



# Next Generation Sequencing ed analisi degli OGM

ROMA, 6 OTTOBRE 2020

KATIA SPINELLA

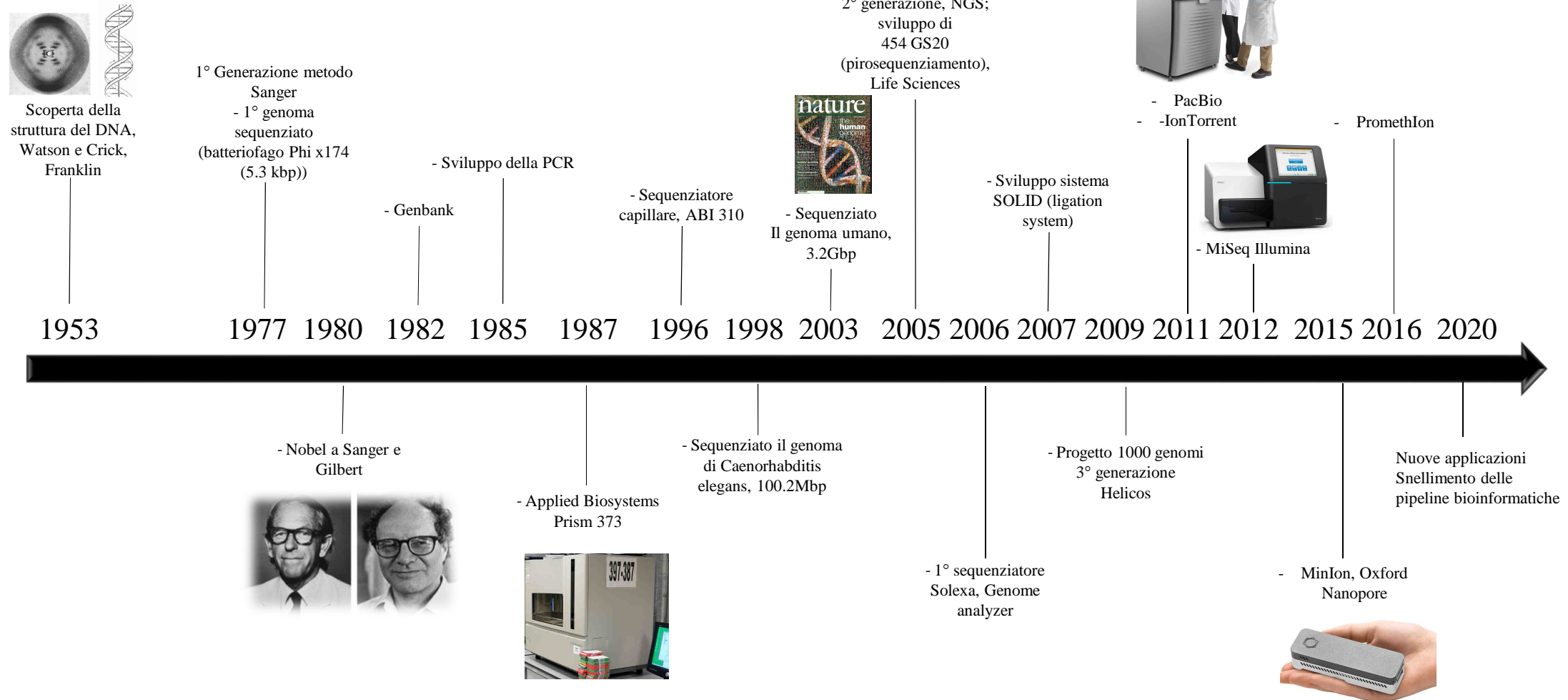


## Di cosa parleremo...

1. Un po' di storia: come si è evoluta la tecnologia del sequenziamento degli acidi nucleici
2. Introduzione alle metodiche di sequenziamento degli acidi nucleici
3. Funzionamento di alcune piattaforme di NGS
4. Workflow ; protocolli operativi
5. Applicazioni
6. Tecnologie NGS per analisi degli OGM (progetti in corso presso il CROGM ed esempi in letteratura) e di MOGM (esempi in letteratura)



# Ripercorrendo la storia...



# Introduzione alle metodiche di sequenziamento degli acidi nucleici

## Generazioni a confronto

### 1<sup>a</sup> generazione

- Sanger sequencing

### 2<sup>a</sup> generazione Short-read sequencing

- Next-Generation sequencing; - Roche 454 (2005, tecnologia pyrosequencing)  
- ABI SOLiD (2007, tecnologia ligation and two-base coding)  
- Illumina (Solexa) (2006, sequencing by synthesis (SBS))

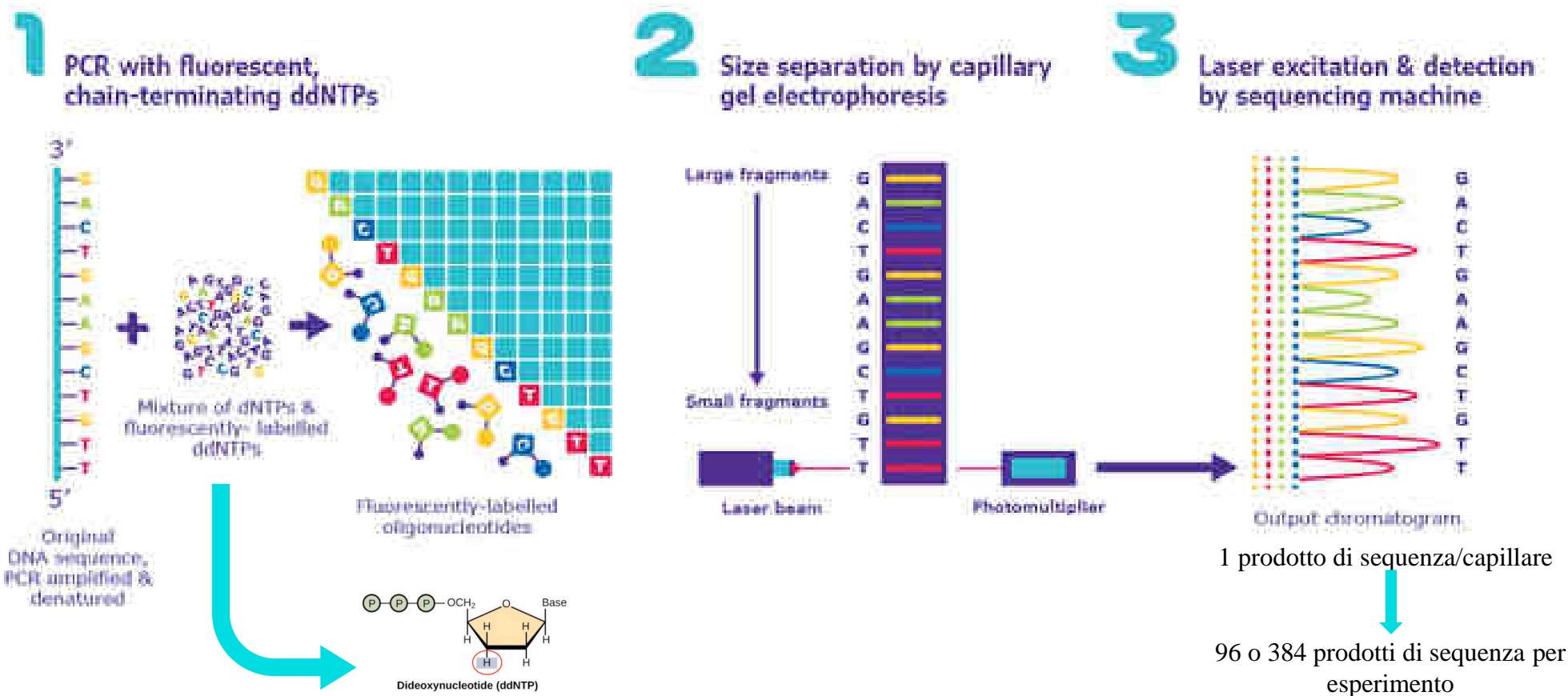
### 3<sup>a</sup> generazione Single- molecule long- read sequencing

- Next-Next (3rd) Generation sequencing; - VisiGen  
- Helicos (2009)  
- PacBio RS System (2011)  
- Oxford Nanopore (2015)



# Sequenziamento di 1° generazione

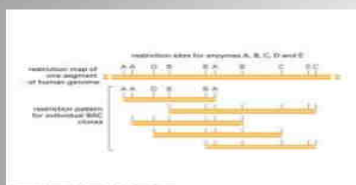
## Metodo Sanger o metodo a terminazione di catena





- ❖ Inizio del progetto 1990
- ❖ Identificare e mappare i circa 30 mila geni del genoma umano
- ❖ Prima bozza rilasciata nel 2000
- ❖ 2 strategie utilizzate : sequenziamento gerarchico e sequenziamento shotgun
- ❖ Progetto terminato nel 2003

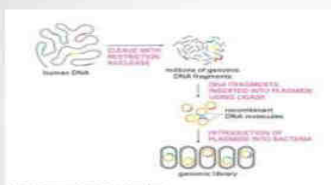
## Clone by Clone



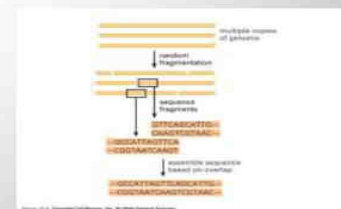
Create a crude physical map of the whole genome by restriction mapping before sequencing

Break the genome into overlapping fragments and insert them into BACs and transfect into *E.coli*

## Shotgun sequencing



Break genome into random fragments, insert fragments into vector, sequence each of the fragments and assemble fragments based on sequence overlaps



- need known flanking region to anneal primer



## Diapositiva 6

---

**T10**

<https://www.slideshare.net/gabrielemusilli/gambari-balez10>

TofaniS; 23/09/2020

# Next-generation sequencing (NGS)

**Next-generation sequencing (NGS):** una serie di metodi che permettono di analizzare sequenze di DNA con strategie che permettono di ottenere milioni di sequenze in parallelo (sequenziamento massivo parallelo).

- Second-generation sequencing: Illumina, Roche 454, SOLiD and IonTorrent (LifeTechnologies), ... letture fino a 600bp
- Third-generation sequencing: MinION (OxfordNanopore), PacBio (Pacific Biosciences), ... letture >1000bp

## STRUMENTI NGS





## Piattaforme di 2° generazione

- ❖ Piattaforme di sequenziamento basate su diverse tecnologie
- ❖ Elevata accuratezza del dato di sequenza
- ❖ Standardizzazione protocolli di sequenziamento e analisi
- ❖ Prodotti di sequenziamento (**reads**) corti (max 600 bp) (Short-read sequencing )
- ❖ Diversi modelli di strumento in grado di generare **output** diversi (in Gb)



# NGS

## Wet-lab

1. preparazione della library
2. sequenziamento e imaging

## Dry-lab

T1

3. analisi bioinformatica dei dati



### T1

Sebbene si differenzino per il tipo di biochimica alla base e per i metodi di acquisizione e di elaborazione dei dati, le piattaforme NGS sono accomunate da un work-flow operativo simile che prevede tre fasi principali: preparazione e amplificazione del campione di DNA (spezzare il genoma in più frammenti di lunghezza contenuta, a seconda del metodo seguito, che possono essere amplificati in vari modi);

sequenziamento e imaging (ogni base in ogni frammento viene identificata, creando delle reads);

analisi bioinformatica dei dati (si sfruttano software bioinformatici per concatenare reads sovrapposte, allungando quindi la lunghezza dei frammenti. Maggiore la lunghezza della sequenza, migliori i risultati).

TofaniS; 09/09/2020

## FORMAZIONE:

*Tecnici di laboratorio* per la **wet part**: estrazione del DNA, preparazione del campione e sequenziamento

✓ Esperienza pregressa (biologia molecolare)

Formazione su:

- Basi e principi di sequenziamento
- Flussi di laboratorio e protocolli
- Identificazione e risoluzione dei problemi

*Bioinformatici* per la **dry part**: analisi post-sequenziamento e gestione degli stessi dati

✓ Esperienza pregressa (biologia molecolare e/o bioinformatica)

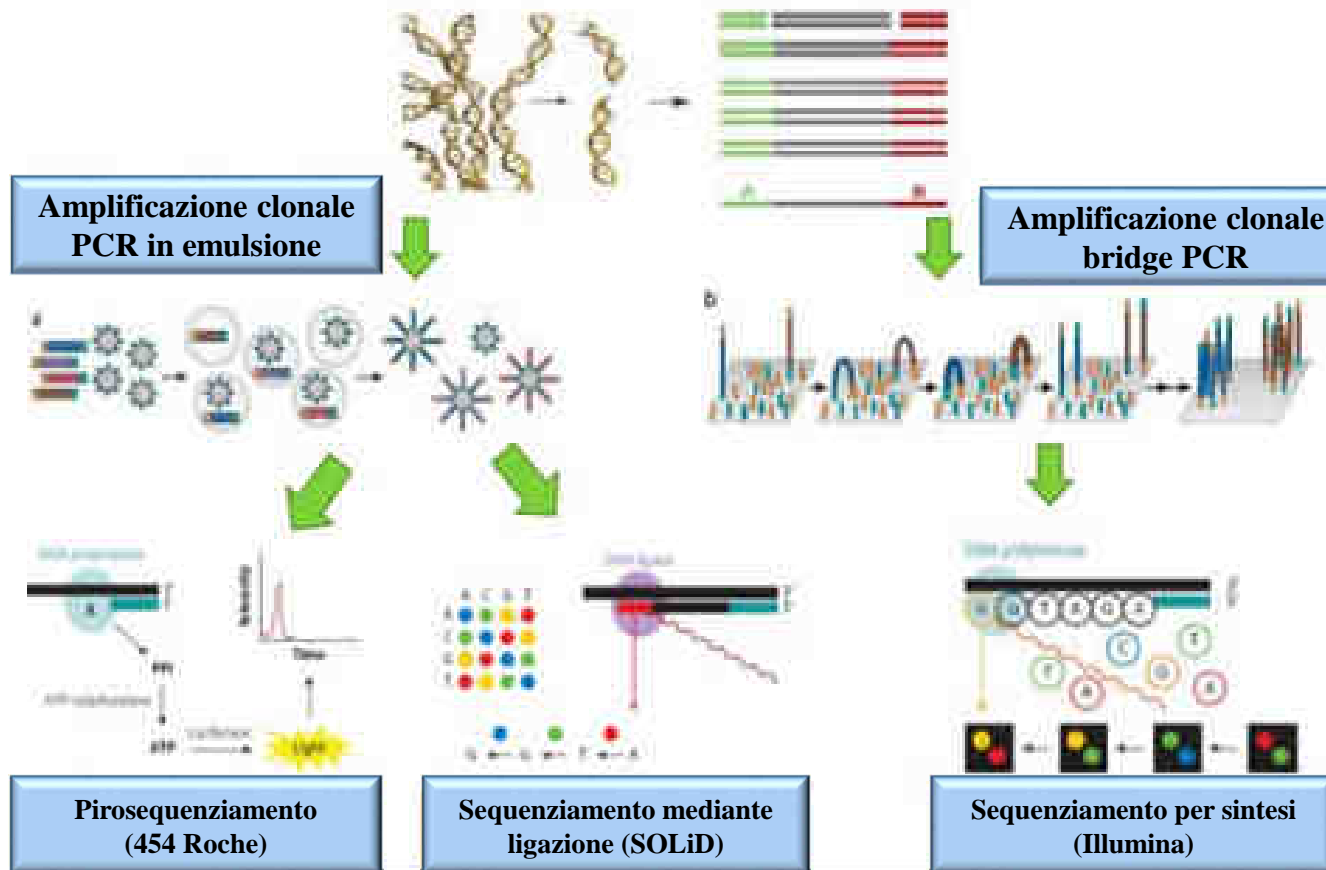
Formazione su:

- Sistemi operative UNIX e Linux
- Utilizzo dei comandi in riga e analisi di pipeline; uso di software commerciali o disponibili online
- Archiviazione dei dati di sequenza
- Identificazione e risoluzione dei problemi



# NGS WORKFLOW

Estrazione del DNA, frammentazione e legame adattatori



## 1. Preparazione della library;

- Frammentazione (fisica o enzimatica)
- Le estremità vengono riparate
- Alle estremità vengono legati degli adattatori

## 2. Amplificazione dei frammenti;

- Sfere in emulsione
- Amplificazione avviene sulla flow cell

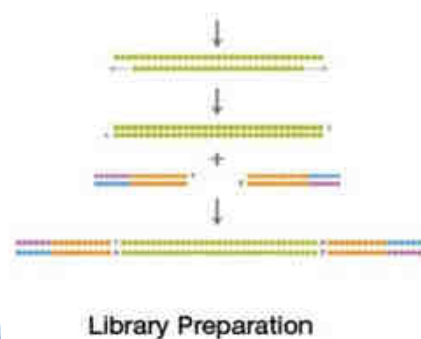
## 3. Sequenziamento;

- Sequenziamento per sintesi
- Sequenziamento per ligazione

# Illumina (tecnologia sequencing by synthesis (SBS))

Il flusso di lavoro NGS su piattaforma Illumina prevede 4 passaggi fondamentali:

A. Preparazione della library



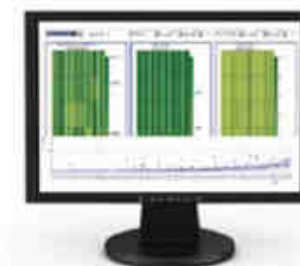
B. Generazione dei cluster



C. Sequenziamento per sintesi



D. Analisi dei dati (dry-lab)



Laboratorio è bello



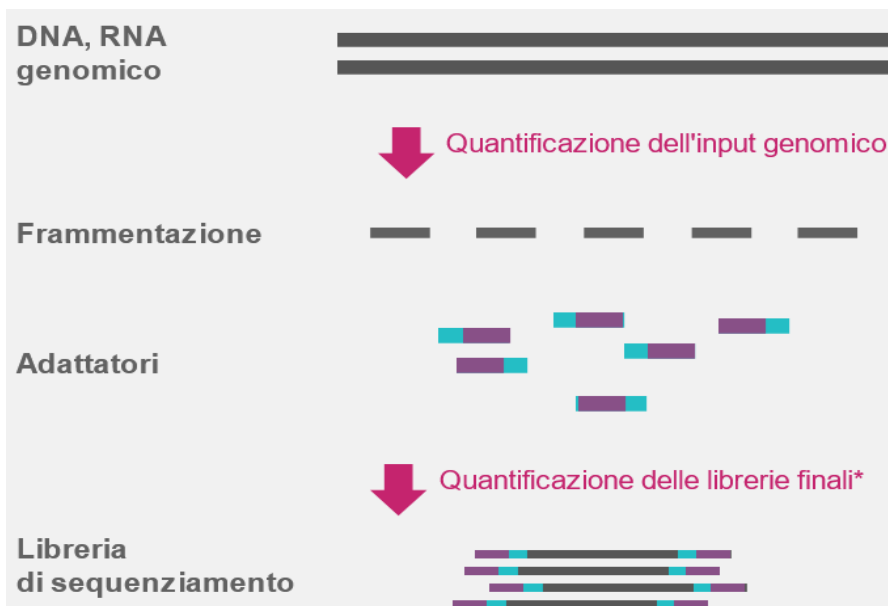


## A. Preparazione della library

**LIBRARY:** Miscela di frammenti di DNA da sequenziare modificati per essere compatibili con la piattaforma di sequenziamento e con la strategia da applicare.

