Molecular Biology (3)

The human genome

Mamoun Ahram, PhD
Walhan Al-Shaer, PhD
Second semester, 2019-2020

Resources

- This lecture
- Cooper, Ch.5, pp. 157-160

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

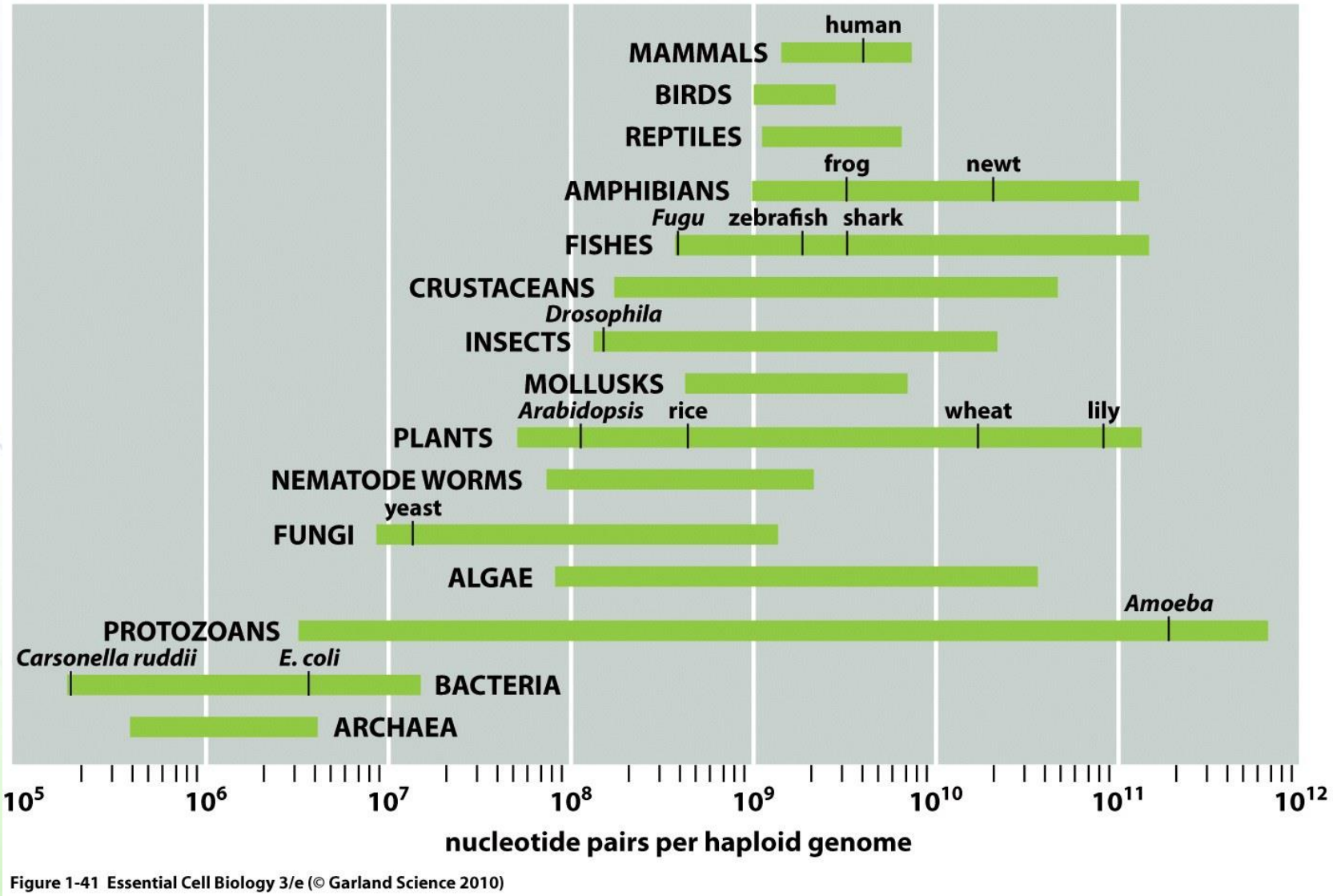
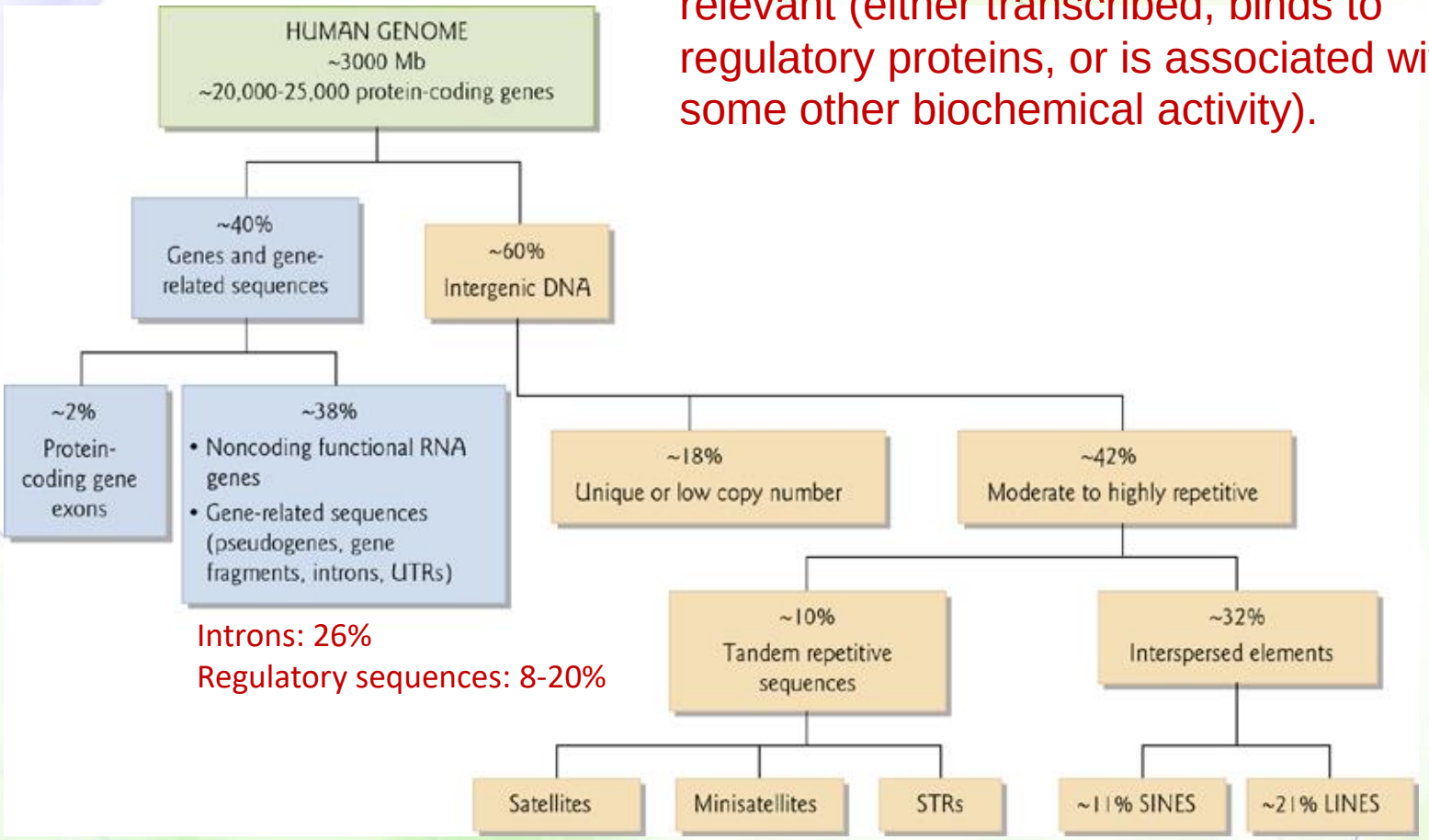


Figure 1-41 Essential Cell Biology 3/e (© Garland Science 2010)

Components of the human genome

80% of the entire human genome is relevant (either transcribed, binds to regulatory proteins, or is associated with some other biochemical activity).

