

ALMA MATER STUDIORUM
UNIVERSITA' DI BOLOGNA
SCUOLA DI AGRARIA E MEDICINA VETERINARIA

Corso di laurea magistrale in
Scienze e Tecnologie Alimentari

**STUDIO DEL PROFILO MICROBICO E DI
GENI FUNZIONALI IN CARCASSE DI BROILER
MEDIANTE
ANALISI METAGENOMICA SHOTGUN**

Tesi di laurea in
“Tracciabilità batterica e sicurezza alimentare”
(Innovazioni e Sicurezza dei prodotti di origine animale, C.I.)

Relatore
Prof. Gerardo Manfreda

Presentata da
Dott. Simone Mazzini

Correlatore
Dott.ssa Alessandra De Cesare

Sessione II
Anno Accademico 2016/2017

INDICE

Glossario	5
Introduzione	7
CAPITOLO 1 - IL SEQUenziAMENTO DEL DNA	9
1.1 Il genoma batterico.....	9
1.1.1 Acidi nucleici	10
1.1.2 Il DNA.....	11
1.1.3 Replicazione del DNA	12
1.2 Tecnologie di sequenziamento del DNA	14
1.2.1 Sequenziamento di I generazione: il metodo Sanger.....	15
1.2.2. NGS – Sequencing by synthesis	18
1.2.2.1 Il sequenziatore 454 Roche Pyrosequencer	18
1.2.2.2 Il sequenziatore Helicos tSMS	22
1.2.2.3 Solexa Illumina	24
1.2.3 Tecnologie tra la seconda e terza generazione - Sequencing by ligation.....	27
1.2.3.1 Il sequenziatore SOLiD di Applied Biosystems	27
1.2.3.2 Ion Torrent Life Technologies	30
1.2.4 Tecnologie di Terza Generazione	31
1.2.4.1 Sequencing by synthesis: SMRT (Single Molecule Real Time sequencing) Pacific Biosciences.....	31
1.2.4.2 Sequencing by ligation: Oxford Nanopore Technologies MinION	32
1.3 Differenze tra le tecnologie di sequenziamento	34
1.4 Sequenziamento applicato allo studio degli alimenti.....	35
CAPITOLO 2 - METAGENOMICA	36
Introduzione	36
2.1 Genomica e metagenomica	37
2.2 Shotgun metagenomic sequencing.....	39
2.2.1 Identificazione della diversità tassonomica	41
2.2.1.1 Geni marker.....	42
2.2.1.2 Binning.....	43
2.2.1.3 Assemblaggio.....	44
2.2.2 Comprensione delle funzioni biologiche	45
2.5 MG-RAST.....	48
2.6 MGmapper	50
2.7 Applicazioni di analisi metagenomiche agli alimenti	51

2.8 Workflow della analisi metagenomica.....	54
2.8.1 Estrazione e purificazione dell'acido nucleico	55
2.8.2 Preparazione delle Library	56
2.8.2.1 Preparazione della library tramite Nextera®XT	58
2.8.3 Sequenziamento	61
2.8.4 Analisi dei dati	62
CAPITOLO 3 - SCOPO DELLA RICERCA.....	65
CAPITOLO 4 - MATERIALI E METODI	67
4.1 Animali dai quali sono state ottenute le carcasse da analizzare.....	67
4.2 Raccolta ed analisi delle carcasse	69
4.2.1 Estrazione del DNA dalla pelle di collo ottenuta dalle carcasse.....	69
4.3 Preparazione delle library per il sequenziamento	70
4.4 Sequenziamento	72
4.5 Analisi delle sequenze.....	72
4.5.1 MG-RAST.....	75
4.5.2 MGmapper	75
4.5.3 Analisi statistica - STAMP	75
CAPITOLO 5 - RISULTATI.....	76
5.1 DNA estratti	76
5.2 Parametri delle library e dei metagenomi ottenuti	77
5.3 Tabelle con i risultati tassonomici	79
5.4 Tabelle con i risultati metabolici.....	95
CAPITOLO 6 - DISCUSSIONE.....	98
CAPITOLO 7 - REFERENCES	102

GLOSSARIO

Abundance	È la rappresentazione relativa una particolare specie all'interno di un ecosistema. Solitamente si misura come il numero di individui presenti in un campione. È un indicatore importante per il calcolo della biodiversità
Adattatore	Un oligonucleotide sintetico a doppio filamento usato per permettere di legare fra loro due molecole
Annealing	Attacco di un oligonucleotide a una molecola di DNA a singolo filamento mediante ibridazione
bp	Unità di misura della lunghezza di un frammento di un acido nucleico a doppio filamento. Indica il numero di basi appaiate (base pair) che si susseguono
Complementare	Il DNA complementare è un DNA a doppia elica sintetizzato a partire da una molecola di mRNA, ad opera dell'enzima trascrittasi inversa.
Denaturazione termica	Rottura dei legami H coinvolti nell'appaiamento delle basi, tramite somministrazione di calore
DNA	L'acido nucleico contiene le informazioni genetiche necessarie alla sintesi di RNA e proteine. È un polimero i cui monomeri sono i nucleotidi (A, G, C e T), l'ordine cui i monomeri sono disposti in sequenza costituisce l'informazione genetica. È strutturato in forma di coppia di filamenti complementari fra loro, detta doppia elica. Negli eucarioti, il DNA si trova all'interno del nucleo cellulare, in strutture dette cromosomi
Flow-cell (piastra di sequenziamento)	Il substrato solido su cui sono posti i campioni di DNA da sequenziare. I campioni vengono preparati e posti sulla piastra di sequenziamento, che viene poi inserita all'interno del sequenziatore, dove avviene il processo di sequenziamento vero e proprio. Può essere divisa in lane, indipendenti fra loro, permettendo il sequenziamento di campioni indipendenti in un'unica run. Alcuni sequenziatori sono in grado di sequenziare in parallelo due flow-cell per run
Lane	Regioni di una flow-cell indipendenti fra loro. Possono essere utilizzate per sequenziare in parallelo più campioni diversi o uno stesso campione ma sequenziato in più volte indipendenti. Generalmente, una flow-cell contiene otto lane
Ligasi	Enzima che ripara nella cellula discontinuità a singolo filamento in molecole di DNA a doppio filamento. Viene usata nel clonaggio dei geni per unire molecole di DNA
nt	Unità di misura della lunghezza di un frammento di acido nucleico a singolo filamento. Indica il numero di nucleotidi che si susseguono
Nucleotidi (dNTP)	Sono monomeri degli acidi nucleici (DNA e RNA), composti da una base azotata, uno zucchero a cinque atomi di C e un gruppo fosfato. I nomi dei nucleotidi sono abbreviati in codici standard di tre lettere, precedute da una "d" nel caso dei nucleotidi del DNA. I nucleotidi sono adenina (dATP), guanina (dGTP), citosina (dCTP) e timina (dTTP). Nell'RNA la timina è sostituita dall'uracile (UTP)

INTRODUZIONE

La scoperta della struttura del DNA è stata una delle più importanti scoperte scientifiche del ventesimo secolo. A metà del '900 lo studio di Erwin Chargaff, insieme agli studi condotti da Rosalind Franklin e Maurice Wilkins, permisero a James Watson e Francis Crick di formulare il modello a doppia elica del DNA pubblicato sulla rivista scientifica Nature il 25 Aprile 1953. Da quel momento, i ricercatori iniziarono a indagare la sequenza del DNA e contribuirono allo sviluppo e messa a punto della tecnologia di sequenziamento di Allan Maxam e Walter Gilbert (1973), basata su modificazioni chimiche del DNA e sul conseguente taglio in posizioni specifiche, e successivamente quella di Frederick Sanger, basata sull'introduzione di terminatori di catena (1975) (Wang Z. *et al.*, 2013; Johnsen JM. *et al.*, 2013). Negli anni '90 e nei primi dieci anni del 2000, parte della comunità scientifica si è dedicata all'identificazione di metodiche di sequenziamento di nuova generazione (NGS) caratterizzate da una più elevata produttività e da costi più contenuti rispetto al sequenziamento Sanger di prima generazione. Sono state quindi sviluppate tecnologie basate su principi molto diversi tra loro, che hanno la possibilità di sequenziare ampi tratti del genoma o interi genomi con tempi e costi ridotti rispetto al sequenziamento tradizionale di prima generazione. (Koboldt DC. *et al.*, 2013; Nguyen L. *et al.*, 2014; Lapunzina P. *et al.*, 2014; Mori A. *et al.*, 2013; Zhou X. *et al.*, 2014; Johansen Taber K.A. *et al.*, 2014; Rabbani B. *et al.*, 2014).

Principali tecnologie NGS

Solexa Illumina

Il sistema Solexa Illumina prevede la frammentazione del DNA da sequenziare. Alle 2 estremità dei frammenti vengono legati degli adattatori che ne permettono l'ancoraggio ad una superficie piana fungendo così da sito di legame per i primer della PCR. L'amplificazione dei frammenti viene effettuata mediante bridge PCR. Vengono così generati dei cluster contenenti circa 1000 copie dello stesso frammento da sequenziare. L'amplificazione mediante bridge PCR è necessaria per rendere il segnale abbastanza forte da poter essere catturato dalla CCD (charge-coupled device) camera. La Mix di reazione per il sequenziamento viene depositata sulla superficie del supporto e contiene i primers, 4 nucleotidi marcati rispettivamente con 4 fluorofori in grado di interrompere temporaneamente la sintesi e la DNA polimerasi. Una volta incorporato il nucleotide avviene la lettura; successivamente viene eliminato il fluoroforo ed il gruppo terminatore al 3' (che impedisce l'attacco di un altro nucleotide) e il ciclo si ripete.

454 Roche Pyrosequencer

Il sistema 454 Roche è basato sulla metodica della PCR in emulsione e del pirosequenziamento, una tecnica di rilevazione della chemiluminescenza derivante dal rilascio del pirofosfato a seguito

dell'incorporazione di deossinucleotidi trifosfato (dNTPs) ad opera di una DNA-polimerasi (sequenziamento contemporaneo e parallelo di frammenti di oltre 400 basi).

Supported Oligonucleotide Ligation and Detection (SOLiD) Life Technologies

Il sistema SOLiD è basato sul sequenziamento di “short reads” di DNA o frammenti di DNA legati a oligonucleotidi adattatori, immobilizzati su biglie e amplificati clonalmente mediante PCR in emulsione. Le biglie con i prodotti di amplificazione clonale vengono fissate sulla superficie di vetro di una cella a flusso, sulla quale ha luogo la reazione di sequenziamento mediante reazioni di ligazione (sequenziamento contemporaneo e parallelo di frammenti circa 35 basi).

Tecnologie NGS terza generazione

Negli ultimi anni sono state presentate ulteriori piattaforme per il sequenziamento del DNA, dette di terza generazione; queste permettono un sequenziamento molto rapido senza richiedere amplificazione tramite reazione PCR del campione. Tra le più promettenti la tecnologia Smrt della Pacific Bioscience e le due piattaforme Nanopore (GridION e MinION) della Oxford Nanopore Technologies. Il sequenziamento del DNA attraverso nanopori offre grandi vantaggi rispetto ad altre metodiche, il DNA viene fatto passare attraverso un nanoporo, inserito all'interno di una membrana lipidica con una differenza di potenziale ai due lati, e ogni base viene letta nel punto più stretto del poro sfruttando la corrente ionica che passa al suo interno per determinare l'identità della base. Sulla flow-cell che viene alloggiata in un semplice supporto della grandezza di una chiavetta USB vi sono più di 500 nanopori che, lavorando contemporaneamente, consentono il sequenziamento di lunghi filamenti di DNA (circa 15.000 bp). (Pepe G. *et al.*, 1997; Pepe G. *et al.*, 2001; Giusti B. *et al.*, 2003; Attanasio M. *et al.*, 2008; Pepe G. *et al.*, 2014).

Considerazioni

Le tecnologie di sequenziamento ad elevata produttività, oltre al sequenziamento di porzioni o interi genomi, forniscono opportunità senza precedenti per la ricerca scientifica e la diagnostica; ad oggi, infatti, esse sono state applicate in una grande varietà di contesti: *metagenomica*, trascrittomica ed epigenetica. Una delle criticità che sono state affrontate in questi anni di enorme diffusione delle tecnologie di sequenziamento è quella relativa allo sviluppo di metodiche computazionali adatte alla analisi di grosse mole di dati (Abel HL. *et al.*, 2013; Del Chierico F. *et al.*, 2015; Magi A. *et al.*, 2010; Magi A. *et al.*, 2013; Magi A. *et al.*, 2014; Pippucci T. *et al.* 2014).

1 IL SEQUENZIAMENTO DEL DNA

1.1 Il genoma batterico

Il genoma batterico è suddiviso in cromosoma batterico ed elementi accessori. Il cromosoma batterico è di forma circolare, unico (infatti i batteri sono cellule aploidi) e dimensioni variabili comprese tra 15000 e 9000000 bp e contiene un gene ogni 1000 bp. Contrariamente al genoma di una cellula eucariote, non possiede introni ed esoni.

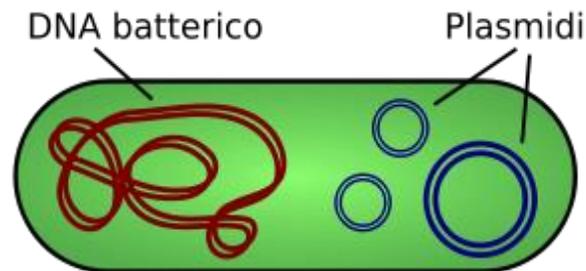


Figura 1.1 Cellula procariote

<http://testdimedicina.altervista.org/blog/la-cellula-procariote-le-proprie-caratteristiche/>

Il genoma della specie è la sommatoria di tutti i geni dei membri di quella specie. All'interno di ogni specie ci sono sia geni conservati (o costitutivi) sia geni accessori. I geni conservati (housekeeping genes) caratterizzano la specie e servono alla cellula per svolgere le funzioni (replicazione del DNA, pathway metabolici). I geni accessori sono posseduti da quasi tutti i membri della specie, i geni ceppo-specifici sono peculiari di ogni ceppo.

Il pool genetico può essere suddiviso in:

- **Core gene pool:** responsabile dell'identità della specie (housekeeping genes) stabile nel tempo e che evolve molto lentamente per a seguito di mutazioni;
- **Flexible gene pool:** responsabile della diversità delle specie (geni accessori), instabile nel tempo ed ha evoluzione rapida a seguito di acquisizione/perdita di geni.

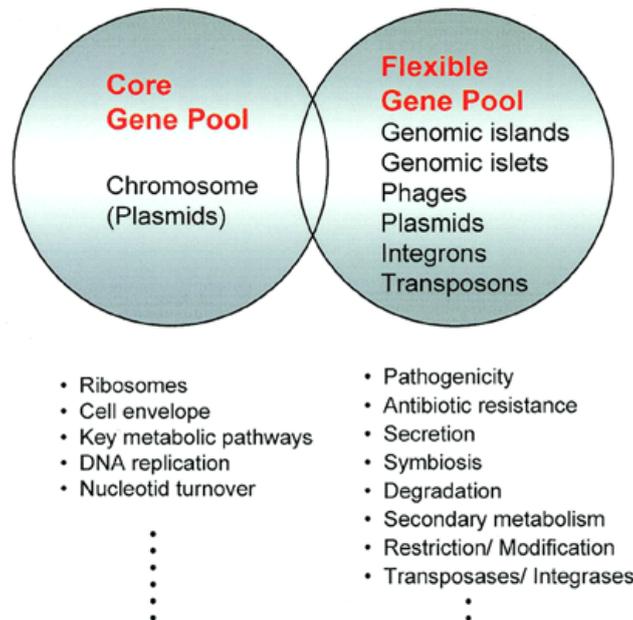


Figura 1.2 DNA nei genomi procarioti

<http://embor.embopress.org/content/2/5/376.Figuraures-only>

1.1.1 Acidi nucleici

Esistono due tipi di acidi nucleici: DNA e RNA. Le unità strutturali degli acidi nucleici sono i nucleotidi, i quali sono caratterizzati da tre componenti: una base azotata, contenente azoto e a struttura eterociclica; uno zucchero pentoso e infine un gruppo fosfato. Quest'ultimo è un gruppo funzionale chimico costituito dall'anione fosfato (PO_4^{3-}) legato a una molecola organica tramite uno dei suoi quattro ossigeni. La molecola formata da zucchero pentoso e base azotata prende il nome di nucleoside. I nucleotidi sono legati tra loro tramite "ponti" covalenti, in cui il gruppo fosfato legato al carbonio dello zucchero in posizione 5' di un'unità nucleotidica reagisce con il gruppo ossidrilico (-OH) legato al carbonio in posizione 3' del nucleotide successivo, formando così un legame fosfodiesterico. Lo scheletro degli acidi nucleici è costituito da un'alternanza di gruppi fosforici e di residui di zucchero pentoso, mentre le basi azotate possono essere considerate come gruppi laterali uniti allo scheletro a intervalli regolari. I gruppi ossidrilici degli zuccheri formano legami idrogeno con l'acqua e per questo lo scheletro di DNA e RNA risulta essere idrofilico (Nelson D.L. *et al.*, 2015). I gruppi fosforici dello scheletro polare sono completamente ionizzati e carichi negativamente a pH 7. Tutti i legami fosfodiesterici delle catene hanno lo stesso orientamento lungo la catena, determinando per ogni acido nucleico una specifica polarità, ed estremità 3' e 5' distinte (Sinden, R. R. 2012).

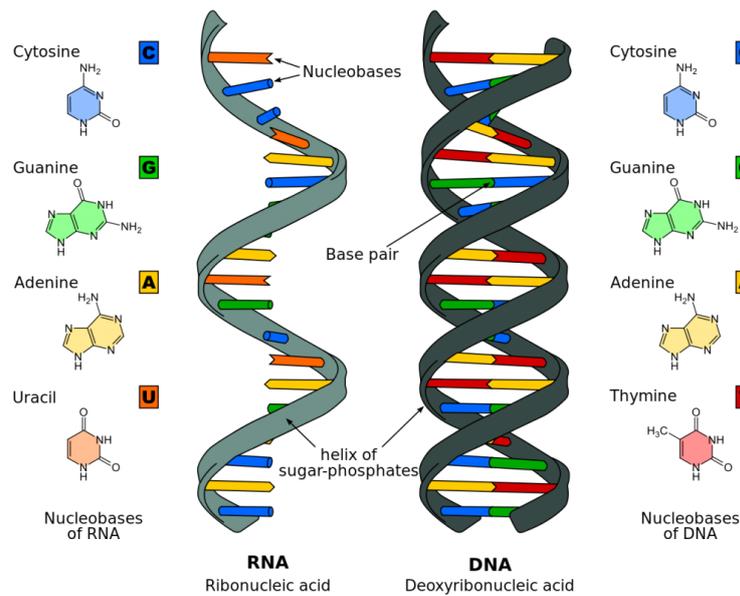


Figura 1.3 Composizione e struttura di RNA e DNA
[\(https://futurism.com/study-shows-why-we-are-made-of-dna-and-not-rna/\)](https://futurism.com/study-shows-why-we-are-made-of-dna-and-not-rna/)

1.1.2 Il DNA

L'acido desossiribonucleico o deossiribonucleico (DNA) è un acido nucleico che ha la funzione di codificare e trasmettere l'informazione genetica. Infatti contiene le informazioni indispensabili per la biosintesi delle proteine (Nelson D.L. *et al.*, 2015).

Chimicamente il DNA è un polimero di nucleotidi legati tra loro. L'ordine nella disposizione sequenziale dei nucleotidi costituisce l'informazione genetica (gene), la quale è tradotta nelle macromolecole che realizzano la specifica funzione codificata, cioè le proteine. I nucleotidi del DNA sono caratterizzati da uno zucchero pentoso, il 2'-deossi-ribosio, chiamato così per la mancanza di un atomo di ossigeno in posizione 2', unito a una delle quattro basi azotate tipiche del DNA e a un gruppo fosfato. Le basi azotate possono essere suddivise in *basi puriniche* e *basi pirimidiniche*, a seconda della struttura biciclica o monociclica che posseggono. Infatti le purine sono basi con struttura ciclica a due anelli, mentre le pirimidine hanno una struttura ciclica con un solo anello. Sono basi puriniche l'*adenina* e la *guanina*, indicate rispettivamente con (A) e (G), mentre la *timina* (T) e la *citocina* (C) rappresentano le basi pirimidiniche del DNA (Nelson D.L. *et al.*, 2015).

Il biochimico Erwin Chargaff, stabilì che le molecole di DNA hanno composizioni in basi distinte. I risultati ottenuti portarono Chargaff a concludere che in tutte le molecole di DNA il numero di residui di adenina è pari al numero di residui di timina ($A=T$); e analogamente il numero dei residui di guanina e citosina coincidono ($G=C$). Ciò porta a concludere che la somma dei residui purinici è pari a quella dei residui pirimidinici: ($A+G=T+C$).

Dal modello messo a punto da Watson e Crick nel 1953 si deduce che il DNA è costituito da due catene elicoidali avvolte attorno allo stesso asse per formare una doppia elica destrorsa. Lo scheletro

covalente idrofilico, costituito da un'alternanza di deossiribosio e gruppi fosforici è all'esterno della doppia elica, mentre le basi puriniche e pirimidiniche sono impilate all'interno della doppia elica, e le loro strutture planari ad anello sono poste in posizione perpendicolare al lungo asse longitudinale della molecola. Si crea una relazione spaziale tra le due catene che genera una scanalatura maggiore e una scanalatura minore sulla superficie della doppia elica. La scanalatura maggiore è larga 22Å, e quella minore 12Å (Watson J.D. *et al.*, 2009).

La differente ampiezza dei due solchi determina una differente accessibilità delle basi. I fattori proteici che regolano il processo della trascrizione genica prendono contatto con le basi presenti nel solco maggiore (Nelson D.L. *et al.*, 2002). I due filamenti costituenti la doppia elica hanno un'estensione antiparallela, ovvero il senso di un filamento è opposto a quello dell'altro filamento, ed entrambi sono orientati in direzione 5'-3'.

Le due catene polinucleotidiche, antiparallele, sono tra loro complementari. Le basi puriniche formano legami idrogeno con le basi pirimidiniche: A può legare solo T e G può legare solo C, con due o tre legami idrogeno rispettivamente. Questi legami idrogeno che si formano implicano la complementarietà tra le catene di DNA. L'accoppiamento di due basi viene comunemente chiamato "paio di basi"; esso costituisce l'unità di misura maggiormente utilizzata per definire la lunghezza di una molecola di DNA.

Un'altra forza che mantiene unite le due catene è dovuta all'interazione tra le basi impilate. Queste interazioni non sono specifiche ma sono coinvolte nella stabilità della doppia elica. Dato che i legami idrogeno sono legami deboli, essi possono esser rotti e riuniti in modo relativamente semplice. I due filamenti possono essere separati tra loro (processo di denaturazione del DNA) sia utilizzando alte temperatura che sfruttando un'azione meccanica. La temperatura alla quale il 50% del DNA è denaturato è detta "temperatura di melting" (T_m) ed è funzione della sequenza di nucleosidi; infatti se sono presenti molte coppie (C)-(G) la T_m risulterà più alta, rispetto a una sequenza con più coppie (A)-(T), perché nel primo caso è necessario rompere tre legami a idrogeno, rispetto ai due del secondo. Indicativamente la T_m di una molecola di DNA è circa uguale a 90°C, con l'abbassamento di temperatura le basi si riappaiano (Sinden R. R., 2012).

1.1.3 Replicazione del DNA

La replicazione del DNA viene definita semiconservativa: le due molecole di DNA prodotte conterranno un filamento parentale, che ha svolto la funzione di stampo, e un filamento ad esso complementare neosintetizzato. Questa ipotesi, proposta da Watson e Crick, fu poi dimostrata con diversi esperimenti dagli studiosi Meselson e Stahl nel 1957. La replicazione è un processo

coordinato, in cui le catene di DNA parentali vengono svolte, separate e replicate simultaneamente, a partire da siti specifici chiamati forcelle di replicazione. La sintesi di una nuova catena di DNA procede sempre in direzione $5' \rightarrow 3'$. Poiché le due eliche di DNA sono antiparallele l'elica usata come stampo viene letta dall'estremità $3'$ verso quella $5'$.

Secondo Reiji Okazaki *et al.*, 1960 una delle due nuove catene di DNA viene sintetizzata sotto forma di brevi frammenti, chiamati “frammenti di Okazaki”, quindi una viene sintetizzata in modo continuo mentre l'altra in modo discontinuo. La catena continua, chiamata anche catena veloce, è quella in cui la sintesi procede nella stessa direzione del movimento della forcella di replicazione, seguendo anche il senso $5' \rightarrow 3'$. L'altra catena, chiamata anche catena lenta, è quella in cui la sintesi procede in senso opposto rispetto al movimento della forcella (Sinden R. R., 2012).

Il DNA viene sintetizzato grazie a un enzima chiamato DNA polimerasi, scoperto nel 1955 da Arthur Kornberg. La DNA polimerasi catalizza la polimerizzazione dei nucleotidi, uno alla volta, in corrispondenza del carbonio in posizione $3'$ del nucleotide della nuova molecola. Gli studi sulla DNA polimerasi portarono a comprendere la necessità di avere delle molecole d'innescò per il processo di sintesi, queste molecole sono chiamate primer. Un primer è un piccolo segmento di acidi nucleici, complementare alla catena stampo, con un gruppo ossidrilico $3'$ libero a cui si possano legare i nucleotidi. I primer sono spesso oligonucleotidi, composti da RNA, sintetizzati da enzimi particolari (Nelson D.L. *et al.*, 2002).

Dopo l'aggiunta di un nucleotide alla catena di DNA neosintetizzata, la DNA polimerasi si muove lungo il filamento stampo e aggiunge un ulteriore nucleotide. Il numero medio di nucleotidi aggiunti prima che la polimerasi si dissocia dal DNA determina un parametro chiamato processività. Esistono in natura diverse DNA polimerasi, ognuna caratterizzata da una propria processività. La replicazione procede per fasi successive. È possibile suddividere la sintesi di una molecola di DNA in tre fasi: inizio, allungamento e terminazione. Queste tre fasi sono distinguibili sia per la reazione che per gli enzimi che intervengono (Nelson D.L. *et al.*, 2002).

Inizialmente si ha l'apertura della doppia elica a causa della rottura dei legami idrogeno operata da enzimi specifici. La denaturazione dei filamenti agisce specialmente nelle regioni ricche di coppie (A)-(T). Le due catene quindi si separano lasciando disaccoppiate le basi azotate di ciascuna di esse. Nella fase di allungamento si ha il processo di sintesi della catena lenta e della catena veloce. Infine si ha il processo di terminazione in cui le forcelle di replicazione si incontrano in una regione terminale, dove è presente una sequenza di basi particolare chiamata Ter. Il processo di replicazione è molto accurato e gli errori sono minimi; infatti anche se un nucleotide sbagliato potrebbe riuscire a formare legami idrogeno con un'altra base nello stampo, in realtà non riesce ad entrare nel sito attivo

della DNA polimerasi, quindi i nucleotidi sbagliati vengono solitamente eliminati prima che si crei il legame fosfodiesterico. Nonostante i minimi errori (un nucleotide sbagliato su 100 milioni appaiati correttamente) possono capitare degli accoppiamenti sbagliati che provocano un cambiamento nella sequenza delle basi azotate. Se il cambiamento è replicato, e trasmesso alle cellule future, si parla di mutazioni e la modifica risulta permanente (Nelson D.L. *et al.*, 2015).

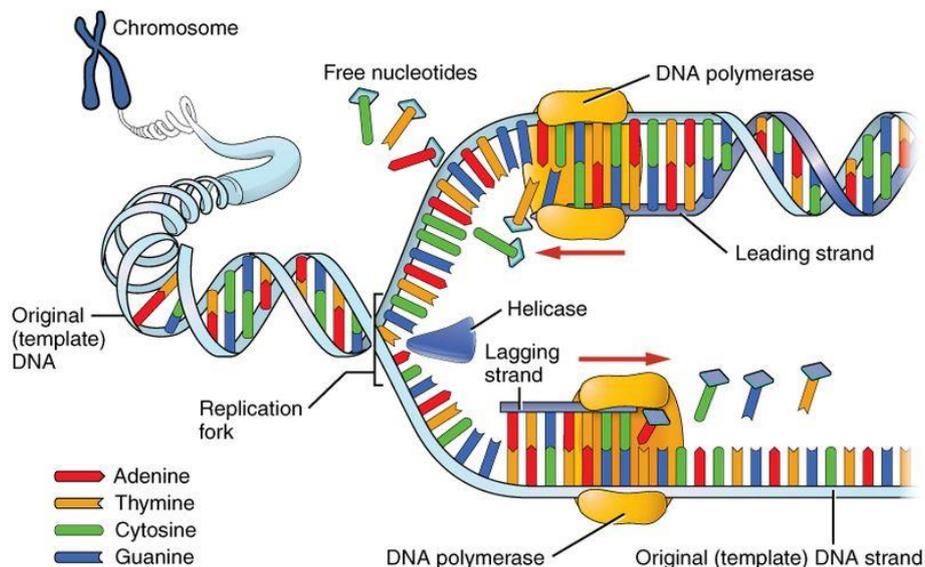


Figura 1.4 Replicazione del DNA

<http://www.newence.com/2017/04/05/nuove-prospettive-nei-meccanismi-cellulari-del-cancro-il-ruolo-degli-errori-da-replicazione-del-dna/>

1.2 Tecnologie di sequenziamento del DNA

Le tecniche NGS hanno promosso nuove tecnologie high-throughput, come la genomica, metagenomica, trascrittomica e metatracrittomica che se confrontate ai metodi precedenti, permettono analisi più massive degli acidi nucleici che consentono di ottenere una descrizione più dettagliata degli ecosistemi microbici. Queste tecnologie possono essere utilizzate in due modi sostanziali: tramite il sequenziamento di tutto l'acido nucleico microbico (*shotgun sequencing*) o tramite il sequenziamento di uno specifico gene (*targeted sequencing*). Per il secondo, segmenti di DNA o cDNA estremamente conservato vengono prima amplificati tramite PCR utilizzando primer universali o specifici. Lo *shotgun sequencing* fornisce informazioni oltre che della composizione filogenetica, anche del numero e la potenziale funzione di specifici geni (Wilmes P. *et al.*, 2008, Solieri L. *et al.*, 2013).

Il termine *sequenziamento*, in biologia molecolare indica il processo per la determinazione dell'esatta struttura primaria di un biopolimero e cioè dell'ordine delle basi nel caso di un *acido nucleico* o degli amminoacidi nel caso di proteine. È il processo che permette di determinare l'esatta sequenza delle

basi nucleotidiche, cioè conoscere l'ordine in cui si susseguono i nucleotidi che compongono il DNA contenuto nel campione in esame. I moderni sviluppi della tecnologia hanno permesso di ottenere strumentazioni (i sequenziatori appunto) in grado di sequenziare in automatico anche genomi complessi. Generalmente, i sequenziatori non sono in grado di sequenziare il DNA input per intero, ma “leggono” delle porzioni di sequenza chiamate “frammenti”, la cui lunghezza dipende dalla tecnologia utilizzata. Una volta avvenuto il ciclo di sequenziamento, lo strumento fornisce delle “reads” che permettono di identificare l'ordine in cui si susseguono le basi nei frammenti di DNA. Sia lo *shotgun sequencing* che il *targeted sequencing* sono state utilizzate in microbiologia per analizzare le fermentazioni in alimenti (Ercolini D., 2013), tuttavia, le ricerche e le applicazioni potenziali di queste metodologie sono in costante aggiornamento.

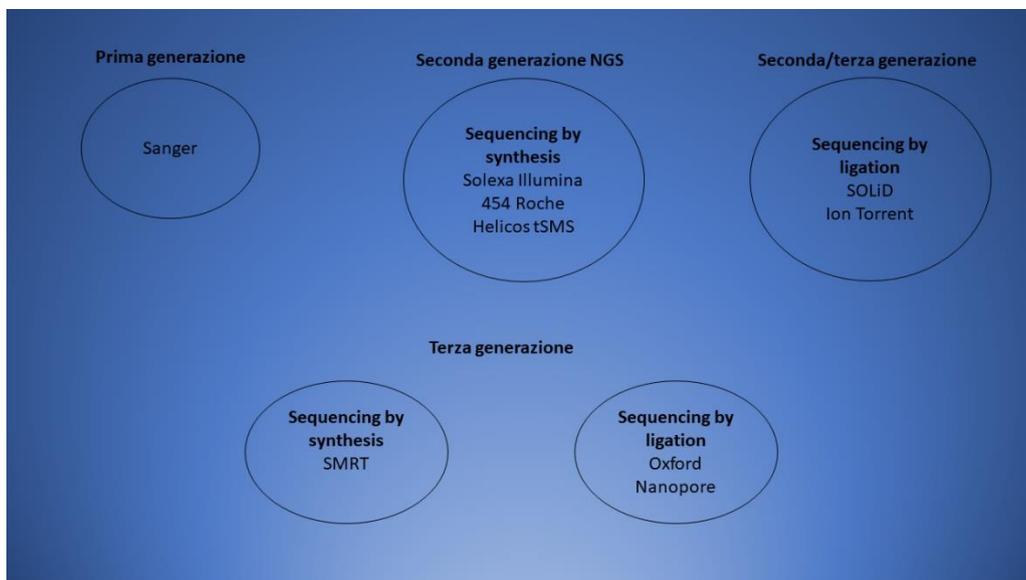


Figura 1.5 Schema delle tecniche di sequenziamento

1.2.1 Sequenziamento di I generazione: il metodo Sanger

Negli anni 70 Sanger portò una svolta nella microbiologia molecolare in quanto, insieme al suo team inventò il metodo a cui venne attribuito il suo stesso nome, per sequenziare il DNA, detto anche “sequenziamento a terminazione della catena” (Sanger F. *et al.*, 1977). Il metodo Sanger è tutt’oggi utilizzato in alcuni ambiti, come i test di paternità e la biologia forense.

Il metodo di sequenziamento del DNA con dideossinucleoteterminatori o metodo di Sanger consiste di 3 fasi: la preparazione del campione, la reazione di sequenziamento, l’elettroforesi (Sanger F. *et al.*, 1977).

Nella prima fase di preparazione del campione, il filamento di DNA che si vuole sequenziare viene amplificato artificialmente, in modo da ottenere diverse copie identiche dello stesso. La tecnica che permette questo processo di duplicazione è la PCR (*Polymerase Chain Reaction*).

Grazie al processo di reazione a catena della polimerasi (PCR), è possibile produrre milioni di copie di una quantità molto ristretta di DNA iniziale. È dunque una tecnica più rapida rispetto al DNA ricombinante, sebbene necessiti delle sequenze nucleotidiche (primer) che possano appaiarsi correttamente con le estremità del segmento da amplificare. La PCR è un processo ciclico, nella prima parte i frammenti di DNA a doppio filamento vengono denaturati attraverso il calore (circa 95°C). Alla miscela viene aggiunto un primer sintetizzato chimicamente (alla temperatura di 50-60°C per favorire il legame del primer col filamento). Vengono aggiunti i quattro deossinucleotidi trifosfato (dNTPs) e la DNA polimerasi, che catalizza la produzione di nuovi filamenti complementari a quelli da copiare. Un singolo ciclo di reazioni impiega pochi minuti per raddoppiare la quantità di DNA iniziale, e il DNA neosintetizzato viene a trovarsi nello stato a doppia elica. Ripetendo il ciclo di reazioni (denaturazione, primer annealing, estensione del filamento) molte volte si ottiene un aumento esponenziale del numero di copie di DNA iniziale. Durante la fase di reazione di sequenziamento, il campione biologico viene sottoposto a quattro processi: denaturazione, primer annealing, estensione del filamento, terminazione. Con la denaturazione i singoli filamenti di DNA vengono separati. Il primer annealing è la fase in cui viene aggiunto un primer all'estremità 3' di uno dei due filamenti. Il primer è sintetizzato artificialmente e appositamente per la sequenza di DNA da sequenziare. Vengono allestite quattro miscele in quattro provette, una per ogni base. In ogni provetta viene aggiunta la DNA polimerasi (fase della copia del filamento), i quattro nucleotidi (dATP, dCTP, dGTP, dTTP), e una piccola quantità di un dideossinucleoside trifosfato (ad esempio ddATP). Un dideossinucleoside trifosfato (ddNTP), è un nucleotide che non ha il gruppo -OH in posizione 3' dello zucchero (Sanger F. *et al.*, 1977).

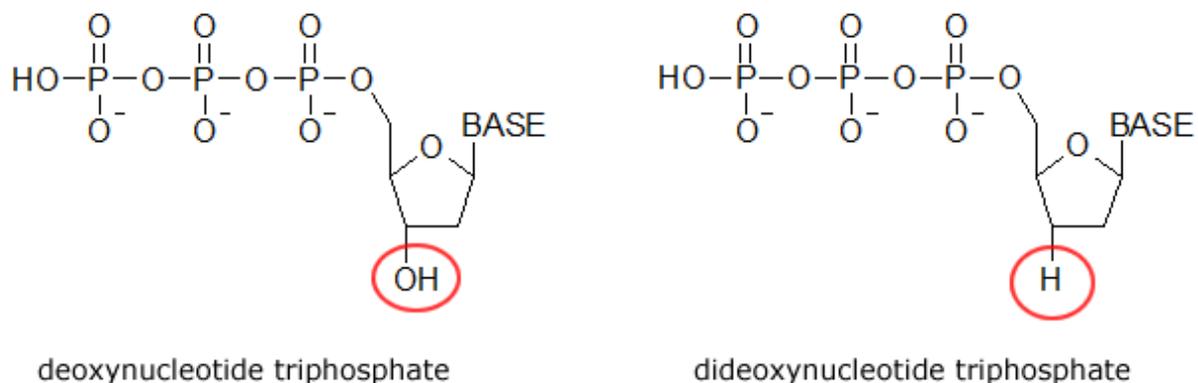


Figura 1.6 Rappresentazione di un deossinucleoside trifosfato, (a sinistra), e di un dideoxinucleoside trifosfato, (a destra). Il dideoxinucleoside trifosfato è un nucleotide che non ha il gruppo -OH in posizione 3' dello zucchero (<http://biochem.co/wp-content/uploads/2008/08/deoxynucleotide-dideoxynucleotide.png>)

Questi comunque potranno esser aggiunti dalla DNA polimerasi a un filamento di DNA in corso di sintesi mediante la formazione di un legame fosfodiesterico tra il suo 5'-fosfato e il 3'-OH del residuo precedente. Tuttavia, poichè i ddNTPs mancano del gruppo -OH in posizione 3', il nucleotide successivo non potrà esser legato come avviene nella replicazione naturale del DNA. Per questo motivo la sintesi si arresta alla posizione in cui un ddNTP è stato incorporato all'estremità in accrescimento di un filamento di DNA (fase di terminazione). A differenza di quanto accade nella PCR questa volta viene copiato solo il filamento specifico per il primer utilizzato, in direzione 5'-3'. La replicazione del DNA procede e nella provetta si viene a trovare una miscela di filamenti stampo del DNA insieme ad una varietà di filamenti neosintetizzati più brevi. I filamenti nuovi, ognuno dei quali termina con un ddATP, per l'esempio riportato, avranno lunghezze differenti. Una volta che la DNA polimerasi incontra una base T sul filamento stampo, essa potrà aggiungere o un dATP o un ddATP. Se viene aggiunto un dATP la crescita del filamento continua, mentre se viene aggiunto un ddATP, la crescita del filamento si arresta. Questo processo viene ripetuto in altre provette rispettivamente per il ddGTP, ddTTP, ddCTP. Dopo aver fatto proseguire per un pò la replicazione del DNA, i filamenti neosintetizzati vengono denaturati e tramite opportune tecniche di laboratorio separati dai filamenti stampo. A questo punto il preparato è pronto per la successiva fase e viene sottoposto ad elettroforesi. L'elettroforesi è un processo elettrocinetico nel quale molecole e particelle cariche, sotto l'influenza di un campo elettrico, migrano in direzione del polo che ha carica opposta. Grazie alla presenza dei gruppi fosfato, le molecole di DNA sono cariche negativamente e quindi migrano verso il polo positivo (anodo) se sottoposte a un campo elettrico, con velocità che dipende dalla loro lunghezza.

