UNIVERSITÀ DEGLI STUDI
DI MODENA E REGGIO EMILIA

Dottorato di ricerca in Scienze, Tecnologie e Biotecnologie Agro-alimentari

Ciclo XXXIII

Genomic and metagenomic approaches for the characterization of bacteria and microbial communities of ecological niches relevant for the human health

Candidato: Francesco Candeliere

Relatore: Prof. Maddalena Rossi

Correlatore: Dott. Stefano Raimondi

Coordinatore del Corso di Dottorato: Prof. Alessandro Ulrici

Abstract

The modern Next Generation Sequencing technologies represent a crucial component in the study of microorganisms and microbial communities thanks to the huge quantity of data they can provide in a short period of time. These technologies allow the identification and characterization of microorganisms exploiting a vast number of bioinformatic tools that can replace the standard *in vitro* typing techniques, resulting in savings of time and resources. Genomics and metagenomics can be applied in different fields and they can provide information on single microorganism or on entire microbial communities. We focused our studies on two ecological niches: food matrices and human gut microbiome, due to their relevance to human health.

The first work of this thesis is a comparative genomic study of 12 *Leuconostoc carnosum* strains isolated from meat-based products. This bacterium is a known colonizer of meat-based food matrices, it plays a role in spoilage, but preservative effects have also been proposed for some strains. In our study we performed whole genome sequencing for all the strains, and after genome assembly we identified their genomic features, the presence of plasmids, and genes related to antibiotic resistance, bacteriocins production, biogenic amines synthesis. We also reconstructed their metabolic pathways. The comparison revealed that the strains are closely related and share most of the metabolic features, confirming the adaptation to the meat environment due to the presence of 23 peptidase genes in their core genome. With this study we provided a deeper insight into the genomic and metabolic features of this bacterium ubiquitous in meat products.

The second project of the thesis aimed to investigate through an inedited metagenomic strategy the presence of beta-glucuronidases (GUS) in the human gut microbiome. Beta-glucuronidases (GUS) produced by gut microbiome bacteria can remove glucuronic acid moieties from a vast range of compounds and metabolites, like drugs and xenobiotics. These molecules are conjugated with glucuronic acid in the liver to be excreted in the gastrointestinal tract, so the action of GUS may reactivate them allowing the reabsorption, with unpredictable and different efficacy of drugs and negative effect on health. GUS are classified in classes by the differences in the catalytic site. 60 shotgun sequenced metagenomic samples from healthy subjects, ascribed to five geographically distinct cohorts, have been retrieved. From this dataset,

bacterial composition has been defined and a novel pipeline to investigate distribution the different GUS has been developed and utilized. Beta-diversity calculated on Bray-Curtis dissimilarity index has been used to determine the distances among samples and determine the differences among samples in terms of GUS type distribution and abundance of bacteria containing GUS. Since the structural differences in the enzyme involve a different substrate specificity, and taking into account the ratio of bacterial community harbouring GUS genes, we can assess that the microbiota composition can alter the excretion of certain drugs or xenobiotics, and determine a wide interindividual variability in terms of response to drugs.

In the third part of the thesis, I present metagenomic analysis carried out in two other different studies, the former aimed to investigate the microbial composition in enriched human faecal samples to identify gut mucin degraders, and the latter focused on the description of the microbiota of *Hermetia illucens* larvae reared for food or feed consumption.

Keywords: genome sequencing; *Leuconostoc carnosum*; metagenomics; gut microbiome; beta-glucuronidase

Sinossi

Le moderne tecnologie di *Next Generation Sequencing* sono un componente cruciale nello studio di microrganismi e comunità microbiche grazie all'enorme quantità di dati fornita in breve tempo. Con queste tecnologie è possibile identificare e caratterizzare i microrganismi utilizzando un ampio numero di *tool* bioinformatici che possono sostituire le classiche tecniche di tipizzazione *in vitro*, portando un risparmio di tempo e risorse. La genomica e la metagenomica possono essere applicate a vari campi e forniscono informazioni sia riguardo un singolo organismo, sia su intere comunità microbiche. I nostri studi sono stati concentrati su due nicchie ecologiche: matrici alimentari e microbiota intestinale umano, entrambi importanti per la salute umana.

Un prima parte della tesi descrive la comparazione genomica di 12 ceppi di *Leuconostoc carnosum* isolati da prodotti a base di carne. Questo batterio è un noto colonizzatore di queste matrici, ricopre un ruolo nel loro deterioramento, ma alcuni ceppi presentano effetti utili alla preservazione. Abbiamo eseguito un sequenziamento *whole genome* per tutti i ceppi, e dopo l'assembly sono state identificate le caratteristiche genomiche, la presenza di plasmidi, di geni responsabili di antibiotico-resistenza, produzione di batteriocine, sintesi di ammine biogene e abbiamo ricostruito i loro *pathway* metabolici. La comparazione ha rivelato che i ceppi sono strettamente correlati e condividono la maggior parte delle caratteristiche metaboliche, evidenziando l'adattamento all'ambiente di isolamento grazie alle 23 peptidasi presenti nel genoma *core*. Questo studio fornisce un approfondimento genomico e metabolico su questo batterio ubiquitario nei prodotti a base di carne.

Un secondo progetto della tesi è stato indirizzato a indagare la presenza delle beta-glucuronidasi (GUS) nel microbiota intestinale umano attraverso una strategia metagenomica inedita. Le GUS prodotte dai batteri del microbiota sono capaci di rimuovere le porzioni di acido glucuronico da molti composti e metaboliti, come farmaci e xenobiotici. Queste molecole vengono coniugate con l'acido glucuronico nel fegato per l'escrezione attraverso il tratto gastrointestinale, quindi l'azione enzimatica potrebbe riattivarle permettendone il riassorbimento comportando un'alterazione dell'efficacia del farmaco ed effetti negativi sulla salute. Le GUS sono organizzate in classi in base alle differenze nel loro sito catalitico. Sono

stati utilizzati 60 metagenomi ottenuti con *shotgun sequencing* provenienti da soggetti sani divisi in coorti in base alla provenienza geografica. Da questo set di dati, è stata definita la composizione batterica ed è stata sviluppata ed impiegata una nuova strategia per indagare la distribuzione delle diverse GUS. La beta diversità calcolata con gli indici di Jaccard e Bray-Curtis è stata impiegata per determinare le distanze tra i campioni e determinare le differenze tra essi in termini di distribuzione del tipo di GUS e abbondanza dei batteri che le contengono. Poiché alle differenze strutturali dell'enzima corrisponde una diversa specificità di substrato, e considerando la proporzione delle comunità batteriche contenenti GUS, possiamo valutare che la composizione del microbiota può alterare l'escrezione di alcuni farmaci o xenobioti e determinare un'ampia variabilità interindividuale in termini di risposta al farmaco.

Un'ultima parte del progetto di tesi descrive l'applicazione dell'approccio metagenomico in due diversi studi: il primo incentrato a indagare la composizione microbica di campioni fecali arricchiti per identificare specie intestinali capaci di degradare la mucina, e il secondo focalizzato sulla descrizione del microbiota di larve di *Hermetia illucens* allevate per consumo umano o animale.

Parole chiave: sequenziamento genomico; *Leuconostoc carnosum*; metagenomica; microbioma intestinale; beta-glucuronidasi

Table of contents

1. OVERALL INTRODUCTION

Next Generation Sequencing (NGS) technologies represented the beginning of a new era in the field of genome deciphering, when the first platform launched in 2005. Previously, the Sanger method, based on the electrophoretic separation of chain-termination products produced in individual sequencing reactions, was the dominant approach for DNA sequencing. With NGS platforms, the approach was different: in a single flow cell, clonally amplicons of single DNA molecules were sequenced massively in parallel. This design allowed to output a huge quantity of data with just a single instrument run and to reduce the costs for genome sequencing. This innovation determined the implementation of NGS-based analysis in many fields like diagnostics, outbreak investigations, antimicrobial resistance, forensics, and food authenticity (Allard et al., 2017; Goodwin et al., 2016; Quainoo et al., 2017).

1.1 THE BASIS OF NGS TECHNOLOGIES

Since the launch of the first NGS sequencer in 2005, various sequencing platforms have been developed and are continuously improved. The common base of these technologies is that, unlike the Sanger methods, the DNA to be sequenced is directly used to construct a library. The library is composed by DNA fragments ligated to synthetic adapters, specific to each platform. The ligated fragments can thus be attached to a solid support, either a bead or a flat microfluidic channel, that is derivatized with adapter sequences complementary to those on the DNA fragments. Each fragment is now amplificated to generate a cluster, to provide sufficient signals to determine the sequencing data.

Contrary to Sanger methods, sequencing and detection occur simultaneously in NGS methods, leading to output a huge quantity of data with just a single instrument run. Read length represents another difference between Sanger and NGS. In the Sanger method, the length of each read depends upon different factors such as the electrophoresis conditions, the

concentration of the polyacrylamide, the time of separation, and the geometry of the gel. In NGS, the read length is limited by the signal-noise ratio and in general, produce shorter reads than Sanger sequencing. Shorter reads limit the length of continuous sequences that can be assembled, but this issue has been addressed by the latest technologies (Goodwin et al., 2016).

Another improvement in NGS is represented by paired-end sequencing. With this technique, the DNA fragments are sequenced from both ends, proving a higher certainty on the result during the alignment steps in data analysis.

Apart from shared features, each platform differs significantly. For example, in the 454 pyrosequencing platform the (Margulies, 2006)) clusters of amplified DNA fragments are pyro sequenced (Ronaghi et al., 1996) and the detected signal is due to the inorganic pyrophosphate released after the addition of an unlabeled nucleotide. Illumina sequencing instead uses reversible dye terminators, the latest technologies, the so-called Third Generation Sequencing like PacBio, even discarded the PCR amplification step, using a single-molecule real-time sequencing approach (Eid et al., 2009). Illumina technology is the sequencing platform used in the project of this thesis, and it is described below.

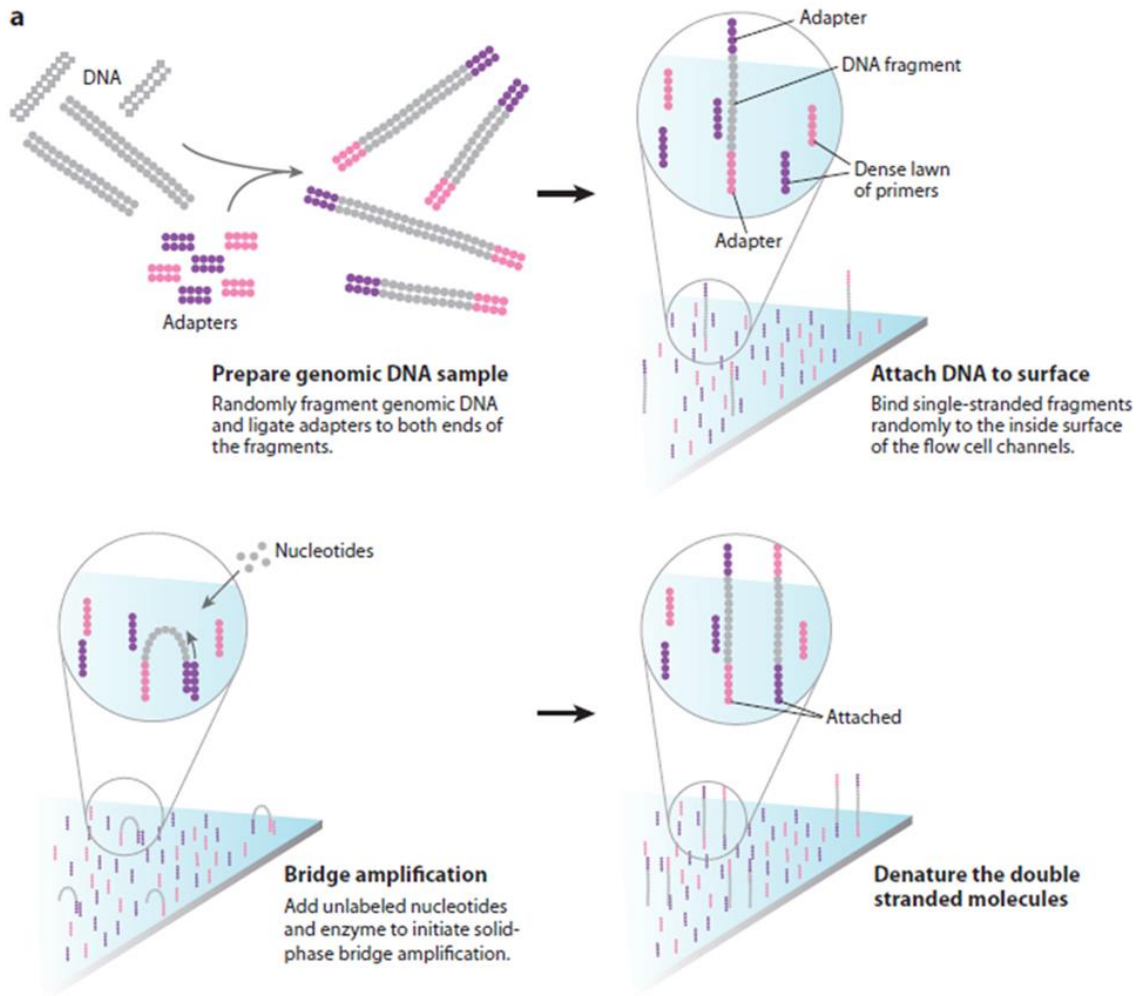
1.2 ILLUMINA TECHNOLOGY

Illumina represents one of the most widely used NGS platform (www.illumina.com). This technology is based on the use of reversible dye terminators that are added to the sequence by repeated cycles of polymerase-directed single base extension (Bentley et al., 2008). Contrary to the classic Sanger method, where the chain-terminator stops the sequence extension, the dye terminators used with Illumina are reversible, so they are removed after each imaging cycle and a new nucleotide is added.

For Illumina sequencing, the DNA template needs to be fragmented, then specific adapters are ligated to each fragment. The adapters allow to link the fragments to the oligonucleotide complementary anchor on the flow cell surface. (Fig. 1.1a) The templates are amplified by “bridge” amplification. Multiple amplification cycles produce a clonally amplified cluster from a single-molecule template. At this point, the sequencing procedure can begin: the clusters are denaturated, and a primer complementary to the adapters and the

reagents (polymerase and mixture of the 4 fluorescent-labeled nucleotides) are added. The nucleotide terminators are included according to the complementary sequence in each clonal cluster, the excess is washed away, and the flow cell is imaged. In the next step, the reversible terminators are unblocked, and a new sequencing cycle is performed (Fig. 1.1b).

At the end of the run, millions of reads are provided with a length depending on which Illumina sequencer has been used, up to 300 bp in the modern machines. The workflow of this technique is reported in the figure adapter from "Next-Generation DNA Sequencing Methods" Elaine R. Mardis (Annu. Rev. Genom. Hum. Genet. 2008.9:387-402).



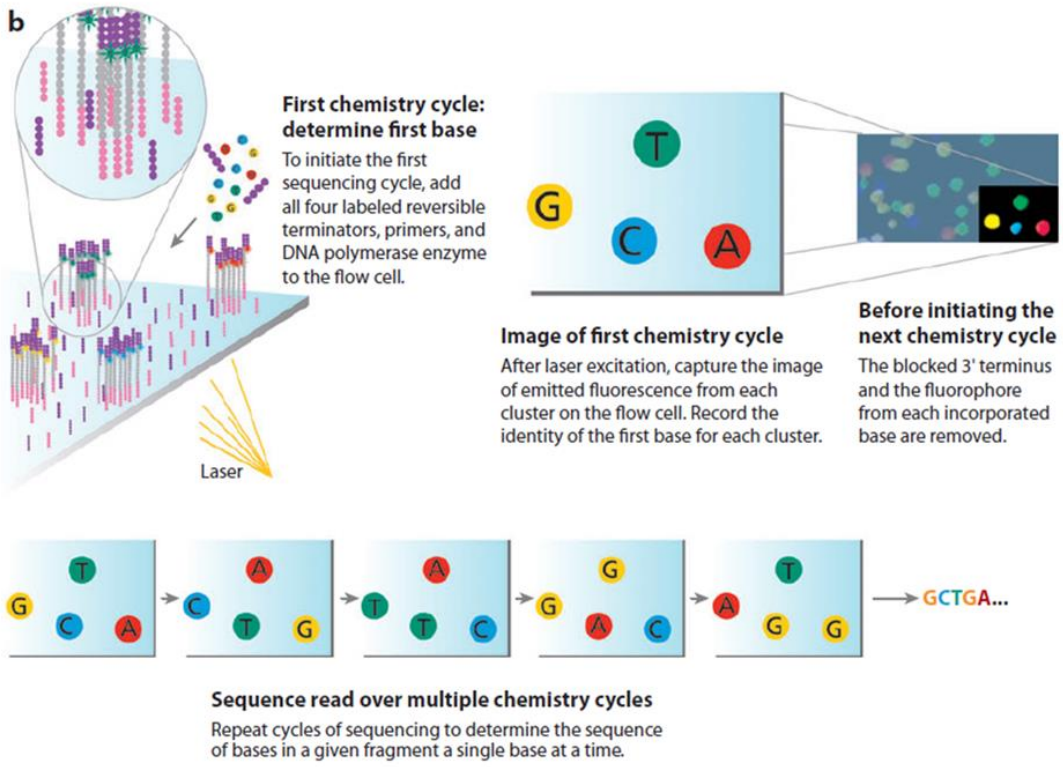


Figure 1.1 The Illumina sequencing approach

(Mardis et al., 2008, Annu. Rev. Genom. Hum. Genet).

