

# Genomica dei Sistemi Modello Vegetali

Simone Ferrari

[Simone.ferrari@uniroma1.it](mailto:Simone.ferrari@uniroma1.it)

Edificio di Botanica

- <https://elearning2.uniroma1.it/course/view.php?id=1987>
- <http://www.plantgenome.uga.edu/links.htm>
- <http://www.arabidopsis.org/>

- Alimentazione
- Regolazione del clima
- Energia
- Farmaci
- Coloranti, aromi, ecc. ecc...



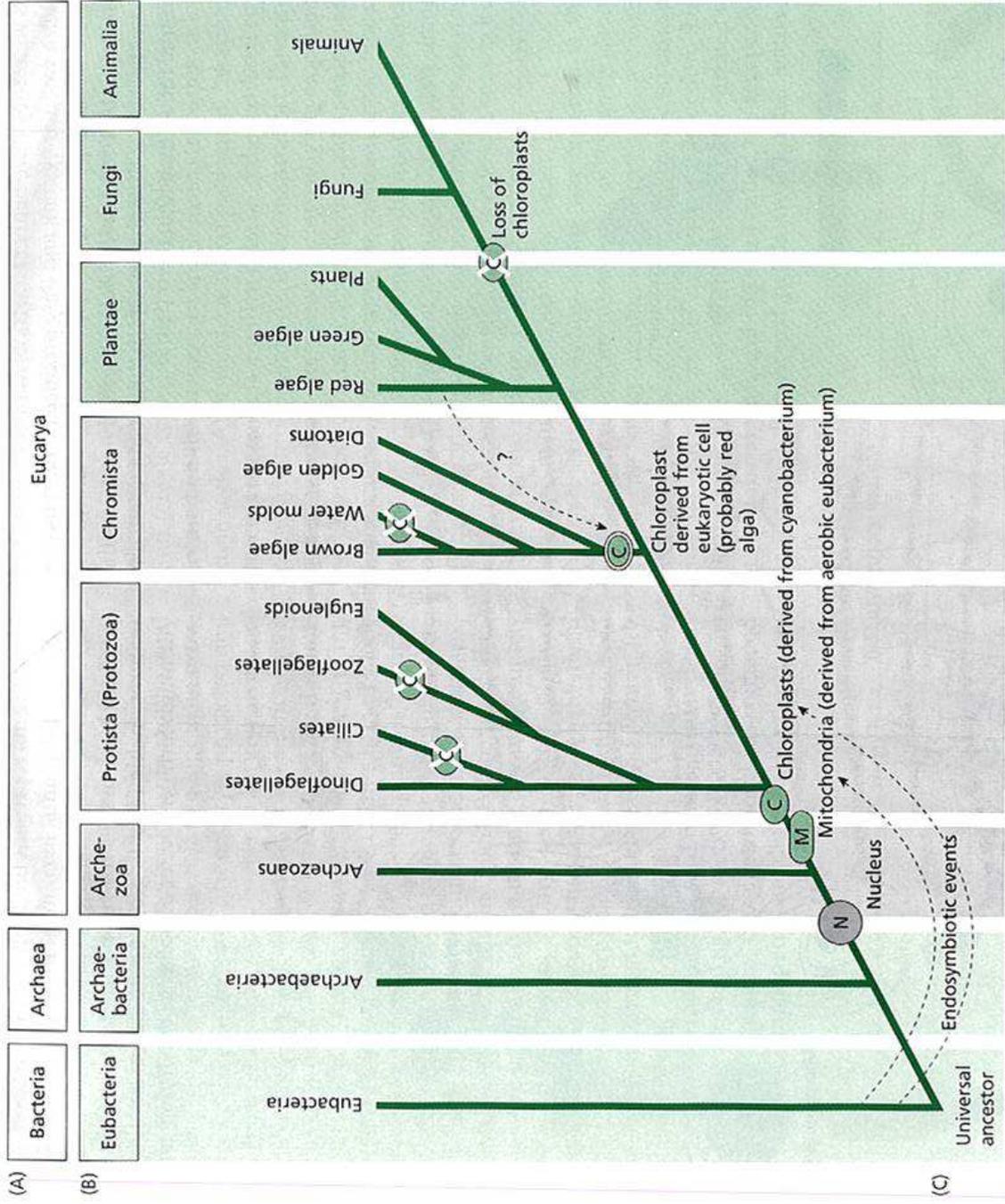
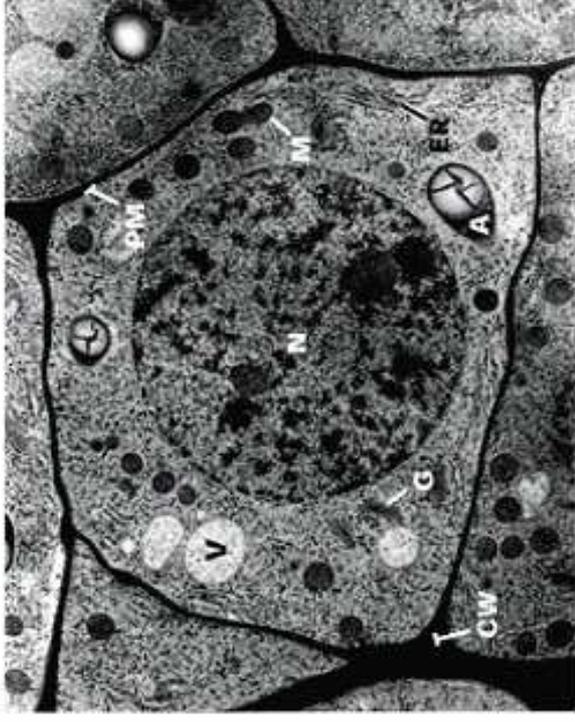
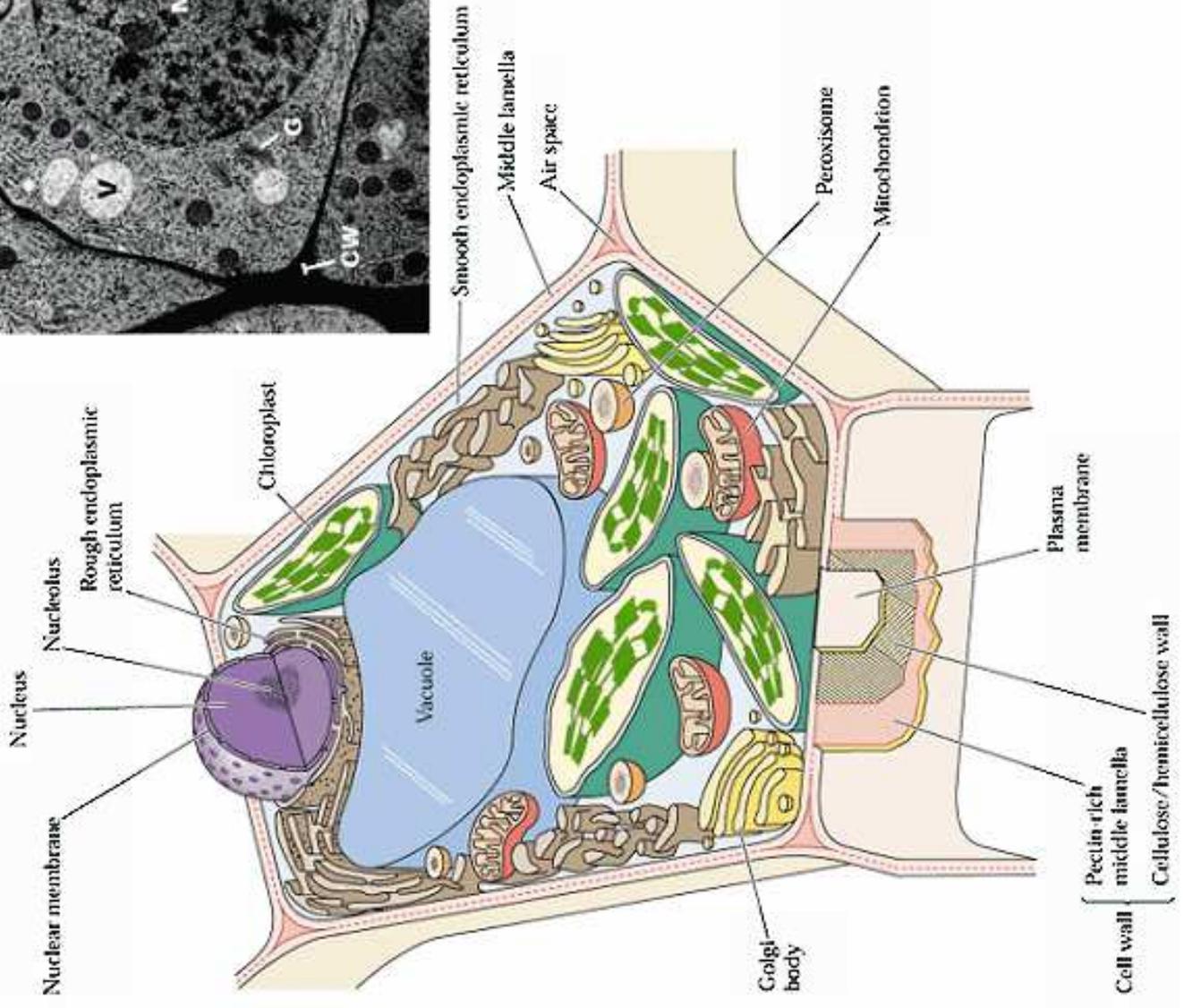


Figure 1.1 Natural classification scheme and phylogeny of living organisms. (A) The three domains of life. (B) Division of organisms into two prokaryotic kingdoms and six eukaryotic kingdoms. (C) Hypothetical phylogeny showing the origin of the major groups of eukaryotes. The appearance of the nucleus and the acquisition of mitochondria and chloroplasts as endosymbiotic organelles are indicated by the dashed lines and arrows. (After Campbell 1996.)

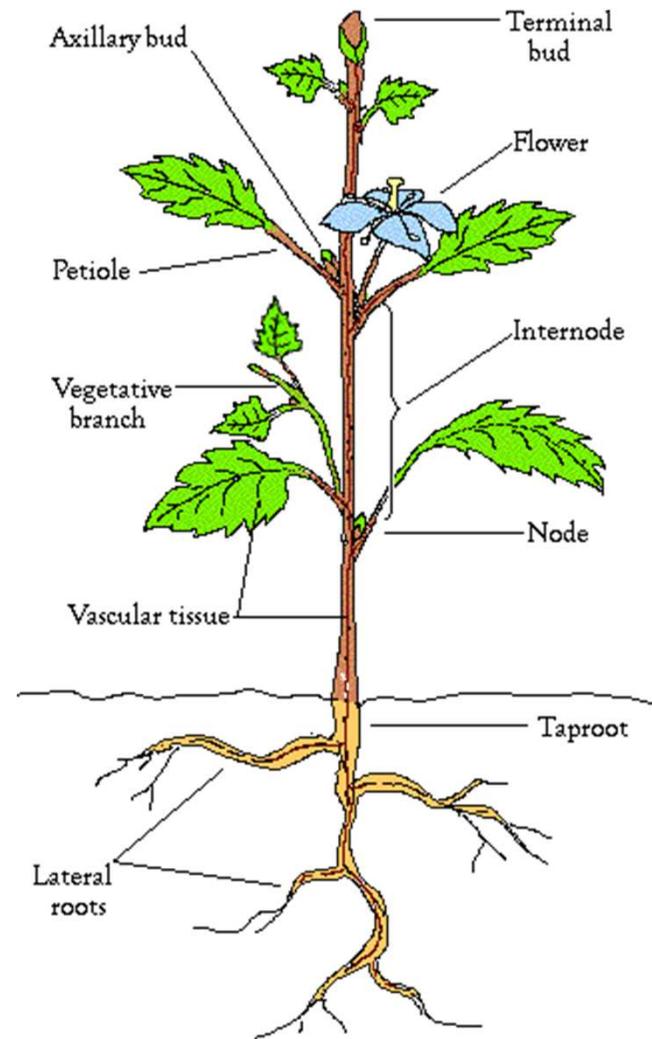
(B)

