

GENOMICA APPLICATA

Evolutionary genomics

Intro

- Understanding
 - the organization,
 - variation,
 - and expression of the **human genome** is central to the principles of genomic and precision medicine.

Evoluzione genomica

Novità evolutive



Figura 7.1

Il pelo e le corna
nei mammiferi, i fiori
e le squame sono ottimi
esempi di novità evolutive.

Come differiscono i genomi



Similarità e differenze tra sequenze genomiche compaiono

- (1) ai livelli di singole basi;
- (2) al livello di geni;
- (3) in blocchi su scala maggiore; e
- (4) al livello di interi genomi che hanno subito duplicazioni complete.

Come si originano le novità evolutive

- Cambiamenti in geni già esistenti
 - Mutazioni puntiformi
 - Inserzioni/delezioni
 - Traslocazioni
 - Riarrangiamenti cromosomici
- Acquisizione di nuovi geni
 - *De novo*
 - Duplicazione
 - Trasferimento orizzontale
 - Ibridazione

Come si originano le novità evolutive: geni *de novo*

Origine da sequenze non codificanti

Copyright © 2008 by the Genetics Society of America

DOI: 10.1534/genetics.107.084491

De Novo Origination of a New Protein-Coding Gene in *Saccharomyces cerevisiae*

Jing Cai,^{*,†,1} Ruoping Zhao,^{*,1} Huifeng Jiang^{*,†} and Wen Wang^{*,2}

ABSTRACT

Origination of new genes is an important mechanism generating genetic novelties during the evolution of an organism. Processes of creating new genes using preexisting genes as the raw materials are well characterized, such as exon shuffling, gene duplication, retroposition, gene fusion, and fission. However, the process of how a new gene is *de novo* created from noncoding sequence is largely unknown. On the basis of genome comparison among yeast species, we have identified a new *de novo* protein-coding gene, *BSC4* in *Saccharomyces cerevisiae*. The *BSC4* gene has an open reading frame (ORF) encoding a 132-amino-acid-long peptide, while there is no homologous ORF in all the sequenced genomes of other fungal species, including its closely related species such as *S. paradoxus* and *S. mikatae*. The functional protein-coding feature of the *BSC4* gene in *S. cerevisiae* is supported by population genetics, expression, proteomics, and synthetic lethal data. The evidence suggests that *BSC4* may be involved in the DNA repair pathway during the stationary phase of *S. cerevisiae* and contribute to the robustness of *S. cerevisiae*, when shifted to a nutrient-poor environment. Because the corresponding noncoding sequences in *S. paradoxus*, *S. mikatae*, and *S. bayanus* also transcribe, we propose that a new *de novo* protein-coding gene may have evolved from a previously expressed noncoding sequence.

Come si originano le novità evolutive: geni *de novo*

Origine da sequenze non codificanti

Letter

On the origin of new genes in *Drosophila*

Qi Zhou,^{1,2,4} Guojie Zhang,^{1,2,3,4} Yue Zhang,^{1,4} Shiyu Xu,¹ Ruoping Zhao,¹
Zubing Zhan,^{1,2} Xin Li,^{1,2} Yun Ding,^{1,2} Shuang Yang,^{1,3} and Wen Wang^{1,5}

Several mechanisms have been proposed to account for the origination of new genes. Despite extensive case studies, the general principles governing this fundamental process are still unclear at the whole-genome level. Here, we unveil genome-wide patterns for the mutational mechanisms leading to new genes and their subsequent lineage-specific evolution at different time nodes in the *Drosophila melanogaster* species subgroup. We find that (1) tandem gene duplication has generated ~80% of the nascent duplicates that are limited to single species (*D. melanogaster* or *Drosophila yakuba*); (2) the most abundant new genes shared by multiple species (44.1%) are dispersed duplicates, and are more likely to be retained and be functional; (3) *de novo* gene origination from noncoding sequences plays an unexpectedly important role during the origin of new genes, and is responsible for 11.9% of the new genes; (4) retroposition is also an important mechanism, and had generated ~10% of the new genes; (5) ~30% of the new genes in the *D. melanogaster* species complex recruited various genomic sequences and formed chimeric gene structures, suggesting structure innovation as an important way to help fixation of new genes; and (6) the rate of the origin of new functional genes is estimated to be five to 11 genes per million years in the *D. melanogaster* subgroup. Finally, we survey gene frequencies among 19 globally derived strains for *D. melanogaster*-specific new genes and reveal that 44.4% of them show copy number polymorphisms within a population. In conclusion, we provide a panoramic picture for the origin of new genes in *Drosophila* species.

Recent de novo origin of human protein-coding genes

David G. Knowles and Aoife McLysaght¹

Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland

The origin of new genes is extremely important to evolutionary innovation. Most new genes arise from existing genes through duplication or recombination. The origin of new genes from noncoding DNA is extremely rare, and very few eukaryotic examples are known. We present evidence for the de novo origin of at least three human protein-coding genes since the divergence with chimp. Each of these genes has no protein-coding homologs in any other genome, but is supported by evidence from expression and, importantly, proteomics data. The absence of these genes in chimp and macaque cannot be explained by sequencing gaps or annotation error. High-quality sequence data indicate that these loci are noncoding DNA in other primates. Furthermore, chimp, gorilla, gibbon, and macaque share the same disabling sequence difference, supporting the inference that the ancestral sequence was noncoding over the alternative possibility of parallel gene inactivation in multiple primate lineages. The genes are not well characterized, but interestingly, one of them was first identified as an up-regulated gene in chronic lymphocytic leukemia. This is the first evidence for entirely novel human-specific protein-coding genes originating from ancestrally noncoding sequences. We estimate that 0.075% of human genes may have originated through this mechanism leading to a total expectation of 18 such cases in a genome of 24,000 protein-coding genes.

[Supplemental material is available online at <http://www.genome.org>. The sequence data from this study have been submitted to GenBank (<http://www.ncbi.nlm.nih.gov/Genbank/>) under accession nos. FJ713693, FJ713696, and FJ713697.]

New genes are a rich substrate for evolution to act upon. New genes frequently arise through duplication of existing genes, or through fusion, fission, or exon shuffling between genes (Long et al. 2003). Origination of genes from noncoding DNA is extremely rare: A few eukaryotic examples are known in yeast and *Drosophila* (Levine et al. 2006; Begun et al. 2007; Cai et al. 2008; Zhou et al. 2008) and

we produced span 91% and 85% of the human and chimp genomes, respectively, and 21,195 (94%) of the 22,568 human protein-coding genes annotated by Ensembl are located within these blocks. Because we only used 1:1 orthologous regions, lineage-specific segmental duplications are excluded from this analysis.

We exploited the extremely high gene order conservation

Duplicazioni

Specie	Geni duplicati (%)
Bacteria (batteri)	
<i>Mycoplasma pneumoniae</i>	44
<i>Helicobacter pylori</i>	17
<i>Haemophilus influenzae</i>	17
Archaea (archei)	
<i>Archaeoglobus fulgidus</i>	30
Eukarya (eucarioti)	
<i>Saccharomyces cerevisiae</i>	30
<i>Caenorhabditis elegans</i>	49
<i>Drosophila melanogaster</i>	41
<i>Arabidopsis thaliana</i>	65
<i>Homo sapiens</i>	38

Duplicazione genica: destino dei geni duplicati

1. Tutte le copie mantengono la stessa funzione.



2. Alcune copie scompaiono.



3. Alcune copie evolvono con una nuova funzione.



Duplicazione genica: destino dei geni duplicati

1. Tutte le copie mantengono la stessa funzione.



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

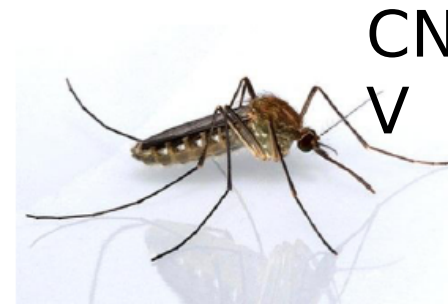
Genome Source	Number of rRNA sets	Number of tRNA genes ^a	Approximate genome size (bp)
Human mitochondrion	1	22	2×10^4
<i>Nicotiana tabacum</i> chloroplast	2	37	2×10^5
<i>Escherichia coli</i>	7	~ 100	4×10^6
<i>Neurospora crassa</i>	~ 100	~ 2,600	2×10^7
<i>Saccharomyces cerevisiae</i>	~ 140	~ 360	5×10^7
<i>Caenorhabditis elegans</i>	~ 55	~ 300	8×10^7
<i>Tetrahymena thermophila</i>	1	~ 800 ^c	2×10^8
<i>Drosophila melanogaster</i>	120-240	590-900	2×10^8
<i>Physarum polycephalum</i>	80-280	~ 1,050	5×10^8
<i>Euglena gracilis</i>	800-1,000	~ 740	2×10^9
Human	~ 300	~ 1,300	3×10^9
<i>Rattus norvegicus</i>	150-170	~ 6,500	3×10^9
<i>Xenopus laevis</i>	500-760	6,500-7,800	8×10^9

Duplicazione genica: destino dei geni duplicati

1. Tutte le copie mantengono la stessa funzione.



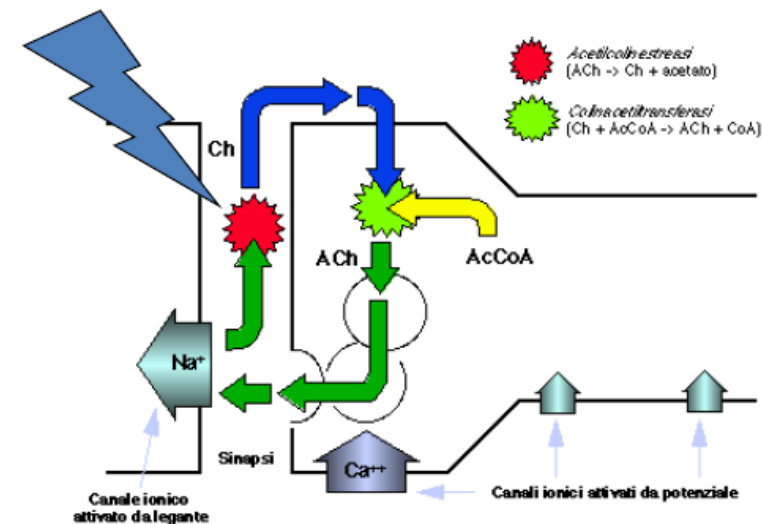
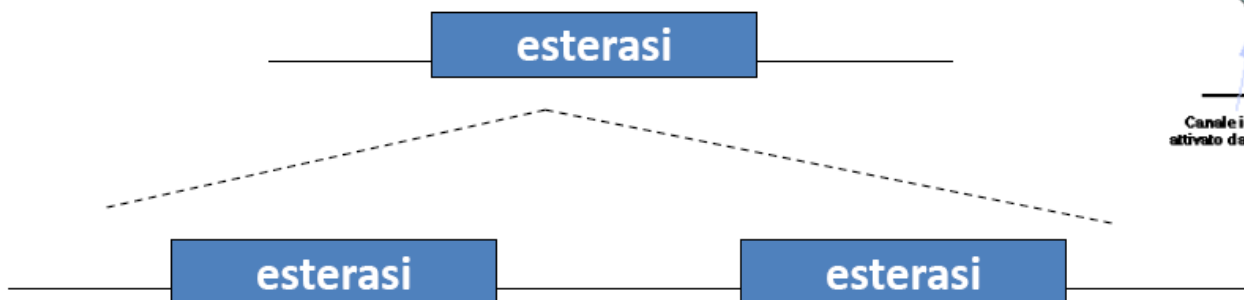
Resistenza ai pesticidi: **zanzare**



Organofosfati: neurotossine che colpiscono il SN degli insetti

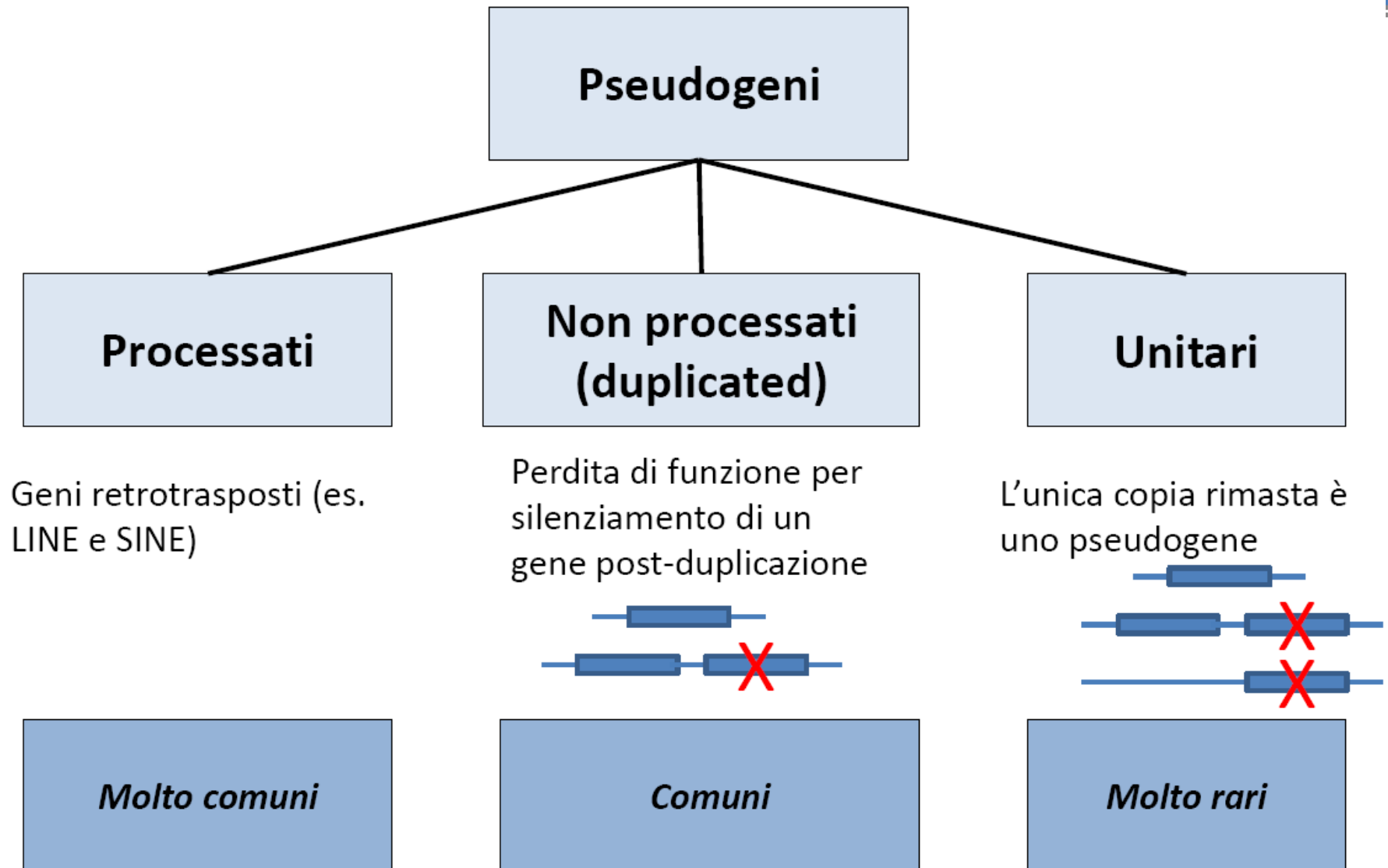
Resistenza:

Amplificazione del gene dell'Esterasi che blocca il pesticida prima dell'arrivo al terminale nervoso



Duplicazione genica: destino dei geni duplicati

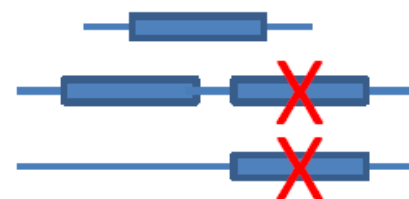
2. Alcune copie scompaiono.



