Organisation physique et génétique des génomes

Table des matières

- Organisation physique de l'ADN
  - Nucléosomes
  - Chromatine
- Principes d'organisation des génomes
  - Exemple d'un génome bactérien
  - Exemple du génome humain
- Evolution des génomes: duplication et recombinaison

Chromosomes

- Structures organisées qui contiennent l'ADN

  Les molécules d'AND sont arrangées avec des protéines appelées histones de manière à tenir dans un espace très réduit. L'ensemble ADN + histones est appelé chromatine

- On peut avoir plusieurs chromosomes dans le noyau

  Ils sont organisés en paires de copies parentales (diploïde) ou seuls (haploïde) ou encore en nombre pair de copies quelconque (polyploïde)

La structure du chromosome eucaryote

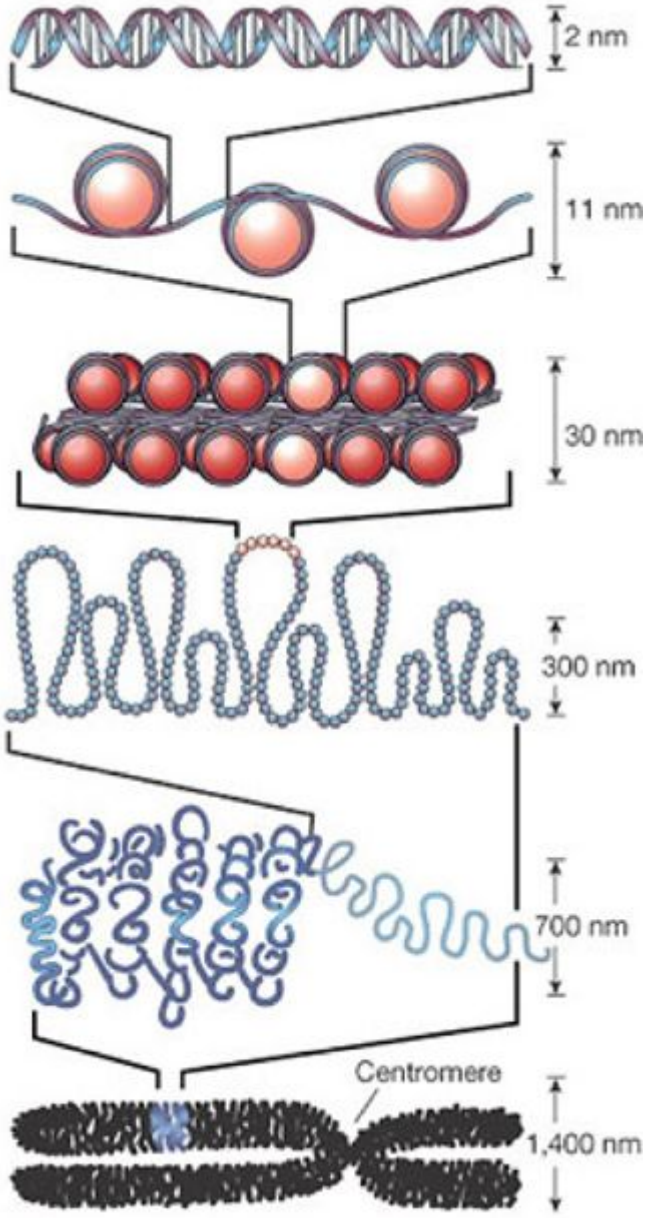
Différents niveaux de compaction

Chromatine =

ADN

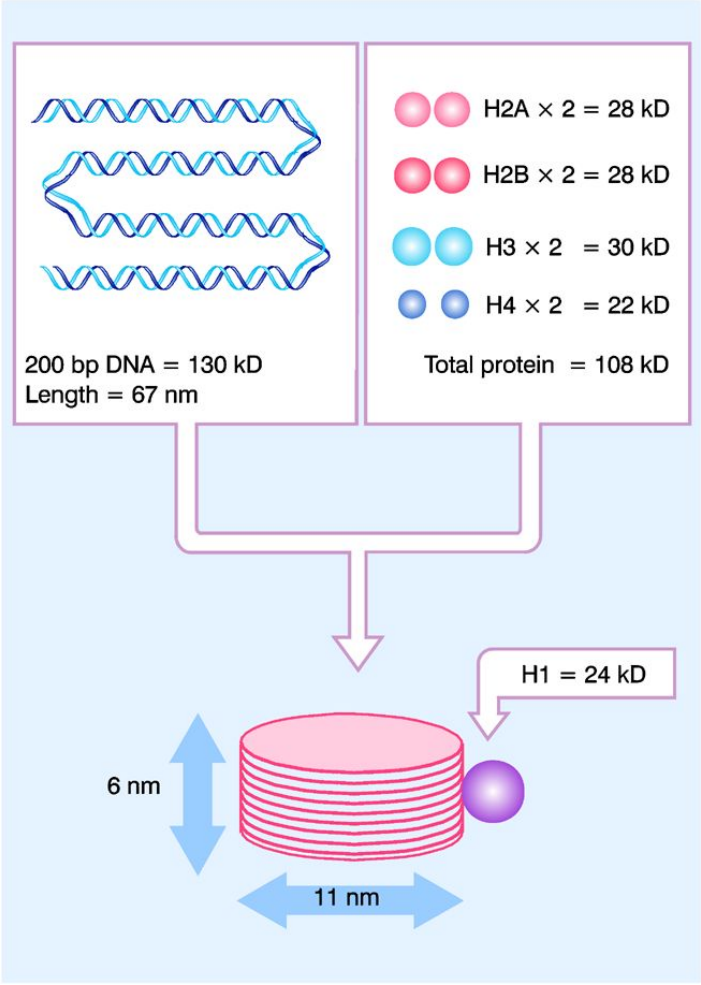
+ protéines

(histones + non-histones)



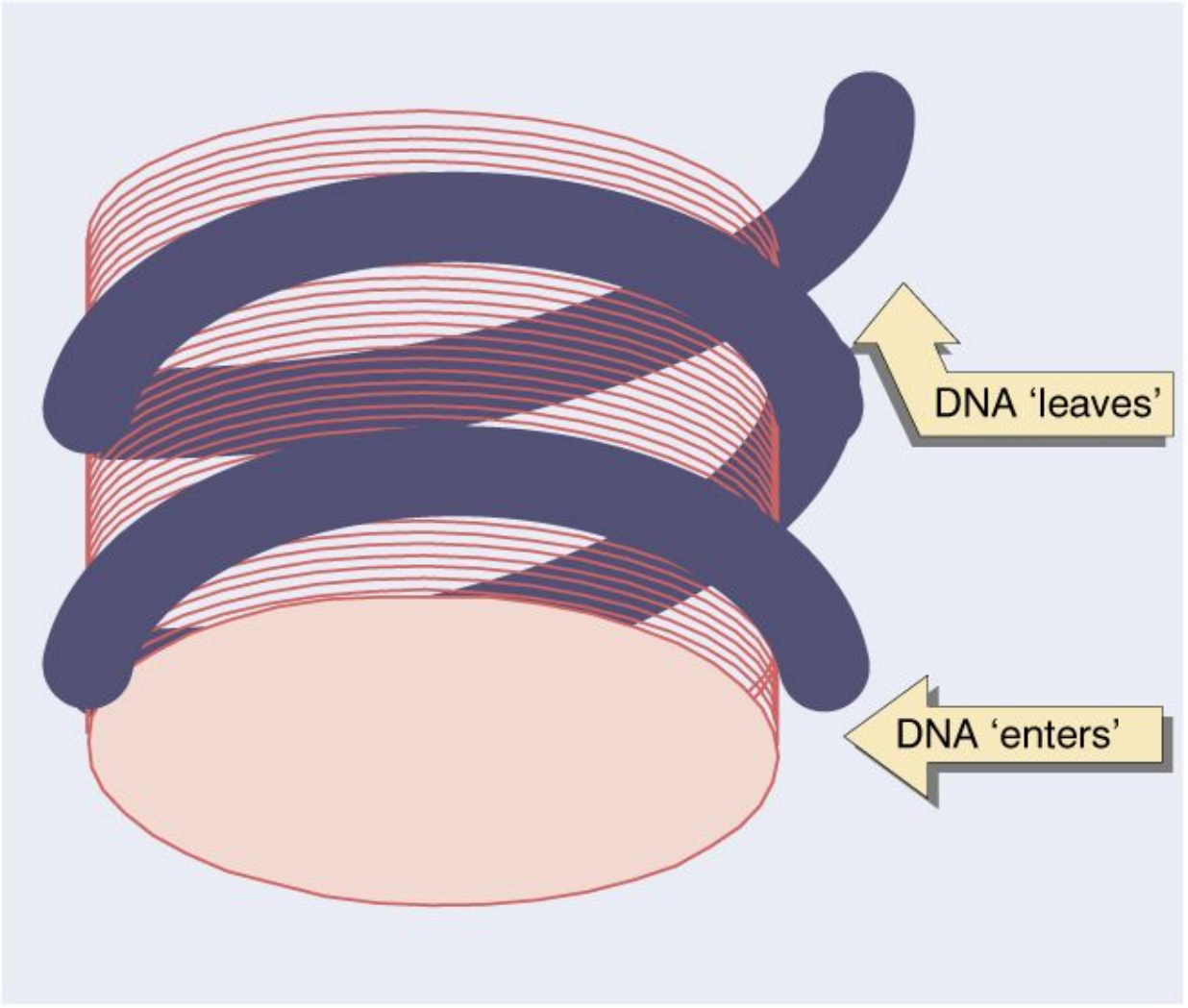
Nucléosomes

Figure 19.3 The nucleosome consists of approximately equal masses of DNA and histones (including H1). The predicted mass of the nucleosome is 262 kD.



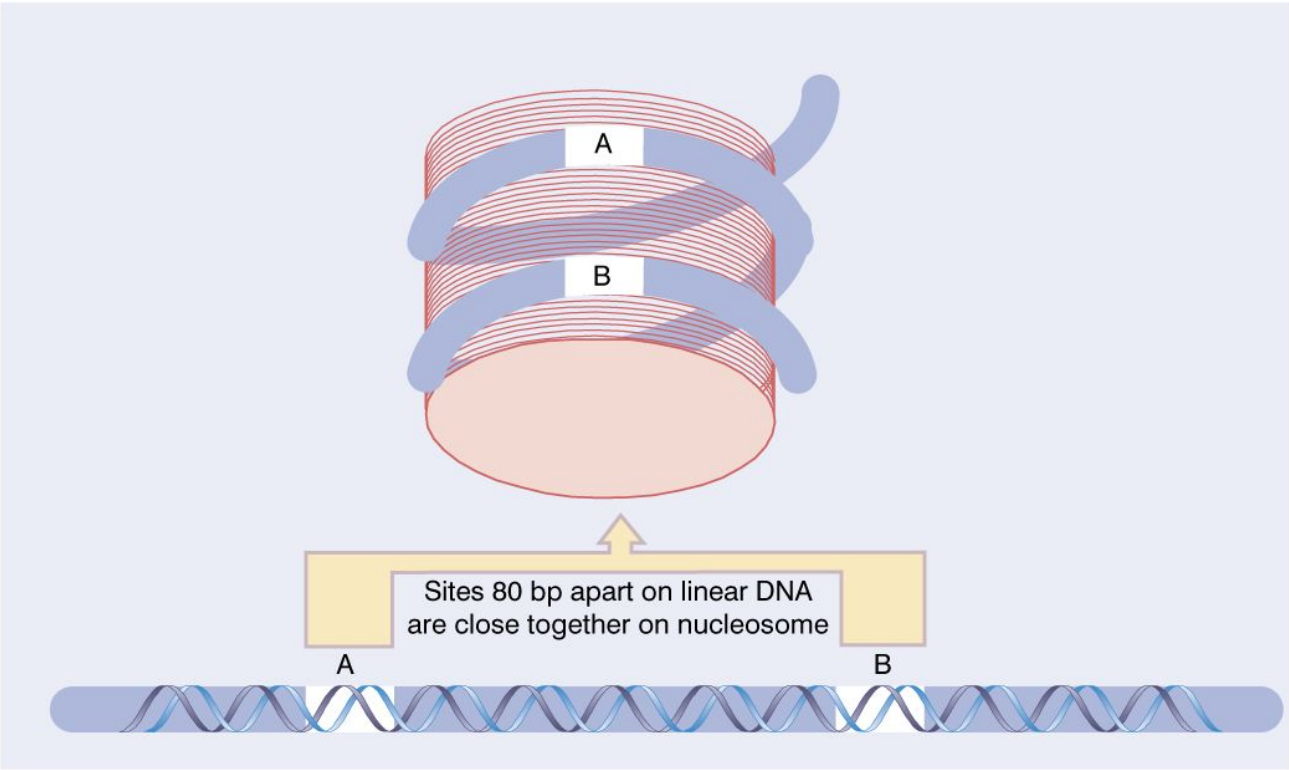
Nucléosomes

Figure 19.4 The nucleosome may be a cylinder with DNA organized into two turns around the surface.



Nucléosomes et organisation de l'ADN

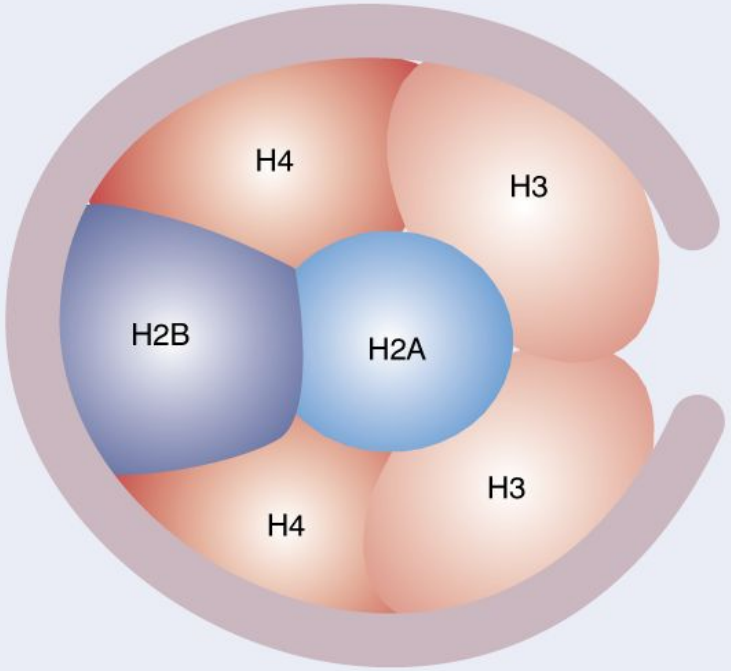
Figure 19.6
Sequences on the DNA that lie on different turns around the nucleosome may be close together.



Nucléosomes

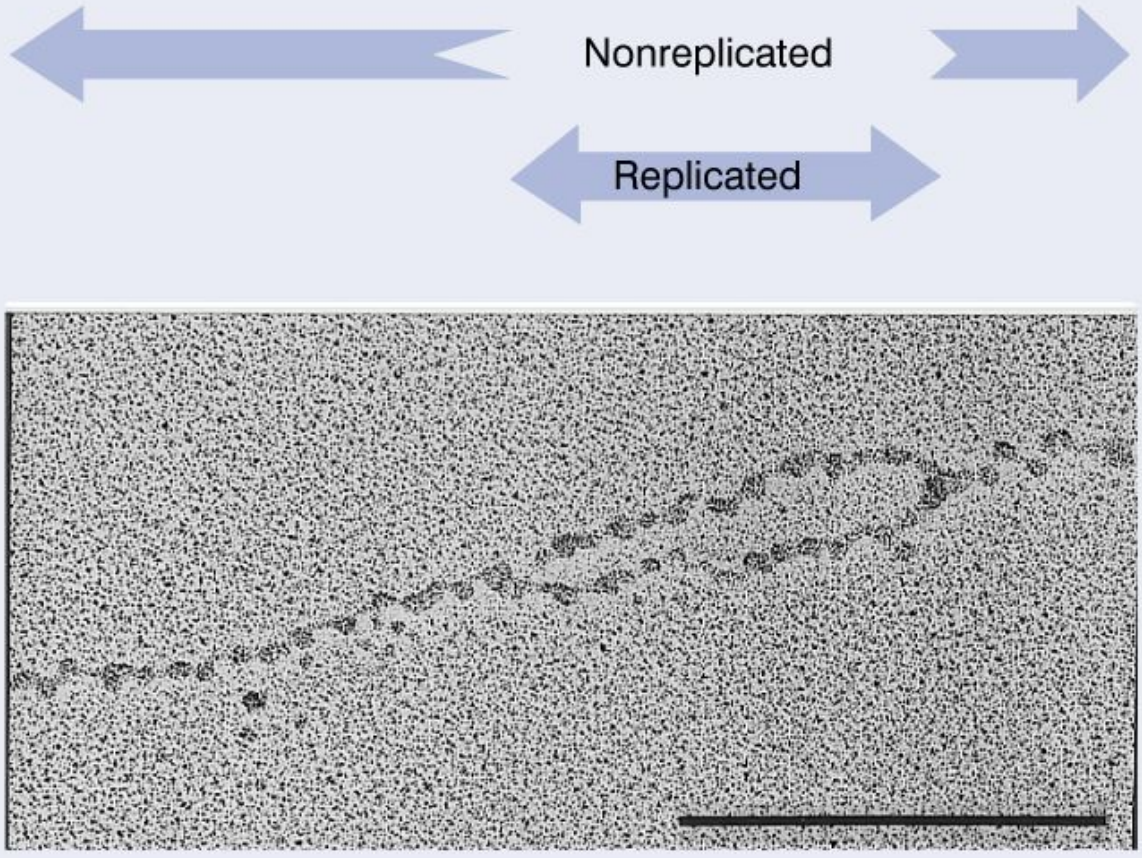
Figure 19.18 The 10 nm fiber is a continuous string of nucleosomes.