Nel nostro caso i prodotti di PCR da sequenziare devono avere tre requisiti:

- 1) Devono aderire al supporto su cui si verificherà la reazione di sequenziamento (**Flow cell**)
- 2) Devono essere riconosciuti dai primer che funzioneranno da innesco per la reazione di sequenziamento
- 3) Devono contenere un **tag** (bandierina) che permetta di risalire al campione di provenienza (**indexing**)



### Per la generazione di cluster

Le librerie devono avere le regioni di legame P5 e P7, che interagiscono con gli oligonucleotidi sulla superficie della cella a flusso.

### Per il multiplex

Le librerie devono avere un indice univoco o una sequenza marcata per miscelare i campioni.

### Per il sequenziamento

Le librerie devono avere regioni di legame per far sì che i primer di sequenziamento attivino Read 1 (Lettura 1) e Read 2 (Lettura 2).



## B. Generazione dei cluster

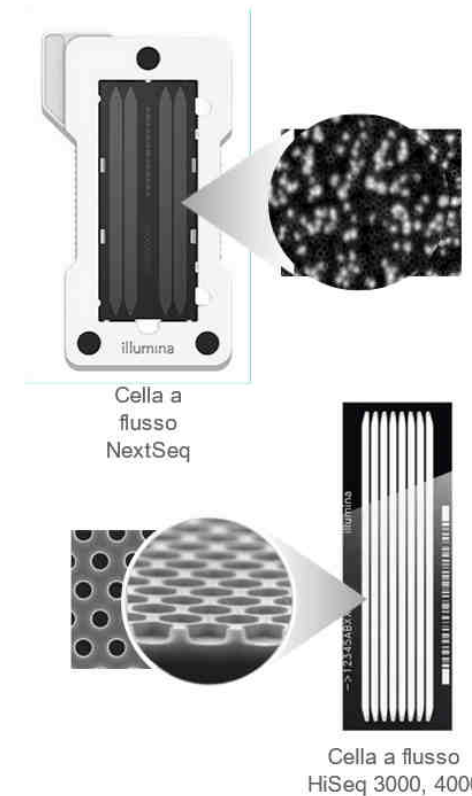
La generazione dei cluster avviene sulla flow cell

### FLOW CELL

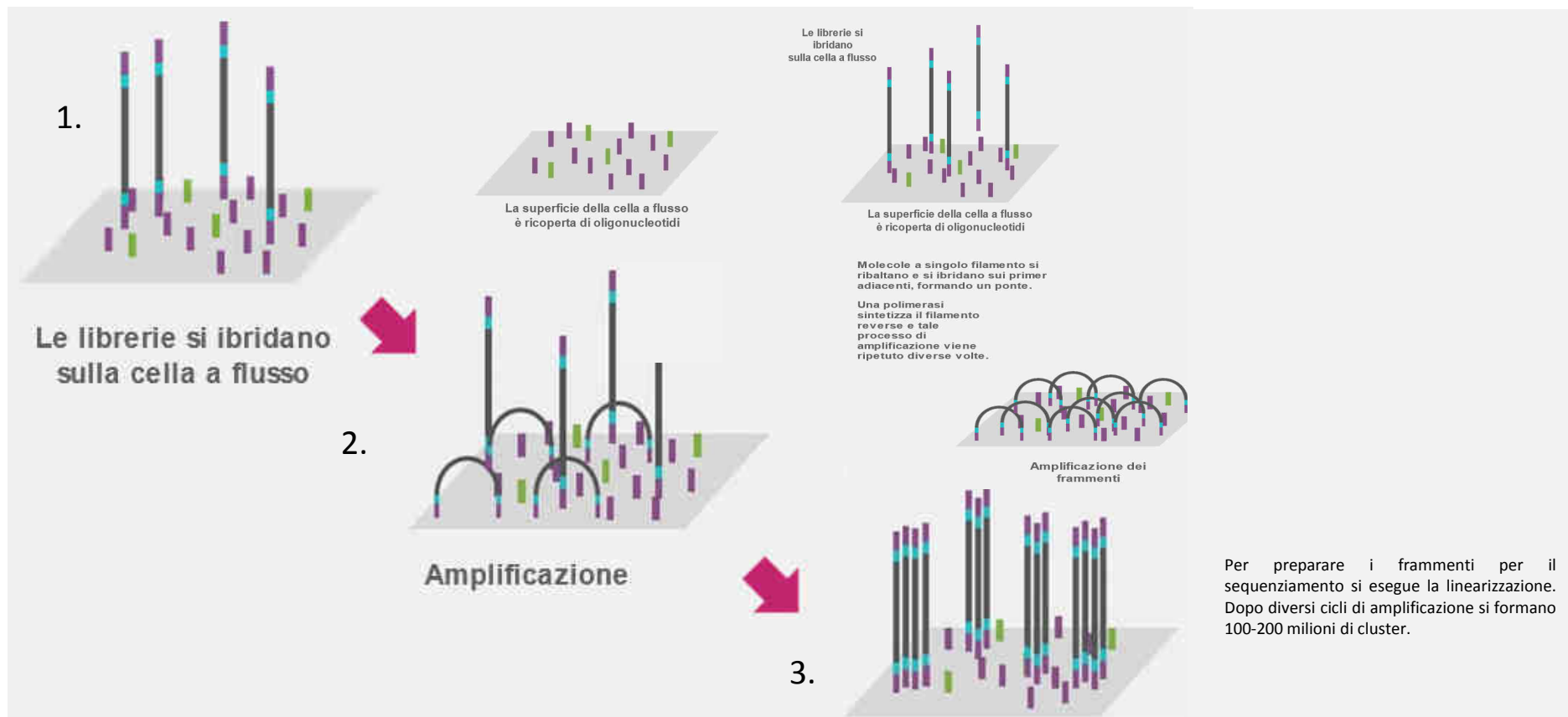
Supporto solido su cui viene immobilizzato il DNA da sequenziare (template). I reagenti liquidi possono fluire all'interno dei nano-pozzetti ed essere ciclicamente rimossi tramite lavaggi.

Una flow-cell è un vetrino suddiviso in corsie composte da milioni di nano-celle dove avviene fisicamente la reazione di sequenziamento.

Ogni strumento di sequenziamento utilizza una determinata flow cell.



## B. Generazione dei cluster (2)

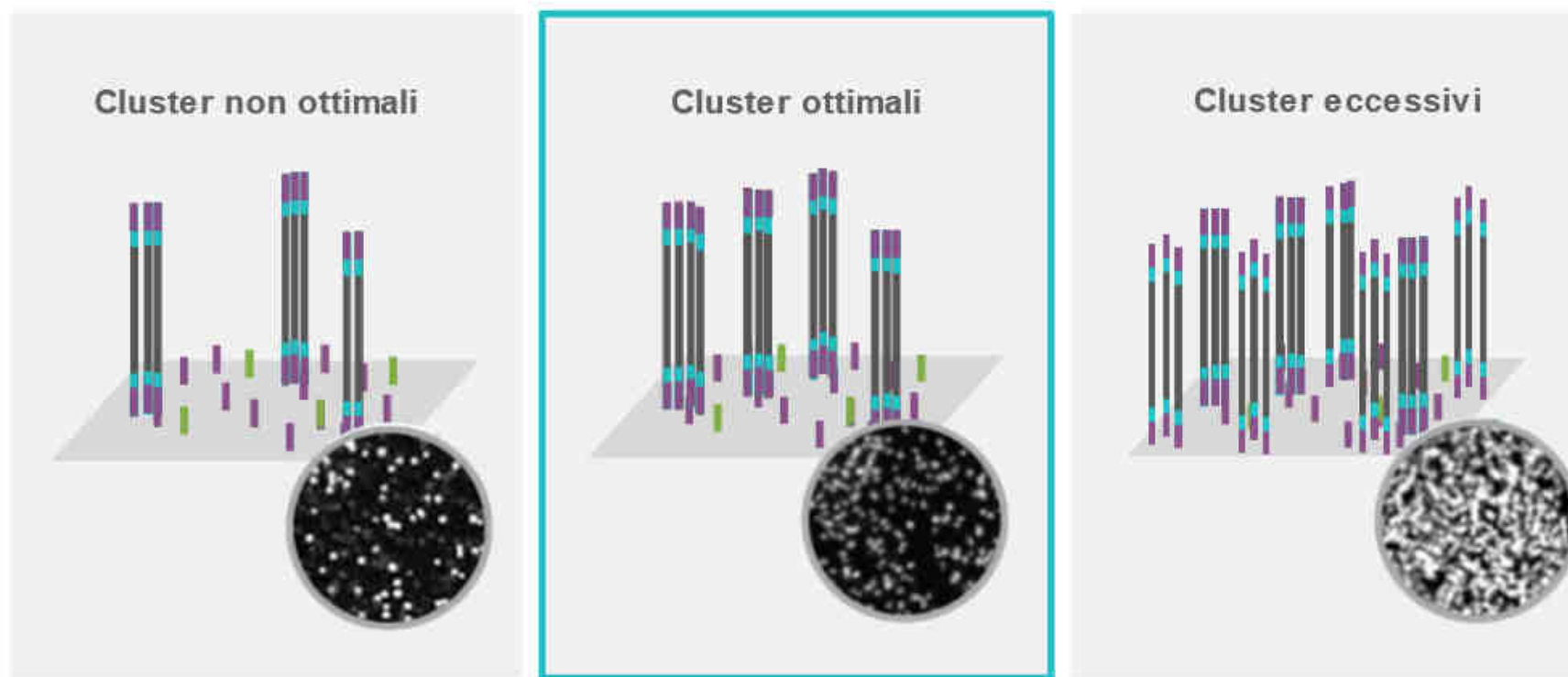


Ciascun frammento della libreria viene clonato in migliaia di copie identiche



## Cluster density

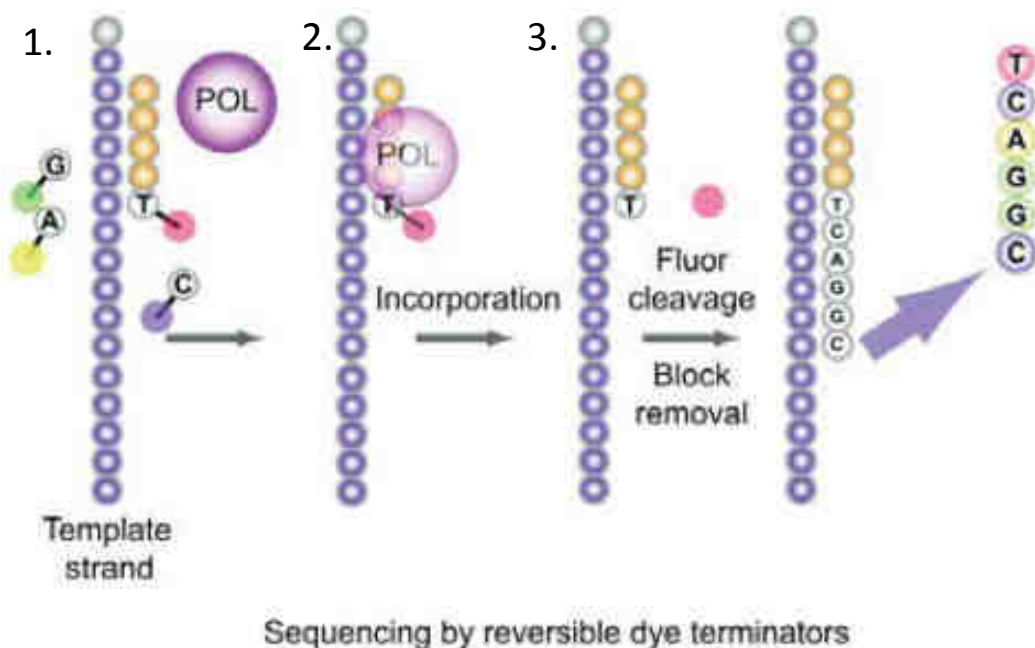
La generazione di cluster sulla flow cell determina la qualità e la resa dei dati. E' importante quantificare la library prima del sequenziamento.





## C. Sequenziamento mediante sintesi (SBS)

Un metodo basato su terminatori reversibili che consente il sequenziamento massivo in parallelo di miliardi di frammenti di DNA, rilevando singole basi mentre vengono incorporate in filamenti di DNA in crescita.



### 1. Incorporazione del nucleotide

Dei nucleotidi marcati con fluorofori legano la base complementare sul template. Si utilizzano 4 fluorofori differenti.

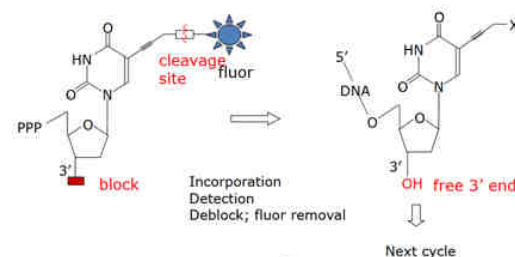
Il fluoroforo blocca il sequenziamento. Ogni cluster può incorporare una base diversa.

### 2. Acquisizione della fluorescenza

Un laser colpisce il cluster. Ogni cluster emette un colore corrispondente al nucleotide incorporato durante questo ciclo.

### 3. Taglio (cleavage)

I fluorofori vengono tagliati e lavati via dalla flow cell rigenerando il 3'OH. Inizia un nuovo ciclo con l'aggiunta di nuovi nucleotidi marcati.



## Diapositiva 17

---

**T7** Tecnologia SBS (Sequencing by Synthesis), un metodo basato su terminatori reversibili che consente il sequenziamento massivo in parallelo di miliardi di frammenti di DNA, rilevando singole basi mentre vengono incorporate in filamenti di DNA in crescita. I coloranti fluorescenti dei terminatori vengono acquisiti via via che ogni dNTP viene aggiunto e poi scisso per consentire l'incorporazione della base successiva. Durante ogni ciclo sono presenti tutti e quattro i dNTP legati ai terminatori reversibili e la competizione naturale riduce al minimo le distorsioni dovute all'incorporazione. Durante ogni ciclo viene eseguita l'identificazione delle basi direttamente dalle misurazioni dell'intensità del segnale.

TofaniS; 21/09/2020

**T8** Ogni ciclo rappresenta l'incorporazione di un nucleotide legato a un fluoroforo al filamento di DNA in via di formazione.

TofaniS; 21/09/2020



# Kit di sequenziamento; Come sceglierlo?

	MiSeq Reagent Kit v2			MiSeq Reagent Kit v3	
Lunghezza delle reads	2 x 25 bp	2 x 150 bp	2 x 250 bp	2 x 75 bp	2 x 300 bp
kit size (cicli)	50	300	500	150	600
Tempo per il sequenziamento	5.5 ore	24 ore	39 ore	21 ore	56 ore
Output	850 Mb	4.5 Gb	7.5 Gb	3.3 Gb	13.2 Gb
n. di reads		15 M			25 M
	MiSeq Reagent Kit v2 Micro			MiSeq Reagent Kit v2 Nano	
Lunghezza delle reads		2 x 150 bp		2 x 250 bp	2 x 150 bp
kit size (cicli)		300		500	300
Tempo per il sequenziamento		19		28 ore	17 ore
Output		1.2 Gb		500 Mb	300 Mb
n. di reads		4 M			1 M



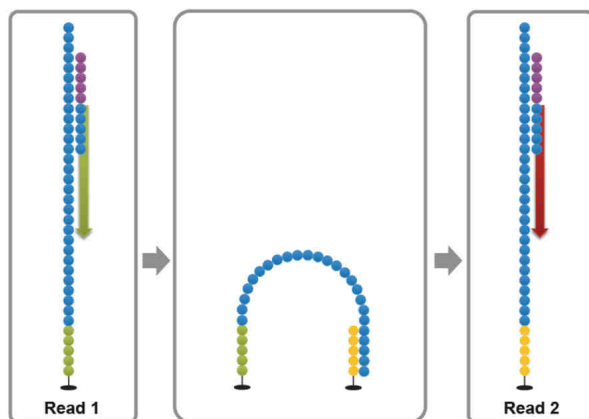
- Flow cell ; Standard, Micro o Nano
- Numero di cicli
- Output
- Coverage



## Come calcolare la lunghezza delle reads?

Tutti i reagenti di sequenziamento Illumina presentano un certo numero di cicli di sequenziamento. Questi cicli sono direttamente correlati alla lunghezza della sequenza. Poiché una base è sequenziata per ciclo, il numero totale di cicli indica il numero massimo di basi che possono essere sequenziate. È possibile utilizzare reagenti di sequenziamento per generare letture continue singole (single end) o per sequenziamento paired-end in entrambe le direzioni. (Ad esempio, un kit da 300 cicli può essere utilizzato per una corsa in lettura singola  $1 \times 300$  bp o una corsa in paired-end  $2 \times 150$  bp.)

