

Genomi di pianta sequenziati



Arabidopsis thaliana (Eukaryota)

[back to top](#)

This was the first plant to be sequenced and is considered *the* species for investigating plant genetics. A member of the mustard family, the plant is popular among researchers because it grows in small spaces, lives about six weeks, and has a small genome.

» Sequenced by: [The Arabidopsis Genome Initiative](#) [Abstract](#)

» Related GNN articles:

[Paranoid but Popular: Mutant mouse-ear cress offers insight into natural plant resistance](#)

[Clickable genomics: Plans for virtual plant posted](#)

[SHATTERPROOF genes in *Arabidopsis* are good news for agriculture](#)

[What makes plants grow? The *Arabidopsis* genome knows](#)

» Image: Peggy Greb/USDA.



Oryza sativa (Eukaryota)

[back to top](#)

A food staple for much of the world's population, rice comes in different varieties. Two strains were sequenced in 2002, the *japonica* (popular in Japan) and the *indica* (grown in China). An international consortium is working on a third rice genome sequence that will be the gold standard.

» Sequenced by: [Syngenta and Myriad Genetics](#) *O. sativa* L. ssp. *indica* [Abstract](#)

[Syngenta](#) *O. sativa* L. ssp. *japonica* [Abstract](#)

» Related GNN article:

[Two Groups Sequence Rice: Combining draft sequences may accelerate completion of finished genome](#)

» Image: Photo by Ma Liwen, Courtesy of Qiu BaoXing. (*Science*)

2000: inizia l'era della genomica funzionale.
Il primo genoma di una pianta superiore è
sequenziato completamente

articles

Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*

The Arabidopsis Genome Initiative*

** Authorship of this paper should be cited as 'The Arabidopsis Genome Initiative'. A full list of contributors appears at the end of this paper*

Arabidopsis thaliana

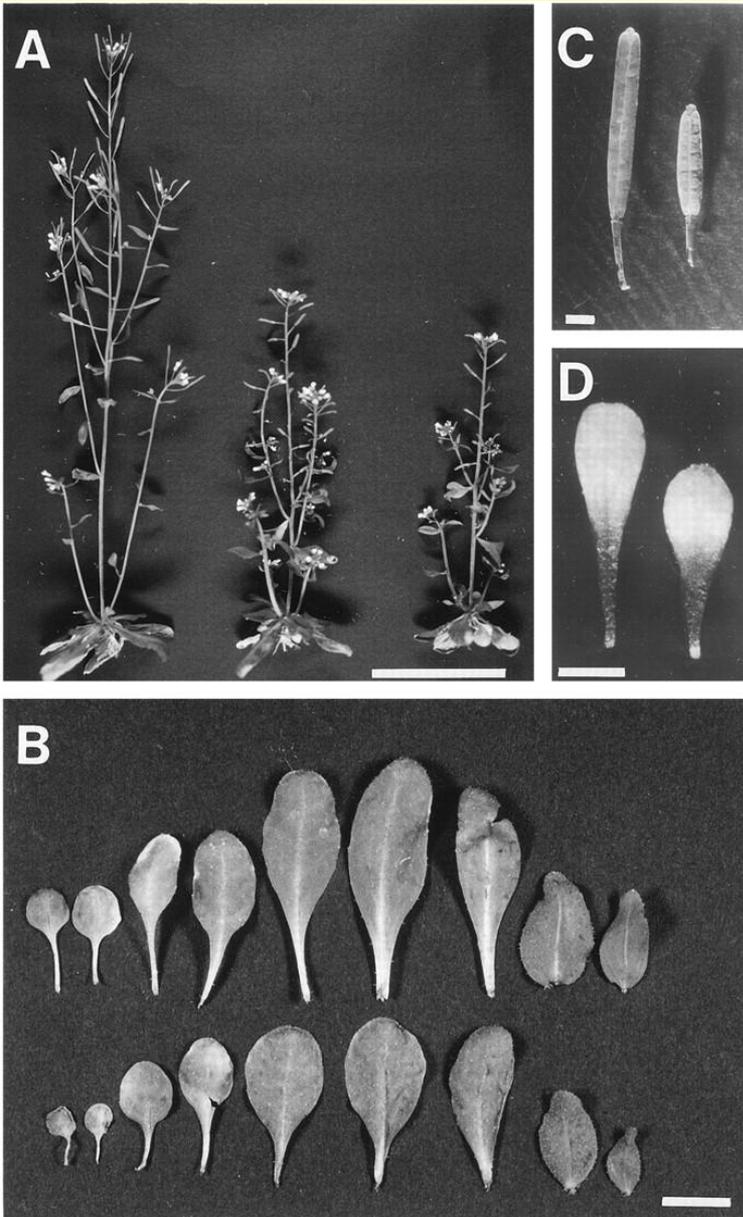


★ *Arabidopsis thaliana* è una piccola angiosperma che viene usata come organismo modello nella biologia vegetale.

★ *Arabidopsis* è un membro della famiglia delle Brassicaceae

★ Questa pianta non ha un'importanza agronomica, ma offre importanti vantaggi per quel che riguarda la ricerca di base, in particolare per l'attribuzione di funzioni ai geni

I vantaggi apportati dall'uso di *Arabidopsis thaliana*:



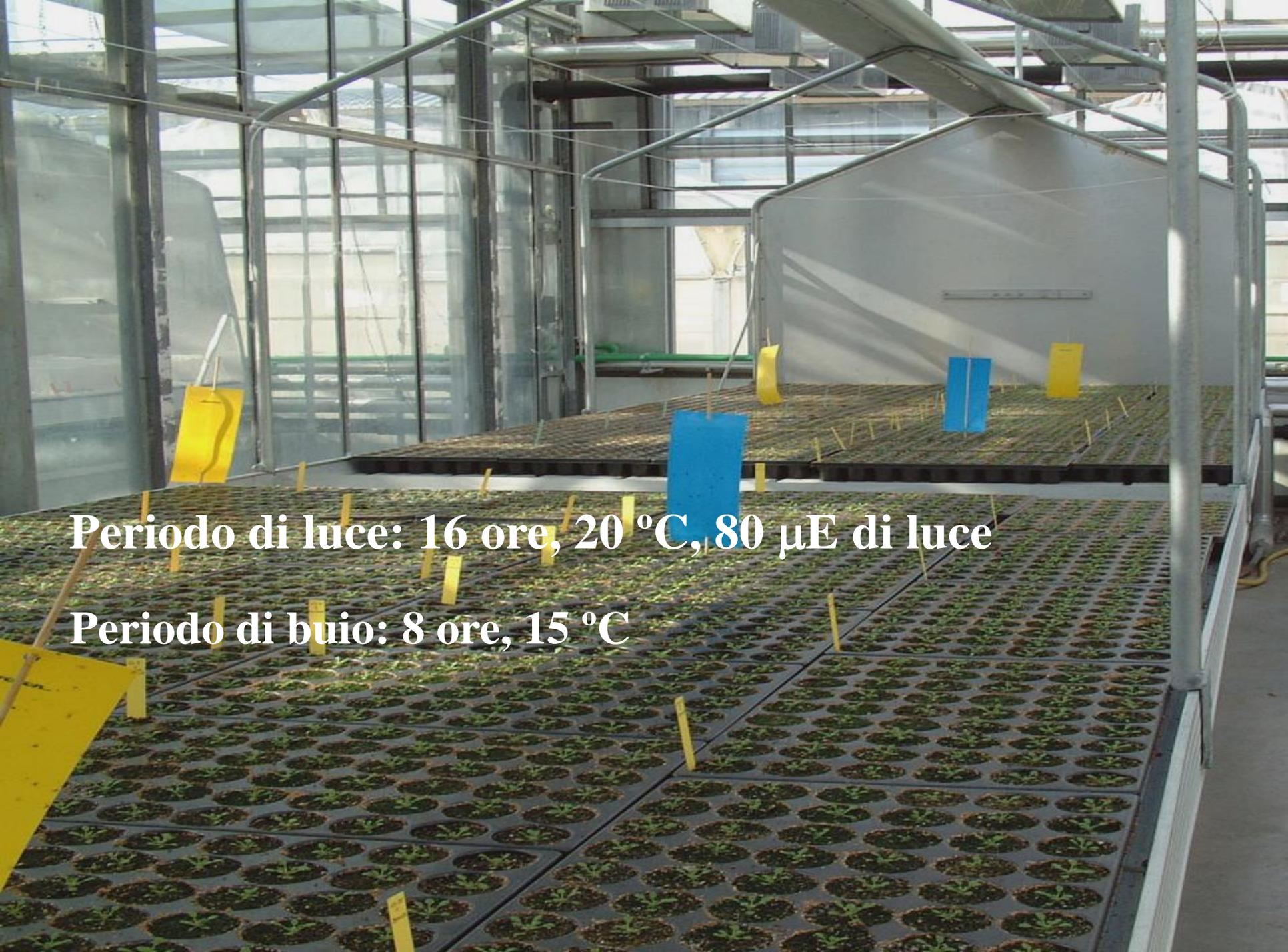
- Genoma piccolo distribuito su 5 cromosomi (114.5 Mb/125 Mb total) e diploide

- Un ciclo vitale molto rapido (circa 6 settimane dalla germinazione al seme maturo)

- E' in grado di produrre molti semi e non richiede molto spazio per essere coltivata

- Può essere trasformata utilizzando la tecnologia dell'Agrobatterio

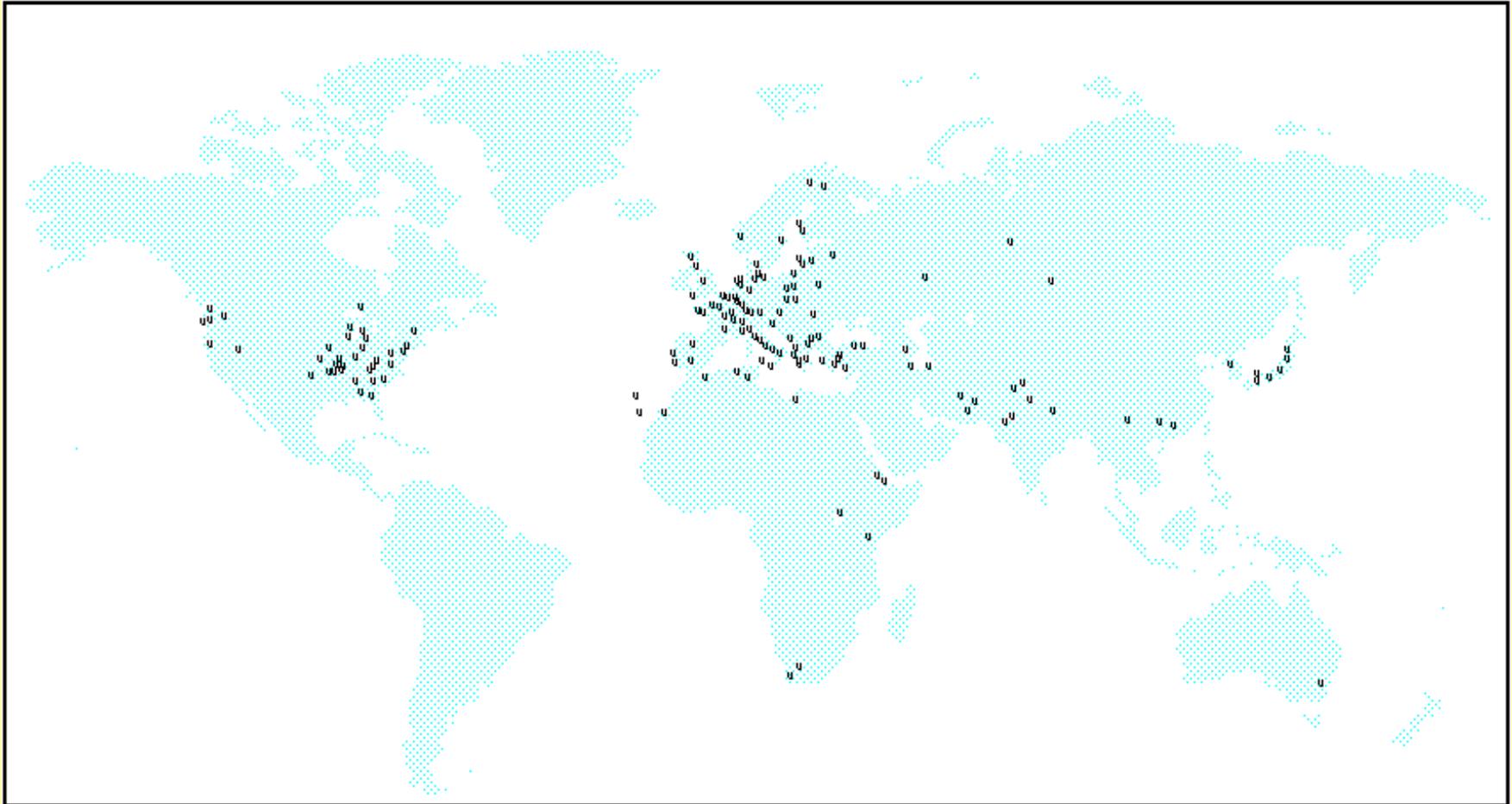
- Disponibilità di mutanti!!



Periodo di luce: 16 ore, 20 °C, 80 μ E di luce

Periodo di buio: 8 ore, 15 °C

Distribuzione geografica delle sottospecie di *Arabidopsis*



Geographical distribution of ecotypes of *Arabidopsis thaliana* (L.) HEYNH.

© Jonathan Clarke

Ampia variabilità raccolte circa 300 varietà, molto utili per la genomica funzionale, comprensione a livello molecolare di molte caratteristiche fenotipiche

▲ *Arabidopsis thaliana* è un'importante sistema modello per identificare geni e determinare la loro funzione. (...).

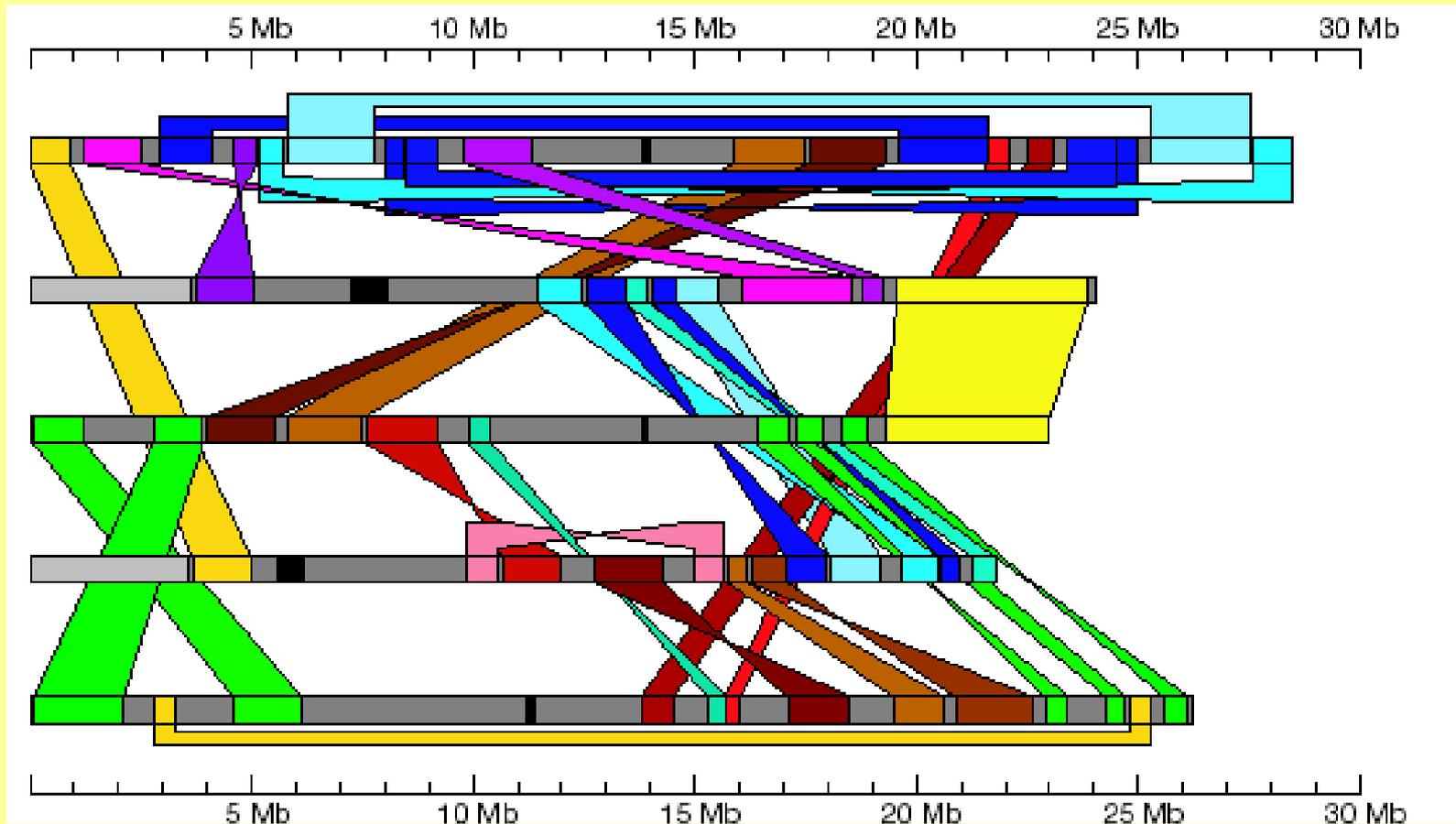
▲ La regione sequenziata comprende 115.4 Mbp delle 125 Mbp che costituiscono il suo genoma. (...).

▲ Il genoma contiene 25,498 geni che codificano proteine appartenenti a 11.000 famiglie, mostrando una diversità funzionale comparabile a quella di *Drosophila* e *C.elegans* (...).

▲ è il primo genoma sequenziato di una pianta e fornisce le basi per comparare processi conservati in tutti gli organismi eucarioti.

▲ Permetterà di identificare le funzioni di geni specifici per le piante e individuare geni importanti per il miglioramento delle piante di valore commerciale

Duplicazioni segmentali



Duplicazioni segmentali intercromosomiche (60 %)

L'attorcigliamento indica che l'evento di duplicazione è accompagnato da un'inversione

Le regioni duplicate non contengono necessariamente le stesse copie di geni....
(mutazioni, delet. ecc.)

Alcune caratteristiche dei tre genomi presenti in *Arabidopsis*

Table 4 General features of genes encoded by the three genomes in *Arabidopsis*

	Nucleus/cytoplasm	Plastid	Mitochondria
Genome size	125 Mb	154 kb	367 kb
Genome equivalent/cell	2	560	26
Duplication	60%	17%	10%
Number of protein genes	25,498	79	58
Gene order	Variable, but syntenic	Conserved	Variable
Density (kb per protein gene)	4.5	1.2	6.25
Average coding length	1,900 nt	900 nt	860 nt
Genes with introns	79%	18.4%	12%
Genes/pseudogenes	1/0.03	1/0	1/0.2–0.5
Transposons (% of total genome size)	14%	0%	4%

Il manifesto sottoscritto dai maggiori ricercatori coinvolti nel settore della biologia vegetale

Editor's Choice

National Science Foundation-Sponsored Workshop Report: "The 2010 Project"

Functional Genomics and the Virtual Plant. A Blueprint for Understanding How Plants Are Built and How to Improve Them

Joanne Chory¹, Joseph R. Ecker¹, Steve Briggs, Michel Caboche, Gloria M. Coruzzi, Doug Cook, Jeffrey Dangl, Sarah Grant, Mary Lou Guerinot, Steven Henikoff, Rob Martienssen, Kiyotaka Okada, Natasha V. Raikhel, Chris R. Somerville, and Detlef Weigel

MISSION STATEMENT

To exploit the revolution in plant genomics by understanding the function of all genes of a reference species within their cellular, organismal, and evolutionary context.

Il progetto 2010 ha lo scopo di comprendere la funzione di tutti i 25,000 geni identificati nel genoma di Arabidopsis...

Il fine ultimo del progetto è conoscere ogni aspetto molecolare dello sviluppo di una pianta, al punto tale da poter simulare la crescita di una pianta virtuale...

...Alla fine saremo in grado di predire la funzione di geni appartenenti a specie di interesse agronomico, attraverso il confronto delle loro sequenze

Quando la funzione di tutti i geni sarà nota saremo in grado di:

1. predire gli effetti di una qualsiasi modificazione genetica

2. modificare a nostro vantaggio specie selvatiche

3. Mantenere ed espandere l'insieme dei germoplasmi

4. comprendere i meccanismi alla base del fenomeno indicato come Eterosi

5. comprendere le basi molecolari alla base della plasticità dei fenotipi

6. conoscere il minor numero di geni necessari alla vita della pianta

7. comprendere le basi genetiche dell'evoluzione

8. comprendere le basi molecolari delle interazioni tra le piante ed altri organismi

1. Predictable outcomes to directed experimental genetic changes.
2. Directed genetic changes that accelerate domestication of wild species.
3. Facile genetic manipulation that ensures maintenance of, and expansion of germplasm bases.
4. A description of the underlying mechanisms of heterosis, and the ability to use this phenomenon more effectively.
5. Enhanced understanding of the genetic basis of phenotypic plasticity, which will have a profound impact not just in plants, but also in animals, including humans.
6. Knowledge of the minimum gene set required for plant life.
7. Understanding of the genetic basis of plant evolution, which will enrich our understanding of the diversity of life on earth.
8. An understanding of interactions between plants and other organisms in their environment, up to the level of ecosystems.

CONFRONTO TRA ACCESSIONI DI ARABIDOPSIS

COLUMBIA

LANDSBERG ERECTA

cambiamenti nella microstruttura dei genomi, frequenti i polimorfismi in regioni codificanti e non codificanti



CONFRONTO TRA ARABIDOPSIS E ALTRE BRASSICACEAE

Arabidopsis

Capsella rubella



Contenuto simile di geni e stessa organizzazione del genoma (duplicazioni)

Allineamenti di frammenti genomici e EST

Arabidopsis

Brassica oleracea



Maggiori divergenze, la poliploidia in Brassica ha portato ad una rapida evoluzione del genoma

IDENTIFICARE LA FUNZIONE DI UN GENE

La vastità del numero di sequenze oggi disponibili ha spostato l'interesse della ricerca verso l'identificazione delle funzioni geniche.

In un genoma completamente sequenziato non si sa, dalla sequenza semplice, quali siano i geni e quali siano le loro funzioni.

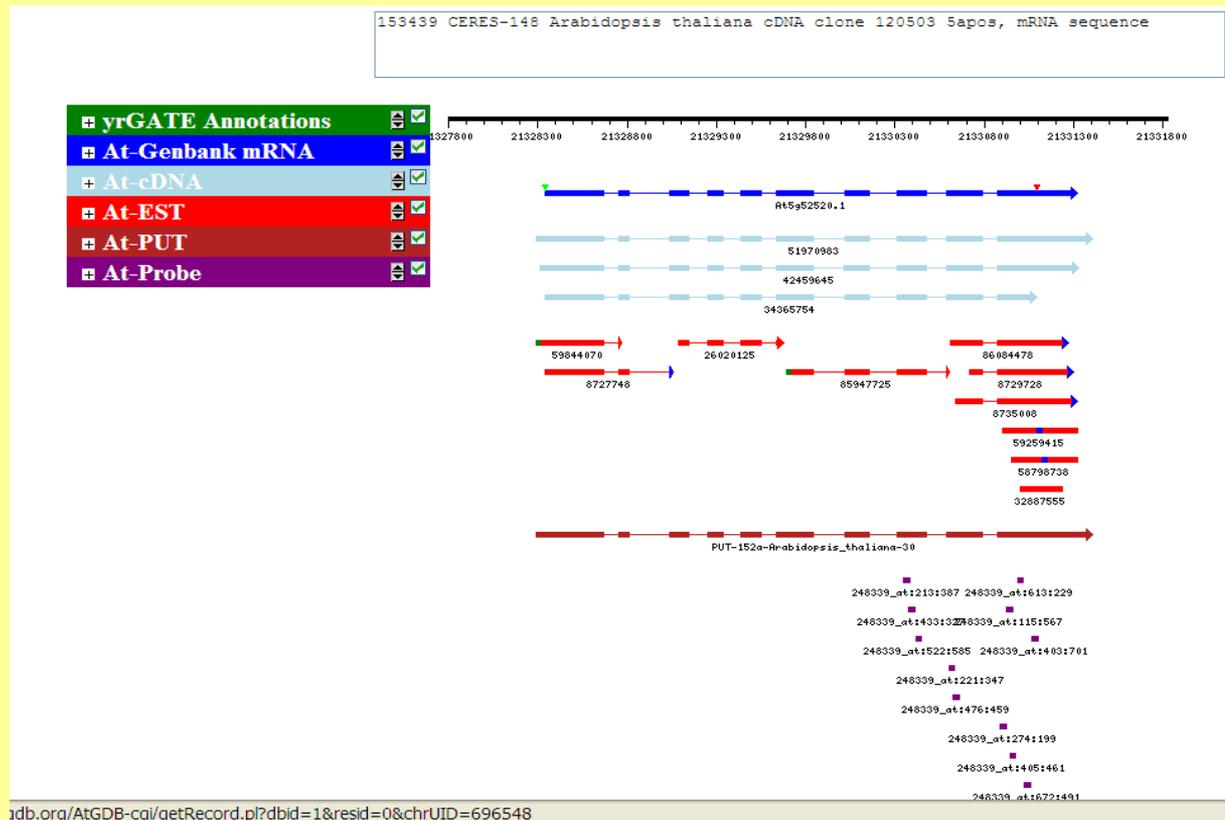
Lo scopo finale della genomica funziona è la definizione del rapporto esistente tra gene e fenotipo:



non è sufficiente definire la funzione biochimica di ciascuna sequenza proteica ma è comunque necessario definire le interazioni tra i diversi prodotti genici

Come si individua un gene e la proteina da esso codificata all'interno di una sequenza genomica

- Si utilizza la sequenza genomica per individuare le molecole di cDNA o anche ESTs (Expressed sequence tags) corrispondenti

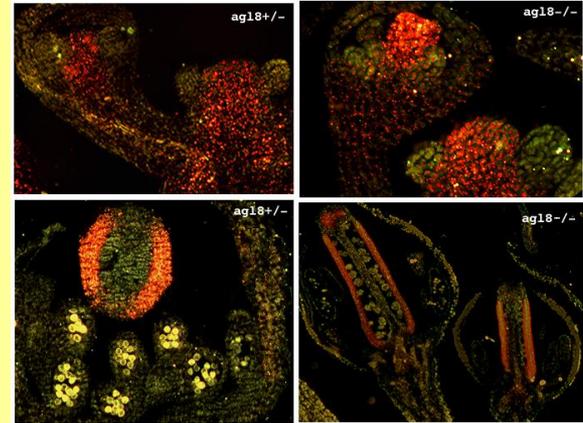
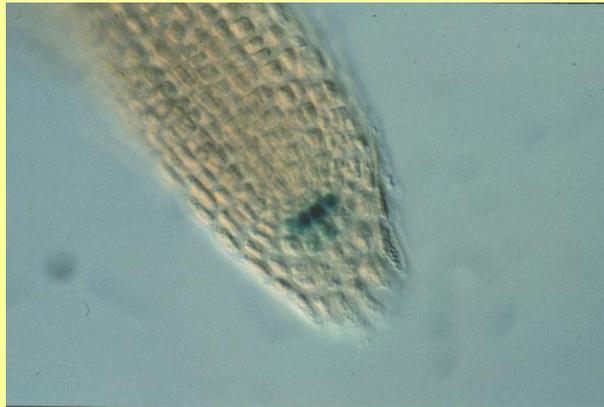


Le EST non "vedono" i geni rari

Guard cell
specific
expression



Root columella
specific
expression



MADS box gene *AGL8*
expression is found in the
carpel wall and inflorescence
meristem...

