

Lawrence Berkeley National Laboratory

Recent Work

Title

Microbial Finishing at DOE Joint Genome Institute (JGI): Sequencing Difficult DNA Templates

Permalink

<https://escholarship.org/uc/item/7065h7wz>

Authors

Martinez, Michele

Richardson, Paul

Lapidus, Alla

Publication Date

2006-04-19

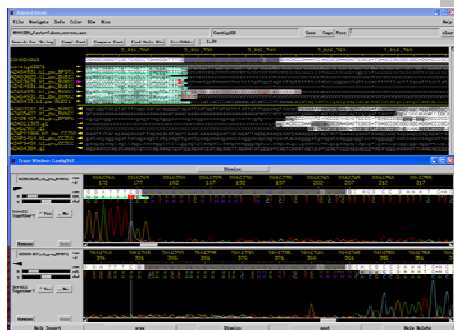
Microbial Finishing at DOE Joint Genome Institute (JGI): Sequencing Difficult DNA Templates

Michele Martinez, Paul Richardson, and Alla Lapidus

Problematic regions include, but are not limited to:

1. GC rich
2. Tandem Repeats
3. AT rich
4. Homopolymer Stretches

The US DOE Joint Genome Institute (JGI) mission is to provide the scientific community with high-quality finished genomes. Approximately 300 microbial genomes are currently in the JGI pipeline and to date, 65 have been completed. The objective of the Microbial Finishing laboratory is to process sequencing reactions in order to close physical gaps, sequence gaps, and to increase quality of reads. Since most of the genomes contain complex regions which are difficult to sequence with standard protocols, the lab must use a multitude of techniques specialized for each project. Problematic regions, for example, can be GC-rich or contain hairpin loops, have long homopolymer stretches, can be AT-rich, and/or contain tandem repeats of variable length. Gap closer in such regions is expensive as well as time-consuming, since it requires extensive troubleshooting strategies. Approaches include, optimizing reaction conditions, applying various sequencing chemistries, sequencing the opposite strand, and additional manual editing. For genomes with $\geq 65\%$ GC content, we use a four step approach to sequence through difficult regions: DMSO, Sequence Finishing Kit (SFK), PCR, and shatter libraries. This strategy has allowed JGI's Microbial Genome Finishing Group to complete a number of complex microbial projects, such as, *Frankia* (~75% GC-rich) and *Thermobifida fusca* (~68% GC rich).

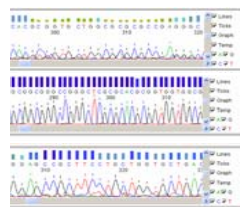


Mycobacterium sp MCS ~ 69% GC content

Identify problematic region

Future Development:

3% of the genomes in the JGI pipeline have greater than 60% AT content. These genomes are more difficult to clone leading to higher number of uncaptured gaps when compared to those with lower AT content. Also, physical gaps and the polishing has proven to be difficult with standard sequencing strategies. For example, *Prochlorococcus* sp 9215 (~70% AT) content, has benefited from 454 data. However, confirming consensus (with 454 only reads) has not been completely successful. Therefore, it is necessary to research and develop new methods of approach in this area.

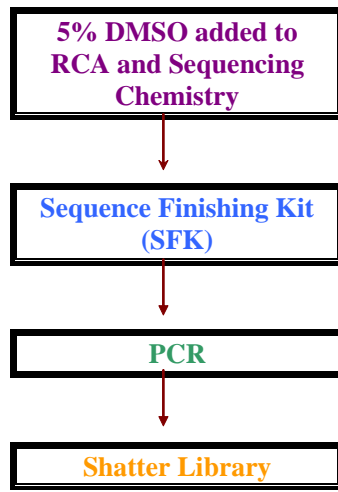


Standard RCA and Sequencing Chemistries

5% DMSO in RCA and in Sequencing Chemistries

SFK and standard Sequencing Chemistries

Four Step Approach

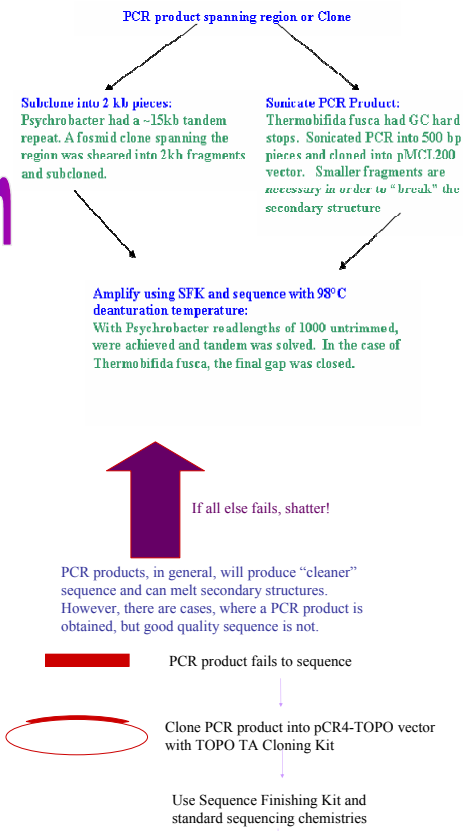


Other options to use during sequencing include:

- 98°C soak for 10 minutes
- 98°C denaturation temperature
- Increase number of cycles to 35

Microbe	Overall GC content of genome	Average Read Length Standard Resection	5% DMSO
<i>Bacillus cereus</i> N10931-50	-36% GC	508	591
<i>Chlamydia trachomatis</i> G3908	-41% GC	541	546
<i>Mycobacterium vanbovenii</i> Pvg-1	-68% GC	599	594

A small sample set was tested to determine how 5% DMSO in sequencing reactions would effect genomes with different GC content. This experiment needs to be conducted on a larger scale, but preliminary results indicate that 5% DMSO can be used on high AT rich genomes as well. The read lengths are not increased, however, for high-throughput production this will allow, in the future, for all projects to be processed with DMSO. In addition, within high AT rich genomes, there are still areas with secondary structures, so this may help reduce the amount of finishing.



Standard sequencing of PCR product failed, after cloning and use of SFK, BHXU14 and BHXU15 closed the gap.