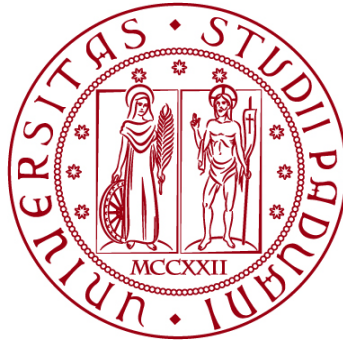


**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI BIOLOGIA**

**Corso di Laurea magistrale in Biologia Sanitaria**



**TESI DI LAUREA**

**STUDIO PRELIMINARE SULLA CARATTERIZZAZIONE DEL  
MICROBIOMA CUTANEO IN AMBITO FORENSE MEDIANTE  
TECNICHE DI NEXT GENERATION SEQUENCING**

**Relatore: Prof. Giovanni Vazza  
Dipartimento di Biologia**

**Correlatore: Magg. inv. sc. Christian Faccineto, PhD**

**Laureanda: Veronica Dal Checco**

**ANNO ACCADEMICO 2022/2023**



## INDICE

ABSTRACT.....	5
1. INTRODUZIONE.....	6
1.1. LE SCIENZE FORENSI	
1.2. LA GENETICA FORENSE	
1.3. MICROBIOLOGIA FORENSE	
1.4. EVOLUZIONE DEL SEQUENZIAMENTO	
1.5. PANNELLI NGS PER L'ANALISI DEL MICROBIOMA	
1.6. SOFTWARE PER ANALISI SECONDARIA	
2. SCOPO DEL LAVORO.....	23
3. MATERIALI E METODI.....	24
3.1. RACCOLTA E CONSERVAZIONE DEI CAMPIONI	
3.2. ESTRAZIONE	
3.3. QUANTIFICAZIONE	
3.4. SEQUENZIAMENTO MEDIANTE NEXT GENERATION SEQUENCING: TECNOLOGIA ION TORRENT	
3.5. SEQUENZIAMENTO MEDIANTE NEXT GENERATION SEQUENCING: TECNOLOGIA ILLUMINA	
3.6. ANALISI STATISTICA CON R-STUDIO	
4. RISULTATI E DISCUSSIONE .....	46
4.1. RICERCA DEL CORE MICROBIOME	
4.2. VERIFICA VARIABILITÀ INTRAINDIVIDUALE	
4.3. VERIFICA VARIABILITÀ INTERINDIVIDUALE	
4.4. RICERCA CORRISPONDENZA TRA TRACCIA E DONATORE	
4.5. ANALISI STATISTICA DEI RISULTATI	
5. CONCLUSIONI.....	71
6. BIBLIOGRAFIA.....	76
7. ALLEGATI.....	78



## ABSTRACT

Innovativi studi in campo forense hanno recentemente iniziato ad occuparsi dell'utilizzo del microbioma come strumento aggiuntivo alle analisi classiche del DNA. Il microbioma umano è estremamente variabile in relazione alla sua funzione e al sito corporeo dal quale deriva; si parla quindi di microbioma cutaneo, intestinale, del cavo orale, del tratto respiratorio, dei capelli, etc.

Le sue possibili applicazioni sono diverse e possono riguardare: l'identificazione personale, ovvero la capacità di ricondurre una traccia al proprio donatore; la geolocalizzazione, ovvero la possibilità di tracciare spostamenti di soggetti grazie alla commistione di microbioma umano e ambientale specifico di determinate aree geografiche; la stima dell'intervallo post-mortem e la previsione delle caratteristiche del donatore (età, sesso e origine biogeografica).

Questo lavoro si è focalizzato sulla raccolta e sullo studio del microbioma cutaneo a scopo forense. I campioni sono stati prelevati dal palmo delle mani e dal collo dei donatori e, una volta caratterizzati, sono stati confrontati tra loro e con i dati riportati in letteratura. In aggiunta, si è provato ad analizzare oggetti venuti a contatto con i distretti cutanei analizzati (i.e. colletto della camicia e volante dell'automobile) per verificare la fattibilità di individuare la compatibilità tra oggetti e abituali utilizzatori. Il microbioma è stato sequenziato tramite Next Generation Sequencing e analizzato con software statistici. Questo è un lavoro preliminare effettuato come studio pilota per valutare l'applicabilità di questa metodica in ambito forense.

## 1. INTRODUZIONE

### 1.1. LE SCIENZE FORENSI

Le Scienze Forensi sono discipline basate su metodi scientifici che hanno la funzione di guidare la risoluzione di questioni legali. Nel corso degli anni, molte attività di indagine criminalistica si sono evolute a formare vere e proprie specialità indipendenti, ognuna con il relativo personale specializzato (Ricci, 2016). Tali discipline comprendono la dattiloscopia, la balistica, l'audio analisi, l'informatica, l'antropologia, la chimica, l'entomologia, la botanica e la genetica. Di queste, la genetica, l'antropologia fisica e la dattiloscopia, vengono utilizzate anche a scopo identificativo in quanto capaci di condurre ad una identificazione personale (Ricci, 2016). Tutte le indagini forensi si basano sul Principio di Locard, sintetizzabile con l'affermazione: "*Ogni contatto lascia una traccia.*" Questo permette agli investigatori di collegare una persona ad un determinato scenario grazie a prove associative come impronte digitali, fluidi corporei, capelli e fibre.

### 1.2. LA GENETICA FORENSE

La Genetica Forense si avvale di moderne tecniche analitiche per risalire all'origine biologica di una traccia con lo scopo di identificare autori di reati criminosi (delitti, violenze sessuali, furti ecc.), resti umani o vittime di disastri di massa, ma anche di studiare i rapporti di parentela in controversie civili. Questa disciplina, oggi, occupa una posizione centrale in ambito forense conseguita a seguito dei numerosi studi che l'hanno vista protagonista negli ultimi decenni ma le sue radici sono molto più antiche. Il primo polimorfismo genetico è stato scoperto nei primi anni del 1900 da Landstainer e consisteva nel sistema del gruppo sanguigno umano ABO ma ha presto rivelato la sua scarsa efficacia nella discriminazione interpersonale a causa della ridotta variabilità e dell'elevata frequenza all'interno della popolazione (Tagliabracci, 2010). Il primo approccio all'identificazione personale è stato, in seguito, sviluppato da Alec Jeffreys quando, nel 1985, risolse il primo caso giudiziario sfruttando i polimorfismi di lunghezza dei frammenti di restrizione (*Restriction Fragment Length Polymorphism o RFLP*) con la tecnica del *DNA fingerprint*. Qualche anno più tardi questa metodica è stata scalzata dal *DNA profiling* basato sull'amplificazione di marcatori microsatelliti (Edwards, *et al.*, 1991).

#### 2.1.1. SHORT TANDEM REPEATS (STR)

Lo scopo dell'analisi genetica in ambito forense è di distinguere un individuo dagli altri sfruttando gli elementi di variabilità genetica interindividuale che costituiscono circa lo 0.1% del nostro genoma (Tagliabracci, 2010). I marcatori maggiormente utilizzati nelle analisi forensi consistono nei polimorfismi di lunghezza definiti come *Short Tandem Repeats* (STRs), in particolare i microsatelliti tetranucleotidici,

la cui iper-variabilità conferisce un elevato potere discriminatorio interindividuale. Gli STRs vengono eletti marcatori più funzionali per semplicità e sensibilità dei test, elevato grado di polimorfismo, informatività, segregazione indipendente, basso tasso di mutazione, ridotta lunghezza nucleotidica e limitati fenomeni di stutter. Questi polimorfismi sono costituiti da quattro nucleotidi che si ripetono in tandem un numero variabile di volte, affiancati da regioni conservate (Tagliabracci, 2010).

Negli Stati Uniti, a partire dal 1996, il laboratorio dell'FBI ha individuato 13 loci STRs da includere nel database nazionale del DNA noto come CODIS (Combined DNA Index System); ai quali ne sono stati, successivamente, aggiunti ulteriori 7 appartenenti all'*European Standard Set of Loci* (ESSL) (Butler, 2012). Attualmente, i kit di ultima generazione per le analisi forensi permettono di analizzare contemporaneamente 24 (ABI, Promega, Qiagen) e 27 (Promega) loci STR.

#### 2.1.2. ALTRI MARCATORI GENETICI (SNP, Y-marker, mtDNA)

Come supporto alle analisi forensi standard o in casi complessi in cui i marcatori STR non portano a risultati utili vengono utilizzati altri marcatori genetici.

Gli SNP sono polimorfismi di sequenza che differiscono per un unico nucleotide distribuiti lungo tutto il genoma. Forniscono numerosi vantaggi, tra cui le dimensioni ridotte (<100 bp), utili quando il DNA è degradato, la possibilità di analisi in multiplex e l'assenza di artefatti *stutter* (scivolamento della DNA polimerasi sul DNA stampo che genera ampliconi più corti delle aspettative) (Butler, 2007). Tuttavia, il tasso di mutazione dei marcatori SNPs risulta pari a  $10^{-8}$  rispetto ai marcatori STRs che risulta pari a  $10^{-4}$ ; ne consegue che sia necessario un numero maggiore di loci per eguagliare il potere discriminativo degli STRs (Butler, 2012).

Nel nostro genoma sono presenti milioni di SNPs ma quelli impiegati in campo forense vengono generalmente raggruppati in quattro categorie sulla base della loro funzione: *Identity SNPs*, *Lineage SNPs*, *Ancestry SNPs* e *Phenotype SNPs*.

Altri marcatori utili risiedono sul cromosoma Y, tra questi, i più abbondanti sono gli Y-STRs e gli Y-SNPs. L'insieme delle mutazioni Y-STRs costituisce un aplotipo ereditato esclusivamente da tutti i discendenti maschi in linea paterna. L'insieme degli aplotipi comuni tra individui forma un aplogruppo che consente di inferire un'ancestralità biogeografica (Caramelli, 2009).

Anche il DNA contenuto nei mitocondri (mtDNA), fornisce marcatori utilizzabili nelle analisi forensi. Si tratta di una molecola circolare a doppia elica costituita da due filamenti complementari definiti *H strand* e *L strand* che contengono una regione *D-loop* caratterizzata da tre segmenti ipervariabili denominati HVR1, HVR2 e HVR3 (HVR= High Variable Region) (Caramelli, 2009). L'aplotipo del mtDNA viene ereditato in linea materna da discendenti maschi e femmine ed è

presente in un elevato numero di copie all'interno della cellula (Butler, 2012). Questi marcatori vengono utilizzati per accertamenti di parentela, per ottenere informazioni quando il materiale biologico di partenza è degradato o limitato (reperti antichi, reperti ossei e formazioni pilifere prive di radice) e per la diagnosi di specie.

### 1.3. MICROBIOLOGIA FORENSE

#### 1.3.1. I BATTERI

I batteri sono microrganismi unicellulari le cui dimensioni si aggirano tra i millimetri e i micrometri. Sono costituiti da una soluzione acquosa gelificata (citoplasma), delimitata da una membrana cellulare composta da fosfolipidi e proteine di trasporto (*Inner Membrane*) e, talvolta, da una seconda membrana costituita da fosfolipidi e lipopolisaccaridi (*Outer Membrane*). Tra *Inner* e *Outer Membrane* è situata una resistente parete cellulare, altamente permeabile agli ioni e alle molecole organiche, e caratterizzata dalla presenza di peptidoglicano, un polimero organico che costituisce quasi l'1% dell'intera massa cellulare (Figura 1). Queste tre componenti vanno a costituire l'*envelope*, una struttura, differente tra batteri Gram positivi e Gram negativi, che ne consente la distinzione. All'interno della cellula, il materiale genetico, contenuto in un unico cromosoma quasi completamente codificante, è situato nel citoplasma, perché le cellule batteriche sono prive di nucleo, ed è organizzato come un sistema di eliche super avvolte denominato nucleoide (Figura 1). Il punto medio del cromosoma è ancorato all'*envelope* e costituisce l'origine di replicazione durante la riproduzione.

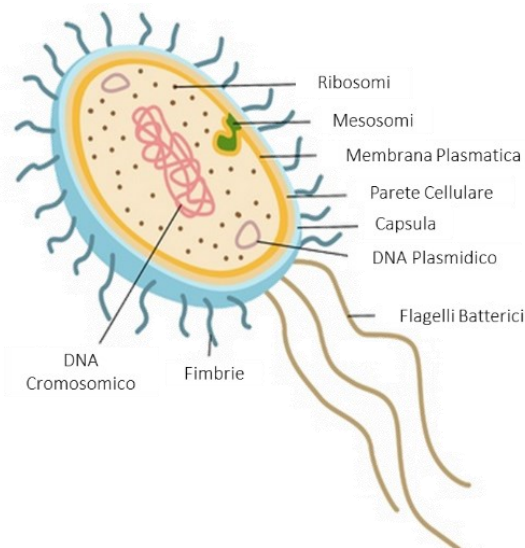


Figura 1. Struttura cellula procariotica e organizzazione del cromosoma batterico.  
(<https://www.gettyimages.it/immagine/cellula-procariotica>)



### 1.3.2. IL GENOMA BATTERICO

Come tutti gli organismi viventi, i microbi sono dotati di materiale genetico, il cui sequenziamento è avvenuto per la prima volta nel 1995 con il metodo Sanger.

L'organizzazione del genoma batterico è semplice ed elegante (Ochman & Davalos, 2006). I batteri possiedono un corredo cromosomico aploide quindi un unico cromosoma circolare della grandezza di 0,6-10 Mb arrangiato tra regioni codificanti ed elementi funzionali. Le regioni codificanti sono necessarie alla sopravvivenza del microrganismo, sono piccole, prive di introni e si susseguono l'un l'altra in modo continuo. La replicazione, come quella eucariotica, è semiconservativa con direzione dei filamenti neosintetizzati 5'-3'. Ha inizio da una regione precisa del DNA ancorata alla parete batterica e utilizza come innesco l'incremento della dimensione cellulare per poi passare sotto il controllo di proteine e metilazione. L'allungamento della catena avviene ad opera della DNA Polimerasi III che ha attività di correzione di bozze in direzione 3'-5' per rimediare ad eventuali errori accumulati durante la polimerizzazione e ridurre il tasso di mutazione.

Uno dei geni fondamentali su cui verrà focalizzata l'attenzione in questa tesi è il *16S ribosomal RNA* (16S rRNA), una delle componenti dell'unità 30S ribosomiale procariotica responsabile del legame con le sequenze di Shine-Dalgarno sul filamento da tradurre. Il gene omonimo viene utilizzato per analisi filogenetiche a partire dal 1977 ad opera di Carl Woese and George E. Fox che per primi ne intuirono le potenzialità. Viene infatti utilizzato come orologio molecolare (*molecular-clock*) per misurare il tempo di divergenza di due o più specie batteriche. Queste analisi basate sul genoma e non più sulla coltura cellulare rientrano nella sfera della metagenomica. I geni che codificano le subunità ribosomiali batteriche (16S, 23S e 5S) vengono inizialmente trascritti come RNA monocistronico e vengono separati solo nella fase post-trascrizionale. La loro evoluzione avviene prevalentemente per accumulo di mutazioni nel corso di trasferimento genico verticale e orizzontale, ovvero per trasferimento di materiale genetico rispettivamente tra cellule discendenti e tra cellule non discendenti.

Il gene batterico 16S (1540 bp), che nel complesso risulta conservato tra le specie batteriche, contiene nove regioni iper-variabili (nominate da V1 a V9) lunghe 30-100bp, coinvolte nella struttura secondaria della piccola subunità ribosomiale. Ognuna di queste ha un diverso grado di conservazione a seconda della specie e del genere ma tutte sono intervallate da regioni altamente conservate che favoriscono la costruzione di primer per l'amplificazione *target*. La sequenza di questo gene è disponibile su database pubblici come il *National Center for Biotechnology Information* (NCBI), *Ribosomal Database Project* (RDP), SILVA e GreenGenes.

Oltre al cromosoma, i batteri sono dotati di elementi mobili che si identificano nei plasmidi e nel genoma di batteriofagi. Questi ultimi codificano proteine che contribuiscono alla fisiologia cellulare: virulenza, capacità di trasferimento genico e resistenza agli antibiotici. I plasmidi sono generalmente piccoli (2 kb - 1,7 Mb),

di forma circolare, super-avvolti, presenti in numero variabile all'interno della medesima cellula e possono essere trasferiti tra specie batteriche, questo ha contribuito fortemente alla spinta evolutiva. I plasmidi presenti in elevato numero di copie garantiscono che tutte le cellule figlie ne abbiano almeno una copia, quelli a basso numero di copie favoriscono il risparmio energetico durante la replicazione. Ognuna di queste molecole ha una propria origine di replicazione, che gli consente di essere indipendente dal DNA cromosomale.

### 1.3.3. MICROBIOMA IN CAMPO FORENSE

Il corpo umano, così come ogni essere vivente e come l'ambiente terrestre, è colonizzato da microrganismi che, per la maggior parte, vivono in maniera simbiotica con l'uomo senza creare danno. L'insieme di tutti questi microrganismi, nella loro totalità o in uno specifico distretto corporeo, è definito microbiota. Il termine microbioma, invece, viene definito per la prima volta nel 1988 come l'intero patrimonio genetico contenuto nei microrganismi appartenenti ad un ambiente definito con proprietà fisico-chimiche distinte (Whipps, *et al.*, 1991).

In ambito clinico, il microbioma è stato ampiamente indagato per il suo ruolo determinante nella salute umana ma, nonostante la lunga storia che ha alle spalle, in ambito forense ha sempre avuto un ruolo marginale.

A partire dalla fine del XIX-XX secolo, la microbiologia veniva applicata da nomi illustri quali L. Pasteur, R. Koch e Locard, pioniere delle scienze forensi, per stabilire la patogenicità e la causa di morte sia in uomo che in animale. In particolare, la prima metà del XX secolo fu un periodo florido per le analisi microbiologiche forensi tanto da meritare menzione dalla scrittrice Agatha Christie nei suoi romanzi gialli (Metcalf, *et al.*, 2017).

Dagli anni '90 si sono registrati i primi utilizzi del microbioma a sostegno delle indagini forensi con l'analisi filogenetica di ceppi appartenenti al virus dell'HIV, poi ammesse come prova in tribunale (State of Louisiana vs. Richard J Schmidt, 1997) negli Stati Uniti (Metzker, *et al.*, 2002); ma solo negli anni 2000, con la paura degli attacchi terroristici biologici, il microbioma ha iniziato ad essere studiato in maniera più intensiva (Robinson, *et al.*, 2021). Nello specifico, dalla curiosità e dall'intuizione che lo studio dei microrganismi residenti all'interno e sulle superfici del corpo umano potesse avere interesse multidisciplinare, nel 2007 nasce "the Human Microbiome Project" (HMP; Institute for Genome Sciences, 2022). Si tratta di un progetto per certi aspetti ispirato al precedente "Human Genome Project" per i genomi umani, creato con lo scopo di individuare tutti i microrganismi appartenenti al microbioma umano e di costruire un database solido dei dati raccolti attraverso il sequenziamento della regione 16S rRNA.

In ambito forense, l'analisi del microbioma potrebbe rivelarsi un'interessante opportunità per ampliare l'elenco dei possibili approcci a supporto delle canoniche profilazioni basate su STR e SNP. In linea teorica, l'opportunità di utilizzare la microbiologia forense potrebbe essere sfruttata quando l'analisi del DNA umano

risulta inconcludente quando si è in presenza di scarse quantità di materiale genetico (*low template DNA*), degradazione del campione o presenza di tracce miste; e quando anche lo studio delle impronte digitali non può essere applicato (Gouello, *et al.*, 2021).

Tra i potenziali vantaggi che le cellule batteriche possono apportare alle indagini forensi troviamo:

- Numerosità.  
I batteri umani sono in numero uguale o maggiore, con un rapporto fino a 1:10, rispetto alle cellule somatiche dell'ospite, questa condizione permette di ottenere un elevato numero di copie di DNA batterico.
- Variabilità.  
Ogni individuo possiede circa 500-1000 specie batteriche differenti, ognuna con il proprio genoma e i propri geni; ne deriva che la diversità genetica del microbioma è potenzialmente molto maggiore rispetto a quella del genoma umano.
- Unicità.  
Il microbioma è variabile da individuo a individuo sia in termini di composizione batterica sia in termini di abbondanza relativa di ciascun taxon. Inoltre, la composizione microbica sembra essere influenzata da caratteristiche proprie dell'ospite come sesso, età, origine biogeografica e stato di salute (Gouello, *et al.*, 2021).

Le possibili applicazioni dello studio del microbioma in questo ambito potrebbero riguardare soprattutto l'identificazione personale, ovvero la capacità di ricondurre una traccia batterica al proprio donatore; la geolocalizzazione, grazie alla commistione tra il microbioma umano e ambientale riconducibile a determinate aree geografiche; la stima dell'intervallo post-mortem e la previsione delle caratteristiche del donatore (Robinson, *et al.*, 2021).

Il microbioma umano può essere prelevato da numerosi siti corporei quali pelle, intestino, cavo orale, tratto respiratorio superiore, regione pubica, capelli, superficie oculare, e fluidi corporei (Neckovic, *et al.*, 2020) tuttavia in questa tesi l'attenzione è stata focalizzata sul microbioma cutaneo.

La cute è il più esteso organo umano il cui ruolo principale è la difesa primaria contro le infezioni causate da microrganismi patogeni e agenti tossici. Si tratta di un ambiente colonizzato da circa  $1 \times 10^7$  microrganismi simbiotici o commensali il cui equilibrio dinamico viene costantemente perturbato da stimoli provenienti dall'interno e dall'esterno del corpo (Edmonds-Wilson, *et al.*, 2015). Il microbioma cutaneo è prevalentemente costituito da batteri Gram positivi, che presentano una spessa parete di peptidoglicano (Figura 1), la quale li rende estremamente resistenti agli agenti perturbanti come i raggi ultravioletti, le variazioni di temperatura e l'umidità (Fierer, *et al.*, 2010).

La presenza di microrganismi cutanei, a composizione unica tra gli individui ed

estremamente abbondanti, potrebbe essere sfruttata in ambito forense per l'analisi del *Touch Microbiome*. Ogni contatto con oggetti o con altri individui favorisce il trasferimento di una traccia microbica specifica e duratura nel tempo attraverso la quale il donatore può essere identificato univocamente (Procopio, *et al.*, 2021).

Inizialmente, con il sequenziamento Sanger, l'analisi del microbioma richiedeva l'allestimento di colture batteriche che consentivano la rilevazione di meno dello 0.001% del microbioma campionato. Con l'avvento del *Massive Parallel Sequencing* (MPS), anche definito *Next Generation Sequencing* (NGS), lo studio del microbioma è diventato più agile perché consente di: ottenere le sequenze di tutto il genoma batterico in parallelo, garantire la possibilità di applicare un unico protocollo per tutti i microrganismi, evitare la fase di clonaggio grazie alla costruzione di librerie e sequenziare direttamente il campione senza bisogno di allestire una coltura riducendo notevolmente costi e tempi.

La tecnologia MPS permette di studiare il microbioma tramite il sequenziamento di regioni genomiche *target*: 16S per i batteri, 18S per i micro-eucarioti e regioni ITS per i funghi; oppure tramite sequenziamento dell'intero genoma (Metcalf, *et al.*, 2017). L'analisi di regioni *target*, metodo di elezione utilizzato dal "Human Microbiome Project", consiste nell'amplificazione del gene selezionato e del successivo sequenziamento alla ricerca di differenze nelle regioni variabili che permettano di discriminare le specie microbiche (Institute for Genome Sciences, 2022).

Per i batteri, il sequenziamento della regione 16S viene usato come *gold standard* per la ricerca in ambito clinico, soprattutto per ceppi aberranti o rari. Per i micro-eucarioti si usa 18S rDNA, omologo di 16S, anch'esso conservato ad eccezione di 9 sequenze ipervariabili presenti all'interno della regione, che, al pari di quelle batteriche, possono essere amplificate selettivamente consentendo la discriminazione tra le specie. Il suo studio risulta piuttosto complesso a causa della presenza di membrane eucariotiche che rendono difficile l'estrazione degli acidi nucleici, mutazioni sulle regioni complementari ai primer di amplificazione e ampliconi di lunghezza variabile (Popovic, *et al.*, 2018). Le regioni ITS fungine (*Internal Transcribed Spacer*) sono due, rispettivamente ITS1 e ITS2, ubiquitarie e altamente variabili in termini di lunghezza e sequenza. Vengono utilizzate soprattutto per l'analisi di funghi nei quali le sequenze 18S non risultano sufficientemente variabili.

In alternativa al sequenziamento *target*, la tecnologia MPS permette di individuare SNP distribuiti lungo tutto il DNA batterico con un sequenziamento *shotgun*, ottenendo una maggiore risoluzione delle specie e dei ceppi batterici (Schmedes, *et al.*, 2017). Un'ulteriore applicazione di questa tecnologia di più recente sviluppo, volta allo studio del microbioma, consiste nell'effettuare un sequenziamento specifico delle regioni CRISPR batteriche che, essendo variabili tra le specie batteriche, favorirebbero la discriminazione microbica (Toyomane, *et al.*, 2020).

Nel corso degli ultimi anni si sono susseguiti numerosi studi preliminari in merito al microbioma cutaneo umano per far luce sulla concreta possibilità di impiegarlo, in futuro, in ambito forense. Tali studi si sono concentrati soprattutto su:

### I. Variabilità intra-individuale tra le diverse regioni cutanee.

Il microbioma cutaneo varia sulla base del sito considerato in funzione di differenze di pH, livello di idratazione, spessore cutaneo, rugosità, densità di follicoli, densità di ghiandole ed esposizione ambientale (Ruuskanen *et al.*, 2022). Infatti, risulta possibile suddividere grossolanamente i siti cutanei secchi, umidi e sebacei, ognuno dei quali caratterizzato da una composizione microbica assestante (Grice, *et al.*, 2009). Generalmente i siti cutanei secchi (palmo della mano, avambraccio, ginocchio, natica, gomito e ginocchio) presentano una prevalenza dei generi *Propionibacterium*, *Staphylococcus* e *Corynebacterium*. I siti sebacei (scalpo, viso, petto e schiena) sono dominati da *Propionibacterium* e *Staphylococcus*. Mentre i siti umidi (ascella, cavo popliteo, interno del gomito, zona interdigitale, inguine e piedi) hanno elevata abbondanza di *Corynebacterium* e *Staphylococcus* (Figura 2).

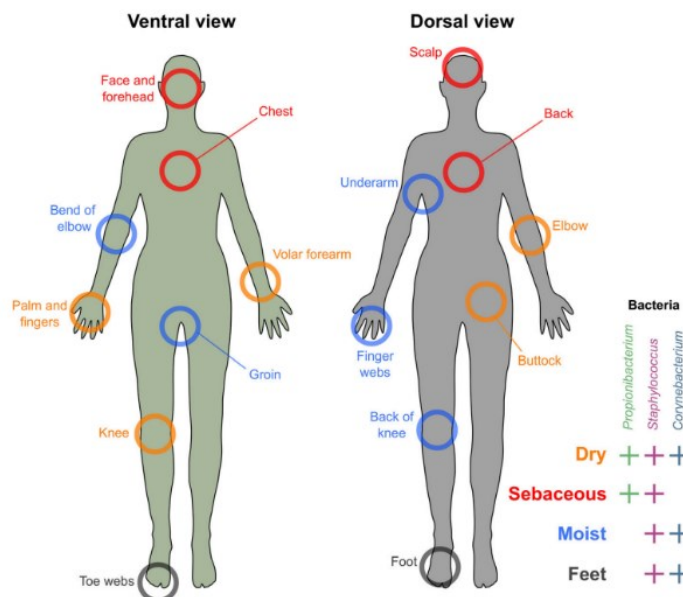


Figura 2. Distribuzione batterica in funzione delle regioni cutanee (Ruuskanen, *et al.*, 2022)

### II. Variabilità interindividuale.

La variabilità microbica caratterizzante la cute di individui differenti è fortemente influenzata da fattori intrinseci come sesso, età, provenienza biogeografica e genetica. Batteri come *Roseomonas* e *Corynebacterium*, sembrano avere elevata ereditabilità. Contribuiscono alla variabilità microbica interindividuale anche fattori estrinseci come il contatto con altri individui, il contatto con animali, l'esposizione a microrganismi cutanei, l'esposizione ai raggi UV, l'esposizione a prodotti chimici, l'utilizzo di cosmetici e le abitudini alimentari (Ruuskanen *et al.*, 2022). La somma di queste variabili favorisce la

costruzione di comunità batteriche specifiche per ogni individuo che, quindi, pongono le prospettive per essere utilizzate come mezzo di discriminazione.

