



«Next Generation Sequencing e Bioinformatica: Nuovi strumenti di supporto per l'analisi di OGM in alimenti e mangimi»

16/05/2019-13/06/2019

Docente: Maria Laura De Marchis



«Next Generation Sequencing per la ricerca di OGM in ambito agroalimentare: Aggiornamenti su protocolli e flussi operativi»

06/10/2020

Docenti: Katia Spinella e Davide La Rocca



«Next Generation Sequencing (NGS) e Bioinformatica: applicazioni nel campo della sicurezza alimentare»

26/10/2021-27/10/2021

Docenti: Katia Spinella, Davide La Rocca, Valeria Russini, Bianca Maria Varcasia,
Maria Laura De Marchis



1° giornata: basi teoriche e nozionistiche

- 1) Principi di sequenziamento e piattaforme NGS - Katia Spinella
- 2) Il wetlab: protocolli analitici per preparazione delle librerie genomiche
Bianca Maria Varcasia (WGS)/ Katia Spinella (Amplicon sequencing)
- 3) Il drylab: strumenti di analisi bioinformatica

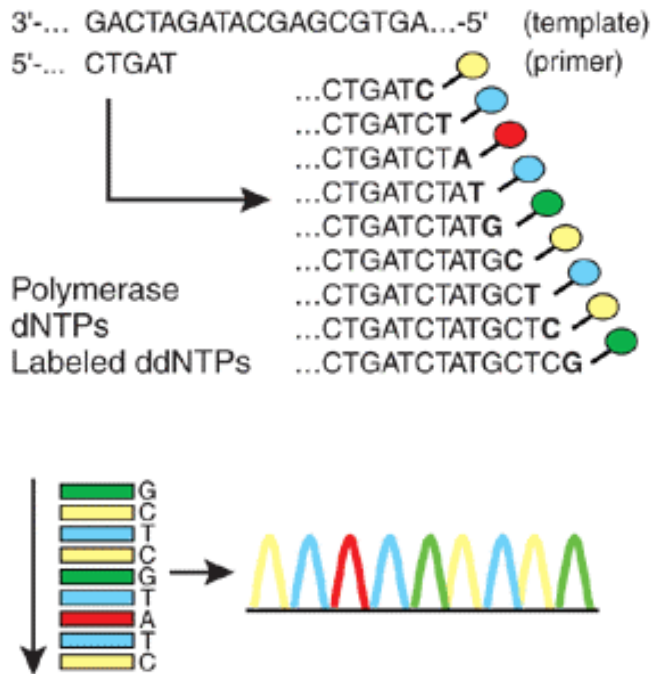
Davide La Rocca (Linux-pipeline di analisi per amplicon sequencing)/ Valeria Russini (pipeline di analisi per WGS)

2° giornata: applicazioni pratiche - «casi di studio» - aggiornamenti

- 1) Identificazione di OGM in alimenti e mangimi: aggiornamenti
Katia Spinella/ Davide La Rocca
- 2) Identificazione di specie microbiche in alimenti e ambiente
Bianca Maria Varcasia/ Maria Laura De Marchis
- 3) Epidemiologia molecolare: indagini correlate alle malattie a trasmissione alimentare
Valeria Russini/Maria Laura De Marchis



Sanger sequencing: sequenziamento di prima generazione



- Approccio in cui deossinucleotidi normali (dNTP) e dideossinucleotidi modificati marcati con 4 composti fluorescenti (ddNTP) sono aggiunti ad una reazione standard di PCR.
- Durante la fase di allungamento alcuni filamenti incorporano i ddNTP, che bloccano l'allungamento e marcano il filamento in posizione terminale.
- Questa miscela di filamenti viene separata tramite elettroforesi capillare e il nucleotide terminale di ogni filamento viene identificato tramite eccitazione con laser e analisi dell'emissione di fluorescenza

1 prodotto di sequenza/capillare ➡ 96 o 384 prodotti di sequenza per esperimento



Next Generation Sequencing (NGS)

