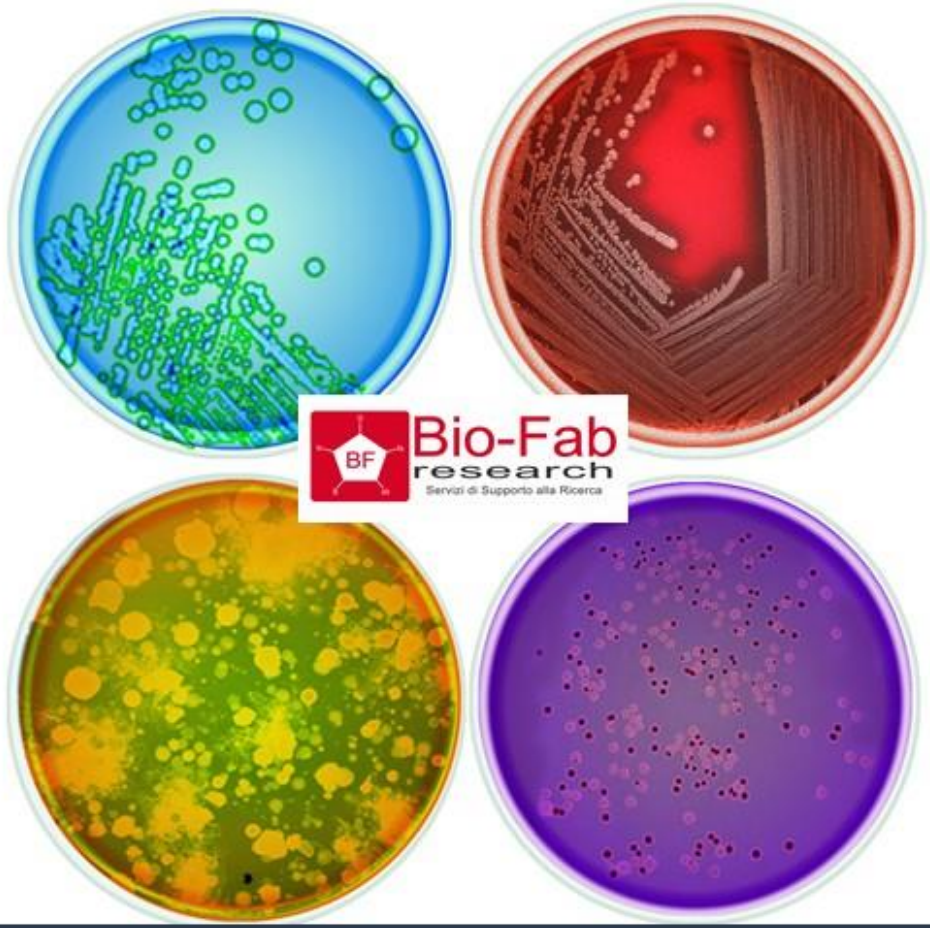
Bio-Fab research
Servizi di Supporto alla Ricerca

Small Genome Sequencing
Servizio di NGS BioFab-ISS



Small Genome Sequencing
Servizio di NGS BioFab-ISS
21 gennaio 2019 aula Bovet ore 10.00

- Introduzione al sequenziamento di interi genomi
- Next generation sequencing: descrizione della tecnologia
- Small genome sequencing: descrizione dell'applicazione
- Strategie di esperimento: reads, coverage e output, SR e PE sequencing
- Quantità e qualità del materiale di partenza
- Preparazione delle librerie, controlli qualitativi
- Invio dei risultati: dati grezzi e possibili analisi
- Analisi bioinformatica: standard e avanzata
- Servizio di supporto all'analisi bioinformatica
- Organizzazione di corse periodiche per ISS

- ✓ Introduzione al sequenziamento di interi genomi
- ✓ Next generation sequencing: descrizione della tecnologia
- ✓ Costi e benefici (corse periodiche)

Bio-Fab research
Servizi di Supporto alla Ricerca

- ✓ Small genome sequencing: descrizione dell'applicazione
- ✓ Strategie di esperimento: reads, coverage, output, SR e PE sequencing
- ✓ Quantità e qualità del materiale di partenza
- ✓ Preparazione delle librerie, controlli qualitativi

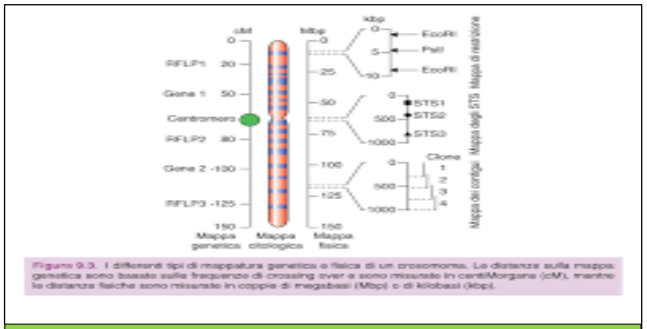
Bio-Fab research
Servizi di Supporto alla Ricerca

- Invio dei risultati: dati grezzi e possibili analisi
- Analisi bioinformatica: Standard, avanzata
- Servizio di supporto all'analisi bioinformatica

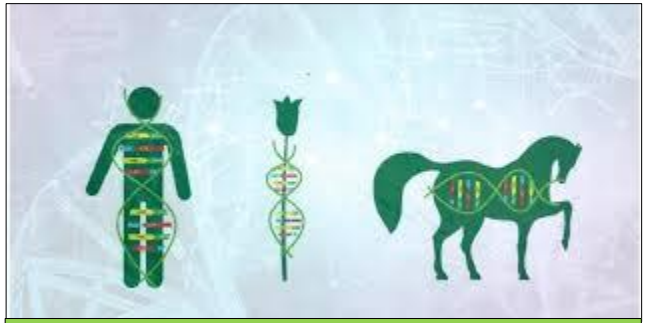
Bio-Fab research
Servizi di Supporto alla Ricerca

Sequenziamento dei genomi: obbiettivi della genomica

Identificazione di tutti i geni e delle altre sequenze significative



Costruire mappe genetiche e fisiche



Confronto tra genomi di specie diverse (evoluzione)

Produzione di un data base per l'accesso alle informazioni

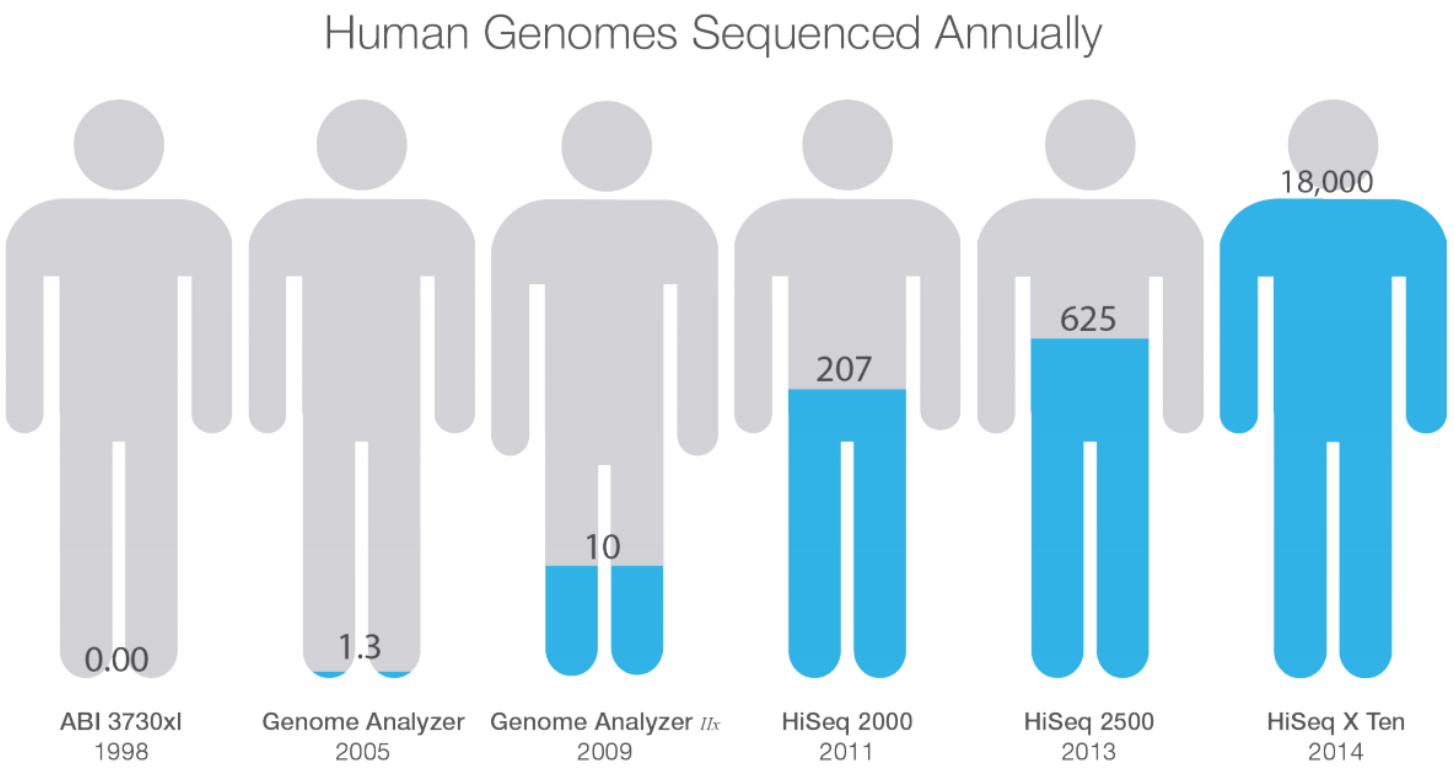
Small genome?

- plasmidi, megaplasmidi
- mitocondriale
- adenovirus
- fagi, cosmidi
- virus
- batteri
- Lieviti
- microfunghi, funghi

Dimensioni?

- Da 600bp a 100Mbp

Evoluzione del sequenziamento



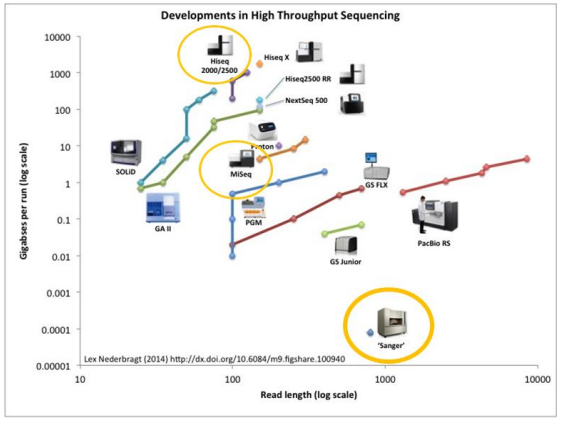
- Per sequenziare il primo genoma umano con i sequenziatori a capillari ci sono voluti 15 anni
- Con HiSeq X Ten si possono sequenziare 45 genomi umani in un giorno

Bio-Fab research
Servizi di Supporto alla Ricerca

Velocità e Throughput

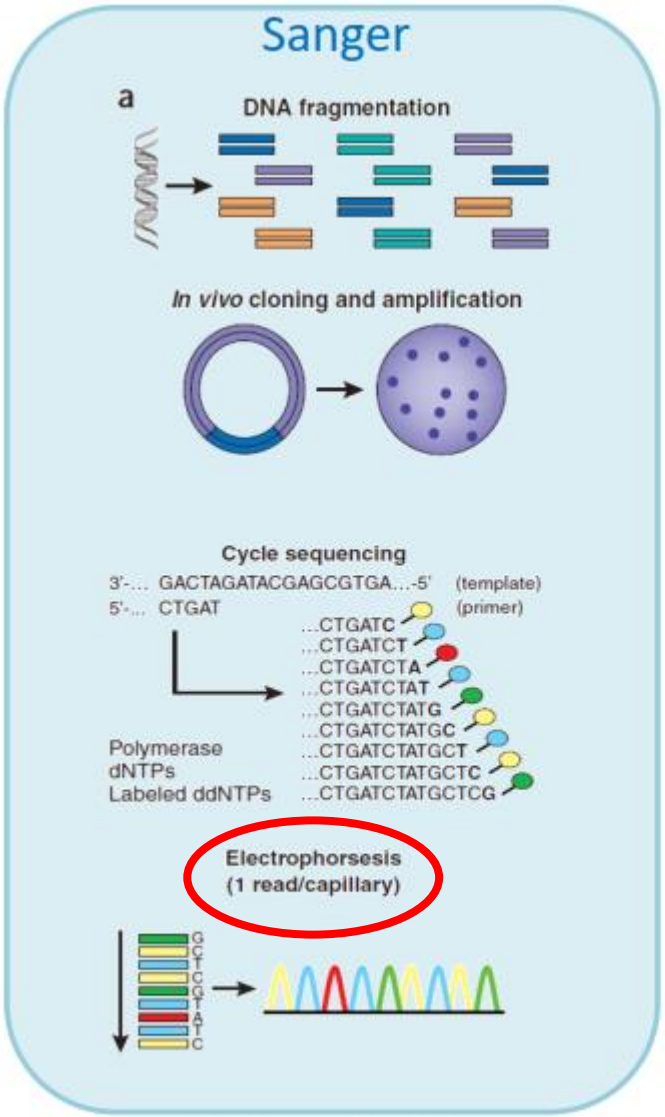
Lo sviluppo della tecnologia ha accelerato enormemente la ricerca e lo studio dei geni

- Sanger 1000 bp X 96 reazioni contemporanee X 10 corse al giorno= 960.000 (0,96 Mb) al giorno
- MiSeq 2X300 bp X 25 milioni di sequenze. Una corsa 15 Gb (PE 30Gb)
- HiSeq 2X150 bp X 5 bilioni di sequenze .Una corsa 1500 Gb



Bio-Fab research
Servizi di Supporto alla Ricerca

Differenza tra il sequenziamento tradizionale e NGS

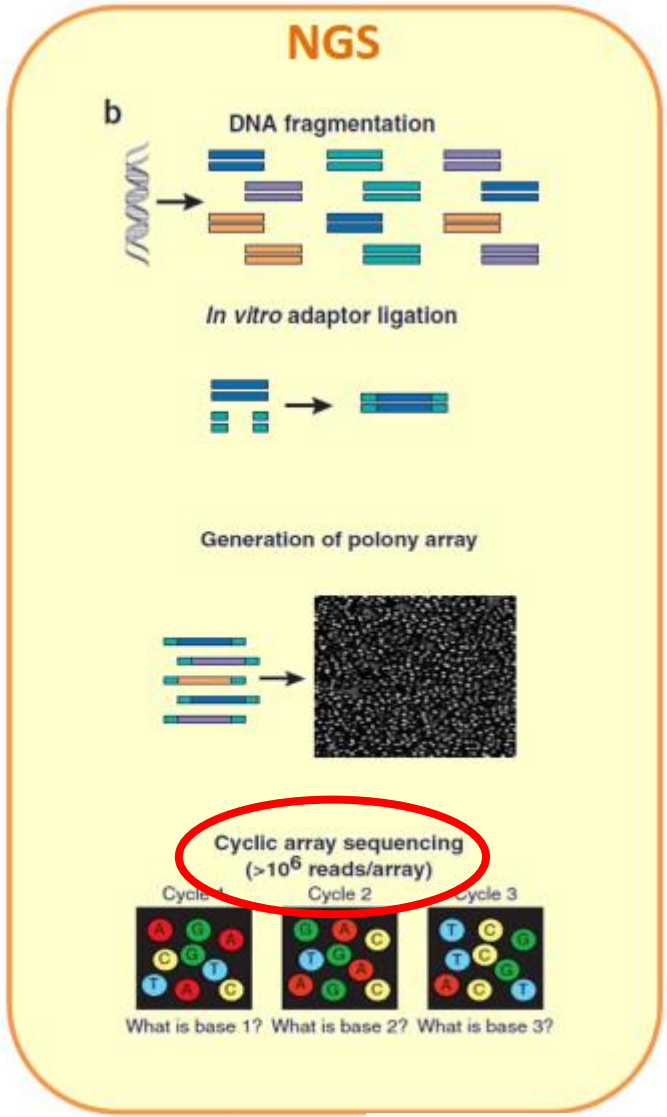


1

Preparazione dei campioni più semplice e veloce per NGS (no clonaggio richiesto)

