Ch. 1

Genomes – Epigenomes – Nuclear organization

L1.1

Genomes

the Human Genome

DNA sequence

Human Genome Project (1991-2003)

Next Generation Sequencing

Databases

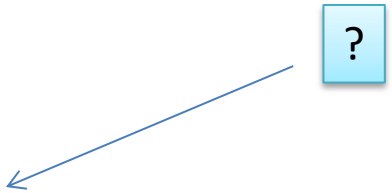
Composition

Human Variation

Moodle website, last section
«Auxiliaries»
you will find videos and text to explain accessing to DNA sequence:

The Human Genome Project
Next Generation Sequencing technologies
Microarrays

HGP (see book, moodle site) 1990-2003



1991-1998 - Physical mapping period (**EST, SST, known genes**)

1998-2003 - Cloning, sequencing (Sanger) and assembly

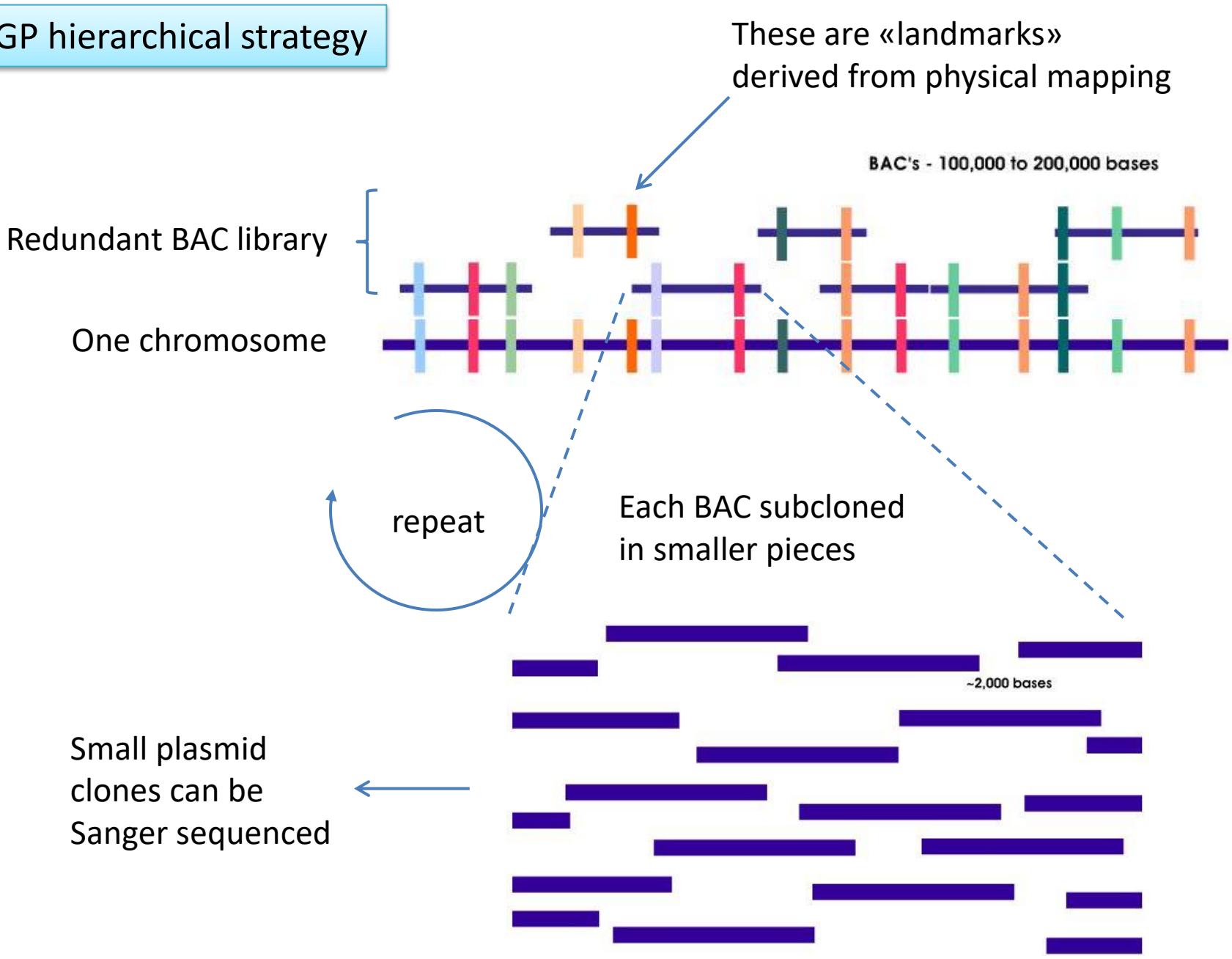
The principle: «hierarchical cloning»