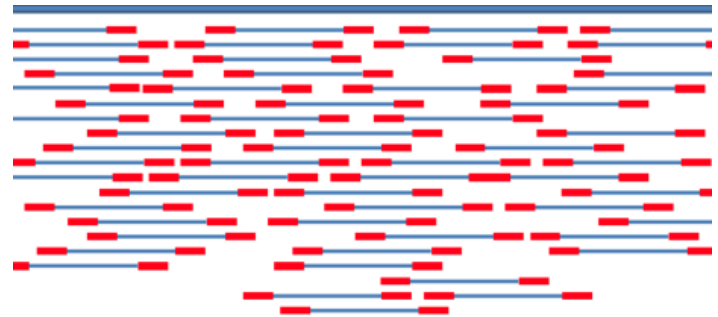
- Il termine “Next Generation Sequencing” si riferisce a un insieme di tecnologie di sequenziamento del DNA ad alta resa basate su reazioni chimiche diverse dal Sanger.
- L’NGS consente il sequenziamento massivo e in parallelo di **milioni** di frammenti di DNA in un unico esperimento.



Sequenziamento Sanger vs NGS

- Lunghezza media dei prodotti di sequenza (fino a 800 bp)
- Pochi sequenziamenti per corsa
- Costi mediamente elevati
- Basso tasso di errore (99,9% di accuratezza)
- Gold standard in determinati settori (es: ricerca clinica) necessario per validare ottenuti in NGS

- Lunghezza variabile dei prodotti di sequenza in base alla tecnologia di NGS (da alcune decine di bp a decine di Kb)
- Milioni di sequenziamenti per corsa
- Costi sempre più accessibili
- Tasso di errore variabile a seconda della tecnologia
- Ridondanza per compensare gli errori di lettura





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

Approcci di sequenziamento in NGS

DNA genomico

- **Whole genome sequencing:** Sequenziamento dell'intero genoma
- **Targeted sequencing:**
 - Amplicon sequencing (prodotti di PCR)
 - Geni mirati (es: exome sequencing)
 - Enrichment sequencing (es: DNA walking)



Approcci di sequenziamento in NGS (2)

• **Metagenoma:** contenuto in DNA totale presente in un campione complesso (es: alimento, ambiente, campione clinico...)

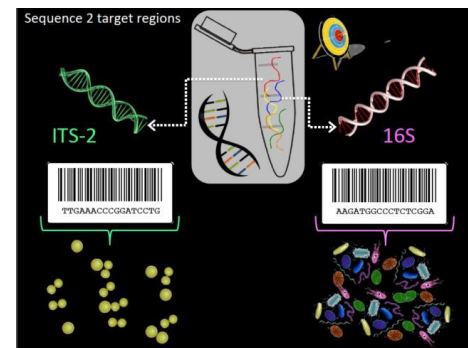
-Shotgun:

Sequenziamento «untargeted» del DNA totale estratto dal campione (WGS di comunità microbiche per analisi tassonomica e funzionale)

-**Metabarcoding:** Metodo di identificazione degli organismi appartenenti ad una comunità (microrganismi, specie vegetali, specie animali) basato sul sequenziamento di specifici marcatori amplificati mediante PCR («targeted»)

Es:

- rRNA 16S per identificazione di specie batteriche
- ITS1-2 o 18S per identificazione specie micotiche
- CO1 per identificazione specie animali
- Mat K o rbcL per identificazione specie vegetali



Approcci di sequenziamento in NGS (3)

Trascrittoma (RNA):

- RNA totale
- mRNA
- small RNA (<30 nt)

Epigenoma (Modifiche del DNA):

