Lawrence Berkeley National Laboratory
LBL Publications

Title
MetaBAT: Metagenome Binning based on Abundance and Tetranucleotide frequency

Permalink
https://escholarship.org/uc/item/3xj1m7nb

Authors
Kang, Dongwan D.
Froula, Jeff
Egan, Rob
et al.

Publication Date
2014-03-18

eScholarship.org Powered by the California Digital Library
University of California

MetaBAT: Metagenome Binning based on Abundance and Tetranucleotide frequency

Dongwan D. Kang*[1, 2], Jeff Froula[1, 2], Rob Egan[1, 2] and Zhong Wang[1, 2]

1Department of Energy Joint Genome Institute, Walnut Creek, CA 94598 USA
2Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720 USA
*Email Address: ddkang@lbl.gov

March 2014

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor The Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or The Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or The Regents of the University of California.

MetaBAT: Metagenome Binning based on Abundance and Tetranucleotide frequency

Dongwan D. Kang* (ddkang@lbl.gov)[1,2], Jeff Froula[1,2], Rob Egan[1,2], and Zhong Wang[1,2]

1. Department of Energy, Joint Genome Institute, Walnut Creek, CA 94598, USA.
2. Genomics Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA.

Grouping large fragments assembled from shotgun metagenomic sequences to deconvolute complex microbial communities, or metagenome binning, enables the study of individual organisms and their interactions. Here we developed automated metagenome binning software, called MetaBAT, which integrates empirical probabilistic distances of genome abundance and tetranucleotide frequency. On synthetic datasets MetaBAT on average achieves 98% precision and 90% recall at the strain level with 281 near complete unique genomes. Applying MetaBAT to a human gut microbiome data set we recovered 176 genome bins with 92% precision and 80% recall. Further analyses suggest MetaBAT is able to recover genome fragments missed in reference genomes up to 19%, while 53 genome bins are novel. In summary, we believe MetaBAT is a powerful tool to facilitate comprehensive understanding of complex microbial communities.