### III. Resistenza sulla sorgente.

La comunità microbica cutanea persiste nonostante la continua esposizione a fattori estrinseci perturbanti. La composizione batterica dei siti secchi e sebacei risulta più stabile rispetto ai siti umidi, nello specifico rispetto alla pianta del piede che, per l'utilizzo di scarpe e per abitudini igieniche, risulta molto più variabile. La stabilità microbica è stata misurata anche a differenti livelli tassonomici, rilevando tendenze simili nel breve e lungo periodo a livello di phylum, mentre sono state rilevate differenze a livello di regno. Un altro fattore che influenza la stabilità microbica risiede nella composizione della stessa. Infatti, questa risulta inversamente correlata alla diversità presente all'interno della comunità. Le specie a bassa ed alta abbondanza risultano essere più stabili, mentre quelle ad abbondanza intermedia subiscono variazioni maggiori (Oh, *et al.*, 2012). Ciò nonostante, il microbioma cutaneo umano risulterebbe abbastanza stabile con la possibilità di ricondurlo al donatore per almeno tre anni (Schmedes, *et al.*, 2017).

### IV. Trasferibilità e permanenza sulla superficie ricevente.

Il microbioma cutaneo è relativamente stabile nel tempo e facilmente reperibile dalle superfici toccate da un potenziale perpetratore. I batteri della pelle possono persistere sulle superfici fino a 30 giorni per via dell'elevata resistenza che hanno nei confronti di stress ambientali, tra cui umidità, temperatura, radiazioni UV. Il *Touch Microbiome*, potenzialmente, può confermare l'associazione con il donatore fungendo da supporto alla canonica analisi degli STR (Procopio, *et al.*, 2021).

### V. Variabili che influenzano le caratteristiche del microbioma cutaneo sul donatore al momento del campionamento.

Risulta molto importante la scelta degli individui e delle caratteristiche del campionamento per evitare di incorrere in bias che potrebbero falsare il risultato finale. A tal proposito, la scelta dei soggetti presi in esame viene vincolata dall'utilizzo di antibiotici, dall'ospedalizzazione, dalla gravidanza, dall'allattamento e dalle allergie (Grice, *et al.*, 2009). Inoltre, in previsione del campionamento, spesso viene chiesto di non attuare comportamenti che possano modificare la composizione microbica come l'utilizzo di sapone, igienizzante e cosmetici (Oh, *et al.*, 2012).

### VI. Predizione delle caratteristiche del donatore.

Il microbioma cutaneo potrebbe essere utilizzato per l'identificazione personale mediante la predizione del sesso, dell'età anagrafica, dell'origine biogeografica e di eventuali malattie. Maschi e femmine vengono contraddistinti da una significativa differenza nella composizione microbica dei batteri residenti (Phan, *et al.*, 2020). Tuttavia, l'origine di questa differenza

è ancora da indagare in quanto potrebbe essere, da un lato, correlata alle abitudini dei soggetti in esame e dall'altro correlata a caratteristiche intrinseche come pH e produzione ormonale. Un altro aspetto interessante riguarda la predizione dell'età anagrafica previa valutazione della composizione microbica cutanea che, probabilmente, si modifica con l'avanzare degli anni in funzione delle modifiche fisiologiche a cui va incontro la cute con l'invecchiamento. In ultima analisi, la presenza di patogeni come *Streptococcus* e *Staphylococcus*, in proporzioni variabili a seconda delle famiglie batteriche, ma anche l'assenza di microrganismi definiti "buoni", può fungere da segnale distintivo per la predizione di una malattia che possa caratterizzare il soggetto in esame (Gupta, *et al.*, 2017).

#### VII. Risorse bioinformatiche e statistiche.

Per poter eseguire un'analisi ottimale c'è la necessità di mantenere in costante aggiornamento i database batterici e di sviluppare software analitici in grado di fornire, in prima analisi, una stima della composizione batterica presente nella zona campionata, con frequenza relativa di ogni famiglia e genere. In seconda analisi, sono necessari strumenti statistici per l'individuazione del possibile match tra sorgente e superficie ricevente, corredato di valori numerici che ne supportino la significatività (Procopio, *et al.*, 2021).

#### VIII. Linee guida per evitare contaminazioni durante il workflow.

Oltre all'utilizzo dei canonici dispositivi di protezione individuale e di strumentazioni sterili e monouso, occorrono accortezze aggiuntive e indicazioni per il monitoraggio delle eventuali contaminazioni. In primo luogo, la disinfezione delle superfici e dei reagenti risulta fondamentale per limitare la possibilità di incorrere in contaminazioni da parte di microrganismi ambientali. In secondo luogo, ad ogni step della procedura analitica, dovrebbero essere allestiti controlli negativi per monitorare le contaminazioni e controlli positivi per monitorare l'efficacia del protocollo (Eisenhofer, *et al.*, 2018).

### 1.4. EVOLUZIONE DEL SEQUENZIAMENTO

Il sequenziamento del DNA è di importanza essenziale in tutti i settori e le applicazioni della biologia moderna. Nel corso degli anni c'è stata un'evoluzione dell'applicazione del sequenziamento automatizzato Sanger a tecnologie di sequenziamento più sofisticate.

Il metodo Sanger ha rappresentato per molto tempo la base della ricerca genetica e genomica ma, a causa delle svariate limitazioni (i.e. metodologia manuale, sequenziamento di un singolo campione, coverage 1x), è stato sostituito da nuove tecnologie (MPS) capaci di sequenziare un grande numero di genomi contemporaneamente (Metzker, *et al.*, 2009).

L'avvento dell'MPS ha permesso di superare la lentezza dell'analisi e la scarsa risoluzione, rappresentando un approccio rivoluzionario per le scienze genomiche.

Questa tecnologia promuove la produzione di un grande numero di dati, con un *coverage* elevato (Metzker, *et al.*, 2009).

#### 1.4.1. METODO SANGER: SEQUENZIAMENTO PRIMA GENERAZIONE

Il metodo Sanger è stato introdotto nel 1975, dall'omonimo premio Nobel, per individuare le sequenze nucleotidiche che costituiscono una determinata molecola di DNA. Questo metodo si basa sulla sintesi di copie del DNA templato a partire da primer specifici che, casualmente, inglobano un dideossinucleotide (ddNTP) terminatore di catena bloccando l'estensione del filamento. Si ottengono, in questo modo, oligonucleotidi di lunghezze differenti che vengono separati tramite elettroforesi e visualizzati grazie a fluorocromi. I ddNTP sono nucleotidi privi di un gruppo ossidrilico ma coniugati ad un fluorocromo che differisce per colore a seconda della base azotata. L'emissione di fluorescenza, emessa dall'ultimo nucleotide del filamento, viene tradotta in un elettroferogramma raffigurante una successione di picchi colorati (Caramelli, 2009).

Il sequenziamento Sanger presenta numerose problematiche soprattutto correlate all'utilizzo di elettroforesi per la lettura delle basi, al ridotto numero di campioni che possono essere processati in parallelo, alla difficoltà nell'automatizzazione, alla scalabilità e ai costi elevati.

#### 1.4.2. NGS: SEQUENZIAMENTO DI SECONDA E TERZA GENERAZIONE

Nel 2005 venne introdotta la seconda generazione di sequenziatori caratterizzati da capacità e rese estremamente elevate. Molte compagnie si impegnarono nello sviluppo di questi macchinari, come Illumina che, con *Solexa*, diventò il primo leader nel campo del sequenziamento genomico. La sua tecnologia sfruttava l'attività della DNA polimerasi, già introdotta da Sanger, aggiungendo l'immobilizzazione del DNA su un supporto fisso e l'utilizzo di terminatori reversibili. Immobilizzare l'acido nucleico mediante adattatori specifici favorisce la lettura di milioni di molecole simultaneamente senza confonderle tra loro. I terminatori reversibili sfruttano una reazione enzimatica per la rilevazione, consentendo di proseguire con l'estensione del filamento.

Altre piattaforme di sequenziamento, differenti da *Solexa* in termini di tecnologia, sono state introdotte da *ABI/Solid* e *Roche/454* (Liu, *et al.*, 2012) e prevedevano:

- Immobilizzazione delle molecole su piccolissime sfere.
- Sostituzione della DNA Polimerasi con la DNA ligasi, che unisce i frammenti invece di sintetizzarli.
- Sequenziamento mediante duplice lettura del medesimo nucleotide al fine di ridurre gli errori.

Il sequenziamento di terza generazione conferisce un'ulteriore evoluzione rispetto al precedente, consentendo la lettura di sequenze lunghe fino a 10 kb. Questa tecnologia consente di costruire librerie in modo veloce e di agire in modo mirato su una molecola *target* senza bisogno di frammentazione. Contestualmente



all'avanzamento delle miglierie aumentano anche l'accuratezza delle letture e l'esecuzione delle analisi successive mediante software (Liu, *et al.*, 2012).

Alcuni esempi di questa nuova tecnologia sono *Ion Torrent*, *Pacific Biosciences* e *Oxford Nanopore Technology*.

### Tecnologia Illumina

I sequenziatori MPS Illumina sfruttano la tecnologia del *Sequencing by Synthesis* (SBS) per il sequenziamento genico massivo. L'SBS utilizza nucleotidi marcati con quattro differenti fluorofori che fungono da terminatori per la reazione di estensione. Dopo l'incorporazione, la fluorescenza viene registrata e successivamente il fluoroforo viene scisso mediante reazione enzimatica per favorire il proseguimento della reazione.

La chimica SBS consente a tutti i macchinari Illumina di effettuare un sequenziamento *paired-end* a partire da entrambe le estremità degli ampliconi *target* con i vantaggi di: diminuire i tempi di amplificazione, favorire l'allineamento con le sequenze di riferimento e facilitare il rilevamento di riarrangiamenti genici, varianti strutturali, fusioni geniche e nuove trascrizioni. Il risultato finale è un accurato sequenziamento base-per-base.

Il flusso di lavoro Illumina prevede quattro passaggi fondamentali:

- I. Preparazione della libreria di sequenziamento mediante frammentazione del genoma o amplificazione della regione *target* e successiva indicizzazione duale per assegnare un riconoscimento univoco ad ogni campione.
- II. Generazione di *cluster*. I *target* di sequenziamento vengono immobilizzati sulla superficie (fase solida) di una *flow cell* (Figura 3) progettata per facilitare l'ingresso dei reagenti e sottoposti ad amplificazione mediante *Bridge PCR* (Figura 4). Vengono prodotti, in questo modo, fino a 1000 copie di DNA per ogni frammento in prossimità della molecola stampo.

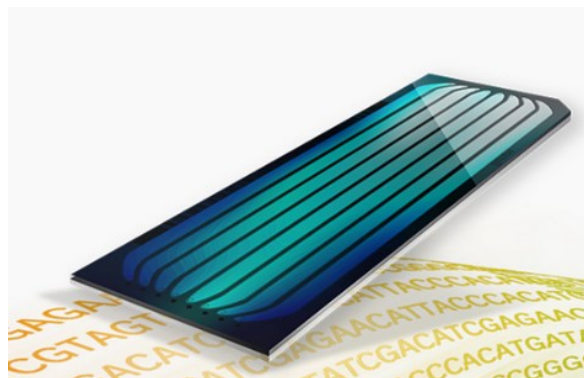


Figura 3. Flow Cell Illumina (illumina.com)

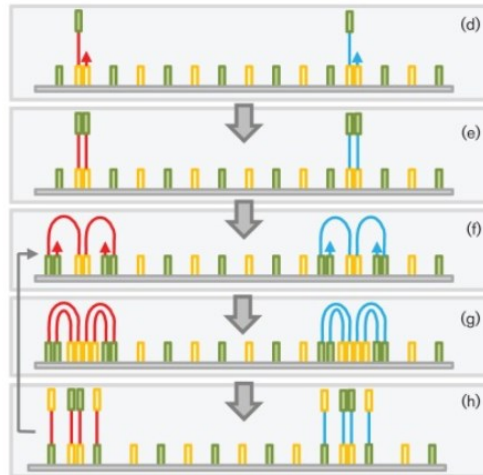


Figura 4. Bridge PCR (Radford, et al., 2012 - [https://www.researchgate.net/publication/225080524\\_Application\\_of\\_next-generation\\_sequencing\\_technologies\\_in\\_virology](https://www.researchgate.net/publication/225080524_Application_of_next-generation_sequencing_technologies_in_virology)).

- III. Sequenziamento mediante rilevamento della fluorescenza generata dall'incorporazione dei nucleotidi marcati.
- IV. Analisi dei dati durante la quale le sequenze vengono allineate con genomi di riferimento.

### Tecnologia Ion Torrent

La tecnologia *Ion Torrent* (Thermo Fisher Scientific), utilizza un sistema di sequenziamento *high-throughput* in grado di generare milioni di sequenze geniche contemporaneamente in tempi molto brevi con un procedimento di SBS. Grazie ad un materiale semiconduttore, questa metodica legge le sequenze di DNA misurando le variazioni di pH generate dal rilascio di ioni idrogeno quando un nucleotide viene incorporato nel filamento in allungamento (Merriman, et al., 2012).

Il sequenziamento *Ion Torrent* prevede diverse fasi consecutive:

- I. Frammentazione del DNA estratto in piccoli segmenti di lunghezza variabile, tramite amplificazione effettuata con l'utilizzo di primer specifici che ricoprono tutta la molecola o regioni *target*.
- II. Digestione parziale delle estremità degli ampliconi per renderle *blunt* (piatte).
- III. Ligazione di adattatori alle estremità 5' e 3' di ciascun amplicone, che fungono da ponte tra le *blunt-ends* e piccole sequenze oligonucleotidiche definite come *barcodes*. I *barcodes* sono codici univoci utili nella distinzione dei differenti campioni quando si lavora in multiplex. In questa fase viene aggiunto un controllo interno di reazione (*Key Sequence*), costituito da 4 basi che vengono incorporate durante la ligazione, il cui ruolo è di favorire la distinzione dei campioni dai *test fragments* (controlli positivi di sequenziamento che invece

hanno una differente sequenza di partenza) e di calibrare ogni singola incorporazione nucleotidica.

#### IV. Purificazione degli ampliconi.

- V. Preparazione del template (amplificazione clonale e arricchimento) e caricamento dei chip che vengono eseguiti automaticamente dallo strumento *Ion Chef<sup>TM</sup>*. L'amplificazione clonale delle sequenze viene effettuata mediante *emulsion PCR (emPCR)* su biglie magnetiche (*Ion Sphere<sup>TM</sup> Particles, ISPs*), disperse in emulsione di olio, acqua e reagenti necessari (*primers*, dNTP, DNA polimerasi e  $MgCl_2$ ). In ciascuna gocciolina è presente una biglia ricoperta di primers complementari agli adattatori a cui si appaia auspicabilmente una sola molecola di DNA che, amplificandosi clonalmente sulla superficie della biglia stessa, andrà a formare milioni di ampliconi dsDNA (Figura 5). In questa fase viene aggiunto un *test fragment* (TF\_1).

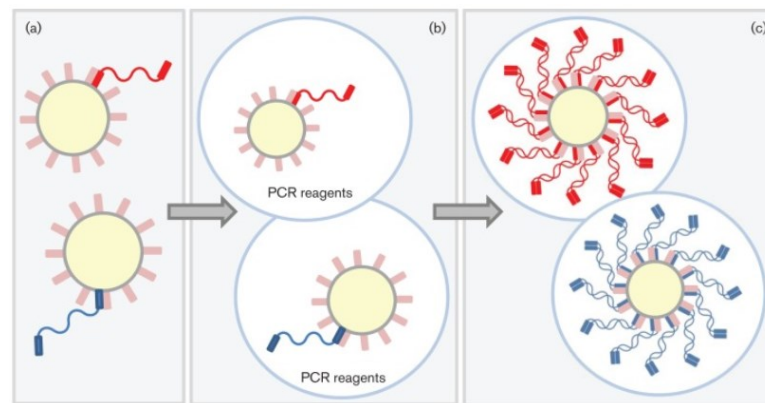


Figura 5. Emulsion PCR (Radford, et al., 2012 - [https://www.researchgate.net/publication/225080524\\_Application\\_of\\_next-generation\\_sequencing\\_technologies\\_in\\_virology](https://www.researchgate.net/publication/225080524_Application_of_next-generation_sequencing_technologies_in_virology)).

L'arricchimento consiste nell'eliminazione delle biglie che non hanno legato alcun frammento di DNA. Per questo motivo vengono aggiunti alla reazione un *primer* di arricchimento biotinilato (*Enrichment Primer*) complementare all'adattatore e delle biglie magnetiche ricoperte di streptavidina (*Capture Beads*). Solo le biglie con DNA legheranno le *Capture Beads* attraverso il legame biotina-streptavidina e verranno attratte da un magnete. Segue poi la denaturazione che favorisce il distacco dell'*Enrichment Primer* e il recupero delle *Capture Beads* con il DNA (ssDNA).

Il template viene caricato su un chip, insieme al secondo *test fragment* (TF\_C). Ogni *chip* è composto da milioni di pozzetti che, nello strato superiore, accolgono le biglie mentre, nei due strati inferiori, contengono i sensori di pH (Figura 6) (Merriman, et al., 2012). I *chip* specifici per l'analisi del microbioma, sulla base del numero di *reads* e del numero di campioni, sono *Ion 540<sup>TM</sup>* e *Ion 550<sup>TM</sup>*. In commercio, sono presenti anche i chip *Ion 510<sup>TM</sup>*,

*Ion 520<sup>TM</sup>* e *Ion 530<sup>TM</sup>*, impiegati in ambito clinico e in ambito forense per l'identificazione umana ma non indicati per l'analisi del microbioma per via della lunghezza degli ampliconi prodotti e del *coverage* necessario alla buona riuscita delle analisi. Sfortunatamente, in fase di arricchimento, non è possibile prevenire il legame delle biglie a più molecole di DNA, pertanto, possono formarsi biglie policlonali, che verranno poi escluse dall'analisi grazie ai filtri impostati nella fase successiva al sequenziamento. Questo può accadere più frequentemente a seguito di un'errata diluizione della libreria caricata su *Ion Chef<sup>TM</sup>*.



Figura 6. Chip Ion 540<sup>TM</sup> e Chip Ion 550<sup>TM</sup> (thermofisher.com)

VI. Sequenziamento del chip eseguito mediante *Ion S5<sup>TM</sup>* (*Ion Proton*) che sfrutta la tecnologia a rilevamento ionico. Nello specifico, lo strumento utilizza il sistema CMSO (*Complementary Metal - Oxide - Semiconductor*), basato sull'invertitore a *transistor ISFET* (*Ion - Sensitive Field Effect Transistor*) che misura il pH in campioni liquidi. Ogni pozzetto contiene verosimilmente una biglia. Il chip viene inondato da un flusso di nucleotidi omogenei che, una volta inglobati, rilasciano di un protone  $H^+$  provocando un accumulo di carica positiva ( $\Delta Q$ ) nel pozzetto. Questo segnale viene trasmesso come variazione di tensione ( $\Delta V$ ) al *gate transistor* e registrato per poi essere utilizzato dall'algoritmo di *base calling* che lo converte in una sequenza leggibile (Gupta, *et al.*, 2020).

Il chip viene sottoposto ad un numero predeterminato di cicli, ognuno dei quali corrisponde a 4 flussi consecutivi di dNTP (*flow*) intervallati da lavaggi. Tutti i dati grezzi vengono immessi nel *Software Torrent Suite* (Thermo Fisher Scientific) che converte il segnale di ogni *read* rilevata, dapprima in uno ionogramma, poi in un file BAM (*Binary Alignment Map*), ossia un formato binario compresso del formato SAM contenente l'allineamento di tutte le letture per un rapido accesso ai dati di sequenziamento.

Vengono definite *reads* le sequenze generate nella lettura della libreria che, grazie ai *barcodes*, possono essere riassemblate nell'intera sequenza genomica, previo allineamento con sequenze di riferimento, da qui il nome *Targeted Resequencing*. In assenza di un genoma di riferimento, la metodica prende il nome di *De Novo Sequencing*.

Dopo l'allineamento, i dati grezzi vengono restituiti dalla macchina in files nel formato FASTQ, un file di testo che contiene i dati di sequenza dei cluster che passano il filtro su una cella a flusso. Ciascuna voce dei FASTQ è composta da 4 righe: un identificatore di sequenza con informazioni sul sequenziamento e sul cluster, la sequenza, un separatore (segno +) e i punteggi di qualità *phred* riferiti all'identificazione di base.

## 1.5. PANNELLI MPS PER L'ANALISI DEL MICROBIOMA

Lo studio di batteri che colonizzano l'uomo ha suscitato molto interesse negli ultimi anni, specialmente in ambito sanitario. Mantenere un microbiota sano è importante perché permette di mantenere il nostro organismo in equilibrio regolando le principali funzioni metaboliche e immunitarie ed evitando l'insorgenza di patologie provocate da disbiosi. La conoscenza della composizione del microbiota umano risulta, quindi, uno strumento essenziale per garantire agli organismi il mantenimento dello stato di salute. Il *target* degli ultimi tempi è sicuramente il microbiota intestinale estremamente abbondante e spesso responsabile di disturbi sia localizzati che sistemici. Le classiche metodiche di coltura cellulare e successivo sequenziamento si sono rivelate obsolete con l'avvento di tecnologie NGS all'avanguardia e semplificate.

Nascono così diversi kit mirati all'analisi del microbioma umano, specialmente intestinale; due esempi consistono in *16S Metagenomic Sequencing Library Preparation* (Illumina - Part # 15044223 Rev. B) e *Ion AmpliSeq Microbiome Health Research Kit* (Thermo Fisher Scientific – MAN0018602) che forniscono reagenti e materiali per analisi mirate mediante NGS.

Il pannello *16S Metagenomic Sequencing Library Preparation* si compone di un unico pool di primers specifici per l'amplificazione delle regioni iper-variabili V3-V4 del 16S rRNA.

Il pannello *Ion AmpliSeq Microbiome Health Research Kit*, invece, si compone di due pool di primer oligonucleotidici di tipo *Ion AmpliSeq*:

- *16S rRNA Gene Pool* per l'amplificazione e il sequenziamento di 8 delle 9 regioni iper-variabili contenute nel gene batterico 16S rRNA (Figura 7).

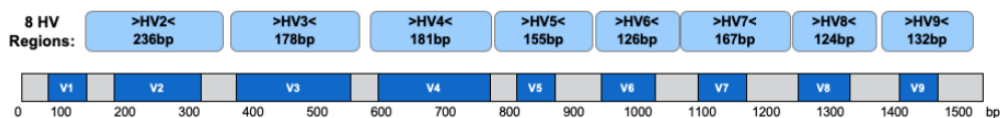


Figura 7. Rappresentazione delle 9 regioni iper-variabili del 16S rDNA e dei rispettivi primers forniti dal kit "Ion AmpliSeq™ Microbiome Health Research Kit" (Characterization of gut microbiota and identification of key bacterial species associated with immune response using a new Ion Torrent next-generation sequencing assay; Thermo Fisher Scientific).

- *Target Species Pool* per l'amplificazione e il sequenziamento di sequenze specie-specifiche appartenenti a 73 specie batteriche responsabili di patologie umane; inclusi *H. pylori*, *B. vulgatus*, *B. adolescentis*, e altri generi come *Lactobacillus* e *Bacteroides*.

L'impiego di primer specie specifici garantisce un'elevata copertura del ceppo batterico rilevato e conferisce un'estrema sensibilità perché necessita di basse quantità di DNA template in input.

Attualmente non sono presenti in commercio kit specifici e validati per l'applicazione forense dell'analisi del microbioma umano. I kit menzionati, infatti, sono stati pensati originariamente per analisi in ambito clinico-sanitario in cui il materiale biologico di partenza, generalmente, è molto abbondante.

## 1.6. SOFTWARE PER L'ANALISI SECONDARIA

L'interpretazione dei dati ottenuti attraverso la tecnologia Illumina viene eseguita mediante *Metagenomics Workflow*, un tipo di analisi secondaria effettuabile attraverso il software *MiSeq Reporter v2.3* (software sul sistema Illumina). Questo sistema esegue un demultiplexing delle *reads* indicizzate, generano files in formato FASTQ e classifica le *reads*. Successivamente, esegue una classificazione tassonomica utilizzando il database di riferimento *Greengenes* e restituendo i taxa batterici sottoforma di grafico.

L'analisi secondaria relativa al protocollo *Ion Torrent* viene eseguita con *Ion Reporter™ Software v5.14* che esegue l'analisi sui file BAM prodotti dal *Torrent Suite™ software*, per ottenere risultati di alta qualità, assegnando priorità e annotando le varianti grazie al supporto di oltre 25 database. Nello specifico, per il microbioma, *Ion AmpliSeq Microbiome Health Research Panel* favorisce la classificazione delle specie microbiche mediante ampliconi 16S e *target* specie-specifici comparando i campioni, singolarmente o in multiplo con differenti database e utilizzando le funzionalità del software QIIME2 (Caporaso, *et al.*, 2010). QIIME2 (*Quantitative Insights Into Microbial Ecology*) è un software ad accesso pubblico nato per agevolare l'interpretazione dei dati grezzi prodotti dal sequenziamento di diverse comunità batteriche e il deposito dei risultati nei database. È in grado di supportare un'ampia gamma di analisi relative alle comunità microbiche e di fornire visualizzazioni grafiche che agevolano l'interazione tra dato e utente. Grazie alle sue proprietà, *Ion Reporter* è, quindi, un software progettato per far fronte a differenti applicazioni nell'ambito della ricerca che spaziano dall'analisi clinica di mutazioni associate all'insorgenza di tumori e alla metagenomica. I risultati vengono poi salvati all'interno del portale stesso sotto forma di grafici e tabelle che mostrano la composizione batterica e le proporzioni relative di ogni specie all'interno del campione.

## 2. SCOPO DEL LAVORO

Il presente lavoro di tesi è stato svolto presso l'Università degli Studi di Padova, in collaborazione con il Laboratorio del Reparto Carabinieri Investigazioni Scientifiche (RIS) di Parma. Lo scopo è stato quello di concentrare l'attenzione sullo studio preliminare del microbioma cutaneo umano, applicato in ambito forense, per valutarne l'utilità nell'identificazione personale a supporto dei marcatori STR oppure quando il DNA umano è degradato o non reperibile.

Questo lavoro analizza la composizione batterica di tracce, prelevate da oggetti di uso quotidiano, e campioni prelevati dalla cute dei soggetti esaminati (confronti), alla ricerca di una possibile corrispondenza. Il procedimento analitico è stato eseguito mediante due moderne tecnologie di *Massive Parallel Sequencing* (MPS), Illumina e Thermo Fisher Scientific, che differiscono sia per la chimica utilizzata, sia per le modalità di rilevazione dei segnali prodotti; entrambi si avvalgono di kit, normalmente impiegati per lo studio del microbioma intestinale. A seguito del sequenziamento sono state ottenute le sequenze geniche relative alle regioni iper-variabili del gene codificante il 16S rRNA che, accumulando mutazioni con elevata frequenza, dovrebbero consentire la discriminazione delle specie batteriche. I profili genetici ottenuti sono stati utilizzati per stabilire la composizione batterica cutanea umana delle aree corporee selezionate. E' stata poi verificata l'attendibilità dei risultati con i dati presenti in letteratura. Affinché i batteri cutanei trovino impiego in ambito forense, infatti, ogni regione corporea e ogni individuo dovrebbero essere caratterizzati da taxa differenti sia in termini di qualità sia in termini di quantità in modo da poter ricondurre una traccia prelevata da un substrato al rispettivo donatore.