Duplicazione genica: destino dei geni duplicati

2. Alcune copie scompaiono. Pseudogeni unitari

Come si può fissare in una popolazione la perdita di un gene?
Per deriva se la selezione è rilassata

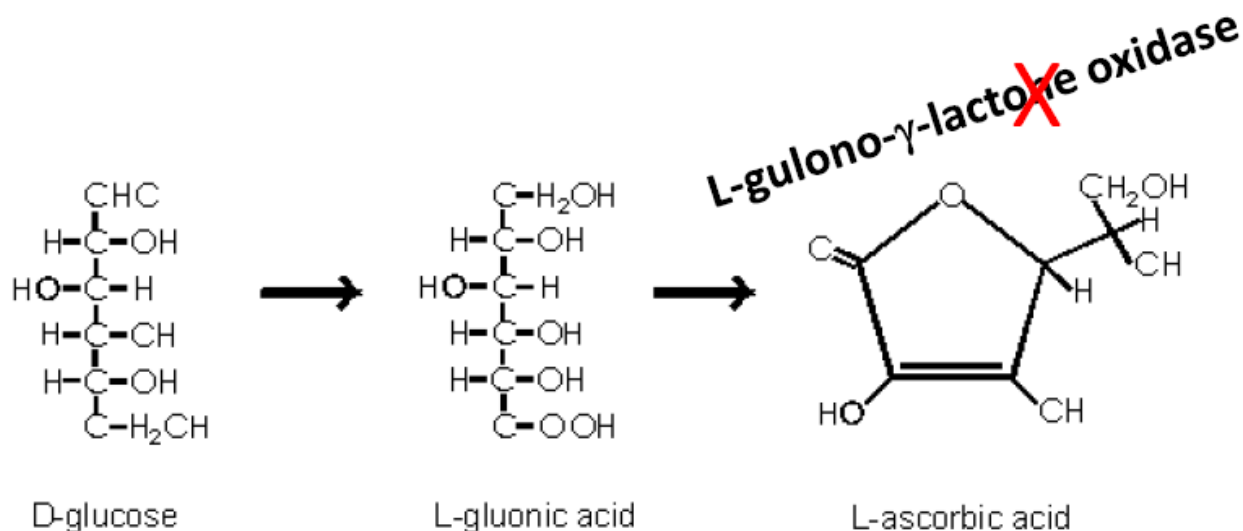


Ipoascorbemia o scorbuto = incapacità di sintetisi dell'acido ascorbico

Cavie

Uomo si ammalano di scorbuto se non assumo acido ascorbico (vitamina C)

Trote

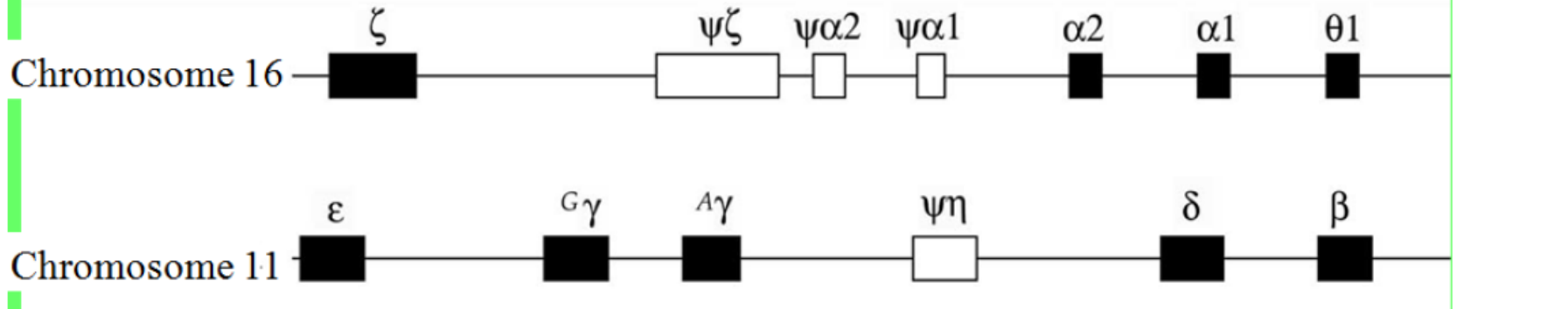


Duplicazione genica: destino dei geni duplicati

2. Alcune copie scompaiono. Pseudogeni non processati



Molti esempi nelle globine



Pseudogene	TATA box	Init. codon	Frame shift	Premature stop	Essential amino acid	Splice GT/AG	Stop codon	Polyadenilation signal AATAAA
Human $\psi\alpha1$		+	+	+	+	+	+	+
Human $\psi\zeta1$				+				
Mouse $\psi\alpha3$	+		+	+		+		
Mouse $\psi\alpha4$			+		+			
Mouse $\beta h3$	?	+	+	+	+	+	?	?
Goat $\psi\beta^x$	+		+	+	+	+	+	+
Goat $\psi\beta^z$	+		+	+	+	+	+	+
Rabbit $\psi\beta2$			+	+	+	+		

Duplicazione genica: destino dei geni duplicati

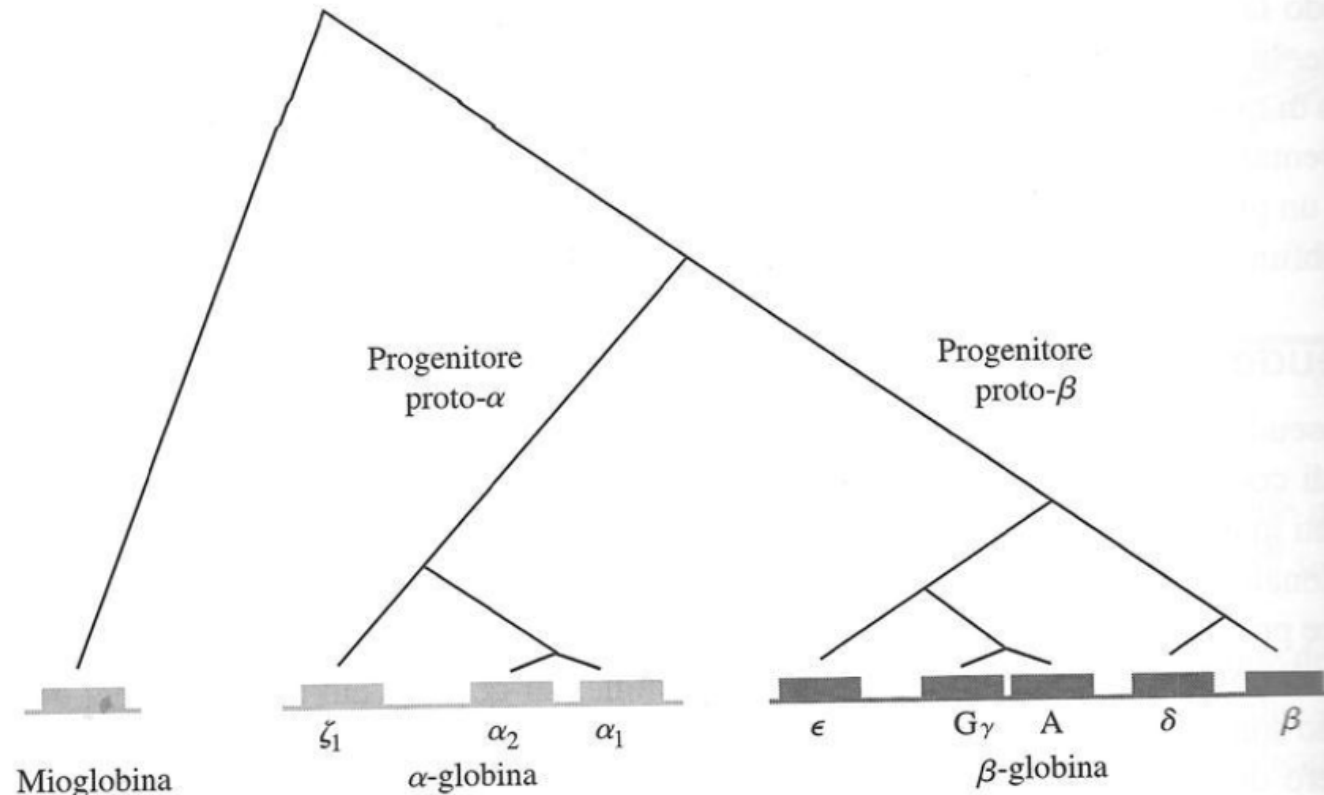
3. Alcune copie evolvono con una nuova funzione.

