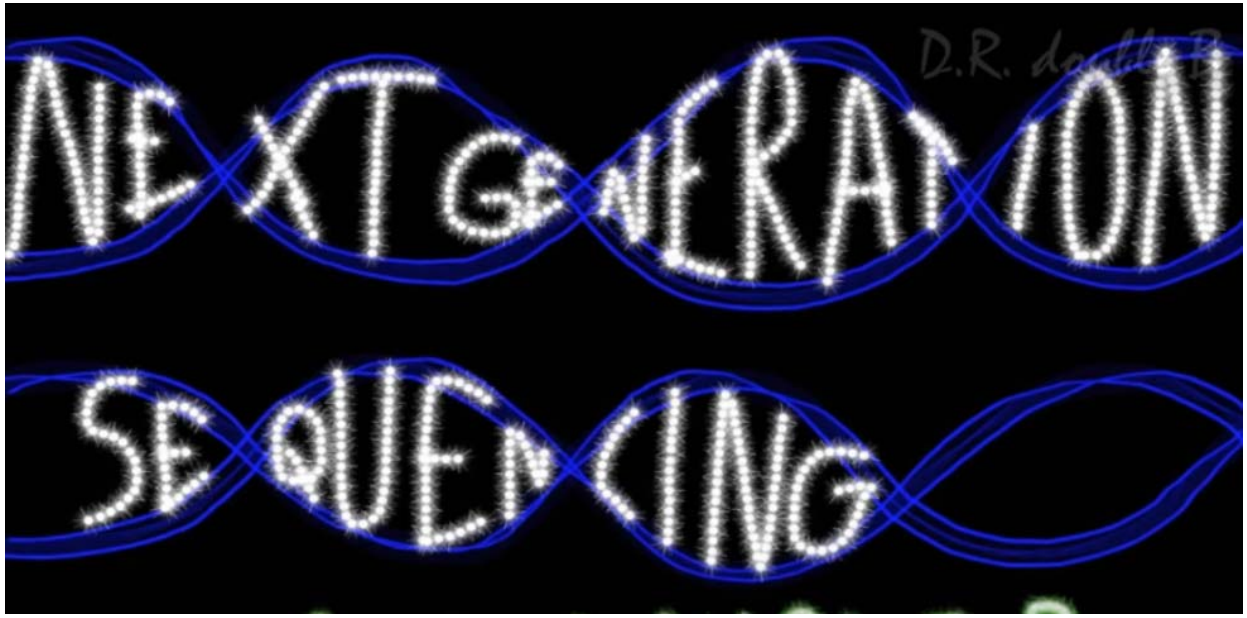
CS123A Bioinformatics Module 5 – Week 15 – Presentation 1

Leonard Wesley

Computer Science Dept

San Jose State Univ



Agenda

- Next Generation Sequencing

NGS: A Working Definition

- **Next Generation Sequencing is the set of technologies and concepts that help determine the nucleotide order (i.e., sequence) of nucleic acids. NGS technologies have resulted in sequencing costs to drop precipitously over the past 4 years (with no end in sight)**

- **NGS is a "technological singularity" happening before us**
  - It is disruptive and hard to predict
  - It has already changed the world we live in
  - It will continue to change our lives, often in unpredictable ways
  - It has fundamentally changed both approaches to and the feasibility of even attempting certain scientific and technological problems

- **NGS will require a corresponding bioinformatics singularity to reach its full potential. This may already be underway…**

- What will YOU be doing in 5 years? It may well be related to NGS!

