

Il Sequenziamento Tecniche e Strategie



Metodi per il sequenziamento del DNA

Metodo della degradazione chimica

(Maxam & Gilbert, 1977)

La sequenza di una molecola di ds DNA viene determinata **mediante trattamento con reagenti chimici** che tagliano la sequenza nelle posizioni di specifici nucleotidi

La molecola di ds DNA viene marcata al 5' di ogni filamento (**gdATP**) e poi denaturata.

Metodo a terminazione di catena

(Sanger et al., 1977)

La sequenza di una molecola di ss DNA viene determinata **mediante sintesi enzimatica** di catene polinucleotidiche complementari che terminano nelle posizioni di specifici nucleotidi

Sequenziamento secondo Maxam e Gilbert

- 1 Marcatura terminale al 5' del DNA a doppio filamento
- 2 Denaturazione e separazione dei due filamenti
- 3 Il DNA a singola elica viene suddiviso in quattro campioni, ognuno dei quali viene trattato con un reagente chimico che degrada **una o due** delle 4 basi del DNA

G = DMS + high T + NaOH

G + A = DMS + HCl

C + T = hydrazine + piperidine

C = hydrazine + piperidine in NaCl 1,5 M

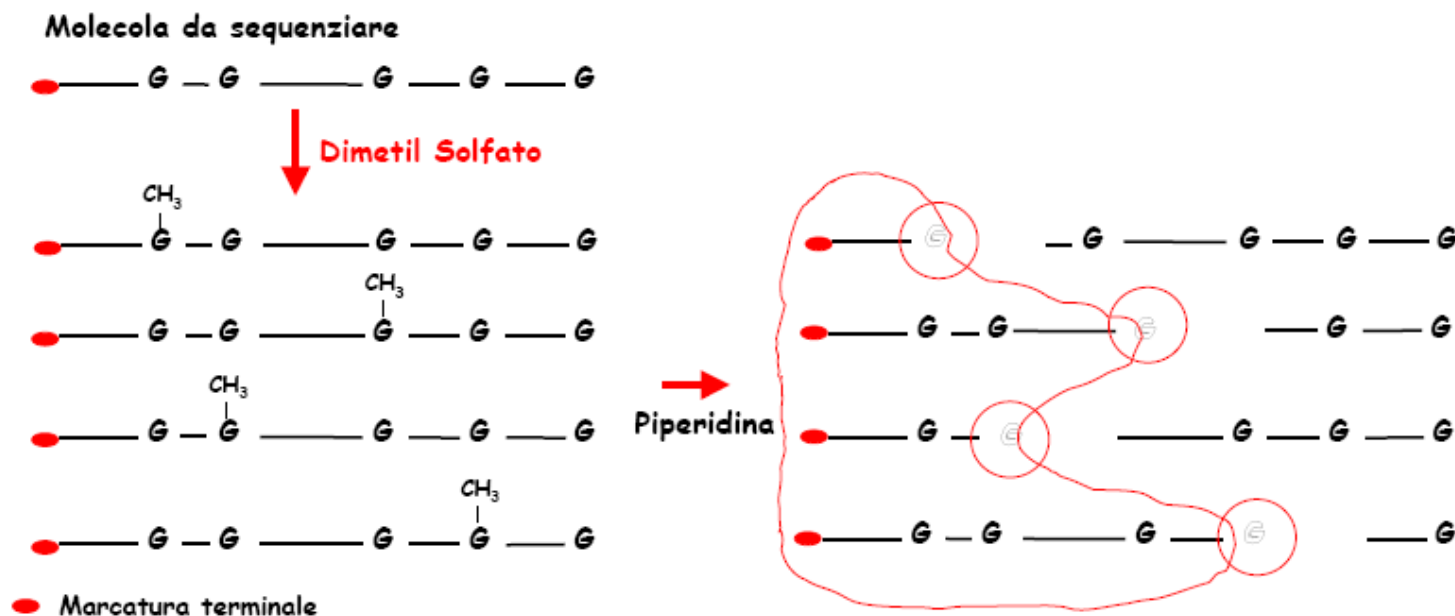
Si utilizzano concentrazioni dei reagenti tali che ogni singola copia del filamento di DNA marcato venga degradata una sola volta; poichè il filamento è presente in molteplici copie, statisticamente, considerando tutte le copie, ogni base del filamento verrà degradata almeno una volta e quindi si produrranno tanti frammenti di lunghezza diversa a seconda del sito di taglio.

- 4 Separazione dei frammenti marcati mediante elettroforesi su gel di poliacrilammide e visualizzazione dei risultati mediante autoradiografia

Sequenziamento secondo Maxam e Gilbert 2

- Dimetil solfato attacca un gruppo metile all'anello purinico delle G. Il reagente è aggiunto in quantità limitante per modificare una G per ogni filamento
- La piperidina rimuove l'anello purinico modificato e taglia la molecola di DNA a monte del sito privo di base che si crea.
- Non essendoci un reagente specifico per A o T, le reazioni effettuate sono:

G, A+G, C, C+T



Sample DNA

- 1 Preparation of homogeneous single-strand DNA
- 2 Addition of ^{32}P as 5' phosphate
- 3 Cleavage at specific nucleotides

$^*\text{ATTGACTTAGCC}'$

$^*\text{ATTGACTTAGCC}$

G reaction A reaction, with some G cleavage (underlined) T reaction, with some C cleavage (underlined) C reaction

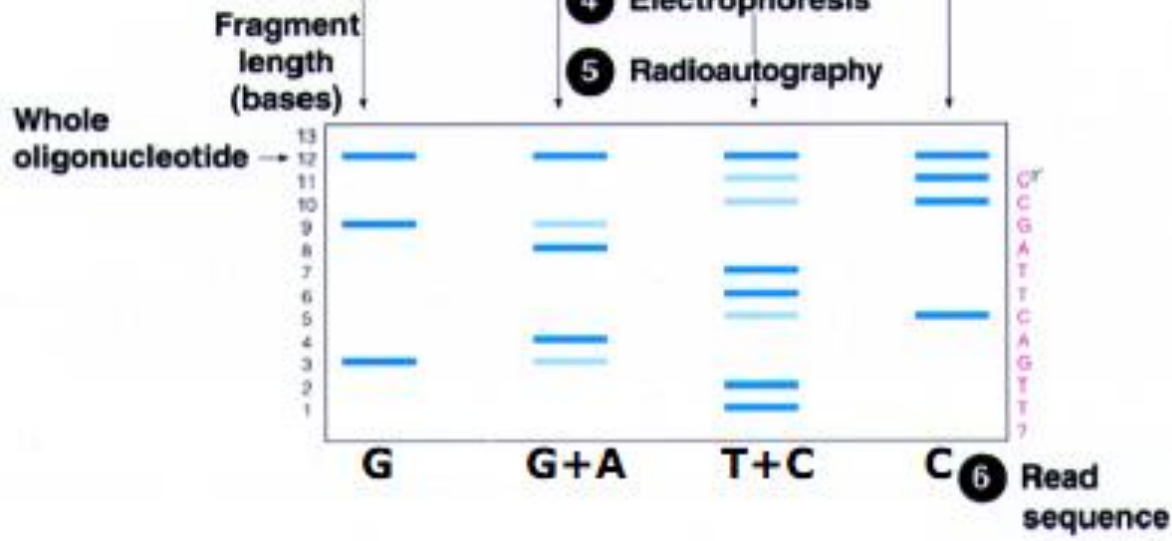
$^*\text{ATTGACTTAGCC}$
 $^*\text{ATTGACTTA}$
 $^*\text{ATT}$

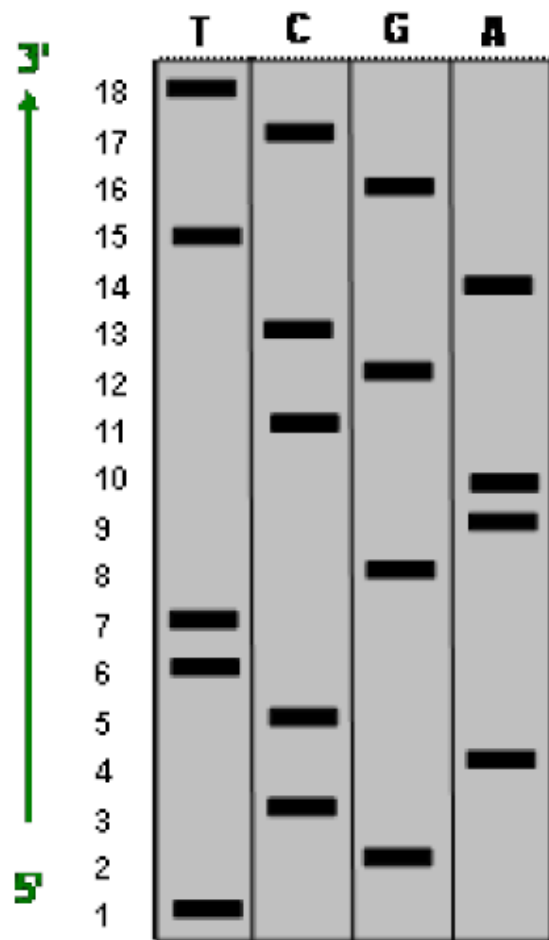
$^*\text{ATTGACTTAGCC}$
 $^*\text{ATTGACTTA}$
 $^*\text{ATTGACTT}$
 $^*\text{ATTG}$
 $^*\text{ATT}$

$^*\text{ATTGACTTAGCC}$
 $^*\text{ATTGACTTAGC}$
 $^*\text{ATTGACTTAG}$
 $^*\text{ATTGACT}$
 $^*\text{ATTGAC}$
 $^*\text{ATTGA}$
 $^*\text{AT}$
 $^*\text{A}$

$^*\text{ATTGACTTAGCC}$
 $^*\text{ATTGACTTAGC}$
 $^*\text{ATTGACTTAG}$
 $^*\text{ATTGA}$

