Next generation DNA sequencing

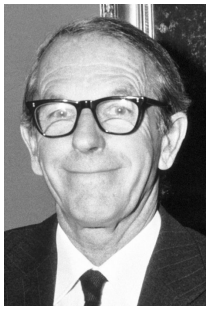
BIOL 426/626
Approaches to Molecular Biology

Class 13: Next generation sequencing

- Learning Goal
  - To understand the "next generation" sequencing methods used to determine genome sequences and the methods for analyzing the information content of those sequences.
- Learning Objectives
  - Explain how next generation sequencing methods work
  - Compare and contrast those methods of analysis
  - Explain the basics of genome-level sequence analysis
- Reading assignment:
  - Dale From Genes to Genomes: Chapter 8

First generation sequencing

Walter Gilbert
Harvard University
Cambridge, MA

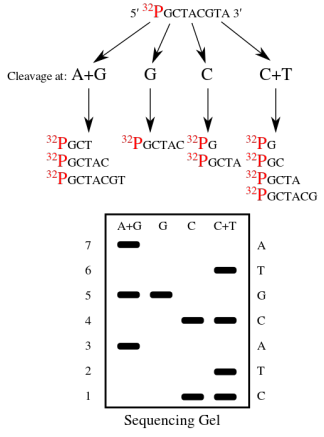


Frederick Sanger
MRC Lab of Molecular Biology
Cambridge, UK

1980 Nobel Prize in Chemistry

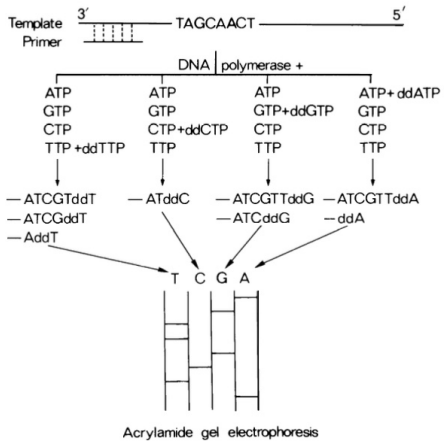
Two types of DNA sequencing

Maxam-Gilbert



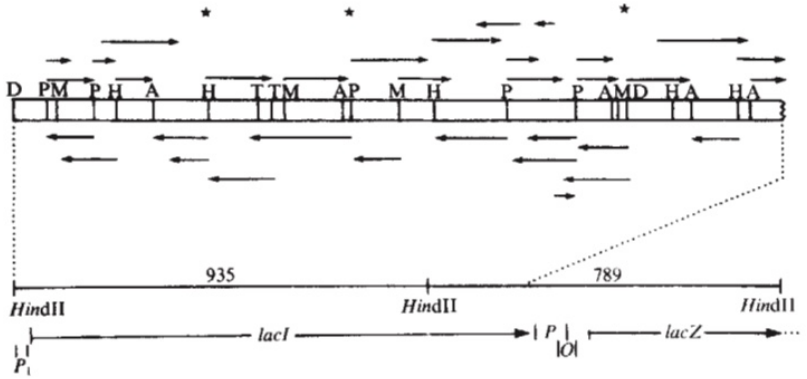
Sequencing Gel

Sanger (Coulson)



Acrylamide gel electrophoresis

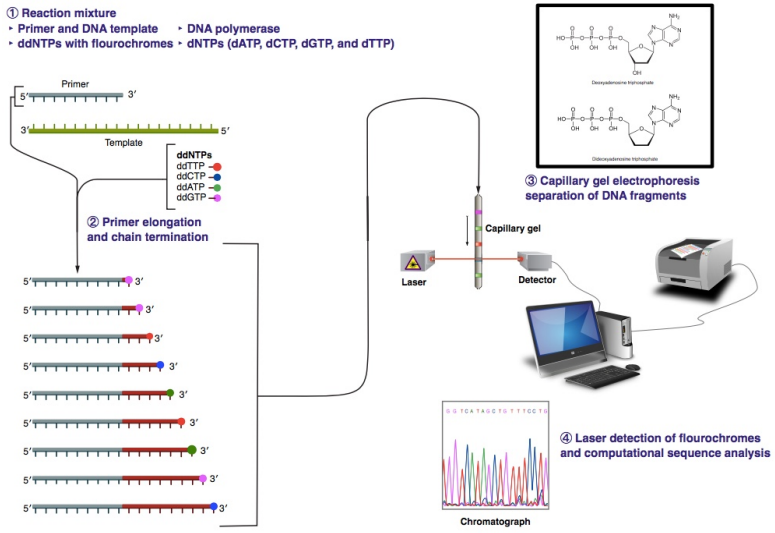
Sequencing strategy for the lacI gene



A problem sequencing genomes…

lacI	Human genome
1080 bp	> 3 billion bp
~ 1 year	~ 3 M years?

