

Lawrence Berkeley National Laboratory

Recent Work

Title

Production workflow tracking and QC analysis

Permalink

<https://escholarship.org/uc/item/27s9q214>

Authors

Kimball, Heather
Trong, Stephan
Kobayashi, Art
[et al.](#)

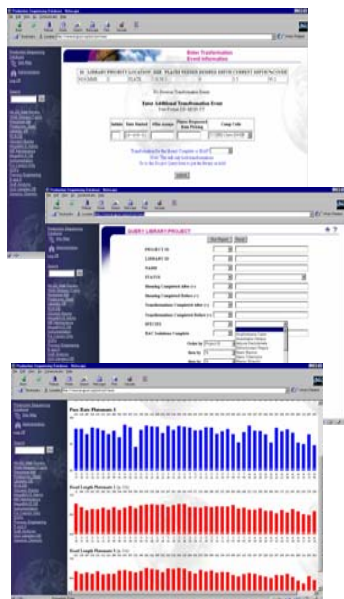
Publication Date

2001-10-23

Abstract

The Joint Genome Institute Production Genomics Facility (PGF) has produced over 2.75 billion bases of draft paired-end sequencing since January 1, 2001. Our sequencing methodologies incorporate two types of DNA template generation: inoculation/SPRI purification and Rolling Circle Amplification (RCA). In order to manage the flow of samples through these processes, a robust database tracking system was developed using ORACLE. Data input and reporting for the workflow has been produced using a combination of commercial database development software and in-house programs. These include ORACLE's WebDB and CGI Perl programming. By leveraging the rapid report and form development cycle using WebDB and augmenting this with the flexibility of in-house programming, we have efficiently deployed a crucial laboratory information management system for our data tracking.

Tools



The PGF database workflow tracking system was implemented using ORACLE 8i. Software development was accomplished using ORACLE's WebDB, PL/SQL and CGI/DBI Perl for generating web-based input forms and reports.

WebDB, an ORACLE web-based applications development tool, was the primary choice for generating data input forms and table-based query reports because of its rapid development time. Although WebDB does provide some Database Administration utilities, its applications development capabilities were the main reason for its use in building our workflow tracking system. Through the use of this tool, we were able to quickly generate forms on tables, packages or procedures using a built-in form wizard. Augmented with PL/SQL, a database procedural language, and html, we incorporated data integrity checks and utilized bar coding for inputs and created sample sheet requests for the MegaBACE sequencers.

Data reporting between each workflow process is also implemented using WebDB. Again, development time is rapid, usually less than 15 minutes depending on complexity. Features within the reports included customizable queries based on input parameters, links between

reports and forms for displaying multiple views of a give data set and an overall standard interface accessible through a web browser.

Because of the ease and quick delivery of forms and reports developed using WebDB, we were able to support the constant change in process flow within the organization. As with any applications development tool, WebDB does have its limitations. For projects that require functionalities outside the scope of this tool, we developed in-house software using Perl to address these issues.

Perl provides the most flexible database interactive solutions. Perl has been used for the transfer of MegaBACE run files from the sequencers to Unix, generation of sample sheets, parsing of barcode files from various instruments and most of the QC related reporting. It has also been used to monitor and clear out disk-space on the MegaBACE instruments and has been preferable to SQL Loader for loading data into the database from excel or tab-delimited files.

By implementing both the rapid report and form development cycle using WebDB and augmenting this with the flexibility of in-house programming, we have effectively deployed a crucial workflow tracking system for our data management.

Project Goals

The goal of the production workflow tracking system was to create a user-friendly, accurate and robust system to track clones, libraries and plates through the PGF sequencing process. The system needed to be flexible, scalable, accessible through a central location and efficient. A web-based system was preferred to avoid requirements for specialized client side software. Completeness and accuracy of data input needed to be controlled while still allowing the users to have primary control of the workflow. Bar coding had to be integrated to avoid errors in manual data entry. The system needed to track multiple sub-processes simultaneously and required easy configuration to support different experimental phases during integration. Finally, any such system used for both troubleshooting and determining project status needed to be accurate, current and robust.

