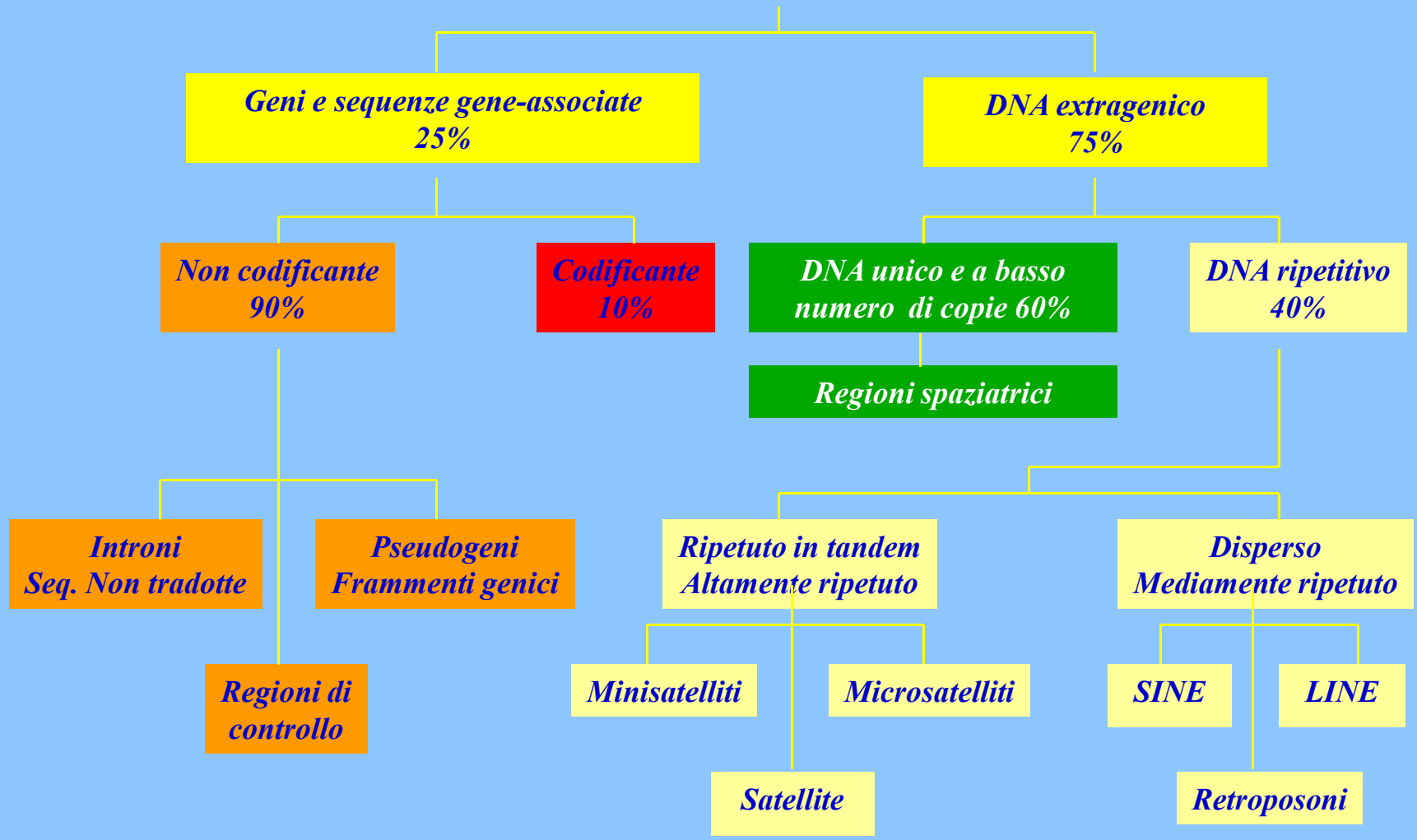


**Genoma Umano
3200 Mb**



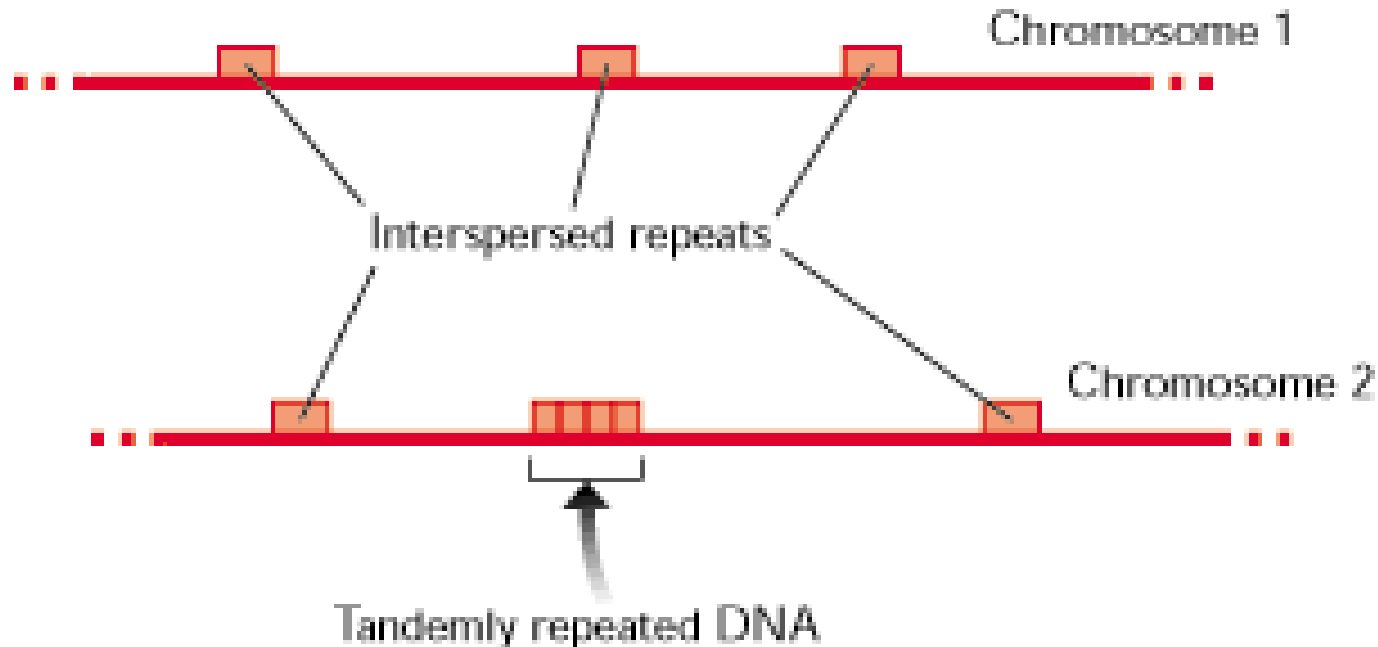
Basato su IHGSC, 2001 e Venter et al., 2001

DNA ripetitivo

E' presente in tutti gli organismi ed in alcuni (incluso l'uomo) costituisce una frazione cospicua del genoma.

Le ripetizioni possono essere

- ripetute in tandem (tandemly repeated)
- ripetizioni disperse nel genoma (genome wide-repeats)

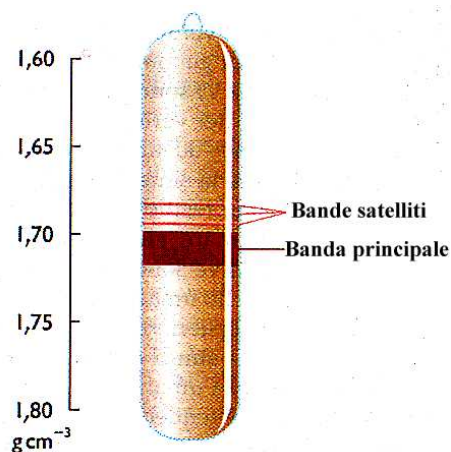


DNA ripetuto in tandem o DNA satellite

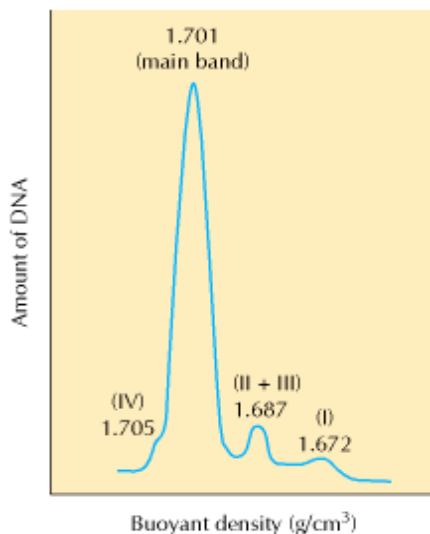
Tipico dei genomi eucariotici (raro nei procarioti)

Satellite

Unità da 5 a 200 bp
Segmenti lunghi fino a qualche centinaio di Kb
localizzato principalmente nei centromeri
(es. DNA alfoide umano)



Gradiente di densità in
cloruro di cesio



Densità di
galleggiamento

Minisatelliti

Unità lunghe fino a 25 bp
Segmenti lunghi fino a 25 kb
(es. DNA telomerico; minisat. Telomerici)
DNA minis. ipervariabile ->DNA fingerprint

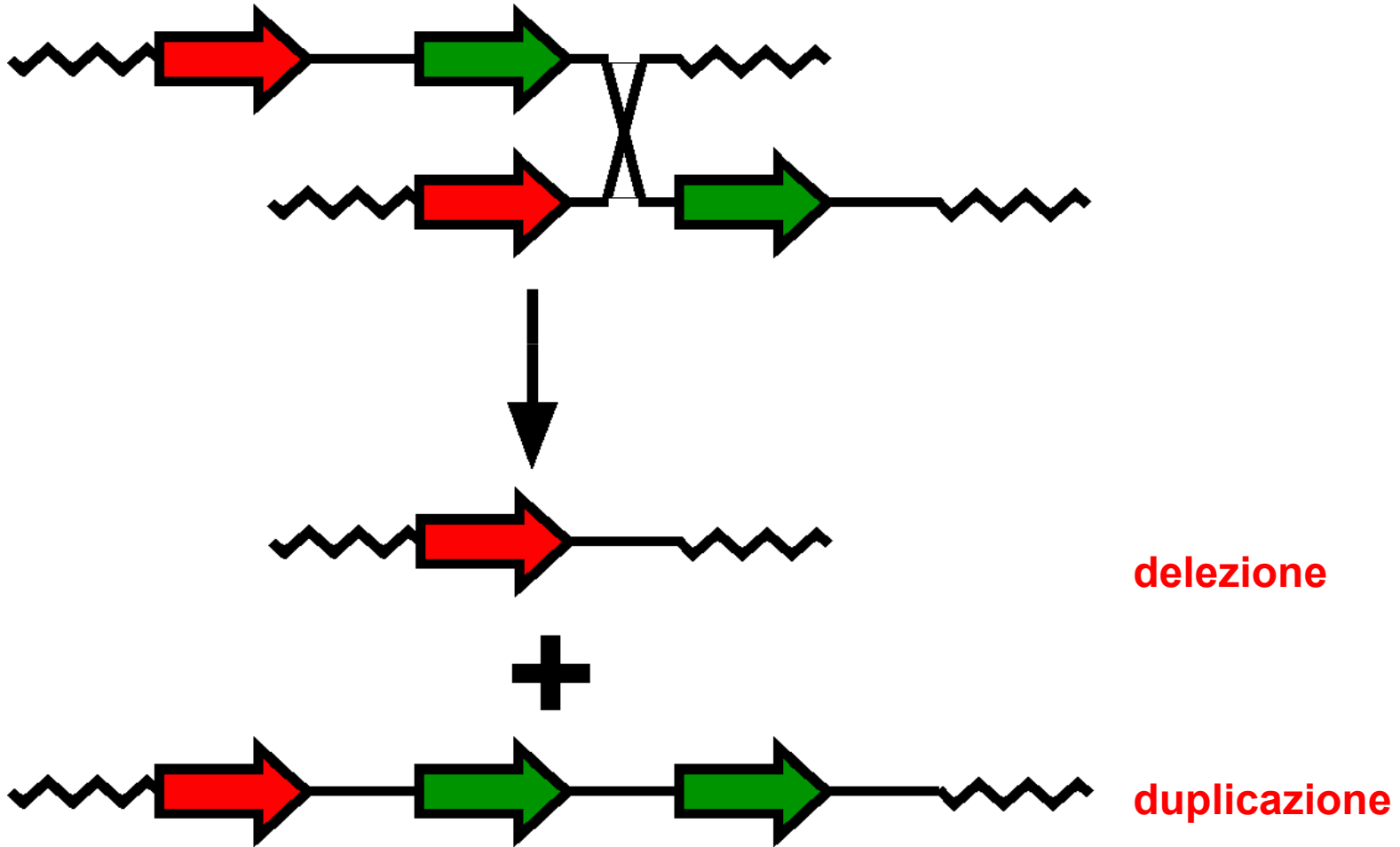
Microsatelliti

Unità < 4bp
Segmenti lunghi fino 150 bp

5'-CACACACACA-3'

Es.: nell'uomo la ripetizione CA
copre lo 0,25% del genoma

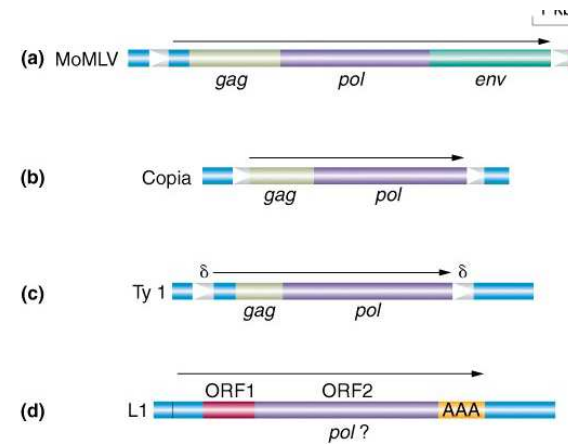
Crossing Over ineguale



DNA ripetuto disperso nel genoma

interspersed genome - wide repeats

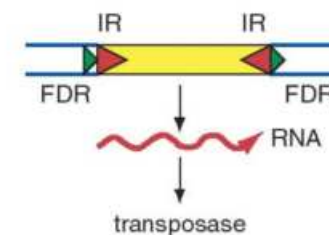
- **Trasposoni ad RNA** (o retrotrasposoni)
 - ⇒ molto + frequenti negli eucarioti
 - ⇒ sono molto simili a **retrovirus**



MoMLV: Moloney murine leukemia virus

- **Trasposoni a DNA**
 - ⇒ molto + frequenti nei procarioti
 - ⇒ **Elementi IS** (es.: IS1, IS186) sequenze d'inserzione

Insertion sequences



DNA ripetuto disperso nel genoma

interspersed genome - wide repeats

- I transposoni si muovono come elementi a DNA
- I retrotransposoni si muovono via un intermedio ad RNA che viene retrotrascritto a DNA da una trascrittasi inversa e mantenuto nel genoma

I TRASPOSONI

I trasposoni sono **elementi mobili** che si trovano nel genoma di tutti gli organismi. Sono in grado, con meccanismi diversi, di saltare da un punto all'altro del genoma. Questi salti possono determinare **drastici cambiamenti nella struttura e nelle modalità di espressione dei geni.**

Appartengono a diverse classi, accomunate dalla caratteristica di avere la sequenza trasposta fiancheggiata da corte sequenze ripetute generate dalla duplicazione di una sequenza nel sito bersaglio (IR)

Sotto certi aspetti possono essere considerati come **elementi parassitari**, e in effetti alcuni di essi sono imparentati con particolari famiglie di virus (retrovirus). Tuttavia rappresentano un formidabile **fattore positivo per i processi evolutivi.** Non a caso circa il 45% del genoma umano e' costituito da residui di trasposoni.

