Title
Finishing of Spirochaeta aurantia M1

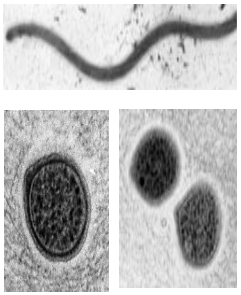
Authors
Clum, Alicia
Lowry, Steve
Rabkin, Brian A.
et al.

Publication Date
2008-05-22

Finishing of Spirochaeta aurantia M1

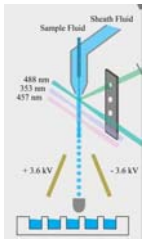
Alicia Clum, Steve Lowry, Brian A. Rabkin, Hui Sun, Alla Lapidus
Lawrence Berkeley National Laboratory
DOE Joint Genome Institute, Walnut Creek, CA 94598

DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE



TEM of whole *S. aurantia* strain M1 showing periplasmic flagellum (stained darkly)
Acknowledge: William Ghiorse (Cornell University)

Shotgun sequencing and finishing of an isolate of the Spirochaeta aurantia M1 genome, a free-living nonpathogenic Spirochete, is in process at the Joint Genome Institute. S. aurantia M1 is being sequenced due to its proximity on the phylogenetic tree to bacteria present in the termite hindgut that were partially sequenced during a metagenomic project at the JGI ,and, for improving the understanding of pathogenic Spirocheates through comparative genomic studies.



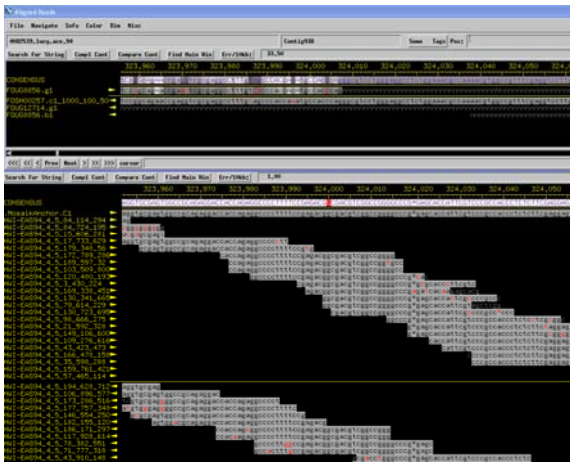
Illumina for polishing

Part of finishing at the JGI is polishing. Our current standards require that each base has to meet quality thresholds. We have begun to use Illumina data to polish in a time and cost effective manner. Our group has developed a tool to align Illumina data to a 454-sanger hybrid assembly. This takes care of a vast number of reactions. However often there are still areas that need additional verification. These remaining areas are verified using traditional Sanger sequencing.

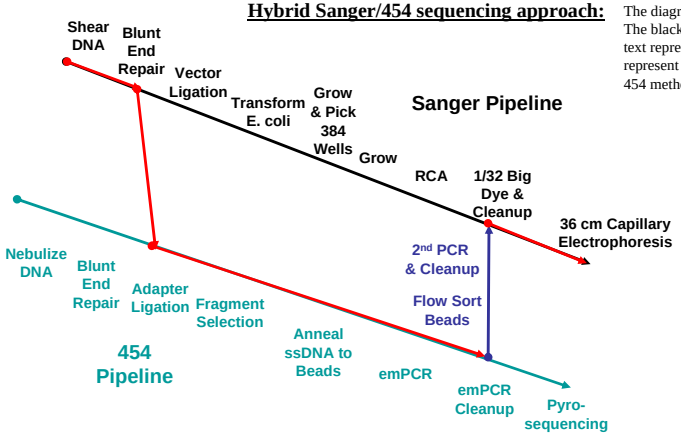
Part A of the screenshot below shows an area of the genome covered by 454 and partially covered by 1 low quality read. Part B shows the Illumina reads aligned to the reference. Bases that are "solexa supported" don't need further verification.

A.

B.

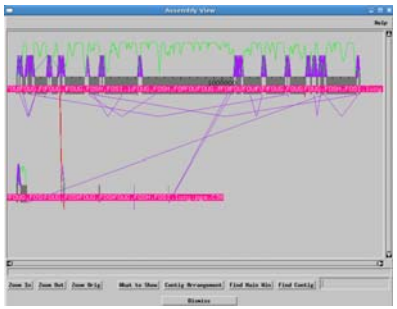


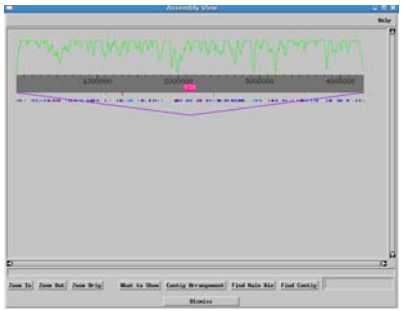
Hybrid Sanger/454 sequencing approach:



The diagram on the left compares 3 different pipelines. The black text represents the Sanger line and the teal text represents the 454 pipeline. The red and blue represent the path utilized for the combined sanger-454 method.

Sorting by flow cytometry
The beads are run through the sorter and are analyzed by lasers which determine the light scattering and fluorescent characteristics of whatever is being sent through (viruses, cells, beads) . Once a target is selected, in this case based on the size of the 454 beads the flow sorter is able to separate each cell into a droplet. These droplets pass by two electrical plates. If the droplet has the proper characteristics it gets charged which diverts the droplet into a plate. For this experiment we sorted the droplets into a 384 well plate so that every well contained one 454 bead. PCR cocktail was then added to these wells followed by cleanup using exo-sap and big dye chemistry.





The initial assembly including fosmids, 454, paired reads from a hybrid 454-Sanger library, and solexa data. The hybrid 454-Sanger approach takes advantage of a bulk cloneless library preparation method involving the amplification of long DNA fragments onto beads, by emulsion PCR, and the subsequent sorting of individual beads by flow cytometry for seeding the Sanger chemistry process.

Almost all gaps are spanned my fosmids. Fosmids were important for gap closing because of their large insert size. The hybrid 454-Sanger library was 2kb so very few gaps were spanned by this library.

Assembly after a few rounds of primer walks. The assembly is in one closed, circular contig. A few areas still remain to be polished. These are the areas were either the solexa depth wasn't high enough to verify the base or the base disagreed with Sanger and 454.

Further improvements:

Further improvements include making a larger insert library. In terms of repeat resolution and gap spanning a 2kb library isn't an ideal size. The sanger/454 hybrid process could also be developed further by sorting the emulsions rather than the 454 beads.

Other applications:

The flow sorter can be used with anything that can be fluorescently labeled or sorted by size. This can allow you to sort a single population out of a community. The flow sorter can also be used in single cell genomics.

FF0056

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.