### Illumina Paired-End Sequencing



### DNA Sequencing Applications

Application	Recommended Read Length
Whole-genome sequencing	$2 \times 150$ bp
Whole-exome sequencing	$2 \times 150$ bp
Targeted enrichment sequencing	$2 \times 150$ bp
Amplicon sequencing	Length of the entire amplicon insert
De novo sequencing	Ranges from $2 \times 150$ to $2 \times 300$ bp

Il numero di cicli determina la lunghezza delle reads ed è uno dei parametri da considerare quando si seleziona il kit di sequenziamento (**KIT SIZE**)



# Output

L'output è il numero di basi totali lette dal sequenziatore, espresso in Mb (mega basi) o Gb (giga basi).

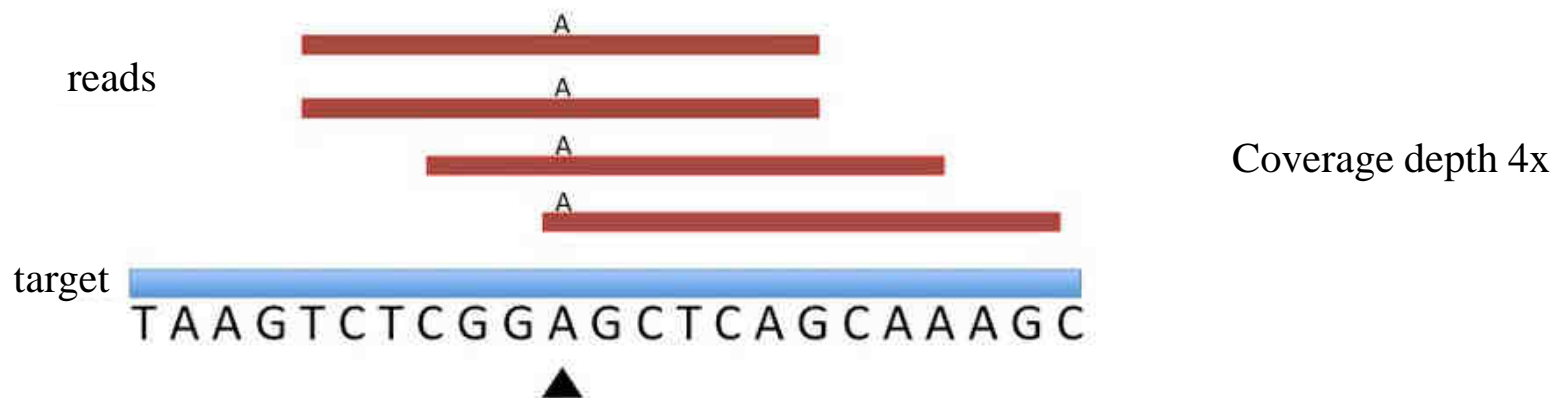
	<u>MiSeq Reagent Kit v2</u>			MiSeq Reagent Kit v3	
Lunghezza delle reads	2 x 25 bp	2 x 150 bp	2 x 250 bp	2 x 75 bp	2 x 300 bp
kit size (cicli)	50	300	500	150	600
Tempo per il sequenziamento	5.5 ore	24 ore	39 ore	21 ore	56 ore
Output	850 Mb	4.5 Gb	7.5 Gb	3.3 Gb	13.2 Gb
n. di reads		15 M		25 M	
	MiSeq Reagent Kit v2 Micro			MiSeq Reagent Kit v2 Nano	
Lunghezza delle reads	2 x 150 bp			2 x 250 bp	2 x 150 bp
kit size (cicli)	300			500	300
Tempo per il sequenziamento	19			28 ore	17 ore
Output	1.2 Gb			500 Mb	300 Mb
n. di reads	4 M			1 M	

Es: Standard Flow Cell con kit da 300 cicli (reads paired- end lunghe 150) ha un output di 4,5Gb pari a 15 milioni di reads.  $150 \text{ bp} \times 2 \times 15.000.000 = 4,5 \text{ Gb}$



## Coverage depth (profondità di lettura)

Il numero medio di letture (reads) ottenute dal sequenziamento per ciascuna base nucleotidica costituente il target. Maggiore è la profondità di lettura più è accurato il dato di sequenza.



Il coverage è un parametro deciso dall' operatore in base; - all' applicazione e all'obiettivo dello studio

- dimensione del genoma da sequenziare o del target
- livello di espressione dei geni
- riferimenti in letteratura di studi simili

## Come scegliere il coverage?

- Un coverage di 20-30X permette di sequenziare in maniera poco profonda il genoma : si può scegliere questo coverage per un resequencing di genomi già noti e per la ricerca di SNP già annotati
- Un coverage di 50-100X permette di sequenziare in profondità i genomi ed analizzare tutte le varianti presenti,



# Come calcolare il coverage

## Esempio

### Amplicon sequencing

- Dimensioni del target\*: 888000 bp
- Numero di reads:  $10^6$
- Tipo di corsa: 2X150

$$\text{Coverage} = \frac{(\text{n. reads}) \times (\text{lunghezza delle reads})}{\text{dimensione del target}} = \frac{150 \times 10^6}{888000} = 169 \times$$

\*Per ottenere il coverage dobbiamo calcolare la lunghezza totale (in bp) dei target da analizzare e moltiplicarli per il numero totale di campioni;

Ogni target è di 250 bp

37 target

96 campioni

Lunghezza totale dei target sequenziati =  $250 \text{ bp} \times 37 \times 96 = 888.000 \text{ bp}$







## D. Analisi dei dati

L'analisi dei dati ottenuti dal sequenziamento, rappresenta un complesso “montaggio di dati” che avviene con l'ausilio di algoritmi bioinformatici ben strutturati e con l'impiego di software particolari. Questo passaggio è specifico a seconda dell'obiettivo da raggiungere.

### Analisi bioinformatica

#### Qualità

- FastQC
- Cutadapt
- Trimmomatic
- ...

#### De-novo Assembly

- Spades
- Abyss
- Trinity RNAseq
- MIRA
- Megahit
- Velvet
- ...

#### Reference mapping

- BWA
- Bowtie
- BBMap
- ...

#### Metabarcoding

- QIIME
- Mothur
- R/Bioconductor

Software per linux

### D. Alignment and Data Analysis

Reads

```

ATGGCATTGCAATTTGACAT
TGGCATTGCAATTTG
AGATGGTATTG
GATGGCATTGCAA
GCATTGCAATTTGAC
ATGGCATTGCAATT
AGATGGCATTGCAATTTG
  
```

Reference Genome

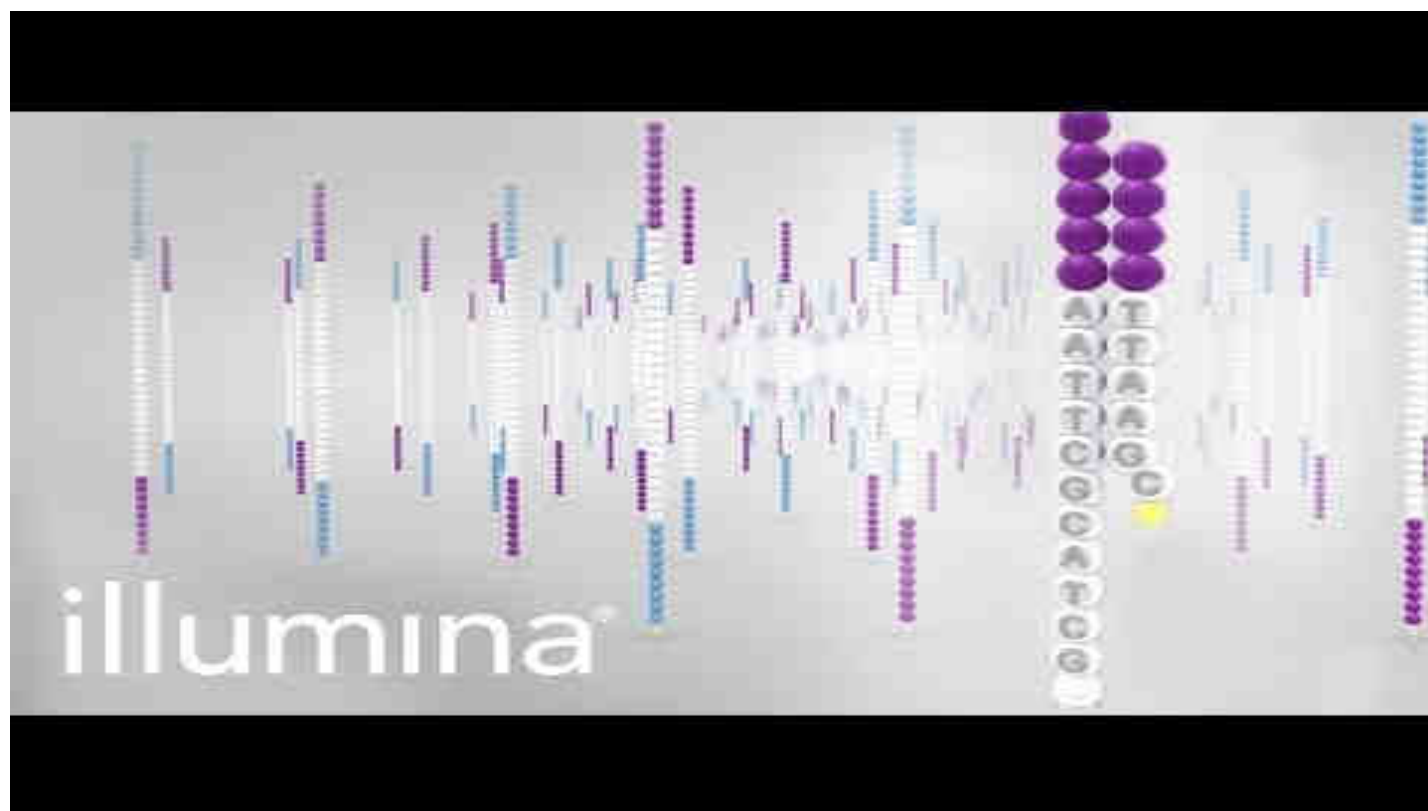
```

AGATGGTATTGCAATTTGACAT
  
```

Reads are aligned to a reference sequence with bioinformatics software. After alignment, differences between the reference genome and the newly sequenced reads can be identified.



## Video; Tecnologia Illumina



## Aspetti critici che influenzano la qualità dei dati



Factors that Contribute to Platform Accuracy





Istituto Zooprofilattico Sperimentale  
del Lazio e della Toscana M. Aleandri

## ILLUMINA; VANTAGGI E SVANTAGGI

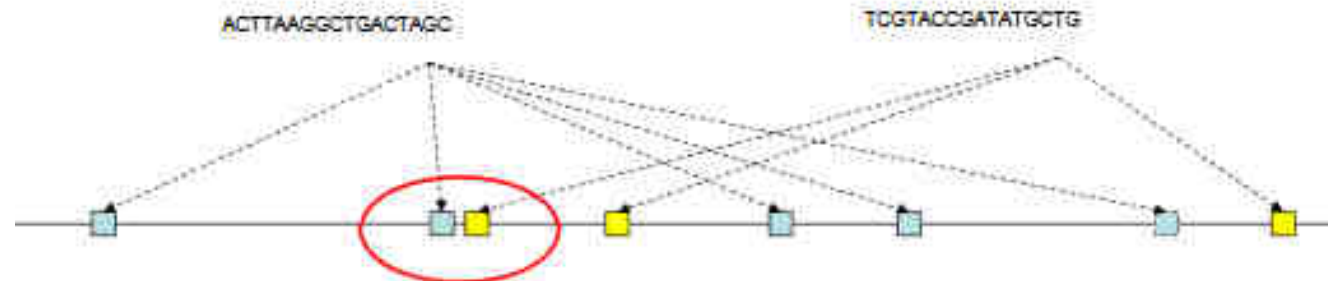


Centro di Riferenza Nazionale  
per la Ricerca di OGM

- Versatilità dello strumento
- Efficienza dello strumento;
  - Costo per corsa
  - Dati per corsa
- Risultati confrontabili perché largamente diffuso
- Sequenze pulite
- Sequenze corte



# Shorts reads



Le reads corte sono problematiche, perchè corte sequenze possono apparire in vari punti del genoma

Soluzione #1: ottenere lunghe reads

Soluzione #2: sequenziare in entrambe le direzioni (paired-end)





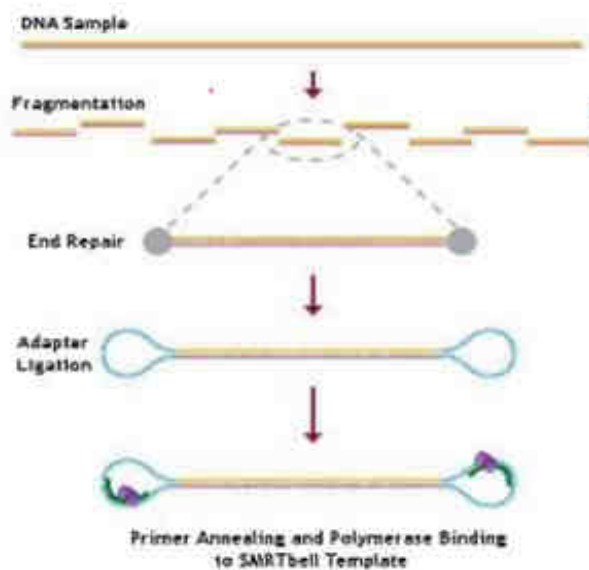
## Sequenziamento di 3° generazione

- ❖ Piattaforme di sequenziamento basate su diverse tecnologie dai costi altamente variabili
- ❖ Accuratezza del dato di sequenza generalmente più bassa
- ❖ Prodotti di sequenziamento (**reads**) lunghi (decine di Kb)
- ❖ Protocolli di analisi complessi, non sempre standardizzabili

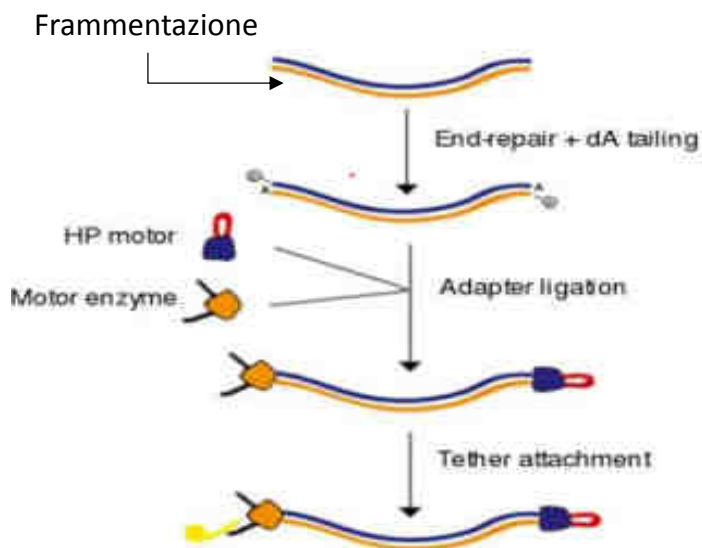


# Preparazione delle libraries

❖ Frammentazione del DNA, end-repair e ligazione



PacBio



Nanopore



**T2** La tecnologia di sequenziamento adottata dal sistema pacBio RS II è detta SMRT sequencing (sigla brevettata che sta per Single Molecule Real-Time sequencing). In questo sistema il filamento di DNA da sequenziare viene ancorato, insieme ad una DNA polimerasi, sul fondo di una zero-mode waveguide (ZMW - una nanostruttura fotonica che è sostanzialmente un minuscolo tubo all'interno del quale è possibile rilevare cambiamenti infinitesimali di luce colorata). Il complesso DNA-polimerasi, ancorato sul fondo della ZMW, viene cimentato con una soluzione contenente i quattro nucleotidi (A, C, G e T) opportunamente marcati ciascuno con un fluorocromo diverso. La DNA polimerasi comincia ad incorporare i nucleotidi per costruire un filamento complementare a quello ancorato sul fondo. Man mano che la polimerasi incorpora un nucleotide, il fluorocromo del nucleotide viene rilasciato e l'apposito sistema di rilevazione registra l'evento luminoso associato. Il software di elaborazione rielabora quindi i segnali luminosi registrati fino a ricostruire la sequenza.

