

Structural genomics

Part 2

Dynamics of the genome

Master 1ère Année

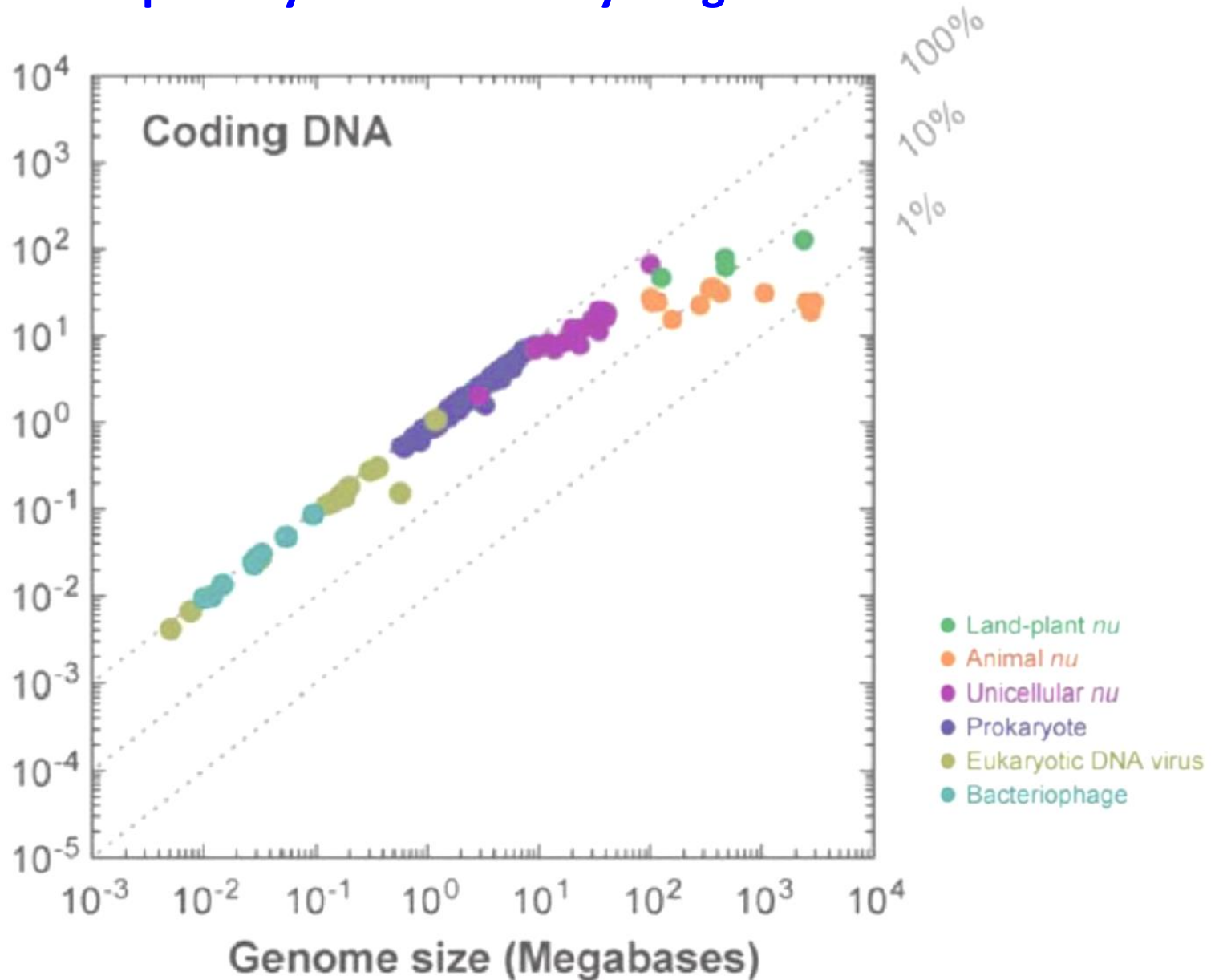
Mention GENIOHME

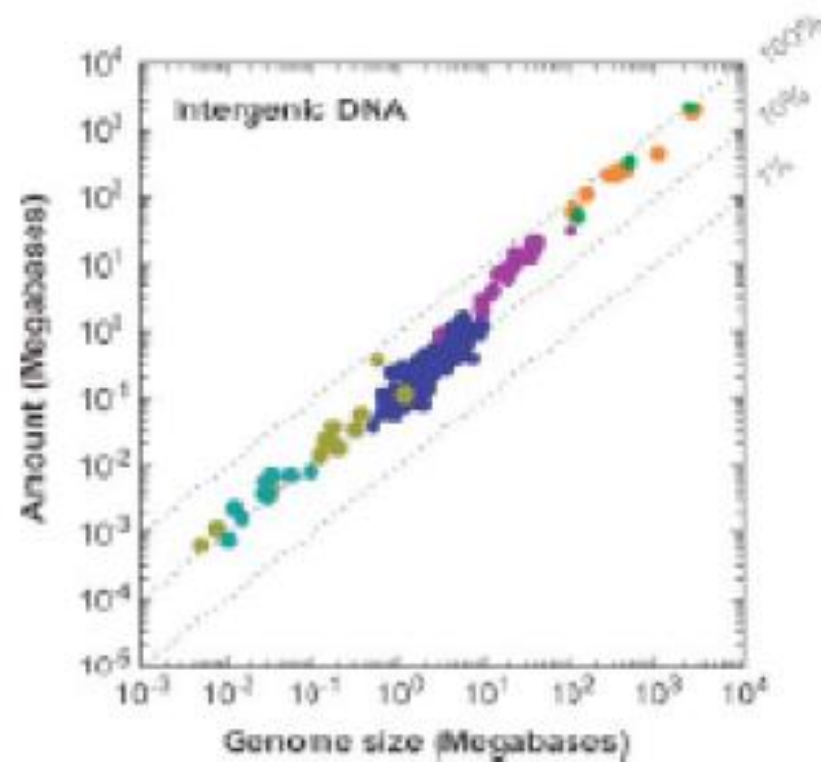
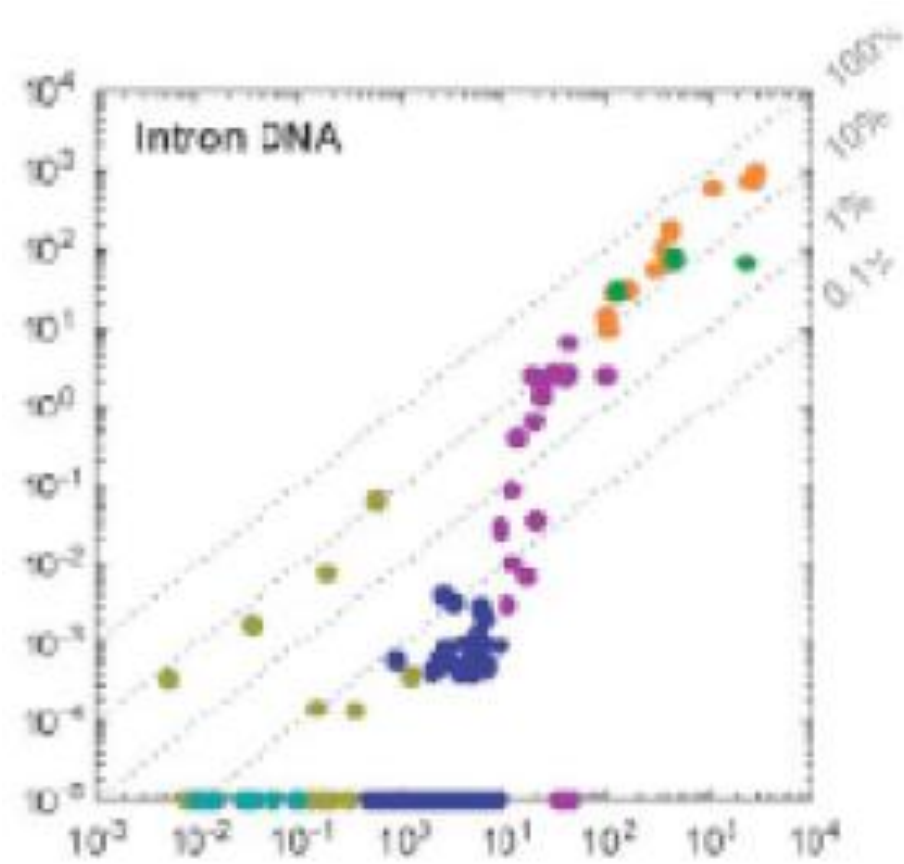
Bénédicte Sturbois

Benedicte.sturbois@univ-evry.fr

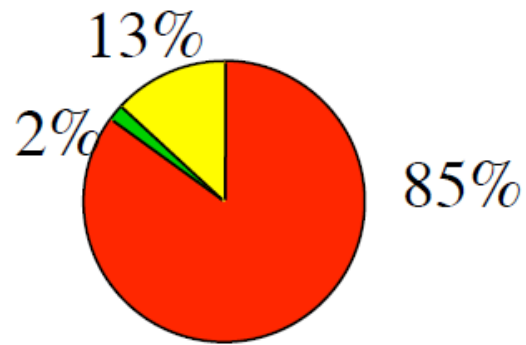
Année 2025-2026

Comparing genomic data from procaryote and eucaryote genomes

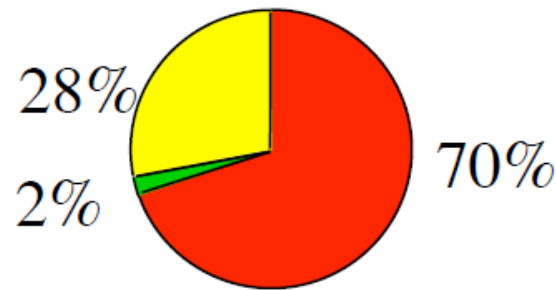




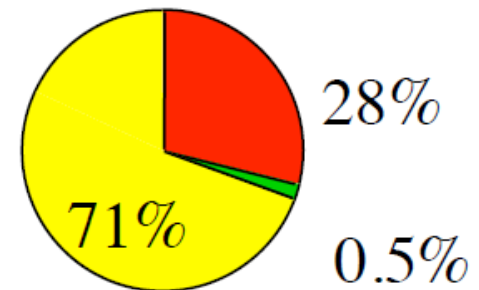
Proportion of different structural elements in entire genomes



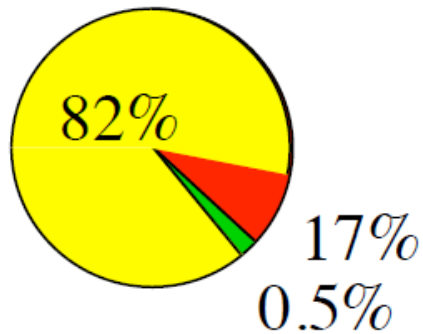
E. coli



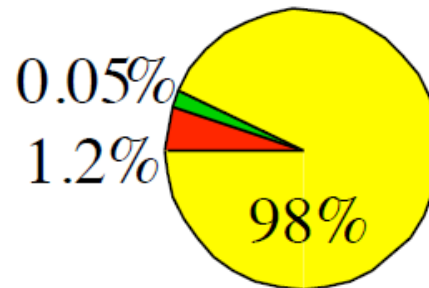
Yeast
S. cerevisiae



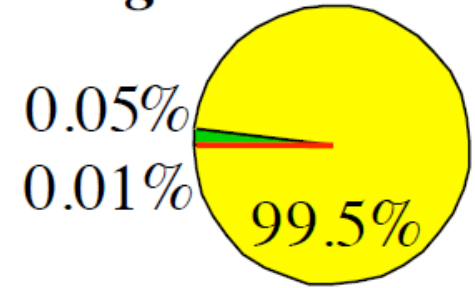
Nematode
C. elegans



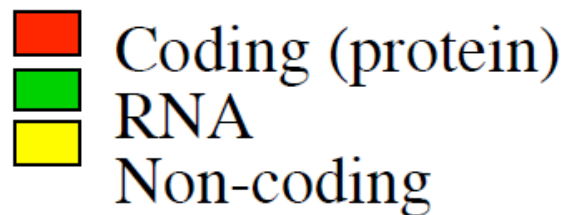
Drosophila



Human



Lungfish
(dipnoi)



Proportion of different structural elements in parts of genomes



Escherichia coli (57 genes)



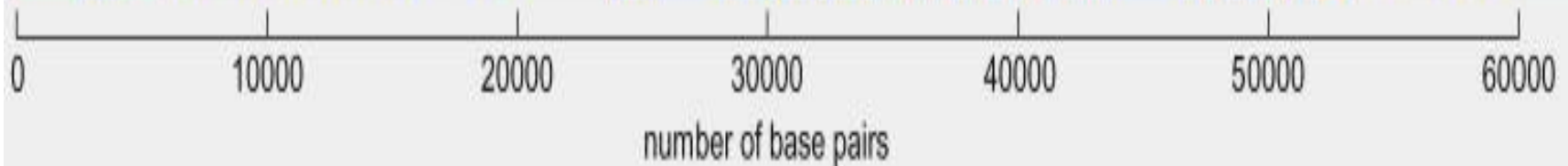
Saccharomyces cerevisiae (31 genes)