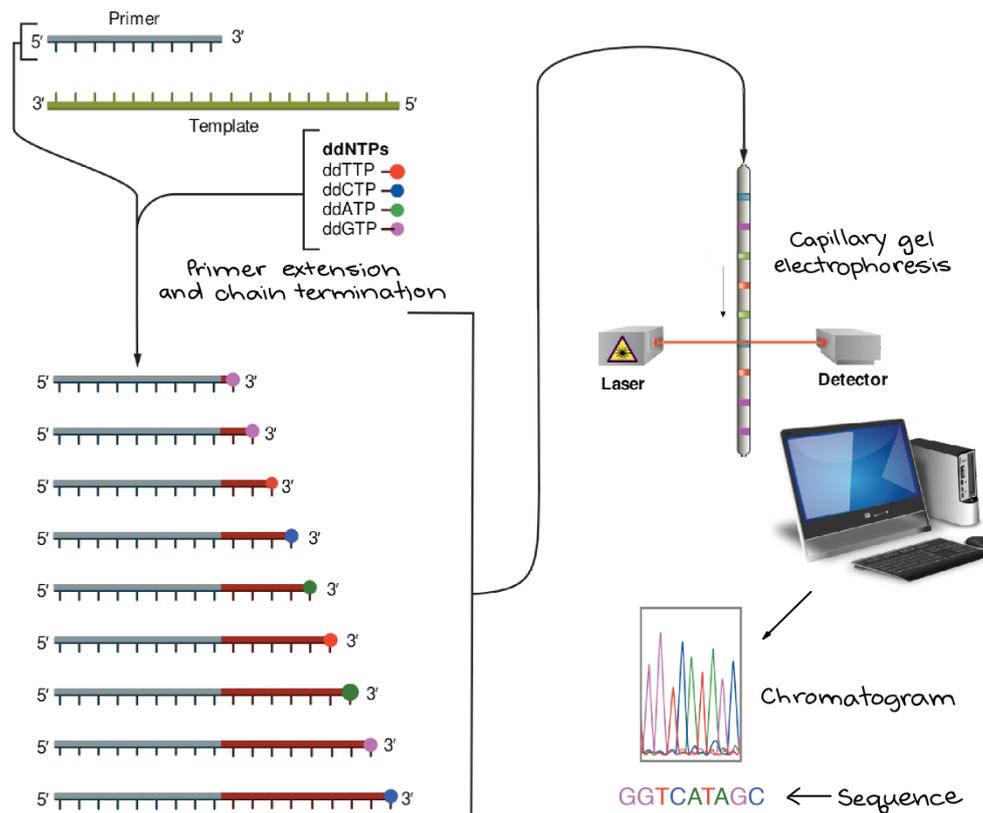


Figura 1.7 Schematizzazione del metodo Sanger

(<https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/dna-sequencing>)

Nel sequenziamento automatico si può allestire una singola reazione inserendo i quattro ddNTPs marcati mediante l'incorporazione di un composto fluorescente, diverso per ogni base, l'elettroforesi avviene lungo dei capillari, all'interno dei quali è sempre ricostruita la rete di gel o in polimero. Durante la corsa elettroforetica, i frammenti vengono letti in ordine di lunghezza crescente da un fascio laser che eccita i marcatori fluorescenti. Quindi l'intensità della luce emessa viene misurata e tale informazione, cioè quale colore di fluorescenza e dunque il tipo di ddNTP è presente all'estremità di ogni filamento di differente lunghezza, viene inviata ad un computer. L'introduzione dell'elettroforesi capillare per la separazione dei frammenti marcati ha consentito un notevole aumento della processività. Sono stati inoltre sviluppati modelli di sequenziatori automatici che sono in grado di eseguire corse elettroforetiche multiple su apparecchi multicapillari (Sanger F. *et al.*, 1977).

1.2.2. NGS – Sequencing by synthesis

1.2.2.1 Il sequenziatore 454 Roche Pyrosequencer

Il sequenziatore Genome Sequencer, commercializzato nel 2005 da 454 Science, ha rappresentato la prima piattaforma di sequenziamento NGS disponibile sul mercato.

La preparazione del campione avviene tramite una frammentazione random del DNA in frammenti di 300 - 800 bp (Margulies M. *et al.*, 2006; Myllykangas S. *et al.*, 2012). Il primo step consiste in un'amplificazione PCR della regione di DNA d'interesse. Durante questa fase si utilizzano dei primer speciali, detti fusion primer, che oltre alla sequenza target per il gene di interesse, hanno due adattatori per il sequenziamento 454 Roche "A" e "B", una breve sequenza chiave (key) e una sequenza "barcode" identificativa del campione (MID) (Figura 1.8). Una volta ottenuti gli ampliconi è necessario effettuare il pooling dei campioni in esame creando un'unica library. Tenendo in considerazione i MID utilizzati in amplificazione, il numero di reads per campione che si vuole ottenere dalla run di sequenziamento e la concentrazione finale, i campioni sono miscelati in modo equimolare fino ad ottenere un unico campione.

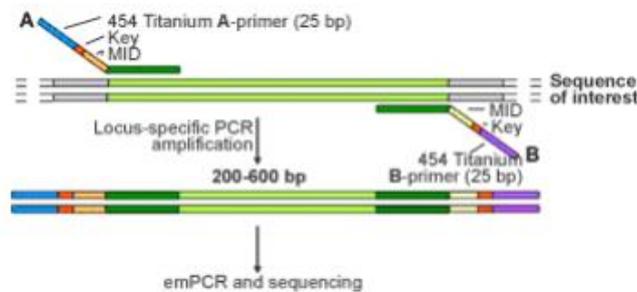


Figura 1.8 Sequenza dei fusion primer

La library di ampliconi è sottoposta ad una seconda amplificazione in una miscela d'olio (Figura 1.9) che, emulsionata, crea delle microgocce che costituiscono dei micro reattori all'interno dei quali è contenuta una sola Capture Bead e tutti i reagenti per l'amplificazione tra cui gli enzimi di amplificazione. La biglia è legata attraverso l'adattatore "B" ad una sola molecola di DNA stampo. Le sfere legate al filamento si trovano quindi nella fase acquosa dell'emulsione di acqua in olio per far avvenire l'amplificazione clonale tramite PCR (PCR a emulsione, em-PCR) (Figura 1.9) Si tratta di un processo ciclico in cui la DNA polimerasi sintetizza il filamento complementare, si denatura la doppia elica e si sintetizzano due nuovi filamenti complementari, fino ad ottenere decine di milioni di copie del frammento per ogni sfera (Shendure J. *et al.*, 2008).

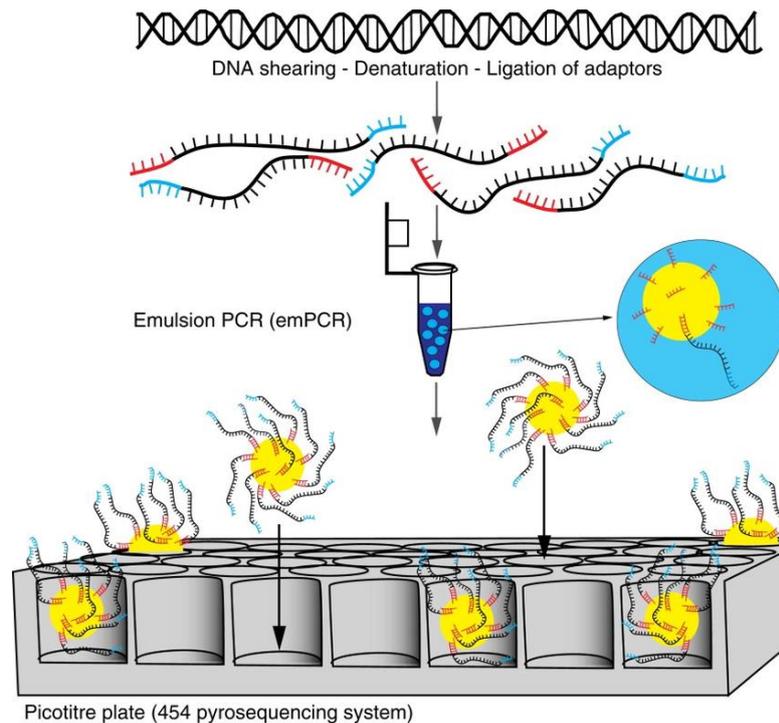


Figura 1.9 Sequenziamento 454, fasi di preparazione della library, PCR in emulsione, caricamento PTP e sequenziamento
(Wood, S. A. et al., 2013)

Al termine del processo in ogni goccia saranno presenti da 10 a 50 milioni di copie di dsDNA per biglia. Dopo il recupero dall'emulsione, per eliminare le biglie che non hanno formato il legame con il DNA, alla reazione sono aggiunti un primer di arricchimento biotinilato (Enrichment Primer) complementare all'adattatore "A" e delle biglie magnetiche ricoperte di streptavidina. Le biglie senza DNA sono eliminate separandole da quelle legate alle biglie magnetiche (che hanno formato un legame streptavidina-biotina con il primer di arricchimento e quindi con le Capture Beads) con un magnete. Avviene quindi la denaturazione del legame tra il doppio filamento con il distacco del primer di arricchimento legato alle biglie magnetiche e il recupero delle Capture Beads con il DNA, che a questo punto è a singolo filamento.

Il pirosequenziamento, prevede il caricamento sulla superficie di una piastra di vetro (Pico Titer Plate) delle biglie cariche di DNA, assieme agli enzimi solforilasi e luciferasi (Enzyme Beads), le PPiase (PPiase Beads), che eliminano il pirofosfato inorganico (PPI), e le Packing Beads, le quale stabilizzano e mantengono tutti i componenti del sistema immobilizzati entro i pozzetti della piastra PTP durante il sequenziamento.

Le sfere ricoperte dai nuovi filamenti sono poste sulla piastra di sequenziamento. Nel sequenziatore 454 Roche, la piastra (detta PicoTiterPlate) ha circa 1.5 milioni di pozzetti, il cui diametro (44 μm) permette di accogliere una sola sfera. Le sfere vi si depositano per gravità e sono ricoperte da numerose sfere più piccole legate a solforilasi. Genome Sequencer utilizza infatti il metodo del

pirosequenziamento (Marguiles M. *et al.*, 2005) che prevede il completamento del filamento e la simultanea lettura del segnale emesso dai nucleotidi incorporati. È un processo ciclico, in cui un singolo dNTP viene aggiunto di volta in volta alla reazione. Se il dNTP è complementare al filamento stampo, la sua incorporazione causa il rilascio di un pirofosfato che viene prelevato dalla sulfurilasi e convertito in ATP. L'ATP ottenuto innesca la reazione luminosa della luciferasi e il segnale luminoso emesso viene captato e misurato da una camera CCD. In base al numero di nucleotidi incorporati sono lasciati più pirofosfati e quindi più molecole di ATP, generando una bioluminescenza più intensa. Misurata l'intensità luminosa è dunque possibile determinare il numero di nucleotidi di uno stesso tipo che si susseguono nella sequenza. Ad esempio, l'incorporazione della sequenza AAA produrrà un'emissione di luce tre volte più intensa rispetto all'incorporazione di un singolo nucleotide A. Se il dNTP non è complementare non verrà incorporato e non verrà generata nessuna emissione luminosa; quindi l'assenza di picchi indica che nella posizione considerata non è presente quel dNTP. Dopo la registrazione dell'emissione luminosa, i rimanenti nucleotidi liberi vengono degradati da apirasi e viene fornito il dNTP successivo, proseguendo la reazione di polimerizzazione. La Figura 1.10 riassume graficamente un singolo ciclo di pirosequenziamento (Breu H., 2010).

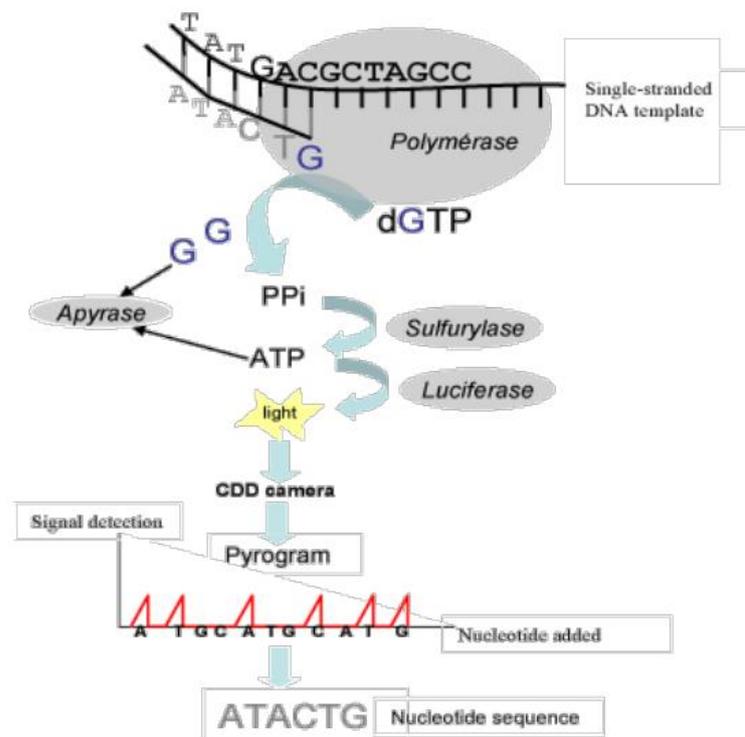


Figura 1.10 Rappresentazione del processo di pirosequenziamento. L'incorporazione del dGTP provoca il rilascio di un pirofosfato (Ppi) che viene convertito in ATP dalla sulfurilasi. L'ATP è coinvolto nella reazione di bioluminescenza ad opera della luciferasi che converte l'energia chimica in energia luminosa. L'emissione di luce è captata da un sensore CCD e elaborata per ottenere la sequenza nucleotidica. I dNTP non incorporati e la molecola di ATP vengono degradati da aspirasi

<http://omicsonline.org>

Ad ogni ciclo è aggiunta una singola base nucleotidica (A, C, G o T); al ciclo successivo sarà poi introdotta una delle basi restanti, e così via ripetendo la sequenza di introduzione per centinaia di cicli. Il tasso di errore medio è compreso tra 10^{-3} e 10^{-4} , superiore al tasso di errore del metodo Sanger. Genome Sequencer permette di sequenziare 1.5 milioni di sfere per run, determinando sequenze di lunghezza variabile in base al numero di cicli di sequenziamento (limitati per mantenimento dell'efficienza di polimerasi e luciferasi).

Il vantaggio del metodo 454 Roche è che le reads lunghe massimo 1000 bp sono più facili da mappare avendo un genoma di riferimento: questo è un aspetto positivo sia per l'assemblamento di un nuovo genoma sia per applicazioni metagenomiche. Il tempo delle run è relativamente veloce, circa 23h. Di contro si ha un throughput relativamente basso di circa 1 milione di reads e 100 mb di dati, unito ad un elevato costo dei reagenti (Van Dijk E. L. *et al.*, 2014).

1.2.2.2 Il sequenziatore Helicos tSMS

La piattaforma Helicos di BioSciences è uno strumento in grado di fornire un totale di 1 Gb/giorno di output. In questo sequenziatore è stato eliminato il processo di amplificazione clonale, il metodo si basa sulla frammentazione del DNA stampo e sulla poliadenilazione all'estremità 3', in cui l'adenosina terminale risulta essere marcata con un fluoroforo. I filamenti di DNA poliadenilati composti da circa 200, 250 bp (Metzker M. L., 2010) vengono uniti ad adattatori universali alle estremità e ibridati a primer oligonucleotidici legati covalentemente al supporto solido (Harris T. D. *et al.*, 2008). In questo modo i filamenti stampo vengono immobilizzati sulla superficie della flow cell (Figura 1.11), con una densità di cattura fino a un massimo di 100×10^6 filamenti di DNA stampo per cm^2 .

Per il sequenziamento, ad ogni ciclo vengono aggiunti in una flow cell una DNA-polimerasi e uno dei 4 dNTP marcati con il fluoroforo Cy5 (Figura 1.11), la cui fluorescenza viene acquisita per determinare l'incorporazione della base nei singoli filamenti di DNA. Le coordinate delle posizioni dei singoli filamenti di DNA catturati sono quindi registrate da una CCD camera e il fluoroforo viene rimosso ed eliminato tramite un lavaggio prima di effettuare la reazione di sequenziamento.

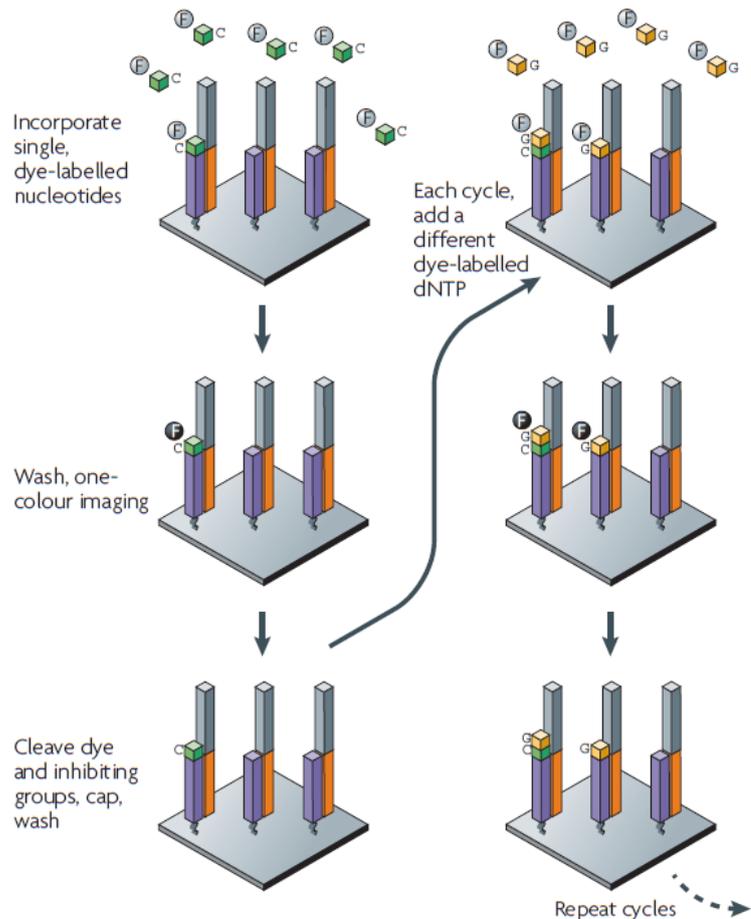
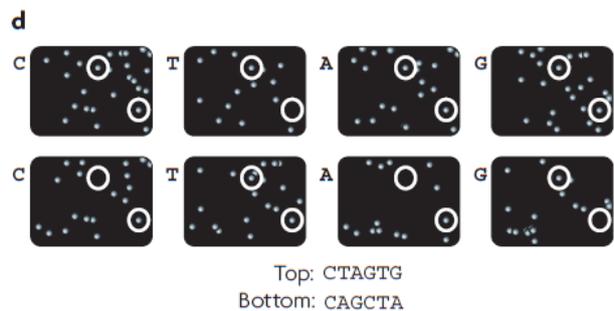


Figura 1.11 Sequenziamento by Helicos: metodo che utilizza un colore per ogni ciclo di sequenziamento con terminazione reversibile (Metzker M. L., 2010)



Dopo rimozione del fluoroforo e lavaggio, tale processo viene ripetuto con il dNTP successivo marcato con Cy5. Ogni ciclo di sequenziamento, che consiste in aggiunte successive di DNA polimerasi e di ciascuno dei 4 dNTP marcati, è chiamato “quad”. In genere, viene effettuato un numero di circa 25-30 “quad”, che permette di ottenere “reads” di 45-50 bp di lunghezza. Helicos ha brevettato dei dNTP marcati definiti terminatori virtuali, che sembra riducano la processività della DNA polimerasi in modo che vengano aggiunte solo singole basi, migliorando l'accuratezza del sequenziamento degli omopolimeri. È interessante notare che la percentuale dei filamenti di DNA con lunghezza intorno ai 50 nucleotidi che può essere sequenziata è notevolmente inferiore rispetto a quella ottenuta sequenziando frammenti di lunghezza minore, intorno ai 25 nucleotidi, probabilmente a causa delle strutture secondarie (es. ad ansa) assunte dalle molecole stampo.

1.2.2.3 Solexa Illumina

Circa a metà degli anni '90, gli scienziati Shankar Balasubramanian e David Klenerman dell'università di Cambridge, stavano utilizzando nucleotidi marcati a fluorescenza per osservare il moto di una polimerasi a livello di una singola molecola in quanto sintetizzava il DNA immobilizzato su una superficie. Tuttavia, la tecnologia SBS (*Sequencing-by-synthesis*) ha generato notevolmente più dati di sequenza, fornendo oltre 3 milioni di basi da una singola esecuzione. Il primo analizzatore genomico Solexa, lanciato nel 2006, aveva una potenza di 1 gigabase (Gb) di dati in una singola run. Solexa è stata acquisita da Illumina all'inizio del 2007. Con questa tecnologia negli anni sono stati sequenziati numerosi genomi di microrganismi, piante, umani e animali. La produzione di dati di sequenziamento di nuova generazione (NGS) è aumentata ad una velocità che supera la legge di Moore. La legge di Moore, coniata dal cofondatore di Intel Gordon Moore nel 1965, descrive la crescita esponenziale nel tempo della complessità nei circuiti integrati, affermando che il numero di transistor in un circuito integrato raddoppia approssimativamente ogni 18 mesi. I costi di sequenziamento e di conseguenza la quantità di dati prodotti hanno un andamento molto più veloce, dimezzando approssimativamente ogni 5 mesi. I dati biologici che vengono generati dal sequenziamento sono quindi diventati più costosi da memorizzare che da generare come è descritto nel grafico 1.1 (Stein L. D., 2010).

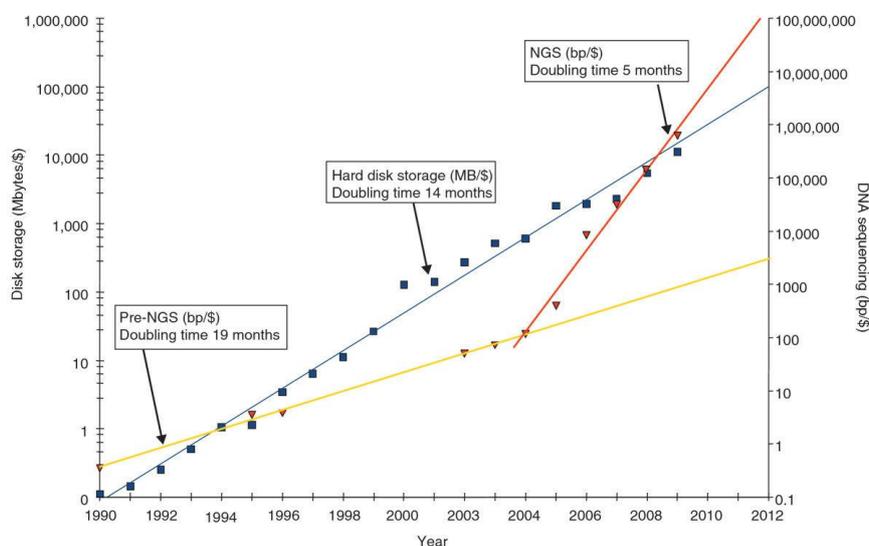


Grafico 1.1 Rappresentazione grafica dei costi di sequenziamento e di storage dal 1990 al 2012
(Stein L. D., 2010)

Tecnologia

La preparazione del campione da analizzare consiste nell'aggiungere specifiche sequenze come adattatori alle estremità dei frammenti di DNA. La produzione delle library inizia con la frammentazione del DNA, in seguito un pull enzimatico ripara le estremità dei frammenti e le adenine vengono aggiunte sulla estremità 3'. In seguito vengono aggiunti gli adattatori la cui struttura è unica

per ogni estremità e permette ai frammenti di DNA di essere immobilizzati sul supporto rigido. L'immobilizzazione avviene grazie all'ibridazione tra gli adattatori e i primer sulla superficie che sono complementari alle library. In questo modo i frammenti di DNA formano strutture a ponte che possono iniziare l'amplificazione tramite appunto la bridge PCR (Myllykangas S. *et al.*, 2012) (Figura 1.12).

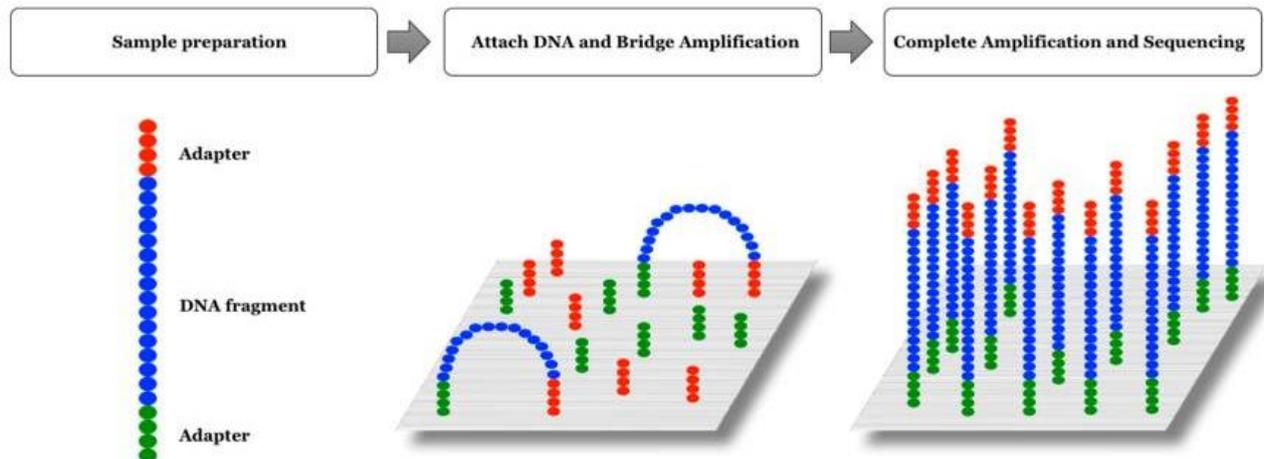


Figura 1.12 I frammenti sono attaccati ai primer immobilizzati su una superficie solida, eseguendo un'amplificazione a ponte e generando cluster di DNA con molecole identiche (Garrido-Cardenas, J. A., *et al.*, 2017).

In questa piattaforma, il sequenziamento del DNA è condotto da nucleotidi marcati a fluorescenza che agiscono come terminatori reversibili della reazione di amplificazione (Morozova O. *et al.*, 2009). In ogni ciclo, i quattro nucleotidi di terminazione reversibile vengono contemporaneamente aggiunti e incorporati dalla polimerasi. Questi nucleotidi sono bloccati chimicamente sostituendo il gruppo 3'-OH con un gruppo 3'-*o*-azidometile per impedire che la polimerasi incorpori più nucleotidi in ciascun ciclo. Con l'incorporazione di un nucleotide viene emesso un segnale di fluorescenza che viene misurato con la fluorescenza totale di riflessione interna (TIRF) usando vari canali laser. Per quanto riguarda il ciclo successivo, i nucleotidi che non sono stati incorporati vengono lavati e il terminatore chimico dell'estremità 3' viene rimosso mediante l'uso di tris-(2-carbossietil)-fosfina per continuare la sintesi della catena (Bentley D. R. *et al.*, 2008). Grazie all'elevato numero di frammenti clonali che si formano durante l'amplificazione, il segnale luminoso è più intenso e quindi più facilmente percepibile. Infatti quando vengono inseriti i nucleotidi terminali marcati, il segnale emesso sarà visibile dal CCD camera. Una volta che il segnale di fluorescenza viene raccolto, inizia un nuovo ciclo, ripetendo questa dinamica fino a quando il sequenziamento di ciascun frammento è terminato.

Come indicato in precedenza, il sistema di rilevazione utilizzato nella piattaforma Illumina è la Total Internal Reflection Fluorescence (TIRF), nota anche come microscopia a onde evanescenti. Il suo

principale vantaggio rispetto ad altri sistemi è che è in grado di rilevare la fluorescenza delle molecole molto vicine a una superficie solida (vetro o plastica) ed è altamente selettiva (Axelrod D., 2001). Con TIRF è possibile illuminare uno strato molto sottile di profondità inferiore a 100 nm, evitando l'eccitazione di altri fluorofori che possono essere vicini ma la cui emissione non è richiesta per la misura (Gould T. J. *et al.*, 2011).

Questo stesso sistema di rilevamento viene utilizzato dalla piattaforma Qiagen GeneReader, lanciata sul mercato nel 2015. Allo stato attuale, i sequenziatori di Illumina sono i più utilizzati nello sviluppo di progetti di sequenziamento massivi. Questo non è solo dovuto alla loro alta precisione nella sequenza e al basso costo per Gb (Gigabyte) ottenuti, ma anche perché dispongono di una grande varietà di apparecchiature sul mercato che possono adattarsi alle esigenze di ogni progetto (Garrido-Cardenas J. A. *et al.*, 2017). Inoltre la lunghezza delle reads è di circa 300 bp, compatibile quindi con la maggior parte delle applicazioni. Per contro la preparazione del campione è particolarmente impegnativa a causa della dispersione dei cluster all'interno delle celle di flusso la concentrazione della library deve essere altamente controllata. Infatti una sovrapposizione di cluster causa risultati sovrapposti con scarsa qualità del risultato di sequenziamento.

HiScanSQ

HiScanSQ (Illumina, San Diego CA) utilizza la chimica SBS per generare delle reads di 200 bp. Data la capacità di sequenziare in parallelo decine di milioni di sequenze, la tecnologia HiScanSQ si rivela particolarmente adatta per il sequenziamento di interi genomi o trascrittomi. Con questo approccio vengono utilizzati terminatori reversibili per rilevare le singole basi nucleotidiche man mano che vengono incorporate nei filamenti di DNA da sequenziare. Durante ciascun ciclo di sequenziamento, alla catena dell'acido nucleico viene aggiunto un solo deossinucleotide trifosfato (dNTP) marcato. Il nucleotide marcato funge da terminatore reversibile per la sintesi del nuovo filamento di DNA nascente, così dopo ogni incorporazione, la fluorescenza viene rilevata da un sistema ottico che permette di identificare la base incorporata. Infine il nucleotide è sottoposto a scissione enzimatica per consentire l'incorporazione del nucleotide successivo. La piattaforma HiScanSQ produce un massimo di 1.5 miliardi di reads totali per ogni run, producendo fino a 150 Gb di output con una media di 60-70 milioni di reads per campione se caricato in pool da 3 campioni per lane (Melle F., 2016).

1.2.3 Tecnologie tra la seconda e terza generazione - Sequencing by ligation

1.2.3.1 Il sequenziatore SOLiD di Applied Biosystems

La tecnologia SOLiD è stata sviluppata sul metodo Sequencing-By-Ligation proposto da George Church nel 2005 (Shendure J. *et al.*, 2005). La piattaforma SOLiD viene introdotta nel mercato a partire dal 2007 da Applied Biosystems. Dopo la denaturazione della doppia elica e l'aggiunta di adattatori universali P1 e P2, i frammenti sono ibridati a primer oligonucleotidici complementari, presenti su delle sfere, come nel 454 Roche. Nella tecnologia SOLiD le sfere sono però più piccole (diametro 0.75 μm invece 28 μm), permettendo una maggiore densità di sfere sequenziabili per run. Le sfere sono messe in emulsione assieme ai reagenti e ai primers della PCR per ottenere l'amplificazione clonale dei filamenti (em-PCR), in modo del tutto simile alla piattaforma 454 Roche. A differenza di quest'ultima tecnologia, la piastra di sequenziamento SOLiD non dispone di pozzetti; le sfere sono invece legate alla superficie opportunamente trattata della piastra di sequenziamento tramite la formazione di un legame covalente con le estremità dei filamenti amplificati, modificate chimicamente. Mentre nel 454 Roche esiste un numero predefinito di pozzetti e quindi di sfere sequenziabili, nel sequenziatore SOLiD il limite del numero di sfere che possono essere deposte sulla superficie è dato solo dal loro diametro (inferiore a 1 μm).

La reazione di sequenziamento non avviene grazie a DNA polimerasi, ma tramite DNA ligasi, un enzima in grado di legare covalentemente due frammenti di DNA. L'acronimo SOLiD significa infatti "Sequencing by Oligonucleotide Ligation and Detection". Ogni ciclo di sequenziamento coinvolge la sfera, un primer di sequenziamento, la ligasi e quattro sonde di dNTP. Le sonde sono composte da 8 basi più un marcatore fluorescente e presentano un sito di rottura fra i nucleotidi 5 e 6. Le prime due basi della sonda sono complementari a due nucleotidi da sequenziare, le basi da 3 a 5 sono degeneri, cioè incapaci di appaiarsi a qualsiasi base della sequenza stampo, le basi 6, 7 e 8 sono anch'esse degeneri e possono essere rimosse, assieme al marcatore fluorescente, tramite la rottura del legame con la base in posizione 5 (Garrido-Cardenas J. A. *et al.*, 2017). Una volta che la sonda si è appaiata alla sequenza stampo, avviene l'eccitazione del marcatore legato alla sonda e la conseguente emissione di fluorescenza. Contemporaneamente, viene rotto il legame tra le posizioni 5 e 6, lasciando libero l'estremità 5' della quinta base della sonda e permettendo il legame con la successiva sonda. L'emissione luminosa è rilevata dalla strumentazione in modo da identificare il colore associato alla coppia delle prime due basi (Li H. *et al.*, 2010). Ogni marcatore produce una specifica emissione luminosa che identifica 4 di 16 possibili coppie di nucleotidi. Perché quindi associare un colore a una coppia di basi, invece di associarne uno a ognuna, rendendo così la corrispondenza biunivoca? Anche se la corrispondenza 1-1 sembra il metodo migliore per procedere in realtà non è così, infatti questa

può portare più facilmente ad errori di sequenziamento, che in alcuni casi, compromettono inevitabilmente i risultati dell'esperimento.

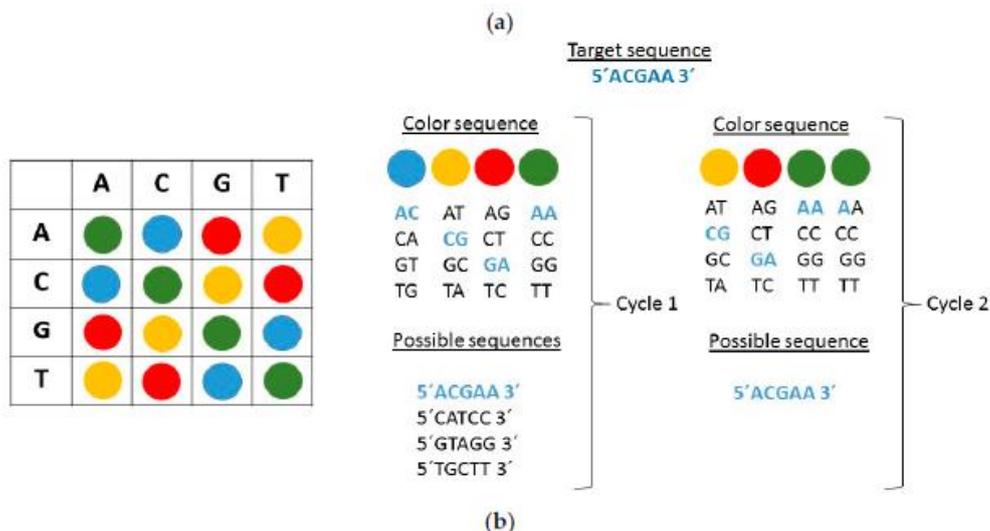


Figura 1.13 Le possibili combinazioni di nucleotidi che i quattro colori dei marcatori fluorescenti rappresentano. Ciascun colore non rappresenta univocamente una singola coppia di basi, ma 4 combinazioni sulle 16 possibili. (Garrido-Cardenas J. A. et al., 2017)

Definita la n posizione della base del primer (appaiato all'adattatore che lega la sequenza stampo alla sfera) le basi n+1 e n+2 saranno quindi complementari alla sequenza stampo, così come le successive basi n+6 e n+7 e così via. Per ciascuna di queste coppie di basi, lo strumento rileva un segnale che può corrispondere a una delle sedici combinazioni possibili. Le basi n+3, n+4 e n+5 sono anch'esse appaiate (in quanto degeneri) ma indeterminate. La loro identificazione sarà resa possibile in un successivo ciclo di sequenziamento, spostando il primer in una posizione definita della sequenza stampo. Dopo l'appaiamento della prima sonda, il processo si ripete con il legame di nuove sonde, l'identificazione e la separazione del marcatore e degli ultimi tre nucleotidi; il numero delle iterazioni (solitamente 7) determina la lunghezza di lettura. Terminato il processo, il filamento ottenuto è rimosso e un nuovo primer viene appaiato al filamento stampo in posizione n-1. Nel nuovo ciclo saranno univocamente appaiate le basi n+2 e n+3, nel ciclo successivo le basi n+3 e n+4 e così via (Garrido-Cardenas J. A. et al., 2017). Nel complesso sono eseguiti cinque cicli di sequenziamento, ciascuno con il primer spostato in una posizione di volta in volta arretrata. I cinque cicli permettono di interrogare ogni base della sequenza per due volte (in due indipendenti reazioni con diverse posizioni dei primer). In Figura 1.14 si vede che la base della posizione di lettura 5 viene coinvolta dal primer numero 2 nel secondo ciclo (indicato in azzurro) e dal primer numero 3 nel primo ciclo (indicato in blu). Questa doppia interrogazione permette di migliorare la bontà del sequenziamento.

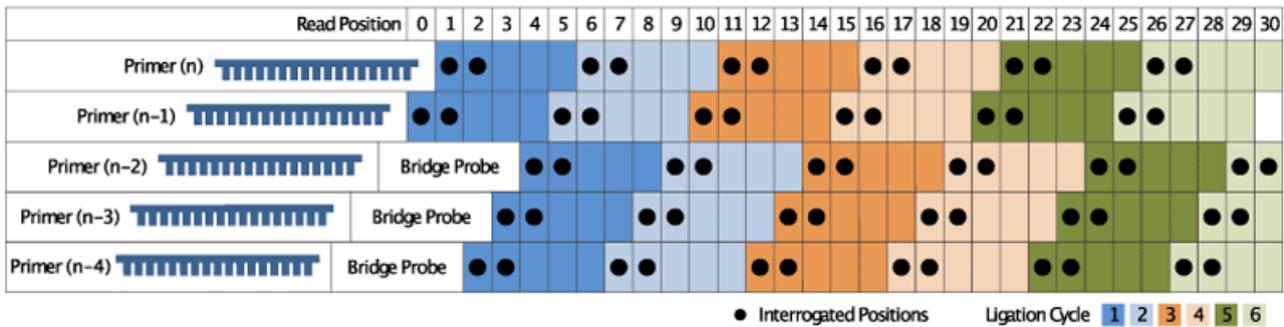


Figura 1.14 Schema del sequenziamento con tecnologia SOLiD. Sono rappresentati i cinque cicli (uno per riga) di sequenziamento, ciascuno dei quali composto da sei passaggi di aggiunta di nuove sonde ad opera della ligasi. I punti neri indicano le coppie di basi interrogate in ogni passaggio (<http://seqanswers.com/forums/showthread.php?t=10>)

Il colore dell'emissione da parte del fluoroforo permette di identificare la coppia delle prime basi, ma non è sufficiente per distinguere un nucleotide dall'altro. Oltre alla specifica emissione per ogni coppia è infatti necessario che una delle basi della sequenza sia nota. La base nota è incorporata nell'ultimo (il quinto) ciclo di sequenziamento e corrisponde all'ultimo nucleotide del primer. Quindi, dato che ciascun colore rappresenta 4 possibili coppie di nNTP e per ognuna di esse il secondo nucleotide coincide con la prima base della coppia successiva, conoscere una base della sequenza permette di interpretare tutta la sequenza nella sua interezza. In Figura 1.15 si propone un esempio di decodifica dei colori rilevati, nota la prima base della sequenza.

