

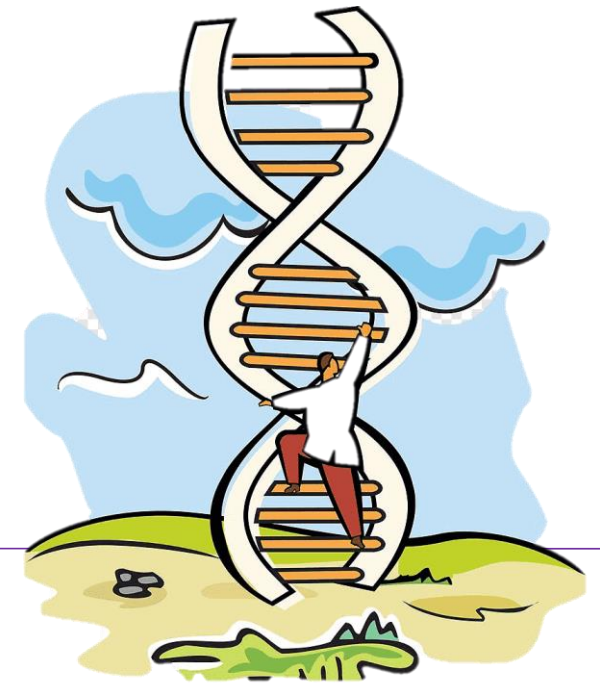
# Analisi molecolari per il fingerprinting delle varietà vegetali

## STATO DELL'ARTE



**Alberto Acquadro**

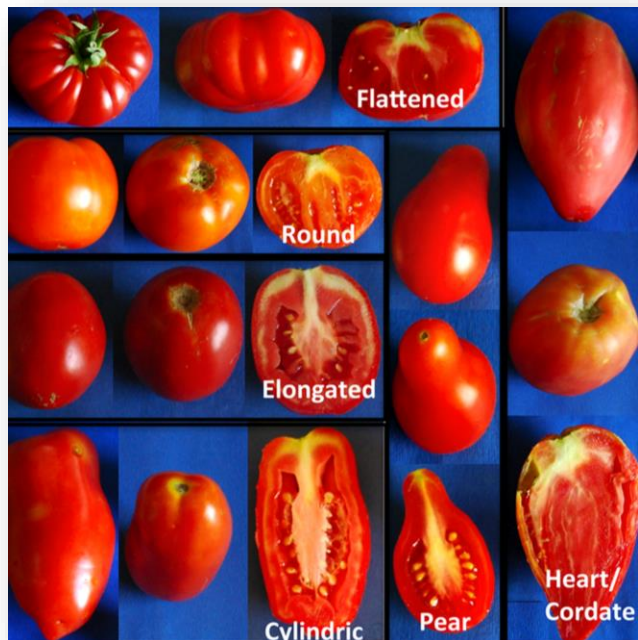
DISAFA – Genetica Vegetale  
Università degli studi di Torino



# MARCATORE GENETICO

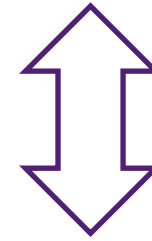
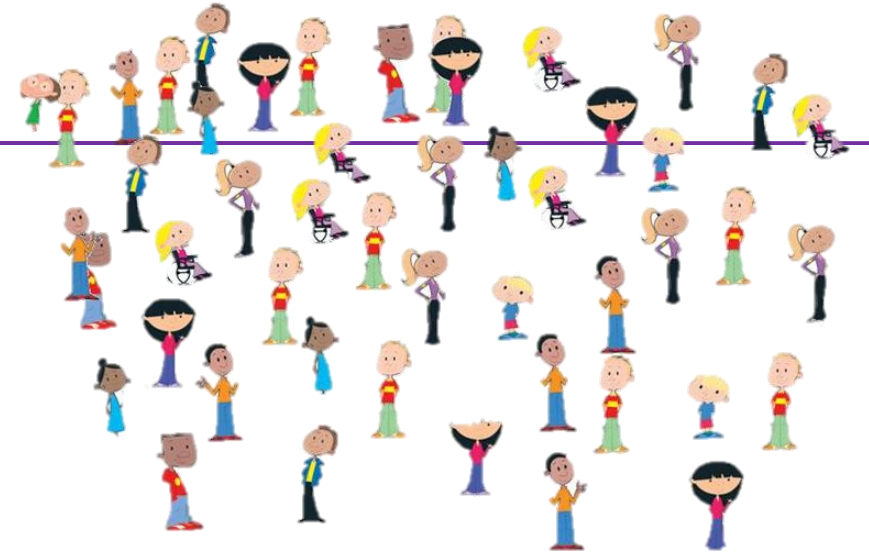
## MARCATORE GENETICO

carattere mendeliano che può essere utilizzato per seguire la segregazione di una particolare regione cromosomica



# MARCATORI GENETICI: **IMPIEGHI**

- Analisi di paternità
- Diagnosi di anomalie genetiche
- **Analisi forensi**
- Identificazione di QTL e geni utili
- Miglioramento delle specie vegetali ed animali (MAS)
- Studio della struttura, evoluzione e biodiversità delle popolazioni
- Studi di epidemiologia
- **Tracciabilità dei prodotti animali e/o dei GMO**
- Conservazione della natura
- Analisi di campioni da erbari e collezioni
- Studi di filogenesi ed evoluzione



# MARCATORI GENETICI: TIPOLOGIE

## Marcatori morfologici - caratteri mendeliani

- colore della cariosside
- forma della foglia
- presenza assenza di corna

## Marcatori biochimici

- gruppi sanguigni
- isoenzimi
- proteine (dell'endosperma, del plasma, del latte ecc.)

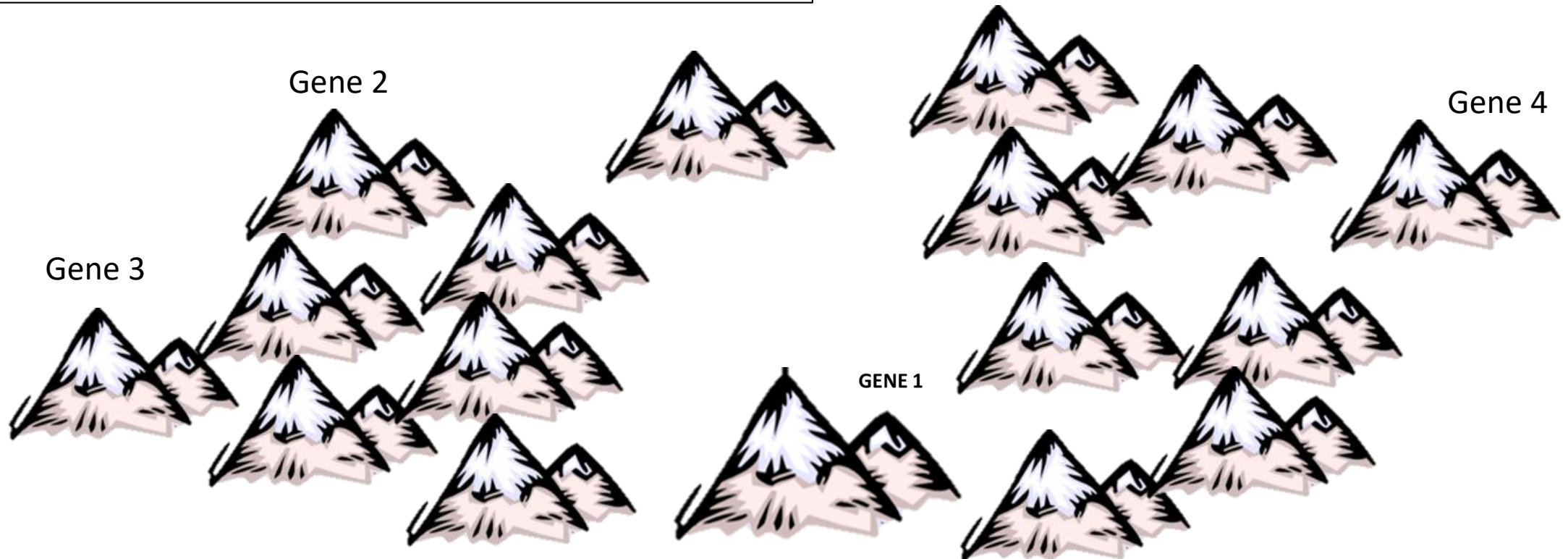
## Marcatori molecolari

una sequenza DNA o proteica facilmente identificabile e la cui ereditabilità sia monitorabile



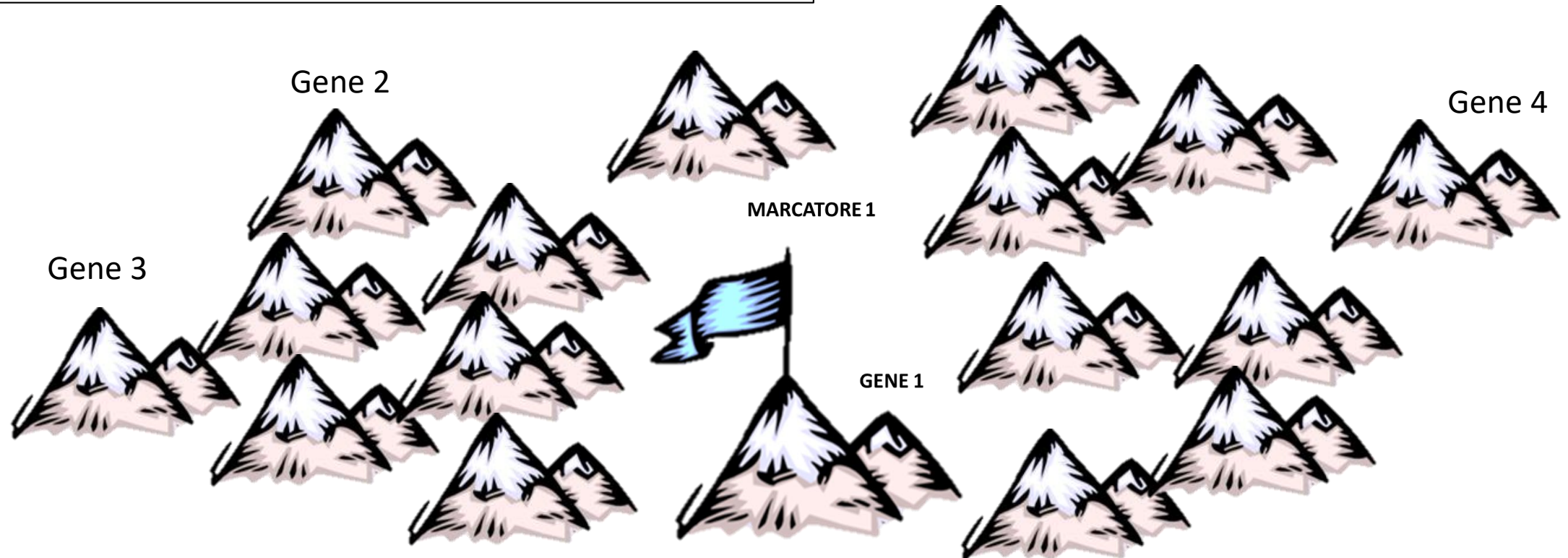
# MARCATORI MOLECOLARI: CHE COSA SONO

Un marcatore è un locus che definisce in modo caratteristico e inequivocabile un tratto genomico



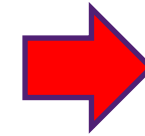
# MARCATORI MOLECOLARI: CHE COSA SONO

Un marcatore è un locus che definisce in modo caratteristico e inequivocabile un tratto genomico

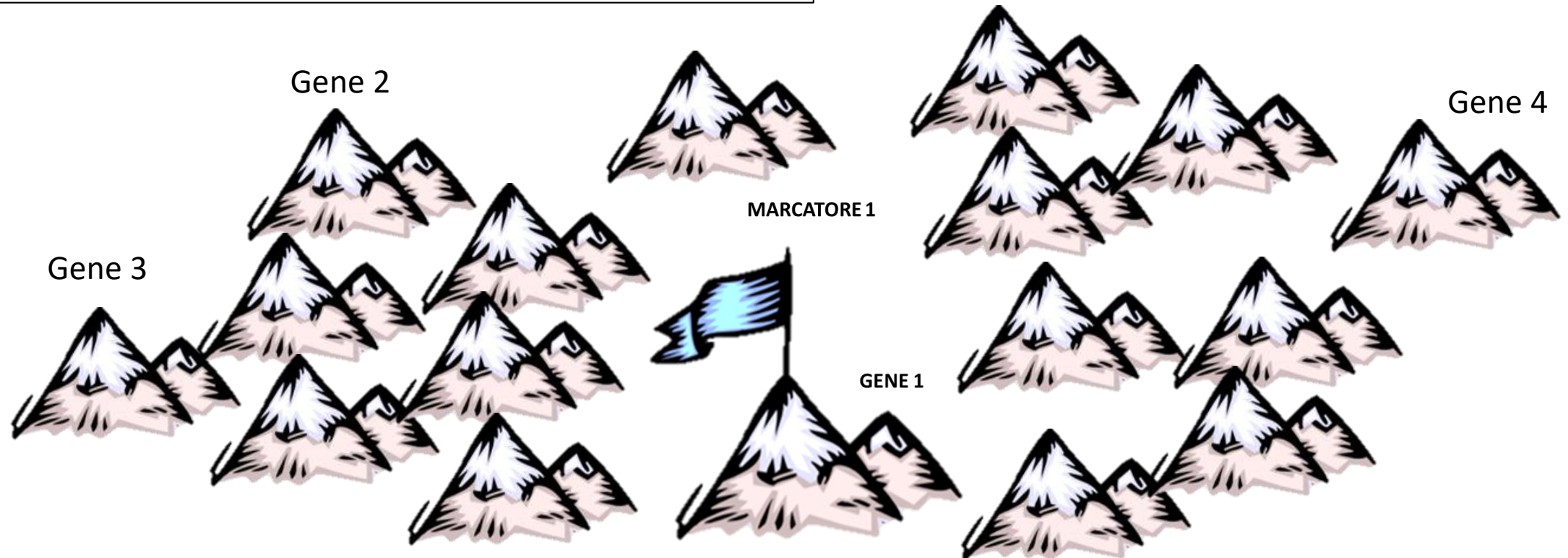


# MARCATORI MOLECOLARI: CHE COSA SONO

Un marcatore è un locus che definisce in modo caratteristico e inequivocabile un tratto genomico



Rilevabile con metodologie molecolari



# IL MARCATORE IDEALE

- Comportamento mendeliano
- Non influenzato dall'ambiente
- Stabile
- Facile da monitorare
- Numeroso
- Codominante
- Polimorfico (incluso anche sequenze non codificanti)
- Presente in qualsiasi tessuto
- Indipendente da sesso ed età
- Analisi automatizzabile
- Analisi veloce ed economica





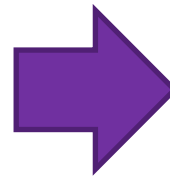
# MARCATORI MOLECOLARI: VANTAGGI

Le caratteristiche morfologiche e biochimiche possono variare in relazione all'ambiente di coltivazione



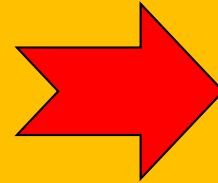
**INFLUENZATI  
DALL'AMBIENTE**

**TECNICHE DI ANALISI DEL DNA**  
IDENTIFICAZIONE DI UNA VARIETÀ  
IN QUALSIASI STADIO DI SVILUPPO

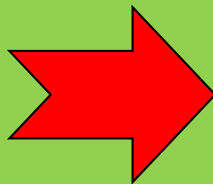


# MARCATORI MOLECOLARI: VANTAGGI

**DNA**



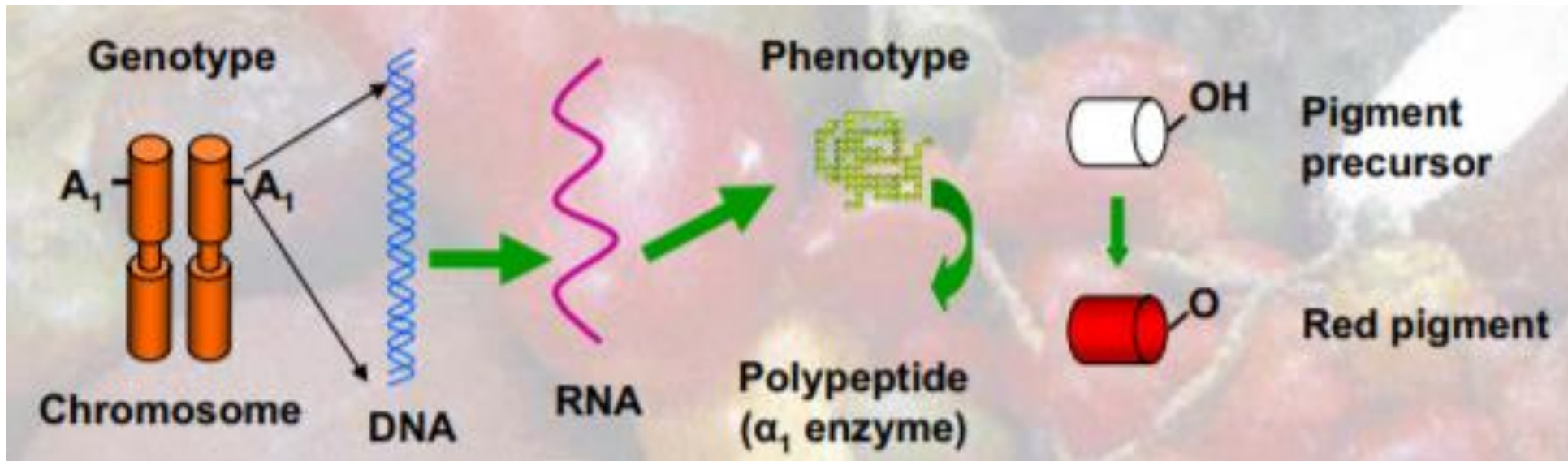
- Marcatori molecolari**
- Potenzialmente infiniti
  - Non influenzati dall'ambiente