1.3 GENOME ASSEMBLY AND ANNOTATION IN WGS

The sequencing procedure outputs raw reads, usually in FASTQ format, based on DNA fragments originated from the genome. The short reads need to be assembled into longer sequences, called contigs, that eventually represent a genome. Many assembly algorithms have been developed, some of which optimized for certain technologies or reads length, or even capable of using reads from multiple sequencing platforms in a hybrid approach. For example, Velvet (Zerbino and Birney, 2008) is an assembler frequently used for Illumina sequencing data and has been applied to a variety of species, e.g. *Enterococcus faecium*, *Staphylococcus aureus*, *Clostridium difficile*, *Escherichia coli* (Feil et al., 2017; Eyre et al., 2012) whereas the algorithm Canu is designed to assemble long reads from a single-molecule

platform such as Pacbio (Koren et al., 2017). The assemblers SPAdes (Bankevich et al, 2012) and RAY (Boisvert et al., 2010) support the use of both short and long reads to assembly the genomes.

In this thesis, we used SPAdes, a de Brujin graph-based assembler. This algorithm breaks down original short reads into smaller sequences called k-mers, which are further reduced into k-1-mers. Then it identifies a Eulerian walk, which is the shortest possible path through the k-1-mers to reconstruct the original sequences. SPAdes algorithm follows four steps: 1) a multisized de Brujin graph is assembled and error correction algorithms are implemented; 2) the distances between the k-mers in the graph are estimated; 3) a paired assembly graph is constructed; 4) contigs are built and initial reads are mapped against them to determine the final sequence. One of the most important features of this tool is that it can use information obtained from k-mers of different lengths. It is in continuous development, so each version offers improvements.

Once the reads are assembled into contigs, it is possible to determine the classification of the bacterial isolate and to identify which genes are present in the genome. Tools like KmerFinder (Larsen et al., 2014; Hasman et al., 2014) or NCBI BLAST (Altschul et al., 1990) offer solution to identify the species of the sequenced genome, and they are both available via command line and as Web-based tools (https://cge.cbs.dtu.dk/services/KmerFinder/; https://blast.ncbi.nlm.nih.gov/Blast.cgi). It is also possible to order those contigs against a reference genome with tools like MUMmer (Kurtz et al., 2004) or Mauve (Darling et al., 2004) to aid possible comparison analysis.

The genome annotation can be performed with different tools, both web-based and command line. Rapid Annotation Using Subsystem Technology (RAST) (https://rast.nmpdr.org/) is a fully automated Web-based tool that can be used to annotate contigs (Aziz et al., 2008). It uses several algorithms, GLIMMER3 (Delcher et al., 2007), BLASTP, BLASTX, BLAST (Altschul et al., 1990), and the SEED subsystem to obtain the best annotation possible. The Web-based interface allows also to view the identified genes and compare them to other genomes. PROKKA (Seemann, 2014) is a rapid command line annotation tool to annotate bacterial genomes. The software combines different tools, like RNAmmer (Lagesen et al., 2007), Aragorn (Laslett et al., 2004), SignalP (Petersen et al., 2011), Infernal (Nawrocki and Eddy, 2013), and Prodigal (Hyatt et al., 2010) to predict features and identify gene location and function. Another option for genomes annotation is represented by curated databases like the Kyoto Encyclopedia of Genes and Genomes

(KEGG) (Kanehisa et al., 2000), or Cluster of Orthologous Groups (COG) (Tatusov et al., 1997), that permit to acquire functional information and to reconstruct metabolic pathways. There are also tools to investigate specific targets, like genes responsible for bacteriocins production (van Heel et al., 2018), antimicrobial resistance genes (Alcock et al, 2020), phages (Arndt et al., 2016), CRISPR determinants (Couvin et al., 2018).

1.4 AMPLICON SEQUENCING

The investigation of microbial communities without bacterial isolation and cultivation can be approached by NGS technologies in two ways: in amplicon sequencing only a specific marker gene is amplified and sequenced, whereas metagenomic is based on the random shotgun sequencing of all the genomes present in a sample.

Both strategies require the extraction of total DNA directly from samples. In the amplicon-based approach the DNA undergoes targeted PCR amplification of phylogenetic marker genes. 16S rRNA gene is generally used for Archaea and Bacteria and the internal transcribed spacer (ITS) of the ribosomal gene cluster sequences for fungi. The sequences of the amplicons are analyzed with dedicated bioinformatics tools, such as Mothur (Schloss et al., 2009) or QIIME2 (Bolyen et al., 2019) to obtain information about the composition of the samples and to study microbial diversity.

The bioinformatic analysis starts with the cleaning and quality filtering of the raw sequences and proceeds with their clustering in Operational Taxonomic Units (OTUs), usually at 97% similarity (Konstantinidis and Tiedje, 2005). OTU clustering can be obtained with three different strategies, closed, *de novo,* or open (Rideout et al.2014), based on the utilization of a reference database to perform the clustering. Common database used for this purpose are Greengenes (greengenes.secondgenome.com) (McDonald et al., 2012), Ribosomal Database Project (RDP) (rdp.cme.msu.edu)(Cole et al., 2009), or SILVA (www.arb-silva.de) (Quast et al., 2013).

This step is useful to detect distinct lineages, to estimate diversity and assess microbial community structure. But it has some downsides, like the fact that a single sequence identity cut-off is inappropriate to delineate true taxonomic lineages such as the species or genus levels, underestimates the number of substitutions compared to a multiple alignment, and

does not consider the variability of the 16S rRNA gene or other conserved targets across the tree or network of life (Nguyen et al., 2016). In the last years, thanks also to the increasing quality of reads, it was possible to develop alternatives that do not rely on clustering or identity thresholds to identify OTUs and enable analysis of the diversity of closely related but distinct bacterial organisms usually grouped into OTUs (Eren et al., 2013). One of these algorithms is DADA2, that allows the identification of Amplicon Sequence Variant (ASVs) and uses them to generate an abundance table (Rosen et al., 2012).

One of the benefits of the amplicon sequencing is the ability to follow the evolution of microbial population over time at various taxonomic level and, compared to metagenomics, this approach provides a cost-efficient analysis of microbial composition.

1.5 METAGENOMIC MICROBIOME PROFILING

The proper metagenomic analysis involves the sequencing of all the genomic content of a sample. This approach has higher costs, but it can provide functional information in addition to the microbial composition. The sequencing output can be aligned to known reference sequences databases to obtain the community profile or can be assembled to reconstruct genomes. *De novo* assembly of the raw sequences allows to identify known species and to detect and describe potentially new ones (Pasolli et al., 2019). Furthermore, it can be exploited to investigate targeted genes.

There are many tools to analyze metagenomics sequences, even without the necessity of genome assembly. MetaPhLan2 (Truong et al., 2015; Segata et al., 2012) profiles the microbial composition in terms of taxonomic attribution and relative abundance mapping the raw reads against its database of marker sequences of specific bacterial clades. Kraken (Wood et al., 2014) and Bracken (Lu et al., 2017) are also a powerful tool that performs fast metagenomic sequence classification and estimate the species abundances. MEGAN (Huson et al., 2007) is another frequently used software to compute and explore the taxonomical content of metagenomic data sets. Web-based pipelines like MG-RAST (Wilke et al., 2016) and Anv'io (Eren et al., 2015) can provide both microbial composition and function annotation.

The *de novo* assembly is the approach that allows to reconstruct of single genomes from metagenomic data sets. The raw reads are assembled in contigs with specific assemblers for metagenomic samples, like metaSPAdes (Nurk et al., 2017) or MEGAHIT (Li et al., 2015). The contigs are binned to obtain metagenome-assembled genomes (MAGs), using tools such as MaxBin2 (Wu et al., 2016) or MetaBAT2 (Kang et al., 2015). The taxonomy attribution of each MAG can be retrieved mapping its contig to a reference database using tools like CAT/BAT (von Meijenfeldt et al., 2019), MEGAN-LR (Huson et al., 2018), Kaiju (Menzel et al., 2016) and DIAMOND (Buchfink et al., 2015), while their abundance can be determined by the proportion of uniquely mapped and correctly paired reads.

1.6 DESCRIBING THE MICROBIAL DIVERSITY

The study of biodiversity is necessary to describe and compare highly diverse communities. Biodiversity is described by richness and evenness. Species richness is defined as the number of the different species found and does not consider the abundances of the single species. Species evenness instead measures the relative abundances of the different species and quantifies how equally distributed the community is.

In the ‘60s, three levels of diversity were introduced by Whittaker: alpha, beta and gamma diversity. The alpha diversity is the within-habitat or intracommunity diversity, the beta-diversity is the between-habitat diversity and the gamma diversity is the diversity of an entire landscape and can be considered a composite of alpha and beta.

1.6.1 Alpha diversity

Alpha diversity is represented by the mean species diversity in sites or habitats at a local scale. The study of the alpha diversity is based on the study of two properties of samples: species numbers (richness of the community) and the relative importance of each species in the community. The analysis of the alpha diversity can be done through the use of different indices, divided into three main groups: species richness, heterogeneity, and evenness.

In metagenomics, rarefaction curves are often generated to analyze alpha diversity. This step determines whether the sampling depth is sufficient to accurately characterize the

bacterial community. The production of more sequences from a single sample gets to the expectation of an increase in the number of observed species until the production of further sequences does not result in new species. With rarefaction curves it is possible to evaluate whether the sequencing was done at a sufficient depth.

Rarefaction analysis involves *in silico* random and repeated subsampling of each community and the average number of OTUs or the observed species can be plotted against the size of the subsamples (Gotelli and Colwell, 2011). In details, rarefaction curves express the expected number of species in n samples randomly chosen from the total number of samples. When the curves flatten, this is the point at which the number of OTUs or observed species does not increase with further sampling, and this means that enough samples have been sequenced to accurately characterize the community (Fig. 1.2).

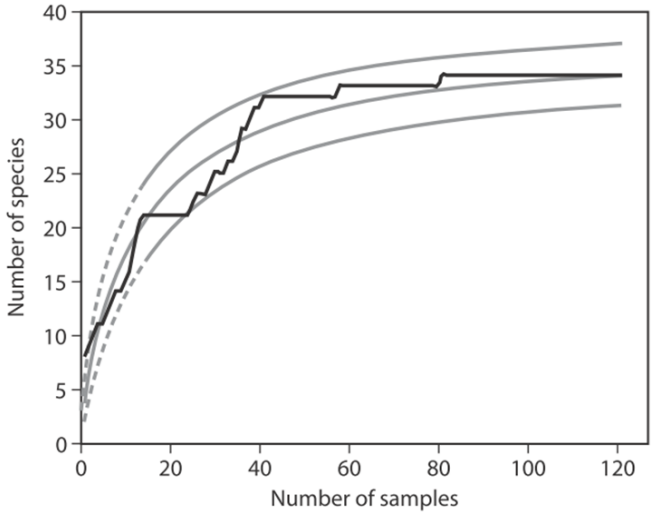


Figure 1.2 Sample- and individual-based rarefaction and accumulation curves (Gotelli and Colwell, 2001)

Species richness: Observed Species Index and Chao1 Index

Species number is the simplest and most fundamental concept of diversity. Richness measure depends on sample size, but it would be better to have a richness index independent of it. To describe species richness within each sample the Species Counts index is applied,

representing the number of distinct OTUs or observed species that were found in a given sample.

Chao1 index is based upon the number of rare classes (OTUs) found in a sample and it uses this number to estimate if there are more undiscovered species (Chao, 1984).

$$S_{es} = S_{obs} + \frac{a^2}{2b}$$

In the equation S_{obs} is the number of species observed in the sample, *a* is the number of species with only a single occurrence in the sample (singletons) and *b* is the number of species with two occurrences in the sample (doubletons). The meaning of this index is that, meanwhile subsampling a community, if more rare species are discovered (singletons), it is likely that more rare species can still be found; Conversely, when all the species have been recovered at least twice (doubletons), there is likely no more species to be discovered.

Heterogeneity index: Shannon's Index

This analysis implies the measurement of two contributing components: the number of species (the richness) and the distribution of individuals among those species (the evenness).

Shannon's Index assumes individuals are randomly sampled from an infinitely large community, and that all species are represented in the sample (Peet, 1974). The Shannon index is calculated from the equation reported below where *S* is the number of species and p_i is the proportion of the species *i* within the sample:

$$H = -\sum_{i=1}^{S} p_i \, log_2 \, (p_i)$$

The index assumes higher values when both the richness and the evenness increase, representing more diverse communities. A community with only one species would have a Shannon index value of 0, conversely, if species are evenly distributed, the index value would be high.

Evenness index: Pielou's index

Pielou's index (Pielou, 1966) measures diversity along with species richness. It is calculated with the following formula.

$$J' = \frac{H'}{H'_{\max}}$$

H' is the number derived from the Shannon diversity index and H'_{max} is the maximum possible value of H'. The values of J' range between 0 and 1, with 0 corresponding to no evenness and 1 to an equal distribution of the species in the sample.

1.6.2 Beta-diversity

The beta-diversity, introduced by R.H. Whittaker in 1960, is defined as "the extent of change in community composition, or degree of community differentiation, in relation to a complex-gradient of environment, or a pattern of environments". More simply, beta-diversity represents and quantifies the effective number of different communities in a certain region, expressed as the ratio between local (alpha) and regional (gamma) diversity and informing about the level of differentiation among biological communities.

The beta-diversity measures can be qualitative (Unweighted) or quantitative (Weighted). The qualitative measures use the presence or absence of data to compare community composition, whereas quantitative measures also consider the relative abundance of each observed organisms.

Beta-diversity metrics can be divided into two groups:

1. Non-phylogenetic beta-diversity metrics, such as Jaccard distance and Bray-Curtis dissimilarity.
2. Phylogenetic beta-diversity metric, such as the unweighted and weighted UniFrac distance.

Phylogenetic metrics use the structure of the phylogenetic tree to include information about the evolutionary distance between taxa. On the contrary, non-phylogenetic metrics only depend on taxa abundance and evenness. Some metrics, such as Jaccard and Unweighted UniFrac, account only for the absence/presence of the taxa, while others, such as Bray-Curtis and Weighted UniFrac, also consider the relative abundance.

Jaccard distance and Bray Curtis dissimilarity

Jaccard and Bray-Curtis measures are widely used in ecology to describe beta-diversity. The Jaccard distance is defined as the measure of the dissimilarity between samples sets. It is complementary to the Jaccard coefficient and is obtained by subtracting the Jaccard coefficient from 1, or, equivalently, by dividing the difference of the sizes of the union and the intersection of two sets by the size of the union, as shown in the formula below.

$$d_J(A,B) = 1 - J(A,B) = \frac{|A \cup B| - |A \cap B|}{|A \cup B|}.$$

The Bray-Curtis dissimilarity (Bray and Curtis, 1957) is used to quantify the differences in terms of species population between two different sites. It is calculated with the following formula, where *i* and *j* are the two sites, S_i and S_j the total number of specimens counted on the respective site, and C_{ij} is the sum of only the lesser count for each species found in both sites.

$$BC_{ij} = 1 - \frac{2C_{ij}}{S_i + S_j}$$

Bray-Curtis dissimilarity values range between 0 and 1, with 0 corresponding to the same composition between the samples and 1 to the case where the sites do not share any species.

UniFrac Measures

UniFrac indices calculate the distances between two communities, incorporating the phylogenetic distances information among observed organisms. All taxa observed in the compared samples are placed on a phylogenetic tree in which branches generated from taxa observed in both samples are called “shared”, and the others “unshared”. The distance can be calculated as:

$$distance = \frac{\sum unshared\ branch\ lengths}{\sum shared + unshared\ branch\ length}$$

In this case, the unweighted distance is considered, where duplicate sequences do not contribute to branch length to the tree, and thus any information on the abundance is not considered.

The Weighted UniFrac is a variant of the unweighted, and it allows to consider the distances based on the abundance information. It is thus a quantitative measure of the beta-diversity between communities. It is represented by the UniFrac value (u):

$$u = \sum_{i}^{n} b_i \times \left| \frac{A_i}{A_T} - \frac{B_i}{B_T} \right|$$

In the equation, the parameters are reported: n as the total number of branches in the tree, b_i as the length of branch i, A_i and B_i the number of sequences that descend from branch i in communities A and B, and A_t and B_t as the total numbers of sequences in both communities. The distance u can be normalized to have a value of 0 for identical communities and 1 for completely different ones. (Lozupone and Knight, 2007).

1.7 MULTIVARIATE DATA ANALYSIS: THE PRINCIPAL COORDINATE ANALYSIS (PCoA)

The distance matrices computed in QIIME2 (Bolyen et al., 2019) from the different beta-diversity metrics can be used to generate Principal Coordinates Analysis (PCoA) plots (Legerde, 1998). PCoA is a popular ordination unconstrained method that makes it possible to explain most of the data variance using fewer variables than the original ones. It is very similar to PCA (Joliffe, 1986), with the main difference being that PCoA allows to visualize

data contained in the distance matrix because it uses distances between samples to calculate positions in the multidimensional framework.

