Biologia Cellulare Molecolare Avanzata

“Functional & regulatory genomics”

1. Complexity of eukaryotic genomes.
2. Sequencing and re-sequencing projects.
3. Basic concepts of gene transcription and regulation
4. Transcriptomes
5. Coding, noncoding and alternative splicing
6. Chromatin as the dynamic environment of genomes.

7. Functional states of chromatin and chromosome territories
8. Epigenetics and gene imprinting
9. Mechanisms and pathways of gene control
10. Evolution of alternative splicing
11. RNA elements that regulate RNA fate and life
12. Other noncoding RNAs and primordial RNA functions

http://biologia.i-learn.unito.it/

Michele De Bortoli
Professor of Molecular Biology
University of Turin

Ogni lezione contiene:

1. ripasso concetti di base (dai corsi del triennio)
2. definizioni di concetti e spiegazioni della metodologia e sperimentazione che ha permesso di chiarirli (slides)
3. studio di uno (o due) lavori scientifici sull'argomento
   (solo alcuni, segnalati e messi a disposizione, devono essere studiati come tali)

Il materiale didattico disponibile a **http://biologia.i-learn.unito.it/:**

1. le slides proiettate a lezione
2. *reviews* da cui sono stati presi gli argomenti
3. lavoro (i) studiato a lezione

Esercizi e tasks sul sito Moodle sono parte del vostro studio!

Farli subito→acquisire conoscenze→comprensione delle lezioni successive !!

Gli studenti devono:

1. conoscere i concetti di **base**
2. conoscere i concetti **specifici** trattati
3. avere compreso metodologie ed approcci **sperimentali**

Prove di profitto:

1) discussione di un **articolo** scientifico recente su uno degli argomenti trattati (durante il corso, a coppie)
2) esame **orale** con discussione sull'intero programma

These background are needed:

1. - Basic Molecular Biology & Genetics

✓DNA replication
✓Transcription
✓Post-transcriptional RNA processing
✓Translation
✓Post-translational protein modification
✓Gene expression regulation (basic mechanisms)
✓Basics of protein structure and molecular representations

Example: Chapters 1 through 10 from “Essential Cell Biology” 2° or 3° Edition – Alberts et al., Garland, 2004 (2°), 2009 (3°)
Italian version – Zanichelli (2005)

2. Recombinant DNA methodology:

- ✓ DNA replication (in vivo and in vitro)
- ✓ PCR, rt-PCR and real-time PCR
- ✓ Basic DNA cloning in plasmids and other vectors
- ✓ Libraries, clones, colonies, storage, propagation, analysis.
- ✓ DNA sequencing, restriction, Southern blot
- ✓ RNA analysis, Northern blot, RNase protection

3. Basic bioinformatics:

- ✓ Database organization
- ✓ Finding gene and protein sequences
- ✓ Basic alignment protocols

Small green spots like the following may appear in the slides:

Sex determination in *D. melanogaster*

These refer to basic knowledge revision at **your charge**

1990-2003	Il Progetto Genoma Umano
2001	Il consorzio HGP e Celera rilasciano la prima bozza del 95% del genoma dell'Uomo
2003	La sequenza è completata
2001-2008	Vengono iniziate (e alcune completate) le sequenze di molti altri genomi di Vertebrati, pesci, uccelli, piante, funghi, batteri
Oggi	Numerosissimi sequenziamenti sono in corso.

Sito NCBI Genomes Eukaryotic (Mammals)

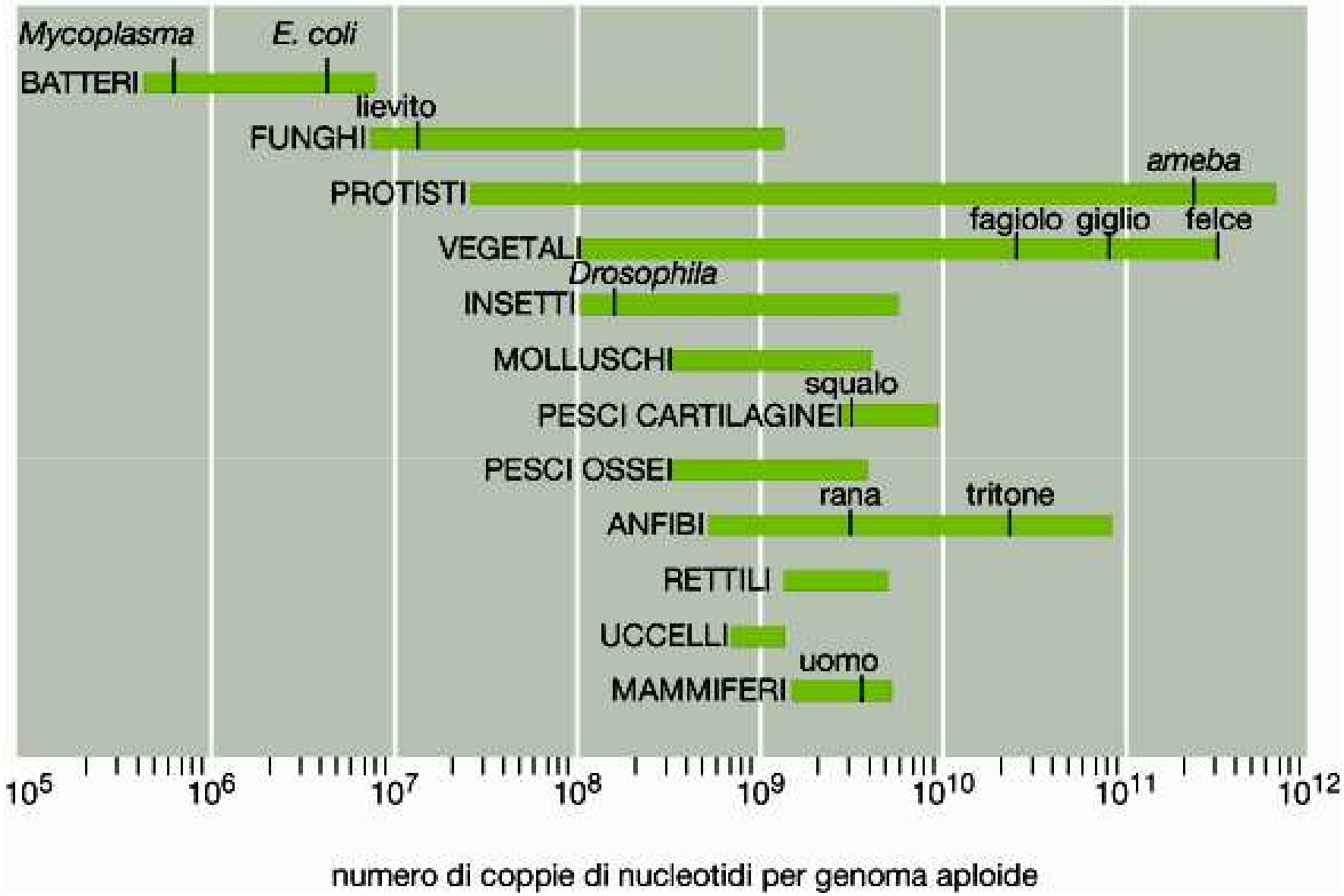
Hundreds of genomes sequenced

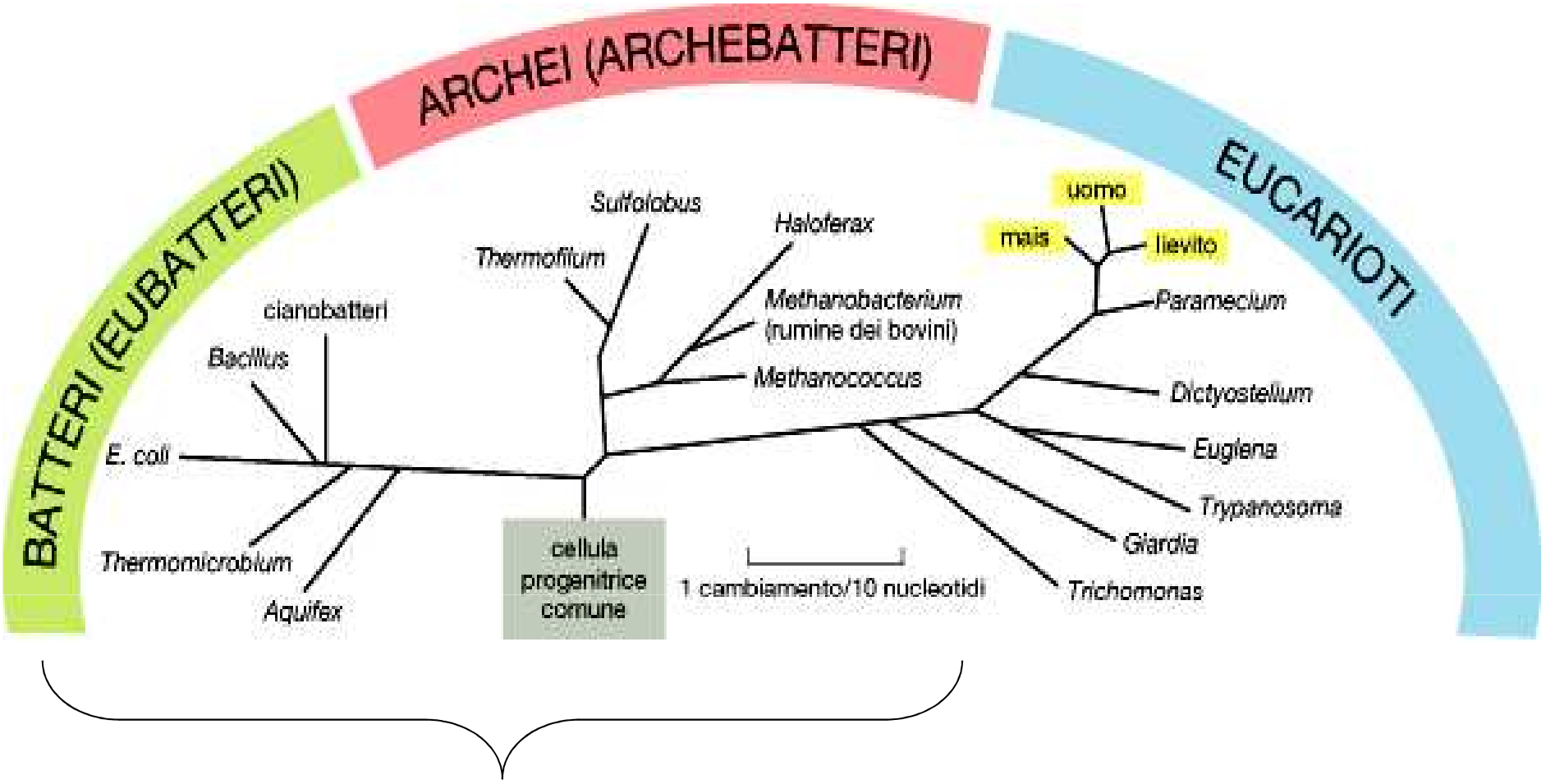
NCBI site Genomic Data
http://www.ncbi.nlm.nih.gov/projects/genome/guide/

How many genomes?

Complete updates for genomic projects (comparative):

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


Mycoplasma
E. coli
BATTERI
lievito
FUNGHI
ameba
PROTISTI
fagiolo
giglio
felce
VEGETALI
Drosophila
INSETTI
MOLLUSCHI
squalo
PESCI CARTILAGINEI
PESCI OSSEI
rana
tritone
ANFIBI
RETTILI
UCCELLI
uomo
MAMMIFERI
10^5
10^6
10^7
10^8
10^9
10^{10}
10^{11}
10^{12}
numero di coppie di nucleotidi per genoma aploide



Da 1000 a 4000 geni

Fino a 30-40,000 geni

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
BACTERIA				
Mycoplasma genitalium	has one of the smallest of all known cell genomes	human genital tract	580	468
Synechocystis sp.	photosynthetic, oxygen-generating (cyanobacterium)	lakes and streams	3573	3168
Escherichia coli	laboratory favorite	human gut	4639	4289
Helicobacter pylori	causes stomach ulcers and predisposes to stomach cancer	human stomach	1667	1590
Bacillus anthracis	causes anthrax	soil	5227	5634
Aquifex aeolicus	lithotrophic; lives at high temperatures	hydrothermal vents	1551	1544
Streptomyces coelicolor	source of antibiotics; giant genome	soil	8667	7825
Treponema pallidum	spirochete; causes syphilis	human tissues	1138	1041
Rickettsia prowazekii	bacterium most closely related to mitochondria; causes typhus	lice and humans (intracellular parasite)	1111	834
Thermotoga maritima	organotrophic; lives at very high temperatures	hydrothermal vents	1860	1877

Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Table 1-1 (part 1 of 2) *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

Table 1–1 Some Genomes That Have Been Completely Sequenced

SPECIES	SPECIAL FEATURES	HABITAT	GENOME SIZE (1000s OF NUCLEOTIDE PAIRS PER HAPLOID GENOME)	ESTIMATED NUMBER OF GENES CODING FOR PROTEINS
ARCHAEA				
Methanococcus jannaschii	lithotrophic, anaerobic, methane-producing	hydrothermal vents	1664	1750
Archaeoglobus fulgidus	lithotrophic or organotrophic, anaerobic, sulfate-reducing	hydrothermal vents	2178	2493
Nanoarchaeum equitans	smallest known archaean; anaerobic; parasitic on another, larger archaean	hydrothermal and volcanic hot vents	491	552
EUCARYOTES				
Saccharomyces cerevisiae (budding yeast)	minimal model eucaryote	grape skins, beer	12,069	~6300
Arabidopsis thaliana (Thale cress)	model organism for flowering plants	soil and air	~142,000	~26,000
Caenorhabditis elegans (nematode worm)	simple animal with perfectly predictable development	soil	~97,000	~20,000
Drosophila melanogaster (fruit fly)	key to the genetics of animal development	rotting fruit	~137,000	~14,000
Homo sapiens (human)	most intensively studied mammal	houses	~3,200,000	~24,000

Genome size and gene number vary between strains of a single species, especially for bacteria and archaea. The table shows data for particular strains that have been sequenced. For eucaryotes, many genes can give rise to several alternative variant proteins, so that the total number of proteins specified by the genome is substantially greater than the number of genes.

Table 1-1 (part 2 of 2) *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

TABELLA 5.6 Caratteristiche generali dei genomi sequenziati

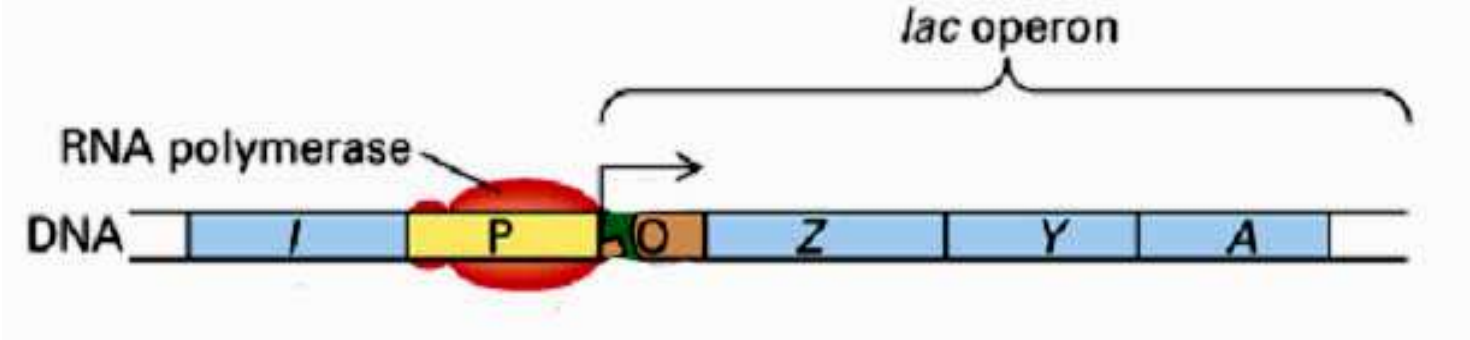
Organismo	Dimensione del genoma (Mb)[a]	Numero di geni	Sequenze codificanti per proteine
Batteri			
Mycoplasma genitalium	0.58	470	88%
H. influenzae	1.8	1743	89%
E. coli	4.6	4288	88%
Lieviti			
S. cerevisiae	12	6000	70%
S. pombe	12	4800	60%
Invertebrati			
C. elegans	97	19,000	25%
Drosophila	180	13,600	13%
Piante			
Arabidopsis thaliana	125	26,000	25%
Riso	390	37,000	12%
Pesci			
Pesce palla	370	20,000–23,000	10%
Uccelli			
Pollo	1000	20,000–23,000	3%
Mammiferi			
Uomo	3200	20,000–25,000	1.2%

[a]Mb = milioni di coppie di basi.

procarioti

Specie	Molecole di DNA	Dimensione (Mb)	Numero di geni
Escherichia coli K-12	1 circolare	4,639	4397
Vibrio cholerae	2 circolari		
	cromosoma princ.	2,961	2770
	Megaplasmide	1,073	1115
Deinococcus radiodurans	4 circolari		
	Cromosoma 1	2,649	2633
	Cromosoma 2	0,412	369
	Megaplasmide	0,177	145
	Plasmide	0.046	40
Borrelia burgdoriferi	7-8 circolari + 11 lineari		
	Cromosoma lineare	0,911	853
	Plasmide circolare cp9	0,009	12
	Plasmide circolare cp26	0,026	29
	Plasmide circolare cp32	0,032	sconosciuti
	Plasmide lineare lp17	0,017	25
	Plasmide lineare lp25	0,024	32
	Plasmide lineare lp28-1	0,027	32
	Plasmide lineare lp28-2	0,030	34
	Plasmide lineare lp28-3	0,029	41
	Plasmide lineare lp28-4	0,027	43
	Plasmide lineare lp36	0,037	54
	Plasmide lineare lp38	0,039	52
	Plasmide lineare lp54	0,054	76
	Plasmide lineare lp17	0,056	sconosciuti

Lactose operon in E. coli illustrates the concept of polycistronic mRNA



LacZ: β-galactosidase cleaves lactose to galactose + glucose

Lac Y: lactose permease, pumps in lactose against electrochemical gradients

LacA: Thiogalactoside transacetylase.