Figure 19.21 In a symmetrical model for the nucleosome, the $H3_2$-$H4_2$ tetramer provides a kernel for the shape. One H2A-H2B dimer can be seen in the top view; the other is underneath.



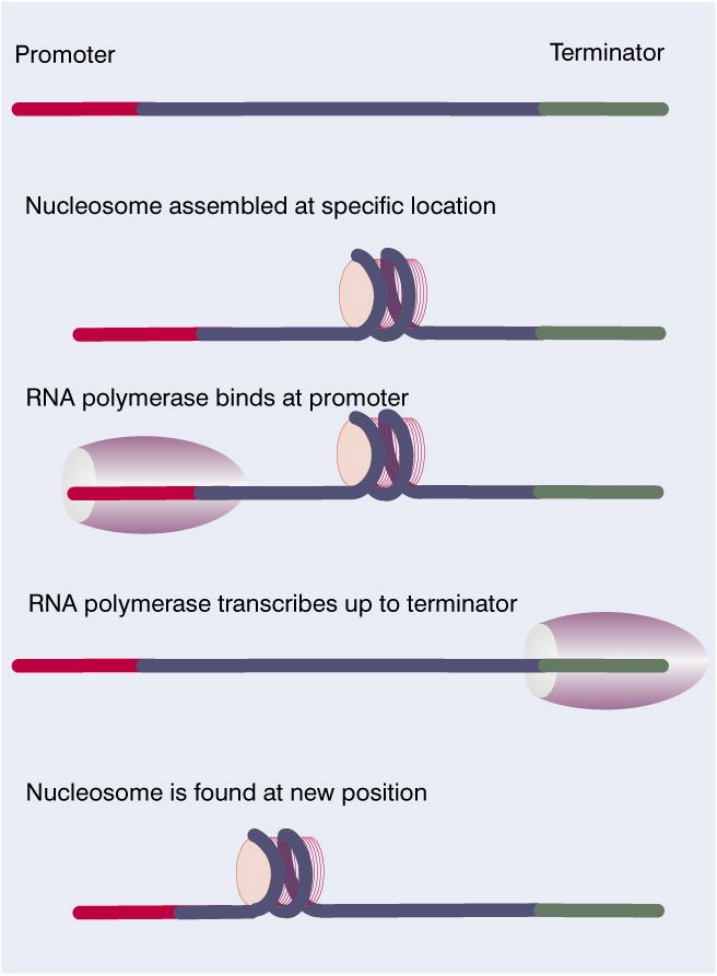
Nucléosomes et réplication

Figure 19.26 Replicated DNA is immediately incorporated into nucleosomes. Photograph kindly provided by S. MacKnight.



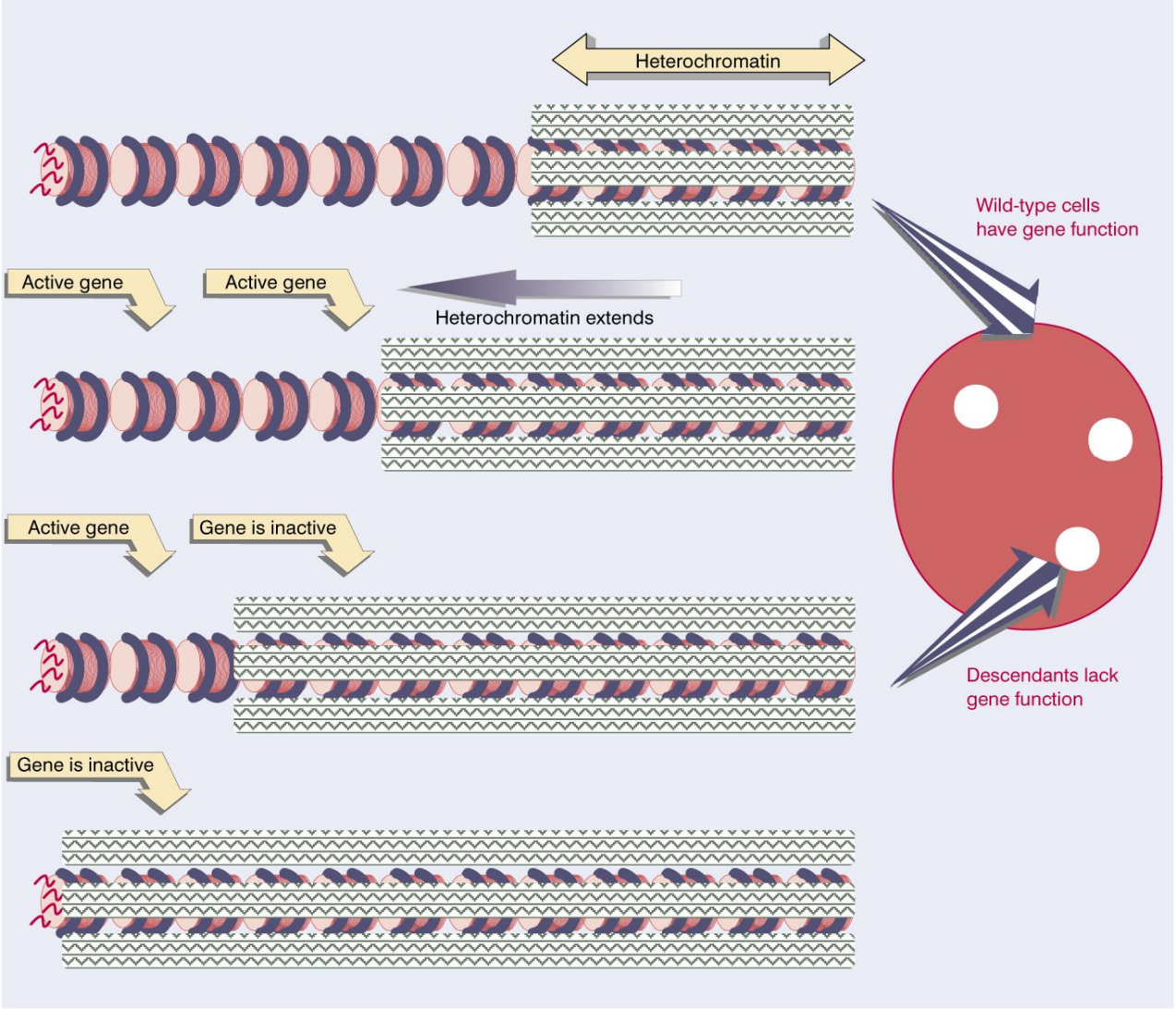
Nucléosomes et transcription

Figure 19.36 A protocol to test the effect of transcription on nucleosomes shows that the histone octamer is displaced from DNA and rebinds at a new position.



Chromatine et inactivation génique

Figure 19.45 Extension of heterochromatin inactivates genes. The probability that a gene will be inactivated depends on its distance from the heterochromatin region.



Chromatine et transcription

Figure 21.16 The pre-emptive model for transcription of chromatin proposes that if nucleosomes form at a promoter, transcription factors (and RNA polymerase) cannot bind. If transcription factors (and RNA polymerase) bind to the promoter to establish a stable complex for initiation, histones are excluded.

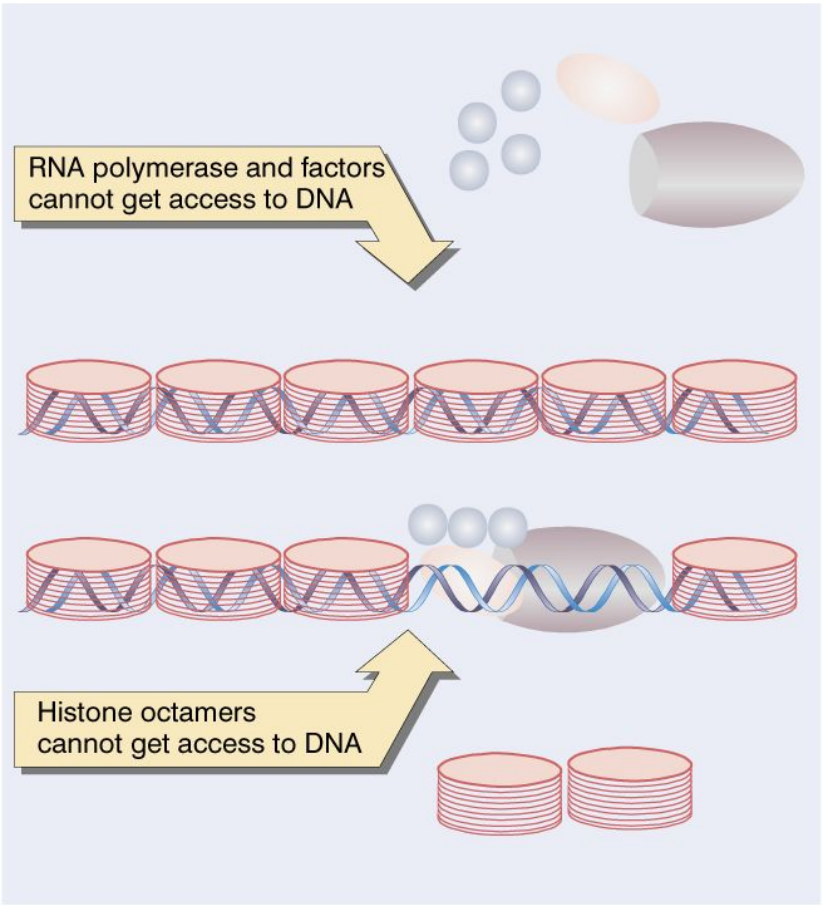
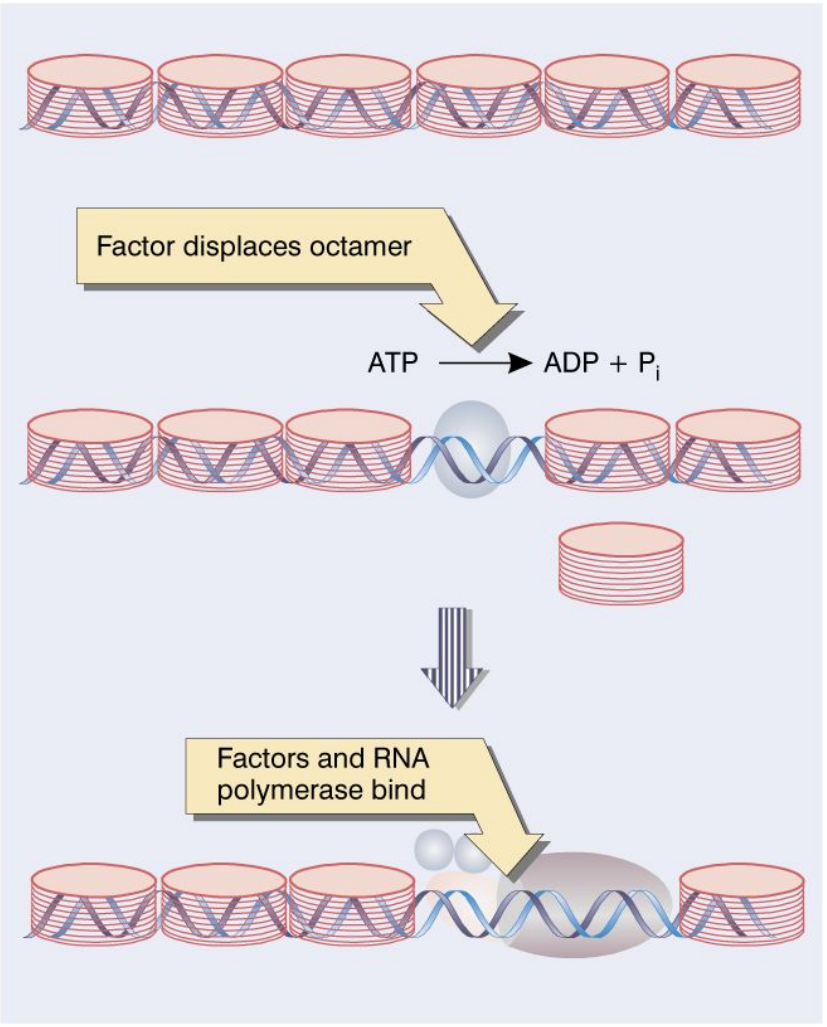


Figure 21.17 The dynamic model for transcription of chromatin relies upon factors that can use energy provided by hydrolysis of ATP to displace nucleosomes from specific DNA sequences.



Organisation génétique des génomes

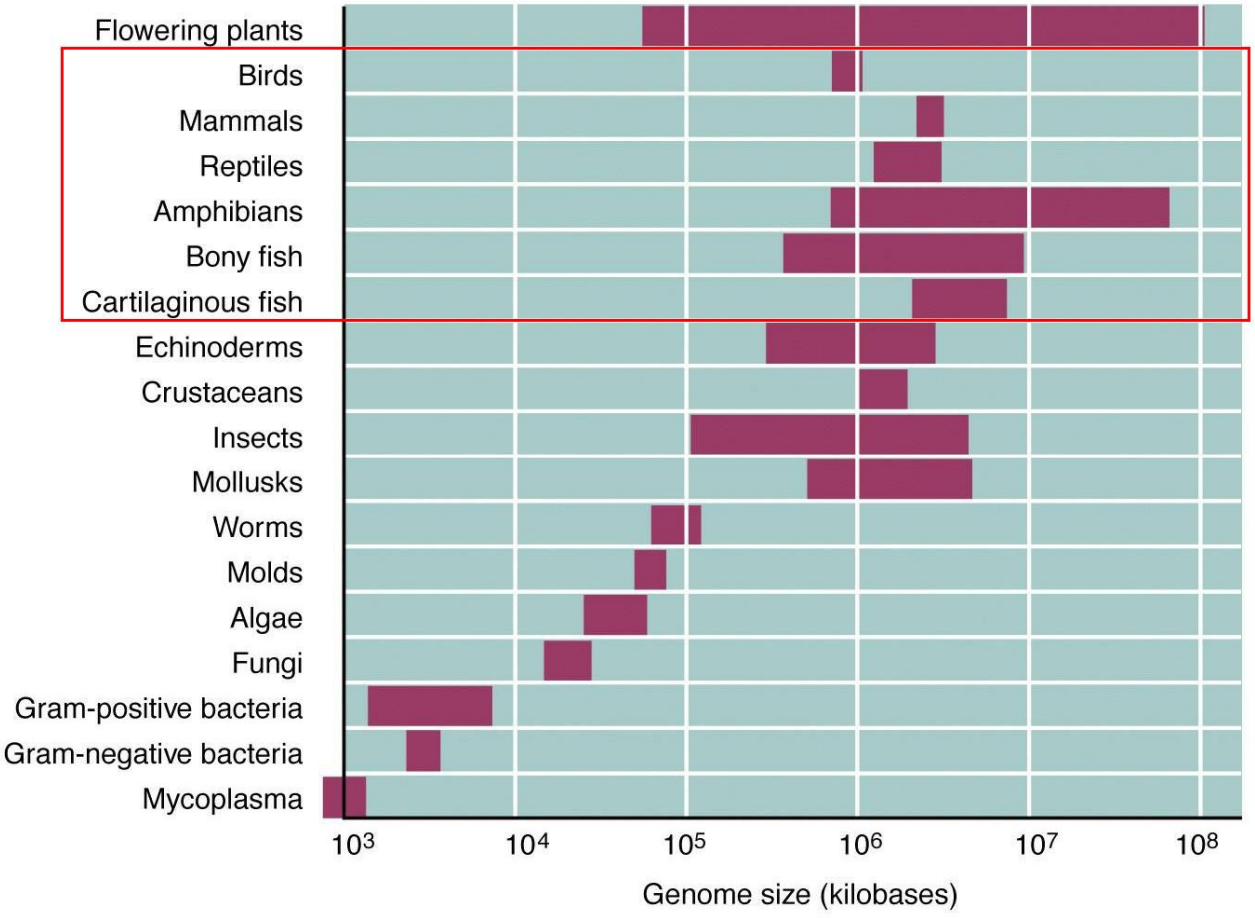
Taille des génomes

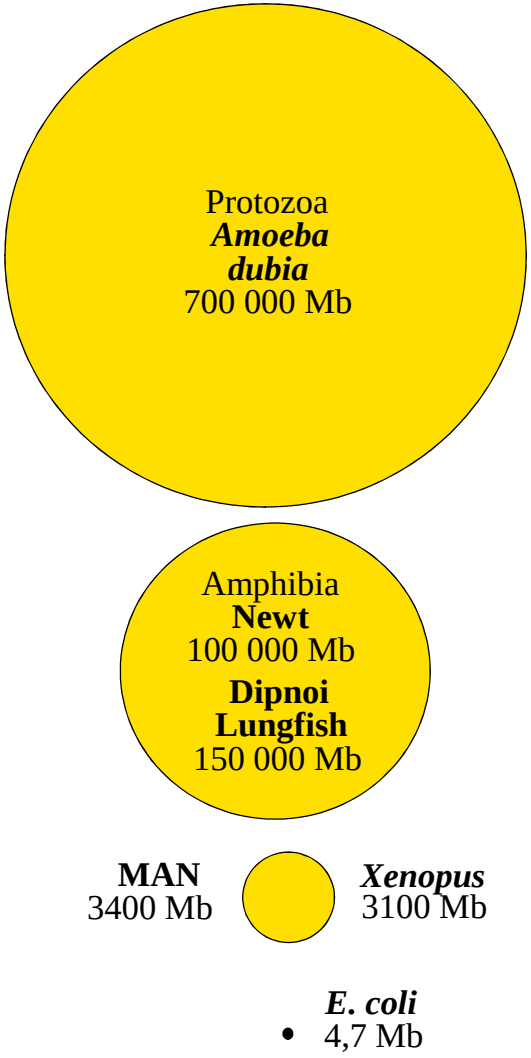
- La longueur du génome et le nombre de gènes sont corrélés chez les procaryotes, pas chez les eucaryotes
- Le pourcentage de séquence codantes (densité) est très variable en fonction des organismes

Figure 1.36 The amount of nucleic acid in the genome varies over an enormous range.