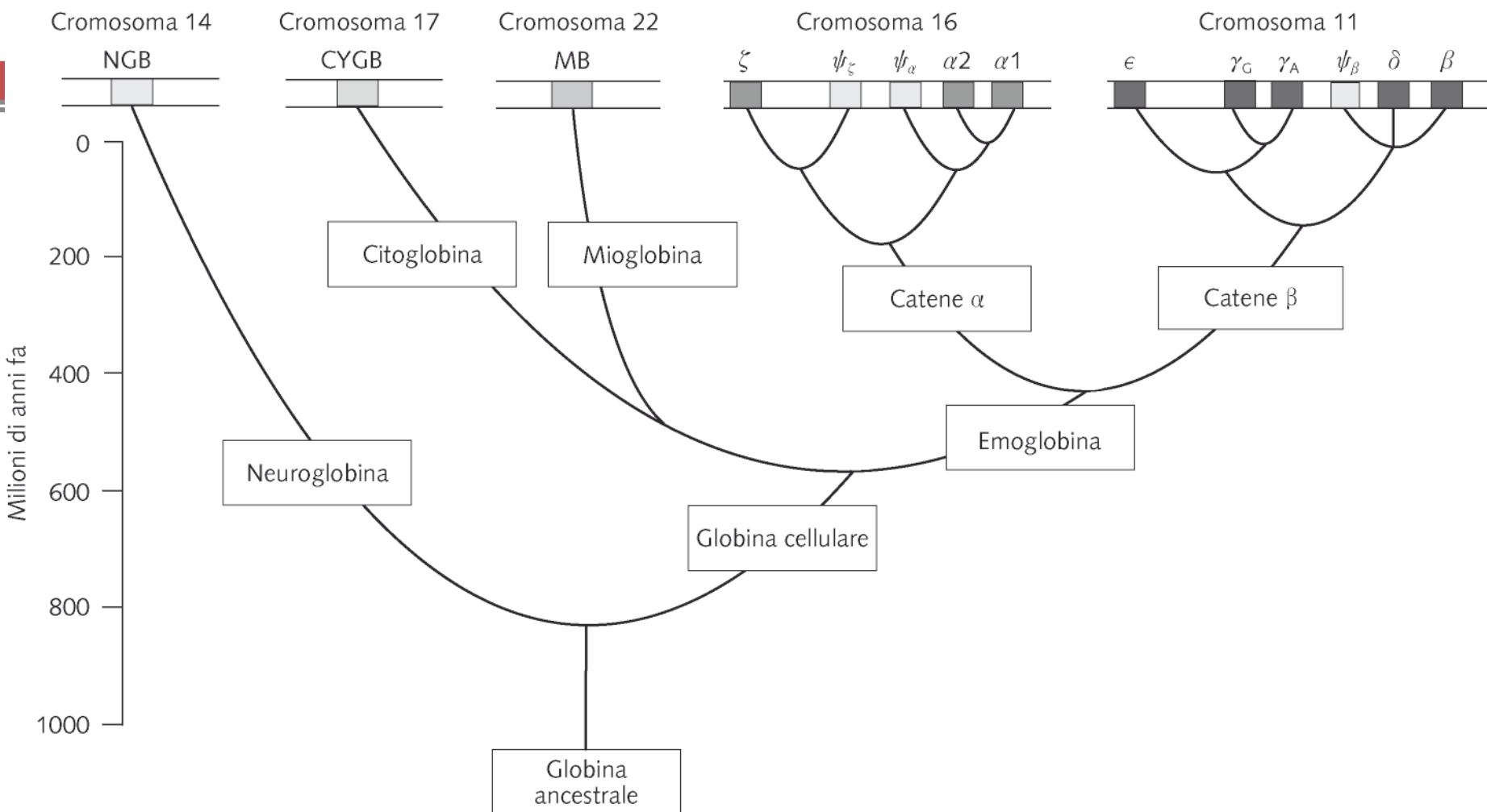


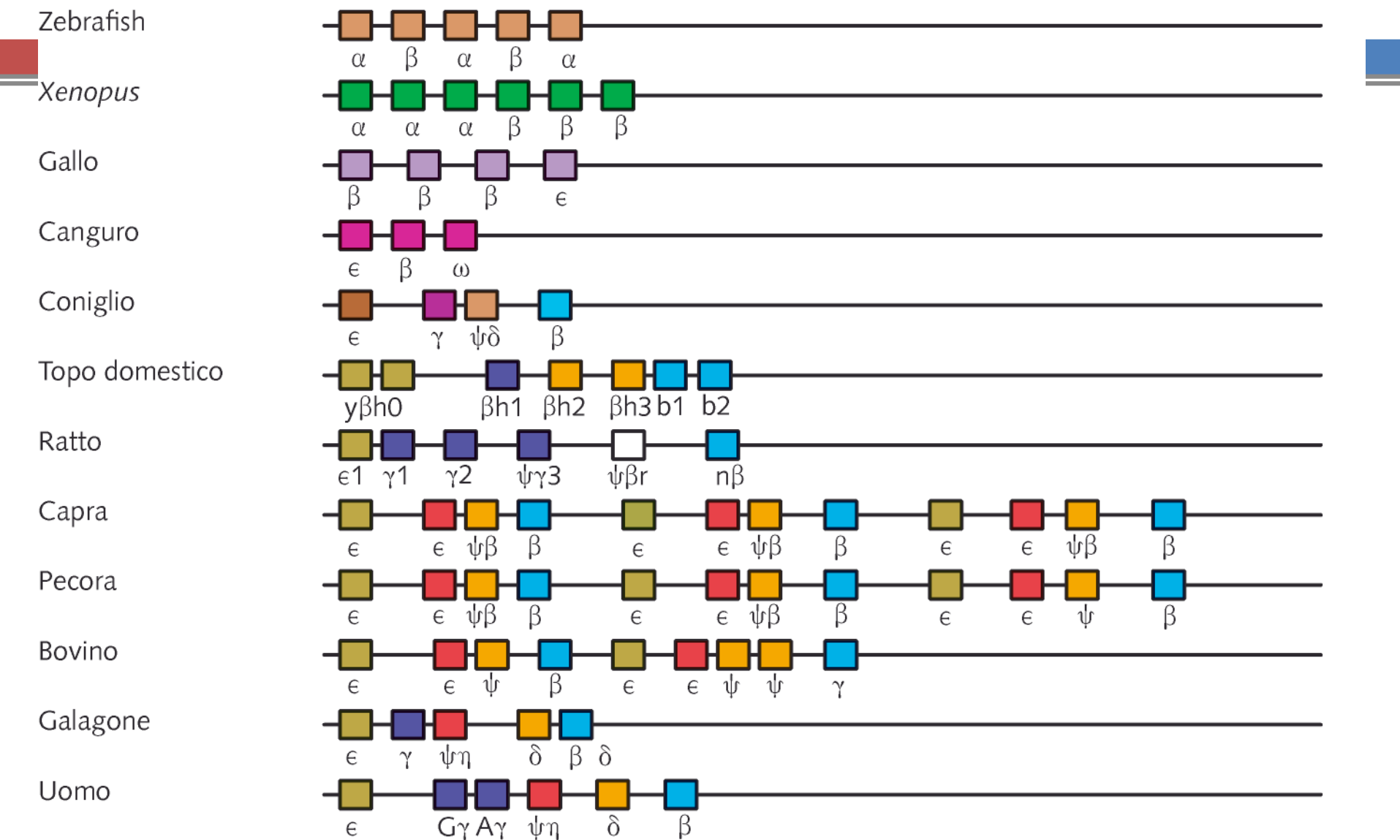
Rosso e blu sono PARALOGHI

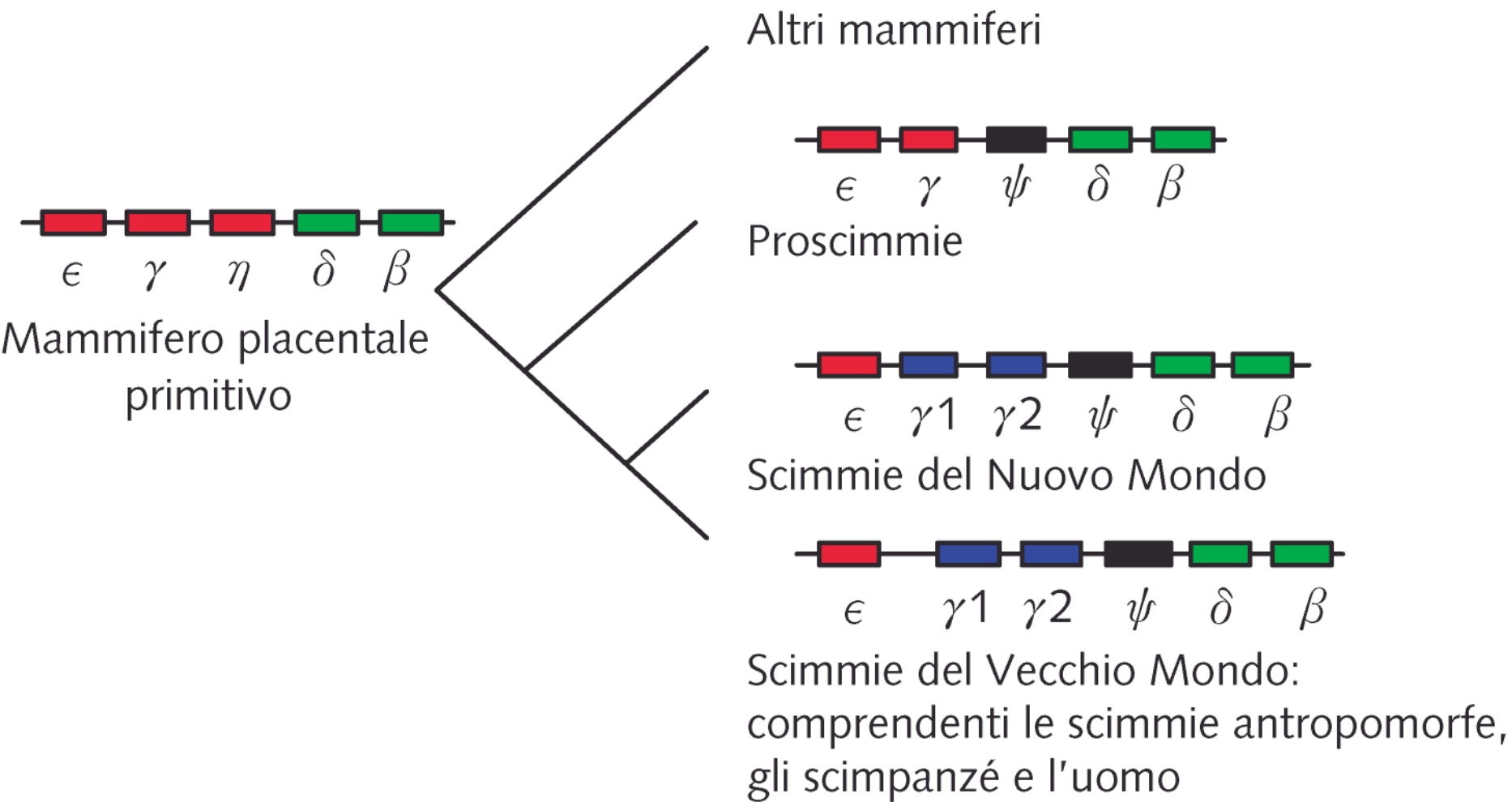
Figura 7.9

Vi sono, per esempio nella nostra specie, diversi geni per le globine, probabilmente derivati da duplicazioni avvenute nel passato. Questo albero mostra la “storia” dei geni per le globine, che spiega la presenza dei paraloghi.

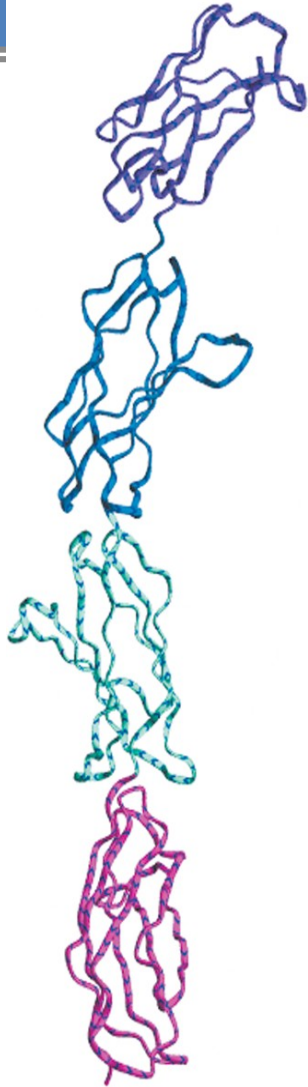








La duplicazione può influenzare singoli esoni



La fibronectina, una grande proteina extracellulare che interviene nell'adesione e nella migrazione cellulare, è una **proteina modulare** contenente più ripetizioni in tandem di tre tipi di dominio denominati F1, F2 e F3. Codificati da singoli esoni

Espansioni delle famiglie

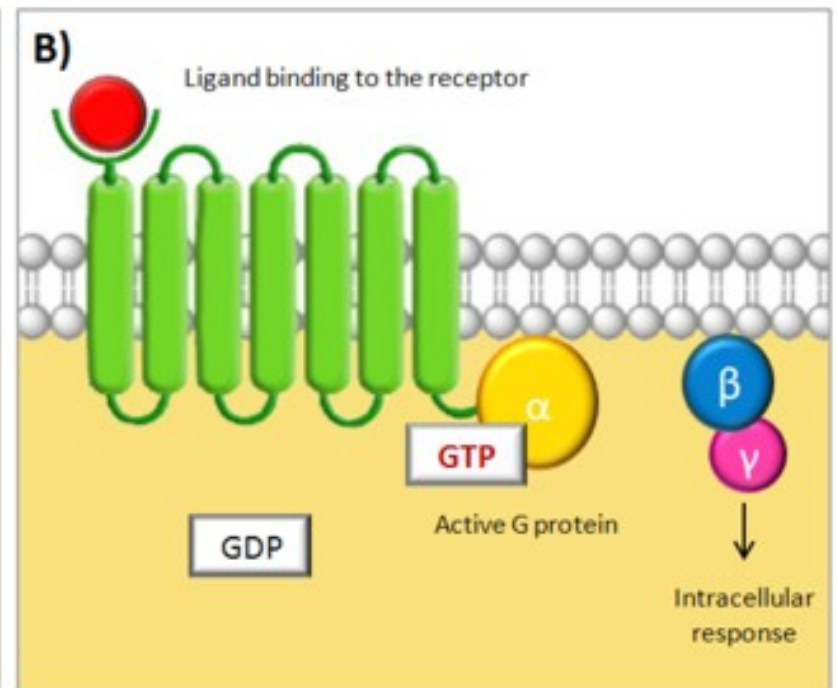
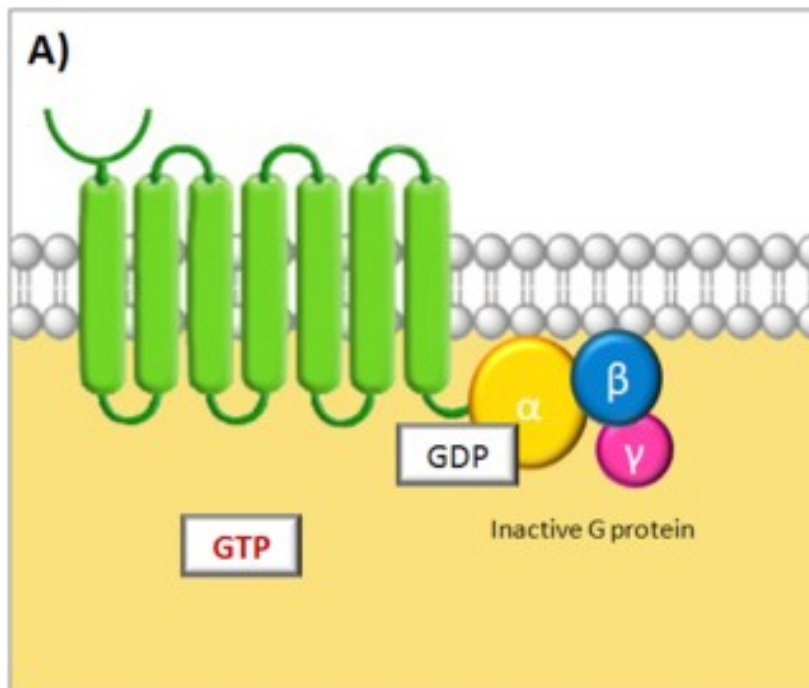
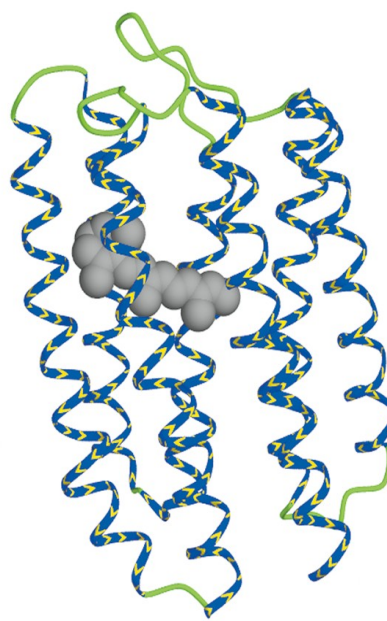
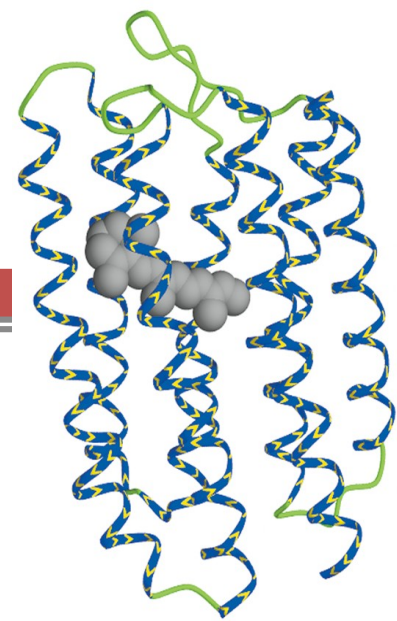
I recettori accoppiati a proteine C (GPCR) sono una grande famiglia di proteine transmembrana che intervengono nella trasduzione dei segnali nelle cellule.

Hanno in comune una struttura contenente 7 eliche transmembrana, disposte in una topologia comune.

Questa famiglia ha la stessa età degli eucarioti ed è grande ed eterogenea.

I genomi dei mammiferi contengono circa 1500 – 2000 GPCR, che costituiscono circa il 3 - 5% del genoma.

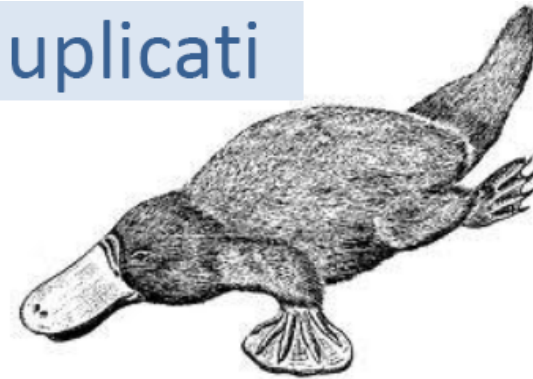
Il genoma umano ha circa 1000 geni per i **recettori olfattivi**, di cui soltanto il 40% sono attivi. Il genoma murino ne ha circa 1300, di cui l'80% sono attivi.



Duplicazione genica: destino dei geni duplicati

3. Alcune copie evolvono con una nuova funzione.

Novità evolutiva nell'ornitorinco: capacità di produrre veleno



Vol 453 | 8 May 2008 | doi:10.1038/nature06936

nature

ARTICLES

Genome analysis of the platypus reveals unique signatures of evolution

A list of authors and their affiliations appears at the end of the paper

We present a draft genome sequence of the platypus, *Ornithorhynchus anatinus*. This monotreme exhibits a fascinating combination of reptilian and mammalian characters. For example, platypuses have a coat of fur adapted to an aquatic lifestyle; platypus females lactate, yet lay eggs; and males are equipped with venom similar to that of reptiles. Analysis of the first monotreme genome aligned these features with genetic innovations. We find that reptile and platypus venom proteins have been co-opted independently from the same gene families; milk protein genes are conserved despite platypuses laying eggs; and immune gene family expansions are directly related to platypus biology. Expansions of protein, non-protein-coding RNA and microRNA families, as well as repeat elements, are identified. Sequencing of this genome now provides a valuable resource for deep mammalian comparative analyses, as well as for monotreme biology and conservation.

Duplicazione genica: destino dei geni duplicati

3. Alcune copie evolvono con una nuova funzione.

Le proteine del veleno derivano da copie duplicate neofunzionalizzate di geni già presenti nel genoma.

Una duplicazione simile ma indipendente si osserva nei rettili (convergenza evolutiva!)

