Molecular Biology (4)

DNA sequencing, PCR, and the human genome

Mamoun Ahram, PhD
Second semester, 2018-2019

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
Human *(Homo sapiens)*	3.2 billion	~ 25,000	46
Mouse *(Mus musculus)*	2.6 billion	~ 25,000	40
Fruit Fly *(Drosophilia melanogaster)*	137 million	13,000	8
Roundworm *(Caenorhabditis elegans)*	97 million	19,000	12
Yeast *(Saccharomyces cerevisia)*	12.1 million	6,000	32
Bacteria (Escherichia coli)	4.6 million	3,200	1
Bacteria (H. influenzae)	1.8 million	1,700	1

Nucleotides per genomes

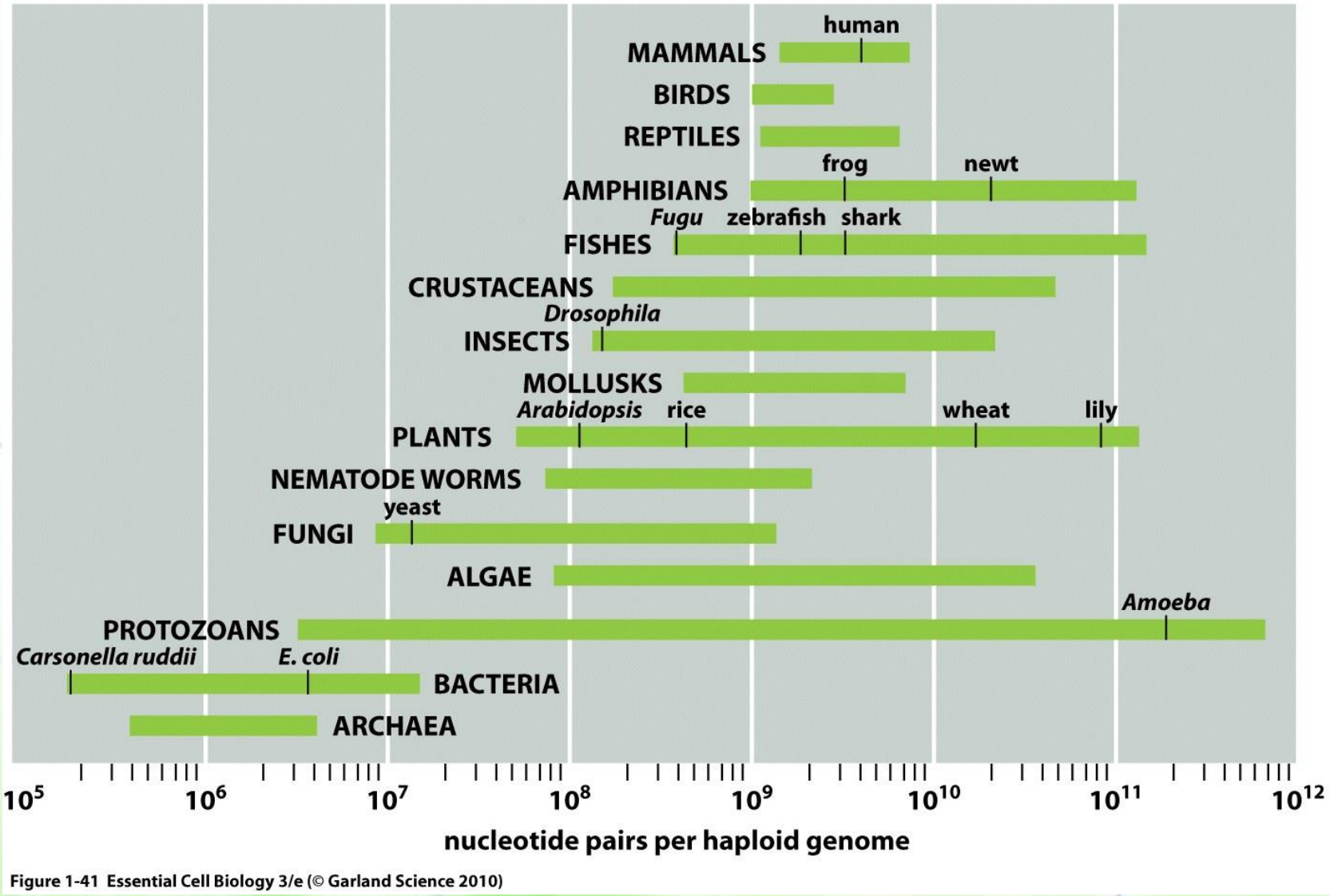
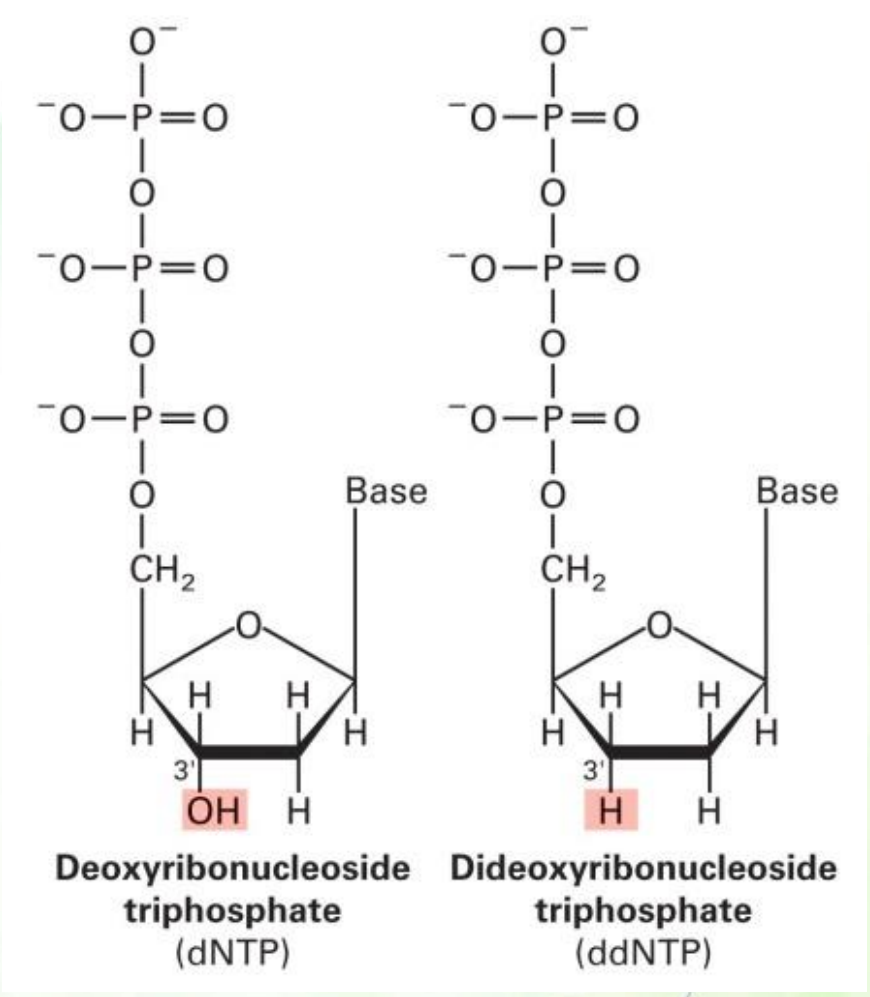