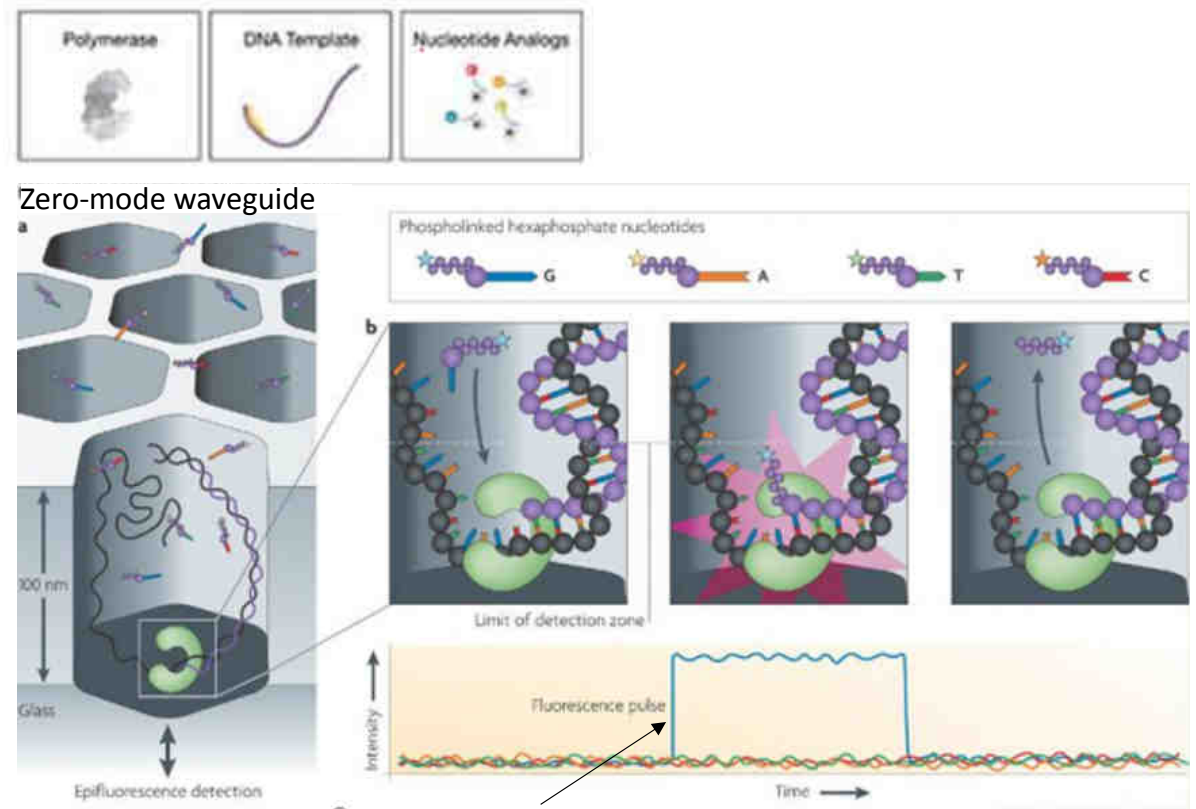
TofaniS; 09/09/2020

**T3** dA-tailing: Viene adenilata l'estremità -3' mediante l'addizione di una singola A. Questo può essere fatto mediante l'utilizzo di una Taq polimerasi[4] o mediante altri enzimi come le transferasi terminali che permettono l'aggiunta di un singolo nucleotide all'estremità -3'. La corrispondente T è presente all'estremità -3' dell'adattatore e ciò permette la formazione di due legami a idrogeno tra le due basi complementari e la conseguente ligazione.

TofaniS; 09/09/2020

## Single Molecule Real Time Sequencing (SMRT)

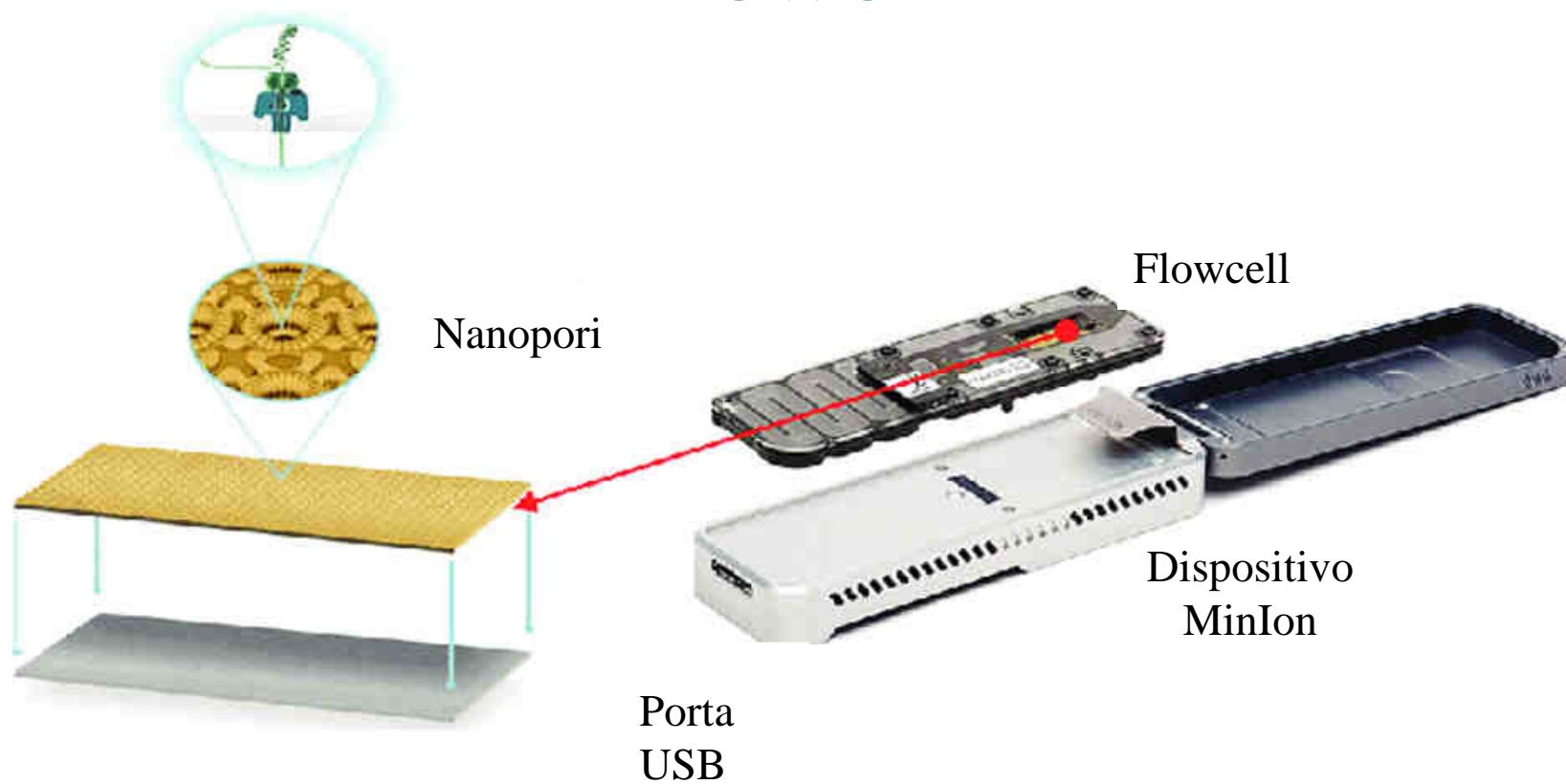
- ❖ Template ancorato insieme alla polimerasi sul fondo di una ZMW
- ❖ 4 nt marcati ognuno con un fluorocromo diverso
- ❖ La polimerasi sintetizza il nuovo filamento, incorpora nuovi nt, il fluorocromo viene rilasciato
- ❖ Il sistema di rilevazione registra l'evento luminoso associato



Incorporazione nuovo nt



# PIATTAFORMA NANOPORE FLOWCELL





# Preparazione della library: Adattatori



- **Leader- adaptor;** 2 oligont semi complementari che appaiandosi formano una struttura ad Y
- **Hairpin- adaptor;** 2 oligont con una regione complementare interna che formano una struttura ad hairpin

Entrambi gli adattatori presentano un complesso enzimatico (**motor protein**) che consente il passaggio del DNA attraverso il nanoporo, Gli adattatori guidano il DNA in prossimità del nanoporo attraverso oligont **tethering** che hanno affinità con i polimeri di membrana.





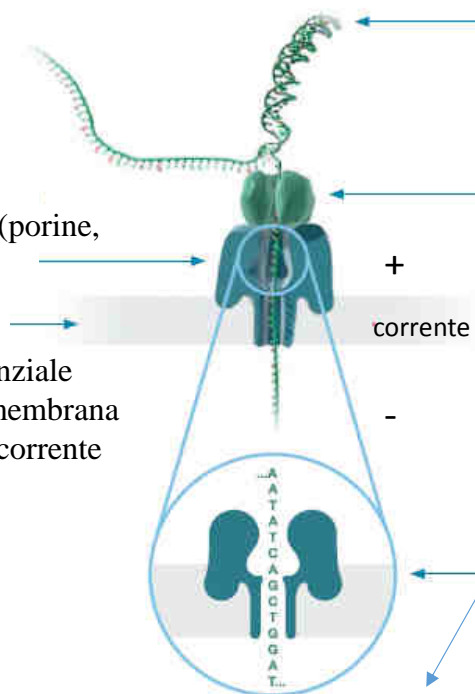
# Come avviene il sequenziamento?

## Nanoporo;

proteine di membrana (porine, es  $\alpha$ -emolisina))

## Membrana;

Una differenza di potenziale applicata ai lati della membrana genera il passaggio di corrente attraverso il poro



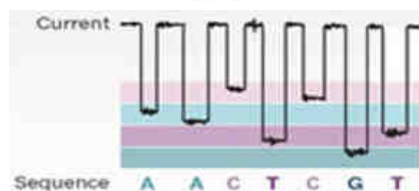
## Template;

dsDNA o cDNA con gli adattatori che consentono l'ancoraggio ai nanopori.

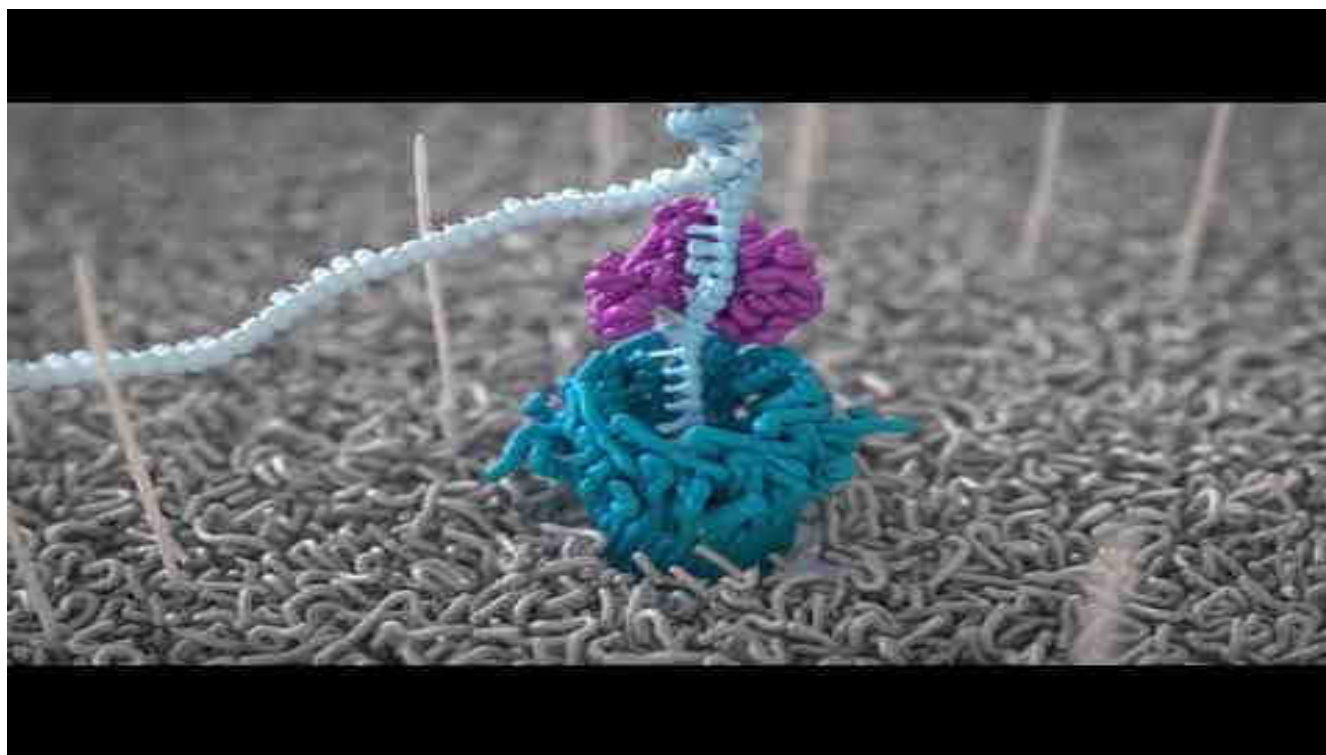
## Complesso enzimatico;

contatta il nanoporo, promuove lo svolgimento della doppia elica di DNA e catalizza il passaggio di uno dei due filamenti (template) attraverso il nanoporo (**single molecule strand sequencing**).

Differenti basi causano una diminuzione più o meno marcata della corrente misurata a seconda dell'ingombro.



# Video; Tecnologia Nanopore



	PacBio RS	Oxford Nanopore
Single Molecule, Real Time sequencing	✓	✓
*Amplification free workflow	✓	✓
**Long reads	✓	✓
***Sequencing by synthesis (DNA polymerase)	✓	X No synthesis

\*Nelle tecnologie precedenti (Sanger, NGS) era necessario **moltiplicare i frammenti in modo da avere un segnale sufficientemente potente per poter essere leggibile**. Riuscire a leggere una **singola molecola di DNA**, monitorando in tempo reale l'aggiunta dei nucleotidi, significa poter usare piccole quantità di DNA risparmiando sui reagenti e sui tempi di preparazione.

\*\***Long reads** offrono un'eccellente opzione di **costruire scaffolds** per assemblaggi *de novo* e per risolvere **regioni ripetitive**, ma in generale mancano di accuratezza.

\*\*\*PacBio utilizza **DNA polimerasi e nucleotidi marcati**. Oxford Nanopore, al contrario, legge una **sequenza di DNA basandosi sui cambiamenti nel flusso di corrente che attraversa dei pori di pochi nanometri**.



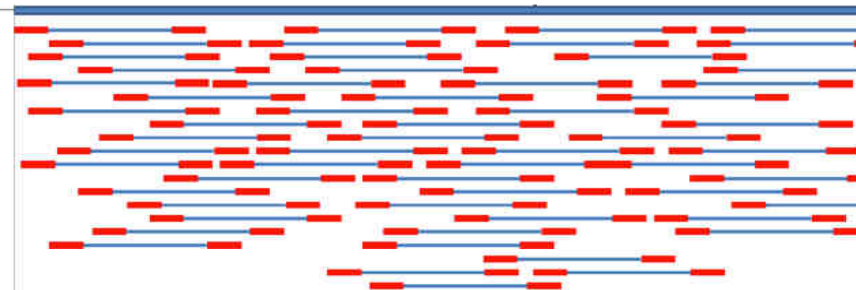


## Ricapitolando...

### Sequenziamento classico (Sanger)



### Sequenziamento Nuova Generazione



**READS:** Dati ottenuti dal sequenziamento, nella forma di una stringa di basi nucleotidiche.

#### Sequenziamento Sanger

- Sequenze lunghe
- Poche sequenze
- **Costo più alto**
- Basso tasso di errore

#### Sequenziamento NGS

- Sequenze corte
- Milioni di sequenze
- **Minor costo**
- Ridondanza per compensare errori di lettura



# Ambiti applicativi



Microbiology



Microbiome



Animal research



Human genetics



Cancer research



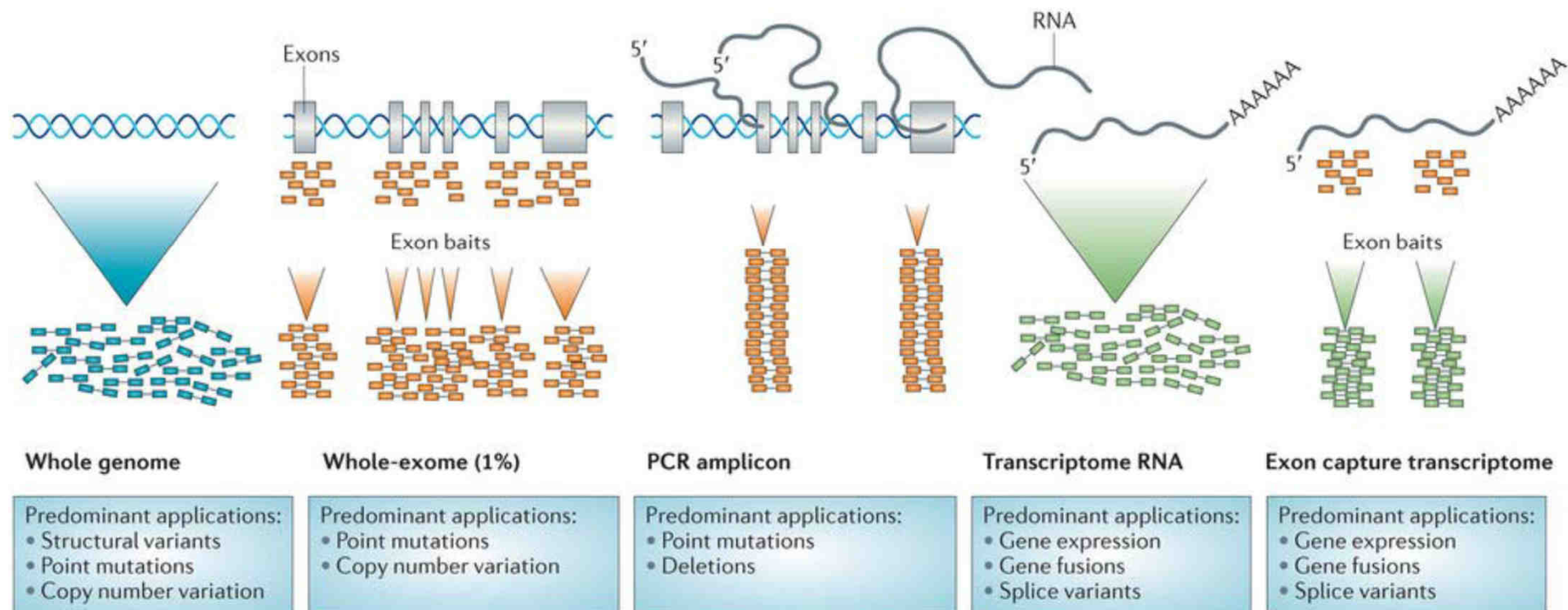
Plant research



Clinical research



# Approcci





**T9**

Whole Genome Sequencing – WGS, noto anche come Whole Genome Shotgun: analisi dell'intero genoma di un individuo

