



# Enzyme-based molecular techniques (part I)

## DNA sequencing

Prof. Mamoun Ahram  
Second semester 2021



# Resources



- This lecture
- Cooper, pp. 127-129, 124-125, Ch. 5, pp.159-162, 166-171



# What is DNA sequencing?



- DNA sequencing is the process of determining the exact order of nucleotides in a genome.
- Importance:
  - Identification of genes and their localization
  - Identification of protein structure and function
  - Identification of DNA mutations
  - Genetic variations among individuals in health and disease
  - Prediction of disease-susceptibility and treatment efficiency
  - Evolutionary conservation among organisms



# DNA sequencing of organism genome



- Viruses and prokaryotes first
- Human mitochondrial DNA
- The first eukaryotic genome sequenced was that of yeast, *Saccharomyces cerevisiae*.
- The genome of a multicellular organism, the nematode *Caenorhabditis elegans*.
- Determination of the base sequence in the human genome was initiated in 1990 and completed in May 2006 via the Human Genome Project





SPECIES	BASE PAIRS (estimated)	GENES (estimated)	CHROMOSOMES
<b>Human</b> ( <i>Homo sapiens</i> )	3.2 billion	~ 25,000	46
<b>Mouse</b> ( <i>Mus musculus</i> )	2.6 billion	~ 25,000	40
<b>Fruit Fly</b> ( <i>Drosophila melanogaster</i> )	137 million	13,000	8
<b>Roundworm</b> ( <i>Caenorhabditis elegans</i> )	97 million	19,000	12
<b>Yeast</b> ( <i>Saccharomyces cerevisia</i> )	12.1 million	6,000	32
<b>Bacteria</b> ( <i>Escherichia coli</i> )	4.6 million	3,200	1
<b>Bacteria</b> ( <i>H. influenzae</i> )	1.8 million	1,700	1