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

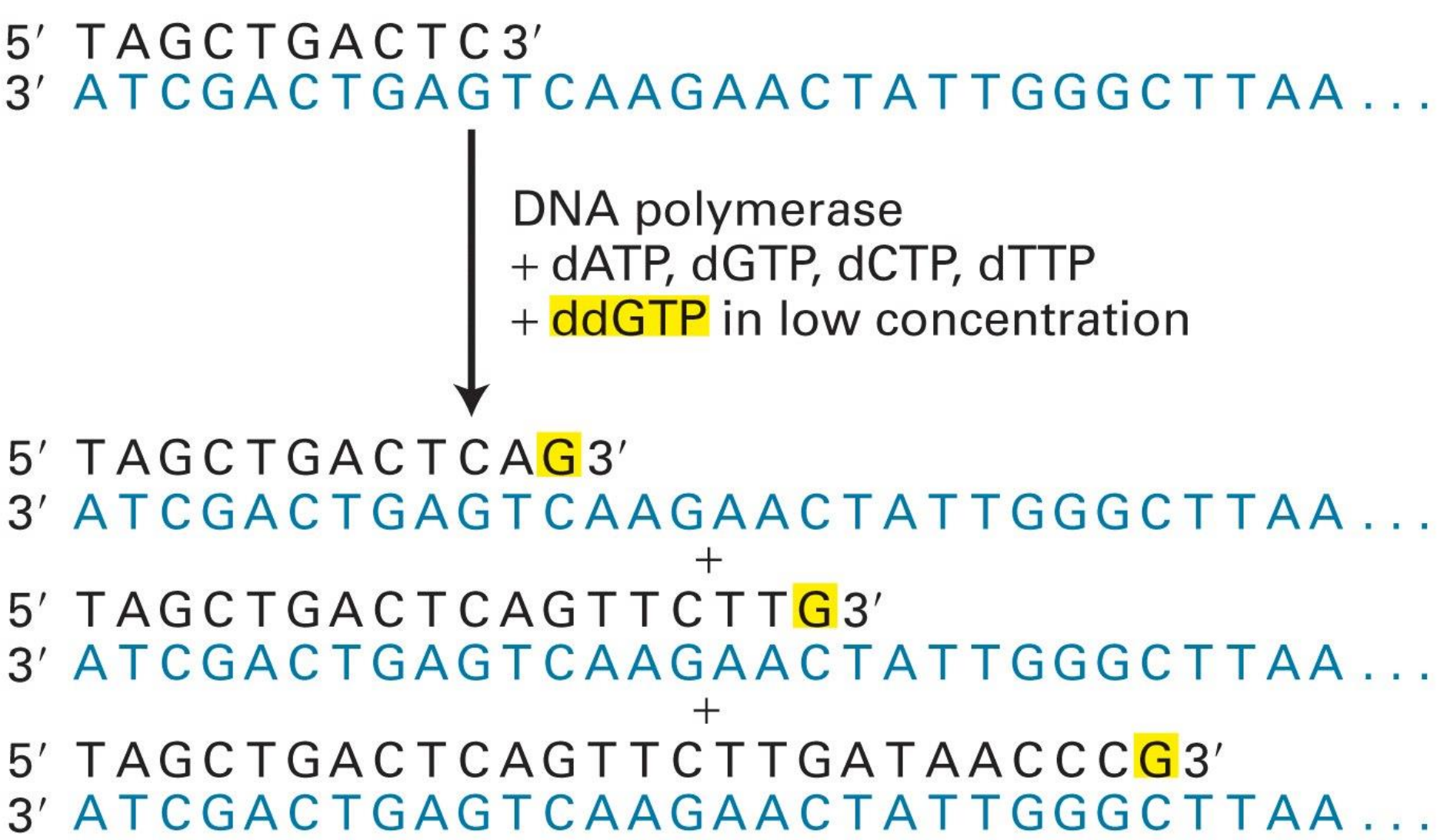
Method of DNA sequencing

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

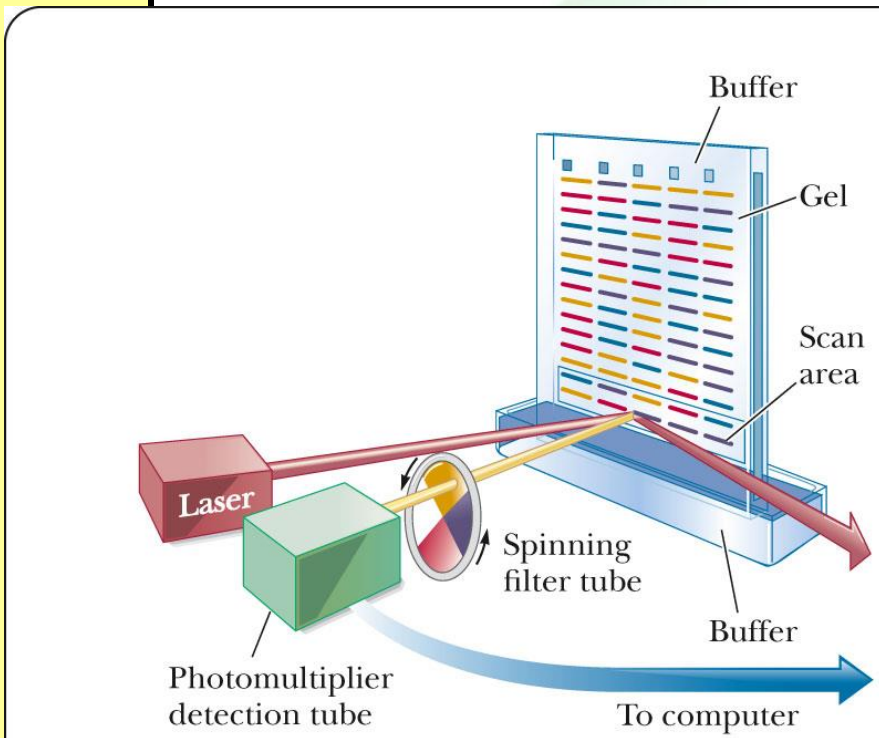


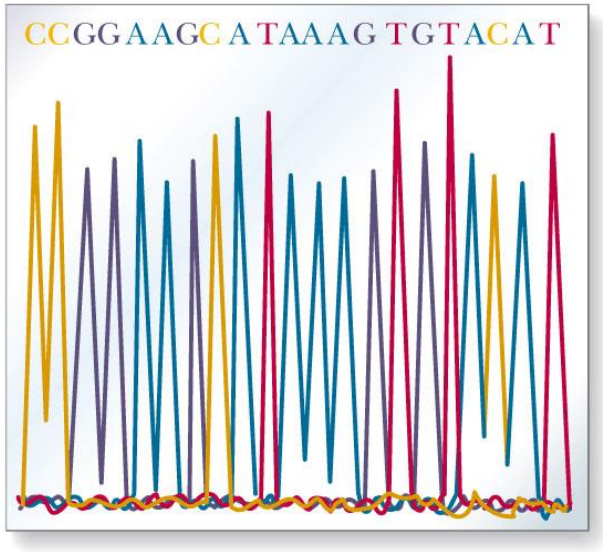
The process…

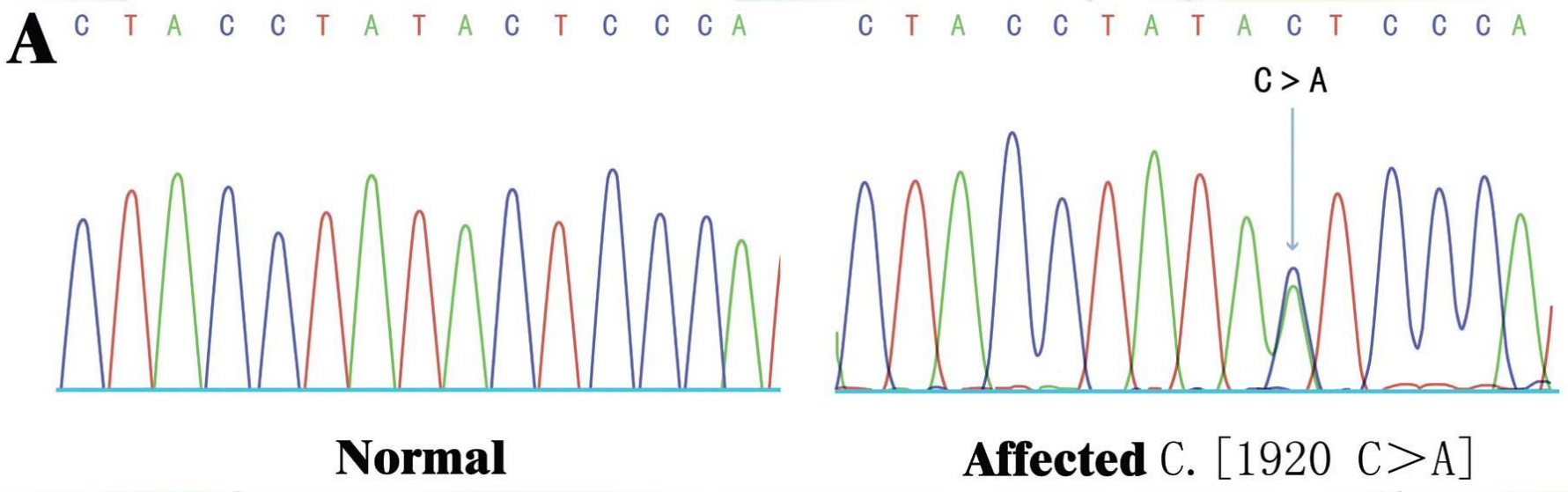
- DNA synthesis is initiated from a primer.
- Reactions include the four deoxynucleotides plus four dideoxynucleotides with ***each type labeled with a unique fluorescent tag***.
- Incorporation of a dideoxynucleotide stops DNA synthesis.
- DNA fragments of different lengths are generated.
- The DNA fragments are separated according to size by gel electrophoresis and their signal is detected.
- The fluorescence output is stored in the form of chromatograms.


5′ TAGCTGACTC 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
DNA polymerase
+ dATP, dGTP, dCTP, dTTP
+ ddGTP in low concentration
5′ TAGCTGACTCAG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTGATAACCCG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .


G A T C


Buffer
Gel
Scan area
Laser
Spinning filter tube
Buffer
Photomultiplier detection tube
To computer

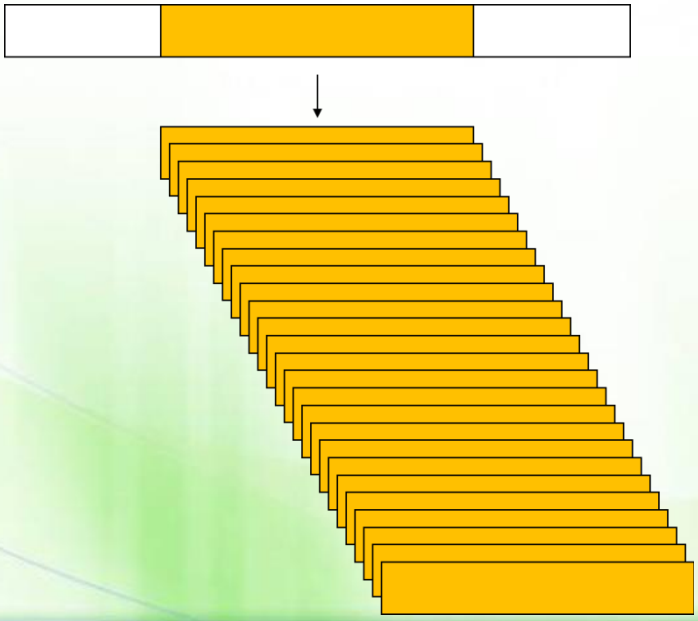

CCGGAAGCATAAAGTGTACAT



What does it mean?

Polymerase Chain Reaction

- Polymerase chain reaction (PCR) allows the DNA from a selected region of a genome to be amplified a billionfold, effectively "purifying" this DNA away from the remainder of the genome.
- It s extremely sensitive; it can detect a single DNA molecule in a sample.

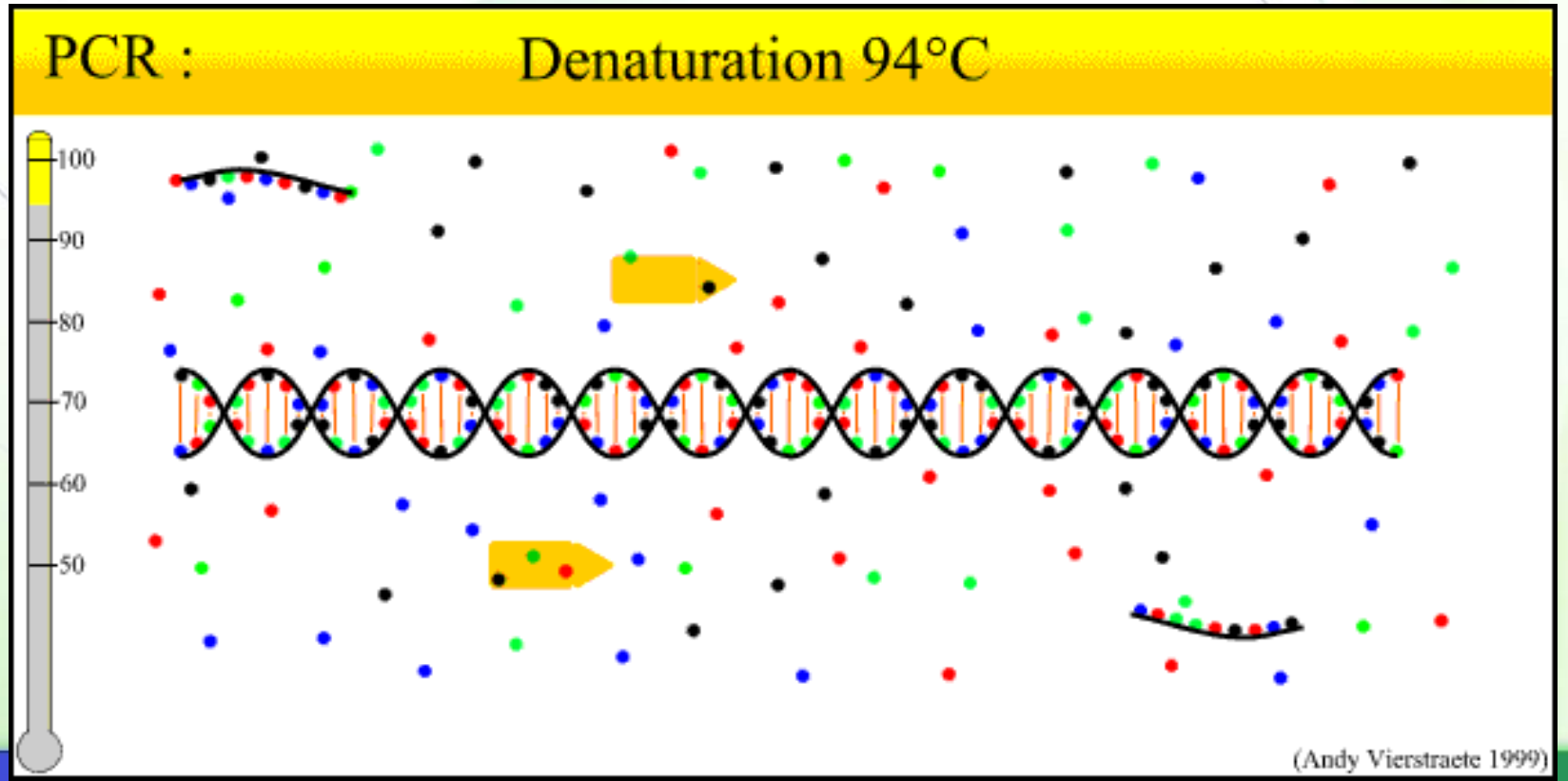


Components of PCR reaction

- The DNA template
- A pair of primers
  - The 15-25 nucleotides-long primers should surround the target sequence.
- All four deoxyribonucleoside triphosphates
- A heat-stable DNA polymerase

