Bio(tech) Interlude: PCR and DNA Sequencing

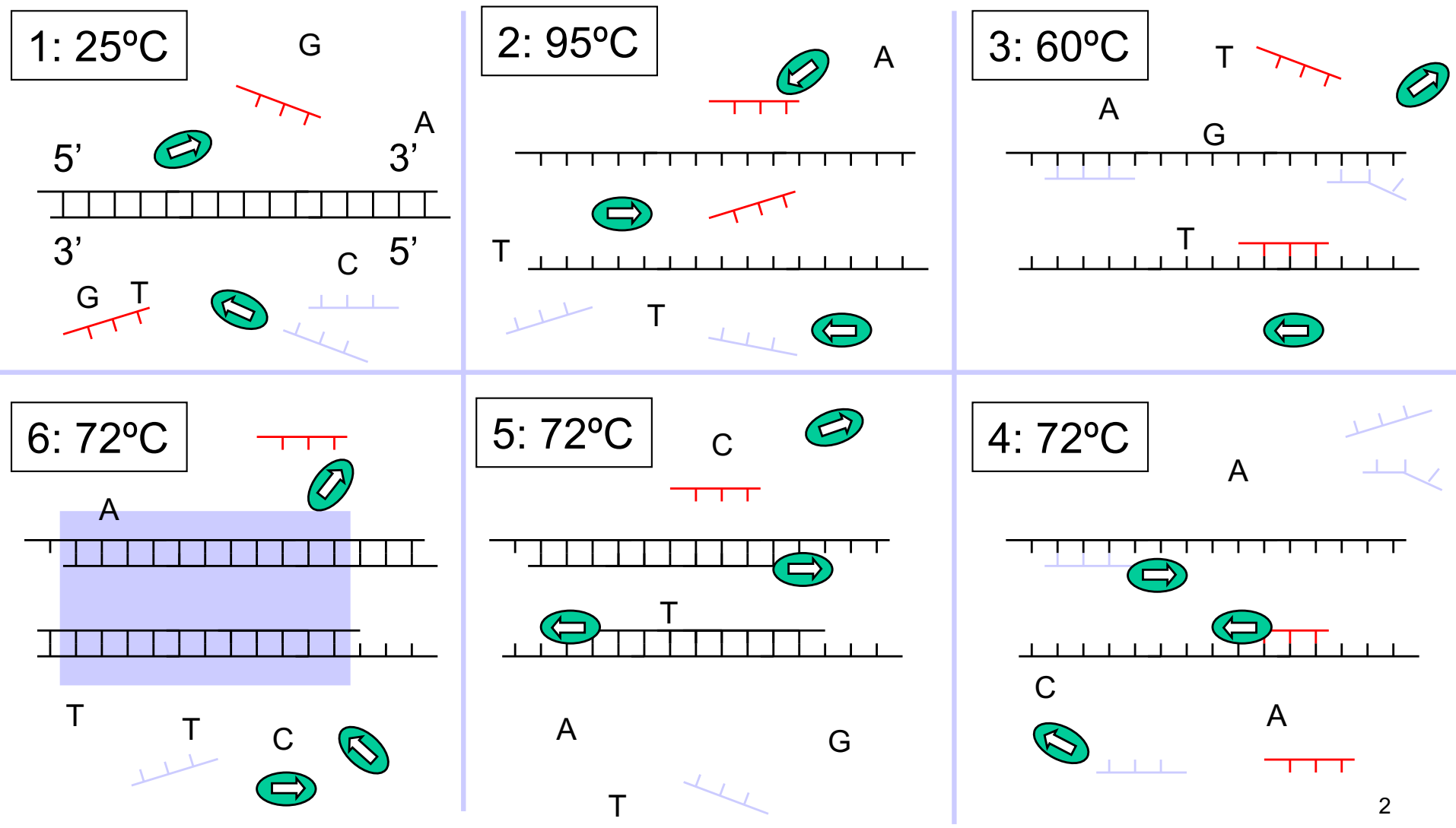
3 Nobel Prizes:

