



Enzyme-based molecular techniques (part I)

DNA sequencing

Prof. Mamoun Ahram

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.

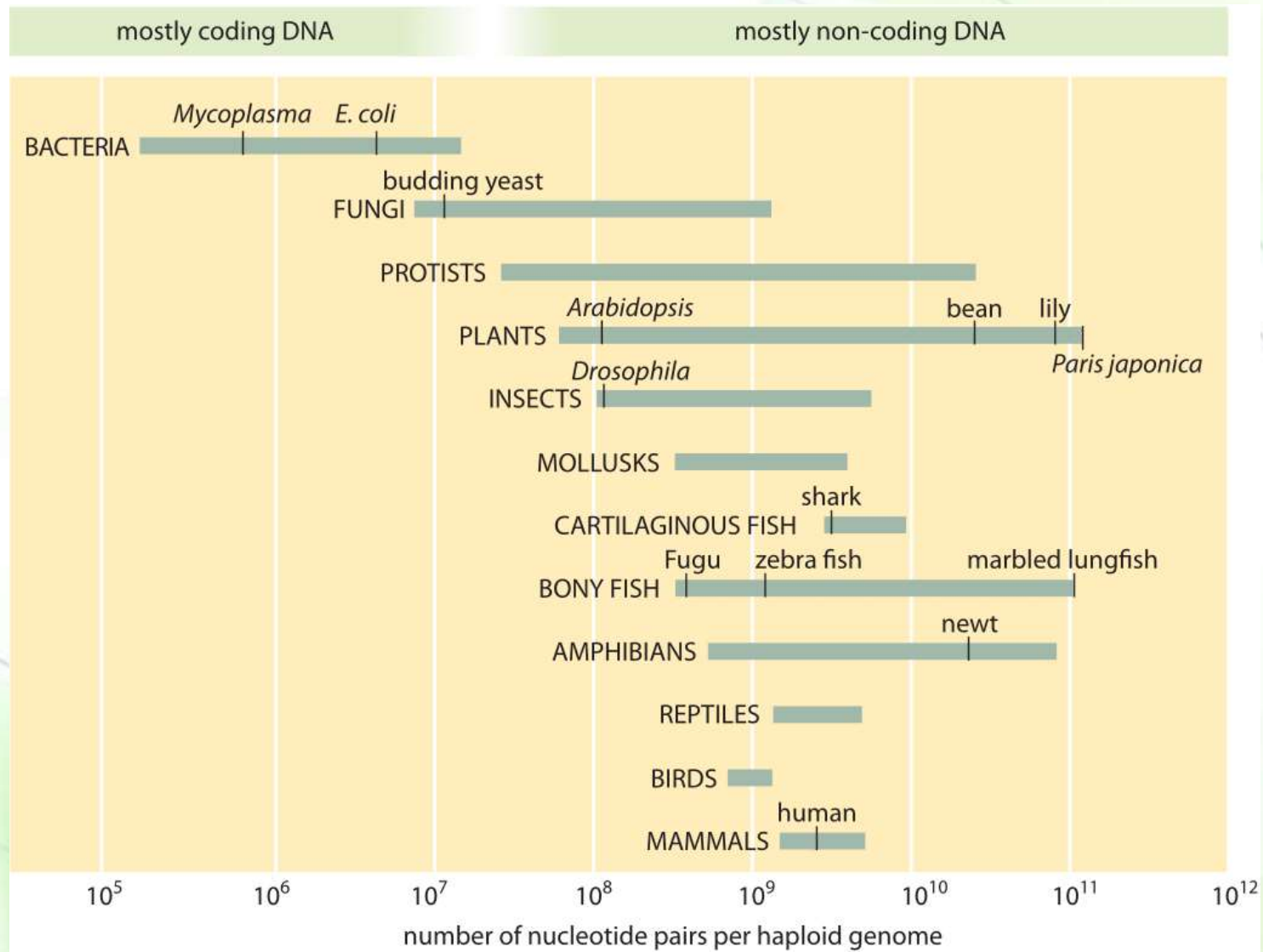




organism	genome size (base pairs)	protein coding genes	number of chromosomes
model organisms			
model bacteria <i>E. coli</i>	4.6 Mbp	4,300	1
budding yeast <i>S. cerevisiae</i>	12 Mbp	6,600	16
amoeba <i>D. discoideum</i>	34 Mbp	13,000	6
nematode <i>C. elegans</i>	100 Mbp	20,000	12 (2n)
fruit fly <i>D. melanogaster</i>	140 Mbp	14,000	8 (2n)
model plant <i>A. thaliana</i>	140 Mbp	27,000	10 (2n)
mouse <i>M. musculus</i>	2.8 Gbp	20,000	40 (2n)
human <i>H. sapiens</i>	3.2 Gbp	21,000	46 (2n)
viruses			
hepatitis D virus (smallest known animal RNA virus)	1.7 Kb	1	ssRNA
<i>HIV-1</i>	9.7 kbp	9	2 ssRNA (2n)
<i>influenza A</i>	14 kbp	11	8 ssRNA
bacteriophage λ	49 kbp	66	1 dsDNA
organelles			
mitochondria - <i>H. sapiens</i>	16.8 kbp	13 (+22 tRNA +2 rRNA)	1
chloroplast - <i>A. thaliana</i>	150 kbp	100	1
eukaryotes - multicellular			
dog <i>C. familiaris</i>	2.4 Gbp	19,000	40
chimpanzee <i>P. troglodytes</i>	3.3 Gbp	19,000	48 (2n)