# Nucleotides per genomes

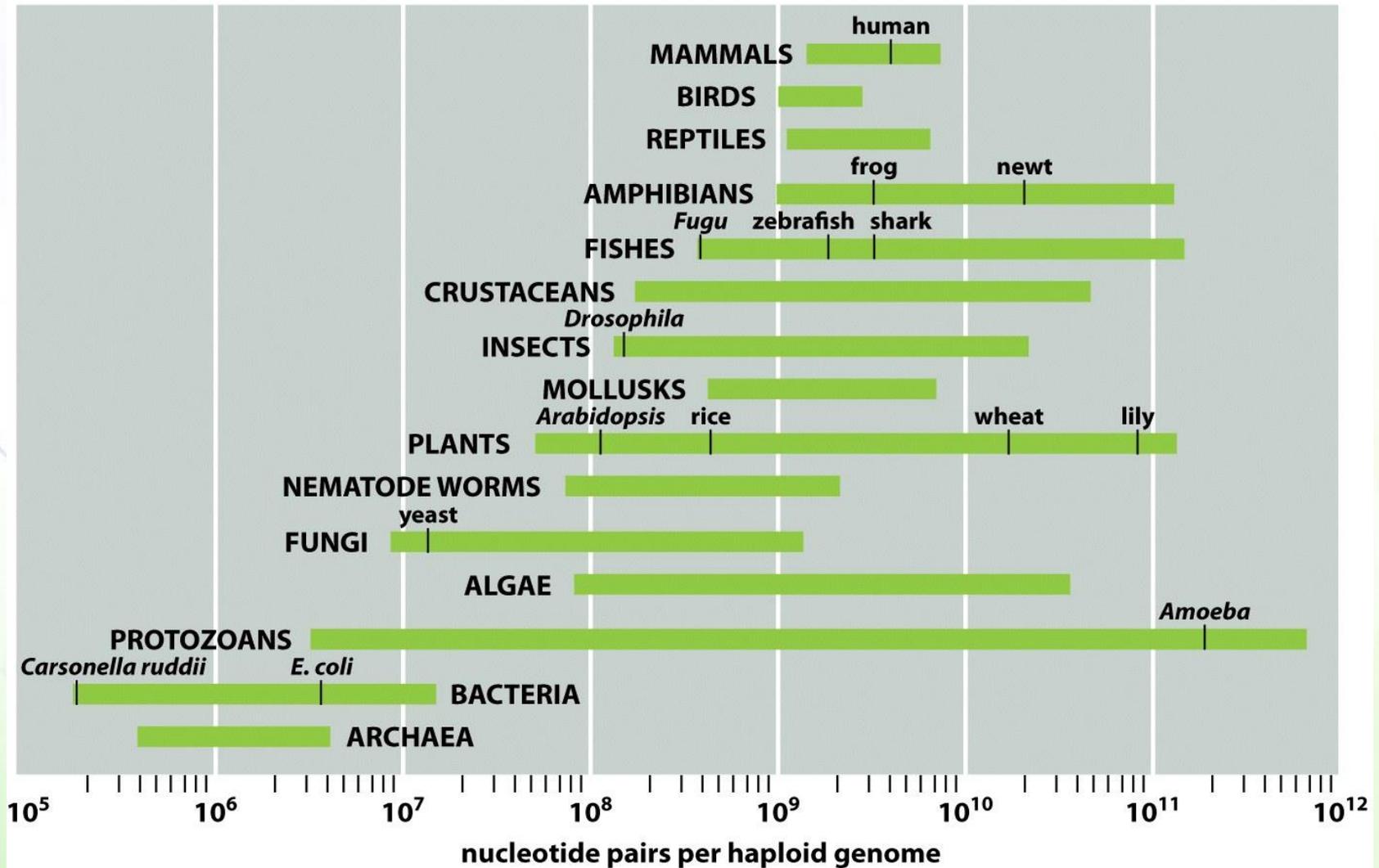


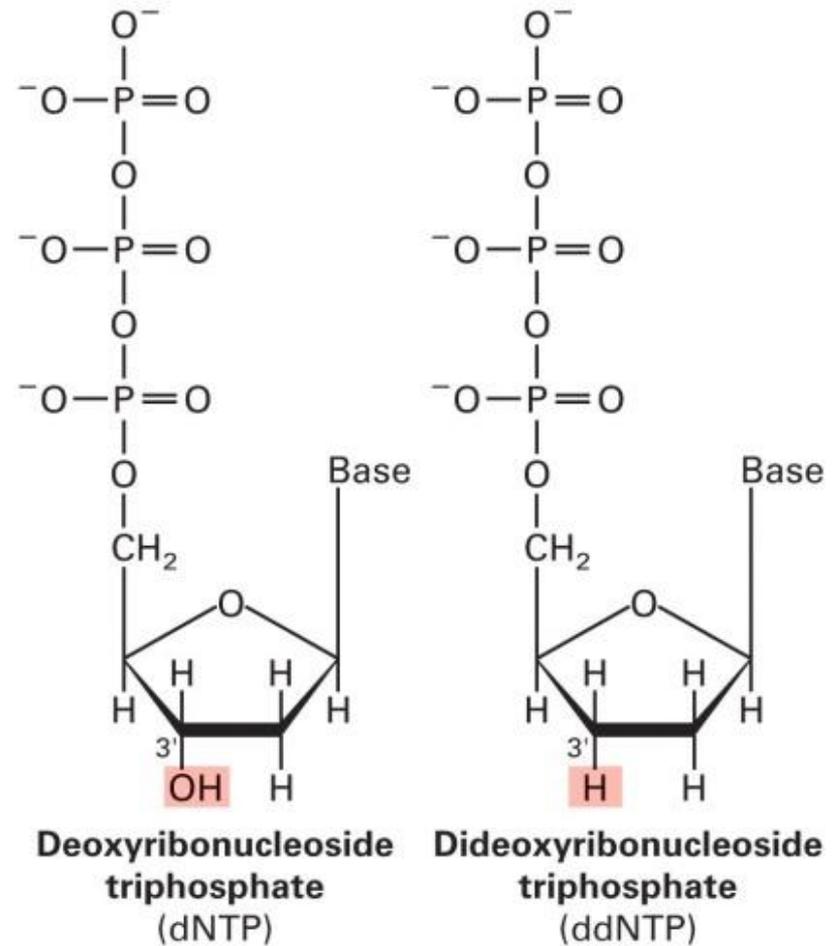
Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)



# Method of DNA sequencing



- The most popular method is based on premature termination of DNA synthesis by dideoxynucleotides.



# The process...



- DNA synthesis is initiated from a primer that has been labeled with a radioisotope
- Four separate reactions are run, each including deoxynucleotides plus one dideoxynucleotide (either A, C, G, or T)
- Incorporation of a dideoxynucleotide stops further DNA synthesis because no 3 hydroxyl group is available for addition of the next nucleotide



# Generation of fragments



- A series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction
- These fragments of DNA are then separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film
- The size of each fragment is determined by its terminal dideoxynucleotide, so the DNA sequence corresponds to the order of fragments read from the gel





5' TAGCTGACTC 3'  
3' ATCGACTGAGTCAAGAACTATTGGGCTTAA ...



DNA polymerase  
+ dATP, dGTP, dCTP, dTTP  
+ **ddGTP** in low concentration

5' TAGCTGACTCA**G** 3'  
3' ATCGACTGAGTCAAGAACTATTGGGCTTAA ...

+

5' TAGCTGACTCAGTTCTT**G** 3'  
3' ATCGACTGAGTCAAGAACTATTGGGCTTAA ...

+

5' TAGCTGACTCAGTTCTT**G**AACCC**G** 3'  
3' ATCGACTGAGTCAAGAACTATTGGGCTTAA ...

