RESEARCH ARTICLE

Genomics and metagenomics technologies to recover ribosomal DNA and single-copy genes from old fruit-body and ectomycorrhiza specimens

Leho Tedersoo[1], Ingrid Liiv[2], Paula Ann Kivistik[3], Sten Anslan[2], Urmas Kõljalg[2], Mohammad Bahram[2,4]

1 *Natural History Museum, University of Tartu, 14A Ravila, 51005 Tartu, Estonia* **2** *Institute of Ecology and Earth Sciences, University of Tartu, 14A Ravila, 51005 Tartu, Estonia* **3** *Estonian Genome Center, University of Tartu, Riia 23b, 51010 Tartu, Estonia* **4** *Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Uppsala, Sweden*

Corresponding author: *Leho Tedersoo* (leho.tedersoo@ut.ee)

Abstract

High-throughput sequencing (HTS) has become a standard technique for genomics, metagenomics and taxonomy, but these analyses typically require large amounts of high-quality DNA that is difficult to obtain from uncultivable organisms including fungi with no living culture or fruit-body representatives. By using 1 ng DNA and low coverage Illumina HiSeq HTS, we evaluated the usefulness of genomics and metagenomics tools to recover fungal barcoding genes from old and problematic specimens of fruit-bodies and ectomycorrhizal (EcM) root tips. Ribosomal DNA and single-copy genes were successfully recovered from both fruit-body and EcM specimens typically <10 years old (maximum, 17 years). Samples with maximum obtained DNA concentration <0.2 ng μl-1 were sequenced poorly. Fungal rDNA molecules assembled from complex mock community and soil revealed a large proportion of chimeras and artefactual consensus sequences of closely related taxa. Genomics and metagenomics tools enable recovery of fungal genomes from very low initial amounts of DNA from fruit-bodies and ectomycorrhizas, but these genomes include a large proportion of prokaryote and other eukaryote DNA. Nonetheless, the recovered scaffolds provide an important source for phylogenetic and phylogenomic analyses and mining of functional genes.

Key words

Fungal fruit-bodies, low-coverage genome reconstruction, metagenome analysis, functional gene mining, Illumina HiSeq

Introduction

DNA sequences of high quality are essential for precise molecular identification of organisms and construction of phylogenies. For these purposes, inclusion of type specimens of the species is of utmost importance, because they carry taxonomic information and anchor the target species amongst potentially multiple cryptic taxa (Federhen 2014). However, type specimens of most taxa are decades or even centuries old and their DNA is often poorly preserved due to unsuitable storage conditions such as high humidity and temperature, insufficient care, etc. Therefore, extraction of high-quality DNA as well as amplification and sequencing from old material is painstaking and often virtually impossible (Pääbo et al. 2004).

Fungi represent one of the most diverse groups of eukaryotes with potentially millions of species and a high incidence of sympatric and allopatric cryptic species (Blackwell 2011). Both fruit-bodies and living cultures may serve as type specimens and form a basis for morphological, biochemical and molecular species recognition. However, the vast majority of fungi form no fruit-bodies and cannot be cultured with available techniques. Molecular identification methods have shed light into the high and undescribed fungal diversity in complex substrates such as roots, soil, sediments, water and foliage that are not represented by sequenced material from specimens (O'Brien et al. 2005; Jones et al. 2011; Tedersoo et al. 2014).

Other co-occurring organisms in voucher specimens may hamper molecular identification and genomic analyses of the target specimen. In living cultures, only endohyphal bacteria are common, but fruit-bodies are often infested with prokaryotes, protists, other fungi and meiofauna (nematodes, collembolans, Diptera larvae, etc.). Ectomycorrhizal (EcM) root tips and lesions on plant leaves are usually dominated by a single causal biotroph, although a vast diversity of microscopic organisms co-occurs (Tedersoo et al. 2009; Yoshida et al. 2013). Because of senescence and other resident taxa, certain fungal species and a substantial fraction (up to 5%) of distinct EcM morphotypes consistently remain unsequenced using the combination of fungal and universal primers and Sanger sequencing (Tedersoo et al. 2008; Nguyen et al. 2013).

DNA sequences from the nuclear ribosomal RNA cistron have been widely used, both for identification and phylogenetics of fungi due to a large number of copies and the level of conservation sufficient for discriminating between individuals (the intergenic spacer; IGS – Guidot et al. 1999), species (the internal transcribed spacer; ITS – Gardes et al. 1991; Kõljalg et al. 2005; Schoch et al. 2012) and higher taxa (large subunit; LSU and small subunit; SSU – Gueho et al. 1989). While the nuclear rDNA is distributed in tens to a few hundred tandem repeats (Baldrian et al. 2013), mitochondrial DNA is also abundant due to the presence of multiple mitochondria in active cells that render both targets easy to amplify and use for phylogenetics and identification purposes. Certain single-copy genes (SCGs) such as Translation Elongation Factor 1 α (TEF1) and RNA Polymerase II subunits (RPB1, RPB2) frequently serve to improve phylogenetic resolution, although their amplification and sequencing may require extra care (Schoch et al. 2012). The amplified size of these markers typically range from 300 to 1500 bases, although both LSU and phylogenetically informative single-copy genes are much longer.

The rationale for using such medium-size fragments is the ease of amplification and the ability of Sanger sequences to cover 1000 bases with high quality.

The rapid development of high-throughput sequencing (HTS) tools has greatly improved our understanding about the phylogeny, genome structure and functioning of fungi (Martin et al. 2008; Dentinger et al. 2016; Kohler et al. 2015). Although the HTS genomics approach (i.e., genome-wide sequencing of a single target organism) is commonly used on living cultures, it also enables to incorporate molecular data from herbarium collections of infected plant leaves and old specimens with degraded DNA (Staats et al. 2013; Yoshida et al. 2013; Dentinger et al. 2016). Single nucleotide polymorphisms (SNPs) and phylogenetically informative marker genes can be rigorously extracted from these genomes and used for phylogenetic reconstruction at the level of isolates to kingdoms (Liti et al. 2010; Capella-Gutierrez et al. 2012; Dentinger et al. 2016). These genomic studies have targeted >100-fold coverage that enables very high accuracy but restricts analysis to a few specimens in a single HTS run. As opposed to genomics, 'metagenomics' is a term for untargeted genome-wide sequencing of all organisms in a sample. This approach is mostly used to study the gene content of environmental samples, but sequencing at the depth of hundreds of millions of reads allows to separate nearly full genomes of the dominant prokaryote taxa (Wrighton et al. 2012).

Using Illumina HiSeq 2x150 paired-end sequencing technology, we evaluate the usefulness of low-coverage genomics and metagenomics analyses for recovering barcoding and other phylogenetically informative genes from voucher specimens of fruit-bodies and mycorrhizas in 85 samples simultaneously. In particular, we aimed to i) develop a protocol for genomics and metagenomics from minute amounts of material; ii) evaluate the possibility to obtain high-quality rDNA and SCG sequence data from old type specimens and root tips; and iii) explain why fruit-bodies and EcM root tips of certain taxa consistently fail to amplify and sequence. The ultimate purpose of this study is to extend the public record of high-quality DNA sequences from taxonomically valuable fruit-body voucher specimens and EcM fungal lineages.

Methods

Specimens

For genomics analysis, we selected 56 voucher specimens of fruit-bodies collected from all continents within the last 54 years (Table 1). These specimens are deposited in the fungaria of Tartu University (TU) and Estonian University of Life Sciences (TAA), with a few additional specimens representing loans from the Plant Pathology herbarium of New South Wales, Australia (DAR). We paid particular attention to cover i) old specimens including holotypes (category 'old': n=21; median age, 17.5 years; range, 10.2–53.5 years since the analysis in January, 2015), ii) species with minute-sized (apothecial Helotiales, sequestrate Endogonales) or corticioid (Thelephorales, Atheliales) fruit-bodies that are all inherently exposed to external contamination ('regular': n=19; median age, 5.6 years; range, 2.0–8.8 years, and iii) species that have consistently failed

Table 1. Fruit-body specimens used for genomic sequencing analysis.

Herbarium code	Identification, EcM lineage	Category	Collection date	Biosample
DAR69412	*Densospora nuda* (holotype)	Old	1989-08-19	SAMN04578188
DAR69419	*Densospora nanospora* (holotype)	Old	1989-08-31	SAMN04578189
DAR69421	*Densospora solicarpa* (holotype)	Old	1989-08-31	SAMN04578190
DAR69441	*Endogone magnospora* (holotype)	Old	1991-09-25	SAMN04578191
TAAM 042608	*Rutstroemia juglandis* (holotype)	Old	1961-xx-xx	SAMN04578222
TAAM 137803	*Sarconiptera vinacea* (holotype)	Old	2000-xx-xx	SAMN04578223
TAAM 159500	*Pseudotomentella atrofusca*	Old	1996-09-03	SAMN04578235
TAAM 166877	*Tomentella ferruginea*	Old	1997-08-18	SAMN04578245
TAAM 181146	*Bankera violascens*	Old	2001-09-25	SAMN04578233
TAAM 182408	*Larissia pyrola* (holotype)	Old	1980-xx-xx	SAMN04578220
TAAM 190020	*Arctomollisia kolymensis* (holotype)	Old	1975-xx-xx	SAMN04578221
TAAM 194916	*Lasiomollisia phalaridis* (holotype)	Old	2003-xx-xx	SAMN04578224
TU100021	*Pseudotomentella sp. nov.*	Old	2004-11-03	SAMN04578243
TU100364	*Odontia cf. fibrosa*	Regular	2006-08-04	SAMN04578228
TU100621	*Amaurodon mustialaensis*	Regular	2006-09-28	SAMN04578251
TU100663	*Sarcodon squamosus*	Regular	2006-10-06	SAMN04578240
TU105081	Thelephorales, *Fam. nov.*	Regular	2006-03-05	SAMN04578226
TU108047	*Pseudotomentella mucidula*	Regular	2008-08-27	SAMN04578242
TU108089	*Phellodon tomentosus*	Regular	2008-09-10	SAMN04578241
TU108144	*Tomentellopsis echinospora*	Regular	2008-09-27	SAMN04578250
TU108291	*Tomentella sp. nov.*	Regular	2009-05-01	SAMN04578247
TU108357	*Pseudotomentella armata, comb.ined*	Regular	2009-05-08	SAMN04578246
TU108377	*Thelephora terrestris*	Regular	2009-08-26	SAMN04578229
TU108482	Thelephorales, *Fam. nov.*	Regular	2010-03-17	SAMN04578248
TU110716	Ceratobasidiaceae, /ceratobasidium1	Regular	2011-12-06	SAMN04578167
TU110838	Thelephorales, *Fam. nov.*	Regular	2012-09-24	SAMN04578168
TU113361	*Endogone*	Unseq.[1]	2014-09-27	SAMN04578192
TU115221	Thelephorales, *Fam. nov.*	Regular	2009-10-19	SAMN04578249
TU115235	Thelephorales, *Fam. nov.*	Old	1997-06-12	SAMN04578230
TU115270	*Pseudotomentella italica, comb.ined.*	Regular	2008-08-09	SAMN04578244
TU115333	*Boletopsis leucomelaena*	Regular	2011-09-09	SAMN04578187
TU115426	Thelephorales, *Fam. nov.*	Regular	2012-08-28	SAMN04578172
TU116148	Atheliales; /atheliales1	Regular	2013-01-14	SAMN04578173
TU116208	*Cantharellus*	Unseq.	2013-07-15	SAMN04578174
TU116326	*Helvella*	Unseq.	2013-09-19	SAMN04578175
TU116380	*Helvella*	Unseq.	2013-10-13	SAMN04578176
TU116400	*Helvella*	Unseq.	2013-11-16	SAMN04578177
TU116448	Pezizaceae	Unseq.	2014-08-09	SAMN04578178
TU116491	*Helvella*	Unseq.	2014-08-11	SAMN04578169
TU116505	*Hydnum*	Unseq.	2014-08-11	SAMN04578179
TU116506	*Cantharellus*	Unseq.	2014-08-11	SAMN04578180
TU116517	*Helvella*	Unseq.	2014-08-11	SAMN04578181
TU116528	*Clavulina*	Unseq.	2014-08-12	SAMN04578182
TU116531	*Helvella*	Unseq.	2014-08-12	SAMN04578171
TU116607	*Coltricia*	Unseq.	2014-08-12	SAMN04578183
TU116615	*Helvella*	Unseq.	2014-08-12	SAMN04578171
TU116680	*Endogone*	Unseq.	2014-10-20	SAMN04578184
TU116699	*Glomus macrocarpum*	Unseq.	2014-10-21	SAMN04578185

Herbarium code	Identification, EcM lineage	Category	Collection date	Biosample
TU118650	*Hydnellum ferrugineum*	Regular	2012-08-28	SAMN04578186
TU115206	*Pseudotomentella humicola*	Old	1997-xx-xx	SAMN04578231
TU123535	*Lenzitopsis oxycedri*	Old	1991-04-26	SAMN04578232
TU100990	*Tomentella subamyloidea* (isotype)	Old	1999-08-24	SAMN04578234
FP133500	*Pseudotomentella fumosa* (holotype)	Old	1972-11-16	SAMN04578236
FP133849	*Pseudotomentella molybdea* (holotype)	Old	1974-11-06	SAMN04578237
FP134609	*Pseudotomentella kaniksuensis* (holotype)	Old	1981-07-23	SAMN04578238
SSMF695-4961	*Pseudotomentella griseopergamacea* (holotype)	Old	1961-10-21	SAMN04578239

[1]Unseq., unsequenced

to amplify or sequence in spite of using different primers and targeting different rDNA regions ('unsequenced': n=16; median age, 0.4 years; range, 0.2–1.5 years; recent collections were used to rule out potentially confounding storage effects). Notably, the 'old' specimens were comprised mainly of Thelephorales, Helotiales and Endogonales, whereas the 'unsequenced' taxa included mostly Pezizales (including *Helvella* spp.) and Cantharellales (including *Cantharellus* spp.) Within the last 10 years, the DNA of these samples has been extracted from 0.05–10 mg fresh or dried material following one of the five protocols outlined in Suppl. material 1.

For metagenomics approach, we selected 29 vouchered EcM root tip specimens from TU-linked collections of L. Tedersoo and M. Bahram (Table 2). These specimens included either i) rarely occurring EcM fungal lineages not represented by fruitbodies or living cultures (cf. Tedersoo and Smith 2013; n=17), or ii) distinct morphotypes that have remained unamplified and unsequenced in spite of multiple attempts and varying primers (n=12). Samples from the latter category primarily originate from Australia (collected in Tasmania in August, 2006; Tedersoo et al. 2008) and Estonia (collected from various hosts and habitats from May to September, 2013; L. Tedersoo, unpublished). The age of EcM samples ranged from 1.2 to 9.5 years (median, 4.7 years). The DNA of EcM root tips was extracted from fresh or CTAB-stored (100 mM Tris-HCI (pH 8.0), 1.4 M NaCl, 20 mM EDTA, 2% cetyltrimethylammonium bromide) material following one of the four protocols given in Suppl. material 1. In addition, we included two composite samples of soil (AV116 and S160; cf. Tedersoo et al. 2014) and a mock community comprised of 24 fruit-body specimens representing different species (cf. Tedersoo et al. 2015) as controls and for evaluating sequence assembly from more complex samples. Negative controls were not included for sequencing, because of DNA concentration below the detection level.

Molecular techniques

The DNA concentration of all samples was measured using Qubit dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA) and Qubit 2.0 Fluorometer (Invitrogen, Carlsbad, CA, USA) in January, 2015. Since the DNA concentration of most samples was <1 ng μl^{-1}, the DNA (300 μl) was concentrated up to three times using 750 μl

Table 2. Ectomycorrhiza specimens used for metagenomic sequencing analysis.