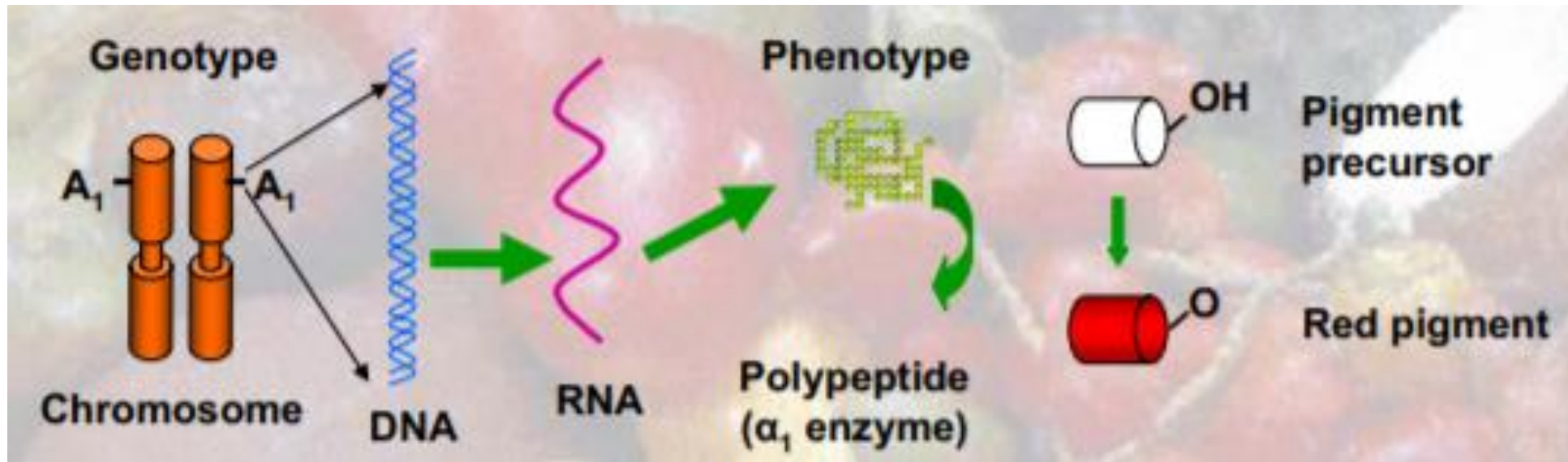
- Caratteri morfologici e fisiologici**
- Pochi caratteri utilizzabili
  - influenzati dall'ambiente

Aumenta il livello di precisione

# IL FENOTIPO (CARATTERE MORFOLOGICO) HA SEMPRE UNA BASE NEL DNA

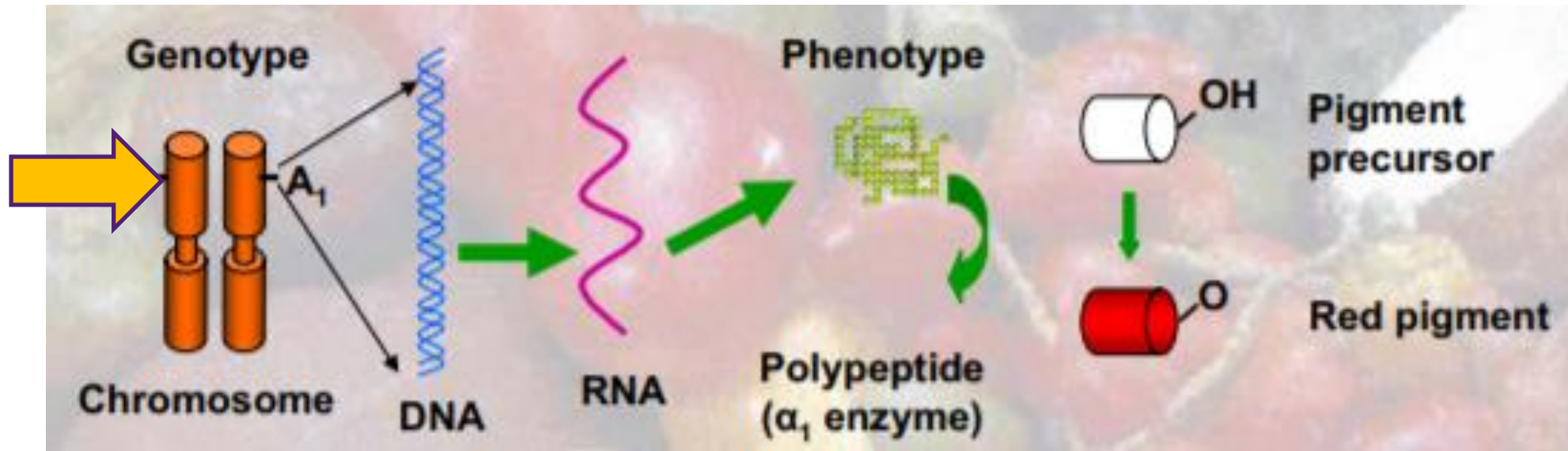


# IL FENOTIPO (CARATTERE MORFOLOGICO) HA SEMPRE UNA BASE NEL DNA



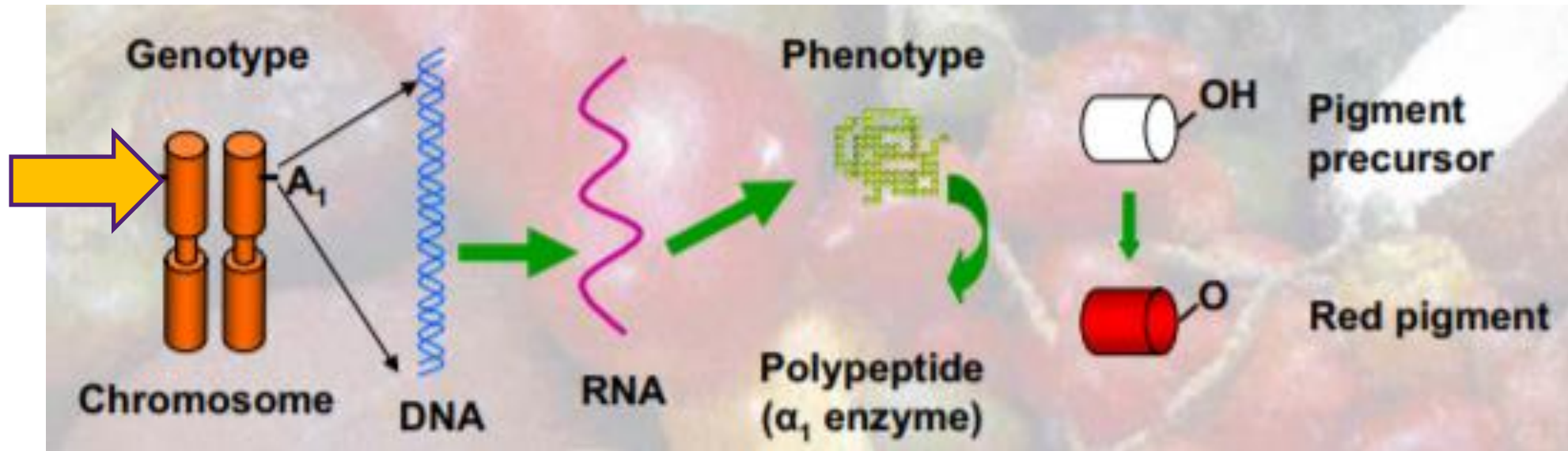
# IL FENOTIPO (CARATTERE MORFOLOGICO) HA SEMPRE UNA BASE NEL DNA

MUTAZIONE in gene chiave



# IL FENOTIPO (CARATTERE MORFOLOGICO) HA SEMPRE UNA BASE NEL DNA

MUTAZIONE in gene chiave



# 1 sola mutazione nel DNA può portare a differenze macroscopiche!

---



**Nel corso della domesticazione  
(circa 10mila anni di storia dell'agricoltura)**

**Singole mutazioni chiave hanno modificato  
uno o pochi geni determinando cambiamenti  
Fenotipici incredibili**

# 1 sola mutazione nel DNA può portare a differenze macroscopiche!

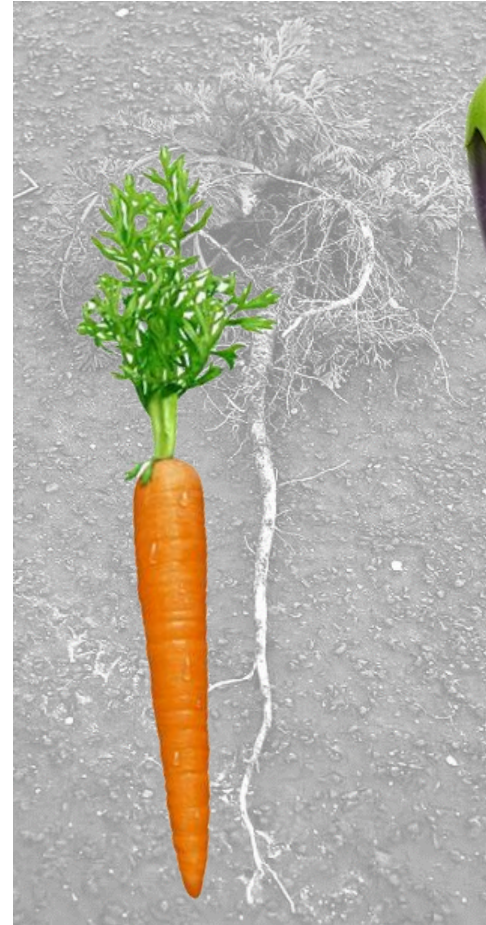
Guardate questi 5 esempi di piante selvatiche



**È molto difficile collegarle con le «crops» di oggi**



# 1 sola mutazione nel DNA può portare a differenze macroscopiche!



# POLIMORFISMO

Il termine **polimorfismo** (dal greco, letteralmente, «avere molte forme») assume significati specifici in diversi contesti

In biologia/genetica si parla di polimorfismo genetico rispetto ad una variazione genetica e in particolare quando questa abbia una frequenza maggiore dell'1% nella popolazione.

## Polimorfismo



Marcatori morfologico: colore

## Monomorfismo



Marcatori morfologico: colore

# POLIMORFISMO (A LIVELLO MOLECOLARE)

**I polimorfismi a livello molecolare si possono classificare in:**

- **polimorfismi di sequenza**  
(conseguenza di differenze nelle sequenze di nucleotidi)
- **polimorfismi di inserzione e delezione**  
(causati da mutazioni puntiformi)
- **polimorfismi nel numero di ripetizioni**

# POLIMORFISMO DI SEQUENZA: **SNP**

**SNP = Single Nucleotide Polymorphism**

seq_1 (A)	ATGCGGC <b>G</b> AT
seq_2 (A)	ATGCGGC <b>G</b> AT
seq_3 (A)	ATGCGGC <b>G</b> AT
seq_1 (B)	ATGCGGC <b>A</b> AT
seq_2 (B)	ATGCGGC <b>A</b> AT
seq_3 (B)	ATGCGGC <b>A</b> AT



## **Definizione**

Un **polimorfismo a singolo nucleotide** è una variazione, del materiale genico a carico di un unico nucleotide tale per cui l'allele polimorfico risulta presente nella popolazione in una proporzione superiore all'1%. Al di sotto di tale soglia si è soliti parlare di variante rara (SNV = single nucleotide variant). Vengono rilevati mediante sequenziamento.

# POLIMORFISMO DI INSERZIONE/DELEZIONE: **INDEL**

**indel = insertion/deletion**

<b>Indiv 1</b>	ATGGATATAGGCTGA-TAGGCTA
<b>Indiv 2</b>	ATGGATATAGGCTGA-TAGGCTA
<b>Indiv 3</b>	ATGGATATAGGCTGAC <b>C</b> TAGGCTA
<b>Indiv 4</b>	ATGGATATAGGCTGAC <b>C</b> TAGGCTA

## **Definizione**

il termine **Indel** si usa come abbreviazione di un evento di mutazione/ricombinazione che può far parte di due classi: una inserzione (insertion) o una cancellazione (deletion). Vengono rilevate mediante sequenziamento (in genere insieme agli SNP)

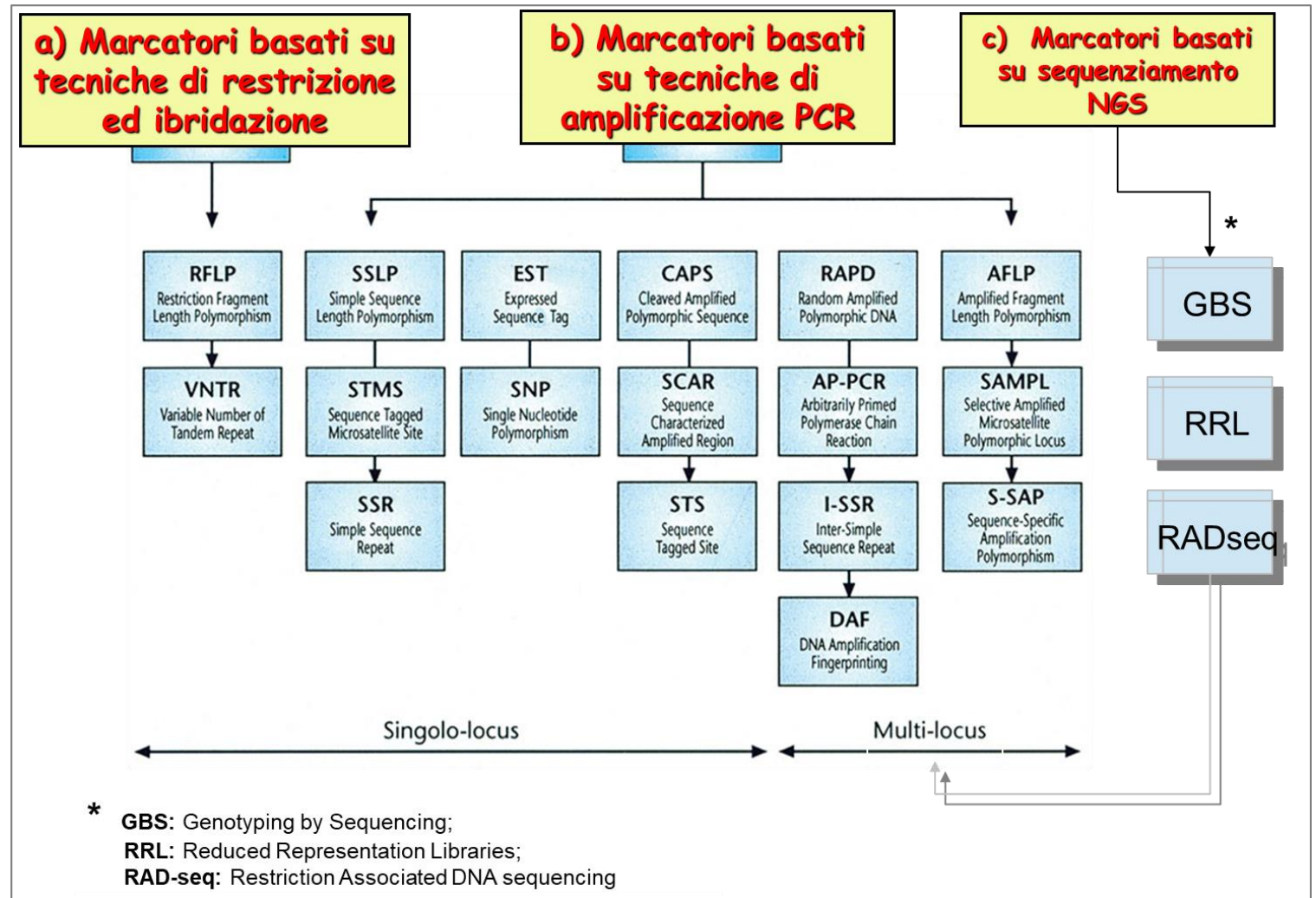
# POLIMORFISMO NEL N° RIPETIZIONI: MICRO- MINI- MIDI-SATELLITI

**Indiv 1** ATGGA**TATATATATA**GGCTGATAGGCTA  
**Indiv 2** ATGGA**TATA**GGCTGATAGGCTA  
**Indiv 3** ATGGA**TATA**GGCTGACTAGGCTA  
**Indiv 4** ATGGA**TATA**GGCTGACTAGGCTA

## Definizione

Il polimorfismi legati al numero di ripetizioni ricadono, ad esempio, nei marcatori SSR (Simple Sequence Repeats) e sono formati da sequenze di DNA che si ripete in tandem. Alla base del loro grande successo c'è l'elevata variabilità intra-locus tra individui della stessa specie.