Lac I: lac repressor

Gene structure in Prokaryotes

1° - Genes are “discontinuous”

Eukaryotes

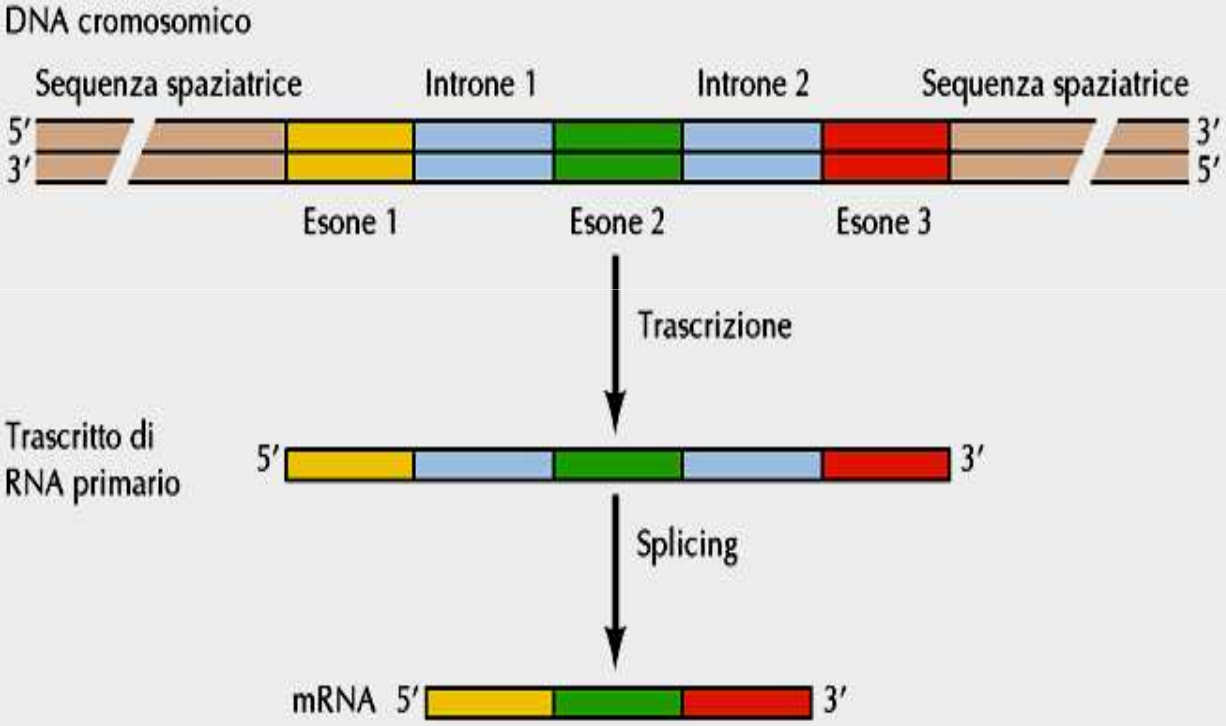
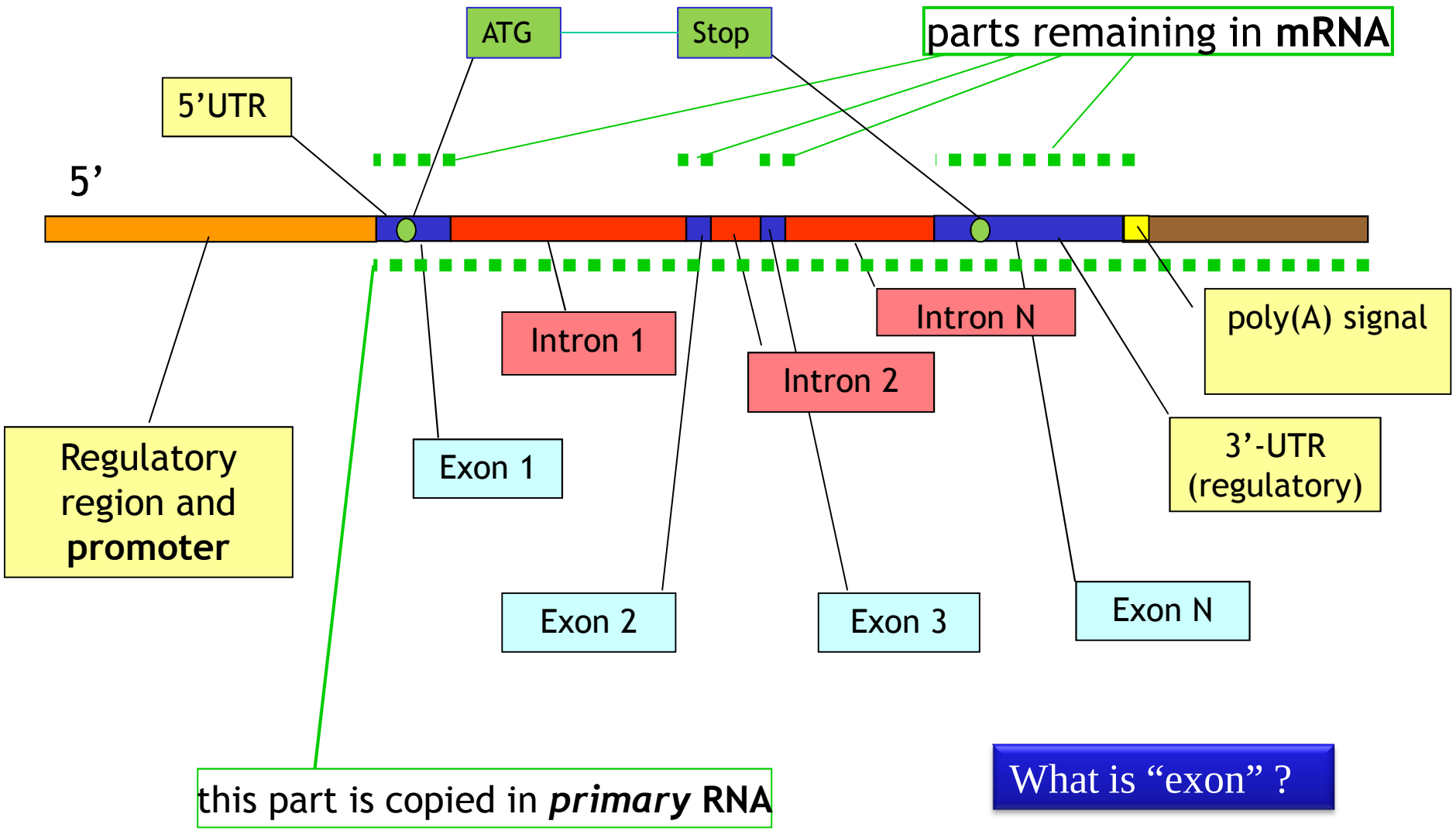
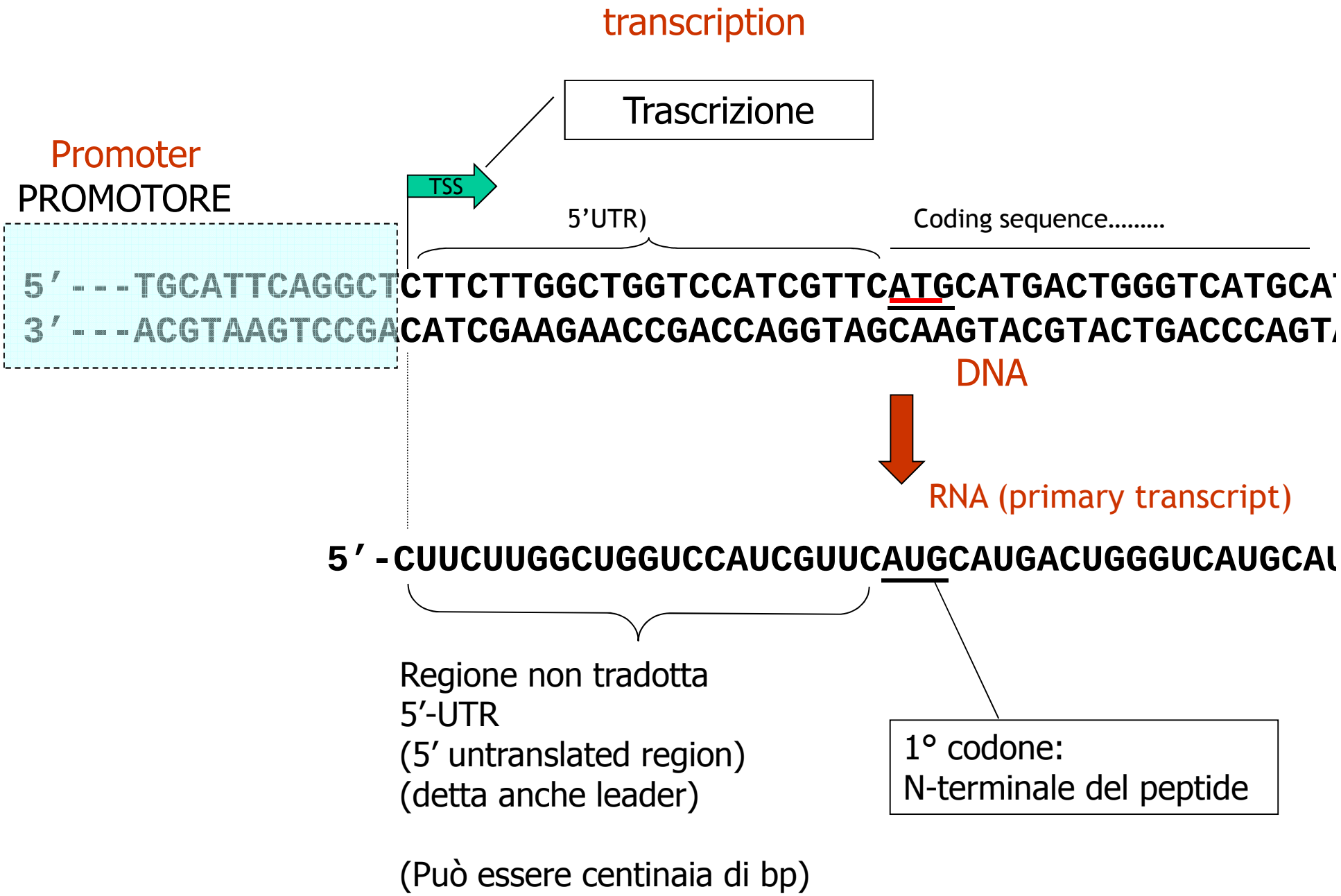


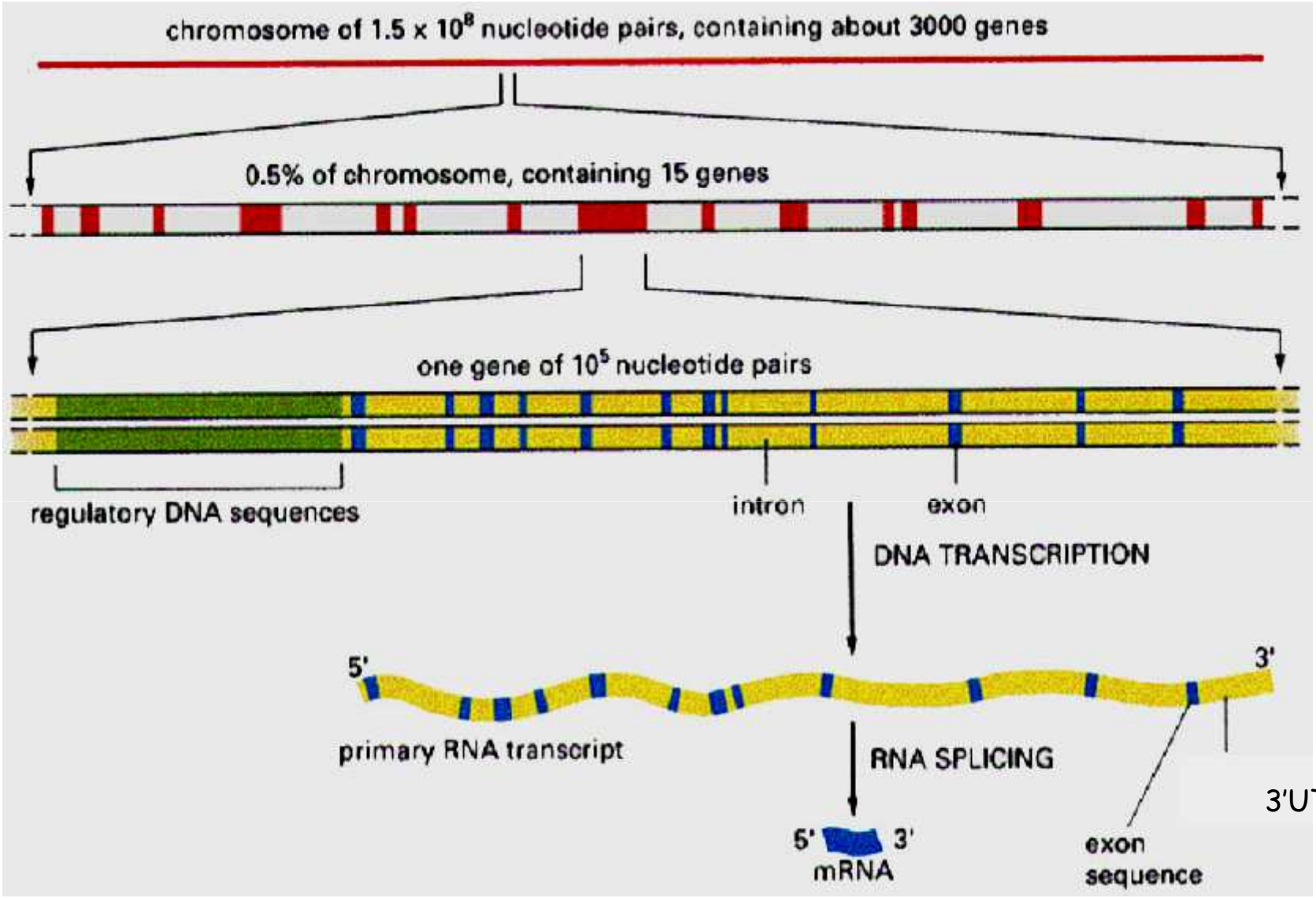
FIGURA 5.2 La struttura dei geni eucariotici La maggior parte dei geni eucariotici è costituita da segmenti di sequenze codificanti (esoni) interrotti da sequenze non codificanti (introni). Sia gli esoni che gli introni vengono trascritti per ottenere un lungo trascritto di RNA primario o precursore. Gli introni sono quindi rimossi attraverso il meccanismo dello splicing per formare il trascritto maturo di mRNA.

Eukaryotes: monocistronic

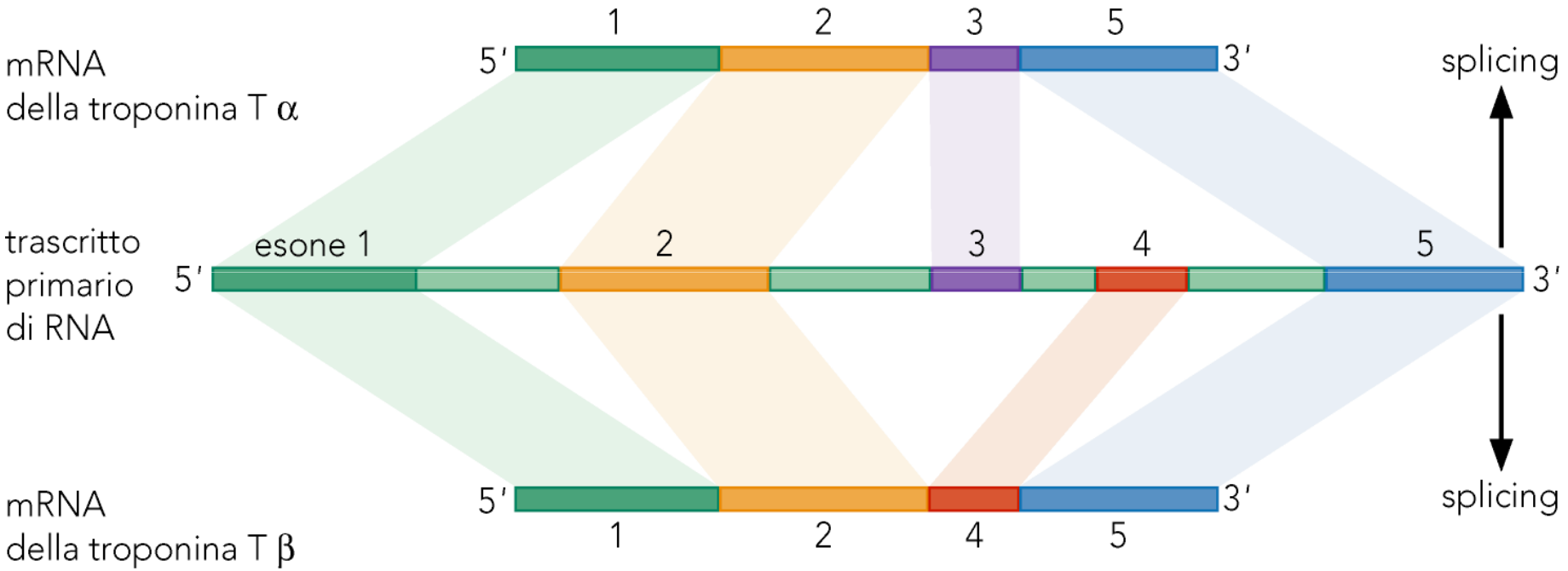


What is "exon" ?


transcription
Trascrizione
Promoter
PROMOTORE
TSS
5'UTR)
Coding sequence.....
5'---TGCATTCAGGCTCTTCTTGGCTGGTCCATCGTTCATGCATGACTGGGTCATGCA
3'---ACGTAAGTCCGACATCGAAGAACCGACCAGGTAGCAAGTACGTACTGACCCAGT
DNA
RNA (primary transcript)
5'-CUUCUUGGCUGGUCCAUCGUUCAUGCAUGACUGGGUCAUGCAU
Regione non tradotta
5'-UTR
(5' untranslated region)
(detta anche leader)
(Può essere centinaia di bp)
1° codone:
N-terminale del peptide


chromosome of 1.5 x 10^8 nucleotide pairs, containing about 3000 genes
0.5% of chromosome, containing 15 genes
one gene of 10^5 nucleotide pairs
regulatory DNA sequences
intron
exon
DNA TRANSCRIPTION
5'
3'
primary RNA transcript
RNA SPLICING
5'
3'
mRNA
exon sequence
3'UTR

In eukaryotes, the peculiar organization of coding sequences in exons, with intervening noncoding intron sequences, allows **alternative splicing** i.e. some exon are either included or excluded from final mRNA, producing different coding sequences and different pepetides from the same gene.



Watson et al., BIOLOGIA MOLECOLARE DEL GENE,
Zanichelli editore S.p.A.

Come sono sistemati i geni in un genoma ?