2. Whole Exome Sequencing – WES: analisi dell'intera regione codificante di tutti i geni di un individuo

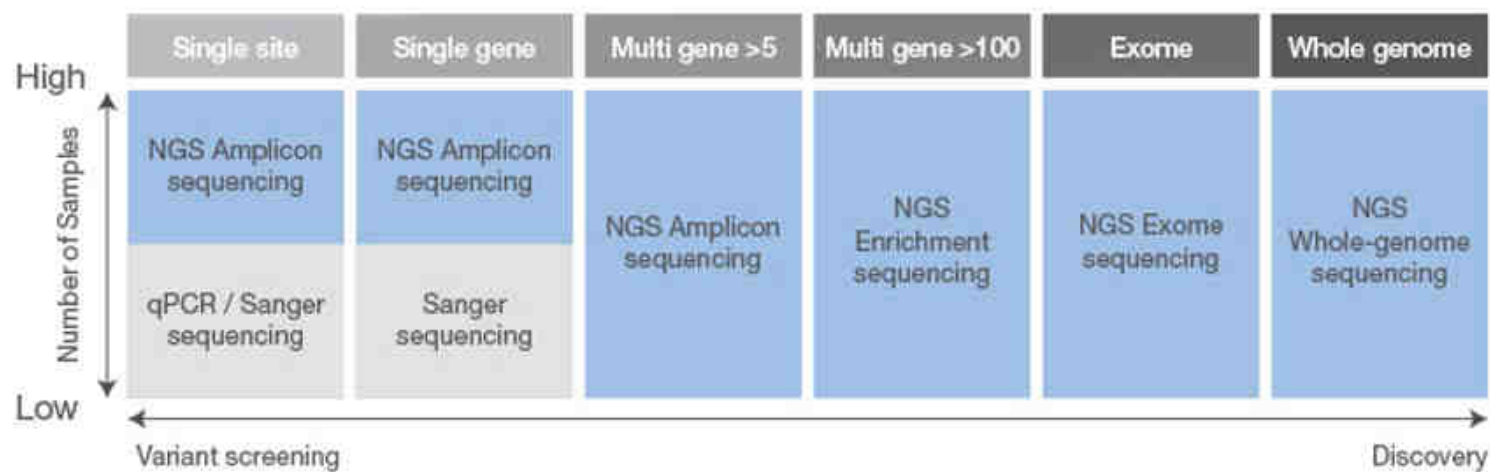
3. Targeted Sequencing: analisi di un gruppo di geni (pannello) o di singolo gene.

4. Transcriptome Analysis: analisi di tutti gli RNA prodotti da una cellula (trascrittoma).

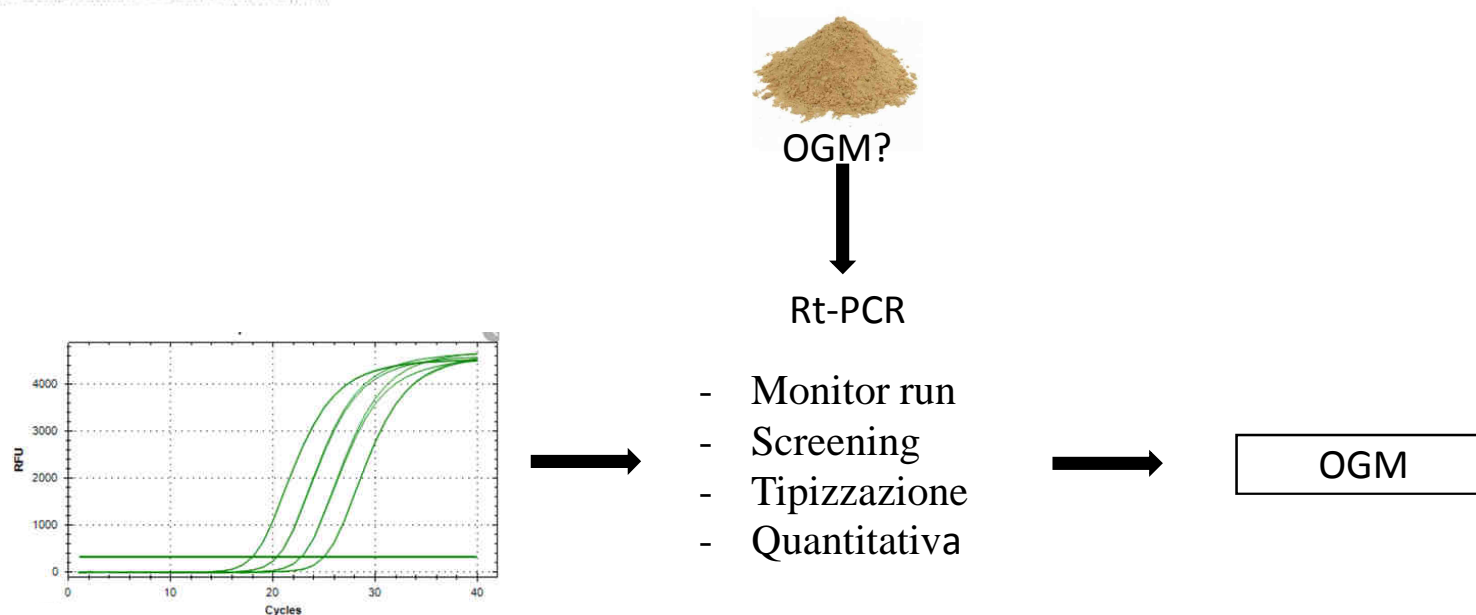
TofaniS; 23/09/2020

# Qual è la migliore piattaforma?

1. Quale applicazione?
2. Quali costi dovrò affrontare?
3. Quanto lunghe devono essere le mie sequenze?
4. Quante sequenze voglio ottenere per ogni ciclo?
5. Quale errore sono disposto a tollerare?
6. Qual è la mole di dati prodotta?



# Analisi di OGM



## Problematiche legate al flusso analitico in rt-PCR per la ricerca di OGM;

- I. Specie GM contaminanti non dichiarate in etichetta
- II. Presenza di organismi donatori (es. virus CaMV)
- III. Presenza di eventi GM non autorizzati



# Applicazioni di metodiche di NGS all'analisi degli OGM

Platform	Output read length	Output no. of reads	Runtime	Type of reads	Comments
Illumina HiSeq	100–150 bp	≤350 million/lane (8 lanes/run)	1–6 days	Paired end	Currently the dominating platform on the market. Insert size <sup>a</sup> 300–600 bp. Lowest cost per sequenced base pair. Sequencing by synthesis
	100–150 bp	≤350 million/lane (8 lanes/run)	1–6 days	Mate pair	Insert size <sup>a</sup> up to several thousand base pairs. Significantly higher costs for library preparation compared with paired-end sequencing
Illumina MiSeq	≤300 bp	≤25 million/run	Hours to 3 days <sup>b</sup>	Paired end/mate pair	Insert size <sup>a</sup> and principle as for HiSeq
Oxford Nanopore Technologies MinION <sup>c</sup>	≤200 kbp	≤2.5 million at 10 kb and standard speed	Minutes to 48 h (sequencing in real time)	Single molecule	Highly flexible read length. Low cost per run. High error rate requires high coverage to obtain consensus sequence. Nanopore sequencing
Pacific Biosciences PacBio	>10 kbp	500 Mbp to 1 Gbp	0.5–6 h/SMRT cell, 1–16 cells/run	Single molecule	Read length highly dependent on input DNA. High cost per sequenced base pair. Single-molecule real-time sequencing with zero-mode waveguide
Roche 454	≤800 bp	≤100,000	18 h	Single end	Withdrawn from the commercial market in 2015/2016. Widely used in studies requiring longer reads. Pyrosequencing.
Thermo Fisher Ion Torrent	≤400 bp	≤80 million/run	A few hours	Single end	Low throughput, low cost per run. Ion semiconductor sequencing

<sup>a</sup>Insert size is the length of two reads plus the distance in base pairs between them.

<sup>b</sup>Runtime is largely dependent on the number of cycles (length of reads).

<sup>c</sup>Read length, runtime, and base calling accuracy are independent according to the manufacturer.

Jensen et al, 2016



# Applicazioni di metodiche di NGS all'analisi degli OGM (Progetti in corso al CROGM)

- ❑ Saggio multiscreening in NGS su piattaforma Illumina, CA5 (progetto in collaborazione con FEM)
- ❖ Sequenziamento in parallelo di un elevato numero di target (37 target di cui 11 geni endogeni, 14 elementi di screening, 8 costrutti, 2 donatori, 2 saggi identificatori di specie)
- ❖ Strategia NGS utilizzata; amplicon sequencing
- ❖ Rilevazione eventi GM di specie vegetali non considerate nella fase di analisi taxon-specifica;
- ❖ Rilevazione eventi GM non autorizzati;
- ❖ Rilevazione presenza nella matrice alimentare di organismi donatori naturali appartenenti a numerose specie (vegetali e non)
- ❖ Identificazione di varianti di sequenza (SNP) correlabili ad uno specifico evento GM



## Esperimento pilota

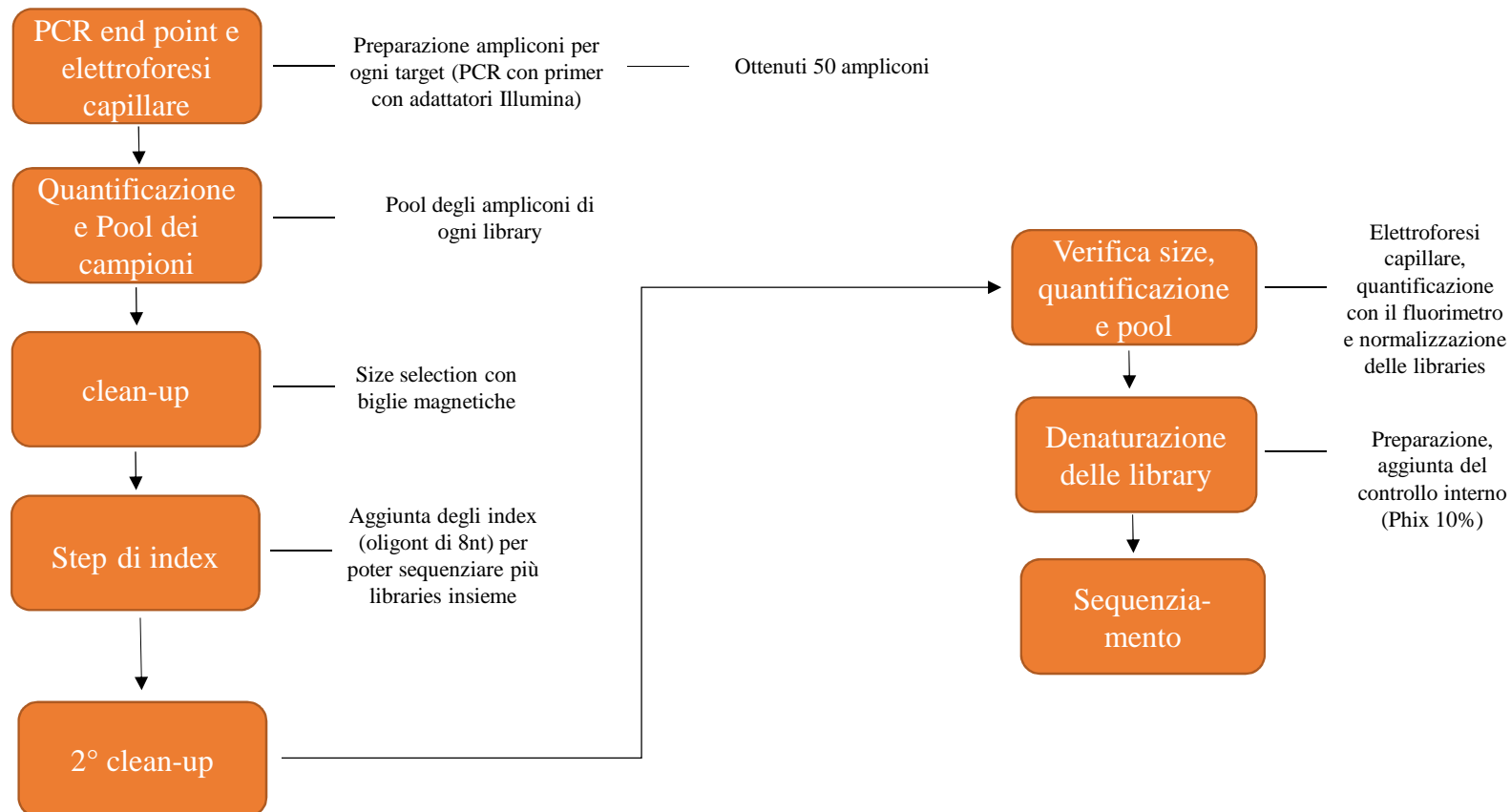
### Step 1. Selezione dei target

- ❖ 5 MDR
- ❖ 3 Campioni di campo
- ❖ 8 Libraries
- ❖ 55 Ampliconi

LIBRARIES	Amplicone/target	Dimensione degli Ampliconi (bp)	Numero ampliconi per library
Cotone MON15985	Acp1	143	5
	NptII	180	
	Cry1Ab/Cry1ac	141	
	23579	261	
	trnI	148-175	
Mais 89034	HMG	146	5
	Tnos	171	
	FMV	145	
	23579	261	
	trnI	148-175	
Soia A5547-127	LEC	141	5
	PAT	135	
	T35s	140	
	23579	261	
	trnI	148-175	
Babrbabietola H7-1	GS	185	7
	CTP-CP4EPS	129	
	CTP2-CP4EPS	243	
	Te9	154	
	FMV	145	
	23579	261	
	trnI	148-175	
Patata EH92-527-1	UGPasi	156	5
	NptII	180	
	JUNCTION P NOS/NPTII	237	
	23579	261	
	trnI	148-175	
20008280 (mangime complementare)	Acp1	143	9
	GS	185	
	Tnos	171	
	CP4EPS	212	
	CTPCP4EPS	129	
	CTP2 CP4EPS	243	
	PAT	135	
	23579	261	
	trnI	148-175	
20008271 (mangime semplice)	HMG	146	8
	Tnos	171	
	CP4EPS	212	
	CTPCP4EPS	129	
	CTP2 CP4EPS	243	
	PAT	135	
	23579	261	
	trnI	148-175	
	HMG	146	
20005309 (mangime complementare)	LEC	141	11
	Acp1	143	
	Tnos	171	
	CP4EPS	212	
	CTPCP4EPS	129	
	CTP2 CP4EPS	243	
	PAT	135	
	BAR	127	
	23579	261	
	trnI	148-175	
	HMG	146	



# Protocollo; preparazione delle libraries

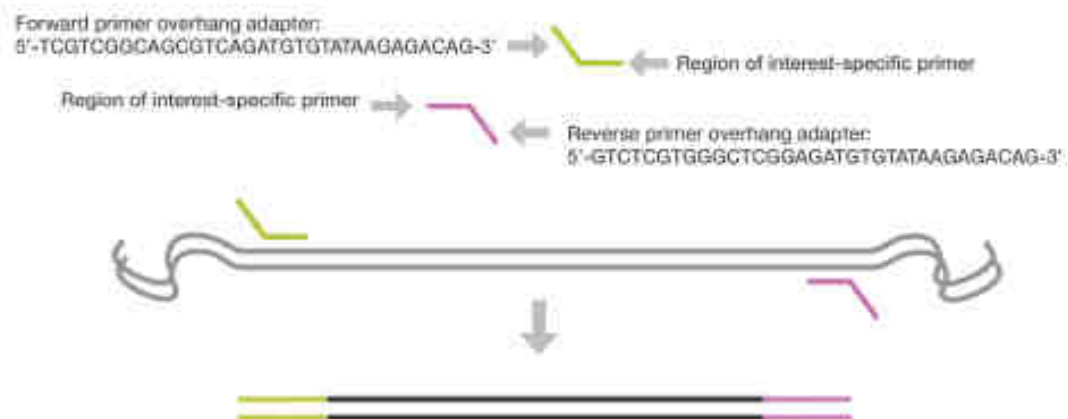


- Gli index consentono di risalire al campione da cui proviene l'amplicone
- Gli adattatori consentono il legame con la flow cell

