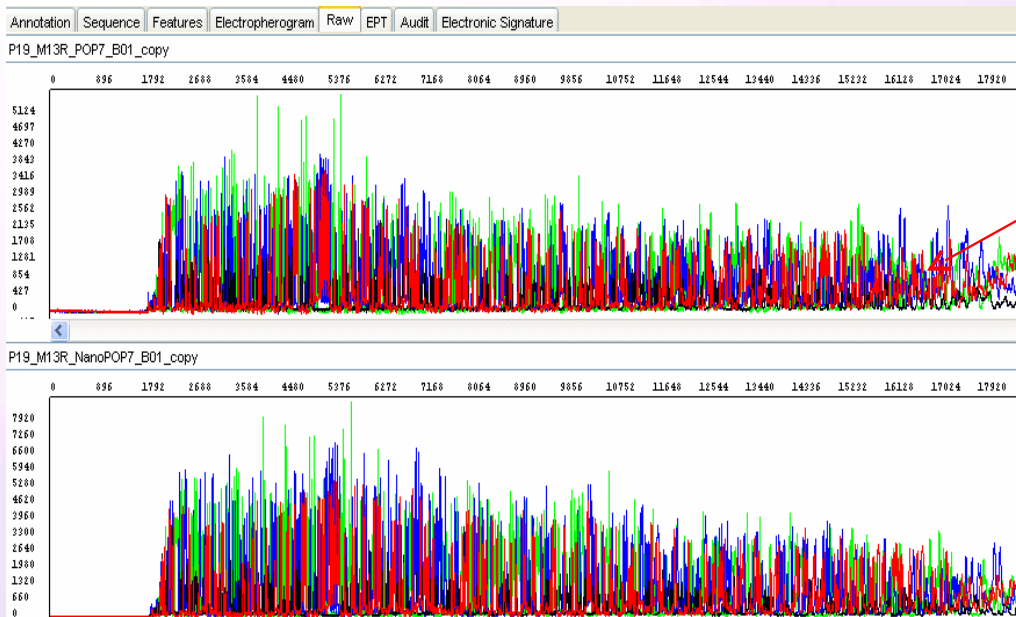


# Metodi di sequenziamento del DNA

**Maurizio Zini**



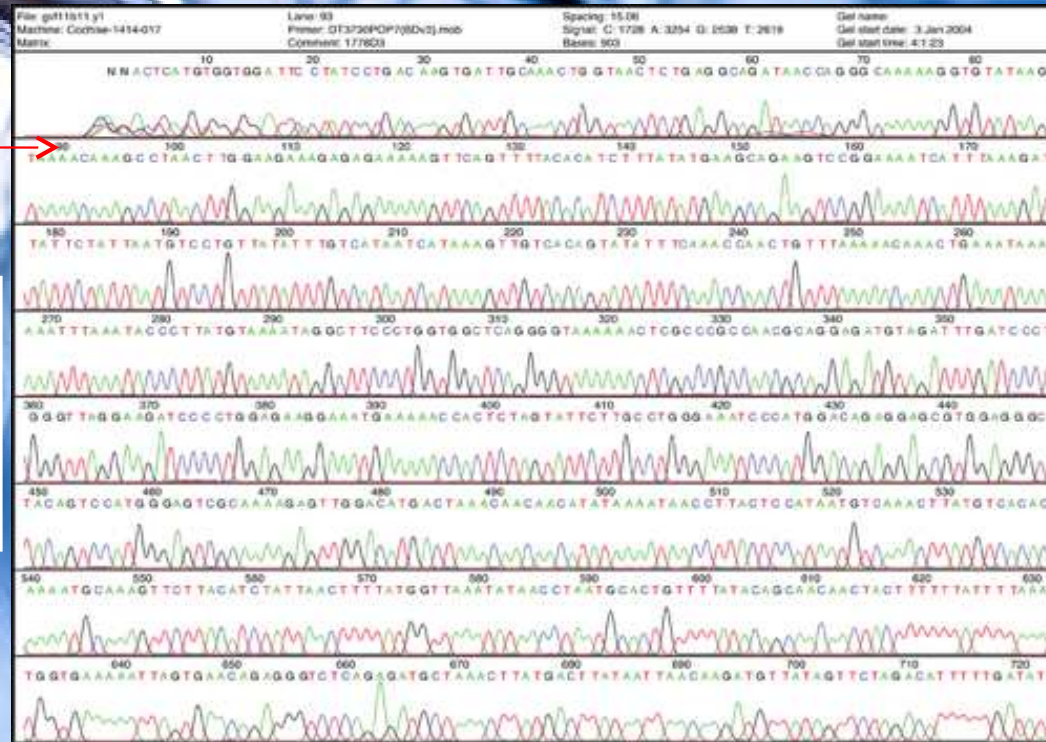


I dati grezzi registrati del sequenziatore

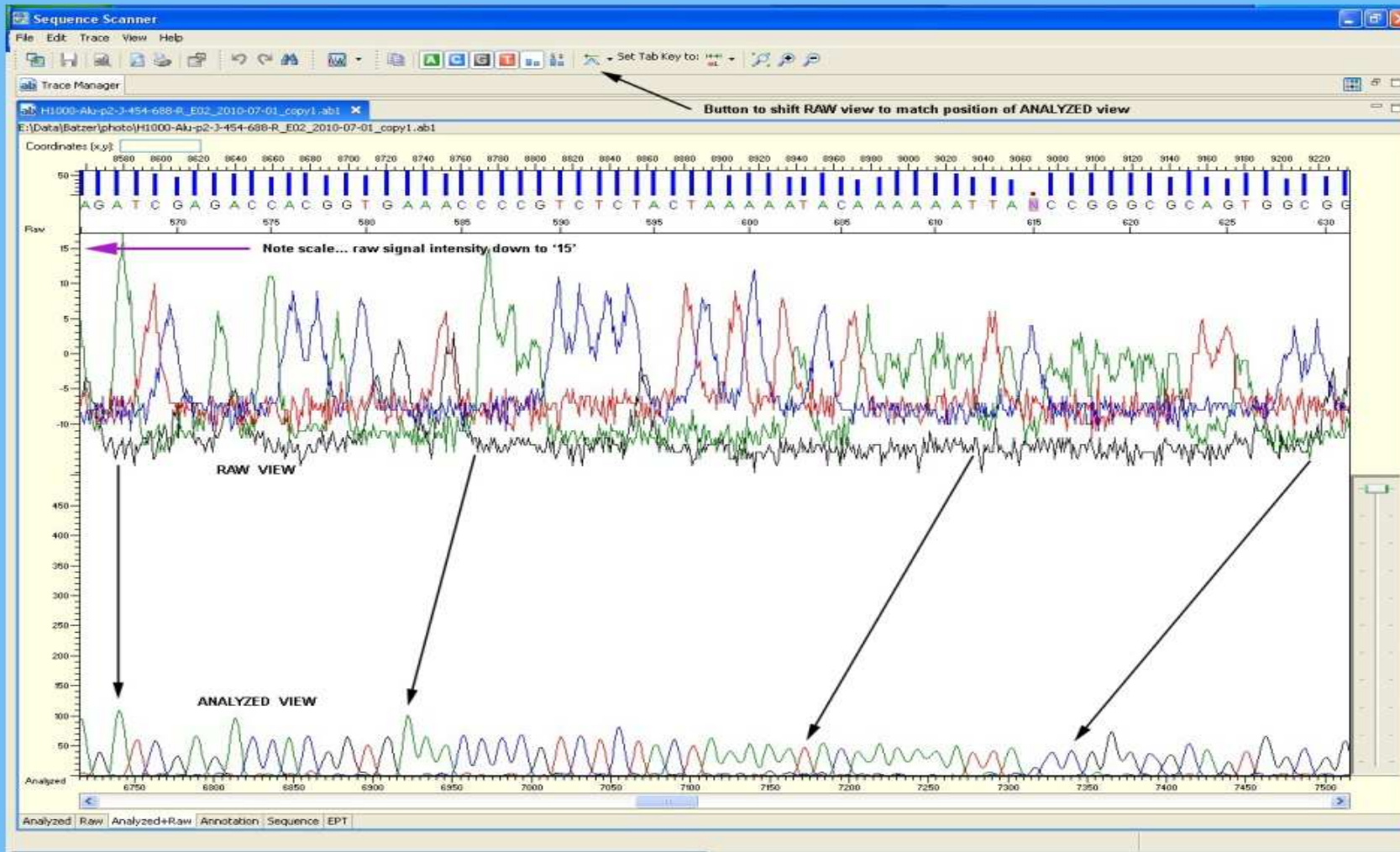
Un software di analisi definito "Base Caller" elabora i dati registrati dal sequenziatore e li rende in forma leggibile.