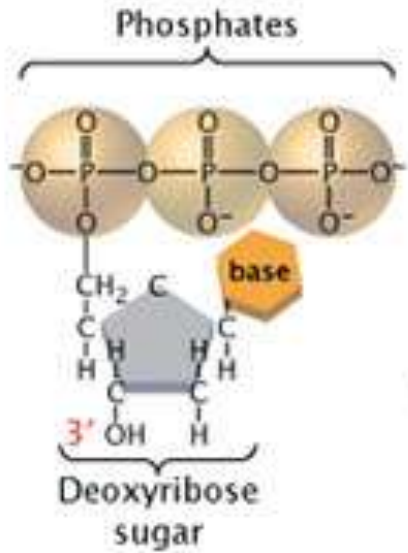
Nucleotides per genomes



DNA synthesis/elongation



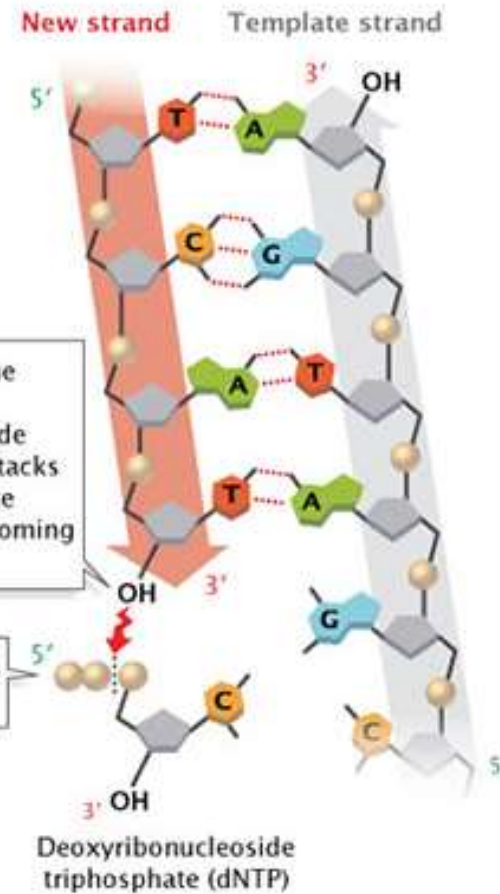
(a)



1 New DNA is synthesized from deoxyribonucleoside triphosphates (dNTPs).

2 In replication, the 3'-OH group of the last nucleotide on the strand attacks the 5'-phosphate group of the incoming dNTP.

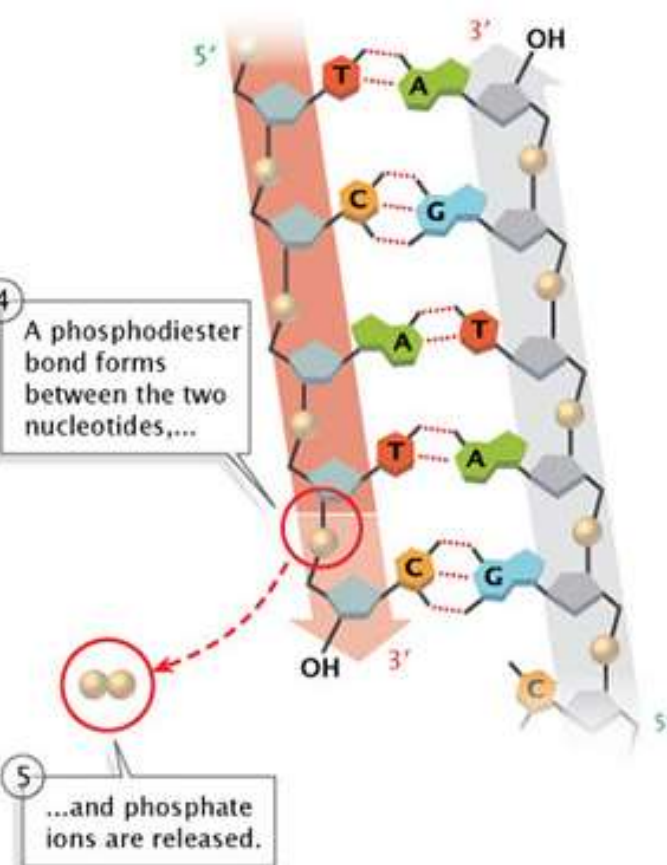
3 Two phosphates are cleaved off.