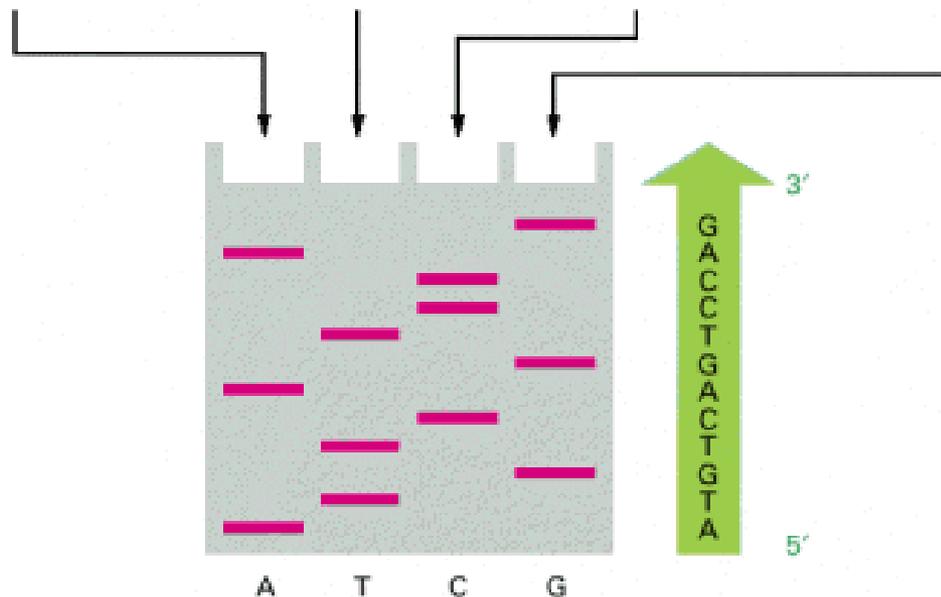
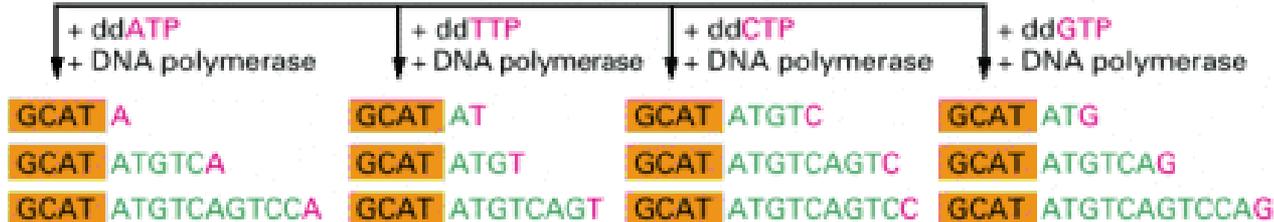




(C)

