

4. Duplication de gènes et évolution des familles de gènes

- 4.1. Duplication et réassortiment de segments de gènes
- 4.2. Evolution de familles de gènes
- 4.3. Evolution concertée de familles de gènes

4. Duplication de gènes et évolution des familles de gènes

- **Duplication de gènes:** évènement fondamental en évolution moléculaire = source du "bricolage moléculaire"
 - Types de duplication:
 - duplication d'un **segment de gène**
 - duplication entière d'un **gène**
 - duplication d'une **partie de chromosome**
 - duplication entière d'un **chromosome**
 - duplication d'un **génomme entier** \Rightarrow polyploidie
- } évolution des génomes
- Cause de duplication: **crossing-over inégal**

4. Duplication de gènes et évolution des familles de gènes

- Crossing-over inégal:

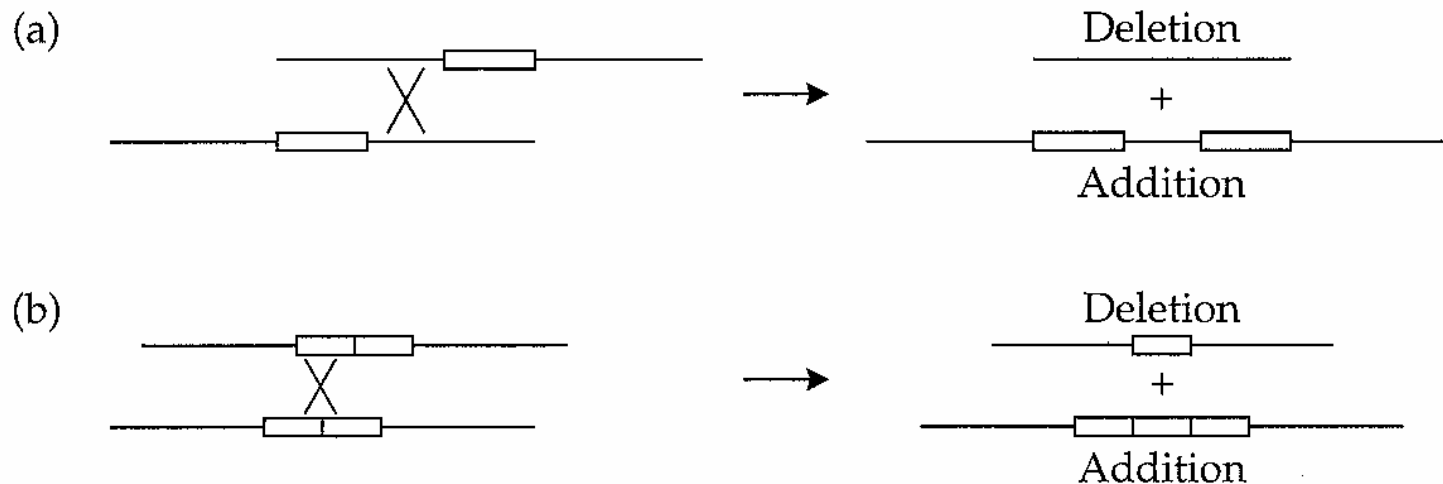


FIGURE 1.16 (a) Unequal crossing over resulting in the deletion of a DNA sequence in one of the daughter strands and the duplication of the same sequence in the other strand. (b) When a DNA segment is duplicated in tandem, the chance of misalignment increases, as does the chance of unequal crossing over. A box denotes a particular stretch of DNA. Modified from Li (1997).

⇒ Processus plus fréquent dans régions présentant des motifs répétés (ex. locus VNTR utilisés dans l'identification de criminels)

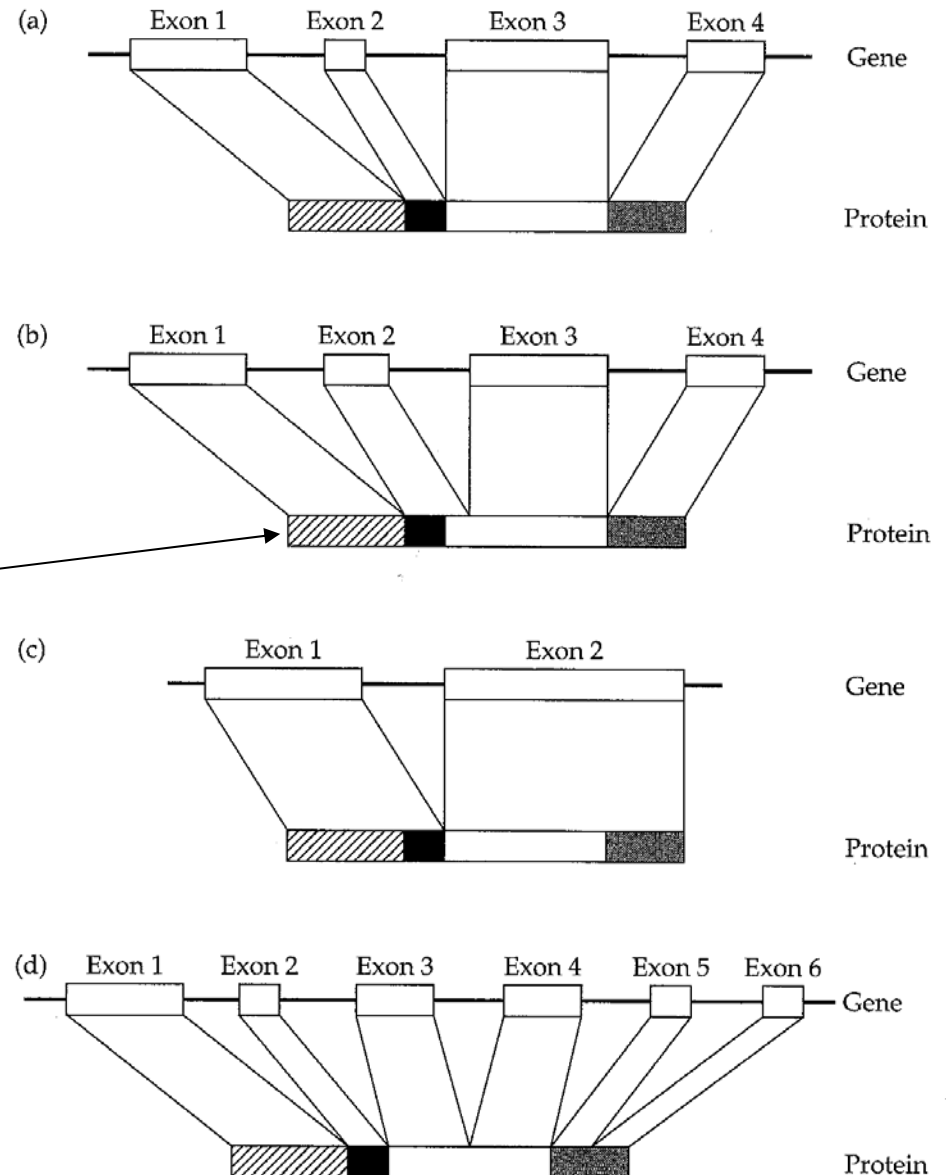
4.1. Duplication et réassortiment de segments de gène

- Domaines et exons
 - **Domaine**: région délimitée au sein d'une protéine qui effectue une fonction précise (ex. liaison au substrat ou à un cofacteur → appelé **domaine fonctionnel**) ou qui constitue une unité structurale compacte distincte des autres parties de la protéine (appelé **domaine structural** ou **module**)
 - **Exon**: partie de la séquence nucléotidique codante d'un gène non interrompue par des séquences non traduites (introns)
 - Très souvent bonne **correspondance** entre domaines et exons
 - Duplication d'un exon \Rightarrow augmentation du nombre de segments fonctionnels \Rightarrow évolution possible de la fonction du gène