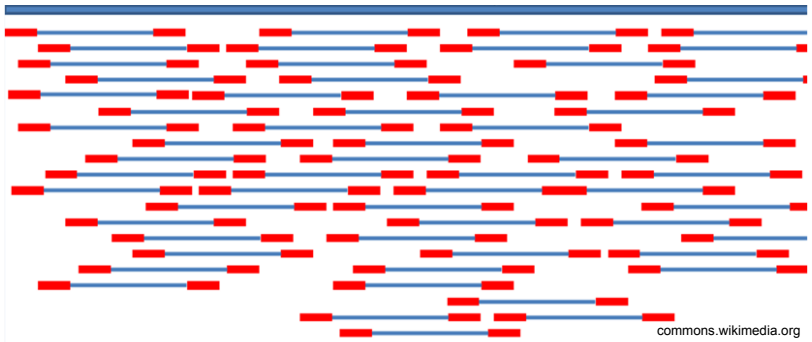
Sanger sequencing can be automated



Problems with genome sequencing using first generation sequencing

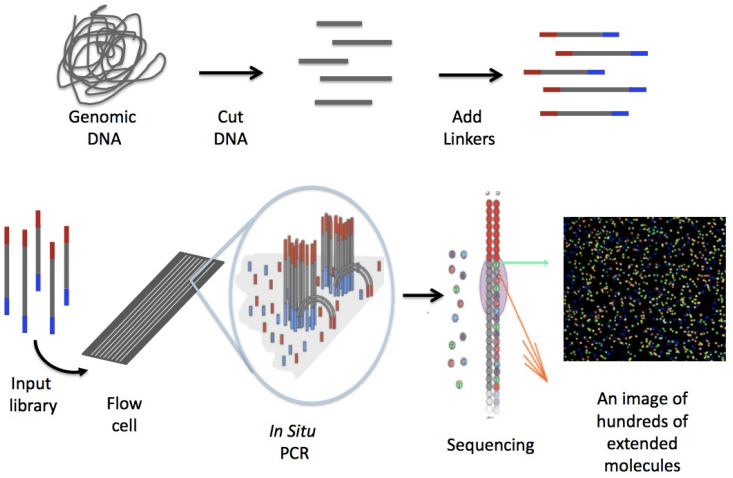
- Slow
  - Preparation
  - Especially the time to separate the fragments (increasing with increasing distance from the primer)
- Expensive
  - Requires massive numbers of sequencing machines
  - Costly reagents
  - Labor intensive
- Time/work intensive
  - Preparation of DNA
  - Generation of bacterial artificial chromosomes (BACs)
  - Problems with filling in final gaps

Paired-end sequencing strategy



Next-generation or second-generation sequencing

- Also known as "massively parallel sequencing"



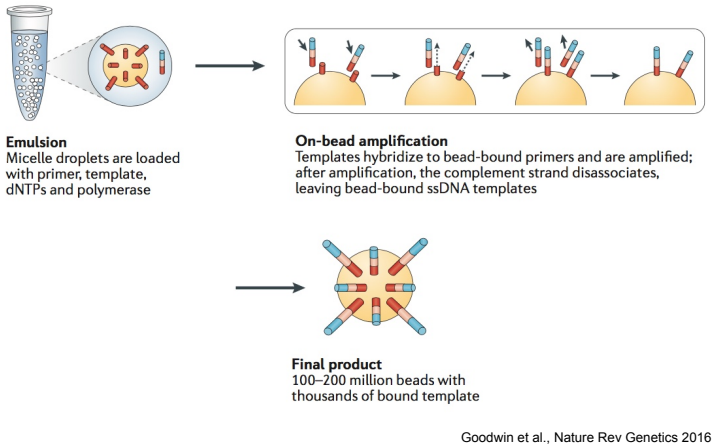
bloodjournal.org

"Higher" generation sequencing procedures

- Pyrosequencing
  - "Parallelized" pyrosequencing: 454 sequencing
- SOLiD sequencing
- Ilumina sequencing
- Ion Torrent semiconductor sequencing
- Nanopore DNA sequencing (third generation method)