- 4 Electrophoresis
- 5 Radioautography





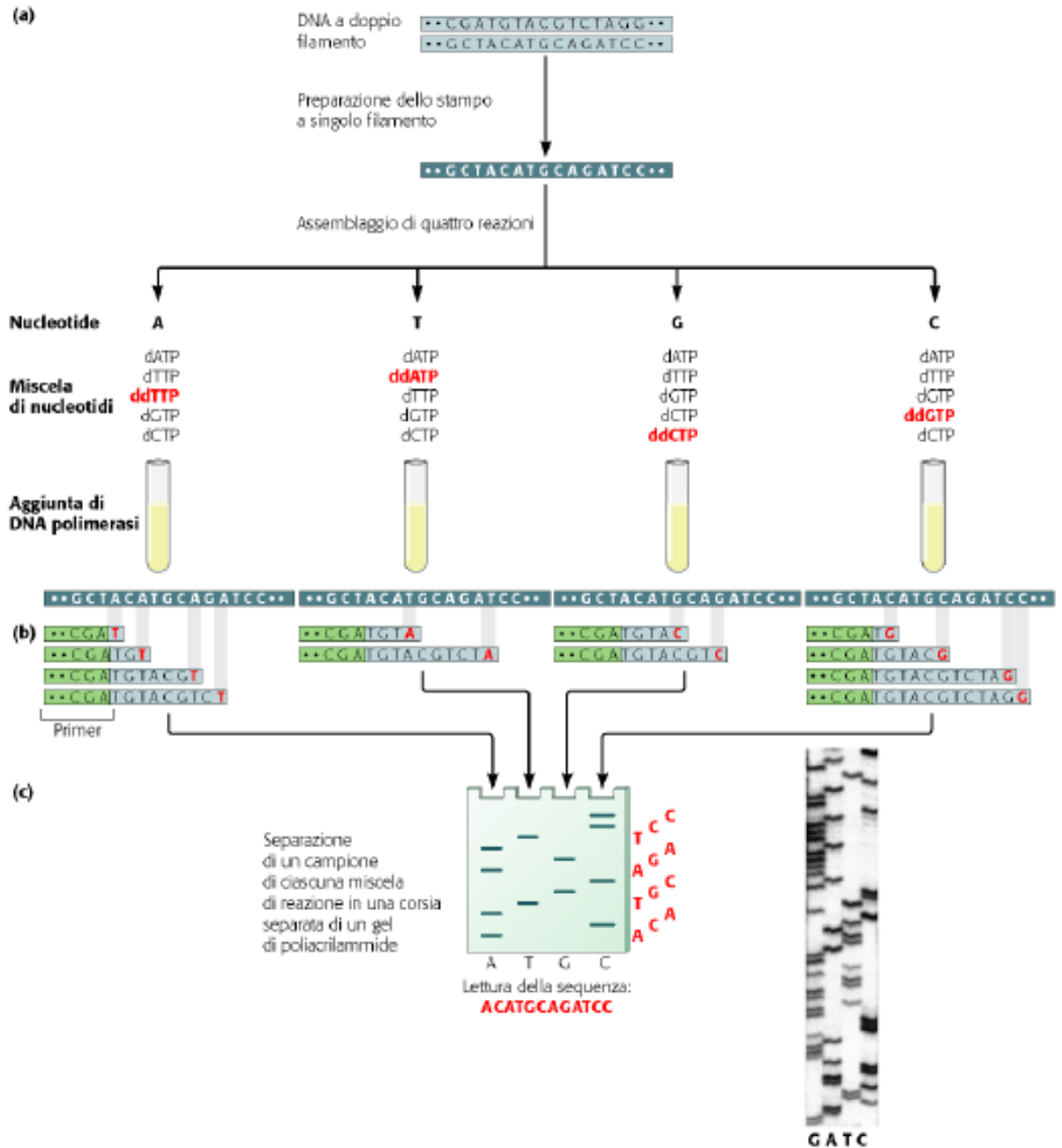
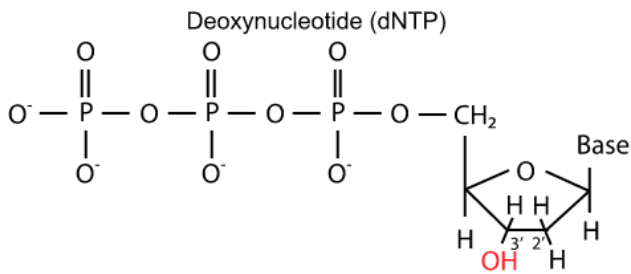
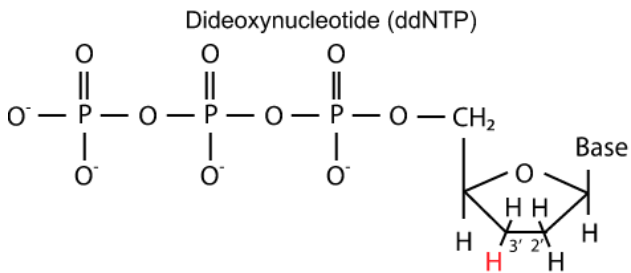
The sequence is read
from bottom to top

5'-TGCACCTTGAACGCATGCT-3'

Direction of
electrophoresis

Sequenziamento secondo Sanger 1

- La reazione si basa sulla sintesi del filamento complementare al ssDNA che serve da stampo (templato)
- La reazione è catalizzata da una DNA pol. con le seguenti caratteristiche:
 - Elevato grado di processività
 - Attività esonucleasica 5'-3' e 3'-5' trascurabile
 - Attività esonucleasica trascurabile
- La reazione richiede dNTP (dATP, dGTP, dCTP, dTTP) come substrato
- Inoltre alla reazione viene aggiunto un **ddNTP** che viene incorporato nella catena dalla polimerasi che non distingue tra dNTP e ddNTP.
- All'incorporazione del ddNTP si arresta la sintesi perché manca il gruppo ossidrilico al 3' necessario per la formazione del legame con il nucleotide successivo
- Siccome è presente sia il dNTP (es. dATP), che il ddNTP (es. ddATP), la sintesi non si interrompe alla prima T dello stampo ma prosegue per X nucleotidi fino alla casuale incorporazione del ddATP.
- Il risultato è un insieme di nuove catene di diversa lunghezza tutte terminanti con un ddNTP
- Si effettuano 4 reazioni distinte ognuna addizionata di un ddNTP (A,G,C,T)



a) Sanger's 'chain-termination' Sequencing method

b)

```

A
ATGCA
ATGCAGCGTTA
ATGCAGCGTTACCA

ATGC
ATGCAGC
ATGCAGCGTTAC
ATGCAGCGTTACC

ATG
ATGCAG
ATGCAGCG
ATGCAGCGTTACCATG

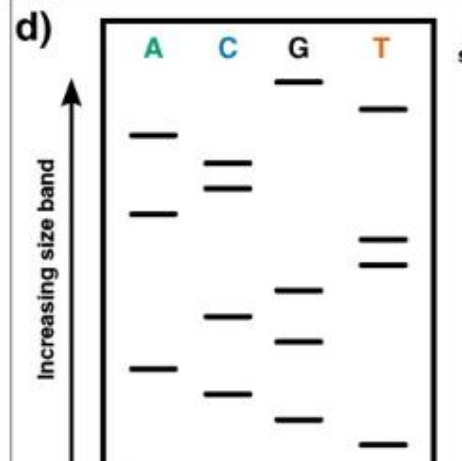
AT
ATGCAGCGT
ATGCAGCGTT
ATGCAGCGTTACCAT
  
```

c)

A+G

C

C+T



Inferred sequence

G
T
A
C
C
C
A
T
T
G
C
G
A
C
G
T

G

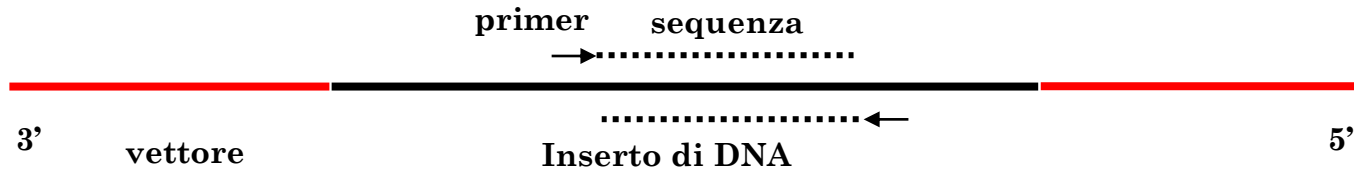
Sequenziamento di DNA clonato

- Il primer (innesco) è necessario per l'inizio della sintesi da parte della DNA pol.
- Il primer stabilisce la regione del DNA stampo che verrà sequenziata:

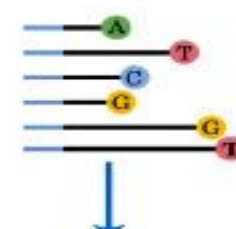
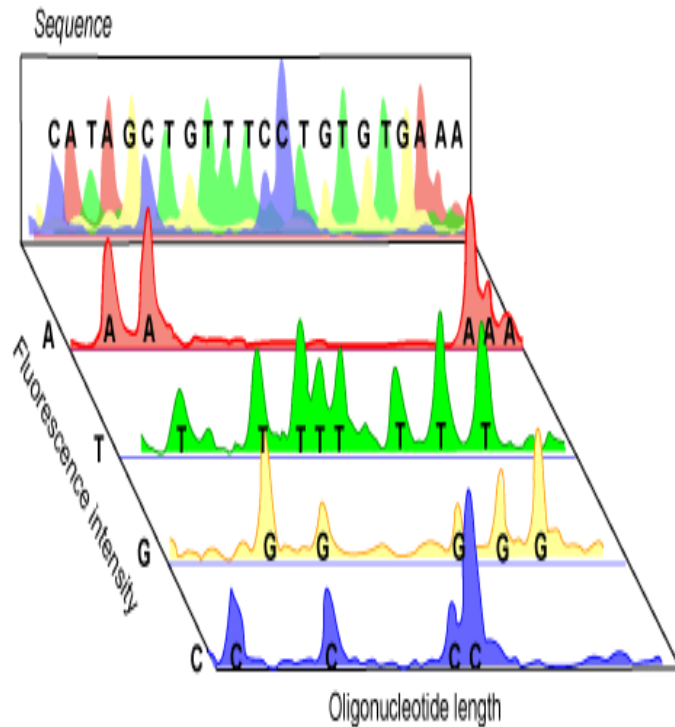
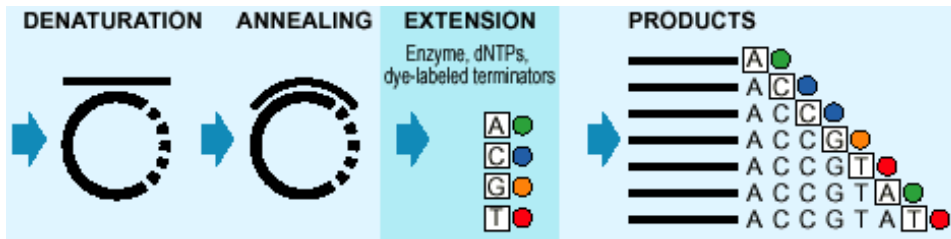
Primer universale



