajimaensis* was most similar to *N. infestans* from which it differed by 18 substitutions in the 950 bp, combined ITS and *TUB* dataset.

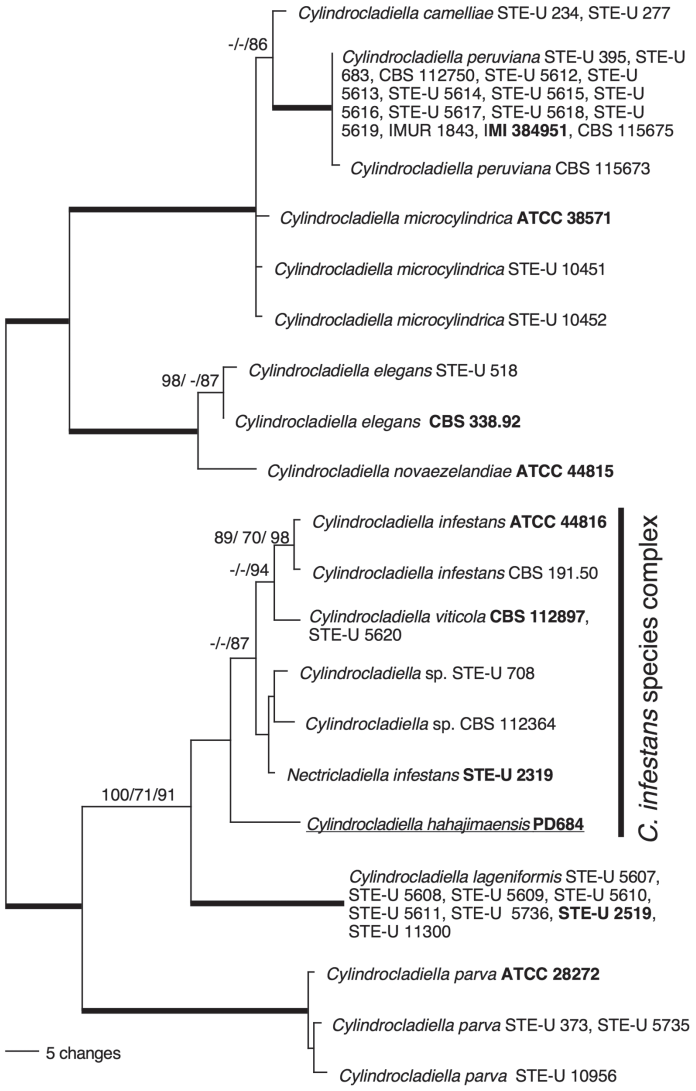


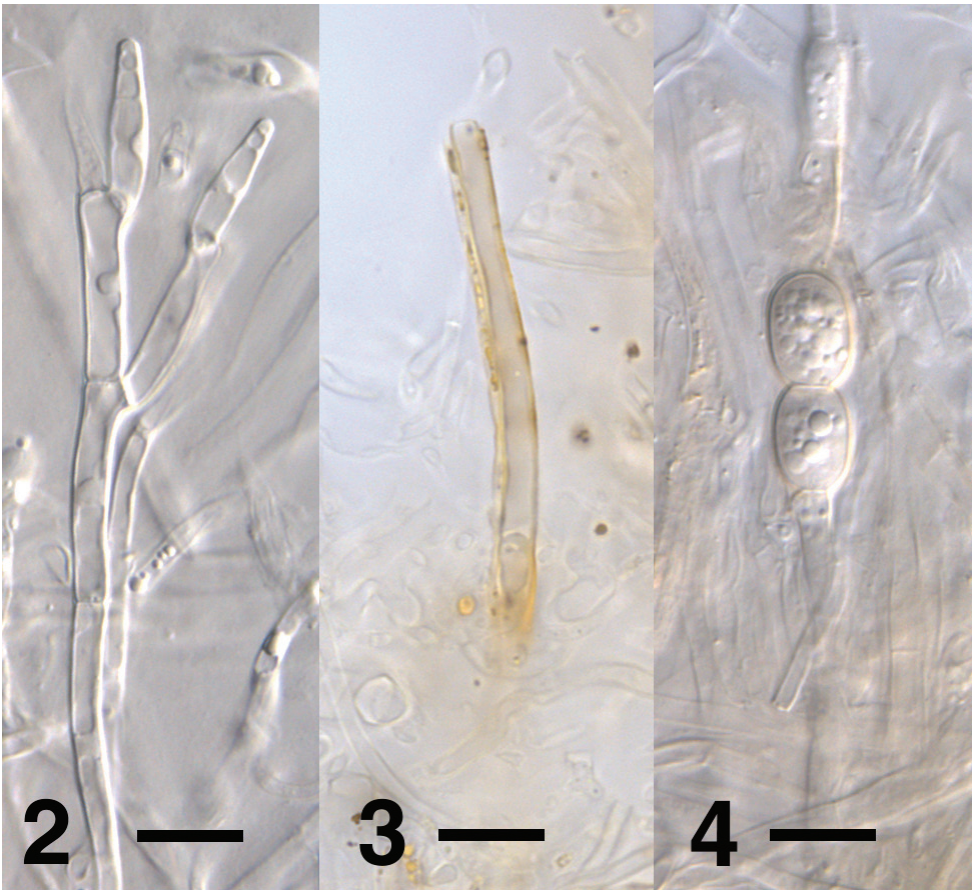
Figure 1. Phylogenetic position of *Cylindrocladiella hahajimaensis* within *Cylindrocladiella* based on a combined ITS and *TUB* dataset of 950 characters and 43 taxa. The tree is midpoint rooted. One of the most parsimonious trees is shown. Species names are followed by strain identifiers for each strain included in this study, identifiers in bold face represent ex-type cultures. *Cylindrocladiella hahajimaensis* is underlined. The *C. infestans* species complex is delimited by a vertical bar on the right. Numbers by the branches are parsimony, Bayesian and likelihood support values above 70 in that order, branches in bold had maximal support in all analyses.

Taxonomy

***Cylindrocladiella hahajimaensis* (Ts. Watanabe) Inderb., R.M.Bostock and K.V.Subbarao, comb. nov.**
Mycobank: MB 483222
Figures 2–4

º*Verticillium hahajimaense* Ts. Watanabe, Mycoscience 42: 594 (2001).

Commentary: *Cylindrocladiella hahajimaensis* was formerly placed in *Verticillium* because of the presence of verticillate conidiophores which were illustrated by Watanabe (2001), together with the chlamydospores and the conidia. We confirmed the presence of these structures, and also found evidence for the presence of penicillate conidiophores (Fig. 2). However, *C. hahajimaensis* strain PD684 conidiated sparsely, and we



Figures 2–4. Select morphological features of *Cylindrocladiella hahajimaensis* strain PD684 (ex-holotype). **2** Potential penicillate conidiophore after eight days on PDA **3** Yellow-pigmented hypha after 15 days on PDA **4** A pair of thick-walled hyaline chlamydospores with brown tinge after 27 days on PDA. Scale bar = 10 µm; Imaging method: DIC.

were unable to conclusively assess the morphology of the penicillate conidiophores. In agreement with Watanabe (2001), no filaments with terminal vesicles were seen. Watanabe (2001) described the colony color as 'Sudan brown or snuff brown', and the chlamydospores were described as brown. However, we found that the brown colony color was caused primarily by yellow to brown-pigmented hyphae (Fig. 2) not mentioned by Watanabe (2001), the chlamydospores remained hyaline with a brown tinge after 27 days of incubation. We did not find any additional cultures of *C. hahajimaensis* at CBS, NIAS or ATCC. More strains of *C. hahajimaensis* will need to be examined to update the description of this fungus.

Discussion

Cylindrocladiella hahajimaensis is the latest member of *Cylindrocladiella* which now consists of eleven species (van Coller et al. 2005; Zhang and Chi 1996). Our phylogenetic analyses showed that *C. hahajimaensis* was nested within *Cylindrocladiella* with high support (Fig. 1), and was thus a member of this genus. Morphology did not contradict this placement. The original description of *C. hahajimaensis* mentioned verticillate conidiophores and chlamydospores which are present in *Cylindrocladiella* (Boesewinkel 1982). However, the original description did not mention other characteristics of *Cylindrocladiella* including penicillate conidiophores and filaments with terminal vesicles. We found evidence for penicillate conidiophores (Fig. 1), but did not find any filaments with terminal vesicles. However, the ex-holotype culture we examined conidiated sparsely, preventing us from corroborating our observations and amending the *C. hahajimaensis* type description.

Our analyses included DNA sequence data derived from ex-type cultures of all known *Cylindrocladiella* species, except for *C. tenuis* which has larger conidia than *C. hahajimaensis* (Zhang and Chi 1996). Also, none of the species described by Lombard et al. (2012) were included in our analyses. This is because we were unable to replicate the topology of the phylogenetic tree in Lombard et al.'s (2012) figure 1, both based on DNA sequence data retrieved from GenBank, and based on a nexus file with MrBayes block provided by the authors. We do not know the cause of this divergence, but it was consistent and independent of the method of analysis and hardware used. Despite excluding the most recent data available, the analyses we presented here showed strong support that *C. hahajimaensis* belonged to *Cylindrocladiella,* mainly to the *C. infestans* species complex (van Coller et al. 2005), a group with uncertain species boundaries. *Cylindrocladiella hahajimaensis* differed from the most similar member of the *C. infestans* species complex included in van Coller et al. (2005) by a considerable 18 nucleotide substitutions at ITS and *TUB*. More analyses with a larger number of isolates, including the species in Lombard et al. (2012) are needed to resolve the branching order and species boundaries in the *C. infestans* species complex, and to determine whether any more recently described species of *Cylindrocladiella* may be synonyms of *C. hahajimaensis*.

Acknowledgements

We would like to thank the National Institute of Agrobiological Sciences Genebank (NIAS), Japan for providing the *C. hahajimaensis* culture free of charge and Lorenzo Lombard and Pedro Crous for providing the nexus file. Funding was provided by the California Leafy Greens Board and the United States Department of Agriculture – National Institute of Food and Agriculture (USDA-NIFA).

References

Boesewinkel HJ (1982) *Cylindrocladiella*, a new genus to accommodate *Cylindrocladium parvum* and other small-spored species of *Cylindrocladium*. Canadian Journal of Botany 60: 2288–2294.

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v4.8.5, Available from http://www.geneious.com

Gams W, Zare R (2001) A revision of *Verticillium* sect. *Prostrata*. III. Generic classification. Nova Hedwigia 72: 329–337.

Gams W, Zare R, Summerbell RC (2005) Proposal to conserve the generic name *Verticillium* (anamorphic Ascomycetes) with a conserved type. Taxon 54: 179.

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

Inderbitzin P, Bostock RM, Davis RM, Usami T, Platt HW, Subbarao KV (2011a) Phylogenetics and taxonomy of the fungal vascular wilt pathogen *Verticillium*, with the descriptions of five new species. PLoS ONE 6: e28341.

Inderbitzin P, Davis RM, Bostock RM, Subbarao KV (2011b) The ascomycete *Verticillium longisporum* is a hybrid and a plant pathogen with an expanded host range. PLoS ONE 6: e18260.

Lombard L, Shivas R, To-Anun C, Crous P (2012) Phylogeny and taxonomy of the genus *Cylindrocladiella*. Mycological Progress: 1–34.

Posada D, Crandall KA (1998) Modeltest: Testing the model of DNA substitution. Bioinformatics 14: 817–818.

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574.

Schoch CL, Crous PW, Wingfield MJ, Wingfield BD (2000) Phylogeny of *Calonectria* and selected hypocrealean genera with cylindrical macroconidia. Studies in Mycology 45: 45–62.

Swofford DL (2002) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts, pp.

van Coller GJ, Denman S, Groenewald JZ, Lamprecht SC, Crous PW (2005) Characterisation and pathogenicity of *Cylindrocladiella* spp. associated with root and cutting rot symptoms of grapevines in nurseries. Australasian Plant Pathology 34: 489–498.

Watanabe T, Watanabe Y, Fukatsu T (2001) New species of *Acremonium*, *Cylindrocarpon* and *Verticillium* from soil in the Bonin (Ogasawara) Islands, Japan. Mycoscience 42: 1340–3540.

White TJ, Bruns TD, Lee SB, 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. Academic Press, Inc., San Diego, 315–322.

Zare R, Gams W (2001a) A revision of *Verticillium* section *Prostrata*. IV. The genera *Lecanicillium* and *Simplicillium* gen. nov. Nova Hedwigia 73: 1–50.

Zare R, Gams W (2001b) A revision of *Verticillium* section *Prostrata*. VI. The genus *Haptocillium*. Nova Hedwigia 73: 271–292.

Zare R, Gams W (2008) A revision of the *Verticillium fungicola* species complex and its affinity with the genus *Lecanicillium*. Mycological Research 112: 811–824.

Zare R, Gams W, Evans HC (2001) A revision of *Verticillium* section *Prostrata.* V. The genus *Pochonia*, with notes on *Rotiferophthora*. Nova Hedwigia 73: 51–86.

Zare R, Gams W, Starink-Willemse M, Summerbell RC (2007) *Gibellulopsis*, a suitable genus for *Verticillium nigrescens*, and *Musicillium*, a new genus for *V. theobromae*. Nova Hedwigia 85: 463–489.

Zhang C, Chi P (1996) A new species of *Cylindrocladiella*, an unrecorded genus from China. Acta Mycologica Sinica 15: 170–172.

MycoKeys 4: 9–21 (2012)
doi: 10.3897/mycokeys.4.2809
www.pensoft.net/journals/mycokeys

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

A further new species in the lichen genus *Arctomia*: *A. borbonica* from Reunion (Mascarene archipelago)

Nicolas Magain[1], Emmanuël Sérusiaux[1]

1 *Evolution and Conservation Biology Unit, University of Liège, Sart Tilman B22, B-4000 Liège, Belgium*

Corresponding author: *Nicolas Magain* (nicolas.magain@doct.ulg.ac.be)

Academic editor: *Imke Schmitt* | Received 1 February 2012 | Accepted 18 June 2012 | Published 20 July 2012

Citation: Magain N, Sérusiaux E (2012) A further new species in the lichen genus *Arctomia*: *A. borbonica* from Reunion (Mascarene archipelago). MycoKeys 4: 9–21. doi: 10.3897/mycokeys.4.2809

Abstract

Arctomia borbonica **sp. nov.** is described as new for science from montane natural and secondary habitats in Reunion in the Mascarene archipelago (Indian Ocean). It has a sterile, foliose, usually wrinkled, thallus whose margins produce goniocysts that disintegrate into a soredioid margin; it looks like a *Leptogium* species. Its phylogenetic position in the Arctomiaceae (Ostropomycetidae, Ascomycota) has been determined with 3 genes (nuLSU, mtSSU, *RPB1*) inferences.

Key words

Ascomycota, Ostropomycetidae, Arctomiaceae, *Arctomia*, phylogenetic inferences, nuLSU, mtSSU, *RPB1*, Reunion, Mascarene archipelago

Introduction

Within the Lecanoromycetes, the subclass Ostropomycetidae Reeb, Lutzoni and Cl. Roux exhibits an impressive diversity of ascomata, thallus forms and ecological requirements. The phylogenetic relationships between genera and families are poorly resolved (Baloch et al. 2010), although impressive progress has been recently achieved for the Graphidaceae (incl. Thelotremataceae), the second largest family of lichenized fungi (Rivas Plata et al. 2012). Many taxa within the subclass still require detailed phylogenetic studies. Indeed, modern statistical methods within a phylogenetical context using several loci sequences yielded interesting and quite unexpected results, such as the polyphyly of two well-known genera. *Graphis* is now resolved into two strongly

supported clades, nested within a large clade comprizing other well-known genera such as *Diorygma, Glyphis* and *Phaeographis* (tribe Graphideae; Rivas Plata et al. 2011). Further, *Pertusaria* is resolved into four strongly supported groups: *Pertusaria* s. str. (incl. the type species *P. pertusa*), *Pertusaria* s. l. 1 including *P. amara*, *P.* s. l. 2 including *P. lactea* and *P. velata*, and a fourth group, comprizing the species with gyalectoid ascomata and recently recognized as the new genus *Gyalectaria* (Schmitt et al. 2010).

Within such a large and very much unresolved variation, the case of the Arctomiaceae is rather simple. The family is strongly supported and includes three genera: *Gregorella* and *Wawea,* each with one species, and *Arctomia* with five species (Henssen 1969; Henssen and Kantvilas 1985; Jørgensen 2003, 2007; Lumbsch et al. 2005; Øvstedal and Gremmen 2001, 2006). They are lichenized with the cyanobacteria genus *Nostoc*, have a corticate thallus, gymnocarpous ascomata, asci with a non-amyloid thallus, and 1-10-septate, hyaline ascospores.

We here report the discovery of a further new species, which we assign to the genus *Arctomia*, found epiphytic in montane habitats in the island of Reunion (Mascarene archipelago, Indian Ocean). The material was first assigned to *Leptogium*, a genus belonging to the Collemataceae in the Lecanoromycetidae (Lumbsch and Huhndorf 2010). It is an unusual species as it has a foliose, sometimes very much crumpled, thallus, producing corticate and easily detached « goniocysts », best developed at the lobes margins, disrupting when mature and then forming a soredioid margin. Three loci were amplified (nuLSU, mtSSU, *RPB1*) and inferences from the sequences produced from two collections left no doubt that the material belongs to the Arctomiaceae, and statistical support to include it in the genus *Arctomia* was found. A new species is thus described in this genus.

Methods

Well-preserved lichen specimens lacking any visible symptoms of fungal infection were used for DNA isolation. Extraction of DNA and PCR amplification were performed following the protocol of Cubero et al. (1999). The primers used were: for nuLSU, LR0R, LR3R, LR3, LR5R and LR6 (following the suggestions available on www.lutzonilab.net/primers), for mtSSU, mtSSU1 and mtSSU3R (Zoller et al. 1999), for *RPB1*, AFasc and 6R1asc (following the suggestions available on www.lutzonilab.net/primers). Amplicons were sequenced by Macrogen®. Sequence fragments were assembled with Sequencher version 4.9 (Gene Codes Corporation, Ann Arbor, Michigan). Sequences were subjected to megaBLAST searches (Wheeler et al. 2006) to detect potential contaminations.

We assembled matrices with most representatives of species included by Lumbsch et al. (2005) in their description of the new genus *Gregorella*, resolved within the strongly supported Arctomiaceae; we further added several other species belonging to the Ostropomycetidae included in the study of the gyalectoid representatives of *Pertusaria* s.l. by Schmitt et al. (2010), assigned to the new genus *Gyalectaria*. All ac-

cessions available on GenBank of representatives of the Arctomiaceae were included; they represent all species assigned to that family, except for both species of *Arctomia* described from subantarctic islands by Øvstedal and Gremmen (2001, 2006). The outgroup species (*Bacidia rosella, Lecanora intumescens* and *Toninia cinereovirens*) were chosen outside the Ostropomycetidae and within the Lecanorales (Miadlikowska et al. 2006) to avoid any putative homoplasy problem. Six new sequences were generated for this study, all belonging to the new species described in this paper (Table 1). The sequences were first aligned using MAFFT (on-line version available at http://mafft.cbrc.jp/alignment/server/) and eventually manually adjusted using MacClade v. 4.05 (Maddison and Maddison 2002). Ambiguous characters have been detected by eye and excluded from the analyses.

Three matrices were assembled: 38 species with 927 included characters for nuLSU, 38 species with 668 included characters for mtSSU and 32 species with 675 included characters for *RPB1* (part 1). Incongruence between the matrices was tested with maximum likelihood analysis using GARLI (Zwickl 2006, version 0.951 for OS X) with gaps treated as missing data, and a single most likely tree was produced. Support for the branches was estimated using bootstrap values from 1000 pseudoreplicates (all other parameters identical to the original ML search). A conflict was considered significant if a clade was supported with bootstrap support > 75% in a one-locus analysis and not in the other two. A further test for conflict was performed with LSU and *RPB1* concatenated in a single matrix versus mtSSU in another. No conflict was detected and therefore the available sequences for the three loci were concatenated. The assembled matrix is deposited in TreeBASE under the accession number 12710.

An unweighted maximum parsimony (MP) analysis was performed in PAUP* 4.0b10 (Swofford 2002). All characters were equally weighted and gaps were treated as missing data. A first heuristic analysis was performed using NNI (Nearest Neighbor Interchange) branch-swapping, with 1000 replicates and saving 10 trees at each step, the functions Steepest descent and MulTrees being in effect. A second analysis was performed with the 10,000 saved trees using TBR (Tree Branch Swapping), with a maximum of 200 trees saved at each step, the function Steepest descent being inactivated. A 50% consensus tree is produced, and the strength of support for individual branches was estimated using bootstrap values (MPBS) obtained from 1000 heuristic bootstrap pseudoreplicates.

A partition of six subsets was implemented in the concatenated matrix: nuLSU, mtSSU, intron in *RPB1*, and three for each *RPB1* codon position. Models of evolution for the maximum likelihood and Bayesian analysis were selected based on the Akaike Information Criterion (Posada and Buckley 2004) as implemented in Mr. Modeltest v2.3 (Nylander 2004). The selected model corresponds to the GTR model of nucleotide substitution (Rodríguez et al. 1990) including a proportion of invariable sites and a discrete gamma distribution of six rates categories. The maximum likelihood analysis was performed using RAxML-HPC2 (Stamatakis 2006) on the Cipres Gateway (Miller et al. 2010), with 1000 bootstrap pseudoreplicates. Bayesian analyses were carried out using the Metropolis-coupled Markov chain Monte Carlo method (MC^3)

Table 1. Species and specimens used for this study, with GenBank accessions numbers for the three loci examined. Newly produced sequences for *Arctomia borbonica* are in bold.