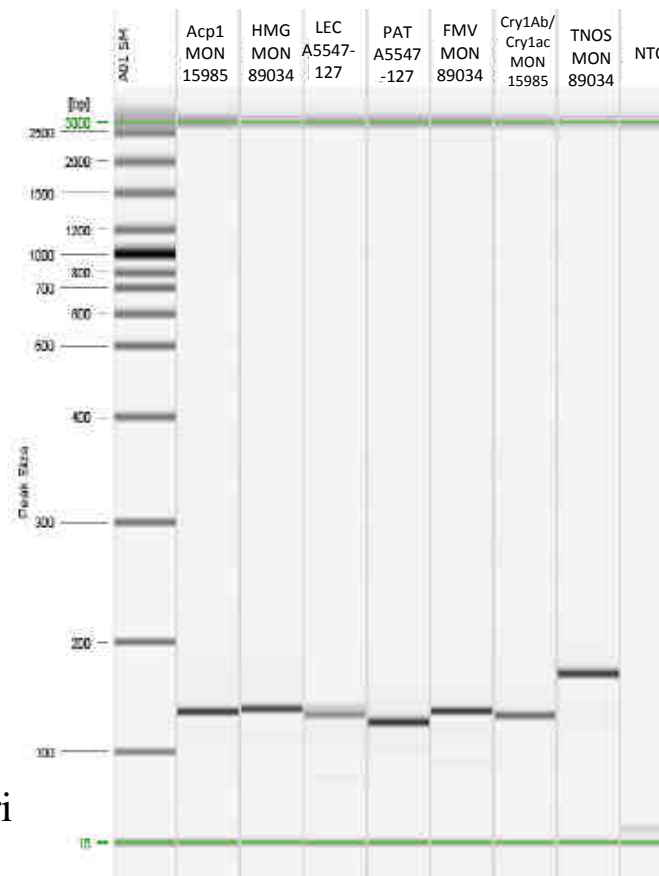


# PCR end-point; preparazione ampliconi

PCR amplify template out of genomic DNA using  
region of interest-specific primers with overhang adapters



Con questo step aggiungiamo alle estremità dell'amplicone gli adattatori  
per i primer di sequenziamento

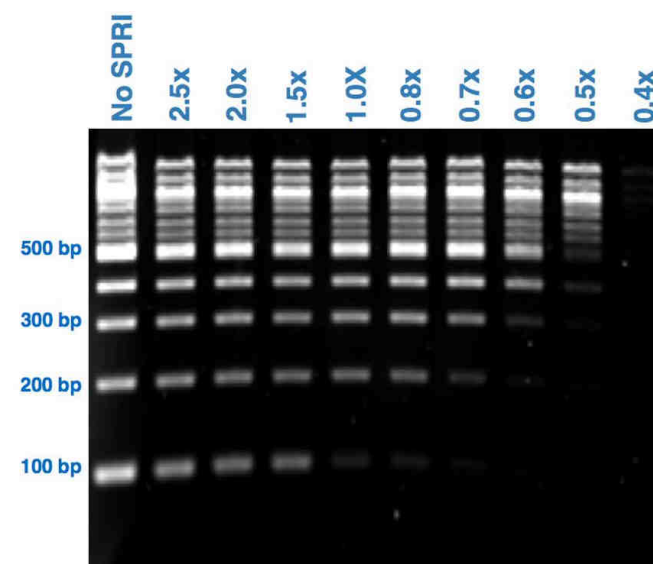
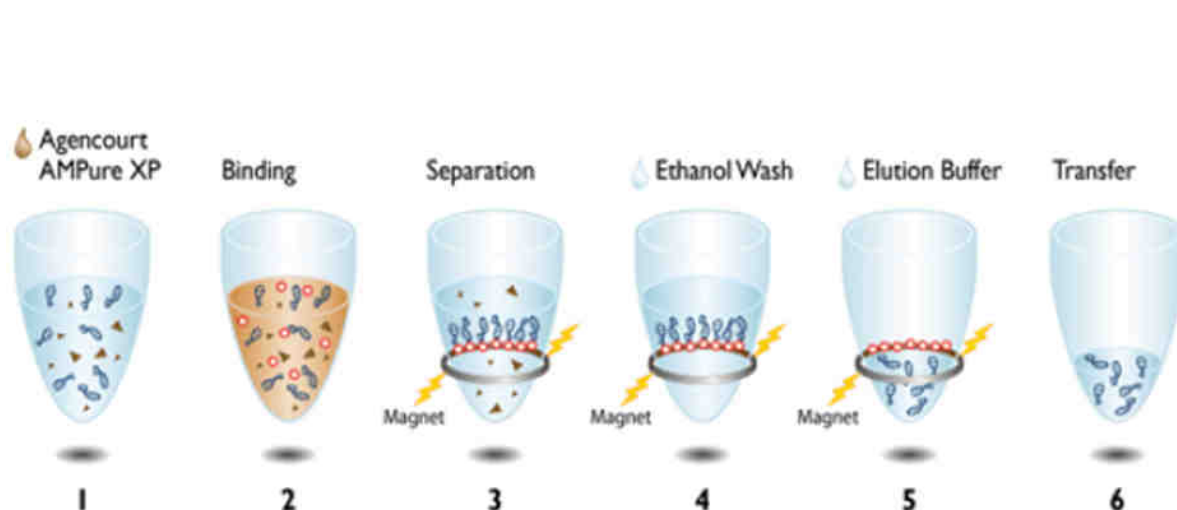


Elettroforesi capillare



# Clean-up con biglie magnetiche

## Solid Phase Reversible Immobilization (SPRI) Methodology

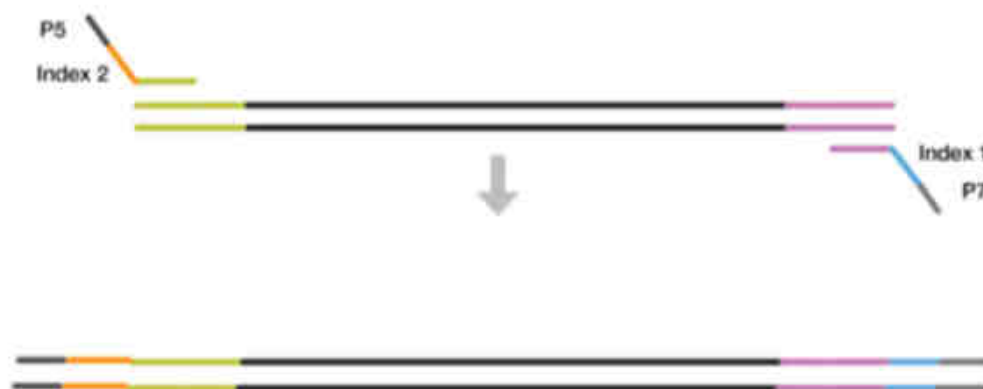
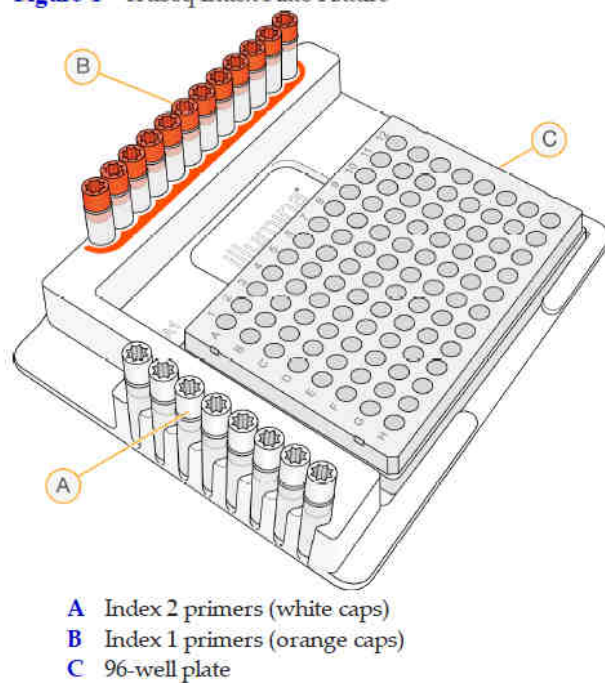


Ratio beads/volume di DNA



## Step PCR-index

**Figure 4** TruSeq Index Plate Fixture



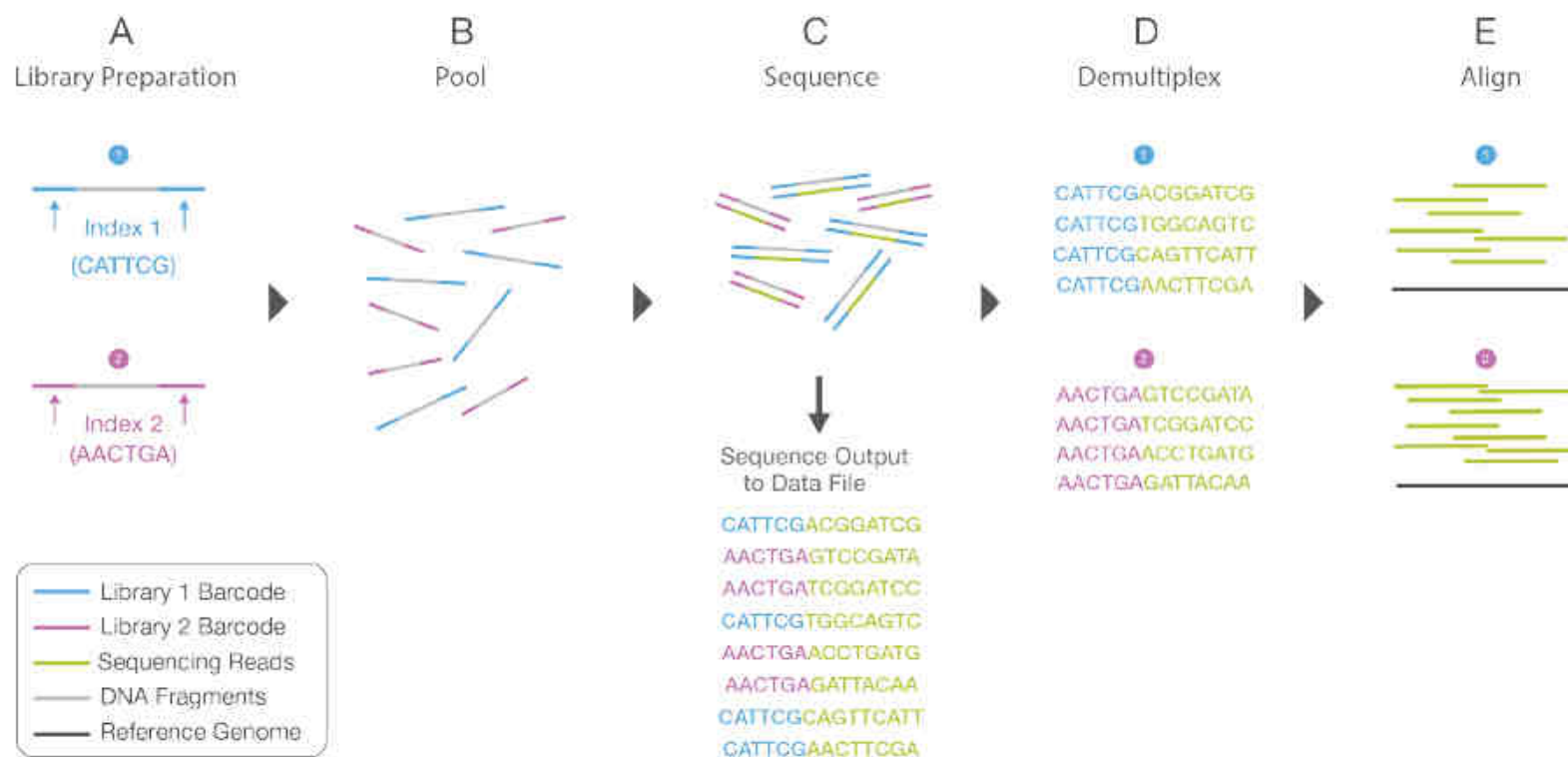
Index 1 (i7)	Sequence	Index 2 (i5)	Sequence
N701	TAAGGCGA	S501	TAGATCGC
N702	CGTACTAG	S502	CTCTCTAT
N703	AGGCAGAA	S503	TATCCTCT
N704	TCCTGAGC	S504	AGAGTAGA
N705	GGACTCCT	S505	GTAAGGAG
N706	TAGGCATG	S506	ACTGCATA
N707	CTCTCTAC	S507	AAGGAGTA
N708	CAGAGAGG	S508	CTAAGCCT
N709	GCTACGCT		
N710	CGAGGCTG		
N711	AAGAGGCA		
N712	GTAGAGGA		

Con questo step aggiungiamo alle estremità dell'amplicone gli indici (index) e gli adattatori per la flow cell (P5 e P7)

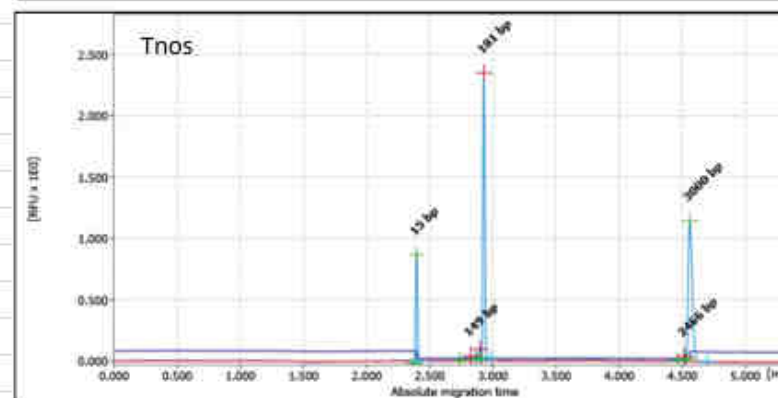
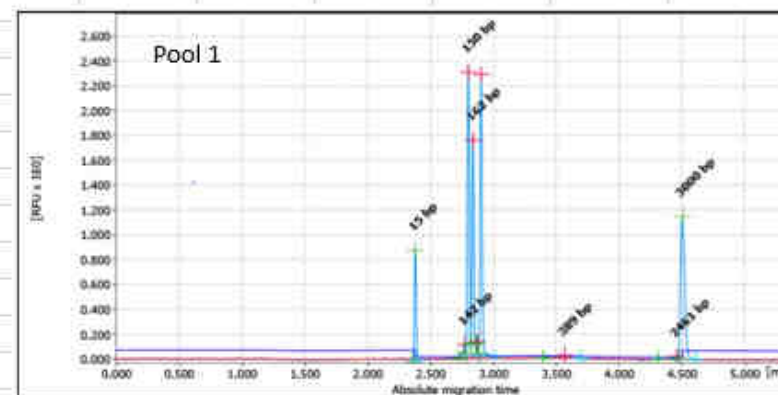
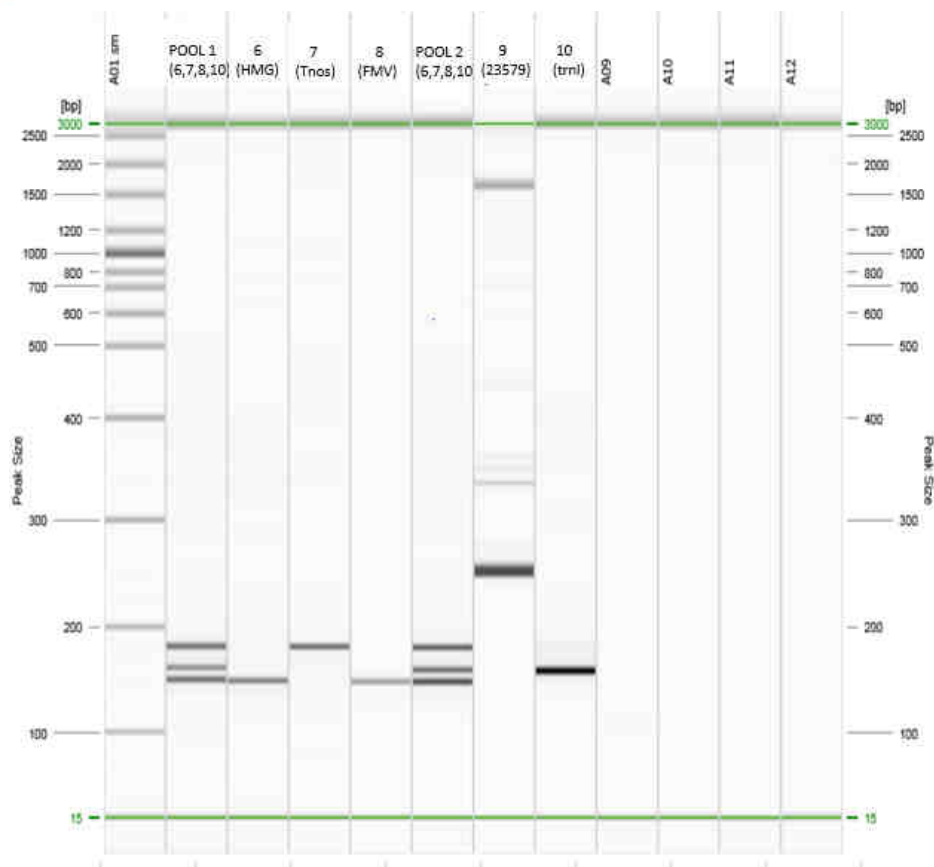