Sample code	Identification and EcM lineage	Category	Collection date	Biosample
IO577	Tulasnellaceae, /tulasnella1	Rare	2010-06-xx	SAMN04578193
KP016	Serendipitaceae, /serendipita1	Rare	2011-07-xx	SAMN04578194
L3043d	*Sebacina*[1]	Unseq.	2006-08-xx	SAMN04578195
L3078g	Tulasnellaceae, /tulasnella2	Rare	2006-08-xx	SAMN04578196
L3136g	unidentified	Unseq.[2]	2006-08-xx	SAMN04578197
L3161g	*Discinella*[1]	Unseq.	2006-08-xx	SAMN04578198
L3185g	*Inocybe*[1]	Unseq.	2006-08-xx	SAMN04578199
L3196a	*Discinella*[1]	Unseq.	2006-08-xx	SAMN04578200
L3196g	*Discinella*[1]	Unseq.	2006-08-xx	SAMN04578201
L3273b	Helotiales, /helotiales5	Rare	2006-08-xx	SAMN04578202
L3289	Helotiales, /helotiales4	Rare	2006-08-xx	SAMN04578203
L3371b	Helotiales, /helotiales3	Rare	2006-08-xx	SAMN04578204
L3581g	Helotiales, /helotiales6	Rare	2006-12-xx	SAMN04578205
L3619g	Endogonales, /densospora	Rare	2006-12-xx	SAMN04578206
L7664	Sordariales, /sordariales1	Rare	2010-03-xx	SAMN04578207
L8253	Pyronemataceae, /pyronemataceae1	Rare	2010-07-xx	SAMN04578208
L8574J	*Tomentella*[1]	Unseq.	2013-05-16	SAMN04578209
L8601L	Pyronemataceae, /pyronemataceae2	Rare	2013-06-10	SAMN04578210
L8623J	*Helvella*[1]	Unseq.	2013-06-11	SAMN04578211
L874	Helotiales, /helotiales2	Rare	2005-07-xx	SAMN04578212
L8748B	Helotiales, /helotiales7	Rare	2013-07-03	SAMN04578213
L8760B	Sordariales, /sordariales2	Rare	2013-07-04	SAMN04578214
L8970d	*Tricholoma fulvum*[1]	Unseq.	2013-08-12	SAMN04578215
L9188J	*Tulasnella*[1]	Unseq.	2013-09-20	SAMN04578216
L9238J	*Fischerula macrospora*[1]	Unseq.	2013-09-22	SAMN04578217
L9302J	*Geopora*[1]	Unseq.	2013-10-08	SAMN04578218
N120	Ceratobasidiaceae, /ceratobasidium2	Rare	2008-09-xx	SAMN04578219
TRON3.1	Agaricomycetes, /agaricomycetes1	Rare	2012-04-xx	SAMN04578225
TS1000	Pyronemataceae, /genea-humaria	Rare	2006-08-xx	SAMN04578227

[1]Identification based on ITS sequence from the metagenome.
[2]Unseq., unsequenced.

96% ethanol, 2 µl Pellet Paint Co-Precipitant (cat no 69049–3; Novagen, Madison, WI, USA) and sodium acetate (0.3 M, pH 5.2). DNA precipitation was performed overight at -20 °C. The pellets were washed once with 75% ethanol (-20 °C) and dissolved into MilliQ water, followed by re-determination of the concentration. The obtained 'maximum concentration' ranged from 0.05 to 8.13 ng µl^{-1} (median, 0.57 ng µl^{-1}). All samples were diluted to the concentration of 0.2 ng µl^{-1} (if below, the maximum concentration was used) and 1 ng of DNA was used as an input to prepare sequencing libraries with Nextera XT kit (Illumina Inc., San Diego, CA, USA) according to the instructions of the manufacturer. The concentration of the libraries was measured with Qubit fluorometer and the libraries were pooled equimolarly. The library pools were concentrated with vacuum evaporation and then the library pools were validated by

TapeStation analysis (Agilent Technologies, Santa Clara, USA) and qPCR with Kapa Library Quantification Kit (Kapa Biosystems, Wilmington, MA, USA) in order to optimize cluster generation. From each library, 22 pg or 54 pg (dilute samples) of DNA was used in the cluster generation and sequenced on the HiSeq2500 rapid flowcell (Illumina Inc.) with 150 bp paired-end reads protocol.

Bioinformatics

The metagenomics reads of individual samples were demultiplexed and quality-filtered using sdm script of the Lotus pipeline (Hildebrand et al. 2014) with the following options: minAvgQuality=27; maxAmbiguousNT=0; maxHomonucleotide=15; QualWindowWidth=30; QualWindowThreshold=0; TrimWindowWidth=15; TrimWindowThreshold=20. The quality-passed reads were assembled in SPAdes (Bankevich et al. 2012) using default options and kmer sizes 27, 33, 55 and 71. We sought to target the phylogenetically informative genes that are in multiple (nuclear rDNA) or single copies (mitochondrial rDNA, RPB1, RPB2 and TEF1) in the genome. The program sortMeRNA (Kopylova et al. 2012) was used to extract rDNA from raw reads, which were assembled into scaffolds in SPAdes. The fragments were subsequently subjected to bulk blastN search against the entire International Nucleotide Sequence Databases consortium (INSDc) to manually inspect the closest matches focusing on scaffolds of 500–12,000 bases. The coverage of the genomes was esitimated using the Core Eukaryotic Mapping Genes Approach (CEGMA), which gives a genome completeness percentage based on partial and full-length alignments of the target genome with 242 core eukaryotic genes (Parra et al. 2007).

The reference database for genomic and metagenomic fragments comprised 46 fungal genomes and 30 bacterial genomes (present in samples according to rDNA analysis). For the selected SCGs, we used a reference data set of James et al. (2006). Scaffolds containing SCGs were double-checked with manual blastN searches against INSDc and downloaded for trimming and quality evaluation. The sequences of confirmed rDNA genes and SCGs were subjected to multiple sequence alignment using MAFFT 7 (Katoh and Standley 2013) along with 2–5 full-length sequences of the respective genes from Ascomycota and Basidiomycota downloaded from INSDc. The alignments were inspected in SeaView 4 (Gouy et al. 2010) and the flanking non-coding regions were removed. Due to the multiple introns and poor alignability, ca 50–100 bases of flanking regions were retained. For rDNA, we retained the entire copy usually comprising of partial Intergenic Spacer (IGS) 2, SSU, ITS1, 5.8S, LSU and partial IGS1. Due to the poor alignability and multiple introns, mitochondrial rDNA was not trimmed. In many cases, both rDNA and SCGs comprised several different copies that were all kept and submitted to the UNITE database (Abarenkov et al. 2010; accessions UDB028495-UDB028830) and INSDc. The genomic and metagenomic scaffolds of fruit-bodies and root tips were submitted to the Short Read Archive (SRA) of INSDc (Bioproject PRJNA308809; biosample accessions SAMN04578167-SAMN04578254).

Statistics

To evaluate the relative performance of genomics and metagenomics approaches for recovering genetic information of fungi from root tip and fruit-body material of different quality, we constructed linear regression and ANOVA models. First, we tested the effects of the maximum DNA concentration, age of specimen and age of DNA as well as DNA extraction method on the number of reads, size of all scaffolds (confirmed fungal and total and proportion of known fungal) and the longest scaffolds representing rDNA by use of general linear models and forward selection of variables as implemented in Statistica (Statsoft Inc., Tulsa, OK, USA). We determined Pearson correlations among the recovered length of ribosomal and mitochondrial rDNA and SCGs. Further, we arbitrarily chose a threshold of 1500 bases as a criterion for 'successful' sequencing of a barcode, because this value roughly corresponds to the size of mitochondrial SSU and LSU, nuclear SSU and the fragment of commonly amplified nuclear LSU (primers ITS3 and LR5 or LR0R and LR7) as well as SCGs. Differences in sequencing success among markers, sample material (fruit-body vs EcM) and fruit-body type ('old', 'regular' and 'unsequenced', see above) were tested using a series of Fisher's exact tests.

To shed light on the potential issues with DNA secondary structure on amplification and sequencing success in Sanger sequencing, we calculated the minimum free energy (MFE) of the secondary structure of ITS1 and ITS2 reads using RNAstructure (default options for DNA; Reuter and Mathews 2010). The MFE provides an approximation for the stability of a given structure, with lower MFE values indicating more stable structures.

Results

Recovery of genomes

DNA extraction methods yielded similar DNA content and concentration that usually required further concentrating efforts given the small size of our samples. Compared with other methods, the simple ammonium sulphate lysis (cf. Anslan and Tedersoo 2015) retained large amounts of polysaccharides that co-precipitated with DNA, but did not interfere with ligation and sequencing. The HiSeq run produced 553,982,778 individual reads (average length, 145.4 bases), of which 86.7% passed initial quality filtering. Individual genomes and metagenomes were covered by 1,366 to 21,288,678 (median, 5,780,997; SD, 3,863,989) reads with no differences among sample types or fruit-body categories. However, specimen age had a significant negative effect on the recovery of reads in fruit-bodies ($F_{1,54}$=4.4; R^2=0.076; P=0.040) but not in EcM root tips (P>0.1; Fig. 1). The time of DNA isolation had no further impact. The total length of all genomic and metagenomic scaffolds averaged 5.5×10^7 bases (SD, 5.6×10^7) across all samples. It was positively related to the maximum DNA concentration
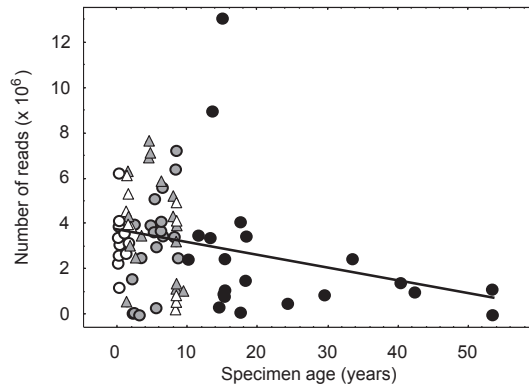
Figure 1. Effect of specimen age on the recovery of reads in the Illumina HiSeq run. Closed circles, 'old' fruit-bodies; shaded circles, 'regular' fruit-bodies; open circles, 'unsequenced' fruit-bodies; shaded triangles, ectomycorrhizal root tips representing unique rare lineages; open triangles, 'unsequenced' ectomycorrhizal root tips.
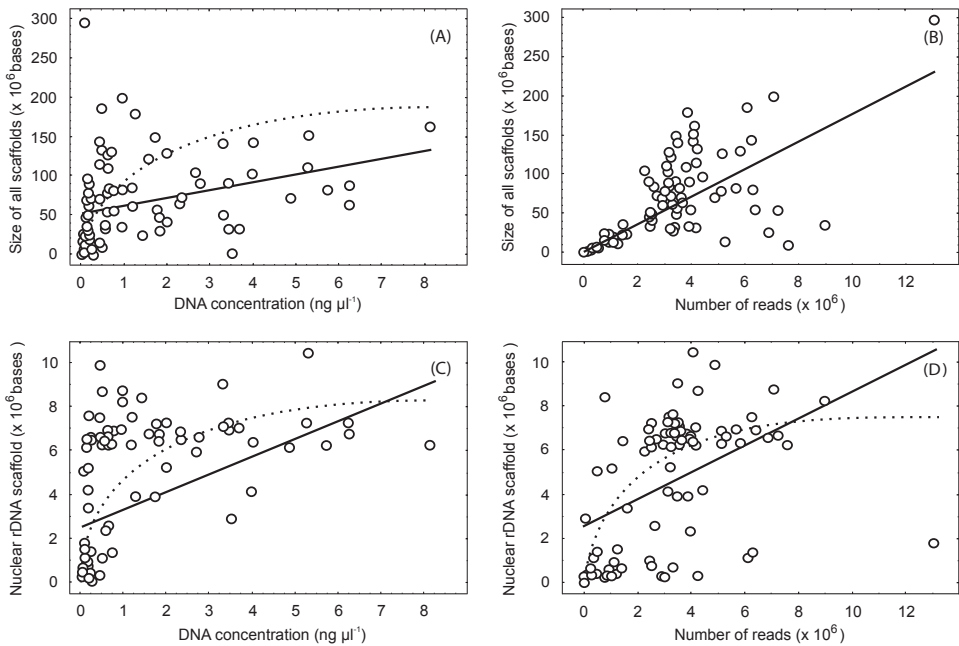
Figure 2. Impact of maximum obtained DNA concentration and number of Illumina HiSeq reads on the size of all scaffolds (**A, B**) and largest nuclear rDNA scaffold (**C, D**). Regular straight lines and dotted lines indicate linear and better fitting logarithmic relationships, respectively.

obtained (partial effect: $F_{1,82}$=10.8; R^2=0.070; P=0.001; Fig. 2A) and total number of reads ($F_{1,82}$=62.1; R^2=0.401; P<0.001; Fig. 2B).

The proportion of genomic and metagenomic sequences belonging strictly to fungi varied greatly across samples, being on average three times lower for EcM root tip

(median, 1.5%; SD, 3.7) compared with fruit-body (median, 4.5%; SD, 15.9) samples ($F_{1,82}$=10.8; R^2=0.098; P=0.001). The lack of closely related reference genomes clearly hampered unequivocal assignment of genomic fragments to fungi or other organisms. Of these, bacteria were the most common organisms in fruit-bodies and EcM root tips, whereas plant scaffolds strongly contributed to the EcM-derived metagenome. However, plant contribution was difficult to establish, because of the large size and ample non-coding regions in plant genomes. Samples of old fruit-bodies and particularly EcM root tips included multiple co-inhabiting fungal species. Their coverage was distinctly lower than that of the target species, but unambiguous separation of these satellite taxa was more difficult for relatively fragmented genomes.

Ribosomal DNA and single copy genes

The coverage of nuclear and mitochondrial DNA and SCGs and their ratio varied greatly across samples independent of sample origin (fruit-body vs. EcM) and category (Suppl. material 1). For the 38 most comprehensively sequenced samples, the standardized ratio of median coverage of nuclear rDNA to mitochondrial rDNA to SCGs was 17.0/4.2/1.0. Notably, *Glomus macrocarpum* (TU116699) and *Endogone* sp. (TU113361) exhibited the corresponding ratios of 1.6/3.6/1.0 and 2.0/2.7/1.0, respectively, indicating low amount of nuclear and mitochondrial rDNA relative to SCGs in their 'fruit-bodies' that are comprised of multinucleate hyphal structures and chlamydospores. In contrast, *Hydnum* sp. (TU116505) and *Cantharellus* sp. (TU116208) stood out as specimens with the highest nuclear rDNA to SCG ratio (43.4/1.0 and 37.6/1.0, respectively).

Among the target regions of fruit-body and EcM samples, nuclear rDNA was relatively more efficiently recovered compared with mitochondrial rDNA and both of these were sequenced with greater success than SCGs (P<0.01 in all cases). There was no difference in the recovery rate among individual SCGs (P>0.5), although RPB1 was completely missing in two samples (ectomycorrhiza of the /genea-humaria lineage TS1000 and *Sarcodon squamosus* TU100663) that exhibited nearly full-length recovery of other SCGs and rDNA. Apart from other taxa, most specimens belonging to Thelephorales contained two highly divergent copies of the TEF1 gene.

