Screening Fungal Genome Sequencing Data and Culture Collections to Better Understand Bacterial:Fungal Interactions

Geoffrey L. House[1]* (ghouse@lanl.gov), Andrea Lohberger[2], Fabio Palmieri[2], La Verne Gallegos-Graves[1], Julia M. Kelliher[1], Demosthenes P. Morales[1], Armand E. K. Dichosa[1], Debora F. Rodrigues[3], Hang N. Nguyen[3], Saskia Bindschedler[2], Jean F. Challacombe[4], Jamey D. Young[5], Pilar Junier[2], and **Patrick S. G. Chain[1]**

[1]Bioscience Division, Los Alamos National Laboratory, Los Alamos, New Mexico; [2] Institute of Biology, University of Neuchâtel, Neuchâtel Switzerland; [3]Civil and Environmental Engineering, University of Houston, Houston, Texas; [4] College of Agricultural Sciences, Colorado State University, Fort Collins, Colorado; [5]School of Engineering Vanderbilt University, Nashville, Tennessee

Project Goals: Interactions between bacteria and fungi are critical to the above- and below-ground functioning of terrestrial ecosystems, yet little is known about these interactions or how they function. It is increasingly clear that these interactions also underpin multiple DOE research priorities, including understanding the possible effects of Earth system change, overcoming current and future energy and environmental challenges, and developing renewable energy sources. This Science Focus Area (SFA) project seeks to better understand of these bacterial:fungal interactions, including the diversity of bacteria and fungi involved, the functions of these interactions in a variety of conditions, and the mechanistic basis of these interactions. Here we outline computational methods we are using as part of this project to better understand the diversity of these bacterial:fungal interactions represented in publicly available DNA sequence data as well as in fungal culture collections.

https://genomicscience.energy.gov/research/sfas/lanlbfi.shtml

The amount of publicly available fungal genome sequencing data is increasing quickly, due in large part to the 1000 Fungal Genomes Project through the Joint Genome Institute (JGI), with data from over 1200 fungal isolates now represented in JGI's Mycocosm database. However, because many fungi have bacteria and viruses associated with them, these DNA sequence datasets from fungi may also begin to provide information about the associated microbiome of these fungal isolates. Furthermore, because the 1000 Fungal Genomes Initiative deliberately seeks to span the full range of known fungal diversity, this presents a unique opportunity to use these DNA sequences to start understanding the diversity of bacteria that may form associations with a wide range of fungi.

To this end, we have developed a bioinformatics pipeline that consists of commonly used tools and custom scripts to identify signals of bacteria that co-occur with hundreds of different fungal isolates. We begin the analysis with raw DNA sequencing reads from fungal genome projects and then remove all identifiable fungal DNA sequencing reads in order to reduce spurious similarities to bacteria. Next, we assemble the remaining reads into longer contigs that contain more information that can aid their taxonomic classification. For this analysis, we discard reads that do not assemble into contigs. We then use multiple taxonomy classifiers to identify contigs that signal the presence of specific bacteria.

Using this bioinformatics pipeline, we have been able to identify specific fungal isolates that have strong signals of likely associated bacteria (e.g. *Rhizobium* sp. and others), while other fungal isolates have strong signals of bacterial contamination (e.g. *Escherichia coli*). However, for other, ambiguous cases, determining whether the identified bacterial signals likely represent contaminants or whether they represent bacteria that may form associations with the fungi remains a challenge.

To help address the problem of differentiating signals of true bacterial associates from likely contaminants, we have also independently used 16S amplicon screens of more than 200 fungal isolates from multiple fungal culture collections for associated bacteria to look for concordance between the results of these screens and the results of the bioinformatics analysis. In addition, we have used the results from screening the fungal cultures in order to guide our bioinformatics screening, finding a clear signal of *Massilia timonae* in the sequencing reads from an isolate of *Aspergillus glaucus*.

By combining DNA sequence data from both fungal collections and publicly available fungal sequencing data, we have identified strong signals of associated bacteria in some fungal isolates. We will continue tuning this analysis pipeline and will also apply it to a wider variety of metagenome samples, including complex soil metagenomes, to identify potential bacterial associates of fungi from a range of environments. This information can then be used to target specific fungi for isolation from environmental samples and to determine how a range of these bacterial:fungal interactions affect both fungal and bacterial phenotypes.

This SFA is supported under the Computational Biosciences Program of the Office of Biological and Environmental Research in the DOE Office of Science through grant DE-AC52-06NA25396

LA-UR-19-20509