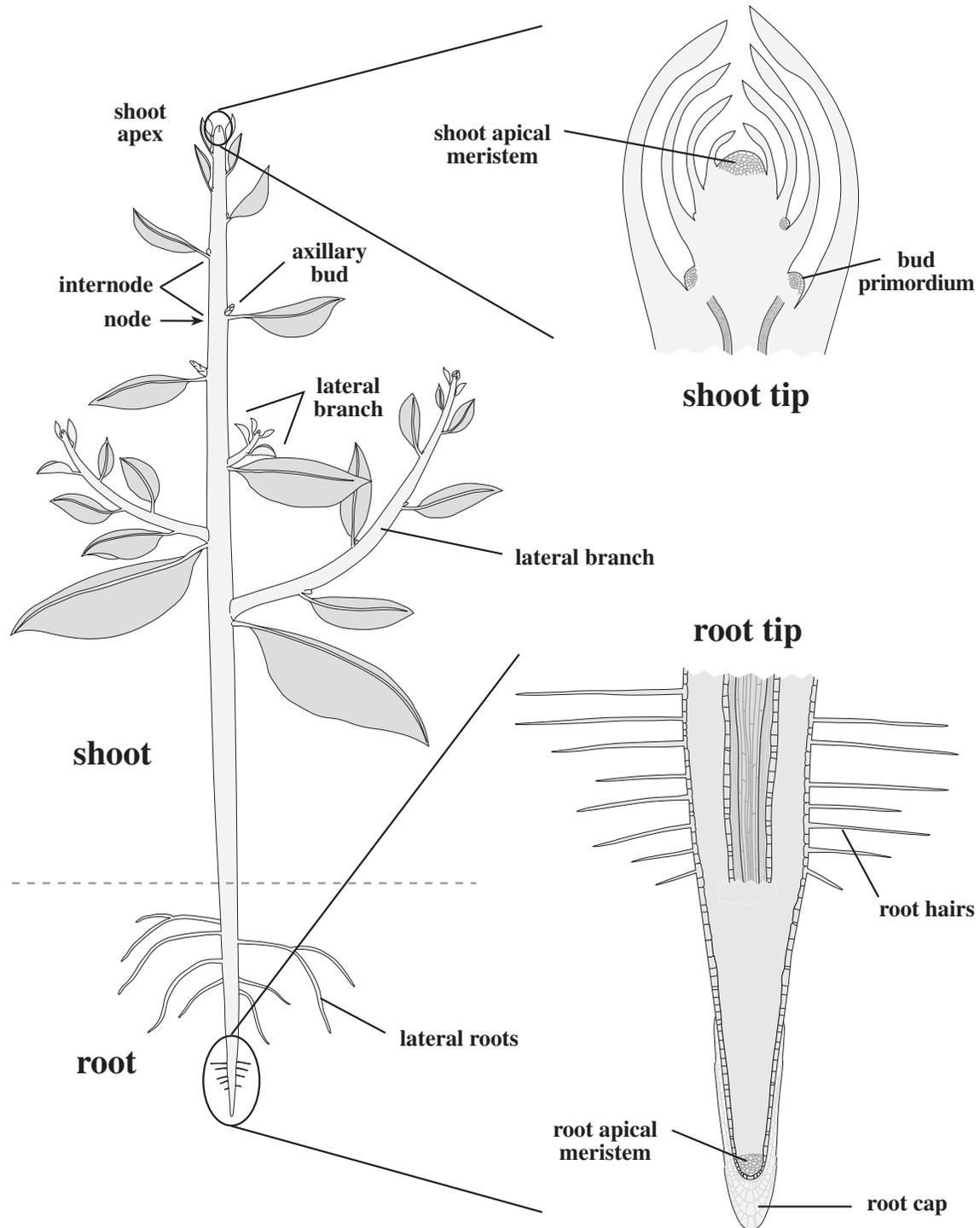


(A) Mesophyll



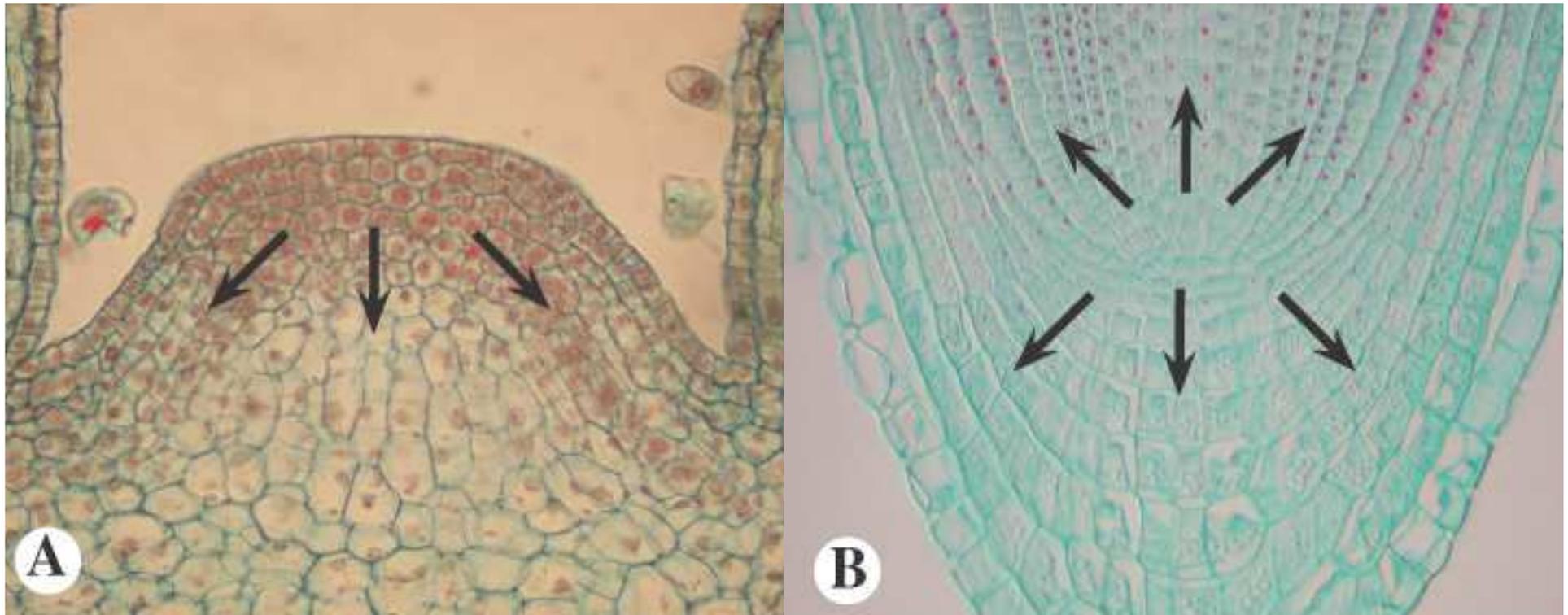
# Gli organi delle piante



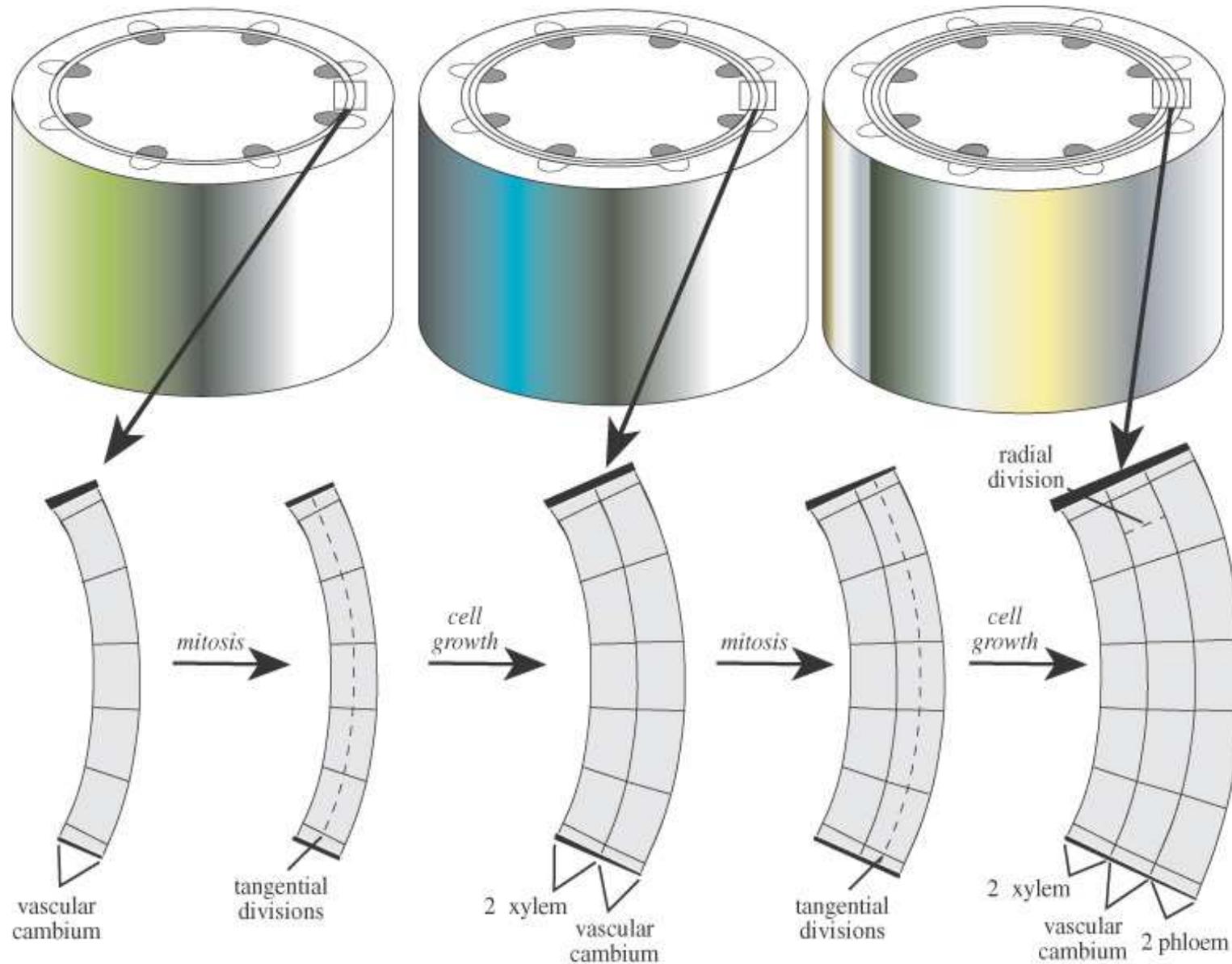


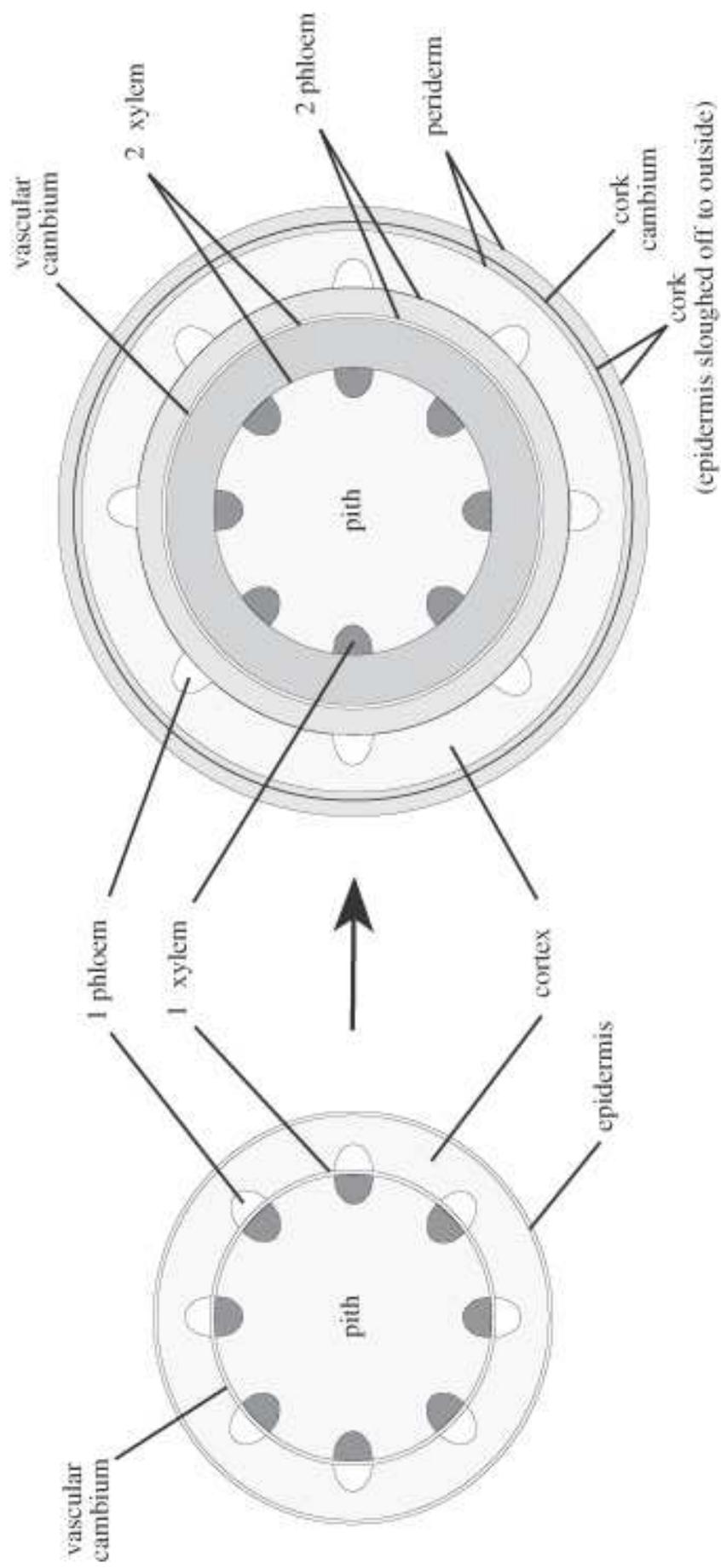
**Meristemi** = regioni in cui c'è attiva divisione cellulare

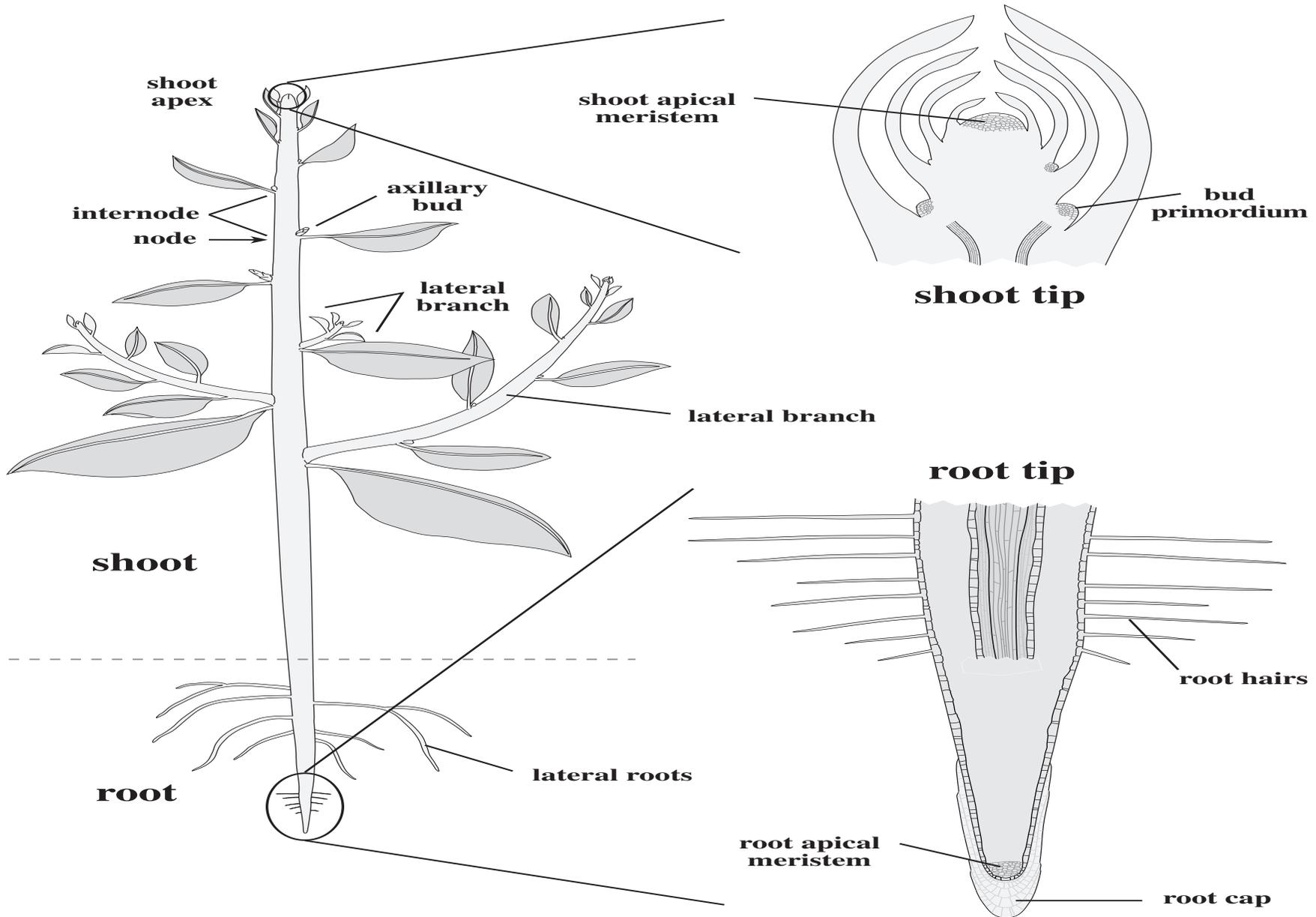
- 1) meristemi **apicali** (germoglio e radice);
- 2) meristemi **lateral**i (cambio cribro-vascolare)