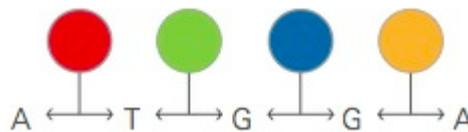


Figura 1.15 Esempio di decodifica dei colori per determinare la sequenza. Si suppongono note la prima base (A) e l'ordine dei colori rilevati (<http://seqanswers.com/forums/showthread.php?t=10>)

Le fonti possibili di errore del sequenziatore SOLiD sono diverse: errore di background nella fase di amplificazione, letture di bassa qualità a causa di presenze di sequenze diverse su una stessa sfera, oltre ad artefatti introdotti nel processo ciclico di ligasi e rimozione del marcatore. La tecnologia SOLiD sequenzia oltre 300 milioni di sfere in parallelo.

Il SOLiD è il sistema più all'avanguardia dopo Illumina in base alle prestazioni in termini di throughput (20 Gb/giorno) e ha tassi di errore molto più bassi rispetto agli altri sistemi con un'accuratezza del 99.94% (Liu L. *et al.*, 2012) dal momento che ogni base viene letta due volte. Per contro i costi di sequenziamento sono superiori a quelle degli altri sequenziatori NGS (Garrido-Cardenas J. A. *et al.*, 2017) inoltre le reads sono più corte rispetto alle altre piattaforme con 75 bp (Van Dijk E. L. *et al.*, 2014) e tempistiche elevate per ogni run. Questo sistema è meno adatto per

l'assemblaggio di genomi *ex novo*. Il sistema SOLiD è molto meno utilizzato rispetto al sistema Illumina e il sistema di preparazione è molto meno sviluppato.

1.2.3.2 Ion Torrent Life Technologies

Per effettuare un sequenziamento con Ion Torrent, il DNA stampo viene amplificato sempre con PCR in emulsione (Nakano M. *et al.*, 2003). Successivamente vengono inseriti i singoli nucleotidi, la tecnologia utilizzata per il sequenziamento del DNA elimina sia la scansione ottica che i dNTP collegati ai fluorofori (Singh P. S. 2015). L'aggiunta di un singolo nucleotide si verifica uno alla volta, quindi non è necessario bloccare i dNTP, come nel caso del sequenziamento per terminazione ciclica reversibile (CTR). Il processo che viene eseguito è lo stesso usato per la costruzione di circuiti integrati in chip informatici. Infatti si sfrutta una tecnologia di semiconduttori (metallo-ossido-semiconduttore, CMOS) (Li H. *et al.*, 2016), in grado di monitorare le variazioni del pH ambientale dovuto al formarsi di ioni quando un singolo nucleotide è incorporato dall'enzima DNA polimerasi nella catena nascente di DNA a doppia elica. Quando il nucleotide aggiunto è incorporato dalla polimerasi nella sintesi del filamento del DNA, viene rilasciato uno ione idrogeno e questo viene rilevato dal sensore CMOS-ISFET (Merriman B. *et al.*, 2012; Rothberg JM. *et al.*, 2011), generando un segnale che viene inviato ad un computer per l'elaborazione.

Il grande successo di questa piattaforma consiste nell'integrazione di un chip che dispone nella sua matrice di milioni di sensori CMOS, in modo che la compilazione di tutti i dati possa essere eseguita in modo economico e semplice (Buermans H. P. J. *et al.*, 2014). La seconda importante innovazione di questa piattaforma è stata l'introduzione di un sensore elettrochimico a livello di ISFET (transistor a sensore di campo ionico) in fondo a ciascun pozzetto (Toumazou C. *et al.*, 2011), che agisce come un misuratore di pH sensibile ai cambiamenti nella concentrazione di H^+ .

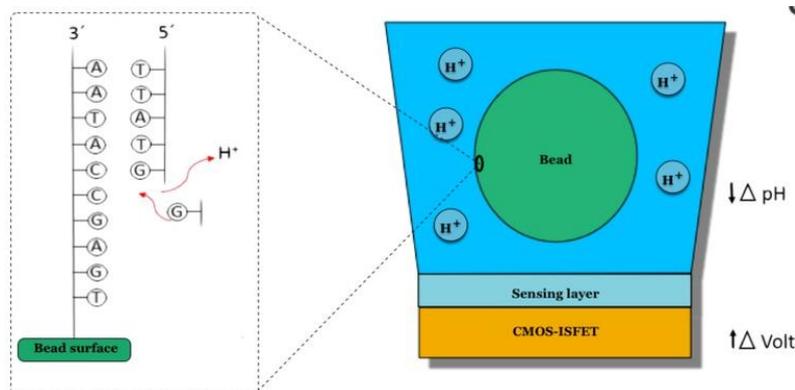


Figura 1.16 Nella piattaforma Ion Torrent, il chip è il sequencer. Ogni pozzetto del chip agisce come un misuratore di pH che è in grado di rilevare i cambiamenti nella concentrazione di H^+ prodotta nella polimerizzazione del DNA (Garrido-Cardenas J. A. *et al.*, 2017)

Quindi come vantaggio questo sistema ha che, a differenza di altre piattaforme, i nucleotidi utilizzati non sono modificati chimicamente con l'utilizzo di fluorofori, quindi non è richiesto uno scanning ottico. I tempi di esecuzioni sono molto veloci infatti un run richiede solo poche ore ed è molto versatile. Per contro questa tecnologia ha, come il 454, un alto tasso di errore (Van Dijk E. L. *et al.*, 2014).

1.2.4 Tecnologie di Terza Generazione

Le piattaforme Pacific Biosciences ed Oxford Nanopore appartengono alle tecniche di sequenziamento di terza generazione e rappresentano un vero e proprio salto di qualità e di potenzialità pur essendo attualmente poco più che dei prototipi. Sicuramente in un prossimo futuro ci si potrà avvalere di strumenti “low cost” di piccole dimensioni, capaci di entrare con più facilità nella pratica medica di routine. Riuscire a leggere una singola molecola di DNA monitorando in tempo reale l'aggiunta dei nucleotidi, significa poter usare piccole quantità di DNA ed avere anche un vantaggio economico, risparmiando sui reagenti. La piattaforma Oxford Nanopore è in grado di leggere una sequenza di DNA basandosi sui cambiamenti nel flusso di corrente che attraversa dei pori di pochi nanometri, che sono caratteristici dello specifico nucleotide che viene aggiunto alla catena (Chessa R., 2014).

1.2.4.1 Sequencing by synthesis: SMRT (Single Molecule Real Time sequencing) Pacific Biosciences

La piattaforma PacBio, commercializzata nel 2010 da Pacific Biosciences, utilizza la tecnologia di sequenziamento in tempo reale di una singola molecola (SMRT). Il sequenziamento si basa sull'utilizzo di nucleotidi marcati con singole molecole fluorescenti, ma il sequenziamento procede senza la necessità di una precedente amplificazione in PCR.

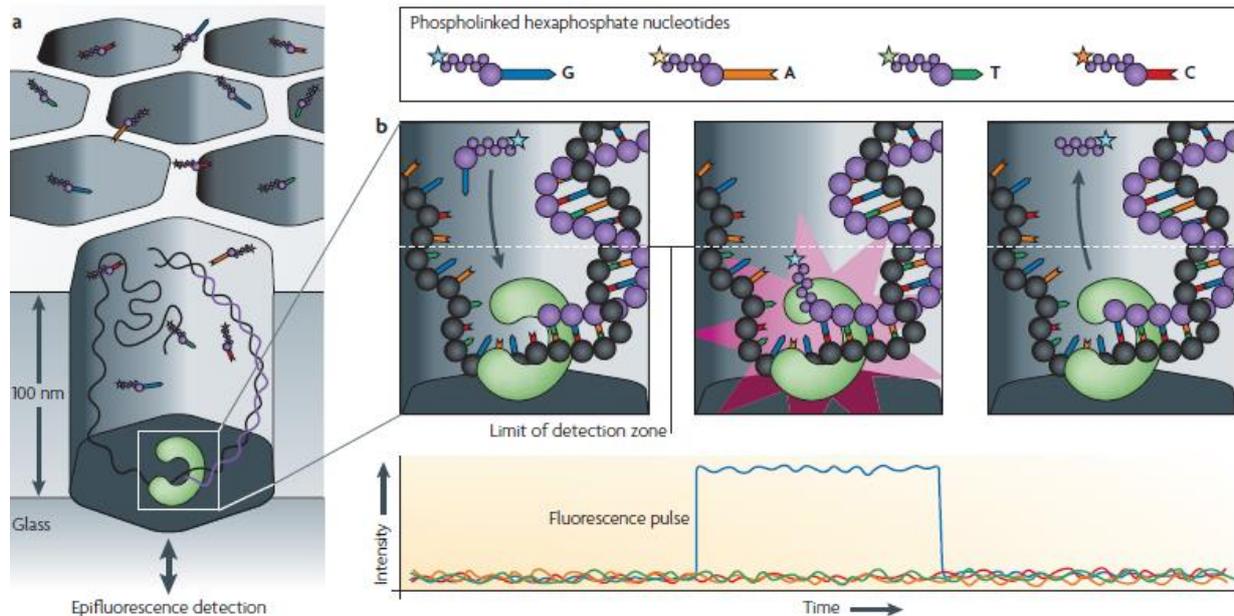


Figura 1.17 Sequenziamento PacBio all'interno di pozzetti ZMW, con dNTP marcate con fluoroforo (Metzker M. L., 2010)

Le molecole di DNA a singolo filamento vengono inserite nel fondo di pozzetti ZMW (zero-mode wave guide detectors) (Levene M. J. *et al.*, 2003), dove è presente una polimerasi immobilizzata (Eid J. *et al.*, 2009) (Figura 1.17). Il sequenziamento utilizzato da questa tecnologia è il real time sequencing (Metzker M. L., 2010): le basi azotate marcate con fluoroforo vengono inserite nei pozzetti e, nel momento in cui vengono incorporate all'interno del filamento di DNA, la molecola fluorescente si stacca dal nucleotide, emettendo un segnale luminoso (Metzker M. L., 2010). Ogni segnale luminoso di colore differente per ogni base, divergerà dagli altri segnali per millisecondi: questa frazione di tempo permette di ricostruire la sequenza nucleotidica. Teoricamente, questa tecnologia riesce a sequenziare grandi molecole di DNA (fino a 30 kb) ed è quindi lo strumento ideale per completare l'assemblaggio di sequenze genomiche o per migliorare quelle già conosciute. Un altro vantaggio è che i tempi di esecuzione sono molto veloci, ovvero di poche ore. Per contro ha alti costi (\$ 2-17 per Mb) e elevati tassi di errore (14%). Inoltre ha la più bassa performance di throughput tra tutte le piattaforme (massimo 500 Mb) e un range limitato di applicazioni (Van Dijk E. L. *et al.*, 2014).

1.2.4.2 Sequencing by ligation: Oxford Nanopore Technologies MinION

L'Oxford Nanopore Technologies MinION è una nuova tecnologia di sequenziamento che offre la possibilità di sequenziare delle reads molto lunghe, di circa 10.000 bp (10 kb) (Urban JM. *et al.*, 2015), sempre in real time. Questa tecnologia è basata su pori (nanopores) che formano una sorta di tunnel tra due compartimenti. La struttura della flow cell è composta da un circuito integrale e

ciascuno dei 512 canali individuali che sono in grado di sequenziare più di 60 bp al secondo. Dopo la frammentazione del DNA campione a singolo filamento, dei nanopori formano strutture a canale (nanopore), che dirigono il campione verso un sensore che permette l'individuazione di ogni nucleotide presente nel filamento di DNA (Branton D. *et al.*, 2008). Il DNA scorre nel poro grazie ad una proteina motrice e in questo modo produce un'alterazione di potenziale tra le due parti del poro, che sarà caratteristica per ogni sequenza di DNA (Clarke J. *et al.*, 2008). La molecola che attraversa il poro causa una temporanea differenza di potenziale tra i due compartimenti, che identifica una precisa sequenza (Eisenstein M., 2012) (Figura 1.18), tramite un segnale come colore, luce o pH. In questo modo la piattaforma nanopore è in grado di riconoscere la struttura del DNA (Mikheyev A. S. *et al.*, 2014). La variazione nel voltaggio misurato dipenderà non solo dal tipo di sequenza ma anche dalla durata in modo che la misura possa essere interpretata in maniera precisa e univoca. Il dispositivo ha un costo relativamente basso, è lo strumento per il sequenziamento del DNA più piccolo mai costruito, è possibile poterlo trasportare dove si intende effettuare lo studio (Goodwin S. *et al.*, 2015). Inoltre, questo tipo di strumento non richiede l'utilizzo da parte di personale addestrato.

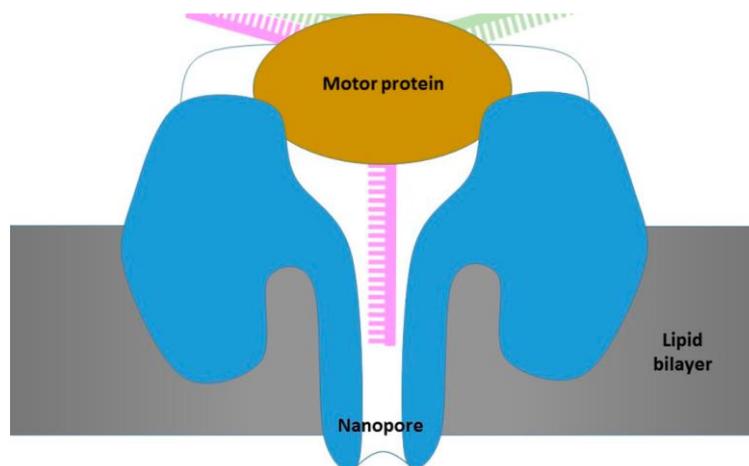


Figura 1.18 Canali utilizzati dalla piattaforma Nanopore Oxford per la sequenza del DNA. Il passaggio del DNA attraverso il nanopore produce alterazioni che vengono misurate grazie alle variazioni di tensione rilevate (Garrido-Cardenas J. A. *et al.*, 2017)

Dal momento che il MinION misura le singole molecole ha una maggiore frequenza di errore (tra il 10-20%) (Quick J. *et al.*, 2014; Jain M. *et al.*, 2015), paragonato agli strumenti high-throughput che leggono coppie clonali di molecole di DNA. Per generare delle sequenze accurate, le regioni genomiche devono essere lette più volte, in modo da eliminare l'errore tramite la media delle misurazioni (Quick J. *et al.*, 2014; Quick, J. *et al.*, 2016).

1.3 Differenze tra le tecnologie di sequenziamento

Dalle caratteristiche delle varie piattaforme descritte nei paragrafi precedenti, emergono differenze significative che vengono riassunte nelle seguenti tabelle:

	Sanger	Genome Analyzer	GS (Genome Sequencer)	SOLiD
Azienda		Solexa/Illumina	454 Life Sciences Roche	Applied Biosistem
Library		Adattatori Lineari	Adattatori Lineari	Adattatori Lineari
Supporto		Flow Cell	Picotiter Plate	Flow Cell
Preparazione del campione	Frammentazione PCR	Frammentazione bridge-PCR	Frammentazione em-PCR	Frammentazione em-PCR
Principio di sequenziamento	Nucleotidi Terminatori fluorescenti	SBS Nucleotidi Terminatori "reversibili" fluorescenti	SBS Pirosequenziamento	SBL
Tempo di run	10 giorni	2.5 giorni	10 ore	7 giorni
Lunghezza delle reads (bp)	400 - 900	75 - 200	300 - 800	35 - 100
Gb per run	0.1	200 - 1000	0.45	30 - 100
Costo dello strumento (\$)	95000	690000 (Hi-Seq)	500000	495000

*Tabella 1.1a Differenze tra piattaforme di sequenziamento
 (Metzker M. L., 2010; Garrido-Cardenas J. A. et al., 2017; Liu L. et al., 2012;
 Myllykangas S. et al., 2012; Voelkerding K. V. et al., 2010)*

	Helicos	Ion Torrent	SMRT	Nanopore
Azienda	Pacific Biosciences	DNA Electronics Ltd	Pacific Biosciences	Pacific Biosciences
Library	Adattatori Lineari	X	Adattatori Circolari	X
Supporto	Flow Cell	Semiconduttori	Pozzetti ZMW	Nanopore
Preparazione del campione	Frammentazione No PCR	Frammentazione em-PCR	Frammentazione No PCR	Frammentazione No PCR
Principio di sequenziamento	SBS Nucleotidi Terminatori "reversibili" fluorescenti	SBL Terminatore Ciclico reversibile (CTR) Proton Detection	SBS Real Time	SBL Real Time
Tempo di run	8 giorni	3.5 ore	1 ora	2 ore
Lunghezza delle reads (bp)	200 - 250	400 - 500	900 - 1000	10000
Gb per run	10 - 40	10 - 100	10	5
Costo dello strumento (\$)	999000	100000	X	X

*Tabella 1.1b Differenze tra piattaforme di sequenziamento
 (Metzker M. L., 2010; Garrido-Cardenas J. A. et al., 2017; Liu L. et al., 2012;
 Myllykangas S. et al., 2012; Voelkerding K. V. et al., 2010)*

1.4 Sequenziamento applicato allo studio degli alimenti

Gli alimenti al loro interno contengono una serie di microrganismi utili che vi conferiscono le caratteristiche organolettiche e tecnologiche desiderate. Vi sono però anche microrganismi indesiderati che ne possono ridurre la qualità e originare problemi di safety. Il metodo tradizionale per determinare la composizione del microbiota alimentare è attraverso la coltura in piastra, tuttavia questa tecnica non è realizzabile per tutti i microrganismi (Amann R. I. *et al.*, 1995). L'isolamento selettivo dei microrganismi richiede fattori di crescita talvolta ignoti e condizioni non riproducibili in laboratorio. Si deve quindi essere consapevoli che i microrganismi sono molti di più rispetto a quelli ottenibili in piastra (Ercolini D. *et al.*, 2001). Nelle ultime decadi molti metodi molecolari, indipendenti dalla coltivazione in piastra, sono stati sviluppati per risolvere questa problematica ed estesi al sistema alimentare. Queste tecniche consentono l'identificazione e in alcuni casi la quantificazione di gruppi microbici alimentari e offrono metodi rapidi e sensibili per determinare la composizione e la diversità delle complesse comunità microbiche (Mayo B. *et al.*, 2014). Queste tecniche si basano sul sequenziamento dell'acido nucleico microbico (DNA e RNA). La maggior parte delle tecniche si basa sull'amplificazione degli acidi nucleici tramite PCR (polymerase chain reaction). Confrontando le sequenze ottenute e conservate in banche dati, si possono costruire relazioni filogenetiche tra i microrganismi. Tra le tecniche che si basano sulla PCR, la qPCR (real-time PCR quantitativa) è stata ampiamente utilizzata per caratterizzare microbiologicamente le fermentazioni degli alimenti (Jany J. L. *et al.*, 2008).

2 METAGENOMICA

Introduzione

Le cellule procariote presenti sulla terra sono nell'ordine di $5 \cdot 10^{30}$ ed utilizzano circa 350-500 petagrammi ($1 \text{ Pg} = 10^{15} \text{ g}$) di carbonio, 85-130 Pg di azoto e 9-14 Pg di fosforo. Questi dati spiegano perché la comunità microbica rappresenti il più grande serbatoio biologico della terra (Whitman W. B. *et al.*, 1998). Sia i batteri che gli archaea riescono a vivere in tutti gli ambienti in grado di sostenere la vita, essi sono in grado di vivere anche in condizioni estreme: dalle profondità marine con temperature di 340°C ad ambienti a 6 km sotto la superficie terrestre. I microorganismi non solo sono ubiquitari, ma sono essenziali per la vita sulla terra, poiché sono la fonte primaria dei nutrienti e riciclano le fonti di C esaurite producendo materia organica disponibile. Come accade per gli animali e le piante, anche la vita dell'uomo è profondamente dipendente dai microorganismi. All'interno del nostro corpo sono presenti un numero molto elevato di batteri, basti pensare che le cellule umane sono nell'ordine di 10^{13} e le cellule batteriche che abitano all'interno dell'essere umano sono circa 10^{14} (Berg R. D., 1996). Anche se il genoma è il primo determinante delle caratteristiche dell'uomo (Collins F. S. *et al.*, 2001) è ormai chiaro che tali caratteristiche sono anche influenzate dalla interazione dell'uomo con il microbiota (O'Hara A. M. *et al.*, 2006) ed è quindi importante studiare anche il genoma microbico che l'uomo ospita (Wooley J. C. *et al.*, 2010).

Lo studio dei genomi microbici è iniziato alla fine degli anni '70 con il sequenziamento dei genomi dei batteriofagi MS2 (Fiers W. *et al.*, 1976) e ϕ -X174 (Air G. *et al.*, 1977). I genomi di 916 specie batteriche, 1987 virali e 67 arcaiche sono depositate all'interno della GenBank release 2.2.6. Avere a disposizione un numero così elevato di genomi microbici ha modificato il settore della microbiologia e degli studi di evoluzione microbica. Ora si è in grado di esaminare il rapporto tra struttura e funzione del genoma in molte specie diverse. Tuttavia, gli studi del genoma del singolo organismo hanno dei limiti. In primo luogo, le limitazioni tecnologiche impongono che un microorganismo debba essere prima coltivato e purificato poi sequenziato. Tuttavia, solo una piccola percentuale dei microorganismi presenti in natura può essere coltivata in laboratorio (Amann R. I. *et al.*, 1995; Rappé M. S. *et al.*, 2003). Inoltre, capita molto raramente che i microorganismi vivano isolati perché fanno parte di una comunità inserita in uno specifico habitat e le loro caratteristiche di ogni ceppo in una popolazione sono diverse da quelle di un ceppo singolo isolato in laboratorio (Wooley J. C. *et al.*, 2010).

2.1 Genomica e metagenomica

Genomica e metagenomica hanno finalità molto diverse: la *genomica* rappresenta il sequenziamento dell'intero genoma di specifici ceppi batterici e trova applicazione negli studi epidemiologici per la tracciabilità dei focolai tossinfettivi. Questo tipo di tecnica ha soppiantato le tecniche genotipiche di tipizzazione batterica (es. PFGE, MLST, AFLP, REP-PCR) perché le informazioni ottenibili dal sequenziamento dell'intero genoma includono quelle ottenibili con le precedenti tecniche genotipiche ed informazioni aggiuntive come il profilo di antibiotico resistenza. La bioinformatica è essenziale per elaborare i dati ottenuti dal sequenziamento. Il limite della genomica, come pure della microbiologia classica, riguarda la possibilità di analizzare soltanto microrganismi coltivabili. Considerando il fatto che il 99% dei microorganismi non può essere coltivato in condizioni di laboratorio (Chitsaz H. *et al.*, 2011) risulta chiara la necessità di studiare i microrganismi direttamente all'interno del loro habitat naturale dove convivono moltissime popolazioni microbiche (Gerlach W. *et al.*, 2011). I dati di sequenza relativi al DNA totale ottenuto da un ecosistema e non soltanto dai suoi microrganismi si definisce metagenoma e lo studio dei metagenomi prende il nome di “**metagenomica**” (Rondon M. R. *et al.*, 2000).

Il termine metagenomica è comparso per la prima volta nel 1998 ed è diventato uno degli strumenti più utilizzati per studi di *ecologia microbica*. È un campo emergente in cui la potenza dell'analisi del genoma (analisi del DNA di un organismo) è applicata ad intere comunità di microorganismi, sorpassando la necessità di isolare e coltivare le singole specie batteriche, evitando così il problema del prelievo e della coltura in laboratorio (Zaltum O. S. A., 2017).

Nel suo approccio e metodo, la metagenomica trascende il genoma individuale e permette di studiare tutto il genoma di una comunità nel suo insieme (Pregolato M., 2010), quantificandone la diversità in termini di ricchezza/*abundance* di specie. Si effettua l'analisi delle sequenze di un ecosistema senza conoscere i microrganismi presenti al suo interno (Marchesi C., 2014). Per questo motivo le informazioni *metagenomiche* permettono una comprensione più approfondita del ruolo ecologico, del metabolismo, della storia evolutiva dei microorganismi in un determinato contesto e di come l'ambiente influisca sulla composizione genomica di intere specie. Combinati con le capacità dei moderni sequenziatori di ottenere molto velocemente frammenti di DNA, gli studi di metagenomica possono generare una quantità enorme di dati. Molte sequenze genomiche presentano similarità con quelle già studiate e sono individuabili mediante l'utilizzo di *algoritmi* di allineamento dotati di elevata sensibilità. Il problema di classificazione e determinazione dell'origine di una sequenza di DNA dato un insieme di genomi conosciuti è comune a molti campi della biologia computazionale molecolare (Marchiori D., 2016).

La lunghezza di ogni frammento prodotto in metagenomica può essere compresa tra 20 e 700 bp, a seconda del metodo di sequenziamento utilizzato e per questo servono algoritmi dedicati per l'analisi dei dati (Wooley J. C. *et al.*, 2010).

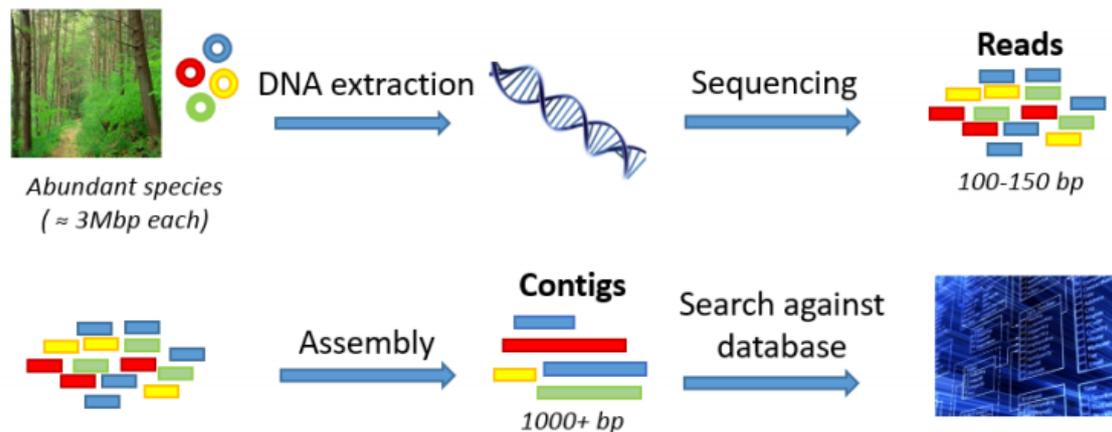


Figura 2.1 Schema riassuntivo di un piano di lavoro generico per l'analisi metagenomica
<https://www.slideshare.net/MadsAlbertsen/20131202-mads-albertsen-extracting-genomes-from-metagenomes>

La metagenomica permette di valutare l'abundance di ogni organismo presente in un sistema, come pure l'abundance dei geni presenti codificanti per enzimi e proteine attivi nei pathway metabolici presenti all'interno della matrice che si sta analizzando. In campo clinico è largamente utilizzata nello studio della microflora presente a livello orale, della microflora intestinale, nello studio di malattie genetiche e nello studio di neoplasie. Esistono due tecniche di sequenziamento HTS (High-throughput sequencing) per studiare i *consortia* microbici: target sequencing e metagenomic shotgun sequencing (De Filippis F. *et al.*, 2017).

Il metodo **targeted sequencing** prevede l'amplificazione e il successivo sequenziamento di una sequenza genetica target. Una comunità microbica viene prelevata (ad esempio da acqua, alimento, suolo, tessuto, biopsia) e il DNA viene estratto dalle cellule del campione. Un marker genomico che fornisce informazioni tassonomiche, comune a tutti gli organismi di interesse viene utilizzato come "target" e amplificato tramite PCR (Sharpton T. J., 2014). Gli ampliconi vengono successivamente sequenziati e caratterizzati a livello bioinformatico per determinare quali microrganismi sono presenti nel campione e la loro abundance. Solitamente per lo studio microbiologico, le sequenze studiate sono le porzioni variabili del 16S ribosomiale (Sharpton T. J., 2014). Questa tecnica consente di ottenere un dato qualitativo relativo alla identificazione delle specie presenti nel campione ed un dato quantitativo relativo alla loro abundance. Il sequenziamento delle regioni variabili del 16S è efficace per identificare complesse comunità microbiche (Lozupone C. A. *et al.*, 2007) ed è stato utilizzato per caratterizzare la biodiversità degli organismi presenti da un ampio range di ecosistemi compreso l'intestino umano (Yatsunenکو T. *et al.*, 2012). Confrontando i profili 16S si determina come la

diversità microbica sia associata all'eterogeneità ambientale e questo ha permesso di approfondire l'interazione tra l'ospite e meccanismi di malattie basati sul microbiota (Smith M. I. *et al.*, 2013). Il numero di reads ottenute per ogni unità tassonomica operativa nel campione (OTU) è proporzionale all'abundance di quel OTU nel campione. Tuttavia, questo tipo di sequenziamento ha alcune limitazioni: può produrre stime estremamente variabili di diversità (Consorzio Jumpstart Human Microbiome Project Data Generation, 2012). Inoltre, l'errore di sequenziamento e di amplificazione può produrre sequenze spesso difficili da identificare (Wylie K. M. *et al.*, 2012). Infine, l'amplificazione di una sequenza target fornisce solitamente solo un approfondimento sulla composizione tassonomica della comunità microbica. Il targeted sequencing è limitato all'analisi di taxa per quei marcatori di genomi conosciuti tassonomicamente a livello informatico e presenti all'interno di database (Sharpton T. J., 2014).

2.2 Shotgun metagenomic sequencing

Il metodo shotgun è un approccio alternativo che studia il microbiota non coltivabile e cerca di evitare i problemi del targeted sequencing. Mentre nel targeted sequencing si effettua l'amplificazione di specifici loci genici, nello shotgun tutto il DNA viene frammentato e sequenziato indipendentemente e non viene effettuata la PCR (De Filippis F. *et al.*, 2017). Dal momento che il genoma viene sequenziato direttamente (dopo frammentazione e preparazione delle library) si evita il bias risultante da una PCR. I risultati che si ottengono tramite il metodo shotgun comprendono anche informazioni ottenibili tramite il metodo targeted. Dopo l'estrazione, il DNA della comunità microbica di interesse viene prima quantificato, poi frammentato e infine si preparano le library per il sequenziamento. Le sequenze ottenute (reads) dallo shotgun vengono sottoposte a controlli di qualità e allineate a varie sequenze genomiche.

Con il metodo shotgun è possibile analizzare i microorganismi presenti nel campione e la loro abundance come pure la tipologia e la abundance dei geni che codificano per determinate proteine, antibiotico resistenza, etc (Figura 2.2). È possibile ricostruire le vie metaboliche presenti all'interno del campione andando a determinare i geni che codificano per un dato enzima di uno specifico pathway metabolico.

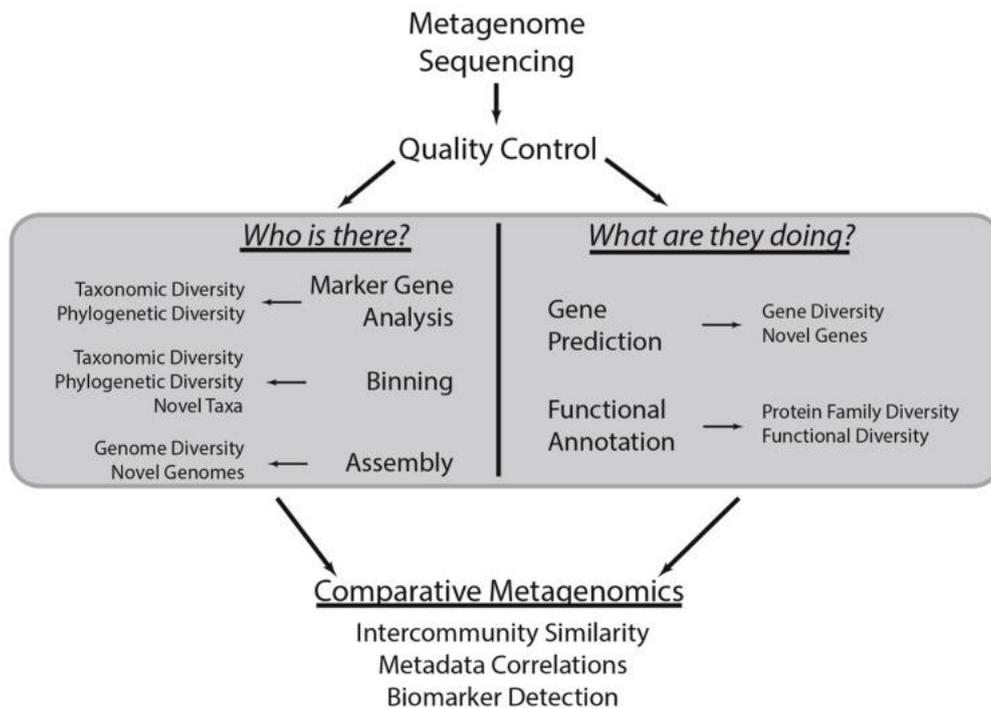


Figura 2.2 Strategie analitiche della metagenomica
(Sharpton T. J., 2014)

Il limite della metagenomica è che talvolta non è possibile conoscere il microorganismo a cui appartiene un determinato gene, ma è solo possibile affermare che all'interno del campione è presente una certa % di geni associati a determinati pathway.

Nonostante le potenzialità della metagenomica, l'analisi dei dati è molto complessa. Molte comunità microbiche sono così differenti che i loro genomi non sono rappresentati dalle reads ottenute dal sequenziamento: infatti se due reads appartenenti allo stesso gene non si sovrappongono, risulta impossibile confrontarle direttamente tramite l'allineamento di sequenza (Sharpton T. J. *et al.*, 2011). Quando le reads si sovrappongono, tuttavia, non significa che appartengano allo stesso genoma: infatti, quando si ha corrispondenza tra reads, esse potrebbero anche derivare da genomi differenti che hanno modificato il loro assemblamento durante il sequenziamento (Mende D. R. *et al.*, 2012). Inoltre, l'analisi metagenomica tende a fornire un ampio volume di dati per ottenere risultati significativi. Questo implica problemi computazionali nella elaborazione delle informazioni. In aiuto ai ricercatori lo sviluppo e il progresso dei software informatici è molto rapido e permette di migliorare l'efficienza delle analisi metagenomiche. La grande sfida che la comunità scientifica sta ora affrontando consiste nel cercare di analizzare l'enorme quantità di dati prodotta dai sequenziatori. La bioinformatica è infatti una nuova disciplina che si occupa dello sviluppo e dell'integrazione delle applicazioni della scienza dell'informazione al servizio della ricerca in campo biotecnologico. Le esplorazioni future di questa tecnica permetteranno di studiare come si modula la microflora

dell'alimento in relazione alle diverse condizioni di temperatura di stoccaggio, trattamento termico, pH, a_w , il tutto grazie ai dati ottenuti dal sequenziamento.

Altro aspetto della metagenomica da non sottovalutare è che il DNA estratto molte volte contiene all'interno DNA indesiderato. Talvolta questo DNA può superare quello di interesse e questo rende più complesso il metodo che deve essere utilizzato. Inoltre, i contaminanti della sequenza genomica sono molto problematici in quando risulta difficile determinare quali reads sono state generate da genoma contaminante e questo potrebbe portare a dedurre informazioni sbagliate. I software bioinformatici sono in grado di filtrare questo DNA da quello metagenomico di interesse (Garcia-Garcerà M. *et al.*, 2013). Queste sfide hanno limitato l'applicazione dell'investigazione metagenomica ma negli ultimi anni, il sequenziamento genomico è stato utilizzato per l'identificazione di nuovi virus (Yozwiak N. L. *et al.*, 2012), per caratterizzare la diversità genomica e la funzione di batteri incoltivabili (Wrighton K. C. *et al.*, 2012).

2.2.1 Identificazione della diversità tassonomica

Una delle prime classificazioni che viene effettuata in una comunità microbica è la quantificazione della sua diversità tassonomica, che consiste nel determinare quali microrganismi sono presenti in una comunità e la loro abundance. La diversità tassonomica serve per identificare la similarità tra due comunità. Inoltre, quando si è a conoscenza delle funzioni biologiche dei taxa, l'analisi tassonomica collega tali funzioni biologiche alla comunità microbica.

La diversità tassonomica viene quantificata tramite differenti strategie (Figura 2.3):

- Analisi di **geni marker** che forniscono informazioni tassonomiche: ogni read viene confrontata ad un database di riferimento in cui sono presenti sequenze che forniscono informazioni tassonomiche o filogenetiche (geni marker), tramite un algoritmo in grado di riconoscere se una read è uguale al gene marker e di classificare le reads in base alla loro similarità con le sequenze dei geni marker.
- **Binning**: raggruppamento delle sequenze in gruppi tassonomici definiti. Si divide in *compositional binning* che utilizza la composizione delle sequenze per classificare le reads in gruppi tassonomici; *similarity binning* che classifica una read in gruppo tassonomico o filogenetico in base alla sua similarità a geni identificati precedentemente; *fragment recruitment* in cui le reads sono allineate a sequenze genomiche molto simili per produrre stime sulla provenienza del genoma.
- **Assemblaggio**: le reads appartenenti a sequenze molto simili vengono unite alle estremità, e a loro volta sono assemblate in supercontigs o genomi completi.

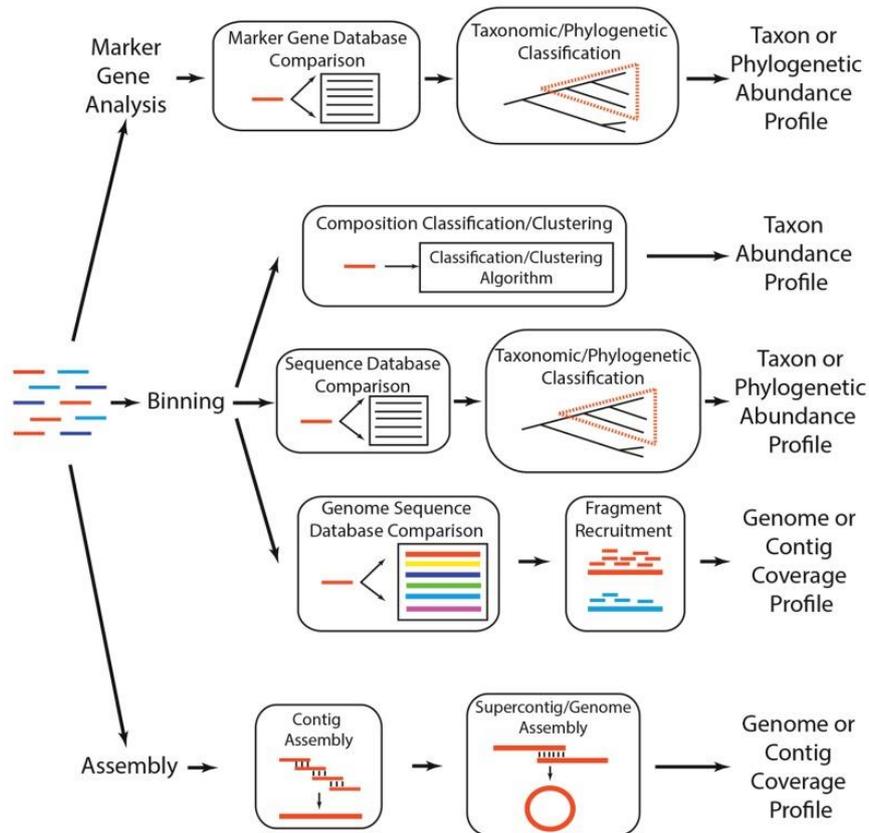


Figura 2.3 Strategie analitiche per determinare quali taxa sono presenti in un metagenoma (Sharpton T. J., 2014)

2.2.1.1 Geni marker

L'analisi con geni marker è uno dei metodi più efficienti per quantificare la diversità tassonomica metagenomica. Questa procedura consiste nella comparazione delle reads metagenomiche ad un database in cui vi sono geni che forniscono informazioni tassonomiche appunto ovvero geni marker. Questo permette di identificare quali reads fanno parte di geni omologhi e di determinare la similarità filogenetica con le sequenze nel database. I geni marker maggiormente utilizzati sono geni rRNA o geni che codificano proteine che tendono ad essere copie singole comuni ai genomi microbici. Dal momento che questo approccio include la comparazione delle reads in piccoli database, l'analisi per stimare la diversità metagenomica è relativamente rapida. Vi sono dei metodi generali con cui vengono utilizzati i geni marker. Il primo fa riferimento alla similarità tra la read da analizzare e i geni marker. Ad esempio MetaPhyler utilizza i risultati della ricerca di una sequenza accoppiata tra reads metagenomiche e un database di geni marker che considerano la classificazione (es famiglia) e proprietà delle reads come la lunghezza. Il secondo metodo utilizza informazioni filogenetiche, più lunghe da calcolare ma forniscono informazioni più accurate. Per esempio AMPHORA (Wu M. *et al.*, 2012), utilizza i modelli HMMs per identificare gli omologhi metagenomici di geni codificanti

proteine o che forniscono informazioni filogenetiche, che sono comuni a genomi sequenziati di batteri e archea.