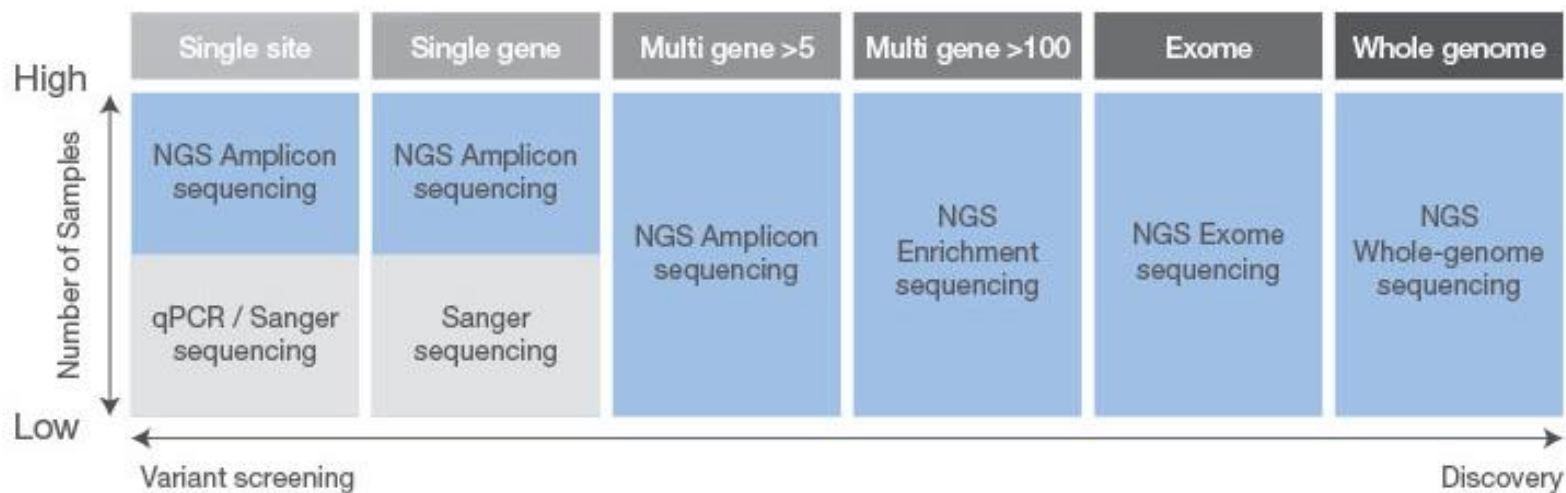
- ChIP-Seq (*Chromatin immunoprecipitation sequencing*: DNA o RNA a cui sono legati specifiche proteine)
- Metil-Seq (studio del pattern di metilazione del DNA, epigenetica)



Come scegliere la tecnologia più adatta?

Dipende da:

- Applicazione
- Numero di campioni
- Numero di target

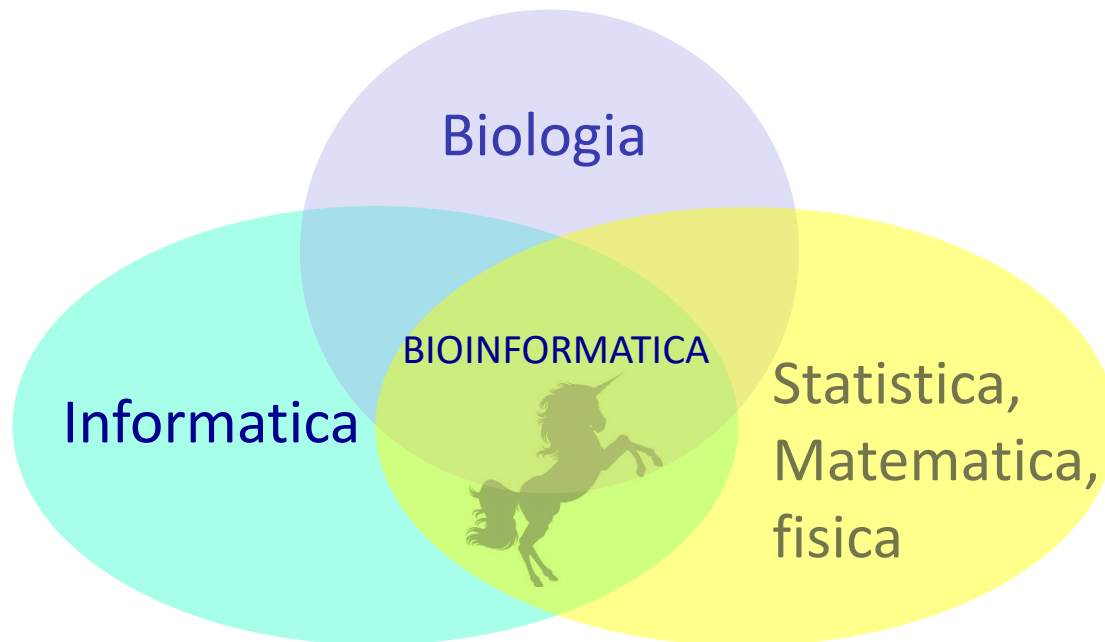




Istituto Zooprofilattico Sperimentale
del Lazio e della Toscana M. Aleandri

BIOINFORMATICA

Disciplina che affronta problematiche biologiche con metodologie e strumenti propri delle scienze dell'informazione e computazionali.