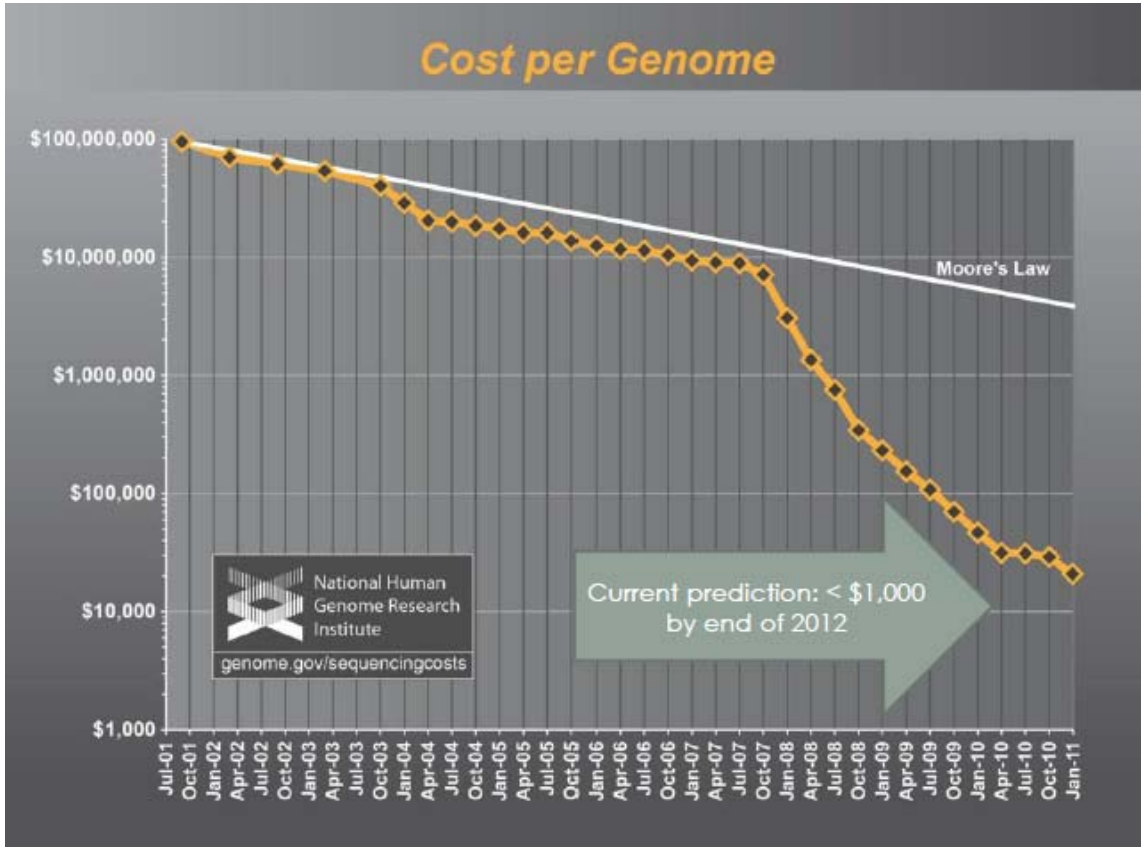
The Promise of NGS

- Personalized medicine
  - Cancer treatment
  - Discovery of causes of rare diseases (& maybe curing them!)
  - Prediction of future health issues, with recommendations for ameliorating them
  - Personalized pharmacogenomics: which drug will work best to treat a given disease in a given individual
- Characterizing entire microbial ecologies
- Molecular archaeology and anthropology
  - Discovering how humans became human & phylogeny relationships to other living organisms..
  - Tracing human population movements
- Plant & animal breeding
- Basic research (e.g. learning what "junk" DNA *really does)*
- ….

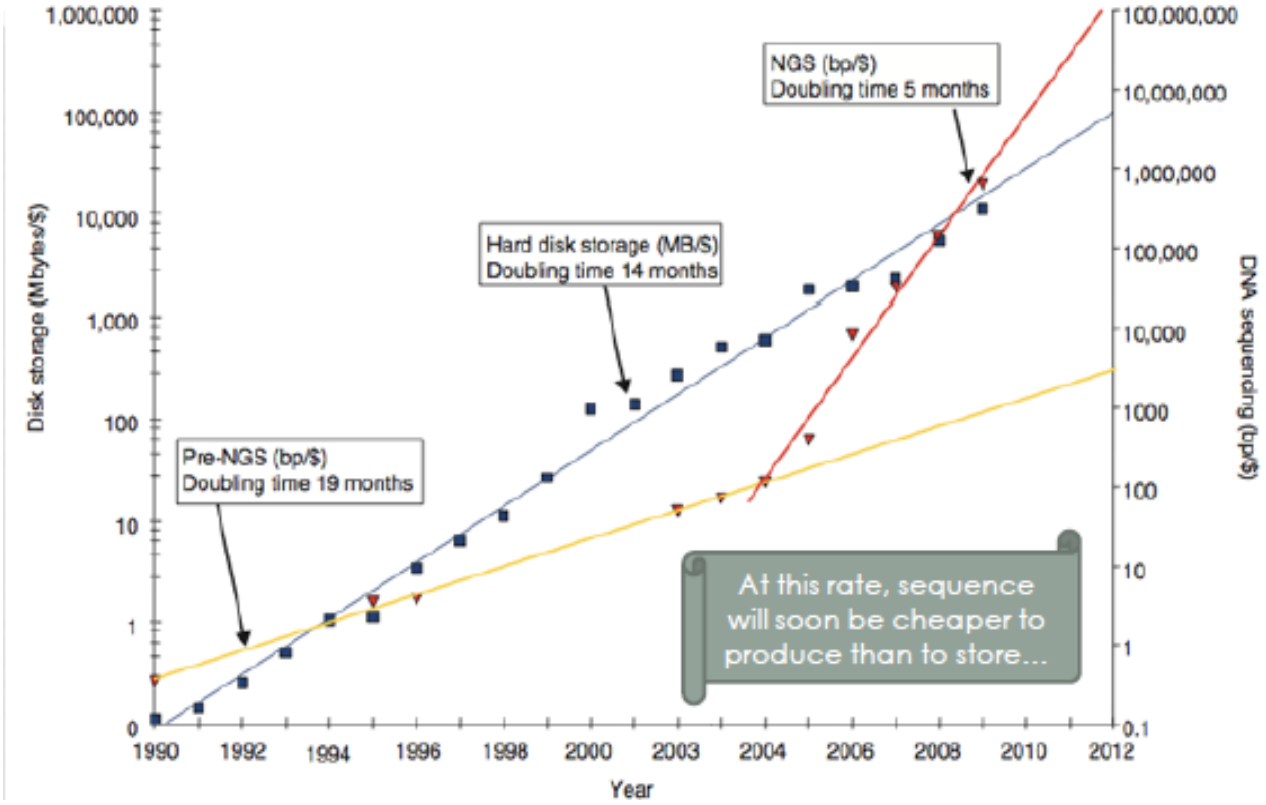
A Little History

- Between late 2007 and the present, DNA sequencing has experienced a “technological singularity”
  - In Sep-2007, the $1,000 human genome sequence was expected to arrive around Sep-2023
  - Instead, it is arriving NOW! Cost now $300 - $1K.
- Numerous business plans, research approaches and sequencing alternatives that made sense in 2007 have rapidly been relegated to the trash bin of history
- The technology is moving so rapidly that many of the “cutting edge” technologies mentioned in the course text (2015/2016) have already been discarded or replaced…