2

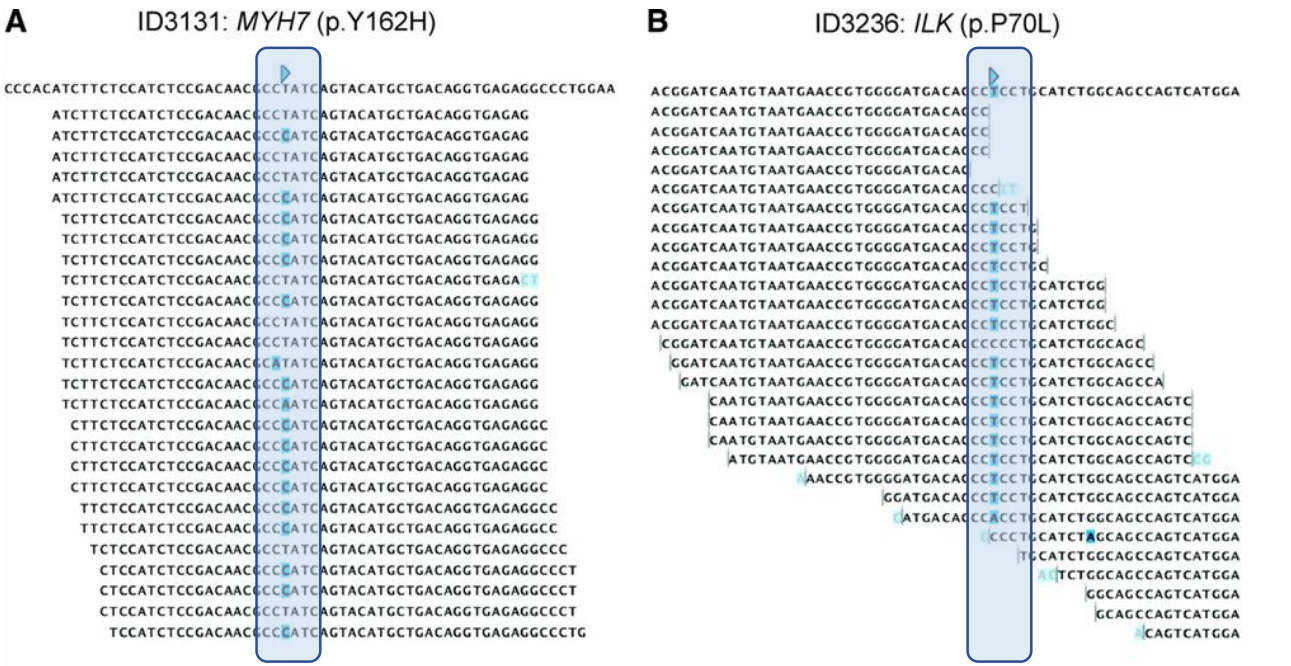
NGS consente una **elevata parallelizzazione** (fino a centinaia di milioni di reazioni di sequenza in parallelo)



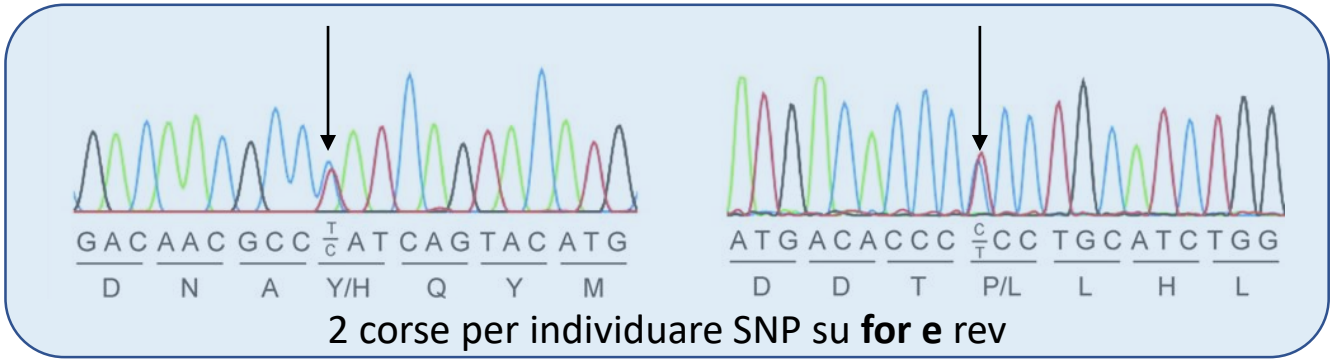
Shendure, J., & Ji, H. (2008). Next-generation DNA sequencing. Nature biotechnology, 26(10), 1135–4

Parallelizzazione = Maggiore accuratezza

NGS



Sanger



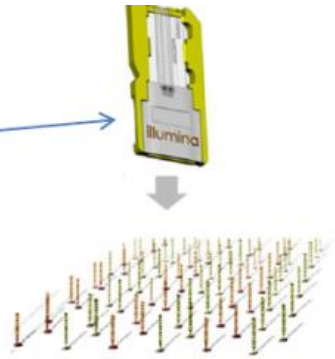
2 corse per individuare SNP su **for e** rev

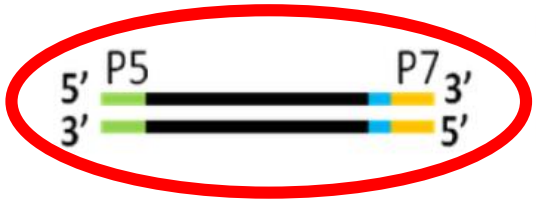
Bio-Fab research
Servizi di Supporto alla Ricerca

Come funziona?

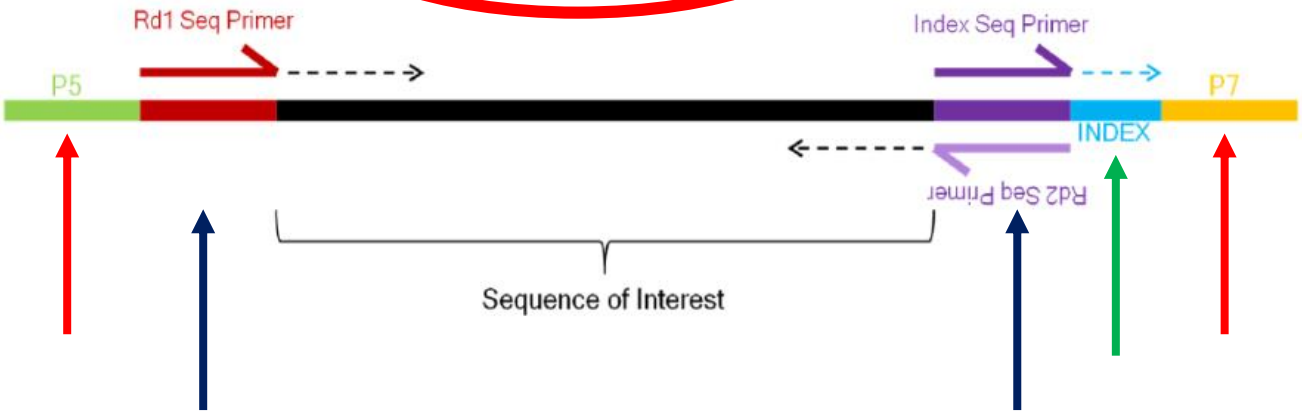
Illumina adaptors

(gli adattatori servono a legare il frammento alla cella dove avviene l'amplificazione e il sequenziamento (P5/P7) e fungono da primer per la reazione di amplificazione prima e di sequenziamento poi)
(Rd1 e Rd2 servono per il paired-end sequencing = sequenziamento a partire da entrambe le estremità di un frammento)

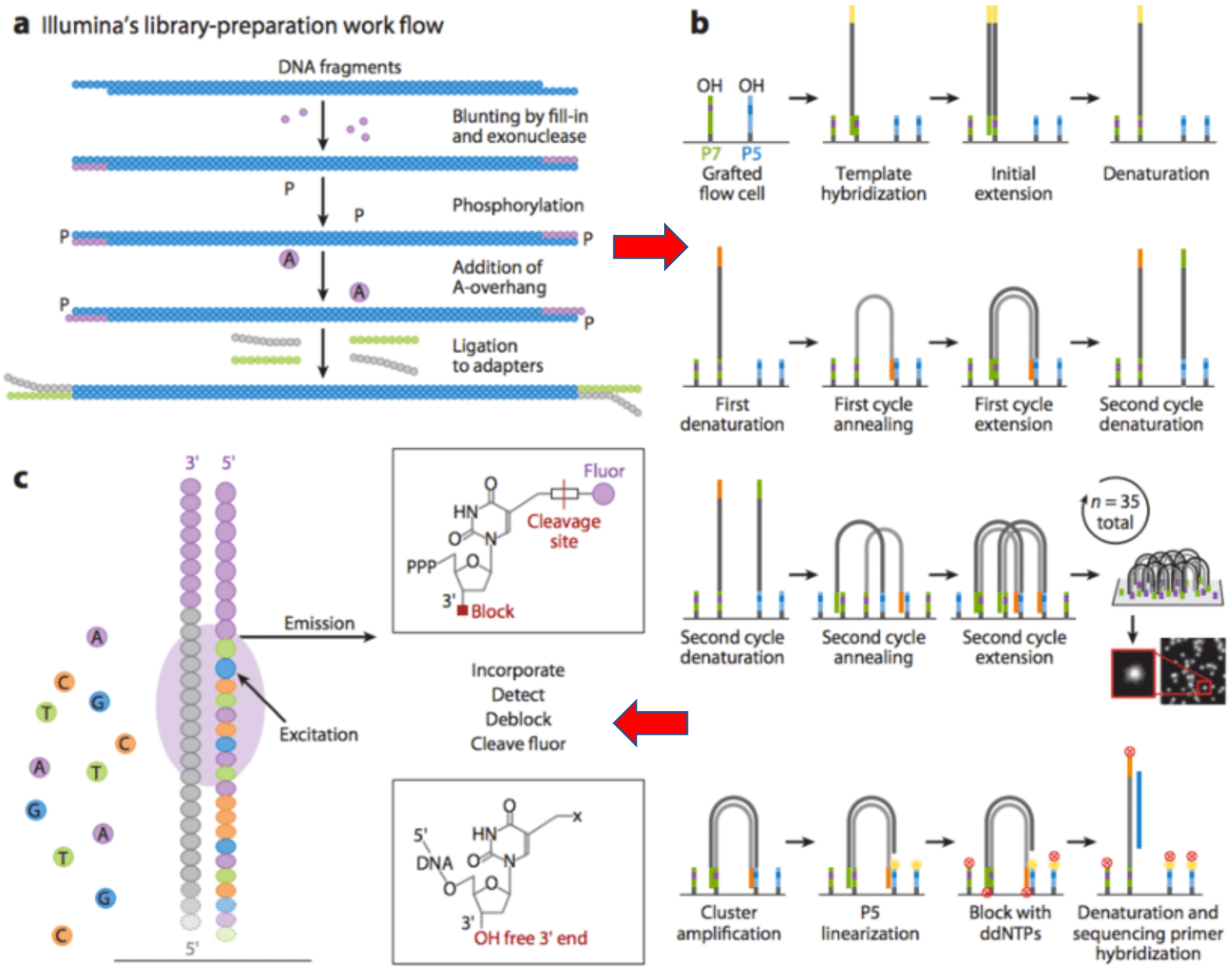




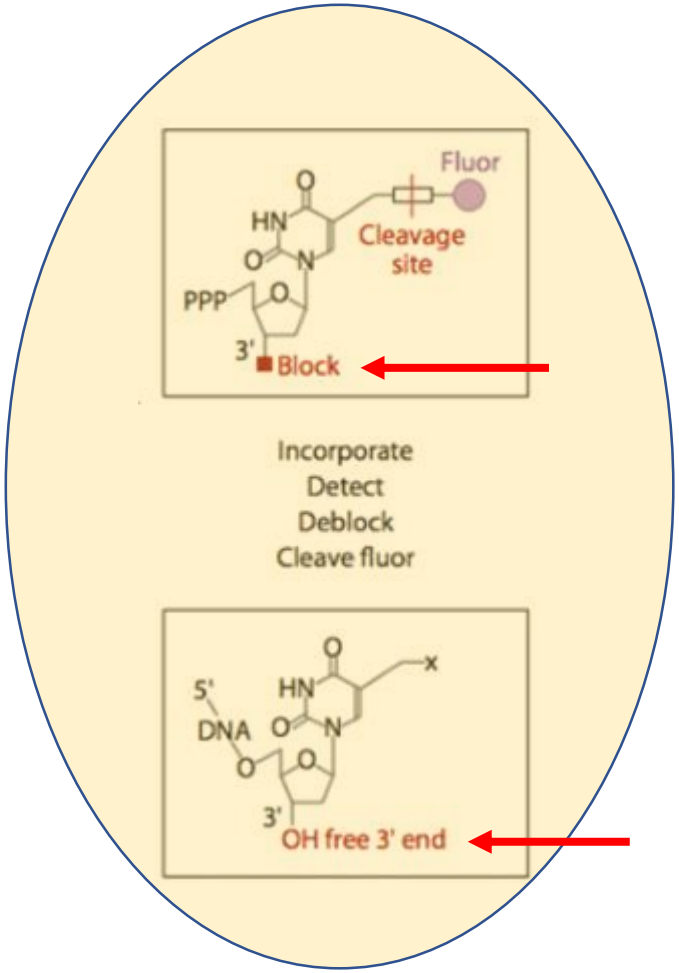
P5/ P7: binding sites to the flow cell
Rd 1 SP: read1 sequencing primer
Rd 2 SP: read2 sequencing primer