Species name	LSU	mtSSU	*1RPB*
Absconditella sp.	AY300825	AY300873	—
Acarosporina microspora	AY584643	AY584612	DQ782818
Agyrium rufum	EF581826	EF581823	EF581822
Arctomia borbonica 1 (holotype)	**JX030030**	**JX030032**	**JX030034**
Arctomia borbonica 2	**JX030031**	**JX030033**	**JX030035**
Arctomia delicatula	AY853355	AY853307	DQ870929
Arctomia interfixa	DQ007345	DQ007348	—
Arctomia teretiuscula	DQ007346	DQ007349	DQ870930
Aspicilia contorta	DQ986782	DQ986876	DQ986852
Bacidia rosella	AY300829	AY300877	AY756412
Chromatochlamys muscorum	AY607731	AY607743	FJ941910
Coccotrema pocillarium	AF274093	AF329166	DQ870940
Conotrema populorum	AY300833	AY300882	—
Diploschistes ocellatus	HQ659183	HQ659172	DQ366252
Gregorella humida	AY853378	AY853329	—
Gyalectaria diluta	GU980982	GU980974	—
Icmadophila ericetorum	DQ883694	DQ986897	DQ883723
Lecanora intumescens	AY300841	AY300892	AY756386
Neobelonia sp.	AY300830	AY300879	—
Ochrolechia parella	AF274097	AF329173	DQ870959
Ochrolechia upsaliensis	GU980986	GU980979	GU981009
Orceolina kerguelensis	AY212830	AY212853	DQ870963
Pertusaria amara	AF274101	AY300900	DQ973048
Pertusaria lactea	AF381557	AF381564	DQ870971
Pertusaria leioplaca	AY300852	AY300903	DQ870973
Pertusaria paramerae	DQ780326	DQ780293	GU981012
Pertusaria pertusa	AF279300	AF381565	DQ870978
Pertusaria pustulata	DQ780332	DQ780297	GU981013
Pertusaria subventosa	AY300854	DQ780302	DQ870981
Placopsis gelida	AY212836	AY212859	DQ870984
Protothelenella corrosa	AY607734	AY607746	DQ870988
Protothelenella sphinctrinoidella	AY607735	AY607747	DQ870989
Thamnolia vermicularis	AY853395	AY853345	DQ915599
Thelotrema subtile	DQ871013	DQ871020	DQ870998
Toninia cinereovirens	AY756365	AY567724	AY756429
Trapelia chiodectonoides	AY212847	AY212873	DQ870999
Trapeliopsis granulosa	AF274119	AF381567	DQ871001
Wawea fruticulosa	DQ007347	DQ871023	DQ871005

in MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003, Altekar et al. 2004). No priors values were assumed and gaps were treated as missing data. Four parallel runs were performed, each using four independent chains (three heated and one cold chain),

with a single tree saved every 100th generation for a total of 6,000,000 generations. The incremental heating scheme was set by default. We used TRACER v1.4.1 (Rambaut and Drummond 2007) to plot the log-likelihood values of the sample points against generation time, and determine when stationarity was achieved. Consequently the first 6,000 sampled trees were deleted as the burn-in of the chain. A majority rule consensus tree with average branch lengths was constructed for the remaining trees using the sumt option of MrBayes. Phylogenetic trees were visualized using FigTree v1.3.1 (Rambaut 2009). Branches support was considered as significant when Maximum Parsimony Bootstrap (MPBS) > 70%, Maximum Likelihood Bootstrap (MLBS) > 70% and Posterior Probabilities (PP) > 0.95.

We tested the monophyly of the genus *Arctomia* by comparing the best unconstrained tree with the best tree obtained by constraining all *Arctomia* sequences to form a monophyletic group. Trees were generated in RaxML and then tested with two methods: the Shimodaira-Hasegawa (SH) test and the Expected Likelihood Weight (ELW) test as implemented in Tree-PUZZLE 5.2. (Shimodaira and Hasegawa 2001, Strimmer and Rambaut 2002, Schmidt et al. 2002).

Results

The concatenated matrix with aligned sequences for nuLSU, mtSSU and *RPB1* has 2781 characters, out of which 511 are excluded (330 for nuLSU out of which 250 represent introns in *Bacidia rosella*, 173 for mtSSU and 8 for *RPB1*), 983 are constant, 276 are parsimony-uninformative and 1011 are parsimony potentially informative. The most parsimonious tree has the following characteristics: length = 6295 steps, CI = 0.336 and RI = 0.428. The ML analysis yielded a tree with a likelihood value of Ln = -28660.4 and length of 6.175. Parameters of the partitions were as follows: LSU — p(A)= 0.2604, p(C)= 0.2216, p(G)= 0.2980, p(T)= 0.2199 a= 0.3134, r(A-C)= 0.7438, r(A-G)= 1.8229, r(A-T)= 0.7430, r(C-G)= 0.7409, r(C-T)= 4.5270, r(G-T)= 1.0000; mtSSU — p(A)= 0.3330, p(C)= 0.1606, p(G)= 0.2136, p(T)= 0.2926, a= 0.4207, r(A-C)= 0.9284, r(A-G)= 2.9298, r(A-T)= 1.6160, r(C-G)= 0.6649, r(C-T)= 3.4571, r(G-T)= 1.0000; *RPB1* intron — p(A)= 0.2349, p(C)= 0.2056, p(G)= 0.2267, p(T)= 0,3287, a= 0.9412, r(A-C)= 6.9358, r(A-G)= 21.9085, r(A-T)= 11.1853, r(C-G)= 8.6280, r(C-T)= 19.3378, r(G-T)= 1.0000; *RPB1*, 1st codon — p(A)= 0.2778, p(C)= 0.2440, p(G)= 0.3318, p(T)= 0.1463, a= 0.4211; r(A-C)= 4.0125, r(A-G)= 5.8268, r(A-T)= 3.1946, r(C-G)= 2.7176, r(C-T)= 2907386, r(G-T)= 1.0000; *RPB1*, 2nd codon — p(A)= 0.3521, p(C)= 0.2038, p(G)= 0.2319, p(T)= 0.2122, a= 0.3474, r(A-C)= 1.7253, r(A-G)= 3.1209, r(A-T)= 0.5159, r(C-G)= 1.9509, r(C-T)= 4.4498, r(G-T)= 1.0000; *RPB1*, 3rd codon — p(A)= 0.2683, p(C)= 0.2056, p(G)= 0.2545, p(T)= 0.2716, a= 0.5667, r(A-C)= 8.7546, r(A-G)= 24.9090, r(A-T)= 4.6296, r(C-G)= 5.8128, r(C-T)= 56.3087, r(G-T)= 1.0000.

All three analyses retrieve the family Arctomiaceae as a strongly supported clade (MPBS= 81%, MLBS = 97%, PP=1) (Fig. 1). All nodes within the Arctomiaceae clade are strongly supported: *A. delicatula* and *A. teretiuscula* form a clade supported

with MLBS= 99% and PP=1.0; they further form a clade with both accessions of *A. borbonica* that is supported with MLBS = 94% and PP=1.0; *Gregorella humida* and *Wawea fruticulosa* form a clade supported with MLBS = 86% and PP= 1.0; and finally the latter is sister to the clade including all accessions of *Arctomia* (except for *A. interfixa*) in a node supported by MLBS= 95% and PP= 1.0.

SH test shows that the likelihood of the topology constraining all *Arctomia* sequences to form a monophyletic group is not significantly worse (at 0.05 significance level) than that with *Arctomia interfixa* being sister to all other accessions of the Arctomiaceae. Following that test, the monophyly of all species assigned to *Arctomia*, incl. *A. borbonica* sp. nov., cannot be rejected. The result of the ELW is the contrary: such a monophyly is rejected at 0.0473 significance level.

Discussion

The lichen family Arctomiaceae is fully recovered in our analysis (Fig. 1) and all other accessions are resolved in positions fully consistent with those published for the Ostropomycetidae (Lumbsch et al. 2005, Baloch et al. 2010, Schmitt et al. 2010), including the polyphyly of representatives of *Pertusaria* that are resolved in three distinct lineages, and the representative of the newly described genus *Gyalectaria* that is resolved as sister to the representative of *Coccotrema*. Our material is resolved without ambiguity within the Arctomiaceae. It is resolved with strong support as sister to a clade comprising the type species of *Arctomia* (*A. delicatula*). The monophyly of the three species of *Arctomia* for which DNA sequences are available, demonstrated with strong support in Lumbsch et al. (2005), is not recovered in our analysis but is not rejected by the topology tests. The assignment of our new species to the genus *Arctomia* can thus be considered legitimate. The apparent dismemberment of *Arctomia* in our analysis (with *A. interfixa* as sister to all other taxa of the Arctomiaceae) may be due to an incomplete dataset (sequences for the three loci are available for all accessions of Arctomiaceae, except for *A. interfixa* which lacks the most informative *RPB1* sequence): indeed, incomplete dataset may produce misleading results in likelihood-based analysis (Simmons 2011). However, separate analyses of LSU and mtSSU sequences yielded the same topology, with *Arctomia* paraphyletic. The status of *Arctomia interfixa* should thus be studied in more details.

Diagnostic characters for the genera recognized within the Arctomiaceae are given by Lumbsch et al. (2005). In the absence of ascomata and conidiomata, they are: thallus crustose, composed of goniocysts for *Gregorella*, fruticose for *Wawea* and crustose to coralloid or squamulose for *Arctomia*. The other two species of *Arctomia*, described by Øvstedal and Gremmen (2001, 2006) and not included in Lumbsch et al. (2005) have a thallus "placodioid" or "foliose, […] squamulose or elongate, forming rosettes". If assigned to *Arctomia*, our new species does not match the thallus description of that genus, as its thallus is foliose and produces typical goniocysts at its margin, disintegrating into a soredioid margin (Fig. 2). We suggest the thallus of *Arctomia borbonica* is much similar to that of *Wawea fruticulosa* which has a "fruticose, olive-grey to brown" thallus (Henssen and Kantvilas 1985)

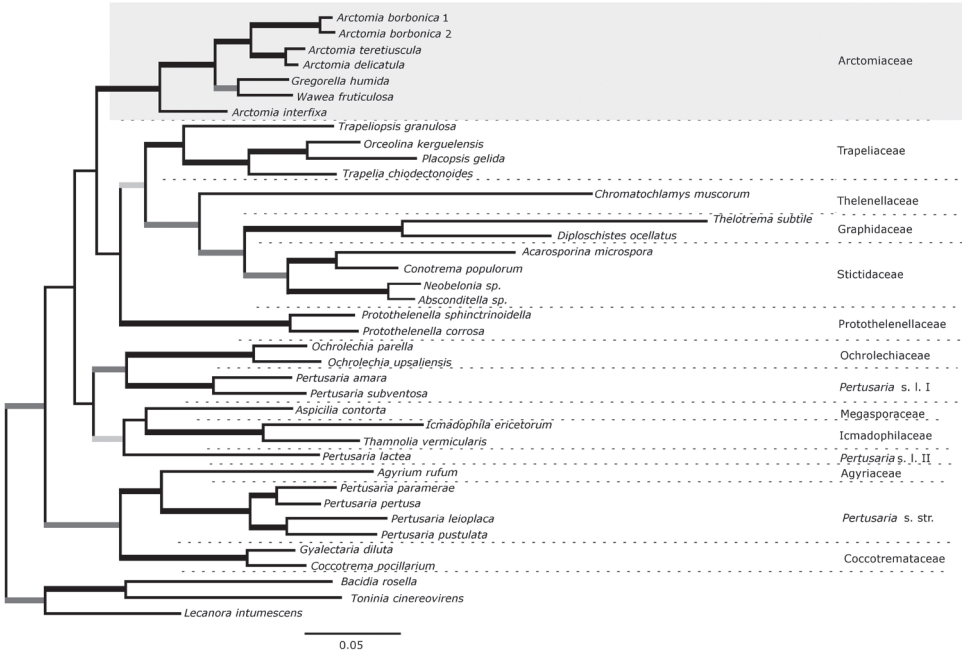


Figure 1. 50% consensus tree produced by the Bayesian analysis of a concatenated matrix with three loci (nuLSU, mtSSU and *RPB1*) with 2531 characters and highlighting the Arctomiaceae and the newly described *Arctomia borbonica*. Branches supported by MPBS and MLBS > 70% and Bayesian posterior probabilities > 0.95 are in black; those supported by MLBS >70% and Bayesian posterior probabilities > 0.95 in dark grey and those only by Bayesian posterior probabilities > 0.95 in light grey.

but with lobes flattened or at least furrowed (see fig. 2 in Henssen and Kantvilas 1985, Kantvilas and Jarman 1999). Further, the structure of the cortex is quite similar in *Wawea* (cross section and surface view: see fig. 3A–B in Henssen and Kantvilas 1985) when compared with *A. borbonica* (Fig. 2C–E). Finally, it is interesting to note that the sister species of *Wawea* is *Gregorella humida* whose thallus is entirely made of goniocysts, very similar to those produced by *Arctomia borbonica* at its thallus margin. As long as ascomata and conidiomata are not found and could provide more information, the thallus characters of *Arctomia borbonica* confuse the generic delimitations within the family.

The hypothesis of describing a new genus for *Arctomia borbonica* has been carefully assessed. Indeed, the genus as circumscribed by Henssen (1969) and Jørgensen (2007) is well-delimited and the inclusion of *A. borbonica* makes it morphologically heterogeneous. We refrained from describing a new genus because of the following points: (a) both subantarctic species recently described by Øvstedal and Gremmen (2001, 2006) in the genus, both assumed not to genuinely belonging to *Arctomia* s. str. and with generic affinities "under study", should be further studied; indeed, several characters put them aside of the genus such as a pluricellular cortex; the description of a new genus within such a small family as the Arctomiaceae is premature

in that context; (b) ascomata and conidiomata are unknown, or not yet discovered, in *A. borbonica* and thus our dataset lacks important characters (Lumbsch et al. 2005, Table 2); (c) morphological and anatomical characters may be very much misleading for phylogenetic reconstruction and sound generic delimitations as demonstrated by many studies in lichenized or unlichenized ascomycetes (Gaya et al. 2008, Lantz et al. 2011, Prieto et al. 2012, Sérusiaux et al. 2010); and (d) two statistical topology tests applied to the likelihood tree gave opposite results to assess the monophyly of *Arctomia* when including all species studied, e.g. *A. borbonica, A. delicatula, A. interfixa* and *A. teretiuscula*.

Taxonomy

***Arctomia borbonica* Magain & Sérus, sp. nov.**
Mycobank: MB 800279
Fig. 2

Diagnosis. Species recognized by its foliose, usually much crumpled, blue grey to brown thallus producing goniocysts at its margins, eventually forming a soredioid margin. Ascomata and conidiomata unknown.

Type. REUNION (Mascarene archipelago). Forêt de Bébour, track starting at Gîte de Bélouve toward Piton des Neiges, 21°4'49"S, 55°31'24"E (DMS), 1850 m alt., 9 Nov 2009, wet montane ericoid tickets, N. Magain & E. Sérusiaux sn (holotype : LG).

Description. Thallus not exceeding 1 cm in diam., with distinct lobes when well-developed, lobes blue-grey to brown when dry, up to 0.2-0.3 mm wide and c. 200-400 µm thick, hardly distinguished in some specimens, with a surface typically wrinkled (even in young lobes), sometimes very much "crumpled", always developing small goniocysts, mainly at the margins but also on the upper surface; cortex (Fig. 2C–E) developed on upper and lower sides, formed by a single layer of small rounded (in cross section) and jigsaw-like (in surface view) cells, less than 5 µm thick; goniocysts (Fig. 2F) 20-80 µm across, always containing compact chains of *Nostoc* cells and covered by a layer of isodiametric to rounded cells, 2–5 µm, best developed at the lobes margins where they eventually form a typical pale brownish soredioid edge, due to cortical disintegration. Photobiont belonging to the cyanobacteria genus *Nostoc* forming chains of small rounded cells 2–5 µm in diam. Ascomata and conidiomata unknown.

Chemistry. No secondary metabolites found by TLC.

Notes. The material looks like a species in *Leptogium*, a genus belonging to the Collemataceae in the Lecanoromycetidae (Lumbsch and Huhndorf 2010). Soredia or soredioid propagules are however unknown in that genus as well as in the closely related *Collema*. *Arctomia borbonica* is easily recognized by its foliose, sometimes very much crumpled, blue grey to brown thallus, producing corticate and easily detached « goniocysts », best developed at the lobes margins, disrupting when mature and then forming a soredioid margin.

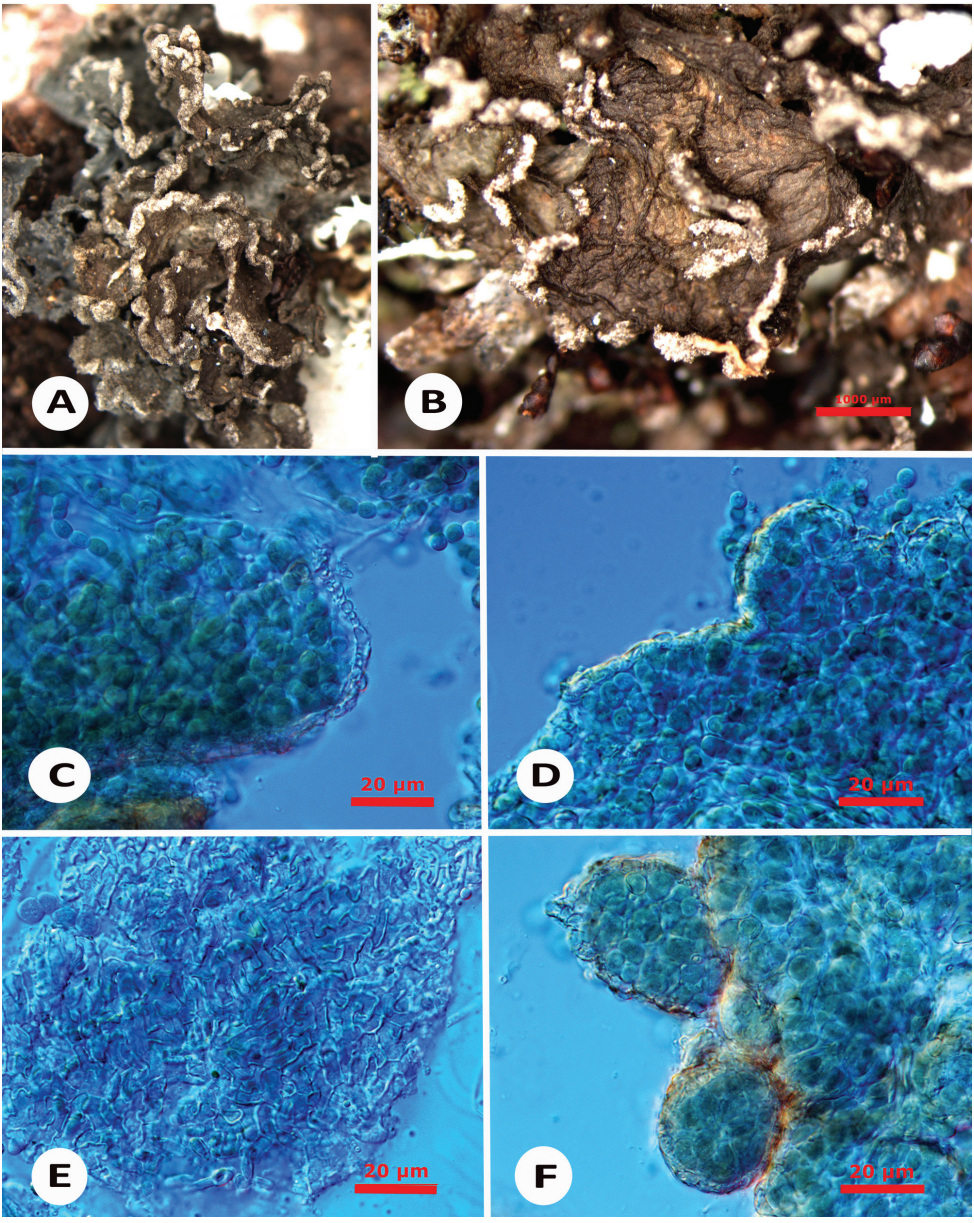


Figure 2. *Arctomia borbonica* (holotype). **A–B** macroscopic view of the thallus, with details of the wrinkled surface **B** and soredioid margin, made of disintegrating goniocysts **C–D** cross section through the thallus, showing the cortex with small, isodiametric cells, and the *Nostoc* chains **E** surface view of the cortex **F** young goniocysts formed at the lobes margins. Scale: **A–B** = 1 mm; **C–E** = 20 µm.

Distribution and ecology. *Arctomia borbonica* has been collected at three different sites on the island of Reunion in the Mascarene archipepago, incl. in highly disturbed secondary tickets with *Eucalyptus* plantations; it grows on trunks (*Eucalyptus*, *Acacia*

heterophylla) or on main stems of *Erica* tickets. It is probably widespread on the island. The two localities with natural vegetation belong to two different and typical habitats. The first one is the margin of the "Forêt de tamarins des hauts" with large boles of the endemic tree *Acacia heterophylla* (locality at the nature reserve "Roche Ecrite", at 1500 m) and corresponds to the "*Acacia* mountain forest" in Strasberg et al. (2005). The other one is the wet upper montane ericoid tickets (type locality; locality in the Bébour forest at 1800–1850 m) and corresponds to the "*Philippia* mountain ticket" in Strasberg et al. (2005). Here the vegetation does not exceed 4–5 m in height and is formed by *Erica arborescens, E. montana, Eugenia buxifolia, Agauria buxifolia, Cordyline mauritiana* (locally very abundant), *Cyathea* sp., *Phylica nitida, Astelia hemichrysa, Blechnum attenuatum*; ground is covered by very thick (up to 80 cm) layer of *Sphagnum* and other bryophytes. It is one of the most rewarding habitat for lichens on Reunion, with many interesting species, including representatives of the austral element (van den Boom et al. 2011), such as *Gomphillus morchelloides, G. pedersenii* and *Sporopodiopsis mortimeriana*.