(c)

4 A phosphodiester bond forms between the two nucleotides,...

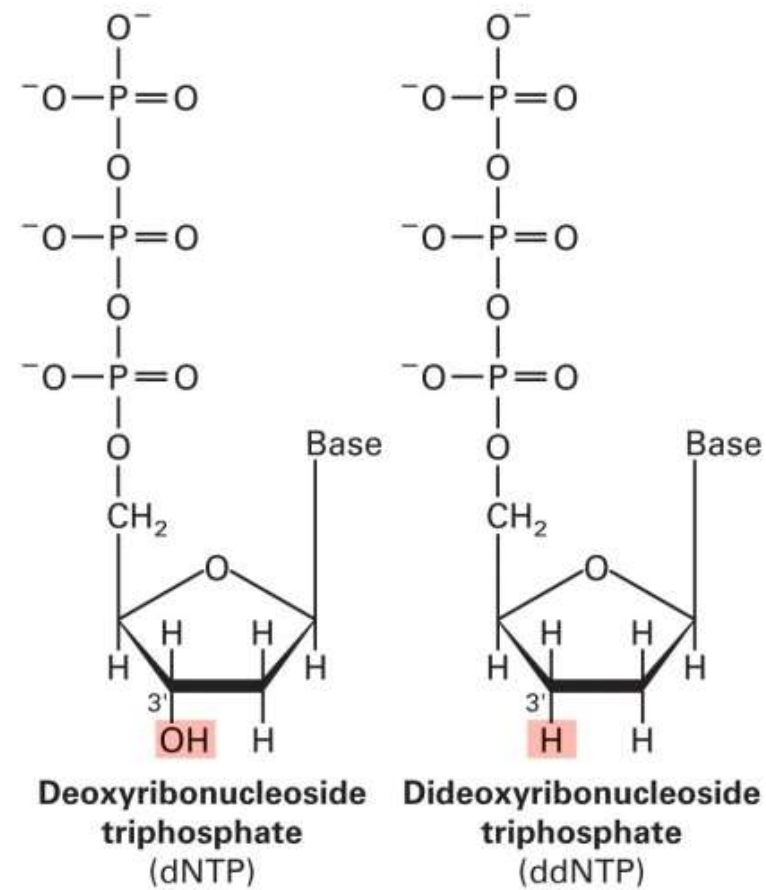
5 ...and phosphate ions are released.



The basic 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 ...

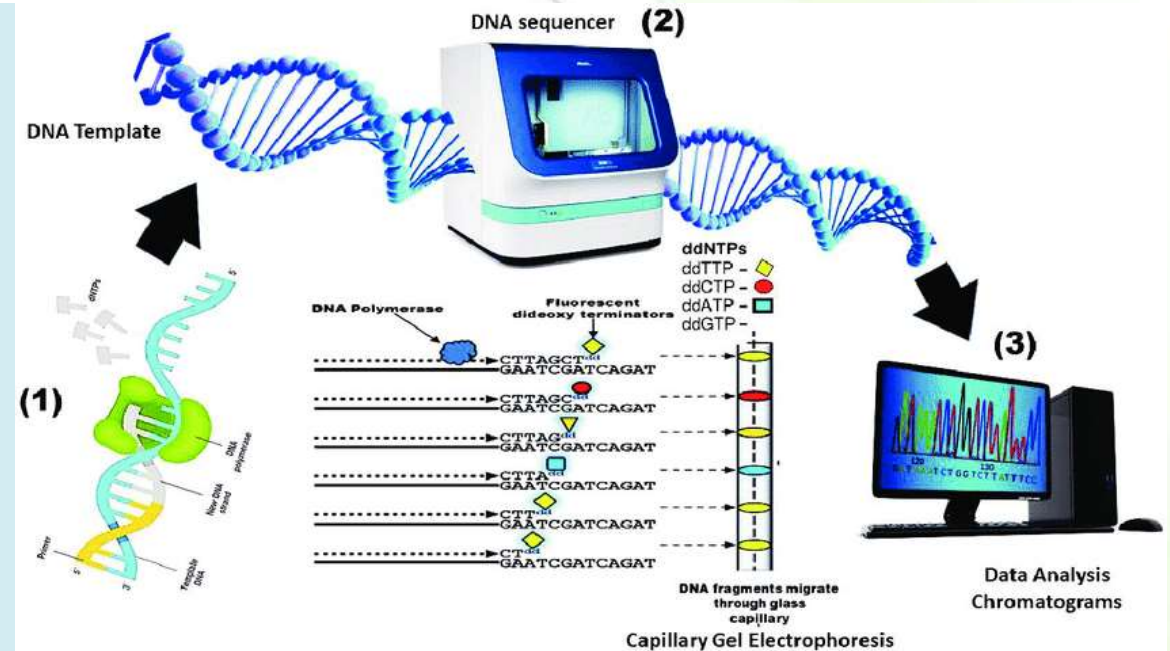
+
5' TAGCTGACTCAGTTCTTGATAACCC**G** 3'
3' ATCGACTGAGTCAAGAACTATTGGGCTTAA ...