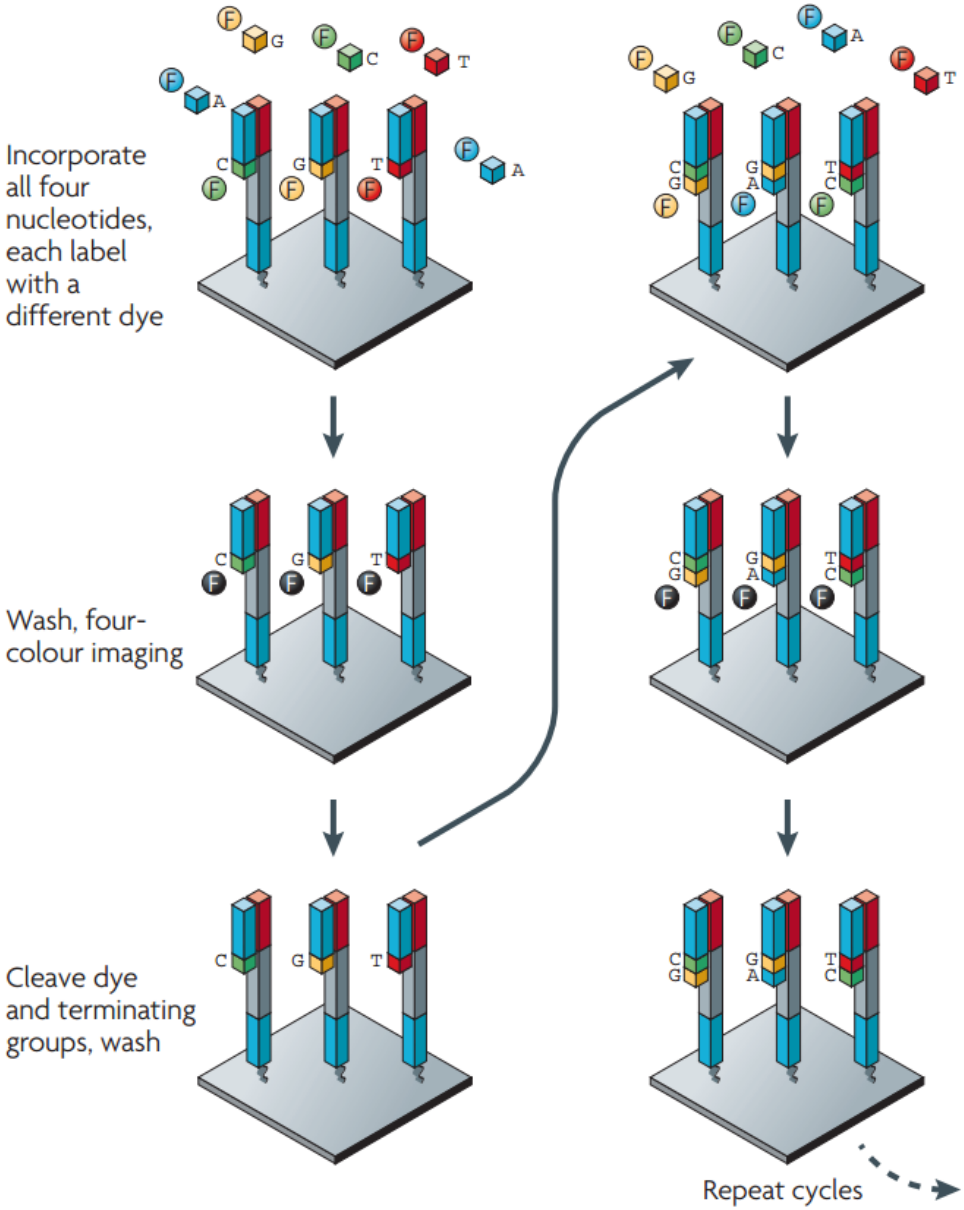


Bio-Fab research
Servizi di Supporto alla Ricerca


a Illumina's library-preparation work flow
DNA fragments
Blunting by fill-in and exonuclease
Phosphorylation
Addition of A-overhang
Ligation to adapters
b
OH OH
P7 P5
Grafted flow cell
Template hybridization
Initial extension
Denaturation
First denaturation
First cycle annealing
First cycle extension
Second cycle denaturation
Second cycle denaturation
Second cycle annealing
Second cycle extension
n = 35 total
Cluster amplification
P5 linearization
Block with ddNTPs
Denaturation and sequencing primer hybridization
c
3' 5'
Emission
Excitation
Fluor
Cleavage site
Block
Incorporate
Detect
Deblock
Cleave fluor
5' DNA
OH free 3' end

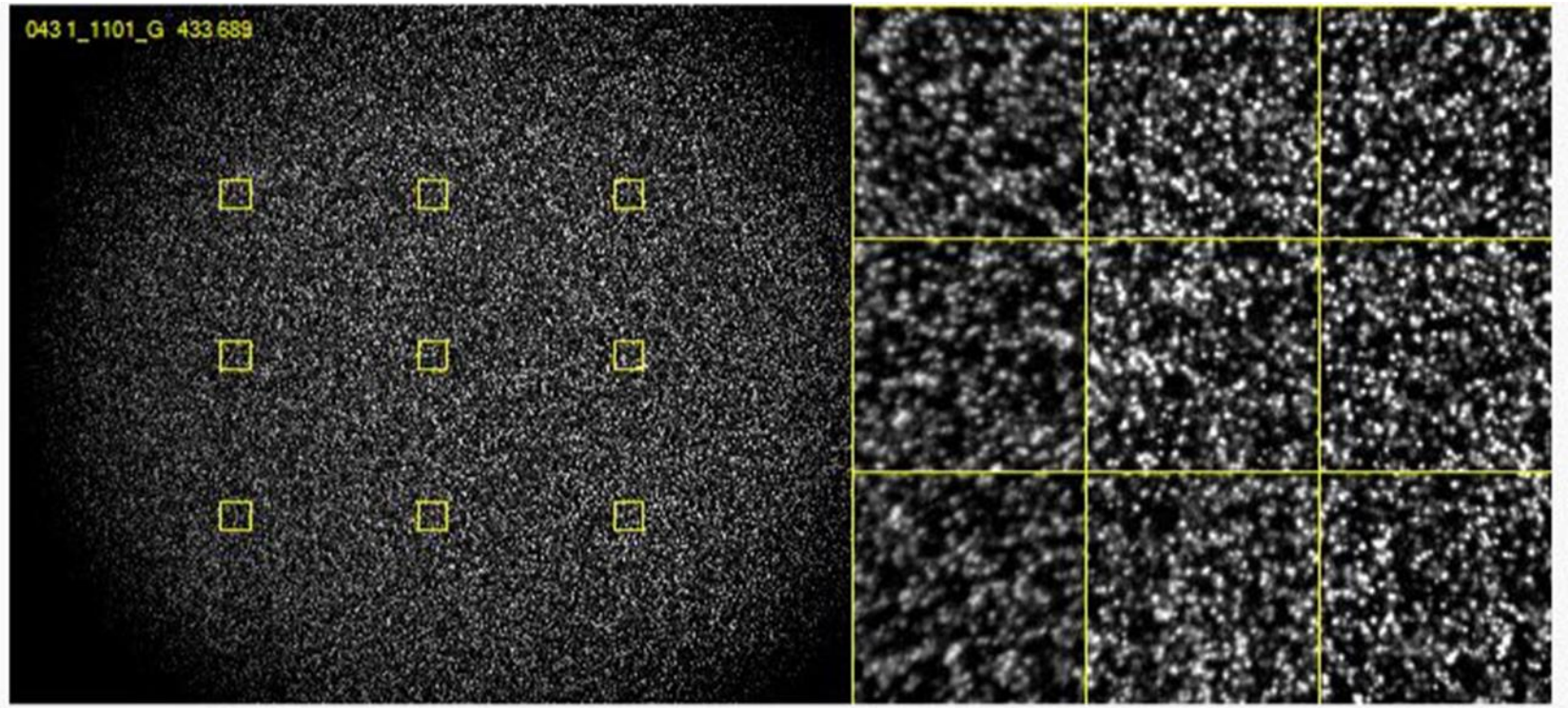
Sequencing By Synthesis





Migliore incorporazione, minore possibilità di errore

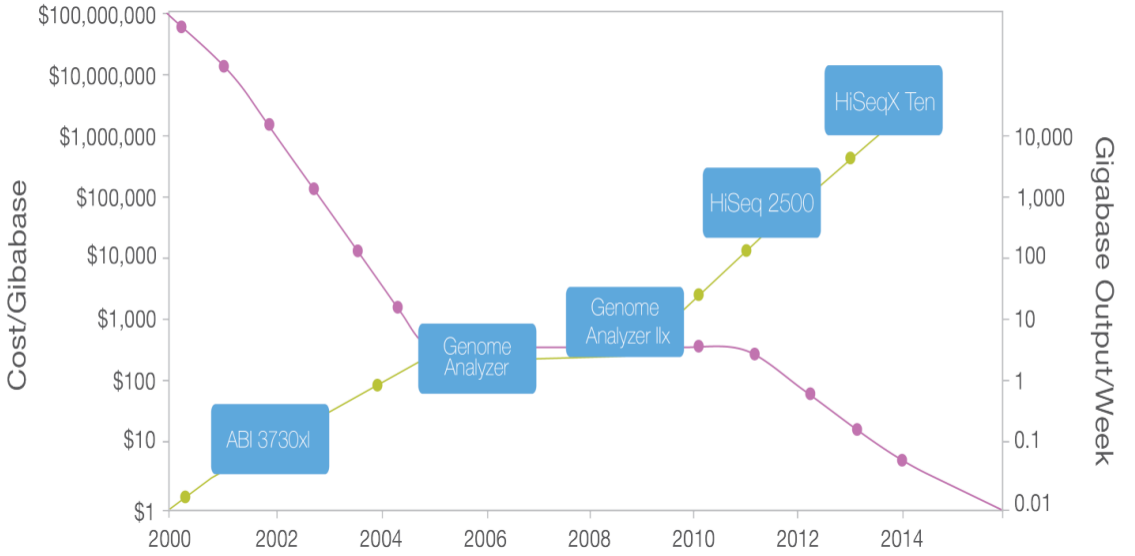
Bio-Fab research
Servizi di Supporto alla Ricerca



Costi e Benefici

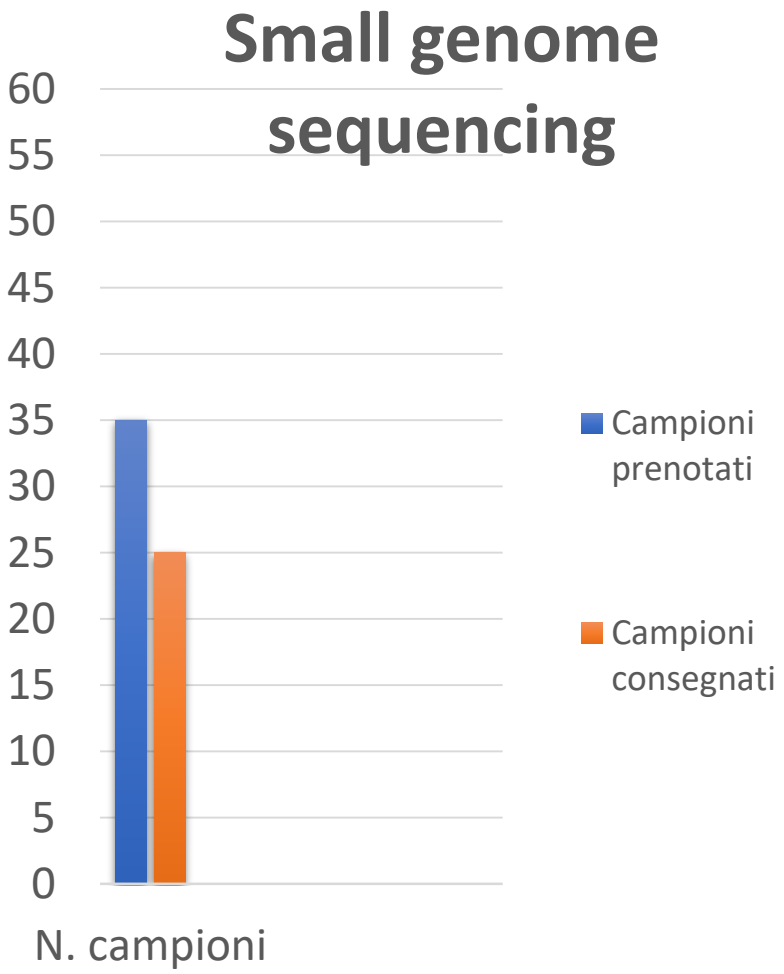
Aumento Gigabasi

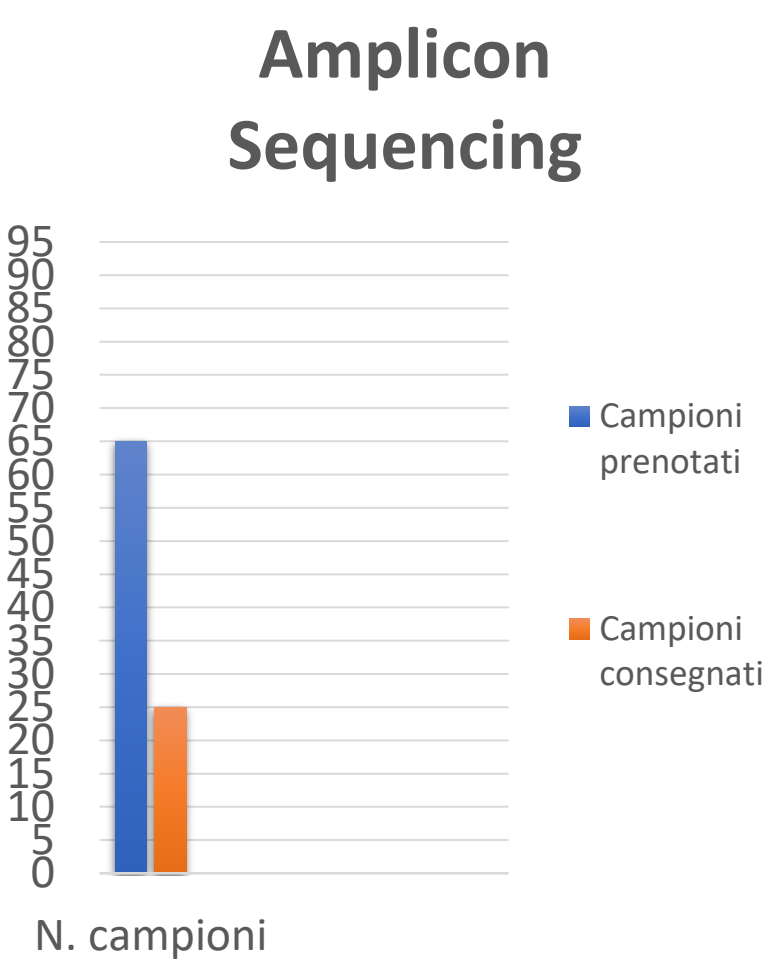
Diminuzione costo



Ma allora perché……

per un sequenziamento NGS ci vogliono tanti……..