4.1. Duplication et réassortiment de segments de gène

- Correspondance entre domaines et exons:

domaines



4.1. Duplication et réassortiment de segments de gène

- Evolution de la correspondance domaine-exon dans la famille des globines

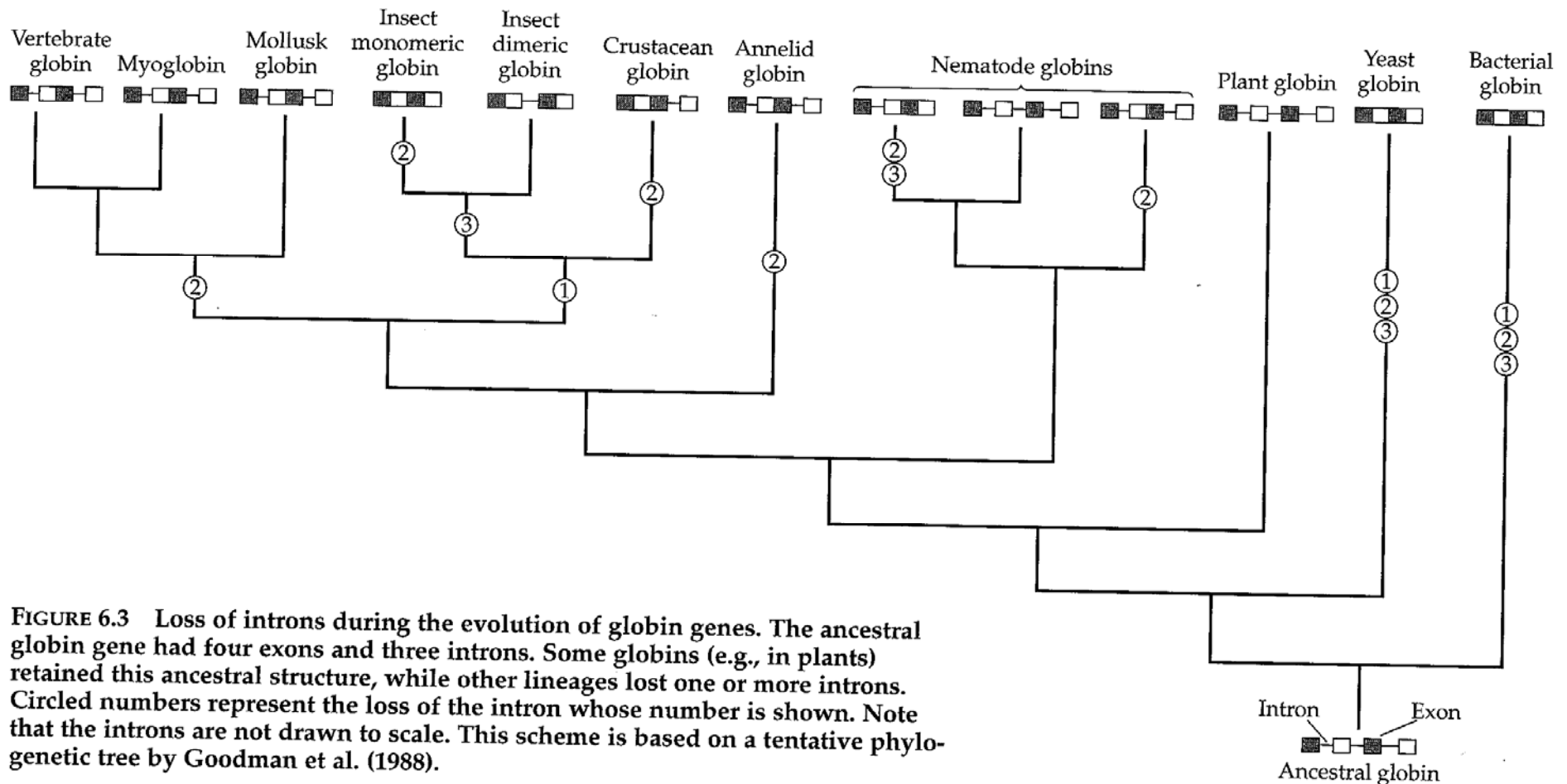


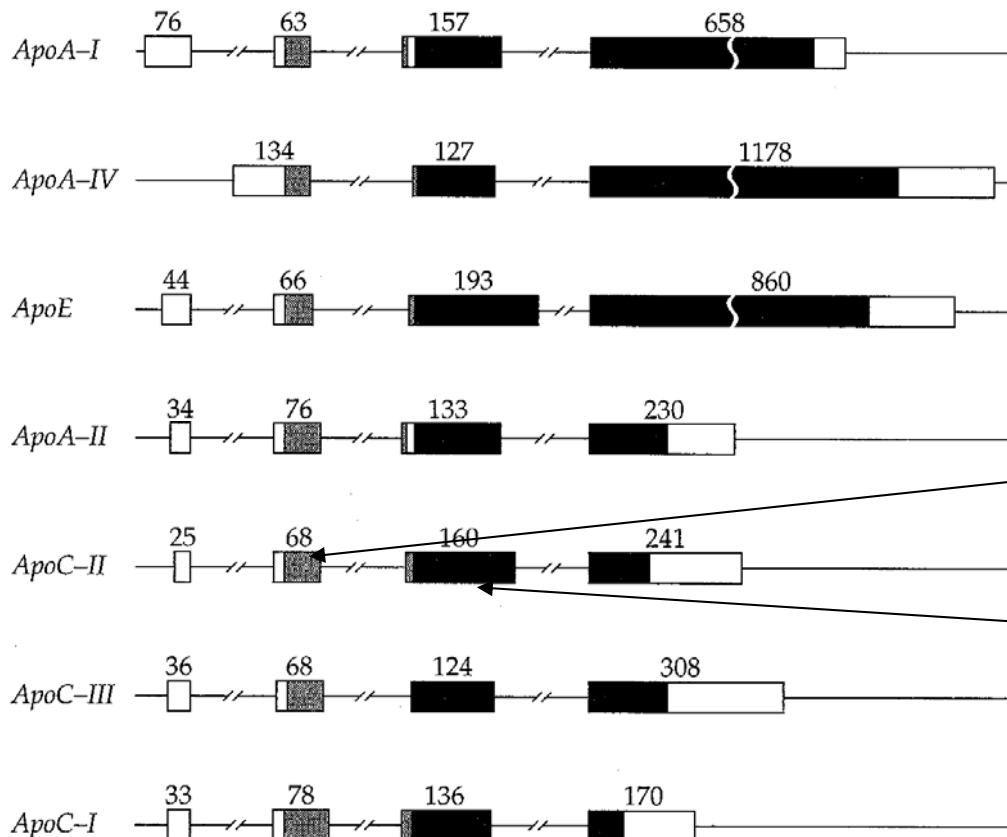
FIGURE 6.3 Loss of introns during the evolution of globin genes. The ancestral globin gene had four exons and three introns. Some globins (e.g., in plants) retained this ancestral structure, while other lineages lost one or more introns. Circled numbers represent the loss of the intron whose number is shown. Note that the introns are not drawn to scale. This scheme is based on a tentative phylogenetic tree by Goodman et al. (1988).

4.1. Duplication et réassortiment de segments de gène

- Duplication de domaines et élongation de gènes: exemple des **apolipoprotéines**
 - Apolipoprotéines = composants protéiniques des complexes lipoprotéiques du plasma
 - rôle des complexes lipoprotéiques: transport du cholestérol, des phospholipides et des triglycérides dans le sang
 - rôle des apolipoprotéines dans ces complexes: liaison aux lipides et solubilisation des complexes lipoprotéiques
 - \exists 9 types différents d'apolipoprotéines dont 7 \in une même famille de gènes

4.1. Duplication et réassortiment de segments de gène

- Duplication de domaines et élongation de gènes: exemple des **apolipoprotéines**
 - Structure des gènes de la famille des apolipoprotéines



- 3 introns sauf *ApoA-IV* (2 introns)
- très grande variabilité dans la **longueur de l'exon 4**