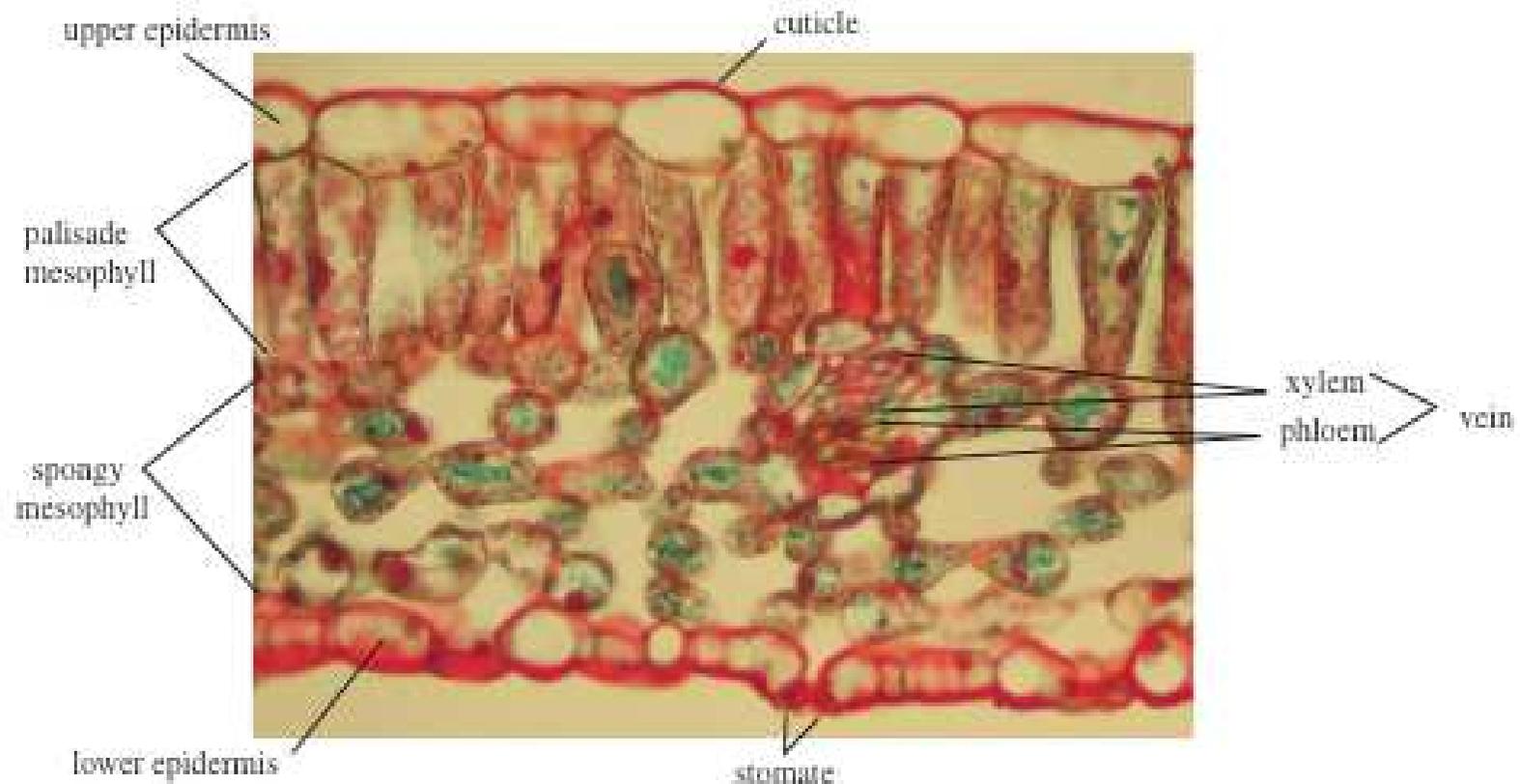
# Cambio cribrovascolare



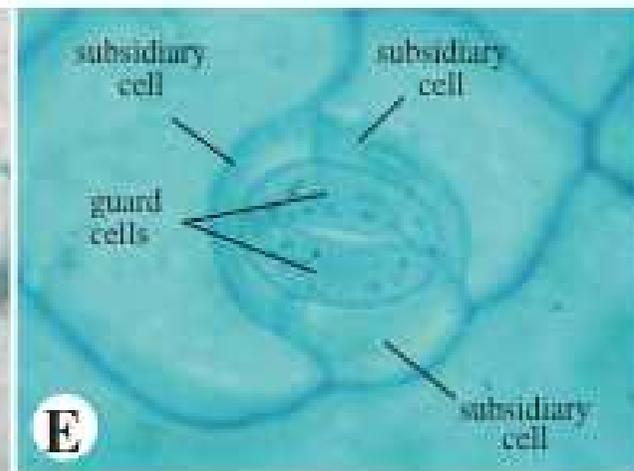
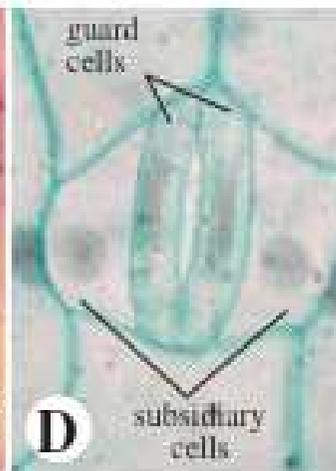
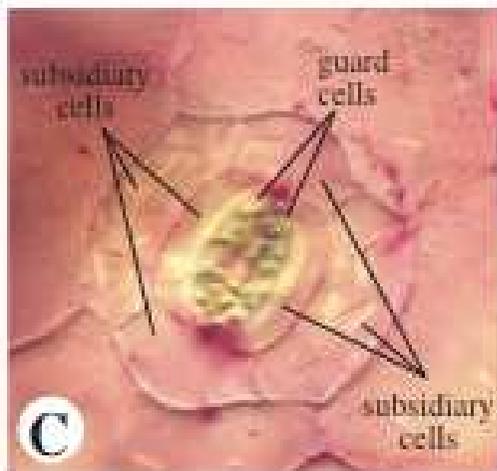
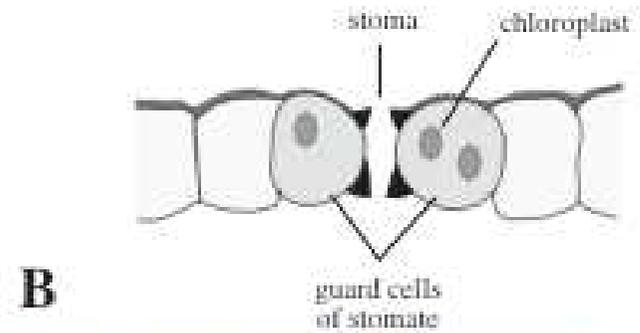
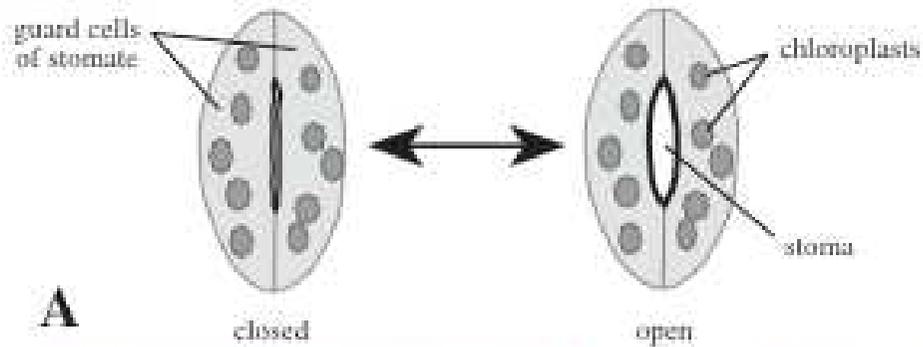




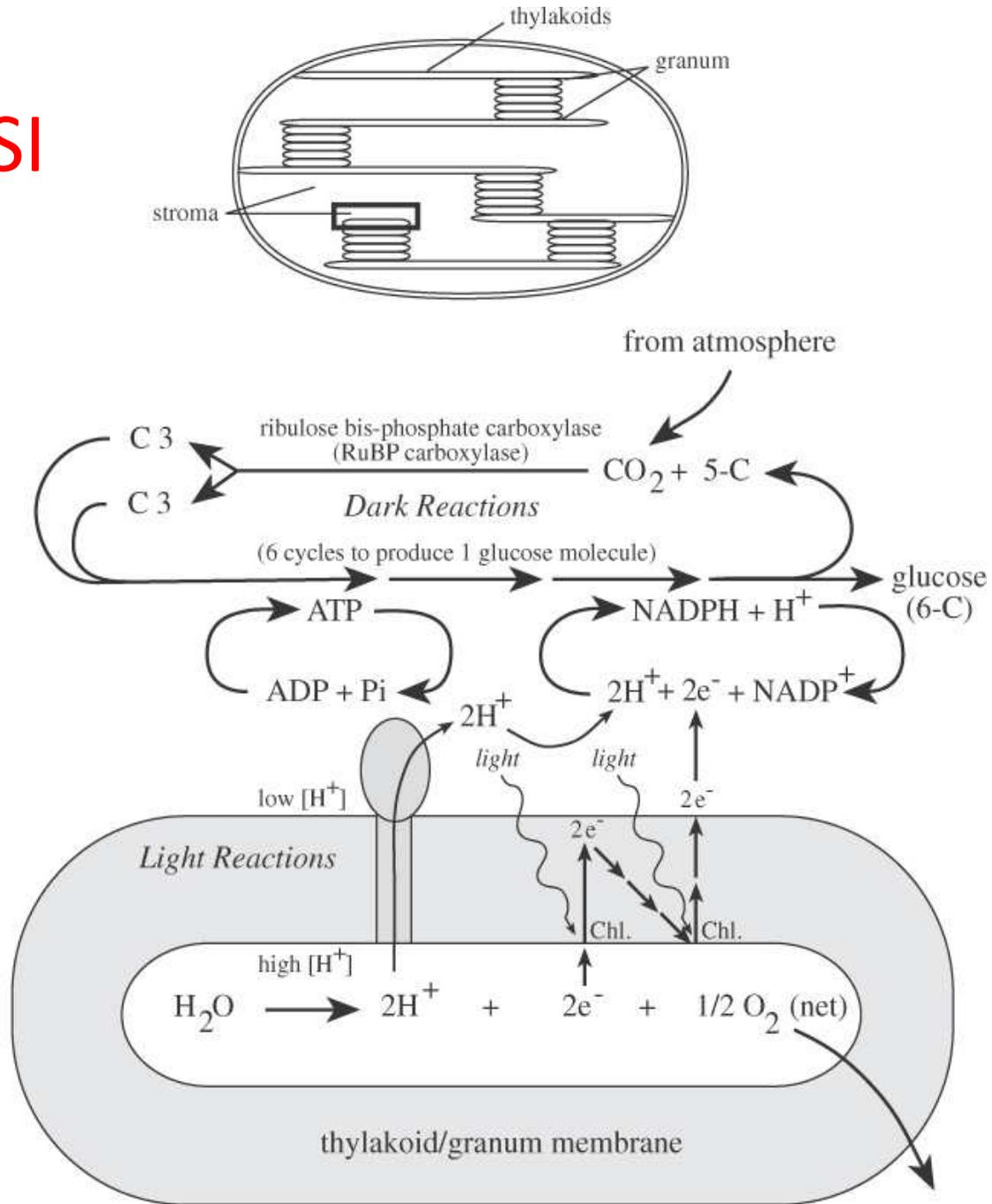
# ANATOMIA DELLA FOGLIA



# STOMI



# FOTOSINTESI



# ORMONI VEGETALI

1. Auxine
2. Gibberelline
3. Citochinine
4. Acido abscissico
5. Etilene
6. Giasmonati
7. Acido salicilico

# 1. AUXINE

- Promuovono la crescita cellulare
- Controllano il differenziamento embrionale, la produzione di radici laterali e lo sviluppo di alcuni frutti



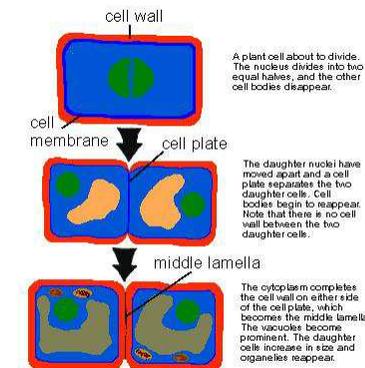
## 2. Gibberelline

- Promuovono l'allungamento dello stelo



## 3. Citochinine

- Promuovono la divisione cellulare ed il differenziamento delle parti aeree



## 4. Acido abscissico

- Promuove la chiusura degli stomi
- Induce la dormienza dei semi



# ETILENE

- Promuove la maturazione di alcuni frutti
- Regola alcune risposte agli stress



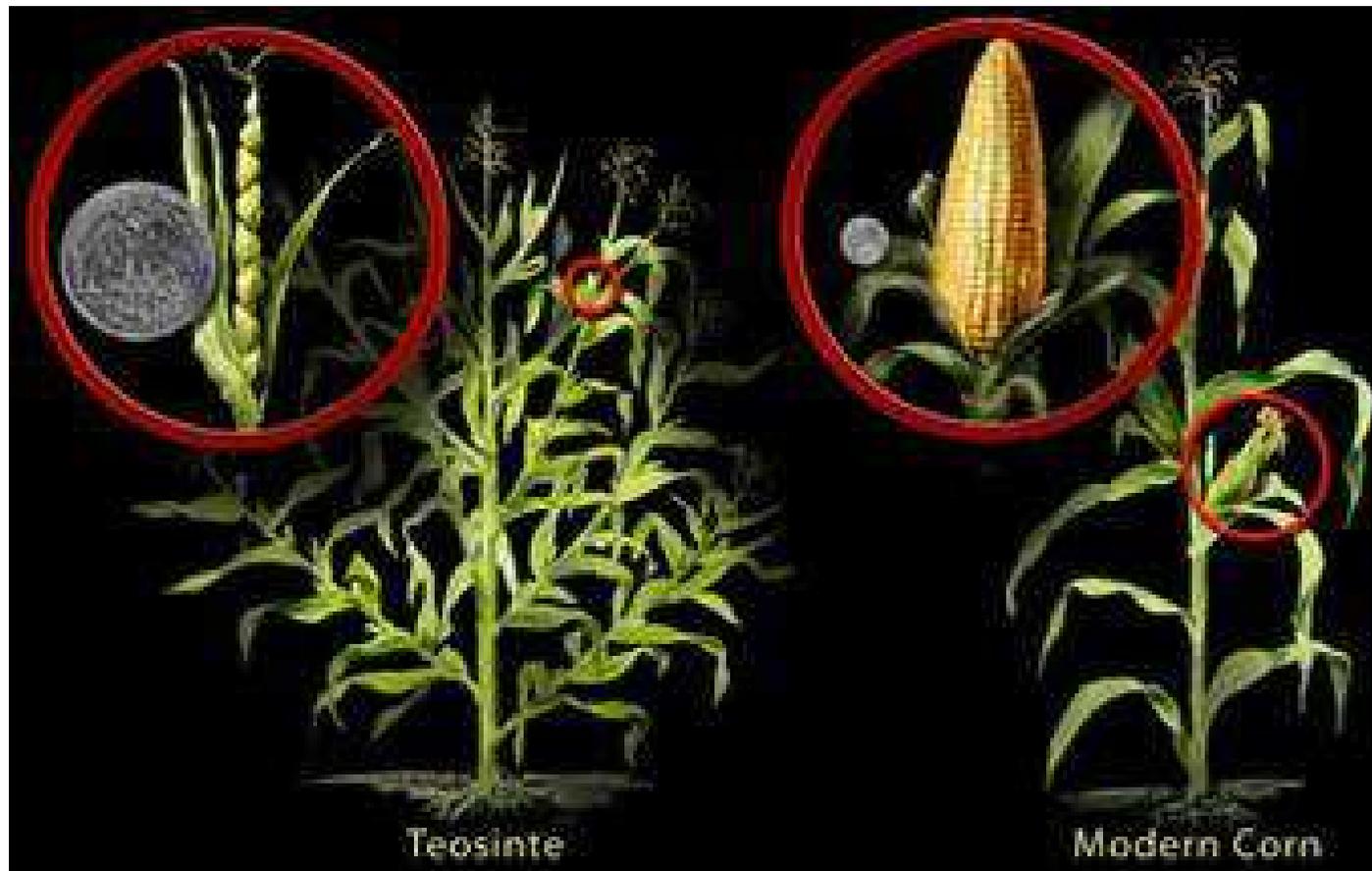
**Ethylene**



Aria

Etilene

## Addomesticamento delle specie vegetali



# **STUDIO DEI GENOMI VEGETALI**

- **Identificazione di geni importanti per caratteri agronomici (produttività, resistenza a stress, proprietà nutrizionali)**
- **Comprensione dell'evoluzione delle piante**

# PIANTE MODELLO

## *Arabidopsis thaliana*

Dicotiledone (*Brassicaceae*)

Piccolo genoma dipolide (C1 = 125 Mbp)

Trasformabile facilmente

5 cromosomi

Piccole dimensioni

Ciclo vitale breve (2 mesi)



<http://www.arabidopsis.org/>

The screenshot shows a web browser window displaying the TAIR (The Arabidopsis Information Resource) homepage. The browser's address bar shows the URL <http://www.arabidopsis.org/>. The page features a navigation menu with links for Home, Help, Contact, About Us, and Login/Register. Below the navigation is a search bar with a dropdown menu set to 'Gene' and a 'Search' button. The main content area is titled 'The Arabidopsis Information Resource' and contains several paragraphs of text describing the database and its resources. On the right side, there are social media links for RSS, Twitter, and Facebook, followed by a 'Breaking News' section with two news items: 'ABRC tool for adding comments about stocks' and 'Notice from ABRC for Chinese Researchers'. The page also includes a logo for Carnegie Institution for Science.