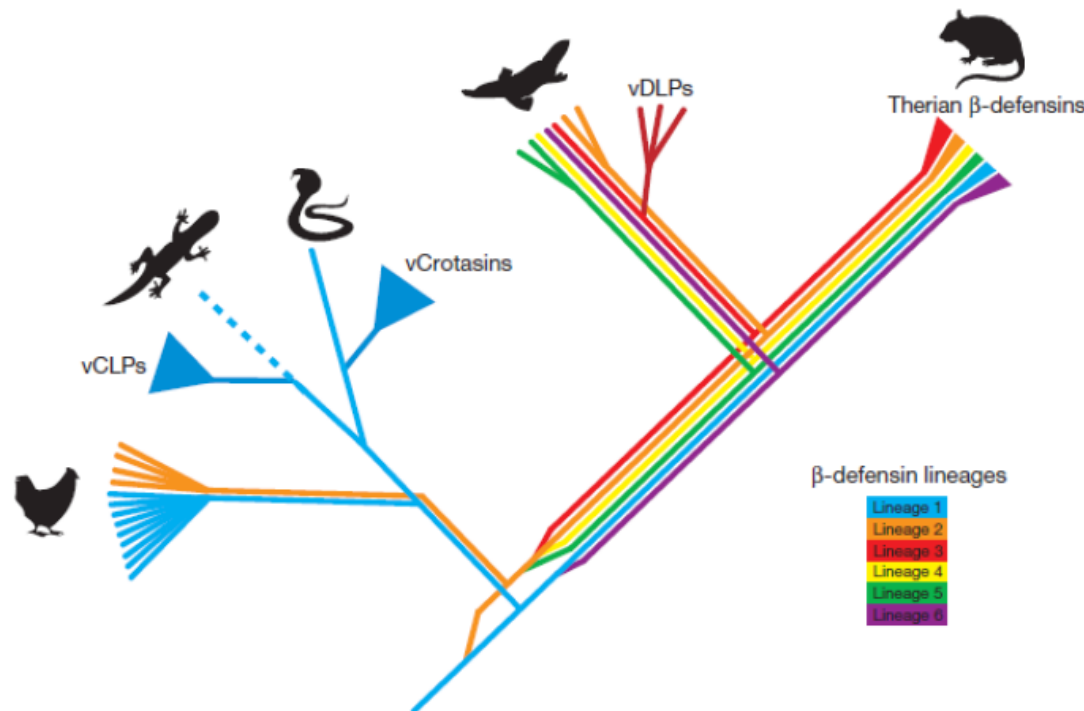
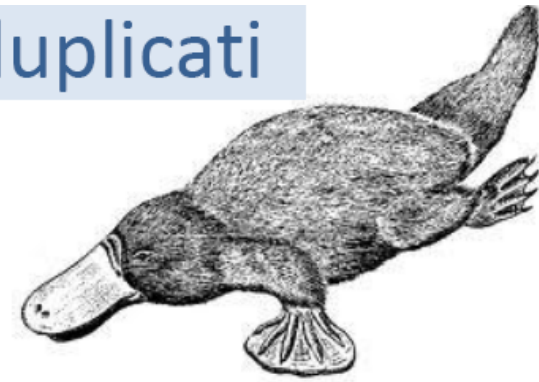
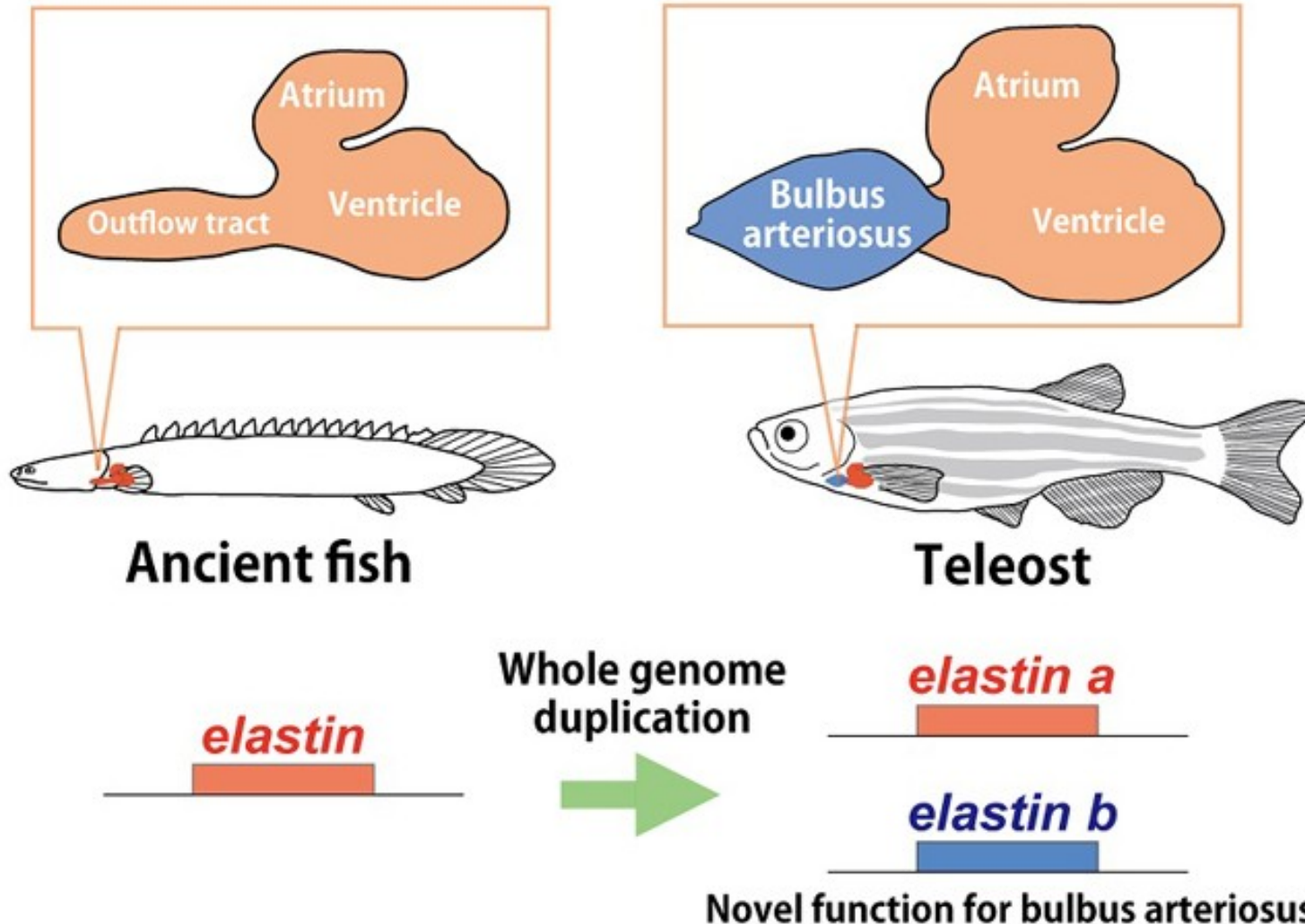


Figure 4 | The evolution of β -defensin peptides in platypus venom gland. The diagram illustrates separate gene duplications in different parts of the phylogeny for platypus venom defensin-like peptides (vDLPs), for lizard

venom crotoamine-like peptides (vCLPs) and for snake venom crotoamines. These venom proteins have thus been co-opted from pre-existing non-toxin homologues independently in platypus and in lizards and snakes⁴⁸.

Duplicazione genica: destino dei geni duplicati



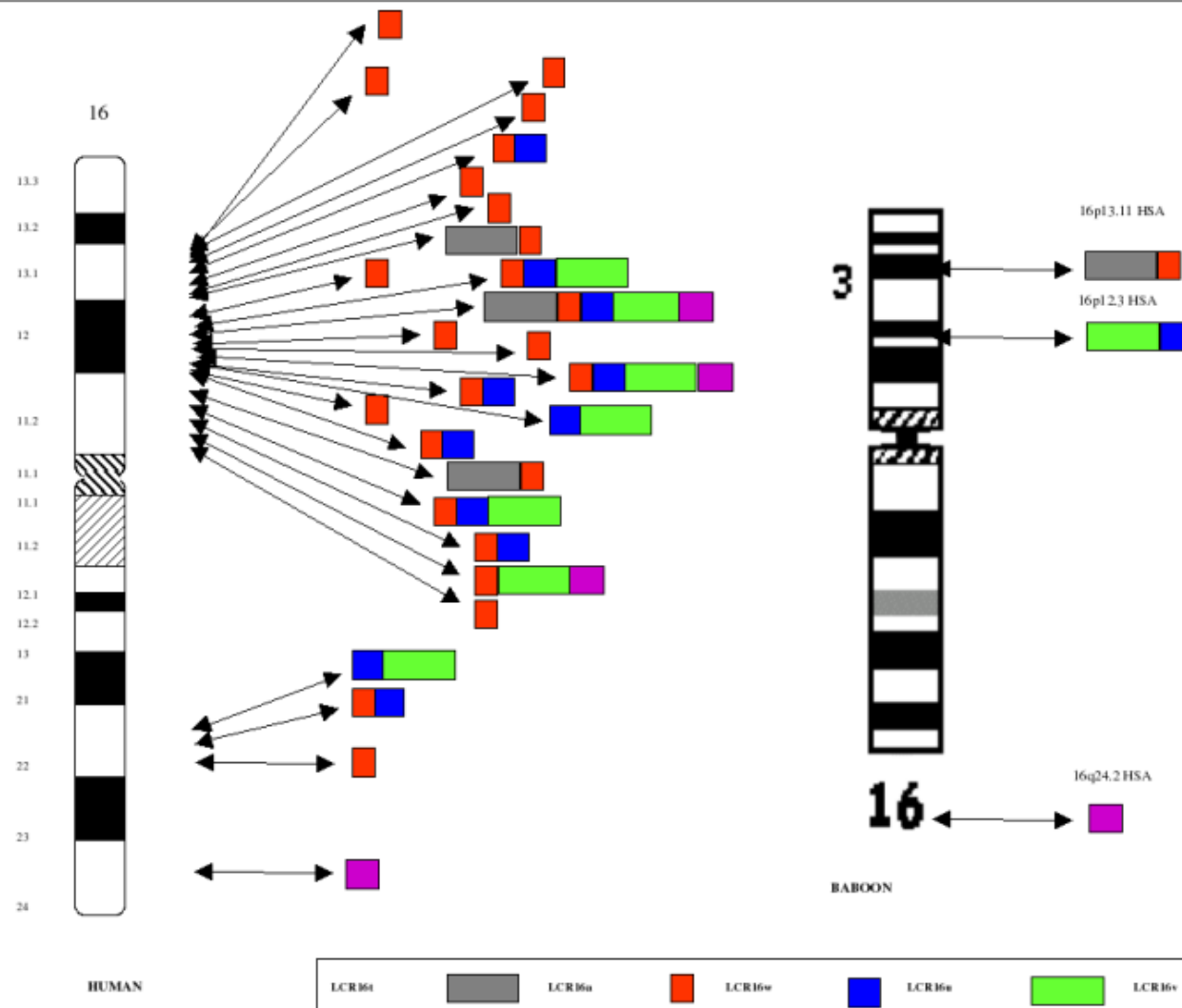
Duplicazioni su larga scala



I genomi di molte specie contengono duplicazioni di regioni multigeniche, la cui lunghezza varia da specie a specie.

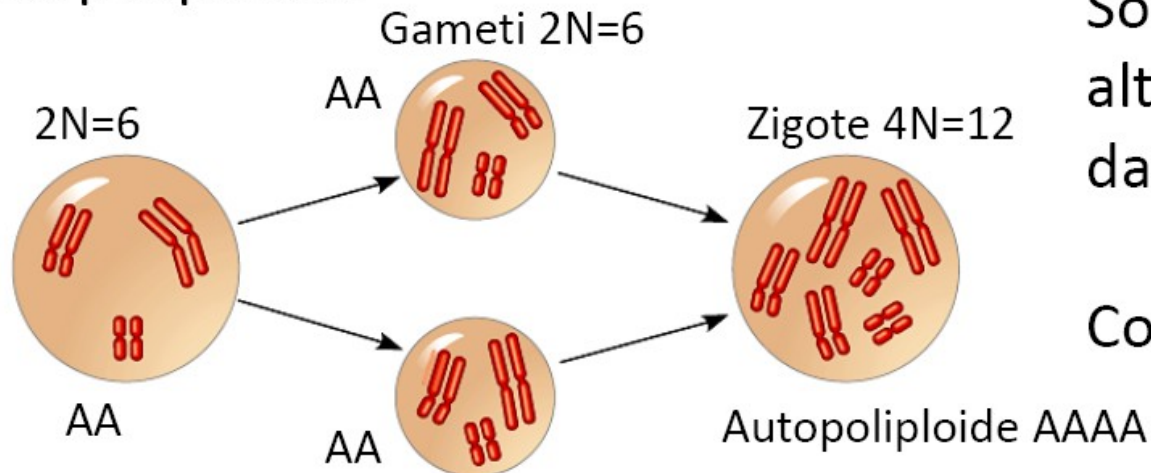
- Le duplicazioni segmentali su grande scala sono una componente importante nella differenza tra genoma dell'uomo e genoma dello scimpanzé, interessando circa il 2,7% del genoma.

Duplicazioni su larga scala



Come si originano le novità evolutive: duplicazioni genomiche

Autopoliploidia

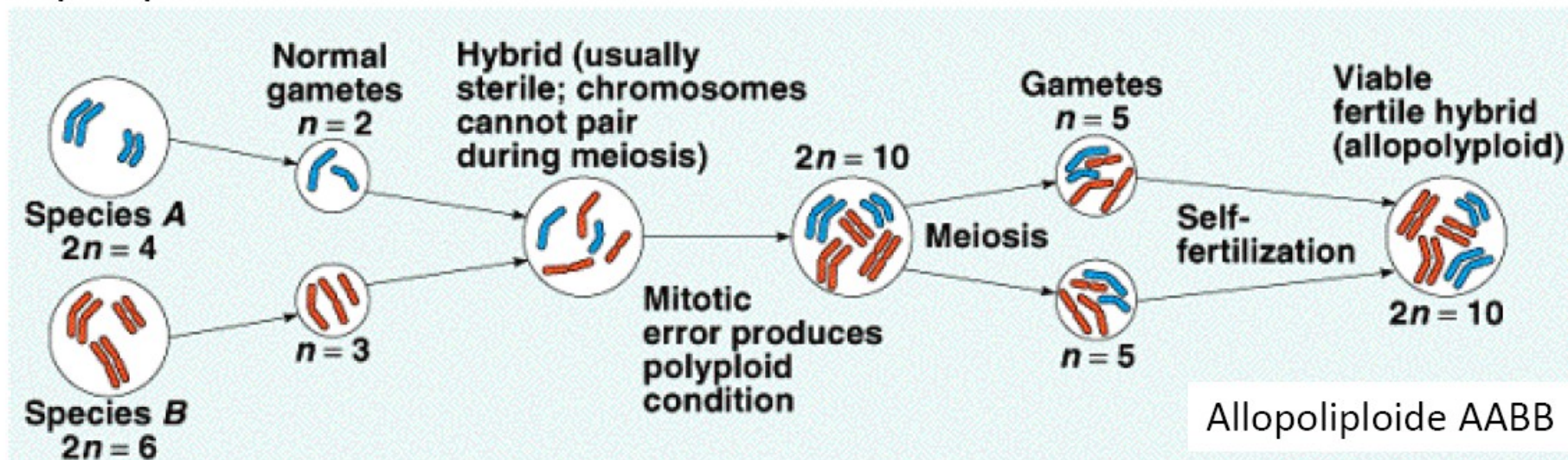


Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Solo l'incrocio con un altro autopoliploide darà prole fertile

Comune fra le piante

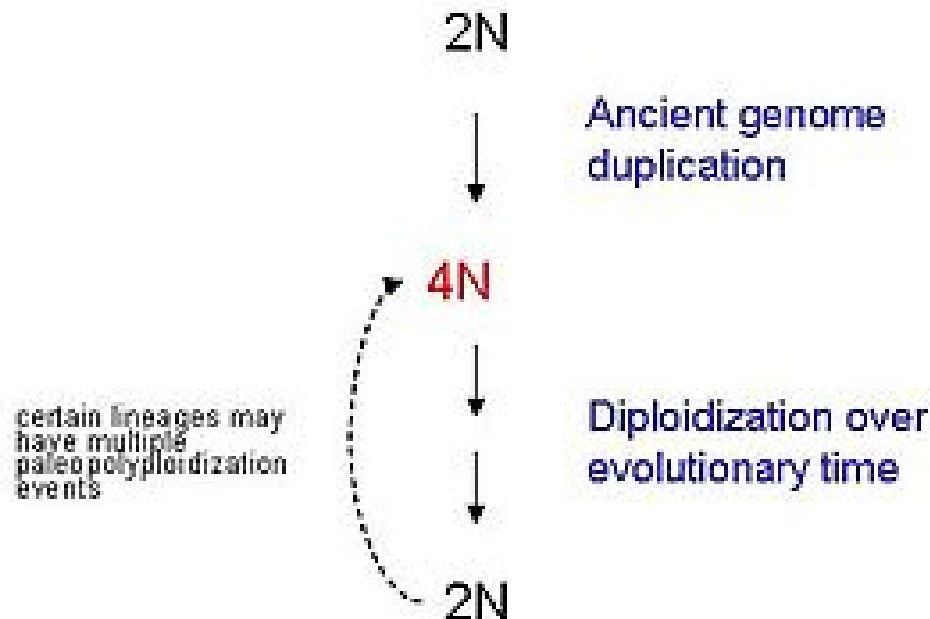
Allopoliploidia



Duplicazioni dell'intero genoma

Paleoploidia.