Trasposoni ad RNA

- **Elementi LTR** (Caratteristici degli eucarioti non-vertebrati)

[B] *Ty1*/copia retrotransposon



- **Retroposoni** (senza LTR alle estremità)

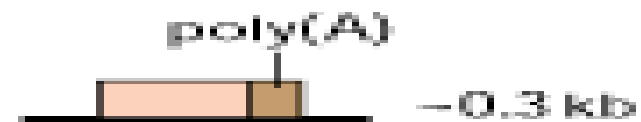
- **LINEs**: hanno trascrittasi inversa (es.: LINE 1 nell'uomo: 6.1 kb, 516.000 copie)

[C] LINE



- **SINEs**: no trascrittasi inversa (es.: Alu, 300bp > 1x10⁶ copie)

[D] SINE

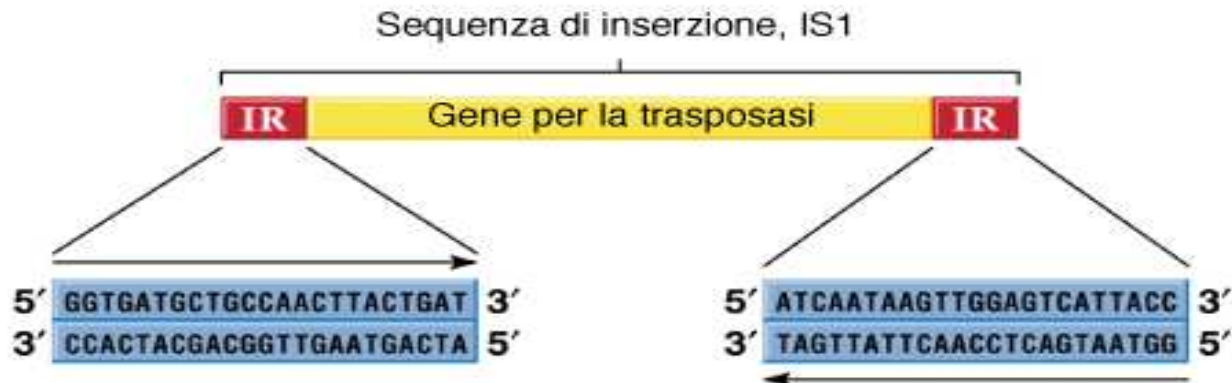


Trasposoni a DNA

- **Elementi IS, Sequenze d'inserzione;** Il genoma di *E.coli* può contenerne ~ 20.

Figura 20.1

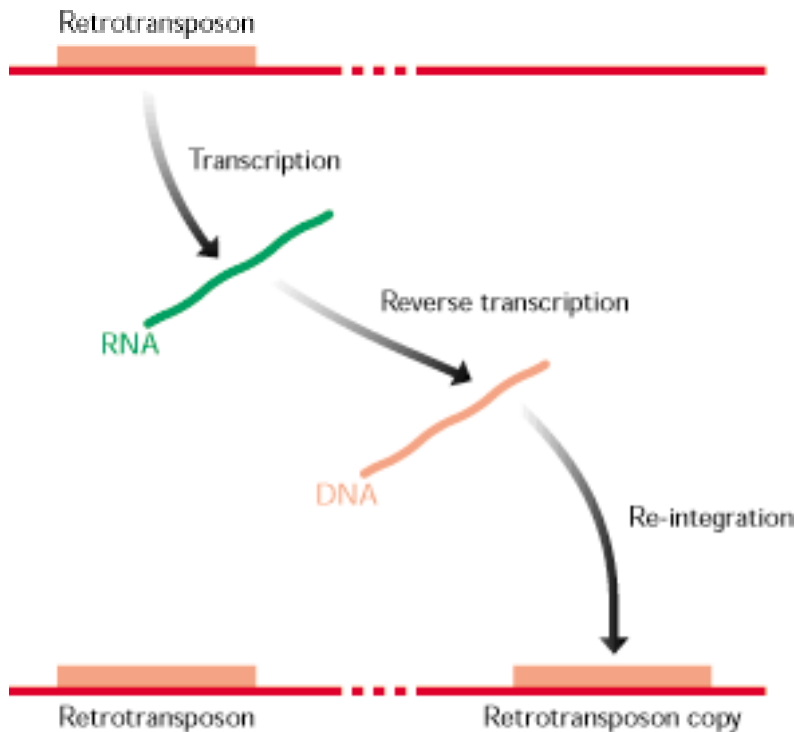
La sequenza di inserzione (IS) dell'elemento trasponibile IS1. L'elemento IS è lungo 768 bp ed ha una ripetizione invertita (IR) a ciascuna estremità. Sotto l'elemento sono indicate le sequenze delle ripetizioni invertite terminali (IR) di 23 bp.



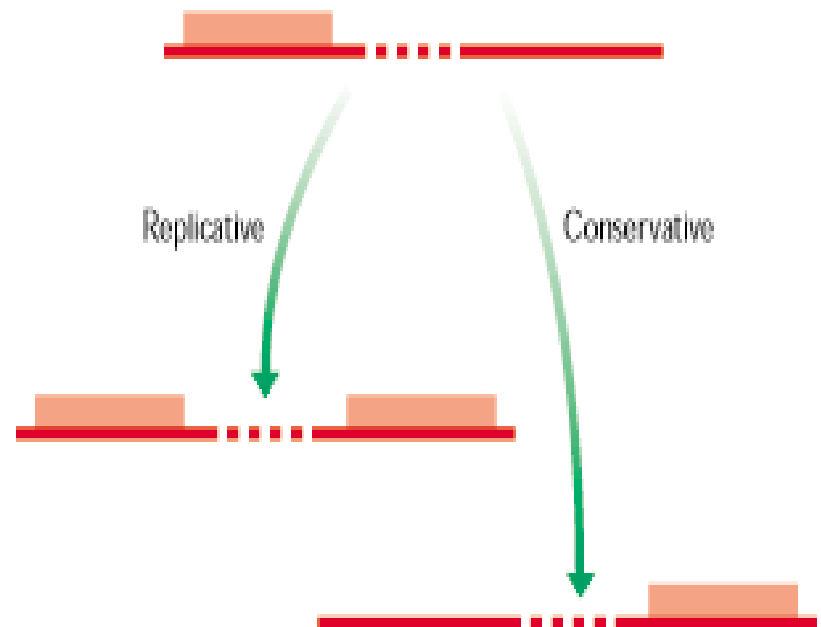
Origine del DNA ripetuto disperso nel genoma

le ripetizioni sparse originano per **trasposizione**

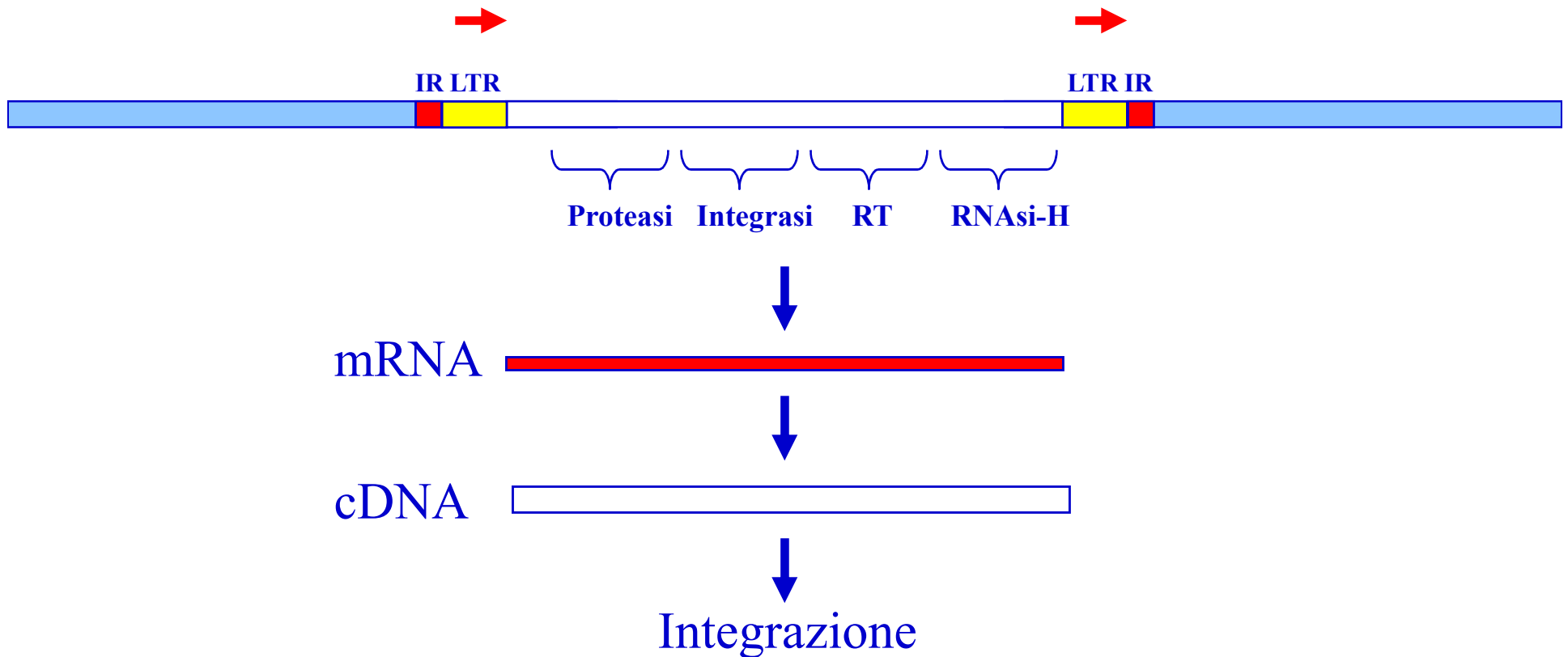
trasposoni ad RNA
(retrotrasposizione)



trasposoni a DNA
trasposizione diretta
(replicativa, conservativa)

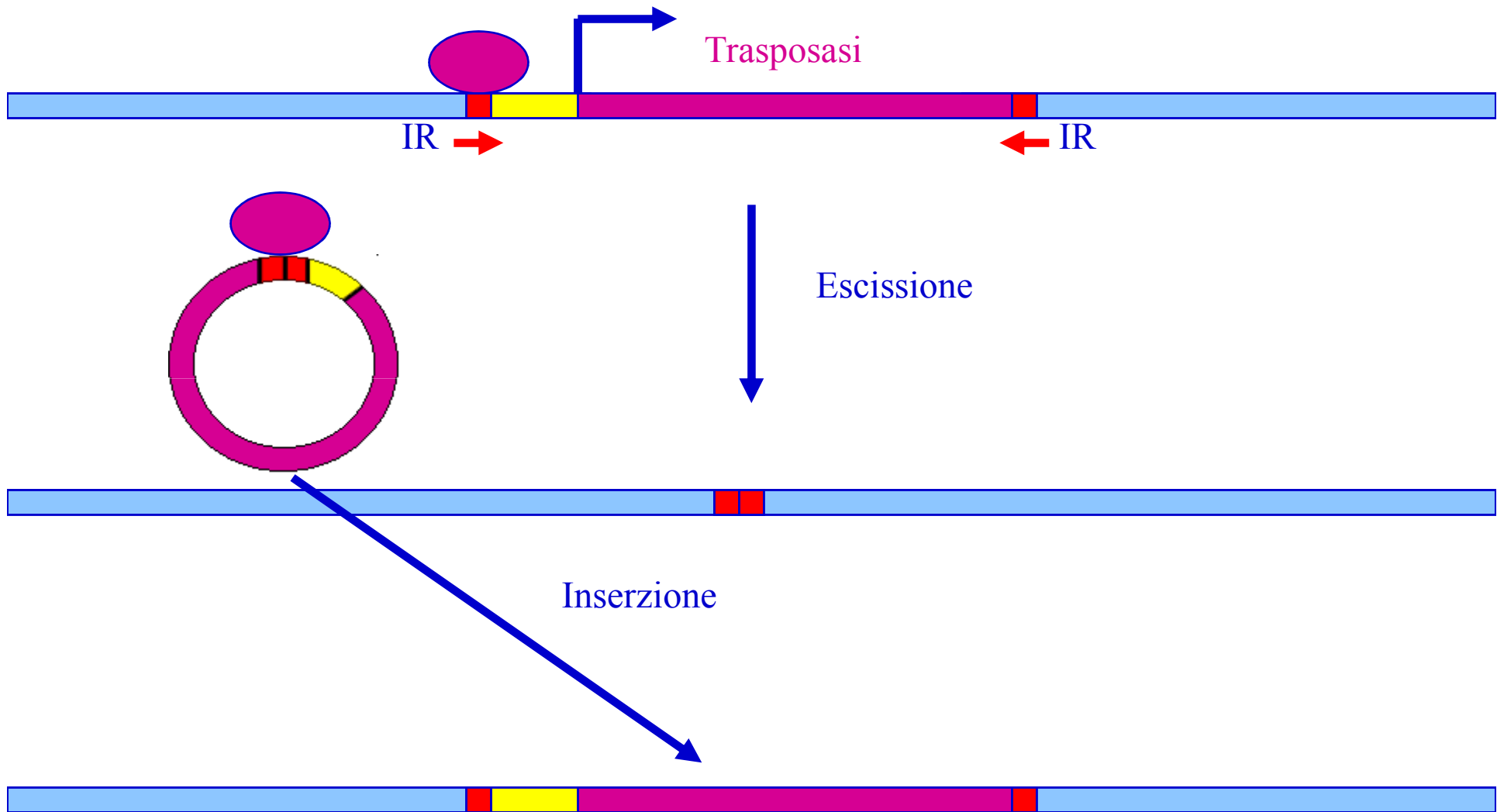


Trasposoni che si muovono indirettamente (retrovirus endogeni e retrotrasposoni o retroposoni)



Gli pseudogeni che si ritrovano nel genoma umano sono spesso derivati da eventi di retrotrasposizione

Trasposoni che si muovono direttamente

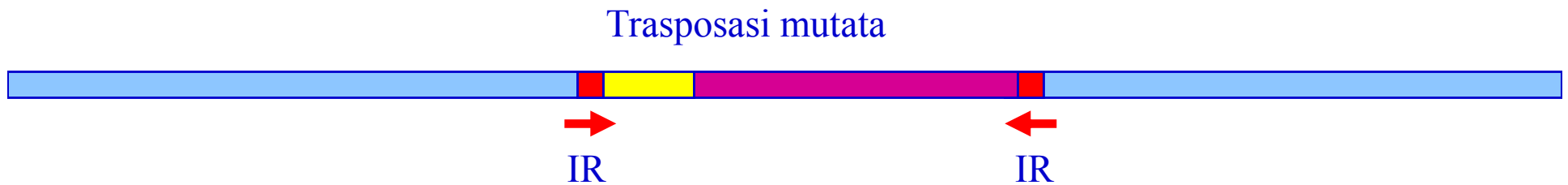


Alcuni trasposoni sono portatori di mutazioni nel gene che codifica la trasposasi, per cui non si possono muovere autonomamente.

