Peeking into the black box – integrated taxonomy of Archaeorhizomycetes

Faheema Kalsoom Khan
Uppsala University
Kerri Kluting
Uppsala University
Jeannette Tångrot
Umeå University
Hector Urbina
Uppsala University
Tea Ammunet
Uppsala University
Shadi Eshghi Sahraei
Uppsala University
Martin Rydén
Uppsala University
Martin Ryberg
Uppsala University
Anna Rosling (Anna.Rosling@ebc.uu.se)
https://orcid.org/0000-0002-7003-5941

Research

Keywords: Archaeorhizomyces victor nom. seq., Archaeorhizomyces secundus nom. seq., biodiversity, dark matter fungi/dark taxa, integrative taxonomy, voucherless taxa

Posted Date: July 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-39531/v1

Version of Record: A version of this preprint was published on November 3rd, 2020. See the published version at https://doi.org/10.1186/s43008-020-00045-9.

Abstract

Due to their submerged and cryptic lifestyle the vast majority of fungal species are difficult to observe and describe morphologically and many remains known to science only from sequences detected in environmental samples. The lack of rules to delimit and name most fungal species is a staggering limitation to communication and interpretation of ecology and evolution in kingdom *Fungi*. Here we use environmental sequence data as taxonomical evidence and take an integrated taxonomic approach by combining phylogenetic and ecological data to generate and test species hypothesis in the class Archaeorhizomycetes (Taphrinomycotina, Ascomycota). Based on environmental amplicon sequencing we recognize 68 distinct phylogenetic species hypotheses (PSHs) of Archaeorhizomycetes, including the two described species in the class, in a well-studied Swedish pine forest podzol soil. Nine of the species hypotheses, including the more abundant ones, are supported by long read data and represent 78% of the sequenced Archaeorhizomycetes community.

Among well supported sister PSHs, significantly differential distribution in the soil profile provide additional ecological evidence supporting the identification of two novel species for which we provide molecular diagnostics and propose names. Our analysis indicates that frequent and abundant taxa can be phylogenetically resolved in environmental samples, while rare taxa remain un-captured at our sampling and sequencing intensity. While environmental sequences cannot be automatically translated to species hypothesis, they can be used to generate phylogenetic and ecological evidence supporting species recognition of abundant species without physical specimens.

Introduction

Species are fundamental units of biodiversity. There is general support for species as entities with their own evolutionary identity and fate, i.e., separately evolving metapopulation lineages, the principle upon which the unified species concept is built (De Queiroz, 1998, 2007). Nevertheless, it remains challenging to operationally delimit species. It is increasingly evident that morphology-based taxonomy is not a feasible approach to systematically classify most eukaryotic microorganisms, including fungi (Hibbett, 2016) and protists (Keeling & Burki, 2019). Diagnostic morphological characters may not be readily available even in some macroscopic animals (Fišer et al., 2018), calling for integrated taxonomy including DNA-based methods for delimitation and classification of cryptic species (Dayrat, 2005; Padial et al., 2010; Yeates et al., 2011). Often, molecular data play a key role in taxonomy, for example, by utilizing phylogenetic evidence to test hypotheses of species boundaries that were developed using morphological evidence. Over the last decade, DNA-based species delimitation has resolved hundreds of cryptic species in algae, and species diagnosis based solely on DNA sequence data is increasingly accepted (Leliaert et al., 2014).

In fungal systematics, the about 140 000 described species are predicted to represent less than one tenth of the true fungal diversity (Hibbett, 2016). The submerged lifestyle and microscopic, morphological simplicity of fungi complicates their discovery and formal nomenclature because the physical specimens

required for naming species in accordance with the International Code of Nomenclature for Algae, Fungi and Plants (the *Code*) can often not be isolated or identified. Many of the undescribed fungi are known from environmental studies using amplicon sequencing (Nilsson, Anslan, et al., 2019) and are among those referred to as “dark taxa” (Ryberg & Nilsson, 2018). Efforts to culture and image “dark taxa” are scarce, and as in the case of the current study, often fruitless, in part because we know so little about these organisms. One approach to identifying hypothetical species using environmental sequence data is via a cluster-based reference system that includes both environmental and ex-type sequences, for example “virtual taxa” of the Maarj*AM* database (Öpik et al., 2010), or “species hypotheses” of the UNITE database (Kõljalg et al., 2013; Nilsson, Larsson, et al., 2019). It is impossible to overstate the value and utility of such databases (for example, as a tool for identification of environmental sequences), however, clustering methods cannot replace taxonomy since outputs are context-dependent and similarity cutoffs remain arbitrary (Ryberg, 2015). Integrated taxonomy may resolve this limitation by using multiple lines of evidence to support or reject species hypothesis, thus building towards well-supported delimitation of species (Leliaert et al., 2014). Naming “dark taxa” is a tool to communicate their presence and gather an understanding of their biology.

One particularly enigmatic lineage filled with “dark taxa” is the fungal class Archaeorhizomycetes (Taphrinomycotina, Ascomycota) (Rosling et al., 2011). Sequences of Archaeorhizomycetes are frequently observed in environmental DNA samples from soil and roots and have been detected in more than 100 environmental studies (Hibbett, 2016; Menkis et al., 2014; Porter et al., 2008). Based on clustering of published environmental DNA sequences, the class is likely to comprise at least 500 species with a wide geographical distribution, occurring in terrestrial environments around the globe (Menkis et al., 2014). The annotated fungal sequence database UNITE (version 8.0) (Kõljalg et al., 2013) currently has 195 species hypotheses classified as Archaeorhizomycetes at the 1.5% dissimilarity threshold. Richness estimates based on clustering at 98.5% of environmental sequences available as Short Read Archive data indicate that Archaeorhizomycetes may encompass as many as 16,231 OTUs (Lücking & Hawksworth, 2018). Previously known as Soil Clone Group 1 based on environmental sequence data (Porter et al., 2008; Schadt et al., 2003), the first living specimen was isolated into pure culture and validly named in 2011: *Archaeorhizomyces finlayi* Rosling & T. James (Rosling et al., 2011). To date, only one additional species, *A. borealis* Menkis, T. James & Rosling (Menkis et al., 2014), has been isolated and described. Even if the species richness estimates based on environmental sequence data are inflated due to sequencing errors or artefacts of OTU clustering methods (Lücking & Hawksworth, 2018), the two name-baring species in the class certainly represent a small fraction of its true diversity. Further, while little is currently known about the ecology and lifestyles of species in the Archaeorhizomycetes, studies using environmental sequences indicate that the class is both abundant and diverse in a range of terrestrial ecosystems (Carrino-Kyker et al., 2016; Clemmensen et al., 2015; Cruz-Paredes et al., 2019; Fernández-Martínez et al., 2017; Geml et al., 2014; He et al., 2019; Kluting et al., 2019; Kyaschenko et al., 2017; Levy-Booth et al., 2019; Maghnia et al., 2017; Pinto-Figueroa et al., 2019; Rosling et al., 2013; Sun et al., 2016). In several studies its abundance is linked to ecological patterns such as succession and nutrient availability highlighting the urgent need to recognize species in the class to facilitate

communication and identify ecological patterns across studies. Despite extensive culturing efforts, no further cultures of Archaeorhizomycetes were successfully isolated in the current study. To estimate diversity of Archaeorhizomycetes, as well as many other under-described fungal lineages, environmental sequence analysis needs to be developed for species hypothesis testing. High abundance in environmental sequence datasets and access to reference sequences for all described species makes Archaeorhizomycetes an excellent case to develop novel taxonomic approaches for describing species known only from environmental sequences, without the risk of creating confusion in respect to species described based on morphology or other kind of characters.

In nature, niche partitioning is an important process that allows co-existence of species with similar resource requirements (Schoener, 1974) such as soil fungal communities (Peay et al., 2008). In accordance with the expectation that closely related species have separate niches in at least one dimension, vertical separation has been demonstrated for sister species of soil fungi as a result of competitive avoidance (Mujica et al., 2016). In the current study we use niche separation as a dimension for ecological species recognition. We designed the current study to advance the field of fungal systematics by using environmental sequence data to generate evidence for taxonomic description of species from a well-studied Swedish pine forest. At this site, previous studies have revealed operational taxonomic units (OTUs) within Archaeorhizomycetes that are both abundant and vertically stratified (Fransson & Rosling, 2014; Lindahl et al., 2007). We combined long and short read amplicon sequences in order to propose phylogenetic species hypotheses (PSH) for the Archaeorhizomycetes diversity at the site. Here, we use environmental sequence data by building phylogenetic species hypotheses that we test using ecological species recognition, i.e. niche separation as defined by read abundance recorded from visually distinct Podzol soil horizons. We hypothesized that sister PSHs would have different realized niches and tested this using short read relative abundance across soil horizons as a proxy for niche distribution. By combining two lines of evidence we firmly delimit and describe two novel species.

Material And Methods

Field site and sampling

Soil samples were collected in mid-October 2013 from Ivantjärnsheden field station close to Jädraås (60°49‘N, 16°30’E, altitude 185 m), a well-documented field site in central Sweden (Persson, 1980) with *Pinus sylvestris* L. overstory and an understory of ericaceous dwarf shrubs (*Calluna vulgaris* (L.) Hull and *Vaccinium vitis-ideae* L.) and mosses [*Dicranum majus* Turner and *Pleurozium schreberi* (Bridel) Mitten]. The stand was naturally regenerated after tree felling in 1957 and thinned before the onset of an experimental field study conducted between 1974 and 1990 [experiment Ih2 (9802)] (Axelsson & Bråkenhielm, 1980) (Fig. S1). Between those years, 30 × 30 m plots received one of four different treatments: irrigation and fertilization (IF), irrigation (I), un-manipulated control (O), as well as clear cut plots (CU). The clear-cut was part of a second study following the initial experiment and represented one plot from each previous treatment. Forest has since started to regenerate in CU plots. To account for

nities, we collected five soil cores (5 cm diameter and 15 cm

deep) in each plot after visually dividing the plot into four quadrants; one core each were taken from the middle of each quadrant and from the middle of the plot after peeling back the top shrub and moss layer (incl. most of the litter layer). Soil cores were separated into visually distinct podzol soil layers: organic soil (O, approximately 0–5 cm depth), mineral elluvial soil (E, 5–8 cm) and mineral illuvial soil (B, 8–15 cm), before pooling the layers for each plot. We sampled four types of plots: three treatments (I: 11, 16, 17; IF: 4, 12, 13; O: 21, 23, 24) as well as the clear cuts (CU: 1, 2, 3) (Fig. S1, Table S1). This sampling rendered a total of 36 soil samples that were separately homogenized in Ziploc bags before separating a 15 mL sample from each that was transported back to the laboratory on ice and stored at -20 °C. Remaining soil was stored cold awaiting fungal isolation efforts.