PCR: Kary Mullis, 1993

Electrophoresis: A.W.K. Tiselius, 1948

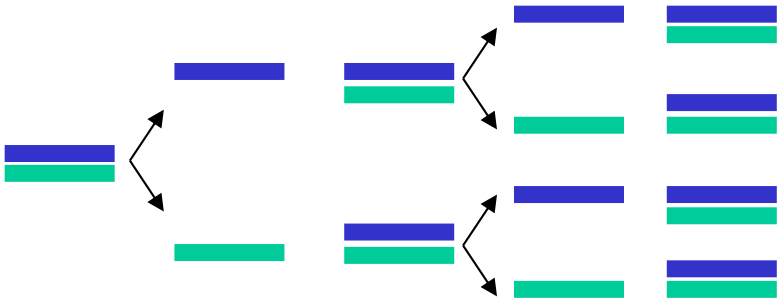
DNA Sequencing: Frederick Sanger, 1980

PCR


1: 25ºC
G
A
5’
3’
3’
5’
C
G
T
2: 95ºC
A
T
T
3: 60ºC
T
A
G
T
6: 72ºC
A
T
T
C
5: 72ºC
C
T
A
G
T
4: 72ºC
A
C
A

Hot spring, near Great Fountain Geyser, Yellowstone National Park

PCR



Ingredients:

- many copies of deoxy nucleotide triphosphates
- many copies of two primer sequences (~20 nt each)
  - readily synthesized
- many copies of Taq polymerase (*Thermus aquaticus*),
  - readily available commercialy
- as little as 1 strand of template DNA
- a programmable “thermal cycler”

Amplification: million to billion fold

Range: up to 2k bp routinely; 50k with other enzymes & care

Why PCR?

PCR is important for all the reasons that filters and amplifiers are important in electronics, e.g., sample size is reduced from grams of tissue to a few cells, can pull out small signal amidst “noisy” background

Very widely used; forensics, archeology, cloning, sequencing, …

And Covid Testing

Table 2. Sensitivity* Mean Estimates of the EUA authorized SARS-CoV-2 molecular diagnostic tests using the FDA SARS CoV-2 Reference Panel

Product LoD (NDU/mL***)	Developer	Test
180	PerkinElmer, Inc.	
540	ScienCell Research Laboratories	
		PCR (RT-qPCR)
		ne PCR Kit
		qPCR Test
		ry SARS-CoV-2 Panel
		ssay
600000	Boston Medical Center	-19 Test

PLOS BIOLOGY

OPEN ACCESS PEER-REVIEWED

METHODS AND RESOURCES

Detection of SARS-CoV-2 RNA by multiplex RT-qPCR

Eriko Kudo, Benjamin Israelow, Chantal B. F. Vogels, Peiwen Lu, Anne L. Wyllie, Maria Tokuyama, Arvind Venkataraman, Doug E. Brackney, Isabel M. Ott, Mary E. Petrone, Rebecca Earnest, Sarah Lapidus, M. Catherine Muenker, [...], Akiko Iwasaki

[view all]

Version 2

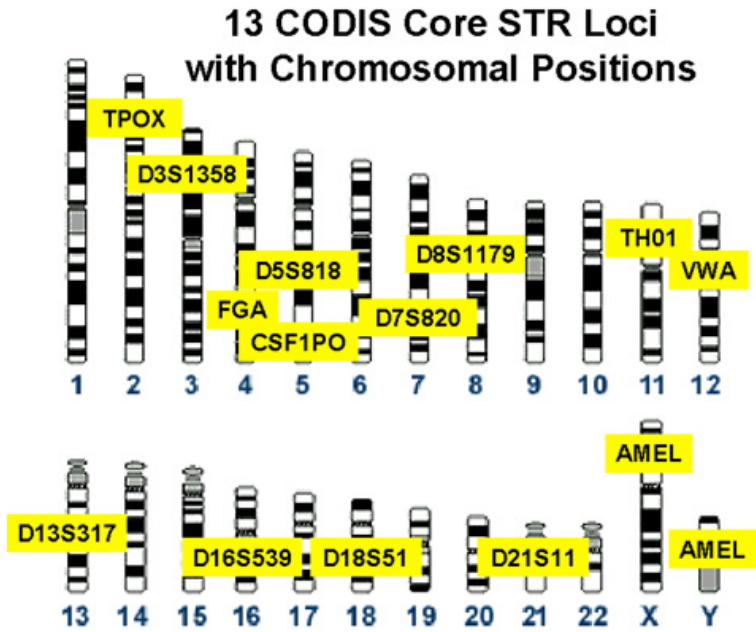
Published: October 7, 2020 • https://doi.org/10.1371/journal.pbio.3000867