Grazie agli index è possibile analizzare un grande numero di campioni simultaneamente



# Verifica insert size

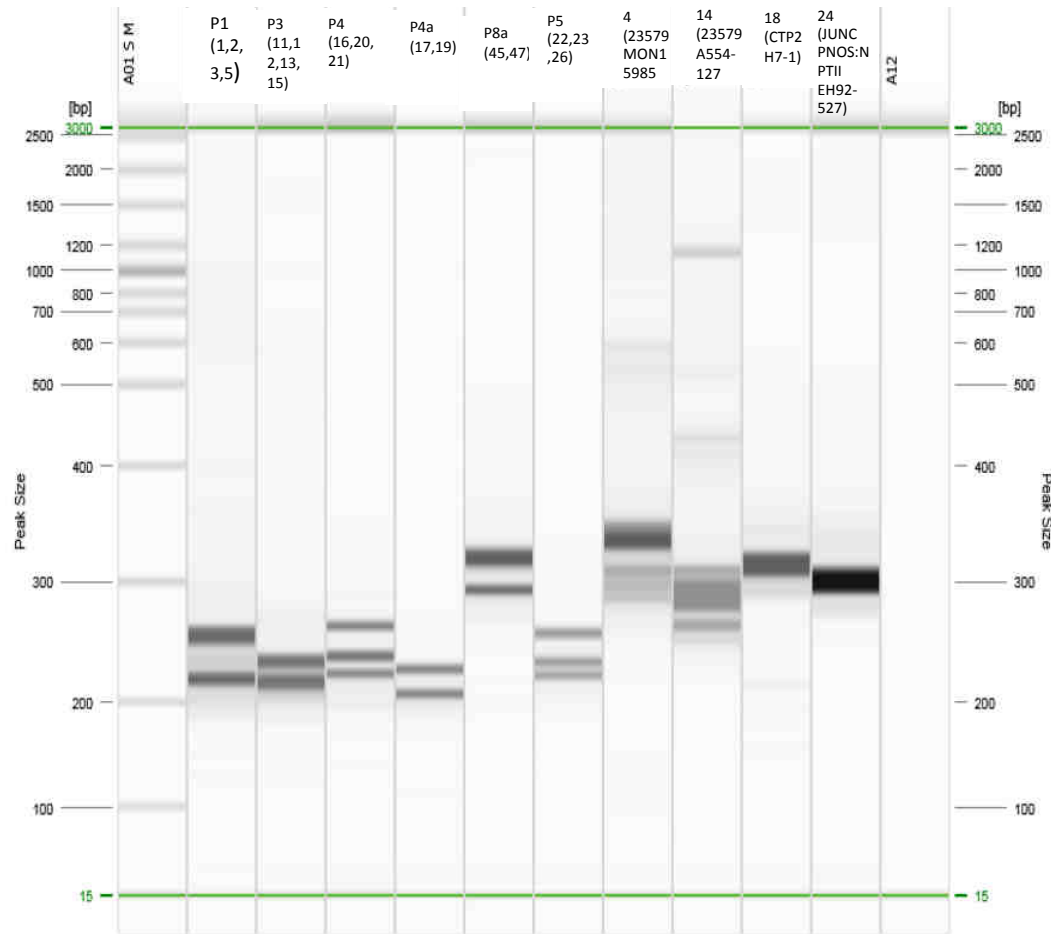


Output elettroforesi capillare

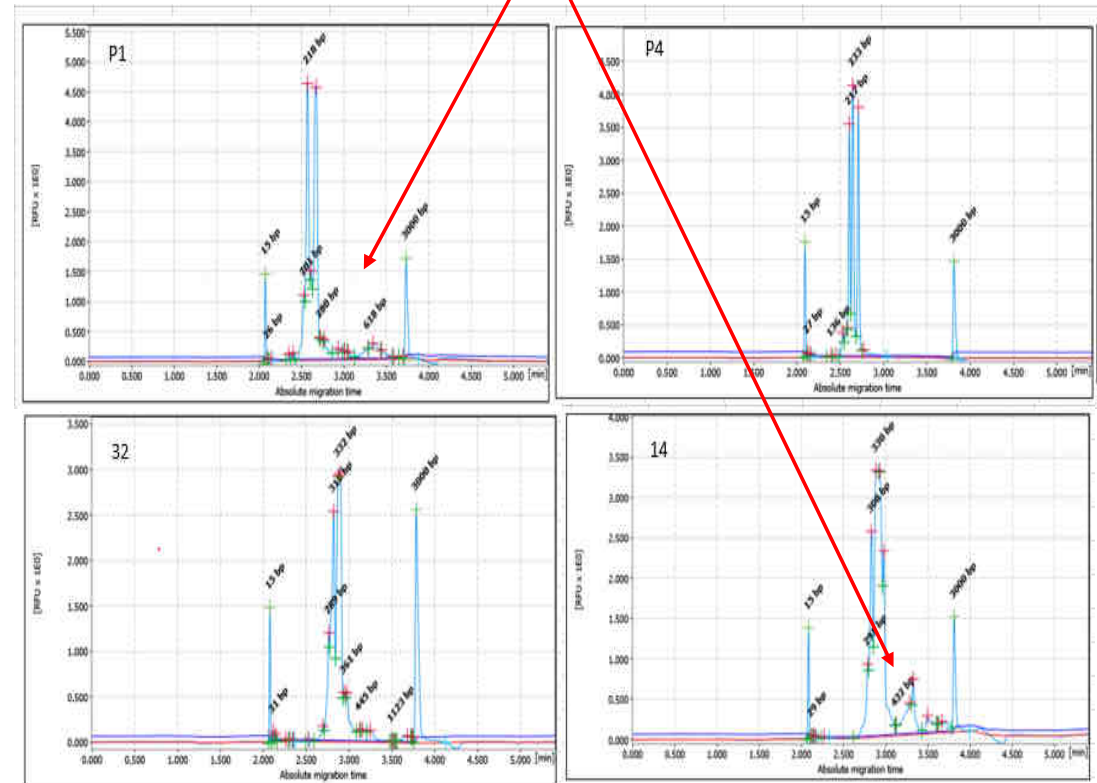




## Verifica insert size (2)



**Contaminazioni, cross-dimer??**



## Normalizzazione e pool delle libraries

$$\frac{(\text{concentration in ng}/\mu\text{l})}{(660 \text{ g/mol} \times \text{average library size})} \times 10^6 = \text{concentration in nM}$$

↓  
dsDNA bp

- Pool libraries
- Denaturazione con NaOH
- Aggiunta controllo positivo Phix (n%)
- Caricamento della library



## ❑ Identificazione di OGM non autorizzati su piattaforme di terza generazione

### Nanopore sequencing technology: a new route for the fast detection of unauthorized GMO

[Marie-Alice Fraiture](#),<sup>#1,2</sup> [Assia Saltykova](#),<sup>#1,3</sup> [Stefan Hoffman](#),<sup>1</sup> [Raf Winand](#),<sup>1</sup> [Dieter Deforce](#),<sup>4</sup> [Kevin Vanneste](#),<sup>1</sup> [Sigrid C. J. De Keersmaecker](#),<sup>1</sup> and [Nancy H. C. Roosens](#)<sup>1</sup>

- OGM non autorizzato; riso Bt
- Cassetta transgenica caratterizzata nel precedente lavoro (siti di inserzione nel cromosoma 2 e 3)
- Disegno di primer specifici per gli elementi di screening rilevati in real-time PCR (target-specifici), P35s,tNOS e t35S pCAMBIA
- DNA walking; - Strategia di pre-arricchimento (Site-Finding PCR)
  - Preparazione library
  - Sequenziamento
- Piattaforma NGS; Oxford Nanopore



**T5**

Tramite DNA walking e sequenziamento andiamo a caratterizzare le regioni (UNKOWN) che fiancheggiano il t-NOS

TofaniS; 15/09/2020

## Strategia di pre-arricchimento (Site-Finding PCR)

Methods	Oligonucleotide names	Oligonucleotide sequences
DNA Walking	p35S-F a	GGGTCTTGCGAAGGATAGTG
	p35S-F b	TGTGCGTCATCCCTTACGTCAGT
	p35S-F c	TATCACATCAATCCACTTGCTTT
DNA Walking	p35S-R a	AAAGCAAGTGGAATTGATGTGATA
	p35S-R b	ACTGACGTAAGGGATGACGCACA
	p35S-R c	CACTATCCTTCGCAAGACCC
DNA Walking	tNOS-F a	GATTAGAGTCCCGCAATTATACATTTAA
	tNOS-F b	TTAATACGCGATAGAAAACAAAAT
	tNOS-F c	AAATATAGCGCGCAAMCTAGGATAA
DNA Walking	tNOS-R a	TTATCCTAGKTTGCGCGCTATATTT
	tNOS-R b	ATTTTGTTTTCTATCGCGTATTAA
	tNOS-R c	TTAAATGTATAATTGCGGGACTCTAATC
DNA Walking	t35S pCAMBIA-F a	CGGGGGATCTGGATTTTAGTA
	t35S pCAMBIA-F b	GGGTTTCTTATATGCTCAACAC
	t35S pCAMBIA-F c	GAGCGAAACCTATAGGAACCCCT
DNA Walking	t35S pCAMBIA-R a	TACTAAAATCCAGATCCCCCG
	t35S pCAMBIA-R b	GTGTTGAGCATATAAGAAACCC
	t35S pCAMBIA-R c	AGGGTTCCTATAGGGTTTCGCTC

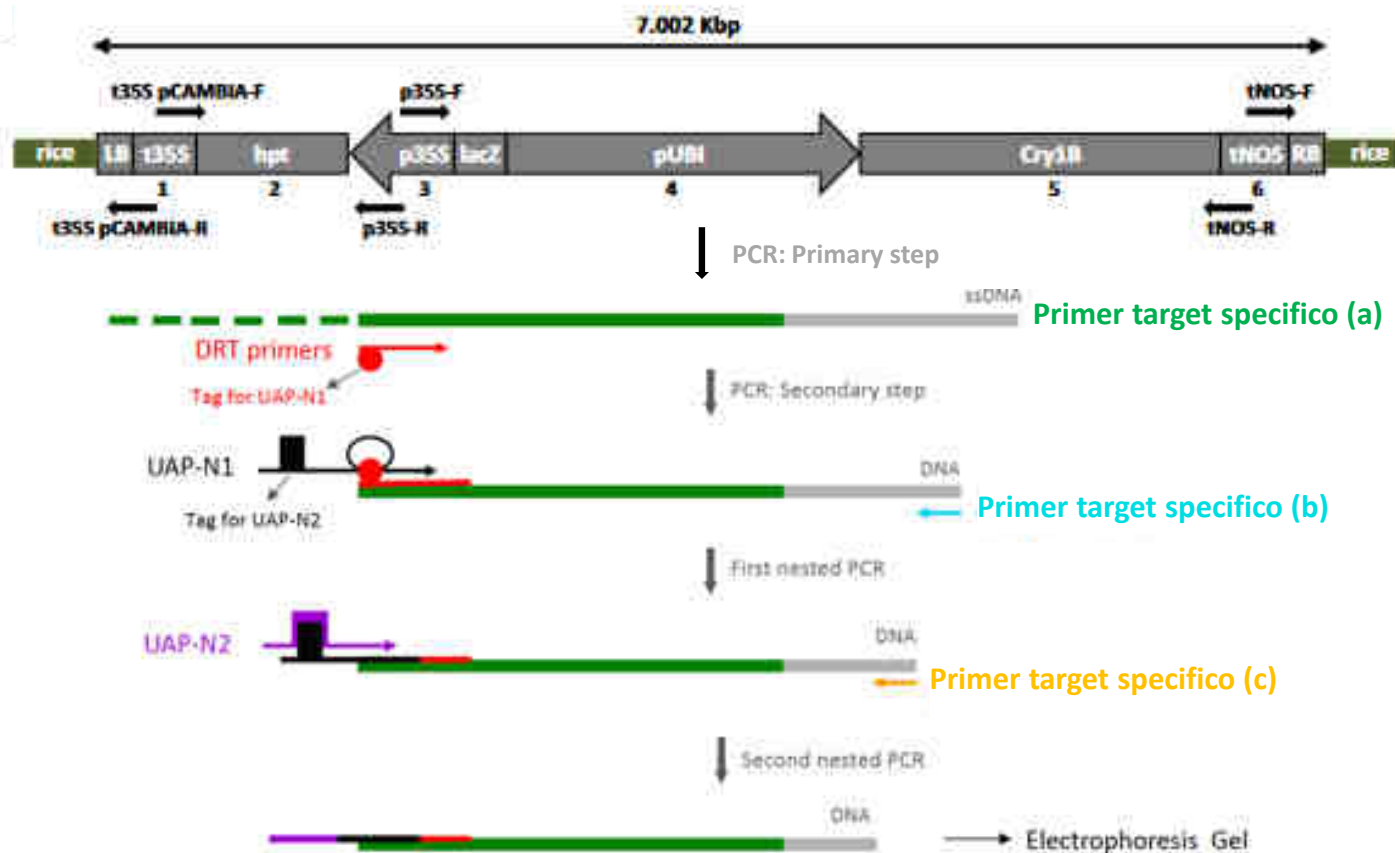
DRT primers	Sequence
DRT primers A	5'-GAACACGCGTCGTTACCTCCXXXXGXXXXTAGT-3'
DRT primers B	5'-GAACACGCGTCGTTACCTCCXXXXGXXXXTAGT-3'
DRT primers C	5'-GAACACGCGTCGTTACCTCCXXXXGXXXXCAT-3'
DRT primers D	5'-GAACACGCGTCGTTACCTCCXXXXGXXXXCCTG-3'
UAP-N1	5'-AGTCGGGAAGCAGTGGTATCAACGCAGAGTGGCCATTACGGCCGAACACGCGTCGTTACCT-3'
UAP-N2	5'-CCUGGAAGCAGTGGTATCAACG-3'

- Disegno di 3 primer target-specifici per ogni elementi di screening rilevati in rtPCR (es: p35s, tNOS, t-35S pCAMBIA)
- PCR con una combinazione di primer degenerati (DRT) e primer target-specifici per amplificare la regione di interesse
- 2 PCR semi-nested con una combinazione di primer target-specifici e primer universali (UAP) per aumentare la specificità e diminuire gli aspecifici
- Verifica degli ampliconi tramite elettroforesi su gel o su strumenti per l'elettroforesi automatizzata





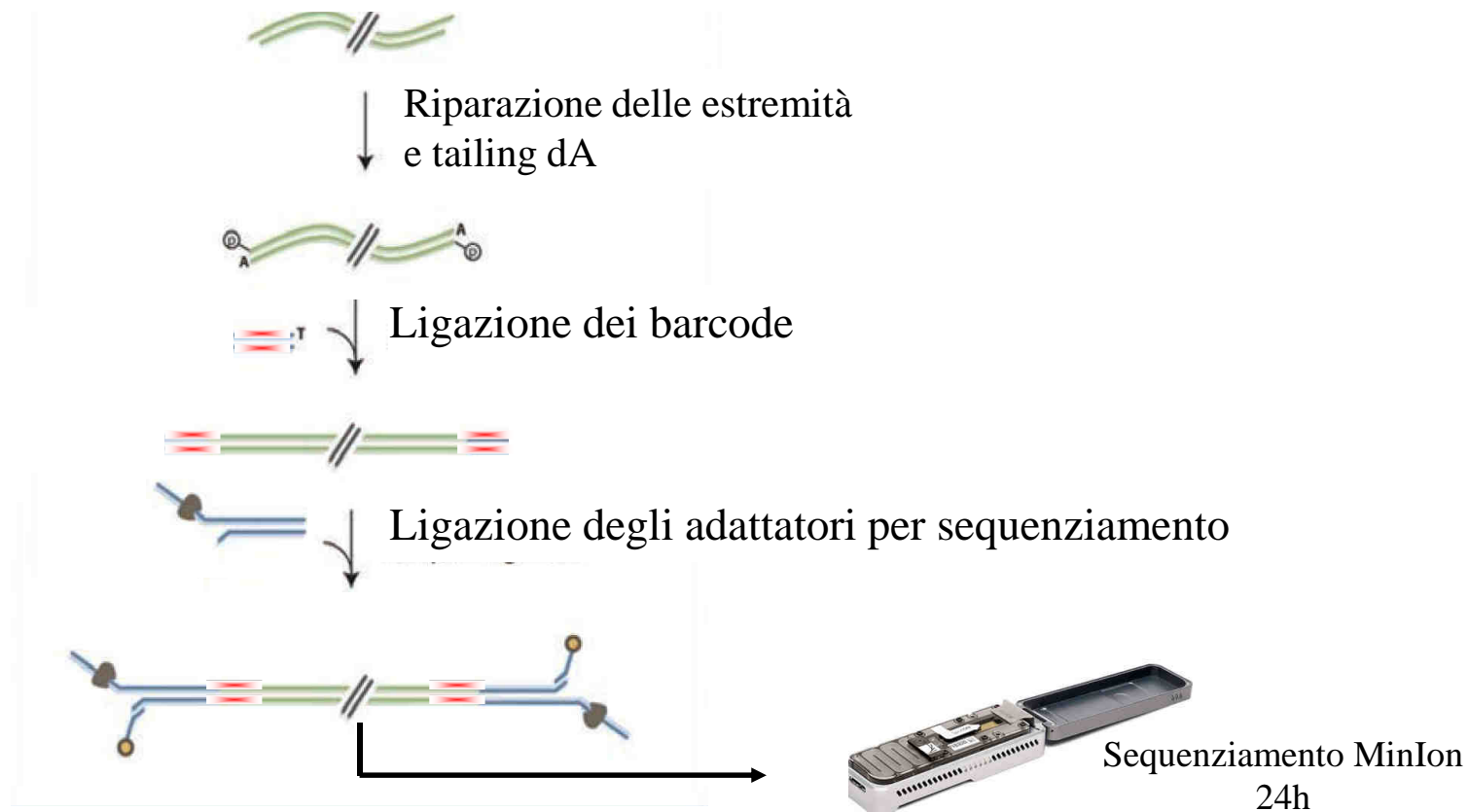
# Cassetta transgenica del riso GM DNA walking





# Preparazione della library

## Pool di ampliconi



# Microrganismi Geneticamente Modificati (MOGM)

## ☐ Definizione e campo di applicazione

### Direttiva 2009/41CE

**MOGM:** un microrganismo il cui materiale genetico è stato modificato in un modo non naturale mediante moltiplicazione e/o ricombinazione naturale;

**IMPIEGO CONFINATO:** ogni attività nella quale i microrganismi sono modificati geneticamente o nella quale tali MGM sono messi in coltura, conservati, trasportati, distrutti, smaltiti o altrimenti utilizzati, e per la quale vengono usate misure specifiche di contenimento al fine di limitare il contatto degli stessi con la popolazione e con l'ambiente e per garantire a questi ultimi un livello elevato di sicurezza

Utilizzati nei processi industriali (panificazione, produzione della birra ecc) per produrre enzimi, additivi.