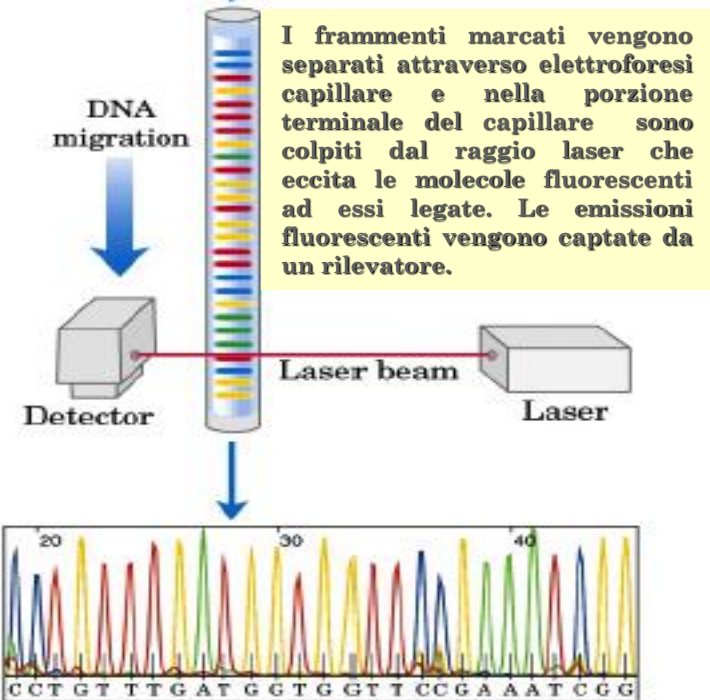
Primer interno, sequenza-specifico



Dye Terminator Labeling Strategy



Coniugando a ciascun ddNTP un diverso marcatore fluorescente, è possibile effettuare le quattro reazioni di sequenziamento in un'unica provetta

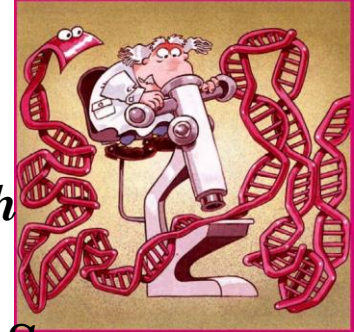


Le informazioni vengono integrate e trasformate in picchi di colore diverso, con aree proporzionali all'intensità di emissione.

Perché si sequenzia???

“... [A] knowledge of sequences could contribute much to our understanding of living matter.”

Frederick Sanger



Nella ricerca di base:

clonaggi

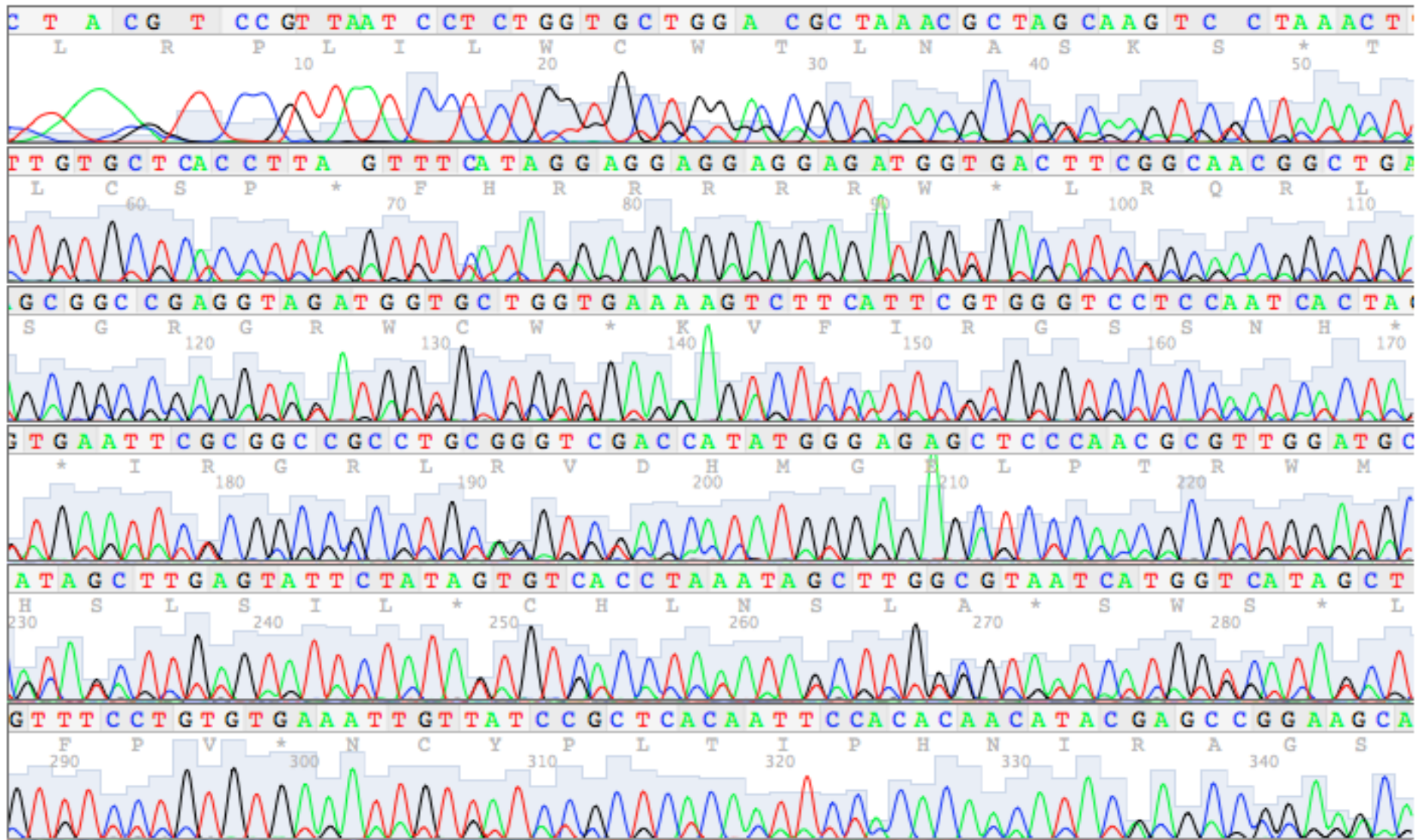
alleli mutanti

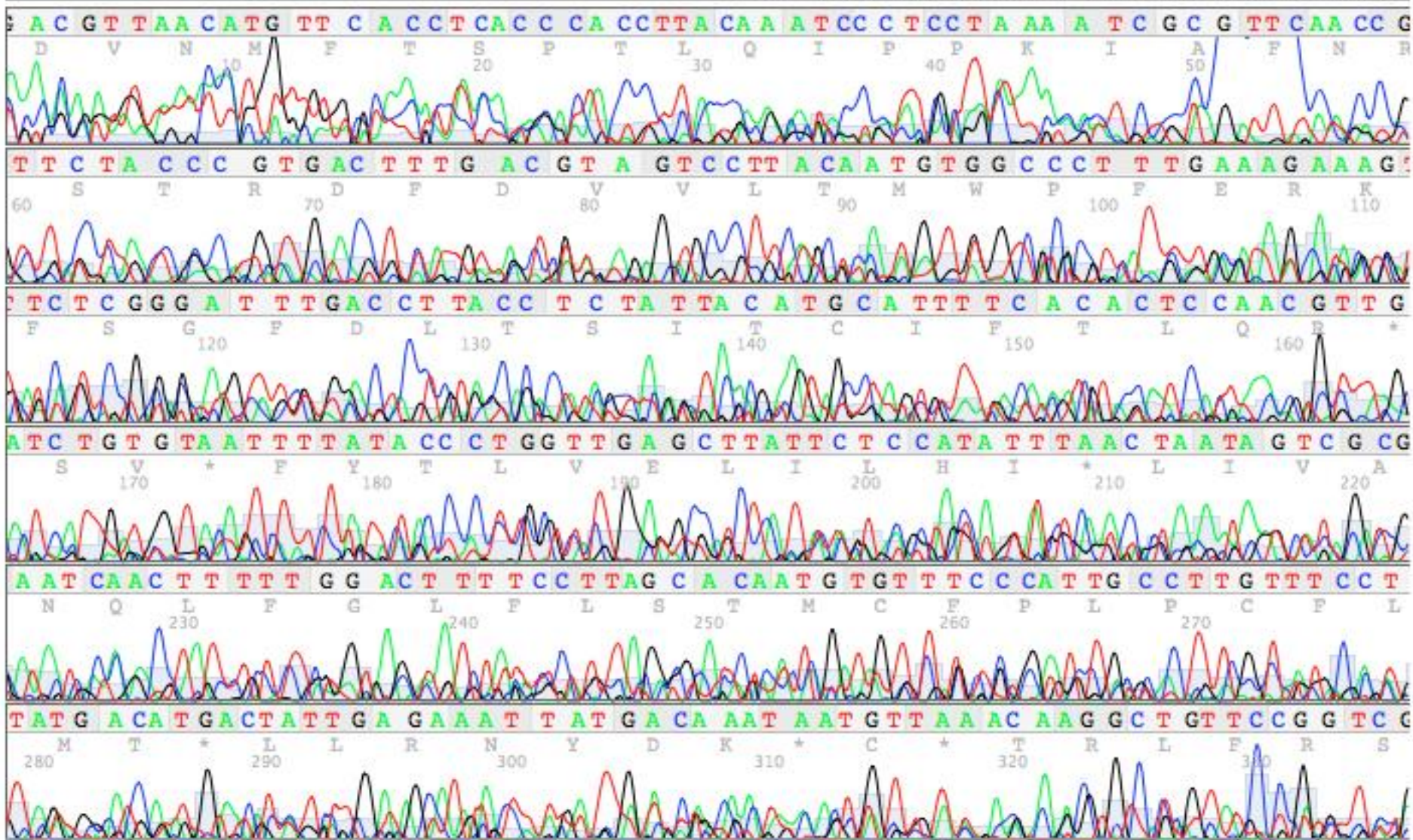
promotori

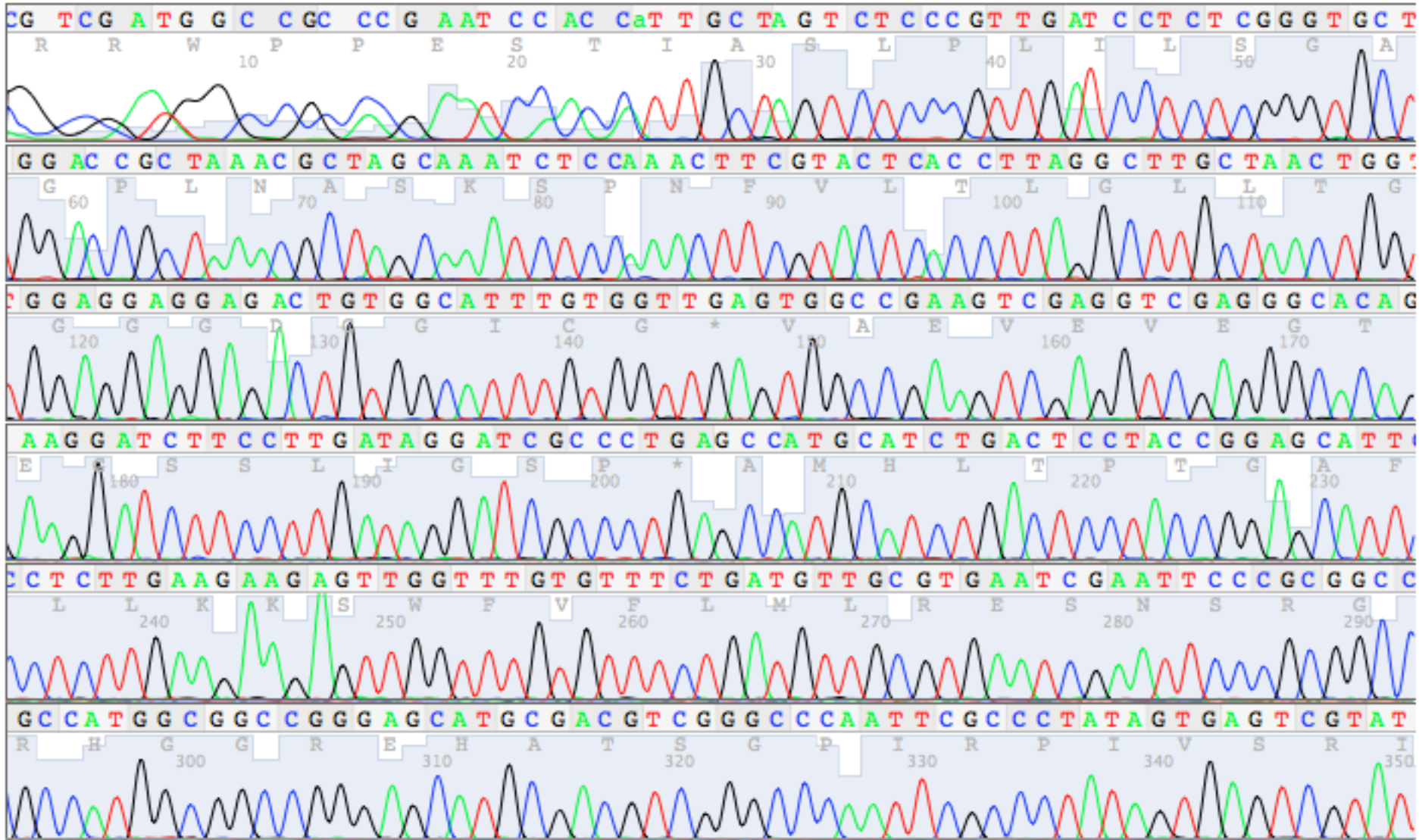
Nella ricerca applicata:

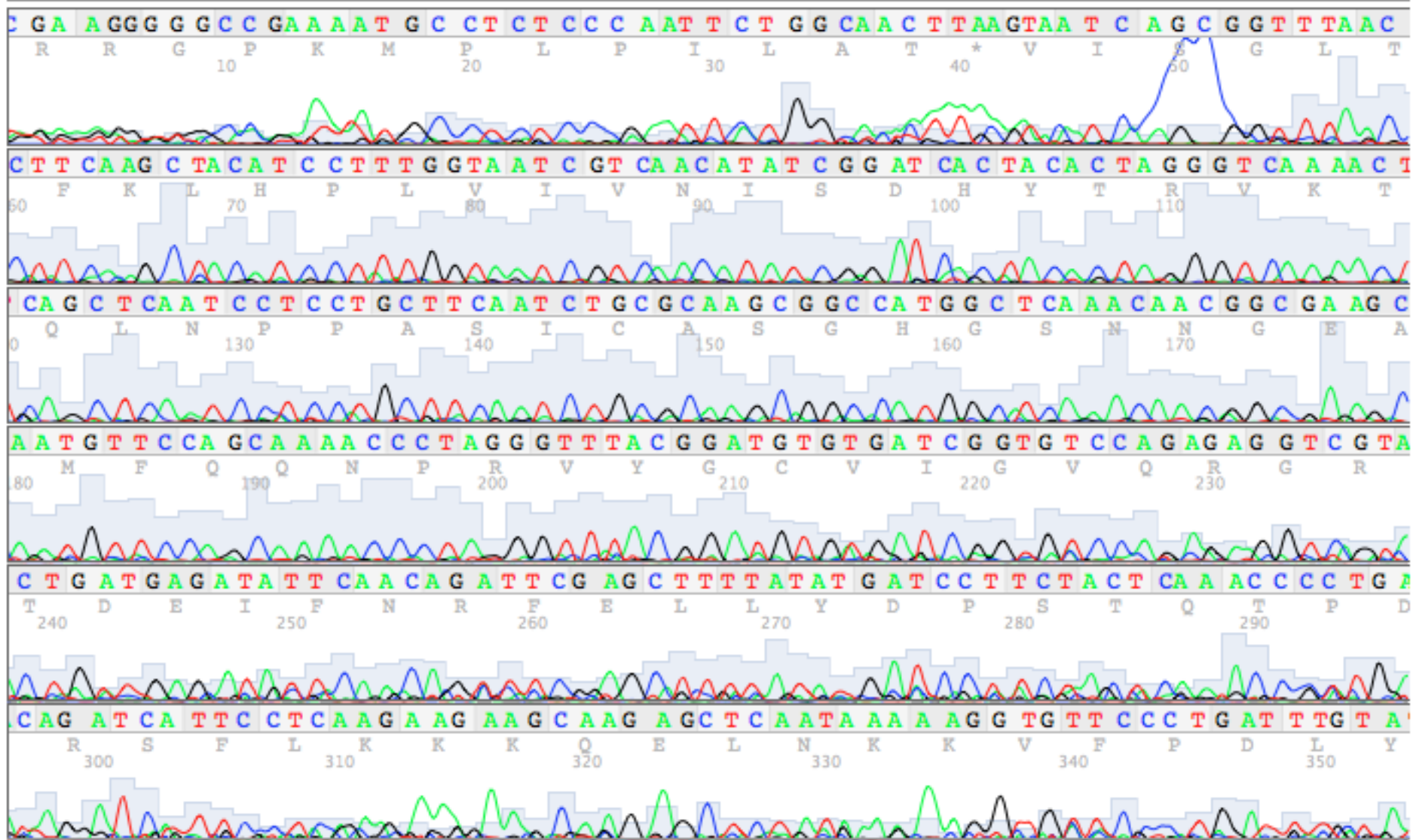
Diagnosi di malattie genetiche

Progetti genome-wide: Sequenziamento genomi, Profili trascrittomici









Mappe genetiche e fisiche

La mappatura si riferisce al processo che permette di determinare le posizioni relative di marcatori (come geni, varianti e altre sequenze di DNA di interesse) all'interno di un cromosoma o di un genoma.

Esistono due approcci per la mappatura:

la mappatura fisica, che permette di stabilire mappe basate sulle distanze fisiche tra marcatori

la mappatura genetica, che permette di stabilire mappe basate sulla frequenza con cui due marcatori vengono ereditati insieme.

Oggi, l'approccio più efficiente per la mappatura prevede il sequenziamento di un genoma e quindi l'utilizzo di programmi per computer per analizzare la sequenza per identificare le posizioni dei marcatori.

Mappatura Fisica

- ✓ Fa uso di tecniche di biologia molecolare per costruire mappe basate su tratti distintivi della sequenza
- ✓ Le tecniche principali sono:
 - le **mappe di restrizione**, che localizzano la posizione relativa dei siti di taglio delle endonucleasi di restrizione
 - l'ibridazione in situ con fluorescenza (**FISH**)
 - la mappatura mediante **STS** (Sequence Tagged Site):
una sequenza relativamente breve **single-copy**, facilmente amplificabile mediante PCR (200-500 bp) che può essere amplificata in modo specifico e la cui posizione nel genoma è mappata.