Genome	Gene Number	Base Pairs
Organisms		
Plants	<50,000	$<10^{11}$
Mammals	100,000	$\sim 3 \times 10^8$
Worms	14,000	$\sim 10^8$
Flies	12,000	1.6×10^8
Fungi	6,000	1.3×10^7
Bacteria	2–4,000	$<10^7$
Mycoplasma	500	$<10^6$
dsDNA Viruses		
Vaccinia	<300	187,000
Papova (SV40)	~6	5,226
Phage T4	~200	165,000
ssDNA Viruses		
Parvovirus	5	5,000
Phage ϕX174	11	5,387
dsRNA Viruses		
Reovirus	22	23,000
ssRNA Viruses		
Coronavirus	7	20,000
Influenza	12	13,500
TMV	4	6,400
Phage MS2	4	3,569
STNV	1	1,300
Viroids		
PSTV RNA	0	359
Scrapie		
Prion	?	?

Taille des génomes et complexité





Gènes essentiels

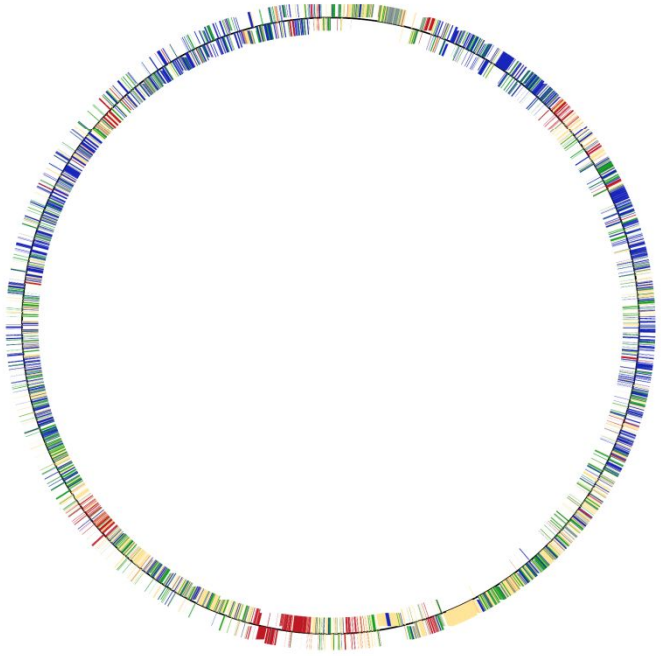
- Un gène dont la délétion est létale est dit essentiel
- La recherche du sous ensemble minimal de gènes pour parvenir à faire vivre un organisme vivant est toujours d'actualité (environ 500 actuellement)

Figure 3.10 Genome sizes, gene numbers and lethal loci.

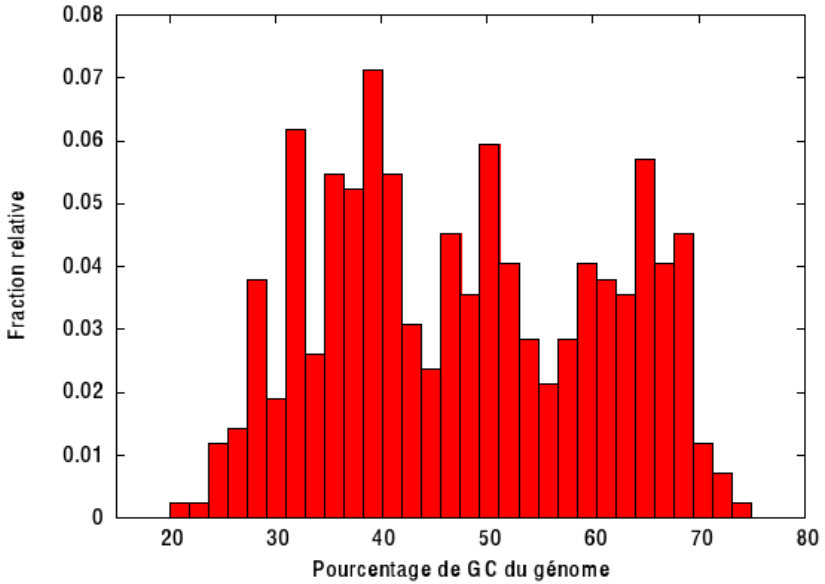
Species	Genome (Mb)	Genes	Lethal loci
Mycoplasma genitalium	0.58	470	
Rickettsia prowazekii	1.11	834	
Haemophilus influenzae	1.83	1,743	
Methanococcus jannaschi	1.66	1,738	
B. subtilis	4.2	4,100	
E. coli	4.6	4,288	1,800
S. cerevisiae	13.5	6,034	3,600
D. melanogaster	165	12,000	3,100
C. elegans	97	19,099	
H. sapiens	3,300	100,000	

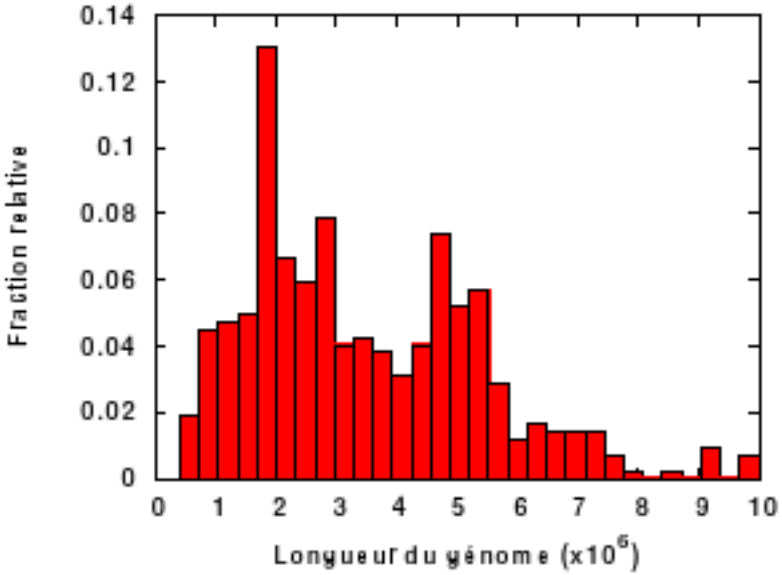
Un génome bacterien : *E.coli*

Genome Info:	Features:
Refseq: NC_000913	Genes: 4493
GenBank: U00096	Protein coding: 4149
Length: **4,639,675 nt**	Structural RNAs: 172
GC Content: **50%**	Pseudo genes: **177**
% Coding: **85%**	Others: **584**

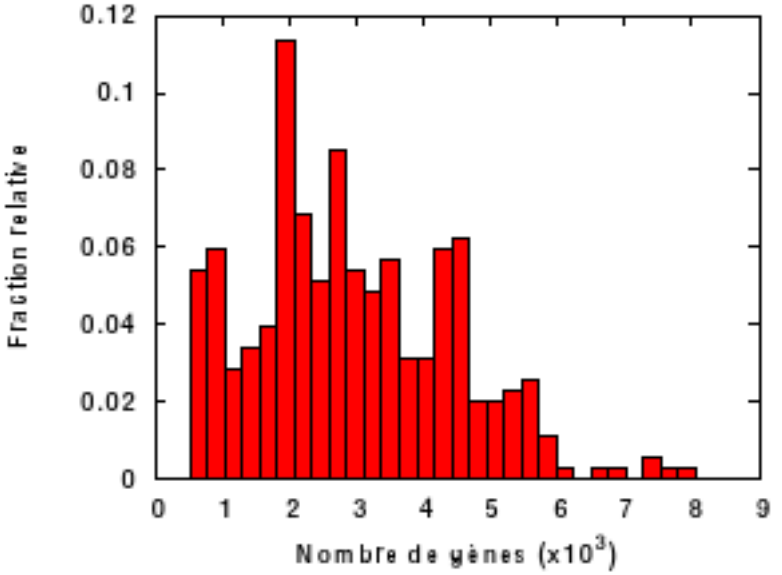


Diversité des génomes bactériens



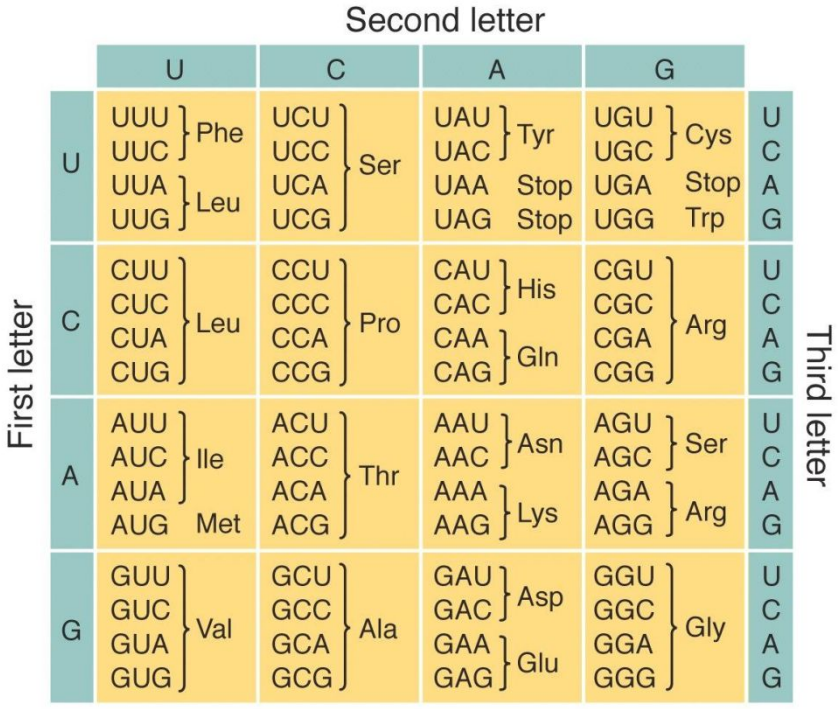


Le nombre de plasmides des génomes bactériens varie de 0 à plus de 20 (dans ce cas 36% du génome)



Les contraintes en GC se traduisent au niveau protéique

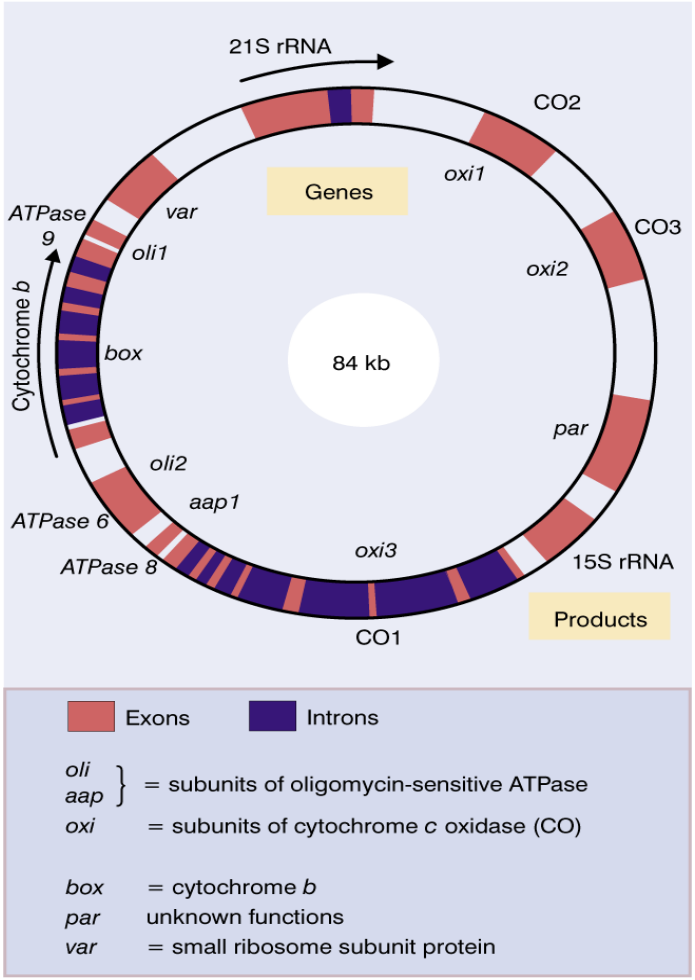
- Les génomes riches ou pauvres en GC ont des possibilités restreintes en termes de séquences protéiques
- Le niveau de GC dans les génomes dépend de facteurs évolutifs et environnementaux



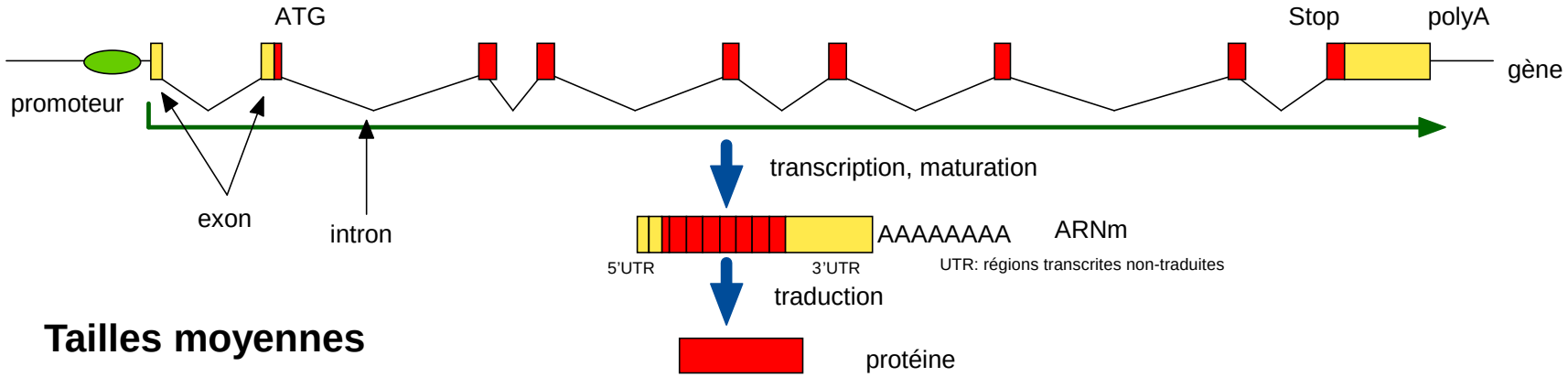
Le génome de la mitochondrie

- Le génome de la mitochondrie ressemble fortement à celui d'une petite bactérie
- Il n'est pas autosuffisant: besoin d'import de certains composés depuis la cellule
- Idem pour les chloroplastes

Figure 3.14 The mitochondrial genome of *S. cerevisiae* contains both interrupted and uninterrupted protein-coding genes, rRNA genes, and tRNA genes (positions not indicated). Arrows indicate direction of transcription.



Structure des gènes humains



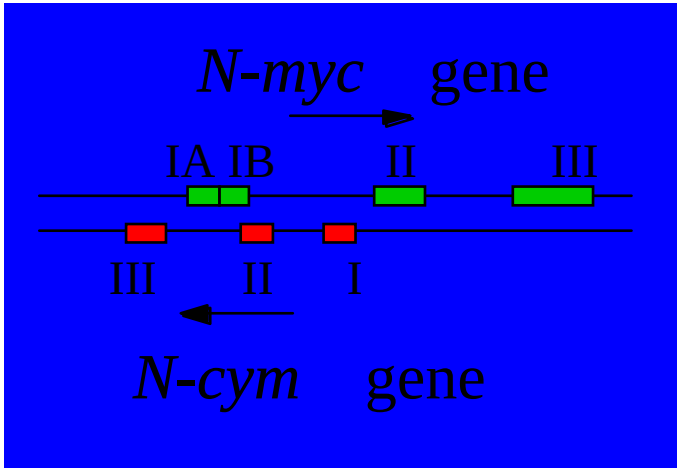
- **Tailles moyennes**
  - **Gene** **45 kb**
  - **CDS** **1500 nt**
  - **Exon (interne)** **145 nt**
  - **Intron** **5200 nt**
  - **5'UTR** **210 nt**
  - **3'UTR** **740 nt**

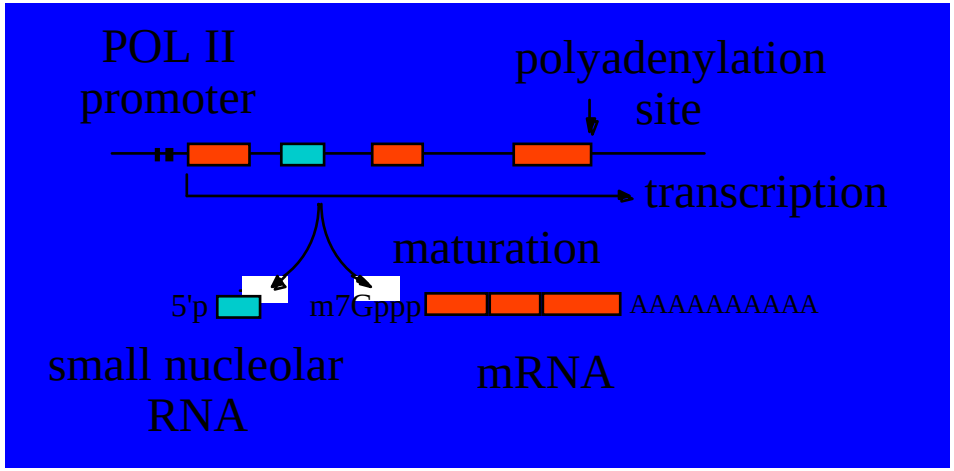
- **Intron / Exon**
  - **Nombres d'introns:** **6 ±3 introns / kb CDS**
  - **Introns / (introns + CDS):** **92%**

- **Epissage alternatif dans plus de 30% des gènes**

Chevauchement?