First, the D_{NxN} matrix of the distances between the N subjects is calculated and then it is transformed into the Δ matrix:

$$\Delta = -\frac{1}{2} D$$

The Δ matrix is centered in such a way that the new origin of the axis system (defined with the new variables) is in the centroid of objects; thus, a P matrix is built:

$$c_{ki} = \partial_{ki} - \frac{1}{N}\sum_{p=1}^{N} \partial_{kp} - \frac{1}{N}\sum_{q=1}^{N} \partial_{qi} - \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N} \partial_{pq}$$

The second and third terms of such matrix are the average row and column of the Δ matrix respectively, while the last term represents the average of this matrix. Furthermore, the eigenvalues λ_i (i=1,2,…,m; m≤ N-1) and the eigenvectors u_{ki} (k=1,2,…,N; i=1,2,…m) of P matrix are calculated. The f_{ki} principal coordinates are obtained by multiplying the eigenvectors by the square root of the corresponding eigenvalues:

$$f_{ki} = \sqrt{\lambda_i}\, u_{ki}$$

Even in this case, it is possible to build an ordination diagram in which samples are represented as points and axes are associated with the new PCoA variables. A distance matrix in which each distance is calculated between every pair of samples expresses the dissimilarity between samples and represents the relationships among diverse microbial communities.

1.8 NGS APPLICATIONS

NGS in microbiology is mainly applied for sequencing the whole genome of a single strain, referred as "whole genome sequencing" (WGS), and for the "metagenomics" approach, where the targets are all the genomes present in a sample. NGS current applications are numerous and spread in various fields. In public health surveillance, WGS of microbial pathogens has been introduced at least in four countries, United Kingdom, Denmark, France, and USA, in order to trace and control foodborne pathogens (Allard et al., 2016; Ashton et al., 2016; Jackson et al., 2016; Kvistholm Jensen et al., 2016; Moura et al., 2016). It was introduced as a replacement technology for previous identification and characterization methods like serotyping, virulence profiling, antimicrobial resistance determination, and other molecular typing methods. One single WGS run with the appropriate tools can replace all those methods in a rapid and cost-efficient way (Allard et al., 2016; Carleton and Gerner-Smidt, 2016).

In food industry, the interest for WGS is increasing in order to improve both food quality and safety, thanks to the benefits associated with the identification of pathogens and species related to spoilage. Some of the applications in this field make it possible to distinguish between new and recurrent introduction of an organism into the production environment, to predict virulence or antimicrobial resistance of a pathogen or the metabolic features that could hinder the preservation of the product, to track the source of microbial contamination (Rantsiou et al., 2017; Van Hoorde and Butler, 2018).

NGS applications had a pivotal role in expanding the knowledge of the gut microbiota ecosystem. The first metagenomic studies described the diversity of gut microbiota from fecal samples and colonic biopsies (Eckburg et al., 2005; Hold et al., 2002; Suau et al., 1999), and the further ones correlated the bacterial diversity on intestinal content to pathologies like obesity, diabetes mellitus, inflammatory bowel disease, and colorectal cancer (Lagier et al., 2012). The metagenomic analysis disclosed the wide diversity of gut bacteria that remained uncultured. To solve this issue, culturomics was developed (Lagier et al., 2018), combining multiple culture conditions with rapid identification of bacteria, and allowed isolation, growth and complete genome sequencing of hundreds of new organisms (Lagier et al., 2016).

The human gut microbiome is one of the major targets of metagenomic studies. However, metagenomics was applied to describe microbial populations in seawater (Venter

et al., 2004), soil (Fierer, 2017), and other human body districts like skin and vaginal mucosa (Gao et al., 2007; Hyman et al., 2005; Turnbaugh et al., 2007).

The metagenomic approach is also widely applied in the food industry in different steps of production (Cao et al., 2017). NGS-based strategies can help to maintain the sanitary standards of the production/processing lines and the quality of the final products with control of pathogenic bacteria that colonize these environments and of cross-contamination of the products (Rodrigues et al., 2017, Kable et al., 2016). Microbial profiling can be used to monitor the surface of raw and ready-to-eat foods (Pothakos et al., 2015; Lo et al., 2016) and in agriculture, NGS-based studies made it possible to investigate how the plant microbiome can be affected by season, irrigation, and soil type (Williams et al., 2013). Through NGS it is possible to monitor the microbiomes associated with food storage condition to evaluate the evolution of the microbial communities during the product's shelf-life (Raimondi et al., 2018, 2019; Ferrocino et al., 2016; De Filippis et al. 2013). Metagenomics can also be used to monitor microorganism during food fermentations to understand the microbial dynamics during this process and identify the presence of possible spoilers (De Filippis, 2016).

2. OBJECTIVES

This thesis is focused on different application of genomics and metagenomics strategies based on NGS technologies. These technologies, that can be applied on different fields and matrices, in this project were focused on the investigation of two different ecological niches, both important for human health: food matrices and human gut microbiome. The main aim of the work proposed in this thesis was to exploit WGS and metagenome technologies to investigate the functional and metabolic potential of food colonizers, to determine microbiota composition of innovative food matrices and of intestinal bacterial population, to develop and apply a new pipeline to search targeted genes in metagenomes. The specific objectives of the work are:

- to assess the potential of the meat colonizer *Leuconostoc carnosum* by comparative genomics;
- to study the interindividual variability of gut microbiota of healthy subjects in terms of β-glucuronidases genes;
- to apply16S rRNA gene profiling for identification of mucin degraders in gut microbiota;
- to exploit 16S rRNA profiling to study the microbiota of the novel food *Hermetia illucens*.

L. carnosum is a lactic acid bacterium often isolated in meat-based products. It is dominant in meat products packed in modified atmosphere, such as cooked ham (Raimondi et al., 2018, 2019). Albeit its main role in meat microbial populations and the high charges detected in these products, a detailed functional analysis of its genome was still lacking. To fill this gap, 12 strains isolated in the UNIMORE MIBI lab have been sequenced with a WGS approach and functional analysis has been performed including 5 publicly available *L. carnosum* genomes. The features of plasmids and the metabolic capabilities of this species have been explored. The sequencing was conducted during the abroad period at Nanyang Technological University of Singapore, under the supervision of Prof. Joergen Schlundt.

A metagenomic investigation of the presence of beta-glucuronidases (GUS) in the human gut microbiome has been performed. These enzymes are produced by some intestinal

bacterial species and may have a role in reabsorption of drugs and xenobiotics. This effect in some cases can be harmful to human health, affecting the biological activity of drugs. The aim of this project has been the assessment of the interindividual variability of the potential of GUS encoded by metagenomes. 60 metagenomes of fecal samples from healthy subjects were retrieved and analyzed. A novel pipeline was projected and applied to extract data from metagenomes and determine the GUS distribution and the main taxa contributing to GUS activity in the gut.

Furthermore, in this PhD project amplicon-based metagenomics was applied to determine microbiota composition in to two different matrices. Mucin degraders of the human gut were identified analyzing fecal populations before and after *in vitro* enrichment steps on a mucin-based medium. Identification of the taxa that get enriched made possible to pick genera and species that take advantage using mucins as carbon and nitrogen sources. Furthermore, 16S rRNA gene profiling was utilized to study microbiota composition of prepupae of *Hermetia illucens,* an insect that recently was accepted as potential food. The metagenomes of larvae reared at three different temperatures was analyzed, since this parameter may be a main driver of microbial population composition and can pose a risk of microbiological hazard.

3. Project 1

Comparative genomics of *Leuconostoc carnosum*

3.1 INTRODUCTION

3.1.1 Lactic acid bacteria

Lactic Acid Bacteria (LAB) are Gram-positive bacteria of the order of Lactobacillales, phylum Firmicutes. LAB encompass 14 genera: *Aerococcus, Alloiococcus, Carnobacterium, Enterococcus, Lactobacillus, Lactococcus, Leuconostoc, Oenococcus, Pediococcus, Streptococcus, Symbiobacterium, Tetragenococcus, Vagococcus,* and *Weissella* (Schleifer and Ludwig, 2006). They share a saccharolytic metabolism based on fermentation of carbohydrates to lactic acid. On the basis of the final products of the fermentation processes, LAB can be distinguished into obligate homofermentative (fermentation of carbohydrates mostly produces lactic acid); obligate heterofermentative (fermentation of carbohydrates results into lactic acid, acetic acid or ethanol, and CO_2); facultative heterofermentative (on the basis of growth conditions they can use different catabolic pathways, homofermentative or heterofermentative).

Despite their wide phylogenetic and functional diversity, all the LAB are anaerobic or microaerophilic, aciduric or acidophilic, non-sporulating bacteria. They inhabit a variety of niches where carbohydrate-based substrates are available, such as plants, plant-derived matrices, silage, fermented foods (e.g. dairy products, fermented dough, milk, vegetables, and meats), spoiled foods, organic matrices, and sewage (Hammes and Hertel, 2006).

LAB are generally auxotrophic for several amino acids, vitamins, and nucleotides, since they have adapted to a very rich nutritional niches (Kandler and Weiss, 1986). Some species are components of the commensal microbiota naturally colonizing diverse districts within the body of humans and animals (Rossi et al. 2016). LAB genome presents a low GC-content. The order Lactobacillales encompasses several families, among which Enterococcaceae, Streptococcaceae, and Lactobacillaceae. The genus *Lactobacillus* includes the highest diversity with more than 215 different species with diverse morphological and functional features.

LAB can be used in food processing, contributing to safety of the final products and to development of positive organoleptic features. Lactic acid resulting from carbohydrate fermentation reduces pH hindering growth of pathogens and spoilage bacteria. Some LAB can produce H_2O_2 and/or bacteriocins, limiting growth of competing microorganisms. In some case they can inhibit growth of *L. monocytogenes*, and other dangerous foodborne pathogens (De Vuyst and Leroy, 2007). The use of properly selected LAB strains as starters can lead to the extension of the shelf-life of food products through inhibition of microbes responsible of deterioration.

3.1.2 The genus *Leuconostoc*

Leuconostoc is a genus of the heterofermentative Lactobacillaceae, according to a recent revision of taxonomy (Zheng et al. 2020). The general traits of *Leuconostoc* species include facultative anaerobiosis, intrinsic vancomycin resistance, catalase negativity, ovococcoid morphology, and dextran production. The genus *Leuconostoc* includes heterofermentative LAB which affect the quality of foods either positively through fermentation, or negatively through spoilage (Hemme and Foucaud-Scheunemann, 2004).

The name of the genus derives from leucos white, and nostoc, algal generic name; *Leuconostoc*, colorless nostoc. Taxonomy of *Leuconostoc* has undergone several rearrangements (Collins et al., 1993; Dicks et al., 1995; Endo and Okada, 2008; Zheng et al. 2020). For instance, the genera *Leuconostoc*, *Weissella*, and *Oenococcus* previously were all ascribed to the genus *Leuconostoc*. According to the "List of prokaryotic names with standing in nomenclature" (http://www.bacterio.net), the genus *Leuconostoc* is currently represented by 15 species, 4 subspecies for *L. mesenteroides* and 4 for *L. gelidum*.

Bacteria belonging to the genus *Leuconostoc* are often associated with plants, vegetables, fermented vegetables, fresh and fermented meat products, and wines (Nieminen et al., 2014;

Chen et al., 2020). They are facultatively anaerobic, catalase-negative, gram-positive cocci arranged in pairs or chains (Garvie, 1986). Some species play a pivotal role in industry, participating in food fermentation processes to produce fermented sausages, fermented vegetables, cereal products, and dairy products such as butter, cream, fresh and raw milk, cheese (Buckenhuskes, 1993; Caplice and Fitzgerald, 1999; Topisirovic et al. , 2006). Most of industrially relevant *Leuconostoc* strains belong to the species *L. mesenteroides* and *L. pseudomesenteroides*. They are important flavor producers for dairy fermentations and participate to most of vegetable fermentations.

Many *Leuconostoc* through the activity of the enzyme dextransucrase produce dextran polymers which have numerous applications in the pharmaceutical area (Zikmanis et al., 2020). As a whole, *Leuconostoc* spp. are of great industrial relevance, being used as new starter cultures for fermented foods and host strains to produce value-added bioproducts in the area of biotechnology (Ghoddusi, 2011).

L. gelidum is a psychrotrophic LAB commonly associated with cold-stored nutrient-rich foods of meat and vegetable origins. It belongs to the predominant microbiota at the end of shelf-life in different kinds of packaged cold-stored food products (Andreevskaya et al., 2016). The heterofermentative metabolism produces CO_2 that can result in blowing of packages also before the end of the shelf-life.

L. mesenteroides is a facultative anaerobe that performs the heterolactic fermentation under microaerophilic conditions. It usually colonizes the skin of many fruits and vegetables and requires complex growth factors and amino acids. *L. mesenteroides* subsp. *mesenteroides* is the main driver of vegetable fermentations, participating to fermentation of coffee beans, sauerkraut, and kimchi. *L. mesenteroides* subsp. *dextranicum* is used as starter culture to produce breads, having a key role in sourdough fermentations (Schleifer, 2009). *L. mesenteroides* is a bacteriocin producer and plays a role as starter also in biopreservation.

3.1.3 *Leuconostoc carnosum*

Leuconostoc carnosum is a known colonizer of food matrices, particularly the meat-related ones (Shaw and Harding, 1989, Raimondi et al., 2019, Raimondi et al., 2018, Geeraerts et al., 2018, Budde et al., 2003, van Laack et al., 1992, Parente et al., 1996).

L. carnosum is reported to grow during the first days of the shelf-life in packaged meat products, that may range up to some weeks. The charge of contaminant bacteria, often

dominated by *L. carnosum*, tends to increase and peaks up to 10^8 cells/g in some foods (Raimondi et al., 2019, Raimondi et al., 2018). These remarkably high levels of *L. carnosum* contamination were observed in modified atmosphere packaged (MAP) cooked ham, where it prevailed over the microbiota well before the expiring date (Björkroth et al. 1998; Vasilopoulos et al 2010; Raimondi et al., 2019), in other meat products such as sausages, vacuum-packaged smoked bacon, and sliced cooked poultry (Li et al., 2019; Geeraerts et al., 2018), and in processed vegetables (Hong et al., 2014; Jung et al., 2014). As a result, a high load of living *L. carnosum* may be ingested with certain foods, including many popular ready-to-eat products the consumption of which is steadily increasing all over in the world. Nonetheless any potential impact of *L. carnosum* on human health has never been assessed.

The high moisture and the low salt content, the near-neutral pH, and the availability of nutrients of meat-related products encourages microbial growth, eventually leading to spoilage. With respect to food wholesomeness, the role of *L. carnosum* is controversial. Many studies describe it as a spoilage bacterium that causes meat deterioration and affects sensorial properties by souring, discoloration, gas production, and slime formation (Shaw and Harding, 1989; Björkroth et al., 1998; Samelis et al., 2006; Raimondi et al., 2019).

On the other hand, likewise other *Leuconostoc* species, *L. carnosum* yields organic acid and hydrogen peroxide, which exert an intrinsic antimicrobial effect, and can produce class II heat-stable bacteriocins (Stiles, 1994; Hastings et al.,1994; van Laack et al., 1992; Felix et al., 1994). Typical bacteriocins produced by *L. carnosum* are leucocins, consisting of small peptides of 30-50 residues with differences due to the length of the prepeptide sequence. Their genetic information generally resides in plasmids but can occur also in the chromosome. Most of leucocins are active against both *Listeria monocytogenes* and *Listeria innocua*, but can also inhibit some lactic acid bacteria, especially at low pH (Felix et al.,1994; Parente et al., 1996; Wan et al., 2015).

Contamination of meat products with *L. carnosum* may restrict the growth of spoiling and pathogenic bacteria (Jacobsen et al., 2003, Lawton et al., 2020). As a matter of fact, the capability of *L. carnosum* to deteriorate meat products seems a strain specific feature. A survey of MAP cooked ham samples from different producers revealed that, at the end of the shelf life, *L. carnosum* dominated the microbiota of both the good and the spoiled ones (Raimondi et al., 2019). These evidences opened the perspective to develop protective starters with specifically selected strains of *L. carnosum* that may preserve meat without impacting on the sensorial properties (Budde et al., 2003).

3.2 AIM OF THE PROJECT

The dual role of *L. carnosum* in spoilage and preservation, the high loads of these bacteria entering the human diet and possibly affecting health, and the lack of studies aimed to highlight strain to strain differences, prompted to deeply explore this species. In the present study, comparative genomics was performed to investigate the main genetic determinants of the 17 strains of *L. carnosum* that have been sequenced so far. Twelve strains were recently isolated from meat matrices and newly sequenced, while five had a genome already publicly available (Table 3.1). Using the dataset of the 17 chromosomal sequences, a comparative genome analysis of the *L. carnosum* taxon was undertaken through the assessment of the phylogeny, and of pan- and core-genome. Genetic diversity, plasmids, phages, CRISPR-Cas systems, bacteriocin, antibiotic resistances, and metabolic capabilities were also investigated.

3.3 MATERIALS AND METHODS

3.3.1 Strain list

12 strains isolated in the UNIMORE MIBI lab have been cultivated for genome extraction and WGS sequencing: WC0318, WC0319, WC0320, WC0321, WC0322, WC0323, WC0324, WC0325, WC0326, WC0327, WC0328, WC0329. For the comparative genomic analysis, five publicly available *L. carnosum* genomes were also included: JB16, CB3620, MFPC16A2803, MFPA29A1405, DSM 5576T.

The strains were isolated from different sources, mostly meat-related products: modified atmosphere packaged (MAP) cooked ham (WC0318, WC0319, WC0320, WC0321, WC0322, WC0323, WC0324, WC0325, WC0329), MAP sausages (WC0326, WC0327, WC0328). vacuum packed beef carpaccio (MFPC16A2803), MAP beef carpaccio (MFPA29A1405), vacuum packaged beef (DSM 5576T) and kimchi (JB16 and CB3620).