TAIR - Home Page

File Modifica Visualizza Preferiti Strumenti ?

Autenticazione Google Ap... Scheda attività didattica S... MIUR - loginmiur.cineca Google blast Basic Local Alignme... Home - PubMed - NCBI Raccolta Web Slice

 Home Help Contact About Us Login/Register

Search Browse Tools Portals Download Submit News ABRC Stocks

### The Arabidopsis Information Resource

The Arabidopsis Information Resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available from TAIR includes the complete genome sequence along with gene structure, gene product information, gene expression, DNA and seed stocks, genome maps, genetic and physical markers, publications, and information about the Arabidopsis research community. Gene product function data is updated every week from the latest published research literature and community data submissions. TAIR also provides extensive linkouts from our data pages to other Arabidopsis resources.

The Arabidopsis Biological Resource Center at The Ohio State University collects, reproduces, preserves and distributes seed and DNA resources of *Arabidopsis thaliana* and related species. Stock information and ordering for the ABRC are fully integrated into TAIR.

 TAIR is located at the Carnegie Institution for Science Department of Plant Biology and funded by the National Science Foundation.

Beginning on October 1, 2013, access to TAIR by researchers at companies or for-profit institutions requires a subscription. Academic / nonprofit access to TAIR will require a subscription after April 1, 2014. ABRC catalog, stock and ordering pages hosted at TAIR will remain accessible without a subscription. Please see our subscription page for further details.



**Subscribe to news feed**

**Follow our Twitter feed**

**Join our Facebook group**

#### Breaking News

**ABRC tool for adding comments about stocks**  
[January 29, 2014]  
We are encouraging users to add comments about stocks that they have ordered. We are collecting information such as validation of insert/sequence, and also problems such as poor germination, lethality, incorrect insert, etc.

**Notice from ABRC for Chinese Researchers**  
[January 15, 2014]  
In an effort to improve successful delivery of orders to China, we will include your address in Chinese on all

# Quali specie sequenziare?

- Impatto economico, sociale e scientifico
- Distanza filogenetica da altre specie sequenziate (-> nuove informazioni)
- Informazioni disponibili (mappe genetiche e fisiche)
- Capacità di persuasione dei ricercatori



# GENOMICA COMPARATIVA

- Analisi e confronto di genomi di specie diverse
- Fornisce informazioni sull'evoluzione delle specie e sulla funzione di geni e sequenze non codificanti
- Es.: funzione di un gene dedotta dallo studio di geni ortologhi in specie modello

# GENOMICA COMPARATIVA

## Cosa si analizza?

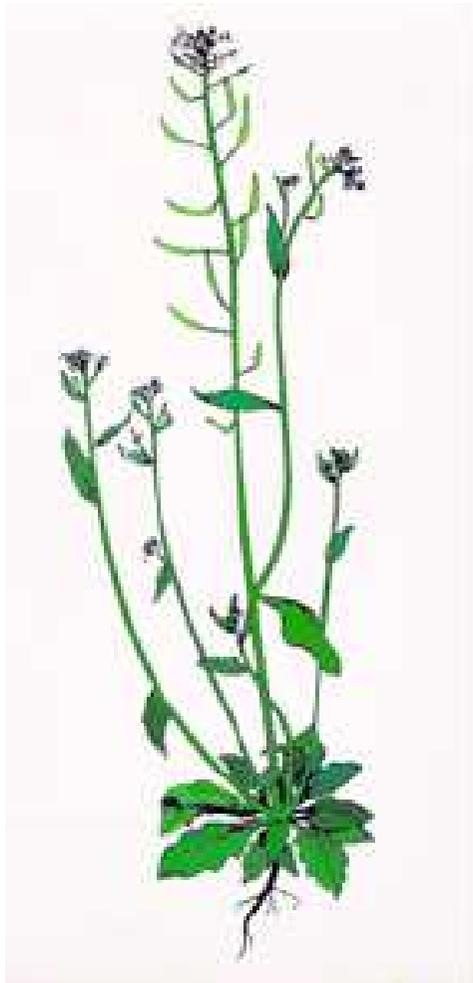
- Similarità di sequenza
- Localizzazione cromosomica dei geni
- Lunghezza e numero esoni
- Quantità di DNA non codificante
- Conservazione di regioni cromosomiche

# **Ostacoli al sequenziamento di specie coltivate**

- **Dimensioni**
- **DNA ripetitivo**
- **Poliploidia**

# Dimensioni del genoma

**Arabidopsis: 125 Mb**



***Fritillaria assyriaca*: 125 Gb!**



## ILLUMINA GENOME ANALYZER

Permette il sequenziamento in parallelo di un numero massiccio di frammenti genomici

-> 1 milione di basi sequenziate per volta!



# **DNA ripetitivo**

**Responsabile per gran parte della  
variabilità nelle dimensioni del  
genoma vegetale**

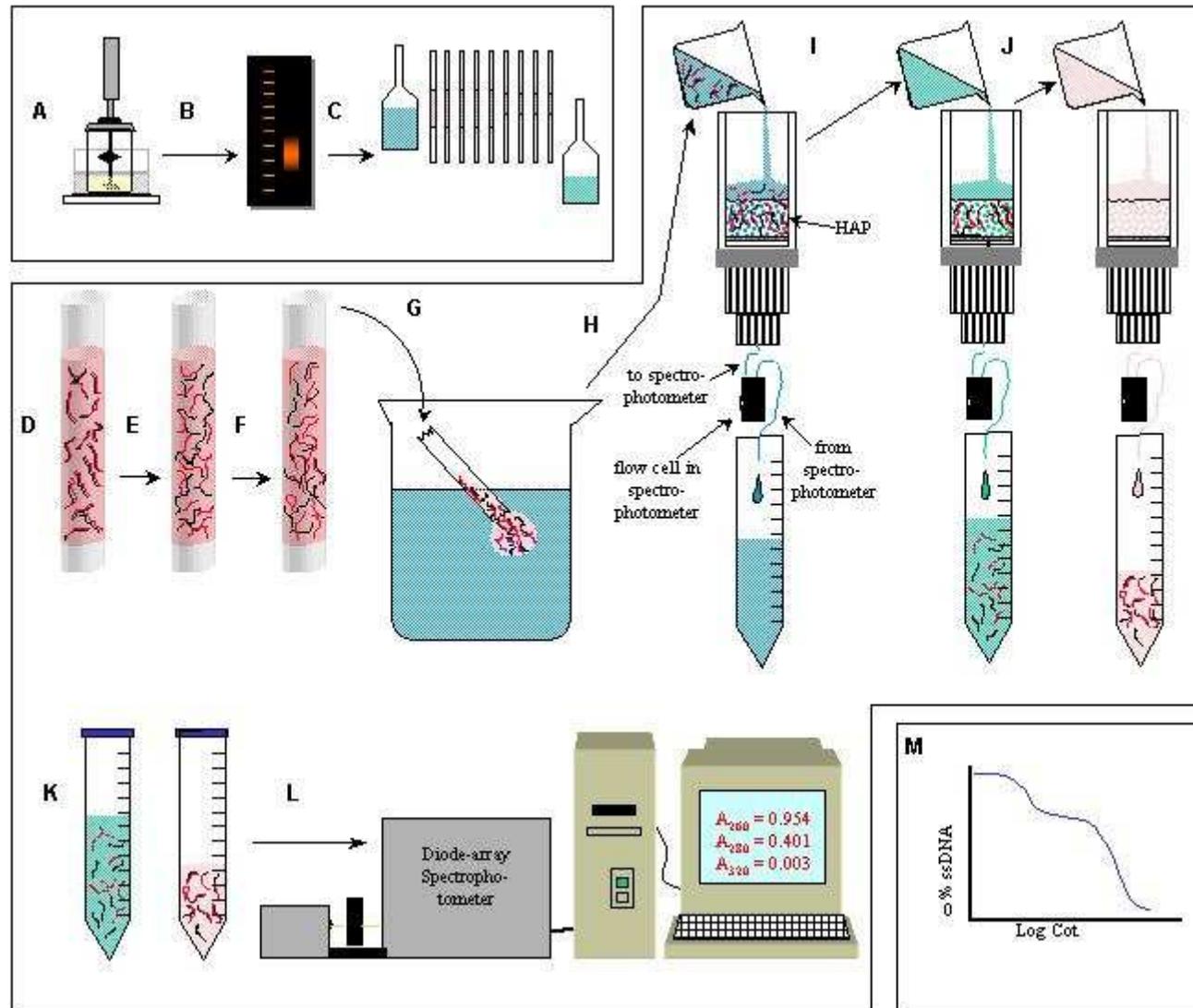
**Complica l'assemblamento delle  
sequenze**

**Sequenze non-ridondanti nel genoma: da  
13% (cipolla) a 77% (pomodoro)**

# **DNA ripetitivo**

**N.B.: le piante hanno più DNA ripetitivo degli animali, e copie individuali possono avere meno mutazioni per distinguerle, perchè più recenti**

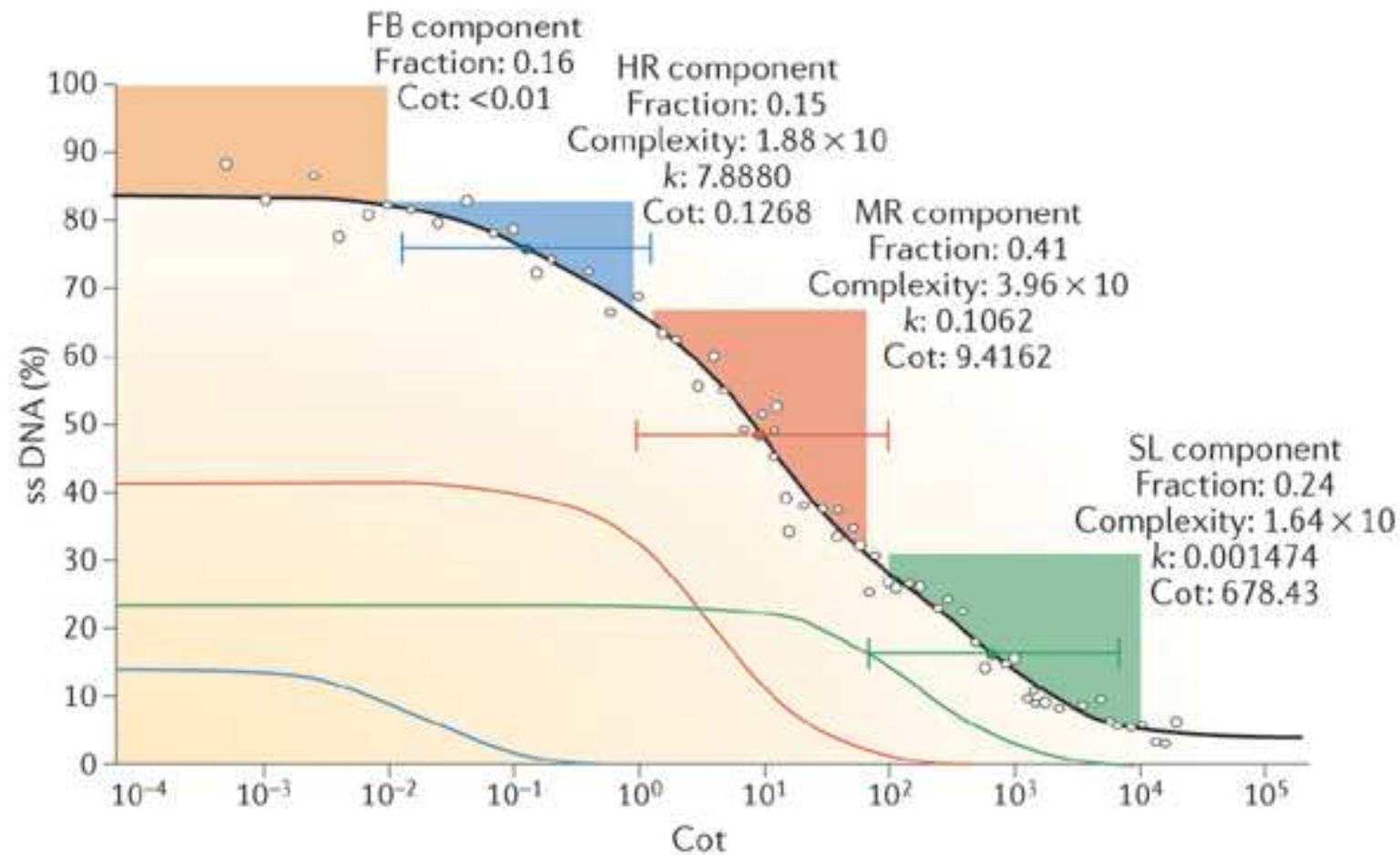
# Cinetica di riassociazione



# Cinetica di riassociazione

- Fornisce il valore  $Cot$ , cioè il prodotto fra la concentrazione dei nucleotidi ( $C_0$ ) ed il tempo di riassociazione (normalizzato per la conc. di cationi nel tampone)
- La cromatografia su colonna di idrossiapatite (che lega il dsDNA) permette di isolare la frazione di DNA che si riassocia ad un particolare valore di  $Cot$ .
- Più il DNA è ripetitivo, più basso sarà il suo valore  $Cot$