The PCR *cycles*

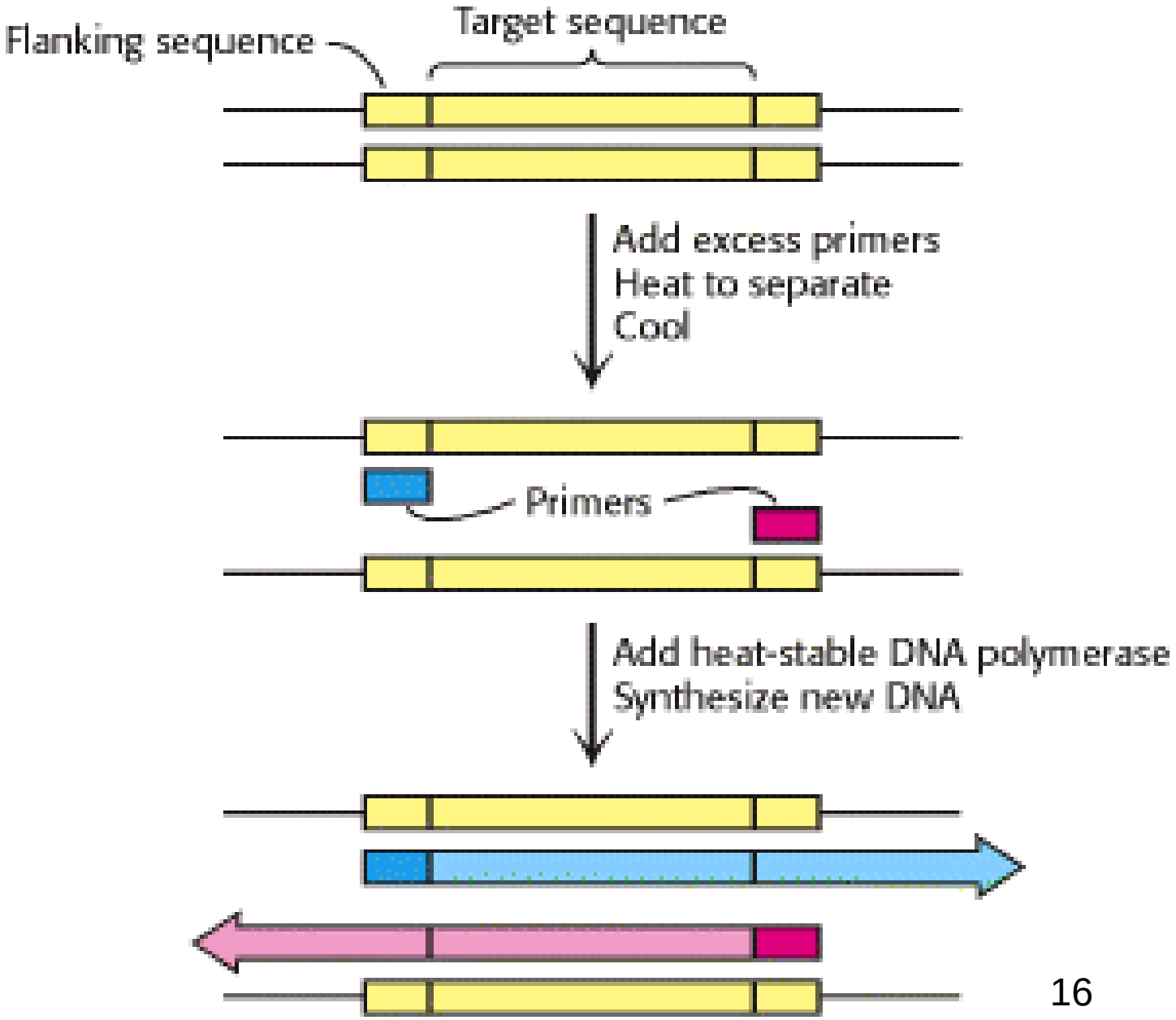
- Denaturation (at 95°C): DNA is denatured into single-stranded molecules.
- Reannealing (50°C to 70°C): the primers anneal to the DNA.
- DNA synthesis (at 72°C): optimal for the polymerase.

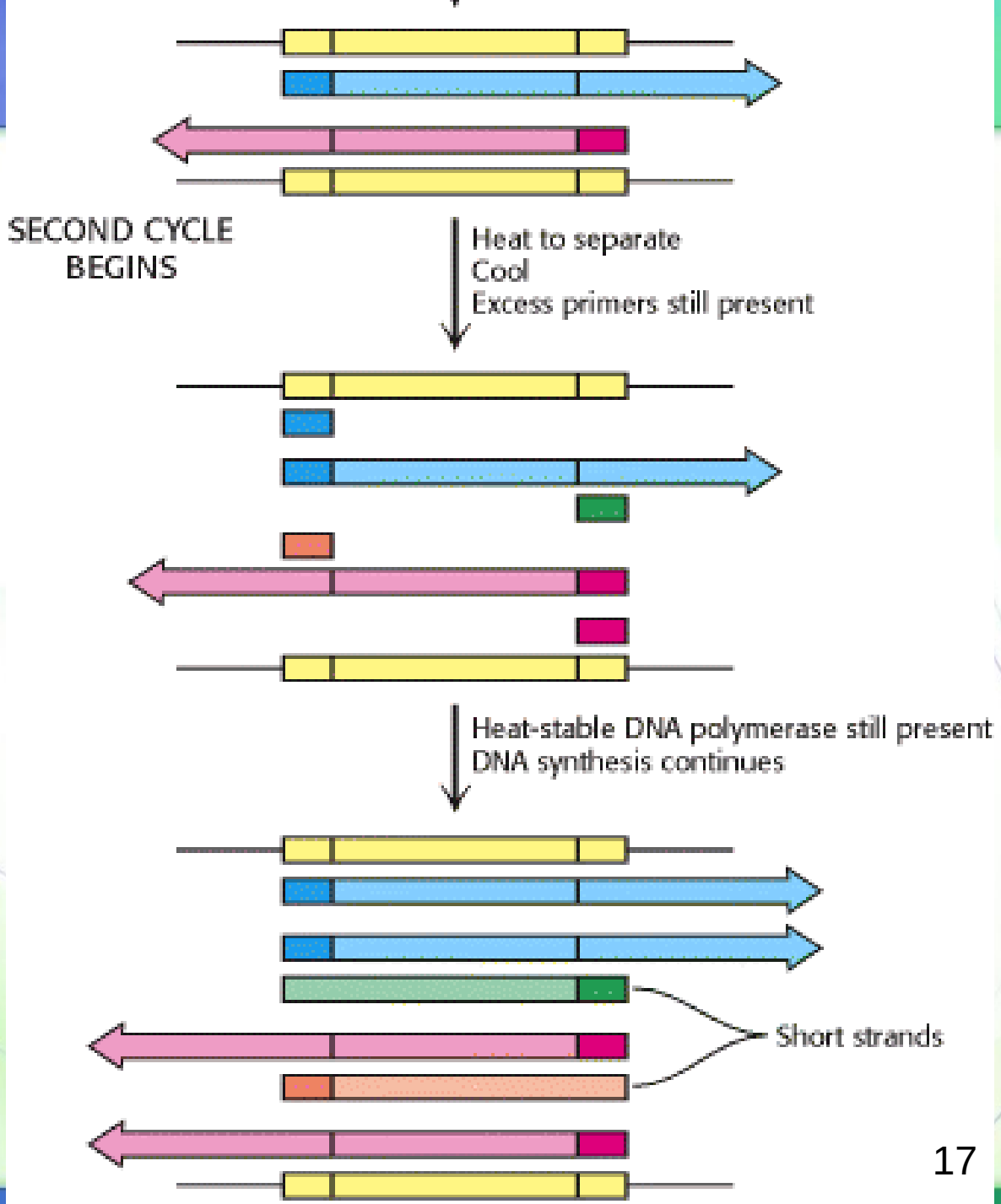


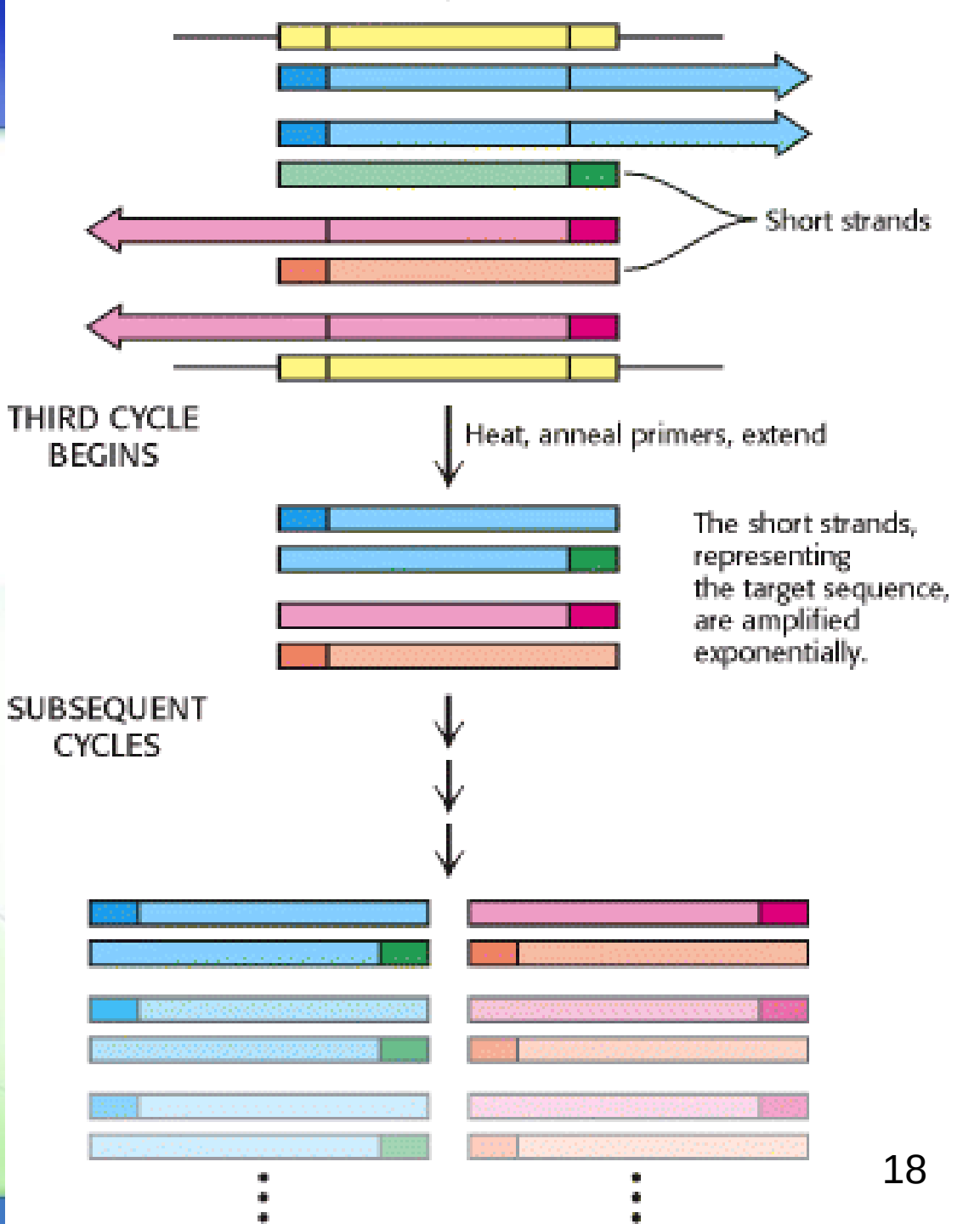
The DNA polymerase

- Suitably heat-stable DNA polymerases have been obtained from microorganisms whose natural habitat is hot springs.
- For example, the widely used Taq DNA polymerase is obtained from a thermophilic bacterium, Thermus aquaticus, and is thermostable up to 95°C.