- Scienza multidisciplinare
- Nascita di una nuova professionalità
- Analisi *in silico* (DRYLAB)

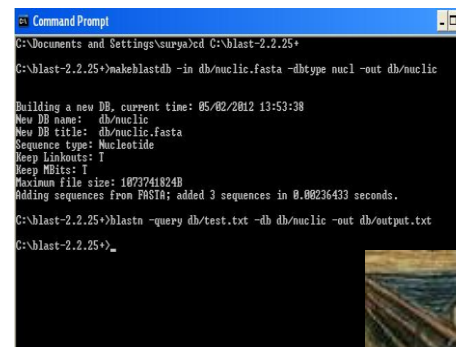
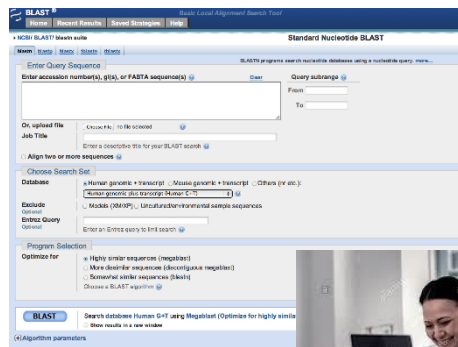




Istituto Zooprofilattico Sperimentale
del Lazio e della Toscana M. Aleandri

Strumenti di analisi bioinformatica

Acquisizione di competenze e
di personale specializzato



“PUSH THE BUTTON”



Pipeline di analisi



“Command line”



BIOINFORMATICO: sviluppo dei servizi informatici necessari per l'accesso e
l'utilizzo dei dati
Utente finale e sviluppatore





Applicazioni

- Comprendere la biologia degli organismi e delle loro comunità con un approccio integrato delle diverse scienze «-omiche»
- Correlare la sequenza, la struttura, le interazioni e le funzioni di biomolecole e loro complessi
- Ricostruzione di eventi evolutivi (analisi filogenetiche) e previsione di future modificazioni (es: machine learning)
- Dare indicazioni precise per la progettazione di molecole bioattive (industria, agricoltura, medicina...)



Analisi dei dati

La Bioinformatica deve fornire:

- strumenti per costruire modelli matematici robusti
- metodi per trattare diverse tipologie di input (algoritmi)
- linguaggi di interrogazione (automatica)
- automazione dei processi di analisi (pipeline/loop)
- consentire una analisi di dati di tipo discreto (es: dati di sequenza) e continuo (es: valori numerici)





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

Software di analisi

OPEN SOURCE

VANTAGGI

- A basso costo/gratis
- Flessibilità e possibilità di personalizzazione
- Molte possibilità di scelta

SVANTAGGI

- Molte applicazioni eseguibili solo su Linux.
- No interfaccia grafica (maggiori competenze)
- Aggiornamenti non regolari (attività sviluppatori)
- Mancanza di un supporto professionale
(FORUM!!)

PROPRIETARI

VANTAGGI

- Affidabili
- Molti servizi di supporto e di formazione disponibili
- Completi, formati modulari
- Regolarmente aggiornati
- A interfaccia grafica

SVANTAGGI

- Costosi
- Meno flessibili



Origine dei dati

DATABASE

- Archivi di informazioni biologiche
- Archivi di informazioni derivate
- Metadati (dati descrittivi, gestionali, strutturali)

