The Finishing Laboratory at the Broad Institute

Daniel Bessette, Nicole Allen, Mostafa Benamara, Chelsea Dunbar, Xiaohong Liu, Tashi Lokyitsang, Rakela Lubonja, Charles Matthews, Anna Montmayeur, Tamrat Negash, Thu Nguyen, Michael FitzGerald

The Broad Institute, Cambridge, MA, USA

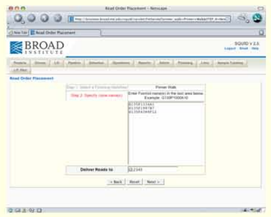
What We Do

Using a combination of high-throughput automation and small-scale custom techniques, the Laboratory Finishing group at the Broad Institute improves rough draft genomes by providing new sequencing data to fill in gaps and resolve conflicts. We work in collaboration with the Computer Finishing and Production Sequencing teams to fulfill work orders that include

- Primer Walking (Fosmid and Plasmid)
- Transposition (Fosmid and Plasmid)
- Resequencing (Fosmid and Plasmid)
- Custom PCR
- Shatter Library creation (PCR products)

Built into these general workflows are special chemistries to address difficult-to-sequence regions. Recently the Shatter Library workflow was revamped and now utilizes a topoisomerase-based cloning system for ease of use and cost reduction. This modest-sized group has the capacity to work on at least one mammal and several fungal- and bacterial-sized genomes simultaneously. The structure of the group also allows it to have significant flexibility and scale-up capability: our ability to address clone tiling path based projects (e.g. mouse), pure whole genome finishing (e.g. Mycobacterium tuberculosis) and a mix of the two (e.g. dog) gives us flexibility, our workflow (rather than project) based tasks allows for increased efficiency, personnel cross-training allows us to shuffle personnel to various workflows according to need, and a mix of automated and manual protocols as well as a seasoned LIMS enable us to tend to large or small orders very efficiently.

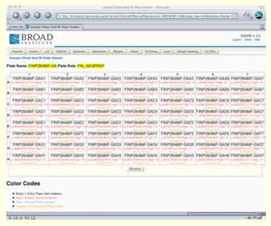
Informatics



Computer Finishing work orders placed in SQUID*

Computer Finishers initiate all work to be done in the lab. They chose from a menu of laboratory techniques and sequencing chemistries.

*Sequencing Query User Interface and Database



Finishing Lab retrieves work orders and performs lab work

Orders are tracked using a network of LIMS, databases, and spreadsheets. Labwork is barcode-driven.

Data available

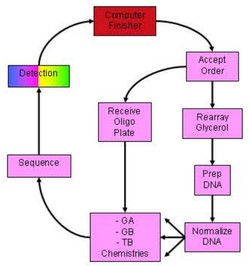
Data are automatically pipelined to Computer Finishers who assemble the sequences and order additional work if necessary.

Data may be viewed in many ways, including by plate and down to an individual trace.

Fosmid/Plasmid Custom Primer Walking

Using custom oligonucleotides, a specified read is extended using the Special Chemistries. Primer walking is ideal for sequence gaps of <1kb in size or for resolving small regions of difficult sequence.

Primer Walk Overview



Glycerol clone stocks are pulled from -80ºC storage

Clones are re-arrayed on a MultiPROBE to match oligo plates

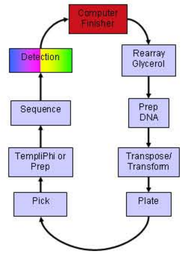
Three sequence chemistries may be used

TB	GB	GA
-5x v3.1 buffer -v3.1 BigDye	-5x v3.1 buffer -v3.1 BigDye -dGTP BigDye	-5x v3.0 buffer -dGTP BigDye -Additive A (*1:1 L-Proline: 4-Methylmorpholine N-Oxide*)
TB is a standard BigDye reaction. This is the workhorse and adequate for the majority of sequences.	GB is a Special Chemistry. dGTP BigDye utilizes a dGTP nucleotide in place of the standard dITP to better handle SSRs, GC-rich, homopolymeric runs, and secondary structures.	GA is a Special Chemistry and a mix of dGTP BigDye and Additive A. This chemistry works well through GC-rich regions.

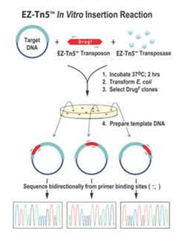
Fosmid/Plasmid Clone Transposition

Transposons offer the possibility of a directed approach to sequence gap closure, where a target DNA up to about 40kb in length can be sequenced with minimal redundancy. Transposons are mobile DNA elements that can be inserted in a reasonably random fashion into the target DNA. A 4kb plasmid insert will generate 64 reads and a 40kb Fosmid insert will generate 1344 reads.

Transposition Overview



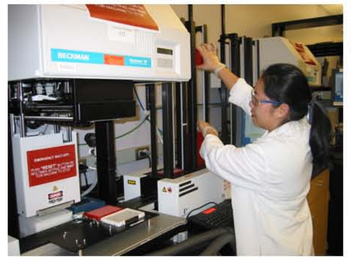
Clones are rearrayed as above and DNA is transposed

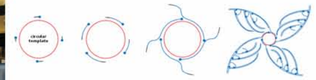


From Epicentre Biotechnologies (http://www.epibio.com)

Clones are transformed and robotically picked (QPix2)

Plasmid DNA is amplified on a PE MultiMek using TempliPhi and then sequenced directly





TempliPhi is a rolling circle amplification method that allows a large amount of template to be generated very easily

From GE Healthcare (http://www6.gelifesciences.com)

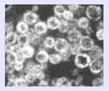
Mixed Projects

We have the additional flexibility of being able to work on multiple project types, e.g. working from whole genomes or assembling tiling paths from any number of clones and clone types.