Attempts to isolate Archaeorhizomycetes from roots

An extensive culturing effort was performed during summer and fall of 2013 and 2014, attempting to isolate species in Archaeorhizomycetes from Ivantjärnsheden field station. New soil samples were collected for isolations in 2014. The isolation protocol was based on previous successful isolations of Archaeorhizomycetes (Grelet et al., 2010; Menkis et al., 2014; Rosling et al., 2011). Roots were separated from soil samples collected as described above. Under a stereomicroscope, healthy and dead root tips about 1 cm long were selected, and superficially disinfected in 30% peroxide for 30 s and rinsed in sterile, deionized water for 2 min. Then, root tips were cut in ~ 2 mm long pieces and placed in 10 cm diameter petri dishes containing modified Melin Norkrans media (MMN) (Marx, 1969) with half strength of glucose. Up to 20 root fragments were placed in each dish and incubated up to five months in darkness at room temperature. Since available Archaeorhizomycetes cultures are slow-growing, all fungal colonies that grew from the root fragment during the first three months were discarded by cutting them out of the plate using a sterile scalpel. Fungal colonies emerging from the root tips after this period were sub-cultivated on new dishes with the same MMN media. Days after plating were recorded for all transfers. Approximately 2,000 root tips were surface-sterilized and plated, resulting in just over 160 cultures somewhat resembling those of *A. borealis* or *A. finlayi*. DNA was extracted from these for amplification of the rDNA ITS and LSU region, using the forward primers ITS1 (White et al., 1990) and reverse LR3 (Hopple Jr & Vilgalys, 1994), and then amplified and sequenced by Sanger technology, also including the reverse primer ITS4 (Gardes & Bruns, 1993). The 98 usable sequences generated were assigned to SH in UNITE (Kõljalg et al., 2013) and BLASTed against GenBank (Altschul et al., 1990). None of the sequences matched Archaeorhizomycetes, but many were identified as root associated fungi in the genera *Coemansia, Meliniomyces*, and *Phialocephala* (among others). The sequences were deposited in GenBank (Sayers et al. 2020), with accession numbers MH843963–MH844060.

Soil DNA extraction

From each composite soil sample, two sub-samples of approximately 0.5 g wet weight were collected into separate 2.0 mL microtube containing 750 µL of lysis buffer (Xpedition™ Soil/Fecal DNA miniprep, Zymo Research Corporation, Irvine, California, USA). Total soil DNA extraction was performed after homogenization for 30 s using a TerraLyser™ (Zymo Research Corporation), following the manufacturer's protocol. DNA concentration and integrity were verified by 0.8% agarose gel electrophoresis on 0.5% Tris

Acetate-EDTA buffer (Sigma-Aldrich, St. Louis, Missouri, USA) stained with 1 × GelRed™ (Biotium Inc., Hayward, California, USA). Two different sequence datasets were generated from these samples: 1) a "phylogenetic" dataset of long amplicons using PacBio SMRT, and 2) an "ecological" dataset based short read metabarcoding using IonTorrent. The longer reads in the “phylogenetic” sequence dataset provides enough information to resolve deeper nodes in the Archaeorhizomycetes tree while the “ecological” dataset provides sufficient sequencing depth and replication to test for species specific realized niches.

Generating the “phylogenetic” sequence dataset

Approximately 1500 bp of the rDNA ITS and LSU region was amplified from all soil DNA extracts using the primer set ITS1F (Gardes & Bruns, 1993) and LR5 (Hopple Jr & Vilgalys, 1994), with Phusion High-Fidelity DNA polymerase (ThermoFisher science, Waltham, Massachusetts, USA). We run a thermo-cycling protocol as follows: an initial denaturation at 95 °C for 10 min followed by 30 cycles of denaturation at 95 °C for 45 s, annealing at 58 °C for 45 s, and elongation at 72 °C for 90 s, with a final elongation at 72 °C for 10 min. A total of 5, 8 and 3 samples successfully amplified for O, E and B horizons, respectively (Table S1). PCR products from the separate soil horizons were pooled and quantified by Nanodrop 2000C (ThermoScientific, Waltham US), and gel electrophoresis was performed to generate three amplicon libraries (SwO, SwE and SwB) for sequencing at SciLifeLab/NGI (Uppsala, Sweden) on a PacBio RS II system (Pacific Biosciences, Menlo Park, CA, USA). Sequences were delivered to us as error-corrected FASTQ files. Raw reads for the current study are available in ENA (samples ERS3508481- ERS3508483).

The “phylogenetic” sequences dataset was filtered in four steps for downstream phylogenetic analysis of the Archaeorhizomycetes diversity at the site. First, the raw sequence reads were filtered and trimmed using the tool cutadapt (version 1.18) (Martin, 2011) to keep only reads with both primers present, and to remove the actual primer sequences from the reads. Amplicons sequenced in reverse, were reverse complemented before continuing the analyses. Secondly, a quality-controlled, long read sequence dataset of all amplified sequence variants (ASVs) was generated using DADA2 (version 1.9.3) (Callahan et al., 2016). Default parameters were used for filtering the reads, but discarding sequences with more than 12 "expected errors" (maxEE = 12). This value was chosen relatively high due to the long reads. In addition, sequences shorter than 50 bp and longer than 3,000 bp were removed as spurious. The option 'pooled' was used both for denoising and chimera detection, in order to consider the sequences in all three samples/data sets together. For chimera removal, the option 'AllowOneOff' was used, meaning that a sequence will be identified as chimeric also if it is one mismatch or indel away from an exact bimera. Thirdly, the tool ITSx (version 1.1-beta) (Bengtsson-Palme et al., 2013) was used to identify different regions of ribosomal RNA within each ASV. Seven of the ASVs were considered incomplete reads and removed from further analysis because no ITS2 region was detected. BLASTing of these sequences suggest that they represent protists. We used Geneious (version 11.1.4) (Kearse et al., 2012) to align the LSU region of the 276 ASVs with the LSU sequences of *A. finlayi* (JF836022), *A. borealis* (KF993708) and the uncultured lineage GS31 (KY687760) which represents a sister class of Archaeorhizomycetes (Tedersoo et al., 2017). The Geneious alignment algorithm set to the Global alignment with free end gaps

option, a cost matrix of 65% similarity (5.0/-4.0) with gap open penalty 12 and gap extension penalty 3. ASV taxonomy assignments with a confidence value of 0.8 or higher was added in the alignment to the ASV name using a customized script using class when assigned, or else phyla or domain. The alignment was visually inspected, and three ASVs were removed because two were truncated (ASV_279 and ASV_196) and one (ASV_280) was chimeric with a possible protist sequence. For the remaining ASVs, taxonomy was predicted with the ITS2 region using the SINTAX classifier (Edgar, 2010) as implemented in VSEARCH (version 2.10.4) (Rognes et al. 2016), and the USEARCH/UTAX reference dataset (version 8.0) (UNITE Community, 2019), available from the UNITE database (Kõljalg et al., 2013). This reference dataset was customized by replacing unassigned species level taxonomy with UNITE SH when available.

A final alignment across 1,186 bp from 273 ASV and 3 reference sequences was uploaded in the CIPRES portal (Miller et al., 2010) and a maximum likelihood tree was built using the online version of RAxML XSED2 (version 8) (Stamatakis, 2014). The GTRGAMMA model was used with 1000 iterations for the calculation of bootstrap support. The tree was visualized in FigTree (version 1.4.4) and rooted with two Rozellomycota sequences (ASV_229 and ASV_237). Taxonomy predictions with a confidence value of 0.8 or higher using class when available, or else phyla or domain was added to the ASV name in the tree file using a customized script. Forty-two ASVs representing Archaeorhizomycetes were identified as those forming a well-supported clade together with *A. borealis* and *A. finlayi*, and distinct from GS31 (Fig. S2). The full-length sequences of these 42 ASVs were aligned in Geneious as described above and a ML tree was generated in the CIPRES portal as described above, and ASVs on long branches were visually inspected in the alignment. This procedure identified ASV_135 a chimera formed from parent sequences ASV106 and ASV8, and was removed from the dataset. With these steps, we generated a high quality Archaeorhizomycetes rDNA sequence dataset consisting of 41 ASVs ranging in length from 1,373-1,801 bp and representing 52,274 reads (Fig. S2, Table S2, S3). In the end, 272 ASV sequences, together with the most specific taxonomy level with a confidence value of 0.8 or above, were published in GBIF (http://doi.org/10.15468/8zymuf).

Generating the “ecological” sequence dataset

The ITS2 region of the rRNA genes was amplified using primers gITS7 *forward* (Ihrmark et al., 2012) and modified ITS4m *reverse* (Rosling et al., 2016) (a 1:1 mixture of ITS4 (White et al., 1990) and a modified version of it: 5'–TCCTC**GC**CTTATTGATATGC–3'), with both primers containing adequate barcode sequences for single-ended sequencing (Table S1). Modifications on the reverse primer ITS4 were included to reduce its known bias against the class Archaeorhizomycetes (Schadt & Rosling, 2015). PCR was performed in a final volume of 20 µL reactions with 10–20 ng DNA, 1 × SSoAdvanced™ Universal SYBR® Green Supermix (Bio-Rad Laboratories, Hercules, California, USA), and 0.8 nM of each primer. PCR amplifications were carried out on a CFR96 Touch™ Real/Time PCR Detection system (Bio-Rad Laboratories) with 10 min pre-denaturation at 95 °C, 1 min DNA denaturation at 95 °C, 45 s annealing at 50, 54 and 58 °C (performed in separate tubes) to compensate for primer binding bias (Schmidt et al., 2013), 50 s of extension at 72 °C and 3 min final extension at 72 °C. Monitoring amplification by qPCR allowed us to adjust the number of cycles between 23–27 for each plate, in order to ensure that

amplification did not plateau and to minimize chimera formation. All reactions were carried out in parallel on the duplicate DNA extracts from each sample, combining all six PCR products before purification using the ZR-96 DNA Clean & Concentrator™-5 (Zymo Research Corporation). Duplicate quantification of PCR products was performed using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Life Technologies Corporation, Carlsbad, California, USA) on a TECAN F500 microplate reader and the band purity and length was checked by electrophoresis in 2% agarose gel 0.5 × TAE buffer. A sequencing library was prepared by pooling 35 ng PCR products from each sample, loaded onto a 318 chip for PGM Ion Torrent sequencing technology (Life Technologies Corporation, Carlsbad, CA, US) and sequenced at the National Genomics Infrastructure (NGI Uppsala; SciLifeLab, Uppsala, Sweden). The current study was sequenced on the same IonTorrent chip as four other studies with samples from different ecosystems including forest soils across Puerto Rico (Urbina et al., 2016), mixed deciduous forest in Indiana, US (Rosling et al., 2016) and two un-published datasets from west Africa (Benin and Burkina Faso) and central Asia (India and Kyrgyzstan), as well as a negative and positive control on this IonTorrent chip. The 4,805,942 raw sequence reads were demultiplexed by the sequencing facility and provided as 96 fastq files. Raw reads for the current study are available in ENA (samples ERS4600640-ERS4600675).