Stocastic: the process required super-extensive and highly redundant cloning

Hierarchical cloning vectors:

- BACs, PACs – 100-200 Kb
- Cosmids and other phage-derived vectors (20-40Kb)
- Plasmids – 2-3 Kb

HGP hierarchical strategy



Sanger di-deoxy-nucleotide terminator method

- Requires isolated DNA fragments (cloned)
- Requires known primer sequence
- Intrinsically limited to 5-700 bp

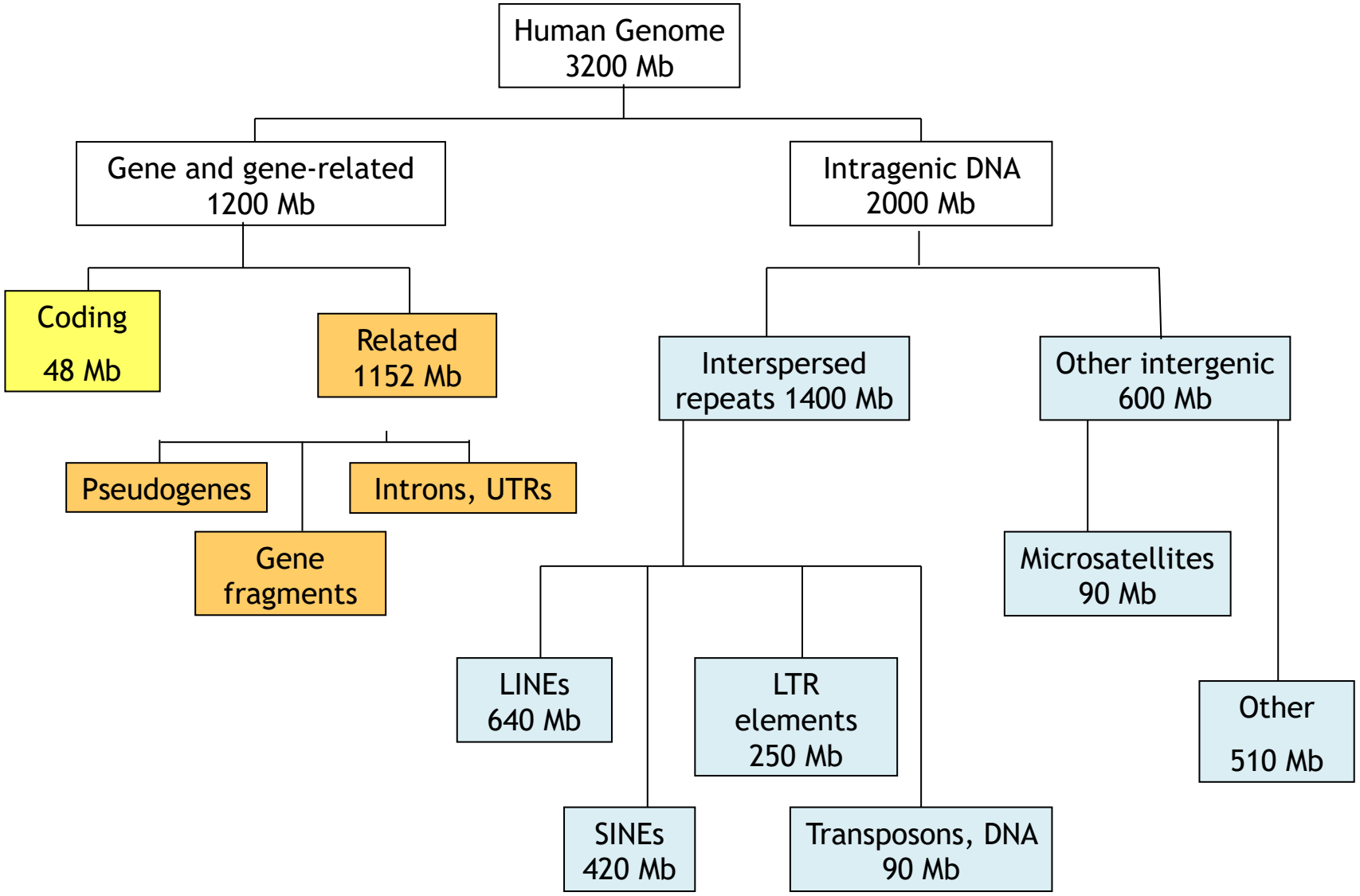
Sanger sequencing

Composition of the Human Genome

Sequence identity was progressively **annotated** in the Human Genome by extensive bioinformatic analysis

