

Introduction to Molecular Methods and ‘Omics Technologies

Dominic Alba

Philippine Genome Center



Molecular Methods

- Detecting/identifying variation
 - What kind of genetic variations exists?
 - What genetic variations are associated to traits and/or diseases?
- Characterizing/functionalization variation
 - What is the consequence of genetic variation?
 - How does genetic variation connect to disease and/or traits?
- Making use of variation
 - How can I use known variation?

Molecular Methods

- Detecting/identifying variation
 - PCR, Sanger Sequencing, Next-Generation Sequencing
- Characterizing/functionalization variation
 - Cloning, Expression, Reporter Assays
- Making use of variation
 - Gene Editing, Directed Evolution, Diagnostics



Laws of genetics
was discovered
1865

DNA was
shown to be the
hereditary
material
1944

	U	C	A	G
U	Phe	Ser	Ser	Leu
C	Ile	Pro	His	Gln
A	Ile	Thr	Asn	Arg
G	Val	Ala	Asp	Gly

Genetic code
was cracked
1966



GenBank was
established
1982

First automated
sequencer
1986



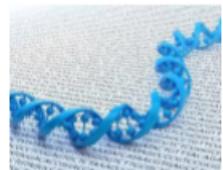
Human Genome
Project launched
1990

Genomes of yeast, *E. coli*,
roundworm, fruit fly, and *A.
thaliana* were completely
sequenced

First completed sequence
of a human chromosome
(Chr. 22)



Draft version of human
genome sequence
completed and published
1996 - 2001



Next Generation
Sequencing
technology was
introduced
2005

1953
DNA structure
was described



1977
Sequencing
methods were
developed

	G	A	T	C
1	-	-	-	-
2	-	-	-	-
3	-	-	-	-
4	-	-	-	-

1985
PCR was
invented



1987
First-gen human
genetic map was
developed



1995
First bacterial
genome was
sequenced



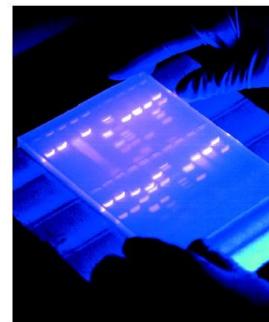
2002
Draft versions of
mouse, rat and rice
genomes were
completed

2003
Human Genome
Project completed

Molecular Approaches in Detecting Variation

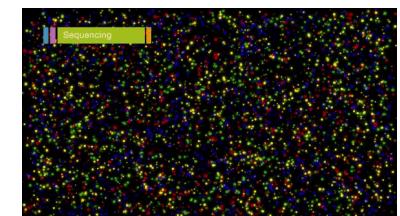
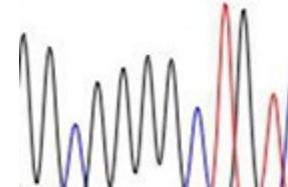
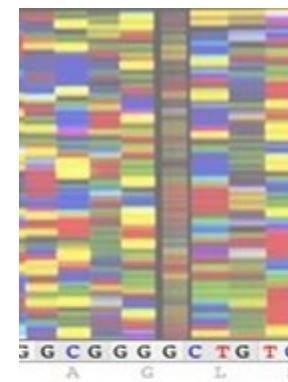
INDIRECT METHODS

PCR, Electrophoresis, Hybridization



DIRECT METHODS

Sequencing

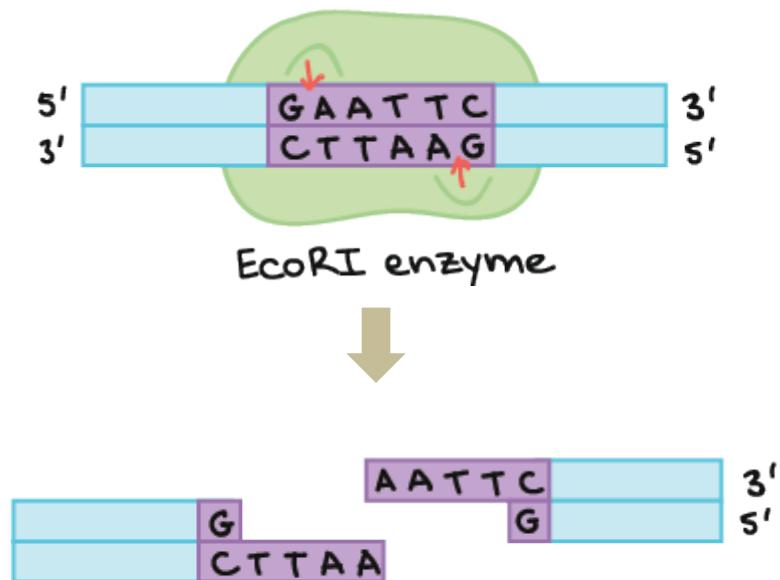


Indirect Methods

Restriction Enzymes

Restriction enzymes are a group of proteins that recognize palindromic **restriction sites**. Once it sees the correct site, it **cuts** the DNA in a region near or within it. This action is **reproducible**.

- A restriction enzyme always **cuts both strands**.
- A restriction enzyme will only **recognize its own restriction site***
- For most REs, the cuts can be **staggered**.
 - Staggered ends produce what are called **sticky ends**.
 - Non-staggered cuts produce **blunt ends**.



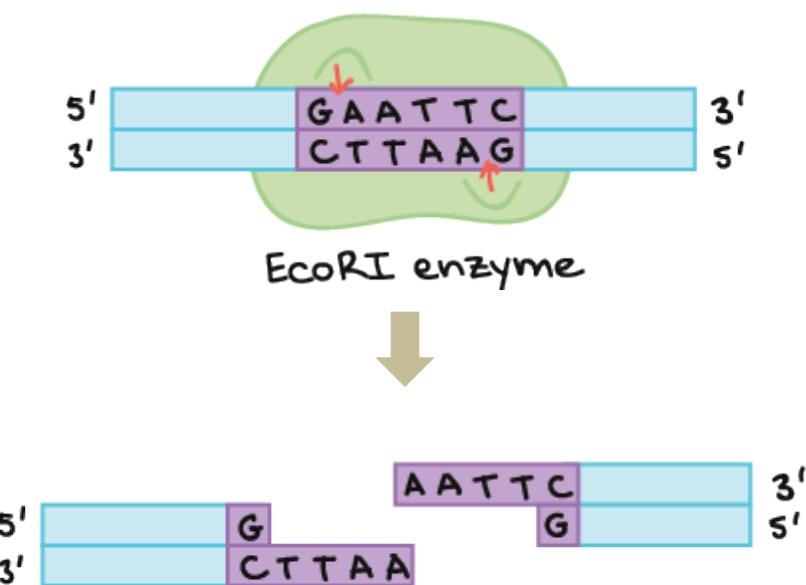
Indirect Methods

Restriction Enzymes

- **Restriction enzymes** will always produce the same **sticky ends***. This means that two different DNA molecules cut with the same enzyme can **stick to one another**.
- Using RE digestion, you can make two different DNA molecules with same sticky ends “stick” to one another.
- How do you join them?



“Sticky ends” stick together,
but gaps remain

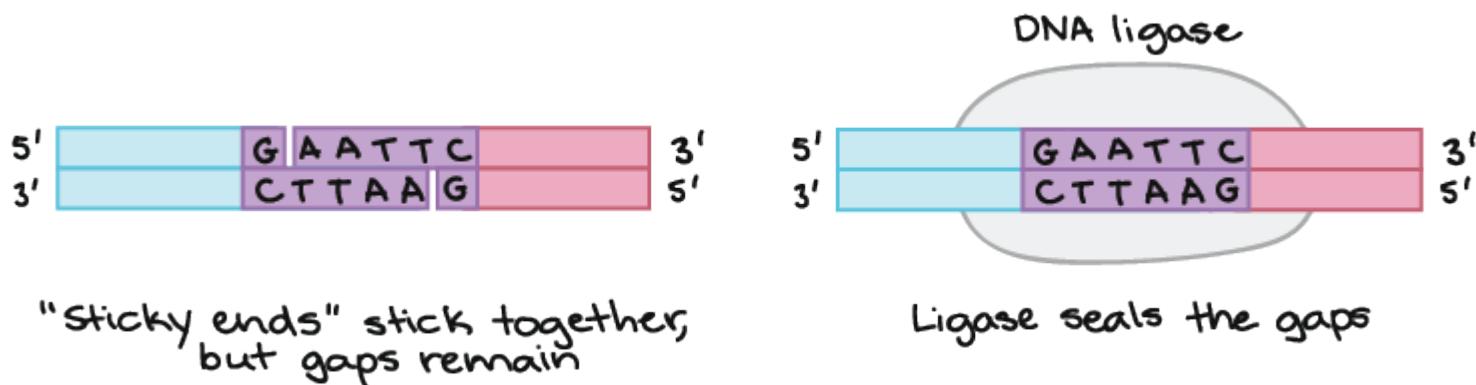


Indirect Methods

DNA Ligase

Remember: **ligase** is an enzyme that is used to join **Okazaki fragments** together.

By using ligase, we can **rejoin the phosphodiester bonds** of the DNA backbone. This reinstates the integrity of the DNA backbone.



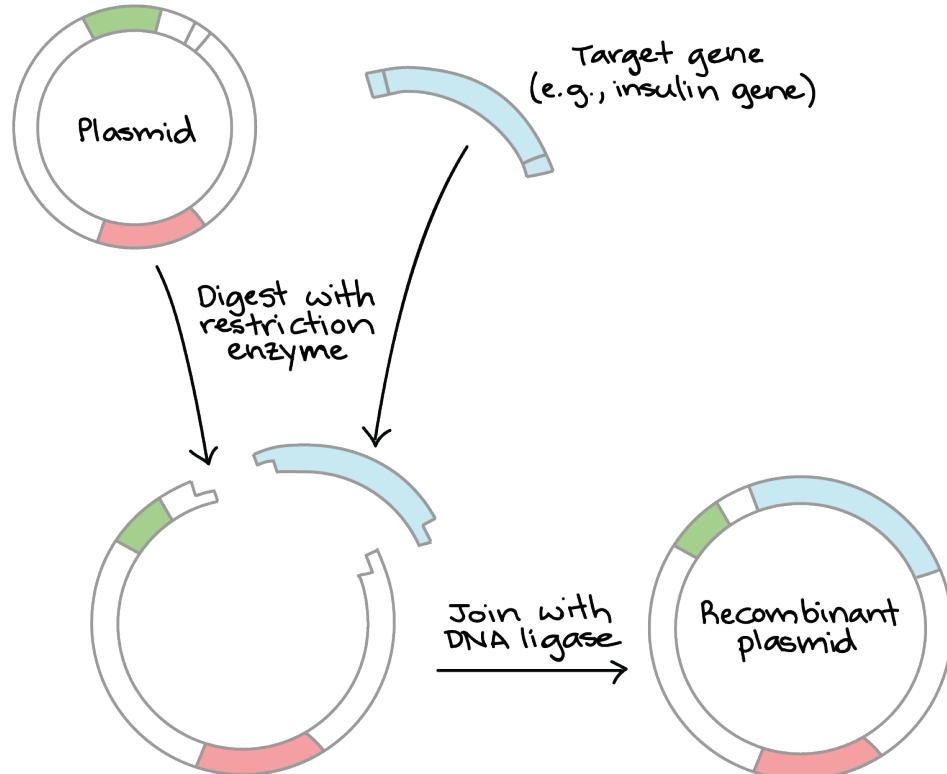
Indirect Methods

Recombinant DNA Technology

Cutting with restriction enzymes and pasting with DNA ligase is the foundation of recombinant DNA technology.

Using REs and ligase, you can cut DNA from two different sources. You can then paste them together into a single **recombinant DNA molecule**.

But how do you make more copies of a recombinant DNA molecule?

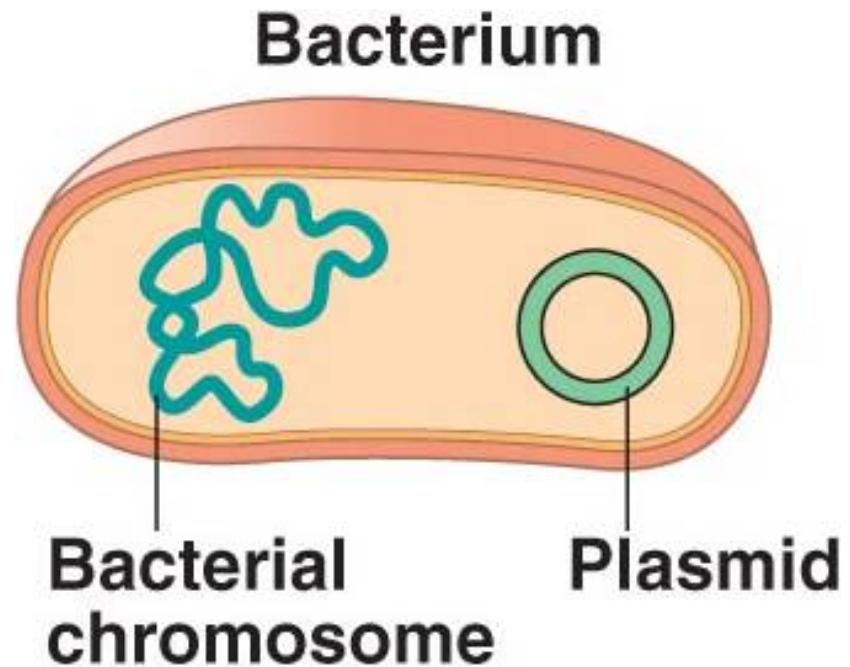


Indirect Methods

Plasmids

Plasmids are short circular molecules of DNA found in bacteria.