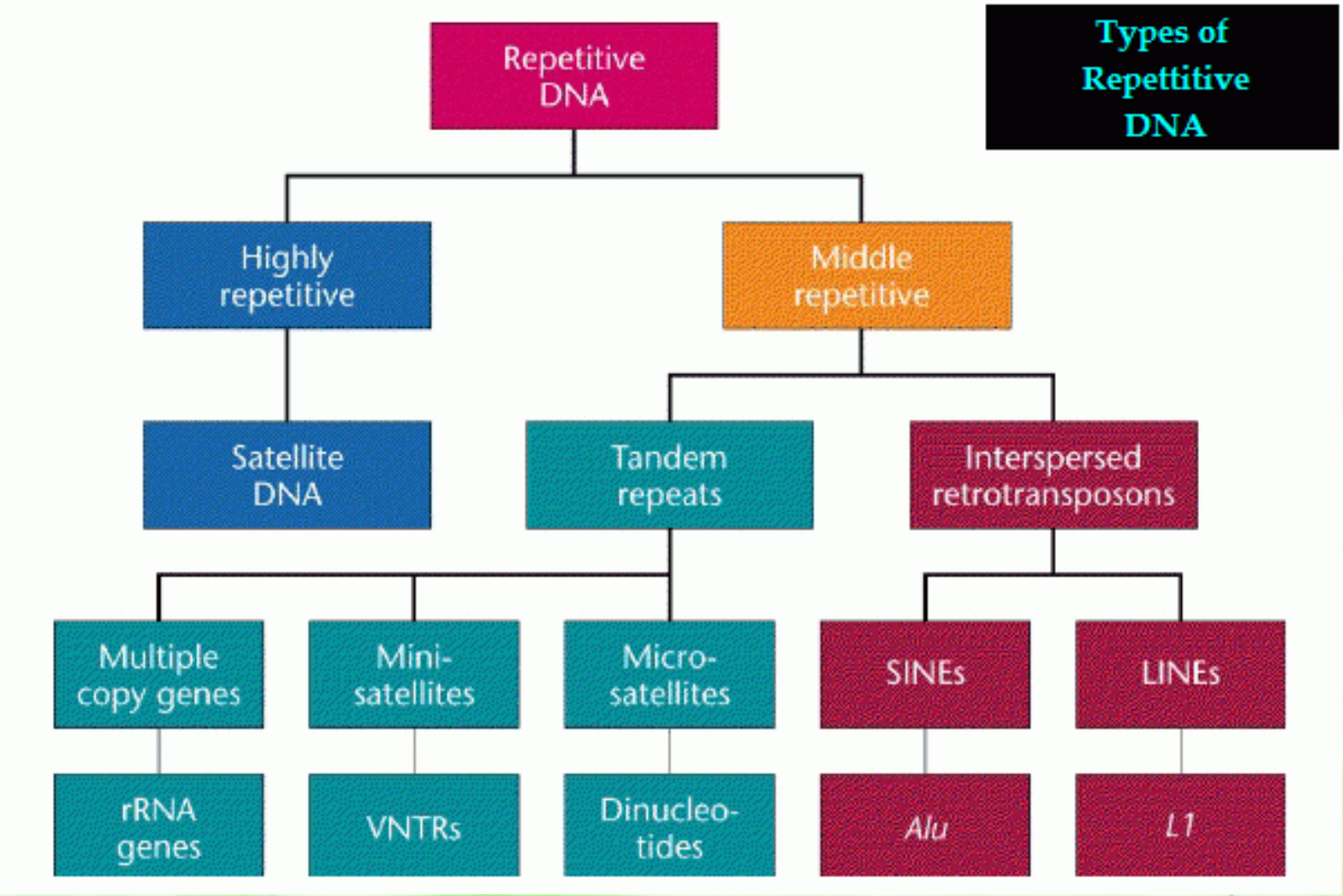


Introns: 26%

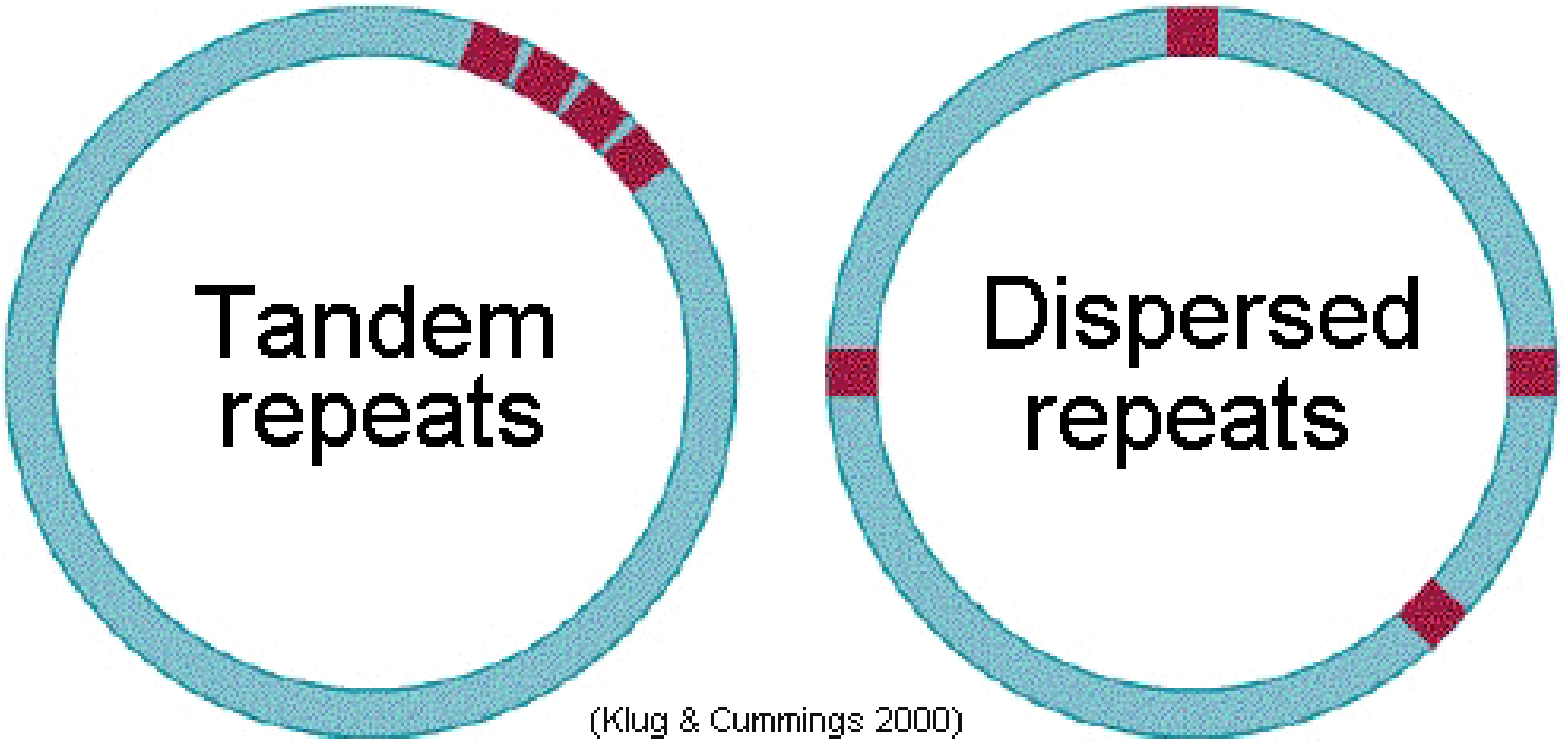
Regulatory sequences: 8-20%

~5% of the genome contains sequences of noncoding DNA that are highly conserved indicating that they are critical to survice.