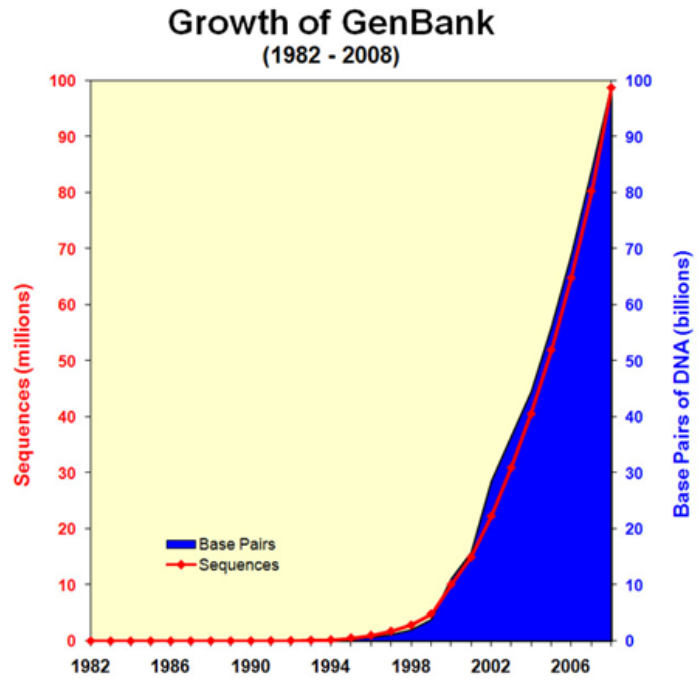
NGS Is Becoming Cheaper … FAST!



A Different View …



Genetic Data Is Accumulating At A Phenomenal Rate

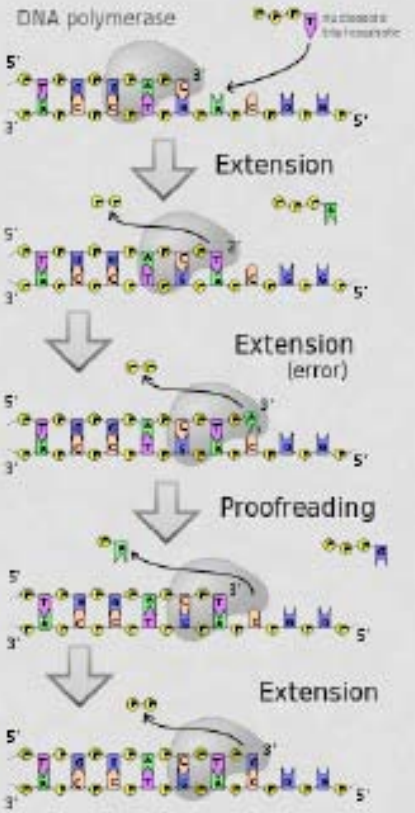


GenBank Data		
Year	Base Pairs	Sequences
1982	680,338	606
1983	2,274,029	2,427
1984	3,368,765	4,175
1985	5,204,420	5,700
1986	9,615,371	9,978
1987	15,514,776	14,584
1988	23,800,000	20,579
1989	34,762,585	28,791
1990	49,179,285	39,533
1991	71,947,426	55,627
1992	101,008,486	78,608
1993	157,152,442	143,492
1994	217,102,462	215,273
1995	384,939,485	555,694
1996	651,972,984	1,021,211
1997	1,160,300,687	1,765,847
1998	2,008,761,784	2,837,897
1999	3,841,163,011	4,864,570
2000	11,101,066,288	10,106,023
2001	15,849,921,438	14,976,310
2002	28,507,990,166	22,318,883
2003	36,553,368,485	30,968,418
2004	44,575,745,176	40,604,319
2005	56,037,734,462	52,016,762
2006	69,019,290,705	64,893,747
2007	83,874,179,730	80,388,382
2008	99,116,431,942	98,868,465

Quick DNA Review

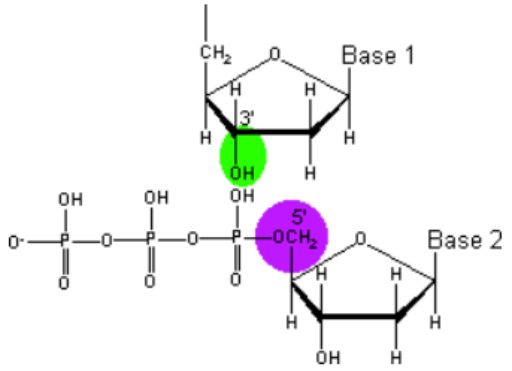
Natural Synthesis of DNA

- 5' to 3'
- By extension of an existing DNA or RNA strand (primer)
- Using a complementary (3' to 5') strand as template
- By addition of dNTP's (releasing pyrophosphate, PP_i)
- IF polymerase has 3'→5' exonuclease activity, it can "proofread" (correct errors)

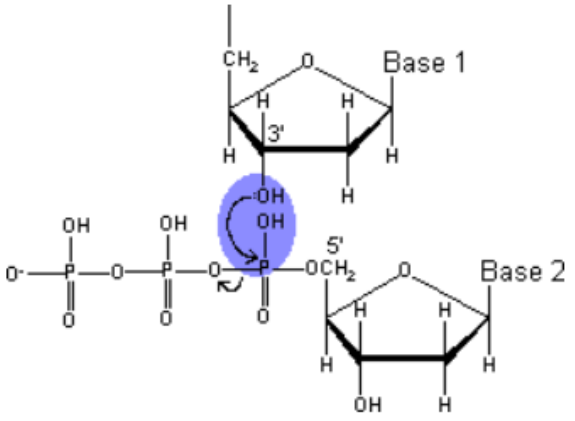


DNA Synthesis Process

- The 5’ group of a nucleotide triphosphate is held close to the 3’ hydroxyl group of a nucleotide chain.