Gènes partiellement superposés



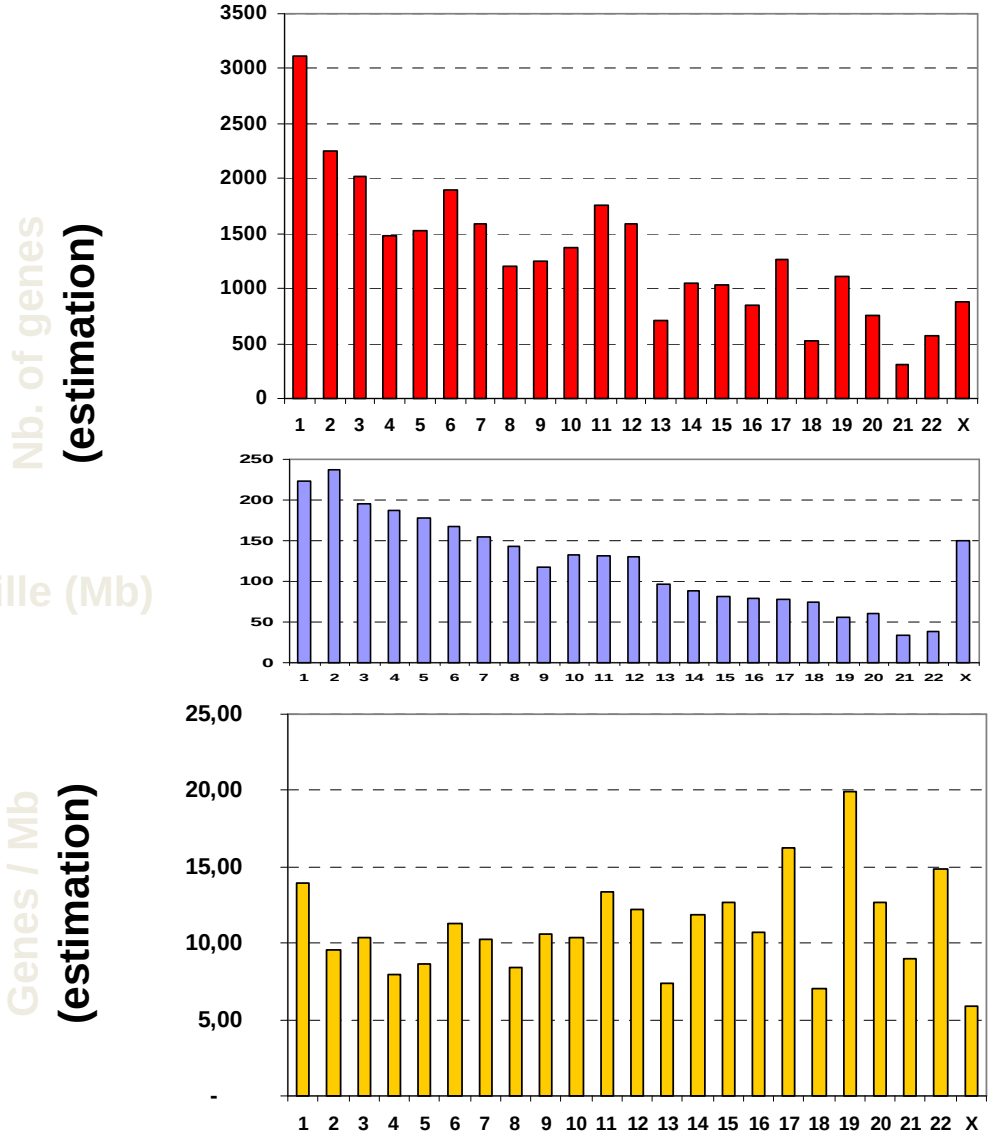


Petit ARN nucléolaire dans un intron de gène codant

Gènes non traduits

- **tRNA (70-100 nt) :** **~ 350**
- **rRNA :**
  - **18S (1.8kb), 5.8S (160 nt), 28S (5kb) :** **150-200**
  - **5S (120 nt):** **200-300**
- **snRNA (small nuclear RNA) (70-200 nt):** **> 100**
  Notamment épissage: U1, U2, U4, ...
- **snoRNA (small nucleolar RNA) (70-200 nt):** **> 100**
  Maturation des rRNA dans le nucléole
- **miRNA (micro RNA) :** **250 identifiés**
  Régulation traduction, stabilité mRNA, transcription
- **ARN interférents:** **??**

Localisation des gènes



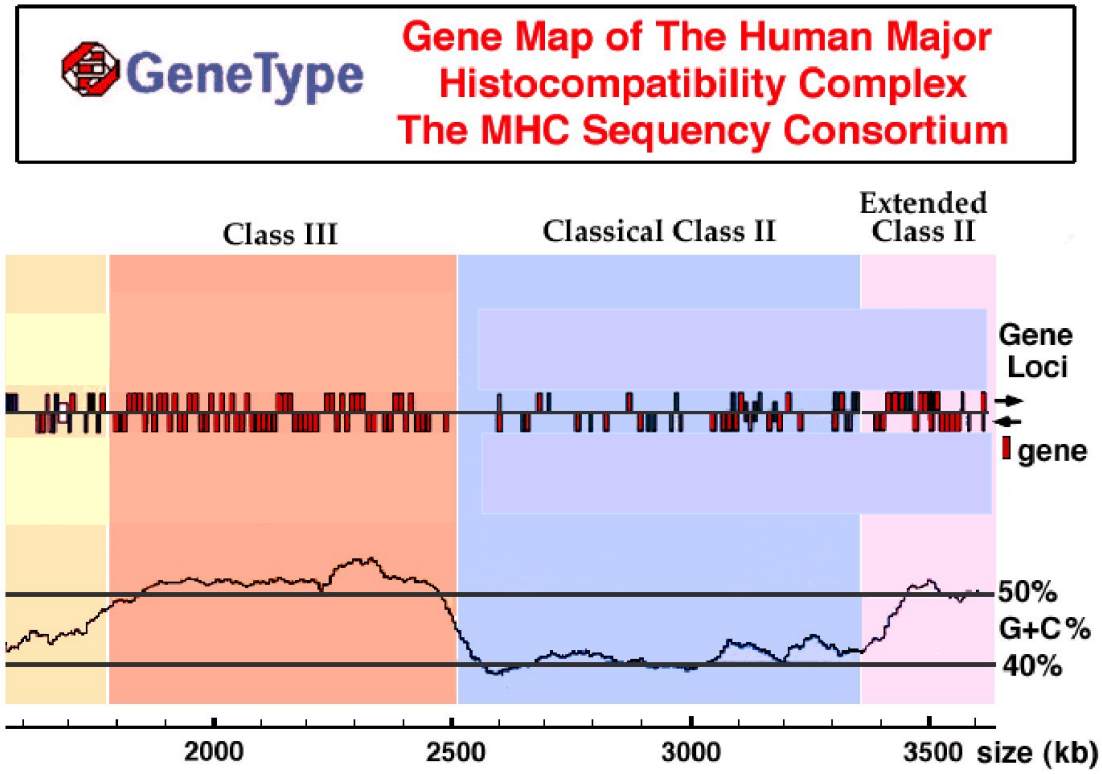
http://www.ncbi.nlm.nih.gov/Science96/

Isochores

Génome de Vertébrés

Variations de la composition en bases le long des chromosomes

Sequence of human MHC

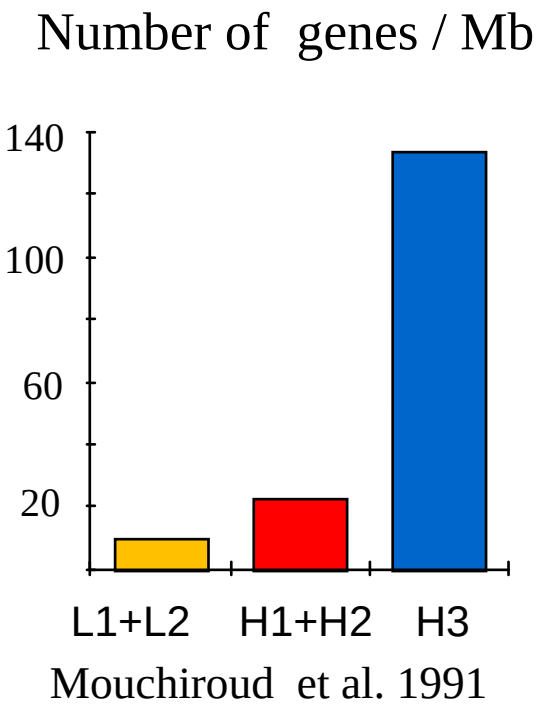


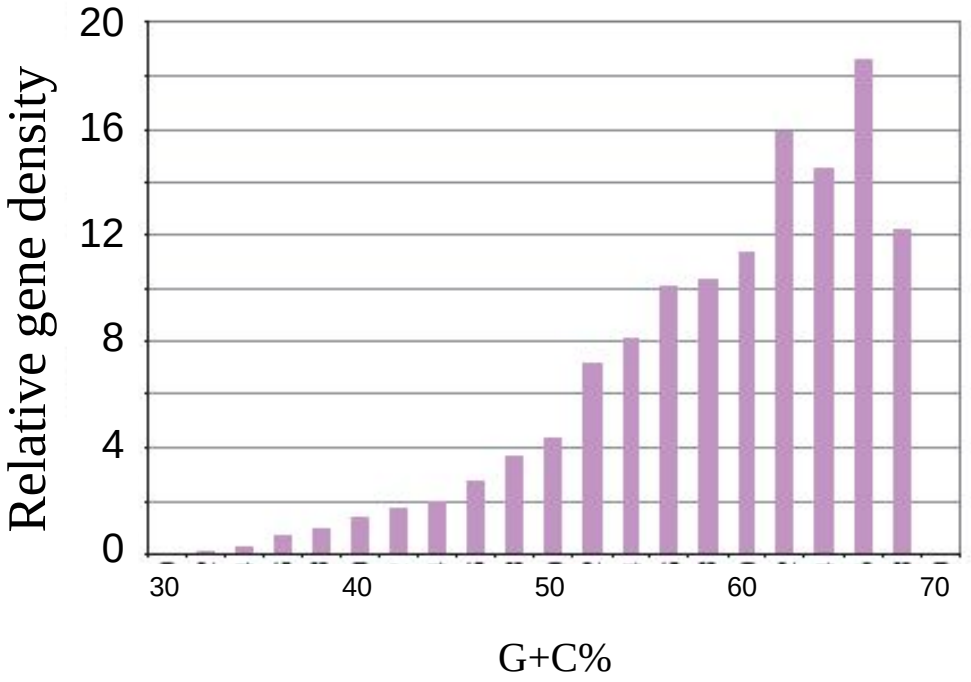
Définition

Régions > 300 kb homogène dans sa composition en bases

Isochores

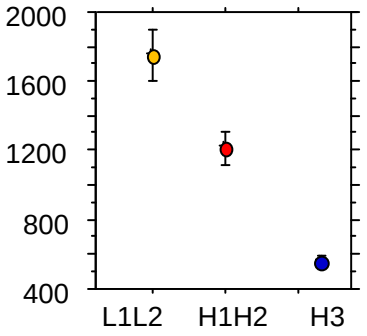
Corrélation isochores riches en G+C / densité génique



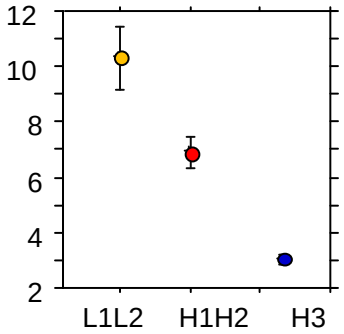


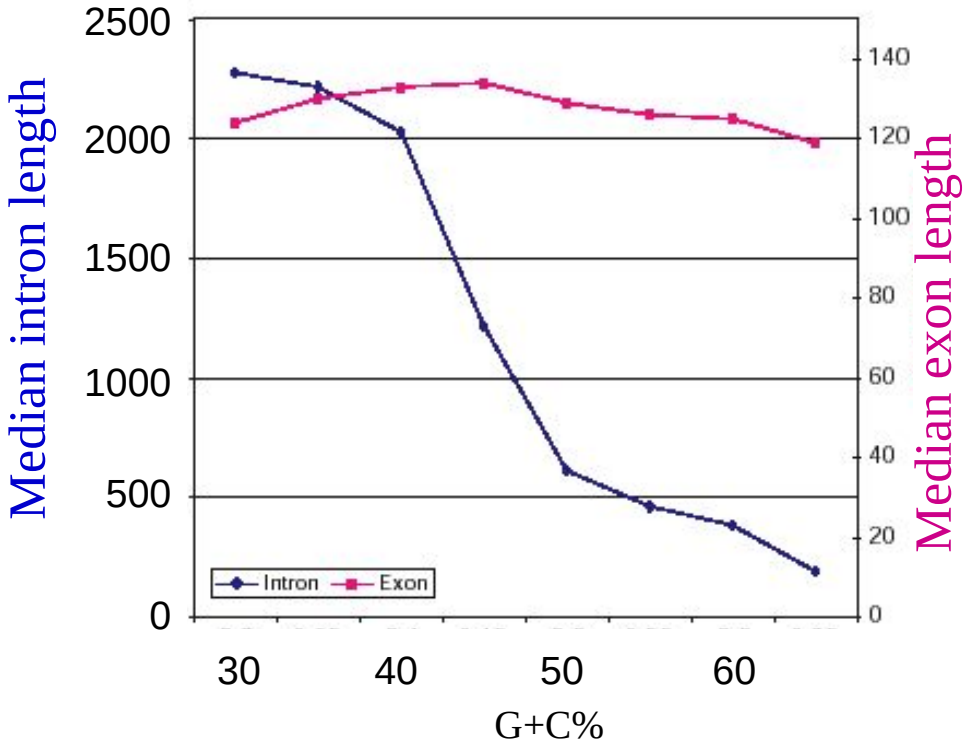
Isochores et introns

Average intron length (bp)



Gene compaction
(intron length/coding region length)





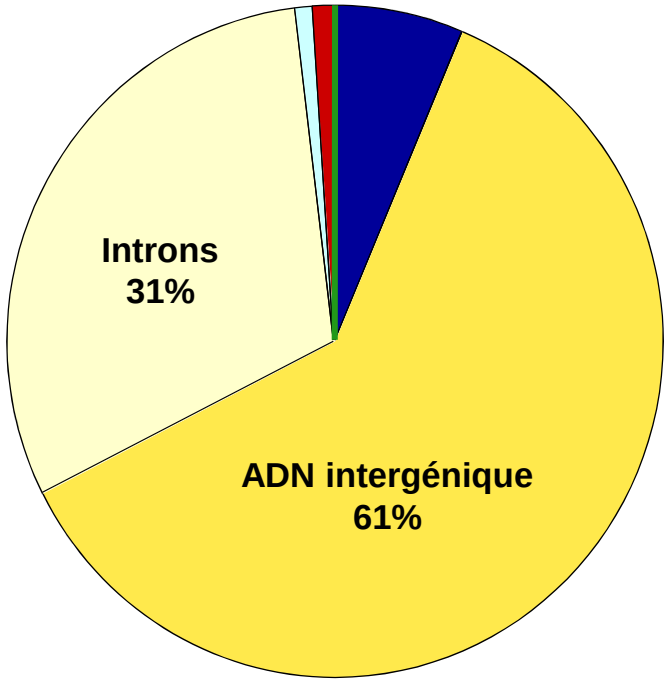
→ **Incidence sur la** structure génique

Et l'ADN non codant?