The SCGs were significantly less efficiently recovered in EcM samples compared with fruit-body samples (by a factor of 1.9 to 6.2; P<0.001), but the recovery of nuclear and mitochondrial rDNA was comparable between sample types (P>0.1). Across all samples, the maximum DNA concentration (partial effect: $F_{1,82}$=26.7; R^2=0.201; P<0.001; Fig. 2C) and the total number of reads ($F_{1,82}$=23.9; R^2=0.180; P<0.001; Fig. 2D) positively affected the length of the largest nuclear rDNA scaffold. The number of reads necessary to yield full-length rDNA became rapidly saturated at the depth of $2 \times 10^6 - 5 \times 10^6$ sequences for samples with maximum DNA concentration > 0.2 ng µl^{-1} (Fig. 2D).

Fruit-body samples

Within fruit-body collections, rDNA and SCGs were better recovered from 'regular' and recent 'unsequenced' collections than 'old' material (Suppl. material 1). Across all samples, the length of largest scaffolds of nuclear rDNA was strongly correlated to that of mitochondrial rDNA (R=0.724; P<0.001) but not SCGs (P>0.05). The length of largest scaffolds was correlated among all SCGs (0.584<R<0.649; P<0.001).

Fruit-body samples displayed great variation in genomic sequencing success. The 'old' samples sequenced most poorly - i.e., nuclear rDNA >1500 bases could be retrieved only for 43.0% of specimens, which is significantly less compared with 'unsequenced' (73.3%) and 'regular' (84.2%) specimens (P<0.01). Mitochondrial rDNA and SCGs were also relatively poorly recovered in 'old' collections, although the differences were less pronounced among the categories (0.01<P<0.15).

The HTS approach highlighted that primer bias and atypically long ITS markers may account for the Sanger sequencing problems in 'unsequenced' fruit-body samples. In particular, several *Helvella* spp. and *Cantharellus* spp. exhibited ITS1 markers of 500-600 bases that exceed the average values three-fold (Tedersoo et al. 2015). In addition, most *Cantharellus* spp. displayed a 3' terminal mismatch or several mismatches to 'universal' and 'fungal' primers (ITSOF, ITS3, ITS4, LR0R). Besides the regular rDNA copy, *Endogone* sp. (TU116680) exhibited two additional copies that were only 87.2% and 77.5% similar in the ITS region and displayed multiple indels and substitutions in the flanking 5.8S and LSU regions including the highly conserved parts. *Lenzitopsis oxycedri* (UK146) possessed one such abnormal copy with 89.8% ITS sequence similarity, whereas *Glomus macrocarpum* (TU116699) had an extra rDNA copy with 96.0% ITS similarity but no mutations in the flanking 5.8S and LSU fragments. These extra copies had 1.9-2.6 times less coverage than the corresponding regular copies, except that of *L. oxycedri* (54.9-fold difference). The potential problems with sequencing *Clavulina* sp. (TU116528) and *Hydnum* sp. (TU116505) could not be tackled, although the former specimen was 'contaminated' by the DNA of Diptera larvae and a chytrid. The secondary structure of recovered ITS1 and ITS2 sequences had a similar minimum free energy (MFE) and MFE per base in the 'unsequenced' and other categories (Suppl. material 1).

While most collections of stipitate fruit-bodies were relatively free from co-colonization by other fungi, specimens of *Helvella* and those with hypogeous and resupinate fruit-bodies were commonly inhabited by multiple putatively saprotrophic or mycoparasitic fungal taxa. Of these, *Tulasnella*, *Rhizoctonia* (syn. *Ceratobasidium*) and unidentified genera of Eurotiales and Sordariales were the most common. Their nuclear rDNA scaffolds were of relatively lower coverage even if the sequences were nearly full-length. Similar patterns but notably shorter satellite sequences were evident in mitochondrial rDNA (up to, 2000 bases) and SCGs (up to 500 bases).

EcM root tip samples

There were no differences in rDNA and SCG recovery among EcM root tip samples that failed determination previously and those representing rare lineages. Out of 12 previously unidentified EcM root tip samples, only one (L3136g) remained further without identification due to low maximum DNA concentration (0.07 ng/µl) and hence low number of retrieved sequences (173,554 reads). Based on the ITS region, the Tasmanian sequences were identified as *Sebacina* sp. (L3043d), *Inocybe australiensis* (L3185g), and *Discinella* sp. (/helotiales4 lineage; L3161g, L3196a, L3196g). The Estonian sequences were identified as *Tomentella* sp. (L8574J), *Helvella* sp. (L8623J), *Geopora* sp. (L9302J), *Tulasnella* sp. (L9188J), *Tricholoma fulvum* (L8970d) and *Fischerula macrospora* (L9238J) based on the full or partial ITS sequences (Suppl. material 1). The DNA of most EcM samples was apparently degraded, because no primer mismatches, excessively long barcodes, paralogues or deviations in the minimum free energy were evident. Only the *Tulasnella* sp. sample (L9188J) exhibited two mismatches to the ITS3 primer and members of the /helotiales4 lineage had ca. 500-base intron between the ITSOF and ITS1 primer sites.

Using the metagenomics approach, three out of 17 EcM root tips with successful Sanger sequences (L848, L8601, L8760b) failed to retrieve high-quality nuclear rDNA sequences >1500 bases. A single EcM fungus always dominated in nuclear rDNA, but the samples were often co-inhabited by a myriad of ascomycetes, in particular Helotiales, Sordariales, Hypocreales and Dothideales. Basidiomycetes were less common, although *Tulasnella*, Ceratobasidiaceae and Tremellales (*Cryptococcus*) occurred in multiple samples. The ratio of plant to fungal nuclear rDNA varied nearly 80-fold, ranging from 0.21 to 16.3 (median, 1.76) with no apparent differences among host taxa.

Across all 29 EcM root tip metagenomes, fungal TEF1, RPB1 and RPB2 scaffolds >1500 bases were successfully obtained for two, five and fourteen samples, respectively. For 13 samples, none of these SCGs were recovered (scaffolds <500 bases). In successfully sequenced EcM samples, individual SCGs typically occurred in several scaffolds located tens to a few hundred bases apart based on mapping to the alignment. BlastN searches against INSDc and comparisons with rDNA revealed that the largest scaffolds obviously belong to the targeted mycobiont. The co-occurrence of other fungi rendered the taxonomic assignment of SCG scaffolds ambiguous.

Soil and mock community samples

The two highly complex soil metagenomes comprised altogether four fungal nuclear rDNA scaffolds >500 bases in size, three of which were obvious chimeras. The mock community sample included 25 scaffolds encompassing ITS or any of the nuclear rDNA genes (>500 bases). Comparisons with respective Sanger sequences revealed that 32% of these sequences were chimeric, some of which comprising >2 parents. Two of the chimeric sequences were 'circular', i.e. comprised of a full-length rDNA

and fragments of another taxon in one of the ends. Most of the chimeric breaks were located in the conserved regions of 3' half of the SSU and 5' end of LSU, but none were evident in the 5.8S rRNA gene. SSU and LSU of certain congeneric taxa (*Lyophyllum* spp., *Tomentella* spp.) were represented by a consensus sequence that matched perfectly to none of the ingredient specimens. In scaffolds with lower coverage, 5' or 3' ends were sometimes highly diverged from the corresponding Sanger sequence or any database sequences, indicating that artefactual sequences are, to some extent, generated by metagenomics methods.

Discussion

Genomic fragments

We recovered partial fungal genomes and metagenomes from <1 ng DNA of fruit-body and EcM root tip samples with variable success, depending on specimen age and DNA quality (see below). This indicates that fungal genomes can be sequenced from minute amounts of DNA if sufficient quality is secured. The current genome sequencing protocols in the 1000 Fungal Genomes project require three to four orders of magnitude more DNA (http://genome.jgi.doe.gov/programs/fungi/1000fungalgenomes.jsf) that cannot be obtained from tiny samples. In comparison, the genomes of prokaryotes are on average ten times smaller and these have been successfully recovered from common species (upwards 1% relative abundance) in the complex environmental material (Wrighton et al. 2012), multiple single cells (Rodrigue et al. 2009), and high-quality starting material of <0.01 ng DNA (Adey et al. 2010). Because of a single DNA molecule and low proportion of repeats and other non-coding regions, bacterial genomes are easier to assemble compared with eukaryotes that tend to possess long non-coding regions, multiple chromosomes and usually one or two organelles. In our study, taxonomic affinity of especially short scaffolds remained undetermined at the kingdom level based on *de novo* assembly. The paucity of closely related fungal reference material, multiple co-inhabiting organisms and moderate sequencing depth complicated scaffold assembly and rendered estimates of genome size and coverage unreliable (not shown).

Our study aimed to recover the most important genetic markers used for barcoding and phylogenetic reconstruction. Nuclear and mitochondrial rDNA sequences were successfully recovered from most fresh and high-quality samples but typically not from fruit-body specimens >10 years old. For these old specimens, the maximum obtained DNA concentration, a proxy for DNA quality and quantity, remained <0.2 ng/µl. Although other DNA samples were further diluted to this level for library preparation, barcoding markers could not be usually obtained from samples with 0.05-0.2 ng/µl maximum DNA concentration. Because Nextera approach uses DNA fragmentation and 12 cycles of PCR in the ligation step ('tagmentation'), the short DNA molecules of degraded material (Allentoft et al. 2012) may have become over-fragmented or poorly amplified and thus lost from further analytical procedures. This speculation is sup-

ported by 2.3-fold lower yield of reads and 2.8-fold lower proportion of known fungi in 'old' samples compared with 'regular' and 'unsequenced' samples taken together. An 18-year old specimen of *Tomentella ferruginea* (TAAM 166877) represented the oldest collection that was successfully sequenced for the full-length of all rDNA genes and SCGs. In comparison, Staats et al. (2013) successfully sequenced the genome of *Pleurotus ostreatus* fruit-body specimen collected in 1931 by taking advantage of 8000-fold greater amount of DNA and relatively clean vegetative material from the interior of a sporocarp. Old fruit-body samples with large initial amounts of degraded DNA can be prepared for Illumina sequencing using fragmentation-free ligation methods (Carpenter et al. 2013).

Across all samples, nuclear and mitochondrial rDNA were more efficiently recovered compared with SCGs, which reflects the results from amplicon sequencing (Schoch et al. 2012) and scaffold coverage. The range of sequence coverage ratio of nuclear rDNA to SCGs (1.6 to 43.4) is somewhat lower than the previously reported rDNA copy numbers based on qPCR (range, 20 to 200; reviewed in Baldrian et al. 2013). Our indirect estimates should be viewed with caution, because the coverage ratio is based on only 2-3 SCGs and does not account for the AT/GC bias (Perisin et al. 2016). The relative amount of mitochondrial DNA certainly depends on the metabolic activity of a fungus, potentially varying between living cultures, fruit-bodies, EcM root tips and natural mycelium. Taken together, our analyses indicate that fungal species exhibit marked differences in the relative amount of nuclear and mitochondrial rDNA that may further affect metabarcoding- and metagenomics-based estimates of diversity. These results explain the relatively low abundance of Glomeromycota in the soil nuclear rDNA pool (Saks et al. 2014; Tedersoo et al. 2014) and support utilization of SCGs as additional barcodes (e.g. Stockinger et al. 2014).

We sought to uncover the causes why certain fungal species and EcM morphotypes have remained unidentified using direct Sanger sequencing of amplicons. We showed that EcM root tip DNA was degraded and/or comprised of multiple fungal species, which may have disabled direct Sanger sequencing. In fruit-body samples, excessive length of ITS1 sequence might have caused low amplification success in several *Cantharellus* spp. and *Helvella* spp. Due to rapid evolution of rDNA genes in *Cantharellus* (Moncalvo et al. 2006), several otherwise conserved primer sites had one or more mismatches to the templates in the commonly used fungal or eukaryote primers. Furthermore, *Lenzitopsis oxycedri*, *Endogone* sp. and *Glomus macrocarpum* possessed several divergent copies of rDNA that is previously known for a small group of Glomeraceae (Stockinger et al. 2010) and is attributed to the multinucleate habit in that group. Potential ITS paralogues with multiple mutations in the conserved region were evident for the two former species, confirming previous implications based on Sanger sequencing of cloned amplicons (Simon and Weiss 2008) and 454 pyrosequencing of amplicons (Tedersoo et al. 2010; Lindner et al. 2013).

Ribosomal DNA scaffolds from soil and mock community metagenomes indicated artificial generation of a high proportion of chimeric scaffolds during DNA assembly. This demonstrates that markers with long conserved regions such as nuclear

rDNA cannot be reliably assembled even in simple fungal communities. Furthermore, artificial consensus sequences were generated for closely related species with nearly identical SSU and LSU. While such artefacts can be relatively easily tracked in mock communities, metagenomic assembly of rDNA is particularly problematic for natural samples from more complex substrates that comprise hundreds to thousands of fungal species. Due to short scaffolds and the paucity of reference data, we cannot estimate the reliability of scaffold assembly in mitochondrial genes and SCGs, but this may be more problematic with closely related species. Such assembly problems are considered of minor importance in prokaryote metagenomes (Wrighton et al. 2012; Parks et al. 2015) because of a single circular chromosome, lower proportion of repeats, more rapid evolution and more relaxed definition of species/OTUs at 97% SSU similarity (Mende et al. 2013).

Conclusions and perspectives

Taxonomically informative rDNA genes and SCGs can be sequenced from <1 ng DNA of fruit-body and EcM root tip specimens using genomics and metagenomics approaches, respectively. However, fruit-body specimens >10 years old need specific care for obtaining high-quality DNA or require fragmentation-free options for ligation. HTS methods also enabled us to recover large fragments of fungal genomes for a majority of EcM root tips and fruit-bodies that could not be sequenced using Sanger method or that represented unique (including type) material. For high-quality DNA samples, two million (meta)genomic reads were sufficient to recover the full-length nuclear rDNA. Recovery of SCGs was more unpredictable among samples, requiring roughly 10 million unpaired reads. This enables sequencing of ca. 50 fungal genomes on a single 2x150 paired-end Illumina HiSeq run at low coverage (5-10 ×; cf. Stajich 2014). As of January, 2016, a commercial Illumina HiSeq run (5.5×10^8 reads) cost between 5000 and 7000 EUR. However, all individual samples need to be separately ligated with a cost 50-100 EUR sample^{-1}. Thus, the cost per low-coverage fungal draft genome amounts ca. 150-250 EUR. We believe that such low-coverage genomics analyses represent a feasible option to generate multi-gene phylogenomic data sets for tens to hundreds of specimens or mining for the presence and diversity of certain gene families such as carbohydrate active enzymes (CAZymes), antibiotics resistance genes and unique metabolic pathways, but not for routine identification. Targeted enrichment using biotin-linked DNA/RNA probes enables even greater throughput and direct focus on selected markers (Carpenter et al. 2013; Moriarty Lemmon and Lemmon 2013; Manoharan et al. 2015). The full metagenome data also enable to construct draft genomes of prokaryotes and viruses associated with the fruit-body 'mycosphere' and soil 'mycorrhizosphere' that shed light on putative functions and metabolic pathways of these co-occurring microorganisms.

The currently available sequence length and error rate combination does not allow reliable large-scale assembly of genetic information of eukaryotes from complex communities using a single HTS platform. Besides tens and hundreds of millions of Illumina HiSeq

reads, metagenomics analyses would benefit from additional low-coverage sequence analysis of long (up to 3000 bases at 5-8 times circular coverage) fragments as routinely implemented by Pacific Biosciences for in-depth genomic reconstructions. Long amplicon-free backbone sequences reduce the incidence of chimeras and assembly artefacts. Combined with targeted marker capture, this approach would allow greater throughput of eukaryote target genes and more efficient utilization of phylogenetics tools in metabarcoding and community-level functional metagenomic analyses.