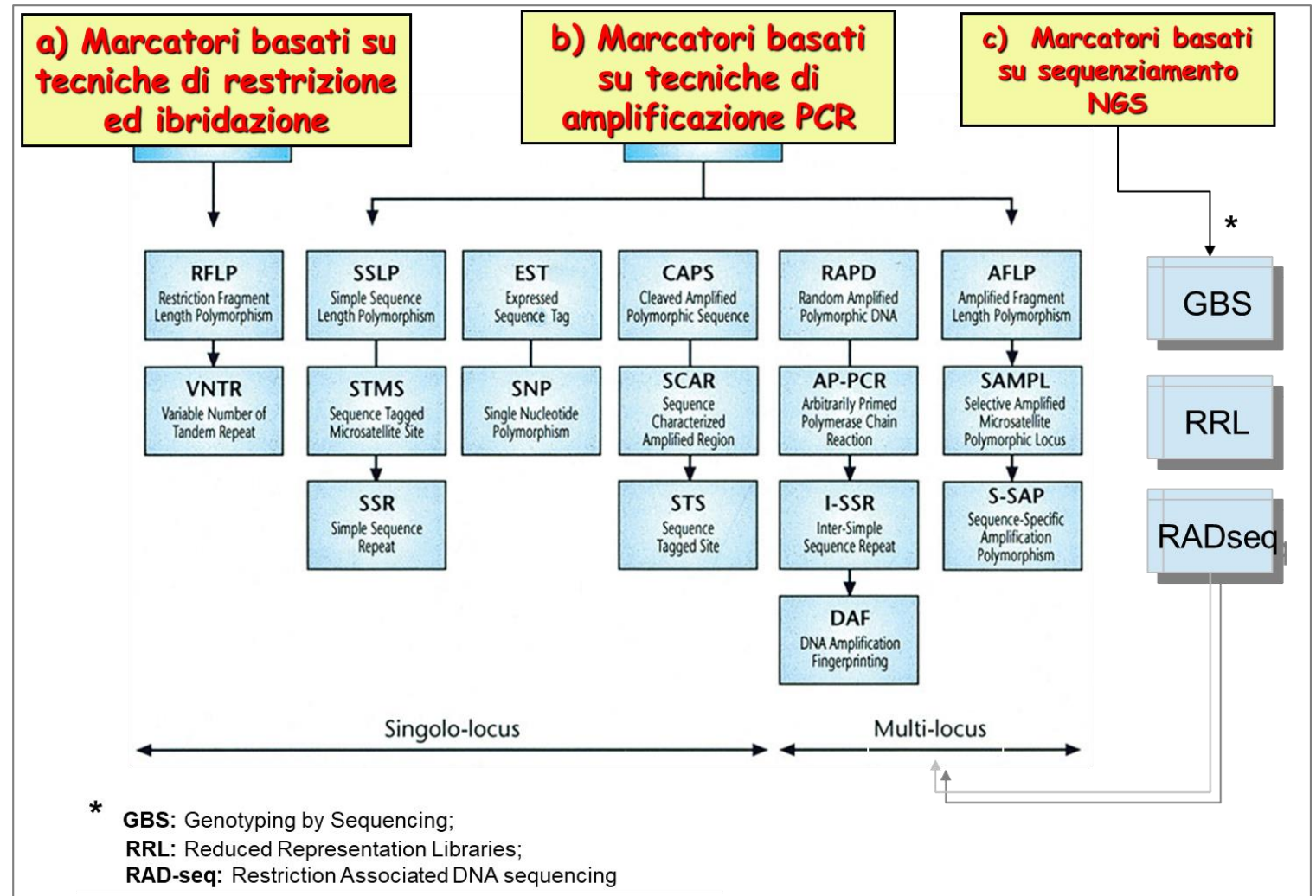
# MARCATORI MOLECOLARI: METODOLOGIE



# MARCATORI MOLECOLARI: METODOLOGIE

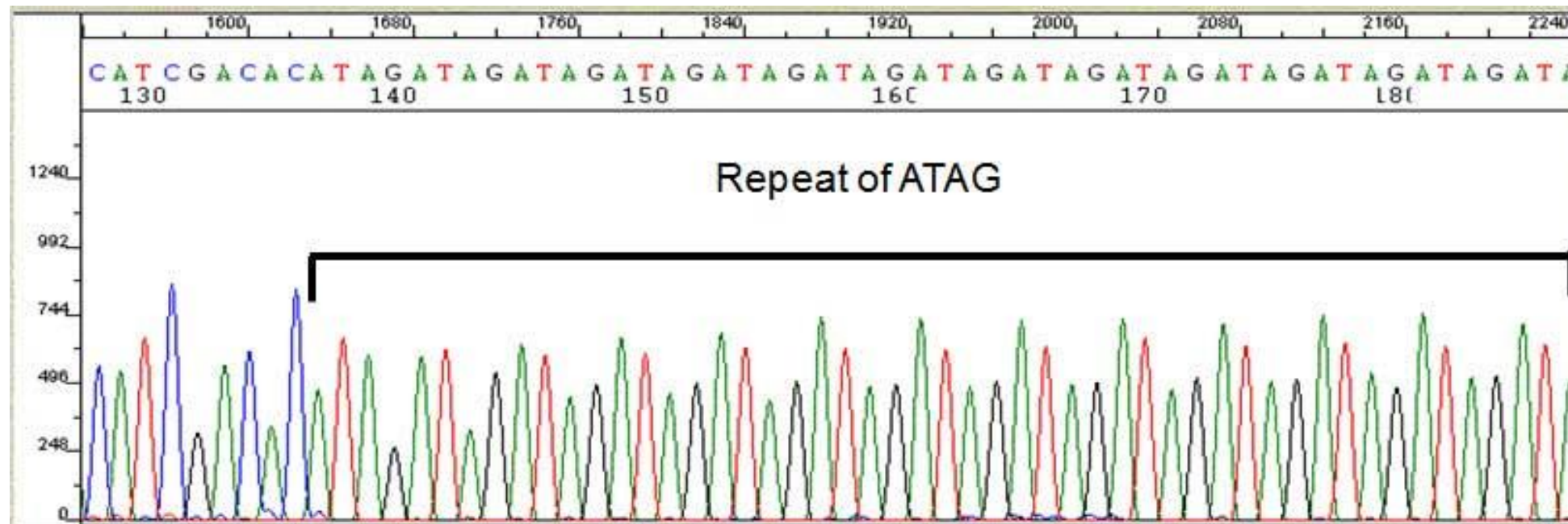


**Molte tecniche**



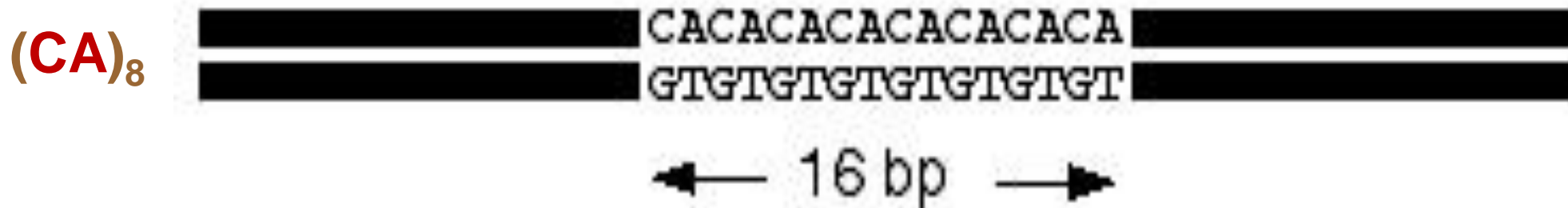


# MARCATORI MICROSATELLITI



# MICROSATELLITI: CHE COSA SONO

- DNA ripetuto in tandem
- Anche chiamati SSR (Simple Sequence Repeats)



# MICROSATELLITI: CHE COSA SONO

## Tipologie di SSR

AAAAAAAAAAAAA = (A)<sub>13</sub>

**MONO-nucleotidici SSR**

ATATATATATAT = (AT)<sub>6</sub>

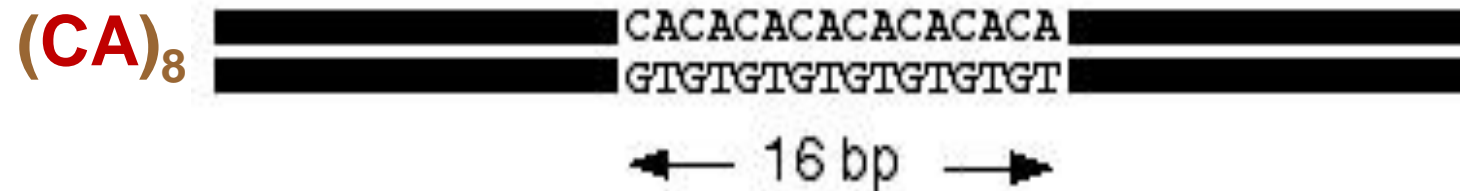
**DI-nucleotidici SSR**

**CGACGACGACGA** = (CGA)<sub>4</sub>

**TRI-nucleotidici SSR**

**ATCGATCGATCG** = (ATCG)<sub>3</sub>

**TETRA-nucleotidici SSR**



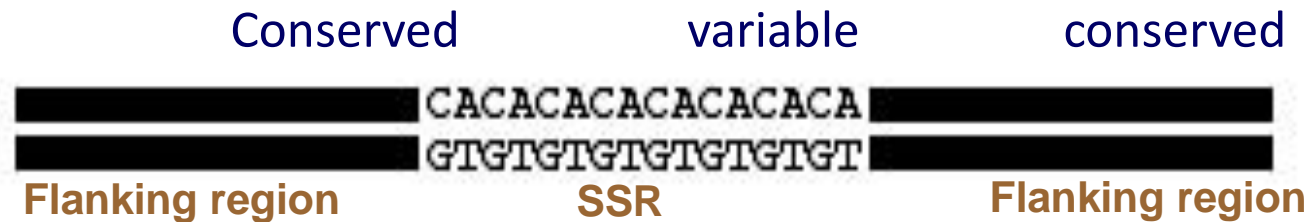
# MICROSATELLITI: CHE COSA SONO

- Ripetizioni ubiquitarie in natura
- Presenti in tutti gli eucarioti:
  - Primati
  - Roditori
  - Altri mammiferi
  - non mammiferi (vertebrati)
  - insetti
  - Piante
- Meno riconosciuti nei procarioti

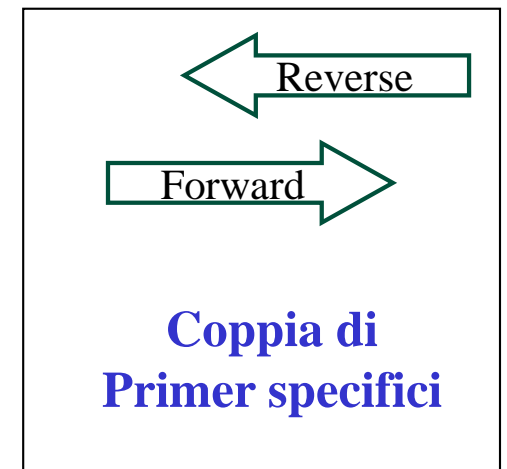
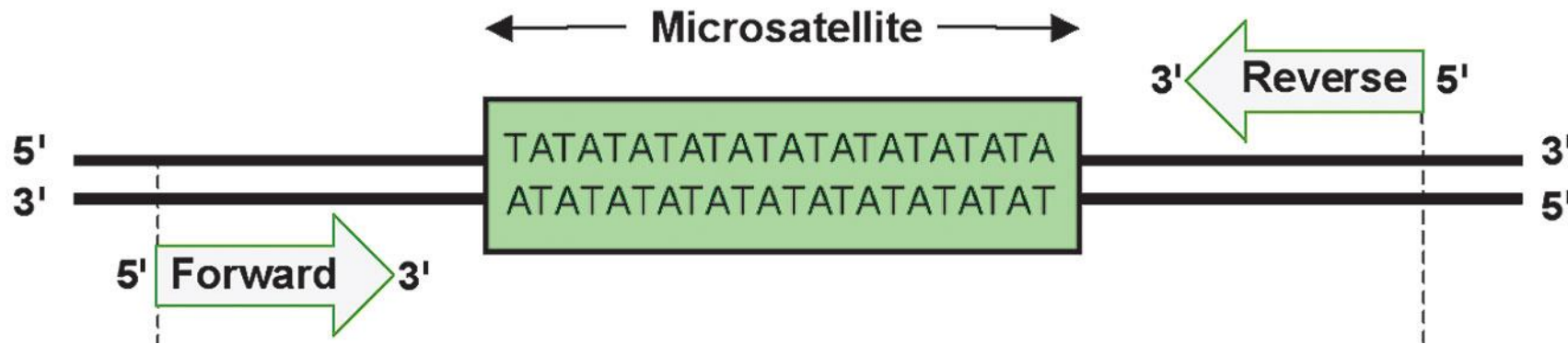
# MICROSATELLITI: CARATTERISTICHE PRINCIPALI

**Altamente variabili** (elevato tasso mutazionale) a livello di:

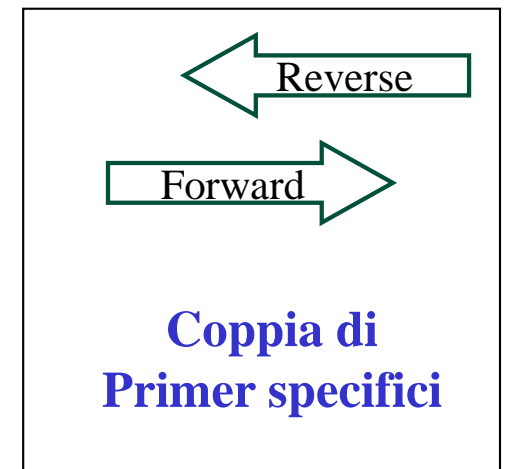
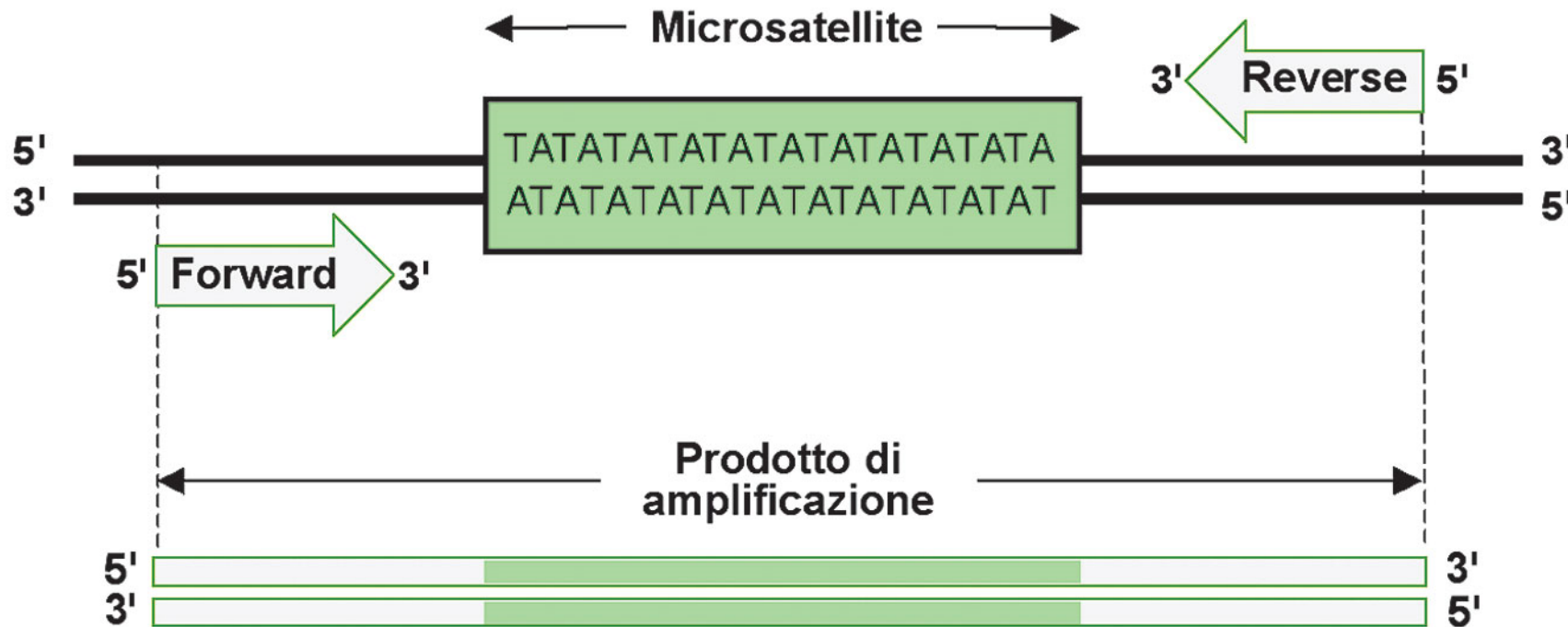
- **Ogni locus >>> molti alleli** (2-100 alleli).
- **Le regioni fiancheggianti sono conservate** entro specie
- **Distribuiti casualmente** nel genoma