Come si individua un gene e la proteina da esso codificata all'interno di una sequenza genomica

- Si utilizza la sequenza genomica per individuare le molecole di cDNA corrispondenti
- Si utilizza la sequenza della molecola di cDNA per suddividere la sequenza genomica in esoni (sequenze codificanti) ed introni (sequenze non codificanti)



Come si individua un gene e la proteina da esso codificata all'interno di una sequenza genomica

- Si utilizza la sequenza genomica per individuare le molecole di cDNA corrispondenti

- Si utilizza la sequenza della molecola di cDNA per suddividere la sequenza genomica in esoni (sequenze codificanti) ed introni (sequenze non codificanti)

- All'interno della sequenza di cDNA si individuano il codone di inizio della traduzione (AUG) e il codone di stop (UAA, UAG, UGA) cioè la CDS

- La traduzione della sequenza CDS risulterà in una sequenza a.a. Il confronto della sequenza aminoacidica con altre sequenze proteiche può permettere l'individuazione della funzione della proteina in esame

In Arabidopsis

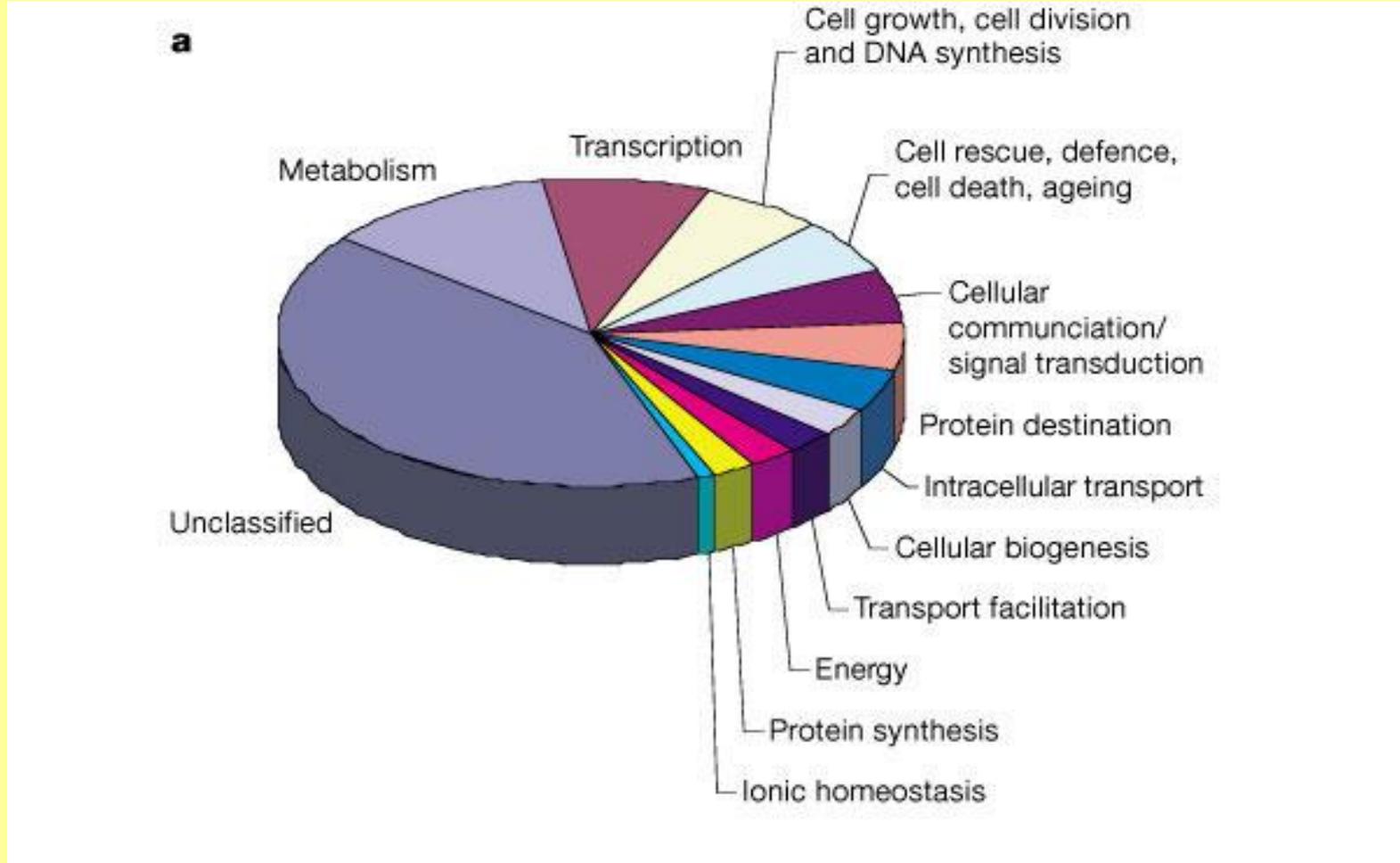
La funzione del 69% dei geni è stata identificata per 'omologia' o 'similarità' con proteine a funzione nota

Solo il 9% è stato caratterizzato sperimentalmente

Molti fattori di trascrizione in Arabidopsis hanno avuto un'evoluzione indipendente da quella di altri eucarioti

Il genoma presenta più omologia con quello di organismi pluricellulari che non con il lievito (trasduzione del segnale, comunicazione cellulare)

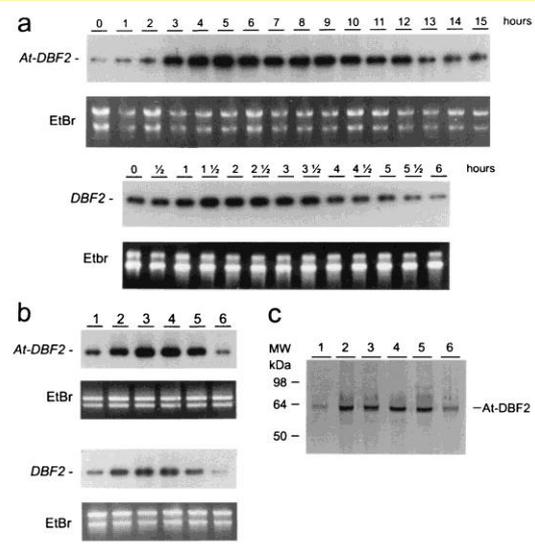
Classificazione dei geni di Arabidopsis in categorie funzionali



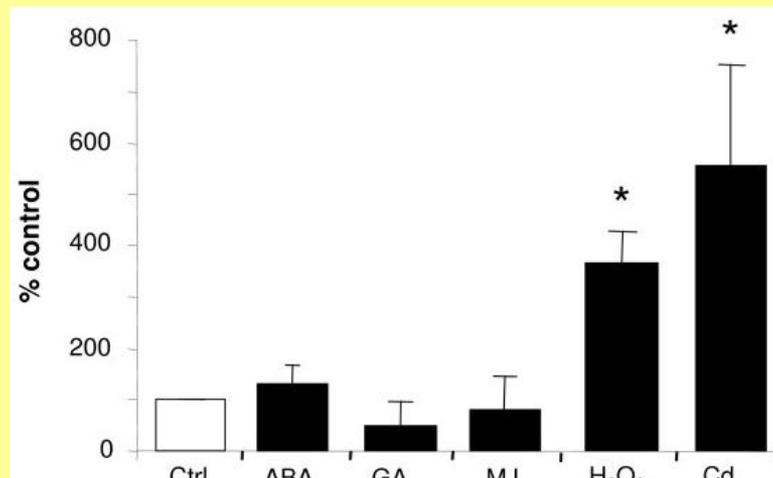
Quasi la metà dei geni di Arabidopsis sono stati classificati come codificanti per proteine sconosciute.

Come si identifica la funzione di un gene codificante una proteina annotata come sconosciuta:

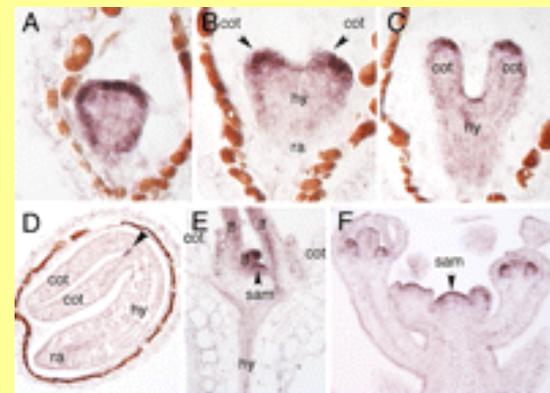
- ➔ Trascrizione del gene (a quale stadio? in quale tessuto? da quale stimolo?)
- ➔ E' ragionevole pensare che geni regolati in modo simile siano coinvolti in processi simili
- ➔ Se l'espressione di alcuni geni è attivata da stress di tipo biotico o abiotico si può pensare che quei geni siano coinvolti ad esempio nella difesa dall'attacco di un patogeno, o nella resistenza alle basse o alte T, siccità, ristagno, salinità ecc.



Northern analysis



Real-time PCR



In situ hybridization

Come si identifica la funzione di un gene codificante una proteina annotata come sconosciuta:

- ➔ La proteina (quanto è abbondante? In quale compartimento cellulare si trova?)
- ➔ Non sempre esiste una correlazione lineare tra quantità di trascritto e quantità di proteina. **Conoscere la quantità della proteina è molto più informativo**
- ➔ Le proteine sono inoltre localizzate in diversi compartimenti cellulari
La localizzazione intracellulare è essenziale per definire la sua funzione.
- ➔ La Proteomica consente di identificare eventuali modificazioni post-traduzionali permettendo di individuare meccanismi regolativi dell'attività proteica

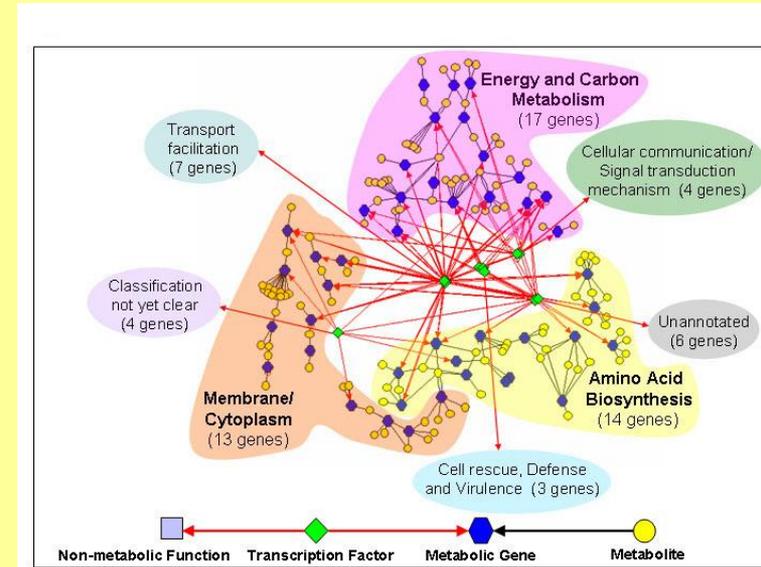
Come si identifica la funzione di un gene codificante una proteina annotata come sconosciuta:

➔ I metaboliti, in quanto prodotti finali dell'attività delle proteine e quindi dei geni, possono essere considerati come una sorta di collante tra il gene ed il fenotipo

➔ Il regno vegetale è in grado di produrre una quantità enorme di metaboliti che oscilla tra le 90,000 e 200,000 molecole, ognuna con le proprie caratteristiche fisico-chimiche

➔ Analizzare le possibili variazioni nella composizione dei metaboliti a seguito di una mutazione o di un qualsiasi tipo di stress è essenziale per l'attribuzione della funzione a determinati geni

➔ La **METABOLOMICA** è la tecnologia deputata a rispondere a questi quesiti



L'omologia non è sufficiente per conoscere la funzione di tutti i geni

tracciare un percorso dal gene alla sua funzione

La dimostrazione ultima del rapporto tra un gene ed il suo fenotipo può essere ottenuto solo tramite l'uso di mutanti

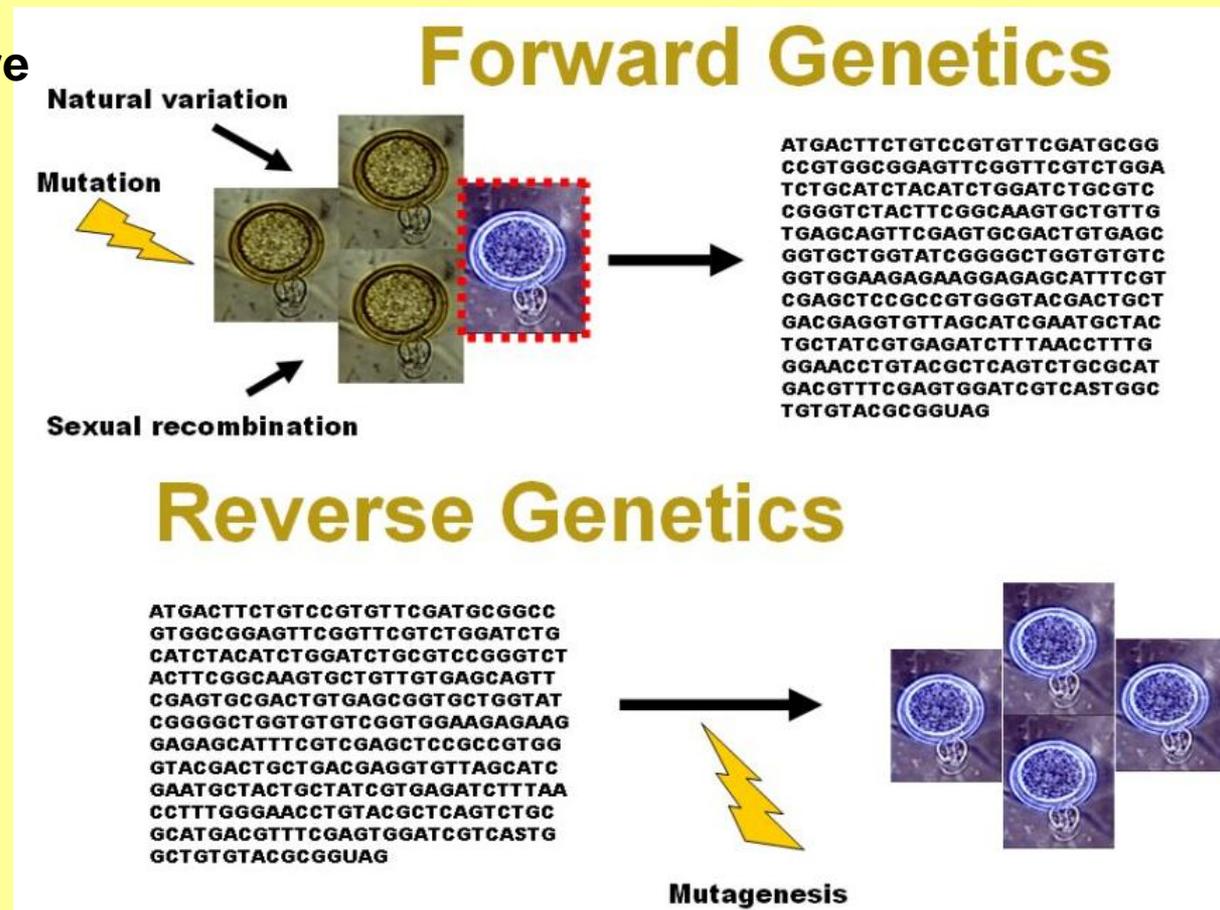
Dal fenotipo per caratterizzare le differenze genetiche

quale è il genotipo?", cioè quale è la sequenza del gene mutante che causa questo fenotipo mutante?

Dalla sequenza del gene alla sua funzione

si parte dal gene di cui si vuole studiare la funzione

"Quale è l'alterazione fenotipica che ne deriva dalla mutazione?"



Genetica Forward (classica)

1. Cercare mutanti che manchino di qualche processo biologico
2. Clonare il gene responsabile per la mutazione usando metodi genetico-molecolari (per esempio map based cloning)
3. Studiare il gene ed il suo prodotto in dettaglio.

Genetica inversa

Studio il gene di interesse (favourite gene FG)

Analisi dei Knockout :

1. Trovare un mutante knockout in un gene
2. Analizzare il mutante per vedere se e dove sta il difetto
3. correlare il difetto (di sequenza?) con i processi biologici

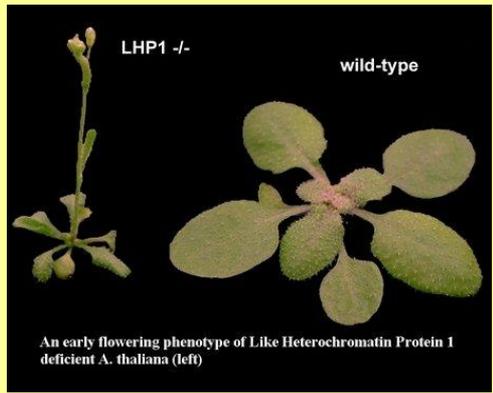
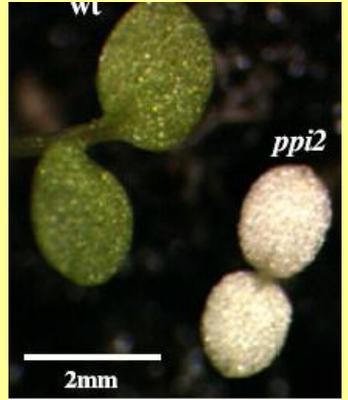
Analisi della sovraespressione ed espressione ectopica

1. Sovraesprimere il FG
2. Correlare il difetto con i processi biologici

Analisi funzionale per mezzo dell'inattivazione genica (reverse genetics)

i maggiori sforzi della genomica sono rivolti allo sviluppo di potenziali mutanti ed alla analisi delle varianti naturali

le basi genetiche di un fenotipo si studiano cercando mutanti in cui quel fenotipo sia alterato



enorme collezione di mutanti

Come ottenere mutanti per la **reverse** genetics

1. Mutazione naturale

2. Mutagenesi random

A. ethylmethane sulfonate (EMS): transizioni $G > A$

TILLING

B. **insertion elements** (T-DNA-tagging o trasposon-tagging):
causano la perdita di funzione del gene interrotto

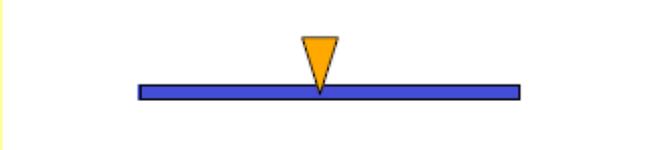
C. activation T-DNA tags: 4x 35S enhancer

in aggiunta alla perdita di funzione, può essere usato per
identificare gain of function.

D. X-ray irradiation

E. metodi basati su RNA interference (VIGS)

La mutazione causata dal trasposone:

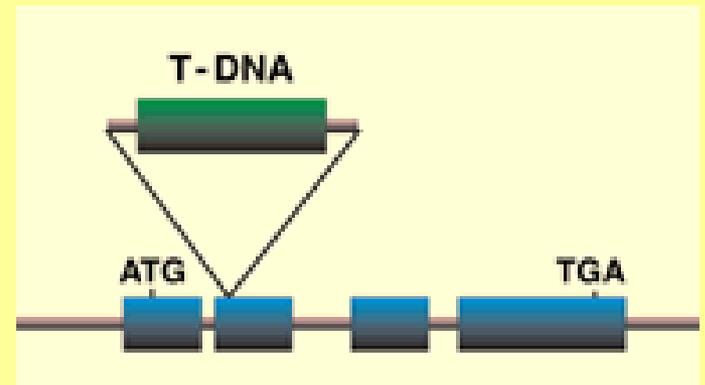
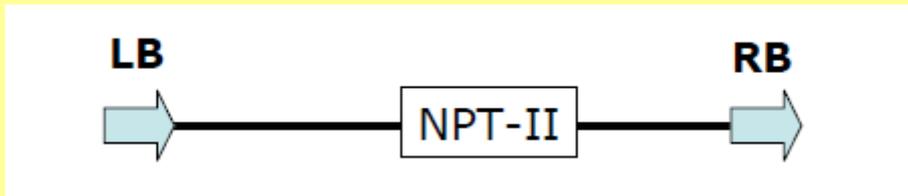


Ac/Ds: Maize transposons.

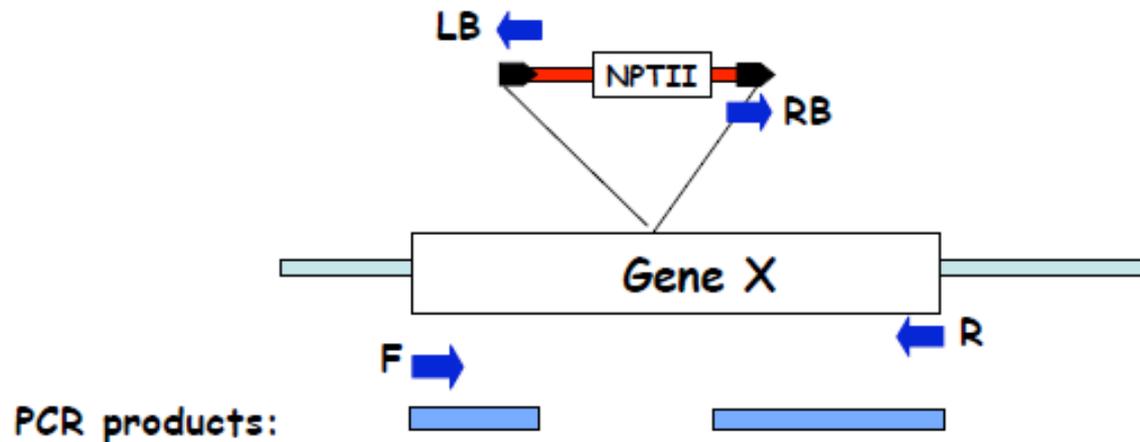
Ac/Ds can function in other plants

both Ac and Ds have 11 bp inverted repeats at the ends, which function in the transposase recognition

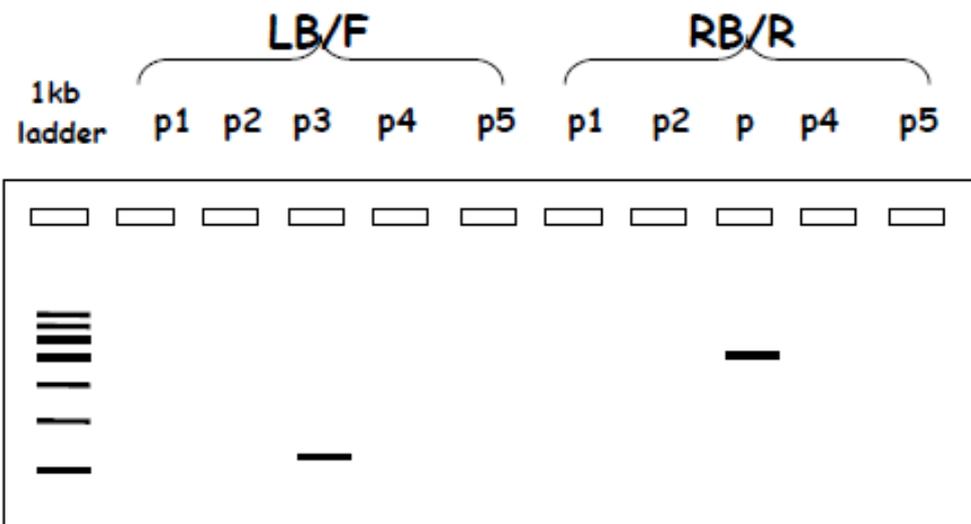
La mutazione causata dall'inserzione del T-DNA



PCR-based screen for T-DNA (or Ds) insertion in specific genes



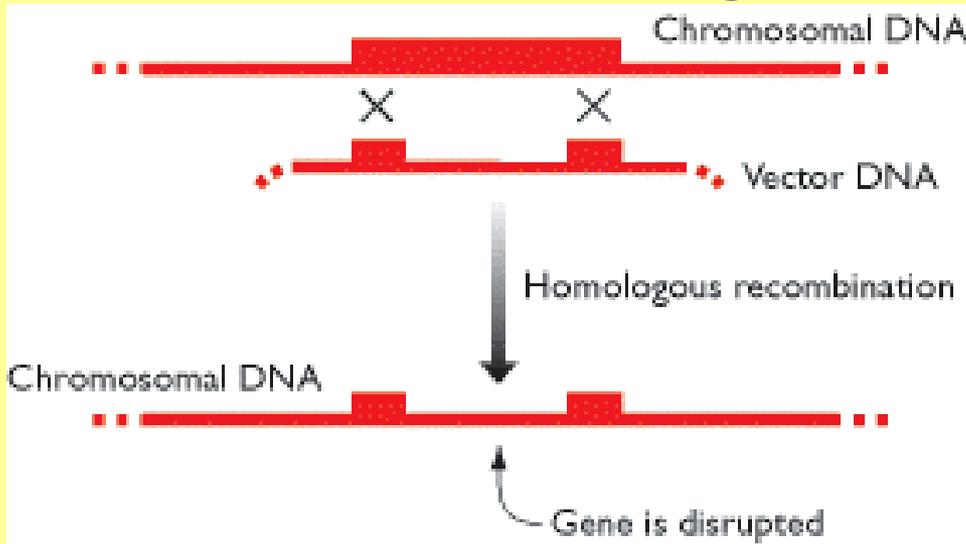
Screening pools (p1 - p5)



identificare la funzione di un gene

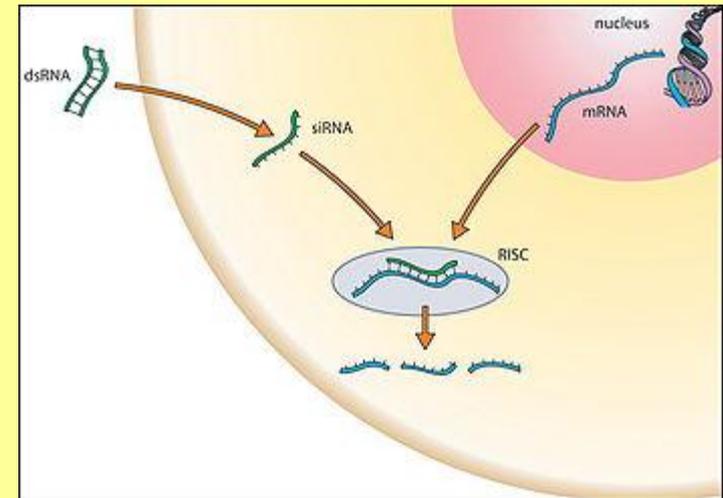
Strumento essenziale è la possibilità di alterare parzialmente o completamente la funzione del gene in esame

ricombinazione omologa



In pianta, a differenza del lievito, la sostituzione di un gene con la sua controparte alterata, non ha portato a risultati apprezzabili

RNA interference



Il sistema del RNA interference (RNAi) non altera completamente l'attività genica, spesso presenta problemi nell'ereditarietà della mutazione

identificare la funzione di un gene

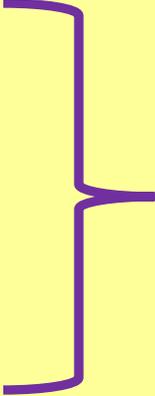
RNAi è molto usato nello studio della funzione di un gene

Prima che la RNAi fosse ben noto, i diversi fenomeni **silenziamento genico** legati all'RNA erano indicati con vari nomi

Silenziamento genico post-trascrizionale

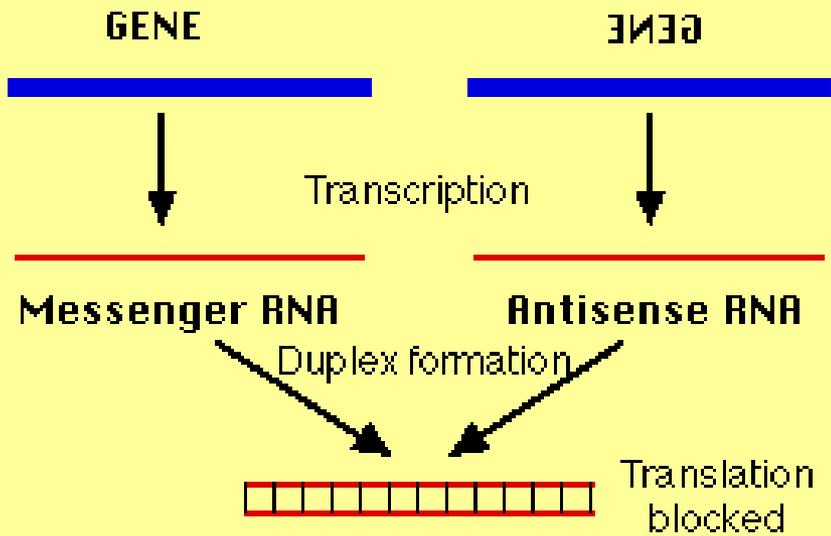
Silenziamento genico indotto da virus

Silenziamento transgenico



Questi fenomeni sono espressione dello stesso meccanismo

L'uso dell'RNA per ridurre l'espressione genica nelle piante:
è stata una procedura usata per molti anni.