Autonomi

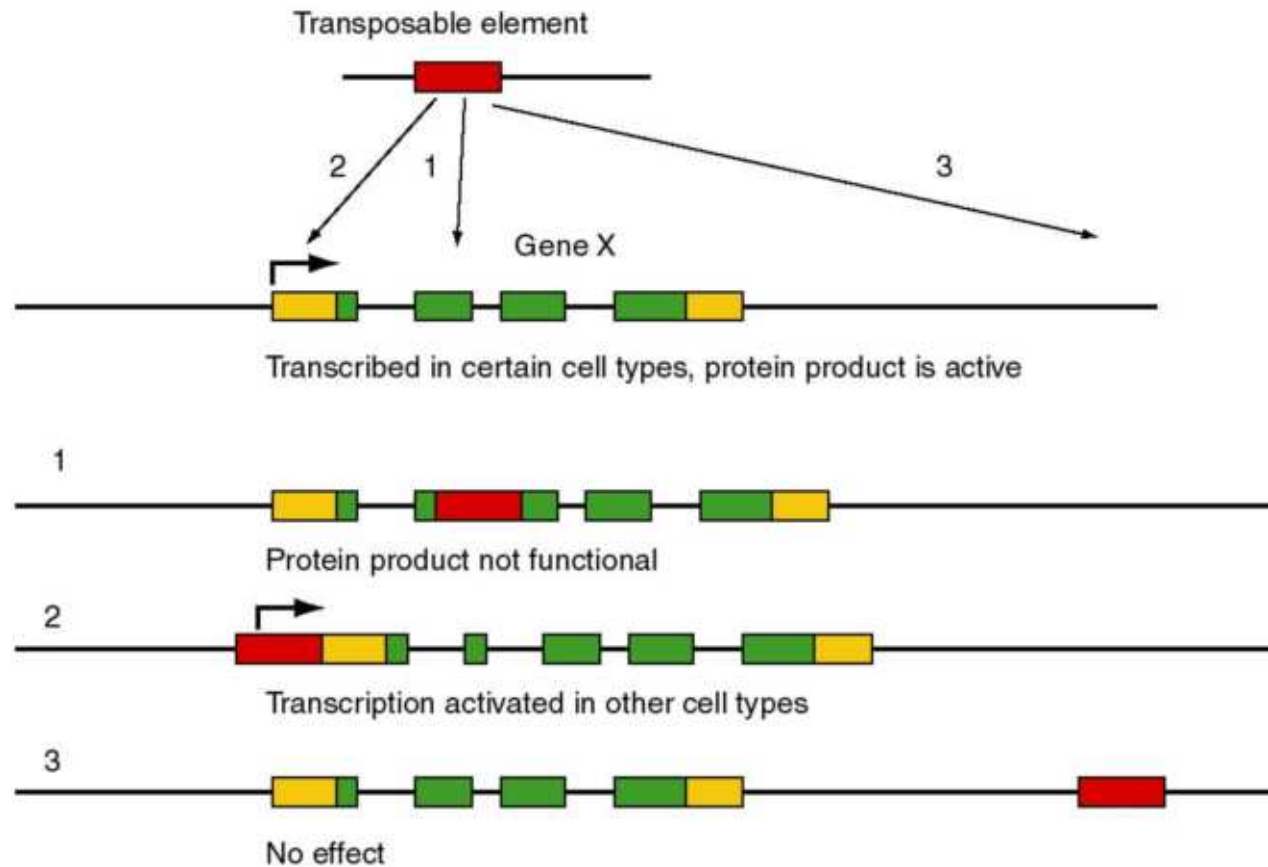


Difettivi



Esempi: Elementi Ac e Ds di mais. Trasposone batterico IS10.

Che cosa possono fare i trasposoni saltando nel genoma?



La mobilità dei trasposoni può essere molto pericolosa per le cellule somatiche, ma la sua attività nella linea germinale aumenta le probabilità di avere organismi con nuove caratteristiche, che possono rivelarsi vincenti in rapporto alle condizioni ambientali



Evoluzione del genoma

Famiglie multigeniche

- famiglie multigeniche **classiche**
- famiglie geniche che codificano prodotti che condividono **grossi domini** altamente conservati
- famiglie geniche che codificano prodotti con **brevissimi motivi** aminoacidici conservati
- **superfamiglie** geniche

Famiglie geniche classiche

Tabella 8.1 Esempi di famiglie con più geni raggruppati

Famiglia	N° di copie	Organizzazione	Localizzazione cromosomica
Raggruppamento dei geni C4 del complemento	2	Unità ripetute composite in tandem lunghe ~ 30 kb; entrambe espresse	6p21.3
Raggruppamento dell'ormone della crescita	5	Raggruppati in 67 kb; uno pseudogene convenzionale	.17q22-24
Raggruppamento dei geni α -globinici	7	Raggruppati in ~ 50 kb; tre geni funzionali, un gene espresso con funzione ignota, tre pseudogeni convenzionali (<i>Figura 8.6</i>)	16p13.3
Geni della catena pesante HLA di classe I	~ 20	Raggruppati in 2 Mb; sei sono espressi, quattro sono pseudogeni convenzionali, sette sono geni troncati o frammenti genici (<i>Figura 8.4</i>)	6p21.3
Geni <i>HOX</i>	38	Organizzati in quattro raggruppamenti (<i>Figura 9.5</i>)	2p, 7, 12, 17
Geni istonici	100	Raggruppati in pochi punti, notevole il raggruppamento composito in 1p21	1p21, 6, 12q

- i membri presentano un elevato grado di omologia per quasi tutta la loro sequenza codificante.

- Quando si presentano raggruppati la loro espressione può essere coordinata da un'unica regione di controllo LCR (Locus Control Region).

- Possono essere **semplici** o **complesse**.

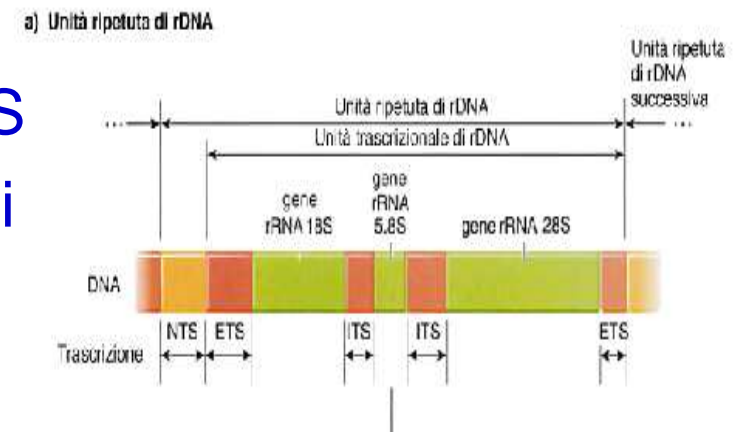
Famiglie geniche classiche semplici

Tutti gli eucarioti e quasi tutti i batteri (tranne i più semplici) hanno copie multiple dei geni per gli rRNA

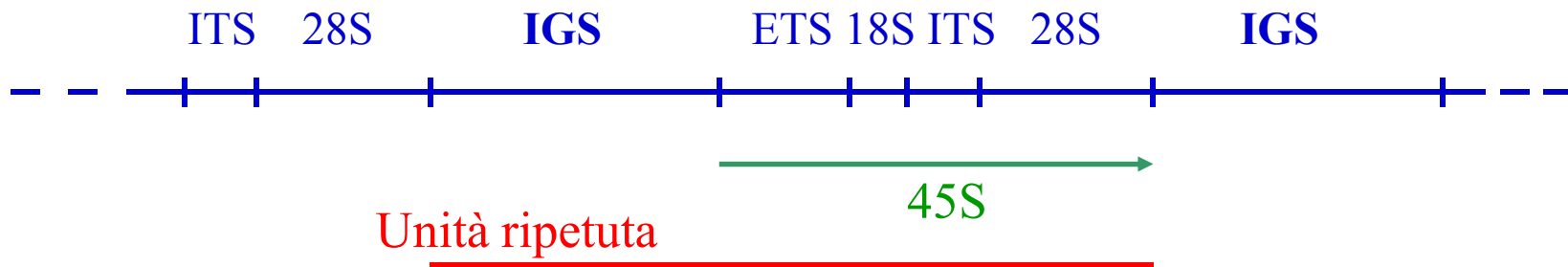
E' un esempio di famiglie geniche SEMPLICI, originatesi per duplicazione genica, in cui **tutti i membri hanno mantenuto la stessa sequenza**

Nell'uomo ci sono 2000 geni per 5S rRNA in un cluster unico sul cromosoma 1

Inoltre: 280 copie di un'unità ripetuta contenente i geni per gli rRNA 28S, 5.8S e 18S, raggruppati in unità di 50-70 geni ciascuna sui cr. 13,14,15, 21, 22.



L'esempio del DNA ribosomale di *Xenopus*

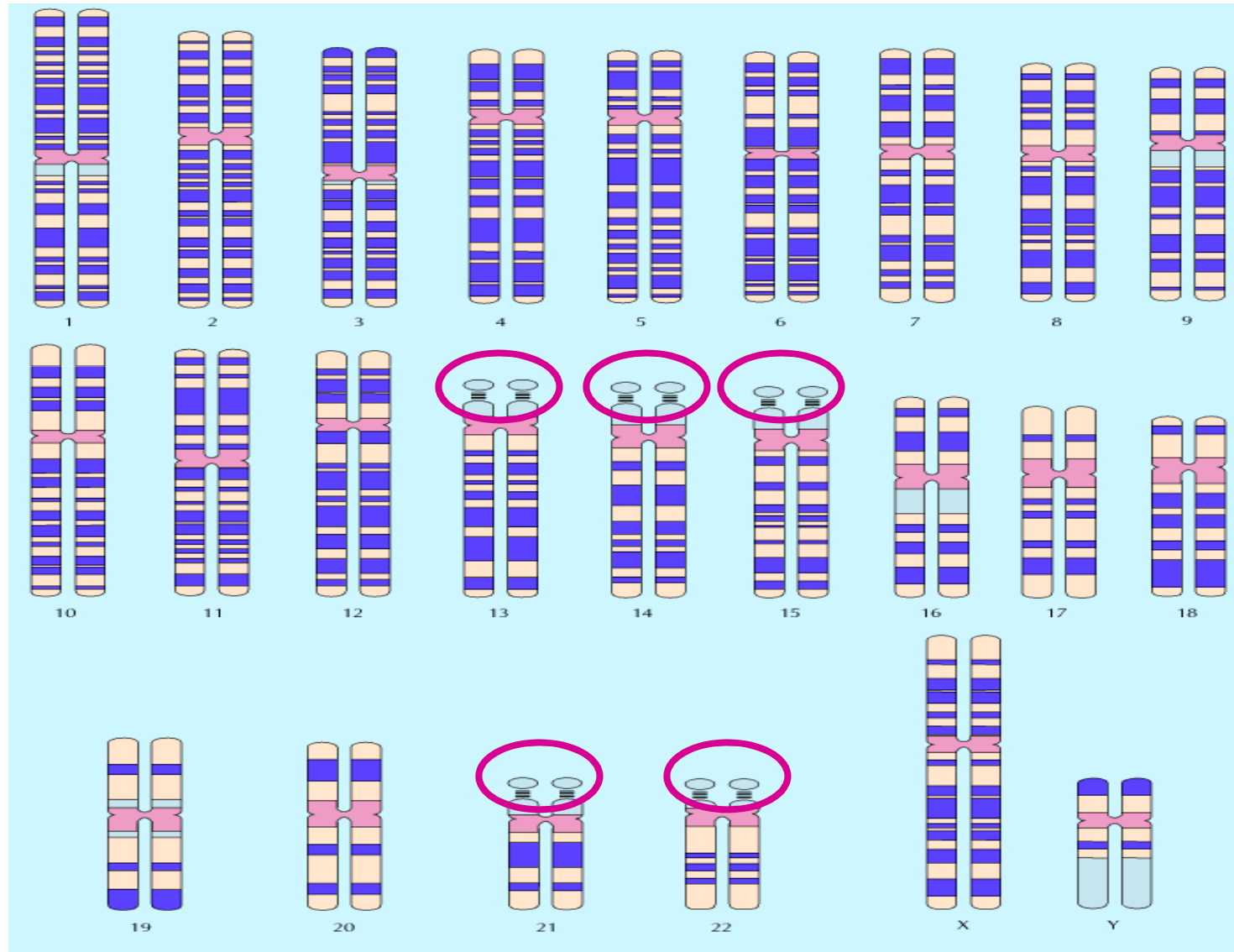


ITS: internal transcribed spacer; ETS: external transcribed spacer;
IGS: intergenic spacer

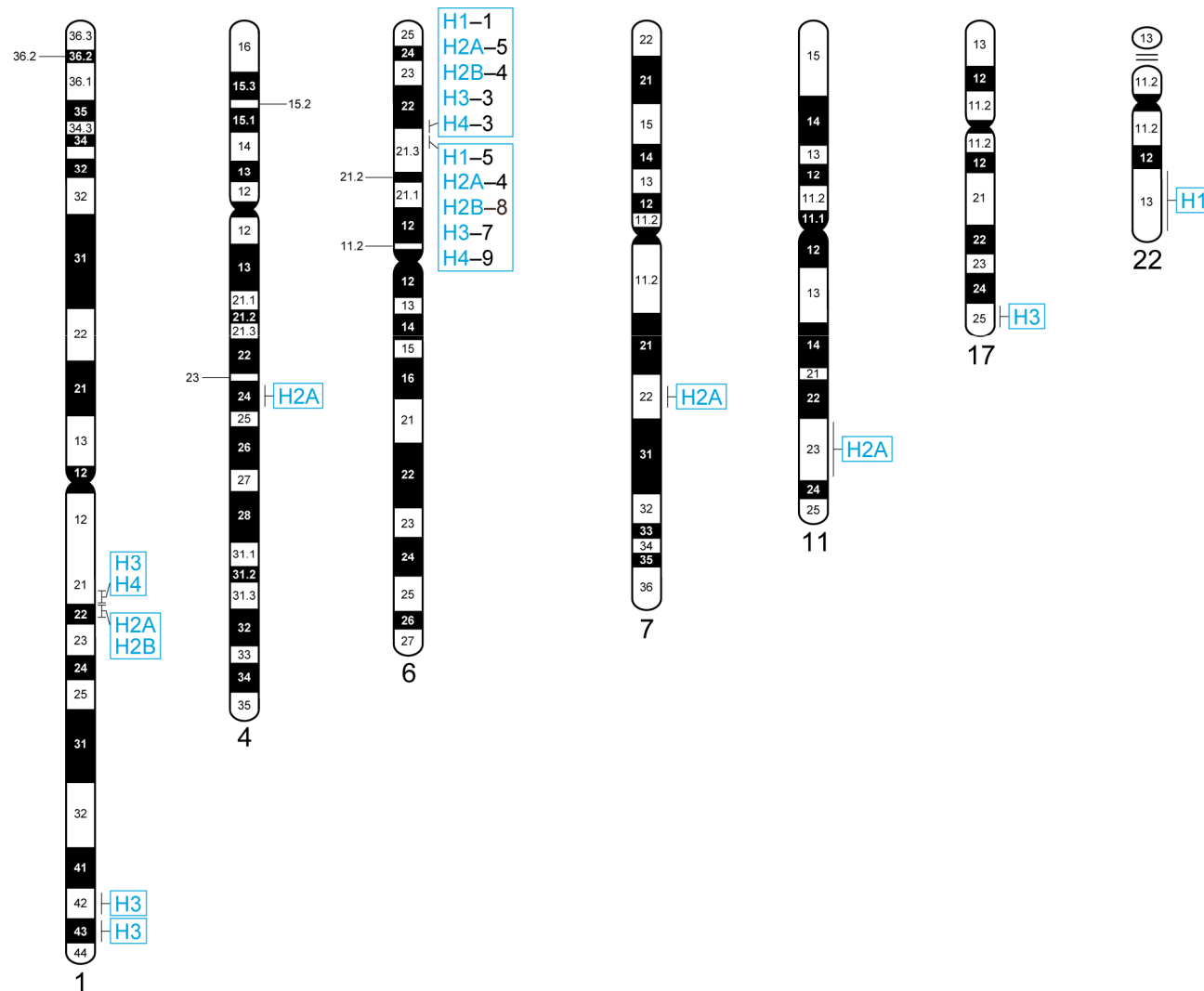
- Ribosomal RNA: Centinaia di copie in gruppo
- IGS: altamente divergenti fra *X. laevis* e *X. borealis* (fra gruppi) = l'IGS ha un alto tasso di mutazione
- IGS: altamente omogenee in *X. laevis* (dentro il gruppo) = non dovuta a pressione selettiva che mantiene l'omogeneità, ma

EVOLUZIONE ORIZZONTALE = EVOLUZIONE CONCERTATA

Distribuzione genomica dell'rDNA



– Geni funzionalmente simili alle volte sono
posizionati in regioni cromosomiche contigue,
ma **solitamente sono disperse** nel genoma



Famiglie geniche classiche complesse

i diversi membri della famiglia, pur con sequenza simile, producono **proteine con diverse proprietà biochimiche**

Es. classico: la famiglia delle globine nei mammiferi

Emoglobina =
 $2 \text{ cat } \alpha + 2 \beta$

Figura 17.19

Mapa di concatenazione dei gruppi dei geni umani delle globine. La funzione, se ve ne è una, del gene per l' α -globina $\Phi 1$ non è nota. (La mappa non è in scala).