A Rapid Test Offers Hope for Community Screening

In a small study in San Francisco, Abbott's BinaxNOW identified infectious people nearly as accurately as a P.C.R. test.

https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-reference-panel-comparative-data

DNA Forensics



E.g. FBI "CODIS" (combined DNA indexing system) data base

As of 2018, over 13,000,000 offender profiles

Picked 13 "short tandem repeats", i.e., variable-length regions of human genome flanked by (essentially) invariant sequences (primer targets), several alleles common at each locus, of which you have 2

Amplify each from, e.g., small spot of dried blood

Measure product lengths (next slides)

http://www.fbi.gov/about-us/lab/biometric-analysis/codis

http://www.dna.gov/solving-crimes/cold-cases/howdatabasesaid/codis/

Gel Electrophoresis

DNA/RNA backbone is negatively charged (they're acids)

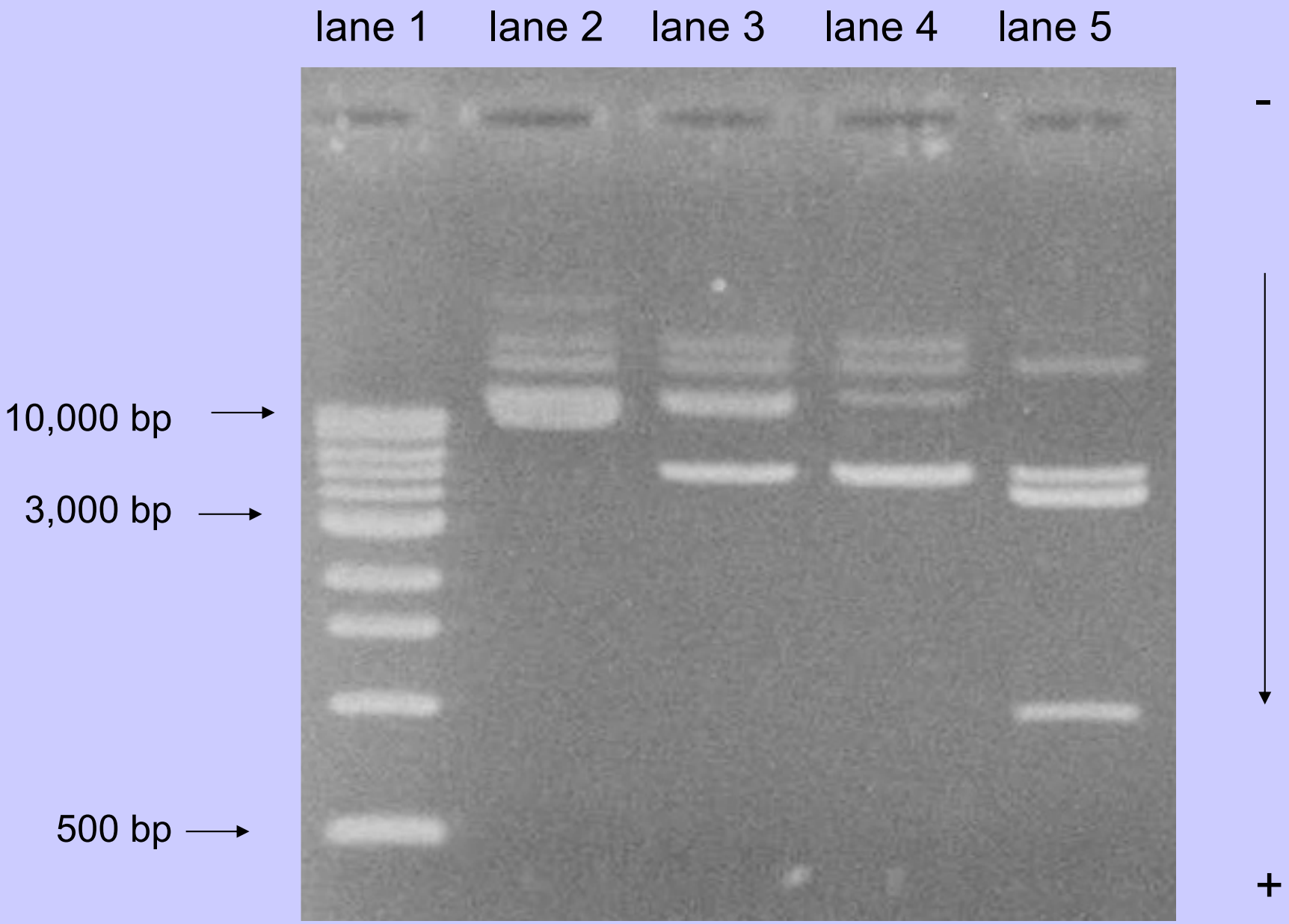
Molecules moves slowly in gels under an electric field