- Sequence similarity (homology)
- Correspondence to known RNA / proteins
- Repeated sequence comparison with known genetic elements
- Knowledge on genomes of different organisms

Genome composition - H. Sapiens (the 2003 version).

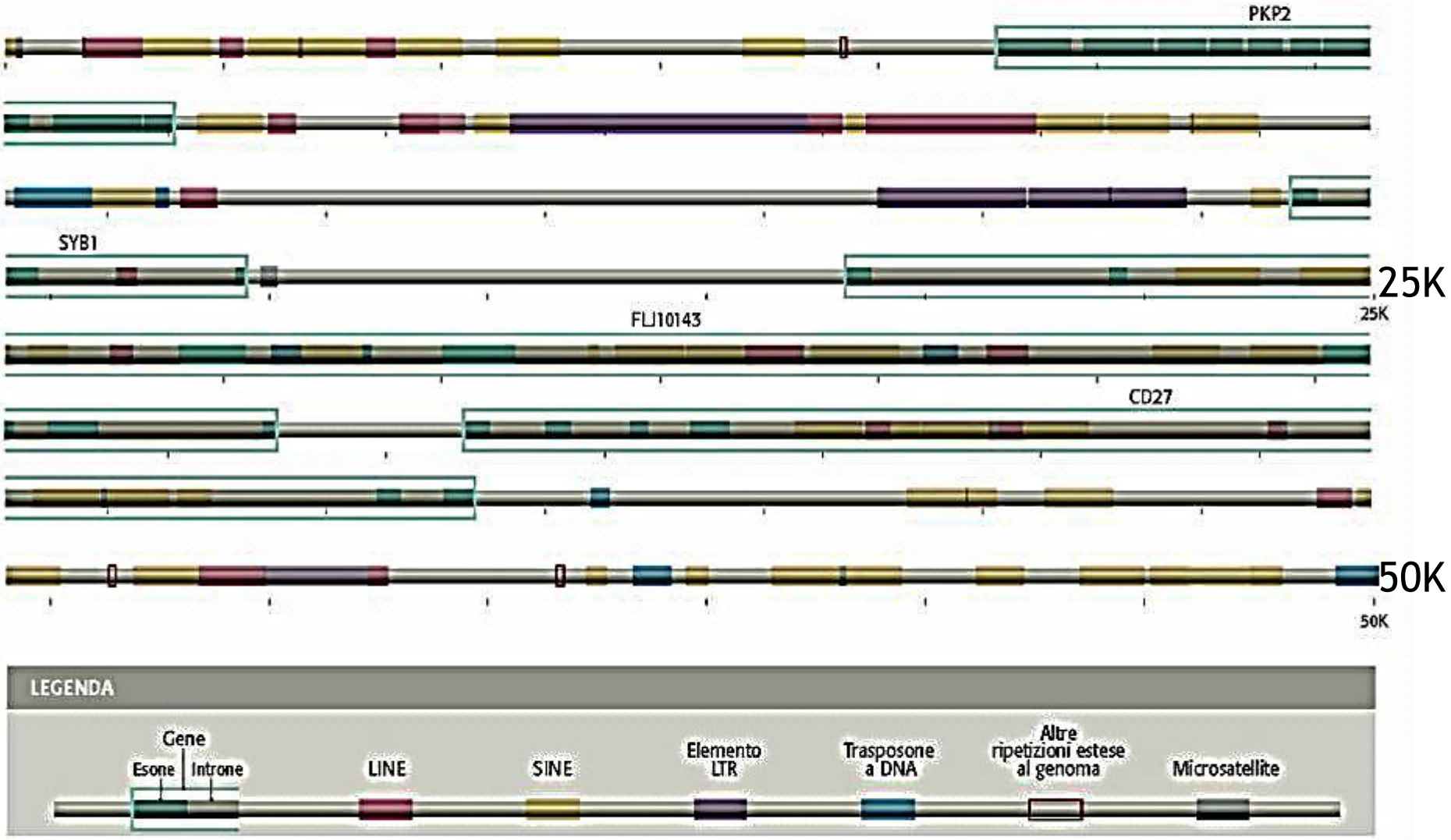


Interspersed repetitive elements - Mobile genetic elements

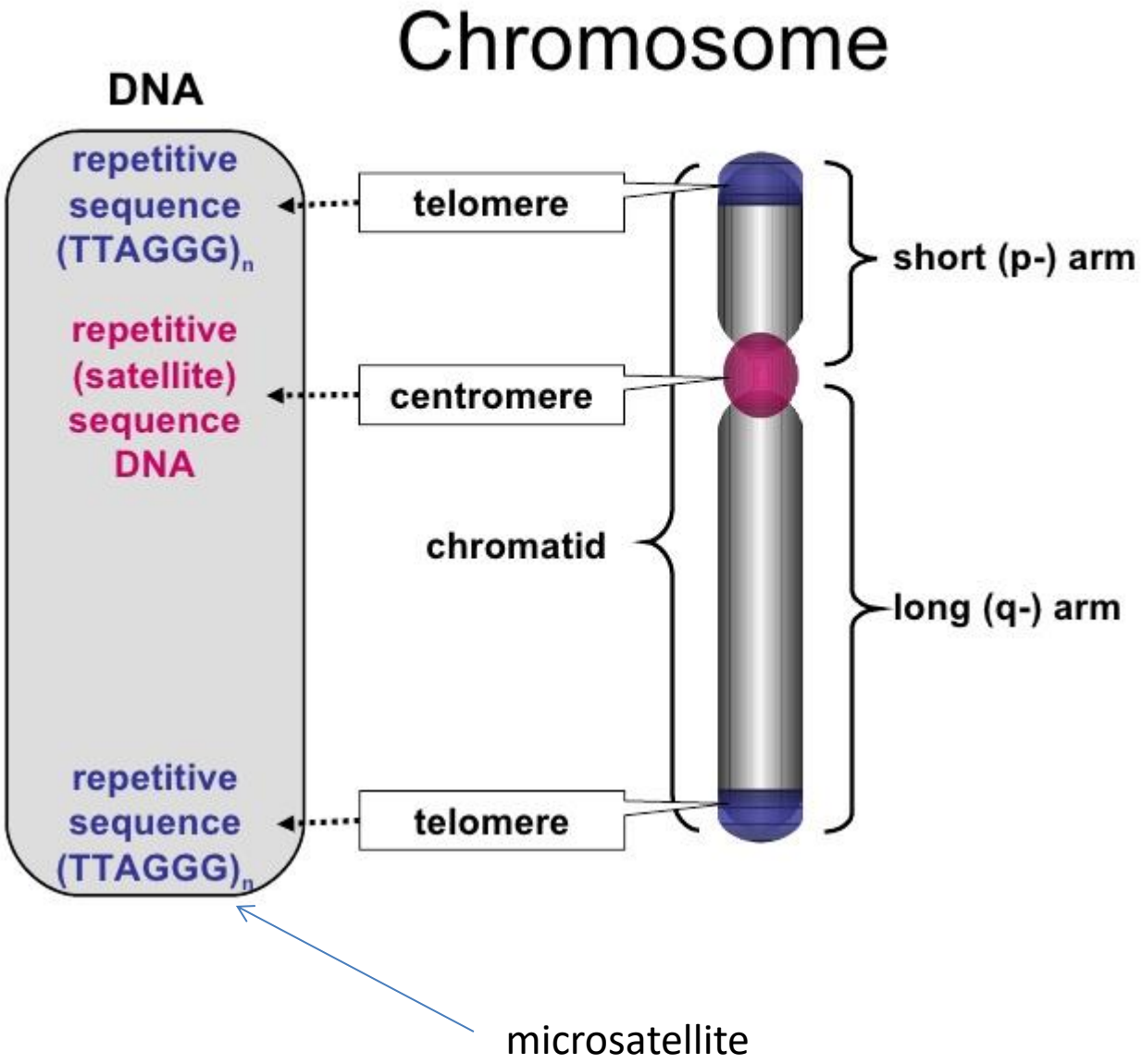
Tabella 1.2 Tipi di ripetizioni estese a tutto il genoma nell'uomo