Repetitive DNA sequences



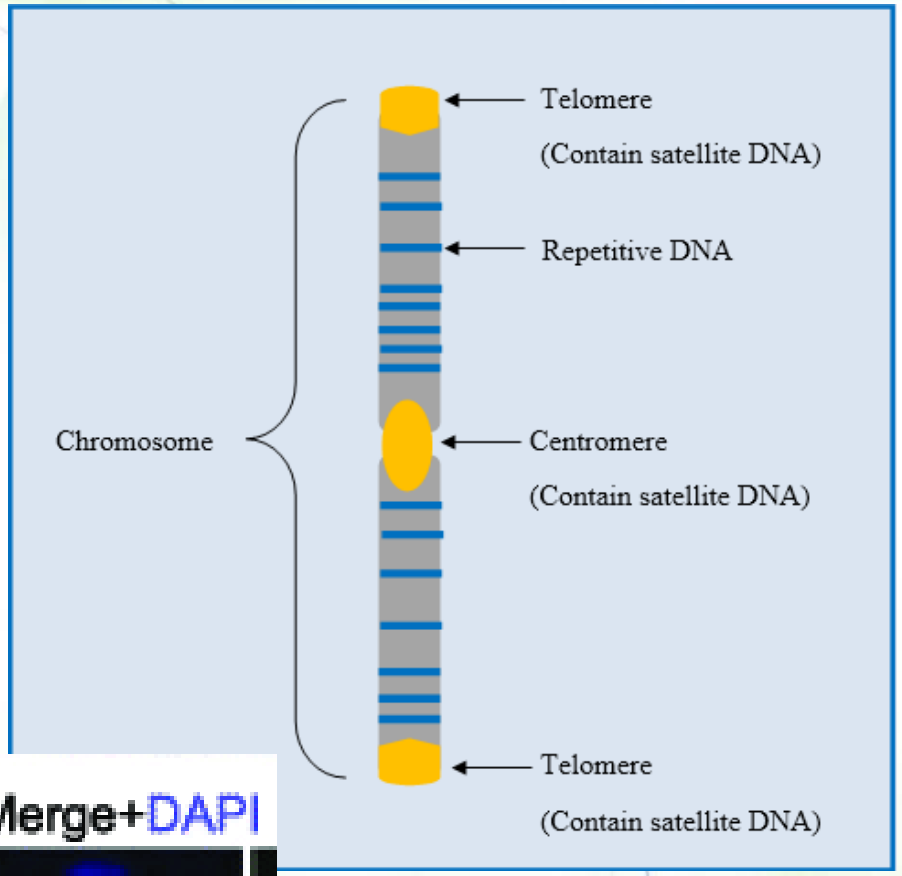
Tandem vs. dispersed

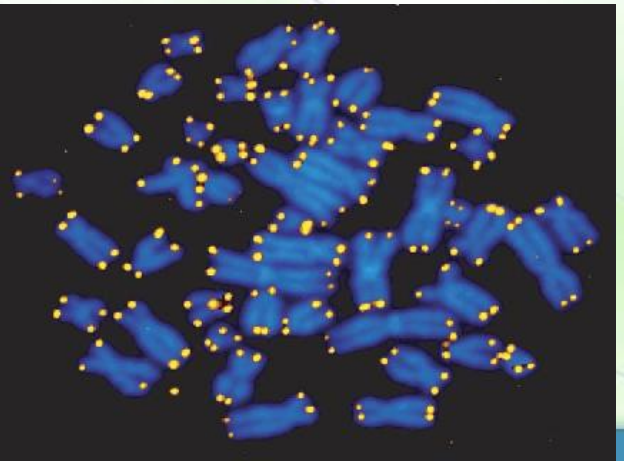


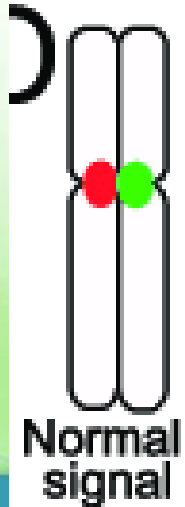
(Klug & Cummings 2000)

