Extrachromosomal Plasmid DNA Project

Ankita Kothari[1*], Yu-Wei Wu[1], Marimikel Charrier[1], Lara Rajeev[1], Andrea M. Rocha [2], Terry C. Hazen[2], Paramvir S. Dehal[1], Dylan Chivian[1], Sarah J. Spencer[3], Eric J. Alm[3], Steven Singer[1], Aindrila Mukhopadhyay[1], Adam P. Arkin[1] and **Paul D. Adams**

[1]*Lawrence Berkeley National Laboratory, Berkeley, CA;* [2]University of Tennessee, Knoxville, TN; [3] Massachusetts Institute of Technology, Cambridge, MA.

http://enigma.lbl.gov

Project Goals: The extrachromosomal plasmid DNA project is one of the discovery projects funded by ENIGMA. It is a short term, high impact, investigatory effort to study the plasmid populations of the ENIGMA wells. The goal of this project is to exclusively explore the prevalence of plasmid DNA in these communities. Additionally, the project aims at studying the relevant functional genes that are typically encoded on the plasmids, conferring advantageous traits to the host, in these communities.

Plasmids are autonomously replicating extra-chromosomal genetic elements that often act as mediators of horizontal gene transfer in the environment. Plasmids host and distribute non-essential genes, independent of the host's chromosome, thereby benefitting the host bacteria in certain specific environmental conditions. Native plasmids typically range from 2 kb to 250 kb in size and have been shown to be present in 10-30% of the cultivated isolates from varied environments. The best-characterized wells at the Oakridge site are now documented to contain several hundred bacterial strains, many of which are likely to contain plasmids.

This is the first study to selectively isolate and analyze the plasmid population from these sites. To optimize a robust method that isolates a range of plasmid sizes, we developed a model system comprising of three strains containing plasmids of sizes – 5kb, 48kb and 202kb in equal proportions, and tested the potential of various alkaline hydrolysis based methods to isolate plasmids from the serial dilutions of the mixed population. The presence of each plasmid was determined by targeting a unique plasmid borne gene via qPCR. In order to get rid of genomic DNA contamination, the isolated DNA was subjected to Plasmid-Safe-ATP-Dependent DNase and the lack of genomic DNA contamination was confirmed using degenerate primers targeting the 16s rRNA coding sequence. The total plasmid DNA thus obtained was amplified using Phi29 DNA polymerase to generate high-quality template for use in DNA sequencing. To increase the sensitivity of the plasmid isolation procedure, the extraction procedures and Phi29 amplification conditions were optimized. Subsequently, plasmid DNA was isolated from the wells GW460 and GW456 and subjected to deep sequencing using the Illumina paired-end protocol. The reads obtained were trimmed using Trimmomatics, assembled using IDBA-UD (Iterative De Bruijn graph Assembler for reads with Highly Uneven Sequencing Depth) and subjected to MG-RAST to produce gene calls, functional annotations and taxonomic classification.

A total of 42543 (including 130 circular contigs) and 32313 contigs (including 760 circular contigs) above 2kb size were detected from the wells GW456 and GW460, respectively. These encode several known plasmid associated genes along with genes involved in secondary metabolism, antibiotic resistance, metal resistance, and nitrogen metabolism, to name a few.

This material by ENIGMA- Ecosystems and Networks Integrated with Genes and Molecular Assemblies (http://enigma.lbl.gov), a Scientific Focus Area Program at Lawrence Berkeley National Laboratory is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Biological & Environmental Research under contract number DE-AC02-05CH11231