DNA sequencing and PCR

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

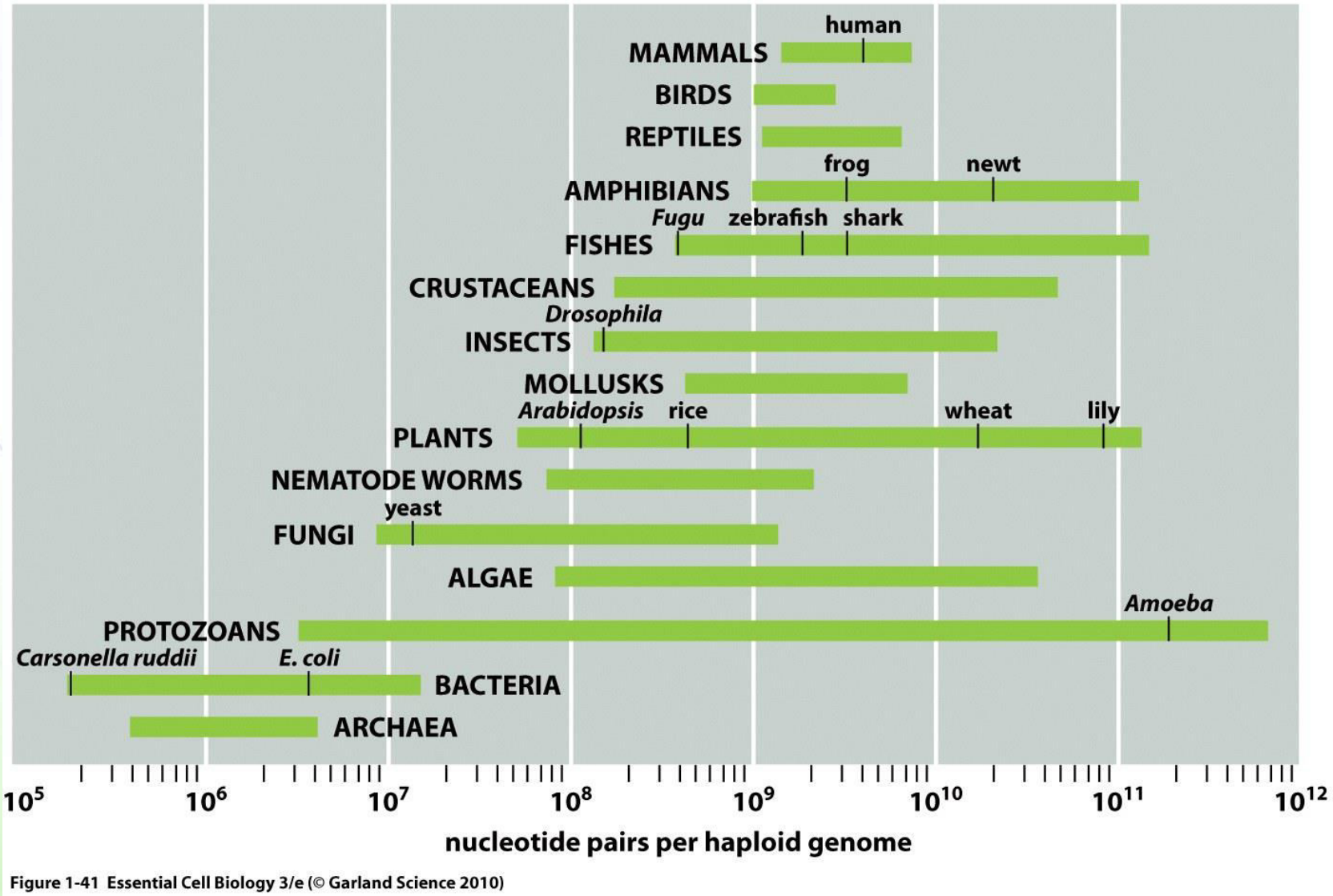
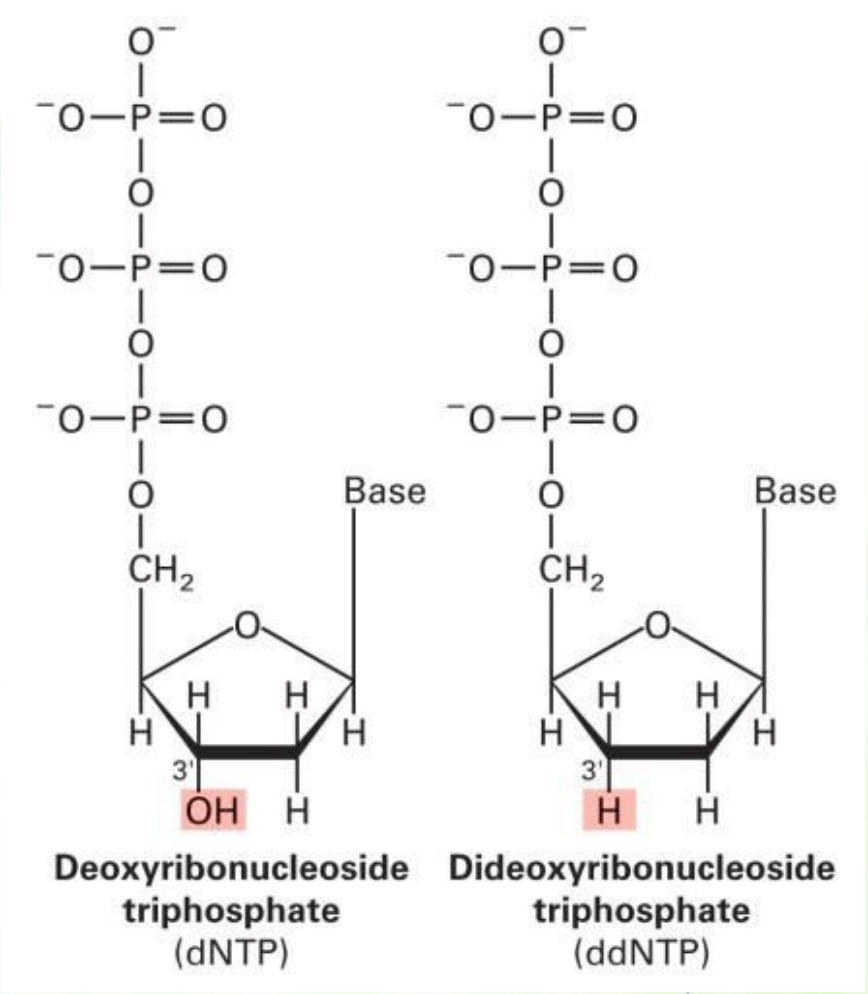