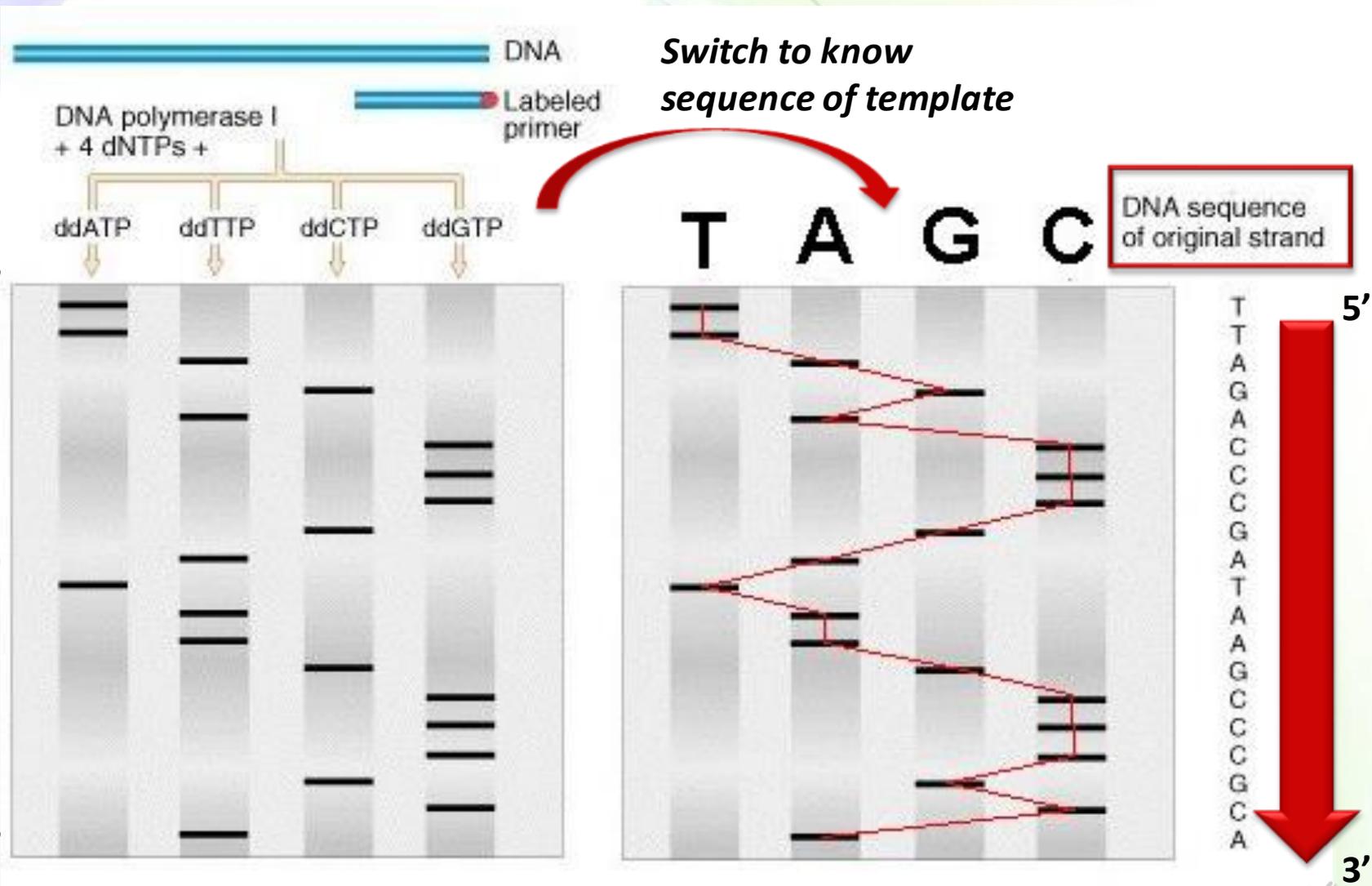


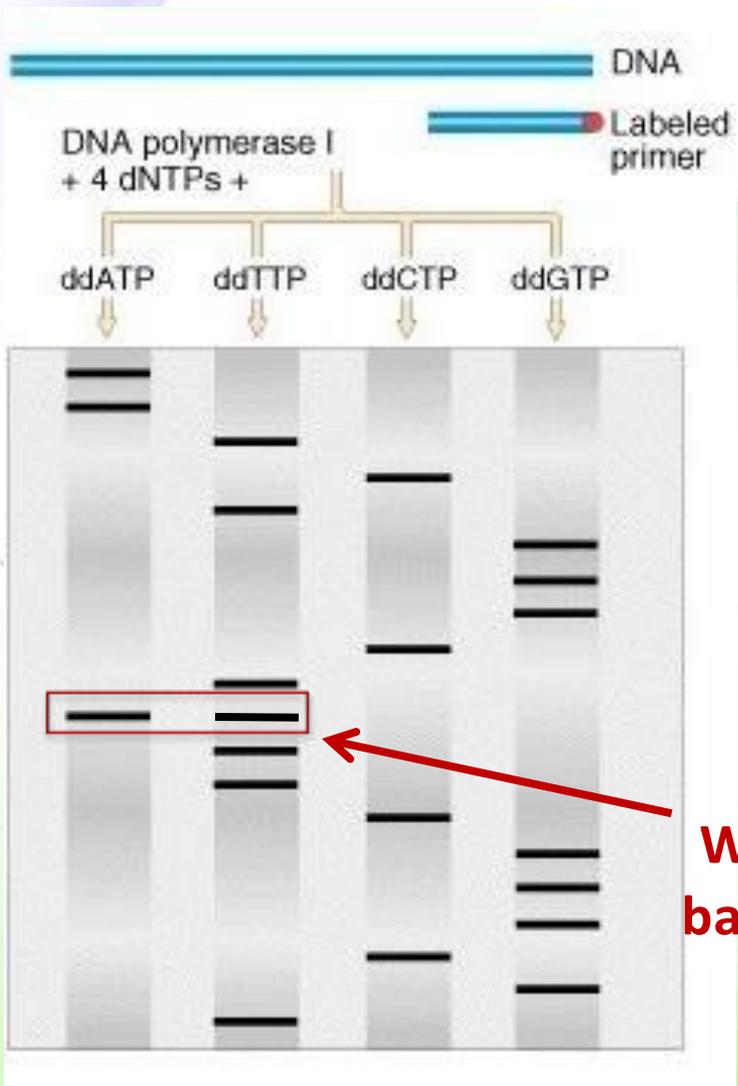
+ excess dATP  
 dTTP  
 dCTP  
 dGTP



DNA sequence reading directly from the bottom of the gel upward, is  
 ATGTCAGTCCAG  
 1 12