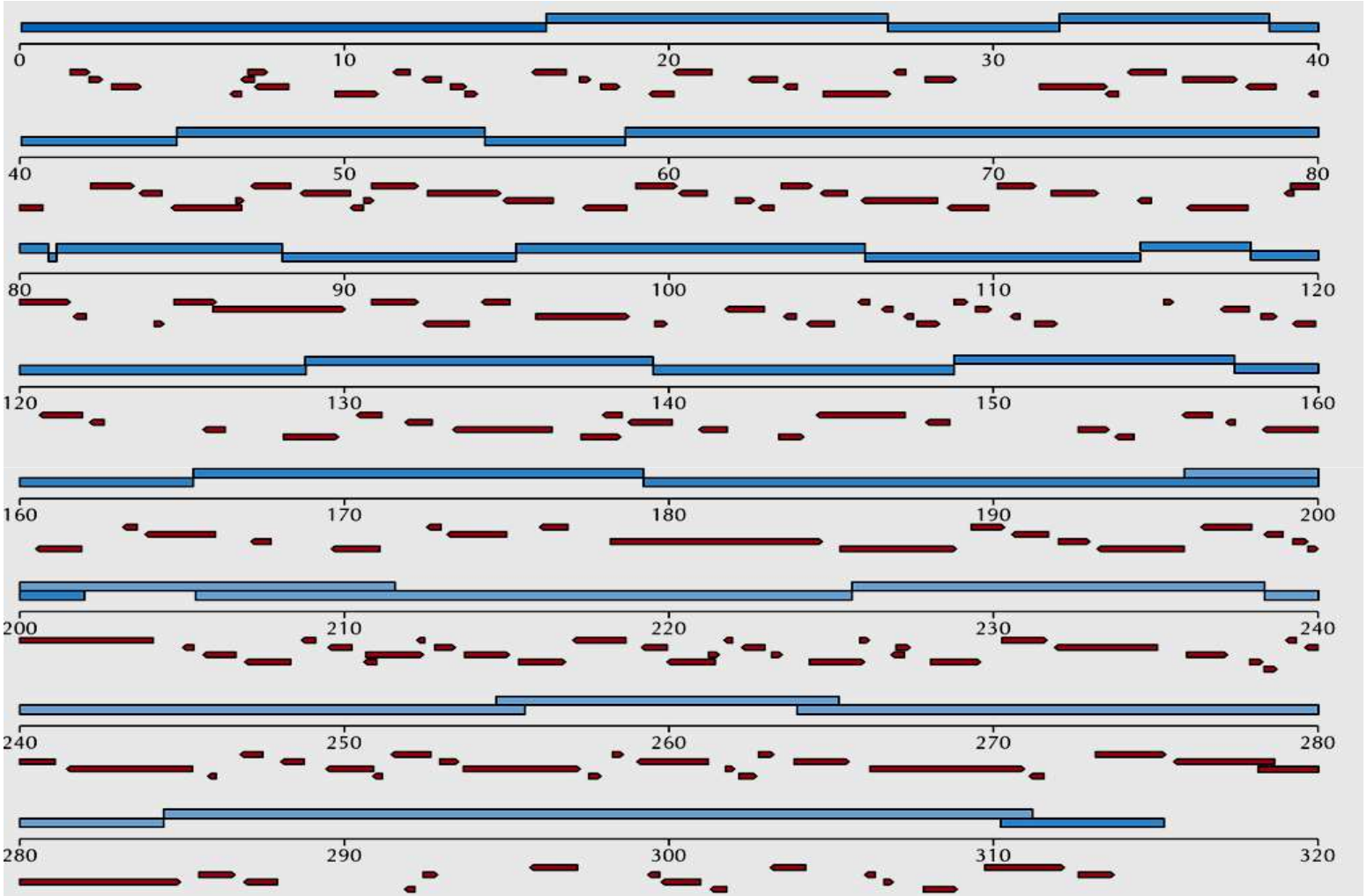


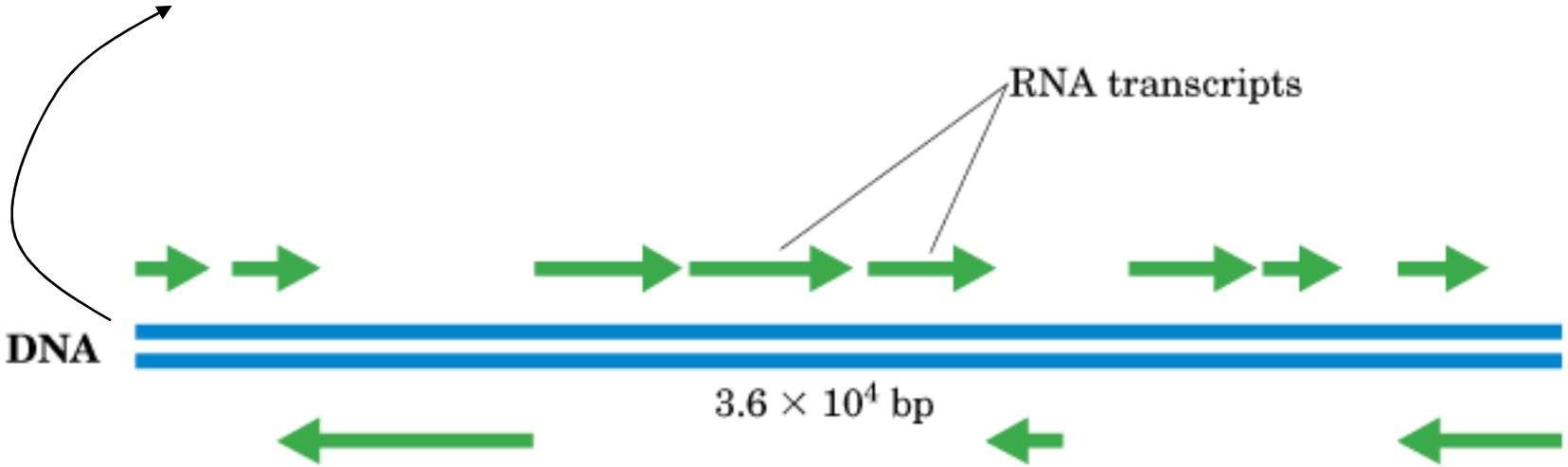
FIGURA 5.24 Il cromosoma III di lievito Le barre blu in alto indicano i cloni utilizzati per il sequenziamento del DNA. Le open-reading frames sono rappresentate da frecce. (Da S.G. Oliver et al., 1992. *Nature* 357: 38).

IMPORTANTE !

La distinzione tra filamento codificante e filamento stampo è **locale** !

Su un cromosoma, i geni sono orientati nei due sensi, senza apparente logica.

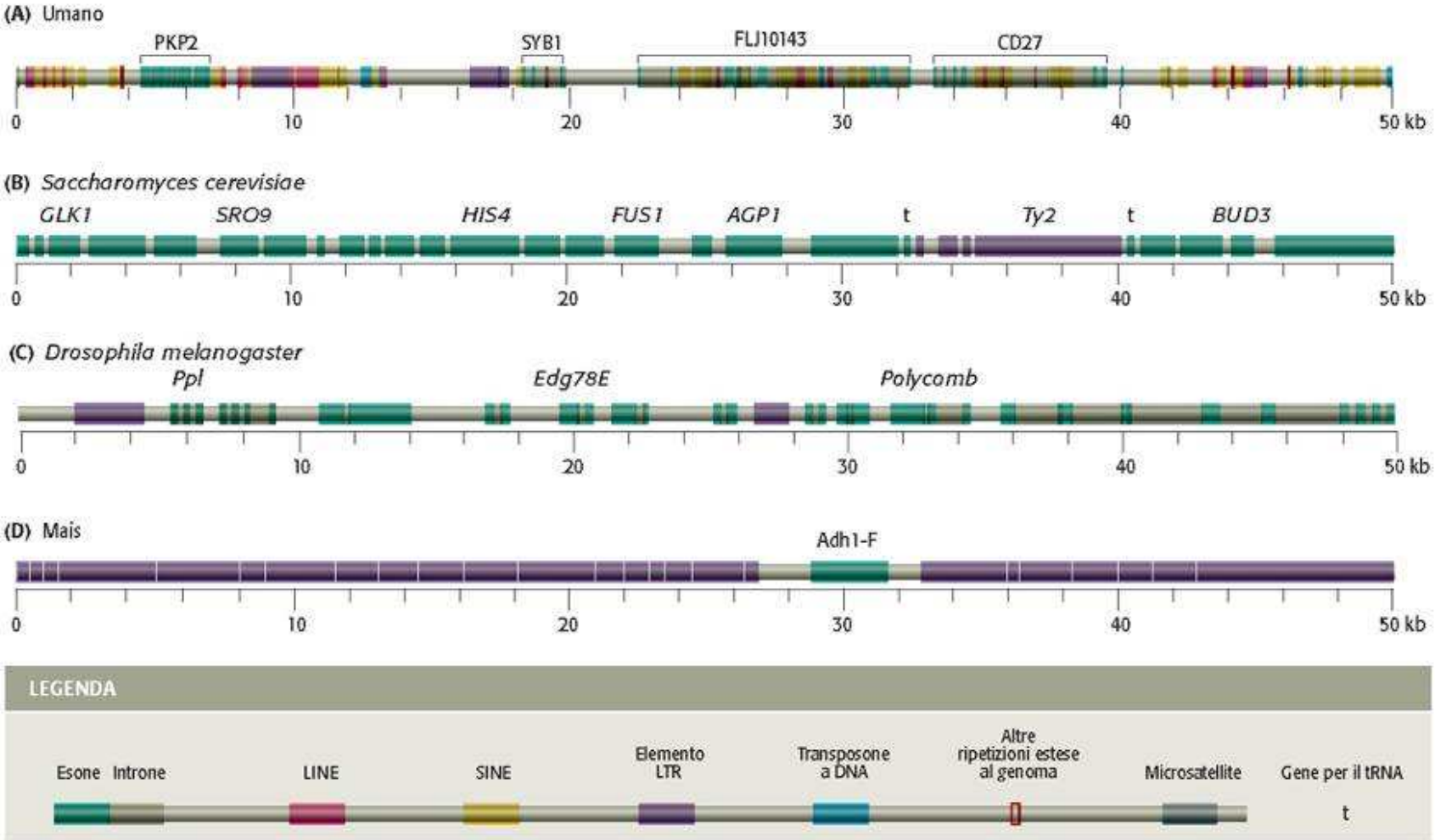
Nel database, la sequenza del cromosoma è relativa ad un filamento, indipendentemente dall'orientamento dei geni.



Le sequenze dei diversi mRNA o RNA funzionali, invece, sono sempre relative al filamento codificante

T.A. Brown
Genomi, III Ed.
EdiSES

Figura 7.15 Confronto tra genoma umano, di lievito, del moscerino della frutta e di mais. (A) Il segmento di 50 kb del cromosoma 12 umano mostrato precedentemente, è confrontato con segmenti di 50 kb derivanti da genomi di (B) *S. cerevisiae*; (C) *Drosophila melanogaster*; (D) mais.



2° - Densità genica

Numero di geni nell'unità di misura del DNA

Diversi organismi (diversi genomi) presentano una diversa distribuzione di geni e di DNA ripetitivo o spaziatore.

	Lievito	moscerino	uomo
Densità genica (geni/Mb)	496	76	11
Introni per gene	0,04	3	9
% di genoma ripetitivo	3,4%	12%	44%

Predicted ORF products mean size in completely sequenced organisms

Organis	size(Mb)	Mean	std	ORFs	min	Max	Tot. aa
SC	1.3	458.8	362.3	6213	25	4910	2850290
CE	97	423.3	371.6	19099	4	7829	8096713
DM	170	497.7	451.2	13695	5	7182	6816125
ATH	100	439.4	318.4	22671	8	5079	9960638
CA		479.6	333.9	6169	21	4162	2958521
HS*	3000	481.4	426.3	21724	16	6669	10484673
SP	15	456.9	353.8	3579	13	4717	1635306
PF+	100	768.9	760	421	54	4981	322400

Average a.a. ~ 150 Da

Il genoma umano 3,200 Mb

23 cromosomi (x2)

The Human Genome Project

Animated tutorials on the Human Genome Project:

http://www.genome.gov/Pages/EducationKit/

(free downloads or on-line view)

How to sequence a genome

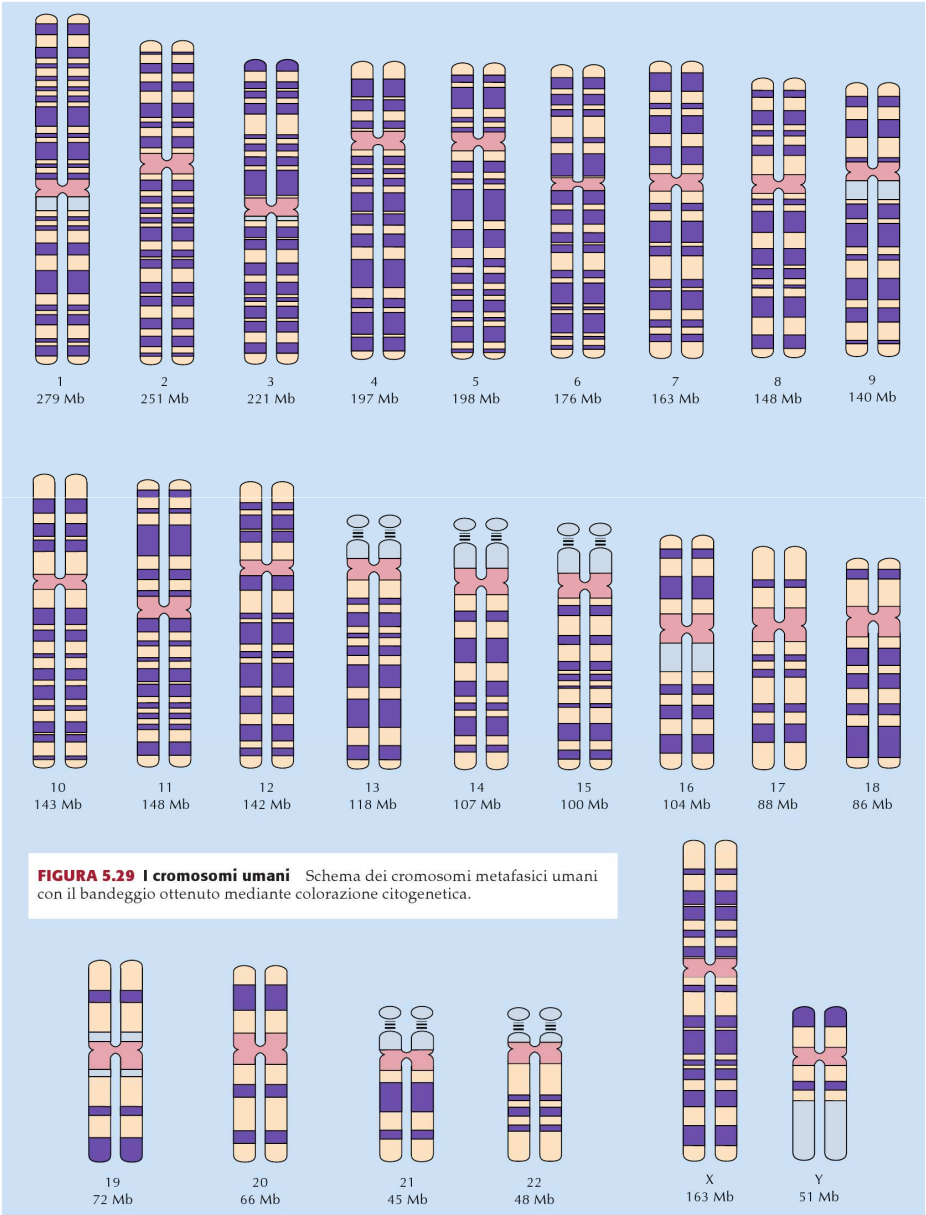
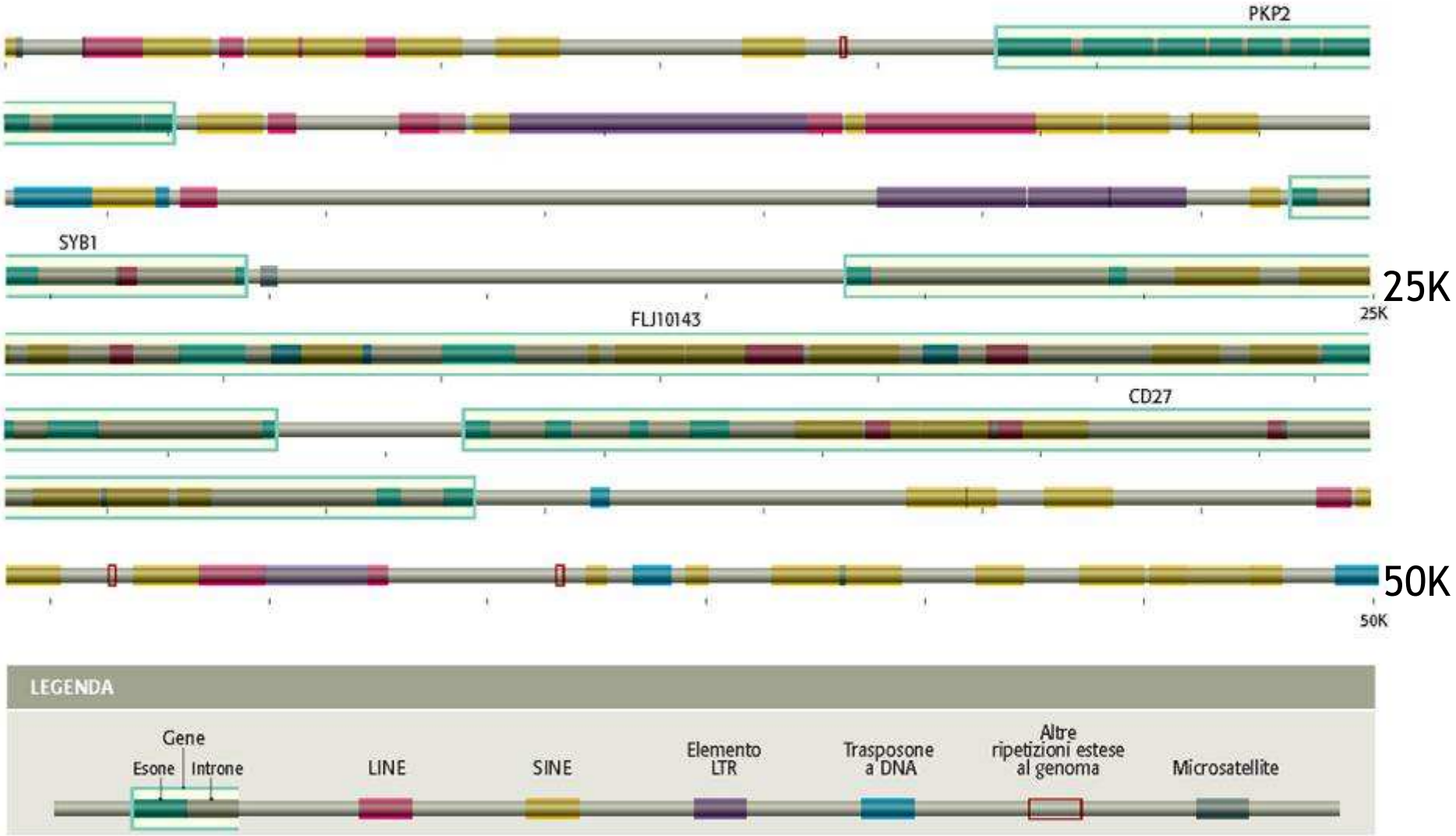


FIGURA 5.29 I cromosomi umani Schema dei cromosomi metafasici umani con il bandeggio ottenuto mediante colorazione citogenetica.

Figura 7.12 Un tratto del genoma umano.
Questa mappa mostra la posizione dei geni, dei segmenti genici, delle ripetizioni estese all'intero genoma e dei microsatelliti in un segmento da 50 kb del cromosoma 12 umano.



Classificazione del DNA eucariotico

Geni codificanti proteine
- geni solitari (in copia unica)
- geni duplicati e diversificatisi (famiglie di geni funzionali e *pseudogeni* non funzionali)

Geni ripetuti (che codificano rRNA, tRNA, rRNA 5S e istoni)

DNA ripetitivo
- DNA a sequenza semplice (satellite)
- DNA a ripetitività intermedia (elementi genetici mobili)
- Trasposoni
- Retrotrasposoni virali
- Elementi dispersi lunghi (LINES; retrotrasposoni non virali)
- Elementi dispersi corti (SINES; retrotrasposoni non virali)

DNA spaziatore non classificato

H. sapiens

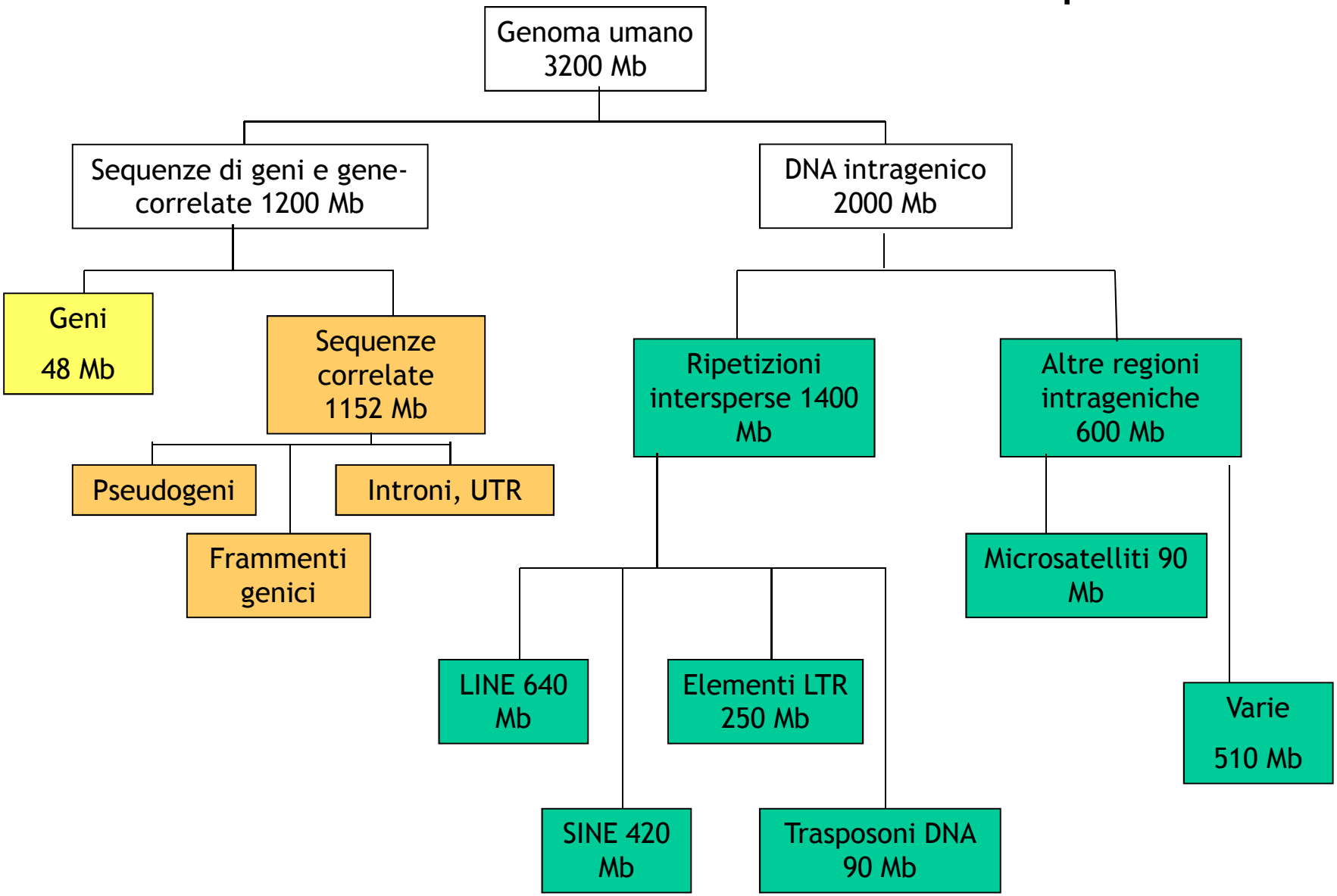

Genoma umano 3200 Mb
Sequenze di geni e gene-correlate 1200 Mb
DNA intragenico 2000 Mb
Geni 48 Mb
Sequenze correlate 1152 Mb
Ripetizioni intersperse 1400 Mb
Altre regioni intrageniche 600 Mb
Pseudogeni
Introni, UTR
Frammenti genici
Microsatelliti 90 Mb
LINE 640 Mb
Elementi LTR 250 Mb
Varie 510 Mb
SINE 420 Mb
Trasposoni DNA 90 Mb