# CARATTERISTICHE DEI MICROSATELLITI: RILEVABILI CON PCR



# CARATTERISTICHE DEI MICROSATELLITI: RILEVABILI CON PCR

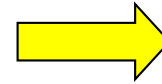


# APPLICAZIONE

**Multiplex:** 14 loci in contemporanea

**Avere la sequenza di DNA**

```
ATGAGCCTAGCACAGCTGATCAATTGCCTGTGTA
GTGTGTGTGTGTGTGTGTGGTGTGCAAGGGTG
GGTGGTGGGATAGAAAAGAGATC
```



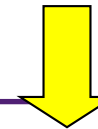
*Forward primer*



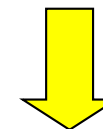
*Reverse primer*



**DISEGNARE 2 primer  
FIANCHEGGIANTI il motivo SSR**



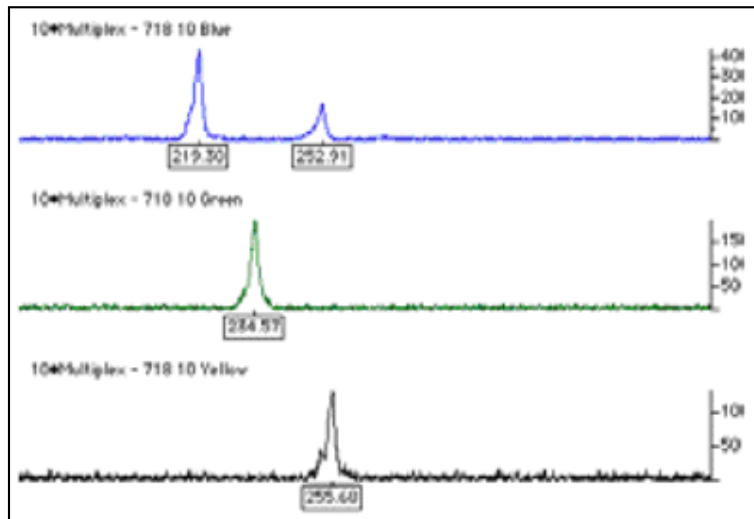
**AMPLIFICAZIONE  
PCR**



**VISUALIZZAZIONE**



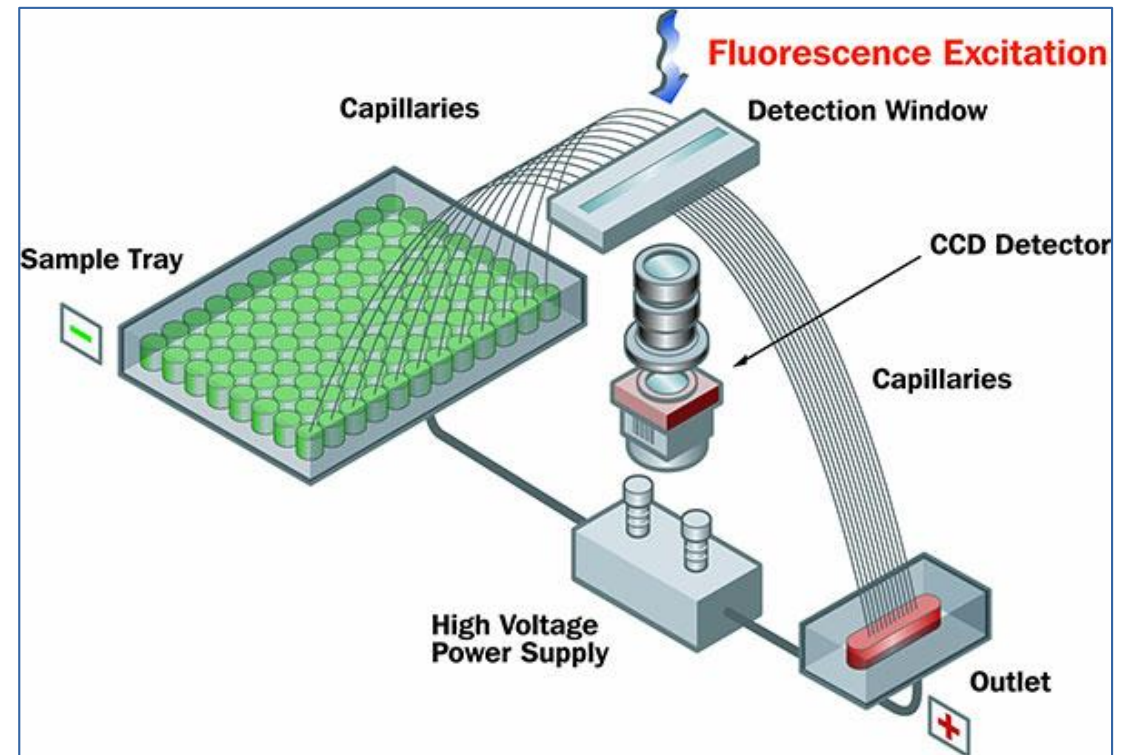
**Elettroforesi capillare**



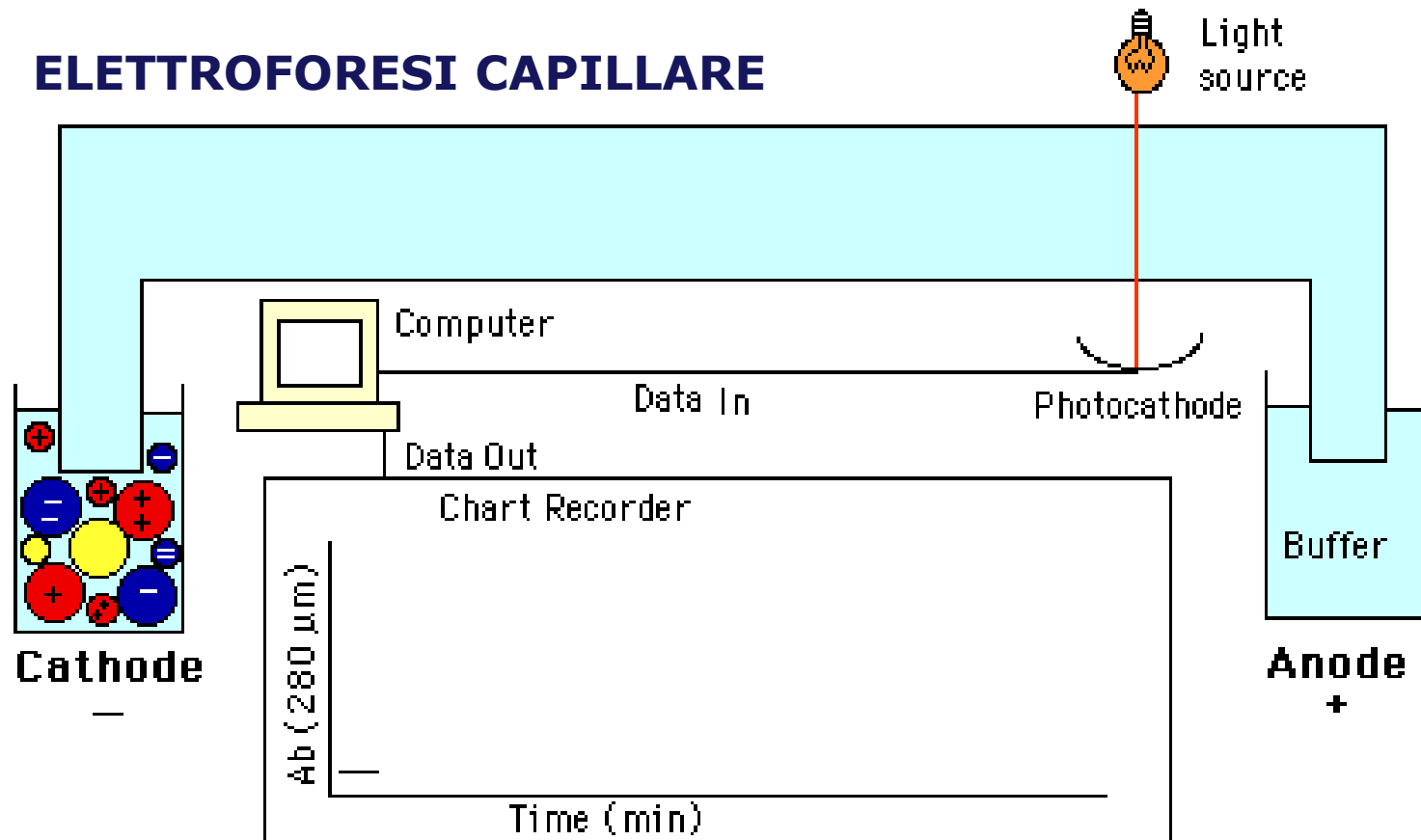


# REQUISITI: **VISUALIZZAZIONE**

## ELETTROFORESI CAPILLARE

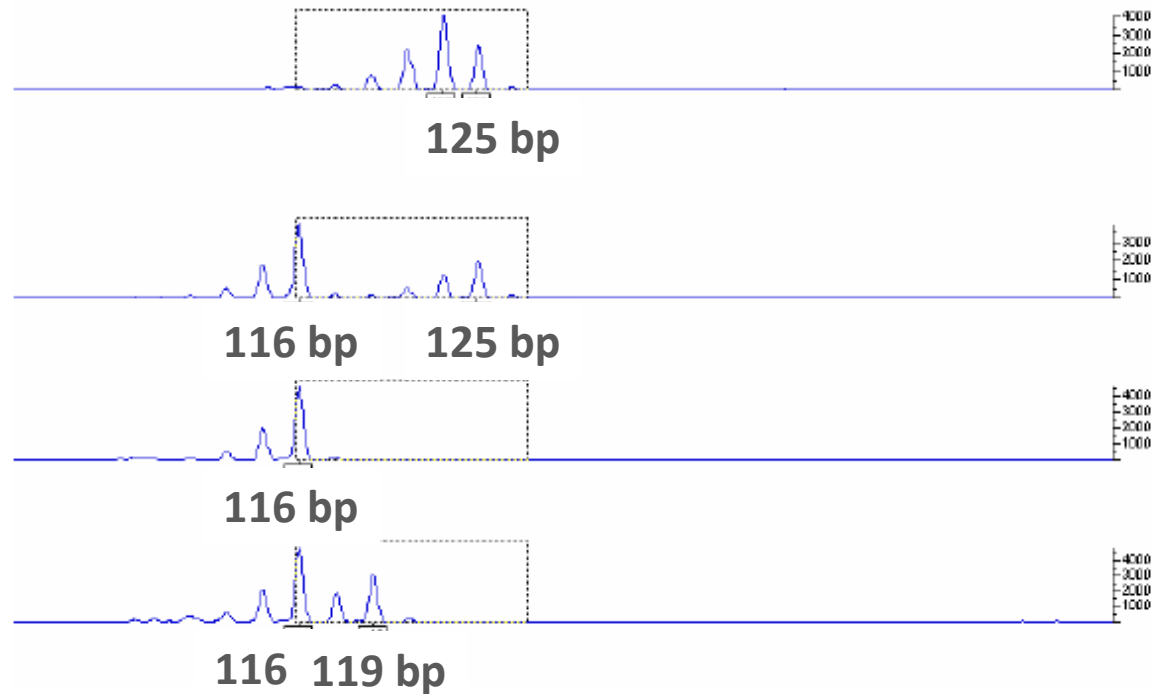


# REQUISITI: CONOSCENZA SPECIE, DNA, PCR



# REQUISITI: CONOSCENZA SPECIE, DNA, PCR

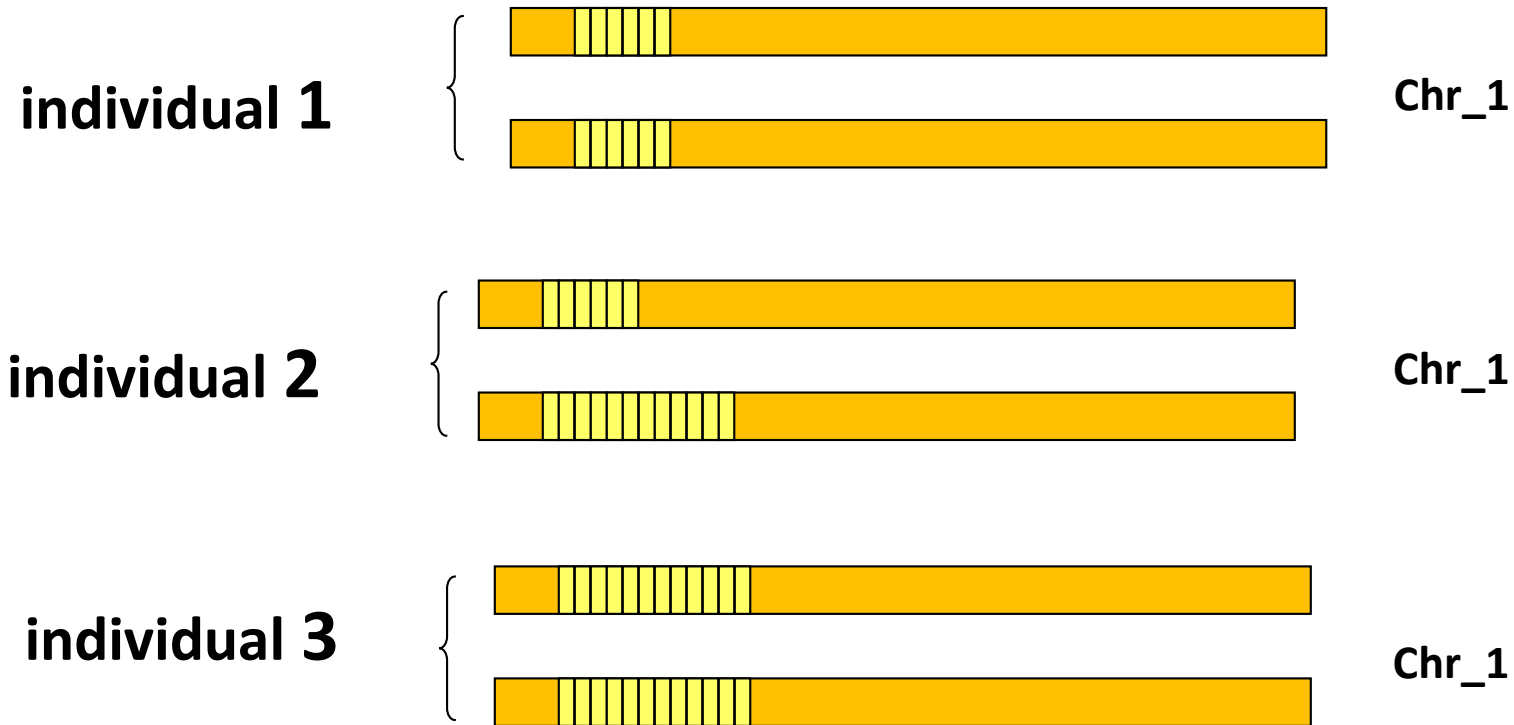
## Elettroferogramma



L'intensità del segnale dipende dalla quantità di frammento e determina l'area e l'altezza dei picchi

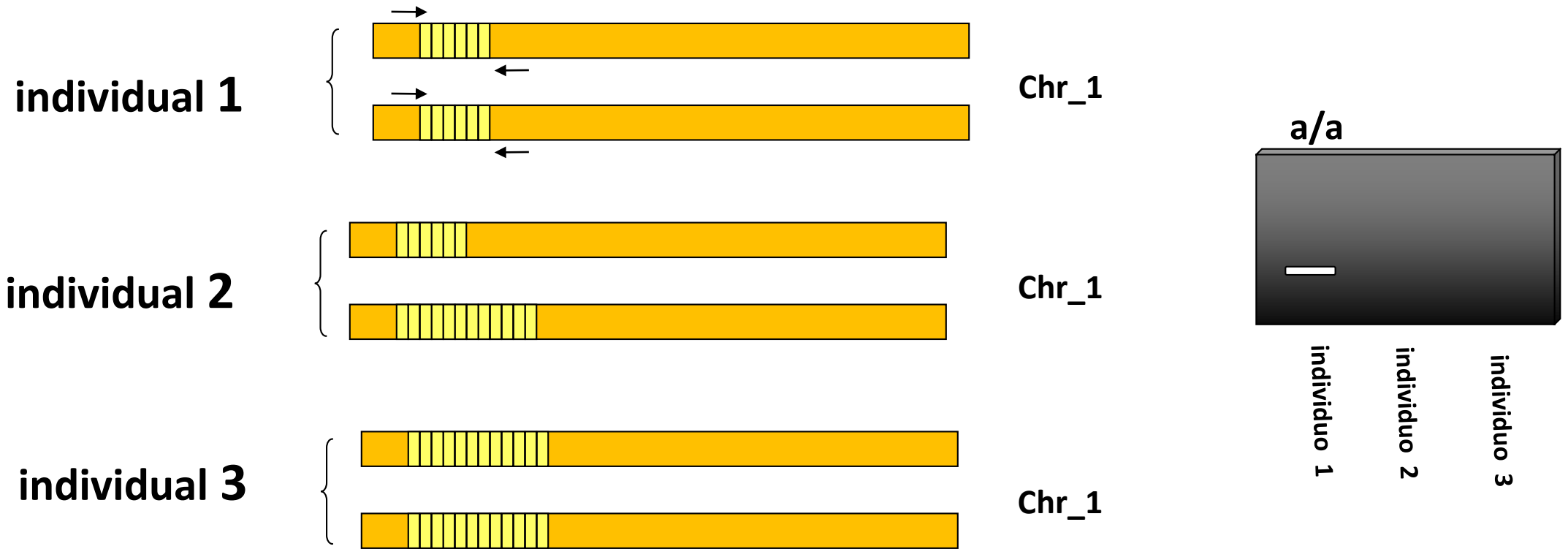
# CARATTERISTICHE DEI MICROSATELLITI: CO-DOMINANTI