Other specimens examined. REUNION (Masarenes archipelago). Nature reserve at Roche Ecrite, track to the summit, 20°58'6"S, 55°26'26"E (DMS), c. 1500 m alt., 4 nov 2009, montane forest dominated by *Acacia heterophylla,* N. Magain & E. Sérusiaux sn (LG). S part of the island, N of St-Philippe, near « gîte Bernard Brice », 21°20'23"S, 55°41'55"E (DMS), 650 m alt., 10 Nov 2009, *Eucalyptus* plantations and secondary tickets, N. Magain & E. Sérusiaux sn (LG).

Acknowledgements

Field studies in Reunion were made possible with the help and advice from the Parc National de La Réunion, especially through the courtesy of Mr B. Lequette and Mr J. M. Pausé. Dr Cl. Ah-Peng and Prof. D. Strasberg of the University of La Réunion in Saint-Denis and Dr. J. Hivert of the Conservatoire Botanique National de Mascarin (St-Leu) were also very helpful. We thank them all most sincerely. We further thank Mr I. Cremasco and L. Gohy for technical assistance in the molecular laboratory and herbarium at the University of Liège. Nicolas Magain is a Ph.D. Student at the University of Liège and acknowledges the financial support by FRIA, an organ of the Belgian Research Foundation. Finally we thank both referees for their critical and helpful notes and suggestions.

References

Altekar G, Aittokallio TJP, Huelsenbeck JP, Ronquist F (2004) Parallel Metropolis-coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. Bioinformatics 20: 407–415. doi: 10.1093/bioinformatics/btg427

Baloch E, Lücking R, Lumbsch HT, Wedin M (2010) Major clades and phylogenetic relationships between lichenized and non-lichenized lineages in Ostropales (Ascomycota: Lecanoromycetes). Taxon 59: 1483–1494.

Cubero OF, Crespo A, Fathi J, Bridge PD (1999) DNA extraction and PCR amplification method suitable for fresh, herbarium-stored, lichenized, and other fungi. Plant Systematics and Evolution 216: 243–249. doi: 10.1007/BF01084401

Henssen A (1969) Eine Studie über die Gattung *Arctomia*. Svensk Botanisk Tidskrift 63: 126–138.

Henssen A, Kantvilas G (1985) *Wawea fruticulosa*, a new genus and species from the Southern Hemisphere. Lichenologist 17: 85–97. doi: 10.1017/S0024282985000093

Gaya E, Navarro-Rosinés P, Llimona X, Hladun N, Lutzoni F (2008) Phylogenetic reassessment of the Teloschistaceae (lichen-forming Ascomycota, Lecanoromycetes). Mycological Research 112: 528–546. doi: 10.1016/j.mycres.2007.11.005

Jørgensen PM (2003) A new species of *Arctomia* from Sichuan Province, China. Lichenologist 35: 287–289. doi: 10.1016/S0024-2829(03)00053-7

Jørgensen PM (2007) Arctomiaceae. Nordic Lichen Flora 3: 9–11.

Kantvilas G, Jarman SJ (1999) Lichens of rainforest in Tasmania and south-eastern Australia. Flora of Australia Supplementary Series 9, 212 pp.

Lantz H, Johnston PR, Park D, Minter DW (2011) Molecular phylogeny reveals a core clade of Rhytismales. Mycologia 103: 57–74. doi: 10.3852/10-060

Lumbsch HT, del Prado R, Kantvilas G (2005) *Gregorella*, a new genus to accommodate *Moelleropsis humida* and a molecular phylogeny of Arctomiaceae. Lichenologist 37: 291–302. doi: 10.1017/S002428290501532X

Lumbsch HT, Huhndorf SM (2010) Myconet Volume 14. Part One. Outline of Ascomycota--2009. Part Two. Notes on Ascomycete Systematics. Nos. 4751–5113. Fieldiana Life and Earth Sciences 1: 1–64. doi: 10.3158/1557.1

Maddison DR, Maddison WP (2002) MacClade version 4.03PPC: analysis of phylogeny and character evolution. Sunderland, MA, Sinauer Associates.

Miadlikowska J, Kauff F, Hofstetter V, Fraker E, Grube M, Hafellner J, Reeb V, Hodkinson BP, Kukwa M, Lücking R, Hestmark G, Otalora MG, Rauhut A, Büdel B, Scheidegger C, Timdal E, Stenroos S, Brodo IM, Perlmutter GB, Ertz D, Diederich P, Lendemer JC, May PF, Schoch C, Arnold AE, Gueidan C, Tripp E, Yahr R, Robertson C, Lutzoni F (2006) New insights into classification and evolution of the Lecanoromycetes (Pezizomycotina, Ascomycota) from phylogenetic analyses of three ribosomal RNA- and two protein-coding genes. Mycologia 98: 1088–1103. doi: 10.3852/mycologia.98.6.1088

Miller MA, Pfeiffer W, Schwartz T (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. Proceedings of the Gateway Computing Environments Workshop (GCE), 14 Nov. 2010, New Orleans, LA, 1–8. doi: 10.1109/GCE.2010.5676129

Nylander JAA (2004) MrModeltest, version 2. Available from the author: http://www.ebc.uu.se/systzoo/staff/nylander.html

Øvstedal DO, Gremmen NJM (2001) The lichens of Marion and Prince Edward Islands. South African Journal of Botany 67: 552–572.

Øvstedal DO, Gremmen NJM (2006) Lichens of sub-Antarctic Heard Island. South African Journal of Botany 72: 353–366. doi: 10.1016/j.sajb.2005.09.008

Posada D, Buckley T (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Systematic Biology 53: 793–808. doi: 10.1080/10635150490522304

Prieto M, Martínez I, Aragón G, Gueidan C, Lutzoni F (2012) Molecular phylogeny of *Heteroplacidium, Placidium*, and related catapyrenoid genera (Verrucariacae, lichen-forming Ascomycota). American Journal of Botany 99: 23–35. doi: 10.3732/ajb.1100239

Rambaut A (2009) FigTree v1.3.1. Available http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Drummond AJ (2007) Tracer v1.5. Available at: http://beast.bio.ed.ac.uk/Tracer

Rodríguez F, Oliver JL, Marín A, Medina JR (1990) The general stochastic model of nucleotide substitution. Journal of Theoretical Biology 142: 485–501. doi: 10.1016/S0022-5193(05)80104-3

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19: 1572–1574. doi: 10.1093/bioinformatics/btg180

Rivas Plata E, Hernández M JE, Lücking R, Staiger B, Kalb K, Cáceres MES (2011) *Graphis* is two genera : A remarkable case of parallel evolution in lichenized Ascomycota. Taxon 60 : 99–107.

Rivas Plata E, Lücking R, Lumbsch HT (2012) A new classification for the family Graphidaceae (Ascomycota: Lecanoromycetes: Ostropales). Fungal Diversity 52: 107–121. doi: 10.1007/s13225-011-0135-8

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartet and parallel computing. Bioinformatics 18: 502–504. doi: 10.1093/bioinformatics/18.3.502

Schmitt I, Frankhauser JD, Sweeney K, Spribille T, Kalb K, Lumbsch HT (2010) Gyalectoid *Pertusaria* species form a sister-clade to *Coccotrema* (Ostropomycetidae, Ascomycota) and comprise the new lichen genus *Gyalectaria*. Mycology 1: 75–83.

Sérusiaux E, van den Boom P, Ertz D (2010) A two-gene phylogeny shows the lichen genus *Niebla* (Lecanorales) is endemic to the New World and does not occur in Macaronesia nor in the Mediterranean basin. Fungal Biology 114: 528–537. doi: 10.1016/j.funbio.2010.04.002

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246–1247. doi: 10.1093/bioinformatics/17.12.1246

Simmons MP (2011) Misleading results of likelihood-based phylogenetic analyses in the presence of missing data. Cladistics, published on-line. doi: 10.1111/j.1096-0031.2011.00375.x

Stamatakis A (2006) RAxML-VI-HPC: Maximum Likelihood-based Phylogenetic Analyses with Thousands of Taxa and Mixed Models. Bioinformatics 22: 2688–2690. doi: 10.1093/bioinformatics/btl446

Strasberg D, Rouget M, Richardson DM, Baret S, Dupont J, Cowling RM (2005) An Assessment of habitat diversity and transformation on La Réunion Island (Mascarene Islands, Indian Ocean) as a basis for identifying broad-scale conservation priorities. Biodiversity and Conservation 14: 3015–3032. doi: 10.1007/s10531-004-0258-2

Strimmer K, Rambaut A (2002) Inferring confidences sets of possibly misspecified gene trees. Proceedings of the Royal Society B, Biological Sciences 269: 137–142. doi: 10.1098/rspb.2001.1862

Swofford DL (2002) PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sunderland, MA, Sinauer Associates.

van den Boom PPG, Brand M, Ertz D, Kalb K, Magain N, Masson D, Schiefelbein U, Sipman HJM, Sérusiaux E (2011) Discovering the lichen diversity of a remote tropical island:

a working list of species collected on Reunion (Mascarene archipelago, Indian Ocean). Herzogia 24: 325–349.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Kapustin Y, Khovayko O, Landsman D, Lipman DJ, Madden TL, Maglott DR, Ostell J, Miller V, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Tatusov RL, Tatusova TA, Wagner L, Yaschenko E (2006) Database resources of the National Center for Biotechnology Information. Nucleic Acids Research, doi: 10.1093/nar/gkj158

Zoller S, Scheidegger C, Sperisen C (1999) PCR primers for the amplification of mitochondrial small subunit ribosomal DNA of lichen-forming ascomycetes. The Lichenologist 31: 511–516.

Zwickl DJ (2006) Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph.D. dissertation, The University of Texas at Austin.

MycoKeys 4: 23–36 (2012)
doi: 10.3897/mycokeys.4.3545
www.pensoft.net/journals/mycokeys

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

A new circumscription of the genus *Varicellaria* (Pertusariales, Ascomycota)

Imke Schmitt[1,2], Jürgen Otte[1], Sittiporn Parnmen[3], Anna D. Sadowska-Deś[1,2], Robert Lücking[3], H. Thorsten Lumbsch[3]

1 *Biodiversity and Climate Research Centre (BiK-F), Forschungsinstitut Senckenberg, Senckenberganlage 25, D-60325 Frankfurt, Germany* **2** *Department of Biological Sciences, Institute of Ecology, Evolution and Diversity, Goethe Universität, Max-von-Laue-Str. 13, D-60438 Frankfurt, Germany* **3** *Department of Botany, Field Museum of Natural History, 1400 S. Lake Shore Drive, Chicago, IL 60605, USA*

Corresponding author: *Imke Schmitt* (imke.schmitt@senckenberg.de)

Academic editor: *Cecile Gueidan* | Received 18 June 2012 | Accepted 26 July 2012 | Published 8 August 2012

Citation: Schmitt I, Otte J, Parnmen S, Sadowska-Deś AD, Lücking R, Lumbsch HT (2012) A new circumscription of the genus *Varicellaria* (Pertusariales, Ascomycota). MycoKeys 4: 23–36. doi: 10.3897/mycokeys.4.3545

Abstract

The lichen-forming genus *Pertusaria* under its current circumscription is polyphyletic and its phylogenetic affiliations are uncertain. Here we study the species of the genera *Pertusaria* and *Varicellaria* which contain lecanoric acid as major constituent, have disciform apothecia, strongly amyloid asci, non-amyloid hymenial gel, 1-2-spored asci, and 1- or 2-celled ascospores with thick, 1-layered walls. We infer phylogenetic relationships using maximum likelihood and Bayesian analyses based on four molecular loci (mtSSU, nuLSU rDNA, and the protein-coding, nuclear *RPB1* and *MCM7* genes). Our results show that the lecanoric acid-containing species form a well-supported, monophyletic group, which is only distantly related to *Pertusaria* s.str. The phylogenetic position of this clade is unclear, but placement in *Pertusaria* s.str. is rejected using alternative hypothesis testing. The circumscription of the genus *Varicellaria* is enlarged to also include species with non-septate ascospores. Seven species are accepted in the genus: *Varicellaria culbersonii* (Vězda) Schmitt & Lumbsch, **comb. nov.**, *V. hemisphaerica* (Flörke) Schmitt & Lumbsch, **comb. nov.**, *V. kasandjeffii* (Szatala) Schmitt & Lumbsch, **comb. nov.**, *V. lactea* (L.) Schmitt & Lumbsch, **comb. nov.**, *V. philippina* (Vain.) Schmitt & Lumbsch, **comb. nov.**, *V. rhodocarpa* (Körb.) Th. Fr., and *V. velata* (Turner) Schmitt & Lumbsch, **comb. nov.** A key to the species of *Varicellaria* is provided.

Key words

Agyriales, Ascomycota, lichen-forming fungi, molecular phylogeny, Ostropomycetidae, *Pertusaria,* Pertusariales, taxonomy

Introduction

Generic classifications in lichen-forming fungi have changed dramatically since the introduction of molecular data. Numerous genera have been shown to be polyphyletic or nested within larger genera (e.g., Amo de Paz et al. 2010a, b; Blanco et al. 2004a, b, 2005, 2006; Crespo et al. 2007, 2010; Crewe et al. 2006; Divakar et al. 2006; Ertz and Tehler 2011; Gueidan et al. 2009; Högnabba 2006; Muggia et al. 2010; Printzen 2010; Rivas Plata and Lumbsch 2011; Rivas Plata et al. 2012; Tehler and Wedin 2008; Wedin et al. 2005; Westberg et al. 2010). A further example of incongruence of current classification and phylogenetic relationships as inferred from DNA sequences is the heterogeneous genus *Pertusaria.* It is the largest genus within Pertusariales, with possibly over 1000 species (Archer and Elix 2011, Messuti and Archer 2009). However, it has been shown to be polyphyletic with species belonging even to different families within the order (Lumbsch and Schmitt 2001, 2002; Schmitt and Lumbsch 2004; Schmitt et al. 2006, 2010).

Schmitt and Lumbsch (2004) identified a combination of phenotypical characters to distinguish between three of the clades of *Pertusaria.* These characters include secondary metabolites, ascoma-morphology, amyloidity of ascus walls and hymenial gel, number of ascospores per ascus, and ascospore wall thickness and layers. Later, Schmitt et al. (2010) identified a fourth clade with gyalectoid ascomata and found it to be related to Coccotremataceae. The latter clade was distinguished as the genus *Gyalectaria* and was placed in Coccotremataceae. However, the two remaining major clades that are not closely related to *Pertusaria* s.str., the *Variolaria* and *Varicellaria* groups identified in Schmitt and Lumbsch (2004), have not yet been reclassified. In continuation of our studies on pertusarialean fungi, we are here addressing the issue of monophyly and classification of the so-called *Varicellaria* clade of *Pertusaria.* This is a group of pertusarialean lichenized fungi characterized by disciform apothecia, non-amyloid hymenial gel, strongly amyloid asci, 1-2-spored asci, and 1- or 2-celled ascospores with more or less thick, 1-layered walls (Schmitt and Lumbsch 2004). Chemically, the clade is characterized by the presence of lecanoric acid as major metabolite. Recent collections of *Pertusaria culbersonii,* a neotropical species with lecanoric acid, prompted us to address the phylogeny of this group and to classify those *Pertusaria* species belonging to the *Varicellaria* group. We have compiled a data set of 29 pertusarialean fungi including all but two species (*P. kasandjeffii* and *P. philippina* – no fresh material available) that were thought to belong to the *Varicellaria* group based on phenotypical evidence.

Materials and methods

Taxon sampling and molecular methods

We assembled a four-locus data set consisting of mtSSU rDNA, nuLSU rDNA, and the protein-coding genes *RPB1* and *MCM7*. The alignment contained 31 species. Specimens and sequences used for molecular analyses are listed in Table 1. Two sequences

Table 1. Species and sequences used in this study. New sequences are indicated in bold.

Name	Phylogenetic lineage	Family	nuLSU	mtSSU	*1RPB*	*7MCM*
*Varicellaria culbersonii**	Varicellaria	?	**JX101871**	**JX101873**	**JX101875**	**JX101874**
Varicellaria hemisphaerica	Varicellaria	?	AF381556	AF381563	DQ902341	GU980998
Varicellaria lactea	Varicellaria	?	AF381557	AF381564	DQ870971	GU981000
Varicellaria rhodocarpa	Varicellaria	?	AF381559	AF381569	N/A	N/A
Varicellaria velata	Varicellaria	?	AY300855	GU980981	DQ870982	GU981005
"Pertusaria" amara	Variolaria	?	AF274101	AY300900	DQ870965	GQ272423
"Pertusaria" corallina	Variolaria	?	AY300850	AY300901	DQ870967	GU980997
"Pertusaria" scaberula	Variolaria	?	AF274099	AF431959	DQ870980	GU981003
"Pertusaria" subventosa	Variolaria	?	AY300854	AY300905	DQ870981	GU981004
Circinaria contorta		Megasporaceae	DQ986782	DQ986876	DQ986852	GU980989
Circinaria hispida		Megasporaceae	DQ780305	HM060722	DQ870933	DQ780273
Lobothallia radiosa		Megasporaceae	DQ780306	DQ780274	DQ870954	GQ272397
Ochrolechia parella		Ochrolechiaceae	AF274097	GU980977	DQ870959	GQ272421
Ochrolechia subpallescens		Ochrolechiaceae	GU980985	GU980978	GU981008	GU980994
Ochrolechia upsaliensis		Ochrolechiaceae	GU980986	GU980979	GU981009	GU980995
Coccotrema cucurbitula		Coccotremataceae	AF274092	AF329161	DQ870939	GU980990
Coccotrema maritimum		Coccotremataceae	AF329164	AF329163	N/A	GU980991
Coccotrema pocillarium		Coccotremataceae	AF274093	AF329166	DQ870940	GU980992
Gyalectaria diluta		Coccotremataceae	GU980982	GU980974	N/A	N/A
Gyalectaria gyalectoides		Coccotremataceae	GU980983	GU980975	GU981006	GU980993
Gyalectaria jamesii		Coccotremataceae	GU980984	GU980976	GU981007	N/A
Thamnolia vermicularis		Icmadophilaceae	AY961599	AY853345	DQ915599	N/A
Icmadophila ericetorum		Icmadophilaceae	DQ883694	DQ986897	DQ883723	N/A
Dibaeis baeomyces		Icmadophilaceae	AF279385	AY300883	DQ842011	N/A
Agyrium rufum		Agyriaceae	EF581826	EF581823	EF581822	GU980988
*Miltidea ceroplasta***		Miltideaceae	HQ391558	HQ391557	**JQ900620**	N/A
*Pertusaria hermaka****	Pertusaria s. str.	Pertusariaceae	DQ780334	DQ780299	**JX101872**	GU980999
Pertusaria paramerae	Pertusaria s. str.	Pertusariaceae	DQ780328	GU980980	GU981012	GU981001
Pertusaria pustulata	Pertusaria s. str.	Pertusariaceae	DQ780332	DQ780297	GU981013	GU981002
Parmeliopsis hyperopta	outgroup	Parmeliaceae	AY607823	AY611167	EF092142	GQ272426
Everniopsis trulla	outgroup	Parmeliaceae	EF108290	EF108289	EF105429	GQ272396

*source: Costa Rica, *R. Lücking 15424* (F)

**source: Australia, *H.T. Lumbsch 20004b, S. Parnmen & T. Widhelm* (F)

***source: Australia, *A. Mangold,* 22 March 2005 (MIN)

of Parmeliaceae (Lecanoromycetes) were used as outgroup, since Lecanoromycetes was shown to be a sister-group of Ostropomycetidae to which Pertusariales belongs (Grube et al. 2004; Miadlikowska et al. 2006; Schmitt et al. 2009). Molecular methods were the same as in a previous study (Schmitt et al. 2010).

Sequence alignments and phylogenetic analysis

We assembled partial sequences using Geneious Pro 5.4.3 (Drummond et al. 2011) and edited conflicts manually. We aligned the sequences using Clustal W (Thompson et al. 1994) (nuLSU, *RPB1*, *MCM7*) or PRANK (Loytynoja and Goldman 2005, 2010) (mtSSU). MtSSU sequences are highly variable and contain substantial length polymorphisms that disrupt the alignment. Thus, we eliminated unreliably aligned sites from the mtSSU alignment using the program Aliscore 2.0 (Misof and Misof 2009). Aliscore settings were: window size of six positions, and gaps treated as ambiguous characters (-N option invoked). After cutting 1084 unreliably aligned positions, 698 positions (39%) of the original mtSSU alignment were left.

We analyzed the alignments using maximum likelihood (ML) and Bayesian inference. To test for potential conflict between data sets, we performed ML analyses on the individual alignments and examined the trees for conflicts supported by 75% bootstrap support. ModelTest (Posada and Crandall 1998) selected the following models as best fits for our data: GTR+G+I for nuLSU, *RPB1*, *MCM7*, and GTR+G for mtSSU. The individual alignments were analyzed in Geneious using MrBayes 3.1 (Huelsenbeck and Ronquist 2001) with the following settings: 1,100,000 generations starting with a random tree and employing 12 simultaneous chains. Two runs were executed, and every 1000th tree was saved into a file. The first 100 trees were discarded as burn in. We checked the traces in Geneious to ensure that stationarity was achieved after the first 100,000 generations. MrBayes settings for the concatenated alignment were the same as above but with 8,000,000 generations and the data split into 8 partitions (mtSSU, nuLSU, and each codon position of *RPB1* and *MCM7*). We used the model GTR+I+G and the burn in was set to 1000. Of the remaining trees, a majority rule consensus tree with average branch lengths was calculated. Posterior probabilities were obtained for each clade. Only clades with posterior probabilities equal or above 0.95 in the Bayesian analysis or bootstrap support equal or above 75 % under ML were considered as strongly supported.