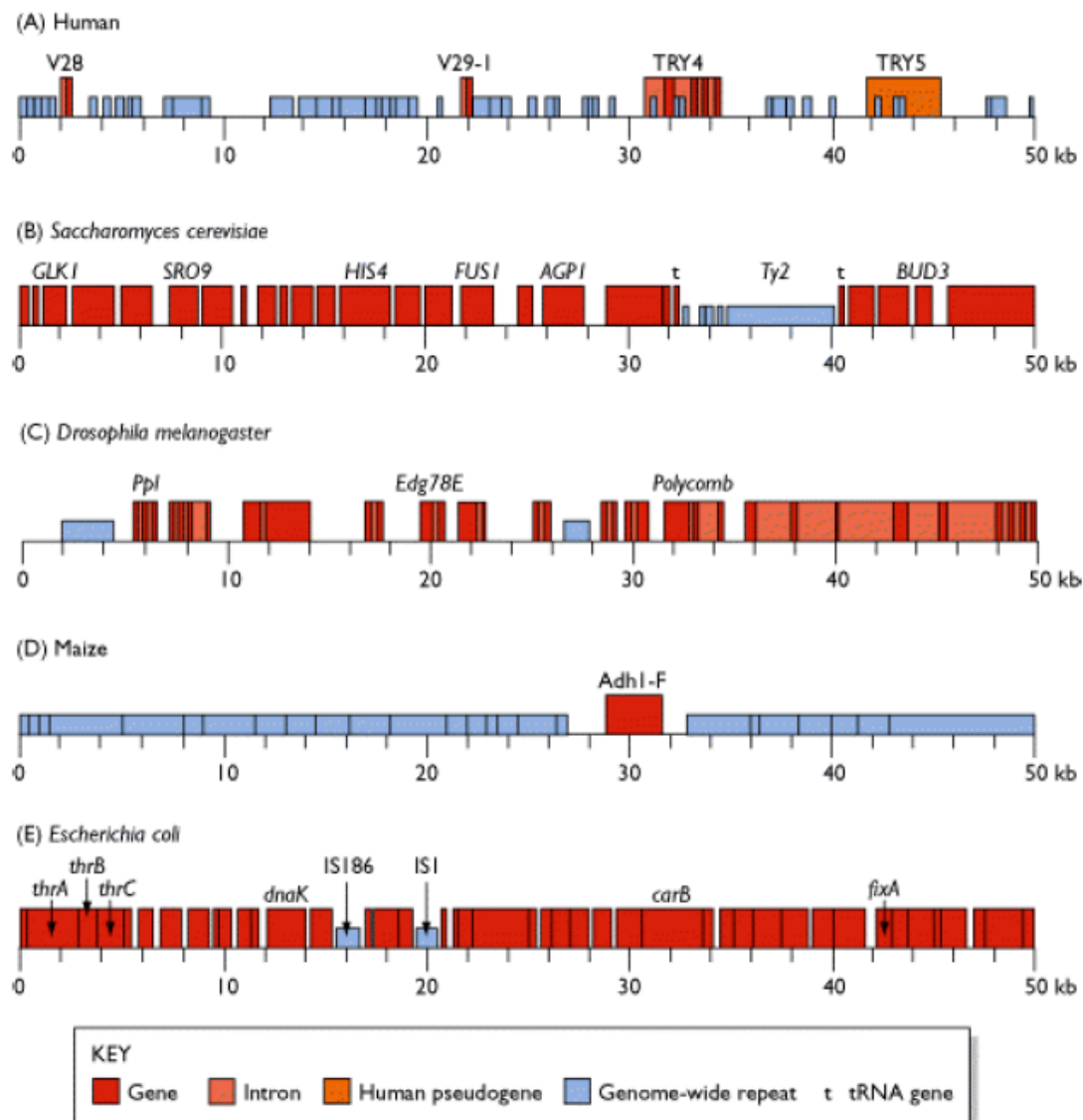
Drosophila melanogaster (9 genes)



Human (2 genes)



Comparison of the structure of genomes



Comparaison d'un segment quelconque de 50 kb des génomes humain, de levure, de drosophile, de maïs et de la bactérie *Escherichia coli*.

Plus les organismes sont simples, plus leur génome est « économe » :

- avec une plus grande densité de gènes et une moins grande densité de séquences répétées
- avec des gènes « simples » c'est-à-dire possédant peu ou pas d'introns.

The human genome

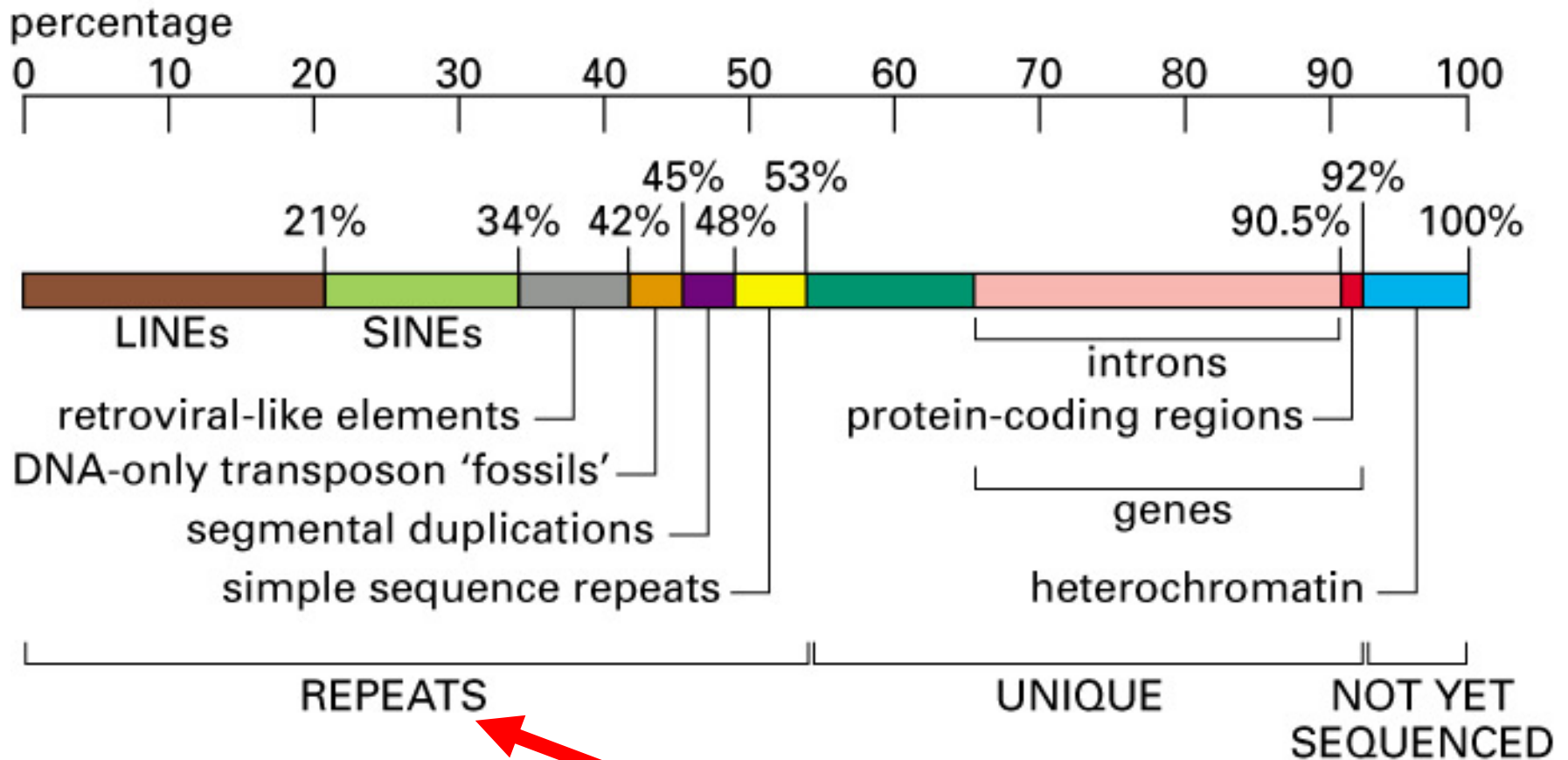


Figure 4-17. Molecular Biology of the Cell, 4th Edition.

A cause de duplications

Comparison between Yeast and Human genomes

	levure	homme
Gènes (codant des protéines)	5 770	~ 23 000
Introns	280	> 100 000
Pseudogènes	10	> 25 000
Éléments mobiles	~ 50	> 1 100 000

Nombre de familles de protéines

~ 4 100

~10 000

Redondance (gènes paralogues)

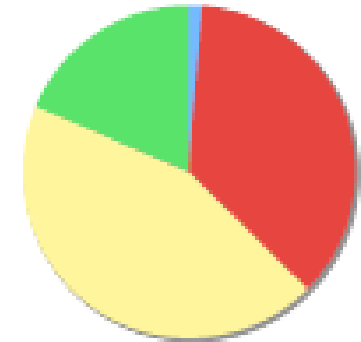
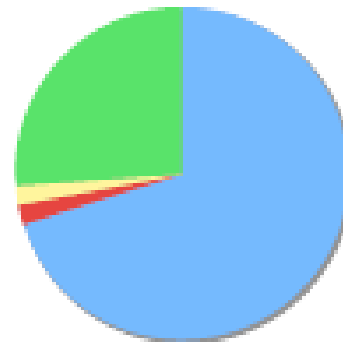
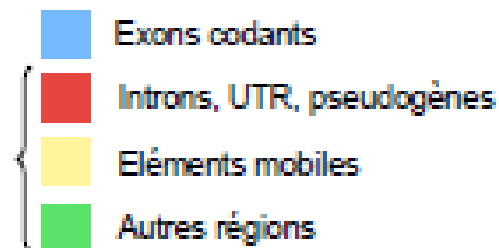
1,4 x

2,3 x

fonctions

régulations

évolution



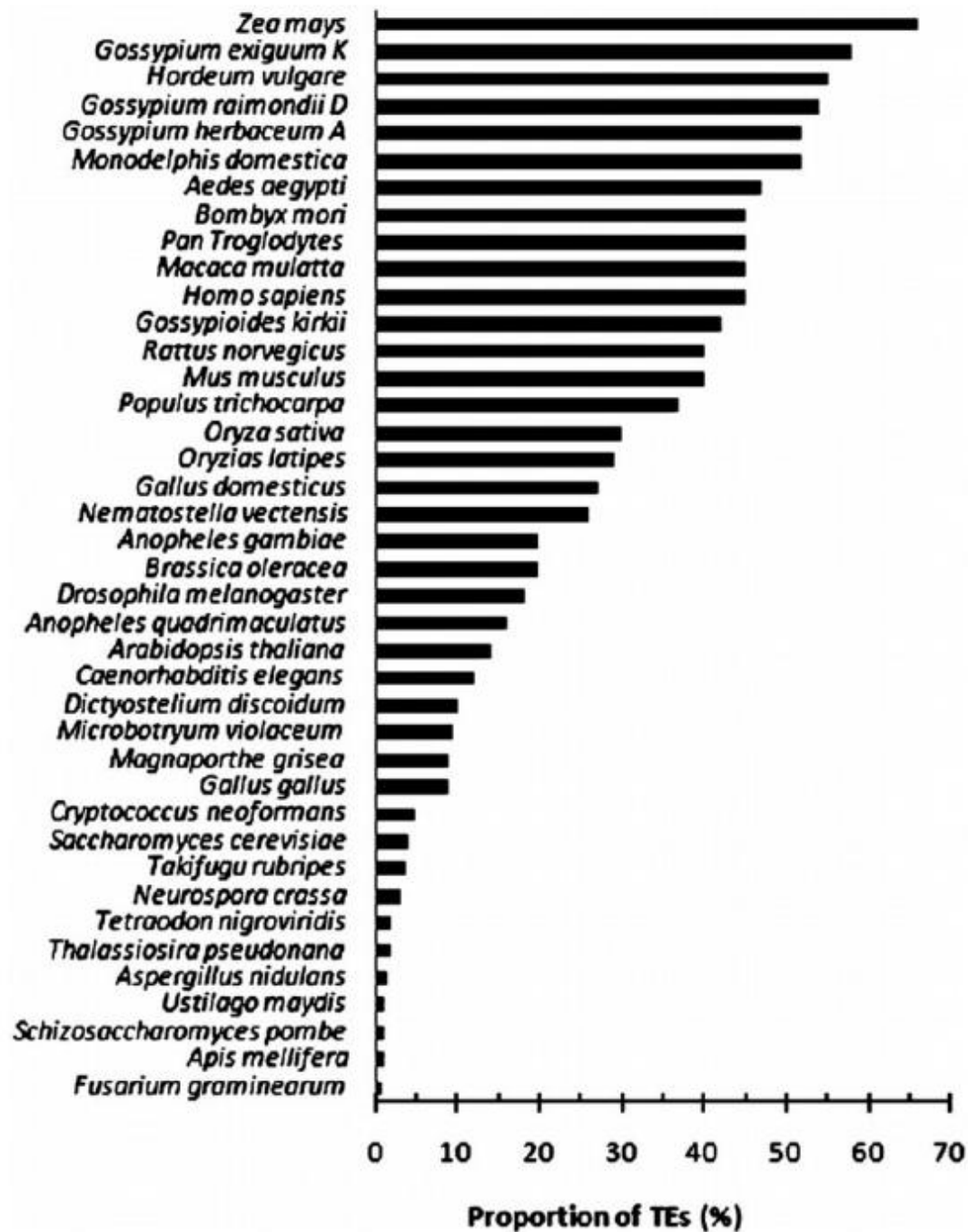
	Yeast	Nematode worm	Drosophile	Arabidopsis	Human
Physical size(Mb)	13	100	180	125	3.000
Average size of 1 cM (kb)	3	500	300	220	800
[G+C] content	38%	36%	nd	41%	41%
Gene number	6.200	19.100	13.600	25.500	~30.000
Coding fraction	68%	27%	13%	29%	1,4%
Average number of exons per gene	1,04	5,5	4,6	5,2	8,7
Gene size (kb)	1,4	2,7	3	2,1	28
Average coding size (without introns)	1.450	1.311	1.497	1.300	1.340
Average size of exons per gene (pb)	1.450	218	150	250	145
Average size of introns per gene (pb)	500	267	487	168	~3.300
Gene frequency (per kb)	2	4,8 / 6	9	4,5	~100
ARNt number	273	584	284	589	535
Chromosomal Localisation of NOR	12	1	X et Y	2, 4	13, 14, 15, 21, 22

