

Chp 5 : La taille des génomes

I – La valeur C

**quantité totale d'ADN dans un génome
haploïde**

caractéristique de chaque espèce

Grande variation dans le monde vivant

Augmentation de la complexité

**→ Augmentation de la taille minimale de C
dans chaque embranchement**

cependant

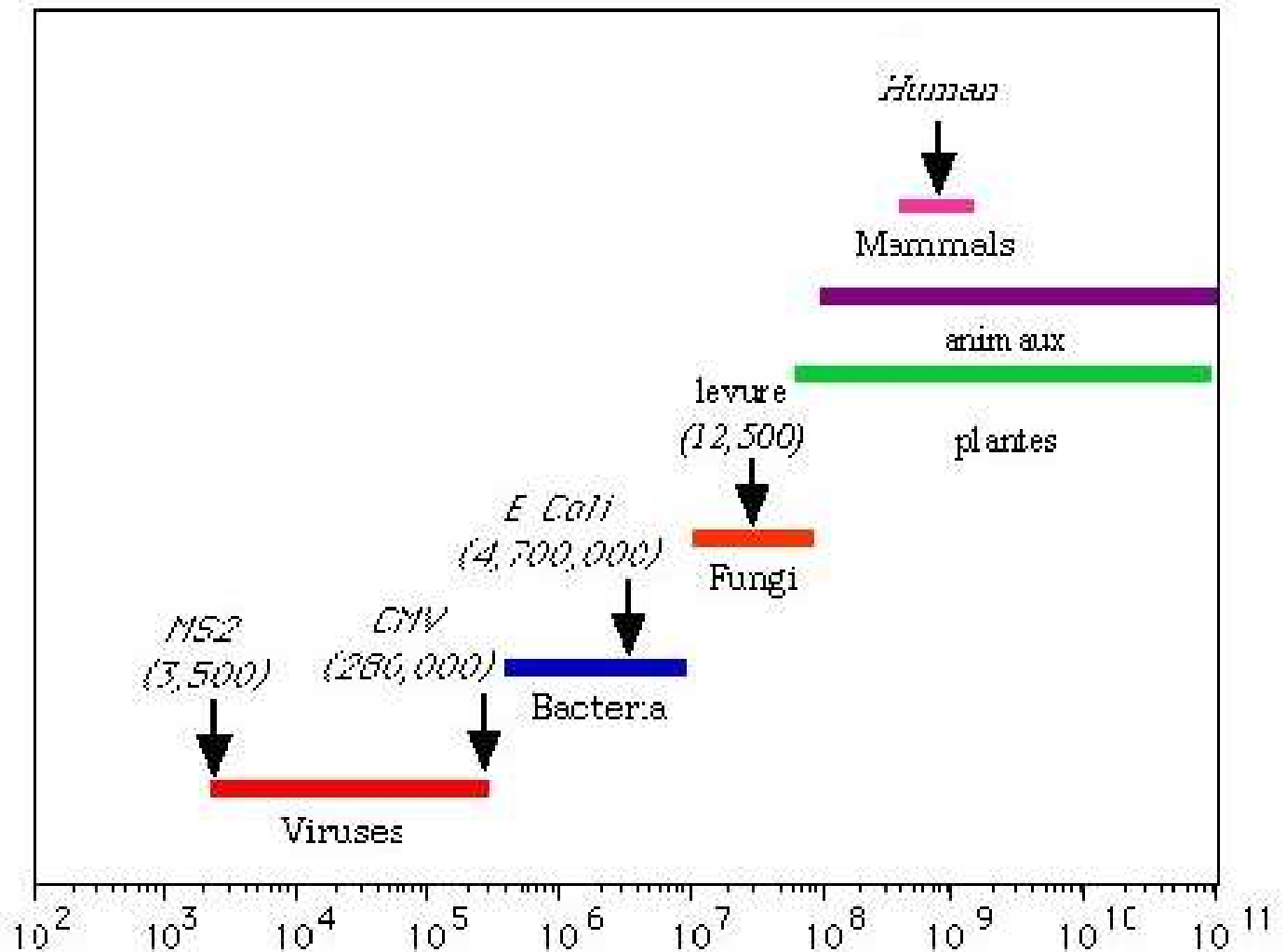
Relation complexité du génome

/ complexité de l'organisme

indéterminable

Grande dispersion dans certains groupes

Comparison of Genome Size:



**La valeur C augmente avec la complexité
des organismes**

MAIS

**grandes variations au sein de certains
taxons**

ex chez les vertébrés :

***H. sapiens*: 3 200 Mb**

***Protopterus sp.* : 140 000 Mb**



**Pas de relation complexité de l'organisme /
valeur C**

**Variation de taille des génomes =
paradoxe de la valeur C**

QUESTION :

**Les génomes les plus grands contiennent-ils
- davantage de gènes différents**

ou

**- davantage de copies de mêmes
gènes ?**

QUESTION :

Hypothèse:

**Les grands génomes contiennent
davantage de gènes différents**



**augmentation de la diversité des gènes =
augmentation de la complexité du
génom**



**Proportion de gènes présents en
exemplaire unique
plus élevée**

II– Composition du génome des eucaryotes

Caractérisation par mesure physique : Co.t
= estimée par cinétique de
ré-association après dénaturation

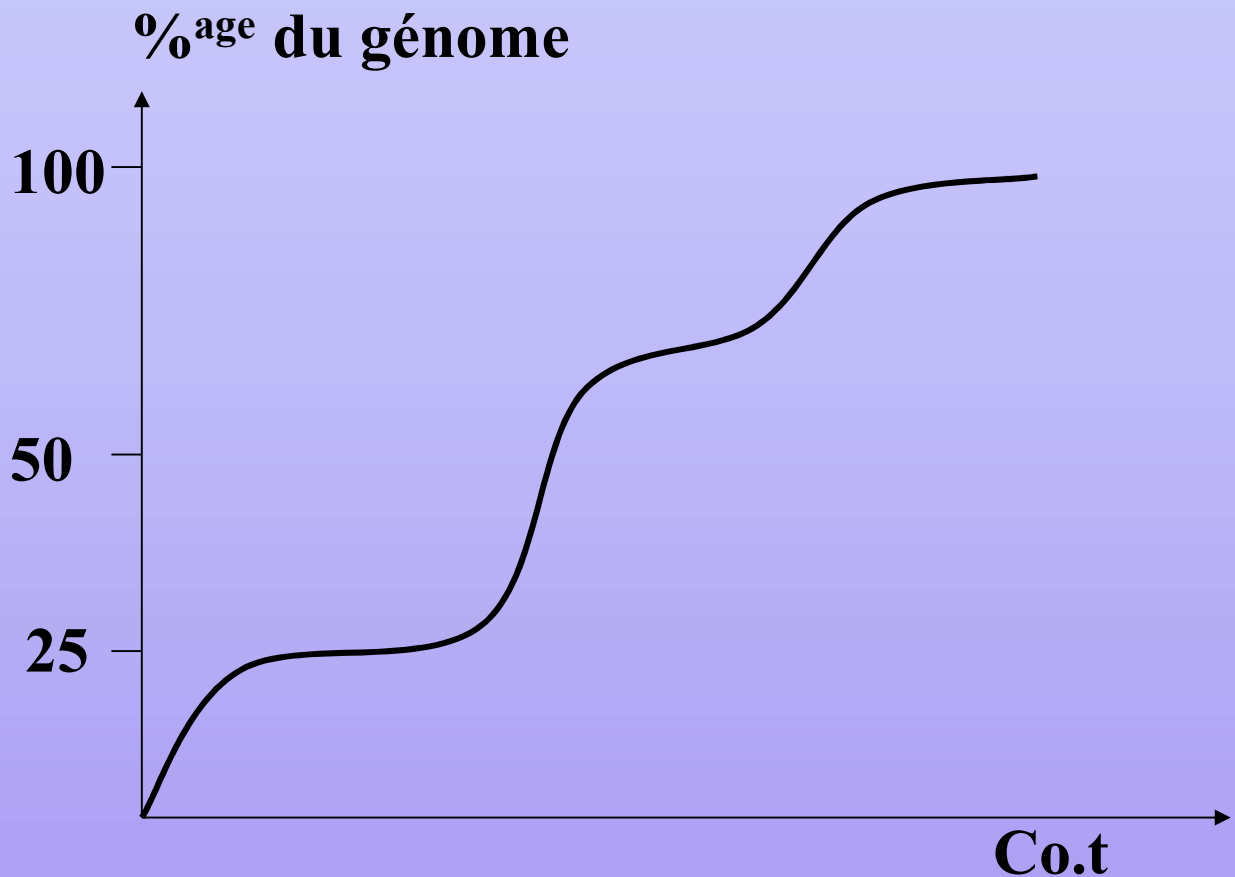
Vitesse d'hybridation : proportionnelle à la
fréquence de répétition d'une sequence

Co.t : inversement proportionnel à la
vitesse de réassociation



Estimation de la complexité du génome
(séquences uniques vs séquences répétées)

