Nuts and bolts of phage genome sequencing

*the 5’5” and 5’8” perspective…*

Allison Johnson & Anneke Padolina
Our role in DNA sequencing

“Rachel, do you want to get on the bus and take your sample to the sequencing core?”
DNA structure

Antiparallel double helix

A-T and G-C base pairs
Phosphate-ribose backbone
5’-3’ directionality
Definitions

Gene: A discrete unit of hereditary information

Genome: An organism’s complete set of DNA (or RNA for some viruses)
- genes + non-coding sequence

Genomics: The study of the complete genetic material of an organism
(an organism can have more than one genome)
Major landmarks in DNA sequencing

1953: Discovery of the structure of the DNA double helix.

1977: Allan Maxam and Walter Gilbert publish DNA Sequencing by chemical degradation. Fred Sanger, independently, publishes DNA sequencing by enzymatic synthesis.
**we will use this method for phage genome finishing**


2005: Next-generation sequencing technology released with potential to produce 100 million to a few billion base pairs at a few days at a cost of a few thousand dollars.
**we used this method for phage genome sequencing**
How does Sanger sequencing work?

Prepare four reaction mixtures; include in each a different replication-stopping nucleotide

Separate products by gel electrophoresis

Read sequence as complement of bands containing labeled strands

http://www.scq.ubc.ca/genome-projects-uncovering-the-blueprints-of-biology/
How does Sanger sequencing work?
454 Pyrosequencing

High throughput
Massively parallel
Sequencing by synthesis

454 Life Sciences GS Flex (Roche) – pyrosequencing
100 million bases per run (~$4000, 6h)
250-400 bases per read
generates image data
image processing = light intensity → numeric encoding of DNA
Sample preparation

DNA is sheared, polished, and denatured

Adaptors A and B and barcode are ligated onto each end

Single DNA strands are attached to DNA capture beads, inside an emulsion bubble

emPCR- PCR inside bubble = amplification in parallel, producing millions of copies of a read
“Massively parallel DNA sequencing” using pyrosequencing (AKA “454”)

hundreds of thousands of beads each carrying millions of copies of a unique single-stranded DNA molecule are sequenced in parallel.

Sequencing by synthesis
“Massively parallel DNA sequencing” using pyrosequencing (AKA “454”)

1. Prepare DNA sample- break into pieces, polish to blunt ends.
2. Denature and add to a PCR mix with one primer immobilized on a nanobead and the other primer soluble.
3. Mix with excess oil to form a water-oil emulsion. All the reactants stay in the aqueous microspheres that form. Template DNA is sufficiently dilute such that each microsphere contains 0 or 1 DNA template.
4. Run the PCR. Anti-template strands will accumulate on the bead as extensions to the immobilized primer.
5. Recover the beads in aqueous solution and let them settle in nanowells.
6. Perform pyrosequencing on the well array, noting light flash intensities with a scanning sensor or CCD camera.

Video...

A 20 minute video is here: http://www.roche-applied-science.com/publications/multimedia/genome_sequencer/flx_multimedia/wbt.htm
1.6 million wells on the PicoTiterPlate™
Signal intensity is proportional to number of nucleotides incorporated.
So what do the reads actually look like?

• Fasta format

> unique header
ATGCGCTTAGCTAG

• 3 GB, one “line” for each read
Genomes assembled by Newbler, viewed in Consed

More about Consed on Friday with Chris & Dan
Computational assembly of DNA sequence

Sequence assembly software was the great contribution of the human genome project

Fig 2: Short fragments of DNA sequence are ordered by overlapping data to recreate the whole genome sequence
Zoom in on a bead!

Each phage is sequenced using a unique barcode
Sequencing a Genome

Shotgun DNA sequencing

Genome

Break into small pieces

Sequence the small pieces, both ways

Assemble into one, based on overlaps
Computational assembly of DNA sequence

Sequence assembly software was the great contribution of the human genome project

Fig 2: Short fragments of DNA sequence are ordered by overlapping data to recreate the whole genome sequence
Shotgun sequencing

Hierarchical vs. Whole genome

Gov’t sponsored project

Celera project

&

how your phage genomes were assembled