Bio-Fab research
Servizi di Supporto alla Ricerca

Prenotazione corsa small genome sequencing Febbraio 2019

- Tipologia di corsa: 2X300
- Dimensioni massime genoma: 5Mbp
- Coverage medio: 30X

Nome e cognome	E-mail	Telefono	Numero campioni	Data prenotazione	Data consegna campioni

N.B.
- Numero massimo di campioni per flow cell: 60
- Prezzo riservato per l'ISS di 200€/campione solo al riempimento totale della flow cell
- La corsa potrebbe subire dei ritardi se non si raggiunge il numero necessario dei campioni

Bio-Fab NGS research

Prenotazione corsa amplicon sequencing (16S-ITS-ampliconi) Febbraio 2019

- Tipologia di corsa: 2X300
- Dimensioni amplicone: 450bp
- Numero di reads consegnate: 100,000

Nome e cognome	E-mail	Telefono	Numero campioni	Data prenotazione	Data consegna campioni

N.B.
- Numero massimo di campioni per flow cell: 96
- Prezzo riservato per l'ISS di 57€/campione solo al riempimento totale della flow cell
- La corsa potrebbe subire dei ritardi se non si raggiunge il numero necessario dei campioni

Bio-Fab NGS research

Small genome sequencing

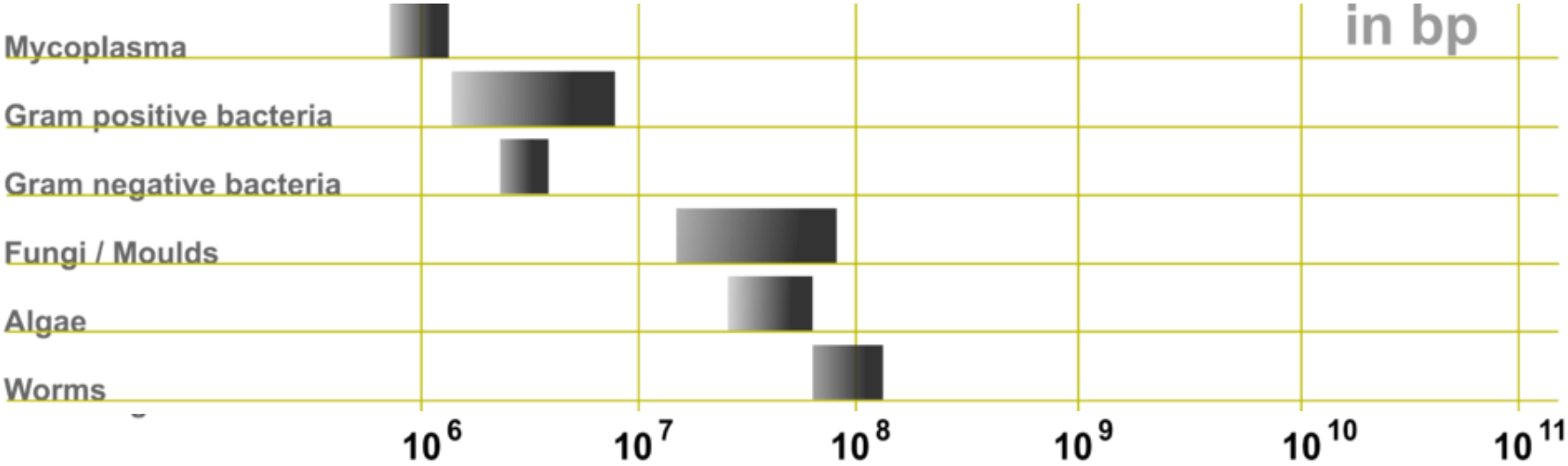
1. Che cosa si intende per small genome sequencing
2. Cosa possiamo fare sequenziando un intero genoma
3. Terminologia
4. Strategie d'esperimento
5. Qualità del materiale da inviare
6. Preparazione libreria
7. Controlli sulla libreria

Bio-Fab research
Servizi di Supporto alla Ricerca

Small genome sequencing

Dimensioni dei genomi sequenziabili con il MiSeq:

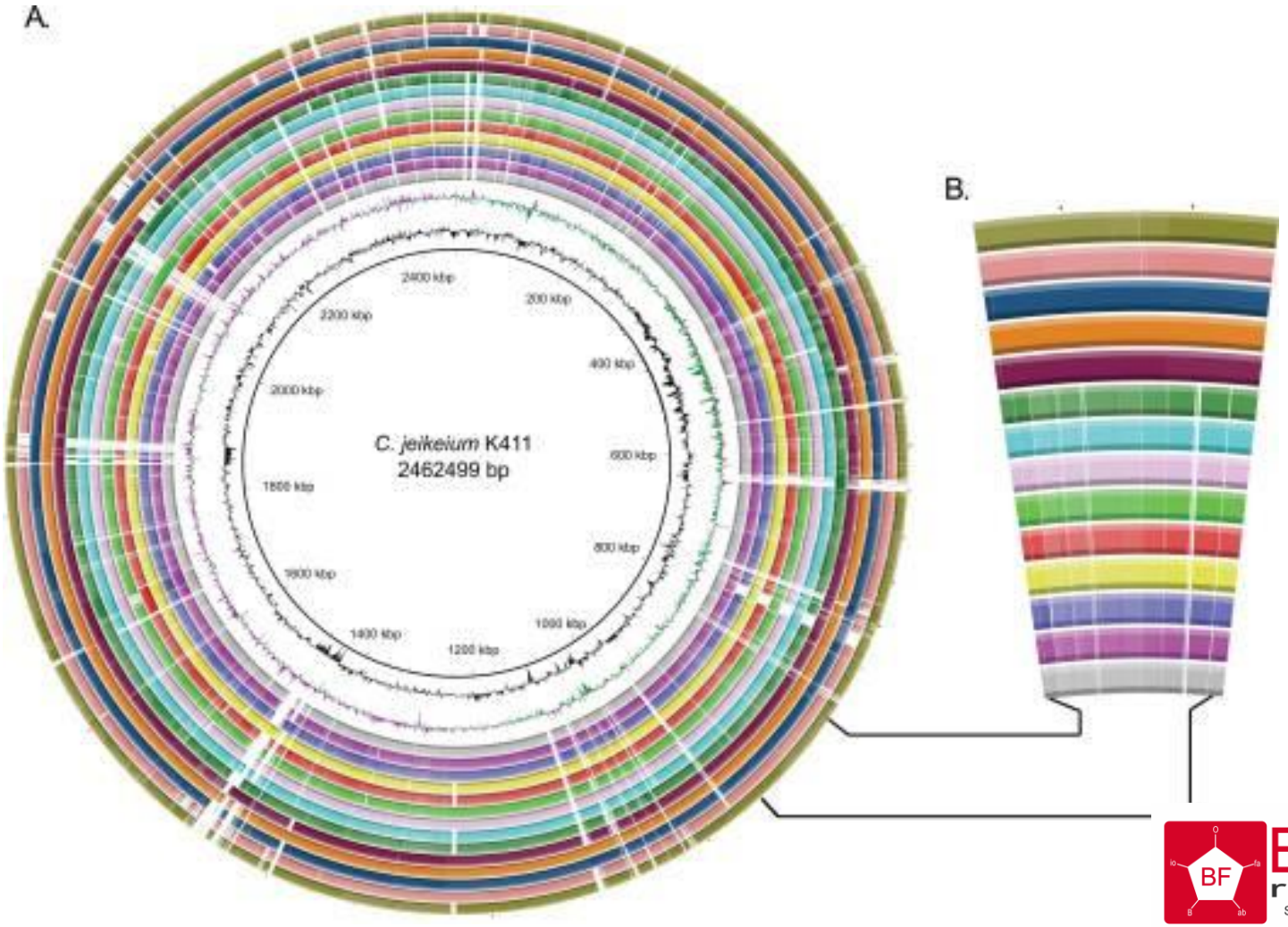
- Plasmidi e mitocondri (4-17 Kbp)
- Virus (10-100 Kbp)
- Adenovirus (30-50 Kbp)
- Batteri (2-6Mbp)
- Lieviti e protozoi (8-10 Mbp)
- Funghi (10-50Mbp)



Bio-Fab research
Servizi di Supporto alla Ricerca

Small genome sequencing

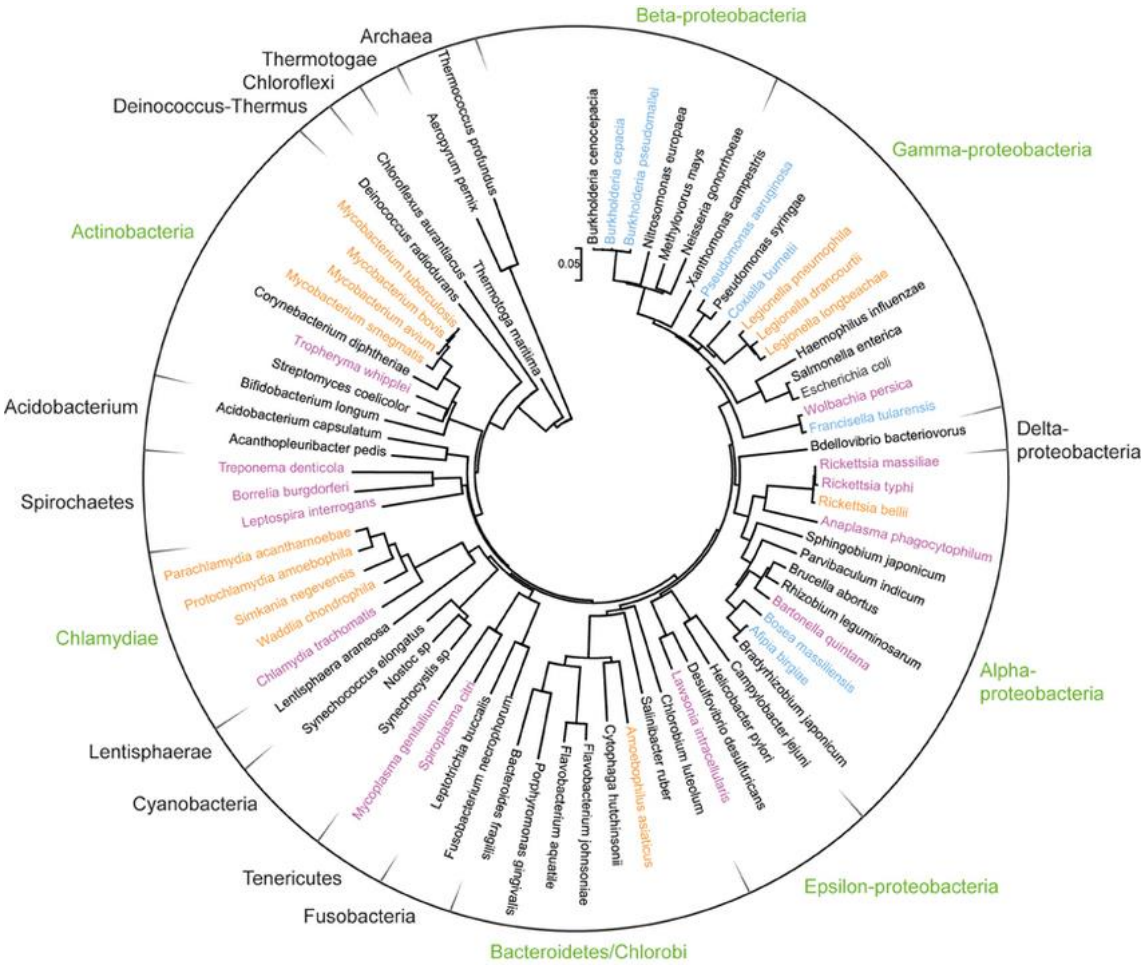
Il sequenziamento di interi genomi permette di sequenziare contemporaneamente tutti i geni noti e non di un dato organismo e di confrontarli con altri organismi della stessa specie o di specie differenti per ricostruire tutte le differenze fra i genomi.



Bio-Fab research
Servizi di Supporto alla Ricerca

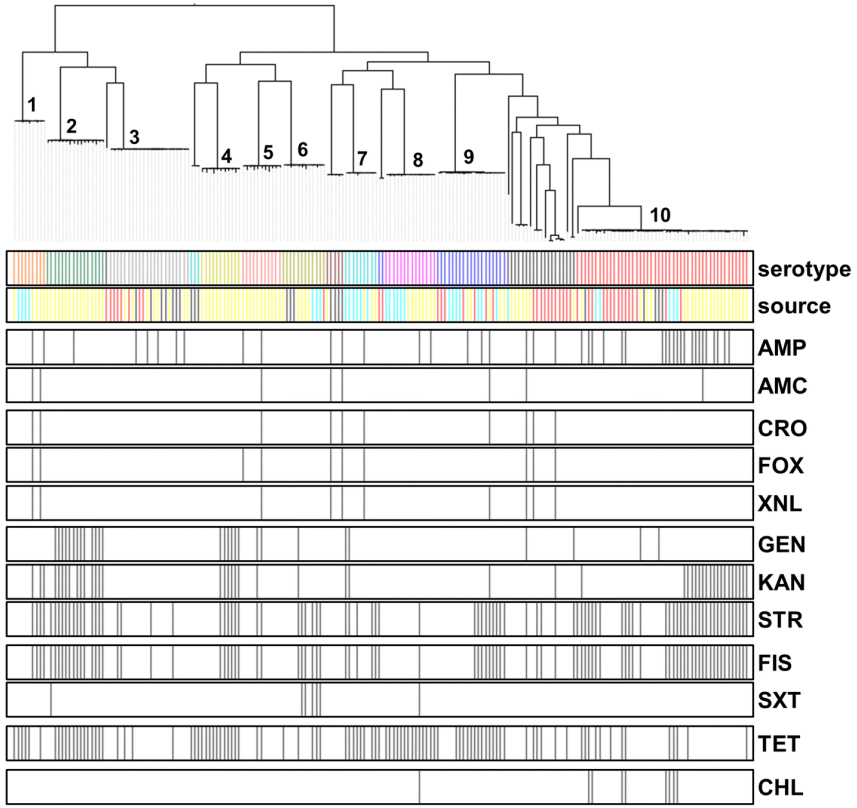
Small genome sequencing

Comprendere l'organizzazione, l'evoluzione e le relazioni evolutive tra diversi organismi
Conoscere i geni e le istruzioni per esprimerli nel tempo e nello spazio



Small genome sequencing

- Ricerca di regioni specifiche del genoma, geni di resistenza, di virulenza o di plasmidi.
- Genotipizzazione dei batteri sequenziati
- Creazione di banche dati personali da poter confrontare con quelle pubbliche o da interrogare in periodi successivi

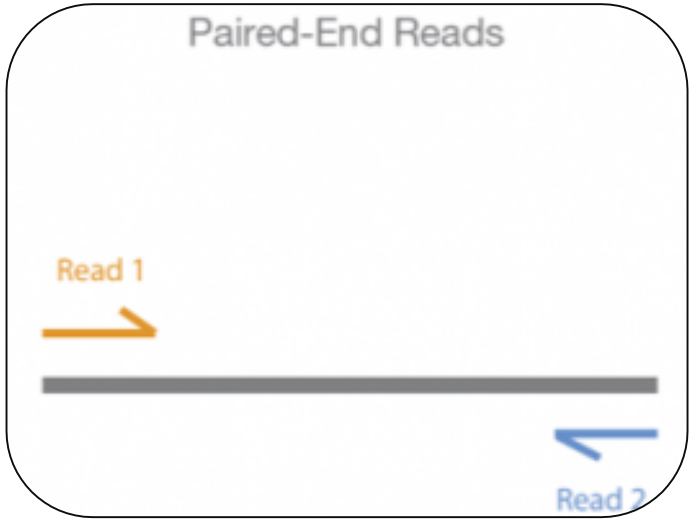


Bio-Fab research
Servizi di Supporto alla Ricerca

Terminologia

- **Read** (lettura): si riferisci ad una stringa di dati che corrisponde ad una data sequenza
- **Numero di reads**: numero di letture effettuate per singolo campione, espresso in M (milioni) di reads
- **Single end** sequencing: sequenziamento di una sola estremità del DNA
- **Paired End** sequencing: vengono sequenziate entrambe le estremità del frammento di DNA, per migliorarne l'accuratezza e l'allineamento

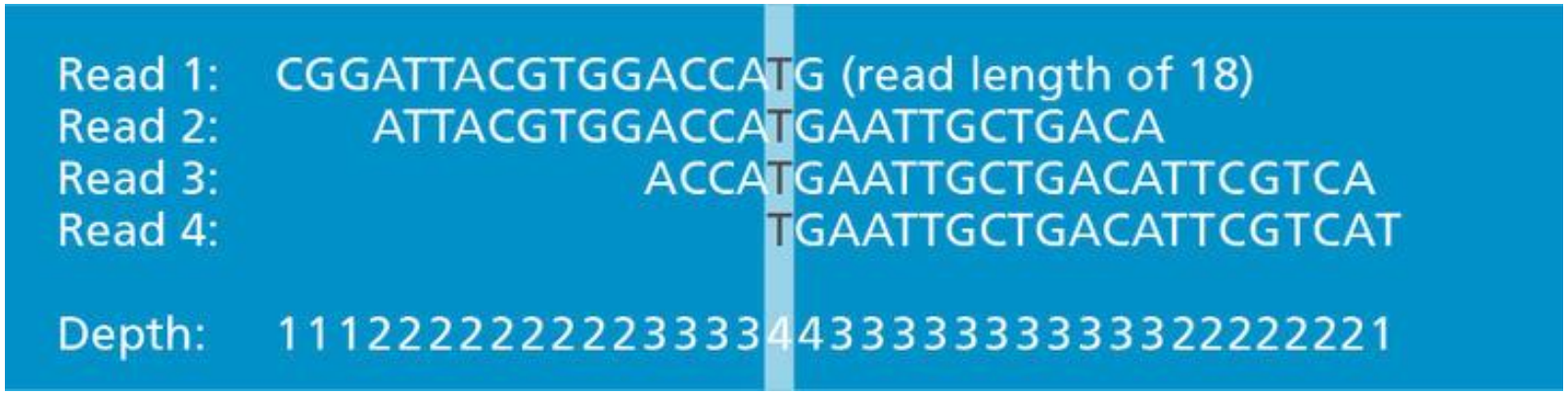
Il sequenziamento PE permette all'algoritmo di mappare meglio le regioni ripetute.