Production Workflow Process

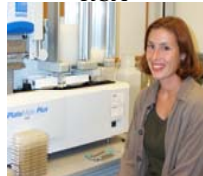
(Process Steps and details tracked)

Library Creation



- Operators
- Batch
- Vector
- Comp.Cells
- LB Lots
- Hydroshears
- Pickers

RCA



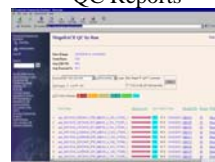
- Operators
- Batch
- Reagents
- Platemates
- Multidrops
- PE 9700 Cyclers

Sequencing Chemistry



- Operators
- Batch
- Primers
- Sequencing Kit
- MJ Cyclers
- PE 9700 Cyclers
- Hydras
- Twisters

QC Reports



- Trace Q Scores
- Plate Q Score Averages
- Plate Q20 Average (Read Length)
- Plate Pass Rate

Capillary Loading



- 84 MegaBACE 1000s
- Operators
- Buffer
- Matrix
- Arrays

TEG Purification

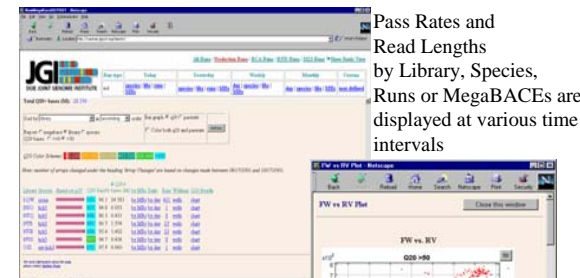


- Operators
- Batch
- Beads
- TEG
- Bio FX

Project Scope

The scope of the project was to track the life cycle of a clone as it went through the PGF workflow. Descriptive information is imported from a preexisting Physical Mapping Database. A clone is then assigned to a queue in the Library Creation subgroup. The result of the Libraries process is a set of 384 well plates linked to a library code and the parent clone. These plates are then tracked through RCA and divided into forward and reverse destination plates. The corresponding plates are then tracked through cycle sequencing. After cycling, each 384 well plate is decondensed into four 96 well plates to be loaded onto the sequencers. After capillary loading and sequencing, the sequence data is base called and sent from the MegaBACE instruments to Unix to be processed through Phred and Phrap for assembly. The QC data is stored in the database at both the plate and trace level and provides the metrics by which all the workflow steps are measured. Tracking of data such as depth and percent cover are stored after each clone assembly. Currently, a workflow system for post sequencing Draft Analysis and Functional Analysis is being developed.

Examples of QC Reports



Pass Rates and Read Lengths by Library, Species, Runs or MegaBACEs are displayed at various time intervals

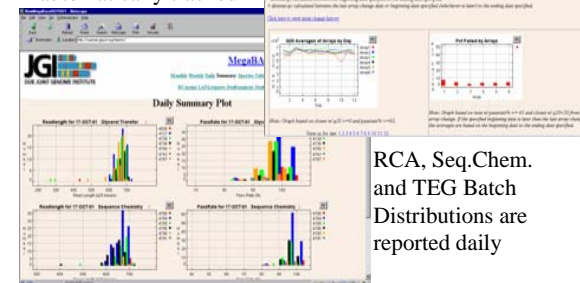
FW and RV Correlations are reported for each plate

On-line trace view is available for each clone



Read Lengths for 96 well or 384 well maps are displayed for each plate

MegaBACE Array Performance is automatically tracked



RCA, Seq.Chem. and TEG Batch Distributions are reported daily

This work was performed under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, under Contracts No. W-7405-Eng-48, No. DE-AC03-76SF00098, and No. W-7405-ENG-36. LBNL-48997