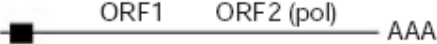
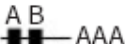




			Size(kb)	nb of genes	Size of genes
Eubactetria		<i>Escherichia coli</i>	4636	4400	1 kb
		<i>Mycoplasma genitalium</i>	580	483	1 kb
Archaeobacteria		<i>Pyrococcus abyssi</i>	1700	1765	1 kb
		<i>Archeoglobus fulgidus</i>	2178	2437	0,8 kb
Eucaryotes	Worm	<i>Giardia lamblia</i>	12 000	?	?
	Sporozoaire	<i>Plasmodium falciparum</i>	25 000	?	?
	Embryophytes	<i>Pinus sp.</i>	68 000 000	?	?
		<i>Arabidopsis thaliana</i>	115 000	25500	3-4 kb
	Mycetes	<i>Saccharomyces cerevisiae</i>	12155	6400	2 kb
		<i>Podospora anserina</i>	34 000	9000 ?	3 kb
	Metazoaires	<i>Caenorhabditis elegans</i>	97 565	18000	3 kb
		<i>Drosophila melanogaster</i>	140 000	13 500	3 kb
		<i>Fugu rubriceps</i>	400 000	30 000	3 kb
		<i>Homo sapiens</i>	3 286 000	30 000	5 kb → 100 Mb

Dynamics of genomes

A. Transposable elements

- DNA sequences that can move within the genome
- More than half of the DNA in eukaryotes
- **Two major classes:**
 - ***Transposons***: Move via a cut and paste system
 - ***Retrotransposons***: Move via an intermediary
- Play a role in the evolutionary capacity of "massive and messy genomes" such as those of plants
 - ✓ Creating new genes
 - ✓ Potentially destructive
 - ✓ Gene editing
 - ✓ Gene programming or re-programming
- Transposition leads to genome expansion
- Transposition is done in real time: differences between lines or descendants



			Length	Copy number	Fraction of genome
LINES	Autonomous		6–8 kb	850,000	21%
SINEs	Non-autonomous		100–300 bp	1,500,000	13%
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Non-autonomous		1.5–3 kb		
DNA transposon fossils	Autonomous		2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

- ❖ TEs are “selfish genes” which when activated can insert copies of themselves into the genome. When this happens in the germline, these insertions are transmitted to the next generation.
- ❖ Four families, based on the mechanism by which they copy themselves:
 - **LINES** (Long Interspersed Nuclear Elements, autonomous)
 - **SINEs** (Short Interspersed Nuclear Elements, use LINE proteins for life cycle)
 - **LTR elements** (Long Terminal Repeats; derived from retroviruses)
 - **DNA transposons** (replicate without RNA intermediary)

	Arabidopsis	Nematode	Drosophile	Mice	Human
LINE/SINE	0,5%	0,4%	4,7% (0,7% + 13,2%)	28%	28%
Retrovirus type Sequences	4,8%	0%	6,4% (1,5% + 16,9%)	10%	7 %
Transposons Type Sequences	5,1%	5,3%	3,6% (0,7% + 9,9%)	1%	3%
Total	10,5%	6,5%	14,9% (3,1% + 40,2%)	38%	38%

Conséquences de l'activité des éléments mobiles sur le génome et l'expression des gènes

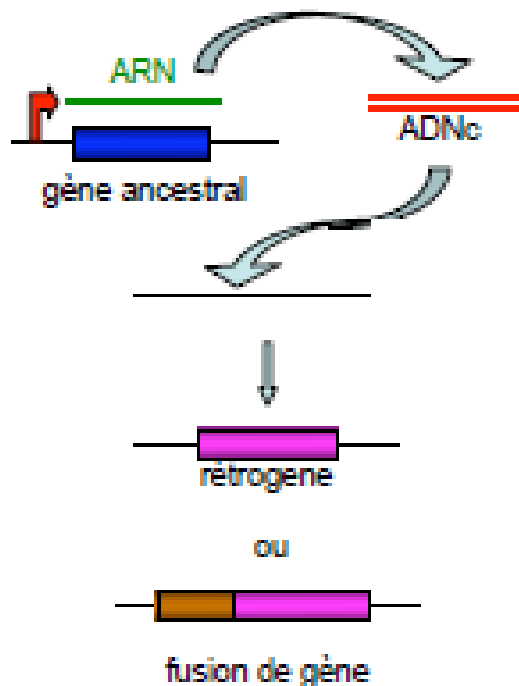
- **Inactivation** d'un gène (exemple historique: le mutant *white* de la Drosophile)
ex. d'insertions délétères de l'élément LINE-1:

gène du facteur VIII: naissance de garçons hémophiles (le gène est sur le chromosome X)
gène APC: cancer du colon
gène de la dystrophine: myopathies

- **activation** de l'expression d'un gène voisin (ex. certaines tumeurs)
- induction de **cassures** chromosomiques (ex. les transposons non duplicatifs laissent des cicatrices)
- induction de **réarrangements** chromosomiques (ex. translocations réciproques entre copies du même élément, cascades de « ponts cassures - fusions »)
- **dysgénèse hybride** (ex. deux lignées de Drosophile qui diffèrent par leur lot de transposons P ou I ne peuvent plus se croiser entre elles, formation d'espèces nouvelles)

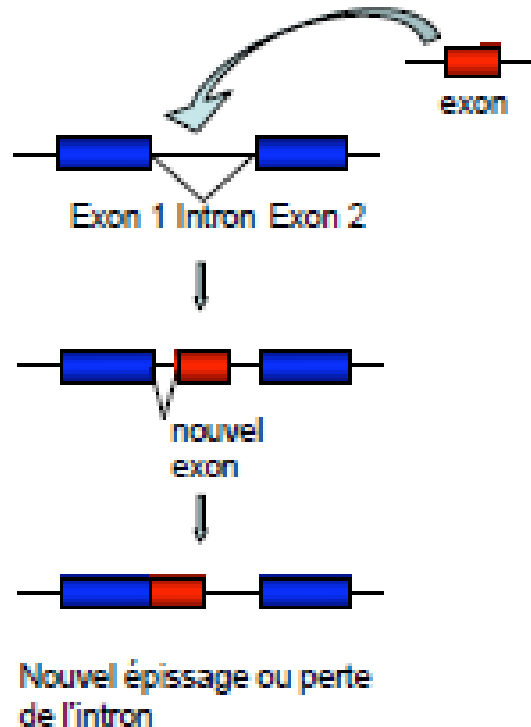
Consequences of TE activity on genome evolution

Retrogene formation



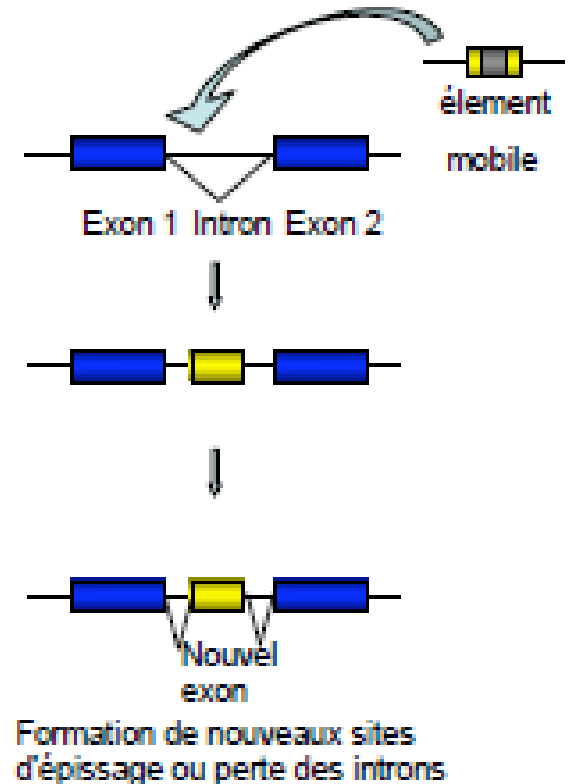
1 % des gènes humains, plus de nombreux pseudogènes, sont issus de ce processus

Exon shuffling



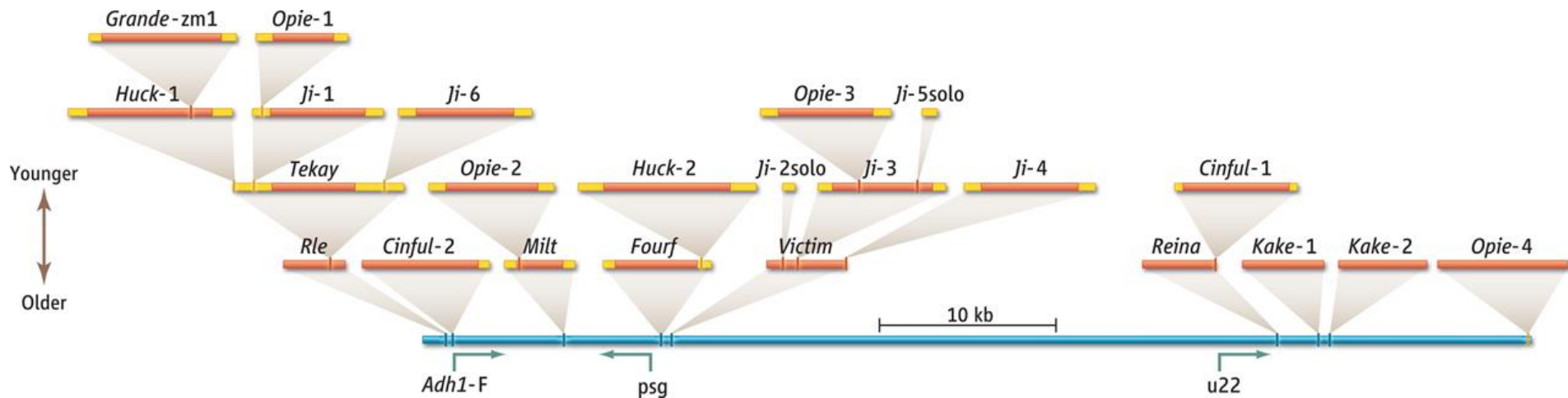
~ 10 % des exons des génomes eucaryotes proviennent de ce processus