Le valutazioni eseguite in questa tesi, sono state sviluppate dapprima attraverso una statistica descrittiva previa osservazione dell'output di sequenziamento e, solo in un secondo momento, sono state rafforzate da un'indagine statistica più approfondita.

In estrema sintesi, gli obiettivi proposti da questo lavoro di tesi sono:

- Verificare la ripetibilità dell'analisi e la possibilità di ottenere dati paragonabili utilizzando due differenti tecnologie di sequenziamento.
- Valutare l'efficacia di kit di sequenziamento, pensati per lo studio del microbioma intestinale, per l'analisi del microbioma cutaneo in ambito forense.
- Identificare le specie batteriche caratteristiche della cute (*Core Microbiome*) avvalendosi del confronto con i dati presenti in letteratura.
- Confermare la presenza di variabilità intra-individuale del microbioma cutaneo.
- Verificare la presenza di variabilità interindividuale del microbioma cutaneo, con la prospettiva di poterlo utilizzare per la discriminazione degli individui.
- Ricercare la corrispondenza tra donatore della traccia e *Touch Microbiome*, rilasciato su una superficie ricevente al momento del contatto, con la prospettiva futura di poter utilizzare questo tipo di analisi a scopo forense.

### 3. MATERIALI E METODI

#### 3.1. RACCOLTA E CONSERVAZIONE DEI CAMPIONI

Per questo esperimento sono stati selezionati 5 soggetti sani (nominati F, G, H, I, L) di età compresa tra 23 e 58 anni, di sesso maschile e femminile. La selezione è avvenuta prediligendo individui che non fossero in corso di gravidanza o allattamento, che nel mese antecedente al prelievo non fossero stati ospedalizzati e che non avessero assunto antibiotici (Procopio, *et al.*, 2021).

Per ognuno di loro sono stati eseguiti quattro prelievi: due tracce raccolte su oggetti di loro proprietà e due campioni di confronto prelevati direttamente dalla cute. Gli oggetti scelti per la sperimentazione consistono nel volante della macchina e nel colletto della camicia. Questa scelta è stata dettata dalla frequente disponibilità di questi ultimi nel corso delle indagini forensi, in quanto possibili fonti di materiale biologico da cui è possibile ricavare il DNA del materiale utilizzatore (Butler, 2012). Ne consegue che, per stabilire la relazione tra “reperito” e utilizzatore, sia stato necessario prelevare i campioni di confronto direttamente dalla cute dei soggetti esaminati, nello specifico dai palmi delle mani e dal collo. Per la raccolta del materiale batterico sono stati utilizzati tamponi cotonati monouso in legno DOC (Neckovic, *et al.*, 2020), umidificati su un unico lato con 25 µl di acqua MilliQ, strofinando e ruotando di 180° il tampone sulla superficie (Schmedes, *et al.*, 2017), seguendo una sola direzione e mantenendo una vigorosa pressione per almeno 30 secondi (Procopio, *et al.*, 2021; Ross, *et al.*, 2017).

Per prima cosa sono state prelevate le tracce, in modo da evitare eventuali contaminazioni che potrebbero falsare il risultato (Butler, 2012):

- Dal volante, il campionamento è stato effettuato scandagliando l'intera superficie, prestando particolare attenzione alle scanalature e ai rilievi che rappresentano possibili zone di accumulo di materiale cellulare.
- Dal colletto, il campionamento è stato effettuato su tutta la superficie interna, partendo dal lembo libero sinistro e terminando al destro.

Con lo scopo di ampliare l'analisi, sono state create artificialmente tre tracce miste (nominate M, N e O). La relativa raccolta è stata eseguita come descritto in precedenza. Nello specifico, il volante di tre autovetture è stato maneggiato da un individuo, differente dal proprietario, per un tempo di almeno 5 minuti simulando la guida (Neckovic, *et al.*, 2020). Il campione:

- M è stato ottenuto dal volante di I maneggiato da F
- N è stato ottenuto dal volante di L maneggiato da I
- O è stato ottenuto dal volante di F maneggiato da L



In seguito, sono stati eseguiti i prelievi dei campioni di confronto sui proprietari dei reperti con lo scopo di eseguire test comparativi:

- Dai palmi, il prelievo è stato eseguito su tutta la porzione compresa tra polso e falangette di entrambe le mani (Neckovic, *et al.*, 2020).
- Dal collo, il prelievo è stato eseguito sulle superfici cutanee laterali e posteriore per un'estensione compresa indicativamente tra l'attaccatura dei capelli sulla nuca e la settima vertebra cervicale, all'incirca 5 cm (Wilkins, *et al.*, 2021).

Dal momento che i prelievi e le analisi sono state eseguite da un unico operatore (nominato V), ne è stato prelevato un campione, da utilizzare per esclusione, relativo ai palmi di entrambe le mani, con la medesima tecnica descritta precedentemente (Neckovic, *et al.*, 2020).

I campioni utilizzati per l'analisi sono stati raccolti nel corso della stessa mattinata, in un arco temporale di un paio di ore, per evitare bias analitici dovuti alle variazioni che può subire il microbiota cutaneo in funzione del tempo (Wilkins, *et al.*, 2021), e dell'interazione con altri individui (Procopio, *et al.*, 2021).

Per simulare una condizione paragonabile alla realtà, ai soggetti in esame non è stata richiesta alcuna modifica delle proprie abitudini in termini di igiene personale, dieta e stile di vita (Procopio, *et al.*, 2021). L'unica accortezza è stata rivolta all'utilizzo del volante dell'automobile, che è stato limitato al solo soggetto preso in esame durante la settimana antecedente al prelievo.

Una volta effettuato il prelievo, i tamponi sono stati inseriti all'interno di provette da 2 ml e sono stati conservati a 4°C in attesa del processamento che è avvenuto meno di 24h dopo la raccolta. I campioni non sono stati congelati per evitare la degradazione del materiale biologico, come da protocollo (Qiagen 2014).

Tabella 1. Riassunto raccolta campioni con rispettiva numerazione attribuita al campione durante il workflow.

CODICE	PALMI	COLLO	VOLANTE	COLLETTO
F	26	27	37	38
G	28	29	39	40
H	30	31	41	42
I	32	33	43	44
L	34	35	45	46
V	36	-	-	-
M	-	-	47	-
N	-	-	48	-
O	-	-	49	-

### 3.2. ESTRAZIONE

L'estrazione è stata eseguita manualmente mediante *QIAamp DNA Microbiome Kit* (Qiagen 2014), che utilizza la tecnologia delle *spin columns* per estrarre uniformemente il DNA batterico di tutte le specie presenti nei tamponi o nei fluidi corporei.

Dal campione di partenza, per mezzo di una lisi differenziale, viene innanzitutto estratto e degradato il materiale genetico relativo alle cellule epiteliali dell'ospite. A seguito di una combinazione tra lisi chimica e meccanica, avviene poi la distruzione delle cellule batteriche in modo tale da portare in soluzione il relativo materiale genetico che verrà arricchito utilizzando le *spin columns* e infine eluito in un volume prestabilito (Figura 8).

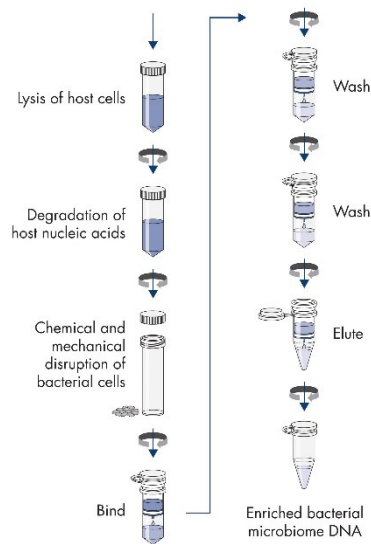


Figura 8. Schema estrazione manuale del DNA batterico mediante *spin-columns* (Qiagen, 2014).

Durante questa fase viene disposto un controllo negativo di estrazione sottoposto al medesimo workflow degli altri campioni e utilizzato per verificare la presenza di eventuali contaminazioni (Eisenhofer, *et al.*, 2018).

Prima di poter procedere con l'estrazione è stato necessario preparare i reagenti e la strumentazione come da specifiche del protocollo. In particolare:

- 25 ml di etanolo (100-80%) sono stati aggiunti al *Buffer AW1*
- 30 ml di etanolo (100-80%) sono stati aggiunti al *Buffer AW2*
- 100  $\mu$ l di *Reagent DX* sono stati aggiunti a 15 ml di *Buffer ATL*

Partendo da uno swab cutaneo, sono state portate in soluzione le cellule batteriche. Per questo motivo, nella provetta in cui è stato raccolto il tampone, è stato aggiunto 1 ml di PBS (*Phosphate Buffered Solution*) e il tutto è stato vortexato e centrifugato. La soluzione così ottenuta è stata poi trasferita in una nuova provetta da 2 ml.

### 3.2.1. ESTRAZIONE E DEGRADAZIONE DEL DNA DELL'OSPITE

Una volta eseguite le operazioni preliminari, è stato possibile procedere con l'estrazione vera e propria, che ha avuto inizio con la lisi differenziale delle cellule epiteliali umane, lasciando intatte le cellule batteriche in quanto più resistenti. Il protocollo prevede di:

1. Aggiungere 500  $\mu$ l di *Buffer AHL* al campione e incubare per almeno 30 minuti a temperatura ambiente (15-25°C) con rotazione sotto-sopra.
2. Centrifugare a 10.000 x g per 10 minuti e rimuovere il surnatante facendo attenzione a non toccare il pellet, contenente le cellule batteriche integre.
3. Aggiungere 190  $\mu$ l di *Buffer RDD* e 2,5  $\mu$ l di *Benzonase*. Quest'ultimo reagente degrada il DNA dell'ospite, portato in soluzione durante gli step precedenti. Vortexare e incubare nel termoshacker impostato a 37°C e 600 rpm per 30 minuti.

### 3.2.2. ESTRAZIONE DEL DNA BATTERICO

Ultimata la degradazione del DNA dell'ospite è stato possibile proseguire con la lisi batterica che prevede una combinazione di step chimici e meccanici per favorire la rottura delle spesse pareti cellulari, ricche di lipidi e polisaccaridi, in modo da evitare distorsioni nella rappresentazione e nella composizione relativa del microbioma. Il protocollo prevede di:

1. Aggiungere 20  $\mu$ l di *Proteinasi K* e incubare nel termoshacker impostato a 56°C e 600 rpm per 30 minuti.
2. Aggiungere 200  $\mu$ l di *Buffer ATL* per favorire la lisi cellulare e mixare bene.
3. Trasferire l'intero volume nei *Pathogen Lysis Tube L* e vortexare per 10 minuti alla massima velocità per assicurare una corretta lisi batterica sfruttando la coesione delle biglie contenute nei tubini stessi. Incubare i campioni a 95°C per 5 minuti.
4. Centrifugare a 10.000 x g per 1 minuto per ridurre la schiuma creatasi dopo la lisi e trasferire il surnatante in nuove provette da 2 ml.
5. Aggiungere 40  $\mu$ l di *Proteinasi K*, vortexare e incubare nel termoshacker impostato a 56°C e 600 rpm per 30 minuti.
6. Aggiungere 200  $\mu$ l di *Buffer APL2* e vortexare.
7. Incubare i campioni a 70°C per 10 minuti. Aggiungere 200  $\mu$ l di etanolo al lisato e vortexare.
8. Trasferire l'intero volume nelle *QIAamp UCP (Ultra Clean Production) Mini spin column* e centrifugare a 6.000 x g per 1 minuto. Questo step favorisce il legame specifico del DNA batterico alla colonnina in silice e l'eliminazione degli scarti derivanti dalla lisi cellulare, riducendo fortemente il rischio di

contaminazioni.

9. Scartare il filtrato.
10. Trasferire le *QIAamp UCP Mini spin column* in un nuovo *collection tube*, quindi aggiungere 500 µl di *Buffer AW1* e centrifugare a 6.000 x g per 1 minuto. Questo step elimina gli aspecifici legati alla colonnina insieme al DNA.
11. Scartare il filtrato e mettere e *QIAamp UCP Mini spin column* in un nuovo *collection tube*, quindi aggiungere 500 µl di *Buffer AW2* e centrifugare a 20.000 x g per 3 minuti. Questo elimina gli aspecifici legati alla colonnina insieme al DNA con un legame più forte rispetto a quelli eliminati nello step precedente.
12. Scartare il filtrato e mettere le *QIAamp UCP Mini spin column* in un nuovo *collection tube*. Centrifugare a 20.000 x g per 1 minuto.
13. Scartare il filtrato. Mettere le *QIAamp UCP Mini spin column* in un nuovo tubino da 1,5 ml e aggiungere 50 µl di *Buffer AVE* di eluizione al centro della membrana. Incubare a temperatura ambiente (15-25°C) per 5 minuti.
14. Centrifugare a 6.000 x g per 1 minuto al fine di ricavare l'eluato di DNA.

Alla fine dell'estrazione è stato ottenuto DNA batterico arricchito ed eluito in un volume finale di 50 µl. I campioni sono stati conservati in congelatore (temperatura compresa tra -15°C e -30°C).

### 3.3. QUANTIFICAZIONE

La quantificazione è stata eseguita mediante *Invitrogen Qubit<sup>TM</sup> 4 Fluorometer* (Thermo Fisher Scientific; Ross, *et al.*, 2017) utilizzando *Qubit<sup>TM</sup> 1X dsDNA HS Assay kit* (Thermo Fisher Scientific; Ross, *et al.*, 2017). Questo kit contiene un *ready-to-use assay buffer* e il DNA standard per la costruzione della curva di taratura.

*Qubit<sup>TM</sup> Fluorometer* è un dispositivo di quantificazione che basa il suo funzionamento sul rilevamento dell'intensità di fluorescenza generata dal legame specifico di un colorante fluorescente con la molecola bersaglio, sia essa dsDNA, ssDNA, RNA o proteina, anche in condizioni di degradazione o in presenza di contaminanti (Thermo Fisher Scientific 2021 - MAN0017209).

La tecnologia *Qubit* permette di calcolare la concentrazione del template interpolando la fluorescenza ottenuta dal campione con la retta di taratura costruita mediante *curve-fitting algorithm* (Figura 9). Lo strumento calcola le concentrazioni basandosi sulla relazione tra i due standards usati negli step di calibrazione con accuratezza e precisione anche a bassissimi range (0.5 ng/mL - 600 ng/mL).

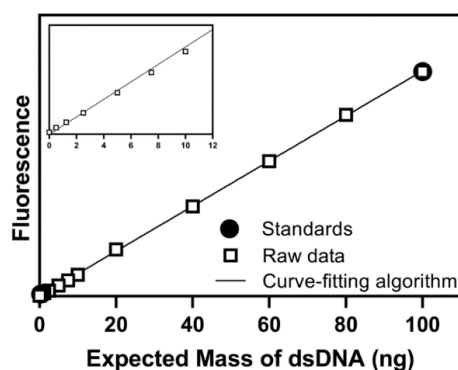


Figura 9. *Qubit™ Assays* retta di taratura (Thermo Fisher Scientific 2021 - MAN0017209).

Il protocollo per la quantificazione prevede di:

1. Preparare i due standard per la costruzione della curva di taratura aggiungendo 10  $\mu\text{l}$  di DNA standard #1 e separatamente DNA standard #2 (Figura 10).
2. Aggiungere 5  $\mu\text{l}$  di DNA template in ogni tubino preparato per la quantificazione (Figura 10). È possibile utilizzare da 1-20  $\mu\text{l}$  di DNA template.
3. Aggiungere 190  $\mu\text{l}$  di *Qubit™ IX dsDNA working solution* ad ogni tubino contenente il DNA standard e 195  $\mu\text{l}$  di *Qubit™ IX dsDNA working solution* ad ogni tubino contenente il DNA template (Figura 9). Il volume di questo reagente va modificato in base al volume di DNA scelto per la quantificazione, per ottenere un volume finale di 200  $\mu\text{L}$ .
4. Vortexare tutti i tubini e incubarli a temperatura ambiente, in condizioni di completa oscurità per almeno 2 minuti. L'assenza di luce è necessaria affinché la fluorescenza del colorante non diminuisca nel tempo, falsando i risultati della lettura.

Una volta preparati i campioni da quantificare, lo strumento è stato settato con i parametri suggeriti dal protocollo:

5. Nella schermata "Home" selezionare le voci di interesse per quanto riguarda il materiale molecolare da quantificare (**DNA**) e il tipo di saggio utilizzato per la quantificazione (**IX dsDNA HS**).
6. Predisporre lo strumento alla lettura degli standard selezionando l'opzione **Read Standards** e inserire uno per volta i tubini contenenti i due standard.
7. Selezionare **Run Samples** e il volume di DNA scelto per la quantificazione (in questo caso 5  $\mu\text{l}$ ), insieme all'unità di misura in cui si vuole ottenere il risultato (ng/ $\mu\text{l}$ ).
8. Inserire il primo tubino contenente il campione da quantificare nell'apposito spazio e chiudere il coperchio. Selezionare l'opzione **Read tube** e, una volta

ottenuto il risultato, rimuovere il tubino. Ripetere l'operazione per tutti i campioni.

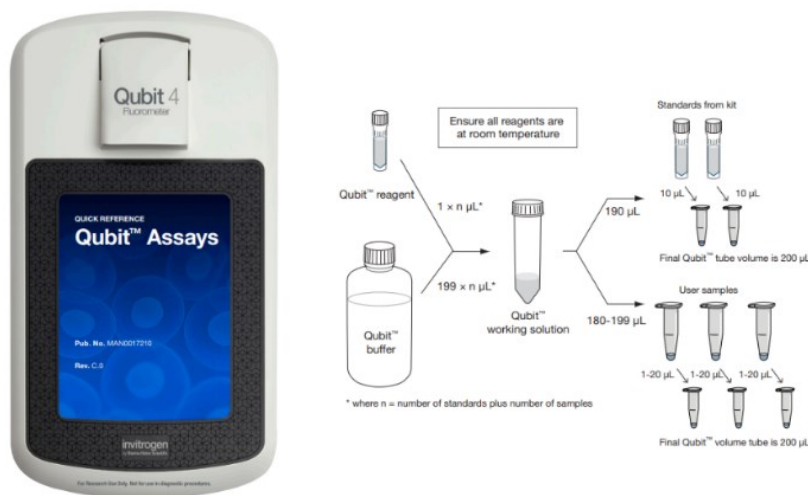


Figura 10. Qubit™ Assays e schema diluizioni per la quantificazione (Thermo Fisher Scientific 2021 - MAN0017209).

### 3.4. SEQUenziAMENTO MEDIANTE NEXT GENERATION SEQUENCING: TECNOLOGIA ILLUMINA

Per la costruzione delle librerie genetiche e per il successivo sequenziamento è stato utilizzato *16S Metagenomic Sequencing Library Preparation* (Illumina – MAN. 15044223 Rev. B). Il kit si compone di primer specifici e di reagenti associati per generare librerie di ampliconi per MPS da effettuare mediante la piattaforma Illumina con l'ausilio di *MiSeq™ System*. L'analisi ha avuto come *target* le regioni iper-variabili V3 e V4 presenti nel gene che codifica il 16S rRNA.

#### 3.4.1. PREPARAZIONE MANUALE DELLE LIBRERIE GENOMICHE

Il protocollo *16S Metagenomic Sequencing Library Preparation* permette la costruzione manuale di librerie genomiche a partire dai campioni di DNA batterico seguendo le fasi descritte successivamente.

##### Amplificazione del DNA Target

La prima fase di preparazione delle librerie genomiche prevede l'amplificazione delle regioni *target* del DNA batterico. Il protocollo prevede di:

1. Preparare la reazione di amplificazione delle regioni *target* in una 96-well plate, dispensando per ogni pozzetto i seguenti reagenti ( $v_{tot} = 25 \mu\text{l}$ ):
  - 5  $\mu\text{l}$  di *Amplicon PCR Forward Primer* 1  $\mu\text{M}$
  - 5  $\mu\text{l}$  di *Amplicon PCR Reverse Primer* 1  $\mu\text{M}$

- 12,5 µl di 2x *KAPA HiFi HotStart ReadyMix*
  - 2,5 µl DNA Microbico ad una concentrazione di 5 ng/µl
2. Chiudere la piastra e azionare la reazione di PCR utilizzando *ProFlex™ PCR System* (Thermo Fisher Scientific) con le impostazioni riportate in Tabella 2:

Tabella 2. Impostazione programma di amplificazione (*Illumina Protocol*)

STAGE	TEMPERATURE	TIME
Hold	95°C	3 minuti
25 Cycles	95°C	30 secondi
	55°C	30 secondi
	72°C	30 secondi
Hold	72°C	5 minuti
Hold	4°C	Hold

### PCR Clean-Up

Questo step utilizza biglie *AMPure XP* per la purificazione degli ampliconi V3 e V4, eliminando i primers che non si sono appaiati e quelli che si sono appaiati tra loro creando dimeri inutilizzabili. Il protocollo prevede di:

1. Centrifugare la piastra a  $1.000 \times g$  a 20°C per 1 minuto.
2. Vortexare le biglie *AMPure XP* per 30 secondi e aggiungerne 20 µl in ogni campione miscelando bene la soluzione con la pipetta. Quindi incubare a temperatura ambiente per 5 minuti.
3. Posizionare la piastra sul supporto magnetico e attendere 2 minuti fino a che la soluzione non risulta limpida e le biglie depositate sul fondo del pozzetto. Quindi rimuovere il surnatante.
4. Mantenendo la piastra sopra al magnete, aggiungere 200 µl di etanolo 80%, incubare per 30 secondi e rimuovere il surnatante.
5. Ripetere le operazioni descritte nello step precedente assicurandosi di rimuovere tutto l'etanolo.
6. Lasciare evaporare l'eccesso di etanolo per 10 minuti a temperatura ambiente.
7. Rimuovere la piastra dal supporto magnetico e aggiungere ad ogni pozzetto 52.5 µl di *Tris* 10 mM a pH 8.5. Mixare bene la soluzione per risospendere le biglie magnetiche e incubare a temperatura ambiente per 2 minuti.
8. Posizionare la piastra sul supporto magnetico e attendere 2 minuti, finché la soluzione non diventa limpida e le biglie non sedimentano sul fondo del pozzetto.
9. Trasferire 50 µl di surnatante in una piastra nuova.

### Indicizzazione degli Ampliconi

Questo step prevede l'attacco di indici duali (realizzati per combinazione di due oligonucleotidi) e adattatori Illumina ai campioni, mediante l'utilizzo di *Nextera XT Index Kit* (Illumina). Il protocollo prevede di:

1. Trasferire 5 µl di ogni campione ottenuto nello step precedente in una piastra nuova. Il volume restante può essere conservato per usi futuri.
2. Ordinare *Index 1* dal numero 1 al numero 12 e *Index 2* dalla A alla Z nell'apposito rack. Quindi inserirvi la piastra (Figura 11).

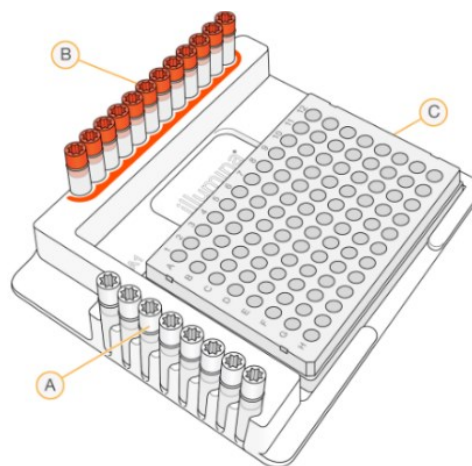


Figura 11. "Nextera XT Index Kit" e piastra per indicizzazione. A: Index 2 primers (white caps); B: Index 1 primers (orange caps); C: 96-well plate (Illumina - MAN 15044223 Rev. B).

3. Aggiungere i seguenti reagenti in ogni pozzetto con il DNA ( $v_{tot} = 50 \mu\text{l}$ ):
  - 5 µl di *Nextera XT Index Primer 1*
  - 5 µl *Nextera XT Index Primer 2*
  - 25 µl 2x *KAPA HiFi HotStart ReadyMix*
  - 10 µl *PCR Grade water*
4. Mixare la soluzione con l'ausilio della pipetta, quindi coprire la piastra e centrifugare a 1.000 x g per 1 minuto.
5. Programmare la reazione di PCR con *ProFlex™ PCR System* (Thermo Fisher Scientific) seguendo le impostazioni riportate in Tabella 3:

Tabella 3. Impostazione protocollo di indicizzazione (Illumina)

STAGE	TEMPERATURE	TIME
Hold	95°C	3 minuti
25 Cycles	95°C	30 secondi
	55°C	30 secondi
	72°C	30 secondi
Hold	72°C	5 minuti
Hold	4°C	Hold



## PCR Clean-Up 2

Come nella precedente purificazione, anche questo step utilizza biglie *AMPure XP* per eliminare tutto ciò che non è utile alla costruzione delle librerie prima della quantificazione. Il protocollo prevede di:

1. Centrifugare la piastra a  $280 \times g$  a  $20^{\circ}\text{C}$  per 1 minuto.
2. Vortexare le biglie *AMPure XP* per 30 secondi e aggiungerne  $56 \mu\text{l}$  in ogni campione miscelando bene la soluzione con la pipetta.
3. Incubare a temperatura ambiente per 5 minuti.
4. Posizionare la piastra sul supporto magnetico e attendere 2 minuti fino a che la soluzione non risulta limpida e le biglie depositate sul fondo del pozzetto. Quindi rimuovere il surnatante.
5. Mantenendo la piastra sopra al magnete, aggiungere  $200 \mu\text{l}$  di etanolo 80%, incubare per 30 secondi e rimuovere il surnatante.
6. Ripetere le operazioni descritte nel punto “5” assicurandosi di rimuovere tutto l’etanolo.
7. Lasciare evaporare l’eccesso di etanolo per 10 minuti a temperatura ambiente.
8. Rimuovere la piastra dal supporto magnetico e aggiungere ad ogni pozzetto  $27,5 \mu\text{l}$  di *Tris* 10 mM a pH 8.5. Mixare bene la soluzione per risospendere le biglie magnetiche e incubare a temperatura ambiente per 2 minuti.
9. Posizionare la piastra sul supporto magnetico e attendere 2 minuti, finché la soluzione non diventa limpida e le biglie non sedimentano sul fondo del pozzetto.
10. Trasferire  $25 \mu\text{l}$  di surnatante in una piastra nuova.

### 3.4.2. QUANTIFICAZIONE, NORMALIZZAZIONE E CREAZIONE DEL POOL

Il metodo di quantificazione consigliato dal protocollo Illumina è di tipo fluorimetrico, e utilizza coloranti progettati per legare la doppia elica di DNA. La concentrazione (nM) del DNA viene determinata mediante *Agilent Technologies 2100 Bioanalyzer trace* sulla base della lunghezza degli ampliconi.

$$\text{concentrazione in nM} = \frac{\text{concentrazione}}{660 * \text{grandezza media librerie}} * 10^6$$

Una volta ottenute le concentrazioni, il protocollo prevede di:

1. Diluire la libreria utilizzando *Resuspension Buffer* (RSB) fino ad ottenere una concentrazione di 4 nM.
2. Aliquotare 5  $\mu$ l di DNA diluito appartenenti ad ogni campione in un unico pool (*Library Pool*).

### 3.4.3. DENATURAZIONE DEL POOL E LOADING

Prima di avviare il sequenziamento, il pool di librerie viene denaturato con NaOH, diluito con buffer di ibridazione e denaturato nuovamente. Ogni corsa deve includere un minimo di 5% di *PhiX* che funge da controllo positivo interno.

*PhiX* è un controllo che fornisce un feedback sull'affidabilità del sequenziamento. Per l'esperimento in questione, le sequenze considerate appartengono al 16S rRNA, ovvero un insieme di regioni molto simili anche tra specie batteriche differenti. Ne consegue che lo strumento, spesso, possa rilevare basi azotate uguali, fornendo un falso segnale di allarme. Per evitare questa problematica e aumentare la variabilità del sequenziamento in questo esperimento è opportuno aggiungere il 10-15% di *PhiX*, che analogamente agli altri ampliconi, si distribuisce in modo casuale sul chip.