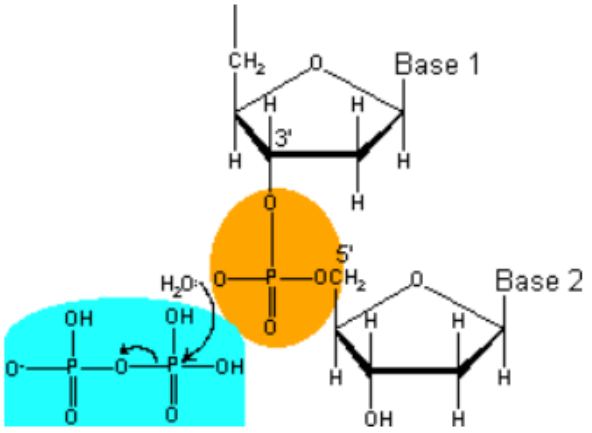


- The 3’hydroxyl group forms a bond to the phosphorous atom of the free nucleotide closest to the 5’ oxygen atom. Meanwhile the bond between the first phosphorus atom and the oxygen atom linking it to the next phosphate group breaks.

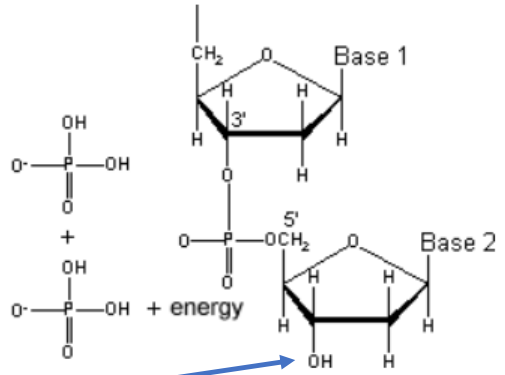


DNA Synthesis Process *(cont.)*

- A new phosphodiester bond now joins the two nucleotides. A pyrophosphate group has been liberated.

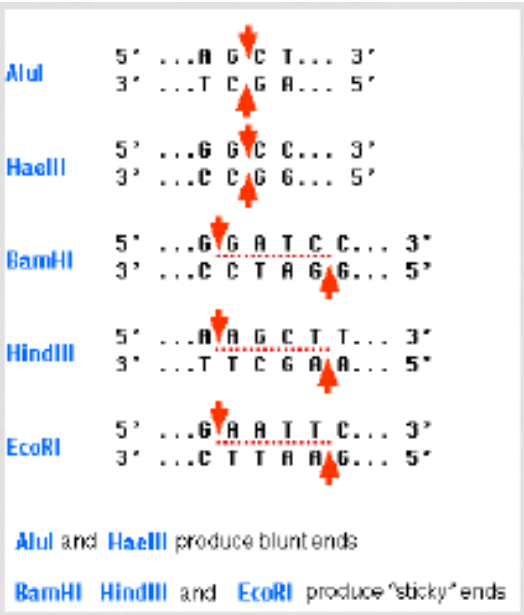


- The pyrophosphate group is hydrolyzed (split by the addition of water), releasing a lot of energy and driving the reaction forward to completion.



Remember this hydroxyl group on the 3' C. If OH is missing it is called A ddN (di-deoxy-nucleotide).

Restriction Enzymes



- Cut DNA at a specific sequence (recognition site)
- Site may be palindromic or non-palindromic
- Cut may be at site, or removed from it
- Cut may leave a 5' or 3' overhang ("sticky end") or a "blunt end"

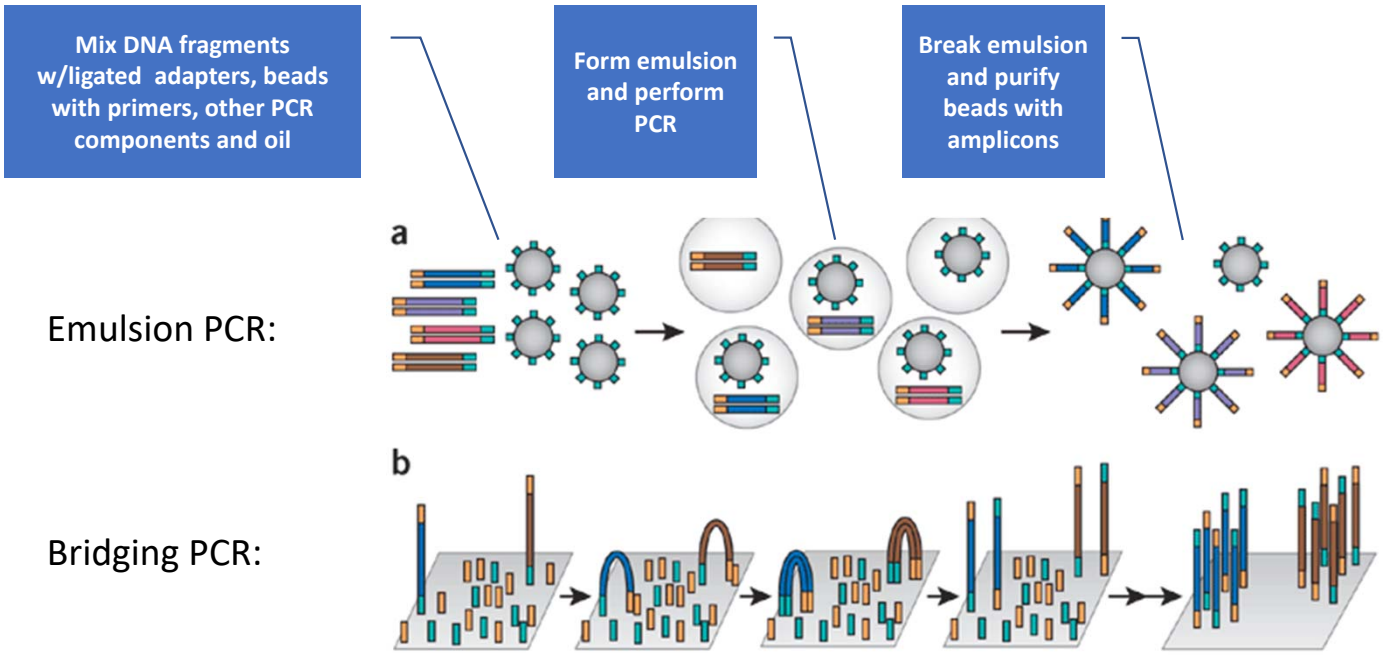
Enzyme	Sequence	Cut Site
FokI	GGATG	G G A T G N N N N N N N N N/N N N N C C T A C N N N N N N N N N N N N N/