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

Method of DNA sequencing

- Based on premature termination of DNA synthesis by dideoxynucleotides

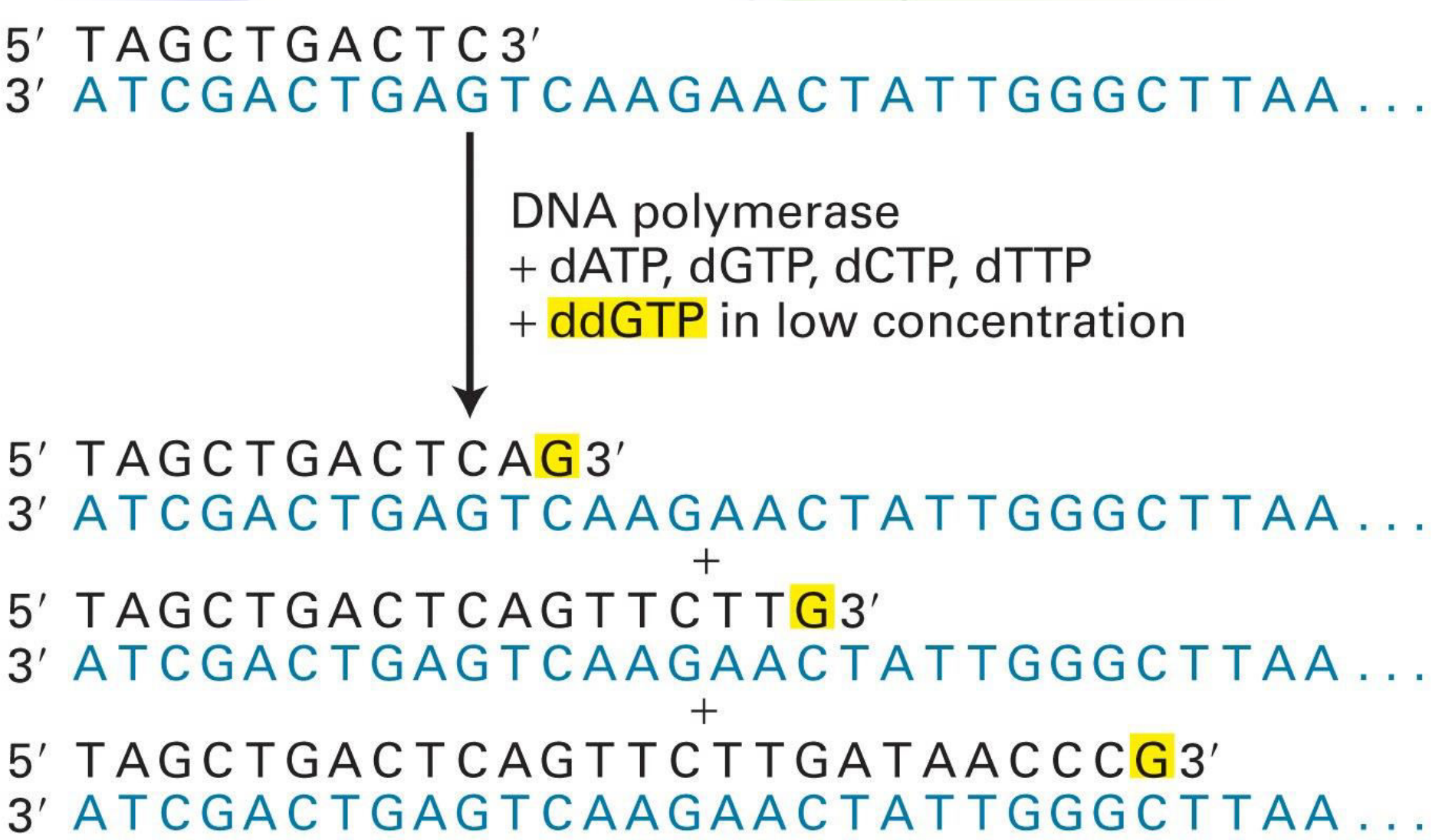


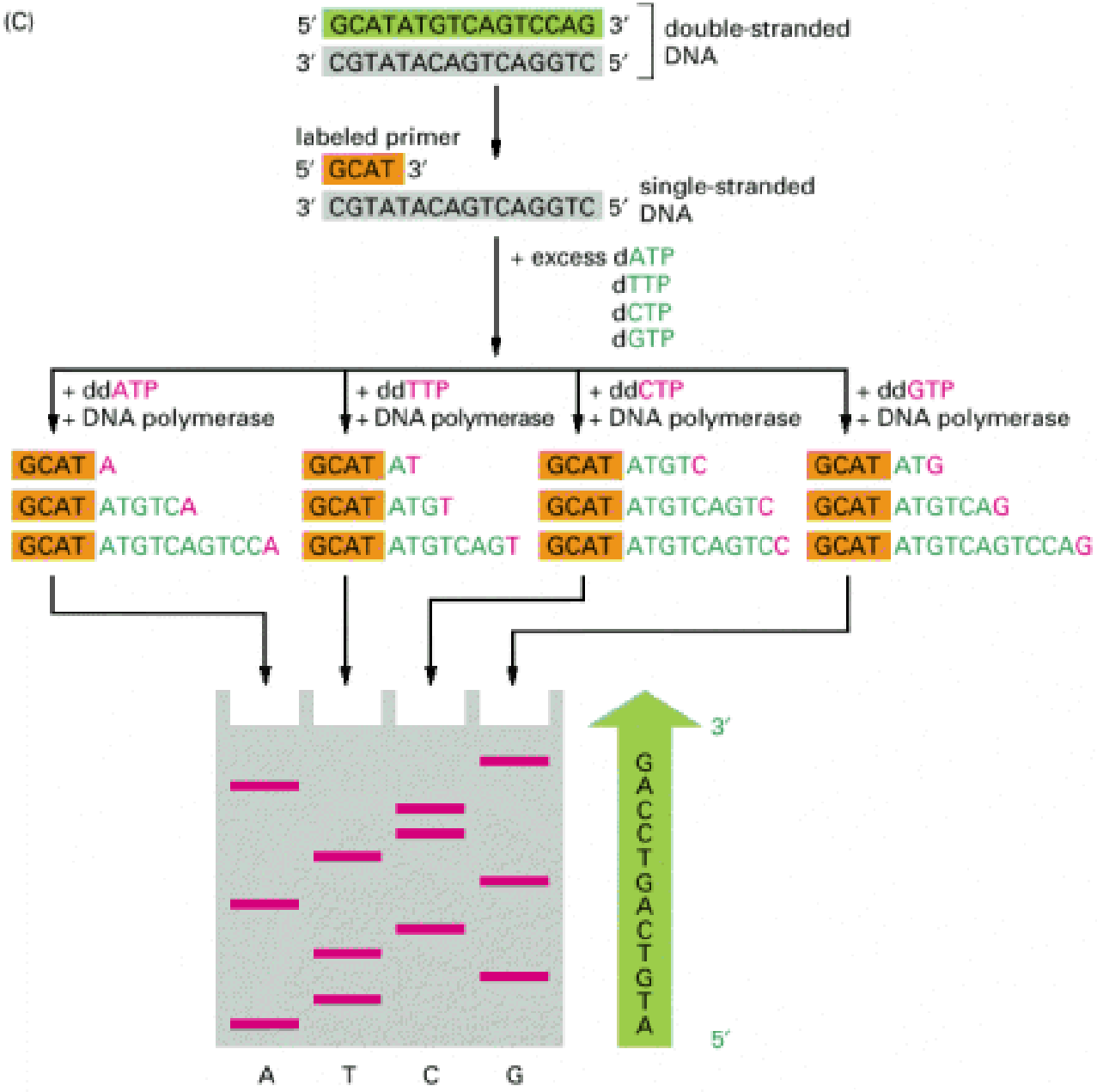
The process...

- 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

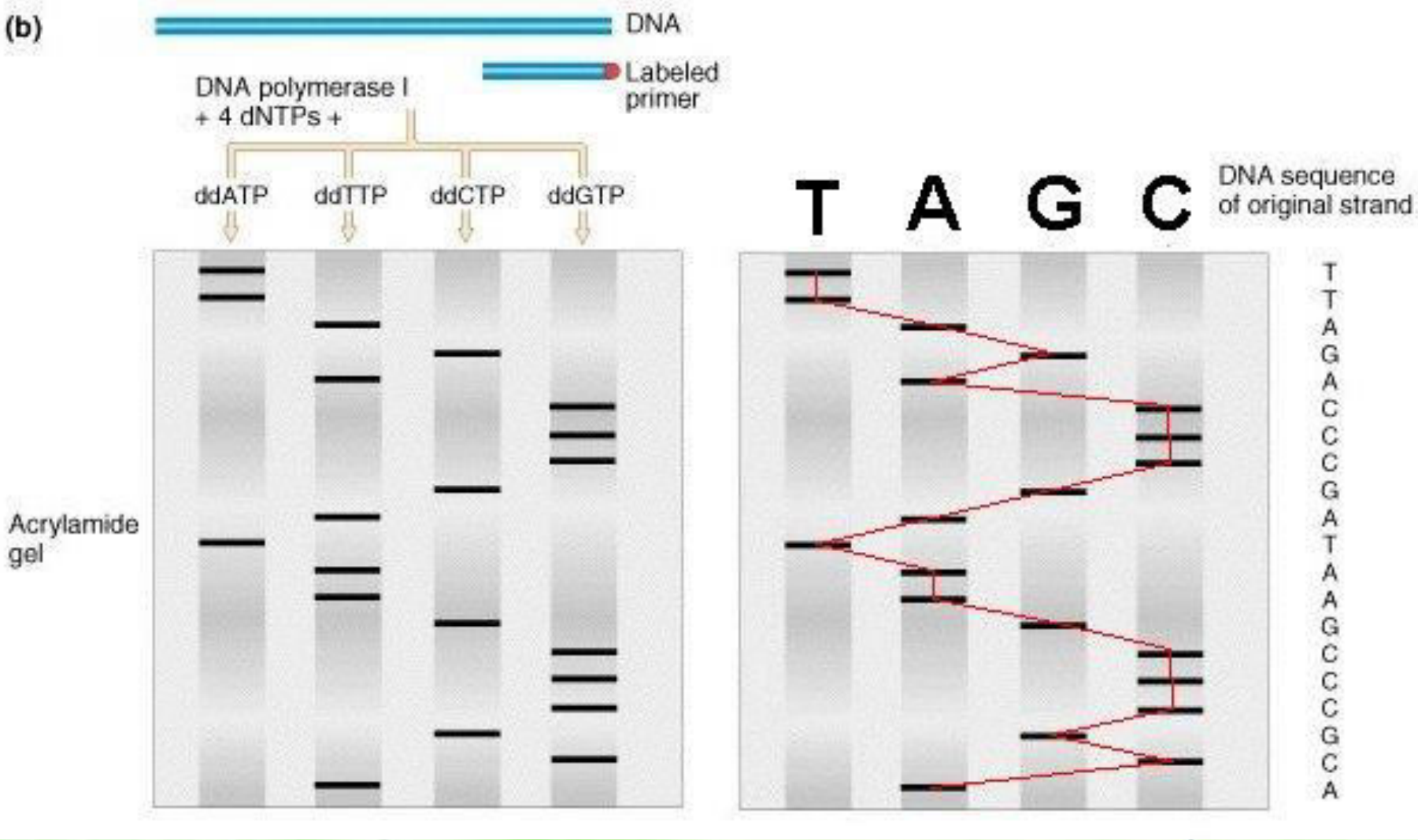
- A series of labeled DNA molecules are generated, each terminated by the dideoxynucleotide in each reaction
- These fragments of DNA are then separated according to size by gel electrophoresis and detected by exposure of the gel to X-ray film
- The size of each fragment is determined by its terminal 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′ TAGCTGACTCAG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .
+
5′ TAGCTGACTCAGTTCTTGATAACCCG 3′
3′ ATCGACTGAGTCAAGAACTATTGGGCTTAA . . .



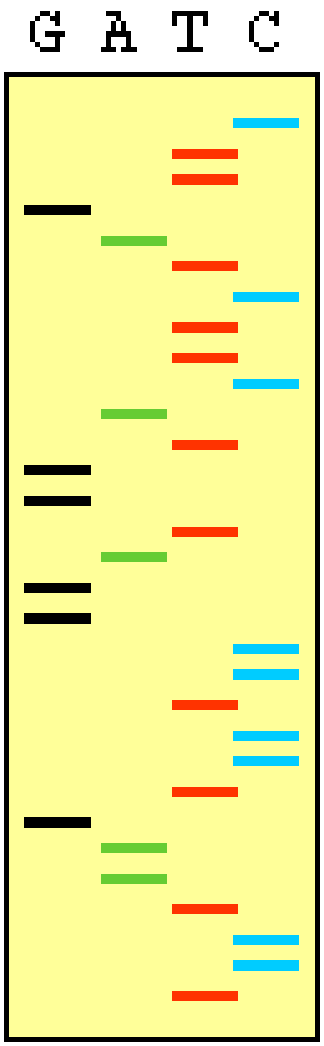
DNA sequence reading directly from the bottom of the gel upward, is

ATGTCAGTCCAG
1 12

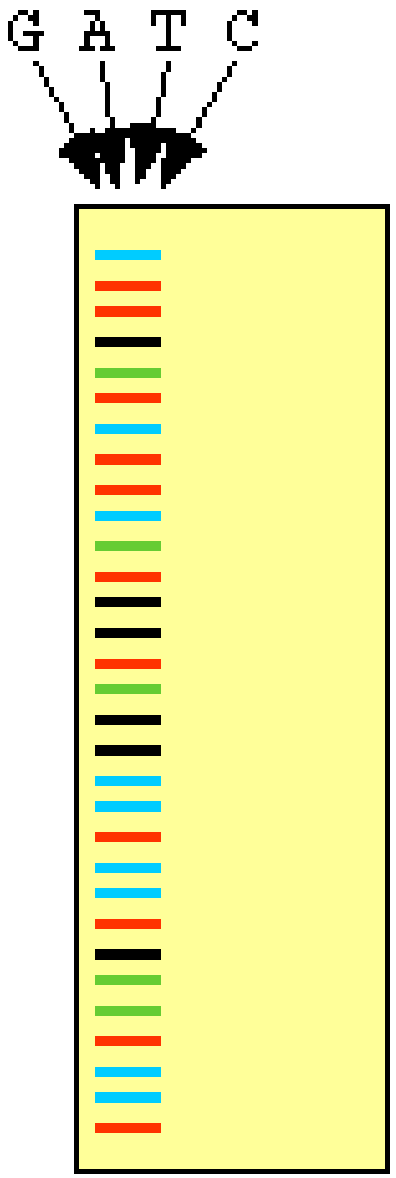

(b)
DNA
Labeled primer
DNA polymerase I
+ 4 dNTPs +
ddATP
ddTTP
ddCTP
ddGTP
Acrylamide gel
T A G C
DNA sequence of original strand
T T A G A C C C G A T A A G C C C G C A

Fluorescence-based DNA sequencing

- Large-scale DNA sequencing is frequently performed using automated systems using fluorescence-based reactions using labeled ddNTPs
- In this case, all four terminators can then be placed in a single tube, and only one reaction is necessary
- The reactions are run into one lane on a gel and a machine is used to scan the lane with a laser



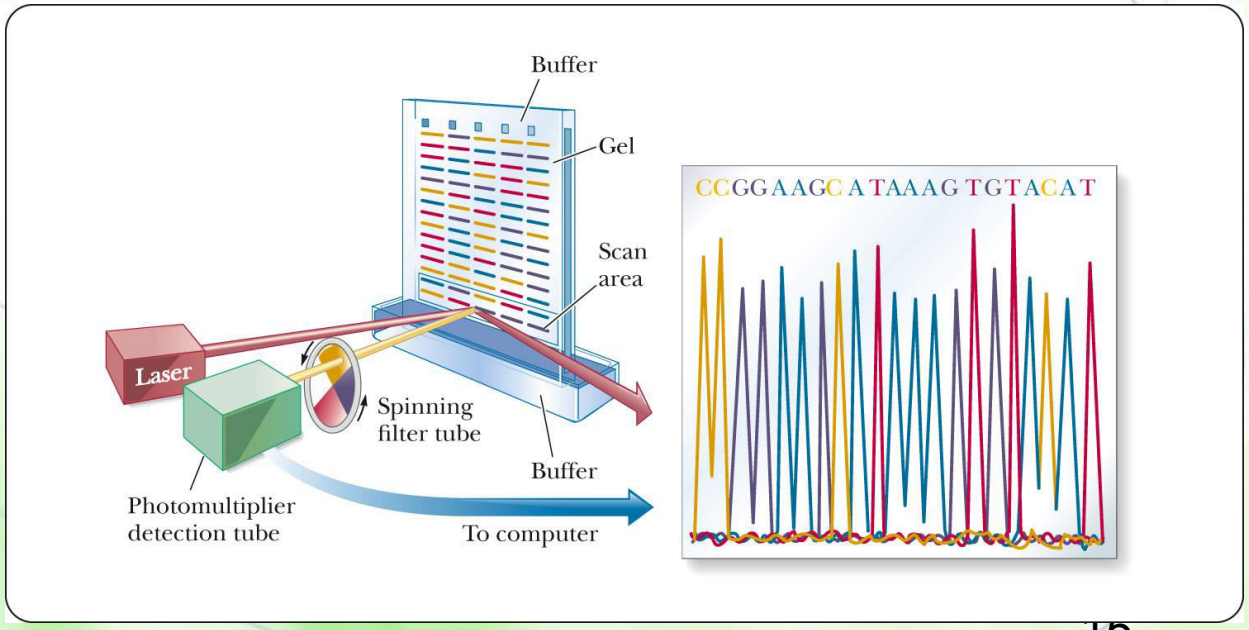
Here's what the products would look like in separate gel lanes.