#### Denaturazione e diluizione del DNA

1. Preparare una soluzione contenente 5  $\mu$ l di *Library Pool* (4 nM) e 5  $\mu$ l di NaOH (0,2 N) in un tubino da microcentrifuga.
2. Diluire il *PhiX* control in modo analogo al *Library Pool*.
3. Vortexare e centrifugare la soluzione a 280 x g per 1 minuto.
4. Incubare per 5 minuti a temperatura ambiente per ottenere DNA denaturato a singolo filamento.
5. Mixare 10  $\mu$ l di DNA denaturato con 990  $\mu$ l di HT1 per interrompere l'azione denaturante dell'NaOH. A questo punto il DNA denaturato ha una concentrazione finale di 20 pM.
6. Diluire il DNA denaturato alla concentrazione finale di interesse aggiungendo HT1.

#### Denaturazione e diluizione del controllo Phix

Viene eseguita la denaturazione e la diluizione della *Phix Library* (10 nM) per portarla alla medesima concentrazione della libreria di ampliconi. Il protocollo prevede di:

1. Combinare 2  $\mu$ l di *Phix Library* (10 nM) e 3  $\mu$ l di *Tris* (10 mM) a pH 8.5 per ottenere una concentrazione finale di *Phix* pari a 4 nM
2. Combinare 5  $\mu$ l di *Phix Library* (4 nM) e 5  $\mu$ l di NaOH (0,2 N) per ottenere

una concentrazione finale di *Phix* pari a 2 nM. Quindi vortexare la soluzione.

3. Incubare la soluzione per 5 minuti a temperatura ambiente per favorire la denaturazione della *Phix Library*.
4. Combinare 10 µl di *Phix Library* denaturata e 990 µl di HT1 per interrompere l'azione denaturante dell'NaOH.

#### Combinazione Pool Library di Ampliconi con Phix Library

1. Combinare 30 µl di *Phix control* denaturato e 570 µl di *Library Pool* denaturata.
2. Incubare la soluzione a 96° C per 2 minuti.
3. Mixare e deporre in ghiaccio per 5 minuti.

#### 3.4.4. PREPARAZIONE DEL SEQUENZIATORE MISEQ (ILLUMINA)

Il sequenziamento Illumina fonda il suo funzionamento sulla tecnologia di *Sequencing By Synthesis* (SBS) che utilizza un set di nucleotidi privati di un gruppo ossidrilico, dideossinucleotidi (ddNTP) ed etichettati con un fluoroforo per ottenere la sequenza genetica di milioni di molecole in parallelo mediante *bridge PCR*. Durante ogni ciclo, viene aggiunto un solo nucleotide ingegnerizzato per terminare la reazione di estensione dei primer, la fluorescenza viene rilevata e infine, il fluoroforo viene rimosso enzimaticamente per consentire l'allungamento del filamento. Il risultato finale è un sequenziamento "base per base" estremamente accurato. Il sequenziamento del microbioma cutaneo è stato effettuato mediante *MiSeq™ Reagent Kits v2* da 300 cicli, selezionato perché compatibile con la lunghezza degli ampliconi in esame:

1. Una volta acceso lo strumento, prima di inserire il chip e di avviare l'analisi è necessario settare le impostazioni del sistema attraverso la schermata del **Main Menu**.
2. Selezionare la scheda **Run** e il **Maintenance Wash**, ovvero il lavaggio di manutenzione che dura circa 90 minuti e favorisce un corretto svolgimento del sequenziamento.
3. Selezionare la scheda **BaseSpace** per la corretta archiviazione dei dati inviati al supporto Illumina e al servizio di monitoraggio.
4. Selezionare **Send Instrument Performance Data to Illumina** per inviare i dati delle prestazioni dello strumento a Illumina al fine di ricevere assistenza immediata qualora dovessero insorgere problematiche durante l'analisi.
5. Rimuovere la cartuccia dei reagenti dal refrigeratore e scongelarla in un bagno d'acqua deionizzata a temperatura ambiente per circa 60 minuti.
6. Miscelare i reagenti della cartuccia per assicurarsi che non si siano formati precipitati o bolle d'aria.

7. Dispensare 600 µl di libreria precedentemente preparate all'interno del serbatoio denominato come *Load Sample* presente sulla cartuccia *MiSeq v2 reagent cartridge* (Figura 13).
8. Impostare le opzioni di sequenziamento utilizzando l'interfaccia di *MiSeq Control Software*.
9. Dopo aver accuratamente pulito la cella a flusso dai sali, inserirla nell'apposito comparto (Figura 13).
10. Miscelare il flacone di PR2 (*Incorporation Buffer*), inserirlo al posto dei flaconi utilizzati per i lavaggi e assicurarsi che il contenitore per gli scarti sia vuoto (Figura 12).
11. Caricare la cartuccia dei reagenti in cui è stato aliquotato il *Library pool* e avviare la corsa.

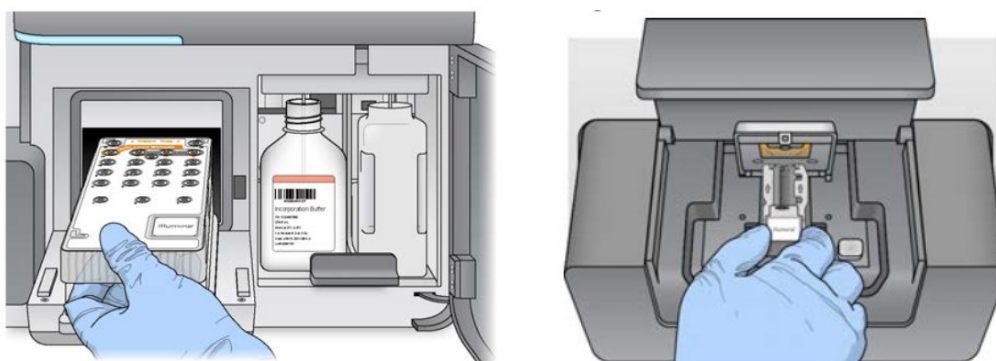


Figura 12. Caricamento cartuccia reagenti, flacone di PR2 e contenitore di scarto (Img sinistra). Inserimento flow-cell (Img destra). (Illumina – 15044223 Rev. B)

### 3.5.5. ANALISI DEI DATI CON MISEQ REPORTER

Dopo aver caricato i campioni sulla piattaforma *MiSeq<sup>TM</sup>*, lo strumento fornisce automaticamente un'analisi secondaria basata su *MiSeq Reporter* (MSR- *MiSeq Reporter Metagenomics Workflow – Reference Guide - Part # 15042317*) quando la corsa viene completata con la modalità “Local Run Manager” o “Sample Sheet”. *MiSeq Reporter* genera file di analisi intermedi, in formato FASTQ, che contengono, per ciascun campione, le letture e i relativi punteggi di qualità. I file FASTQ vengono utilizzati come input per l'allineamento.

I dati di sequenziamento si basano sul confronto delle sequenze V3-V4 appartenenti al gene 16S rDNA con il database *Greengenes* (<http://greengenes.lbl.gov/>). L'analisi fornisce in output il grafico dei cluster, la tabella dei campioni e una rappresentazione grafica mediante *pie-chart* relativa alla classificazione dei campioni.

### 3.5. SEQUENZIAMENTO MEDIANTE NEXT GENERATION SEQUENCING: TECNOLOGIA ION TORRENT

Analogamente a quanto descritto in precedenza, la costruzione delle librerie genetiche e il successivo sequenziamento è stato realizzato anche mediante l'utilizzato di *Ion AmpliSeq™ Microbime Health Research Kit* (Thermo Fisher Scientific; Bjerre, *et al.*, 2019). Il kit si compone di due pools di *Ion AmpliSeq oligonucleotide primers* e di reagenti associati per generare librerie di ampliconi da sequenziare attraverso la piattaforma *Ion Torrent* con l'ausilio dei sistemi *Ion S5 Chef™* e *Ion S5 Sequencer™* (Thermo Fisher Scientific - MAN0018602).

#### 3.5.1. PREPARAZIONE MANUALE DELLE LIBRERIE GENOMICHE

*Ion AmpliSeq™ Microbime Health Research Kit* permette di costruire manualmente le librerie genomiche a partire dai campioni di DNA batterico.

##### Amplificazione del DNA Target

La prima fase di preparazione delle librerie genomiche è l'amplificazione delle regioni *target* del DNA batterico precedentemente estratto e quantificato. Nello specifico, i 24 campioni estratti più 1 controllo negativo di estrazione sono stati utilizzati per svolgere le due differenti analisi previste dal kit: l'analisi di *Target Specie* e l'analisi del *16S rDNA* (Figura 13).

Il protocollo prevede di:

1. Normalizzare il DNA template diluendolo fino al raggiungimento delle concentrazioni richieste da protocollo: 0,167 ng/μl per l'analisi del 16S e 1,67 ng/μl per l'analisi di *Target Specie*.
2. Preparare la reazione di amplificazione delle regioni *target* in una *96-well plate* (Figura 13), dispensando per ogni pozzetto i seguenti reagenti ( $v_{tot} = 20 \mu l$ ):
  - 4 μl di *5X Ion AmpliSeq™ HiFi Mix*, miscela di reazione per amplificazione
  - 4 μl di *5X Ion AmpliSeq™ Microbiome Health Research – Target Species Pool* oppure *16S rRNA Gene Pool*
  - 12 μl di DNA target, il massimo volume di DNA consentito dal protocollo.

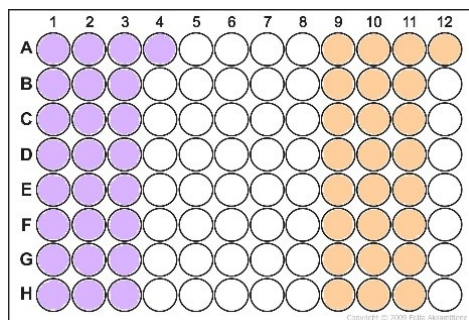


Figura 13. Disposizione piastra di amplificazione per “Ion AmpliSeq™ Microbime Health Research Kit” (Thermo Fisher Scientific). I medesimi estratti sono stati divisi in due e utilizzati per due analisi distinte: in viola i campioni ordinati da 1 a 25 per Target Species, in arancione i campioni ordinati da 26 a 50 per 16S rRNA.

3. Chiudere la piastra, vortexare e centrifugare.
4. Amplificare i campioni con l’ausilio di *ProFlex™ PCR System* (Thermo Fisher Scientific) impostando il programma di amplificazione riportato in Tabella 4:

Tabella 4. Impostazione programma di amplificazione

STAGE	STEP	TEMPERATURE	TIME
Hold	Attivazione enzimatica	99°C	2 minuti
20 Cycles	Denaturazione	99°C	15 secondi
	Annealing ed estensione	60°C	4 minuti
Hold	-	10°C	Hold

### Digestione Parziale degli Ampliconi

Al termine dell’amplificazione delle regioni *target* viene preparata la reazione di digestione parziale degli ampliconi. Questo step viene realizzato mediante *FuPa Reagent (Ion AmpliSeq™ Microbime Health Research Kit)*, un enzima in grado di digerire in modo parziale i primer, fosforilare gli ampliconi e favorire la successiva reazione di ligazione degli *adapters*. Il protocollo prevede di:

1. Aggiungere ai campioni amplificati 2 µl di *FuPa Reagent* ad ogni pozzetto ( $v_{tot} = 22 \mu\text{l}$ ).
2. Chiudere la piastra, vortexarla, centrifugarla e inserirla all’interno del *ProFlex™ PCR System* impostando il programma descritto in Tabella 5:

Tabella 5. Impostazione protocollo di digestione

TEMPERATURE	TIME
50°C	20 minuti
55°C	20 minuti
60°C	20 minuti
10°C	Hold (for up to 1 hour)

### Ligazione degli Adattatori

In seguito, è stata effettuata la ligazione di specifiche sequenze numerate che permettono l'identificazione univoca di ciascun campione. In questo caso viene utilizzato il kit *Ion Xpress™ Barcode X*, contenente gli *adapters* e i *barcodes*. Gli *adapters* sono piccoli oligonucleotidi che vengono ligati alle estremità piatte e fosforilate dei frammenti genetici da analizzare e fungono da ponte per l'attacco dei *barcodes*. I *barcodes* sono sequenze oligonucleotidiche univoche che servono per distinguere i differenti campioni quando si lavora in multiplex. Il protocollo prevede di:

1. Preparare le diluizioni 1:4 dei *barcodes* in un volume finale di 8 µl:
  - 2 µl di *Ion P1 Adapter*
  - 2 µl di *Ion Xpress™ Barcode X*
  - 4 µl di *Nuclease-free Water*
2. Aggiungere ad ogni pozzetto contenente i campioni i seguenti reagenti (volume finale = 30 µl), nell'ordine in cui vengono menzionati per evitare la formazione di aspecifici:
  - 4 µl *Switch Solution*
  - 2 µl diluted *Ion Xpress™ barcode adapter mix*
  - 2 µl *DNA Ligase*
3. Disporre la piastra inserirla all'interno del *ProFlex™ PCR System* impostando il programma descritto in Tabella 6:

Tabella 6. Impostazione protocollo di ligazione

TEMPERATURE	TIME
22°C	30 minuti
68°C	5 minuti
72°C	5 minuti
10°C	Hold (for up to 24 hour)

### Purificazione ed Eluizione

A questo punto le librerie vengono trasferite in una piastra apposita, *Shaped-Bottom Microplate*, che permette la purificazione dei campioni barcodati mediante biglie magnetiche *Agencourt™ AMPure™ XP Reagent* (Thermo Fisher Scientific). Questa tecnologia sfrutta biglie paramagnetiche che legano selettivamente il materiale genetico da analizzare, eliminando i dimeri di adattatori e tutto ciò che non viene legato al supporto. Il protocollo prevede di:

1. Aggiungere 45 µl di *Agencourt™ AMPure™ XP Reagent* ad ogni pozzetto; miscelare per favorire la sospensione delle biglie e del DNA in soluzione e incubare per 5 minuti a temperatura ambiente.

2. Posizionare la piastra sopra al supporto magnetico (*DynaMag™-96 Side Magnet* - Thermo Fisher Scientific) e incubare per 2 minuti fino a che la soluzione non risulta trasparente. Scartare il surnatante.
3. Aggiungere 150 µl di etanolo al 70% e muovere la piastra sul magnete “side-to-side” per almeno tre volte in diagonale per favorire il lavaggio delle biglie. Scartare il surnatante. Quindi, ripetere l’operazione per un secondo lavaggio.
4. Lasciare asciugare le biglie a temperatura ambiente (15-25°C) per 5 minuti senza andare in *over-dry*. La condizione di *over-dry* risulta problematica in quanto, seccandosi, le biglie magnetiche si depositano sul fondo del pozzetto e vi rimangono saldamente adese rendendo difficile il successivo step di eluizione.
5. Rimuovere la piastra dal magnete e risospendere le biglie in 50 µl di *Low TE*.
6. Incubare a temperatura ambiente per almeno 2 minuti.
7. Disporre la piastra sul supporto magnetico per almeno 2 minuti finché la soluzione non diventa trasparente, quindi rimuovere il surnatante e trasferirlo in una nuova *96-well plate*.

### 3.5.2. QUANTIFICAZIONE DELLE LIBRERIE

La quantificazione viene eseguita mediante *Ion Library TaqMan™ Quantitation Kit* (Thermo Fisher Scientific) che utilizza sonde *TaqMan*. Questo step permette di calcolare le pM (picoMoli) presenti nel campione incognito per interpolazione con la retta di taratura creata amplificando diluizioni seriali di un campione standard a concentrazione nota. Questa retta è la rappresentazione grafica dei valori di Ct generati per ciascuna curva del DNA standard in rapporto alla concentrazione logaritmica, considerando un intervallo di confidenza del 95%. Il protocollo prevede di:

1. Diluire le librerie 1:500 combinando 2 µl di template e 998 µl di *Nuclease-free Water* in una *96-well deep-well plate* nuova. Chiudere la piastra, vortexare e centrifugare.
2. Preparare tre diluizioni seriali (6.8 pM, 0.68 pM, e 0.068 pM) degli standard di *E. coli DH10B Ion Control Library* per la qPCR come in Tabella 7:

Tabella 7. Diluizione degli Standard

STANDARD	CONTROL LIBRARY	NUCLEASE FREE WATER	DILUTION FACTOR	CONCENTRATION
1	5 µl E.coli	45 µl	1:10	6.8 pM
2	5 µ Std 1	45 µl	1:100	0.68 pM
3	5 µ Std 2	45 µl	1:1000	0.068 pM



3. Preparare la *PCR Master Mix* per la qPCR considerando che per ogni campione occorrono:
  - 5 µl di *2X TaqMan™ Master Mix*
  - 1 µl di *20X Ion TaqMan™ Assay*
4. Dispensare la *PCR Master Mix* in una nuova 96-well deep-well plate e unire i campioni diluiti (volume finale = 20 µl):
  - 11 µl di *PCR Master Mix*
  - 9 µl di campione diluito 1:500
5. Settare *HID Real-Time PCR Analysis Software* (version 1.3) per il layout della qPCR, nell'*AB7500 Real-Time PCR System* (Thermo Fisher Scientific) come illustrato in Tabella 8.

Tabella 8. Impostazione protocollo di qPCR

STAGE	TEMPERATURE	TIME
Hold (UDG incubation)	50°C	2 minuti
Hold (polymerase activation)	95°C	20 secondi
Cycle (40 cycles)	95°C	1 secondo
	60°C	20 secondi

6. Una volta ottenuti i risultati della quantificazione, se necessario, normalizzare le librerie in modo da avere una concentrazione ottimale di 50 pM per ognuna, quantità di templato ottimale per la performance dell'esperimento:

$C_{qPCR}$ =concentrazione ottenuta da qPCR

$C_{real}$ =concentrazione reale della libreria

F=fattore di diluizione

$$C_{real} = 500 * C_{qPCR}$$

$$F = C_{real}/50$$

7. Combinare 5 µl di ogni libreria normalizzata in un'unica provetta da 1.5 ml *Eppendorf LoBind™ tube* (Thermo Fisher Scientific) allo scopo di creare il pool delle librerie.

### 3.5.3. PREPARAZIONE DEL TEMPLATING

La preparazione del templating, ovvero gli step di amplificazione clonale effettuata mediante *emulsion-PCR*, l'arricchimento con le particelle *Ion Sphere™ Particles* (ISP) e il caricamento del chip, è stato eseguito da *Ion Chef™* in modo automatizzato. Le librerie sono state caricate su *Ion 540™ Chip Kit* (Thermo Fisher Scientific). Per questa fase è stato necessario:

1. Creare la **Planned Run Template** e la **Planned Run del Torrent Suite™ Software** (versione 5.12.3) impostando le specifiche come da protocollo. Nello specifico, alcune informazioni di fondamentale importanza sono:
  - Tipo di pannello utilizzato per la creazione delle librerie (*Ion AmpliSeq™ Microbime Health Research Kit*).
  - Numero di campioni analizzati in associazione con i rispettivi *barcodes*.
  - Tipo di chip utilizzato selezionato sulla base della lunghezza degli ampliconi e del *coverage* che si vuole ottenere.
2. Caricare *Ion S5™ Chef* con i consumabili contenuti nel kit di *templating Ion S5™ Chef Supplies* (Thermo Fisher Scientific) seguendo le istruzioni del macchinario (Figura 14). In particolare, sono necessari:
  - *Ion 540™ Chef Reagents*, cartuccia che va scongelata per 45 minuti prima dell'uso in cui sono presenti 4 provette. Le prime due dotate di *QRcode* sono quelle in cui sono stati messi i pool di librerie, la terza contiene NaOH; la quarta è vuota per diluizioni o mix che fa lo strumento durante il processo.
  - *Ion S5™ Chef Solution*, cartuccia che contiene le soluzioni necessarie per la preparazione del *templating*.
  - *Ion S5™ Chef Supplies*, che a sua volta contiene gli adattatori del chip, *Enrichment Cartridge v2*, *Tip Cartridge v2*, *PCR Plate*, *Frame Seal v2*, *Recovery Station Disposable Lid v2*, *Recovery Tube v2*
  - *Ion 540™ Chip*

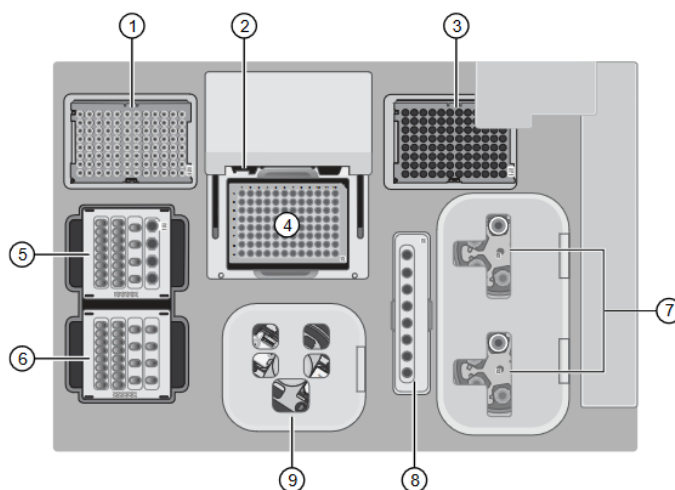


Figura 14. Ion S5 Chef instrument (Thermo Fisher Scientific 2021 - MAN0018668)

- |  |   |
|--|---|
| 1. Puntali per pipette nuovi                                 | 6. Stazione soluzioni   |
| 2. Coperchio riscaldato                                      | 7. Centrifughe di recupero – eseguono la centrifugazione delle Ion Sphere™ Particelle |
| 3. Rack per puntali di scarto (usati)                        | 8. Stazione di arricchimento  |
| 4. Blocco campione termociclatore per piastra da 96 pozzetti | 9. Centrifuga di caricamento dei chip   |
| 5. Stazione reagenti   |   |

3. Trasferire 25 µl di pool di librerie nel primo tubino di *Ion 540™ Chef Reagents*.
4. Azionare *Ion S5™ Chef* (scegliendo l'opzione *Caricamento del chip*) e attendere il tempo necessario affinché avvenga l'amplificazione e l'arricchimento del template, nonché il caricamento del chip (circa 17 ore).

#### 3.5.4. PREPARAZIONE DELLO STRUMENTO PER IL SEQUENZIAMENTO

Per il sequenziamento del chip viene utilizzato *Ion S5™ Sequencer* (Thermo Fisher Scientific). La piattaforma *Ion Torrent* misura la differenza di pH causata dal distacco di uno ione idrogeno dai nucleotidi che vengono aggiunti al filamento in esame e, in questo modo, garantisce la lettura della sequenza genetica.

La fase preliminare di preparazione dello strumento dura circa 45 minuti, mentre il sequenziamento vero e proprio dura circa 2 ore. Il protocollo prevede di:

1. Eseguire l'inizializzazione dello strumento selezionando **Settings - Clean Instrument**. Quindi svuotare la *waste reservoir*, inserire una nuova *Ion S5™ Wash Solution*, *Ion S5™ Cleaning Solution (Ion S5™ Sequencing Solutions – Thermo Fisher Scientific)*, *Ion S5™ Sequencing Reagents* (che vanno scongelati 20 minuti in anticipo) e un chip usato. I *Sequencing Reagents* contengono anche i nucleotidi necessari all'analisi, mentre la *Wash Solution* serve per i lavaggi che lo strumento effettua dopo ogni flusso di nucleotidi.
2. Estrarre il chip da *Ion S5™ Chef* e scartare l'eventuale liquido in eccesso presente nel pozzetto di caricamento.
3. Trasferire il chip su *Ion S5™ Sequencer* e far partire il sequenziamento.

#### 3.5.5. ANALISI DEI DATI CON ION REPORTER SOFTWARE ONLINE

La prima analisi è effettuata dalla *Torrent Suite* (TS) che, a mano a mano che vengono eseguiti i flow di nucleotidi, genera un file con le specifiche di caricamento/loading e sequenze utilizzabili. Il TS software è organizzato sulla base delle tre fasi principali del sequenziamento: il *Planning*, il *Monitor* e i *Dati*.

Il *Plan* racchiude sia i modelli di esperimento riutilizzabili, sia le *planned runs*, ovvero le istruzioni eseguibili per il singolo sequenziamento. Selezionando il modello sperimentale riutilizzabile è possibile selezionarne i dettagli come l'applicazione, il riferimento, i file BED, il progetto, i *plug-in* e le destinazioni di esportazione per i file dei risultati.

Il monitor favorisce la visualizzazione dello stato del sistema e del lavoro in esecuzione, inclusi i grafici di qualità che forniscono informazioni quasi in tempo reale sulle corse.

Una volta completata l'operazione, i dati ottenuti vengono visualizzati attraverso il riepilogo dell'esecuzione, il rapporto sulle corse e i risultati dei *plug-in*. A tal proposito, il software permette anche di scaricare i file di output e scaricare il report della corsa.

Per ogni campione vengono generati files in formato BAM che contengono l'elenco delle *reads* mappate con il loro allineamento al genoma di riferimento (solitamente associati a file BAI, file riassuntivi delle informazioni contenute nei file BAM) e in formato FASTQ, generati tramite l'uso di un *plug-in*, che riportano le chiamate nucleotidiche per ogni sequenza, associate al valore della qualità di lettura.

Le analisi successive vengono eseguite mediante *Ion Reporter™ Software online version* (Thermo Fisher Scientific), una versione gratuita *Ion Reporter™ Software*.

Per questo step la procedura prevede di:

1. Scaricare i file formato BAM dalla *Torrent Suite (TS)*.
2. Caricare i medesimi file, adeguatamente nominati, in *Ion Reporter™ Software* con la seguente procedura: **Samples - Define Sample – Manual – Upload BAM**.
3. Selezionare i due file BAM relativi alla *Target Species* e al 16S del medesimo campione e cliccare **Add to Sample**.
4. Nominare il campione che a questo punto conterrà i due file prima menzionati e cliccare **Add to Sample List**.
5. Assegnare ad ogni file il **Bacterial Marker Type** (rispettivamente *Target Species* o *16S*). Quest'ultima operazione è di fondamentale importanza per poter effettuare l'analisi.
6. Nella sezione **Workflow** selezionare **Microbiome Analysis** con il workflow name **AmpliSeq Microbiome Health – vv 1.0**, quindi modificare i parametri di analisi in modo da adattarli alle condizioni sperimentali adottate. Rinominare il nuovo workflow come **AmpliSeq Microbiome Health – vv 1.0\_custom**.
7. Far partire l'analisi contemporanea di tutti i campioni caricati, come da protocollo: **Analyses – Launch Analyses – Manual** – nella sezione **Workflow** selezionare **AmpliSeq Microbiome Health – vv 1.0\_custom** – nella sezione **Samples** selezionare i campioni precedentemente aggiunti alla lista – confermare e lanciare l'analisi nominando ogni singola analisi in modo tale da poterle distinguere una volta terminato il processo.

I risultati sono visibili e scaricabili nella sezione: **Analyses – Overview**.

### 3.6. ANALISI STATISTICA CON R-STUDIO

Una volta ottenuta la composizione delle famiglie batteriche all'interno dei campioni analizzati, corredata di abbondanza relativa e numero di *reads*, è stata eseguita un'analisi statistica mediante *R-Studio v4.2.2* con lo scopo di supportare le osservazioni effettuate sui dati di output (Watanabe, *et al.*, 2018). Si tratta di un ambiente di sviluppo integrato per il linguaggio di programmazione R che,

attraverso una serie di pacchetti specifici, consente l'elaborazione dei dati per effettuare calcoli statistici e ottenere rappresentazioni grafiche.