Author contribution

LT planned and designed research; UK provided material; IL and PAK performed laboratory analyses; MB, LT and SA analysed data; LT wrote the manuscript with others' input.

Acknowledgements

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 B, Vrålstad T, Liimatainen K, Peintner U, Kõljalg U (2010) The UNITE database for molecular identification of fungi – recent updates and future perspectives. New Phytologist 186: 281–285. doi: 10.1111/j.1469-8137.2009.03160.x

Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, Stackhouse M, MacKenzie AP, Caruccio NP, Zhang X, Shendure J (2010) Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. Genome Biology 11: R119. doi: 10.1186/gb-2010-11-12-r119

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert TP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, Bunce M (2012) The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B 279: 4724–4733. doi: 10.1098/rspb.2012.1745

Anslan S, Tedersoo L (2015) Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. European Journal of Soil Biology 69: 1–7. doi: 10.1016/j.ejsobi.2015.04.001

Baldrian P, Vetrovsky T, Cajthaml T, Dobiasova P, Petrankova M, Snajdr J, Eichlerova I (2013) Estimation of fungal biomass in forest litter and soil. Fungal Ecology 6: 1–11. doi: 10.1016/j.funeco.2012.10.002

Bankevich A, Nurk S, Antipov S, Gurevich AA, Dvorkin M, Kulikov AS (2012) SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455–477. doi: 10.1089/cmb.2012.0021

Blackwell M (2011) The Fungi: 1, 2, 3 … 5.1 million species? American Journal of Botany 98: 426–428. doi: 10.3732/ajb.1000298

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft M, Sikora M (2015) The American Journal of Human Genetics 93: 852–864. doi: 10.1016/j.ajhg.2013.10.002

Capella-Gutierrez S, Marcet-Houben M, Gabaldon T (2012) Phylogenomics supports microsporidia as the earliest diverging clade of sequenced fungi. BMC Biology 10: 47. doi: 10.1186/1741-7007-10-47

Dentinger BTM, Gaya E, O'Brien H, Suz LM, Lachlan R, Diaz-Valderrama JR, Koch RA, Aime MC (2016) Tales from the crypt: genome mining from fungarium specimens improves resolution of the mushroom tree of life. Biological Journal of the Linnean Society 117: 11–32. doi: 10.1111/bij.12553

Federhen S (2014) Type material in the NCBI Taxonomy Database. Nucleic Acids Research 2014: 1–13.

Gardes M, White TJ, Fortin JA, Bruns TD, Taylor JW (1991) Identification of indigenous and introduced symbiotic fungi in ectomycorrhizae by amplification of nuclear and ribosomal DNA. Canadian Journal of Botany 69: 180–190. doi: 10.1139/b91-026

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: 221–224. doi: 10.1093/molbev/msp259

Gueho EP, Kurtzman CP, Peterson SW (1989) Evolutionary affinities of heterobasidiomycetous yeasts estimated from 18S and 25S ribosomal RNA sequence divergence. Systematic and Applied Microbiology 12: 230–236. doi: 10.1016/S0723-2020(89)80067-0

Guidot A, Lumini E, Debaud J-C, Marmeisse R (1999) The nuclear ribosomal DNA intergenic spacer as a target sequence to study intraspecific diversity of the ectomycorrhizal basidiomycete Hebeloma cylindrosporum directly on Pinus root systems. Applied and Environmental Microbiology 65: 903–909.

Hildebrand F, Tadeo R, Voigt AY, Bork P, Raes J (2014) LotuS: an efficient and user-friendly OTU processing pipeline. Microbiome 2: 30. doi: 10.1186/2049-2618-2-30

James TY, Kauff F, Schoch CL (2006) Reconstructing the early evolution of fungi using a six-gene phylogeny. Nature 443: 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: 200–203. doi: 10.1038/nature09984

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30: 772–780. doi: 10.1093/molbev/mst010

Kohler A, Kuo A, Nagy LG, Morin E, Barry KW, Buscot F, Canbäck B, Tunlid A, Grigoriev IV, Hibbett DS, Martin F (2015) Convergent losses of decay mechanisms and rapid turnover of symbiosis genes in mycorrhizal mutualisms. Nature Genetics 47: 410–415. doi: 10.1038/ng.3223

Kopylova E, Noé L, Touzet H (2012) SortMeRNA: Fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28: 3211–3217. doi: 10.1093/bioinformatics/bts611

Kõljalg U, Larsson K-H, Abarenkov K, Nilsson RH, Alexander IJ, Eberhardt U (2005) UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. New Phytologist 166: 1063–1068. doi: 10.1111/j.1469-8137.2005.01376.x

Lindner DL, Carlsen T, Nilsson RH, Davey M, Schumacher T, Kauserud H (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecology and Evolution 3: 1751–1764. doi: 10.1002/ece3.586

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA (2010) Population genomics of domestic and wild yeasts. Nature 458: 337–341. doi: 10.1038/nature07743

Manoharan L, Kushwaha SK, Hedlund K, Ahren D (2015) Captured metagenomics: large-scale targeting of genes based on 'sequence capture' reveals functional diversity in soils. DNA Research 22(6): 451–460. doi: 10.1093/dnares/dsv026

Martin F, Aerts A, Ahren A, Brun A, Danchin EGJ, Dochaussoy F (2008) The genome of *Laccaria bicolor* provides insights into mycorrhizal symbiosis. Nature 452: 88–92. doi: 10.1038/nature06556

Mende DR, Sunagawa S, Zeller G, Bork P (2013) Accurate and universal delineation of prokaryotic species. Nature Methods 10: 881–884. doi: 10.1038/nmeth.2575

Moncalvo J-M, Nilsson RH, Koster B (2006) The cantherelloid clade: dealing with incongruent gene trees and phylogenetic reconstruction methods. Mycologia 98: 937–948. doi: 10.3852/mycologia.98.6.937

Moriarty Lemmon E, Lemmon AR (2013) High-throughput genomic data in systematics and phylogenetics. Annual Reviews in Ecology, Evolution and Systematics 44: 99–121. doi: 10.1146/annurev-ecolsys-110512-135822

Nguyen NH, Landeroz F, Garibay-Oriel R, Hansen K, Vellinga EC (2013) The *Helvella lacunosa* species complex in western North America: cryptic species, misapplied names and parasites. Mycologia 105: 1275–1286. doi: 10.3852/12-391

O'Brien HE, Parrent JL, Jackson JA, Moncalvo J-M, Vilgalys R (2005) Fungal community analysis by large-scale sequencing of environmental samples. Applied and Environmental Microbiology 71: 5544–5550. doi: 10.1128/AEM.71.9.5544-5550.2005

Parks DH, Imelfort M, Skannerton CT, Hugenholtz P, Tyson GW (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Research 25: 1043–1055. doi: 10.1101/gr.186072.114

Parra G, Bradnam K, Korf I (2007) CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061–1067. doi: 10.1093/bioinformatics/btm071

Perisin M, Vetter M, Gilbert JA, Bergelson J (2016) 16Stimator: statistical estimation of ribosomal gene copy numbers from draft genome assemblies. The ISME Journal 10: 1020–102. doi: 10.1038/ismej.2015.161

Pääbo S, Poinar H, Serre D, Jaenicke-Despres, Hebler J, Rohland N (2004) Genetic analyses from ancient DNA. Annual Reviews in Genetics 38: 645–679. doi: 10.1146/annurev.genet.37.110801.143214

Reuter JS, Mathews DH (2010) RNAstructure: software for RNA secondary structure prediction and analysis. BMC Bioinformatics 11: 129. doi: 10.1186/1471-2105-11-129

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR (2009) Whole Genome Amplification and De novo Assembly of Single Bacterial Cells. PLoS ONE 4: e6864. doi: 10.1371/journal.pone.0006864

Saks Ü, Davison J, Öpik M, Vasar M, Moora M, Zobel M (2014) Root-colonizing and soil-borne communities of arbuscular mycorrhizal fungi in a temperate forest understorey. Botany 92: 277–285. doi: 10.1139/cjb-2013-0058

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA (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: 6241–6246. doi: 10.1073/pnas.1117018109

Simon UK, Weiss M (2008) Intragenomic variation of fungal ribosomal genes is higher than previously thought. Molecular Biology and Evolution 25: 2251–2254. doi: 10.1093/molbev/msn188

Staats M, Erkens RHJ, van de Vossenberg B, Wieringa JJ, Kraijeveld K, Stielow B, Geml J, Richardson JE, Bakker FT (2013) Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. PLoS ONE 8: e69189. doi: 10.1371/journal.pone.0069189

Stajich J (2014) Phylogenomics enabling genome-based mycology. The Mycota 7B: 279–294.

Stockinger H, Krüger M, Schüssler A (2010) DNA barcoding of arbuscular mycorrhizal fungi. New Phytologist 187: 461–474. doi: 10.1111/j.1469-8137.2010.03262.x

Stockinger H, Peyret-Guzzon M, Koegel S, Bouffaud M-L, Redecker D (2014) The largest subunit of RNA polymerase II as a new marker gene to study assemblages of arbuscular mycorrhizal fungi in the field. PLoS ONE 9: e107783. doi: 10.1371/journal.pone.0107783

Tedersoo L, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson RH, Bork P, Hildebrand F, Abarenkov K (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys 10: 1–43. doi: 10.3897/mycokeys.10.4852

Tedersoo L, Bahram M, Põlme S (2014) Global diversity and geography of soil fungi. Science 234: 1078. doi: 10.1126/science.1256688

Tedersoo L, Jairus T, Horton BM, Abarenkov K, Suvi T, Saar I, Kõljalg U (2008) Strong host preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA barcoding and taxon-specific primers. New Phytologist 180: 479–490. doi: 10.1111/j.1469-8137.2008.02561.x

Tedersoo L, Pärtel K, Jairus T, Gates G, Põldmaa K, Tamm H (2009) Ascomycetes associated with ectomycorrhizas: molecular diversity and ecology with particular reference to the Helotiales. Environmental Microbiology 11: 3166–3178. doi: 10.1111/j.1462-2920.2009.02020.x

Tedersoo L, Smith ME (2013) Lineages of ectomycorrhizal fungi revisited: foraging strategies and novel lineages revealed by sequences from belowground. Fungal Biology Reviews 27: 83–99. doi: 10.1016/j.fbr.2013.09.001

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerknoes NC, Wilkins MJ, Hettich RL, Lipton MS, Williams KH, Long PE, Banfield JF (2012) Fermentation, hydrogen, and sulfur Metabolism in multiple uncultivated bacterial phyla. Science 337: 1661–1665. doi: 10.1126/science.1224041

Yoshida K, Schuenemann VJ, Cano LM, Pais M, Mishra B, Sharma R, Lanz C, Martin FN, Kamoun S, Krause J, Thines M, Weigel D, Burbano HA (2013) The rise and fall of the *Phytophthora infestans* lineage that triggered the Irish potato famine. eLife 2: e00731.

Supplementary material 1

Full information and metadata about the genomic and metagenomic samples
Authors: Leho Tedersoo, Ingrid Liiv, Paula Ann Kivistik, Sten Anslan, Urmas Kõljalg, Mohammad Bahram
Data type: table
Explanation note: Detailed information about metadata, DNA quality and genomic/ metagenomic results of fruit-body and EcM root tip samples.

RESEARCH ARTICLE

Detection of signal recognition particle (SRP) RNAs in the nuclear ribosomal internal transcribed spacer 1 (ITS1) of three lineages of ectomycorrhizal fungi (Agaricomycetes, Basidiomycota)

Magnus Alm Rosenblad[1], María P. Martín[2], Leho Tedersoo[3], Martin Ryberg[4], Ellen Larsson[5], Christian Wurzbacher[5], Kessy Abarenkov[6], R. Henrik Nilsson[5]

1 *Department of Marine Sciences, University of Gothenburg, Bioinformatics Infrastructure for Life Sciences, Box 460, SE-405 30 Göteborg* 2 *Departmento de Micología, Real Jardín Botánico, RJB-CSIC, Plaza de Murillo 1, 28014 Madrid, Spain* 3 *Institute of Ecology and Earth Sciences, University of Tartu, Tartu, Estonia* 4 *Department of Organismal Biology, Evolutionary Biology Centre, Uppsala University, Sweden* 5 *Department of Biological and Environmental Sciences, University of Gothenburg, Box 463, 405 30 Göteborg, Sweden* 6 *Natural History Museum, University of Tartu, Vanemuise 46, Tartu 510 14, Estonia*

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

Abstract

During a routine scan for Signal Recognition Particle (SRP) RNAs in eukaryotic sequences, we surprisingly found in silico evidence in GenBank for a 265-base long SRP RNA sequence in the ITS1 region of a total of 11 fully identified species in three ectomycorrhizal genera of the Basidiomycota (Fungi): *Astraeus*, *Russula*, and *Lactarius*. To rule out sequence artifacts, one specimen from a species indicated to have the SRP RNA-containing ITS region in each of these genera was ordered and re-sequenced. Sequences identical to the corresponding GenBank entries were recovered, or in the case of a non-original but conspecific specimen differed by three bases, showing that these species indeed have an SRP RNA sequence incorporated into their ITS1 region. Other than the ribosomal genes, this is the first known case of non-coding RNAs in the eukaryotic ITS region, and it may assist in the examination of other types of insertions in fungal genomes.

Key words

SRP RNA, non-coding RNA, ribosomal RNA, internal transcribed spacer 1 (ITS1), ectomycorrhizal fungi

Introduction

The nuclear ribosomal internal transcribed spacer (ITS) region is part of the ribosomal DNA cistron. The ITS region is transcribed together with the 18S, 5.8S, and 28S genes but removed in the post-transcriptional processing of the rRNA. The ITS region has three separate subregions: the ITS1, the 5.8S gene, and the ITS2. The ITS1 is situated between the 18S and the 5.8S genes, and the ITS2 is situated between the 5.8S and the 28S genes. The ITS region varies significantly in length among fungal species (Taylor and McCormick 2008, Tedersoo et al. 2015), and both of ITS1 and ITS2 form secondary structures with stems, bulges, and loops (Freire et al. 2012, Rampersad 2014). The secondary structure is important for correct processing of the rRNA; although the ITS1 and ITS2 regions are not expressed in the ribosome, there are constraints on the evolution of the ITS region, and it has both fast evolving and more conserved regions (Nazar 2004, Mullineaux and Hausner 2009). However, there is still much to learn about the function of the ITS region, especially for ITS1 (Rampersad 2014, Coleman 2015). The fast evolving regions of ITS has made it a cornerstone in species/genus-level phylogenetic inference in fungi and other organisms for more than 20 years, and it is the formal fungal barcode used for molecular species identification (Álvarez and Wendel 2003, Schoch et al. 2012).

One element that has never been implicated in the context of the eukaryotic rDNA cluster and ITS evolution is the existence of non-coding RNAs (ncRNA) other than the 18S, 5.8S, and 28S rRNAs. Based on the recent identification of a ubiquitous eukaryotic ncRNA in the fungal phylum Basidiomycota, viz. the Signal Recognition Particle RNA (SRP RNA; Dumesic et al. 2015), we discovered a ~265 bases long homologue of this gene in a set of fungal ITS1 sequences (Fig. 1). The SRP RNA is an essential component of the SRP, a ribonucleoprotein particle that co-translationally directs proteins to the endoplasmic reticulum (ER) membrane. The SRP RNA acts both as a scaffold for the SRP proteins and as a regulator of the SRP by mediating a global reorganization of the SRP in response to cargo binding (Rosenblad et al. 2009, Akopian et al. 2013).

