

How can metagenomics be used to address questions in ecology?

Jay T. Lennon, Indiana University

Metagenomics is a potential tool for addressing ecological questions. Three ways were described. (1) Description and discovery. Metagenomics provides information about who is there from the perspective of taxonomy and what functions might be present. In some instances it may be possible to combine these through the assembly of some genomes without cultivation. (2) Identification of patterns based on sequences obtains. This could include finding new patterns as well as testing established ones. An example of the latter is testing whether theories developed for macroecology apply to microbial communities and their function. (3) Prediction. Knowledge about the genetic potential of a soil might be incorporated into ecosystem models and provide a stronger mechanistic basis for these models, which would potentially enhance their predictive power.

Nuts and bolts of DNA sequencing approaches and bioinformatic tools

Dionysios Antonopoulos, Argonne National Laboratory

The current status and challenges relating to DNA sequencing technologies and analysis of sequence data were reviewed. The dynamic nature of emerging DNA sequencing technologies has produced a myriad of challenges not only for the analysis and interpretation of the data itself but for the underlying computation as well. The application of pre-existing approaches for data handling and analysis did not scale with the volume of data that current next-generation DNA sequencing platforms produced leading to a dramatic and untenable computational requirement, specifically for the analysis of shotgun metagenomic datasets. As a result, new strategies have been developed based on reducing data redundancy, adoption of different search algorithms, and taking advantage of cloud computing capabilities. In generating data handling and processing resources for the research community, the ability to augment individual experimental datasets by comparing to others has highlighted the importance of capturing experimentally-relevant information (“metadata”) with the sequence data. Enabling larger-scale studies will create unique opportunities for generating more robust models of microbial community dynamics in the environment.

Scratching the surface of the forest soil metagenome: Facing some challenges

William W. Mohn, University of British Columbia

Metagenomic analysis promises to be a powerful approach to elucidate the functional potential of communities, particularly their capacity to cycle nutrients. A case study was presented analyzing metagenomes from forest soils of the Long-term Soil Productivity Study, a large-scale field experiment. A large Illumina HiSeq data set (200 Gb) yielded very little assembly (0.5 to 5 %). Thus, unassembled data were used to detect experimental treatment effects, focusing on genes encoding lignocellulose decomposition. Short (75-b) HiSeq reads permitted effective analysis of genes of interest. Forest harvesting 12 years prior to our sampling reduced the relative abundance of lignocellulose degradation genes but increased their diversity. Profiles of these genes differed significantly among soil horizons and treatments with varied organic matter removal. Curated databases of lignocellulose degradation genes/proteins were critical to the above analysis, but such databases are lacking for a majority of microbial catabolic genes. Analysis of terpene degradation genes in the data set highlights the difficulty in identifying catabolic genes in metagenomes. The lack of functionally-characterized reference proteins in conjunction with automated annotation of genomes and metagenomes has propagated errors in

gene identification in databases. Complex phylogenies further complicate analysis of catabolic gene/protein families. Bioinformatic and biochemical analyses are required to enable informative analyses of catabolic potential in metagenomes.