Insertion of TE



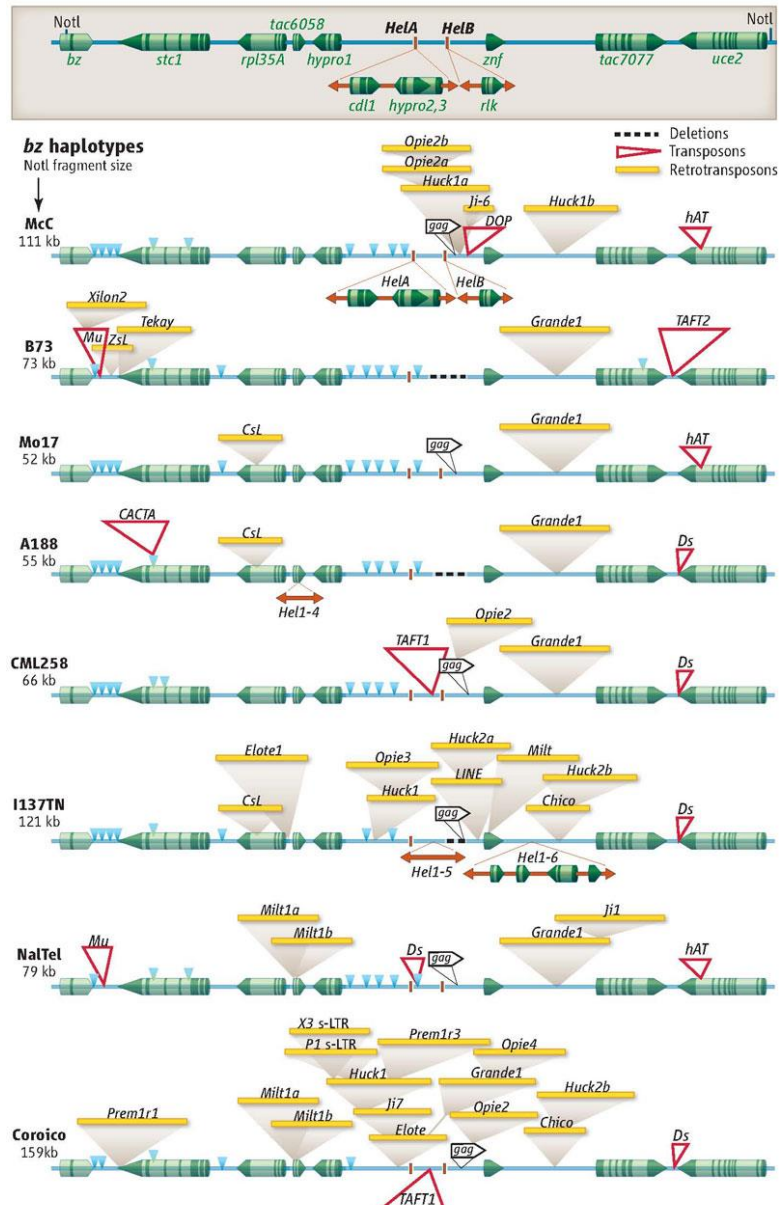
~ 4 % des nouveaux exons du génome humain proviennent de ce processus

Fig. 6. The arrangement of retrotransposons in the maize *adh1-F* region.

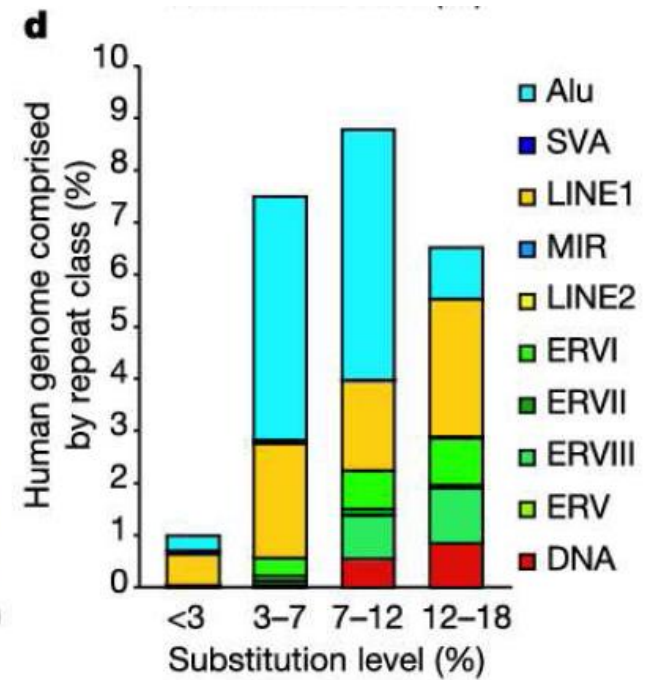
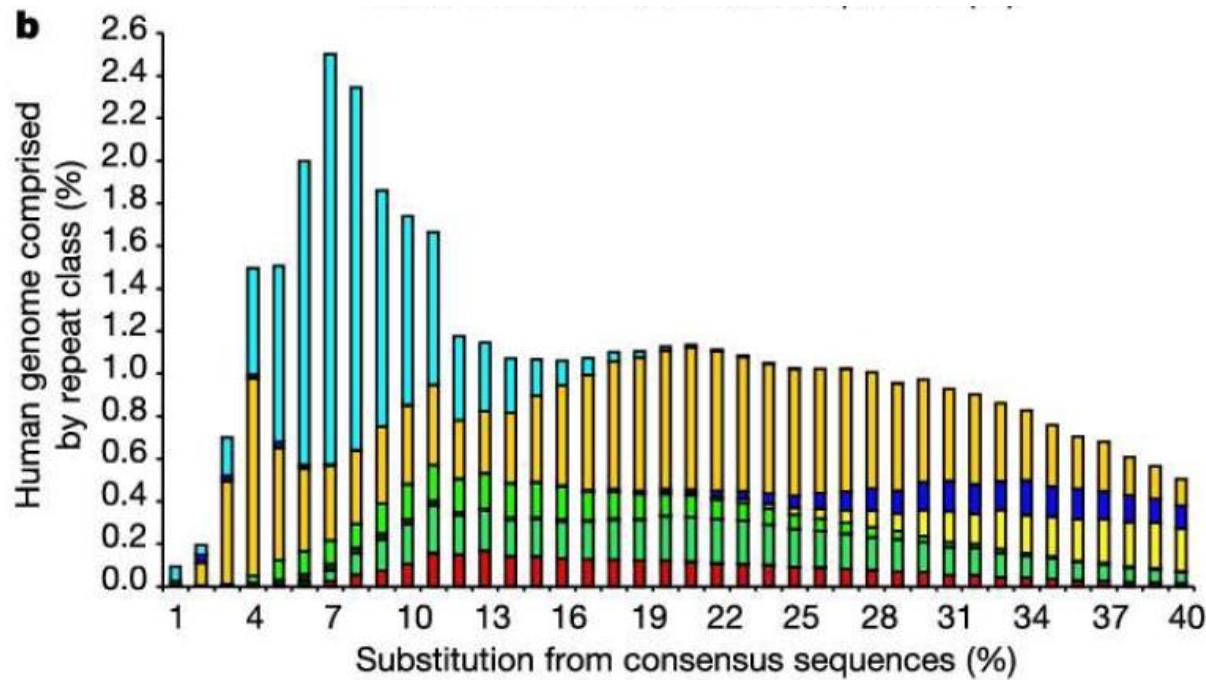


The short lines represent retrotransposons, with the internal domains represented in orange and the LTRs in yellow. Younger insertions within older insertions are represented by the successive rows from the bottom to the top of the diagram. Small arrows show the direction of transcription of the genes

Fig. 7. The organization of the sequence adjacent to the bronze (bz) gene in eight different lines (haplotypes) of maize.



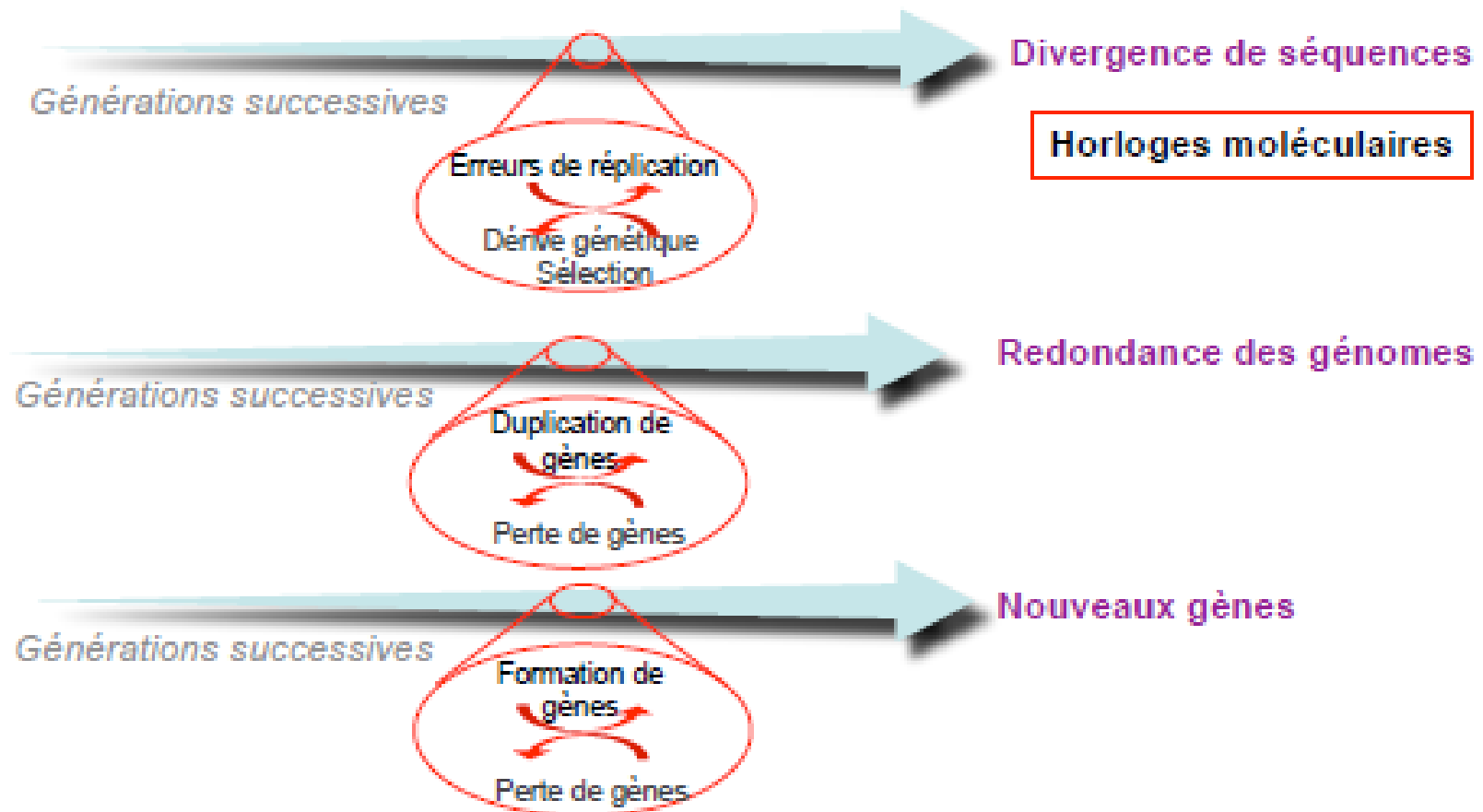
The genes in this region are shown in the top diagram: *bz*, *stc1*, *rpl35A*, *tac6058*, *hypro1*, *znf*, *tac7077*, and *uce2*. The orientation of the gene is indicated by the direction of the green pentagon, pointing in the direction of transcription; exons are represented in dark green and introns in light green. Each haplotype is identified by its name and the size of the cloned NotI fragment. The same symbols are used for gene fragments carried by Helitrons (Hels), which are represented as bidirectional arrows below the line for each haplotype. Vacant sites for *HelA* and *HelB* are provided as reference points and marked by short vertical red bars. Dashed lines represent deletions. Retrotransposons are represented by yellow bars. DNA transposons and TAFTs (TA-flanked transposons), which are probably also DNA transposons, are represented by red triangles; small insertions are represented by light blue triangles.



- ❖ Age of a TE can be determined (approximately) by counting average number of substitutions from the consensus sequence, supposed to be the ancestral state.
- ❖ Histogram of TEs versus age shows the activity over time. Alus have been very active, but recently things have quieted down in human.