Human
tiling path

Mouse
mixed

Coccidioides immitis
whole genome



Dog
mixed

Magnaporthe grisea
whole genome

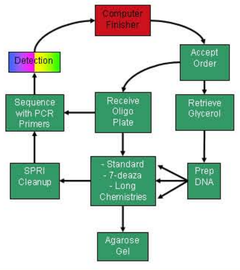
Mycobacterium tuberculosis
whole genome

Tick
mixed

Custom PCR

Difficult-to-clone regions maybe captured by a PCR product using BAC or genomic templates. The PCR product may be end-sequenced to close the gap if it is <1kb.

PCR Overview



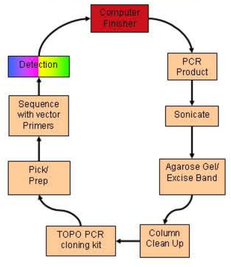
Three PCR chemistries may be used

Standard	7-deazaG	Long
•Taq •Pfu •dNTPs •reaction buffer •H_2O	•Taq •Pfu •**7-deazaG dNTPs** •reaction buffer •H_2O	•**Expand Long Template PCR System** •dNTPs •**Buffer 1** •H_2O
Useful for 0.5 – 5 kb amplifications of standard templates	Useful for 0.5 – 5 kb amplifications of GC-rich templates	Useful for 0.5 – 10 kb amplifications of standard templates

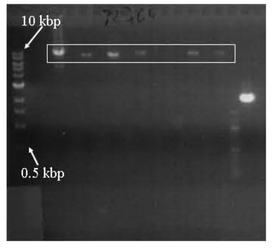
Shatter Library

Difficult-to-sequence regions may be shattered into smaller (~500 bp) fragments. This process is usually the last resort method for finishing.

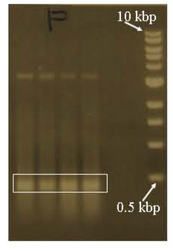
Shatter Library Overview



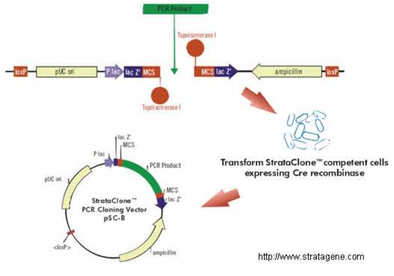
~10 kb Long PCR products



~0.5 kb Shatter products



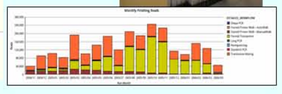
We recently implemented Stratagene's 'StrataClone™ Blunt PCR Cloning Kit' that utilizes topoisomerases linked to arms of the cloning vector. This greatly increases the ease of the process and shaves at least a day off of the total time.



http://www.stratagene.com

Flexibility

We have designed our workflows to adapt to variable order volumes. We have automation in place for times of high-throughput and personnel trained to run low volume orders by hand.



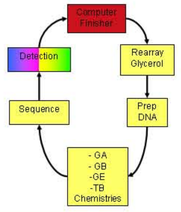
Resequencing

Clones that failed in Production Sequencing, need additional coverage, or require Special Chemistry go to the Resequencing Workflow.

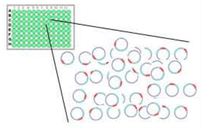
After feedback from many at the last 'Finishing in the Future' conference, we tested and implemented the GE Finishing Kit. It worked very well and GE may soon become the main resequencing chemistry.

The GE Finishing Kit is a modification of their TempliPhi Kit

Resequencing Overview

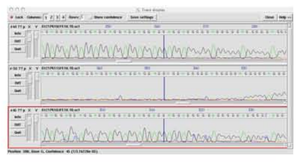


One of the key features of our Resequencing system is that we can also resequence the transposon libraries made here in the lab.

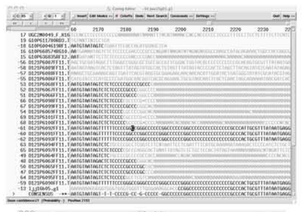


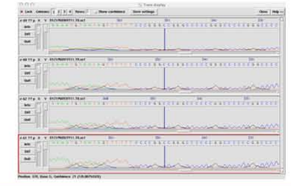
Special Sequencing Chemistries

GA rich sequence gaps: GA

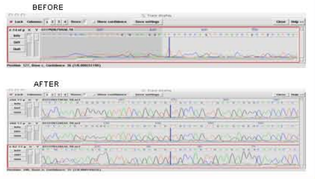


Mono-nucleotide sequence gaps: GB





Secondary Structure sequence gap: GA



The Lab

Back (l to r): Tamrat Negash, Charles Matthews/ Daniel Bessette, Michael FitzGerald *Front (l to r):* Xiaohong Liu, Anna Montmayeur, Nicole Allen, Tashi Lokyitsang, Mostafa Benamara, Rakela Lubonja, Chelsea Dunbar, Thu Nguyen