

Rapid automated curation of genomes from metagenomes and integration into KBase

Rohan Sachdeva^{1*} (rohansach@berkeley.edu), Livia Moura², Dylan Chivian³, Jennifer Pett-Ridge⁴, and **Jillian F. Banfield**^{1,3}

¹Innovative Genomics Institute, University of California, Berkeley; ²University of São Paulo, São Paulo, ³Lawrence Berkeley National Laboratory, Berkeley; ⁴Lawrence National Livermore National Laboratory, Livermore

<http://genomicscience.energy.gov/research/sfas/llnlsoil.shtml>

<http://kbase.us>

Project Goals:

The large scale recovery of metagenome assembled genomes (MAGs) has often resulted in the placement of the genome as the operational unit through which microbial biology and systems are understood. The underlying assembled sequences of these mass-produced genomes have characteristic errors from the assembly process that can result in MAGs that do not represent a true biological organism. Ultimately, erroneous genomes limit their utility for making robust ecological, evolutionary, and energetic inferences about microbial systems. Assembly errors can be fixed, but repair necessitates time consuming manual human-guided curation that ultimately restricts the number of curation targets. To address this, we are developing FixAME, an automated curation system that is being integrated into KBase. FixAME rapidly identifies and repairs many common assembly errors across entire metagenomic assemblies and can substantially increase the accuracy and completion of MAGs. FixAME also improves the overall contiguity of metagenomic assemblies while simultaneously sidestepping the bottleneck of costly human-guided manual curation.

Funding statement.

Funding for this project was provided by the Department of Energy - Microbes Persist: Microbes Persist: Building a KBase Foundation for Microbial and Viral Ecogenomics in Soil” under the Lawrence Livermore National Laboratory Soil Microbiome Science Focus Area.