Vi sono alcune lacune in questo metodo: la strategia opera sotto l'assunzione che la piccola frazione metagenomica omologa ai geni marker considerati sia rappresentativa della distribuzione tassonomica intera della comunità considerata. Mentre i ricercatori cercano di identificare i geni marker che sono presenti in più genomi, le sequenze genomiche disponibili ai ricercatori durante l'identificazione di un gene marker può non essere adeguata per riflettere la diversità dei genomi presenti nella comunità investigata. Inoltre l'analisi con geni marker non è appropriata per quei taxa per cui non sono ancora stati identificati geni marker certi. Grazie alle scoperte recenti sui geni marker specifici per una classificazione tassonomica (Darling A. E. *et al.*, 2014) e l'espansione delle sequenze all'interno dei database per identificare la diversità filogenetica (Wu D. *et al.*, 2009) questo problema è in via di risoluzione. È chiaro che la maggiore disponibilità di sequenze che discriminano la diversità filogenetica di una comunità, limita sempre di più queste problematiche.

2.2.1.2 Binning

Con il Binning, generalmente, ogni sequenza è classificata (binned) in un gruppo tassonomico (genere, famiglia) attraverso il confronto con geni di riferimento o viene raggruppata con altre sequenze, in base ad esempio al contenuto di GC: in questo modo le sequenze possono essere ricondotte ad uno specifico gruppo tassonomico con caratteristiche condivise (es il contenuto di GC). In primo luogo, a seconda del metodo utilizzato, il binning può fornire indicazioni sulla presenza di nuovi genomi difficilmente identificabili. In secondo luogo, può fornire informazioni sul distinto numero e tipo di taxa presenti nella comunità. Inoltre, il binning fornisce un modo per ridurre la complessità dei dati, in modo tale che le analisi post-binning (ad es. montaggio) possano essere eseguite indipendentemente per ogni lettura piuttosto che sulla totalità dei dati. Il binning può essere eseguito su dati assemblati o non assemblati, anche se la maggior parte degli algoritmi riesce ad essere più preciso nel momento in cui aumenta la lunghezza delle reads. Gli algoritmi di binning generalmente utilizzano tre modalità operative:

- Composizione di sequenza
- Somiglianza di sequenza
- Reclutamento dei frammenti

La composizione delle sequenze utilizza caratteristiche di sequenza del metagenoma (es frequenza di tetrameri) per raggruppare o classificare sequenze in gruppi tassonomici. Questi metodi generalmente non richiedono l'allineamento a un database di sequenze di riferimento, e di conseguenza, possono

elaborare in maniera relativamente rapida grandi metagenomi. Alcuni di questi metodi invece analizzano le sequenze intere del genoma a priori per formare classificazioni che stratificano le sequenze in gruppi tassonomici. Ad esempio Phymm (Brady A. *et al.*, 2009; Brady A. *et al.*, 2011) utilizza modelli interpolati Markov (Sharpton T. J., 2014), che combinano la probabilità predittiva di appartenenza ad una linea filogenetica in base alla lunghezza delle reads. Altri metodi utilizzano le sequenze caratteristiche per raggruppare le reads in gruppi senza l'utilizzo di database di riferimento. In questo modo si potrebbero identificare organismi precedentemente ignoti. Ad esempio ESOMs (Emergent Self-Organizing Maps) può essere utilizzato per raggruppare le reads in base alla frequenza di quattro nucleotidi e calcolare la distribuzione dell'abundance (Dick G. J. *et al.*, 2009). Questa metodologia non fornisce direttamente l'identificazione tassonomica, tuttavia è utile per raggruppare le reads e assemblarle in genomi quasi completi che rappresentano microrganismi ancora non tipizzati (Wrighton K. C. *et al.*, 2012). Il raggruppamento prima si basa sul contenuto di GC poi sulla frequenza tetranucleotidica.

Le reads possono essere anche classificate in base alla similarità con sequenze presenti all'interno di database. Questo metodo richiede maggiori ricerche computazionali dal momento che ogni read viene allineata e confrontata con un ampio numero di sequenze. Questo metodo ovviamente non è ideale per l'identificazione di nuovi genomi, tuttavia un metodo basato sulla similarità delle reads permette di avere una maggiore accuratezza e risoluzione. Uno dei maggiori programmi utilizzati è MEGAN che effettua la comparazione delle reads con un database in cui le sequenze sono inserite in base alla classificazione tassonomica NCBI (Huson D. H. *et al.*, 2011). Viene utilizzato anche MG-RAST che fa riferimento alla ricostruzione filogenomica di sequenze presenti all'interno di database a cui è simile la read analizzata.

2.2.1.3 Assemblaggio

In alcuni casi, i genomi completi o quasi completi possono essere assemblati e questo può fornire una panoramica sulla composizione genomica dei microrganismi non coltivabili (Ruby J. G. *et al.*, 2013). Il rischio di questo metodo è l'assemblaggio di genomi distinti in chimere a causa della somiglianza. Il rischio di generare chimere è maggiore in comunità complesse (Luo C. *et al.*, 2012).

L'assemblamento è limitato al taxa più abbondante della comunità microbica che si sta analizzando: senza un sequenziamento estensivo, risulta difficile assemblare genomi di microrganismi rari. Inoltre, l'assemblaggio può produrre chimere che devono essere utilizzate con cautela. Le regioni ripetute all'interno di un genoma sono difficoltose da assemblare. La combinazione di reads lunghe (PacBio) e read piccole (Illumina) nella stessa operazione di assemblaggio può limitare l'errore, nonostante vi

siano pochi strumenti oggi in grado di combinare questi dati (Deshpande V. *et al.*, 2013). Infine, l'assemblaggio richiede una elaborazione di dati molto complessa e molta RAM. Utilizzare un'operazione di binning prima dell'assemblaggio può ridurre la complessità computazionale.

2.2.2 Comprensione delle funzioni biologiche

La metagenomica cerca di chiarire quali siano le funzioni collettive che sono codificate all'interno dei genomi degli organismi che fanno parte di una comunità. La diversità funzionale di una comunità può essere quantificata correlando le sequenze metagenomiche a particolari funzioni (Figura 2.4). Le reads che contengono sequenze che codificano per precise proteine vengono identificate tramite particolari algoritmi (*gene prediction*). Ogni sequenza codificante viene poi confrontata con altri geni, proteine, famiglie di proteine o pathway metabolici presenti nel database, di cui sono conosciute alcune informazioni funzionali (*functional annotation*) (Figura 2.4).

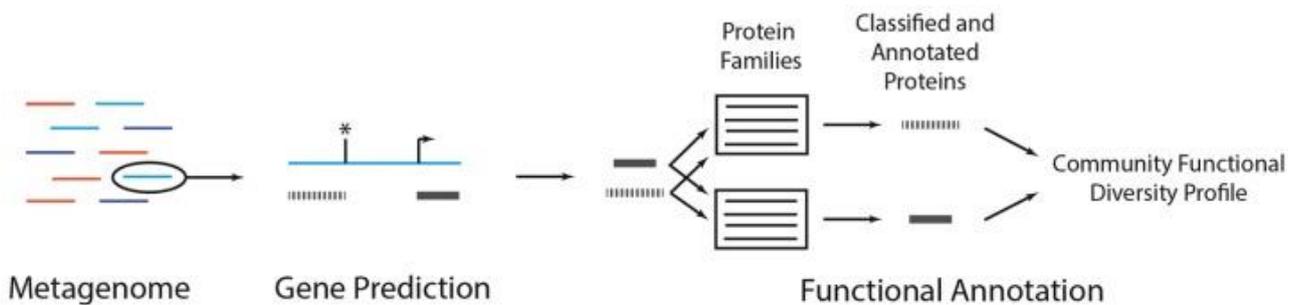


Figura 2.4 Workflow per l'analisi funzionale in metagenomica (Sharpton T. J., 2014)

Questa analisi metagenomica produce un profilo che descrive il numero di funzioni distinte e la loro relativa abundance e può essere utilizzata per confrontare vari metagenomi e identificare comunità microbiche funzionalmente simili (Human Microbiome Project Consortium, 2012), sapendo che specifici trattamenti influenzano la composizione di una comunità microbica (Looft T. *et al.*, 2012) e che le funzioni biologiche cambiano in base all'ambiente o a variabili fisiologiche (Morgan X. C. *et al.*, 2012). I metagenomi derivanti dal sequenziamento possono anche permettere di scoprire nuovi geni (Nacke H. *et al.*, 2012) o fornire informazioni sulle condizioni ecologiche associate a quei geni di cui non si conosce ancora la funzione (Buttigieg P. L. *et al.*, 2013).

2.2.2.1 Previsione genica

La previsione del gene determina quali reads metagenomiche contengono sequenze codificanti. Una volta identificate, queste sequenze codificanti possono essere correlate ad una precisa funzione. La previsione genica può essere condotta su sequenze metagenomiche assemblate o non assemblate. Per le sequenze assemblate, la previsione genica è molto simile all'analisi di sequenza che viene effettuata

sull'intero genoma, tramite algoritmi previsionali che richiedono parametri specie-specifici. Le sequenze possono essere anche non sottoposte ad assemblaggio, nel momento in cui si hanno sequenze parzialmente codificanti in cui la parte codificante, ovvero il gene, si trova all'estremità iniziale o finale della read. Uno dei metodi più semplici per identificare una sequenza codificante è il binning che confronta le reads con sequenze geniche presenti all'interno di un database. Questo metodo permette anche di conoscere la funzionalità del gene analizzato (Deshpande V. *et al.*, 2013). Questa metodica è stata utilizzata per quantificare la diversità genetica nelle comunità microbiche marine (Rusch D. B. *et al.*, 2007), il microbiota intestinale (Qin J. *et al.*, 2010) ed è utile per catalogare i geni specifici presenti in un metagenoma. Questo metodo è una previsione genica high-throughput poiché si affida ad algoritmi che riescono rapidamente ad associare un frammento ad una sequenza molto simile presente nel database. Se l'obiettivo è identificare nuovi geni o geni altamente differenti all'interno di una comunità microbica questo non è il metodo migliore.

Un altro metodo per la previsione genica prevede la trasduzione di ogni read in sei possibili proteine codificanti e il confronto di ogni proteina risultante in un database di sequenze proteiche mediante l'allineamento delle sequenze. L'allineamento delle sequenze può essere utile per identificare quindi quelle sequenze metagenomiche che codificano per peptidi omologhi alle proteine presenti nel database.

La previsione genica può anche identificare geni nuovi. I modelli di previsione genica valutano le proprietà dei geni microbici (lunghezza delle reads, contenuto di GC, utilizzo di codoni) al fine di verificare se la read che si sta analizzando non è simile ad una sequenza presente nel database. Quindi questo metodo permette di identificare geni che hanno proprietà comuni ad altri geni microbici ma che tuttavia differiscono da geni che sono già stati scoperti. Vi sono diversi software che permettono di effettuare una predizione di nuovi geni come MetaGene (Noguchi H. *et al.*, 2006), Glimmer-MG (Kelley D. R. *et al.*, 2012), MetaGeneMark (Zhu W. *et al.*, 2010).

2.2.2.2 Associazione genica ad una funzione biologica

Una volta individuate le sequenze codificanti, esse possono essere associate ad una precisa funzione biologica. Il metodo più comune è tramite la classificazione delle proteine codificate dai geni individuati in famiglie di proteine. Una famiglia di proteine è un gruppo di proteine correlate nella sequenza codificante dal punto di vista evolutivo (Finn R. D. *et al.*, 2013). Dal momento che le proteine all'interno di una famiglia hanno un predecessore comune, si pensa che proteine simili abbiano funzioni biologiche simili. La classificazione di una proteina metagenomica all'interno di una famiglia di proteine richiede la sua comparazione all'interno di un database in cui sono presenti

sequenze note che codificano per determinate proteine (database di sequenza) o il confronto della sequenza con un modello probabilistico che descriva la diversità delle proteine in una famiglia (database HMMs). I database di sequenza come SEED o MG-RAST permettono di confrontare le reads con quelle presenti nelle loro banche dati e consentono di collegarle a specifiche funzioni biologiche (Overbeek R. *et al.*, 2013). Questo fa in modo che l'analisi sia relativamente veloce rispetto all'utilizzo di database HMMs. La classificazione proteica delle reads richiede calcoli complessi in quanto tutti i peptidi vengono confrontati a tutte le sequenze della proteina presenti nel database. Ogni confronto è indipendente per cui i cluster di calcolo e i server possono distribuire parallelamente il carico computazionale e migliorare il throughput ottenuto. I server web che interagiscono con questi cluster per effettuare una predizione genica, la ricerca nel database e la classificazione in una famiglia sono ad esempio MG-RAST (Meyer F. *et al.*, 2011) e IMG/M (Markowitz V. M. *et al.*, 2013).

Mentre la classificazione proteica delle reads metagenomiche è utile per riuscire a identificare una funzione comune, il metodo è imperfetto. Dal momento che una comunità microbica è assai eterogenea, la diversità funzionale codificata nel metagenoma può solo approssimare l'attività funzionale di una comunità. Inoltre la maggior parte dei database contengono famiglie la cui funzione biologica non è ancora nota per cui le reads omologhe a queste famiglie di proteine non riescono a trovare ancora un'associazione funzionale. Inoltre ogni database utilizza approcci differenti quindi ognuno può inserire differenti porzioni geniche che possono poi produrre profili funzionali differenti per descrivere una comunità microbica. Gli strumenti web che gestiscono i dati metagenomici forniscono ai ricercatori mezzi per confrontare le sequenze e attribuirgli una funzione (Tabella 2.1). Le analisi da condurre, il focus della ricerca, il progetto sperimentale e gli obiettivi rimangono ancora talvolta motivo di confusione per i ricercatori. Ad esempio se si devono conoscere i pathway metabolici che maggiormente rappresentano una comunità microbica, è bene utilizzare database come KEGG o Meta Cyc. Un'analisi finalizzata alla conta dei differenti tipi di proteine farà uso di un database che permetta di risaltare la diversità filogenetica. Se si sa che la comunità contiene linee filogeneticamente divergenti in base ai database, è meglio utilizzare tecniche tassonomiche meno rigide sulla divergenza di sequenza. Se l'obiettivo principale è quello di caratterizzare il genoma di un organismo molto presente (abundance) nella comunità microbica, allora è preferibile utilizzare l'assemblaggio.

Resource	Methods
AmphoraNet	Marker gene analysis: phylogeny
CAMERA	Various: taxonomic and functional annotation, comparative analyses
Comet	Functional annotation, comparative analyses
LEfSe (Galaxy)	Comparative analyses
IMG/M	Various: taxonomic and functional annotation, comparative analyses
MG-RAST	Various: taxonomic and functional annotation, comparative analyses
MALINA	Various: taxonomic and functional annotation, comparative analyses
METAGENassist	Various: taxonomic annotation, comparative analyses
MetaPhlAn (Galaxy)	Marker gene analysis: similarity
NBC	Binning: compositional classification
Orphelia	Gene prediction
PhyloPithia webservice	Binning: compositional classification
Real time metagenomics	Functional annotation
WebCARMA	Binning: sequence similarity
WebMGA	Various: taxonomic and functional annotation

*Tabella 2.1 Ricerche metagenomiche basate su web-databases
(Sharpton T. J., 2014)*

2.5 MG-RAST

Per analizzare i dati biologici che descrivono sequenze di geni, composizione e struttura delle proteine, processi biochimici nelle cellule, etc. vengono utilizzati strumenti informatici. Diversi software applicativi, anche *open source*, sono stati elaborati e messi a disposizione in internet. Essi possono essere utilizzati per l'analisi dei dati di sequenziamento derivati dalle tecnologie NGS. MG-RAST, è un software bioinformatico open source, con all'interno un database, in grado di effettuare la previsione genica e l'associazione funzionale. Nel momento in cui viene effettuato l'upload dei dati ottenuti dal campione, questi sono confrontabili con quelli contenuti nel database. Il software riesce a dare informazioni relative ad abundance e metabolic pathway. Le sequenze vengono normalizzate, elaborate e i risultati vengono generati automaticamente. Dopo aver effettuato l'upload del dataset il sistema inizia un controllo di qualità e permette all'utilizzatore di settare i parametri per l'analisi filogenetica. Il sistema rende visibili nel display i risultati e permette all'utilizzatore di modificare i parametri (Figura 2.5). Dal momento che i sistemi di *genome annotation* sono in continua evoluzione, per fare in modo di inserire nuovi criteri che potrebbero essere sviluppati, la pipeline è stata progettata con un framework modulare che consente di aumentare gli step di analisi dei

frammenti. Il server offre diversi metodi come la ricostruzione filogenetica e l'individuazione di pathway metabolici con cui accedere ai dati diversi. Inoltre, il server offre una completa capacità di ricerca. L'accesso ai dati è protetto da password e tutti i dati generati dalla pipeline automatica sono disponibili per il download e l'analisi in vari formati comuni (Meyer F. *et al.*, 2011).

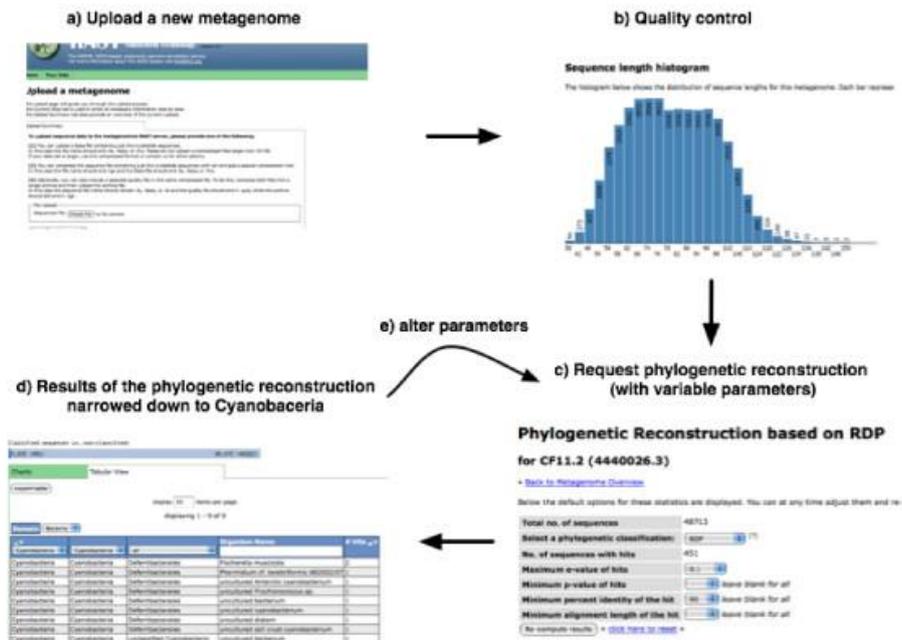


Figura 2.5 Step di analisi dei dati in MG-RAST
(Meyer F. *et al.*, 2011)

I primi dati prodotti da MG-RAST sono i profili di abundance. MG-RAST inoltre effettua la previsione dei geni presenti nel metagenoma e identifica gli omologhi più vicini a quei geni presenti nei genomi isolati tramite BLAST (BLAST-like alignment tool). BLAST non vede somiglianze al di sotto del 70%, in modo da non effettuare associazioni con altri geni. Nel momento in cui si ottiene la migliore sovrapposizione di geni derivanti da un genoma isolato, tutte le analisi successive vengono fatte utilizzando i geni dei genomi isolati e non i geni metagenomici. Questo crea molte limitazioni poiché l'analisi non viene eseguita sui geni originali del metagenoma ma sui geni “proxy”. Il vantaggio di questo metodo è la sua velocità computazionale. L'unico step informatico intenso è trovare le corrispondenze migliori del metagenoma da analizzare con gli isolati. Una volta fatto, tutte le comparazioni sono già preimpostate (Oulas A. *et al.*, 2015; Kent W. J., 2002).

2.6 MGmapper

MGmapper è un portale web sviluppato da Petersen T. N. *et al.*, (2017) per la mappatura di sequenze metagenomiche. MGmapper funziona con i file fastq e una ha varietà di pipeline con strumenti personalizzabili così da creare degli script che forniscono annotazioni tassonomiche e file statistici riepilogati.

MGmapper elabora i dati del sequenziamento ottenuti tramite le tecnologie NGS assegnando ogni read al genoma di riferimento ogni read e consentendo una classificazione tassonomica affidabile fino al livello di specie.

MGmapper è stato creato per permettere di effettuare una analisi di routine di set di dati complessi, tramite l'uso di molti database con sequenze di riferimento per batteri, virus, funghi, piante, mammiferi vertebrati, invertebrati. Consente anche l'uso di banche dati genetiche che contengono geni di resistenza antimicrobica, rRNA 16S o di qualsiasi database personalizzato basato su un set di sequenze fasta. Inoltre, la maggior parte delle analisi viene eseguita in parallelo per un'esecuzione rapida.

Il software MGmapper consiste in tre fasi principali: pre-elaborazione delle reads, una loro mappatura sui database che contengono le sequenze di riferimento e post-elaborazione dei risultati. La fase di pre-elaborazione viene eseguita utilizzando cutadapt. Il cutadapt taglia le estremità di bassa qualità dalle reads e in seguito esegue la rimozione dell'adattatore. Nella seconda fase, le reads già tagliate sono allineate a un set di database di sequenze di riferimento usando bwa mem. Le reads possono essere identificate in fullmode o bestmode. In modalità full, tutte le reads vengono sottoposte ad allineamento. Nel bestmode, viene eseguita la mappatura in tutti i database, simultaneamente, ma viene scelta la coppia di reads di allineamento migliore. Una coppia di reads è considerata valida solo se la somma dei punteggi di allineamento (SAS) è superiore a qualsiasi valore SAS del database. La versione attuale di MGmapper comprende 18 databases (Figura 2.6) (Petersen T. N. *et al.*, 2017).

Click to 'Show' available databases:

Id	Database	Update date	Species	Sequences	Remark
1	Archaea	20151124	228	3487	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome,rep genome
2	Bacteria	20151124	1847	6969	genome_rep=Full; assembly_lvl=Complete Genome,Chr
3	MetaHitAssembly	20140111	373	41126	MetaHitAssembly project
4	HumanMicrobiome	20140409	456	42645	Human Microbiome Project
5	Bacteria_draft	20151124	1964	137980	genome_rep=Full; refseq_cat=rep genome
6	ResFinder	20160103	-	2128	Resistance genes
7	Human	20151021	1	25	GRCh38.p3 chr and Mt
8	Virus	20151124	2515	6640	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome,rep genome
9	Fungi	20151124	628	2235299	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome, rep genome
10	Protozoa	20151124	2404	2959545	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome, rep genome
11	Plasmid	20151124	683	4085	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome,rep genome
12	Plant	20151124	172	26338014	genome_rep=Full; assembly_lvl=Complete Genome,Chr or refseq_cat=ref genome,rep genome
13	Vertebrates_mammals	20151125	119	15301228	genome_rep=Full; assembly_lvl=Complete Genome or refseq_cat=ref genome,rep genome
14	Vertebrates_other	20151125	12556	14798671	genome_rep=Full; assembly_lvl=Complete Genome or refseq_cat=ref genome,rep genome
15	Invertebrates	20151125	358	24978725	genome_rep=Full; assembly_lvl=Complete Genome or refseq_cat=ref genome,rep genome
16	GreenGenes	20140708	-	1012495	ver 13.5
17	Silva	20150722	-	571510	release 123
18	NonFluViruses	20150902	-	51328	Virus pathogen resource (http://www.viprbrc.org/brc/)

Reference and representative genomes from genbank or refseq include genomes assigned as either 'reference' or 'representative'. However, some species are still not assigned with a status as a reference or representative genome. In those cases one species is selected such that all species are present with at least one genomic sequence.

Figura 2.6 Elenco dei database attualmente disponibili su MGmapper (Petersen T. N. et al., 2017)

2.7 Applicazioni di analisi metagenomiche agli alimenti

Le tecniche HTS (High-throughput sequencing) hanno modificato il modo di approcciarsi all'ecologia microbica, considerando la popolazione microbica come *consortia* (Cocolin L. et al., 2015). In campo alimentare, lo scopo dell'analisi può essere riferita a patogeni, contaminazioni associate o microrganismi benefici e colture starter. Ad esempio, il sequenziamento di ampliconi di sequenze target, in particolare di rRNA è stato ampiamente utilizzato nella decade precedente per monitorare le comunità microbiche durante la fermentazione ed è il metodo più comune per studiare l'ecologia microbica associata agli alimenti. Gli ampliconi di rRNA sono stati utilizzati per studiare la fermentazione e la contaminazione di alimenti tra cui, in particolare, i prodotti lattiero-caseari. Ad esempio il microbiota nel formaggio varia a seconda dell'origine animale del latte, a seconda degli ingredienti utilizzati e in base al latte, se crudo o pastorizzato (Quigley L. et al., 2012). Inoltre, l'elevato contenuto di sale in alcuni formaggi produce l'assenza di *Leuconostoc* e *Pseudomonas*. In alcuni studi è stato condotto il monitoraggio della fermentazione della cagliata (De Filippis F. et al., 2014), l'analisi della stagionatura di un formaggio (De Pasquale I. et al., 2014; De Filippis F. et al., 2016) e della distribuzione microbica in differenti parti del formaggio stesso (De Pasquale I. et al., 2016).

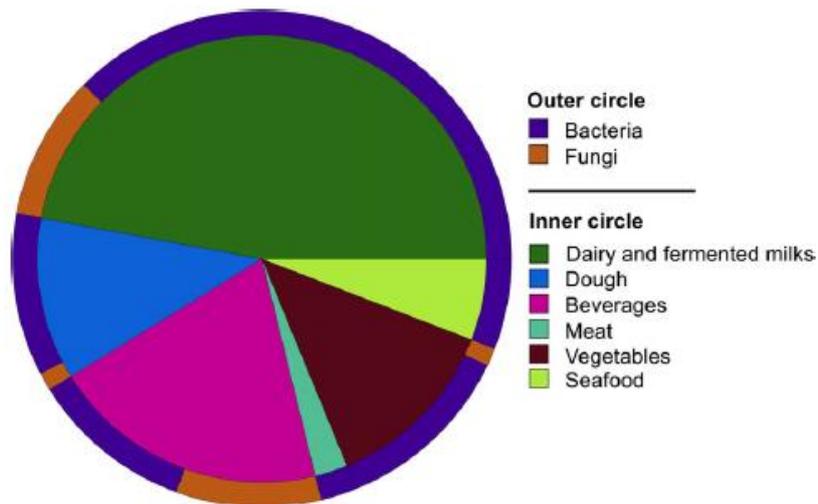


Grafico 2.1 Grafico che mostra l'abbondanza di studi HTS di alimenti e bevande fermentati raggruppati secondo la matrice alimentare. Per ciascun prodotto nel cerchio esterno si evidenzia la proporzione di studi che analizzano le comunità batteriche o fungine
(De Filippis F. et al., 2017)

È stata studiata anche l'evoluzione microbica durante la fase di maturazione: in un formaggio americano all'inizio del processo fermentativo il microbiota dominante è risultato composto da *Enterobacteriaceae* mentre negli stadi successivi dominavano lieviti e *Lactobacillales* (Bokulich N. A. et al., 2012).

L'analisi della flora microbica tramite sequenze target (HTS) è stata utilizzata anche per monitorare i cambiamenti durante la conservazione di carne per 40 giorni in atmosfera modificata (MAP): *B. thermosphacta* e *C. divergens* sono risultati i batteri predominanti. Il sequenziamento risulta utile anche per scoprire contaminazioni patogene: un ceppo non necessariamente patogeno di *E. coli* metabolicamente attivo (tramite pyrosequencing del DNA) è stato trovato in un formaggio danese, mentre la sua concentrazione diminuiva durante la fermentazione (Masoud W. et al., 2012).

I patogeni che si sviluppano all'interno degli alimenti sono spesso analizzati per riuscire a monitorare il loro destino e percorso durante il processo produttivo, durante la fermentazione o durante la conservazione (De Filippis F. et al., 2017). Per questo motivo molti saggi PCR gruppo-, genere, specie-specifici sono stati sviluppati per determinare i patogeni tramite appunto sequenziamento di geni target. Nei dati ottenuti tramite sequenziamento di ampliconi si deve sempre tener presente l'errore della PCR (Sipos R. et al., 2010), determinato ad esempio dall'amplificazione preferenziale dei primer nei confronti di alcune specie (Pinto A. J. et al., 2012).

Tramite il sequenziamento del metagenoma ottenuto con shotgun, si ottengono moltissimi vantaggi in più rispetto al metodo basato sul target di ampliconi. Infatti, si riesce ad ottenere un quadro generale dell'intera comunità microbica presente, tracciando e confrontando contemporaneamente l'abundance di batteri e altri microrganismi, direttamente all'interno della matrice alimentare.

L'output dell'analisi metagenomica, basata sul sequenziamento dell'intero DNA frammentato è l'insieme delle attività potenziali delle comunità microbiche e lo studio dell'ecologia microbica, risulta fondamentale sia per lo sviluppo delle fermentazioni sia per la comprensione delle interazioni microbiche che permettono di avere processi produttivi di qualità all'interno di alimenti fermentati (De Filippis F. *et al.*, 2017). L'analisi metagenomica permette di studiare i microrganismi alimentari e di capire come le loro potenziali funzioni possono essere modificate o modulate per migliorare la qualità degli alimenti. Lo shotgun permette di recuperare genomi microbici completi raggiungendo un'identificazione a livello di ceppo (De Filippis F. *et al.*, 2017), tuttavia, anche se metagenomica e metatrascrittomica permettono di avere più informazioni, sono tecniche ancora molto costose. Alcune applicazioni metagenomiche al settore alimentare riguardano ad esempio l'espressione genica microbica durante la maturazione di formaggio in superficie (Dugat-Bony E. *et al.*, 2015): si è evidenziato che geni implicati della produzione di flavour a partire dal metabolismo aminoacidico sono maggiormente espressi nella prima fase di maturazione. Erkus O. *et al.* (2013) ha caratterizzato la coltura starter indefinita di un formaggio e, nonostante siano state trovate solo due specie, è stata identificata un'elevata diversità a livello di ceppo. La comparazione genomica ha permesso di dedurre che fosse presente un meccanismo di competizione: la sensibilità ai fagi del ceppo più resistente dipendeva dalla sua densità e questo ha impedito l'eradicazione di altre linee genetiche. La genomica comparata dei ceppi industriali sta fornendo una comprensione più ricca e approfondita della composizione genetica e variabilità in questi importanti microrganismi e questo promette di individuare rapidamente i loci genici che formano tratti significativi dal punto di vista industriale. Lo shotgun è stato utilizzato anche per definire comunità di virus in gamberi fermentati e crauti (Park E. J. *et al.*, 2011): sono state trovate alcune discrepanze tra la previsione dei batteriofagi con un confronto di omologia effettuato nello shotgun e il sequenziamento di rRNA 16S. Questa discrepanza mostra le lacune presenti nei database pubblici dei genomi virali degli alimenti fermentati (Ercolini D. 2013). Tramite lo shotgun sequencing è stata studiata l'ecologia e la fonte di contaminazione in un macello, cercando di capire se la contaminazione delle carcasse derivasse dalle feci o dalla pelle. Questo lavoro ha prodotto circa due milioni di sequenze di DNA, che hanno permesso di identificare 3641 specie differenti tra pelle, carcassa e feci. Lo studio ha dimostrato che la metagenomica può essere un potente strumento per comprendere l'ecologia microbica nel macello. L'analisi ha dimostrato che i batteri sulle carcasse sono probabilmente derivati dalla pelle (60%) e altre fonti ambientali. Gli organismi di origine fecale sono la maggior parte e mostrano il maggiore livello di diversità senza specie dominanti nelle popolazioni. Non sembra che le pelli altamente contaminate, contaminano la carcassa ma più probabilmente che contribuiscano ad una contaminazione

generalizzata nell'ambiente di lavorazione. L'elevato livello di contaminazione di microrganismi non fecali delle carcasse a partire dalla pelle, l'elevata presenza di *S. epidermidic* e la discrepanza tra le pelli altamente contaminate da microrganismi fecali e le carcasse associate dimostra che ulteriori esperimenti potrebbero delucidare chiarire i risultati dell'analisi. Nel lavoro è stato chiarito che le acque di lavaggio sono una fonte assolutamente importante per la contaminazione microbica, derivano dalla contaminazione fecale e molti batteri colonizzano l'ambiente a partire dalle acque di lavaggio domestiche contaminate. I batteri patogeni sono presenti in acque di lavaggio perché qui sono in grado di sviluppare rapidamente ed un esempio sono le acque di lavaggio dei vegetali WWTP (Wastewater treatment plants) (Cai L. *et al.*, 2013). La presenza di patogeni in WWTP può rappresentare un rischio molto grave per la salute pubblica e tecniche basate su PCR sono applicate di routine per queste acque ma sono progettate specificatamente per certi target e non possono fornire un quadro completo di tutti i patogeni (Toze S., 1999). Nello studio di Cai L. *et al.*, (2013) è stato effettuato il sequenziamento metagenomico tramite Hi-Seq di Illumina di 15 campioni di DNA da WWTP e sono stati scoperti 24 data set di DNA metagenomico. Ogni set è stato normalizzato a 12500000 sequenze di DNA con una lunghezza di 150-190 bp. Per individuare la distribuzione, diversità e l'abundance dei patogeni umani sono stati usati BLASTN, BLASTX e MetaPhlAn. In questo modo sono stati identificati nove patogeni; l'abundance complessiva dei batteri patogeni è stata stimata al 0,06-3,20% nel totale della popolazione di batteri utilizzando il fingerprinting del gene rRNA 16S. I batteri patogeni rilevati nei fanghi attivi e negli effluenti condivisi avevano profili analoghi, ma erano diversi in base al gene di rRNA 16S e sul fingerprinting basato su fattori di virulenza. Questi risultati hanno fornito un profilo completo di patogeni umani comunemente interessati in WWTP e si è dimostrato che lo shotgun sequencing è un attuabile ed efficace approccio per la rilevazione ambientale di patogeni batterici umani.

2.8 Workflow della analisi metagenomica

Indipendentemente dalla tecnologia utilizzata, è possibile riconoscere tre fasi comuni a tutte le metodiche di sequenziamento. A partire dal campione alimentare, sono necessari una serie di step, che devono essere ottimizzati:

- Estrazione e purificazione dell'acido nucleico
- Costruzione della library
- Sequenziamento
- Analisi delle sequenze (Figura 2.7).

Nel caso di campioni alimentari, si effettuano trattamenti che combinano azioni meccaniche ed enzimatiche (Linnarsson S., 2010). Inoltre, alcuni componenti alimentari come proteine e lipidi, possono interferire con il DNA, non permettendo una analisi accurata, quindi questi componenti devono essere rimossi durante l'estrazione dell'acido nucleico (Bonaïti C. *et al.*, 2006).

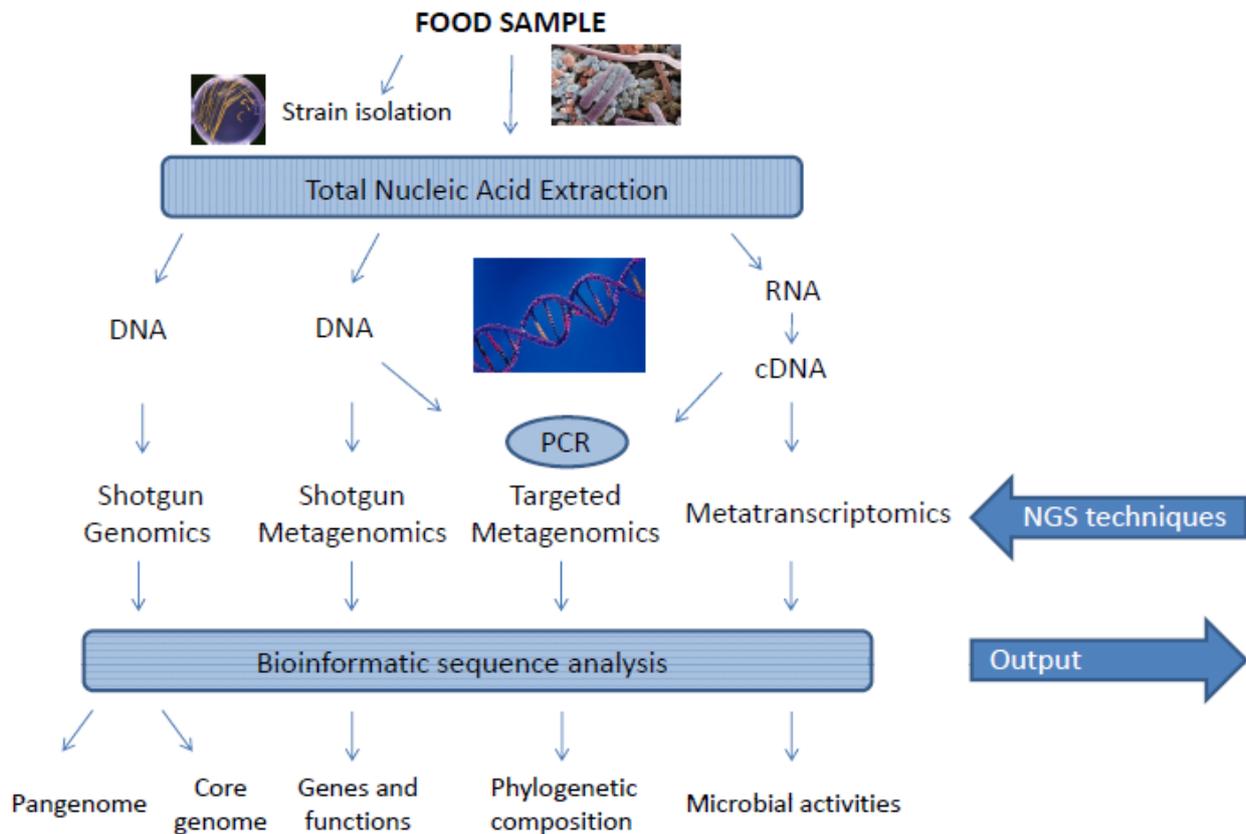


Figura 2.7 Schema di lavoro per il sequenziamento in food
(Mayo B. *et al.*, 2014)

2.8.1 Estrazione e purificazione dell'acido nucleico

Negli studi microbiologici, la manipolazione e la conservazione degli alimenti è critica. È sempre consigliabile che il processo sia rapido e devono essere evitate le condizioni che potrebbero alterare la concentrazione di cellule microbiche originale. I sistemi alimentari sono sempre eterogenei, quindi le strategie devono essere attentamente progettate per garantire la rappresentatività dei campioni nelle diverse condizioni che si ritrovano negli alimenti e quindi deve essere effettuata una buona omogeneizzazione. L'incompleta lisi cellulare e la conseguente errata estrazione del DNA può portare a risultati irrealistici sulla comunità microbica alimentare. Questo può essere evitato eseguendo step di lisi successivi per migliorare l'efficienza di estrazione, ad esempio tramite trattamenti enzimatici e meccanici (Linnarsson S., 2010).

