

Lawrence Berkeley National Laboratory

LBL Publications

Title

Metagenomic Bacterial Finishing at JGI

Permalink

<https://escholarship.org/uc/item/7815r2zj>

Authors

Lowry, Stephen
Clum, Alicia
Goltsman, Eugene
et al.

Publication Date

2008-05-28

Metagenomic Bacterial Finishing at JGI.

Stephen Lowry¹, Alicia Clum¹, Eugene Goltsman¹, Hector Garcia Martin², Phil Hugenholtz¹, Alla Lapidus¹

¹ DOE Joint Genome Institute, ² Lawrence Berkeley National Laboratory

DOE Joint Genome Institute, Walnut Creek, CA
2800 Mitchell Drive, Walnut Creek CA 94598

The sequencing to completion of uncultured bacterial genomes from mutualistic communities is a demanding process in the best of cases.

The complexity of the community, the quantity of genomic DNA available, the fraction of the total DNA collected representing the organism under study are all added to the normal difficulties of establishing a complete sequence.

At the JGI we have managed to complete the sequence of three metagenomic organisms, and are investigating a fourth, presenting a considerable range of difficulty.

Candidatus Korarchaeum cryptofilum OPF8, (NC_010482; GI:170289627), is the first of this apparently ancient hyperthermophilic phyletic group to be sequenced (3). The ability to obtain ample DNA of near-monocultural purity and low strain complexity made this the most straightforward sort of metagenomic subject. The target organism constitutes ~40% of the Yellowstone thermal Obsidian Pool community. The community could be maintained in culture, and it was found that *K. cryptofilum* was the most resistant member to SDS lysis, thus allowing DNA purification to better than 90%. Its strain complexity was low as indicated by a SNP rate of ~0.2%.

Some organisms have remote or difficult habitats limiting the availability of source material. This is the case with the thermophile *Desulforudis audaxviator*, (NC_010424; GI:169830219), from fractures in the earth's crust at a depth of 2800 meters in a South African gold mine (4).

The bacteria were collected on filters through which large amounts of subterranean water was passed. The surprising fact that this ecosystem contained but one species fortunately meant that the DNA yield of this one-time-only collection was sufficient to complete the genome.

Considerably more complex situations are the rule, as illustrated by the case of *Candidatus Accumulibacter phosphatis Type IIA str. CU-1*. This and closely related species are the principal actors in the sequestration of inorganic phosphate as intracellular polyphosphate in wastewater treatment facilities.

Bioreactor sludge derived from a working facility in Wisconsin is a physically unresolvable mixture of organisms, with *A. phosphatis* predominating at about 40%. The entire DNA sample was sequenced and the resulting data then subjected to phylogenetic parsing using the Phytopythia (2) binning technique, greatly reducing the complexity of the subclone libraries.

A single 5Mb chromosome was successfully sequenced, along with 3 plasmids of 167, 42 and 38kb.

We are now approaching a still more difficult genome. *Candidatus Endomicrobium trichonymphae* is an intracellular symbiont of a flagellate protist, itself part of the normal hindgut

community of a termite host. It is of interest in the pursuit of the efficient breakdown of cellulose and lignin necessary in the hoped-for use of bulk plant materials as CO₂-neutral fuel stocks.

Again partitioning with Phytopythia was absolutely necessary. So far this has yielded some 42,000 Sanger reads. Of these about 40% of the assembled contigs are similar to related organisms. Also contigs from 454 pyrosequencing contributed about 750,000 bp of additional coverage (~0.5x). Additional difficulty arises from sample size and strain complexity. There are several closely related organisms with substantial representation. There are several hindguts in the sample, and the lines of descent may be relatively independent even at the protist level.

(1) Martin, H.G., Hugenholtz, P., et al. *Nature Biotechnology*, v24, no.10, 1263-69.

2006. *Metagenomic Analysis of two Enhanced Biological Phosphorous Removal*

(EBPR) sludge communities

(2) McHardy et al. *Appl, Environ Microbiol* 2000, 66(3):1175-1182, *Accurate phylogenetic classification of variable-length DNA fragments.*

(3) James G. Elkins, et al., *PNAS*. 2008 (in Press), *The Korarchaeota: Archaeal orphans representing an ancestral lineage of life*

(4) Dylan Chivian, et al., *Science* 2008 (in Press), *Environmental genomics reveals a single species ecosystem deep within the Earth.*

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

LLNL-ABS-XXXXXX

LBNL-XXXX Abs.