L'analisi applicata al microbioma, in questo caso, consente di sviluppare una metodica per l'identificazione personale effettuando valutazioni sulla variabilità tassonomica che intercorre tra due diversi campioni (in questo caso traccia da contatto e donatore). Nello specifico, R-Studio è stato utilizzato per calcolare la distanza di Canberra per coppie di campioni (Watanabe, *et al.*, 2018), attraverso la normalizzazione della differenza assoluta tra le abbondanze relative ad ogni taxon:

$$D_{canberra}(X, Y) = \sum_{i=1}^n \frac{|X_i - Y_i|}{|X_i| + |Y_i|}$$

in cui  $X_i$  e  $Y_i$  rappresentano le abbondanze, espresse in numero di *reads*, per ogni taxon contenuto nei campioni rispettivamente X e Y. Minore è la distanza di Canberra, maggiore è la similarità tra la composizione tassonomica dei due campioni messi a confronto.

La distanza di Canberra applicata all'identificazione personale è stata calcolata sul numero di *reads* prodotte dal sequenziamento del microbioma cutaneo umano sfruttando uno *script* di R (Watanabe, *et al.*, 2018) ([https://github.com/yamada-lab/human\\_skin\\_personal\\_identification](https://github.com/yamada-lab/human_skin_personal_identification)). Questa tecnica particolarmente innovativa è stata selezionata perché consente di effettuare l'analisi statistica sfruttando anche i taxa con abbondanza relativa inferiore all'1% che, secondo recenti studi, aumenterebbero la risoluzione del risultato. Per prima cosa è stata creata una *OTU Table*, ovvero una tabella in formato *.xlsx*, contenente la composizione microbica organizzata per famiglie batteriche dei campioni da analizzare. Per ogni famiglia è stato riportato il corrispondente numero di *reads* estrapolato dai dati di output di *Ion Reporter Software*. Successivamente, sono stati scelti i *reference sample* contro cui confrontare i *query sample* mediante il calcolo della distanza di Canberra, per stabilire una correlazione tra una traccia e il suo potenziale donatore.

Questa distanza viene calcolata:

$$D_{individual}(q, r_{individual}) = \frac{\sum_{i=1}^n D_{canberra}(q, r_i)}{n}$$

in cui  $q$  e  $r_i$  rappresentano rispettivamente la composizione tassonomica del *query sample* e del *reference sample*. Il campione di riferimento con il valore minimo di  $D_{individual}$ , confrontato con il *query sample* dovrebbe, plausibilmente, essere il proprietario del *query sample* stesso.

I risultati ottenuti sono, in seguito, stati rappresentati graficamente.

#### 4. RISULTATI E DISCUSSIONE

In questo studio preliminare sono stati estratti, quantificati e sottoposti a sequenziamento con tecniche NGS sia le tracce prelevate dal colletto della camicia e dal volante della macchina, sia i campioni di confronto prelevati dai palmi delle mani e dal collo, relativi ai soggetti presi in esame.

I campioni (Tabella 1), estratti con colonnine in silice, sono stati quantificati mediante strumentazione *Invitrogen Qubit 4 Fluorometer* (Thermo Fisher Scientific) con il *ready-to-use assay buffer* e il *DNA standard* presenti nel *Qubit™ IX dsDNA HS Assay kit* (Thermo Fisher Scientific). Le concentrazioni ottenute hanno una media di 0,1 ng/μl (Tabella 9), inferiore rispetto ai dati riscontrati in letteratura che riportano una media di 5 ng/μl (Procopio, *et al.*, 2021). È comunque possibile osservare una certa omogeneità nella quantità di DNA batterico estratto dai prelievi, sia per quanto riguarda i campioni prelevati sulla cute, sia per quanto riguarda i campioni prelevati sulle superfici riceventi. Nonostante si tratti di low template DNA, questa evidenza potrebbe indicare che, fino a questo momento, il processamento è avvenuto in parallelo senza subire bias analitici campione-correlati.

Tabella 9. Concentrazioni dei campioni post-estrazione ottenute mediante strumentazione *Qubit*.

NUMERO CAMPIONE	CONCENTRAZIONE (ng/μl)	NUMERO CAMPIONE	CONCENTRAZIONE (ng/μl)
F1	0.0788	I1	0.0524
F2	0.104	I2	0.285
F3	0.111	I3	0.0928
F4	0.106	I4	0.0876
G1	0.237	L1	0.120
G2	0.127	L2	0.174
G3	0.086	L3	0.176
G4	0.0668	L4	0.0904
H1	0.128	V1	0.062
H2	0.130	M	0.086
H3	0.131	N	0.0908
H4	0.0796	O	0.110
		NTC	0.0424

Per le successive fasi di amplificazione e sequenziamento sono stati utilizzati due approcci analitici differenti: la tecnologia Illumina tramite l'utilizzo del sequenziatore *MiSeq™* e la tecnologia *Ion Proton* tramite sequenziatore *Ion S5™* (Thermo Fisher Scientific).

Per quanto riguarda il sequenziamento Illumina, è stata seguita la procedura prevista dal protocollo *16S Metagenomic Sequencing Library Preparation*, ideata e generalmente utilizzata per studi sul microbioma intestinale in campo clinico-sanitario, che risulta molto più abbondante rispetto a quello cutaneo. È stato utilizzato questo protocollo perché, in commercio, non sono presenti kit validati specificamente per il microbioma cutaneo né ad uso clinico, né ad uso forense. Il protocollo prevede di utilizzare una concentrazione di materiale genetico batterico di partenza pari a 5 ng/μl, quantità necessaria affinché l'appaiamento dei primer e l'innesco della reazione di amplificazione avvengano senza incorrere in effetti

stocastici. Come è possibile osservare dalla Tabella 9, nessuno dei campioni analizzati raggiunge la concentrazione minima di DNA richiesta. Nonostante ciò, la procedura analitica è stata portata avanti effettuando alcuni adattamenti volumetrici atti ad ottimizzarla per poterla applicare in modo efficace allo studio del microbioma cutaneo.

La fase seguente consiste nell'allestimento delle librerie genomiche utilizzando *Nextera XT Index Kit* (Illumina) per l'indicizzazione e *MiSeq v2 reagent kit* (Illumina) per il sequenziamento. Come da specifiche del protocollo, una volta preparate le librerie, è stata effettuata una seconda quantificazione per poter tarare le quantità di reagenti da utilizzare per i successivi step di analisi, in base alle concentrazioni dei campioni. Infatti, le librerie sono state unite in un unico pool di librerie utilizzando volumi di NaOH (per la denaturazione del DNA) e di HT1 (per l'interruzione della denaturazione e la diluizione delle librerie) rapportati in funzione della quantificazione. In seguito, i campioni sono stati caricati sullo strumento *MiSeq™* per il sequenziamento vero e proprio. Durante quest'ultima fase lo strumento ha interrotto la procedura, dopo alcuni cicli dal suo inizio, fornendo come output una notifica di errore.

I medesimi campioni (Tabella 1) con le stesse concentrazioni (Tabella 9), sono stati utilizzati per un secondo esperimento effettuato per mezzo della tecnologia *Ion Torrent* (Thermo Fisher Scientific). Per la costruzione delle librerie genomiche è stato utilizzato *Ion AmpliSeq™ Microbime Health Research Kit* (Thermo Fisher Scientific), costituito da due pool di primer per l'amplificazione selettiva rispettivamente della regione 16S rRNA (*16S rRNA Gene Pool*) e di regioni specie specifiche (*Target Species Pool*) atte al riconoscimento di 73 ceppi batterici.

Alla luce delle concentrazioni degli estratti, ottenute mediate Qubit™ e consultabili in Tabella 9, che risultano ottimali per l'analisi della regione 16S (1 ng di DNA) ma esigue per l'analisi specie specifica (10 ng di DNA), non è stato necessario eseguire alcuna diluizione nella fase antecedente al sequenziamento. I campioni sono stati utilizzati tal quali per la costruzione delle librerie con entrambi i pool e successivamente sono state quantificate mediante qPCR (Tabella 10) e sequenziate per mezzo di *Ion Proton S5™*.

Tabella 10. Concentrazioni delle librerie ottenute mediante strumentazione qPCR, espressa in ng/ul

CAMPIONE	TEC. ILLUMINA	TEC. ION TORRENT		CAMPIONE	TEC. ILLUMINA	TEC. ION TORRENT	
		TARGET SPECIE	16S			TARGET SPECIE	16S
F1	0.0564	1,704	0,0897	I1	0.101	0,2037	0,0114
F2	0.0532	0,426	0,159	I2	0.070	0,1116	0,7504
F3	0.0504	0,0220	0,012	I3	0.0732	0,0121	0,1303
F4	0.0592	0,0897	0,3599	I4	0.0684	0,0226	0,1942
G1	0.0668	0,419	0,1034	L1	0.0756	0,00383	0,0104
G2	0.0556	0,6485	0,0349	L2	0.0744	0,007	0,0322
G3	0.0592	0,0318	0,00344	L3	LT	0,0198	0,0516
G4	0.0528	0,0246	0,0372	L4	0.060	0,00598	0,0363
H1	0.0675	0,0058	0,0201	V1	0.060	0,1572	0,0995
H2	0.0596	0,07096	0,0071	M	0.060	0,07495	0,028
H3	0.0564	0,005	0,1424	N	0.0632	0,1161	0,3353
H4	0.0712	0,155	0,528	O	0.058	0,0211	0,0366
				NTC	0.0468	0,0881	0,0426

Una volta terminato il sequenziamento, il *Torrent Suite (TS) Software* ha generato un *Run Report* che fornisce l'andamento e la qualità della corsa tenendo in considerazione vari parametri. I valori ottenuti in output dal sequenziamento, utili per le fasi successive dell'esperimento, sono stati riportati tra parentesi a fianco di ogni voce menzionata o all'interno del rispettivo paragrafo:

- *ISP density*

Riassume le caratteristiche della distribuzione delle ISPs sul chip Ion 540™ effettuata da *Ion Chef™* (Thermo Fisher Scientific) e i segnali che vengono registrati in fase di sequenziamento. La principale voce da consultare è *ISP Loading* (in questo caso 11%) che indica la percentuale di pozzetti del chip che contengono ISPs legate al DNA amplificato. Questo valore, per essere accettabile, in situazioni standard e validate, dovrebbe essere maggiore del 50%. Un caricamento inferiore potrebbe essere determinato da errori durante la fase di sequenziamento per mancato appaiamento dei primer o disfunzioni della DNA polimerasi. Ad affiancare questo valore, una rappresentazione grafica e colorimetrica della distribuzione delle biglie. La scala di colori rispecchia lo spettro della luce visibile e va dal rosso, che indica il massimo caricamento possibile, al blu (Figura 15).

- *ISP summary*

Riassume schematicamente i parametri che hanno favorito la selezione delle *reads* utilizzabili per il sequenziamento, tra cui le *Usable reads*, ovvero quelle effettivamente utili una volta applicati i filtri per le ISPs policlonali, per la qualità e per i dimeri di adattatori. In questo caso specifico, le *reads* utilizzabili sono il 23% del totale prodotto, questo valore risulta fortemente influenzato dalla loro scarsa qualità (Figura 15).

- *Read Length*

Riporta sottoforma di istogramma la lunghezza media delle *reads* (espressa in bp) sequenziate in relazione al *coverage* (Figura 15). Nel caso specifico di questa tesi, il protocollo di sequenziamento (Thermo Fisher Scientific - MAN0018602) indica come lunghezza media degli ampliconi 100-150 bp, pertanto, i segnali corrispondenti a lunghezze maggiori o minori possono rappresentare aspecifici di PCR, ISPs policlonali sfuggite al controllo o concatenazioni di DNA formatesi a seguito di eventi stocastici. Come è possibile vedere dalla Figura 15, la lunghezza delle *reads* prodotte dal sequenziamento ha un picco molto consistente in corrispondenza delle 100 bp, pertanto, gli ampliconi prodotti, in media, rispettano le previsioni sulla lunghezza. Tuttavia, rimane presente un rumore di fondo inferiore a 50 bp, probabilmente dovuto ad un mancato appaiamento dei primer (Figura 15).



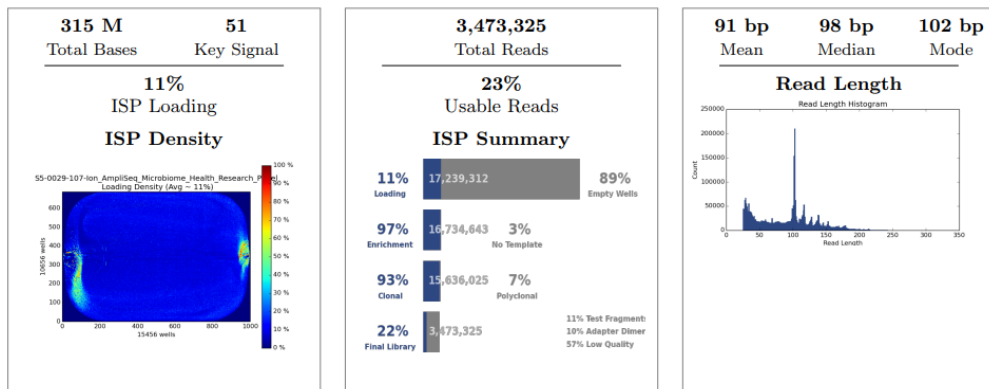


Figura 15. Riassunto di ISP density, ISP summary e Read Length, estrapolato dal Run Report prodotto dal Torrent Suite Software dopo il sequenziamento.

- Riassunto esito sequenziamento per campione

Il *Run Summary* prodotto dalla TS contiene anche l'elenco dei campioni analizzati, associati ai rispettivi *barcodes*. I *barcodes* sono stati scelti con un numero progressivo che va da 001 a 050 e i campioni sono stati nominati, in fase di sequenziamento, allo stesso modo. I campioni da 1 a 25 sono stati sottoposti a sequenziamento con l'utilizzo del *Target Species Pool*, mentre quelli da 26 a 50 sono stati sequenziati con *16S rRNA Gene Pool*. Per ognuno di questi è stato riportato il numero, la lunghezza media e la distribuzione grafica delle lunghezze relative alle *reads* ottenute.

Un altro parametro fornito in output dalla TS, per ogni campione, è il numero di basi con *Predicted Quality*  $\geq Q20$ , ovvero una precisione di chiamata corrispondente al 99%. Questo valore, confrontato con il numero totale di basi chiamate, permette di verificare la buona riuscita del sequenziamento e di essere confidenti nel fatto che le sequenze in output siano corrette e non soggette ad errori.

- Test Fragment

L'ultima informazione reperibile attraverso il *Run Report* riguarda il *Test Fragment* (Figura 16) che rappresenta il controllo di sequenziamento. Dall'istogramma riportato in Figura 16 si intuisce che il sequenziamento di TF\_1 è avvenuto correttamente, con una qualità del 94% espressa come percentuale di *reads* con un minimo di 50 bp e un phred score di 17 (tasso di errore 1 su 50, considerato accettabile da protocollo Thermo Fisher Scientific - MAN0017598).

Test Fragment	Reads	Percent 50AQ17	Read Length Histogram
TF_1	540,461	94	

Figura 16. Output Test Fragment estrapolato dal Run Report.

• Alignment Summary

Una volta effettuato il sequenziamento, la TS allinea le sequenze ottenute con un pannello specifico per il microbioma *Microbiome health panel reference* e successivamente valuta la qualità dell'allineamento (Figura 17). In questo esperimento, sono state allineate l'80% delle basi rilevate. Considerando che, in condizioni ottimali, questo valore non dovrebbe essere inferiore a 95%, ci si aspetta che alcuni campioni abbiano prodotto *reads* di bassa qualità che non sono state allineate; pertanto, il *coverage* complessivo risulterà minore rispetto alle aspettative, probabilmente per la presenza di inibitori, contaminanti o aspecifici.

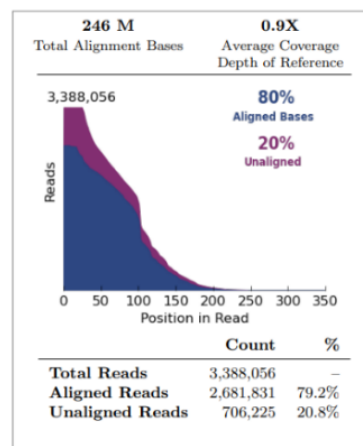


Figura 17. Alignment Summary estrapolato dal Run Report.

Una prima osservazione dei risultati del sequenziamento (mostrati in Figura 18 prendendo come esempio un solo campione) evidenzia l'esito non ottimale dell'analisi effettuata con *Target Species Pool* (campioni 1 – 25). Il fatto che questa analisi non sia andata a buon fine è intuibile dalla distribuzione grafica della lunghezza delle *reads* che in tutti i campioni risulta inferiore a 50 bp oppure omogeneamente distribuite lungo l'asse delle ascisse del relativo istogramma. Al contrario, in condizioni ottimali, ci si aspetterebbe di ottenere un picco ad una precisa lunghezza le sequenze prodotte dovrebbero essere tutte uguali. Le distribuzioni ottenute riflettono una mancata amplificazione o un'amplificazione soggetta ad errori stocastici, probabilmente dovuta alla scarsa quantità di materiale genetico di partenza (la quantità richiesta per questi primer era nettamente superiore a quanto possibile rilevare da tracce di interesse forense). Considerata la scarsa qualità dei risultati di sequenziamento, si è deciso di non proseguire con l'analisi dei dati riferiti a questo pool di primer.

Barcode Name	Sample	Bases	≥ Q20	Reads	Mean Read Length	Read Length Histogram
IonXpress.017	Sample 17	6,242,228	5,168,539	85,089	73 bp	

Figura 18. Run Summary esemplificativo dei campioni da 1 a 25 (*Target Specie Analysis*) estrapolato dal Run Report.

Per quanto riguarda l'analisi con *16S rRNA Gene Pool* (campioni 26 – 50), è stato possibile ottenere, da alcuni campioni, dati utili per le successive analisi informatiche mediante l'utilizzo di *Ion Reporter Software* (mostrati in Figura 19 prendendo come esempio un solo campione). I parametri che hanno consentito di selezionare i campioni idonei al proseguimento dell'esperimento sono stati definiti dall'operatore sulla base delle informazioni presenti in letteratura. Nello specifico i campioni 27, 28, 29, 30, 31, 35, 36 38, 40, 41, 44, 45, 46, 48, 49 (evidenziati in giallo in Tabella 11) hanno:

- Numero di *reads* superiore a 900 (Williams, *et al.*, 2019). Si tratta di un valore indicativo di quante volte un amplicone è stato letto durante il sequenziamento. Ottenere questa numerosità è necessario per avere una quantità di segnale sufficiente per essere captato e per coprire il segnale delle *reads* che contengono errori.
- Lunghezza media delle *reads* attorno alle 100-150 bp, indicativamente fornita dal kit di sequenziamento *Ion AmpliSeq™ Microbime Health Research Kit* (Thermo Fisher Scientific)
- Istogramma della distribuzione delle lunghezze delle *reads* con un evidente picco tra 100-150 bp che indica l'amplificazione selettiva della regione desiderata con limitati effetti stocastici dovuti all'appaiamento dei primer
- *Quality score*  $\geq$  Q20 che comprenda l'intero numero di basi rilevate durante il sequenziamento. Lo scarto esistente tra il numero di basi totali per campione e il numero di basi che hanno superato il controllo di qualità  $\geq$  Q20 rappresenta l'ammontare di basi che possono essere state lette in modo errato dallo strumento.


Barcode Name	Sample	Bases	$\geq$ Q20	Reads	Mean Read Length	Read Length Histogram
IonXpress_035	Sample 35	47,184,840	39,440,164	443,618	106 bp	

Figura 19. Run Summary esemplificativo dei campioni da 26 a 50 (*16S rDNA Analysis*) estrapolato dal Run Report.

I campioni numero 32, 34, 37 e 39 (evidenziati in verde in Tabella 11), pur non raggiungendo il numero minimo di *reads* richiesto dal protocollo, rispettano gli altri parametri di qualità, pertanto, dopo attenta valutazione dell'operatore sono stati ritenuti utili e si è deciso di sottoporli comunque alle successive analisi.

I campioni 26, 42 e 43 (evidenziati in blu in Tabella 11) sono caratterizzati da un numero medio di basi per *reads* inferiore a 100 bp. Tuttavia, grazie al loro elevato numero di *reads* registrate, superiore rispetto al limite prestabilito, e per via della presenza di un picco ben definito a 100bp nel grafico della distribuzione della lunghezza delle *reads*, nel caso del campione 43, sono stati ritenuti idonei al proseguimento delle analisi.

Il campione 50 (NTC) rappresenta il controllo negativo di estrazione, che è stato portato avanti per tutto il flusso di lavoro e sottoposto ad analisi informatiche al pari dei campioni che hanno superato il controllo qualità. NTC è stato quantificato dopo l'estrazione, evidenziando una concentrazione di 0,0424 ng/μl, paragonabile ad altri campioni. In fase di quantificazione, non è stato possibile risalire alla natura dell'eventuale contaminazione in quanto lo strumento *Qubit<sup>TM</sup>* utilizza come rilevatore un intercalante del DNA a doppia elica che non consente di distinguere tra batterico, umano o altro. Dal *Run Report*, invece, è stato possibile apprezzare la negatività del campione in quanto le 807 *reads* prodotte risultano di lunghezza inferiore a 50bp, con una media di 33 bp. Analisi successive effettuate mediante *Ion Reporter Software* hanno confermato tale ipotesi, pertanto, è stato possibile escludere la presenza di contaminazioni batteriche. Alla luce di questa evidenza, la piccola quantità di materiale genetico, rinvenuta in fase di quantificazione è stata plausibilmente attribuita ad una contaminazione di origine non batterica, alla presenza di residui di DNA umano che non sono stati adeguatamente eliminati durante la fase di estrazione o ad un errore dello strumento causato dall'estrapolazione dei valori considerati.

I restanti campioni (33 e 47, non evidenziati in Tabella 11) non hanno superato i parametri necessari per poter proseguire con le analisi informatiche e non sono più stati presi in considerazione. Nello specifico, il campione 33, ha una lunghezza media di *reads* inferiore a 50 bp ed un *coverage* inferiore a 900, caratteristiche che non gli hanno consentito di superare il controllo qualità utile per la buona riuscita dell'esperimento. Da queste osservazioni si può dedurre che l'amplificazione delle regioni *target* del campione 33, che da protocollo dovrebbero aggirarsi attorno alle 100-150 bp, non sia andata a buon fine.

Tabella 11. Riassunto caratteristiche campioni analizzati per il 16S rDNA. In giallo sono evidenziati i campioni che hanno superato il controllo qualità, in verde quelli che non raggiungono il numero minimo di reads ma superano gli altri parametri di qualità, in blu quelli che non hanno superato il controllo qualità ma contengono un numero di reads molto elevato e in bianco quelli che non hanno superato il controllo qualità per alcun parametro e che, pertanto, non sono stati presi in considerazione per l'analisi secondaria.

NUMERO CAMPIONE	DESCRIZIONE	READS	MEDIA (bp)	NUMERO CAMPIONE	DESCRIZIONE	READS	MEDIA (bp)
26 (F1)	F mani	46729	41	32 (I1)	I mani	177	89
27 (F2)	F collo	1106	88	33 (I2)	I collo	727	41
37 (F3)	F volante	273	86	43 (I3)	I volante	57596	62
38 (F4)	F colletto	4405	102	44 (I4)	I colletto	102893	99
28 (G1)	G mani	418119	110	34 (L1)	L mani	889	69
29 (G2)	G collo	26394	107	35 (L2)	L collo	443618	106
39 (G3)	G volante	358	107	45 (L3)	L volante	27199	106
40 (G4)	G colletto	24057	106	46 (L4)	L colletto	11286	80
30 (H1)	H mani	10271	105	36 (L5)	V mani	3022	87
31 (H2)	H collo	1629	103	47 (M)		1284	57
41 (H3)	H volante	71093	104	48 (N)		116261	84
42 (H4)	H colletto	66124	51	49 (O)		8403	85
				50 (NTC)	negativo	807	33

I file BAM relativi ai campioni selezionati (26, 27, 28, 29, 30, 31, 32, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 48, 49, 50) sono stati analizzati mediante il software *Ion Reporter online versione v18.8*. Le sequenze batteriche ipervariabili del 16S rDNA, contenute all'interno dei suddetti file, sono state allineate con il

database di riferimento consigliato dal protocollo di analisi (*Curated Greengenes v13.5*), con lo scopo di ottenere un elenco dettagliato delle famiglie e dei generi batterici contenuti in ogni campione, corredati delle rispettive frequenze relative. Unitamente alla composizione microbica, vengono fornite anche informazioni in merito alle sequenze iper-variabili che sono state utilizzate per l'allineamento. Da una prima osservazione di questo output, è stato possibile apprendere l'esito del controllo di qualità del campione (*Pool QC*), il numero totale di *reads* (*Total Reads*), la lunghezza media espressa in bp (*Mean Read Length*) e il numero di quelle effettivamente idonee ad essere utilizzate (*Total Valid Mapped Reads*). In molti dei campioni analizzati queste informazioni hanno consentito di riscontrare una discrepanza tra il numero totale di *reads* e il numero totale di quelle ritenute "valide" dal software, anomalia

probabilmente dovuta ad una selezione effettuata per mezzo dei parametri settati nel workflow analitico.

Trattandosi di uno studio preliminare e innovativo, soprattutto mediante l'utilizzo della tecnologia *Ion Torrent*, l'analisi secondaria dei dati ha previsto diversi step di settaggio per ottenere un workflow ottimale. I parametri di interesse, con i relativi valori selezionati, appartenenti al *Microbiome Analysis workflow*, sono stati riportati in Tabella 12 e corredati di definizione.

Tabella 12. Parametri workflow customizzato per Microbiome Analysis in Ion Reporter

PARAMETRO	VALORE	DEFINIZIONE
Minimum Local Alignment Score	25	Lunghezza minima di regioni locali con alto livello di somiglianza.
Minimum Normalized Read count for Family, Genus and Species	0.25 0.25 0.12	Valore soglia oltre il quale la Famiglia, il Genere e la Specie batterica viene riportata come PRESENT.
Minimum read count for Family, Genus and Species	80 80 20	
Minimum Read Length e Maximum Read Length	80-350	Range di lunghezza delle reads all'interno del quale dovrebbero trovarsi le sequenze di interesse. Per eliminare aspecifici.
Minimum value per species per signature.	5	Minimo valore di firme (sequenze di riconoscimento) che ha ogni specie batterica all'interno del 16S.
Minimum count per signature	100	Minimo numero di reads per firma batterica appartenente al 16S.
Minimum fraction of signature match	0.20	Frazione minima ottenuta da firme osservate fratto attese. Quindi minore è questo valore, minore è il conteggio delle firme osservate rispetto alle attese.
Minimum total counts per sequence	5	
Minimum count per signature for 16S Analysis rescue step e minimum fraction on unique counts per signature for 16S Analysis rescue step.	50 0,20	Conteggi totali minimi e conteggi unici minimi per firma batterica per l'analisi 16S per la fase di ripristino in base alla percentuale di conteggi univoci.
Minimum total valid mapped reads per library for 16S Pool QC (1) e minimum mean read length for 16S Pool QC (0).	100 6	Per il quality control delle reads e consiste nel valore minimo oltre di reads valide e di lunghezza di sequenza per considerare il controllo qualità PASS (passato).

Il software *Ion Reporter* consente, inoltre, di visualizzare i taxa batterici attraverso rappresentazioni grafiche e alberi tassonomici interattivi.

Una prima panoramica sugli esiti di questo lavoro sperimentale consente di dedurre che la tecnologia Illumina sia meno efficace rispetto all'analogia *Ion Proton* di *Thermo Fisher Scientific* nell'analisi del microbioma cutaneo umano, in quanto la prima non ha prodotto alcun output di sequenziamento. Una possibile spiegazione a questo problema è da ricercarsi nella discrepanza tra la quantità minima di materiale richiesto dal protocollo e la quantità di materiale genetico di partenza sia al momento dell'amplificazione delle sequenze bersaglio, sia al momento del caricamento sulla cartuccia di sequenziamento. Va ricordato, inoltre, che i reagenti e la metodologia utilizzati in questo lavoro di tesi sono stati settati sul microbioma intestinale in ambito sanitario, e che non sono ancora stati ottimizzati per il microbioma cutaneo in ambito forense; anche questo aspetto può aver influito sull'esito dell'esperimento.