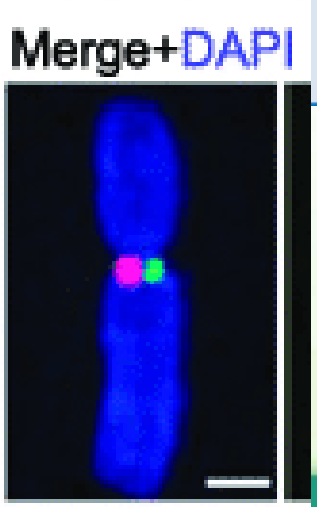
Satellite (macro-satellite) DNA

- 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

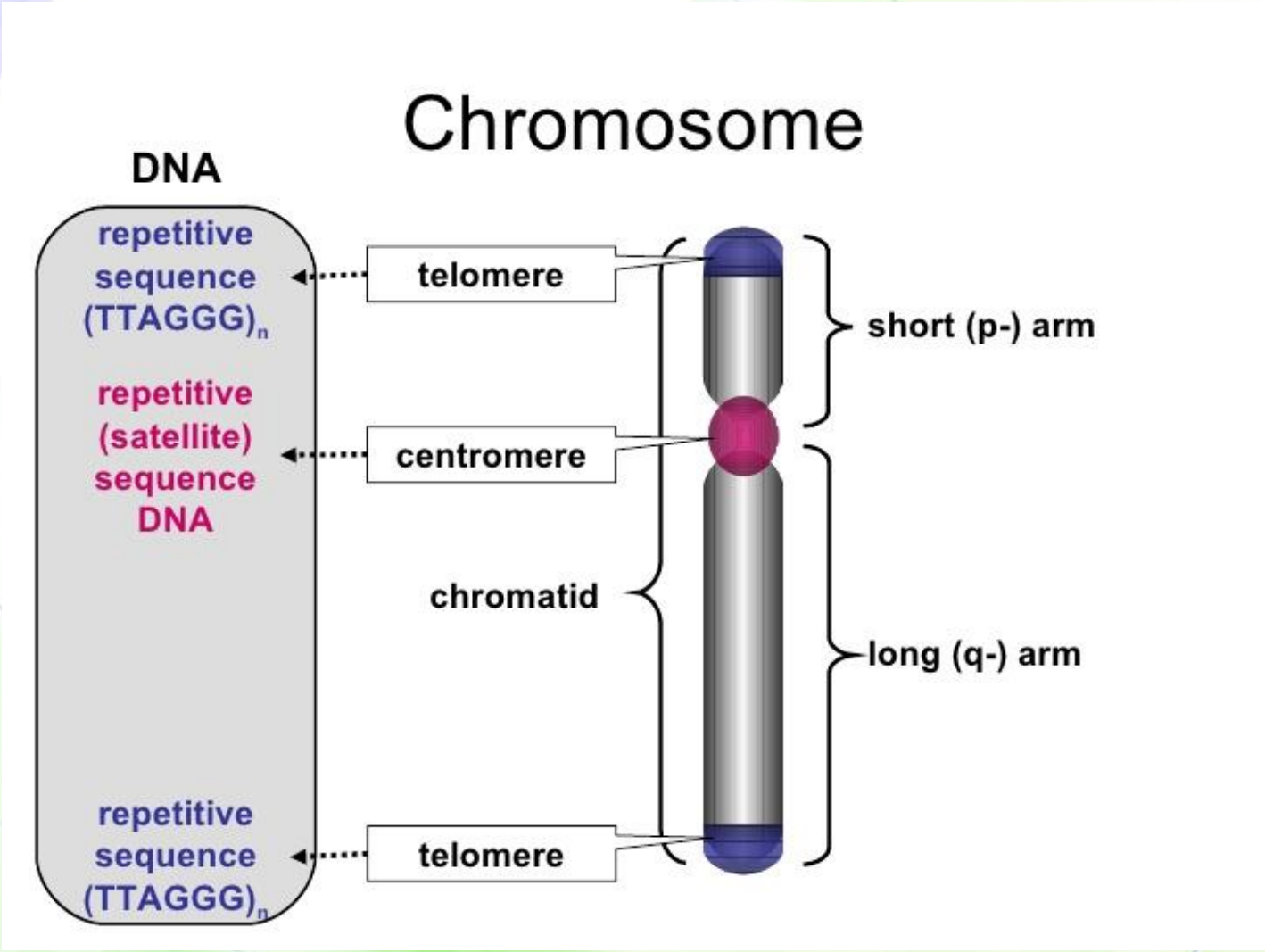






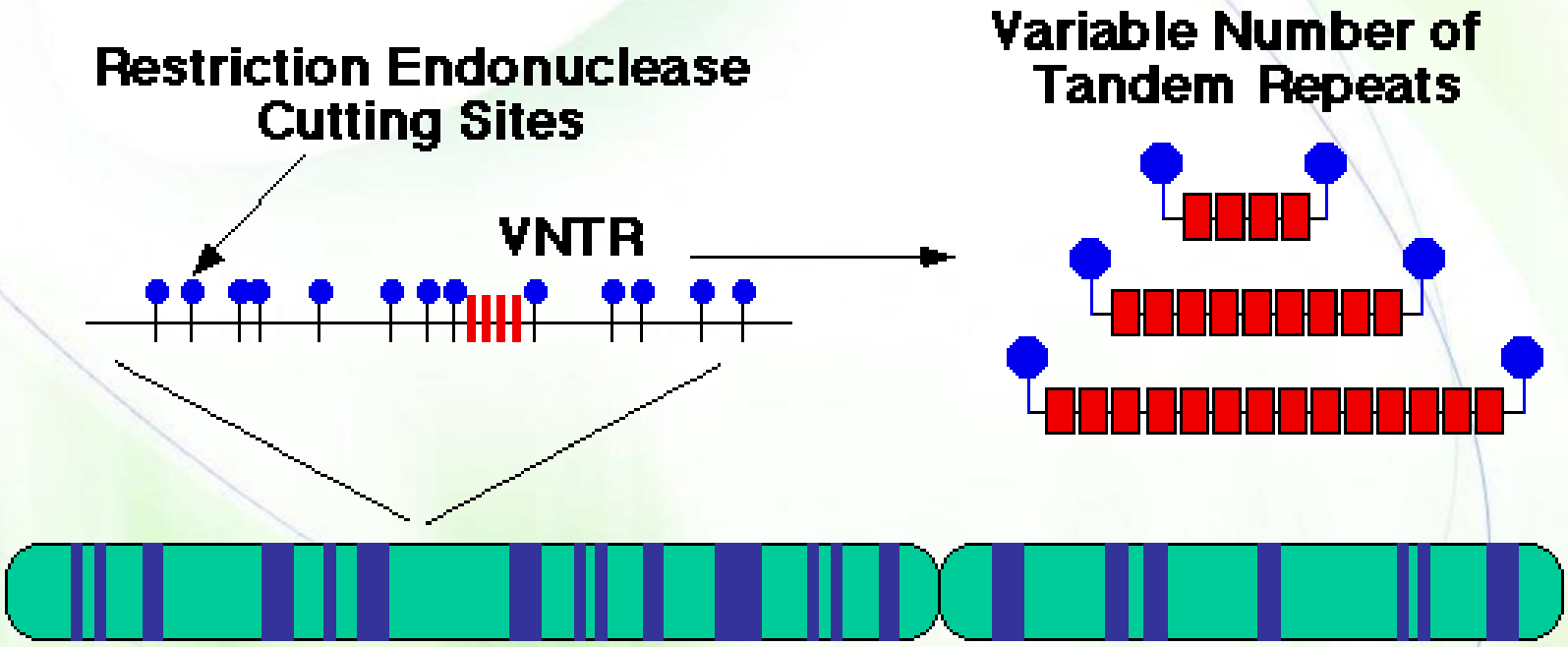


Centromeric and telomeric repeats



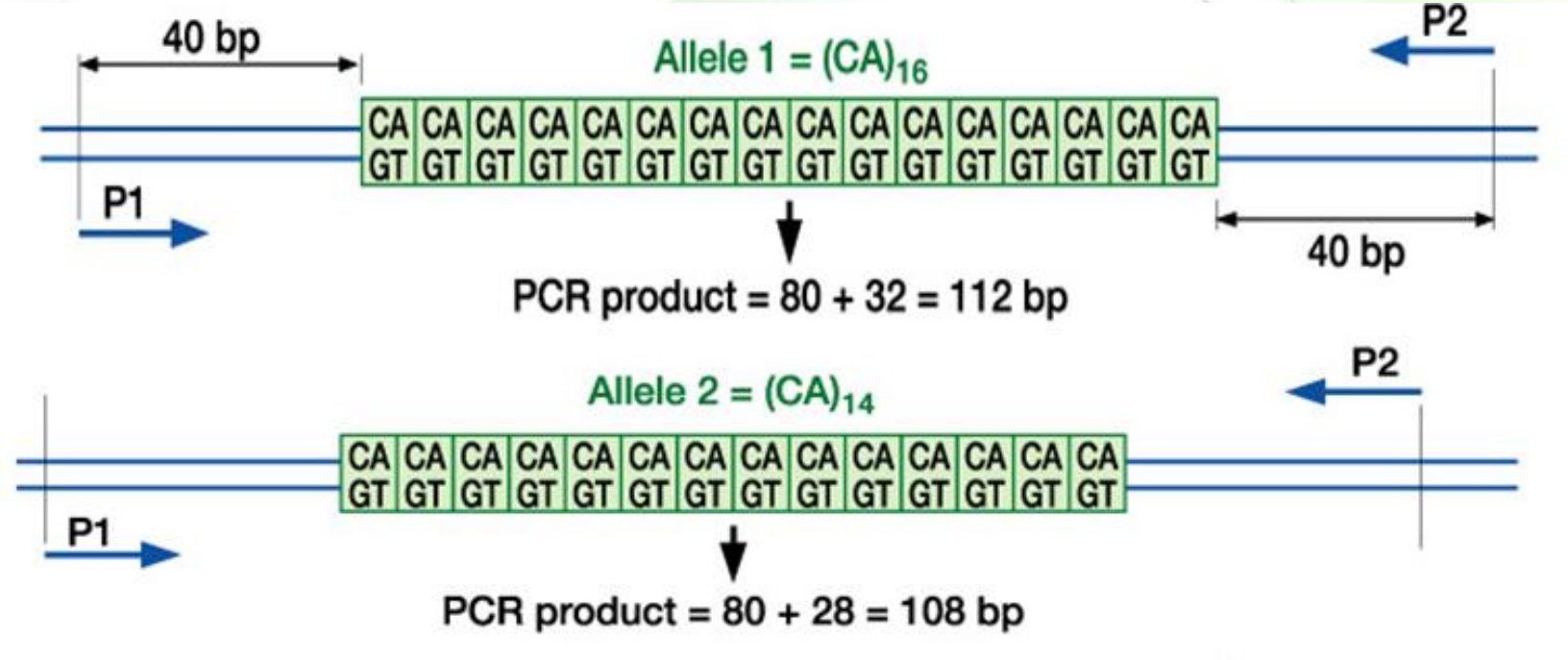
Mini-satellite DNA

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