- **Genes, regulatory elements: ~ 2-5%**
  20 000 – 25 000 proteic genes
  tRNA, rRNA, snRNA, miRNA
  UTR

- **Non-coding sequences: ~ 95-98%**
  - **Satellite DNA (centromeres) ~ 6-7%**
  - **Microsatellites ~ 2%**
  - **Transposable elements ~ 46%**
  - **Pseudogenes ~ 1%**
  - **Other (ancient -telements?) ~ 42%**

- **Variations in gene, repeat density and base composition (isochores) along chromosomes**

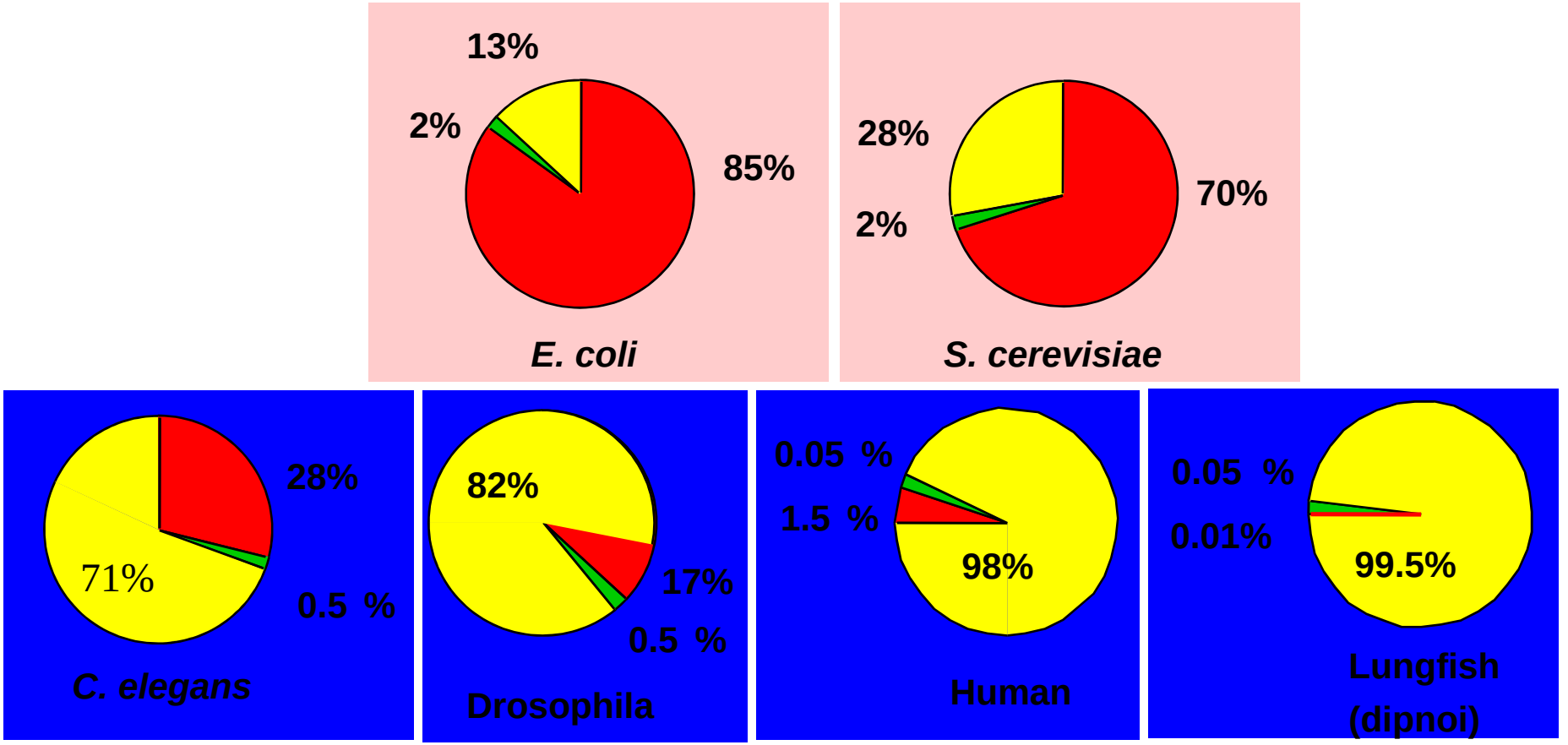


>90% sans fonction connue

Elements fonctionnels non-quantifiés:
- **ARN non-traduits**
- **Promoteurs, enhancers**
- **ORI, MAR, télomères**

Eléments non-fonctionnels:
- **Pseudogènes : 1.2%**
- **Elements transposables : 46%**

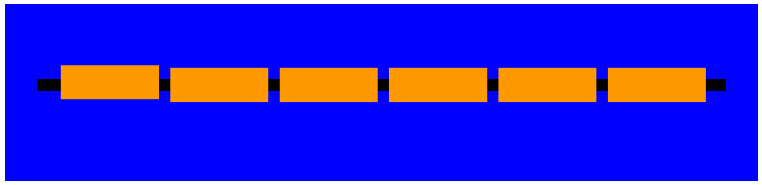
ADN non codant



Les séquences répétées

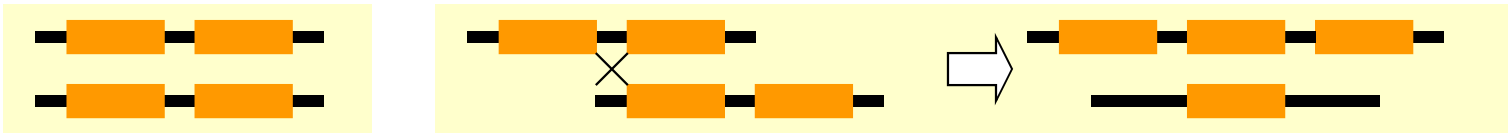
- **Tandem repeats**

	Motif(nt)	L(repeat)	%
Satellite	**2-2000**	**→ 10Mb**	**6.5**
Minisatellite	**2-64**	**100-20000**	**0.3**
Microsatellite	**1-6**	**10-100**	**2**



Slippage of the DNA-pol: **CACACACACACA**

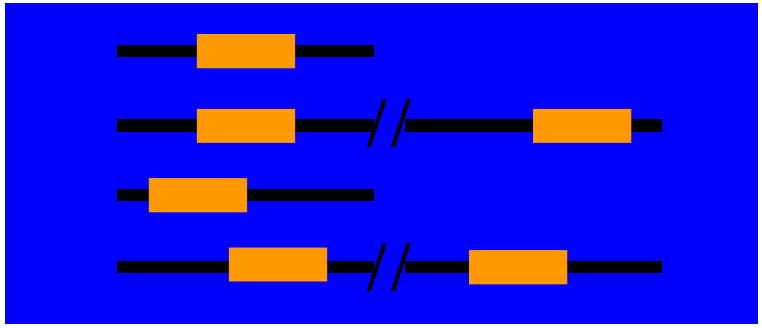
Unequal crossing-over:



- **Interspersed repeats**

  - **DNA transposon (rares)**
  - **Retroelement**

} **46%**



Evolution des génomes

Le modèle Jukes-Cantor

- Probabilité de substitution *par* site *par* untié de temps est **α**

- Substitution : 3 remplacements possibles
(e.g. C → {A,G,T})

- Non-substitution: 1 seule possibilité
(e.g. C → C)

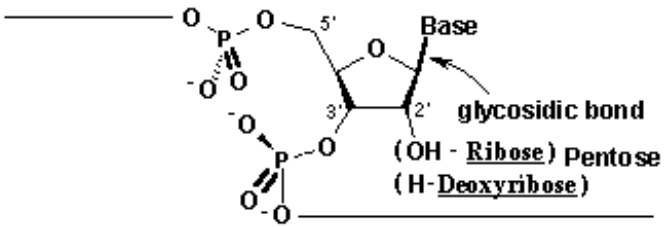
Le modèle Jukes-Cantor

La matrice de transition de la chaîne de Markov a la forme suivante:

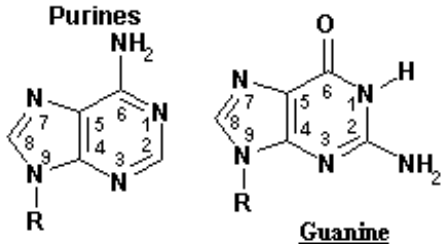
$M_{JC} =$

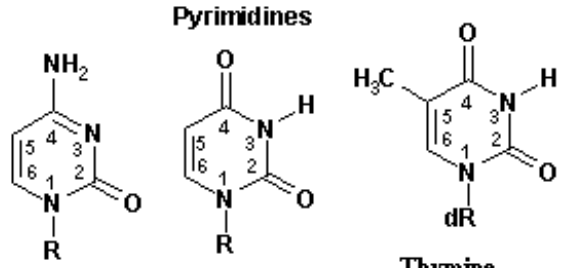
	A	C	G	T
A	1-α	α/3	α/3	α/3
C	α/3	1-α	α/3	α/3
G	α/3	α/3	1-α	α/3
T	α/3	α/3	α/3	1-α

Purines et pyrimidines

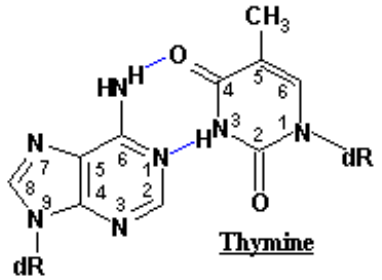


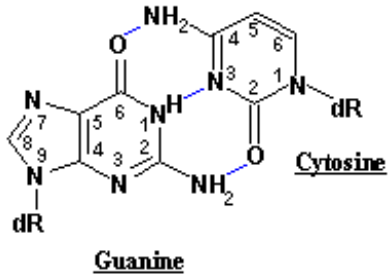
Nucleotide monophosphate



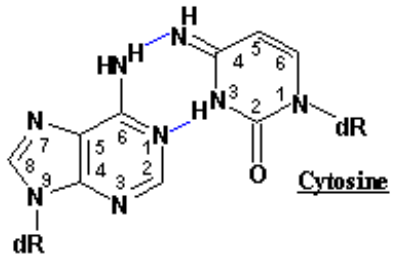


CORRECT PAIRINGS

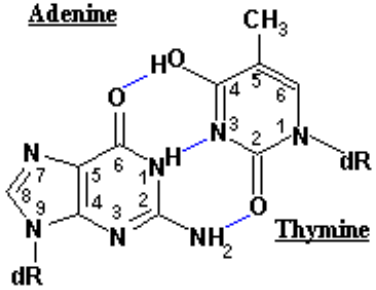




WRONG PAIRINGS



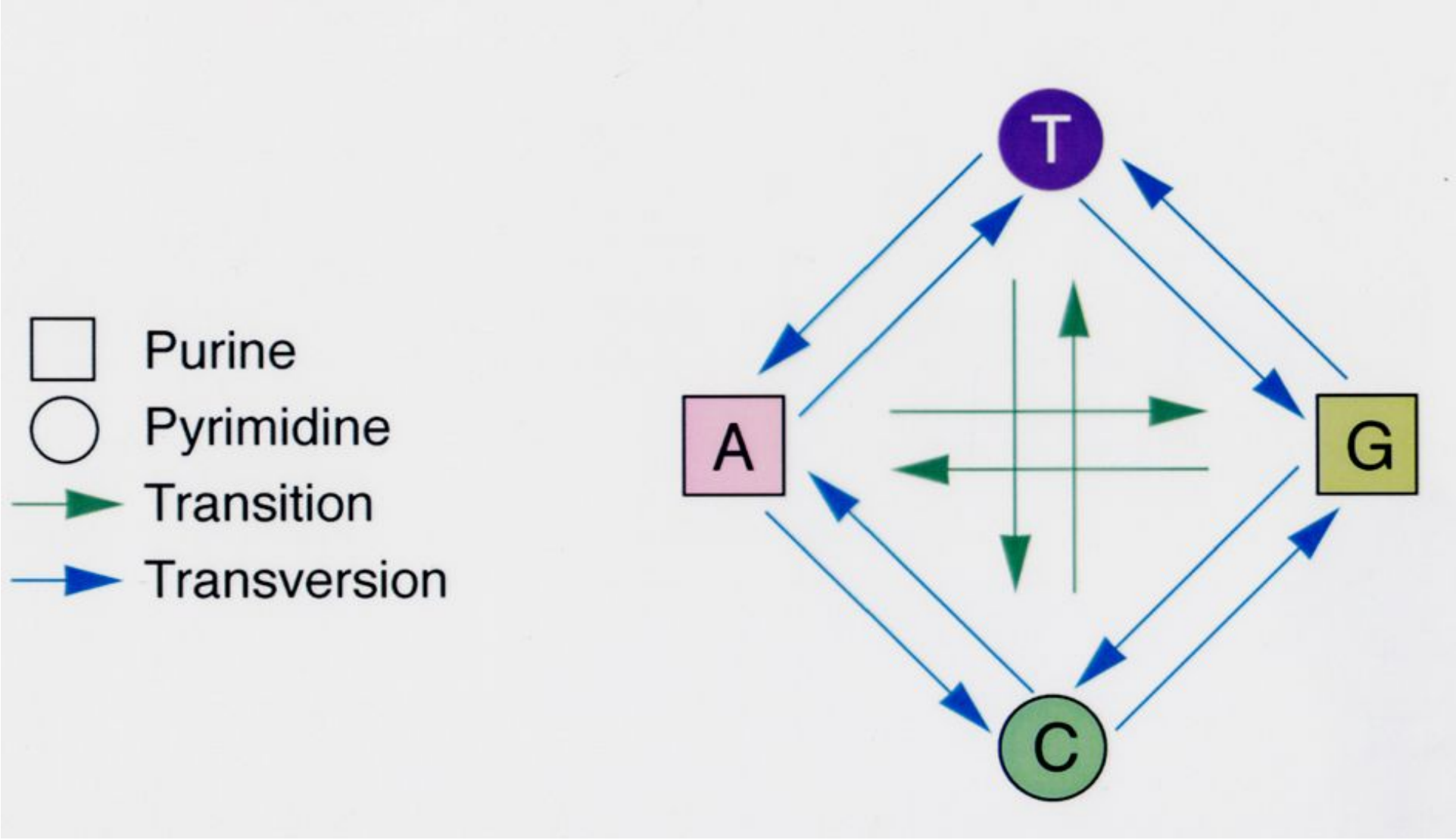
Cytosine in its imino form can pair wrongly with adenine



Thymine in its enol form can pair wrongly with guanine

blue lines = hydrogen bonds

Transitions et transversions



Le modèle de Kimura

Inclut le biais de substitution entre transistions et transversions

Probabilité de transition (G↔A et T↔C) *par* site *par* unité de temps est α

Probabilité de transversion (G↔T, G↔C, A↔T, and A↔C) *par* site *par* unité de temps est β

Le modèle de Kimura

La matrice de transition de la chaîne de Markov a alors la forme suivante:

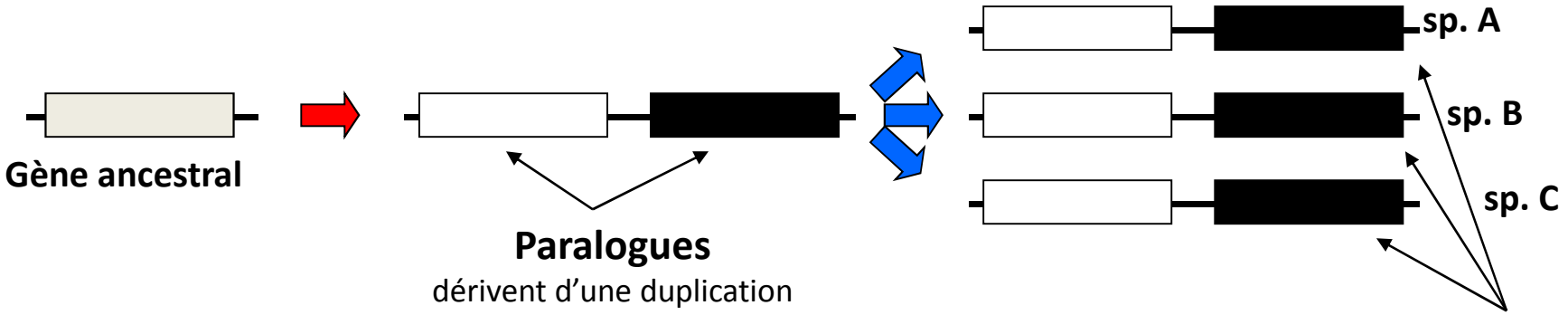
$M_{K2P} =$

	A	C	G	T
A	$1-\alpha-\beta$	β	α	β
C	β	$1-\alpha-\beta$	β	α
G	α	β	$1-\alpha-\beta$	β
T	β	α	β	$1-\alpha-\beta$

Modèles de substitution plus avancés

- Plus réalistes
- Le modèles de Kimura a une distribution uniforme stationnaire comme solution, ce qui n'est pas vrai
- Une possibilité: la probabilité de substitution d'une purine à une pyrimidine est différente de celle d'une pyrimidine à une purine (solution dynamique).
- Le modèle HKY capture ce biais

Duplications géniques



Paralogues
dérivent d'une duplication

Orthologues
Assument la même fonction chez des espèces différentes

Devenir des gènes dupliqués

Divergence & perte de fonction (pseudogène)