- Short
- Extrachromosomal
- Independently replicating
- Contain genes that allow bacteria to deal with stress
- Not necessary for day-to-day living of bacteria
- Can be transferred to another bacterium without a plasmid

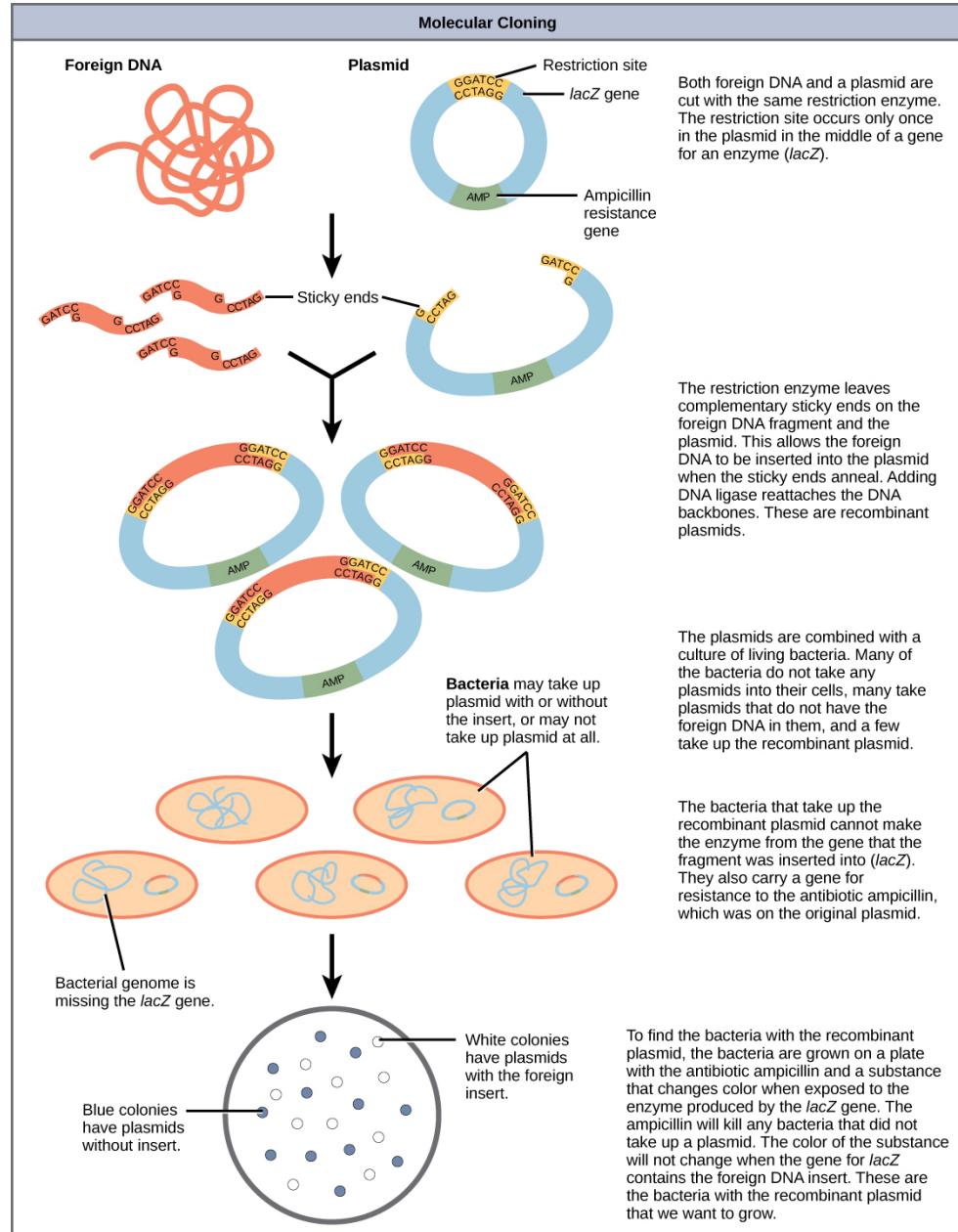


Indirect Methods

DNA Cloning

We can use plasmids to make more of a recombinant DNA molecule. This process is called **DNA cloning**.

1. Cut the desired gene from the DNA of the source organism.
2. Cut a plasmid with the same RE.
3. Paste the gene into the plasmid.
4. Insert the plasmid back into a bacterium (a transformation)
5. Profit.



Electrophoresis

Separation of DNA fragments according to:

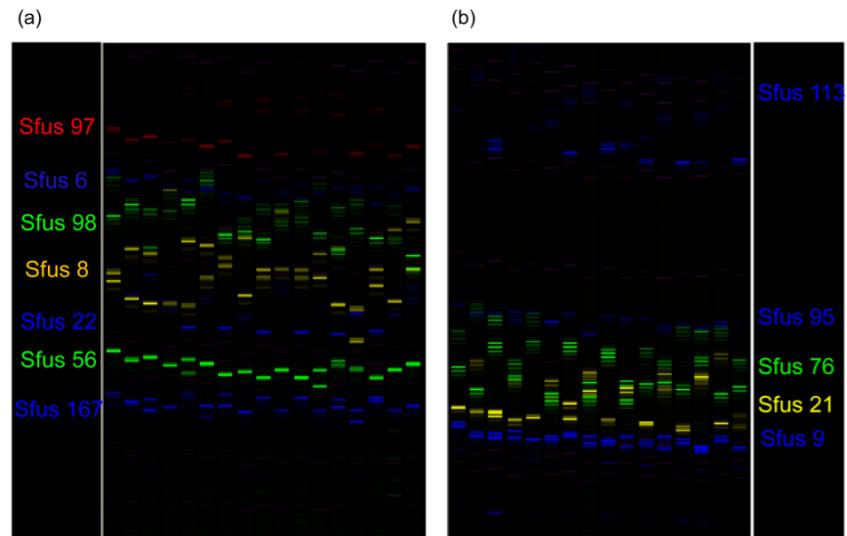
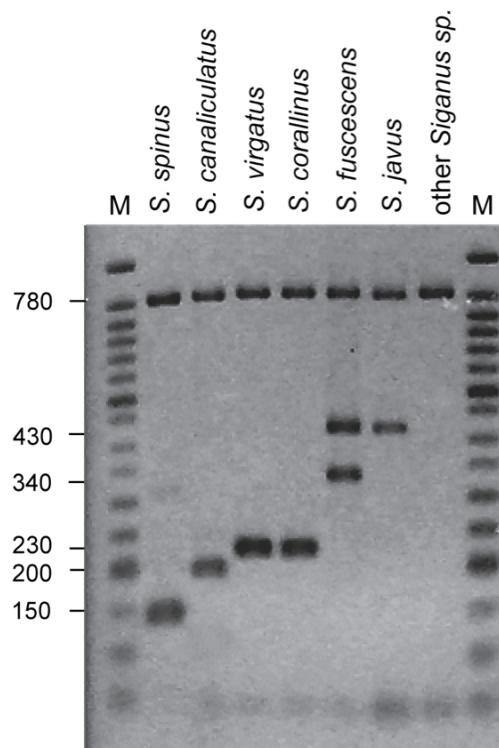
SIZE

PHYSICAL properties influenced by sequence
(melting temperature, secondary structure)

Electrophoresis

SIZE separation

- Migration in an electric field (moving towards + electrode)
- Smaller fragments migrate faster through the gel
- Detection via staining or labelled fragments (fluorophores)
- Size standards to infer fragment sizes



Fragment analysis of fluorophore-labelled microsatellite loci

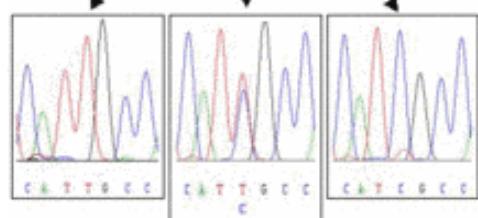
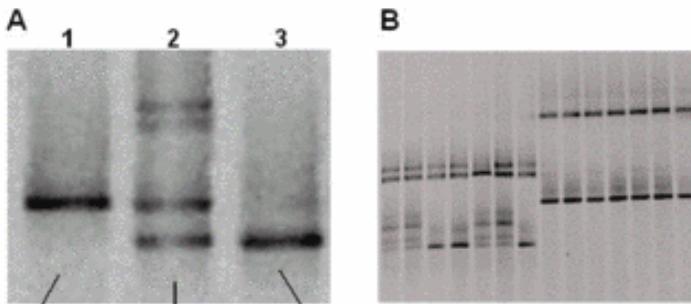
EtBr stained multiplex PCR amplicons

Indirect Methods

Electrophoresis

Separation based on DENATURATION temperature

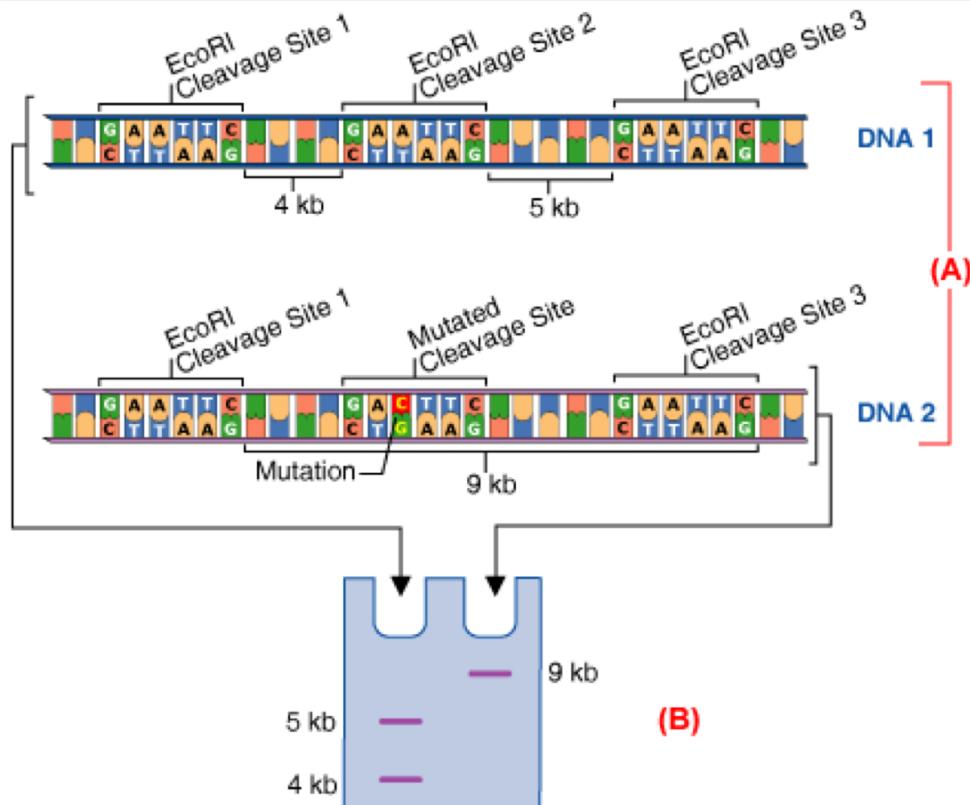
- Melting temp = $f(\text{sequence})$
- Denaturing gradient applied across the gel
 - Temperature (TGGE)
 - Chemical agent (urea; DGGE)
- As DNA is denatured (ssDNA), migration slows; location of fragment an indication of melting temperature
- Detection via staining (EtBr; silver)



Silver-stained DGGE showing homozygote and heterozygote DNA

Restriction Fragment Length Polymorphisms

Detection of mutations based on cleavage of DNA by restriction enzymes at specific recognition sites

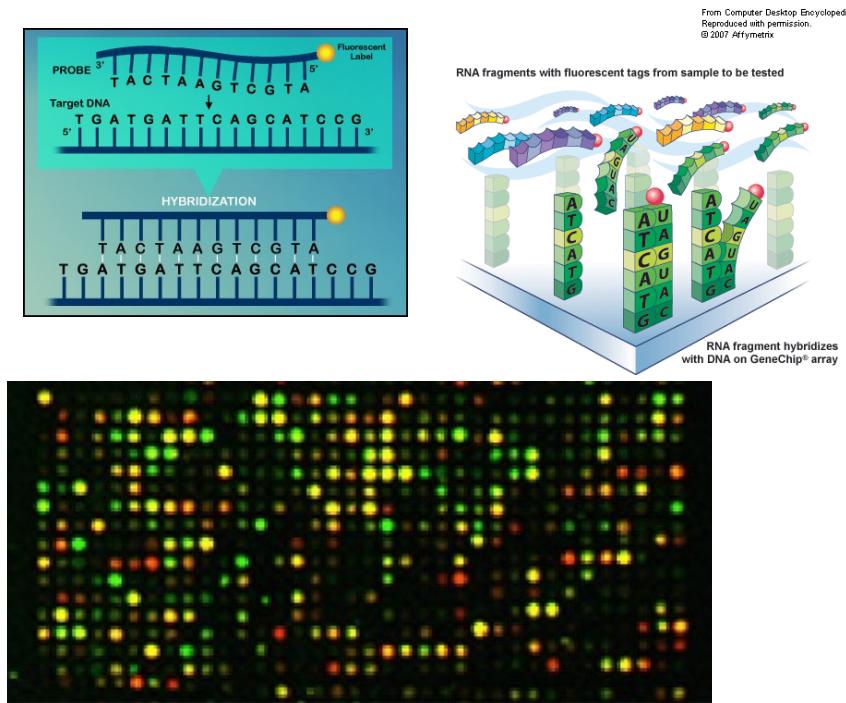


- Digestion yields variable-sized fragments (restriction fragment length polymorphism; RFLPs)
- Information on presence/absence of restriction site

Indirect Methods

Hybridization and Blotting

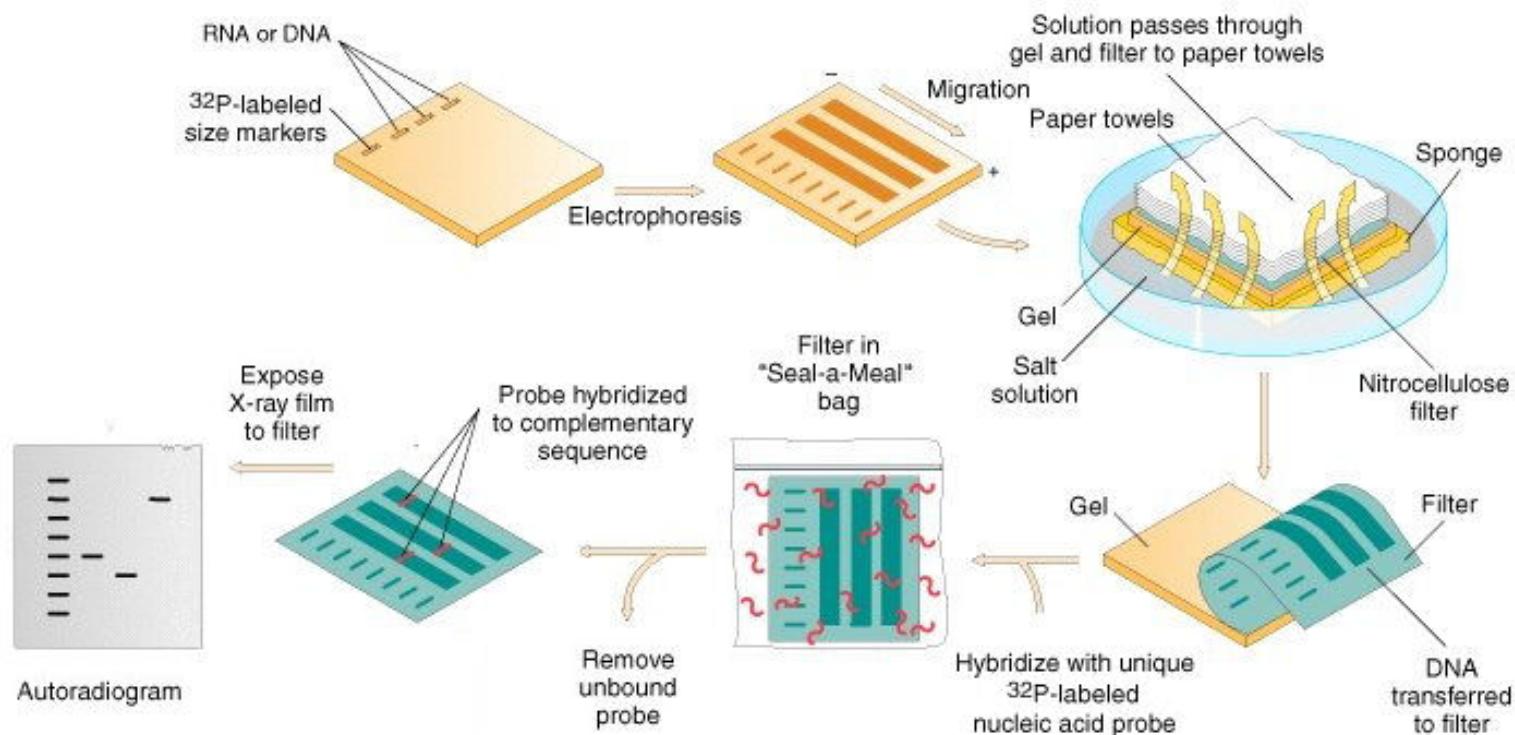
- Principle: Base-pairing between complementary nucleotides
- Template + labelled probe → hybridization = signal production
 - Quantitative: signal intensity defines degree of hybridization
 - Qualitative: presence or absence of signal



Section of a microarray image, courtesy of Eric Jeffery, Corixa Corporation

Indirect Methods

Hybridization and Blotting



Polymerase Chain Reaction

Polymerase chain reaction is a method widely used to make many copies of a specific DNA segment.