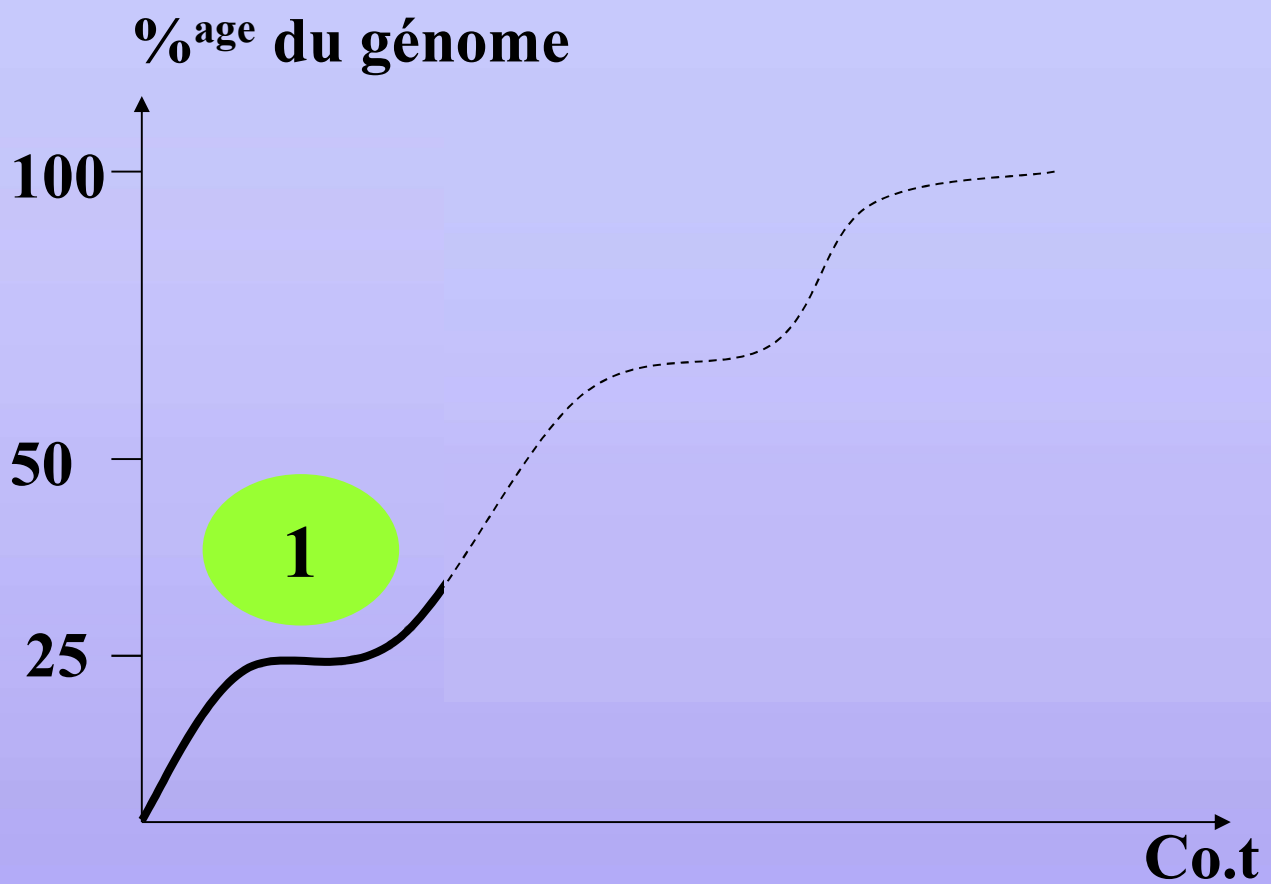
Co.t d'un génome complet:



3 phases :