- agarose gels for large molecules
- polyacrylamide gels for smaller ones

Smaller molecules move faster

So, you can *separate DNAs & RNAs by size*

Nobel Chem prize, 1948 Arne Wilhelm Kaurin Tiselius

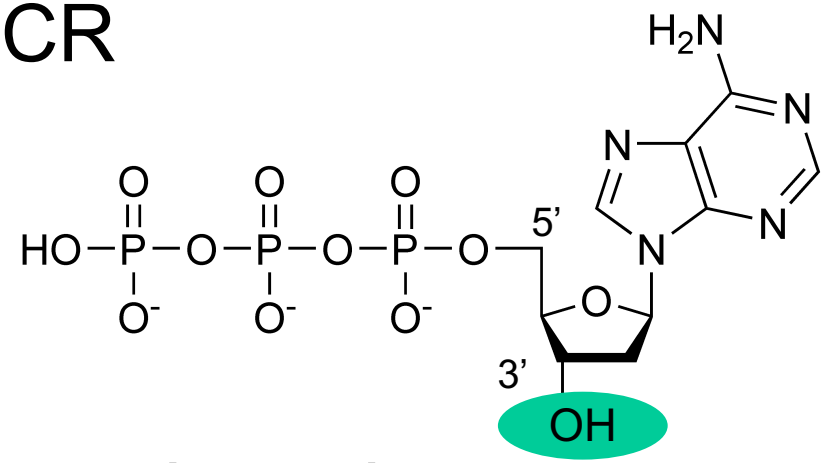

lane 1
lane 2
lane 3
lane 4
lane 5
10,000 bp
3,000 bp
500 bp
-
+

DNA Sequencing – Sanger Method

Like one-cycle, one-primer PCR

Suppose 0.1% of A's:

- are *di*-deoxy adenosine's; backbone can't extend
- carry a green florescent dye



