Lawrence Berkeley National Laboratory
Recent Work

Title
JGI Microbial Sequencing Process

Permalink
https://escholarship.org/uc/item/2hz6536f

Authors
Rio, Tijana Glavina del
Daum, Chris
Deshpande, Shweta
et al.

Publication Date
2009-05-27
Microbial Quality Analysis (Fig 1, Step 4)
The QA group works closely with finishers to ensure that projects have met internal quality specifications before they receive final contamination and genome builds before final contamination and genome builds. The QA group identifies potential problems, such as contamination and genome builds, and ensures that the genome builds meet all quality standards.

Initial QC of microbial projects:
This stage begins with the received microbial sample. The sample is first screened for contamination and quality control before being subjected to further analysis. The sample is then sequenced, and the resulting data are analyzed to determine the presence of contaminants and to assess the quality of the sequence data. The final results are used to determine whether the microbial sample is suitable for further analysis.

Microbial finishing (Fig 1, Step 10)
JGI finishing process: Once a draft assembly has been generated and QC’d, it is ready for finishing to begin. An in-house developed software tool creates subprojects for each gap. In-silico attempts are made to close gaps using existing unassembled scaffolds and illumina data. Any remaining gaps are tackled by PCR-based techniques. Each scaffold is then amplified using PCR and the resulting products are sequenced. Once all gaps are closed, illumina data is used to polish the genome. Any areas that are still unknown are subjected to resequencing.

LAMB: Finishing: The initial draft assembly contains many contigs and scaffolds. During the first round of finishing, the initial scaffold assembly is converted into a phrap program by making small reads from the consensus, collecting the read pairs in the 454 paired-end library. The Perseus/Consed software package is used for sequence assembly and quality assessment in the following finishing process. After the shotgun stage, the final consensus scaffold is assembled using parallel phrap. Post-quality control, the scaffold is filtered using gap resolution (CFF Han, unpublished), Delfiner (Han, 2006), or sequencing cloned bridging PCR products with subclones. Contigs are finally assembled into contig sequences using CAP3 (Huang, 1999). Gaps between contigs are closed by editing in Consed, by PCR and by Bobo PCR primer walks. Additional reactions are necessary to close gaps and to ensure the quality of the final scaffold. The final scaffold sequence is completed with an error rate less than 1 in 100,000.

IMG: (Fig 1, Step 13)
IMG is a powerful data management platform that supports timely analysis of genomes from a comparative, functional, and evolutionary perspective. The major motivation of IMG is to provide a data management platform that can store and process large amounts of genomic data efficiently. The platform is designed to support data analysis and annotation of bacterial, plant, and fungal genomes. The IMG platform can be used to analyze genome assemblies, to perform comparative genomics, and to generate comprehensive summaries of genomic data.