Serve come prova la presenza, nell'intero genoma, di coppie di geni omologhi che *si presentano nello stesso ordine (colinearità)*; o l'evidenza di un «orologio molecolare» che indica uguali tempi di divergenza in molte coppie di omologhi.



Yeast

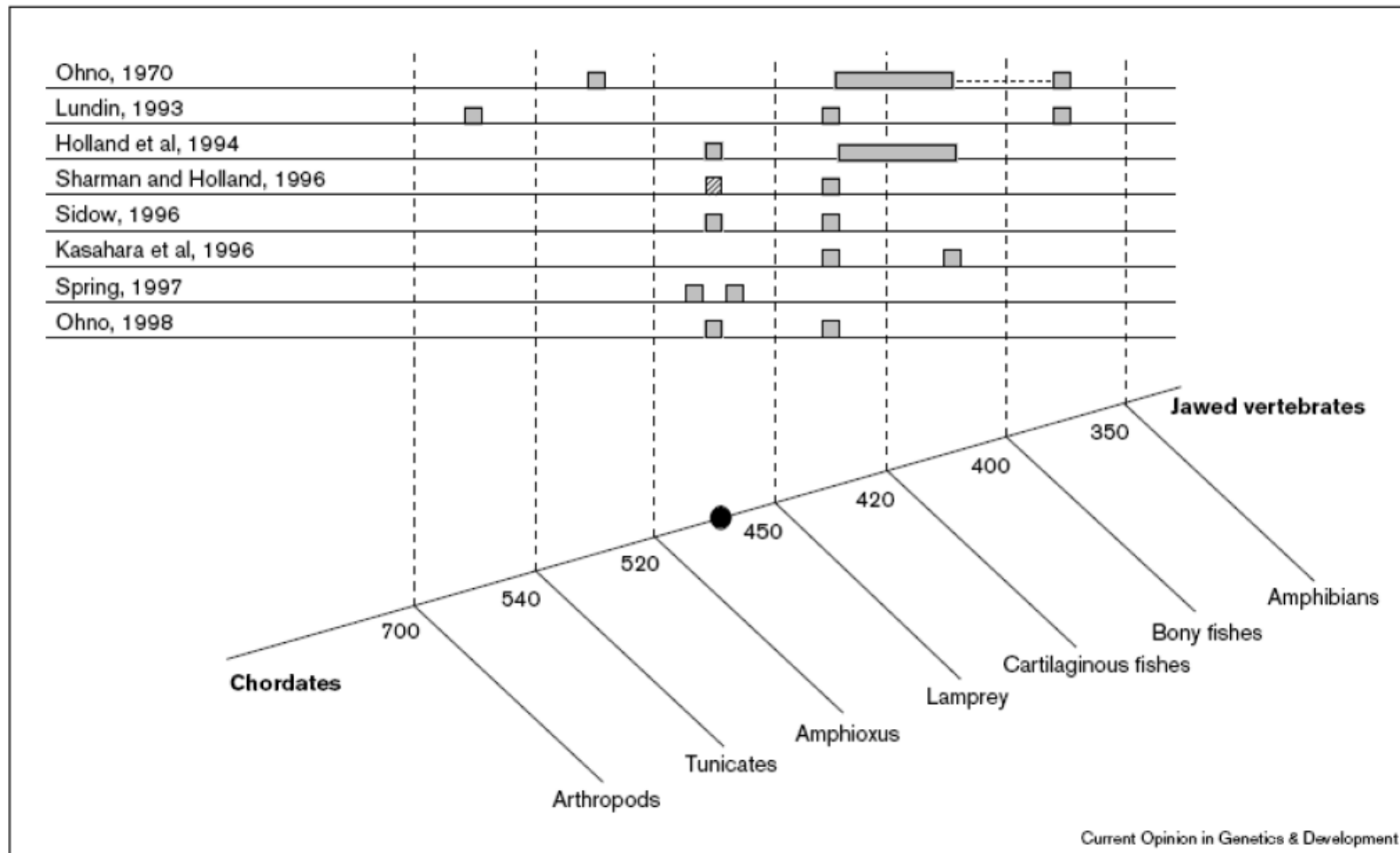


Il genoma del lievito subì una duplicazione circa 10^8 anni fa. Gli effetti sono oscurati dai successivi riarrangiamenti cromosomici e dalla massiva perdita di materiale duplicato.

Comunque il genoma del lievito contiene 55 regioni duplicate, lunghe in media 55 kb, che coprono insieme il $\sim 50\%$ del genoma e comprendono 376 coppie di geni omologhi.

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis Susumu Ohno



Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

1R: il genoma raddoppia, **2R**: il genoma quadruplica (aggiungiamo una fish-specific genome duplication: **3R** per la linea dei pesci)

Ancora molta discussione sul numero e la datazione delle duplicazioni

Si vedono delle tracce della 2R (esempio: geni Hox), ma altre sono cancellate da eventi successivi

Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

Si vedono delle tracce della 2R (esempio: geni Hox)

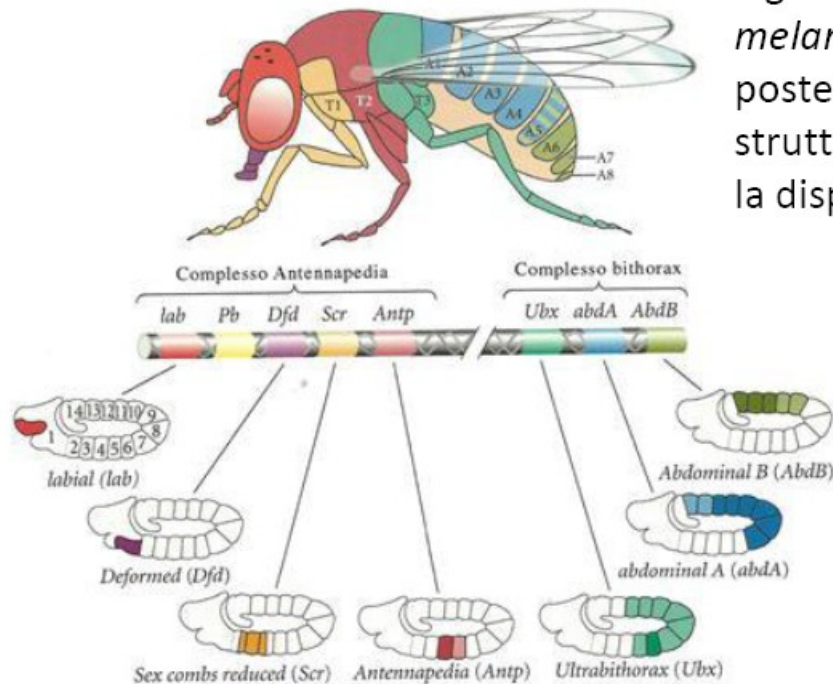
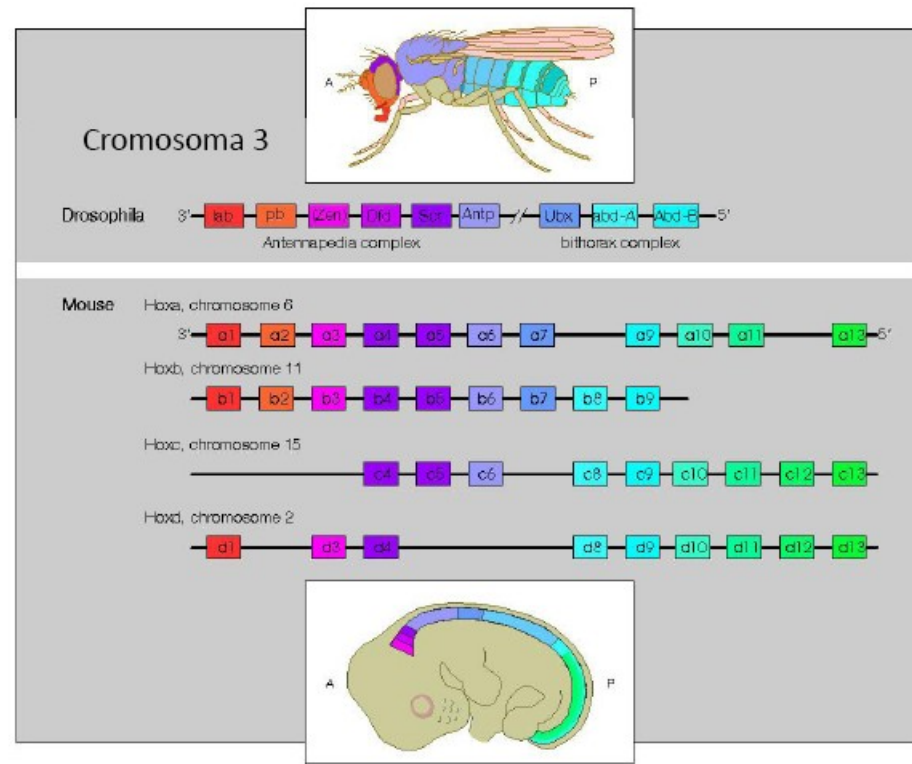


Fig. 9.10: L' espressione dei geni Hox nell'embrione di *D. melanogaster* (in basso) in specifici domini dell'asse antero-posteriore del corpo definisce la posizione di specifiche strutture lungo lo stesso asse nell'adulto (in alto). Al centro la disposizione dei geni Hox sul cromosoma



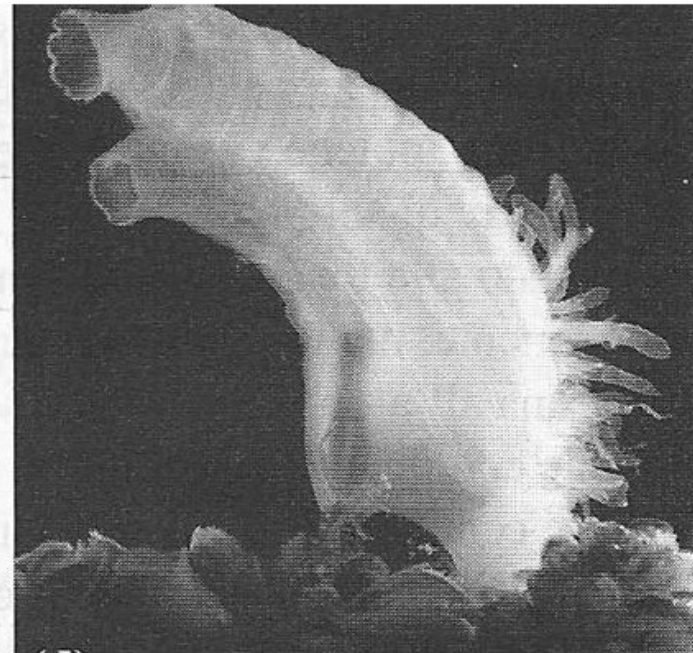
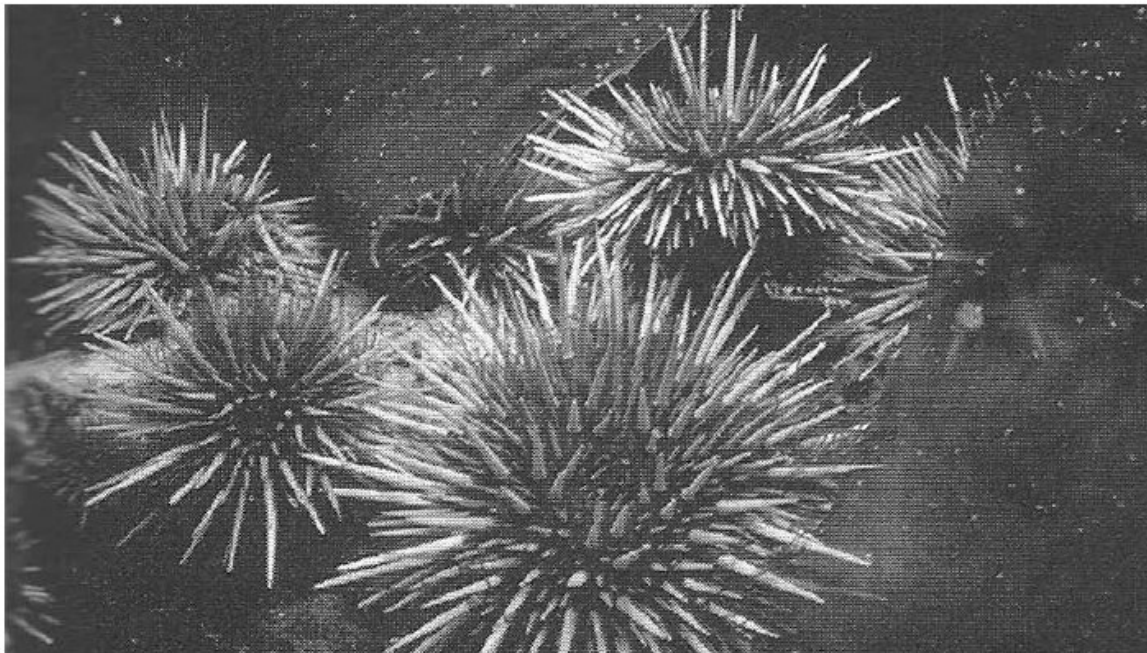
>Complessità
>Nr geni omeotici

4 clusters nei vertebrati
Geni con similarità di
sequenza con quelli di
Drosophila

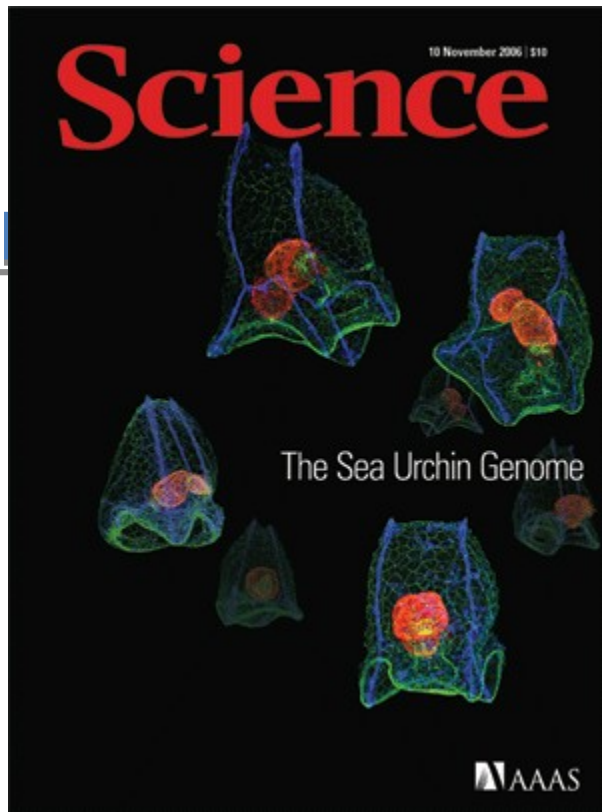
Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

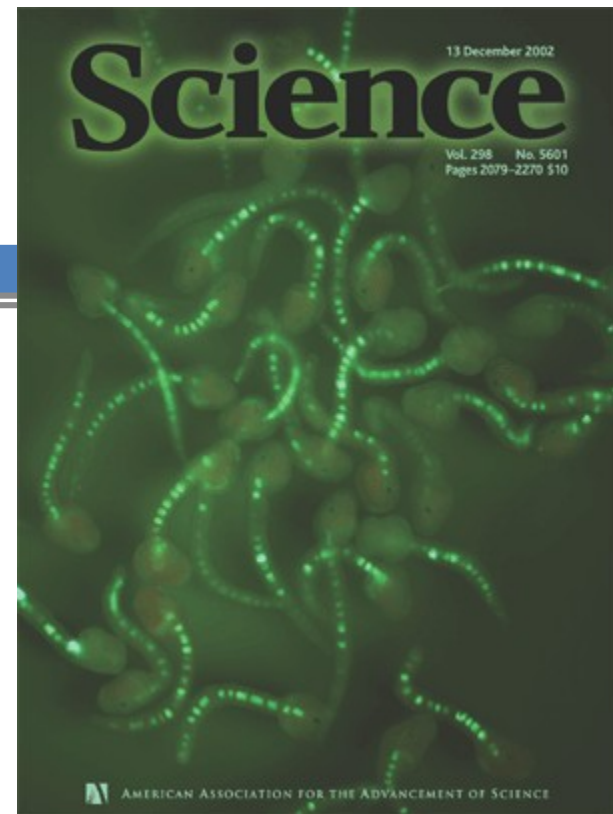
Si vedono delle tracce della 2R (esempio: sequenziamento di genomi)