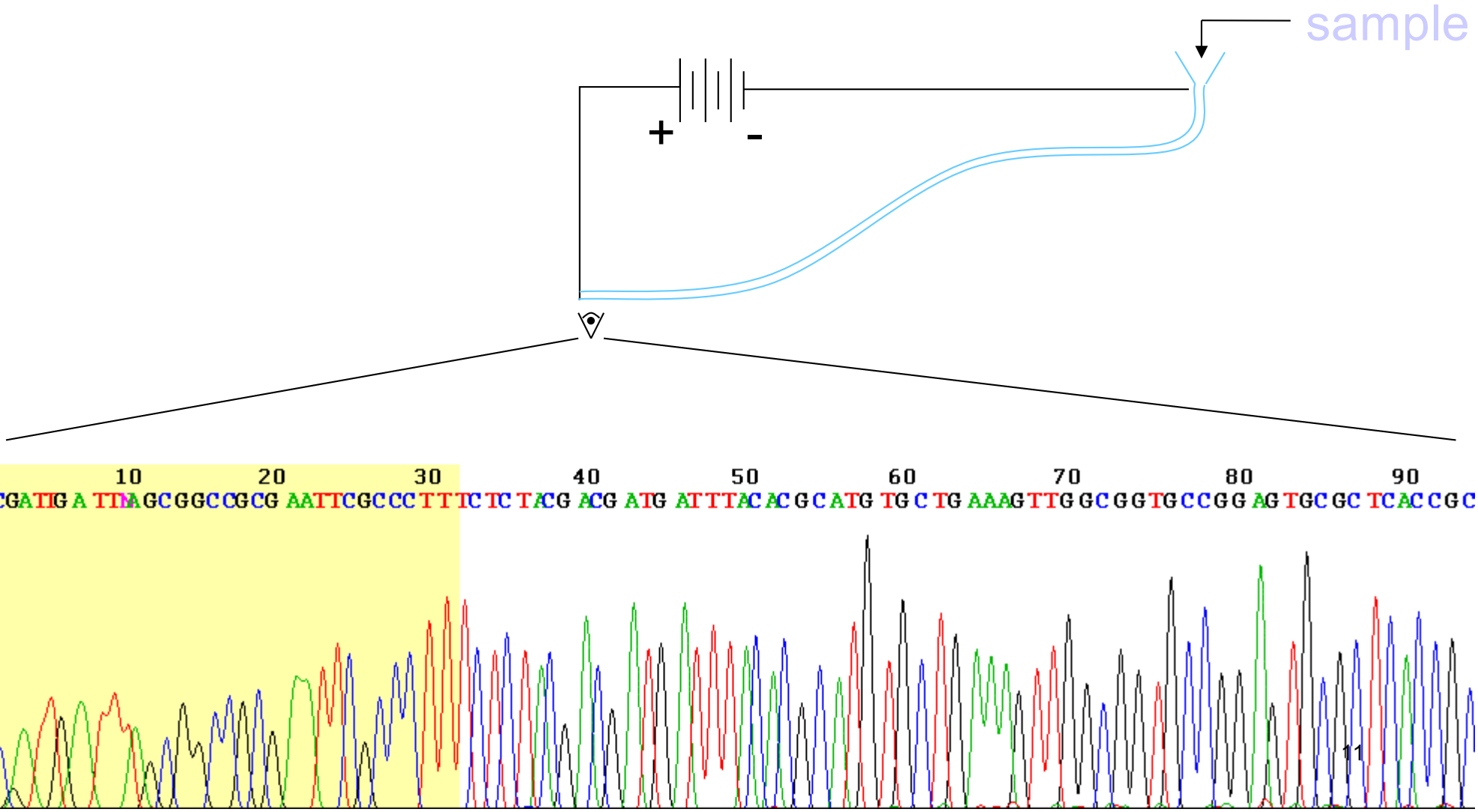
Separate by capillary gel electrophoresis

If frags of length 42, 49, 50, 55 … glow green, those positions are A's

Ditto C's (blue), G's (yellow), T's (red)

DNA Sequencing

Sanger with capillary electrophoresis



Sequencing A Genome

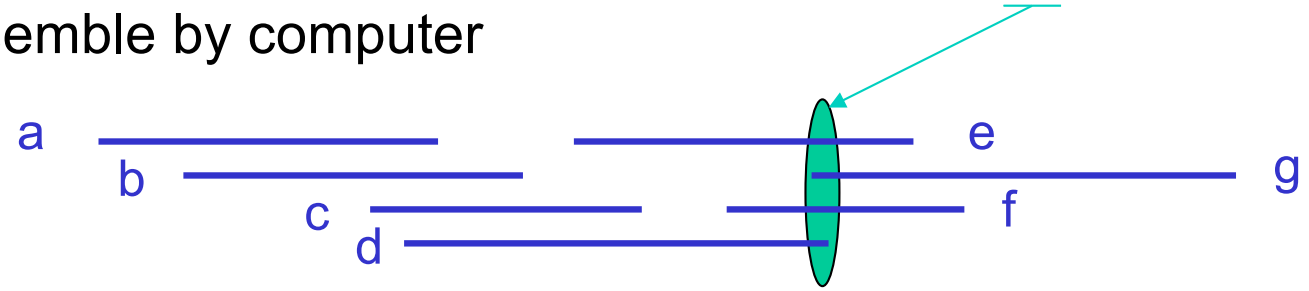
Highly automated

Typical Sanger "read" about 600 nt

"Whole Genome Shotgun" approach:

- randomly fragment (many copies of) genome
- sequence many, enough to cover each base 10x or more times
- reassemble by computer

E.g., human genome project:
≈ 30Gbases and
≈ $3 \times 10^9 / 600 \times 10$
= 5×10^7 reads



Complications: repeated region, missed regions, sequencing errors, chimeric DNA fragments, …