Discovered by Kary Mullis in 1983. Kary Mullis won the Nobel Prize in Chemistry in 1993 for his discovery of PCR



Kary Mullis

Indirect Methods

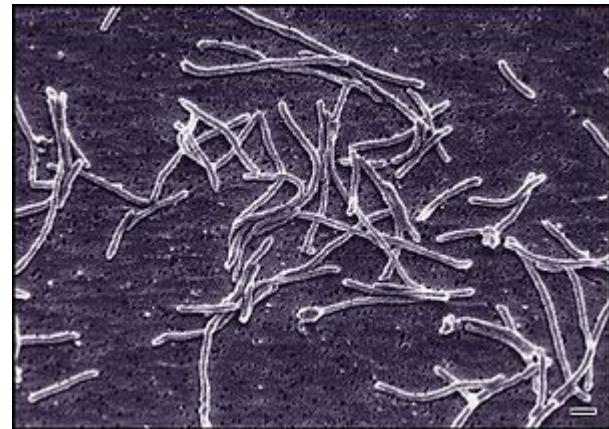
Thermus aquaticus

Species of **Archaea**. Archaea are prokaryotes which have the ability to thrive in extreme conditions.

Thermus aquaticus is an Achaean species with the capability to tolerate extremely high temperatures.

Discovered in 1969 by **Thomas Brock** and **Hudson Freeze** (Indiana University) from geysers at the Yellowstone National Park.

Normally thrives at **70°C**, but can survive from 50°C to 80°C.



Indirect Methods

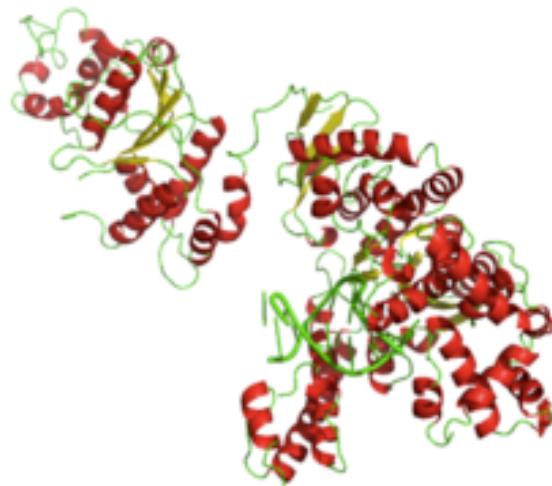
Taq Polymerase

DNA polymerase from *Thermus aquaticus*.

Optimum activity at 72°C. No activity at 90°C, but does not denature. Also no activity at 60°C.

Used by **Kary Mullis** (Cetus Corporation) in 1983 to synthetically replicate DNA exponentially.

The discovery of Taq polymerase enabled the development of **Polymerase Chain Reaction (PCR)**. Mullis together with Michael Smith were awarded the **1993 Nobel Prize in Chemistry**.

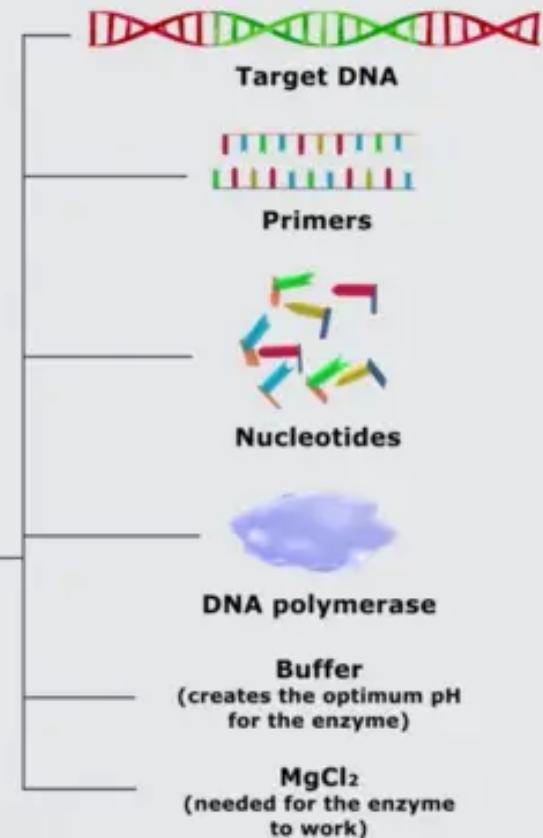


Indirect Methods

PCR Polymerase Chain Reaction

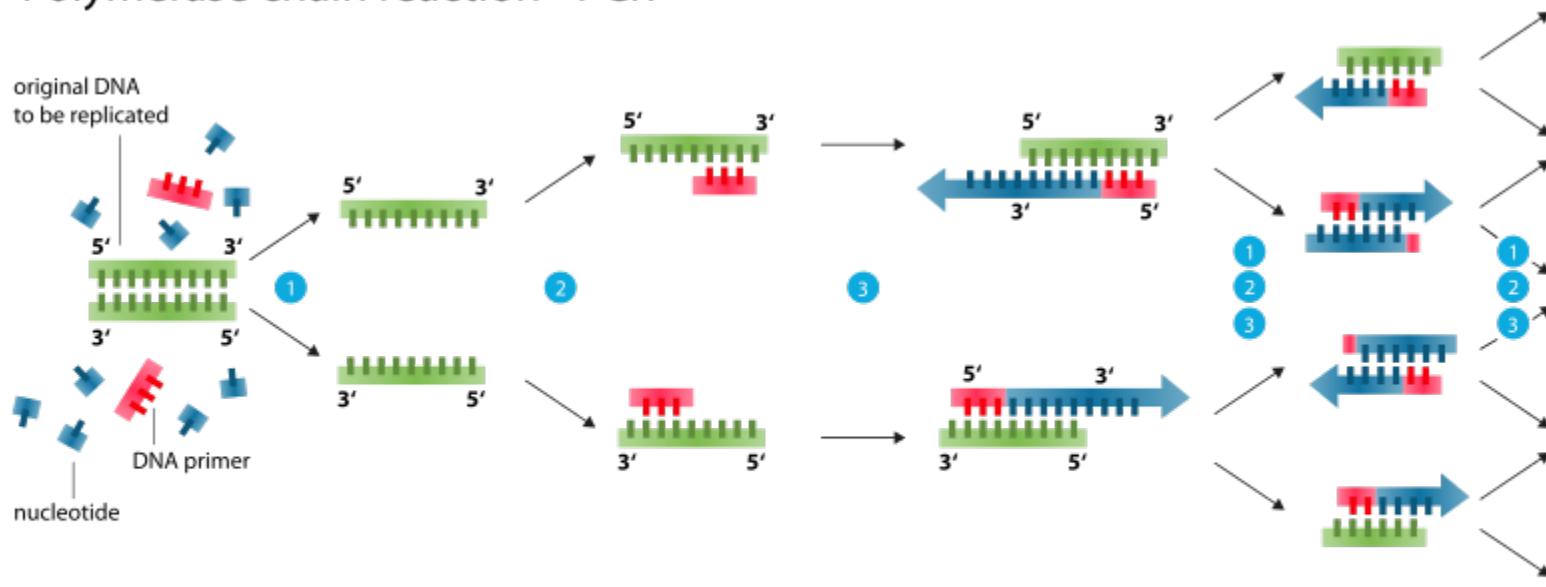


PCR Machine



Indirect Methods

Polymerase chain reaction - PCR



Polymerase Chain Reaction

Two functions:

1. Amplifying specific loci for further analysis
 - Amplification of specific genes
 - Amplification of particular loci with size differences for fragmentanalysis
2. PCR-based assays (selective amplification)
 - Usually done for identification

Polymerase Chain Reaction

PCR-based assays (selective amplification)



Volume 354, Issue 2

May 2014

Article Contents

Abstract

Introduction

Materials and methods

Detecting the presence of bacterial DNA by PCR can be useful in diagnosing culture-negative cases of infection, especially in patients with suspected infection and antibiotic therapy

Maria M. Lleo, Valentina Ghidini, Maria Carla Tafi, Francesco Castellani, Ilaria Trento, Marzia Boaretti Author Notes

FEMS Microbiology Letters, Volume 354, Issue 2, May 2014, Pages 153–160,
<https://doi.org/10.1111/1574-6968.12422>

Published: 01 May 2014 Article history ▾



PDF



Split View



Cite



Permissions



Share ▾

Abstract

Indirect Methods

Polymerase Chain Reaction

PCR-based assays (selective amplification)

Primers and target genes used to detect bacterial DNA in culture-negative clinical samples

Primer	Sequence	Amplified gene	Gene product	Amplicon size (in bp)	Primer specificity (detected bacteria)
BR1	GGACTACCAGGGTATCTAAT AGAGTTGATCCTGGCT	16S rRNA gene	Ribosomal DNA 16S	804	All
EntB	TGAATCACAAAGTGGTAAGCG TGGGGATGACGTCAAGTCAT	16S rRNA gene	Ribosomal DNA 16S	300	<i>Enterobacteriaceae</i>
Eco	ATCATGGAAGTAAGACTGC TTGCTGTGCCAGGCAGTTT	<i>uidA</i>	β -glucuronidase	356	<i>E. coli</i>
pbp5	CATGAGCAATTAAATCGG CATAGCCTGTCGCAAAAC	<i>pbp5</i>	Penicillin-binding protein 5	444	<i>E. faecalis</i>

Indirect Methods

Polymerase Chain Reaction in Diagnostics

Biotek-M™ Dengue aqua by the team of Dr. Raul Destura (UP Manila).



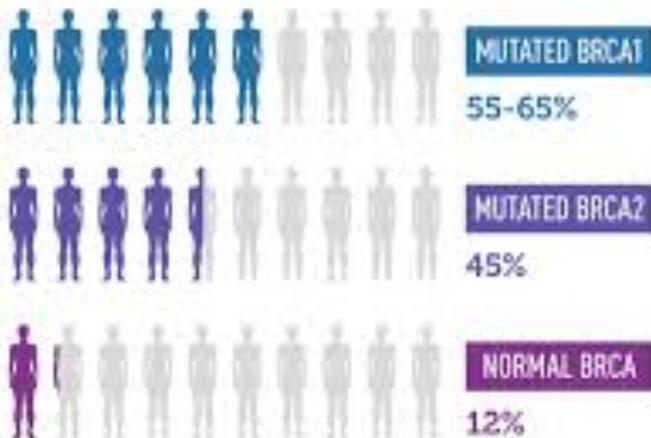
Indirect Methods

Polymerase Chain Reaction in Diagnostics

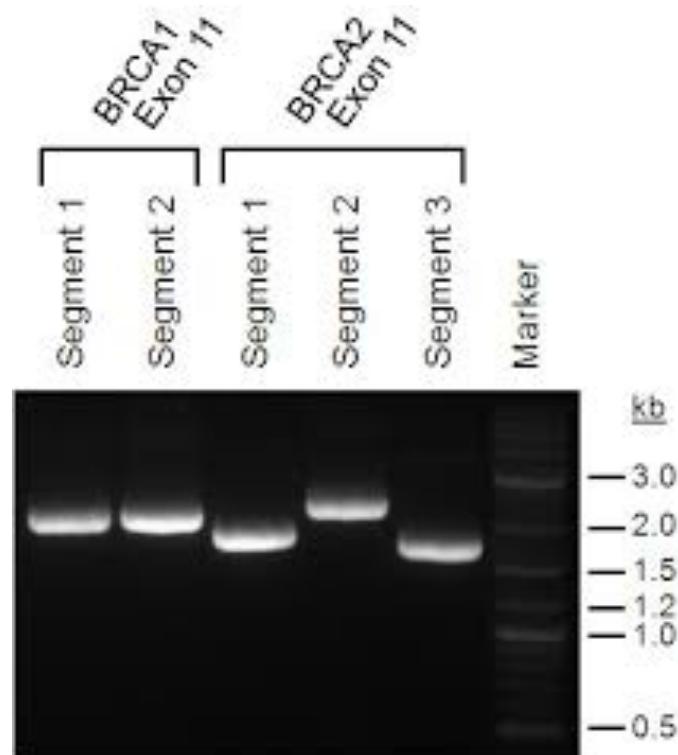
BRCA gene test – diagnostic test for familial Breast Cancer

NATIONAL CANCER INSTITUTE CHANCES OF DEVELOPING BREAST CANCER BY AGE 70

Specific inherited mutations in the BRCA1 and BRCA2 genes increase the risk of breast and ovarian cancers. Testing for these mutations is usually recommended in women without breast cancer only when the person's medical or family history suggests a possible presence of a heritable mutation in BRCA1 or BRCA2. Testing is often recommended in younger women mainly if diagnosed with breast cancer because it can influence treatment decisions and have implications for their family members.