FIRST CYCLE BEGINS




SECOND CYCLE
BEGINS
Heat to separate
Cool
Excess primers still present
Heat-stable DNA polymerase still present
DNA synthesis continues
Short strands

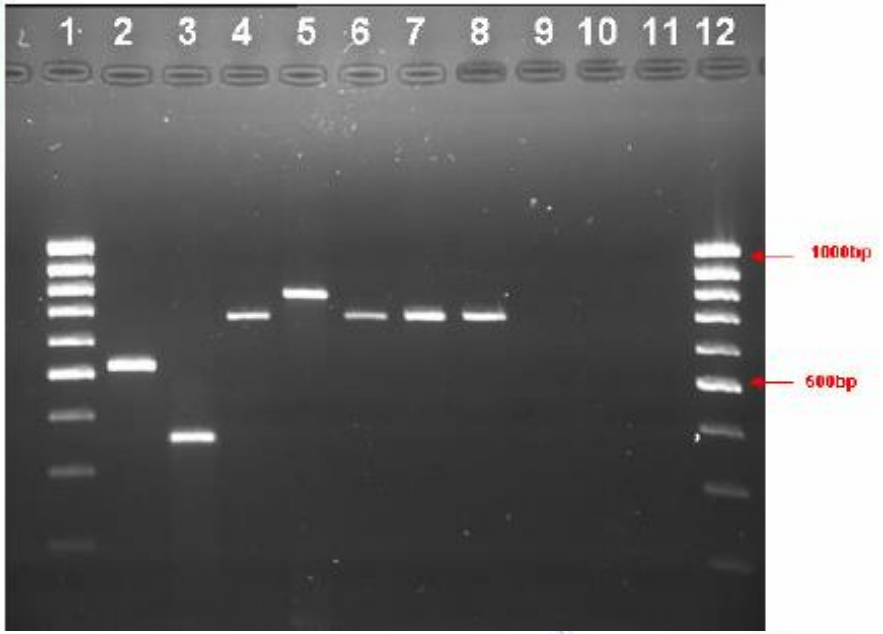

Short strands
THIRD CYCLE BEGINS
Heat, anneal primers, extend
The short strands, representing the target sequence, are amplified exponentially.
SUBSEQUENT CYCLES

PCR cycles

- 20-30 cycles of reaction are required for DNA amplification.
  - the products of each cycle serving as the DNA templates for the next-hence the term polymerase "chain reaction“.
- Every cycle doubles the amount of DNA.
- After 30 cycles, there will be over 250 million short products derived from each starting molecule.

Detection of DNA fragments

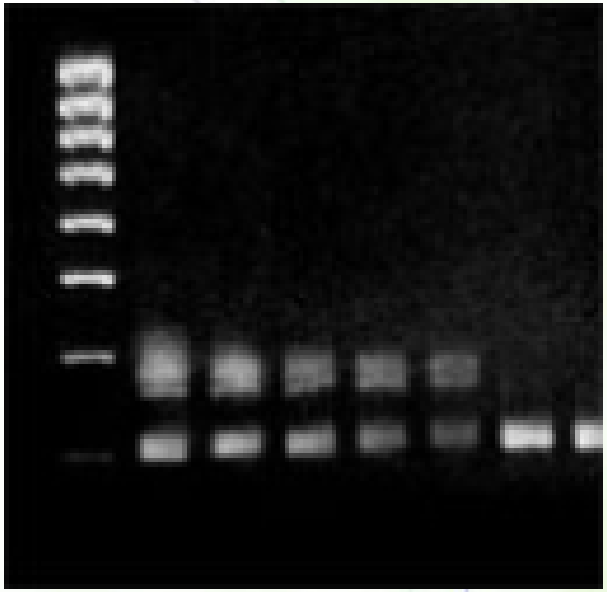
- This DNA fragment can be easily visualized as a discrete band of a specific size by agarose gel electrophoresis.



Importance of primers

- The specificity of amplification depends on the specificity of the primers to not recognize and bind to sequences other than the intended target DNA sequences
- How can you prevent it?
- How can you take advantage of it?

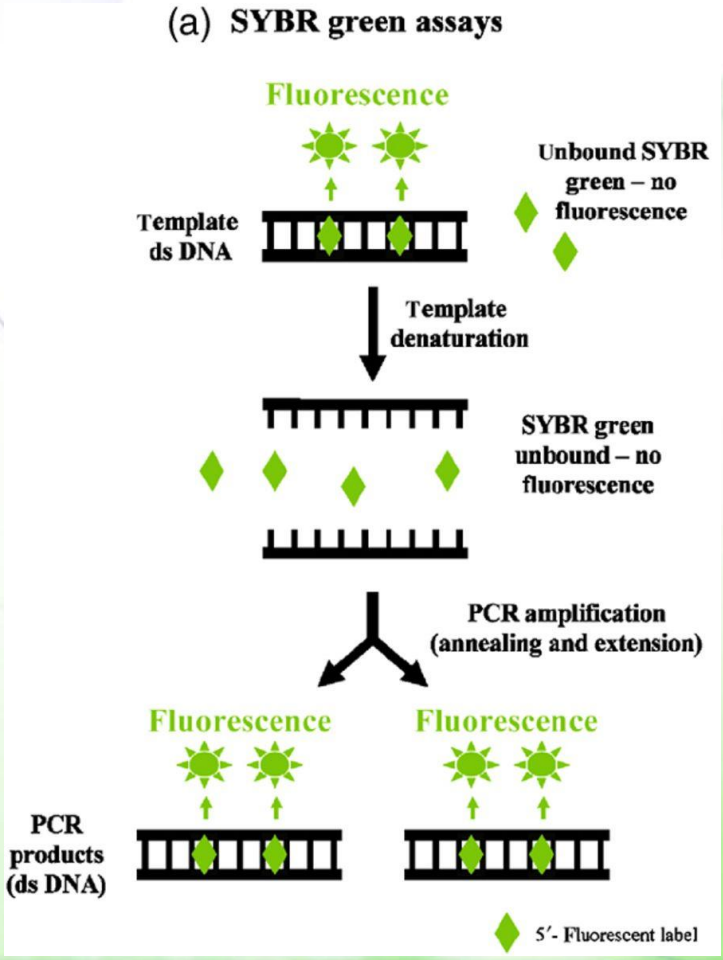
Annealing temperature



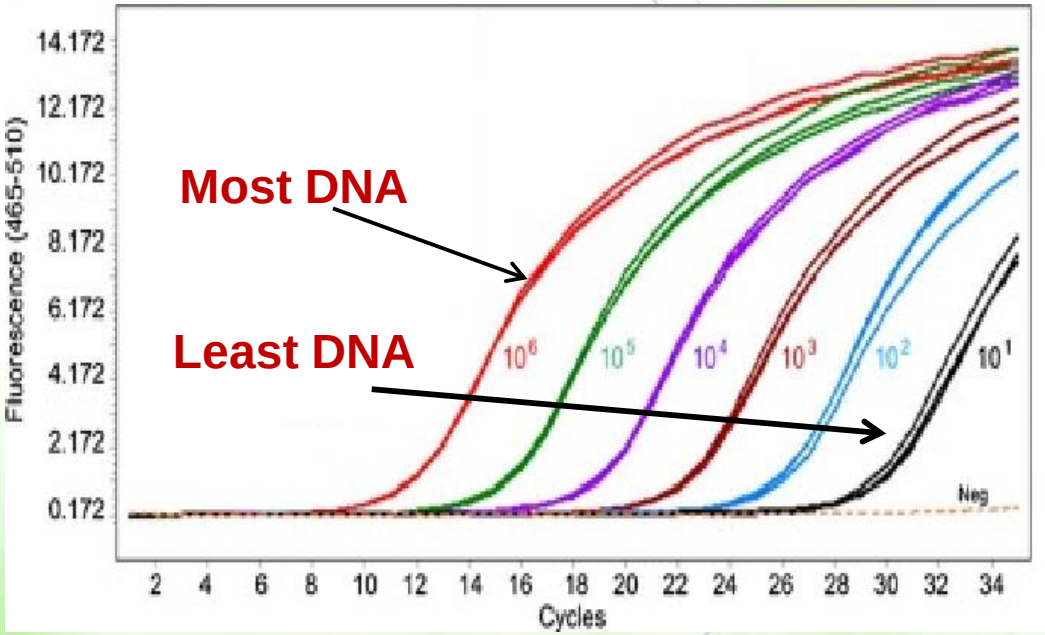
Uses of PCR

- Discovery of gene families
- Disease diagnosis
- Paternity and criminal cases. Why?
  - An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population.
- Viral and bacterial load: the quantity of virus in a given volume. How?
  - Quantitative PCR

Quantitative PCR (qPCR)

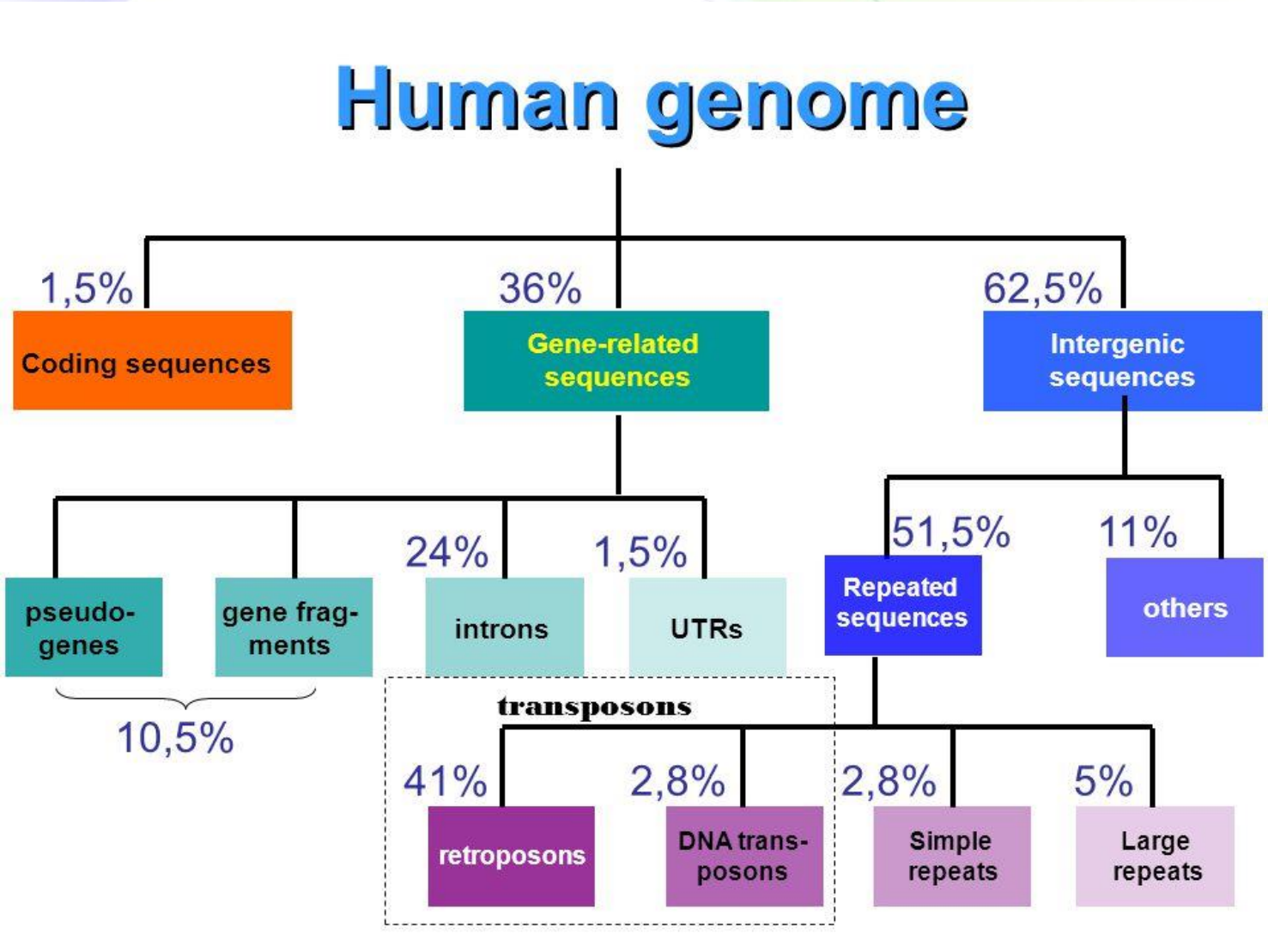


- SYBR green binds to double-stranded DNA and fluoresces only when bound.
- A way of relative quantitation of amount of DNA in a sample is by amplifying it in the presence of SYBR green.
- The higher the amount of DNA, the sooner it is detected.