But overall accuracy ~10^{-4}, if careful

Illumina Sequencing

~1 billion microscopic PCR “colonies” on 1x2” slide

“Read” ~50-150bp of sequence from (1 or 2) ends of each

Ends fluorescently labeled, blocked, chemically cycled

“Reversible dye terminators”

Automated: takes a few days; ~ 100 G bases/day

Costs a few thousand dollars

Generates terabytes of data (mostly images)

I,e., ~ 30x human genome/day

(you need 25x-50x to assemble)

(equal to all of pre-2008 Genbank)

“Next Generation” Sequencing

Many technical improvements to Sanger approach over many years, culminating in highly automated machines used for the HGP

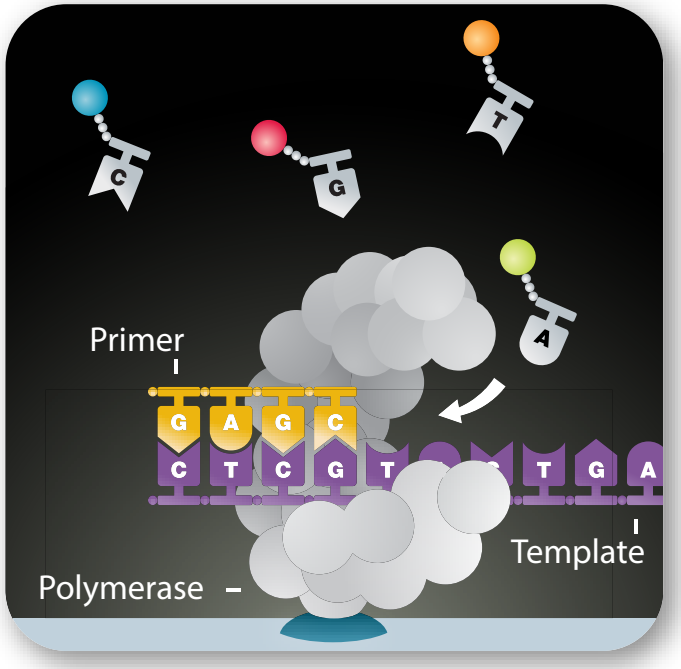
Since then, many innovative new ideas/products:

- Helicos: single molecule flourescence tethered to flow cell
- Ion Torrent: semiconductor detection of ions released by polymerase
- Roche 454: emulsion PCR; pyro sequencing
- ABI SOLiD: emulsion PCR, sequence by ligation, “color-space”
- Illumina: colony PCR; reversible dye terminator
- Oxford Nanopore
- Pacific Biosciences: single tethered polymerases in “zero mode waveguide” nano-wells, circularized DNA, “real time”
- Complete Genomics: rolling circle replication/DNA nanoballs

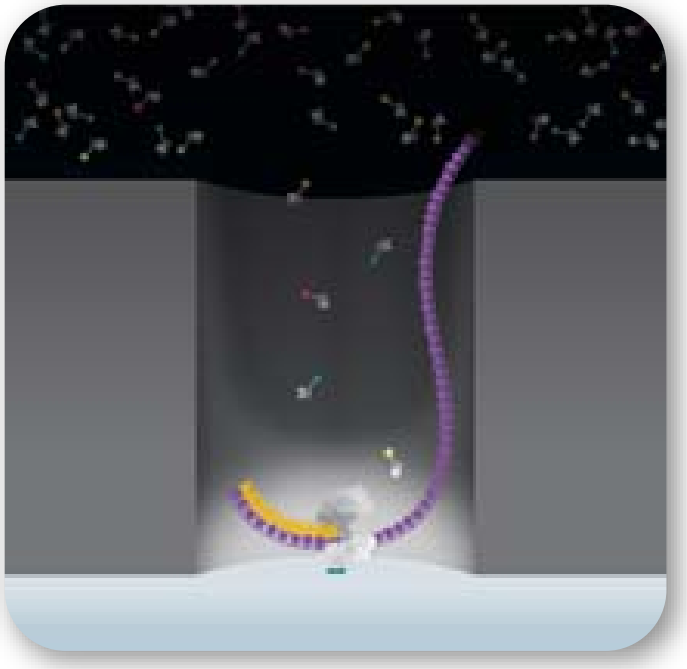
Off-market?