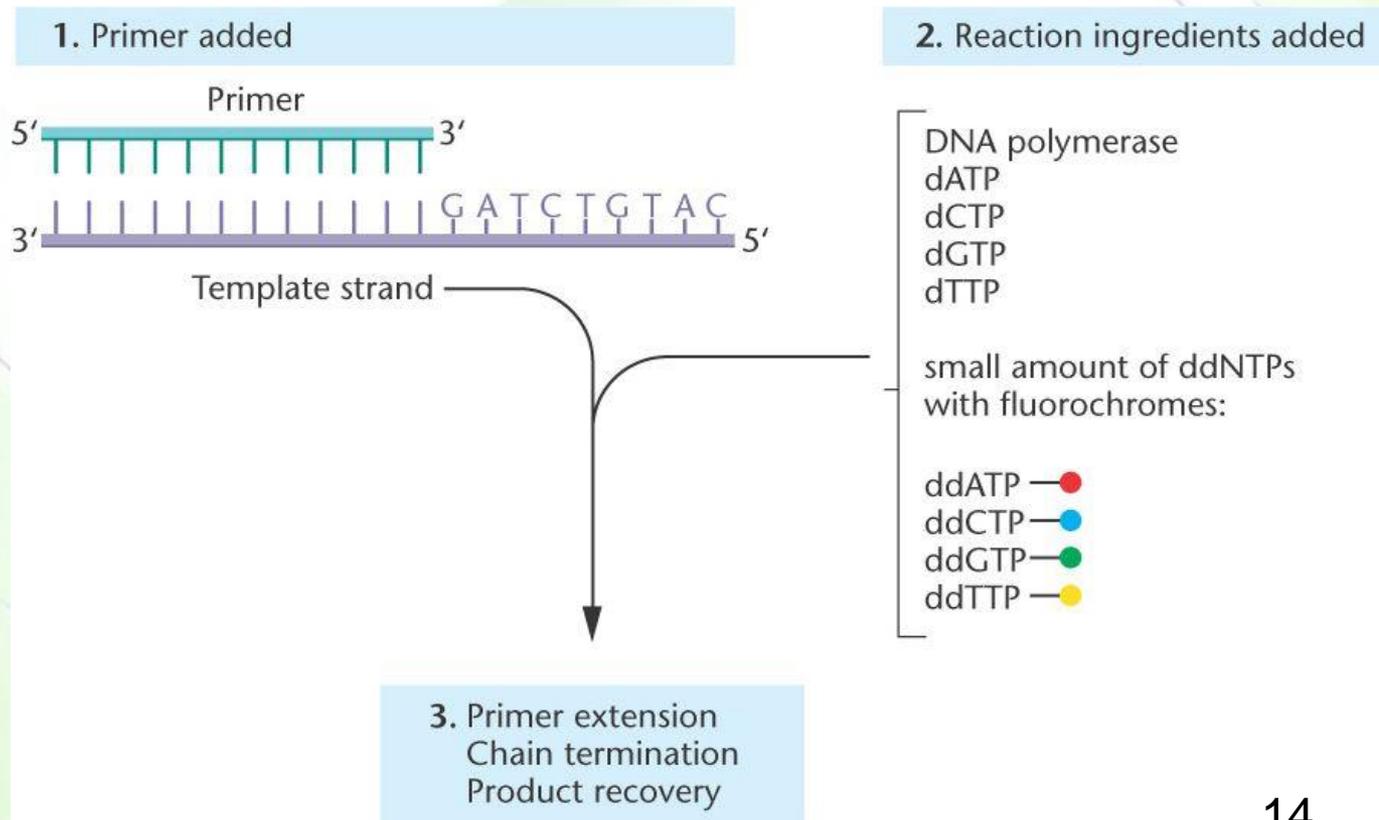
What do the band patterns mean?