The ML analysis of the concatenated alignment was performed with the program RAxML (Stamatakis 2006) using the default rapid hill-climbing algorithm. The model of nucleotide substitution chosen was GTRMIX. The data set was partitioned into eight parts (mtSSU, nLSU and each codon position of *RPB1* and *MCM7*). Rapid bootstrap estimates were carried out for 2000 pseudoreplicates. Phylogenetic trees were visualized using the program TreeView (Page 1996).

As in previous studies (e.g. Schmitt and Lumbsch 2004) the lecanoric acid-containing species of *Pertusaria* clustered outside *Pertusaria* s.str., and instead with the genus *Varicellaria*, hence contradicting current classification. Thus, we tested whether our data

are sufficient to reject monophyly of *Pertusaria* s.str. + lecanoric acid containing *Pertusaria* spp. For hypothesis testing, we used two different methods: i) Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 2001) and ii) expected likelihood weight (ELW) test (Strimmer and Rambaut 2002). The SH and ELW test were performed using Tree-PUZZLE 5.2 (Schmidt et al. 2002) with the combined data set, comparing the best tree agreeing with the null hypotheses, and the unconstrained ML tree. These trees were inferred in Tree-PUZZLE using the GTR+I+G nucleotide substitution model.

Results

We obtained six new sequences indicated in Table 1. The combined alignment of the nuLSU, mtSSU rDNA, *RPB1,* and *MCM7* included 2790 unambiguously aligned nucleotide position characters, 1226 of which were variable. The single locus ML topologies did not show any conflicts and hence a concatenated analysis was performed. The maximum likelihood tree did not contradict the Bayesian tree topologies and thus only the majority-rule consensus tree of the Bayesian tree sampling is shown here (Fig. 1). In the phylogenetic tree, species of the *Varicellaria*-group form a strongly supported monophyletic group, including *P. culbersonii.* The *Varicellaria*-group is sister to the *Variolaria*-group, but this relationship lacks support. The genus *Ochrolechia* is a well-supported sister-group to Megasporaceae (*Circinaria* and *Lobothallia*), and this clade is sister to the *Varicellaria*- and *Variolaria*-groups, but again this relationship lacks support. *Agyrium* and *Miltidea* form a supported sister-group, which is strongly supported sister to the well-supported, monophyletic *Pertusaria* s.str. The well-supported, monophyletic genera *Coccotrema* and *Gyalectaria* have a well-supported sister-group relationship. The sister-group relationship of Coccotremataceae and the clade including *Agyrium, Miltidea,* and *Pertusaria* s.str. lacks support. A placement of the *Varicellaria* clade in *Pertusaria* s.str. is rejected significantly ($p \leq 0.001$ in both tests) using alternative hypothesis testing.

Discussion

The current study confirms previous results on the polyphyly of *Pertusaria* (Lumbsch and Schmitt 2001, 2002; Lumbsch et al. 2006; Schmitt and Lumbsch 2004; Schmitt et al. 2006, 2010). It also confirms that species with lecanoric acid as major constituent and disciform apothecia are closely related to *Varicellaria rhodocarpa* and therefore should be included in the genus *Varicellaria*. Our taxon sampling included all but two species putatively belonging to the *Varicellaria*-group and hence we feel confident to draw formal nomenclatural consequences.

We will address the issue of the phylogeny and classification of the species-rich *Variolaria*-group in the future using an extended and geographically balanced taxon sampling. Our study shows that additional, molecular markers will be necessary to elu-

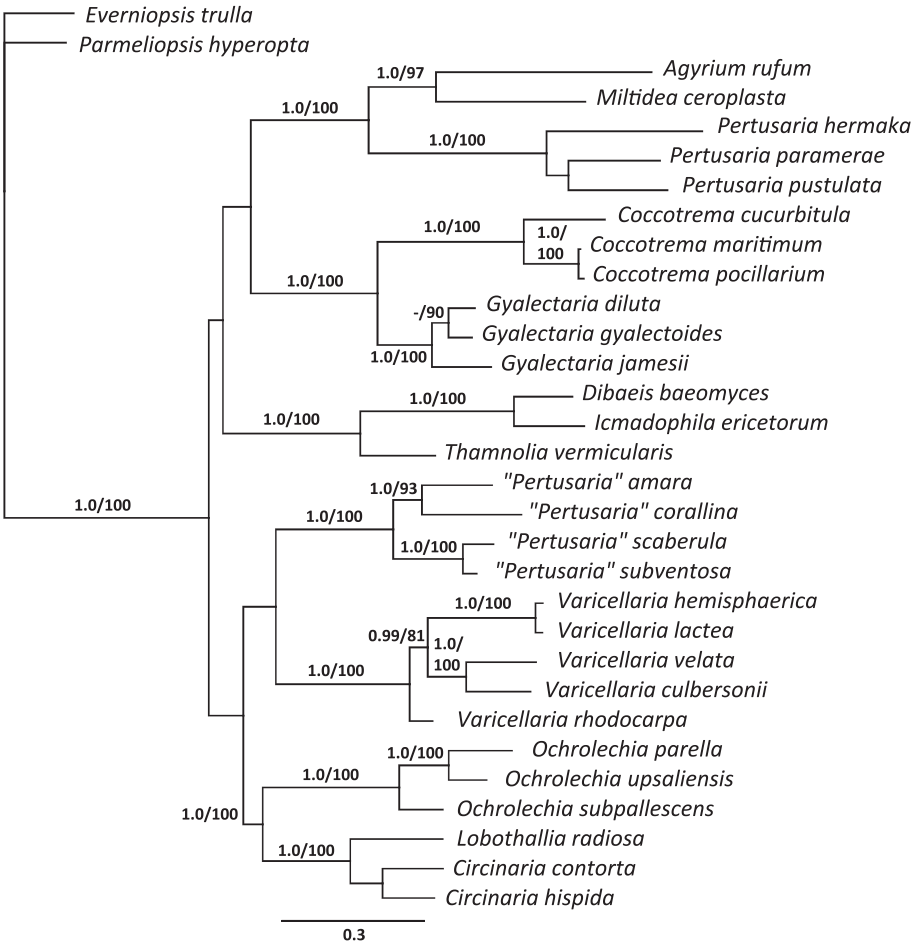


Figure 1. Phylogeny of pertusarialean fungi based on mtSSU, nuLSU, *RPB1* and *MCM7* sequences. This is a 50% majority rule consensus tree based on 14,000 trees from a Bayesian analysis. Values above the branches are posterior probabilities/ML bootstrap support (ML based on 2000 replicates).

cidate the phylogenetic relationships of major clades within Pertusariales (incl. Agyriales) (Hodkinson and Lendemer 2011), since the backbone of the phylogeny of the order almost entirely lacks support.

Taxonomic consequences and key to the species

Varicellaria **Nyl. Mém. Soc. Imp. Sci. Nat. Cherbourg 5: 119. 1858.**
http://species-id.net/wiki/Varicellaria

Type species. *Varicellaria microsticta* Nyl. Mém. Soc. Imp. Sci. Nat. Cherbourg 5: 119. 1858. [=*V. rhodocarpa* (Körb.) Th.Fr.]

=*Clausaria* Nyl. Annls Sci. Nat., Bot., sér. 4 15: 45. 1861.

Type species. *Clausaria fallens* Nyl., Ann. Sci. Nat., Bot., sér. 4 15: 45. 1861. [=*Varicellaria velata* (Turner) Schmitt & Lumbsch]

The genus in its enlarged circumscription includes species with disciform ascomata, non-amyloid hymenial gel, strongly amyloid, 1-2-spored asci, and 1- or 2-celled ascospores with thick, 1-layered walls. All species contain lecanoric acid, and may also contain lichexanthone or variolaric acid. Currently, we accept seven species in this genus. The accepted names and authorities are listed below.

***Varicellaria culbersonii* (Vězda) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800038
http://species-id.net/wiki/Varicellaria_culbersonii

Basionym. *Pertusaria culbersonii* Vězda. Lich. sel. exs. 60: 4 (no. 1487). 1977. Type. Costa Rica, San José, Cerro de la Muerte, 3330m alt., 1976, on soil, *W.L. Culberson 13195J* (holotype PRA-V).

***Varicellaria hemisphaerica* (Flörke) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800039
http://species-id.net/wiki/Varicellaria_hemisphaerica

Basionym. *Variolaria hemisphaerica* Flörke. Deutsche Lich. 2: 6. 1815. Type. Germany, Berlin [Flörke, Deutsche Lichenen exs. 29] (isotype BM).

Synonym. *Pertusaria hemisphaerica* (Flörke) Erichsen. Hedwigia 72: 85. 1932.

***Varicellaria kasandjeffii* (Szatala) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800040
http://species-id.net/wiki/Varicellaria_kasandjeffii

Basionym. *Pertusaria kasandjeffii* Szatala. Magy. Bot. Lapok 29: 83. 1930. Type. Bulgaria, Cepelarska planina, in monte Turluka, par Pamsakli, 1500m alt., 6.1929, *Szatala* (isotype HBG-1233).

This species is only known from a few localities in Bulgaria and Romania (Hanko 1983). Since no fresh material was available, we could not generate molecular data. However, the species agrees morphologically and chemically with the *Varicellaria*-group (Fig. 2) and in fact its distinction from *P. lactea* is not entirely clear. Both taxa contain lecanoric and variolaric acid, but *P. kasandjeffii* differs in being esorediate and having a thick, bulbate thallus. Additional collections are required to test whether *P. kasandjeffii* is indeed different from *P. lactea.*

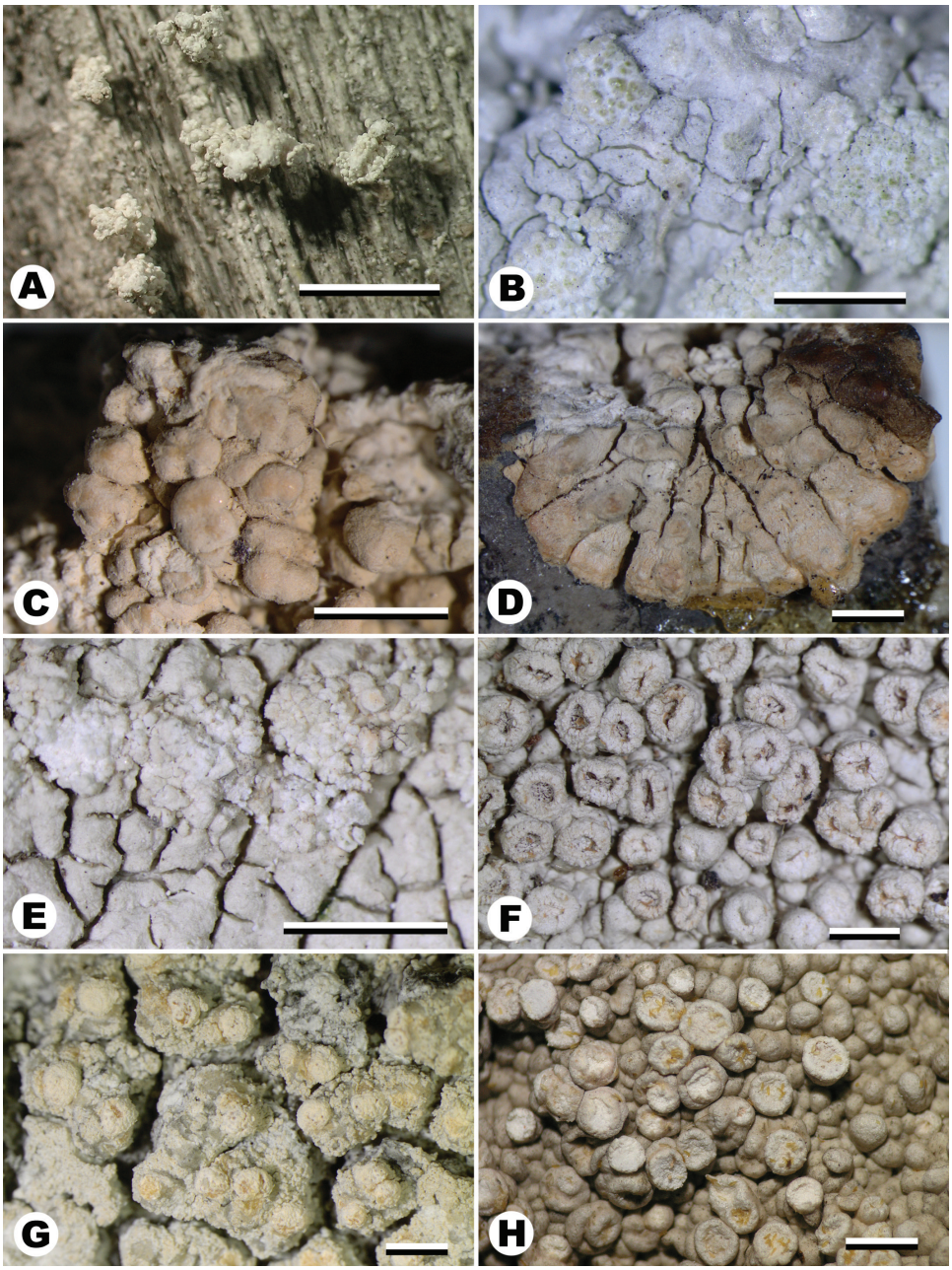


Figure 2. The species of *Varicellaria*. **A** *V. culbersonii.* Costa Rica, Buck 44182 (F) **B** *V. hemisphaerica*. Germany, 15.4.2004, Schmitt (FR) **C, D** *V. kasandjeffii*. Isotype. Bulgaria, Cepalarska planina: in monte Turluka, par Pasmakali, 1500 m, 9.6.1929, *Szatala* (HBG-1233) **E** *V. lactea*. Spain, Schmitt 5.6.2003 (FR) **F** *V. philippina*. Holotype. Philippines, Mindanao Dist. Lanao, Camp Keithley by lake Lanao, Sept. 1907, M.S. Clemens, (TUR-V-0006709) **G** *V. rhodocarpa*. Sweden, Printzen 6908 (FR) **H** *V. velata.* Colombia, Moncada & Davila 1537 (F). Scale bar: 1mm. Images were taken with an Olympus SC30 camera under an Olympus SZX7 stereomicroscope.

***Varicellaria lactea* (L.) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800041
http://species-id.net/wiki/Varicellaria_lactea

Basionym. *Lichen lacteus* L., Mant. Pl. 1: 132. 1767. Type. Sweden, Västergötland, Mularp, 6.08.1922, *Vrang* [=Malme, Lich. Suec. Exs. 848] (neotype UPS, designated by Jørgensen et al. (1994)).

Synonyms. *Lepra lactea* (L.) F.H.Wigg. Prim. fl. Holsat.: 97. 1780. *Variolaria lactea* (L.) Pers. Ann. Bot. 1: 24. 1794. *Psora lactea* (L.) P.Gaertn., G.Mey. & Scherb. Ökonom.-techn. Fl. Wetterau 3: 214. 1801. *Zeora lactea* (L.) Arnold. Flora, Jena 53: 214. 1870. *Pertusaria lactea* (L.) Arnold. Verh. zool.-bot. Ges. Wien 22: 283. 1872. *Ochrolechia lactea* (L.) Matzer & Hafellner. Bibl. Lichenol. 37: 101. 1990.

***Varicellaria philippina* (Vain.) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800589
http://species-id.net/wiki/Varicellaria_philippina

Basionym. *Pertusaria philippina* Vain. Philipp. J. Sci., C, Bot. 8: 131. 1913. Type. Philippines, Mindanao, Lanao, Castra Keithley at Lake Lanao, 1907, *Clemens 1302* (holotype TUR-V 6391!).

This species is only known from the Philippines (Wainio 1913) and Papua New Guinea (Elix et al. 1997). We could not generate molecular data since no fresh material was available. Morphologically and chemically the species agrees with *P. velata* (Fig. 2), but differs in having 2-spored asci.

***Varicellaria rhodocarpa* (Körb.) Th.Fr. Lich. Scand. (Uppsala) 1: 322. 1871.**
http://species-id.net/wiki/Varicellaria_rhodocarpa

Basionym. *Pertusaria rhodocarpa* Körb. Syst. lich. germ.: 384. 1855.

Synonyms. *Varicellaria microsticta* Nyl. Mém. Soc. Imp. Sci. Nat. Cherbourg 5: 119. 1858. *Varicellaria kemensis* Räsänen. Ann. Soc. zool.-bot. Fenn. Vanamo 3: 295. 1926.

***Varicellaria velata* (Turner) Schmitt & Lumbsch, comb. nov.**
Mycobank: MB 800042
http://species-id.net/wiki/Varicellaria_velata

Basionym. *Parmelia velata* Turner. Trans. Linn. Soc. London 9: 143. 1808. Type. Great Britain, England, Sussex, 1805, *Borrer* (holotype BM-4109).

Synonyms. *Lichen velatus* (Turner) Sm. & Sowerby. Engl. Bot. 29: tab. 2062. 1809. *Variolaria velata* (Turner) Ach. Lich. univ.: 696. 1810. *Pertusaria velata* (Turner) Nyl. Lich. Scand. (Uppsala): 179. 1861.

Pertusaria conglobata (Ach.) Th.Fr. Lichenogr. Scand. 1: 321. 1871. *Variolaria conglobata* Ach. Syn. Lich.: 132. 1814.

Pertusaria haematommoides Zahlbr., Feddes Rep. 33: 50. 1933. Type. Taiwan, Rengechi, *Asahina 263* (W – holotype!).

Pertusaria obvelata Nyl. Bih. K. svenska Vetensk. Akad. Handl. 3: 1–156. 1888.

Key to the species of Varicellaria

1a Ascospores 2-celled, thallus esorediate or rarely sorediate, containing lecanoric acid, growing on soil, detritus or mosses in arctic-alpine habitats of the northern Hemisphere ..... ***V. rhodocarpa***

1b Ascospores 1-celled, thallus esorediate or sorediate, chemistry and habitat various **2**

2a Thallus esorediate **3**

2b Thallus sorediate **6**

3a Thallus thin, coarsely wrinkled to rimose-cracked, containing lecanoric acid, ± lichexanthone, and ± variolaric acid **4**

3b Thallus thick, bullate, apothecia rare or unknown, when present 1-1.5 mm in diam., lacking lichexanthones, Neotropical or restricted to eastern Europe 5

4a Asci 1-spored, cosmopolitan ..... ***V. velata***

4b Asci 2-spored, so far only known from Philippines and Papua New Guinea ... . ***V. philippina***

5a Growing on siliceous rocks, known only from the Balkan region of Europe . ..... ***V. kasandjeffii***

5b Growing on soil, detritus or mosses, known from high altitudes in Central America ..... ***V. culbersonii***

6a Thallus containing lecanoric acid, on bark, rarely on rocks .... ***V. hemisphaerica***

6b Thallus containing lecanoric acid and variolaric acid, on rocks, rarely on bark ..... ***V. lactea***

Acknowledgements

We thank Armin Mangold (Berlin) and Todd Widhelm (Omaha) for collecting material used in this study, and Miklós Bálint (Frankfurt) for assisting with running phylogenetic analyses on the "FUCHS" computing cluster of the Center for Scientific Computing, Frankfurt am Main. Matthias Schultz (Hamburg), Seppo Huhtinen (Turku), and Anton Igersheim (Vienna) kindly sent type material on loan, and Klaus Kalb (Neumarkt) provided literature. This study was funded by 'LOEWE, Landes-Offensive zur Entwicklung Wissenschaftlich-ökonomischer Exzellenz' of Hesse's Ministry of Higher Education, Research, and the Arts. A.S-D. was supported by a stipend from Deutsche Bundesstiftung Umwelt (DBU).

References

Amo de Paz G, Lumbsch HT, Cubas P, Elix JA, Crespo A (2010a) The morphologically deviating genera *Omphalodiella* and *Placoparmelia* belong to *Xanthoparmelia* (Parmeliaceae). Bryologist 113: 376–386. doi: 10.1639/0007-2745-113.2.376

Amo de Paz G, Lumbsch HT, Cubas P, Elix JA, Crespo A (2010b) The genus *Karoowia* (Parmeliaceae, Ascomycota) includes unrelated clades nested within *Xanthoparmelia*. Australian Systematic Botany 23: 173–184. doi: 10.1071/SB09055

Archer AW, Elix JA (2011) A preliminary world-wide key to the lichen genus *Pertusaria*. http://www.rbgsyd.nsw.gov.au/science/Plant_Diversity_Research/Key_to_Pertusaria

Blanco O, Crespo A, Divakar PK, Elix JA, Lumbsch HT (2005) Molecular phylogeny of parmotremoid lichens (Ascomycota, Parmeliaceae). Mycologia 97: 150–159. doi: 10.3852/mycologia.97.1.150

Blanco O, Crespo A, Divakar PK, Esslinger TL, Hawksworth DL, Lumbsch HT (2004a) *Melanelixia* and *Melanohalea*, two new genera segregated from *Melanelia* (Parmeliaceae) based on molecular and morphological data. Mycological Research 108: 873–884. doi: 10.1017/S0953756204000723

Blanco O, Crespo A, Elix JA, Hawksworth DL, Lumbsch HT (2004b) A molecular phylogeny and a new classification of parmelioid lichens containing *Xanthoparmelia*-type lichenan (Ascomycota: Lecanorales). Taxon 53: 959–975. doi: 10.2307/4135563

Blanco O, Crespo A, Ree RH, Lumbsch HT (2006) Major clades of parmeliold lichens (Parmeliaceae, Ascomycota) and the evolution of their morphological and chemical diversity. Molecular Phylogenetics and Evolution 39: 52-69. doi: 10.1016/j.ympev.2005.12.015

Crespo A, Kauff F, Divakar PK, Amo G, Arguello A, Blanco O, Cubas P, del Prado R, Elix JA, Esslinger TL, Ferencova Z, Hawksworth DL, Lutzoni F, Millanes AM, Molina MC, Perez-Ortega S, Wedin M, Ahti T, Bungartz F, Calvelo S, Aptroot A, Barreno E, Candan M, Cole M, Ertz D, Goffinet B, Lindblom L, Lücking R, Mattsson JE, Messuti MI, Miadlikowska J, Piercey-Normore M, Rico V, Sipman HJM, Schmitt I, Spribille T, Thell A, Thor G, Lumbsch HT (2010) Phylogenetic generic classification of parmelioid lichens (Parmeliaceae, Ascomycota) based on molecular, morphological and chemical evidence. Taxon 59: 1735–1753.