3.3.2 Cultivation

The *L. carnosum* strains were previously isolated from cooked ham packaged in a modified atmosphere and from fresh sausages (Raimondi et al., 2018, Raimondi et al., 2019) and were stored in our laboratory's collection. For DNA extraction, each strain was cultivated in brain heart infusion broth (infusion from brain heart 8 g/L; peptic digest of animal tissue 5 g/L; pancreatic digest of casein 16 g/L; sodium chloride 5 g/L; glucose 2 g/L; disodium hydrogen phosphate 2.5 g/L, Becton Dickinson, USA) at 30 °C for 48 hours in microaerophilic conditions. Biomass was collected by centrifugation for 5 minutes at 5000 x g.

3.3.3 Genome extraction

The genomic DNA was extracted with QiAmp DNA Mini Kit (QIAGEN GmbH, Germany) following the manufacturer instruction and the concentration was determined with a Qubit fluorometer (Invitrogen, USA).

3.3.4 Sequencing

The DNA samples were submitted to Singapore Centre for Environmental Life Sciences Engineering (SCELSE), where they were tagged with Illumina TruSeq HT DNA dual barcodes for library preparation and sequenced on an Illumina HiSeq 2500 instrument (Illumina, USA). For each sample, 250-bp paired-end reads were obtained.

3.3.5 Genome assembly

The raw reads were first checked for quality with FastQC 0.11.7 (Andrews, 2010). Due to the presence of Illumina adapters and low quality reads, Cutadapt v1.16 was used to trim adapters and to remove reads with a quality score lower than 20, using the following parameters: --overlap=15 --minimum-length=30 -- quality-cutoff=20 (Martin, 2021). The postprocessed reads were checked again with FastQC to be sure that the procedure was conducted successfully. The trimmed reads were assembled with SPAdes v3.12 with the parameters --careful --cov-cutoff auto -k auto (Bankevich et al, 2012). The quality of the assemblies was evaluated using QUAST 5.0.2 (Gurevich et al., 2013). Trimming, quality checking, and assembly were performed on Galaxy platform (usegalaxy.eu) (Afgan et al., 2018). The completeness of the assemblies was determined with CheckM 1.0.8 (Parks et at., 2015). The

taxonomy was confirmed with KmerFinder on CGE server (cge.cbs.dtu.dk/services/KmerFinder/) (Hasman et al., 2014). The assembled genomes were ordered with Mauve v2.4.0 (Darling et al., 2004), using the sequence of *L. carnosum* JB6 as reference.

3.3.6 Genomes

The genome of the 12 strains of *L. carnosum* herein analyzed, published by Candeliere et al., 2020, is available with the following Genbank accessions: SAMN11618753, SAMN11618754, SAMN11618755, SAMN11618756, SAMN11618757, SAMN11618758, SAMN11618759, SAMN11618761, SAMN11618763, SAMN11618764, SAMN11618767, and SAMN11618768. Average Nucleotide Identity (ANI), digital-DNA/DNA hybridization (dDDH), and pangenome analysis also included the genomes of other *L. carnosum* strains, i.e. JB16, CBA3620, MFPC16A2803, MFPA29A1405 and DSM 5576T, available with the accessions SAMN02603179, SAMN11843679, SAMEA104699786, SAMEA104699785, and SAMN14908560 respectively.

3.3.7 Genomes analysis

Plasmids contigs were identified using plasmidSPAdes v 3.12 (Nurk et al., 2017). ANI and dDDH were calculated utilizing the web tools ANI Matrix (http://enve-omics.ce.gatech.edu/g-matrix/) and Genome-to-Genome Distance Calculator GGDC 2.1 (https://www.dsmz.de/services/online-tools/genome-to-genome-distance-calculator-ggdc), by all-against-all approach (Rodriguez-R and Kostantinidis, 2016; Meier-Kolthoff et al., 2013). The threshold for species demarcation was 95% and 70% for ANI and dDDH, respectively (Richter and Rosselló-Móra, 2009).

Gene clusters such as prophages, CRISPR/Cas systems, insertion sequences, bacteriocins genes, and antimicrobial resistances were identified with specific web tools. Prophage sequences were searched with PHASTER (http://phaster.ca/) (Arndt et al., 2016). CRISPRs and Cas genes were searched utilizing CRISPRCasFinder (https://crisprcas.i2bc.paris-saclay.fr/CrisprCasFinder/Index) (Couvin et al., 2018), with default settings and subtype clustering of Cas genes. Bacteriocins genes were searched with BAGEL 4 server (http://bagel4.molgenrug.nl/index.php) (van Heel et al., 2018). Insertion sequences were identified with ISFinder (https://isfinder.biotoul.fr/) (Siguier et al., 2006). Antimicrobial

resistance was assessed by Resistance Gene Identifier (RGI) tool of Comprehensive Antibiotic Resistance Database (CARD), processing the contigs file for "Perfect, Strict and Loose hits" (https://card.mcmaster.ca/analyze/rgi) (Alcock et al., 2020).

3.3.8 Genomes annotation and functional characterization

Genomes were annotated with Prokka, utilizing default parameters (Seemann, 2014). Roary was utilized to calculate the pangenome and to delineate core, soft core, accessory, shell, and cloud genes, utilizing Prokka annotation files (Page et al., 2015). To differentiate plasmids and chromosomal accessory genes and highlight the plasmid contribution, Roary was used to calculate the pangenome with and without the plasmid contigs. COG annotation was conducted utilizing WebMGA server (http://weizhong-lab.ucsd.edu/webMGA/server/) (Wu et al. 2011), using as input the protein files predicted by Prokka. A phylogenetic tree based on core genome alignment was constructed with FastTree (Price et al., 2009) utilizing the alignment file produced by Roary and was visualized with iTOL (Letunic and Bork, 2019).

To investigate the similarity among genomes and among plasmids, Jaccard's distance was computed based on the presence / absence of predicted genes and subjected to Principal Coordinate Analysis (PCoA).

The functional prediction of the genomes was carried out with the KEGG tools BlastKOALA (https://www.kegg.jp/blastkoala) and Mapper (https://www.genome.jp/kegg/tool/map_pathway.html) with the aim to predict the metabolic functions, such as transporters, sugars catabolism, and biosynthetic pathways of aminoacids, vitamins, and bases (Kanehisa et al., 2016; Kanehisa and Sato, 2020).

The presence of specific genes that are not included in KEGG annotation was investigated with a BLASTp search of cognate proteins recognized in a closely related taxon (i.e. in *L. carnosum*, in the genus *Leuconostoc*, in other Lactobacillales, or in other bacteria). The following enzymes for citrate metabolism and acetoin pathway were searched (Garcia-Quintanas et al. 2008): citrate permease (AAA60396) of *Leuconostoc mesenteroides* subsp. *mesenteroides,* citrate lyase alpha and beta subunit (CAA71633, CAA71632) of *Leuconostoc mesenteroides* subsp. *cremoris,* oxaloacetate decarboxylase (AFS39629) of *Leuconostoc gelidum,* alpha-acetolactate synthase (SPJ44178) of *L. carnosum*), alpha-acetolactate decarboxylase (AFT82058) of *L. carnosum*, diacetyl reductase (SPJ42929) of *L. carnosum*, 2,3-butanediol dehydrogenase (WP_135197409) of *L. carnosum*. The genes encoding the

carboxilases yielding biogenic amines were searched (Li et al. 2018; Rodrigo-Torres et al. 2019): tyrosine decarboxylase (AAN77279) and agmatine deaminase (ABS19476 and ABS19477) of *Lactobacillus brevis* (proposed *Levilactobacillus brevis*), ornithine decarboxylase (ANJ65946) of *Lactobacillus rossiae* (proposed *Furfurilactobacillus rossiae*), histidine decarboxylase of (N877767) of *Lactobacillus reuteri* (proposed *Limosilactobacillus reuteri*), and lysine decarboxylase (NP_414728) of *Escherichia coli*. For menaquinone biosynthesis, the presence of 1,4-dihydroxy-2-naphthoyl-CoA hydrolase from *Leuconostoc mesenteroides* subsp. *mesenteroides* ATCC 8293 (ABJ61187) was searched.

3. 4 RESULTS

3.4.1 General genome features

The 12 newly assembled genomes of *L. carnosum* had a mean size of 1.77 Mbp, laying in the range between 1.65 and 1.85 Mbp of strains *L. carnosum* WC0329 and *L. carnosum* WC0322, respectively, a GC content of 37.0–37.2% and an average completeness of 99.43% (Table 3.1). The genomes of the 5 already published strains had a size of 1.59–1.79 Mbp and a GC content of 37.0–37.4% (Table 3.1). A >455-fold coverage was obtained for all the 12 genomes, with a mean of 535.3-fold. The draft genomes encompassed from 11 to 40 contigs, on average 21 per genome. For all the assemblies, 1–3 contigs together constituted more than 50% of the genome size (L50 value, mean = 1.92; N50 value, mean = 645,510 bp) (Table 3.1). A total of 20 natural plasmids, ranging in size from 2.4 to 63.5 kbp, were identified in the new 12 draft genomes, 1–3 per strain (Table 3.1, Table 3.2). Twelve other plasmids occurred in 4 already available genomes (Table 3.2), while the information was not available for *L. carnosum* DSM 5576T. For all the pairwise comparisons between 17 genomes, both ANI and dDDH were higher than the corresponding threshold for species demarcation (95 and 70%, respectively), thus all the strains were confirmed to belong to the same species with only minor intra-species differences (Table 3). *L. carnosum* WC326 and WC327 resulted the same strain according to both ANI and dDDH.

The number of coding sequences (CDS) in the 12 new genomes ranged between 1,644 and 1,898, with a mean of 1,790. Single plasmids harbored from 3 to 67 genes, thus each draft

genome included 14–124 plasmid genes, accounting for 0.8–6.7% of the CDS content. The number of CDS in the 5 genomes already available ranged from 1,635 and 1,846, with plasmids of 4 strains harboring from 3 to 69 genes. All the CDS in the 17 genomes were compared by a blast all-against-all approach to identify orthologous gene groups and construct pan- and core genome matrices. The pangenome of *L. carnosum* encompassed 3,221 orthologous genes, while the core genome included 1,383 chromosomal genes (on average 43% of total genes). The accessory genome was composed of 689 shell genes (548 of which in the chromosome and 141 in plasmids) and 1,149 cloud genes (868 of which in the chromosome and 281in plasmids) (Fig. 3.1A). As per the Heap's law, the value of γ was 0.35 for 17 genomes, and thus *L. carnosum* pangenome was considered open (Fig. 3.1B) (Tettelin et al., 2008).

The Jaccard distance matrix among genomes, due to the presence/absence of the genes, is displayed in the PCoA plot of Fig. 3.2A. Five genomes laying at negative values of PCo1 (i.e. WC0319, WC0324, WC0328, JB16, and CBA3620) were separated from the others dispersed at positive values of PCoA1, among which 3 lay at positive values of PCo2 (WC0318, WC0326, and WC0327), and the others grouped together at lower values. The phylogenetic relationship between the 17 strains was constructed using relative hierarchical clustering based on core genome alignment (Fig. 3.2B). The low phylogenetic distances confirmed a close relatedness among the *L. carnosum* strains, even though sequence differences within the core genome revealed three major clades. All the genomes that lay at positive PCo1 values were in the same phylogenetic clade, while the five genomes that lay at negative PCo1 values were clearly separated in two different clades.

A comparative analysis of chromosomal sequences was carried out for all the genomes. The 1,383 genes of the core genome were ascribed to 24 COG families (Fig. 3.3). Functional distribution of the COGs of the core showed that the majority encoded components of the information processing systems associated to translation, ribosomal structure, and biogenesis (13.1%) and of amino acid transport and metabolism (10.1%). The core also included 87 genes of function unknown (6.1%) and 114 genes with only a general prediction of biochemical activity (8.0%).

Table 3.1 General genomic features of the 17 strains of *Leuconostoc carnosum* analyzed in this work.

Strains	Genome size (bp)	No. of contigs	N50	L50	Coverage	G+C (%)	No. of CDS	No. of tRNA	No. of prophages	Completeness (%)	Plasmids
WC0318	1,745,630	15	1142374	1	638	37.2	1739	51	2	99.476	pFRA18, pALB18
WC0319	1,700,071	11	1137580	1	521	37.1	1679	49	1	99.476	pELI19
WC0320	1,812,114	18	1109738	1	561	37.1	1853	51	3	99.476	pFRA20, pALB20, pICCOLO
WC0321	1,804,293	40	1106035	1	532	37.1	1854	51	2	99.476	pFRA21, pGLO21, pALB21
WC0322	1,853,239	16	306690	2	515	37.0	1898	50	3	99.476	pALB22, pLQ22
WC0323	1,773,698	23	256026	3	560	37.2	1802	51	2	99.476	pFRA23, pGLO23
WC0324	1,765,760	13	1137276	1	460	37.1	1768	49	3	99.476	pELI24
WC0325	1,830,248	27	229291	3	502	37.2	1881	51	4	99.476	pLQ25, pICCOLO
WC0326	1,770,048	22	412621	2	574	37.2	1767	49	1	99.476	pSTE
WC0327	1,769,683	20	412621	2	568	37.2	1768	49	1	99.476	pSTE
WC0328	1,815,949	36	239739	3	538	37.1	1826	51	3	98.953	pALAN28, pUFO, pELI28
WC0329	1,650,966	14	256123	3	455	37.2	1644	51	1	99.476	pFRO29
JB16	1,645,096	5	1645096	1	-	37.2	1686	64	1	-	pKLC1, pKLC2, pKLC3, pKLC4
CBA3620	1,590,008	3	1590008	1	-	37.4	1635	66	3	-	unnamed1, unnamed2
MFPC16A2803	1,786,865	50	242386	3	-	37.0	1846	32	2	-	pMFPC16A2803A, pMFPC16A2803B, pMFPC16A2803C, pMFPC16A28E
MFPA29A1405	1,634,774	22	165377	4	-	37.3	1660	46	2	-	pMFPA29A1405B, pMFPA2A1405C
DSM 5576T	1,820,660	21	147749	4	-	37.0	1740	47	1	-	Not yet identified

Table 3.2 Size and number of CDS predicted in the plasmids.

Plasmid	Size (bp)	No. of CDS
pALAN28	63454	72
pALB18	27911	32
pALB20	42988	45
pALB21	29994	37
pALB22	41757	49
pELI19	18422	21
pELI24	13067	14
pELI28	9860	12
pFRA18	32446	34
pFRA20	63088	67
pFRA21	42529	47
pFRA23	48616	51
pFRO29	50548	57
pGLO21	41385	40
pGLO23	39823	41
pICCOLO	2351	3
pLQ22	36356	44
pLQ25	42858	41
pSTE	39016	37
pUFO28	18346	16
pKLC1	21990	21
pKLC2	29615	34
pKLC3	40165	45
pKLC4	36602	39
unnamed1	57926	69
unnamed2	53399	67
pMFPC16A2803A	78165	88
pMFPC16A2803B	22807	36
pMFPC16A2803C	11421	12
pMFPC16A28E	2344	3
pMFPA29A1405B	18286	24
pMFPA2A1405C	13778	14

Table 3.3 ANI and dDDH values calculated between the 12 strains analyzed in this work and the 5 publicly available. The blank diagonal line marks 100% value.