What is DNA sequencing?



- DNA sequencing is the process of determining the exact order of the chemical building blocks, that are the A, T, C, and G bases, that make up the genome

1. DNA molecules (template) are extracted from the sample.
2. The molecules of a DNA template are mixed with DNA polymerase, a primer, nucleotides, and reaction-terminating, four-fluorescent nucleotides.
3. DNA fragments of different lengths and colors are generated each ending with a fluorescent signal.
4. The DNA fragments are separated and read by an instrument.
5. The fluorescent color of the terminating nucleotide is converted into a letter, then a sequence is created.

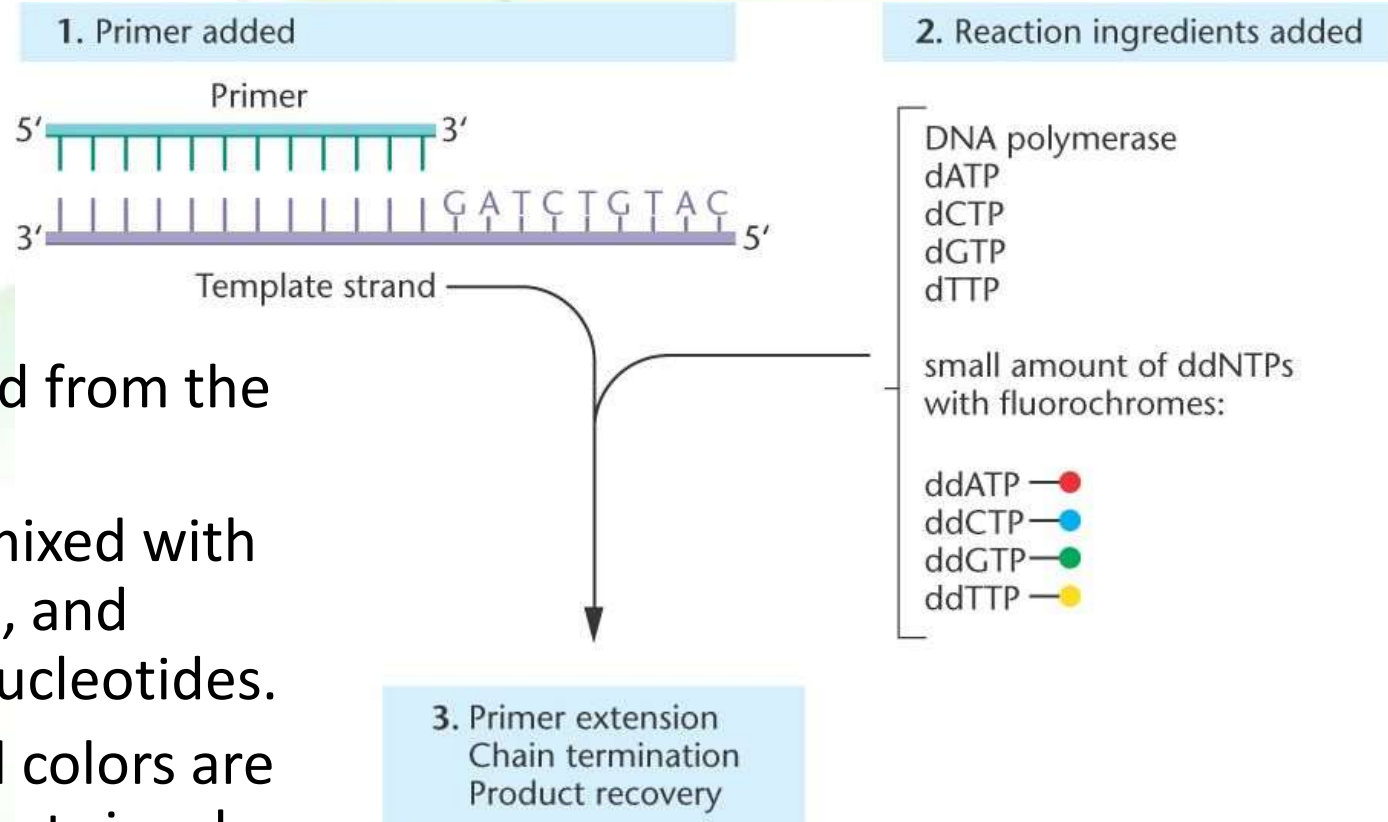


It can identify mutations and polymorphisms in a genome.