TABELLA 5.1 Caratteristiche generali di un gene umano

		Range
Numero di esoni	9	1-cent.
Numro di introni	8	0-cent.
Sequenze esoniche:		Range
regione non tradotta al 5′ (5′UTR)	300 paia di basi	30- some Kb
sequenza codificante	1400 paia di basi	300- some Mb
regione non tradotta al 3′ (3′UTR)	800 paia di basi	100- some Kb
TOTALE	2500 paia di basi	-
Sequenze intron iche:	27,000 paia di basi	2Kb- 100Mb

I valori mostrati in tabella sono i valori **medi**

Genes

Protein coding (mRNA)

noncoding: ncRNA

- ribosomal RNAs & tRNA
- snRNA (small nuclear)
- snoRNA (small nucleolar)
- miRNA (micro RNA)
- antisense
- enhancer RNAs
- retroelement RNAs

Genes present in multiple copies
•rRNA, tRNA encoding genes
•histone encoding genes

Table 19 Number of tRNA genes in various organisms

Organism	Number of canonical tRNAs	SeCys tRNA
Human	497	1
Worm	584	1
Fly	284	1
Yeast	273	0
Methanococcus jannaschii	36	1
Escherichia coli	86	1

Number of tRNA genes in each of six genome sequences, according to analysis by the computer program tRNAscan-SE . Canonical tRNAs read one of the standard 61 sense codons; this category excludes pseudogenes, undetermined anticodons, putative supressors and selenocysteine tRNAs. Most organisms have a selenocysteine (SeCys) tRNA species, but some unicellular eukaryotes do not (such as the yeast *S. cerevislae*).

The human genome contains:

2,000 genes
encoding **5S rRNA** in one cluster on chromosome 1

280 copies
of the **transcription unit** encoding **28S, 5,8S and 18S rRNA**,

(organized in 5 clusters of 50-70 units on chromosomes 13, 14, 15, 21, 22)

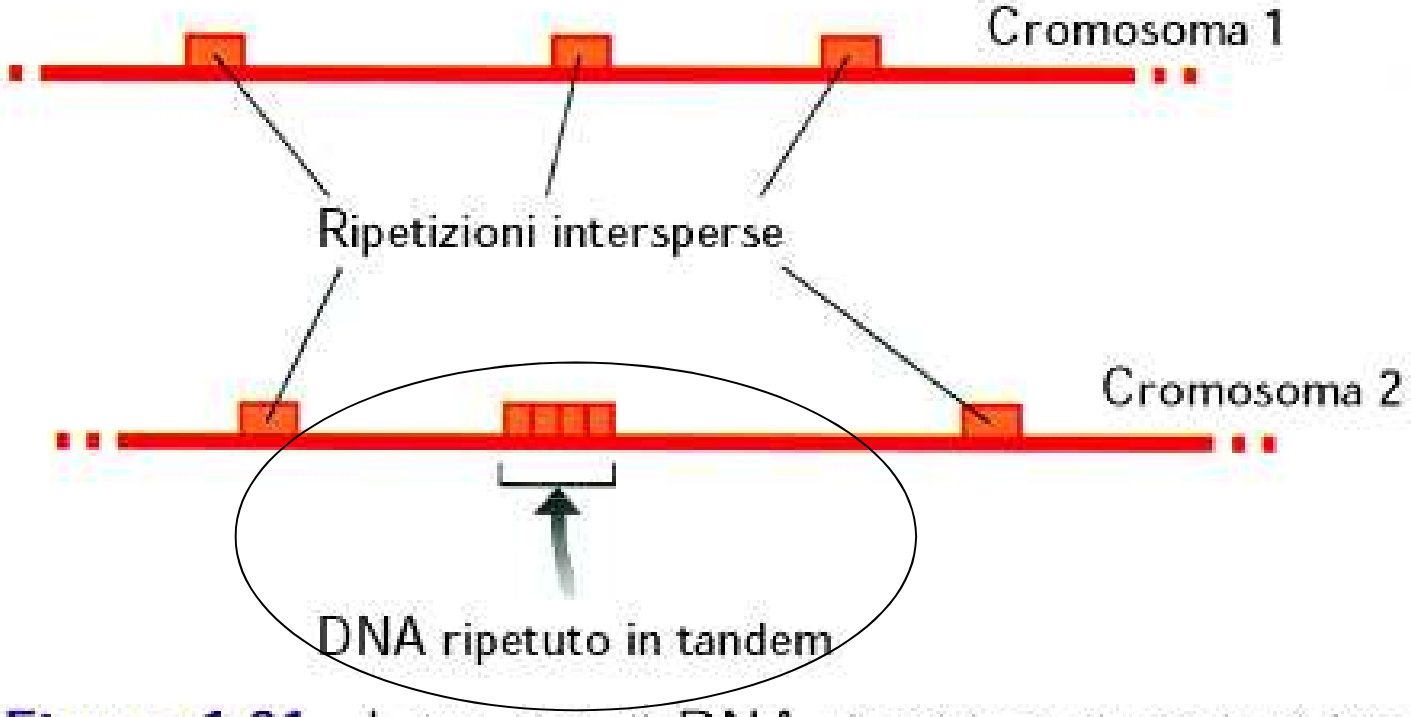
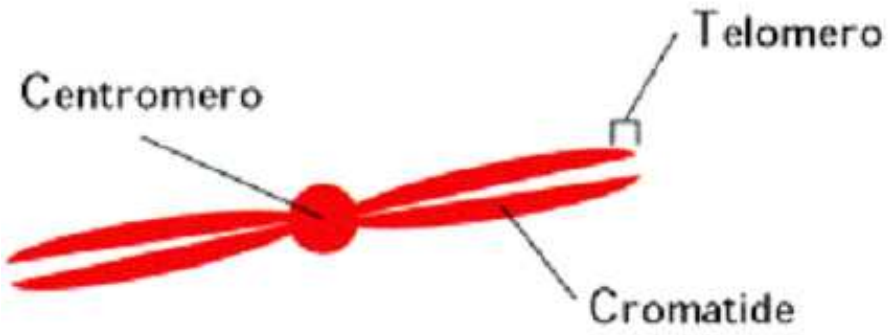
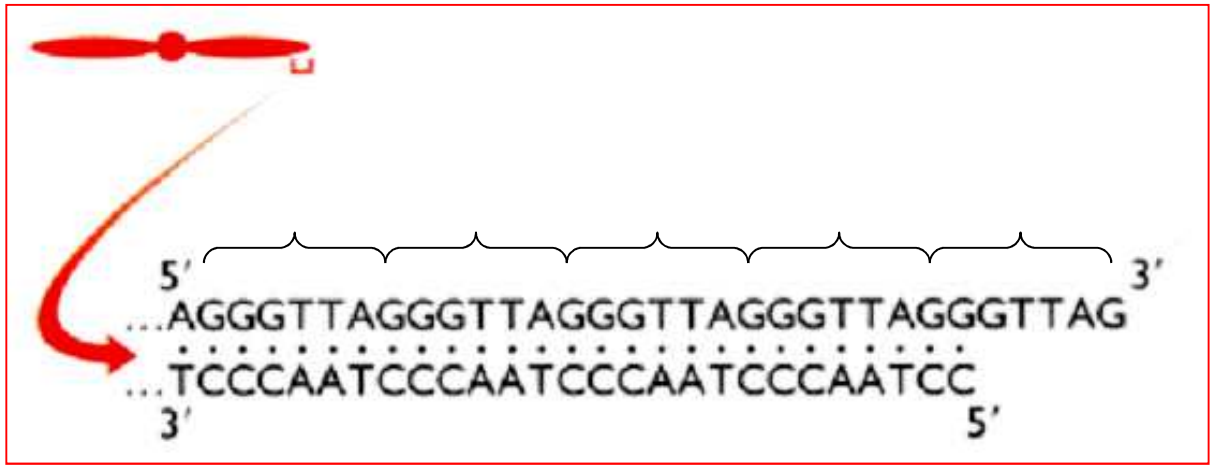


Figura 1.21 I due tipi di DNA ripetitivo: ripetizioni intersperse e DNA ripetuto in tandem.

T.A. Brown
Genomi
EdiSES





Un esempio di ripetizioni microsatellitari: i telomeri dei cromosomi

Telomers and telomerase

TABELLA 5.5 DNA telomerici

Organismo	Sequenza telomerica ripetuta
Lieviti	
Saccharomyces cerevisiae	$G_{1-3}T$
Schizosaccharomyces pombe	$G_{2-5}TTAC$
Protozoi	
Tetrahymena	GGGGTT
Dictyostelium	$G_{1-8}A$
Piante	
Arabidopsis	AGGGTTT
Mammiferi	
Uomo	AGGGTT

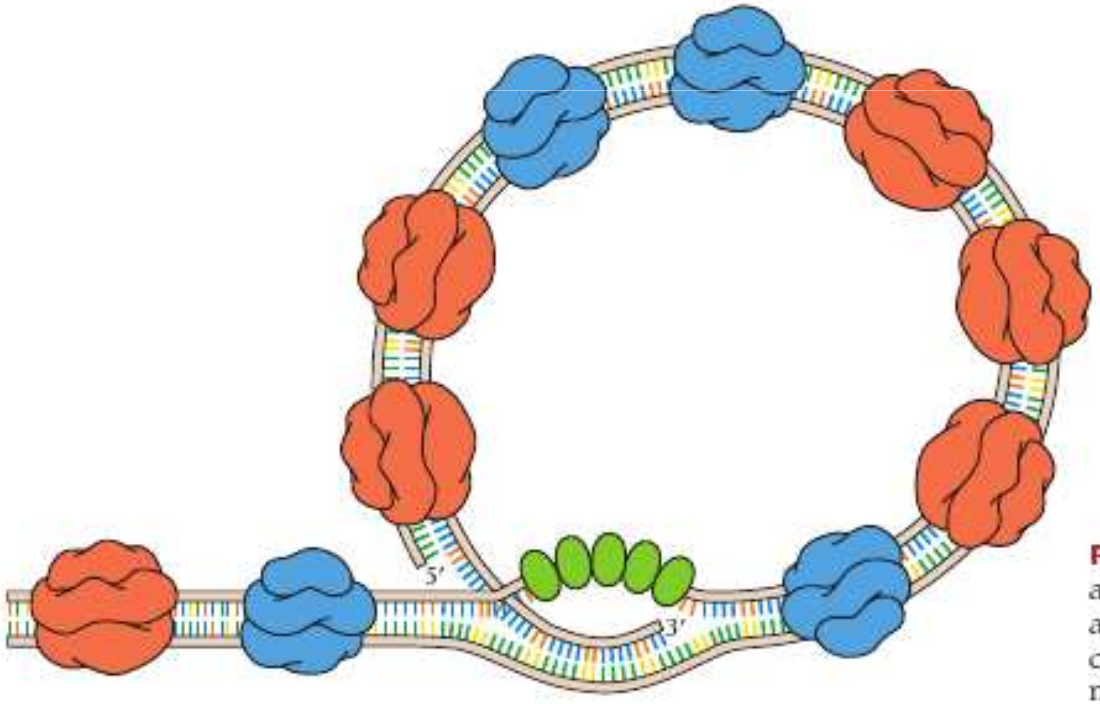


FIGURA 5.22 Struttura di un telomero Le anse di DNA telomerico si ripiegano su se stesse a formare una struttura circolare e si associano con diverse proteine che proteggono le estremità dei cromosomi.

Tabella 1.3 Microsatelliti nel genoma umano

Lunghezza della unità ripetitiva	Numero approssimativo delle copie nel genoma umano
1	120.000
2	140.000
3	37.500
4	105.000
5	56.000
6	49.000
7	27.000
8	35.500
9	27.500
10	27.500
11	28.000

Da IHGSC (2001).

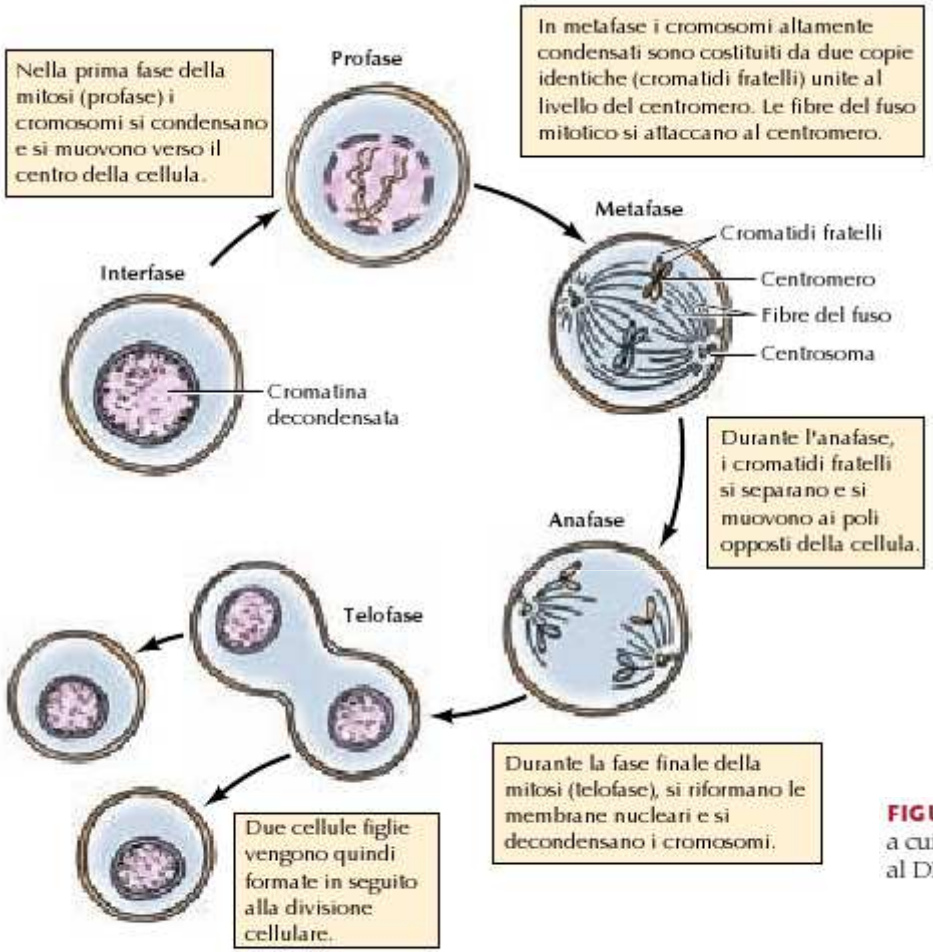


FIGURA 5.18 I cromosomi durante la mitosi Dato che il DNA viene replicato durante l'interfase, la cellula contiene due copie duplicate identiche di ciascun cromosoma prima di entrare in mitosi.

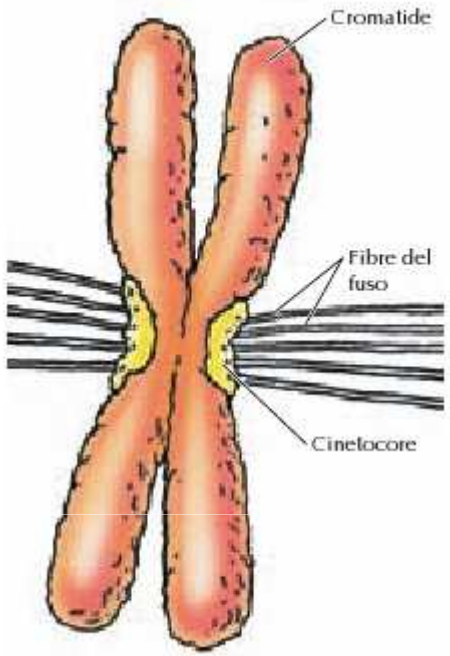
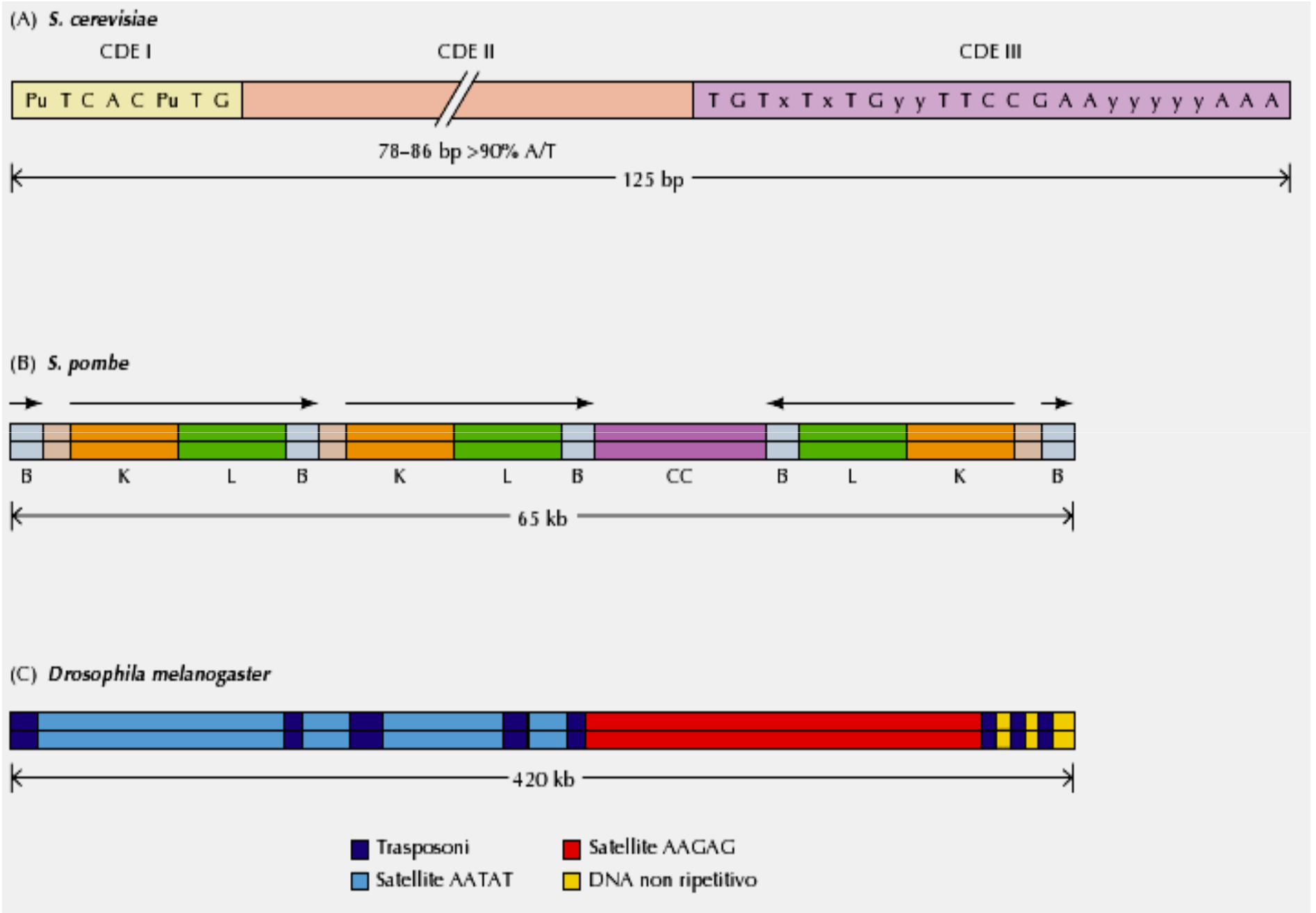


