NLPhy: New Leotiomycetes Phylogeny with missing taxa from Siberian peatlands

Project leader: Dr. Danny Haelewaters (dhaelewa@purdue.edu), Purdue University, Department of Organismic and Evolutionary Biology, West Lafayette, Indiana 47907, USA.
Other project members: Tina Melie (tina.jones@colorado.edu), University of Colorado-Boulder, Boulder, Colorado 80309, USA; Nathan Schoutteten (nathan.schoutteten@ugent.be), Ghent University, Ghent, Belgium; M. Catherine Aime (maime@purdue.edu), Purdue University, West Lafayette, Indiana, USA.
Duration: 22 person days during summer 2020, 1 visit (= 88 person-days total).
Station: Mukhrino Field Station (Russia).

Background

Nothing in biology makes sense except in the light of evolution (Dobzhansky 1973), and perhaps nothing in community ecology makes sense except in the light of systematics. Leotiomycetes is an understudied class of fungi that are frequently identified as major constituents of environmental sampling studies. Yet, very little is known about subclass-level relationships *within* this clade, many described taxa remain *incertae sedis* within the class, and their functional roles in these environments is typically based on speculation. Leotiomycetes play key roles in nutrient cycling, ecosystem functioning, and are found in all environments where researchers have explored.

Known ecologies from this clade are highly diverse, including economically and ecologically important pathogens (including powdery mildews on various plants and the causal agent of the white nose syndrome in bats), endophytes, saprobes, mycorrhizae, and species capable of growing in jet fuel and catabolizing hydrocarbons. Researchers are beginning to understand how microbial community composition and function shift over time, across different environments, and across changing climatic conditions. However, interpretation of these results is impeded when the taxonomic status of these fungi is uncertain.

Objectives

An accurate understanding of phylogenetic placement is needed to understand their putative functional roles, and the consequences of their presence or absence in particular environments. Here, we propose to use whole-genome sequencing from fruiting bodies and single cell isolates in addition to genomic-scale phylogenetic analyses to address two specific questions related to our understanding of Leotiomycetes systematics and diversity:

1) What is the diversity of environmentally abundant members of Leotiomycetes in undersampled habitats in the Siberian taiga?
2) Based on improved sampling, how are major clades related to one another in Leotiomycetes?

Specific Aim 1: Describe the diversity of Leotiomycetes in undersampled habitats in the Siberian taiga
With the rise of culture-independent techniques, namely amplicon-based high throughput sequencing, we now recognize the Leotiomycetes as one of the most abundant and diverse groups in a variety of environments (Vrålstad et al. 2002, Tedersoo et al. 2014, Voříšková et al. 2014, Röhl et al. 2017). Due to the meticulous work of researchers doing culture-based studies, the diversity of Leotiomycetes from some environments is not surprising (e.g., agroecosystems [Lauber et al. 2008], freshwater streams

[Baschien et al. 2013], and the high alpine [Schadt et al. 2001]). Yet in other environments (e.g., peat bogs [Lamit et al. 2017], mangroves [Fryar et al. 2019], seagrasses [Gnavi et al. 2014], and animals [Godinho et al. 2019] from the arctic/antarctic) their presence is unexpected and ecologies largely unknown.

Hypothesis 1: The incorporation of molecular phylogenetic data will lead to the description of several species and higher taxa from Siberian taiga.

Specific Aim 2: Present an improved genomic-scale phylogeny of Leotiomycetes
Many studies show that a high diversity of Leotiomycetes taxa cannot be definitively placed below the class or ordinal level. This is true more-so for the Leotiomycetes than most other classes in the phylum Ascomycota, due to the lack of support for the phylogenetic relationships within Leotiomycetes. Poor sampling of loci has led to unsupported/conflicting phylogenetic hypotheses (e.g., Ekanayaka et al. 2019). Whole-genome datasets, however, have the power to resolve *intra-class* and intraordinal-level relationships (Johnston et al. 2019).

Currently, most orders have no genomic-scale data. By including whole genomes from a broader set of taxa in our phylogenetic analyses, deep evolutionary relationships can be resolved. We will also be able to test for conflicts between gene tree *versus* species tree – to assess whether there is incomplete lineage sorting or hybridization, which may have led to the traditional marker loci (such as SSU, ITS, LSU) being uninformative for this class.

Hypothesis 2: The large amount of species *incertae sedis* is an artefact of poor understanding of *intra-class* relationships; the inclusion of broader taxonomic sampling and genomic-scale data in phylogenetic analyses will result in better supported relationships among major clades of Leotiomycetes.

