

# Evoluzione del genoma

Silvia Fuselli, [fss@unife.it](mailto:fss@unife.it)

29 novembre 2011

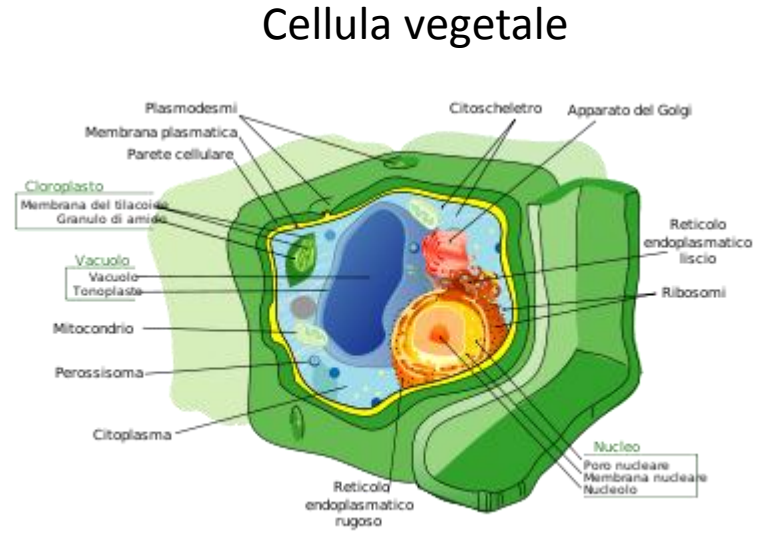
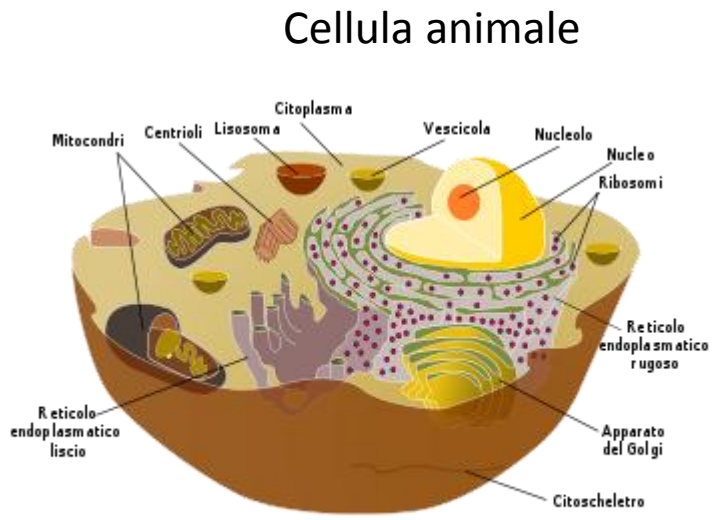
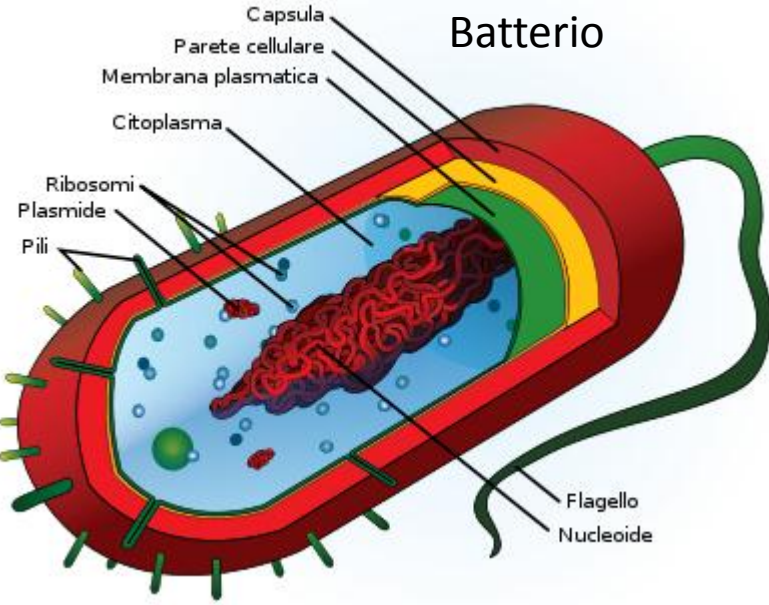
# In questa lezione parleremo di

- Meccanismi di evoluzione del genoma
- Formazione di nuovi geni
- Dimensioni del genoma e complessità degli organismi

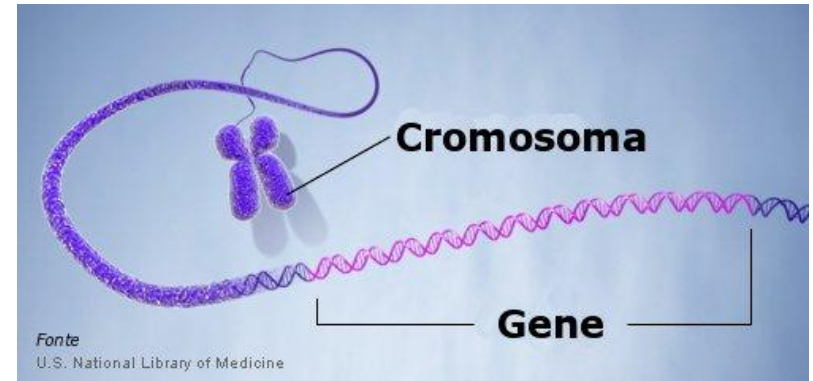
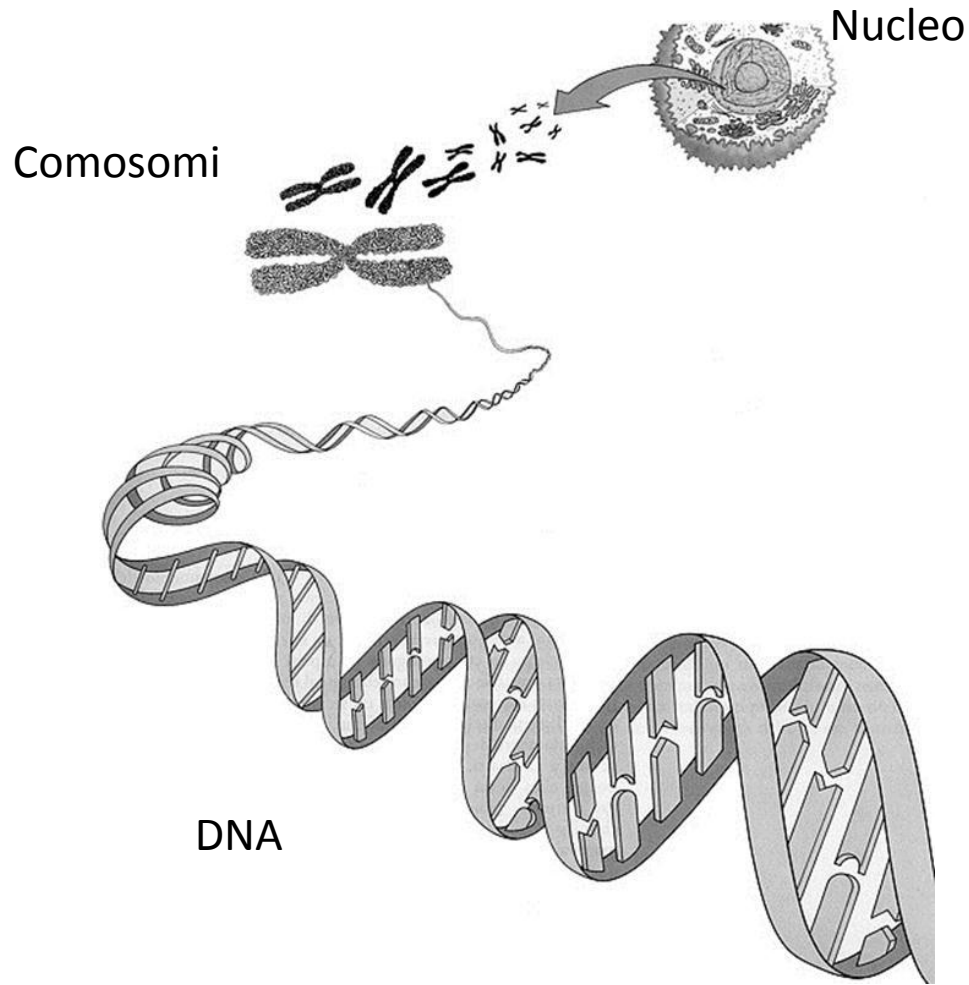
Bibliografia:

Evoluzione. Modelli e processi. A cura di Ferraguti e Catellacci. Pearson.  
Capitolo 7.