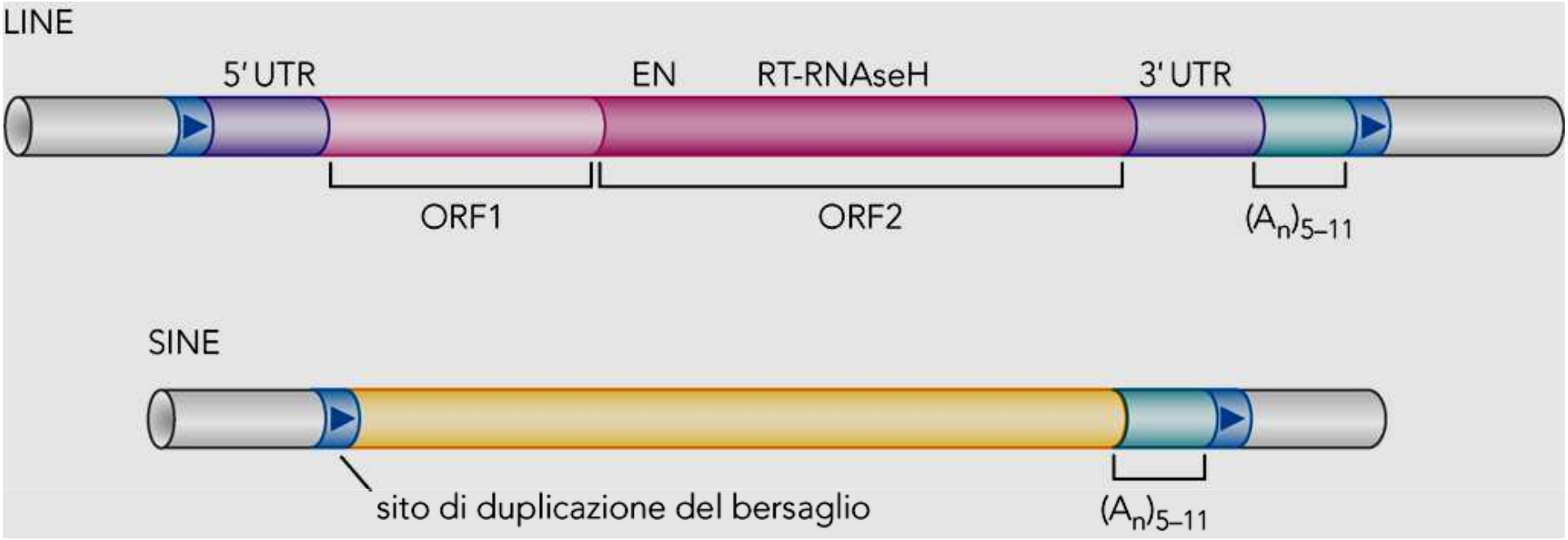
FIGURA 5.19 Il centromero di un cromosoma metafasico Il centromero è la regione a cui si legano i due cromatidi fratelli durante la metafase. Proteine specifiche sono legate al DNA centromerico, formando il cinetocore, che è il sito di attacco delle fibre del fuso.


(A) S. cerevisiae
CDE I
CDE II
CDE III
Pu T C A C Pu T G
T G T x T x T G y y T T C C G A A y y y y y y A A A
78–86 bp >90% A/T
125 bp
(B) S. pombe
B K L B K L B CC B L K B
65 kb
(C) Drosophila melanogaster
420 kb
Trasposoni
Satellite AATAT
Satellite AAGAG
DNA non ripetitivo

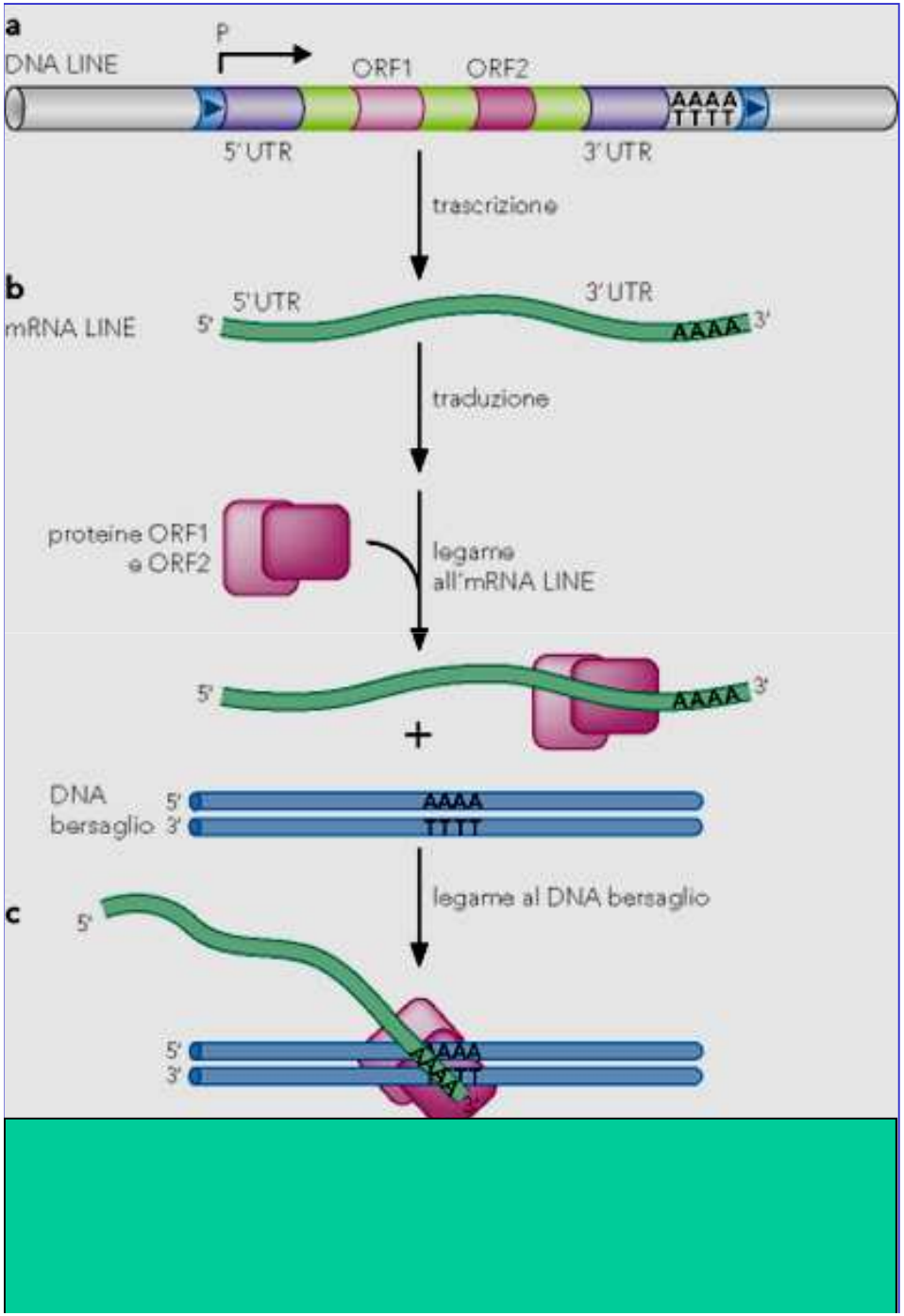
Interspersed repetitive elements - Mobile genetic elements

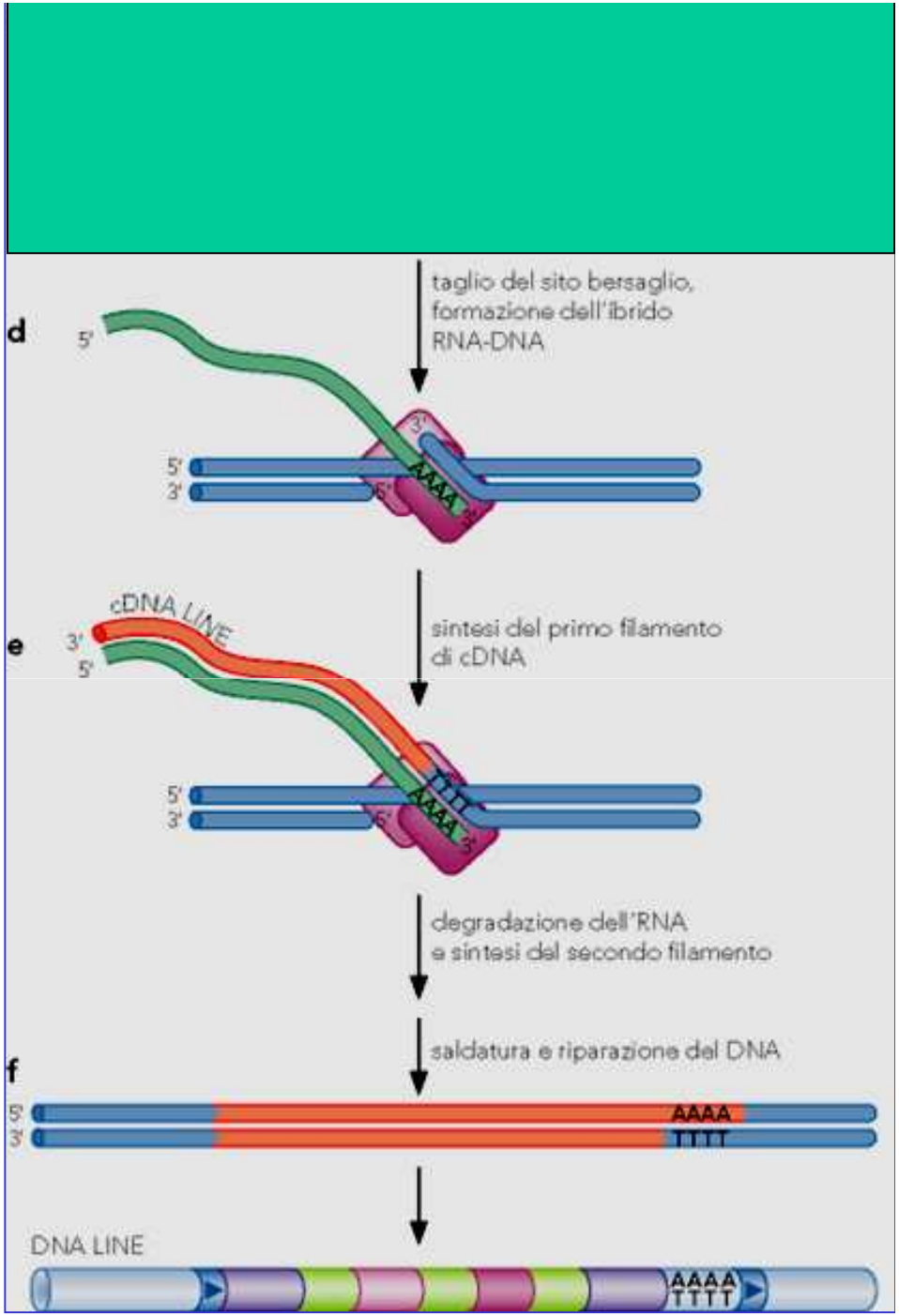
Tabella 1.2 Tipi di ripetizioni estese a tutto il genoma nell'uomo

Tipo di ripetizione	Sottotipo	Numero approssimativo delle copie nel genoma umano
SINE		1.558.000
	Alu	1.090.000
	MIR	393.000
	MIR3	75.000
LINE		868.000
	LINE-1	516.000
	LINE-2	315.000
	LINE+3	37.000
Elementi LTR		443.000
	Classe I ERV	112.000
	Classe II ERV(K)	8.000
	Classe III ERV(L)	83.000
	MaLR	240.000
Trasposoni DNA		294.000
	hAT	195.000
	Tc-1	75.000
	PiggyBac	2.000
	Non classificato	22.000

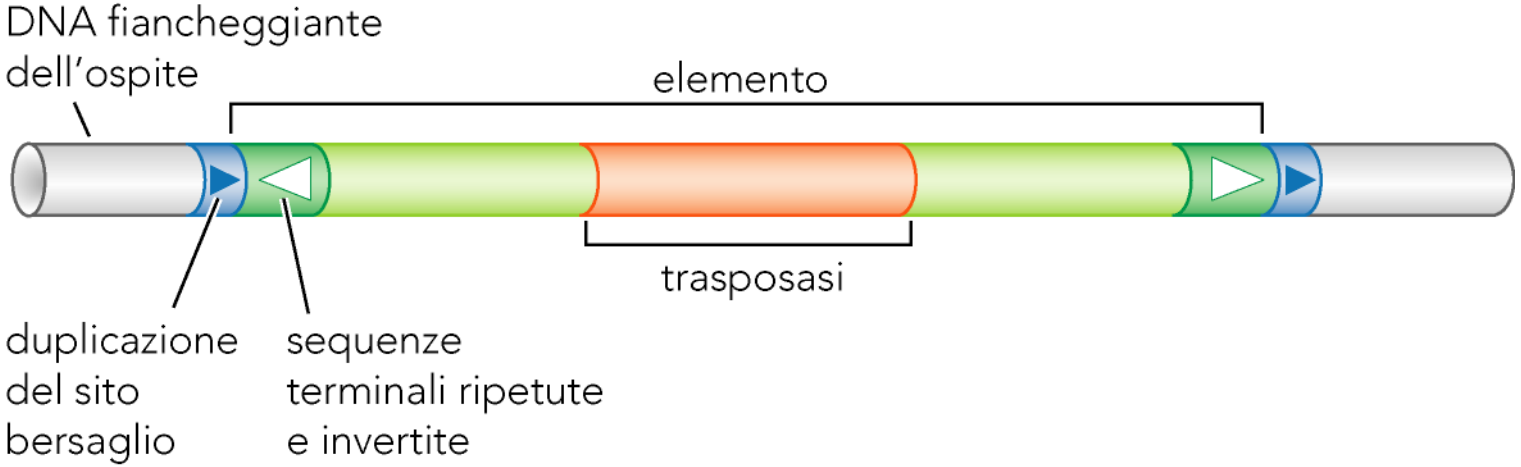


Trasposizione e retrotrasposizione

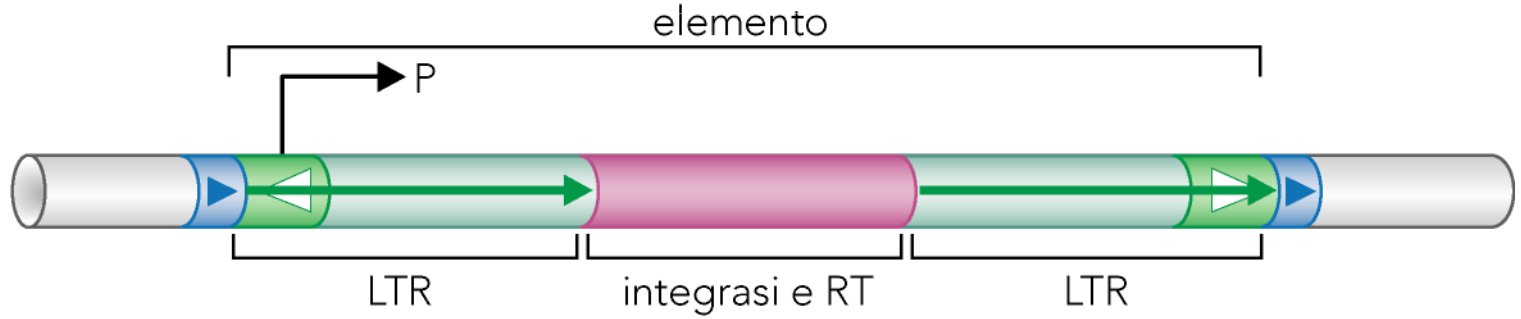

a
DNA LINE
P
ORF1
ORF2
AAAA
TTTT
5' UTR
3' UTR
trascrizione
b
mRNA LINE
5' UTR
3' UTR
5'
AAAA
3'
traduzione
proteine ORF1
e ORF2
legame
all'mRNA LINE
5'
AAAA
3'
+
DNA
bersaglio
5'
3'
AAAA
TTTT
legame al DNA bersaglio
c
5'
5'
3'
AAAA
TTTT


taglio del sito bersaglio,
formazione dell'ibrido
RNA-DNA
d
5'
3'
5'
3'
AAAA
cDNA LINE
sintesi del primo filamento
di cDNA
e
3'
5'
5'
3'
TTTT
AAAA
degradazione dell'RNA
e sintesi del secondo filamento
saldatura e riparazione del DNA
f
5'
3'
AAAA
TTTT
DNA LINE
AAAA
TTTT