Crespo A, Lumbsch HT, Mattsson JE, Blanco O, Divakar PK, Articus K, Wiklund E, Bawingan PA, Wedin M (2007) Testing morphology-based hypotheses of phylogenetic relationships in Parmeliaceae (Ascomycota) using three ribosomal markers and the nuclear *RPB1* gene. Molecular Phylogenetics and Evolution 44: 812–824. doi: 10.1016/j.ympev.2006.11.029

Crewe AT, Purvis WO, Wedin M (2006) Molecular phylogeny of Acarosporaceae (Ascomycota) with focus on the proposed genus *Polysporinopsis*. Mycological Research 110: 521–526. doi: 10.1016/j.mycres.2006.01.010

Divakar PK, Crespo A, Blanco O, Lumbsch HT (2006) Phylogenetic significance of morphological characters in the tropical *Hypotrachyna* clade of parmelioid lichens (Parmeliaceae, Ascomycota). Molecular Phylogenetics and Evolution 40: 448–458. doi: 10.1016/j.ympev.2006.03.024

Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, Markowitz S, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2011) Geneious v5.4. Available from http://www.geneious.com.

Elix JA, Aptroot A, Archer AW (1997) The lichen genus *Pertusaria* (lichenised Ascomycotina) in Papua New Guinea and Australia: twelve new species and thirteen new reports. Mycotaxon 64: 17–35.

Ertz D, Tehler A (2011) The phylogeny of Arthoniales (Pezizomycotina) inferred from nucLSU and *RPB2* sequences. Fungal Diversity 49: 47–71. doi: 10.1007/s13225-010-0080-y

Grube M, Baloch E, Lumbsch HT (2004) The phylogeny of Porinaceae (Ostropomycetidae) suggests a neotenic origin of perithecia in Lecanoromycetes. Mycological Research 108: 1111–1118. doi: 10.1017/S0953756204000826

Gueidan C, Savic S, Thus H, Roux C, Keller C, Tibell L, Prieto M, Heidmarsson S, Breuss O, Orange A, Fröberg L, Wynns AA, Navarro-Rosines P, Krzewicka B, Pykala J, Grube M, Lutzoni F (2009) Generic classification of the Verrucariaceae (Ascomycota) based on molecular and morphological evidence: recent progress and remaining challenges. Taxon 58: 184–208.

Hanko B (1983) Die Chemotypen der Flechtengattung *Pertusaria* in Europa. Bibliotheca Lichenologica 19: 1–297.

Hodkinson BP, Lendemer JC (2011) The orders of Ostropomycetidae (Lecanoromycetes, Ascomycota): recognition of Sarrameanales and Trapeliales with a request to retain Pertusariales over Agyriales. Phytologia 93: 407–412.

Högnabba F (2006) Molecular phylogeny of the genus *Stereocaulon* (Stereocaulaceae, lichenized ascomycetes). Mycological Research 110: 1080–1092. doi: 10.1016/j.mycres.2006.04.013

Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17: 754–755. doi: 10.1093/bioinformatics/17.8.754

Jørgensen PM, James PW, Jarvis CR (1994) Linnaean lichen names and their typification. Biological Journal of the Linnean Society 115: 261–405.

Loytynoja A, Goldman N (2005) An algorithm for progressive multiple alignment of sequences with insertions. Proceedings of the National Academy of Sciences of the United States of America 102: 10557–10562. doi: 10.1073/pnas.0409137102

Loytynoja A, Goldman N (2010) webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. BMC Bioinformatics 11: 579. doi: 10.1186/1471-2105-11-579

Lumbsch HT, Schmitt I (2001) Molecular data suggest that the lichen genus *Pertusaria* is not monophyletic. Lichenologist 33: 161–170. doi: 10.1006/lich.2000.0309

Lumbsch HT, Schmitt I (2002) Molecular data shake the Pertusariaceae tree into order. Lichenology 1: 37–43.

Lumbsch HT, Schmitt I, Barker D, Pagel M (2006) Evolution of micromorphological and chemical characters in the lichen-forming fungal family Pertusariaceae. Biological Journal of the Linnean Society 89: 615–626. doi: 10.1111/j.1095-8312.2006.00696.x

Messuti MI, Archer AW (2009) ¿Cuántos taxones pueden incluirse teóricamente en el género *Pertusaria*? Glalia 2: 1–9.

Miadlikowska J, Kauff F, Hofstetter V, Fraker E, Grube M, Hafellner J, Reeb V, Hodkinson BP, Kukwa M, Lücking R, Hestmark G, Otalora MG, Rauhut A, Büdel B, Scheidegger C, Timdal E, Stenroos S, Brodo I, Perlmutter GB, Ertz D, Diederich P, Lendemer JC, May P, Schoch CL, Arnold AE, Gueidan C, Tripp E, Yahr R, Robertson C, Lutzoni F (2006)

New insights into classification and evolution of the Lecanoromycetes (Pezizomycotina, Ascomycota) from phylogenetic analyses of three ribosomal RNA- and two protein-coding genes. Mycologia 98: 1088–1103. doi: 10.3852/mycologia.98.6.1088

Misof B, Misof K (2009) A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Systematic Biology 58: 21–34. doi: 10.1093/sysbio/syp006

Muggia L, Gueidan C, Grube M (2010) Phylogenetic placement of some morphologically unusual members of Verrucariales. Mycologia 102: 835–846. doi: 10.3852/09-153

Page RDM (1996) TreeView: an application to display phylogenetic trees on personal computers. Computer Applied Biosciences 12: 357–358.

Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. Bioinformatics 14: 817–818. doi: 10.1093/bioinformatics/14.9.817

Printzen C (2010) Lichen systematics: The role of morphological and molecular data to reconstruct phylogenetic relationships. In: Lüttge U, Beyschlag W, Büdel B, Francis D (Eds). Progress in Botany 71, 233–275.

Rivas Plata E, Lumbsch HT (2011) Parallel evolution and phenotypic divergence in lichenized fungi: a case study in the lichen-forming fungal family Graphidaceae (Ascomycota: Lecanoromycetes: Ostropales). Molecular Phylogenetics and Evolution 61: 45–63. doi: 10.1016/j.ympev.2011.04.025

Rivas Plata E, Lücking R, Lumbsch HT (2012) A new classification for the family Graphidaceae (Ascomycota: Lecanoromycetes: Ostropales). Fungal Diversity 52: 107–121. doi: 10.1007/s13225-011-0135-8

Schmidt HA, Strimmer K, Vingron M, von Haeseler A (2002) TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics 18: 502–504. doi: 10.1093/bioinformatics/18.3.502

Schmitt I, del Prado R, Grube M, Lumbsch HT (2009) Repeated evolution of closed fruiting bodies is linked to ascoma development in the largest group of lichenized fungi (Lecanoromycetes, Ascomycota). Molecular Phylogenetics and Evolution 52: 34–44. doi: 10.1016/j.ympev.2009.03.017

Schmitt I, Fankhauser JD, Sweeney K, Spribille T, Kalb K, Lumbsch HT (2010) Gyalectoid *Pertusaria* species form a sister-clade to *Coccotrema* (Ostropomycetidae, Ascomycota). Mycology 1: 75–83.

Schmitt I, Lumbsch HT (2004) Molecular phylogeny of the Pertusariaceae supports secondary chemistry as an important systematic character set in lichen-forming ascomycetes. Molecular Phylogenetics and Evolution 33: 43–55. doi: 10.1016/j.ympev.2004.04.014

Schmitt I, Yamamoto Y, Lumbsch HT (2006) Phylogeny of Pertusariales (Ascomycotina): Resurrection of Ochrolechiaceae and new circumscription of Megasporaceae. Journal of the Hattori Botanical Laboratory: 753–764.

Shimodaira H, Hasegawa M (2001) CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17: 1246–1247. doi: 10.1093/bioinformatics/17.12.1246

Stamatakis A (2006) RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22: 2688–2690. doi: 10.1093/bioinformatics/btl446

Strimmer K, Rambaut A (2002) Inferring confidence sets of possibly misspecified gene trees. Proceedings of the Royal Society of London Series B, Biological Sciences 269: 137–142. doi: 10.1098/rspb.2001.1862

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research 22: 4673–4680. doi: 10.1093/nar/22.22.4673

Tehler A, Wedin M (2008) Systematics of lichenized fungi. In: Nash TH III (Ed). Lichen Biology Second Edition. Cambridge University Press, Cambridge. viii+486 pages, 336–352.

Wainio EA (1913) Lichenes insularum philippinarum. Philippine Journal of Science, Section C Botany 8: 99–137.

Wedin M, Döring H, Könberg K, Gilenstam G (2005) Generic delimitations in the family Stictidaceae (Ostropales, Ascomycota): the *Stictis-Conotrema* problem. Lichenologist 37: 67–75. doi: 10.1017/S0024282904014653

Westberg M, Crewe AT, Purvis OW, Wedin M (2010) *Silobia*, a new genus for the *Acarospora smaragdula* complex (Ascomycota, Acarosporales) and a revision of the group in Sweden. Lichenologist 43: 7–25. doi: 10.1017/S0024282910000617

MycoKeys 4: 37–63 (2012)
doi: 10.3897/mycokeys.4.3606
www.pensoft.net/journals/mycokeys

RESEARCH ARTICLE

A peer-reviewed open-access journal
MycoKeys
Launched to accelerate biodiversity research

Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences

R. Henrik Nilsson[1], Leho Tedersoo[2,3], Kessy Abarenkov[2], Martin Ryberg[4], Erik Kristiansson[5], Martin Hartmann[6,7], Conrad L. Schoch[8], Johan A. A. Nylander[9], Johannes Bergsten[10], Teresita M. Porter[11], Ari Jumpponen[12], Parag Vaishampayan[13], Otso Ovaskainen[14], Nils Hallenberg[1], Johan Bengtsson-Palme[15], K. Martin Eriksson[1], Karl-Henrik Larsson[16], Ellen Larsson[1], Urmas Kõljalg[2,3]

1 *Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, 405 30 Göteborg, Sweden* **2** *Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia* **3** *Natural History Museum, University of Tartu, Tartu, Estonia* **4** *Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996-1610, USA* **5** *Department of Mathematical Statistics, Chalmers University of Technology, 412 96 Göteborg, Sweden* **6** *Molecular Ecology, Agroscope Reckenholz-Tänikon Research Station ART, Zurich, Switzerland* **7** *Forest Soils and Biogeochemistry, Swiss Federal Research Institute WSL, Birmensdorf, Switzerland* **8** *National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20892, USA* **9** *Department of Biodiversity Informatics, Swedish Museum of Natural History, Box 50007, 104 05 Stockholm, Sweden* **10** *Department of Entomology, Swedish Museum of Natural History, Box 50007, 104 05 Stockholm, Sweden* **11** *Department of Biology, McMaster University, Life Sciences Bldg., 1280 Main Street West, Hamilton, ON L8S 4K1, Canada* **12** *Division of Biology, Kansas State University, Manhattan, KS 66506, USA* **13** *Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA* **14** *Department of Biosciences, University of Helsinki, PO Box 65, FI-00014 Helsinki, Finland* **15** *Department of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Box 434, 405 30 Göteborg, Sweden* **16** *Department of Research and Collections, University of Oslo, Natural History Museum, Postboks 1172, Blindern, 0318 Oslo, Norway*

Corresponding author: *R. Henrik Nilsson* (henrik.nilsson@bioenv.gu.se)

Academic editor: *Thorsten Lumbsch* | Received 29 June 2012 | Accepted 14 August 2012 | Published 5 September 2012

Citation: Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JAA, Bergsten J, Porter TM, Jumpponen A, Vaishampayan P, Ovaskainen O, Hallenberg N, Bengtsson-Palme J, Eriksson KM, Larsson K-H, Larsson E, Kõljalg U (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys 4: 37–63. doi: 10.3897/mycokeys.4.3606

Abstract

Molecular data form an important research tool in most branches of mycology. A non-trivial proportion of the public fungal DNA sequences are, however, compromised in terms of quality and reliability, contribut-

ing noise and bias to sequence-borne inferences such as phylogenetic analysis, diversity assessment, and barcoding. In this paper we discuss various aspects and pitfalls of sequence quality assessment. Based on our observations, we provide a set of guidelines to assist in manual quality management of newly generated, near-full-length (Sanger-derived) fungal ITS sequences and to some extent also sequences of shorter read lengths, other genes or markers, and groups of organisms. The guidelines are intentionally non-technical and do not require substantial bioinformatics skills or significant computational power. Despite their simple nature, we feel they would have caught the vast majority of the severely compromised ITS sequences in the public corpus. Our guidelines are nevertheless not infallible, and common sense and intuition remain important elements in the pursuit of compromised sequence data. The guidelines focus on basic sequence authenticity and reliability of the newly generated sequences, and the user may want to consider additional resources and steps to accomplish the best possible quality control. A discussion on the technical resources for further sequence quality management is therefore provided in the supplementary material.

Key words
ITS, sequence reliability, sequence quality control, fungi, databases, barcoding

Introduction

The inconspicuous and largely subterranean or endophytic nature of much of fungal life presents a challenge to mycology. Many fungal lineages do not seem to produce tangible fruiting bodies, and for those that do, the factors promoting - and acting against - fruiting body formation are only partly understood. As a result, most sampling sites and habitats host a much greater fungal diversity than the above-ground view offered by fruiting bodies would lead the observer to believe (Porter et al. 2008; Hibbett et al. 2011). Furthermore, discriminatory yet easily assessed morphological characters are something of a rare commodity in mycology, and morphology alone often falls short of providing unequivocal species identification and delimitation. For these and other reasons, mycologists were quick to embrace molecular (DNA sequence) data as a research tool in the early 1990s (Horton and Bruns 2001; Anderson and Cairney 2004). Today, DNA sequences represent a key source of information in nearly all branches of mycology, including systematics, taxonomy, and ecology (Stajich et al. 2009), and the landmarks include the establishment of a phylogenetic backbone and a classification system for the fungal kingdom (Blackwell et al. 2006; James et al. 2006; Hibbett et al. 2007).

For all their advantages, molecular data do not solve all open research questions in mycology, and examples of where the misuse and misinterpretation of molecular data hampered mycological progress are easy to point out (Nilsson et al. 2006). Sequences of compromised technical quality or of incorrect taxonomic or ecological annotations are major contributors in this respect in that they may lead researchers to erroneous results and conclusions. When such entries are made publicly available through the international sequence databases, their compromised integrity becomes a problem not only for the researcher who generated them in the first place but for the entire mycological - indeed, scientific - community. Several studies have reported on the various shortcomings of the public DNA sequence corpus (e.g., Gilks et al. 2002; Harris 2003; Bidartondo et

al. 2008), but none have succeeded in halting the continual submission of substandard entries to the databases. On the contrary, there are indications that the proportion of several classes of compromised sequences - such as chimeras and reverse complementary sequences - increases over time (Abarenkov et al. 2010b). While very experienced users may perhaps be able to look through such broken data, many others may not be in a position to do so, particularly not since a growing number of people from outside mycology - even outside the academia - now use fungal sequence data as a part of their work. The highly automated nature of many sequence analysis pipelines similarly makes software suites susceptible to several kinds of sequence errors - such as incorrect taxonomic annotations - since these automata are often built to accept certain classes of information at face value.

The most popular genetic marker for mycological research questions at and below the genus level is the nuclear ribosomal internal transcribed spacer (ITS) region, a ca. 450–650 base pair (bp.) region consisting of the two variable spacers ITS1 and ITS2 and the intercalary, highly conserved 5.8S gene (Begerow et al. 2010). In addition to being widely used for phylogenetic inference and in systematics, the ITS region is proposed as the formal fungal barcode and forms the primary choice for molecular identification of fungi from environmental samples (Vrålstad 2011; Schoch et al. 2012). Several of the present authors have spent significant time pursuing compromised ITS sequences in the International Nucleotide Sequence Databases (INSD: GenBank, ENA, and DDBJ; Karsch-Mizrachi et al. 2012) and UNITE (Abarenkov et al. 2010a; http://unite.ut.ee) or have worked with sequence reliability in other respects. Over time we have noticed several features that signal high-quality, as well as substandard, ITS entries. The most striking observation is probably that, in nearly all cases, severely compromised ITS sequences can be detected manually using just a few simple guidelines (Table 1), without the assistance of technical software packages or access to significant computational power. Many of these guidelines have been put in writing by us and others, but they are scattered across the literature and often mentioned just in passing. In addition, several of them are published in outlets rarely consulted by mycologists. The present publication aims to bring those guidelines and observations on how to establish basic authenticity and reliability of newly generated ITS sequences together in a single, easily digestible publication. The guidelines are simple and straightforward to apply; substantial bioinformatics expertise is not required, and only on-line resources of the paste-and-click type are used. Their simple nature notwithstanding, we believe that these guidelines would have caught the vast majority of the present severely compromised fungal ITS sequences in the public corpus, had they been available and applied at the time of data generation and accessioning.

We would like to stress that the guidelines described here focus on basic sequence authenticity and reliability; they are certainly no panacea for sequence quality management. Their purpose is to assist in pruning severely compromised entries from newly generated, nearly full-length (typically, but not exclusively, Sanger-derived) fungal ITS datasets before those sequences are put to scientific use. The target audience comprises researchers who have just started to use molecular tools (e.g., students) as well as those who otherwise would have taken little action in the direction of quality management.

Table 1. Overview of the five guidelines.

Target of guideline	Way of getting there
1. Establish that the sequences come from the intended gene or marker	Do a multiple alignment of the sequences and verify that they all feature some suitable, conserved sub-region (here the 5.8S gene)
2. Establish that all sequences are given in the correct (5' to 3') orientation	Examine the alignment for any sequences that do not align at all to the others; re-orient these; re-run the alignment step; and examine them again
3. Establish that there are no (bad cases of) chimeras in the dataset	Run the sequences through BLAST in INSD/UNITE and verify that the best match comprises more or less the full length of the query sequences
4. Establish that there are no other major technical errors in the sequences	Examine the BLAST results carefully, particularly the graphical overview and the pairwise alignment, for anomalies
5. Establish that any taxonomic annotations given to the sequences make sense	Examine the BLAST hit list to see that the species names produced make sense

For the user wishing to apply the most advanced and technical quality control solutions to a new dataset right from the start, we provide an account of the bioinformatics of ITS sequence quality control in Appendix. One is nevertheless mistaken to believe that sequence reliability is a matter of bioinformatics only; taxonomic knowledge and common sense are just as important, if much more difficult to algorithmize. What follows is an attempt at a joint treatment of these three aspects.

A word on the query and reference datasets

The sequences in INSD and UNITE are often used as reference datasets to which newly generated ("query") sequences are compared in pursuit of taxonomic and ecological annotation. Neither INSD nor UNITE seek to store full ITS sequence datasets generated by next-generation sequencing (NGS) technologies such as 454 pyrosequencing (Margulies et al. 2005), at least not as primary sequences. The sheer volume and the high frequency of platform-generated sequencing errors derived from NGS approaches necessitate extensive, elaborate quality control measures (Gilles et al. 2011; Quince et al. 2011), and the guidelines presented here should certainly not be used as a replacement for those. Indeed, the present paper primarily targets ITS sequences derived through traditional Sanger sequencing, that is, ITS sequences that usually cover more or less the full length of the ITS region (≥500 bp.). The guidelines thus apply first and foremost to research endeavours where full-length ITS sequences are used, including most ITS-borne studies in systematics, taxonomy, and ecology. Many data mining efforts also fall within the scope of the guidelines, as do the core ITS sequences of INSD/UNITE.

Much of the following will apply also to genes and markers other than the ITS region – particularly the neighbouring ribosomal small subunit (SSU) and large subu-

nit (LSU) genes - and it will certainly apply to the ITS region in groups of organisms other than fungi. Nevertheless, for the sake of example, the user is assumed to have a newly generated fungal ITS dataset (with chromatograms), ideally of near-full-length sequences or at least sequences covering approximately the same part of the ITS region. A proportion of the sequences is assumed to be annotated to various hierarchical classification levels, such as "Uncultured chytridiomycete", *"Penicillium* sp.", and *"Amanita muscaria"*. To avoid overly simplified examples, we will furthermore assume that the data offer some degree of taxonomic complexity and span several fungal phyla and multiple orders. If the dataset is small - say fewer than 50 sequences - the user should probably consider each sequence individually. For datasets up to a few hundred sequences, the user could use a clustering tool such as the BLASTclust implementation at http://toolkit.tuebingen.mpg.de/blastclust to reduce the dataset to one representative sequence per "species" or operational taxonomic unit (OTU; Blaxter et al. 2005). The BLASTclust settings of 97–98% similarity over at least 90% of the length of the shortest sequence in a pairwise alignment will do a reasonable job at approximating the species level in mixed-fungi datasets. For the remainder of this document, the user would then only have to consider one (representative) sequence per such OTU, bypassing the need to address large numbers of near-identical entries. For larger datasets still, the user could further reduce the BLASTclust settings to 85% clustering similarity or even somewhat lower, provided that the length criterion is kept at 90%. The clustering step is optional and only meant as a way to reduce the number of sequences in need of examination; the present paper does not seek to give advice on how to cluster sequences into OTUs for purposes of richness estimation or similar endeavours. While the clustering step removes the user one level from the actual sequence data, we have found the difference to be negligible in terms of basic sequence authenticity and reliability. If any of the clusters contain two or more sequences with full or partial taxonomic annotations, the user should take the opportunity to skim through these to verify that they make approximate sense, meaning that the sequences in the cluster are expected to be annotated as closely related taxa. A cluster with the confamilial ascomycete genera *Penicillium* and *Aspergillus* would probably make sense under the relaxed clustering settings discussed here; a cluster with *Penicillium* (*Ascomycota*) and *Amanita (Basidiomycota)* would not. In the latter case, one or more of the sequences are mislabelled or otherwise deficient, e.g., chimeric. The truly impatient user may now make use of the fact that severely compromised sequences tend to be unique in the nature of their misfortune and thus come out as singletons (clusters of only one sequence) in the clustering process (cf. Huse et al. 2010). However, we argue that checking singletons only is a low-resolution approach that should be reserved for the largest of datasets (more than ~5,000 sequences), and that each sequence or at least representative OTU sequence (preferably the most common sequence type, rather than the consensus sequence or the longest sequence, of each OTU) in smaller datasets should be individually scrutinized using the guidelines provided below.

