

Organism-resolved Metagenomics using ggKbase: Recovering and analyzing thousands of genomes from metagenomic samples

*Brian C. Thomas¹, Jillian F Banfield²

1.Dept. of Earth and Planetary Science, University of California, Berkeley, 2.Dept. of Earth and Planetary Science and Environmental Sciences, Policy and Management, University of California, Berkeley

Metagenomics has the potential to recover near complete and complete genomes for the majority of microbial members present at > 0.2 % abundance level in natural communities, to track changes in community composition across space and time, and to document evolution *in situ*. The field has advanced tremendously since the first community metagenomic study of an acid mine drainage biofilm, more than a decade ago. Metagenomics is now a tool widely used by microbial ecologists. Thousands of research groups worldwide are now collecting complex metagenomic datasets, yet expertise in how to effectively and efficiently use the information is lacking. Currently most metagenomic data analysis methods yield only fragmented, partial genomes, or worse, they attempt to analyze only the read data from a sample and forego any genome assembly. Although such studies provide information about the representation of genes in an environment, much information is also lost. The result is often unsuitable for the development of useful, organism-resolved metabolic models.

Here we describe a multi-faceted approach for genomic analyses using ggKbase. ggKbase is a platform for storing, integrating, managing and analyzing metagenomic data. ggKbase provides intuitive visual binning tools which display key binning traits in a dynamic fashion allowing for the quick creation and assessment of organism bins. ggKbase provides a rapid functional profiling of genomes from a community and generates a detailed analysis of the overall community composition. ggKbase integrates multiple sources of annotation (e.g. KEGG, UniRef etc.) and provides high quality metabolic pathway information. Using ggKbase we have analyzed tens of thousands of genomes from almost 1000 metagenome samples. We will present the ggKbase framework and highlight several use cases.

Keywords: metagenomics