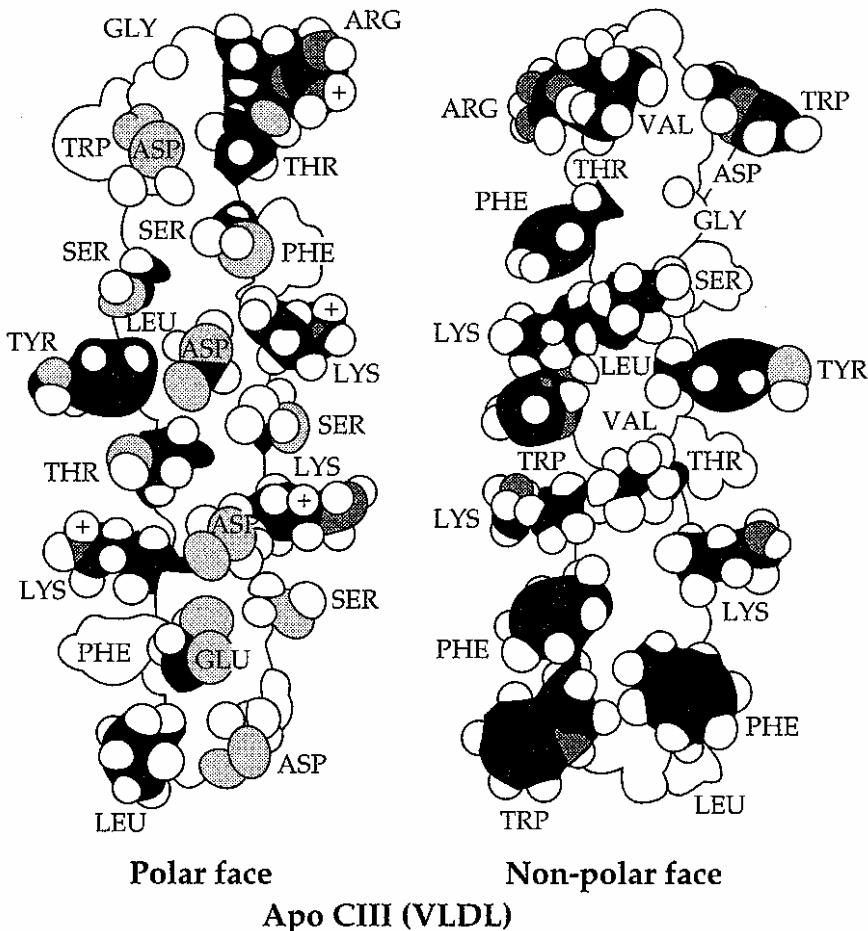
peptide de signal

protéine finale

4.1. Duplication et réassortiment de segments de gène

- Exemple des **apolipoprotéines**

- Découverte d'un grand nombre de répétitions d'un motif de base de 22 a.a.



- Structure secondaire de type **hélice α**
à 2 faces distinctes

- face **polaire**: ac. aspartique, ac. glutamique (-); lysine, arginine (+) \Rightarrow interface avec phase aqueuse; ségrégation des résidus (+) à la périphérie et des résidus (-) au centre de la face polaire

- face **hydrophobe**: leucine, tryptophane, phénylalanine, tyrosine \Rightarrow se glisse entre les chaînes hydrophobes des lipides cibles

4.1. Duplication et réassortiment de segments de gène

• Exemple des apolipoprotéines

– Découverte d'un grand nombre de répétitions d'un motif de base de 22 a.a.

A-I	44	L	K	L	L	D	N	W	D	S	V	T	S	T	F	S	K	L	R	E	Q	L	G	65
	66	P	V	T	Q	E	F	W	D	N	L	E	K	E	T	E	G	L	R	Q	E	M	S	87
	88											K	D	L	E	E	V	K	A	K	V	Q	98	
	99	P	Y	L	D	D	F	Q	K	K	W	Q	E	E	M	E	L	Y	R	Q	K	V	E	120
	121	P	L	R	A	E	F	Q	E	G	A	R	Q	K	L	H	E	L	Q	E	K	L	S	142
	143	P	L	G	E	E	M	R	D	R	A	R	A	H	V	D	A	L	R	T	H	L	A	164
	165	P	Y	S	D	E	L	R	Q	R	L	A	A	R	L	E	A	L	K	E	N	G	G	186
	187	A	R	L	A	E	Y	H	A	K	A	T	E	H	L	S	T	L	S	E	K	A	K	208
	209	P	A	L	E	D	L	R	Q	G	L	L												219
	220	P	V	L	E	S	F	K	V	S	F	N	S	A	L	E	E	Y	T	K	K	L	N	241
A-IV	40	A	L	F	Q	D	K	L	G	V	F	L	S	A	L	E	E	Y	T	K	K	L	N	61
	62	P	F	A	T	E	L	H	E	R	L	A	K	D	S	E	K	L	K	E	E	I	V	83
	84											K	E	L	E	E	L	R	A	R	L	L		94
	95	P	H	A	N	E	V	S	Q	K	I	G	D	N	L	R	E	L	Q	Q	R	L	E	116
	117	P	Y	A	D	Q	L	R	T	Q	V	N	T	Q	A	E	Q	L	R	Q	Q	L	T	138
	139	P	Y	A	Q	R	M	E	R	V	L	R	E	N	A	D	S	L	Q	A	S	L	R	160
	161	P	H	A	D	E	L	K	A	K	I	D	Q	N	V	E	E	L	K	G	R	L	T	182
	183	P	Y	A	D	E	F	K	V	K	I	D	Q	T	V	E	E	L	R	R	S	L	A	204
	205	P	Y	A	Q	D	T	Q	E	K	I	L	N	H	Q	L	E	G	L	T	F	Q	M	226
	227	K	N	A	E	E	L	V	K	R	I	S	A	S	A	E	E	G	L	R	Q	S	L	248
	249	P	L	A	E	D	V	R	G	N	L	R	G	N	T	E	E	G	L	R	S	L	A	269
	270			A	E	L	G	G	H	L	D	Q	Q	V	E	E	F	R	R	R	V	E		288
	289	P	Y	G	E	N	F	N	K	A	L	V	Q	Q	M	E	Q	L	R	T	K	L	G	310
	311	P	H	A	G	D	V	E	G	H	L	S	F	L	E	K	D	L	R	D	K	V	N	332
E	62	A	L	M	D	E	T	M	K	E	L	K	A	Y	K	S	E	L	E	E	Q	L	T	83
	84	P	V	A	E	E	T	R	A	R	L	S	K	E	L	Q	A	A	Q	A	R	L	G	105
	106											A	D	M	E	D	V	C	C	G	R	L	V	116
	117	Q	Y	R	G	E	V	Q	A	M	L	G	Q	S	T	E	E	L	R	V	R	L	A	138
	139	S	H	L	R	K	L	R	K	R	L	L	R	D	A	D	D	L	Q	K	R	L	A	160
	161	V	Y	Q	A	G	A	R	E	G	A	E	R	G	L	S	A	I	R	E	R	L	G	182
	183	P	L	V	E	Q	G	R	V	R	A	A	T	V	G	S	L	A	G	Q			201	
	202	P	L	Q	E	E						A	Q	A	W	G	R	L	A	R			217	
	218		M	E	E	M	G	S	R	T	R	D	R	L	D	E	V	K	E	Q	V	A	237	
	241	A	K	L	E	E	Q	A	Q	Q	I	R	L	Q	A	E	A	F	Q	A	R	L	K	262
	267	P	L	V	E	D	M	Q	R	Q	W	A	G	L	V	E	K	V	Q	A	A	V	G	288
A-II	40											S	Y	F	E	K	S	K	E	Q	L	T	50	
	51	P	L	I	K	K	A	G	T	E	L	V	N	F	L	S	Y	F	V	E	L	G	71	

-P: proline

-L: leucine; Y: tyrosine; V: valine
(hydrophobes)

-D: ac. asp.; E: ac. glut. (polaires)

-K: lysine; R: arginine (polaires)

⇒ Très grande variabilité dans la longueur de l'exon 4 est due à variation du nombre de motifs répétés de 22 a.a.