Bio-Fab research
Servizi di Supporto alla Ricerca

Terminologia

- **Coverage (profondità):** indica il numero medio di basi sequenziate che si allineano ad una data base sulla reference
- **Output:** numero di basi totali lette dal sequenziatore, espresso in **Gbp** (giga basi)



Bio-Fab research
Servizi di Supporto alla Ricerca

Terminologia

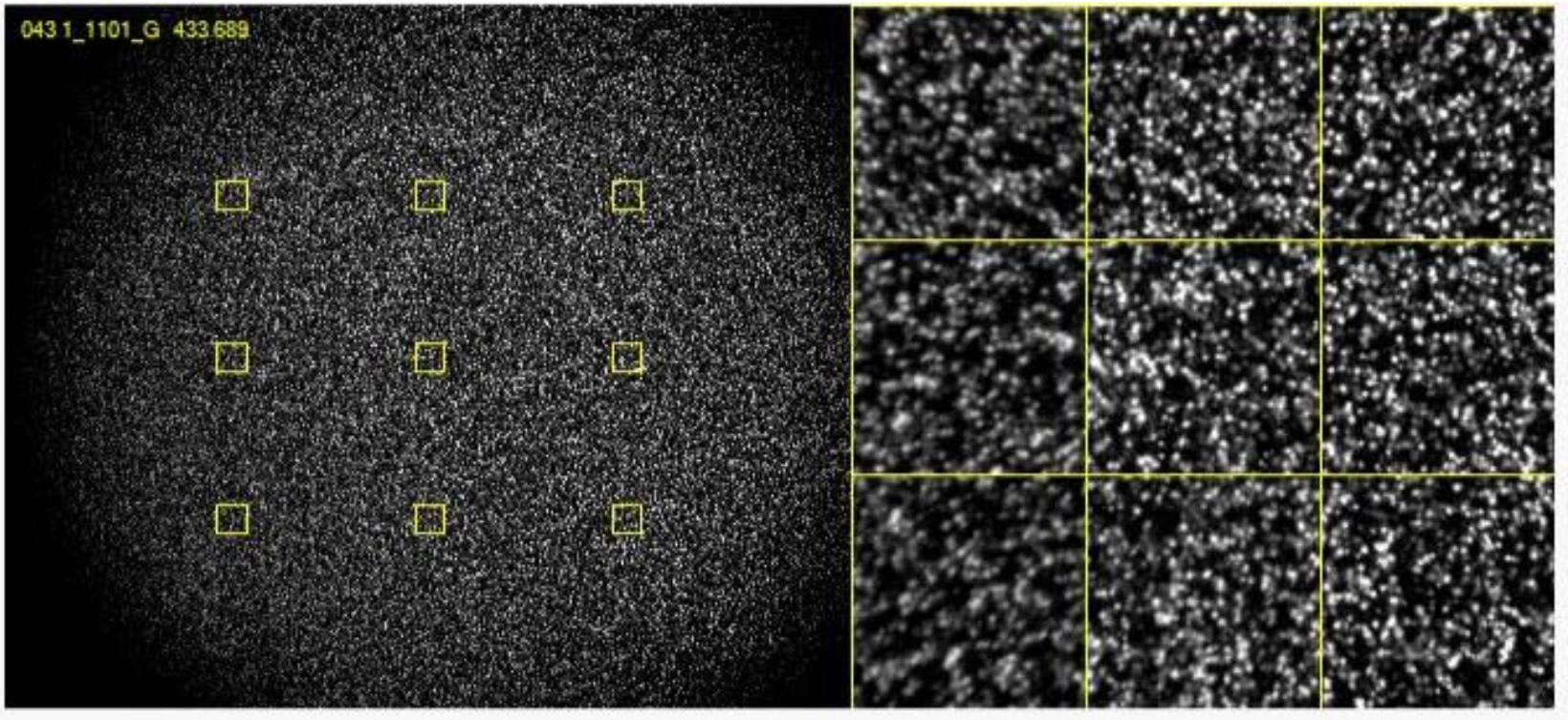
- **Index**: sequenza nucleotidica che si aggiunge per PCR o ligasi ad ogni campioni corso sul sequenziatore
- **Multiplexing**: è un processo tipico delle corse Illumina, in cui si aggiunge ad ogni campione un index interno per aumentare il numero di campioni corsi su una flowcell.
- **Flowcell**: il vetrino dove avviene la lettura del DNA

Flow cell Illumina

Bio-Fab research
Servizi di Supporto alla Ricerca

Terminologia

Cluster: raggruppamento clonale di DNA derivante dal DNA templato legato ad una flowcell



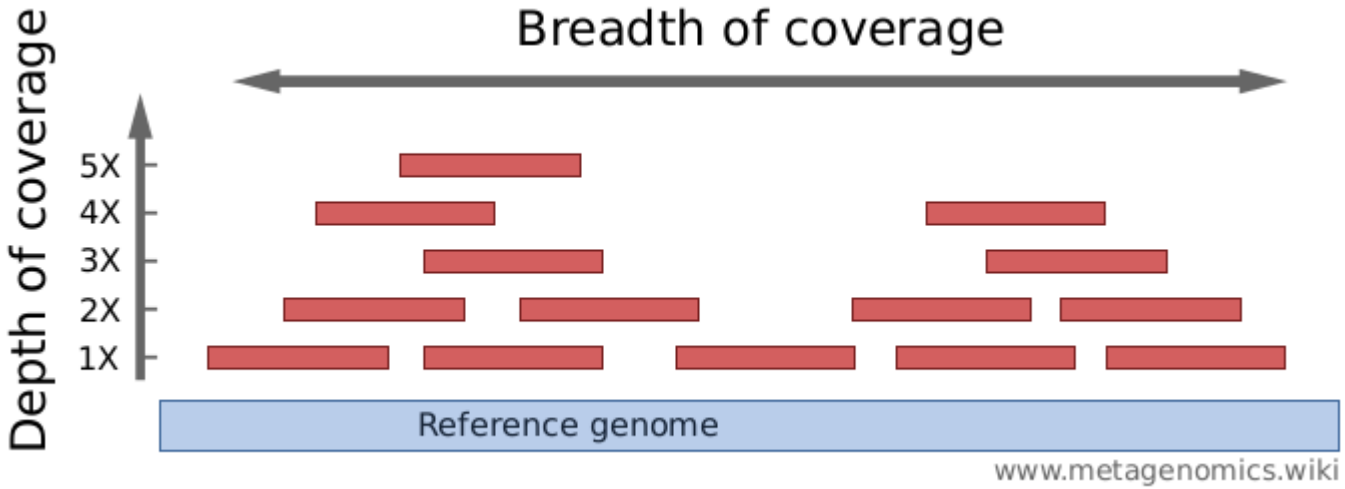
Bio-Fab research
Servizi di Supporto alla Ricerca

Strategie di esperimento

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, anche per genomi non annotati.



Bio-Fab research
Servizi di Supporto alla Ricerca

Quanti campioni posso caricare in una corsa?

Dipende da:

- Dimensioni del genoma
- Coverage
- Output dal sequenziatore

MiSeq - Illumina

- Reads number: **25M** di reads in SR o **50M** di reads in PE
- Output: **15Gbp**
- Letture: **1X50** fino a **2X300** (600 nucleotidi)

Bio-Fab research
Servizi di Supporto alla Ricerca

Quanti campioni posso caricare in una corsa?

Voglio sequenziare *E. coli*

- Dimensioni genoma E. coli: circa 4 Mbp
- Coverage richiesto: 50X
- Tipo di corsa: 2X300

$$C = LN / G$$

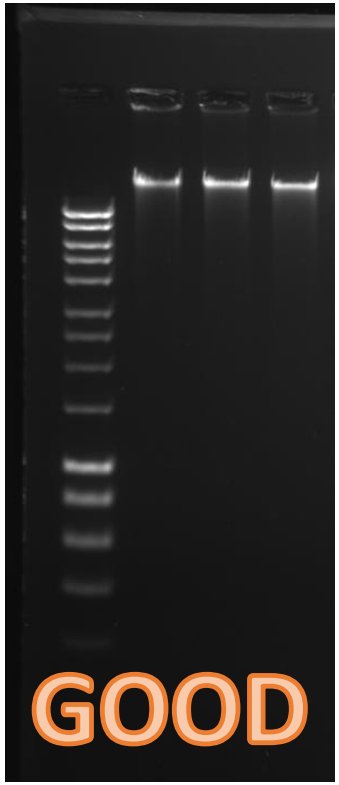
- C stands for coverage
- G is the haploid genome length
- L is the read length
- N is the number of reads

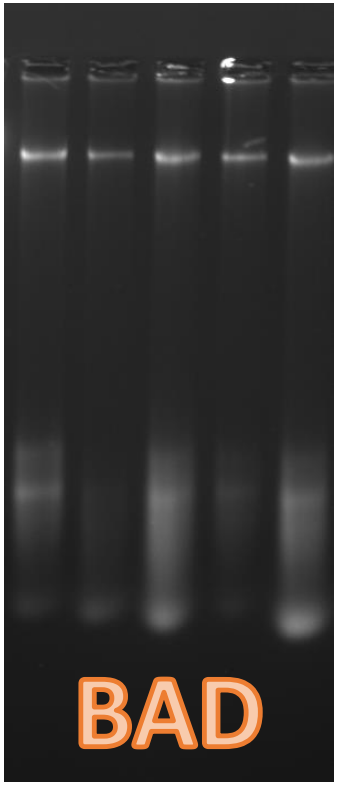
Per sequenziare un genoma di *E. coli* con una profondità 50X ed una corsa 2X300, saranno necessari all'incirca 0,2 Gbp

Bio-Fab research
Servizi di Supporto alla Ricerca

Quantità e qualità del DNA genomico da inviare

1. Controllare il DNA genomico con una corsa elettroforetica (gel di Agarosio 0,8%)





Bio-Fab research
Servizi di Supporto alla Ricerca

Quantità e qualità del DNA genomico da inviare

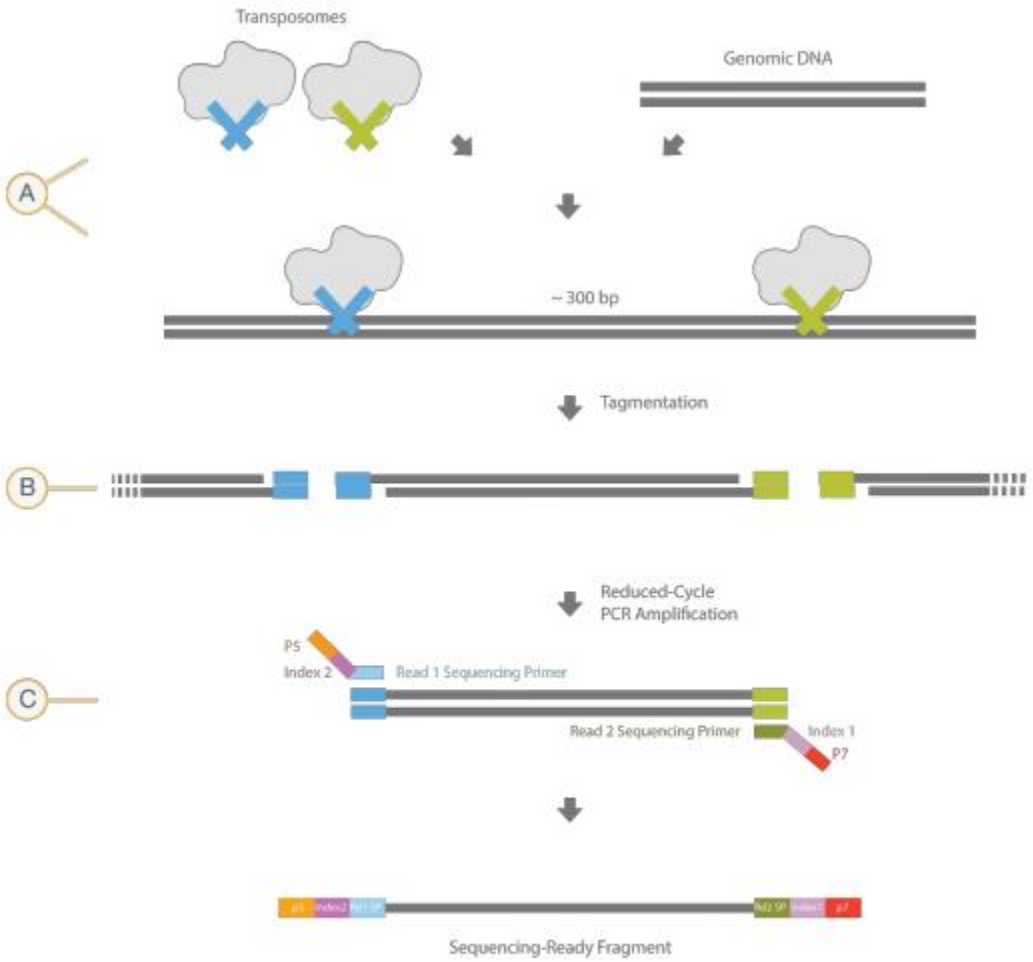
2. Controllare la qualità e quantità del DNA con uno spettrofotometro