Principes généraux de l'évolution des génomes

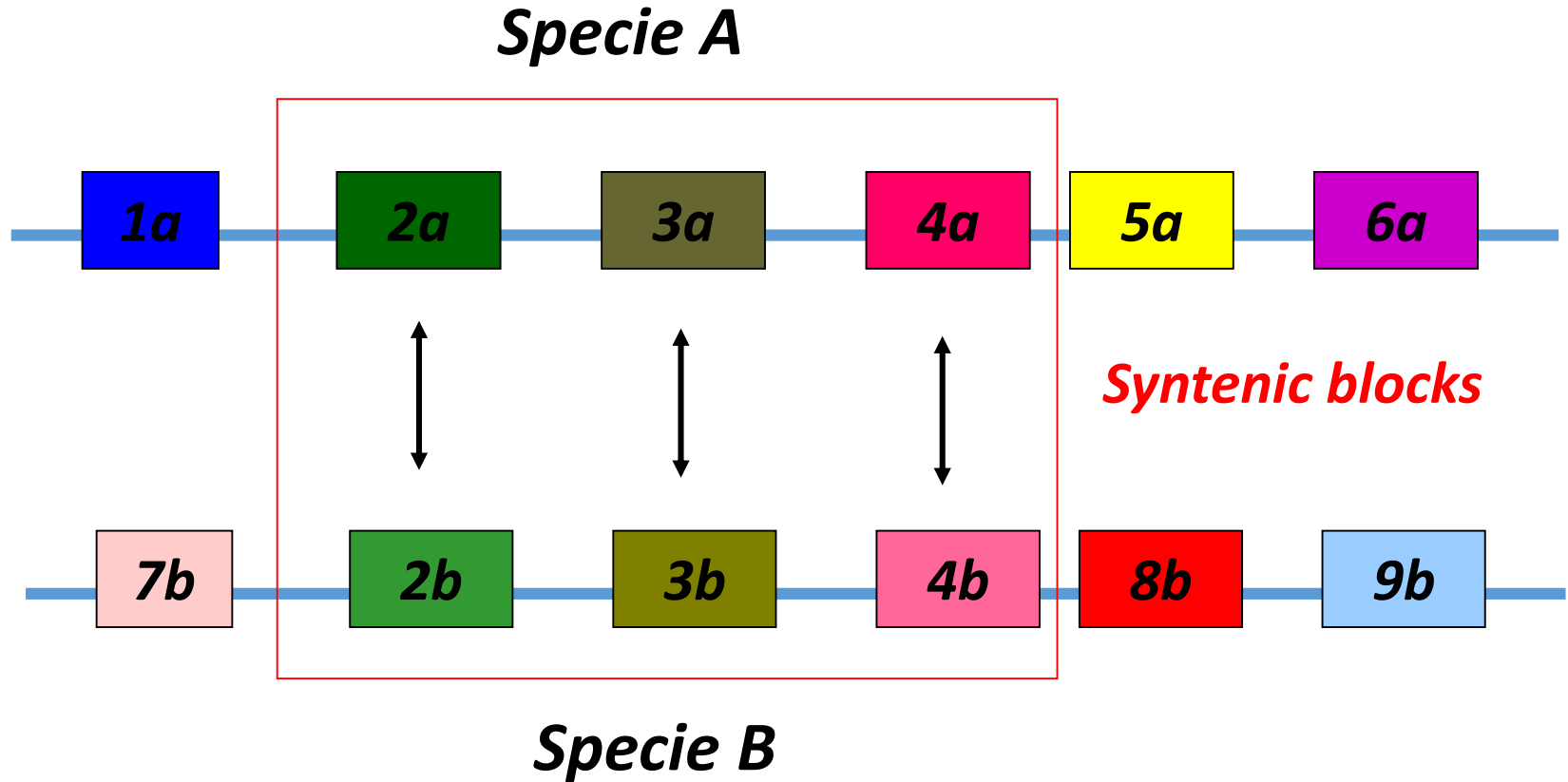
Les organismes vivants n'évoluent pas, ils se reproduisent. Ce sont les populations qui évoluent.



- Les génomes ne sont que des instantanés de structures en évolution constante à chaque génération.
- Les génomes des individus d'une même espèce ne sont pas strictement identiques.

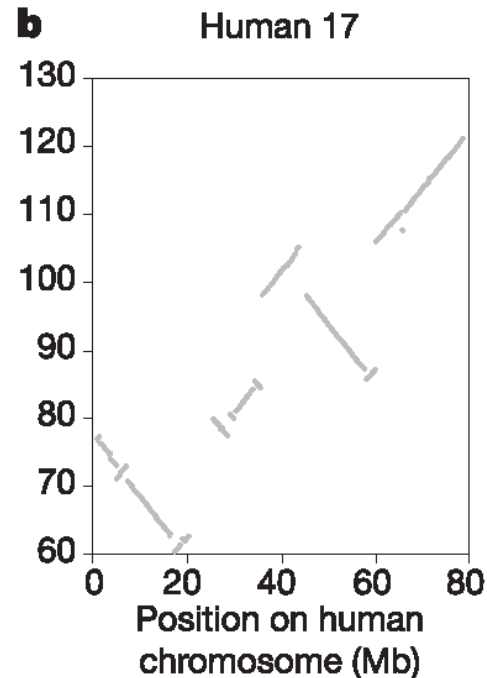
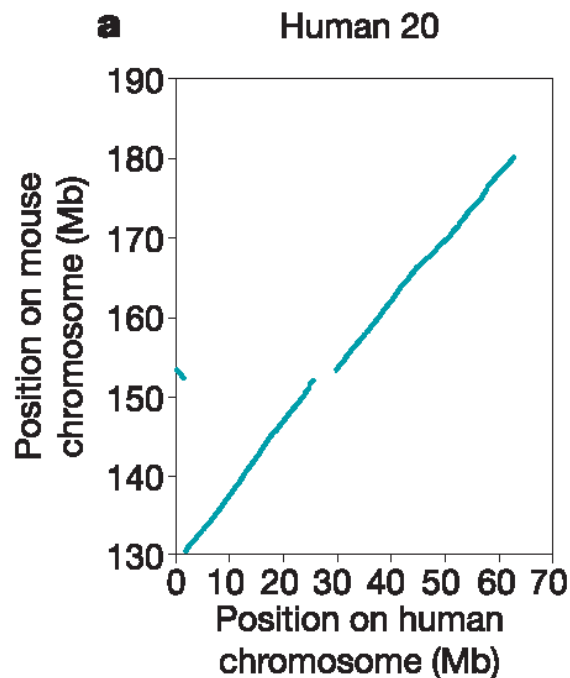
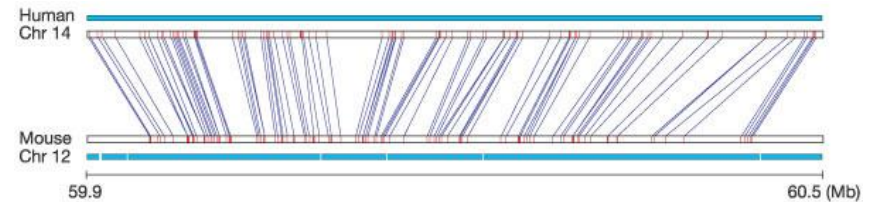
B. Duplication

❖ Synteny / genome evolution



❖ Whole genome comparison

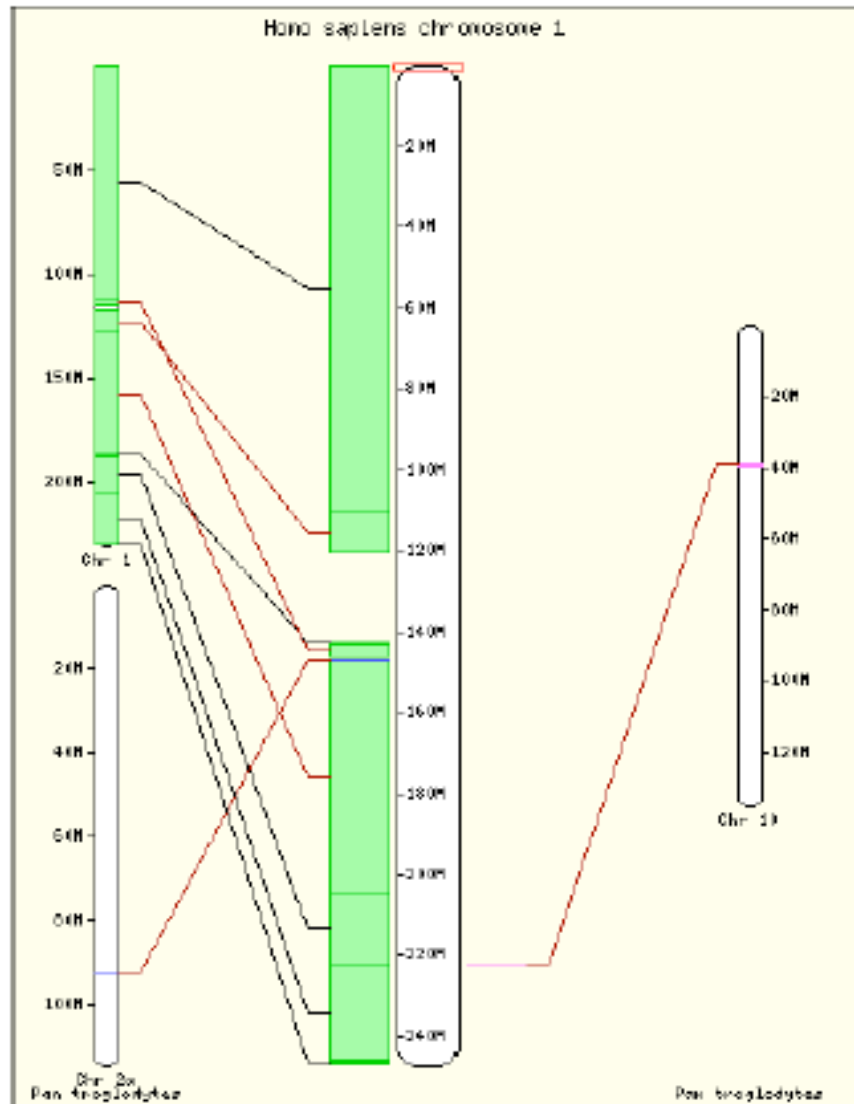
- ❖ Large stretches of sequences = **Macrosynteny**
- ❖ Divergence until 450Mya (*fugu*-human) with sufficient similarity