Template amplification: emulsion PCR

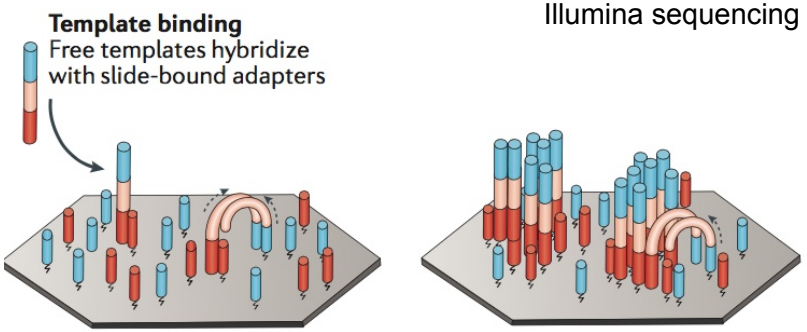
454, SOLiD & Ion Torrent sequencing



Goodwin et al., Nature Rev Genetics 2016 17:33

Template amplification: solid phase bridge amplification

Illumina sequencing



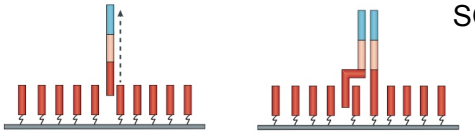
Bridge amplification
Distal ends of hybridized templates interact with nearby primers where amplification can take place

Cluster generation
After several rounds of amplification, 100–200 million clonal clusters are formed

Goodwin et al., Nature Rev Genetics 2016 17:33

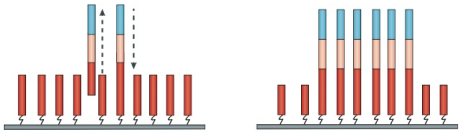
Template amplification: solid phase template walking

SOLiD Wildfire sequencing



Template binding
Free DNA templates hybridize to bound primers and the second strand is amplified

Primer walking
dsDNA is partially denatured, allowing the free end to hybridize to a nearby primer

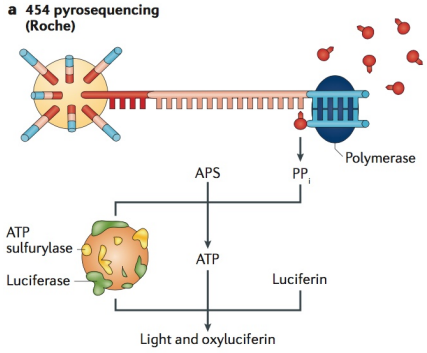


Template regeneration
Bound template is amplified to regenerate free DNA templates

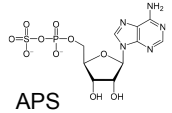
Cluster generation
After several cycles of amplification, clusters on a patterned flow cell are generated

Goodwin et al., Nature Rev Genetics 2016 17:33

Mechanism of 454/Pyrosequencing

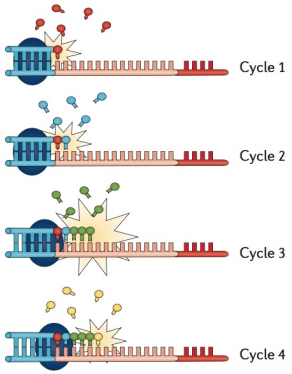


Pyrosequencing
As a base is incorporated, the release of an inorganic pyrophosphate triggers an enzyme cascade, resulting in light



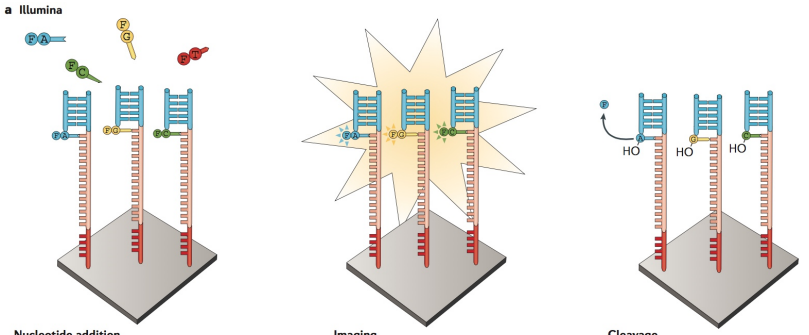
APS

Single nucleotide addition
Only one dNTP species is present during each cycle; multiple identical dNTPs can be incorporated during a cycle, increasing emitted light