The software package DADA2 (version 1.14.0) (Callahan et al., 2016) for R (version 3.6.1) (R Core Team, 2019) was used to quality filter the raw reads and infer ASVs. Prior to ASV inference, primer and barcode sequences were trimmed from the reads using cutadapt (version 2.6 with Python version 3.7.5) (Martin, 2011), allowing for up to two mismatches, and requiring detection of both forward and reverse primers, as well as a minimum primer/read overlap of 10 bp. After primer trimming, 1,905,188 reads (39.6% of raw reads) were retained. The $fi < erAndTrim$ function of the DADA2 package was then used to trim the first 15 bp ($trimLeft = 15$), truncate the reads at any quality score of two ($tuncQ = 2$), and filter out reads that are less than 75 bp in length ($\min Len = 75$) or have more than five expected errors ($\max \exists = 5$). Next, error learning and denoising (via the $\leq arnErr$ or s and $dada$ functions, respectively) of the reads were done using default parameters, except that a homopolymer gap penalty of 1 was specified for the denoising step ($HOMOPOLYMER_{G}AP_{P}ENALTY = -1$), and the alignment band size was increased to 32 for both error learning and denoising steps ($BAND_{S}IZE = 32$). All 96 samples from the chip were pooled for ASV inference ($p\infty l = TRUE$) and $removeBimeraDenovo$ function was used to detect and remove chimeras ($method = "p\infty \leq d"$) and allowing for a one-off mismatch (`allowOneOff = TRUE), resulting in a total of 4,822 ASVs, representing 1,804,814 reads (37.6% of raw reads). For clarity ASVs generated from IonTorrent data are called itASVs though out the text.

Two methods were used to identify itASVs as putative Archaeorhizomycetes sequences. In the first method, taxonomy was predicted using a bootstrap support cutoff of 0.8 for all itASVs using the SINTAX classifier in USEARCH (v11.0.667_i86osx32) (Edgar, 2010) with a modified version of the UNITE USEARCH/UTAX reference dataset for all Eukaryotes (version 8.0) (UNITE Community, 2019). Since the SINTAX classifier cannot handle blank fields in taxonomy strings, the UNITE reference dataset was modified to replace blank fields with '<lowest_name_ided>_<rank>_Incertae_sedis', where < lowest_name_ided > is the identity of the lowest identified supra-taxonomic rank, and < rank > is the

taxonomic rank of the blank field, using the custom python script add_Incertae_sedis.py. These taxonomy predictions were also used to detect non-fungal sequences (based on a 0.8 bootstrap cutoff). In the second method, itASVs were BLASTed against a reference dataset consisting of the 41 Archaeorhizomycetes ASVs from the phylogenetic dataset, plus five sequences of the undescribed class-level lineage GS31 (GenBank Accession numbers KY687669, KY687744, KY687760, KY687776, and KY687785) (Tedersoo et al., 2017), using the blastn command line tool (version 2.9.0) (Camacho et al., 2009) with a minimum of 70% query coverage per high-scoring sepment pair (-$qcov_hsp_perc$70). An itASV was identified as putatively Archaeorhizomycetes if it was detected via either method (i.e., any itASV with a blast hit to one of the Archaeorhizomycetes ASVs of the phylogenetic dataset or with a taxonomic prediction to Archaeorhizomycetes). A total of 182 and 282 itASVs were identified across the 96 samples via the SINTAX method and the BLAST method, respectively, and all 182 itASVs detected via the SINTAX method were also identified via the BLAST method.

After removing non-fungal itASVs as well as the sequence from the positive control DNA (itASV_1), the dataset consisted of 4,461 itASVs and 1,680,044 reads, of which 282 itASVs/425,349 reads were putatively Archaeorhizomycetes (25.3% of the fungal reads) across the entire sequencing run. The $m\eta MDS$ function from the R package vegan (version 2.5-6) (Oksanen et al., 2013) was used to conduct an nMDS ordination of the complete itASV occurrence matrix across 96 samples (Fig. S3). The ordination was based on a distance matrix, calculated after transforming the matrix to per-sample relative abundances using the Bray-Curtis dissimilarity index, and a maximum of 200 random starts was specified. The results indicate that sample 24B may suffer from tag-switching with one of the samples from Puerto Rico, however we still choose to include this sample in all downstream analyses, since only seven rare Archaeorhizomycetes itASV were unique to sample 24B, and therefore the Archaeorhizomycetes community was not notably different from the other Jädraås samples. Finally, the ”ecological” dataset (an itASV count per sample matrix for the Jädraås samples) was generated by removing all samples from other studies and positive/negative controls. Further itASVs occurring only once across the 36 Jädraås samples were filtered out. This “ecological” dataset consisted of 1,664 itASVs (619,176 reads) (Supplementary datafile 1), and 123 of these itASVs (233,667 reads; 37.7% of Jädraås fungal reads) were putatively Archaeorhizomycetes (Table S4, Supplementary datafile 2).

Delimiting phylogenetic species hypotheses based on environmental sequences

We combined the two sequence datasets described above to delimitate species supported by both phylogenetic and ecological species recognition. In brief, we generated phylogenetic species hypothesis (PSHs) from a ML tree including both long ASVs and itASVs. Read counts per sample were combined for all itASVs merged into a PSH. Well supported sister PSHs, ie. pairs of PSHs supported by long read data, were tested for having different realized niches using their relative abundance in the total fungal community. The method is specified below.

Phylogenetic species recognition

We generated an alignment with the 123 itASVs from the “ecological” dataset and the 41 ASVs from the “phylogenetic” dataset including also reference rDNA sequences of *A. borealis* (KF993708), *A. finlayi* (JF836022) and the uncultured lineage GS31 (KY687760) as well as several 5.8 and LSU rDNA sequences representing outgroup taxa (*Saitoella complicata* (AY548296), *Schizosaccharomyces pombe* (EU916982), *Taphrina deformans* (DQ470973) and *Taphrina wiesneri* (NG_027620)) using Geneious as above. To avoid duplication of data, we identified and removed 25 itASVs that were identical to the ITS2 regions of a long read ASV in the alignment (Supplementary datafile 3). A final alignment with 146 sequences was uploaded in the CIPRES portal (Miller et al., 2010)and a maximum likelihood tree was built using the online version of RAxML-HPC2 on XSED version 8 (Stamatakis, 2014) using the GTRGAMMA model and 1000 iterations for the calculation of bootstrap support. PSHs were delimited using both Bayesian and Maximum likelihood implementations of the Poisson tree processes (PTP) model, using the online web server (Zhang et al., 2013) (Supplementary datafiles 4). FigTree v1.4.4 was used to visualize the resulting tree, and the Maximum likelihood solution PSHs were collapsed to their stem node for display. All alignments and trees are made available in TreeBASE, Study ID S26320.

Analysis of relative sequence read abundance and distribution of Archaeorhizomycetes

We used a 2-way ANOVA (function aov in R 3.4.1) (R Core Team, 2019) to test for general trends in total fungal reads, number of Archaeorhizomycetes itASVs, number of Archaeorhizomycetes PSH, and relative abundance of Archaeorhizomycetes across soil horizons and treatments administered in previous studies (Table S5). Terms were added sequentially and post-hoc Tukey test with HSD p-value correction for multiple comparisons were performed when significant effects were detected. Read counts for itASVs were combined into PSHs before calculating relative abundance of PSHs in the total sequenced fungal community for each sample (Supplementary datafile 5). The overall Archaeorhizomycetes community composition in relation to soil horizon was visualized using nMDS ordination analysis (metaMDS in package vegan) based on the relative abundance of the 68 delimited Archaeorhizomycetes PSHs using the default Bray-Curtis distance. Two dimensions were assigned to the analysis with a maximum of 100 starts. Further, the significance of soil horizon, plot and treatments in shaping the Archaeorhizomycetes community composition was further tested using permutational multivariate analysis of variance (PERMANOVA) implemented in the anova.cca function in the vegan package of R (Oksanen et al., 2013) (Table S6). This analysis was based on relative abundance of the nine PSHs represented by long read, 999 permutations were performed. We also tested for soil horizon specificity of the nine PSH using a Kruskal Wallis test in R (kruskal.test in R version 3.4.1) (R Core Team, 2019). The multiple comparisons *p*-values were corrected with Benjamini & Hochberg-method (Table S7).

IonTorrent sequencing generated 4.5 times more reads than PacBio sequencing and we thus expected the “ecological” dataset to represent a more exhaustive sampling compared to the phylogenetic dataset. Based on the “ecological” dataset, we estimated the proportion of Archaeorhizomycetes species richness and abundance was represented by the “phylogenetic” dataset by calculating both reads in and counts of PSHs including also unassigned ASVs as proportion of all PSHs (Fig. 1, Table S4). We used “ecological”

dataset read abundance of PSHs to estimate the detection limit of the “phylogenetic” dataset by determining the rarest PSH including long read ASVs. However, detection limits are not absolute since not all soil samples were included in the “phylogenetic” dataset (Table S1).

Ecological species recognition

Based on the phylogenetic analysis of PSHs detected at the site, we selected two pairs of sister PSHs (PSH_2 vs. PSH_3 and PSH_7 vs. PSH_8) that were supported by long reads from the “phylogenetic” dataset and frequently observed in the “ecological” dataset (Fig. S4). We tested these pairs for niche specific distribution using relative abundance in the “ecological” dataset with soil horizons as operationally identified niches. Orthogonal contrasts for each pair were given as explanatory variables against soil horizon relative abundances in the statistical model, which was run with 200 permutations (Table S8).