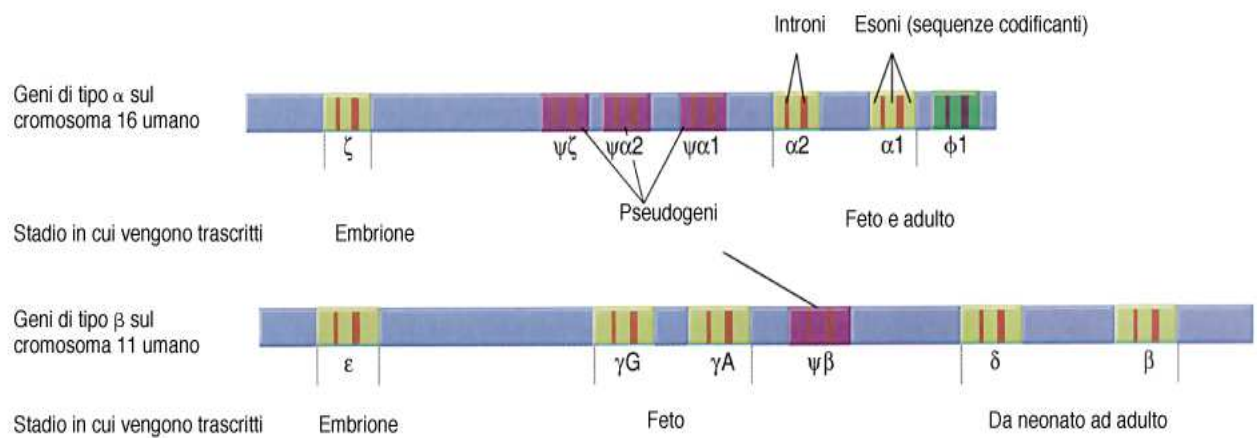
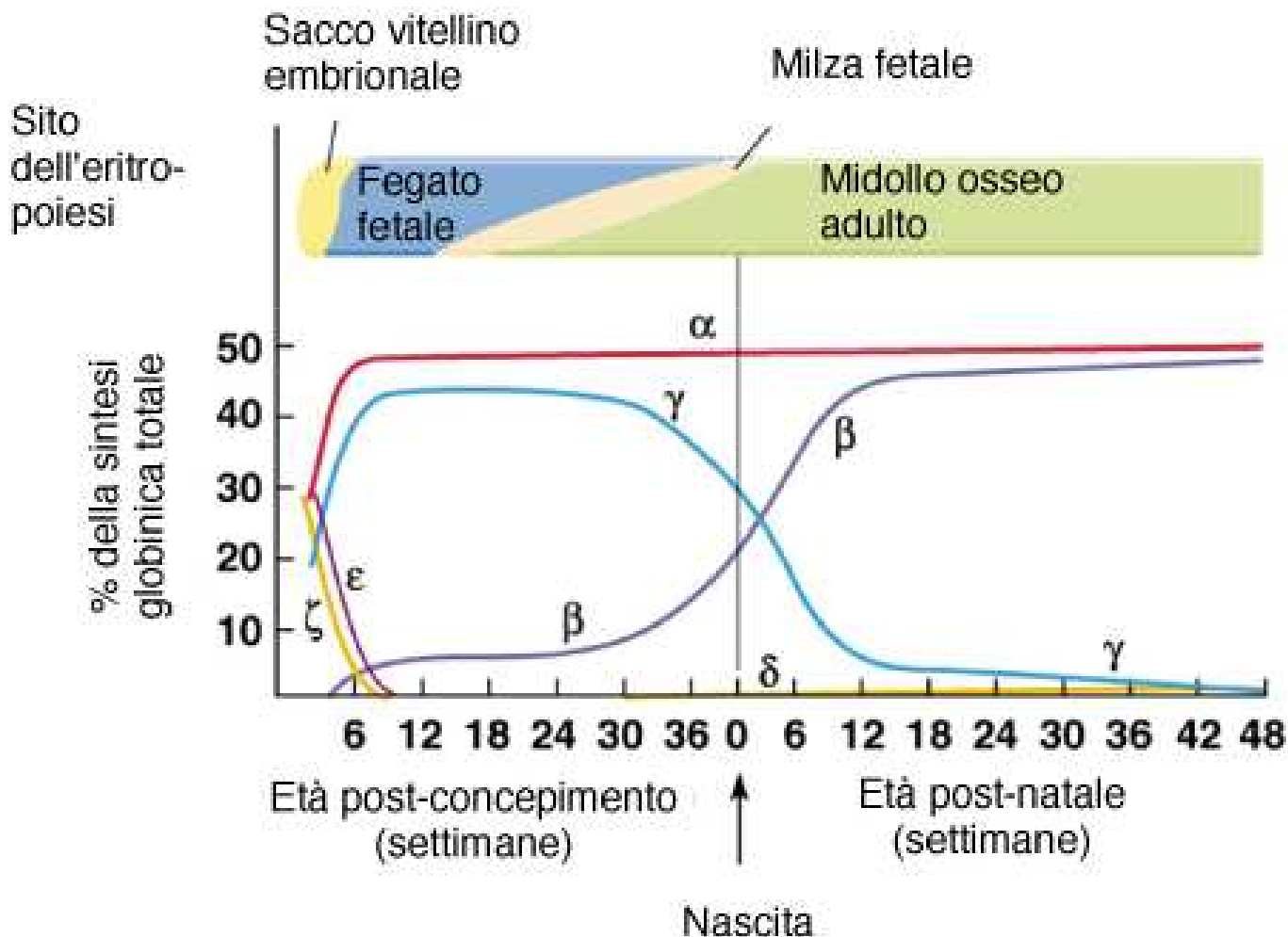
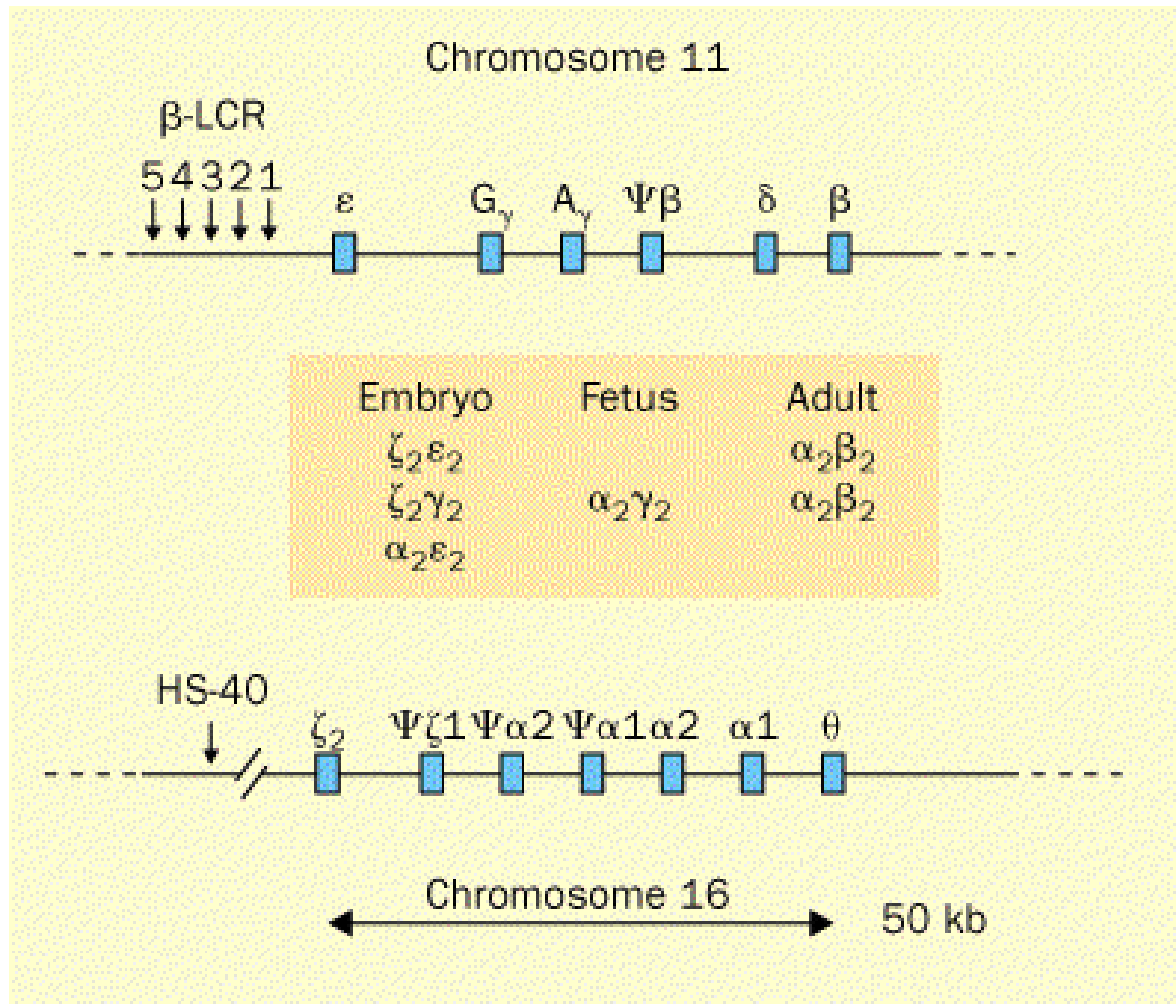


Figura 17.18

Confronto della sintesi di differenti catene globiniche a determinati stadi di sviluppo embrionale, fetale e post-natale.





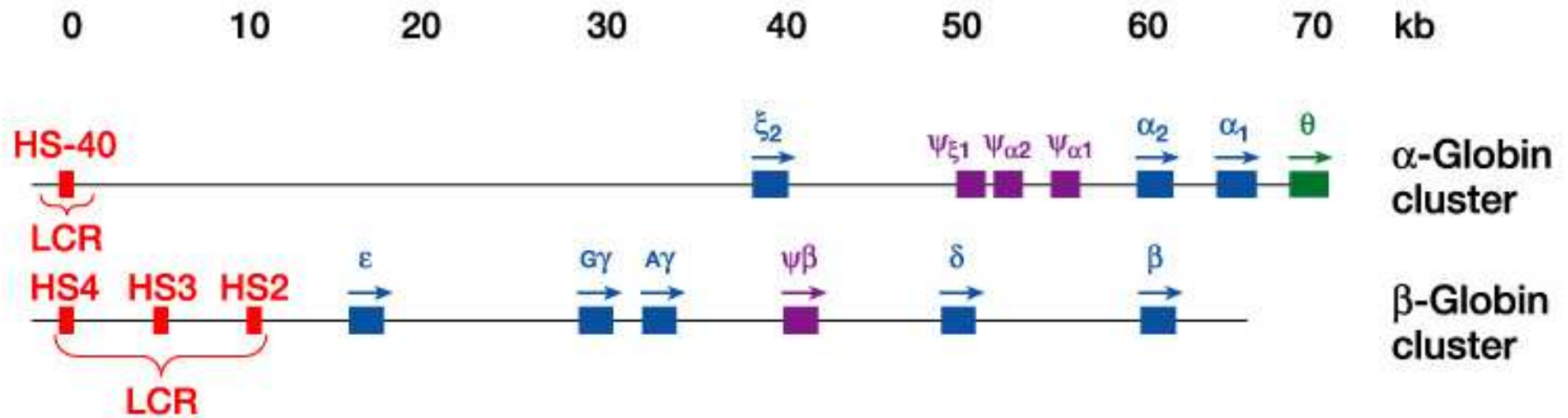
Controllo genetico e struttura dell'emoglobinina umana
 Regioni di controllo: locus control region [LCR]; hypersensitive site [HS]-40.
 (Weatherall and Proven, Lancet 2000;355:1169-1175)

$\zeta_2\varepsilon_2$ Emoglobina Embrionale (espressione nel sacco vitellino)

$\alpha_2\gamma_2$ HbF Emoglobina fetale (espressione nel fegato e nella milza)

$\alpha_2\delta_2$ HbA₂ Emoglobina dell'adulto

$\alpha_2\beta_2$ HbA



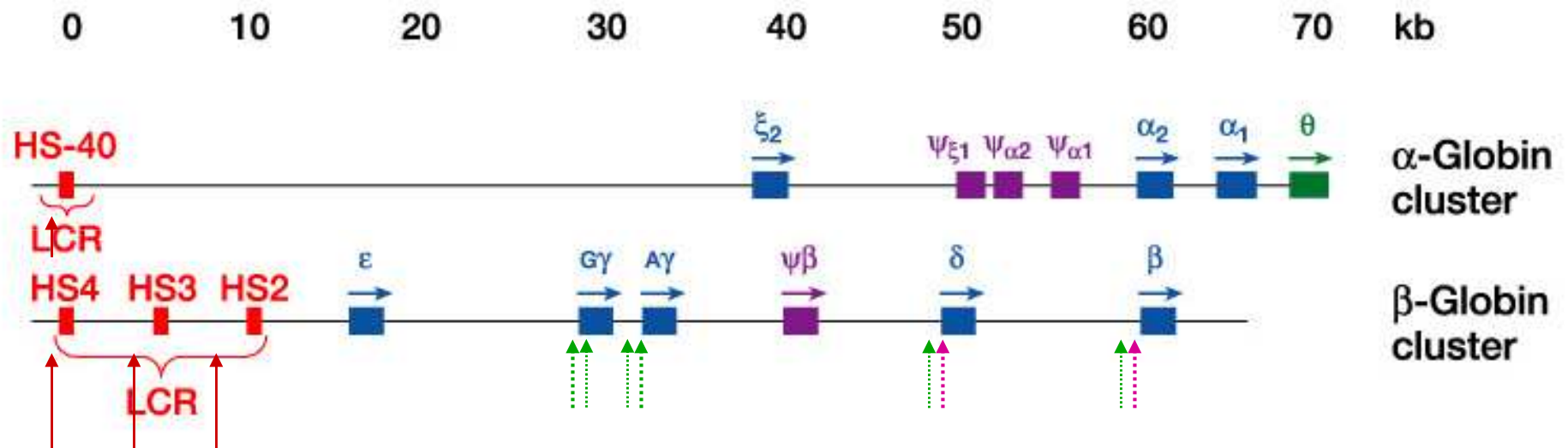
L'emoglobina nell' adulto è di tipo HbA 96% $\alpha_2\beta_2$, HbA₂ $\alpha_2\delta_2$ 3% e HbF $\alpha_2\gamma_2$ 1%
Le LCR (locus control region) funzionano da enhancer per la trascrizione dei geni globinici

$\zeta_2\varepsilon_2$ Emoglobina Embrionale (espressione nel sacco vitellino)

$\alpha_2\gamma_2$ HbF Emoglobina fetale (espressione nel fegato e nella milza)

$\alpha_2\delta_2$ HbA₂ Emoglobina dell'adulto

$\alpha_2\beta_2$ HbA



Siti ipersensibili a DNAsi I
Eritroide-specifici (enhancer)

Altri siti ipersensibili nella regione dei promotori dei singoli geni ->
specificità dello stadio di sviluppo.