PSEUDOGENISATION

Conservation de la même fonction (redondance) - Spécialisation

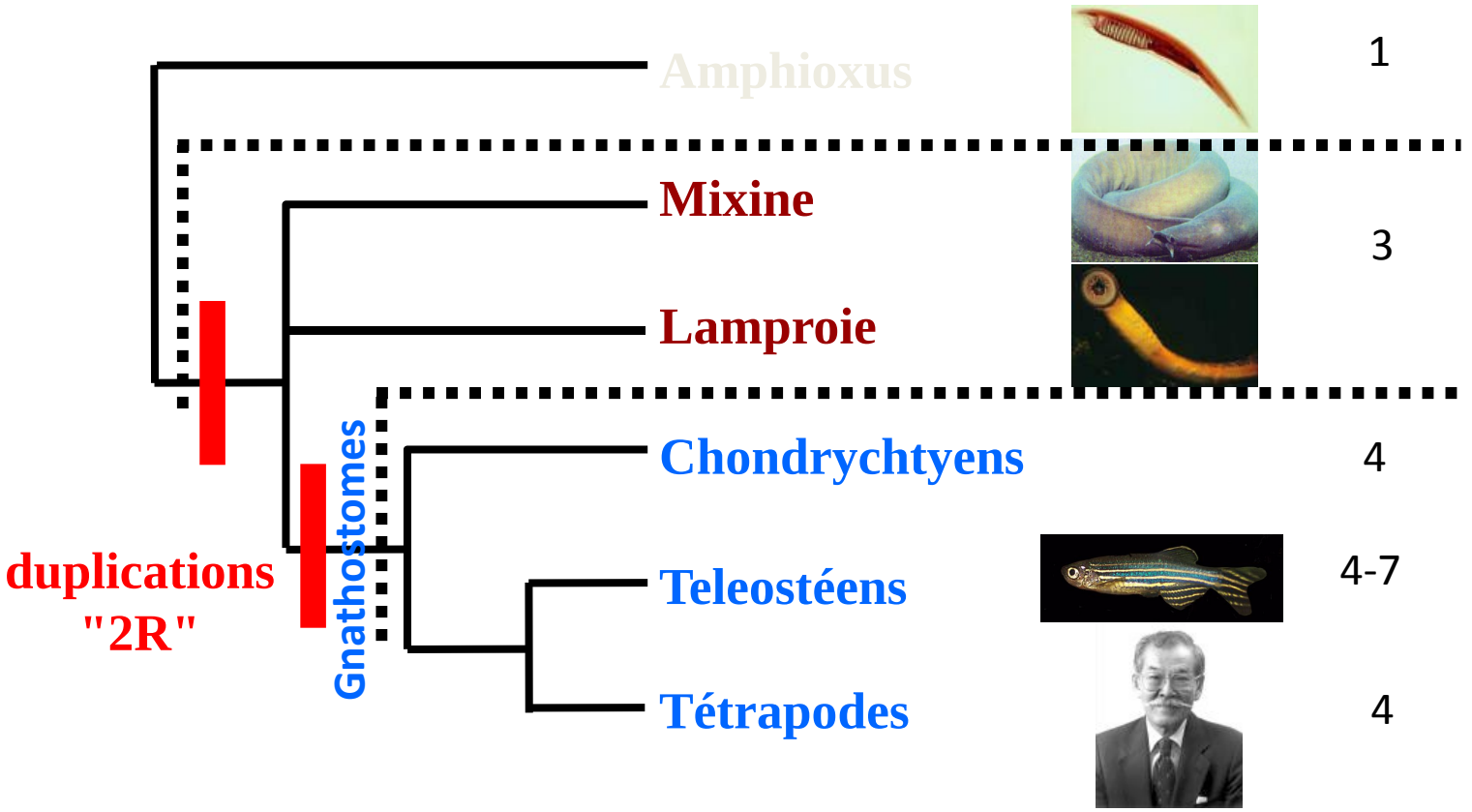
SUB-FONCTIONNALISATION

Acquisition d'une nouvelle fonction

NEO-FONCTIONNALISATION

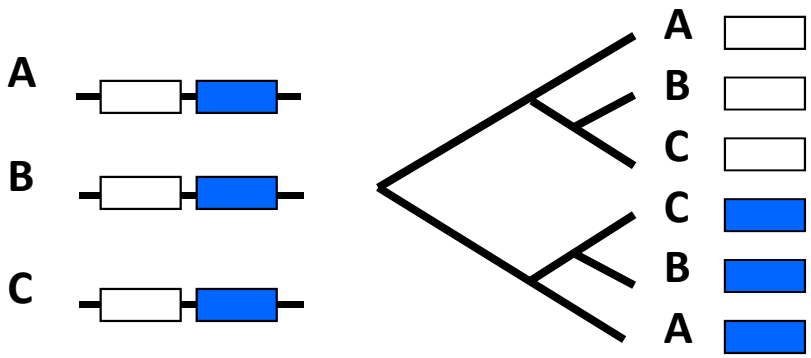
Duplications de groupes de gènes

Evolution des espèces



Hox clusters

Duplications, paralogie et phylogénie



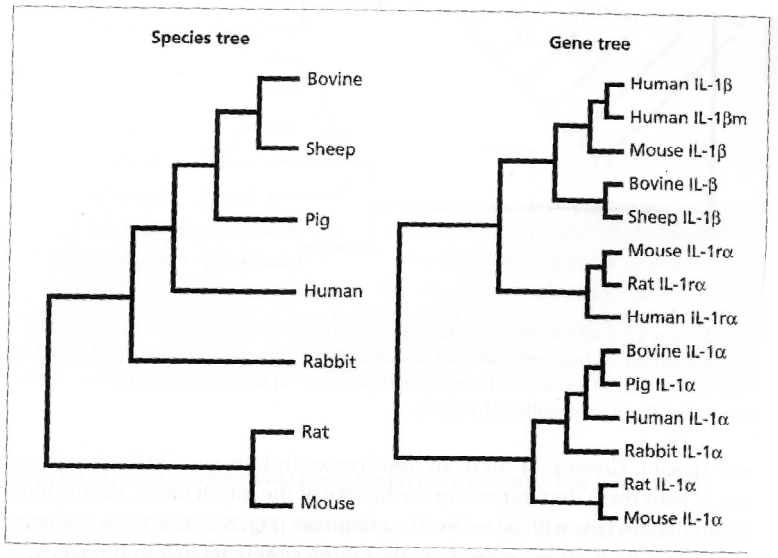


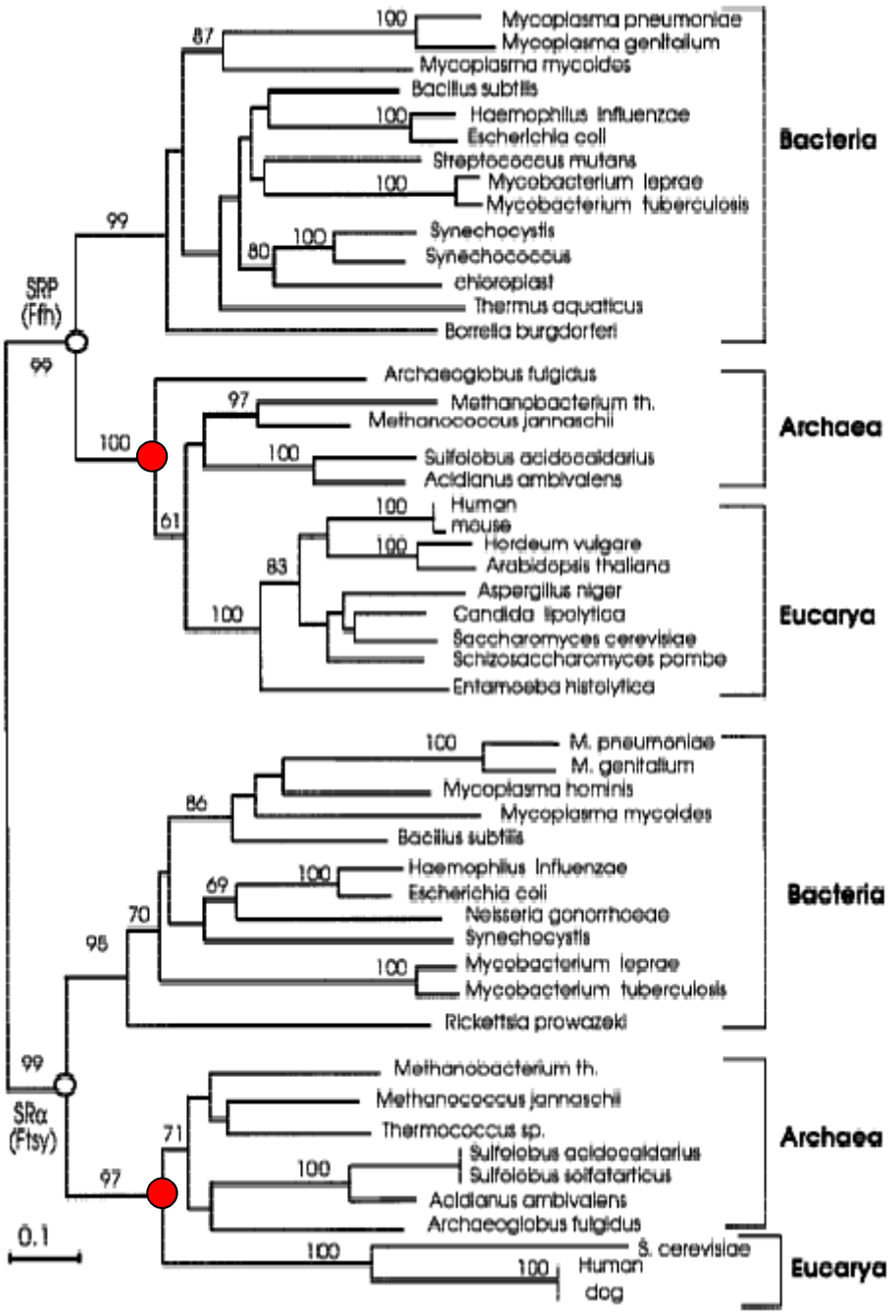
Fig. 8.5 Tree for mammals and for mammalian interleukin-1 genes.

The Root of the Universal Tree of Life Inferred from Anciently Duplicated Genes Encoding Components of the Protein-Targeting Machinery

Simonetta Gribaldo, Piero Cammarano

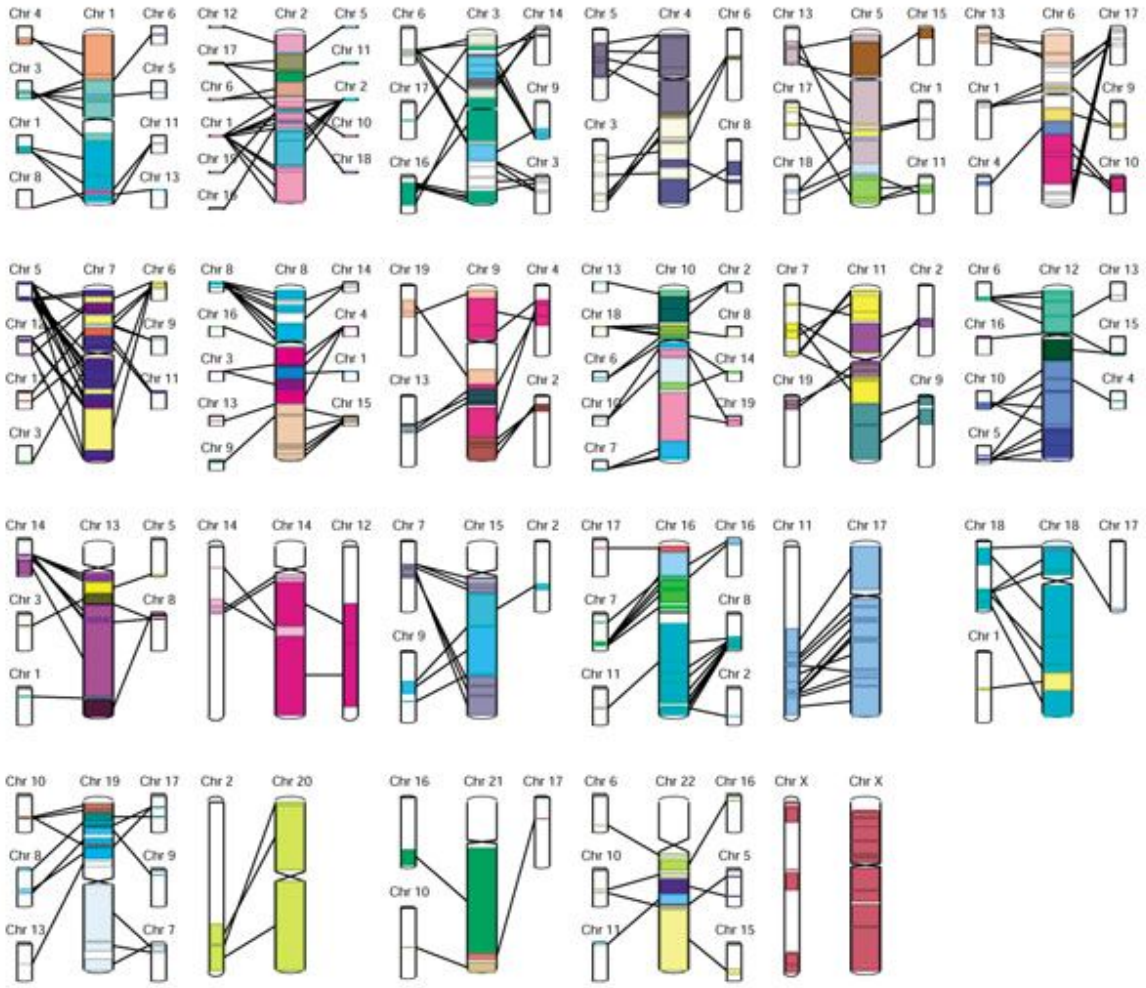
J Mol Evol (1998) 47:508–516

Fig. 3. Neighbor-joining tree constructed from the SRP54(Ffh)/SRα(Ftsy) alignment. Numbers attached to internal nodes are BCL based on 100 bootstrap replicates of the original alignment. BCLs <50% are not shown. Evolutionary distance matrices were generated by the PROTDIST program with the "Dayhoff" option (see Methods). Scale bar represents 0.1 amino acid substitution per site.



ACQUISITION DU NOYAU

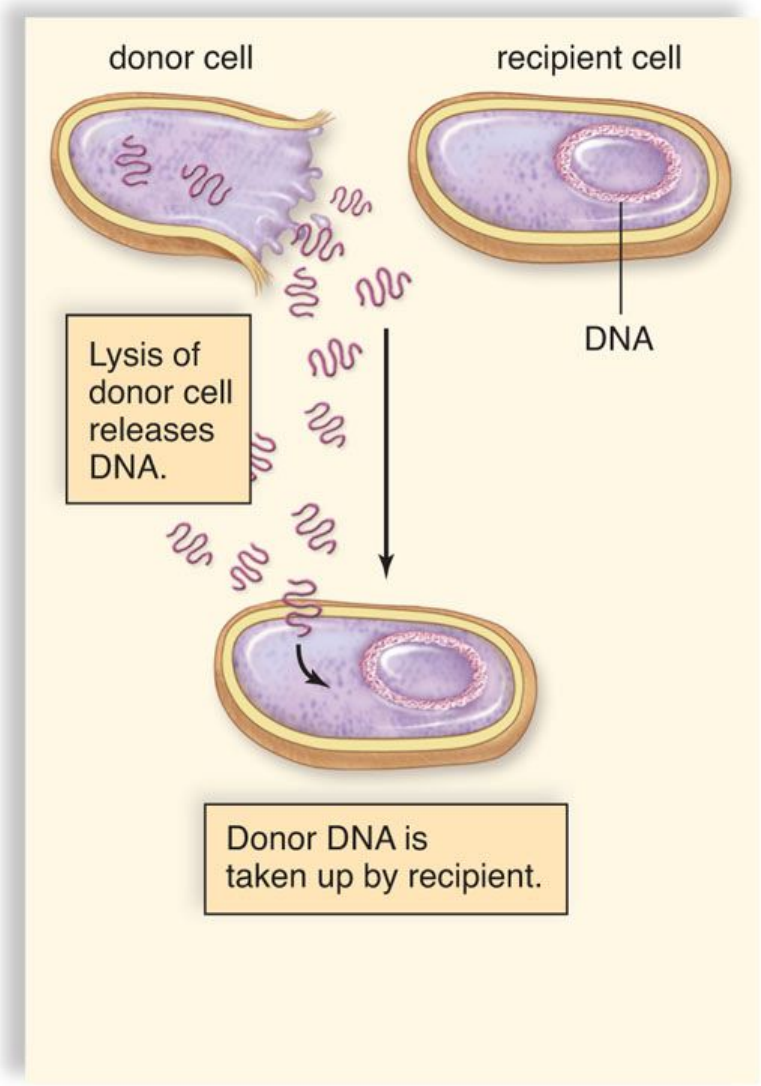
Synténie


Chr 4 Chr 1 Chr 6
Chr 3 Chr 5
Chr 1 Chr 11
Chr 8 Chr 13
Chr 12 Chr 2 Chr 5
Chr 17 Chr 11
Chr 6 Chr 2
Chr 1 Chr 10
Chr 18
Chr 6 Chr 3 Chr 14
Chr 17 Chr 9
Chr 16 Chr 3
Chr 5 Chr 4 Chr 6
Chr 3 Chr 8
Chr 13 Chr 5 Chr 15
Chr 17 Chr 1
Chr 18 Chr 11
Chr 13 Chr 6 Chr 17
Chr 1 Chr 9
Chr 4 Chr 10
Chr 5 Chr 7 Chr 6
Chr 9
Chr 11
Chr 3
Chr 8 Chr 8 Chr 14
Chr 16 Chr 4
Chr 3 Chr 1
Chr 13 Chr 15
Chr 9
Chr 19 Chr 9 Chr 4
Chr 13 Chr 2
Chr 13 Chr 10 Chr 2
Chr 18 Chr 8
Chr 6 Chr 14
Chr 19
Chr 7
Chr 7 Chr 11 Chr 2
Chr 19 Chr 9
Chr 6 Chr 12 Chr 13
Chr 16 Chr 15
Chr 10 Chr 4
Chr 5
Chr 14 Chr 13 Chr 5
Chr 3 Chr 8
Chr 1
Chr 14 Chr 14 Chr 12
Chr 7 Chr 15 Chr 2
Chr 9
Chr 17 Chr 16 Chr 16
Chr 7 Chr 8
Chr 11 Chr 2
Chr 11 Chr 17
Chr 18 Chr 18 Chr 17
Chr 1
Chr 10 Chr 19 Chr 17
Chr 8 Chr 9
Chr 13 Chr 7
Chr 2 Chr 20
Chr 16 Chr 21 Chr 17
Chr 10
Chr 6 Chr 22 Chr 16
Chr 10 Chr 5
Chr 11 Chr 15
Chr X Chr X