DATI PRODOTTI *IN HOUSE*



WETLAB





Istituto Zooprofilattico Sperimentale
del Lazio e della Toscana M. Aleandri

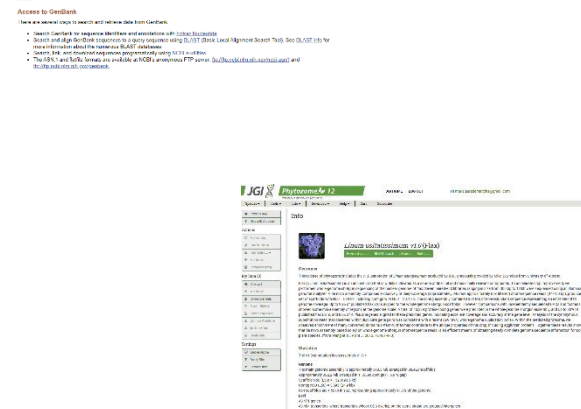
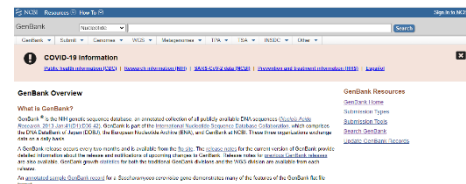
DATABASE BIOLOGICI



DATABASE PRIMARI

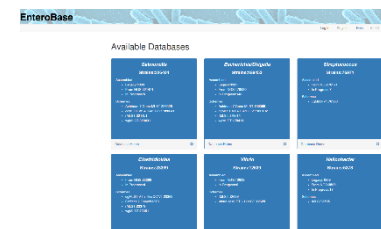
- Archivi di dati grezzi, sequenze annotate, sequenze genomiche, sequenze proteiche...
- Alimentati esclusivamente da chi sottomette il dato (laboratori di sequenziamento, pubblicazioni)
- Contengono i link per i database secondari

<https://www.ncbi.nlm.nih.gov/genbank/>



DATABASE SECONDARI

- Archivi di informazioni derivate
- Dati curati
- Metadati
- Contengono tool e pipeline di analisi specifiche
- Possono essere pubblici (registrazione) o privati (con accesso limitato)



Istituto Superiore di Sanità



Processamento e storage dei dati



IZSLT:

RAM: 500 Gb con 64 threads

Spazio di archiviazione: 9 Tb

OS: Ubuntu 20.04

- Potenza di calcolo
- Elevato numero di CPU /threads
- Elevata RAM
- Accesso ad un server per processare e conservare i dati
- Centri di calcolo (es:CINECA)



Attività di NGS e bioinformatica presso l'IZSLT

**ISTITUTO ZOOPROFILATTICO SPERIMENTALE
DEL LAZIO E DELLA TOSCANA M. ALEANDRI**

DELIBERAZIONE DEL DIRETTORE GENERALE

n. 282..... del 3/08/2020

OGGETTO: Organizzazione delle attività di Sequenziamento massivo (High Throughput Sequencing, HTS) e dell'attività bioinformatica

- Coordinamento delle attività di HTS e delle relative attività bioinformatiche della UOC D.O. Diagnostica Generale, attraverso la sua articolazione UOS Diagnostica e Caratterizzazione Molecolare, NRL-AR e CRN-AR
- Gruppo di lavoro costituito dalla UOC Ufficio di Staff Ricerca, Innovazione e Cooperazione Internazionale e dalle strutture coinvolte in prima istanza nelle attività di HTS e bioinformatica:
 - UOC DO Diagnostica generale
 - UOC DO Virologia
 - UOC DO Microbiologia degli alimenti
 - UOSD Ricerca e Controllo degli Organismi Geneticamente Modificati



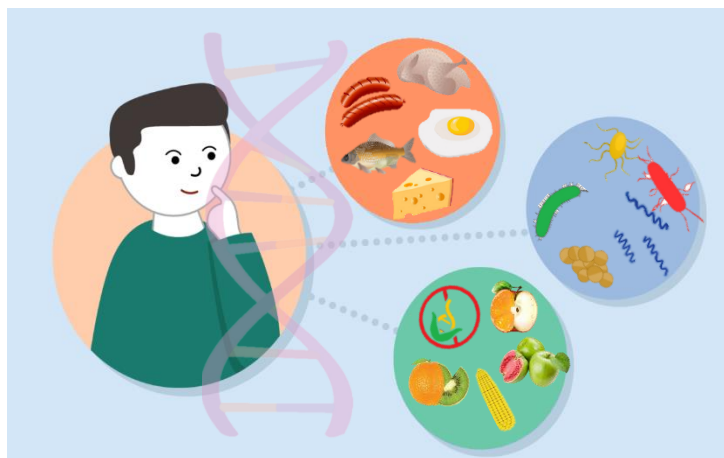


Aspetti trattati nella delibera

- Addestramento e formazione continua del personale coinvolto nelle attività di wetlab e drylab
- Organizzazione del lavoro (acquisto reagenti, impiego sequenziatore NGS, fruizione del servizio di sequenziamento in out-sourcing)
- Organizzazione area bioinformatica (coordinamento attività, utilizzo del server centrale di calcolo, acquisto di software proprietari)
- Rapporti IZSLT con altri Enti



