21. Molecular tools and databases to monitor and classify soil fungal and bacterial communities

John Dunbar[1*] (dunbar@lanl.gov), Jason Gans[1], Andrea Porras-Alfaro[1], Gary Xie[1], Kuan-Liang Liu[1,3], Cedar Hesse[1], La Verne Gallegos-Graves[1], and **Cheryl R. Kuske[1]**

[1] Los Alamos National Laboratory, Los Alamos, NM; [2] Western Illinois University, Macomb; [3] Institute of Information Management, National Cheng Kung University, Taiwan

Project Goals: One goal of our program is to develop resources that broadly support the microbial ecology research community. We are conducting three activities that contribute to this goal: 1) production of hand curated fungal large subunit rRNA (LSU) and internal transcribed spacer (ITS) sequence databases, 2) validation of their classification accuracy using a naïve Bayesian fungal classifier and BLAST, and 3) production of quantitative PCR assays targeting bacterial and fungal taxa important in carbon cycling in soils from multiple biomes.

Fungal rRNA sequence databases and classifier validation. Reliable, publicly available sequence databases are critical for interpretation of ecosystem surveys and for design of molecular assays that can be used to monitor specific genes in large-scale ecosystem experiments and regional surveys. In 2011, we published a hand-curated LSU database that spanned the Ascomycota and Basidiomycota and demonstrated the classification accuracy of this database using a leave-one-out-classification scheme (1). This database was delivered to the Ribosomal Database Project (RDP) and has been publicly available. As of fall 2013, the RDP was averaging about 2,200 classifier runs and 925,000 query sequences per month against the LSU database. In the past year, we considerably improved the database in multiple ways. We increased coverage of the Glomeromycota, Chytridiomycota and other basal lineages, added representative non- fungal Eukarya, and resolved taxonomy discrepancies (cases where prior publications had called the same sequence by different names). The updated LSU database will soon be publicly available at the Ribosomal Database Project (RDP) website (http://rdp.cme.msu.edu/) and may be downloaded for local use (2).

Although the LSU gene is superior for phylogenetic placement of unknown sequences, the ITS region is widely used as a taxonomic barcode to identify fungal isolates, resolve species and strain differences, and to characterize diversity. We generated a curated ITS database (about 9,000 sequences) that spans all Fungi and demonstrated the accuracy with which the naïve Baysian classifier can correctly place sequences (3). Using parallel datasets for ITS and LSU, where both regions were sequenced from the same isolate (about 1100 sequences), we compared the classification accuracy of the two regions and found that, from phylum to genus, they gave comparable results across a range of sequence sizes and PCR anchor points. With both genes, species-level classification of sequences from environmental surveys is hampered by the lack of sufficient sequence coverage at this level in the databases.

Our soil surveys from multiple forest ecosystems show that a noteworthy proportion of the LSU sequences retrieved from soils represent substantially divergent, novel clades that are yet not classifiable to finer scale taxonomy. Clearly, both the LSU and ITS databases need to be expanded to improve coverage of environmental sequences that are not currently represented in culture collections.

Quantitative PCR assays for climate change responsive fungi and bacteria. Use of qPCR assays to detect and quantify microbial taxa and gene sequences within a complex background of microorganisms is a scalable, rapid and statistically rigorous approach to track populations of interest. However, the fragmentary nature and growing quantity of DNA-sequence data make group-specific assay design

challenging. We solved this problem by developing a software platform that enables PCR-assay design at an unprecedented scale. The platform provides a powerful capacity to address previously intractable assay design problems. The software accommodates the use of thousands of target and non-target sequences, allows degeneracies, applies sophisticated rejection criteria, and attempts to produce the minimum number of assays to detect a target group. Previously, we used the software to design qPCR assays for Acidobacteria Group 1 (4). We are currently designing assays for other bacterial groups (7 Actinomycete suborders and 10 families) and for fungal groups (11 Ascomycete suborders) of interest in terrestrial carbon cycling.

(1) Liu K-L, A Porras-Alfaro , CR Kuske, SA Eichorst, G Xie (2012) *Accurate, rapid taxonomic classification of fungal large subunit rRNA genes.* Appl Environ Microbiol 78:1523-33.

(2) Cole JR, Q Wang, JA Fish, B Chai, D McGarrell, Y Sun, CT Brown, A Porras-Alfaro, CR Kuske, JM Tiedje (2013) *Ribosomal database project: data and tools for high throughput rRNA analysis*. Nucl Acids Res, doi:10.1093/nar/gkt1244, pg 1-10.

(3) Porras-Alfaro A, K-L Liu, CR Kuske, G Xie (2013) From genus to phylum: large-subunit and internal transcribed spacer rRNA operon regions show similar classification accuracies influenced by database composition. Appl Environ Microbiol, doi:10.1128/AEM.o2894-13.

(4) Gans JD, J Dunbar, SA Eichorst, LV Gallegos-Graves, M Wolinsky, CR Kuske (2012) *A robust primer design platform applied to detection of Acidobacteria Group 1 in soil*. Nucleic Acids Res 1-11 DOI:10.1093/nar/gks238.

Funding statement: The information in this poster was supported by the U.S. Department of Energy Biological System Science Division, through a Science Focus Area Grant (2009LANLF260), and by the Los Alamos National Laboratory LDRD program.