A more thorough in silico analysis verified the presence of SRP RNAs in the ITS1 region of a total of 11 fully identified fungal species (separate Latin binomials) distributed over three lineages of ectomycorrhizal basidiomycetes (Boletales: *Astraeus* (1 species: *A. sirindhorniae*), Russulales: *Russula* (1 species: *R. olivacea*), and Russulales: *Lactarius* (9 species: *L. argillaceifolius*, *L. aspideus*, *L. brunneoviolaceus*, *L. luridus*, *L. nanus*, *L. pallescens*, *L. pseudouvidus*, *L. uvidus*, and *L. violascens*). The notion of an additional ncRNA element in the ITS1 region is novel and would seem – at least at a first glance – to compromise the function of the ITS1. Hypothetically, any of contamination, chimeric unions, or other laboratory or data analysis artifacts could explain this finding. In this study we apply DNA sequencing and bioinformatics to verify the presence of SRP RNA sequences in the ITS1 region of representatives of these fungi.

Materials and methods

SRP RNA bioinformatics

The bioinformatic analysis of non-coding RNAs such as the SRP RNA is not trivial, as the primary sequence may vary substantially as long as the secondary structure is preserved. To enable searches for SRP RNAs without requiring exact sequence matches across the full length of the SRP RNA, we used a secondary structure covariance model constructed from the full set of available ascomycete SRP RNAs with the basidiomycete SRP RNAs from Dumesic et al. (2015) added, as well as a second dataset containing all covariance models from Rfam (Nawrocki et al. 2014). These models were used in an INFERNAL v1.1 cmsearch (Nawrocki and Eddy 2013) run against flatfiles of the International Nucleotide Sequence Database Collaboration (INSDC; Nakamura et al. 2013; February 2015 release). After observing several highly significant matches to what seemed to be the fungal ITS1 region, we re-ran the search on the flatfiles from the manually curated fungal ITS database UNITE (Abarenkov et al. 2010; release 2015-08-01) using ITSx 1.0.11 (Bengtsson-Palme et al. 2013) to identify the ITS1 region. A total of 63 matches and 11 fully identified species from three ectomycorrhizal basidiomycete lineages were recovered (Suppl. material 1). All matches were examined manually to verify that they displayed all the universally conserved motifs and nucleotides. The sequences were found to stem from more than 20 different published and unpublished studies. The fact that these sequences had been found multiple times independently is highly suggestive of technically sound, authentic sequence data (Nilsson et al. 2012), but to further confirm the authenticity of the sequences we re-sequenced one herbarium specimen from each of these lineages, either from the original material or from other conspecific specimens.

PCR and sequencing

To rule out systematic PCR artifacts as sources of false positives in the bioinformatics analyses, we retrieved the original, or conspecific, specimens underlying one representative from each of the genera (Table 1): the conspecific collection *Astraeus sirindhorniae* MA-Fungi 47735 (collected in the Philippines; herbarium MA), the authentic collection *Lactarius luridus* TU118993 (collected in Estonia; herbarium TU), and the authentic collection *Russula olivacea* TU101845 (collected in Estonia; herbarium TU). We specifically sought to use a different primer pair combination and PCR conditions than did the original sequence authors in a further attempt at generalizing our findings. For *A. sirindhorniae*, DNA extractions, PCR reactions, and sequencing were performed as described in Martín and Winka (2000), however using DNeasy Plant Mini Kit (Qiagen) with overnight incubation for DNA extraction. Primers used for amplification were ITS5/ITS4 (White et al. 1990; Suppl. material 1). For the specimens of

Lactarius and *Russula*, the DNA was extracted and amplified using the primers ITS1f and ITS4b following Anslan and Tedersoo (2015). Sequences were edited and assembled using Sequencher 4.2 (Gene Codes, Ann Arbor). All sequences were examined for sequence quality following Nilsson et al. (2012). Chimera detection was undertaken using UCHIME (Edgar et al. 2011) and the UNITE chimera reference dataset (Nilsson et al. 2015; release 2015-03-11).

Results

The ITS sequences recovered from the sequencing round passed all quality control measures we exercised. In addition, no sequence was found to have the multiple DNA ambiguity symbols suggestive of the presence of several information-wise distinct ITS copies in the individuals at hand (Hyde et al. 2013). The resulting sequences manifested the SRP RNA sequence in the ITS1 region of all three re-sequenced lineages. The two authentic specimens of *Russula* and *Lactarius* produced identical ITS sequences to those already extant. The sequence from the conspecific *Astraeus sirindhorniae* specimen differed by three bases from the extant sequence, which is well within the expected intraspecific variation when conspecific isolates are compared across geographical distances (Thailand and the Philippines in this case). The sequences were deposited in the INSDC as accessions KU356730-KU356732.

SRP RNA-containing sequences of *Russula* and *Lactarius* were found to have an average length of some 890 bases; the corresponding average length for the SRP RNA-containing *Astraeus* sequences was 840 bases. When using BLAST to find the most similar sequences of *Russula*, *Lactarius*, and *Astraeus* that did not contain the SRP RNA, we found that their ITS region was on average 616 bases (*Russula*), 644 bases (*Lactarius*), and 620 bases (*Astraeus*). This corresponds well to the length of the SRP RNA (~265 bases) for all of *Astraeus*, *Russula,* and *Lactarius*, allowing for some few bases of divergence considering the cross-species comparison. The distances between the SRP RNA and the surrounding genes 18S and 5.8S were almost the same within each lineage, but differed somewhat among the three lineages: 80 and 174 bases (*Lactarius*), 150 and 80 bases (*Russula*), and 55 and 132 bases (*Astraeus*).

Discussion

The finding that ncRNAs are located in tandem is not novel. Apart from the highly conserved nuclear rDNA cluster, some ncRNAs have been found to cluster in several protist species, e.g., SRP RNA together with U6 snRNA, 5S rRNA, SL RNA, and tRNAs in dinoflagellates (Zhang et al. 2013). Regarding the transcription of the SRP RNA in fungi, the transcriptional promoters of the SRP RNA in *Saccharomyces cerevisiae* (the TFIIIC-binding A- and B-box) are internal, and the SRP RNAs have a polypyrimidine termination sequence similar to other RNA polymerase III genes. Although
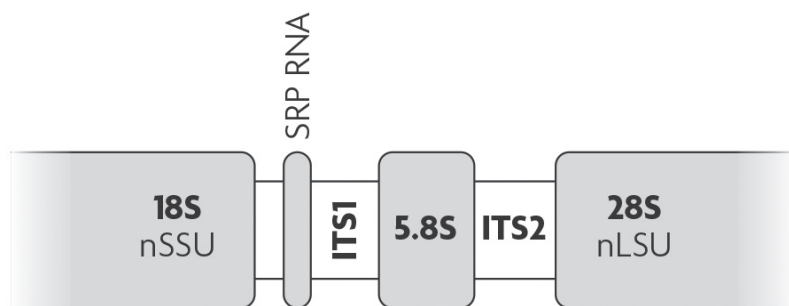
Figure 1. Schematic illustration of the fungal ITS region and neighboring rDNA genes. The subregions ITS1, 5.8S, and ITS2 of the ITS region are indicated along with the SRP RNA in the first part of the ITS1. The absolute positions of the subregions and the SRP RNA are provided in Suppl. material 2.

yeast SRP RNA also has an upstream TATA box – a motif we could not clearly identify in our alignments – inactivation of this region does not result in a significant effect on transcription (Dieci et al. 2002). Therefore it is possible that our identified SRP RNAs could be transcribed independently of the transcription of the rDNA cluster. If so, the SRP RNAs found in the ITS1 region most probably are fully functional, but there may be another copy in the genome that constitutes the major transcript. Whether the SRP RNA could be a product of the rDNA processing seems less likely. Although the A_3 cleavage site that is the closest to the 5.8S should be downstream of the SRP RNA 3' end since the distance between the SRP RNA and the 5.8S is at least 80 nucleotides in the 63 sequences, the SRP RNA needs correct 5' and 3' ends to fold into the proper secondary structure. Therefore the processing of the tricistronic ribosomal transcript most probably leads to non-functional SRP RNA. The surprising lack of mutations, as compared to other identified basidiomycete SRP RNAs (Suppl. material 4), could be explained not only by the insertion being a recent event, but also by the need to preserve the secondary structure of the SRP RNA region and thus the remaining parts of ITS1.

The three species from which the SRP RNA was recovered are all ectomycorrhizal basidiomycetes and come from two different orders and two different families. Two of these lineages are closely related (*Russula* and *Lactarius*, both in Russulaceae (Russulales)); the third one – *Astraeus sirindhorniae* (Boletales) – comes from the same subphylum (Agaricomycotina) as the former two. Even so, the Russulales and the Boletales are separate orders, such that the presence of SRP RNAs in these fungi must be considered independent gains. In the case of *Russula* and *Lactarius* – two very speciose genera – the vast majority of the known species do not have the SRP RNAs in their ITS1. Similarly, none of the other species in *Astraeus* treated by Phosri et al. (2014) were found to have the SRP RNA. It would appear far more realistic to view these SRP RNAs as independent insertion events than as a plesiomorphic ITS state where the ancestor contained the SRP RNA, but where all species except the few considered here lost it (Miller et al. 2006; Suppl. materials 1–4). Indeed, we found no evidence for SRP RNA in the ITS region of any other fungus.

Table 1. Data on the underlying specimens and PCR primers. The already sequenced specimens of *Russula* and *Lactarius* were re-sequenced with a different primer pair compared to the extant sequences. Our Philippines specimen of *Astraeus sirindhorniae* had never been sequenced before, but we used a different primer pair compared to the *Astraeus sirindhorniae* sequence generated by Phosri et al. (2014) from a Thailand collection. Primer sequences are available in Suppl. material 1B.

Species	Original entry	Resequenced entry	Specimen	Herbarium	Country	Original primers	Resequencing primers	ITS1	5.8S	ITS2
Astraeus sirindhorniae	(not sequenced before)	KU356730	MA-Fungi 47735	Madrid	Phillipines	(not sequenced before)	ITS5 / ITS4	1–442	443–444	–
Lactarius luridus	UDB023551 (UNITE)	KU356731	TU118993	Tartu	Estonia	ITS0F / LB-W	ITS1F / ITS4b	1–482	483–640	641–896
Russula olivacea	UDB016000 (UNITE)	KU356732	TU101845	Tartu	Estonia	ITS0Fr / LB-W	ITS1F / ITS4b	1–457	458–615	616–891

The three previously identified SRP RNAs in the Russulales are not located in or close to the rDNA cluster (Dumesic et al. 2015), and we argue that the SRP RNA must be considered as an independently inserted element in these ITS1 sequences. Although this does not cause any problems in terms of molecular identification of these species, it does present a potential difficulty to the uncritical use of ITS sequences in phylogenetic inference in these fungal lineages. Under the assumption that the SRP RNA is found in the exact same position in the ITS1 region among species, the sequences could still be aligned jointly as long as the SRP RNA part is kept as a separate element in the multiple sequence alignment. Any failure to realize that the SRP RNA should be treated as a separate element to be scored as gaps in species that do not have the SRP RNA is certain to give rise to very noisy multiple sequence alignments and skewed inferences of phylogeny. In other words, there is a risk that alignment tools will try to align other parts of the ITS1 region onto the SRP RNA part in large alignments, which would violate homology assumptions. We briefly examined whether several of the most commonly used alignment programs were able to recognize the unique nature of the SRP RNA and not try to stack other parts of the ITS1 onto the SRP RNA. The results were generally encouraging as long as the number of non-SRP RNA containing species was kept reasonably low, with only minor manual adjustments needed a posteriori. In the worse situation where the SRP RNA insertion is not found in the exact same position across species, it will not be possible to maintain position homology in the multiple sequence alignment. In that scenario, the sequences containing the SRP RNA must be excluded from the alignment process, or the SRP RNA element must be removed. Interestingly, Eberhardt (2002) reported an unexpected 250-base insertion in the ITS1 of *Russula olivacea* – one of the species examined in the present study – and chose to exclude it from her multiple sequence alignment due to alignment difficulties. In hindsight it seems probable that this 250-base region indeed represents the SRP RNA. As demonstrated by the Eberhardt (2002) example, there is widespread (although not necessarily universal) awareness of the importance of examining multiple sequence alignments manually before they are put to scientific use. However, the increasing use of fully automated solutions to data harvesting and phylogenetic inference may present a concern here (cf. Antonelli et al. 2014).

Our findings are not without potential shortcomings though. The number of ITS copies per fungal cell can approach 200 or more (Bellemain et al. 2010; Black et al. 2013). Whereas the process of concerted evolution is thought to homogenize the array of ITS copies (Álvarez and Wendel 2003), it is not uncommon to find evidence of two or more distinct ITS copies during sequencing work (Hyde et al. 2013). A recent pyrosequencing-based study found evidence for multiple, information-wise distinct ITS copies in 3–5% of the 99 examined species of Ascomycota and Basidiomycota (Lindner et al. 2013). The extent to which the three species examined here contain multiple distinct ITS copies is unknown but may well be low, given that we obtained single, clean PCR products and sequence chromatograms for all three species. Even so, it is conceivable that we – much like the original sequence authors – in fact amplified a rare and perhaps non-functional ITS copy. Although this would not disqualify our

finding of an SRP RNA in the fungal ITS region, it would raise questions regarding whether the inclusion of the SRP RNA ruined the function of this particular ITS copy, essentially rendering the corresponding rRNA non-functional. In either case, we view this multiple-copy scenario as unlikely given our consistent obtainment, and the more than 20 independent recoveries, of the SRP RNA-containing ITS1 sequences.

Unfortunately, none of the species of the present study have a complete genome published, so a detailed analysis of the SRP RNA in the context of the genomes of these and closely related fungi will have to wait. That said, the trend that published fungal genomes tend to come without the ribosomal operon for reasons of convenience is most unfortunate (Schoch et al. 2014). We join the barcoding community in extending a plea that whenever a genome is sequenced and assembled, the ribosomal operon should be assembled into the genome as a part of that process. If this undertaking proves to be too complex, then at least the full ribosomal operon should be bundled with the genome, even if its assembly into the genome cannot be accomplished.

Conclusions

We found evidence of Signal Recognition Particle (SRP) RNAs in the ITS1 region of a total of 11 fully identified species in three ectomycorrhizal genera: *Astraeus*, *Russula*, and *Lactarius*. Other than the ribosomal genes, this is the first known case of noncoding RNAs in the fungal ITS region. Our finding is small step towards explaining the many insertions found throughout fungal genomes, and it adds a new element to the field of fungal ITS evolution.