Siti ipersensibili del fegato fetale

Siti ipersensibili nel midollo osseo adulto

Geni omologhi

I geni **omologhi** condividono un antenato comune e questo fatto è rivelato dalle **similarità di sequenza**, sulle quali si basa la filogenesi molecolare

I geni omologhi si dividono in due categorie:

- Geni **PARALOGHI**
- Geni **ORTOLOGHI**

Sequenze ortologhe

Due sequenze si dicono **ortologhe** quando rappresentano lo stesso gene in specie differenti. Il gene era presente già nell'antenato comune delle due specie considerate. Le sequenze **ortologhe** *derivano pertanto da un processo di speciazione*

Sequenze paraloghe

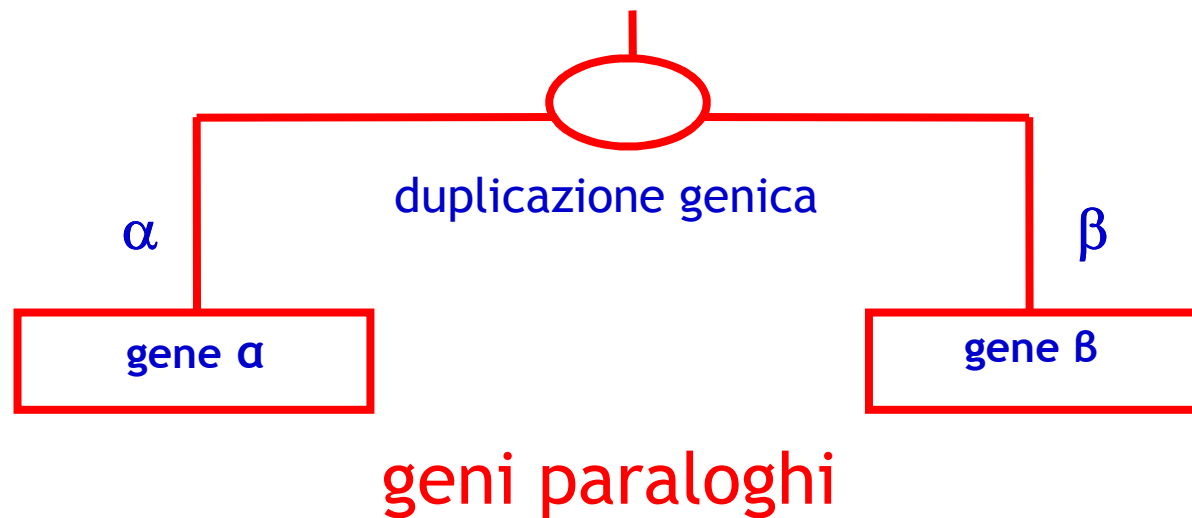
Due sequenze si dicono **paraloghe** quando derivano da un comune gene ancestrale e *si sono prodotte per un processo di duplicazione genica*

Geni paraloghi

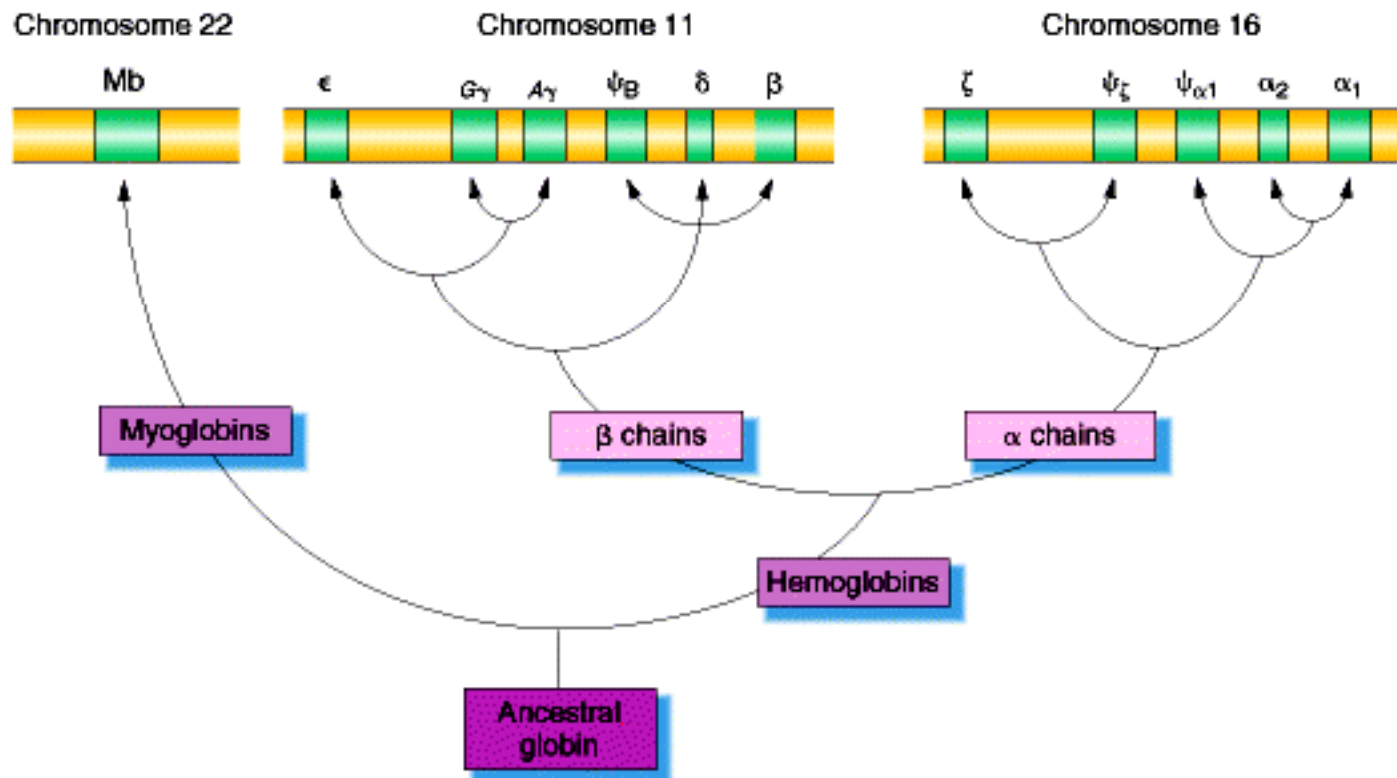
Si definiscono **paraloghi** i geni, **localizzati in uno stesso organismo**, derivanti da eventi di **duplicazione**. Questi eventi possono portare alla produzione di **famiglie di proteine** correlate con strutture e funzioni biologiche simili.

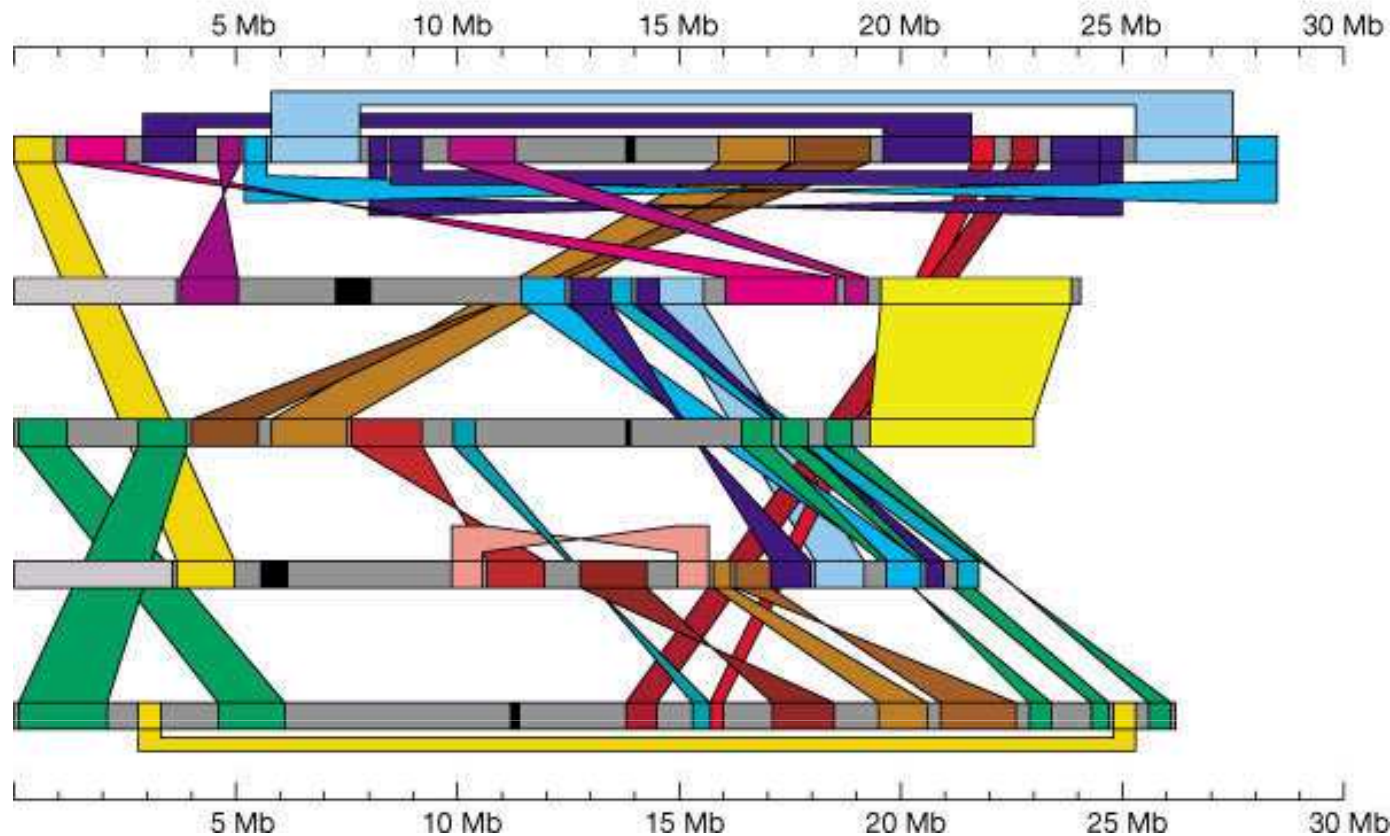
A volte però i paraloghi possono diversificarsi moltissimo, ed **acquisire funzioni molto diverse** nonostante la struttura simile.

L'antenato comune ai due geni può essere o non essere presente in specie precedenti a quelle in cui si trovano attualmente.



I meccanismi di duplicazione e diversificazione consentono di generare facilmente nuove varianti di una proteina





Regioni duplicate nel genoma di *Arabidopsis* (il cromosoma 1 è in alto). I centromeri sono colorati in nero.

Le bande colorate collegano segmenti duplicati corrispondenti. I segmenti duplicati con orientamento invertito sono collegati con bande colorate “attorcigliate”.

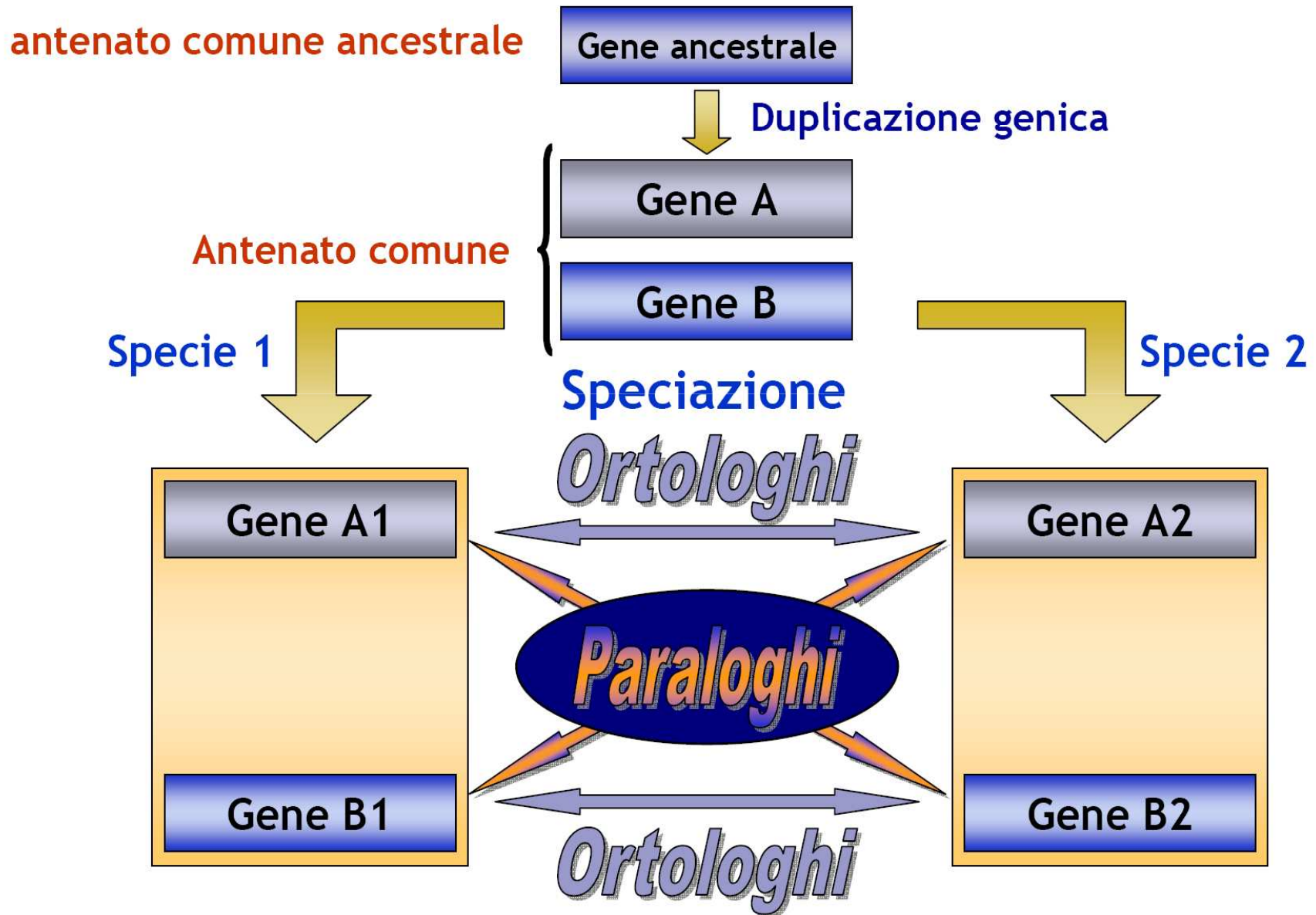
Non sono riportate le regioni di similarità a carico dell'rDNA.

Geni ortologhi

Due geni trovati **in specie diverse** sono **ortologhi** quando le proteine che codificano mostrano un 60-80% di identità. Quasi certamente queste proteine hanno la **stessa struttura tridimensionale**, gli stessi domini funzionali e la stessa funzione biologica.

I geni codificanti hanno avuto origine da un **antenato comune** che esisteva in un periodo evolutivo precedente all'evento di speciazione.

Ortologhi e Paraloghi

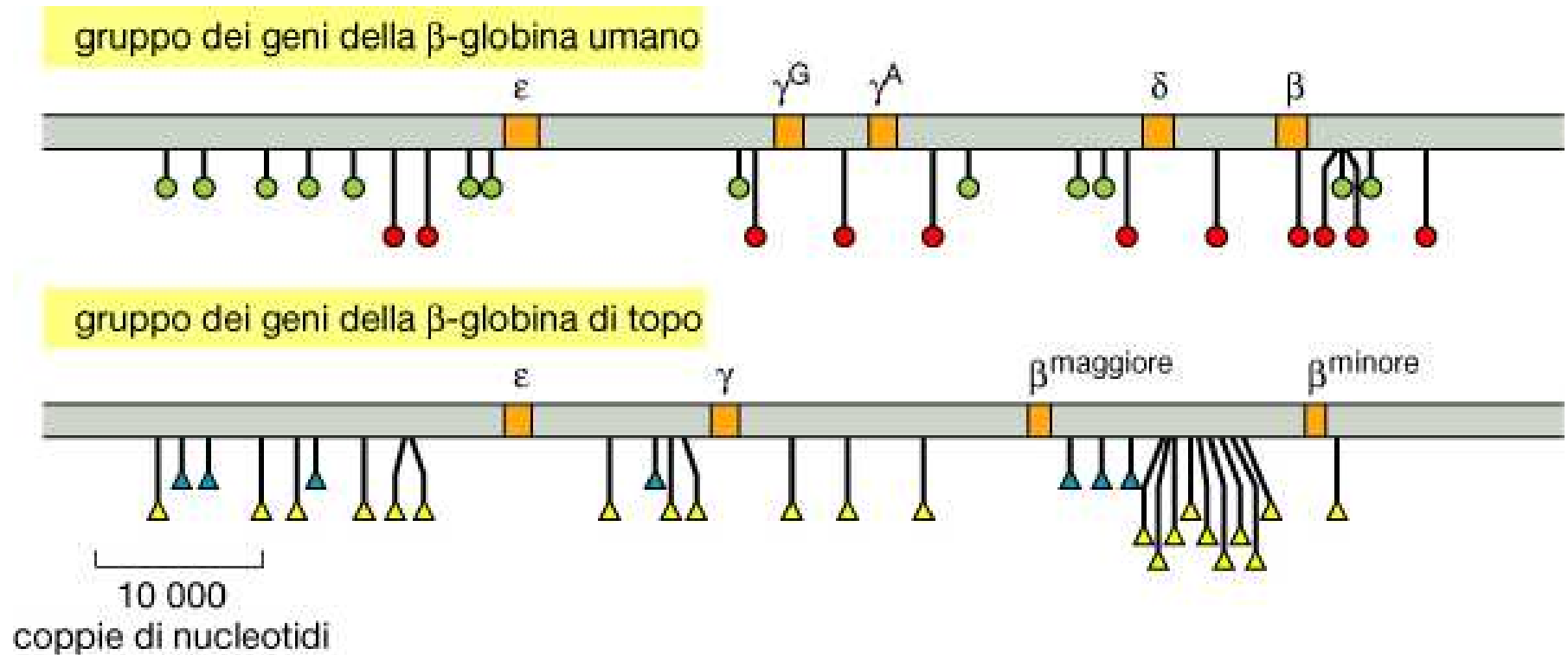


Geni sintenici

Un gruppo di geni viene definito **sintenco** se i singoli componenti, che possono essere strutturalmente molto diversi tra di loro, vengono mantenuti **nello stesso ordine nel genoma di specie diverse**.