Rapporto 260/280 > 1,8 – 2,0
Rapporto 260/230 > 2,0 – 2,2

Quanto DNA ci dovete inviare?
Circa 200 ng, concentrato almeno 20 ng/µl

Bio-Fab research
Servizi di Supporto alla Ricerca

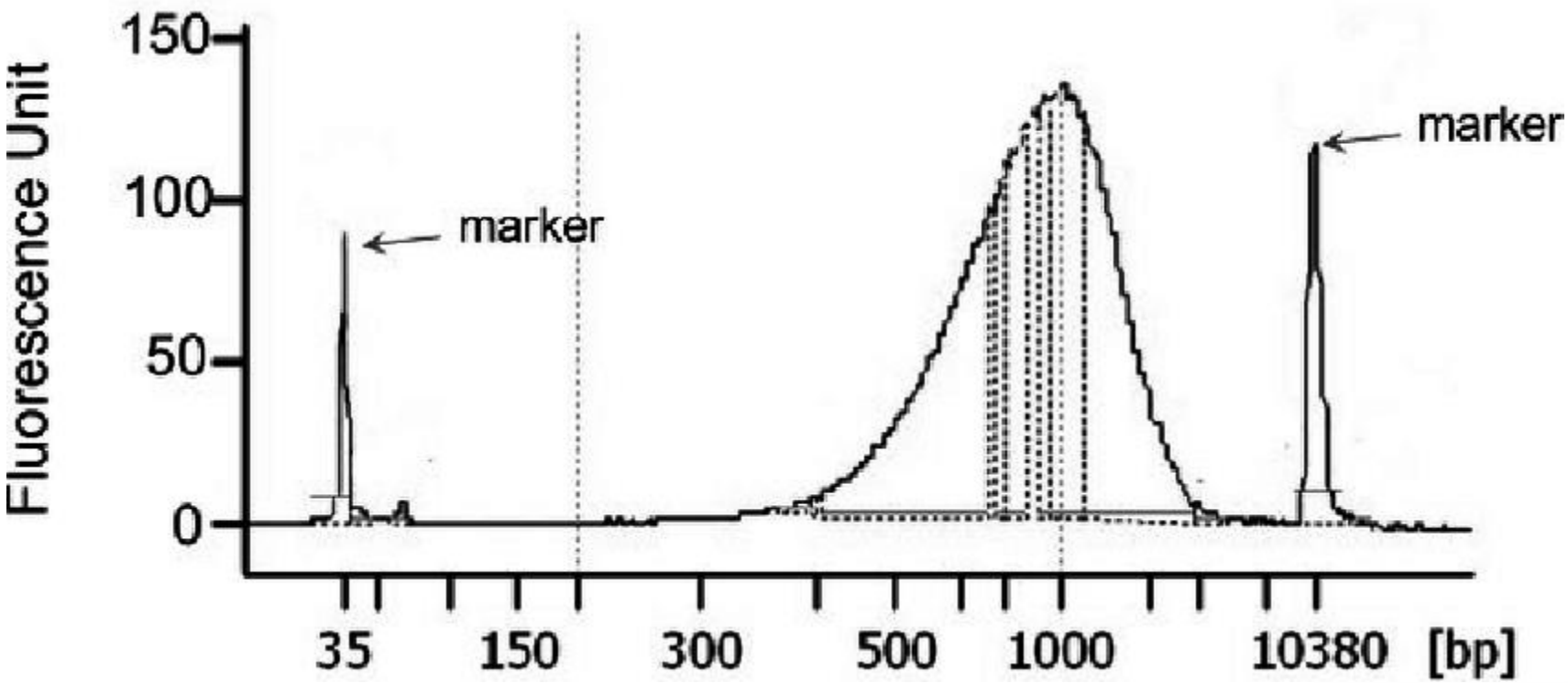
Preparazione della libreria



A Nextera XT transposome with adapters combined with template DNA
B Tagmentation to fragment and add adapters
C Limited-cycle PCR to add index adapter sequences

Bio-Fab research
Servizi di Supporto alla Ricerca

Controllo e quantificazione delle librerie
Agilent 2100 Bioanalyzer

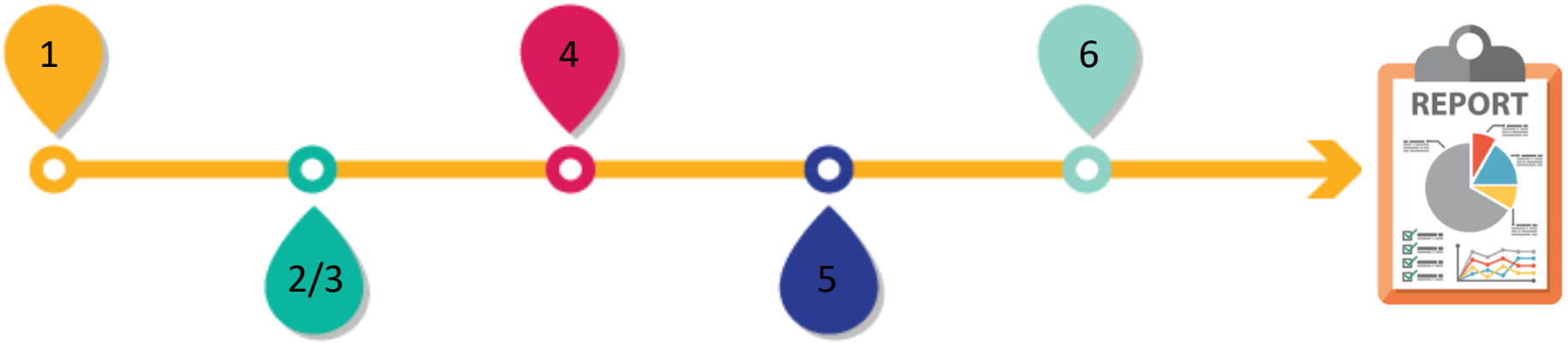


Bio-Fab research
Servizi di Supporto alla Ricerca

Perché la bioinformatica

- File di testo MOLTO GRANDI (migliaia di milioni di righe)
  - Non si possono usare gli strumenti “soliti”
  - Enorme utilizzo di memoria e tempi di corsa
  - Gestire, analizzare, accumulare, trasferire ed archiviare file giganteschi
- Necessità di computer potenti e di competenze
  - Computer clusters
  - Necessità di nuovi algoritmi e software spesso open source Unix/Linux based.
  - Collaborazione tra chi sviluppa la tecnologia, i bioinformatici e i biologi

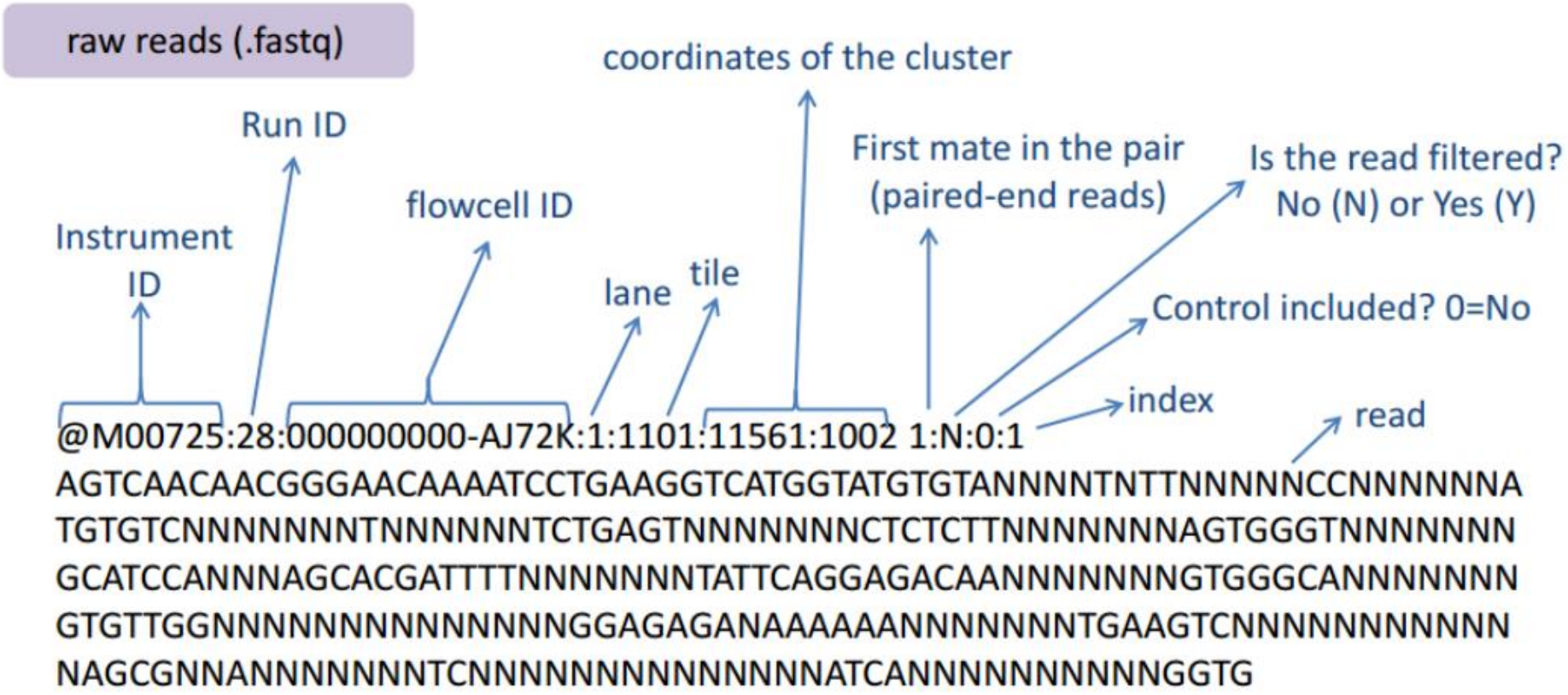
Workflow



1. Raw data
2. Assembly de novo
3. Chiamata delle varianti
4. Resistenze e altro
5. Filogenomica
6. Analisi personalizzate

Bio-Fab research
Servizi di Supporto alla Ricerca

Dati grezzi



Bio-Fab research
Servizi di Supporto alla Ricerca

Dati grezzi

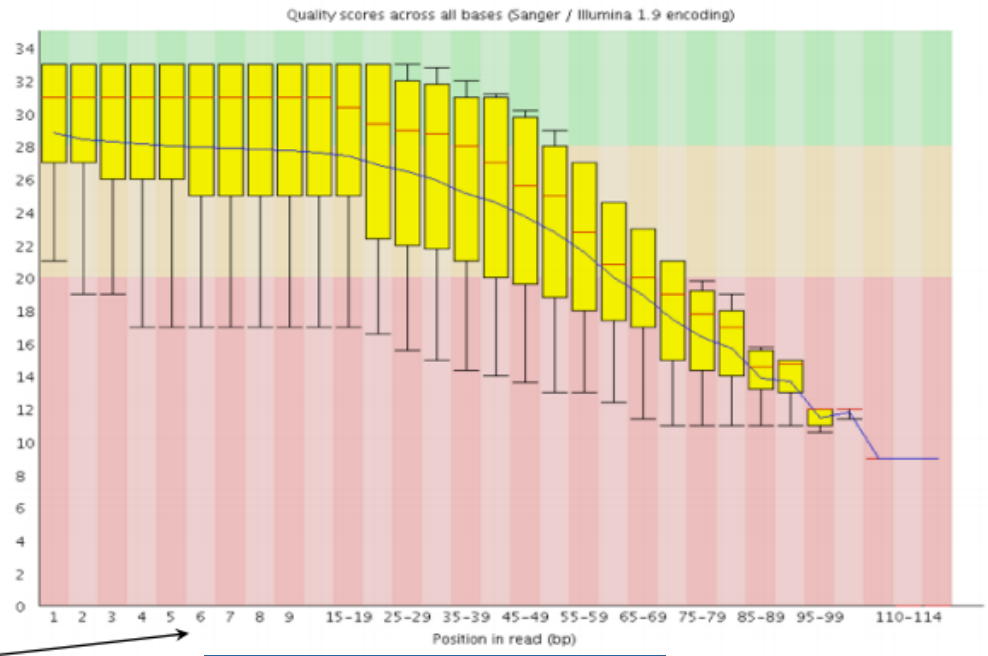
Basic Statistics

Measure	Value
Filename	B13_212.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	522099
Filtered Sequences	0
Sequence length	16-115
%GC	50

quality (Scale 0-40)

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

Per base sequence quality



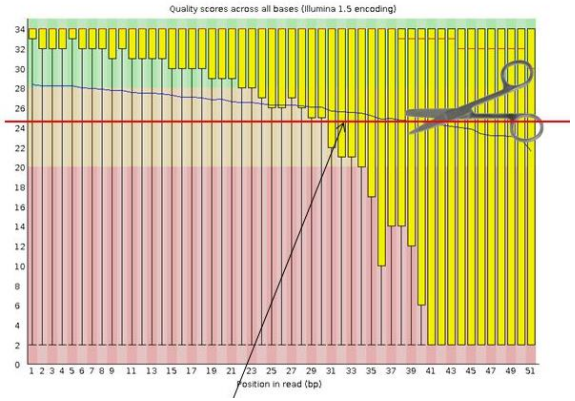
read position

Babraham Bioinformatics

Bio-Fab research
Servizi di Supporto alla Ricerca

Controllo qualità delle reads

- Non tutte le basi che compongono una reads hanno lo stesso livello di qualità
- La qualità generalmente tende a diminuire più ci avviciniamo al 3’
- E' necessario verificare in ogni esperimento di sequenziamento come varia la qualità al variare della posizione sulla read
- Basi con qualità < 20 vengono generalmente rimosse mediante un processo di trimming



- Procedendo dal 3' verso il 5' si rimuovono nucleotidi da ogni reads fino a raggiungere una qualità minima (Phred quality score >=20)