# Fluorescence-based DNA sequencing

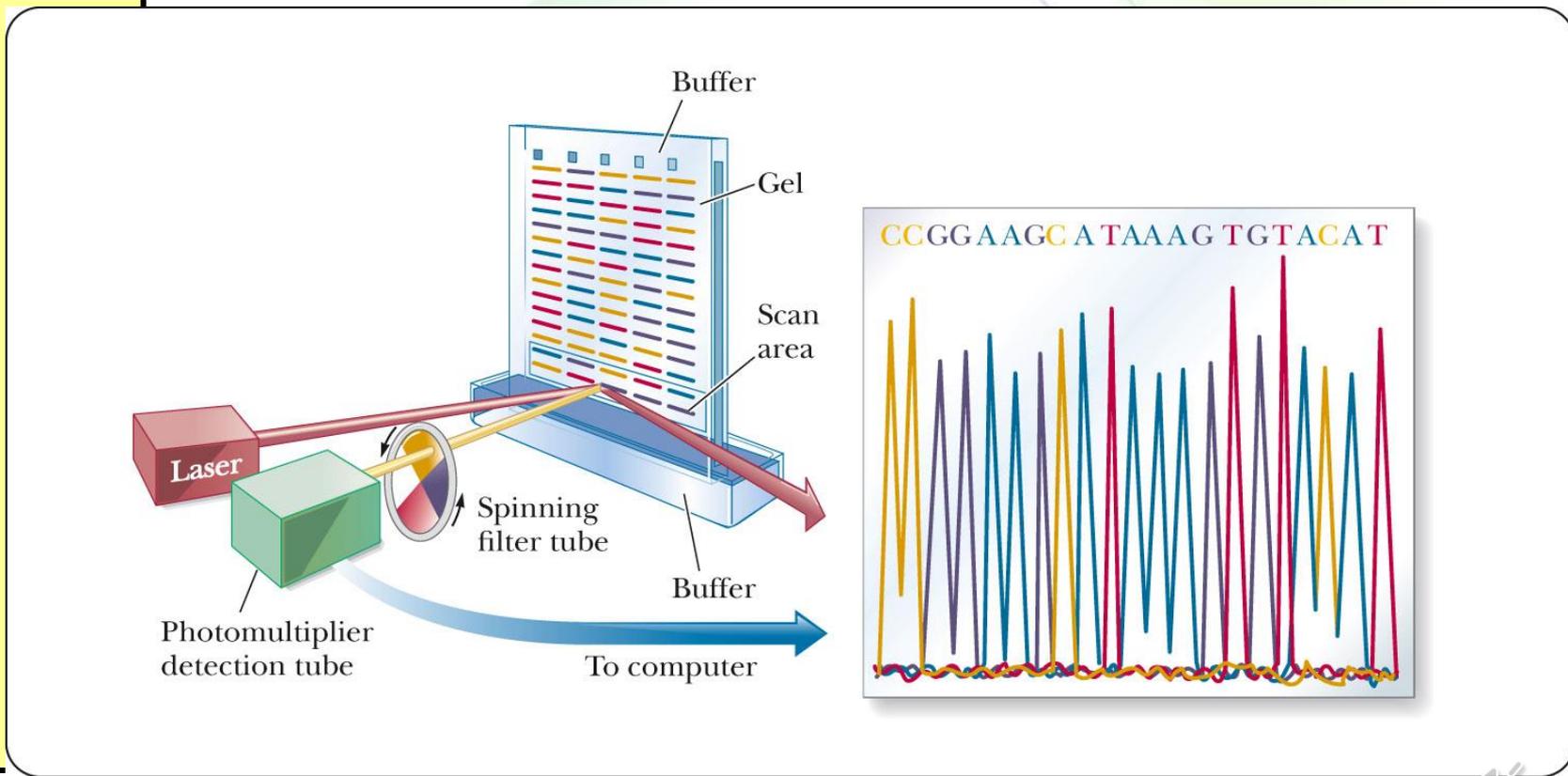
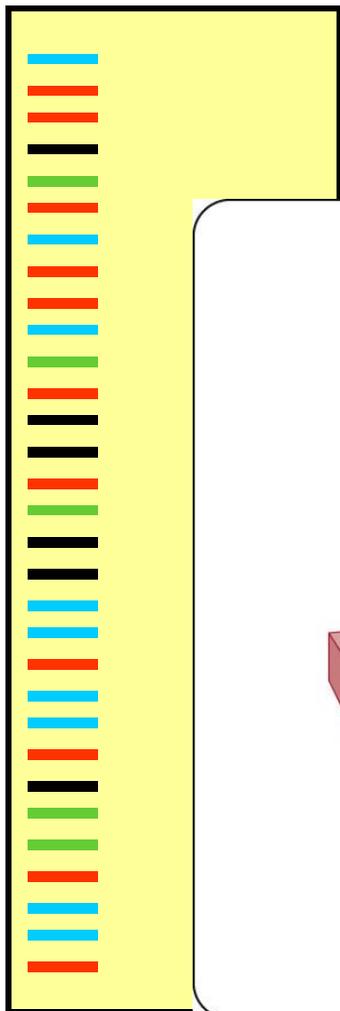


- Reactions include the four deoxynucleotides plus the four dideoxynucleotides in the same reaction with **each ddNTP labeled with a unique fluorescent tag**.