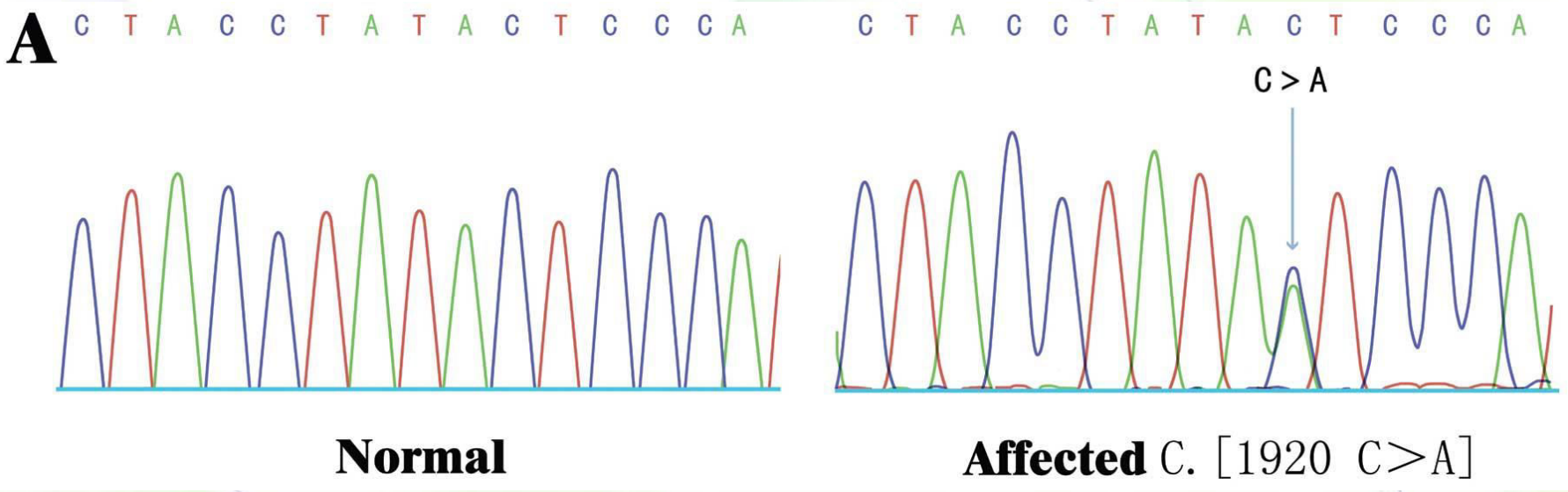


Here's what the products would look like in a single gel lanes.

Detection of fragments

- The wavelength of fluorescence can be interpreted by the machine as an indication of which reaction (ddG, ddA, ddT, or ddC) the product came from
- The fluorescence output is stored in the form of chromatograms

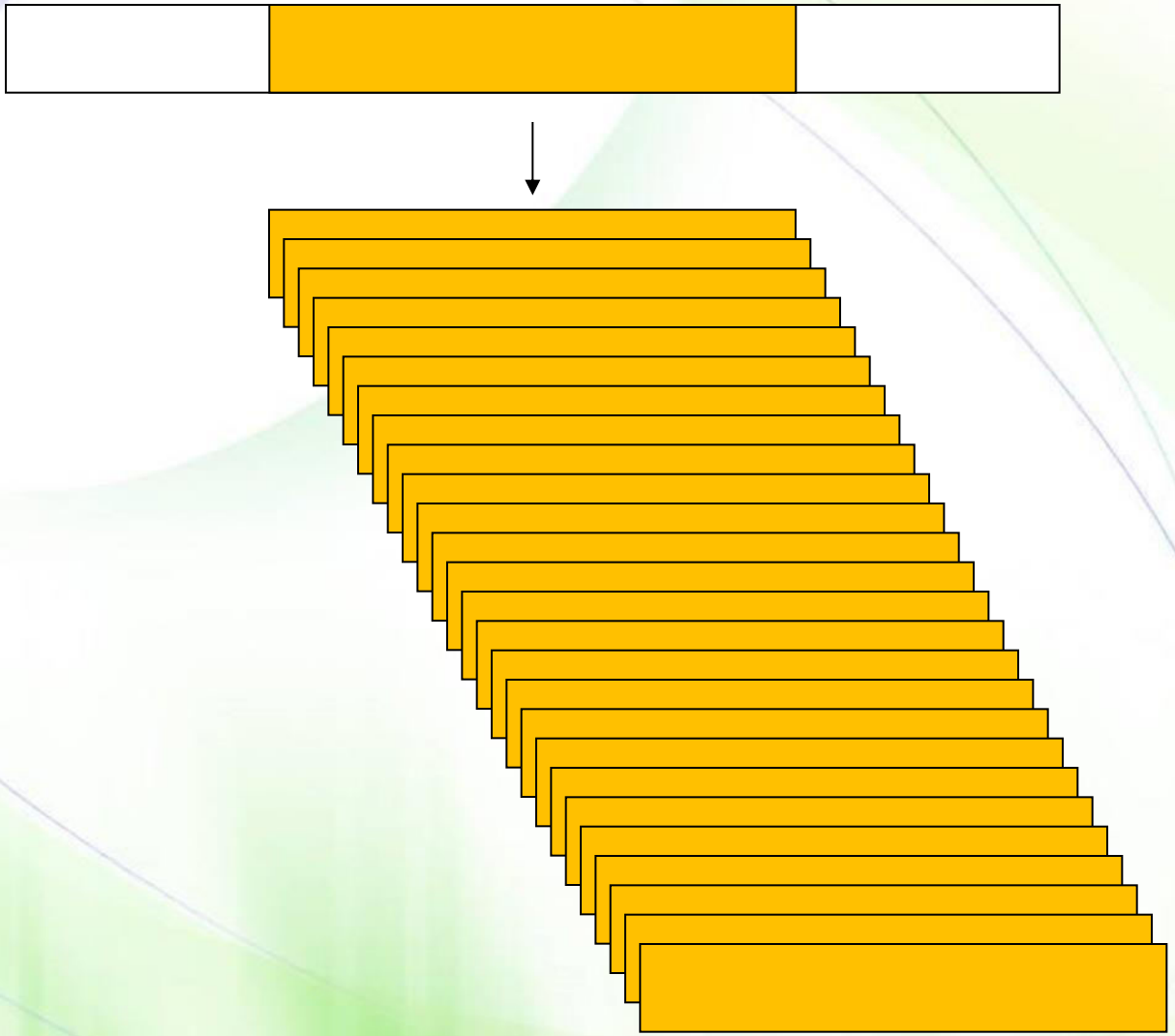




What does it mean?

Polymerase Chain Reaction

- In 1984, Kary Mullis devised a method called the polymerase chain reaction (PCR) for amplifying specific DNA sequences
- 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
- The PCR method is extremely sensitive; it can detect a single DNA molecule in a sample



Components of PCR reaction

- A pair of primers that hybridize to the target DNA.
  - These primers should be specific for the target sequence and which are often about 15-25 nucleotides long. The region between the primers is amplified
- All four deoxyribonucleoside triphosphates (dNTPs: dATP, dCTP, dGTP and dTT)
- A heat-stable DNA polymerase