# Cinetica di riassociazione



# Clonaggio basato sul valore Cot (CBCS)

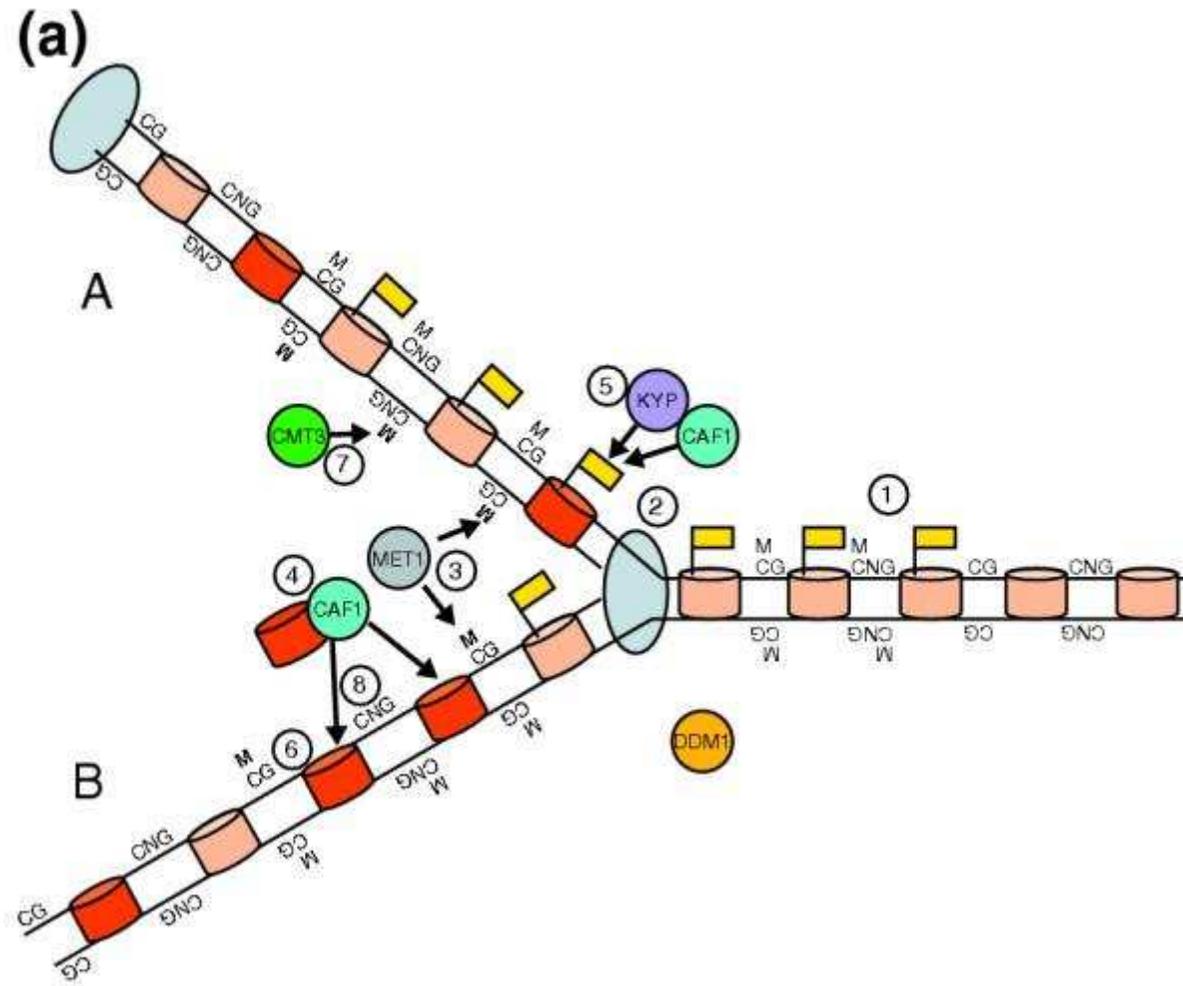
L'analisi Cot permette di isolare specificamente frazioni più o meno ripetitive

DNA meno ripetitivo viene sequenziato

-> più facile da assemblare in contigs

-> maggiore percentuale di geni

il DNA più ricco in geni è ipometilato rispetto a quello non codificante (inclusa una parte di DNA ripetitivo)



# Methylation filtration (MF)

clonaggio del DNA genomico totale in ceppi di *E. coli* che degradano il DNA metilato -> sequenziamento dei cloni e assemblaggio in contigs

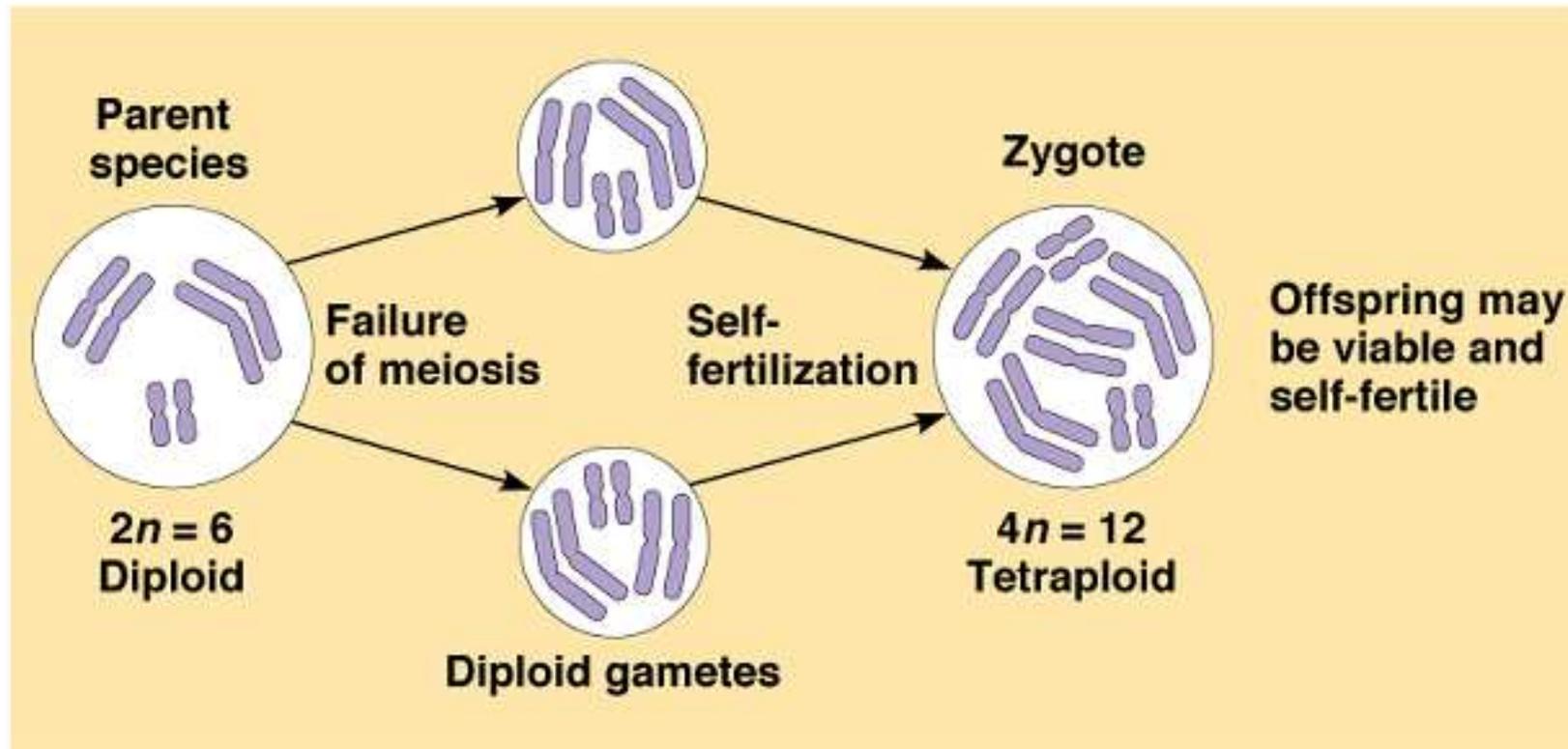
Svantaggio: non sempre il DNA codificante è ipometilato (es. metilazione indotta da stress, o in colture cellulari)

# POLIPLOIDIA

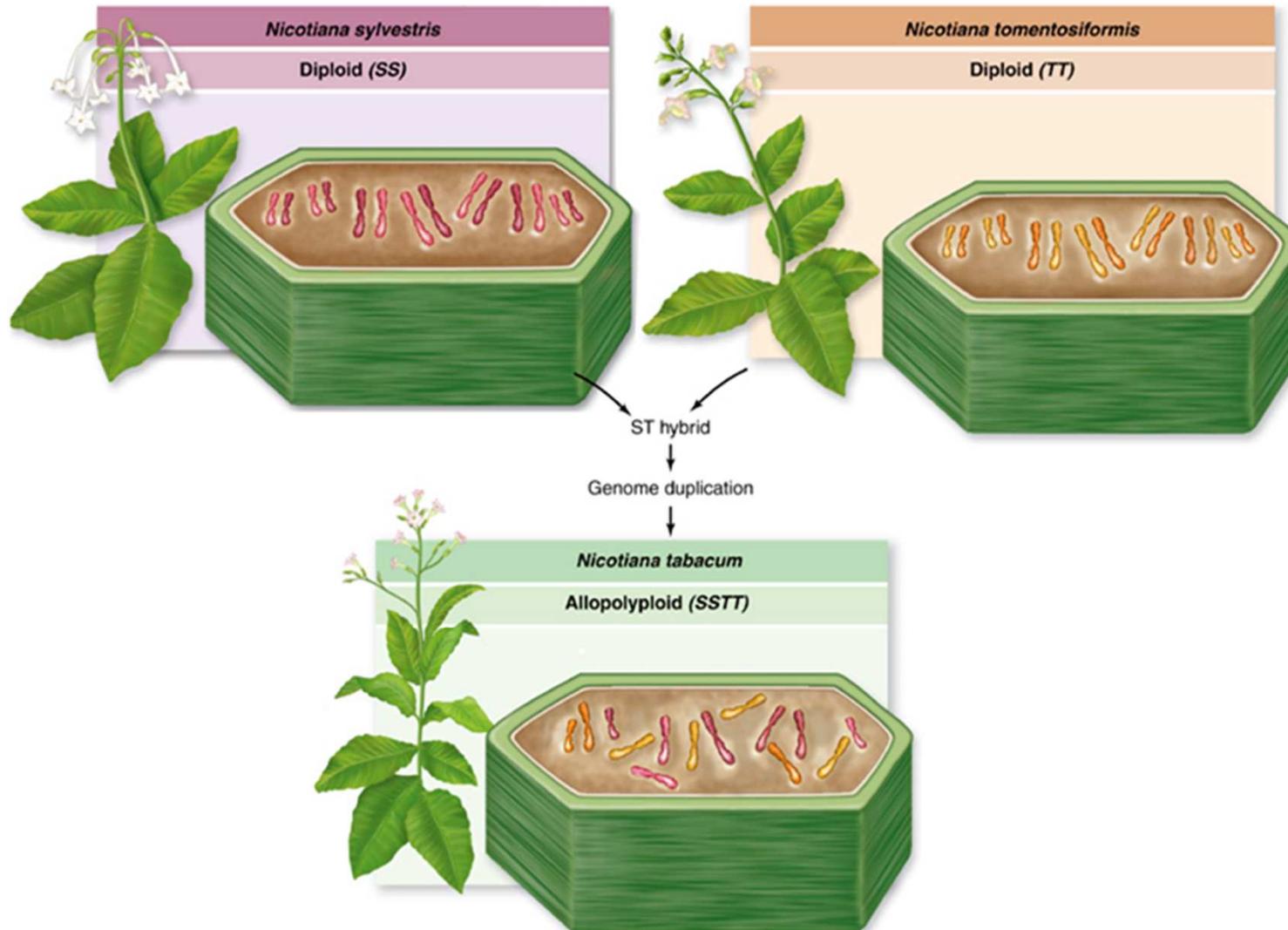
- Duplicazione del genoma in una specie (**autopoliploidia**), attraverso errore meiotico (4 copie di ogni cromosoma)
- Ibridazione di due specie diverse (**allopoliploidia**)

# AUTOPOLIPLOIDIA

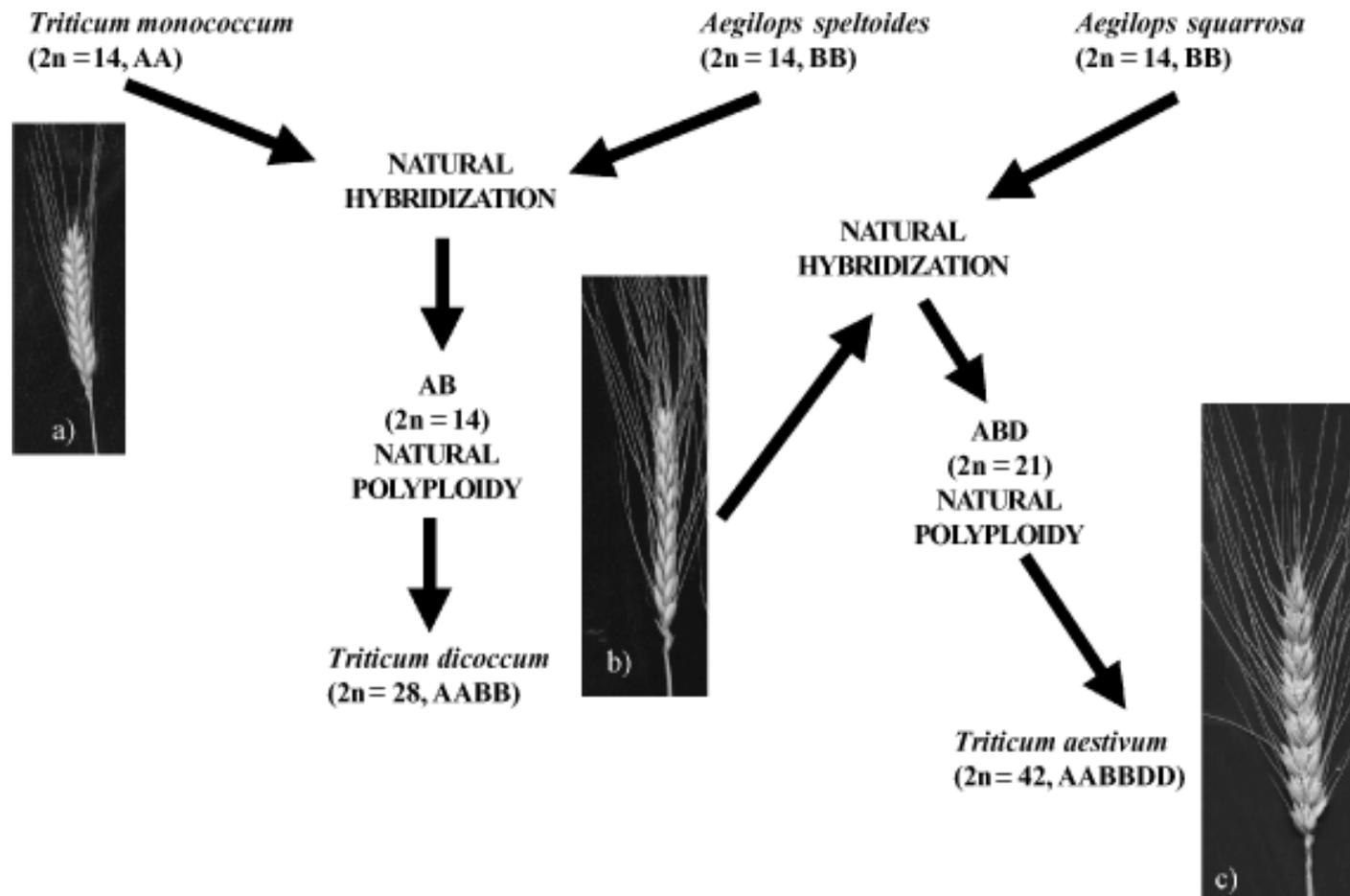
(es. Canna da zucchero, patata, erba medica, caffè)