La sintenia suggerisce una **relazione funzionale** tra i diversi geni.

Sintenia conservata tra genoma umano e di topo

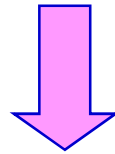


L'ordine dei geni su un cromosoma tende a conservarsi per milioni di anni: questo costituisce il fenomeno della

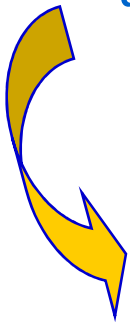
SINTENIA

Duplicazione genica

La duplicazione genica non è sempre seguita dalla divergenza della sequenza e dall'evoluzione di una famiglia di geni con funzioni differenti



In alcune famiglie di geni o sequenze di DNA può esserci maggiore correlazione tra sequenze dei membri della famiglia in una specie (paraloghi) che tra ortologi di specie differenti



EVOLUZIONE CONCERTATA

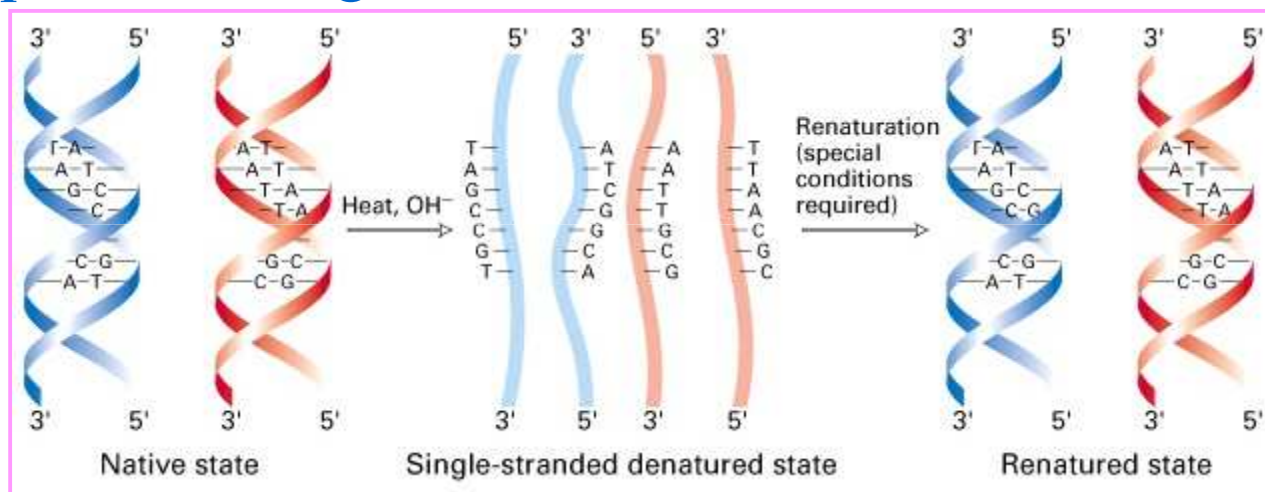
Evoluzione concertata

I membri delle famiglie di geni multipli **tendono ad assomigliarsi** in misura maggiore di quanto ci si aspetti.

Essi evolvono **contemporaneamente** sotto la spinta di meccanismi (**C.O. ineguale, Conversione genica**) che rendono il più somiglianti possibile le sequenze della famiglia

Evoluzione concertata

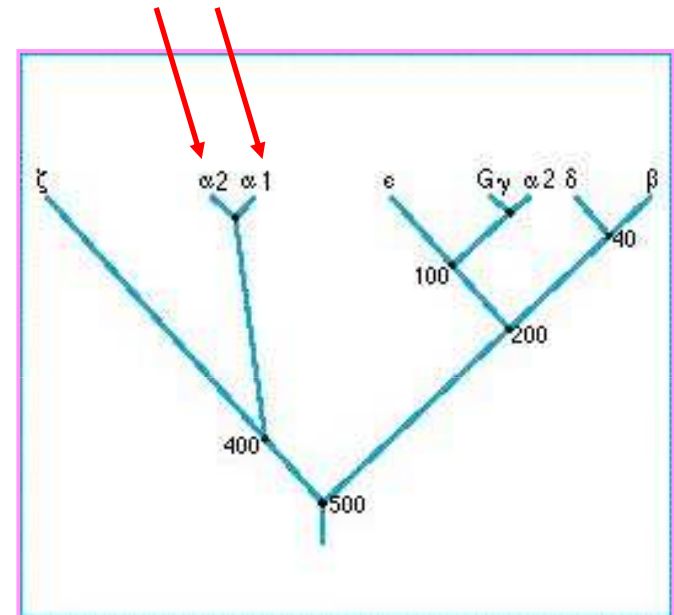
- Lo sviluppo di tecniche di DNA reannealing e di ibridazione (anni '60) conducono all'osservazione che frammenti denaturati di DNA di una specie formano *duplexes* molto più rapidamente di quanto facciano con quelli di specie affini dalle quali, quindi, divergono molto



- Coevoluzione del DNA:** i membri di una famiglia di sequenze evolvono insieme mantenendo alta la similarità divergendo sempre più da quelli di altre specie

L'esempio delle globine

- Tutti i primati hanno 2 α globine. Assumiamo quindi che l'antenato comune dei primati avesse due geni α -globinici.
- Se $\alpha 1$ e $\alpha 2$ della stessa specie si sono separati circa 300 milioni di anni fa dovrebbero aver accumulato molti cambiamenti AA.
- Si osserva un'alta omogeneità intra-specifica: la conclusione è che i geni $\alpha 1$ e $\alpha 2$ non si sono evoluti in modo indipendente, ma in maniera **concertata**.

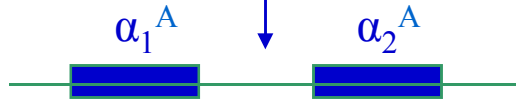


Antenato comune

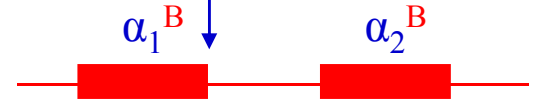


Speciazione

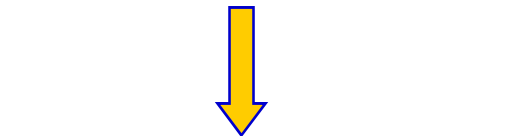
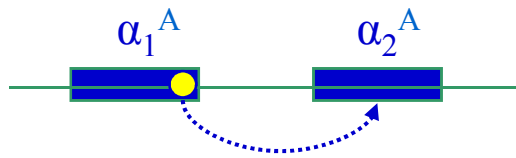
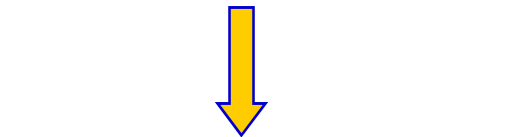
Specie A



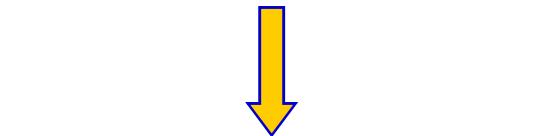
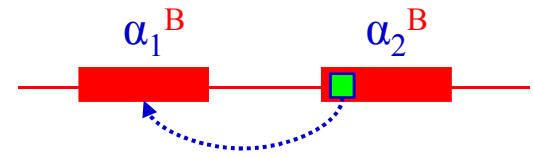
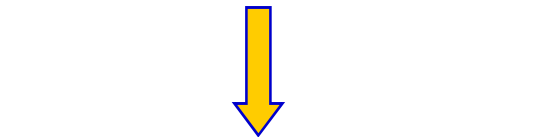
Specie B



Mutazione



Scambio



Evoluzione concertata

L'evoluzione concertata avviene grazie allo scambio tra sequenze di DNA non alleliche all'interno di un genoma.

Diversi meccanismi:

- Crossing-over ineguale
- Scambio ineguale tra cromatidi fratelli
- Conversione genica

spesso sono coinvolte sequenze di DNA ripetute in tandem, questo può produrre una **omogeneizzazione delle sequenze**

Conversione genica

E' una conseguenza del processo ricombinativo.

Il crossing over porta alla formazione di una regione di DNA ibrida costituita dall'unione di due filamenti provenienti da cromatidi non identici.

Se la rottura del DNA avviene in una regione in cui i due cromatidi non portano la stessa informazione, il cromatide ricombinato avrà una regione eteroduplex in cui vi sono appaiamenti errati (AG, AC, TG , TC)

Conversione genica

Gene conversion is a nonreciprocal transfer of genetic information. The following figure illustrates the difference between gene conversion and DNA crossover. An origin of gene conversion is explained in Figure 8-D-5.

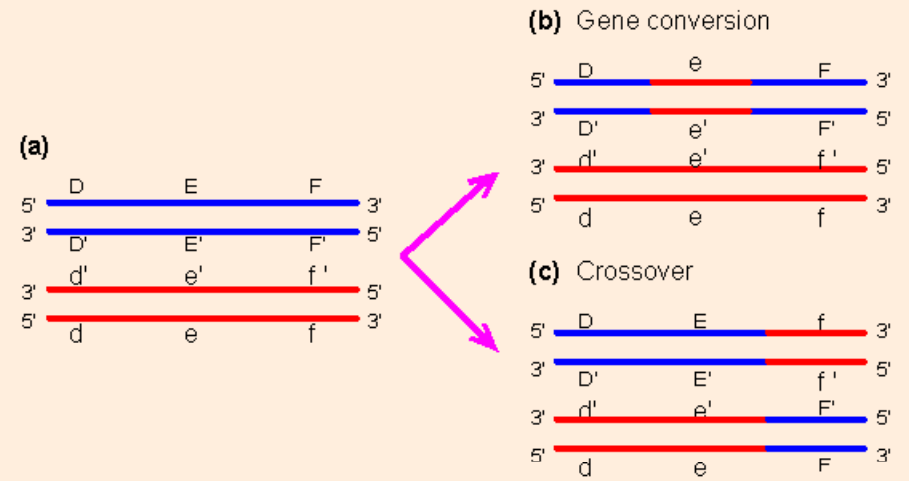
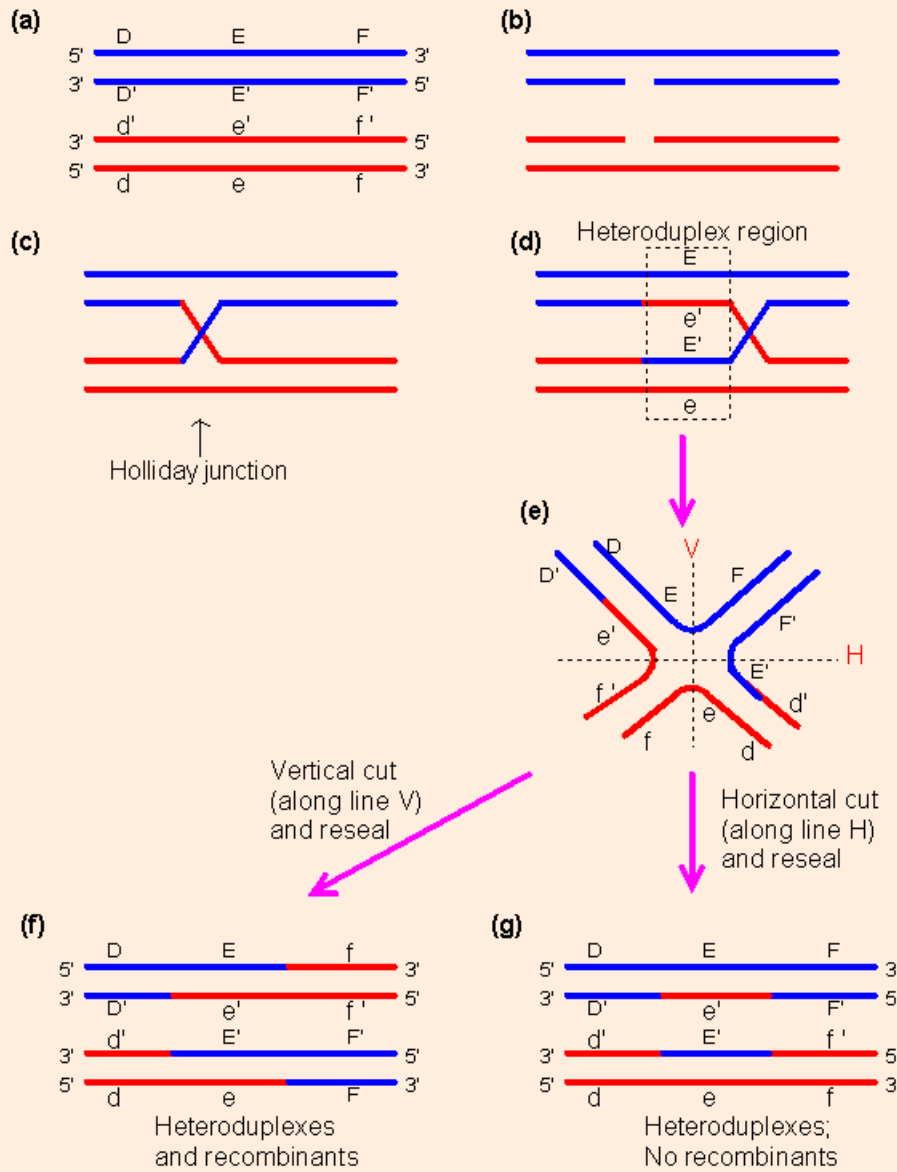


Figure 8-D-4. Comparison between gene conversion and DNA crossover. (a) Two DNA molecules. (b) Gene conversion - the red DNA donates part of its genetic information (e-e' region) to the blue DNA. (c) DNA crossover - the two DNAs exchange part of their genetic information (f-f' and F-F').

Gene conversion is a nonreciprocal transfer of genetic information. The following figure illustrates the difference between gene conversion and DNA crossover. An origin of gene conversion is explained in Figure 8-D-5.