3 vitesses de ré-association

3 catégories d'ADN à taux de répétition différents

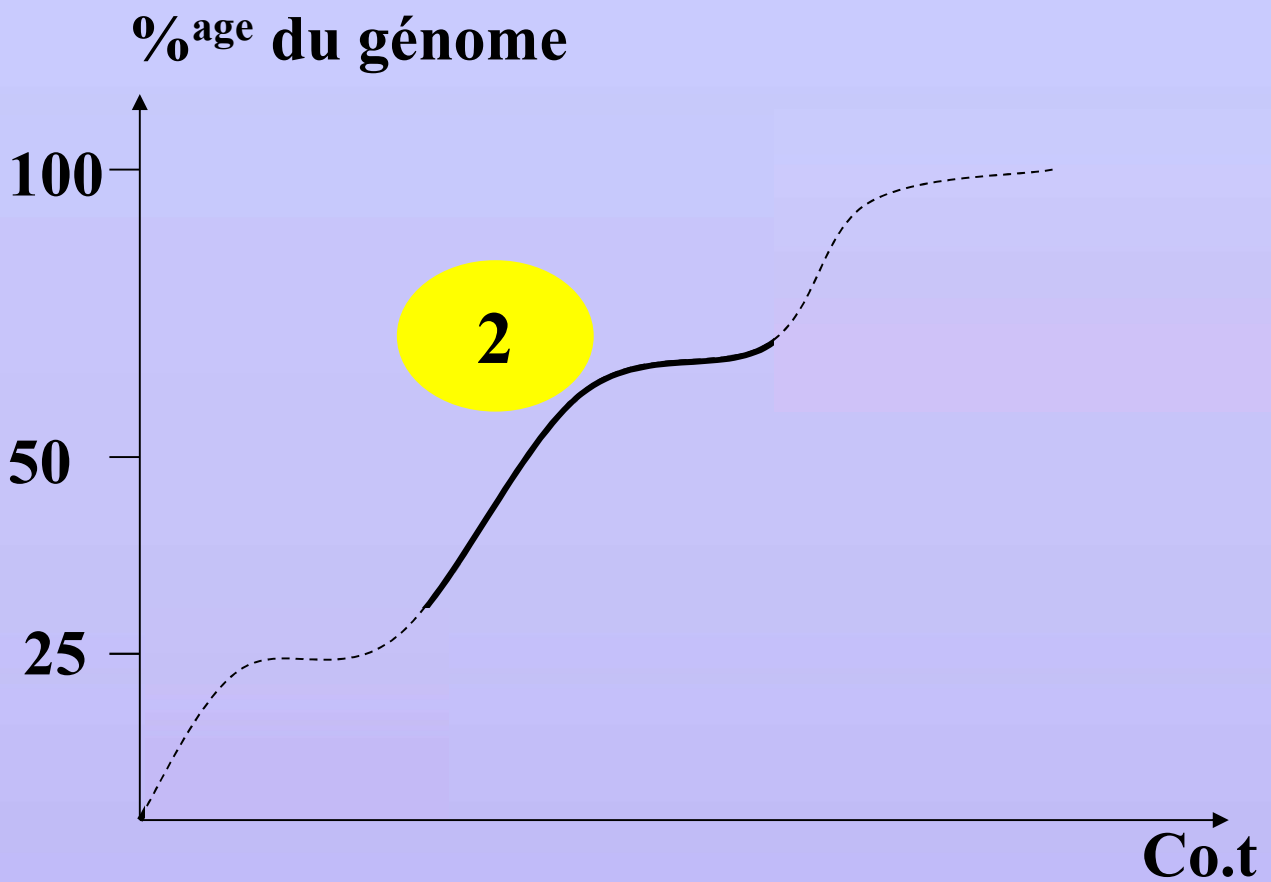


- Co.t faible : ré-association rapide



séquences hautement répétées

25% du génome total



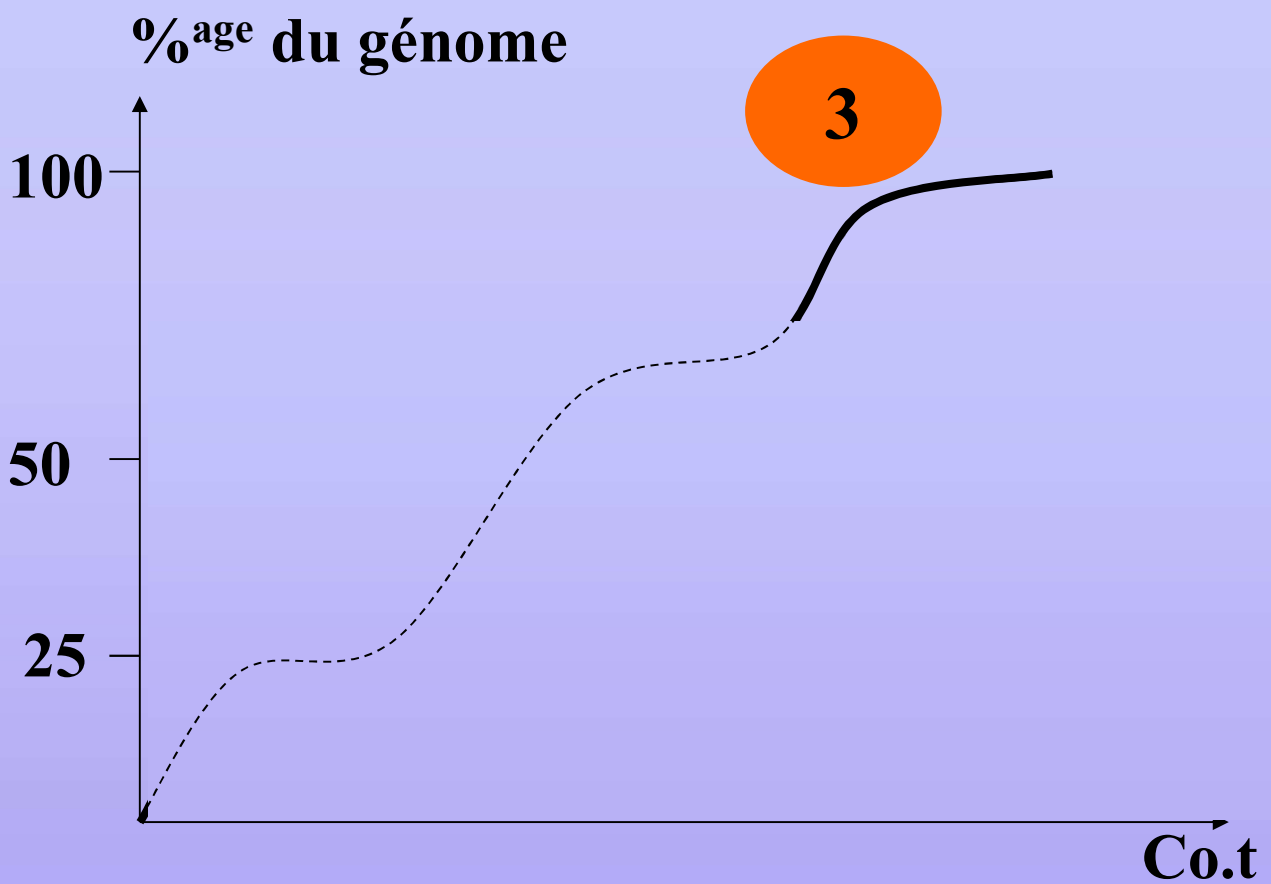
- **Co.t moyen : ré-association**

intermédiaire



**séquences moyennement
répétées**

30% du génome total



- Co.t élevé : ré-association

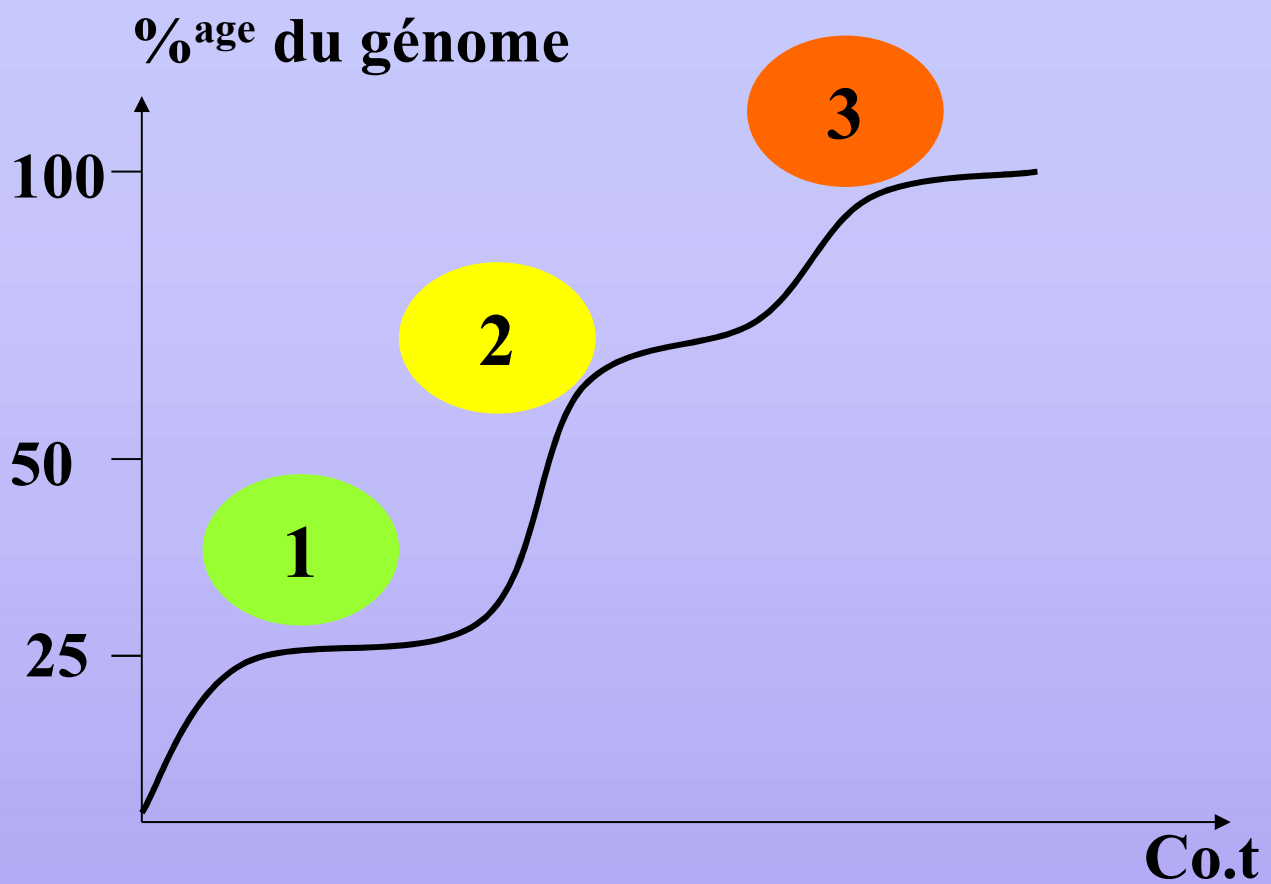
lente

séquences



uniques

45% du génome total



Portion 1 isolée par un plateau

→ nettement différente

Portions 2 et 3 : +/- fusionnées

A – Les sequences répétitives

**composées de séquences d'ADN
présentes en plusieurs exemplaires**

**Nombre de copies :
fréquence de répétition**



**2 groupes distincts par leur fréquence
de répétition**

- moyennement répétées**
- hautement répétées**

**1 – Séquences moyennement répétées
= séquences apparentées**

Souvent groupées en famille

en nombre variable :  X.10³

**nombre variable d'une espèce à
l'autre**

homogénéités +/- marquées

surtout pour les exons

Origine : phénomènes de duplication

variations au sein d'une famille de gène

 famille polymorphe

Gènes dérivant d'une séquence ancestrale:

 gènes homologues

même avec des divergences fortes 13

3 types de famille de gènes :

- famille multigénique simple =

Répétition de gènes en tandem dont l'expression répond à un même signal

- famille multigénique complexe =

Répétition de gènes transcrits séparément (réponse à des signaux différents)

- famille multigénique régulée =

Gènes aparentés dont l'expression varie au cours du temps et parfois selon les tissus

Dans une même famille :

→ perte de fonctionnalité de séquences par accumulation de mutations

→ Pseudogènes

α) Exemple de gènes de structure

**ensemble de gènes à fonction voisine ou
identique**

Expression à des temps différents



Famille multigénique régulée

Les globines :

a) Structure et expression des gènes

Protéines constituanes de l'hémoglobine

**Transporteur d'oxygène dans le règne
animal**

**Nombreux gènes fonctionnels codant les
globines**

→ structure identique

3 exons + 2 introns

Différents gènes dans une même espèce :

Même séquence origine ancestrale

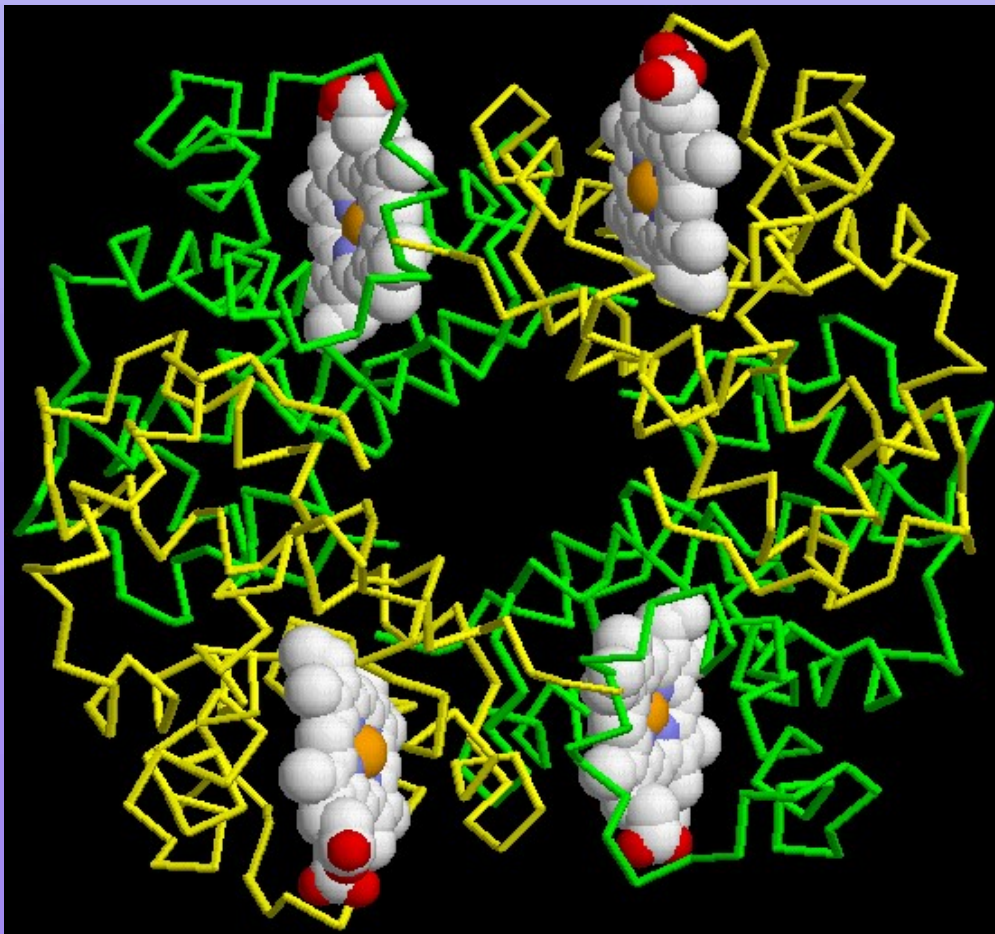


→ Modèles complexes

2 types de globines (α et β)