Principali applicazioni del sequenziamento di piccoli genomi

De novo Assembly

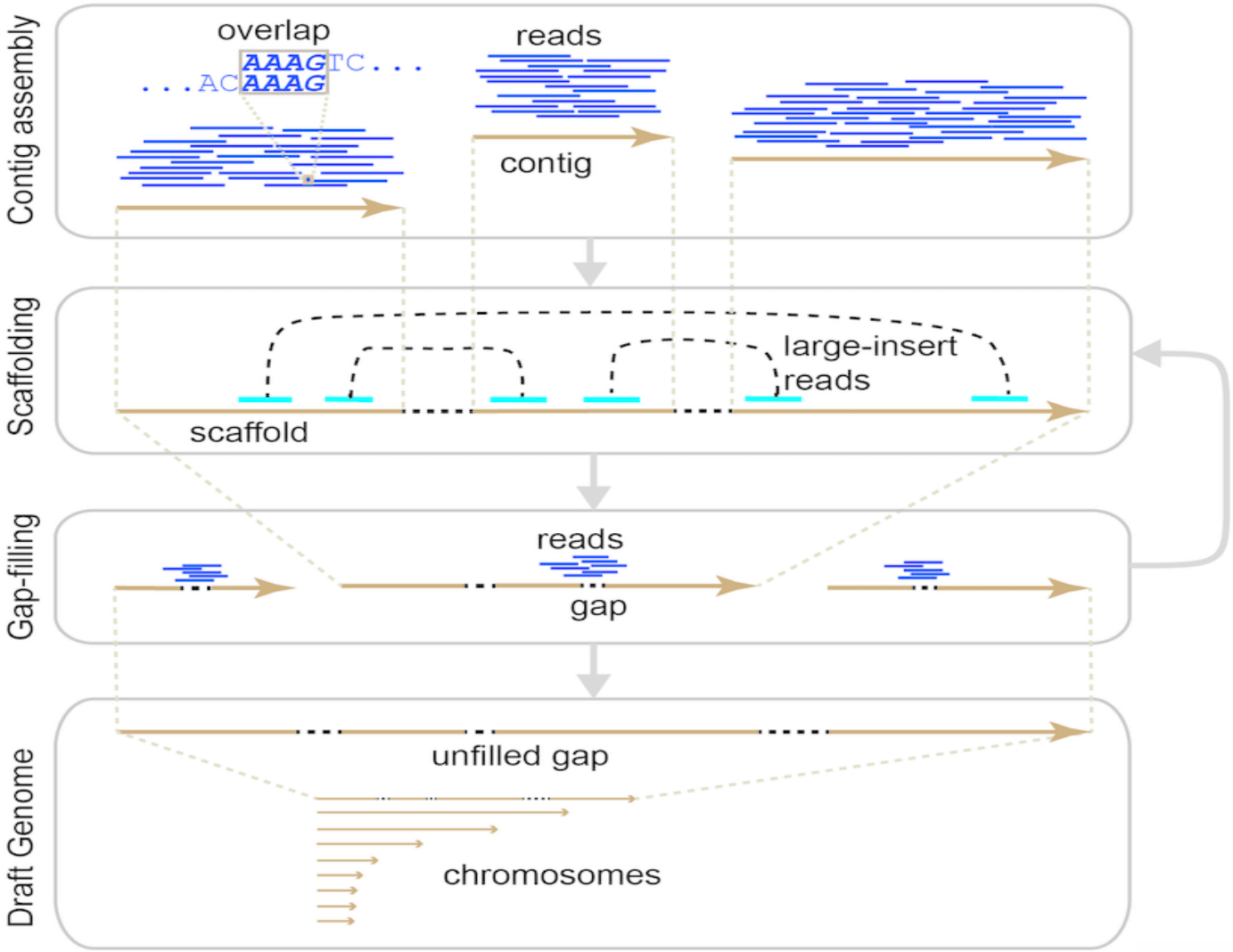
Microbial Genomes
Viral Genomes
Non-model Organisms
BAC/YAC Screening
Functional annotation
Detection of Recombination Events
Plasmids

Resequencing /Variant Calling

SNP Discovery
Detection of indel and recombination events
Plasmids

Bio-Fab research
Servizi di Supporto alla Ricerca

Assembly de novo

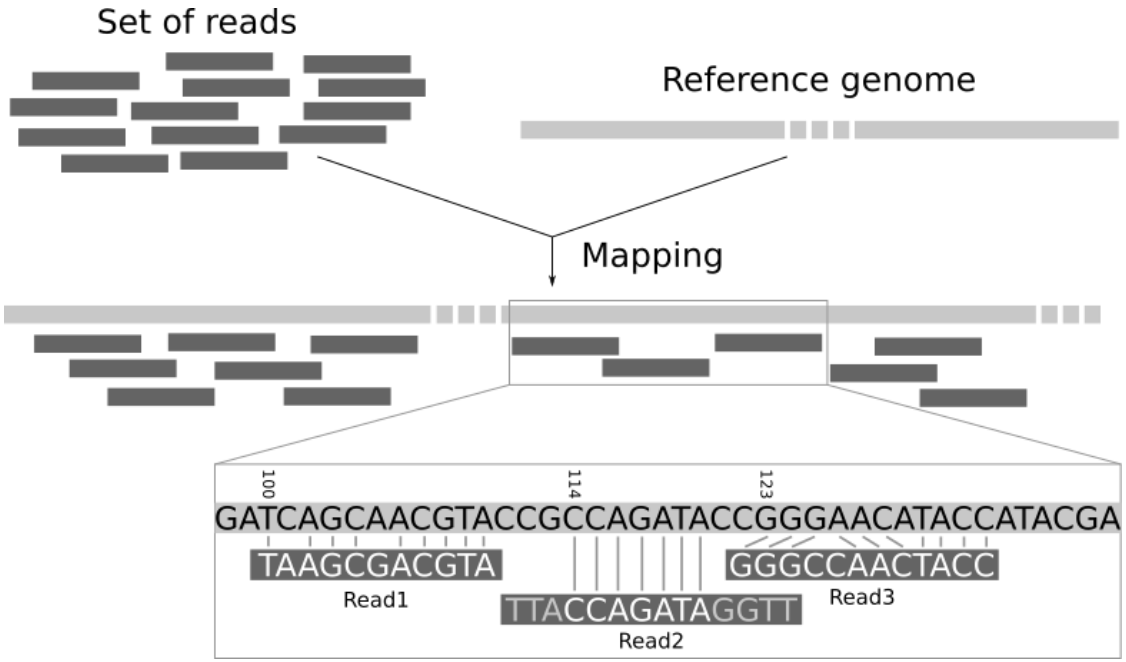


Bio-Fab research
Servizi di Supporto alla Ricerca

Allineamento al reference

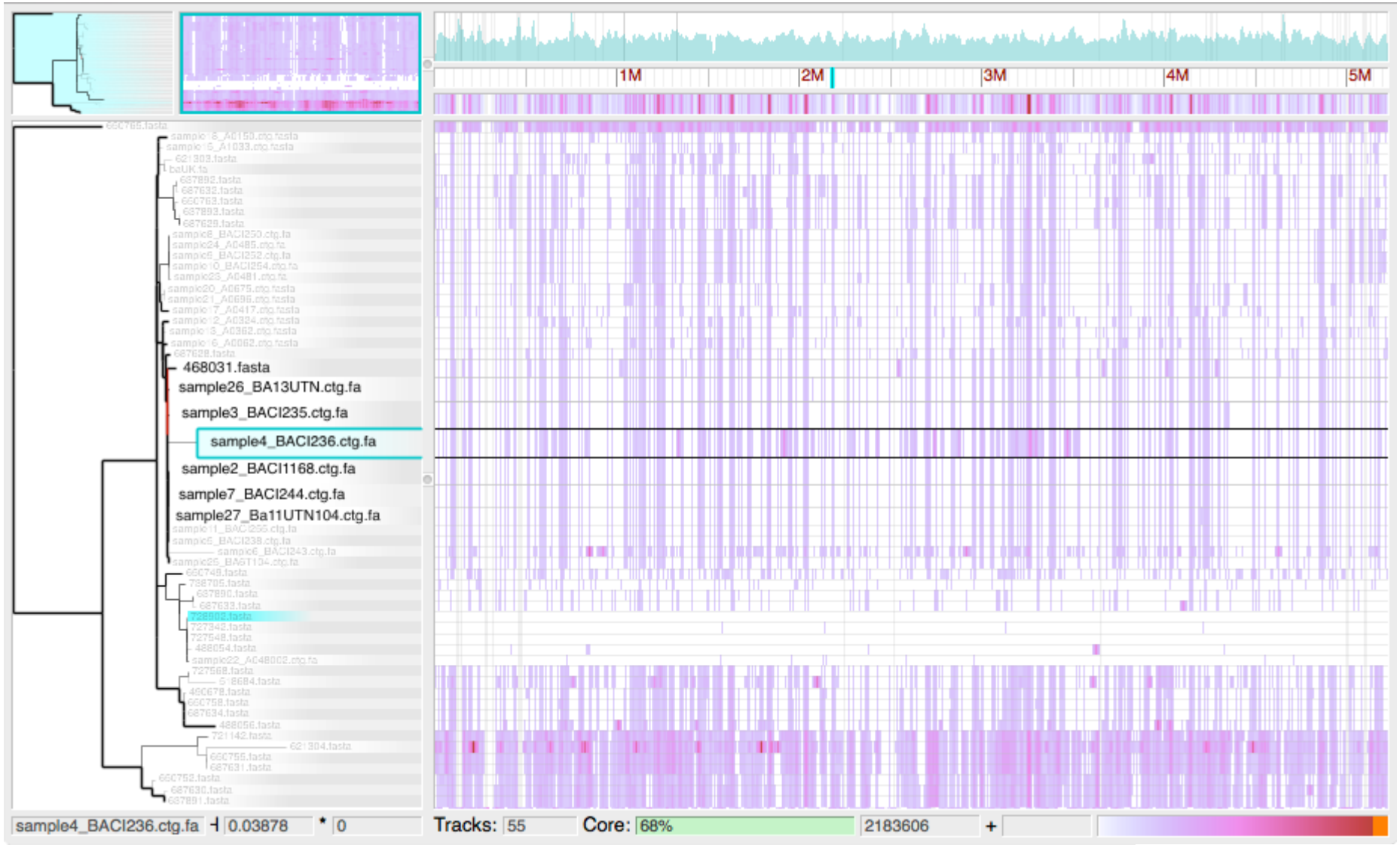
- **Genoma di riferimento**: una o più sequenze di DNA che rappresentano il genoma di un organismo

- **Allineamento**: identificare la posizione delle reads rispetto al genoma di riferimento



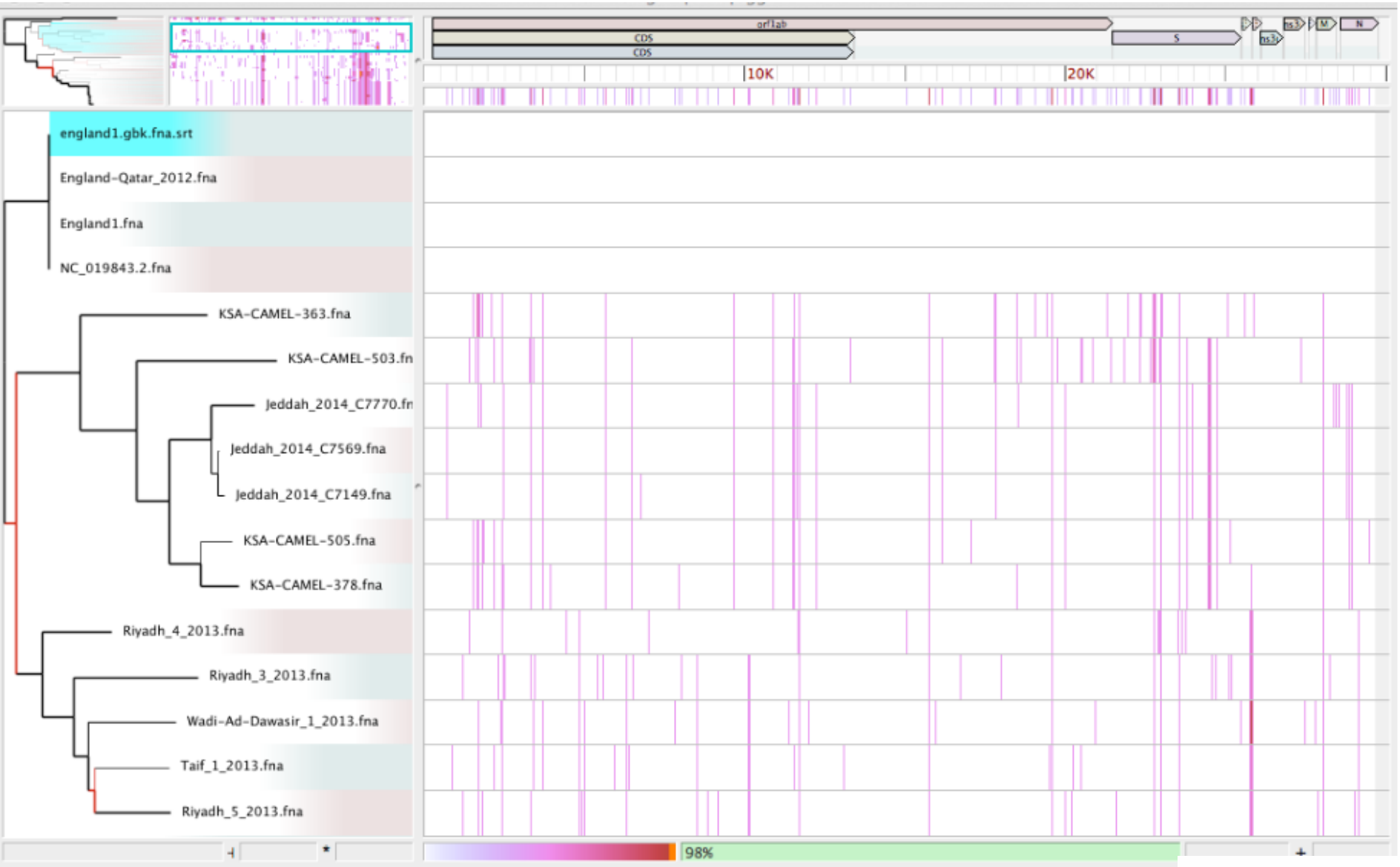
- Processo nel quale si determina **la posizione di provenienza più probabile** di una read all'interno del genoma