Syntenic maps indicating relationship between chromosomal regions of two species

Chimpanzee

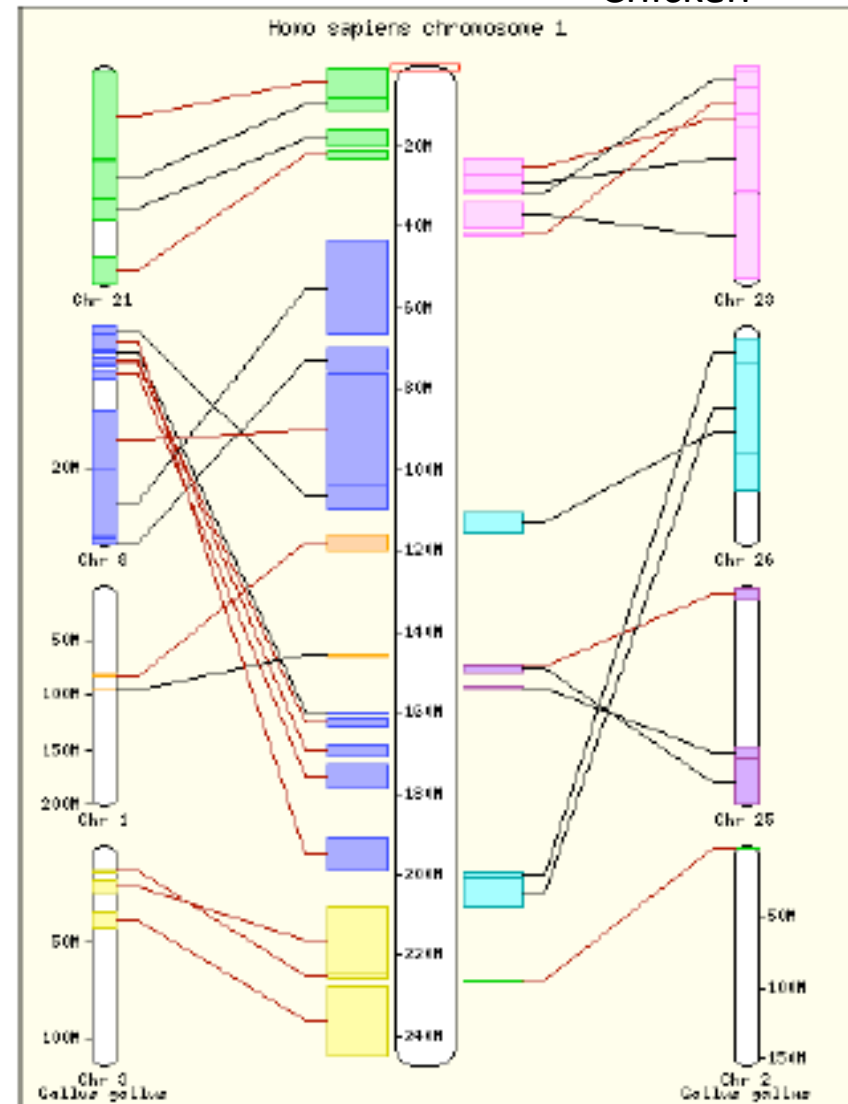
Chimpanzee



Human

Chicken

Chicken

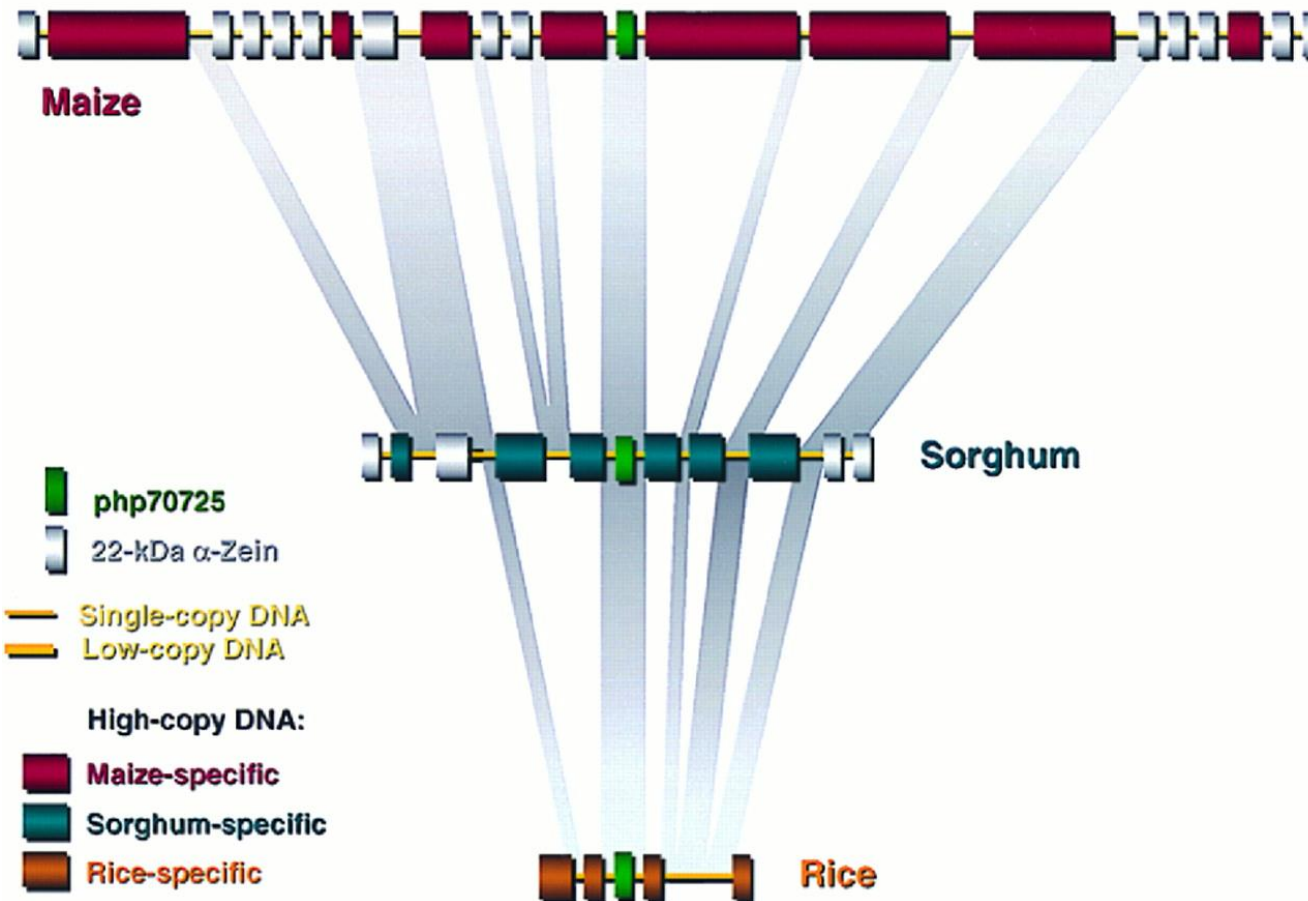


Human

❖ Whole genome comparison

❖ Conservation of **microsynteny**

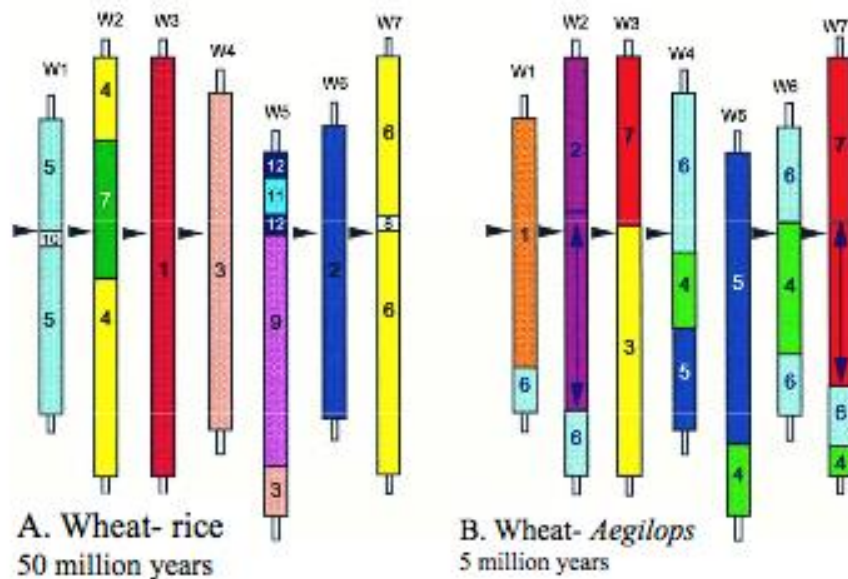
A MODEL OF MICROSYNTENY IN THE 22-KDa α -ZEIN CLUSTER



Tailles par rapport
au riz

Arabidopsis	0.34
Sorghum	1.74
Maize	5.81
Barley	11.4
Wheat	37.22

❖ Chromosomal rearrangements



Different organisms show different rates of chromosome rearrangements

Comparative maps of the wheat genome described in terms of the rice genome (A) and the *Aegilops umbellulata* genome (B).

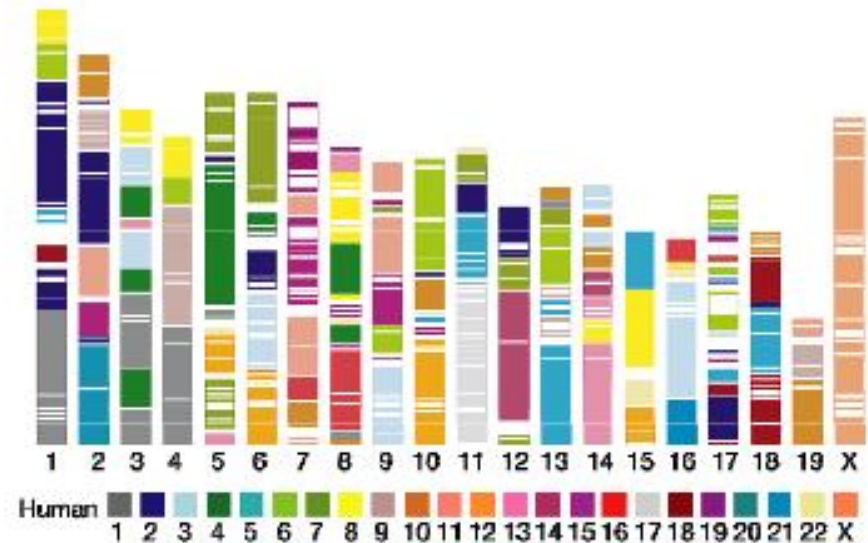
(A) Data from Kurata et al. (1994), Van Deynze et al. (1995)

(B) Data from Zhang et al. (1998).

Human – Mouse (80 million years): Many rearrangements (180 blocks) but some large conserved regions!

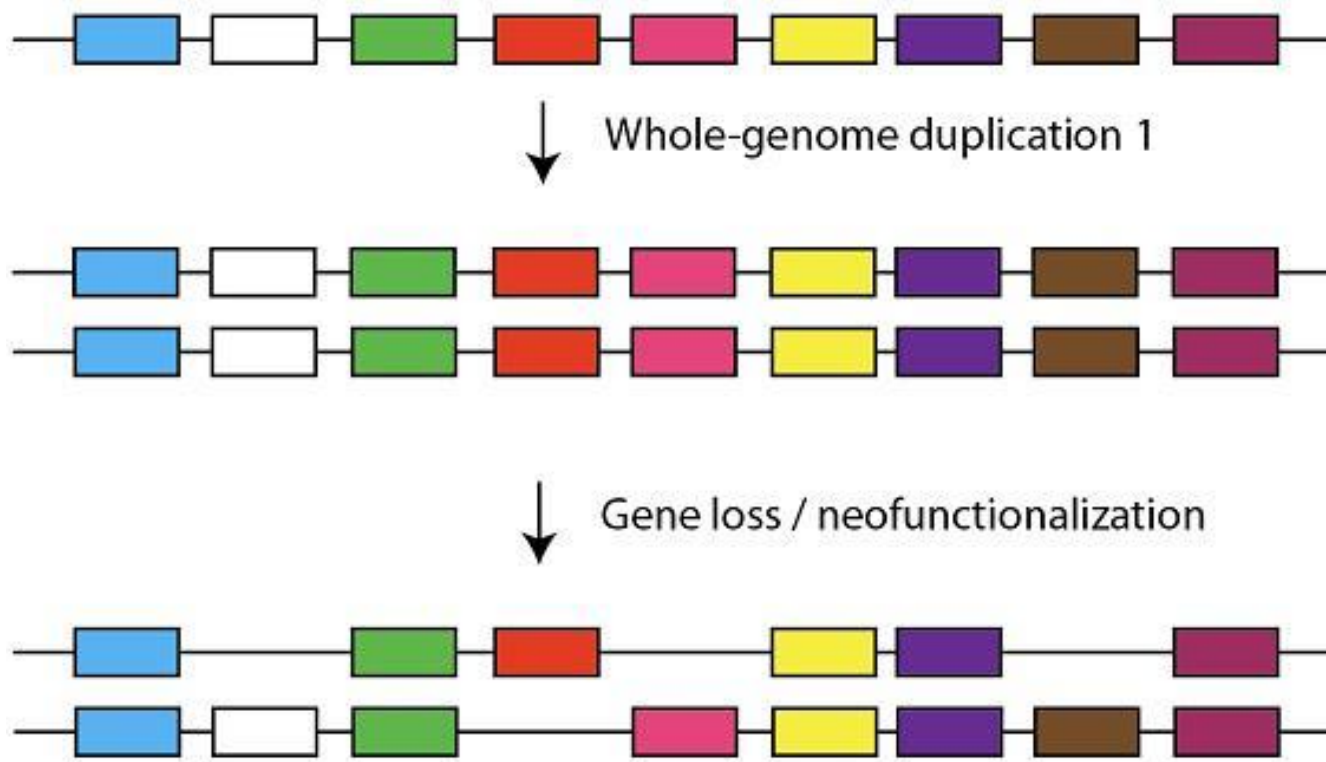
Human – Chimpanzee (5 million years) almost completely colinear.

- Allows transfer of genetic mapping data between organisms.
- Allows functional analysis of human genes via precise deletion of their colinear regions in the mouse genome



❖ Genome expansion and contraction

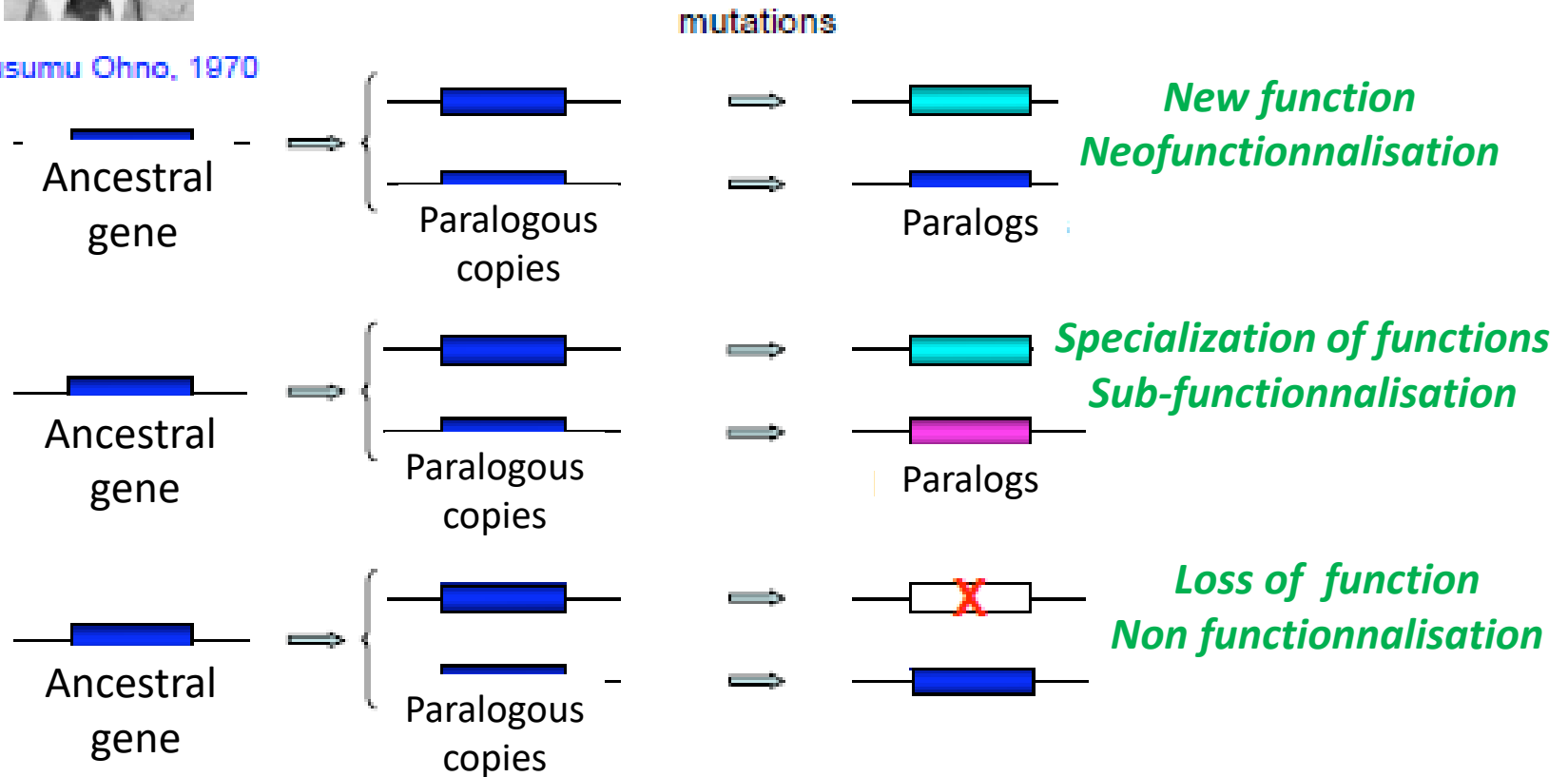
- ❖ Genome duplications, segmental duplications: important mechanism for generating new genes.