Guideline 1. It is simple to check that all query sequences represent the ITS region

Upwards of five hundred public sequences are, or have previously been, annotated as ITS sequences when they in fact have been shown to represent other genes or markers or are noise (seemingly random nucleotide letters) throughout. The reasons could be many and range from primer matches to unexpected parts of the genome at hand to the mixing up of test tubes, files, or individual sequences. These sequences contribute significant noise to any data-mining effort targeting the fungal ITS sequence corpus by, e.g., inflating diversity estimates. For molecular identification of fungi, these sequences pose something of an indirect problem, since they are very unlikely to show up in ITS-based BLAST searches (Altschul et al. 1997; documentation at http://www.ncbi.nlm.nih.gov/books/NBK1762/). Nevertheless, a user - knowing that a particular species is present through an ITS sequence in the reference database - may want to confirm the hypothesized taxonomic affiliation of a newly generated ITS sequence, only to arrive at what seems to be a proof that the newly generated sequence does not belong to that very species. In other words, it is a matter of database integrity that genetic annotations really reflect the true marker in question.

An expedient way to ensure that all query sequences represent the ITS region is to compute a multiple sequence alignment in any of a number of on-line multiple alignment services, notably MAFFT (http://mafft.cbrc.jp/alignment/server/ ; Katoh and Toh 2010). Such quickly derived, manually unedited multiple alignments of the ITS region are of limited scientific usefulness save one aspect: the highly conserved, ca. 160 bp. 5.8S gene of the ITS region will form a firm anchor in the middle part of nearly any such alignment. Thus all sequences for which the 5.8S is aligned in this way must be ITS sequences; it is inconceivable that they would produce a good alignment to the 5.8S if they in fact represent a different gene or marker altogether or if they were composed of stochastic, artefactual nucleotide data. Figure 1 shows an alignment featuring five sequences each of the fungal phyla *Ascomycota*, *Basidiomycota, Glomeromycota, Chytridiomycota,* and *Zygomycota* s.l.*;* the reader will probably agree that the 5.8S is easy to spot, despite the disparate taxonomic scope of the sequences. The obvious conclusion is that all sequences in that alignment represent the ITS region. The user is recommended to have MAFFT order the sequences in the alignment by similarity ("Output order: Aligned"), which normally has the effect of forcing any deviant sequences to the bottom of the alignment (or to produce separate sequence blocks that do not align well together). The separation of non-deviant from deviant sequences makes the former much easier to look at and the latter much easier to spot in the first place. The MAFFT server usually returns even large alignment jobs within half an hour, and to scroll down the alignment along the characteristic 5' ("left") end of the 5.8S (cf. Figure 1 or Hibbett et al. 1995) in an alignment editor to check for alignment compliance should not take more than one minute. After that minute – if the 5.8S was found in all sequences - the user can be

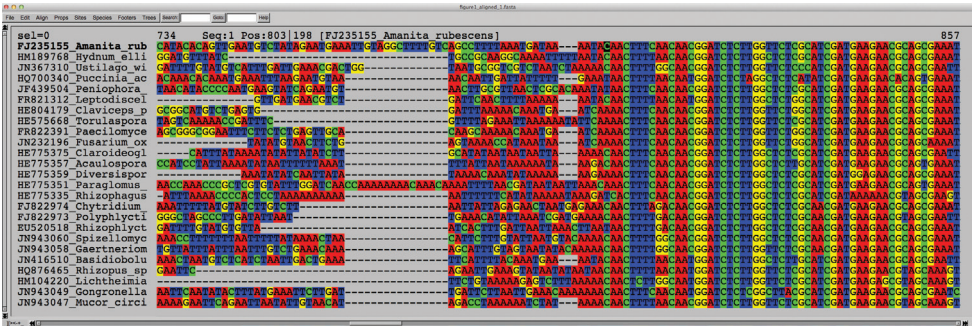


Figure 1. An ITS alignment featuring five random species each of the fungal phyla *Ascomycota, Basidiomycota, Glomeromycota, Chytridiomycota,* and *Zygomycota* s.l. The left half of the screen represents the ITS1 and the right half the 5.8S. Whereas the ITS1 alignment appears more or less chaotic, the 5.8S stands out as a very conserved element throughout these five phyla. The 5.8S starts at position 803 (indicated by the black cursor in the uppermost sequence). Seaview (Gouy et al. 2010) was used to display the alignment.

sure that all sequences in the alignment indeed are ITS sequences. (Strictly speaking they need not be fungal ITS sequences however; oomycete, metazoan, and plant ITS sequences are sometimes retrieved with so-called "fungus-specific" ITS primers (e.g., Tedersoo et al. 2010). The process of verifying hypothesized taxonomic affiliations is discussed in Guideline 5.)

Sequences that do not produce any noteworthy similarity to the 5.8S region of the alignment are likely to belong to one of four categories: 1) they may be partial ITS sequences, containing nothing, or very little, of the 5.8S; 2) they may represent genes or markers other than the ITS (comprising, for example, the 3' SSU intron); 3) they may be of very low read quality or even feature random sequence data altogether; and 4) they may be reverse complementary. The case of reverse complementary sequences is handled separately below (Guideline 2); for the other three - and for the few fungi with truly divergent 5.8S/ITS region sequences, such as *Cantharellus* and *Tulasnella* (Feibelman et al. 1994; Taylor and McCormick 2008) - a simple manual NCBI-BLAST search in INSD is likely to reveal the nature of the complication. The user is advised to pay attention to any sequences for which the 5.8S cannot be located, and not to make scientific use of those sequences until their nature has been clarified.

As an alternative to the alignment-based approach, the user may choose to subject the query sequences - individually or, more likely, in batches - to BLAST searches in INSD. Whether or not a sequence is an ITS sequence can usually be inferred from the annotation of the top five matches alone. As a rule of thumb, a high-quality fungal ITS sequence that features the full 5.8S gene will always produce at least 100 ITS-related BLAST (blastn) matches of a bitscore of about 200 or greater (if only to the 5.8S itself) in INSD under default settings. A sequence that, in contrast, produces just a handful of matches most certainly requires further scrutiny and is, in our experience, very unlikely to qualify as a high-quality ITS sequence in the end.

Guideline 2. A single alignment step can assess the orientation of the query sequences

While it perhaps would seem natural to assume that all newly generated sequences come in the correct (5' to 3') orientation, this is in practice not always the case. A study by Nilsson et al. (2011b) showed that about 1% of the fungal ITS sequences in INSD in fact were given backwards and with all purines and pyrimidines transposed (e.g., ...TAGC... instead of the correct ...GCTA...), that is, they are reverse complementary. Whereas some software tools account for the presence of reverse complementary entries - notably the sequence similarity search engine BLAST - most tools for, e.g., multiple alignment and sequence clustering do not, at least not by default. Reverse complementary sequences can become a tangible problem when sequences are downloaded from sequence databases for use in, e.g., phylogenetic inference or diversity assessments. If the user recognizes the disparate nature of these entries - which the user is likely to do when viewing a multiple alignment but not when working with sequence clustering - the problem is easy to fix through any of a number of web services for sequence reorientation (e.g., http://www.bioinformatics.org/sms/rev_comp.html). However, if the user does not recognize these entries as problematic, they are certain to introduce significant noise into the study.

It would seem likely that most reverse complementary sequences are produced during the contig assembly, a semi-to-fully-automated step where the sequence data produced by each primer employed are brought together to form the full sequence – a contig (cf. Miller and Powell 1994). Whereas the assembly software - such as Sequencher (GeneCodes Corp., Ann Arbor, MI, USA) - usually get sequence orientation and general assembly right, the user sometimes has to step in and provide assistance. Failure of man or machine to account for the read direction of the individual primers may lead to sequence data in the reverse complementary orientation (suggesting that it may be a good idea to add the name of the primer to the name of each primer read to facilitate manual identification of mistakes). Fortunately, the process of establishing read orientation for a set of newly generated ITS sequences is straightforward. A multiple alignment of all query sequences as outlined under Guideline 1, preferably ordered by sequence similarity, is normally enough. By locating the 5.8S gene in that alignment, the user will quickly find any entries that do not seem to contain the 5.8S (Figure 2). By reorienting those seemingly anomalous entries and re-running the multiple alignment step, the user will find out whether any of the sequences in fact were reverse complementary initially. In locating the 5.8S, the user should make sure to check for the characteristic 5' end of the gene (CAACTTTC... or various minor variations thereof in nearly all fungi; see Figure 1 or Hibbett et al. (1995)). Verifying the presence of the 5' end is a necessary precaution against the (unlikely) case that most or all sequences in the alignment in fact are reverse complementary (in the former case, the correctly oriented sequences would be in the minority and appear "anomalous" at the end of the alignment). Excluding the time it takes for the server to compute the multiple alignment, the time consumption of this step is very small - even for large datasets it should be less than five minutes.

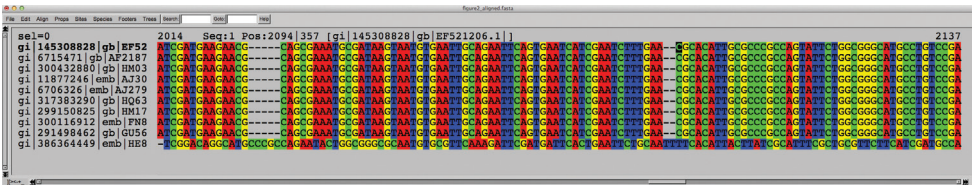


Figure 2. A reverse complementary sequence (bottom) aligned to its nine best BLAST matches, all of which were nearly identical to the query sequence based on BLAST scores, and all of which were given in the correct orientation by their respective authors.

An alternative, and perhaps less advisable, approach to reverse complementary control involves BLAST in INSD. By default, BLAST offers native support for reverse complementary queries (as well as reference sequences) and makes very little noise if a reverse complementary sequence is found. In fact, the user has to scroll down several pages of BLAST output - to the actual alignment produced by BLAST - to get an idea of whether a query sequence is reverse complementary or not. Here, the item "Strand=Plus/Plus" indicates that both the query and the reference sequence are in the same read direction. If the five to ten best matches are all "Strand=Plus/Plus" (and particularly if they come from two or more different studies), the user can be reasonably certain that the query sequence is given in the correct orientation. Similarly, several consecutive "Strand=Plus/Minus" suggest that the query sequence is reverse complementary (Figure 3). Problematically, but logically, a reverse complementary sequence in INSD will produce a "Strand=Plus/Plus" BLAST result to a reverse complementary query, with the second match hopefully showing "Strand=Plus/Minus". In other words, based on the BLAST output alone it is not always easy to conclude which sequence is reverse complementary and which is given in the regular orientation. Indeed, the hypothetical existence of large batches of reverse complementary INSD sequences for some particular species would interfere with the above observations, suggesting that the best way to approach reverse complementary control is by looking at the actual sequence data in a multiple alignment. A special case of reverse complementary sequences - the reverse complementary chimera - is treated under Guideline 4 below.

Guideline 3. PCR chimeras tend to lack full counterparts in the sequence databases and are therefore usually easy to spot through BLAST

The traditional view of a PCR chimera is an artificial sequence resulting from the joining of two (or occasionally more) sequence fragments that do not originate from the same species (see Guideline 4 for a wider definition). In a typical fungal ITS chimera, either the ITS1 comes from one species and 5.8S plus ITS2 come from another, or ITS1 plus 5.8S come from one species and ITS2 from another (Figure 4). In other words, the chimeric breakpoint often seems to be located in the first – and more conserved - part of the 5.8S. These traditional chimeras can unintentionally be

```
>gb|EF521206.1|  Uncultured fungus clone OTU4 18S ribosomal RNA gene, partial
sequence; internal transcribed spacer 1, 5.8S ribosomal RNA
gene, and internal transcribed spacer 2, complete sequence;
and 28S ribosomal RNA gene, partial sequence
Length=646

 Score = 1136 bits (615),  Expect = 0.0
 Identities = 633/641 (99%), Gaps = 6/641 (1%)
 Strand=Plus/Minus

Query  1    TATTGATATGCTTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCAACATTTGCATGA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| || ||
Sbjct  635  TATTGATATGCTTAAGTTCAGCGGGTATTCCTACCTGATCCGAGGTCAACATTT-CA-GA  578

Query  61   AGTTGGGTGTTTTACGGACGTGGACGCGCCGCGCTCCCGGTGCGAGTTGTGCAAACTACT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  577  AGTTGGGTGTTTTACGGACGTGGACGCGCCGCGCTCCCGGTGCGAGTTGTGCAAACTACT  518

Query  121  GCGCATGAGAGGCTGCGGCGAGACCGCCACTGTATTTCGGGGCCGGGATCCCGTCTTAGG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  517  GCGCATGAGAGGCTGCGGCGAGACCGCCACTGTATTTCGGGGCCGGGATCCCGTCTTAGG  458

Query  181  GGTTCCCGAAGTCCCCAACGCCGACCCCCCGGAGGAGGGGTTCGAGGGTTGAAATGACGC  240
            |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Sbjct  457  GGTTCCCGAAGTCCCCAACGCCGACCCCCC---GGAGGGGTTCGAGGGTTGAAATGACGC  401

Query  241  TCGGACAGGCATGCCCGCCAGAATACTGGCGGGCGCAATGTGCGTTCAAAGATTCGATGA  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  400  TCGGACAGGCATGCCCGCCAGAATACTGGCGGGCGCAATGTGCGTTCAAAGATTCGATGA  341

Query  301  TTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTTCATCGATGCC  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  340  TTCACTGAATTCTGCAATTCACATTACTTATCGCATTTCGCTGCGTTCTTCATCGATGCC  281
```

Figure 3. "Strand=Plus/Minus" indicates that the query and reference sequence come in opposing read directions. Another hint comes from the observation that the alignment starts at the first base (1) in the query sequence and progresses upwards to base 60 in the first alignment line; however, for the reference sequence, the alignment starts at base 635 and progresses downwards to base 578.

produced in the PCR step when the DNA of two or more species are present and when the gene or marker in question features a highly conserved segment (here the 5.8S; cf. Fonseca et al. 2012). If the conserved segment in the extending strand is similar enough to the corresponding segment in the contaminant species, this strand can re-anneal to the contaminant DNA instead, with a chimeric sequence as the result. (The risk of producing chimeras in mixed-template PCRs can be reduced by optimizing the PCR protocol, see Wang and Wang 1997 and Qiu et al. 2001.) Chimeras form a particularly treacherous class of compromised sequences, because at a cursory glance they often seem like perfectly fine ITS sequences: all of ITS1, 5.8S, and ITS2 are typically present in their full length and in the expected order. One of the two underlying species dominates the sequence by comprising the ITS1+5.8S or 5.8S+ITS2, and it is the dominant species that tends to prevail in BLAST searches. The scientific (Latin) name given to a chimeric sequence is wrong by definition, but the name is particularly troublesome in cases where the dominant species formed the contaminant (non-targeted) species initially. Such sequences invite BLAST-based misannotations, often spanning fungal orders or even phyla (cf. Hugenholtz and Huber 2003). Chimeric sequences without species names (e.g., "Unidentified fungus")

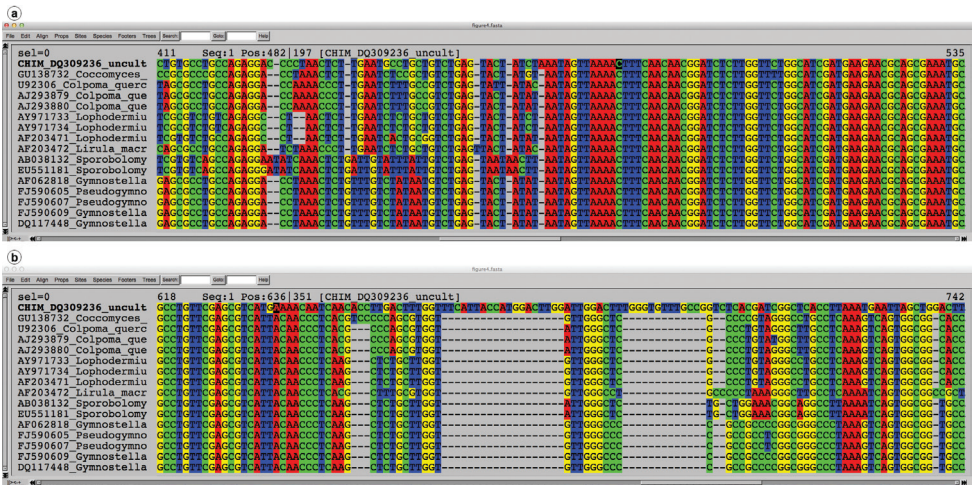


Figure 4. A multiple alignment where the topmost sequence is chimeric and the remaining sequences represent its best BLAST matches. The alignment is fine in ITS1 and 5.8S (**a**; the 5.8S starts at position 479), but the alignment in ITS2 (**b**; position 637 and on) falls far short of scientific rigour. Alignments like these bespeak chimeric unions.

are perhaps less of a problem to molecular species identification, but like all chimeras they inflate diversity assessments such as sequence/OTU richness, estimated richness, and phylogenetic diversity measures (in the latter case for the reason that chimeric sequences tend to form long branches; cf. Tedersoo et al. 2011). Chimeras may however also be detrimental to endeavours other than diversity assessment, for example through skewing multiple alignments.

UNITE has a record of about 1,000 chimeric fungal ITS sequences in the public corpus, corresponding to 0.4% of the number of such sequences. The real number of chimeras is probably significantly higher, since chimeras between closely related species are much more difficult to find than chimeras between distantly related ones. The vast majority of the 1,000 known chimeras are of the "distantly related" type; the chimera in Figure 4 is such an example. Cloning of PCR amplicons is a component in many studies in which chimeras were subsequently reported, suggesting that studies employing cloning should be particularly vigilant against chimeric unions. Fortunately, finding at least bad cases of chimeras in newly generated datasets is fairly straightforward. The solution draws from the observation that chimeric sequences tend to be unique in datasets of small to moderate sizes, i.e., that any given illegitimate union of sequence fragments happened only once in the study. This somewhat rough approximation means that the user can cluster the query dataset at approximately the species level (97-98% similarity, 90% sequence coverage; see above) and then focus on the singletons (or all small-sized OTUs) only. By subjecting the singleton sequences to BLAST searches and keeping an eye on the graphical summary of the BLAST hits provided by NCBI-BLAST (http://blast.ncbi.nlm.nih.gov/), the user will be able to

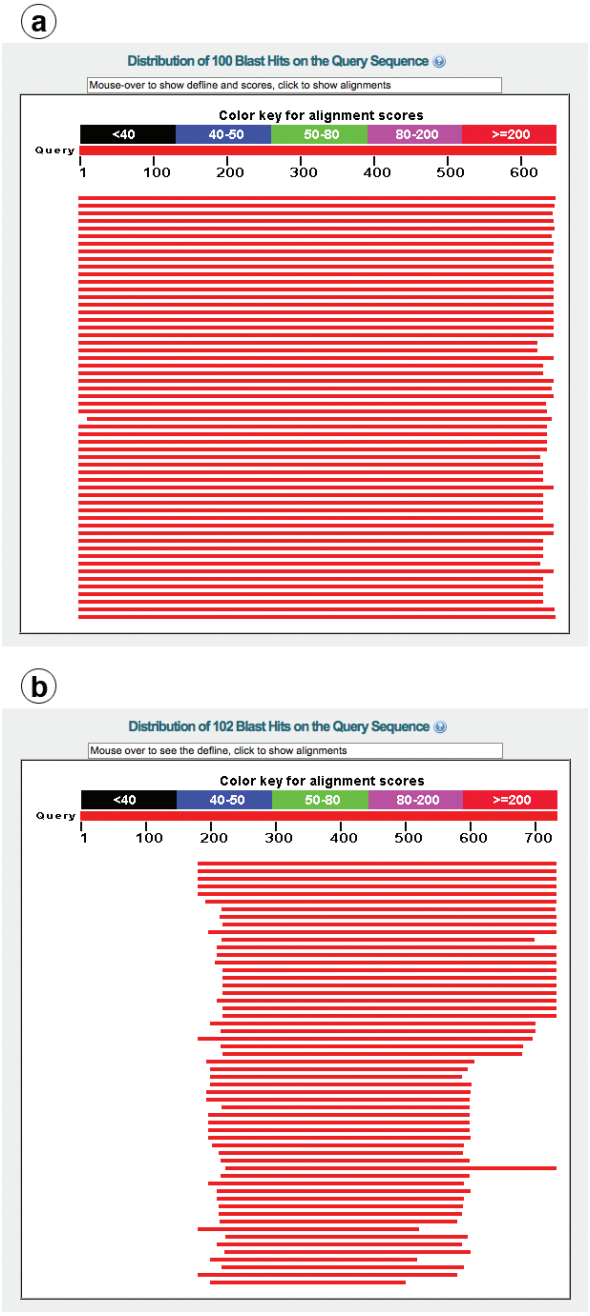


Figure 5. a Graphical overview of the BLAST results of a regular sequence **b** BLAST results of a chimeric sequence where the ITS1 comes from another species, such that the ITS1 is not involved in the alignment featuring the 5.8S+ITS2 (hence the lack of a match for the first ca. 180 bp.). Obviously, a severely compromised sequence that is already in INSD will always find a perfect match through BLAST in INSD: itself. In that case, the presence of a 100% similar reference sequence cannot be used as a testimony to the authenticity of the query sequence.

identify sequences in need of further scrutiny. Figure 5a shows a BLAST run where a query sequence was well matched across its full length by the topmost hits. Figure 5b, in contrast, shows a chimeric sequence where the 5.8S and ITS2 were well matched by the topmost hit, whereas the ITS1 could not be aligned at all to it. This corresponds to the case where the ITS1 comes from a distantly related species with respect to the remainder of the sequence. All cases where ITS1+5.8S - or 5.8S+ITS2 - produce nearly perfect matches, whereas ITS2 or ITS1, respectively, produces an unexpectedly poor match, call for closer scrutiny.