⇒ permet adaptation de la spécificité à ≠ molécules lipidiques

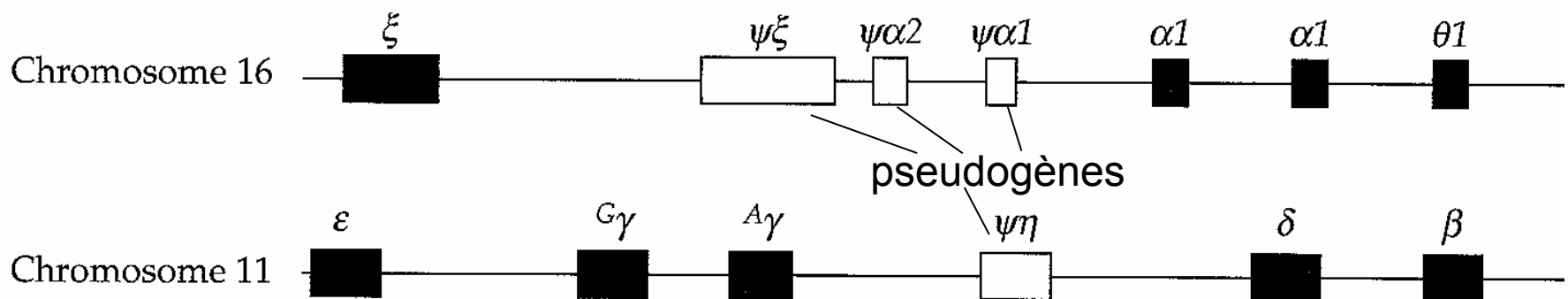
4.2. Evolution de familles de gènes

- Evolution de gènes dupliqués en entier:
 - conservation de la fonction originale \Rightarrow copies invariantes: production en masse de certains ARN et protéines
 - émergence d'une nouveauté génétique: une copie conserve sa fonction initiale, l'autre acquiert une nouvelle fonction
 - perte de la fonction originale et dégradation de la séquence (codon stop, perte TATA box, décalage de cadre de lecture, ...): évolution en pseudogène

4.2. Evolution de familles de gènes

- Evolution de gènes dupliqués en entier:
 - L'ensemble des gènes appartenant à un même groupe de séquences ayant évolué par le processus de duplication = **famille de gènes** (incluant copies fonctionnelles et non fonctionnelles)
 - Les membres d'une famille de gènes se trouvent généralement à proximité l'un de l'autre sur le même chromosome, en raison de leur évolution par crossing-over inégal

Exemple: les familles de gènes de l' α et de la β -globine chez l'homme



4.2. Evolution de familles de gènes

- Conservation de la fonction originale après duplication :
 - "répétitions de dose" favorisées quand justifié par un besoin métabolique intense pour une molécule donnée:
 - ARN ribosomiaux, ARN de transfert → répétés en **tandem**
 - protéines ubiquistes comme les histones (lors de divisions cellulaires)

TABLE 10.2 Numbers of rRNA and tRNA genes per haploid genome in various organisms

Genome source	Number of rRNA genes^a	Number of tRNA genes	Approximate genome size (bp)
Human mitochondrion	1	22	1.7×10^4
<i>Mycoplasma genitalium</i>	2	33	5.8×10^5
<i>Escherichia coli</i>	7	~100	4×10^6
<i>Saccharomyces cerevisiae</i>	~140	320–400	1.3×10^7
<i>Tetrahymena thermophila</i>	1	ND ^b	2×10^8
<i>Drosophila melanogaster</i>	130–250	~750	2×10^8
Human	~300	~1,300	3×10^9
<i>Xenopus laevis</i>	400–600	~7,800	8×10^9

4.2. Evolution de familles de gènes

Emergence d'une nouveauté génétique après duplication: une copie conserve sa **fonction initiale**, l'autre diverge et acquiert une **nouvelle fonction**

- Exemple: les protéines **opsines** des **pigments rétiniens** responsables de la vision des couleurs

- Chez l'homme + singes de l'ancien monde → ∃ 3 opsines différentes:

- opsine "bleue" (iodopsine S) située sur le chromosome 7 (=autosome)
- opsines "rouge" et "verte" (iodopsine M et L) situées sur le chrom. X

⇒ vision **trichromatique**

Apparus par 2 évènements de duplication:

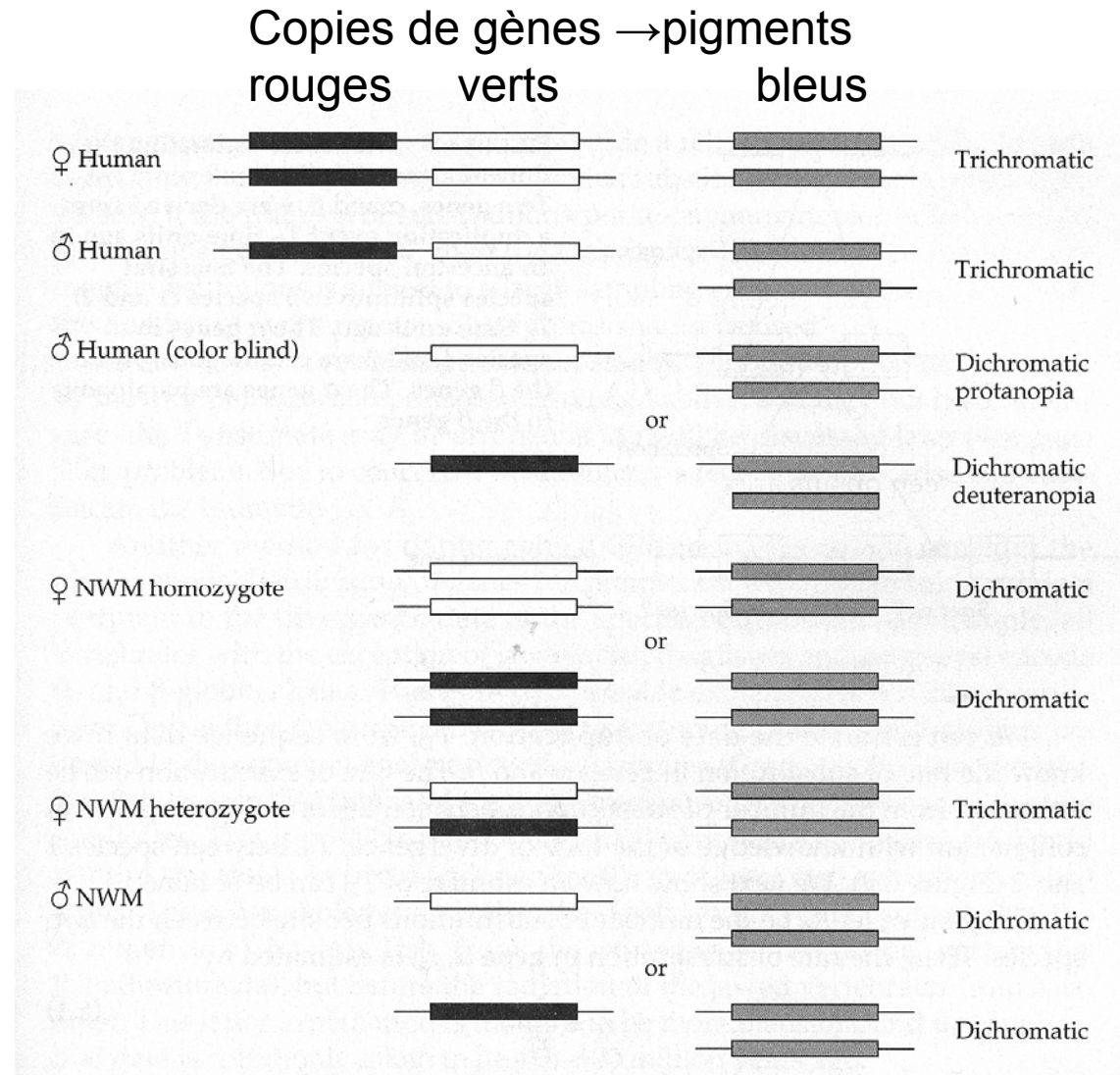
- 600 Ma: séparation entre iodopsine S et iodopsine "ML" (43% identité séq. a.a.)
⇒ vision **dichromatique** après divergence de S et "ML"
- 35 Ma: séparation entre iodopsine M et L (96% identité séq. a.a.)
⇒ vision **trichromatique** après divergence de M et L

4.2. Evolution de familles de gènes

- Exemple: les protéines **opsines** des **pigments rétiniens** responsables de la vision des couleurs
 - Homme + singes de l'ancien monde $\rightarrow \exists 3$ opsines $\neq \Leftrightarrow$ vision **trichromatique** :
 - opsine "bleue" (iodopsine S) située sur le chromosome 7 (=autosome)
 - opsines "rouge" et "verte" (iodopsine M et L) situées sur le chrom. X
 - Troubles génétiques de vision des couleurs (daltonisme): altération gène M ou L
 - individu ♀: si 1 seule copie porte allèle délétère (hétéroz.) \Leftrightarrow vision normale
 - individu ♀: caractère exprimé seulement si homozygote pour allèle délétère
 - individu ♂: caractère toujours exprimé car une seule copie du chrom. X
 - Singes du Nouveau Monde : une seule iodopsine (ML) sur le chrom. X
 - vision dichromatique
 - si allèles "M" et "L" chez ♀ hétérozygotes \Leftrightarrow trichromatiques;
♀ homozygotes + ♂ \Leftrightarrow dichromatiques

4.2. Evolution de familles de gènes

- Acquisition d'une nouvelle fonction après duplication:
 - Ex.: vision des couleurs liée à coexistence de 3 types de gènes de pigmentation de la rétine, issus de **duplication**.