RNA antisense a singolo filamento venivano introdotti nelle cellule vegetali che si appaiavano con l'mRNA a singolo filamento ad esso complementare.

Si credeva che il doppio filamento di RNA risultante dall'appaiamento non potesse venir tradotto in una proteina,

è chiaro che l'RNA a doppio filamento avvia invece il meccanismo dell'interferenza.

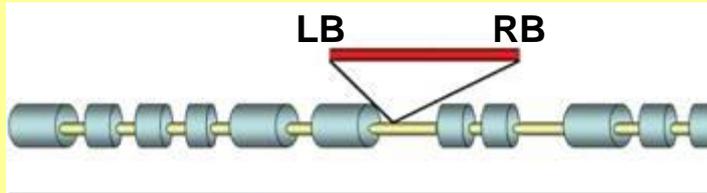
In seguito....

identificare la funzione di un gene

è necessario approntare strategie per inattivare il > numero possibile di geni e studiare i fenotipi derivanti, per comprendere quali siano le funzioni dei geni quando attivi

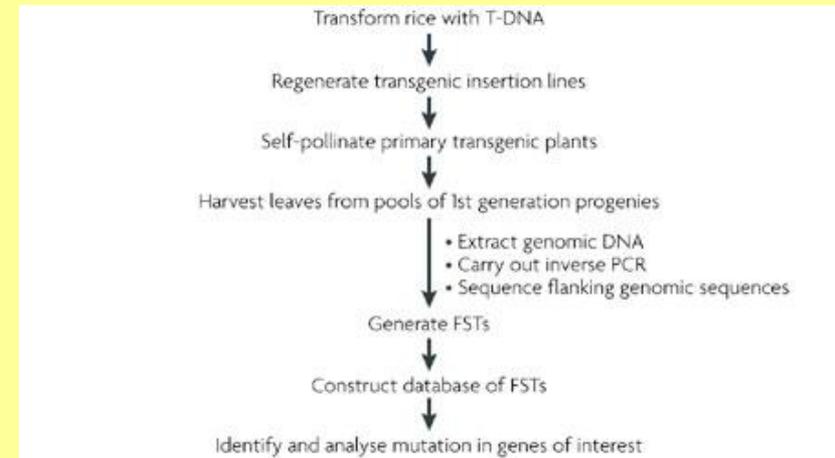
Inserimento casuale nel genoma di Arabidopsis

elementi di T-DNA che inserendosi all'interno dei geni ne impediscono l'attività



In Arabidopsis un'insezione per ciascun gene. A disposizione un'enorme collezione di mutanti.

Ci permetterà in tempi relativamente brevi di attribuire una funzione a ciascuno dei 25,000 geni contenuti nel genoma di Arabidopsis



Mutagenesi inserzionale

T-DNA: uno strumento valido sia per la forward che per la reverse genetics

Un singolo trattamento genera molte inserzioni indipendenti di T-DNA e relativi mutanti.

- 1. trasformato con T-DNA**
- 2. in modo casuale**
- 3. stabile**
- 4. uso la PCR per individuare dove è avvenuta l'integrazione**

Mutagenesi T-DNA

1. Inserire il T-DNA in Arabidopsis e fissare le linee
2. Cercare i fenotipi mutanti
3. Caratterizzare i geni alla base del fenotipo

Il principale svantaggio della mutagenesi mediante T-DNA:

- pattern di integrazione complessi, es. trasferimento di sequenze di vettore che fiancheggiano il T-DNA,
- le inserzioni multiple e la elevata frequenza di concatameri inseriti, complica l'identificazione di sequenze fiancheggianti.

SIGnAL

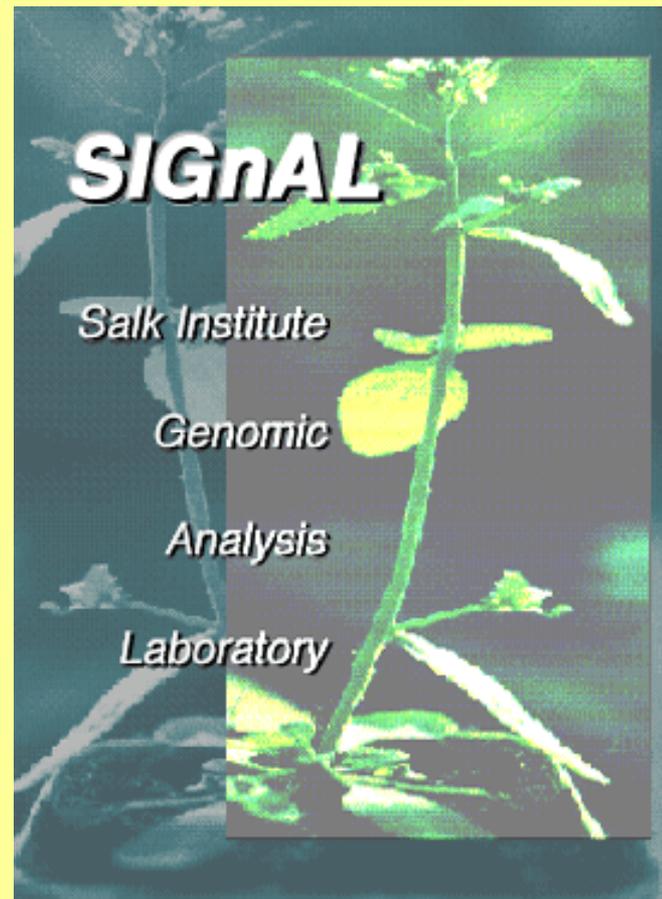
Salk Institute Genomic Analysis Laboratory

Funded by the National Science Foundation

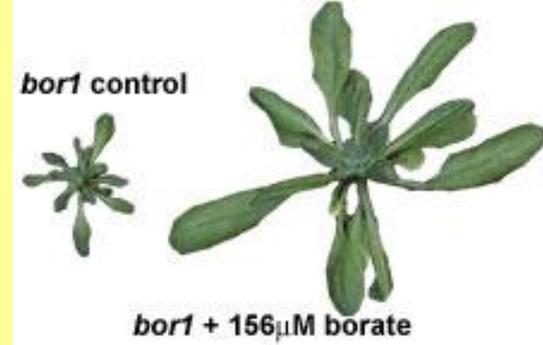


*Arabidopsis
Gene Collection*

*Salk Insertion
Sequence Database*



Mutanti in Arabidopsis



tra tutti i geni predetti in Arabidopsis soltanto il 10% daranno origine ad un fenotipo visibile se alterati nella loro funzione.

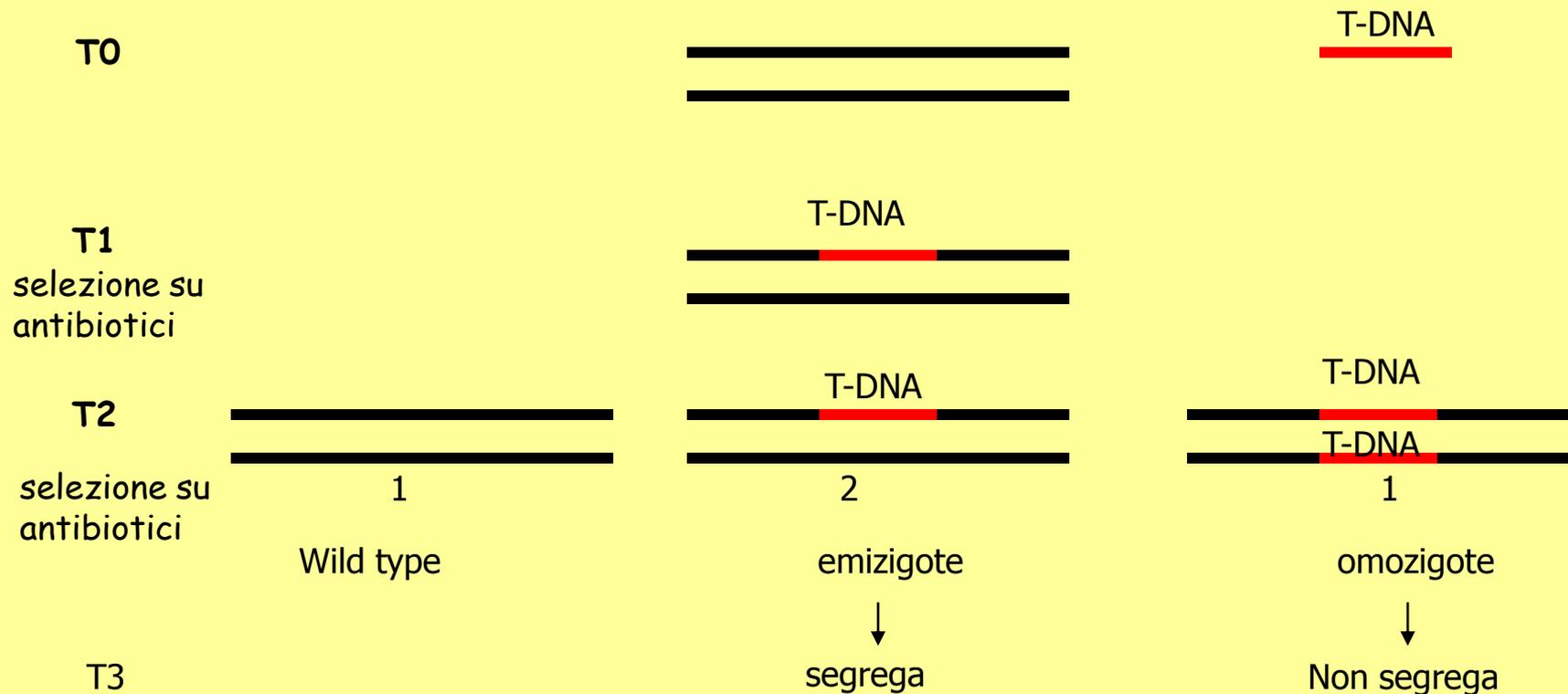
↓
circa il 65% dei geni di Arabidopsis fanno parte di **famiglie geniche** dove geni con alta omologia codificano per proteine con funzioni molto simili o le stesse



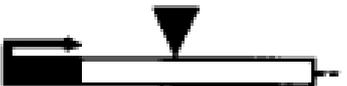
Superare il problema attraverso la creazione di mutanti multipli.
Incroci di piante mutate nei geni che codificano per proteine con un'identica funzione porteranno alla completa soppressione di quella funzione e molto probabilmente origineranno un fenotipo visibile

Come si fa lo screening di mutazioni per inserzione

Lo screening per mutanti inizia in generazione M2 (o in T2) con le poche eccezioni relative allo screening di mutanti dominanti in M1 (o T1)



“Knockology”

<i>Name</i>	<i>Location</i>	<i>Result</i>
 <i>Knock-out</i>	coding region or promoter	null
 <i>Knock-down</i>	promoter or 3' UTR	reduced expression
 <i>Knock-on</i>	promoter	increased expression
 <i>Knock-about</i>	coding region	not null
 <i>Knock-knock</i>	coding region or promoter	multiple KOs in one plant
 <i>Knock-worst</i>	coding region or promoter	chromosomal rearrangement

← es. introne

← incrocio

RNA interference

RNAi scoperto nel tentativo di aumentare l'attività del gene della **calcione sintasi** in petunia

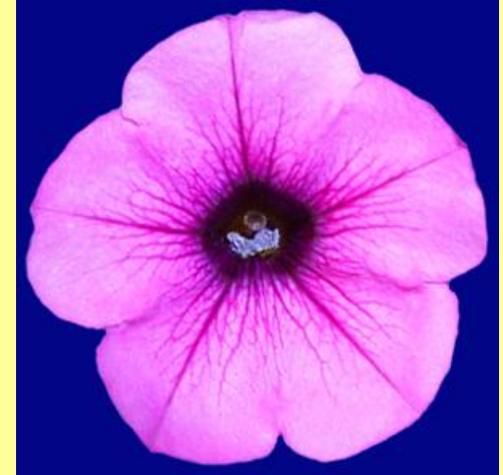


Enzima chiave nella sintesi dei flavonoidi

In Olanda e USA fine anni 80



sovraespresso il gene per modificare l'intensità del colore dei petali



Inaspettato:

la pigmentazione non aumentò, ottennero una variegazione di colore ed in alcuni casi la perdita totale dello stesso

In Petunia

(nel 1990)

Parentale



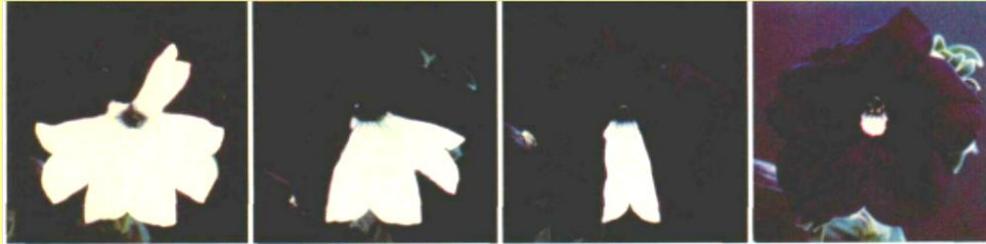
Questo fenomeno definito

co-soppressione

Intendendo la soppressione sia del gene introdotto che del gene endogeno

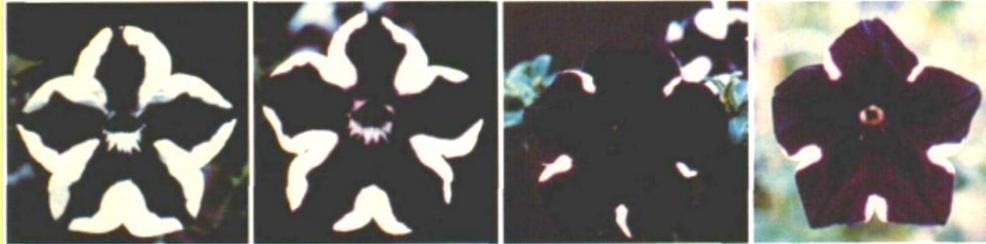
Transgene

218.11



Transgene

218.43



Transgene

218.56



Transgene

218.38



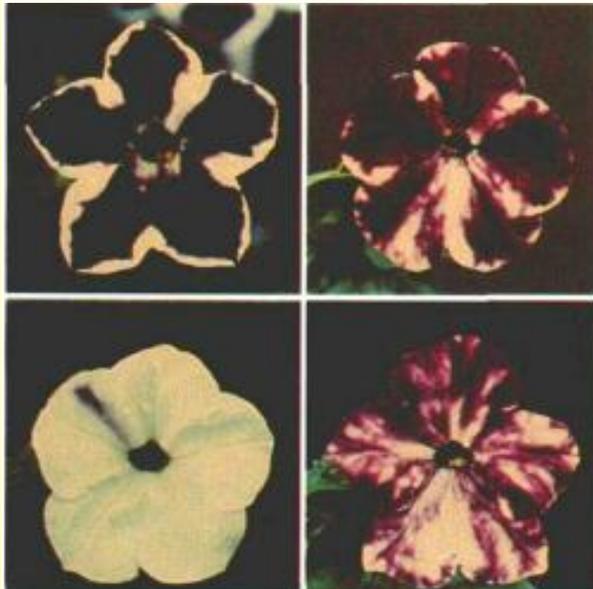
Transgene
218.43



Transgene
218.56



P
r
o
g
e
n
i
e

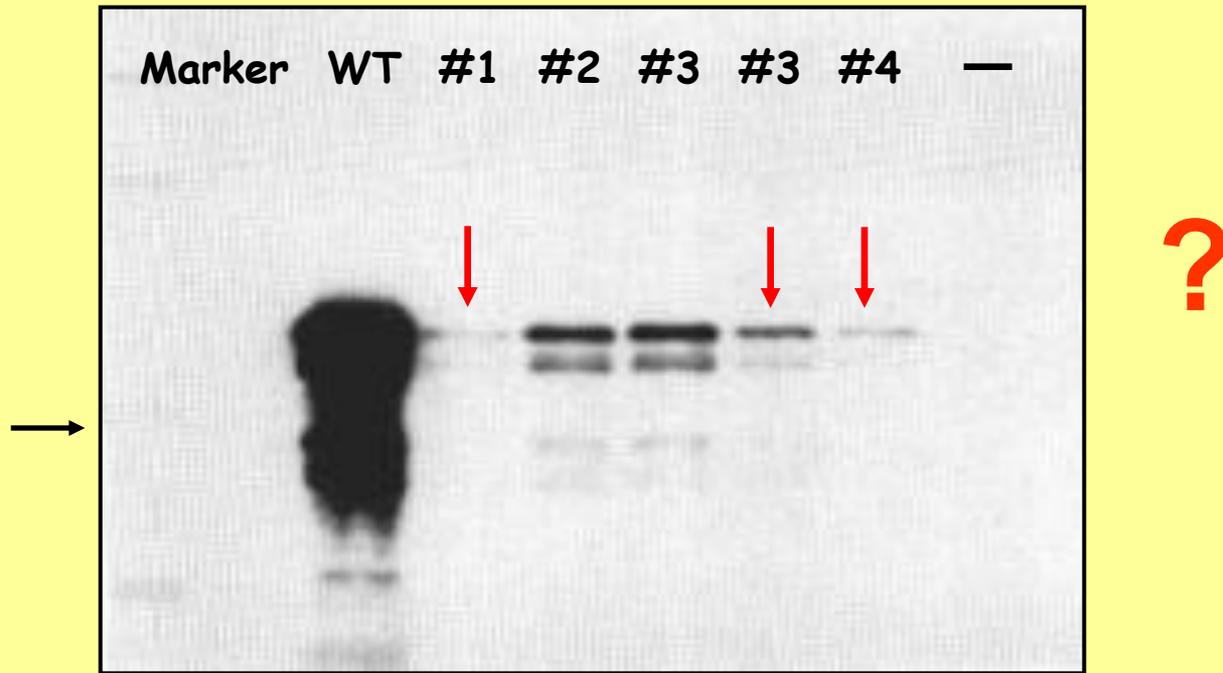


P
r
o
g
e
n
i
e



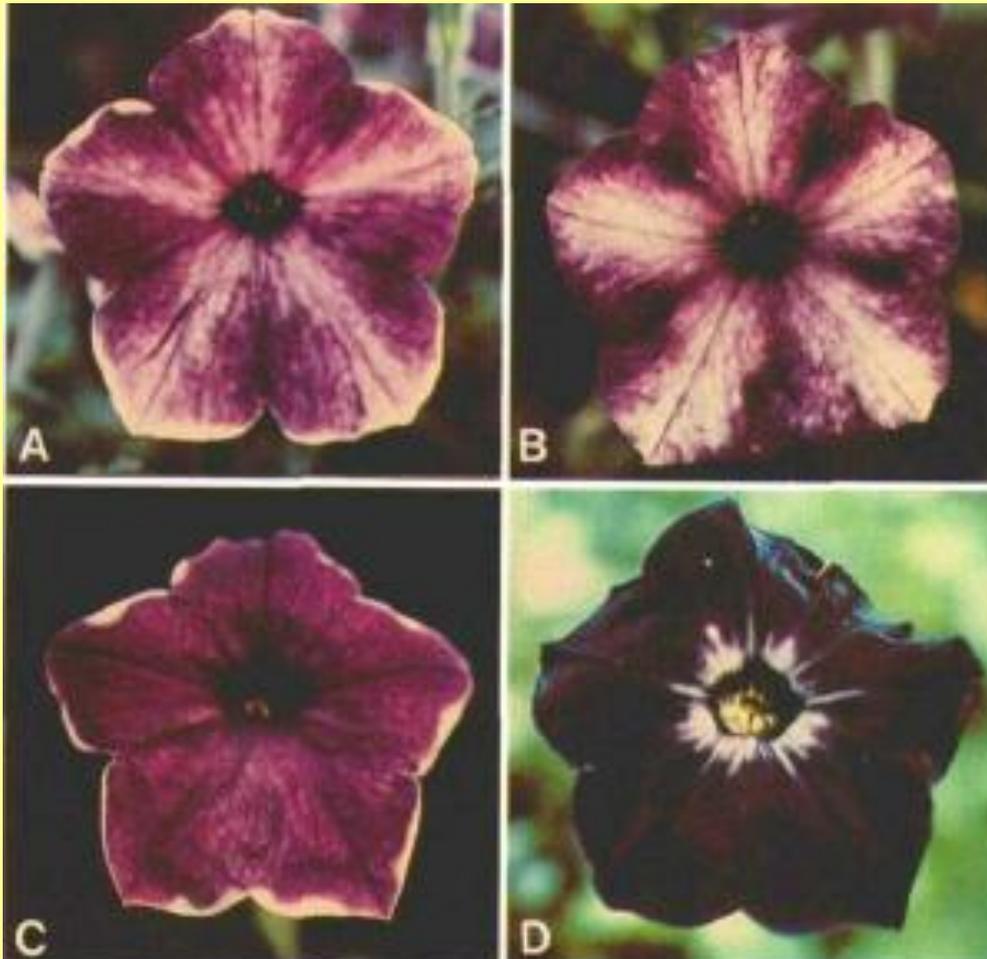
Il fenotipo era ereditabile

Analisi Northern:



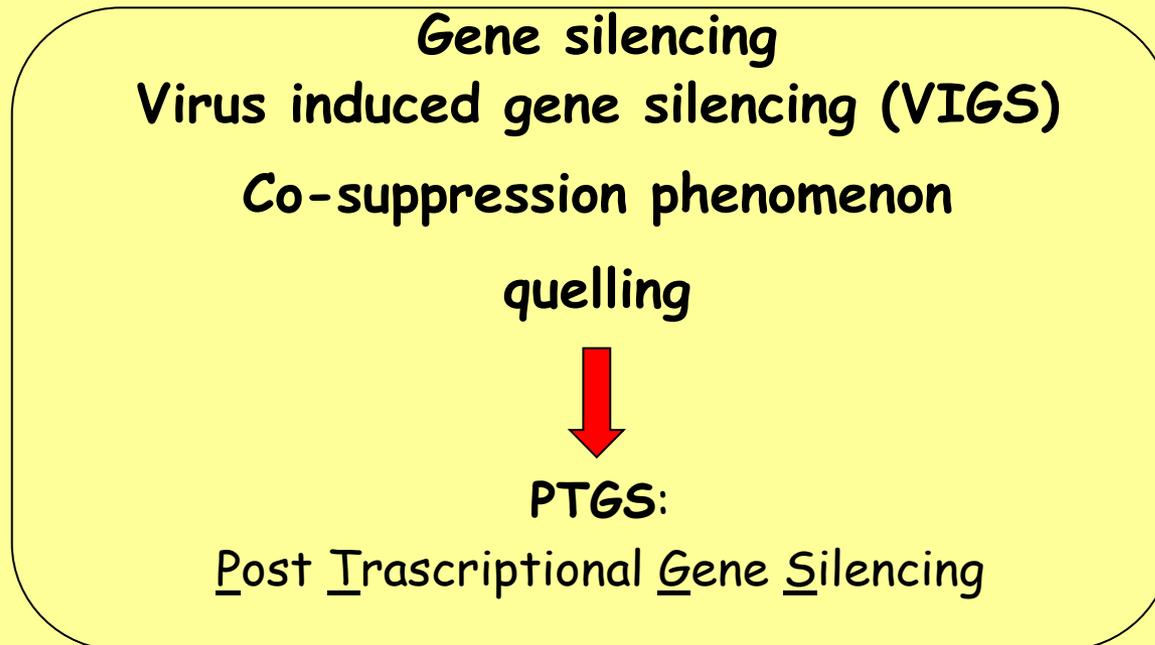
L'RNA del transgene e del gene endogeno era assente, o comunque in quantità molto inferiore rispetto al WT

Phenotypes of CHS antisense Transgenic flowers



“co-suppression of gene expression”

Qualche anno dopo...