Mappatura Genetica

- ✓ Fa uso di tecniche genetiche che mostrano la posizione di geni o altri marcatori. Le tecniche consistono in **esperimenti di incrocio**
- ✓ I marcatori genici sono utilizzabili quando sono noti alleli mutanti dotati di fenotipo. I fenotipi possono essere **macroscopici** (colore degli occhi in *Drosophila*) ma anche **biochimici** (gruppi sanguigni)
- ✓ I marcatori di DNA devono anch'essi esser presenti in almeno 2 forme alleliche. Se ne conoscono 3 tipi:
 - i Polimorfismi della Lunghezza dei Frammenti di Restrizione (**RFLP**)
 - i Polimorfismi della Lunghezza di Sequenze Semplici (**SSLP**)
 - i Polimorfismi di Singoli Nucleotidi (**SNP**)

**La mappatura genetica è basata
sull'analisi di associazione genetica**

⇒ sulle leggi di Mendel

- ✓ **La risoluzione di una mappa genetica dipende dal numero di crossing over che sono stati individuati. Ciò rappresenta un grosso limite per gli eucarioti superiori x cui si ottiene una progenie poco numerosa**
- ✓ **Le mappe genetiche hanno un'accuratezza limitata, a causa dell'esistenza di “hot spots” di ricombinazione**

3.12 CONFRONTO TRA MAPPE GENETICHE E MAPPE FISICHE

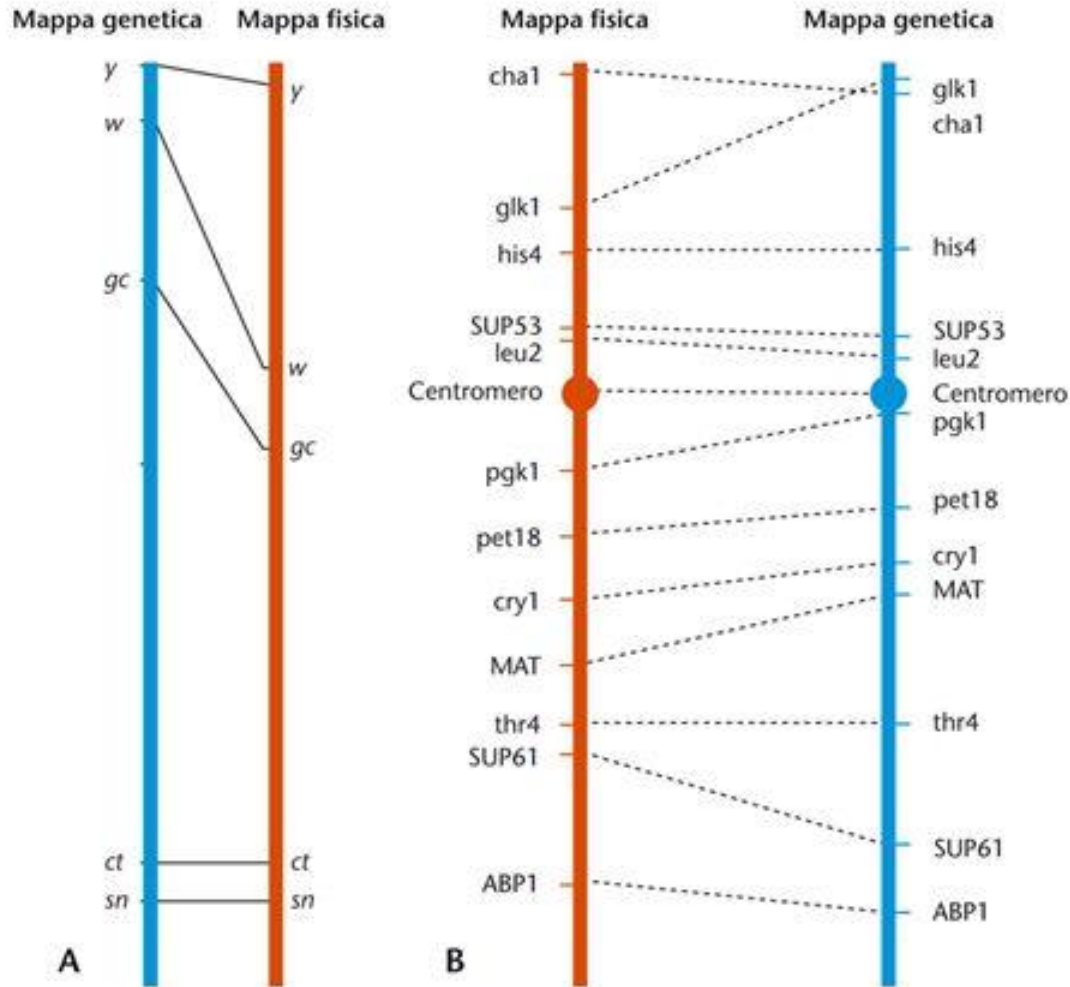


Figura 3.26
Relazione tra distanze genetiche e fisiche in un tratto del cromosoma X politenico di *Drosophila* (A).
Relazione tra distanze genetiche e fisiche in *S. cerevisiae* (B).

POLIMORFISMO

La presenza nella popolazione di due o più varianti (alleli, fenotipi, varianti di sequenza, varianti di struttura cromosomica) con frequenze significative.

Un locus è considerato polimorfico se presenta almeno **due alleli** dei quali il **più raro** ha una **frequenza maggiore dell' 1%**.

Sito a Sequenza Etichettata STS

- è una breve sequenza di DNA, **100-500bp**, facilmente riconoscibile e presente come **sequenza unica** nel cromosoma/genoma
- rappresentano un collegamento diretto tra mappe genetiche e mappe fisiche

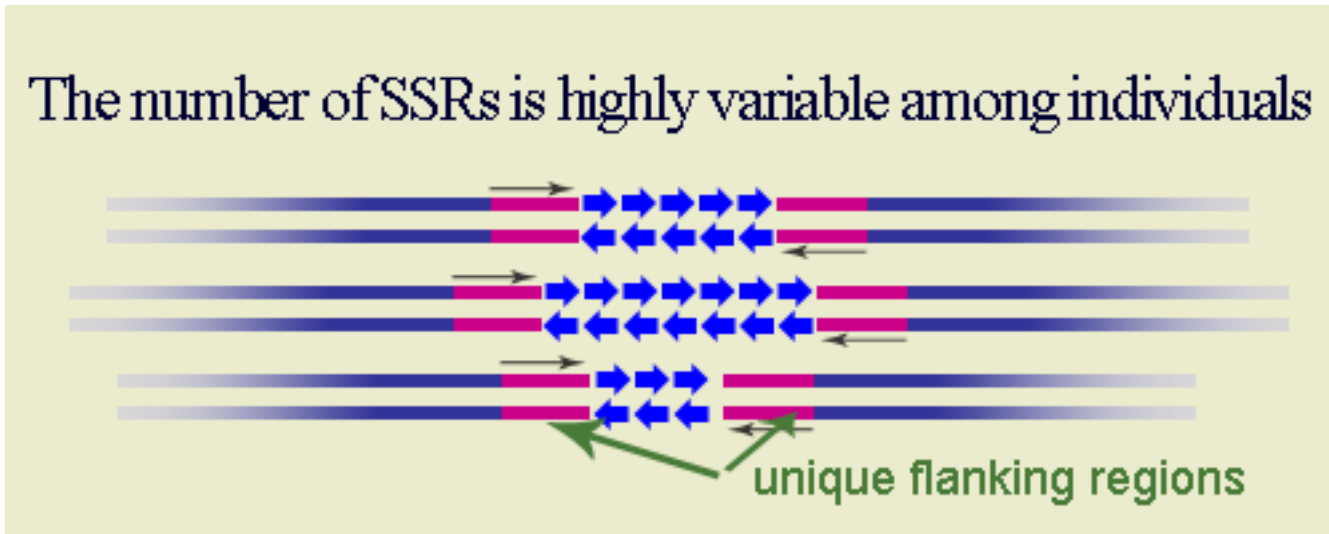
Una sequenza di DNA per essere considerata una STS deve rispondere a 2 caratteristiche:

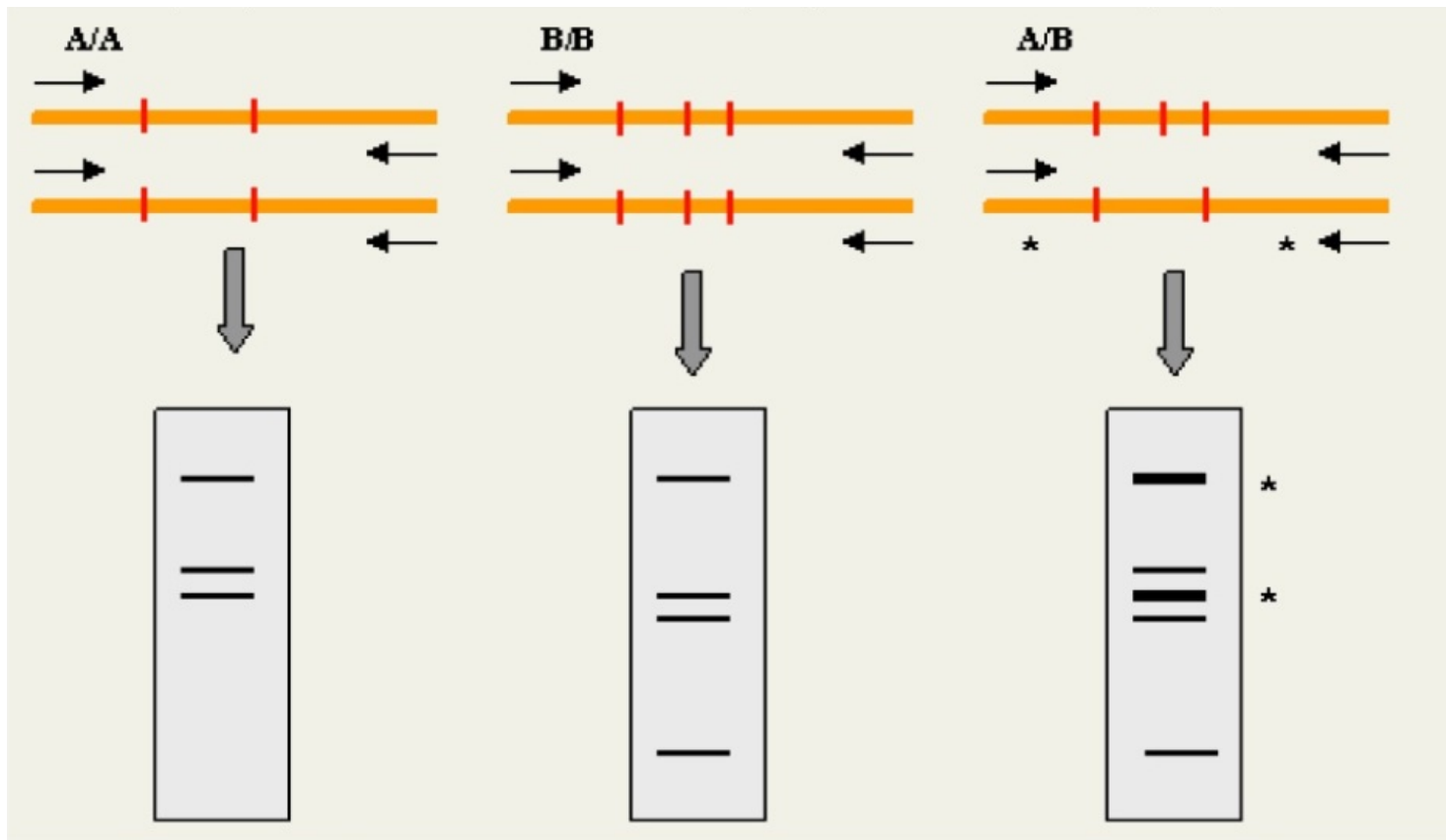
- ✓ **la sua sequenza deve essere nota**, così da verificarne la presenza nei frammenti di DNA mediante PCR
 - ✓ deve avere **localizzazione unica nel cromosoma/genoma** (single copy) in esame
- Le fonti più comuni di STS sono:
 - ✓ le **EST** (Expressed Sequence Tags), sequenze espresse corrispondenti a porzioni di cDNA di geni presenti in singola copia
 - ✓ gli **SSLP** (Simple Sequence Length Polymorphisms)
 - ✓ elementi inserzionali stabili (**trasposoni**)

Polymorphic loci chiamati anche Simple Sequence Repeats/Simple Sequence Repeats Polymorphisms (SSR/SSLP/SSRP). La sequenza del DNA di un STS può contenere elementi ripetitivi, sequenze che compaiono altrove nel genoma, ma fintanto che le sequenze ad entrambe le estremità del sito sono uniche e conservate, ogni STS rappresenta un marcatore.

Questi polimorfismi vengono identificati costruendo primer per PCR sulle *flanking sequences* della regione dei microsatelliti.

SSR/SSLP





Cleaved Amplified Polymorphic Sequences (CAPS)

i polimorfismi sono dovuti a differenze nelle lunghezze dei frammenti di restrizione causate da SNP o INDEL che creano o aboliscono i siti di riconoscimento delle endonucleasi di restrizione negli ampliconi PCR prodotti da primer oligonucleotidici locus-specifici.

How It Works

CAPS assay utilizza frammenti di DNA amplificati che vengono digeriti con un'endonucleasi di restrizione per visualizzare RFLP.

Polimorfismi della Lunghezza dei Frammenti di Restrizione (RFLP)

- Gli RFLP sono conseguenza di una mutazione puntiforme ma in un sito di restrizione
- Per ogni RFLP ci sono solo **2 possibili alleli**, con o senza sito di restrizione
- Vengono individuati sia mediante analisi Southern che mediante PCR

Polimorfismi della Lunghezza di Sequenze Semplici (SSLP)

- Gli SSLP sono insiemi di **sequenze ripetute** che presentano **variazioni di lunghezza**
- Per ogni SSLP ci sono **più possibili alleli**, a seconda di quante unità ripetute presentano
- esistono 2 tipi di SSLP:
 - ☺ **minisatelliti o Ripetizioni in Tandem a Numero Variabile (VNTR)**, in cui la lunghezza dell'unità ripetuta è di 25bp max. Principalmente localizzati nelle regioni telomeriche
 - ☺ **microsatelliti o Ripetizioni in Tandem Semplici (STR)**, le cui ripetizioni sono generalmente di unità di/tetranucleotidiche. Dispersi in tutto il genoma

SSLP

Simple Sequence Length Polymorphism

Numero variabile di ripetizioni in tandem

Gli alleli differiscono per il numero variabile delle unità ripetute in tandem

Ereditarietà mendeliana

Molti alleli

Elevata eterozigosità

MINISATELLITI

VNTR

Variable Number Tandem Repeat

Unità ripetuta: 20bp→centinaia bp

Dimensione alleli: poche centinaia →kb

Localizzazione subtelomerica

MICROSATELLITI

STR

Short Tandem Repeat

Unità ripetuta: 2-4 bp

Dimensione alleli: 70-400 bp

Ben distribuiti nel genoma

Tri- e tetra-nucleotidici

Multiplexing / automazione

Polimorfismi di Singoli Nucleotidi (SNP)

- **Gli SNP sono mutazioni puntiformi che però non causano RFLP**
- **Per ogni SNP in genere sono presenti solo 2 alleli**
- **Tra i vantaggi degli SNP ci sono sia la loro abbondanza (1.42 milioni nel genoma umano), che la loro rapida identificazione mediante ibridazione**
- **Tipizzati in grande scala con sistemi automatizzati (microchip)**

RFLP

Sito restrizione

2 alleli

E. Mendeliana

PCR restrizione

SNP

un nucleotide

2 alleli

E. mendeliana

seq.automatici

VNTR

n-ripetizioni

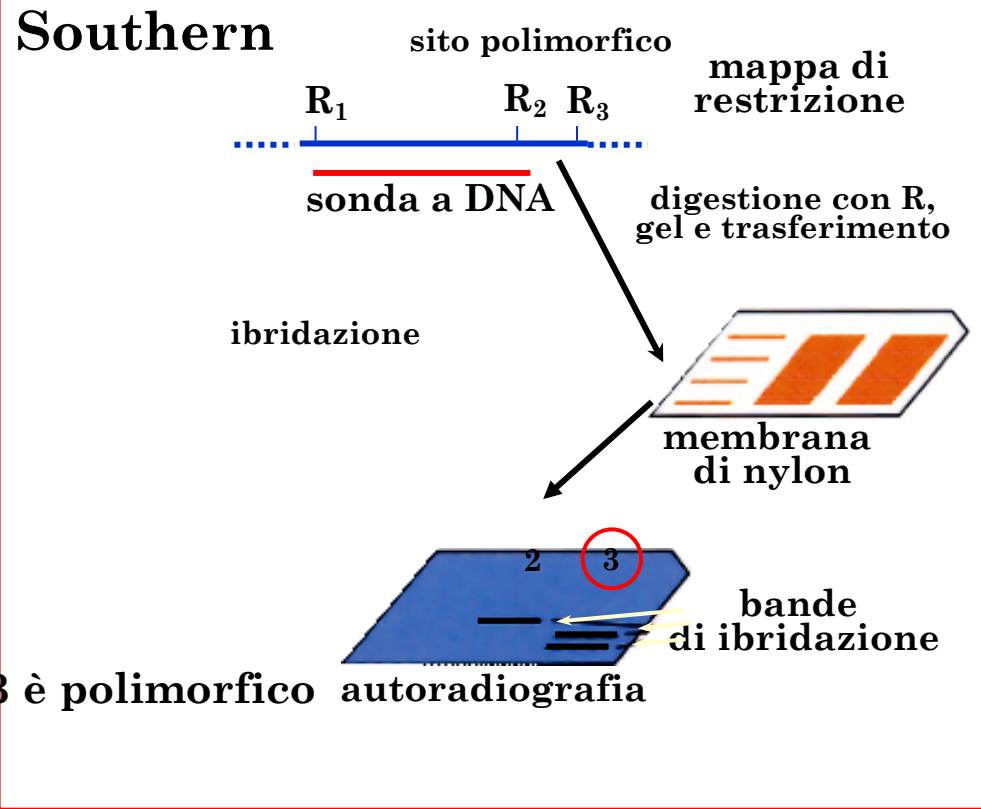
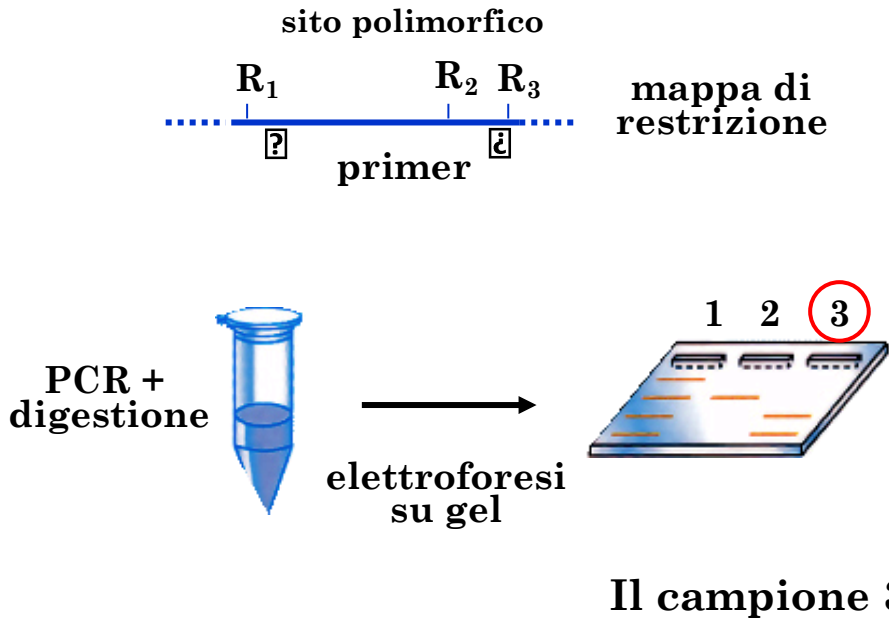
molti alleli