Proposed materials and methods

Sample collection: Targeted collection will happen around Mukhrino Field Station up to the village of Shapsha. From *Andromeda polifolia* leaves in these bog habitats, two species have been recovered that are potentially undescribed: *Hyalopeziza* sp. (?Hyaloscyphaceae) and Hyaloscyphaceae gen. sp. (Filippova & Thormann 2015). These have only thus far been studied based on morphology; we will recollect and isolate tissue for DNA extraction. In the forest litter in mixed coniferous–deciduous forests very close to the village of Shapsha, collaborator N.V. Filippova has collected *Stamnaria* sp. here in 2008–2015, which we later formally described based as *Stamnaria yugrana* on combined morphology and molecular data (Haelewaters et al. 2018).

As mentioned before, Leotiomycetes are dominant taxa in peat bogs (Lamit et al. 2017, Vašutová et al. 2019) and in the mixed coniferous–deciduous forests in the vicinity of the Research Station (*Pinus sibirica*, *Picea obovata*, *Abies sibirica*, with admixture of *Betula pubescens*, *Populus tremula*, *Sorbus sibirica*, *Salix* spp.), Leotiomycetes are common habitants of litter. Where possible, collected samples will be filtered and "washed" over 5 and 2 µm filters multiple times to reduce chances for contamination and maximize target material captured in the DNA isolation process. Rice-sized pieces of tissue or single fruiting bodies will be isolated in 1.5 mL Eppendorf tubes with CTAB buffer for DNA extraction at Purdue. The rest of the samples will be gently dried for long-term preservation.

Single Cell Genome Sequencing: DNA of pieces of tissue or individual fruiting bodies will be extracted using commercial kits that are available at Purdue University (QIAamp DNA Micro Kit, REPLI-g Single Cell Kit, QIAGEN). Whole-genomic Illumina NextSeq sequencing will then be performed. As quality control, Sanger sequencing of the internal transcribed spacer (ITS) barcode will be performed to confirm extracted DNA is from the target organism prior to genome sequencing. Library preparation will be performed using the Illumina Nextera DNA Flex Library Prep kit. Finally, 25-35 libraries will be multiplexed on each high output Illumina NextSeq 2x150 run at the Purdue Genomics Core Facility.

Analysis: All genomes will be quality assessed and trimmed using Trimmomatic (Bogler et al. 2014) and assembled using the SPAdes program (Bankevich et al. 2012) or the metaSPAdes version of that program (Nurk et al. 2017) for fruiting body-derived genomes. For these metagenomic samples, the emergent self-organized map (ESOM) approach will be taken to use kmer frequencies to group the parts of the assembly physically which can then be identified using BLAST or coverage-based annotations. Gene and functional annotation will be performed using the *funannotate* pipeline (Palmer 2016). Proteinortho (Lechner et al. 2011) will be used to identify and obtain single copy orthologous protein data with which we will then align and trim using Muscle (Edgar 2004) and TrimAl (Capella-Guitérrez et al. 2009). IQ-TREE (Nguyen et al. 2015, Chernomor et al. 2016) will be used for maximum likelihood phylogenetic reconstruction.

Expected results and possible risks

It is anticipated that the proposed research will generate new multilocus and genomic-scale data of Leotiomycetes taxa collected from peatlands in western Siberia, contributing to the almost nonexistent data currently available. Newly sequences genomes are crucial for providing resolution and accurate placement of the many families and orders within the class. For example, no genomes or 15-locus (sensu Johnston et al. 2019) sequence data are available for the orders Cyttariales, Lahmiales, Lauriomycetales, Leotiales, Medeolariales, Triblidiales, for 15–20 “clades” within the large order Helotiales, and for the many genera (36) that are currently placed in the class as “incertae sedis”.

We expect that we will find undescribed taxa (species, genera, families) that we will characterize adequately – using morphology, DNA barcoding (ITS), multilocus phylogeny (15 loci), and genomic-scale data. The proposed research will result in both stand-alone taxonomic papers in field-specific journals and scientific contributions in high-impact factor journals (*Persoonia*, *IMA Fungus*, or *Fungal Diversity*), when possible *open access*. We will also share results in talks and posters during conferences, so the research will be more easily disseminated to the international mycological community. All data will be made available in open-access sequence databases and data-sharing portals such as FigShare.

Three risk factors can be identified. The first potential risk is that we may not collect material in the field to make a significant contribution. This will be the first field study combining traditional methods with modern analysis tools in the Siberian region. There is always a risk, but we collaborate with Nina V. Filippova who has extensive knowledge of the area and of the subject organisms. If material is scarce, the visit will lead to better collaboration among the group members – bringing together researchers from Ghent University (Belgium), the University of Colorado-Boulder (USA), Purdue University (USA), and Yugra State University (Russia) – and Dr. Filippova has agreed to share some of her previously collected materials for sequencing. The second potential risk is that DNA extractions may fail. DNA

extractions have been problematic when starting from sporocarp tissue. We will culture samples in the field (after filtering and "washing" over 5 and 2 µm filters multiple times) as to maximize our chances to successful extract DNA. We are actively collaborating with Dr. C. Alisha Quandt regarding her expertise with cutting-edge microculturing techniques and genome sequencing.