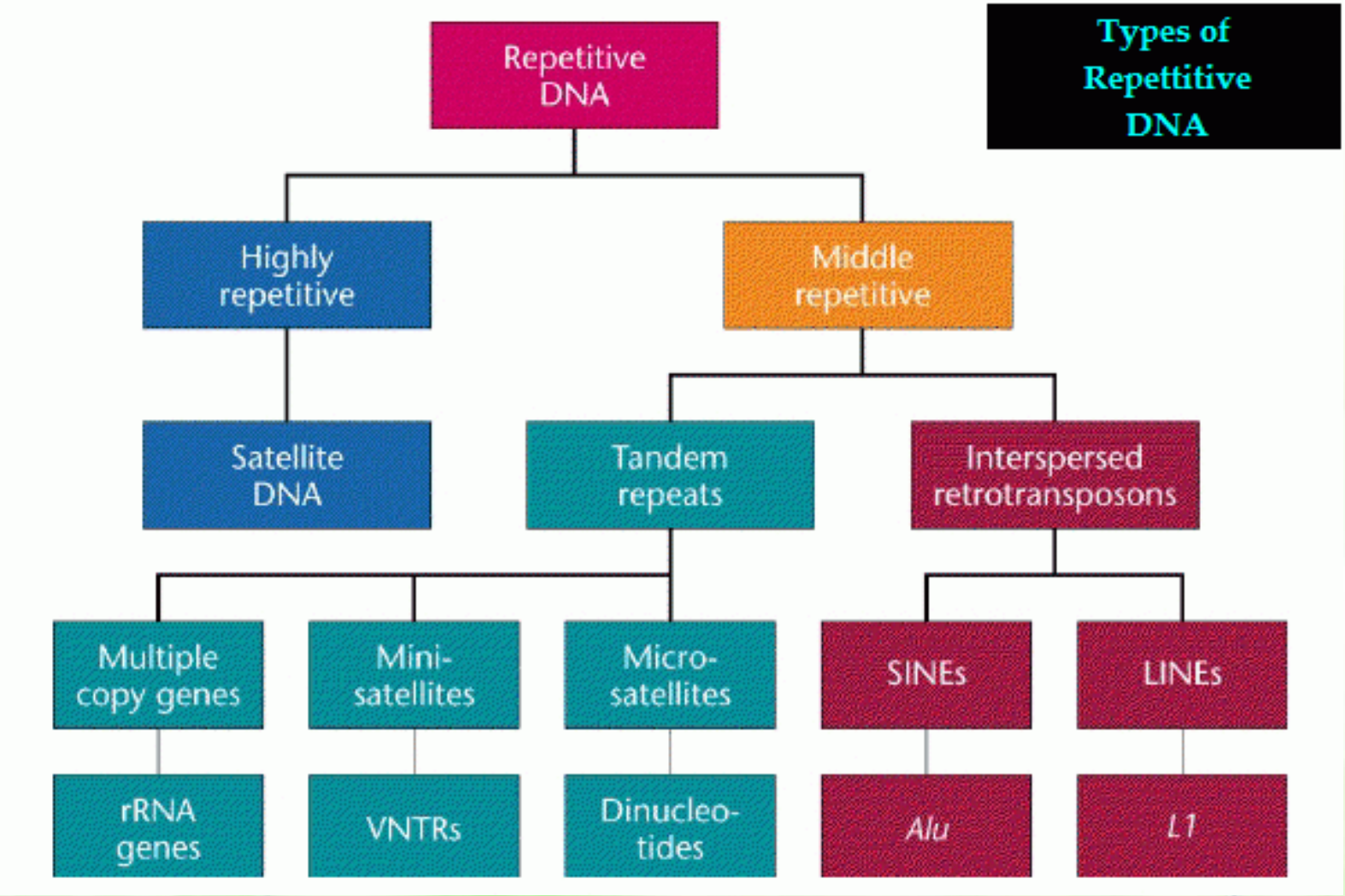


The human genome

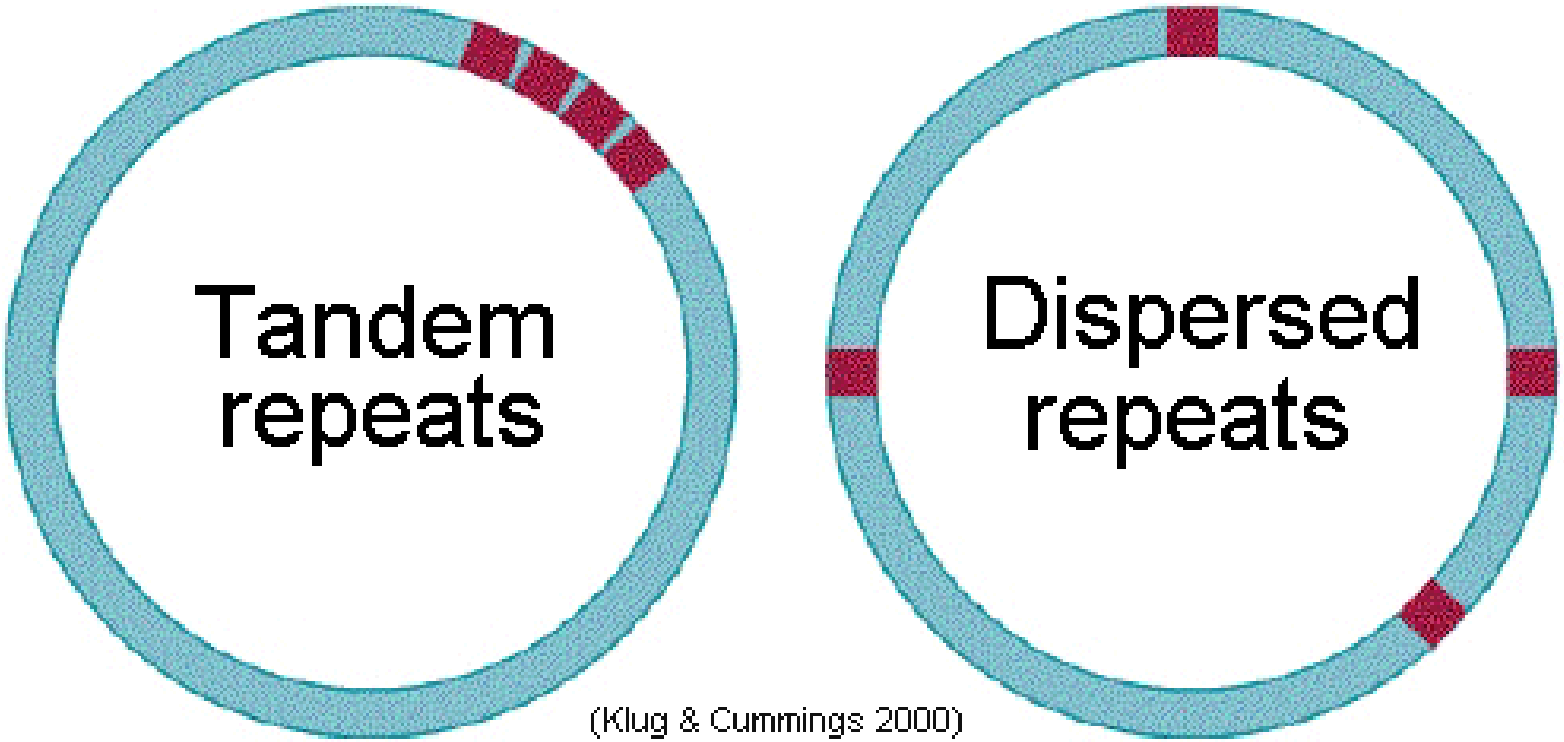
Components of the human genome



Repetitive DNA sequences



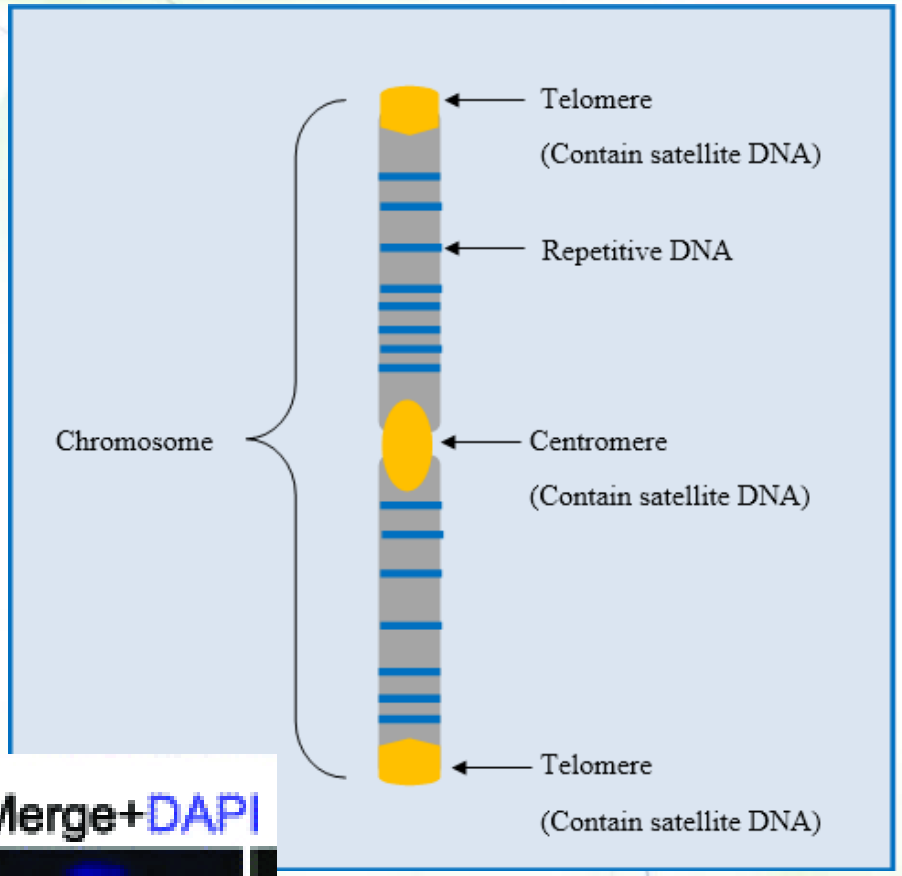
Tandem vs. dispersed

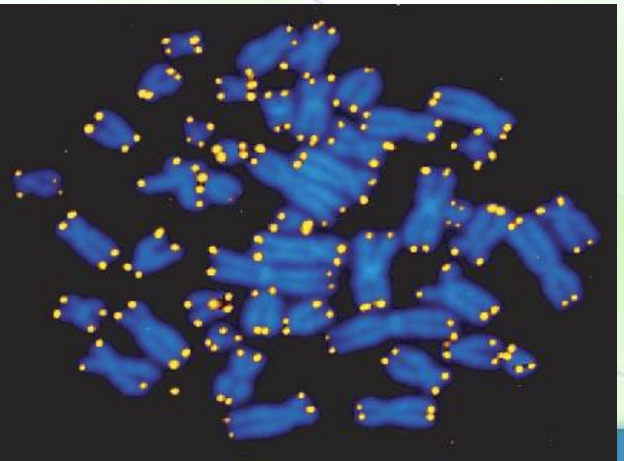


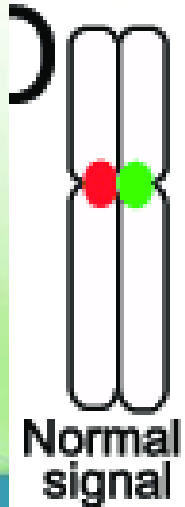
(Klug & Cummings 2000)