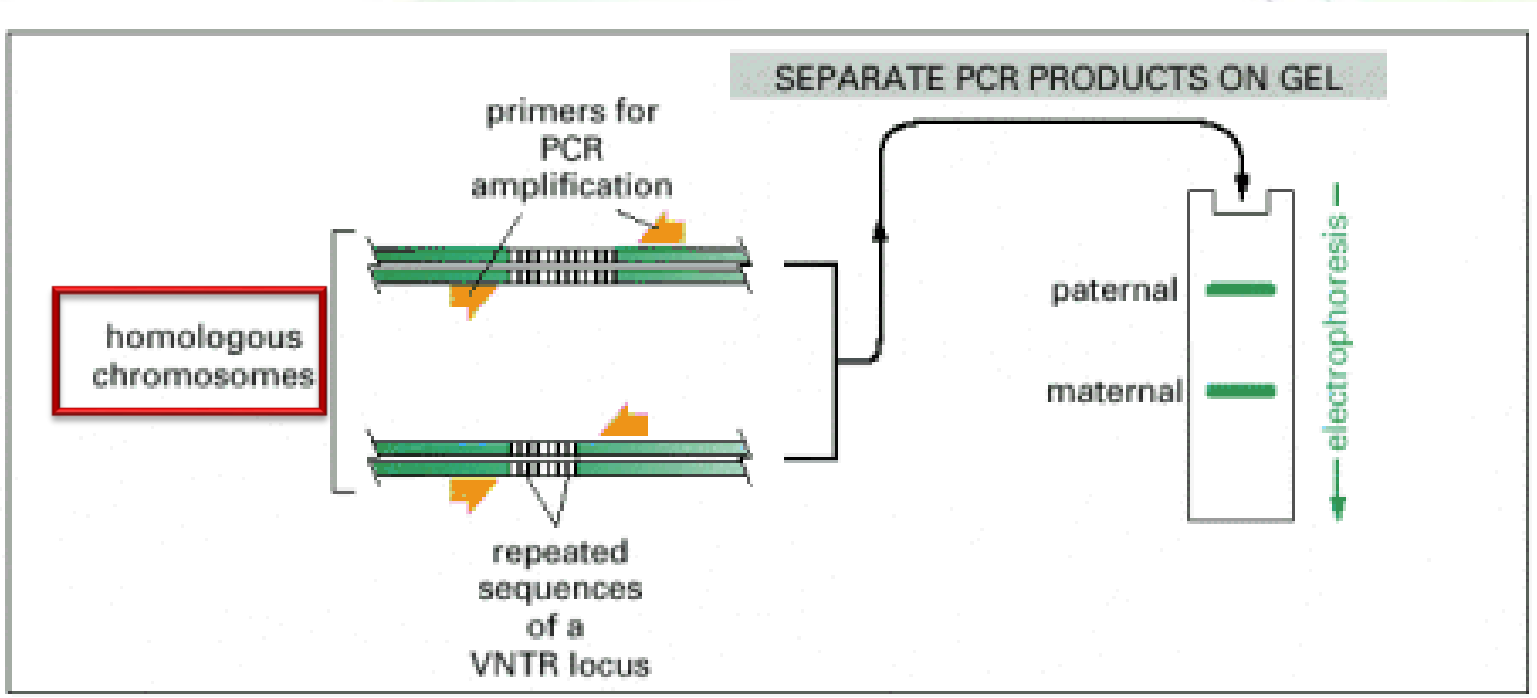
Micro-satellite DNA

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

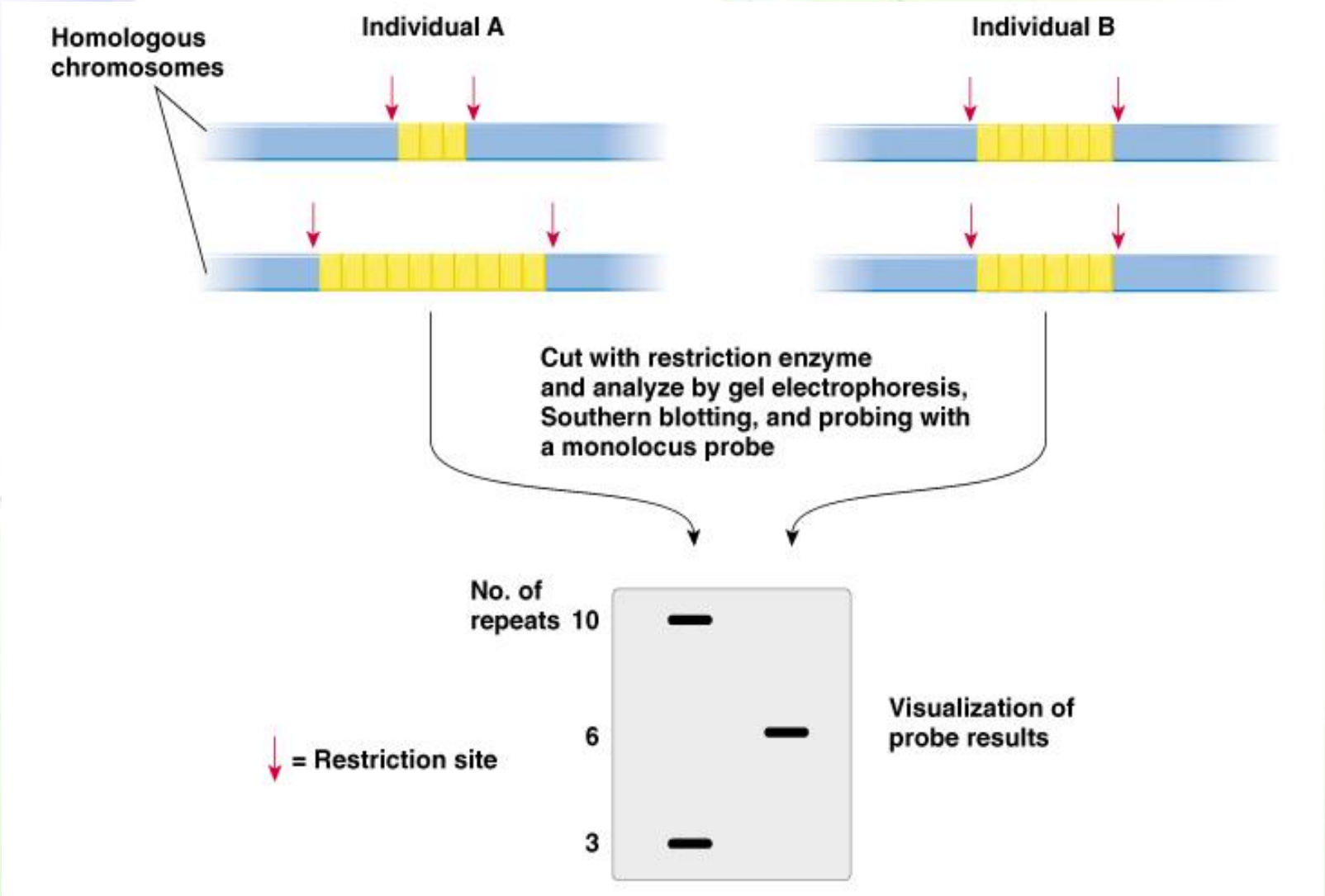


Polymorphisms of VNTR and STR

- STRs and VNTRs are highly variable among individuals (polymorphic).
  - They are useful in DNA profiling for forensic testing.

