The Pyrosequencing Process

Christopher Hack, Doug Smith, Ed Kirton, Zhiying Jean Zhao, Andy Yuen, Jamie Jett, Feng Chen, and Paul Richardson
DOE Joint Genome Institute

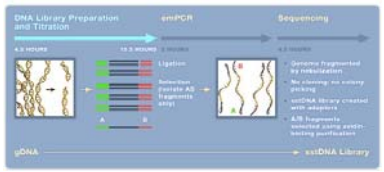
DOE JOINT GENOME INSTITUTE
US DEPARTMENT OF ENERGY
OFFICE OF SCIENCE

Introduction

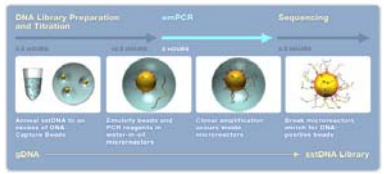
Sanger sequencing has been the predominant method for sequencing DNA since it was introduced in 1972. Recently, however, newer technologies have surfaced that have the potential to reshape the sequencing world. One such technology is pyrosequencing, as developed by 454 Life Sciences. The 454 sequencing platform is an integrated system of emulsion-based PCR amplification of hundreds of thousands of DNA fragments linked to high throughput parallel pyrosequencing in picoliter-sized wells. The system can potentially deliver over 100 million bases per run, with readlengths over 200 bases, from a 6-hour run.

Overview of the 454/Roche Pyrosequencing Process

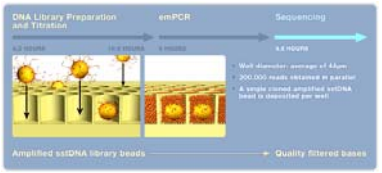
Figures below courtesy of 454 Life Sciences



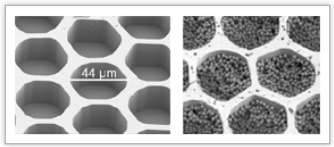
Step 1: Library Creation
- Shear genomic DNA into ~500bp fragments by nebulization
- Attach adapters A and B to the ends of the fragments
- Select only fragments that have adapters A and B attached
- Quantify DNA fragments for downstream processing

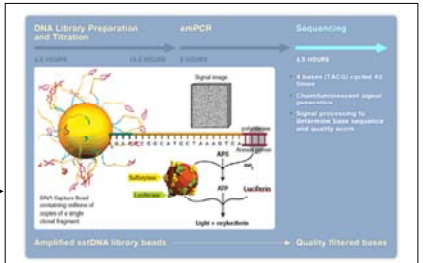


Step 2: Emulsion PCR
- Anneal DNA fragments to sequencing beads
- Emulsify in oil to create isolated aqueous droplets containing single sequencing beads with PCR amplification reagents
- Perform PCR by thermocycling
- Recover beads with amplified DNA from the emulsion
- Anneal sequencing primer to the amplified DNA



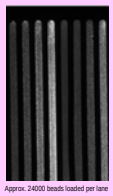
Step 3: Insert sequencing beads into PicoTiter Plate
- PicoTiter Plate wells are 44 μM in diameter
- Sequencing beads are 25-36 μM in diameter
- Maximum of 1 sequencing bead is deposited per well
- Sequencing beads are surrounded and held in place by smaller enzyme beads.
- PicoTiter Plate can hold up to 1.8 million sequencing beads





Step 4: Sequencing of DNA fragments in parallel
- Reagents for each base are passed sequentially over PicoTiter Plate
- As bases are added to the complementary strand, the luciferase reaction emits light
- An image is taken at each step by a CCD camera
- Sequence and quality data is determined from the light intensity observed at each step

Data Output from 454 Sequencing Run



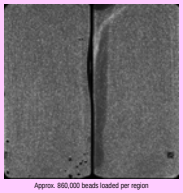


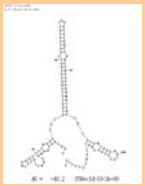
Figure A (left): Image taken during a 8-lane (titration) sequencing run. Figure B (middle): Image taken during a 2-region (bulk) sequencing run. Each pinpoint of light in figures A and B represents a bead where the addition of a base to the DNA sequence is causing the light-emitting luciferase reaction.