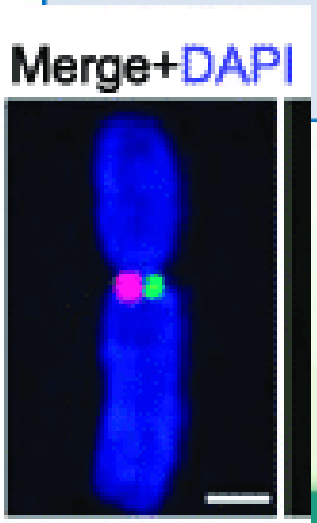
Satellite (macro-satellite) DNA

- Regions of 5-300 bp repeated 10^6-10^7 times (10% of genome)
- Centromeric repeats (171 bp) unique to each chromosome (you make chromosome-specific probes) by **fluorescent in situ hybridization**.
- Telomeric repeats



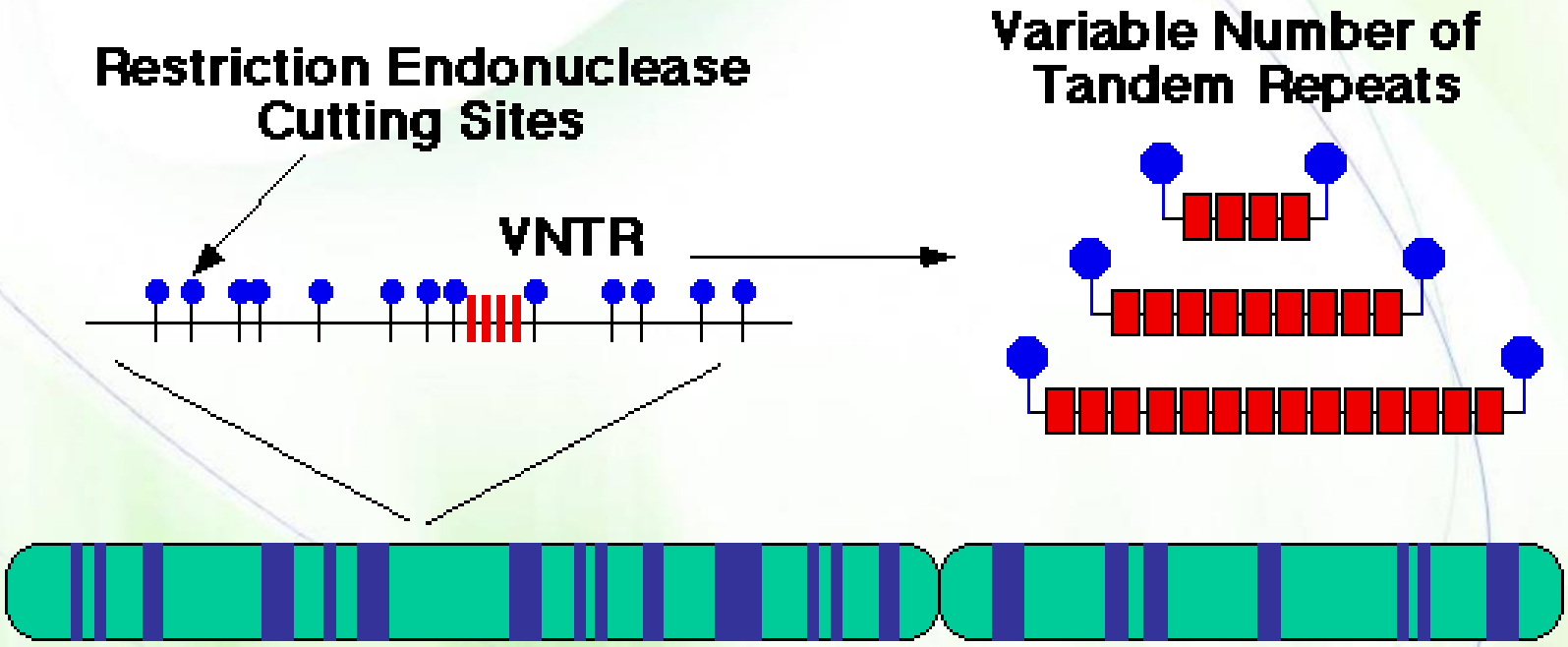






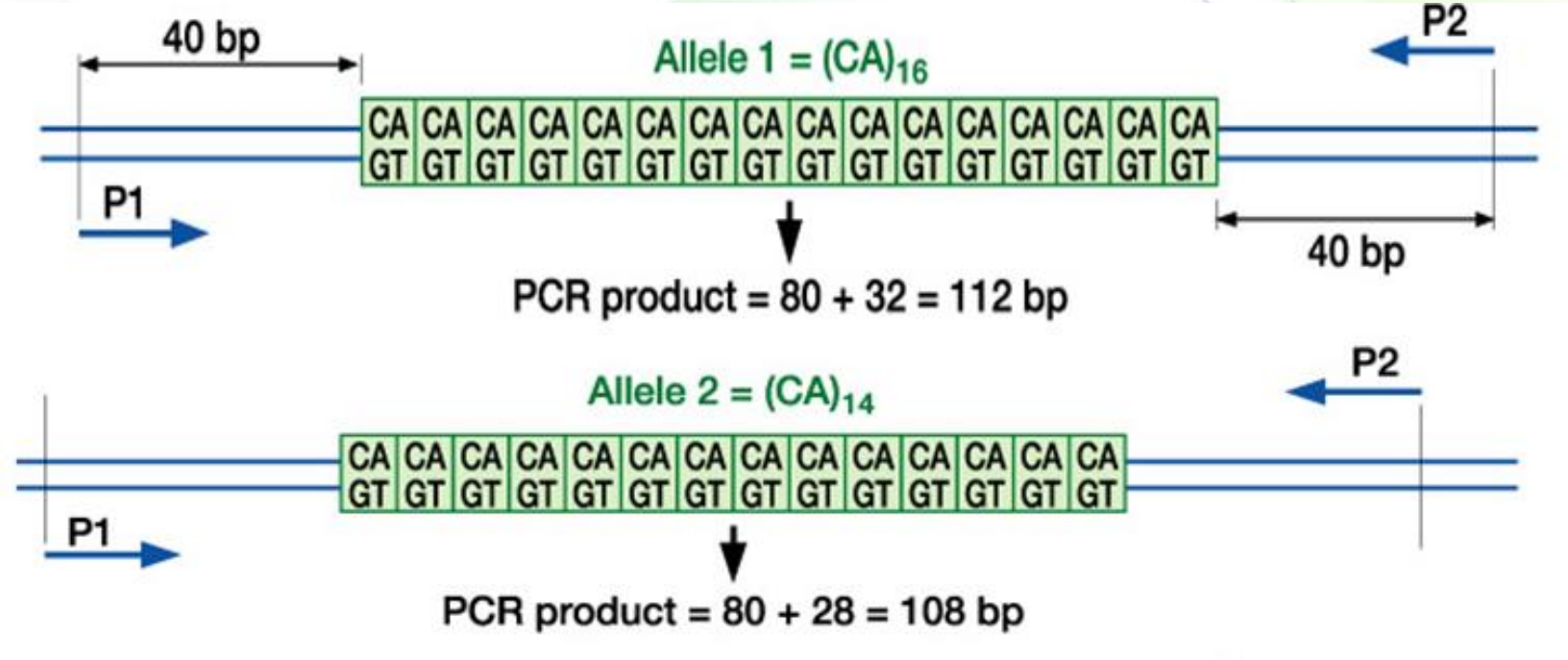
Mini-satellite DNA

- Mini satellite sequences or VNTRs (variable number of tandem repeats) of 20 to 100 bp repeated 20-50 times