# Enzimi alimentari utilizzati nell'industria alimentare






## □ Definizioni

**Reg 1331/2008, 1332/2008, 1333/2008**

**ENZIMA ALIMENTARE:** Un prodotto ottenuto da vegetali, animali o microrganismi o ottenuto mediante fermentazione utilizzando microrganismi:

- i) contenente uno o più enzimi in grado di catalizzare una specifica reazione biochimica; e
- ii) aggiunto ad alimenti per uno scopo tecnologico in una qualsiasi fase di fabbricazione, trasformazione, preparazione, trattamento, imballaggio, trasporto o conservazione degli stessi

## Enzymes used in food industries

Dairy production	Brewing	Baking	Wine and fruit juice	Meat
Rennet	$\beta$ -Glucanase	Maltogenic amylase	Pectinase	Protease
Lactase	$\alpha$ -Amylase	Glucose oxidase	$\beta$ -Glucanase	Papain
Protease	Protease	Pentosenase		
Catalases 	Amyloglucosidase 			



## ☐ Criteri per l'autorizzazione

**Reg EC 1829/2003, 1830/2003**

- ❖ EFSA deve valutare la sicurezza alimentare
- ❖ Il dossier riguardante il MOGM utilizzato per la produzione dell'enzima è confidenziale -> No accesso alle sequenze
- ❖ Linee guida dell' EFSA;
  - Assenza del ceppo produttore;
  - Assenza del DNA del MOGM (<10 ng di DNA per grammo o ml di prodotto; lunghezza dell'amplicone/ AMR gene)



- Nel 2014,2018,2019,2020 contaminazioni da MOGM in additivi alimentari (vitamina B2, enzimi) destinati all'alimentazione animale
- Bollettini RASFF (2014.1249, 2014.1360, 2014.1657, 2018.2755, 2019.0793, 2019.3332, 2019.3216, 2020.2579)
- Presenza di marcatori di selezione (geni per la resistenza agli antibiotici (canamicina, neomicina, cloramfenicolo, tetraciclina))
- Pericolo per la salute umana

**RASFF Portal**

European Commission > RASFF Portal

**Notification details - 2014.1249**

unauthorised genetically modified (Bacillus subtilis) bacteria in vitamin B2 from China, via Germany

Reference:	2014.1249	Notification type:	feed - information for follow-up - company's own check
Notification date:	05/09/2014	Action taken:	detained by operator
Last update:	21/09/2016	Distribution status:	distribution to other member countries
Notification from:	Germany (DE)	Product:	bacteria in vitamin B2
Classification:	information for follow-up	Product category:	feed additives
Risk decision:	undecided	Published in RASFF Consumers' Portal:	has never been published

**RASFF Portal**

European Commission > RASFF Portal

**Notification details - 2019.3332**

unauthorised genetically modified bacteria (Bacillus velezensis) in food enzyme from China

Reference:	2019.3332	Notification type:	food - alert - monitoring of media
Notification date:	20/09/2019	Action taken:	withdrawal from recipient(s)
Last update:	14/05/2020	Distribution status:	distribution to other member countries
Notification from:	Belgium (BE)	Product:	food enzyme
Classification:	alert	Product category:	food additives and flavourings
Risk decision:	undecided	Published in RASFF Consumers' Portal:	has never been published

**RASFF Portal**

European Commission > RASFF Portal

**Notification details - 2020.2579**

unauthorised genetically modified micro-organism DNA in food enzymes from the United Kingdom, dispatched from New Zealand, with raw material from China

Reference:	2020.2579	Notification type:	food - information for follow-up - monitoring of media
Notification date:	24/06/2020	Action taken:	
Last update:	20/08/2020	Distribution status:	product traded online
Notification from:	Belgium (BE)	Product:	food enzymes
Classification:	information for follow-up	Product category:	food additives and flavourings
Risk decision:	undecided	Published in RASFF Consumers' Portal:	has never been published



GM *B.subtilis* nella vitamina B2 80%  
(additivo per mangimi)

GM *B.velezensis* in un enzima alimentare  
(proteasi)

GM *B.licheniformis* in un enzima  
alimentare (alfa-amilasi)





# Problemi legati alle contaminazioni accidentali di MOGM

→ Presenza di geni per l'antibiotico resistenza (AMR genes) come marker di selezione



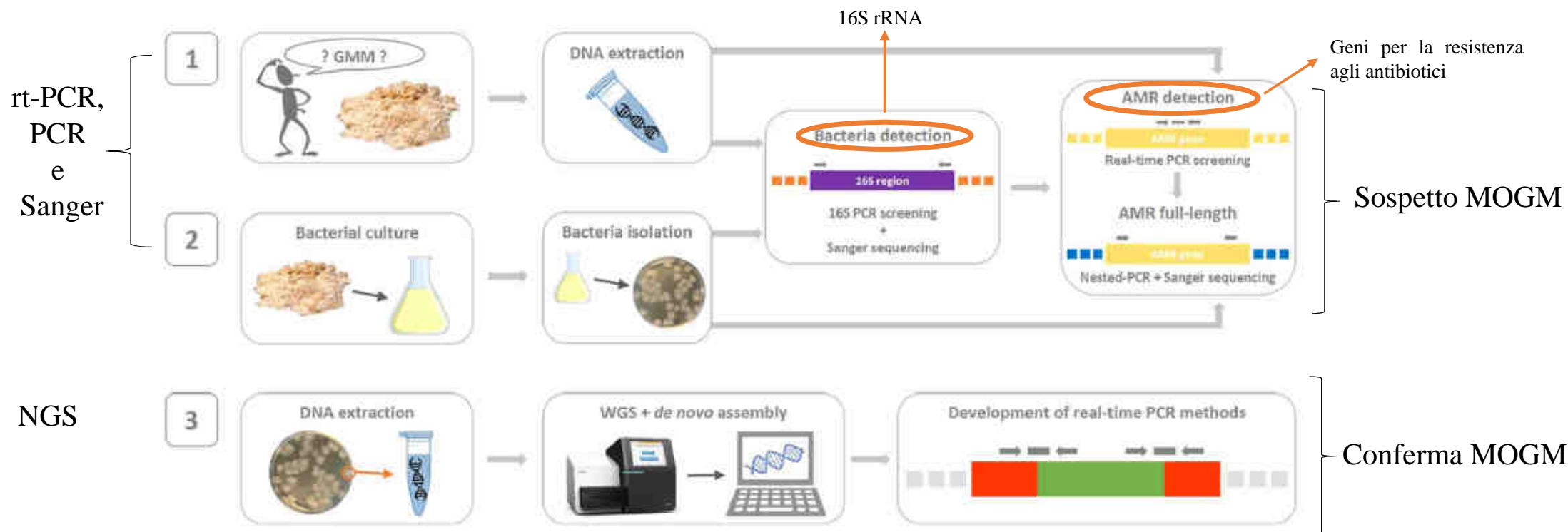
- Trasferimento orizzontale dell'antibiotico resistenza al microbiota intestinale attraverso l'ingestione di prodotti contaminati

Sviluppo di una strategia per ricercare MOGM e determinare la presenza di AMR genes





# Strategia



Fraiture et al,2020, «Identification of an unauthorized genetically modified bacteria in food enzyme through whole genome sequencing». Sci Rep, 10:7094.



## Quali informazioni ricavo dalla PCR e dal sequenziamento?

### ID del MOGM

- 1) Tecniche di PCR → Genere
- 2) PCR/Sanger → Geni AMR
- 3) NGS → Conferma MOGM

### ID del MOGM (RASFF 2019.3332)

- 1) *Bacillus*
- 2) Kanamicina
- 3) *Bacillus velezensis*  
Pub110  
Kan R Bleo R



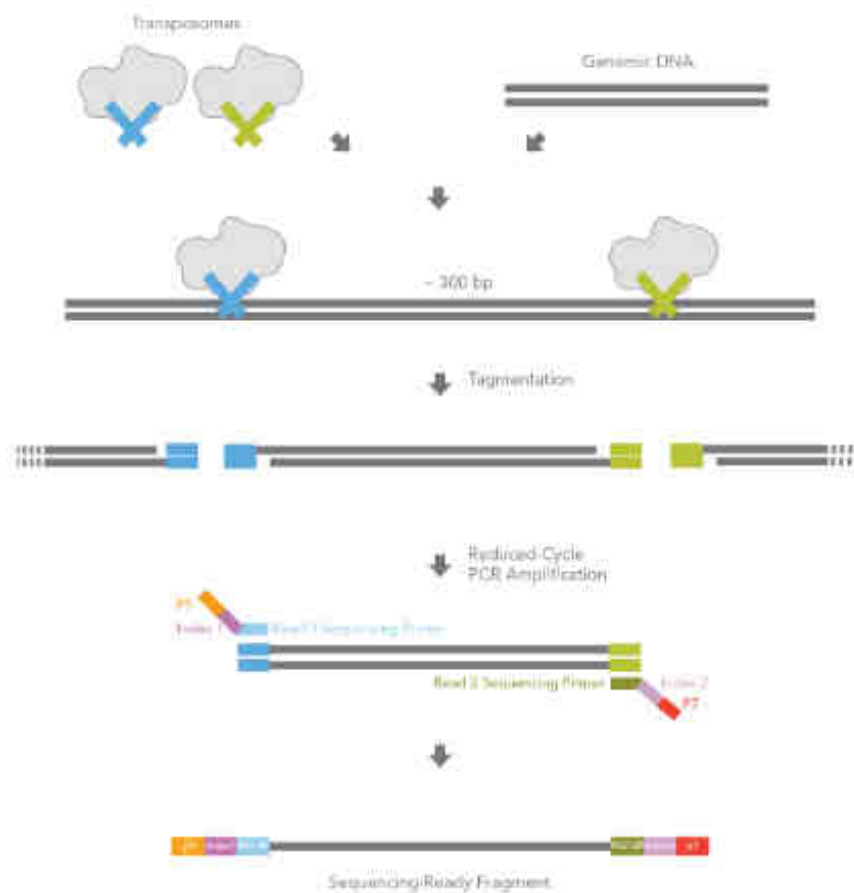


Istituto Zooprofilattico Sperimentale  
del Lazio e della Toscana *M. Aleandri*

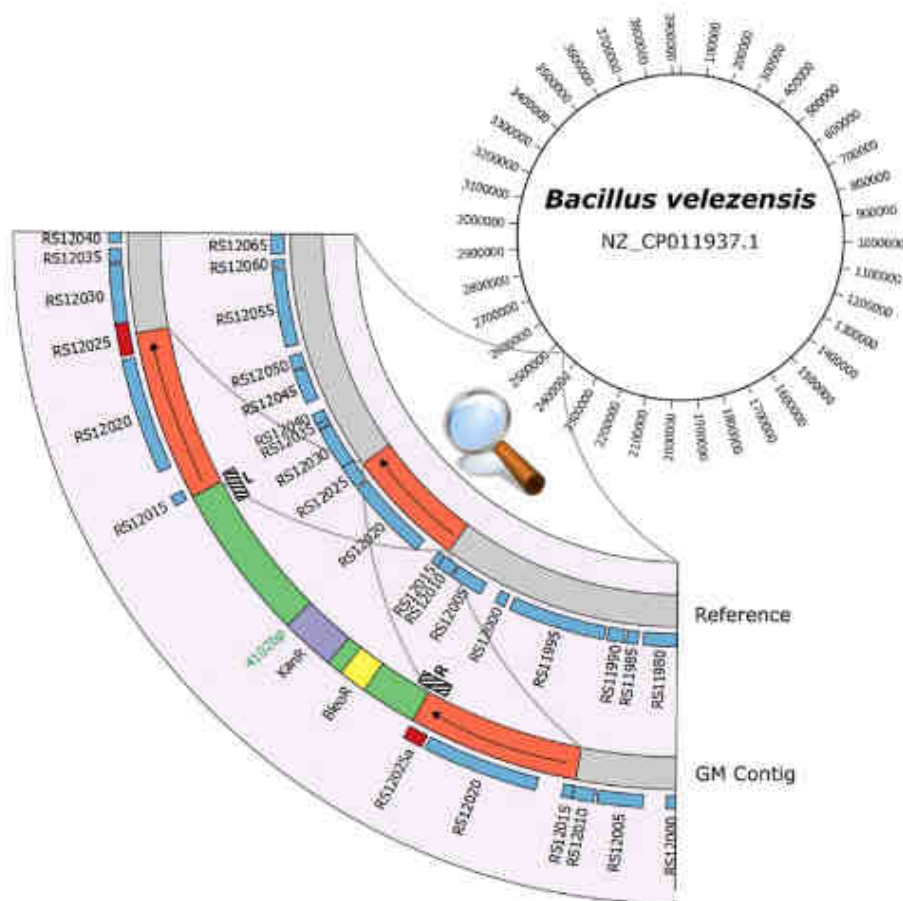
# Workflow WGS



Centro di Riferenza Nazionale  
per la Ricerca di OGM



# Cassetta transgenica caratterizzata mediante NGS



- PUB110 nel gene codificante per la proteasi
- Presenza di AMR genes (KanR e BleoR)

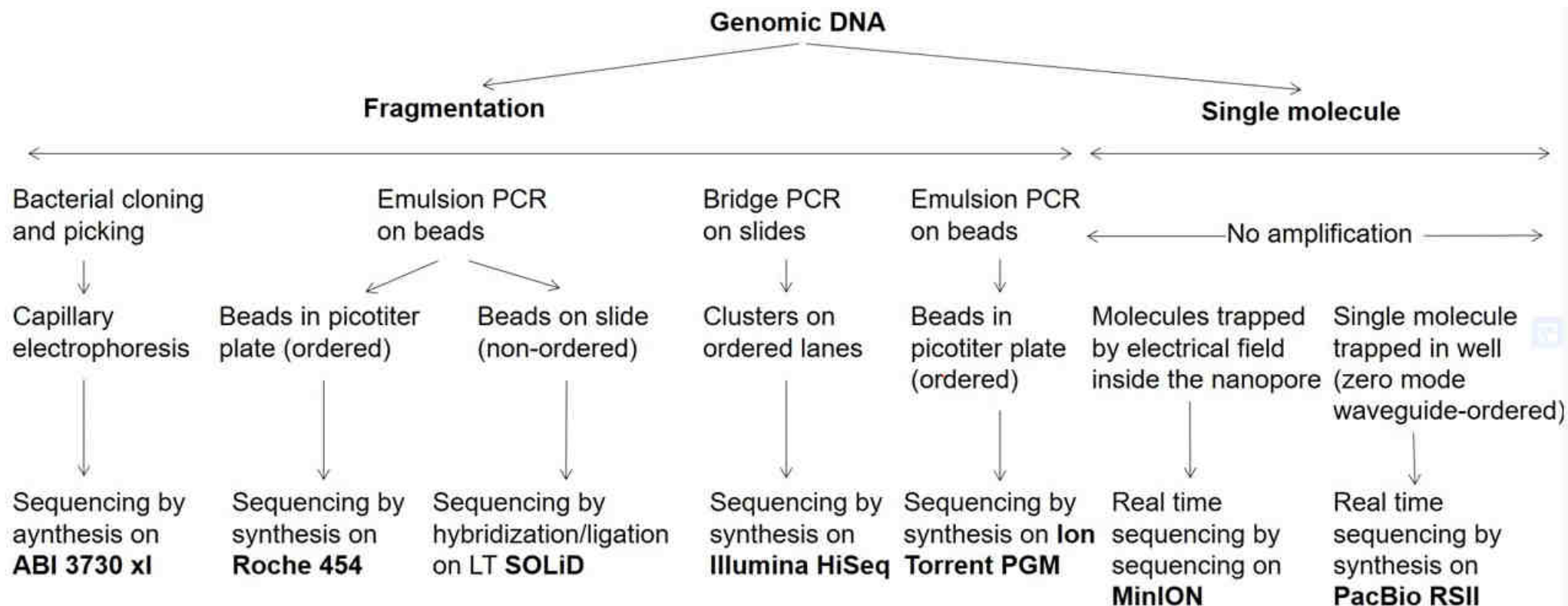


**MOGM NON AUTORIZZATO**





# Ricapitolando per concludere...



**Grazie per non esservi!  
Addormentati!!!**