# ALLOPOLIPLOIDIA: TABACCO



# ALLOPOLIPOIDIA: FRUMENTO

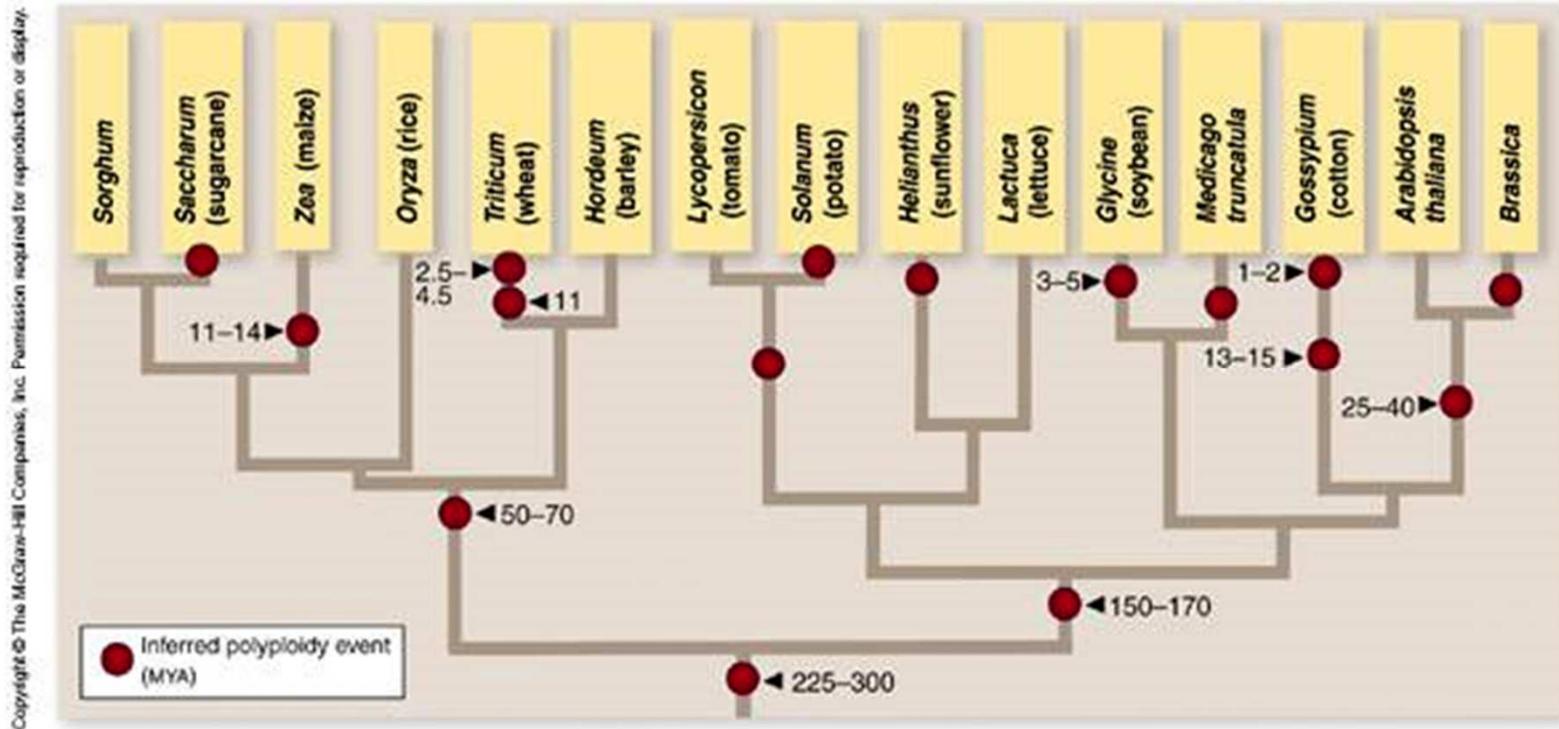


**Figure 1** - Synoptic chart of cultivated wheats evolution: the diploid (2n = 14, AA) forms of *Triticum monococcum* (a) were naturally pollinated by weed species, possible *Aegilops speltoides* (2n = 14, BB?), in about 10,000 B.C. primitive farms. The subsequent genome duplication of hybrids by natural polyploidy gave rise to several wild and cultivated tetraploid species (2n = 28, AABB) like *Triticum dicoccum* (b) and *Triticum durum* (Figure 2a); again, the natural pollination of the tetraploid *T. dicoccum* (b) by another weed species, *Aegilops squarrosa* (2n = 14, DD) gave rise to the hexaploid (2n = 42, AABBDD) species (c).

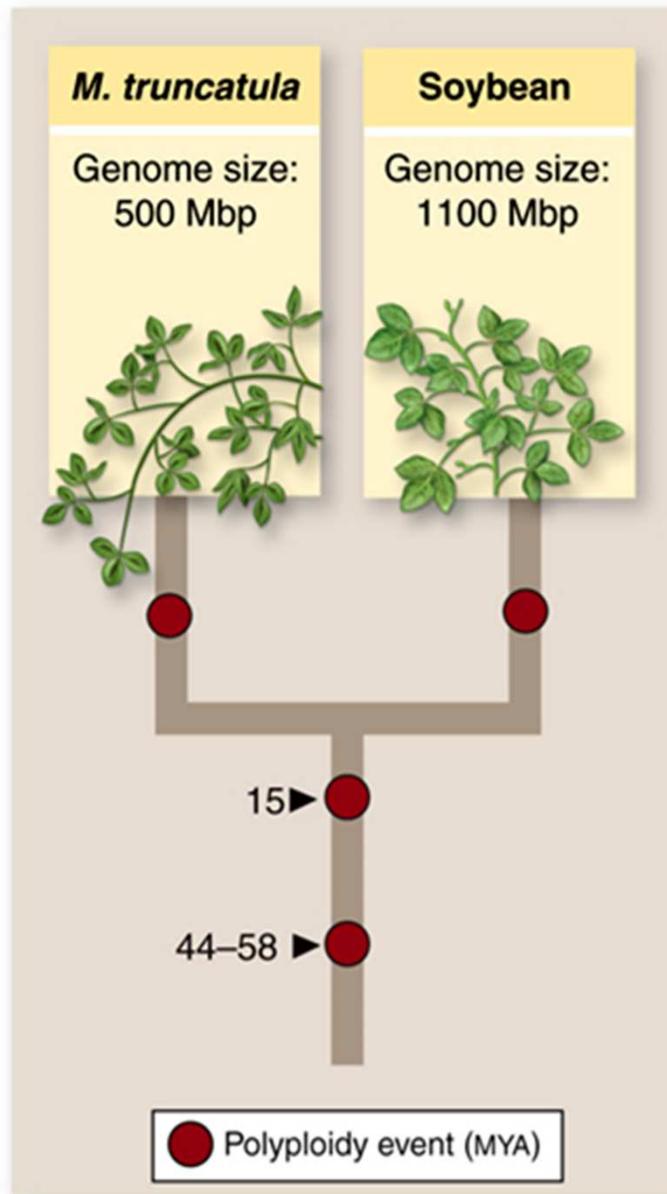
## La poliploidia guida lo studio dell'evoluzione dei genomi

- **Paleopoliploidia:** confronto degli eventi di ploidizzazione
  - Divergenza di sequenze duplicate
  - Presenza o assenza di coppie di geni duplicati in seguito a ibridazione

# Evoluzione dei genomi

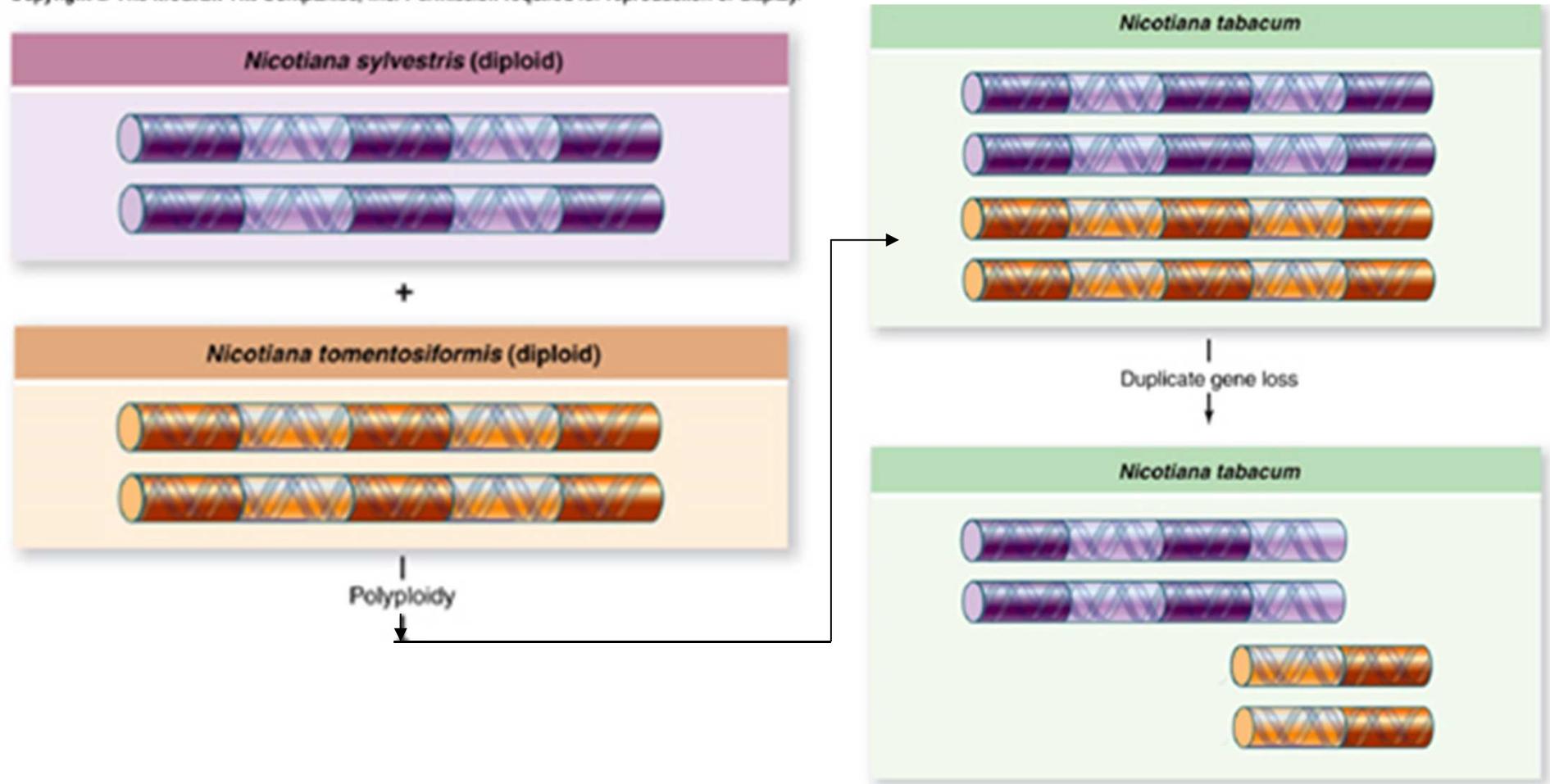


La poliploidia è diffusa nelle piante e ha avuto origini multiple durante l'evoluzione



Riduzione delle dimensioni del genoma

- Destino dei geni duplicati
  - Perdita di funzione per mutazione
  - Nuove funzioni
  - Suddivisione delle funzioni tra le due copie



Perdita di geni duplicati-> problema anche per identificare geni ortologhi in specie diverse

# POLIPLOIDIA E SEQUENZIAMENTO DEI GENOMI

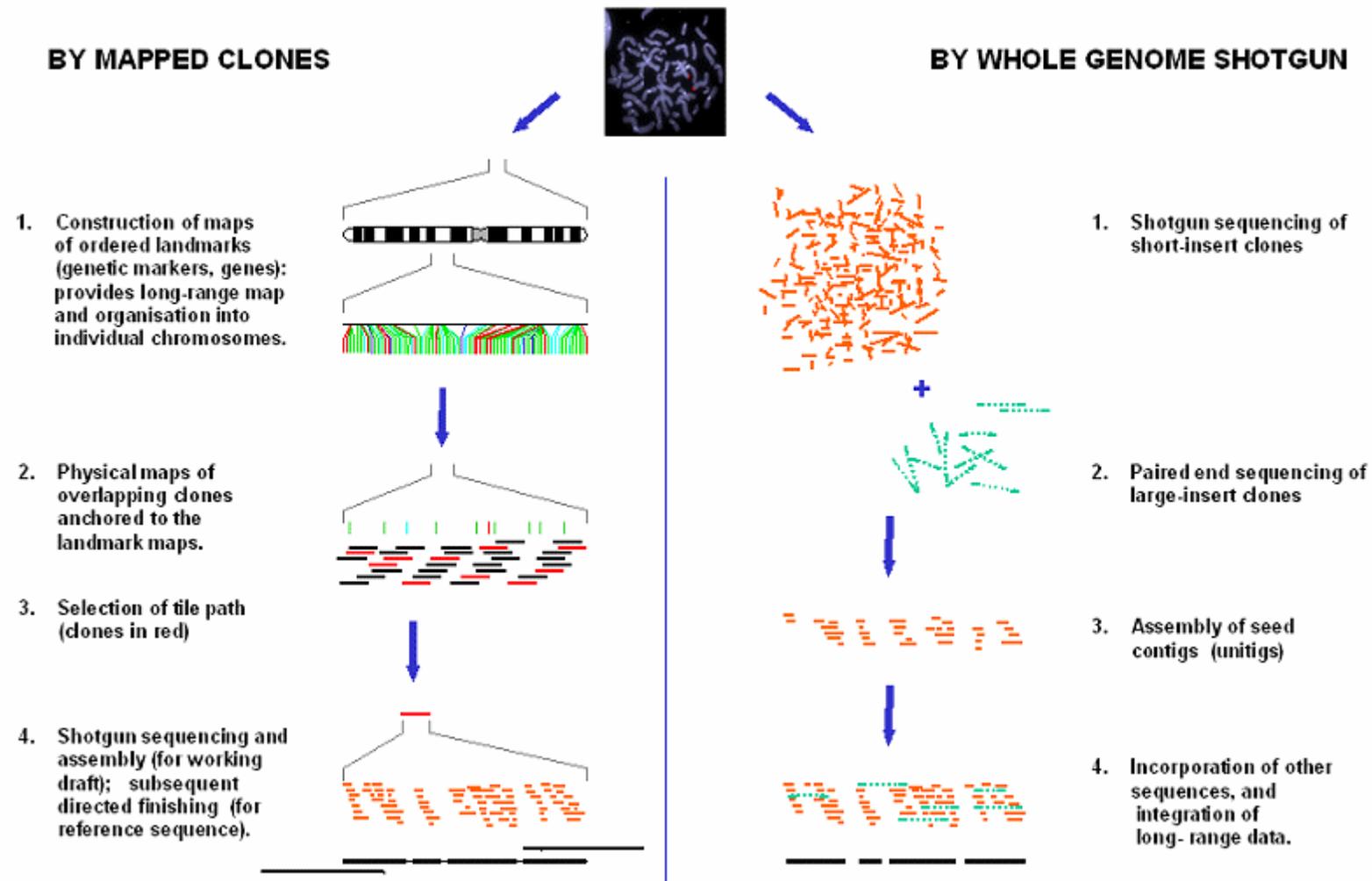
Molte specie autopoliploidi sono intolleranti all'INBREEDING, e hanno alti livelli di eterozigosità, importanti per la produttività