a trasposoni a DNA

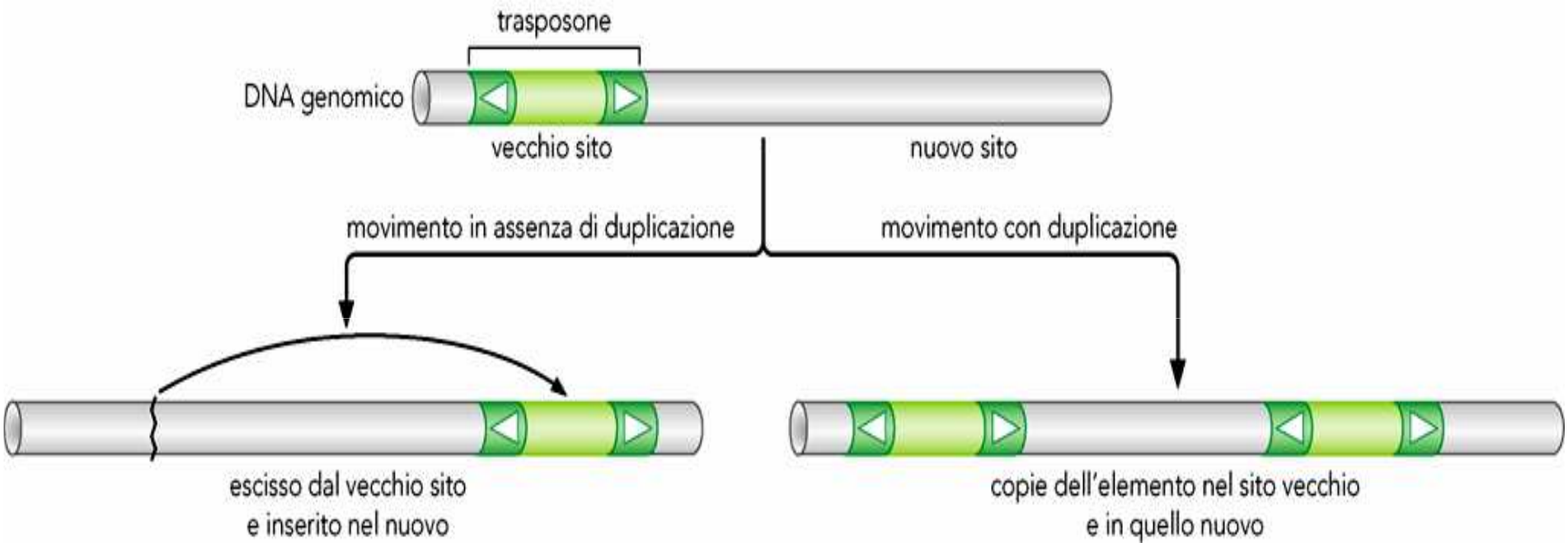


b retrotrasposoni tipo virus/retrovirus



c retrotrasposoni poli-A

Trasposoni



abortivo

replicativo

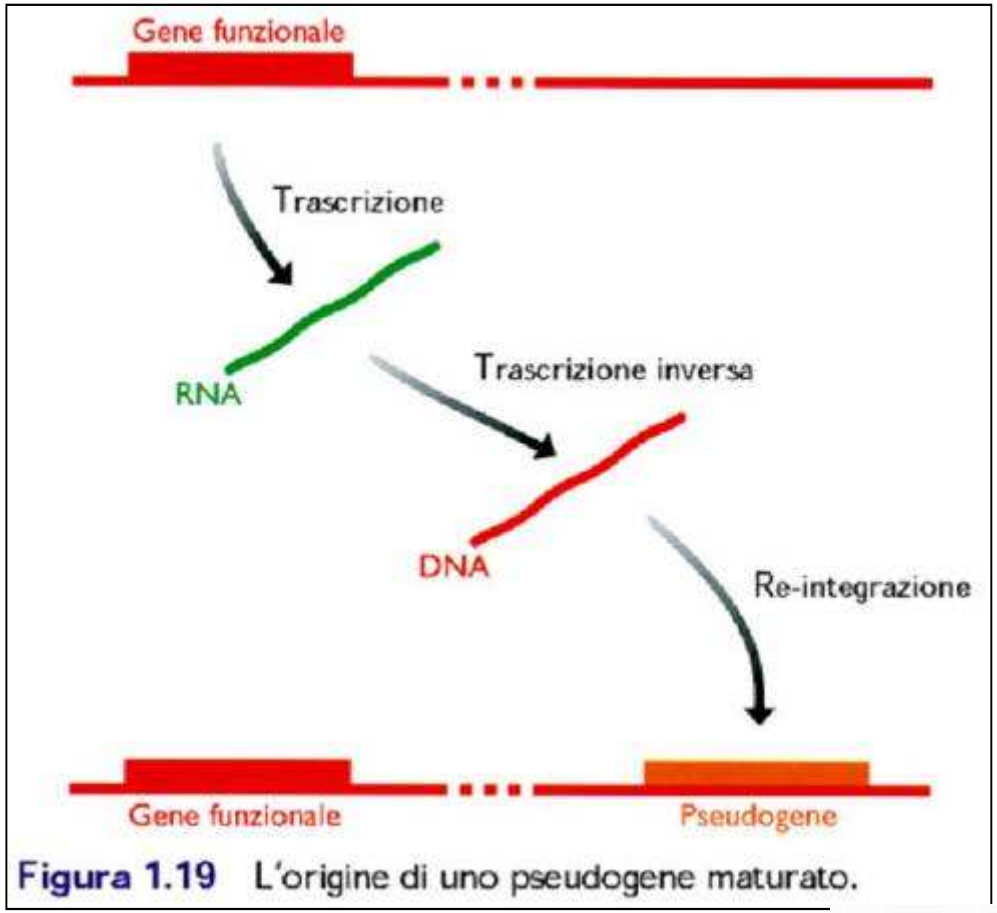
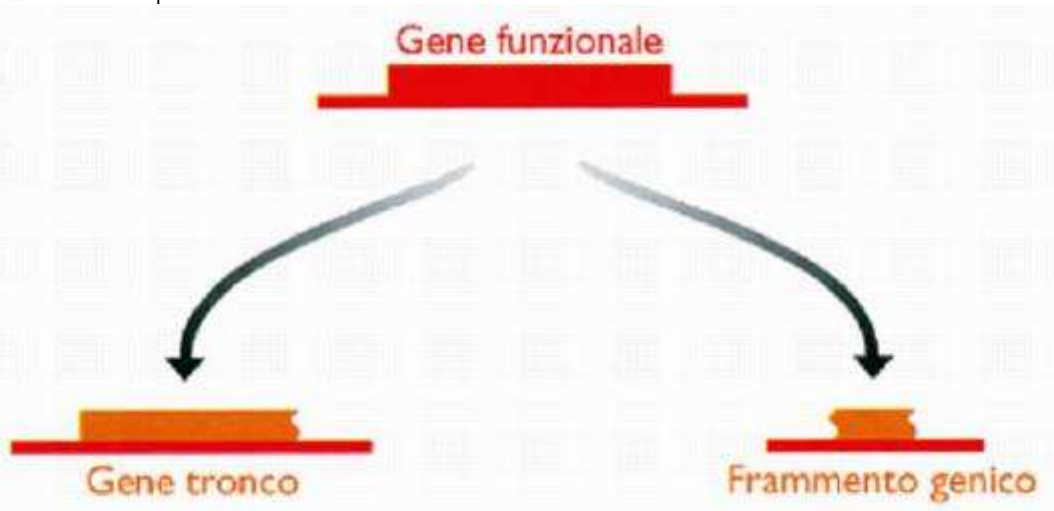


Figura 1.19 L'origine di uno pseudogene maturato.

Second class of pseudogenes are gene copies inactivated by multiple mutations, or:



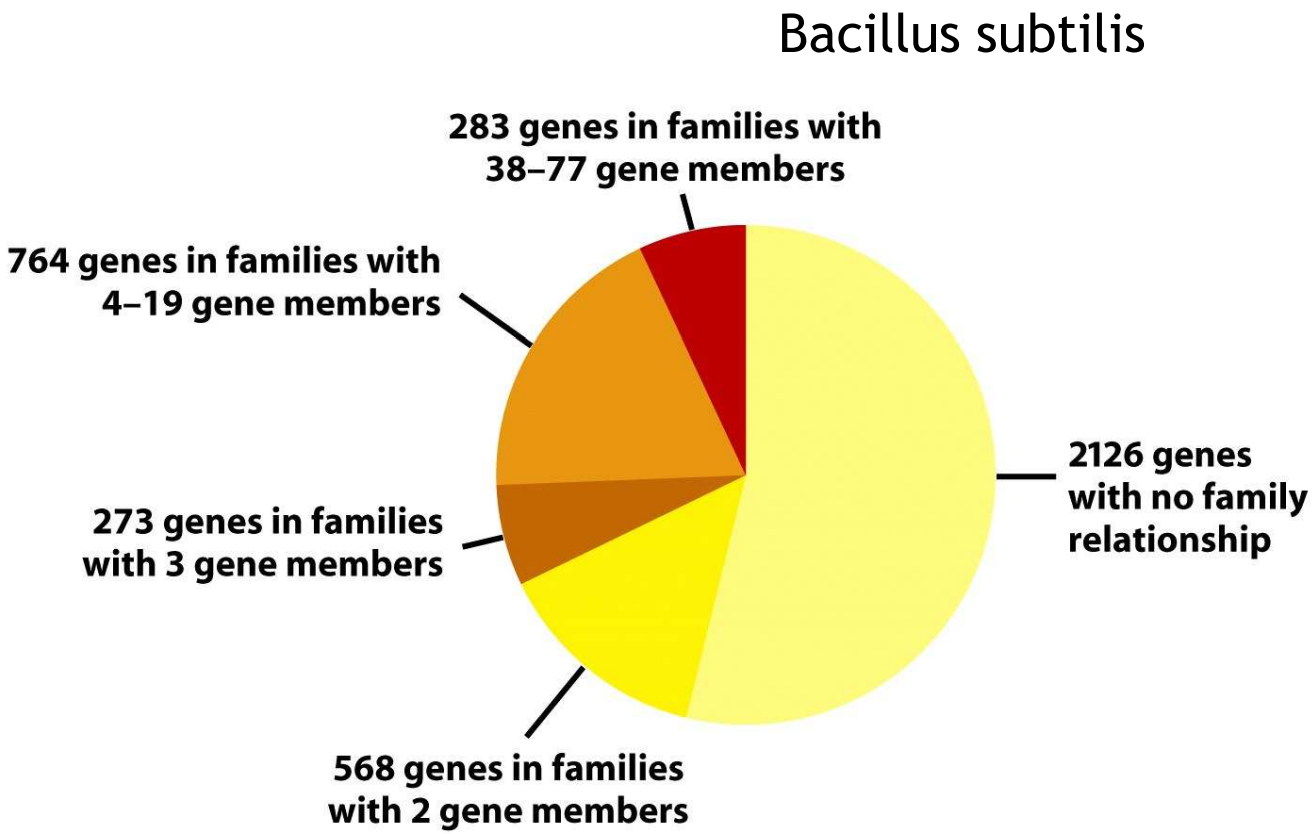


Figure 1-24 *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

catalogo

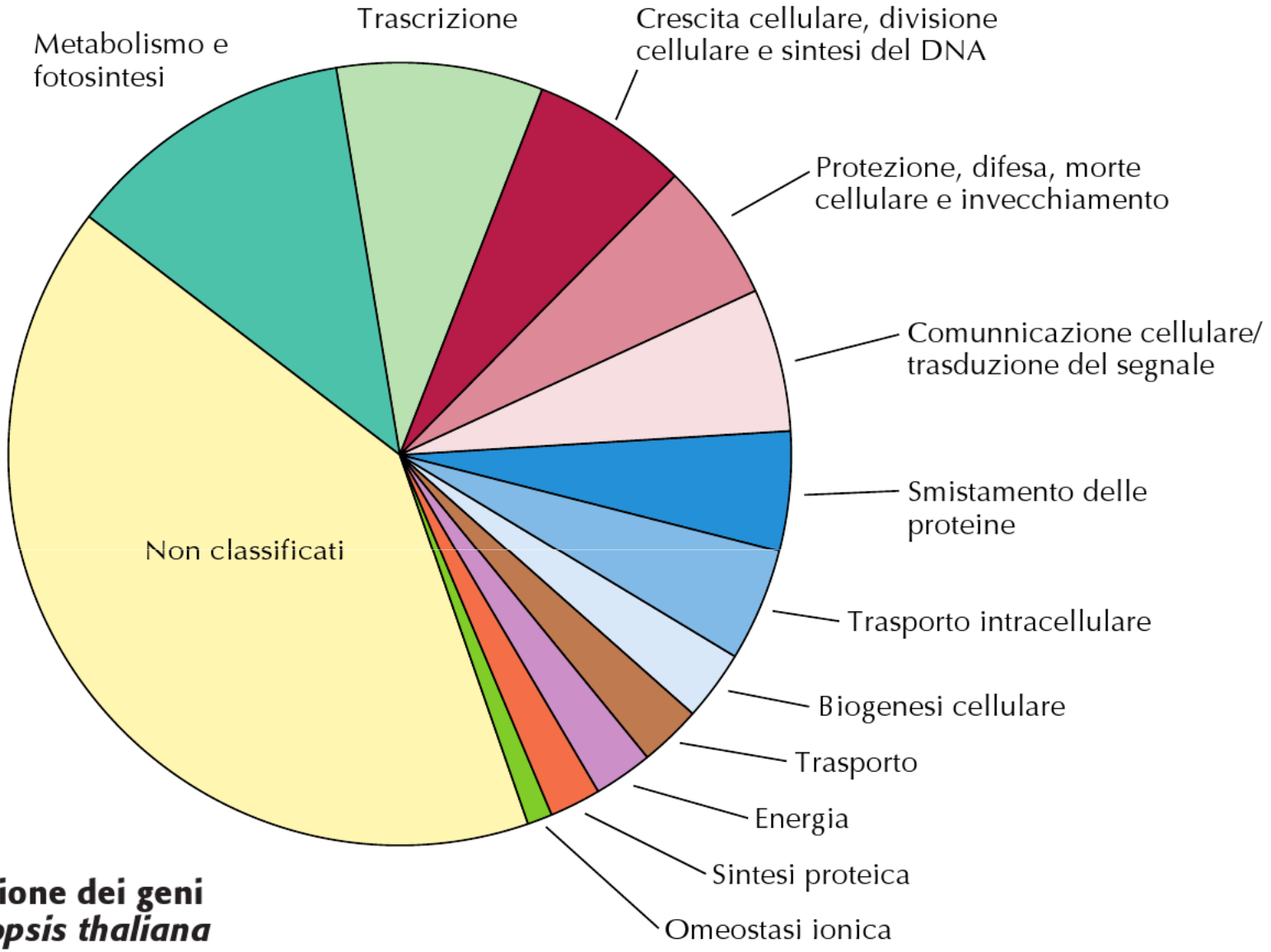


FIGURA 5.28 Funzione dei geni predetti in *Arabidopsis thaliana*
Il grafico illustra la proporzione dei geni di *Arabidopsis* in differenti categorie funzionali. (Da: The *Arabidopsis* Genome Initiative, 2000. *Nature* 408: 796).

Gene functions

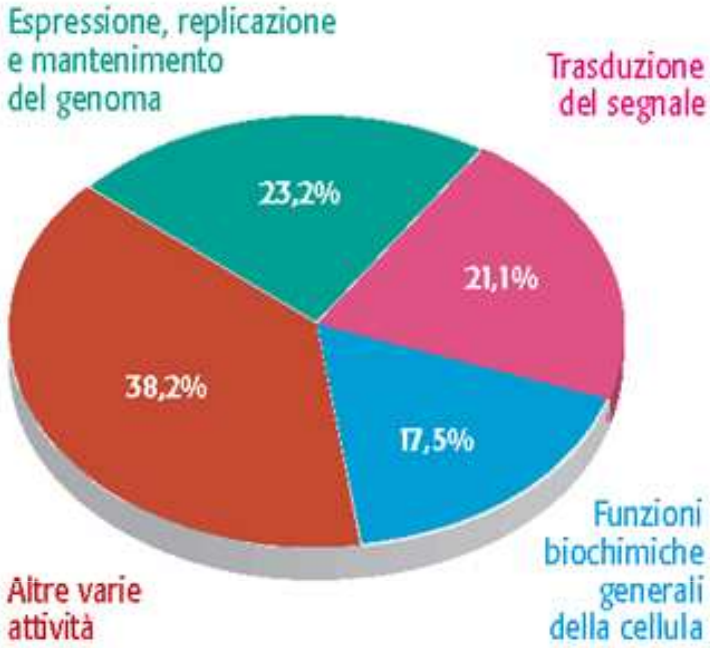
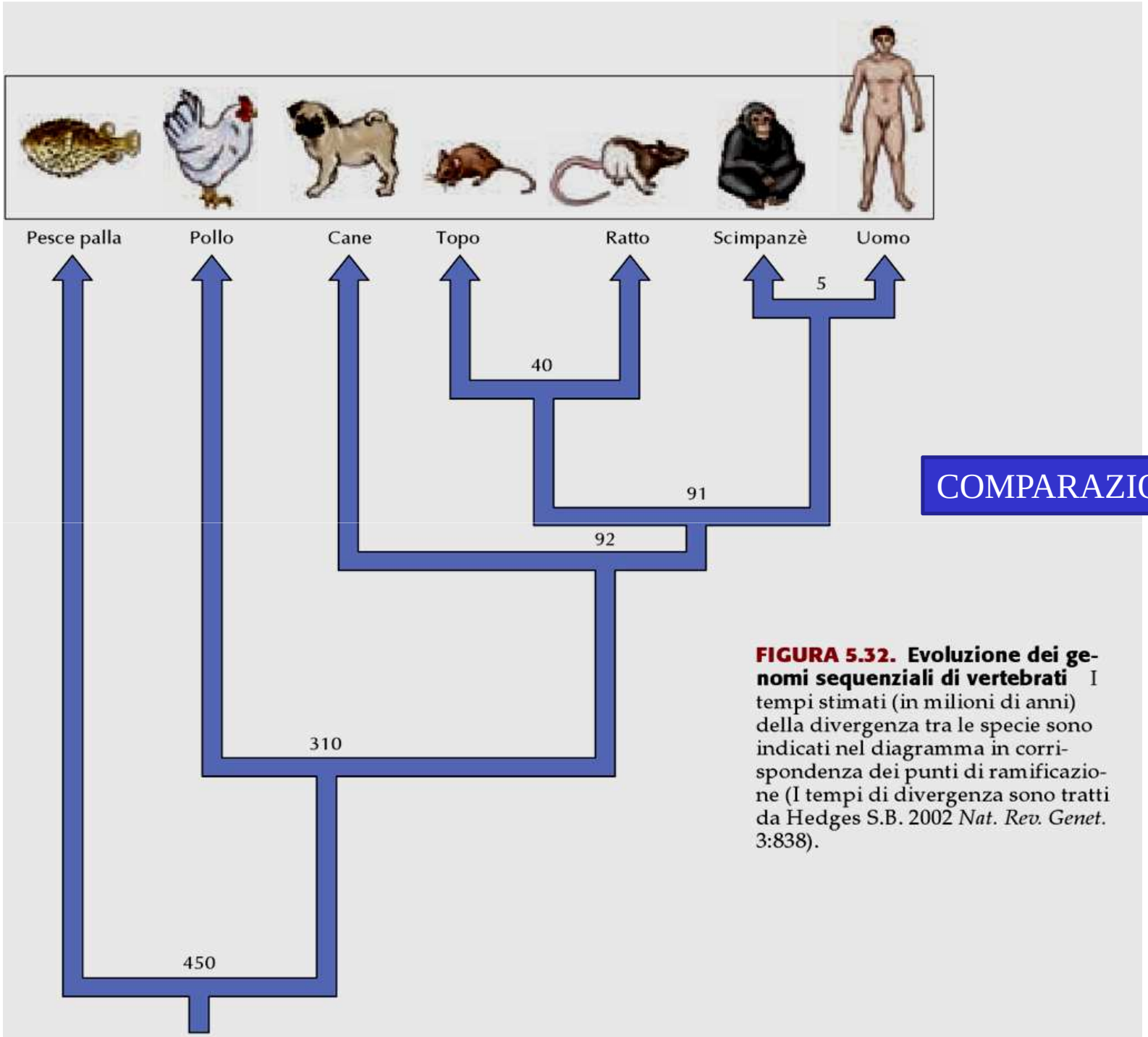


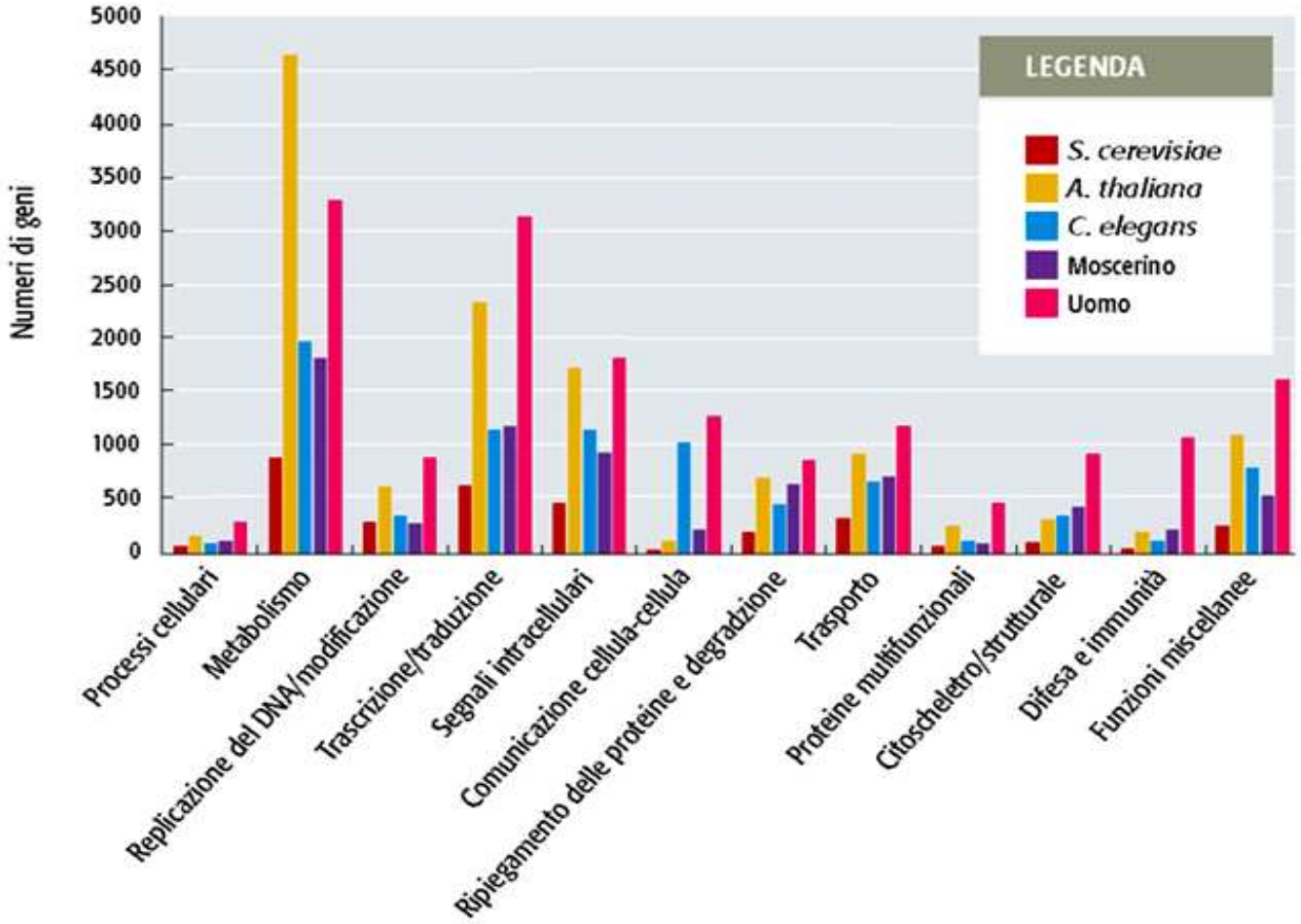
Figura 7.16 Classificazione del catalogo dei geni umani. Il diagramma a torta mostra la suddivisione dei geni umani codificanti proteine finora identificati. Mancano approssimativamente 13.000 geni, le cui funzioni sono sconosciute. La porzione corrispondente a varie altre attività include, tra gli altri, le proteine coinvolte nei processi di trasporto biochimico e nel ripiegamento, le proteine immunologiche e le proteine strutturali.