Enzyme	Sequence	Cut Site
AcuI	CTGAAG	C T G A A G N N N N N N N N N N N N N N N/ G A C T T C N N N N N N N N N N N N N N/N N

Polymerase Chain Reaction (PCR)

PCR Video

Emulsion PCR



NGS Technologies

Sanger Sequencing

Sanger sequence video

https://www.youtube.com/watch?v=e2G5zx-OJIw&feature=em-share_video_user

Naked Replication (No Labels)

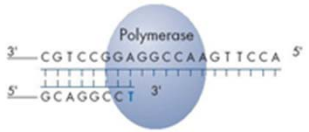
DNA Replication Complex$_n$ + dNTP

$\rightarrow$ DNA Replication Complex$_{n+1}$ + PP$_i$ + H$^+$

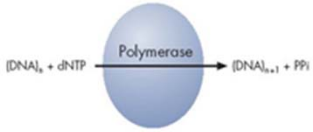
Pyrosequencing

Pyrosequencing

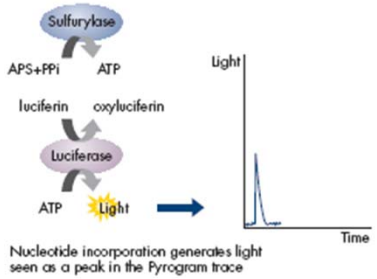
Step 1
A sequencing primer is hybridized to a single-stranded PCR amplicon that serves as a template, and incubated with the enzymes, DNA polymerase, ATP sulfurylase, luciferase, and apyrase as well as the substrates, adenosine 5' phosphosulfate (APS), and luciferin.



Step 2
The first deoxribonucleotide triphosphate (dNTP) is added to the reaction. DNA polymerase catalyzes the incorporation of the deoxyribo-nucleotide triphosphate into the DNA strand, if it is complementary to the base in the template strand. Each incorporation event is accompanied by release of pyrophosphate (PPi) in a quantity equimolar to the amount of incorporated nucleotide.

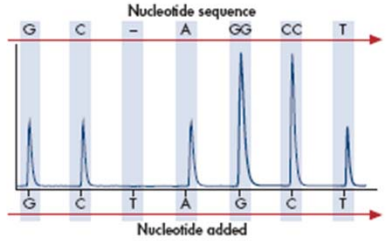


Step 3
ATP sulfurylase converts PPi to ATP in the presence of adenosine 5' phosphosulfate (APS). This ATP drives the luciferase-mediated conversion of luciferin to oxyluciferin that generates visible light in amounts that are proportional to the amount of ATP. The light produced in the luciferase-catalyzed reaction is detected by a charge coupled device (CCD) chip and seen as a peak in the raw data output (Pyrogram). The height of each peak (light signal) is proportional to the number of nucleotides incorporated.



Step 4
Apyrase, a nucleotide-degrading enzyme, continuously degrades unincorporated nucleotides and ATP. When degradation is complete, another nucleotide is added.

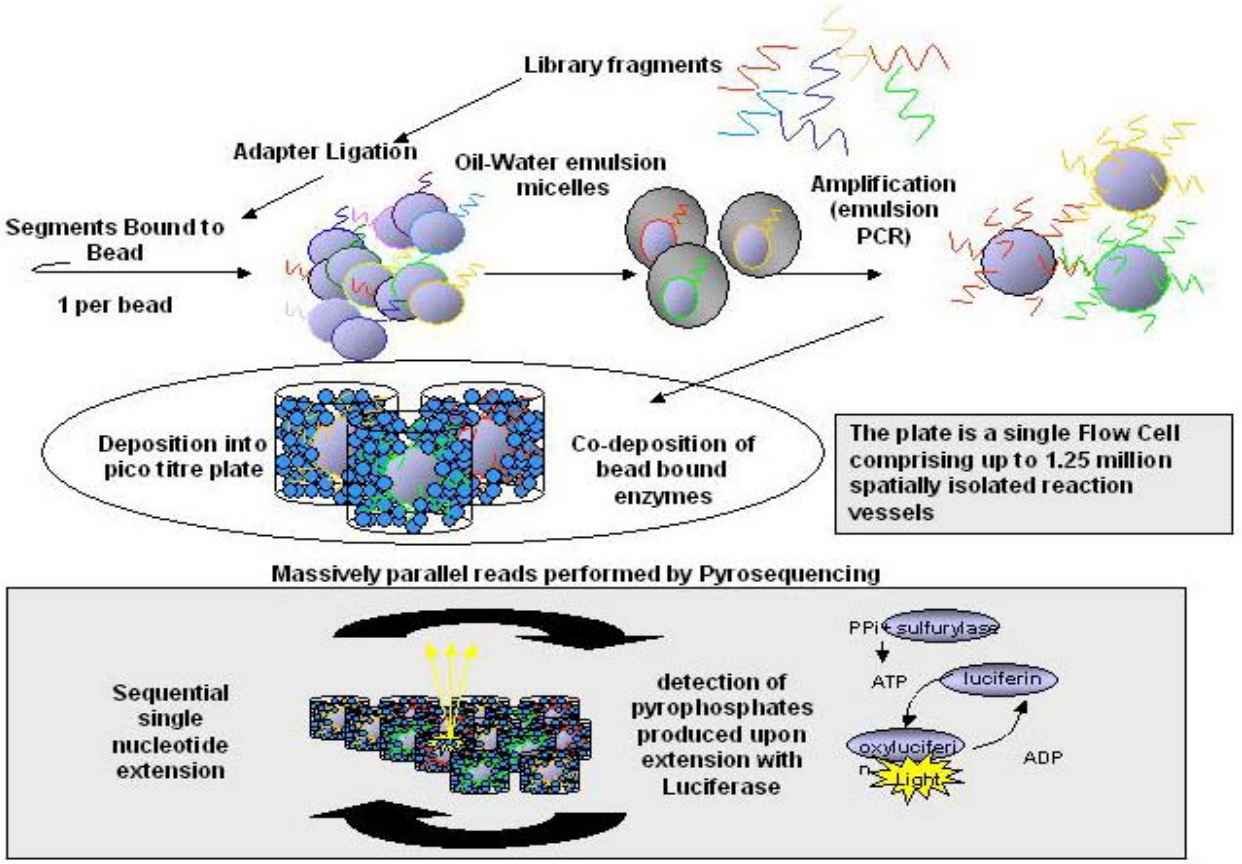
Step 5
Addition of dNTPs is performed sequentially. It should be noted that deoxyadenosine alfa-thio triphosphate (dATP·S) is used as a substitute for the natural deoxyadenosine triphosphate (dATP) since it is efficiently used by the DNA polymerase, but not recognized by the luciferase. As the process continues, the complementary DNA strand is built up and the nucleotide sequence is determined from the signal peaks in the Pyrogram trace.