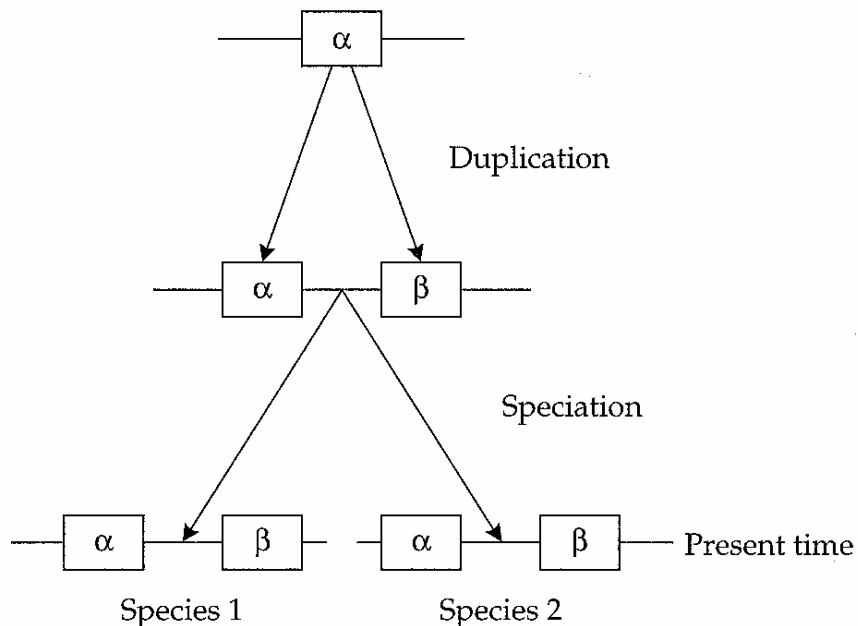
4.2. Evolution de familles de gènes

- Perte de la fonction originale après duplication et dégradation de la séquence:
 - une des 2 copies conserve la fonction originale, l'autre devient non-fonctionnelle (mutation délétère) \Rightarrow évolution en **pseudogène** (gène A \Rightarrow pseudogène noté ΨA)
 - mécanisme de perte de fonction: codon stop, perte TATA box, décalage de cadre de lecture, perte sites d'épissage,...
 - forte probabilité d'évolution en pseudogène après duplication (la première des 2 copies qui mute \Rightarrow pseudogène; l'autre devient un gène à copie unique) \Rightarrow pseudogènes très fréquents (\exists presque toutes familles de gènes)

4.2. Evolution de familles de gènes

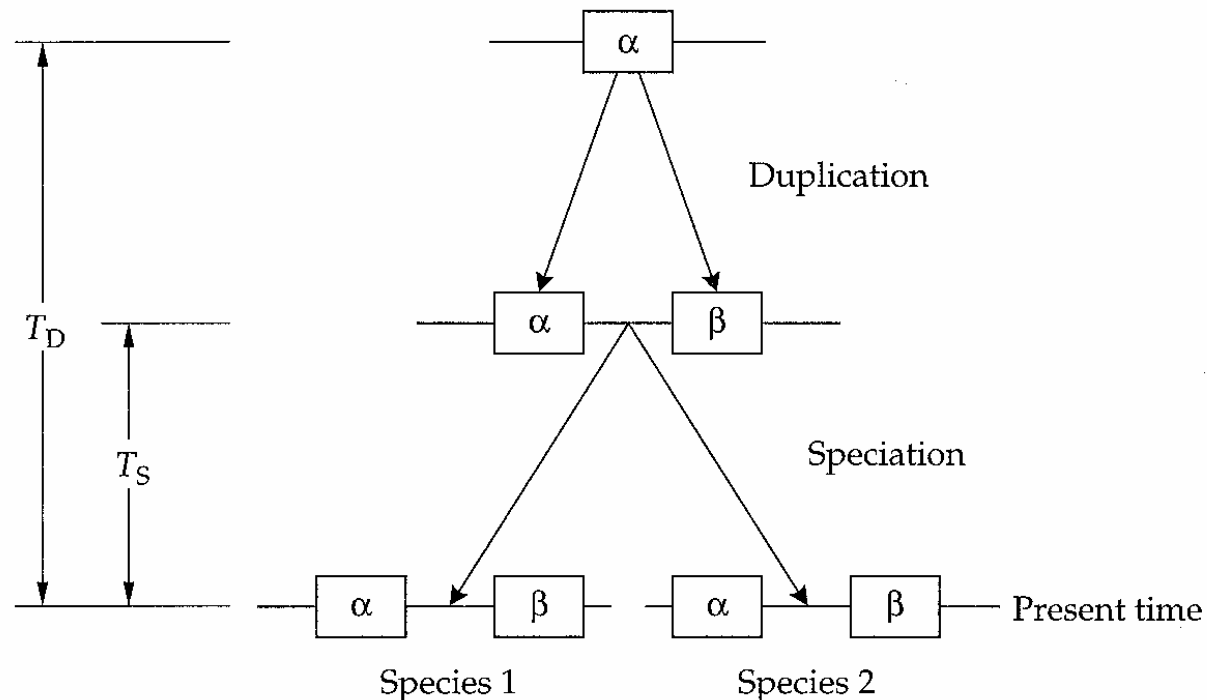
- Datation d'un évènement de duplication:
 - 2 gènes sont dits **paralogues** s'ils dérivent d'un évènement de **duplication** (\in 2 locus distincts dans même génome: α et β)
 - 2 gènes sont dits **orthologues** s'ils dérivent d'un évènement de **spéciation** (\in même locus dans 2 génomes distincts: α espèce 1 et α espèce 2)

terme "homologue"
imprécis!



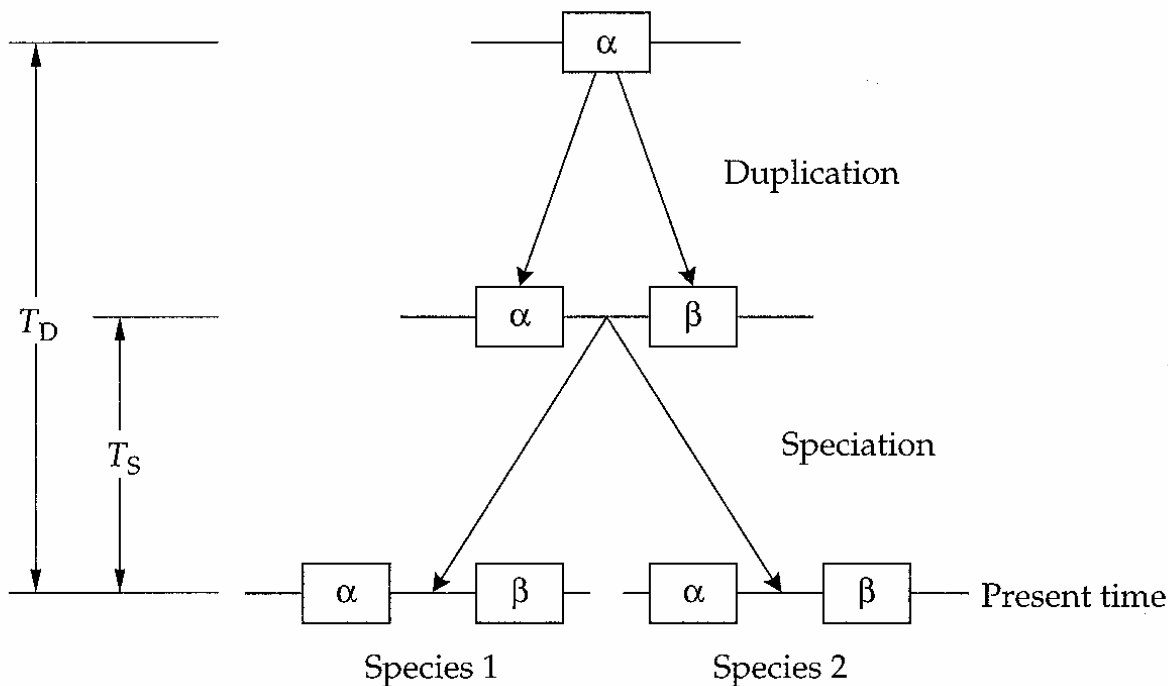
4.2. Evolution de familles de gènes

- Datation d'un évènement de duplication:
 - on veut estimer T_D = temps écoulé depuis duplication menant à α et β (2 gènes paralogues)
 - on connaît T_S = temps écoulé depuis spéciation menant aux espèces 1 et 2; et les séquences des 2 paires de copies orthologues de α et β