NGS e bioinformatica nel campo della sicurezza alimentare



Tutto ciò che possiede un genoma è potenzialmente trattabile con strumenti di HTS e con tool di analisi bioinformatica

Uno sguardo alla letteratura...



Ambiti applicativi di NGS e bioinformatica nel campo della sicurezza alimentare

Strumenti di screening per la caratterizzazione della **composizione** di un alimento o di un mangime

- Ricerca degli allergeni in alimenti e mangimi

Wang Y, Zhou J, Peng H, Ma J, Li H, Li L, Li T, Fang Z, Ma A, Fu L. **High-Throughput Identification of Allergens in a Food System via Hybridization Probe Cluster-Targeted Next-Generation Sequencing.**

J Agric Food Chem. 2021 Oct 13;69(40):11992-12001

- Identificazione di specie in alimenti altamente processati

Muñoz-Colmenero M, Martínez JL, Roca A, Garcia-Vazquez E.

NGS tools for traceability in candies as high processed food products: Ion Torrent PGM versus conventional PCR-cloning.

Food Chem. 2017 Jan 1;214:631-636. doi: 10.1016/j.foodchem.2016.07.121.





Istituto Zooprofilattico Sperimentale
del Lazio e della Toscana M. Aleandri

Ambiti applicativi di NGS e bioinformatica nel campo della sicurezza alimentare (2)

Strumenti di identificazione e caratterizzazione per la **tracciabilità** di un alimento o di un mangime

- Tracciabilità geografica

Milan M, Maroso F, Dalla Rovere G, Carraro L, Ferraresso S, Patarnello T, Bargelloni L, Cardazzo B, Fariselli P. **Tracing seafood at high spatial resolution using NGS-generated data and machine learning: Comparing microbiome versus SNPs.** Food Chem. 2019 Jul 15;286:413-420. doi: 10.1016/j.foodchem.2019.02.037.

- Autenticazione di specie per il controllo delle frodi alimentari

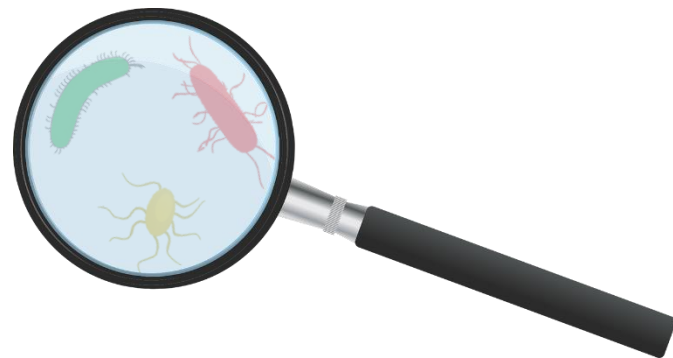
Kappel K, Haase I, Käppel C, Sotelo CG, Schröder U. **Species identification in mixed tuna samples with next-generation sequencing targeting two short cytochrome b gene fragments.** Food Chem. 2017 Nov 1;234:212-219.

Cavin C, Cottenet G, Cooper KM, Zbinden P. **Meat Vulnerabilities to Economic Food Adulteration Require New Analytical Solutions.** Chimia (Aarau). 2018 Oct 31;72(10):697-703.