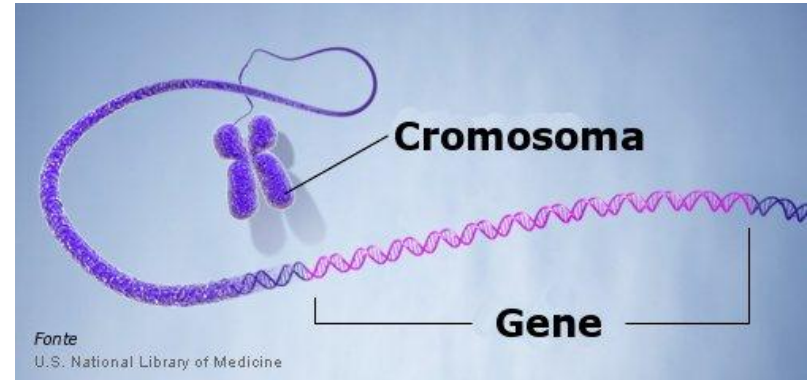
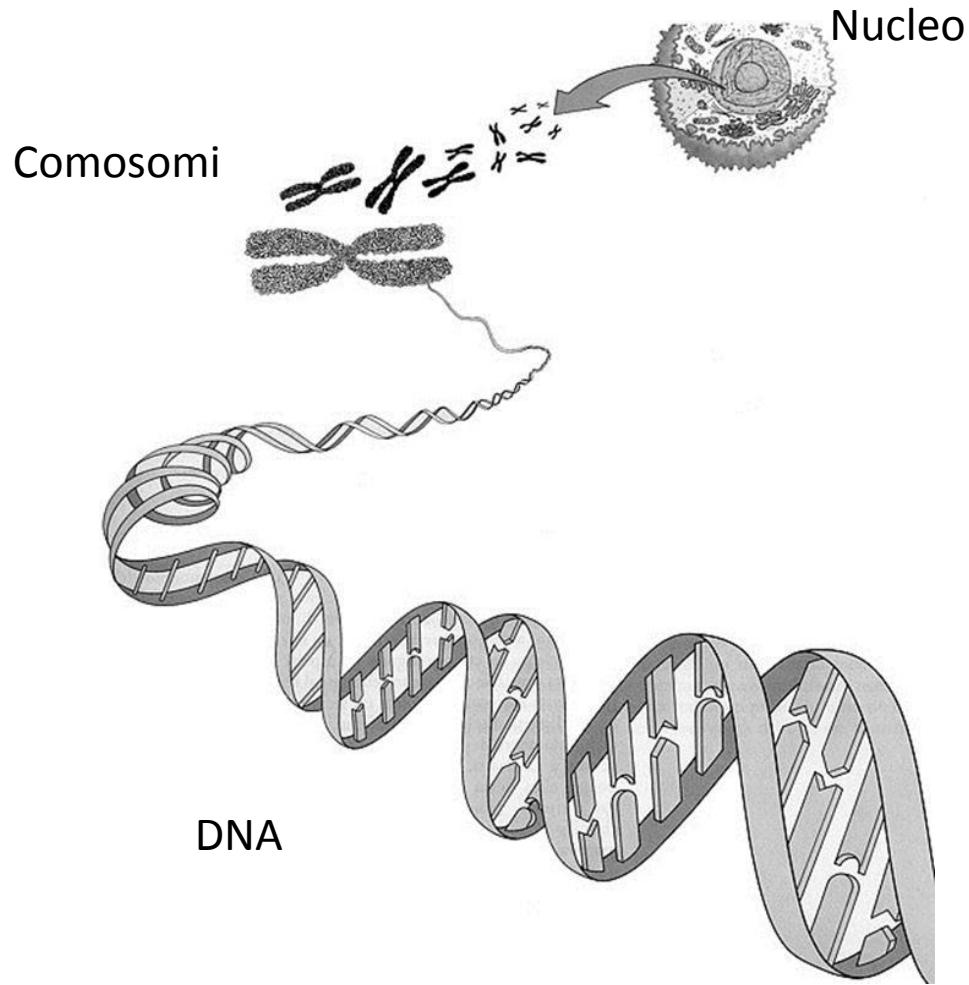
# Un particolare tipo di 'ecosistema': la cellula



# Il materiale genetico



# Il materiale genetico



# I 3 paradossi del genoma

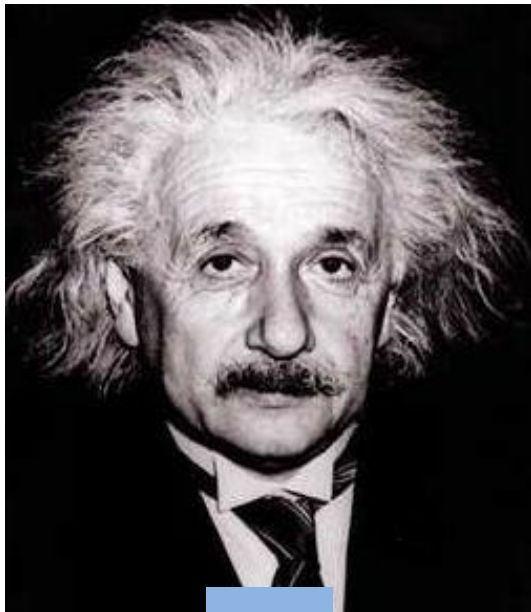
**K**

**N**

**C**

# Paradosso del valore **K**: la complessità degli organismi non correla con il numero di cromosomi.

*Homo sapiens*



46

*Lysandra atlantica*



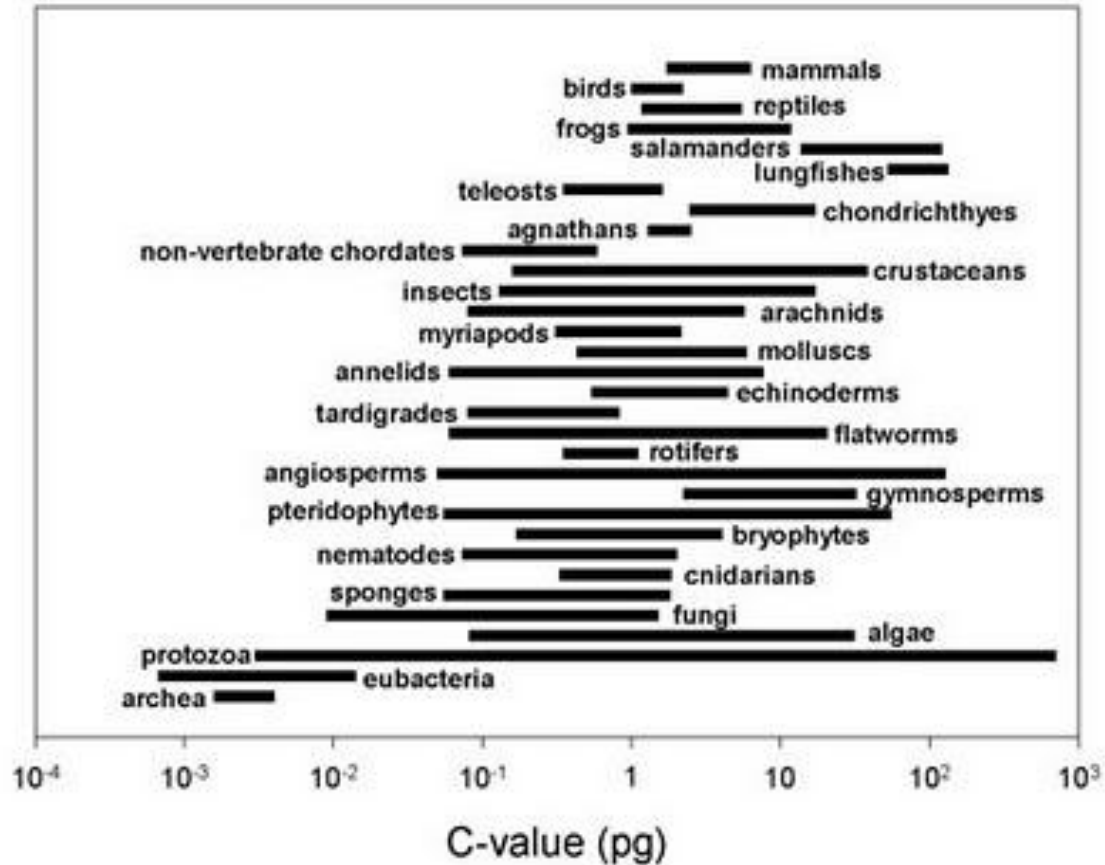
250

*Ophioglossum reticulatum*



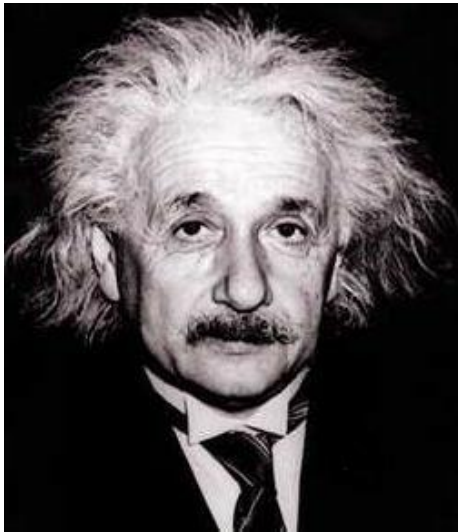
~1260

# Il valore C (quantità di DNA contenuto nel nucleo di una cellula aploide di un organismo)





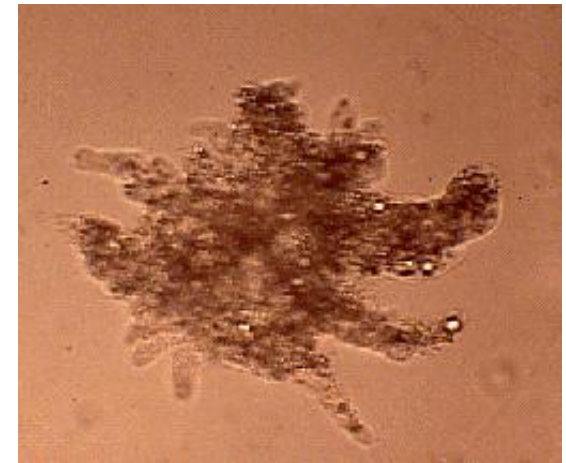
# Paradosso del valore C: la complessità non correla con la grandezza del genoma.