Il dato leggibile :  
Elettroferogramma

Presenta in formato sia testuale che grafico i dati della sequenza assegnando anche un valore QV ad ogni chiamata.

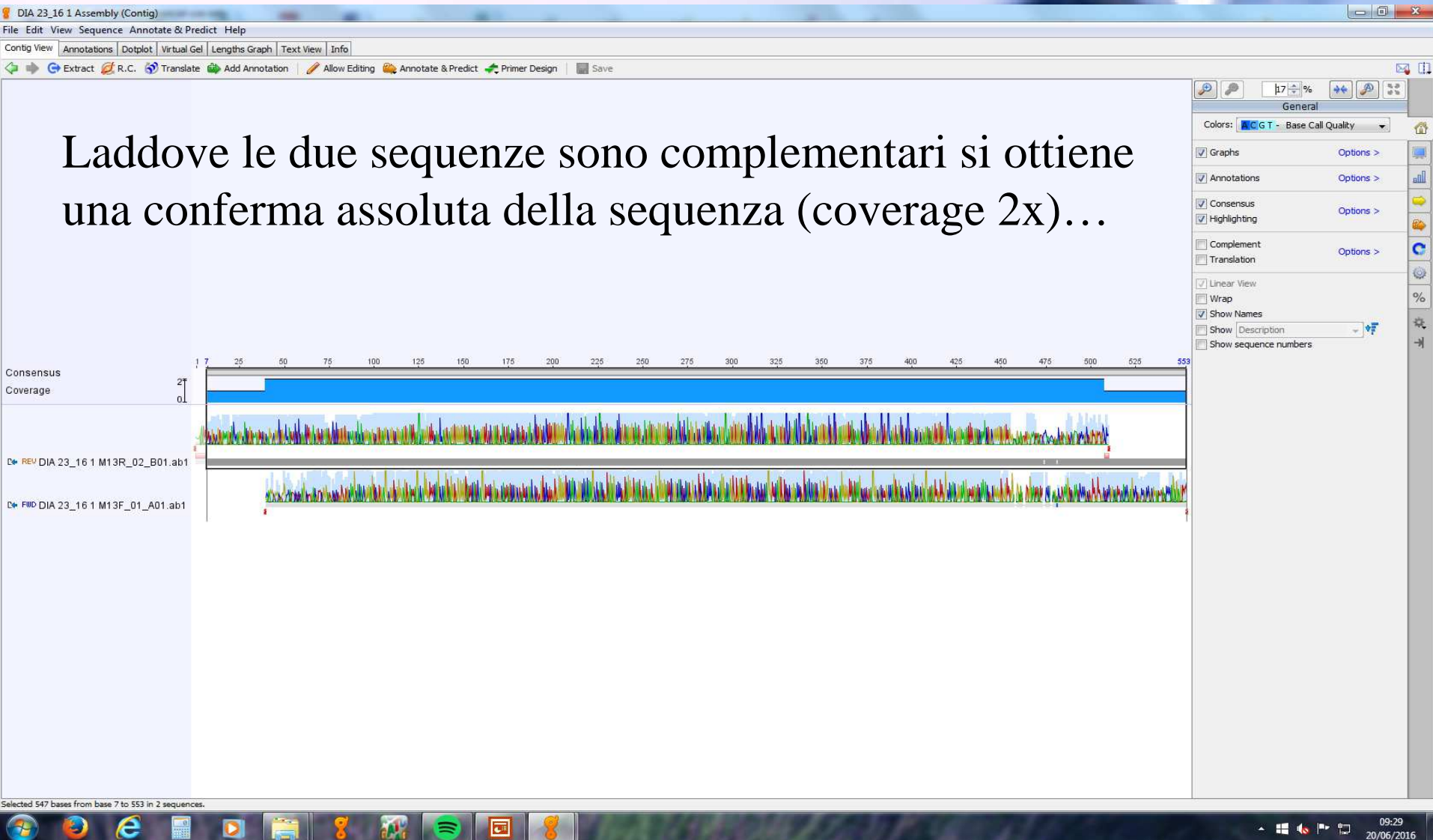


Il QV assegna un valore di attendibilità mostrato da colori e ampiezza delle colonne ad ogni singola chiamata e lascia all'utente l'interpretazione del risultato.



# Allineamento delle due sequenze ottenute dai due filamenti di DNA complementari

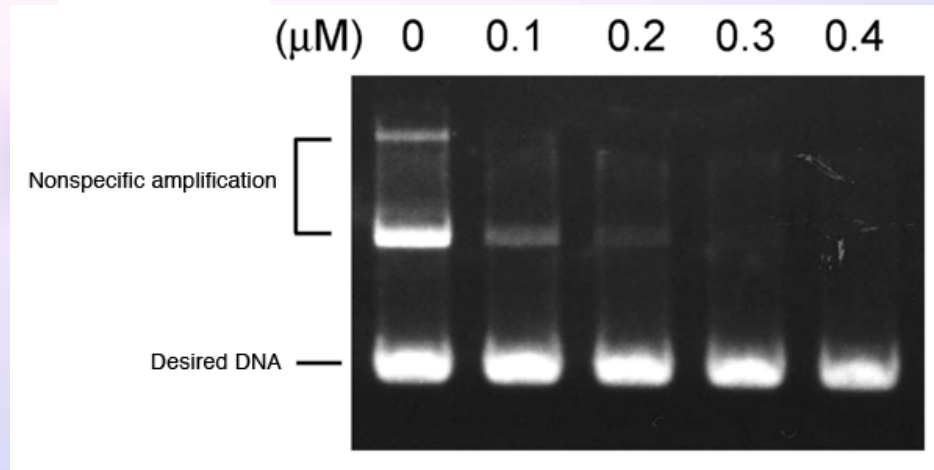
Laddove le due sequenze sono complementari si ottiene una conferma assoluta della sequenza (coverage 2x)...





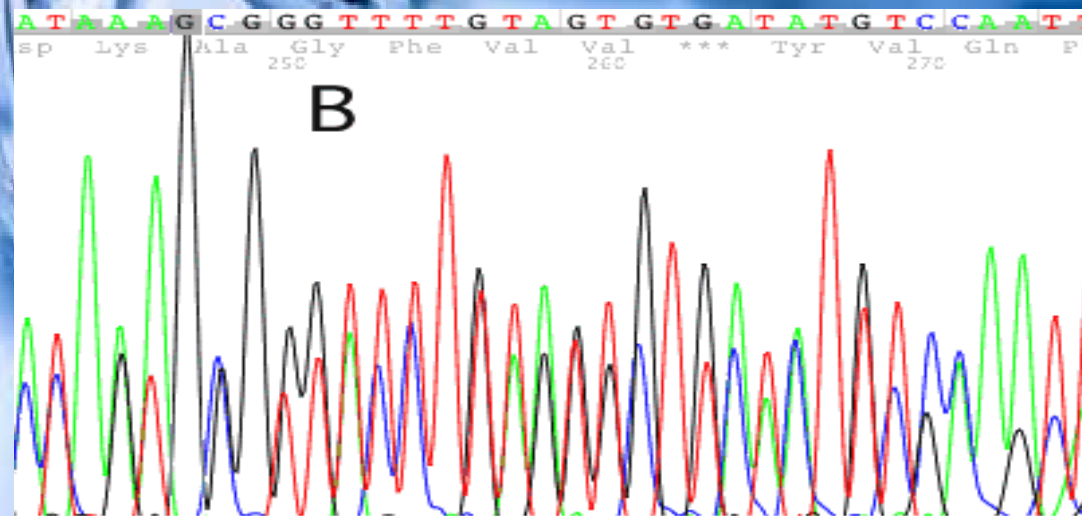
# I requisiti più importanti del campione per un buon sequenziamento

## 1 Amplificato puro (no aspecifici di PCR)



In alcuni casi si può purificare da tagliando la banda da gel.