Duplication of genes and their consequences

Susumu Ohno, 1970



❖ Different levels of duplication in genomes (1)

Whole genome (polyploidy)



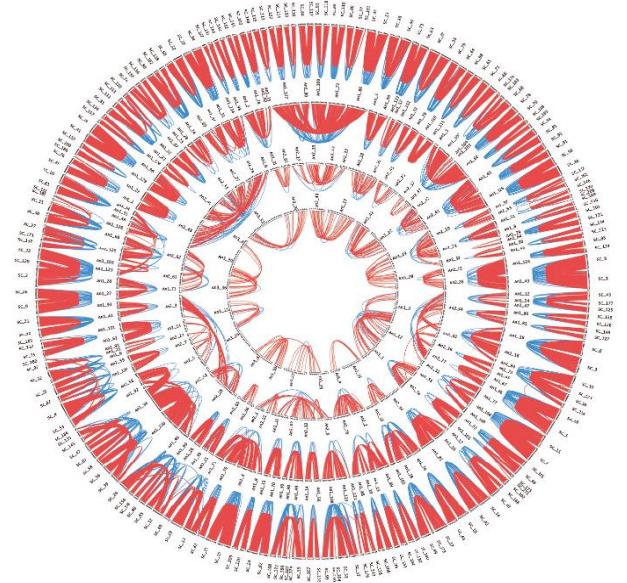
Xenopus laevis: 36 chr diploids

Xenopus vestitus: 72 chr. tetraploids

Xenopus ruwenzoriensis: 108 chr. hexaploids

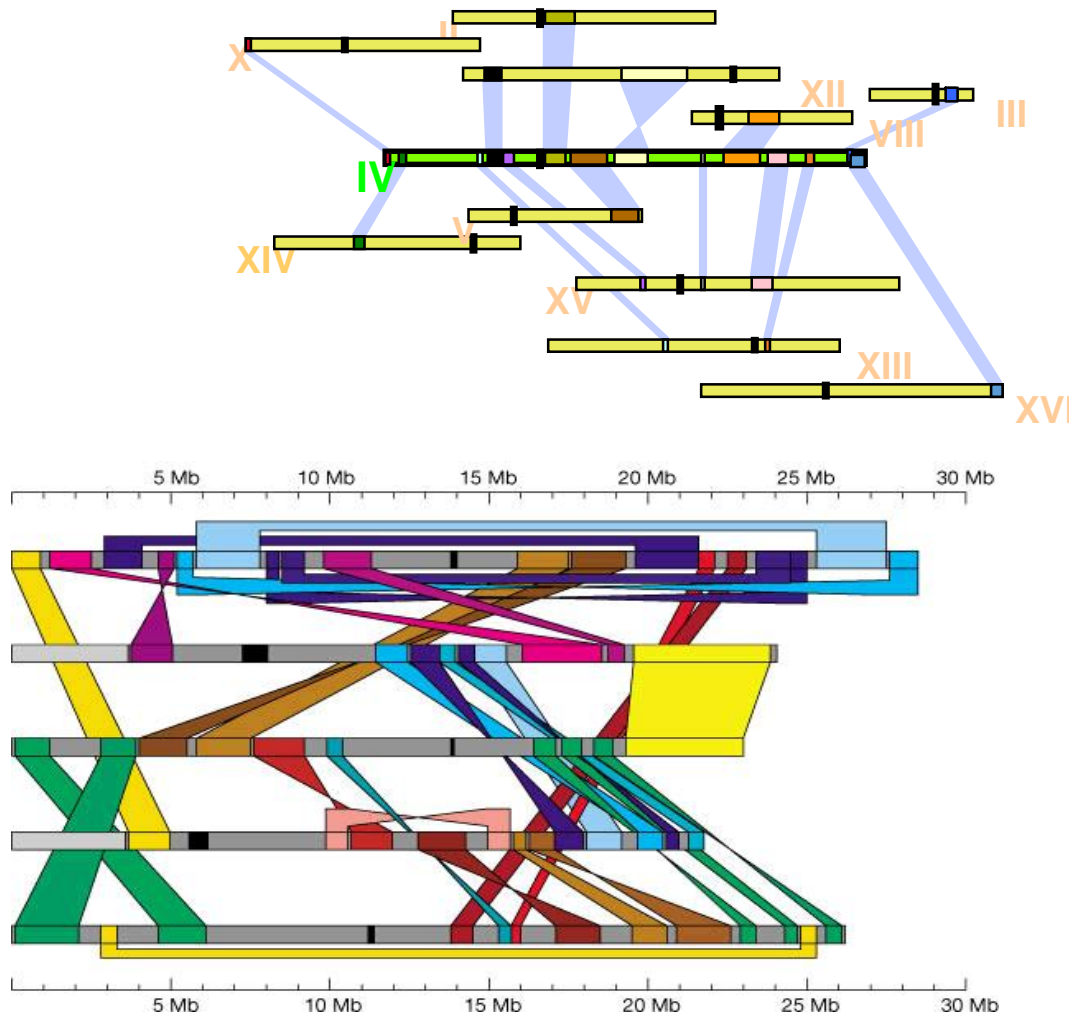
Whole genome duplication

Ex.: Le génome de la paramécie (49 000 gènes) révèle les traces de trois évènements successifs de duplication (ancêtre ~ 6 000 gènes)



❖ Different levels of duplication in genomes (2)

Segments of genome



Saccharomyces cerevisiae

- 12 Mb
- 16 chromosomes
- 5 800 genes.

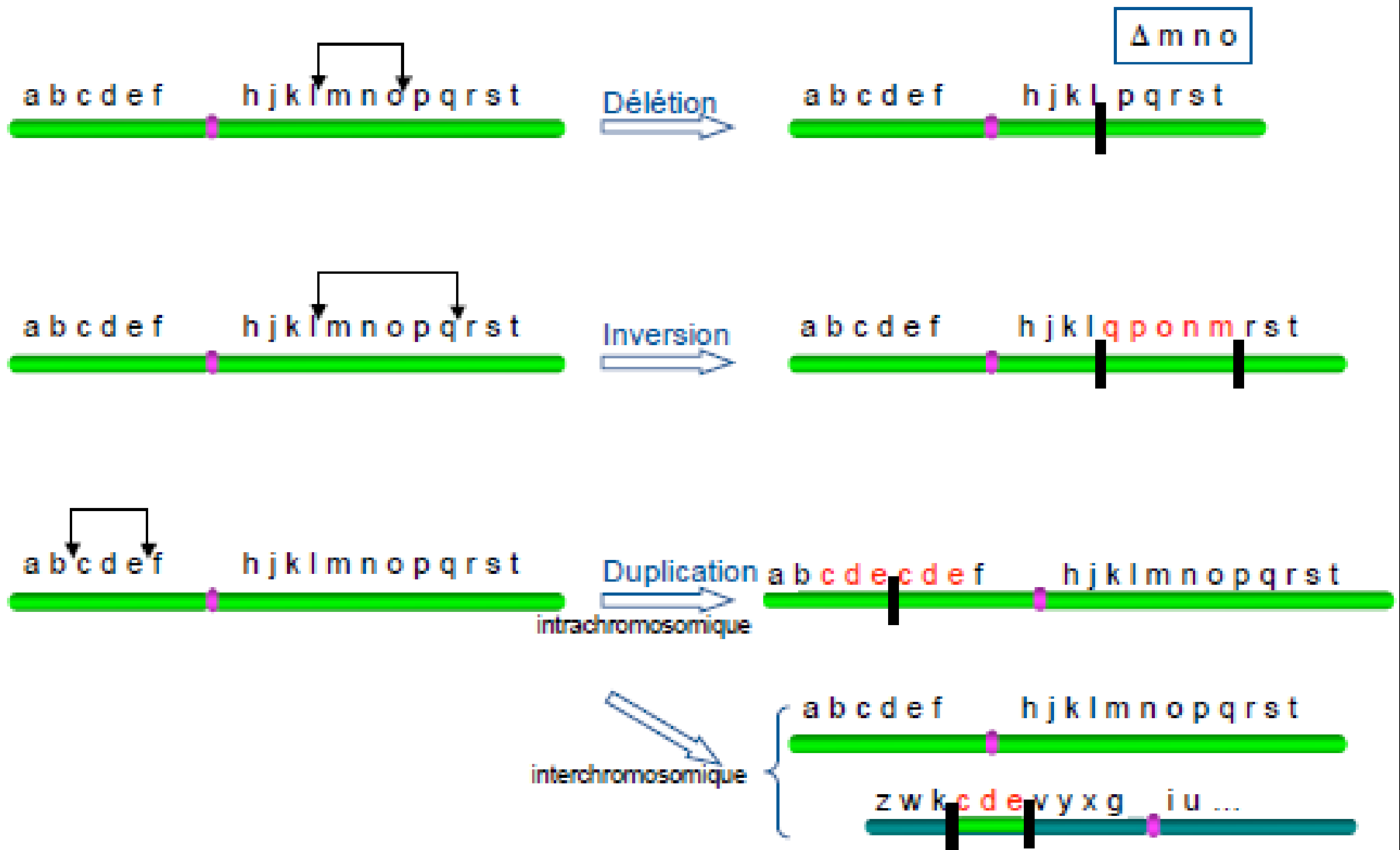
-> 50% of the genome

- 125 Mb
- 5 chromosomes
- 28 000 genes.

-> + 50% of the genome

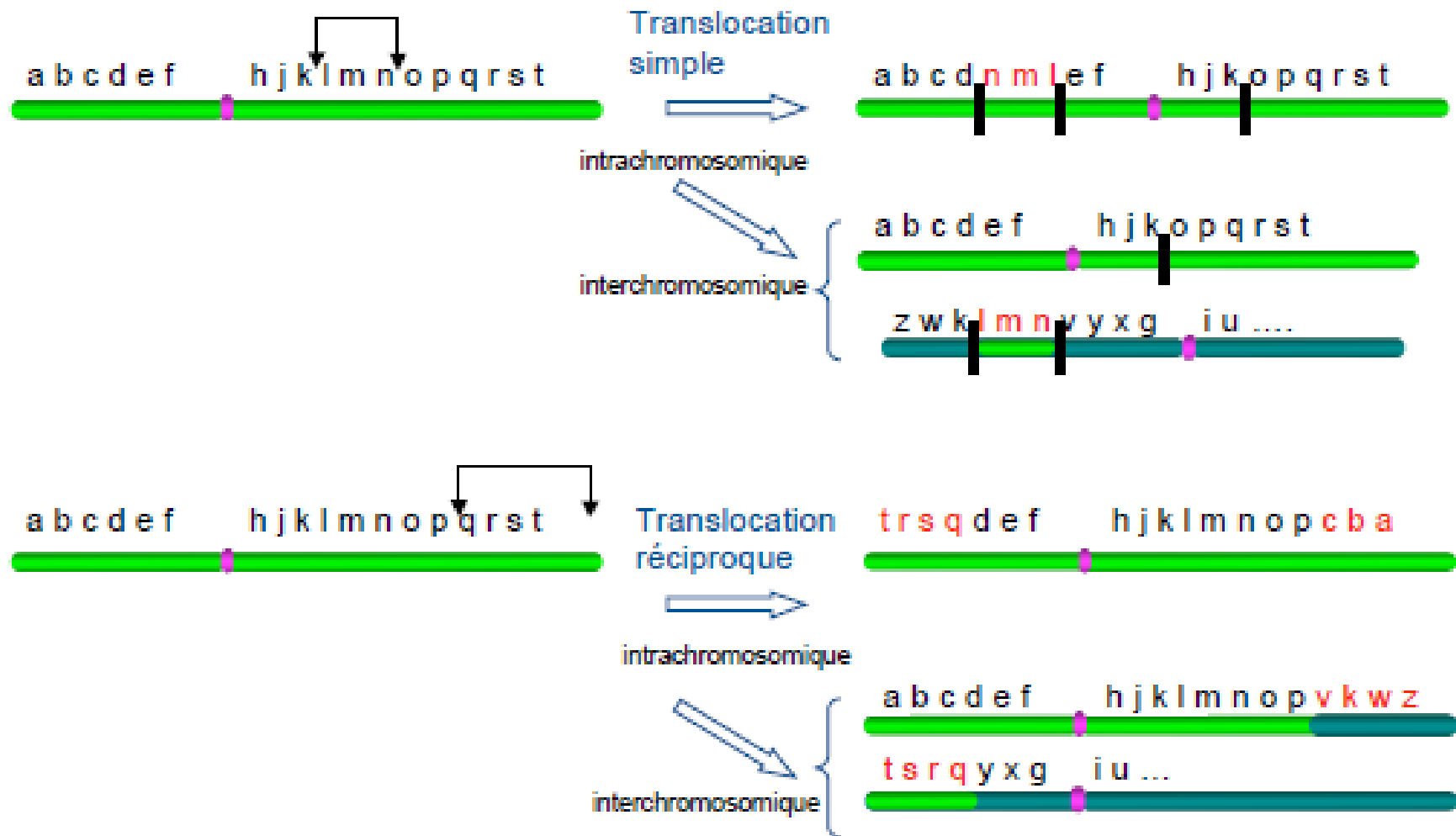
The human genome shows hundreds of duplicated segments (> 5kb) ---> 150kb (5% of genome)