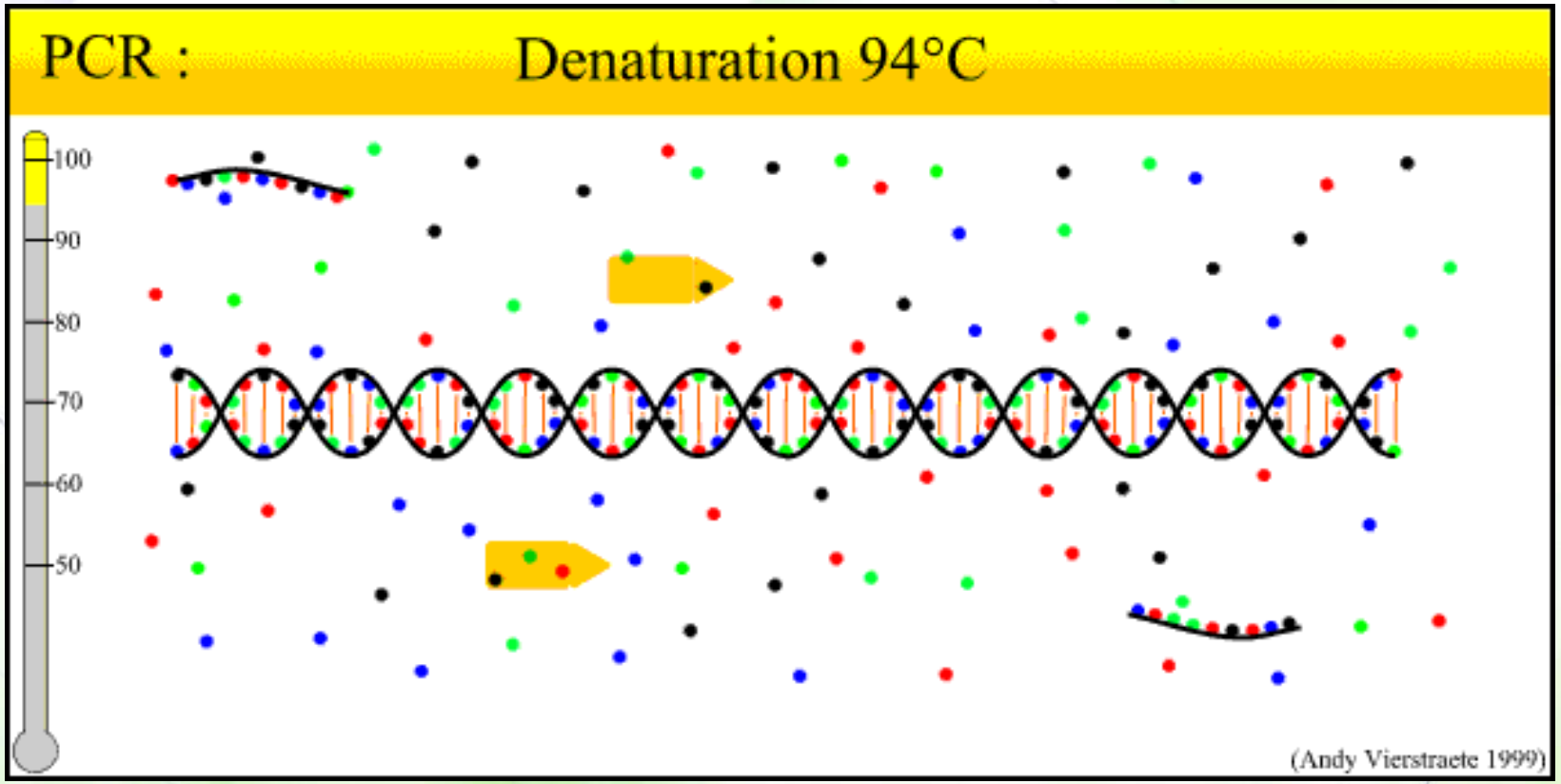
DNA polymerases

- 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 94°C

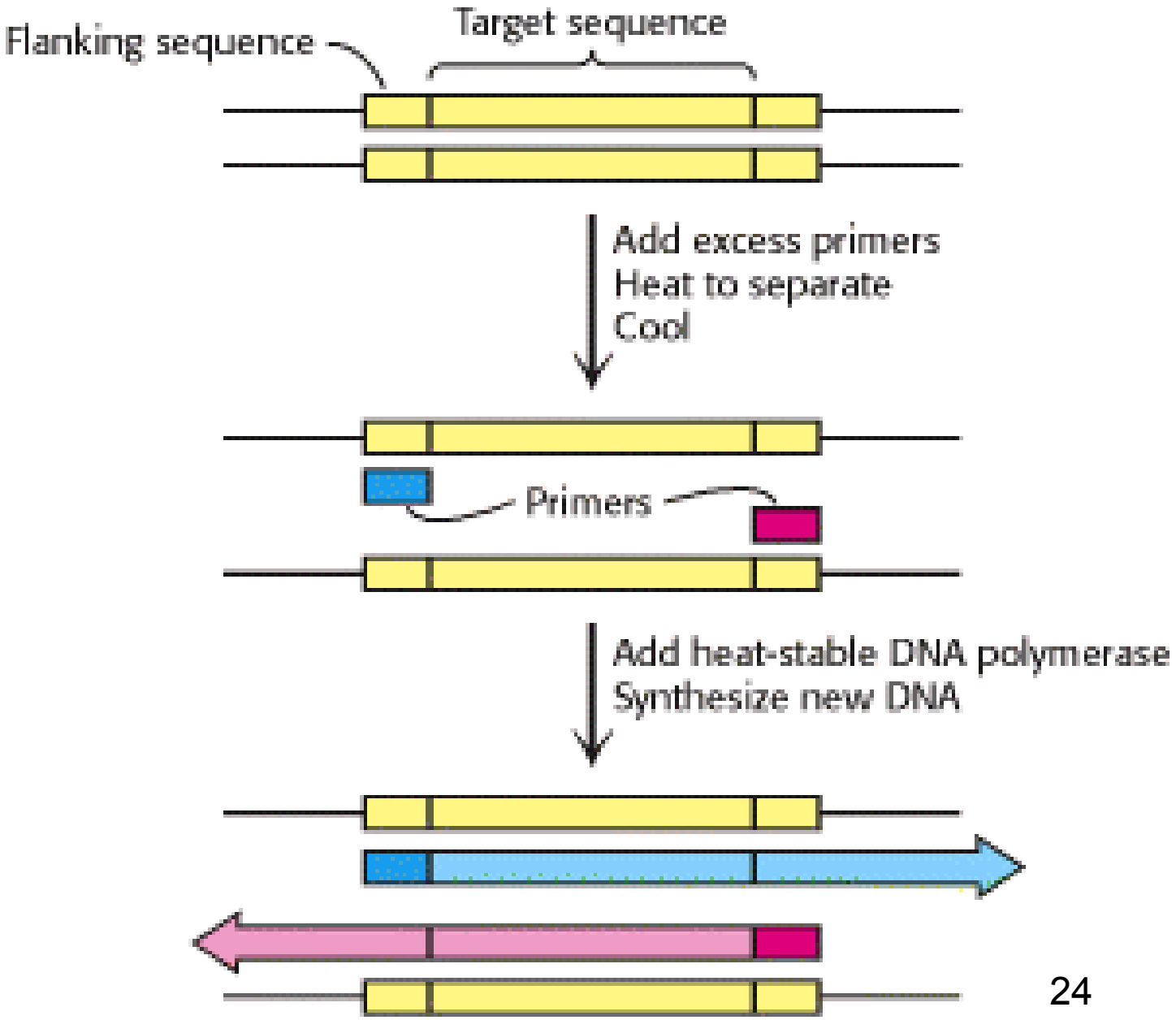
PCR cycle

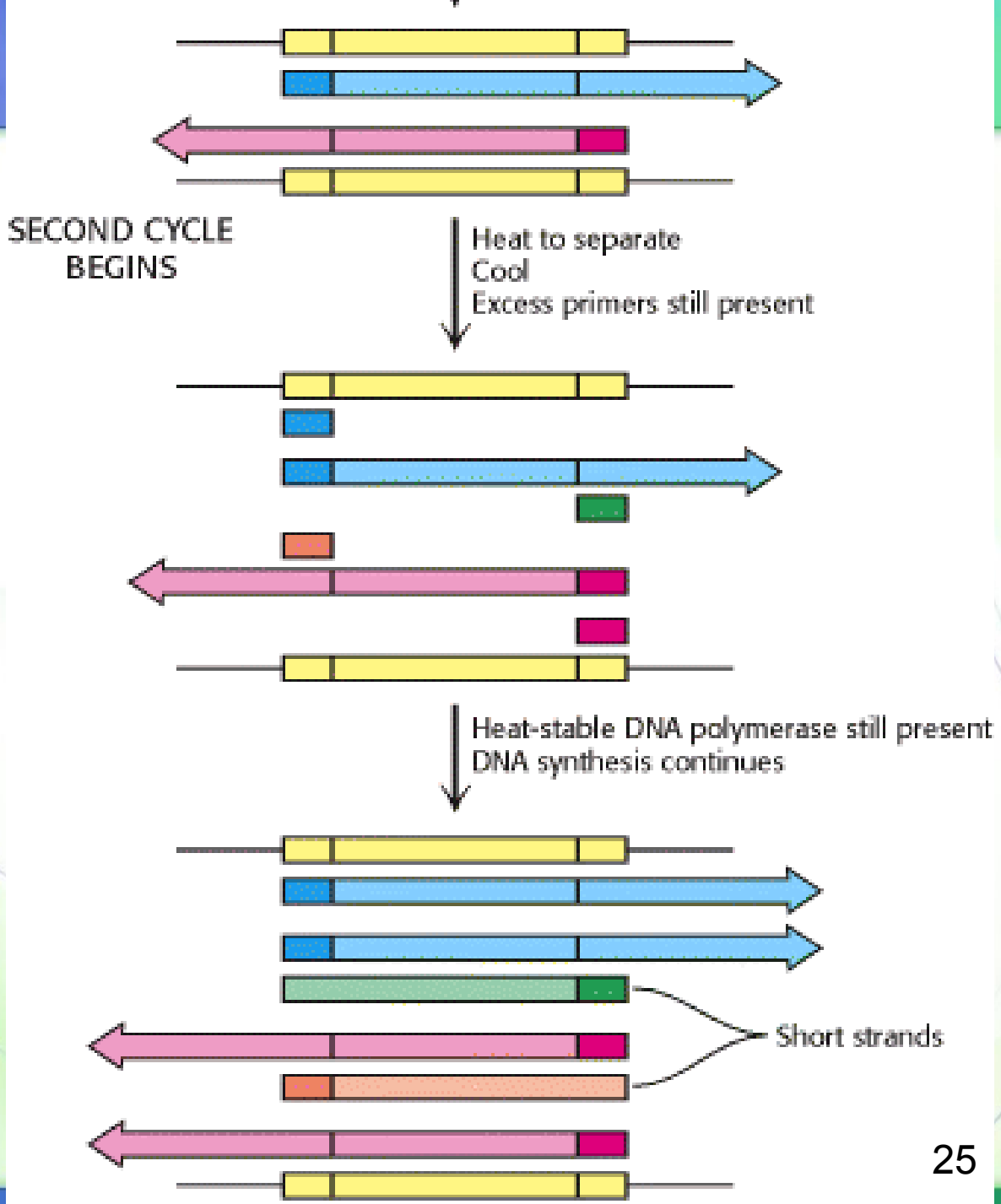
- Denaturation, typically at about 93-95°C. At this temperature the hydrogen bonds that hold together the two polynucleotides of the double helix are broken, so the target DNA becomes denatured into single-stranded molecules
- Reannealing at temperatures usually from about 50°C to 70°C where the primers anneal to the DNA
- DNA synthesis, typically at about 70-75°C, the optimum for Taq polymerase

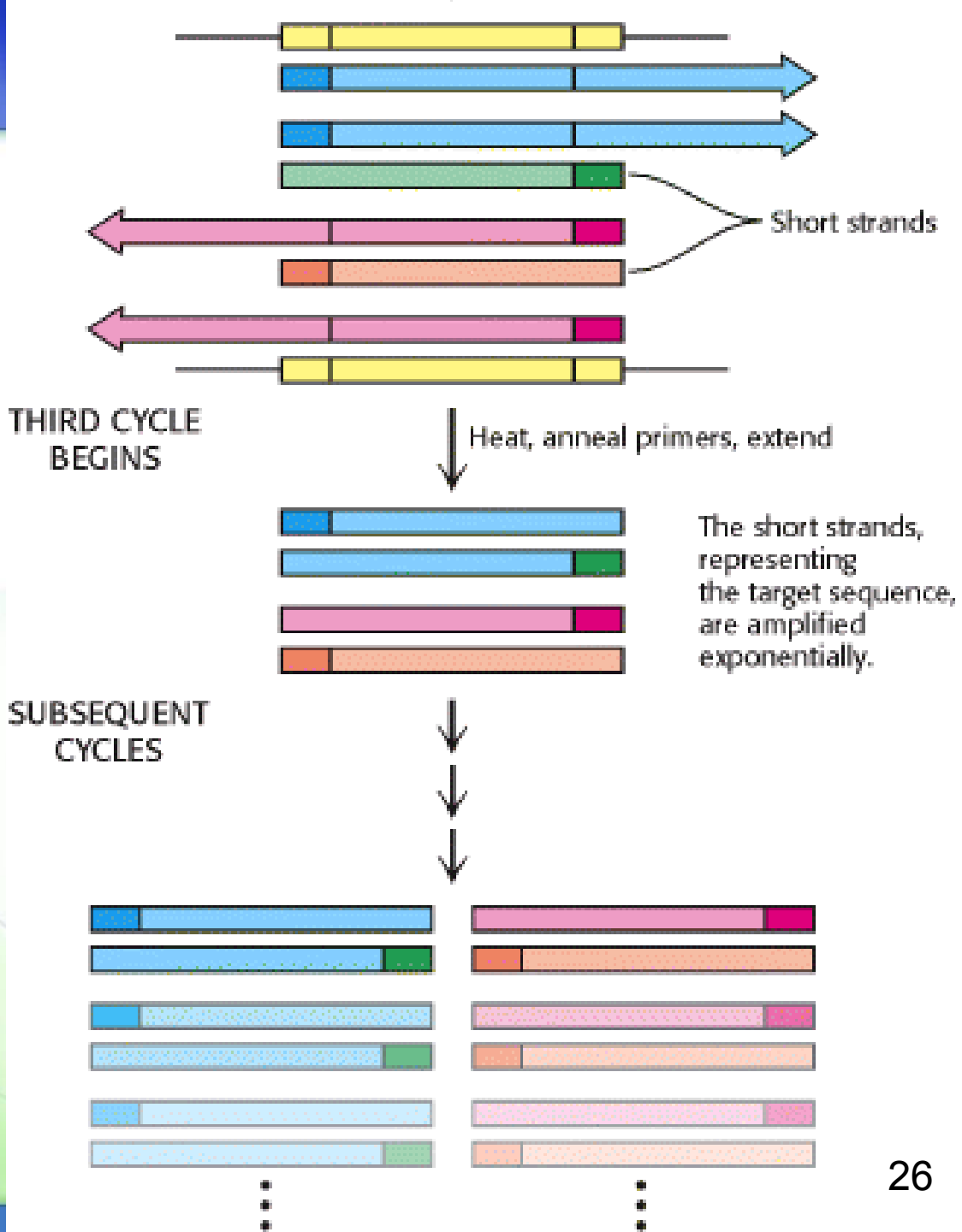

PCR : Denaturation 94°C
100
90
80
70
60
50
(Andy Vierstraete 1999)

A template sequence with 5 ATT repeats.