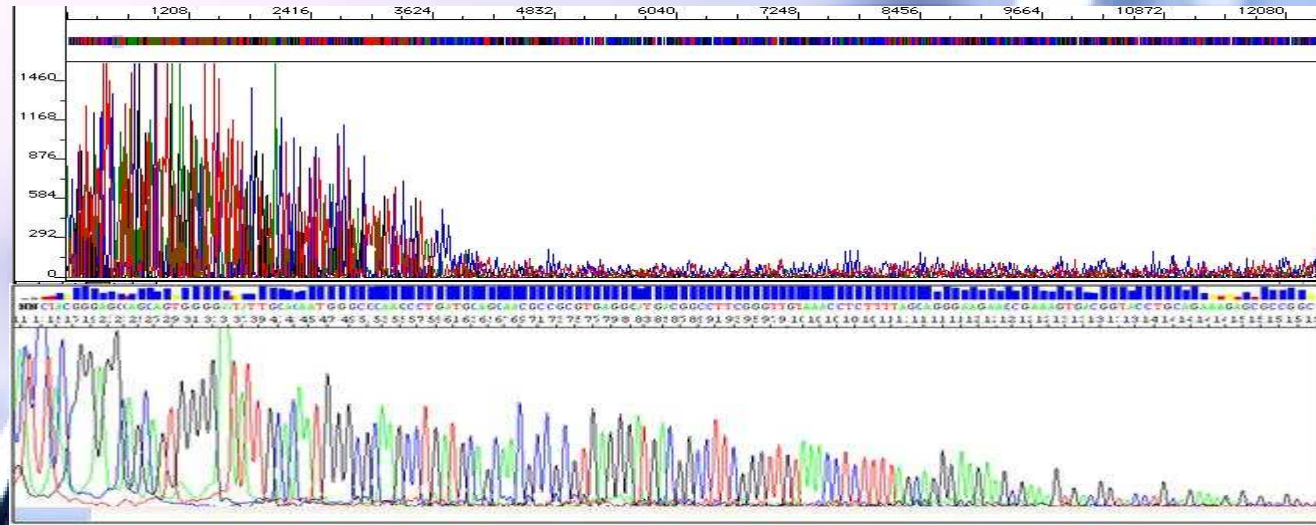
Altrimenti nel tubo di reazione avvengono due sequenziamenti contemporaneamente e non si può assegnare la chiamata all'uno o all'altro amplificato. Sequenza illeggibile.



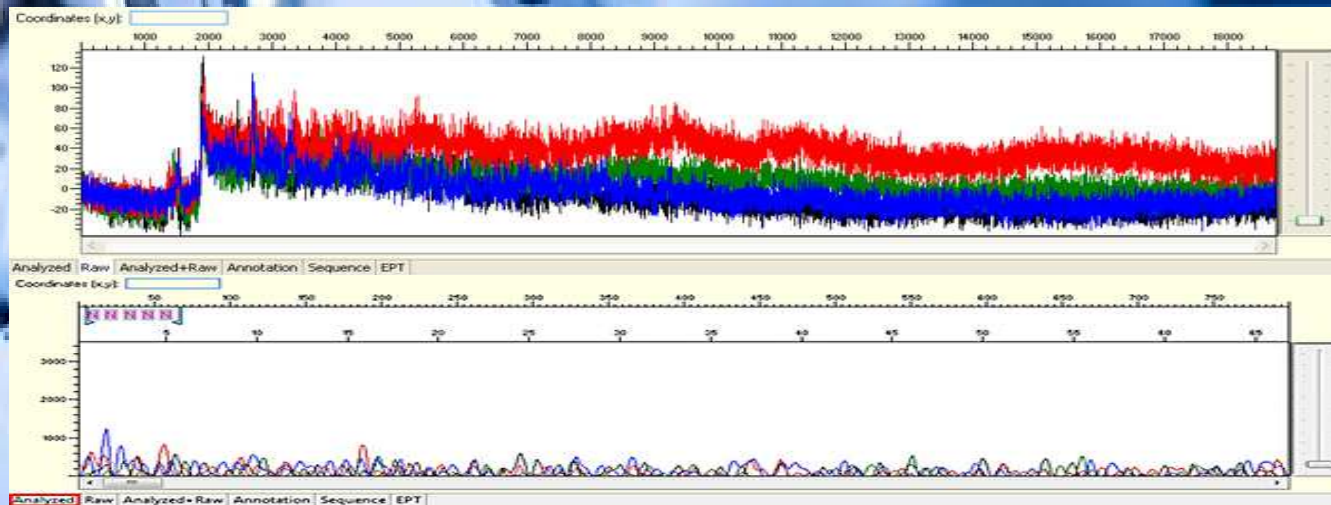
# I requisiti più importanti del campione

## 2 Buona quantificazione dell'amplificato

*Troppo DNA*  
stampo consumo dei reagenti (primers e nucleotidi) nei primi cicli di reazione... la sequenza è troncata.



*Poco DNA stampo...*



# Sequenziamento automatico con metodo Sanger

- ♠ Estrazione di DNA (dipende dalla matrice).
- ♠ PCR
- ♠ Purificazione del campione di DNA (per eliminare Sali, nucleotidi liberi, ma soprattutto residui dei primers usati per l'amplificazione), a volte è necessario purificare la banda di DNA dal gel di agarosio.
- ♠ Quantificazione (la quantità di DNA da usare per il Cycle seq. è un fattore critico).
- ♠ Cycle Sequencing (2 sequenziamenti... per i due filamenti di DNA del campione per ottenere una sequenza "consensus").
- ♠ Purificazione dei prodotti di cycle sequencing (per eliminare i sali ma soprattutto i fluorocromi "liberi").
- ♠ Elettroforesi capillare.
- ♠ Analisi dei dati (software e banche dati).

Ad oggi sono disponibili diversi metodi di sequenziamento del DNA ed ognuno di questi ha le sue peculiarità...