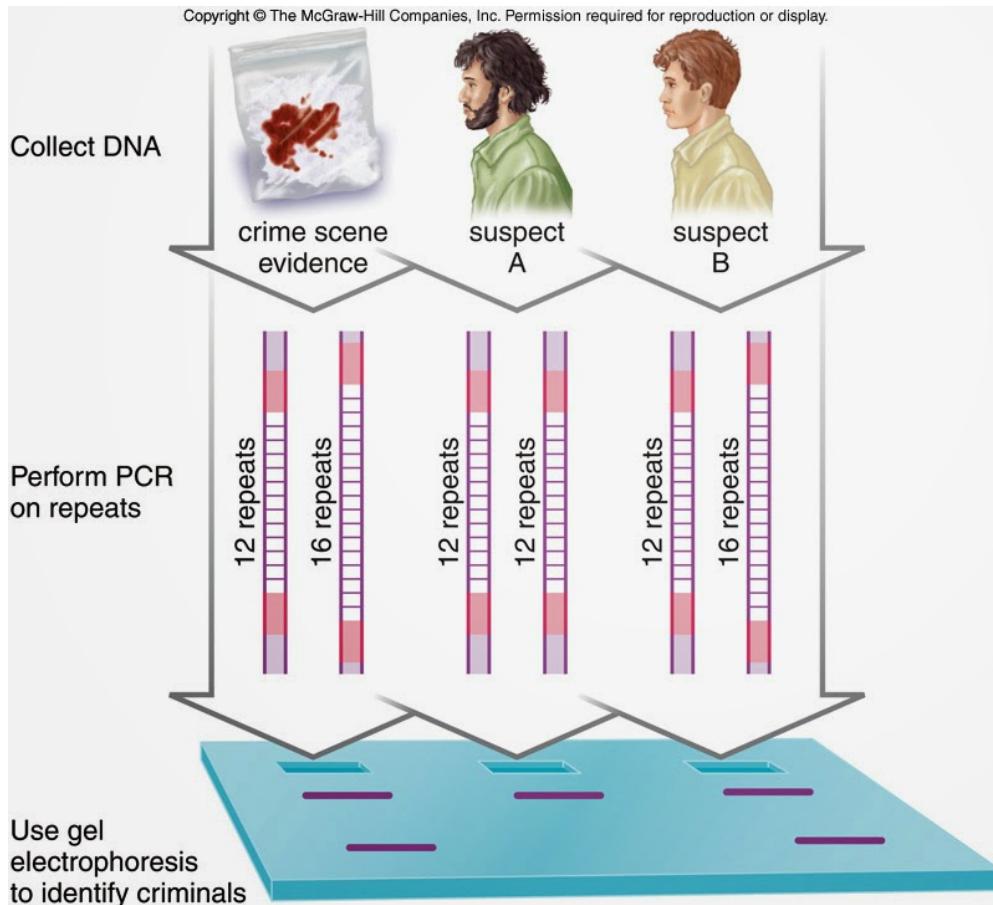


www.cancer.gov/breca-fact-sheet



Indirect Methods

Polymerase Chain Reaction in Forensics

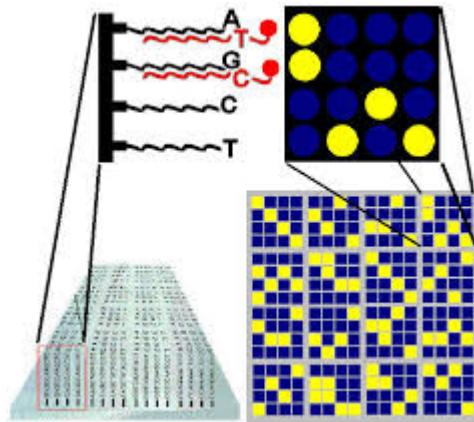


DNA sequences between individuals **are different**. Some regions have **repetitive sequences**.

The length of **repetitive sequences can differ between individuals**. By amplifying the repetitive sequences and **comparing their sizes**, we can infer which DNA sources match.

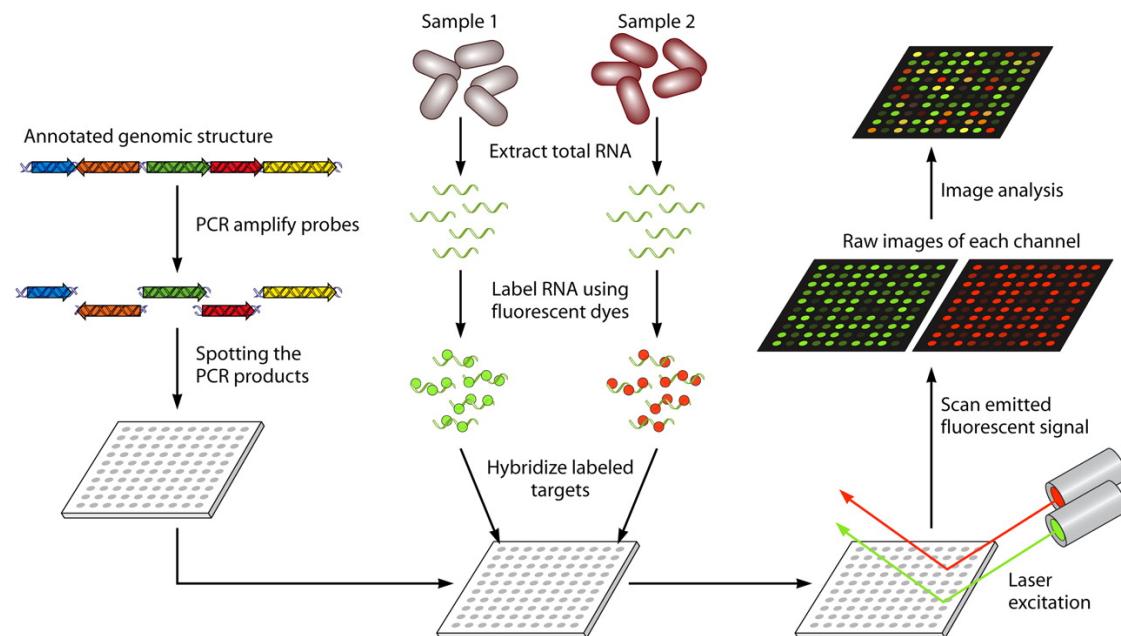
Indirect Methods

High-Throughput Hybridization: Microarrays



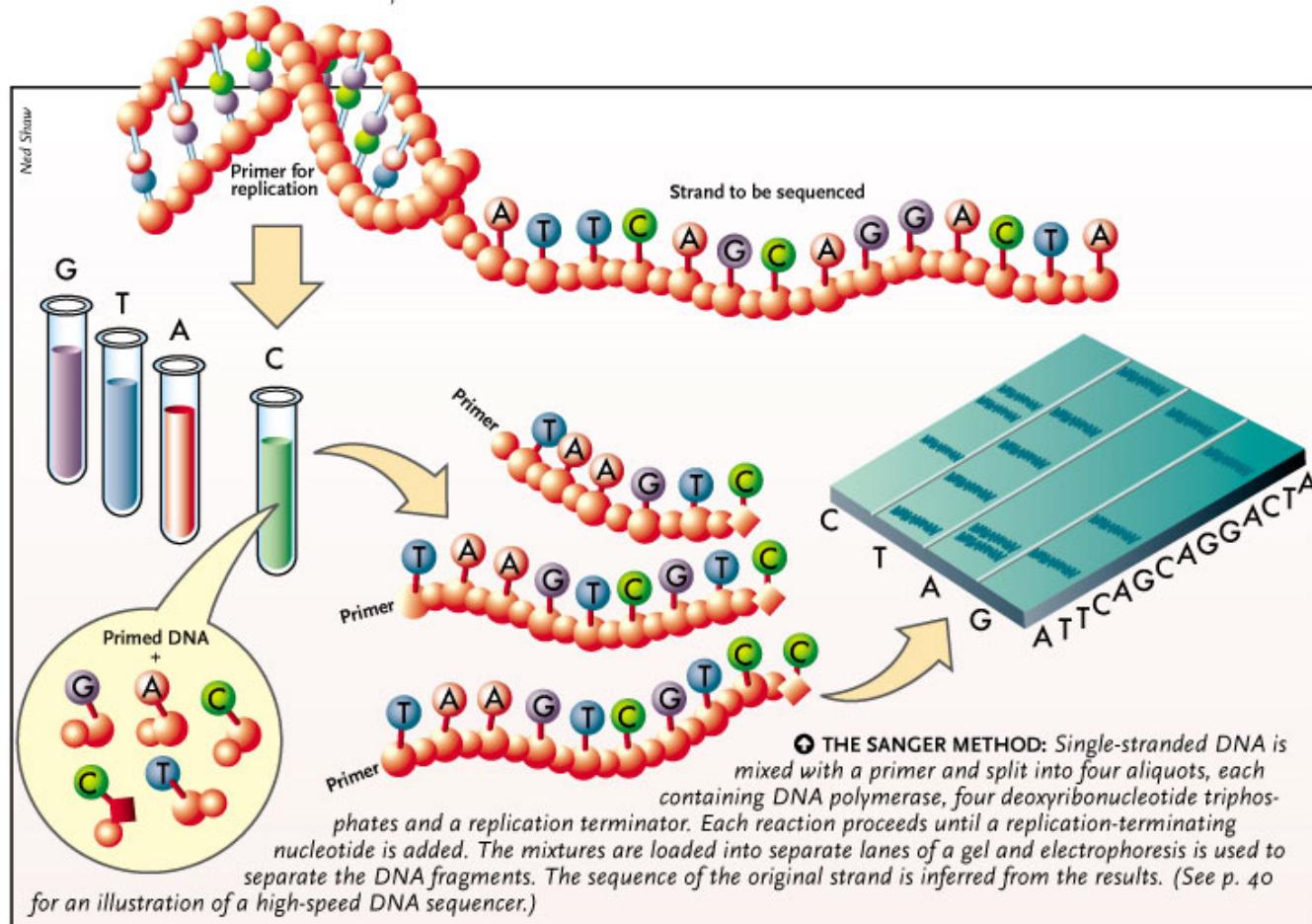
- Microarrays are chips with multiple unique probes.
- Each probe is specific to a certain region of DNA.
- Depending on the presence or absence of the DNA sequences, the chip changes color.

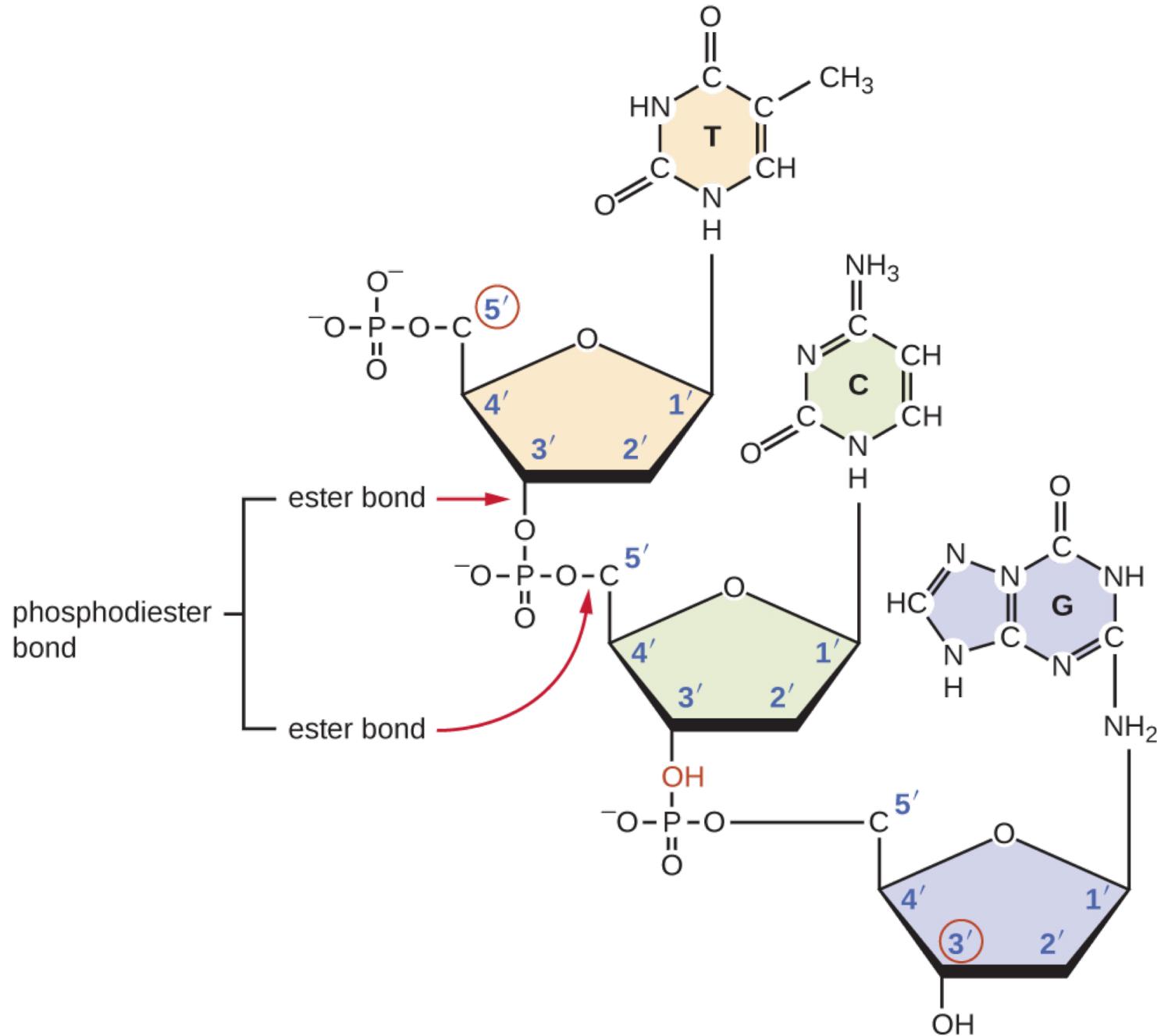
- Microarrays can be used to compare DNA quantities



Direct Methods

DNA Sequencing: Sanger Method (1977)