Acknowledgements

References

Abarenkov K, Nilsson RH, Larsson K-H, Alexander IJ, Eberhardt U, Erland S, Høiland K, Kjøller R, Larsson E, Pennanen T, et al. (2010) 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

Akopian D, Shen K, Zhang X, Shan SO (2013) Signal Recognition Particle: An essential protein targeting machine. Annual Review of Biochemistry 82: 693–721. doi: 10.1146/annurev-biochem-072711-164732

Álvarez I, Wendel JF (2003) Ribosomal ITS sequences and plant phylogenetic inference. Molecular Phylogenetics and Evolution 29(3): 417–434. doi: 10.1016/S1055-7903(03)00208-2

Anslan S, Tedersoo L (2015) Performance of cytochrome c oxidase subunit I (COI), ribosomal DNA Large Subunit (LSU) and Internal Transcribed Spacer 2 (ITS2) in DNA barcoding of Collembola. European Journal of Soil Biology 69: 1–7. doi: 10.1016/j.ejsobi.2015.04.001

Antonelli A, Condamine FL, Hettling H, Nilsson K, Nilsson RH, Oxelman B, Sanderson MJ, Sauquet H, Scharn R, Silvestro D, et al. (2014) SUPERSMART: ecology and evolution in the era of big data. PeerJ PrePrints 2: e501v1. doi: 10.7287/peerj.preprints.501v1

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H (2010) ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology 10: 189. doi: 10.1186/1471-2180-10-189

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F, et al. (2013) Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods in Ecology and Evolution 4810: 914–919. doi: 10.1111/2041-210X.12073

Black J, Dean T, Byfield G, Foarde K, Menetrez M (2013) Determining fungi rRNA copy number by PCR. Journal of Biomolecular Techniques 24(1): 32–38. doi: 10.7171/jbt.13-2401-004

Coleman AW (2015) Nuclear rRNA transcript processing versus internal transcribed spacer secondary structure. Trends in Genetics 31(3): 157–163. doi: 10.1016/j.tig.2015.01.002

Dieci G, Giuliodori S, Catellani M, Percudani R, Ottonello S (2002) Intragenic promoter adaptation and facilitated RNA polymerase III recycling in the transcription of SCR1, the 7SL RNA gene of *Saccharomyces cerevisiae*. Journal of Biological Chemistry 277(9): 6903–6914. doi: 10.1074/jbc.M105036200

Dumesic PA, Rosenblad MA, Samuelsson T, Nguyen T, Moresco JJ, Yates JR, Madhani HD (2015) Noncanoncial signal recognition particle RNAs in a major eukaryotic phylum revealed by purification of SRP from the human pathogen *Cryptococcus neoformans*. Nucleic Acids Research 43(18): 9017–9027. doi: 10.1093/nar/gkv819

Eberhardt U (2002) Molecular kinship analyses of the agaricoid Russulaceae: correspondence with mycorrhizal anatomy and sporocarp features in the genus *Russula*. Mycolological Progress 1(2): 201–223. doi: 10.1007/s11557-006-0019-6

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27(16): 2194–2200. doi: 10.1093/bioinformatics/btr381

Freire MCM, da Silva MR, Zhang X, Almeida ÁMR, Stacey G, de Oliveira LO (2012) Nucleotide polymorphism in the 5.8 S nrDNA gene and internal transcribed spacers in *Phakopsora pachyrhizi* viewed from structural models. Fungal Genetics and Biology 49(2): 95–100. doi: 10.1016/j.fgb.2011.12.010

Hyde KD, Udayanga D, Manamgoda DS, Tedersoo L, Larsson E, Abarenkov K, Bertrand YJK, Oxelman B, Hartmann M, Kauserud H, et al. (2013) Incorporating molecular data in fungal systematics: a guide for aspiring researchers. Current Research in Environmental and Applied Mycology 3(1): 1–32.

Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular Biology and Evolution 30(4): 772–780. doi: 10.1093/molbev/mst010

Lindner DL, Carlsen T, Nilsson RH, Davey M, Schumacher T, Kauserud H (2013) Employing 454 amplicon pyrosequencing to reveal intragenomic divergence in the internal transcribed spacer rDNA region in fungi. Ecology and Evolution 3(6): 1751–1764. doi: 10.1002/ece3.586

Maddison DR, Swofford DL, Maddison WP (1997) NEXUS: an extensible file format for systematic information. Systematic Biology 46(4): 590–621. doi: 10.1093/sysbio/46.4.590

Martín MP, Winka K (2000) Alternative methods of extracting and amplifying DNA from lichens. Lichenologist 32(2): 189–196. doi: 10.1006/lich.1999.0254

Miller SL, Larsson E, Larsson K-H, Verbeken A, Nuytinck J (2006) Perspectives in the new Russulales. Mycologia 98(6): 960–970. doi: 10.3852/mycologia.98.6.960

Mullineaux T, Hausner G (2009) Evolution of rDNA ITS1 and ITS2 sequences and RNA secondary structures within members of the fungal genera *Grosmannia* and *Leptographium*. Fungal Genetics and Biology 46(11): 855–867. doi: 10.1016/j.fgb.2009.08.001

Nakamura Y, Cochrane G, Karsch-Mizrachi I (2013) The international nucleotide sequence database collaboration. Nucleic Acids Research 41(D): D21–D24. doi: 10.1093/nar/gks1084

Nawrocki EP, Eddy SR (2013) Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29(22): 2933–2935. doi: 10.1093/bioinformatics/btt509

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, Finn RD (2014) Rfam 12.0: updates to the RNA families database. Nucleic Acids Research 43(D1): D130–D1373. doi: 10.1093/nar/gku1063

Nazar RN (2004) Ribosomal RNA processing and ribosome biogenesis in eukaryotes. IUBMB Life 56(8): 457–465. doi: 10.1080/15216540400010867

Nilsson RH, Tedersoo L, Abarenkov K, Ryberg M, Kristiansson E, Hartmann M, Schoch CL, Nylander JAA, Bergsten J, Porter TM, et al. (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

Nilsson RH, Tedersoo L, Ryberg M, Hartmann M, Unterseher M, Porter TM, Bengtsson-Palme J, Walker DM, de Sousa F, et al. (2015) A comprehensive, automatically updated fungal ITS sequence dataset for reference-based chimera control in environmental sequencing efforts. Microbes and Environments 30(2): 145–150. doi: 10.1264/jsme2.ME14121

Phosri C, Watling R, Suwannasai N, Wilson A, Martín MP (2014) A new representative of star-shaped fungi: *Astraeus sirindhorniae* sp. nov. from Thailand. PLoS ONE 9(5): e71160. doi: 10.1371/journal.pone.0071160

Rampersad SN (2014) ITS1, 5.8S and ITS2 secondary structure modelling for intra-specific differentiation among species of the *Colletotrichum gloeosporioides* sensu lato species complex. SpringerPlus 3: 684. doi: 10.1186/2193-1801-3-684

Rosenblad MA, Larsen N, Samuelsson T, Zwieb C (2009) Kinship in the SRP RNA family. RNA Biology 6(5): 508–516. doi: 10.4161/rna.6.5.9753

Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, Chen W, et al. (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

Schoch CL, Robbertse B, Robert V, Vu D, Cardinali G, Irinyi L, Meyer W, Nilsson RH, Hughes K, Miller AN, et al. (2014) Finding needles in haystacks: linking scientific names, reference specimens and molecular data for Fungi. Database (Oxford) 2014: bau061. doi: 10.1093/database/bau061

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, Anslan S, Bahram M, Põlme S, Riit T, Liiv I, Kõljalg U, Kisand V, Nilsson H, Hildebrand F, Boork P, Abarenkov K (2015) Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoding analyses of fungi. MycoKeys 10: 1–43. doi: 10.3897/mycokeys.10.4852

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. PCR protocols: a guide to methods and applications 18: 315–322.

Zhang H, Campbell DA, Sturm NR, Rosenblad MA, Dungan CF, Lin S (2013) Signal Recognition Particle RNA in dinoflagellates and the perkinsid *Perkinsus marinus*. Protist 164(5): 748–761. doi: 10.1016/j.protis.2013.07.004

Supplementary material 1

Output from cmsearch and primers used

Authors: Magnus Alm Rosenblad, María P. Martín, Leho Tedersoo, Martin Ryberg, Ellen Larsson, Christian Wurzbacher, Kessy Abarenkov, R. Henrik Nilsson

Data type: text/computer output

Explanation note: A) The output from cmsearch showing all 63 relevant matches to the three ectomycorrhizal lineages. B) Detail of the primers used to re-amplify the specimens.

Supplementary material 2

ITS multiple sequence alignment

Authors: Magnus Alm Rosenblad, María P. Martín, Leho Tedersoo, Martin Ryberg, Ellen Larsson, Christian Wurzbacher, Kessy Abarenkov, R. Henrik Nilsson

Data type: text/DNA sequence data

Explanation note: A multiple sequence alignment in the NEXUS format (Maddison et al. 1997) comprising all 63 matching ITS sequences, plus the three newly generated ones (KU356730, KU356731, and KU356732). The alignment was produced in MAFFT without manual adjustment (Katoh and Standley 2013). The alignment is composed of partial nSSU (bases 1-34 in the alignment), the full ITS1 (bases 35-678), the full 5.8S (bases 679-838), the full ITS2 (bases 839-1395), and partial nLSU (bases 1396-end). The SRP RNA occupies position 203-474 in the alignment. The alignment is provided for overview purposes only; the two-order nature of the taxa (Boletales and Russulales) coupled with the high variability of the ITS region jointly mean that the alignment will not be suited for phylogenetic inference.

Supplementary material 3

ITS/SRP RNA multiple sequence alignment
Authors: Magnus Alm Rosenblad, María P. Martín, Leho Tedersoo, Martin Ryberg, Ellen Larsson, Christian Wurzbacher, Kessy Abarenkov, R. Henrik Nilsson
Data type: text/DNA sequence data
Explanation note: Multiple sequence alignment comprising the 63 public ITS1 sequences with SRP RNA found in them, the three newly generated sequences, and the SRP RNA sequences from Dumesic et al. (2015) (*Stereum hirsutum*, *Heterobasidion irregulare*, and *Heterobasidion annosum*).

Supplementary material 4

SRP RNA multiple sequence alignment
Authors: Magnus Alm Rosenblad, María P. Martín, Leho Tedersoo, Martin Ryberg, Ellen Larsson, Christian Wurzbacher, Kessy Abarenkov, R. Henrik Nilsson
Data type: text/DNA sequence data
Explanation note: Multiple sequence alignment with the SRP RNA sequences of Dumesic et al. (2015; *Stereum hirsutum*, *Heterobasidion irregulare*, and *Heterobasidion annosum*) aligned to our newly generated ITS sequences of *Russula* and *Lactarius*.

RESEARCH ARTICLE

Psora altotibetica (Psoraceae, Lecanorales), a new lichen species from the Tibetan part of the Himalayas

Einar Timdal[1], Walter Obermayer[2], Mika Bendiksby[3]

1 *Natural History Museum, University of Oslo, P.O. Box 1172 Blindern, N-0318 Oslo, Norway* **2** *Institut für Pflanzenwissenschaften, Karl-Franzens-Universität Graz, Holteigasse 6, A-8010 Graz, Austria* **3** *NTNU University Museum, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*

Corresponding author: *Einar Timdal* (einar.timdal@nhm.uio.no)

Abstract

In the present study, we describe the new species, *Psora altotibetica*, from nine localities in China (Tibetan area) and Nepal. The study includes analyses of anatomy, secondary chemistry, and DNA sequence data of *P. altotibetica* and presumed close relatives. *Psora altotibetica* resembles *P. indigirkae* morphologically, but is phylogenetically closer to *P. tenuifolia* and *P. vallesiaca*. It differs from *P. indigirkae* in the colour of the apothecia, the size of the ascospores, and in the secondary chemistry. The species is terricolous and was collected in the alpine zone of the Great Himalayas between 4230 and 5000 m altitude. *Psora tenuifolia* and *P. vallesiaca* are here reported as new to China and the Himalayas.

Key words

Asia, Lecanorales, lichenized ascomycetes, taxonomy

Introduction

The genus *Psora* consists of c. 30 species growing on soil and rock, mainly in arid areas, from the arctic to the subtropical regions of the world (Timdal 2002). The current concept of the genus was proposed by Schneider (1980) and emended by Timdal (1984). No world monograph exists, but there are two revisions of the genus in North America (Timdal 1986, 2002) and keys to the species in Europe (Poelt and Vězda 1981) and Asia (Timdal and Zhurbenko 2004). Ekman and Blaalid (2011) published a molecular phylogeny of the Psoraceae, including 18 species of *Psora*.

In the current Chinese checklist (Wei 1991), five species of *Psora* are listed: (1) *P. asahinae* (Zahlbr.) J.C.Wei, (2) *P. asiae-centralis* (H.Magn.) N.S.Golubk., (3) *P. crenata* (Taylor) Reinke, (4) *P. decipiens* (Hedw.) Hoffm., and (5) *P. lurida* (Ach.) DC. However, three species on this list (1, 2, and 5) are currently excluded from *Psora*, and for one (3), the single Chinese record is based on incorrectly identified material [*Psora asahinae* was synonymised with *Psorula rufonigra* (Tuck.) Gotth.Schneider (Schneider 1980), *P. asiae-centralis* was placed in *Toninia* as *T. tristis* ssp. *asiae-centralis* (H.Magn.) Timdal by Timdal (1992), *P. lurida* was placed in *Romjularia* as *R. lurida* (Ach.) Timdal by Timdal (2008), and the report of *P. crenata* is based on the holotype of *Lecidea undulata* H.Magn. (Bohlin 82, S!) which was placed in synonymy with *Anamylopsora pulcherrima* (Vain.) Timdal by Timdal (1991)]. Hence, *P. decipiens* (4) is the only currently known species of *Psora* in China. It is listed from Xizang by Wei (1991) and from Yunnan by Wang (2012, including a beautiful photograph), and is, together with *Psora himalayana* (C.Bab.) Timdal (described from Uttar Pradesh, India, at 4700 ft), the only known *Psora* species in the Himalayas.

On a visit to the herbarium of Institut für Pflanzenwissenschaften, Karl-Franzens-Universität Graz (GZU) in 1992, one of us (ET) discovered an apparently undescribed species of *Psora* collected near the Khumbu Glacier south of Mt Everest (Nepal) by Josef Poelt in 1962. Two years later, ET was asked by Torstein Engelskjøn to identify some specimens collected by him in 1993 in the Rongbuk Valley north of Mt Everest (Tibet) during a joint Chinese-Norwegian scientific expedition. The material comprised the same apparently undescribed species of *Psora*. The material from both expeditions was rather poor and put aside pending richer collections. In 2013, during a visit to GZU, MB was made aware of the *Psora* collection that WO had made in the Himalayas in 1994 and 2000. This collection contained the putative new species, and a renewed study, including DNA analysis, has led us to describe it here as *Psora altotibetica*.

Material and methods

The specimens

This study is based on 15 collections of *Psora* made by WO in the Himalayas in 1994 and 2000, the collection of *Psora altotibetica* by J. Poelt (GZU), the two by T. Engelskjøn

(TROM), two by G. & S. Miehe made in the Himalayas in 1993 (GZU), four collections from the Karakoram Range (Pakistan) in GZU which are here identified as *P. himalayana* and *P. vallesiaca* (Schaer.) Timdal, and two collections of *P. indigirkae* Timdal & Zhurb. from Yakutia in O. Voucher data for these 26 specimens are given in the list of examined specimens and in Table 1. The table also contains voucher data for the specimens of which DNA was downloaded from GenBank or obtained through the Norwegian Barcode of Life project (NorBOL). With the exception of one collection (*P. himalayana*, Zhurbenko 98161), we have examined all collections listed in Table 1 by morphology during this project or previously.

Anatomy

Microscope sections were cut on a freezing microtome and mounted in water, 10% KOH (K), lactophenol cotton blue, a modified Lugol's solution in which water was replaced by 50% lactic acid, and chlor-zinc-iodine. Amyloid reactions were observed in the modified Lugol's solution after pretreatment in K. Chlor-zinc-iodine was used to locate remnants of algae in the cortex, and polarized light was used to locate crystals of secondary metabolites and calcium oxalate. Calcium oxalate was identified by adding 25% sulphuric acid to the section; the oxalate crystals dissolve and needle shaped crystals of calcium sulphate precipitate. Ascospore measurements are given as $X \pm 1.5 \times SD$, where X is the arithmetic mean and SD – the standard deviation.

Secondary chemistry

Thin-layer chromatography (TLC) was performed in accordance with the methods of Culberson (1972), modified by Menlove (1974) & Culberson and Johnson (1982). All specimens were examined by TLC, with the exception of Miehe & Miehe 9573/23/03 and Obermayer 4502, which were omitted due to being represented by richer material from the same localities.