1998: Fire and Mello hanno coniato il termine **RNAi**

RNA interference (RNAi)

RNAi compreso solo dopo gli esperimenti di Fire e Mello (Fire et al., Nature 1998)

RNA a singolo filamento è poco efficiente come inibitore

RNA a doppio filamento (dsRNA) è un inibitore potente

In *C. elegans*, alterazioni fenotipiche possono essere indotte solo con sequenze complementari a regioni esoniche, suggerendo un silenziamento post-trascrizionale

RNA interference (RNAi)

RNA a doppio strand silenzia, in maniera sequenza-specifico, l'espressione genica attraverso l'appaiamento con l' mRNA bersaglio seguito dalla sua degradazione



Silenziamento post-trascrizionale

Meccanismo biochimico altamente conservato che coinvolge più di 10 geni

RNAi viene utilizzato per silenziare selettivamente un gene

Questa **loss of function** permette di identificare la funzione del gene mediante il suo "spegnimento"

Non si parla di **Knock-out** ma di **Knock-down**

Meccanismo base dell'RNAi

Il modello funzionale prevede tre fasi:

Fase di "iniziazione"

Taglio del dsRNA in siRNA (21-23nt)

Fase "effettrice"

Il siRNA viene incorporato in un complesso proteico definito **RISC** "RNA-induced silencing complex"

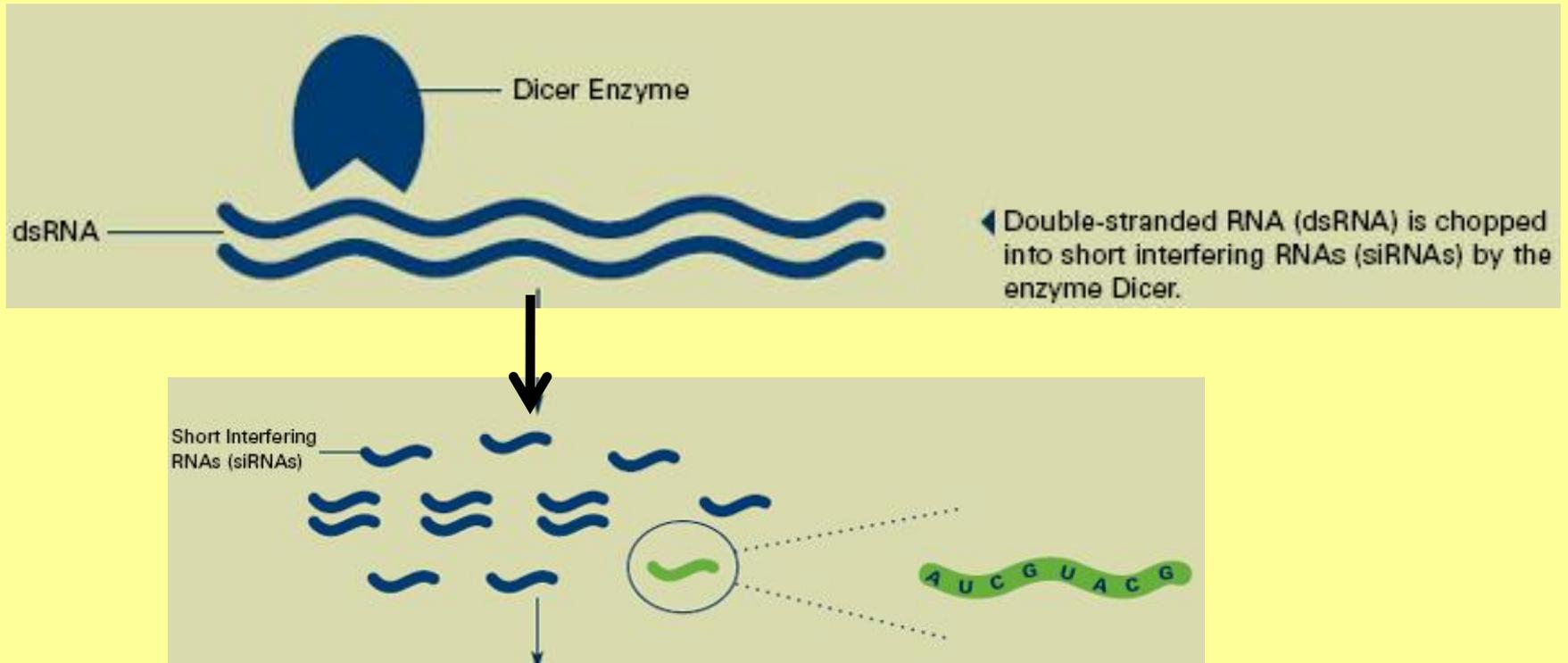
Fase di "amplificazione"

Sintesi di nuovo dsRNA da parte di una RNA-polimerasi RNA-dipendente

MECCANISMO

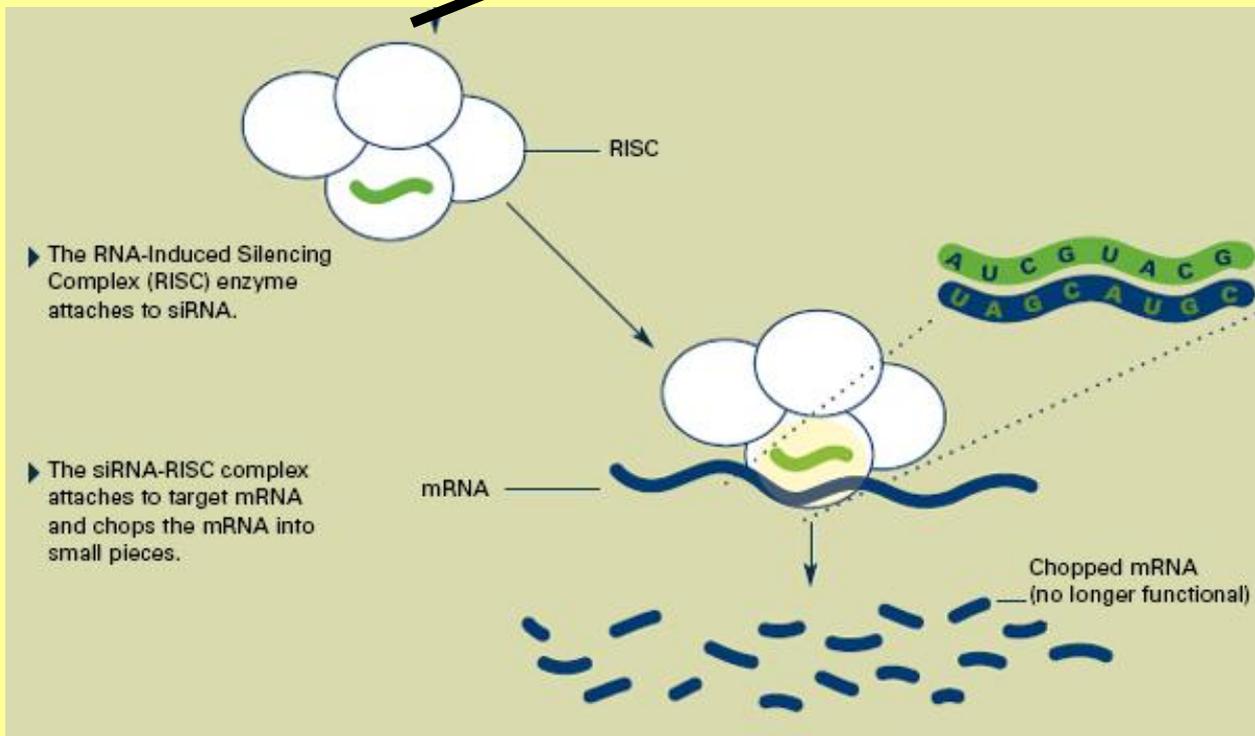
FASE DI INIZIAZIONE

Enzima ribonucleasico della famiglia delle RNAsi III.
Genera siRNA di 21-23 bp



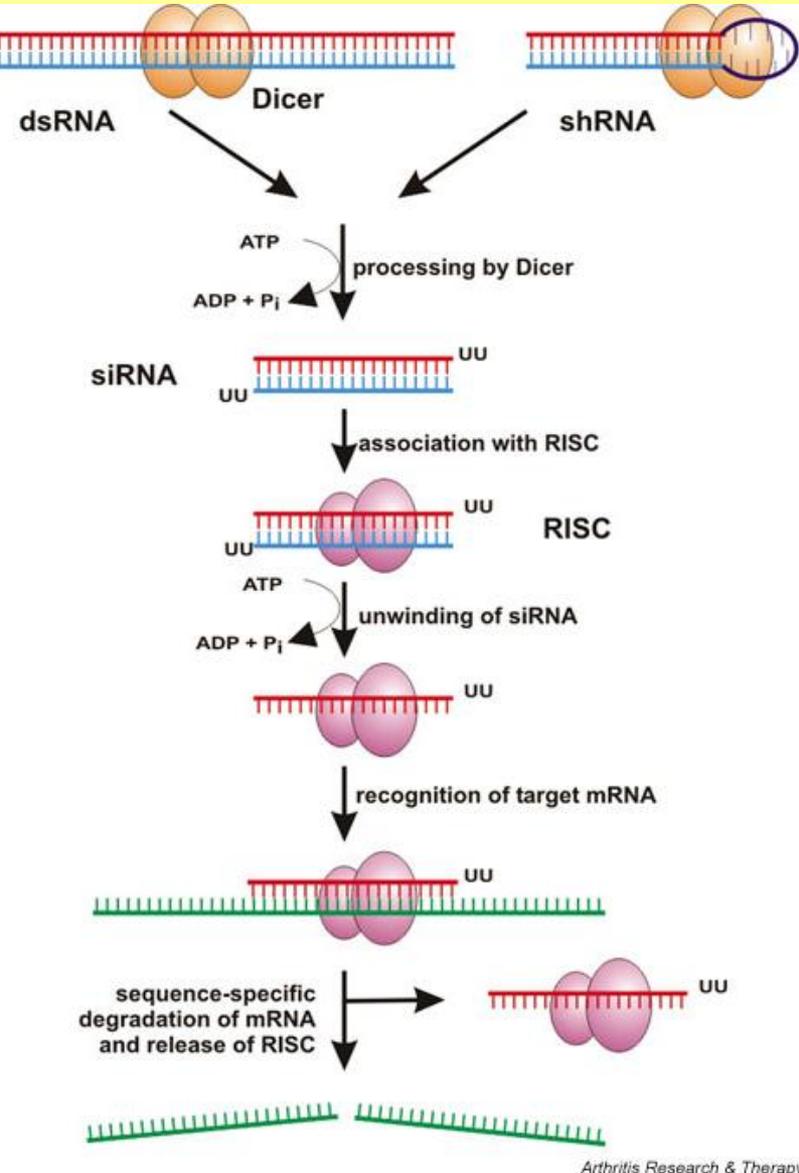
RISC costituito da due RNA binding proteins e da una nucleasi chiamata Slicer

FASE EFFETRICE



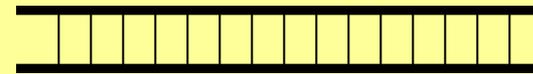
FASE DI AMPLIFICAZIONE

Identificati 2 meccanismi di amplificazione, entrambi operati da una (RdRP)



1) Formazione dsRNA

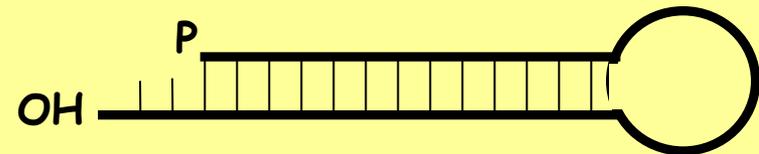
- RNA polimerasi RNA dipendente



siRNA

- *formazione hairping*

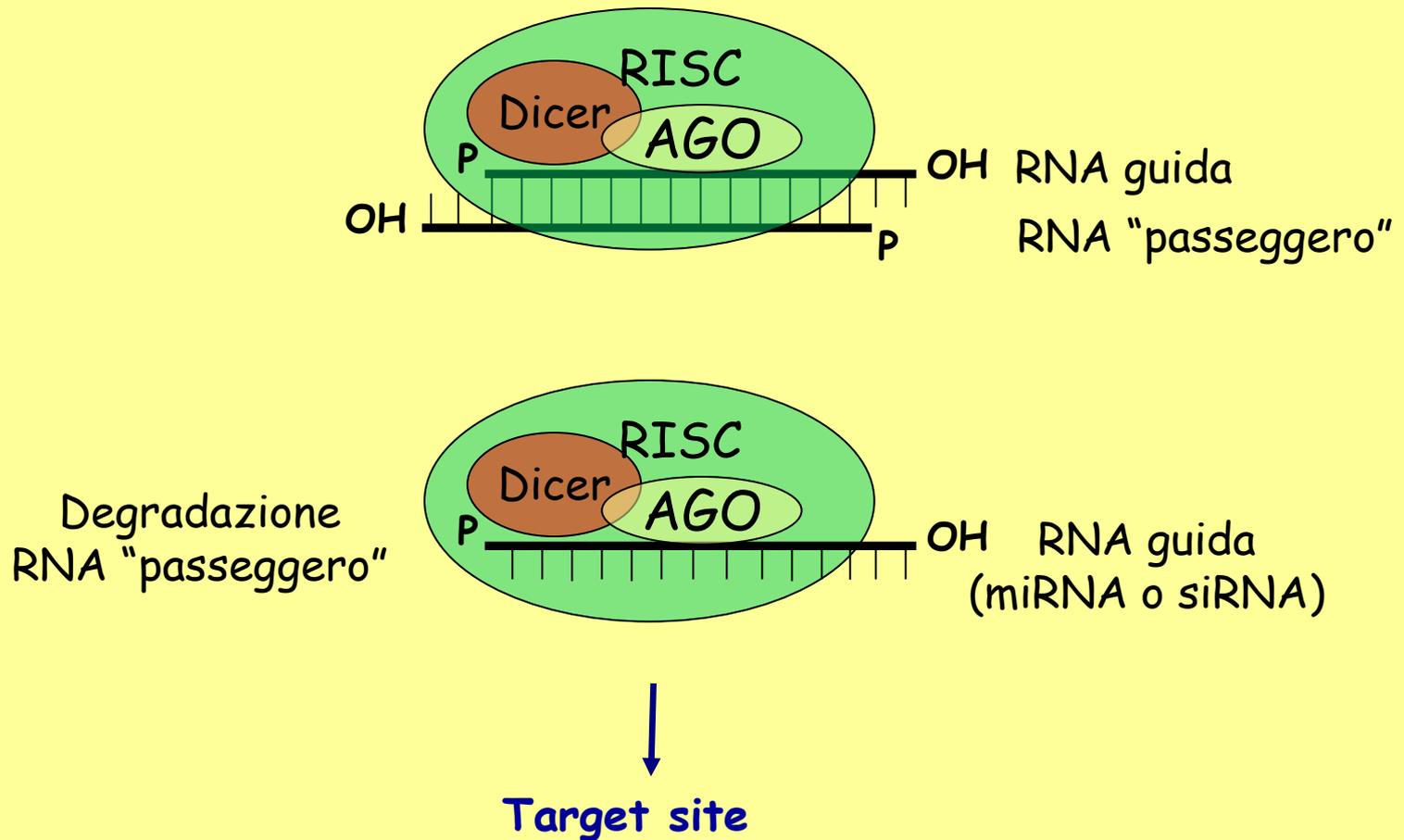
(appaiamento intramolecolare del trascritto)



miRNA

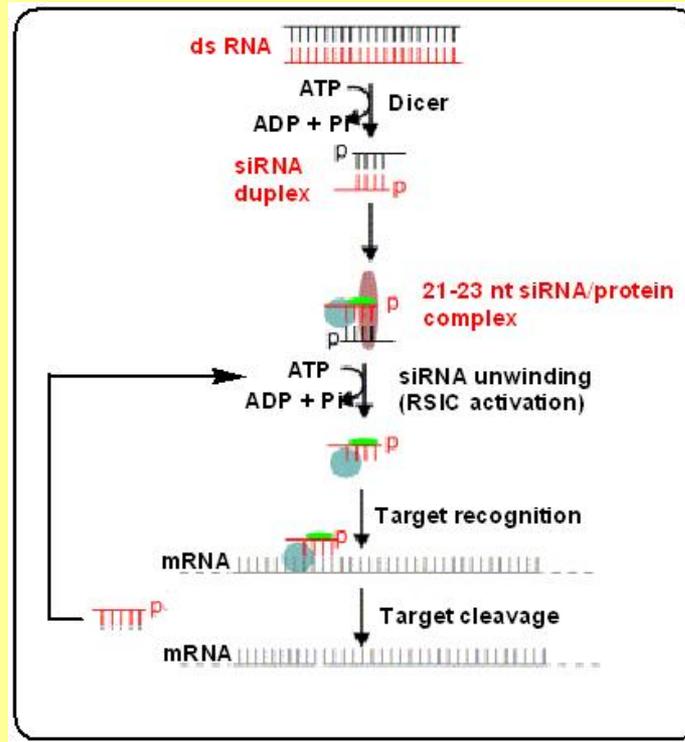
La biogenesi dei miRNA e siRNA è diversa ma non è diverso il loro meccanismo d'azione.

RISC (RNA Induced Silencing Complex)



Nel complesso RISC si lega la proteina Argonauta (AGO), si muove al sito target e può indurre la proteina AGO a tagliare la sequenza prevenendone così la traduzione

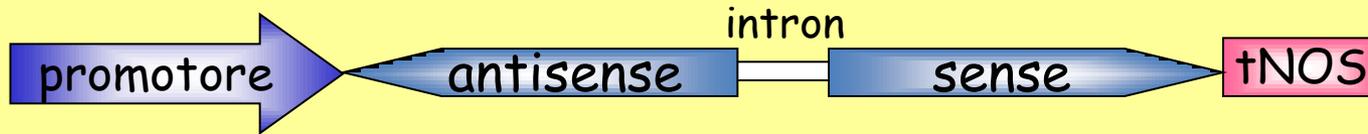
Target site:



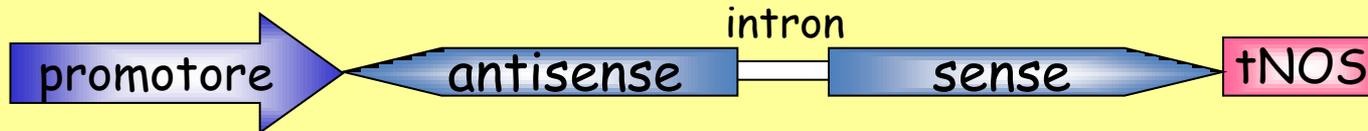
L'uso di RNA a doppio filamento come strumento per studiare la funzione dei geni si è diffuso dopo la scoperta del meccanismo molecolare dell'RNAi, inizialmente in *petunia* e più tardi in *Caenorhabditis*.

Studio dei geni coinvolti in tutti i processi che portano alla formazione della pianta, la resistenza agli stress

Temporal and spatial control gene silencing



1) **CaMV 35S**



2) **Promotore inducibile: trattamento chimico**

3) **Promotore regolato dallo sviluppo**

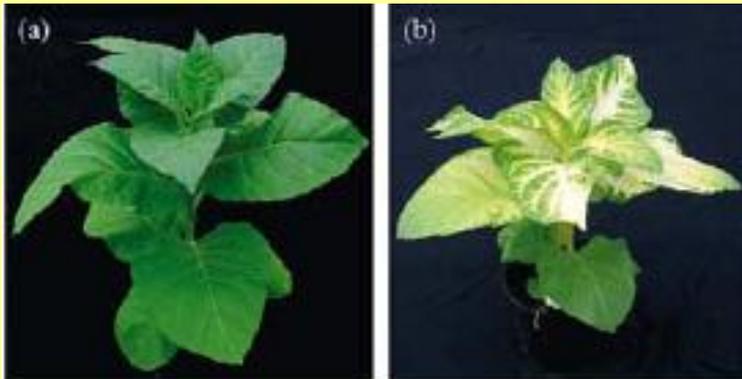
4) **Promotore tessuto specifico**

Temporal and spatial control of (Mg)-chelataase subunit I (*Chl I*) gene silencing

Temporal *Chl I* silencing

WT

dsChl

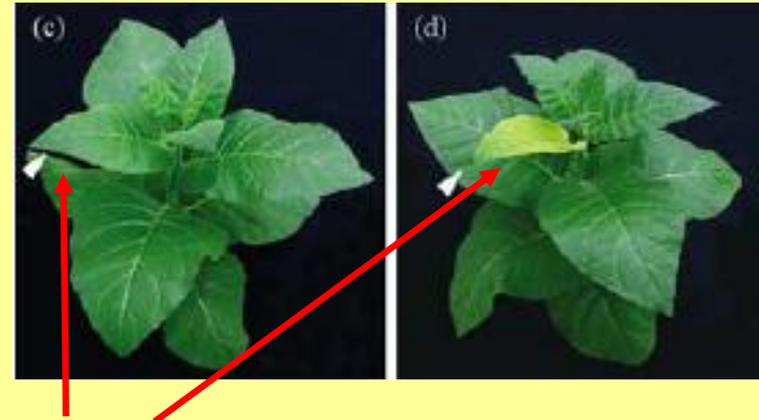


Dopo 15 gg di trattamento
con Etanolo

Local *Chl I* silencing

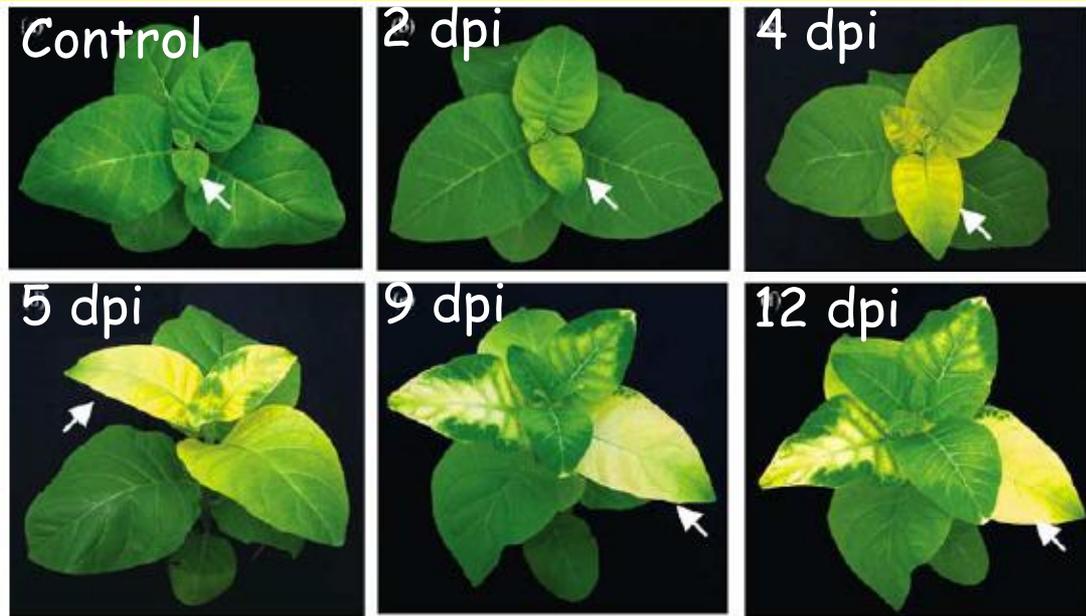
WT

dsChl



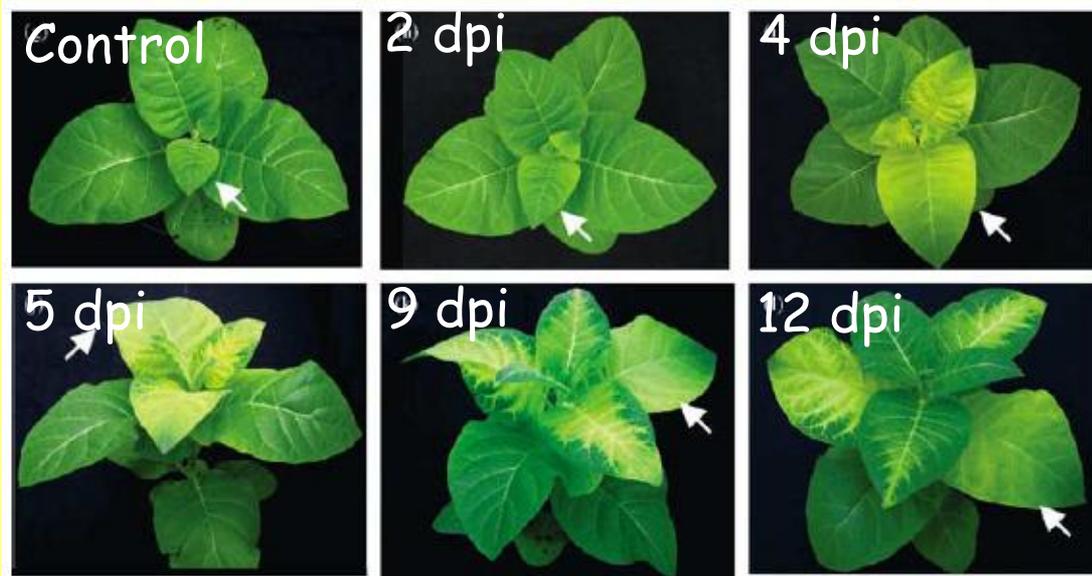
Foglie a contatto con Etanolo

Linea A



**dpi: days
post-induction**

Linea B



Riassumendo...

Gene silencing e genomica funzionale

Sequenziamento dei genomi di
Arabidopsis , riso ecc.

Genetica inversa (transposon tagging, T-DNA tagging)



Collezione di mutanti Syngenta Arab Insertion Library (SAIL)

Salk Institute Genomic Analysis Laboratory (SIGnAL)

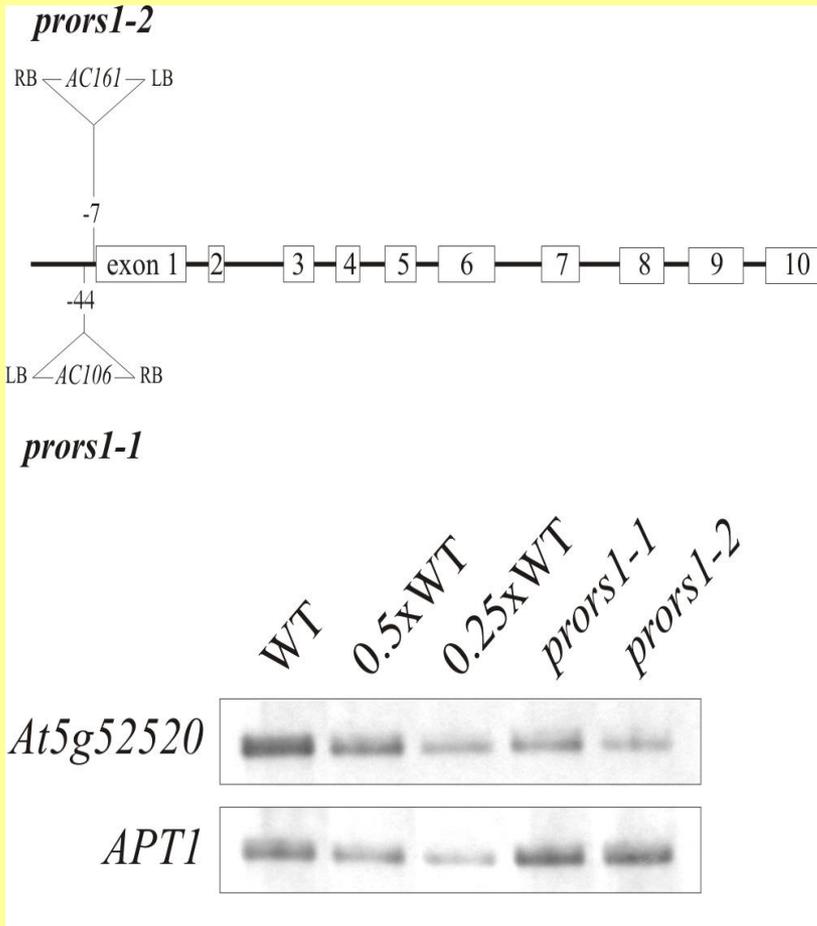
**Gene silencing: introduzione di dsRNA in una cellula per indurre
la degradazione del trascritto del gene che si vuol silenziare**

Effetti: knock-out

knock-down

Alcuni esempi:

mutazioni "leaky" dovute ad inserzioni di T-DNA nella regione del promotore



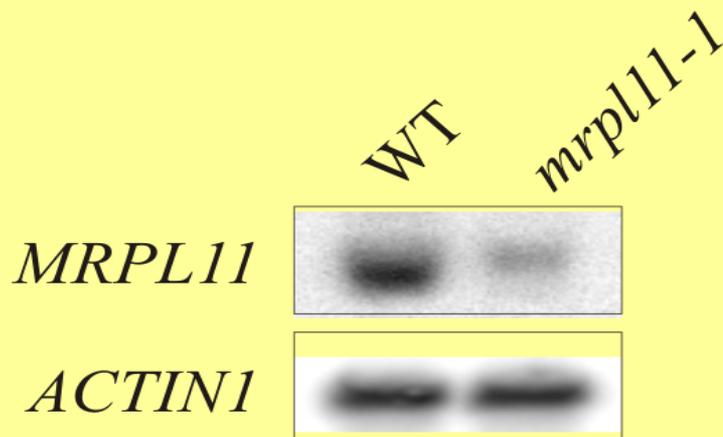
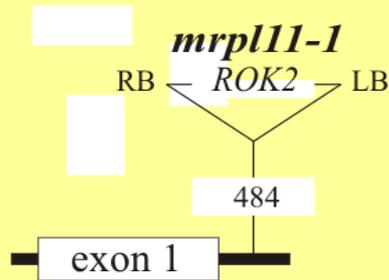
Due inserzioni di T-DNA avvenute a 44 e 7 bp a monte del codone d'inizio della traduzione

causano una riduzione della trascrizione nei due alleli mutati *prors1-1* e *prors1-2*,