Dai risultati ottenuti con la tecnologia Ion Proton, successivi all'analisi secondaria svolta con *Ion Reporter*, si evince che i campioni di confronto prelevati direttamente sulla cute umana sono stati processati meno efficacemente rispetto ai campioni delle tracce prelevati sulle superfici, nonostante le concentrazioni iniziali fossero paragonabili. I campioni prelevati dai palmi delle mani hanno fornito l'effettiva composizione batterica per i soggetti F, G e H (Tabella 11). Le variabili che possono aver influito sul risultato possono essere ricondotte allo stile di vita dell'individuo, all'utilizzo più frequente, considerato il periodo storico, di gel igienizzante o al semplice lavaggio delle mani con acqua e sapone; nonché la presenza, sulla cute, di sostanze, ad oggi non ancora note, che possono aver inibito il processo di sequenziamento solo per il DNA batterico prelevato dalla cute umana. Il medesimo ragionamento può essere fatto per i campioni prelevati dalla pelle in prossimità del collo. Anche in questo caso, le operazioni di igiene personale e le sostanze fisiologiche proprie della cute, possono aver interferito con l'analisi microbica del distretto esaminato, consentendo la produzione di dati interpretabili solo per il soggetto G e dati di scarsa utilità per i soggetti H e L (Tabella 11). Il campione V, prelevato sui palmi delle mani dell'operatore con lo scopo di utilizzarlo come campione di esclusione per eventuali contaminazioni che possono essere avvenute durante il workflow analitico, non ha prodotto dati sufficienti per adempiere al suo scopo, pertanto, non è più stato tenuto in considerazione per le operazioni successive.

Considerando, invece, i campioni delle tracce, prelevate sulle superfici inanimate, la situazione appare differente. Infatti, i prelievi effettuati sul volante e sul colletto hanno dato risultati positivi per 3 campioni su 5 (H, I, L e G, I, L rispettivamente). Questo fenomeno può trovare diverse spiegazioni plausibili:

- Background microbico della superficie ricevente.
- Contaminazioni ambientali.

- Frequente utilizzo dei substrati prima della sanificazione che può aver consentito l'accumulo e la proliferazione batterica.
- Substrato favorevole alla proliferazione.

I campioni riguardanti le misture (M, N e O), effettuati per ampliare il dataset e per provare ad introdurre lo studio dei misti nell'analisi del microbioma cutaneo a scopo forense, non hanno prodotto dati sufficienti per poter effettuare le opportune analisi comparative, pertanto i campioni, da qui in avanti non verranno più menzionati.

Per le considerazioni successive è stata utilizzata una statistica descrittiva, mediante osservazione e comparazione dei dati, relativi alle Famiglie batteriche, ottenuti come output dall'analisi secondaria effettuata per mezzo di *Ion Reporter*. In un secondo momento è stato assegnato un valore statistico alle osservazioni menzionate in precedenza, utilizzando le distanze di Canberra per i confronti tra le coppie di campioni.

#### 4.1. RICERCA DEL CORE MICROBIOME

Per prima cosa, in questo lavoro di tesi si è cercato di individuare un *Core Microbiome* rilevabile in tutti i soggetti esaminati, ovvero la presenza di famiglie batteriche caratteristiche del microbiota cutaneo, avvalendosi del software *Ion Reporter* e del confronto con i dati, ancora preliminari, presenti in letteratura. In tutti i campioni cutanei che hanno prodotto un risultato analizzabile, sia per quanto riguarda i prelievi effettuati sul collo (G collo – 29, H collo – 31, L collo – 35), sia per quanto riguarda quelli effettuati sui palmi (F mani – 26, G mani – 28, H mani – 30), sono state rilevate famiglie batteriche ricorrenti (evidenziate in arancione in Allegato 1) quali: *Propionibacteriaceae*, *Corynebacteriaceae*, *Staphylococcaceae* e *Streptococcaceae*. Queste osservazioni risultano coerenti con quelle riportate in letteratura da Ruuskanen (Ruuskanen, *et al.*, 2022), nel cui lavoro evidenzia la presenza di questa stessa composizione microbica nei distretti cutanei presi in considerazione, ad indicazione del fatto che, il microbioma cutaneo, presenta alcune famiglie batteriche peculiari che permettono di differenziarlo da quello riscontrabile in altre matrici biologiche di partenza.

In Tabella 13 è stata riportata la composizione del *Core Microbiome* cutaneo individuato in questo lavoro di tesi, con annessi i range di abbondanza relativa a cui vengono riscontrati i vari taxa, ovvero il valore percentuale minimo e massimo di prevalenza all'interno della popolazione microbica.

Tabella 13. *Core Microbiome* cutaneo umano

FAMIGLIE	RANGE ABBONDANZA RELATIVA SUL COLLO	RANGE ABBONDANZA RELATIVA SUI PALMI
Propionibacteriaceae	34.98 % - 92.49 %	34.36 % - 50.93 %
Corynebacteriaceae	2.54 % - 9.49 %	4.95 % - 17.49 %
Staphylococcaceae	1.05 % - 6.02 %	6.8 % - 12.31%
Streptococcaceae	0.42 % - 4.94 %	2.40 % - 8.65 %

Tra le famiglie batteriche riportate in Tabella 13 ci sono:

- *Propionibacteriaceae*, che presentano differenti classi, tra cui *Propionibacterium*, un gram-positivo bastoncellare anaerobico in grado di sintetizzare acido propionico che funge da parassita facoltativo o commensale in uomo e animali. Sono microrganismi onnipresenti e innocui per gli ospiti che si localizzano in prossimità delle ghiandole sudoripare e sebacee della pelle.
- *Staphylococcaceae*, tra cui *Staphylococcus*, sono batteri gram-positivi e catalasi positivi che vivono in condizioni di anaerobiosi. In natura colonizzano prevalentemente la pelle degli animali a sangue caldo e dell'uomo. Alcuni membri di questa famiglia possono fungere da agente eziologico di numerose patologie cutanee nell'uomo.
- *Streptococcaceae*, batteri gram-positivi che includono il genere *Streptococcus*. Molti streptococchi possono vivere in condizioni di anaerobiosi e sono conosciuti per la loro presenza caratterizzante nel microbioma salivare umano.
- *Corinebacteriaceae*, tra cui *Corynebacterium*, sono batteri gram-positivi bastoncellari, aerobi, che parassitano piante, animali e uomini. Molti corinebatteri fanno parte, infatti, della flora commensale dell'uomo.

Un aspetto interessante risiede nell'abbondanza relativa delle suddette famiglie batteriche. In particolare, quelle riscontrate a livello del collo presentano le medesime proporzioni in tutti i campioni analizzati. Viceversa, le stesse trovate sui palmi delle mani hanno proporzioni variabili da individuo a individuo. Questo fenomeno potrebbe trovare spiegazione nel fatto che i due distretti vengono a contatto con l'ambiente esterno in modo completamente differente. Per quanto riguarda il collo, si tratta di un distretto che verosimilmente è difficilmente accessibile ad agenti contaminanti che possono arricchirne la flora batterica residente, inoltre risulta anche meno a contatto con superfici riceventi a cui poter trasferire i batteri; di conseguenza non avvengono fluttuazioni significative dell'abbondanza relativa. Per quanto riguarda i palmi, invece, come già accennato, esiste una maggiore possibilità di esposizione e trasferimento, nonché un utilizzo di agenti igienizzanti che possono influire sulla composizione microbica e causarne fluttuazioni personalizzate in base alle attività quotidiane.

#### **4.2. VERIFICA VARIABILITÀ INTRAINDIVIDUALE**

Dopo aver cercato di definire il *Core Microbiome* cutaneo con il fine di poterlo discriminare da quello appartenente ad altre matrici umane, è stata ricercata l'esistenza di variabilità microbica tra mani e collo, confrontando il microbioma ottenuto da prelievi effettuati sui diversi distretti cutanei di tutti gli individui e, successivamente, in uno stesso individuo. Questo aspetto risulta particolarmente interessante perché, se ogni distretto corporeo dovesse possedere una propria nicchia batterica, si potrebbe aprire la possibilità di descrivere il tipo di contatto



avvenuto tra il donatore e la sorgente ricevente anche in applicazione ad analisi di tipo forense.

I risultati ottenuti dal confronto di tutti i campioni relativi al collo ha mostrato la ricorrenza delle famiglie batteriche appartenenti al *Core Microbiome* (evidenziate in arancione in Allegato 1). In quanto studio iniziale, in letteratura non è segnalata la composizione del microbioma cutaneo del collo. È stato, quindi, stilato un elenco di famiglie batteriche rilevate su questa superficie corporea (Tabella 13) che riflette quello caratteristico della cute umana nel suo insieme. Queste osservazioni risentono della debolezza dovuta alla limitata disponibilità dei dati impiegati per questa analisi. Ciononostante, il presente studio può fungere da modello pionieristico per il futuro nell'ipotesi che, con un numero maggiore di campioni, sia possibile ottenere evidenze più approfondite in merito all'utilizzo del microbioma cutaneo del collo in ambito forense.

I risultati ottenuti dal confronto di tutti i campioni relativi ai palmi, oltre al *Core Microbiome*, hanno rilevato la presenza ricorrente di ulteriori famiglie batteriche, presenti in tutti gli individui (evidenziati in verde in Allegato 1). Tra queste ci sono:

- *Rhodobacteraceae*, gram-negativi, che colonizzano l'ambiente acquatico. Comprendono, tra i generi, il *Paracoccus* che si ritrova frequentemente nel suolo in condizioni di ossigeno ridotto.
- *Pasteurellaceae*, batteri commensali e potenzialmente patogeni per i mammiferi vertebrati (tra cui uomo) e gli uccelli. Presentano esponenti del genere *Haemophilus*, gram-negativi che possono causare infezioni lievi e gravi nell'uomo.
- *Micrococcaceae*, di cui fa parte il genere *Micrococcus*, sono cocchi immobili gram-positivi e catalasi positivi che abitano la cute e le mucose sia animali che umane.
- *Sphingomonadaceae*, tra cui *Sphingomonas*, gram-negativi strettamente aerobi, la cui presenza risulta di particolare interesse. Sono veicoli di endotossine e sono caratterizzati da un involucro cellulare è costituito da glicosfingolipidi utili a conferire resistenza alle sostanze antibatteriche. Le sfingomonadi includono basi nucleotidiche distintive all'interno del gene codificante il *16S rRNA*. Sono ampiamente distribuite in natura perché sono in grado di sopravvivere a basse concentrazioni di nutrienti e a metabolizzare una numerosa varietà di fonti di carbonio (tra cui anche composti tossici). Alcuni esponenti di questo taxa sono coinvolti in infezioni nosocomiali nell'uomo.

Al contrario del microbioma del collo, quello dei palmi delle mani è stato ampiamente studiato e, in letteratura, sono riportati diversi dati utili per poter effettuare un confronto. In questo modo è stato possibile dapprima paragonare le famiglie batteriche ottenute da ogni campione alla ricerca di una composizione batterica specifica per i palmi delle mani e, successivamente, confrontare tale risultato con quelli già pubblicati da Edmonds-Wilson et al., (Edmonds-Wilson et

al., 2015). I batteri presenti in questa regione corporea rispecchiano sia i taxa comuni a tutta la cute (evidenziati in arancione in Allegato 1), sia alcuni taxa specifici (evidenziati in verde in Allegato 1). Il confronto con i dati presentati da Edmonds-Wilson et al., che riassumono la composizione microbica delle mani rilevata in tutti gli studi precedenti, conferma i risultati ottenuti.

Una volta effettuata un'osservazione generale sulla composizione microbica di tutti gli individui selezionati, discriminando le regioni corporee, è stata effettuata la stessa analisi limitatamente ai singoli soggetti alla ricerca di variabilità intra individuale. Con i dati ottenuti dal sequenziamento, è stato possibile analizzare le differenze tra il microbioma cutaneo dei vari distretti corporei esclusivamente per i campioni prelevati al soggetto G, che risultano idonei secondo gli standard di qualità e pertanto sono confrontabili. Nei due campioni, i taxa condivisi appartenenti al *Core Microbiome* (evidenziati in arancione in Allegato 2a) presentano comunque differenze significative per quanto riguarda l'abbondanza relativa. Per esempio, le *Staphylococcaceae*, hanno un'abbondanza relativa dell'11.8% sulle mani, e del 6.16% sul collo. Oltre al *Core Microbiome*, sono presenti famiglie batteriche che in un distretto risultano abbondanti e caratteristiche, mentre nell'altro risultano marginali, talvolta con abbondanza relativa inferiore all'1%.

Alcune famiglie batteriche (*Pasteurellaceae*, *Porphyromonadaceae*, *Fusobacteriaceae*, *Neisseriaceae*, *Enterobacteriaceae*, *Flavobacteriaceae* e *Peptoniphilaceae*), evidenziate in blu in Allegato 2a, sono maggiormente rappresentate a livello delle mani. La differenza più significativa è indicata dal taxon delle *Pasteurellaceae*, che compone il microbioma delle mani per il 9.35%, mentre quello del collo per solo l'1.57%. Viceversa, taxa come *Sphingomonadaceae*, *Micrococcaceae*, *Pseudomonadaceae*, *Moraxellaceae* e *Rhodobacteraceae* (evidenziati in viola in Allegato 2a), sembrano essere più abbondanti all'interno del microbiota del collo. Questo aspetto è interessante, soprattutto per quanto riguarda le famiglie delle *Rhodobacteraceae*, *Sphingomonadaceae* e *Micrococcaceae* che, inizialmente, sono state ricondotte alla composizione microbica tipica delle mani.

Infine, sono state riscontrate alcune famiglie batteriche esclusivamente attribuibili al microbiota delle mani (*Peptoniphilaceae* evidenziata in giallo in Allegato 2a), e alcune riconducibili esclusivamente al microbiota del collo, tra cui *Intrasporangiaceae* e *Caulobacteraceae* che hanno abbondanza relativa >1% (evidenziati in verde in Allegato 2a). Altre famiglie batteriche, evidenziate in verde in Allegato 2a, sono state rilevate esclusivamente a livello del collo ma, a differenza delle precedenti, con un'abbondanza relativa <1%. Degna di menzione è soprattutto la famiglia delle *Oxalobacteraceae*, un batterio patogeno opportunista per l'uomo.

Gli altri taxa riportati in tabella e non evidenziati risultano condivisi tra le due regioni corporee del medesimo individuo sia per quanto riguarda la loro presenza,

sia per l'abbondanza con cui sono rappresentati, motivo per cui non sono utili per la determinazione della variabilità intra-individuale.

#### 4.3. VERIFICA VARIABILITÀ INTERINDIVIDUALE

In seguito, si è provato a verificare la presenza di variabilità interindividuale, in termini di microbioma cutaneo, al fine di valutarne un possibile utilizzo nel corso di indagini forensi per l'identificazione personale. Con i dati ottenuti, è stato possibile effettuare una valutazione del microbioma cutaneo esclusivamente relativo alle mani dei soggetti F, G e H (consultabile in allegato 2b). In prima analisi, il microbioma cutaneo delle mani sembra avere alcune famiglie batteriche ricorrenti che caratterizzano la suddetta regione in tutti gli individui analizzati (evidenziate in verde in Allegato 1 e in verde in Allegato 2b). Tuttavia, ogni individuo presenta anche taxa batterici specifici che, potenzialmente, potrebbero consentirgli di differenziarsi dagli altri:

- Il microbioma cutaneo delle mani dell'individuo F risulta caratterizzato da taxa non riscontrati negli altri campioni analoghi, tra cui *Lactobacillaceae* e *Caulobacteraceae* che presentano un'abbondanza relativa superiore all'1% e sono stati evidenziati in viola in Allegato 2b.
- Il microbioma cutaneo proprio delle mani dell'individuo G, evidenziato in indaco in Allegato 2b, è caratterizzato dalle *Fusobacteriaceae*, *Flavobacteriaceae*, *Porphyromonadaceae* e *Enterobacteriaceae* che presentano un'abbondanza relativa superiore all'1%. Altre famiglie batteriche presenti solo nel campione delle mani di G, invece, sono rappresentati da un'abbondanza relativa inferiore all'1%: *Carnobacteriaceae*, *Pseudomonadaceae*, *Prevotellaceae*, *Anaeroplasmataceae*, *Leptotrichiaceae*, *Peptostreptococcaceae* e *Nocardoidaceae*.
- Il microbioma cutaneo delle mani dell'individuo H, evidenziato in fucsia in Allegato 2a, è caratterizzato dalle famiglie delle *Lachnospiraceae*, *Ruminococcaceae*, *Aerococcaceae*, *Succinivibrionaceae*, *Erysipelotrichaceae* e *Thermoanaerobacteraceae*, che presentano un'abbondanza relativa inferiore all'1%.

Particolarmente degna di menzione è la famiglia delle *Neisseriaceae*, gram-negativi commensali dei mammiferi e aerobi che fanno parte dei proteobatteri, il cui genere *Neisseria* è stato rilevato sia nel soggetto G (abbondanza relativa >1%), sia nel soggetto H (abbondanza relativa <1%). Grazie all'albero filogenetico circolare interattivo, fornito come output dal software di analisi secondaria *Ion Reporter* (Figura 20), è stato possibile indagare più a fondo la presenza di questo taxa batterico. Per prima cosa è stata osservata la mappa filogenetica relativa ai palmi del soggetto G (Figura 20) in cui i proteobatteri rappresentano in totale il 3% di tutti i batteri rilevati, di cui il 20% è occupato dalle *Neisseria*. Il campione in questione

racchiude numerose specie appartenenti a questa famiglia, come ad esempio *N. meningitidis*, *N. flavescens*, *N. subflava* e *N. mucosa*.

In seguito, è stata studiata la mappa filogenetica relativa ai palmi del soggetto H (Figura 21) in cui i *Proteobacteria* rappresentano l'1% dei taxa totali, tra questi, il 24% di *Neisseria*. Tra le specie di spicco, è stato possibile osservare, nel suddetto campione, *N. Snodgrassella alvi* che è rappresentata da un'abbondanza relativa superiore all'1%. Si tratta di un ceppo particolarmente interessante perché, di norma, colonizza l'intestino delle api mellifere motivo per cui non è mai stato precedentemente segnalato in letteratura come componente della flora microbica cutanea dell'uomo (Edmonds-Wilson *et al.*, 2015). Una possibile spiegazione a questo fenomeno è l'assunzione o il contatto, da parte del soggetto esaminato, di sostanze che possono contenere il suddetto microrganismo.

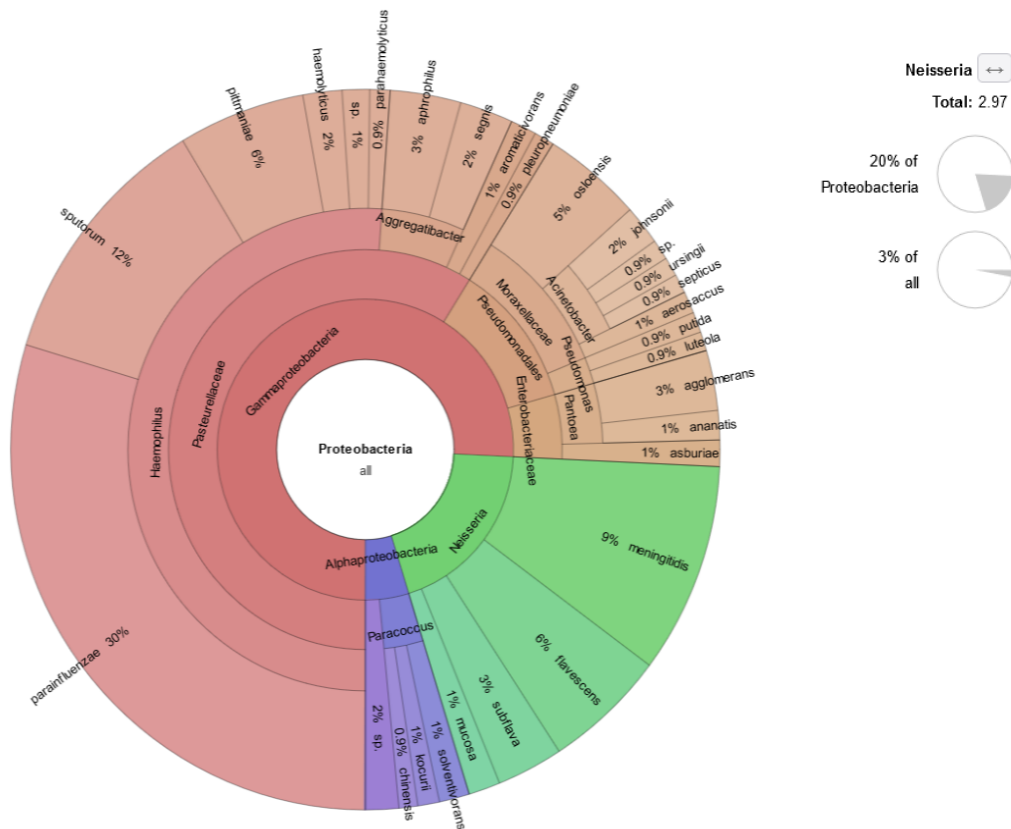


Figura 20. Rappresentazione grafica interattiva della famiglia delle Neisseriaceae e dei suoi componenti riscontrati nel campione prelevato dalle mani del soggetto G (colorata in verde) ottenuta mediante Ion Reporter Software.



Figura 21. Rappresentazione grafica interattiva della famiglia delle Neisseriaceae e dei suoi componenti riscontrati nel campione prelevato dalle mani del soggetto H (colorata in verde ed indicato con la freccia rossa) ottenuta mediante Ion Reporter Software.

Nel presente lavoro di tesi, i campioni cutanei sono stati prelevati nell'arco di una mattinata, prestando attenzione a mantenere una condizione il più possibile simile a quella potenzialmente riscontrata nei casi reali ed evitando l'introduzione di variabili che avrebbero potuto complicare le valutazioni finali. La presenza di variabilità microbica, infatti, è da ritenersi correlata all'unicità del microbioma cutaneo di ciascun individuo, fermo restando che l'analisi svolta in questo esperimento è di tipo preliminare e si basa su osservazioni effettuate su un ridotto numero di campioni.

#### 4.4. RICERCA CORRISPONDENZA TRA TRACCIA E DONATORE

Oltre alla discriminazione personale, una delle analisi fondamentali in campo forense, effettuata di routine anche mediante DNA umano attraverso l'utilizzo degli STR, è la ricerca di corrispondenza tra una traccia, lasciata su una superficie ricevente, ed il relativo donatore. Tuttavia, qualora non fosse disponibile il DNA umano o fosse eccessivamente degradato, il microbioma potrebbe fornire una potenziale soluzione per riuscire ugualmente a ricollegare i due campioni al fine di supportare le indagini forensi. Il presente studio, in modo preliminare e pionieristico, è stato pensato per valutare la possibilità di utilizzare i batteri della

pelle per ricercare una eventuale corrispondenza (match) tra donatore della traccia e *Touch Microbiome*, rilasciato sulla superficie ricevente al momento del contatto.

Nonostante le difficoltà incontrate durante il sequenziamento, e la limitata quantità di dati ottenuti alla fine delle analisi, è stata comunque eseguita una valutazione del possibile match sui campioni G collo-colletto e H mani-volante che sono stati ritenuti idonei al controllo qualità.

#### Match G collo - G colletto

Paragonando i taxa batterici, è stata riscontrata correlazione tra la traccia prelevata sul colletto della camicia del soggetto G e il confronto prelevato sul collo del medesimo soggetto (proprietario della camicia). Tutte le famiglie batteriche trovate sul colletto sono state riscontrate anche sulla relativa regione cutanea (evidenziati in giallo in Allegato 3a), appartenente alla persona che le ha depositate, fatta eccezione per le *Peptoniphilaceae* (evidenziati in viola in Allegato 3a) che sono state rilevate, con un'abbondanza relativa inferiore all'1%, esclusivamente sulla superficie ricevente probabilmente perché appartenenti al microbioma ambientale o dovute a qualche forma di contaminazione. Al contrario, tutti i taxa presenti solo sulla cute, che non trovano corrispondenza in quelli della superficie ricevente, vengono rappresentati con un'abbondanza relativa inferiore all'1% (non evidenziati in Allegato 3a). Risulta quindi plausibile che la scarsa rappresentanza di questi microrganismi abbia costituito un ostacolo al proprio trasferimento, motivo per cui non sono state rilevate sulla superficie ricevente; oppure potrebbero essere essi stessi frutto di trasferimento da parte di oggetti terzi.

È invece interessante notare la variabilità nell'abbondanza relativa delle diverse specie batteriche sulla sorgente e sulla superficie ricevente. Nella maggior parte dei casi l'abbondanza relativa di una data famiglia batterica è paragonabile tra i due substrati o leggermente inferiore sulla superficie ricevente. Questo può essere spiegato con la scarsa trasferibilità del microrganismo, con la difficile adesione al substrato ricevente o con la modalità di contatto adottata. In quest'ottica, è degna di nota l'abbondanza relativa delle *Pseudomonadaceae* (evidenziata in rosso in Allegato 3a), pari a 30.19% sulla superficie ricevente e 4.6% sul collo del donatore. Questa evidenza può essere spiegata con la presenza pregressa di tali microrganismi sul colletto della camicia (trasferimento da colletto a collo), con la sua elevata trasferibilità, o con una spiccata proliferazione causata da condizioni ambientali favorevoli sulla superficie ricevente. Le *Pseudomonadaceae Pseudomonas* sono Bacilli gram-negativi, aerobi obbligati, mobili e non-fermentanti, che vengono considerati patogeni opportunisti. Avvalendosi delle rappresentazioni grafiche interattive fornite in output dal software *Ion Reporter* sono state individuate, nel dettaglio, le specie appartenenti al genere *Pseudomonas* presenti nel campione prelevato dal colletto del soggetto G (Figura 22). Tra queste, particolarmente interessanti sono: *P. putida* (facente parte della flora normale dell'orofaringe e del tratto respiratorio ma considerato un patogeno opportunisto, agente eziologico di

infezioni delle vie urinarie, di ferite e di malattia infiammatoria pelvica.), *P. stutzeri* (commensale del tratto respiratorio e urogenitale nell'uomo, in poche occasioni viene associato direttamente a processi infettivi) e *P. aeruginosa* (riscontrabile in condizioni normali nelle feci umane che possono rappresentare fonti di epidemie e di contaminazione cutanea, occasionalmente può trovarsi nelle zone più umide della cute e nella saliva.). Le medesime sono state rilevate anche sul collo del soggetto G (Figura 23). La peculiarità di queste specie batteriche, non propriamente riconducibili al microbioma cutaneo umano e la loro sovrapposizione tra i due campioni potrebbe, in sintesi, contribuire a rafforzare l'associazione tra donatore e superficie ricevente.



Figura 22. Rappresentazione grafica interattiva della famiglia delle Pseudomonadaceae e dei suoi componenti riscontrati nel campione prelevato dal collo del soggetto G (colorata in rosso/arancione) ottenuta mediante Ion Reporter Software.

