Recent Work

Title
Jgi sequencing projects: statistics and timelines

Permalink
https://escholarship.org/uc/item/5tr034fr

Authors
Rio, Tijana Glavina del
Barry, Kerrie
Goodwin, Lynne
et al.

Publication Date
2007-03-28
JGI Sequencing Projects: Statistics and Timelines

Authors: Tijana Glavina del Rio, Kerrie Barry, Lynne Goodwin, Miranda Harmon-Smith, Harris Shapiro, Susan Lucas and David Bruce.

The DOE of Energy’s (DOE) Joint Genome Institute (JGI) is one of the major publicly funded high throughput sequencing centers. The current capacity of the Production Genomics Facility (PGF) in Walnut Creek, California is approximately three billion bases per month, generating a total of 55 million lanes this year. JGI sequencing projects are initiated through one of three peer reviewed programs: Community Sequencing Program (CSP), DOE Microbial Program and the Laboratory Science Program (LSP). This poster will present an overview of project statistics for 2006 and current projects for 2007. In 2006, the JGI processed a collection of DOE mission relevant sequencing projects ranging from prokaryotes to eukaryotes as well as several microbial communities. The poster will also describe how projects are scheduled for production sequencing and display tools used for tracking projects to their completion. Project timeline from initiation to completion will also be presented.

Introduction

The DOE Joint Genome Institute (JGI) is a “virtual institute” that integrates the genomic capabilities of six partner institutions: Lawrence Berkeley National Laboratory, Lawrence Livermore National Laboratory, Los Alamos National Laboratory, Oak Ridge National Laboratory, Stanford University, and Pacific Northwest National Laboratory. After completing the sequencing of the human Genome project (Chromosomes 5, 16 and 19), JGI has shifted its focus to the non-human components of the biosphere, particularly those relevant to the science mission of the Department of Energy. The menu of completed projects includes wide variety of microbes and microbial communities as well as many important eukaryotic model systems such as fish (Fugu rubripes) and sea squirt (Ciona intestinalis). JGI has also sequenced a frog ( Xenopus tropicalis), a green algae (Chlamydomonas reinhardtii), two diatoms (Phaeodactylum tricornutum and Thalasiosira pseudonana), a white rot fungus (Phanerochaete chrysosporium), many filamentous fungi (Trichoderma reesei, Aspergillus niger, Laccaria bicolor, Myxospherae graminicola, Nectria haematococca, Phyomyces blakesleeanus), poplar tree (Populus trichocarpa) as well as a number of fungal-like plant pathogens (Phytophthora ramorum and Phytophthora sojae). There are three major DOE-directed sequencing programs that utilize the high-throughput sequencing of the JGI: Community Sequencing Program (CSP), DOE Office of Biological and Environmental Research (BER) Microbial Sequencing Program Laboratory Science Program (LSP) and DOE Microbial Program. You can find more information about these programs on our website: http://www.jgi.doe.gov/programs/index.html

Sequencing Strategy

The PGF sequencing strategy employs the whole-genome shotgun sequencing method to produce high-quality draft sequence. For each project a 3-Kb, 8-Kb, and 40-Kb DNA library is created and sequenced from both sides of the library insert, producing paired ends, resulting in approximately 8-9x depth. In support of the eukaryotic projects, cDNA libraries are generated in order to allow for the construction of a large and with annotation efforts. For each large genome 15 fosmid clones are selected and subconed for better size estimation. Recently, we have changed the sequencing strategy for the prokaryotic projects by replacing a 3-kb library with a 454 library for additional coverage and increased coverage for the 8-kb library 4x to 6x. For all projects, sequenced reads are deposited in the GenBank Trace Archive at NCBI. Reads are aligned by using various genome assemblers to produce the primary draft assembly, which consists of contigs linked into larger scaffolds by paired-end information. The PGF and several partner institutions (Stanford University, Los Alamos National Laboratory, Lawrence Livermore National Laboratory and Lawrence Berkeley National laboratory) perform finishing work (gap closing, quality improvement, and assembly verification) for organisms that require this level of refinement. All genomes have at least a minimal automated annotation, and most may be searched on genome browsers via the Genome Portal. Eukaryotic Genome Portal link: http://genome.jgi-psf.org/.

For Prokaryotic Genome Portal link: http://genome.jgi-psf.org/.