E. mendeliana

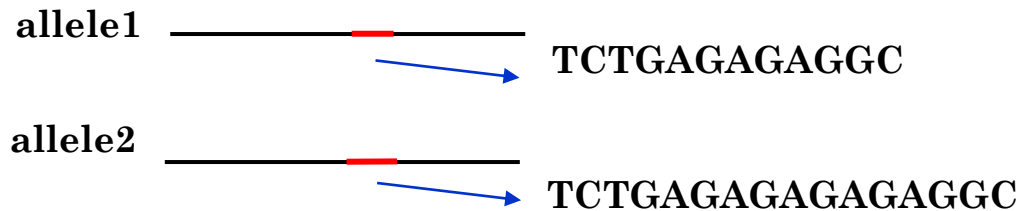
PCR microchip

RFLP

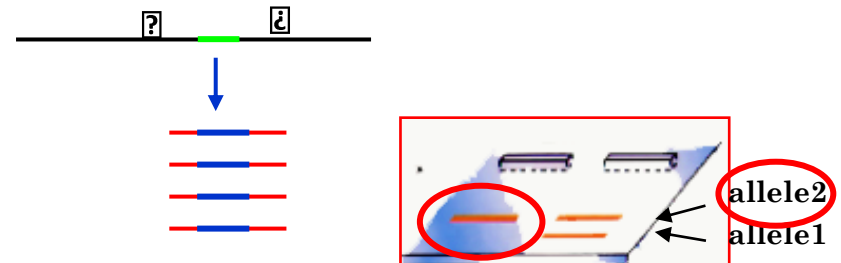
PCR



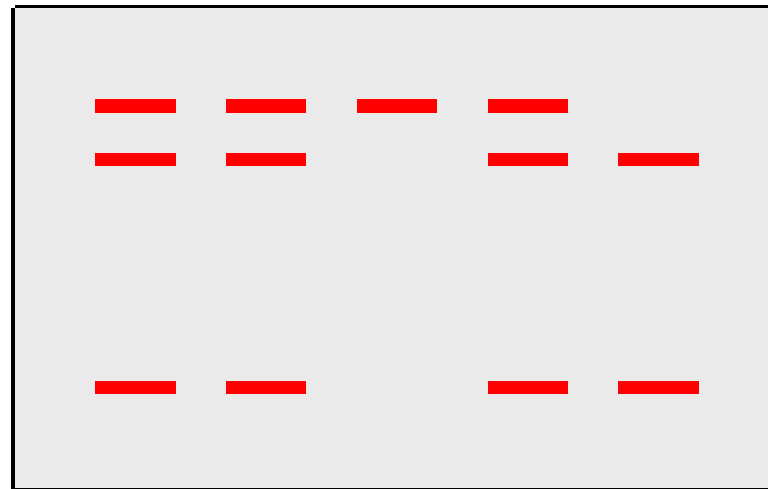
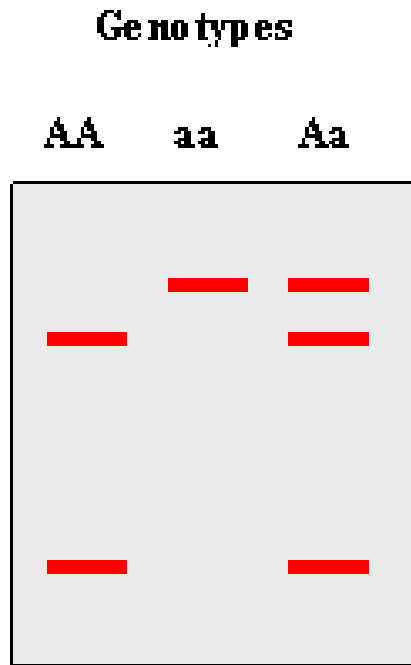
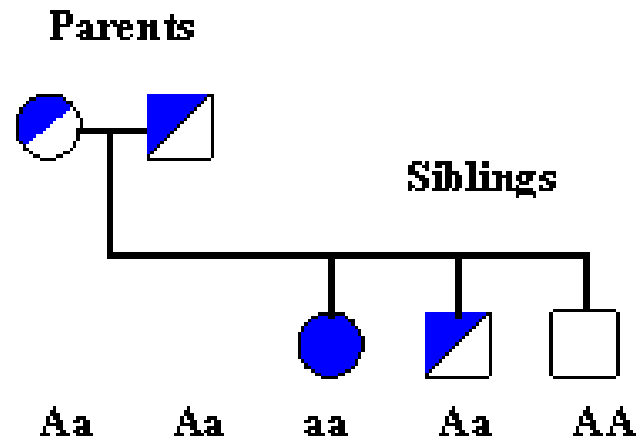
SSLP