Note

We chose to do this project with 4 researchers. One of them (project leader, Dr. Haelewaters) is a recent graduate from Harvard University (2018), whereas two other researchers are Ph.D. students (Melie and Schoutteten). Dr. Aime (graduated 2001) is an established and well-respected professor. We chose this team as to combine different expertise; this will increase taxon sampling success.

References

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. Journal of Computational Biology 19: 455-477.

Baschien C, Tsui CKM, Gulis V, Szewzyk U, Marvanová L. 2013. The molecular phylogeny of aquatic hyphomycetes with affinity to the Leotiomycetes. Fungal Biology 117: 660-672.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120.

Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25: 1972-1973.

Chernomor O, Von Haeseler A, Minh BQ. 2016. Terrace aware data structure for phylogenomic inference from supermatrices. Systematic Biology 65: 997-1008.

Dobzhansky T. 1973. Nothing in biology makes sense except in the light of evolution. The American Biology Teacher 35: 125-129.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792-1797.

Ekanayaka AH, Hyde KD, Gentekaki E, McKenzie EHC, Zhao Q, Bulgakov TS, Camporesi E. 2019. Preliminary classification of Leotiomycetes. *Mycosphere* 10: 310-489.

Filippova NV, Thormann MN. 2015. The fungal consortium of *Andromeda polifolia* in bog habitats. *Mires and Peat* 16: 6.

Fryar SC, **Haelewaters D**, Catcheside DE. 2019. *Annabella australiensis* gen. & sp. nov. (Helotiales, Cordieritidaceae) from South Australian mangroves. Mycological Progress 18: 973-981.

Gnavi G, Ercole E, Panno L, Vizzini A., Varese GC. 2014. Dothideomycetes and Leotiomycetes sterile mycelia isolated from the Italian seagrass *Posidonia oceanica* based on rDNA data. SpringerPlus 3: 508.

Godinho VM, de Paula MTR, Silva DAS, Paresque K, Martins AP, Colepicolo P, Rosa CA, Rosa LH. 2019. Diversity and distribution of hidden cultivable fungi associated with marine animals of Antarctica. Fungal Biology 123: 507-516.

Haelewaters D, Filippova NV, Baral H-O. 2018. A new species of *Stamnaria* (Leotiomycetes, Helotiales) from Western Siberia. *Mycokeys* 32: 49-63.

Johnston PR, Quijada L, Smith CA, Baral H-O, Hosoya T, Baschien C, Pärtel K, Zhuang W-Y, **Haelewaters D**, Park D, López-Giráldez F, Wang Z, Townsend JP. 2019. A multigene phylogeny toward a new phylogenetic classification for the Leotiomycetes. *IMA Fungus* 10: 1.

Lamit LJ, Romanowicz KJ, Potvin LR, Rivers AR, Singh K, Lennon JT, Tringe SG, Kane ES, Lilleskov EA. 2017. Patterns and drivers of fungal community depth stratification in *Sphagnum* peat. FEMS Microbiology Ecology 93: fix082.

Lauber CL, Strickland MS, Bradford MA, Fierer N. 2008. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. Soil Biology and Biochemistry 40: 2407-2415.

Lechner M, Findeiß S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (co-) orthologs in large-scale analysis. BMC Bioinformatics 12: 124.

Nguyen L-T, Schmidt HA, Von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Molecular Biology and Evolution 32: 268-274.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Research 27: 824-834.

Palmer JM. 2016. Funannotate: pipeline for genome annotation. https://github.com/nextgenusfs/funannotate

Röhl O, Peršoh D, Mittelbach M, Elbrecht V, Brachmann A, Nuy J, Boenigk J, Leese F, Begerow D. 2017. Distinct sensitivity of fungal freshwater guilds to water quality. Mycological Progress 16: 155-169.

Schadt CW, Mullen RB, Schmidt SK. 2001. Isolation and phylogenetic identification of a dark-septate fungus associated with the alpine plant *Ranunculus adoneus*. New Phytologist 150: 747-755.

Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, et al. 2014. Global diversity and geography of soil fungi. Science 346: 1256688.

Vašutová M, Martin Jiroušek M, Hájek M. 2019. How to estimate fungal diversity in bogs? Oral presentation, Congress of European Mycologists, Warsaw, Prague. 18 September 2019.

Voříšková J, Brabcová V, Cajthaml T, Baldrian P. 2014. Seasonal dynamics of fungal communities in a temperate oak forest soil. New Phytologist 201: 269-278.

Vrålstad T, Myhre E, Schumacher T. 2002. Molecular diversity and phylogenetic affinities of symbiotic root-associated ascomycetes of the Helotiales in burnt and metal polluted habitats. New Phytologist 155: 131-148.