Polypeptides actifs

association 2 à 2 : $\alpha + \beta$



Différents gènes dans une même espèce :

Même séquence origine ancestrale



→ Modèles complexes

2 types de globines (α et β)

Polypeptides actifs

association 2 à 2 : $\alpha + \beta$

Polypeptides α et β

↪ 2 groupes de gènes

Chez *H. sapiens* : sur les chromosomes

respectivement 16 et 11

→ évolution indépendante

Groupe des globines α

compacté sur 28 kb

3 gènes fonctionnels :

α_1

α_2

ζ

+ 3 gènes non fonctionnels

$\Psi_{\alpha 1}$

$\Psi_{\alpha 2}$

Ψ_{ζ}

Groupe des globines β

réparti sur 65 kb

5 gènes fonctionnels

ϵ

G^γ

A^γ

β

+ 1 gène non fonctionnel

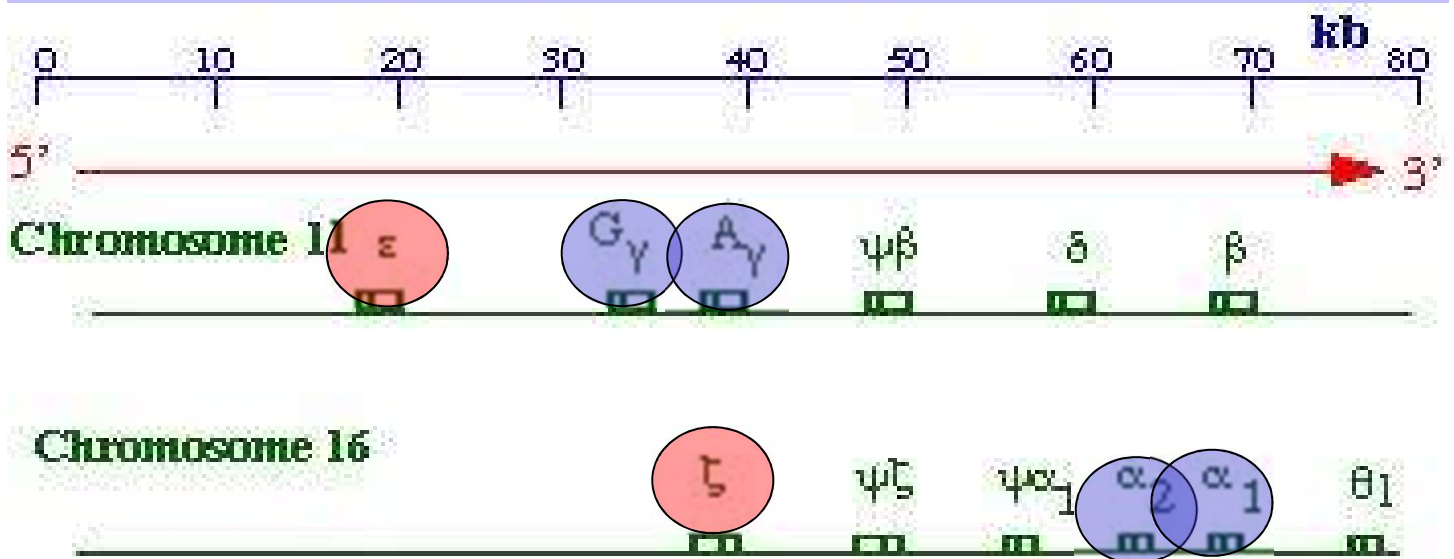
δ

Ψ_{β}

b) Répartition des gènes codant les polypeptides de globine

Ordre des gènes sur les chromosomes

= ordre d'expression au cours du temps



Vie embryonnaire: expression de

ζ

ϵ

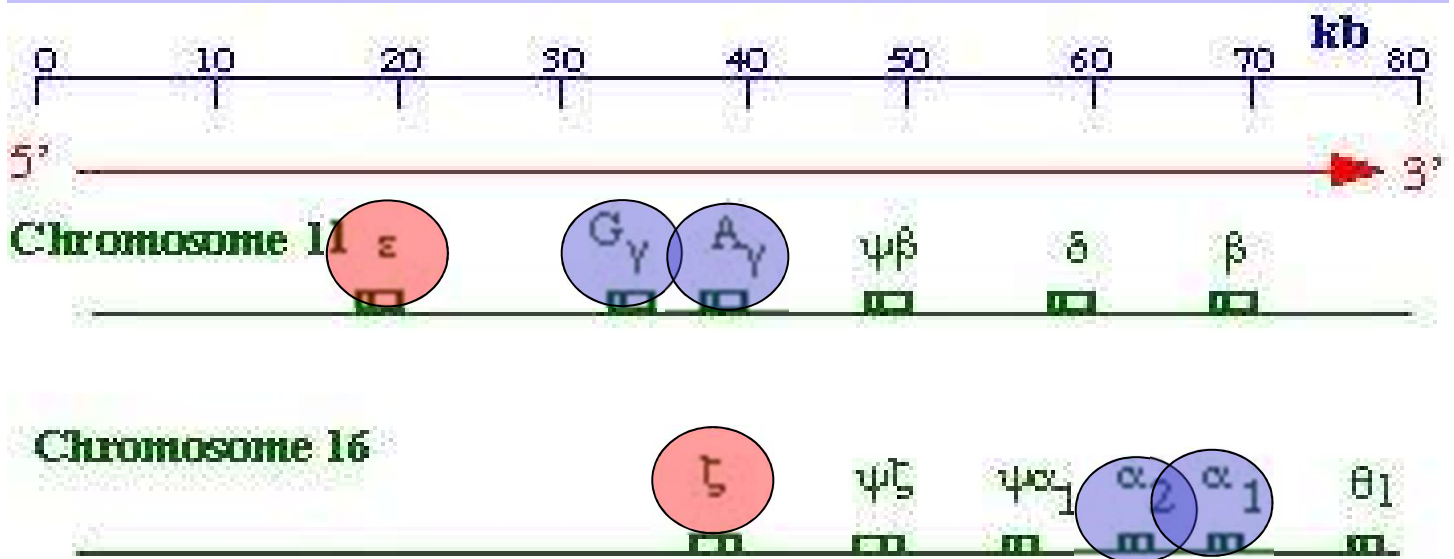
Après 10 semaines de gestation :

expression des gènes α et γ 19

b) Répartition des gènes codant les polypeptides de globine

Ordre des gènes sur les chromosomes

= ordre d'expression au cours du temps



α_1 et α_2 polypeptides identiques
exprimés en même temps

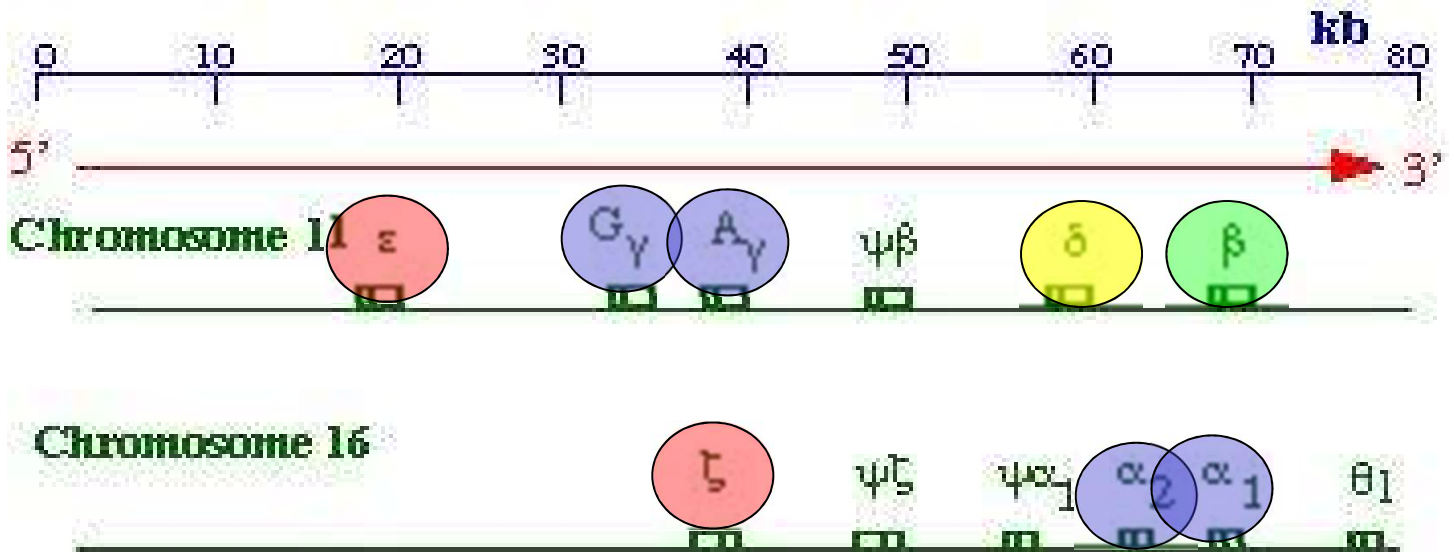
Mais

α_2 plus transcrit plus efficacement que α_1

↪ donne la majeure partie des chaînes α

Après la naissance

β remplace γ



A l'âge adulte : expression du gène δ

Mais

Faible portion des globines

Régulation temporelle associée à des séquences de régulation activatrices ou inhibitrices

c) Evolution des gènes

**2 groupes issus d'un gène ancestral
commun**

Origine du gène ancestral : 600 à 800 MA

**Groupes α et β
présents chez
tous les
vertébrés
Gnathostomes**



**Mais absents chez
les Agnathes**



duplication en α et β :