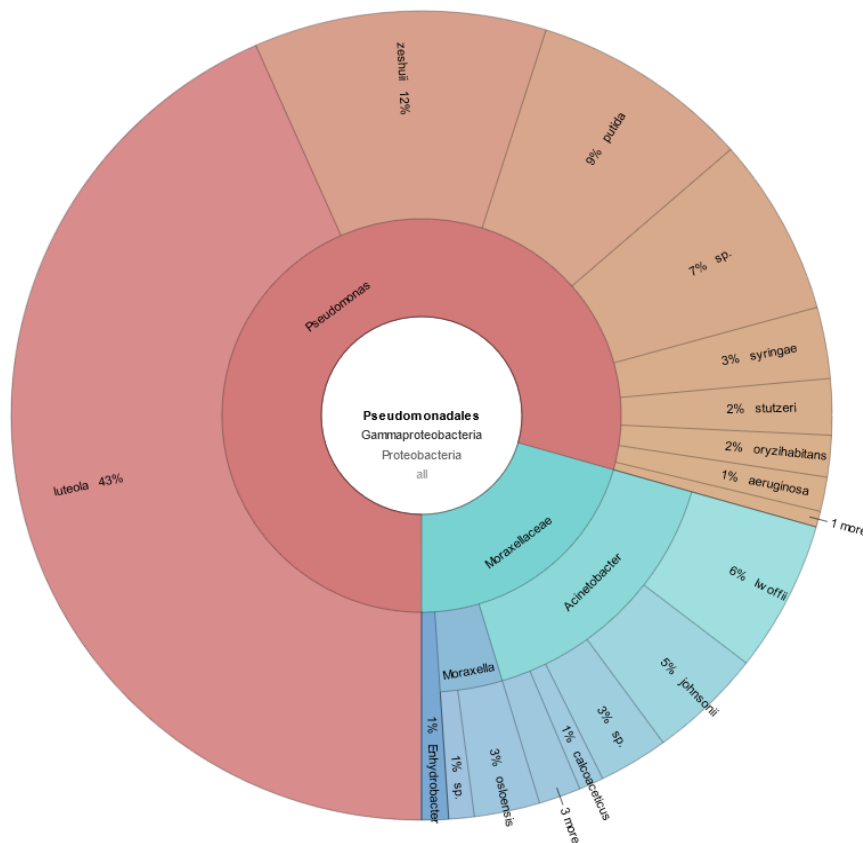


Figura 23. Rappresentazione grafica interattiva della famiglia delle Pseudomonadaceae e dei suoi componenti riscontrati nel campione prelevato dal colletto del soggetto G (colorata in rosso/arancione) ottenuta mediante Ion Reporter Software.

### Match H mani – H volante

Nel corso dell'analisi è stata valutata anche la corrispondenza tra il microbioma prelevato sul volante della macchina del soggetto H e il relativo campione di confronto prelevato dalle mani del medesimo soggetto. Tutti i taxa presenti sul campione di confronto sono stati ritrovati anche sulla traccia e sono stati evidenziati in giallo in Allegato 3b. Fanno eccezione le *Erysipelotrichaceae*, *Ruminococcaceae*, *Streptomyetaceae* e *Succinivibrionaceae*, che sono state rilevate solo sulla superficie delle mani del soggetto H, con un'abbondanza relativa inferiore all'1%. Probabilmente, il motivo per cui il trasferimento sul volante non è avvenuto in modo efficiente è da ricercarsi nella scarsa rappresentazione di questi taxa sulla cute. Più numerosi, invece, sono i taxa rinvenuti sul volante in uso al soggetto H. Per quanto riguarda quelli con abbondanza relativa inferiore all'1% (evidenziati in viola in Allegato 3b), la spiegazione può essere analoga a quella indicata per i campioni precedentemente analizzati; ci sono, però, anche taxa con abbondanza relativa maggiore che vanno a costituire a tutti gli effetti il microbioma del substrato. Questi ultimi sono stati evidenziati in verde in Allegato 3b ed appartengono alle famiglie delle *Prevotellaceae*, *Carnobacteriaceae*, *Anaeroplasmataceae*, *Pseudomonadaceae* e *Brevibacteriaceae*. La loro presenza



sul volante può essere giustificata con una contaminazione ambientale oppure può essere la conseguenza dell'utilizzo precedente dell'automobile da parte di altri utenti.

Anche dal confronto tra palmi delle mani di H e volante della mano di H, l'abbondanza relativa delle famiglie che risultano sovrapposte nei due campioni è paragonabile o comunque leggermente inferiore sulla superficie ricevente, fatta eccezione per *Neisseriaceae*, *Pasteurellaceae*, *Streptococcaceae* e *Veillonellaceae*, la cui maggior rappresentazione a livello del substrato ricevente può, ancora una volta, essere spiegata con la presenza pregressa di tali microrganismi sul volante della macchina (trasferimento da palmo a volante), con l'elevato grado di trasferibilità o con una spiccata proliferazione causata da condizioni ambientali favorevoli sulla superficie ricevente.

In ogni caso, sia per quanto riguarda i campioni G collo – G colto, sia per quanto riguarda i campioni H mani – H volante, è possibile utilizzare i dati forniti in output dal software *Ion Reporter* per osservare una sovrapposizione tra i taxa batterici delle due coppie di campioni. A causa del limitato numero di campioni a disposizione e del limitato numero di repliche per ogni misurazione è difficile riuscire ad affermare l'esistenza di un'associazione univoca. In questo caso, è quindi raccomandabile focalizzare l'attenzione su specifici taxa batterici che, per intrinseche proprietà, non siano comuni nella popolazione e nelle zone studiate e che per questo possano indirizzare l'operatore verso un'associazione traccia – donatore più robusta. Un altro aspetto da tenere in considerazione durante l'analisi descrittiva è la presenza di fluttuazione del microbioma, sulle superfici riceventi, nell'arco della giornata (Wilkins, *et al.*, 2021). Nello specifico caso di questo lavoro sperimentale si è cercato di raccogliere i campioni da analizzare nel medesimo arco temporale per limitare quanto più possibile eventuali variazioni microbiche che potevano verificarsi in tal senso. Ciononostante, il riscontro di taxa sulle superfici riceventi, non sovrapponibili con quelli appartenenti alle mani del relativo donatore, può essere correlato a questo tipo di variabilità non prevedibile e non controllabile.

Oltre ad una valutazione statistica descrittiva, è stato possibile calcolare gli indici di distanza tra i campioni definiti sia sulla base dei taxa batterici, sia sulla base dell'abbondanza relativa, che hanno permesso di supportare statisticamente le osservazioni fatte fino a questo punto. La possibilità di effettuare questo tipo di analisi secondaria sui dati a disposizione è stata di fondamentale importanza dal momento che, trattandosi di uno studio volto ad introdurre proposte per elementi di innovazione in campo forense, la mera osservazione di un'evidenza senza supporto statistico non può essere utilizzata.

#### 4.5. ANALISI STATISTICA DEI RISULTATI

Le osservazioni fatte fino ad ora, mediante confronto delle famiglie batteriche presenti in ogni campione che abbia superato il flusso di lavoro, sono state integrate attraverso un'analisi statistica effettuata su R-Studio per mezzo del calcolo della distanza di Canberra.

Lo scopo è di verificare l'eventuale correlazione, i campioni che hanno fornito dati interpretabili a seguito dell'analisi con *Ion Reporter Software* sono stati raggruppati per regioni corporee coinvolte nel trasferimento:

- F mani, H mani, H volante, I volante, L volante
- G collo, G colletto, H collo, I colletto, L colletto

Dopo una prima fase di preparazione dei dati, in cui ad ogni famiglia batterica è stato attribuito il numero di *reads* ottenuto dal sequenziamento di ogni campione, è stato applicato lo script descritto da Watanabe, et al., (Watanabe, *et al.*, 2018) nel suo recente articolo. Il risultato (Tabella 14, Tabella 15) mostra i valori della distanza di Canberra ottenuti dal confronto tra *query sample*, riportati sulle righe e *reference sample* riportati sulle colonne. Queste matrici sono utilizzate per produrre le rispettive rappresentazioni grafiche (*heatmap*) riportate sotto alla Tabella 14 e alla Tabella 15 che esprimono i medesimi risultati attraverso una scala colorimetrica. Ricordando che minore è il valore della distanza di Canberra, maggiore è la somiglianza tra i campioni confrontati, sulla diagonale delle tabelle è possibile apprezzare l'esito del confronto della composizione microbica di ogni campione con sé stesso che dà come risultato un valore pari a zero (evidenziate in arancione in Tabella 14 e 15). La medesima osservazione può essere fatta sui grafici, in cui maggiore è la somiglianza tra due campioni messi a confronto, maggiore è l'intensità del colore riportato sulla cella corrispondente. Anche in questo caso la diagonale rappresenta il confronto di un campione con sé stesso, motivo per cui risulta di un colore molto intenso ad indicare la massima similarità possibile.

Osservando le distanze di Canberra in Tabella 14, ottenute confrontando il primo cluster di campioni (F mani, H mani, H volante, I volante, L volante), è emerso che:

- Il campione F mani, confrontato con H mani, ha fornito un valore di Canberra maggiore rispetto al confronto di quest'ultimo con tutti gli altri campioni del cluster. Questo ha consentito di verificare la presenza di variabilità tra campioni cutanei appartenenti ad individui differenti. La stessa osservazione può essere fatta mediante il grafico corrispondente che mostra, per il confronto tra questi due campioni, un colore tendente al bianco (assenza di somiglianza).
- Il campione H mani, paragonato con il rispettivo volante (H volante), ha una distanza di Canberra inferiore rispetto al confronto effettuato contro tutti gli altri volanti appartenenti ad altri soggetti (tranne per volante di L). In questo modo si percepisce una sorta di correlazione tra la composizione microbica di una traccia

prelevata dalla superficie ricevente e il rispettivo donatore perché la distanza di Canberra risulta minore rispetto al confronto del donatore con campioni casuali. La colorazione arancione intermedio raffigurato in corrispondenza del confronto tra i due campioni analizzati confermata la somiglianza.

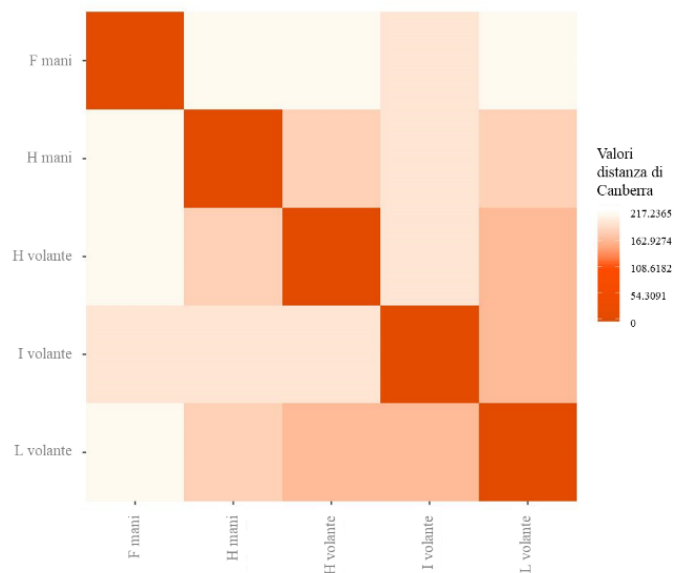
- Il campione F mani, con cui non è stato possibile paragonare il rispettivo volante, confrontato con tutti i volanti a disposizione ha ottenuto un indice di Canberra maggiore rispetto a quello riscontrato tra mani e volante del soggetto H. La stessa osservazione può essere fatta sulla base del relativo grafico che mostra una colorazione tendente al bianco, indice del fatto che non sia possibile ricondurre il soggetto F all'utilizzo di un volante casuale tra quelli analizzati.
- Il campione L volante, di cui non è disponibile il rispettivo prelievo cutaneo, sia paragonato con il riferimento cutaneo di H (H mani) sia confrontato con tutti gli altri prelievi effettuati sui volanti, ha fornito una distanza di Canberra inferiore rispetto alle altre coppie create in modo randomico. Correlazione inattesa, visivamente apprezzabile anche dal colore arancione intermedio del grafico corrispondente.
- Il campione I volante, di cui non è disponibile il rispettivo prelievo cutaneo, sia paragonato con il riferimento cutaneo di H (H mani) e di F (F mani) sia confrontato con il prelievo effettuato su H volante ha fornito un valore di Canberra inferiore rispetto alle altre coppie create in modo randomico. Dal grafico corrispondente è possibile osservare la leggera somiglianza tra i campioni in questione perché rappresentati da una debole colorazione arancione intermedia. Essendo molto debole, questa correlazione potrebbe essere ricondotta ad eventi stocastici che possono aver influenzato il sequenziamento dei campioni oppure ad una ridotta variabilità del microbioma presente nel campione I che lo rende parzialmente simile agli altri.

Trattandosi di uno studio preliminare in cui è stato preso in considerazione un numero limitato di campioni con una singola replica ognuno, per provare spiegare le corrispondenze inattese sono state formulate differenti ipotesi. La prima prende in considerazione il fatto che la composizione batterica dei volanti potrebbe aver risentito dell'influenza del microbioma ambientale, condiviso tra tutti i soggetti, con cui l'automobile può essere venuta a contatto durante l'utilizzo. Una seconda ipotesi potrebbe, invece, attribuire la somiglianza della composizione batterica sui volanti ad una proliferazione microbica preferenziale provocata dalla natura del materiale di cui è costituito il volante. Un'altra spiegazione riguarda la mancanza di dati relativi ai campioni cutanei dei soggetti I ed L. Dal momento che il workflow analitico non ha permesso di caratterizzare la composizione batterica di questi due campioni, non è possibile prevedere in che modo questi avrebbero condizionato l'analisi di Canberra qualora fossero stati disponibili. Il loro inserimento avrebbe potuto provocare una fluttuazione delle distanze tale da ridefinire l'assetto del match per tutti i campioni raggruppati nel cluster prescelto. In ultima analisi, è

plausibile ipotizzare che questi campioni mancanti, per aver prodotto tracce da contatto simili tra loro, possano essere caratterizzati da una composizione microbica sovrapponibile data dal fatto che i soggetti presi in esame condividono il medesimo ambiente lavorativo e, di conseguenza, come noto dalla letteratura, questo potrebbe aver provocato una diminuzione della variabilità interindividuale del microbioma cutaneo (Wilkins, *et al.*, 2021).

Tabella 14: Distanze di Canberra per il confronto tra i campioni prelevati dalle mani e dal volante dei soggetti analizzati. In giallo sono evidenziati i valori di Canberra calcolati sulle coppie di campioni relativi ad H. Di seguito è riportata la rappresentazione grafica (heatmap) realizzata a partire dalla tabella, la scala colorimetrica varia da rosso (massima somiglianza) a bianco (minima somiglianza).

	<b>F mani</b>	<b>H mani</b>	<b>H volante</b>	<b>I volante</b>	<b>L volante</b>
<b>F mani</b>	0	204.145	217.236	194.808	213.381
<b>H mani</b>	204.145	0	181.715	184.762	177.917
<b>H volante</b>	217.236	181.715	0	185.847	158.930
<b>I volante</b>	194.808	184.762	185.847	0	165.655
<b>L volante</b>	213.381	177.917	158.930	165.655	0



Osservando le distanze di Canberra in Tabella 15, ottenute confrontando il secondo cluster di campioni (G collo, G colletto, H collo, I colletto, L colletto), è emerso che:

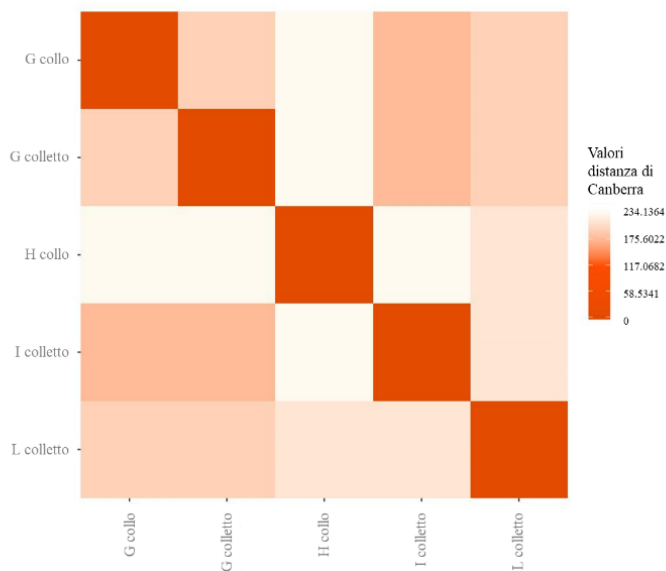
- Il campione G collo, confrontato con H collo, ha ottenuto un valore di Canberra maggiore rispetto agli altri confronti, ad indica una dissimilarità tra i due campioni. Conseguentemente, è stata quindi osservata una verosimile variabilità interindividuale nel microbioma cutaneo. Queste osservazioni sono confermate dalla colorazione tenue, tendente al bianco, del grafico corrispondente.
- Il campione G collo, paragonato con il rispettivo colletto (G colletto), ha ottenuto un valore di Canberra minore rispetto al confronto casuale effettuato contro tutti i colletti appartenenti ad altri soggetti (tranne per colletto di I). In questo modo, si percepisce una verosimile correlazione tra la composizione microbica di una traccia prelevata dalla superficie ricevente e il rispettivo donatore, visibile anche dalla colorazione arancione intermedia del grafico corrispondente.
- Il campione H collo, di cui non è disponibile il rispettivo prelievo sulla superficie ricevente, paragonato con tutti gli altri prelievi effettuati sui colletti, ha fornito, una distanza di Canberra maggiore rispetto a quella riscontrata tra collo e colletto di G. Sembrerebbe, di conseguenza, avere una composizione microbica differente da tutti i campioni parallelamente analizzati, come è possibile vedere anche dalla colorazione tenue, tendente al bianco, del grafico corrispondente.
- Il campione I colletto, di cui non è disponibile il rispettivo prelievo cutaneo, sia paragonato con il riferimento cutaneo di G (G collo) sia confrontato con il prelievo effettuato sul colletto di G (G colletto), ha fornito un indice di Canberra inferiore rispetto alle altre coppie create in modo randomico. Correlazione osservabile anche dal colore arancione intermedio derivante dal confronto grafico.
- Il campione L colletto, di cui non è disponibile il rispettivo prelievo cutaneo, sia paragonato con il riferimento cutaneo di G (G collo) sia confrontato con il prelievo effettuato sul colletto di G (G colletto), ha fornito un indice di Canberra inferiore rispetto alle altre coppie create in modo randomico ma superiore rispetto alle coppie effettivamente correlate. Infatti, il grafico corrispondente evidenzia, per questo confronto, una colorazione arancione intermedia-chiara.

Come nel caso precedente, sono state formulate alcune ipotesi per provare a dare una spiegazione ai valori di Canberra inattesi, riscontrati durante l'analisi statistica, non compatibili con l'attribuzione di un match univoco tra traccia da contatto e confronto cutaneo. Sono plausibili le ipotesi di una qualche contaminazione di tipo ambientale e di una proliferazione preferenziale di determinate famiglie batteriche, dovuta alla tipologia di tessuto con cui è stato realizzato il colletto della camicia. Un'ulteriore variabile che può aver influito sulla proliferazione di specifiche famiglie batteriche è stata suggerita dalla natura stessa della cute. Il collo, essendo una zona soggetta ad elevata sudorazione e secrezione sebacea, potrebbe essere

costituita da una composizione batterica simile anche in campioni provenienti da soggetti diversi. Inoltre, anche in questo caso non sono stati ottenuti i riscontri delle composizioni microbiche relative a tutti gli individui quindi l'analisi risulta parziale perché non è stato possibile stabilire con certezza la somiglianza tra i vari microbiomi del collo e le variazioni che avrebbero potuto comportare alle distanze statistiche, qualora fosse stato possibile inserirle nell'analisi.

Tabella 15: Distanze di Canberra per il confronto tra i campioni prelevati dal collo e dal colletto dei soggetti analizzati. In giallo sono evidenziati i valori di Canberra calcolati sulle coppie di campioni relative a G. Di seguito è riportata la rappresentazione grafica (heatmap) realizzata a partire dalla tabella, la scala colorimetrica varia da rosso (massima somiglianza) a bianco (minima somiglianza).

	<b>G collo</b>	<b>G colletto</b>	<b>H collo</b>	<b>I colletto</b>	<b>L colletto</b>
<b>G collo</b>	0	183.763	228.598	170.885	192.347
<b>G colletto</b>	183.763	0	223.109	176.268	196.875
<b>H collo</b>	228.598	223.109	0	234.136	215.094
<b>I colletto</b>	170.885	176.268	234.136	0	200.968
<b>L colletto</b>	192.347	196.875	215.094	200.968	0



## 5. CONCLUSIONI

Il microbioma umano è una delle nuove frontiere scientifiche così innovativa che, negli ultimi anni, ha aperto numerose opportunità in diversi ambiti, primo tra tutti quello clinico-sanitario. Questa linea di ricerca nascente ha mostrato un forte potenziale anche nell'ambito forense in cui è stata approfondita grazie alle moderne tecniche di sequenziamento già impiegate per le classiche indagini genetiche. Caratteristiche come la numerosità batterica, superiore a quella delle cellule umane, la variabilità genetica e l'unicità nei diversi individui rendono, questo tipo di analisi, ottimale e promettente in ambito forense anche se gli studi circa nel suo impiego sono ancora agli inizi.

Con il presente lavoro di tesi si è voluto evidenziare il potenziale che questo strumento potrebbe fornire come supporto delle canoniche analisi del DNA nelle indagini forensi, focalizzando l'attenzione sul microbioma cutaneo di mani e collo. I reperti più frequentemente rilevati sulla scena del crimine, infatti, sono costituiti da guanti, indumenti e oggetti di uso personale, correlati con le suddette regioni corporee. La scelta è stata suggerita da un precedente studio in letteratura in cui sono stati riassunti numerosi lavori che coinvolgono il microbioma cutaneo delle mani; viceversa, il microbioma del collo è stato studiato in questa tesi perché non ancora menzionato, a scopo forense, nei precedenti lavori in letteratura.

Lo sviluppo di tecnologie all'avanguardia, come il *Massive Parallele Sequencing* (MPS) ha consentito il progresso negli studi genetici, ampliando il campo della ricerca e le possibilità di analisi specifiche. Per questo esperimento, infatti, è stato utilizzato il sequenziamento mediante MPS delle regioni *16S rDNA* del genoma batterico. Si tratta di un metodo veloce, paragonato al sequenziamento shotgun dell'intero genoma batterico, e agile in termini di allineamento delle sequenze con le banche dati. Il medesimo approccio è stato precedentemente adottato anche per l'*Human Microbiome Project* al fine di costruire una banca dati ampia ed aggiornata relativa alla caratterizzazione batterica.

Trattandosi di uno studio preliminare risente ancora della mancanza di un metodo analitico unico e validato che comprenda tutte le fasi dal campionamento all'analisi secondaria. Di conseguenza, sono stati utilizzati protocolli e kit originariamente ideati per i batteri intestinali a scopo sanitario che richiedono abbondante materiale di partenza. Nonostante il campionamento sia stato realizzato seguendo le indicazioni riportate in letteratura, la successiva fase di estrazione effettuata con *QIAamp DNA Microbiome Kit* ha restituito basse quantità di materiale genetico (0,1 ng/μl). Studi paralleli hanno riscontrato una maggiore efficienza di estrazione con il kit *QIAamp PowerFecal Pro DNA Kit* (QIAGEN; Procopio, *et al.*, 2021). Per la successiva fase di sequenziamento è stata testata l'efficacia di due differenti tecnologie MPS: Illumina con lo strumento *MiSeq<sup>TM</sup>* e Ion Proton con lo strumento *Ion S5<sup>TM</sup>*. Attualmente non sono disponibili in commercio kit specifici per analisi di microbioma in ambito forense; per questo motivo ne sono stati testati due con lo

scopo di individuare quello più efficacemente adattabile a questo tipo di indagine. I risultati ottenuti hanno permesso di concludere che la tecnologia *Ion S5<sup>TM</sup>*, con il relativo kit *Ion AmpliSeq<sup>TM</sup> Microbime Health Research*, risulta essere più efficiente ed efficace per l'analisi del microbioma cutaneo rispetto alla tecnologia *MiSeq<sup>TM</sup>* e relativo *MiSeq Reagent kit v2*. Nello specifico, il kit di sequenziamento, *Ion AmpliSeq<sup>TM</sup> Microbime Health Research* è costituito da due pool di primer, uno per l'analisi della regione *16S rDNA* e uno per l'analisi di 73 specie batteriche. Il primo, che analizza 8 delle 9 regioni iper-variabili del segmento 16S batterico, ha consentito la produzione di dati paragonabili a quelli descritti in letteratura per studi analoghi, pur partendo da una quantità di DNA in input molto limitata (1 ng). Questa metodica, quindi, potrebbe dimostrarsi vantaggiosa nei casi di low template DNA che costituiscono la maggior parte delle tracce riscontrabili in casi forensi. Viceversa, il secondo pool, che dovrebbe permettere di rilevare 73 specie batteriche mediante amplificazione e confronto di sequenze *target* batteriche specie/ceppo specifiche con i dati presenti nei database, si è dimostrato poco efficace per via dell'elevata quantità di DNA batterico richiesta in input (10 ng) e non raggiunta dal materiale di partenza ottenuto dal campionamento (trattandosi appunto di tracce). Risulta, di conseguenza, difficilmente applicabile a casi reali in cui il *Touch DNA* è molto limitato e necessita di una maggiore sensibilità di rilevamento. A questo proposito, la costruzione di protocolli specifici per il microbioma cutaneo o l'affinamento dei protocolli attuali sono aspetti fondamentali su cui porre l'attenzione con lo scopo di ottenere un unico metodo riproducibile e ripetibile specifico per analisi forensi. Un altro aspetto da migliorare è l'inserimento di dati all'interno di database batterici ampliandoli sia in termini di numerosità della popolazione di riferimento, sia in termini di numerosità di regioni cutanee analizzate; inserendo, inoltre, informazioni legate alla variabilità del microbioma nel tempo e metadati associati agli individui esaminati (origine biogeografica, età, dieta, informazioni sociali) di cui, però, le pubblicazioni forensi sono ancora carenti.

Con i dati del sequenziamento MPS sono stati ottenuti i profili relativi a 12 campioni sui quali è stato possibile effettuare le seguenti valutazioni conclusive:

#### I. Ricerca del *Core Microbiome*.

Il confronto effettuato su tutti i campioni prelevati in corrispondenza della cute umana di individui diversi ha evidenziato la presenza di un microbioma comune e caratteristico denominato *Core Microbiome*; ovvero un insieme di taxa batterici condivisi tra i campioni esaminati. Il *Core Microbiome* è costituito da una serie di microrganismi, fondamentali per il corretto equilibrio e benessere cutaneo, in questo caso di palmi delle mani e collo, che vivono in simbiosi con l'ospite. La sua composizione in taxa e la rispettiva abbondanza relativa, risultano pressoché costanti e durature nel tempo, anche dopo esposizione ad agenti perturbanti provenienti dall'interno e dall'esterno dell'organismo. Lo studio del *Core Microbiome* non mira, quindi, a trarre



informazioni da sfruttare per identificazione personale, bensì potrebbe rivelarsi uno strumento ausiliario per la discriminazione della sorgente di una traccia riscontrata a seguito di un ipotetico evento criminoso. Trattandosi di batteri cutanei, è stato possibile descrivere alcune famiglie batteriche caratteristiche che sono state riscontrate sia sui palmi delle mani sia sul collo e che conservano approssimativamente le medesime proporzioni in tutti i campioni: *Propionibacteriaceae*, *Corynebacteriaceae*, *Staphylococcaceae* e *Streptococcaceae*.

## II. Verifica variabilità intra-individuale.

Il confronto tra campioni provenienti da diverse parti del corpo realizzato, dapprima, sulla totalità degli individui selezionati e, in seguito, su ogni singolo individuo (compatibilmente con i dati ottenuti dal sequenziamento) ha consentito di effettuare differenti osservazioni riguardo alla composizione microbica sito-specifica. Per prima cosa è stata osservata la ricorrenza nella composizione microbica relativa al collo di tutti i soggetti esaminati. In questo caso, questa composizione tassonomica rispecchia quella precedentemente definita nel *Core Microbiome* cutaneo complessivo. Saranno, quindi, necessari ulteriori studi basati su campioni cutanei più numerosi e su più repliche per ognuno di essi al fine di riuscire a definire con maggiore precisione una composizione caratteristica del microbioma del collo che consenta di essere individuato come sito donatore in modo univoco in un potenziale evento criminoso. La medesima osservazione è stata eseguita sui campioni microbici derivanti dai palmi delle mani di tutti i soggetti presi nel loro insieme. Seppur partendo da quantità di DNA batterico estratto molto esigue, non conformi a quelle richieste dai kit di sequenziamento, è stato possibile, individuare una composizione microbica ricorrente sulle mani che ha poi trovato riscontro anche nei dati riportati in letteratura da Edmonds-Wilson et al., (Edmonds-Wilson *et al.*, 2015).