Il controllo della qualità e la quantificazione dell'acido nucleico sono fondamentali in ogni metodo di estrazione di acidi nucleici. Si preferisce effettuare un'accurata quantificazione con un fluorimetro (Linnarsson S., 2010), come il Fragment Analyzer™ (Quail M. A. *et al.*, 2008). La qualità e la quantità di DNA richiesto per il sequenziamento totale (metagenomica shotgun) è di solito superiore a quella necessaria per il sequenziamento 16S (targeted sequencing). Negli alimenti, è fondamentale distinguere tra cellule attive, inattive e vitali: le cellule morte possono contribuire al pool di DNA totale che verrà sequenziato.

Prima di eseguire l'estrazione del DNA, le cellule di comunità microbiche complesse possono essere arricchite con propidio monoazide (PMA) o etidio monoazide (EMA). Questi composti entrano nelle cellule, danneggiandone le membrane, si inseriscono tra la doppia elica di DNA, formando legami covalenti sotto esposizione a luce intensa, evitando così l'amplificazione con PCR (Nocker A. *et al.*, 2006).

2.8.2 Preparazione delle Library

Uno step fondamentale in ogni schema di lavoro NGS è la frammentazione dell'acido nucleico del campione in una library di sequenziamento. La library è quindi l'insieme di frammenti che deve essere sequenziato. Esistono molti protocolli di preparazione delle library ma essi hanno in comune la frammentazione del DNA in frammenti da 50 – 500 bp e l'inserimento conseguente di adattatori alle estremità dei frammenti da sequenziare (Van Dijk E. L. *et al.* 2014). In Figura 1.23 sono illustrate le tecniche principali di preparazione delle library. Gli adattatori lineari (Figura 2.8a) vengono usati in Genome Analyzer e SOLiD in cui le sequenze degli adattatori vengono aggiunte ad entrambe le estremità dei frammenti di DNA. Gli adattatori circolari (Figura 2.8b) vengono utilizzati nelle piattaforme CGA dopo quattro sequenze di adattatori distinti vengono internalizzate in un campione circolare di DNA. Gli adattatori bubble (Figura 2.8c) sono utilizzati nel sistema PacBio, in cui tali adattatori vengono aggiunti ai doppi filamenti di DNA per generare una molecola circolare (Myllykangas S. *et al.*, 2012).

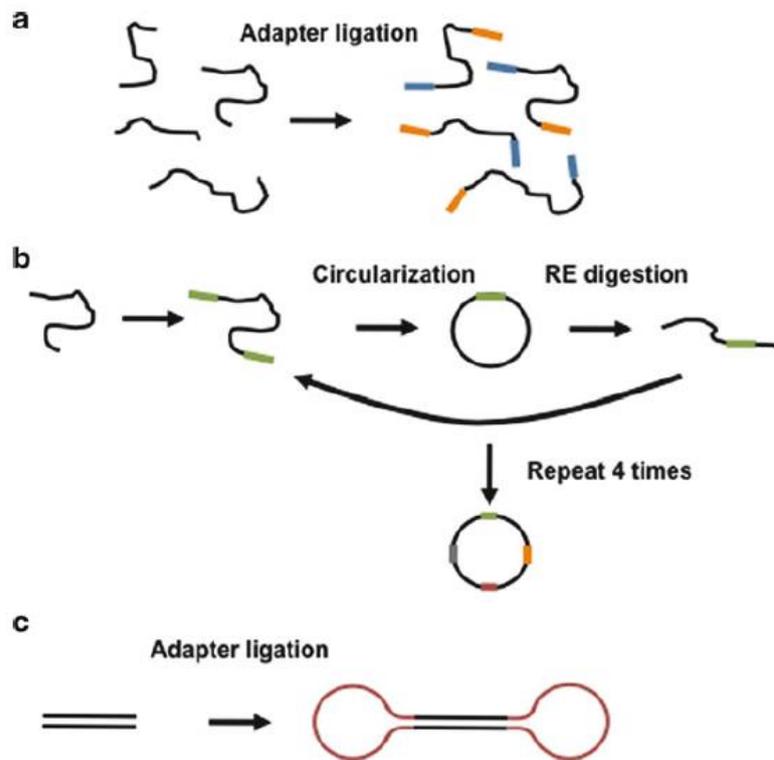


Figura 2.8 Preparazione della library
(Mylykangas S. *et al.*, 2012)

Una differenza fondamentale nella costruzione delle library nel sequenziamento dell'intero DNA (shotgun metagenomic) e di sequenze specifiche (targeted metagenomic) è la necessità per quest'ultima di amplificare prima le sequenze di DNA conservate. Per le sequenze target metagenomic, vi sono piattaforme di adattatori specifici disegnati come parte dei primer di PCR, che danno come risultato una molecola pronta per essere sequenziata (Binladen J. *et al.*, 2007). È possibile sequenziare campioni differenti se si utilizzano piccoli frammenti di DNA specifico detti barcodes (Binladen J. *et al.*, 2007). Questi vengono disegnati come parte dei primer di PCR. Le miscele normalizzate degli ampliconi devono essere fatte prima del sequenziamento, assicurando così che ogni amplicone sia adeguatamente rappresentato nella reazione di sequenziamento (Linnarsson S., 2010).

Nel metodo shotgun, l'amplificazione del DNA non è sempre necessaria, lo diventa solo quando sono disponibili piccole quantità di DNA (Kuehn J. S. *et al.*, 2013). Un metodo con cui è possibile effettuare la preparazione della library di sequenziamento è il NexteraXT (Illumina), successivamente le library vengono normalizzate attraverso Fragment Analyzer™.

La preparazione della library delle tecniche shotgun comincia con la frammentazione del DNA tramite forza meccanica o tramite enzimi di restrizione (Linnarsson S., 2010). In entrambi i casi gli adattatori sono legati ai frammenti di DNA risultanti.

Aggiungere gli adattatori è necessario per permettere ai frammenti di DNA di ancorarsi ad una superficie solida e definire il sito in cui la reazione di sequenziamento avrà inizio. I sistemi di sequenziamento, eccetto PacBio RS, richiedono l'amplificazione della sequenza della library per formare frammenti distinti rilevabili a livello spaziale (Figura 2.9). L'amplificazione può essere effettuata *in situ*, in emulsione o in soluzione per generare cluster di DNA clonale. La PCR in emulsione viene applicata al sistema Genome Sequencing e SOLiD, in cui la PCR avviene sulla superficie di biglie (Figura 2.9a). La bridge PCR è utilizzata per generare cluster di sequenze amplificate *in situ* su un supporto solido (Figura 2.9b). L'amplificazione circolare viene utilizzata per creare lunghi tratti di DNA che si ripiegano in strutture a spirale (nanoballs) e vengono utilizzate nella tecnologia CGA (Figura 2.9c). La polimerasi di DNA biotinilato è utilizzata nel sistema PacBio RS e si lega al campione adattato a forma di bolla (bubble) (Figura 2.9d). Il complesso polimerasi/campione viene immobilizzato sulla parte inferiore di una guida d'onda a modalità zero (ZMW) (Myllykangas S. *et al.*, 2012).

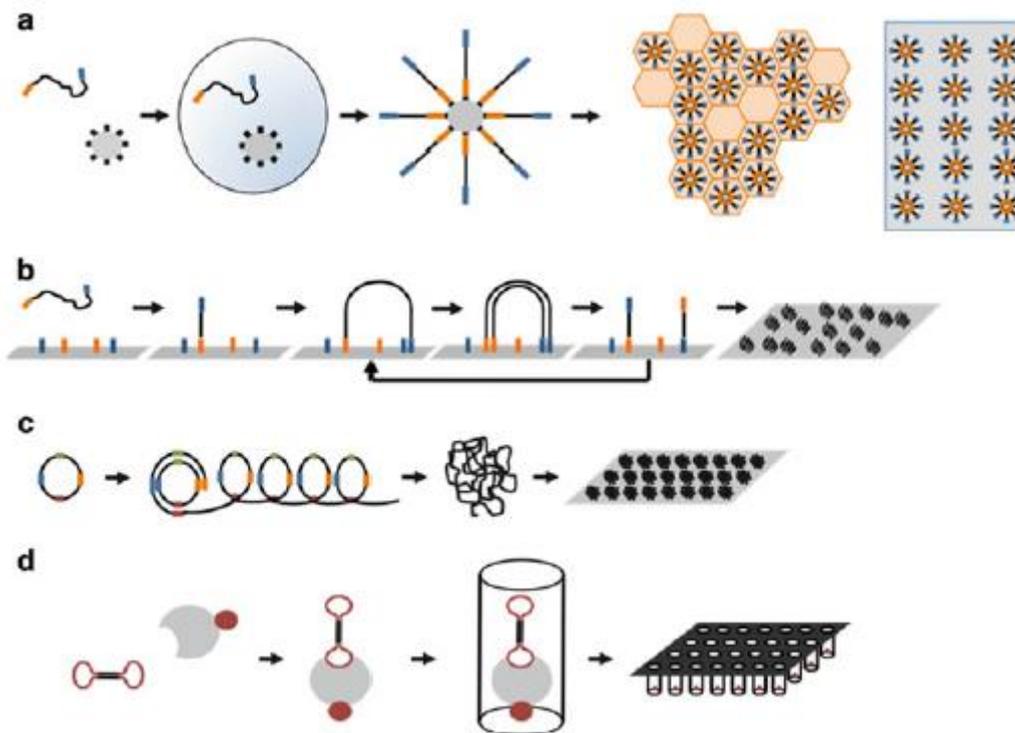


Figura 2.9 Preparazione delle library per le diverse piattaforme di sequenziamento
(Myllykangas S. *et al.*, 2012)

2.8.2.1 Preparazione della library tramite Nextera®XT

La preparazione della library segue ad una amplificazione ad ampio raggio. Il kit di preparazione della library Nextera® XT (Illumina Inc.) è creato per sequenziare ampliconi, genomi di piccole dimensioni e plasmidi. La preparazione della library Nextera® XT utilizza la frammentazione

enzimatica chiamata “tagmentation” in cui gli ampliconi di DNA sono frammentati dalla trasposasi Nextera® XT e simultaneamente legati ad adattatori che permettono 126 cicli di amplificazione in PCR. Può essere sequenziata una vasta gamma di library che offrono frammenti da 200 bp a 1000 - 1500 bp (McElhoe J. A. *et al.*, 2014) (Figura 2.10). In questo modo la procedura riduce il tempo di preparazione e il materiale di input. Infatti è il metodo più veloce per la preparazione dei campioni per ogni piattaforma di sequenziamento Illumina, permette di analizzare fino a 96 campioni e richiede la quantità più bassa di campione di DNA rispetto agli altri metodi (1 ng) (McElhoe J. A. *et al.*, 2014).

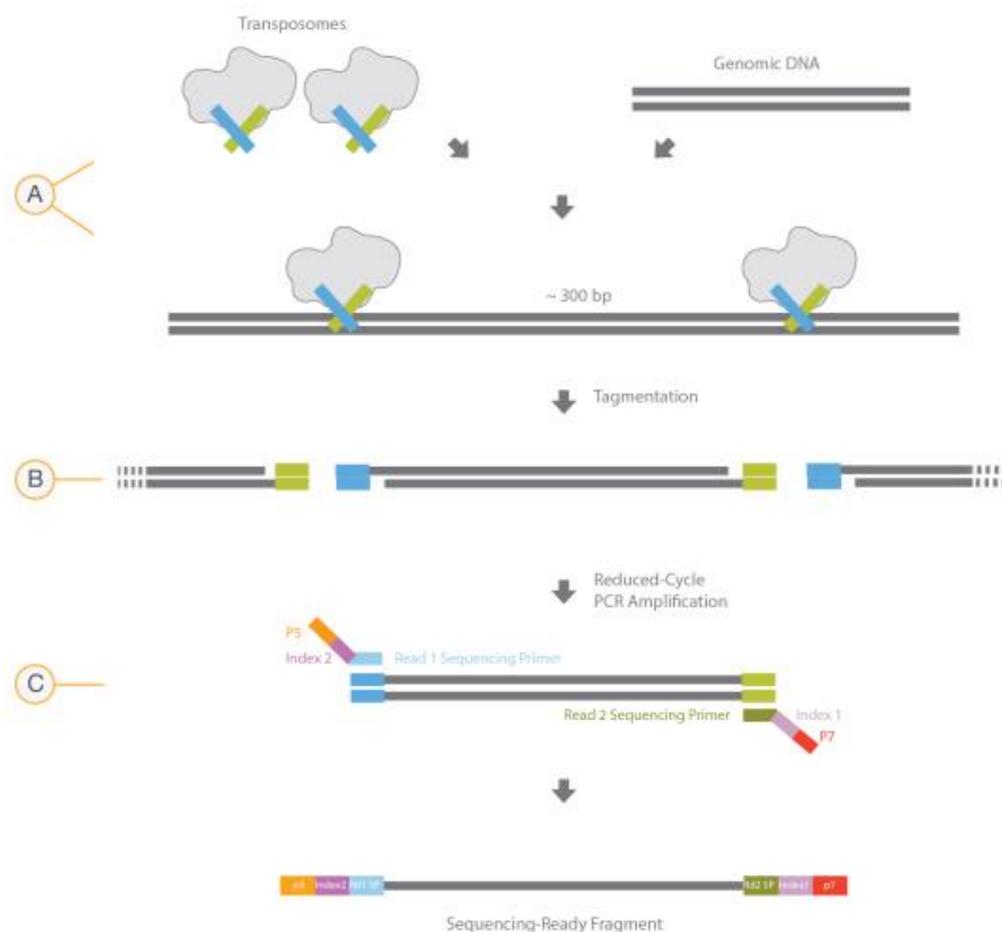


Figura 2.10 Preparazione delle library con Nextera® XT.
A: trasposoma con gli adattatori combinati con il campione di DNA.
B: tagmentation per effettuare la frammentazione e aggiungere gli adattatori.
C: ciclo di PCR con aggiunta dei primer
 (Nextera XT DNA Library Prep Kit Reference Guide)

Il work flow della preparazione della library comprende la *tagmentation* del DNA, l’amplificazione delle library, la pulizia, la normalizzazione e unione delle library. (Figura 1.26). Dopo la *tagmentation* del DNA con Amplicon Tagment Mix (ATM) e TD (Tagment DNA Buffer), il campione di materiale genomico viene amplificato utilizzando un programma di PCR a cicli limitati in cui vengono inserite

le sequenze degli adattatori già dal primo step così come i primer (Index 1 e 2) per l'amplificazione in PCR (Nextera XT DNA Library Prep Kit Reference Guide). Gli amplificati ottenuti dalla PCR vengono purificati tramite biglie AMPure XP ovvero particelle magnetiche che effettuano un'immobilizzazione reversibile in fase solida del DNA in presenza di polietilene glicolica (PEG) e sale (McElhoe J. A. *et al.*, 2014).

Successivamente viene effettuato il sequenziamento delle library che solitamente sono composte da 250 – 1000 bp con tutti i sequenziatori Illumina. La concentrazione delle library successivamente vengono normalizzate in modo da avere di ognuna una concentrazione equimolare. Successivamente vengono unificate tramite il pooling. Il pooling delle library consiste nell'unificazione di volume uguale delle library normalizzate in un singolo tubo. La normalizzazione antecedente garantisce una rappresentazione più equa del pool di campioni. Prima del sequenziamento, il pool viene diluito e denaturato e successivamente caricato per il sequencing.



Figura 2.11 Workflow di Nextera XT
(Nextera XT DNA Library Prep Kit Reference Guide)

2.8.3 Sequenziamento

Quando il campione è pronto viene inserito all'interno dello strumento, il sequenziamento viene effettuato in modo differente a seconda del metodo utilizzato (Tabella 2.2).

Table 2.1 High-throughput sequencing platforms

Platform	Company	Sequencing library	Support	Feature generation	Sequencing reaction	Detection method
GS FLX	454 Life Sciences, Roche	Linear adapters	Picotiter plate	Emulsion PCR	Synthesis	Pyrosequencing
Genome analyzer	Solexa, Illumina	Linear adapters	Flow cell	Bridge PCR	Synthesis	Fluorophore labeled reversible terminator nucleotides
SOLiD	Applied Biosystems	Linear adapters	Flow cell	Emulsion PCR	Ligation	Fluorophore labeled oligonucleotide probes
CGA platform	Complete genomics	Circular adapters	DNA nanoball arrays	Rolling circle amplification	Ligation	Fluorophore labeled oligonucleotide probes
PacBio RS	Pacific biosciences	Bubble adapters	Zero mode waveguide	Single molecule	Synthesis, real-time	Phospholinked fluorophore labeled nucleotides

Tabella 2.2 Piattaforme di sequenziamento High-throughput
(Van Dijk E. L. et al., 2014)

Le piattaforme di sequenziamento high-throughput integrano nei loro sistemi tecnologie fluide o ottiche per effettuare e monitorare la reazione molecolare di sequenziamento. I sistemi fluidi sono in grado di generare la reazione di sequenziamento in parallelo e di immobilizzare e sequenziare il DNA. Questo viene fatto tramite meccanismi automatici che erogano i reagenti sul campione immobilizzato di acido nucleico, tra cui i dNTP. I sistemi in parallelo utilizzano sistemi ottici per estrapolare informazioni sugli eventi molecolari che definiscono la reazione di sequenziamento. Ogni ciclo di sequenziamento consiste nella sintesi dell'acido nucleico complementare al campione immobilizzato: questo acido nucleico complementare viene ricostruito tramite l'inserimento di ogni nucleotide per volta. Ogni base del campione immobilizzato si legherà solo con la sua base complementare e questo per tutta la sequenza nucleotidica, andando così a ricostruire tutto il filamento complementare al campione (Myllykangas S. et al., 2012). Il sequenziamento *by synthesis* viene effettuato aggiungendo DNA polimerasi e i dNTP fluorescenti in ordine uno dopo l'altro: la non incorporazione determina l'immediata degradazione del dNTP, viceversa l'incorporazione nel filamento nascente innesca una reazione che emette chemiluminescenza la quale viene rilevata da un CCD. Le letture sono tipicamente di 200 – 400 bp, ma molti campioni possono essere trattati in parallelo all'interno dei sequenziatori NGS. Il sistema pyrosequencing del sequenziamento GS FLX si basa sulla registrazione dell'emissione di luce nel momento in cui i nucleotidi vengono incorporati (Figura 2.12a). Nel Genome Analyzer vengono utilizzati nucleotidi reversibili: ogni nucleotide ha una sorta di etichetta fluorescente e una terminazione che impedisce l'aggiunta di altri nucleotidi sul filamento sintetizzato. Tutti e quattro i nucleotidi sono analizzati in parallelo (Figura 2.12b). Nel sistema PacBio RS vengono utilizzati nucleotidi con fluorofori rimovibili: ogni nucleotide ha un fluoroforo specifico che si

inserirsi nella fase di incorporazione. La reazione di polimerizzazione del DNA in parallelo viene monitorata tramite l'emissione di una luce pulsata per ogni nucleotide inserito. Il sequenziamento *by ligation* consiste invece nell'incorporazione di oligonucleotidi fluorescenti (Figura 2.12d) e si applica nel sistema SOLiD. Gli oligonucleotidi sono marcati da un fluoroforo e corrispondono ad una base specifica nel filamento (Myllykangas S. et al. 2012).

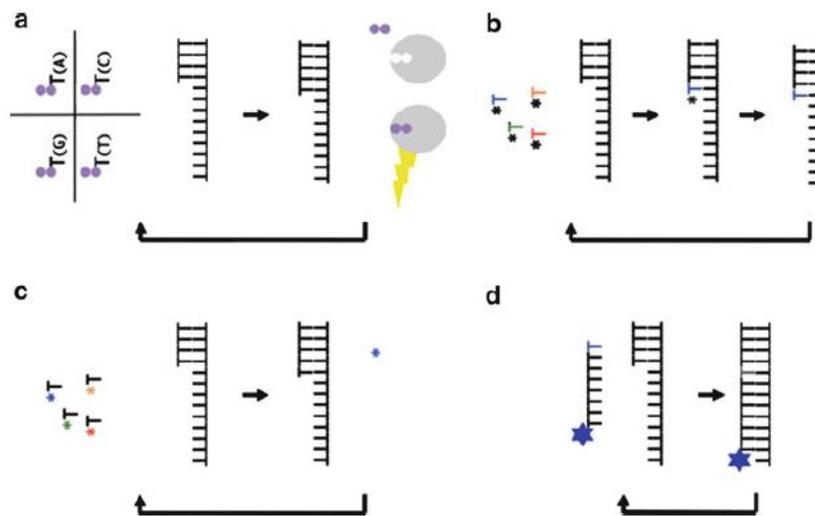


Figura 2.12 Reazione ciclica di sequenziamento nelle varie piattaforme (Myllykangas S. et al., 2012)

Le varie tecnologie di sequenziamento, seguendo strade diverse, forniscono come output delle sequenze di lunghezza variabile da alcune decine fino a diverse centinaia di basi. Queste vengono combinate insieme, di solito in un file FASTQ, un tipo di file utilizzato in biologia per memorizzare sequenze genetiche e i relativi quality scores, ossia il punteggio che l'algoritmo che effettua l'analisi assegna alla stringa, e che viene poi utilizzato per scegliere il match migliore contro il genoma di riferimento (Myllykangas S. et al., 2012).

2.8.4 Analisi dei dati

L'analisi bioinformatica si divide in tre fasi: l'allineamento per ricercare corrispondenze tra le reads e un genoma di riferimento, variant calling, per separare differenze dovute ad errori genetici dagli errori strumentali compiuti nell'analisi, filtraggio e annotazione (Dolled-Filhart M. P. et al., 2013), che tentano di allineare le reads al genoma di riferimento. L'allineamento è il processo attraverso cui si mappano reads corte ad un genoma di riferimento. Si tratta di un compito complesso, in quanto il software deve confrontare ogni reads con ogni posizione del DNA di riferimento. Si tratta di un passaggio computazionalmente impegnativo, e dispendioso in termini temporali (Myllykangas S. et

al., 2012). Per ottenere un'adeguata individuazione tassonomica i fattori chiave sono la lunghezza delle reads, la loro qualità e l'abundance. Nella targeted metagenomic, le reads sono raggruppate in unità tassonomiche operative (OTU) (Figura 2.13), tramite due metodi che portano a risultati simili (Sul W. J. *et al.*, 2011). Il primo metodo si basa sulla similarità delle basi delle reads con quelle presenti nelle banche dati (*similarità di basi*). Il secondo metodo effettua l'assegnazione tassonomica, classificando le reads a vari livelli tassonomici (phylum, classe, ordine, famiglia, genere) in base alla *similarità delle frequenze* in riferimento alle frequenze presenti nelle banche dati. L'assegnazione tassonomica ha il vantaggio di fornire informazioni sulla relazione delle reads identificate con gruppi di microrganismi conosciuti e permette la comparazione tra studi differenti anche quando le sequenze derivano da regioni differenti del gene target.

Il ribosomiale (RDP) (Cole J. R. *et al.*, 2008), Silva (Quast C. *et al.*, 2012) e Greengenes projects (DeSantis T. Z. *et al.*, 2006) forniscono banche dati aggiornate per sequenze di rDNA di piccole dimensioni (16S e 18S) e di grandi dimensioni (23S e 28S). Se non si riesce a identificare la specie genica di appartenenza, si considera che un'omologia di sequenza del 95% con quella di riferimento identifica lo stesso genere, mentre un'omologia del 97% identifica la stessa specie. Queste percentuali potrebbero non essere raggiunte nel caso in cui vi siano reads troppo corte, per questo motivo, riuscire ad ottenere reads sufficientemente lunghe è fondamentale per l'analisi dati (Mayo B. *et al.*, 2014).

Quando si lavora con altri marcatori filogenetici (*rpoB*, *gyrG*, *recA*) o geni funzionali (codificanti per antibiotico-resistenza, attività di biodegradazione), vengono utilizzati altri database come FunGene (<http://fungene.cme.msu.edu>). Infine possono essere utilizzate risorse web quali RAST (Aziz R. K. *et al.*, 2008) e MG-RAST (Meyer F. *et al.*, 2008), che offrono database progettati per aiutare i dati genomici e metagenomici, rispettivamente. Tutti questi database hanno pipeline user-friendly in grado di essere utilizzate anche da utenti poco esperti, tuttavia non consentono di modificare e personalizzare i dati. Il problema è che molti microrganismi non hanno una classificazione tassonomica e non vi sono genomi di riferimento nei database per cui questi organismi devono essere catalogati *ex novo* (Rachid C. T. *et al.*, 2013). Quando si incontra questo tipo di problematica, un metodo molto comune che si utilizza per analizzare le informazioni ottenute dal sequenziamento si basa sulla somiglianza genetica tra le reads. In questo modo si costruiscono cluster infatti basati sulla dissimilarità delle sequenze (normalmente 1-3% per i batteri) (Rachid C. T. *et al.*, 2013). Tuttavia questo metodo non fornisce informazioni su chi è quel preciso microrganismo, limitando l'interpretazione dei dati.

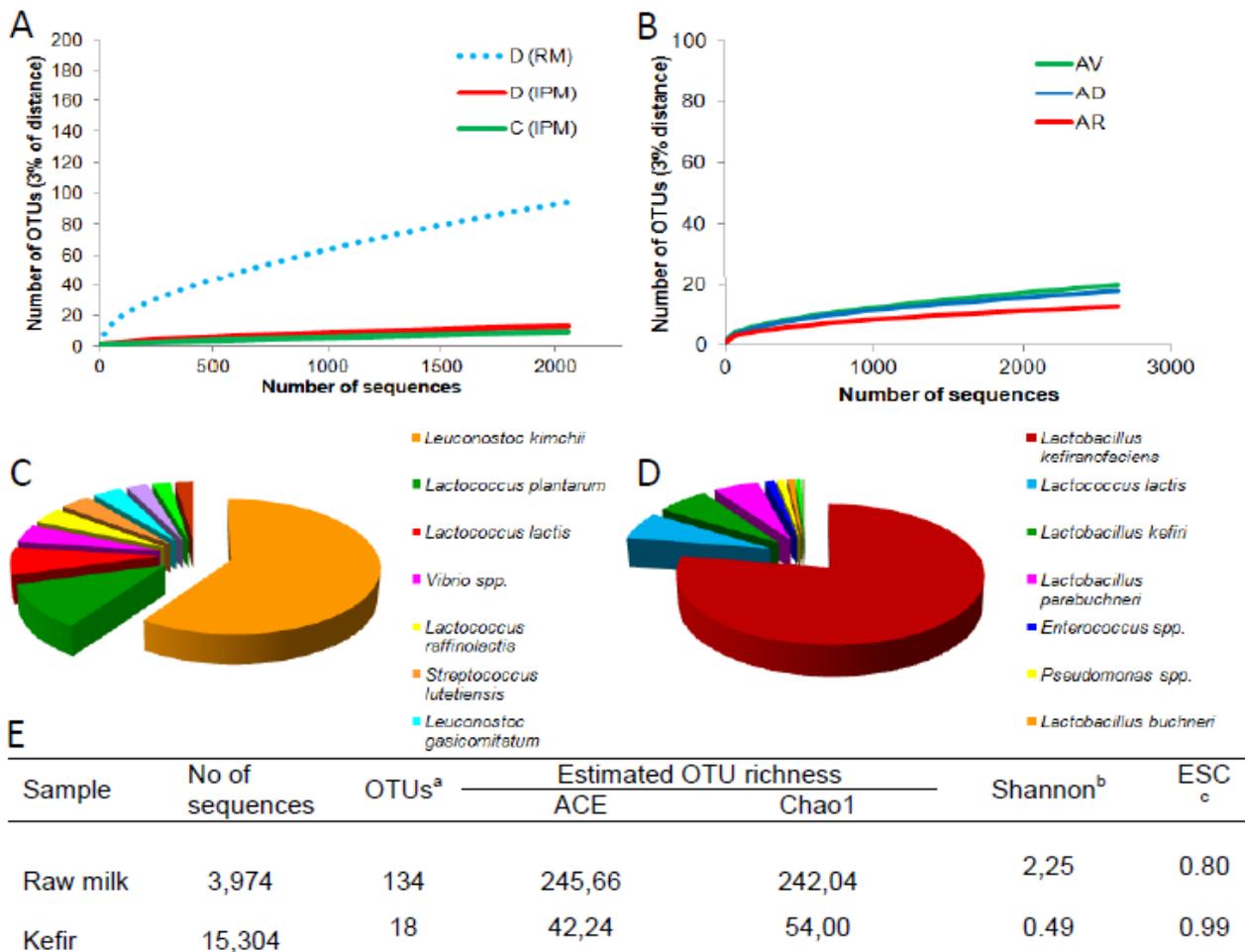


Figura 2.13 Esempio di curva di rarefazione di latte crudo (linea blu) e latte pastorizzato, incubato a 42°C per 24h (linee verde e rossa) (A) e di kefir (B). Rappresentazione schematica del microbiota identificato da latte crudo (C) e kefir (D). Tabella in cui viene stimata la quantità di OTU e gli indici di qualità (E) (Leite A. M. O. et al., 2012; Delgado S. et al., 2013)

Alcuni degli strumenti bioinformatici che integrano l'approccio tassonomico e tassonomico-indipendente offrono strategie per personalizzare le analisi e fornire dati di alta qualità che possono essere presentati con tabelle e grafici (Figura 2.13). I software più popolari sono Mothur (Schloss P. D. et al., 2009) e QIMME (Caporaso J. G. et al., 2010), ad accesso libero, con manuali ed esempio di data analysis. Questi programmi vengono periodicamente aggiornati, migliorando le opzioni analitiche.

Un altro metodo promettente per l'elaborazione dei dati NGS collega la filogenesi e la funzionalità molecolare ed è il software PICRUSt che si basa su un vasto numero di dataset genomici e metagenomici (<https://img.jgi.doe.gov/cgi-bin/w/main.cgi>). PICRUSt è in grado di estrapolare il numero di geni condivisi da microrganismi appartenenti a diversi taxa (Mayo B. et al., 2014). Questo programma fornisce informazioni sulle relazioni fenotipiche all'interno di una comunità microbica, basandosi esclusivamente sulla sua composizione filogenetica (Langille M. G. et al., 2013).

3 SCOPO DELLA RICERCA

Nelle ultime decadi le tecniche per la determinazione dei batteri patogeni negli alimenti si sono modificate radicalmente (Valderrama W.B. *et al.*, 2016; Nugen S. R. *et al.*, 2008; Gracias K. S. *et al.*, 2004). Le tecniche tradizionali, basate sulla crescita dei patogeni in terreni colturali selettivi, hanno prodotto risultati interessanti che hanno permesso di chiarire molti aspetti relativi alla sicurezza degli alimenti. Tuttavia, il limite di tali tecniche deriva dal fatto che possono essere applicate solo a gruppi di batteri patogeni e degradativi coltivabili. Inoltre, tali microrganismi vengono studiati *in vitro* e non all'interno dell'ecosistema alimento.

Approcci molecolari più moderni, come saggi immunoenzimatici e reazione a catena della polimerasi, risultano più specifici ma permettono di determinare una o poche specie per volta. Tuttavia, tali specie nell'alimento possono cambiare in funzione di molteplici stress di varia natura (temperatura, atmosfera, presenza di altri gruppi microbici) che causano un dinamico ri-arrangiamento delle popolazioni, sia di batteri patogeni che di batteri degradativi (Larsen M. H. *et al.*, 2014; Pricope L. *et al.*, 2013). Per questo motivo è sicuramente più interessante studiare, almeno in fase esplorativa, non singoli gruppi microbici ma intere popolazioni contaminati gli alimenti. Mediante analisi metagenomica shotgun sul DNA totale estratto da una matrice alimentare è possibile studiare una popolazione microbica nel suo complesso e nel momento dell'interazione tra i molti gruppi microbici presenti.

La metagenomica shotgun comincia a trovare oggi molteplici applicazioni perché i costi sono diminuiti, sono disponibili protocolli bioinformatici per la piena valorizzazione dei risultati raccolti ed esistono siti web nei quali è possibile stoccare i metagenomi raccolti (Manichanh C. *et al.*, 2008). In confronto al sequenziamento di ampliconi, soprattutto rappresentato dal sequenziamento dei geni codificanti il 16S rRNA, la metagenomica shotgun permette un'identificazione dei gruppi tassonomici che non risente degli errori che potrebbero derivare dallo sbilanciamento dell'amplificazione verso un gruppo rispetto ad un altro (Shah N. *et al.*, 2011). Inoltre, la metagenomica shotgun oltre all'identificazione di tutti i gruppi microbici (non solo batteri ma anche eucarioti e virus) permette di identificare in un campione la composizione e l'abundance dei geni. Tale composizione può permettere di predire un insieme di proteine ed enzimi che si possono poi associare nelle potenziali vie metaboliche che caratterizzano un campione (Li J. *et al.*, 2014). In generale gli obiettivi di un'analisi metagenomica shotgun sono i seguenti:

- Identificare le popolazioni microbiche presenti e la loro abundance, dal livello di phylum al livello di specie;

- Identificare il contributo in geni fornito da ogni gruppo microbico presente nel sistema oggetto dello studio;
- Quantificare la variabilità tra i geni intra-specie ed intra-popolazione (Scholz M. B. *et al.*, 2012).

Una delle più comuni applicazioni della metagenomica shotgun nel settore delle produzioni di origine animale, è legata alla conoscenza degli effetti di enzimi, miscele probiotiche ed additivi integrati nella dieta sul microbioma intestinale degli animali. In quest'ottica anche la modulazione delle secrezioni intestinali e pancreatiche di fitasi nei polli allevati rappresenta un approccio fondamentale per un sistema di allevamento ecosostenibile. Per ottenere questo risultato è possibile aggiungere alle diete fitasi che tamponano l'effetto anti-nutrizionale dell'acido fitico, aumentando la digeribilità delle proteine e modulando il microbiota intestinale. In questo modo si può evitare di integrare il mangime con fosforo inorganico (Selle P. H. *et al.*, 2007) che determinerebbe un aumento dei costi di produzione e un incremento del rilascio di fosforo nell'ambiente (Kiarie E. *et al.*, 2013).

Nella mia ricerca ho applicato la metagenomica shotgun per valutare le popolazioni microbiche ed i geni presenti su carcasse di polli da carne raccolte al macello al termine del tunnel di raffreddamento mettendo a confronto due metodologie di analisi dei dati. Le carcasse analizzate sono state ottenute da animali alimentati con una dieta di controllo, una dieta addizionata con 1500 unità di fitasi per kg di mangime ed una dieta addizionata con la stessa concentrazione di fitasi e 3 g/kg di inositolo. La scelta di analizzare carcasse ottenute da animali alimentati con diete diverse è stata fatta al fine di aumentare la probabilità di osservare differenze significative tra le carcasse, anche se tale eventuale differenza non può essere attribuita soltanto alla dieta degli animali ma anche alle contaminazioni crociate delle carcasse durante la macellazione, fino al termine del tunnel di raffreddamento. Inoltre sono stati messi a confronto due diversi tools bioinformatici ossia MG-RAST e MGmapper al fine di valutare la loro capacità di identificazione delle specie microbiche in un ecosistema microbico complesso.

4 MATERIALI E METODI

4.1 Animali dai quali sono state ottenute le carcasse da analizzare

Le carcasse analizzate sono state raccolte al macello e sono state ottenute da animali che hanno seguito un normale ciclo di allevamento di 35 giorni. Le carcasse sono state ottenute da polli maschi Ross 308, tutti appartenenti al medesimo gruppo di origine ma alimentati con diete diverse. Infatti, le carcasse del gruppo A sono state ottenute da animali alimentati con la dieta base, descritta in tabella 4.1. Le carcasse del gruppo B sono state ottenute da animali alimentati con la dieta base addizionata con 1500 unità di fitasi per kg mangime. Infine, le carcasse del gruppo C sono state ottenute da animali alimentati con la dieta base addizionata con 1500 unità di fitasi per kg di mangime e 3 g di inositolo per kg di mangime. Il programma di alimentazione in allevamento ha previsto una fase starter fino al giorno 14, una fase di crescita dal giorno 15 al giorno 28 ed una fase di finissaggio dal giorno 29 al giorno 35. Le diverse formulazioni delle diete, come pure l'acqua, sono sempre state somministrate agli animali *ad libitum*. I periodi di luce e buio e le temperature di allevamento sono state decise in accordo con il Regolamento EU 43/2007 sul benessere animale (European Union, 2007).

Tabella 4.1 Dieta base somministrata agli animali dai quali sono state ottenute le carcasse analizzate.

	Mangime starter (0-14 g)	Mangime Grower (15-28 g)	Mangime Finisher (29-35 g)
Mais	34.96	12.73	
Mais bianco	0.00	0.00	15.00
Grano	10.00	20.00	25.01
Sorgo	0.00	0.00	5.00
Farina di soia	23.11	20.63	17.60
Expanded soybean	10.00	10.00	13.00
Semi di girasole	3.00	3.00	3.00
Farina di glutine di mais	4.00	3.00	0.00
Olio di semi di soia	3.08	4.43	5.48

Fosfato bicalcico	1.52	1.20	0.57
Calcium carbonate	0.91	0.65	0.52
Carbonato di calcio	0.15	0.10	0.15
Sale	0.27	0.27	0.25
Colina cloridrato	0.10	0.10	0.10
Solfato di Lisina	0.59	0.55	0.46
DL-metionina	0.27	0.29	0.30
Treonina	0.15	0.14	0.14
Xylanase	0.08	0.08	0.08
Fitasi	0.10	0.10	0.10
Vitamin-mineral premix	0.50	0.50	0.50
Sostanza secca	88.57	88.65	88.64
Proteine	22.70	21.49	19.74
Lipidi	7.06	8.24	9.74
Fibre	3.08	3.04	3.07
Cenere	5.85	5.17	4.49
ME (kcal/kg)	3,076	3,168	3,264

Le fitasi sono enzimi di derivazione animale o vegetale, si possono utilizzare nei mangimi al fine di idrolizzare l'acido fitico a monofosfato inorganico ed esteri fosforici degradare (Dvoráková J., 1998). È stato dimostrato che l'attività fitasica aumenta con il diminuire del numero dei gruppi fosfato. Ciò suggerisce che l'idrolisi dell'acido fitico a livello intestinale possa essere favorita dalle fitasi presenti nello stomaco (Sandberg A. S. *et al.*, 2002).

L'aggiunta di fitasi potrebbe essere un'alternativa valida per ottimizzare la defosforilazione del fitato, infatti, fino ad ora le fitasi sono state principalmente usate come additivo per l'alimentazione animale nelle diete di suini e pollame. Numerosi studi hanno dimostrato l'efficacia di fitasi microbiche

aggiunte nelle diete degli animali (Stahl C. A. *et al.*, 2000; Rodehutschord M. *et al.*, 1995). Infatti si è visto che migliorano la biodisponibilità di rame e zinco nel pollame (Yi Z. *et al.*, 1996). Oltretutto incrementano la concentrazione di minerali come magnesio, fosforo, calcio, manganese e zinco nel plasma e nelle ossa (Vielma J. *et al.*, 1998). Quindi includendo adeguate quantità di fitasi nelle diete di animali, si riduce la necessità di aggiungere ortofosfato nell'alimentazione.