Placement of local PSHs among available Archaeorhizomycetes sequences globally

To place the Archaeorhizomycetes PSHs from the current study in a larger phylogenetic context, an alignment was generated including publicly available environmental sequences previously identified as belonging to the Archaeorhizomycetes (Menkis et al., 2014), as well as new sequences affiliating with the class identified by BLAST search in UNITE (Altschul et al., 1990). Environmental sequences were included if they covered at least two of the three rDNA regions ITS1, ITS2 or LSU. Duplicate sequences from individual studies were excluded. Sequences were aligned including the outgroup previously described and visually inspected in Geneious version 11.1.4 (Kearse et al., 2012) to remove suspected chimeras. The alignment was truncated after the LSU D3 region where some sequences had long inserts (Holst-Jensen et al., 1999) and insertions present in only two or less sequences were deleted. Maximum likelihood trees were built using the online version of RAxML XSED2 (version 8) (Stamatakis, 2014) in CIPRES portal (Miller et al., 2010), with the GTRGAMMA model and 1000 iterations for the calculation of bootstrap support. To focus the analysis on PSHs from the current study a series of alignments and trees were generated by stepwise removing published sequences that separated on deep nodes without including any of the PSHs identified in the current study. The final alignment included a total of 172 Archaeorhizomycetes sequences in addition to the 41 ASVs generated in the current study and 6 outgroup sequences (alignment and tree are available in TreeBASE Study S26320). PSHs were delimitated across the tree using the bPTP portal as above and referred to as global PSHs and visualized in TreeView by collapsing nodes corresponding to PSHs according to the maximum likelihood solution. All included Archaeorhizomycetes sequences were mapped to UNITE species hypotheses at 98.5% by massBLAST of their ITS region (Supplementary datafile 6). The generated tree allowed us to evaluate the robustness of phylogenetic species delimitation in our local dataset and to visualize global sister clade relationships. Further, the larger Archaeorhizomycetes alignment was used to visually identify diagnostic sequences regions in both the ITS1 and ITS2 region, for two novel species first hypothesized as PSH_7 and PSH_8 (Fig. 1).

Results

Across all samples in the studied pine forest, 38% of the sequenced fungal community was assigned to class Archaeorhizomycetes in the short amplicon “ecological” dataset (Table S4). In the “phylogenetic” dataset on the other hand, the class represents 26% of the fungal reads (Table S2). Despite high relative abundance based on total fungal reads and intense cultivation efforts no isolates of Archaeorhizomycetes were successfully obtained. The relative abundance of class Archaeorhizomycetes was not significantly affected by soil horizon (Fig. S5, Table S5) or treatment (Table S5). However, number of Archaeorhizomycetes itASVs and PSHs were significantly affected by soil horizon (Table S6), with higher richness detected in B horizon samples compared to samples from both O and E horizon (Fig. S6). Across all samples, 68 PSHs of Archaeorhizomycetes were delimited, nine of which were supported by reads in the “phylogenetic” dataset, one of which included the reference sequence of *A. finlayi* (Fig. 1, Fig. S7). The most abundant PSHs were supported by long read data in the “phylogenetic” dataset (Fig. S4), together accounting for 78% of the Archaeorhizomycetes reads identified in the “ecological” dataset. Likely due to the lower sequencing depth of the “phylogenetic” dataset, taxa making up less than 1% of the sequenced fungal community were not consistently recovered in long read ASVs (Fig. S4).

The Archaeorhizomycetes community composition was structured by soil horizons (Fig. S8) and model testing showed that the relative abundance in the “ecological” dataset of nine PSHs supported by long reads was significantly affected by soil horizon (PERMANOVA: PSH vs Horizon df = 2, F = 4.6070, P < 0.0001) but not by treatment or plot (Table S6). The niche distribution also varied with plot*treatment interactions (df = 3, F = 2,0344, P > 0.005) (Table S6). This indicates that soil horizons may reflect niches explored differently by these fungi. While most of the nine PSHs were not significantly associated with a single horizon, we found that *A. finlayi* was tentatively associated with the O horizon (Table S7). For two sister pairs of phylogenetically well supported PSHs that were both abundant and frequently observed (Fig. S4, Table S8), we found significant differences in realized niche for PSHs within each pair (Fig. 1, Table S9). Based on their relative sequence read abundance in the fungal community across soil horizons, PSH_7 and PSH_8 had significantly different niche distributions ($p < 0.05$). Differential niche distribution of PSH_1 and PSH_2 was marginally significant ($p < 0.1$; Table S9). This provides further evidence that the phylogenetically distinct PSH_7 and PSH_8 are also ecologically separate species.

Global perspective on recognized taxa

The site-specific Archaeorhizomycetes diversity (Fig. 1) was analyzed in a global perspective by populating an alignment of ASVs from our “phylogenetic” dataset with publicly available environmental sequences that formed well supported clades with long reads Archaeorhizomycetes ASVs from the current study (Fig. 2, Supplementary datafile 6). A total of 78 global PSHs were delimited and supported using a bPTP model, 41 of which are represented by a single sequence (Fig. 2). However, some of the PSHs delimited in the local dataset (Fig. 1) were not stable with the addition of published environmental sequences. Most notably, PSH_2, PSH_5 and PSH_9 (identified as *A. finlayi*) each split in the global [illegible] into two global PSHs, both containing sequences previously

recovered from the studied field site, with PSH_2:2 being the more frequently observed of the two and including long sequences in UNITE SH1566367.08FU, while sequences in PSH 2:1 did not map to an existing UNITE SH (Supplementary datafile 6). Similarly, global PSH_5:1 contained previously published sequences from the same field site, as well as from Ireland and the US, while global PSH_5:2 contained only two ASVs from the current study (Supplementary datafile 6). Both global PSH_5:1 and PSH_5:2 cluster on a well-supported branch with four global PSHs that include previously published sequences (Fig. 2). Sequences in three of these PSHs map to the same UNITE SH, demonstrating that the boundaries of phylogenetically delimited PSHs and cluster-based SHs are not always the same. Further, ASVs and published sequences mapping to SH1556760.08FU, *Archaeorhizomyces finlayi*, split into two global PSHs (PSH_9:1 and PSH_9:2) in the current analysis (Fig. 2, Supplementary datafile 6). These two PSHs cluster on a well-supported branch with five single sequence PSHs mapping to four different UNITE SHs. Limited taxon sampling and low representation of intra species genetic variation of potential sister taxa as well as possible chimeric sequences within the clade may obscure PSH delimitation.

The included sequences represent only a fraction of the global Archaeorhizomycetes diversity, including sequences from 90 of the 181 UNITE SH currently identified as belonging to Archaeorhizomycetes (Kõljalg et al., 2013). With the exception of PSH_4 and PSH_2:1, long read ASVs mapped to UNITE SH meaning that the majority of our ASVs where highly similar to previously observed environmental sequences (Supplementary datafile 6). *Archaeorhizomycetes borealis* is distributed across the Eurasian boreal biome (Menkis et al., 2014) but was rare at our study site and was only detected as a short amplicon itASV.

In the global analysis, the sister taxa PSH_7 and PSH_8 cluster together with three single sequences (Fig. 2), two of which map to SH1566388.08FU (Supplementary datafile 6). Based on the globalfungi.com database (Vetrovsky et al., 2020) this SH restricted to North America, indicating that closely related taxa exist globally but that these are likely geographically separated from those captured in our dataset. While addition of publicly available environmental sequences weakened the bPTP ML support for these two PSHs, both remain intact. In the global tree, the delimitation of global PSH_7 and global PSH_8 both include sequences previously recovered from the studied field station as well as sequences collected throughout Europe (Supplementary datafile 7). These two PSHs are distinct based on both local (Fig. 1) and global (Fig. 2) phylogenetic analysis as well as ecological evidence demonstrating that they have significantly different realized niches across soil horizons at the study site (Fig. S9, Table S9). Overall, relative abundance of PSH_8 is higher at 8.9 ± 1,7% across all three horizons compared to PSH_7 at 3,5 ± 1,2%. Relative abundance of PSH_8 is highest in O horizon and decreases towards deeper soil layers while the relative abundance of PSH_7 is stable throughout the soil profile (Fig. S9a). When co-existing in a sample, relative abundance of PSH_7 is higher when the relative abundance of PSH_8 is low (Fig. S9b) indicating competitive avoidance between the two species similar to patterns of vertical separation previously observed for soil fungal sister species (Mujica et al., 2016).

Based on combined phylogenetic and ecological evidence we propose two novel species 1) PSH_7 and 2) *Archaeorhizomyces victor nom. sEq.* for

PSH_8 with names appended with *nom. sEq.* to indicate that the names are based on a sequence in the absence of acceptable type material.

Taxonomy

Archaeorhizomyces secundus *nom. sEq.* Kluting, M. Ryberg & Rosling *sp. nov*.

MykoBank MB826774

Diagnosis: Separated from other species in the genus by ribosomal sequences possessing the following distinctive characters in ITS1: CCGAGTCGCCACAT, at position homologous to bases 103–116 in ASV_4; and in ITS2: CCATACCTTTTTGGTGTGT, at position homologous to bases 317–335 in ASV_4.

Ecological notes

In DNA extracts from soil and from roots, often from ectomycorrhizal roots of *Pinus sylvestris* but also from roots of *Calluna vulgaris*. Found mostly in pine forest but also in other coniferous forest including spruce and heatland, in temperate, boreal and alpine climate. Found in both organic and mineral soil horizons but is less abundant in upper soil layers compared to *Archaeorhizomyces victor*.

Etymology

The species is outnumbered by its sister species for colonization of organic soil.

Type sequence

ASV_4 based on sequence reads from Ivantjärnsheden field station, Jädraås, SWEDEN, Uppsala, October-2013, **UDB0779127** in UNITE

Distribution

Sweden, Norway, Finland, United Kingdom, Austria.

Additional sequences

AB560514, DQ309209, FM992980, HM069470, JN032483, KX289979 (Alignment in Supplementary datafile 8).

Archaeorhizomyces victor *nom. sEq.* Kluting, M. Ryberg & Rosling sp. nov.

MykoBank MB827624

Diagnosis: Distinct from other species in the genus by ribosomal sequences possessing the following distinctive characters in ITS1: CGAATGGCTTTT at position homologous to bases 48–59 in ASV3, and

ATGTGCTTTGGCGCCAAGT at position homologous to bases 93–111 in ASV_3; and in ITS2: TCATACCTTCTT at position homologous to 323–333 in ASV_3.

Ecological notes

In DNA extracts from soil and from roots, often from ectomycorrhizal roots of *Pinus sylvestris* but also from *Calluna vulgaris*. Most frequently found in coniferous forests but also in deciduous forest, in temperate, boreal and alpine climate. Found in both organic and mineral soil horizons, outnumbers Archaeorhizomyces secundus in upper soil layers.