	WC0318	WC0319	WC0320	WC0321	WC0322	WC0323	WC0324	WC0325	WC0326	WC0327	WC0328	WC0329	JB16	CBA3620	MFPC16A2803	MFPA29A1405	DSM 5576
WC0318		99.0	99.8	99.7	99.6	99.7	99.0	99.7	99.7	99.7	99.1	99.7	99.1	99.1	99.8	99.8	99.6
WC0319	91.7		98.9	99.0	99.0	99.0	100	99.0	98.9	98.9	99.2	99.0	99.2	99.2	99.0	99.0	98.9
WC0320	98.2	90.8		99.7	99.7	99.8	98.9	99.7	99.8	99.8	99.0	99.8	99.1	99.0	99.8	99.8	99.7
WC0321	97.7	91.6	96.8		99.8	100	99.0	99.8	99.8	99.8	99.1	99.9	99.1	99.2	99.9	100	99.8
WC0322	97.5	91.7	95.5	97.7		99.8	99.0	99.8	99.7	99.7	99.0	99.9	99.1	99.1	99.8	99.9	99.8
WC0323	98.3	91.5	97.4	99.6	97.9		99.0	99.8	99.8	99.8	99.1	99.9	99.1	99.2	99.9	100	99.9
WC0324	91.7	99.7	92.2	90.7	91.4	91.4		99.0	98.9	98.9	99.1	99.0	99.2	99.2	99.0	99.0	98.9
WC0325	97.4	92.1	97.1	97.2	96.9	97.6	91.7		99.7	99.7	99.1	99.9	99.2	99.1	99.9	99.9	99.8
WC0326	98.0	91.0	98.0	98.1	98.2	98.2	91.0	97.6		100.0	99.1	99.8	99.1	99.1	99.9	99.8	99.7
WC0327	98.0	91.0	98.0	98.1	98.2	98.2	91.0	97.7	100		99.1	99.8	99.1	99.1	99.9	99.8	99.7
WC0328	92.6	93.4	90.5	91.8	90.6	92.3	92.9	91.3	92.3	92.3		99.1	100	99.8	99.2	99.2	99.1
WC0329	98.3	91.7	98.3	99.6	99.4	99.6	91.8	99.2	98.4	98.4	93.3		99.2	99.2	99.9	100	99.9
JB16	93.2	94.2	92.9	93.4	92.1	93.4	94.2	93.4	92.7	92.7	99.7	93.8		99.9	99.2	99.2	99.1
CBA3620	92.9	93.8	92.2	93.8	92.5	93.8	93.8	93.4	92.7	92.7	98.4	93.8	99.1		99.1	99.2	99.1
MFPC16A2803	98.8	92.1	98.1	99.1	97.4	98.7	91.9	98.9	99.2	99.2	92.6	99.5	93.6	93.4		99.9	99.9
MFPA29A1405	98.5	92.2	98.7	99.7	99.3	99.6	92.2	99.0	98.9	98.9	93.6	100	93.7	93.8	100		99.9
DSM 5576T	96.7	91.3	96.7	98.2	97.2	98.7	91.5	98.6	97.1	97.1	92.1	99.1	91.9	93.5	99.6	99.4	

dDDH

ANI

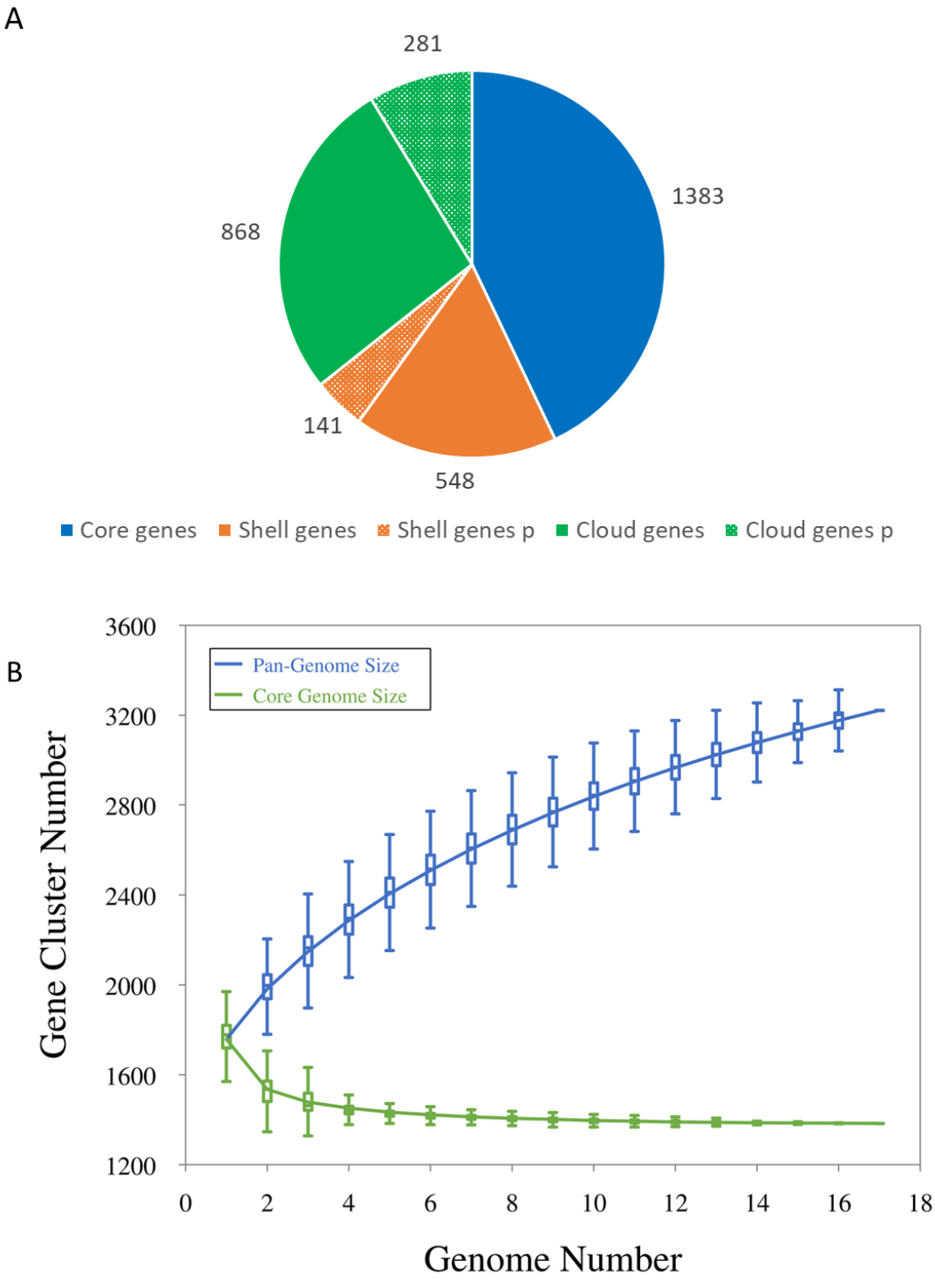


Figure 3.1 (A) Distributions of CDS found in the pangenome of *L. carnosum*: core genes (blue), shell genes (orange), cloud genes (green) in chromosome (plain colors) and plasmids (dotted colors). (B) Estimation of the pangenome (blue) and the core genome (green) by including genomes one by one.

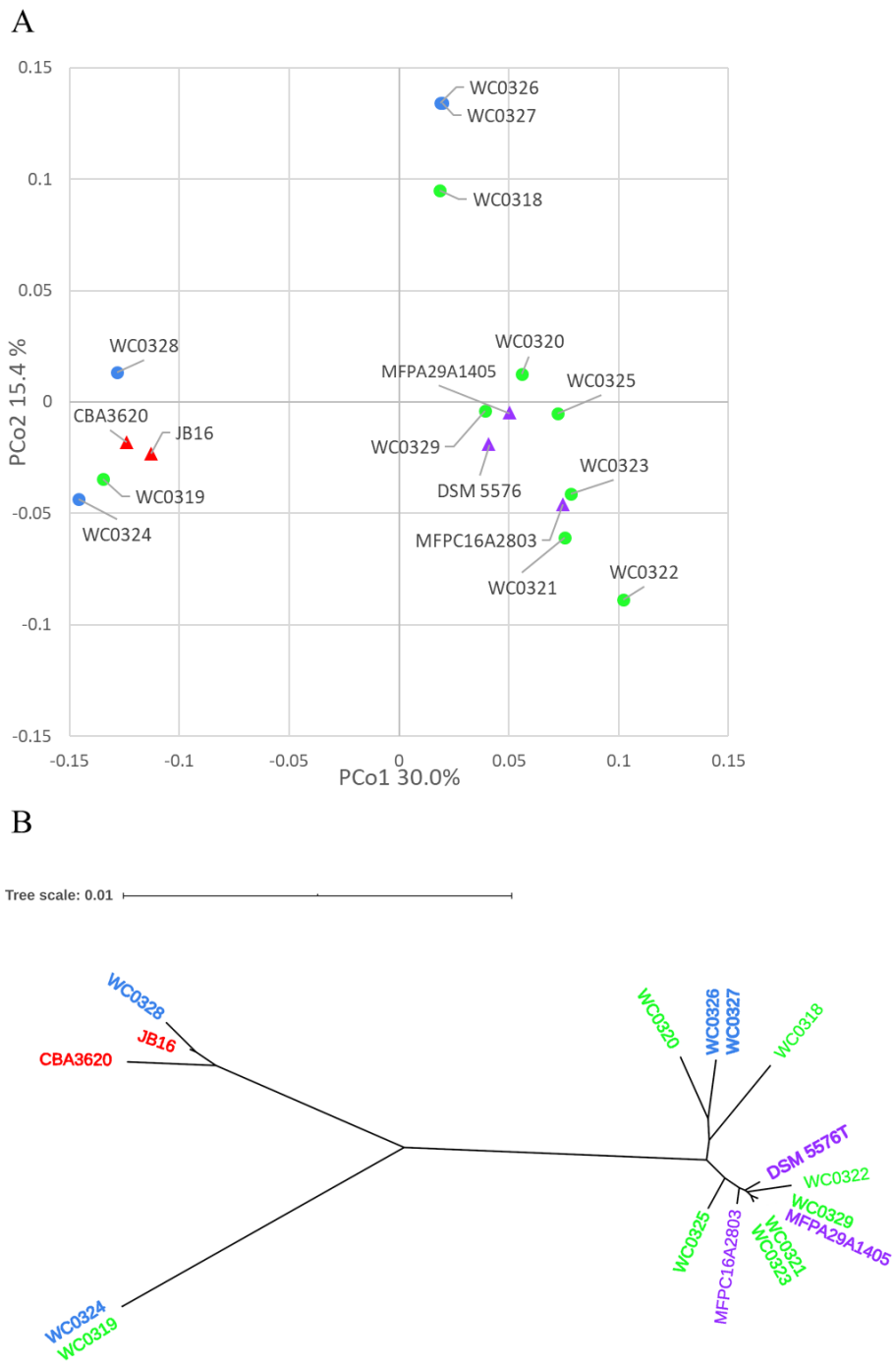


Figure 3.2 (A) PCoA visualization of Jaccard distances among genomes, based on shared genes. Circles are the genomes published by Candeliere et al. (2020); triangles are the genomes already published. (B) Neighbor joining unrooted phylogenetic tree based on core genome alignment. In both panels, colors indicate the source of the strain: MAP cooked ham, green; MAP sausage, blue; packaged beef products, purple; kimchi, red.

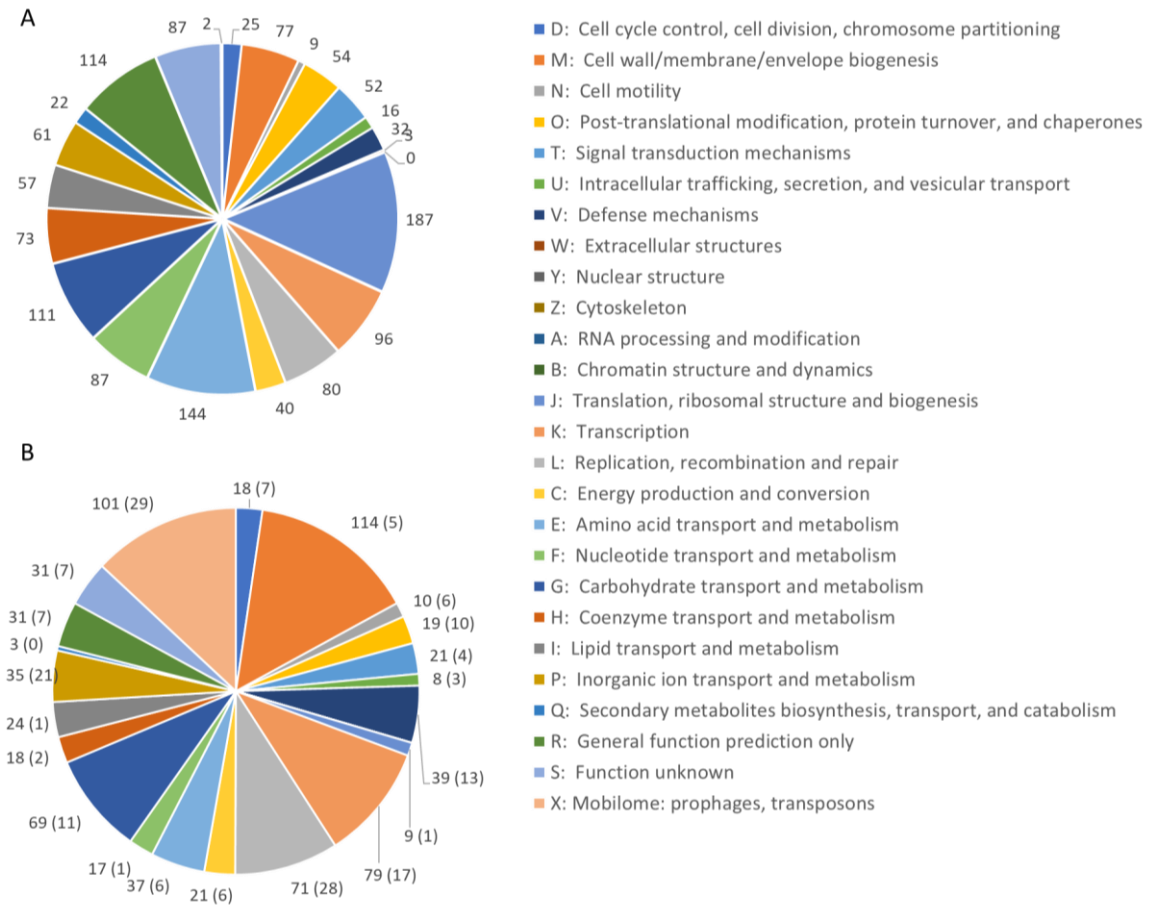


Figure 3.3 COG class abundances in core genome (A) and in the accessory genome (B). The number of CDS predicted within each class is reported. The CDS in plasmids are in brackets.

3.4.2 Mobilome

The chromosome sequences were analyzed for the presence of mobile elements, such as genes putatively encoding insertion sequences (IS) and transposases. 29 genes encoding transposases were identified in the accessory genome, 20 of them as part of IS. Except for *L. carnosum* MFPA29A1405 that did not harbor any, all the strains harbored at least one IS. The IS belonged to the families IS3, IS5, IS6, IS30, IS256, and IS1182. The IS presented similarities with counterparts of other Lactobacillallales: IS3, *Lactobacillus sakei* (proposed *Latilactobacillus sakei* by Zheng et al., 2020) and *Weisella cibaria*; IS5 and IS1182, *Lactobacillus plantarum* (proposed *Lactiplantobacillus plantarum* by Zheng et al., 2020); IS6, *Enterococus faecium* and *Leuconostoc mesenteroides*; IS30, *L. plantarum*, *Leuconostoc*

lactis, and *Pediococcus pentosaceus*; IS256, *Enterococcus hirae* and *Lactobacillus helveticus*. The most represented were IS3 and IS30 transposases (7 and 6 genes, respectively).

3.4.3 Plasmids

The 32 plasmids were annotated. The Jaccard distance matrix among plasmids, due to the presence of shared genes and predicted functions, is displayed in the PCoA plot of Fig. 3.4. pALB plasmids, pKLC4, and pMFPC16A2803B harbored genes encoding all the functions for conjugative transfer, including mating pair formation and DNA replication, mobilization, and transfer.

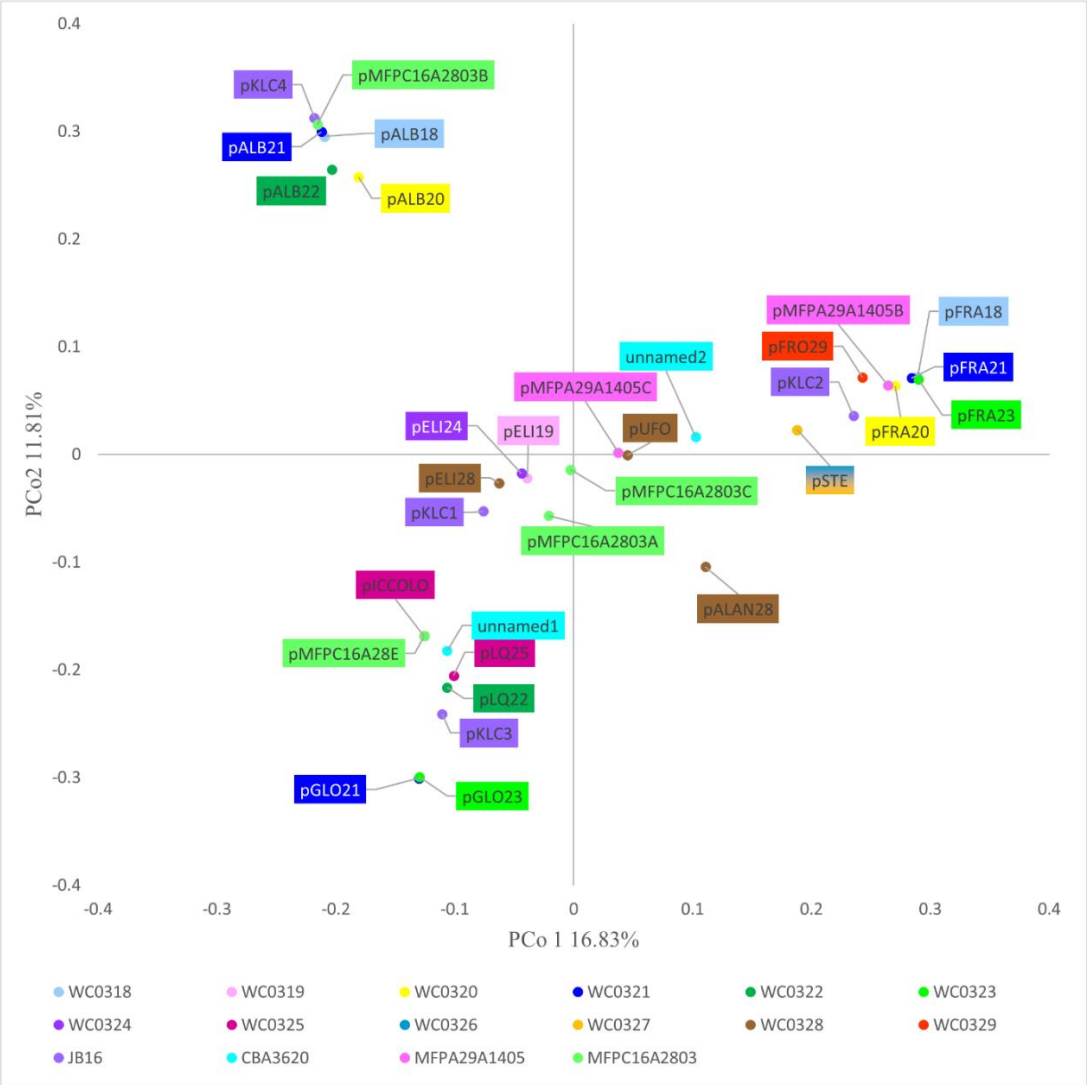


Figure 3.4 PCoA visualization of Jaccard distances among plasmids, based on shared genes.