Goodwin et al., Nature Rev Genetics 2016 17:33

Mechanism of Illumina sequencing



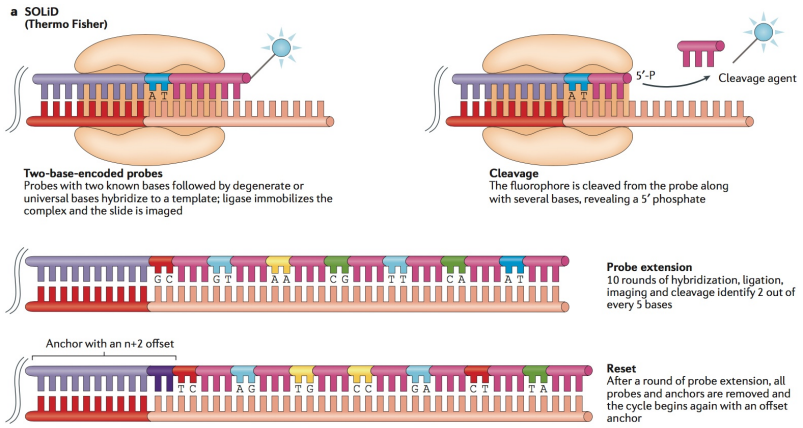
Nucleotide addition
Fluorophore-labelled, terminally blocked nucleotides hybridize to complementary base. Each cluster on a slide can incorporate a different base.

Imaging
Slides are imaged with either two or four laser channels. Each cluster emits a colour corresponding to the base incorporated during this cycle.

Cleavage
Fluorophores are cleaved and washed from flow cells and the 3′-OH group is regenerated. A new cycle begins with the addition of new nucleotides.

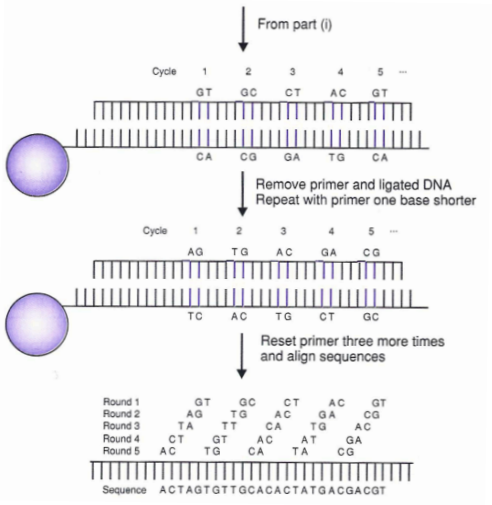
Goodwin et al., Nature Rev Genetics 2016 17:33