DNA extraction, PCR and sequencing

We performed DNA extraction, PCR amplification, PCR purification, and cycle sequencing as described by Bendiksby and Timdal (2013). DNA was extracted from apothecia of 13 specimens (Table 1; GenBank Accession Numbers KU863631–KU863656). All DNA isolates produced for the present study are deposited in the DNA collection at Natural History Museum, University of Oslo. We amplified and sequenced the nuclear ribosomal internal transcribed spacer (ITS) and the mitochondrial ribosomal small subunit (mtSSU) using the primer pairs ITS5/ITS4 (White et al. 1990) and mtSSU1/mtSSU3R (Zoller et al. 1999), respectively.

Table 1. *Psora* specimens used in this study with voucher information, major lichen substances, and GenBank accession numbers. New sequences are indicated by accession numbers in bold.

Taxon, Specimen	Voucher Information	Major Lichen Substances	GenBank Accession Number ITS	mtSSU
P. altotibetica 1	China, Xizang, *Obermayer 5282* (GZU), holotype	gyrophoric acid	**KU863638**	**KU863651**
P. altotibetica 2	China, Xizang, *Miehe & Miehe 9573/23/02* (GZU)	gyrophoric acid	**KU863639**	**KU863652**
P. altotibetica 3	China, Xizang, *Obermayer 5223* (GZU)	gyrophoric acid	**KU863640**	**KU863653**
P. altotibetica 4	China, Xizang, *Obermayer 4365* (GZU)	gyrophoric acid	**KU863642**	**KU863655**
P. altotibetica 5	China, Xizang, *Obermayer 3967* (GZU)	gyrophoric acid	**KU863641**	**KU863654**
P. altotibetica 6	China, Xizang, *Obermayer 4485* (GZU)	gyrophoric acid	**KU863643**	**KU863656**
P. altotibetica	China, Xizang, *Engelskjon* T-030b (TROM L-42812)	gyrophoric acid	-	-
P. altotibetica	China, Xizang, *Engelskjon* T-036 (TROM L-42813)	gyrophoric acid	-	-
P. altotibetica	China, Xizang, *Miehe & Miehe 9573/23/03* (GZU)	-	-	-
P. altotibetica	China, Xizang, *Obermayer 4350* (GZU)	gyrophoric acid	-	-
P. altotibetica	China, Xizang, *Obermayer 4502* (GZU)	-	-	-
P. altotibetica	China, Xizang, *Obermayer 4981* (GZU)	gyrophoric acid	-	-
P. altotibetica	Nepal, *Poelt* 1138 (GZU)	gyrophoric acid	-	-
P. californica	USA, California, *Timdal* SON139/04 (O-L-60112)	bourgeanic acid, gyrophoric acid	EF524322	EF524292
P. globifera 1	Greenland, *Timdal* 10149 (O-L-139171)	no substances	EF524323	EF524294
P. globifera 2	Norway, *Klepsland* JK11-L619 (O-L-183774)	no substances	**KU873928**	-
P. globifera 3	Norway, *Bendiksby et al.* 12914 (O-L-184327)	no substances	**KU873930**	-
P. globifera 4	Norway, *Klepsland* JK11-L213 (O-L-177145)	no substances	**KU873929**	-
P. globifera 5	Norway, *Hjelmstad* s.n. (O-L-184143)	no substances	**KU873932**	-
P. himalayana	Russia, Yakutia, *Zhurbenko* 98161 (M-0066792)	-	AY425635	-
P. himalayana	Pakistan, *Miehe & Miehe* 3529 (GZU)	no substances	-	-
P. himalayana	Pakistan, *Poelt* K91-416	no substances	-	-
P. hyporubescens	USA, California, *Bratt & Timdal* 7052 (O-L-22483), holotype	anthraquinones, gyrophoric acid	EF524311	EF524295
P. indigirkae 1	Russia, Yakutia, *Haugan & Timdal* YAK19/03 (O-L-19148), holotype	bourgeanic acid, gyrophoric acid	EF524302	-
P. indigirkae 2	Russia, Yakutia, *Haugan & Timdal* YAK17/24 (O-L-19086), paratype	bourgeanic acid, gyrophoric acid	**KU863631**	**KU863644**

Taxon, Specimen	Voucher Information	Major Lichen Substances	GenBank Accession Number	
			ITS	mtSSU
P. indigirkae 3	Russia, Yakutia, *Zhurbenko* 92185 (O-L-118686), paratype	bourgeanic acid, gyrophoric acid	KU863632	KU863645
P. nitida	Mexico, Baja California, *Timdal* SON33/06 (O-L-15546)	gyrophoric acid	EF524313	EF524296
P. pacifica	USA, California, *Rosentreter* 14580 (O-L-126265)	gyrophoric acid, unknown accessory	EF524314	EF524297
P. peninsularis	Mexico, Baja California, *Timdal* SON32/07 (O-L-15539), holotype	norstictic acid	EF524320	EF524298
P. russellii	Mexico, Baja California, *Timdal* SON31/03 (O-L-15531)	norstictic acid	EF524321	EF524300
P. tenuifolia 1	Russia, Yakutia, *Haugan & Timdal* YAK17/26 (O-L-19088)	norstictic acid, zeorin	EF524309	EF524303
P. tenuifolia 2	China, Xizang, *Obermayer* 4487 (GZU)	norstictic acid, zeorin	KU863636	KU863649
P. tenuifolia 3	China, Xizang, *Obermayer* 5236 (GZU)	zeorin	KU863637	KU863650
P. tenuifolia	China, Sichuan, *Obermayer* 9791 (GZU)	norstictic acid, zeorin	-	-
P. tenuifolia	China, Xizang, *Obermayer* 4525 (GZU)	zeorin	-	-
P. testacea	Greece, *Rui & Timdal* TH06/04 (O-L-59263)	atranorin	EF524315	EF524301
P. tuckermanii	USA, Arizona, *Rui & Timdal* US240/05 (O-L-59926)	no substances	EF524317	EF524304
P. vallesiaca 1	Greece, *Rui & Timdal* 7993 (O-L-15186)	norstictic acid	EF524324	EF524291
P. vallesiaca 2	China, Xizang, *Obermayer* 3227 (GZU)	norstictic acid	KU863633	KU863646
P. vallesiaca 3	China, Xizang, *Obermayer* 5279 (GZU)	no substances	KU863635	KU863648
P. vallesiaca 4	Pakistan, *Poelt* K91-705 (GZU)	norstictic acid	KU863634	KU863647
P. vallesiaca 5	Norway, *Bendiksby et al.* 12979 (O-L-184392)	norstictic acid	KU873926	-
P. vallesiaca 6	Norway, *Klepsland* JK11-L624 (O-L-183778)	norstictic acid	KU873927	-
P. vallesiaca 7	Norway, *Klepsland* JK11-L601 (O-L-183760)	norstictic acid	KU873931	-
P. vallesiaca	China, Xizang, *Obermayer* 4482 (GZU)	norstictic acid	-	-
P. vallesiaca	Pakistan, *Poelt* s.n. (GZU)	norstictic acid	-	-

Data analyses

Sequences were assembled and edited using SEQUENCHER v.4.1.4 (Gene Codes Corporation, Ann Arbor, Michigan, U.S.A.). Alignments were established in BIO-EDIT 7.2.3 (Hall 1999) using the "ClustalW/Multiple alignment" option with subsequent manual adjustments. We analysed and summarized the data with parsimony and Bayesian phylogenetic methods, including model testing, as described in Bendiksby et al. (2015). The nuclear and mitochondrial datasets were analysed separately and in combination (concatenated) with indels treated as missing data.

Results

Species identifications

The 24 Central Asian specimens were identified by morphology and secondary chemistry as *Psora altotibetica* (13), *P. himalayana* (2), *P. vallesiaca* (5), and *P. tenuifolia* Timdal (4).

Anatomy

The following key characters for including *P. altotibetica* in *Psora* were observed in the new species: the upper cortex contained remnants of algae throughout both the lower stainable layer and the upper epinecral layer ('Scheinrindentyp' of Poelt 1958); the hypothecium contained calcium oxalate crystals; the epihymenium contained orange crystals which dissolved in K with a purple diffusion (assumed to be anthraquinones); and the ascus contained a well-developed, amyloid tholus with a central, deeper amyloid tube structure (*Porpidia*-type).

The following species level characters were observed in *P. altotibetica*: Upper cortex composed of thin-walled hyphae with rounded lumina; lower cortex composed of mainly periclinally oriented hyphae; crystals of calcium oxalate and assumedly gyrophoric acid (dissolving in K) present both in upper cortex and medulla; no crystals in lower cortex; ascospores 9–14 × 5–7 μm.

Secondary Chemistry

The results of the TLC examinations are given in Table 1. All examined specimens of *P. altotibetica* contained gyrophoric acid; no traces of fatty acids were detected.

Molecular data

Altogether 26 DNA sequences were generated from 13 specimens for the present study (13 ITS and 13 mtSSU), including two specimens of *P. indigirkae* from Yakutia in O. In addition, seven unpublished ITS sequences of *P. globifera* (Ach.) A.Massal. and *P. vallesiaca* from Norway were generated by the lichen DNA barcode project, OLICH, at the Norwegian Barcode of Life (NorBOL). Moreover, 24 ITS and mtSSU sequences from 13 *Psora* specimens were downloaded from GenBank. GenBank accession numbers of all 57 sequences are given in Table 1.

Alignments and phylogenetic analyses

The ITS matrix of 32 accessions was 676 basepairs long and contained 181 parsimony-informative characters. The basepairs and parsimony-informative characters for the mtSSU matrix of 25 accessions were 881 and 28, respectively. The estimated best fit model of evolution for ITS was SYM+G and for mtSSU it was HKY+I+G. Both parsimony jackknife and Bayesian trees of ITS vs mtSSU were congruent but resolved to various extents (not shown). Therefore, for the final analyses, a concatenated dataset of 1557 bp was used. In the Bayesian analysis of the concatenated dataset, the average standard deviation of split frequencies had fallen to 0.0045 at termination (four million generations) and the first 1000 saved trees (i.e. 25%) were discarded as burn-in. The Bayesian 50% majority rule consensus tree, rooted with *P. testacea*, is presented with both Bayesian and parsimony branch support superimposed (Fig. 1). The molecular data group, with high support, multiple accessions of each species according to species determination based on morphology. The single exception is one accession of *P. himalayana*, which falls out nested within a *P. vallesiaca* clade. The latter consists of strongly supported subclades. *Psora tenuifolia* is strongly supported as phylogenetic sister species to *P. altotibetica*. The *P. tenuifolia* – *P. altotibetica* clade is in turn sister to the *P. vallesiaca* clade. A clade consisting of *P. hyporubescens* Timdal and *P. pacifica* Timdal is also strongly supported. Apart from this, the molecular data does not support any further inter-species relationships.

Discussion

Psora tenuifolia is the sister species of *P. altotibetica* in our phylogeny (Fig. 1). It differs in having thinner, ascending, less pruinose, more white-edged squamules containing zeorin and usually norstictic acid, and in having a well-developed lower cortex composed of mainly anticlinally oriented hyphae which are densely covered by calcium oxalate crystals (Timdal 1986). *Psora tenuifolia* was previously known from Alaska and arctic Canada (Timdal 1986) and from Yakutia (Zhurbenko 2003). In two specimens (Obermayer 4525 and 5236) norstictic acid was not detected by TLC; these specimens
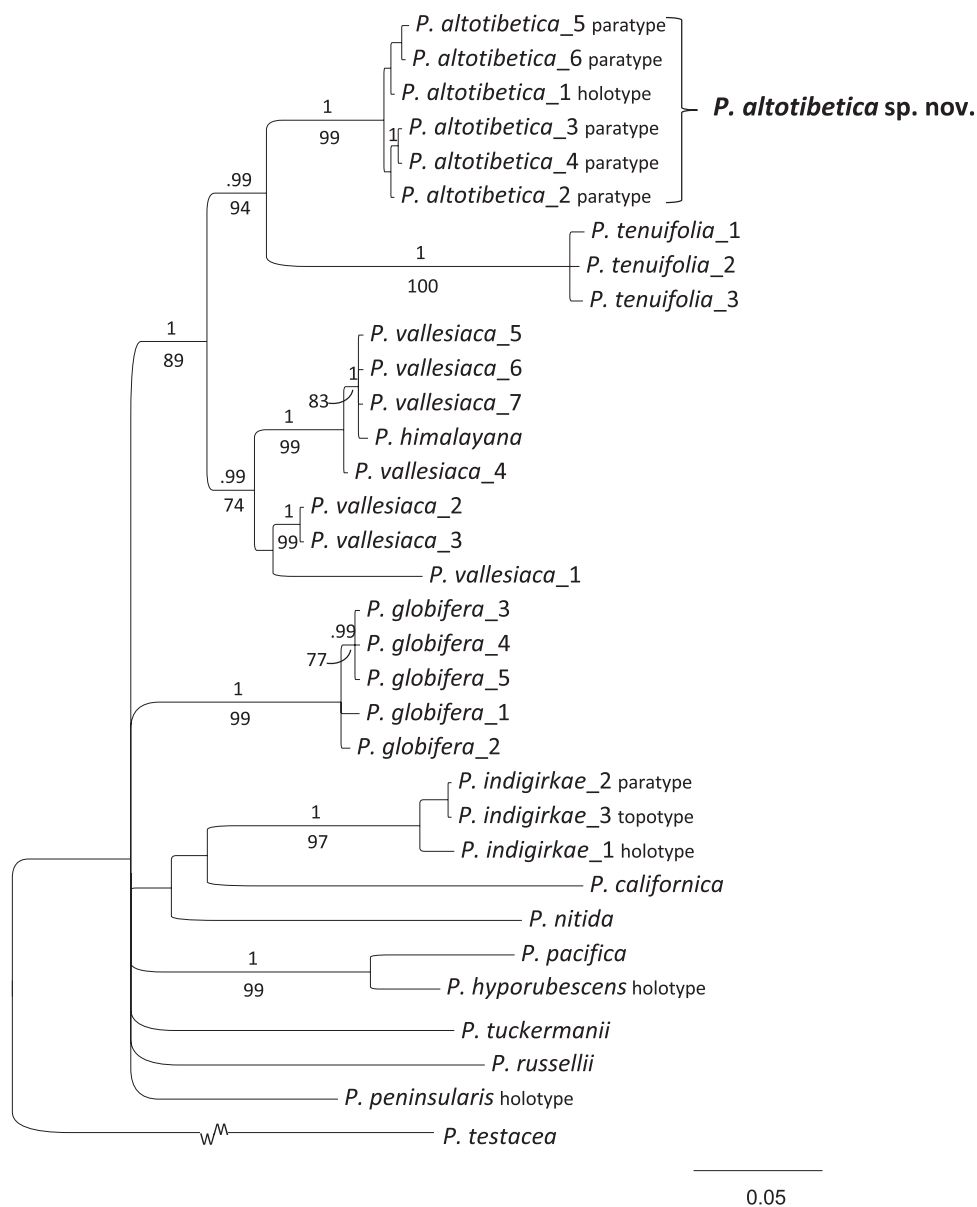
Figure 1. The Bayesian 50% majority rule consensus tree based on a concatenated alignment of ITS and mtSSU sequences of 33 accessions of 14 *Psora* species (see Table 1). Parsimony jackknife support values above 50% are shown below branches and Bayesian posterior probabilities above. The curly branch leading to *P. testacea* has been shortened to reduce the size of a broad figure.

represent a previously unknown chemotype of *P. tenuifolia*. The species is here reported as new to China (Sichuan and Xizang) and the Himalayas. One collection (Obermayer 4487) is a mixture of *P. altotibetica* and *P. tenuifolia*.