- postérieure à l'apparition des
vertébrés : 550 MA
- antérieure au développement des
gnathostomes : 450 MA

c) Evolution des gènes

**2 gènes issus d'un gène ancestral
commun**

Depuis 450 MA



2 groupes de gènes



divergences par série de mutations

Parfois acquisition de nouvelles fonctions

ou

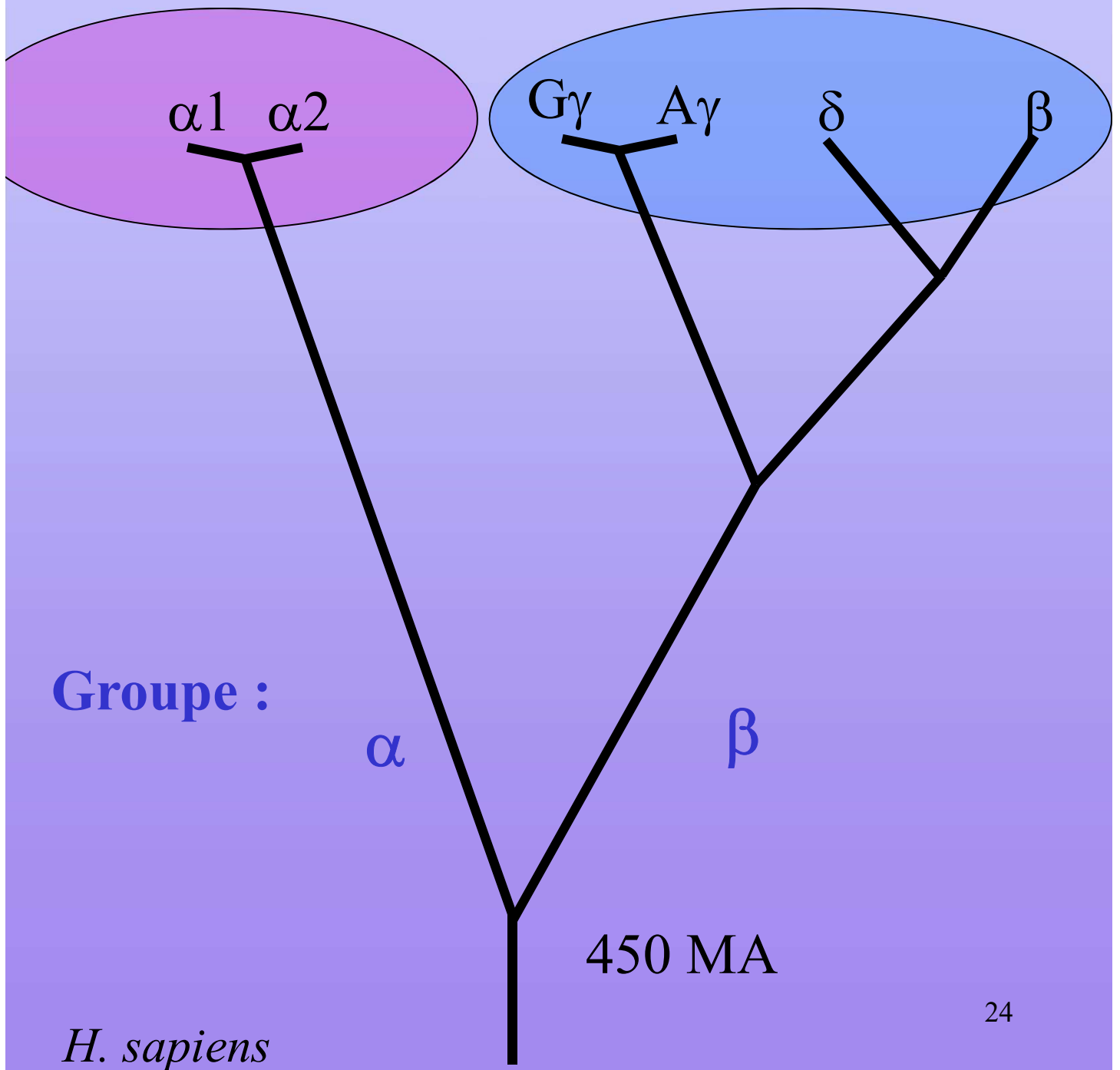
perte de fonction



pseudogènes

Histoire évolutive de la famille des globines

- comparaison d'espèces dont la date de divergence est connue
- estimation des phases évolutives des globines




Gènes α et β :

organisés en tandem chez

- les poissons**
- les batraciens**

Indépendants chez

- les oiseaux**
- les mammifères**

 **Duplications des gènes en tandem
variable selon le groupe systématique**



datation de la 1ere divergence

intra-groupe

Groupe des gènes β :