COMPARAZIONE

FIGURA 5.32. Evoluzione dei genomi sequenziali di vertebrati I tempi stimati (in milioni di anni) della divergenza tra le specie sono indicati nel diagramma in corrispondenza dei punti di ramificazione (I tempi di divergenza sono tratti da Hedges S.B. 2002 *Nat. Rev. Genet.* 3:838).

7.17 – Confronto tra i cataloghi genici di S. cerevisiae, A. thaliana, C. elegans, D. melanogaster e H. sapiens. I geni sono raggruppati in base alle loro funzioni, dedotta dai domini proteici che sono specificati da ciascun gene.



T.A. Brown
Genomi, III Ed.
EdiSES

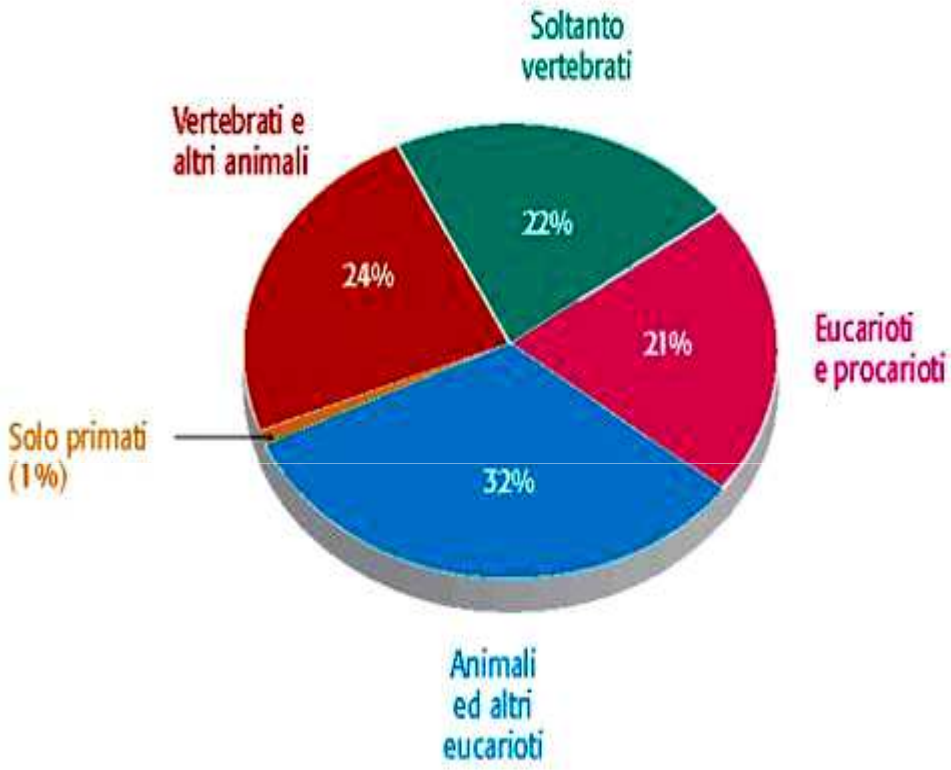


Figura 7.18 Relazione tra il catalogo genico umano e quello di altri gruppi di organismi. Il grafico a torta classifica il catalogo genico umano in base alla distribuzione dei singoli geni negli altri organismi. Il grafico mostra, per esempio, che il 22% del catalogo genico umano è costituito da geni che sono specifici per i vertebrati, mentre un altro 24% comprende geni specifici per i vertebrati e altri animali.

EdiSES

T.A. Brown
Genomi, III Ed.
EdiSES

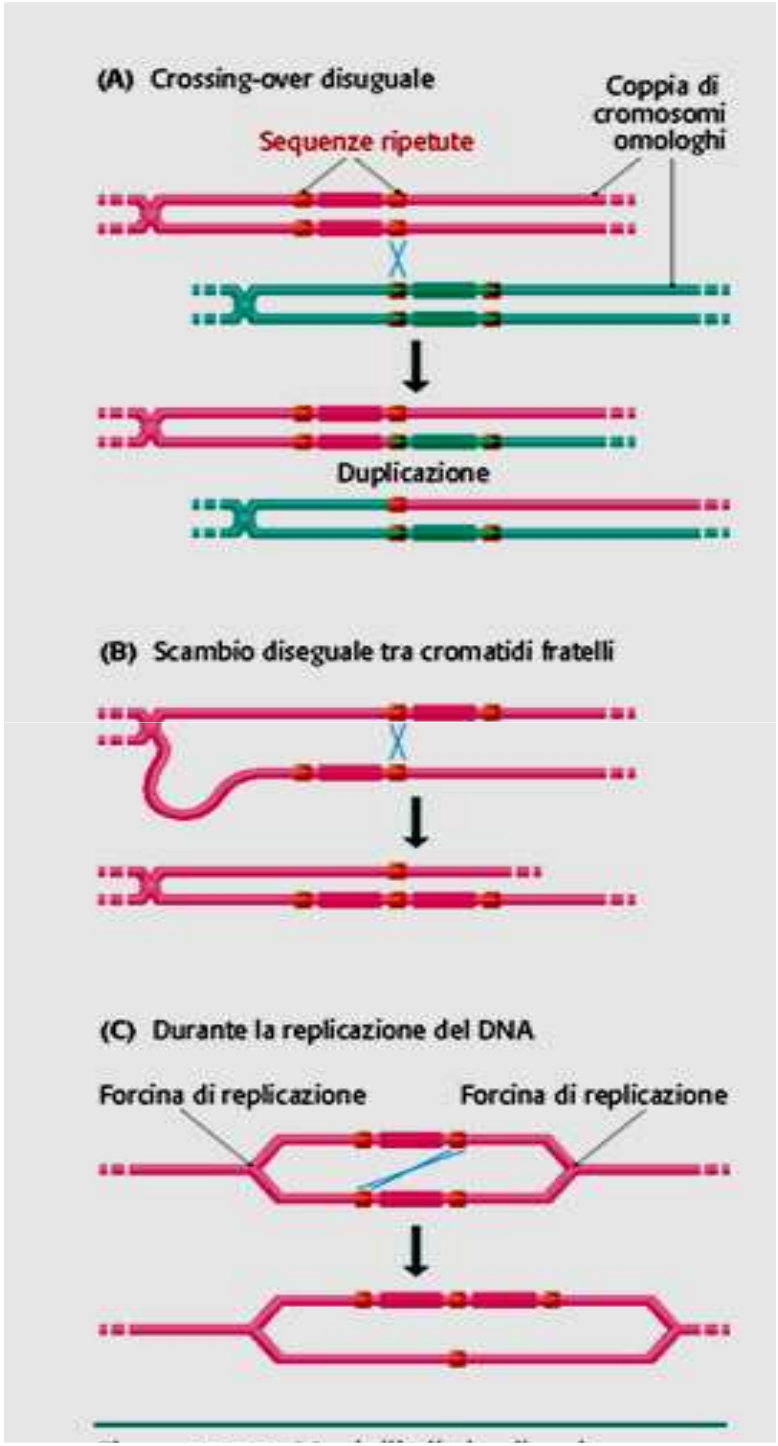
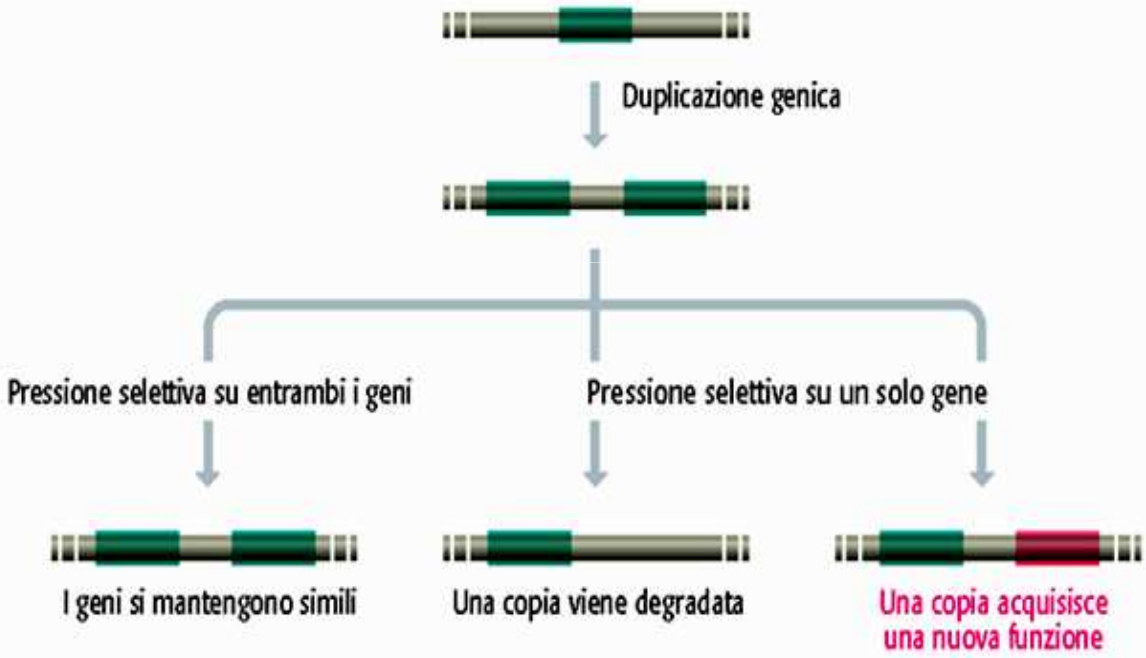


Figura 18.11 Modelli di duplicazione genica tramite (A) crossing-over disuguale tra cromosomi omologhi, (B) scambio disuguale tra cromatidi fratelli e (C) durante la replicazione di un genoma batterico. In tutti i casi la ricombinazione avviene tra due copie di brevi sequenze ripetute e porta alla duplicazione della sequenza tra le ripetizioni. Il crossing-over disuguale e lo scambio disuguale tra cromatidi fratelli sono identici, eccetto che per il fatto che il primo utilizza cromatidi appartenenti a una coppia di cromosomi omologhi e il secondo i cromatidi di un singolo cromosoma. In (C), la ricombinazione avviene tra due doppie eliche figlie che sono state appena sintetizzate tramite replicazione del DNA.

T.A. Brown
Genomi, III Ed.
EdiSES

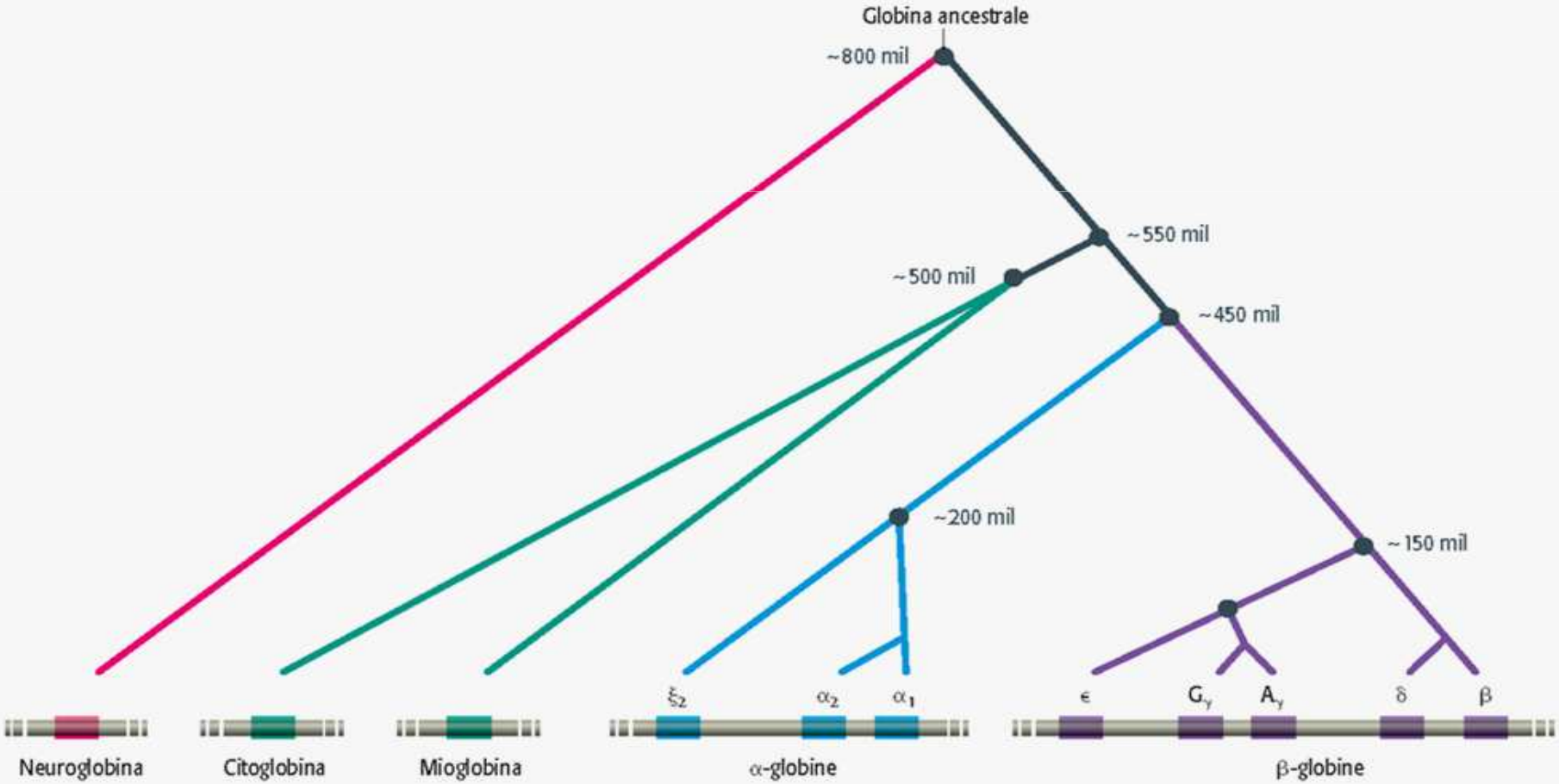
Figura 18.7 Tre possibili conseguenze della duplicazione genica.



T.A. Brown
Genomi, III Ed.
EdiSES

T.A. Brown
Genomi, III Ed.
EdiSES

Figura 18.8 Evoluzione della famiglia genica delle globine umane. I diversi membri della famiglia si trovano ora su cromosomi diversi: il gene della neuroglobina è sul cromosoma 14, il gene della citoglobina è sul cromosoma 17 e il gene della mioglobina è sul cromosoma 22. Il cluster dei geni delle α-globine è sul cromosoma 16 e quello delle β-globine sul cromosoma 11. Abbreviazioni: Mil, milioni di anni fa.