-> problema nell'assemblaggio dei contigs (più alleli diversi per ogni gene)

Negli allopoliploidi i cromosomi duplicati hanno subito sufficiente divergenza per non appaiarsi tra loro -> le sequenze delle coppie geniche sono distinguibili

N.B.: tutte le angiosperme sono PALEOPOLIPLOIDI, ma i geni "paleologhi" sono normalmente ben differenziati

# Sequenziamento “whole-genome shotgun” o “clone-by-clone”?



# “whole-genome shotgun”

## Vantaggi

Rapido  
Meno costoso  
Utile per sequenziare regioni  
refrattarie alla mappatura  
fisica (es. regioni ripetitive)

## Svantaggi

Assemblaggio complicato se ci  
sono molte regioni ripetitive  
In autoploidi, non distingue  
aplotipi diversi di geni identici

# “clone-by-clone”

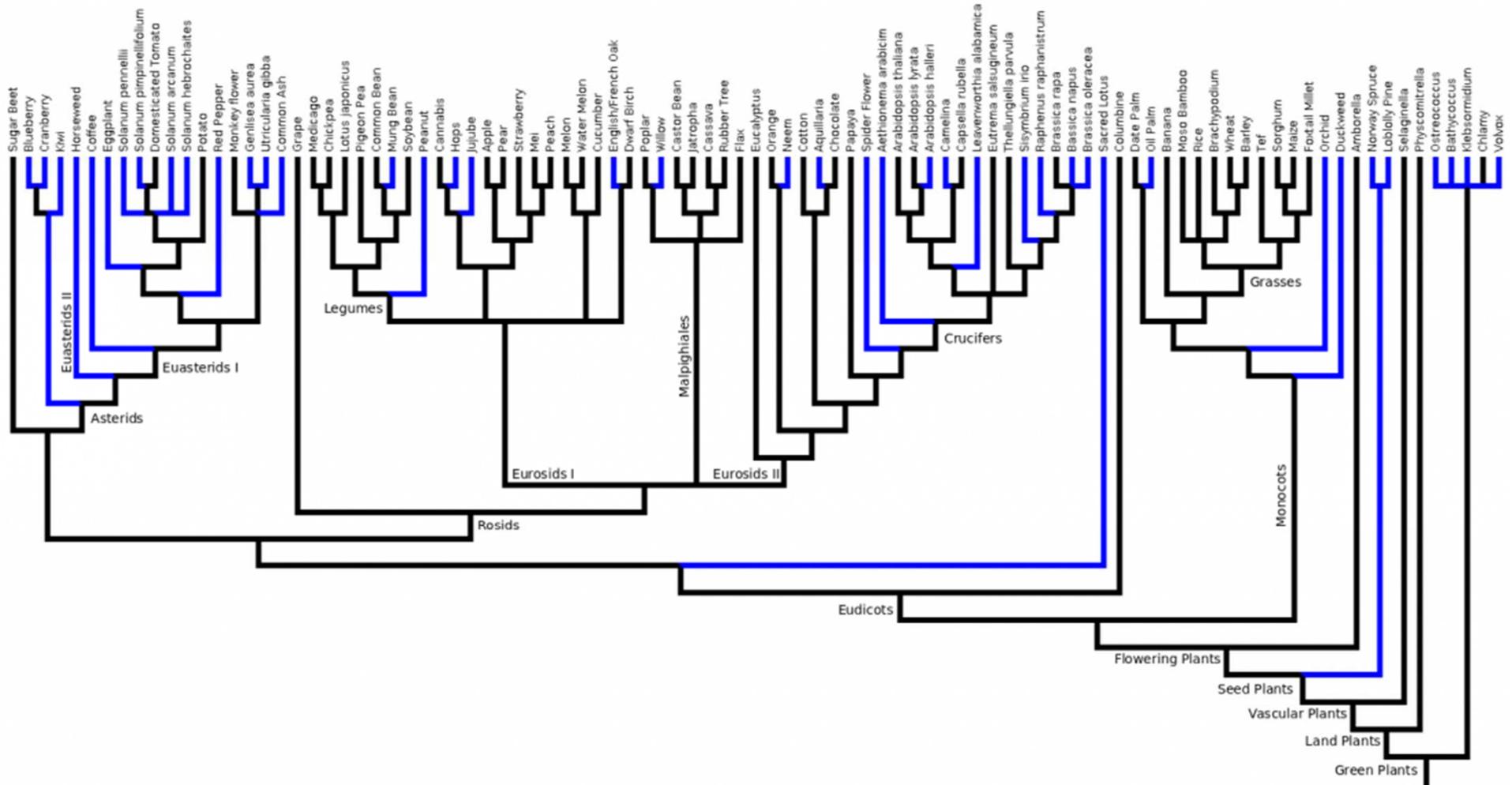
## Vantaggi

Delimita l'incertezza a intervalli  
piccoli (100Kb)  
Un allele alla volta -> no problema  
di eterozigotità

## Svantaggi

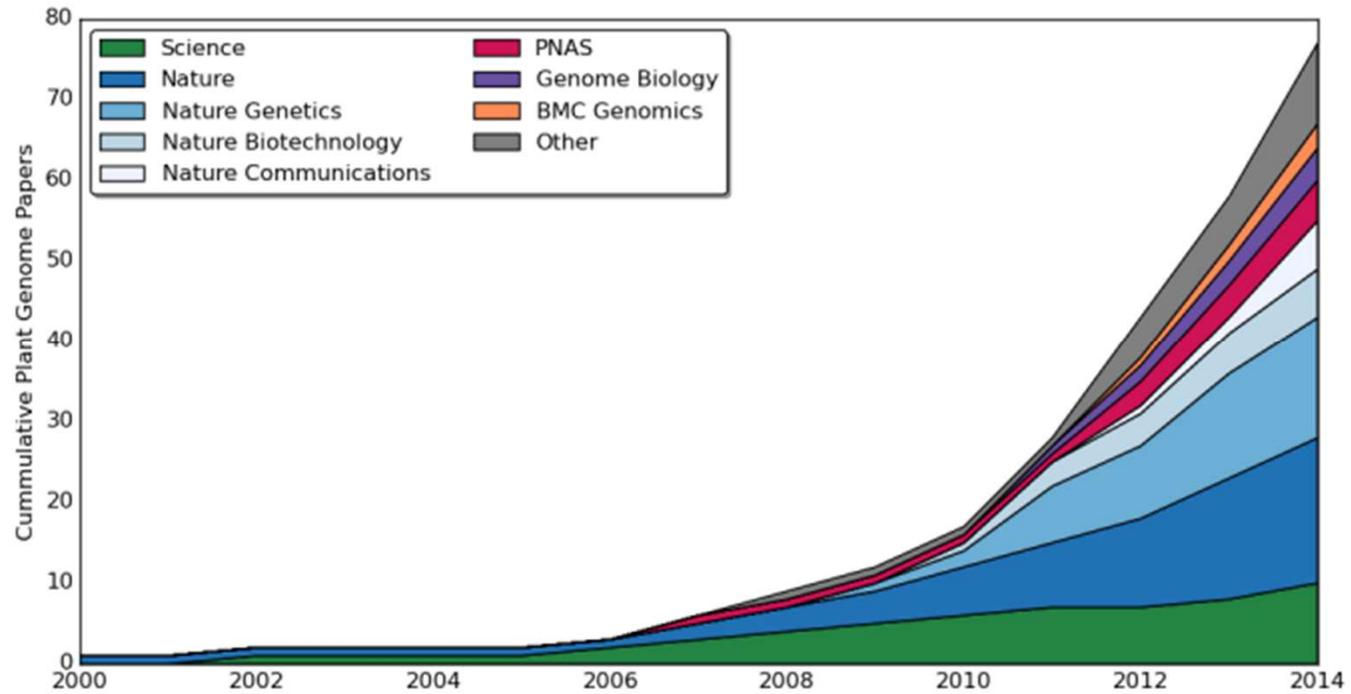
Costo dell'assemblaggio della  
library e dell'ordinamento dei  
contigs

# PROGETTI DI SEQUENZIAMENTO DI GENOMI DI PIANTE COMPLETATI



[https://genomeevolution.org/wiki/index.php/Sequenced\\_plant\\_genomes#Phylogenetic\\_Tree](https://genomeevolution.org/wiki/index.php/Sequenced_plant_genomes#Phylogenetic_Tree)

# PROGETTI DI SEQUENZIAMENTO DI GENOMI DI PIANTE COMPLETATI



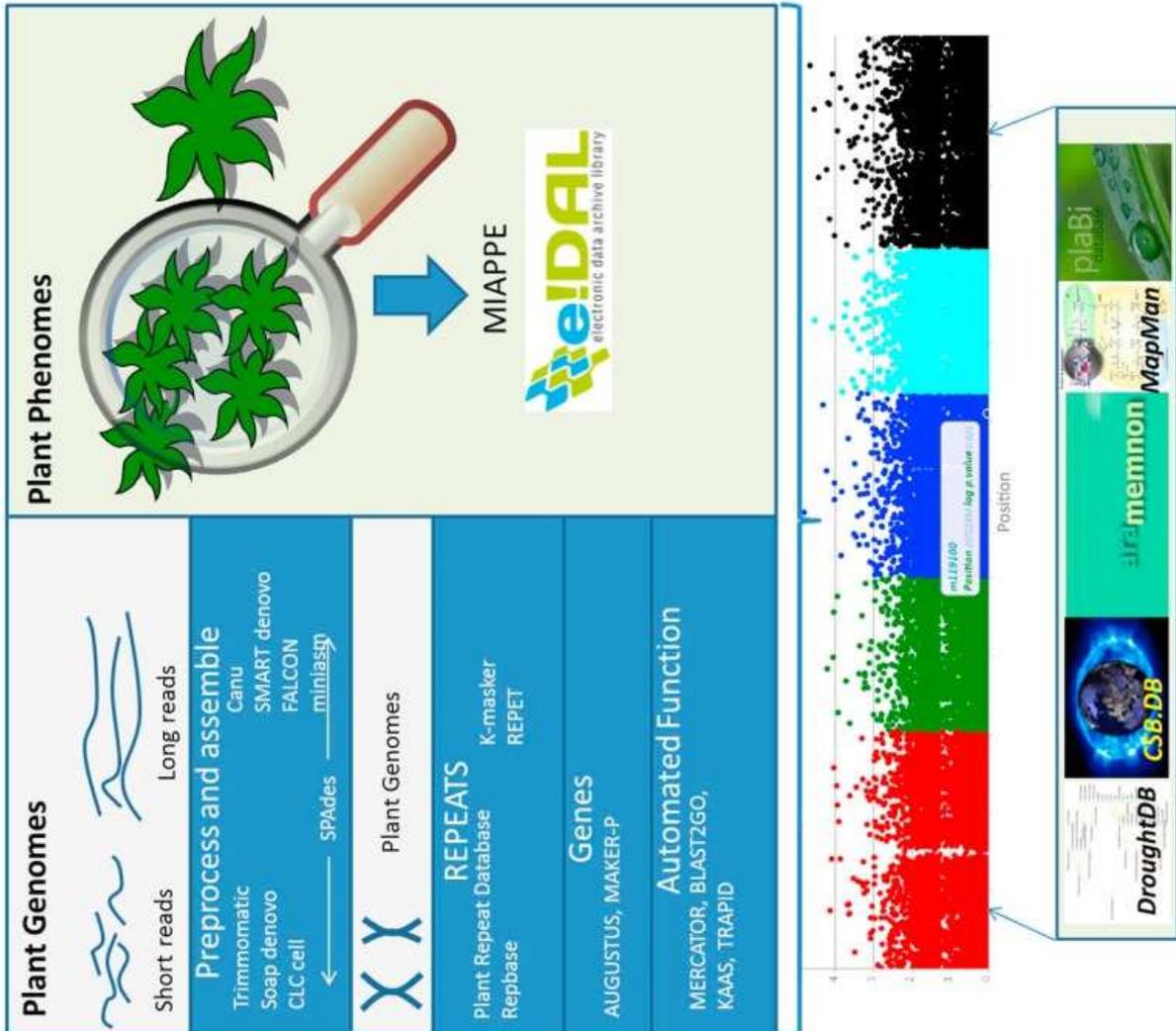
# **PROGETTI DI SEQUENZIAMENTO DI GENOMI DI PIANTE COMPLETATI**

<http://www.plabipd.de/portal/web/guest/angiosperm-phylogenetic-view>

# EST = Expressed Sequence Tags

Creata sequenziando l'estremità 5' e/o 3' di mRNA isolati a caso e convertiti in cDNA (di solito 200–900 nt)

- > veloce e poco costoso
- > scoperta geni nuovi
- > marcatori per mappatura
- > base per futuri progetti di sequenziamento genomico
- > parziale copertura della porzione codificante del genoma



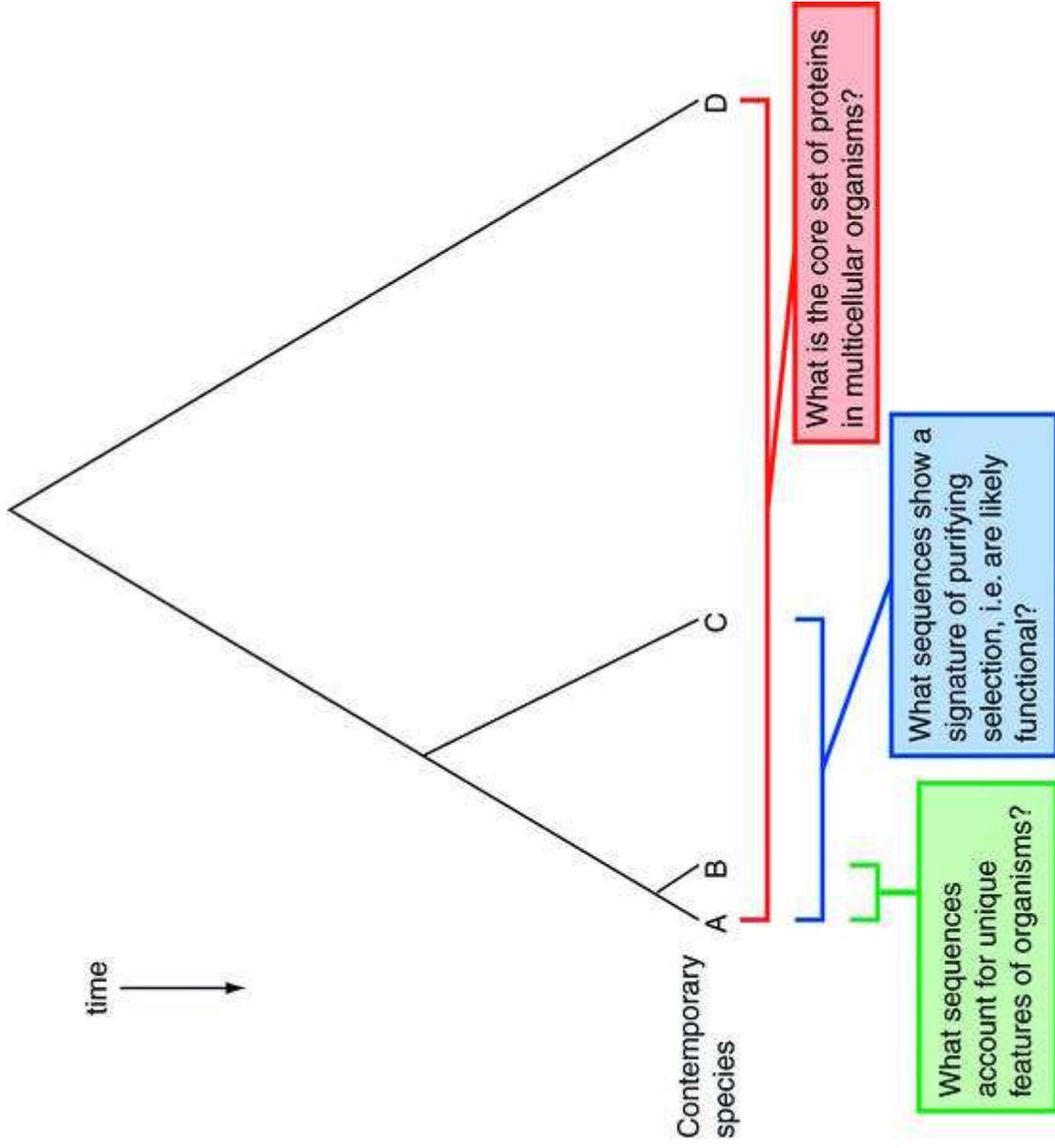
# GENOMICA COMPARATIVA

- Analisi e confronto di genomi di specie diverse
- Fornisce informazioni sull'evoluzione delle specie e sulla funzione di geni e sequenze non codificanti
- Es.: funzione di un gene dedotta dallo studio di geni ortologhi in specie modello