Fluorescence-based DNA sequencing

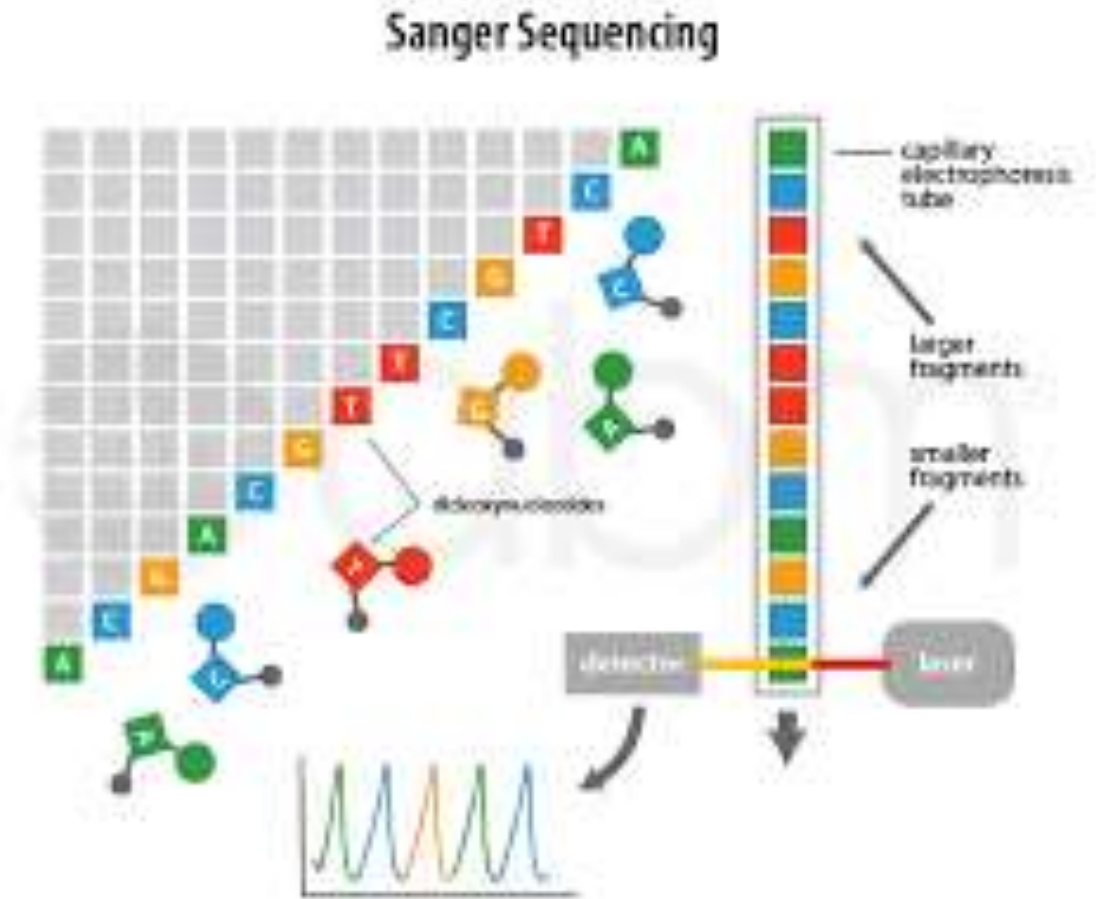


1. DNA molecules (template) are extracted from the sample.
2. The molecules of a DNA template are mixed with DNA polymerase, a primer, nucleotides, and reaction-terminating, four-fluorescent nucleotides.
3. DNA fragments of different lengths and colors are generated each ending with a fluorescent signal.
4. The DNA fragments are separated by a gel and read by an instrument.
5. The fluorescent color of the terminating nucleotide is converted into a letter, then a sequence is created.