The series pGLO and pLQ and the plasmids pMFPA29A1405B, pMFPC16A28E, and pKLC1 had some conjugation genes, but lacked the genes encoding mobilization proteins (relaxases), resulting not mobilizable. The series pELI and pSTE, pFRO29, pALAN28, pUFO, pKLC2, pKLC3, pMFPC16A2803A, pMFPC16A2803C, pMFPA2A1405C, unnamed1, and unnamed2 were mobilizable but not conjugative plasmids, for the presence of 2-4 mobility-related genes.

pICCOLO and pMFPC16A28E were likely the same small plasmid of approximately predicted 2,350 bp, occurring in the strains *L. carnosum* WC0320, WC0325 and MFPC16A2803. They showed a very high identity (> 97.8%) with the small plasmid pM411 (2303 bp) of *L. plantarum*, replicating via the rolling circle mechanism. They encoded a N-acetyltransferase and the rolling circle initiator protein RepB. Genes encoding RepB proteins occurred in 16 of the 32 replicons, such as in the pELI series.

IS3 transposase was found in the series pFRA and in pSTE, pFRO29, pKLC2, pKLC3, pKLC4, pMFPC16A2803A, pMFPC16A2803C, pMFPA29A1405B, unnamed1, and unnamed2. pSTE and pKLC2 also shared IS30 transposase while pFRO29, pMFPA29A1405B, and the series pFRA shared IS110, that contributed to their clustering. The series of pGLO plasmids harbored the IS6 transposase and other common recombinases and transposases, whereas the pSTE plasmids and pKLC2 encoded both IS3 and IS30 transposases.

Peculiar was the scarce presence or absence of genes associated to transposases in pALB plasmids, albeit the large size of these replicons, likely due to their conjugative feature.

The plasmid unnamed 2 and series pALB and pELI were also characterized by genes encoding DNA-methyl transferase and/or restriction endonucleases.

Known antibiotic resistance genes were not identified in the new plasmids. Genes encoding ImmA/IrrE and RelE/ParE toxin-antitoxin systems were present in the series pSTE, pLQ25, and pGLO and in pKLC3, pMFPC16A2803A, and unnamed1. The latter also harbored RelB/DinJ system. ImmA/IrrE, RelE/ParE, and RelB/DinJ thus added to HicB family and PemK/MazF toxin-antitoxin systems located in the core genome.

The plasmids often harbored genes involved in stress responses and participating to arsenic and heavy metal detoxification. Genetic determinants of thioredoxin were identified in pALB and pGLO series and in pKLC1, pKLC3, pKLC4, pMFPC16A2803A,

pMFPC16A2803B, unnamed1, and unnamed2 plasmids. Interestingly, only pKLC1, pMFPC16A2803A, and pGLO series also encoded a glutathione reductase gene, which product can contribute for regulating cellular redox homeostasis, in concert with the glutathione peroxidase present in the core genome.

Genes associated to bacteriocin production were detected in 4 plasmids. pFRA21, pFRA23, and pFRO29 held the genes encoding the leucocin-B and an immunity protein. Moreover, the genes encoding an additional immunity protein and an accessory secretion protein for the bacteriocin were predicted in pFRA21 and pFRA23. Conversely, pFRA18 did not harbor any gene associated to bacteriocin production. pALAN28 had 3 other genes linked to antibacterial activity: a lactococcin, a mesentericin, and another immunity protein.

3.4.4 Reconstructed metabolic pathways

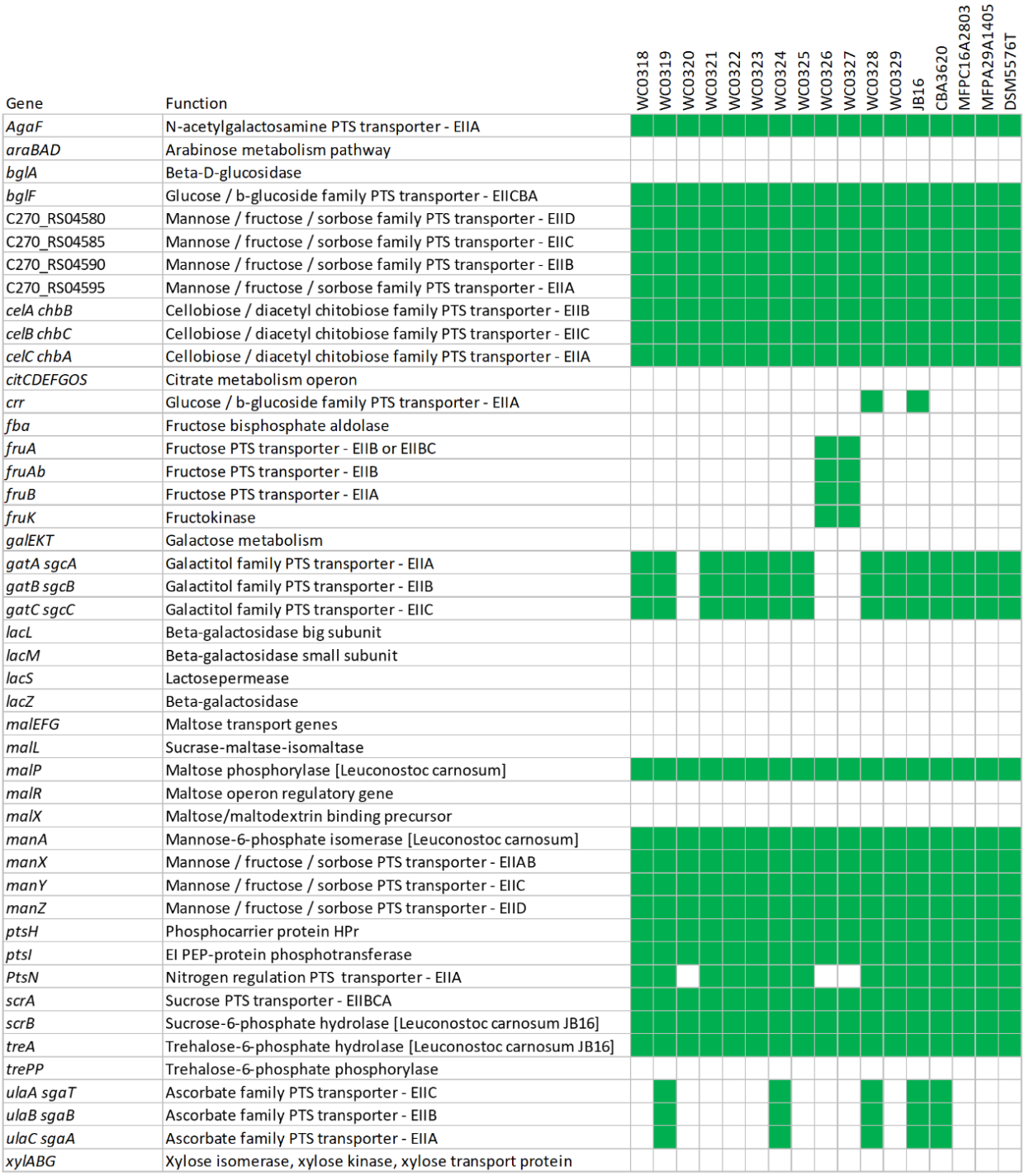
Pathway reconstruction based on KEGG gene annotation of the core genome predicted at least 5 complete PTS transporters, i.e. one for sucrose, two belonging to mannose-fructose-sorbose family, and one each to glucose–β-glucoside and cellobiose–diacetyl chitobiose families. On the other hand, a complete PTS transporter for fructose and two belonging to ascorbate and galactitol families were accessory. Genes encoding other PTS components were also recognized as part of incomplete transporters in the core or in the accessory genome, such as EIIA components (or proteins with EIIA domains) with similarity to *agaF*, *crr*, and *ptsN* that may have a role in the uptake of substrates if they could interact with unidentified EIIB and EIIC components (Figure 3.5).

Complete ABC transporters for the uptake of ribose and xylose, oligopeptides, many aminoacids, and other compounds or nutrients such as nucleosides, amines, biotin, phosphate, manganese, were predicted in the core genome. The accessory genome included a complete ABC transporter for the uptake of choline and two exporters presenting similarity to those encoded by the operons *TagGH* and *NodIJ* for the export of teichoic acid and capsular polysaccharides, respectively. Three eukaryotic-type ABC exporters were also recognized in the core genome. One is involved in cytochrome export and assembly, while two have similarity to the multidrug efflux systems *mdlAB* and *lmrCD*. The accessory genome included a further eukaryotic type ABC transporter, similar to *blpA* that is involved in bacteriocin export. *L. carnosum* is heterofermentative, with the sugars being metabolized via pentose phosphate and phosphoketolase pathway (PKP), yielding lactate, acetate or ethanol, and CO_2. The pivot enzyme of PKP, D-xylulose-5-phosphate phosphoketolase (EC

4.1.2.9), was encoded in the core genome. Consistently, the main bacterial glycolytic pathways were incomplete, the Embden-Meyerhof's one lacking 6-phosphofructokinase and fructose-bisphosphate aldolase, and the Entner-Doudoroff's one lacking phosphogluconate dehydratase and 2-dehydro-3-deoxyphosphogluconate aldolase. Leloir's pathway for galactose utilization was incomplete in all the strains, even though some strains harbored the genes encoding galactose epimerase (EC 5.1.3.3) and UDP-glucose 4-epimerase (EC 5.1.3.2).

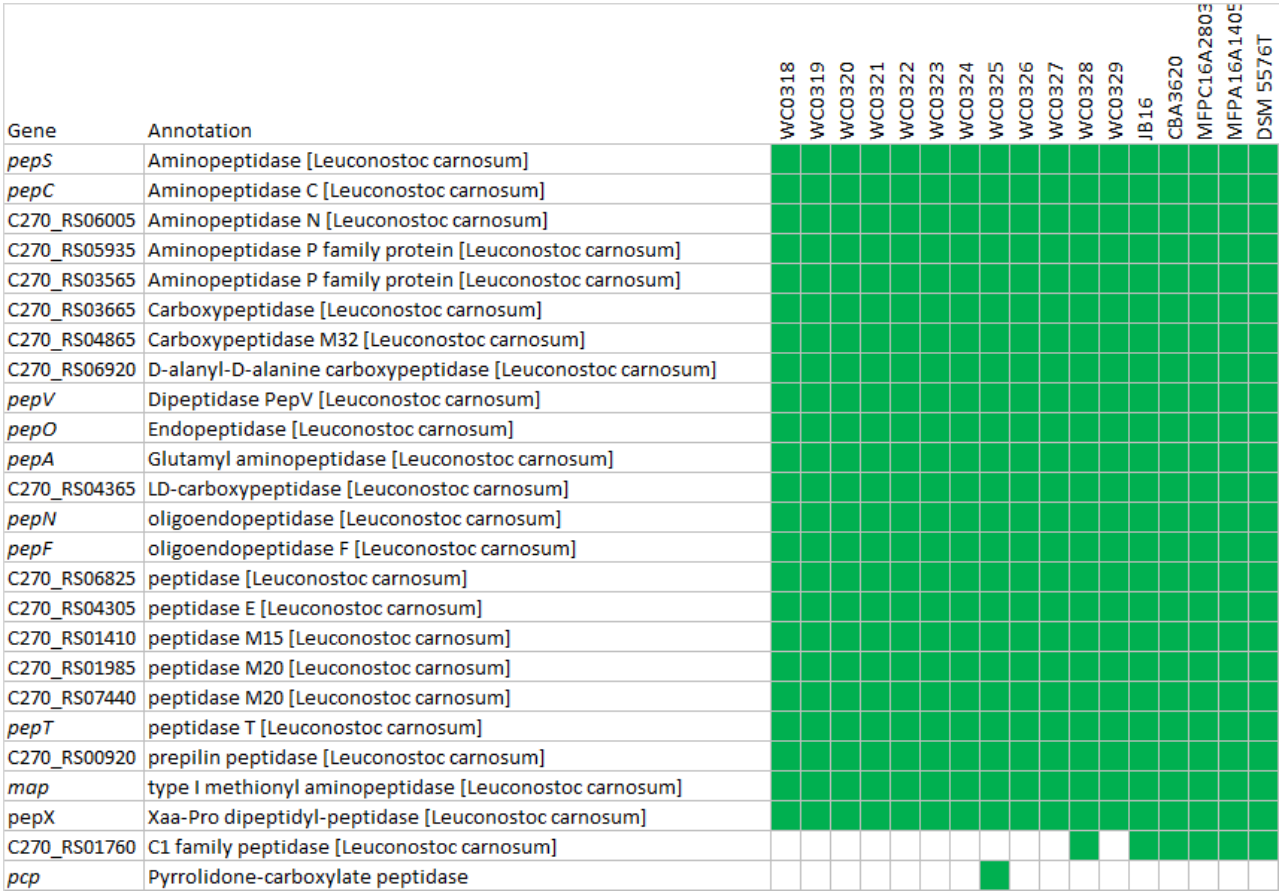
The whole route for ascorbate transformation into xylulose 5-phosphate and channeling into the pentose phosphate pathway was accessory, the seven enzymes being encoded in some strains, or all missing in the others. The citrate pathway was absent in all the strains, due to the lack of the *Cit* genes encoding citrate transporter, citrate lyase, and oxaloacetate decarboxylase. Nonetheless, all the strains harbored the genes encoding α-acetolactate synthase, α-acetolactate decarboxylases, and diacetyl acetoin reductases, and may yield acetoin and 2,3 butanediol from pyruvate.

A total of 25 peptidase genes was identified (Figure 3.6). The vast majority were encoded by core genes, while a minority were accessory. Reconstruction of amino acid biosynthesis pathways suggested that the strains did not differ regarding the ability to produce amino acids, except for tryptophan. Some metabolic routes involved in the addition of an amino group to carbon backbones seemed interrupted, thus hampering the synthesis of some amino acids. In particular, the genes encoding glutamate dehydrogenase and glutamate synthase for the biosynthesis of glutamate from 2-oxoglutarate were not identified. The lack of alanine dehydrogenase and aspartate aminotransferase interrupted the synthesis of alanine and aspartate, respectively, while the pathway of serine production from glycolysis intermediates seemed ineffective due to the lack of phosphoserine phosphatase.



Gene	Function
AgaF	N-acetylgalactosamine PTS transporter - EIIA
araBAD	Arabinose metabolism pathway
bglA	Beta-D-glucosidase
bglF	Glucose / b-glucoside family PTS transporter - EIICBA
C270_RS04580	Mannose / fructose / sorbose family PTS transporter - EIID
C270_RS04585	Mannose / fructose / sorbose family PTS transporter - EIIC
C270_RS04590	Mannose / fructose / sorbose family PTS transporter - EIIB
C270_RS04595	Mannose / fructose / sorbose family PTS transporter - EIIA
celA chbB	Cellobiose / diacetyl chitobiose family PTS transporter - EIIB
celB chbC	Cellobiose / diacetyl chitobiose family PTS transporter - EIIC
celC chbA	Cellobiose / diacetyl chitobiose family PTS transporter - EIIA
citCDEFGOS	Citrate metabolism operon
crr	Glucose / b-glucoside family PTS transporter - EIIA
fba	Fructose bisphosphate aldolase
fruA	Fructose PTS transporter - EIIB or EIIBC
fruAb	Fructose PTS transporter - EIIB
fruB	Fructose PTS transporter - EIIA
fruK	Fructokinase
galEKT	Galactose metabolism
gatA sgcA	Galactitol family PTS transporter - EIIA
gatB sgcB	Galactitol family PTS transporter - EIIB
gatC sgcC	Galactitol family PTS transporter - EIIC
lacL	Beta-galactosidase big subunit
lacM	Beta-galactosidase small subunit
lacS	Lactosepermease
lacZ	Beta-galactosidase
malEFG	Maltose transport genes
malL	Sucrase-maltase-isomaltase
malP	Maltose phosphorylase [Leuconostoc carnosum]
malR	Maltose operon regulatory gene
malX	Maltose/maltodextrin binding precursor
manA	Mannose-6-phosphate isomerase [Leuconostoc carnosum]
manX	Mannose / fructose / sorbose PTS transporter - EIIAB
manY	Mannose / fructose / sorbose PTS transporter - EIIC
manZ	Mannose / fructose / sorbose PTS transporter - EIID
ptsH	Phosphocarrier protein HPr
ptsI	EI PEP-protein phosphotransferase
PtsN	Nitrogen regulation PTS transporter - EIIA
scrA	Sucrose PTS transporter - EIIBCA
scrB	Sucrose-6-phosphate hydrolase [Leuconostoc carnosum JB16]
treA	Trehalose-6-phosphate hydrolase [Leuconostoc carnosum JB16]
trePP	Trehalose-6-phosphate phosphorylase
ulaA sgaT	Ascorbate family PTS transporter - EIIC
ulaB sgaB	Ascorbate family PTS transporter - EIIB
ulaC sgaA	Ascorbate family PTS transporter - EIIA
xylABG	Xylose isomerase, xylose kinase, xylose transport protein

Figure 3.5 Genetic potential for metabolism of carbohydrates based on the presence (green) or absence (white) of predicted transporters or enzymes.