$3.4 \times 10^9$  bp  
*Homo sapiens*



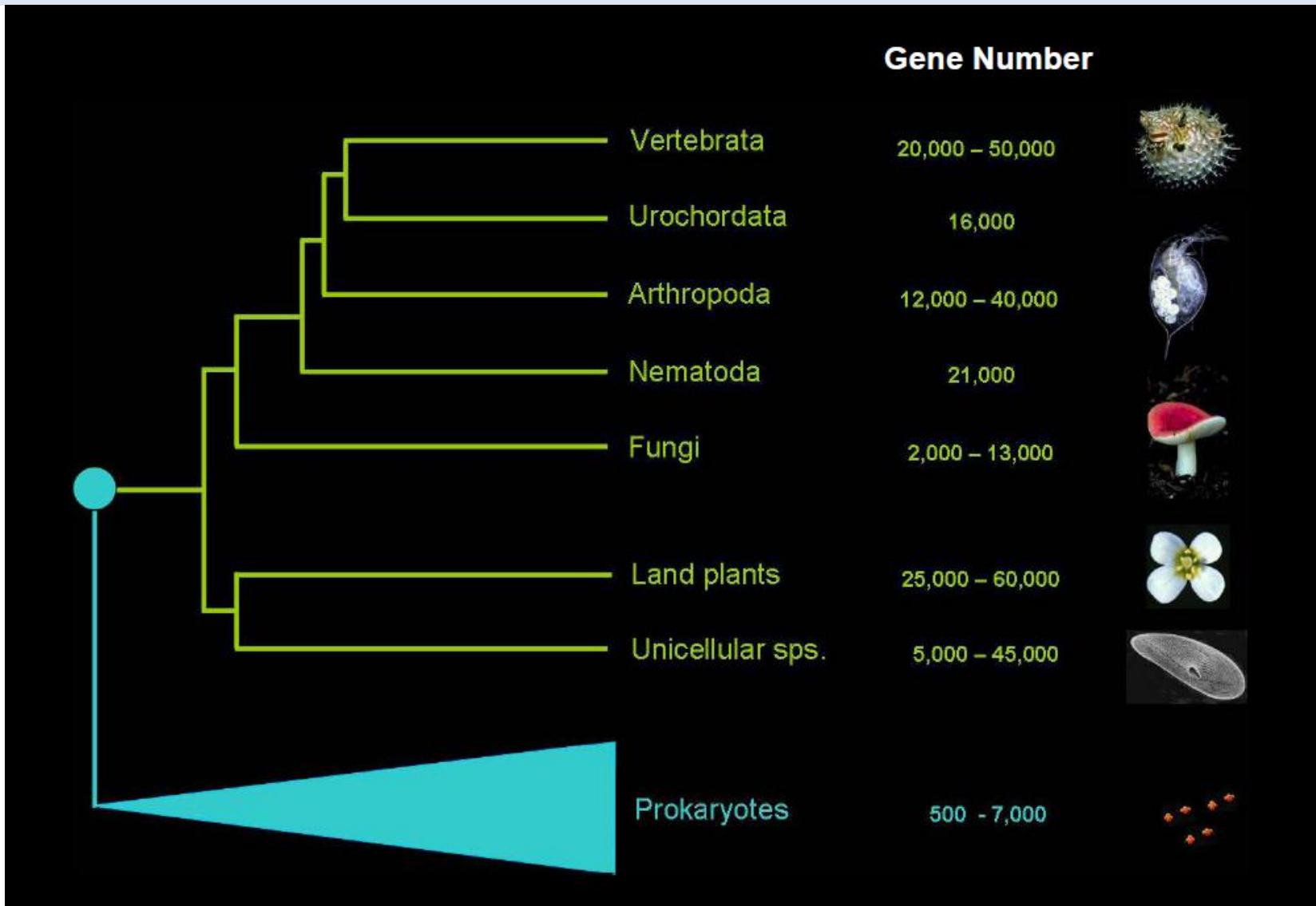
$1.5 \times 10^{10}$  bp  
*Allium cepa*



$6.8 \times 10^{11}$  bp  
*Amoeba dubia*

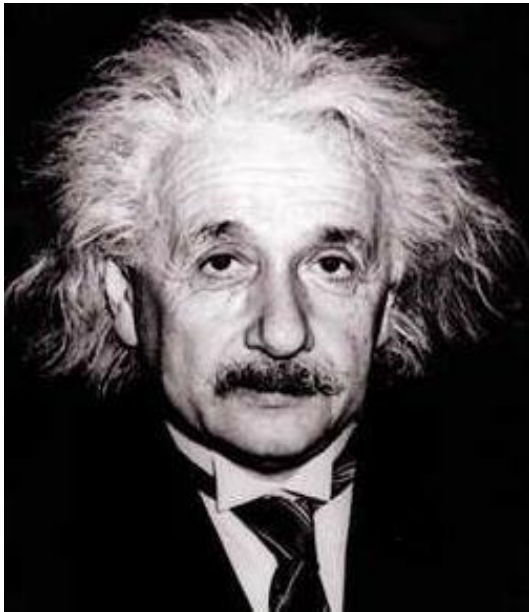
# Paradosso del valore N:

Il **numero di geni** e la complessità degli organismi non sono correlati



# Paradosso del valore N:

Il **numero di geni** e la complessità degli organismi non sono correlati



~24000 geni



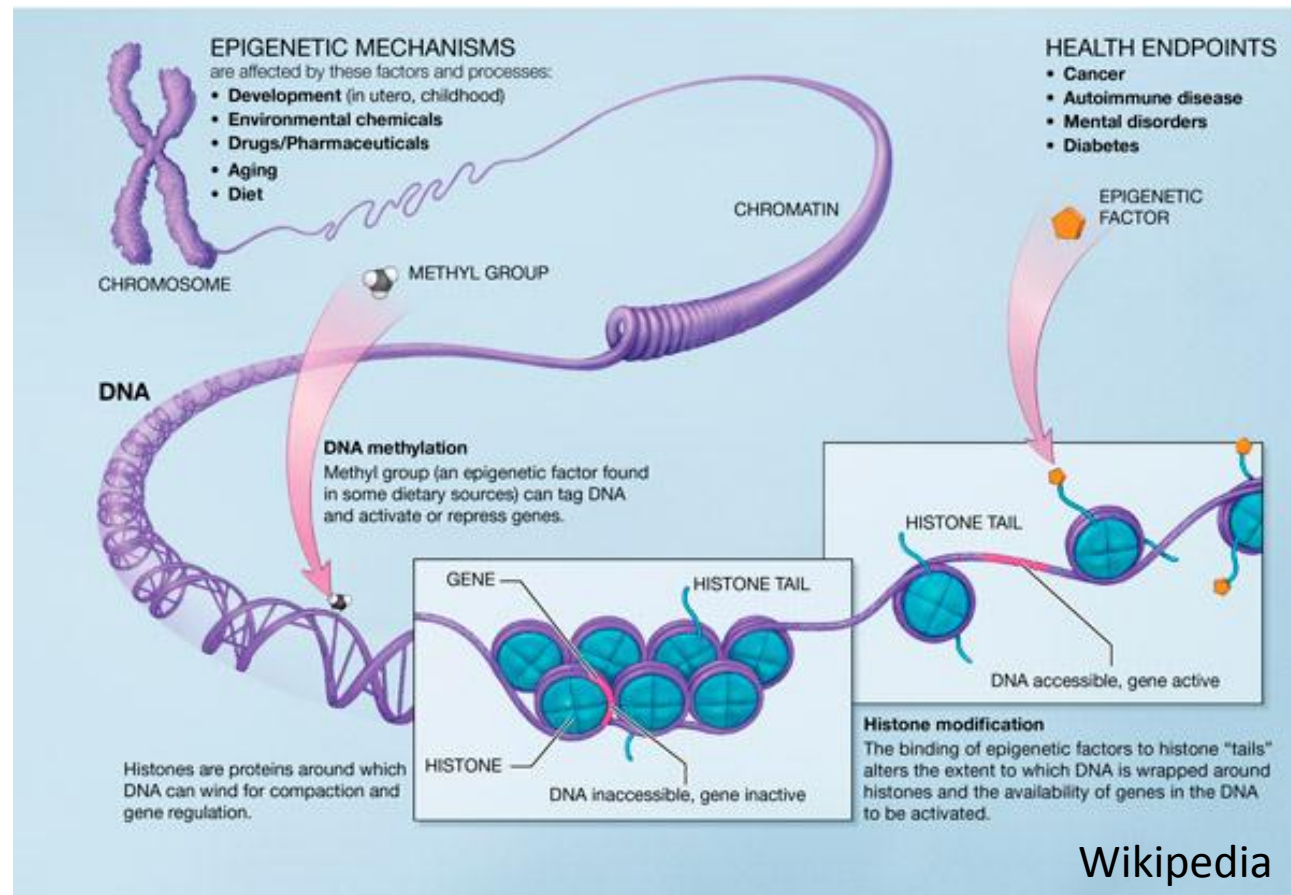
~25000 geni



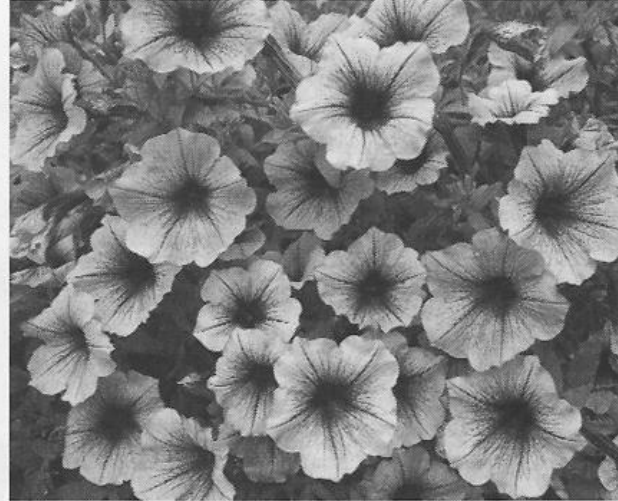
~60000 geni

# La complessità del genoma correla con la complessità degli organismi

- Organizzazione in cromosomi
- Organizzazione in geni strutturali e regolatori
- Regolazione genica
- Fattori epigenetici

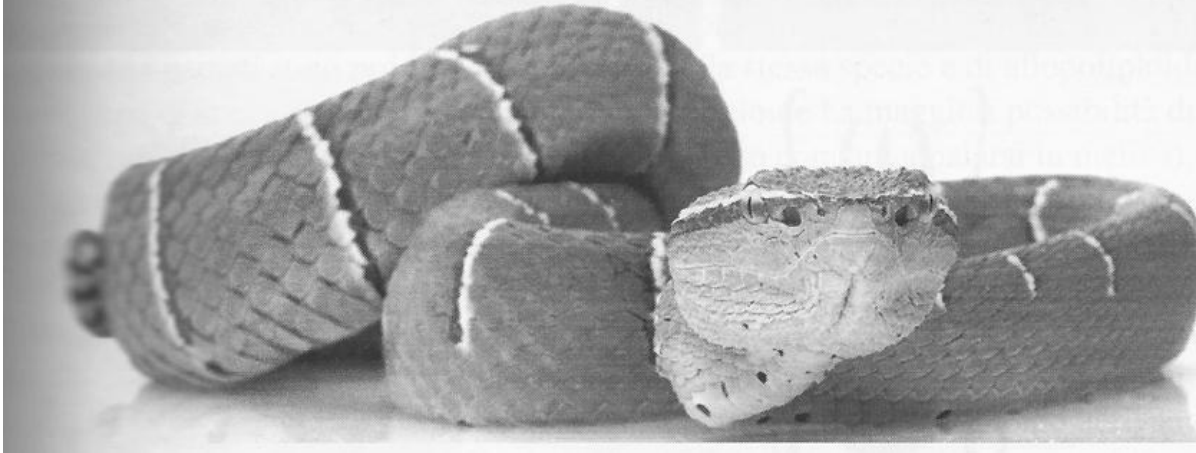


# Novità evolutive



**Figura 7.1**

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



# 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,<sup>\*,†,1</sup> Ruoping Zhao,<sup>\*,1</sup> Huifeng Jiang<sup>\*,†</sup> and Wen Wang<sup>\*,2</sup>**