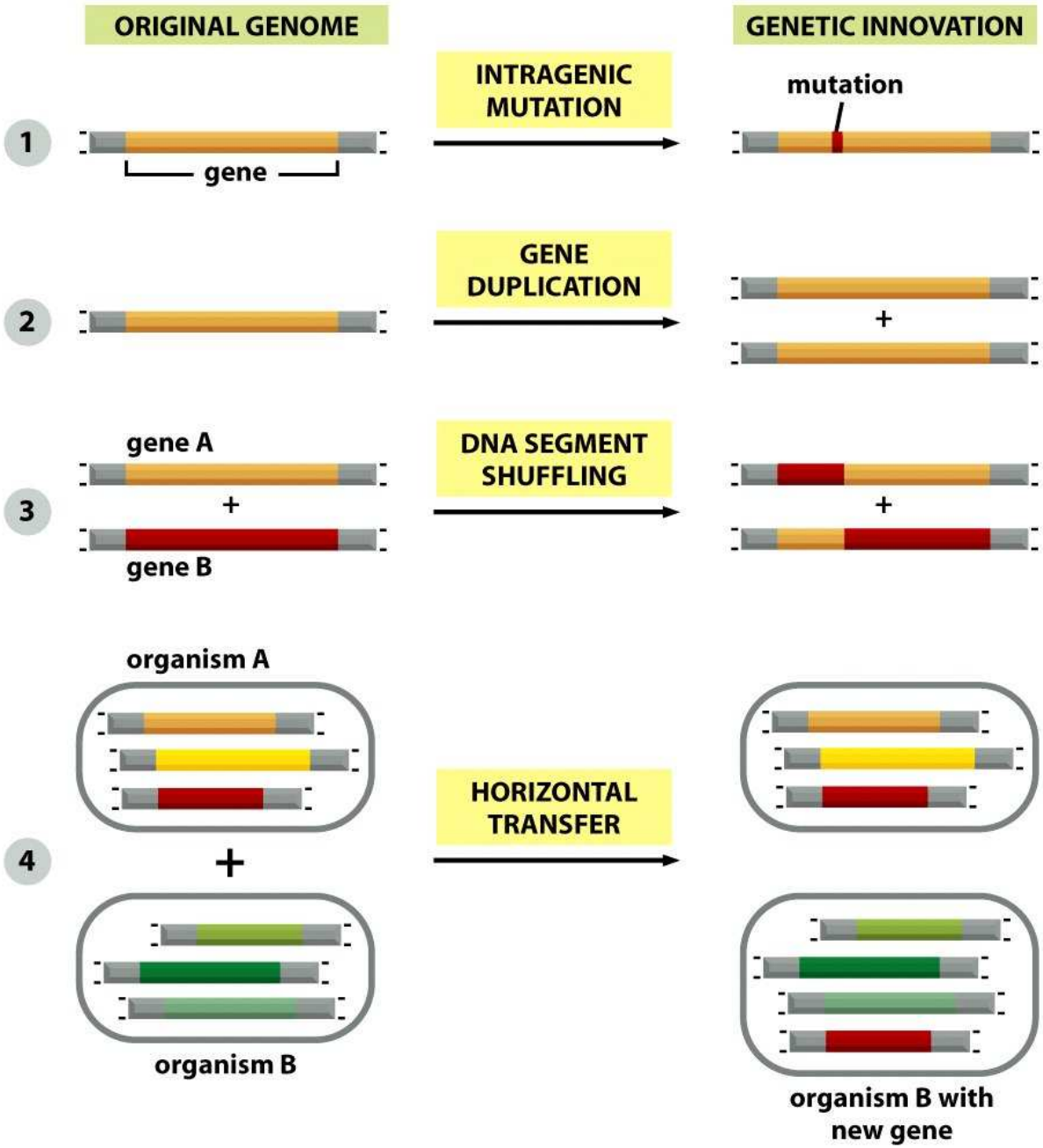


Figure 1-23 *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

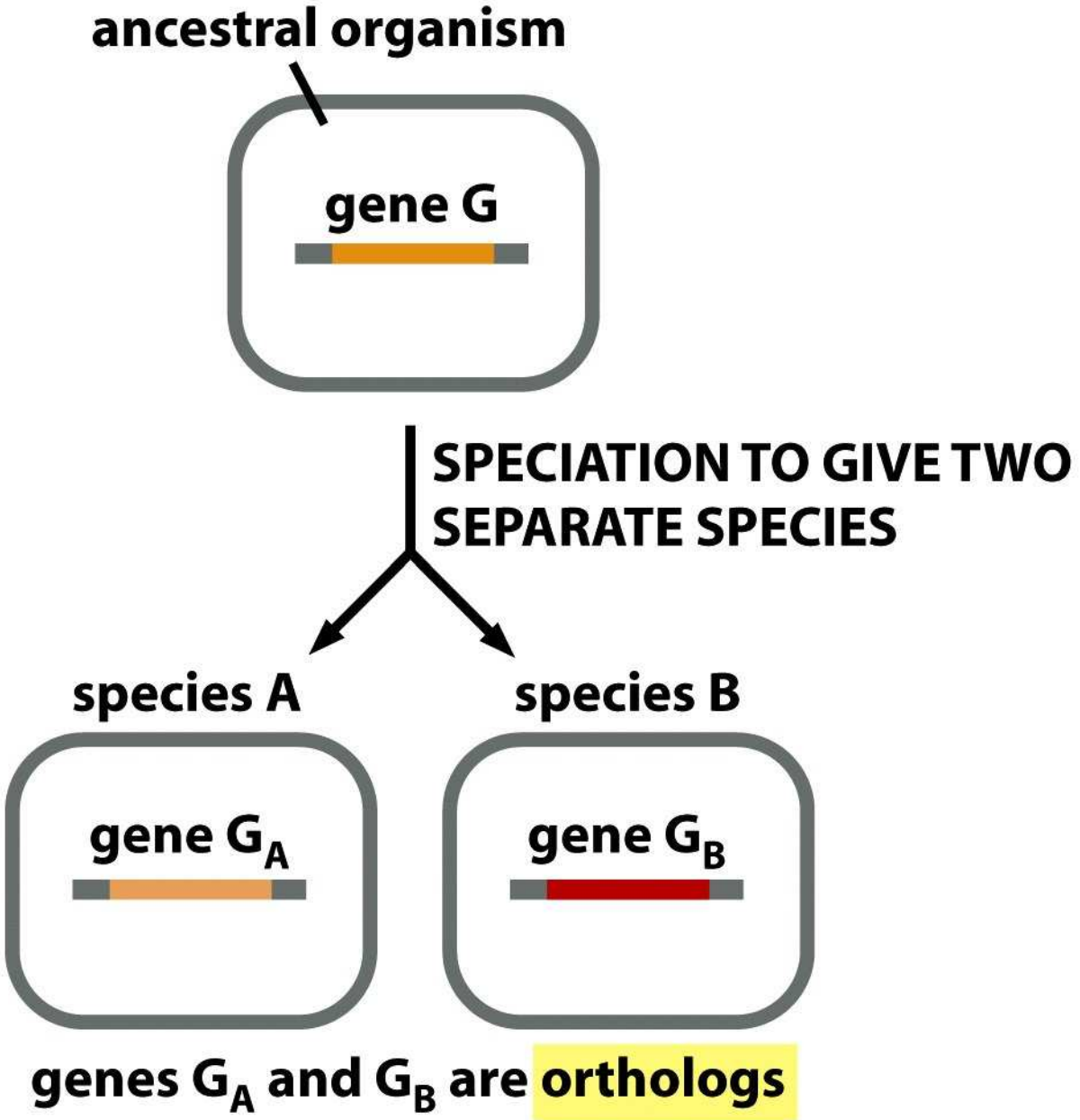


Figure 1-25a *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

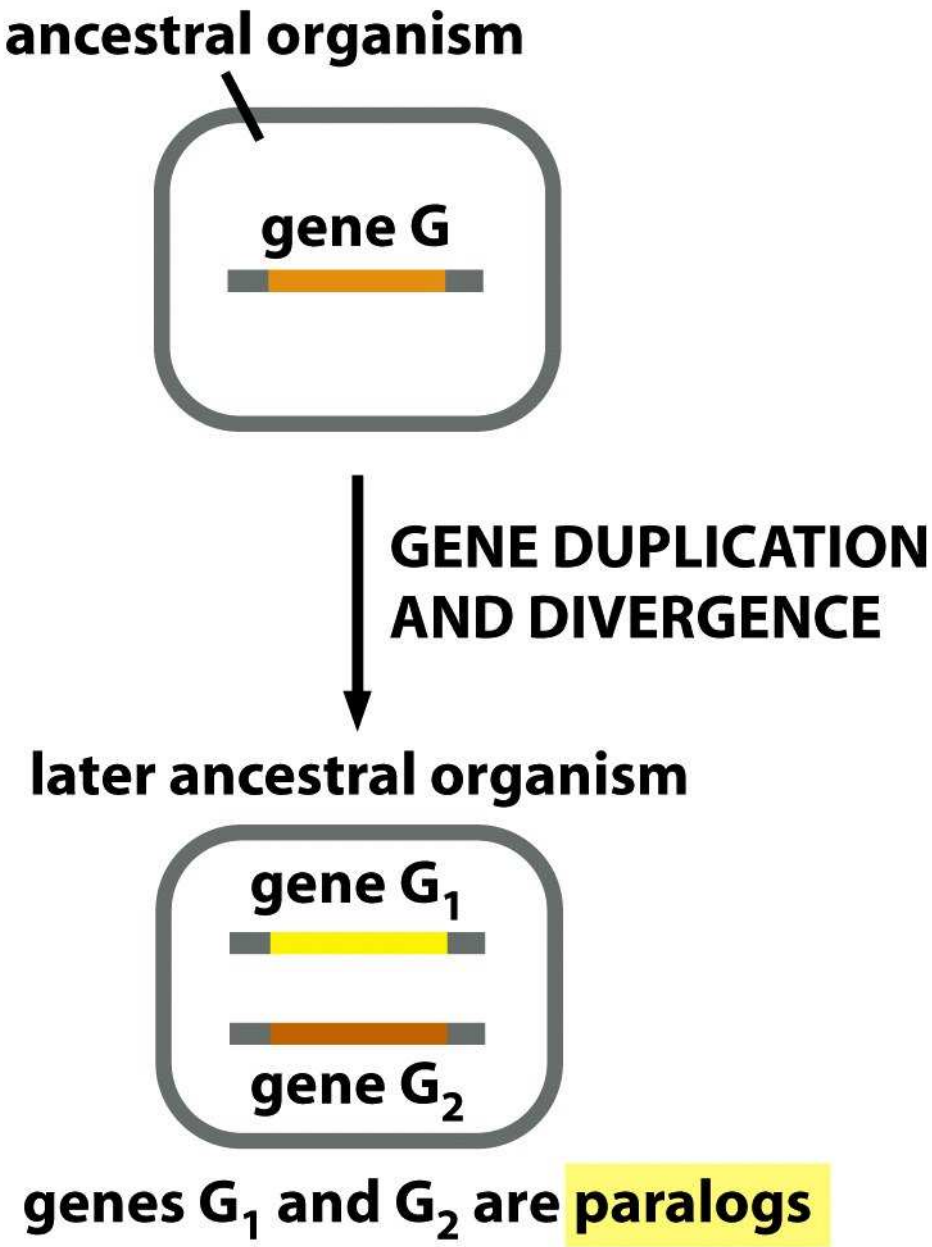


Figure 1-25b *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

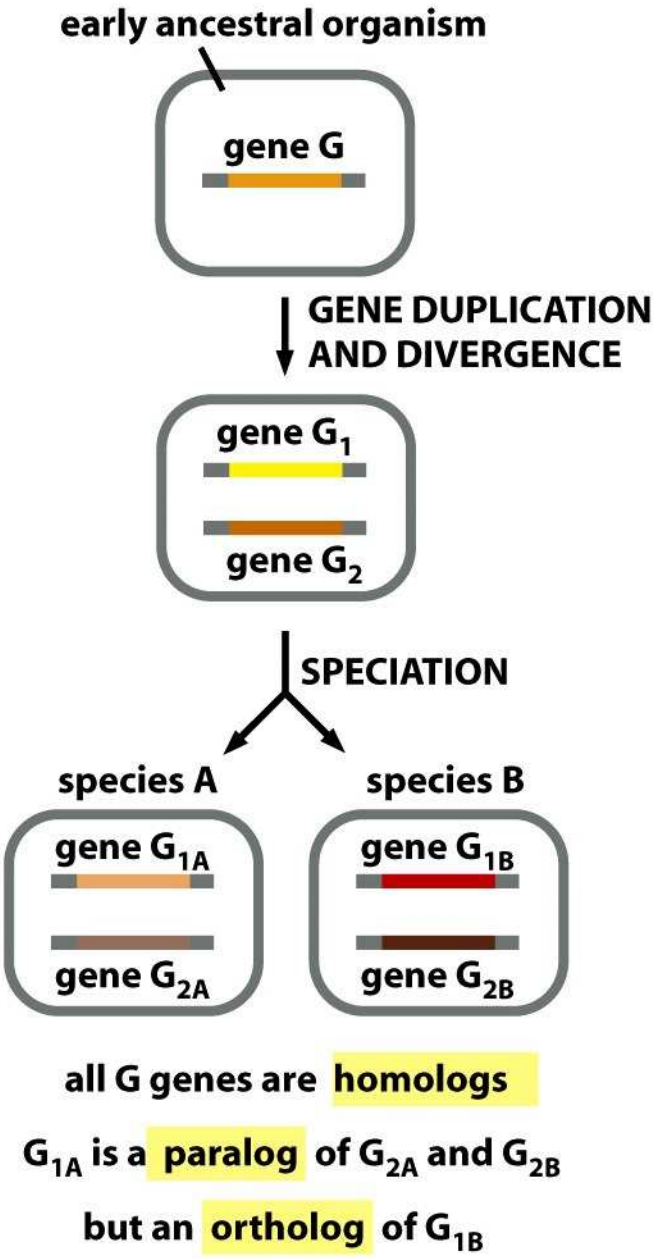


Figure 1-25c *Molecular Biology of the Cell*, Fifth Edition (© Garland Science 2008)

Le similitudini tra proteine hanno rivelato un addizionale livello di organizzazione:

il **dominio**

Il dominio è una sottostruttura prodotta da qualunque parte del polipeptide che si possa ripiegare in una conformazione stabile indipendentemente dal resto della proteina.

Il concetto di dominio è molto importante in genomica, perchè spesso i domini delle proteine sono codificati da singoli esoni, giustificando la teoria dell'"exon shuffling" per l'evoluzione delle proteine.

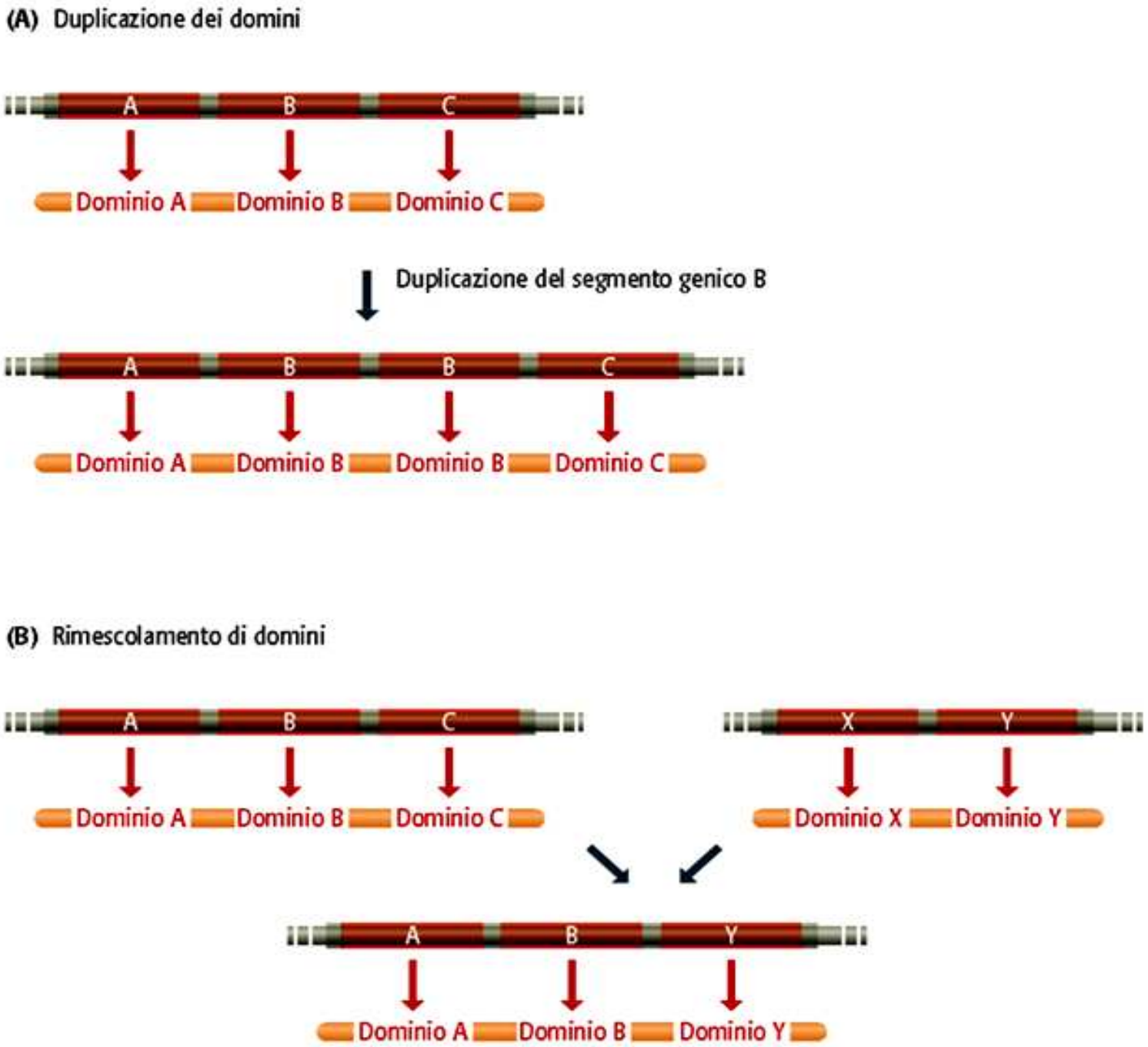
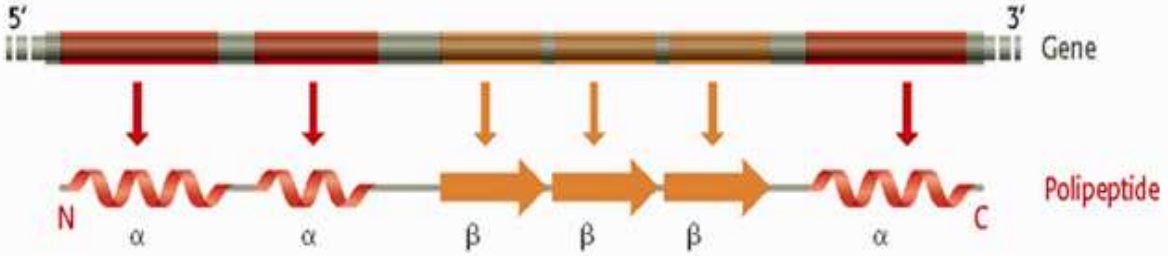


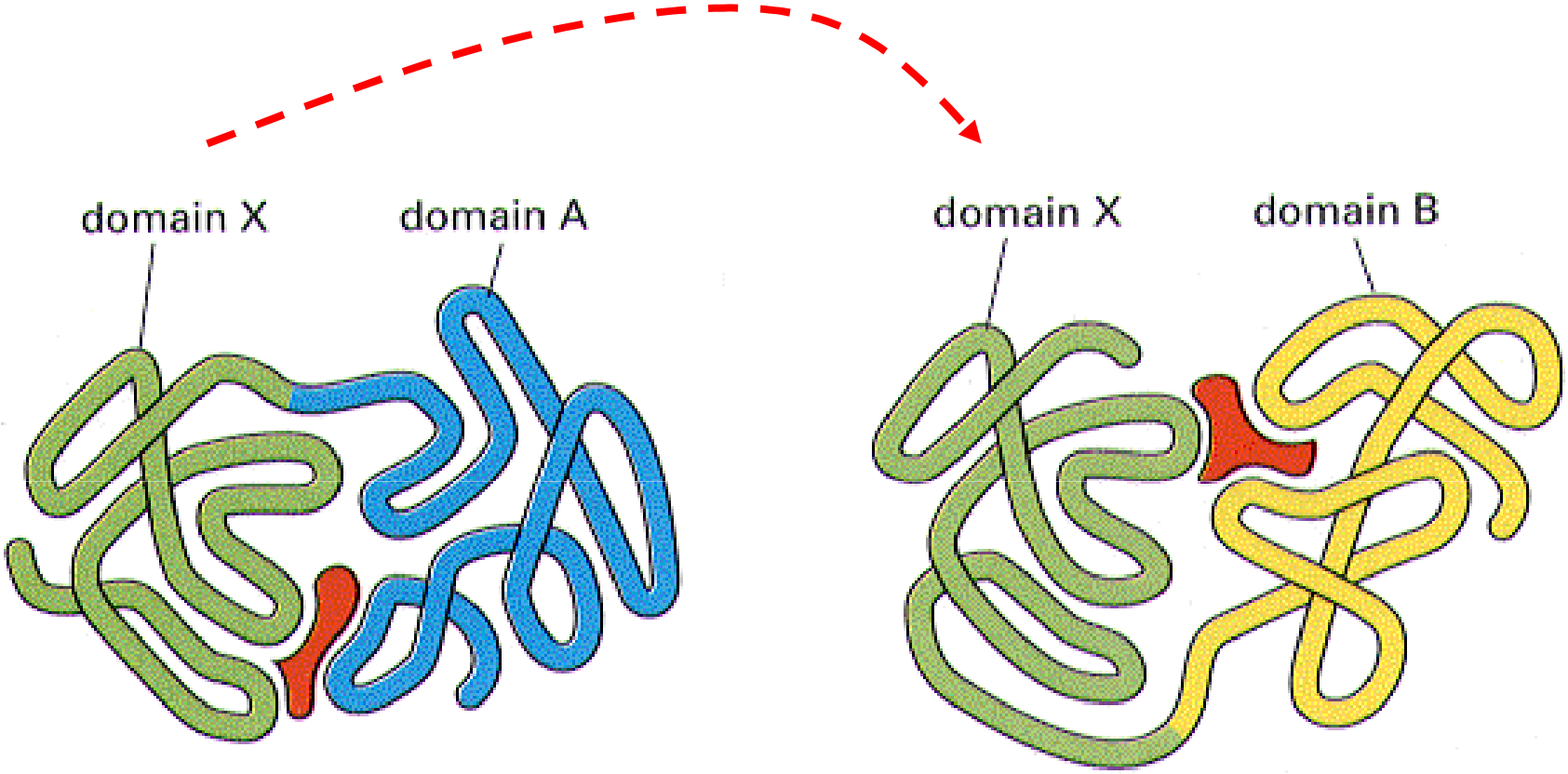
Figura 18.18 L
tramite (A) du
rimescolamen

Figura 18.17 I domini strutturali sono unità di una catena polipeptidica codificate da una serie contigua di nucleotidi. In questo esempio semplificato, ogni struttura secondaria del polipeptide rappresenta un singolo dominio strutturale. In realtà, la maggior parte dei domini strutturali comprende due o più strutture secondarie.



T.A. Brown
Genomi, III Ed.
EdiSES

Due proteine diverse presentano un dominio simile



Proteina 1

Proteina 2