È stato dimostrato che l'acido fitico complessa gli ioni metallici multivalenti che vanno ad interferire sull'assimilazione di importanti minerali come ferro, zinco, calcio, magnesio, rame, con carenze nella dieta sia umana che animale. Inoltre, il fitato inibisce enzimi digestivi, quali α -amilasi, pepsina e tripsina, e rende le proteine più resistenti alla digestione proteolitica.

Il fitato non degradato e parti di fosfato inorganico non sono digeribili dai monogastrici compreso l'uomo (Simell M. *et al.*, 1989), quindi dopo essere passati attraverso il tratto intestinale sono escreti all'esterno, causando seri problemi ambientali come l'eccesso di fosfati nelle acque, che man mano portano all'eutrofizzazione (sovrabbondanza di nitrati e fosfati), specialmente in aree in cui c'è una produzione intensiva di bestiame.

In conclusione, l'utilizzo di fitasi migliora la digeribilità ed anche il valore nutritivo dei mangimi e contribuisce ad abbassare il livello di fosforo escretato nell'ambiente (Dvoráková J., 1998) fino al 50% (Zhang Z. B. *et al.*, 2000).

4.2 Raccolta ed analisi delle carcasse

Le 15 carcasse da analizzare sono state raccolte al macello al termine del tunnel di raffreddamento ed appoggiate su un tavolo di acciaio inox accuratamente lavato e disinfettato prima dei prelievi e tra i prelievi delle carcasse appartenenti ai diversi gruppi.

Da ciascuna carcassa sono stati prelevati 10 g circa di pelle di collo in accordo con le norme di campionamento per le carcasse di pollame riportate nel Regolamento (CE) 2073/2005 sui criteri microbiologici applicabili ai prodotti alimentari. Ogni campione di pelle di collo è stato inserito in un sacchetto da stomacher sterile e trasportato in laboratorio tra 2 e 4°C.

4.2.1 Estrazione del DNA dalla pelle di collo ottenuta dalle carcasse

All'arrivo in laboratorio ogni campione è stato diluito 1:5 (peso/volume) con soluzione fisiologica sterile (0.85% NaCl in acqua sterile) ed omogenato nello stomacher Mayo homogenius a velocità normale per 1 minuto. Al termine della fase di omogeneizzazione, tutta la parte liquida del campione è stata trasferita in una provetta sterile da 50 ml e centrifugata a 6800 rpm per 20 minuti a 4°C. Il pellet ottenuto al termine della centrifugazione è stato poi risospeso in 5 ml di soluzione fisiologica

fredda e centrifugato nuovamente a 6800 rpm per 20 minuti a 4°C. Al termine della centrifugazione 250 mg del pellet ottenuto sono stati sottoposti ad estrazione DNeasy Blood & Tissue Kit (Qiagen, Milano, Italia) che prevede la separazione del DNA dal campione mediante suo legame ad una colonnina di silice. Ad ogni campione di pellet sono stati aggiunti 200 µl di una soluzione al 10% di proteinasi k prima di un periodo di incubazione a 56°C per 12 ore. Al termine del periodo di incubazione ad ogni campione sono stati aggiunti 410 µl di una soluzione contenente etanolo prima di una omogenizzazione per 15 secondi. Al termine della omogenizzazione un volume massimo di 900 µl di lisato a campione è stato centrifugato per 10 minuti a 6000 rpm (3800 x g). Al pellet così ottenuto sono stati aggiunti 500 µl di buffer e si è centrifugato nuovamente a 6000 rpm per 5 minuti. Il surnatante ottenuto è stato aggiunto in una apposita colonnina, addizionato con 200 µl di buffer, incubato a T ambiente per 1 min e raccolto mediante centrifugazione a 6000 rpm per 2 minuti. Al termine della fase di estrazione il DNA estratto da ogni campione di pelle di collo è stato quantificato mediante BioSpectrometer® (Eppendorf, Milano, Italiay) e Fragment Analyzer (Advanced Analytical).

4.3 Preparazione delle library per il sequenziamento

Le library per il sequenziamento sono state preparate con Kit NexteraXT Illumina. La costruzione delle library è necessaria per frammentare il DNA estratto ed aggiungere ad ogni frammento sequenze specifiche, chiamate adattatori. Le library sono state preparate in modalità “paired-end” al fine di ottenere una sequenza da ciascuna delle due estremità di ogni frammento.

La quantità di DNA necessaria per la preparazione delle library è 1 ng (5 µl alla concentrazione di 0,2 ng/ µl) il livello di purezza necessario corrisponde ad un rapporto dei valori di assorbanza a 260 e 280 nm compreso tra 1.8 e 2.0. La prima fase della preparazione delle library prevede la simultanea frammentazione del DNA genomico estratto e l’aggiunta dei primers di sequenziamento Illumina.

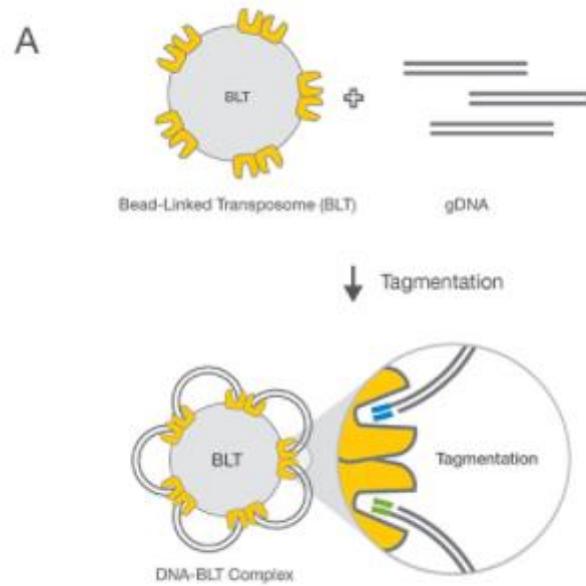


Figura 4.1 Tagmentazione dei frammenti
<https://www.illumina.com/>

La fase successiva è rappresentata dalla PCR attraverso la quale si aggiungono ai frammenti da sequenziare indici ed adattatori.



Figura 4.2 PCR per legare adattatori e indici
<https://www.illumina.com/>

La terza e ultima fase prevede la separazione delle sequenze pronte per il sequenziamento da tutti i reagenti PCR ed il pooling delle sequenze.



Figura 4.3 Purificazione dei frammenti delle library pronti per il sequenziamento
<https://www.illumina.com/>

4.4 Sequenziamento

Il sequenziamento è stato effettuato per sintesi mediante HiScan™ SQ System (Illumina) con TruSeq SBS Kit v3-HS che permette di ottenere un output di 600 Gb. Il sequenziamento è stato effettuato in modalità paired-end a 100 bp.



Figura 4.4 HiScan™ SQ System

https://support.illumina.com/sequencing/sequencing_instruments/hiscansq.html

4.5 Analisi delle sequenze

I sequenziatori di ultima generazione high throughput generano centinaia di milioni di sequenze in una singola analisi. Prima di analizzare le sequenze, è necessario eseguire un loro controllo qualità. I metagenomi ottenuti in questa ricerca sono stati sottoposti ad analisi qualitativa mediante il software FastQC versione “fastqc_v0.11.5”. FastQC effettua una serie di analisi grazie al quale è possibile avere una rapida impressione per comprendere la qualità delle reads ottenute durante il sequenziamento.

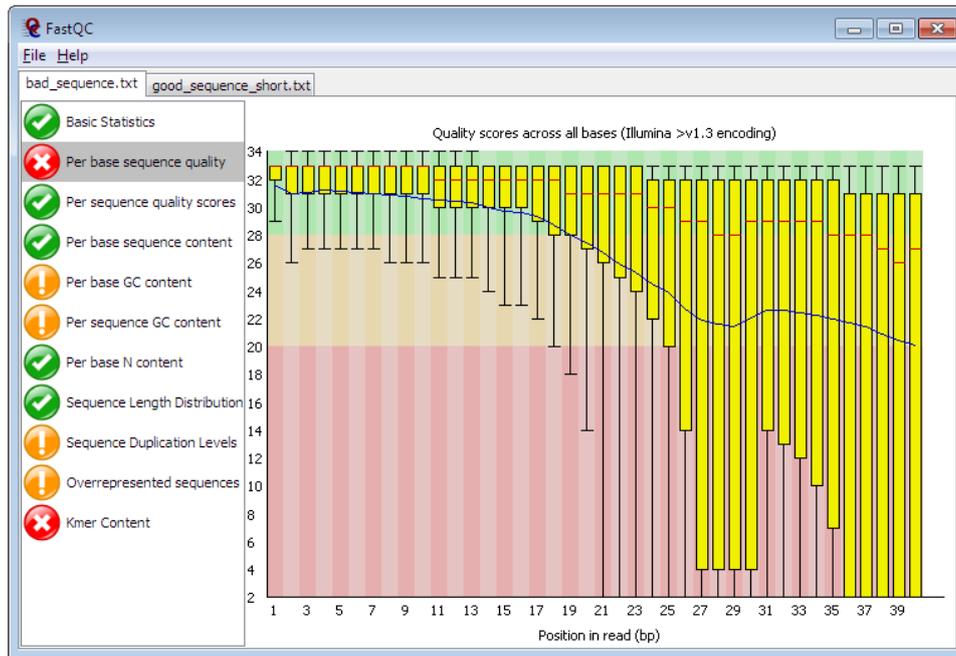


Figura 4.5 Esempio di schermata del software bioinformatico *FastQC*

I file fastq descrivono ogni sequenza con 4 righe. La prima riga mostra l'identificativo della sequenza. La seconda riga contiene la sequenza di nucleotidi. La terza riga generalmente contiene solo un simbolo / "+". La quarta riga mostra dei simboli corrispondenti al punteggio di qualità assegnato ad ogni nucleotide mostrato sulla seconda riga (Figura 4.6).

```
@SEQ_ID
GATTTGGGGTTCAAAGCAGTATCGATCAAATAGTAAATCCATTTGTTCAACTCACAGTTT
+
!' '*((( (***) )%%%++) (%%%) .1***-+*' ')) **55CCF>>>>>CCCCCCC65
```

Figura 4.6 Esempio di una reads elaborata con *FASTQC*

I punteggi di qualità sulla quarta riga rappresentano la probabilità di un errore di sequenza (p) in ciascuna posizione nucleotidica. Conoscendo la probabilità di tale errore, è possibile calcolare la qualità complessiva della sequenza (Q) attraverso la seguente formula:

$$Q = -10 \log_{10} p$$

Ad esempio, se la probabilità di un errore (p) è uguale a 0,01, il punteggio di qualità (Q) corrispondente sarà 20; se $p = 0,001$, $Q = 30$.

Tuttavia, come possiamo vedere nella quarta riga dell'esempio, il file fastq non contiene numeri a due cifre ma solo alcuni caratteri dispari. Questi sono caratteri ASCII speciali che vengono utilizzati per

codificare i valori di qualità con un singolo simbolo, piuttosto che con una doppia o tripla cifra. Tale codifica a singolo carattere è necessaria per avere una corrispondenza “uno-a-uno” tra ciascun nucleotide sulla riga 2 e ogni punteggio sulla linea 4. Ciò è possibile solo con l'uso di un singolo carattere su entrambe le linee.

I primi 32 elementi nella tabella ASCII sono caratteri non stampabili. Pertanto, per assegnare un carattere stampabile ad un valore di qualità dovremmo iniziare dal 33° elemento. Questo tipo di codifica di qualità si chiama *Phred33* ed è il modo più comune di valutare la qualità delle sequenze. Il *Phred33* viene utilizzato nei file di output delle piattaforme di sequenziamento Illumina. FastQC analizza i file in formato fastq come input e restituisce un report HTML contenente informazioni ben strutturate e illustrate graficamente su diversi aspetti qualitativi delle reads. Il rapporto che genera il software è diviso in diverse sezioni. Nella prima sezione si trovano informazioni generali sul numero e sulla lunghezza delle reads, chiamate "Statistiche di base". L'elenco degli adattatori potenzialmente non eliminati è nella sezione "sequenze sovra-rappresentate". In questa pagina sono disponibili spiegazioni dettagliate relative al significato e ai potenziali problemi di ciascuna sezione del rapporto FASTQC.

È possibile accedere ai comandi sopra descritti, selezionandoli dalla barra operativa posizionata a sinistra (Figura 4.5). Ogni elemento nel sommario è contrassegnato da un segno di spunta verde, un punto esclamativo arancione o una croce rossa. Un segno di spunta verde indica che questo aspetto dei dati è di qualità soddisfacente, un punto esclamativo arancione indica che qualcosa è leggermente anomalo e una croce rossa indica che il parametro corrispondente dei dati di origine è molto insolito e necessita di correzione. Si deduce che le croci rosse sono di grande interesse al fine di migliorare la qualità dei dati di sequenziamento.

Ad esempio se si considera la prima voce con croce rossa, “Per base sequence quality” (Figura 4.5), significa che la qualità delle basi è molto bassa. In effetti la figura mostra che le basi dalla posizione 20 a 25 hanno un valore di qualità (*Phred33*) compreso tra 28 e 20 mentre le basi dalla posizione 26 in poi hanno un valore di qualità troppo basso. Quando le basi a bassa qualità si trovano alla fine delle reads possono essere eliminate ma devono essere in numero basso (es. 10-15 basi).

Per l'analisi bioinformatica i metagenomi ottenuti in questa ricerca sono stati caricati sui seguenti portali:

- MG-RAST <http://metagenomics.anl.gov/>
- MGmapper <https://cge.cbs.dtu.dk/services/MGmapper/>

4.5.1 MG-RAST

In MG-RAST i filtri applicati nella mia analisi sono stati i seguenti:

- Rimozione delle reads relative a DNA eucariote
- Rimozione delle reads di bassa qualità (Phred ≤ 15)
- Best Hit

Come database per la classificazione tassonomica ho utilizzato il database ribosomiale SSU. Come output dell'analisi dei metagenomi MG-RAST fornisce una serie di grafici, tabelle, file csv ed excel che possono essere esportati in software per l'analisi biostatistica. MG-RAST è un sistema molto versatile e permette di esportare i fastq anche durante i passaggi intermedi dell'analisi.

4.5.2 MGmapper

MGmapper è un portale web molto simile al precedente, ma più selettivo, con un Phred di default ≤ 30 . L'accesso a MG mapper è attualmente disponibile attraverso l'invio di dati all'Università Tecnica della Danimarca. Per questo motivo i setting dell'analisi non sono modificabili. Anche in questo portale si sono elaborate le informazioni "Best Hit" ed i dati in output sono stati esportati in formato csv per la successiva elaborazione.

4.5.3 Analisi statistica - STAMP

I risultati relativi alle abundances dei gruppi tassonomici e dei geni funzionali sono stati analizzati attraverso il t-test non parametrico di White utilizzando il software Statistical Analysis of Metagenomic profile Software v 2.0.9 (STAMP) (Parks D.H. *et al.*, 2014). Il primo passaggio è stato quello di convertire il formato cvs nel formato txt (MS-DOS). Successivamente, è stato creato il file metadata necessario per associare le reads ottenute ai campioni di appartenenza (Controllo, Fitasi e Fitasi + Inositolo).

È stata effettuata l'elaborazione statistica tramite test "ANOVA" e Post-hoc test "Tukey-Kramer", impostando un valore di significatività del dato p-value < 0.05 . È stato chiesto al software di eliminare le reads non classificate.

Il software crea una tabella con la media della abundance di reads e il di p-value per ogni profilo richiesto. Questa tabella viene esportata per una successiva elaborazione su foglio di calcolo Excel. L'operazione è stata ripetuta per tutti i dati ottenuti con entrambi i portali web (MG-RAST e MGmapper).

5 RISULTATI

5.1 DNA estratti

I DNA estratti dai campioni di pelle di collo hanno mostrato una concentrazione media di 33,47 ng/μl, con valori massimi e minimi pari a 80 e 11 ng/μl in due campioni del gruppo B, rappresentati dal 368 e 367, rispettivamente (Tabella 5.1). Tutti i campioni di DNA estratti hanno mostrato quantità sufficiente per la preparazione di library da sottoporre a sequenziamento, corrispondente a 1 ng. Inoltre, i campioni di DNA hanno mostrato tutti un livello di purezza sufficientemente alto, corrispondente ad un rapporto tra valore di assorbanza a 260 e 280 nm sempre compreso tra 1,8 e 2 (Tabella 5.1).

Tabella 5.1 Concentrazione e qualità del DNA estratto da ogni campione

Origine della carcassa	ID campione	Concentrazione DNA estratto (ng/μl)	Rapporto 260/280 nm
Gruppo Controllo	357	55,80	1,89
Gruppo Controllo	358	70,40	1,97
Gruppo Controllo	359	20,40	1,75
Gruppo Controllo	360	35,90	1,95
Gruppo Controllo	361	19,90	1,99
Gruppo Fitasi	367	11,0	1,73
Gruppo Fitasi	368	80,0	1,99
Gruppo Fitasi	369	37,20	1,95
Gruppo Fitasi	370	32,80	1,83
Gruppo Fitasi	371	14,60	1,91
Gruppo Fitasi + Inositolo	377	36,90	1,89
Gruppo Fitasi + Inositolo	378	20,10	1,97
Gruppo Fitasi + Inositolo	379	17,80	1,90
Gruppo Fitasi + Inositolo	380	24	1,90
Gruppo Fitasi + Inositolo	381	25,30	1,90

5.2 Parametri delle library e dei metagenomi ottenuti

Le library associate ad ogni singolo campione estratto hanno mostrato frammenti di lunghezza media pari a 1340 bp (Tabella 5.2) mentre i metagenomi ottenuti da ogni singola library hanno mostrato un numero medio di reads pari a 6816563 (Tabella 5.2) ed un indice Phread sempre maggiore di 30. Tutti i metagenomi sono stati depositati nel sistema MG-RAST con le denominazioni (ID) indicate in tabella 5.2 e resi pubblici. I metagenomi sono stati analizzati sia con MG-RAST che con MGmapper.

Tabella 5.2 Parametri delle library e delle sequenze ottenute da ogni campione

Origine della carcassa	ID campione	ID Metagenoma	Lunghezza frammenti library (bp)	Concentrazione library per il sequenziamento (mM)	Reads ottenute dal sequenziamento
Gruppo Controllo	357	XT_357	1150	9,45	9445326
Gruppo Controllo	358	XT_358	1256	4,49	7636834
Gruppo Controllo	359	XT_359	1235	5,89	7728898
Gruppo Controllo	360	XT_360	1066	6,10	6850052
Gruppo Controllo	361	XT_361	1238	2,73	6664934
Gruppo Fitasi	367	XT_367	1465	4,15	5226314
Gruppo Fitasi	368	XT_368	1206	7,14	6363476
Gruppo Fitasi	369	XT_369	1313	5,72	6544524
Gruppo Fitasi	370	XT_370	1285	4,10	5160700
Gruppo Fitasi	371	XT_371	1315	8,65	5485326
Gruppo Fitasi + Inositolo	377	XT_377	1551	3,20	7974236
Gruppo Fitasi + Inositolo	378	XT_378	1501	2,35	5492496
Gruppo Fitasi + Inositolo	379	XT_379	1322	2,30	7335068
Gruppo Fitasi + Inositolo	380	XT_380	1444	3,02	7171638
Gruppo Fitasi + Inositolo	381	XT_381	1549	2,51	7168624

5.3 Tabelle con i risultati tassonomici

Tabella 5.3 Valori medi di abundance (%) ottenuti con MG-RAST relativi alle prime 30 classi identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi e inositolo (* = 0,05 < p < 0,01)

Classe	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
Gammaproteobacteria	0,184	90,931	2,436	88,637	1,545	87,785	2,814
Bacilli	0,598	3,044	1,331	3,571	1,050	3,869	1,011
Betaproteobacteria	0,305	1,637	0,048	1,712	0,527	1,342	0,262
Flavobacteria	0,168	1,424	0,344	1,044	0,138	2,155	1,306
Clostridia	0,191	1,023	0,874	2,846	0,858	2,479	2,085
Alphaproteobacteria	0,192	0,560	0,049	0,461	0,069	0,506	0,090
Actinobacteria	0,058	0,423	0,106	0,766	0,229	0,688	0,208
Deltaproteobacteria	0,241	0,171	0,013	0,143	0,022	0,161	0,030
Bacteroidia	0,242	0,158	0,024	0,336	0,224	0,263	0,095
Sphingobacteria	0,227	0,148	0,039	0,104	0,012	0,204	0,129
Cytophagia	0,198	0,113	0,023	0,079	0,009	0,150	0,086
Epsilonproteobacteria	0,100	0,070	0,010	0,053	0,012	0,060	0,008
Chlorobia	0,016	0,067*	0,007	0,050*	0,008	0,054*	0,007
Fusobacteria	0,273	0,038	0,016	0,043	0,012	0,088	0,077
Erysipelotrichi	0,657	0,026	0,017	0,019	0,004	0,026	0,011
Spirochaetes	0,299	0,019	0,004	0,016	0,002	0,021	0,006
Planctomycetacia	0,036	0,018*	0,003	0,013*	0,001	0,015*	0,003
Deinococci	0,275	0,016	0,005	0,012	0,002	0,014	0,003
Negativicutes	0,438	0,015	0,004	0,013	0,003	0,017	0,005
Chloroflexi	0,140	0,011	0,002	0,008	0,002	0,009	0,002
Opitutae	0,046	0,009*	0,002	0,006*	0,001	0,007*	0,001
Deferribacteres	0,048	0,008*	0,001	0,005*	0,002	0,006*	0,001
Thermotogae	0,250	0,008	0,003	0,006	0,001	0,009	0,002
Nitrospira	0,023	0,006*	0,001	0,004*	0,001	0,005*	0,001
Mollicutes	0,376	0,006	0,002	0,005	0,001	0,007	0,002
Synergistia	0,054	0,006	0,002	0,006	0,001	0,011	0,004
Chrysiogenetes	0,865	0,005	0,001	0,005	0,001	0,005	0,001
Aquificae	0,160	0,005	0,001	0,004	0,001	0,006	0,001
Solibacteres	0,138	0,005	0,001	0,004	0,001	0,006	0,001
Chlamydiae	0,032	0,005*	0,001	0,003*	0,000	0,005*	0,001
Totale		99,974		99,978		99,973	

Tabella 5.4 Classi identificate in seguito ad analisi MG-RAST significativamente aumentate (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05 < p < 0,01)

Classe	Fitasi	Fitasi + Inositolo
Chlorobia	-*	-*
Planctomycetacia	-*	-*
Opitutae	-*	-*
Deferribacteres	-*	-*
Nitrospira	-*	-*
Chlamydiae	-*	+*

La tabella 5.3 riassume i valori medi di abundance ottenuti con MG-RAST delle prime 30 classi batteriche identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. I Gammaproteobacteria rappresentano la classe più abbondante in tutti le carcasse testate, con valori % sempre superiori all'87%. Tutte le altre classi mostrano una abundance inferiore al 4 %. Le classi con valori di abundance compreso tra 1 e 4 % sono Bacilli, Betaproteobacteria, Flavobacteria e Clostridia. Tutte le classi con valori di abundance superiore all'1% non hanno mostrato differenze significative tra le carcasse appartenenti ai tre gruppi analizzati. Tuttavia, i Gammaproteobacteria sono diminuiti nei gruppi trattati rispetto al controllo e i Flavobacteria sono diminuiti rispetto al controllo soltanto nel gruppo trattato con fitasi. Al contrario, le classi Bacilli e Clostridia sono aumentate rispetto al controllo in entrambi i gruppi trattati, mentre i Betaproteobacteria sono aumentati soltanto nel gruppo trattato con fitasi.

Tra tutte le classi con abbondanza inferiore all'1%, quelle che hanno mostrato differenze significative nei tre gruppi sono rappresentate da: Chlorobia, Planctomycetacia, Opitutae, Deferribacteres, Nitrospira e Chlamydiae (Tabella 5.4). In particolare, le classi Chlorobia, Planctomycetacia, Opitutae, Deferribacteres e Nitrospira sono diminuite significativamente nei gruppi trattati rispetto al controllo, mentre la classe Chlamydiae è diminuita significativamente soltanto nel gruppo trattato con fitasi.

Tabella 5.5 Valori medi di abundance (%) ottenuti con MGmapper delle prime 17 Classi identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi e inositolo (* = 0,05<p<0,01)

Classe	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
Gammaproteobacteria	0,373	94,608	2,607	92,146	1,453	92,515	3,289
Bacilli	0,954	3,142	1,568	3,215	1,408	3,422	0,965
Flavobacteriia	0,180	0,901	0,204	0,550	0,109	0,829	0,392
Clostridia	0,183	0,850	1,039	3,237	1,028	2,669	2,712
Betaproteobacteria	0,107	0,354	0,081	0,507	0,235	0,257	0,094
Bacteroidia	0,207	0,050	0,016	0,231	0,232	0,119	0,037
Actinobacteria	0,045	0,038*	0,013	0,063*	0,020	0,100*	0,049
Alphaproteobacteria	0,082	0,020	0,007	0,017	0,002	0,029	0,010
Erysipelotrichia	0,179	0,009	0,006	0,004	0,003	0,003	0,002
Sphingobacteriia	0,152	0,008	0,003	0,005	0,001	0,023	0,022
Deltaproteobacteria	0,208	0,003	0,000	0,004	0,001	0,004	0,001
Deinococci	0,517	0,002	0,004	0,000	0,000	0,001	0,001
Epsilonproteobacteria	0,225	0,002	0,001	0,005	0,003	0,005	0,001
Tissierellia	0,647	0,002	0,001	0,002	0,001	0,002	0,001
Fusobacteriia	0,386	0,001	0,001	0,005	0,005	0,007	0,008
Negativicutes	0,266	0,001	0,001	0,002	0,001	0,003	0,001
Chlamydiia	0,802	0,001	0,000	0,001	0,000	0,001	0,000
Totale							

Tabella 5.6 Classe identificate in seguito ad analisi MGmapper significativamente aumentate (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05<p<0,01)

Classe	Fitasi	Fitasi + Inositolo
Actinobacteria	+*	+*

La tabella 5.5 riassume i valori medi di abundance ottenuti con MGmapper delle prime 17 classi batteriche identificate nei campioni di pelle dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. Le uniche classi con valori di abundance superiore all'1% sono rappresentate da Gammaproteobacteria, Bacilli e Clostridia. Nessuna di queste classi ha mostrato valori abundance significativamente diversi tra le carcasse appartenenti ai tre gruppi analizzati, anche se i Clostridia sono aumentati in entrambi i gruppi trattati.

Tra tutte le classi con abundance inferiore all'1% soltanto Actinobacteria sono risultati significativamente più alti nei gruppi trattati rispetto al controllo (Tabella 5.6).

Tabella 5.7 Valori medi di abundance (%) ottenuti con MG-RAST delle prime 30 Famiglie identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi + inositolo (* = 0,05<p<0,01) (= p<0.01)**

Famiglie	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
Moraxellaceae	0,000	56,396**	7,556	30,084**	9,389	27,760**	3,240
Enterobacteriaceae	0,038	18,965*	7,354	36,378*	13,079	33,267*	3,641
Aeromonadaceae	0,035	10,508*	2,342	15,573*	3,075	19,112*	5,915
Streptococcaceae	0,679	1,653	0,788	1,298	0,638	1,610	0,320
Flavobacteriaceae	0,164	1,352	0,325	0,986	0,135	2,056	1,247
Shewanellaceae	0,000	1,134**	0,198	1,691**	0,475	2,389**	0,186
Vibrionaceae	0,046	1,069*	0,158	1,425*	0,254	1,714*	0,471
Pseudomonadaceae	0,957	0,949	0,058	0,915	0,231	0,929	0,148
Clostridiaceae	0,177	0,780	0,746	2,588	0,879	2,227	2,037
Bacillaceae	0,145	0,632	0,246	0,724	0,106	0,477	0,102
Comamonadaceae	0,080	0,562	0,073	0,823	0,328	0,466	0,135
Pasteurellaceae	0,937	0,440	0,047	0,433	0,089	0,453	0,095
Burkholderiaceae	0,088	0,376	0,028	0,307	0,066	0,303	0,039
Enterococcaceae	0,066	0,337	0,200	0,390	0,104	0,733	0,334
Neisseriaceae	0,541	0,247	0,011	0,215	0,063	0,217	0,042
Micrococcaceae	0,039	0,184*	0,071	0,420*	0,149	0,307*	0,107
Staphylococcaceae	0,252	0,183	0,076	0,494	0,466	0,220	0,031
Xanthomonadaceae	0,306	0,135	0,007	0,112	0,022	0,122	0,025
Alteromonadaceae	0,376	0,131	0,012	0,112	0,022	0,122	0,022
Rhodocyclaceae	0,010	0,115*	0,012	0,080*	0,020	0,081*	0,011
Bacteroidaceae	0,184	0,108	0,018	0,283	0,207	0,191	0,058
Sphingobacteriaceae	0,232	0,105	0,030	0,073	0,009	0,151	0,101
Rhodobacteraceae	0,129	0,101	0,008	0,076	0,017	0,090	0,020
Cytophagaceae	0,193	0,101	0,020	0,070	0,008	0,134	0,078
Lactobacillaceae	0,252	0,093	0,034	0,485	0,515	0,627	0,571
Alcaligenaceae	0,034	0,092*	0,009	0,073*	0,017	0,066*	0,010
Pseudoalteromonadaceae	0,256	0,089	0,009	0,110	0,029	0,118	0,028
Peptostreptococcaceae	0,359	0,085	0,073	0,071	0,037	0,036	0,016
Bradyrhizobiaceae	0,073	0,082	0,008	0,067	0,010	0,073	0,005
Oxalobacteraceae	0,020	0,072*	0,007	0,054*	0,014	0,051*	0,005
Totale		97,076		96,407		96,104	

Tabella 5.8 Famiglie identificate in seguito ad analisi MG-RAST significativamente aumentate (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05<p<0,01) (= p<0.01)**

Famiglia	Fitasi	Fitasi + Inositolo
Moraxellaceae	-**	-**
Enterobacteriaceae	+*	+*
Aeromonadaceae	+*	+*
Shewanellaceae	+**	+**
Micrococcaceae	+*	+*
Rhodocyclaceae	-*	-*
Alcaligenaceae	-*	-*
Oxalobacteraceae	-*	-*

La tabella 5.7 riassume i valori medi di abundance ottenuti con MG-RAST delle prime 30 famiglie batteriche identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. Moraxellaceae, Enterobacteriaceae Aeromonadaceae rappresentano le famiglie più abbondanti in tutte le carcasse analizzate. In particolare, le Moraxellaceae sono diminuite significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (dal 56,396 al 30,084 %) che nel gruppo trattato con fitasi e inositolo (da 56,396 a 27,760 %). Al contrario, le Enterobacteriaceae sono aumentate significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 18,965 a 36,378) che nel gruppo trattato con fitasi e inositolo (da 18,965 a 33,267). Analogamente le Aeromonadaceae sono aumentate significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 10,508 a 15,573) che nel gruppo trattato con fitasi e inositolo (da 10,508 a 19,112).

Le famiglie che hanno mostrato una abundance compresa tra l'1 e il 2 % sono risultate le Streptococcaceae, Flavobacteriaceae, Clostridiaceae, Shewanellaceae e Vibrionaceae e tra queste sia le Shewanellaceae che le Vibrionaceae sono risultate significativamente superiori nei gruppi trattati rispetto al controllo (Tabella 5.8). Tra le famiglie con valori di abundance inferiori all'1 %, quelle significativamente diverse nei gruppi analizzati sono risultati Micrococcaceae, Rhodocyclaceae, Alcaligenaceae e Oxalobacteraceae. In particolare, le Micrococcaceae sono aumentate rispetto al controllo rispetto a tutti i gruppi trattati, mentre le Rhodocyclaceae, Alcaligenaceae e Oxalobacteraceae sono diminuite rispetto al controllo in tutti i gruppi trattati.

Tabella 5.9 Valori medi di abundance (%) ottenuti con MGmapper delle prime 30 Famiglie identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi + inositolo (* = 0,05<p<0,01) (= p<0.01)**

Famiglie	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
Moraxellaceae	0,000	51,364**	10,172	23,715**	8,597	22,051**	1,479
Enterobacteriaceae	0,087	22,836	11,823	42,005	15,085	37,562	5,880
Aeromonadaceae	0,110	19,479	3,758	25,178	5,991	30,952	9,892
Streptococcaceae	0,372	1,663	0,974	0,959	0,677	1,090	0,397
Flavobacteriaceae	0,180	0,900	0,204	0,550	0,109	0,827	0,391
Clostridiaceae	0,172	0,773	0,977	3,182	1,037	2,610	2,691
Planococcaceae	0,105	0,571	0,286	0,676	0,161	0,361	0,067
Shewanellaceae	0,006	0,471**	0,156	0,819**	0,381	1,484**	0,482
Enterococcaceae	0,029	0,334*	0,309	0,455*	0,164	1,173*	0,626
Pseudomonadaceae	0,889	0,306	0,039	0,299	0,115	0,327	0,079
Comamonadaceae	0,062	0,295	0,080	0,451	0,218	0,191	0,067
Bacillaceae	0,096	0,243	0,173	0,110	0,048	0,077	0,018
Staphylococcaceae	0,374	0,186	0,098	0,559	0,719	0,183	0,009
Paenibacillaceae	0,018	0,106*	0,031	0,061*	0,014	0,064*	0,012
Pasteurellaceae	0,412	0,066	0,049	0,049	0,024	0,034	0,014
Peptostreptococcaceae	0,314	0,053	0,051	0,032	0,027	0,015	0,001
Bacteroidaceae	0,169	0,043	0,015	0,207	0,203	0,086	0,020
Neisseriaceae	0,487	0,038	0,013	0,025	0,010	0,032	0,019
Idiomarinaceae	0,041	0,030*	0,013	0,017*	0,009	0,011*	0,002
Lactobacillaceae	0,293	0,023	0,012	0,377	0,478	0,462	0,498
Vibrionaceae	0,011	0,017*	0,003	0,020*	0,005	0,027*	0,004
Micrococcaceae	0,055	0,015	0,006	0,033	0,013	0,024	0,009
Bradyrhizobiaceae	0,145	0,012	0,005	0,009	0,001	0,013	0,001
Corynebacteriaceae	0,401	0,010	0,004	0,012	0,004	0,013	0,002
Erysipelotrichaceae	0,179	0,009	0,006	0,004	0,003	0,003	0,002
Burkholderiaceae	0,679	0,008	0,001	0,009	0,002	0,010	0,003
Eubacteriaceae	0,068	0,008	0,004	0,006	0,002	0,013	0,004
Sphingobacteriaceae	0,153	0,008	0,003	0,005	0,001	0,022	0,021
Xanthomonadaceae	0,306	0,008	0,002	0,011	0,004	0,016	0,012
Carnobacteriaceae	0,577	0,007	0,004	0,007	0,004	0,005	0,002
Totale		99,880		99,841		99,740	

Tabella 5.10 Famiglie identificate in seguito ad analisi MGmapper significativamente aumentate (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05<p<0,01) (= p<0.01)**

Famiglia	Fitasi	Fitasi + Inositolo
Moraxellaceae	-**	-**
Shewanellaceae	+**	+**
Enterococcaceae	+*	+*
Paenibacillaceae	-*	-*
Idiomarinaceae	-*	-*
Vibrionaceae	+*	+*

La tabella 5.9 riassume i valori medi di abundance ottenuti con MGmapper delle prime 30 famiglie batteriche identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. Moraxellaceae, Enterobacteriaceae Aeromonadaceae rappresentano le famiglie più abbondanti in tutte le carcasse analizzate. In particolare, le Moraxellaceae sono diminuite significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (dal 51,364 a 23,715 %) che nel gruppo trattato con fitasi e inositolo (da 51,364 a 22,051 %). Al contrario, le Enterobacteriaceae sono aumentate significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 22,836 a 42,005) che nel gruppo trattato con fitasi e inositolo (da 22,836 a 37,562). Analogamente, le Aeromonadaceae sono significativamente aumentate rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 19,479 a 25,178 %) che nel gruppo trattato con fitasi e inositolo (da 19,479 a 30,952).

Le famiglie che hanno mostrato una abundance compresa tra l'1 e il 3 % sono risultate le Streptococcaceae, Clostridiaceae, Shewanellaceae e Enterococcaceae. Tra queste sia le Shewanellaceae che le Enterococcaceae sono risultate significativamente superiori nei gruppi trattati rispetto al controllo (Tabella 5.10). Tra le famiglie con valori di abundance inferiori all'1 %, quelle significativamente diverse nei gruppi analizzati sono risultati Paenibacillaceae, Idiomarinaceae Vibrionaceae. In particolare, le Paenibacillaceae e Idiomarinaceae sono diminuite rispetto al controllo in tutti i gruppi trattati, mentre le Vibrionaceae sono aumentate rispetto al controllo in tutti i gruppi trattati (Tabella 5.10).