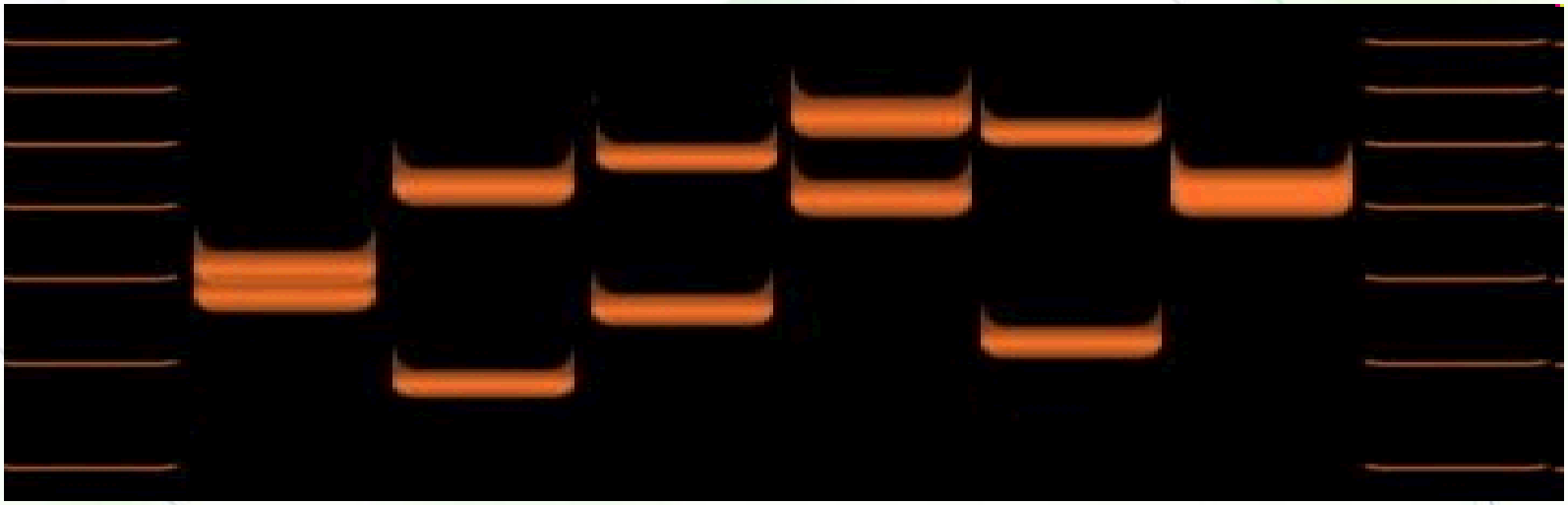


Microsatellites and VNTRs as DNA Markers



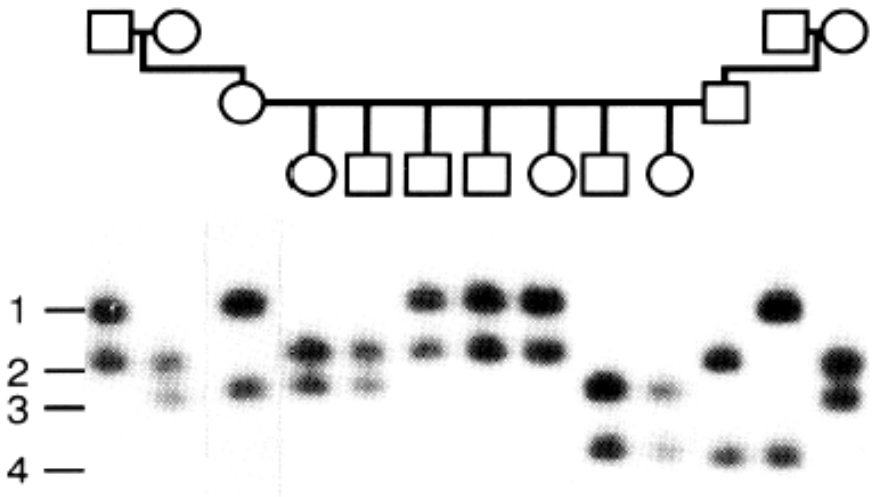
VNTR in medicine and more

- VNTR allelic length variation among 6 individuals.

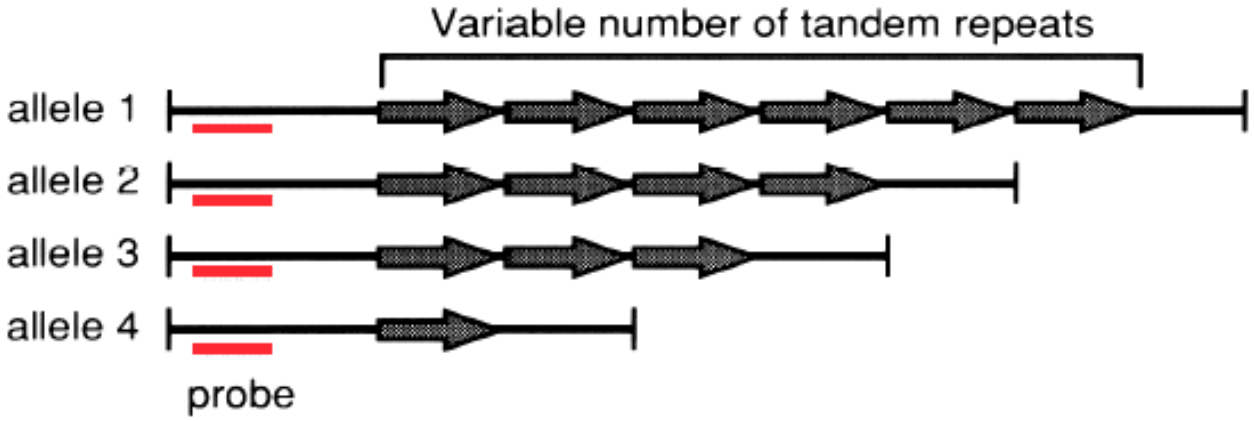


The likelihood of 2 unrelated individuals having same allelic pattern is extremely improbable.