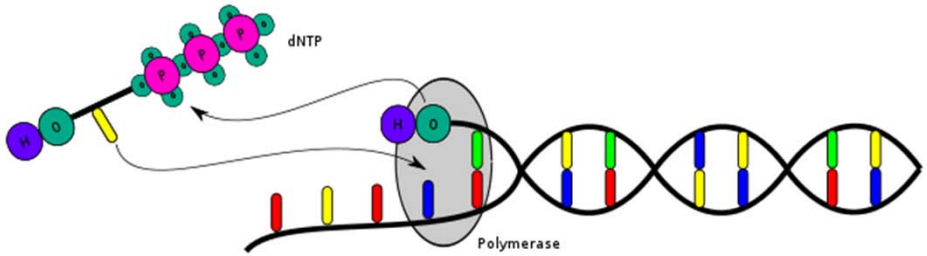
[Pyrosequencing video](https://www.youtube.com/watch?v=bNKEhOGvcaI)

https://www.youtube.com/watch?v=bNKEhOGvcaI

454 Sequencing

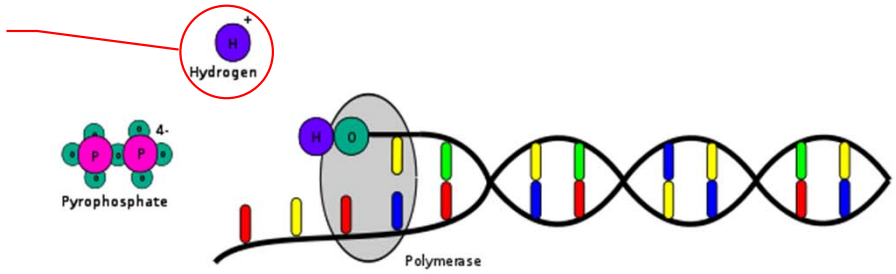


Ion Torrent



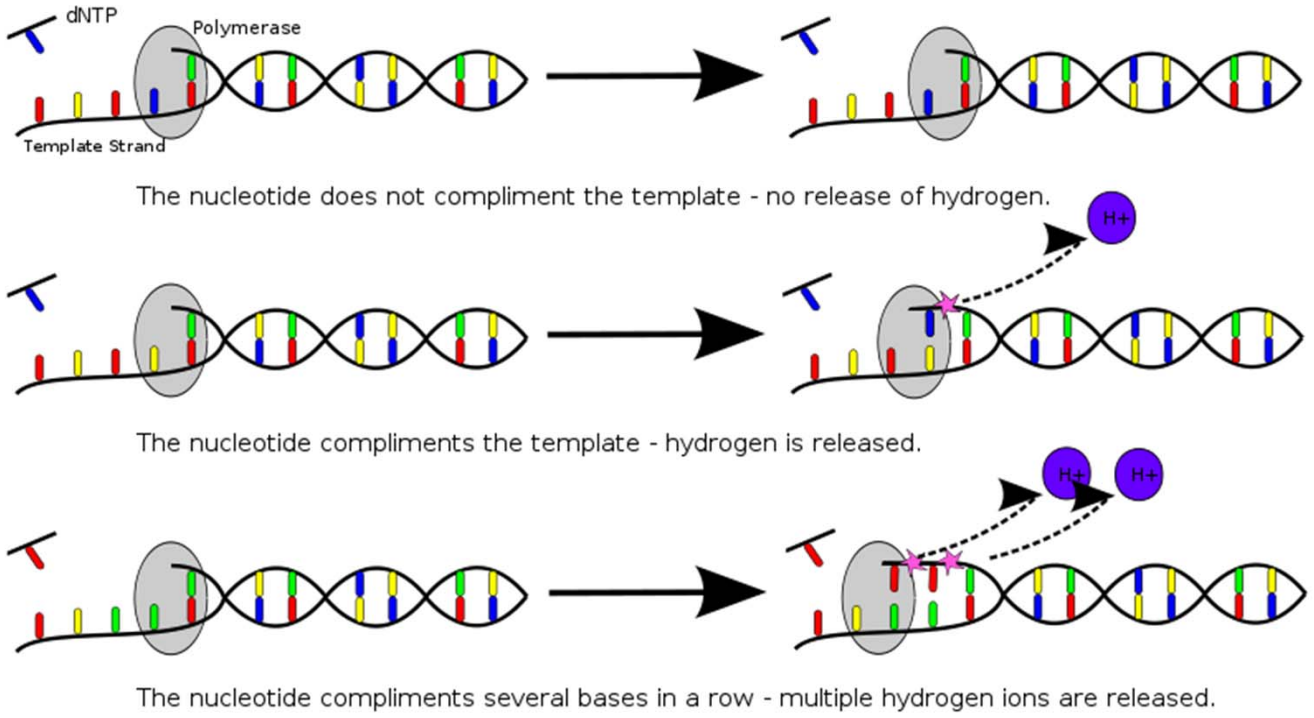
Polymerase integrates a nucleotide.