Tipo di ripetizione	Sottotipo	Numero approssimativo delle copie nel genoma umano
SINE		1.558.000
	Alu	1.090.000
	MIR	393.000
	MIR3	75.000
LINE		868.000
	LINE-1	516.000
	LINE-2	315.000
	LINE+3	37.000
Elementi LTR		443.000
	Classe I ERV	112.000
	Classe II ERV(K)	8.000
	Classe III ERV(L)	83.000
	MaLR	240.000
Trasposoni DNA		294.000
	hAT	195.000
	Tc-1	75.000
	PiggyBac	2.000
	Non classificato	22.000

A 50 Kb tract of the Human genome showing gene position, repeats, microsatellites (taken from Chr. 12)



tandem repeats

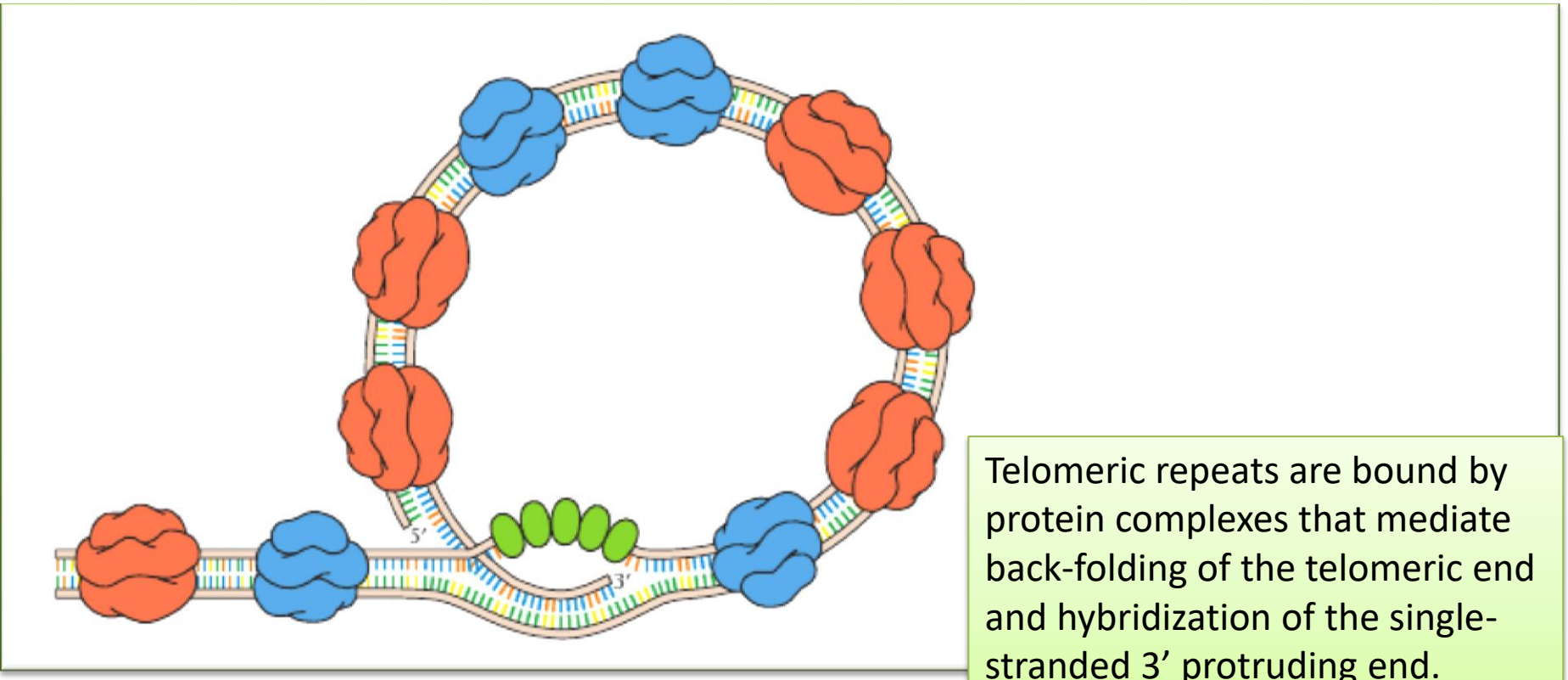


Telomers and telomerase

Short tandem sequence repeats at telomeres.