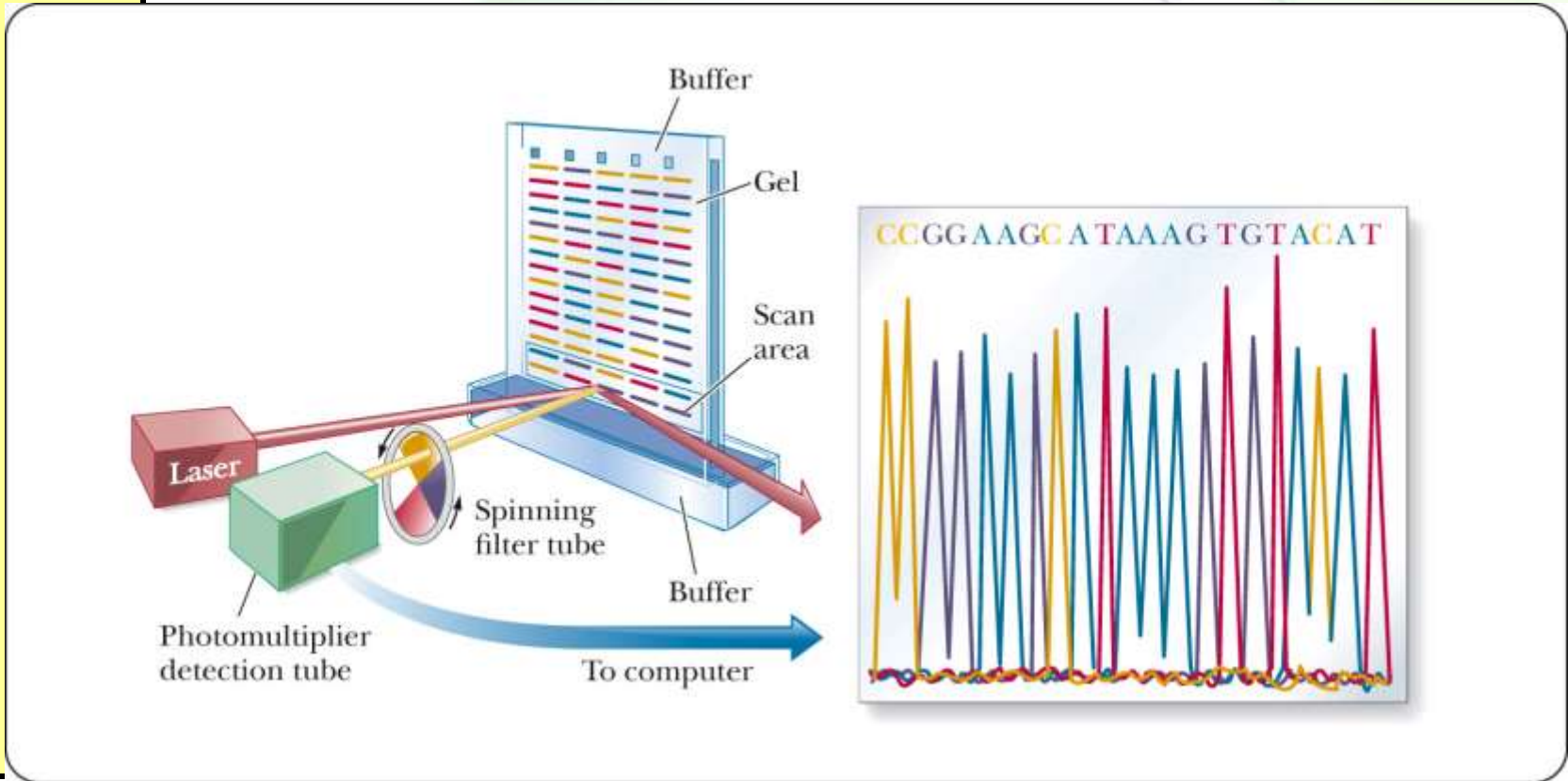
Sequencing a DNA Strand

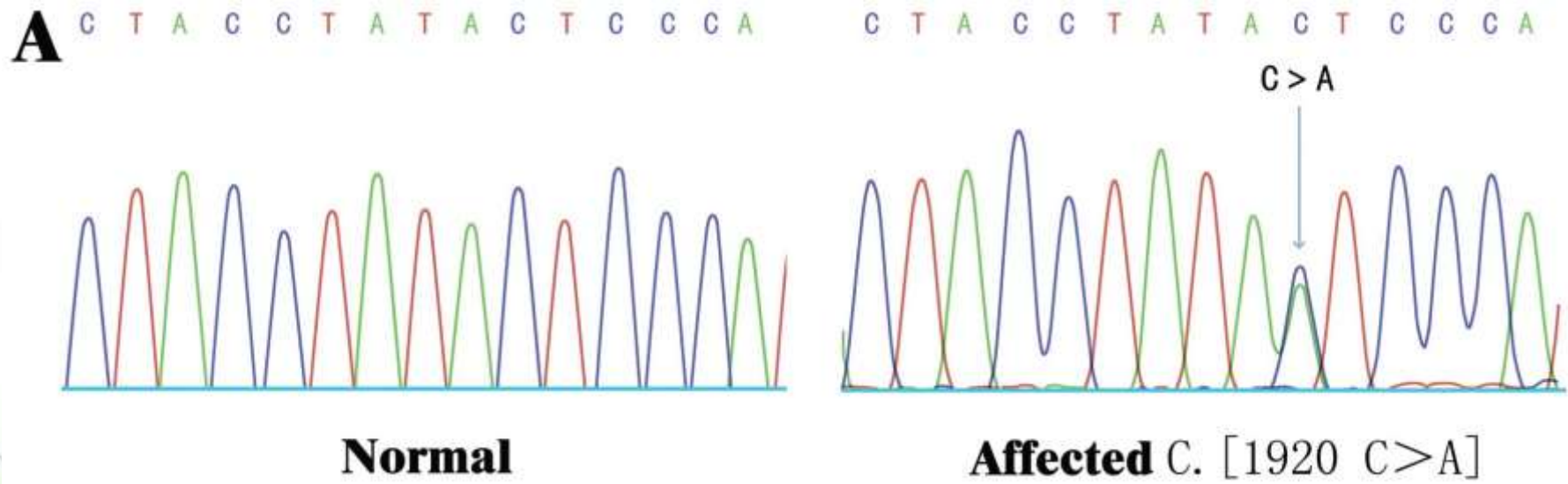
© Sinauer Associates, Inc.