Tabella 5.11 Valori medi di abundance (%) ottenuti con MG-RAST delle prime 30 Specie identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi + inositolo (* = 0,05<p<0,01) (= p<0.01)**

Specie	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
<i>Acinetobacter johnsonii</i>	0,000	19,763**	2,944	9,934**	2,996	9,757***	1,223
<i>Acinetobacter baumannii</i>	0,000	17,120**	2,470	9,027**	2,890	8,396**	0,994
<i>Escherichia coli</i>	0,064	9,005	6,513	22,640	9,966	18,688	5,256
<i>Acinetobacter sp. Adp1</i>	0,000	8,752**	1,220	4,543**	1,426	4,273**	0,496
<i>Aeromonas hydrophila</i>	0,034	5,485*	1,210	8,166*	1,643	10,063*	3,137
<i>Aeromonas salmonicida</i>	0,035	4,812*	1,095	7,125*	1,372	8,675*	2,660
<i>Acinetobacter sp. Dr1</i>	0,000	3,120**	0,458	1,488**	0,490	1,453**	0,201
<i>Acinetobacter lwoffii</i>	0,001	1,646**	0,246	1,000**	0,323	0,818**	0,077
<i>Salmonella enterica</i>	0,015	1,148*	0,197	1,810*	0,438	1,993*	0,395
<i>Acinetobacter junii</i>	0,013	1,093*	0,172	0,765*	0,320	0,550*	0,087
<i>Acinetobacter radioresistens</i>	0,001	0,810**	0,118	0,505**	0,160	0,409**	0,034
<i>Acinetobacter haemolyticus</i>	0,002	0,725**	0,083	0,466**	0,167	0,374**	0,034
<i>Enhydrobacter aerosaccus</i>	0,015	0,711*	0,232	0,479*	0,052	0,363*	0,071
<i>Klebsiella pneumoniae</i>	0,419	0,694	0,073	0,790	0,198	0,841	0,162
<i>Shigella flexneri</i>	0,054	0,593	0,424	1,431	0,584	1,232	0,301
<i>Enterobacter sp. 638</i>	0,291	0,592	0,052	0,497	0,033	0,563	0,130
<i>Citrobacter koseri</i>	0,146	0,495	0,044	0,559	0,080	0,871	0,454
<i>Clostridium perfringens</i>	0,151	0,481	0,559	2,293	0,886	1,931	1,968
<i>Chryseobacterium gleum</i>	0,127	0,474	0,126	0,375	0,062	0,707	0,349
<i>Enterobacter cloacae</i>	0,329	0,457	0,031	0,384	0,039	0,470	0,136
<i>Citrobacter sp. 30_2</i>	0,140	0,439	0,080	0,399	0,062	0,923	0,655
<i>Acinetobacter sp. Atcc</i>	0,009	0,407**	0,051	0,276**	0,097	0,241**	0,028
<i>Psychrobacter sp. Prwf-1</i>	0,007	0,395**	0,090	0,249**	0,081	0,199**	0,037
<i>Proteus mirabilis</i>	0,455	0,394	0,172	0,541	0,177	0,469	0,130
<i>Escherichia sp. 3_2_53faa</i>	0,085	0,380	0,256	1,077	0,585	0,848	0,300
<i>Streptococcus pyogenes</i>	0,424	0,367	0,203	0,236	0,142	0,251	0,075
<i>Providencia rustigianii</i>	0,898	0,343	0,433	0,309	0,288	0,245	0,067
<i>Shigella dysenteriae</i>	0,058	0,308	0,209	0,741	0,315	0,632	0,154
<i>Shewanella baltica</i>	0,002	0,301**	0,076	0,537**	0,163	0,853**	0,236
<i>Acinetobacter genomosp</i>	0,001	0,284**	0,042	0,171**	0,061	0,143**	0,009
Totale		81,595		78,813		77,229	

Tabella 5.12 Specie identificate in seguito ad analisi MG-RAST aumentate significativamente (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05<p<0,01) (= p<0.01)**

Specie	Fitasi	Fitasi + Inositolo
<i>Acinetobacter johnsonii</i>	-**	-**
<i>Acinetobacter baumannii</i>	-**	-**
<i>Acinetobacter sp. ADP1</i>	-**	-**
<i>Aeromonas hydrophila</i>	+*	+*
<i>Aeromonas salmonicida</i>	+*	+*
<i>Acinetobacter sp. DRI</i>	-**	-**
<i>Acinetobacter lwoffii</i>	-**	-**
<i>Salmonella enterica</i>	+*	+*
<i>Acinetobacter junii</i>	-*	-*
<i>Acinetobacter radioresistens</i>	-**	-**
<i>Acinetobacter haemolyticus</i>	-**	-**
<i>Enhydrobacter aerosaccus</i>	-*	-*
<i>Acinetobacter sp. ATCC</i>	-**	-**
<i>Psychrobacter sp. PRwf-1</i>	-**	-**
<i>Shewanella baltica</i>	+**	+**
<i>Acinetobacter genomosp. 3</i>	-**	-**

La tabella 5.11 riassume i valori medi di abundance ottenuti con MG-RAST delle prime 30 specie batteriche identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. Le specie con abundance superiore al 10 % in almeno un gruppo sono rappresentate da *Acinetobacter johnsonii*, *Acinetobacter baumannii* *Escherichia coli* *Aeromonas hydrophila*. Tutte queste specie ad eccezione di *Escherichia coli* (p=0,064) sono risultate significativamente diverse nei gruppi analizzati. In particolare *Acinetobacter johnsonii* è significativamente diminuito rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (dal 19,763 a 9,934 %) che nel gruppo trattato con fitasi e inositolo (da 19,763 a 9,757 %). Analogamente, *Acinetobacter baumannii* è significativamente diminuito rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 17,120 a 9,027 %) che nel gruppo trattato con fitasi e inositolo (da 17,120 a 8,396 %). Al contrario, *Aeromonas hydrophila* è significativamente aumentata

rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (da 5,485 a 8,166 %) che nel gruppo trattato con fitasi e inositolo (da 5,485 a 10,063 %).

Le specie che hanno mostrato una abundance compresa tra l'1 e il 9 % in almeno un gruppo sono risultate le *Acinetobacter sp. ADP1*, *Aeromonas salmonicida*, *Acinetobacter sp. DR1*, *Acinetobacter lwoffii*, *Salmonella enterica*, *Acinetobacter junii*, *Shigella flexneri*, *Clostridium perfringens*, *Escherichia sp. 3_2_53FAA*. Tra queste specie sia *Acinetobacter sp. ADP1*, *Acinetobacter sp. DR1*, *Acinetobacter lwoffii* e *Acinetobacter junii* sono risultate significativamente inferiori nei gruppi trattati rispetto al controllo (Tabella 5.12). Al contrario *Aeromonas salmonicida* e *Salmonella enterica* aumentano significativamente nei gruppi trattati rispetto la controllo.

Tra le specie con valori di abundance inferiori all'1 %, quelle significativamente diverse nei gruppi analizzati sono risultati *Acinetobacter radioresistens*, *Acinetobacter haemolyticus*, *Enhydrobacter aerosaccus*, *Acinetobacter sp. ATCC*, *Psychrobacter sp. PRwf-1*, *Shewanella baltica* e *Acinetobacter genomosp. 3*. In particolare tra queste specie *Acinetobacter radioresistens*, *Acinetobacter haemolyticus*, *Enhydrobacter aerosaccus*, *Acinetobacter sp. ATCC*, *Psychrobacter sp. PRwf-1* e *Acinetobacter genomosp 3* sono diminuite significativamente rispetto al controllo in tutti i gruppi trattati, mentre *Shewanella baltica* è aumentata rispetto al controllo in tutti i gruppi trattati (Tabella 5.12).

Tabella 5.13 Valori medi di abundance (%) ottenuti con MGmapper delle prime 30 Specie identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi + inositolo (* = 0,05<p<0,01) (= p<0.01)**

Specie	p-values	Controllo (%)	Controllo std. Dev. (%)	Fitasi (%)	Fitasi std. Dev. (%)	Fitasi + inositolo (%)	Fitasi + inositolo std. Dev. (%)
<i>Acinetobacter johnsonii</i>	0,000	32,860**	7,072	14,250**	4,773	14,209**	0,988
<i>Escherichia coli</i>	0,100	16,266	13,038	36,799	15,013	30,467	8,837
<i>Aeromonas veronii</i>	0,099	13,832	2,432	18,649	4,679	22,283	6,964
<i>Acinetobacter lwoffii</i>	0,000	9,837**	1,992	4,352**	1,440	4,231**	0,337
<i>Aeromonas hydrophila</i>	0,177	3,777	0,874	4,457	0,970	5,767	2,101
<i>Citrobacter freundii</i>	0,208	1,320	0,364	0,966	0,227	2,530	2,077
<i>Aeromonas media</i>	0,129	1,074	0,247	1,283	0,258	1,671	0,569
<i>Morganella morganii</i>	0,476	1,053	0,836	0,736	0,538	0,538	0,191
<i>Acinetobacter junii</i>	0,033	0,991*	0,280	0,675*	0,386	0,389*	0,091
<i>Acinetobacter townneri</i>	0,004	0,847**	0,271	0,426**	0,223	0,222**	0,092
<i>Acinetobacter baumannii</i>	0,000	0,801**	0,152	0,402**	0,151	0,336**	0,034
<i>Klebsiella oxytoca</i>	0,008	0,739**	0,206	0,355**	0,098	0,452**	0,110
<i>Clostridium perfringens</i>	0,165	0,730	0,948	3,151	1,039	2,569	2,674
<i>Providencia rustigianii</i>	0,754	0,696	1,076	0,425	0,439	0,352	0,121
<i>Streptococcus iniae</i>	0,234	0,667	0,301	0,361	0,285	0,383	0,203
<i>Kurthia sp. 11kri321</i>	0,094	0,556	0,279	0,662	0,158	0,347	0,064
<i>Aeromonas salmonicida</i>	0,260	0,534	0,385	0,464	0,082	0,813	0,339
<i>Acinetobacter bereziniae</i>	0,001	0,531**	0,075	0,299**	0,123	0,236**	0,032
<i>Empedobacter brevis</i>	0,112	0,505	0,130	0,272	0,070	0,367	0,201
<i>Moraxella bovoculi</i>	0,107	0,486	0,436	0,067	0,019	0,179	0,137
<i>Acinetobacter bouvetii</i>	0,585	0,475	0,377	0,359	0,185	0,290	0,100
<i>Streptococcus dysgalactiae</i>	0,173	0,433	0,409	0,180	0,171	0,079	0,022
<i>Acinetobacter gernerii</i>	0,005	0,406**	0,106	0,214**	0,106	0,160**	0,021
<i>Acinetobacter haemolyticus</i>	0,001	0,395**	0,078	0,201**	0,082	0,178**	0,013
<i>Acinetobacter sp. Tth0-4</i>	0,000	0,393**	0,079	0,175**	0,060	0,170**	0,013
<i>Acinetobacter venetianus</i>	0,001	0,382**	0,083	0,180**	0,075	0,169**	0,018
<i>Acinetobacter radioresistens</i>	0,001	0,374**	0,076	0,200**	0,083	0,143**	0,013
<i>Acinetobacter tandoii</i>	0,001	0,368**	0,094	0,170**	0,065	0,158**	0,008
<i>Acinetobacter parvus</i>	0,020	0,324*	0,058	0,199*	0,091	0,184*	0,034
<i>Enterobacter cloacae</i>	0,003	0,274**	0,027	0,170**	0,051	0,179**	0,029
Totale		91,925		91,096		90,051	

Tabella 5.14 Specie identificate in seguito ad analisi MGmapper significativamente aumentate (+) o diminuite (-) nel gruppo trattato con fitasi e fitasi e inositolo rispetto al gruppo di controllo (* = 0,05<p<0,01) (= p<0.01)**

Specie	Fitasi	Fitasi + Inositolo
<i>Acinetobacter johnsonii</i>	_**	_**
<i>Acinetobacter lwoffii</i>	_**	_**
<i>Acinetobacter junii</i>	_*	_*
<i>Acinetobacter townneri</i>	_**	_**
<i>Acinetobacter baumannii</i>	_**	_**
<i>Klebsiella oxytoca</i>	_**	_**
<i>Acinetobacter bereziniae</i>	_**	_**
<i>Acinetobacter gernerii</i>	_**	_**
<i>Acinetobacter haemolyticus</i>	_**	_**
<i>Acinetobacter sp. TTH0-4</i>	_**	_**
<i>Acinetobacter venetianus</i>	_**	_**
<i>Acinetobacter radioresistens</i>	_**	_**
<i>Acinetobacter tandoii</i>	_**	_**
<i>Acinetobacter parvus</i>	_*	_*
<i>Enterobacter cloacae</i>	_**	_**

La tabella 5.13 riassume i valori medi di abundance ottenuti con MGmapper delle prime 30 specie batteriche identificate nei campioni di pelle ottenuti dagli animali del gruppo di controllo, del gruppo trattato con fitasi e del gruppo trattato con fitasi ed inositolo. Le specie con abundance superiore al 10 % in almeno un gruppo sono rappresentate da *Acinetobacter johnsonii*, *Escherichia coli* e *Aeromonas veronii*. Tra queste specie soltanto *Acinetobacter johnsonii* è diminuito significativamente rispetto al gruppo di controllo sia nel gruppo trattato con fitasi (dal 32,860 a 14,250 %) che nel gruppo trattato con fitasi e inositolo (da 32,860 a 14,209 %).

Le specie che hanno mostrato una abundance compresa tra l'1 e il 9 % in almeno un gruppo sono risultate *Acinetobacter lwoffii*, *Aeromonas hydrophila*, *Citrobacter freundii*, *Aeromonas media* e *Morganella morganii*. Tra queste specie soltanto *Acinetobacter lwoffii* è risultato significativamente più abbondante nel controllo rispetto ad entrambi i gruppi trattati (Tabella 5.14).

Tra le specie con valori di abundance inferiori all'1 %, quelle significativamente diverse nei gruppi analizzati sono risultate *Acinetobacter junii*, *Acinetobacter townneri*, *Acinetobacter baumannii*, *Acinetobacter bereziniae*, *Acinetobacter gernerii*, *Acinetobacter haemolyticus*, *Acinetobacter sp. TTH0-4*, *Acinetobacter venetianus*, *Acinetobacter radioresistens*, *Acinetobacter tandoii*, *Acinetobacter parvus* e *Enterobacter cloacae*. Tutte queste specie sono diminuite significativamente rispetto al controllo in tutti i gruppi trattati (Tabella 5.14).

Tabella 5.15 Valori medi della frequenza relativa (%) della abundance di classe e famiglia prelevati da pelle del collo di polli appartenenti al gruppo di controllo, fitasi e fitasi + inositolo, elaborati con MG-RAST (* = 0,05<p<0,01) (= p<0.01)**

Classe	Famiglia	p-value	Controllo	Fitasi	Fitasi + inositolo
Gammaproteobacteria		0,184	90,931	88,637	87,785
	Moraxellaceae	0,000	56,396**	30,084**	27,760**
	Enterobacteriaceae	0,038	18,965*	36,378*	33,267*
	Aeromonadaceae	0,035	10,508*	15,573*	19,112*
	Shewanellaceae	0,000	1,134**	1,691**	2,389**
	Vibrionaceae	0,046	1,069*	1,425*	1,714*
	Pseudomonadaceae	0,957	0,949	0,915	0,929
	Pasteurellaceae	0,937	0,440	0,433	0,453
	Xanthomonadaceae	0,306	0,135	0,112	0,122
	Alteromonadaceae	0,376	0,131	0,112	0,122
			89,728	86,722	85,868
Bacilli		0,598	3,044	3,571	3,869
	Streptococcaceae	0,679	1,653	1,298	1,610
	Bacillaceae	0,145	0,632	0,724	0,477
	Enterococcaceae	0,066	0,337	0,390	0,733
	Staphylococcaceae	0,252	0,183	0,494	0,220
			2,805	2,906	3,041
Betaproteobacteria		0,305	1,637	1,712	1,342
	Comamonadaceae	0,080	0,562	0,823	0,466
	Burkholderiaceae	0,088	0,376	0,307	0,303
	Neisseriaceae	0,541	0,247	0,215	0,217
	Rhodocyclaceae	0,010	0,115*	0,080*	0,081*
			1,300	1,424	1,068
Flavobacteria		0,168	1,424	1,044	2,155
	Flavobacteriaceae	0,164	1,352	0,986	2,056
			1,352	0,986	2,056
Clostridia		0,191	1,023	2,846	2,479
	Clostridiaceae	0,177	0,780	2,588	2,227
			0,780	2,588	2,227
Actinobacteria		0,058	0,423	0,766	0,688
	Micrococcaceae	0,039	0,184*	0,420*	0,307*
			0,184	0,420	0,307

Tabella 5.16 Valori medi della frequenza relativa (%) della abundance di classe e famiglia prelevati da pelle del collo di polli appartenenti al gruppo di controllo, fitasi e fitasi + inositolo, elaborati con MGmapper (* = 0,05<p<0,01) (= p<0.01)**

Classe	Famiglia	p-value	Controllo	Fitasi	Fitasi + inositolo
Gammaproteobacteria		0,373	94,608	92,146	92,515
	Moraxellaceae	0,000	51,364**	23,715**	22,051**
	Enterobacteriaceae	0,087	22,836	42,005	37,562
	Aeromonadaceae	0,110	19,479	25,178	30,952
	Shewanellaceae	0,006	0,471**	0,819**	1,484**
	Pseudomonadaceae	0,889	0,306	0,299	0,327
	Pasteurellaceae	0,412	0,066	0,049	0,034
	Idiomarinaceae	0,041	0,030*	0,017*	0,011*
			94,552	92,082	92,423
Bacilli		0,954	3,142	3,215	3,422
	Streptococcaceae	0,372	1,663	0,959	1,090
	Planococcaceae	0,105	0,571	0,676	0,361
	Enterococcaceae	0,029	0,334*	0,455*	1,173*
	Bacillaceae	0,096	0,243	0,110	0,077
	Staphylococcaceae	0,374	0,186	0,559	0,183
	Paenibacillaceae	0,018	0,106*	0,061*	0,064*
	Lactobacillaceae	0,293	0,023	0,377	0,462
			3,125	3,196	3,411
Flavobacteriia		0,180	0,901	0,550	0,829
	Flavobacteriaceae	0,180	0,900	0,550	0,827
			0,900	0,550	0,827
Clostridia		0,183	0,850	3,237	2,669
	Clostridiaceae	0,172	0,773	3,182	2,610
	Peptostreptococcaceae	0,314	0,053	0,032	0,015
			0,826	3,214	2,625
Betaproteobacteria		0,107	0,354	0,507	0,257
	Comamonadaceae	0,062	0,295	0,451	0,191
	Neisseriaceae	0,487	0,038	0,025	0,032
			0,332	0,476	0,223
Bacteroidia		0,207	0,050	0,231	0,119
	Bacteroidaceae	0,169	0,043	0,207	0,086
			0,043	0,207	0,086

Le tabelle 5.15 e 5.16 riassumono quanto descritto nelle tabelle precedenti, associando le famiglie più abbondanti identificate in ogni gruppo di carcasse alle corrispondenti classi, in funzione dei dati ottenuti, rispettivamente, con MG-RAST e MGmapper.

5.4 Tabelle con i risultati metabolici

Tabella 5.17 Geni con % di abundance superiore allo 0.2% e significativamente maggiore nel gruppo di controllo rispetto ai gruppi trattati con fitasi

Funzione	p-values	Controllo (%)	SD (%)	Fitasi (%)	SD (%)	Fitasi+inositolo: (%)	SD (%)
Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	0,000	0,553	0,026	0,448	0,017	0,437	0,015
Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)	0,030	0,243	0,020	0,210	0,018	0,210	0,013
Sam-dependent methyltransferases	0,000	0,277	0,011	0,242	0,007	0,243	0,011
Gamma-aminobutyrate permease and related permeases	0,005	0,290	0,020	0,238	0,026	0,219	0,029
Glutamate synthase domain 3	0,003	0,236	0,014	0,199	0,022	0,186	0,012
Dihydroxyacid dehydratase/phosphogluconate dehydratase	0,000	0,202	0,006	0,162	0,016	0,161	0,010
Malic enzyme	0,006	0,219	0,019	0,173	0,018	0,183	0,015
Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (e3) component, and related enzymes	0,000	0,273	0,020	0,222	0,015	0,217	0,004
Outer membrane receptor proteins, mostly fe transport	0,004	0,586	0,074	0,449	0,051	0,449	0,011
Acetyl-coa acetyltransferase	0,000	0,291	0,026	0,216	0,025	0,203	0,009
Acyl-coa dehydrogenases	0,000	0,660	0,078	0,424	0,078	0,398	0,013
Acyl-coenzyme a synthetases/amp-(fatty) acid ligases	0,003	0,222	0,017	0,177	0,021	0,174	0,013
Enoyl-coa hydratase/carnithine racemase	0,000	0,217	0,016	0,167	0,017	0,166	0,004
Acyl-coa synthetases (amp-forming)/amp-acid ligases ii	0,000	0,342	0,035	0,249	0,030	0,225	0,009
Type iia topoisomerase (dna gyrase/topo ii, topoisomerase iv), b subunit	0,014	0,240	0,019	0,218	0,017	0,202	0,008
Guanosine polyphosphate pyrophosphohydrolases/synthetases	0,001	0,245	0,018	0,210	0,014	0,197	0,003
Glutamyl- and glutaminyl-trna synthetases	0,011	0,208	0,014	0,179	0,014	0,184	0,005
Soluble lytic murein transglycosylase and related regulatory proteins (some contain lysm/invasin domains)	0,003	0,214	0,006	0,194	0,008	0,194	0,009
Type i restriction-modification system methyltransferase subunit	0,001	0,217	0,028	0,147	0,024	0,150	0,010
Atpases with chaperone activity, atp-binding subunit	0,021	0,324	0,026	0,285	0,010	0,283	0,021
Glutathione s-transferase	0,001	0,217	0,014	0,176	0,013	0,179	0,010

Tabella 5.18 Geni con % di abundance superiore allo 0.2% e significativamente maggior nei gruppi trattati con fitasi rispetto al controllo

Funzione	p-values	Controllo (%)	SD (%)	Fitasi (%)	SD (%)	Fitasi+inositolo: (%)	SD (%)
Abc-type oligopeptide transport system, periplasmic component	0,011	0,232	0,027	0,274	0,012	0,293	0,030
Amino acid permeases	0,002	0,171	0,035	0,258	0,035	0,259	0,014
Abc-type sugar transport system, atpase component	0,021	0,142	0,036	0,216	0,049	0,211	0,011
Glycosidases	0,000	0,174	0,028	0,252	0,012	0,269	0,006
Thiamine pyrophosphate-requiring enzymes [acetolactate synthase, pyruvate dehydrogenase (cytochrome), glyoxylate carboligase, phosphonopyruvate decarboxylase]	0,028	0,179	0,018	0,208	0,013	0,198	0,003
Anaerobic dehydrogenases, typically selenocysteine-containing	0,022	0,400	0,047	0,457	0,037	0,480	0,012
Pyruvate-formate lyase	0,001	0,141	0,024	0,214	0,028	0,212	0,004
Integrase	0,009	0,207	0,017	0,238	0,017	0,247	0,012
Response regulators consisting of a chey-like receiver domain and a winged-helix dna-binding domain	0,016	0,373	0,008	0,373	0,007	0,387	0,005
Superfamily ii dna and rna helicases	0,000	0,238	0,012	0,270	0,010	0,297	0,020
Abc-type amino acid transport/signal transduction systems, periplasmic component/domain	0,001	0,235	0,021	0,285	0,018	0,327	0,036
Fog: eal domain	0,003	0,188	0,043	0,259	0,018	0,275	0,023
Fog: ggdef domain	0,001	0,359	0,044	0,428	0,019	0,465	0,023
Fog: pas/pac domain	0,046	0,219	0,020	0,250	0,026	0,282	0,044

Le tabelle 5.17 e 5.18 riassumono i geni funzionali identificati utilizzando il Kyoto Encycloperdia of Genes and Genomes (KEGG) database (www.genome.jp/kegg) (Kanehisa M. *et al.*, 2002) con abundance superiore allo 0,2 % in almeno un gruppo. In particolare la tabella 5.17 riporta i geni significativamente diminuiti nei gruppi trattati con fitasi e fitasi e inositolo rispetto al controllo mentre la tabella 5.18 riporta i geni significativamente aumentati nei gruppi trattati.

Tra i geni significativamente diversi nei gruppi analizzati e con abundance superiore allo 0,2% in almeno un gruppo, i geni *Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)*, *Predicted hydrolases or acyltransferases (alpha/beta hydrolase superfamily)*, *SAM-dependent methyltransferases* sono stati identificati in modo incerto e l'analisi bioinformatica ha consentito soltanto una loro predizione.

Tra i geni associati al trasporto e metabolismo degli aminoacidi (AA), un gene associato alla componente periplasmatica del sistema di trasporto degli oligopeptidi e un gene codificante per AA permeasi sono aumentati nei gruppi trattati mentre un gene per gamma aminobutirrato permeasi e glutamato sintasi sono diminuiti nei gruppi trattati.

Tra i geni associati al metabolismo e trasporto dei carboidrati, un gene codificante glicosidasi ed una componente ATPasica di un sistema di trasporto degli zuccheri sono significativamente aumentati nei gruppi trattati, mentre un gene codificante *Dihydroxyacid dehydratase/phosphogluconate dehydratase* è significativamente diminuito nei gruppi trattati.

Tra i geni codificanti per sistemi di conversione e produzione energetica, un gene codificate per deidrogenasi anaerobiche e piruvato formato liasi sono significativamente aumentati nei gruppi trattati mentre un gene codificante per l'enzima malico e per il complesso *Pyruvate/2-oxoglutarate dehydrogenase complex, dihydrolipoamide dehydrogenase (E3) component, and related enzymes* sono significativamente diminuiti.

Tutti i geni i geni associati a trasporto e metabolismo lipidico, rappresentati da *Acetyl-CoA acetyltransferase*, *Acyl-CoA dehydrogenases*, *Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases*, *Enoyl-CoA hydratase/carnithine racemase* sono significativamente diminuiti nei gruppi trattati rispetto al gruppo di controllo.

Tra i geni per processi cellulari e trasduzione del segnale due geni codificanti per chaperoni e un gene associato a meccanismi di difesa cellulare sono risultati significativamente inferiori nei gruppi trattati mentre quattro geni associati a meccanismi di trasduzione del segnale FOG sono significativamente aumentati nei gruppi trattati.

6 DISCUSSIONE

I dati ottenuti in questa ricerca riassumono la composizione tassonomica e la composizione in geni funzionali associata a carcasse di pollo da carne raccolte al macello al termine del tunnel di raffreddamento. Le carcasse analizzate erano ottenute da animali alimentati con una dieta di controllo, una dieta addizionata con 1500 unità di fitasi per kg di mangime ed una dieta addizionata con la stessa concentrazione di fitasi e 3 g/kg di inositolo. Anche se la composizione intestinale degli animali è soltanto uno dei fattori che può influenzare la composizione microbiologica delle carcasse, differenze significative sono state identificate sia a livello tassonomico, di classe, famiglia e specie, che a livello di geni funzionali tra le carcasse del gruppo di controllo che le carcasse dei gruppi alimentati con fitasi.

Gli aspetti innovativi di questa ricerca sono principalmente due: il primo è rappresentato dalla applicazione della metagenomica shotgun ad una matrice alimentare; il secondo è rappresentato dall'analisi bioinformatica dei dati tassonomici eseguita comparativamente con due sistemi diversi, rappresentati da MG-RAST e MGmapper.

I risultati ottenuti, sia in relazione alla composizione tassonomica dei campioni, che in relazione alla composizione in geni funzionali, sono molti di più di quelli presentati, che fanno riferimento, per i gruppi microbici alle prime 30 classi, famiglie e specie identificate in termini di abundance; per i geni funzionali a quelli con abundance superiore 0,2 in almeno un gruppo. Anche applicando questi criteri di selezione i risultati ottenuti sono tanti e non tutti spiegabili attraverso lo studio della letteratura nota, specialmente in relazione ai geni funzionali. L'ottenimento mediante analisi metagenomica di dati non ancora interpretabili può essere considerato un limite. Tuttavia l'identificazione di differenze significative tra gruppi in relazione a determinati microorganismi o geni funzionali rappresenta il primo passo per comprendere i meccanismi biologici alla base di tali differenze.

Il grande potenziale dell'analisi metagenomica riguarda la possibilità di studiare le popolazioni microbiche (non solo batteri ma anche virus ed eucarioti) nel momento della loro interazione. La composizione in geni funzionali fornisce un'indicazione quantitativa dei geni presenti. Tuttavia l'associazione di specifici geni a ciascun microorganismo rappresenta una sfida ancora aperta a livello di analisi bioinformatica. Inoltre, l'analisi metagenomica mappa i geni presenti in un campione ma per sapere quali di questi geni verranno trascritti e tradotti in proteine ed enzimi attivi nelle vie metaboliche dei batteri è necessario seguire un approccio multi-omico integrato, che prevede l'impiego di metagenomica (analisi dei geni) metatrascrittomica (analisi dell'mRNA) e metaproteomica (analisi delle proteine).

Il sistema di analisi MG-RAST di default analizza sequenze con un Phred =15; ne consegue che le reads elaborate sono risultate maggiori rispetto a quelle elaborate da MGmapper che lavora con un

Phred superiore. Per questo motivo le reads identificate nei metagenomi di questa ricerca ed analizzate con MG-RAST sono state assegnate a più gruppi tassonomici e geni funzionali ma con una abundance inferiore rispetto a quelli ottenuti con analisi MGmapper.

In relazione ai dati tassonomici, la classe dei Gammaproteobacteria è risultata la più abbondante in tutti i gruppi testati ed analizzati, sia con MG-RAST che con MGmapper. Tale classe è rappresentata principalmente dalle famiglie Moraxellaceae, Enterobacteriaceae e Aeromonadaceae.

La famiglia delle Moraxellaceae appartiene all'ordine delle Pseudomonadales all'interno della classe dei Gammaproteobacteria all'interno del phylum Proteobacteria. La famiglia delle Moraxellaceae è risultata significativamente più elevata nel controllo rispetto ai gruppi trattati con fitasi sia con MG-RAST che con MGmapper. Questa famiglia è stata proposta nel 1991 dal Rossau *et al.*, per includere i generi *Moraxella*, *Acinetobacter* e *Psychrobacter*. Le specie di *Moraxella* identificate nelle carcasse sono risultate: *Moorella thermoacetica*, *Moraxella catarrhalis* e *Moraxella bovis* ma con abundance sempre inferiori allo 0,2%. Inoltre nessuna di queste specie è risultata significativamente diversa nei gruppi analizzati. Al contrario la specie *Enhydrobacter aerosaccus* ha mostrato valori di abundance compresi tra lo 0,3 e lo 0,7 % ed è risultata significativamente più elevata nei gruppi trattati rispetto al controllo. Questa specie è già stata descritta nelle anatre come contaminante ambientale legato alla fase di eviscerazione degli animali (Huawei W. *et al.*, 2013). Il ruolo delle specie di *Moraxella* nella degradazione dei prodotti a base di carne non è chiaro perché non producono prodotti di degradazione degli AA come ad esempio *Pseudomonas* e *Shewanella*. Inoltre in passato sono state identificate come specie di *Moraxella* specie in realtà appartenenti a *Psychrobacter* e *Acinetobacter*.

Il genere *Acinetobacter* fa parte dei batteri degradativi della carne durante la conservazione. Inoltre può causare infezioni nosocomiali e mostra una attitudine a sviluppare forti resistenze a molteplici antibiotici. La proporzione di *Acinetobacter baumannii* isolati in Inghilterra con resistenza multipla agli antibiotici è aumentata da meno del 5% nel 1990 al 24% nel 2007 dimostrando così l'attitudine della specie ad acquisire geni di resistenza antibiotica con particolare riferimento ai beta lattami. Alcune specie sono in grado di degradare gli idrocarburi e quindi vengono utilizzate per la bonifica dei terreni.

Le specie di *Acinetobacter* identificate in questo studio sono: *Acinetobacter baumannii*, *Acinetobacter calcoaceticus*, *Acinetobacter genomosp. 13TU*, *Acinetobacter genomosp. 3*, *Acinetobacter haemolyticus*, *Acinetobacter johnsonii*, *Acinetobacter junii*, *Acinetobacter lwoffii*, *Acinetobacter radioresistens*, *Acinetobacter sp. ADPI*, *Acinetobacter sp. ATCC 27244*, *Acinetobacter sp. DRI*, *Acinetobacter venetianus* mentre quelle descritte in letteratura in relazione ai polli sono *Acinetobacter baumannii*, *Acinetobacter lwoffii* e *Acinetobacter calcoaceticus* (Yang X., 2014). In questa ricerca entrambe queste specie sono risultate significativamente inferiori nelle

carcasse ottenute da polli alimentati con fitasi rispetto alle carcasse ottenute dai polli del gruppo di controllo. Le specie identificate in questa ricerca sono: *Psychrobacter arcticus*, *Psychrobacter cryohalolentis*, *Psychrobacter sp. PRwf-1*. Il genere *Psychrobacter* fa parte dei batteri degradativi della carne ma tende a diminuire nel tempo perché non è un buon competitore con gli altri batteri degradativi.

La famiglia delle Enterobacteriaceae è risultata la seconda in termini di abundance ed è aumentata significativamente sia nelle carcasse ottenute da animali alimentati con fitasi che sulle carcasse ottenute da animali alimentati con fitasi e inositolo ma soltanto in funzione dei risultati ottenuti con MG-RAST. Tuttavia, anche i dati con MGmapper dimostrano un incremento della abundance di questa famiglia nei gruppi delle carcasse trattate con fitasi. All'interno della famiglia le specie significativamente diverse nei gruppi considerati sono *Salmonella enterica* ed *Enterobacter cloacae*. Tuttavia nella tabella che riporta le specie significativamente diverse nei gruppi in funzione dei dati analizzati con MGmapper (Tabella 5.14) *Salmonella enterica* non è riportata perché il p-value è leggermente maggiore al limite di significatività (0,068). L'aumento della abundance di *Salmonella enterica* nelle carcasse degli animali alimentati con fitasi rappresenta un elemento di preoccupazione per l'elevato numero di tossinfezioni alimentari legate a questo patogeno che in Italia nel 2016 sono state 2983 (relazione annuale Ministero della Salute). Al contrario di *Salmonella enterica*, *Enterobacter cloacae*, che tuttavia appartiene ai batteri degradativi della carne, nei gruppi trattati con fitasi diminuisce molto significativamente in funzione dell'analisi effettuata con MGmapper (Tabella 5.14). Una netta diminuzione si è osservata anche in relazione al patogeno *Klebsiella oxytoca* che nei dati analizzati con MGmapper ha dimostrato una diminuzione di abundance da 0,739 nel controllo a 0,355 e 0,452 nei gruppi trattati con fitasi e fitasi e inositolo.

Le Aeromonadaceae sono la terza famiglia in termini di abundance presente in tutti i gruppi analizzati. All'interno della famiglia, le specie *Aeromonas hydrophila* e *Aeromonas salmonicida* sono aumentate significativamente nei gruppi trattati con fitasi in funzione dell'analisi effettuata con MG-RAST. Tale aumento si evidenzia anche nei dati analizzati con MGmapper ma non è significativo. Mentre *Aeromonas hydrophila* è già stata descritta in prodotti a base di carne, *Aeromonas salmonicida* solitamente è associata a prodotti ittici. *Aeromonas hydrophila* nel gruppo di controllo ha mostrato una abundance pari a 5,485 e nei gruppi trattati con fitasi e fitasi e inositolo la abundance è aumentata a 8,166 e 10,063 rispettivamente.

L'unica specie non appartenente alle famiglie indicate e significativamente maggiore nei gruppi trattati è *Shewanella baltica*. Anche se questa specie è spesso associata a prodotti ittici ma è già stata descritta in prodotti a base di carne (Coton M. *et al.*, 2013).

L'analisi dei geni funzionali ha mostrato differenze significative tra geni associati al trasporto e metabolismo degli aminoacidi (AA), geni associati al metabolismo e trasporto dei carboidrati, geni codificanti per sistemi di conversione e produzione energetica, geni associati a trasporto e metabolismo lipidico e geni per processi cellulari e trasduzione del segnale. Il dato più significativo dal punto di vista biologico riguarda alcuni geni associati a trasporto e metabolismo lipidico, rappresentati da *Acetyl-CoA acetyltransferase*, *Acyl-CoA dehydrogenases*, *Acyl-coenzyme A synthetases/AMP-(fatty) acid ligases*, *Enoyl-CoA hydratase/carnithine racemase* significativamente diminuiti nei gruppi trattati rispetto al gruppo di controllo. La relazione tra l'aggiunta di fitasi nella dieta di polli da carne ed il metabolismo lipidico è già stata descritta (Liu N. *et al.*, 2010).

7 REFERENCES

- **Abel, H. J., & Duncavage, E. J. (2013).** Detection of structural DNA variation from next generation sequencing data: a review of informatic approaches. *Cancer genetics*, 206(12), 432-440.
- **Air, g., Sanger, F., Barrell, B., Brown, N., Coulson, A., Fiddes, J., Smith, M. (1977, January).** Nucleotide-Sequence of DNA of Bacteriophage-phix174. In *proceedings of the Australian Biochemical Society* (vol. 10, pp. 60-60). Monash univ dept Biochemistry, Clayton Victoria 3168, Australia: proc aust biochemical soc.
- **Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995).** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143-169.
- **Amann, R. I., Ludwig, W., & Schleifer, K. H. (1995).** Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological reviews*, 59(1), 143-169.
- **Apolloni, A. (2012).** Confronto di metodi statistici per la misura dell'espressione differenziale in dati di RNA sequencing.
- **Attanasio, M., Lapini, I., Evangelisti, L., Lucarini, L., Giusti, B., Porciani, M. C., Pepe, G. (2008).** FBN1 mutation screening of patients with Marfan syndrome and related disorders: detection of 46 novel FBN1 mutations. *Clinical genetics*, 74(1), 39-46.
- **Axelrod, D. (2001).** Total internal reflection fluorescence microscopy in cell biology. *Traffic*, 2(11), 764-774.
- **Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Meyer, F. (2008).** The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1), 75.
- **Bentley, D. R. (2006).** Whole-genome re-sequencing. *Current opinion in genetics & development*, 16(6), 545-552.
- **Bentley, D. R., Balasubramanian, S., Swerdlow, H. P., Smith, G. P., Milton, J., Brown, C. G., Boutell, J. M. (2008).** Accurate whole human genome sequencing using reversible terminator chemistry. *nature*, 456(7218), 53.
- **Berg, R. D. (1996).** The indigenous gastrointestinal microflora. *Trends in microbiology*, 4(11), 430-435.
- **Binladen, J., Gilbert, M. T. P., Bollback, J. P., Panitz, F., Bendixen, C., Nielsen, R., Willerslev, E. (2007).** The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PloS one*, 2(2), e197.

- **Bokulich, N. A.,** Bamforth, C. W., & Mills, D. A. (2012). Brewhouse-resident microbiota are responsible for multi-stage fermentation of American coolship ale. *PLoS One*, 7(4), e35507.
- **Bokulich, N. A.,** Lewis, Z. T., Boundy-Mills, K., Mills, D. A. (2016). A new perspective on microbial landscapes within food production. *Current opinion in biotechnology*, 37, 182-189.
- **Bokulich, N. A.,** Mills, D. A. (2012). Next-generation approaches to the microbial ecology of food fermentations. *BMB reports*, 45(7), 377-389.
- **Bonaïti, C.,** Parayre, S., & Irlinger, F. (2006). Novel extraction strategy of ribosomal RNA and genomic DNA from cheese for PCR-based investigations. *International journal of food microbiology*, 107(2), 171-179.
- **Boon, E. M.,** & Faas, B. H. (2013). Benefits and limitations of whole genome versus targeted approaches for noninvasive prenatal testing for fetal aneuploidies. *Prenatal diagnosis*, 33(6), 563-568.
- **Brady, A.,** Salzberg, S. L. (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature methods*, 6(9), 673-676.
- **Branton, D.,** Deamer, D. W., Marziali, A., Bayley, H., Benner, S. A., Butler, T., Jovanovich, S. B. (2008). The potential and challenges of nanopore sequencing. *Nature biotechnology*, 26(10), 1146-1153.
- **Breu, H.** (2010). A theoretical understanding of 2 base color codes and its application to annotation, error detection, and error correction. *White paper, Life Technologies*.
- **Buermans, H. P. J.,** & Den Dunnen, J. T. (2014). Next generation sequencing technology: advances and applications. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1842(10), 1932-1941.
- **Buttigieg, P. L.,** Hankeln, W., Kostadinov, I., Kottmann, R., Yilmaz, P., Duhaime, M. B., Glöckner, F. O. (2013). Ecogenomic perspectives on domains of unknown function: correlation-based exploration of marine metagenomes. *PLoS One*, 8(3), e50869.
- **Cai, L.,** Zhang, T. (2013). Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environmental science & technology*, 47(10), 5433-5441.
- **Caporaso, J. G.,** Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Huttley, G. A. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 7(5), 335-336.
- **Chessa, R.** (2014). *Diagnosi molecolare della Distrofia Muscolare di Duchenne e della Neurofibromatosi tipo 1 con la tecnologia di sequenziamento Next Generation* (Doctoral dissertation, Università degli Studi di Cagliari).

