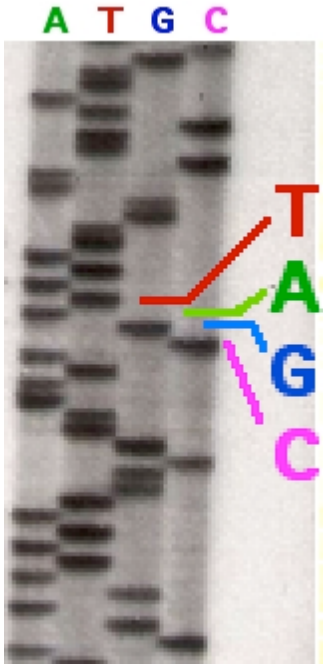
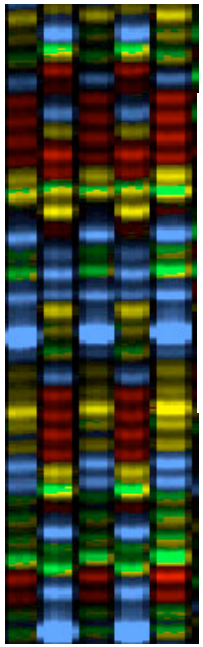
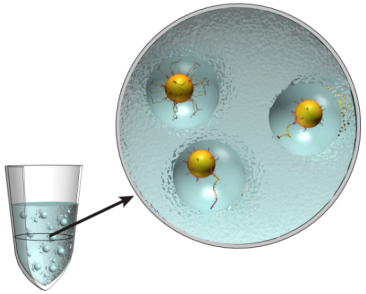
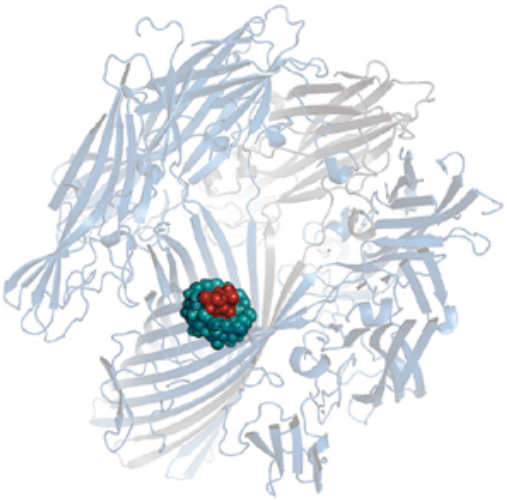
The Evolution of Sequencing

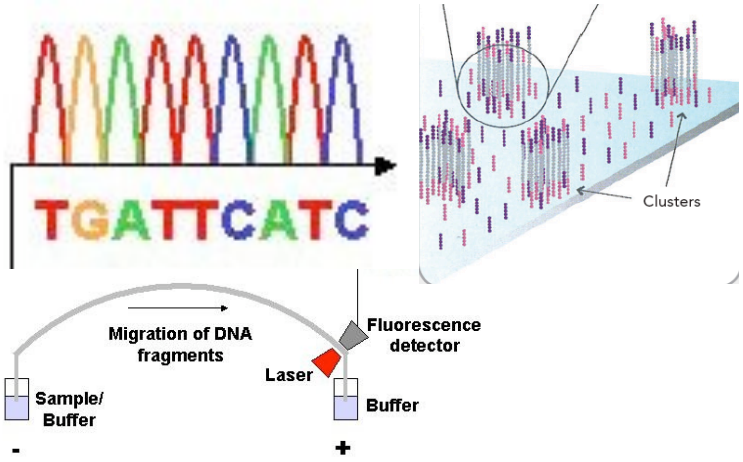
- Sanger sequencing
  - Gels
  - Cycle sequencing
  - Fluorescence
  - Capillary electrophoresis
- Sequencing, the “Next Generation”
  - “Sequencing by synthesis”
    - Pyrosequencing (Roche/454)
    - Cluster sequencing (Illumina/Solexa)
  - Sequencing by ligation (ABI/SOLiD)
- Next “Next Generation”
  - 4th Generation?
  - Single molecule sequencing