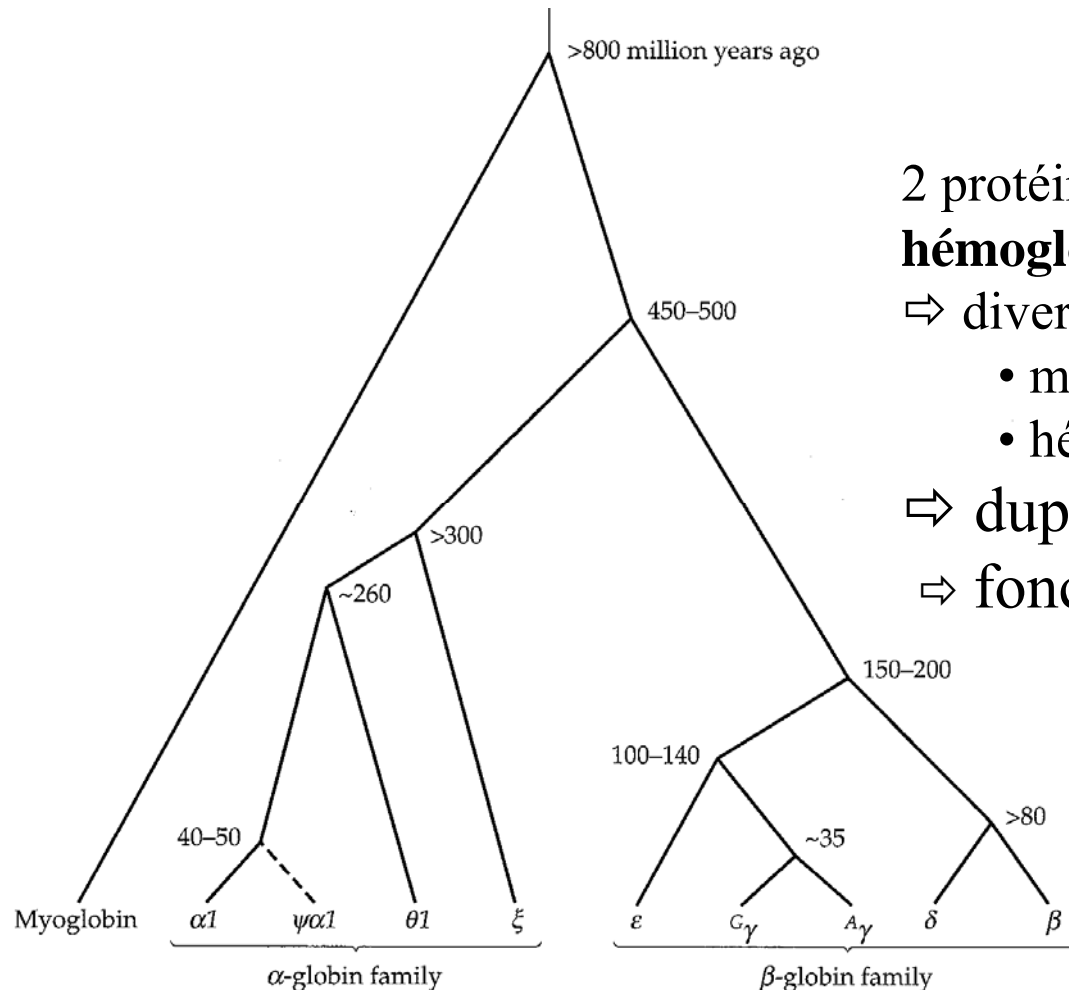
4.2. Evolution de familles de gènes

- Datation d'un évènement de duplication:
 - estimation taux substitution moyen: $r_{\alpha} = K_{\alpha} / (2T_S)$ avec K_{α} entre 2 espèces;
 $r_{\text{moy}} = (r_{\alpha} + r_{\beta}) / 2 \Rightarrow$ hypothèse que même taux substitution entre α et β
 - estimation nombre substitutions entre α et β : $K_{\alpha\beta} =$ moyenne 4 combinaisons
 - calcul final: $T_D = K_{\alpha\beta} / (2r_{\text{moy}})$



4.2. Evolution de familles de gènes

- Exemple d'évolution d'une famille de gènes : les **globines**
 - Composé de 3 sous-familles: myoglobine; α et β -globines



2 protéines : **myoglobine** (monomérique);
hémoglobine (tétramérique: $2\alpha+2\beta$)

⇒ divergence il y a 600-800 10^6 années

- myoglobine: tissus musculaires

- hémoglobine: sang

⇒ duplication puis différenciation

⇒ fonction nouvelle

4.2. Evolution de familles de gènes

- Exemple d'évolution d'une famille de gènes : les **globines**

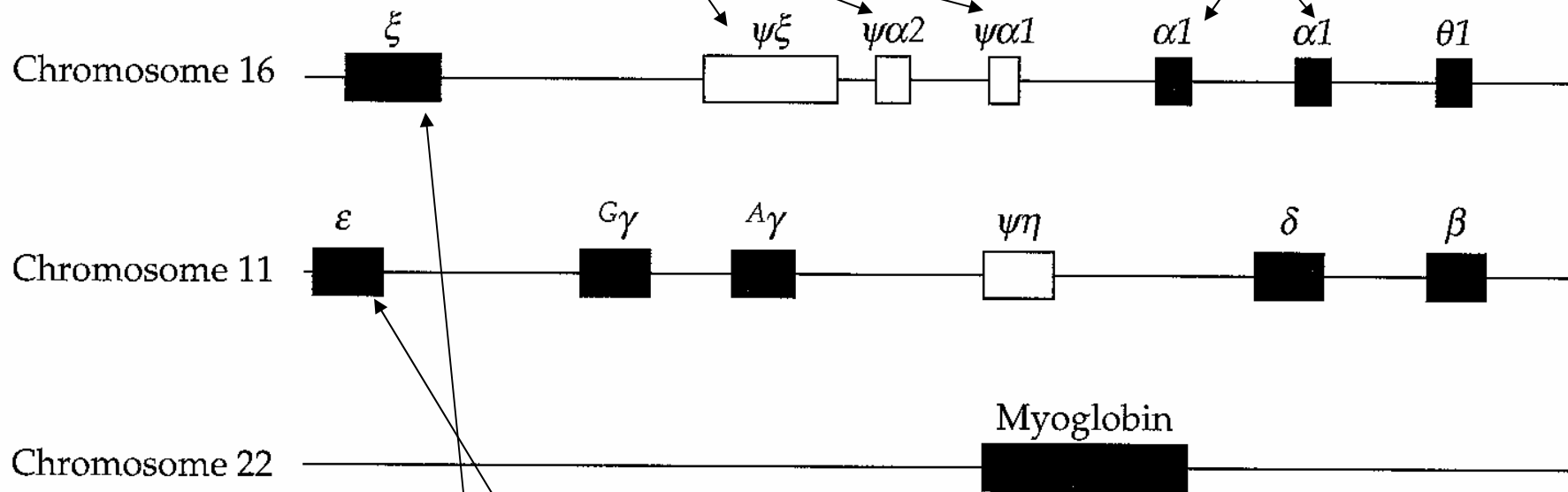
- Composé de 3 sous-familles: myoglobine; α et β -globines