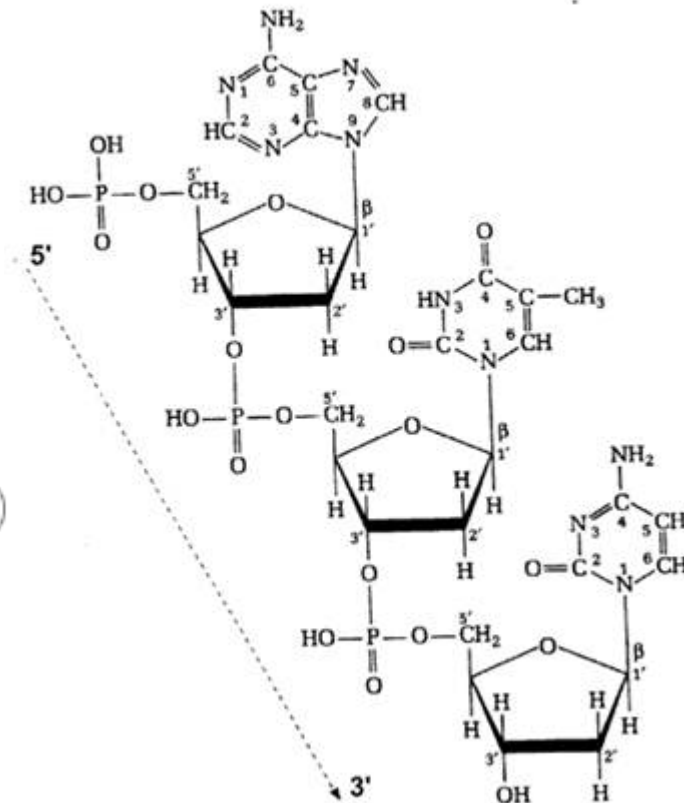
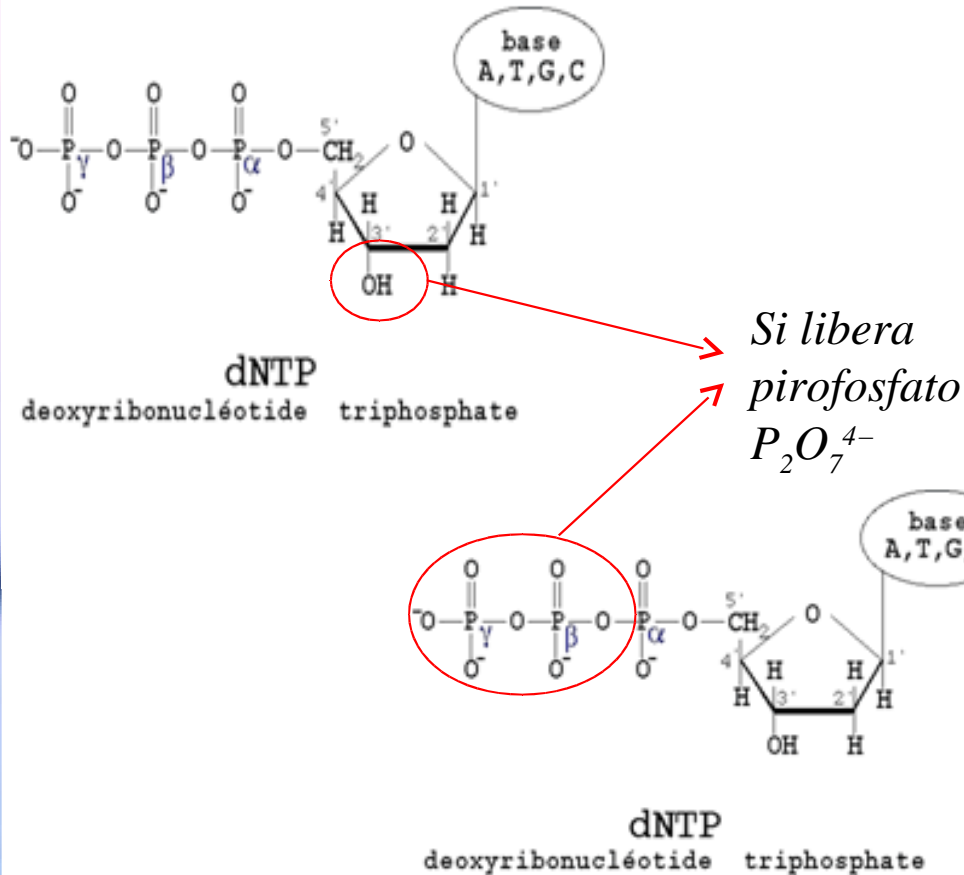
- Sequenziamento DNA con metodo Sanger su capillare (max 900 nucleotidi)
- **Pirosequenziamento (max 80 nucleotidi)**
- Sequenziamenti «Next Generation Sequencing» (fino a 600 Gb)





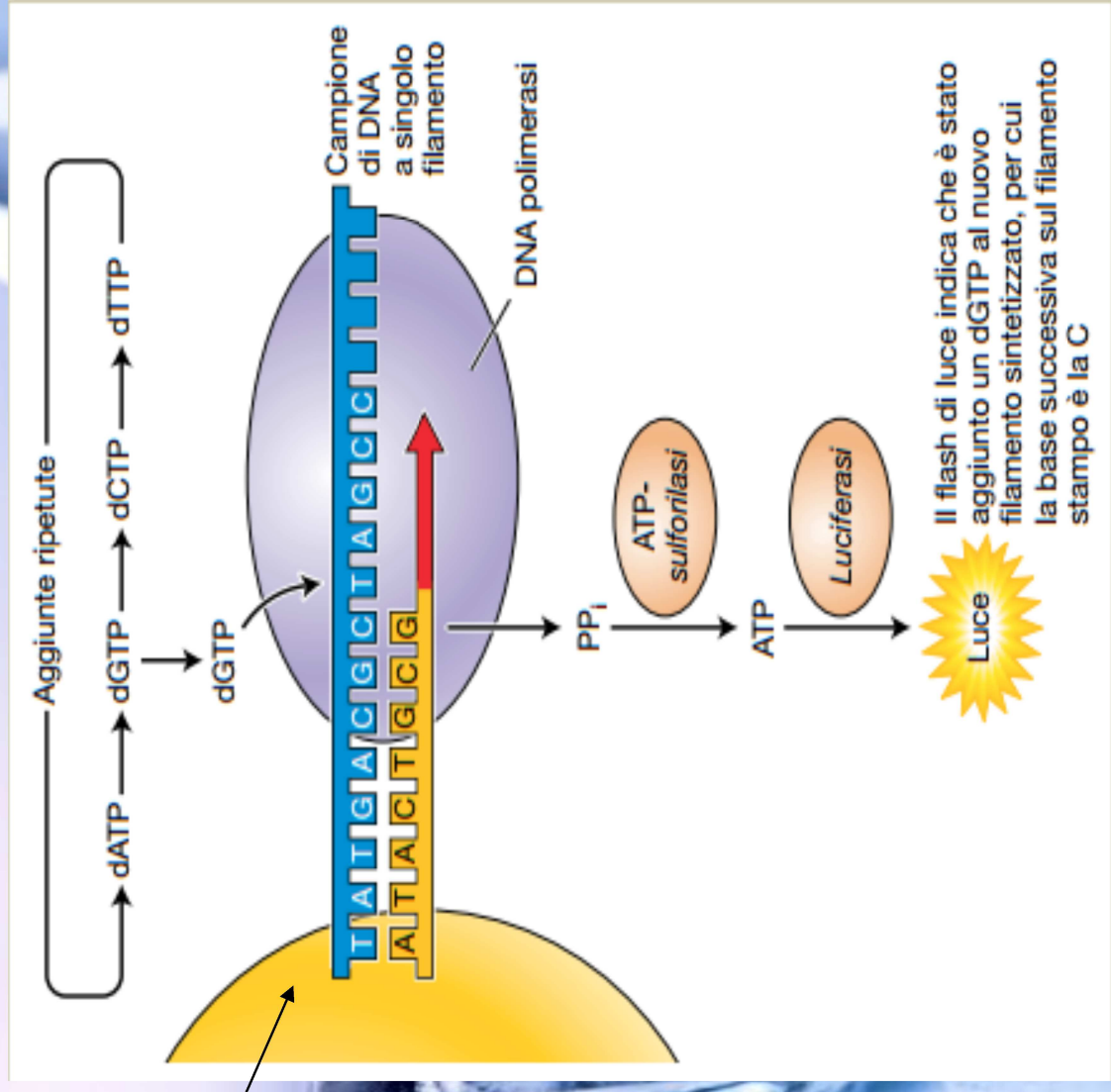
# Pirosequenziamento: un'idea per sequenziare DNA senza elettroforesi su gel

Sfrutta la chimica della reazione della DNA polimerasi per evidenziare l'incorporazione di un nucleotide attraverso il prodotto *pirofosfato*.



## Pirosequenziamento

Il filamento DNA stampo è immobilizzato su una resina attraverso un legame streptavidina-biotina...