In the case of Figure 5b, it is the ITS1 that does not harmonize with the remainder of the sequence. Doing a BLAST search based on ITS1 alone shows that it is a polypore (100% similarity); the 5.8S+ITS2 BLAST, in contrast, shows that those parts belong to an agaric (100% similarity). By doing separate BLAST searches like this, the user will come fairly close to practical proof that the sequence in question is chimeric. Such sequences should be pruned from the query dataset, and they should similarly not be submitted to the sequence databases. However, the user should keep in mind that legitimate query sequences - particularly long ones - can also produce BLAST results similar to that in Figure 5b for the reason that the most similar reference sequences were much shorter due to, e.g., primer choice. The BLAST alignment indicates at what base in the query and the reference sequence the alignment starts. For example, if the alignment start is "1" in the reference sequence but "350" in the query sequence, then the seemingly odd BLAST results simply reflect the absence of reference data. Introns such as the one at the 3' end of the SSU may produce similar results. However, also in these situations, subjecting the non-matching part of the query sequence to a BLAST search is likely to reveal the nature of the problem.

Problematically, not all cases of chimera detection will be as straightforward as the example in Figure 5b, and the user will sometimes face difficult decisions. After all, ITS sequence data are available for a mere 1% of the hypothesized 1.5 million extant species of fungi (Hawksworth 2001; Hibbett et al. 2011), and some newly generated sequences will be singletons, and perhaps look odd, for the reason that they have not been sequenced before, such that no fit objects of comparison are available. To routinely exclude sequences that differ from known sequences would obviously not be a good way to expand our knowledge of the fungal kingdom. The user is probably best advised to delete the sequences she feels sure are chimeric and leave the rest of the sequences in the dataset; it would still be a major improvement over not checking the dataset for chimeras at all. If these dubious sequences are of particular relevance to the study, and if there is fungal material left from which to regenerate those sequences, then the user would have the opportunity to verify the biological, or artefactual, origin of those sequences through another round of sequencing. A further complication is that in studies with great sequencing depth, more or less identical chimeras between the most common OTUs may occur more than once in the dataset. A solution to this problem could be to check a representative sequence also from OTUs that are not singletons (focusing, as needed, on all OTUs with few constituent sequences).

Guideline 4. Sequences can be broken in other, puzzling ways; BLAST, again, will tell

BLAST also has the capacity to indicate several other classes of compromised entries. Figure 6 shows an assembly chimera, which is the product of incorrect assembly of two or more sequence fragments (primer reads) into a single contig. The dotted vertical line in the reference sequences indicates a break in the alignment between these and the query sequence. The user will have to scroll down to the BLAST alignment to learn of the exact nature of the break. Often one finds that such sequences were assembled with the ITS1 and the ITS2 in the wrong order. The resulting BLAST alignment will be divided into sections, and the user might find that, e.g., base 285 to 614 in the query sequence are matched by bases 1 to 330 in the reference sequence. Bases 1-284 in the query are, however, best matched by bases 331-614 in the reference sequence; although it may not always be straightforward to see exactly what the problem is, the non-contiguous nature of these alignment segments at least makes it easy to see that there indeed seems to be a problem to begin with. If all alignment sections are in the Strand=Plus/Plus orientation, and the next few reference sequences similarly produced such sectioned alignments with respect to the query, then the user can be certain that the query sequence is an assembly chimera. It is easy to see that assembly chimeras may follow as a result of minimal overlap between the fragments under assembly and the subsequent failure of the contig software – under the settings applied - to pick the correct ends for merger. If there is no overlap at all between the fragments - such that there should have been additional sequence data between two fragments that are now joined - the corresponding BLAST results will look something like Figure 6. Such bridged sequences may also be produced inadvertently in, e.g., the phylogenetic analysis package PAUP (Swofford 2003) when the user excludes certain alignment regions from the analysis due to, e.g., poor alignability using the generic "EXCLUDE" command. If the user then exports the alignment analysed for INSD or TreeBase (Sanderson et al. 1994) deposition, the individual sequences will lack the parts excluded from the analysis and therefore qualify as chimeric. Alternatively, if an extraneous sequence segment was assembled into a position where it should not have been, such as in the middle of the 5.8S, the BLAST results tend to look similar to those shown in Figure 7. Finally, reverse complementary chimeras are produced when a sequence is assembled to contain one or more fragments in the regular orientation and one or more fragments in the reverse complementary orientation (cf. Hartmann et al. 2011). The BLAST results of such sequences often look like Figure 6, and the BLAST alignment will indicate that one or more of the sections are in the opposite direction, "Strand=Plus/Minus".

The distal (5' and 3') ends of newly generated sequences are typically of lower read quality than the interior parts of the sequence. It is the job of the contig assembly software to highlight poorly read bases clearly enough that the user can address them before the final sequence is produced from the contig. Untrimmed sequences tend to look like the one in Figure 8 when run through BLAST; note that the match does not include the first ca. 20, and the last ca. 30, base-pairs. Unless all of the reference

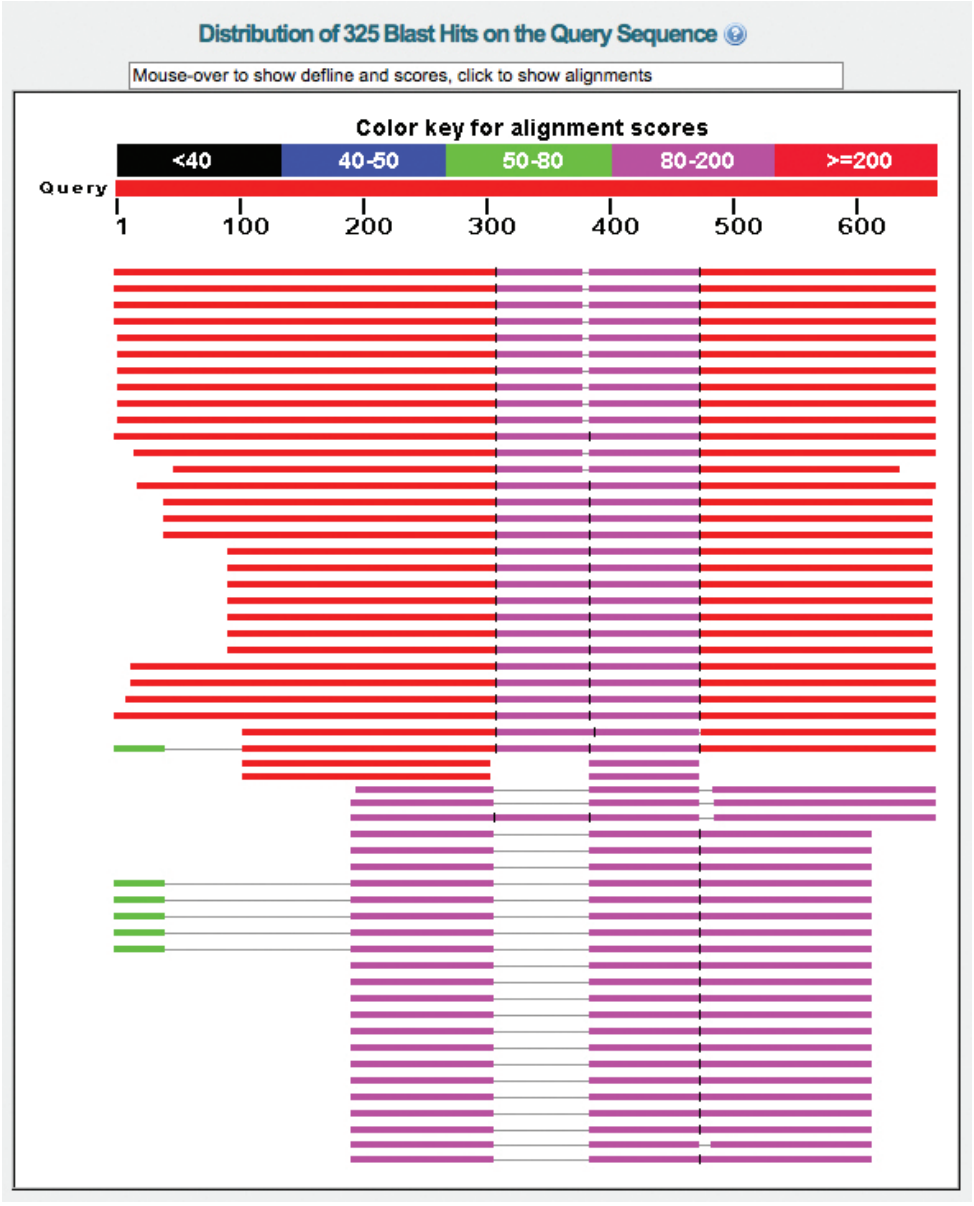


Figure 6. An assembly chimera. The black dashed lines indicate breaks in the BLAST alignment and should always be taken to mean that manual examination is needed.

sequences are in fact shorter than the query sequence, the user should probably re-check the chromatograms in the distal parts of the sequence - and consider trimming regions of poor quality - at this stage. Many public ITS sequences, in turn, are poorly trimmed, sometimes leaving the process of telling whether it is the query or the reference sequence that features the low-quality bases all but intractable. This speaks to the

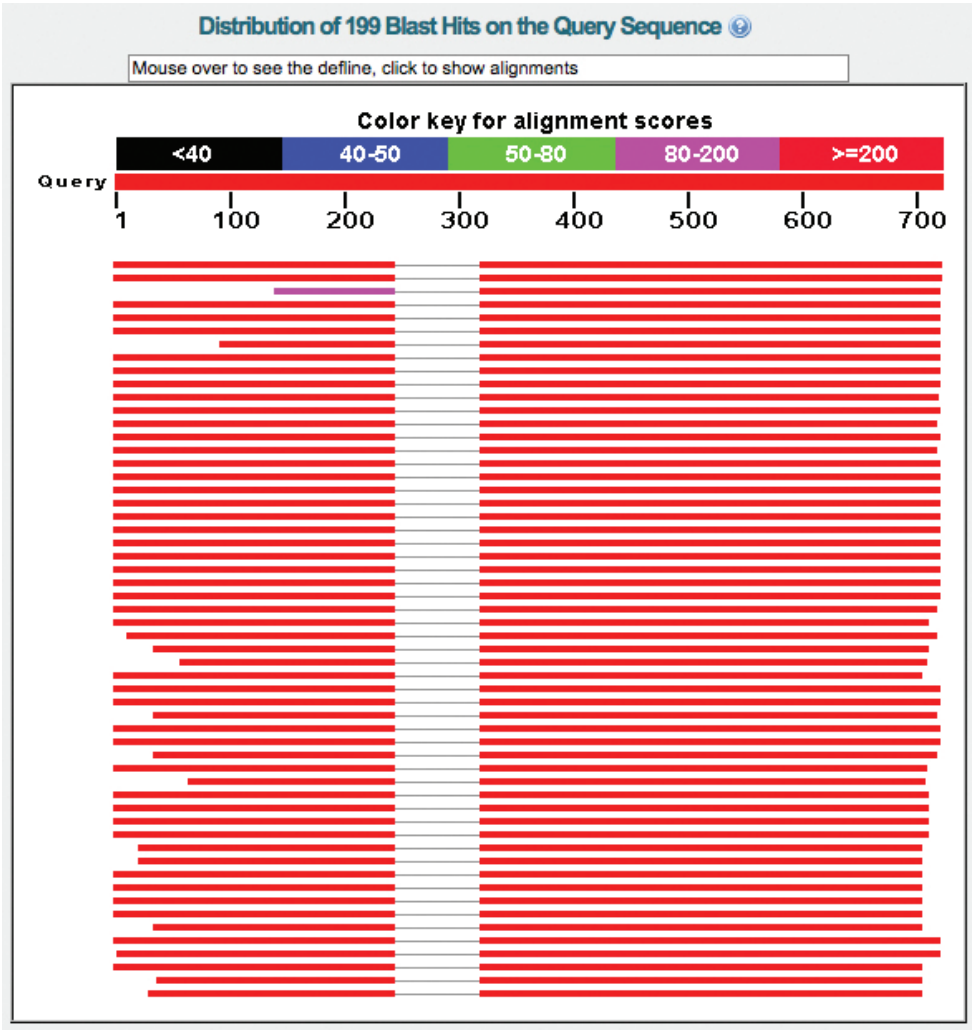


Figure 7. An assembly chimera. An extraneous sequence segment was assembled into a position where it should not have been, such as in the middle of the 5.8S. The white area in the reference sequences indicates the absence of sequence data for this particular part of the query sequence. Manual examination is always needed in cases like this.

importance of always taking the sequence assembly step seriously and of paying special attention to any region where the chromatograms appear substandard. Other newly generated sequences are of reduced read quality throughout. One obvious sign is that they may feature IUPAC DNA ambiguity symbols (e.g., N and S; Cornish-Bowden 1985). If these are scattered along the full length of the sequence, our experience is that the sequence should be discarded altogether. If they, on the other hand, are clustered in some single region of the sequence - typically at either distal end - and the chromatograms look satisfactory in the remaining regions of the sequence, then the sequence

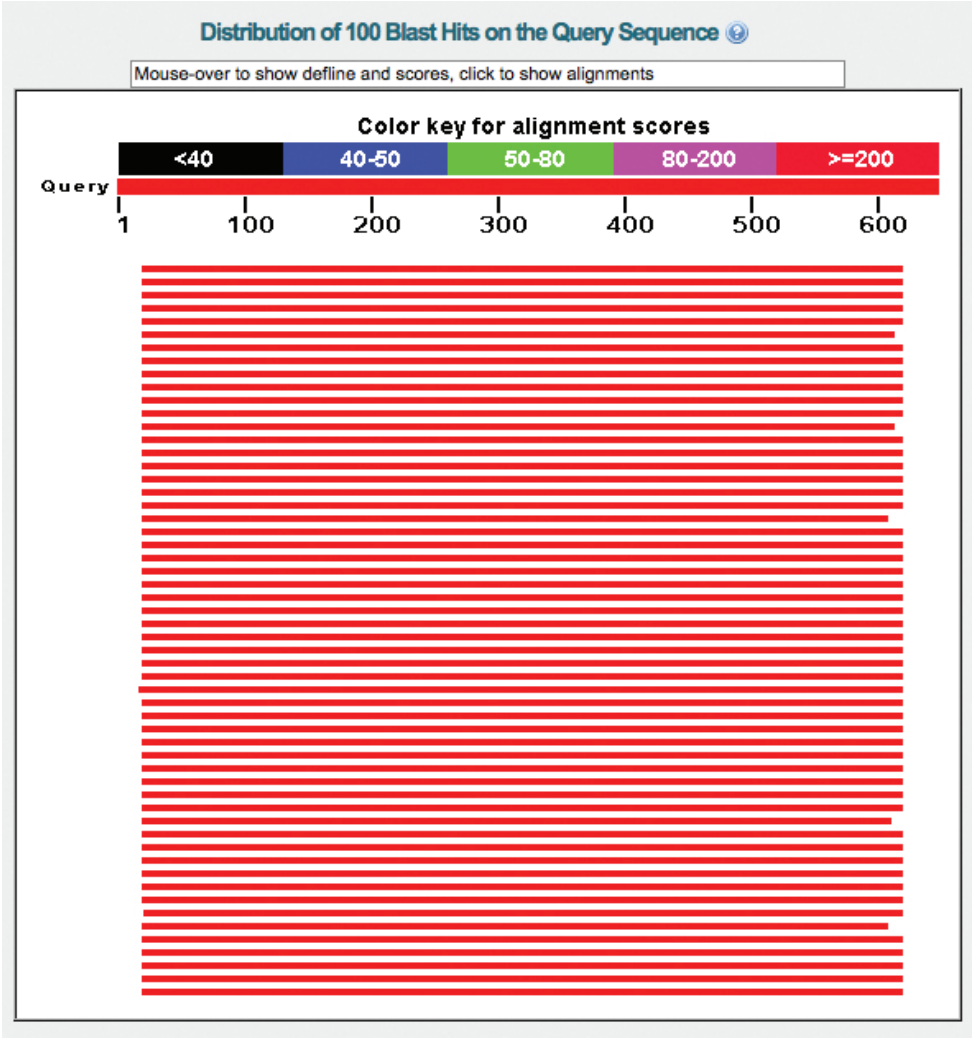


Figure 8. Untrimmed sequences tend to look like this when run through BLAST. Note how the first ca. 20 bp., and the last ca. 30 bp., of the query sequence (represented by the red bar with scale marks every 100 bp.) do not align to any of the BLAST hits. The use of different but closely situated primers may give a similar pattern, however, pointing at the need to also look at the BLAST alignments for start and end positions of the reference sequences.

is probably reliable (although in need of distal trimming). Another tell-tale sign may be suspiciously large homopolymer regions (e.g., ...AAAAAAAAAA...); again the user should go back to the chromatograms to scrutinize these regions. A complication is that the underlying fungal individual may have alleles of different lengths in these regions, making exact base-calling hard. Of particular difficulty are those sequences in which neither ambiguity symbols nor suspicious homopolymer regions are present, but that still are very distant from the closest BLAST hit. The BLAST alignment may

offer some tentative clues here. If the mismatches are scattered more or less evenly across the full length of the query sequence, it is likely that the general sequence quality is substandard, such that the sequence should be discarded. If, on the other hand, there are no - or significantly fewer - mismatches in the region corresponding to the 5.8S in the BLAST alignment, this would suggest that the sequence is authentic, if very deviant from everything else. Indeed, several large groups of previously unknown fungi have been described in recent years (e.g., Jones et al. 2011; Rosling et al. 2011).

Guideline 5. Taxonomic annotations should be verified before the sequences are used

About half of the 250,000 public, full-length fungal ITS sequences are annotated to the level of species (Hibbett et al. 2011). Several studies have, however, shown that the taxonomic reliability of the entries in the public sequence databases has yet to reach perfection, and more than 10% of the public fungal ITS sequences that carry a species name may in fact carry an incorrect species name (e.g., Nilsson et al. 2006). It is easy to see that morphology-based species identification procedures sometimes go wrong among closely related or otherwise highly similar species, and these misidentifications would then carry over to the taxonomic annotation of the sequence generated from the specimen. Many of the misidentifications we have come across, however, span orders, classes, and frequently also phyla of the fungal tree of life. Indeed, more than 20 fungus-related cases of misidentification at the kingdom level are indexed in UNITE. This suggests that taxonomic competence is only one of several processes leading to incorrect taxonomic annotations of public sequences. Unintended sequencing of epifungal - or intrasporocarp - parasites, mutualists, or commensalists appears common, for example. PCR contaminations and the mixing up of test tubes, computer files, and labels stand out as other major sources of error. Incorrectly identified or contaminated cultures – even in the major international culture collections – form an additional, serious concern. The conclusion is obvious: nobody - regardless of degree of taxonomic competence - should by default assume that their taxonomic annotations are correct and not in need of verification.

We take the position that all sequences in a newly generated dataset should be verified for taxonomic affiliation, even if they are annotated only to kingdom level (e.g., "Uncultured fungus"). The process of verifying a hypothesized taxonomic annotation - or at least ruling out the possibility that the annotation is way off - is usually trivial and amounts to a simple BLAST run. A sequence annotated as *Penicillium* is expected to hit other *Penicillium* sequences (usually in a chaotic list of anamorphic and teleomorphic names, species complexes, and numerous environmental sequences; a visit to Index Fungorum (http://www.indexfungorum.org/) or MycoBank (http://www.mycobank.org/) may be needed to establish the relations of the names obtained). A quick check of some degree of consistency among the top ten matches is normally enough to confirm the basic authenticity of the taxonomic affiliation, particularly if the top ten matches stem from two or more different studies. The INSD keyword "BARCODE" (specified in the description of the entry)

indicates that a sequence complies with a number of quality criteria (http://barcoding.si.edu/pdf/dwg_data_standards-final.pdf) and so should be weighted as a more reliable reference sequence. However, looking at BLAST hit lists is often more difficult than one might think. The following five basic principles may be good to keep in mind. **a)** BLAST is sensitive to the length and level of sequence conservation of the query and reference sequences, and the user is advised to prune any large parts of the SSU and LSU from the ITS sequences before doing BLAST searches (cf. Kang et al. 2010). It sometimes pays off to use only the ITS1 or ITS2 for the searches. **b)** BLAST does local alignment and so will base its core statistics on the part of the query sequence it managed to align rather than the full length of the query sequence. Thus, even if a match says "100% similar", it will typically not apply to the full length of the query sequence, and confirming the proportion of the sequence aligned requires examination of the coverage statistics reported in the BLAST searches. If the user is concerned with the absolute similarity of the query sequence to the best match, a second alignment step (in, e.g., MAFFT) and a pocket calculator may be needed. **c)** In the case of identical BLAST bitscores (matches), the order of the hits is for all practical purposes uninformative. This cautions against looking only at the very topmost match; if there are several equally good matches, they are all equally relevant. **d)** The degree to which the ITS region is species specific differs among fungal lineages, as does the average distance to the closest species for any given species (Nilsson et al. 2008). It is a good idea to refrain from oversimplified approaches to species identification and sequence annotation, such as enforcing a strict 97% similarity criterion at all times. Indeed, BLAST reports on similar sequences rather than species names. **e)** The taxon sampling of fungi is still very much incomplete (Brock et al. 2009; Nilsson et al. 2011a). Thus, even if some particular query sequence does not hit any of the species the user had expected - but more remotely related ones instead - it does not have to mean that anything is wrong; it could just be a case of thin taxon sampling. The GenBank Nucleotide (http://www.ncbi.nlm.nih.gov/nuccore/) query string "Amanita[ORGN] AND 5.8S[TITL]" will show whether there are any ITS sequences annotated as belonging to the genus *Amanita* in GenBank. Such simple queries permit the user to examine and establish whether the expected species are present among the available ITS sequences in the database. Ross et al. (2008) and Ovaskainen et al. (2010) provide interesting statistics on the performance of BLAST under varying conditions, including incomplete database coverage.