**È POSSIBILE DISTINGUERE L'OMOZIGOTE DALL'ETEROZIGOTE**



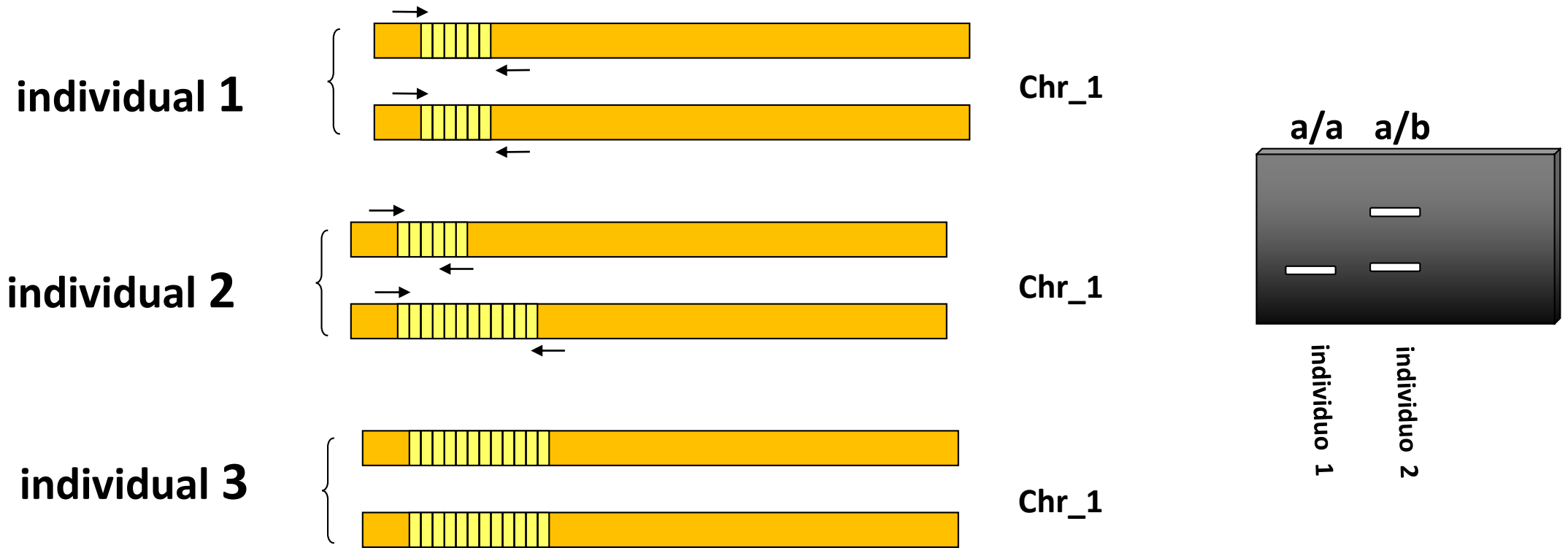
# CARATTERISTICHE DEI MICROSATELLITI: CO-DOMINANTI

**È POSSIBILE DISTINGUERE L'OMOZIGOTE DALL'ETEROZIGOTE**



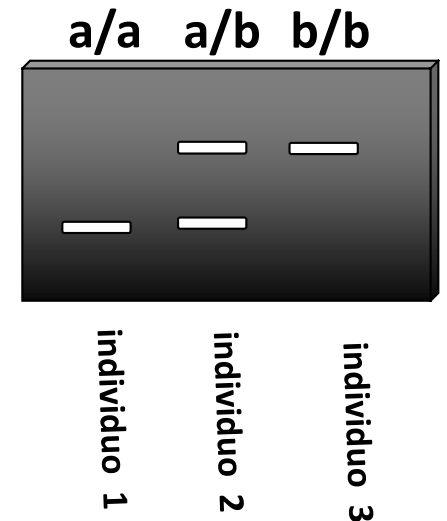
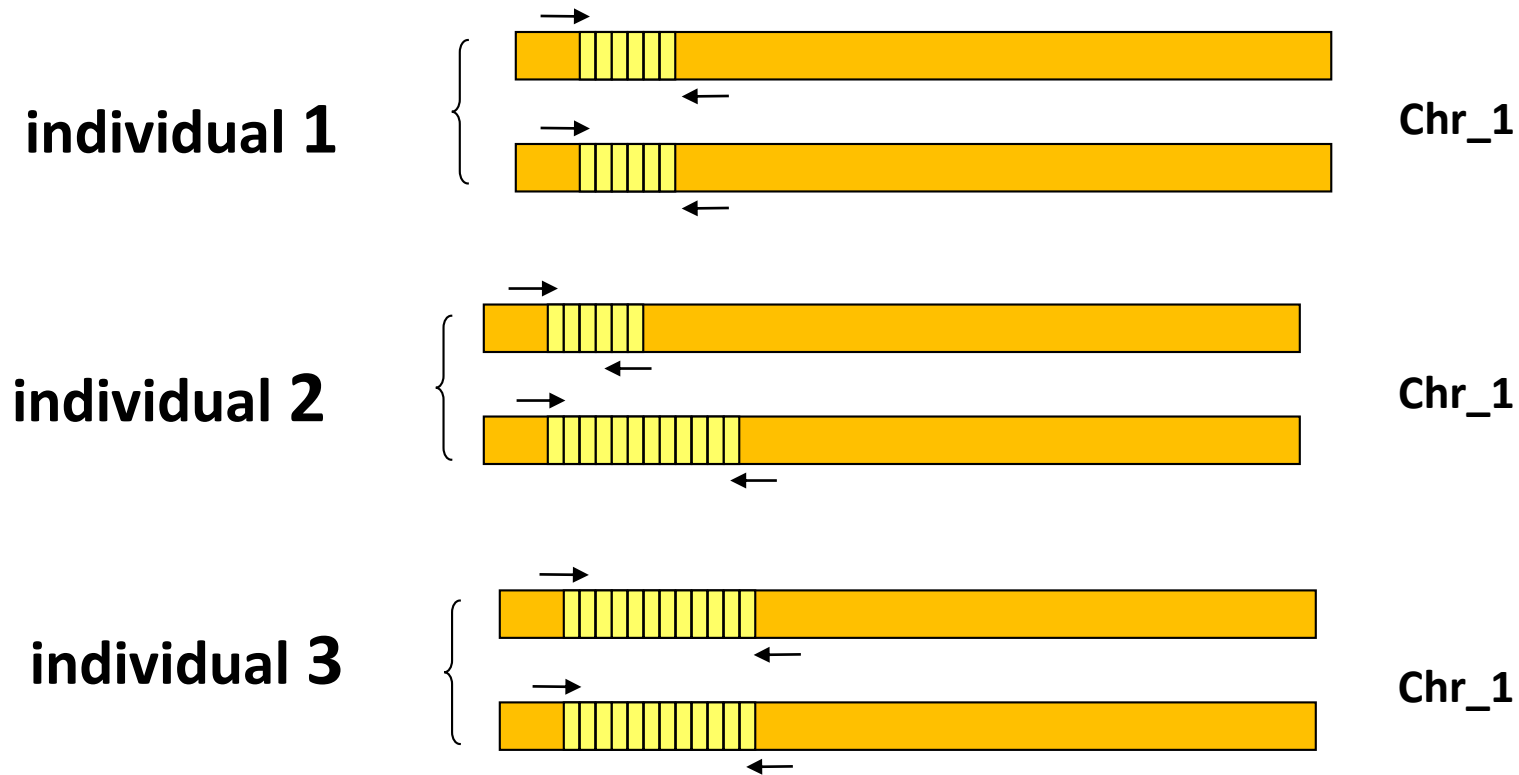
# CARATTERISTICHE DEI MICROSATELLITI: CO-DOMINANTI

**È POSSIBILE DISTINGUERE L'OMOZIGOTE DALL'ETEROZIGOTE**



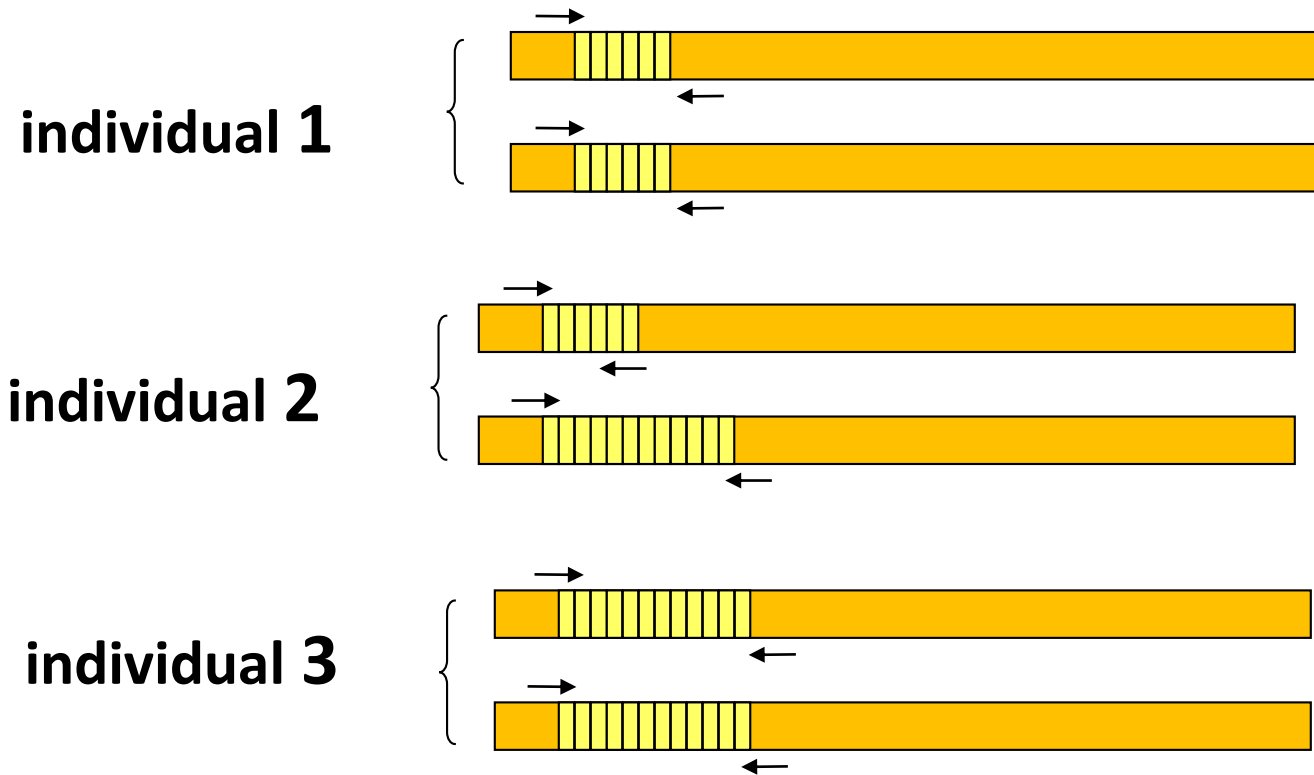
# CARATTERISTICHE DEI MICROSATELLITI: CO-DOMINANTI