Caratterizzazione mediante PCR



Inheritance of RFLP markers

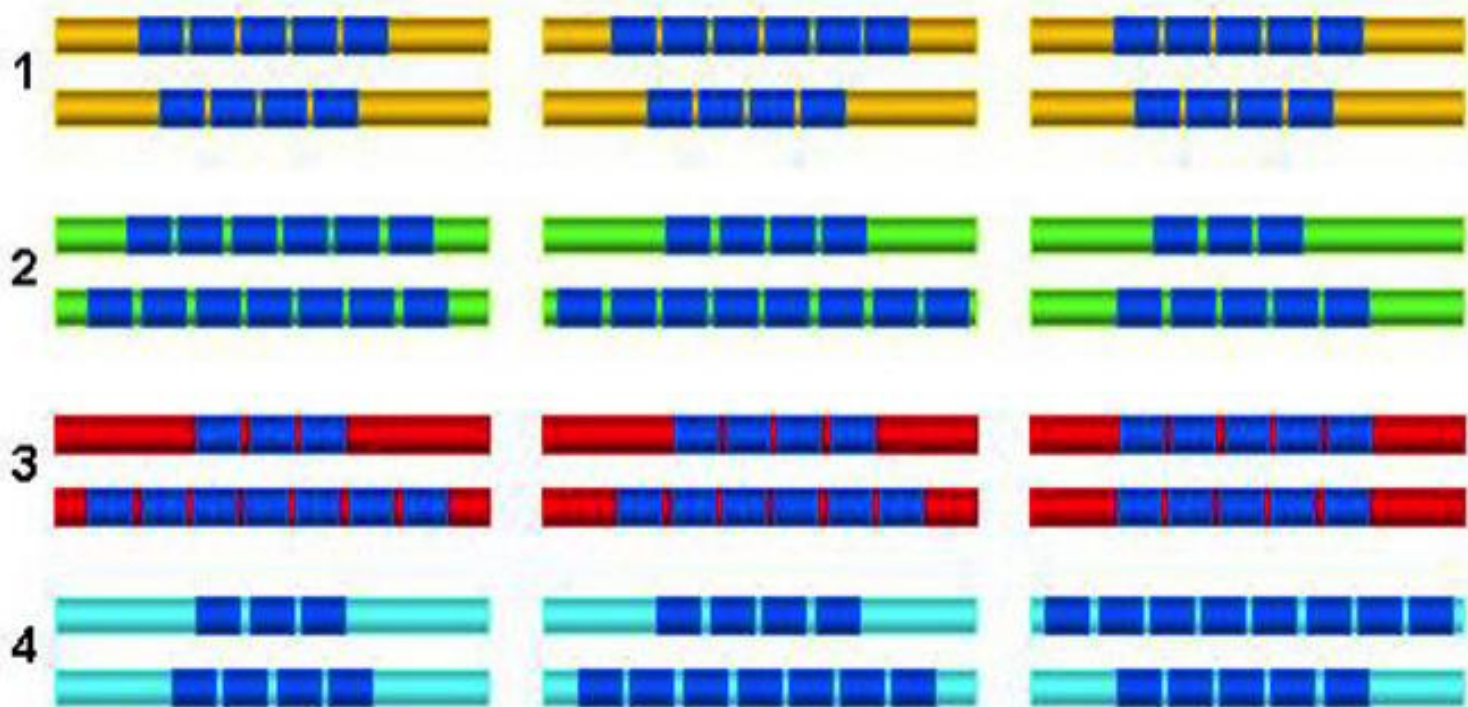


VNTRs: 4 pares de cromosomas homólogos en tres personas

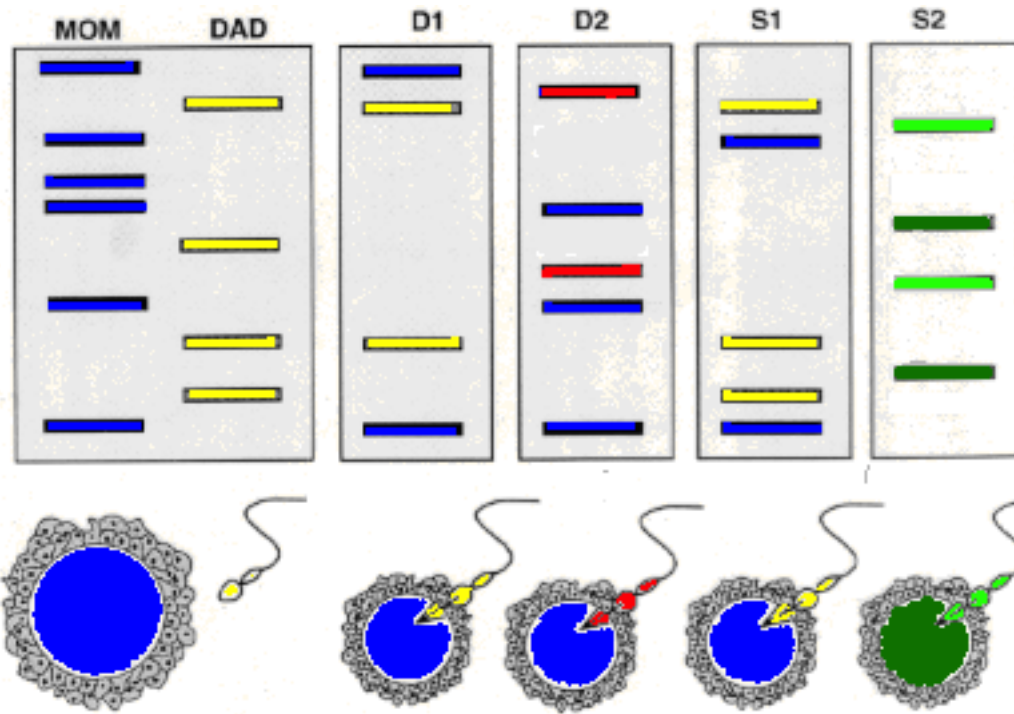
Individuo A

Individuo B

Individuo C



Analisi Paternità



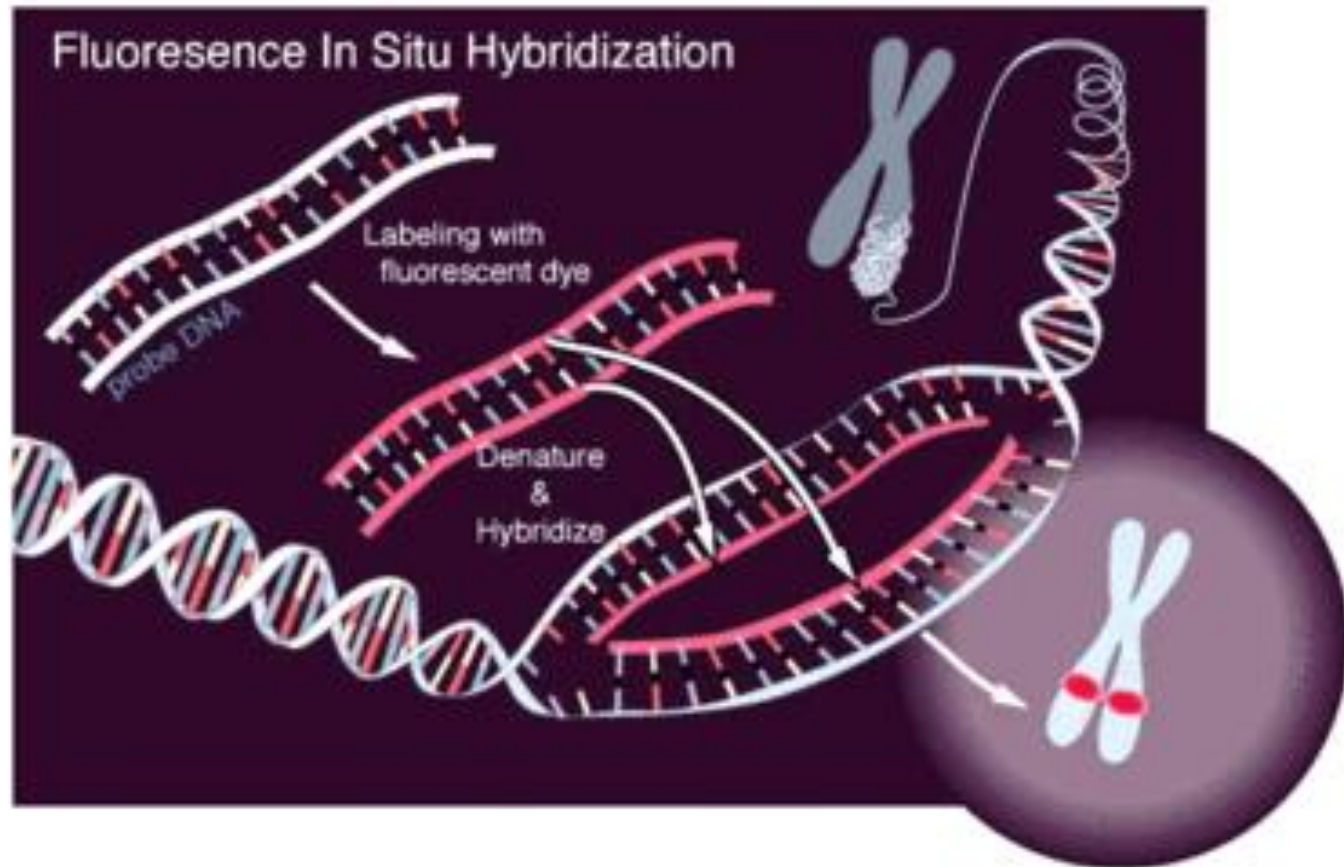
Ibridazione in situ con fluorescenza FISH

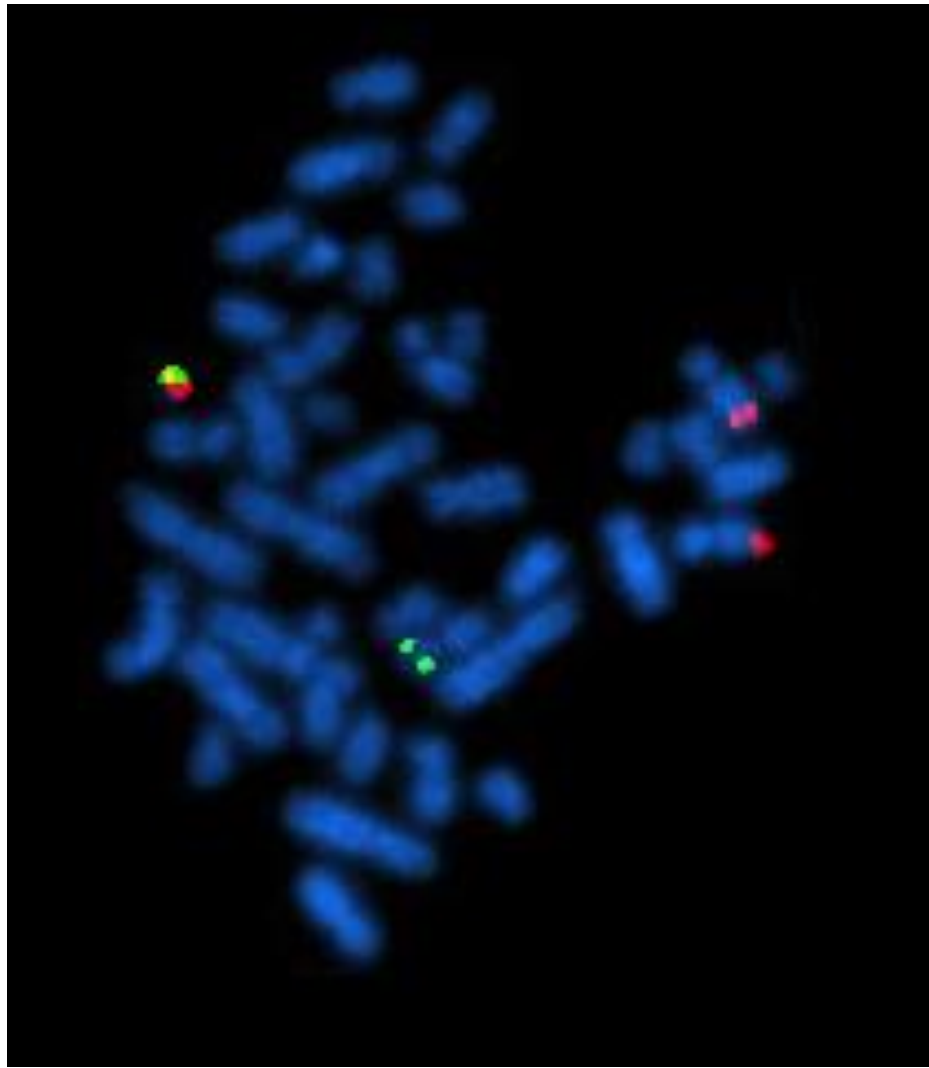
- consente di visualizzare **direttamente** la posizione di un marcatore all'interno di un cromosoma o su una molecola di DNA distesa
- il marcatore è una sequenza di DNA che viene osservata ibridando con una sonda fluorescente
- utilizza sonde fluorescenti che combinano un'alta sensibilità con un'alta risoluzione

Applicazioni della FISH

- **su cromosomi metafasici**, in cui il segnale è misurato rispetto all'estremità del braccio corto del cromosoma (valore FLpter). La risoluzione è pari ad **1Mb**
- **su cromosomi metafasici distesi meccanicamente**, mediante un passaggio di centrifugazione. La risoluzione è pari a **2/300Kb**
- **su cromosomi in interfase**, lo stadio del ciclo in cui i cromosomi sono meno condensati. La risoluzione è pari a **25Kb**

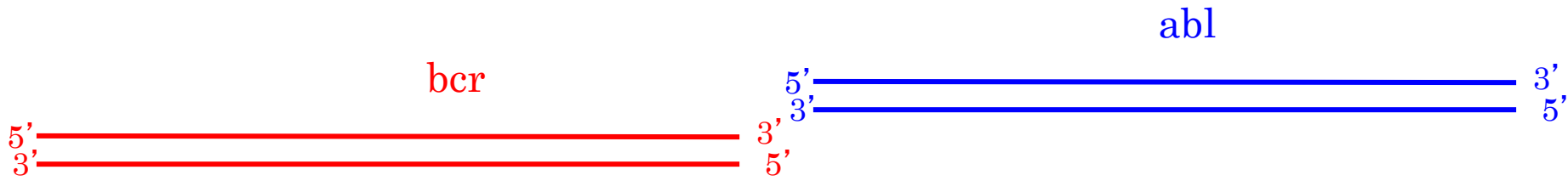
Fluorescence In Situ Hybridization





A metaphase cell positive for the bcr/abl rearrangement (associated with chronic myelogenous leukemia) using FISH. The chromosomes can be seen in blue. The chromosome that is labeled with green and red spots (upper left) is the one where the rearrangement is present.

Cellula normale



Traslocazione
tra i cromosomi

9 e 22

Cellula trasformata



Amplificazione
per PCR



L' amplificazione di una banda specifica da una coppia di primers **bcr/abl** consente di diagnosticare una forma di leucemia mieloide cronica dovuta a traslocazione tra cromosomi