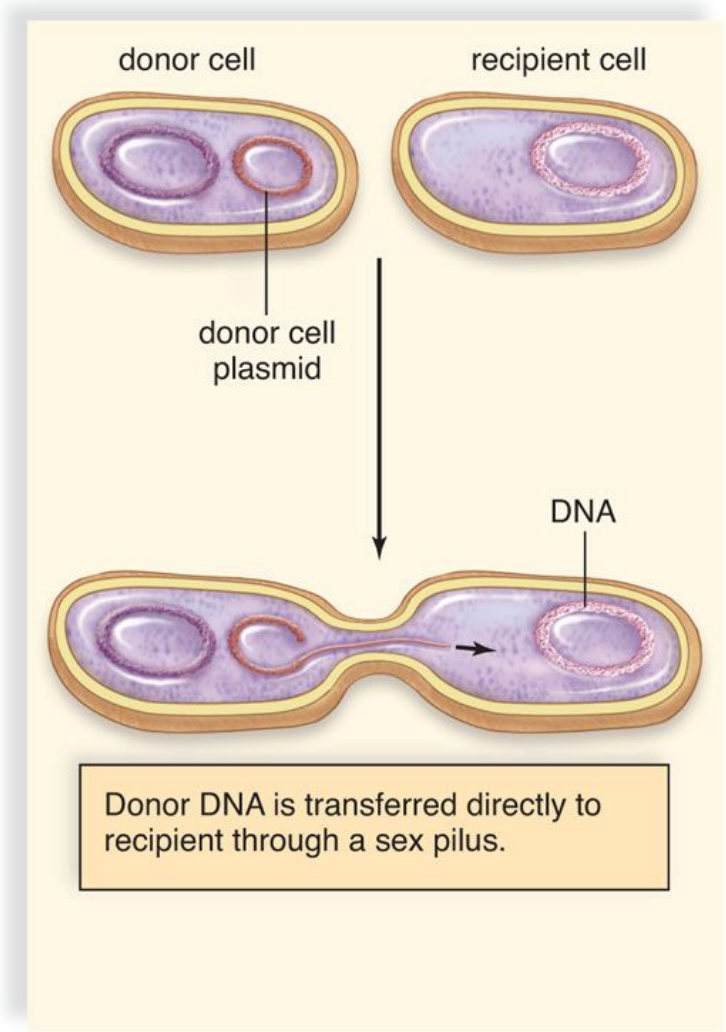
Transfert horizontal

- **Transformation –** une bactérie ingère de l'ADN provenant du milieu environnant.
- **Conjugaison –** La bactérie donneuse envoie de l'ADN à une autre bactérie par l'intermédiaire d'un pilus sexuel.
- **Transduction –** Un bactériophage transfère une portion d'ADN d'une bactérie à l'autre au cours d'infections multiples.

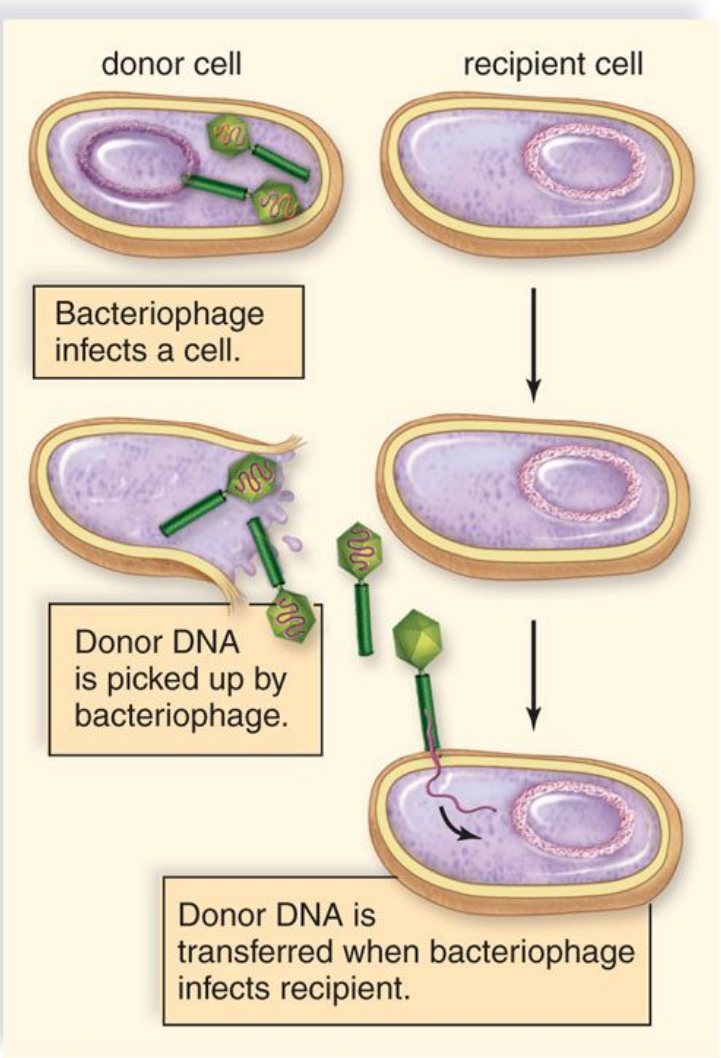
Transformation



Conjuguaison



Transduction



Chromosome Packing

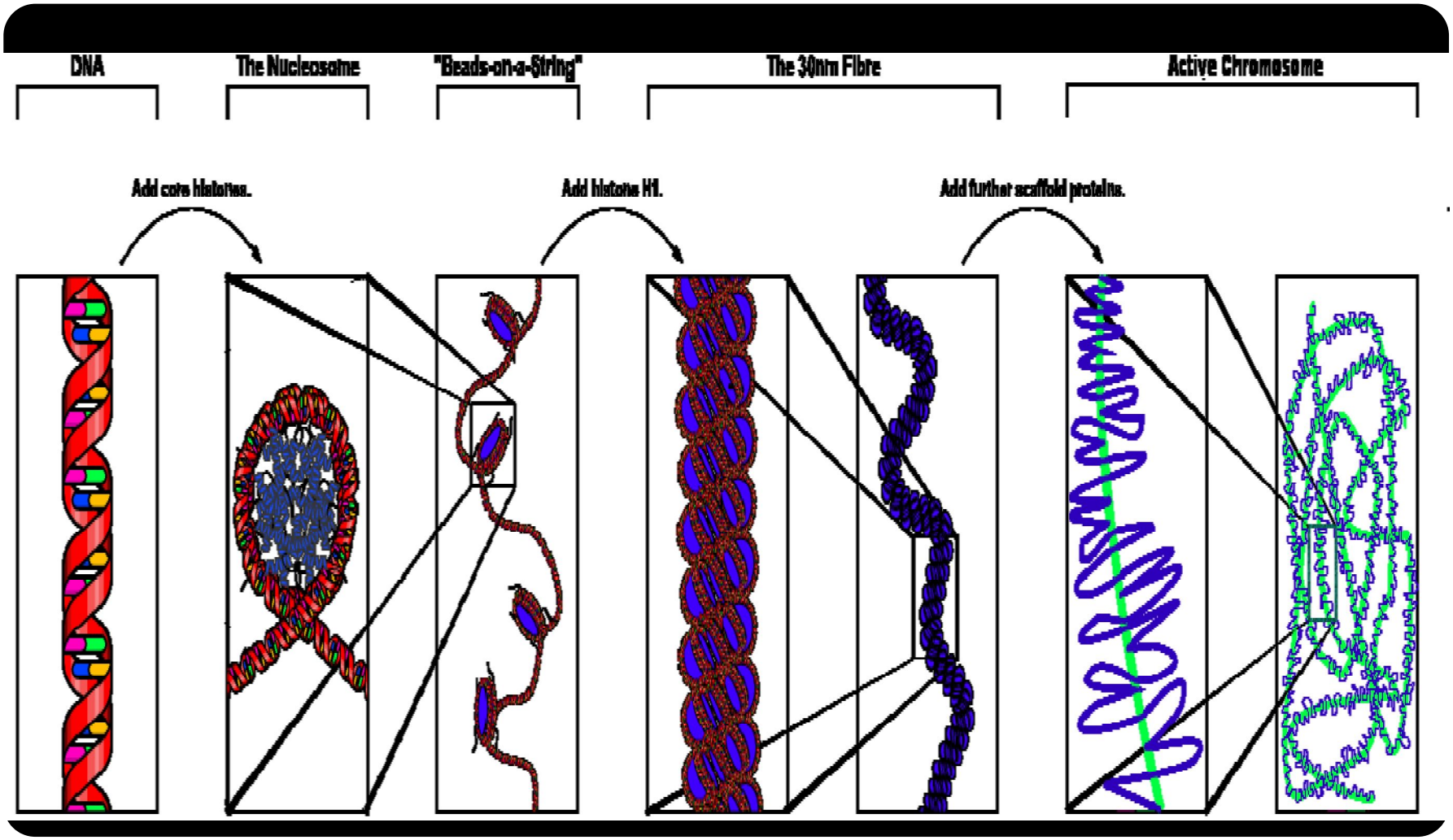


Figure 3.15 The chloroplast genome codes for 4 rRNAs, 30 tRNAs, and ~50 proteins.

Genes	Types
RNA-coding	
16S rRNA	1
23S rRNA	1
4.5S rRNA	1
5S rRNA	1
tRNA	30
Gene Expression	
r-proteins	19
RNA polymerase	3
Others	2
Thylakoid Membranes	
Photosystem I	2
Photosystem II	7
Cytochrome *b*/*f*	3
H^+-ATPase	6
Others	
NADH dehydrogenase	6
Ferredoxin	3
Ribulose BP Cblase	1
Unidentified	29
Total	110

Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications

(Archaea/Bacteria/eukaryote/phylogeny)

JAMES R. BROWN* AND W. FORD DOOLITTLE

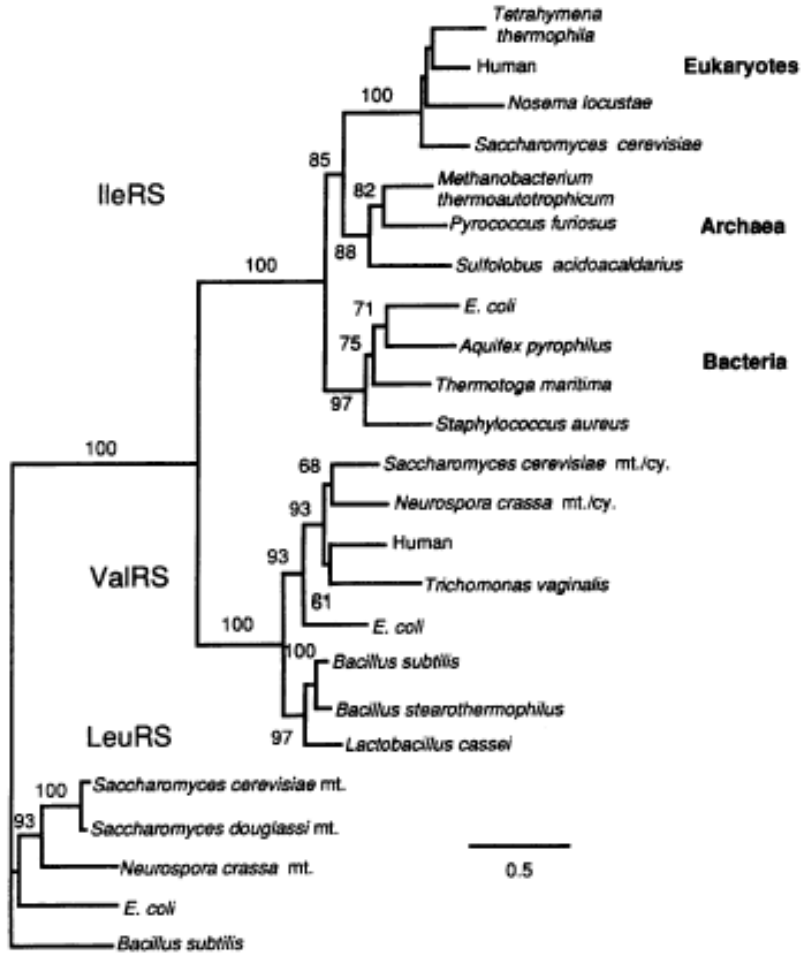


FIG. 3. Neighbor-joining tree of IleRS, ValRS, and LeuRS genes using the program NEIGHBOR (26). The scale represents 0.5 expected number of amino acid replacements per position as determined with the program PROTDIST. Numbers are the frequency of occurrence of nodes that exceeded 50% of 300 bootstrap replicates.

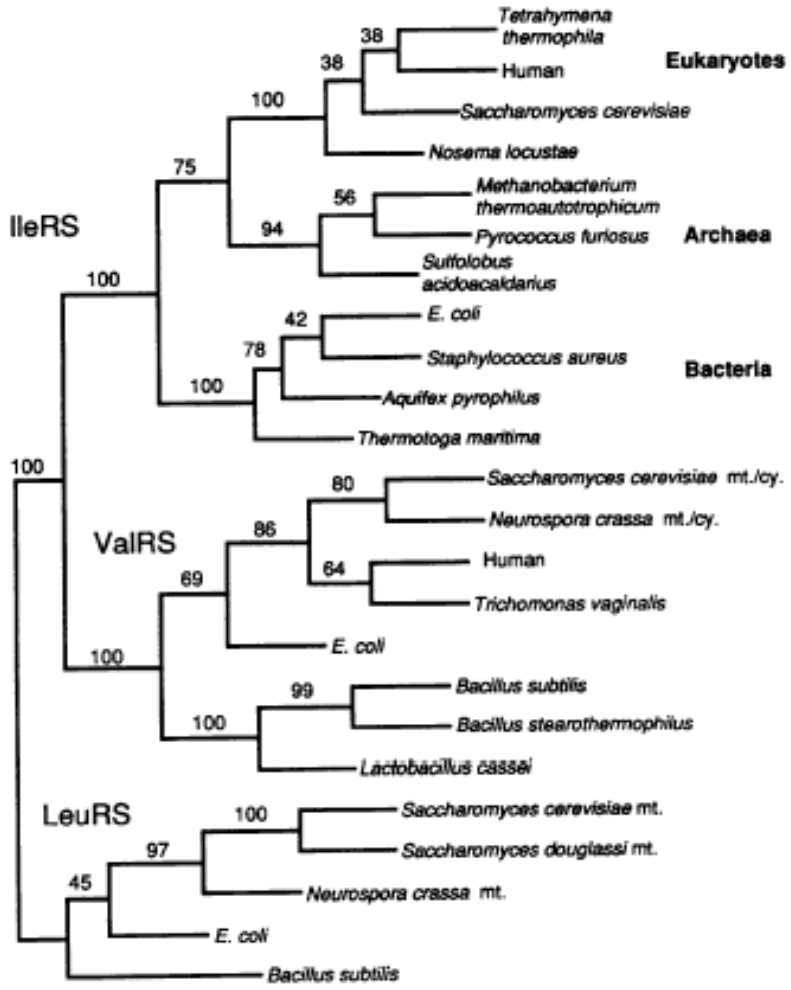
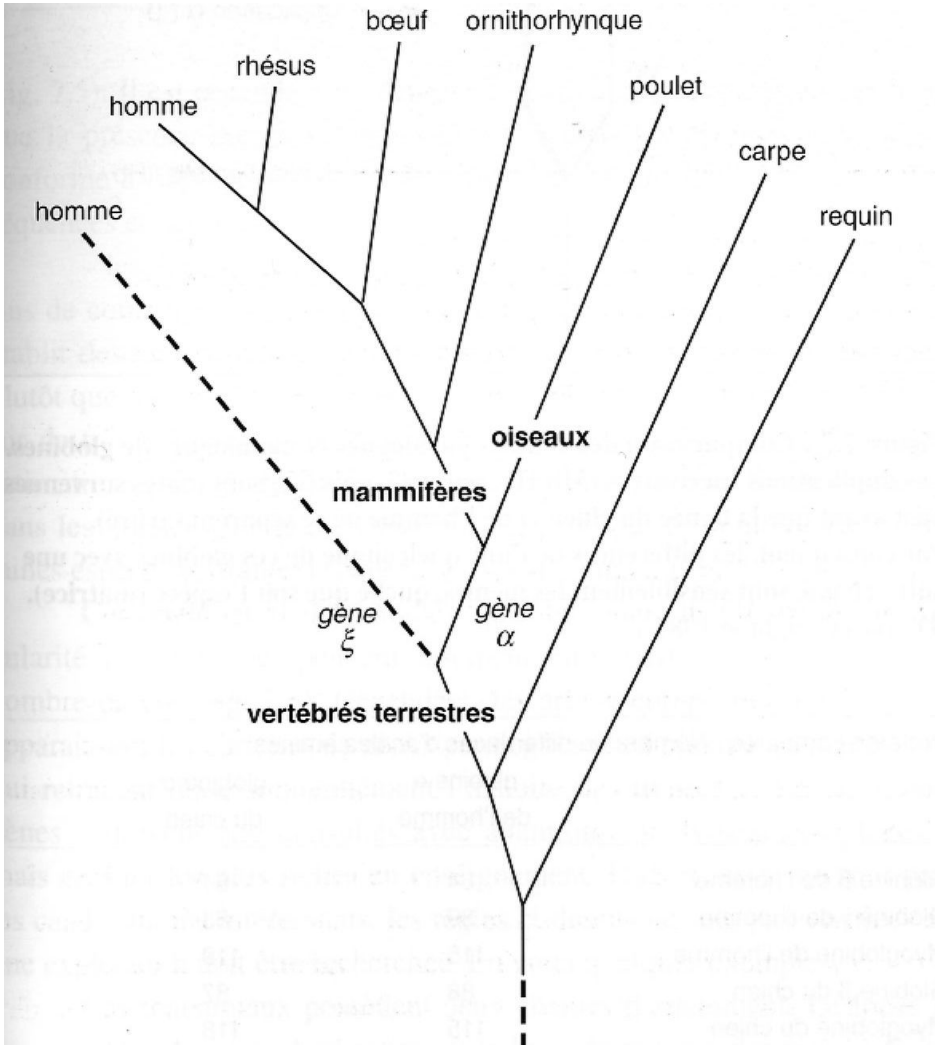


FIG. 2. Consensus maximum parsimony tree of IleRS, ValRS, and LeuRS genes using the program PROTPARS (26). Numbers are the frequency of occurrence of nodes in 300 bootstrap replicates.