Conversione genica

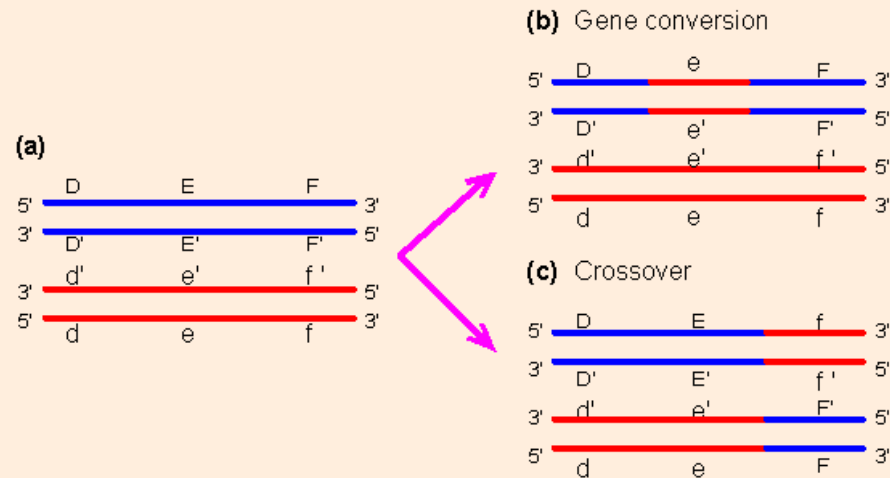


Figure 8-D-4. Comparison between gene conversion and DNA crossover. **(a)** Two DNA molecules. **(b)** Gene conversion - the red DNA **donates** part of its genetic information (e-e' region) to the blue DNA. **(c)** DNA crossover - the two DNAs **exchange** part of their genetic information (f-f' and F-F').

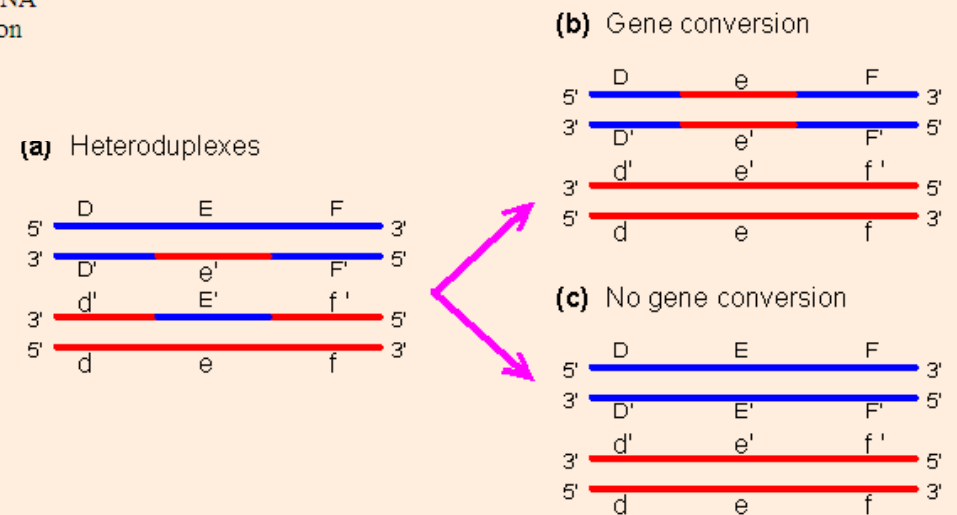
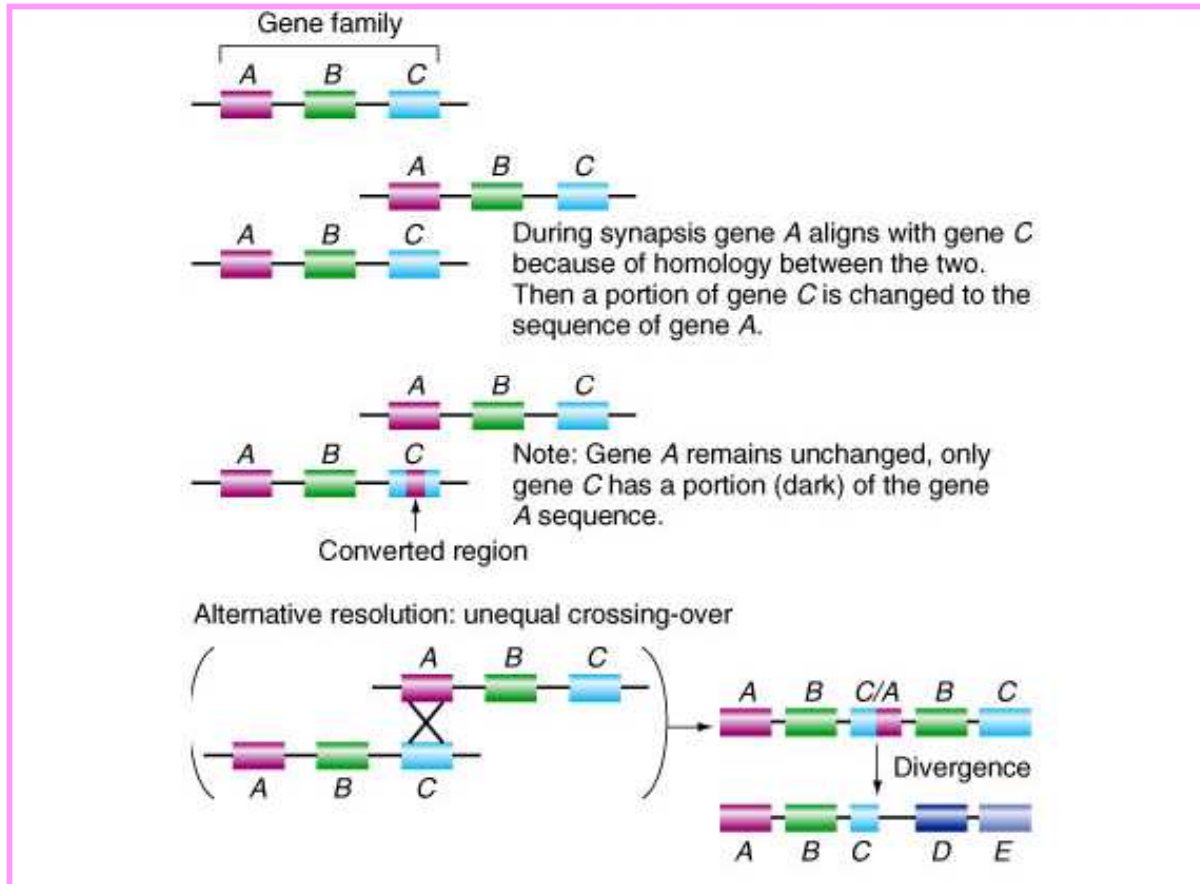


Figure 8-D-5. An origin of gene conversion. **(a)** Heteroduplexes formed by the **resolution of Holliday structure** or by other mechanisms. **(b)** The blue DNA uses the invaded segment (e') as template to "correct" the mismatch, resulting in gene conversion. **(c)** Both DNA molecules use their original sequences as template to correct the mismatch. Gene conversion does not occur.

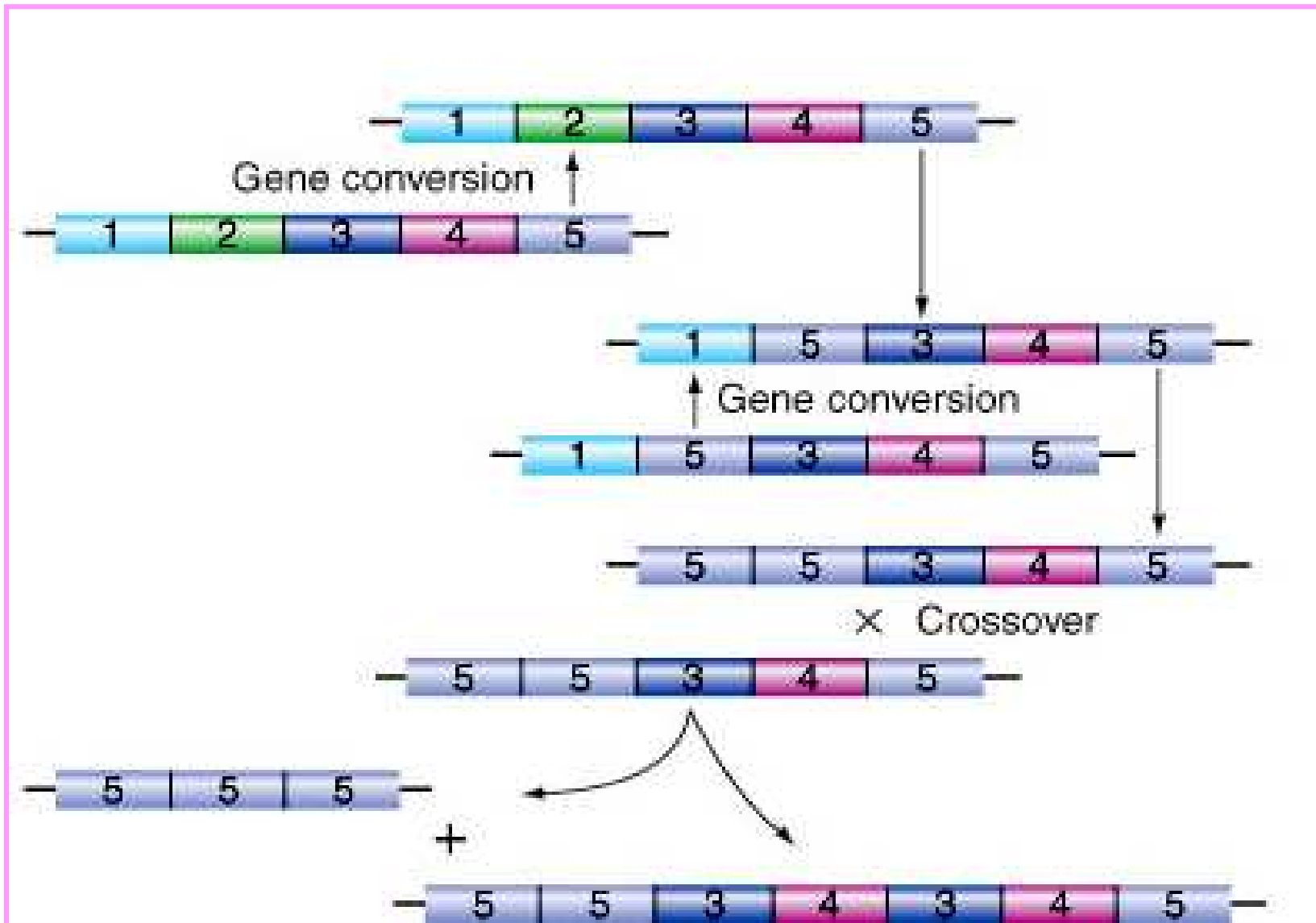
Evoluzione concertata



Conversione genica

Crossing-over
inequale

L'evoluzione concertata può portare ad omogeneizzazione



Conversione genica

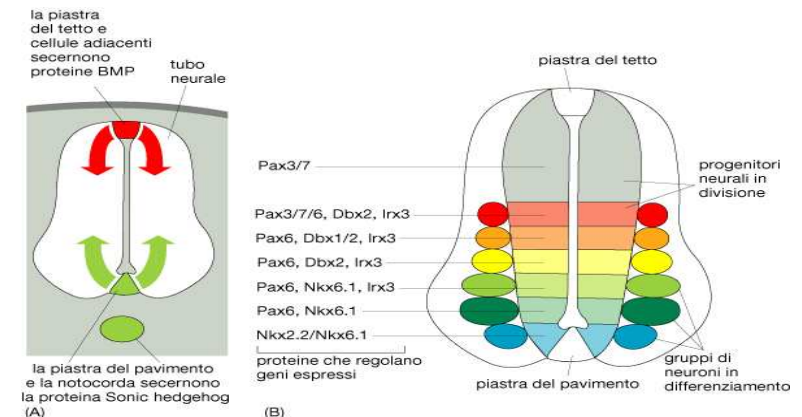
Crossing-over ineguale

Famiglie multigeniche

- famiglie multigeniche classiche
- famiglie geniche che codificano prodotti che condividono **grossi domini** altamente conservati
- famiglie geniche che codificano prodotti con **brevissimi motivi** aminoacidici conservati
- **superfamiglie** geniche

Famiglie geniche che codificano prodotti che condividono grossi domini altamente conservati

- ✓ spesso la porzione rimanente oltre il dominio condiviso non ha alta similarità
- ✓ spesso codificano prodotti con funzioni importanti nello sviluppo
- ✓ **Esempi ne sono:**
 - Geni delle omeobox (30 geni HOX), dominio di 60 aa;
 - Geni PAX (9) la paired box codifica un dominio di 130aa



What are the PAX genes?

The PAX gene family plays a critical role in the **formation of tissues and organs during embryonic development**. The PAX gene family is also important for **maintaining the normal function of certain cells after birth**. To carry out these roles, the PAX genes provide instructions for making proteins that attach to specific areas of DNA and help control the activity (expression) of particular genes. On the basis of this action, PAX proteins are called **transcription factors**.

After birth, the PAX genes are usually turned off (inactivated), but in some tissues the PAX genes are active. These active genes aid in tissue re-growth and protect against cell death caused by cellular stress (such as increased sodium levels in the cell).

There are nine genes in the PAX gene family. These genes are divided into subgroups based on various aspects of similarity. Subgroup I includes PAX1 and PAX9; subgroup II includes PAX2, PAX5, and PAX8; subgroup III includes PAX3 and PAX7; and subgroup IV includes PAX4 and PAX6.

Mutations in PAX genes lead to disorders that involve the incomplete development of tissues in which a particular PAX gene is expressed. Additionally, the overexpression of PAX genes has been noted in a variety of cancers. It is thought that the cell protection function of PAX genes prevents cell death and permits tumor growth (proliferation).

L'evoluzione dei geni HOX offre importanti esempi di geni ortologhi, paraloghi e sintenici

WT



Antennapedia



Geni omeotici

Geni omeotici: geni che regolano l'identità delle regioni corporee. Le mutazioni nei geni omeotici causano la trasformazione di una struttura corporea intera (o di una sua parte) in un'altra.

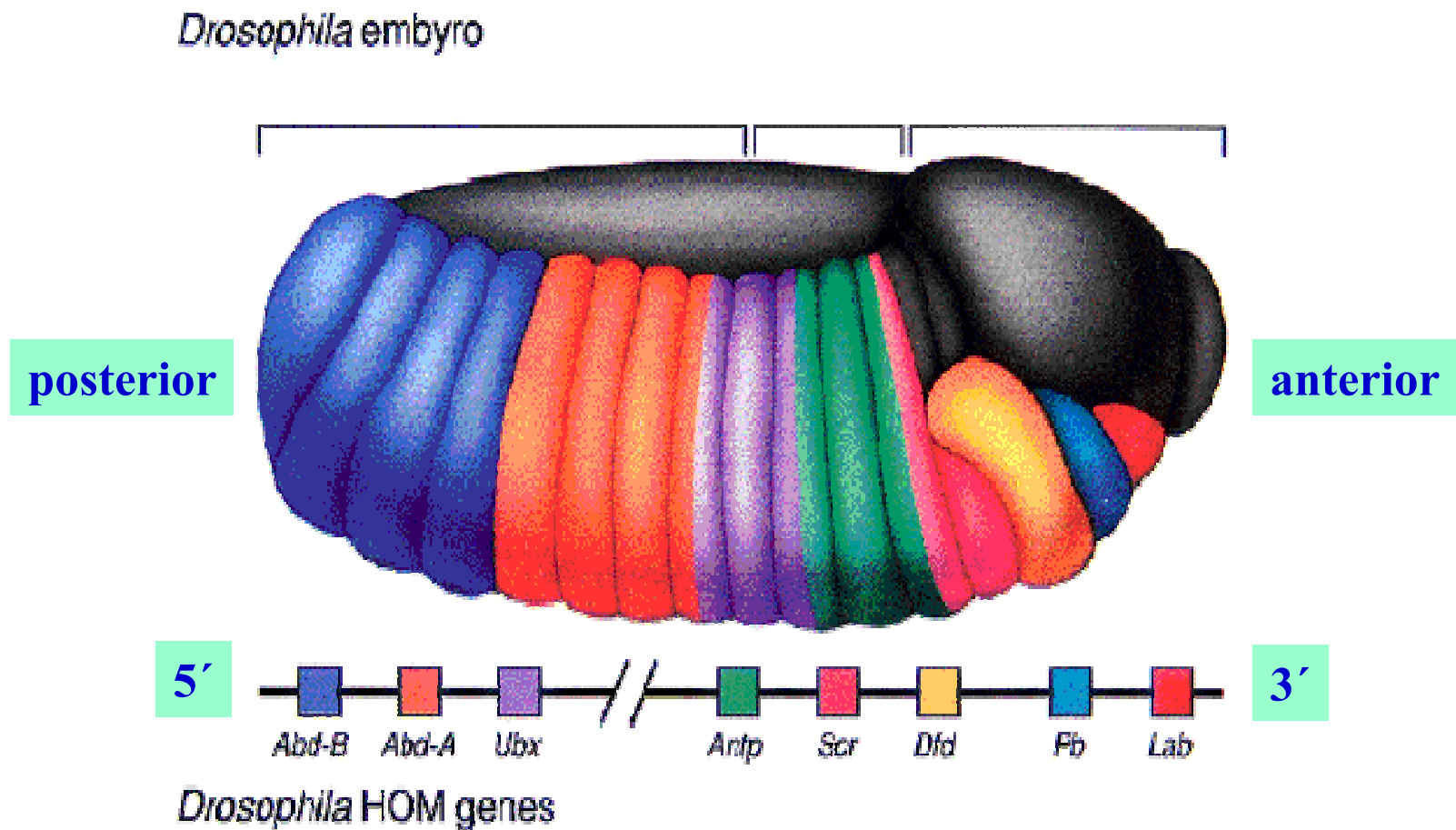
WT



Antennapedia



L'embrione di *Drosophila* si sviluppa in maniera segmentale



Geni HOX

Geni HOX: geni omeotici contenenti un **omeobox** e localizzati in gruppi associati sul cromosoma.