- ⇒ sur des chromosomes différents

- 2 copies quasi-identiques

- Présence de **pseudogènes**

- ⇒ duplication sans différenciation

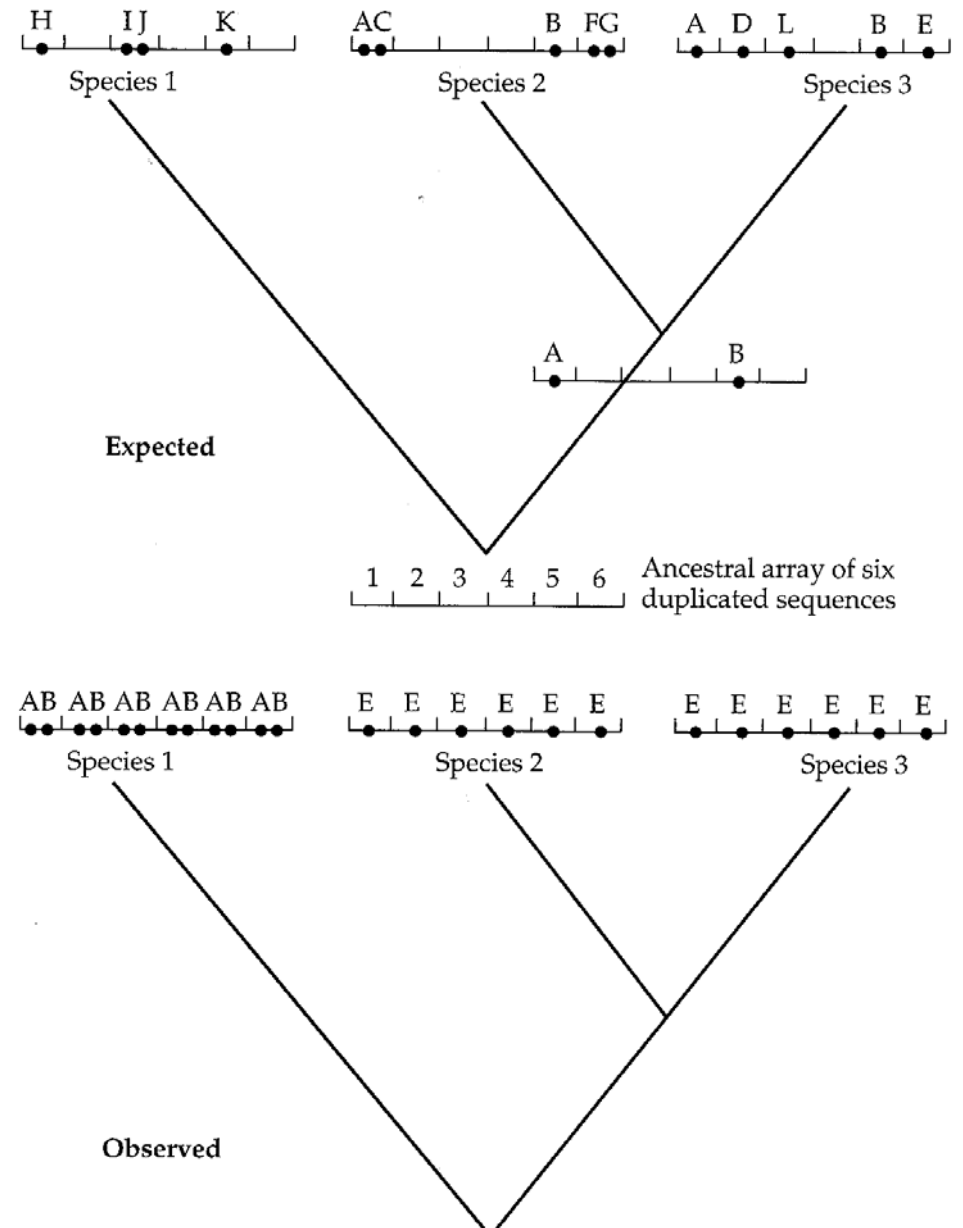


chaînes $\xi + \varepsilon$ → hémoglobines de l'**embryon** et du **fetus**

⇒ duplication puis différenciation ⇒ fonction nouvelle

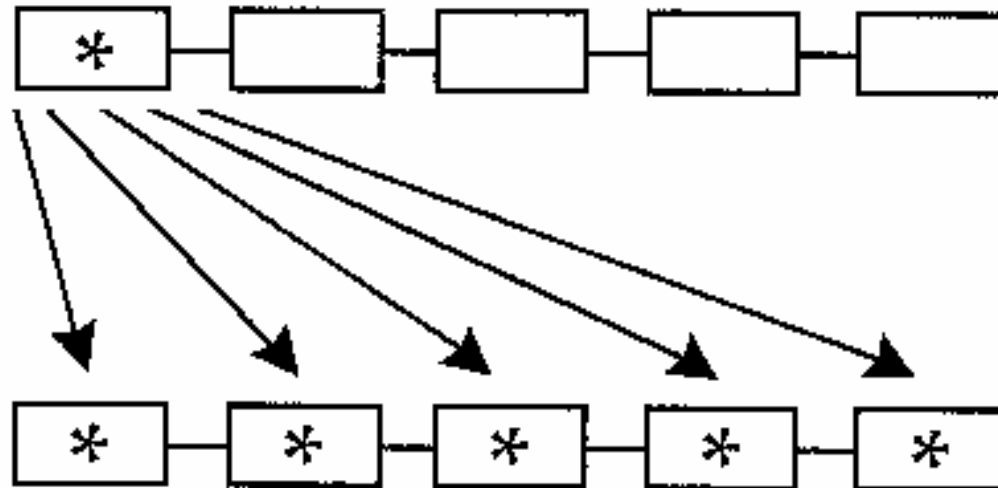
4.3. Evolution concertée de familles de gènes

- Patrons attendus et observés de variation au sein d'une famille de gènes comportant 6 gènes dupliqués en tandem:
 - au sein d'une même espèce, les membres d'une famille de séquences répétées sont fortement similaires, alors que les membres de la même famille provenant d'espèces apparentées peuvent être fort différents
- ⇒ évolution **concertée** des membres d'une famille de gènes



4.3. Evolution concertée de familles de gènes

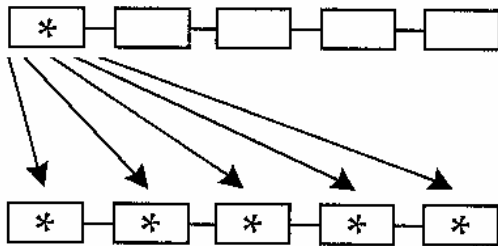
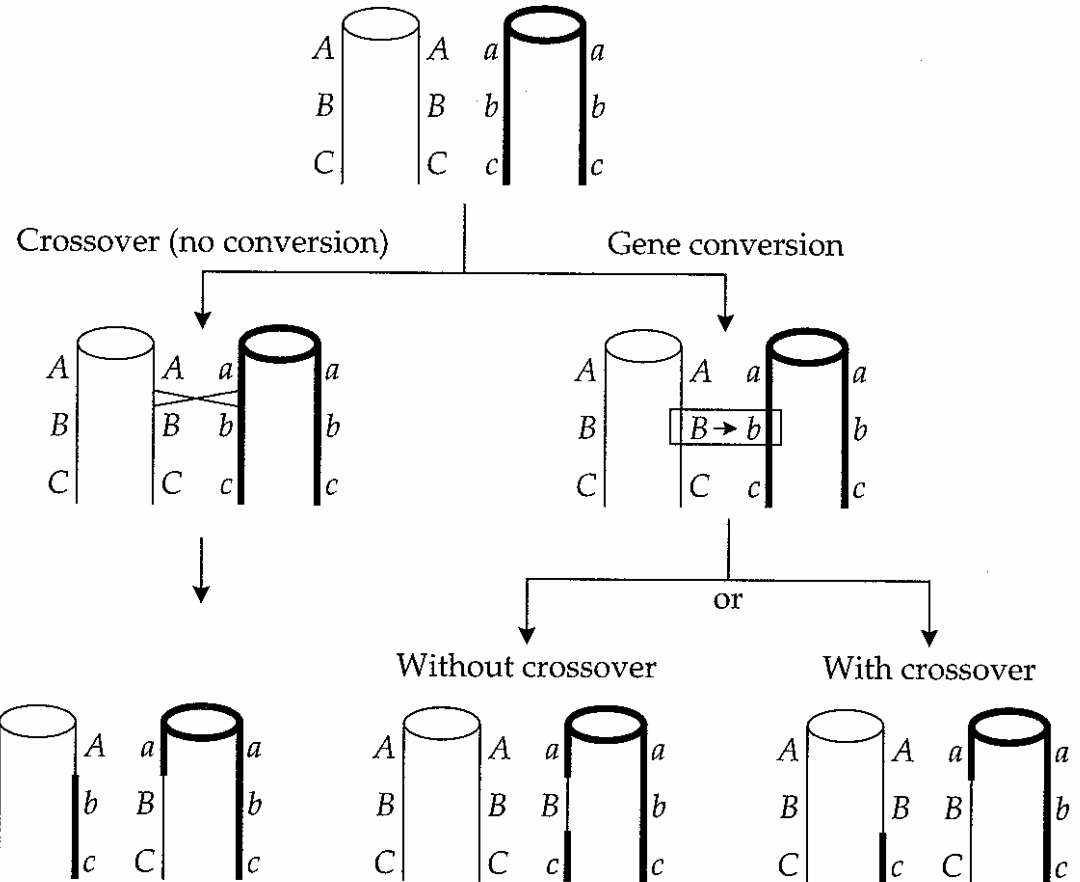
- homogénéisation de la variation au sein d'une famille de gènes dupliqués en tandem:
 - \exists mécanismes qui copient une nouvelle mutation apparue dans 1 élément répété vers les autres:
 - conversion génique
 - crossing-over inégal