Gene	Annotation	WC0318	WC0319	WC0320	WC0321	WC0322	WC0323	WC0324	WC0325	WC0326	WC0327	WC0328	WC0329	JB16	CBA3620	MFPC16A2803	MFPA16A1405	DSM 5576T
pepS	Aminopeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepC	Aminopeptidase C [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS06005	Aminopeptidase N [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS05935	Aminopeptidase P family protein [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS03565	Aminopeptidase P family protein [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS03665	Carboxypeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS04865	Carboxypeptidase M32 [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS06920	D-alanyl-D-alanine carboxypeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepV	Dipeptidase PepV [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepO	Endopeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepA	Glutamyl aminopeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS04365	LD-carboxypeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepN	oligoendopeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepF	oligoendopeptidase F [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS06825	peptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS04305	peptidase E [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS01410	peptidase M15 [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS01985	peptidase M20 [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS07440	peptidase M20 [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepT	peptidase T [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS00920	prepilin peptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
map	type I methionyl aminopeptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
pepX	Xaa-Pro dipeptidyl-peptidase [Leuconostoc carnosum]	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
C270_RS01760	C1 family peptidase [Leuconostoc carnosum]											■		■	■	■	■	■
pcp	Pyrrolidone-carboxylate peptidase								■									

Figure 3.6 Predicted genes encoding peptidases (green) in the genome of *L. carnosum*.

The pathways depending on pre-formed serine to yield glycine and cysteine, on glutamate to yield glutamine, arginine, and proline, and on aspartate to yield asparagine, threonine, and methionine were complete in the core genome. Many genes encoding enzymes of lysine biosynthesis from aspartate were predicted in the core genome. The biosynthetic pathway seemed complete from aspartate to 2,3,4,5-tetrahydrodipicolinate. Some enzymes ascribable to the bacterial routes transforming this intermediate into lysine (i.e. the succinylase, acetylase, and dehydrogenase pathways) could be predicted, yet none of the routes seemed complete due to the lack of one or two enzymes. Interestingly, the carboxy-lyase catalyzing lysine formation from the last intermediate onto which the three routes converge (meso-2,6-diaminoheptanedioate) was predicted.

The anabolic pathways leading to branched chain amino acids were all complete in the core genome. With regards to aromatic amino acids, the route from 5-phosphoribosyl diphosphate to histidine and the shikimate pathway from erythrose 4-phosphate to chorismate were complete. The pathway leading to tyrosine seemed interrupted due to the lack of chorismate mutase, while the pathway leading to phenylalanine also lacked prephenate dehydratase. Nonetheless, the other enzymes necessary for the synthesis of tyrosine and phenylalanine were predicted. The metabolic pathway for tryptophan was complete in most of the strains while it was absent in *L. carnosum* WC0318, WC0319, and WC0324.

All the strains harbored the whole set of genes and operons for the synthesis of adenine and guanine ribonucleotides from inosine monophosphate, and of pyrimidine ribonucleotides from UMP. On the other hand, the pathway for uridine monophosphate biosynthesis from glutamine missed the gene encoding the aspartate carbamoyl transferase regulatory subunit, whereas the catalytic subunit was present. For pyrimidine deoxyribonucleotide biosynthesis from CDP/CTP, all the strains lacked the gene encoding the dCTP deaminase.

All the strains seemed unable to synthetize any biogenic amine, based on the absence of the genes encoding histidine, lysine, tyrosine, and ornithine decarboxylases (*hdcA*, *ldc*, *tyrDC*, and *odc* respectively), and agmatine deiminase (*aguD* and *aguA*).

The metabolic pathways for the biosynthesis of most vitamins and cofactors were missing or largely incomplete in all the strains, that seemed unable to synthetize thiamine, riboflavin, pyridoxal, NAD, pantothenate, biotin, folate, heme, cobalamin, and ubiquinone. Only the pathway for menaquinone biosynthesis was complete in all the strains. In this pathway, the module encoding 1,4-dihydroxy-2-naphthoyl-CoA hydrolase, a thioesterase of PaaI family, was not predicted using KEGG but was found utilizing the sequence of the cognate protein of *L. mesenteroides* ATCC 8293 as query for a BLASTp search against the proteins of *L. carnosum*. A PaaI family thioesterase presenting 68% of identity was detected in all the strains, suggesting that they all have the whole set of genes for the biosynthesis of menaquinone.

3.4.5 Stress responses and two-components signal transduction systems

The pangenome of *L. carnosum* lacked genetic determinants associated to superoxide dismutase, catalase, and *de novo* biosynthesis of glutathione. On the other hand, the core genome harbored other genes for coping with oxidative stress and reactive oxygen species,

such as those encoding thioredoxin and thioredoxin reductase, a peroxidase, and two peroxiredoxins. Moreover, genes encoding thioredoxin-related proteins were harbored by 14 out of 32 plasmids. The core genome harbored the gene encoding a glutathione peroxidase (EC 1.11.1.9), that relies on pre-formed glutathione for H_2O_2 detoxification, while two plasmids encoded the counterpart encoding the glutathione reductase.

The genes of the stringent response mediated by the alarmone nucleotides (P)PPGPP were present in the core genome. Putative members of the two-component system were identified: 6 histidine kinases (HK) and 7 response regulators (RR) were located in the core genome, and single HK and RR in the accessory portion. Four complete two-components systems presented similarity to known regulators such as PhoR/PhoP, VicK/VicR, AgrC/AgrA, and CiaH/CiaR.

3.4.6 Antibiotic resistance

The search for antibiotic resistance genes with RGI tool did not return any hit marked as "Perfect" or "Strict". Many lactic acid bacteria possess intrinsic resistance to vancomycin due to D-alanine D-alanine ligase encoded by the gene *ddl*. This resistance is correlated to the presence of a conserved phenylalanine residue (F) in the active site of the enzyme, while the sensitive strains possess a tyrosine residue (Y) in this position (Campedelli et al., 2019). The multiple sequence alignment of the *ddl* enzyme highlighted that all the17 strains of *L. carnosum* possess the F type enzyme, resulting in intrinsic vancomycin resistance.

3.4.7 Bacteriophages and phage defense

The genome sequences were investigated for prophages using PHASTER. For each strain, the tool identified 1 to 4 prophage gene clusters. Several intact prophages and a few incomplete ones were detected. Complete sequences of putatively active prophages were found in 9 out of 17 genomes (WC0319, WC0320, WC0321, WC0322, WC0323, WC0324, WC0325, WC0328, MFPC16A2803). In all the strains without active prophages, sequences of prophages likely inactive were identified. In 3 strains, WC0321, WC0328 and CBA3620, questionable regions have also been found.

A type II CRISPR-Cas (Clustered Regularly Interspaced Short Palindromic Repeats) locus was found in the chromosomes of *L. carnosum* WC0319 and WC0324. The CRISPR array of WC0319 and WC0324 contained 16 and 11 spacers, respectively, the 11 of WC0324

identified as a first string of WC319 array, followed by other 5 unshared spacers. This arrangement suggested a iterative acquisition of the spacers and a common origin of the two strains, in agreement with the negligible phylogenetic distance, with the close position in the PCoA plot, with the hold of single similar plasmids (pELI19 and pELI24).

3.4.8 Bacteriocin production

Screening of the entire genomes of *L. carnosum* using the bacteriocin database BAGEL 4 revealed the presence of 5 types of putative bacteriocin encoding loci (Table 3.4). Plasmids of *L. carnosum* WC0321, WC0323, and WC0329 harbored the genes for the synthesis of leucocin B. The genome of *L. carnosum* WC0321 and WC0323 encompassed also the gene encoding the transporter LanT, missing in WC0329. Most of the strains encoded a chromosomal lactococcin-like protein with double glycine leader peptide. No further information on the type of lactococcin was yielded by the prediction tool. Genes encoding mesentericin B105, mesentericin Y105, and/or a mesentericin-like protein were found in the genome of *L. carnosum* WC0328 and *L. carnosum* MFPC16A2803.

Table 3.4 Putative bacteriocins genes present in the strains.
Genes are chromosomal unless a plasmid indication is given in brackets.
– indicates absence of putative bacteriocins.

Strain	Bacteriocin
WC0318	Lactococcin-like bacteriocin
WC0319	–
WC0320	Lactococcin-like bacteriocin
WC0321	Leucocin B (pFRA21), lactococcin-like bacteriocin
WC0322	Lactococcin-like bacteriocin
WC0323	Leucocin B (pFRA23), lactococcin-like bacteriocin
WC0324	–
WC0325	Lactococcin-like bacteriocin
WC0326	Lactococcin-like bacteriocin
WC0327	Lactococcin-like bacteriocin
WC0328	Mesentericin B105, mesentericin Y105
WC0329	Leucocin B (pFRO29), lactococcin-like bacteriocin
JB16	–
CBA3620	–
MFPC16A2803	Lactococcin-like bacteriocin, mesentericin-like protein, mesentericin B105
MFPA29A1405	Lactococcin-like bacteriocin
DSM 5576T	–

3.5 DISCUSSION

In this study, all the genomes available for *L. carnosum* were compared to decipher their metabolic and functional potential and to determine the ecological role in food matrices (mainly meat) and the impact on food transformations. Most strains (15) were isolated from packed meat products of various origin (cooked ham, sausages, and beef), while a minority were isolated from kimchi (Raimondi et al., 2018; Raimondi et al., 2019; Jung et al. 2012).

Genome comparison indicated that the 17 strains of *L. carnosum* are a compact group of bacteria, with values of sequence similarity much higher than the threshold required for species demarcation (ddDH > 90.5%; ANI > 98.9%), albeit their pangenome remained open.

Differences within the core genome, revealed by sequence alignment, indicated three clades of strains. The clade encompassing most of the strains, all isolated from meat, was also separated in PCoA plot computed from the matrix of presence/absence of genes. The two strains isolated from kimchi were closely related, laying in the same clade and PCoA cluster, together with meat strains. Differences of location in PCoA plot with respect to phylogenetic distances can be attributed to accessory genes, with some contribution of genes encoded by plasmids. For instance, *L. carnosum* WC0328 is in a different clade than *L. carnosum* WC0324 and WC0319, but they all cluster together in PCoA plot and harbor a plasmid of the pELI series. *L. carnosum* WC326 and WC327, that resulted to be the same strain, were isolated from the same factory, in fresh sausages in diverse lots sampled two months apart (Raimondi et al., 2018), suggesting the persistence of this species in the environment of production plants.

L. carnosum harbored the genes for heterolactic fermentation sugars, yielding L-lactate, ethanol, acetate, and CO_2 (Posthuma et al., 2002). The availability in meat of a small range of carbohydrates, mainly glucose deriving from glycogen hydrolysis (Pereira and Vicente, 2013), is consistent with the small potential to ferment sugars, restricted to glucose and few others. With a sole exception, carbohydrate metabolic capabilities were shared by all the strains and were not strain-dependent. The genes for ribose assimilation and fermentation were present only in *L. carnosum* WC0322, encoded by plasmid pLQ22.

Unlike other lactic acid bacteria, including *Leuconostoc* species, *L. carnosum* seemed unable to uptake and metabolize citrate via citrate lyase and oxaloacetate decarboxylase

(Bekal et al. 1998; Garcia Quintans et al. 2008). However, all the strains may yield flavored four carbon metabolites such as acetoin and butanediol from pyruvate, even though the role of *L. carnosum* in food ripening and aroma development remains to be clarified.

Meat is a rich substrate mainly composed of proteins, albeit an abundant fraction of amino acids or small peptides is also available (Cobos and Diaz, 2015). Despite the ability to synthetize some amino acids, all the strains seemed unable to grow utilizing ammonium as the sole nitrogen source and must rely on a pool of few pre-formed amino acids, that should include at least glutamate, aspartate, serine, and alanine. Since the biosynthetic pathway of other amino acids (e.g. tyrosine, phenylalanine, lysine) is present but incomplete, it remains to be clarified whether some genes are not correctly annotated, and prediction could make up for the interruption. Adaptation of *L. carnosum* to a nitrogen-rich environment explains the autotrophy not only for several amino acids, but also for some ribonucleotide and deoxyribonucleotide, and for several vitamins and cofactors. On the other side, *L. carnosum* clearly took advantage of the 23 peptidase genes identified in the core genome, and other 2 encoded by accessory genes.

The presence of a choline transporter indicates that *L. carnosum* may use this compound abundant in meat to cope with salt and other osmotic stress. However, while in other bacteria (e.g. *Bacillus subtilis*) choline uptake is functional to the biosynthesis of the osmoprotectant betaine (Kappes et al., 1999), any gene involved in choline transformation to betaine is missing.

All the strains herein compared harbored 1 to 4 plasmids, that presented main functions associated to: hydrolysis of proteins; transport and metabolism of amino acids, oligopeptides, and carbohydrates; production of bacteriocin and exopolysaccharide; resistance to bacteriophage, heavy metal, and other stress responses; DNA restriction-modification systems.

A peculiar trait of these replicons was the absence of genetic determinants for antibiotic resistance, that were not detected either in the chromosomes. However, the genes *mdlAB* and *lmrCD,* putatively encoding the multidrug efflux systems, were located in the genome. There is no information on whether *mdlAB* confers any antibiotic resistance to bacteria, whereas *lmrCD* confers to *Lactococcus lactis* the resistance to some toxic compounds, including daunomycin (Kobayashi et al., 2001; Lubelsky et al., 2006). The lack of genes encoding antibiotic resistance may be due to the specificity of *L. carnosum* for meat matrices, that seclude the species in an environment where the selective pressure is limited and the

interaction with gut bacteria, the most exposed to the antibiotics utilized in animal production, is low.

The safety of *L. carnosum* is a key point, considering the high load of bacteria belonging to this species that are ingested mostly consuming products such as MAP sliced meat. Safety of this species was associated to the absence not only of any antibiotic resistance genetic determinants, but also of decarboxylase genes for biogenic amine production, assessing the incapability of these strains to produce biogenic amine.

In most of the strains, the potential of bacteriocin synthesis, secretion, and immunity was identified, including the production of leucocin B, a class IIa bacteriocin effective against *L. monocytogenes* (Felix et al., 1994). It is quite interesting the fact that, at the end of the shelf-life, MAP cooked ham generally presented an abundant population of *L. carnosum*, regardless of the sensorial evolution of the product (Raimondi et al., 2019). Indeed, *L. carnosum* dominated both in seriously spoiled products and in those that maintained good sensorial properties. However, studies of biochemistry and physiology of *L. carnosum*, also focused on molecular mechanisms affecting meat spoilage, are still lacking. The utilization of *L. carnosum* strains as bioprotective starters could be encouraged by the ability to produce bacteriocins, thus a deeper investigation is required to determine their activity spectrum against spoilage and pathogen bacteria. Production and activity of the predicted lactococcin-like proteins need to be analyzed further.

As a whole, this study provided a deep insight into the genomic and metabolic features of this important bacterium ubiquitous in meat products, revealing that it is safe and opening new possibilities of exploitation in food conservation industry.

These results of this study are presented in two papers published in:

1. *Microbiology Resource Announcements* as "Draft genomes of 12 *Leuconostoc carnosum* isolated from cooked ham packaged in modified atmosphere and fresh sausages"

GENOME SEQUENCES

Check for updates

Draft Genome Sequences of 12 *Leuconostoc carnosum* Strains Isolated from Cooked Ham Packaged in a Modified Atmosphere and from Fresh Sausages

Francesco Candeliere,[a] Stefano Raimondi,[a] Gloria Spampinato,[a] Moon Yue Feng Tay,[c,d] Alberto Amaretti,[a,b] Joergen Schlundt,[c,d] Maddalena Rossi[a,b]

[a]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy
[b]BIOGEST—SITEIA, University of Modena and Reggio Emilia, Modena, Italy
[c]Nanyang Technological University Food Technology Centre (NAFTEC), Singapore
[d]School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore

2. *Frontiers in Microbiology* as "Comparative genomics of *Leuconostoc carnosum*"

ORIGINAL RESEARCH ARTICLE
Front. Microbiol. | doi: 10.3389/fmicb.2020.605127

Comparative genomics of Leuconostoc carnosum

Provisionally accepted The final, formatted version of the article will be published soon. Notify me

Francesco Candeliere[1], Stefano Raimondi[1], Gloria Spampinato[1], Moon Y. Tay[2, 3], Alberto Amaretti[1, 4], Jorgen Schlundt[2, 3] and Maddalena Rossi[1, 4*]

[1]University of Modena and Reggio Emilia, Italy
[2]Nanyang Technological University, Singapore
[3]School of Chemical and Biomedical Engineering, College of Engineering, Nanyang Technological University, Singapore
[4]Biogest Siteia, University of Modena and Reggio Emilia, Italy

4. THE HUMAN GUT MICROBIOME

The gastrointestinal tract (GIT) is a complex and specialized ecosystem, densely populated by trillions of organisms (10^{14}), mainly anaerobic bacteria but also viruses and eukaryotic organisms.

The number of bacteria harboured in the GIT is 10-folds higher than the number of host cells and the collective genome of the microbiota, is 100-fold greater than that of the human genome (Bäckhed et al., 2005; Huttenhower et al., 2012; Ley et al., 2011; Qin et al., 2010). The instauration of a mutually beneficial relationship between host and gut bacteria results in essential contributions to human health, e.g., by producing beneficial metabolites and vitamins (Salonen and de Vos, 2014). Moreover, the microbiota modulates the maturation of the immune system, regulates innate and adaptive immune responses, and acts as a barrier to prevent colonization of the GIT by pathogens (Maynard et al., 2012).