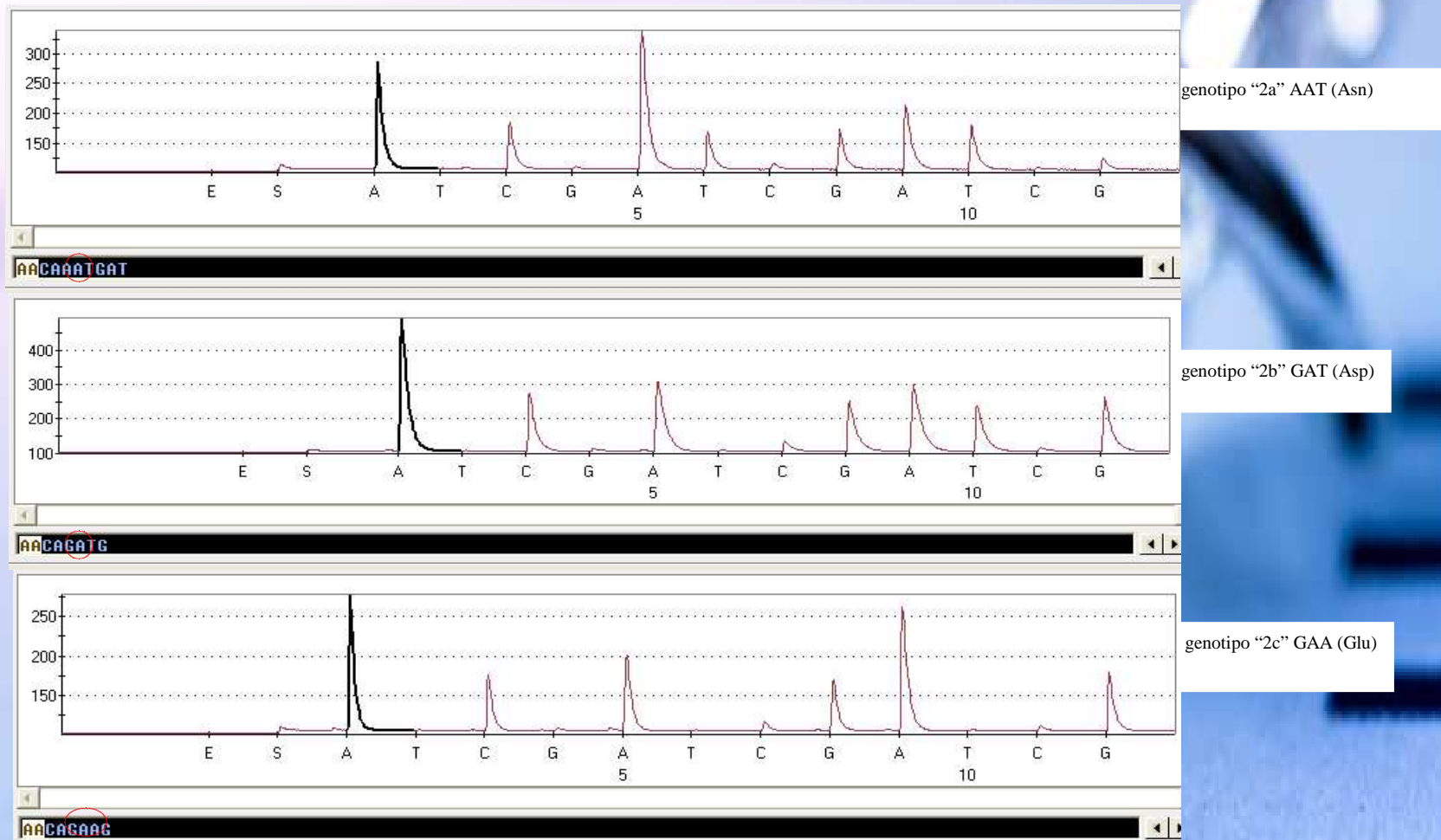


## Pirosequenziamento

- Ad ogni incorporazione di un nucleotide si libera un pirofosfato ( $P_2O_7^{4-}$ ).
- Il pirofosfato è coinvolto in una cascata enzimatica che porta all'emissione di luce.
- In presenza di un primer e di DNA polimerasi (che incorpora i nucleotidi usando come stampo il filamento complementare), è possibile somministrare al campione un solo nucleotide per volta separatamente (ACGT) e leggere quindi l'emissione di luce in questi quattro momenti diversi.
- Si determina così l'esatta sequenza del filamento in estensione (pirogramma).
- Nel caso di nucleotidi ripetuti (ad es. AATTT...) l'emissione di luce sarà pressochè proporzionale al n di pirofosfati liberati (quindi emissione di luce tripla se si incorporano tre T).
- Il meccanismo viene ripetuto fino alla determinazione della sequenza necessaria (max 80 nucleotidi...)



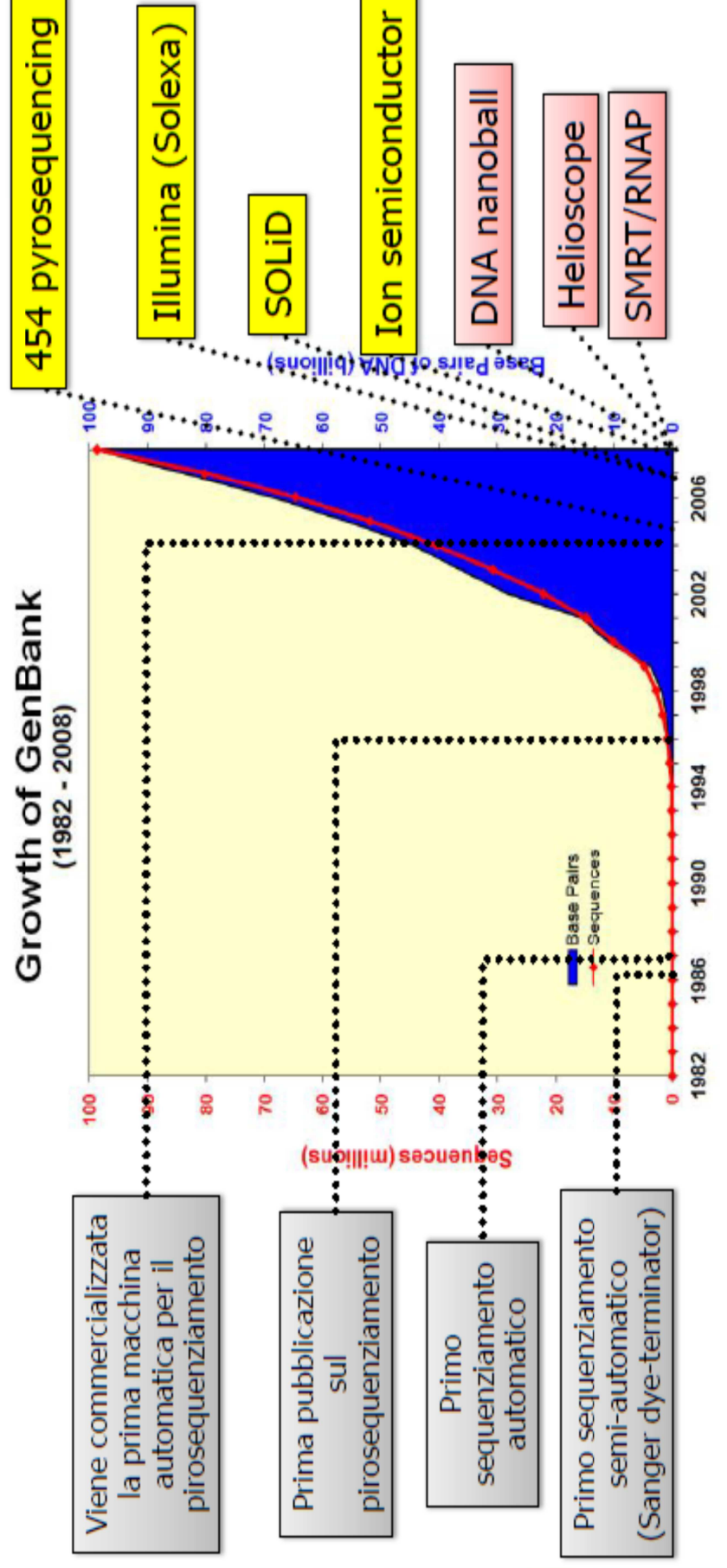
## Pirogramma Genotipizzazione CPV attraverso sequenza di una tripletta



**Figura 2:** dall'alto in basso sono riportati i profili di sequenza ottenuti per i tre genotipi. La sequenza di dispensazione (3xATCG) è riportata subito sotto ciascun pirogramma mentre la sequenza nucleotidica rilevata è evidenziata in nero.