#### 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*

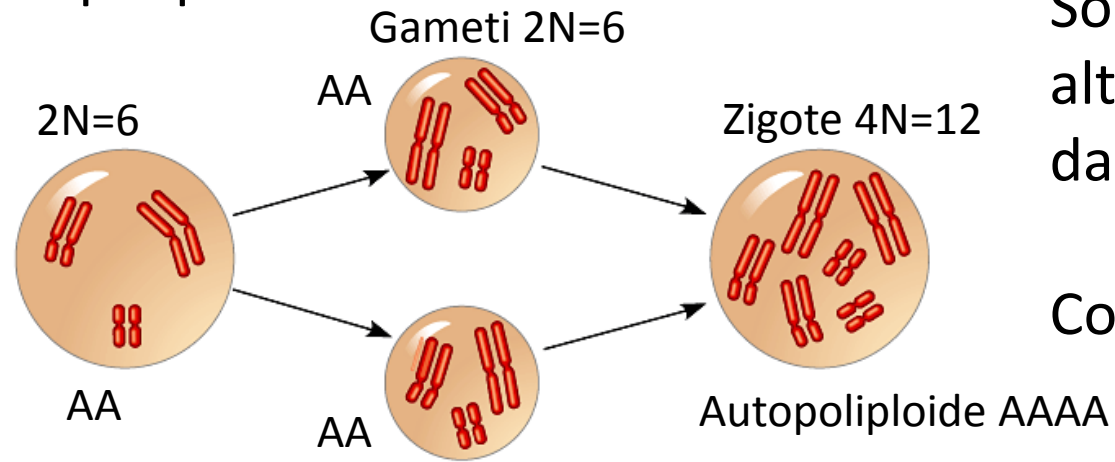
Qi Zhou,<sup>1,2,4</sup> Guojie Zhang,<sup>1,2,3,4</sup> Yue Zhang,<sup>1,4</sup> Shiyu Xu,<sup>1</sup> Ruoping Zhao,<sup>1</sup>  
Zubing Zhan,<sup>1,2</sup> Xin Li,<sup>1,2</sup> Yun Ding,<sup>1,2</sup> Shuang Yang,<sup>1,3</sup> and Wen Wang<sup>1,5</sup>

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.



# Come si originano le novità evolutive: duplicazioni genomiche

## Autopoliploidia

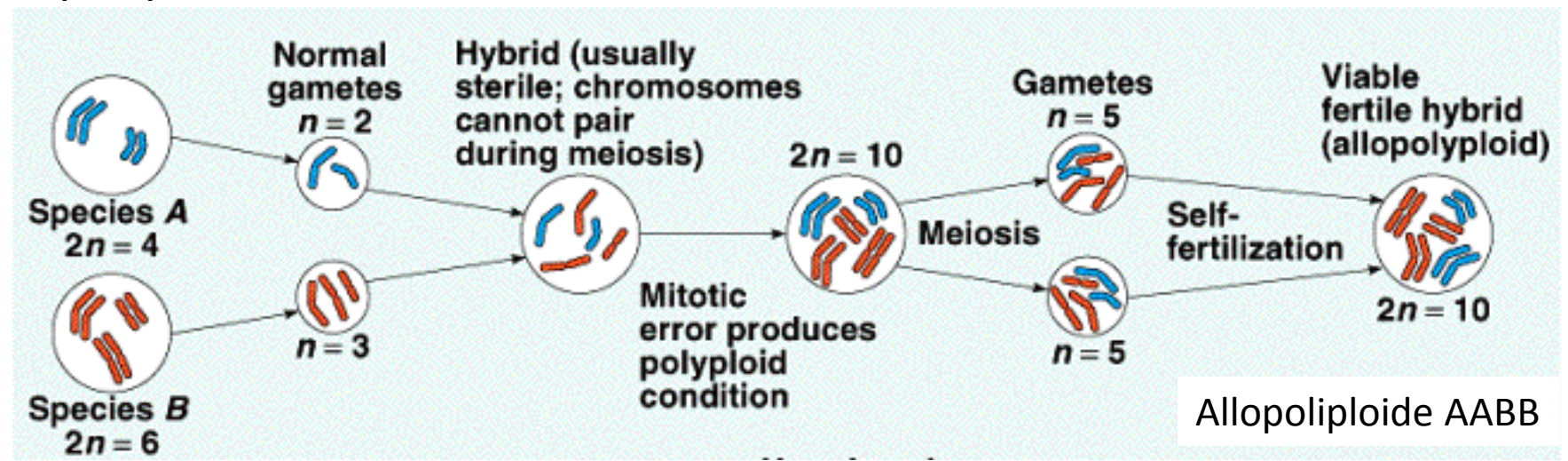


Copyright © Pearson Education, Inc., publishing as Benjamin Cummings.