Il sequenziamento dei genomi di riccio di mare (*Strongylocentrotus purpuratus*, echinoderma) e dell'ascidia (*Ciona intestinalis*, tunicato) mostrano numerosi geni in **proporzione 1 a 4 rispetto agli omologhi vertebrati**



The sea urchin, *Strongylocentrotus purpuratus*, has been a premier model for understanding the gene networks that underlie development. Shown are sea urchin pluteus larvae stained to reveal the ectoderm (green), mesoderm

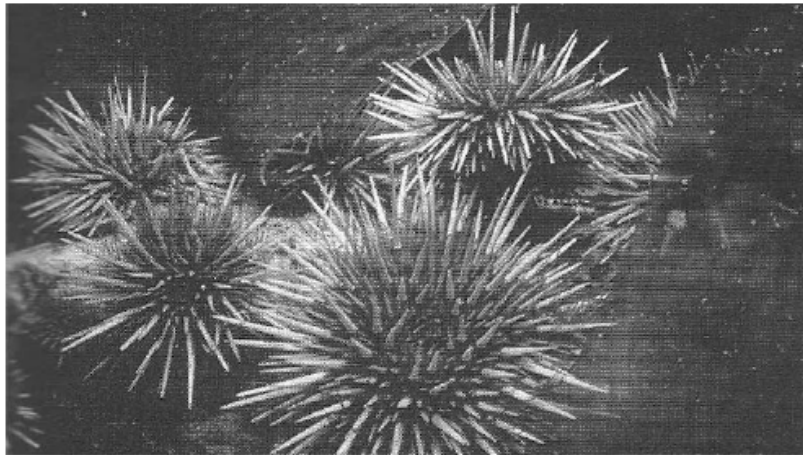


A group of about 30 tadpole-stage *Ciona intestinalis* embryos 12 to 14 hours after fertilization. The embryos were electroporated at the one-cell stage with a fusion gene containing the 5' cis-regulatory region of the *Brachyury* gene attached to a green fluorescent protein (GFP) reporter gene. The *Brachyury* regulatory DNA directs GFP expression in the

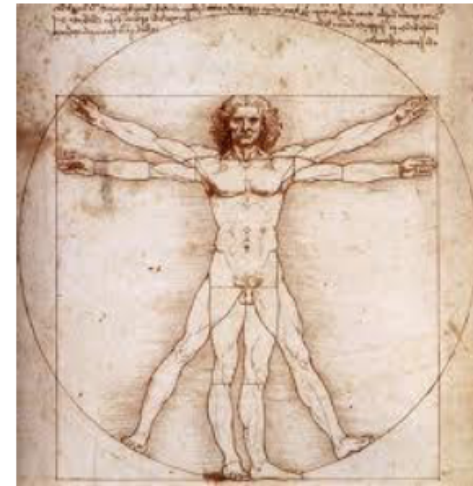
Come si originano le novità evolutive: duplicazioni genomiche

2R Hypothesis

Tuttavia....



~ 23.000 geni



~ 24.000 geni

Nel complesso il riccio di mare ha tanti geni quanti l'uomo: **dopo le duplicazioni genomiche molte copie non sopravvivono come geni funzionanti**

Nelle piante

Il sequenziamento di *A. thaliana* ha rilevato almeno 2 duplicazioni genomiche.

La maggior parte delle piante sono poliploidi.

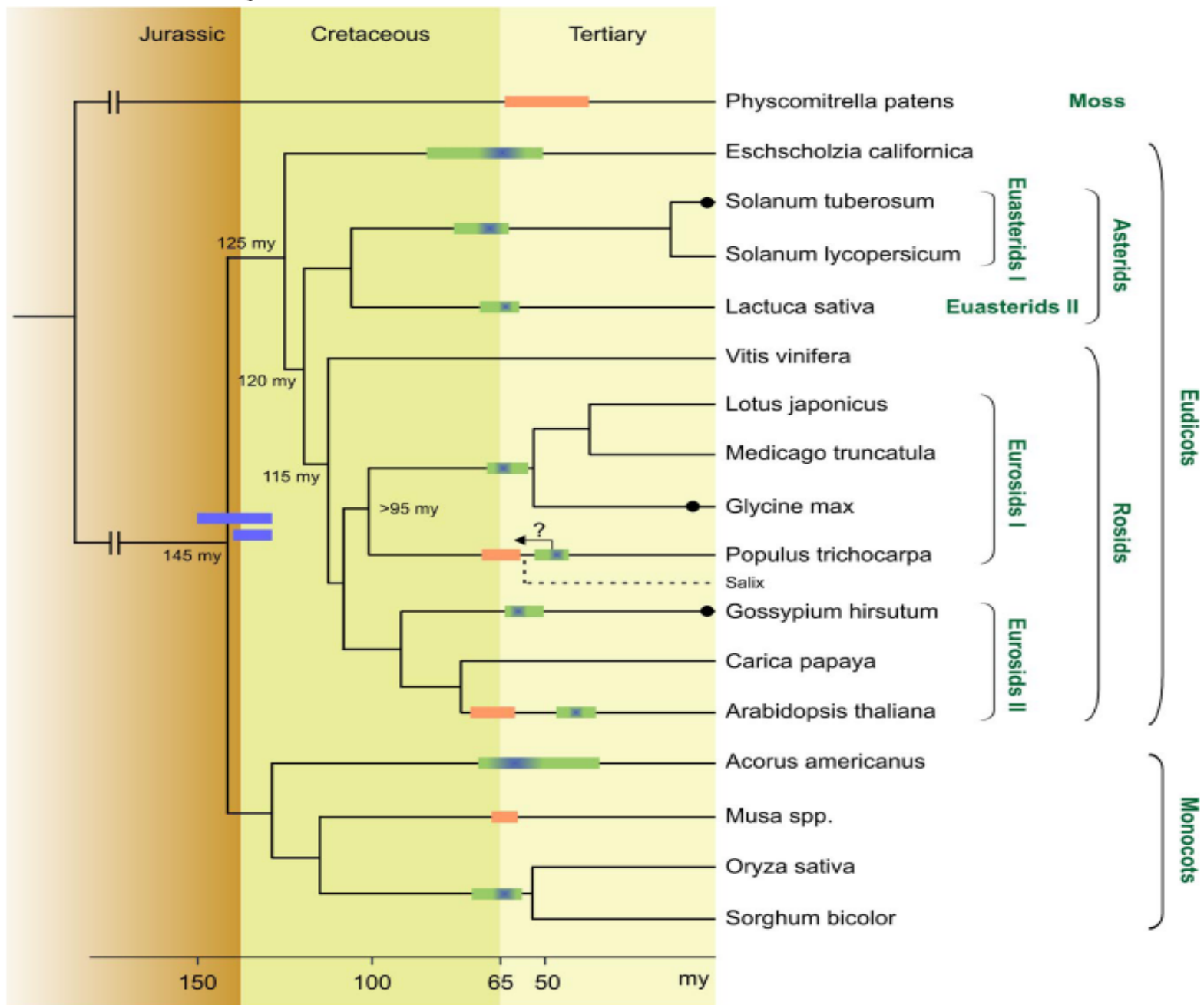
In studi condotti sulla flora artica, si è osservato che la frazione di specie vegetali diploidi e poliploidi aumenta all'aumentare della latitudine.

Plants with double genomes might have had a better chance to survive the Cretaceous–Tertiary extinction event

Jeffrey A. Fawcett^{a,b,1}, Steven Maere^{a,b,1}, or

^aDepartment of Plant Systems Biology, Menden Institute Ghent University, 9002 Ghent, Belgium

Communicated by Marc C. E. Van Montagu, Ghent Univ



Poliploidia nel frumento

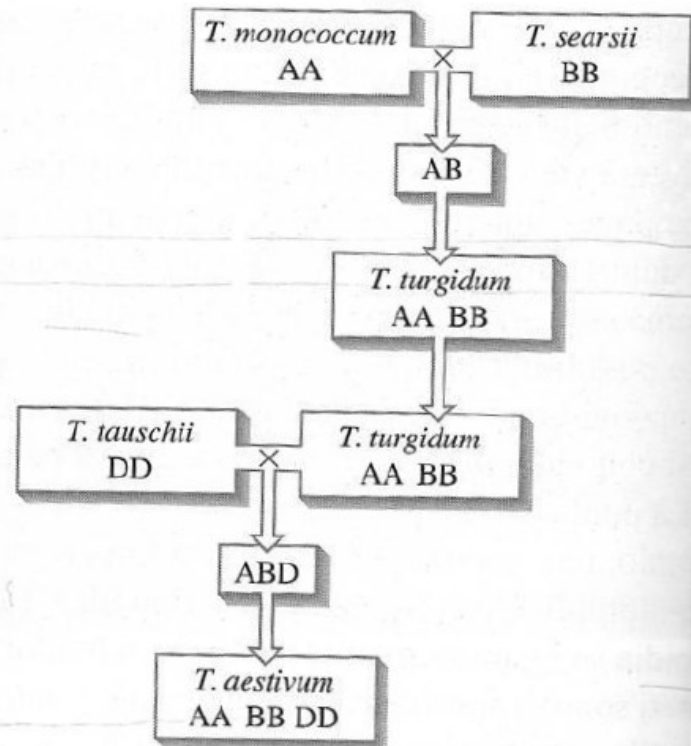
<i>Triticum monococcum</i>	farro piccolo, farro monococco, einkorn	AA
<i>Triticum dicoccum</i>	farro medio, farro dicocco	AABB
<i>Triticum turgidum</i>	grano duro, frumento duro	AABB
<i>Triticum spelta</i>	farro grande, granfarro, spelta	AABBDD
<i>Triticum aestivum</i>	grano tenero, frumento tenero	AABBDD
<i>Triticosecale</i>	triticale	AABBRR

legenda: A, genoma del frumento diploide originale o di una specie affine; B, genoma di una graminacea selvatica, *Aegilops speltoides*, o di una specie affine; D, genoma di un'altra graminacea selvatica, *T. tauschii*, o di una specie affine; R, genoma di segale (*Secale cereale*).

Come si originano le novità evolutive: duplicazioni genomiche

Figura 7.4

Il frumento, *Triticum aestivum*, è esaploiploide e deriva dalla fusione di tre specie separate: *T. monococcum*, *T. searsii* e *T. tauschii*.



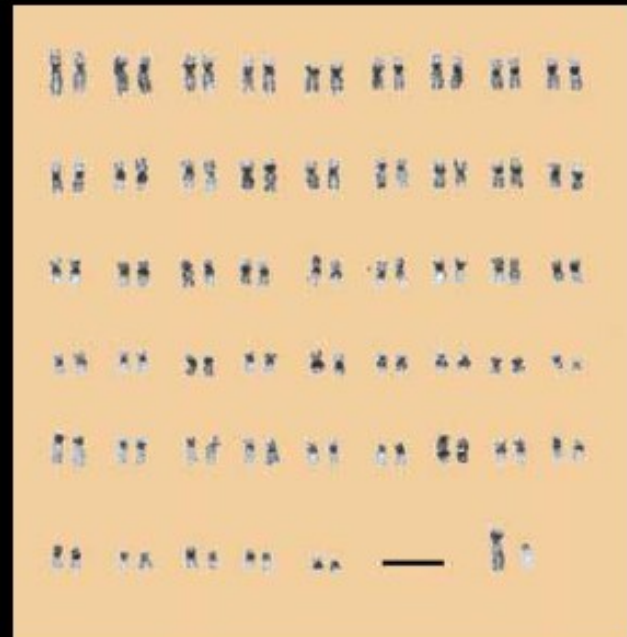
POLIPLOIDIA NEI MAMMIFERI



Tympanoctomys barrerae

Nel 1999 viene identificato un mammifero poliploide, il topo rosso di Mendoza (Argentina), $4N= 102$.

Il successo evolutivo di questa specie poliploide è legato al mantenimento di un unico set di cromosomi sessuali XY.



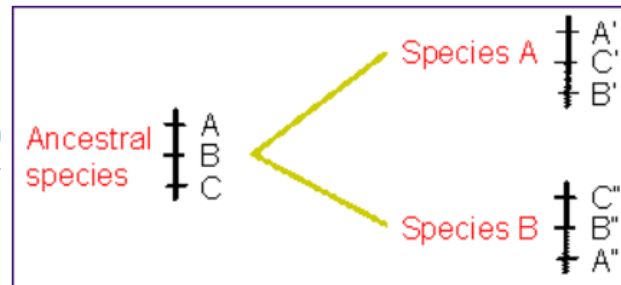
Sintenia

Blocchi di sintenia: regioni del genoma dove l'ordine dei geni viene conservato, come risultato della discendenza da un antenato comune.

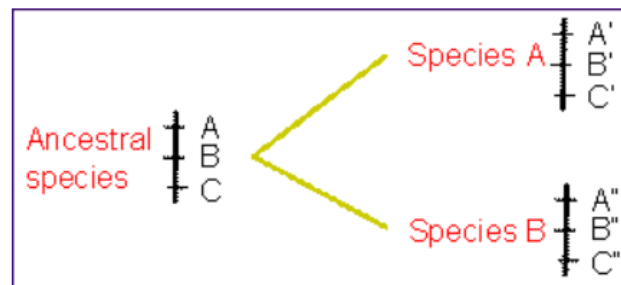
Macrosintenie posso essere evidenziate anche con bandeggi cromosomici

Definition: syntenic and collinear

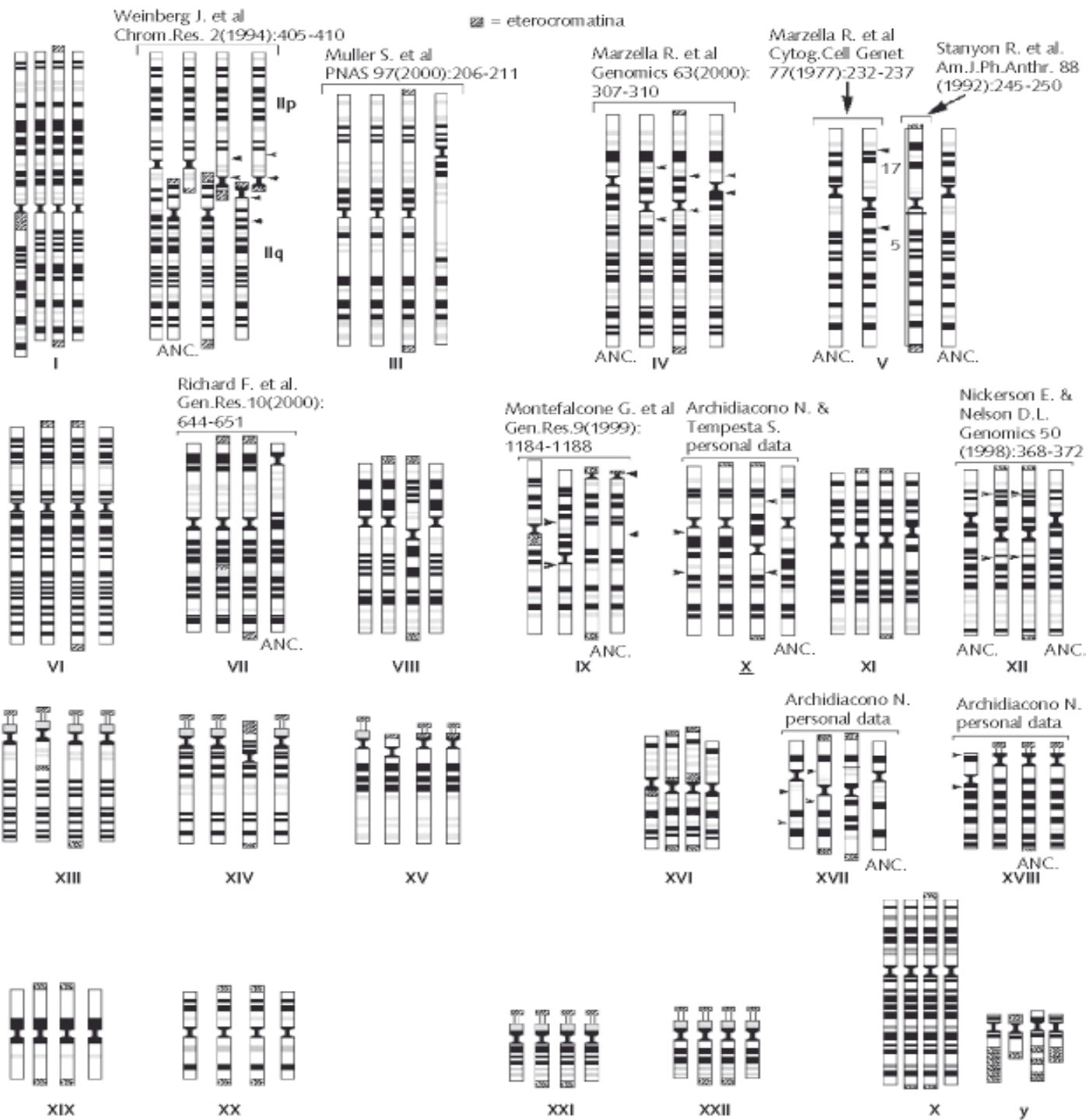
Syntenic* = a set of loci in two different species which is located on the same chromosome in each (not necessarily in the same order).