Omeobox: regione di 180 bp di DNA che codifica per una particolare classe di domini proteici che legano il DNA. Esistono circa 20 famiglie di geni contenenti omeobox.

Omeodomini geni HOX di drosofila

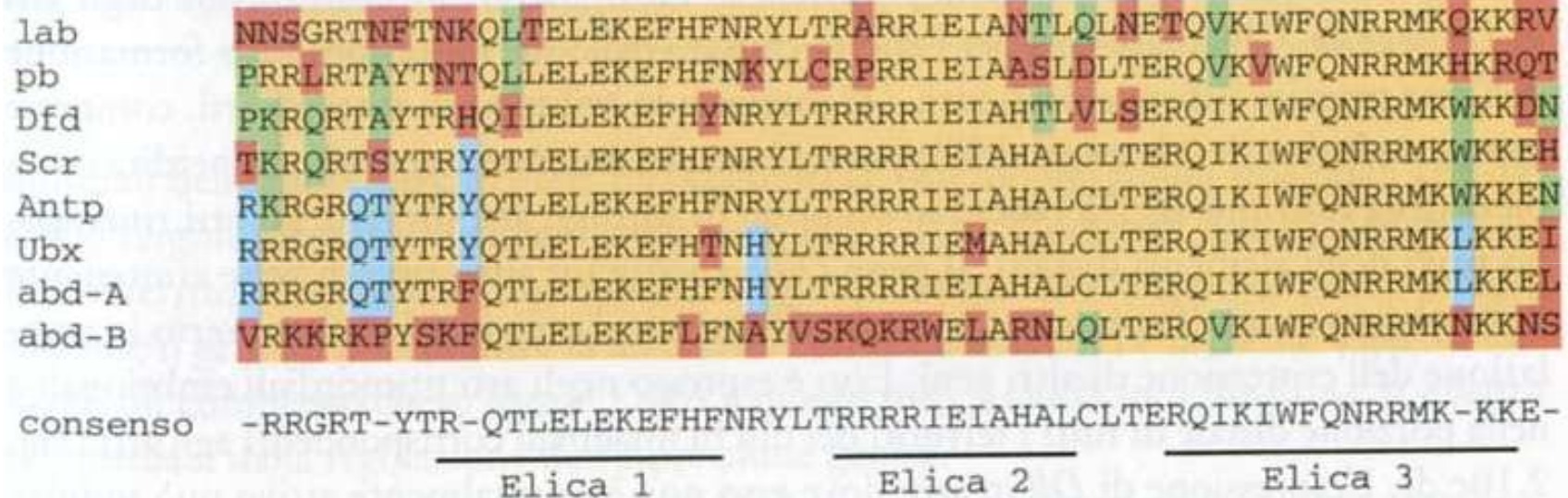
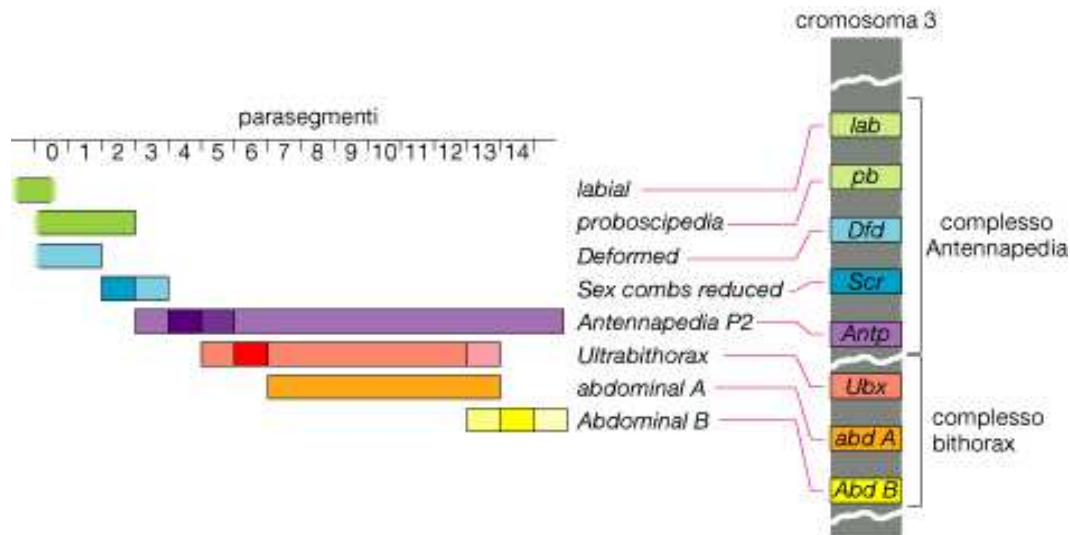


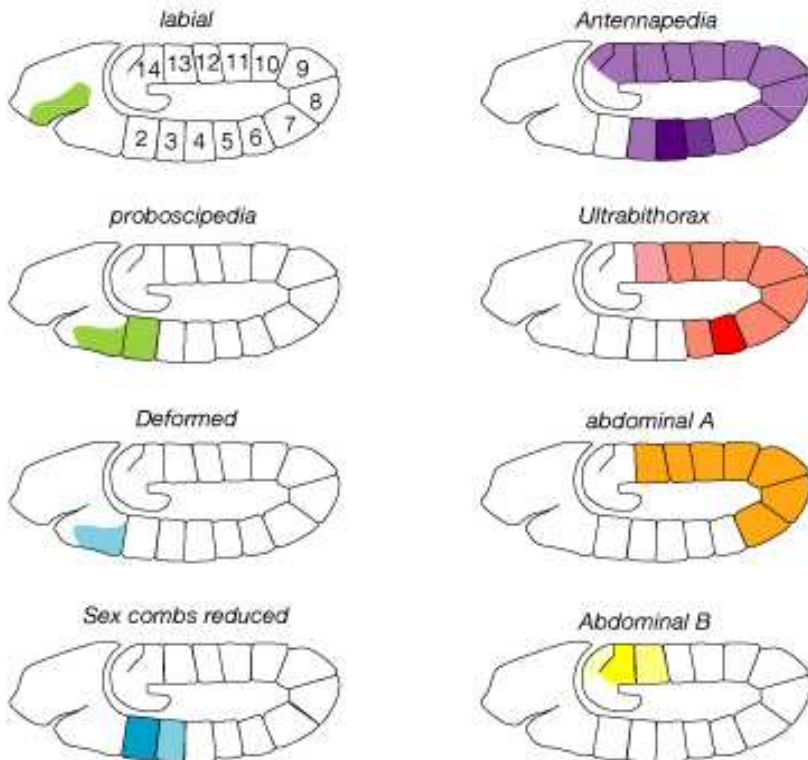
Figura 2.8

Omeodomini dei geni *Hox* di *Drosophila*

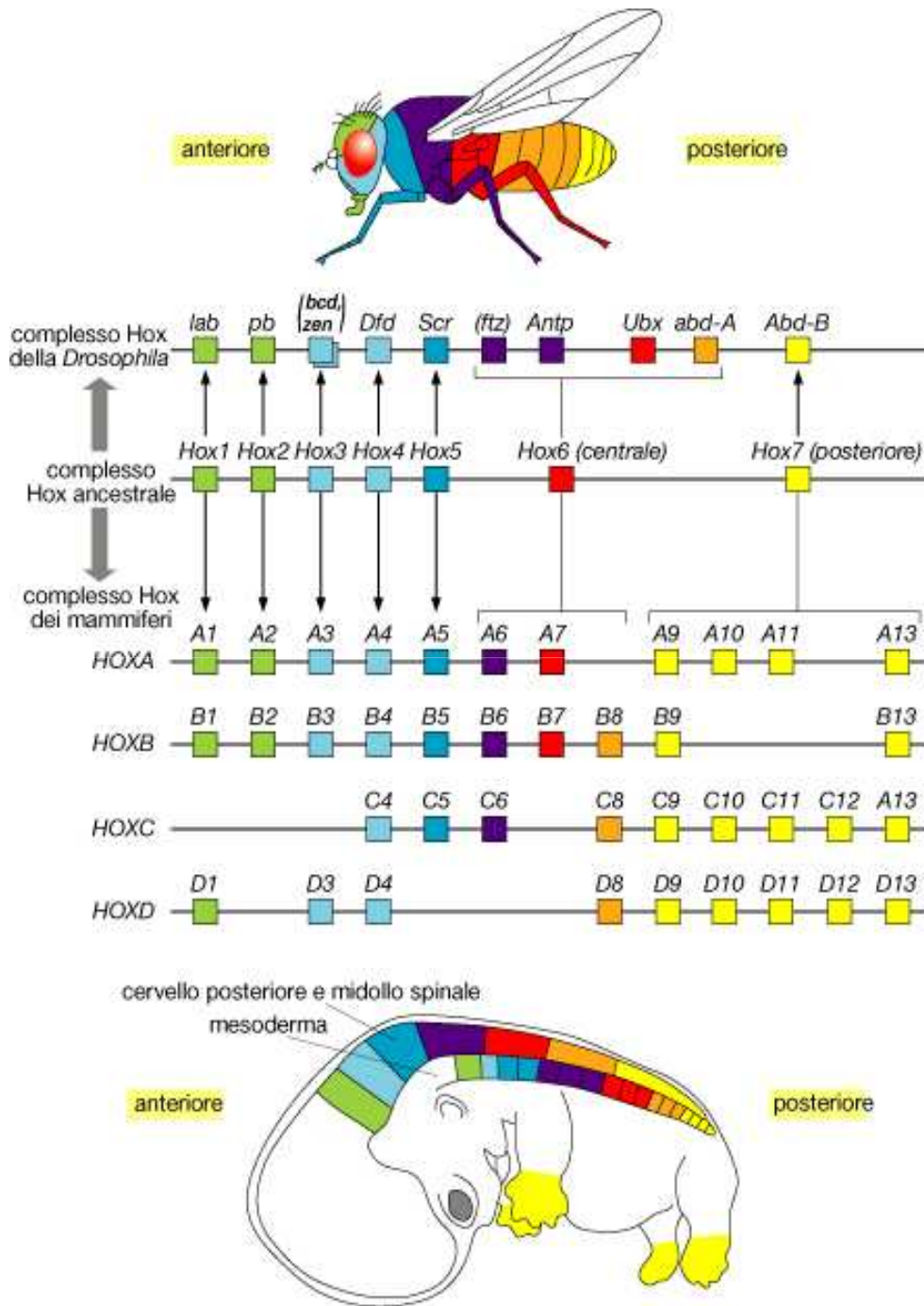
Ciascuno degli otto geni *Hox* di *Drosophila* codifica per proteine che contengono un dominio altamente conservato di 60 aminoacidi che si lega al DNA, l'omeodominio, composto da tre alfa eliche. La terza elica è la più conservata come sequenza. I residui divergenti sono colorati in rosso; quelli condivisi tra sottoinsiemi di proteine sono colorati in azzurro o verde.



La sequenza dei geni sul Cr. 3 di *Drosophila* corrisponde alla sequenza spaziale con cui i geni vengono espressi nei parasegmenti della larva.



La maggior parte dei geni è espressa a livello massimo in un segmento (colore scuro) e ad un livello più basso nei segmenti adiacenti. In regioni in cui i domini di espressione sono sovrapposti è il più “posteriore” dei geni localmente attivi che determina il fenotipo locale.



Si pensa che il complesso si sia originato come segue: nell'antenato comune a vermi, mosche e vertebrati, un singolo gene selettore omeotico primordiale ha subito duplicazioni ripetute originando una serie di questi geni in tandem ("HOX ancestrale").

Nella linea della Drosophila il complesso si è poi diviso nei complessi separati antennapedia e bitorax.

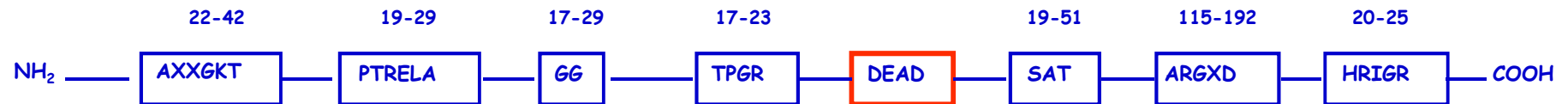
Nella linea che ha portato ai mammiferi, l'intero complesso è stato duplicato ripetutamente originando i 4 complessi HOX

Famiglie multigeniche

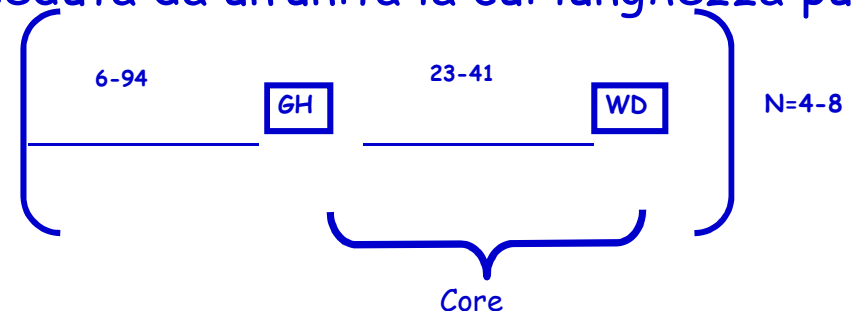
- famiglie multigeniche classiche
- famiglie geniche che codificano prodotti che condividono **grossi domini** altamente conservati
- famiglie geniche che codificano prodotti con **brevissimi motivi** aminoacidici conservati
- superfamiglie geniche

Famiglie multigeniche che codificano prodotti con brevissimi motivi aminoacidici conservati

Famiglia della DEAD box: geni per enzimi coinvolti nella maturazione dell' RNA come inizio della traduzione e splicing . Vi sono 8 motivi aminoacidici conservati compresa la DEADbox: Asp-Glu-Ala-Asp



Famiglia dell'unità ripetuta WD: geni che codificano per prodotti con funzioni diverse, ma tutti coinvolti nella regolazione cellulare (divisione, trascrizione). I prodotti presentano da 4 a 8 unità ripetute in tandem, contenenti un "core" di lunghezza fissa (27-45AA) che termina con il dipeptide WD: Trp-Asp. Il WD è preceduta da un'unità la cui lunghezza può variare nelle diverse ripetizioni



Famiglie multigeniche

- famiglie multigeniche classiche
- famiglie geniche che codificano prodotti che condividono **grossi domini** altamente conservati
- famiglie geniche che codificano prodotti con **brevissimi motivi** aminoacidici conservati
- **superfamiglie** geniche

Superfamiglie

Le famiglie presentano variabilità nell'omologia con la sequenza di riferimento e nel grado con cui le sequenze conservate definiscono la famiglia.

L'appartenenza ad una famiglia è data non solo dall'omologia complessiva, ma dal presentare un dominio condiviso o **organizzazione e funzione simile**