In un secondo momento, sono stati messi a confronto i campioni relativi ad un unico individuo ma appartenenti a regioni corporee differenti. Escludendo i taxa appartenenti al *Core Microbiome*, sono state osservate differenze significative tra il microbioma dei due distretti, sia a livello di taxa microbici, sia a livello di abbondanza relativa. Il riscontro di variabilità intra individuale è un elemento di forte interesse per la comunità forense, in quanto potrebbe consentire di ipotizzare il tipo di contatto sulla base della diversità microbica, favorendo lo sviluppo di una *linkage theory* robusta a supporto delle indagini. Alcuni studi precedenti hanno verificato la variabilità batterica in relazione al pH, livello di idratazione, spessore cutaneo, rugosità, densità ghiandolare ed esposizione ambientale, andando così a spiegare le motivazioni che spingono determinati batteri a proliferare preferenzialmente su alcune regioni corporee.

### III. Verifica variabilità interindividuale.

Il confronto effettuato tra campioni cutanei di individui differenti ha messo in luce alcune differenze tra le composizioni microbiche che vanno così a costituire la variabilità interindividuale, la quale consentirebbe di distinguere un individuo dall'altro. Questo aspetto risulta fondamentale per una futura applicazione forense dell'analisi del microbioma. In linea teorica, secondo i dati riportati in letteratura, due individui indipendenti condividerebbero solo il 13 % dei taxa cutanei osservati, coerentemente con quanto osservato nel presente lavoro di tesi, ad indicare una probabile riproducibilità delle analisi. Nella maggior parte dei lavori riportati in letteratura, i taxa che presentano un'abbondanza relativa <1% vengono eliminati dalle analisi perché ritenuti molto spesso riconducibili a contaminazioni, ad aspecifici di sequenziamento o ad errori di appaiamento con le sequenze di riferimento (Ross, *et al.*, 2017). Tuttavia, esistono evidenze più recenti secondo cui i taxa batterici con bassa abbondanza relativa (<1%) possano contribuire più efficacemente all'identificazione personale rispetto ai taxa con abbondanza relativa più elevata. Taxa abbondanti, infatti, risulterebbero maggiormente soggetti al trasferimento e di conseguenza più comuni nella popolazione di riferimento, inficiandone l'utilità nella discriminazione individuale (Watanabe, *et al.*, 2018). Il presente esperimento di tesi sembrerebbe mostrare dati compatibili con la teoria di Watanabe, *et al.* (Watanabe, *et al.*, 2018), nel cui lavoro sperimentale viene incrementata l'accuratezza del match includendo anche i taxa con abbondanza relativa bassa.

### IV. Ricerca corrispondenza tra traccia e donatore.

Il confronto tra traccia contenente *Touch Microbiome*, microbioma trasferito dalla cute al substrato mediante contatto diretto e campione cutaneo prelevato dal donatore ha permesso di rilevare una sovrapposizione tra i batteri presenti nei campioni. Tuttavia, nei casi in cui l'associazione univoca tra i campioni non sia stata confermata, è stata approfondita la ricerca di specie "anomale", non classicamente riconosciute come appartenenti al microbioma cutaneo nel tentativo di individuare un'associazione tra i campioni. La possibilità di effettuare questo tipo di analisi secondaria sui dati a disposizione è stata di fondamentale importanza dal momento che questo studio è volto a proporre innovazioni in campo forense in cui la mera osservazione di un'evidenza, senza supporto statistico, non è sufficiente. In linea teorica, l'analisi del microbioma per la ricerca di corrispondenza tra traccia e confronto risulterebbe più facile e vantaggiosa rispetto all'analisi del DNA umano in condizioni di low template. Se questo dovesse essere effettivamente confermato da ulteriori studi, l'analisi del microbioma potrebbe sancire un inizio importante per l'ingresso nella nuova era delle scienze forensi.

Tuttavia, vanno considerati alcuni aspetti critici che sarà necessario

approfondire e chiarire in futuro quali la presenza di microbioma di background sulle superfici riceventi, la presenza di microrganismi cutanei con abbondanza relativa insufficiente per essere trasferiti al substrato ricevente e la possibilità di avere contaminazioni da parte di DNA batterico esogeno o proveniente da una fonte diversa dello stesso donatore. Non va tralasciato, inoltre, uno degli aspetti più critici che mettono a dura prova l'applicazione dello studio del microbioma per l'identificazione personale ovvero la variabilità nel tempo. Sebbene numerosi studi presenti in letteratura abbiano dimostrato la stabilità del microbioma cutaneo sulla cute per almeno tre anni (Schmedes, *et al.*, 2017), non è chiaro come questo vari dopo questo arco temporale e come si comporti su superfici riceventi di varia natura dal momento che fluttuazioni della composizione batterica, sebbene leggere, sono state riscontrate anche nel corso di una giornata (Wilkins, *et al.*, 2021). In sintesi, per garantire l'applicabilità dell'analisi del microbioma, il profilo genetico batterico ottenuto dovrebbe essere persistente e coerente indipendentemente dalle modalità di contatto; aspetto indispensabile per ogni tipologia di campione forense, anche per quanto riguarda l'analisi del DNA.

Oltre ad una statistica descrittiva, è stato possibile applicare all'analisi anche un approccio basato sul calcolo di indici di distanza statistica (distanza di Canberra), calcolata per coppie di campioni mediante R-Studio. La scelta di questa metodica è stata influenzata dal fatto che, utilizzando il numero di *reads*, permette di tenere in considerazione anche taxa con un'abbondanza relativa <1%, particolarmente utili per incrementare la sensibilità dell'analisi. I risultati ottenuti applicando questa statistica ai dati disponibili nel presente lavoro sperimentale indicano, verosimilmente, l'esistenza di un trasferimento microbico dal donatore alla superficie ricevente, sia nel caso di quello residente sui palmi delle mani, sia nel caso di quello residente sul collo. Tuttavia, a causa dei bias analitici legati alle basse concentrazioni di DNA di partenza, al sequenziamento soggetto ad eventi stocastici e alla mancanza di repliche, non è possibile stabilire con certezza l'efficacia del sistema statistico prescelto.

Trattandosi di un campo del tutto innovativo e ancora agli albori, il presente studio rappresenta un approccio preliminare effettuato con lo scopo di evidenziare i vantaggi e le criticità dell'utilizzo del microbioma cutaneo in ambito forense applicando una metodica normalmente diffusa in ambito sanitario e utilizzando un protocollo non ancora specifico. I dati ottenuti da questa analisi si limitano ad una iniziale esplorazione del potenziale relativo al microbioma in campo scientifico-forense. Nonostante siano emersi alcuni elementi di criticità che necessitano di essere approfonditi e superati, questo studio dimostra come il microbioma umano cutaneo potrebbe realmente costituire un'importante risorsa per lo sviluppo futuro delle analisi forensi.

## 7. BIBLIOGRAFIA

- BUSHNELL, Brian. BMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2014.
- BUTLER, J. DNA extraction methods. Dalam: Butler J, penyunting. Advanced topics in forensic DNA typing: Methodology. Edisi ke-1. 2012.
- BUTLER, John M. Forensic DNA typing: biology, technology, and genetics of STR markers. Elsevier, 2005.
- BUTLER, John M. Short tandem repeat typing technologies used in human identity testing. *Biotechniques*, 2007, 43.4: Sii-Sv..
- CAPORASO, J. Gregory, et al. QIIME allows analysis of high-throughput community sequencing data. *Nature methods*, 2010, 7.5: 335-336.
- CARAMELLI, David. Antropologia molecolare: manuale di base. *Antropologia molecolare*, 2009, 0-0.
- EDMONDS-WILSON, Sarah L., et al. Review of human hand microbiome research. *Journal of dermatological science*, 2015, 80.1: 3-12.
- EDWARDS, Albert, et al. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *American journal of human genetics*, 1991, 49.4: 746.
- EISENHOFER, Raphael, et al. Contamination in low microbial biomass microbiome studies: issues and recommendations. *Trends in microbiology*, 2019, 27.2: 105-117.
- FIERER, Noah, et al. Forensic identification using skin bacterial communities. *Proceedings of the National Academy of Sciences*, 2010, 107.14: 6477-6481.
- GOUELLO, Audrey, et al. Analysis of microbial communities: An emerging tool in forensic sciences. *Diagnostics*, 2022, 12.1: 1.
- GRICE, E. A., et al. NCS Program, GG Bouffard, RW Blakesley, PR Murray, ED Green, ML Turner, JA Segre. *Science*, 2009, 324: 1190.
- GUPTA, V. K. Paul S. Dutta C: Geography ethnicity or subsistence-specific variations in human microbiome composition and diversity. *Front Microbiol*, 2017, 8.1162: 10.3389.
- LIU, Lin, et al. Comparison of next-generation sequencing systems. *J Biomed Biotechnol*, 2012, 2012.251364: 251364.
- MERRIMAN, Barry; R&D TEAM, Ion Torrent; ROTHBERG, Jonathan M. Progress in ion torrent semiconductor chip based sequencing. *Electrophoresis*, 2012, 33.23: 3397-3417.
- METCALF, Jessica L., et al. Microbiome tools for forensic science. *Trends in biotechnology*, 2017, 35.9: 814-823.
- METZKER, Michael L., et al. Molecular evidence of HIV-1 transmission in a criminal case. *Proceedings of the National Academy of Sciences*, 2002, 99.22: 14292-14297.

- NECKOVIC, Ana, et al. Investigation of direct and indirect transfer of microbiomes between individuals. *Forensic Science International: Genetics*, 2020, 45: 102212.
- OCHMAN, Howard; DAVALOS, Liliana M. The nature and dynamics of bacterial genomes. *Science*, 2006, 311.5768: 1730-1733.
- OH, Julia, et al. Shifts in human skin and nares microbiota of healthy children and adults. *Genome medicine*, 2012, 4: 1-11.
- PHAN, Katherine, et al. Retrieving forensic information about the donor through bacterial profiling. *International journal of legal medicine*, 2020, 134: 21-29.
- POPOVIC, Ana; PARKINSON, John. Characterization of eukaryotic microbiome using 18S amplicon sequencing. *Microbiome analysis: methods and protocols*, 2018, 29-48.
- PROCOPIO, Noemi, et al. "Touch microbiome" as a potential tool for forensic investigation: A pilot study. *Journal of Forensic and Legal Medicine*, 2021, 82: 102223.
- RICCI, Ugo. (2015) *D.N.A. – Oltre ogni ragionevole dubbio*. Edizione Nerbini.
- ROBINSON, Jake M., et al. Forensic applications of microbiomics: a review. *Frontiers in Microbiology*, 2021, 11: 3455.
- ROSS, Ashley A.; DOXEY, Andrew C.; NEUFELD, Josh D. The skin microbiome of cohabiting couples. *MSystems*, 2017, 2.4: e00043-17.
- RUUSKANEN, Suvi, et al. Ecosystem consequences of herbicides: The role of microbiome. *Trends in Ecology & Evolution*, 2022.
- SCHMEDES, Sarah E.; WOERNER, August E.; BUDOWLE, Bruce. Forensic human identification using skin microbiomes. *Applied and environmental microbiology*, 2017, 83.22: e01672-17.
- TAGLIABRACCI, Adriano. *Introduzione alla genetica forense: Indagini di identificazione personale e di paternità*. Springer Science & Business Media, 2009.
- TOYOMANE, Kochi, et al. Evaluation of CRISPR diversity in the human skin microbiome for personal identification. *Msystems*, 2021, 6.1: e01255-20.
- TRINGALI, G., et al. Rapid and efficacious real-time quantitative PCR assay for quantitation of human DNA in forensic samples. *Forensic science international*, 2004, 146: S177-S181.
- WATANABE, Hikaru, et al. Minor taxa in human skin microbiome contribute to the personal identification. *PLoS One*, 2018, 13.7: e0199947.
- WHIPPS, John M.; LUMSDEN, Robert D. Biological control of *Pythium* species. *Biocontrol Science and Technology*, 1991, 1.2: 75-90.
- WILANTHO, Alisa, et al. Diversity of bacterial communities on the facial skin of different age-group Thai males. *PeerJ*, 2017, 5: e4084.
- WILKINS, David, et al. Diurnal variation in the human skin microbiome affects accuracy of forensic microbiome matching. *Microbiome*, 2021, 9.1: 129.

## 8. ALLEGATI

*Allegato 1: Composizione batterica relativa ai campioni cutanei ottenuti al termine dell'analisi, corredata dei valori di abbondanza relativa per ogni famiglia. In arancione sono evidenziate le famiglie batteriche attribuite al "Core Microbiome" cutaneo perché presenti in tutti i campioni. In verde sono evidenziate le famiglie batteriche attribuite al microbioma cutaneo specifico dei palmi delle mani, in quanto riscontrate in tutti i campioni prelevati, in quella zona, a tutti i soggetti.*

<b>Famiglia</b>	<b>G collo (G2) Abb rel Fam</b>	<b>H collo (H2) Abb rel Fam</b>	<b>L collo (L2) Abb rel Fam</b>	<b>F mani (F1) Abb rel Fam</b>	<b>G mani (G1) Abb rel Fam</b>	<b>H mani (H1) Abb rel Fam</b>
Acetobacteraceae	0.79	0	0	0	0	0
Actinomycetaceae	0.5	0	0	8.12	0	0.39
Aerococcaceae	0	0	0	0	0	0.4
Anaeroplasmataceae	0.53	0	0	0	0.44	0
Aurantimonadaceae	0.33	0	0	0	0	0
Bacillaceae	0.23	0	0	0	0.4	0.39
Bradyrhizobiaceae	0	0	0	2.21	0	0
Brevibacteriaceae	0.74	0	0	0	0	0
Carnobacteriaceae	0.83	0	0	0	0.55	0
Caulobacteraceae	1.32	0	0	2.08	0	0
Chitinophagaceae	0.26	0	0	0	0	0
<b>Corynebacteriaceae</b>	<b>9.77</b>	<b>5.88</b>	<b>2.54</b>	<b>4.95</b>	<b>11.16</b>	<b>17.49</b>
Cytophagaceae	0.46	0	0	0	0	0
Dermabacteraceae	0.39	0	0	0	0	0
Dermacoccaceae	0.36	0	0	0	0	0
Dietziaceae	0.55	0	0	0	0	0
Enterobacteriaceae	0.26	0	0	0	1.04	0
Erysipelotrichaceae	0	0	0	0	0	0.28
Flavobacteriaceae	1.78	0	0	0	3.58	0
Fusobacteriaceae	0.52	0	0	0	3.88	0
Gaiellaceae	0.18	0	0	0	0	0
Geodermatophilaceae	0.29	0	0	0	0	0
Gordoniaceae	0.34	0	0	0	0	0
Hyphomicrobiaceae	0.44	0	0	0	0	0
Intrasporangiaceae	1.81	0	0	0	0	0
Lachnospiraceae	0.19	0	0	0	0	0.68
Lactobacillaceae	0	0	0	2.62	0	0
Leptotrichiaceae	0.28	0	0	0	0.4	0
Methylobacteriaceae	0.64	0	0	0	0	0
Microbacteriaceae	0.16	0	0	0	0	0
<b>Micrococcaceae</b>	<b>3.39</b>	<b>4.95</b>	<b>0</b>	<b>6.55</b>	<b>1.27</b>	<b>3.26</b>
Moraxellaceae	2.61	0	0	0	1.81	0.54
Neisseriaceae	0.6	0	0.29	0	3.08	1.94
Nocardiaceae	0.36	0	0	0	0	0
Nocardioideae	0.95	0	0	0	0.26	0
Oxalobacteraceae	0.95	0	0	0	0	0
<b>Pasteurellaceae</b>	<b>1.57</b>	<b>0</b>	<b>0</b>	<b>2.11</b>	<b>9.35</b>	<b>1.82</b>
Peptoniphilaceae	0.86	0	0	0	1.92	2.2
Peptostreptococcaceae	0	0	0	0	0.29	0
Porphyromonadaceae	0.54	0	0	0	2.05	0
Prevotellaceae	0.58	0	0	0	0.48	0
<b>Propionibacteriaceae</b>	<b>34.89</b>	<b>80.09</b>	<b>92.49</b>	<b>50.93</b>	<b>34.44</b>	<b>45.37</b>
Pseudomonadaceae	4.61	0	0	0	0.49	0
Pseudonocardiaceae	0.57	0	0	0	0	0
Rhizobiaceae	0.35	0	0	0	0	0
<b>Rhodobacteraceae</b>	<b>8.45</b>	<b>0</b>	<b>0</b>	<b>3.92</b>	<b>1.34</b>	<b>0.8</b>
Ruminococcaceae	0	0	0	0	0	0.44
Sphingobacteriaceae	0.67	0	0	0	0	0
<b>Sphingomonadaceae</b>	<b>1.72</b>	<b>0</b>	<b>0.31</b>	<b>5.18</b>	<b>0.42</b>	<b>0.58</b>
Staphylococcaceae	6.16	5.16	1.05	5.68	11.8	12.42
<b>Streptococcaceae</b>	<b>5.0</b>	<b>3.92</b>	<b>0.42</b>	<b>2.4</b>	<b>4.16</b>	<b>8.65</b>
Streptomycetaceae	0.14	0	0	2.62	0	0.56
Succinivibrionaceae	0	0	0	0	0	0.31
Thermoanaerobactera	0.22	0	0	0	0	0.42
Veillonellaceae	0.39	0	0	0	0.95	0
Xanthomonadaceae	0.41	0	0	0	0	0

Allegato 2: Composizione batterica relativa ai campioni cutanei ottenuti al termine dell'esperimento, raggruppati in modo differente sulla base delle esigenze analitiche, e corredata dei valori di abbondanza relativa per ogni famiglia. In Allegato 2a (a sinistra) è rappresentato il confronto tra mani e collo del soggetto G; le famiglie batteriche attribuite al "Core Microbiome" cutaneo sono evidenziate in arancione; in verde sono evidenziate le famiglie batteriche attribuite al microbioma del collo, in giallo al microbioma delle mani; in blu sono evidenziate le famiglie le mani e in viola quelle batteriche presenti in entrambi i campioni ma maggiormente rappresentate a livello del collo. In Allegato 2b (a destra) è rappresentato il confronto tra le mani di tutti i soggetti analizzati: le famiglie batteriche attribuite al "Core Microbiome" cutaneo sono evidenziate in arancione; in rosa sono evidenziate le famiglie batteriche attribuite al soggetto H, in indaco quelle attribuite al soggetto G, in viola quelle attribuite al soggetto F; in verde quelle attribuite al microbioma cutaneo specifico dei palmi delle mani, in quanto riscontrate in tutti i campioni prelevati, in quella zona, a tutti i soggetti.

Famiglia	G mani (G1)	G collo (G2)
	Abb rel Fam	Abb rel Fam
Acetobacteraceae	0	0.79
Acetobacteraceae	0	0.79
Actinomycetaceae	0	0.5
Anaeroplasmataceae	0.44	0.53
Aurantimonadaceae	0	0.33
Bacillaceae	0.4	0.23
Brevibacteriaceae	0	0.74
Carnobacteriaceae	0.55	0.83
Caulobacteraceae	0	1.32
Chitinophagaceae	0	0.26
Corynebacteriaceae	11.16	9.77
Cytophagaceae	0	0.46
Dermabacteraceae	0	0.39
Dermacoccaceae	0	0.36
Dermacoccaceae	0	0.36
Dietziaceae	0	0.55
Enterobacteriaceae	1.04	0.26
Flavobacteriaceae	3.58	1.78
Fusobacteriaceae	3.88	0.52
Gaiellaceae	0	0.18
Geodermatophilaceae	0	0.29
Gordoniaceae	0	0.34
Hyphomicrobiaceae	0	0.44
Intrasporangiaceae	0	1.81
Lachnospiraceae	0	0.19
Leptotrichiaceae	0.4	0.28
Methylobacteriaceae	0	0.64
Microbacteriaceae	0	0.16
Micrococcaceae	1.27	3.39
Moraxellaceae	1.81	2.61
Neisseriaceae	3.08	0.6
Nocardiaceae	0	0.36
Nocardiodiaceae	0.26	0.95
Oxalobacteraceae	0	0.95
Pasteurellaceae	9.35	1.57
Peptoniphilaceae	1.92	0.86
Peptostreptococcaceae	0.29	0
Porphyromonadaceae	2.05	0.54
Prevotellaceae	0.48	0.58
Propionibacteriaceae	34.44	34.89
Pseudomonadaceae	0.49	4.61
Pseudonocardiaceae	0	0.57
Rhizobiaceae	0	0.35
Rhodobacteraceae	1.34	8.45
Sphingobacteriaceae	0	0.67
Sphingomonadaceae	0.42	1.72
Staphylococcaceae	11.8	6.16
Streptococcaceae	4.16	5.0
Streptomycetaceae	0	0.14
Thermoanaerobacteraceae	0	0.22
Veillonellaceae	0.95	0.39
Xanthomonadaceae	0	0.41

Famiglia	F mani (F1)	G mani (G1)	H mani (H1)
	Abb rel Fam	Abb rel Fam	Abb rel Fam
Actinomycetaceae	8.12	0	0.39
Aerococcaceae	0	0	0.4
Anaeroplasmataceae	0	0.44	0
Bacillaceae	0	0.4	0.39
Bradyrhizobiaceae	2.21	0	0
Carnobacteriaceae	0	0.55	0
Caulobacteraceae	2.08	0	0
Corynebacteriaceae	4.95	11.16	17.49
Enterobacteriaceae	0	1.04	0
Erysipelotrichaceae	0	0	0.28
Flavobacteriaceae	0	3.58	0
Fusobacteriaceae	0	3.88	0
Lachnospiraceae	0	0	0.68
Lactobacillaceae	2.62	0	0
Leptotrichiaceae	0	0.4	0
Micrococcaceae	6.55	1.27	3.26
Moraxellaceae	0	1.81	0.54
Neisseriaceae	0	3.08	1.94
Pasteurellaceae	2.11	9.35	1.82
Peptostreptococcaceae	0	0.29	0
Porphyromonadaceae	0	2.05	0
Prevotellaceae	0	0.48	0
Propionibacteriaceae	50.93	34.44	45.37
Pseudomonadaceae	0	0.49	0
Rhodobacteriaceae	3.92	1.34	0.8
Ruminococcaceae	0	0	0.44
Sphingomonadaceae	5.18	0.42	0.58
Staphylococcaceae	5.68	11.8	12.42
Streptococcaceae	2.4	4.16	8.65
Streptomycetaceae	2.62	0	0.56
Succinivibrionaceae	0	0	0.31
Thermoanaerobacteraceae	0	0	0.42
Veillonellaceae	0	0.95	0

Allegato 3: Composizione batterica relativa ai campioni di traccia e donatore, utilizzati per effettuare il confronto forense, corredata dei valori di abbondanza relativa per ogni famiglia. In Allegato 3a (a sinistra) è rappresentato il confronto tra famiglie batteriche presenti sul collo e sul colletto del soggetto G. In Allegato 3b (a destra) è rappresentato il confronto tra le famiglie batteriche presenti sulle mani e sul volante del soggetto H. Per entrambe le tabelle: in giallo sono evidenziate le famiglie batteriche sovrapponibili nei due campioni messi a confronto e in viola quelle riscontrate solo sulla superficie ricevente con un'abbondanza relativa <1%. In Allegato 3a, la famiglia batterica identificata mediante riquadro rosso risulta presente in entrambi i campioni, con significative differenze per quanto riguarda l'abbondanza relativa. In Allegato 3b le famiglie batteriche evidenziate in verde rappresentano le famiglie batteriche riscontrate solo sulla superficie ricevente con un'abbondanza relativa >1%.

Famiglia	G collo (G2)		G colletto (G4)	
	Abb rel Fam		Abb rel Fam	
Acetobacteraceae	0.79		0	
Actinomycetaceae	0.5		0	
Anaeroplasmataceae	0.53		0	
Aurantimonadaceae	0.33		0	
Bacillaceae	0.23		0	
Brevibacteriaceae	0.74		0.37	
Carnobacteriaceae	0.83		0	
Caulobacteraceae	1.32		0.61	
Chitinophagaceae	0.26		0	
Corynebacteriaceae	9.77		8.91	
Cytophagaceae	0.46		0	
Dermabacteraceae	0.39		0.12	
Dermacoccaceae	0.36		0	
Dietziaceae	0.55		0.25	
Enterobacteriaceae	0.26		0.35	
Flavobacteriaceae	1.78		0.56	
Fusobacteriaceae	0.52		0	
Gaiellaceae	0.18		0	
Geodermatophilaceae	0.29		0	
Gordoniaceae	0.34		0.25	
Hyphomicrobiaceae	0.44		0	
Intrasporangiaceae	1.81		0.97	
Lachnospiraceae	0.19		0	
Lactobacillaceae	0		0.14	
Leptotrichiaceae	0.28		0	
Methylobacteriaceae	0.64		0	
Microbacteriaceae	0.16		0	
Micrococcaceae	3.39		2.52	
Moraxellaceae	2.61		7.84	
Neisseriaceae	0.6		0	
Nocardiaceae	0.36		0	
Nocardiodiaceae	0.95		0	
Oxalobacteraceae	0.95		0.82	
Pasteurellaceae	1.57		0.28	
Peptoniphilaceae	0.86		0.47	
Porphyromonadaceae	0.54		0	
Prevotellaceae	0.58		0	
Propionibacteriaceae	34.89		33.34	
<b>Pseudomonadaceae</b>	<b>4.61</b>		<b>30.19</b>	
Pseudonocardiaceae	0.57		0	
Rhizobiaceae	0.35		0	
Rhodobacteraceae	8.45		5.64	
Sphingobacteriaceae	0.67		0.41	
Sphingomonadaceae	1.72		0.51	
Staphylococcaceae	6.16		4.16	
Streptococcaceae	5.0		1.01	
Streptomycetaceae	0.14		0	
Thermoanaerobacteraceae	0.22		0	
Veillonellaceae	0.39		0	
Xanthomonadaceae	0.41		0	

Famiglia	H mani (H1)		H volante (H3)	
	Abb rel Fam		Abb rel Fam	
Actinomycetaceae	0.39		1.01	
Aerococcaceae	0.4		1.03	
Anaeroplasmataceae	0		1.43	
Bacillaceae	0.39		0.39	
Bradyrhizobiaceae	0		0.16	
Brevibacteriaceae	0		1.1	
Burkholderiaceae	0		0.3	
Carnobacteriaceae	0		2.54	
Chitinophagaceae	0		0.18	
Clostridiaceae	0		0.21	
Corynebacteriaceae	17.49		4.99	
Dermabacteraceae	0		0.14	
Enterobacteriaceae	0		0.54	
Erysipelotrichaceae	0.28		0	
Flavobacteriaceae	0		0.77	
Fusobacteriaceae	0		0.61	
Geodermatophilaceae	0		0.28	
Halomonadaceae	0		0.73	
Hyphomicrobiaceae	0		0.5	
Lachnospiraceae	0.68		0.58	
Leptotrichiaceae	0		0.23	
Methylobacteriaceae	0		0.16	
Microbacteriaceae	0		0.45	
Micrococcaceae	3.26		1.77	
Moraxellaceae	0.54		0.77	
Neisseriaceae	1.94		3.49	
Nocardiodiaceae	0		0.19	
Oxalobacteraceae	0		0.25	
Pasteurellaceae	1.82		6.04	
Peptoniphilaceae	2.2		1.01	
Porphyromonadaceae	0		0.85	
Prevotellaceae	0		4.47	
Propionibacteriaceae	45.37		31.53	
Pseudomonadaceae	0		1.18	
Rhodobacteraceae	0.8		0.7	
Ruminococcaceae	0.44		0	
Sanguibacteraceae	0		0.14	
Sphingomonadaceae	0.58		0.76	
Staphylococcaceae	12.42		4.84	
Streptococcaceae	8.65		18.54	
Streptomycetaceae	0.56		0	
Succinivibrionaceae	0.31		0	
Thermoanaerobacteraceae	0		0.29	
Veillonellaceae	0.42		1.66	
Xanthomonadaceae	0		0.21	