G A T C

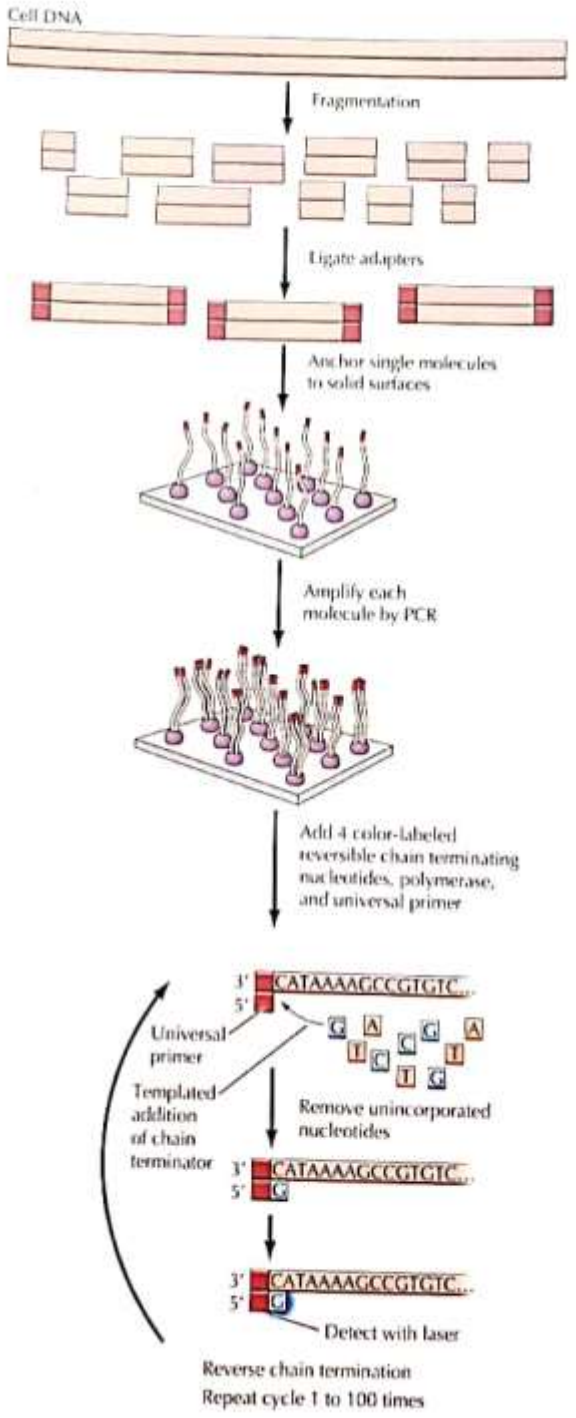




What does it mean?

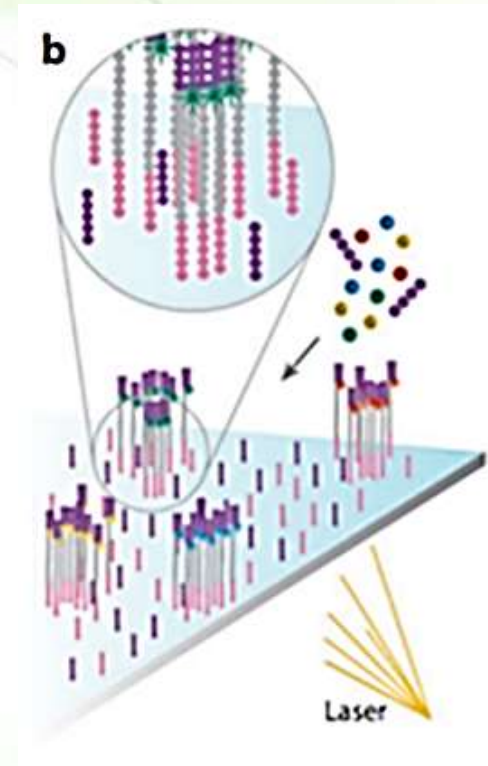
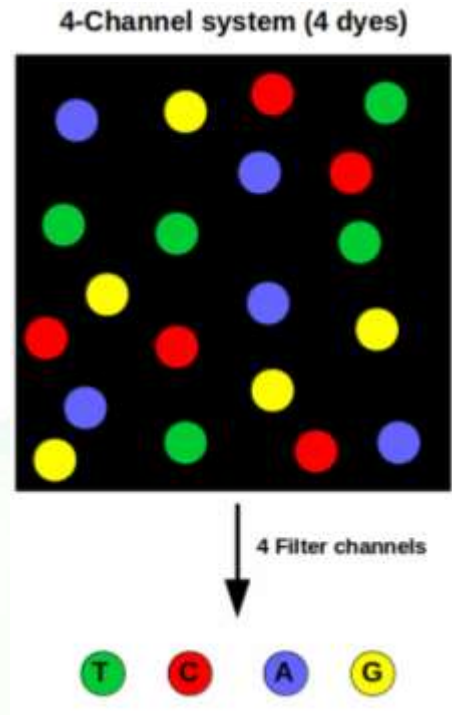