Project Flow

Once a project is approved through one of the three programs, it is ready to enter the sequencing process. The starting material can be converted to four different library types, depending on the project scope: WGS library, 16/18S library, EST library or 454 library. Samples will go through the following process steps: library construction, sequencing, quality assurance, assembly, finishing, annotation and analysis. Various quality control measures have been implemented throughout the process in order to ensure utmost quality of the DNA sample and constructed library before large scale sequencing begins. Figure 1 shows the JGI sequencing project pipeline on a high level. If a project requires more than one library type to be constructed, the ultra high throughput strategy is to have all libraries for each project run concurrently through the process so that all of the different data is ready for final assembly and analysis at the same time.

Fig. 1. JGI Sequencing Project Pipelines

Project Management

As high throughput sequencing centers move from managing a small number of large projects to managing many simultaneous small projects, the ability to govern schedule, cost, quality, and project management is essential. To better handle the tracking, organization and flow of sequencing projects through the JGI sequencing pipeline, the JGI Project Management Office (PMO) was formed in January, 2006. The PMO consists of a team of five project managers. Each project manager is dedicated to a group of projects based on specific microbial communities. Throughout the life cycle of a sequencing project, the project managers oversee its stepwise progression through the JGI sequencing pipeline with particular emphasis on active, ongoing communications with the PI and associated collaborators. The JGI Global Project Tracking System (GPTS) database is the project manager’s primary resource for communicating project progress through the sequencing pipeline process to both the internal and external collaborators. For tracking projects on a library level, a simple excel spreadsheet is used for this purpose (Fig 2 and 3). This is necessary in order to know the status of a project and all of its many components.

Fig. 2.

Fig. 3.

Prioritization and Scheduling of Projects

The following considerations and guiding principles are used to determine the order in which projects enter the pipeline for sequencing:

1. Project Initiation completed-User Agreement signed, Statement of Work (SOW) finalized, Required Information (RI) sheet received from collaborator.
2. DNA sample quality- DNA is received and passed DNA QC.
3. Library Creation-Library is received and initial QC is performed.
4. Quality Control- DNA library is QC’d and qualified.
5. Sequencing- DNA library is sequenced.
6. Assembly and Analysis- Data is archived, assembled, annotated, and analyzed.
7. Finishing, Annotation and Data Release- Data is finalized, annotated, and released.

Fig. 4.

Fig. 5.

Status of 2006/2007 Large Eukaryotic Projects

<table>
<thead>
<tr>
<th>Program/Year</th>
<th>Genus Species</th>
<th>Status</th>
</tr>
</thead>
<tbody>
<tr>
<td>CSP 2006</td>
<td>Arabidopsis lyrata</td>
<td>Production Final 8x</td>
</tr>
<tr>
<td>CSP 2006</td>
<td>Capsella rubella</td>
<td>Production Lib QC</td>
</tr>
<tr>
<td>CSP 2006</td>
<td>Mitulus guttuliflorus</td>
<td>Production Final 8x</td>
</tr>
<tr>
<td>DOE 2006</td>
<td>Pseudorizhiza multiseris</td>
<td>CLN-47 Awaiting Sample</td>
</tr>
<tr>
<td>DOE 2006</td>
<td>Saccharum bicolor</td>
<td>Final Assembly</td>
</tr>
<tr>
<td>DOE 2006</td>
<td>Daturastrum nigrum UTEX-1644</td>
<td>Awaiting Sample</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Aquilegia formosa</td>
<td>Awaiting Sample</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Brachypodium distachyon</td>
<td>Production Lib QC</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Gossypium raimondii</td>
<td>Production Lib QC</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Mammoth esculenta</td>
<td>Production Lib QC</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Glycine max Williams 82</td>
<td>in 4x Assembly</td>
</tr>
<tr>
<td>CSP 2007</td>
<td>Thelungiella saltirea</td>
<td>Awaiting Sample</td>
</tr>
</tbody>
</table>

Fig. 6. shows the status of 2006 projects. We have 28 projects in production, 9 in final assembly and 25 pending. Fig. 7. shows the status of 2007 projects. We have 20 projects in production, 91 pending and 2 completed.

Fig. 6.

Fig. 7.

JGI Project Status 2006

JGI Project Status 2007

Conclusion

JGI has gone through a major transition from sequencing human genome using BAC by BAC approach to sequencing different genomes using the shotgun approach. The three major scientific programs allow a wide variety of projects to enter the sequencing pipeline and address DOE mission as well as provide the scientific community with high throughput DNA sequencing capability. All projects are entered into the data base and tracked and scheduled through the process. Scheduling and tracking of projects ensures meeting established sequencing timelines and that no project is left behind. Quality control measures implemented at different steps ensure sample quality and prevent contaminants from being present in the final product. As we grow, JGI will continue to provide integrated, high-throughput sequencing and computational analysis to enable genomic-scale/systems-based scientific approaches to DOE-relevant challenges in energy and the environment.

UCRL-POST-229322