Psora vallesiaca is the phylogenetic sister species of the *P. altotibetica - P. tenuifolia* clade (Fig. 1). It differs from *P. altotibetica* in having less pruinose squamules with a more up-turned and white-edged margin, and in containing norstictic acid. It is morphologically more similar to *P. tenuifolia* than to *P. altotibetica*; see Timdal (1986) for discussion on the differences between *P. tenuifolia* and *P. vallesiaca*. Two specimens here identified as *P. vallesiaca* (Obermayer 4482 and 5279) do not contain lichen substances and were first thought to represent *P. himalayana*. We obtained sequences from the latter, and it clusters with *P. vallesiaca* in our phylogeny (Fig. 1, specimen *P. vallesiaca* 3). The sequence of *P. himalayana* downloaded from GenBank (specimen not examined by us, chemistry unknown) also clusters with those of *P. vallesiaca* in our phylogeny. Timdal (1986) remarked that *P. himalayana* and *P. vallesiaca* are sometimes difficult to distinguish morphologically, but that the chemistry is diagnostic. It now seems that there is a norstictic acid deficient chemotype of *P. vallesiaca*, making the taxonomic status of *P. himalayana* in need of revision. *Psora vallesiaca* is here reported as new to China (Sichuan and Xizang) and the Himalayas.

The other species of *Psora* known from the Himalayas, *P. decipiens*, differs in having orange to red or rose, more regularly rounded squamules with a usually more upturned and crenulate margin, in having strictly marginal apothecia, and in lacking lichen substances or more rarely containing norstictic acid or very rarely hyposalazinic and hypostictic acids (see, e.g., Timdal 2002).

Nine other *Psora* species contain gyrophoric acid (Timdal 1986, 2002, Timdal and Zhurbenko 2004). Five of those are in our phylogeny (Fig. 1; i.e. *P. californica* Timdal, *P. hyporubescens*, *P. indigirkae*, *P. nitida* Timdal, and *P. pacifica*), but none are closely related to *P. altotibetica*. *Psora indigirkae* is the morphologically most similar *Psora* species, but differs from *P. altotibetica* in having brown apothecia, often with a reddish hue, larger ascospores (14–17 × 7–8 µm; n=50), and in its secondary chemistry: gyrophoric acid is accompanied by bourgeanic acid (Timdal and Zhurbenko 2004). None of the four remaining gyrophoric acid containing *Psora* species are morphologically similar to *P. altotibetica*. *Psora nipponica* (Zahlbr.) Gotth.Schneider and *P. rubiformis* (Ach.) Hook. have, e.g., a lower cortex similar to that of *P. tenuifolia* (Timdal 1986), *P. montana* Timdal has medium to castaneous brown squamules and brown, more plane apothecia, and *P. russellii* (Tuck.) A.Schneider has larger, more rounded squamules, often with a central depression, and almost always contains norstictic acid as the major compound (gyrophoric acid being minor to trace or lacking).

Taxonomy

Psora altotibetica Timdal, Obermayer & Bendiksby, sp. nov.
Mycobank: MB 816840
Fig. 2

Diagnosis. Similar to *Psora indigirkae*, but apothecia black, ascospores shorter, and bourgeanic acid absent from the thallus.

Figure 2. *Psora altotibetica*, part of holotype. Scale bar = 1 mm.

Type. CHINA. Xizang: Himalaya Range, 165 km SSE of Lhasa, 40 km W of Lhünze, little village on way to Nera Tso (=Ni La Hu), 28°23'N, 92°05'E, 4300–4400 m alt., dry-valley, N-exposed dry slopes, on the ground, 1 Aug 1994, W.Obermayer 5282 (holotype: GZU!).

Description. Thallus squamulose; squamules up to 3 mm wide, rounded, adnate, dispersed to adjacent, weakly concave to plane; upper surface medium brown, dull, becoming moderately to densely pruinose, smooth when young, later with fissures in the cortex; margin concolorous with upper side or partly white, straight to slightly down-turned or slightly up-turned, entire; upper cortex 40–60 μm thick, composed of pale brown, thin-walled hyphae with rounded lumina, containing remnants of algae throughout (Chlor-zinc-iodine!), containing crystals of gyrophoric acid (assumedly, dissolving in K) and (in pruinose squamules) crystals of calcium oxalate; epinecral layer hardly developed. Medulla not amyloid, containing both lichen substances (dissolving in K) and calcium oxalate; lower cortex poorly developed, composed of mainly periclinally oriented, pale brown hyphae, not containing crystals; lower surface brown. Apothecia up to 1.2 mm diam., marginal or submarginal on the squamules, plane and indistinctly marginate when young, soon becoming convex and immarginate, black, epruinose or faintly white pruinose at the margin. Hypothecium colourless, containing crystals of calcium oxalate; epihymenium yellowish brown, containing orange crystals dissolving in K, K+ purple. Ascus clavate, with a well-developed, amyloid tholus containing a deeper amyloid tube, lacking an ocular chamber (*Porpidia*-type); ascospores ellipsoid, simple, hyaline, 9–14 × 5–7 μm (n = 30). Conidiomata not seen.

Figure 3. *Psora altotibetica*, known distribution. Open circle = holotype locality.

Chemistry. Gyrophoric acid (by TLC); upper cortex and medulla K–, C+ faintly red, KC+ faintly red, P–.

Habitat and distribution. The species is terricolous and known from nine localities in China (Tibet) and Nepal at altitudes between 4230 and 5000 m (Fig. 3).

Etymology. The name refers to its occurrence at high altitude in the Tibetan part of the Himalayas.

Other specimens examined. *Psora altotibetica*. CHINA. Xizang: Shegar, Rongbuk Valley, S of Rongpu-si, 5000 m alt., moraine hill, 7 Jul 1993, T.Engelskjøn T-030b (TROM L-42812); 5000 m alt., rock base, patchwise on silt, 7 Jul 1993, T.Engelskjøn T-036 (TROM L-42813); Upper Tsangpo basin, N of Saga, 29°34'N, 85°15'E, alt. 4760 m, 28°W, grazed stony relicts of *Juniperus* dwarf-scrub, on ground, 26 Aug 1993, G.Miehe. & S.Miehe 9573/23/02 (GZU) & 9573/23/03 (GZU); 120 km SSW of Quamdo (=Changtu), 10 km S of Bamda, 30°09'N, 97°17'E, 4500–4700 m alt., on mosses, overhang, NNW-exp., 6 Aug 1994, W.Obermayer 3967 (GZU); Himalaya Range, 160 km S of Lhasa, dry valley of Kuru river, 10 km NW Lhozag, 28°24'N, 90°39'E, 4230 m alt., N-exposed steep rocks in a glen, on soil, 17 Jul 1994, W.Obermayer 4350 (GZU) & 4365 (GZU); Himalaya Range, 170 km S of Lhasa, between Lhozhag and Lhakhang Dzong, W-facing slopes of Dhalari mountain, 28°20'N, 90°58'E, 4300 m alt., NNW-exposed, ±underhang, 20 Jul 1994, W.Obermayer 4485 (GZU) & 4502 (GZU); Himalaya Range, 170 km SE of Lhasa, 110 km SSE of Tsetang (Nedong), 28°35'N, 92°23'E, 4700 m alt., alpine meadows with *Kobresia pygmaea*, ground with *Ochotona*-burrows, 26 Jul 1994, W.Obermayer 4981 (GZU); Himalaya Range, 130 km SE of Lhasa, 50 km SSE of Tsetang (Nedong), on way to the pass Putrang La, 28°52'N, 92°06'E, 4400 m alt., dry slope, on soil, 2 Aug 1994,

W.Obermayer 5223 (GZU). NEPAL. Mahalangur Himal, Khumbu, Moränen des Khumbu-Gletschers bei Lobuche, 4950–5000 m alt., Sep 1962, J.Poelt 1138 (GZU).

Psora himalayana. PAKISTAN. Karakorum, Naz Bar (Yasin), 36°17–25'N, 73°0–17'E, 3400–3470 m alt., subalpine Juniperus communis dwarf-scrub with Juniperus macropoda trees in Seriphidium maritimum steppe; on silt in rock crevices, 11 Sep 1990, G.Miehe & S.Miehe 3529 (GZU); Karakorum, Baltistan, Haramosh Range, "Alm" Pakora SE Ganto La, 35°41'N, 75°21'E, 3600–3800 m alt., pasture and rocks around the alm, rocky slopes, 3 Jul 1991, J.Poelt K91-416 (GZU).

Psora indigirkae. RUSSIA. Sakha Republic: Momskii region, along the river Indigirka, c. 48 km NNW of Tyubelyakh, 65°48'N, 142°53'E, 200–300 m alt., on calcareous soil in limestone cliffs, 20 Jul 1992, R.Haugan & E.Timdal YAK17/24 (O L-19086); c. 54 km N-NNW of Tyubelyakh, 65°51'N, 143°01'E, 200–300 m alt., 20 Jul 1992, M.P.Zhurbenko 92185 (O L-118686).

Psora tenuifolia. CHINA. Sichuan: Tibetan fringe mountains (=Hengduan Shan), Shaluli Shan, on the outskirts of Yajiang, 200 m E of the river Yalong Jiang, 30°02'22"N, 101°00'16"E, 2610 m alt., NE-exposed dry slopes with schist outcrops, on thin soil crust (over schist), 12 Aug 2000, W.Obermayer 9791 (GZU); Xizang: Himalaya Range, 170 km S of Lhasa, between Lhozhag and Lhakhang Dzong, W-facing slopes of Dhalari mountain, 28°20'N, 90°58'E, 4300 m alt., NNW-exposed, ±underhang, 20 Jul 1994, W.Obermayer 4487 (GZU); Himalaya Range, 175-180 km S of Lhasa, between Lhozhag and Lhakhang Dzong, Kuru river valley, pass, 28°12'N, 91°00'E, 3600 m alt., on soil, 21 Jul 1994, W.Obermayer 4525 (GZU); Himalaya Range, 210 km SE of Lhasa, 15 km ESE of Lhünze, way to Qayü, dry-valley of Subansiri, 28°24'N, 92°37'E, 4100–4200 m alt., on soil (+ mosses), 31 Jul 1994, W.Obermayer 5236 (GZU).

Psora vallesiaca. CHINA. Sichuan: Shalui Shan Mts, 30 km NE Batang, S Yidun, 30°16'N, 99°25'E, 3750–3800 m alt., on marble outcrops, soil, 25 Jun 1994, W.Obermayer 3227 (GZU); Xizang: Himalaya Range, 165 km SSE of Lhasa, 40 km W of Lhünze, little village on way to Nera Tso (=Ni La Hu), 28°23'N, 92°95'E, 4300–4400 m alt., dry-valley, N-exposed dry slopes, on the ground, 1 Aug 1994, W.Obermayer 5279 (GZU); Himalaya Range, 170 km S of Lhasa, between Lhozhag and Lhakhang Dzong, W-facing slopes of Dhalari mountain, 28°20'N, 90°58'E, 4300 m alt., NNW-exposed, ±underhang, 20 Jul 1994, W.Obermayer 4482 (GZU); PAKISTAN. Karakorum, Baltistan, Haramosh Range, between "Alm" Matumdus (3620 m, 35°42'N, 75°23'E) and Hemasil in the Basna Valley, 3100–3200 m alt., 7 Jul 1991, J.Poelt K91-705 (GZU); Karakorum, Baltistan, Basna valley, Basnald, 2500 m alt., 10 Jul 1991, J.Poelt s.n. (GZU).

Acknowledgments

References

Bendiksby M, Timdal E (2013) Molecular phylogenetics and taxonomy of *Hypocenomyce* sensu lato (Ascomycota, Lecanoromycetes) – extreme polyphyly and morphological/ecological convergence. Taxon 62: 940–956. doi: 10.12705/625.18

Bendiksby M, Haugan R, Spribille T, Timdal E (2015) Molecular phylogenetics and taxonomy of the *Calvitimela aglaea* complex (Tephromelataceae, Lecanorales). Mycologia 107: 1172–1183. doi: 10.3852/14-062

Culberson CF (1972) Improved conditions and new data for the identification of lichen products by a standardized thin-layer chromatographic method. Journal of Chromatography 72: 113–125. doi: 10.1016/0021-9673(72)80013-X

Culberson CF, Johnson A (1982) Substitution of methyl tert.-butyl ether for diethyl ether in the standardized thin-layer chromatographic method for lichen products. Journal of Chromatography 238: 483–487. doi: 10.1016/S0021-9673(00)81336-9

Ekman S, Blaalid R (2011) The devil in the details: interactions between the branch-length prior and likelihood model affect node support and branch lengths in the phylogeny of the Psoraceae. Systematic Biology 60: 541–561. doi: 10.1093/sysbio/syr022

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

Menlove JE (1974) Thin-layer chromatography for the identification of lichen products. Bulletin of the British Lichen Society 34: 3–5.

Poelt J (1958) Die lobaten Arten der Flechtengattung *Lecanora* Ach. sensu ampl. in der Holarktis. Mitteilungen der Botanischen Staatssammlung München 2: 411–589.

Poelt J, Vězda A (1981) Bestimmungsschlüssel europäischer Flechten. Ergänzungsheft II. Bibliotheca Lichenologica 16: 1–390.

Schneider G (1980) Die Flechtengattung *Psora* sensu Zahlbruckner – Versuch einer Gliederung. Bibliotheca Lichenologica 13: 1–291.

Timdal E (1984) The delimitation of *Psora* (Lecideaceae) and related genera, with notes on some species. Nordic Journal of Botany 4: 525–540. doi: 10.1111/j.1756-1051.1984.tb02059.x

Timdal E (1986) A revision of *Psora* (Lecideaceae) in North America. The Bryologist 89: 253–275. doi: 10.2307/3243197

Timdal E (1991) *Anamylopsora*, a new genus in the Lecideaceae. Mycotaxon 42: 249–254.

Timdal E (1992, '1991') A monograph of the genus *Toninia* Massal. (Lecideaceae, Ascomycetes). Opera Botanica 110: 1–137.

Timdal E (2002) *Psora*. In: Nash III TH, Ryan BD, Gries C, Bungartz F (Eds) Lichen Flora of the Greater Sonoran Desert Region, Vol. 1. Lichens Unlimited, Arizona State University, Tempe, 418–430.

Timdal E (2008) *Romjularia*. In: Nash III TH, Gries C, Bungartz F (Eds) Lichen Flora of the Greater Sonoran Desert Region, Vol. 3. Lichens Unlimited, Arizona State University, Tempe, 287–289.

Timdal E, Zhurbenko MP (2004) *Psora indigirkae*, a new species from eastern Siberia, with key to the Asian species of *Psora*. Symbolae Botanicae Upsalienses 34: 25–29.

Zhurbenko MP (2003) New and rare lichen species (Lichenes) from Sakha-Yakutiya Republic and Magadan region. Botanicheskii Zhurnal 88: 111–118.

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. doi: 10.1017/S0024282999000663

Wang L-S (2012) 中国云南地衣 [Zhongguo Yunnan di yi; Lichens of Yunnan in China]. Shanghai ke xue ji shu chu ban she, Shanghai, 1–238. [In Chinese]

Wei JC (1991) An enumeration of lichens in China. International Academic Publishers, Beijing, 1–278.

White TJ, Bruns T, Lee S, Taylor JW (1990) Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In: Innis MA, Gelfand DH, Sninsky JJ, White TJ (Eds) PCR protocols: A guide to methods and applications. Academic Press, New York, 315–322. doi: 10.1016/b978-0-12-372180-8.50042-1