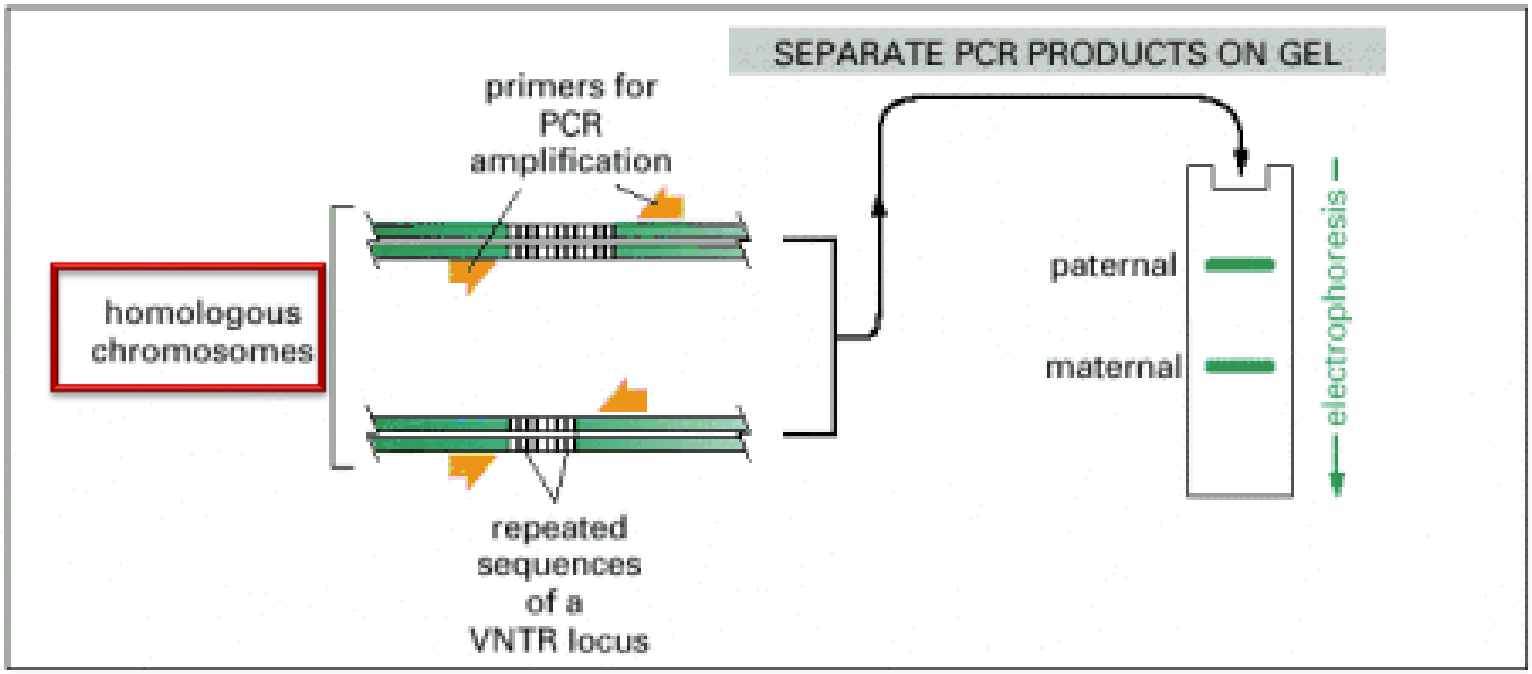
Micro-satellite DNA

- STRs (short tandem repeats) of 2 to 10 bp repeated 10-100 times

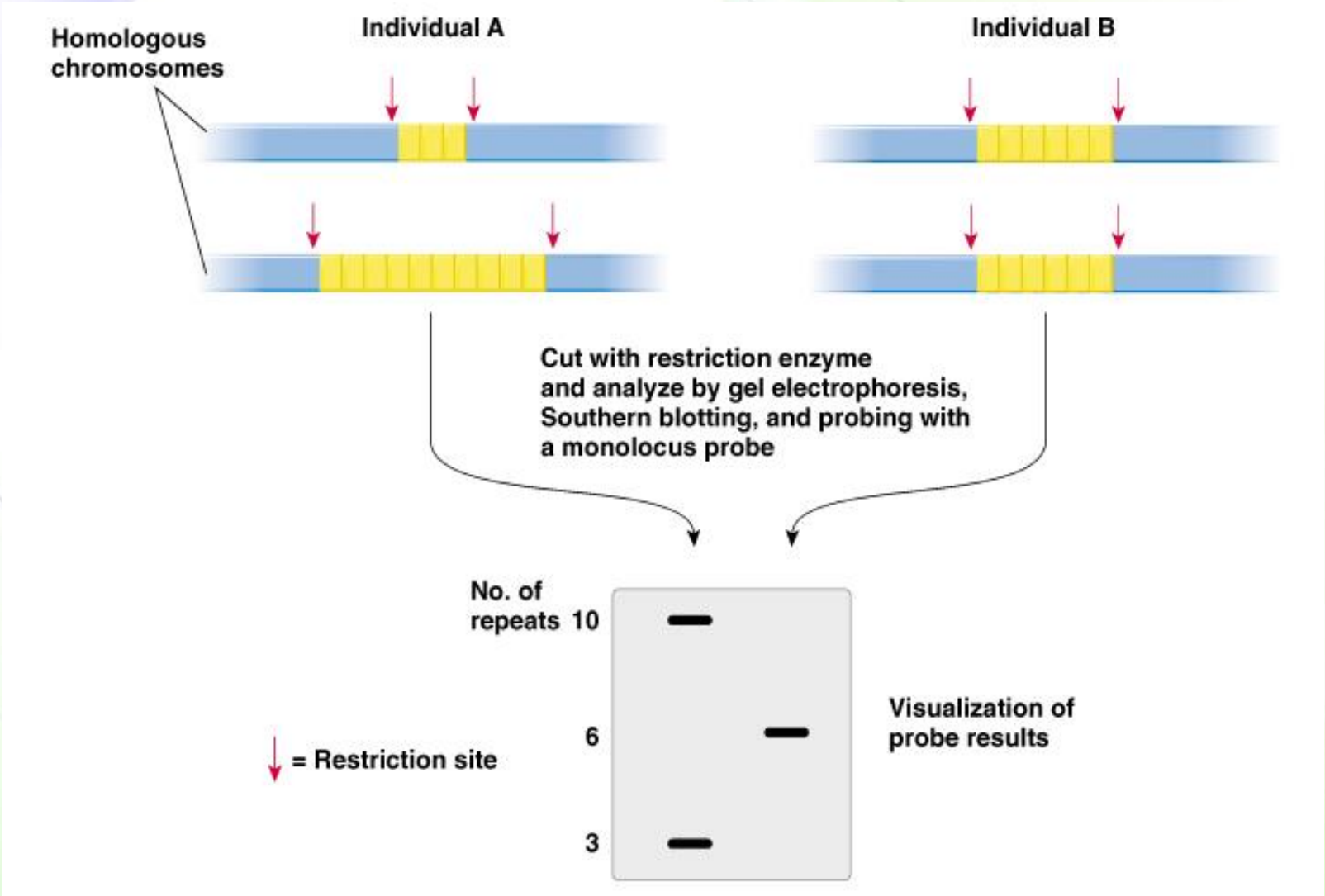


Polymorphisms of VNTR and STR

- STRs and VNTRs are highly variable among individuals (polymorphic)
  - Thus, they are useful in DNA profiling for forensic testing

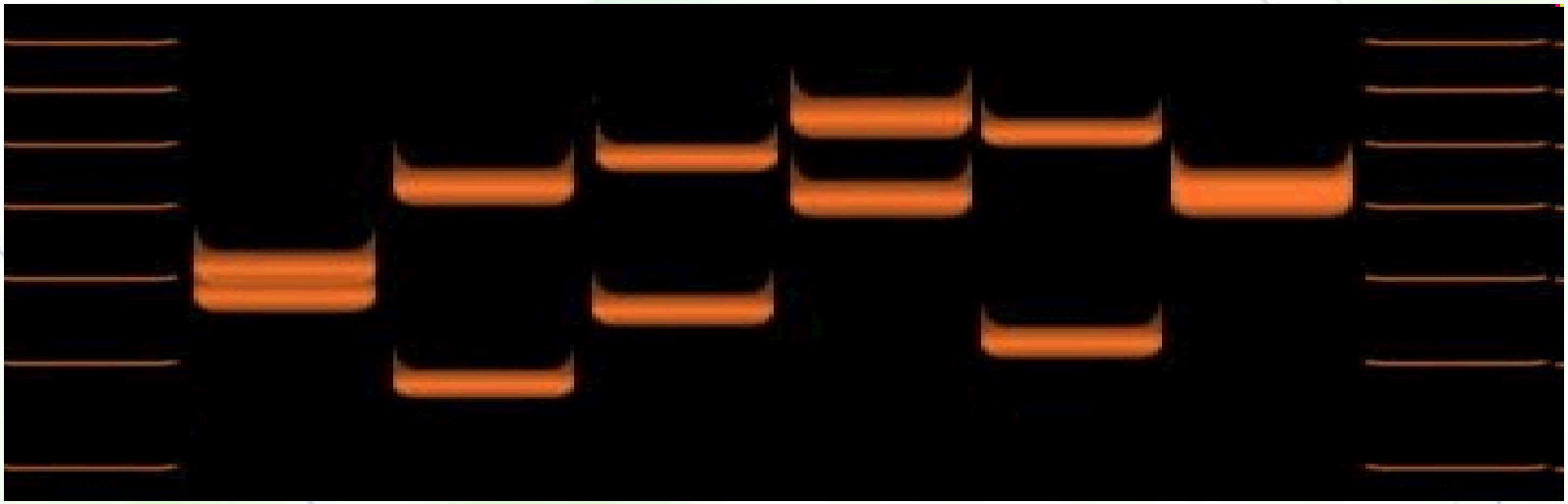


Microsatellites and VNTRs as DNA Markers



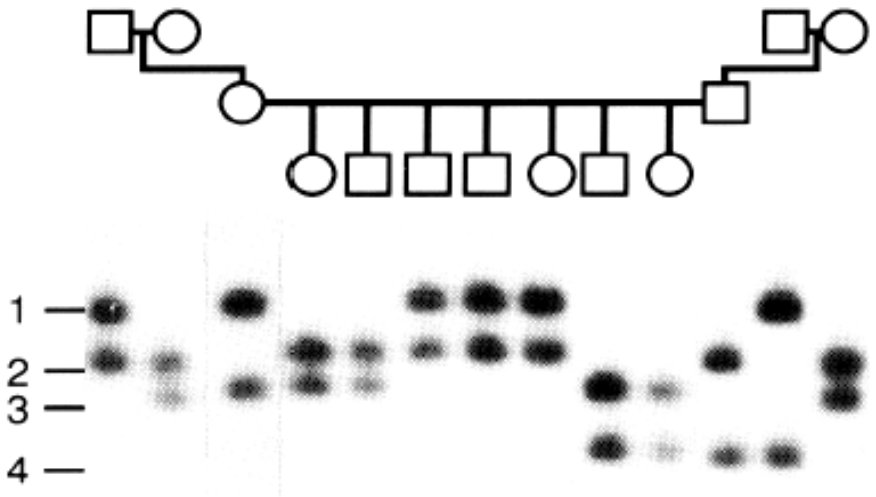
VNTR in medicine and more

- The picture below illustrates VNTR allelic length variation among 6 individuals.

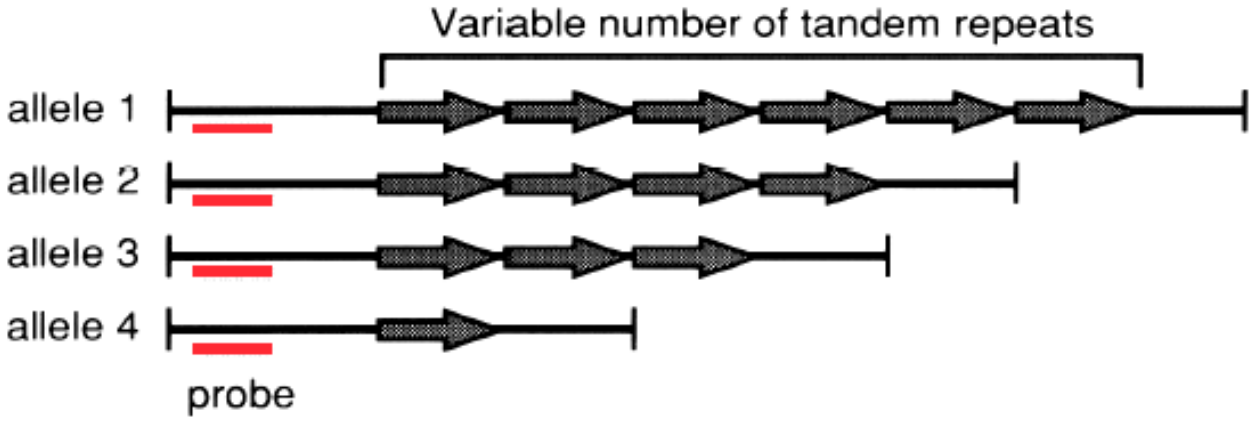


The likelihood of 2 unrelated individuals having same allelic pattern extremely improbable