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

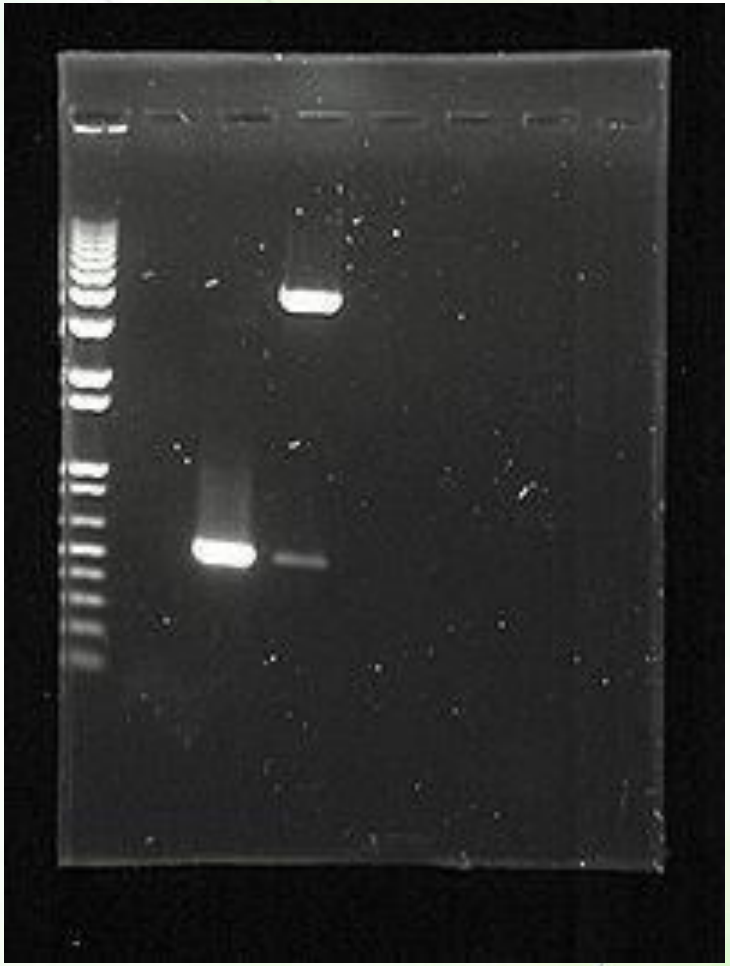

Short strands
THIRD CYCLE BEGINS
Heat, anneal primers, extend
The short strands, representing the target sequence, are amplified exponentially.
SUBSEQUENT CYCLES

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

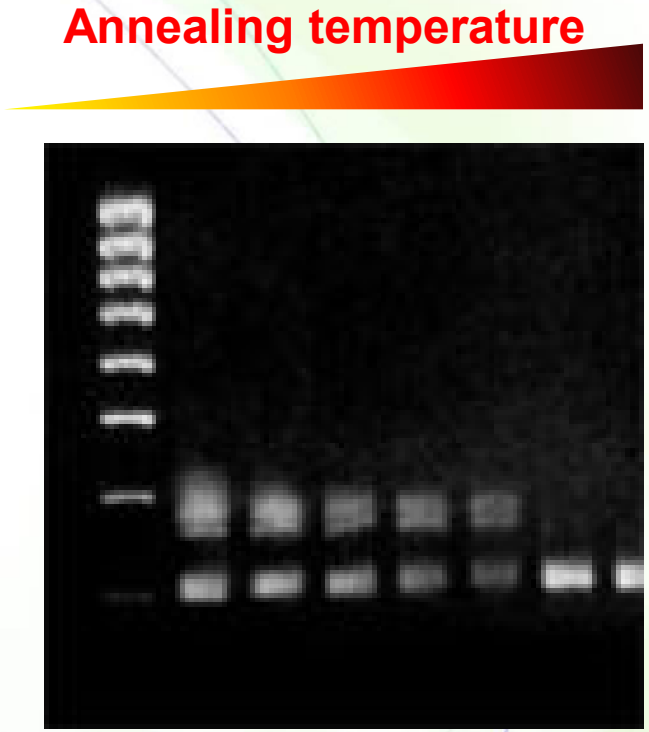
Detection of DNA fragments

- 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?



Advantages

- Easy, fast, sensitive, robust
- Discovery of gene families
- Disease diagnosis

Disadvantages

- Primers must be known
- Contamination
- Product length is limited (usually <5 Kb)
- Accuracy is an issue
- Not quantitative

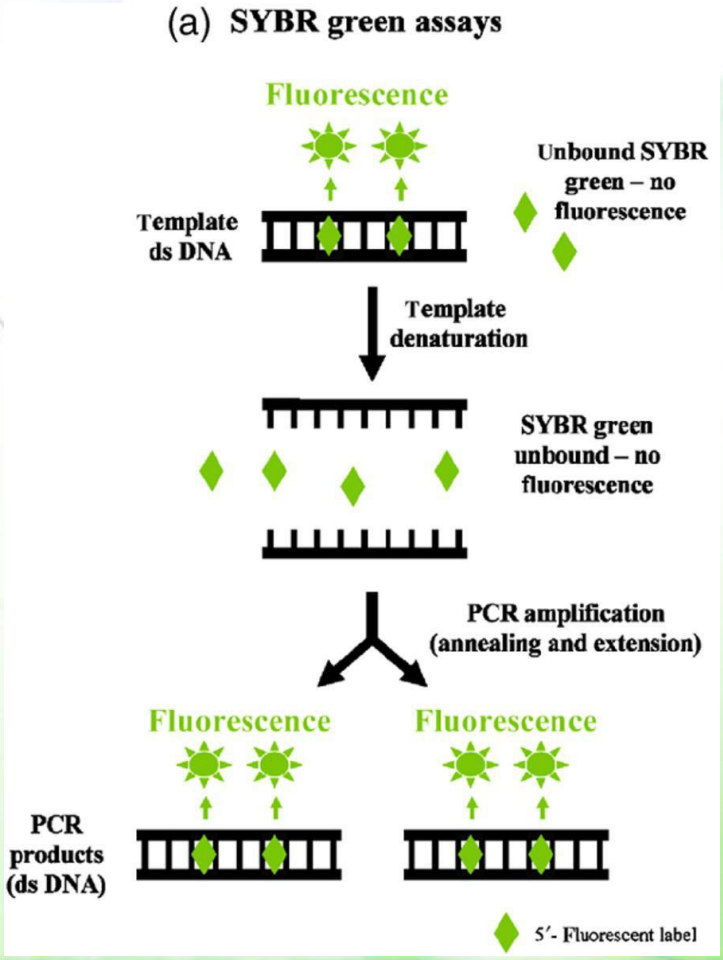
Exploring genomes

Forensic medicine

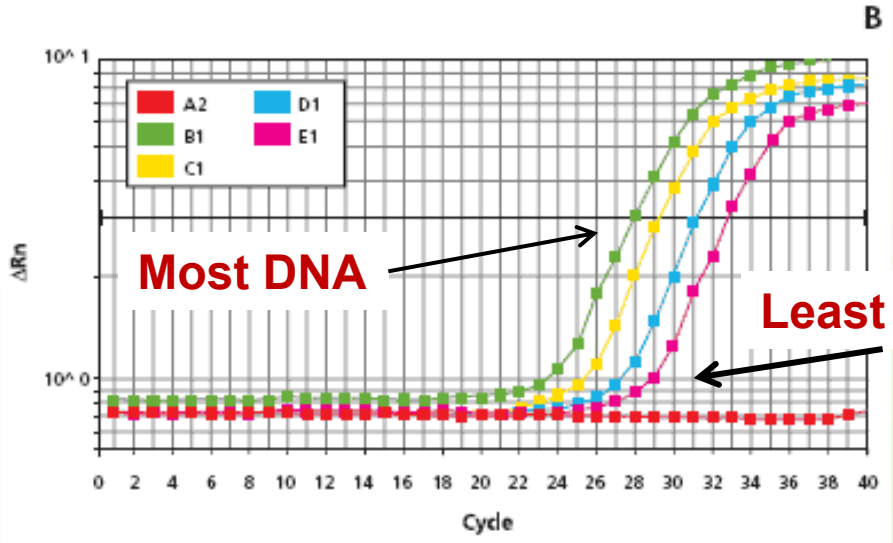
- Viral and bacterial load: the quantity of virus in a given volume. How?
  - Quantitative PCR

- Paternity and criminal cases. Why?
  - An individual DNA profile is highly distinctive because many genetic loci are highly variable within a population.

Quantitative PCR (qPCR)

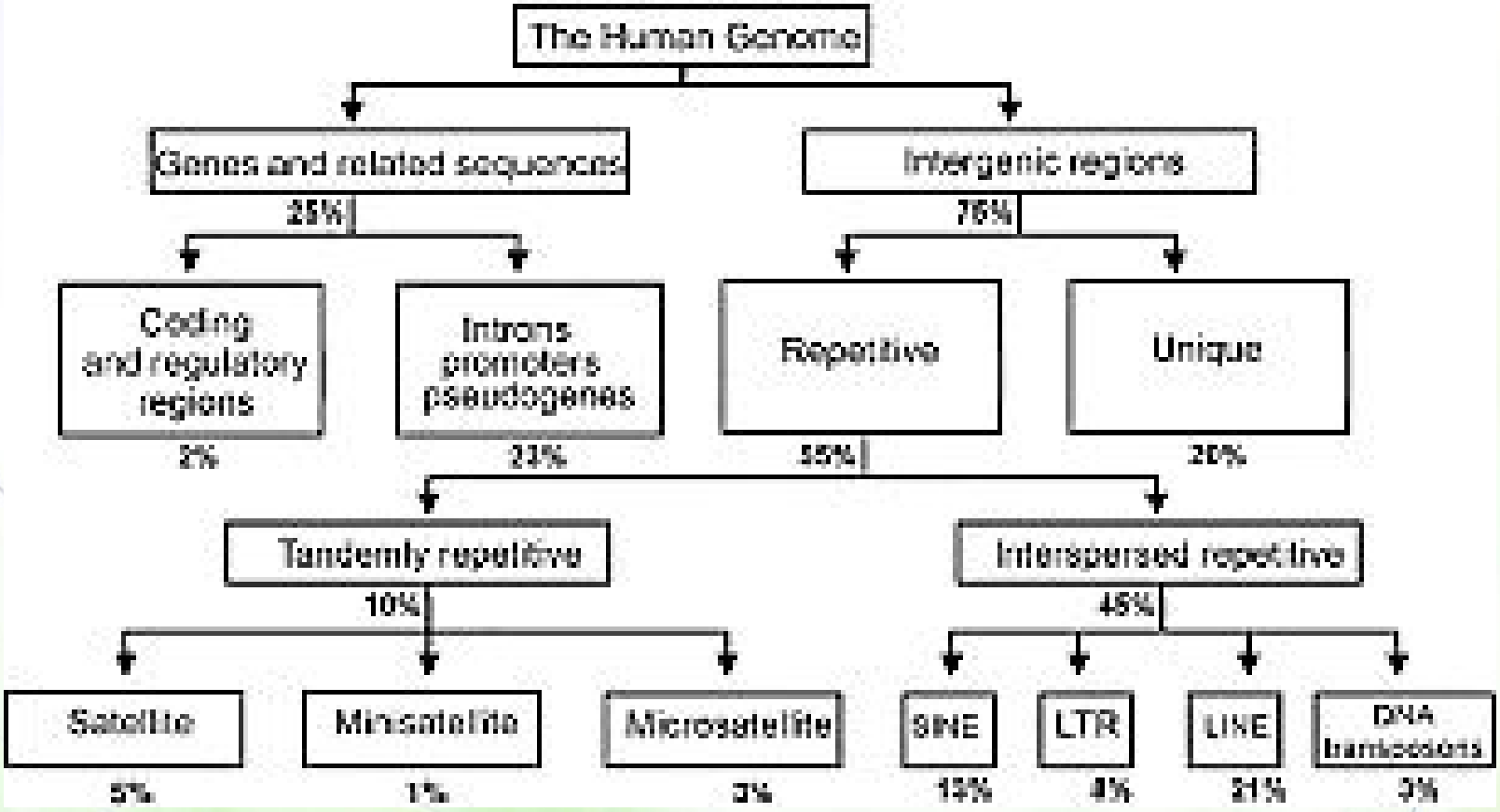


- SYBR green binds to double-stranded DNA and fluoresces only when bound.
- A way of relative quantitation of amount of DNA in a sample is by amplifying it in the presence of SYBR green.
- The higher the amount of DNA, the sooner it is detected.