- **Chin, C. S.**, Sorenson, J., Harris, J. B., Robins, W. P., Charles, R. C., Jean-Charles, R. R., Paxinos, E. E. (2011). The origin of the Haitian cholera outbreak strain. *New England Journal of Medicine*, 364(1), 33-42.
- **Chitsaz, H.**, Yee-Greenbaum, J. L., Tesler, G., Lombardo, M. J., Dupont, C. L., Badger, J. H., Schulz-Trieglaff, O. (2011). Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nature biotechnology*, 29(10), 915-921.
- **Clarke, J.**, Wu, H. C., Jayasinghe, L., Patel, A., Reid, S., & Bayley, H. (2009). Continuous base identification for single-molecule nanopore DNA sequencing. *Nature nanotechnology*, 4(4), 265-270.
- **Cocolin, L.**, Ercolini, D. (2015). Zooming into food-associated microbial consortia: a 'cultural' evolution. *Current Opinion in Food Science*, 2, 43-50.
- **Cole, J. R.**, Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., Tiedje, J. M. (2008). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic acids research*, 37(suppl_1), D141-D145.
- **Collins, F. S.**, & McKusick, V. A. (2001). Implications of the Human Genome Project for medical science. *Jama*, 285(5), 540-544.
- **Costanza, J.** (2010). Algoritmi per l'analisi ed il troubleshooting di segnali di sequenziamento sanger del DNA.
- **Coton, M.**, Joffraud, J. J., Mekhtiche, L., Leroi, F., Coton, E. (2013). Biodiversity and dynamics of the bacterial community of packaged king scallop (*Pecten maximus*) meat during cold storage. *Food microbiology*, 35(2), 99-107.
- **Darling, A. E.**, Jospin, G., Lowe, E., Matsen IV, F. A., Bik, H. M., & Eisen, J. A. (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ*, 2, e243.
- **De Filippis, F.**, Genovese, A., Ferranti, P., Gilbert, J. A., & Ercolini, D. (2016). Metatranscriptomics reveals temperature-driven functional changes in microbiome impacting cheese maturation rate. *Scientific reports*, 6, 21871.
- **De Filippis, F.**, La Storia, A., Stellato, G., Gatti, M., & Ercolini, D. (2014). A selected core microbiome drives the early stages of three popular Italian cheese manufactures. *PLoS One*, 9(2), e89680.
- **De Filippis, F.**, Parente, E., & Ercolini, D. (2017). Metagenomics insights into food fermentations. *Microbial biotechnology*, 10(1), 91-102.
- **De Pasquale, I.**, Calasso, M., Mancini, L., Ercolini, D., La Storia, A., De Angelis, M., Gobbetti, M. (2014). Causal relationship between microbial ecology dynamics and proteolysis

- during manufacture and ripening of protected designation of origin (PDO) cheese Canestrato Pugliese. *Applied and environmental microbiology*, 80(14), 4085-4094.
- **De Pasquale, I.**, Di Cagno, R., Buchin, S., De Angelis, M., & Gobbetti, M. (2016). Spatial Distribution of the Metabolically Active Microbiota within Italian PDO Ewes' Milk Cheeses. *PloS one*, 11(4), e0153213.
 - **Del Chierico, F.**, Ancora, M., Marcacci, M., Camma, C., Putignani, L., & Conti, S. (2015). Choice of next-generation sequencing pipelines. *Bacterial Pangenomics: Methods and Protocols*, 31-47.
 - **Delgado, S.**, Rachid, C. T., Fernández, E., Rychlik, T., Alegría, Á., Peixoto, R. S., & Mayo, B. (2013). Diversity of thermophilic bacteria in raw, pasteurized and selectively-cultured milk, as assessed by culturing, PCR-DGGE and pyrosequencing. *Food microbiology*, 36(1), 103-111.
 - **DeSantis, T. Z.**, Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and environmental microbiology*, 72(7), 5069-5072.
 - **Deshpande, V.**, Fung, E. D., Pham, S., & Bafna, V. (2013). Cerulean: A hybrid assembly using high throughput short and long reads. In *International Workshop on Algorithms in Bioinformatics* (pp. 349-363). Springer, Berlin, Heidelberg.
 - **Deshpande, V.**, Fung, E. D., Pham, S., Bafna, V. (2013). Cerulean: A hybrid assembly using high throughput short and long reads. In *International Workshop on Algorithms in Bioinformatics* (pp. 349-363). Springer, Berlin, Heidelberg.
 - **Devlieghere, F.**, Lefevre, I., Magnin, A., & Debevere, J. (2000). Growth of *Aeromonas hydrophila* in modified-atmosphere-packed cooked meat products. *Food Microbiology*, 17(2), 185-196.
 - **Dick, G. J.**, Andersson, A. F., Baker, B. J., Simmons, S. L., Thomas, B. C., Yelton, A. P., Banfield, J. F. (2009). Community-wide analysis of microbial genome sequence signatures. *Genome biology*, 10(8), R85.
 - **Dolled-Filhart, M. P.**, Lee, M., Ou-yang, C. W., Haraksingh, R. R., & Lin, J. C. H. (2013). Computational and bioinformatics frameworks for next-generation whole exome and genome sequencing. *The Scientific World Journal*, 2013.
 - **Dugat-Bony, E.**, Straub, C., Teissandier, A., Onesime, D., Loux, V., Monnet, C., Fraud, S. (2015). Overview of a surface-ripened cheese community functioning by meta-omics analyses. *PLoS One*, 10(4), e0124360.

- **Dvořáková, J. (1998).** Phytase: sources, preparation and exploitation. *Folia microbiologica*, 43(4), 323-338.
- **Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., Bibillo, A. (2009).** Real-time DNA sequencing from single polymerase molecules. *Science*, 323(5910), 133-138.
- **Eisenstein, M. (2012).** Oxford Nanopore announcement sets sequencing sector abuzz.
- **Ercolini, D. (2013).** High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied and environmental microbiology*, 79(10), 3148-3155.
- **Ercolini, D. (2013).** High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. *Applied and environmental microbiology*, 79(10), 3148-3155.
- **Ercolini, D., Moschetti, G., Blaiotta, G., & Coppola, S. (2001).** The potential of a polyphasic PCR-DGGE Approach in evaluating microbial diversity of natural whey cultures for water-buffalo mozzarella cheese production: bias of culture-dependent and culture-independent analyses. *Systematic and Applied Microbiology*, 24(4), 610-617.
- **Erkus, O., De Jager, V. C., Spus, M., van Alen-Boerrigter, I. J., Van Rijswijck, I. M., Hazelwood, L., Smid, E. J. (2013).** Multifactorial diversity sustains microbial community stability. *The ISME journal*, 7(11), 2126-2136.
- **Fiers, W., Contreras, R., Duerinck, F., Haegeman, G., Iserentant, D., Merregaert, J., Volckaert, G. (1976).** Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature*, 260(5551), 500-507.
- **Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., Sonnhammer, E. L. (2013).** Pfam: the protein families database. *Nucleic acids research*, 42(D1), D222-D230.
- **Garcia-Garcerà, M., Garcia-Etxebarria, K., Coscollà, M., Latorre, A., Calafell, F. (2013).** A new method for extracting skin microbes allows metagenomic analysis of whole-deep skin. *PloS one*, 8(9), e74914.
- **Garrido-Cardenas, J. A., Garcia-Maroto, F., Alvarez-Bermejo, J. A., & Manzano-Agugliaro, F. (2017).** DNA sequencing sensors: an overview. *Sensors*, 17(3), 588.
- **Gerlach, W., Stoye, J. (2011).** Taxonomic classification of metagenomic shotgun sequences with CARMA3. *Nucleic acids research*, 39(14), e91-e91.
- **Giusti, B., Porciani, M. C., Brunelli, T., Evangelisti, L., Fedi, S., Gensini, G. F., Pepe, G. (2003).** Phenotypic variability of cardiovascular manifestations in Marfan Syndrome: Possible

- role of hyperhomocysteinemia and C677T MTHFR gene polymorphism. *European heart journal*, 24(22), 2038-2045.
- **Goodwin, S.**, Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. C., & McCombie, W. R. (2015). Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. *Genome research*, 25(11), 1750-1756.
 - **Gould, T. J.**, Myers, J. R., & Bewersdorf, J. (2011). Total internal reflection STED microscopy. *Optics express*, 19(14), 13351-13357.
 - **Gracias, K. S.**, McKillip, J. L. (2004). A review of conventional detection and enumeration methods for pathogenic bacteria in food. *Canadian journal of microbiology*, 50(11), 883-890.
 - **Harris, T. D.**, Buzby, P. R., Babcock, H., Beer, E., Bowers, J., Braslavsky, I., Giladi, E. (2008). Single-molecule DNA sequencing of a viral genome. *Science*, 320(5872), 106-109.
 - **Huawei, W.**, Fang, L., & Daoying, W. (2013). Microflora changes of chilled duck at processing plants. Journal of Nanjing Agricultural University.
 - Human Microbiome Project Consortium. (2012). A framework for human microbiome research. *Nature*, 486(7402), 215-221.
 - **Huse, S. M.**, Huber, J. A., Morrison, H. G., Sogin, M. L., & Welch, D. M. (2007). Accuracy and quality of massively parallel DNA pyrosequencing. *Genome biology*, 8(7), R143.
 - **Huson, D. H.**, Mitra, S., Ruscheweyh, H. J., Weber, N., Schuster, S. C. (2011). Integrative analysis of environmental sequences using MEGAN4. *Genome research*, 21(9), 1552-1560.
 - **Jain, M.**, Fiddes, I. T., Miga, K. H., Olsen, H. E., Paten, B., & Akeson, M. (2015). Improved data analysis for the MinION nanopore sequencer. *Nature methods*, 12(4), 351-356.
 - **Jany, J. L.**, & Barbier, G. (2008). Culture-independent methods for identifying microbial communities in cheese. *Food Microbiology*, 25(7), 839-848.
 - **Jay, J. M.**, Loessner, M. J., & Golden, D. A. (2009). *Microbiologia degli alimenti*. Springer Science & Business Media.
 - **Johnsen, J. M.**, Nickerson, D. A., & Reiner, A. P. (2013). Massively parallel sequencing: the new frontier of hematologic genomics. *Blood*, 122(19), 3268-3275.
 - Jumpstart Consortium Human Microbiome Project Data Generation Working Group. (2012). Evaluation of 16S rDNA-based community profiling for human microbiome research. *PLoS ONE* 7:e39315 10.1371/journal.pone.
 - **Kanehisa, M.**, Goto, S., Kawashima, S., & Nakaya, A. (2002). The KEGG databases at GenomeNet. *Nucleic acids research*, 30(1), 42-46.

- **Kelley, D. R.**, Liu, B., Delcher, A. L., Pop, M., Salzberg, S. L. (2011). Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering. *Nucleic acids research*, 40(1), e9-e9.
- **Kent, W. J.** (2002). BLAT—the BLAST-like alignment tool. *Genome research*, 12(4), 656-664.
- **Kiarie, E.**, Romero, L. F., Nyachoti, C. M. (2013). The role of added feed enzymes in promoting gut health in swine and poultry. *Nutrition research reviews*, 26(1), 71-88.
- **Koboldt, D. C.**, Steinberg, K. M., Larson, D. E., Wilson, R. K., & Mardis, E. R. (2013). The next-generation sequencing revolution and its impact on genomics. *Cell*, 155(1), 27-38.
- **Kuehn, J. S.**, Gorden, P. J., Munro, D., Rong, R., Dong, Q., Plummer, P. J., Phillips, G. J. (2013). Bacterial community profiling of milk samples as a means to understand culture-negative bovine clinical mastitis. *PLoS One*, 8(4), e61959.
- **Langille, M. G.**, Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., Beiko, R. G. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*, 31(9), 814-821.
- **Lapunzina, P.**, López, R. O., Rodríguez-Laguna, L., García-Miguel, P., Martínez, A. R., Martínez-Glez, V. (2014). Impact of NGS in the medical sciences: Genetic syndromes with an increased risk of developing cancer as an example of the use of new technologies. *Genetics and molecular biology*, 37(1), 241-249.
- **Larsen, M. H.**, Dalmasso, M., Ingmer, H., Langsrud, S., Malakauskas, M., Mader, A., Wallace, R. J. (2014). Persistence of foodborne pathogens and their control in primary and secondary food production chains. *Food Control*, 44, 92-109.
- **Leite, A. M. O.**, Mayo, B., Rachid, C. T. C. C., Peixoto, R. S., Silva, J. T., Paschoalin, V. M. F., Delgado, S. (2012). Assessment of the microbial diversity of Brazilian kefir grains by PCR-DGGE and pyrosequencing analysis. *Food microbiology*, 31(2), 215-221.
- **Levene, M. J.**, Korlach, J., Turner, S. W., Foquet, M., Craighead, H. G., & Webb, W. W. (2003). Zero-mode waveguides for single-molecule analysis at high concentrations. *Science*, 299(5607), 682-686.
- **Li, H.**, Homer, N. (2010). A survey of sequence alignment algorithms for next-generation sequencing. *Briefings in bioinformatics*, 11(5), 473-483.
- **Li, H.**, Liu, X., Li, L., Mu, X., Genov, R., & Mason, A. J. (2016). CMOS Electrochemical Instrumentation for Biosensor Microsystems: A Review. *Sensors*, 17(1), 74.

- **Li, J.**, Jia, H., Cai, X., Zhong, H., Feng, Q., Sunagawa, S., Juncker, A. S. (2014). An integrated catalog of reference genes in the human gut microbiome. *Nature biotechnology*, 32(8), 834-841.
- **Lindblom, A.**, & Robinson, P. N. (2011). Bioinformatics for human genetics: promises and challenges. *Human mutation*, 32(5), 495-500.
- **Linnarsson, S.** (2010). Recent advances in DNA sequencing methods—general principles of sample preparation. *Experimental cell research*, 316(8), 1339-1343.
- **Liu, L.**, Li, Y., Li, S., Hu, N., He, Y., Pong, R., Law, M. (2012). Comparison of next-generation sequencing systems. *BioMed Research International*, 2012.
- **Liu, N.**, Ru, Y., Wang, J., Xu, T. (2010). Effect of dietary sodium phytate and microbial phytase on the lipase activity and lipid metabolism of broiler chickens. *British Journal of nutrition*, 103(6), 862-868.
- **Looft, T.**, Johnson, T. A., Allen, H. K., Bayles, D. O., Alt, D. P., Stedtfeld, R. D., Hashsham, S. A. (2012). In-feed antibiotic effects on the swine intestinal microbiome. *Proceedings of the National Academy of Sciences*, 109(5), 1691-1696.
- **Lozupone, C. A.**, Knight, R. (2007). Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences*, 104(27), 11436-11440.
- **Luo, C.**, Tsementzi, D., Kyrpides, N. C., Konstantinidis, K. T. (2012). Individual genome assembly from complex community short-read metagenomic datasets. *The ISME journal*, 6(4), 898-901.
- **Magi, A.**, Benelli, M., Gozzini, A., Girolami, F., Torricelli, F., & Brandi, M. L. (2010). Bioinformatics for next generation sequencing data. *Genes*, 1(2), 294-307.
- **Magi, A.**, Tattini, L., Cifola, I., D’Aurizio, R., Benelli, M., Mangano, E., Magini, P. (2013). EXCAVATOR: detecting copy number variants from whole-exome sequencing data. *Genome biology*, 14(10), R120.
- **Manichanh, C.**, Chapple, C. E., Frangeul, L., Gloux, K., Guigo, R., & Dore, J. (2008). A comparison of random sequence reads versus 16S rDNA sequences for estimating the biodiversity of a metagenomic library. *Nucleic Acids Research*, 36(16), 5180-5188.
- **Marchesi, C.** (2014). Identificazione di nuove epossido idrolasi da librerie metagenomiche e loro espressione in forma ricombinante.
- **Marchiori, D.** (2016). Implementazione e analisi di metodi per la scoperta di k-mers discriminativi nella classificazione di sequenze metagenomiche.
- **Mardis, E. R.** (2008). The impact of next-generation sequencing technology on genetics. *Trends in genetics*, 24(3), 133-141.

- **Margulies, M.**, Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bembien, L. A., Dewell, S. B. (2006). Corrigendum: Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 441(7089), 120-121.
- **Markowitz, V. M.**, Chen, I. M. A., Chu, K., Szeto, E., Palaniappan, K., Pillay, M., Huntemann, M. (2013). IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Research*, 42(D1), D568-D573.
- **Masoud, W.**, Vogensen, F. K., Lillevang, S., Al-Soud, W. A., Sørensen, S. J., & Jakobsen, M. (2012). The fate of indigenous microbiota, starter cultures, *Escherichia coli*, *Listeria innocua* and *Staphylococcus aureus* in Danish raw milk and cheeses determined by pyrosequencing and quantitative real time (qRT)-PCR. *International journal of food microbiology*, 153(1), 192-202.
- **Mayo, B.**, TCC Rachid, C., Alegría, Á., MO Leite, A., S Peixoto, R., & Delgado, S. (2014). Impact of next generation sequencing techniques in food microbiology. *Current genomics*, 15(4), 293-309.
- **McElhoe, J. A.**, Holland, M. M., Makova, K. D., Su, M. S. W., Paul, I. M., Baker, C. H., Young, B. (2014). Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Science International: Genetics*, 13, 20-29.
- **Melle, F.** (2016). *Il sequenziamento massivo dell'esoma rivela nuovi target molecolari nella neoplasia blastica delle cellule dendritiche plasmocitoidi (BPDCN)* (Doctoral dissertation, alma).
- **Mende, D. R.**, Waller, A. S., Sunagawa, S., Järvelin, A. I., Chan, M. M., Arumugam, M., Bork, P. (2012). Assessment of metagenomic assembly using simulated next generation sequencing data. *PloS one*, 7(2), e31386.
- **Merriman, B.**, Torrent, I., Rothberg, J. M., & R&D Team. (2012). Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 33(23), 3397-3417.
- **Metzker, M. L.** (2010). Sequencing technologies—the next generation. *Nature reviews genetics*, 11(1), 31-46.
- **Meyer, F.**, Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M. (2011). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics. Mathematics and Computer Science Division, Argonne National Laboratory*, 9700.
- **Meyer, F.**, Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M. (2011). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional

- analysis of metagenomes. *BMC Bioinformatics. Mathematics and Computer Science Division, Argonne National Laboratory, 9700.*
- **Meyer, F.**, Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Wilkening, J. (2008). The metagenomics RAST server—a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC bioinformatics*, 9(1), 386.
 - **Mikheyev, A. S.**, Tin, M. M. (2014). A first look at the Oxford Nanopore MinION sequencer. *Molecular ecology resources*, 14(6), 1097-1102.
 - **Morgan, X. C.**, Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., Bousvaros, A. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome biology*, 13(9), R79.
 - **Mori, A.**, Deola, S., Xumerle, L., Mijatovic, V., Malerba, G., & Monsurrò, V. (2013). Next generation sequencing: new tools in immunology and hematology. *Blood research*, 48(4), 242-249.
 - **Morozova, O.**, Hirst, M., & Marra, M. A. (2009). Applications of new sequencing technologies for transcriptome analysis. *Annual review of genomics and human genetics*, 10, 135-151.
 - **Myllykangas, S.**, Buenrostro, J., & Ji, H. P. (2012). Overview of sequencing technology platforms. In *Bioinformatics for high throughput sequencing* (pp. 11-25). Springer New York.
 - **Nacke, H.**, Engelhaupt, M., Brady, S., Fischer, C., Tautzt, J., Daniel, R. (2012). Identification and characterization of novel cellulolytic and hemicellulolytic genes and enzymes derived from German grassland soil metagenomes. *Biotechnology letters*, 34(4), 663-675.
 - **Nagalakshmi, U.**, Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, 320(5881), 1344-1349.
 - **Nakano, M.**, Komatsu, J., Matsuura, S. I., Takashima, K., Katsura, S., & Mizuno, A. (2003). Single-molecule PCR using water-in-oil emulsion. *Journal of biotechnology*, 102(2), 117-124.
 - **Nelson, D. L.**, & Cox, M. M. (2002). I principi di biochimica di Lehninger (sesta edizione).
 - **Nelson, D. L.**, & Cox, M. M. (2015). *Introduzione alla biochimica di Lehninger*. Zanichelli.
 - **Nguyen, L.**, & Burnett, L. (2014). Automation of molecular-based analyses: a primer on massively parallel sequencing. *The Clinical Biochemist Reviews*, 35(3), 169.
 - **Nocker, A.**, Cheung, C. Y., & Camper, A. K. (2006). Comparison of propidium monoazide with ethidium monoazide for differentiation of live vs. dead bacteria by selective removal of DNA from dead cells. *Journal of microbiological methods*, 67(2), 310-320.

- **Noguchi, H.**, Park, J., Takagi, T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic acids research*, 34(19), 5623-5630.
- **Nugen, S. R.**, Baeumner, A. J. (2008). Trends and opportunities in food pathogen detection. *Analytical and bioanalytical chemistry*, 391(2), 451.
- **O'Hara, A. M.**, & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO reports*, 7(7), 688-693.
- **Oulas, A.**, Pavludi, C., Polymenakou, P., Pavlopoulos, G. A., Papanikolaou, N., Kotoulas, G., Iliopoulos, I. (2015). Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinformatics and Biology insights*, 9, 75.
- **Overbeek, R.**, Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Vonstein, V. (2013). The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*, 42(D1), D206-D214.
- **Park, E. J.**, Kim, K. H., Abell, G. C., Kim, M. S., Roh, S. W., & Bae, J. W. (2011). Metagenomic analysis of the viral communities in fermented foods. *Applied and environmental microbiology*, 77(4), 1284-1291.
- **Parks, D. H.**, Tyson, G. W., Hugenholtz, P., Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 30(21), 3123-3124.
- **Pepe, G.**, Giusti, B., Attanasio, M., Comeglio, P., Porciani, M. C., Giurlani, L., Abbate, R. (1997). A major involvement of the cardiovascular system in patients affected by Marfan syndrome: novel mutations in fibrillin 1 gene. *Journal of molecular and cellular cardiology*, 29(7), 1877-1884.
- **Pepe, G.**, Giusti, B., Evangelisti, L., Porciani, M. C., Brunelli, T., Giurlani, L., Abbate, R. (2001). Fibrillin-1 (FBN1) gene frameshift mutations in Marfan patients: genotype-phenotype correlation. *Clinical genetics*, 59(6), 444-450.
- **Pepe, G.**, Nistri, S., Giusti, B., Sticchi, E., Attanasio, M., Porciani, C., Gensini, G. F. (2014). Identification of fibrillin 1 gene mutations in patients with bicuspid aortic valve (BAV) without Marfan syndrome. *BMC medical genetics*, 15(1), 23.
- **Petersen, T. N.**, Lukjancenko, O., Thomsen, M. C. F., Sperotto, M. M., Lund, O., Aarestrup, F. M., & Sicheritz-Pontén, T. (2017). MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads. *PloS one*, 12(5), e0176469.
- **Pinto, A. J.**, Raskin, L. (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PloS one*, 7(8), e43093.

- **Pippucci, T.**, Magi, A., Gialluisi, A., & Romeo, G. (2014). Detection of runs of homozygosity from whole exome sequencing data: state of the art and perspectives for clinical, population and epidemiological studies. *Human heredity*, 77(1-4), 63-72.
- **Pregolato, M.** (2010). Biodiversità nel Corpo Fisico Umano. Le nuove frontiere della Metagenomica e del Quorum Sensing.
- **Pricope, L.**, Nicolau, A., Wagner, M., & Rychli, K. (2013). The effect of sublethal concentrations of benzalkonium chloride on invasiveness and intracellular proliferation of *Listeria monocytogenes*. *Food Control*, 31(1), 230-235.
- **Qin, J.**, Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., Mende, D. R. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *nature*, 464(7285), 59-65.
- **Quail, M. A.**, Kozarewa, I., Smith, F., Scally, A., Stephens, P. J., Durbin, R., Turner, D. J. (2008). A large genome center's improvements to the Illumina sequencing system. *Nature methods*, 5(12), 1005-1010.
- **Quast, C.**, Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41(D1), D590-D596.
- **Quick, J.**, Ashton, P., Calus, S., Chatt, C., Gossain, S., Hawker, J., De Pinna, E. (2015). Rapid draft sequencing and real-time nanopore sequencing in a hospital outbreak of *Salmonella*. *Genome biology*, 16(1), 114.
- **Quick, J.**, Loman, N. J., Duraffour, S., Simpson, J. T., Severi, E., Cowley, L., Ouédraogo, N. (2016). Real-time, portable genome sequencing for Ebola surveillance. *Nature*, 530(7589), 228.
- **Quick, J.**, Quinlan, A. R., & Loman, N. J. (2014). A reference bacterial genome dataset generated on the MinION™ portable single-molecule nanopore sequencer. *Gigascience*, 3(1), 22.
- **Quigley, L.**, O'Sullivan, O., Beresford, T. P., Ross, R. P., Fitzgerald, G. F., Cotter, P. D. (2012). High-throughput sequencing for detection of subpopulations of bacteria not previously associated with artisanal cheeses. *Applied and environmental microbiology*, 78(16), 5717-5723.
- **Rabbani, B.**, Tekin, M., & Mahdieh, N. (2014). The promise of whole-exome sequencing in medical genetics. *Journal of human genetics*, 59(1), 5-15.

- **Rachid, C. T.**, Santos, A. L., Piccolo, M. C., Balieiro, F. C., Coutinho, H. L., Peixoto, R. S., Rosado, A. S. (2013). Effect of sugarcane burning or green harvest methods on the Brazilian Cerrado soil bacterial community structure. *PLoS One*, 8(3), e59342.
- **Rappé, M. S.**, & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Reviews in Microbiology*, 57(1), 369-394.
- **Rhoads, A.**, & Au, K. F. (2015). PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*, 13(5), 278-289.
- **Robinson, R. K.** (2014). *Encyclopedia of food microbiology*. Academic press.
- **Rodehutschord, M.**, Pfeffer, E. (1995). Effects of supplemental microbial phytase on phosphorus digestibility and utilization in rainbow trout (*Oncorhynchus mykiss*). *Water Science and Technology*, 31(10), 143-147.
- **Rondon, M. R.**, August, P. R., Bettermann, A. D., Brady, S. F., Grossman, T. H., Liles, M. R., Tiong, C. L. (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and environmental microbiology*, 66(6), 2541-2547.
- **Rossau, R.**, Van Landschoot, A., Gillis, M., De Ley, J. (1991). Taxonomy of Moraxellaceae fam. nov., a new bacterial family to accommodate the genera Moraxella, Acinetobacter, and Psychrobacter and related organisms. *International Journal of Systematic and Evolutionary Microbiology*, 41(2), 310-319.
- **Rothberg, J. M.**, Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., ... & Hoon, J. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature*, 475(7356), 348.
- **Ruby, J. G.**, Bellare, P., DeRisi, J. L. (2013). PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3: Genes, Genomes, Genetics*, 3(5), 865-880.
- **Rusch, D. B.**, Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., Beeson, K. (2007). The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS biology*, 5(3), e77.
- Sandberg, A. S., & Andlid, T. (2002). Phytogenic and microbial phytases in human nutrition. *International journal of food science & technology*, 37(7), 823-833. Stahl, C.A.,
- **Sandberg, A. S.**, Andlid, T. (2002). Phytogenic and microbial phytases in human nutrition. *International journal of food science & technology*, 37(7), 823-833.
- **Sanger, F.**, Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the national academy of sciences*, 74(12), 5463-5467.

- **Schloss, P. D.**, Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... & Sahl, J. W. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and environmental microbiology*, 75(23), 7537-7541.
- **Schmidtke, J.**, & Cassiman, J. J. (2010). The EuroGentest clinical utility gene cards. *European Journal of Human Genetics*, 18(9).
- **Scholz, M. B.**, Lo, C. C., & Chain, P. S. (2012). Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Current opinion in biotechnology*, 23(1), 9-15.
- **Selle, P. H.**, Ravindran, V. (2007). Microbial phytase in poultry nutrition. *Animal Feed Science and Technology*, 135(1), 1-41.
- **Shah, N.**, Tang, H., Doak, T. G., Ye, Y. (2011). Comparing bacterial communities inferred from 16S rRNA gene sequencing and shotgun metagenomics. In *Biocomputing 2011* (pp. 165-176).
- **Sharpton, T. J.** (2014). An introduction to the analysis of shotgun metagenomic data. *Frontiers in plant science*, 5.
- **Sharpton, T. J.**, Riesenfeld, S. J., Kembel, S. W., Ladau, J., O'Dwyer, J. P., Green, J. L., Pollard, K. S. (2011). PhylOTU: a high-throughput procedure quantifies microbial community diversity and resolves novel taxa from metagenomic data. *PLoS computational biology*, 7(1), e1001061.
- **Shendure, J.**, & Ji, H. (2008). Next-generation DNA sequencing. *Nature biotechnology*, 26(10), 1135.
- **Shendure, J.**, Porreca, G. J., Reppas, N. B., Lin, X., McCutcheon, J. P., Rosenbaum, A. M., Church, G. M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*, 309(5741), 1728-1732.
- **Simell, M.**, Turunen, M., Piironen, J., Vaara, T. (1989). Feed and food applications of phytase. *Lecture-3rd Meet. Industrial Applications of Enzymes, Barcelona (Spain)*.
- **Sinden, R. R.** (2012). *DNA structure and function*. Elsevier.
- **Singh, P. S.** (2015). From Sensors to Systems: CMOS-Integrated Electrochemical Biosensors. *IEEE Access*, 3, 249-259.
- **Sipos, R.**, Székely, A., Révész, S., & Márialigeti, K. (2010). Addressing PCR biases in environmental microbiology studies. *Bioremediation: Methods and Protocols*, 37-58.

- **Smith, M. I.**, Yatsunenkov, T., Manary, M. J., Trehan, I., Mkakosya, R., Cheng, J., Liu, J. (2013). Gut microbiomes of Malawian twin pairs discordant for kwashiorkor. *Science*, 339(6119), 548-554.
- **Solieri, L.**, Dakal, T. C., & Giudici, P. (2013). Next-generation sequencing and its potential impact on food microbial genomics. *Annals of microbiology*, 63(1), 21-37.
- **Srinivasan, S.**, Batra, J. (2014). Four generations of sequencing: Is it ready for the clinic yet?. *Journal of Next Generation Sequencing & Applications*, 1(107).
- **Stahl, C. H.**, Roneker, K. R., Thornton, J. R., & Lei, X. G. (2000). A new phytase expressed in yeast effectively improves the bioavailability of phytate phosphorus to weanling pigs. *Journal of animal science*, 78(3), 668-674.
- **Stein, L. D.** (2010). The case for cloud computing in genome informatics. *Genome biology*, 11(5), 207.
- **Sul, W. J.**, Cole, J. R., Jesus, E. D. C., Wang, Q., Farris, R. J., Fish, J. A., & Tiedje, J. M. (2011). Bacterial community comparisons by taxonomy-supervised analysis independent of sequence alignment and clustering. *Proceedings of the National Academy of Sciences*, 108(35), 14637-14642.
- **Taber, K. A. J.**, Dickinson, B. D., & Wilson, M. (2014). The promise and challenges of next-generation genome sequencing for clinical care. *JAMA internal medicine*, 174(2), 275-280.
- **Thudi, M.**, Li, Y., Jackson, S. A., May, G. D., & Varshney, R. K. (2012). Current state-of-art of sequencing technologies for plant genomics research. *Briefings in Functional Genomics*, 11(1), 3-11.
- **Timp, W.**, Mirsaidov, U. M., Wang, D., Comer, J., Aksimentiev, A., & Timp, G. (2010). Nanopore sequencing: electrical measurements of the code of life. *IEEE transactions on nanotechnology*, 9(3), 281-294.
- **Tomasetti, C.**, Li, L., & Vogelstein, B. (2017). Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. *Science*, 355(6331), 1330-1334.
- **Toumazou, C.**, & Georgiou, P. (2011). Piet Bergveld-40 years of ISFET technology: From neuronal sensing to DNA sequencing. *Electronics Letters*, 47(26), S7-S12.
- **Toze, S.** (1999). PCR and the detection of microbial pathogens in water and wastewater. *Water Research*, 33(17), 3545-3556.
- **Urban, J. M.**, Bliss, J., Lawrence, C. E., & Gerbi, S. A. (2015). Sequencing ultra-long DNA molecules with the Oxford Nanopore MinION. *bioRxiv*, 019281.

- **Valderrama, W. B.**, Dudley, E. G., Doores, S., & Cutter, C. N. (2016). Commercially available rapid methods for detection of selected food-borne pathogens. *Critical reviews in food science and nutrition*, 56(9), 1519-1531.
- **Van Dijk, E. L.**, Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, 30(9), 418-426.
- **Van Hijum, S. A.**, Vaughan, E. E., & Vogel, R. F. (2013). Application of state-of-art sequencing technologies to indigenous food fermentations. *Current opinion in biotechnology*, 24(2), 178-186.
- **Vielma, J.**, Lall, S. P., Koskela, J., Schöner, F. J., Mattila, P. (1998). Effects of dietary phytase and cholecalciferol on phosphorus bioavailability in rainbow trout (*Oncorhynchus mykiss*). *Aquaculture*, 163(3), 309-323.
- **Vitale, G.** (2013). Analisi critica di tecniche di sequenziamento di nuova generazione.
- **Voelkerding, K. V.**, Dames, S. A., Durtschi, J. D., & Brisci, A. (2010). Nextgeneration sequencing”: dalla ricerca di base alla diagnostica. *Biochim Clin*, 34, 287-303.
- **Wang, Z.**, Liu, X., Yang, B. Z., & Gelernter, J. (2013). The role and challenges of exome sequencing in studies of human diseases. *Frontiers in genetics*, 4.
- **Watson, J. D.**, Baker, T. A., Bell, S. P., Gann, A., Levine, M., & Losicke, R. (2015). *Biologia molecular do gene*. Artmed Editora.
- **Whitman, W. B.**, Coleman, D. C., & Wiebe, W. J. (1998). Prokaryotes: the unseen majority. *Proceedings of the National Academy of Sciences*, 95(12), 6578-6583.
- **Wilmes, P.**, Simmons, S. L., Deneff, V. J., & Banfield, J. F. (2008). The dynamic genetic repertoire of microbial communities. *FEMS microbiology reviews*, 33(1), 109-132.
- **Wood, S. A.**, Smith, K. F., Banks, J. C., Tremblay, L. A., Rhodes, L., Mountfort, D., Pochon, X. (2013). Molecular genetic tools for environmental monitoring of New Zealand's aquatic habitats, past, present and the future. *New Zealand Journal of Marine and Freshwater Research*, 47(1), 90-119.
- **Wooley, J. C.**, Godzik, A., & Friedberg, I. (2010). A primer on metagenomics. *PLoS computational biology*, 6(2), e1000667.
- **Wrighton, K. C.**, Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Long, P. E. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*, 337(6102), 1661-1665.
- **Wrighton, K. C.**, Thomas, B. C., Sharon, I., Miller, C. S., Castelle, C. J., VerBerkmoes, N. C., Long, P. E. (2012). Fermentation, hydrogen, and sulfur metabolism in multiple uncultivated bacterial phyla. *Science*, 337(6102), 1661-1665.

- **Wu, D.**, Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N. N., Hooper, S. D. (2009). A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea. *Nature*, 462(7276), 1056-1060.
- **Wu, M.**, Scott, A. J. (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics*, 28(7), 1033-1034.
- **Wylie, K. M.**, Truty, R. M., Sharpton, T. J., Mihindukulasuriya, K. A., Zhou, Y., Gao, H., Pollard, K. S. (2012). Novel bacterial taxa in the human microbiome. *PloS one*, 7(6), e35294.
- **Yatsunenkov, T.**, Rey, F. E., Manary, M. J., Trehan, I., Dominguez-Bello, M. G., Contreras, M., Heath, A. C. (2012). Human gut microbiome viewed across age and geography. *Nature*, 486(7402), 222-227.
- **Yi, Z.**, Kornegay, E. T., Denbow, D. M. (1996). Supplemental microbial phytase improves zinc utilization in broilers. *Poultry Science*, 75(4), 540-546.
- **Yozwiak, N. L.**, Skewes-Cox, P., Stenglein, M. D., Balmaseda, A., Harris, E., DeRisi, J. L. (2012). Virus identification in unknown tropical febrile illness cases using deep sequencing. *Plos neglected tropical diseases*, 6(2), e1485.
- **Zaltum, O. S. A.** (2017). Mortalità anomala in molluschi bivalvi: approcci diagnostici tradizionali e di metagenomica.
- **Zhang, Z. B.**, Kornegay, E. T., Radcliffe, J. S., Wilson, J. H., Veit, H. P. (2000). Comparison of phytase from genetically engineered *Aspergillus* and canola in weanling pig diets. *Journal of Animal Science*, 78(11), 2868-2878.
- **Zhou, X.**, & Rokas, A. (2014). Prevention, diagnosis and treatment of high-throughput sequencing data pathologies. *Molecular ecology*, 23(7), 1679-1700.
- **Zhu, W.**, Lomsadze, A., & Borodovsky, M. (2010). Ab initio gene identification in metagenomic sequences. *Nucleic acids research*, 38(12), e132-e132.

Sitografia

- <http://biochem.co/wp-content/uploads/2008/08/deoxynucleotide-dideoxynucleotide.png>
- <http://seqanswers.com/forums/showthread.php?t=10>
- <http://seqll.com/technology-information/>
- http://uvmgg.wikia.com/wiki/DNAseq:454_pyrosequencing
- <http://www.454.com>
- <http://www.appliedbiosystems.com>
- http://www.biology.arizona.edu/molecular_bio/problem_sets/nucleic_acids/15t.html
- <http://www.illumina.com/>

- <http://www.newence.com/2017/04/05/nuove-prospettive-nei-meccanismi-cellulari-del-cancro-il-ruolo-degli-errori-da-replicazione-del-dna/>
- <http://www.pacb.com/>
- <https://bredagenetics.com/sequencing-machines-illumina/?lang=it>
- <https://futurism.com/study-shows-why-we-are-made-of-dna-and-not-rna/>
- https://support.illumina.com/sequencing/sequencing_instruments/hiscansq.html
- <https://www.illumina.com/>
- <https://www.khanacademy.org/science/biology/biotech-dna-technology/dna-sequencing-pcr-electrophoresis/a/dna-sequencing>
- <https://www.nature.com/scitable/topicpage/the-order-of-nucleotides-in-a-gene-6525806>
- <https://www.slideshare.net/MadsAlbertsen/20131202-mads-albertsen-extracting-genomes-from-metagenomes>
- <https://www.thermofisher.com/it/en/home/brands/product-brand/5500-series-genetic-analysis-systems.html>
- <https://www2.le.ac.uk/projects/vgec/highereducation/topics/dna-genes-chromosomes>

Sono sul solito treno, in una fredda mattinata di inizio dicembre, stesso treno che da diversi mesi, tutte le mattine mi porta ad Ozzano. Ovviamente sono in ritardo, devo imparare la presentazione e finire di sistemare la grafica, ma ho deciso comunque di dedicare questo viaggio ai ringraziamenti.

Era l'estate del 2012, mi trovavo in Sicilia, ospite a casa di Marco, avvolto in un clima fantastico. Un clima che faticherò a ritrovare in vita, lì ho deciso di iscrivermi all'università. Il mio passato mi dava torto, non ho mai brillato come studente e stavo per intraprendere una direzione che avrebbe condizionato il resto della mia vita.

Ho deciso, non sarebbero stati 3, ma assolutamente 5 anni. Al pit-sto della triennale ero più convinto che mai che avrei voluto continuare, ora, non vorrei più smettere di studiare.

Non ho mai azzeccato le scelte giuste da fare nella vita, è vero che l'adolescenza è un momento difficile, ma io ho deciso, senza scusanti. Quello che posso dire ora, è che con l'iscrivermi all'università e smettere di fumare, sono state delle scelte azzeccate. Quello che provo ora è indescrivibile, sono a pochi giorni dal completare un percorso lungo e complesso, non rimpiango nulla di questi ultimi 5 anni.

Innanzitutto vorrei ringraziare il team di sicurezza alimentare del DISTAL di Ozzano, coordinato dal prof. Manfreda, che insieme ai suoi collaboratori, Alessandra, Alex, Frédérique, Pietro ed infine Paola, mi hanno fornito un bagaglio di conoscenze molto ampio.

Grazie alla mia famiglia, genitori, fratelli, nonni e zii, senza di loro non mi sarebbe stato possibile tutto questo, grazie veramente. Un pensiero speciale ai nonni che sono venuti a mancare in questi anni, sono sicuro che ora sarebbero fieri di me.

Grazie alla famiglia di Giulia, siete stati un supporto molto importante in tanti sensi.

Grazie a tutti i miei amici, spaziando da Forlì a Cesena, da Ravenna a Riccione. Il 12 saremo tutti insieme a festeggiare, quasi tutti.

Grazie a Giulia, senza ombra di dubbio la persona più importante del mio presente, e lo dico sfidando la superstizione, anche del mio futuro. Insieme stiamo completando questo percorso, insieme è stato possibile completarlo. Sono convinto che i risultati che sto ottenendo senza di lei non sarebbero stati così fantastici.