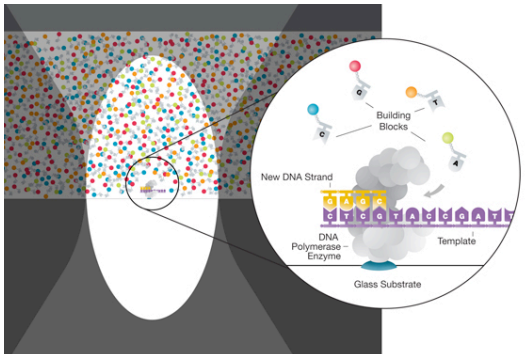

A T G C
T
A
G
C

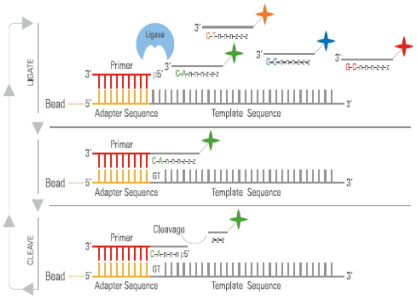


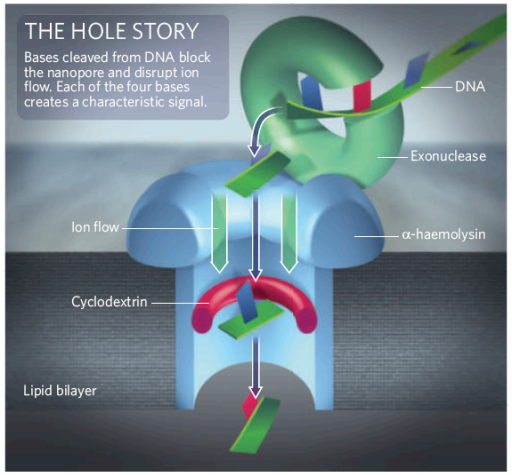




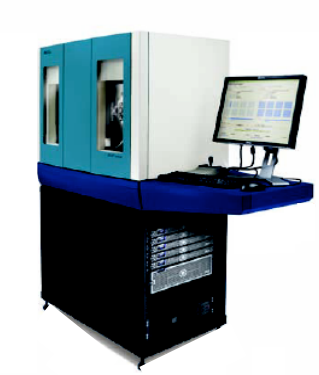

TGATTCATC
Clusters
Migration of DNA fragments
Fluorescence detector
Laser
Sample/ Buffer
Buffer
-
+


Building Blocks
New DNA Strand
DNA Polymerase Enzyme
Template
Glass Substrate


Ligase
Primer
Bead
Adapter Sequence
Template Sequence
LIGATE
CLEAVE
Cleavage


THE HOLE STORY
Bases cleaved from DNA block the nanopore and disrupt ion flow. Each of the four bases creates a characteristic signal.
DNA
Exonuclease
Ion flow
α-haemolysin
Cyclodextrin
Lipid bilayer

Macro versus Micro Reads



Applied Biosystems
SOLiD

Read Length
35 - 75bp **<=>** 250 - 450bp

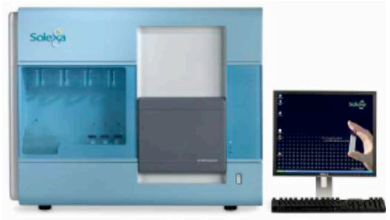
Base Pairs Per Run
3 - 10 Gb **<=>** 0.1 - 0.5 Gb

Base Pairs Per Day
1 - 1.5 Gb **<=>** 0.2 - 1.0 Gb

Number of Sequences
100 M **<=>** 1.2 M

Roche / 454 FLX

Run Time
3 - 7 days **<=>** 0.5 days



Illumina / Solexa
Genetic Analyzer

Reagent Cost per Run
~$4K - $12K **<=>** $6K

Error Rate
Varies, different characteristics

Technology and Informatics

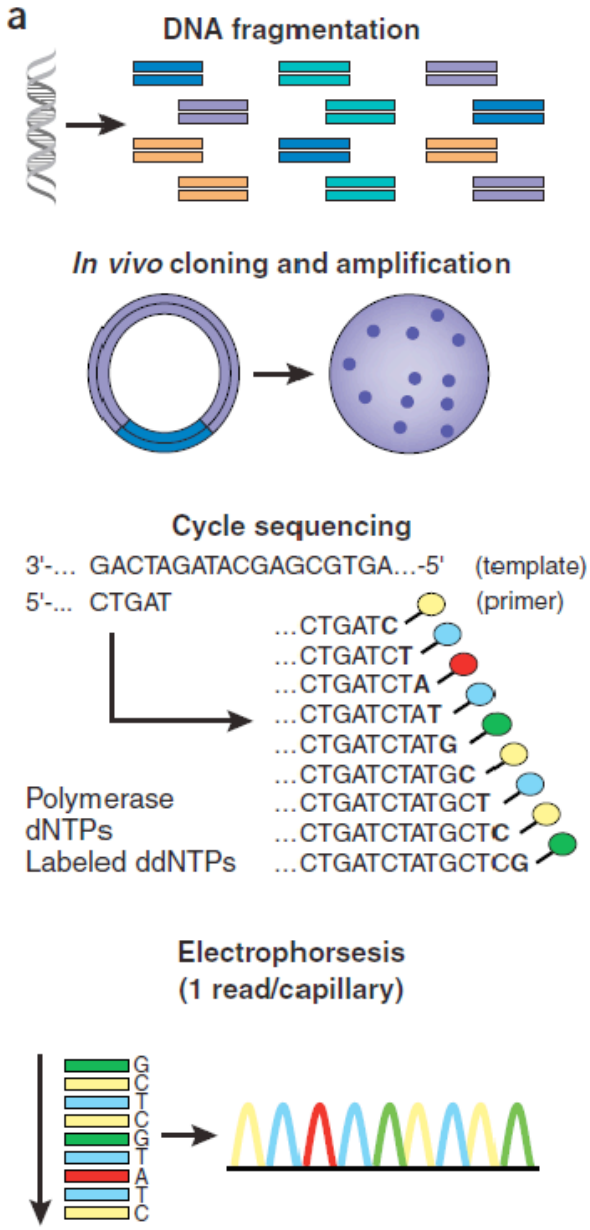
PR Space versus Science Space

- Flow and phasing
- Data quality and Error rate
  - Variation along sequence
  - Quality scores (Equivalence?)
- Length distribution versus average
- Raw versus recovered sequence
  - How much coverage with different methods?
- Tagging (barcodes) and multiplexing
  - Variation in coverage

Next-Gen Basics

- Library creation
  - Shearing, size selection
    - Size distribution
  - Specific primer sequences (adaptors) flank target sequence
    - Allows amplification
      - Opportunity for extra “mutation”
- Tagging (barcodes)
  - Proportion of sequence wasted
- Ligation or amplification (454)
- Paired ends

Sanger Sequencing



- DNA is fragmented
- **Cloned** to a vector
  - Plasmid, BAC
  - Linkage
- Cyclic sequencing
- Separation by electrophoresis
- Read fluorescent tags