# GENOMICA COMPARATIVA

## Cosa si analizza?

- Similarità di sequenza
- Localizzazione cromosomica dei geni
- Lunghezza e numero esoni
- Quantità di DNA non codificante
- Conservazione di regioni cromosomiche



# Predizione della funzione di un gene a partire dalla sequenza di geni in altre specie

**Gene con funzione ignota**



**Trasferimento di annotazione**

**Specie modello**

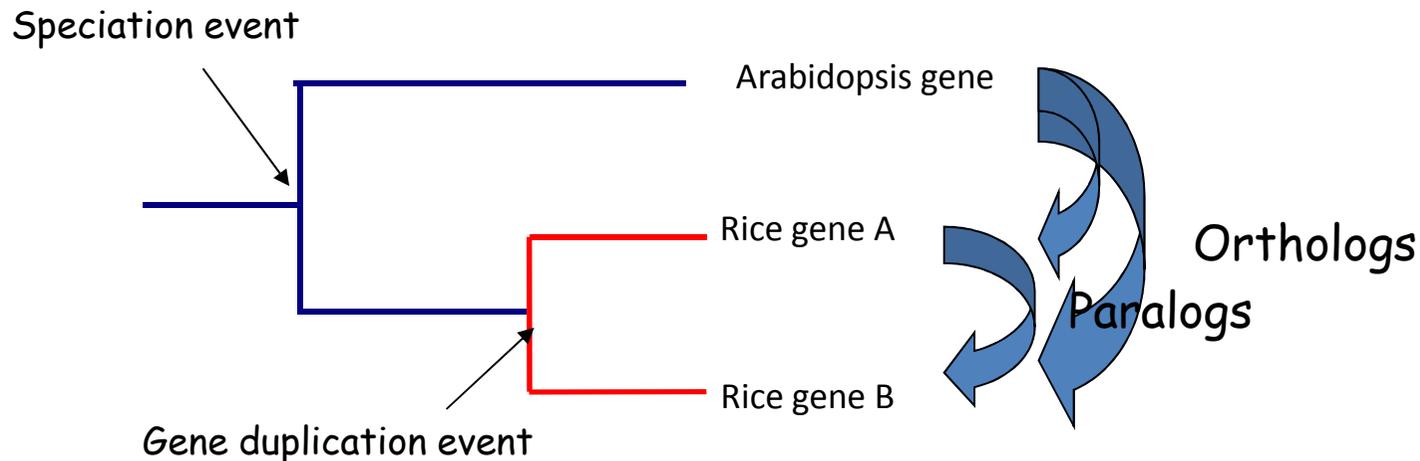


**Geni omologhi**

**Gene con funzione X**

# Geni omologhi

- **Geni ortologhi** sono geni omologhi che discendono dall'ultimo ancestore comune attraverso speciazione
- Molto probabilmente codificano per proteine con funzione simile



- **Geni paraloghi** sono geni omologhi che si sono evoluti per duplicazione e possono codificare proteine con funzioni più divergenti
- **Geni inparaloghi**: geni ortologhi che hanno subito duplicazione

**Come trovare in una specie un gene ortologo ad un gene noto in un'altra specie?**

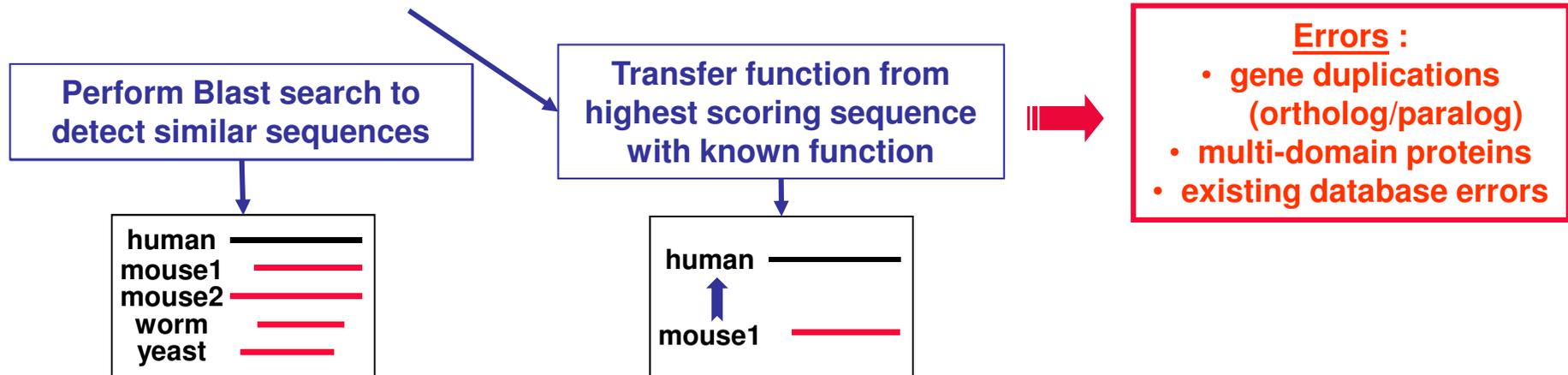
# Come predire l'omologia?

**Similarità e omologia non sono la stessa cosa!**

**Geni simili si assomigliano sulla base di un'osservazione empirica**

**Geni omologhi sono geneticamente correlati (fatto storico: hanno antenato comune)**

# Metodo classico : annotazione funzionale basata sulla somiglianza (Blast)



# Predizione dell'omologia sulla base della similarità

Es. BLAST

## Vantaggi:

- Facile
- Veloce
- Direttamente sul genoma completo

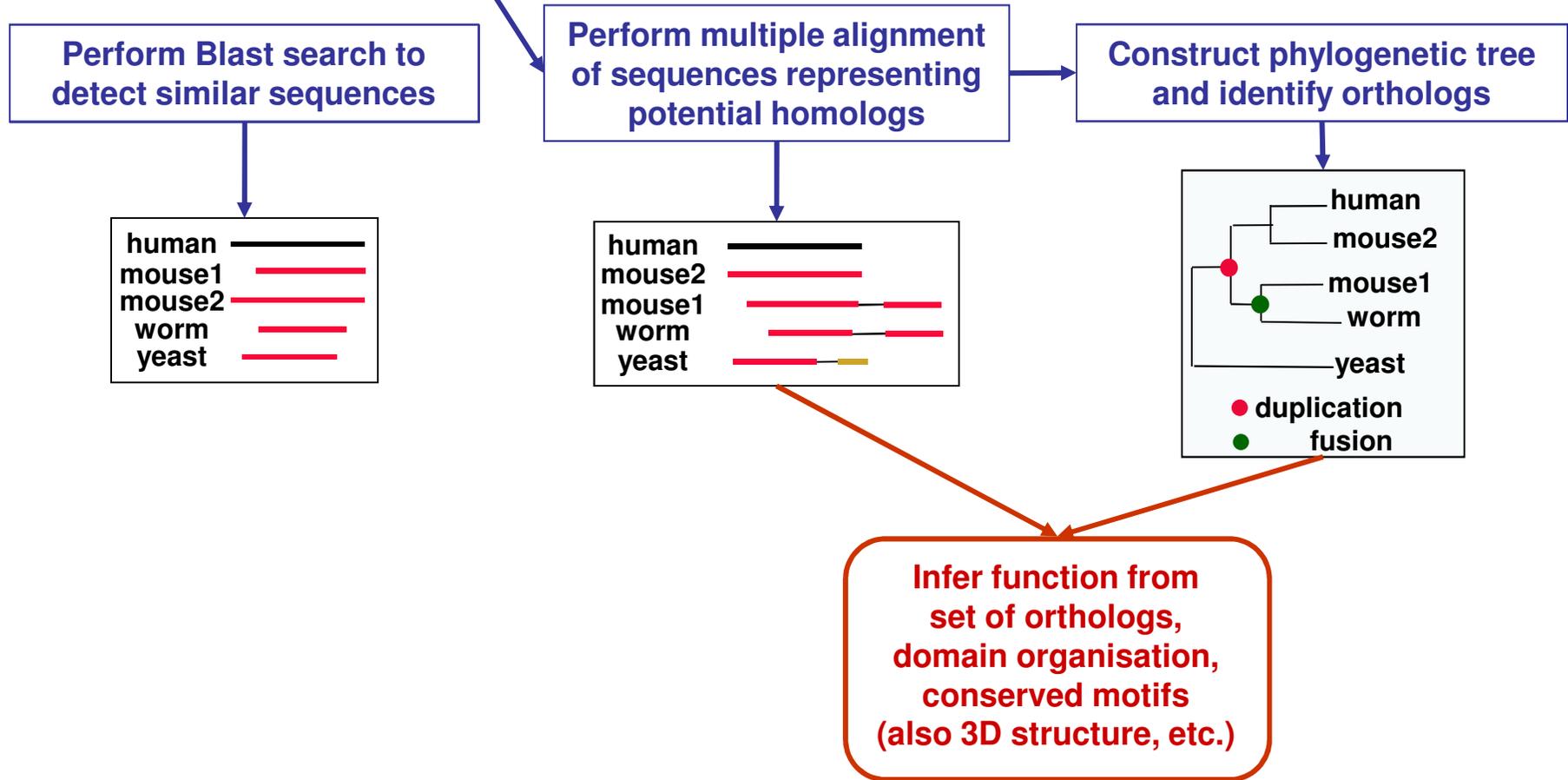
## Svantaggi:

- Come stabilire la soglia di E-value per trasferire l'annotazione del gene da una specie all'altra?

**Due sequenze possono presentare similarità senza essere evolutivamente correlate!**

- Non identifica eventi di duplicazione genica

# Phylogeny-based inference



# Predizione dell'omologia sulla base della filogenesi

## Vantaggi:

- Efficiente per identificare duplicazioni (paraloghi e ortologhi)

## Svantaggi:

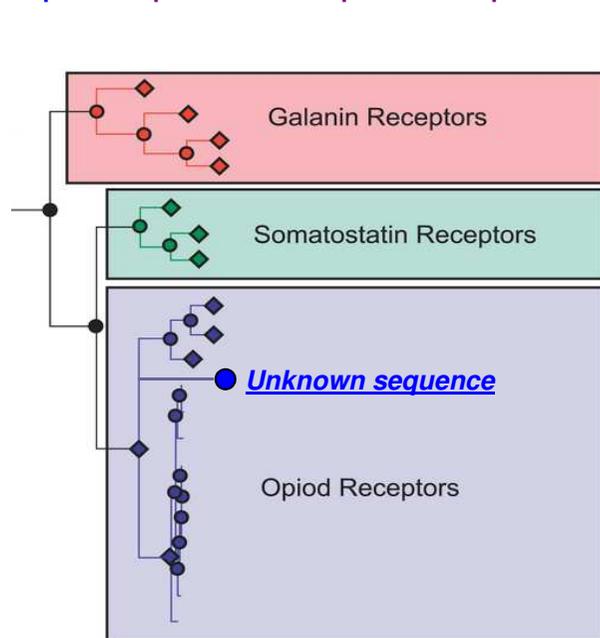
- Lento
- Richiede raggruppamento dei geni in famiglie

## Metodi correnti

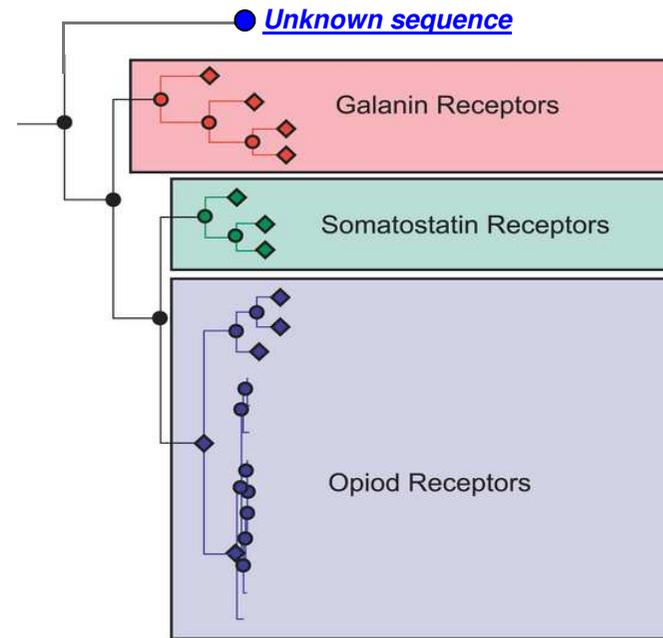
- RIO e Orthostrapper : solo per 1900 famiglie di geni vegetali (Pfam)
- GOST (usa GreenPhylDB family : 6420 famiglie geniche vegetali)

- Tree-based orthology: build a phylogenetic tree of a group of genes and compare gene tree to species tree to define speciation, duplication events
  - Resampled Inference of Orthologs (RIO) (Zmasek and Eddy, 2002)
  - Orthostrapper (Storm and Sonnhammer, 2002)
  - Levels Of Orthology From Trees (LOFT) (Van de Heijden et al, 2007)

■ Example: G protein-coupled receptors



Prediction: Opioid receptor



More general prediction: GPCR of unknown specificity

# GreenPhylDB

A phylogenomic platform for plant comparative genomics



## Due specie modello



- *Oryza sativa* e *Arabidopsis thaliana*
- Genoma completo
- Alta qualità dell'annotazione (TAIR release 7, TIGR release 5)
- Evidenze funzionali disponibili
- In futuro integrerà altre specie

# InParanoid

- <http://inparanoid.sbc.su.se>
- Database per identificare geni ortologhi e inparaloghi tra specie diverse di eucarioti (animali, piante, funghi, protisti)

**<http://www.phytozome.net/>**

**[http://genomeevolution.org/wiki/index.php/Sequenced\\_plant\\_genomes](http://genomeevolution.org/wiki/index.php/Sequenced_plant_genomes)**