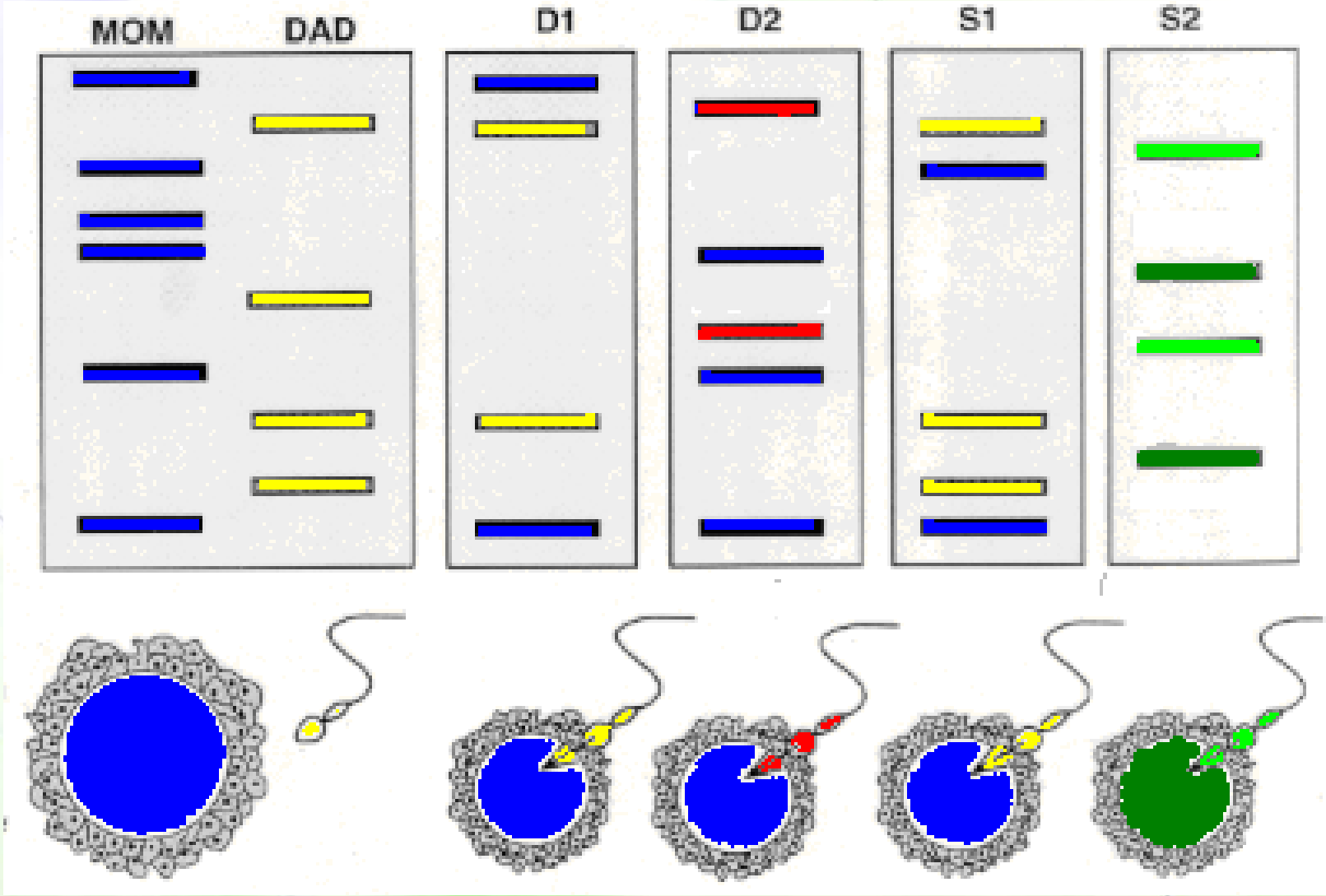
Real example



single-locus probe but multiple alleles



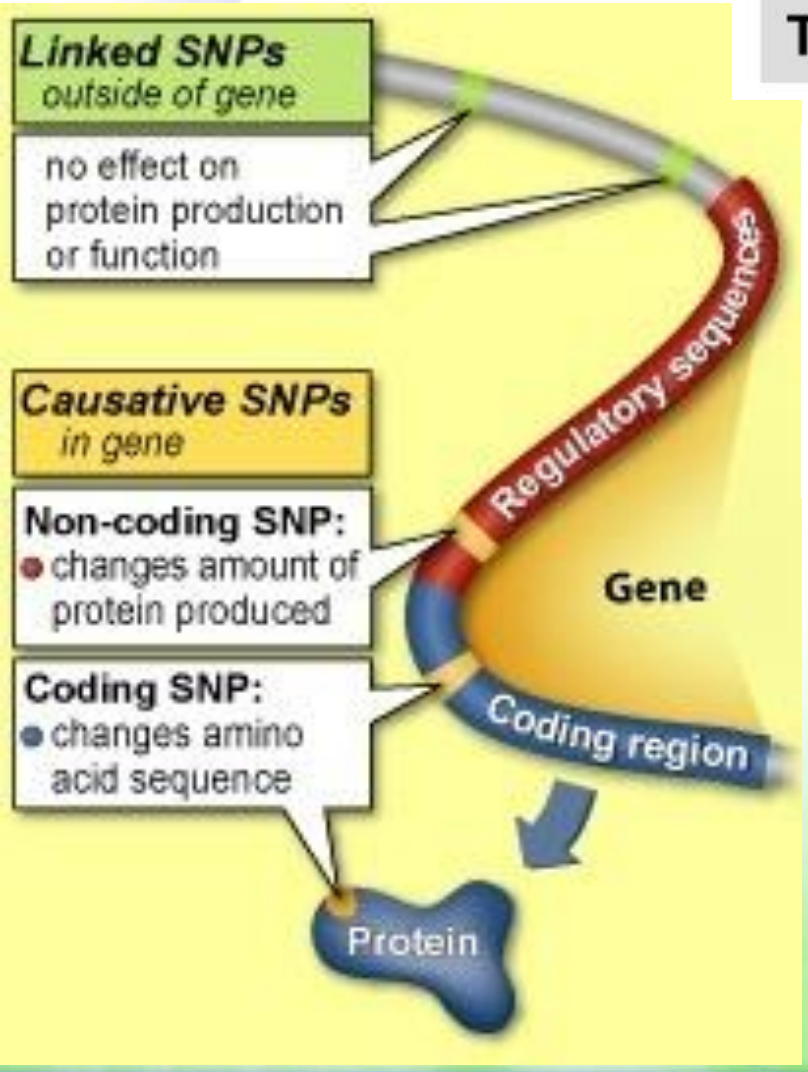
Thompson & Thompson Genetics in Medicine, p. 130, 1991


MOM
DAD
D1
D2
S1
S2

Single nucleotide polymorphism (SNPs)

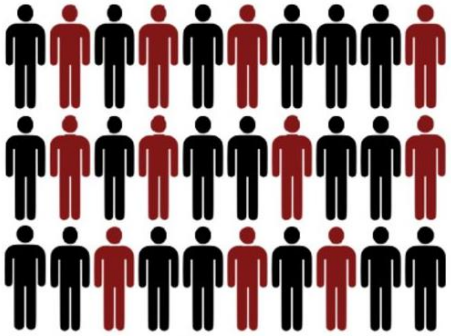
- Another source of genetic variation
- Single-nucleotide substitutions of one base for another
- Two or more versions of a sequence must each be present in at least one percent of the general population
- SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs.
  - ~10 million SNPs within the 3-billion-nucleotide human genome
  - Only 500,000 SNPs are thought to be relevant

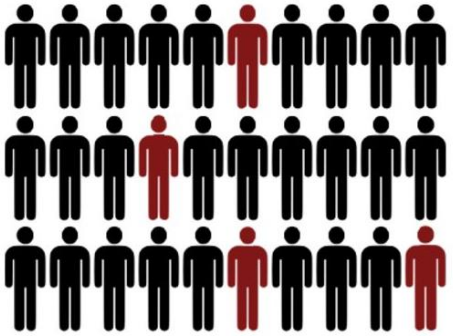
Categories of SNPs



TTGGCCAGCTGGA**C**GAGGGCGATGAC

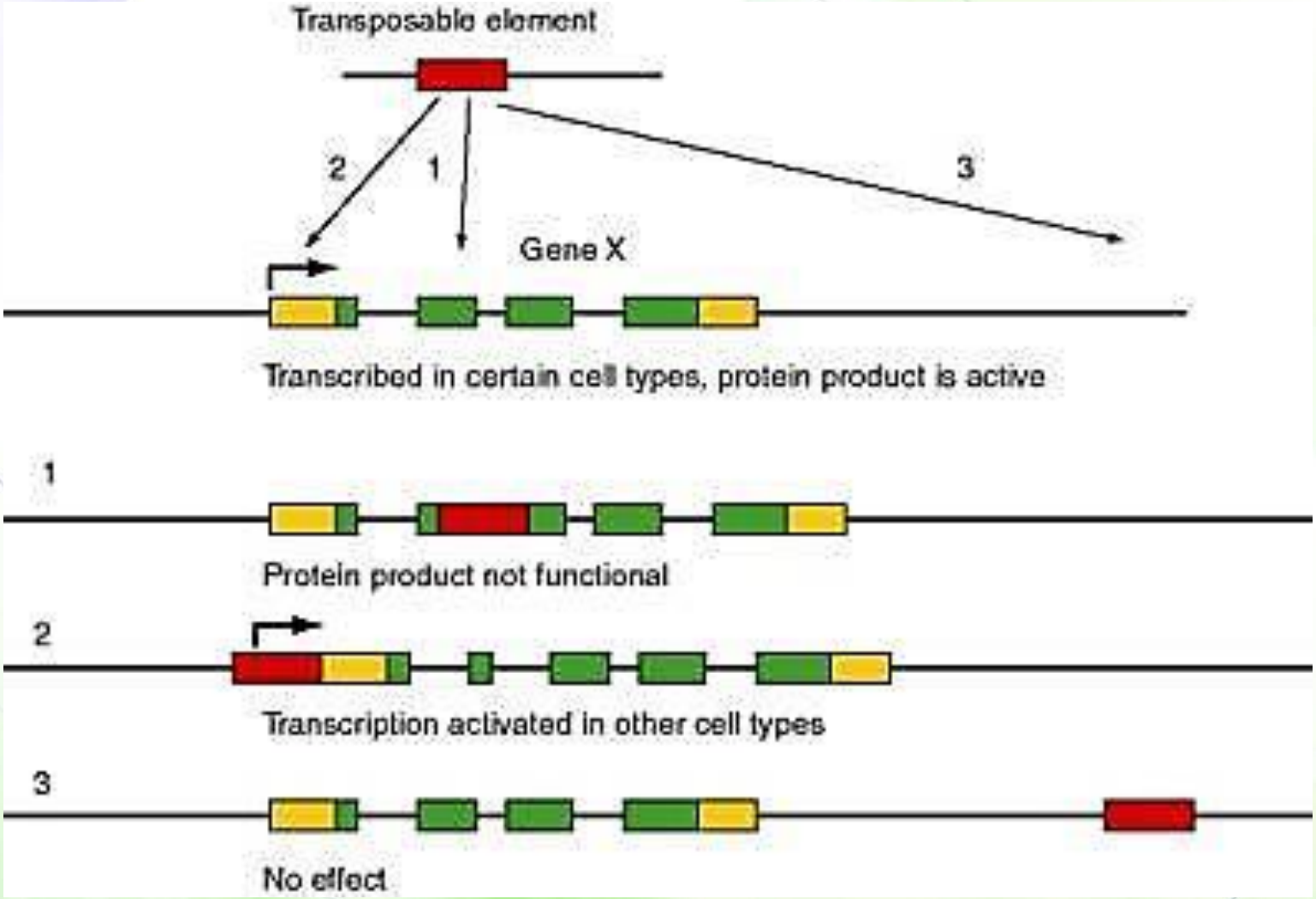
TTGGCCAGCTGGA**T**GAGGGCGATGAC





Transposons (jumping genes)

- They are segments of DNA that can move from their original position in the genome to a new location.
- Two classes:
  - DNA transposons (2-3% of human genome)
  - RNA transposons or retrotransposons (40% of human genome).
    - Long interspersed elements (LINEs)
    - Short interspersed elements (SINEs) – An example is Alu (300 bp)
- Diseases often caused by transposons include hemophilia A and B, severe combined immunodeficiency, porphyria, predisposition to cancer, and Duchenne muscular dystrophy.


Transposable element
2
1
3
Gene X
Transcribed in certain cell types, protein product is active
1
Protein product not functional
2
Transcription activated in other cell types
3
No effect