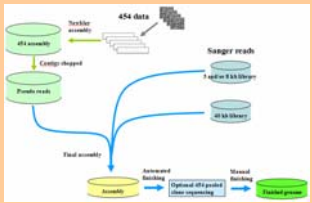
Figure C (right): Flowgram of a single well in a PicoTiter Plate over the course of an entire run. Each peak represents the intensity of light emitted at each step of the sequencing run, normalized to the light intensity emitted by the addition of the 4-base control sequence tag at the beginning of the run. From the light signals emitted, the sequence of the DNA fragment on the sequencing bead in the given well is determined.

Strategy for Whole Genome Shotgun Sequencing for Microbial Genomes

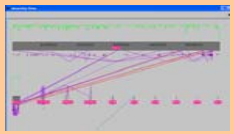
Problem: Sanger sequencing has difficulty sequencing through certain regions of DNA, such as hairpin turns. This leaves gaps in the final assembly.

Solution:
1. Sequence and assemble with 454 data.
2. Fragment the 454 assembly *in silico* to 1000-base regions with 100-base overhangs.
3. Re-assemble genome using Phrap with both Sanger reads and 454 fragments.





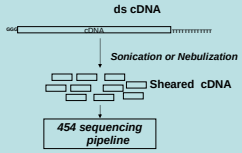
Microbial genome assembled with Sanger data only. Several gaps exist in the assembly.



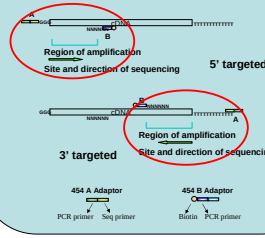
The same genome assembled with both Sanger and 454 data. Gaps that appeared using only Sanger data are closed.

Strategy for Sequencing cDNA with 454

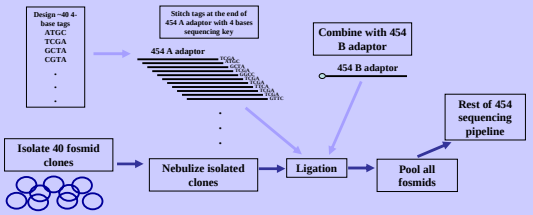
- If sequence over the whole cDNA region is desired, the approach is the same as for whole genome shotgun sequencing.



- If sequence is desired only at the 5' or 3' end of the cDNA, incorporate 454 adapter sequence into PCR primers to amplify region of interest, then use 454 to sequence the amplified PCR products.



Strategy for Sequencing 40-kb Fosmid Libraries using Sequence-tagged 454 Adapters



For each 454 read from the tagged fosmid shotgun library, the tag sequence (first four bases) was separated from the remaining sequence and the total error probability of the barcode was calculated. If the barcode sequence had an expected error rate of less than 5%, the unique read sequence was placed into the appropriate bin, otherwise the read was discarded. The reads in each bin belong to a particular fosmid and were assembled using Phrap assembler separately.

454 Pyrosequencing Accomplishments at the JGI

- 65 Microbes Sequenced
- 4.87 billion bases sequenced
- Average of 33 million bases of data per bulk sequencing run

Currently in Development

- A vacuum and centrifuge-based system to recover sequencing beads from emulsion PCR, replacing ergonomically hazardous 2-3 hour syringe-based recovery system.
- Validation of 454 GS FLX instrument to increase average readlengths to >200.
- Development of LIMS to manage process and data tracking.
- Integration of 454 pyrosequencing technology into the JGI production line.

LBNL-62591 Poster

This work was performed under the auspices of the US Department of Energy's Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Livermore National Laboratory under Contract No. W-7405-Eng-48, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231 and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.