Etymology

The species wins in competition with sister species for colonization of organic soil.

Type sequence

ASV_3 based on sequence reads from Ivantjärnsheden field station, Jädraås, SWEDEN, Uppsala, October-2013, **UDB0779126** in UNITE.

Distribution

Austria, Finland, Germany, the Netherlands, Sweden, United Kingdom.

Additional sequences

AB560521, DQ309123, HM069408, HQ873359, JF300381, JN006467, JN006468, JN032485 (Alignment in Supplementary datafile 8).

Discussion

The class Archaeorhizomycetes is highly diverse, ubiquitous, and often highly abundant in environmental DNA samples from around the world (Porter et al., 2008). With only two formally named and thus recognized species very little is known about the ecology, morphology, and life styles of the species within Archaeorhizomycetes (Menkis et al., 2014; Rosling et al., 2011). Species within the class are likely to have different ecological roles and occupy different niches (Rosling et al., 2013) but intra-taxonomic boundaries have only begun to be characterized within this lineage. The goal of this study was to investigate the diversity of *Archaeorhizomyces* spp. at Ivantjärnsheden field station, a well-studied site in which occurrences of OTUs representing species of *Archaeorhizomyces* have been previously documented (Fransson & Rosling, 2014; Lindahl et al., 2007). We found that it remains difficult to obtain pure cultures of *Archaeorhizomyces* even at a site with high abundance based on amplicon sequencing. This prevents us from collecting traditional sources of evidence, such as morphological characterization, for species descriptions and have instead delineated two novel species of *Archaeorhizomyces* using alternative lines of evidence based on environmental sequences. By combining phylogenetic and species are different from each other and previously

described species and we choose to name them *A. secundus nom. sEq.* and *A. victor nom. sEq.* As there is no material isolating the single species, we choose to not designate type material, and recognize that these names are invalid according to the *Code*. We do however assign a sequence describing the fungal barcoding region that together with type locality serve the purpose of type. We have appended them with *nom. sEq.* to represent their status, in line with the recommendations of Lücking and co-authors (Lucking et al., 2018). It is important to remember the function of a name: as a tool to label and communicate a defined object or concept. Naming “dark taxa” is a tool to communicate their presence and gather an understanding of their biology. We hope that by applying and using these names, regardless of validity, we might be able to accumulate knowledge about these species faster, and potentially expediate discovery of living specimens that can be linked to them. They also serve as suggested names, for if and when they can be validated by specimen or image.

Names are essential for communication, and species without a valid name are ignored in many contexts. However, the vast majority of fungal species cannot be named in accordance with the *Code* because we know them only from environmental sequences. While molecular data have proven useful in taxonomy for instance for the delimitation and characterization of taxa, the role that molecular data should play in nomenclature has been much debated as of recent, particularly the role of environmental DNA sequences (Hibbett, 2016; Lücking & Hawksworth, 2018; Lucking et al., 2018; Sarma, 2018; Seifert, 2017; Thines et al., 2018; Zamora et al., 2018).

Some researchers site concerns with the use of cluster-based analysis to detect taxa in complex samples (Thines et al., 2018; Zamora et al., 2018) although proponents of sequence-based nomenclature have argued that this is largely a misconception (Lücking & Hawksworth, 2018; Lucking et al., 2018). Indeed, the application of clustering methods to environmental metagenomic sequence data is useful tools for community studies (Clemmensen et al., 2013; Tedersoo et al., 2014), but insufficient for taxonomy (Ryberg & Nilsson, 2018). Sequence variation within OTUs is both biological and artifactual, and although the boundaries of OTUs are generally decent proxies of species, they are entirely study and method dependent. Sequence similarity threshold levels are often applied uniformly (for example, 98.5% similarity), although the degree of inter and intraspecific variation is known to differ between groups of fungi. An OTU may thus contain sequences from more than one species, or only part of a single species. Here, we circumvent this problem by applying a method to distinguish between biological variation and sequencing error, and compare only amplicon sequence variants ASVs. Including sequence quality data alongside the raw reads and ASV sequences will facilitate cross-comparison between studies in the future. Further, with ASVs we are able to get better insight into the intra- vs inter- specific sequence variation, which helps us determine phylogenetic species hypotheses. OTU sequences in reference databases are of varying quality, complicating the use of such data for species delineation. This combination, a blurring of the intraspecific variation, and potential quality issues of publicly available sequences may be a contributing factor to why phylogenetic species hypothesis support dropped drastically in our global analysis (Fig. 2) compared to the local analysis (Fig. 1). There are also shortcomings in using a single locus, for instance the ITS, as a universal locus for species delimitation. It

TS variation could actually represent multiple species, as well

as examples of high within species variability. However, this limitation applies to any single characteristic be it molecular or morphological.

Another concern with using environmental sequences in taxonomy is the potential for describing artifacts as species, or otherwise applying a name to an organism that is already described (Thines et al., 2018; Zamora et al., 2018). Sequencing errors (such as large indels or chimeric sequences) could result in a number of “ghost taxa”, and species being repeatedly described, (for example, if a species that is described but not sequenced gets erroneously described a second time based on environmental data) are real possibilities. We are careful here to mind the fine line between this detrimental effect, and the benefit of using sequence-based nomenclature as a vessel for taxonomic discovery. In the case of Archaeorhizomyces, we are certain that sequences from all described species was included in our analysis. Further, we used two sequencing technologies, PacBio and Ion Torrent, and used a method to detect and correct sequencing errors.

Considering that the vast majority of fungal species are currently unidentified and visible only through the lens of environmental sequences, we urgently need to explore new routes for discovering and documenting the world’s fungal biodiversity. There is compelling arguments in favor of sequence-based nomenclature for dark taxa of fungi arguing that names are essential for communication of these organisms, and that alternative proposed methods using alphanumerically coded names is simply not as accessible as latin binomial names (Hibbett, 2016; Ryberg & Nilsson, 2018). We agree that *Archaeorhizomyces secundus* is much less abstract than for example ‘SH1566369.08FU’ for specialists and non-specialists alike. A name also conveys biologically informative information. By simply seeing the genus name of this species, one immediately can infer something about its phylogenetic placement in the fungal tree of life. Formal proposals have been put forth to amend the *Code* so as to allow the use of DNA sequences as type material for formally describing fungal taxa, largely to facilitate taxonomic discovery (Hawksworth, 2015), and a few examples of how this could be done exist (De Beer et al., 2016; Lücking & Moncada, 2017). Not surprisingly, however, there are also many within the mycological community who oppose the idea of DNA sequences as acceptable type material (Thines et al., 2018; Zamora et al., 2018). It is clear that both perspectives have potential benefits and potential consequences (Ryberg & Nilsson, 2018) but it seems that sequence based nomenclature in some capacity is the only way forward for discovering a documenting fungal biodiversity, especially for lineages like Archaeorhizomycetes.

We feel confident of the sequence data upon which we build phylogenetic species hypotheses and by combining different lines of evidence we find that two of the most abundant and frequently observed species at the site can be described and named. *Archaeorhizomycetes secundus* and *A. victor* represent two species with distinct realized niches at this site. In the future they may need to be delimited against additional - hitherto undescribed - species, or amended to include further variation in light of the new data. However, by naming them we facilitate communication about their ecology and distribution and hope to thereby stimulate further research into taxonomic and functional diversity within the

Declarations

Ethics approval and consent to participate

Not applicable

Consent of publication

Not applicable

Availability of Data and Materials

Sequences generated to screen new cultures are available on GenBank, with accession numbers MH843963–MH844060, (PopSet: 1472857015). Raw reads for both the “phylogenetic” and “ecological” datasets are available in ENA under the accessions ERS3508481- ERS3508483 and ERS4600640-ERS4600675. The complete “phylogenetic” dataset with 272 fungal ASVs is available in UNITE through the GBIF repository, http://doi.org/10.15468/8zymuf. Sequence alignments and phylogenetic trees are available in the TreeBASE repository, http://purl.org/phylo/treebase/phylows/study/TB2:S22994. Additional datasets supporting the conclusions of this article are included within the article and as supplementary datafiles available in our published OSF repository at: https://osf.io/96dkz/?view_only=2a166dd33fa04e95a05208f800fefb90.

Competing interests

The authors declare that they do not have any competing interests

Funding

Funding, VR (2012-03950), Carl Trygger foundation and Lars Hiertas Memorial Foundation. The project is conducted as part of the Biodiversity Atlas Sweden which is made possible by its partners and by grants from the Swedish Research Council. Sequencing was performed by the SNP&SEQ Technology Platform at NGI Sweden and SciLife Laboratory, Uppsala, supported by the VR and the KAW. Computations were performed on resources provided by SNIC through UPPMAX.

Authors’ contributions

FKK performed first round of bioinformatic analysis and KK performed the final round of bioinformatic analysis together with JT who also handled all data publication. HU designed and performed field sampling, molecular work and all culturing efforts. TA performed statistical analysis, SES generated the phylogenetic dataset and MRyd assisted with bioinformatic analysis. MRyb and AR developed the conceptual framework together with FK and KK and did the phylogenetic analysis. MRyb, AR and KK wrote the manuscript with input from all authors who read and approved the final manuscript.

Acknowledgements

We thank C. Karlsson for assistance at the field site, T. Tajrin for isolation efforts, M. Bahram for statistics discussion. We are greatly thankful to Tomáš Větrovský, Petr Baldrian, K. Paey, S. Glassman and K. Clemmensen for sharing sequence data that helped us develop our analysis. We also thank for data and S. Ekman, S. Redhead and S. Pennycook for nomenclature guidance.