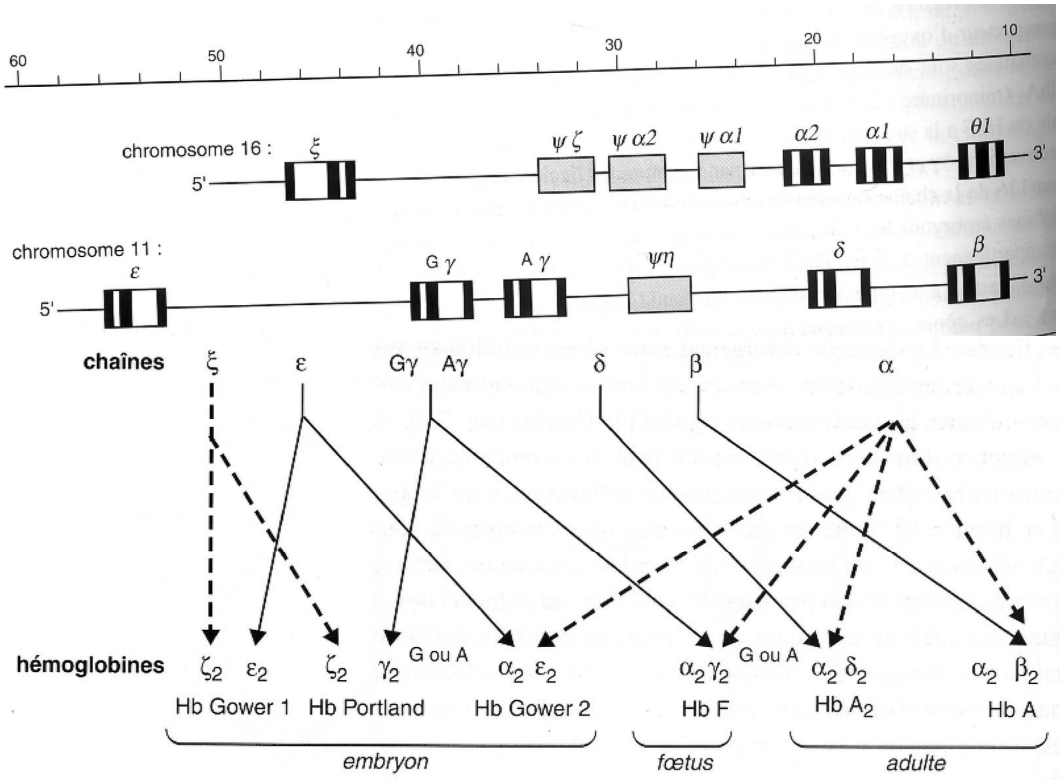
Duplications géniques & génomiques

Identification des orthologues et des paralogues de globine



Duplications géniques & génomiques

Evolution des globines



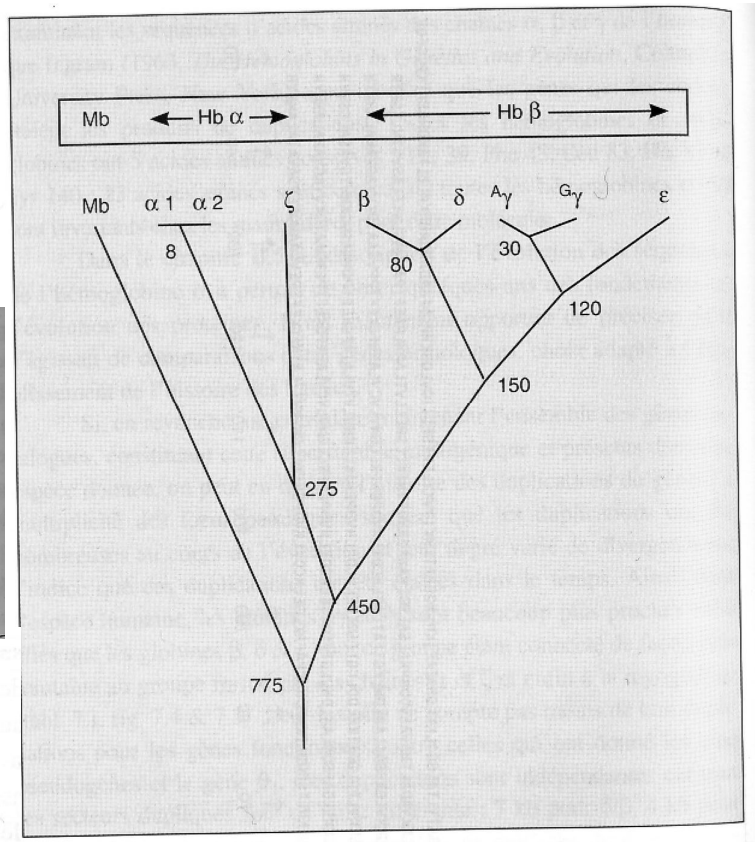


Figure 7.5 : Phylogénie des globines humaines.
L'âge approximatif des duplications est indiqué. Les duplications myoglobine/hémoglobine et Hbα/ Hbβ sont très anciennes. La lignée α comprend deux duplications : la dernière a donné α_1 et α_2, il y a environ 8 MA ; les deux chaînes sont identiques. La diversification des gènes β comprend une première duplication, générant un proto-β et un proto-ε vers 150 ou 200 MA. Les marsupiaux n'ont que ces deux gènes. À la base du rameau euthérien, vers 80-100 MA, la duplication de proto-β a fourni β et δ et la duplication de proto-ε a fourni ε, γ et η. Ce dernier est devenu un pseudogène chez les primates. Les chaînes $^{A}\gamma$ et $^{G}\gamma$ (duplication il y a 25 à 35 MA) diffèrent par 1 seul acide aminé (Ala ou Gly en position 136).

Figure 18.12 The human X chromosome can be divided into distinct regions by its banding pattern.The short arm is *p* and the long arm is *q*; each arm is divided into larger regions that are further subdivided. This map shows a low resolution structure; at higher resolution, some bands are further subdivided into smaller bands and interbands, e.g. p21 is divided into p21.1, p21.2, and p21.3.

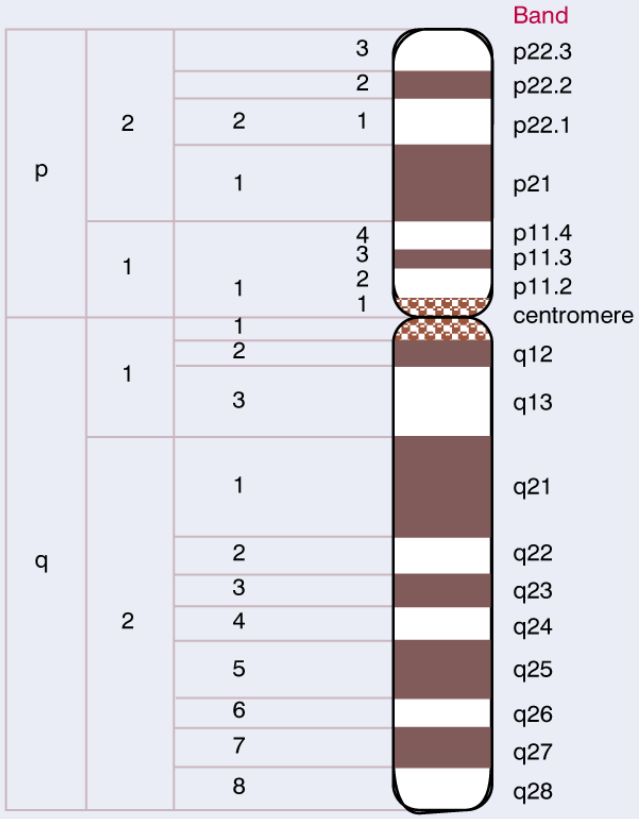


Figure 18.11 G-banding generates a characteristic lateral series of bands in each member of the chromosome set. Photograph kindly provided by Lisa Shaffer.

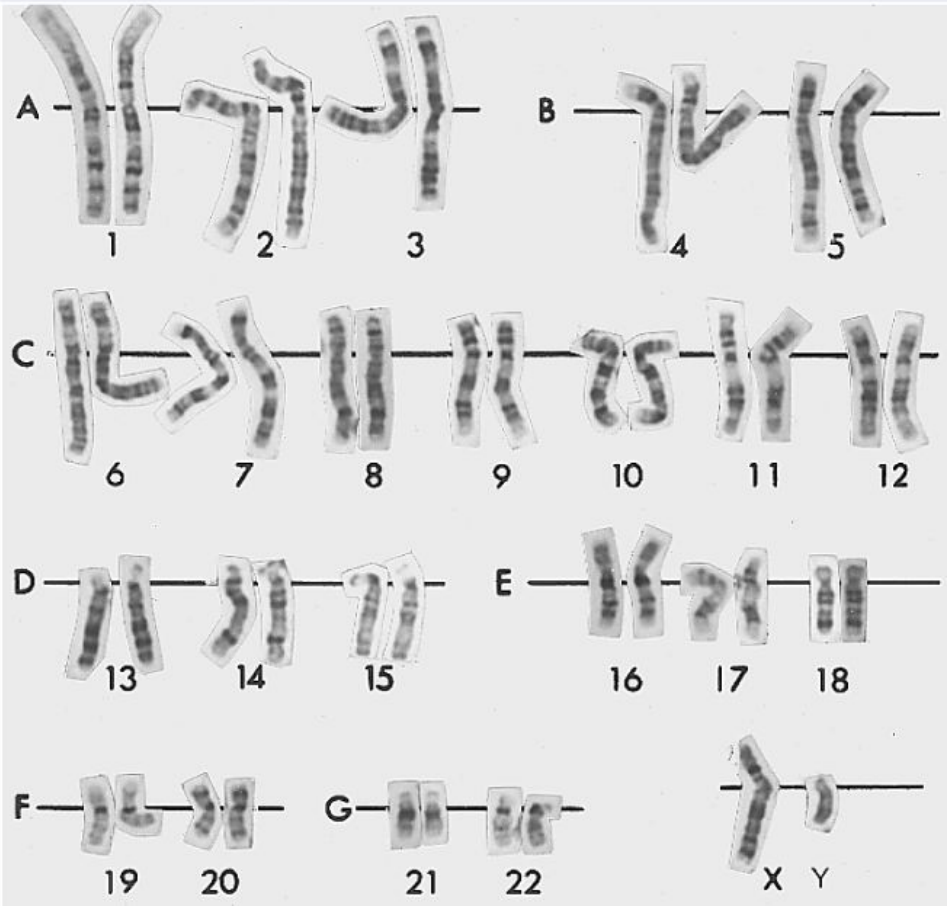
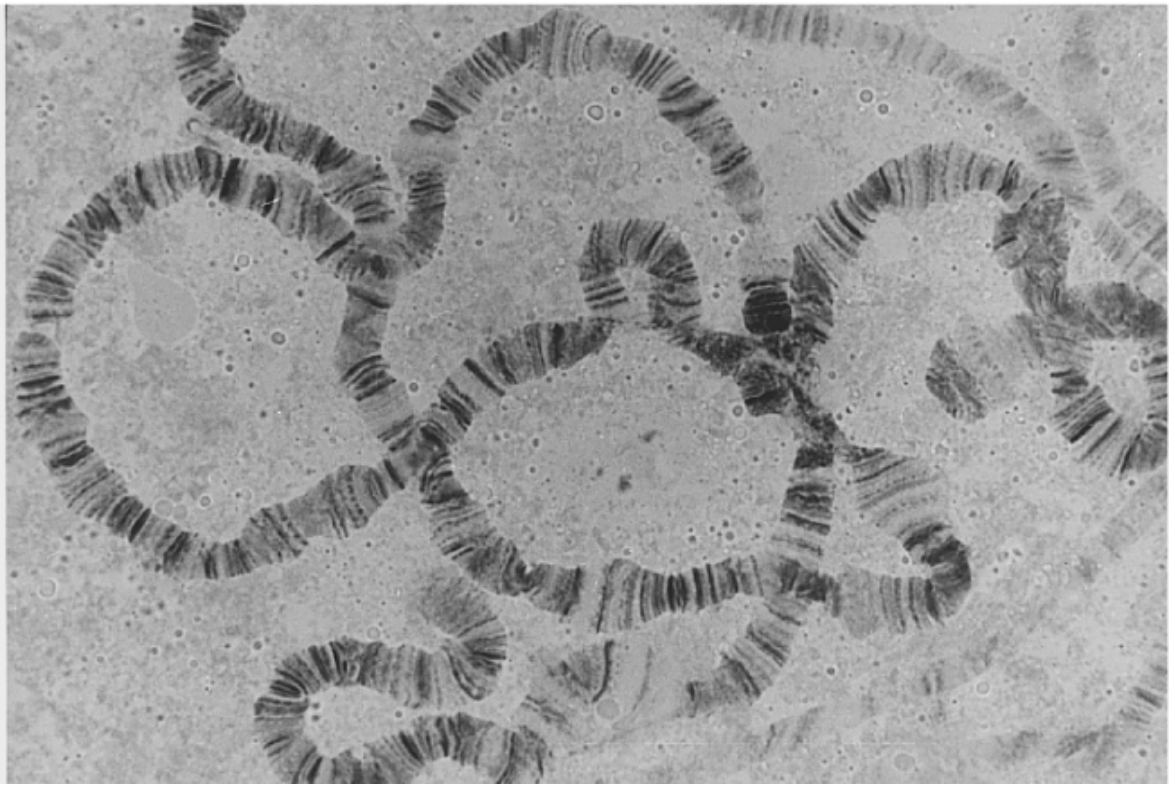


Figure 18.15 The polytene chromosomes of *D. melanogaster* form an alternating series of bands and interbands. Photograph kindly provided by Jose Bonner.

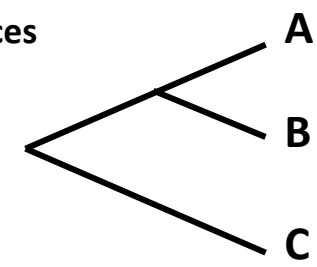


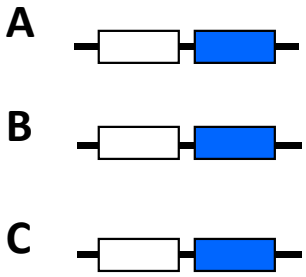
Les bandes chromosomiques

Duplications géniques & génomiques

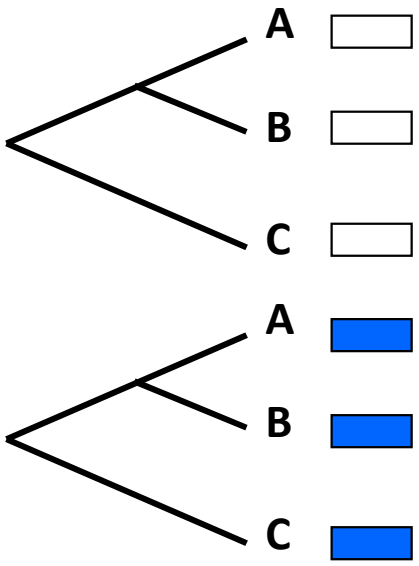
Identification des orthologues et des paralogues

Les espèces

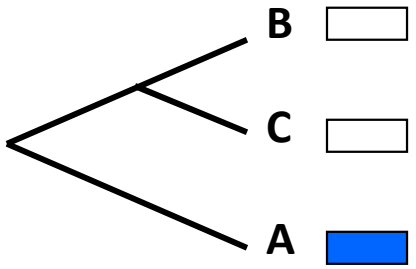




Phylogénies obtenues en comparant des orthologues



Phylogénie obtenues en comparant des orthologues et des paralogues
(Si le paralogue chez A a fortement divergé)



Phylogénie d'orthologues = Phylogénie des espèces
Phylogénie de paralogues = Phylogénie différente