Collinear = a set of loci in two different species which are located on the same chromosome in each, and are conserved in the same order.



Next Page



Valutare l'effetto della selezione

Nell'analizzare la divergenza di geni correlati, come si riesce a distinguere l'effetto della selezione dalla deriva genetica?

Date due sequenze geniche allineate, si possono calcolare K_s , il numero di sostituzioni sinonime, e K_a , il numero di sostituzioni non sinonime.

Valutare l'effetto della selezione

Il calcolo di K_a e di K_s implica più che un semplice conteggio a causa della necessità di stimare i possibili cambiamenti multipli e di apportare le appropriate correzioni.

Dal rapporto tra i due tipi di sostituzioni possiamo identificare se esiste ed il tipo di selezione

$K_a/K_s = 1$ Evoluzione neutrale

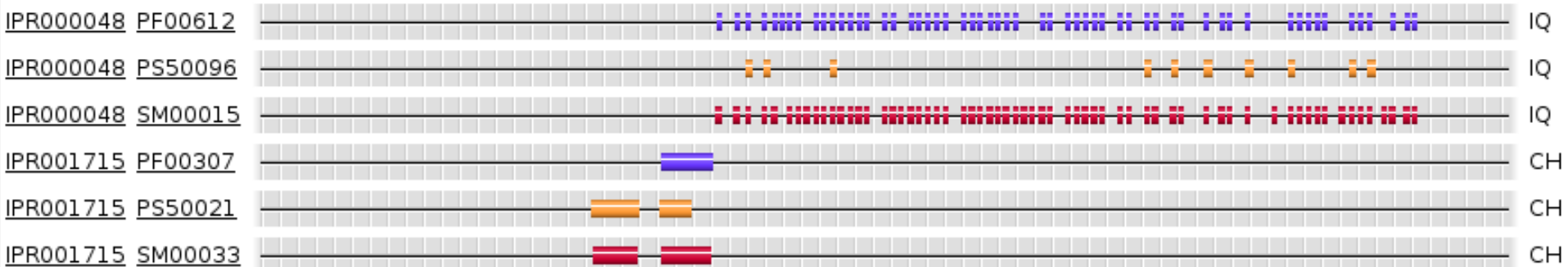
$K_a/K_s > 1$ Selezione positiva.

$K_a/K_s < 1$ Selezione purificante.

Struttura e domini di ASPM

<http://www.genecards.org/cgi-bin/carddisp.pl?gene=ASPM>

- **ASPM è una proteina di 3477 aa** e contiene:
un dominio di legame-ai-microtubuli N-terminale
2 domini CH (Calponin-homology, legame con l'actina)
39 motivi ripetuti Ile-Gln (IQ)
1 regione C-terminale con funzione ignota.



Struttura e domini di ASPM

- Inoltre, **ASPM** presenta **28 esoni**.

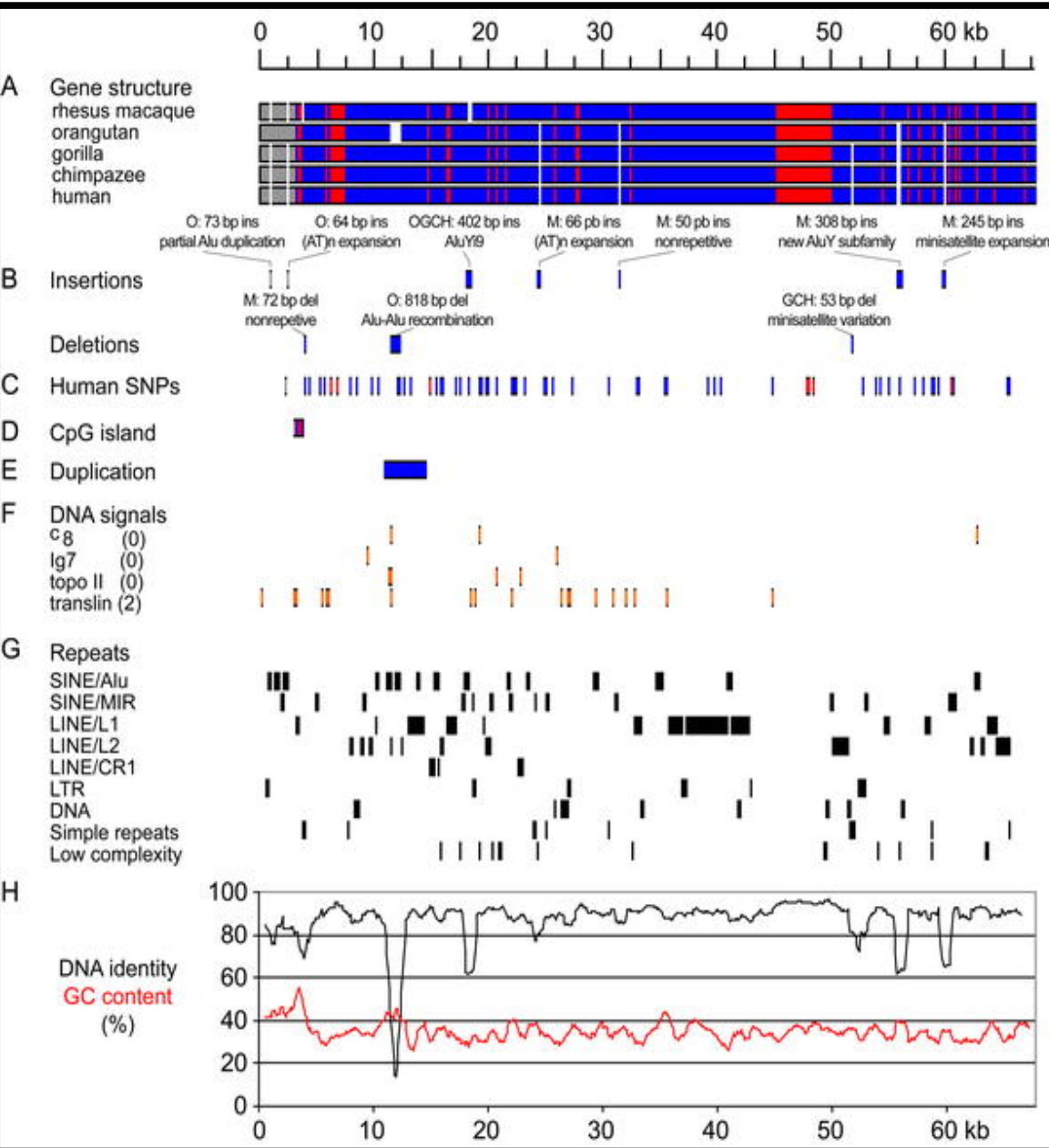
- Domanda:

l'evoluzione di ASPM è costante per tutta la proteina?

O alcuni domini sono più conservati?

-> *'Accelerated Evolution of the ASPM Gene Controlling Brain Size Begins Prior to Human Expansion'*,
Kouprina et al.

Evoluzione di ASPM



- Nella figura è mostrato l'allineamento tra le copie di ASPM di uomo, macaco, orango, gorilla, scimpanzè.
- Le delezioni più grandi avvengono nelle zone non codificanti (in blu), mentre il resto è conservato.

Motif	Percentage
Putative microtubule	~10%
Calponin	~10%
Repeated IG domains	~10%
C-terminal	~10%



In conclusione:



- ASPM è un gene coinvolto nella mitosi dei neuroni del cervello, che ha subito forte selezione positiva nell'uomo, specie nelle **regioni a IQ e nell'esone 18**.
- La sua selezione comincia addirittura prima della separazione tra uomo e scimmie.
- Studi recenti hanno identificato **una variante positiva (A44871G)** evolutasi 5800 +/- 5000 anni fa.

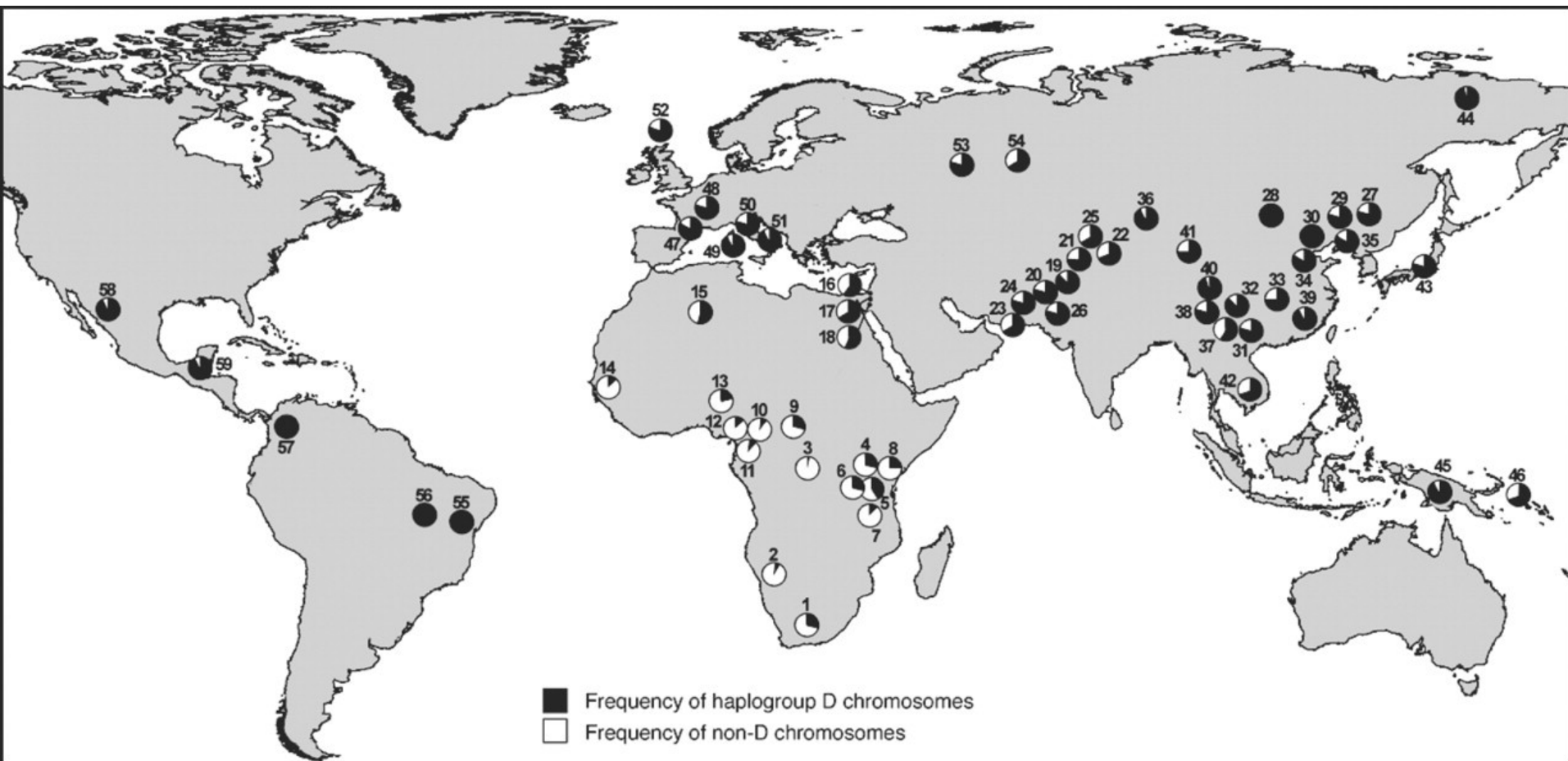
SUMMARY

- *Microcephalin* and *ASPM* are genes involved in regulating brain size
- A variant of *Microcephalin* has reached very high frequencies in non-Sub-Saharan Africans in the last 37,000 years.
- A variant of *ASPM* has reached very high frequencies especially in Caucasoids but also in some southern Mongoloids and Australoids in the last 5,800 years.

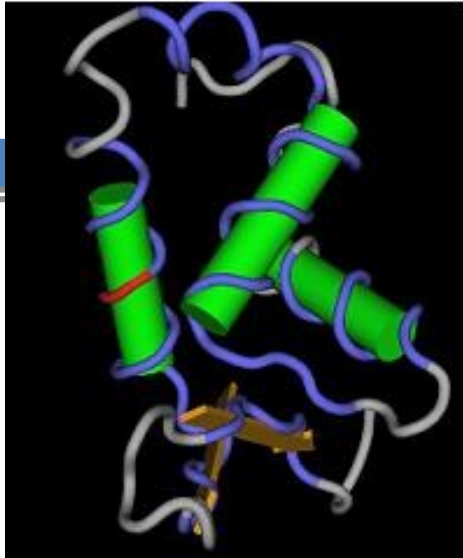
It is almost inconceivable that these two factors were caused by random factors (drift).

Therefore *selection* has acted on these two genes, favoring the new *Microcephalin* variant in non-Sub-Saharan Africans and the *ASPM* especially in Caucasoids, but also to a lesser extent in some southern Mongoloid and Australoid groups.

We know absolutely nothing about what the new *Microcephalin* and *ASPM* variants actually do. What we *do* know is that they confer some substantial advantage that has caused them to grow in numbers. Perhaps, they confer some cognitive or behavioral ability.