References

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of molecular biology, 215*(3), 403-410.
2. Axelsson, B., & Bråkenhielm, S. (1980). Investigation sites of the Swedish Coniferous Forest Project: Biological and physiographical features. In T. Persson (Ed.), *Structure and function of northern coniferous forests; an ecosystem study* (Vol. 32, pp. 25-64). Berlings.
3. Bengtsson-Palme, J., Ryberg, M., Hartmann, M., Branco, S., Wang, Z., Godhe, A., De Wit, P., Sánchez-García, M., Ebersberger, I., & de Sousa, F. (2013). Improved software detection and extraction of ITS1 and ITS 2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in ecology and evolution, 4*(10), 914-919.
4. Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: high-resolution sample inference from Illumina amplicon data. *Nature methods, 13*(7), 581.
5. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC bioinformatics, 10*(1), 421.
6. Carrino-Kyker, S. R., Kluber, L. A., Petersen, S. M., Coyle, K. P., Hewins, C. R., DeForest, J. L., Smemo, K. A., & Burke, D. J. (2016). Mycorrhizal fungal communities respond to experimental elevation of soil pH and P availability in temperate hardwood forests. *Fems Microbiology Ecology, 92*(3), fiw024.
7. Clemmensen, K., Bahr, A., Ovaskainen, O., Dahlberg, A., Ekblad, A., Wallander, H., Stenlid, J., Finlay, R., Wardle, D., & Lindahl, B. (2013). Roots and associated fungi drive long-term carbon sequestration in boreal forest. *Science, 339*(6127), 1615-1618.
8. Clemmensen, K. E., Finlay, R. D., Dahlberg, A., Stenlid, J., Wardle, D. A., & Lindahl, B. D. (2015). Carbon sequestration is related to mycorrhizal fungal community shifts during long-term succession in boreal forests. *New Phytologist, 205*(4), 1525-1536.
9. Cruz-Paredes, C., Svenningsen, N. B., Nybroe, O., Kjøller, R., Frøslev, T. G., & Jakobsen, I. (2019). Suppression of arbuscular mycorrhizal fungal activity in a diverse collection of non-cultivated soils. *Fems Microbiology Ecology, 95*(3), fiz020.
10. Dayrat, B. (2005). Towards integrative taxonomy. *Biological journal of the Linnean society, 85*(3), 407-417.
11. De Beer, Z. W., Marincowitz, S., Duong, T. A., Kim, J.-J., Rodrigues, A., & Wingfield, M. J. (2016). Hawksworthiomyces gen. nov.(Ophiostomatales), illustrates the urgency for a decision on how to name novel taxa known only from environmental nucleic acid sequences (ENAS). *Fungal biology, 120*(11), 1323-1340.

12. De Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. *Endless forms: species and speciation*.
13. De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic biology, 56*(6), 879-886.
14. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics, 26*(19), 2460-2461.
15. Fernández-Martínez, M. A., Pérez-Ortega, S., Pointing, S. B., Green, T. A., Pintado, A., Rozzi, R., Sancho, L. G., & de los Ríos, A. (2017). Microbial succession dynamics along glacier forefield chronosequences in Tierra del Fuego (Chile). *Polar Biology, 40*(10), 1939-1957.
16. Fišer, C., Robinson, C. T., & Malard, F. (2018). Cryptic species as a window into the paradigm shift of the species concept. *Molecular Ecology, 27*(3), 613-635.
17. Fransson, P., & Rosling, A. (2014, DEC 2014). Fungal and bacterial community responses to Suillus variegtus extraradical mycelia and soil profile in Scots pine microcosms. *Plant and Soil, 385*(1-2), 255-272. https://doi.org/10.1007/s11104-014-2231-5
18. Gardes, M., & Bruns, T. D. (1993). ITS primers with enhanced specificity for basidiomycetes-application to the identification of mycorrhizae and rusts. *Molecular Ecology, 2*(2), 113-118.
19. Geml, J., Gravendeel, B., van der Gaag, K. J., Neilen, M., Lammers, Y., Raes, N., Semenova, T. A., de Knijff, P., & Noordeloos, M. E. (2014). The contribution of DNA metabarcoding to fungal conservation: diversity assessment, habitat partitioning and mapping red-listed fungi in protected coastal Salix repens communities in the Netherlands. *Plos One, 9*(6).
20. Grelet, G. A., Johnson, D., Vrålstad, T., Alexander, I. J., & Anderson, I. C. (2010). New insights into the mycorrhizal Rhizoscyphus ericae aggregate: spatial structure and co-colonization of ectomycorrhizal and ericoid roots. *New Phytologist, 188*(1), 210-222.
21. Hawksworth, D. L. (2015). Proposals to clarify and enhance the naming of fungi under the International Code of Nomenclature for algae, fungi, and plants. *Ima Fungus, 6*(1), 199-205.
22. He, W., Detheridge, A., Liu, Y., Wang, L., Wei, H., Griffith, G. W., Scullion, J., & Wei, Y. (2019). Variation in soil fungal composition associated with the invasion of *Stellera chamaejasme* in Qinghai–Tibet plateau grassland. *Microorganisms, 7*(12), 587.
23. Hibbett, D. (2016). The invisible dimension of fungal diversity. *Science, 351*(6278), 1150-1151.
24. Holst-Jensen, A., Vaage, M., Schumacher, T., & Johansen, S. (1999). Structural characteristics and possible horizontal transfer of group I introns between closely related plant pathogenic fungi. *Molecular biology and evolution, 16*(1), 114-126. https://doi.org/10.1093/oxfordjournals.molbev.a026031
25. Hopple Jr, J. S., & Vilgalys, R. (1994). Phylogenetic relationships among coprinoid taxa and allies based on data from restriction site mapping of nuclear rDNA. *Mycologia, 86*(1), 96-107.
26. Ihrmark, K., Bödeker, I., Cruz-Martinez, K., Friberg, H., Kubartova, A., Schenck, J., Strid, Y., Stenlid, J., Brandström-Durling, M., & Clemmensen, K. E. (2012). New primers to amplify the fungal ITS2 region–

evaluation by 454-sequencing of artificial and natural communities. *Fems Microbiology Ecology, 82*(3), 666-677.

27. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., & Duran, C. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics, 28*(12), 1647-1649.
28. Keeling, P. J., & Burki, F. (2019). Progress towards the Tree of Eukaryotes. *Current Biology, 29*(16), R808-R817.
29. Kluting, K., Clemmensen, K., Jonaitis, S., Vasaitis, R., Holmström, S., Finlay, R., & Rosling, A. (2019). Distribution patterns of fungal taxa and inferred functional traits reflect the non-uniform vertical stratification of soil microhabitats in a coastal pine forest. *Fems Microbiology Ecology, 95*(11), fiz149.
30. Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F., Bahram, M., Bates, S. T., Bruns, T. D., Bengtsson-Palme, J., & Callaghan, T. M. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology, 22*(21), 5271-5277.
31. Kyaschenko, J., Clemmensen, K. E., Hagenbo, A., Karltun, E., & Lindahl, B. D. (2017). Shift in fungal communities and associated enzyme activities along an age gradient of managed Pinus sylvestris stands. *The ISME journal, 11*(4), 863-874.
32. Leliaert, F., Verbruggen, H., Vanormelingen, P., Steen, F., López-Bautista, J. M., Zuccarello, G. C., & De Clerck, O. (2014). DNA-based species delimitation in algae. *European journal of phycology, 49*(2), 179-196.
33. Levy-Booth, D. J., Giesbrecht, I. J., Kellogg, C. T., Heger, T. J., D'Amore, D. V., Keeling, P. J., Hallam, S. J., & Mohn, W. W. (2019). Seasonal and ecohydrological regulation of active microbial populations involved in DOC, CO 2, and CH 4 fluxes in temperate rainforest soil. *The ISME journal, 13*(4), 950-963.
34. Lindahl, B. D., Ihrmark, K., Boberg, J., Trumbore, S. E., Högberg, P., Stenlid, J., & Finlay, R. D. (2007). Spatial separation of litter decomposition and mycorrhizal nitrogen uptake in a boreal forest. *New Phytologist, 173*(3), 611-620.
35. Lücking, R., & Hawksworth, D. L. (2018). Formal description of sequence-based voucherless Fungi: promises and pitfalls, and how to resolve them. *Ima Fungus, 9*(1), 143.
36. Lucking, R., Kirk, P. M., & Hawksworth, D. L. (2018). Sequence-based nomenclature: a reply to Thines et al. and Zamora et al. and provisions for an amended proposal "from the floor" to allow DNA sequences as types of names. *Ima Fungus, 9*(1), 185.
37. Lücking, R., & Moncada, B. (2017). Dismantling Marchandiomphalina into Agonimia (Verrucariaceae) and Lawreymycesgen. nov.(Corticiaceae): setting a precedent to the formal recognition of thousands of voucherless fungi based on type sequences. *Fungal Diversity, 84*(1), 119-138.
38. Maghnia, F. Z., Abbas, Y., Mahé, F., Kerdouh, B., Tournier, E., Ouadji, M., Tisseyre, P., Prin, Y., El Ghachtouli, N., & Yakhlef, S. E. B. (2017). Habitat-and soil-related drivers of the root-associated fungal community of Quercus suber in the Northern Moroccan forest. *Plos One, 12*(11).

39. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *journal, 17*(1), 10-12.
40. Marx, D. H. (1969). Influence of ectotrophic mycorrhizal fungi on resistance or pine roots to pathogenic infections. I Antagonisum of mycorrhizal fungi to root pathogenic fungi and soil bacteria. *Phytopathology, 59*, 153-163.
41. Menkis, A., Urbina, H., James, T. Y., & Rosling, A. (2014, 2014-Dec). Archaeorhizomyces borealis sp. nov. and a sequence-based classification of related soil fungal species. *Fungal biology, 118*(12), 943-955. https://doi.org/10.1016/j.funbio.2014.08.005
42. Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 gateway computing environments workshop (GCE),
43. Mujica, M. I., Saez, N., Cisternas, M., Manzano, M., Armesto, J. J., & Perez, F. (2016, Jul). Relationship between soil nutrients and mycorrhizal associations of two Bipinnula species (Orchidaceae) from central Chile. *Annals of Botany, 118*(1), 149-158. https://doi.org/10.1093/aob/mcw082
44. Nilsson, R. H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P., & Tedersoo, L. (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nature Reviews Microbiology, 17*(2), 95-109.
45. Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., Bengtsson-Palme, J., Jeppesen, T. S., Schigel, D., Kennedy, P., Picard, K., Glöckner, F. O., & Tedersoo, L. (2019). The UNITE database for molecular identification of fungi: handling dark taxa and parallel taxonomic classifications. *Nucleic acids research, 47*(D1), D259-D264.
46. Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R., Simpson, G. L., Solymos, P., Stevens, M. H. H., & Wagner, H. (2013). Package 'vegan'. *Community ecology package, version, 2*(9), 1-295.
47. Öpik, M., Vanatoa, A., Vanatoa, E., Moora, M., Davison, J., Kalwij, J., Reier, Ü., & Zobel, M. (2010). The online database MaarjAM reveals global and ecosystemic distribution patterns in arbuscular mycorrhizal fungi (Glomeromycota). *New Phytologist, 188*(1), 223-241.
48. Padial, J. M., Miralles, A., De la Riva, I., & Vences, M. (2010). The integrative future of taxonomy. *Frontiers in zoology, 7*(1), 16.
49. Peay, K. G., Kennedy, P. G., & Bruns, T. D. (2008). Fungal community ecology: a hybrid beast with a molecular master. *Bioscience, 58*(9), 799-810.
50. Persson, T. (1980). *Structure and function of northern coniferous forests; an ecosystem study* (Vol. 32). Berlings.
51. Pinto-Figueroa, E. A., Seddon, E., Yashiro, E., Buri, A., Niculita-Hirzel, H., Van Der Meer, J. R., & Guisan, A. (2019). Archaeorhizomycetes spatial distribution in soils along wide elevational and environmental gradients reveal co-abundance patterns with other fungal saprobes and potential weathering capacities. *Frontiers in Microbiology, 10*, 656.
52. Porter, T. M., Schadt, C. W., Rizvi, L., Martin, A. P., Schmidt, S. K., Scott-Denton, L., Vilgalys, R., & currence and phylogenetic placement of a soil clone group

adds a prominent new branch to the fungal tree of life. *Molecular Phylogenetics and Evolution, 46*(2), 635-644.