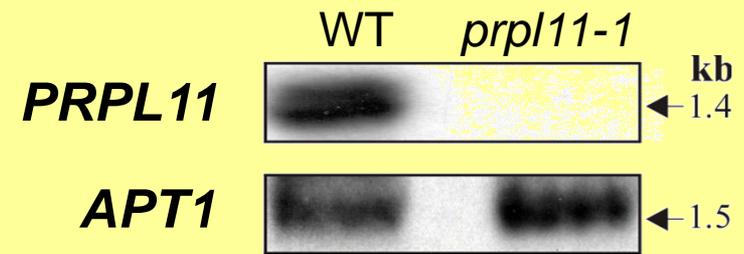
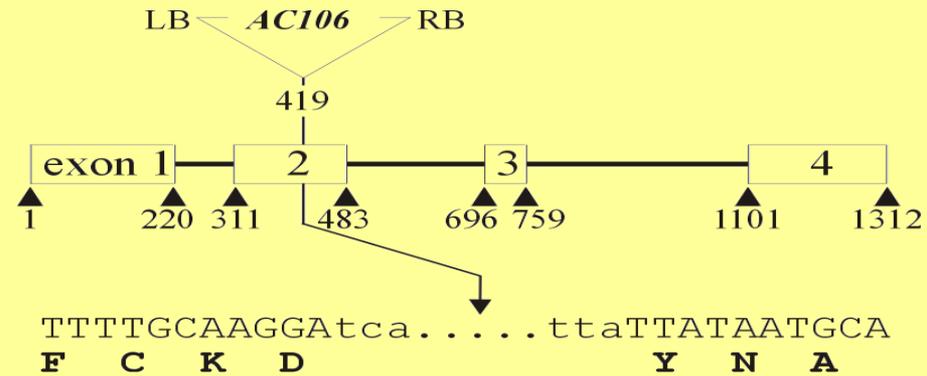
analisi di RT-PCR

Esempi di mutazioni "leaky" e Knock-out causate dal T-DNA nella regione 3' UTR o all'interno dell'ORF

Inserzione nel 3'UTR



Inserzione nel secondo esone



Mutagenesi

T-DNA: uno strumento valido sia per la forward che per la reverse genetics

- 1. trasformato con T-DNA contenente un trasposone**
- 2. in modo casuale**
- 3. non è stabile**
- 4. fenotipi visibili in T2**

Le mutazioni con trasposone attivo non sono stabili

Il trasposone può venir escisso ed il gene torna ad essere trascritto



Durante il processo di escissione lascia dei segni sul genoma (duplicazioni o delezioni) si può alterare il frame e produrre proteine incomplete, non funzionali e degradate

Esempi di modificazioni indotte dall'escissione del trasposone *En*

PetE1 (At1g76100) GATCGTT**TT**AGCTGGAAA

3' *En-1* 5'

198

pete1-1

GATCGTT**TTA** ttaGCT

+ TTA

pete1-1.1

GATCGTT¹⁹⁵ (-52bp)²⁴⁸CCC

- 52 bp

PetE2 (At1g20340) ATGGCCT**TC**AGTAACCTCA

5' *En-1* 3'

10

pete2-1

ATGGCCT**CA** tcaGTAACCTCA

pete2-1.1

ATGG...A tcaGTAACCTCA

pete2-1.2

ATGGCCT**C**. .gaGTAACCTCA

pete2-1.3

ATGGCCT**CA** tgaGTAACCTCA

Stop codon

Reversione somatica indotta dall'escissione del trasposone

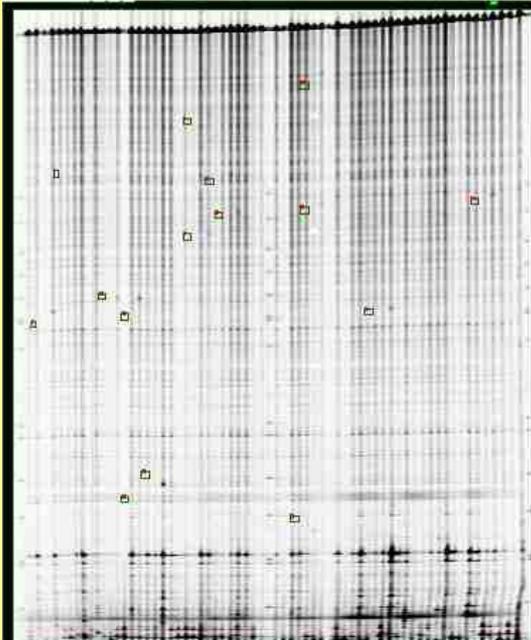


soltanto una parte della pianta contiene cellule con un trasposone inserito in un gene essenziale allo sviluppo dei cloroplasti.

Nel resto della pianta il trasposone è saltato fuori dal gene senza alterarne la capacità di produrre una proteina totalmente funzionale

identificare la funzione di un gene

Spesso la completa distruzione di un gene non risulta essere particolarmente informativa;



si pensi al caso in cui la distruzione del gene risulti in un fenotipo letale



avere geni con mutazioni puntiformi e codificanti per proteine parzialmente attive è molto più utile

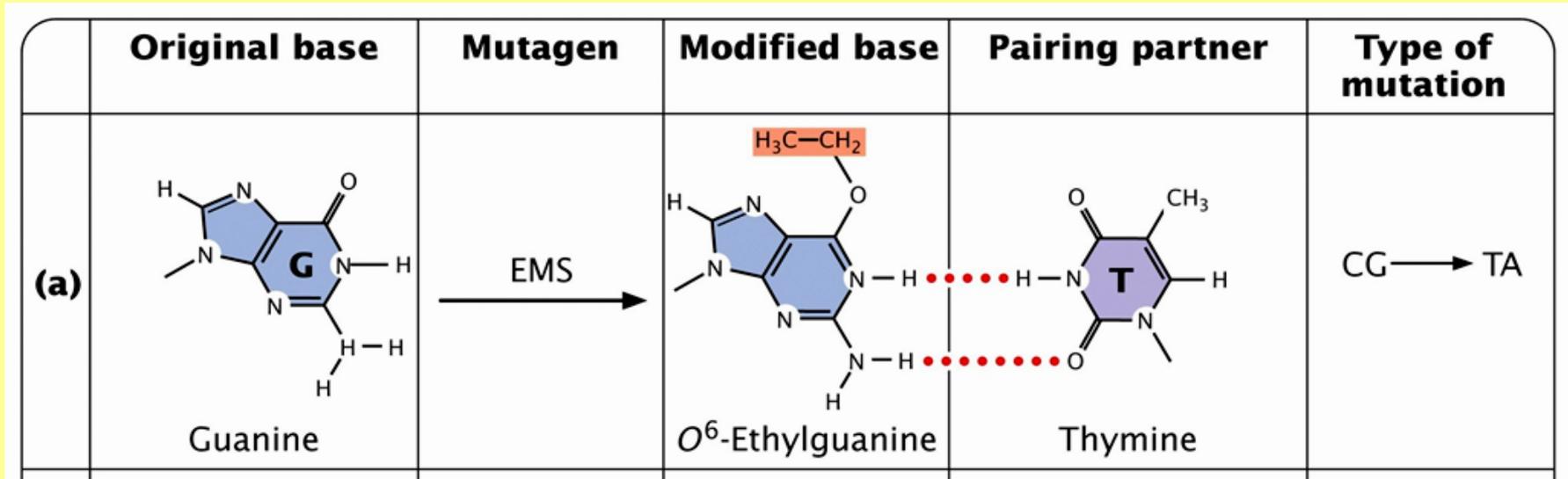
Il **"TILLING"** è la tecnologia che permette in tempi relativamente rapidi l'identificazione di questi alleli mutati.

TILLING

- Targeting **I**nduced **L**ocal **L**esions **I**n **G**enomes
- Una strategia di **genetica inversa** applicabile a qualsiasi organismo
- Combina mutagenesi chimica e PCR-screening per l'identificazione di mutazioni in sequenze di interesse
- Consolidata in *Arabidopsis*.
- Oggi applicata anche in altre specie vegetali e non

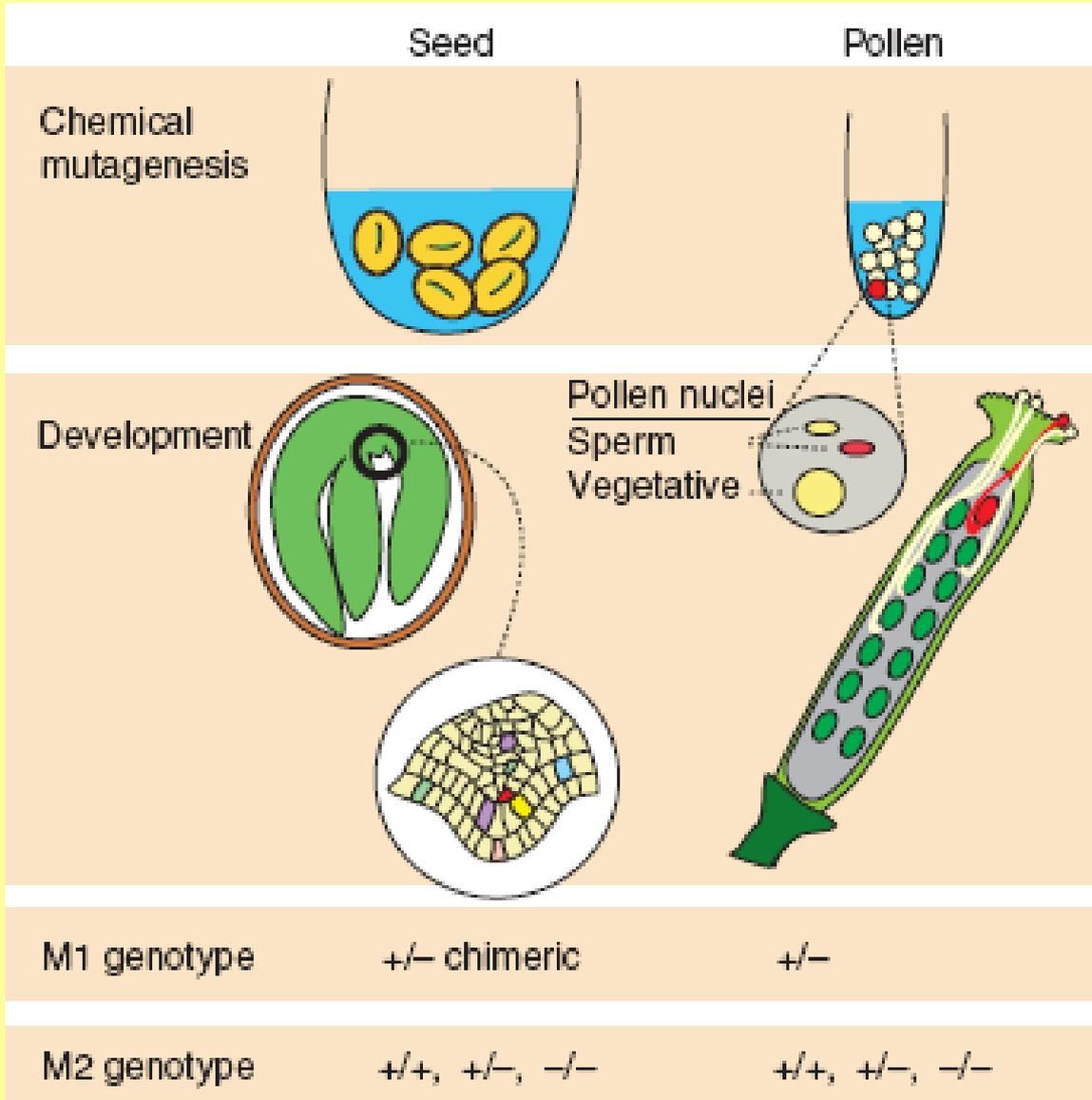
Mutagenesi chimica con Ethyl-methyl sulfonato (EMS)

- ◆ Ethyl-methyl sulfonato è un agente alchilante delle basi azotate
- ◆ La mutazione più frequente consiste nella transizione da *G/C* a *A/T*

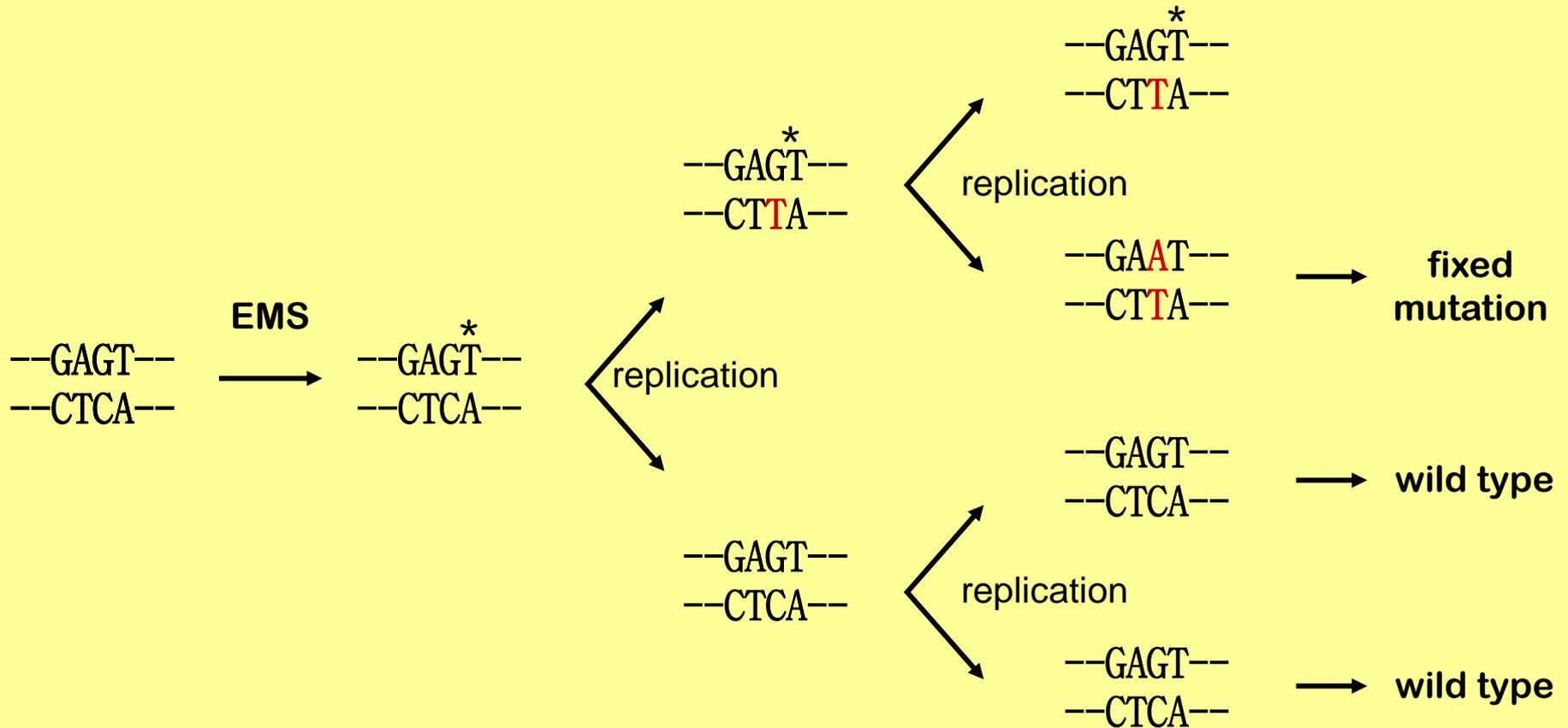


- ◆ La mutagenesi con EMS provoca mutazioni ad alta densità, rendendo possibile l'analisi di un numero ridotto d'individui per isolare il gene mutato d'interesse
- ◆ produzione di diverse sequenze alleliche, cioè di versioni mutate del locus che portano ad avere una proteina con una serie di diverse funzioni

Questo tipo di mutagenesi consente di individuare quali parti di un determinato enzima svolgono un ruolo importante, quali invece hanno semplicemente lo scopo di unire tra loro i vari domini funzionali



Le mutazioni saranno fissate solo dopo replicazione



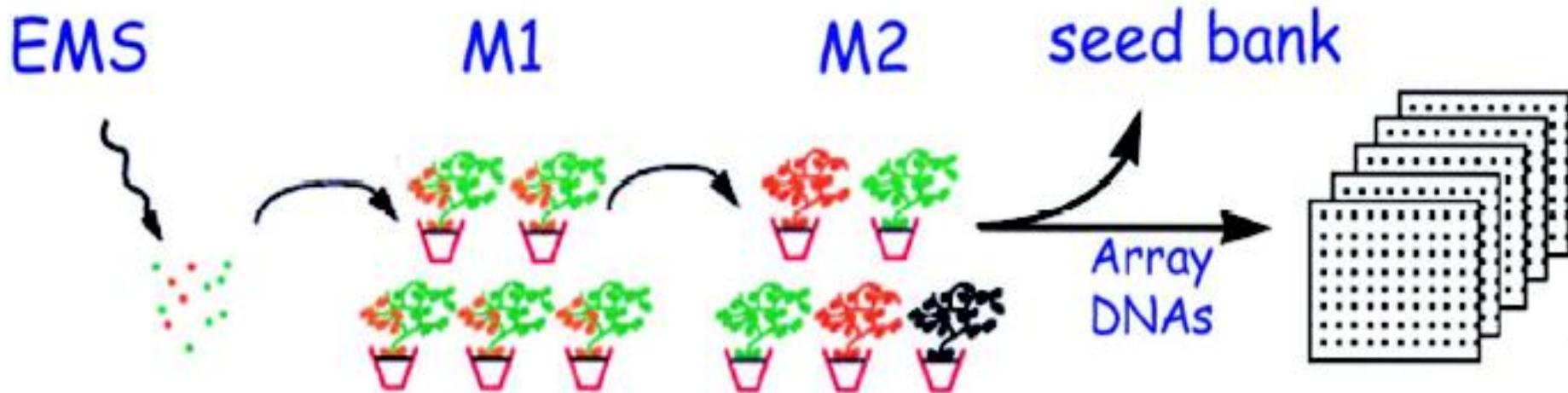
Nella generazione F1 avremo dei mosaicismi:

alcune cellule avranno la mutazione, altre cellule no.

Normalmente vengono mutagenizzati i semi con ethylmethylsulfonato (EMS)

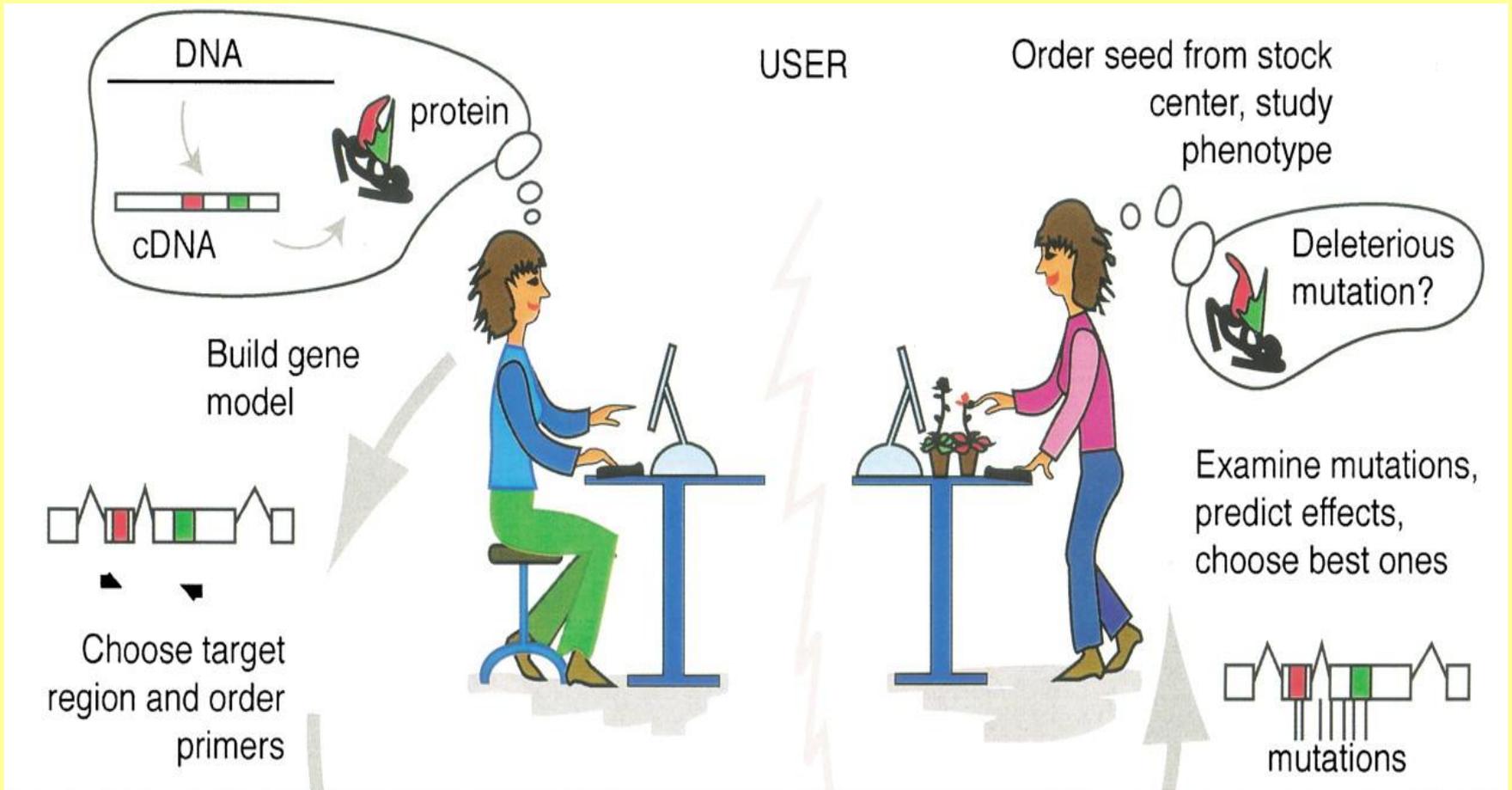
Le piante M1 ottenute vengono lasciate autofecondare

dagli individui M2 viene estratto il DNA per lo screening delle mutazioni
(la progenie di ogni pianta M2 viene catalogata e conservata)



I campioni di DNA vengono riuniti in pool e organizzati in microtiter-plates

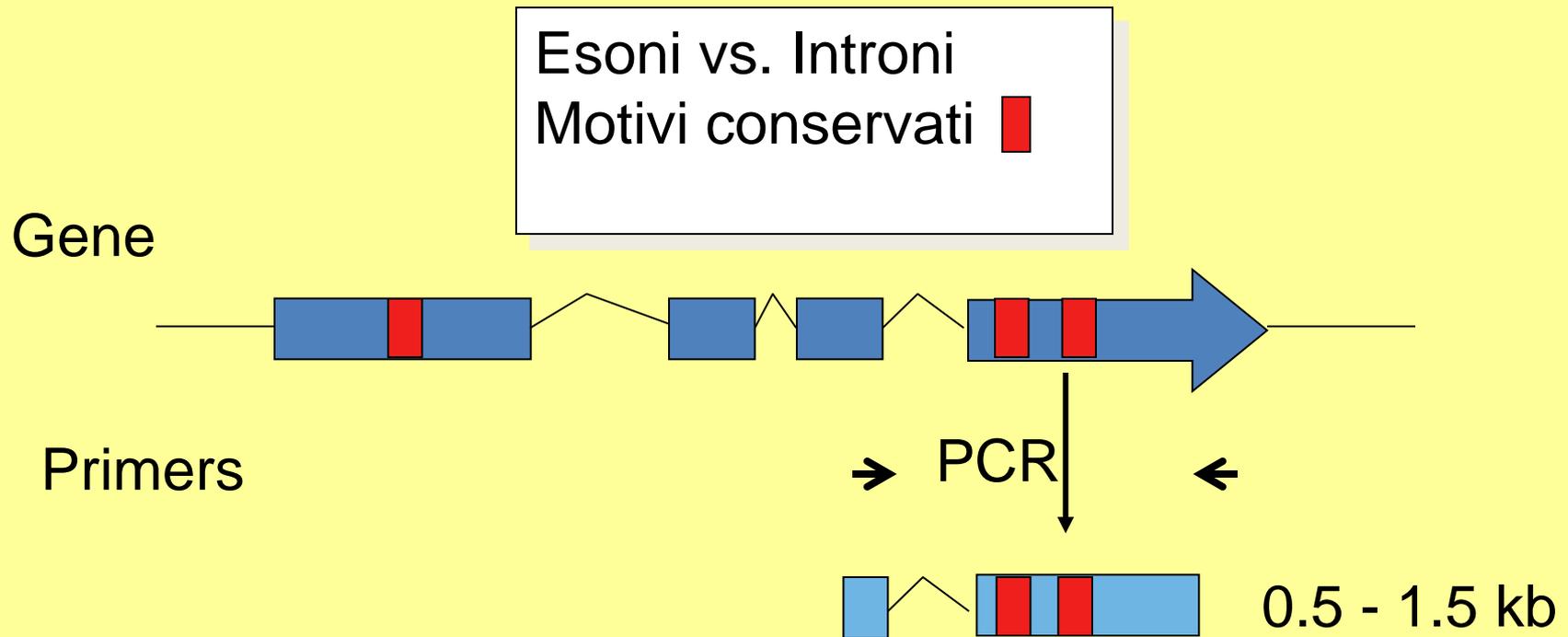
96 pozzetti x 8 plates (768 pools)



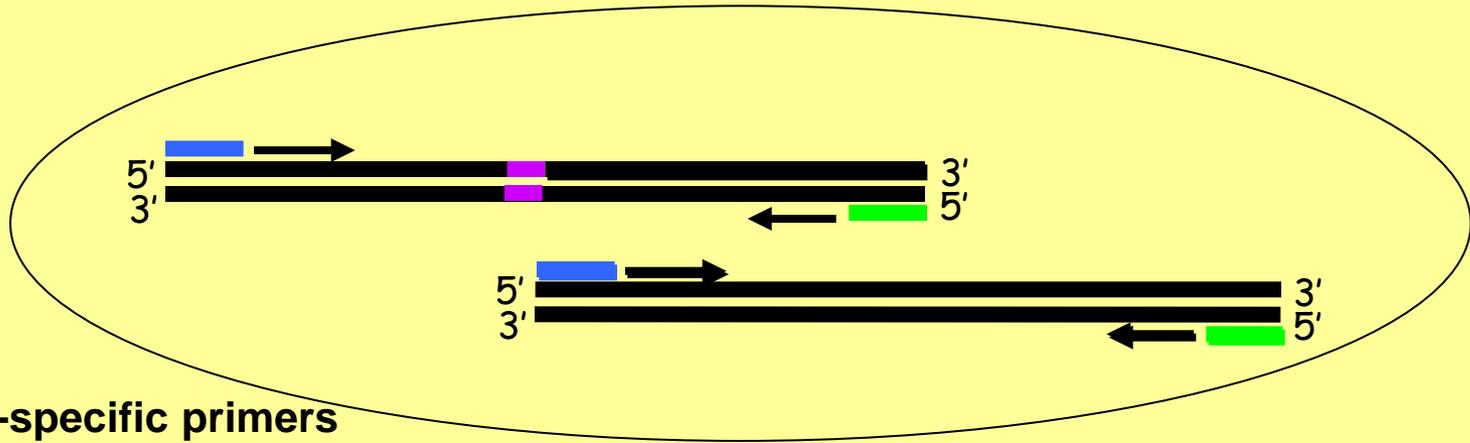
- ◆ Tilling è utile **1** individuare i fenotipi visibili prodotti dalla mutazione
- ◆ **2** individuare all'interno di una proteina residui amminoacidici che hanno un ruolo molto importante all'interno di domini funzionali
- ◆ quindi la scelta dei primers per lo screening dipenderà anche da ciò che si vuole analizzare