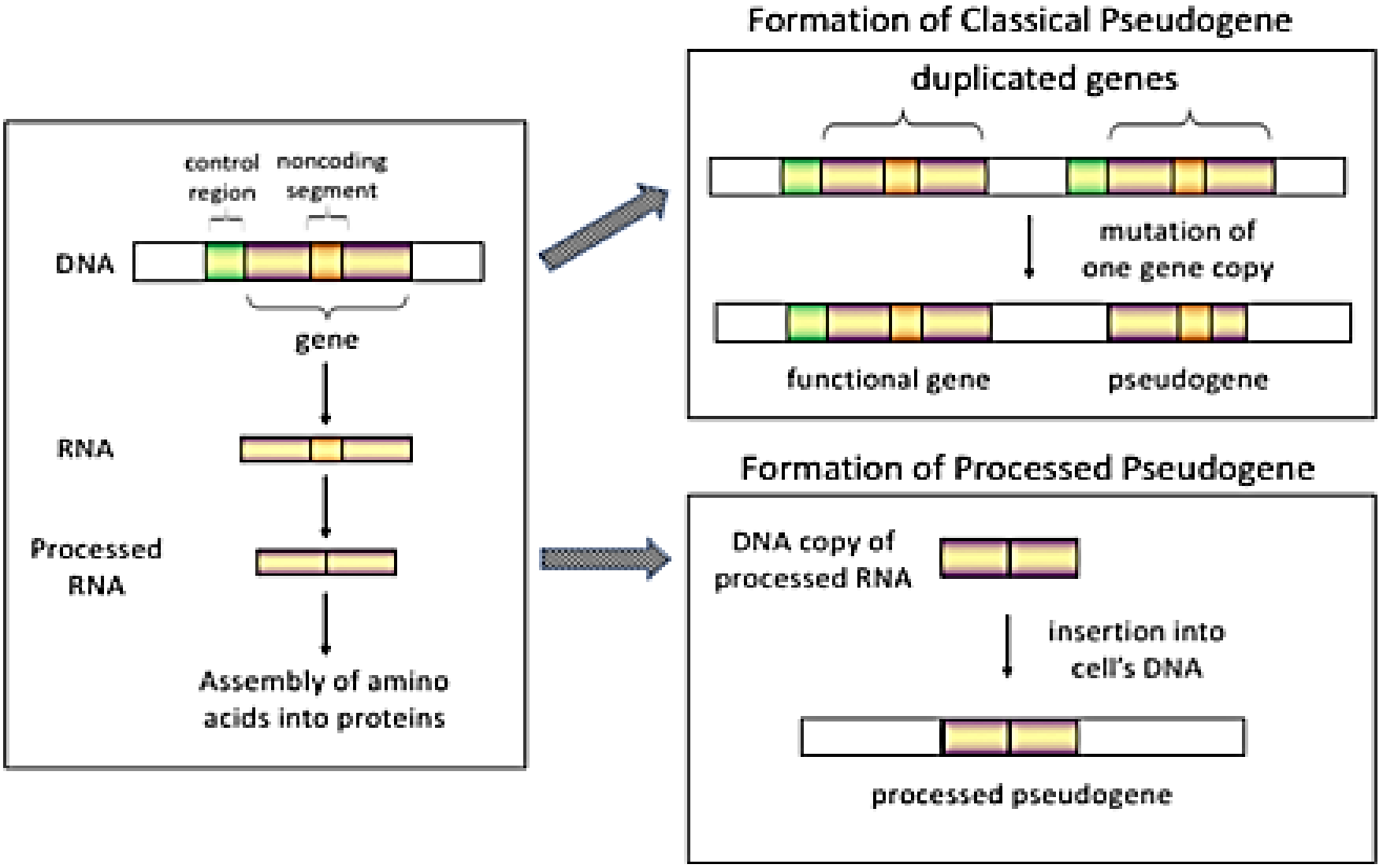
In H. sapiens: TTAGGG (2,500 repeats)

Repeat sequence differs in different organisms



Telomeric repeats are bound by protein complexes that mediate back-folding of the telomeric end and hybridization of the single-stranded 3’ protruding end.