53. [Record #1697 is using a reference type undefined in this output style.]
54. Rosling, A., Cox, F., Cruz-Martinez, K., Ihrmark, K., Grelet, G.-A., Lindahl, B. D., Menkis, A., & James, T. Y. (2011, Aug 12). Archaeorhizomycetes: Unearthing an Ancient Class of Ubiquitous Soil Fungi. *Science, 333*(6044), 876-879. https://doi.org/10.1126/science.1206958
55. Rosling, A., Midgley, M. G., Cheeke, T., Urbina, H., Fransson, P., & Phillips, R. P. (2016, FEB 2016). Phosphorus cycling in deciduous forest soil differs between stands dominated by ecto- and arbuscular mycorrhizal trees. *New Phytologist, 209*(3), 1184-1195. https://doi.org/10.1111/nph.13720
56. Rosling, A., Timling, I., & Taylor, D. L. (2013). Archaeorhizomycetes: Patterns of distribution and abundance in soil. In *Genomics of Soil-and Plant-Associated Fungi* (pp. 333-349). Springer.
57. Ryberg, M. (2015). Molecular operational taxonomic units as approximations of species in the light of evolutionary models and empirical data from Fungi. *Molecular Ecology, 24*(23), 5770-5777.
58. Ryberg, M., & Nilsson, R. H. (2018). New light on names and naming of dark taxa. *MycoKeys*(30), 31.
59. Sarma, V. V. (2018). More than a handful of dirt: sequence-based species description and the role of the ICN (a response to Seifert). *Ima Fungus, 9*(1), 36-39.
60. Schadt, C. W., Martin, A. P., Lipson, D. A., & Schmidt, S. K. (2003). Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science, 301*(5638), 1359-1361.
61. Schadt, C. W., & Rosling, A. (2015, 2015-Jun-26). FUNGAL BIOGEOGRAPHY. Comment on "Global diversity and geography of soil fungi". *Science (New York, N.Y.), 348*(6242), 1438. https://doi.org/10.1126/science.aaa4269
62. Schmidt, P.-A., Bálint, M., Greshake, B., Bandow, C., Römbke, J., & Schmitt, I. (2013). Illumina metabarcoding of a soil fungal community. *Soil Biology and Biochemistry, 65*, 128-132.
63. Schoener, T. W. (1974). Resource partitioning in ecological communities. *Science, 185*(4145), 27-39.
64. [Record #1702 is using a reference type undefined in this output style.]
65. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics, 30*(9), 1312-1313.
66. Sun, H., Terhonen, E., Kovalchuk, A., Tuovila, H., Chen, H., Oghenekaro, A. O., Heinonsalo, J., Kohler, A., Kasanen, R., & Vasander, H. (2016). Dominant tree species and soil type affect the fungal community structure in a boreal peatland forest. *Environ. Microbiol., 82*(9), 2632-2643.
67. Tedersoo, L., Bahram, M., Puusepp, R., Nilsson, R. H., & James, T. Y. (2017). Novel soil-inhabiting clades fill gaps in the fungal tree of life. *Microbiome, 5*(1), 42.
68. Tedersoo, L., Bahram, M., Ryberg, M., Otsing, E., Koljalg, U., & Abarenkov, K. (2014, Aug). Global biogeography of the ectomycorrhizal/sebacina lineage (Fungi, Sebacinales) as revealed from comparative phylogenetic analyses. *Molecular Ecology, 23*(16), 4168-4183. https://doi.org/10.1111/mec.12849

69. Thines, M., Crous, P. W., Aime, M. C., Aoki, T., Cai, L., Hyde, K. D., Miller, A. N., Zhang, N., & Stadler, M. (2018). Ten reasons why a sequence-based nomenclature is not useful for fungi anytime soon. *Ima Fungus, 9*(1), 177-183.

70. UNITE Community. (2019). Unite usearch/utax release for eukaryotes. Version 18.11.2018. In

71. Urbina, H., Scofield, D. G., Cafaro, M., & Rosling, A. (2016, MAY 2016). DNA-metabarcoding uncovers the diversity of soil-inhabiting fungi in the tropical island of Puerto Rico. *Mycoscience, 57*(3), 217-227. https://doi.org/10.1016/j.myc.2016.02.001

72. Vetrovsky, T., Morais, D., Kohout, P., Lepinay, C., Gallardo, C. A., Holla, S. A., Bahnmann, B. D., Bilohneda, K., Brabcova, V., & D'Alo, F. (2020). GlobalFungi: Global database of fungal records from high-throughput-sequencing metabarcoding studies. *bioRxiv*.

73. White, T. J., Bruns, T., Lee, S., & Taylor, J. (1990). Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In *PCR protocols: a guide to methods and applications* (Innis, MA Gelfand, DH Sninsky, JJ White, Thomas J ed., Vol. 18, pp. 315-322). Academic Press.

74. Yeates, D. K., Seago, A., Nelson, L., Cameron, S. L., Joseph, L., & Trueman, J. W. (2011). Integrative taxonomy, or iterative taxonomy? *Systematic entomology, 36*(2), 209-217.

75. Zamora, J. C., Svensson, M., Kirschner, R., Olariaga, I., Ryman, S., Alberto Parra, L., Geml, J., Rosling, A., Adamcik, S., Ahti, T., Aime, M. C., Ainsworth, A. M., Albert, L., Alberto, E., Garcia, A. A., Ageev, D., Agerer, R., Aguirre-Hudson, B., Ammirati, J., Andersson, H., Angelini, C., Antonin, V., Aoki, T., Aptroot, A., Argaud, D., Sosa, B. I. A., Aronsen, A., Arup, U., Asgari, B., Assyov, B., Atienza, V., Bandini, D., Baptista-Ferreira, J. L., Baral, H.-O., Baroni, T., Barreto, R. W., Baker, H., Bell, A., Bellanger, J.-M., Bellu, F., Bemmann, M., Bendiksby, M., Bendiksen, E., Bendiksen, K., Benedek, L., Beresova-Guttova, A., Berger, F., Berndt, R., Bernicchia, A., Biketova, A. Y., Bizio, E., Bjork, C., Boekhout, T., Boertmann, D., Bohning, T., Boittin, F., Boluda, C. G., Boomsluiter, M. W., Borovicka, J., Brandrud, T. E., Braun, U., Brodo, I., Bulyonkova, T., Burdsall, H. H., Jr., Buyck, B., Burgaz, A. R., Calatayud, V., Callac, P., Campo, E., Candusso, M., Capoen, B., Carbo, J., Carbone, M., Castaneda-Ruiz, R. F., Castellano, M. A., Chen, J., Clerc, P., Consiglio, G., Corriol, G., Courtecuisse, R., Crespo, A., Cripps, C., Crous, P. W., da Silva, G. A., da Silva, M., Dam, M., Dam, N., Dammrich, F., Das, K., Davies, L., De Crop, E., De Kesel, A., De Lange, R., Bonzi, B. D. M., dela Cruz, T. E. E., Delgat, L., Demoulin, V., Desjardin, D. E., Diederich, P., Dima, B., Dios, M. M., Divakar, P. K., Douanla-Meli, C., Douglas, B., Drechsler-Santos, E. R., Dyer, P. S., Eberhardt, U., Ertz, D., Esteve-Raventos, F., Salazar, J. A. E., Evenson, V., Eyssartier, G., Farkas, E., Favre, A., Fedosova, A. G., Filippa, M., Finy, P., Flakus, A., Fos, S., Fournier, J., Fraiture, A., Franchi, P., Molano, A. E. F., Friebes, G., Frisch, A., Fryday, A., Furci, G., Marquez, R. G., Garbelotto, M., Garcia-Martin, J. M., Otalora, M. A. G., Sanchez, D. G., Gardiennet, A., Garnica, S., Benavent, I. G., Gates, G., Gerlach, A. d. C. L., Ghobad-Nejhad, M., Gibertoni, T. B., Grebenc, T., Greilhuber, I., Grishkan, B., Groenewald, J. Z., Grube, M., Gruhn, G., Gueidan, C., Gulden, G., Gusmao, L. F. P., Hafellner, J., Hairaud, M., Halama, M., Hallenberg, N., Halling, R. E., Hansen, K., Harder, C. B., Heilmann-Clausen, J., Helleman, S., Henriot, A., Hernandez-Restrepo, M., Herve, R., Hobart, C., Hoffmeister, M., Hoiland, K., Holec, J., Holien, H., Hughes, K., Hubka, V., Huhtinen, S., Ivancevic, B., Jagers, M., Jaklitsch, W., Jansen, A., Jayawardena, ...ston, P., Jorgensen, P. M., Karnefelt, I., Kalinina, L. B.,