L'avvento dell'NGS ha portato all'aumento esponenziale delle sequenze depositate....



Tuttavia ancora oggi il sequenziamento su capillare è ritenuto il «gold standard» tra tutte queste ed è il maggiormente diffuso.

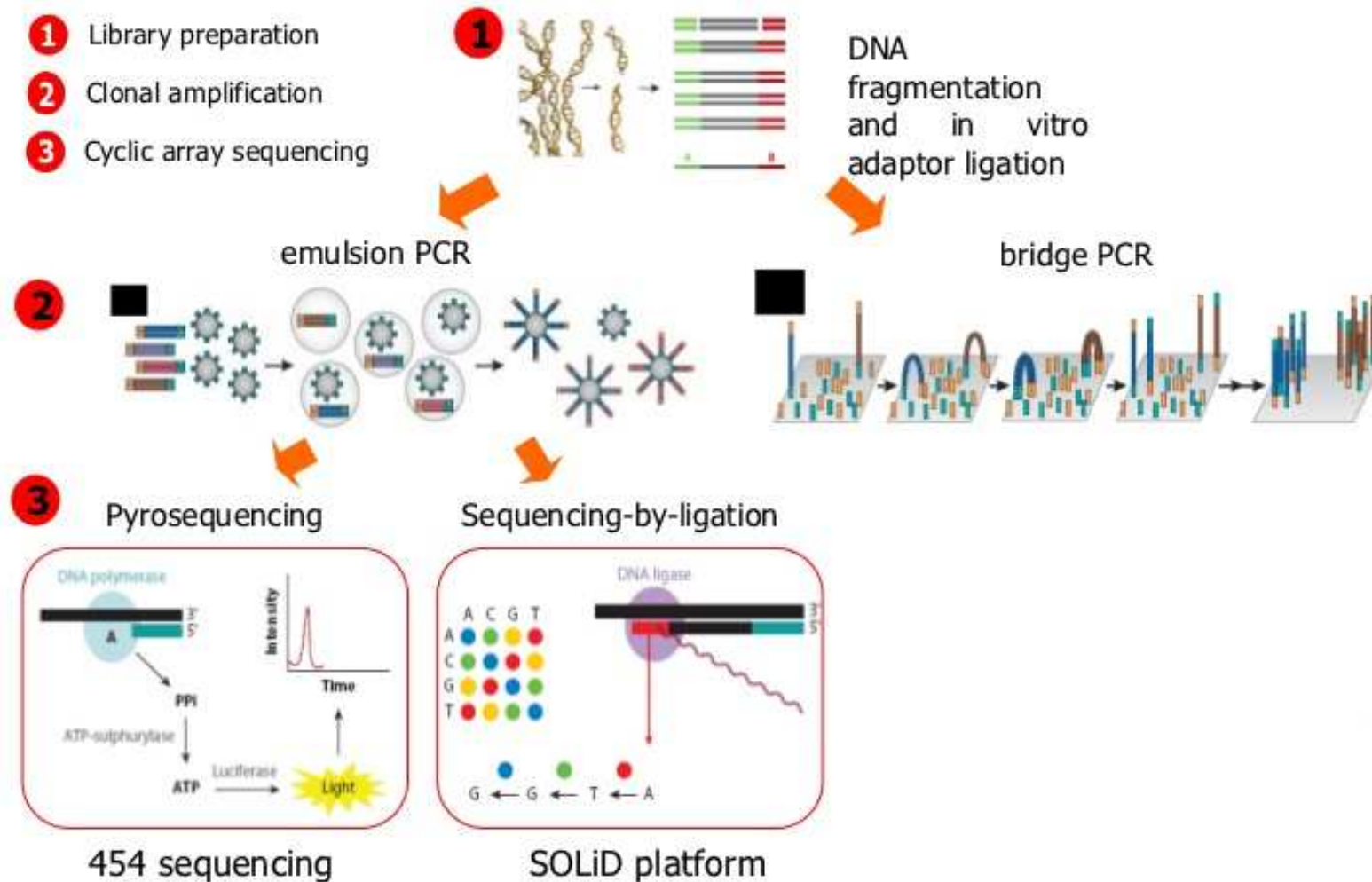
# Confronto dei metodi di Sequenziamento

Table 2: Comparison of different sequencing technologies

Parameters	Sequencing Technologies	Ion Torrent (Ion semi-conductor)	Roche 454 (Pyrosequencing)	Illumina (Sequencing by synthesis)	SOLiD (Sequencing by ligation)	Pacific Bio (Single molecule real-time sequencing)	Helicos (True Single Molecule Sequencing)	Sanger method (Chain termination)
Sequencing chemistry		Detection of released H <sup>+</sup>	Pyrosequencing	Reversible terminators	Ligation	Fluorescently labelled dNTPs	Reversible terminators	Di-deoxy Chain termination
Adapter used		Adapters	Adapters	Adapters	Adapters	Hairpin adapters	Poly(A) adapter	N/A
Amplification method		Emulsion PCR	Emulsion PCR	Bridge amplification in situ	Emulsion PCR	Linear amplification	No amplification	Sequencing PCR
Separation method		Ion Spheres and high density array	Microbeads and 'picotitre' plate	Glass slide hybridization	Beads on glass slide	Captured by DNA polymerase in microcell	Flow-cell hybridization	Electrophoresis
Read length		200-400 bp	700 bp	50 to 250 bp	50-75 bp	1000 bp	25 bp	400 to 900 bp
Reads per run		up to 5 million	1 million	up to 3 billion	1.2 to 1.4 billion	35-75 thousand	1 billion	Not available
Maximum data output per run		1 Gb	700 Mb	600 Gb	20 Gb	Not available	35 Gb	Not available
Accuracy		98%	99.9%	96%	99.9%	99%	99%	99.9%
Time per run		2 hours	24 hours	1 to 10 days	1 to 2 weeks	30 minutes to 2 hours	5-10 days	20 minutes to 3 hours
Cost per 1 million bases (in US\$)		\$1	\$10	\$0.05 to \$0.15	\$0.13	\$2	Not available	\$2400
Advantages		<ul style="list-style-type: none"> <li>• Equipment relatively less expensive</li> <li>• Fast reaction</li> </ul>	<ul style="list-style-type: none"> <li>• Long read size</li> <li>• Fast reaction</li> </ul>	<ul style="list-style-type: none"> <li>• High sequence yield</li> </ul>	<ul style="list-style-type: none"> <li>• Low cost per base of sequencing</li> </ul>	<ul style="list-style-type: none"> <li>• Longest read length</li> <li>• Less time consuming</li> </ul>	<ul style="list-style-type: none"> <li>• No PCR induced bias and errors</li> <li>• Tolerates degraded samples</li> </ul>	<ul style="list-style-type: none"> <li>• Long individual reads.</li> <li>• Applied in many sequence based research</li> </ul>
Disadvantages		<ul style="list-style-type: none"> <li>• Homopolymer error</li> </ul>	<ul style="list-style-type: none"> <li>• Homopolymer error</li> <li>• Runs relatively expensive</li> </ul>	<ul style="list-style-type: none"> <li>• Needs high DNA concentration</li> <li>• Very expensive equipment.</li> </ul>	<ul style="list-style-type: none"> <li>• Slower than other sequencing methods.</li> </ul>	<ul style="list-style-type: none"> <li>• Low yield at high accuracy</li> <li>• Equipment very expensive.</li> </ul>	<ul style="list-style-type: none"> <li>• Time to sequence a single nucleotide is high</li> <li>• High error rate</li> </ul>	<ul style="list-style-type: none"> <li>• Higher cost per base of sequencing</li> <li>• Impractical for whole genome sequencing projects</li> </ul>

## Metodi sequenziamento generazione 2

### Next-generation DNA sequencing



## Metodi sequenziamento generazione 2

NGS con pre-amplificazione (Pyrosequencing 454, Illumina, Solid, Ion Torrent...)