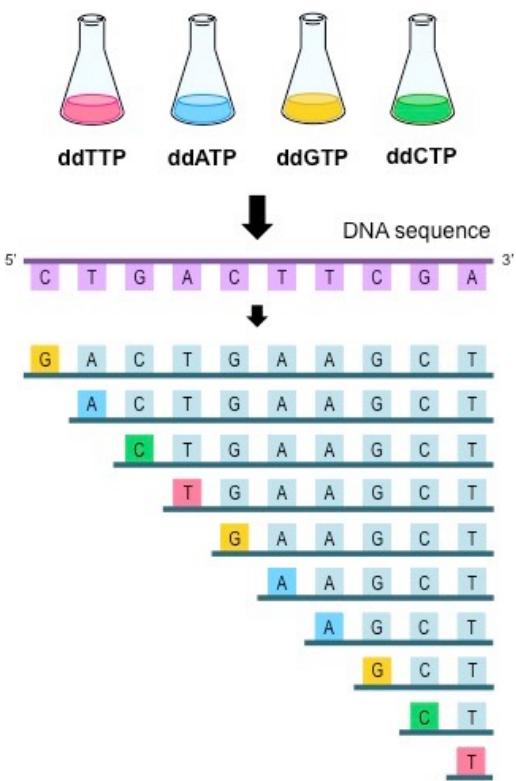


Direct Methods

Sanger Sequencing

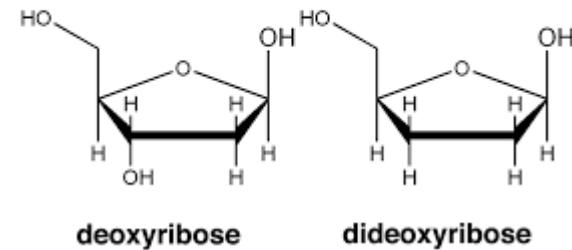
Automated

4 x PCR (+ one dideoxynucleotide)



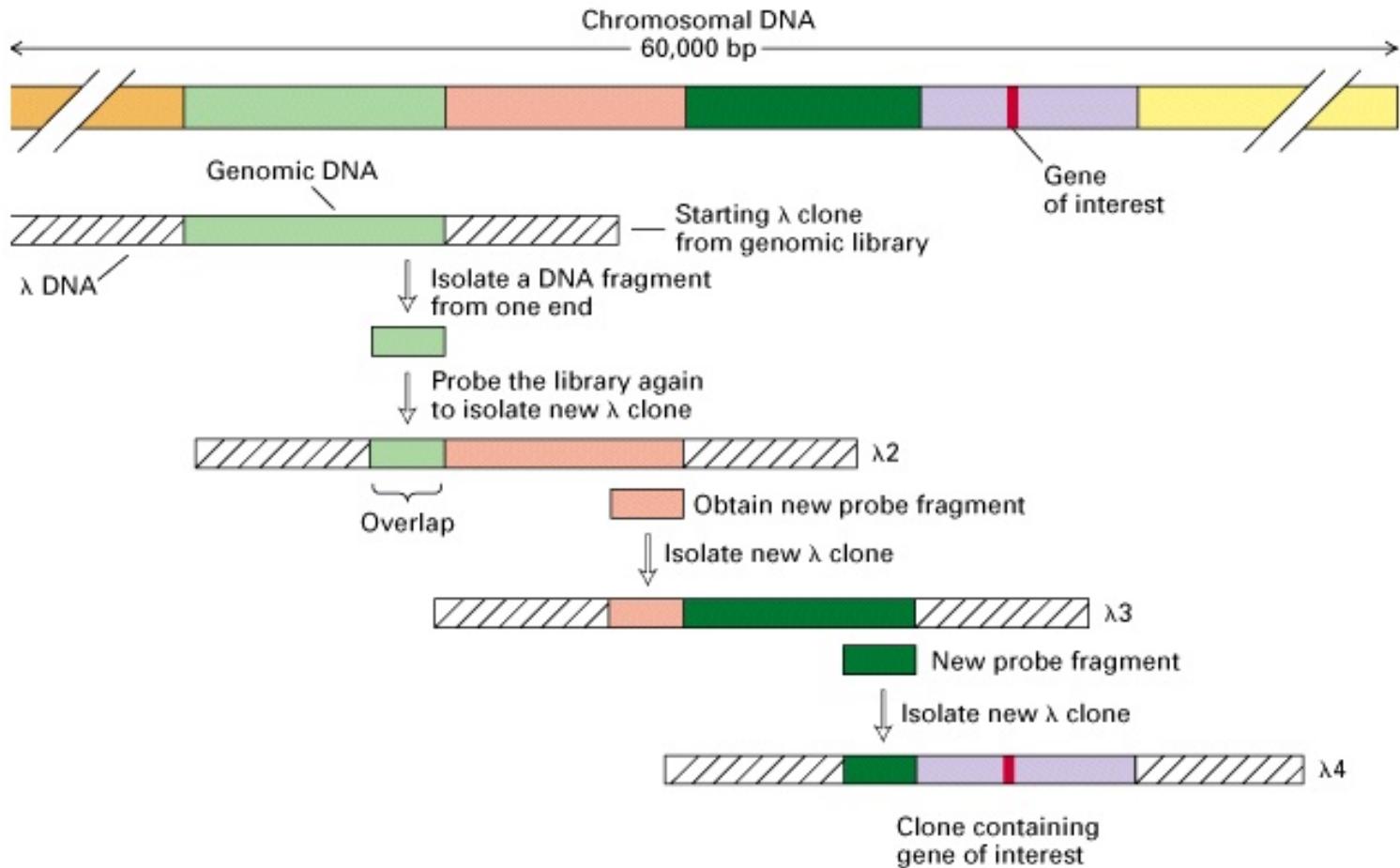
Dye-terminator chemistry
Fluorescent-labelled ddNTPs

single strand DNA template
+ primer
+ dNTPs (dATP, dTTP, dGTP, dCTP)
+ polymerase
+ labelled ddNTPs (limiting amounts)



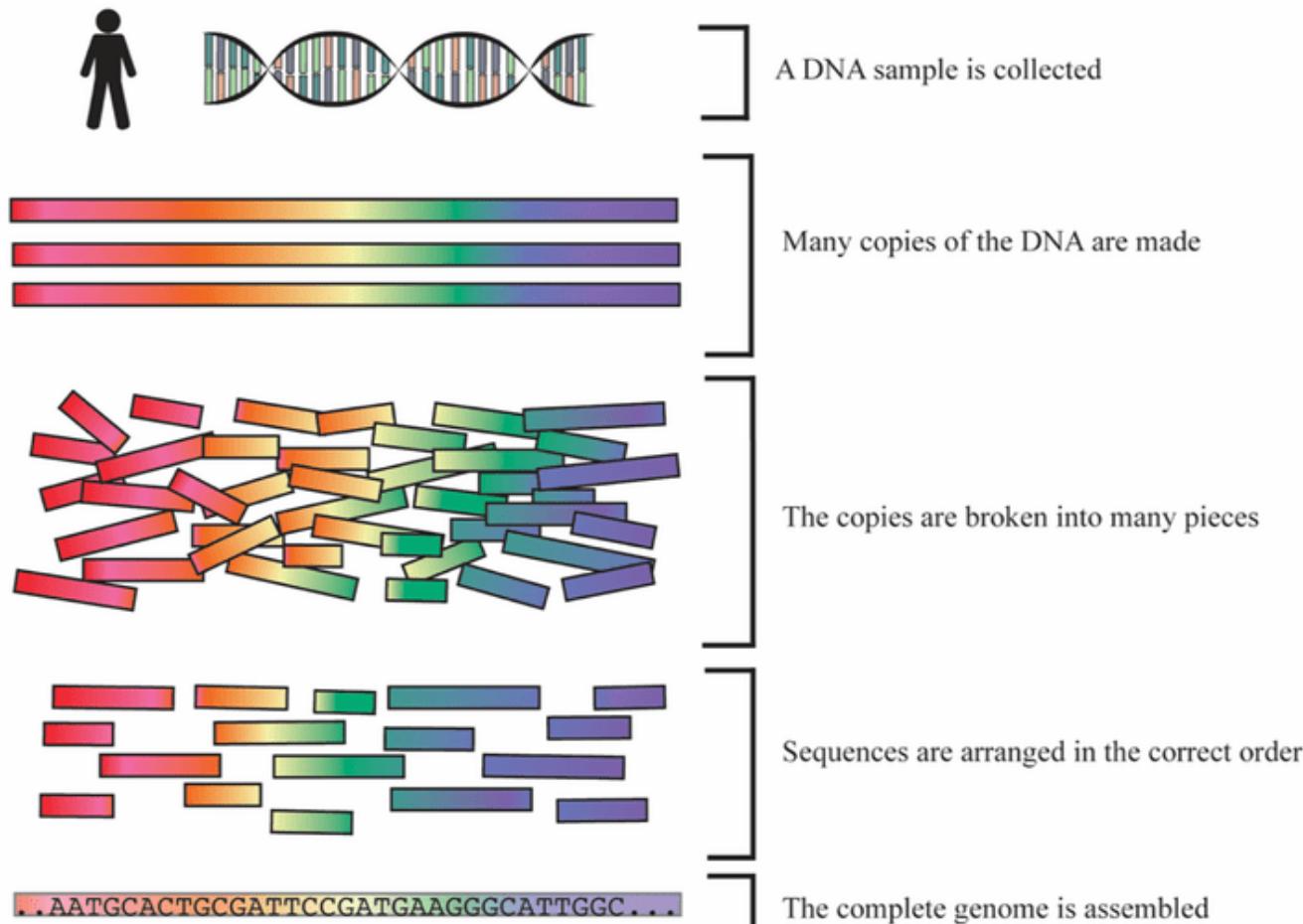
Direct Methods

Sequencing the Genome: Chromosome Walking



Direct Methods

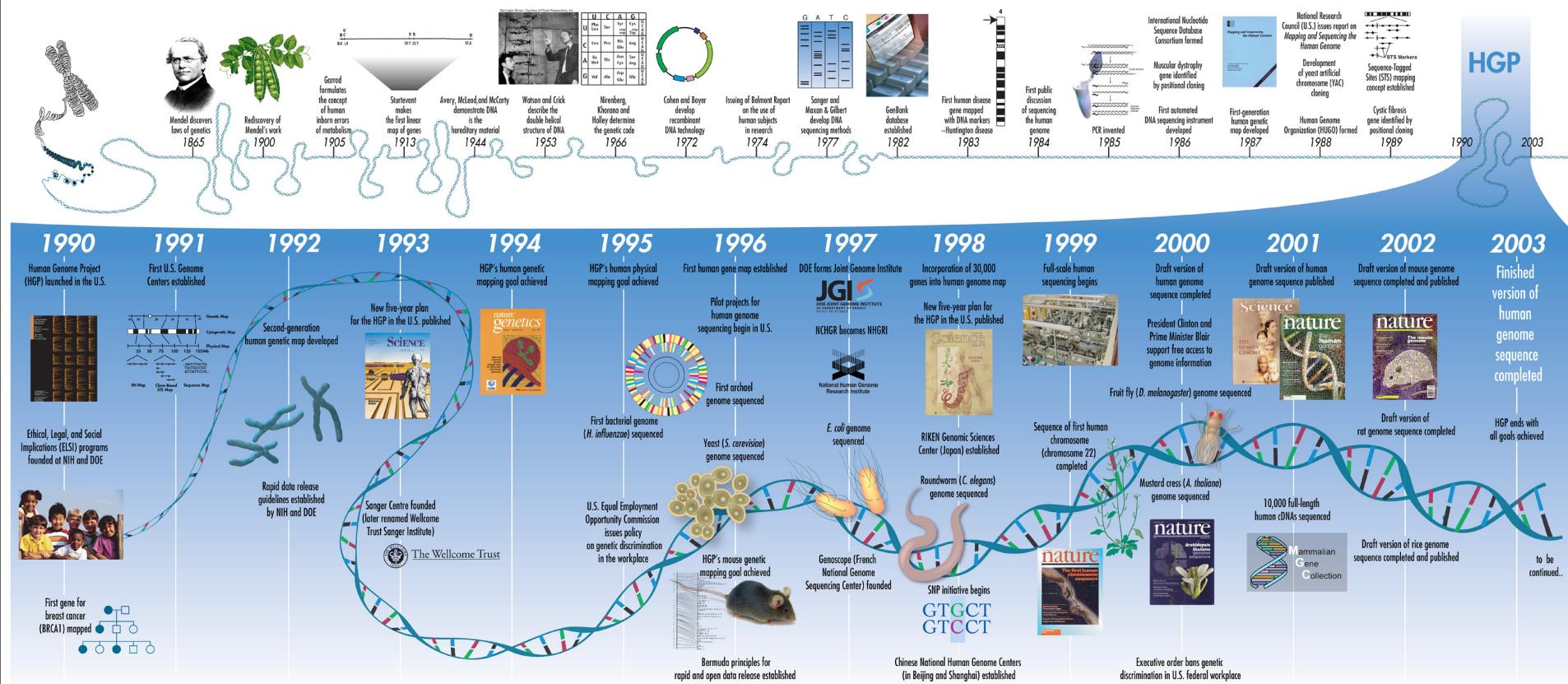
Sequencing the Genome: Shotgun Sequencing



Direct Methods

The Human Genome Project

\$3.8B Investment in Human Genome Project Drove \$796B in Economic Impact Creating 310,000 Jobs and Launching the Genomic Revolution

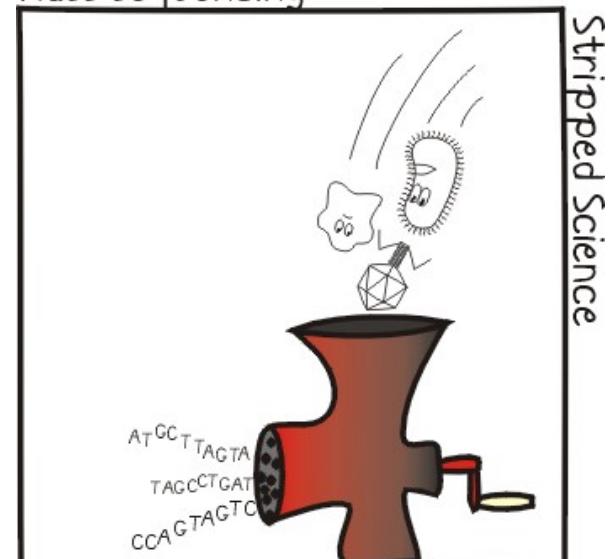


Next-generation DNA Sequencing:

Massively Parallel, High-Throughput Sequencing

- Based on the concept of shotgun sequencing
- High throughput due to massive parallelization (10^6 sequencing reactions run simultaneously)
- Short reads (35 - 400 bp) for assembly
- Typical yields per run: 0.5 - 4 Gigabases
- Different platforms
 - 454 pyrosequencing
 - Illumina (Solexa)
 - SOLid sequencing
 - Ion torrent