Real example

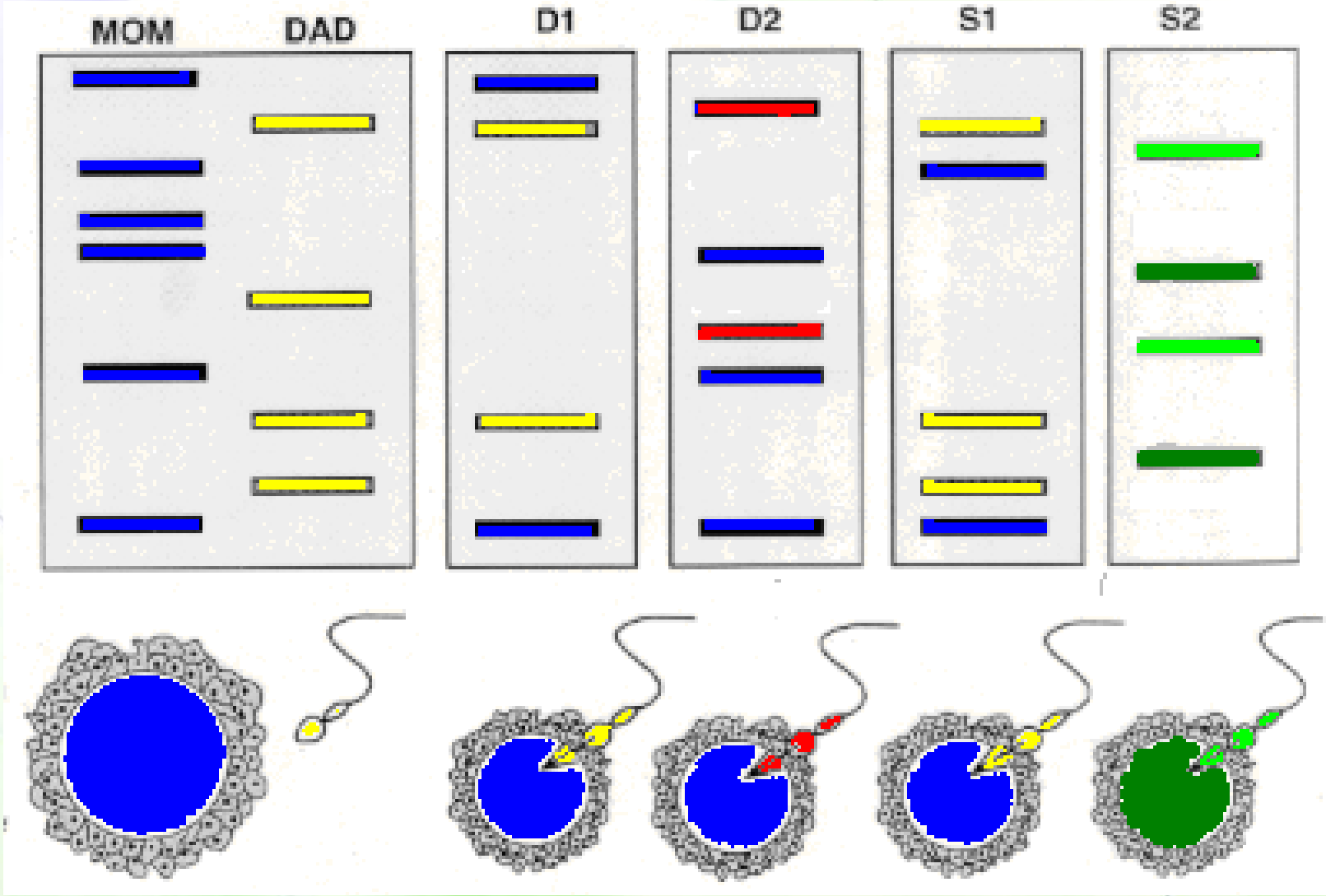


single-locus probe but multiple alleles



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

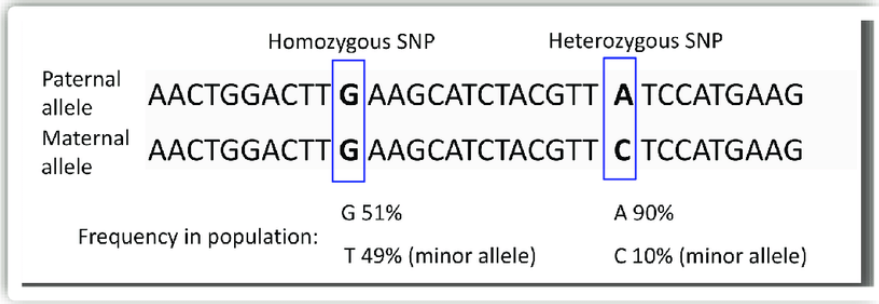
Paternity testing



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

Examples



Individual 1

Chr 2 ...CGATATTCCTATCGAATGTC...
copy1 ...GCTATAAGGATAGCTTACAG...

Chr 2 ...CGATATTCCCATCGAATGTC...
copy2 ...GCTATAAGGGTAGCTTACAG...

Individual 2

Chr 2 ...CGATATTCCCATCGAATGTC...
copy1 ...GCTATAAGGGTAGCTTACAG...

Chr 2 ...CGATATTCCCATCGAATGTC...
copy2 ...GCTATAAGGGTAGCTTACAG...

Individual 3

Chr 2 ...CGATATTCCTATCGAATGTC...
copy1 ...GCTATAAGGATAGCTTACAG...

Chr 2 ...CGATATTCCTATCGAATGTC...
copy2 ...GCTATAAGGATAGCTTACAG...

Individual 4

Chr 2 ...CGATATTCCTATCGAATGTC...
copy1 ...GCTATAAGGATAGCTTACAG...

Chr 2 ...CGATATTCCCATCGAATGTC...
copy2 ...GCTATAAGGGTAGCTTACAG...

Individual 5

Chr 2 ...CGATATTCCCATCGAATGTC...
copy1 ...GCTATAAGGGTAGCTTACAG...