Principi di base comuni

1. Frammentazione del DNA in milioni di sequenze corte,
2. Immobilizzazione di questi frammenti su sfere o su chip.
3. Amplificazione per preparare la library da sequenziare (PCR emulsione, Bridge PCR)
4. Ottenimento di milioni di piccole sequenze lette singolarmente con sistemi chimico-fisici diversi, che saranno successivamente assemblate e "ripulite" con software e conoscenze adeguati.



## Bioinformatica

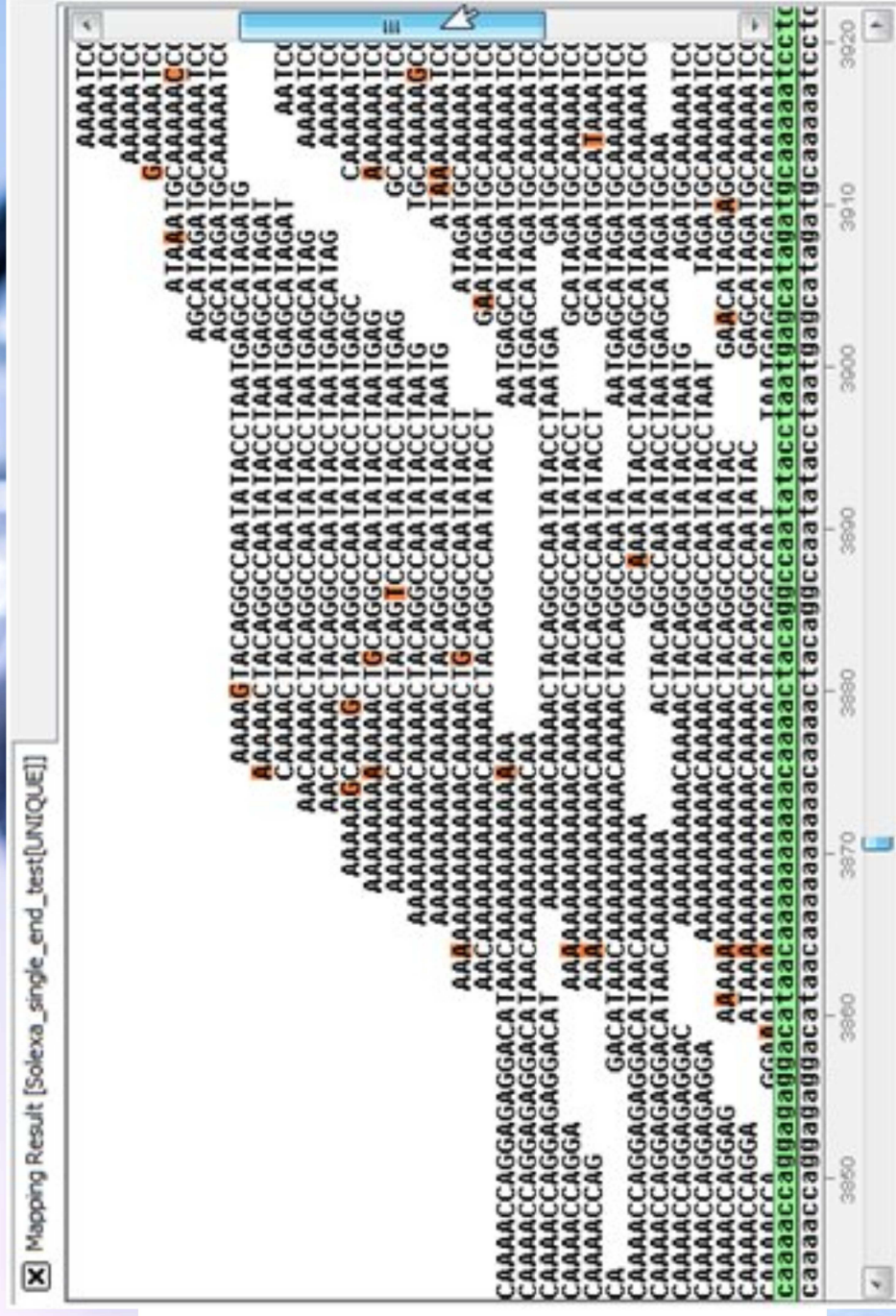
La parte più complicata dei sequenziamenti NGS  
è l'allineamento di milioni di piccole sequenze per creare contig.