Chromosomal rearrangements



Nouvelles jonctions, possibilité d'altération ou de fusion de gènes

Chromosomal rearrangements

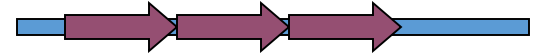


❖ Different levels of duplication in genomes (3)

Satellites (mini, micro, alu)



Tandem multicopie repetition at the centromere
-170 pb



-> 7% of the *Cercopithecus aethiops* genome

Transposons



Zea Mays

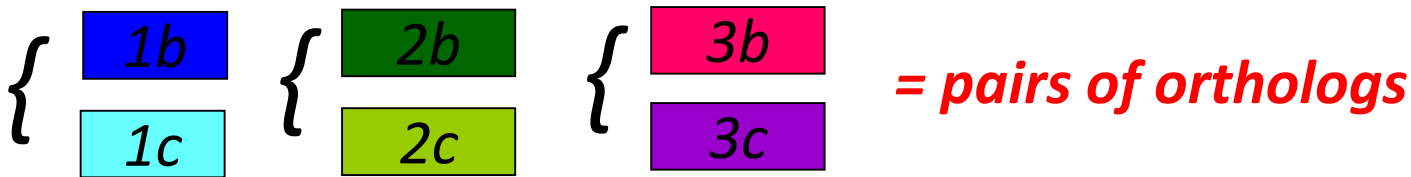
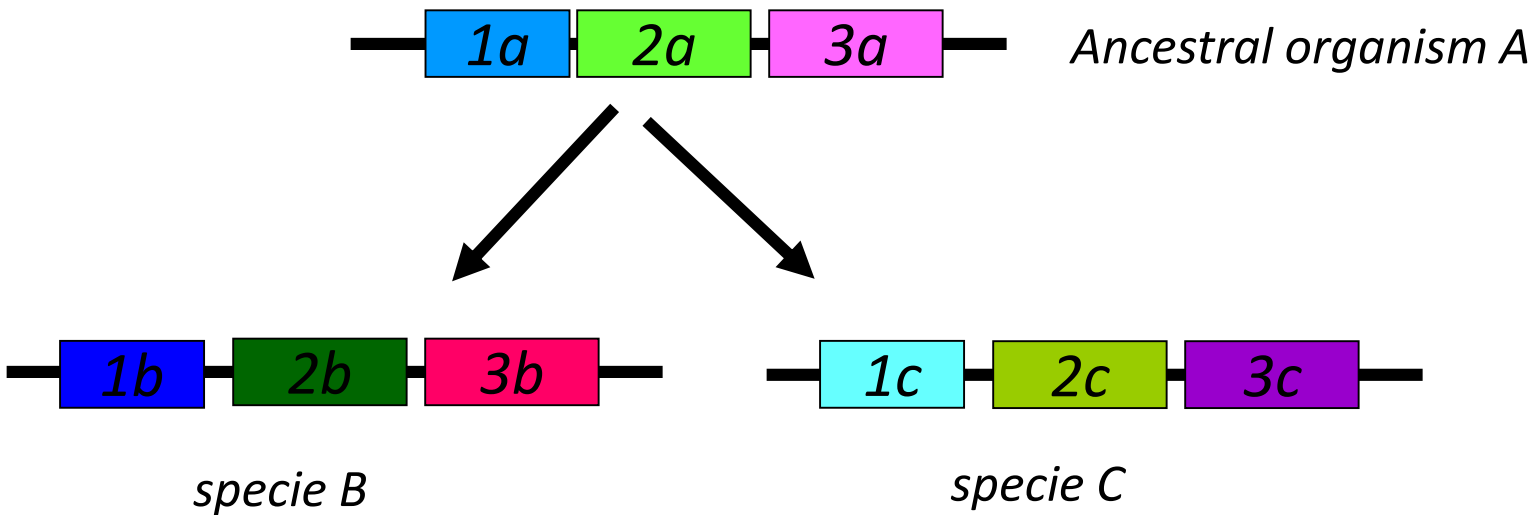
2,4 Gb

10 pairs of chr

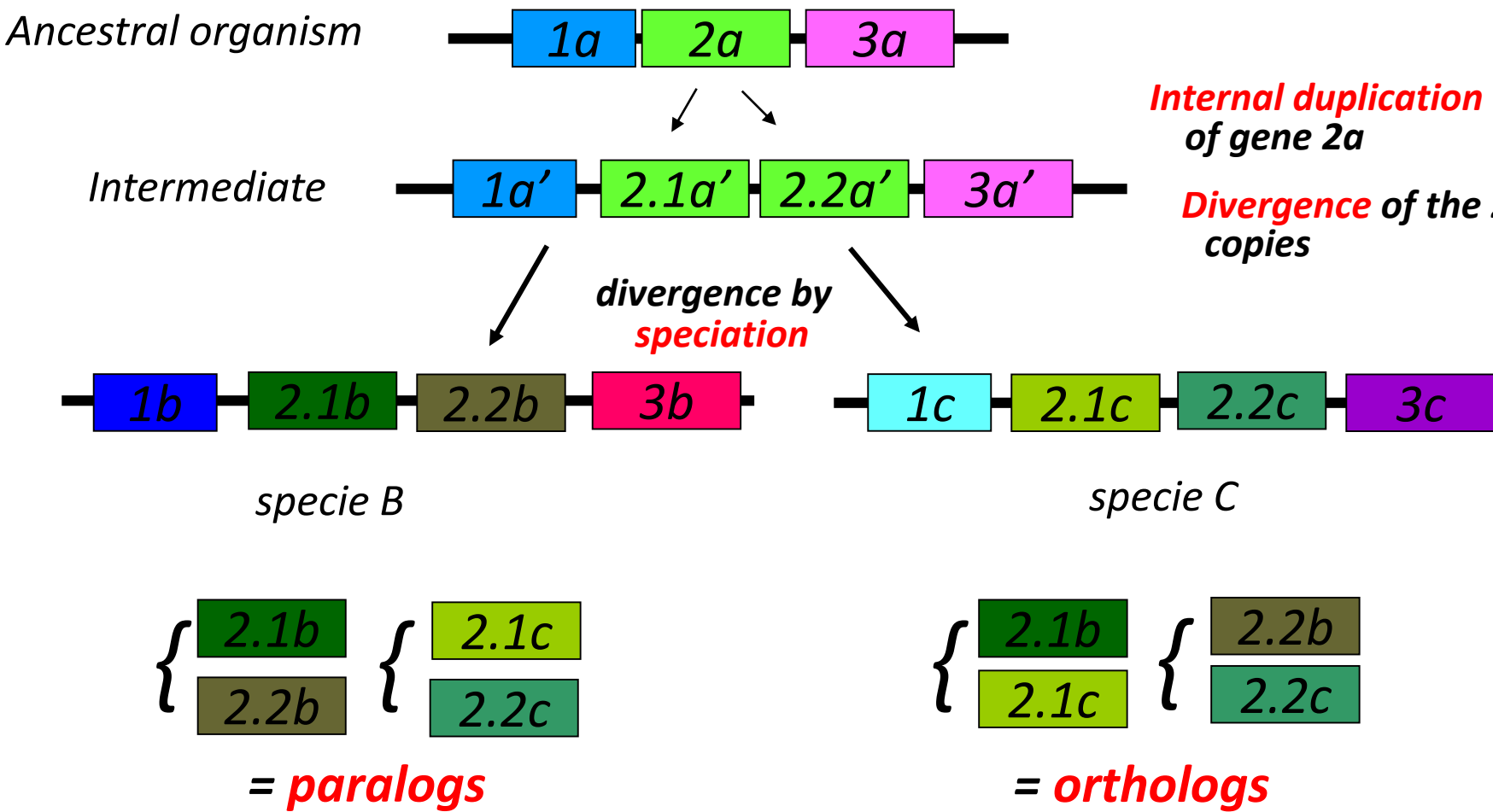
> 60% of transposons

Orthologs / Paralogs

LUCA = Last Common Universal Ancestor



Orthologs / Paralogs



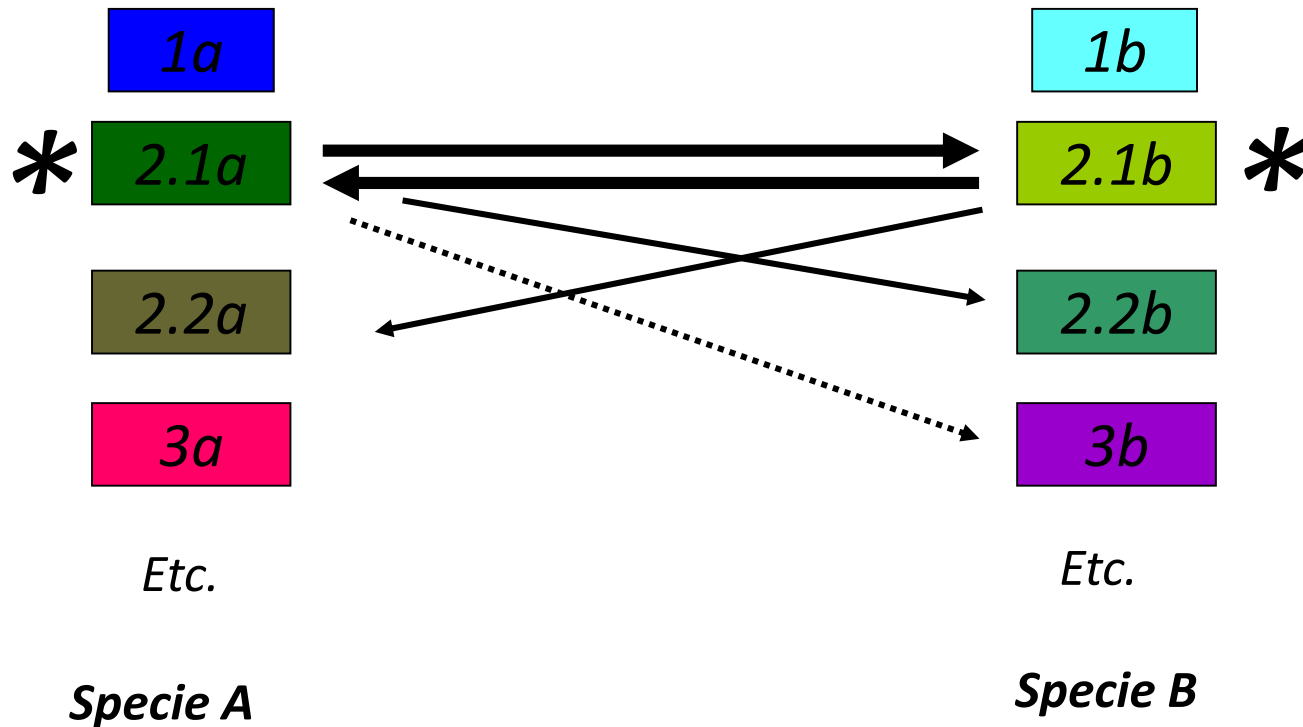
Speciation: the evolutionary process by which new living species appear

Duplication: the process of doubling genetic material on a chromosome, a large chromosomal portion, or an entire genome.

Orthologs / Paralogs

❖ How orthologs can be detected?

❖ Best way : best reciprocal hit



The genome *is only a snapshot* of the continuous processes of duplication, loss of genes and movements of transposable elements over successive generations.