**È POSSIBILE DISTINGUERE L'OMOZIGOTE DALL'ETEROZIGOTE**



# CARATTERISTICHE DEI MICROSATELLITI: CO-DOMINANTI

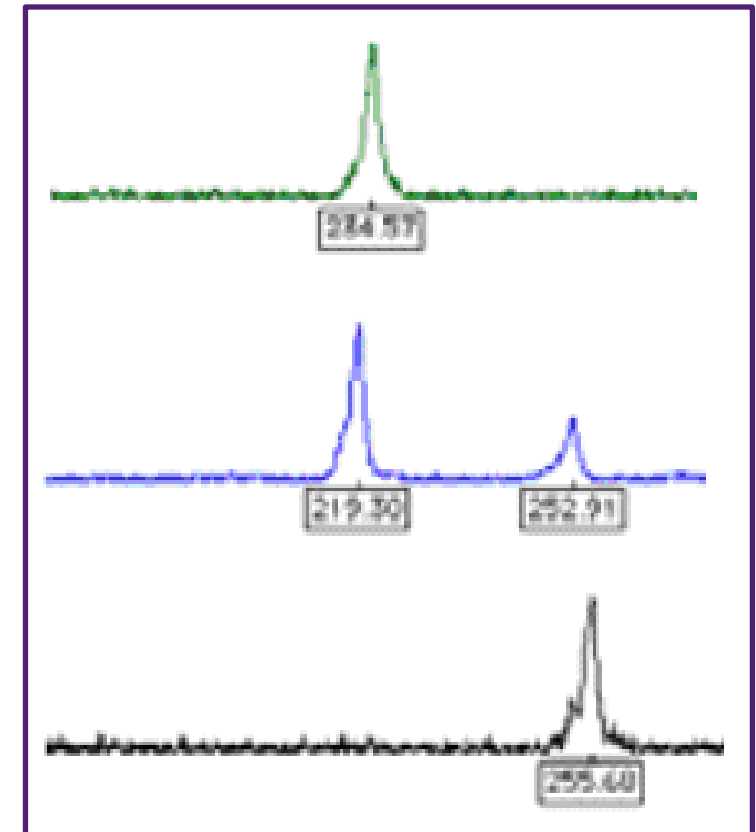
**È POSSIBILE DISTINGUERE L'OMOZIGOTE DALL'ETEROZIGOTE**



Chr\_1

Chr\_1

Chr\_1

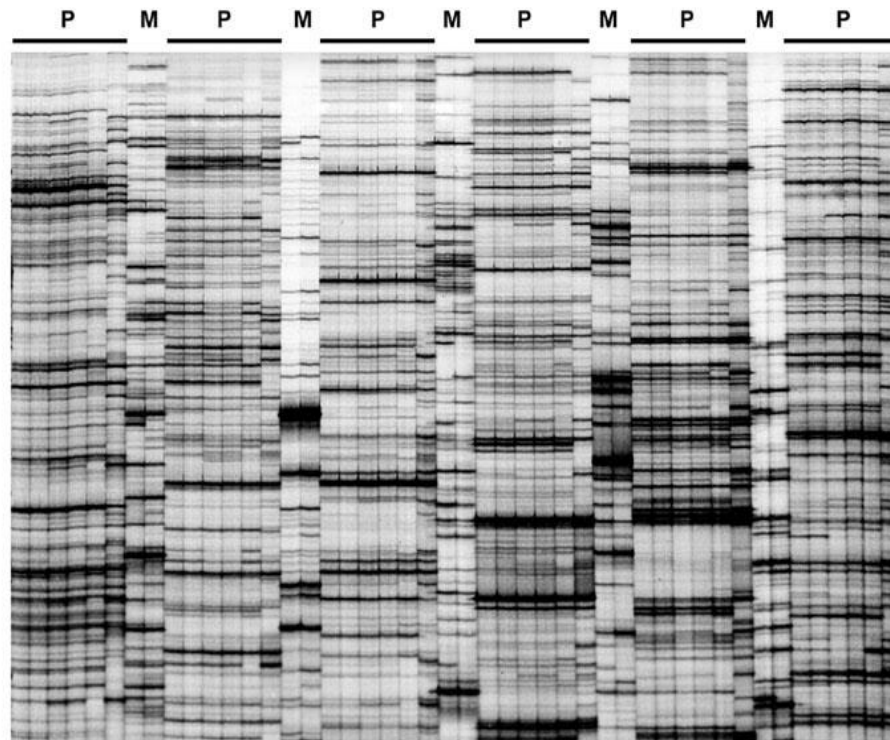




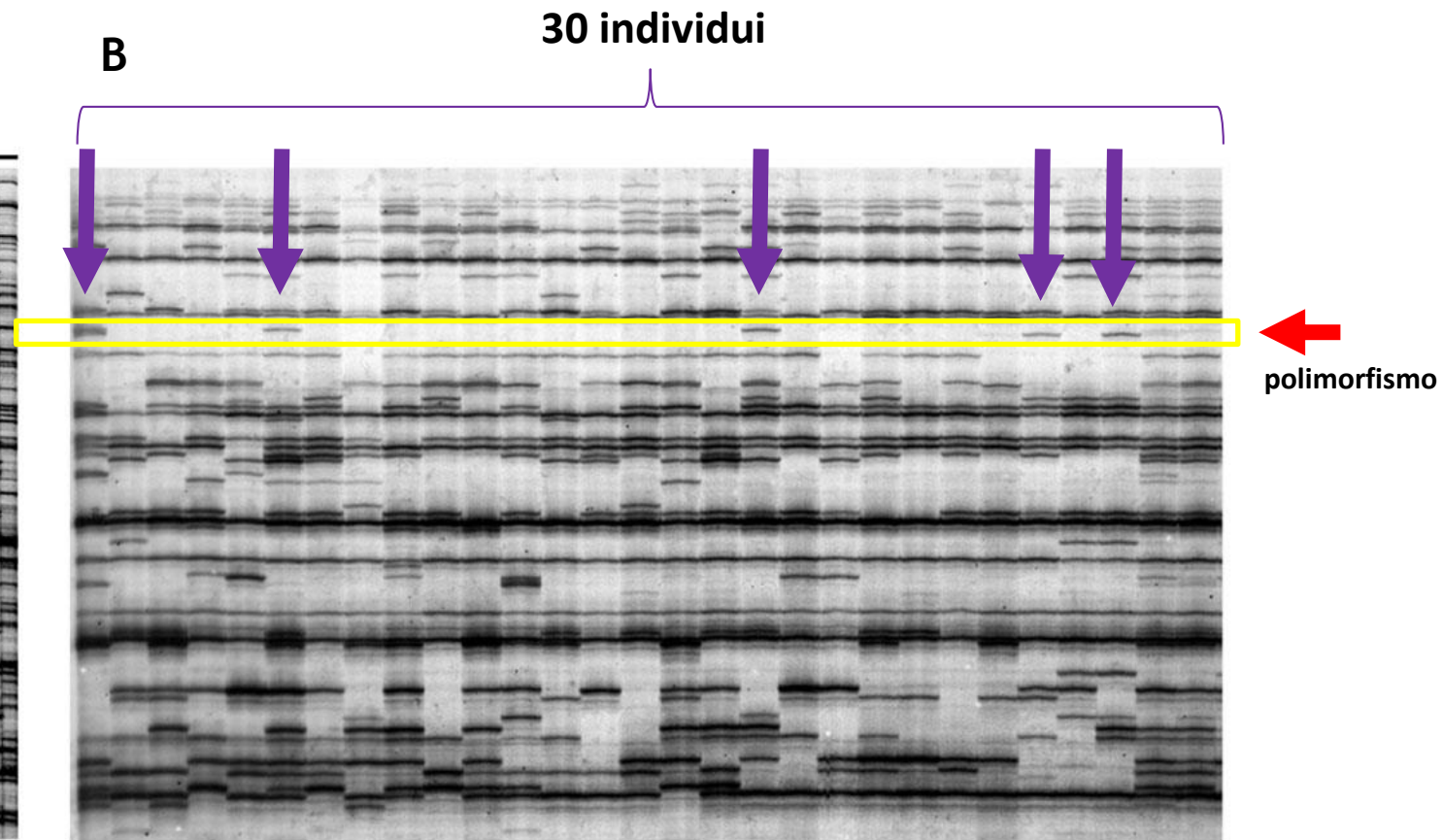
# MARCATORI AFLP

(AMPLIFIED FRAGMENT LENGTH POLYMORPHISM)

A



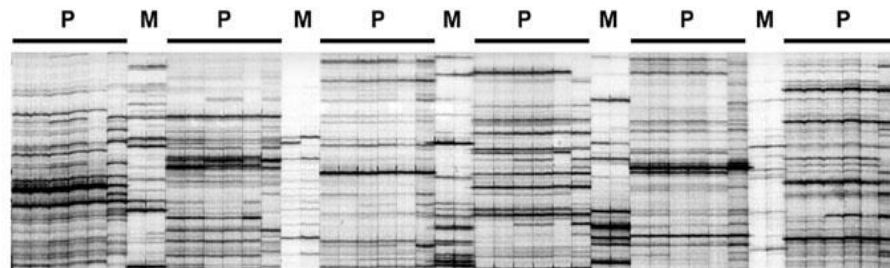
B



# MARCATORI AFLP

(AMPLIFIED FRAGMENT LENGTH POLYMORPHISM)

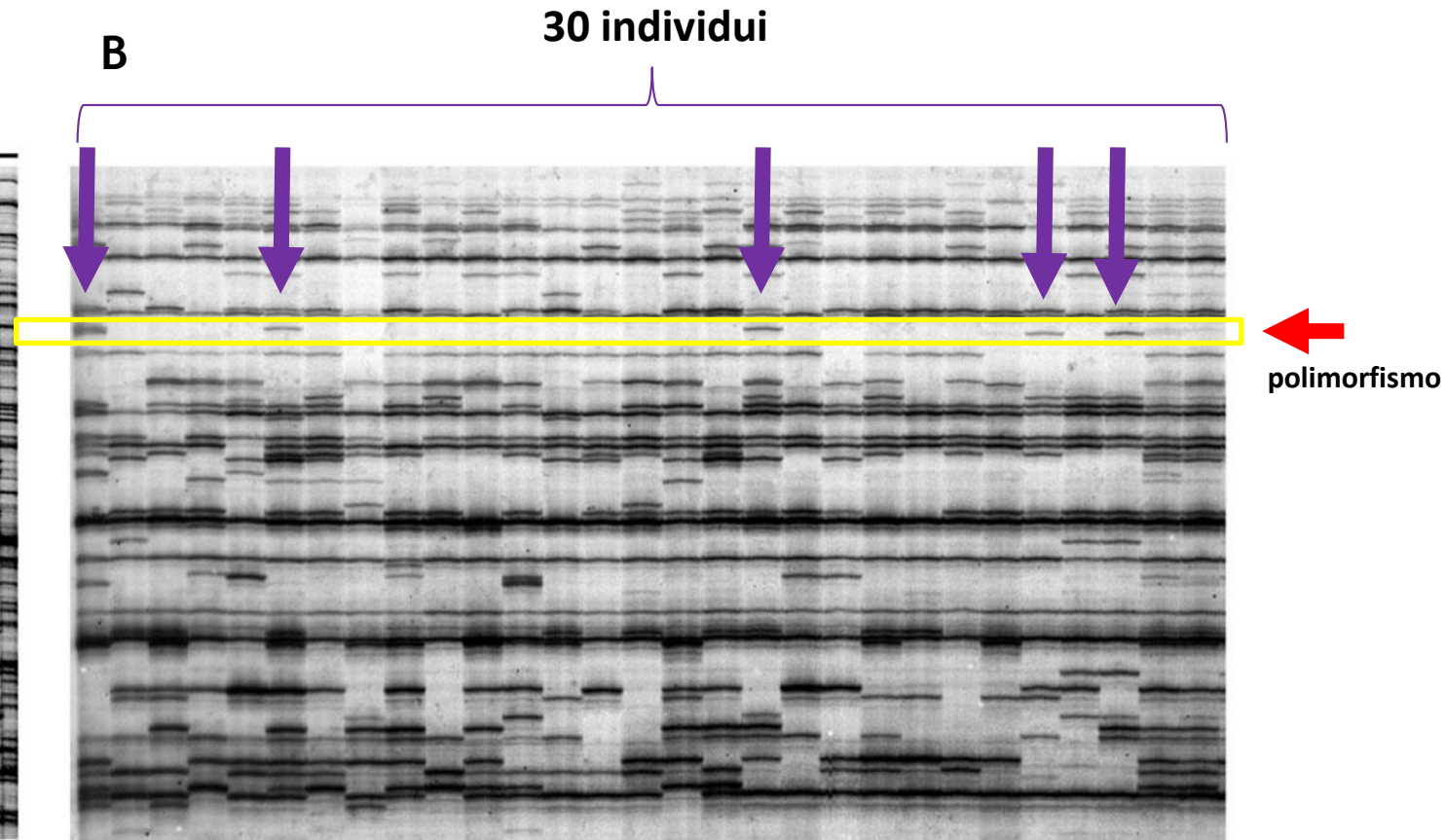
A



## AFLP: peculiarità

- MULTI-LOCUS (50-100 per esperimento)
- NO CONOSCENZE apriori del GENOMA
- COSTI RIDOTTI


B



polimorfismo

# MARCATORI SNP (SINGLE NUCLEOTIDE POLYMORPHISM)

seq_1 (A)	ATGCGGC	G	AT
seq_2 (A)	ATGCGGC	G	AT
seq_3 (A)	ATGCGGC	G	AT
seq_1 (B)	ATGCGGC	A	AT
seq_2 (B)	ATGCGGC	A	AT
seq_3 (B)	ATGCGGC	A	AT



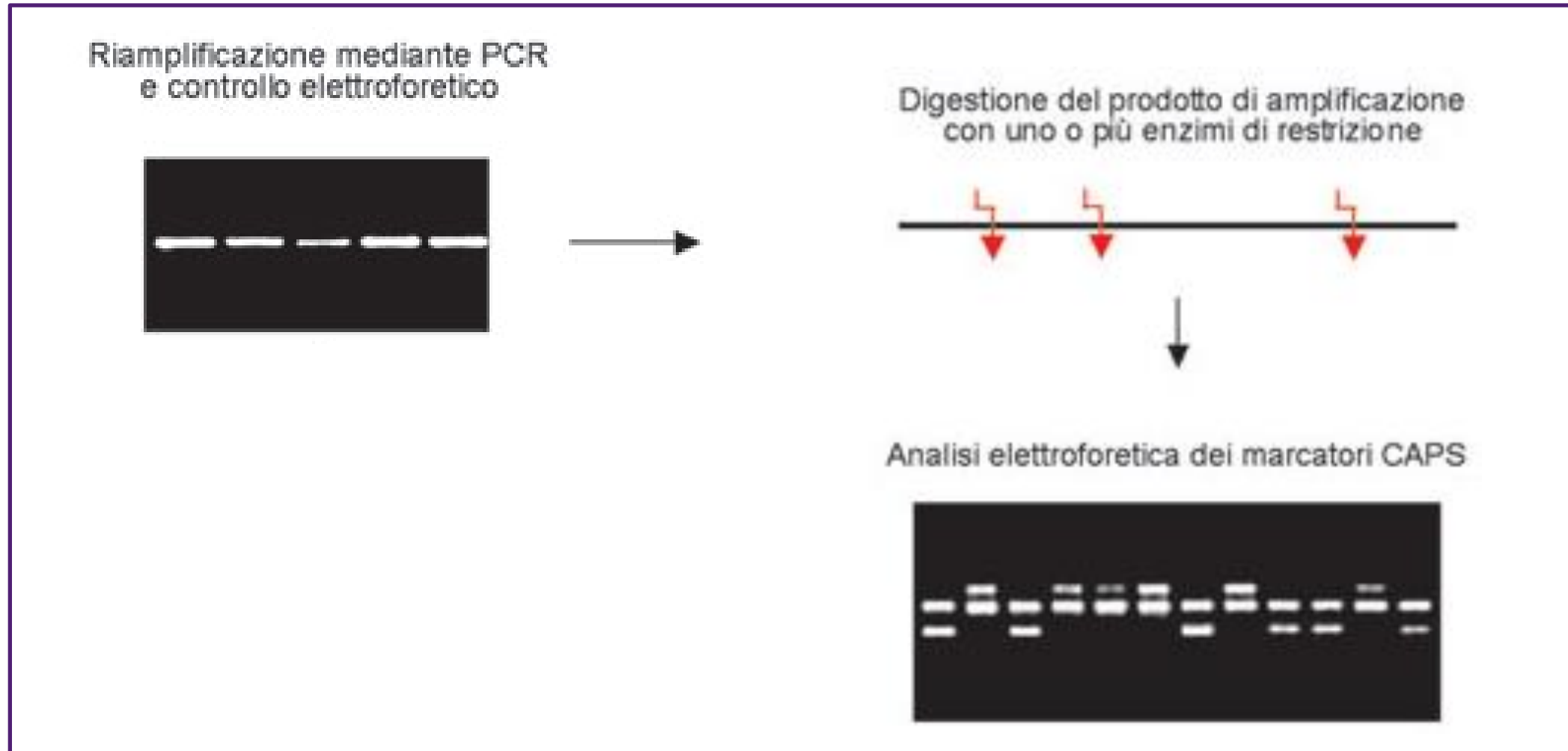
## Definizione

Un polimorfismo a singolo nucleotide è una variazione, del materiale genico a carico di un unico nucleotide tale per cui l'allele polimorfico risulta presente nella popolazione in una proporzione superiore all'1%. Al di sotto di tale soglia si è soliti parlare di variante rara (SNV = single nucleotide variant). Vengono rilevati mediante sequenziamento.

Identificabile mediante allineamento multiplo della stessa sequenza derivante da individui diversi

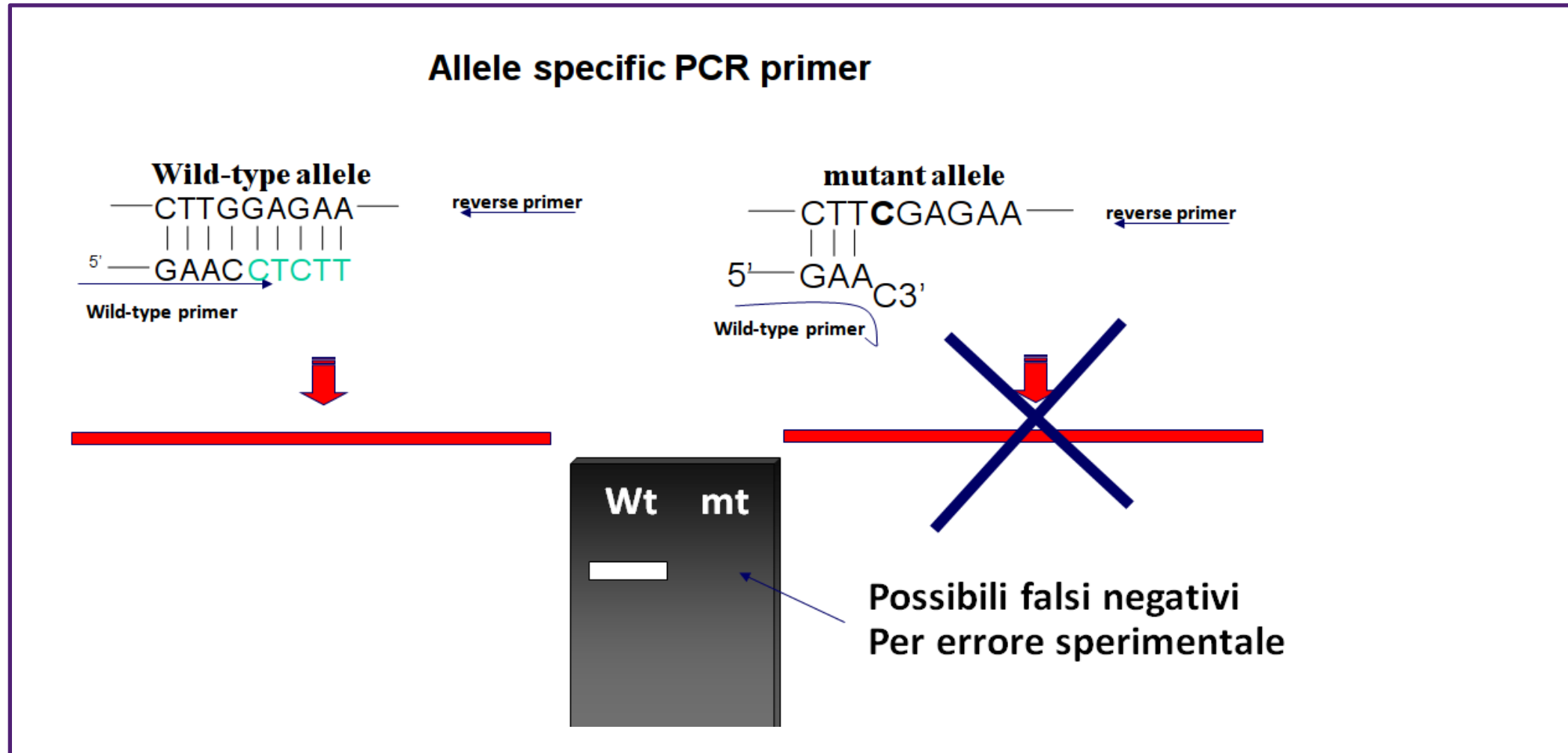
# MARCATORI SNP: RILEVAZIONE

CAPS (cleaved amplified polymorphic sequence)



# MARCATORI SNP: RILEVAZIONE

ARMS (Allele Refractory Mutation System)



# MARCATORI SNP: RILEVAZIONE

## Tetra-primer ARMS

### Mwg2062

Consensus sequence

```
.....TTGTG TCAAGCATAT CGGTTGCTCT TGCTATCGTT
TTGGTGCACA A TCTATGGGG G TAT GCATAC AGCAA CGACA AGGAAAGT GGT GGAGT ACATT
TCAAGAATTA TGCCAATTA T TGGC GTGGC R TTC TTGT TTG ATGACATGCA GTGTGTTCTT
TCAAGGTATTA TA AGGAGGAT CCTA TTGTTT TCGAACGTGC TG GAAATA TG CGACTTTA TT
AATA AGCATA TTTTTCAGG TA TTGTTA GG GGCTGCGGCT TTCAAAAGAT TGGCTCCTA T
GTCAAT CTTA GTGCGTACTA CCT.....
```

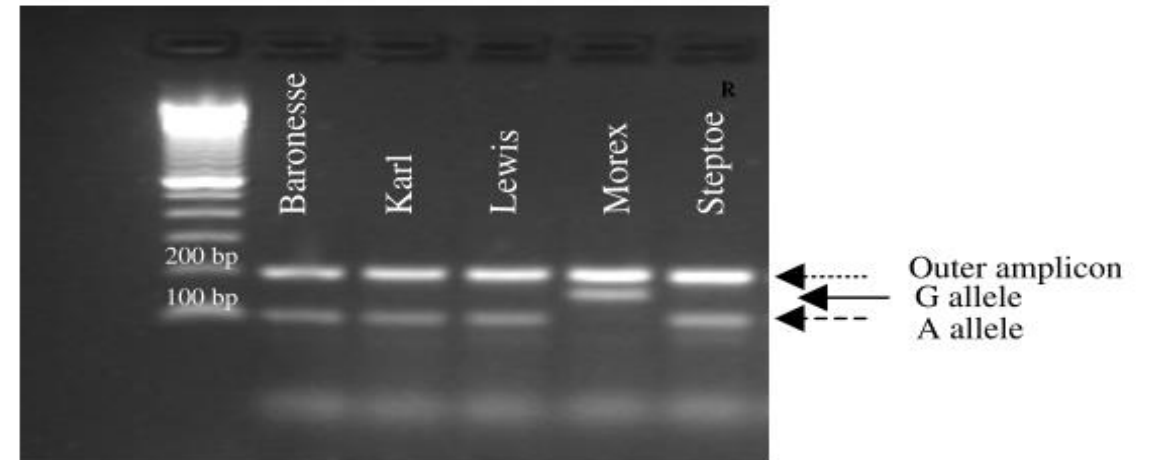
Morex: allele G

Baronesse, Karl, Lewis: allele A

@325 .....gccaattattggcgtggc(**g/a**)ttctgtttgat....