Mass sequencing

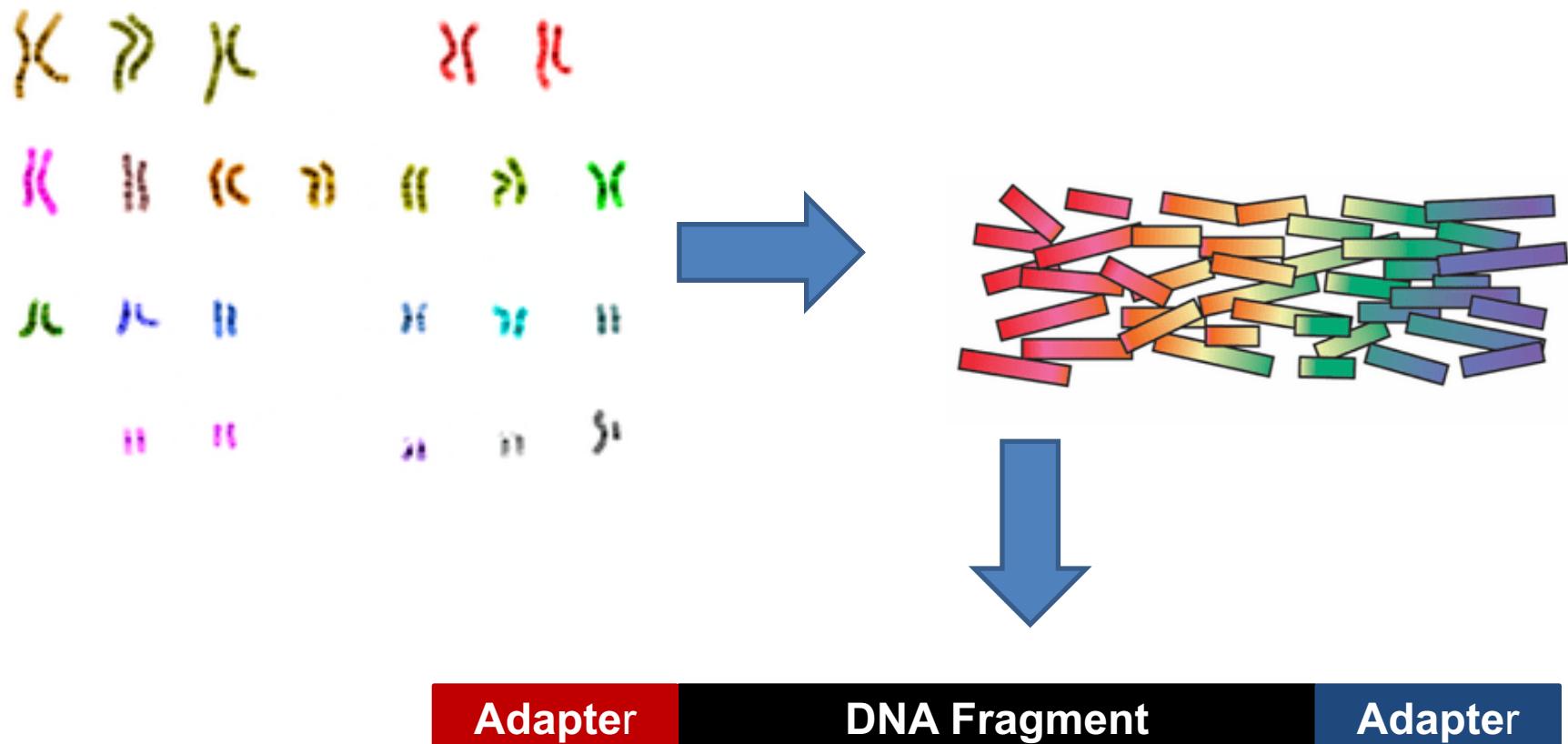


by Viktor S. Poór

Direct Methods

Next-generation DNA Sequencing:

Massively Parallel, High-Throughput Sequencing



Next-generation DNA Sequencing:

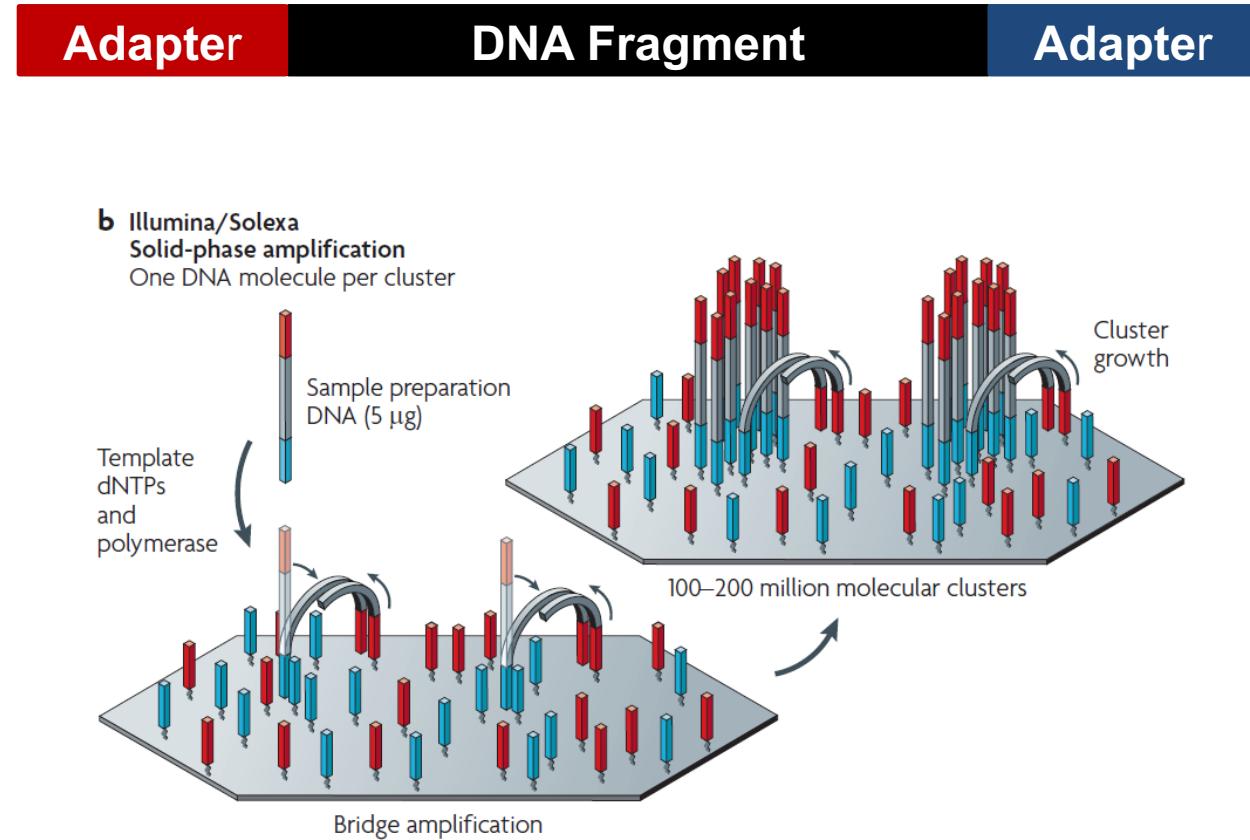
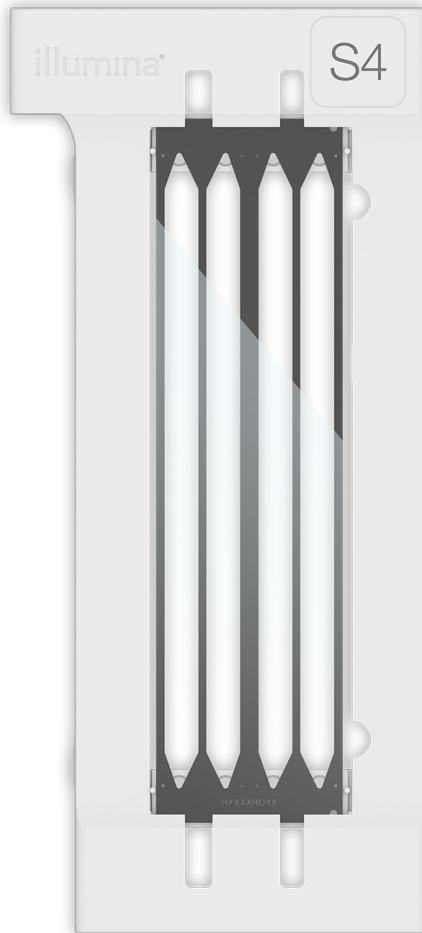
Massively Parallel, High-Throughput Sequencing



- Sequencing Adapters contain the following:
 - An **index**, that allows for multiplexing
 - An **oligo** that hybridizes with the flow cell
 - A **primer binding site** to start sequencing by synthesis

Direct Methods

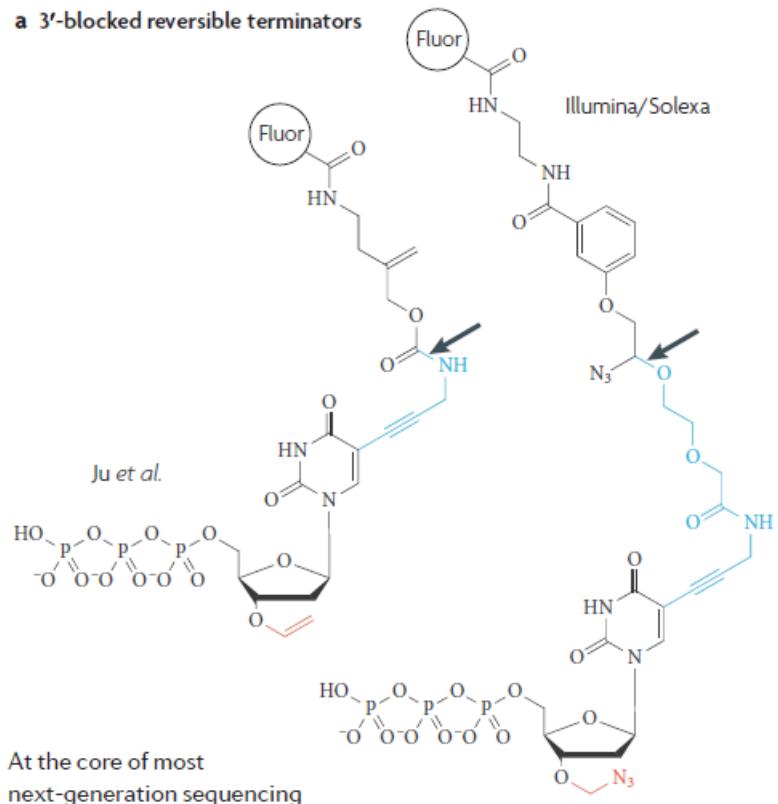
Next-generation DNA Sequencing: Bridge Amplification (Illumina)



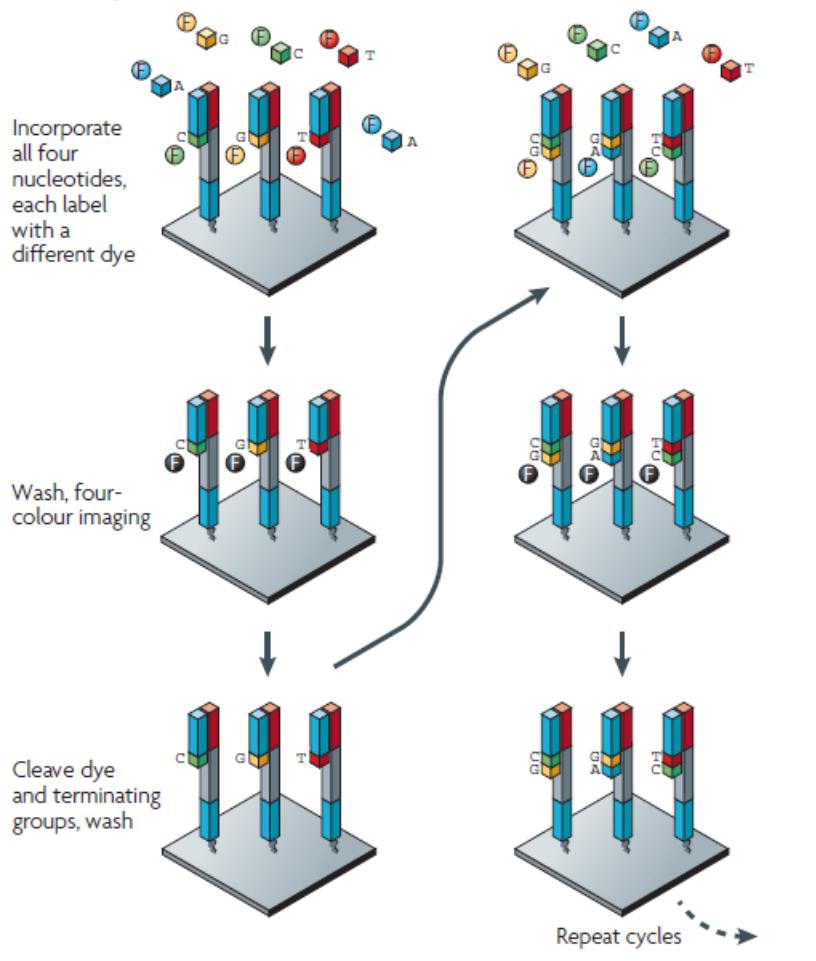
Direct Methods

Next-generation DNA Sequencing: Reversible Termination

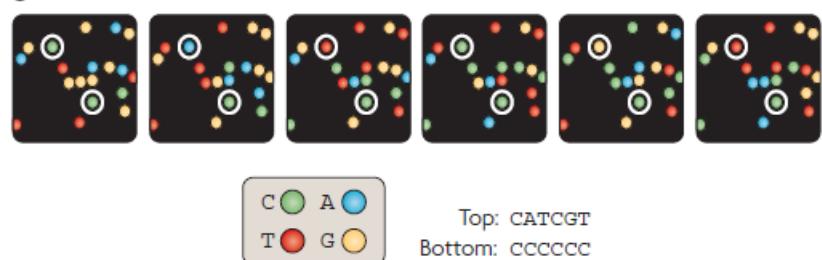
a 3'-blocked reversible terminators



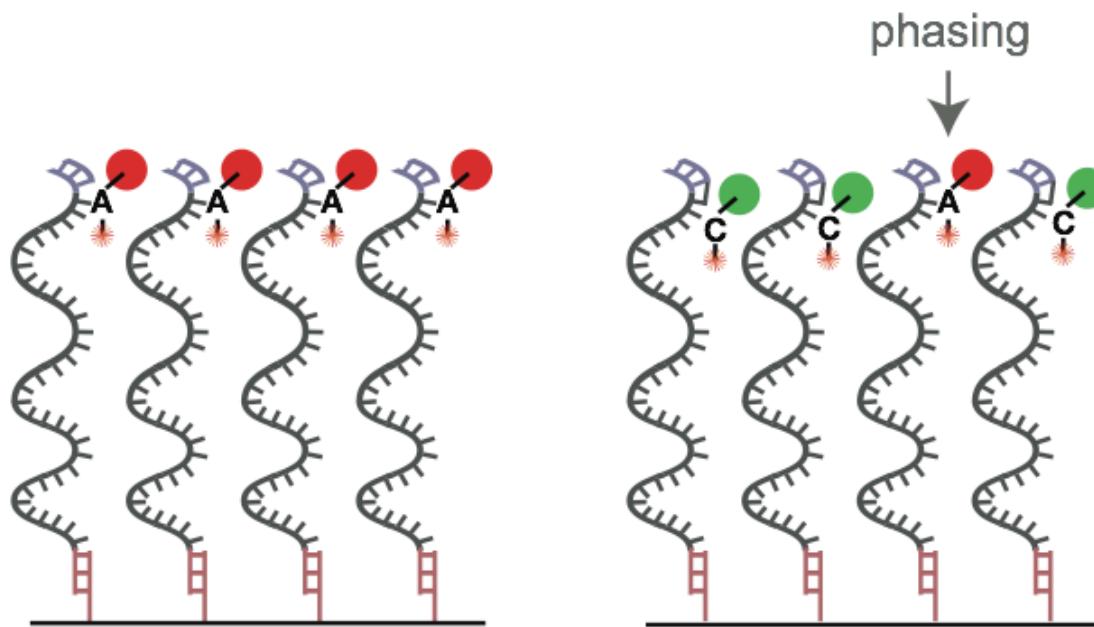
a Illumina/Solexa — Reversible terminators