Solo l'incrocio con un altro autopoliploide darà prole fertile

Comune fra le piante

## Allopoliploidia

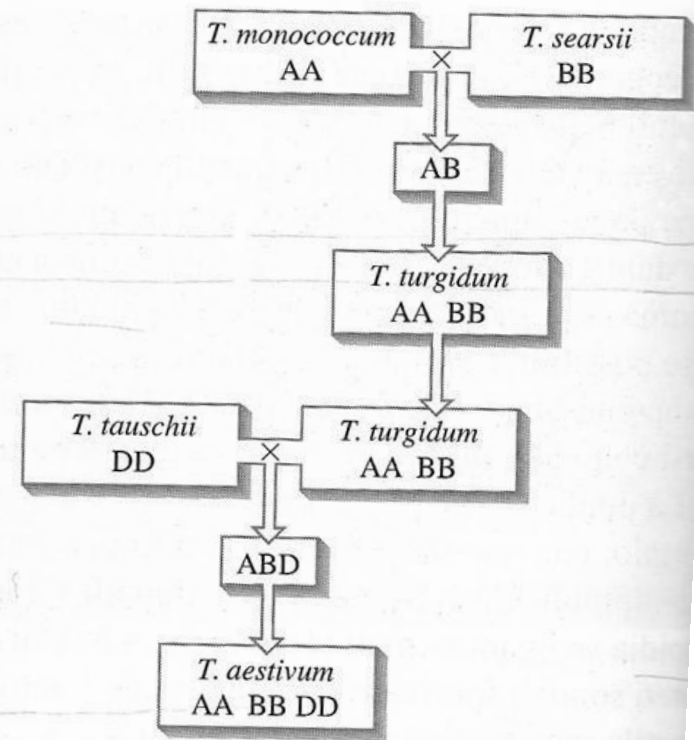


Allopoliploide AABB

# Come si originano le novità evolutive: duplicazioni genomiche

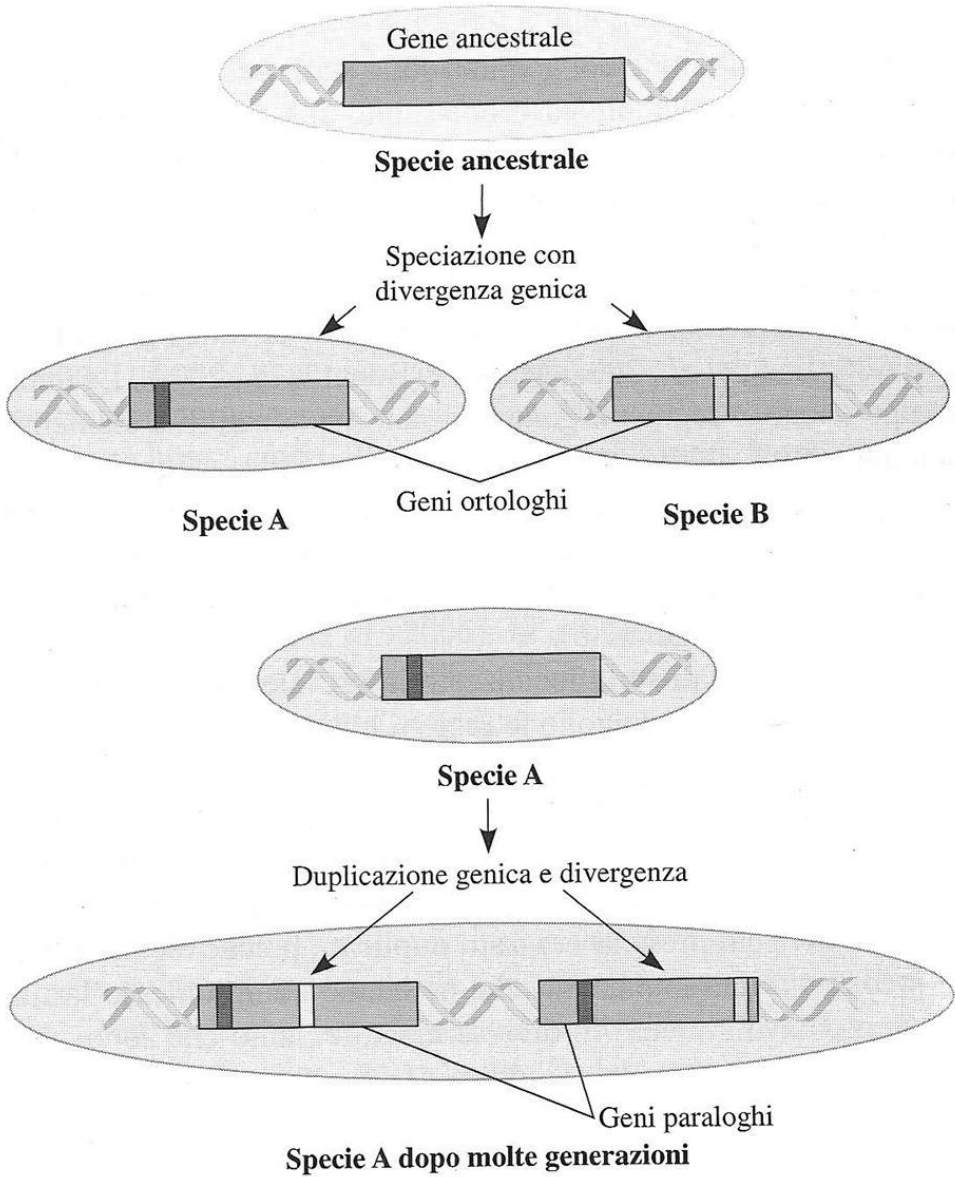
## Figura 7.4

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



# GENI OMOLOGHI

↓                      ↓  
**Ortologi**            **Paraloghi**



**Figura 7.7**

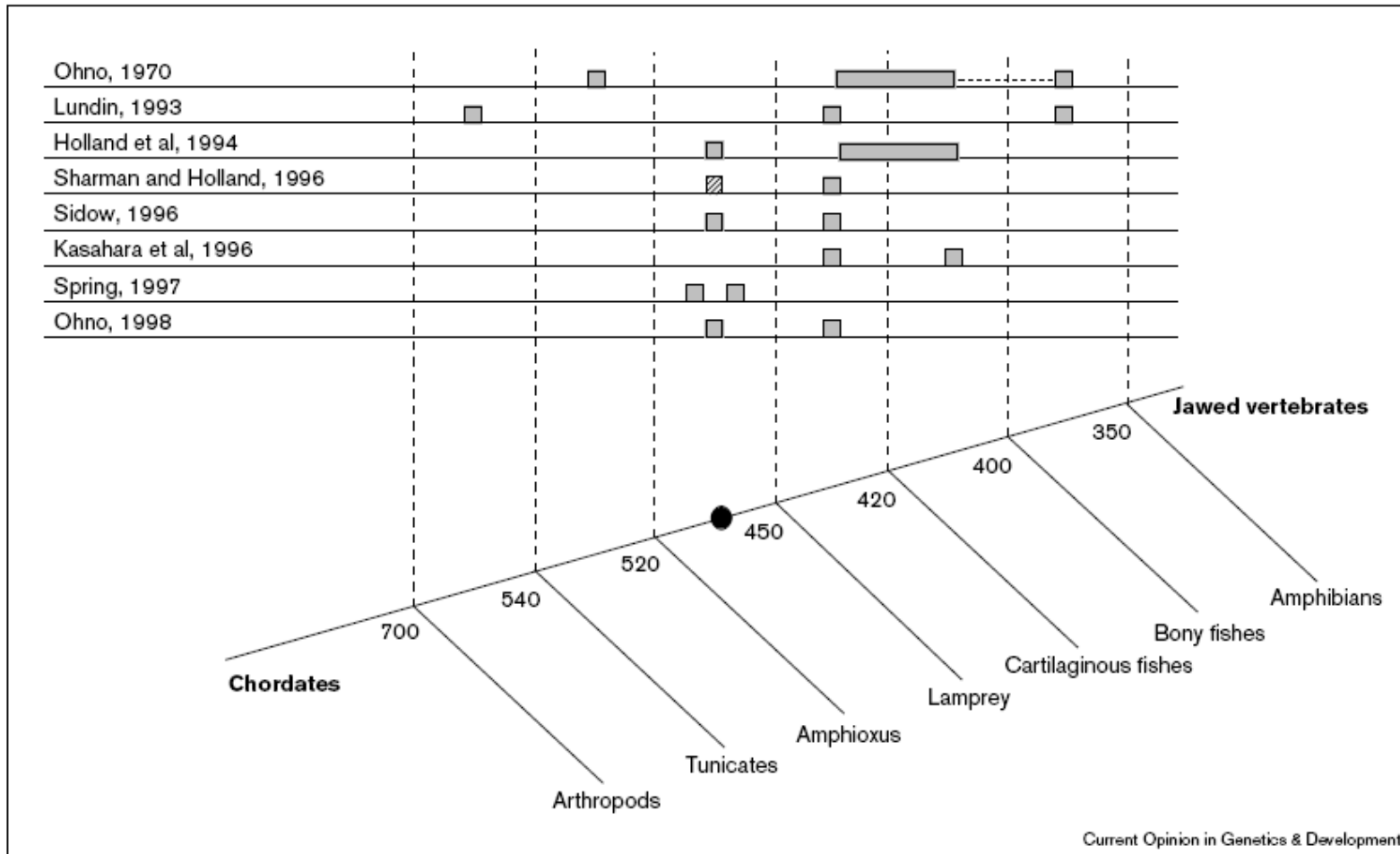
Geni ortologi (a) e geni paraloghi (b).

Sopra: divergenza tra geni ortologi, ossia presenti in specie diverse e derivati da uno stesso gene nell'antenato comune alle due specie.

Sotto: divergenza tra geni paraloghi, ossia presenti in uno stesso organismo a seguito di un evento di duplicazione genica.

# 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)

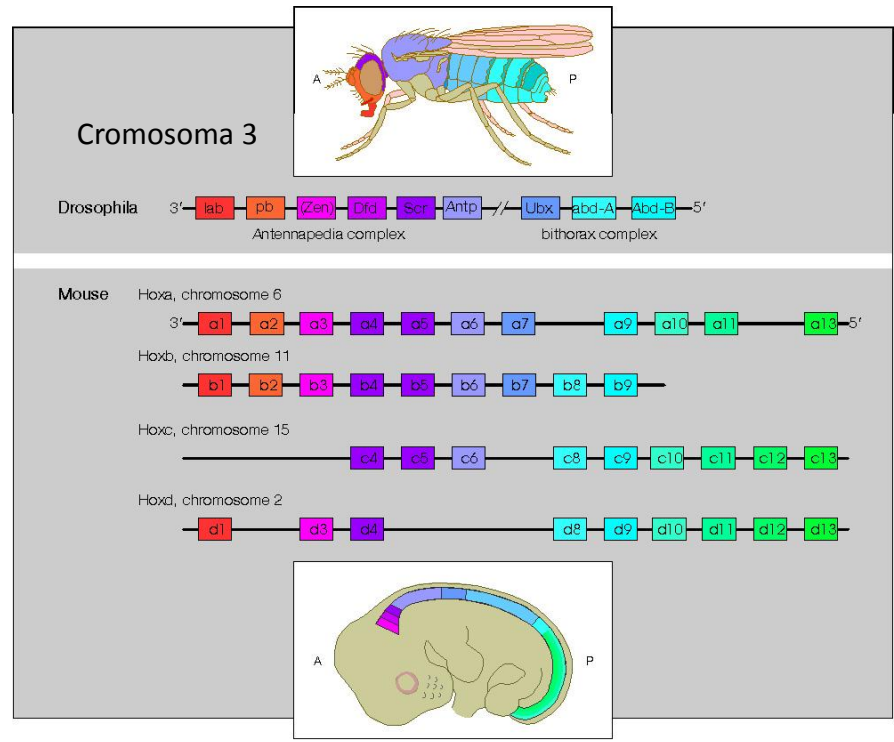
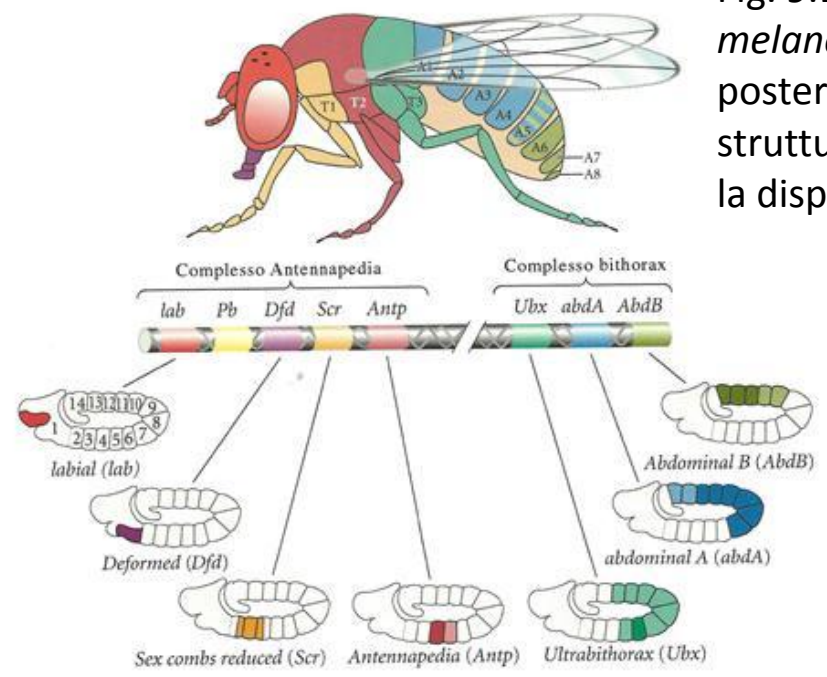
Ancora molta discussione sul numero e la datazione delle duplicazioni