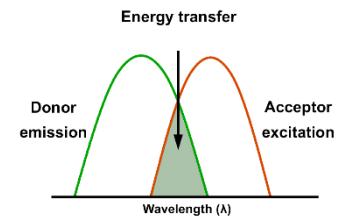
PCR and gel electrophoresis

**B**



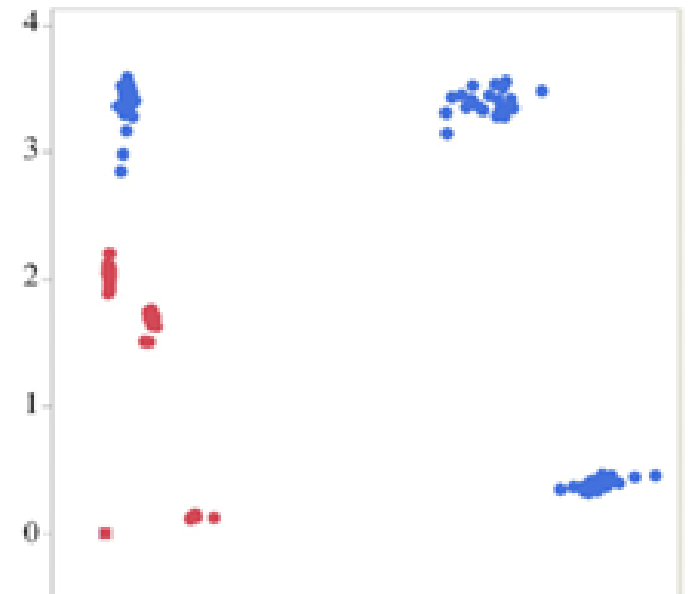
# MARCATORI SNP: RILEVAZIONE

## SAGGI SNP MULTIPLEXED



Più loci  
in contemporanea  
*(fluorescenza)*

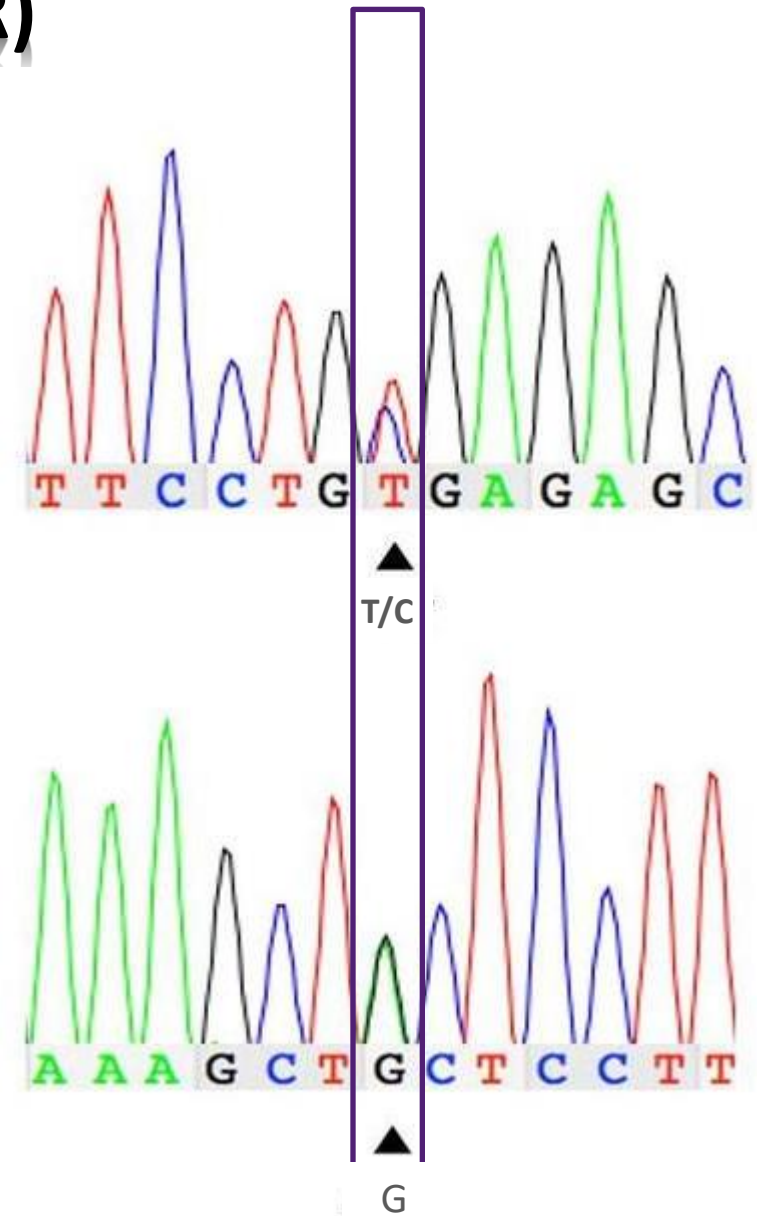
- Allele Specific Hybridisation
- Taqman
- Primer Extension
- Oligonucleotide Ligation assay (OLA)



# MARCATORI SNP: RILEVAZIONE (SANGER)

## Rilevazione dei marcatori SNPs

Sono visualizzati attraverso sequenziamento di prodotti di PCR di campioni di DNA isolati da un certo numero di individui geneticamente differenziati



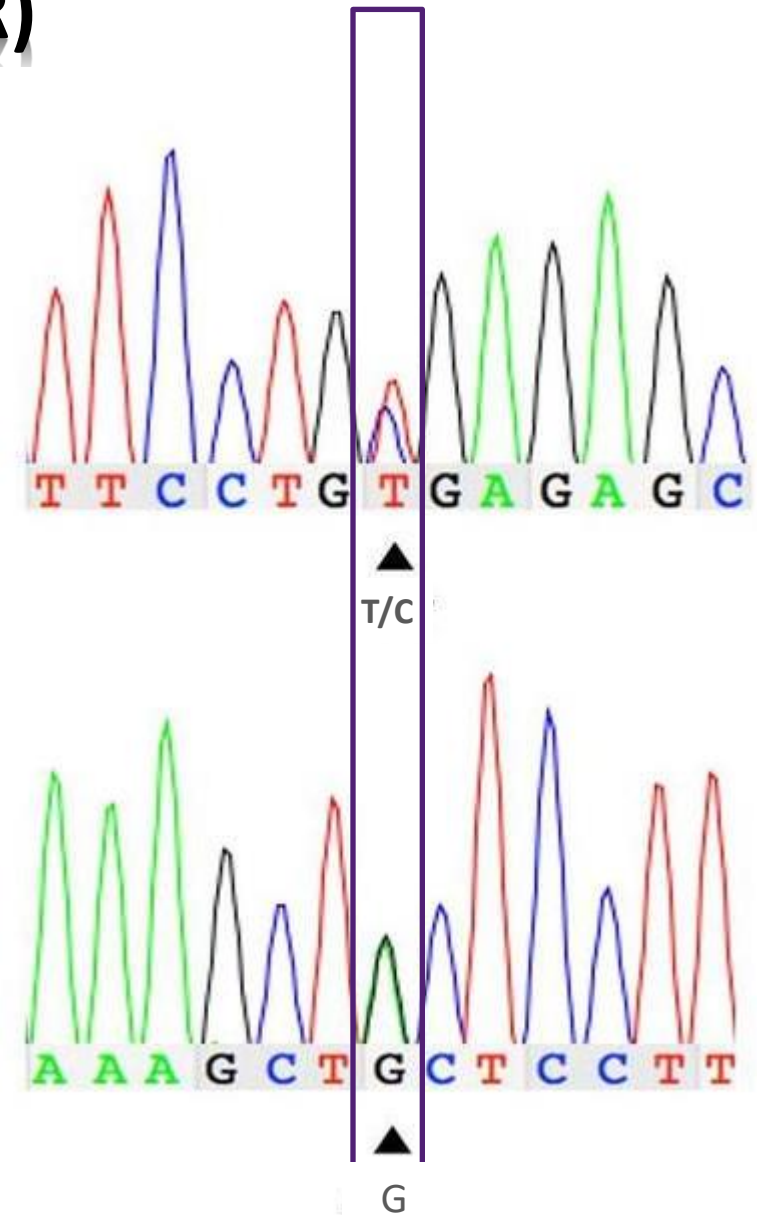


# MARCATORI SNP: RILEVAZIONE (SANGER)

## Rilevazione dei marcatori SNPs

Sono visualizzati attraverso sequenziamento di prodotti di PCR di campioni di DNA isolati da un certo numero di individui geneticamente differenziati

5 euro per frammento di 500 bp



# MARCATORI SNP: RILEVAZIONE (ILLUMINA)

Oggi è possibile sequenziare direttamente i campioni da analizzare (anche se essi sono in numero di > 100) e a **prezzi contenuti**



**15 euro per un miliardo di basi**



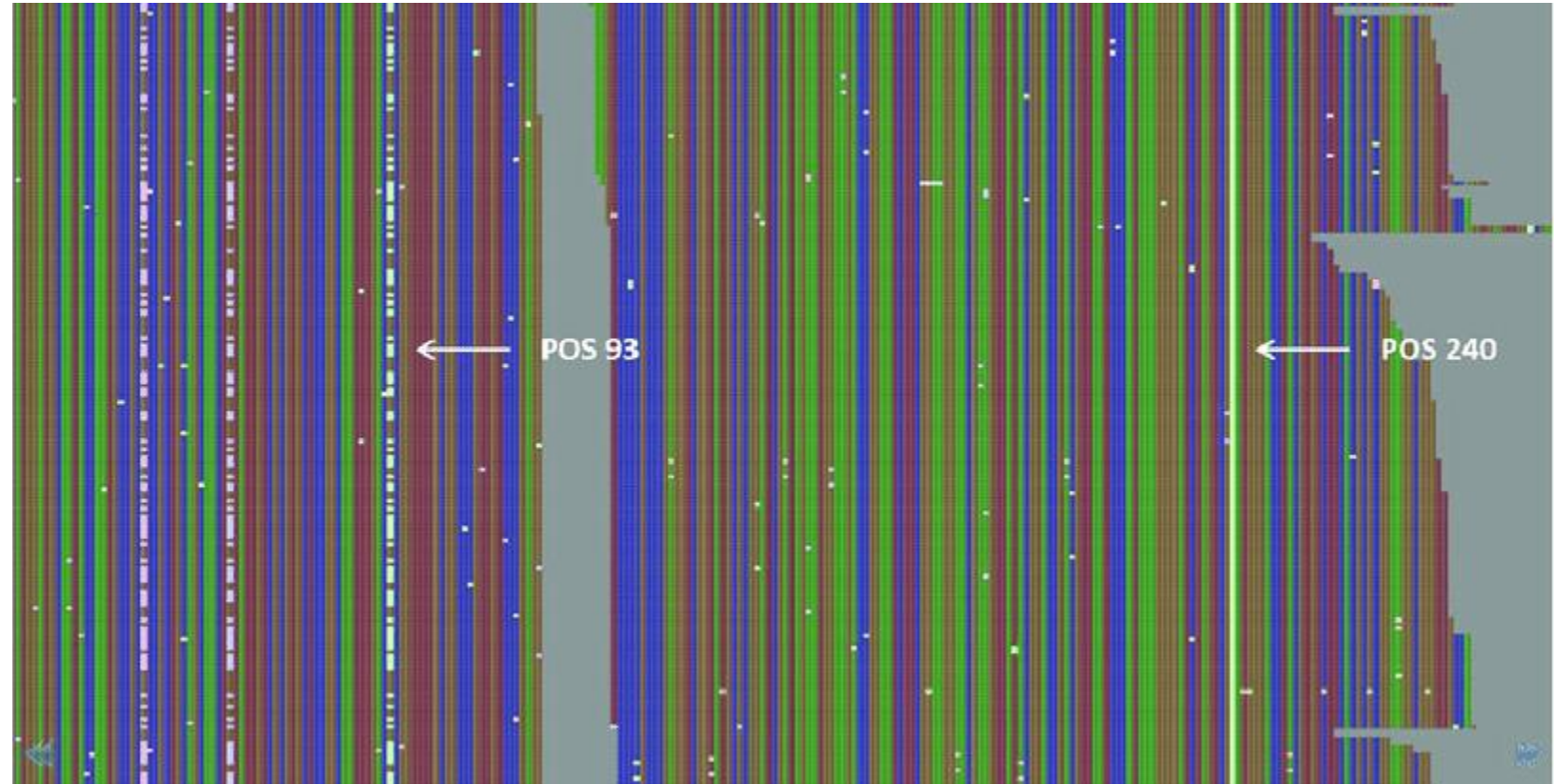
**450 euro per sequenziare  
un individuo  
(con genoma di 1Gbp, 30X)**

# SNP: SEQUENZIAMENTO «WHOLE GENOME»

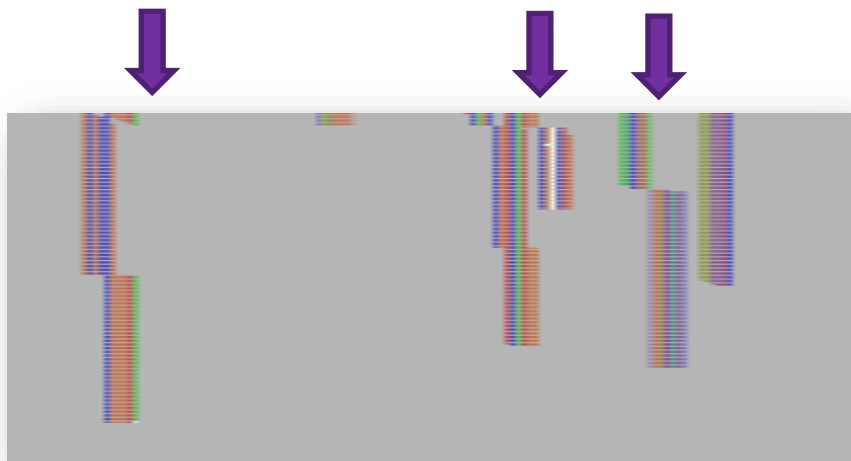
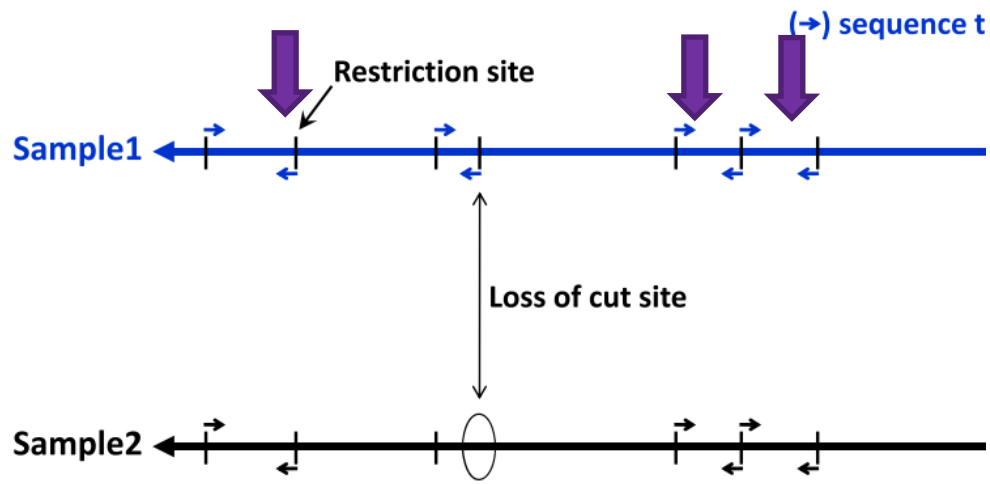
Serie di 3 SNP consecutive