**differentiation du sous-groupe γ
+ modification des séquences de
régulation**



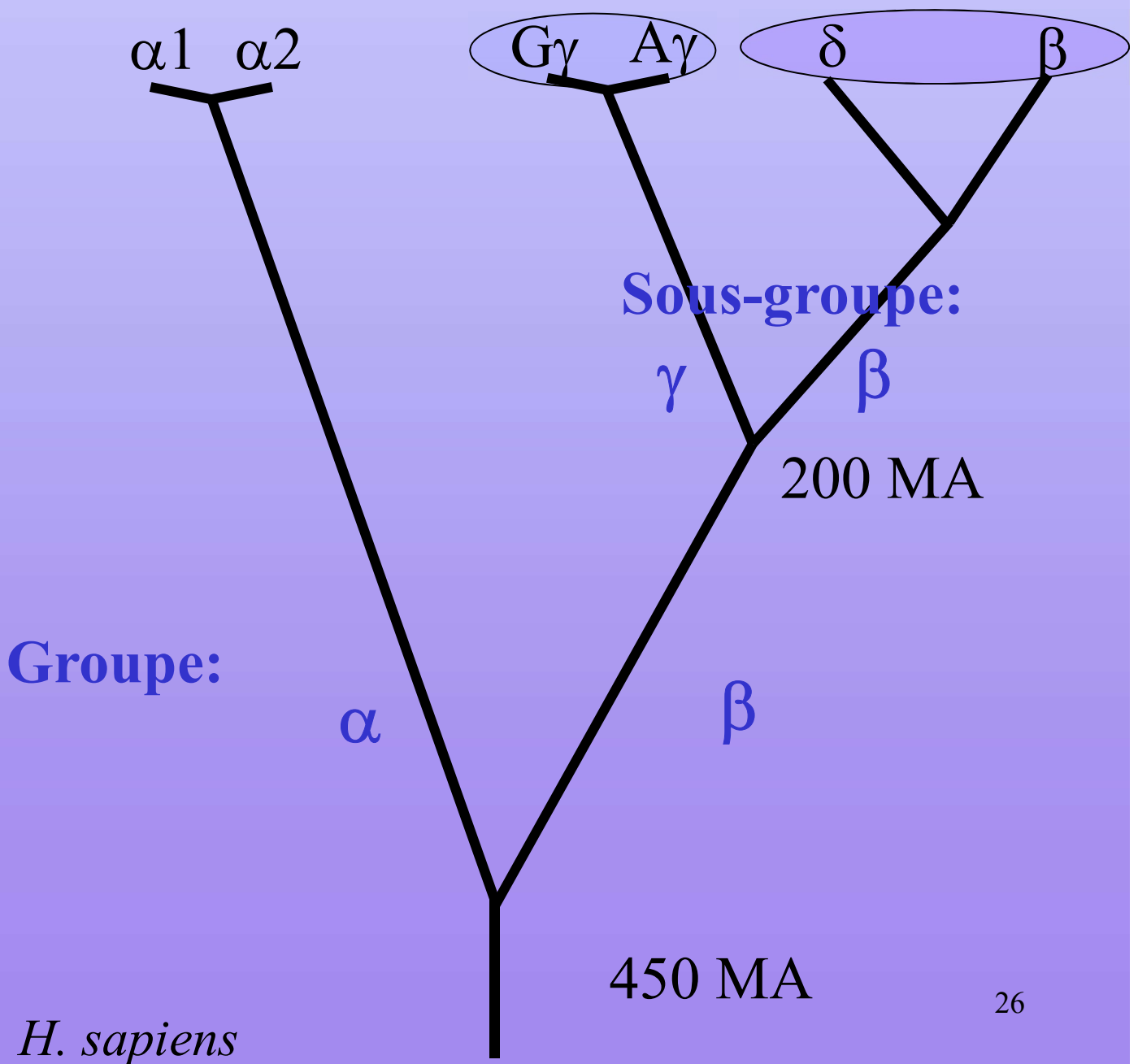
Expression différentielle

des gènes β et γ

Dans les taxons proches de l'*H. sapiens*

1^{ère} Duplication dans le groupe β

→ 2 sous-groupes de gènes : β et γ



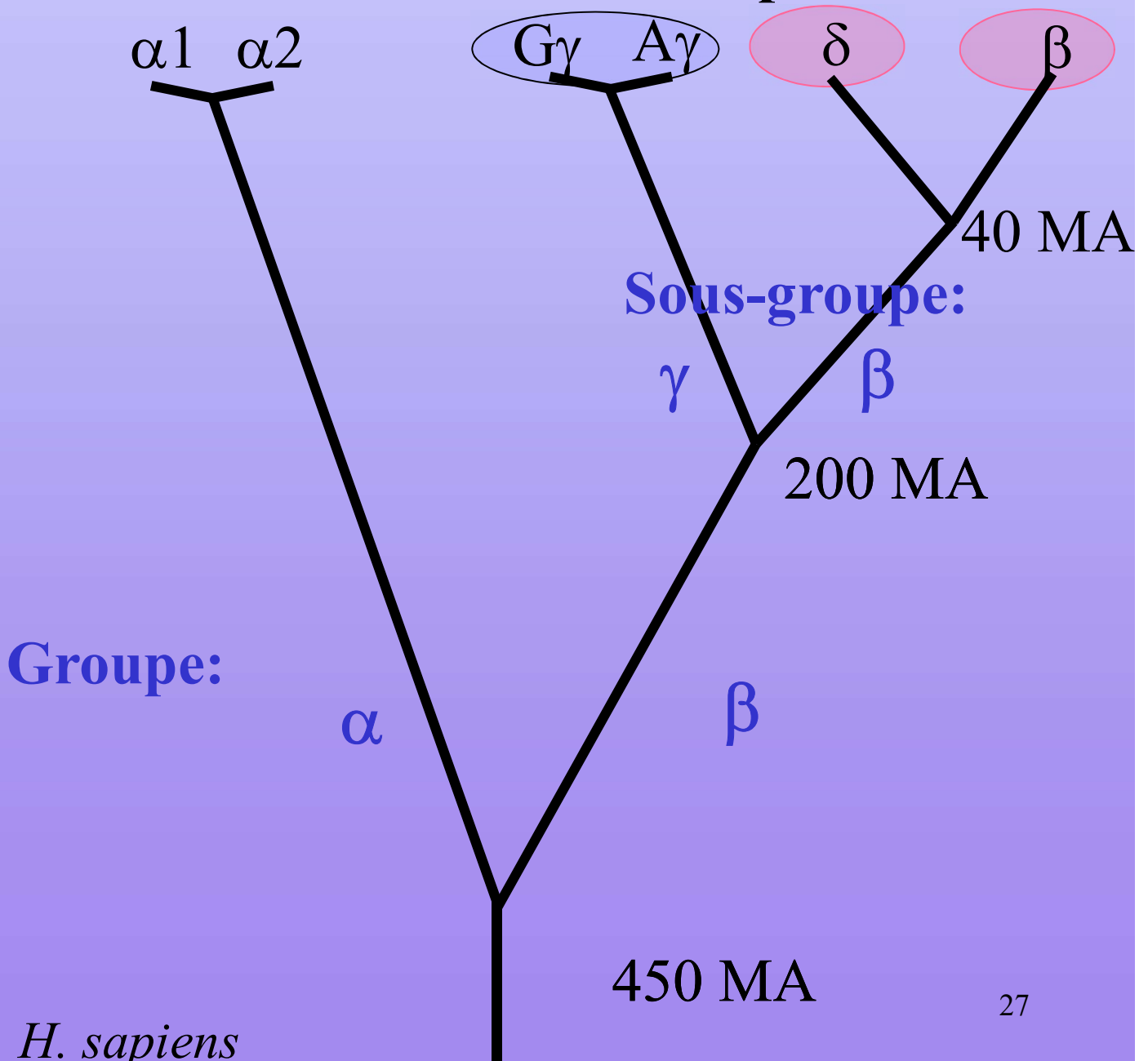
2eme duplication du groupe de gènes β chez les primates uniquement

→ entre 60 et 40 MA

→ 3 gènes : γ – δ – β

= arrangement de base chez tous les

primates

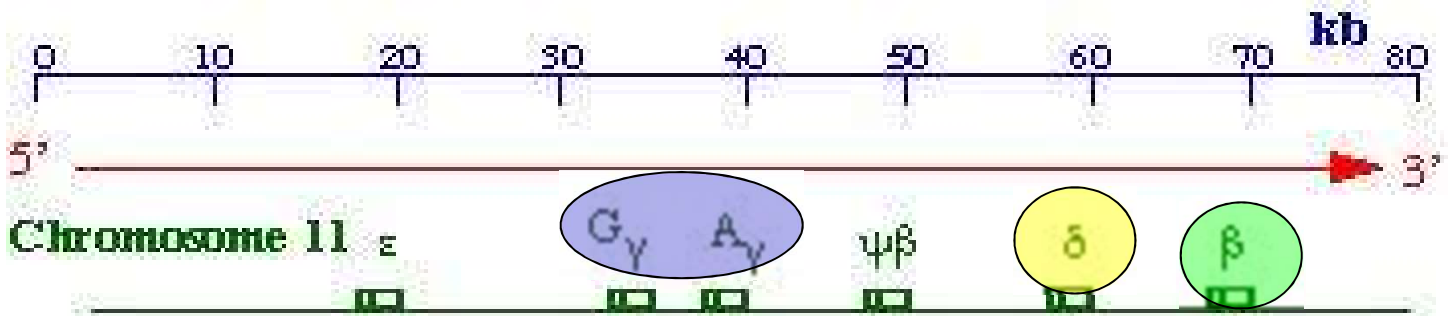


2eme duplication du groupe de gènes β chez les primates uniquement

→ entre 60 et 40 MA

→ 3 gènes : γ – δ – β

= arrangement de base chez tous les primates



3eme duplication du groupe de gènes β :

Gène γ

→ G^γ et A^γ

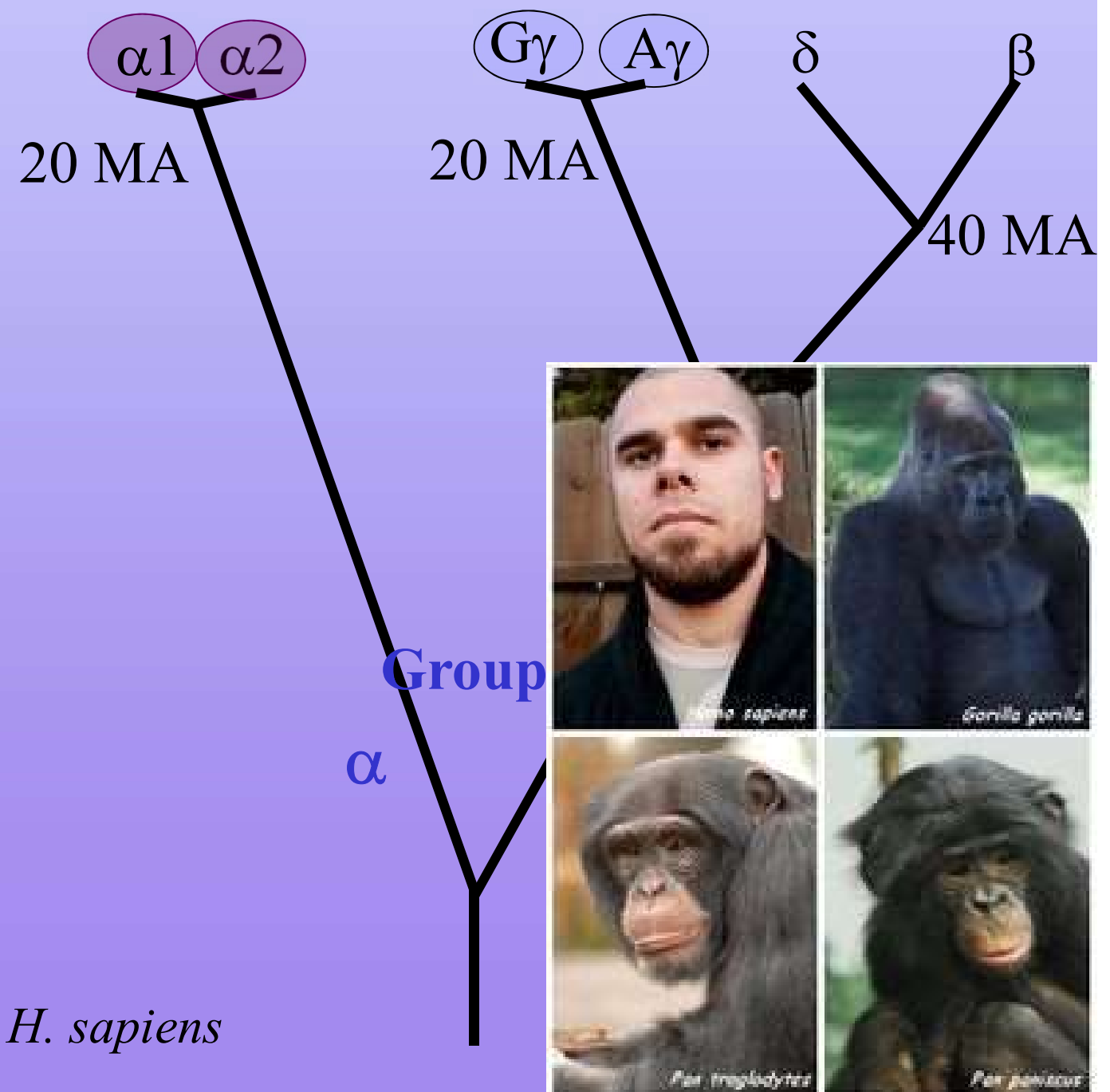
caractéristique des primates

de “Est-Atlantique”