Next-generation sequencing

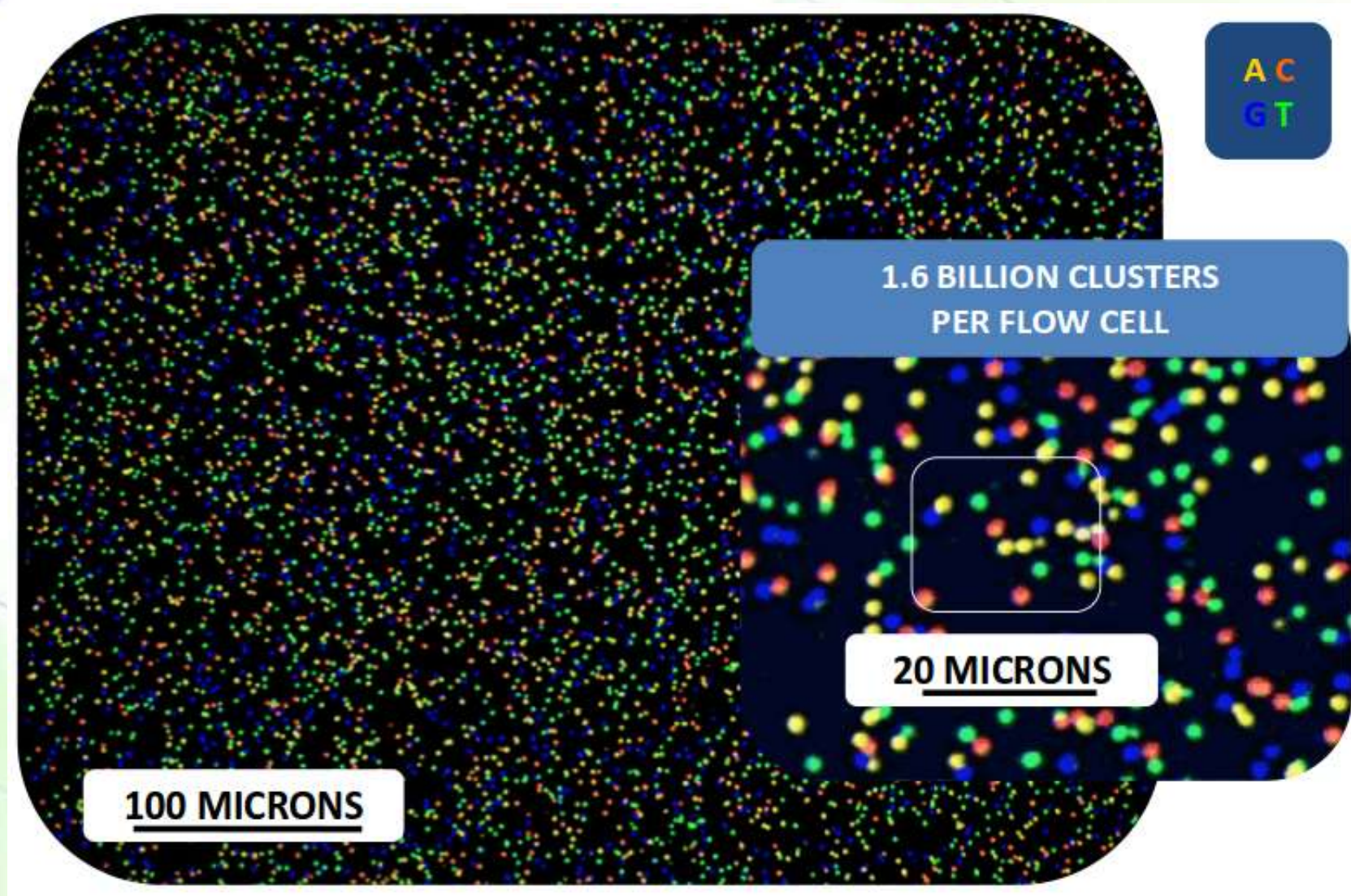


- Cellular DNA is fragmented.
- DNA adapters are added to ends of each DNA fragment.
- Each DNA fragment is attached to a solid surface and amplified like PCR using primers that anneal to the adapter sequences.
- Four-color nucleotides with terminating ends are added.
- A single nucleotide is incorporated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

The detection



A real look





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

 Data Analysis
Create contiguous sequences