G A T C



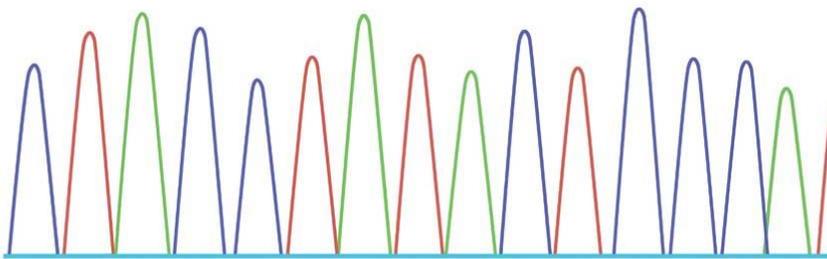


**A**

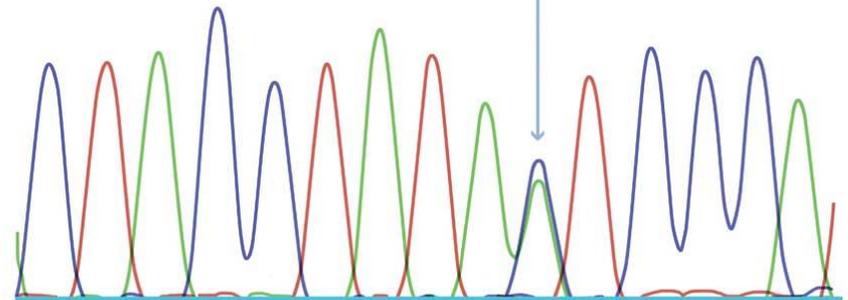
C T A C C T A T A C T C C C A

C T A C C T A T A C T C C C A

C > A



**Normal**



**Affected C. [1920 C > A]**

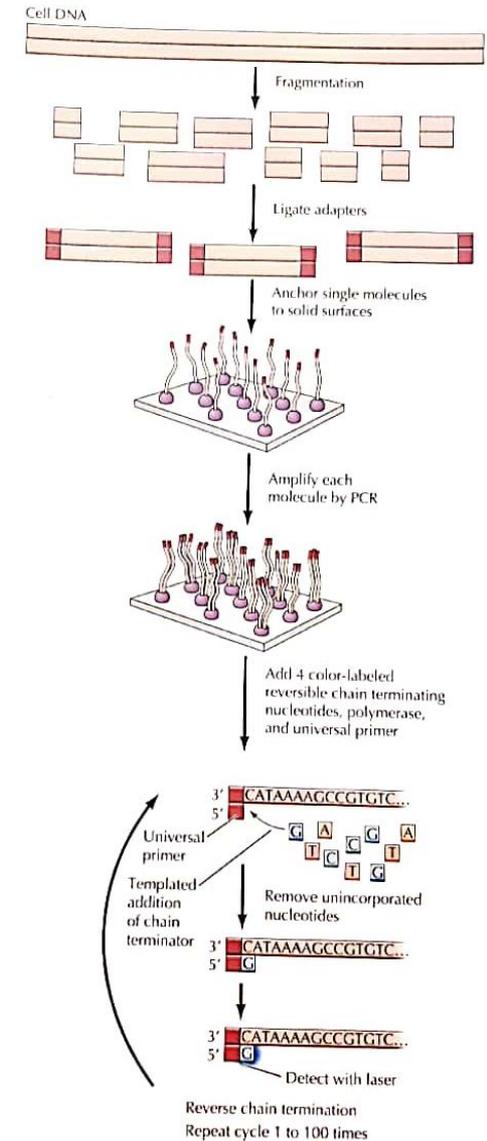
**What does it mean?**



# Next-generation sequencing



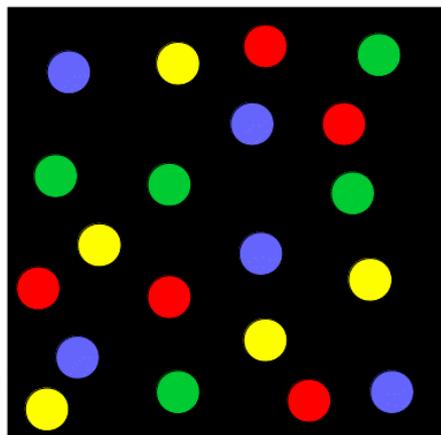
- Cellular DNA is fragmented.
- DNA adapters are added to ends of each DNA fragment.
- Each DNA fragment is attached to a solid surface and amplified like PCR using primers that anneal to the adapter sequences.
- Four-color nucleotides with terminating ends are added.
- A single nucleotide is incorporated and unincorporated nucleotides are removed.
- The incorporated nucleotide is modified for two reasons:
  - It is activated and *detected* by a special camera.
  - A new nucleotide can be added to it.
- The cycle is repeated.