Ambiti applicativi di NGS e bioinformatica nel campo della sicurezza alimentare (3)

Strumento di screening e detection per la caratterizzazione della **composizione** di un alimento o di un mangime



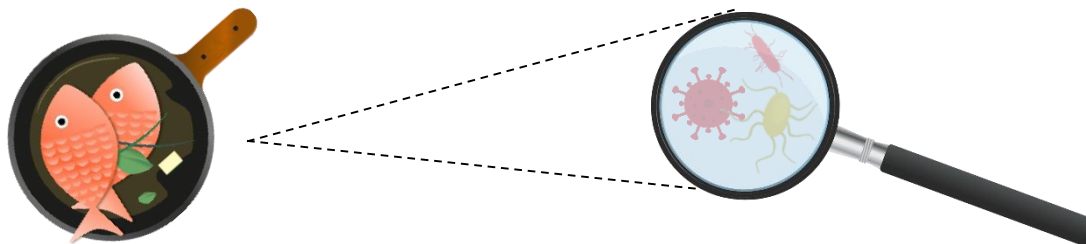
- Ricerca di OGM di origine vegetale non autorizzati in alimenti e mangimi
- Ricerca di microrganismi geneticamente modificati (MOGM) in alimenti e mangimi



Ambiti applicativi di NGS e bioinformatica nel campo della sicurezza alimentare (4)

Strumento per la detection di **specie contaminanti** (virus e batteri)

- Ricerca di patogeni a trasmissione alimentare negli alimenti e l'ambiente

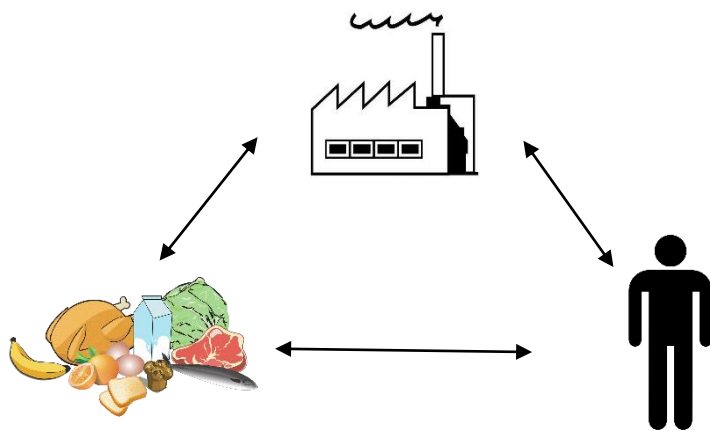


Jagadeesan B, Gerner-Smidt P, Allard MW, Leuillet S, Winkler A, Xiao Y, Chaffron S, Van Der Vossen J, Tang S, Katase M, McClure P, Kimura B, Ching Chai L, Chapman J, Grant K.
The use of next generation sequencing for improving food safety: Translation into practice. Food Microbiol. 2019 Jun;79:96-115.



Ambiti applicativi di NGS e bioinformatica nel campo della sicurezza alimentare (5)

Strumento per la **subtipizzazione** di patogeni a trasmissione alimentare isolati da alimenti, uomo e ambiente



- Attività di sorveglianza
- Caratterizzazione di patogeni a trasmissione alimentare nel corso di indagini epidemiologiche (individuazione di cluster epidemici, source attribution)



Dall'immensamente piccolo all'immensamente grande



GENOMI VIRALI

- genomi di piccole dimensioni (3500 bp-1,2 Mbp)
- a RNA o DNA
- a singolo o doppio filamento
- elevata variabilità nella composizione genica
- elevatissimi tassi di mutazione:
(RNA-virus 10^{-3} - 10^{-5} /bp/generazione
DNA-virus 10^{-6} - 10^{-8} /bp/generazione)



GENOMI BATTERICI

- genomi di piccole dimensioni (0,2-9 Mbp)
- assetto aploide
- elementi extracromosomici (plasmidi)
- elevata variabilità nella composizione genica (geni accessori)
- tasso di mutazione (0,003 mut/genoma/ generazione cellulare)



GENOMI VEGETALI

- genomi di grandi dimensioni (300-17000 Mbp)
- presenza di regioni ripetute
- fenomeno della poliploidia (errori meiotici, ibridi interspecifici)



Le variabili da considerare per la scelta della strategia di wetlab e di drylab da utilizzare sono molteplici:

- ✓ Specie indagata (batterio, virus, fungo, animale, vegetale...)
- ✓ Dimensioni target (intero genoma, esoma, amplicone...)
- ✓ Caratteristiche target (RNA, DNA...)
- ✓ Applicazione (screening, detection, tipizzazione...)

Procedure
operative
standard

Linee guida
Specialist

Forum
Dati di letteratura