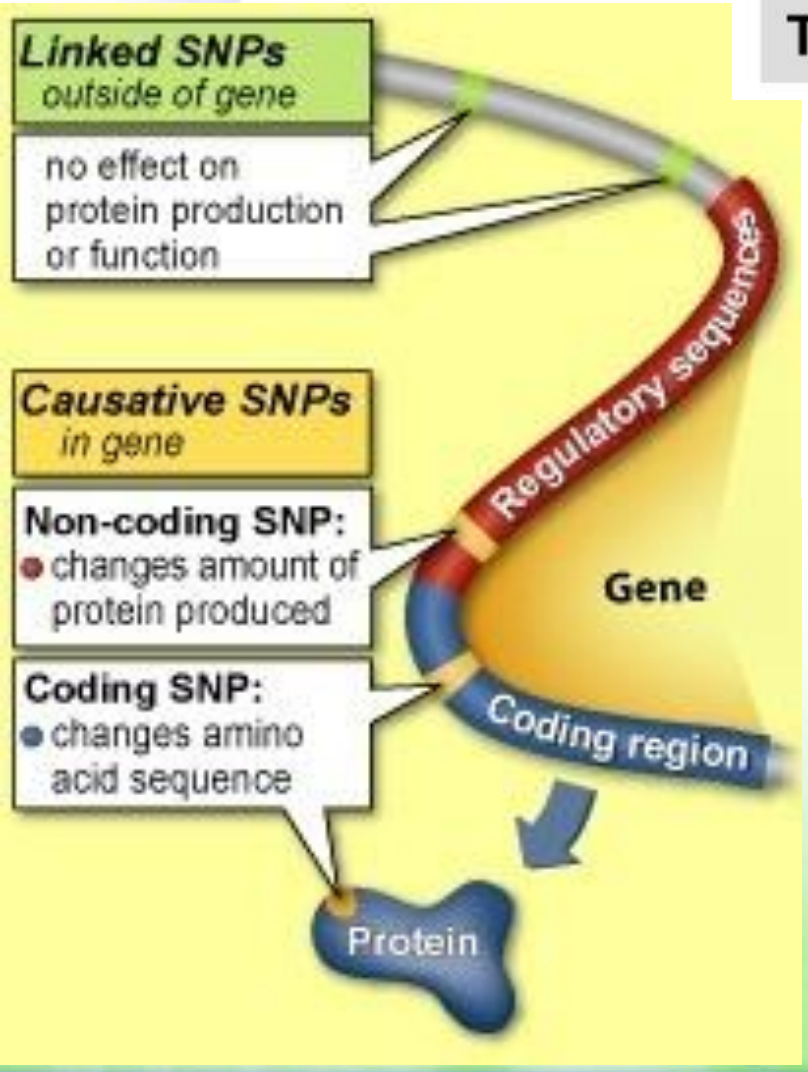
Chr 2 ...CGATATTCCTATCGAATGTC...
copy2 ...GCTATAAGGATAGCTTACAG...

Individual 6

Chr 2 ...CGATATTCCCATCGAATGTC...
copy1 ...GCTATAAGGGTAGCTTACAG...

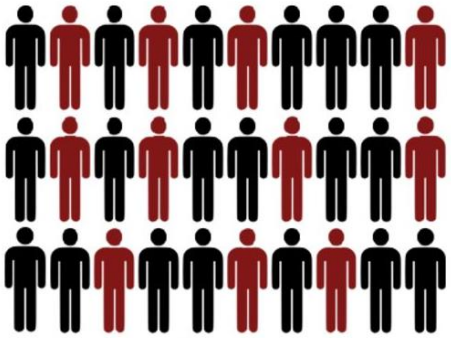
Chr 2 ...CGATATTCCTATCGAATGTC...
copy2 ...GCTATAAGGATAGCTTACAG...

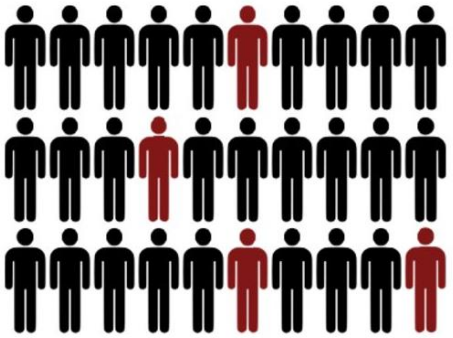
Categories of SNPs



TTGGCCAGCTGGA**C**GAGGGCGATGAC

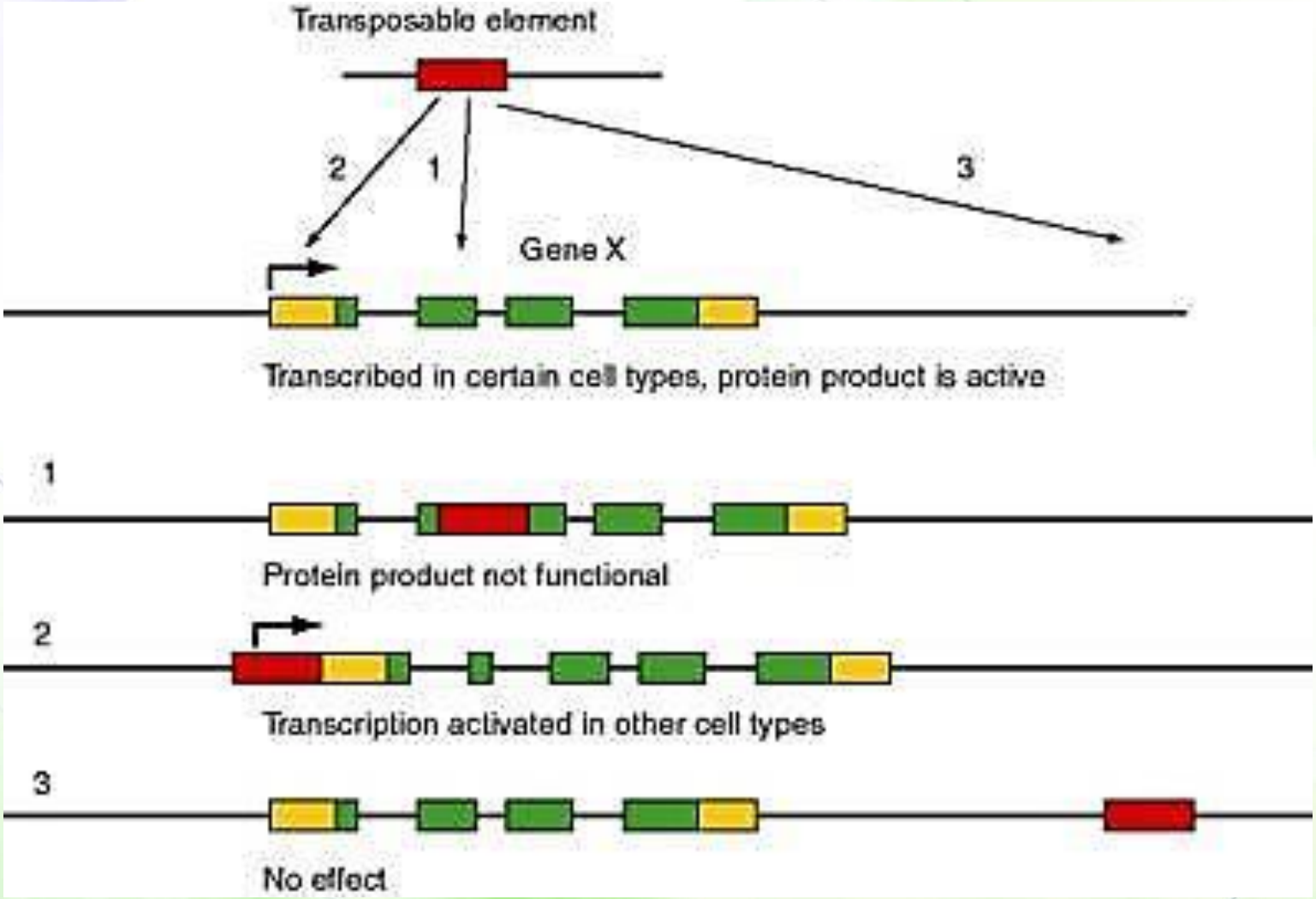
TTGGCCAGCTGGA**T**GAGGGCGATGAC





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)
- Over 99% of the transposons in the human genome lost their ability to move, but we still have some active transposable elements that can sometimes cause disease.
  - 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