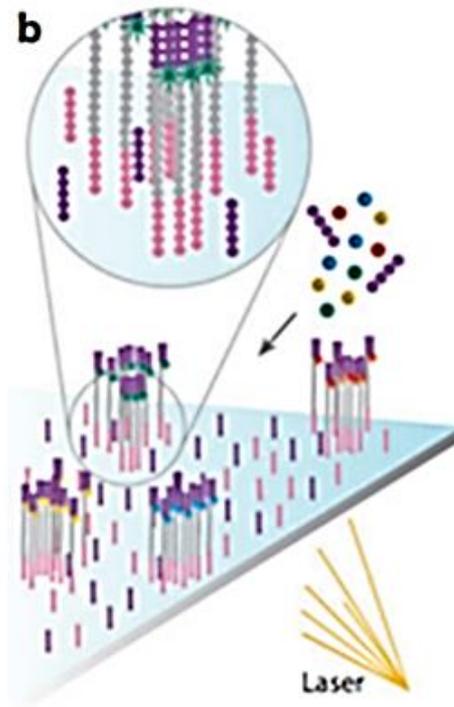
# The detection



4-Channel system (4 dyes)



4 Filter channels



Round 1

Round 2

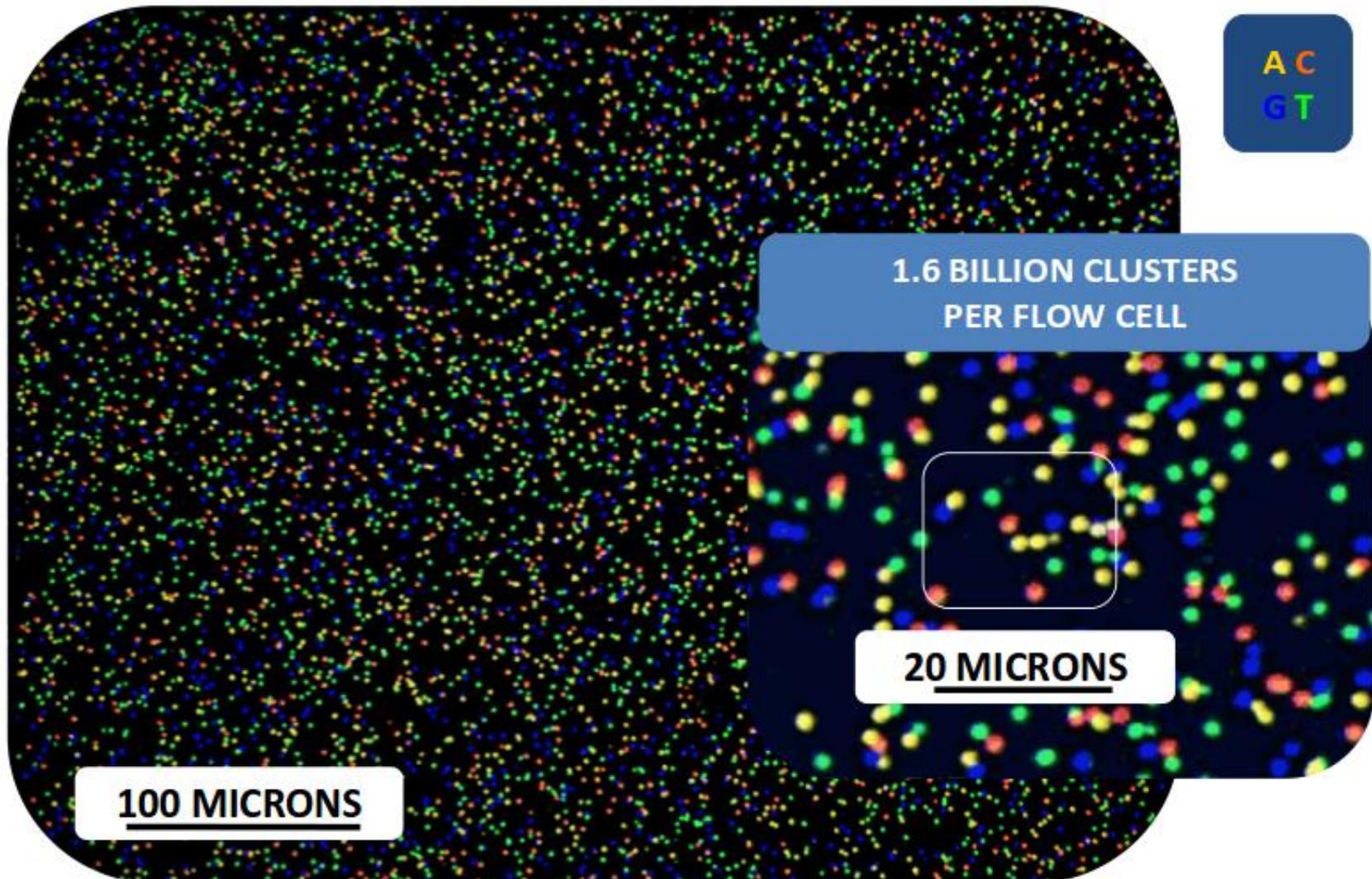
Round 3

Round 4

Round 5



# A real look



100 MICRONS

A C  
G T

1.6 BILLION CLUSTERS  
PER FLOW CELL

20 MICRONS





<https://www.youtube.com/watch?v=womKfikWlxM>

Data Analysis  
Create contiguous sequences