Scegliere la porzione del gene che si vuol analizzare

- ◆ ottenere degli effetti prodotti dalle mutazioni importante considerare le parti codificanti,
e tra queste le regioni più conservate durante il processo evolutivo, più alta
- ◆ sarà la probabilità che queste regioni svolgano un importante ruolo funzionale

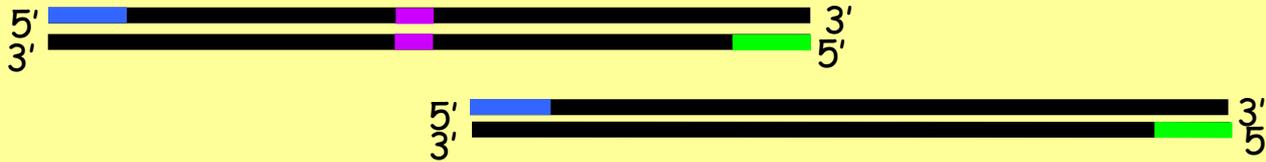


Amplificazione per PCR del gene desiderato da **pool di DNA**



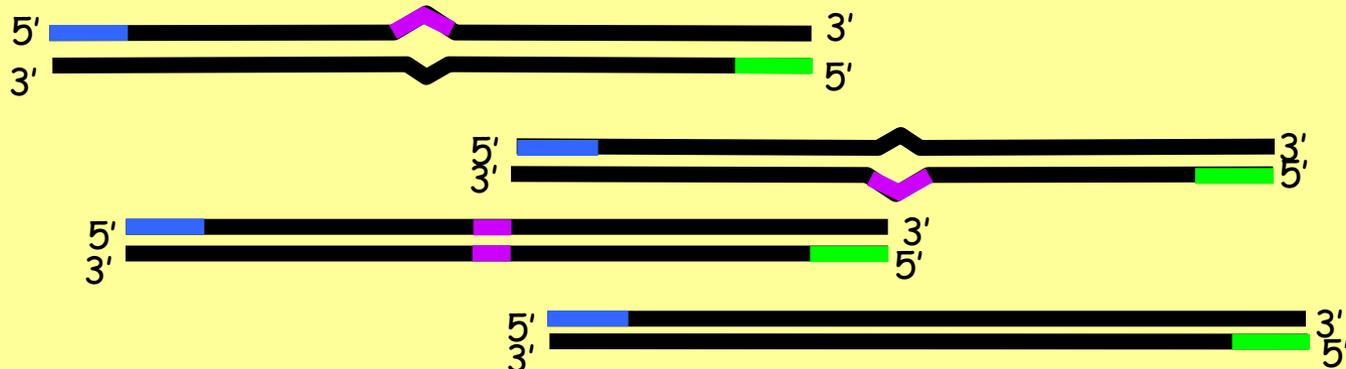
two gene-specific primers
labeled with different IRDyes

↓ PCR

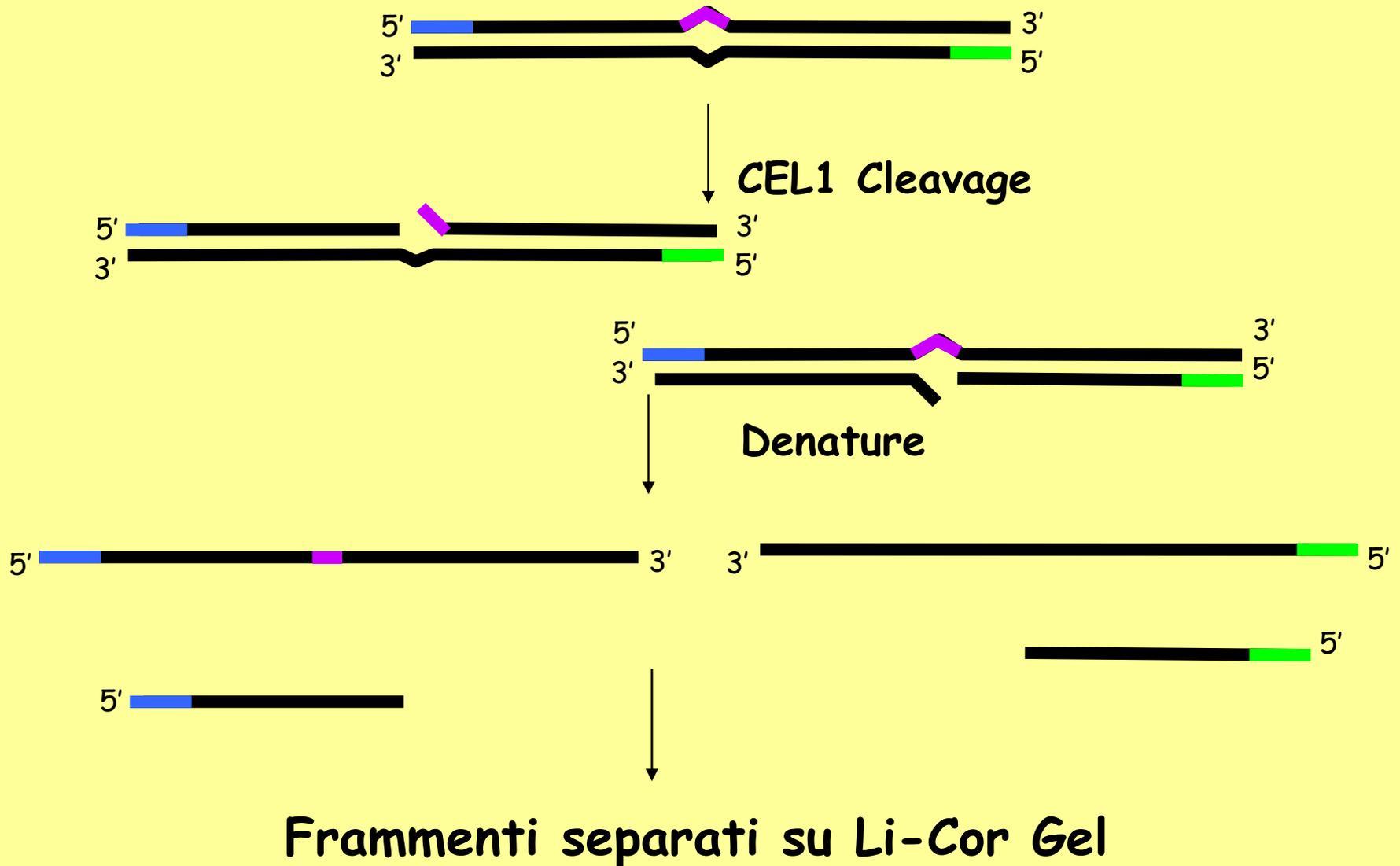


denaturazione e rinaturazione al
fine di ottenere l'eteroduplex

↓ heat,
re-nature



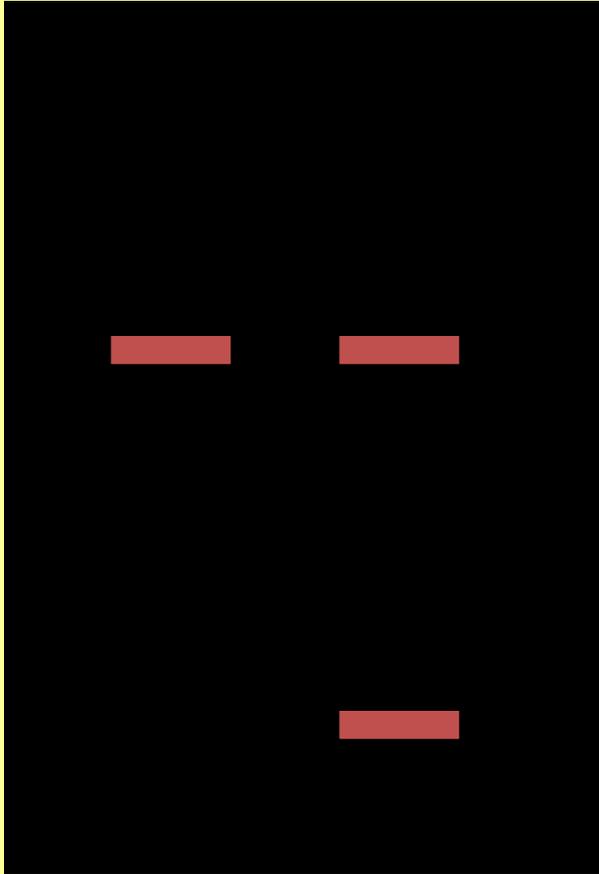
Screening basato su gel di Acrilamide per individuare etero-duplex



LI-COR Scanning Results:

Normal

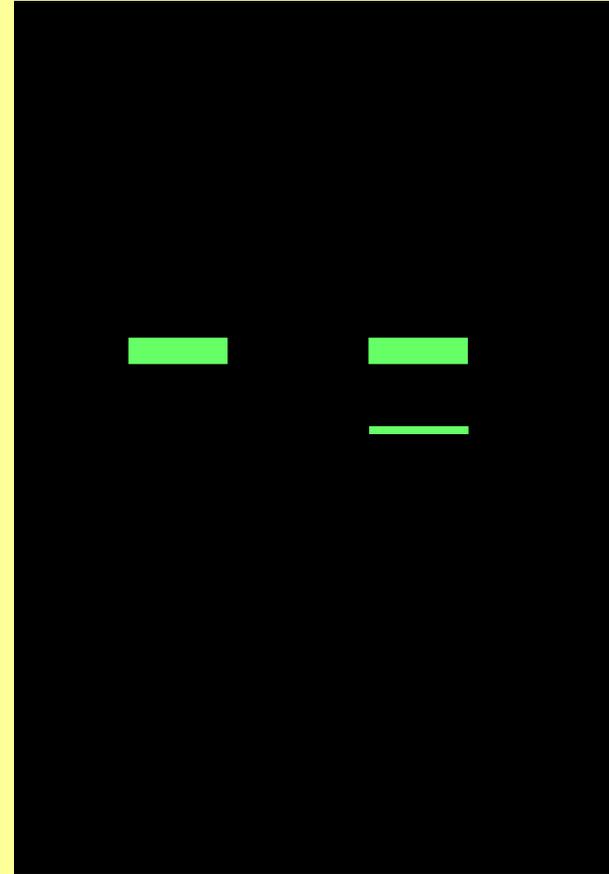
Mutant



IR DYE 700

Normal

Mutant



IR DYE 800

1.0kb

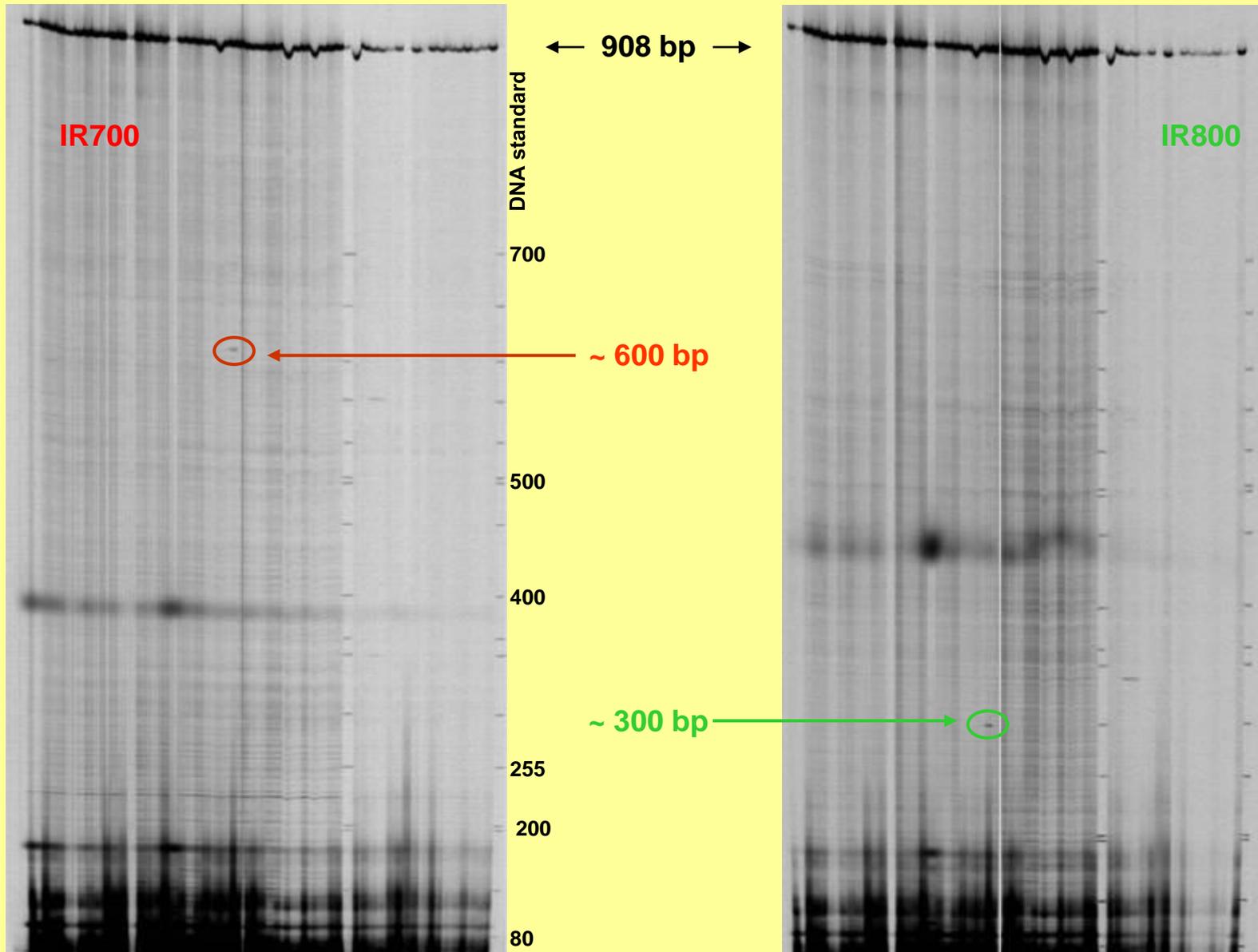
0.8kb

0.2 kb

Li-cor



TILLING barley



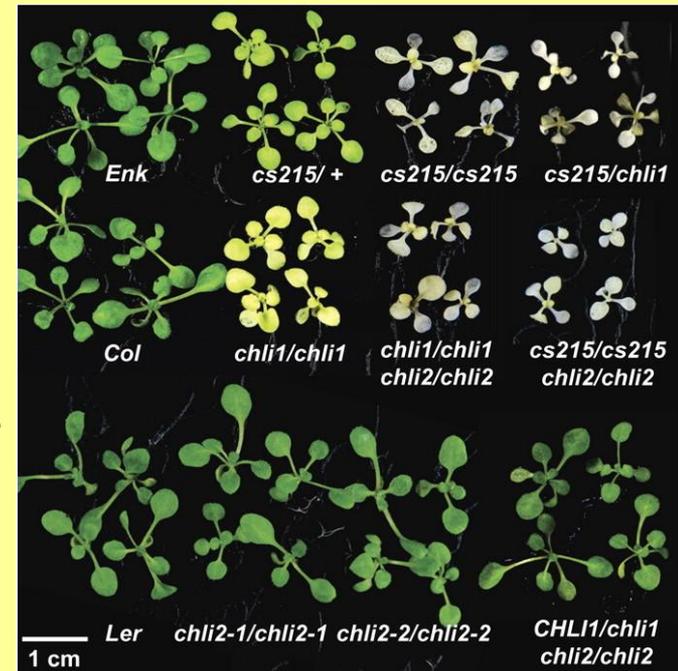
TILLING

Vantaggi:

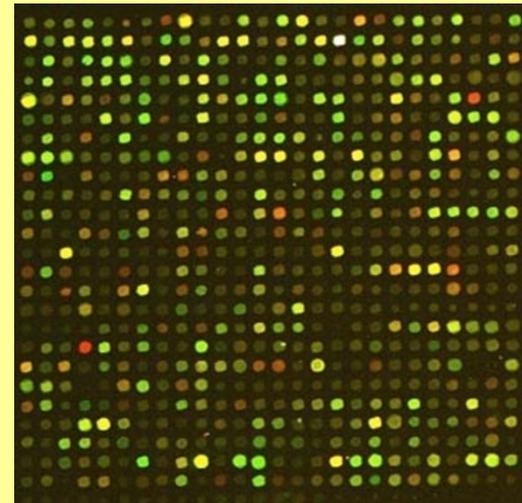
- Trasformazione non necessaria, quindi applicabile a gran parte degli organismi
- Ottenimento di serie alleliche (varianti) e non solo di knock-out
- Molto utile nel caso in cui il silenziamento completo del gene risulta in un fenotipo letale

Il **knock out** di un gene è di solito il modo migliore per comprendere la sua funzione

Quindi è necessario approntare strategie per inattivare il maggior numero possibile di geni e studiare i fenotipi derivanti, per comprendere quali siano le funzioni dei geni quando attivi.



Gli strumenti della genomica funzionale (che è fortemente automatizzata) sono molto importanti per raggiungere questi obiettivi



Analisi fenotipica

La base di qualunque strategia di mutagenesi è l'analisi fenotipica.

Forse alcuni geni, per caso, non avranno subito alcuna mutazione, mentre in altri casi si potrebbero essere prodotti numerosi alleli mutanti.

Alcuni fenotipi mutanti sono evidenziabili solo allo stato omozigote, dove, però non sono fertili; è importante in questi casi, mantenere e propagare la mutazione allo stato eterozigote.

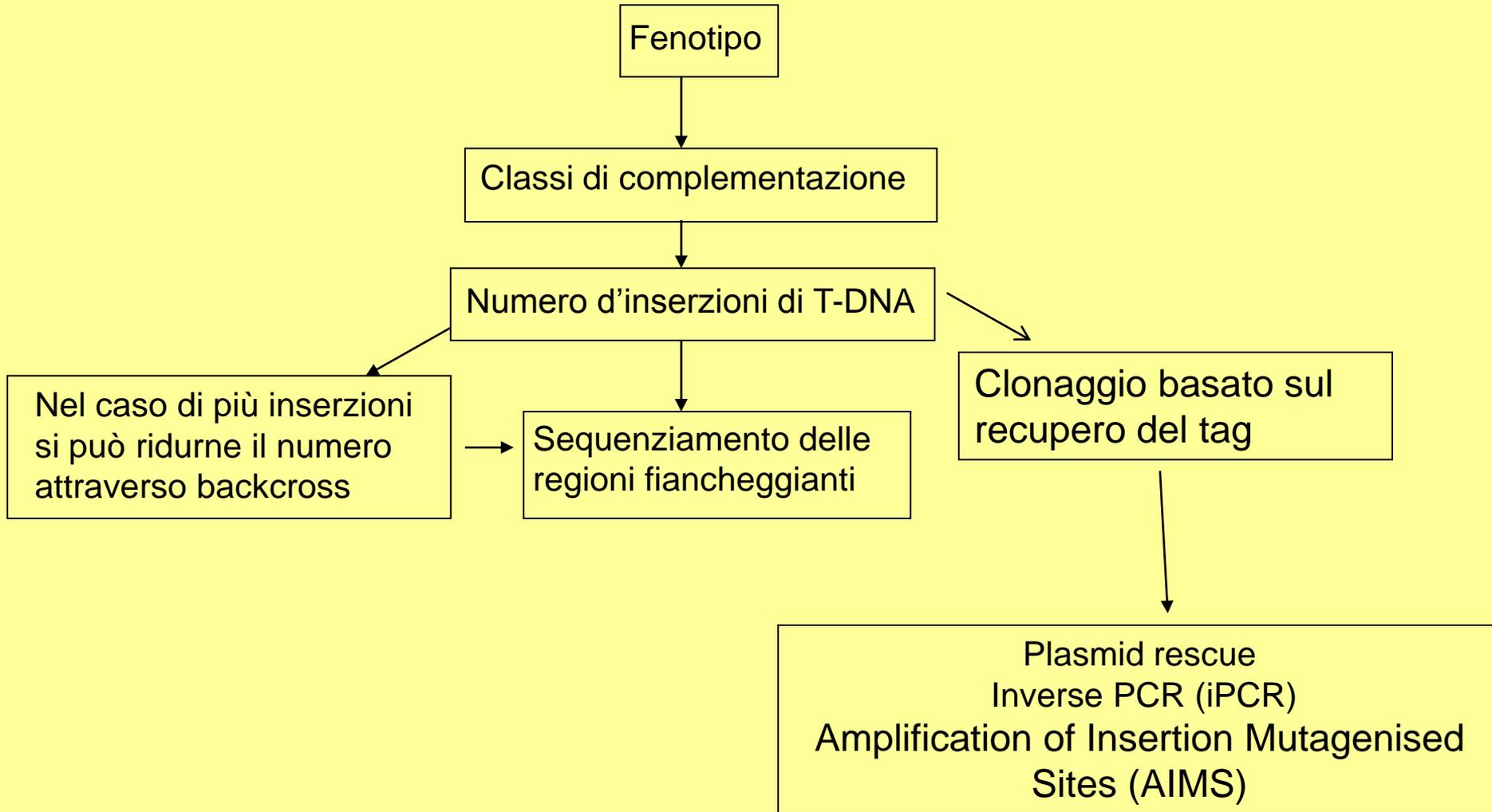
Per identificare geni mutanti si ricorre a “**screening**”, ovvero a condizioni di crescita particolari tali da poter, in qualche modo, selezionare positivamente o negativamente i geni mutati.

In alcuni casi le mutazioni sono facilmente visualizzabili;
es. mutazioni in geni che regolano il tempo di fioritura.

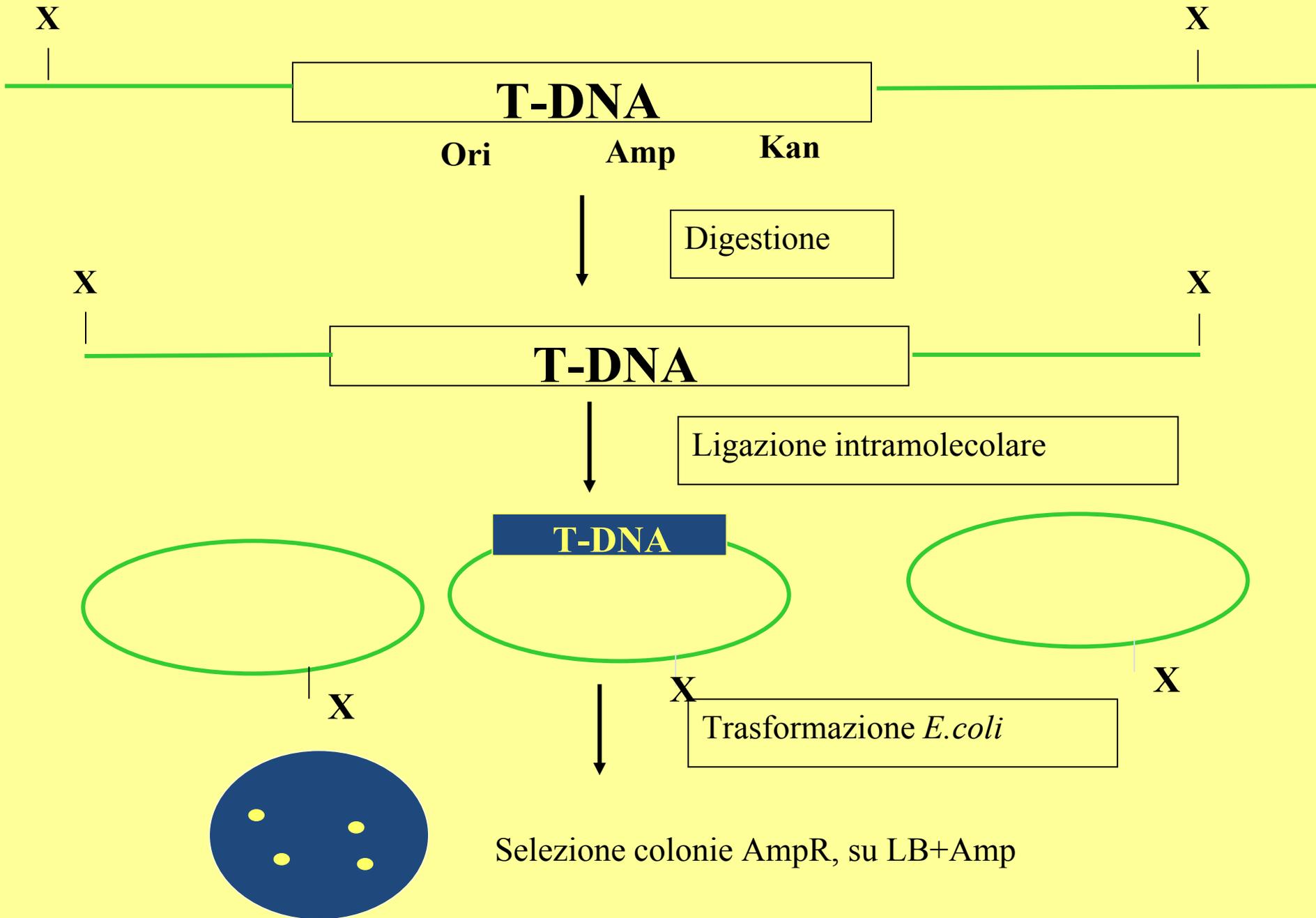
mutazioni in geni legati alla percezione o alla sintesi delle giberrelline, daranno fenotipi nani ecc.



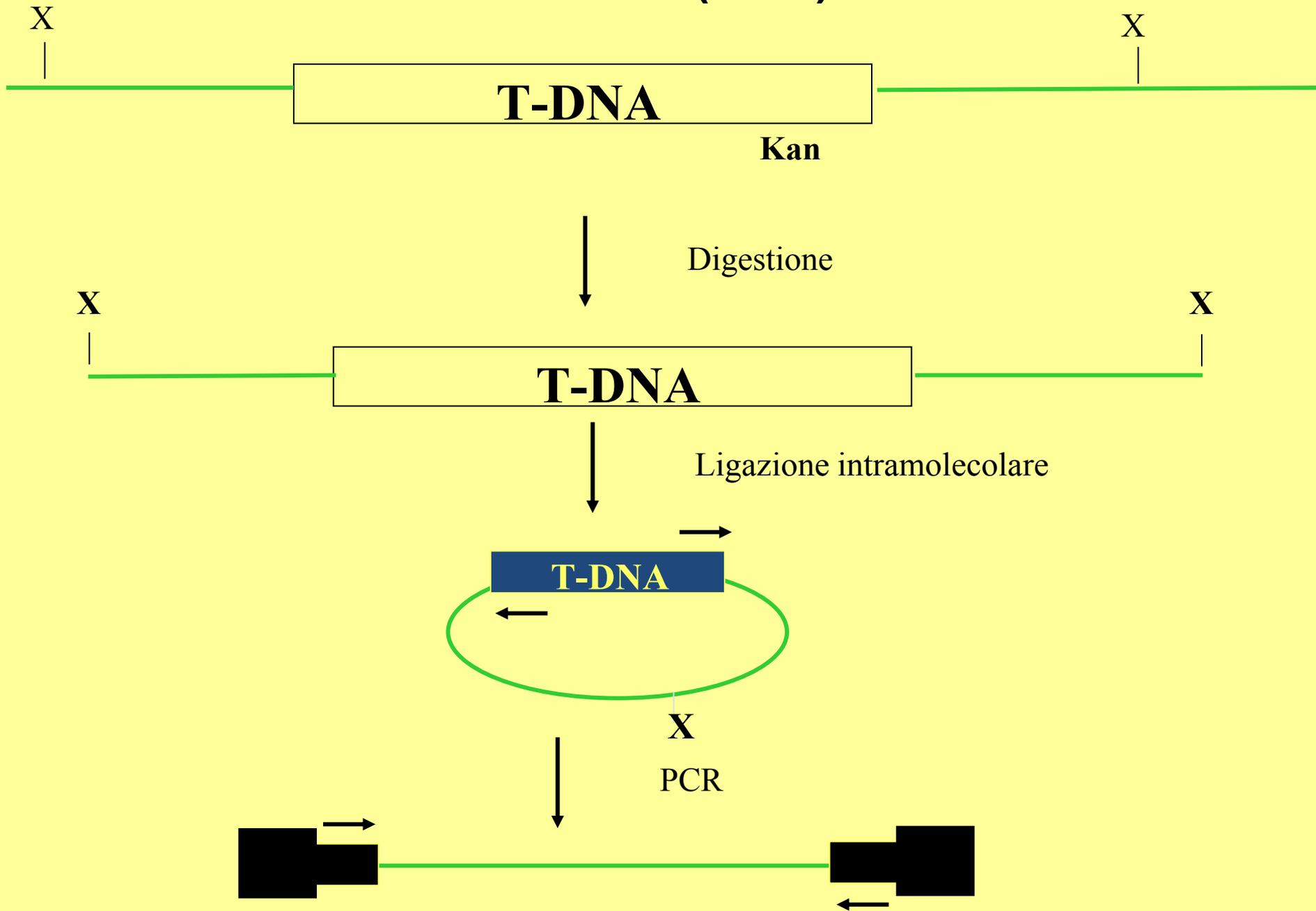
Come si procede una volta identificato il fenotipo?