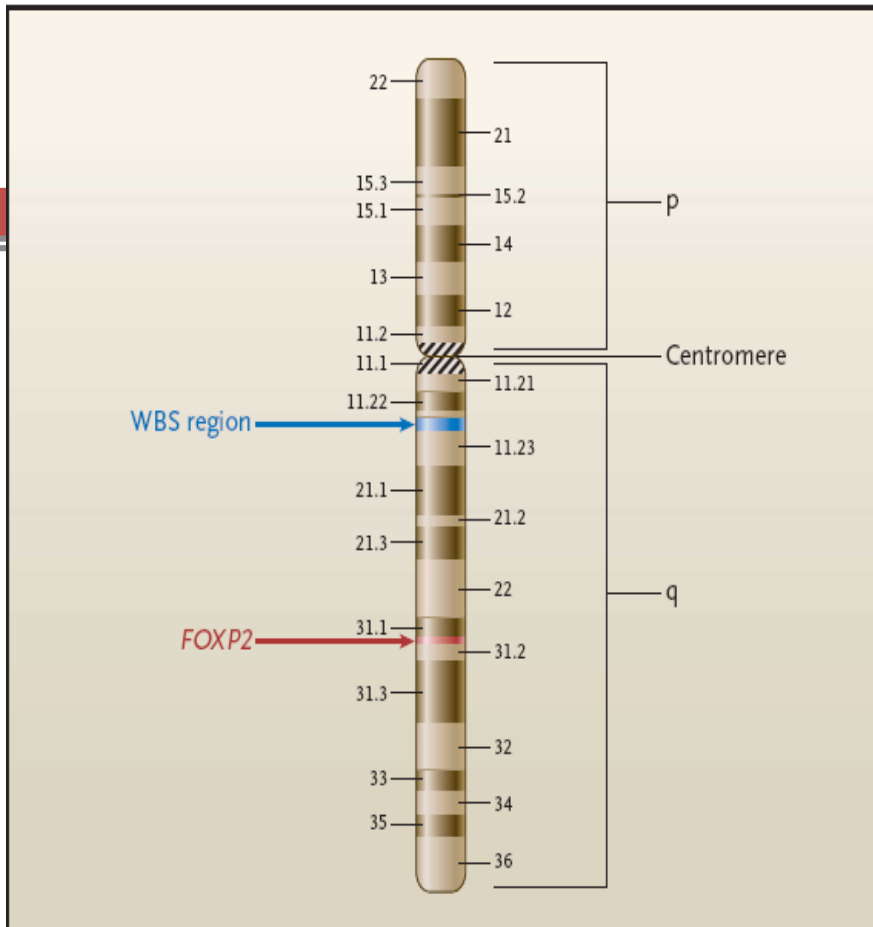


La Proteina FOXP2



In rosso è mostrata la
posizione della
mutazione dell' aa 325

- Il gene Foxp2 (cromosoma 7) ✉ Proteina Foxp2 :
- ❖ costituita da 715 aa
 - ❖ fattore di trascrizione
 - ❖ contiene una regione ricca di poli-glutammine
 - ❖ è tra il 5% delle proteine più conservate ✉
mostra particolari similitudini in tutti i primati

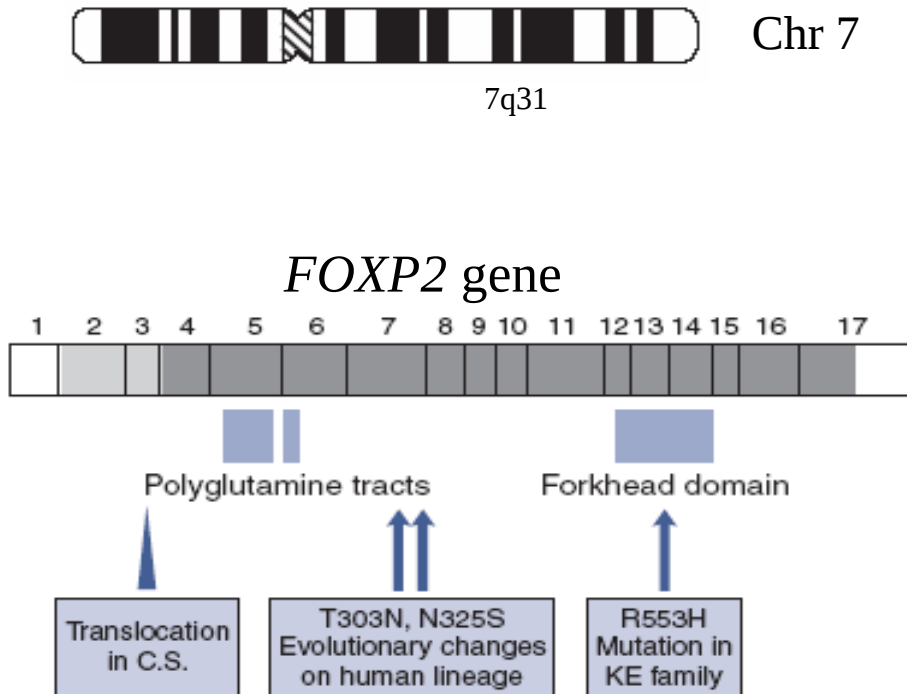


Chromosome 7 Gene Dosage and Speech and Language Disorder.

An ideogram of chromosome 7 shows cytogenetic banding (with two regions of the long q arm highlighted), a 600-kb region containing the *FOXP2* gene, and a 1.5-Mb interval that is commonly deleted in the Williams–Beuren syndrome (WBS). Point mutations and chromosomal abnormalities involving *FOXP2* are one cause of developmental verbal dyspraxia. Duplication of the WBS region may also be associated with a delay in speech.

La famiglia genica **Forkhead-box (FOX)** nell'uomo consiste di almeno 43 membri, suddivisi in 17 sottofamiglie (FOX A- FOX Q). Oltre all'uomo molteplici altre specie risultano possedere nel loro corredo geni della famiglia Fox. Tali geni risultano essere caratterizzati dal dominio Fox, ossia un dominio di binding al DNA monomerico di 80-100 aminoacidi.

Data la capacità di legare il DNA, le proteine Fox si comportano da fattori di trascrizione e risultano essere direttamente coinvolte nella regolazione dell'espressione di geni target correlati per esempio allo sviluppo embrionale e al controllo metabolico.



Nello specifico il dominio di binding al DNA di Foxp2 è mutato negli individui affetti da deficit nella comunicazione orale.

Il gene Foxp2 è localizzato nel braccio lungo del cromosoma 7 (7q31)

Il Gene FOXP2

Si è osservato che **due diverse mutazioni** in questo gene sono associate a disordini nello sviluppo del linguaggio:

R553H (KE Family)

Mutazione missense che porta ad una condizione che colpisce il sistema del linguaggio, della scrittura e della comunicazione. Gli individui affetti hanno problemi nella pronuncia delle parole, non hanno il controllo della grammatica, non riescono ad avere particolari movimenti delle labbra e della lingua, e non capiscono bene la scrittura ✉ malattia ad eredità autosomica dominante.

R328X

Il gene è interrotto da una traslocazione che coinvolge i cromosomi 5 e 7, che porta a disordini del linguaggio simili a quelli sopra elencati. (No relazioni familiari)

Allineamento di sequenze aminoacidiche di Foxp2

Human	MMQESATETI	SNSSMNQNGM	STLSSQLDAG	SRDGRSSGDT	SSEVSTVELL	HLQQQALQA	ARQLLLQQQT	SGLKSPKSSD	KQRPLQVPVS	VAMTIPQVIT
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	PQQMQQILQQ	QVLSPQQQLQA	LLQQQQAVML	QQQQQLQEFYK	KQDEQLHLQL	QQQQQQQQQQ	QQQQQQQQQQ	QQQQ-QQQQQ	QQQQQQQQQQ	QHHPGKQAKE
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	QQQQQQQQQQ	LAAQQLVFQQ	QLLQMQLLQQ	QQHLLSLQRQ	GLISIPFGQA	ALFVQSLPQA	GLSPAEIQQL	WKEVTGVHSM	EDNGIKHGEL	DLTTNNSST
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	TSSNTSKASP	PITHHSIVNG	QSSVLSARRD	SSSHEETGAS	HTLYGHGVCK	WPGCESICED	FGQFLNHLNN	EHALDDNSTA	QCRVQMQVVQ	QLEIQLSKER
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	ERLQAMMTHL	HMRPSEPSPS	PKPLNLVSSV	TMSKNMLETS	PQSLPQTPTT	PTAPVTPITQ	GPSVITPASV	PNVGAIARRH	SDKYNIPMSS	EIAPNYEFYK
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	NADVRRPFTY	ATLIRQAIME	SSDRQLTINE	IYSWFTRTFA	YFRRNAATWK	NAVRHNLSLH	KCFVRVENVK	GAYMTVDEVE	YQKRRSQKIT	GSPTLVKNIP
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

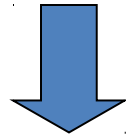
Human	TSLGYGAALN	ASLQAALAES	SLPLLSNPGL	INNASSGLLQ	AVHEDLNGSL	DHIDSNNGSS	PGCSPQPHIH	SIHVKEEPVI	AEDEDCPMSL	VTTANHSPEL
Chimp										
Gorilla										
Orang										
Rhesus										
Mouse										

Human	EDDREIEEEP	LSEDLE*
Chimp		
Gorilla		
Orang		
Rhesus		
Mouse		

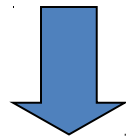
2 mutazioni uomo-specifiche :

□ In posizione 303 (da treonina ad asparagina)

□ In posizione 325 (da asparagina a serina)



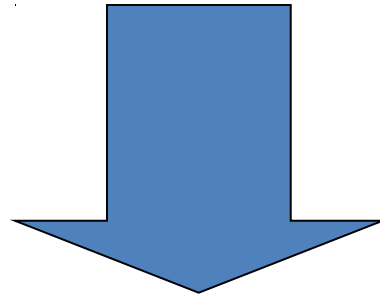
sito per la fosforilazione ➡ cambio nella struttura secondaria predetta. La fosforilazione di questo FT può essere un importante meccanismo che media la regolazione trascrizionale



Nuova sequenza genica ➡ aumento n° proteine
➡ perfezionamento bocca e laringe ➡ articolazione di suoni complessi

Per studiare se i 2 cambiamenti aminoacidici codificati dall'esone 7 sono polimorfici negli umani, è stato sequenziato questo esone in 44 cromosomi umani presi da distinte popolazioni di tutto il mondo

In nessuno di questi casi si osservò la presenza di polimorfismi e/o sostituzioni aminoacidiche ✉ nessuna variazione genetica apparsa fra uomini di diversa origine etnica



FISSAZIONE

(120.000-200.000 ANNI FA)

Non solo gli uomini, ma anche gli scimpanzè, altri primati e perfino i topi dispongono del gene Foxp2.

Qual è allora il particolare che ci differenzia dalle altre specie e ci permette di parlare?

Il gene del linguaggio (Svante Paabo et al.) ha subito una mutazione che si è fissata nella nostra specie circa 120-200 mila anni fa.

Grazie alle proteine prodotte in più dalla nuova sequenza genica, bocca e laringe si sono perfezionate tanto da permettere l'articolazione di suoni complessi.

Dal confronto tra i geni Foxp2 dello scimpanzè, del gorilla, dell'orango, del macaco rhesus e del topo, tenuto conto delle varie differenze, è stata dedotta l'evoluzione genetica delle specie.

Subito dietro all'uomo nella scala evoluzionistica, ci sono scimpanzè (capaci secondo molti scienziati di comunicare a gesti), gorilla e macaco. Più distaccati orango e topi.

Nessuna variazione genetica è invece apparsa fra uomini di diversa origine etnica.

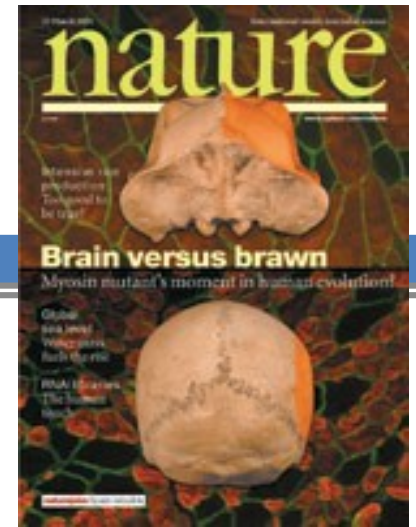
Le persone con una sola copia di questo gene funzionante hanno problemi non solo nell'articolare il linguaggio, ma anche nel comprenderlo, nel seguire le regole grammaticali e nel muovere i muscoli del viso.

La miosina

Stedman et al. (2004) the loss of the sarcomeric **MYOSIN gene MYH16** in the human lineage lead to smaller masticatory muscles. They estimated that mutation that lead to the inactivation (**a two base pair deletion**) occurred **2.4 MYA** right before Homo ergaster/erectus showed up in Africa.

This period that followed was marked by **a strong increase in cranial capacity**.

The authors put up the hypothesis that the loss of that gene removed an evolutionary constrain on brain size in the genus homo.



Pattern di gracializzazione dell'apparato masticatorio associato ad un simultaneo processo di encefalizzazione accelerata

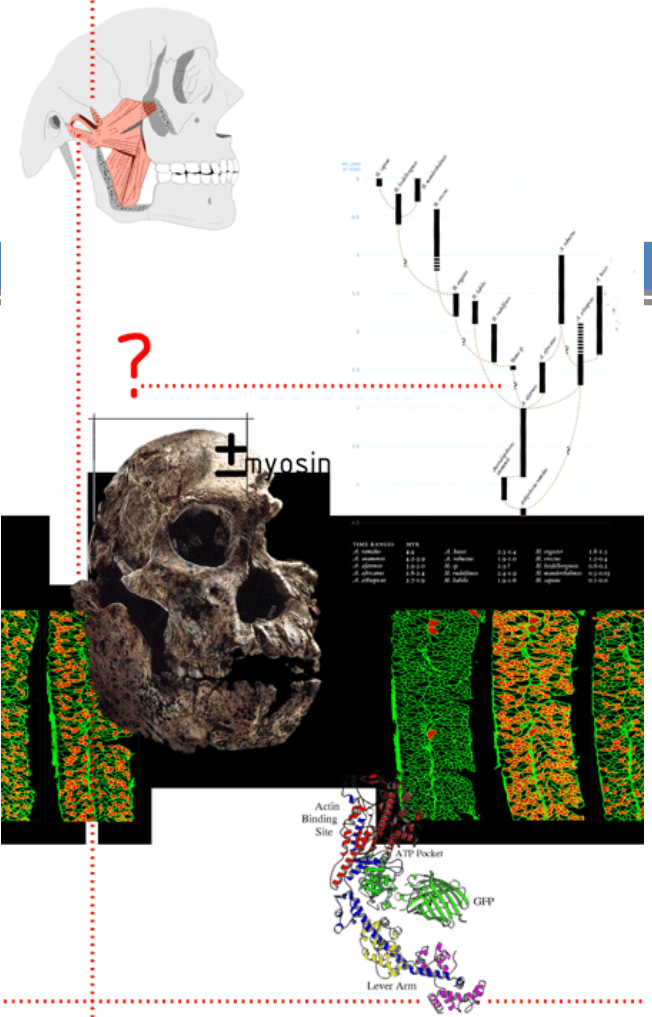
Mutazione porta a riduzione della taglia delle fibre muscolari e degli interi muscoli masticatori.

MYH16 evoluzione sotto pressione di selezione negativa in tutte le linee ancestrali eccetto che in quella che ha portato direttamente a *H. sapiens*

However Perry et al. (2005) estimated that the MYH16 gene has been lost 5.3 MYA, long before the genus homo appeared. They also find conflicting estimates of nonsynonymous fixation rates across different regions of this gene, revealing a complex pattern inconsistent with a simple model of pseudogene evolution for human MYH16.



Figure 1 News of chews — the jaw muscles of apes, such as this mountain gorilla, and humans could reflect a profound evolutionary divergence.



Pattern di crescita dello scheletro craniofaciale può essere radicalmente alterato cambiando l'anatomia dei muscoli. La riduzione della taglia dei muscoli masticatori e della forza di contrazione come risultato della mutazione MYH16 negli antenati potrebbe aver avuto un considerevole impatto sulla morfologia del cranio. Questi effetti potrebbero aver incluso una riduzione di stress attorno alle ossa della scatola cranica, permettendo loro di diventare più larghe.