Ion Torrent

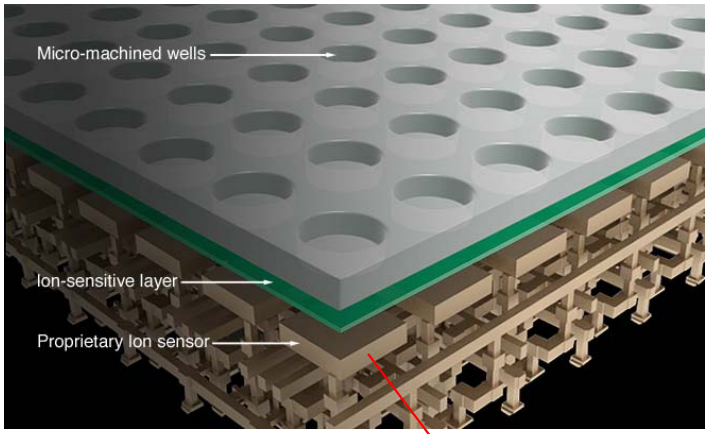


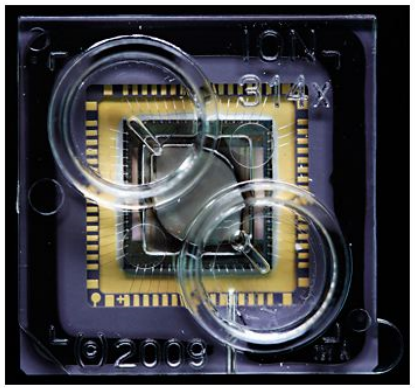
Hydrogen and pyrophosphate are released.

Synthesis With 1 dNTP



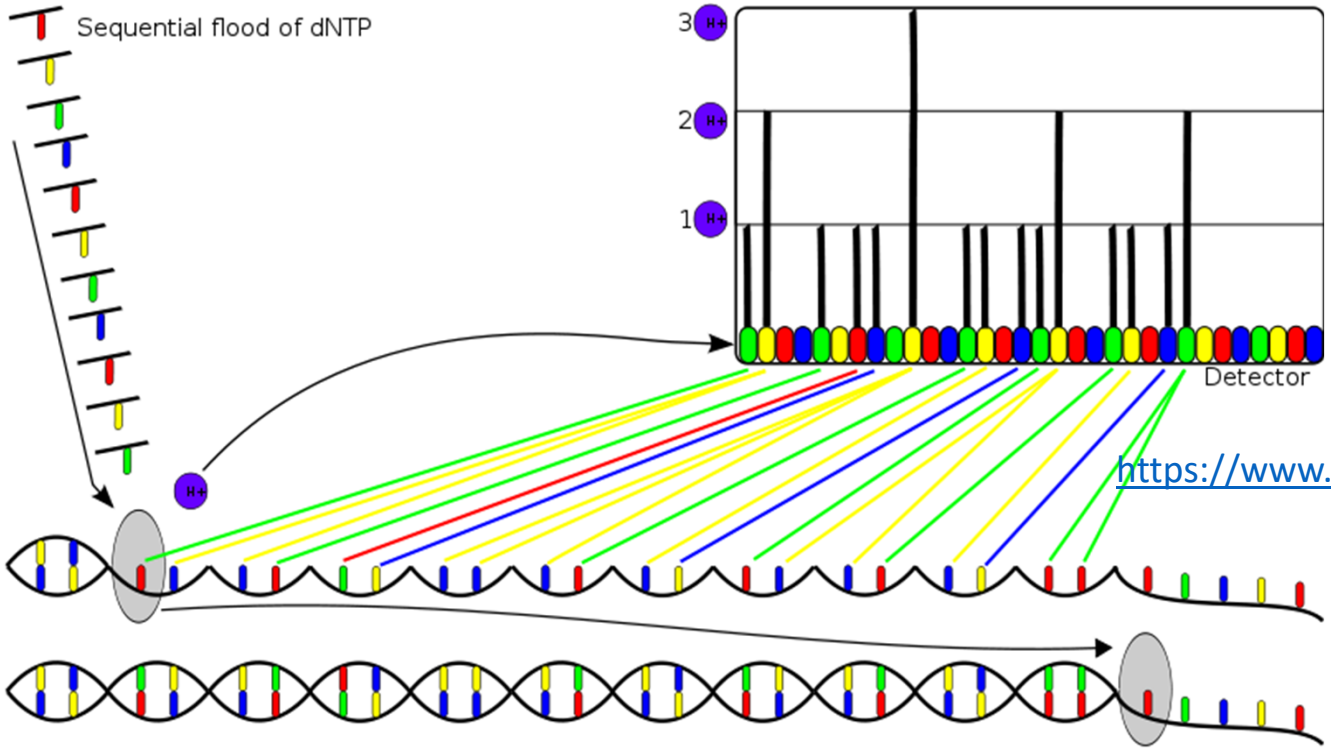
Measuring pH Changes





Use Ion-Sensitive Field Effect Transistor (ISFET) technology to measure pH change electronically