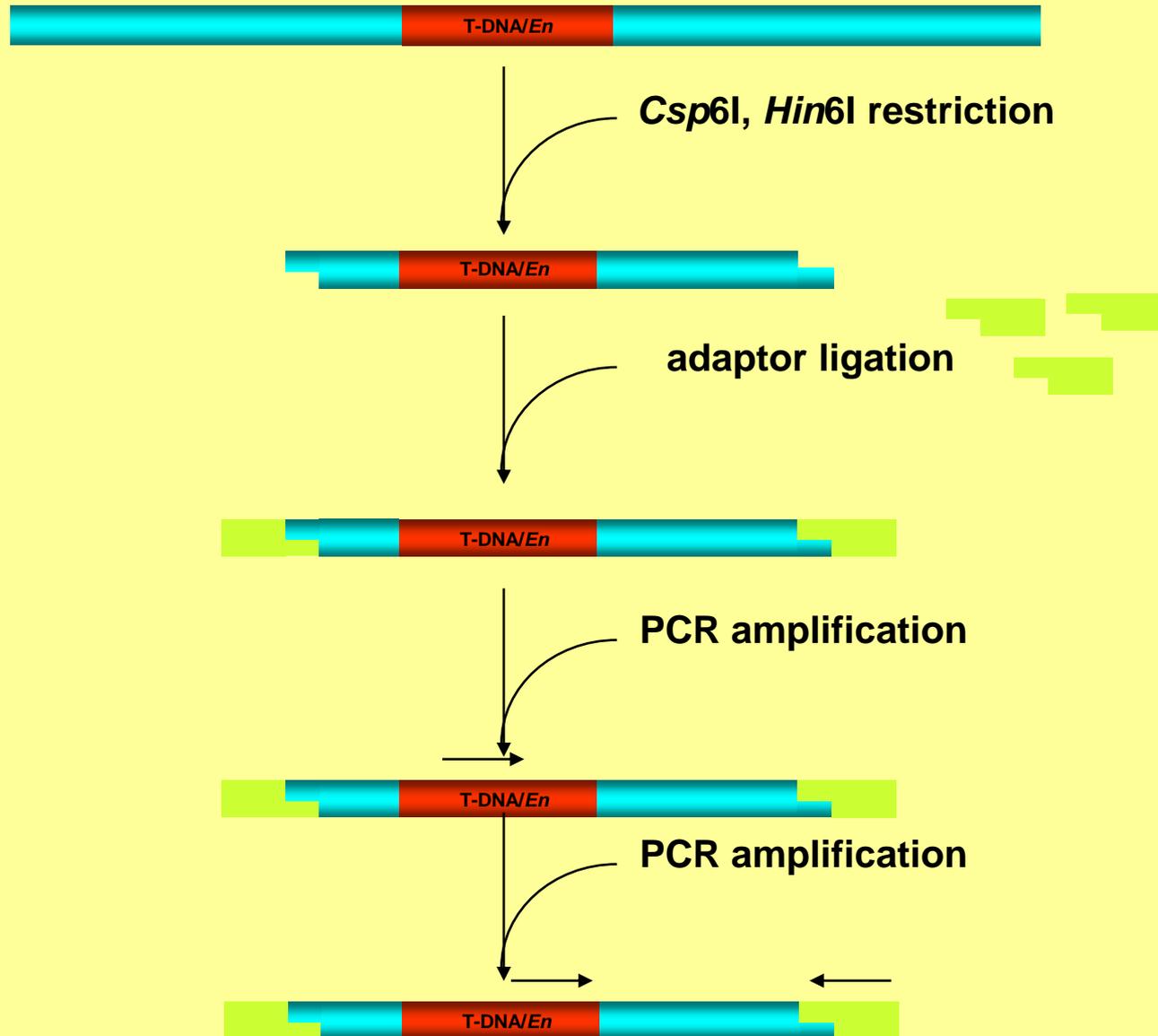
Plasmid Rescue



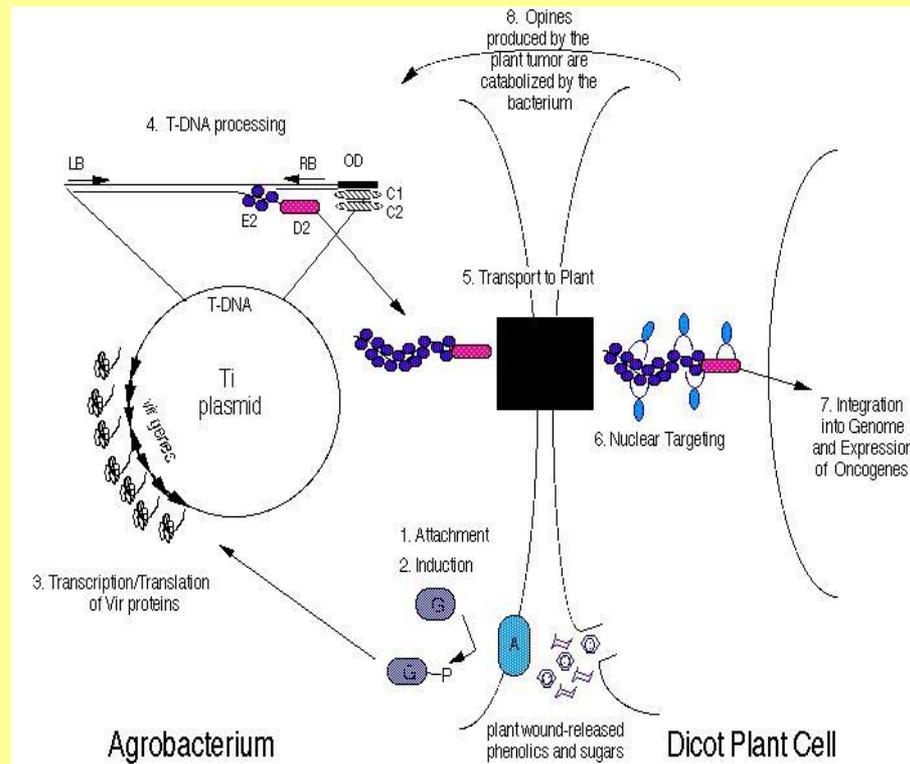
Inverse PCR (iPCR)



Amplification of Insertion Mutagenised Sites (AIMS)



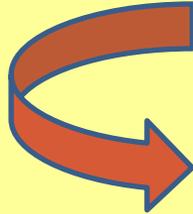
Complementazione con la versione selvatica del gene



La reintroduzione del gene selvatico in un mutante alterato in quello stesso gene dovrebbe ripristinare il fenotipo selvatico.

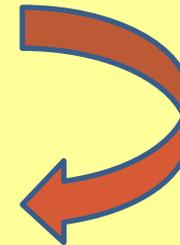
Questa è la prova ultima che il fenotipo che si osserva è dovuto all'alterazione del gene d'interesse

Una volta identificato il gene mutato e il fenotipo della pianta corrispondente,



Si procede con tutta una serie di esperimenti mirati alla caratterizzazione molecolare dell'attività della proteina codificata

Le analisi da fare saranno specifiche per ciascun tipo di gene



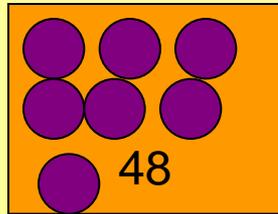
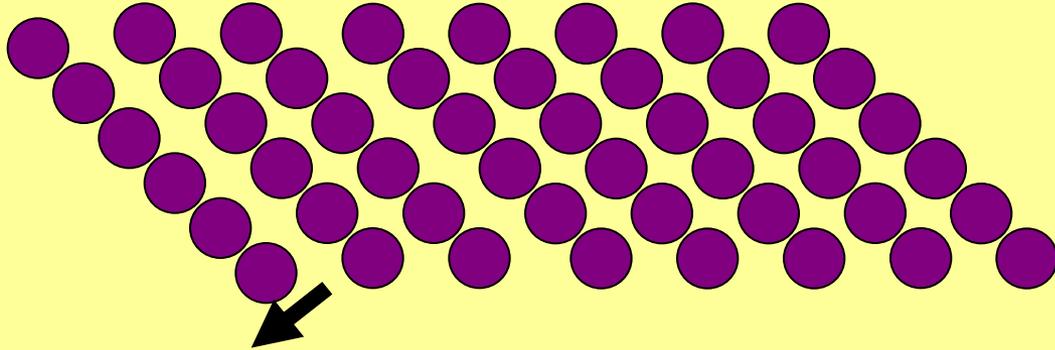
Suggerimenti circa gli esperimenti da fare potranno venire dalle analisi di Trascrittomica, Proteomica e Metabolomica

Strategie di screening per la Reverse Genetics

**Strategia basata sulla creazione di pool di DNA e
analisi PCR per cercare l'inserzione nel gene
d'interesse**

Creazione di pool di DNA

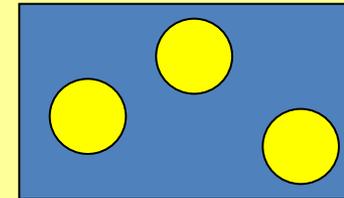
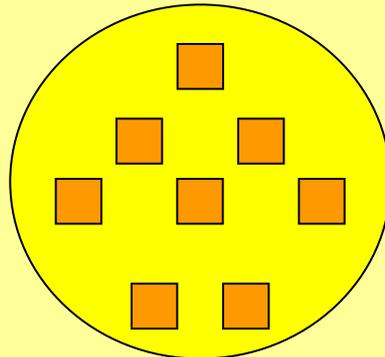
48.384 linee



1008 pools

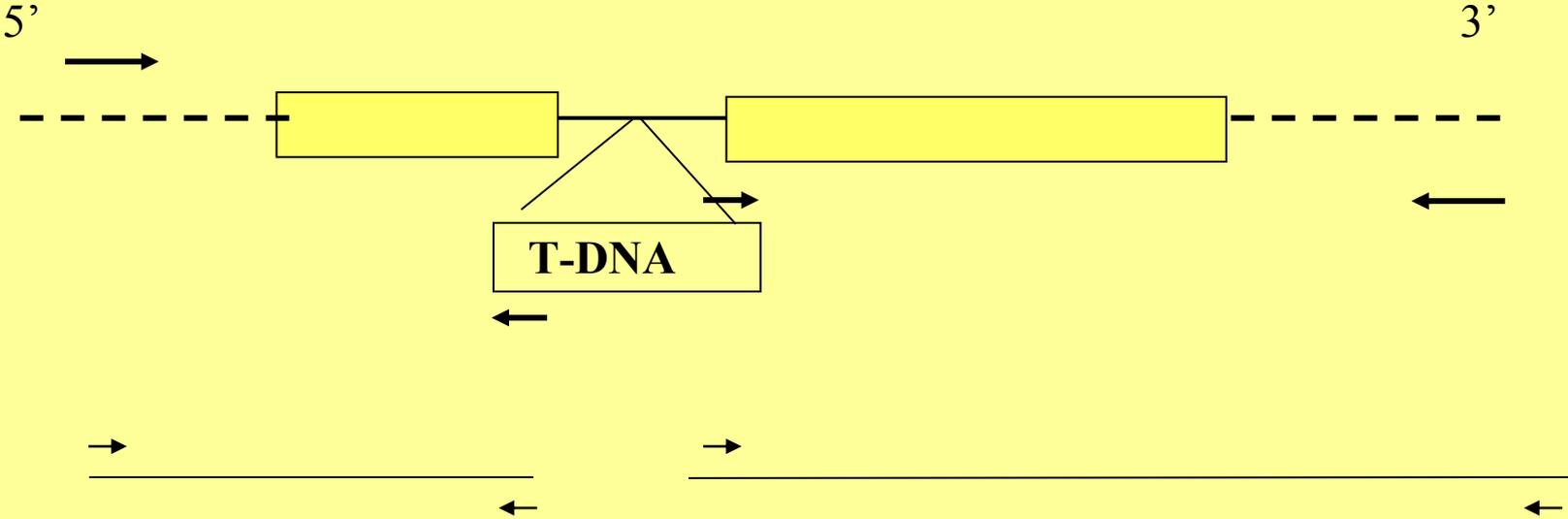
Ogni hyperpool contiene
il DNA di 1152 linee

126 Superpools
Ognuno contenente
8 pools (384 linee)



42 hyperpools
ognuno contenente
3 superpools

Strategia generale di amplificazione per PCR



Procedura di Screening

Valutazione della qualità dei primers:

specificità+sensibilità

background con i primers del T-DNA

Screening per PCR di 42 Hyperpools (HP):

combinazione di un primer gene-specifico (F/R)

con un primer del T-DNA (RB/LB)

Screening per PCR dei 3 Superpools

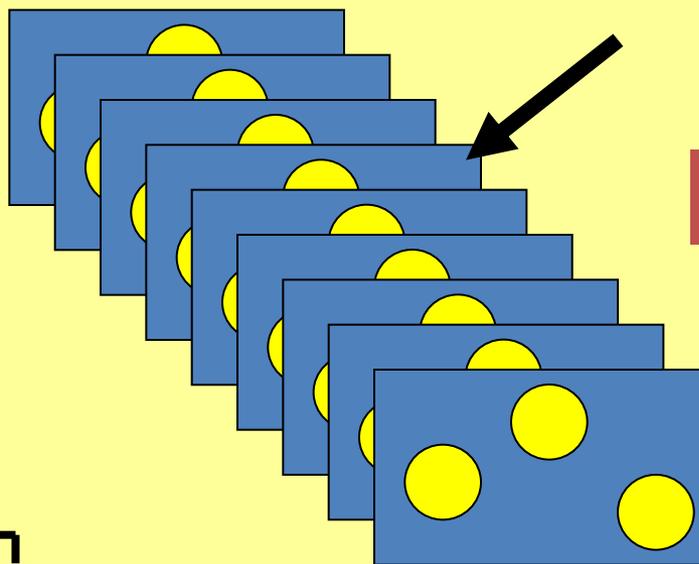
Screening per PCR degli 8 pools

Screening per PCR delle 48 linee

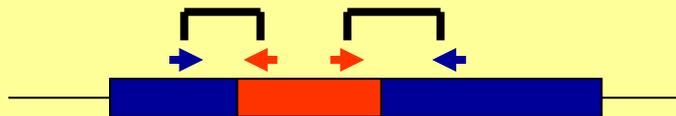
Identificazione della linea

Identificazione della singola pianta

Hyperpool
(3 Superpools)

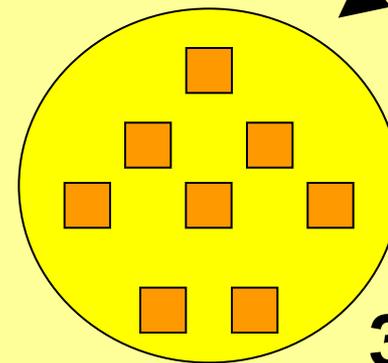
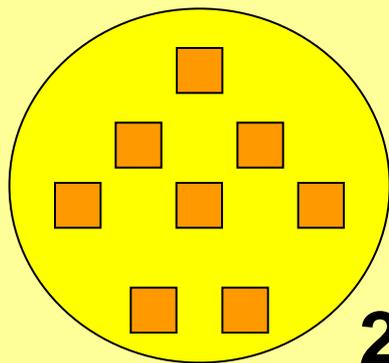
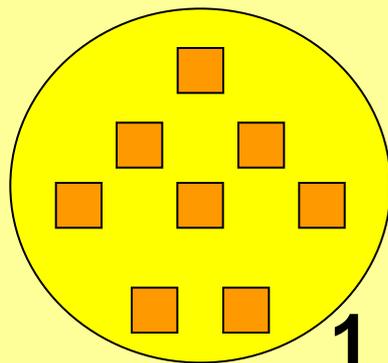


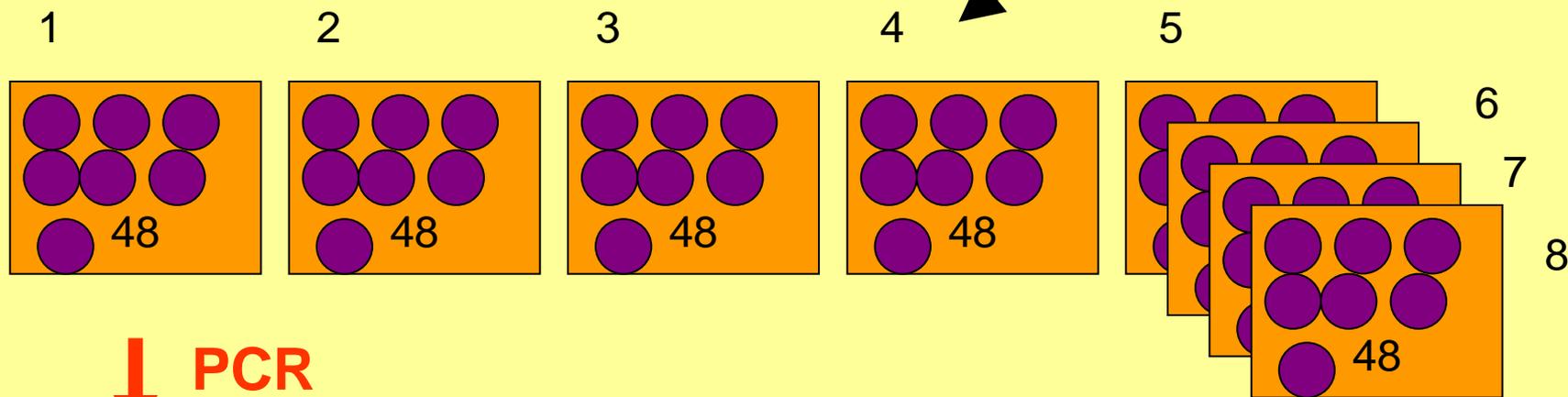
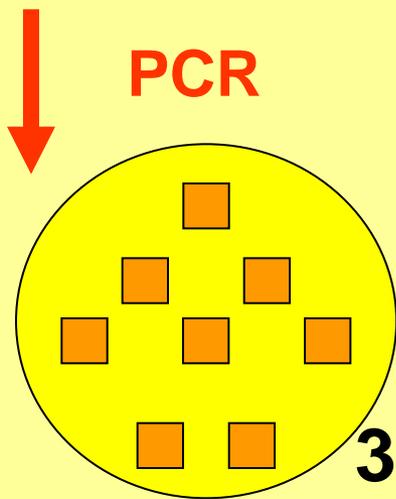
42X



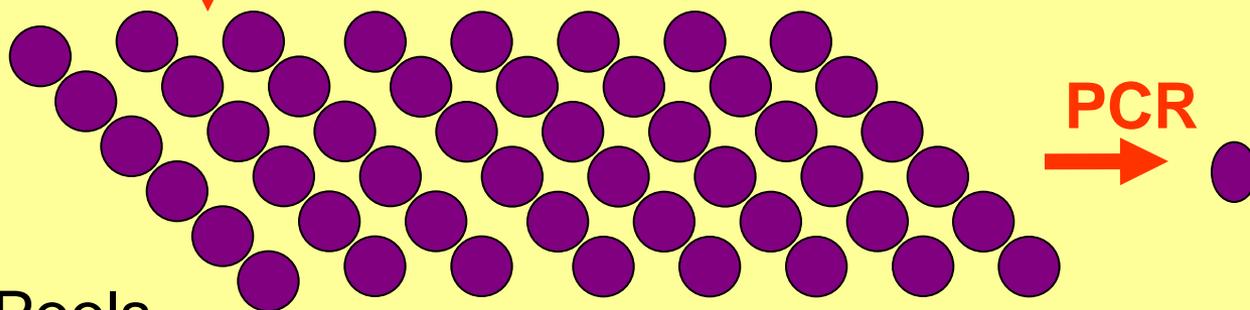
PCR

Superpool
(8 pools)





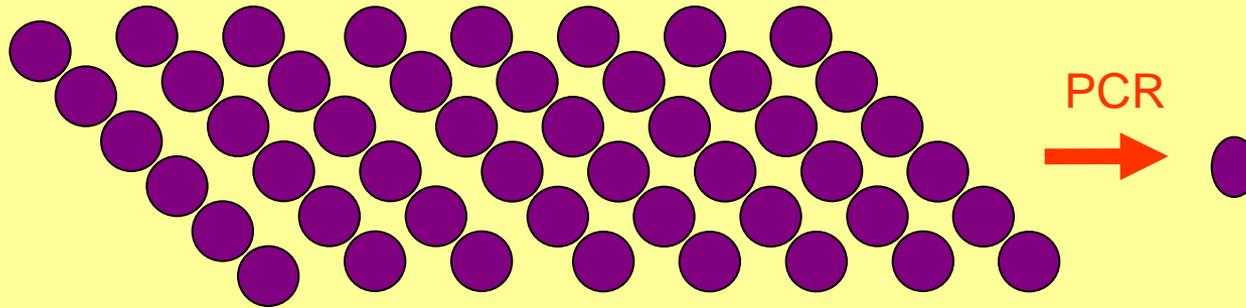
PCR



Mutante dovuto
all'inserzione del
TDNA nel gene di
interesse

Pools
(48 linee)

4



Dagli Hyperpools si arriva sino alla singola pianta

**Insertion
Library**



**DNA Pools
& Super-pools**

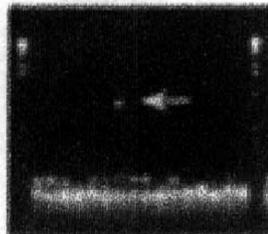


PCR

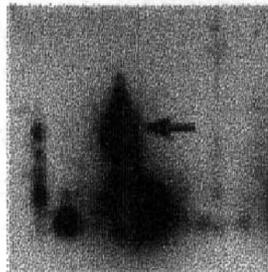


Hybridization

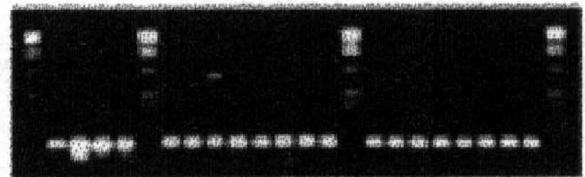
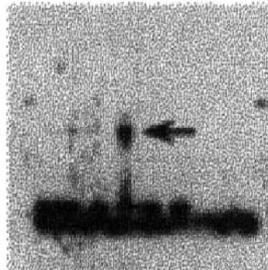
AGAROSE GEL



T-DNA PROBE



GENE PROBE



Sub-pools



Individual lines



Mutant line

Collezioni di Mutanti

Esistono istituzioni internazionali che generano continuamente mutanti messi a disposizione della comunità internazionale.

Il livello di caratterizzazione di questi mutanti è molto basso e di solito si limita alla identificazione di piante resistenti , al recupero e al parziale sequenziamento delle sequenze fiancheggianti il tag.

In molti casi le sequenze dei mutanti presunti sono depositate in apposite banche dati che è possibile interrogare utilizzando programmi dedicati.

Tra le più importanti banche di mutanti ricordiamo la Salk, La Riken, la Gabi-Kat

Attualmente sono state sequenziate più di 400.000 regioni fiancheggianti le inserzioni di T-DNA

Le inserzioni sono localizzate in più del 90% dei geni di Arabidopsis

Quasi tutte le inserzioni possono essere cercate al T-DNAexpress website e i semi corrispondenti possono essere ordinati nei centri di stoccaggio di Nottingham (per l'Europa), e al Arabidopsis Biological Resource Center (ABRC) che si trova all'Ohio State University (per l'America)

Pregi e difetti della Forward e Reverse Genetics

La **Forward Genetics** è più dispendiosa , occorre eseguire uno screening di diverse migliaia di linee mutate e poi isolare il gene responsabile del fenotipo d'interesse.

la **Forward Genetics** garantisce un fenotipo di partenza che è un elemento essenziale per la comprensione della funzione genica

La **Reverse Genetics** è molto più semplice, con un click si possono avere i semi del mutante desiderato. Tuttavia, ciò non garantisce un fenotipo rendendo quindi la caratterizzazione della funzione genica molto più complicata

Riso *Oryza sativa*



- è alimento base per >3 miliardi di persone, e fornisce il 50-80% del loro apporto calorico
- la domanda in Asia è prevista crescere del 25%
- è modello per altre importanti colture
- diploide
- dimensione del genoma: 430 Mb
- 12 cromosomi
- *si può trasformare con agrobatterio*

The rice sequencing story

1997/1998: **International Rice Genome Sequencing Project (IRGSP)** formed;

aim to sequence japonica rice by 2008 using **clone-by-clone shotgun** approach.

Apr. 2000: **Monsanto** announces draft sequence.

Jan. 2001: **Syngenta** announces that it used **shotgun** approach to sequence *japonica* rice.

Oct. 2001: **The Beijing Genomics Institute (BGI)** completes sequencing of *indica* rice using **whole-genome shotgun** approach.

RISO: La specie modello per lo studio del genoma dei cereali



Specie coltivata, autogama

Rapido ciclo vitale (in serra 90 giorni da seme a seme)

Il genoma completamente sequenziato (Novembre 2002) da un consorzio internazionale di laboratori di ricerca (Giappone, Brasile, USA, Cina, Francia)

Forte similarità con la struttura del genoma di altri cereali (mais, grano duro/grano tenero sorgo, miglio)

A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *japonica*)

Stephen A. Goff,^{1*} Darrell Ricke,¹ Tien-Hung Lan,¹
Gernot Presting,¹ Ronglin Wang,¹ Molly Dunn,¹
Jane Glazebrook,¹ Allen Sessions,¹ Paul Oeller,¹ Hemant Varma,¹
David Hadley,¹ Don Hutchison,¹ Chris Martin,¹ Fumiaki Katagiri,¹
B. Markus Lange,¹ Todd Moughamer,¹ Yu Xia,¹ Paul Budworth,¹
Jingping Zhong,¹ Trini Miguel,¹ Uta Paszkowski,¹ Shiping Zhang,¹
Michelle Colbert,¹ Wei-lin Sun,¹ Lili Chen,¹ Bret Cooper,¹
Sylvia Park,¹ Todd Charles Wood,² Long Mao,³ Peter Quail,⁴
Rod Wing,⁵ Ralph Dean,⁵ Yeisoo Yu,⁵ Andrey Zharkikh,⁶
Richard Shen,^{6†} Sudhir Sahasrabudhe,⁶ Alun Thomas,⁶
Rob Cannings,⁶ Alexander Gutin,⁶ Dmitry Pruss,⁶ Julia Reid,⁶
Sean Tavtigian,⁶ Jeff Mitchell,⁶ Glenn Eldredge,⁶ Terri Scholl,⁶
Rose Mary Miller,⁶ Satish Bhatnagar,⁶ Nils Adey,⁶
Todd Rubano,^{6†} Nadeem Tusneem,⁶ Rosann Robinson,⁶
Jane Feldhaus,⁶ Teresita Macalma,⁶ Arnold Oliphant,^{6†}
Steven Briggs¹

The genome of the japonica subspecies of rice, an important cereal and model monocot, was sequenced and assembled by whole-genome shotgun sequencing. The assembled sequence covers 93% of the 420-megabase genome. Gene predictions on the assembled sequence suggest that the genome contains 32,000 to 50,000 genes. Homologs of 98% of the known maize, wheat, and barley proteins are found in rice. Synteny and gene homology between rice and the other cereal genomes are extensive, whereas synteny with *Arabidopsis* is limited. Assignment of candidate rice orthologs to *Arabidopsis* genes is possible in many cases. The rice genome sequence provides a foundation for the improvement of cereals, our most important crops.