Kantvilas, G., Karadelev, M., Kasuya, T., Kautmanova, I., Kerrigan, R. W., Kirchmair, M., Kiyashko, A., Knapp, D. G., Knudsen, H., Knudsen, K., Knutsson, T., Kolarik, M., Koljalg, U., Kosuthova, A., Koszka, A., Kotiranta, H., Kotkova, V., Koukol, O., Kout, J., Kovacs, G. M., Kriz, M., Kruys, A., Kudera, V., Kudzma, L., Kuhar, F., Kukwa, M., Kumar, T. K. A., Kunca, V., Kusan, I., Kuyper, T. W., Lado, C., Laessoe, T., Laine, P., Langer, E., Larsson, E., Larsson, K.-H., Laursen, G., Lechat, C., Lee, S., Lendemer, J. C., Levin, L., Lindemann, U., Lindstrom, H., Liu, X., Hernandez, R. C. L., Llop, E., Locsmandi, C., Lodge, D. J., Loizides, M., Lokos, L., Luangsa-ard, J., Luderitz, M., Lumbsch, T., Lutz, M., Mahoney, D., Malysheva, E., Malysheva, V., Manimohan, P., Mann-Felix, Y., Marques, G., Martinez-Gil, R., Marson, G., Mata, G., Matheny, P. B., Mathiassen, G. H., Matocec, N., Mayrhofer, H., Mehrabi, M., Melo, I., Mesic, A., Methven, A. S., Miettinen, O., Romero, A. M. M., Miller, A. N., Mitchell, J. K., Moberg, R., Moreau, P.-A., Moreno, G., Morozova, O., Morte, A., Muggia, L., Gonzalez, G. M., Myllys, L., Nagy, I., Nagy, L. G., Neves, M. A., Niemela, T., Nimis, P. L., Niveiro, N., Noordeloos, M. E., Nordin, A., Noumeur, S. R., Novozhilov, Y., Nuytinck, J., Ohenoja, E., Fiuza, P. O., Orange, A., Ordynets, A., Ortiz-Santana, B., Pacheco, L., Pal-Fam, F., Palacio, M., Palice, Z., Papp, V., Partel, K., Pawlowska, J., Paz, A., Peintner, U., Pennycook, S., Pereira, O. L., Daniels, P. P., Capella, M. A. P.-D.-G., del Amo, C. M. P., Gorjon, S. P., Perez-Ortega, S., Perez-Vargas, I., Perry, B. A., Petersen, J. H., Petersen, R. H., Pfister, D. H., Phukhamsakda, C., Piatek, M., Piepenbring, M., Pino-Bodas, R., Esquivel, J. P. P., Pirot, P., Popov, E. S., Popoff, O., Alvaro, M. P., Printzen, C., Psurtseva, N., Purahong, W., Quijada, L., Rambold, G., Ramirez, N. A., Raja, H., Raspe, O., Raymundo, T., Reblova, M., Rebriev, Y. A., Garcia, J. d. D. R., Ripoll, M. A. R., Richard, F., Richardson, M. J., Rico, V. J., Robledo, G. L., Barbosa, F. R., Rodriguez-Caycedo, C., Rodriguez-Flakus, P., Ronikier, A., Casas, L. R., Rusevska, K., Saar, G., Saar, I., Salcedo, I., Martinez, S. M. S., Montoya, C. A. S., Sanchez-Ramirez, S., Sandoval-Sierra, J. V., Santamaria, S., Monteiro, J. S., Schroers, H. J., Schulz, B., Schmidt-Stohn, G., Schumacher, T., Senn-Irlet, B., Sevcikova, H., Shchepin, O., Shirouzu, T., Shiryaev, A., Siepe, K., Sir, E. B., Sohrabi, M., Soop, K., Spirin, V., Spribille, T., Stadler, M., Stalpers, J., Stenroos, S., Suija, A., Sunhede, S., Svantesson, S., Svensson, S., Svetasheva, T. Y., Swierkosz, K., Tamm, H., Taskin, H., Taudiere, A., Tedebrand, J.-O., Lahoz, R. T., Temina, M., Thell, A., Thines, M., Thor, G., Thus, H., Tibell, L., Tibell, S., Timdal, E., Tkalcec, Z., Tonsberg, T., Trichies, G., Triebel, D., Tsurykau, A., Tulloss, R. E., Tuovinen, V., Sosa, M. U., Urcelay, C., Valade, F., Garza, R. V., van den Boom, P., Van Vooren, N., Vasco-Palacios, A. M., Vauras, J., Santos, J. M. V., Vellinga, E., Verbeken, A., Vetlesen, P., Vizzini, A., Voglmayr, H., Volobuev, S., von Brackel, W., Voronina, E., Walther, G., Watling, R., Weber, E., Wedin, M., Weholt, O., Westberg, M., Yurchenko, E., Zehnalek, P., Zhang, H., Zhurbenko, M. P., & Ekmani, S. (2018, JUN 2018). Considerations and consequences of allowing DNA sequence data as types of fungal taxa. *Ima Fungus, 9*(1), 167-175. https://doi.org/10.5598/imafungus.2018.09.01.10

76. Zhang, J., Kapli, P., Pavlidis, P., & Stamatakis, A. (2013). A general species delimitation method with applications to phylogenetic placements. *Bioinformatics, 29*(22), 2869-2876.

Figures

Figure 1

Maximal likelihood phylogenetic inference demonstrating the diversity of Archaeorhizomycetes in mid-Sweden Pine forest podzol soil based on environmental ribosomal long and short amplicon ASVs representing 68 phylogenetic species hypotheses (PSH) based on Maximum likelihood solutions in bPTP. Two of these include sequences of the described species A. borealis and A. finlayi. The tree includes the GS31 (Tedersoo, et al. 2017) and four Taphrinamycotina

species as outgroup. PSHs were cartooned to their stem node to visualize PSH represented by more than one ASV. Nine PSHs (numbers 1-9) and were represented by long read ASVs from the current study and are highlighted in large bold font. Remaining cartooned terminal nodes were represented only by short read itASVs from the current study and are labeled with PSH_number or Archaeorhizomycetes borealis. Bootstrap support values over 75 are shown on the branches (calculated from 1000 iterations). Average relative sequence read abundance in soil horizons O, E and B are inserted in colored boxes for two pairs of sister PSHs. Bent arrows indicate tested orthogonal contrasts with p-values shown in the middle. (See Fig. S4 for corresponding tree with all ASV labels displayed).

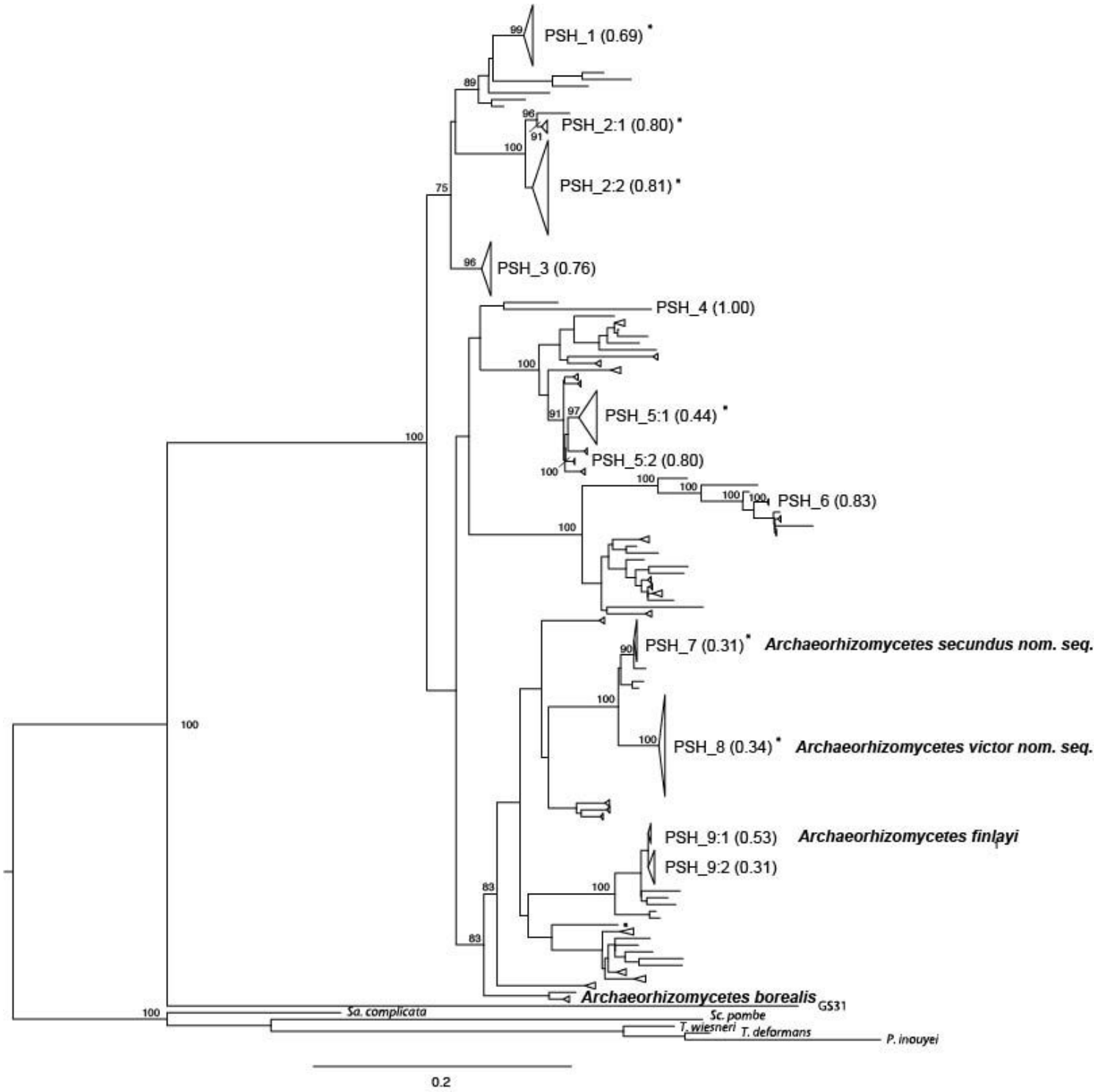

Figure 2

Maximum likelihood tree including all long read ASVs from the current study and publicly available environmental sequences of Archaeorhizomycetes covering at least two of the rDNA regions ITS1, ITS2 and LSU. The tree is limited to environmental sequences that cluster on well supported basal nodes with sequences from the current study. Nodes represent the 78 phylogenetic species hypothesis (PSH) based on Maximum likelihood solutions in bPTP. Nodes including ASVs from the current study are cartooned to visualize how many sequences are included and labeled with PSH_number according to Figure 1, followed by ML support value for species delineation based on bPTP in parenthesis. Names for two described and two novel species are included in bold for their corresponding PSHs. Following addition of environmental sequences three local PSHs split, these are indicated by adding :1 and :2 after the PSH number. PSHs including only publicly available sequences are collapsed. Placement of previously published sequences from the studied field station is indicated by * after labeled node. Bootstrap supports are calculated from 1000 iterations and indicated only when 80 or higher on branches leading up to PSH from the current study. Full set of support values and sequence names are available in supplementary figure S9.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementarymaterialArchyaeorhizomycetes.pdf