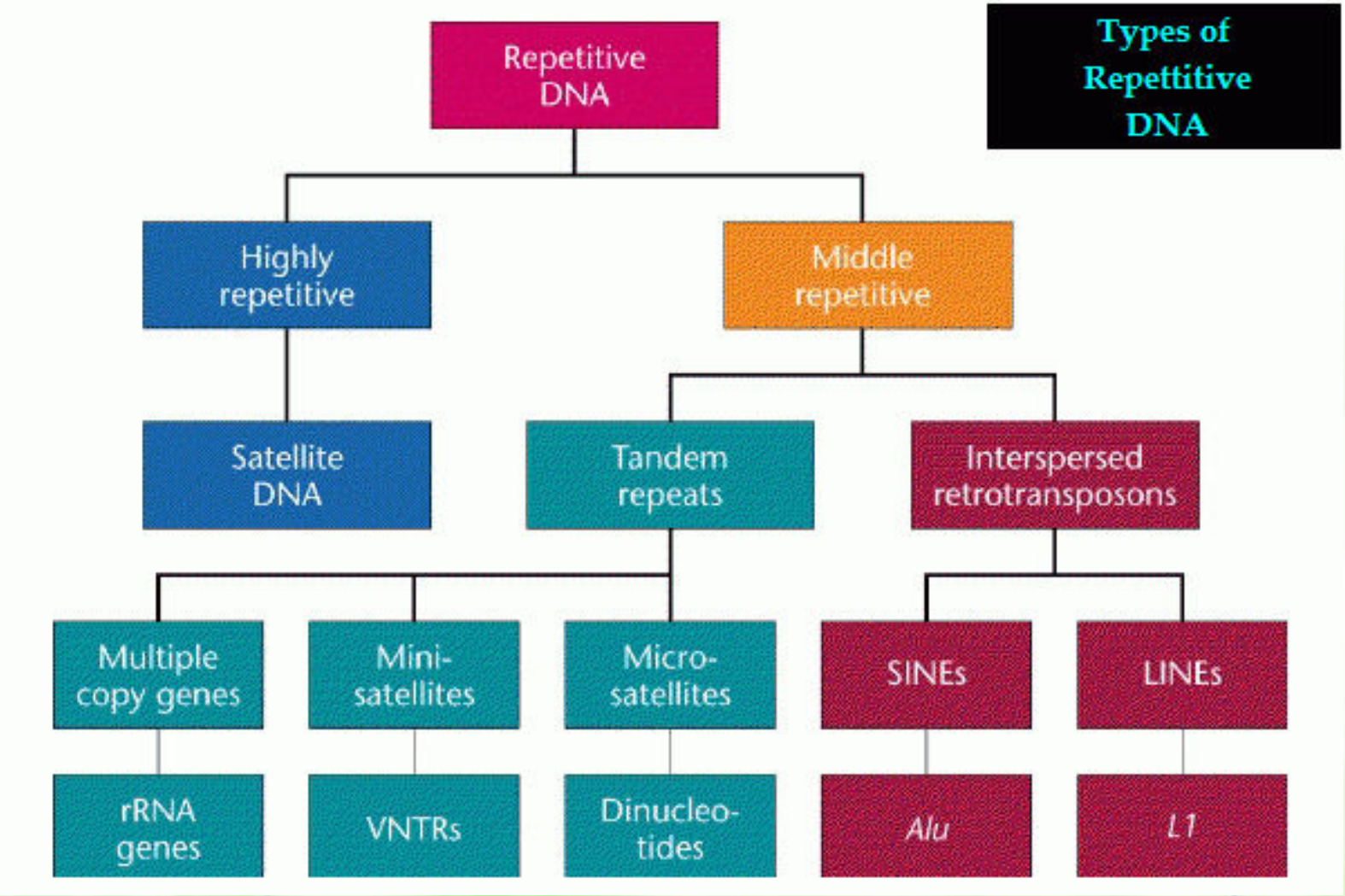


- http://www.youtube.com/watch?v=kvQWKcMdyS4
- http://www.bio.davidson.edu/courses/immunology/flash/rt_pcr.html

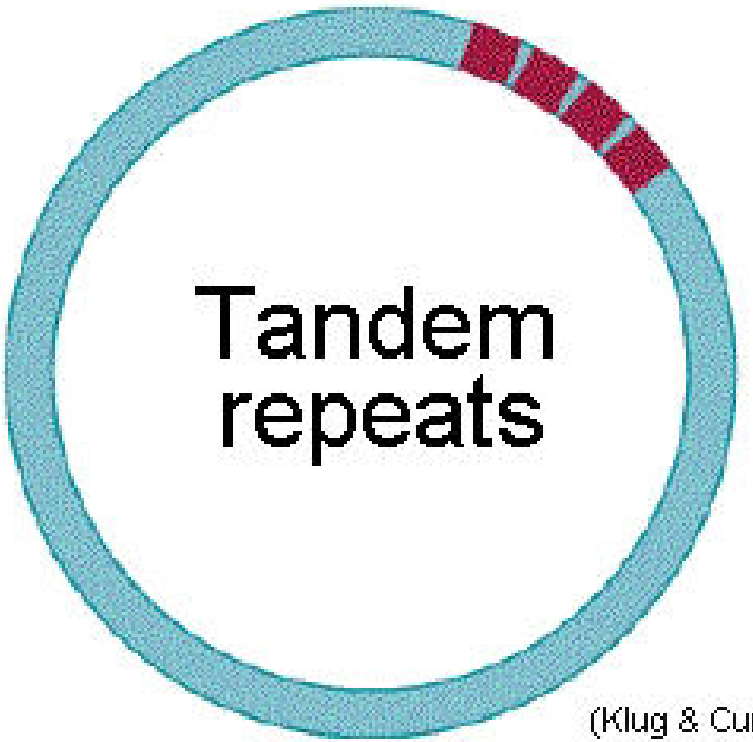
Components of the human genome

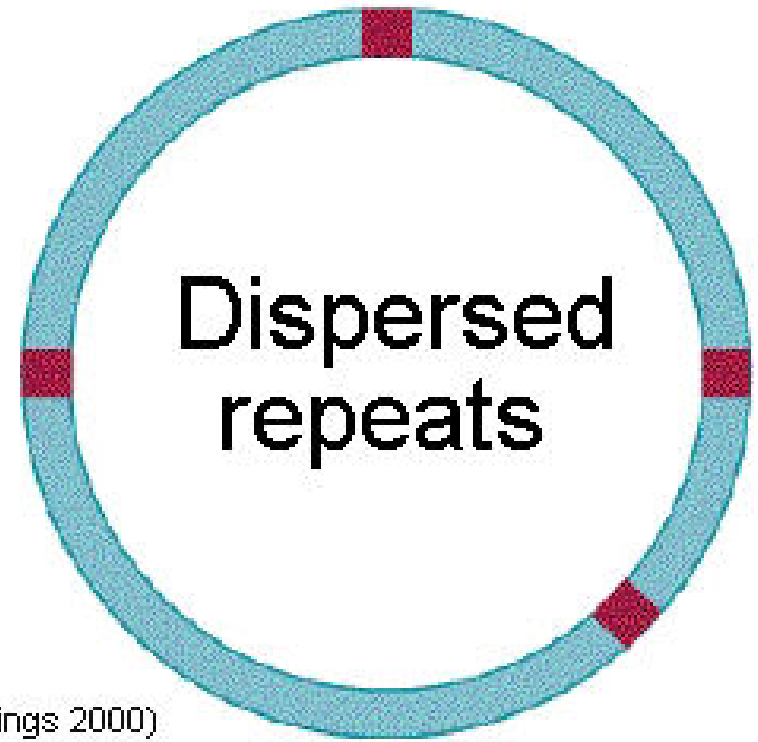


Repetitive DNA sequences



Tandem vs. dispersed





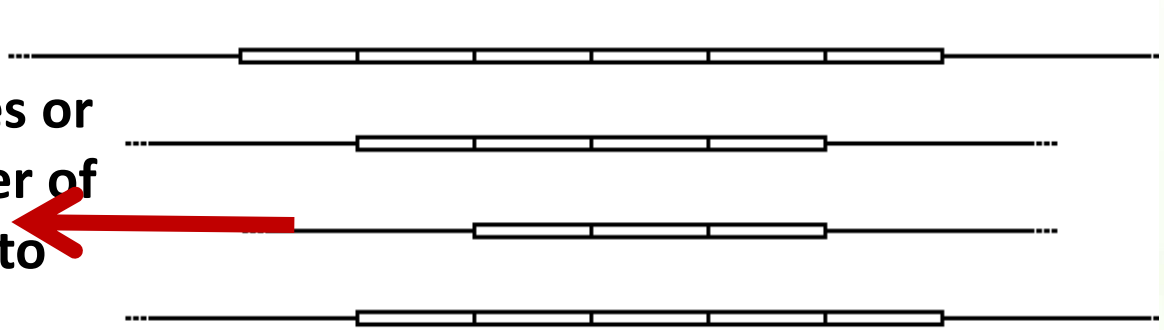
(Klug & Cummings 2000)

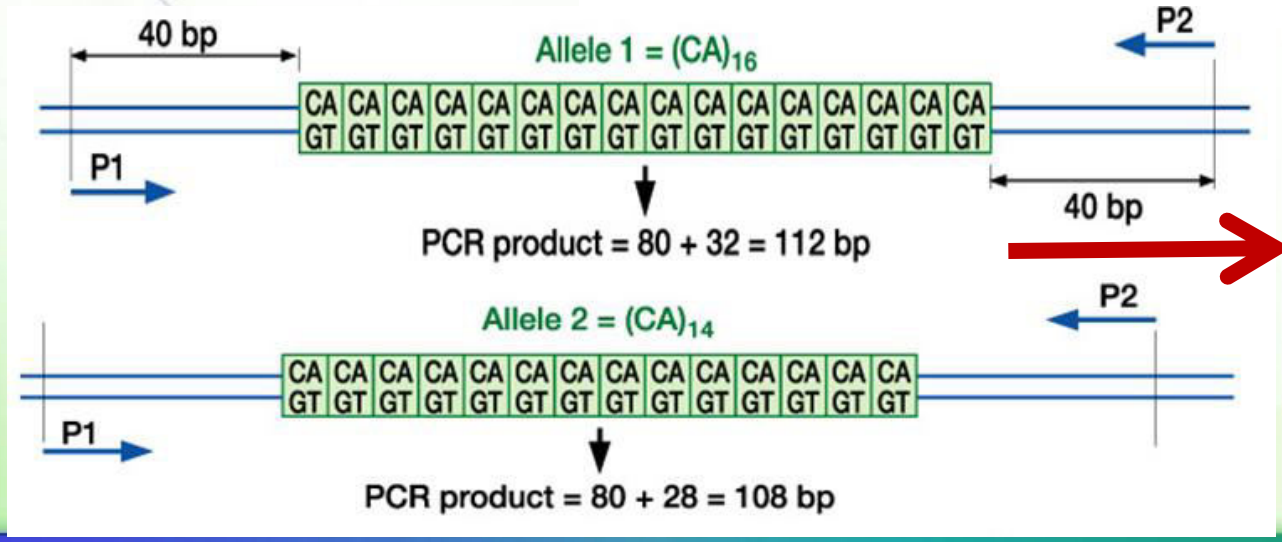
Types of DNA repeats

Satellite (macro-satellite) DNA

- **Centromeric repeats (171 bp) unique to each chromosome (you make chromosome-specific probes)**
- **Telomeric repeats**