It is typically simple to establish basic authenticity of the taxonomic annotations for a set of query sequences. The process described above will often take the user to the genus level or even the species level in some cases, at which stage one can rule out severe misannotation. Going all the way to actually verifying the species-level annotation is a trickier objective, and one that will not always be possible based on BLAST and the public sequence databases alone. A phylogenetic analysis of the query sequence and the 20-30 best BLAST matches (or as many as alignability allows) is a good starting point for a more robust examination of the taxonomic affiliation of the query sequence (cf. Taylor et al. 2000). The alignment/phylogenetic analysis combination may also be helpful in locating otherwise anomalous sequence data; (single) sequences that are found on unusually long branches or that do not find well-supported positions may be worth looking closer at.

Taking action on bad sequences in INSD / UNITE

Anyone using the public sequence databases to pursue low-quality entries in a newly generated dataset will sooner or later find low-quality entries also in these databases. When skimming through BLAST hit lists, for instance, one regularly sees entries whose taxonomic annotation simply has to be wrong for one reason or the other - a single *Betula* (birch) in a list of *Amanita* (fly agaric), for instance. It is easy to feel that some mistakes are so far off and absurd as to be harmless. In reality they are harmless only to a limited number of people, namely those with a relevant taxonomic background; with a reasonable insight into how BLAST operates; and with enough time on their hands to interpret their sequence similarity searches manually. Everyone else may be in harm's way. We did an informal evaluation of 20 fungal ITS sequences whose taxonomic annotation was off at the ordinal or class level by simply running the accession numbers through Google. Three of the sequences (15 %) had been used under their original (incorrect) name in at least one other scientific publication than the one through which they were released. Even taxonomic experts would be hard put to spot many such derived mistakes since they are published one level removed from the original data, suggesting a route through which errors and mistakes can be cited and re-cited enough to eventually be accepted as truths. There is thus every reason to take some form of action when one comes across a public DNA sequence associated with significant error.

Hartmann et al. (in press) discuss several ways to take action on compromised public sequences. We will assume here that the user is very pressed for time and unwilling to spend more than a minute on the matter; we also assume that the nature of the complication is severe enough to be beyond questioning or interpretation. A quick, friendly email to the original sequence authors is in fact likely to solve the problem altogether, because few scientists would presumably like their names to be associated with persistent, broken data. Sequence authors have considerable say over their entries in INSD, and a request from them to the INSD staff (e.g., http://www.ncbi.nlm.nih.gov/About/glance/contact_info.html) is unlikely to go unheeded. It is however notoriously difficult to find people and their present contact information over the web (cf. Wren 2008). Another path to take action on broken sequences is therefore to write an email to the INSD staff directly. We have found the INSD staff to be very friendly and service-minded in these matters. Several options are open to the INSD staff to deal with misidentified sequences. One recent example is to add an UNVERIFIED keyword to highly problematic sequences and exclude the sequence from BLAST, although the sequence will still be archived in INSD. Finally, it is possible to use the third-party sequence annotation feature of UNITE/PlutoF (Abarenkov et al. 2010b) to simply replace the incorrect species name with the correct one, or to mark the entry as chimeric, or to take whatever other action appropriate. Third-party annotations of sequences via PlutoF are visible to users in the European Nucleotide Archive of the INSD through a link-out function. We feel that the exact way in which the user chooses to take action is less important compared to whether or not the user chooses to take action in the first place, and we hope that the mycological community will be able to set a high standard here.

Concluding remarks

The present document brings together a set of guidelines, recommendations, and observations towards identifying severely compromised sequences before they are put to scientific use. While they were written with the non-bioinformatician in mind and aim to be non-technical and straightforward to apply, we still believe they are powerful enough to have prevented the deposition of the vast majority of severely compromised fungal ITS sequences in the public sequence databases, had they been applied at the time of sequence accessioning. Importantly, however, these guidelines would not have caught all cases of badly damaged sequence data. Thus, the application of the principles presented here will not guarantee - but rather just increase the chance - that the dataset at hand will be of reasonable standard after processing. Furthermore we would like to stress that these guidelines offer little in way of fine-grained authenticity and reliability. Misidentification among closely related species, somewhat reduced levels of general sequence read quality, and base-inflation in homopolymer regions are all examples of problems that are only partly addressed by this document. We certainly do not want our guidelines to be used as replacements for more advanced, technical solutions; we rather hope that they will be used by those who, for one reason or the other, do not have access to or would not consider running any advanced, technical solutions in the first place (e.g., Appendix).

Our guidelines come with no other software requirement than a web browser. They still require something else of the user too: a critical, inquisitive, and perhaps imaginative mind. It would seem impossible to lay down firm rules to which all high-quality sequences would comply and that all low-quality sequences would violate. Rather the user should expect to find herself in situations where the user herself is the best arbiter of what is correct and what isn't. Although such a situation would not be unfamiliar to anyone in systematics or taxonomy, we would still like to point out the importance of common sense in pursuing broken sequence data. The present authors spent considerable time trying to make this document as rich and multi-faceted as possible, but it goes without saying that additional, relevant observations and advice are to be found among the remaining members of the scientific community. We hope that anyone in the position to improve or add to the present set of guidelines will take the time and opportunity to do so. The potential outlets are many and range from the "Add comments" feature of the present journal to separate publications in this or any other journal. The ever-increasing weight assigned to molecular data in mycology - and the life sciences as a whole - suggests that any such move may have positive ramifications extending far beyond the datasets of each individual user.

Acknowledgements

RHN acknowledges financial support from Swedish Research Council of Environment, Agricultural Sciences, and Spatial Planning (FORMAS, 215-2011-498). LT

and UK received financial support from ETF grants 8235, 9286, and FIBIR. EK acknowledges support from the Life Science Area of Advance at Chalmers University of Technology, the Swedish Research Council (VR), and FORMAS. CLS was supported in part by the Intramural Research Program of the NIH, National Library of Medicine. TMP acknowledges financial support from the Government of Canada through Genome Canada and the Ontario Genomics Institute through the Biomonitoring 2.0 project (OGI-050). AJ acknowledges financial support from Department of Energy's Biological and Environmental Research Program (Award #ER65000). Yann Bertrand and Chris Quince are acknowledged for valuable input on the manuscript. Support from the Gothenburg Bioinformatics Network is gratefully acknowledged.

References

Abarenkov K, Nilsson RH, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, Sen R, Taylor AFS, Tedersoo L, Ursing BM, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U (2010a) The UNITE database for molecular identification of fungi - recent updates and future perspectives. New Phytologist 186(2): 281–285. doi: 10.1111/j.1469-8137.2009.03160.x

Abarenkov K, Tedersoo L, Nilsson RH, Vellak K, Saar I, Veldre V, Parmasto E, Prous M, Aan A, Ots M, Kurina O, Ostonen I, Jõgeva J, Halapuu S, Põldmaa K, Toots M, Truu J, Larsson K-H, Kõljalg U (2010b) PlutoF - a web-based workbench for ecological and taxonomical research, with an online implementation for fungal ITS sequences. Evolutionary Bioinformatics 6: 189–196. doi: 10.4137/EBO.S6271

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25(17): 3389–3402. doi: 10.1093/nar/25.17.3389

Anderson IC, Cairney JWG (2004) Diversity and ecology of soil fungal communities: increased understanding through the application of molecular techniques. Environmental Microbiology 6(8): 769–779. doi: 10.1111/j.1462-2920.2004.00675.x

Begerow D, Nilsson RH, Unterseher M, Maier W (2010) Current state and perspectives of fungal DNA barcoding and rapid identification procedures. Applied Microbiology and Biotechnology 87(1): 99–108. doi: 10.1007/s00253-010-2585-4

Bidartondo M, Bruns TD, Blackwell M et al. (2008) Preserving accuracy in GenBank. Science 319(5870): 1616. doi: 10.1126/science.319.5870.1616a

Blackwell M, Hibbett DS, Taylor JW, Spatafora JW (2006) Research Coordination Networks: a phylogeny for kingdom Fungi (Deep Hypha). Mycologia 98(6): 829–837. doi: 10.3852/mycologia.98.6.829

Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R, Abebe E (2005) Defining operational taxonomic units using DNA barcode data. Philosophical Transactions of the Royal Society of London Series B – Biological Sciences 360(1462): 1935–1943. doi: 10.1098/rstb.2005.1725

Brock PM, Döring H, Bidartondo MI (2009) How to know unknown fungi: the role of a herbarium. New Phytologist 181(3): 719–724. doi: 10.1111/j.1469-8137.2008.02703.x

Cornish-Bowden A (1985) Nomenclature for incompletely specified bases in nucleic acid sequences: recommendations 1984. Nucleic Acids Research 13(9): 3021–3030. doi: 10.1093/nar/13.9.3021

Feibelman T, Bayman P, Cibula WG (1994) Length variation in the internal transcribed spacer of ribosomal DNA in chanterelles. Mycological Research 98(6): 614–618. doi: 10.1016/S0953-7562(09)80407-3

Fonseca VG, Nichols B, Lallias D, Quince C, Carvalho GR, Power DM, Creer S (2012) Sample richness and genetic diversity as drivers of chimera formation in nSSU metagenetic analyses. Nucleic Acids Research 40(9): e66. doi: 10.1093/nar/gks002

Gilks WR, Audit B, De Angelis D, Tsoka S, Ouzounis CA (2002) Modeling the percolation of annotation errors in a database of protein sequences. Bioinformatics 18(12): 1641–1649. doi: 10.1093/bioinformatics/18.12.1641

Gilles A, Meglécz E, Pech N, Ferreira S, Malausa T, Martin J-F (2011) Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing. BMC Genomics 12: 245. doi: 10.1186/1471-2164-12-245

Gouy M, Guindon S, Gascuel O (2010) SeaView Version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Molecular Biology and Evolution 27(2): 221–224. doi: 10.1093/molbev/msp259

Harris JD (2003) Can you bank on GenBank? Trends in Ecology and Evolution 18(7): 317–319. doi: 10.1016/S0169-5347(03)00150-2

Hartmann M, Howes CG, Veldre V, Schneider S, Vaishampayan PA, Yannarell AC, Quince C, Johansson P, Johanna Björkroth K, Abarenkov K, Hallam SJ, Mohn WW, Nilsson RH (2011) V-RevComp: Automated high-throughput detection of reverse complementary 16S ribosomal RNA gene sequences in large environmental and taxonomic datasets. FEMS Microbiology Letters 319(2): 140–145. doi: 10.1111/j.1574-6968.2011.02274.x

Hartmann M, Bengtsson J, Nilsson RH (in press) High-throughput quality control of next-generation sequencing datasets from environmental studies using hidden Markov models.

Hawksworth DL (2001) The magnitude of fungal diversity: the 1.5 million species estimate revisited. Mycological Research 105(12): 1422–1432. doi: 10.1017/S0953756201004725

Hibbett DS, Tsuneda A, Fukumasa-Nakai Y, Donoghue MJ (1995) Phylogenetic diversity in shiitake inferred from nuclear ribosomal DNA sequences. Mycologia 87(5): 618–638. doi: 10.2307/3760806

Hibbett DS, Bindera J, Bischoffb JF, Blackwellc M, Cannond PF, Erikssone OE, Huhndorff S, Jamesg T, Kirkd PM, Lückingf R, Lumbschf HT, Lutzonig F, Mathenya PB, McLaughlinh DJ, Powelli MJ, Redheadj S, Schochk CL, Spataforak JW, Stalpersl JA, Vilgalysg R, Aimem MC, Aptrootn A, Bauero R, Begerowp D, Bennyq GL, Castleburym LA, Crousl PW, Dair Y-C, Gamsl W, Geisers DM, Griffitht GW, Gueidang C, Hawksworthu DL, Hestmarkv G, Hosakaw K, Humberx RA, Hydey KD, Ironsidet J-E, Kõljalgz U, Kurtzmanaa CP, Larssonab K-H, Lichtwardtac R, Longcoread J, Miądlikowskag J, Millerae A, Moncalvoaf J-M, Mozley-Standridgeag S, Oberwinklero F, Parmastoah E, Reebg V, Rogersai JD, Roux-

aj C, Ryvardenak L, Sampaioal JP, Schüßleram A, Sugiyamaan J, Thornao RG, Tibellap L, Untereineraq WA, Walkerar C, Wanga Z, Weiras A, Weisso M, Whiteat MM, Winkae K, Yaoau Y-J, Zhang N (2007) A higher-level phylogenetic classification of the Fungi. Mycological Research 111(5): 509–547. doi: 10.1016/j.mycres.2007.03.004

Hibbett DS, Ohman A, Glotzer D, Nuhn M, Kirk P, Nilsson RH (2011) Progress in molecular and morphological taxon discovery in Fungi and options for formal classification of environmental sequences. Fungal Biology Reviews 25(1): 38–47. doi: 10.1016/j.fbr.2011.01.001

Horton TR, Bruns TD (2001) The molecular revolution in ectomycorrhizal ecology: peeking into the black-box. Molecular Ecology 10(8): 1855–1871. doi: 10.1046/j.0962-1083.2001.01333.x

Hugenholtz P, Huber T (2003) Chimeric 16S sequences of diverse origin are accumulating in the public databases. International Journal of Systematic and Evolutionary Microbiology 53(1): 289–293. doi: 10.1099/ijs.0.02441-0

Huse SM, Welch DM, Morrison HG, Sogin ML (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. Environmental Microbiology 12(7): 1889–1898. doi: 10.1111/j.1462-2920.2010.02193.x

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung G-H, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüßler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies RD, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lücking R, Büdel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R (2006) Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature 443(7113): 818–822. doi: 10.1038/nature05110

Jones MDM, Forn I, Gadelha C, Egan MJ, Bass D, Massana R, Richards TA (2011) Discovery of novel intermediate forms redefines the fungal tree of life. Nature 474(7350): 200–203. doi: 10.1038/nature09984

Kang S, Mansfield MAM, Park B, Geiser DM, Ivors KL, Coffey MD, Grünwald NJ, Martin FN, Lévesque CA, Blair JE (2010) The promise and pitfalls of sequence-based identification of plant pathogenic fungi and oomycetes. Phytopathology 100(8): 732–737. doi: 10.1094/PHYTO-100-8-0732

Karsch-Mizrachi I, Nakamura Y, Cochrane G (2012) The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research 40(D1): D33–D37. doi: 10.1093/nar/gkr1006

Katoh K, Toh H (2010) Parallelization of the MAFFT multiple sequence alignment program. Bioinformatics 26(15): 1899–1900. doi: 10.1093/bioinformatics/btq224

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Hel-

gesen S, Ho CH, Irzyk JP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu HL, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057): 376-380. doi: 10.1038/nature03959

Miller MJ, Powell JI (1994) A quantitative comparison of DNA sequence assembly programs. Journal of Computational Biology 1(4): 257–269. doi: 10.1089/cmb.1994.1.257

Nilsson RH, Ryberg M, Kristiansson E, Abarenkov K, Larsson K-H, Kõljalg U (2006) Taxonomic reliability of DNA sequences in public sequence databases: a fungal perspective. PLoS ONE 1: e59. doi: 10.1371/journal.pone.0000059

Nilsson RH, Kristiansson E, Ryberg M, Hallenberg N, Larsson K-H (2008) Intraspecific ITS variability in the kingdom Fungi as expressed in the international sequence databases and its implications for molecular species identification. Evolutionary Bioinformatics 4: 193–201. doi: 10.4137/EBO.S653

Nilsson RH, Ryberg M, Sjökvist E, Abarenkov K (2011a) Rethinking taxon sampling in the light of environmental sequencing. Cladistics 27(2): 197-203. doi: 10.1111/j.1096-0031.2010.00336.x

Nilsson RH, Veldre V, Wang Z et al. (2011b) A note on the incidence of reverse complementary fungal ITS sequences in the public sequence databases and a software tool for their detection and reorientation. Mycoscience 52(4): 278–282. doi: 10.1007/s10267-010-0086-z

Ovaskainen O, Nokso-Koivistoa J, Hottolaa J, Rajalab T, Pennanenb T, Ali-Koveroa H, Miettinenc O, Oinonenc P, Auvinend P, Paulind L, Larssone K-H, Mäkipää R (2010) Identifying wood-inhabiting fungi with 454 sequencing – what is the probability that BLAST gives the correct species? Fungal Ecology 3(4): 274–283. doi: 10.1016/j.funeco.2010.01.001

Porter TM, Skillman JE, Moncalvo J-M (2008) Fruiting body and soil rDNA sampling detects complementary assemblage of Agaricomycotina (Basidiomycota, Fungi) in a hemlock-dominated forest plot in southern Ontario. Molecular Ecology 17(13): 3037–3050. doi: 10.1111/j.1365-294X.2008.03813.x

Qiu X, Wu L, Huang H, McDonel PE, Palumbo AV, Tiedje JM, Zhou J (2001) Evaluation of PCR-generated chimeras, mutations, and heteroduplexes with 16S rRNA gene-based cloning. Applied and Environmental Microbiology 67(2): 880-887. doi: 10.1128/AEM.67.2.880-887.2001

Quince C, Lanzen A, Davenport RJ, Turnbaugh PJ (2011) Removing noise from pyrosequenced amplicons. BMC Bioinformatics 12: 38. doi: 10.1186/1471-2105-12-38

Rosling A, Cox F, Cruz-Martinez K, Ihrmark K, Grelet G-A, Lindahl BD, Menkis A, James TY (2011) Archaeorhizomycetes: Unearthing an ancient class of ubiquitous soil fungi. Science 333(6044): 876–879. doi: 10.1126/science.1206958

Ross HA, Murugan S, Li WLS (2008) Testing the reliability of genetic methods of species identification via simulation. Systematic Biology 57(2): 216–230. doi: 10.1080/10635150802032990

Sanderson MJ, Donoghue MJ, Piel W, Eriksson T (1994) TreeBASE: a prototype database of phylogenetic analyses and an interactive tool for browsing the phylogeny of life. American Journal of Botany 81(6): 183.

Schocha CL, Seifertb KA, Huhndorfc A, Robertd V, Spougea JL, Levesqueb CA, Chenb W, Fungal Barcoding Consortium (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. Proceedings of the National Academy of Sciences USA 109(16): 6241–6246. doi: 10.1073/pnas.1117018109

Stajich JE, Berbee ML, Blackwell M, Hibbett DS, James TY, Spatafora JW, Taylor JW (2009) The fungi. Current Biology 19(18): R840–855. doi: 10.1016/j.cub.2009.07.004

Swofford DL (2003) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, Fisher MC (2000) Phylogenetic species recognition and species concepts in fungi. Fungal Genetics and Biology 31(1): 21–32. doi: 10.1006/fgbi.2000.1228

Taylor DL, McCormick MK (2008) Internal transcribed spacer primers and sequences for improved characterization of basidiomycetous orchid mycorrhizas. New Phytologist 177(4): 1020–1033. doi: 10.1111/j.1469-8137.2007.02320.x

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I, Bahram M, Bechem E, Chuyong G, Kõljalg U (2010) 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. New Phytologist 188(1): 291–301. doi: 10.1111/j.1469-8137.2010.03373.x

Tedersoo L, Abarenkov K, Nilsson RH, Arthur Schüssler, Grelet G-A, Kohout P, Oja J, Bonito GM, Veldre V, Jairus T, Ryberg M, Larsson K-H, Kõljalg U (2011) Tidying up International Nucleotide Sequence Databases: ecological, geographical, and sequence quality annotation of ITS sequences of mycorrhizal fungi. PLoS ONE 6: e24940. doi: 10.1371/journal.pone.0024940

Vrålstad T (2011) ITS, OTUs and beyond—fungal hyperdiversity calls for supplementary solutions. Molecular Ecology 20(14): 2873–2875. doi: 10.1111/j.1365-294X.2011.05149.x

Wang GC, Wang Y (1996) The frequency of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from different bacterial species. Microbiology 142(5): 1107–1114. doi: 10.1099/13500872-142-5-1107

Wren JD (2008) URL decay in MEDLINE—a 4-year follow-up study. Bioinformatics 24(11): 1381–1385. doi: 10.1093/bioinformatics/btn127

Appendix

Technical considerations. (doi: 10.3897/mycokeys.4.3606.app) File format: PDF.

Explanation note: Discussion on sequence quality and reliability assessment for the more technically inclined user.

Citation: Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JAA, Bergsten J, Porter TM, Jumpponen A, Vaishampayan P, Ovaskainen O, Hallenberg N, Bengtsson-Palme J, Eriksson KM, Larsson K-H, Larsson E, Kõljalg U (2012) Five simple guidelines for establishing basic authenticity and reliability of newly generated fungal ITS sequences. MycoKeys 4: 37–63. doi: 10.3897/mycokeys.4.3606.app