Lawrence Berkeley National Laboratory
LBL Publications

Title

Metagenomic Characterization of Compost and Rain Forest Soil Microbial Communities

Permalink

https://escholarship.org/uc/item/5r33t2vj

Authors

Allgaier, Martin
Reddy, Amitha M.
VanderGheynst, Jean S.
et al.

Publication Date

2009-12-18

Metagenomic Characterization of Compost and Rain Forest Soil Microbial Communities

Martin Allgaier[1]*, Amitha Reddy[12], Jean VanderGheynst[2], Alex Copeland[3], Victor Kunin[3], Patrik D'haeseleer[14], Kristen DeAngelis[15], Julian Fortney[15], Dylan Chivian[15], Paramvir S. Dehal[15], Blake Simmons[16], Terry C. Hazen[15], and Phil Hugenholtz[13]

[1]Joint BioEnergy Institute, Emeryville, CA
[2]University of California Davis, Department of Biological & Agricultural Engineering, Davis, CA
[3]DOE Joint Genome Institute, Walnut Creek, CA
[4]Lawrence Livermore National Laboratory, Livermore, CA
[5]Lawrence Berkeley National Laboratory, Berkeley, CA
[6]Sandia National Laboratories, Livermore, CA

*presenting author
eMail: Phil Hugenholtz (PI) PHugenholtz@lbl.gov
Martin Allgaier (presenter) MAllgaier@lbl.gov

Microorganisms are a promising source of novel carbohydrate-active enzymes (cellulases, hemicellulases, and lignases) since they are primarily responsible for plant biomass degradation in nature. However, most carbohydrate-active enzymes used by industry come from only a few model organisms. To identify new lignocellulolytic enzymes we shotgun sequenced genomic DNA from two sources: a sample of pristine Puerto Rican rain forest soil and a sample obtained from a solid state fermentation experiment in which switchgrass was incubated 30 days in a lab under mesophilic and thermophilic conditions after inoculating with green waste compost from an industrial facility. Both ecosystems display high rates of plant biomass degradation and are therefore prime targets for novel carbohydrate-active enzyme discovery. 454-Titanium pyrosequencing was used to generate metagenome data sets from the two samples resulting in a total of 1,412,492 reads (rain forest: 863,759; compost: 548,733) with an average read length of 424 bases. Reads were quality filtered and trimmed in preparation for comparative analyses with the metagenome analysis tools IMG/M and MicrobesOnline. The complexity of the rain forest soil metagenome precluded assembly, so sequence data were analyzed unassembled. However, for the compost sample, significant assembly occurred resulting in contigs up to 50 kb in length. Preliminary comparative analysis of a fraction of the rain forest soil sequence data revealed more than 2,700 cellulases, hemicellulases, and ligninases including glycoside hydrolases as well as glycosyl transferases, representing ~1% of all predicted protein-coding genes (e.g. compared to 1.2% or 0.03% identified in metagenomic data sets from termite guts or silage surface soil, respectively). The enzyme repertoires present in the two metagenome data sets will be further analyzed to identify new deconstruction enzymes and compared to assess differences in activity and community composition between compost and rain forest soil.