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

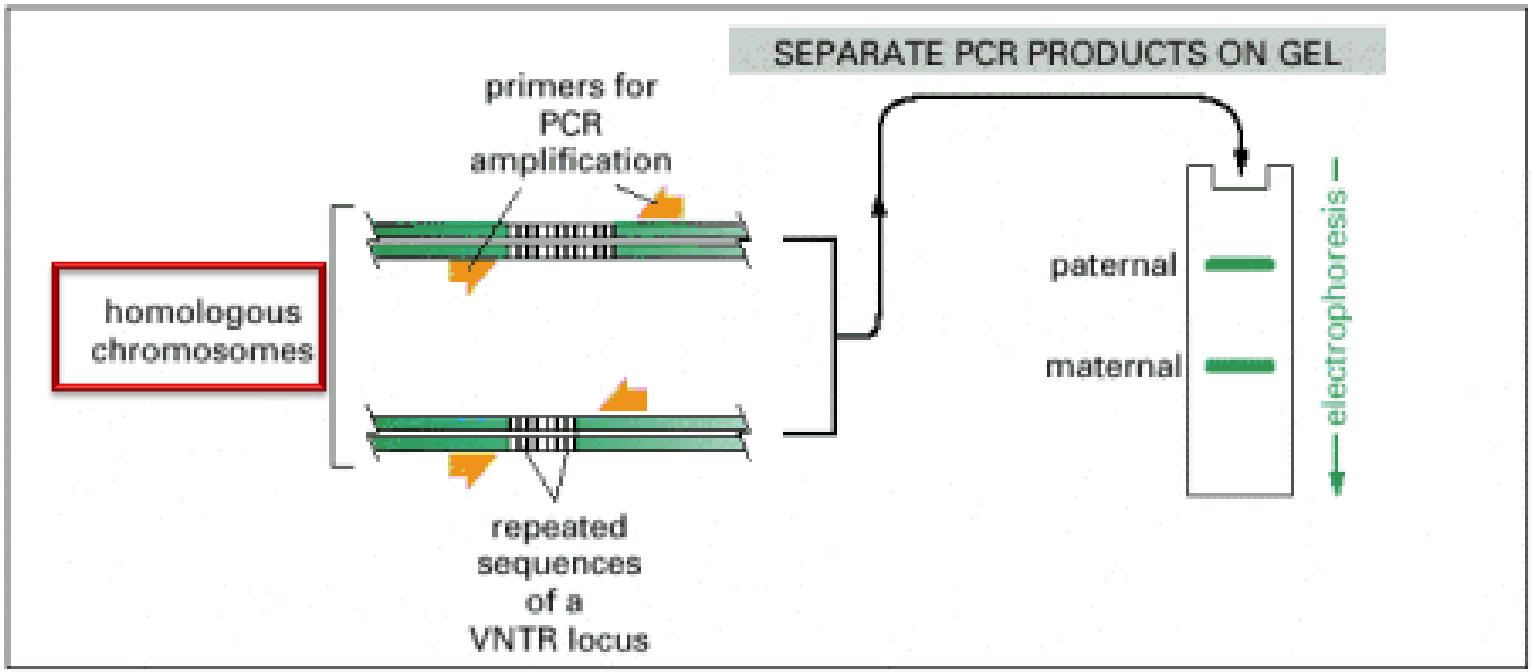




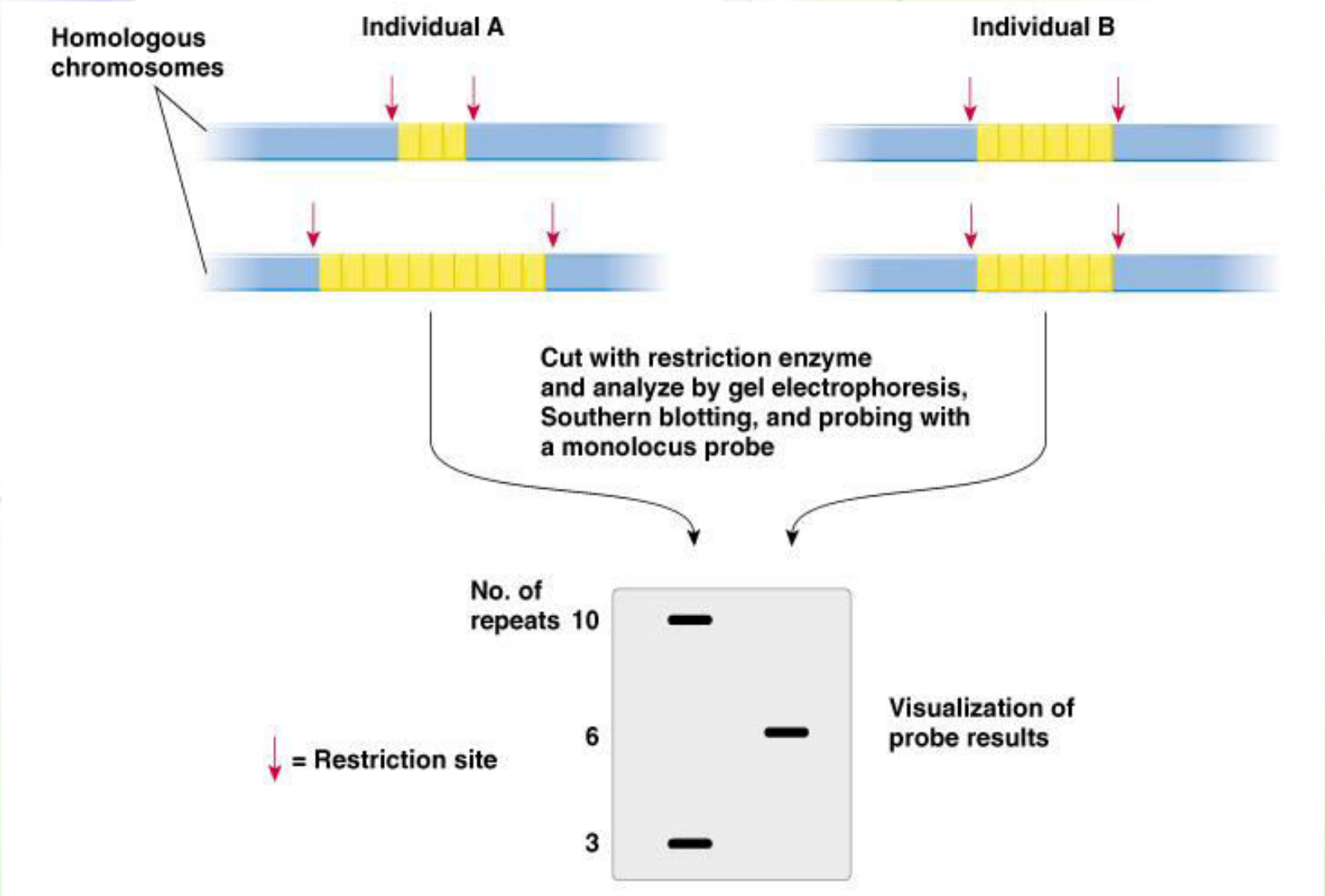
STRs (short tandem repeats) of 2 to 10 bp

Polymorphisms of VNTR and STR

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

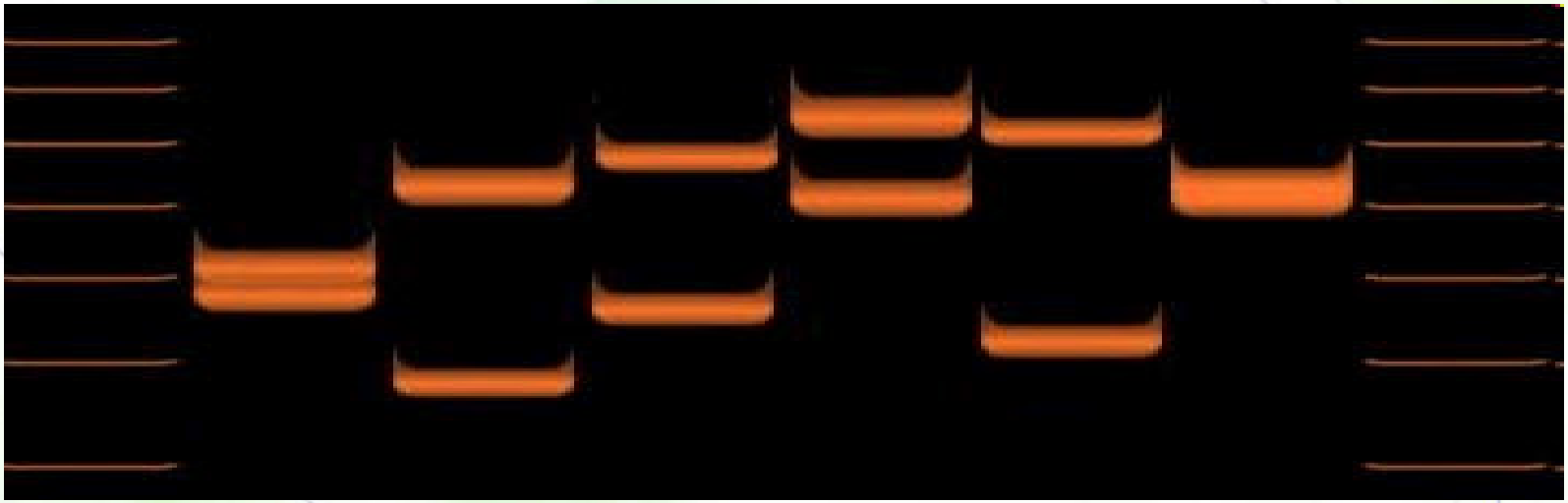


Microsatellites and VNTRs as DNA Markers



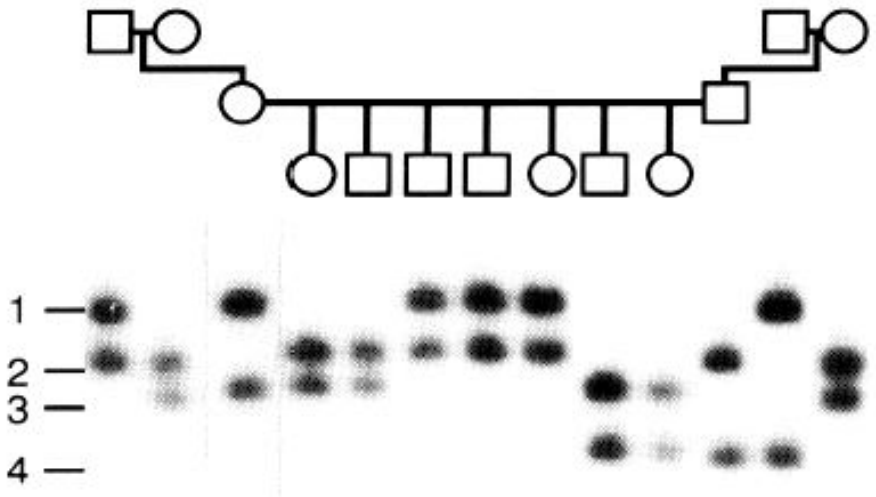
VNTR in medicine and more

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

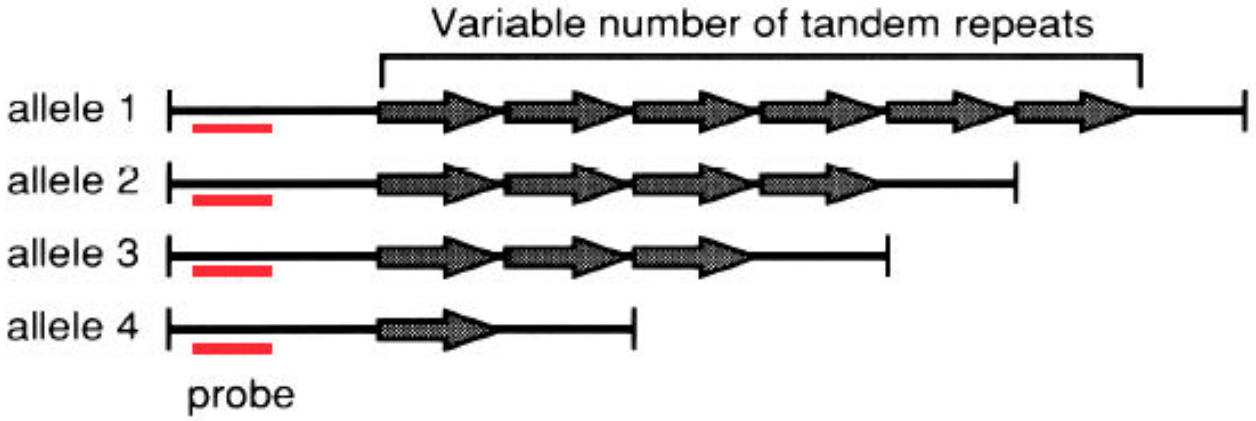


The likelihood of 2 unrelated individuals having same allelic pattern extremely improbable