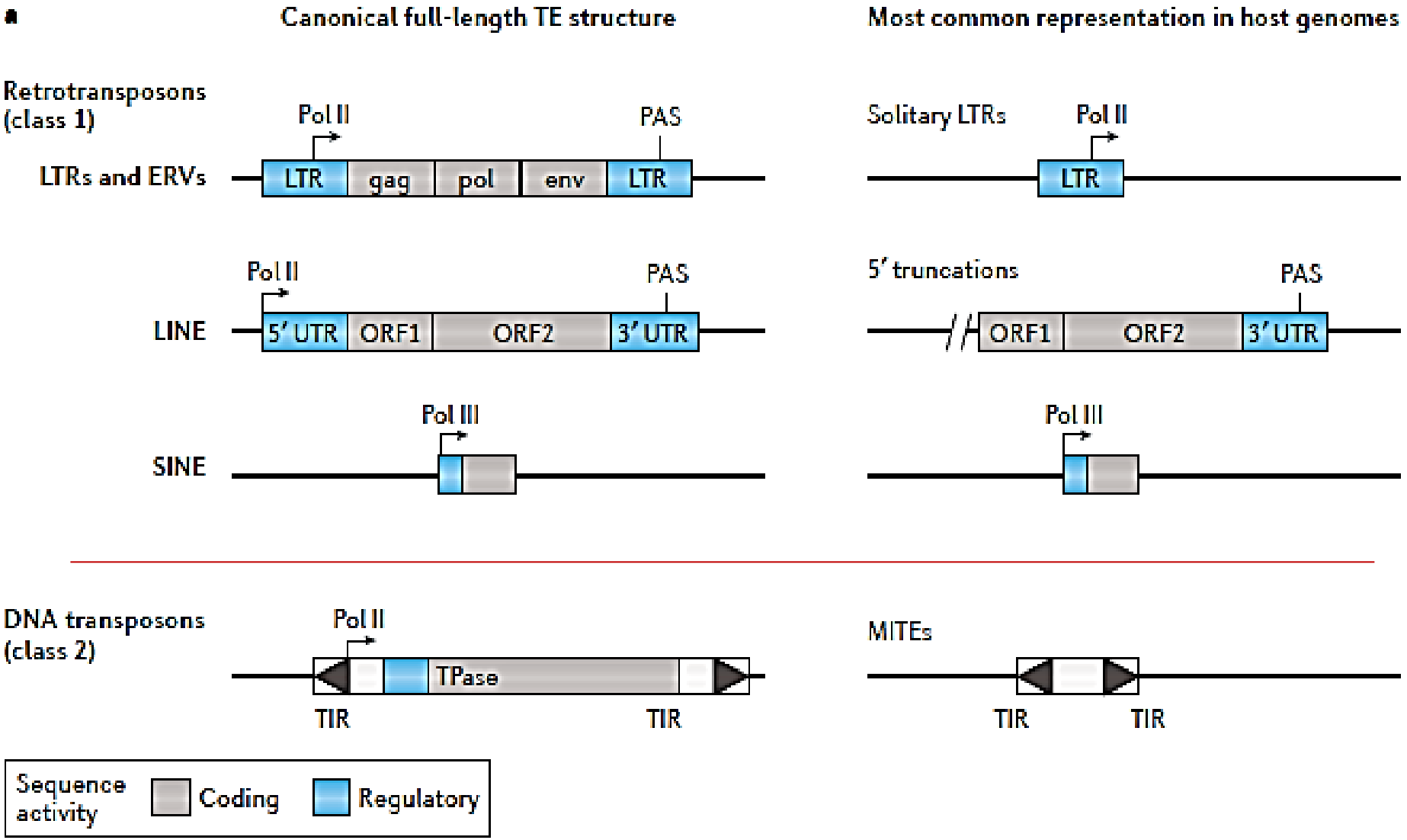
pseudogenes



Transposable Elements (TE)

«mobile» elements deriving from either retrovirus infection or retrocopies of cell genes or even autonomousy replicating DNA tracts that can move to different genomic positions from the original.