Technology is changing rapidly!

Pacific Biosciences



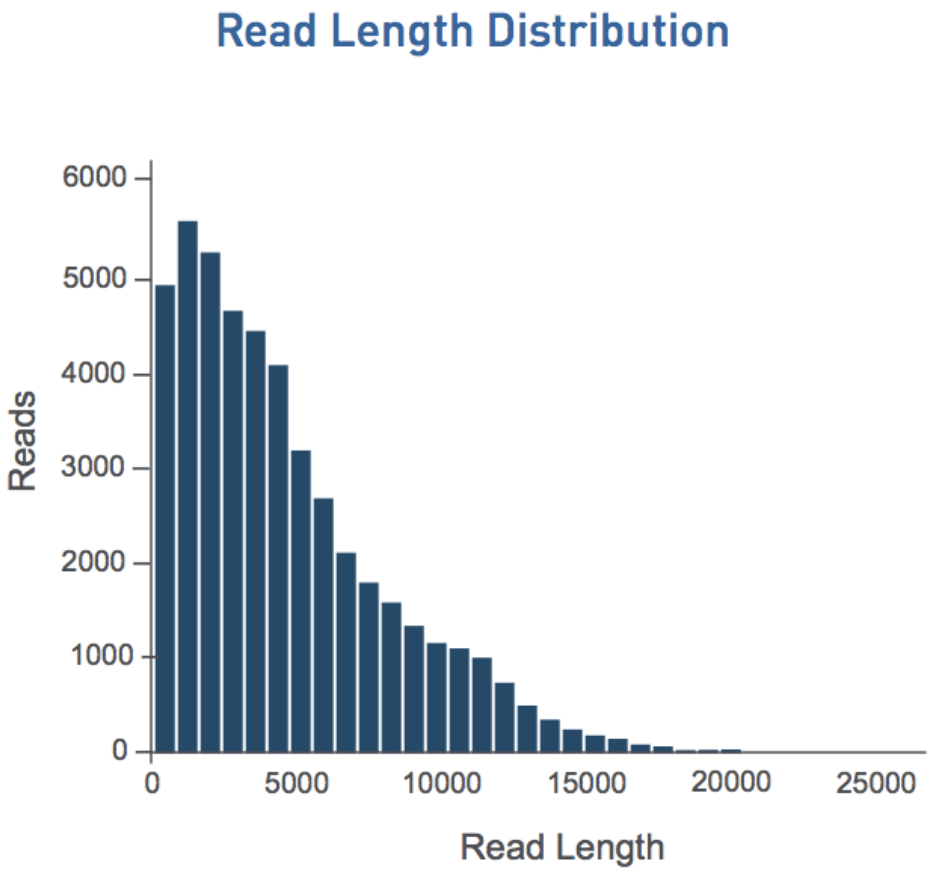
Phospholinked nucleotides



Zero-Mode Waveguides

http://files.pacb.com/pdf/PacBio_RS_II_Brochure.pdf http://www.globenewswire.com/NewsRoom/Attachment/18068