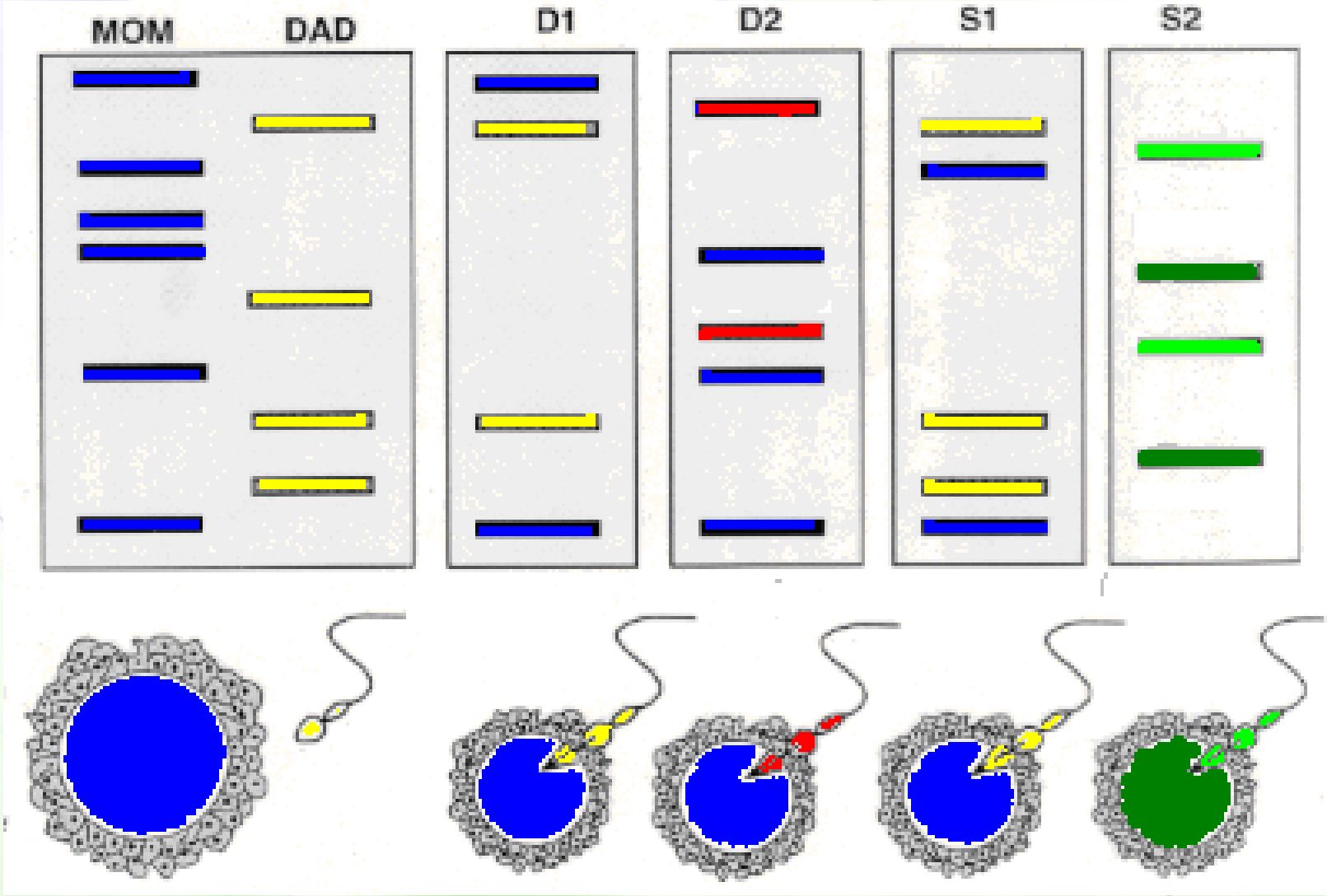
Real example



single-locus probe but multiple alleles



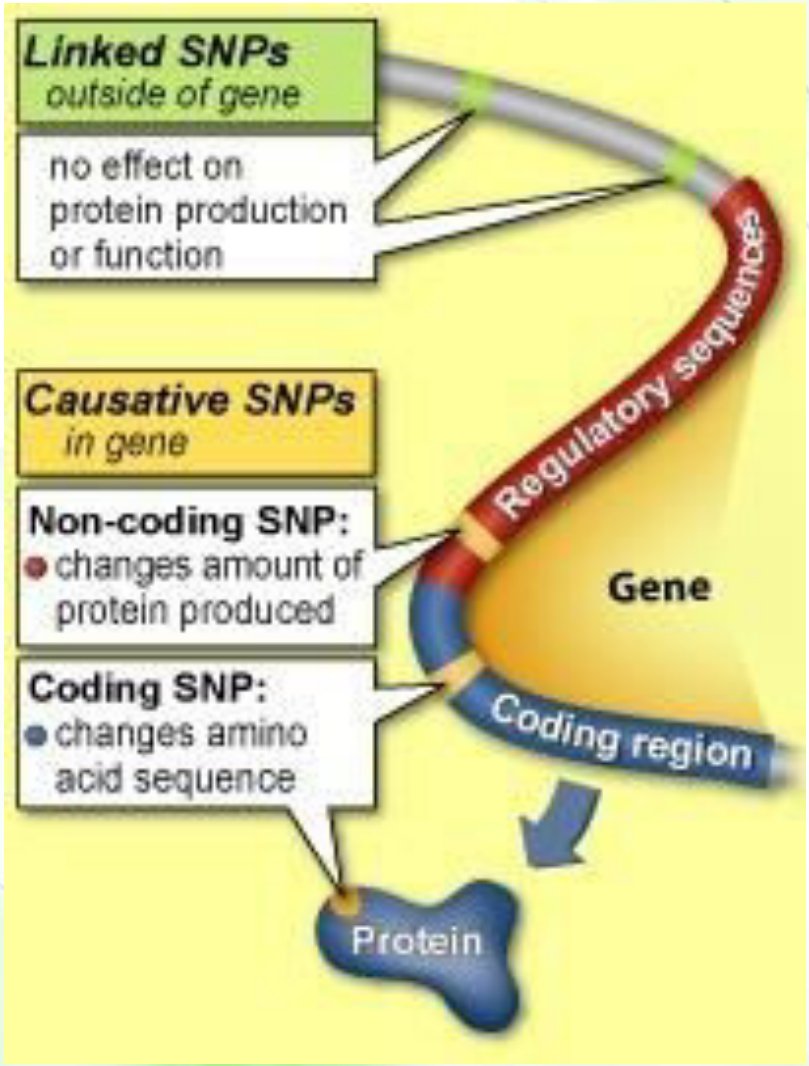
Thompson & Thompson Genetics in Medicine, p. 130, 1991


MOM
DAD
D1
D2
S1
S2

Single nucleotide polymorphism (SNPs)

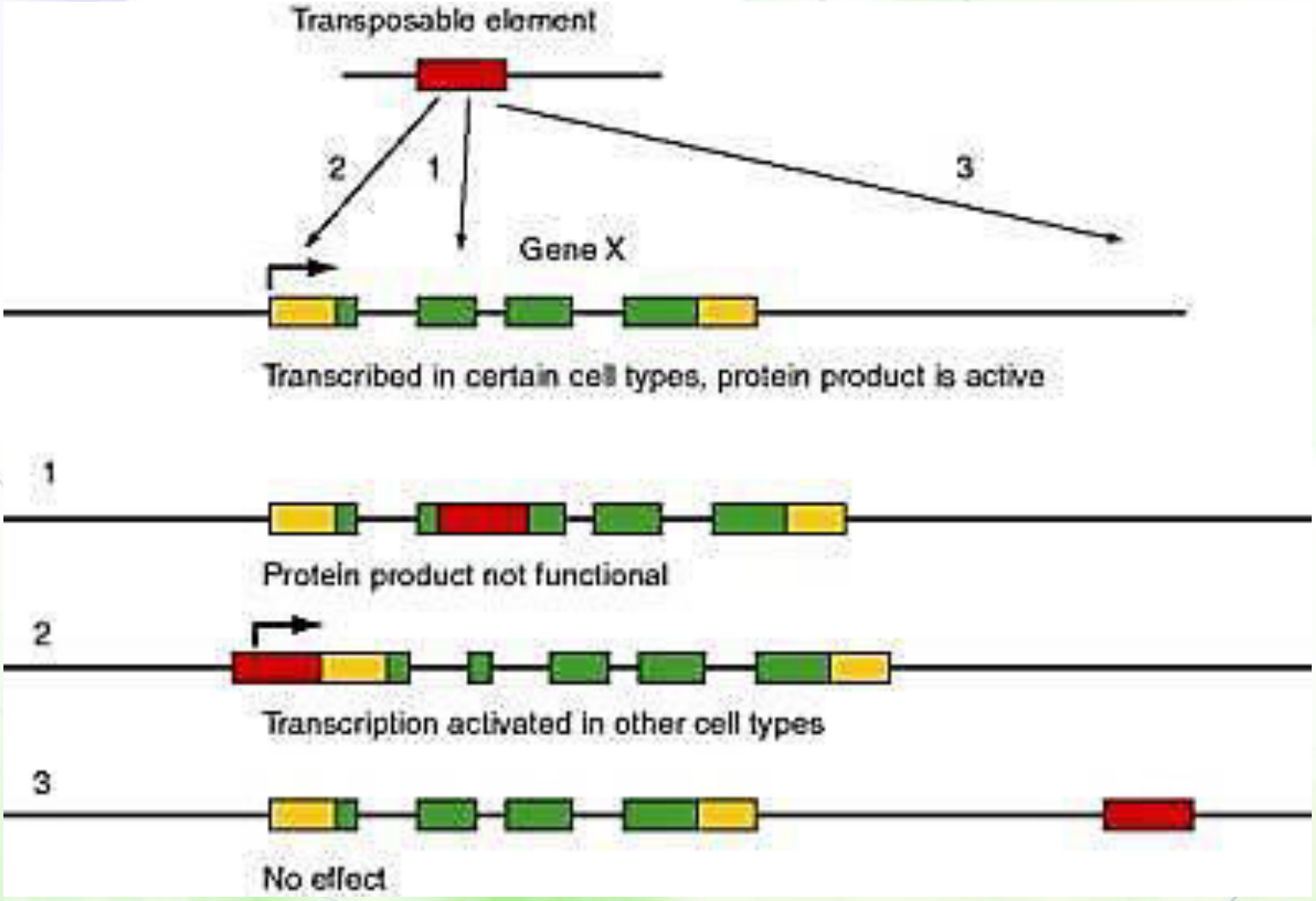
- Another source of genetic variation
- Single-nucleotide substitutions of one base for another
- Two or more versions of a sequence must each be present in at least one percent of the general population
- SNPs occur throughout the human genome - about one in every 300 nucleotide base pairs.
  - ~10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

Categories of SNPs



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