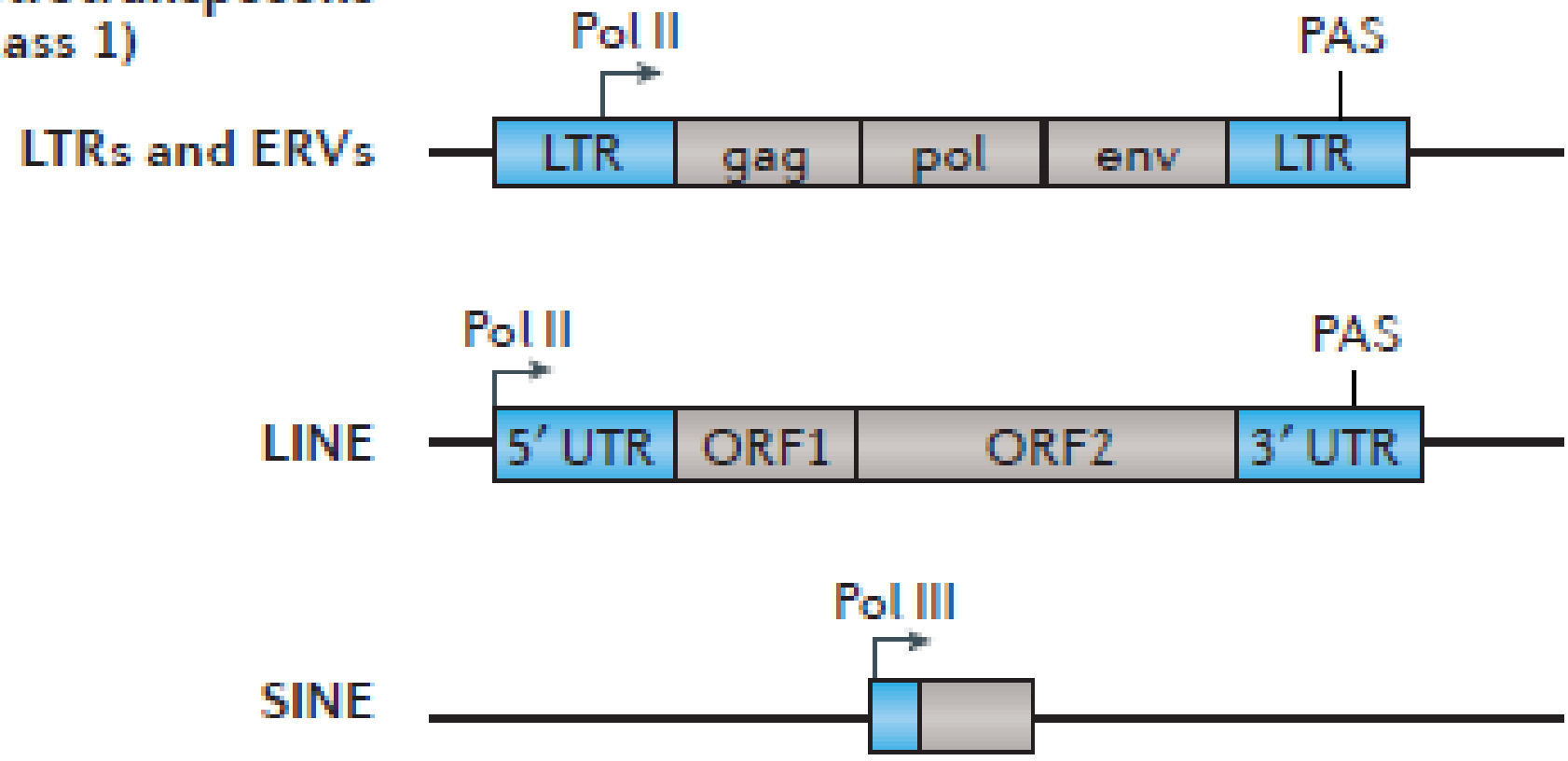
DNA transposons
Retrotransposons
LTR
LINEs
SINEs



TPase = Transposase

from Chuong et al., 2017, Nat Rev Genet

Retrotransposons (class 1)



Pol II & Pol III

Conclusions

In **2003**, only a thiny fraction of the Human Genome sequence could be attributed with a function.

Most of the sequence was thought to be redundant, repetitive and essentially «junk» DNA.

This conclusion, though, was adversed by scientists that studied the phylogenetic conservation, showing that many regions with no apparent function are indeed extremely conserved between organisms (the «dark matter» theory).

For this reason, scientists started several projects to systhematically analyze every regions of the Human (and mouse) genomes to unravel any possible functional role.