4.3. Evolution concertée de familles de gènes

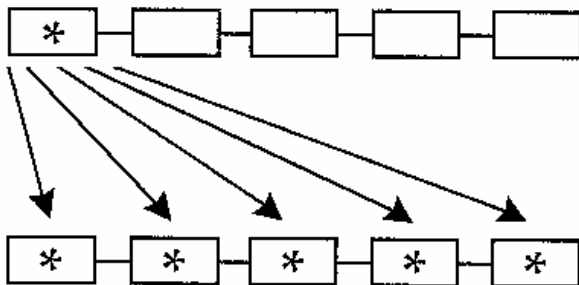
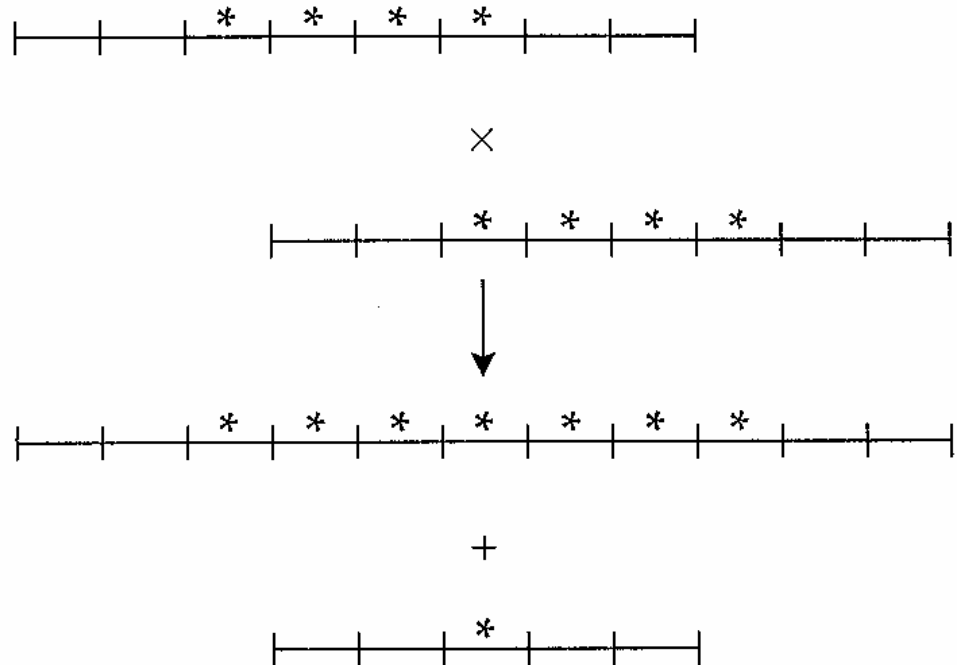
- homogénéisation de la variation au sein d'une famille de gènes dupliqués en tandem:

– conversion génique



4.3. Evolution concertée de familles de gènes

- homogénéisation de la variation au sein d'une famille de gènes dupliqués en tandem:
 - crossing-over inégal:



4.3. Evolution concertée de familles de gènes

- Evolution concertée des gènes $A\gamma$ et $G\gamma$ chez les grands singes:
 - duplication date de 55 millions d'années, nettement antérieure à la séparation Homme/Chimpanzé/Gorille
 - exon 3: montre patron phylogénétique correspondant au temps de divergence
 - exons 1 et 2: évolution concertée dans chaque lignée

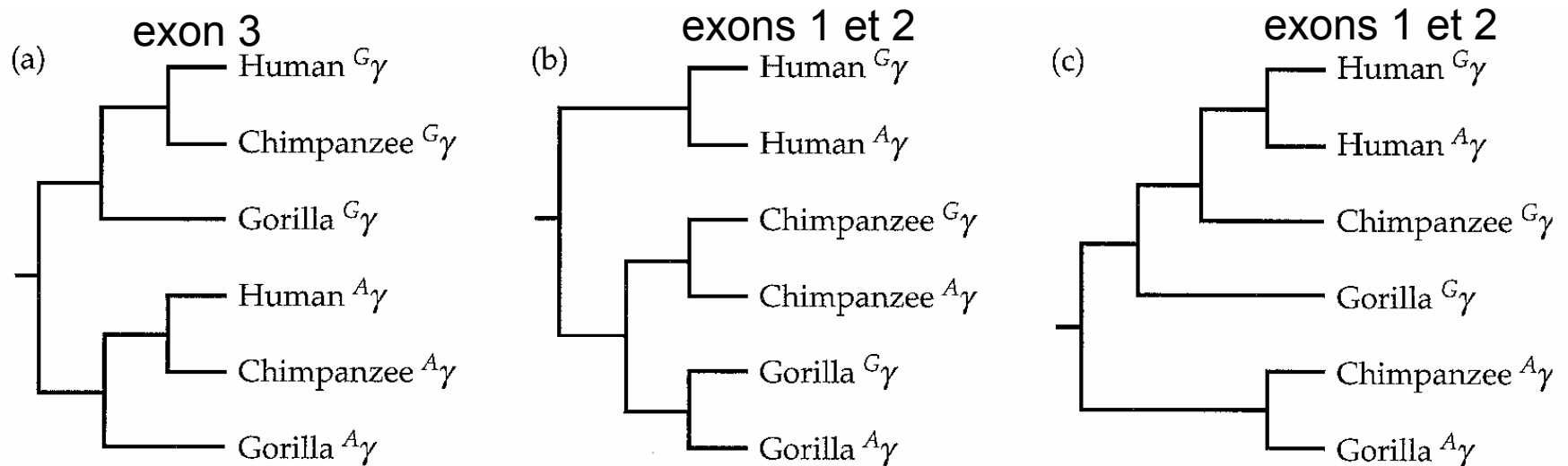


FIGURE 6.33 Phylogenetic trees for (a) exon 3 and (b) exons 1 and 2 of the $G\gamma$ - and $A\gamma$ -globin genes from human, chimpanzee, and gorilla. (c) Expected phylogenetic tree for exons 1 and 2 if gene conversion had only occurred in the human lineage.

⇒ évolution concertée avec intensités \neq dans \neq régions du gène (contraintes sélectives \neq)

4.3. Evolution concertée de familles de gènes

- Conséquences évolutives de l'évolution concertée:
 - propagation de mutations favorables à l'ensemble des membres d'une famille de gènes
 - retarde la divergence de copies d'un même gène
 - complique la datation des événements de duplication de gènes