**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



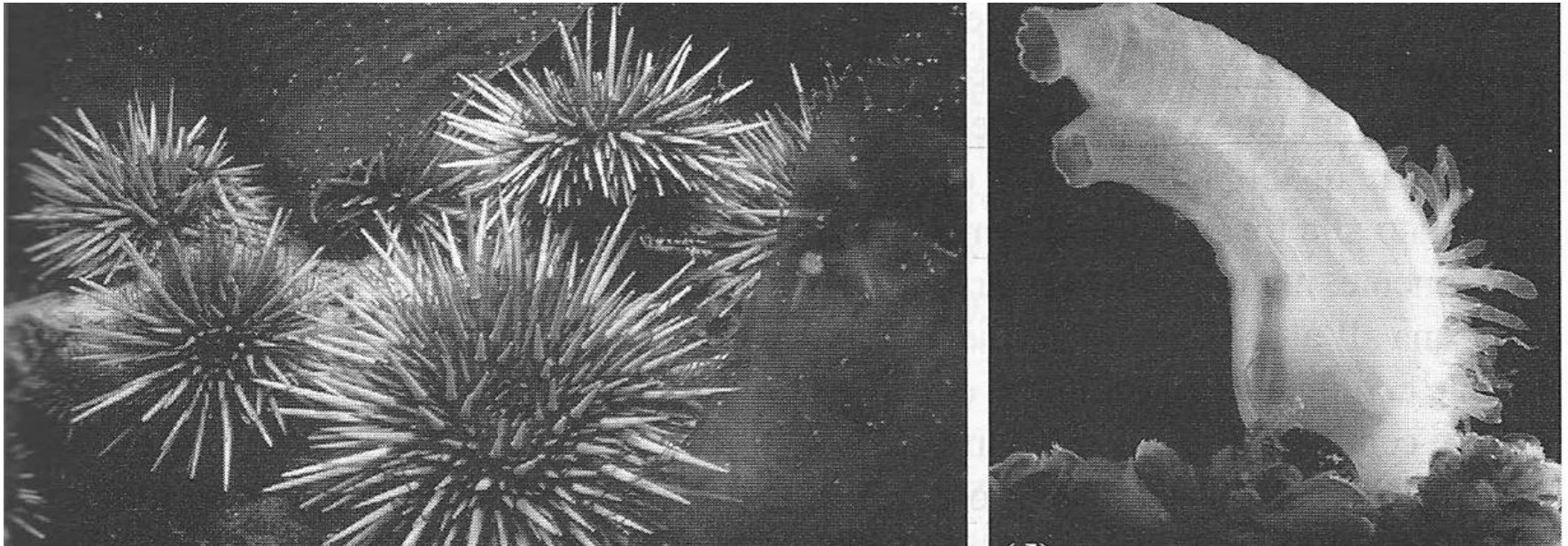
>Complessità  
>Nr geni omeotici

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)

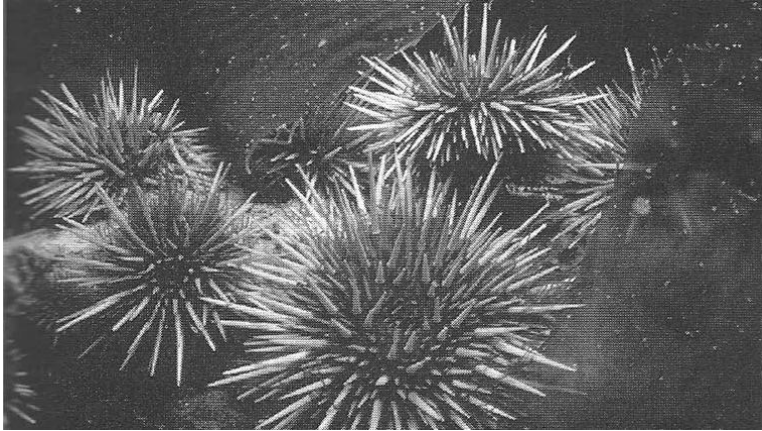


Il sequenziamento dei genomi di riccio di mare (*Strongylocentrotus purpuratus*, echinoderma) e dell'ascidia (*Ciona intestinalis*, tunicato) mostrano numerosi geni in **proporzione 1 a 4 rispetto agli omologhi vertebrati**

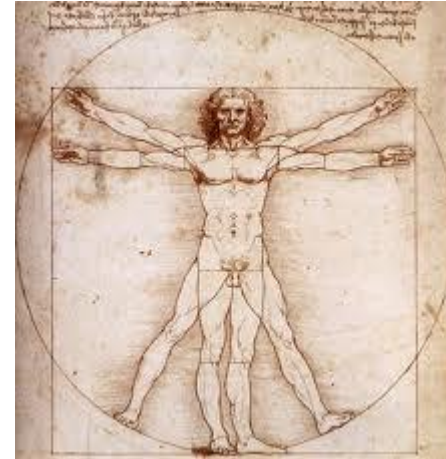
# Come si originano le novità evolutive: duplicazioni genomiche

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**



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

# Duplicazione genica: destino dei geni duplicati

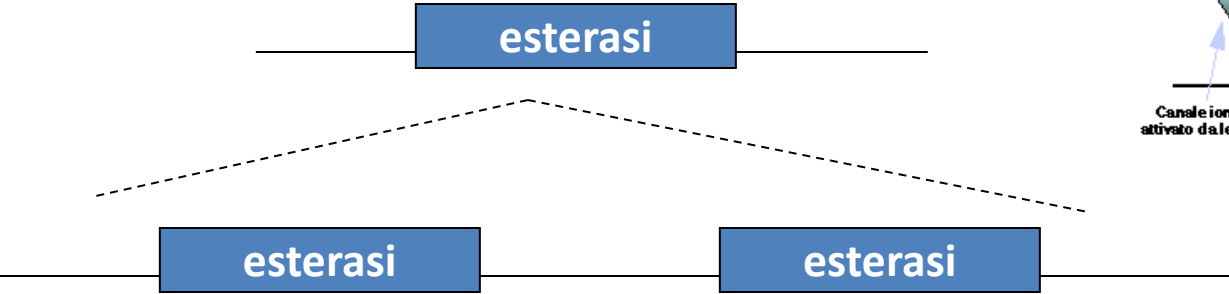
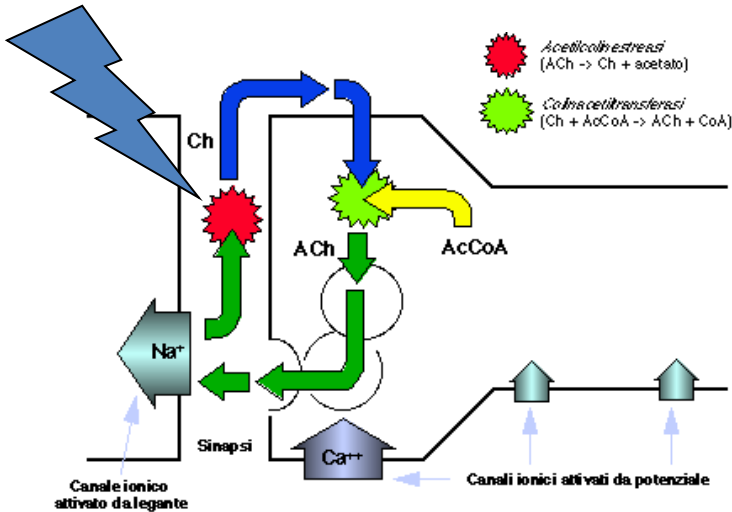
1. Tutte le copie mantengono la stessa funzione.



Resistenza ai pesticidi: **zanzare**

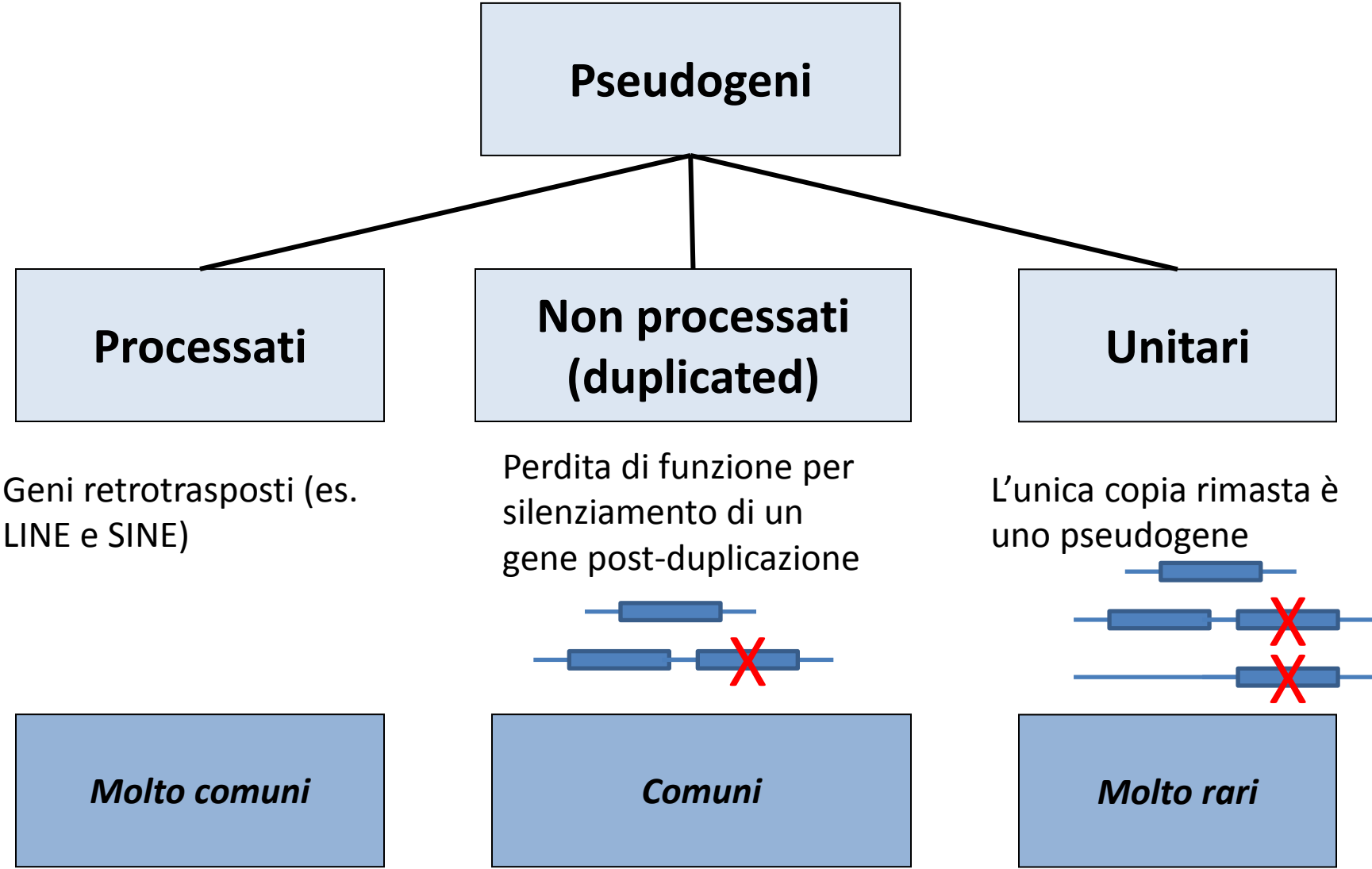
Organofosfati: neurotossine che colpiscono il SN degli insetti