--ACCGT

---CGTGC

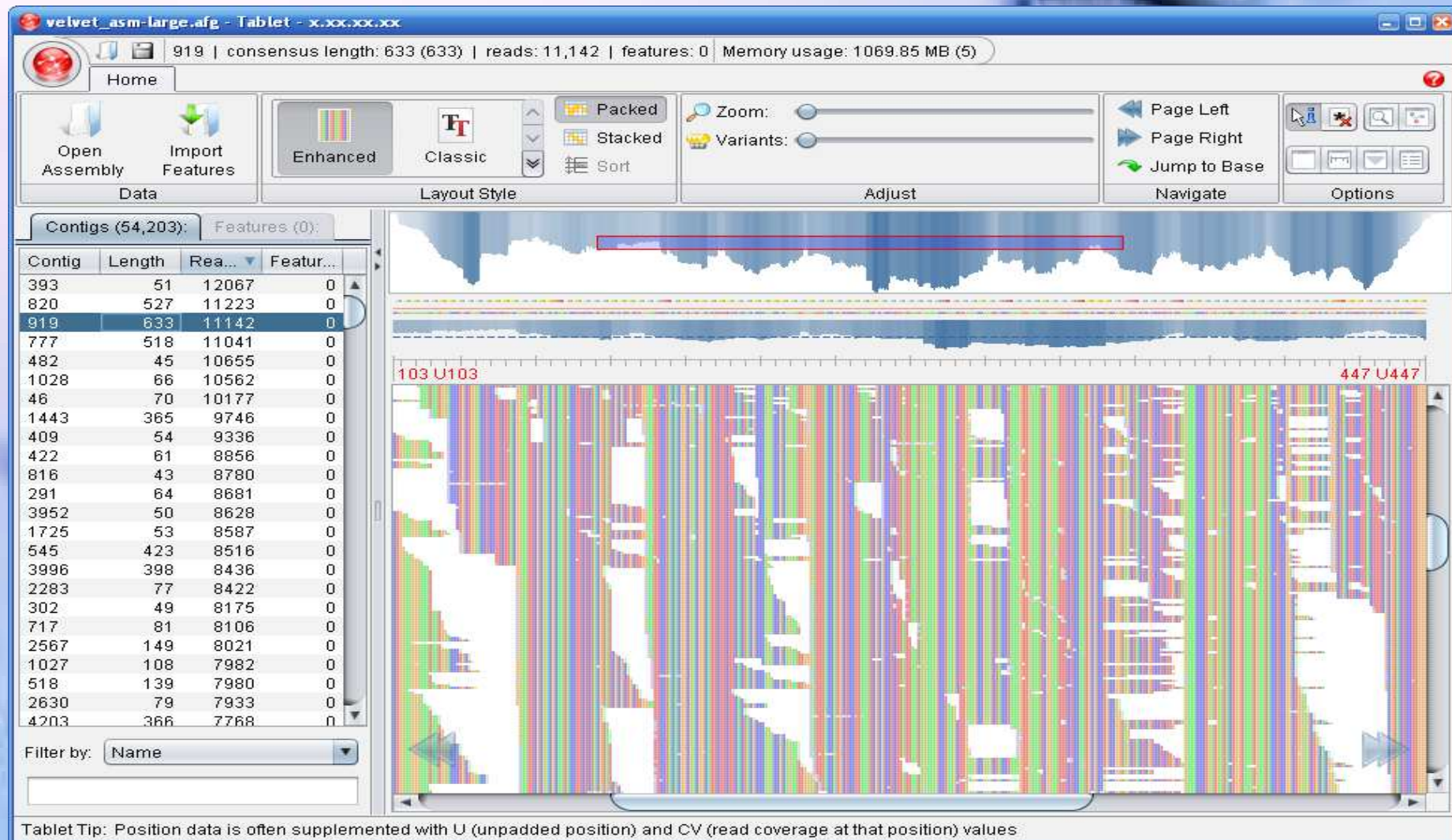
TTAC

TTACCGTGC



# Bioinformatica

la parte più importante dei sequenziamenti NGS è  
l'allineamento di milioni di piccole sequenze per creare sequenze lunghe e  
continue



# Next Generation Sequencing

## Applicazioni:

- Sequenziamento di interi genomi
- *De Novo* Sequencing (DNA sconosciuti)
- Trascrittomica : Sequenziamento di tutto l'mRNA
- Metagenomica
- Altro...



## **Perché il sequenziamento su capillare è ancora il metodo più importante**

1. Il sequenziamento su capillare è ancora la metodica più affidabile nell'accuratezza delle chiamate (>99,99 % accuratezza, ancora superiore all'NGS). In pubblicazioni anche recenti dove si è fatto uso di NGS per identificare SNP su vasta scala, si è proceduto alla conferma di quelli ritenuti di interesse con sequenziamento Sanger su capillare.

1. Lunghezza delle sequenze fino a 900 nucleotidi (senza quindi dover procedere a successivi assemblaggi di più sequenze)

1. Evoluzione dell'automazione con sequenziatori anche a 96 capillari (noi abbiamo un 8 capillari).



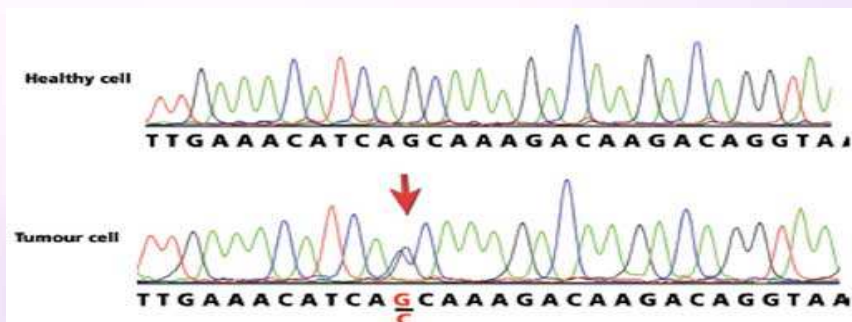
## **Perché il sequenziamento su capillare è ancora il metodo più importante**

4. In diagnostica clinica dove si lavora su sequenze di DNA note come prodotti di PCR diagnostiche, è un metodo rapido di conferma del risultato o di caratterizzazione del campione.

4. Un sequenziatore a capillare viene utilizzato anche per altre analisi come analisi dei frammenti in forense (es. analisi di paternità o identificazione individuo) o analisi sulle singole mutazioni puntiformi SNP (es. primer extension) dove tecniche NGS sarebbero inappropriate e comunque meno affidabili.

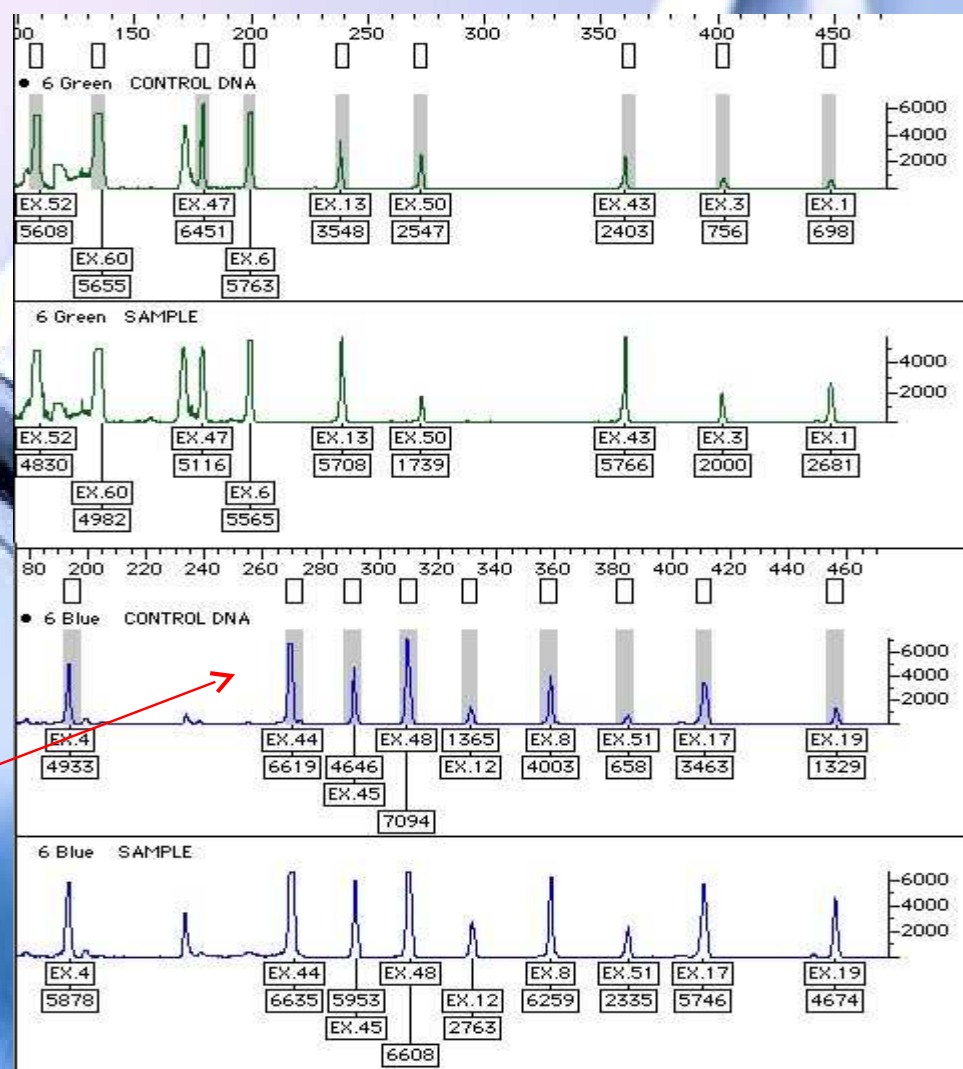
4. Geni particolarmente informativi possono essere amplificati e sequenziati, per risalire ad esempio alla specie animale o tipizzare un batterio o **ottenere dati per una filogenesi.**

## Flessibilità del sequenziatore a capillare



Analisi SNPs già caratterizzati

Microsatelliti (Forense) analisi  
frammenti per diagnosi molecolari di  
malattie genetiche





AATTCGATGAATTACAATT  
T**GRAZIE**AACCACTCGACT  
TAGTTACTCCATA**PER**ATC  
CGGTCAAGGTTAGTCTTAC  
CTCL'**ATTENZIONE**GGTAC  
CTAGGTTATAGAATTCAGT  
ACCG