Ion Torrent Sequence Calling



Ion Torrent Video

https://www.youtube.com/watch?v=WYBzbxIfuKs

Illumina Sequencing

https://www.youtube.com/watch?v=fCd6B5HRaZ8

Illumina Sequencing Technology

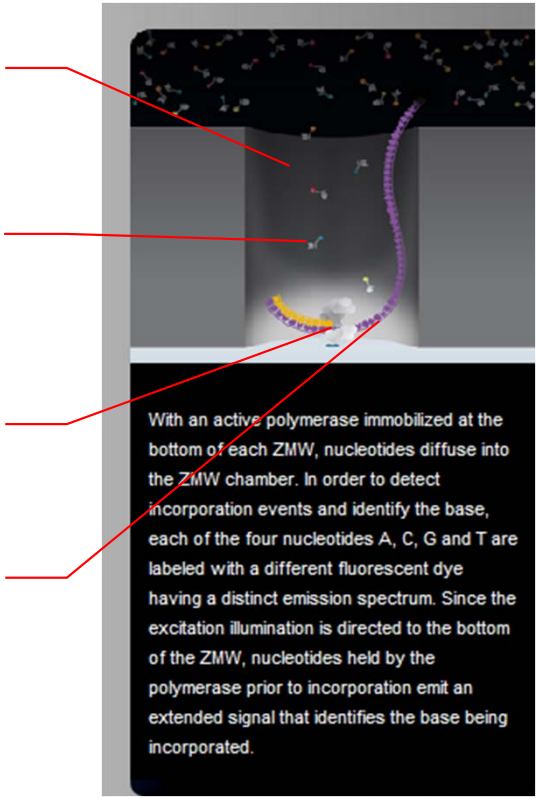
PacBio Sequencing

Zero Mode Waveguide: restricts fluorescence detection to a very narrow layer at bottom of nanowell

Phospho-labelled dNTP's: each dNTP is labelled with a uniquely colored fluorophore at the γ-phosphate; incorporation releases the label, which then diffuses out of the detection zone

Anchored DNA Polymerase: holds template and nucleotide being incorporated in detection zone for several milliseconds

Long, Circular Template: allows template to be read many times ("rolling circle" mechanism)



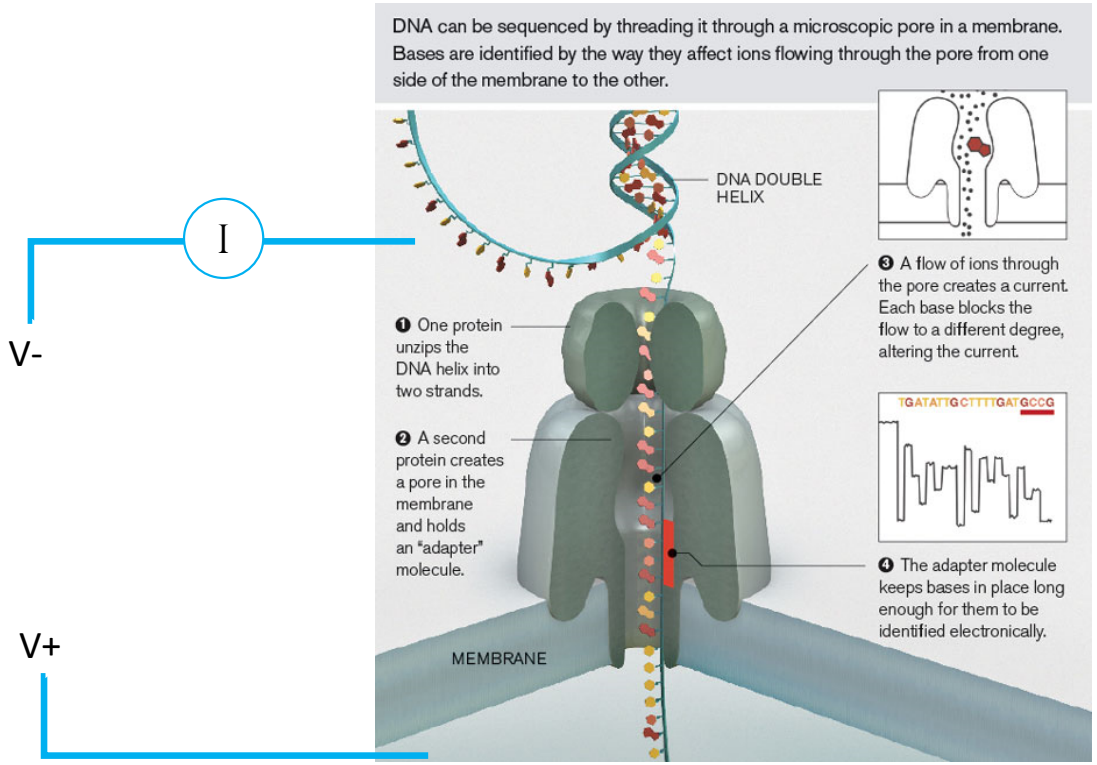
PacBio Video

http://www.youtube.com/watch?v=v8p4ph2MAvI

Oxford Nanopore Sequencing

https://www.youtube.com/watch?v=E9-Rm5AoZGw

Nanopore Sequencing Video



What Is Next?

Assembly Of Sequenced Data (if time permits)