b



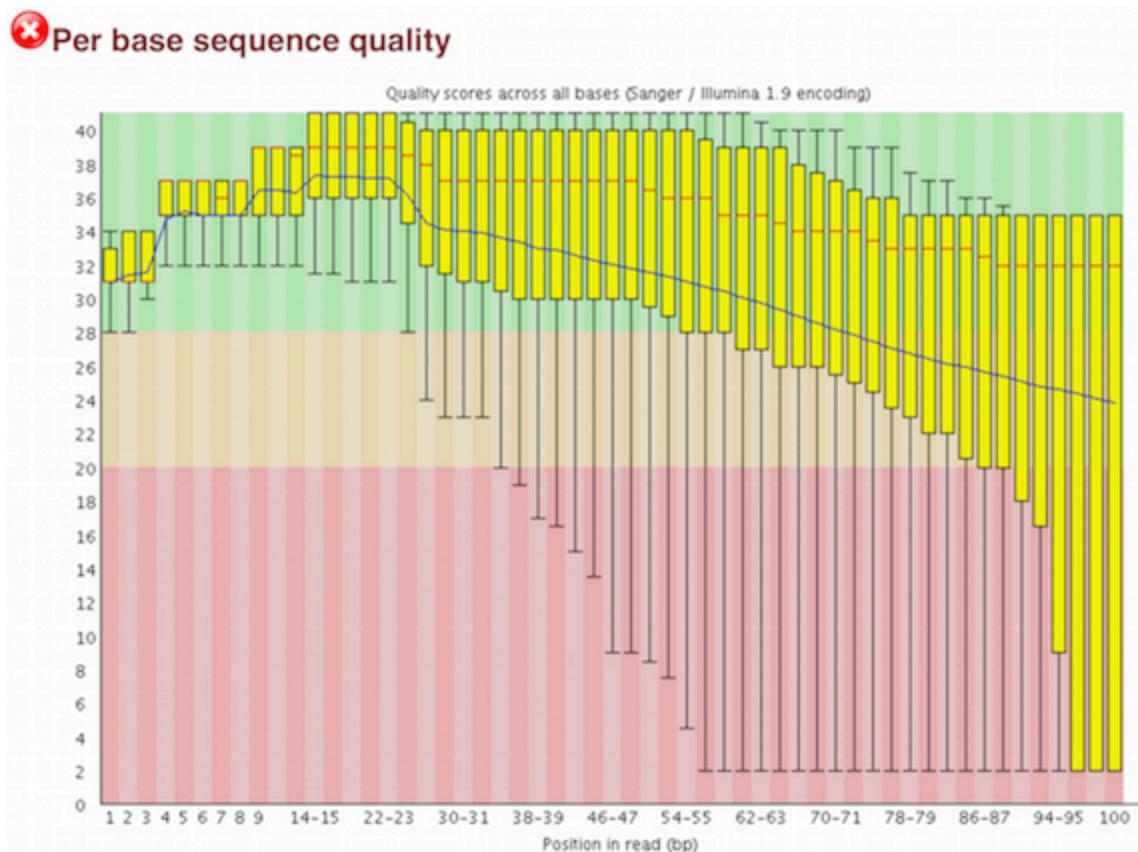
Next-generation DNA Sequencing: Phasing



- As sequencing proceeds, the cluster slowly goes out of phase (as some fragments are synthesized slower than others).
- Thus, the quality of the fluorescent signals degrade as read length increases

Direct Methods

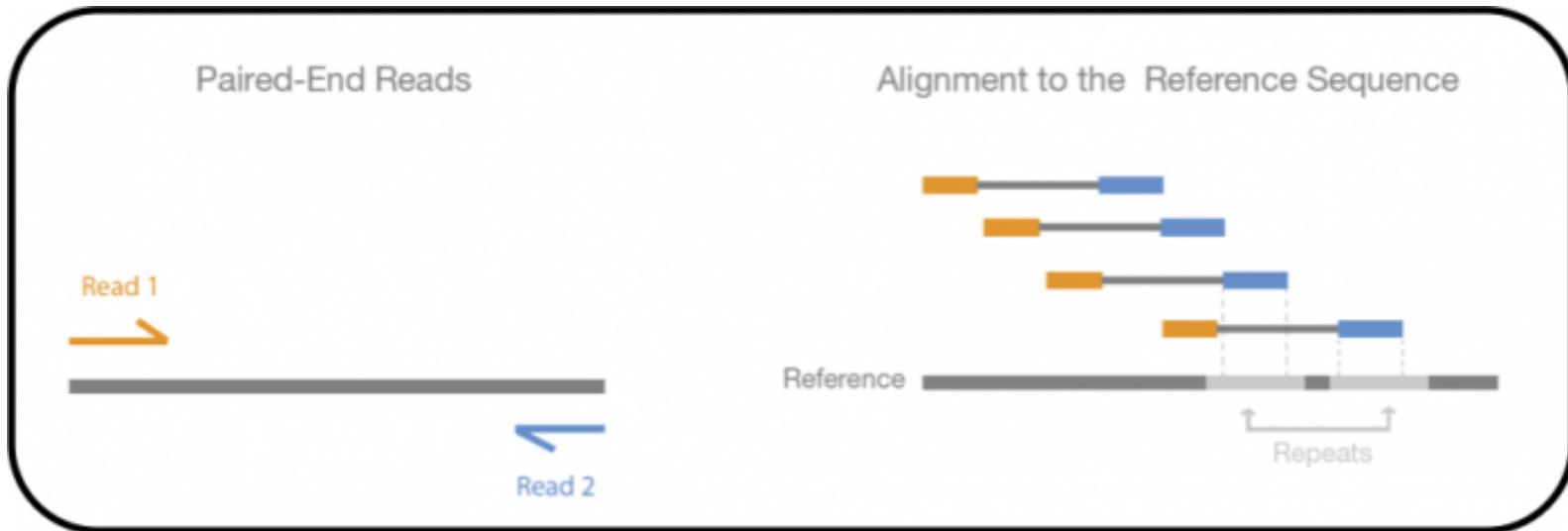
Next-generation DNA Sequencing: Phasing



- As sequencing proceeds, the cluster slowly goes out of phase (as some fragments are synthesized slower than others).
- Thus, the quality of the fluorescent signals degrade as read length increases

Direct Methods

Next-generation DNA Sequencing: Single- vs Paired-End Sequencing

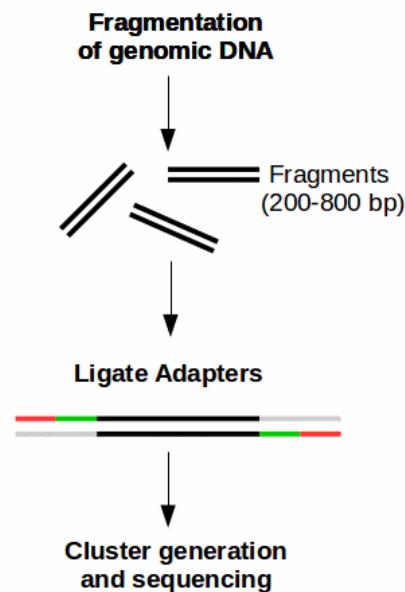


- For the same amount of DNA fragments, paired-end sequencing allows us to generate more sequence and structural information

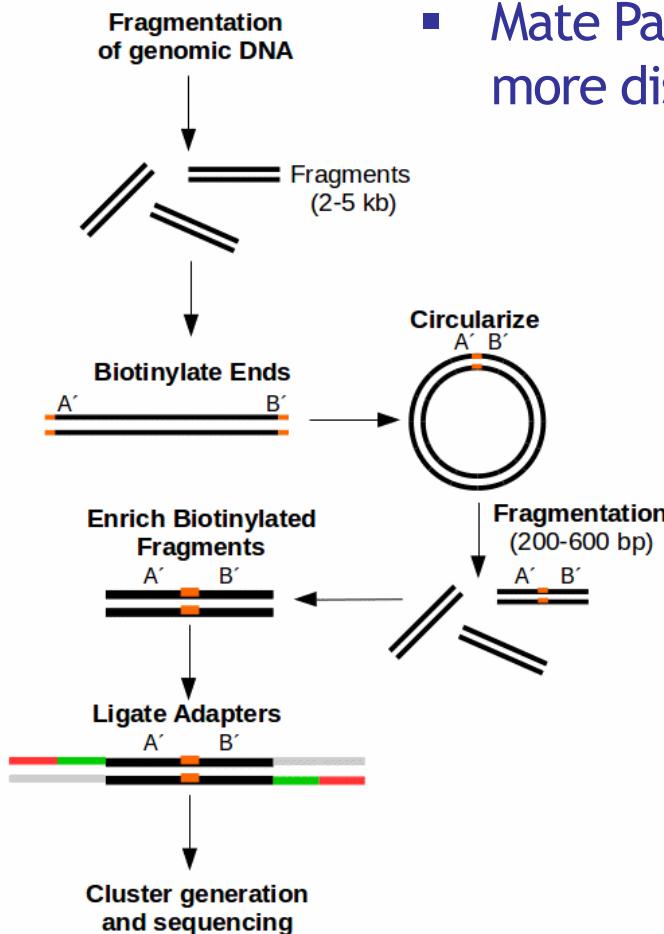
Direct Methods

Next-generation DNA Sequencing: Paired-End vs Mate Pair Sequencing

Paired-End Sequencing (Short-insert paired-end reads)



Mate Pair Sequencing



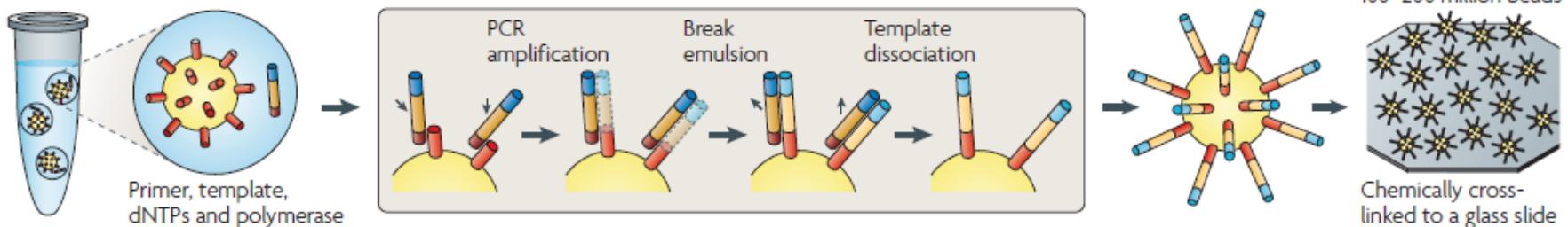
- Mate Pair sequencing allows for more distant paired-ends.

Direct Methods

Next-generation DNA Sequencing: Emulsion Amplification (Roche, Life)

a Roche/454, Life/APG, Polonator Emulsion PCR

One DNA molecule per bead. Clonal amplification to thousands of copies occurs in microreactors in an emulsion



Adapter

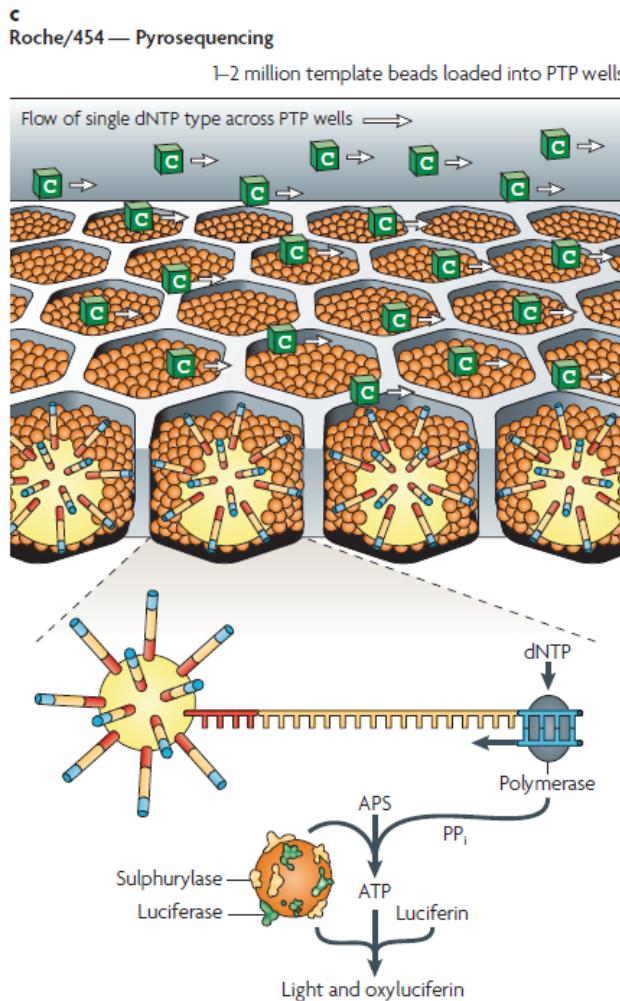
DNA Fragment

Adapter

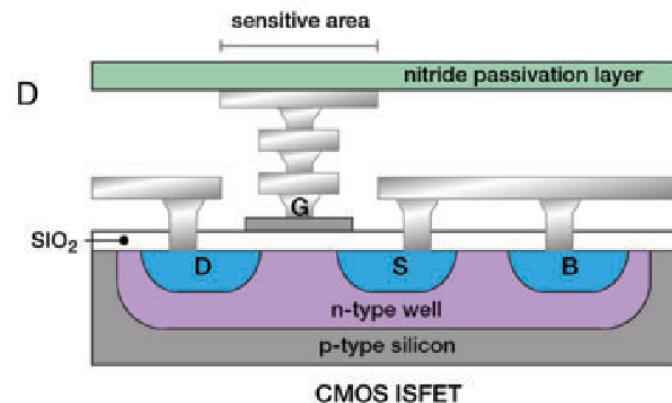
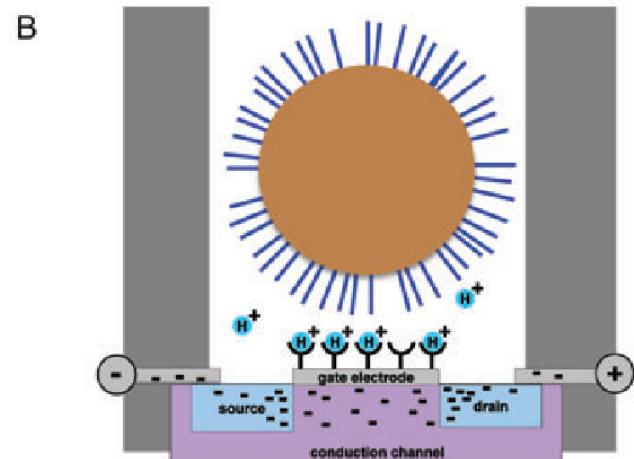
Direct Methods