Pila di 50 reads



# METODOLOGIE “REDUCED COMPLEXITY”

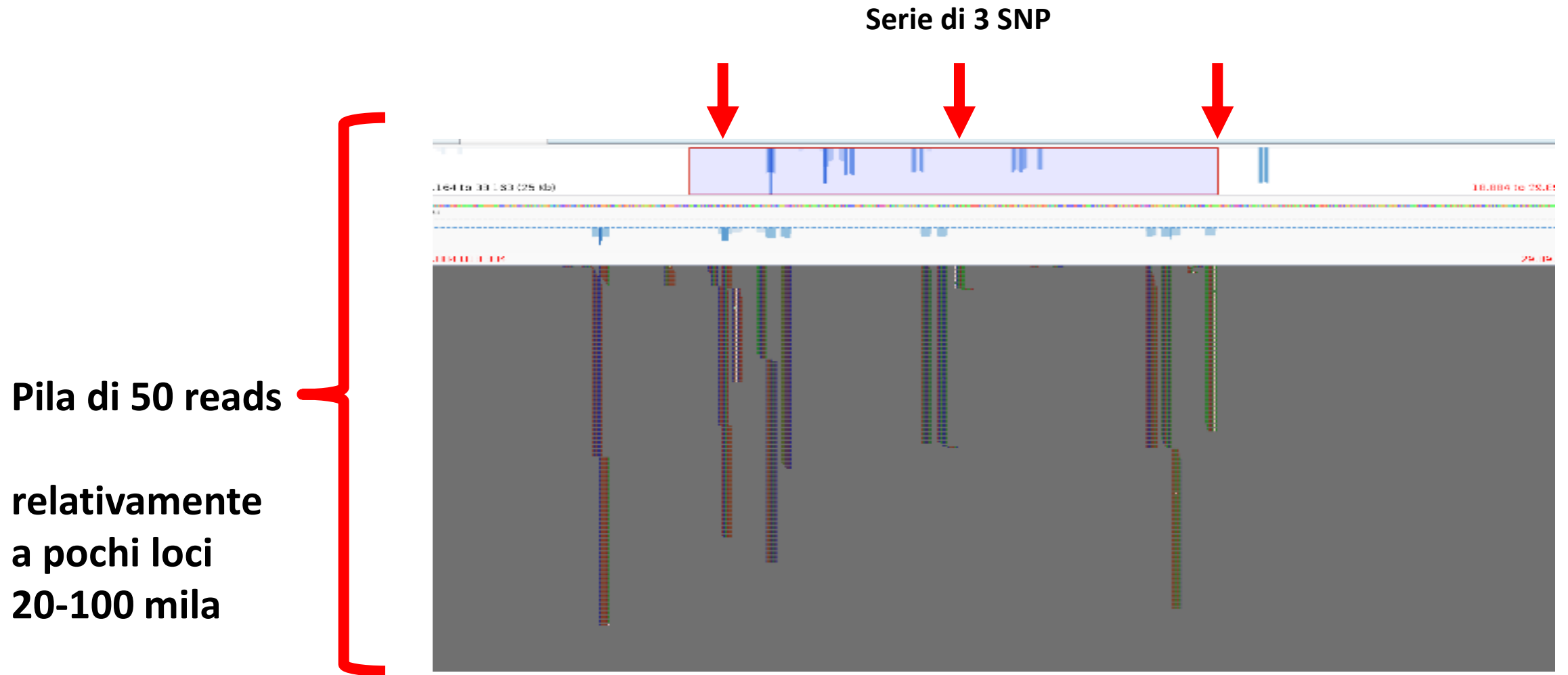


Tramite metodologie “reduced complexity” possono accadere due cose:

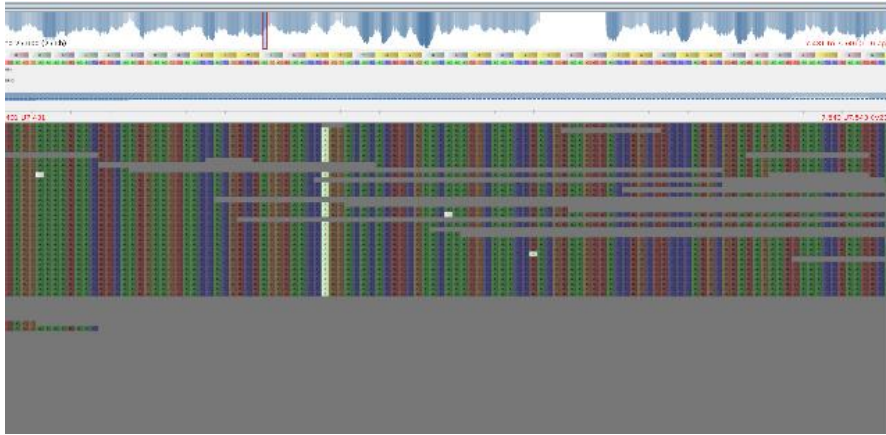
1) Si può perdere un sito di restrizione (e pertanto il frammento non viene sequenziato). E' è un polimorfismo che non viene visto direttamente ma che si può contare

2) Si osservano SNP nei frammenti sequenziati

# SNP: SEQUENZIAMENTO A RIDOTTA COMPLESSITÀ



# «WHOLE GENOME» VS RIDOTTA COMPLESSITÀ

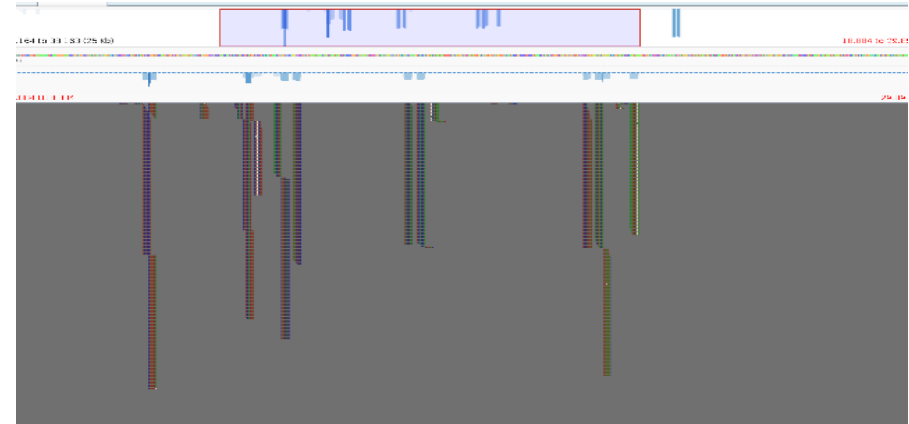


whole genome seq



15 euro / 1 Gbp

**450 euro per sequenziare**  
un individuo  
(con genoma di 1Gbp, 30X)



Reduced complexity seq



15 euro / 1 Gbp

**30 euro per sequenziare**  
un individuo  
(con genoma di 1Gbp, 30X)

# METODOLOGIE “REDUCED COMPLEXITY”

Le più importanti “campionano” la situazione nell’intorno di un tipo di siti di restrizione

**Le tecniche oggi disponibili per questo sono:**

- Radseq (a uno o a due enzimi)
- CROPS (Complexity reduction of polymorphic sequences)
- GBS (genotyping by sequencing)

OPEN ACCESS Freely available online

 PLoS one

## A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species

Robert J. Elshire<sup>1</sup>, Jeffrey C. Glaubitz<sup>1</sup>, Qi Sun<sup>2</sup>, Jesse A. Poland<sup>3</sup>, Ken Kawamoto<sup>1</sup>, Edward S. Buckler<sup>1,4</sup>, Sharon E. Mitchell<sup>1\*</sup>

# METODOLOGIE “REDUCED COMPLEXITY”

Altre “campionano” la situazione nell’intorno di regioni selezionate dall’utente  
(sequenze esoni/introniche)

**Le tecniche oggi disponibili per questo sono:**

- SPET (Single Primer Enrichment Technology)
- ARRAY (infinium, Illumina)



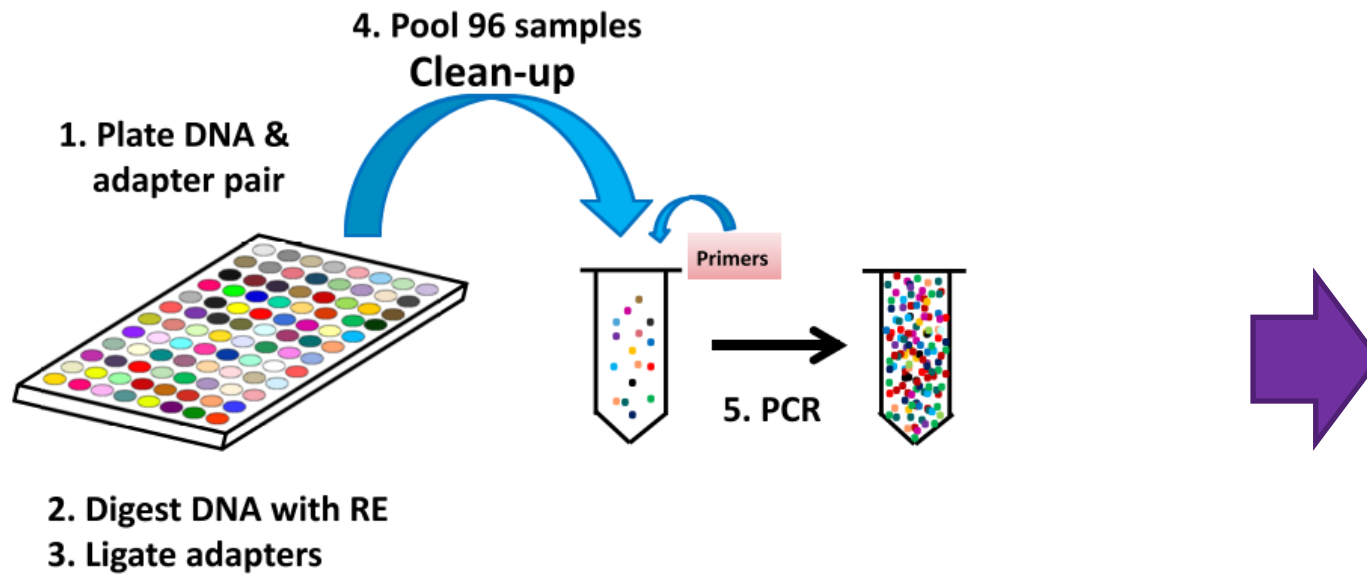
## Single Primer Enrichment Technology (SPET) for High-Throughput Genotyping in Tomato and Eggplant Germplasm

*Lorenzo Barchi<sup>1</sup>, Alberto Acquadro<sup>1</sup>, David Alonso<sup>2</sup>, Giuseppe Aprea<sup>3</sup>,  
Laura Bassolino<sup>4</sup>, Olivia Demurtas<sup>3</sup>, Paola Ferrante<sup>3</sup>, Pietro Gramazio<sup>2</sup>, Paola Mini<sup>3</sup>,  
Ezio Portis<sup>1</sup>, Davide Scaglione<sup>5</sup>, Laura Toppino<sup>4</sup>, Santiago Vilanova<sup>2</sup>, María José Díez<sup>2</sup>,  
Giuseppe Leonardo Rotino<sup>4</sup>, Sergio Lanteri<sup>1\*</sup>, Jaime Prohens<sup>2\*</sup> and Giovanni Giuliano<sup>3</sup>*

OPEN ACCESS



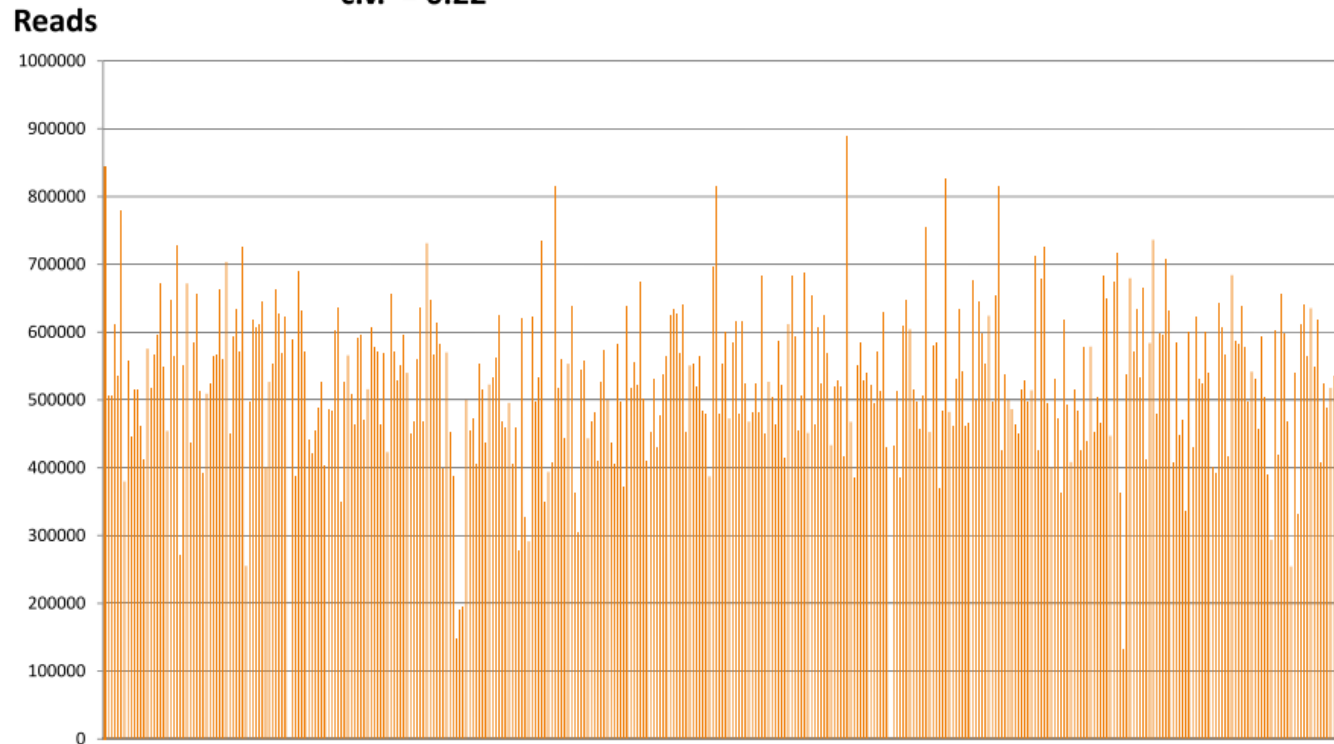
# METODOLOGIE “REDUCED COMPLEXITY”



# METODOLOGIE “REDUCED COMPLEXITY”

## 384-plex GBS Results for Maize

Mean read count per line = 528,000  
c.v. = 0.22



**Fino a 384 piante  
sequenziate insieme**

# EDV E SEQUENZIAMENTO

Da un punto di vista legislativo una nuova varietà deve essere

**DUS**

**Distinta – Uniforme - Stabile**



Quali soglie per definire una varietà **DISTINTA**  
in caso di sequenziamento Illumina?

# EDV E SEQUENZIAMENTO

Da un punto di vista legislativo una nuova varietà deve essere

## DUS

### Distinta – Uniforme - Stabile



Quali soglie per definire una varietà **DISTINTA**  
in caso di sequenziamento Illumina?



## DIPENDE

# EDV E SEQUENZIAMENTO

**Le varietà hanno caratteristiche diverse in relazione al sistema riproduttivo**

**autogamia**



linea pura

**allogamia**



popolazioni di piante selezionate che si inter-incrociano

**prop.vegetativa**



clone



**Un ulteriore categoria di varietà è rappresentata dagli ibridi F1** (linee inbred – eterosi)

# EDV E SEQUENZIAMENTO

GAMIA	Struttura varietà	SOGLIA EDV	1gb genome
Allogama	popolazioni di piante selezionate che si inter-incrociano	?	< 5 su 5.000 LOCI
Autogama	Linea pura	?	?
Prop. Vegetativ.	Clone	?	?
Ibrido F1	Piante uniformi	?	?

- Occorre portare avanti un lavoro preliminare (es. g2pSOL)
- SPECIE per SPECIE
- Occorre sequenziare più individui della stessa varietà per verificarne la variabilità
- Non ancora in vigore in UPOV



# GRAZIE PER L'ATTENZIONE