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



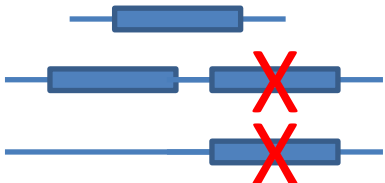
# Duplicazione genica: destino dei geni duplicati

2. Alcune copie scompaiono.



# Duplicazione genica: destino dei geni duplicati

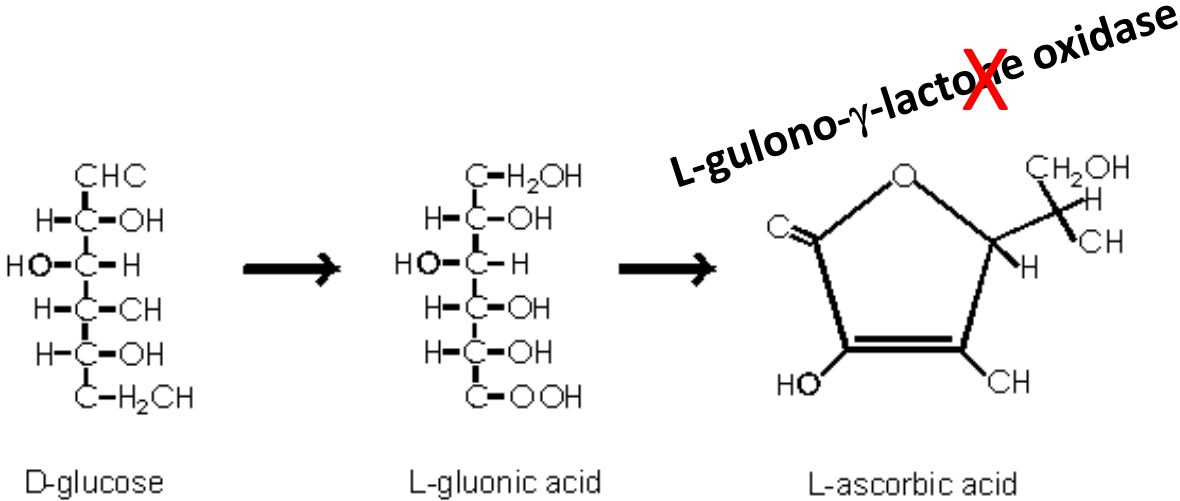
## 2. Alcune copie scompaiono. Pseudogeni unitari



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

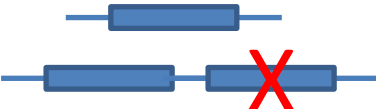
- Cavie
  - Uomo
  - Trote
- si ammalano di scorbuto se non assumo acido ascorbico (vitamina C)



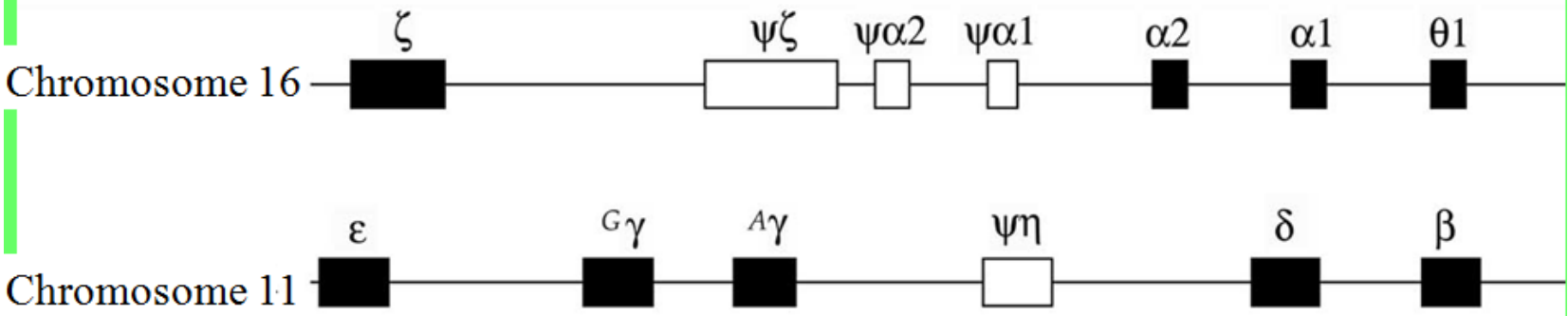
Lo pseudogene non ha una controparte funzionante: pseudogene unitario  
Mutazioni diverse in uomo e cavia: perdita di funzione in momenti diversi

# 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 $\psi\alpha 1$ |          | +           | +           | +              | +                    | +            | +          | +                             |
| Human $\psi\zeta 1$  |          |             |             | +              |                      |              |            |                               |
| Mouse $\psi\alpha 3$ | +        |             | +           | +              |                      | +            |            |                               |
| Mouse $\psi\alpha 4$ |          |             | +           |                | +                    |              |            |                               |
| Mouse $\beta h 3$    | ?        | +           | +           | +              | +                    | +            | ?          | ?                             |
| Goat $\psi\beta^x$   | +        |             | +           | +              | +                    | +            | +          | +                             |
| Goat $\psi\beta^z$   | +        |             | +           | +              | +                    | +            | +          | +                             |
| Rabbit $\psi\beta 2$ |          |             | +           | +              | +                    | +            |            |                               |

# Duplicazione genica: destino dei geni duplicati

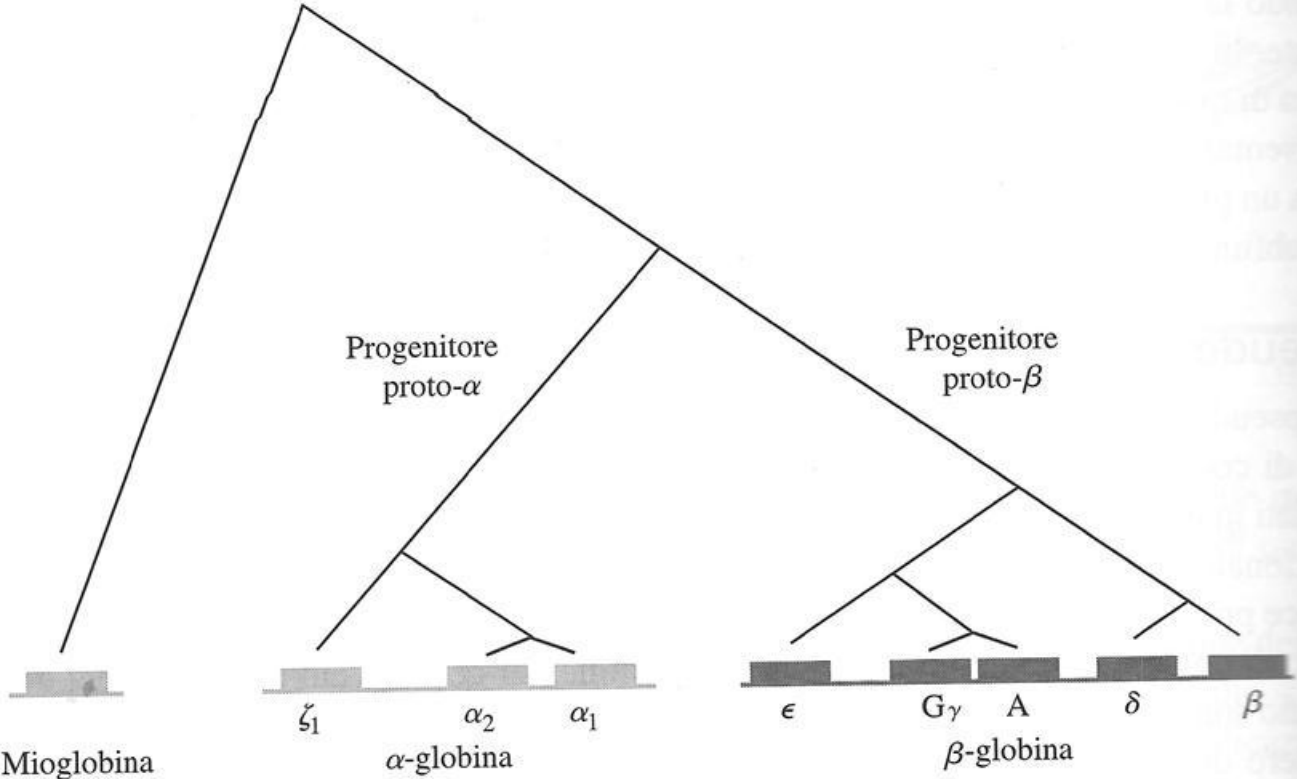
3. Alcune copie evolvono con una nuova funzione.



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.