Chiamata delle varianti

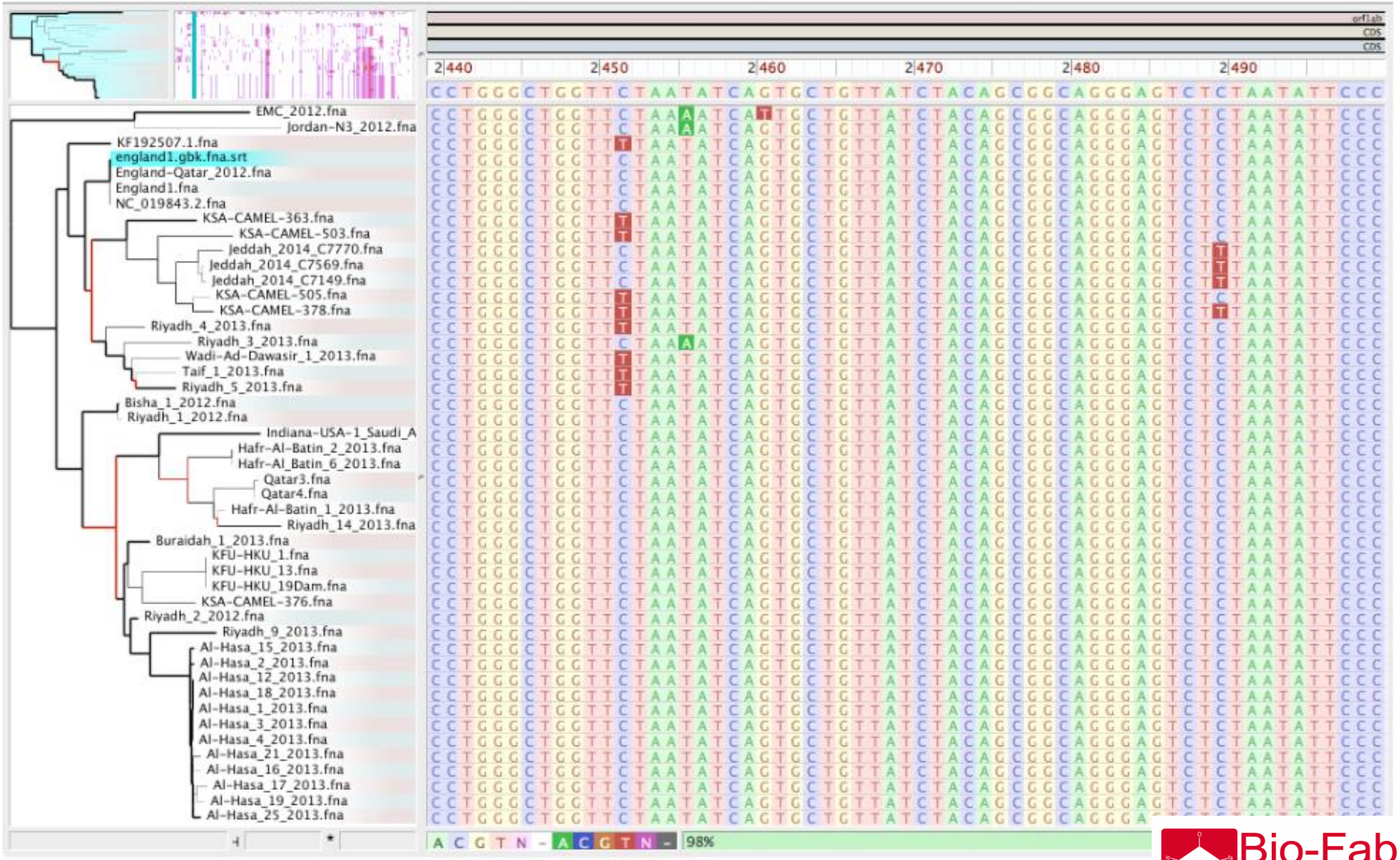


Bio-Fab research
Servizi di Supporto alla Ricerca

Chiamata delle varianti

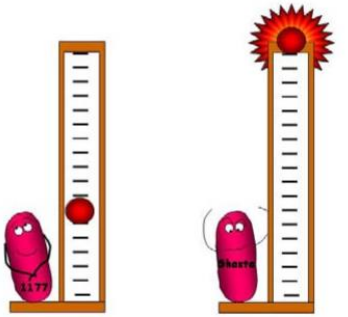


Chiamata delle varianti



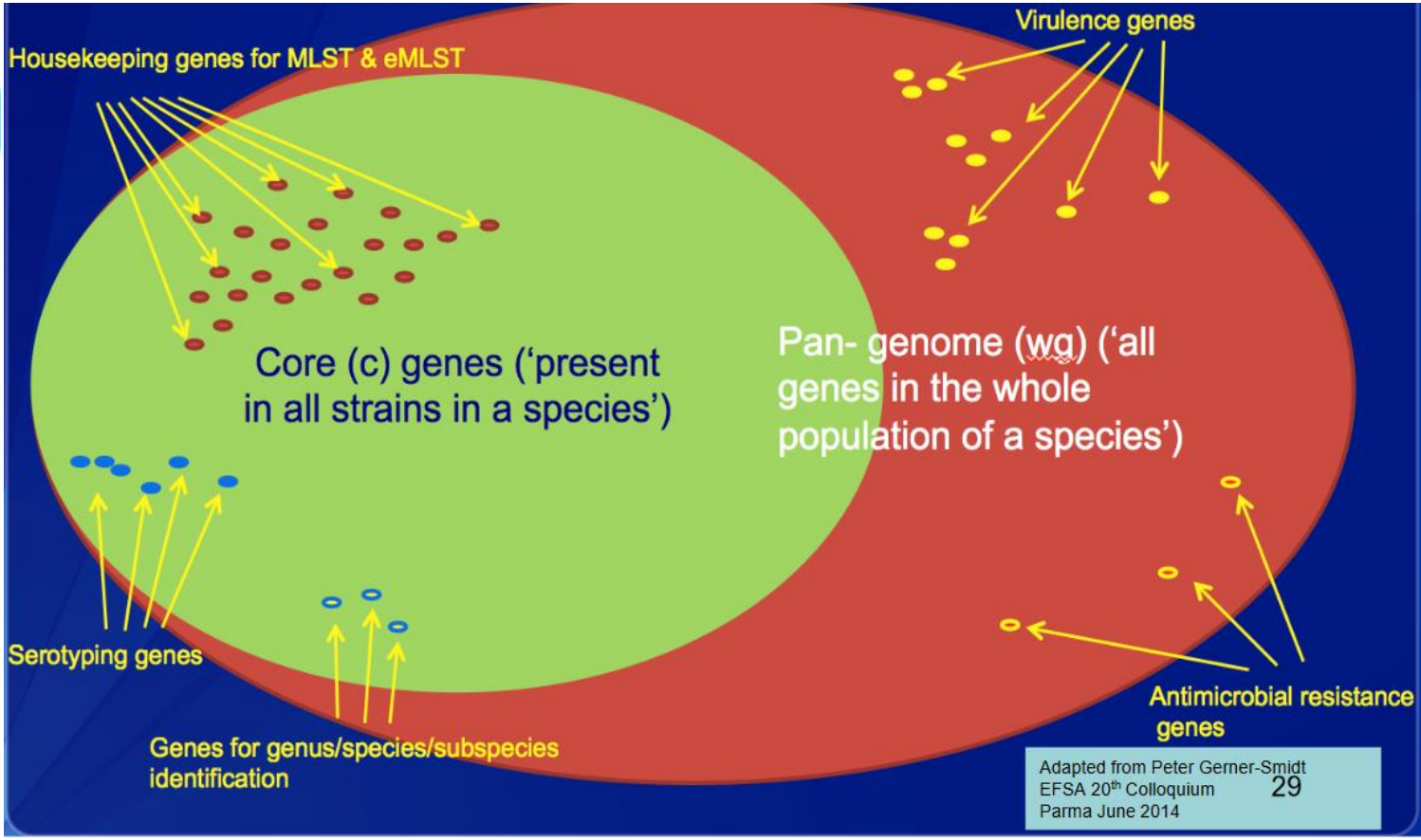
Bio-Fab research
Servizi di Supporto alla Ricerca

Resistenze e altro



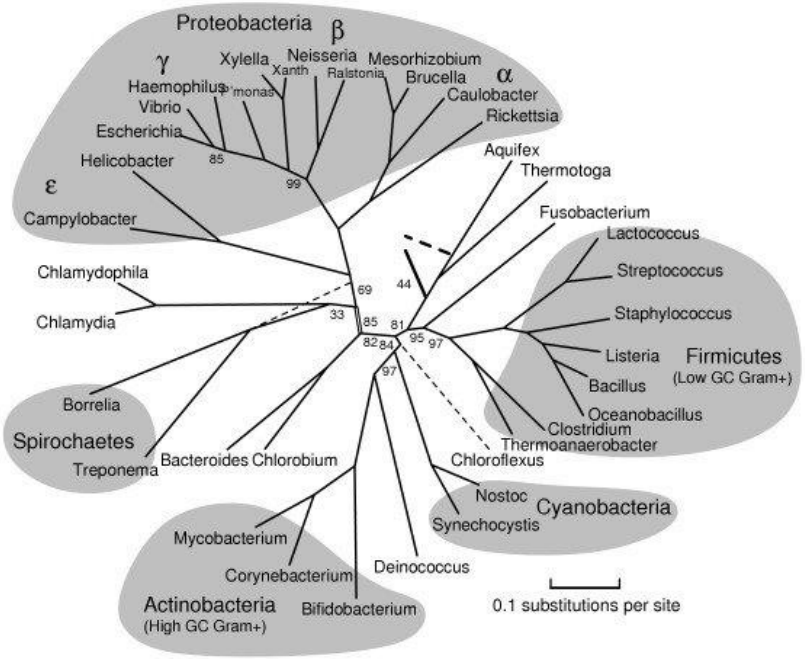
SPECIFIC GENOTYPING

- MLST in silico
- Antibiotic resistance profile
- Virulence profile
- Secondary metabolite biosynthesis profile
- PLASMID ANALYSIS
- Analysis of specific functions (enzymatic activities, regulators, transposases..)



Bio-Fab research
Servizi di Supporto alla Ricerca

Analisi filogenomica

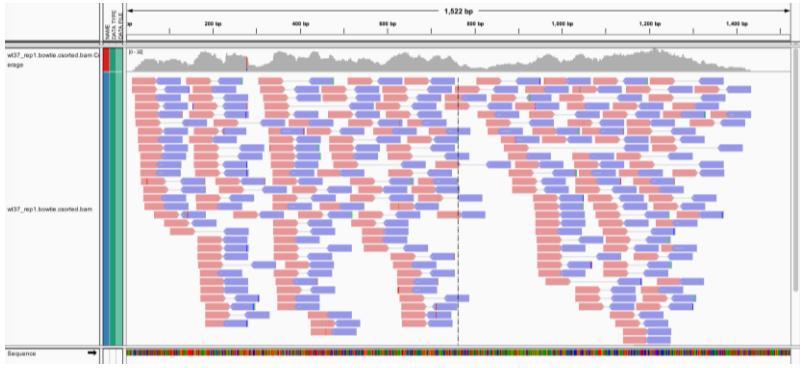


Identità fra genomi

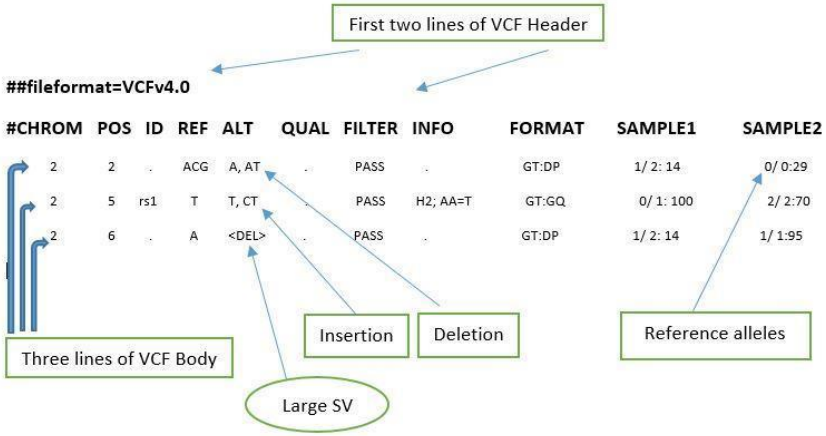
Bio-Fab research
Servizi di Supporto alla Ricerca

File di output

File di allineamento delle reads .bam/.bai

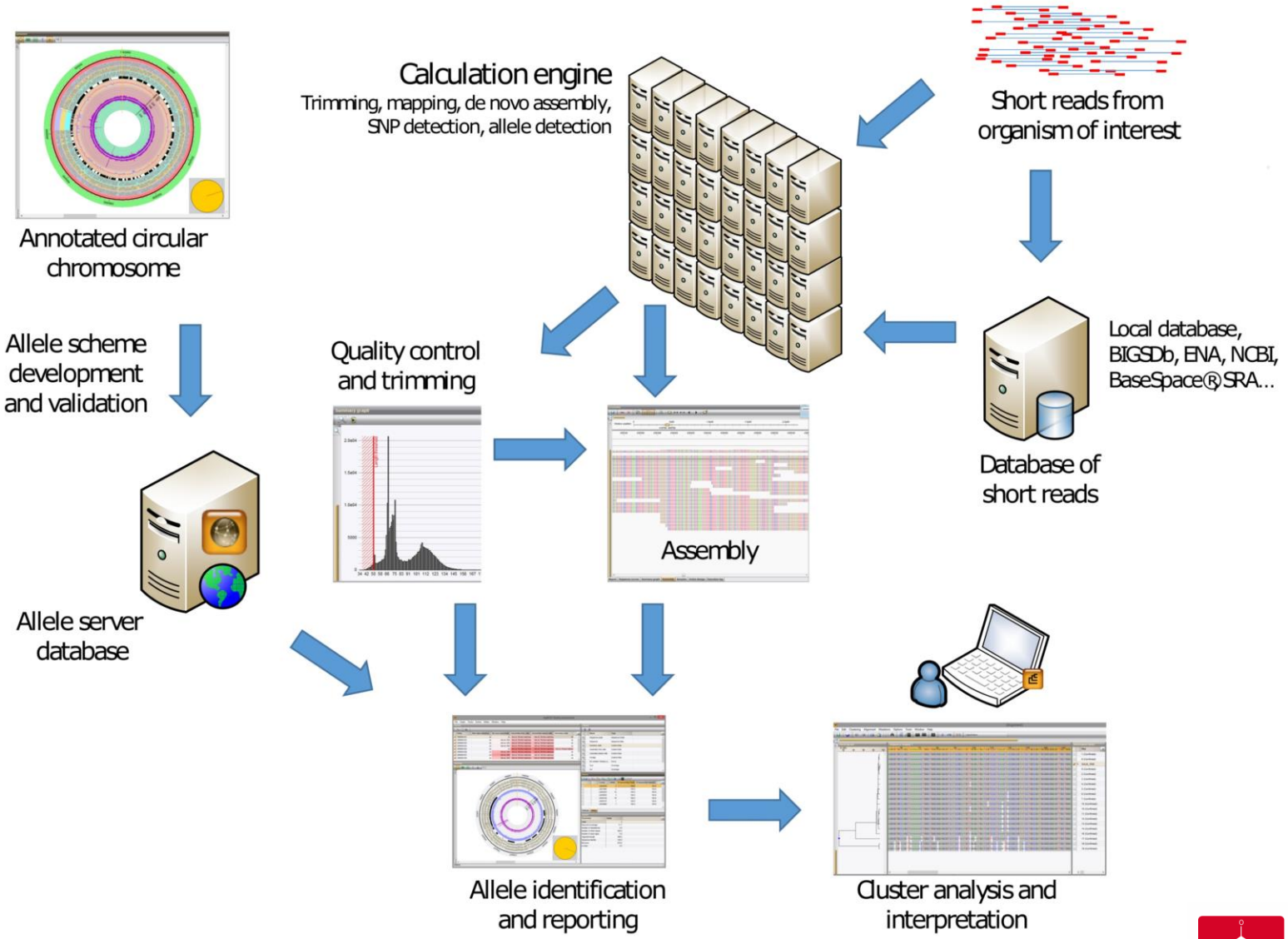


File con le varianti rispetto al reference e tra i vari allineamenti .vcf/.gff



Sequenza Consensus e allinamento del genoma .fasta

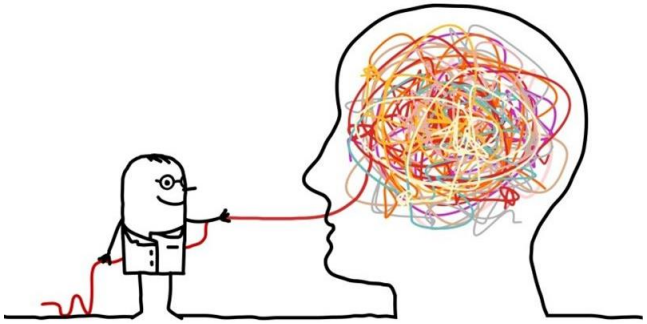
Costruzione database dedicato per velocizzare le analisi



Bio-Fab research
Servizi di Supporto alla Ricerca

Servizi di assistenza bioinformatica

Bio-Fab offre un supporto per personalizzare l'analisi bioinformatica; non solo per i dati elaborati sui nostri sequenziatori, **ma anche su dati grezzi forniti direttamente dal ricercatore.**



- **Servizio standard** Supporto bioinformatico incluso in tutte le pipeline. Viene fornita un'assistenza per la comprensione dei dati.

- **Servizio avanzato** Include la manipolazione personalizzata dei dati e la rappresentazione grafica ad-hoc nel servizio standard, con supporto per la pubblicazione dei dati.

- **Servizio top** Comprende l'assistenza per la progettazione dell'esperimento. Sviluppo di pipeline personalizzate e la valutazione di diverse metodologie per garantire la fornitura di dati affidabili per l'interpretazione biologica

Bio-Fab research
Servizi di Supporto alla Ricerca

Bio-Fab research
Servizi di Supporto alla Ricerca