Mechanism of SOLiD sequencing

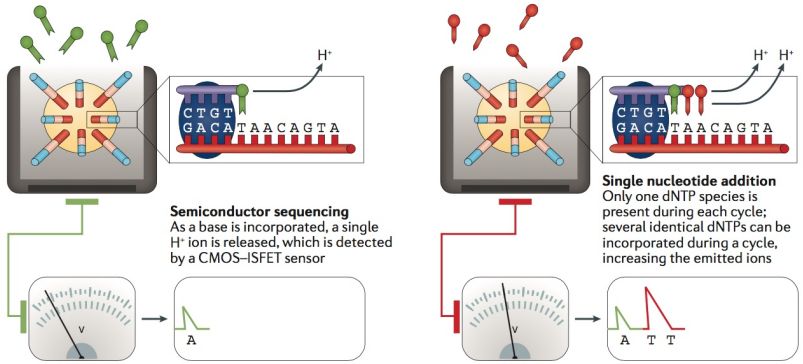


Goodwin et al., Nature Rev Genetics 2016 17:33

SOLiD sequencing: repeated synthesis

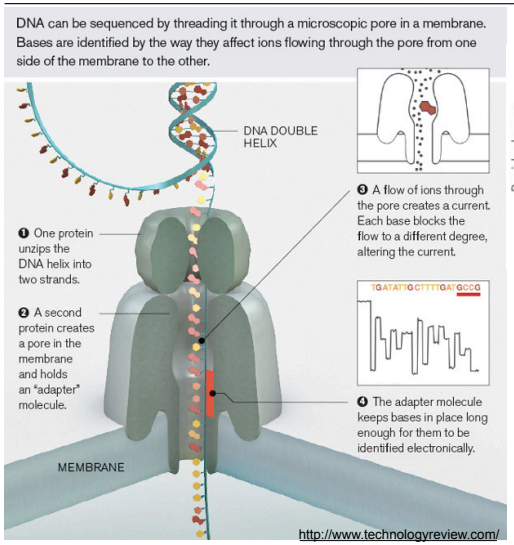


Mechanism of Ion Torrent sequencing

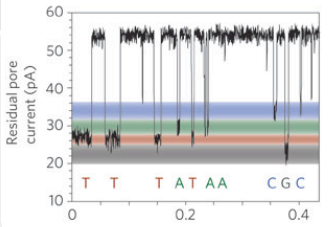


Goodwin et al., Nature Rev Genetics 2016 17:33

Mechanism of Oxford Nanopore sequencing



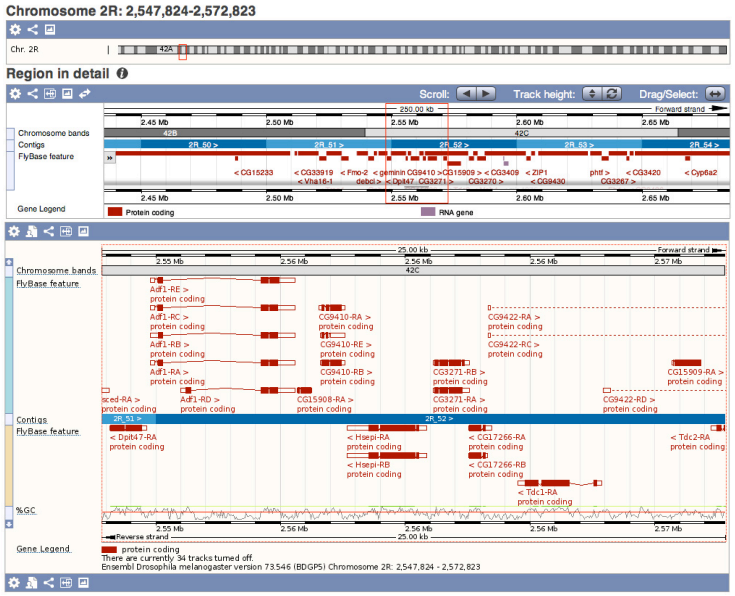
http://www.technologyreview.com/



Genome sequencing by methods

Method	Read length	Reads	Error (%)	Cost/Gb
SOLiD	50	700 M	< 0.1%	$130
Illumina	75-100	20-50 M	< 1%	$30-250
454	400-700	0.1 - 1 M	1%	$10-40K
Ion	200-400	0.5 - 5 M	1%	$700-1000
Nanopore	200 Kb	>100 K	12%	$750

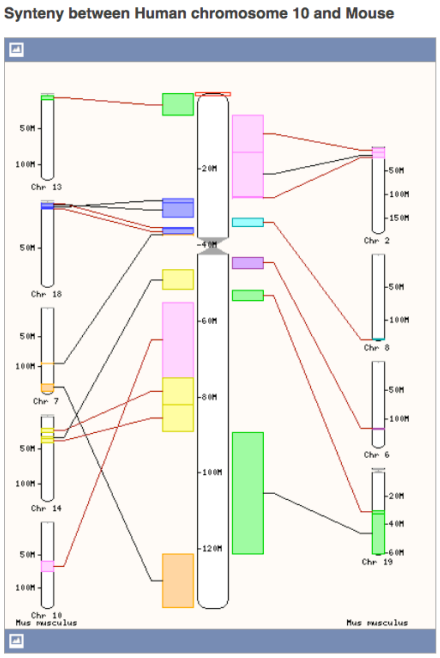
Genome browsing...



Synteny viewer...

Synteny: the conservation of blocks of order within two sets of chromosomes that are being compared with each other.

Notice how these blocks on one human chromosome are spread across nine mouse chromosomes

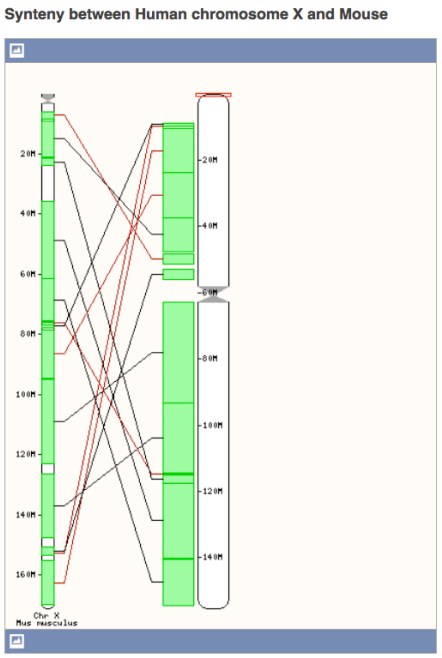


Synteny viewer...

Compare the Chr 10 synteny to synteny between the human and mouse X chromosome.

Although the order of blocks has scrambled, the genes all still reside on the X.

Why is that?



Reading for next time:

- Glazov et al. (2010) Whole-exome re-sequencing in a family quartet identifies *POP1* mutations as the cause of a novel skeletal dysplasia. PLOS Genetics 7:e1002027.