# Nell'uomo

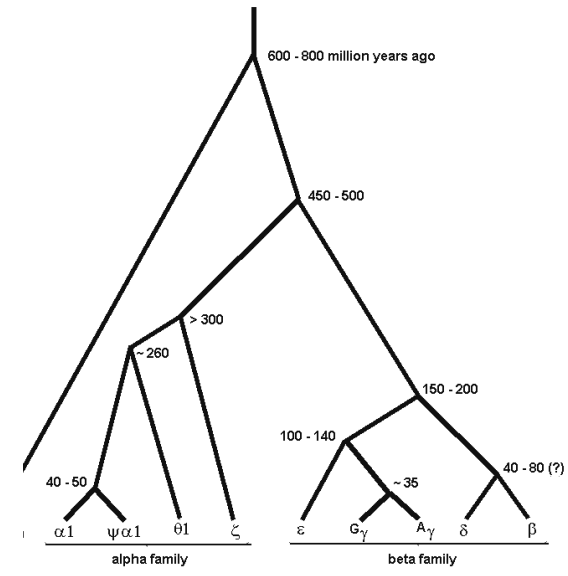
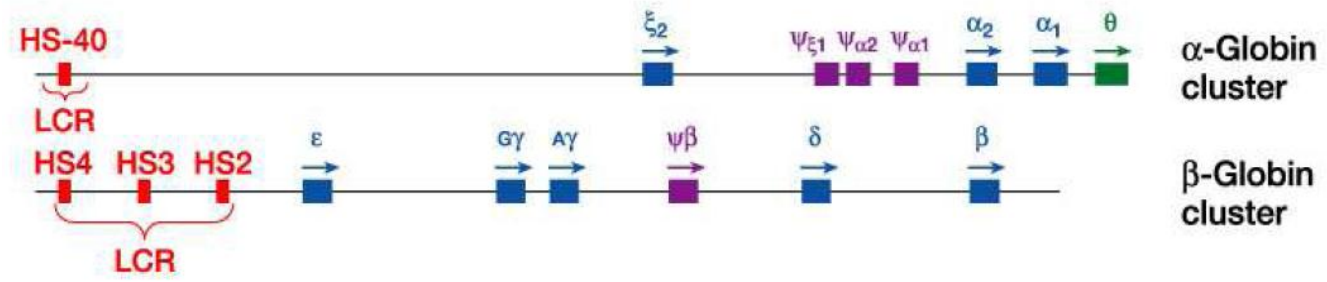
$\zeta_2\varepsilon_2$  Emoglobina Embrionale (espressione nel sacco vitellino)

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

$\alpha_2\delta_2$  HbA<sub>2</sub> Emoglobina dell'adulto

$\alpha_2\beta_2$  HbA

0 10 20 30 40 50 60 70 kb



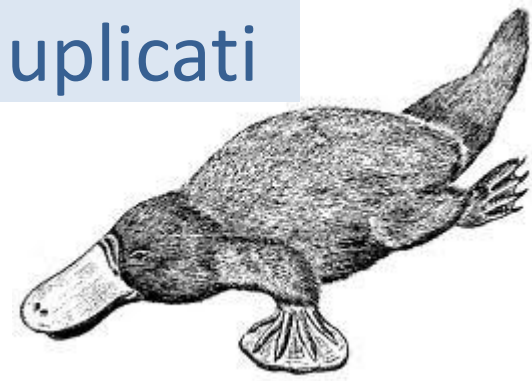
L'emoglobina nell'adulto è di tipo HbA 96%  $\alpha_2\beta_2$ , HbA<sub>2</sub>  $\alpha_2\delta_2$  3% e HbF  $\alpha_2\gamma_2$  1%



# 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

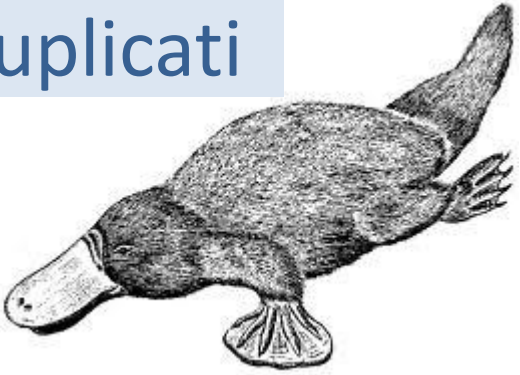
---

## 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!)

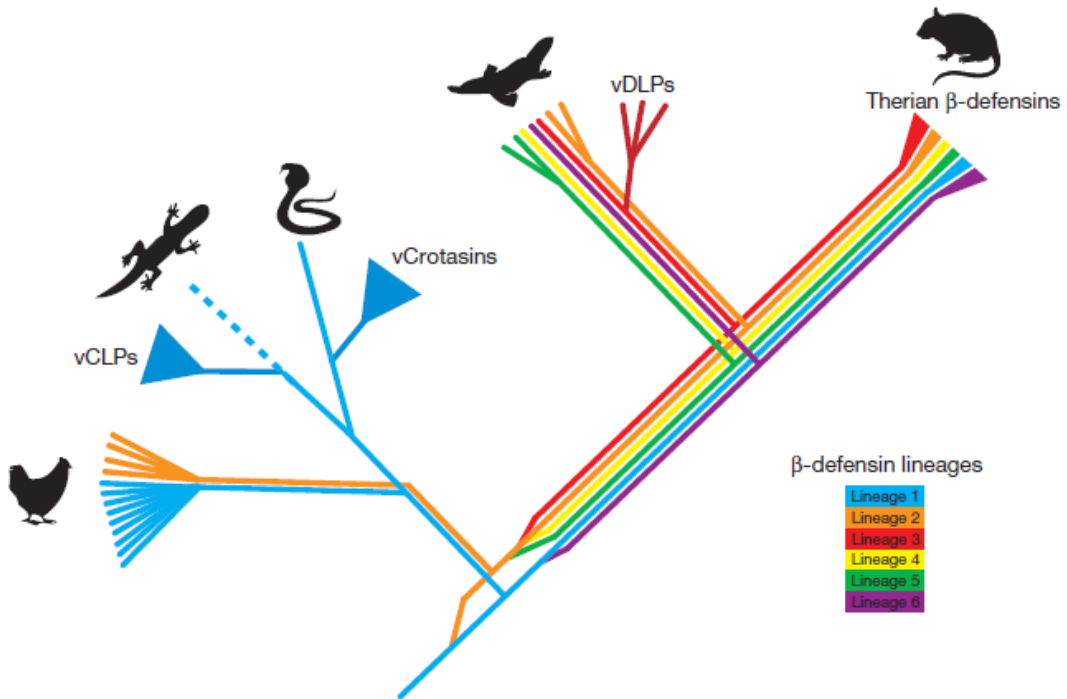
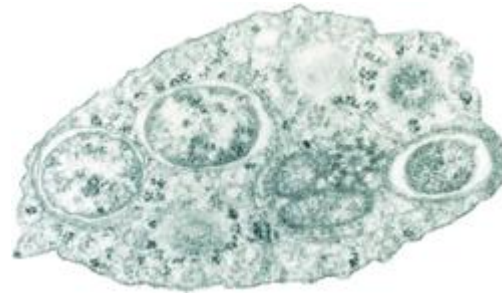
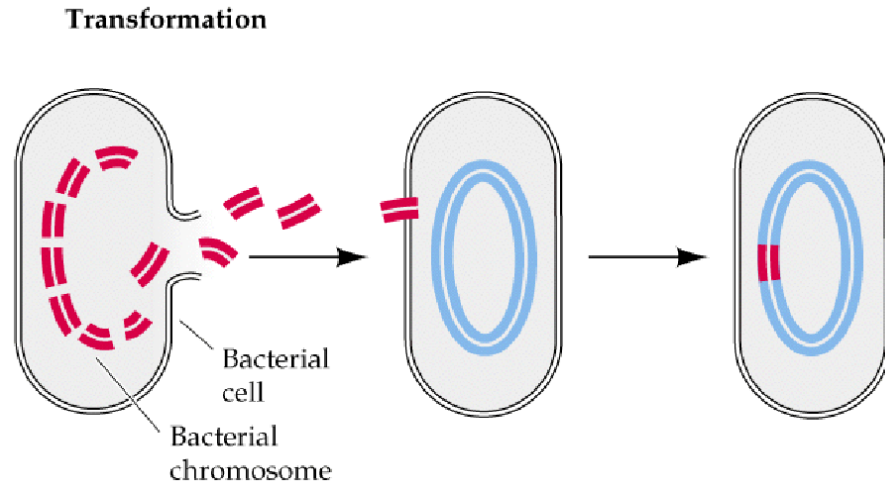
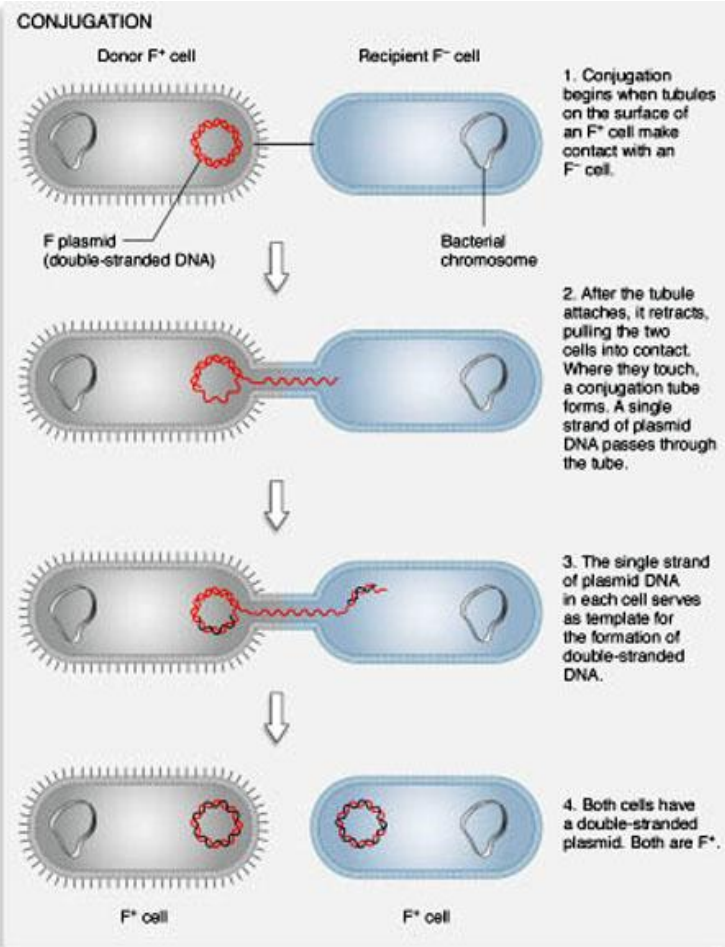


Figure 4 | The evolution of  $\beta$ -defensin peptides in platypus venom gland. The diagram illustrates separate gene duplications in different parts of the phylogeny for platypus venom defensin-like peptides (vDLPs), for lizard

venom crotoxin-like peptides (vCLPs) and for snake venom crotoxinamines. These venom proteins have thus been co-opted from pre-existing non-toxin homologues independently in platypus and in lizards and snakes<sup>48</sup>.

# Come si originano le novità evolutive: trasferimento orizzontale



*Wolbachia* ( $\alpha$ -proteobatterio) all'interno di una cellula di insetto: trasmessa per generazioni tramite eredità materna