Pacific Biosciences



Advantages:
- single molecules
- long reads
- direct CH_3 detection

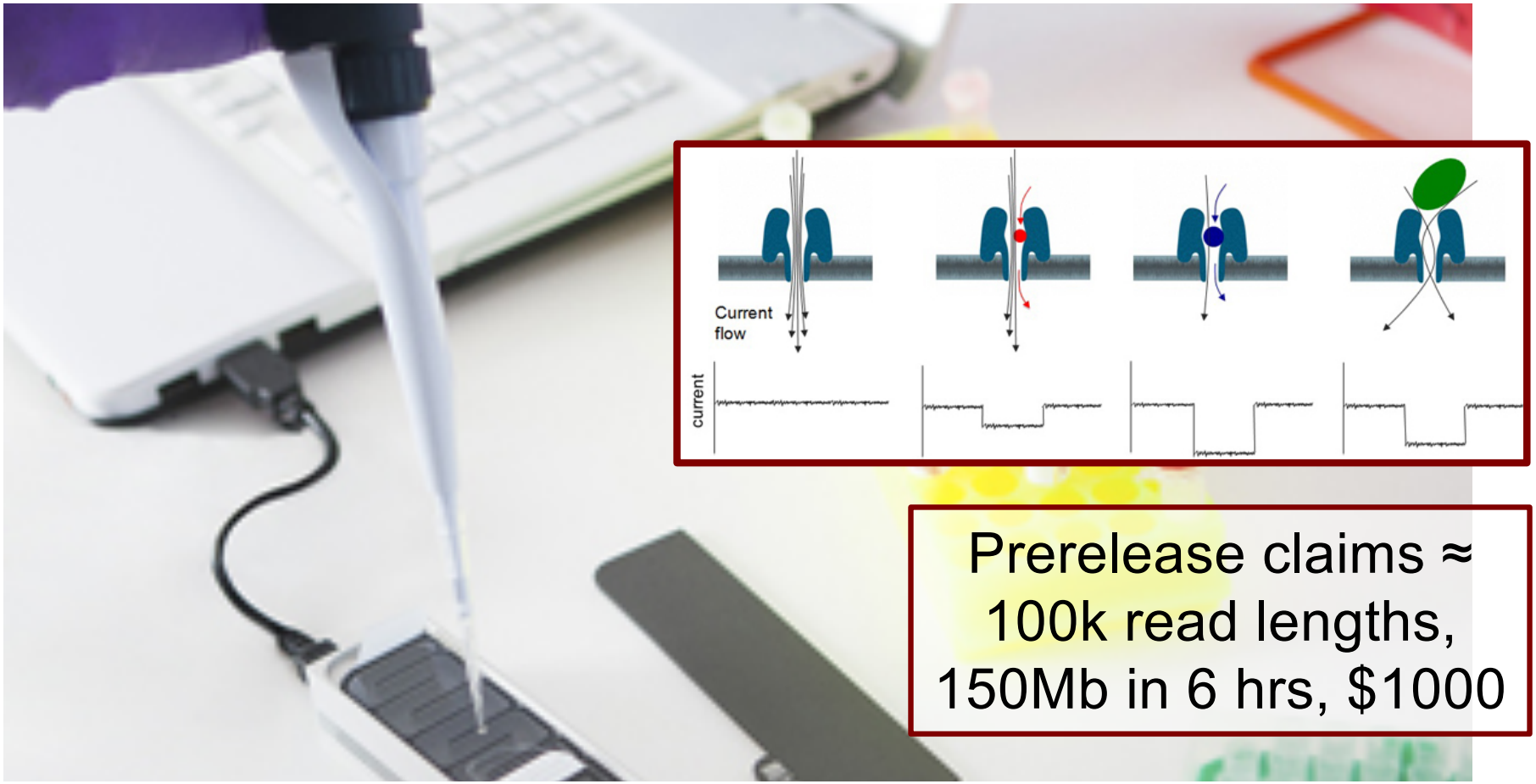
Disadvantages:
- throughput
- error rate; (circularize?)

Read Length:	
Average:	4,606 bp
95th Percentile:	11,792 bp
Maximum:	23,297 bp
Throughput per SMRT® Cell:	216 Mb
	47,197 reads

Based on data from 11 kb plasmid library using a 120 minute movie

http://www.pacificbiosciences.com/img/assets/smrt_sequencing_advantage_readlength_lg.png

Oxford Nanopore



Prerelease claims ≈
100k read lengths,
150Mb in 6 hrs, $1000

http://www.nanoporetech.com/uploads/Technology_New/MinION/MinION_117.jpg

http://www.nanoporetech.com/uploads/Technology_New/Introduction_To_Nanopore_Sensing/Nanopore_sensing_101_0_rs.jpg

Personal Genomes

2001: ~$2.7 billion (Human Genome Project)

2003: ~$300 million

2007: ~$1 million

2008: ~$60 thousand

2009: ~$4400

Now : ≪$1000

bioinformatics not included…

Summary

PCR allows simple *in vitro* amplification of minute quantities of DNA (having pre-specified boundaries)

Sanger sequencing uses

- a PCR-like setup with modified chemistry to generate varying length prefixes of a DNA template with the last nucleotide of each color-coded
- gel electrophoresis to separate DNA by size, giving sequence

Sequencing random overlapping fragments allows genome sequencing (and many other applications)

"Next Gen" sequencing: many innovations

- throughput up, cost down (lots!)