20 MA

Histoire évolutive du groupe des globines α

Organisation génomique
caractéristiques des hominidés
➔ moins de 20 MA



β) Cas des histones : une famille multigénique régulée

a) Structure et expression des gènes

*** structure protéique très conservée**

peu de variations

associées à des différences d'expression

- au cours du développement

- dans différents tissus

*** structure nucléique : variations
interspécifiques**

**Organisation en gènes liés mais
transcription indépendante**

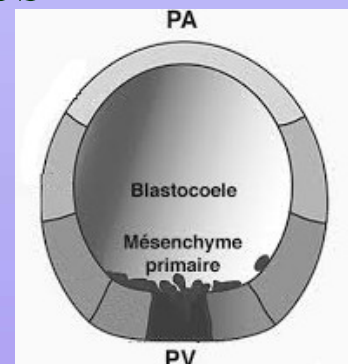
Exemple de l'oursin :

2 groupes de gènes

expression différente dans le temps

- 1er : avant la blastula

→ gènes précoces



- 2eme : après la blastula

→ gènes tardifs

Organisation différente selon le groupe



*** groupe précoce : 5 gènes sur le même
brin**

**lus dans le même sens
séparés par des espaceurs**



Répétition en tandem plusieurs

100^{aines} de fois

*** groupe tardif :**

- répétition quelques dizaines de fois

- groupes de gènes non organisés:

**parfois gènes liés et géographiquement
proches**

ou

parfois copies isolées

**Ordre des gènes et sens de transcription
indépendants**

Exemple de la drosophile :

pas de structure en deux groupes

Cependant

**Différence de taux de synthèse liée à des
phénomènes de régulation**

Toutes les copies : chromosome 2

liées en une seule unité

répétée environ 100 fois

Exemple de *S. cerevisiae*

2 copies d'un segment : $H_2A + H_2B$

+

2 copies d'un segment : $H_3 + H_4$

γ) Conséquences de la duplication des gènes

a) Influence sur les crossing-over

comparaison des génomes

mise en évidence du rôle de

- la duplication

- le réarrangement

comme facteurs d'évolution



Rôle aussi important que l'accumulation lente des mutations ponctuelles

Duplication : favorise

l'expansion/contraction

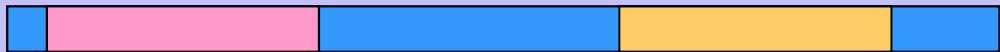
par crossing-over inégal



gain/perte d'un segment d'ADN

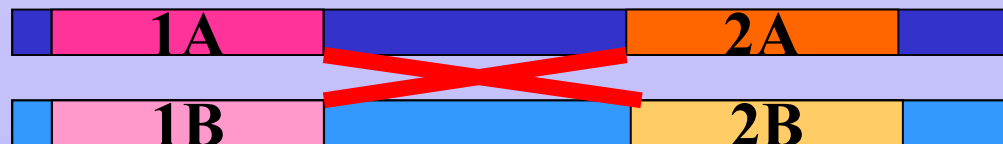
Crossing-over : échange de matériel génétique entre chromosomes homologues

Chromosome maternel A

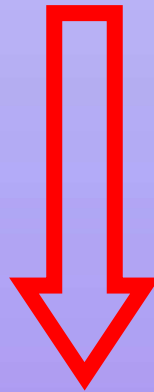


Chromosome paternel B

Crossing-over : échange de matériel génétique entre chromosomes homologues

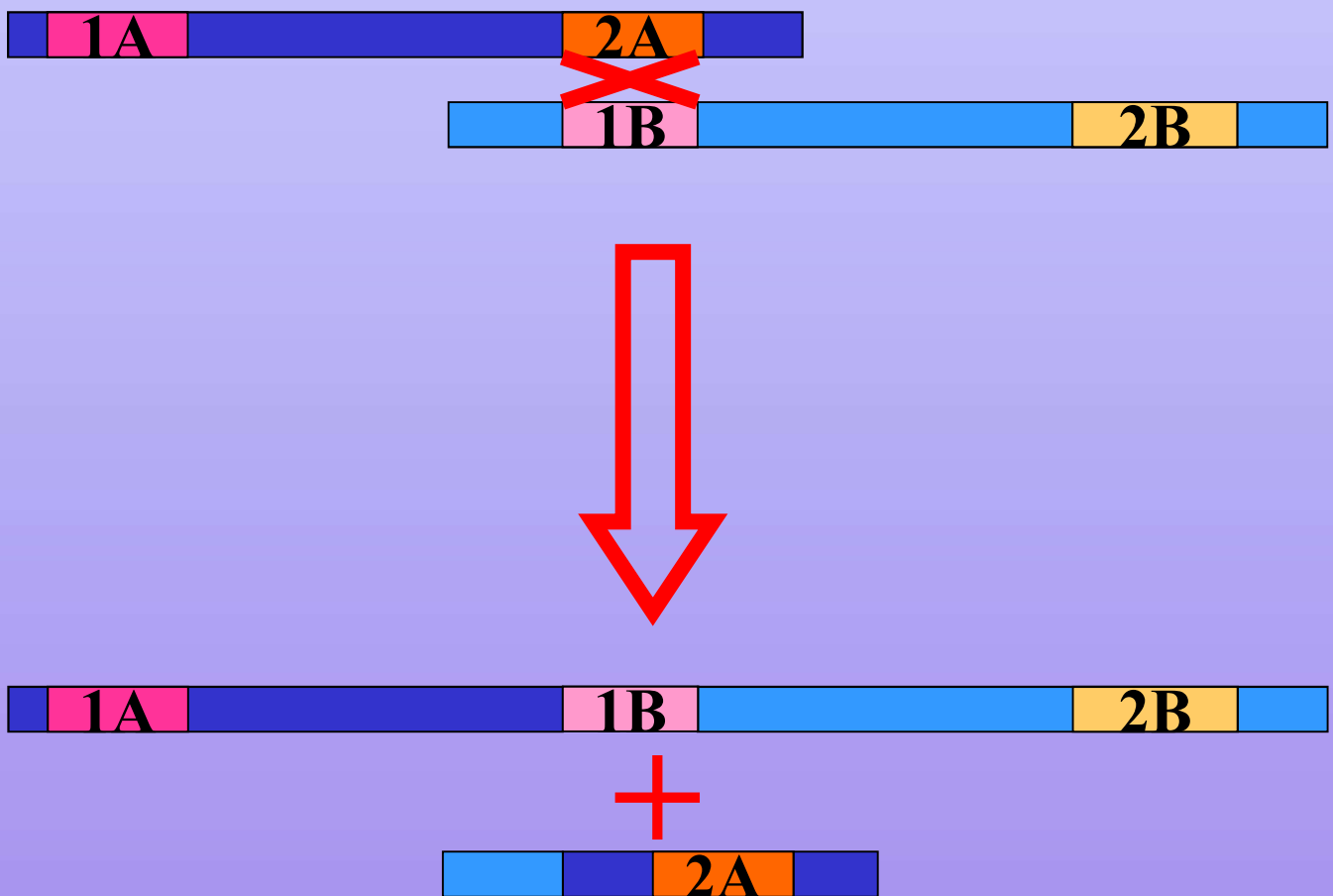


↑ **Crossing-over** ↑
Copie 1 **Copie 2**



Crossing-over inégal : régions flanquantes non appariées

↪ chromosomes à
recombinaison non réciproque



1 chromosome avec duplication

+

1 chromosome avec délétion

Pour 1 chromosome :
augmentation du nombre de copies
pour l'autre : perte d'une copie

Chez les eucaryotes:

présence des introns
→ divergence élevée

↪ limite l'homologie des
sequences géniques

↓
limite les crossing-over inégaux

Présence d'introns

↪ augmentation de la stabilité des
groupes de gènes

b) Les pseudogènes

**séquences proches des séquences de
gènes fonctionnels**

Cependant

→ Jamais traduit

**Parenté variable avec les gènes fonctionnels
de la même famille**



parfois structure générale

Exons/introns

parfois structure éloignée



identification de la famille

difficile

Inactivation par mutation à un stade précoce de l'expression

- - initiation de la transcription**
- excision-épissage**
- terminaison précoce**

pseudogènes = anciens gènes actifs



duplication

puis perte d'activité par mutation:



**Accumulation de mutations car absence de
pression de sélection**

**→ polymorphisme nucléotidique
délétion de paires de bases**

**Plusieurs pseudogènes dans une même
famille**

→ évolution indépendante

**Pseudogènes souvent présents à proximité
des gènes fonctionnels de la famille**

Cependant

**Pseudogènes hors de la famille
ex: cytochrome b**

Gène mitochondrial

dans la famille des Arvicolinae

**forme pseudogène dans le
génomme nucléaire**


17,6 % de divergence

4 indels




translocation : 6 MA



 **translocation : sortie d'un gène de sa famille**

**Souvent liée à une nouvelle copie
localisation différente dans le génome**

**Divergence gène actif (origine)
pseudogène (copie inactive)**

 **estimation de la période de la
duplication**

**Pseudogènes :
présents dans presque toutes les
familles de gènes**

2 – les séquences hautement répétées

Existence dans le génome de séquences répétées à hautes fréquences

= séquences hautement répétitives

→ = ADN satellite

Séquences très courtes :

de quelques bases (bp)

à quelques dizaines de bp



Grands éléments: plusieurs

centaines (milliers) de bp

**Long segment : multiples copies identiques
ou +/- apparentées**

**Souvent : composition nucléotidiques
particulière**

**distincte de la composition moyenne
du génome**

Séquences non fonctionnelles

Présents chez pratiquement tous les
eucaryotes supérieurs
en proportions très variables

Souvent localisés autour du
centromère


Hétérochromatine

Zones étroitement enroulées
inerte d'un point de vue fonctionnel

Localisation centromérique
→ fonction structurale

liée à la formation des
kinétochores

ADN satellite : peut représenter un pourcentage important du génome

 **pourcentage de génome non fonctionnel**

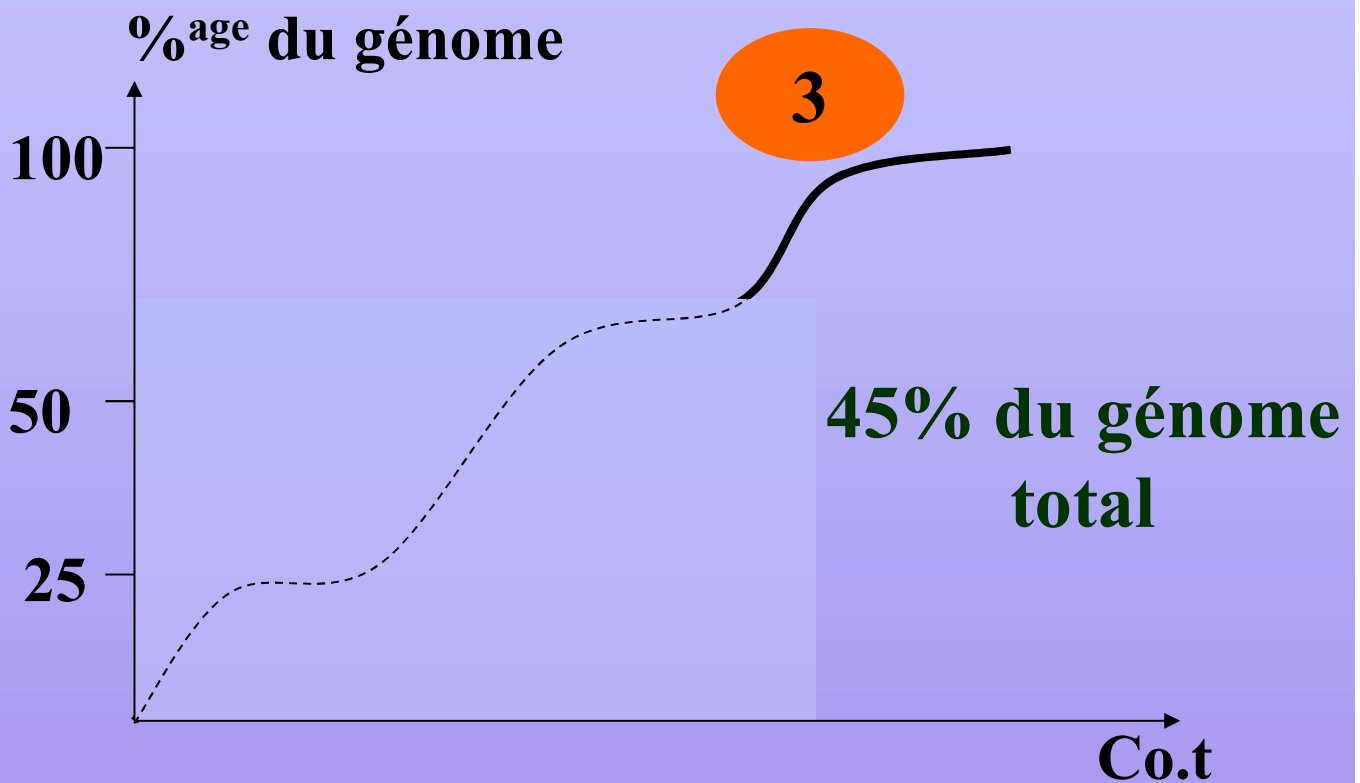
ex : Drosophile

3 types de séquences hautement répétées identifiées :

séquence	Séquence prédominante	Nb de copies / Génomes	% du génome
I	ACAAACT	$1,1 \cdot 10^7$	25%
II	ATAAACT	$3,6 \cdot 10^6$	8%
III	ACAAATT	$3,6 \cdot 10^6$	8%

B – Les sequences non répétitives

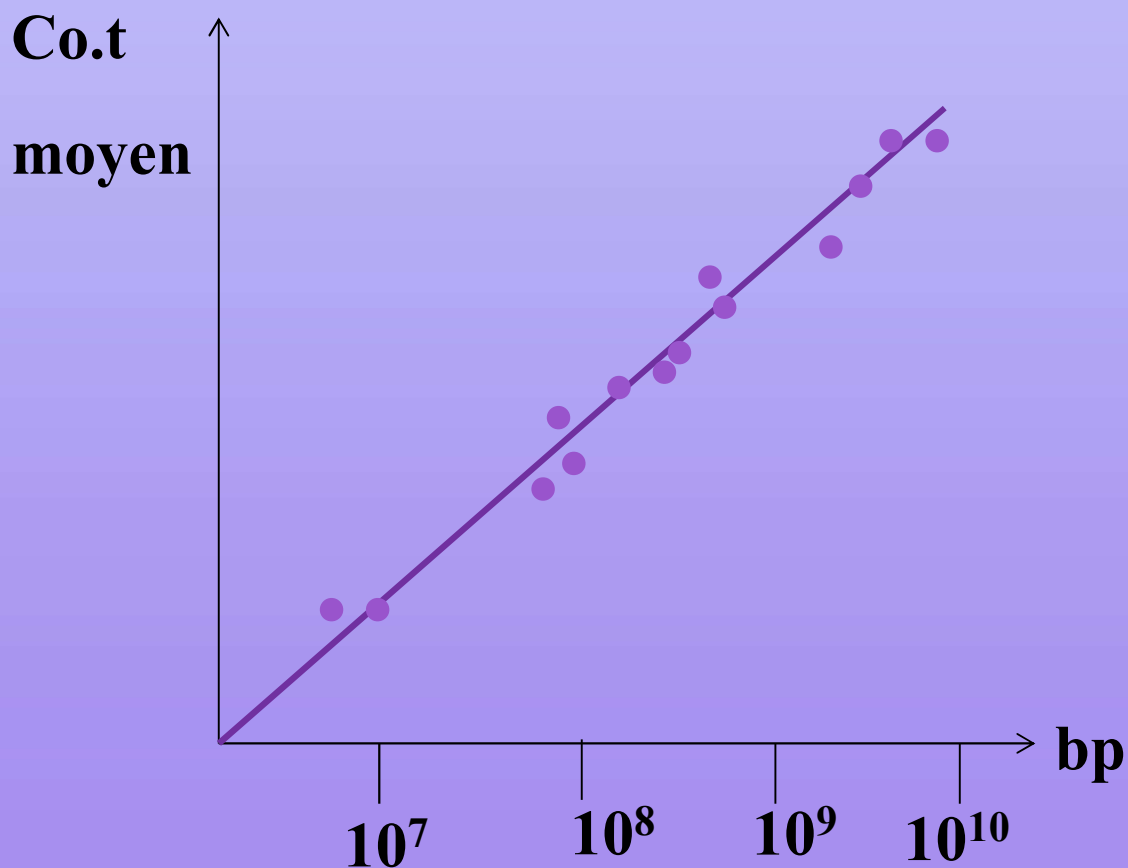
Généralement : composant principal du
gèneome des eucaryotes



Seul type d'ADN présent chez les
eubactéries

Séquences individuelles présentes en 1 seul exemplaire

**Analyse du génome : corrélation
contenu en ADN /
proportion de séquences non répétées**



Contenu en ADN du génome haploïde

Grands génomes



Séquences uniques nombreuses

Augmentation de taille

→ ne résulte pas seulement d'une
multiplication des séquences

Mais surtout

**Augmentation du nombre de gènes en
exemplaire unique**

Augmentation de taille



**Augmentation de la
complexité du génome**

QUESTION :

**Les génomes les plus grands contiennent-ils
- davantage de gènes différents**

ou

**- davantage de copies de mêmes
gènes ?**



**Différence de taille des génomes s'explique
par une plus grande diversité des séquences**



Séquences codantes

**Gènes codant : inclus dans l'ADN non
répétitif à 80%**