A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. *indica*)

Jun Yu,^{1,2,3,4*} Songnian Hu,^{1*} Jun Wang,^{1,2,5*}
Gane Ka-Shu Wong,^{1,2,4*} Songgang Li,^{1,5} Bin Liu,¹ Yajun Deng,^{1,6}
Li Dai,¹ Yan Zhou,^{2,7} Xiuqing Zhang,^{1,3} Mengliang Cao,⁸ Jing Liu,²
Jiandong Sun,¹ Jiabin Tang,^{1,3} Yanjiong Chen,^{1,6}
Xiaobing Huang,¹ Wei Lin,² Chen Ye,¹ Wei Tong,¹ Lijuan Cong,¹
Jianing Geng,¹ Yujun Han,¹ Lin Li,¹ Wei Li,^{1,9} Guangqiang Hu,¹
Xiangang Huang,¹ Wenjie Li,¹ Jian Li,¹ Zhanwei Liu,¹ Long Li,¹
Jianping Liu,¹ Qihui Qi,¹ Jinsong Liu,¹ Li Li,¹ Tao Li,¹
Xuegang Wang,¹ Hong Lu,¹ Tingting Wu,¹ Miao Zhu,¹
Peixiang Ni,¹ Hua Han,¹ Wei Dong,^{1,3} Xiaoyu Ren,¹
Xiaoli Feng,^{1,3} Peng Cui,¹ Xianran Li,¹ Hao Wang,¹ Xin Xu,¹
Wenxue Zhai,³ Zhao Xu,¹ Jinsong Zhang,³ Sijie He,³
Jianguo Zhang,¹ Jichen Xu,³ Kunlin Zhang,^{1,5} Xianwu Zheng,³
Jianhai Dong,² Wanyong Zeng,³ Lin Tao,² Jia Ye,² Jun Tan,²
Xide Ren,¹ Xuwei Chen,³ Jun He,² Daofeng Liu,³ Wei Tian,^{2,6}
Chaoguang Tian,¹ Hongai Xia,¹ Qiyu Bao,¹ Gang Li,¹ Hui Gao,¹
Ting Cao,¹ Juan Wang,¹ Wenming Zhao,¹ Ping Li,³ Wei Chen,¹
Xudong Wang,³ Yong Zhang,^{1,5} Jianfei Hu,^{1,5} Jing Wang,^{1,5}
Song Liu,¹ Jian Yang,¹ Guangyu Zhang,¹ Yuqing Xiong,¹ Zhijie Li,¹
Long Mao,³ Chengshu Zhou,⁸ Zhen Zhu,³ Runsheng Chen,^{1,9}
Bailin Hao,^{2,10} Weimou Zheng,^{1,10} Shouyi Chen,³ Wei Guo,¹¹
Guojie Li,¹² Siqi Liu,^{1,2} Ming Tao,^{1,2} Jian Wang,^{1,2} Lihuang Zhu,^{3†}
Longping Yuan,^{8†} Huanming Yang^{1,2,3†}

We have produced a draft sequence of the rice genome for the most widely cultivated subspecies in China, *Oryza sativa* L. ssp. *indica*, by whole-genome shotgun sequencing. The genome was 466 megabases in size, with an estimated 46,022 to 55,615 genes. Functional coverage in the assembled sequences was 92.0%. About 42.2% of the genome was in exact 20-nucleotide oligomer repeats, and most of the transposons were in the intergenic regions between genes. Although 80.6% of predicted *Arabidopsis thaliana* genes had a homolog in rice, only 49.4% of predicted rice genes had a homolog in *A. thaliana*. The large proportion of rice genes with no recognizable homologs is due to a gradient in the GC content of rice coding sequences.

Predizione della classificazione dei geni trovati in riso

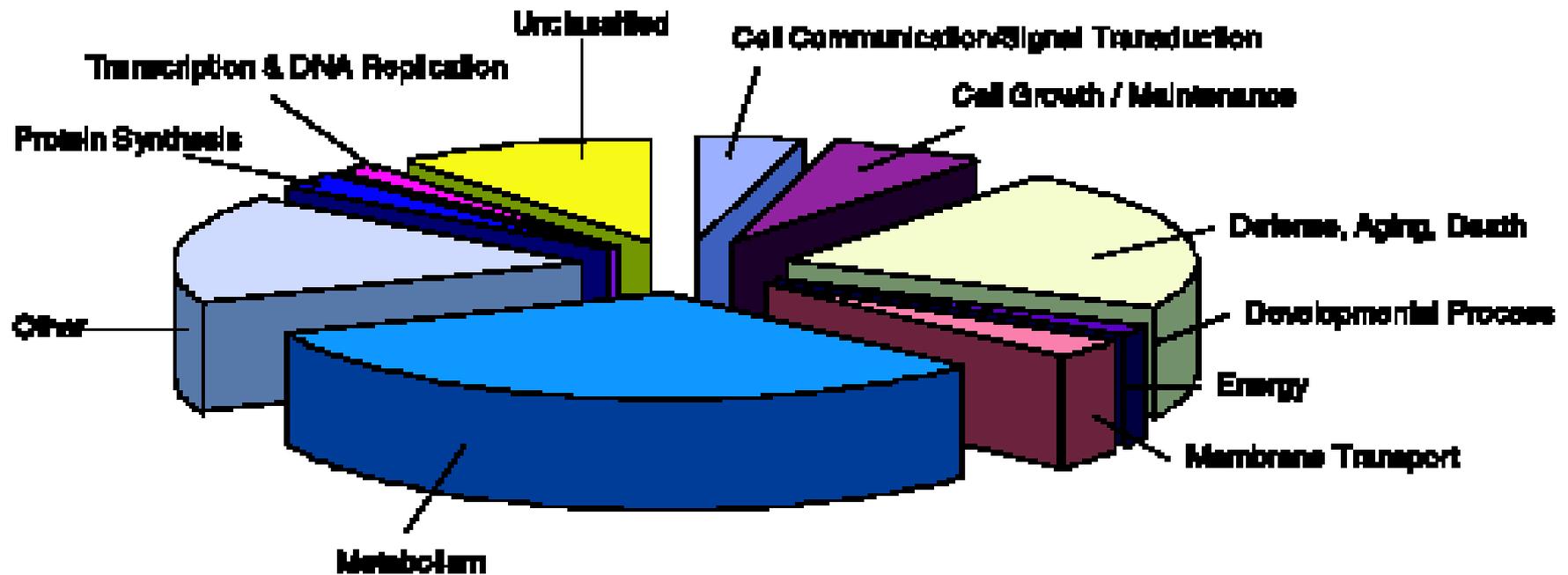
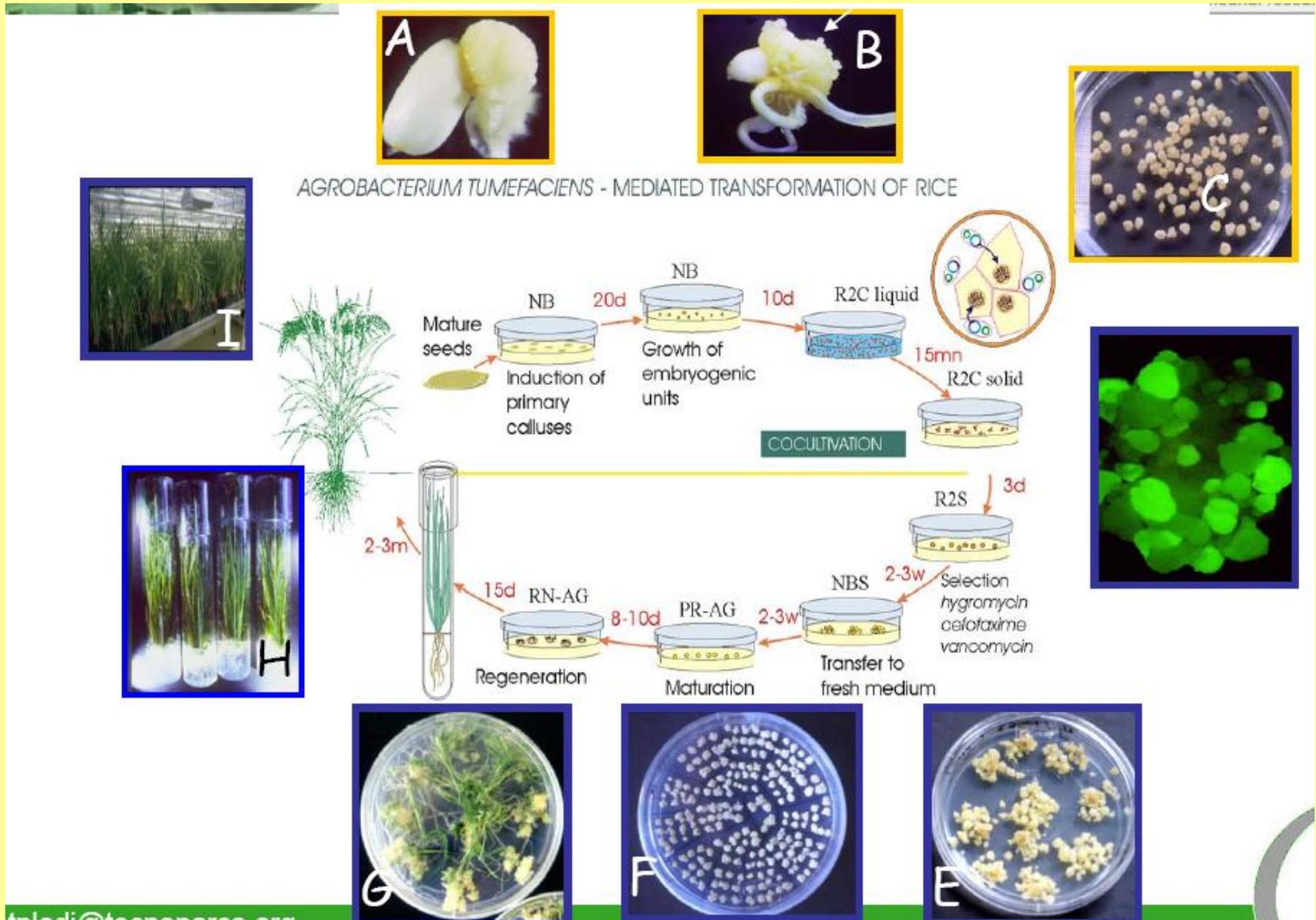


Fig. 1. Rice gene prediction classifications. *HMLgenes*³⁰⁰ were classified with Interpro and GO software (27–29); the categories generated are shown.

Highlights of rice genome sequencing: genome comparisons

- 85% of the *Arabidopsis* genes have an homolog in rice, but more than 50% of predicted rice genes had no homolog in *Arabidopsis*. 98% of the genes known in cereals are found in the rice genome
- 22% of the IRGSP's japonica sequence is not recognized in the incomplete BGI's indica sequence. Frequency of SNPs is about 1 every 268 bp + Presence of InDels.
- 8,000 genes shared by Rice and *Arabidopsis* do not exist in other organisms
- Order of genes in blocks of few hundred kbp preserved between Rice and *Arabidopsis* but with lacking/additional genes in each block.

La trasformazione genetica in riso facilita la caratterizzazione della funzione genica



The Genome of Black Cottonwood, *Populus trichocarpa* (Torr. & Gray)

G. A. Tuskan,^{1,3*} S. DiFazio,^{1,4†} S. Jansson,^{5†} J. Bohlmann,^{6†} I. Grigoriev,^{9†} U. Hellsten,^{9†} N. Putnam,^{9†} S. Ralph,^{6†} S. Rombauts,^{10†} A. Salamov,^{9†} J. Schein,^{11†} L. Sterck,^{10†} A. Aerts,⁹ R. R. Bhalerao,⁵ R. P. Bhalerao,¹² D. Blaudez,¹³ W. Boerjan,¹⁰ A. Brun,¹³ A. Brunner,¹⁴ V. Busov,¹⁵ M. Campbell,¹⁶ J. Carlson,¹⁷ M. Chalot,¹³ J. Chapman,⁹ G.-L. Chen,² D. Cooper,⁶ P. M. Coutinho,¹⁹ J. Couturier,¹³ S. Covert,²⁰ Q. Cronk,⁷ R. Cunningham,¹ J. Davis,²² S. Degroeve,¹⁰ A. Déjardin,²³ C. dePamphilis,¹⁸ J. Detter,⁹ B. Dirks,²⁴ I. Dubchak,^{9,25} S. Duplessis,¹³ J. Ehlting,⁷ B. Ellis,⁶ K. Gendler,²⁶ D. Goodstein,⁹ M. Gribskov,²⁷ J. Grimwood,²⁸ A. Groover,²⁹ L. Gunter,¹ B. Hamberger,⁷ B. Heinze,³⁰ Y. Helariutta,^{12,31,33} B. Henrissat,¹⁹ D. Holligan,²¹ R. Holt,¹¹ W. Huang,⁹ N. Islam-Faridi,³⁴ S. Jones,¹¹ M. Jones-Rhoades,³⁵ R. Jorgensen,²⁶ C. Joshi,¹⁵ J. Kangasjärvi,³² J. Karlsson,⁵ C. Kelleher,⁶ R. Kirkpatrick,¹¹ M. Kirst,²² A. Kohler,¹³ U. Kalluri,¹ F. Larimer,² J. Leebens-Mack,²¹ J.-C. Leplé,²³ P. Locascio,² Y. Lou,⁹ S. Lucas,⁹ F. Martin,¹³ B. Montanini,¹³ C. Napoli,²⁶ D. R. Nelson,³⁶ C. Nelson,³⁷ K. Nieminen,³¹ O. Nilsson,¹² V. Pereda,¹³ G. Peter,²² R. Philippe,⁶ G. Pilate,²³ A. Poliakov,²⁵ J. Razumovskaya,² P. Richardson,⁹ C. Rinaldi,¹³ K. Ritland,⁸ P. Rouzé,¹⁰ D. Ryaboy,²⁵ J. Schmutz,²⁸ J. Schrader,³⁸ B. Segerman,⁵ H. Shin,¹¹ A. Siddiqui,¹¹ F. Sterky,³⁹ A. Terry,⁹ C.-J. Tsai,¹⁵ E. Uberbacher,² P. Unneberg,³⁹ J. Vahala,³² K. Wall,¹⁸ S. Wessler,²¹ G. Yang,²¹ T. Yin,¹ C. Douglas,^{7‡} M. Marra,^{11‡} G. Sandberg,^{12‡} Y. Van de Peer,^{10‡} D. Rokhsar^{9,24‡}

We report the draft genome of the black cottonwood tree, *Populus trichocarpa*. Integration of shotgun sequence assembly with genetic mapping enabled chromosome-scale reconstruction of the genome. More than 45,000 putative protein-coding genes were identified. Analysis of the assembled genome revealed a whole-genome duplication event; about 8000 pairs of duplicated genes from that event survived in the *Populus* genome. A second, older duplication event is indistinguishably coincident with the divergence of the *Populus* and *Arabidopsis* lineages. Nucleotide substitution, tandem gene duplication, and gross chromosomal rearrangement appear to proceed substantially more slowly in *Populus* than in *Arabidopsis*. *Populus* has more protein-coding genes than *Arabidopsis*, ranging on average from 1.4 to 1.6 putative *Populus* homologs for each *Arabidopsis* gene. However, the relative frequency of protein domains in the two genomes is similar. Overrepresented exceptions in *Populus* include genes associated with lignocellulosic wall biosynthesis, meristem development, disease resistance, and metabolite transport.

45.000 putativi geni codificanti proteine

Evento di duplicazione dell'intero genoma

8000 geni presenti in duplice copia

Duplicazione di geni in tandem,
sostituzioni nucleotidiche,
riarrangiamenti cromosomici si sono verificati più lentamente in
pioppo rispetto ad Arabidopsis

n. > di geni in pioppo rispetto ad Arabidopsis

Sviluppo di meristema secondario

Sistema di trasporto

Resistenza ai patogeni

GENOMA DEL PIOPPO



- 45.000 geni codificanti proteine
- 485 Mb
- Trasformazione
- Raggiunge la maturità riproduttiva in 4-6 anni
- 91% dei geni di Arabidopsis sono omologhi a geni del pioppo
- Un notevole aumento nel numero di geni codificanti cellulosa e lignina

GENOMA DEL PIOPPO



Disease resistance. The likelihood that a perennial plant will encounter a pathogen or herbivore before reproduction is near unity. The long-generation intervals for trees make it difficult for such plants to match the evolutionary rates of a microbial or insect pest. Aside from the formation of thickened cell walls and the synthesis of secondary metabolites that constitute a first line of defense against microbial and insect pests, plants use a variety of disease-resistance (*R*) genes.



GENOMA DEL PIOPPO

Molto informativo rispetto a quello che possiamo imparare da *Arabidopsis*

Formazione del legno,
Movimento di acqua e nutrienti
Resistenza a patogeni in specie perenne

GENOMA DELLA VITE (2007)



The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla

The French–Italian Public Consortium for Grapevine Genome Characterization*

The analysis of the first plant genomes provided unexpected evidence for genome duplication events in species that had previously been considered as true diploids on the basis of their genetics^{1–3}. These polyploidization events may have had important consequences in plant evolution, in particular for species radiation and adaptation and for the modulation of functional capacities^{4–10}. Here we report a high-quality draft of the genome sequence of grapevine (*Vitis vinifera*) obtained from a highly homozygous genotype. The draft sequence of the grapevine genome is the fourth one produced so far for flowering plants, the second for a woody species and the first for a fruit crop (cultivated for both fruit and beverage). Grapevine was selected because of its important place in the cultural heritage of humanity beginning during the Neolithic period¹¹. Several large expansions of gene families with roles in aromatic features are observed. The grapevine genome has not undergone recent genome duplication, thus enabling the discovery of ancestral traits and features of the genetic organization of flowering plants. This analysis reveals the contribution of three ancestral genomes to the grapevine haploid content. This ancestral arrangement is common to many dicotyledonous plants but is absent from the genome of rice, which is a monocotyledon. Furthermore, we explain the chronology of previously described whole-genome duplication events in the evolution of flowering plants.

All grapevine varieties are highly heterozygous; preliminary data showed that there was as much as 13% sequence divergence between alleles, which would hinder reliable contig assembly when a whole-genome shotgun strategy was used for sequencing. Our consortium therefore selected the grapevine PN40024 genotype for sequencing. This line, originally derived from Pinot Noir, has been bred close to full homozygosity (estimated at about 93%) by successive selfings, permitting a high-quality whole-genome shotgun assembly.

A total of 6.2 million end-reads were produced by our consortium, representing an 8.4-fold coverage of the genome. Within the assembly, performed with Arachne¹², 316 supercontigs represent putative allelic haplotypes that constitute 11.6 million bases (Mb). These values are in good fit with the 7% residual heterozygosity of PN40024 assessed by using genetic markers. When considering only one of the haplotypes in each heterozygous region, the assembly (Table 1a) consists of 19,577 contigs ($N_{50} = 65.9$ kilobases (kb)), where N_{50} corresponds to the size of the shorter supercontig or contig in a subset representing half of the assembly size) and 3,514 supercontigs ($N_{50} = 2.07$ Mb) totalling 487 Mb. This value is close to the 475 Mb previously reported for the grapevine genome size³.

Using a set of 409 molecular markers from the reference grapevine map⁴, 69% of the assembled 487 Mb, arranged into 45 ultracontigs

Table 1 | Global statistics on the genome of *Vitis vinifera*

(a) Assembly						
	Status	Number	N_{50} (kb)	Longest (kb)	Size (Mb)	Percentage of the assembly
Contigs	All	19,577	65.9	557	467.5	–
Supercontigs	All	3,514	2,065	12,675	487.1	100
	Anchored on chromosomes	191	3,189	12,675	335.6	68.9
	Anchored on chromosomes and oriented	143	3,827	12,675	296.9	60.9
(b) Annotation						
	Number	Median size (bp)	Total length (Mb)	Percentage of the genome	%GC	
Gene	30,434	3,399	225.6	46.3	36.2	
Exons CDS	149,351	130	33.6	6.9	44.5	
Introns CDS	118,917	213	178.6	36.7	34.7	
Intergenic	30,453	3,544	261.5	34.7	33.0	
tRNA*	600	73	0.04	NS	43.0	
miRNA†	164	103.5	0.002	NS	35.9	
(c) Orthology						
	Number of orthologous proteins		Mean identity (%)			
<i>P. trichocarpa</i>	12,996		72.7			
<i>A. thaliana</i>	11,404		65.5			
<i>O. sativa</i>	9,731		59.8			
Common to eudicotyledons‡	10,547					
Common to Magnoliophyta§	8,121					

*Transfer RNA (tRNA) values were computed on exons.

†Micro RNAs (miRNAs) are members of known conserved miRNA families.

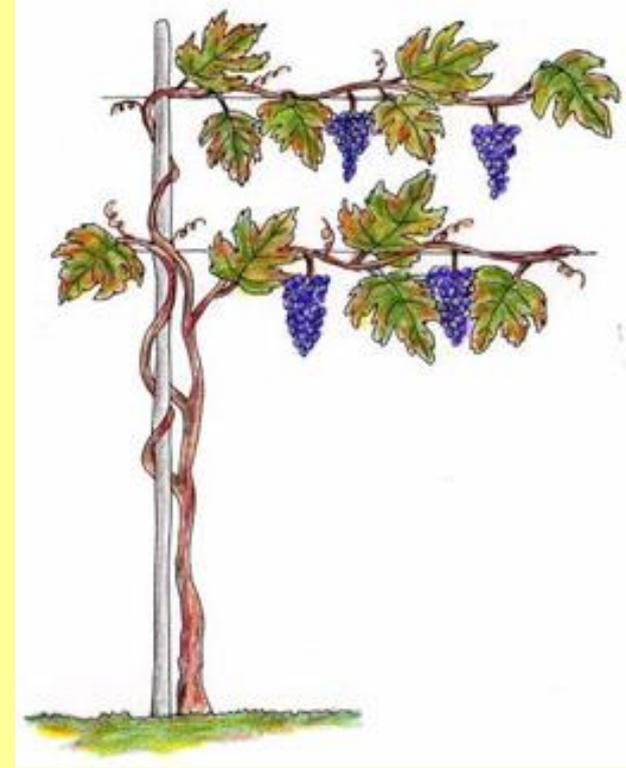
‡Eudicotyledons are represented by *P. trichocarpa* and *A. thaliana*.

§Magnoliophyta (most flowering plants) are represented by *P. trichocarpa*, *A. thaliana* and *O. sativa*.

*A list of participants and their affiliations appears at the end of the paper.

GENOMA DELLA VITE

- Specie coltivata diploide
- Piccolo genoma - 475 Mb
- 30.434 geni codificanti proteine
- 19 Cromosomi



• “A striking feature of the grapevine genome lies in the existence of large families related to wine characteristics, which have a higher gene copy number than in other sequenced plants” (Terpenoid genes are 89 in Grapevine while in rice, poplar and Arabidopsis are 30-40)

this species, including domestication traits. A selective amplification of genes belonging to the metabolic pathways of terpenes and tannins has occurred in the grapevine genome, in contrast with other plant genomes. This suggests that it may become possible to trace the diversity of wine flavours down to the genome level. Grapevine is also a crop that is highly susceptible to a large diversity of pathogens including powdery mildew, oidium and Pierce disease. Other *Vitis* species such as *V. riparia* or *V. cinerea*, which are known to be resistant to several of these pathogens, are interfertile with *V. vinifera* and can be used for the introduction of resistance traits by advanced backcrosses²⁷ or by gene transfer. Access to the *Vitis* sequence and the exploitation of synteny will speed up this process of introgression of pathogen resistance traits. As a consequence of this, it is hoped that it will also prompt a strong decrease in pesticide use.

Confronto:

Genomi delle piante e Genomi degli animali

I genomi delle piante contengono numerose classi di geni assenti o scarsamente rappresentati nei genomi animali.

I prodotti di questi geni “specifici” delle piante comprendono:

gli enzimi richiesti per la biosintesi della parete cellulare

Alcune proteine di trasporto, che spostano tra una cellula e l'altra nutrienti, composti tossici, metaboliti, proteine e ac. nucleici

Alcuni enzimi e altre macromolecole necessarie per la fotosintesi, come la Rubisco e le proteine di trasporto degli elettroni

I prodotti che sono coinvolti nel turgore cellulare e nelle risposte tipiche di un sistema di vita sessile, come il fototropismo ed il geotropismo

Numerosi enzimi e citocromi coinvolti nella produzione di centinaia di migliaia di metaboliti secondari delle piante in fioritura

Un numero molto elevato di geni R, legati alla resistenza ai patogeni e ai fattori associati

Il genoma dei vegetali condivide con quello degli animali molte famiglie geniche:

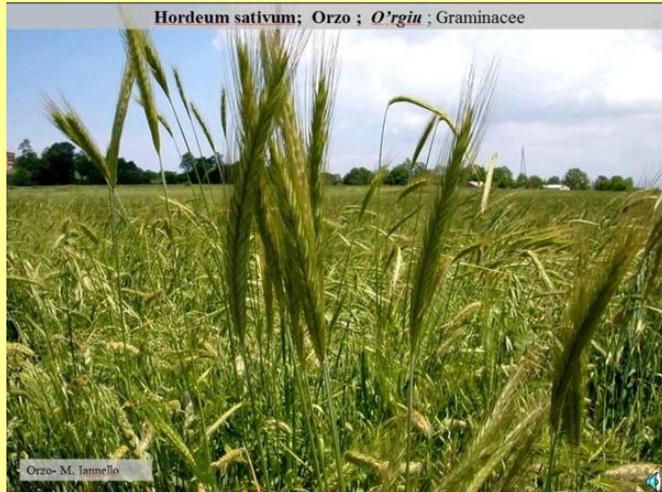
-  coinvolte nella comunicazione intracellulare
-  nella regolazione trascrizionale
-  nella trasduzione dei segnali durante lo sviluppo

Altre famiglie di fattori di trascrizione sono tipiche delle piante.

composti a funzione ormonale esclusivi delle piante

PROGETTO GENOMA

Graminacee e Leguminose



Sono in corso progetti riguardanti il genoma di più di 60 specie di piante

Dal punto di vista economico, i più importanti di questi progetti riguardano le principali piante alimentari

Orzo

Mais

Miglio

Riso

Frumento

Lolium

Vite

Alfalfa

Soia

Fagiolo

Pomodoro

Patata

Melo

Cotone

Alcuni di questi genomi sono molto grandi (poliploidia e DNA ripetitivo)

ed il sequenziamento dell'intero genoma non è al momento realizzabile

Le ricerche si concentrano quindi sulla
genomica comparativa

Entrambi i genomi

Riso e mais hanno genomi relativamente piccoli e sono così importanti per le economie agricole dei Paesi Sviluppati che è stata data priorità al sequenziamento completo