Next-generation DNA Sequencing: Pyrosequencing and Ion Semiconductor Sequencing

Pyrosequencing



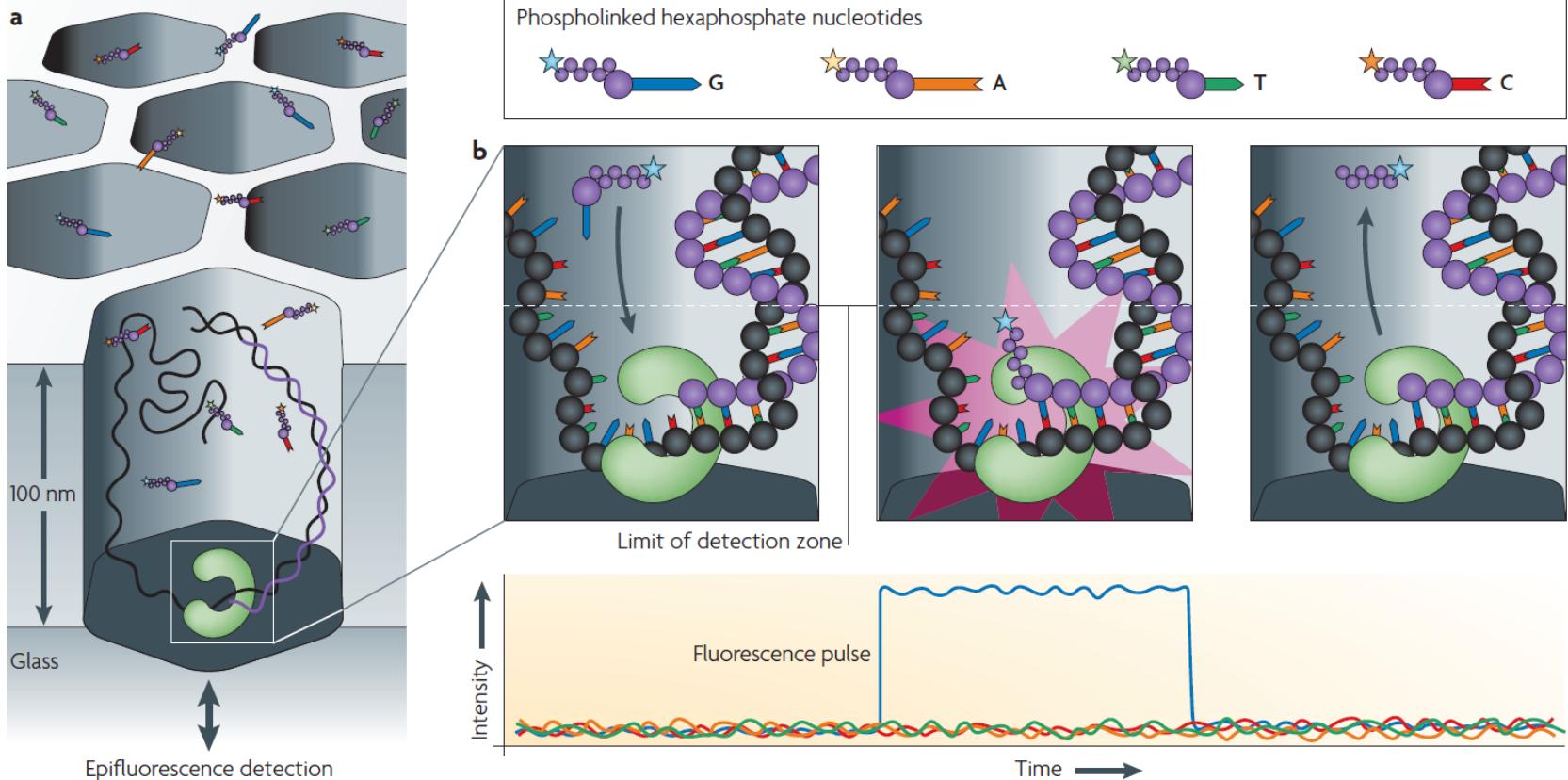
Ion Semiconductor Sequencing



Direct Methods

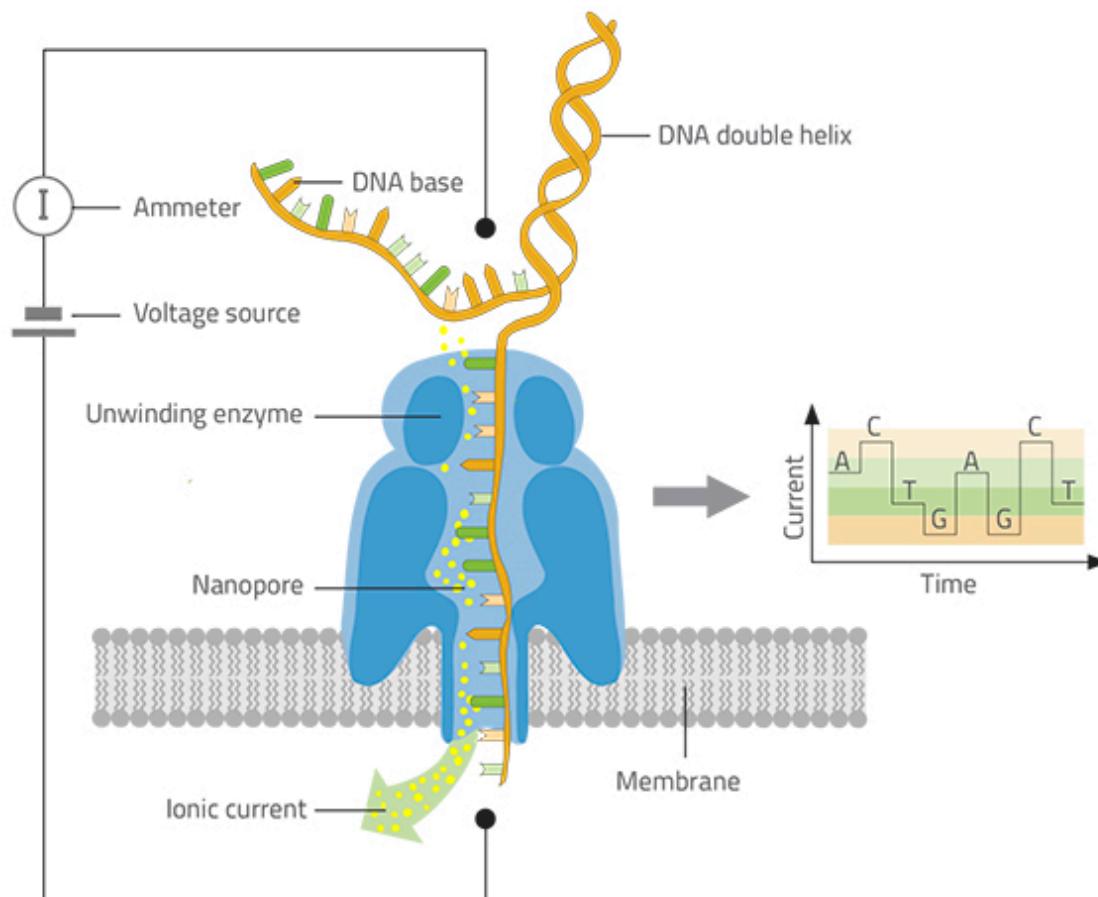
Next-generation DNA Sequencing: Real-time Sequencing (Pacific Biosciences)

Pacific Biosciences — Real-time sequencing



Direct Methods

Next-generation DNA Sequencing: Nanopore Sequencing (Oxford Nanopore)



Sequencing James Watson

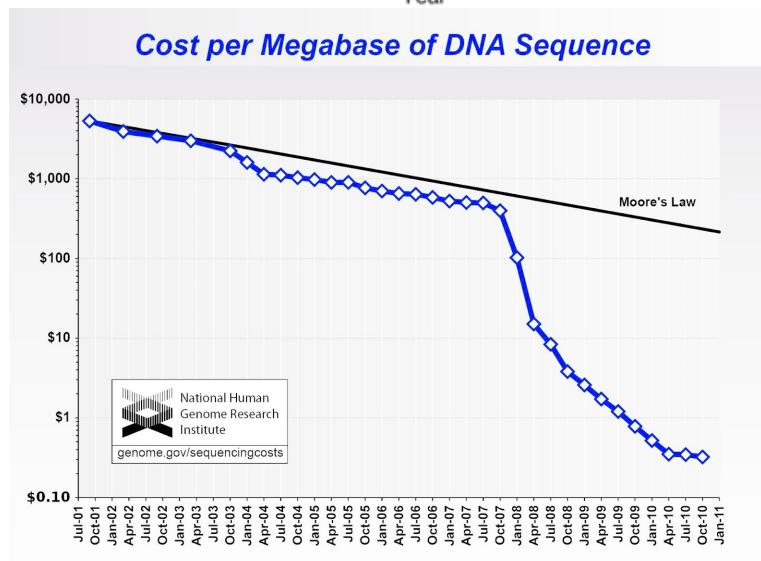
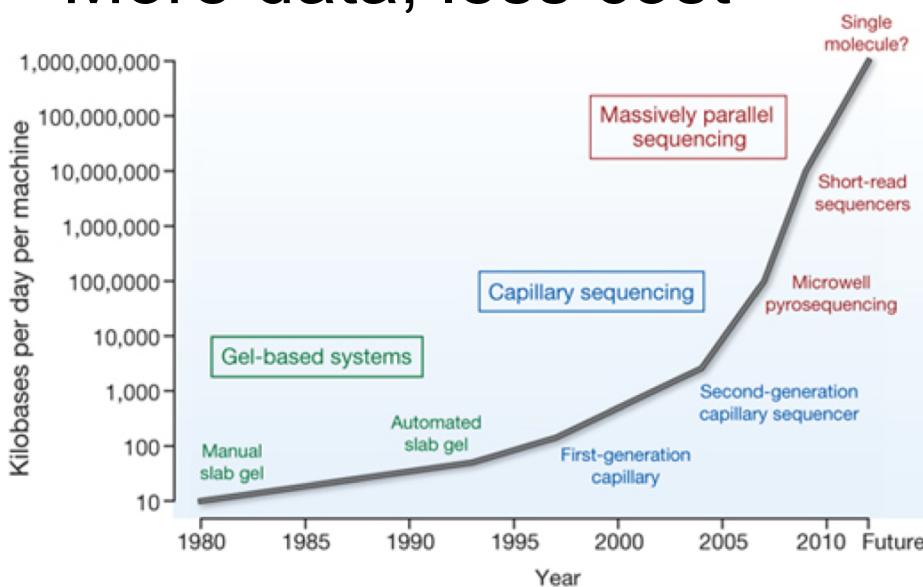
Human genome size = 3×10^9 bp = 3 Gb

- First whole human genome to be sequenced with next-generation technology
 - 24.5 Billion bases of genomic DNA sequence generated at the 454 Sequencing Center
 - 3.6 Million variants detected, including several disease susceptibility gene associations

Jim Watson	Human Genome Project
454 Life Sciences, A Roche Company	Sanger
2 months, 3 instruments	10-13 years
<\$1 million \$250,000 with Titanium	\$100 million - \$2.7 billion
7.4x coverage	7.5x coverage
250 bp read length 400bp with Titanium	500-800 bp read length

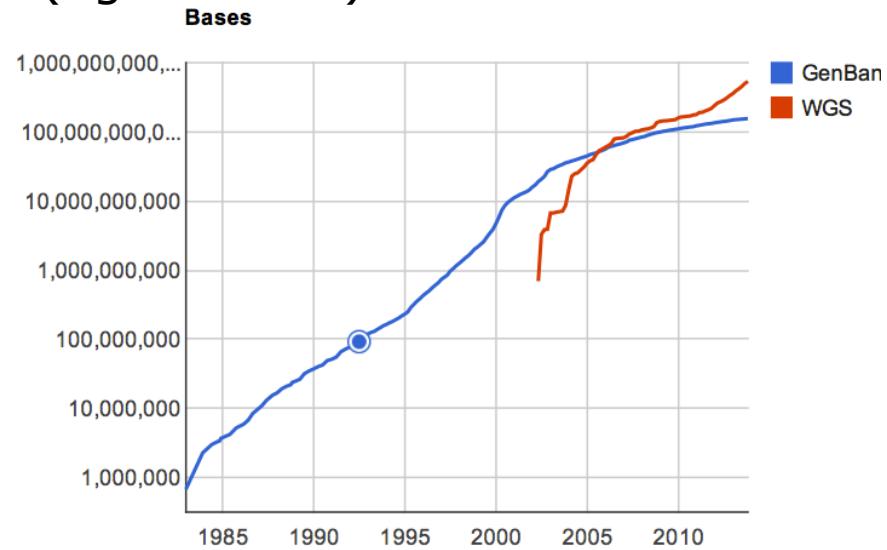
James Watson decoded.

Increasing capabilities in DNA sequencing: More data, less cost



<http://genome.wellcome.ac.uk/>

A wealth of information in public domain databases
(e.g. GenBank)



<http://www.ncbi.nlm.nih.gov/genbank/statistics>