4.1 COLONIZATION OF THE GUT

The microbiota composition depends on early life events, diet and pharmaceutical treatments that affect both stability and function of the gut ecosystem. The prenatal GIT of the fetus is virtually sterile and the microbial colonization of the newborn is reported to start during and after birth (Fanaro et al., 2003, Dominguez-Bello et al., 2010). However, recent studies reported that the placenta contains a low charge of non-pathogenic bacteria belonging to *Firmicutes, Tenericutes, Proteobacteria, Bacteroidetes,* and *Fusobacteria* (Aagaard et al., 2014). Placental bacteria contribute to differentiate the early colonization of preterm infants in comparison to term infants (Doyle et al., 2014). Moreover, during the delivery, the neonate comes in contact with the microorganisms of the vagina, the faeces and the skin of the mother which colonize the oral cavity and the stomach of the infant. For this reason, the first colonization it is considered to take place via the placenta and the second one during birth

. After few hours from birth microbial colonization of the GIT is still limited and the first colonizers are *Escherichia coli*, streptococci, and staphylococci because of the positive

potential redox of the area and the presence of oxygen. Afterwards, these bacteria consume the oxygen, leading to a decreased redox potential and enabling the colonization by strictly anaerobic bacteria (Orrhage and Nord, 1999). In the breastfed infants, the population of lactobacilli, bifidobacteria and enterococci consolidate in the first 4 days of life and after 10 days the *Bacteroidetes* appear and increase their concentration rapidly. The most represented genera during the breastfeeding phase are, in decreasing order, *Bifidobacterium* (10^{10}-10^{11} cells/g), *Bacteroidetes* and *Clostridium* (10^{8}-10^{9} cells/g). With the weaning, bifidobacteria stop to be predominant and after 2 years of life the microbiota of the child become similar to the adult one.

4.2 THE COLONIC MICROBIOTA OF ADULTS

The microbial diversity and composition of the microbiome vary greatly long the GIT, that is 7 meters long from the oral cavity up to the anus and is divided into diverse anatomic regions harboring different concentrations of bacteria (Fig. 4.1). The bacterial load increases along the intestinal tract as the pH switches from acidic towards neutrality. In particular, the colon hosts most of the bacteria, especially anaerobes, reaching loads between 10^{12} to 10^{13} bacteria per gram of feces, due to the high availability of nutrients, the low redox potential, and the absence of oxygen. Species belonging to the phyla *Bacteroidetes* and *Firmicutes* are the most represented, followed by *Actinobacteria*, which is mainly comprised of the genus *Bifidobacterium*, *Proteobacteria*, and *Verrucomicrobia* (Li et al., 2012; Qin et al., 2010). The microorganisms of the intestine thrive utilizing undigested carbon and nitrogen sources reaching the colon. These nutrients are represented by those carbohydrates and proteins, introduced with diet, that escape the hydrolysis by the host digestive enzymes and absorption through gut epithelia. The microbiota metabolism yields specific products that cause different metabolic responses in the host (David et al., 2014). For this reason, a well-balanced microbiota, characterized by a high diversity contributes to maintain the homeostasis of the entire organism.

A healthy intestinal microbiota is characterized by temporal stability, diversity, and conserved key functions (Martinez et al., 2013; Flores et al., 2014). However, despite the high individuality in composition, it is plausible that there are similarities among different healthy subjects. Qin et al in 2010, demonstrated that a “core microbiome” is composed of approximately 160 bacterial species, shared between healthy individuals (Claesson et al.,

2010; Martinez et al., 2013; Qin et al., 2010; Rajilic-Stojanovic et al., 2009; Salonen et al., 2014; Sekelja et al., 2011; Tap et al., 2009; Turnbaugh et al., 2009; Turnbaugh et al., 2007; Willing et al., 2010). The core microbiome is composed by approximately 30% of the detected phylotypes consisting mainly of *Firmicutes (Ruminococcus, Eubacterium and Faecalibacterium spp.), Bacteroides spp.* and *Actinobacteria (Bifidobacterium ssp.)* (Qin et al., 2010; Salonen et al., 2014).

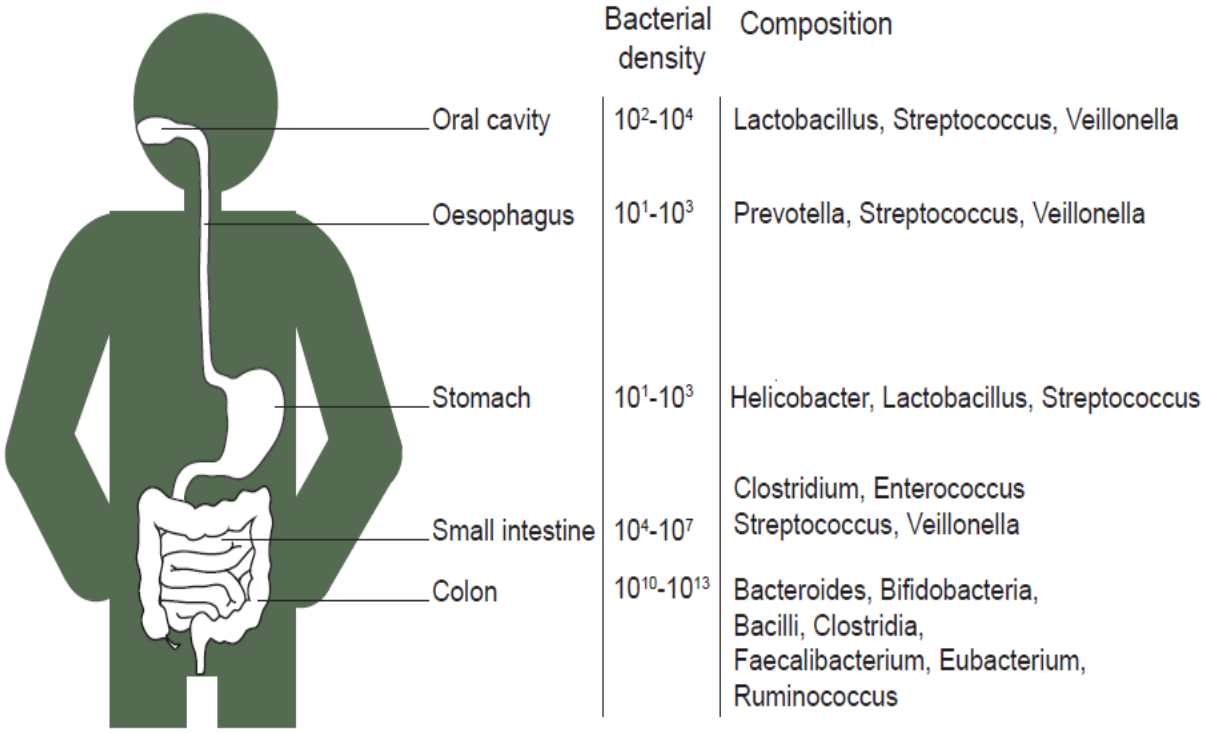


Figure 4.1 Characteristics of the adult intestinal tract and microbiota (Jalanka J, PhD thesis, 2014)

Perturbations and changes of the intestinal microbiota may alter the healthy status of the host as they can correlate with common disorders such as inflammatory bowel disease, obesity, diabetes, and atopic disease (Dethlefsen et al., 2008; Jernberg et al., 2007; Vrieze et al., 2013).

For this reason, the interest in deeply studying this environment is increasing day by day and the techniques that allow reaching this goal are implementing every day. To date, culture dependent methods allowed to characterize over 1000 different bacterial species of the human GIT (Rajilic-Stojanovic and de Vos, 2014), while the majority of the intestinal bacteria are still considered “uncultivable” and have been characterized only by their 16S rRNA sequence. Over 3000 more microorganisms and 9.8 million bacterial genes have been identified from the intestines of healthy adults with the aid of the NGS technologies, increasing our understanding of the bacterial diversity and its relation with health and disease (Li et al., 2012).

4.3 THE METABOLISM OF THE GUT MICROBIOTA

The bacteria residing in the intestinal tract have to utilize non-digestible dietary substrates and endogenous mucus as energy sources (Scott et al., 2013). The dietary carbohydrates and proteins represent the main substrates for the bacterial fermentative processes in the large intestine. Many studies described the fermentation of polysaccharides or dietary fibers by the colon microbiota yielding short chain fatty acids (SCFA). However, also a high amount of proteins and peptides, approximately 13 grams, enter the human large intestine every day, where they represent nitrogen sources for bacteria. Apart from diet, proteins in the large intestine have also endogenous origins including host mucus secretions and the constant turnover of colonic epithelial and bacterial cells (Scott et al., 2013).

5. Project 2

Metagenomic investigation of β-glucuronidases (GUS) in the human gut microbiome

5.1 INTRODUCTION

5.1.1 Beta-glucuronidases

Humans and their colon microbiota have evolved together, establishing close symbiotic interrelationship, fruitful for both. The gut microbiota as a whole is implicated in a number of essential biological processes such as resistance to colonization (Mandoli et al., 2010), immune system modulation (Sharon et al., 2014), synthesis of essential vitamins and nutrients, and digestion of polysaccharides (Koh et al., 2016). The gut microbiota can process foreign compounds (xenobiotics), including dietary components, environmental pollutants, and pharmaceuticals, changing their chemical structure, and thus modifying their lifetimes, bioavailability, and biological effects (Koppel et al., 2016).

The human gut microbiota encodes a broad diversity of enzymes, many of which are exclusively microbial, adding significant chemical diversity to the metabolic reactions carried out by the human body (Koppel et al., 2016). The study of the influence of gut microbiota on human health has gained prominence in recent years, focusing on the mechanisms by which microbial metabolic activity can affect human biology, on how these

metabolisms vary between individuals and on the exploitation of these chemical transformation to improve health (Koppel et al., 2016). The main challenge in this sense is to connect microbial chemistry to specific organisms, genes, and enzymes (Sharon et al., 2014).

The deglucuronidation of glucuronides is part of this metabolic symbiosis between the microbiome and the human host. β-glucuronidase (GUS) enzymes expressed by the gut microbes mediate the reactivation of important endogenous molecules such as hormone glucuronides and neurotransmitters (Lombardi et al., 1978; Winter and Bokkenheuser, 1987; Asano et al., 2012). Once released by hydrolysis, glucuronic acid became a bacterial substrate that enters the Entner-Doudoroff pathway, along which sugar acids are catabolized to pyruvate and shunted into the TCA cycle (Peekhaus and Conway, 1998). Mammals also express a GUS enzyme ortholog that is localized to lysosomes in tissues like liver and intestines and plays an essential role in the degradation of endogenous glycosaminoglycans (Jain et al., 1996).

The bacterial deglucuronidation activity allows the regeneration of the original molecule that was eliminated by the host and facilitates reuptake by the gut epithelia and recirculation in the bloodstream (Wilson et al., 1992). This process, composed of glucuronidation in the liver, delivery to the GI lumen via the bile duct, reactivation and absorption via the intestinal epithelia, and transport back to the liver, is called enterohepatic circulation (Roberts et al., 2002) (Fig. 5.1), and it can affect the pharmacokinetics of many drugs and also regulates the levels of endogenous compounds (Winter and Bokkenheuser, 1987; Roberts et al., 2002; Vítek and Carey, 2003).

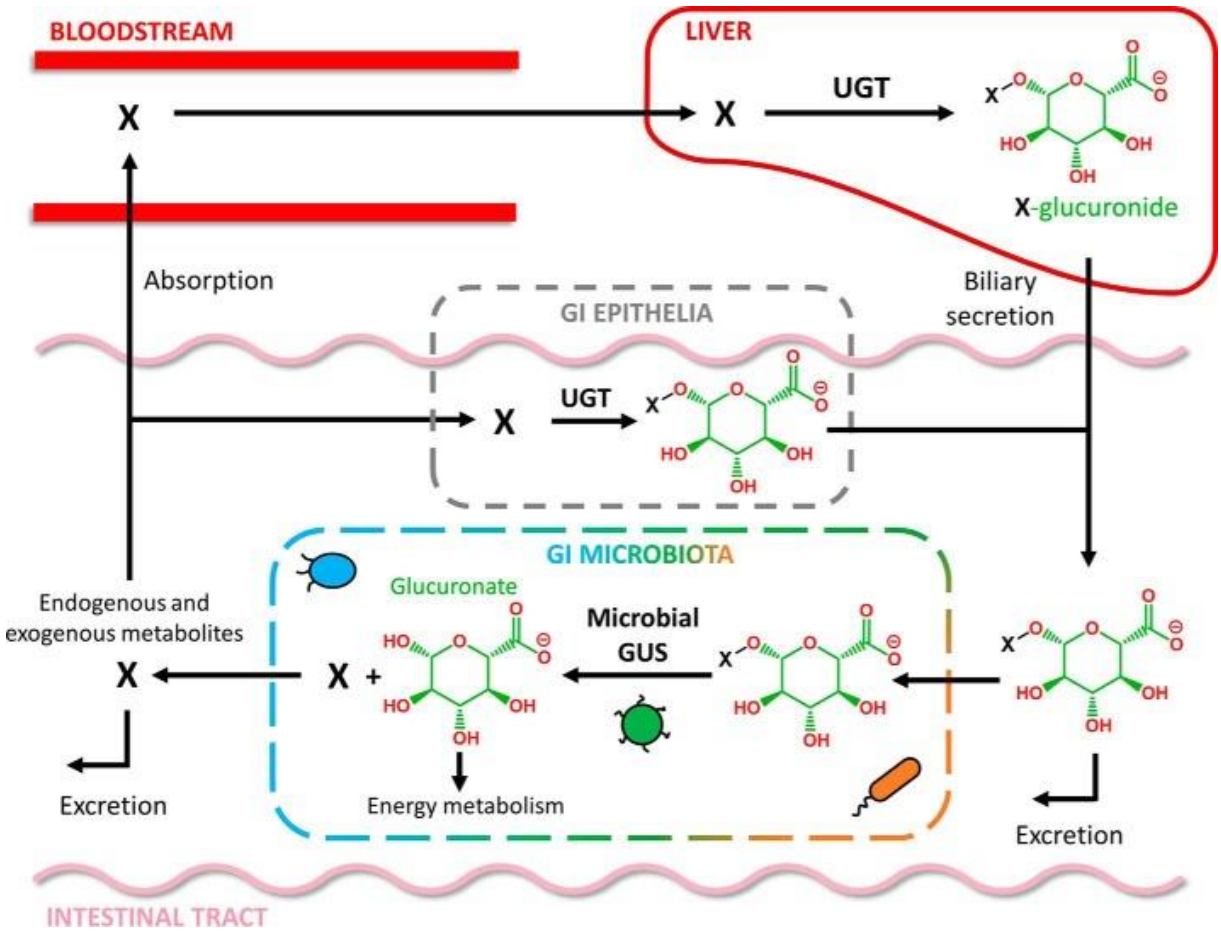


Figure 5.1 Graphical representation of the Enterohepatic circulation of chemically distinct molecules (X), from Pellock and Redinbo, (2017)

5.1.2 Glucuronidation of endogenous substrates

Endogenous glucuronides were clearly the driving force for the symbiotic evolution of host-associated bacterial GUS enzymes. The glucuronidation reaction is catalyzed by UDP-glucuronosyltransferase (UGT), which append glucuronic acid, derived from UDP-glucuronate, to hydroxyl, carboxylate, and other nucleophilic functional groups of aglycons (Liston et al., 2001; Meech et al., 2012).

One of the most heavily glucuronidated endogenous molecules is bilirubin, a breakdown product of heme (Fevery et al., 1983). Bilirubin glucuronides are generated in the liver and enter the GI tract from the bile duct. In the GI tract, stercobilin, which gives feces its brown colour, and urobilin, which is responsible for the yellow colour of urine and the

yellow complexion of jaundiced subjects (Vítek, and Ostrow, 2009), are the products of the metabolism of bilirubin glucuronides carried out by the intestinal microbiota. The deconjugated bilirubin that manages to escape further metabolism by bacteria is reabsorbed through the GI epithelia and undergoes enterohepatic circulation (Vítek and Carey, 2003).

Hormones are also subject to glucuronidation. For example, the estrogen metabolites estradiol, estrone, and estriol are glucuronidated by multiple UGT isoforms (Raftogianis et al., 2000). The androgen testosterone and the thyroid hormone thyroxine can also be glucuronidated (Kreek et al., 1963; Yamanaka et al., 2007) and their glucuronides can be hydrolysed by bacterial GUS enzymes (Graed et al., 1977; Hazenberg et al., 1988; DiStefano et al., 1993).

The neurotransmitters dopamine, norepinephrine, and serotonin are other endogenous molecules that are glucuronidated in the body and metabolized by bacterial GUS. Around 50% of all dopamine is generated in the GI tract (Eisenhofer et al., 1997), where it acts as a regulator of gut motility and water absorption (Flemström et al., 1993; Haskel and Hanani, 1994). Recently, it has been shown that microbes have a significant role in the processing of dopamine glucuronide in the GI lumen of mice (Asano et al., 2012).

Bile acids are important compounds to gut health and are significantly processed by the microbiota. Much like bilirubin, bile acids are heavily metabolized by the microbiota. In the liver, bile acids are conjugated to sulfate, taurine, and glycine moieties, all of which can be removed by GI microbial sulfatases and bile salt hydrolases, and they can also be conjugated to glucuronic acid (Matern et al., 1984). Glucuronidated bile acids likely provide a significant energy source to bacteria capable of processing them.

In addition, endogenous polysaccharides also represent a critical source of glucuronides in the gut. Chondroitin sulfate and hyaluronic acid are glucuronic acid-containing polysaccharides that can be found in the GI tract (Dutton, 2012). Bacteria express many endo- and exo-glycosidases that work in concert to break down complex polysaccharides. With a mechanism analogous to human GUS, which catabolizes extracellular matrix polysaccharides in lysosomes, bacterial GI GUS enzymes play similar roles with substrates like chondroitin sulfate that enter the GI from host cells sloughed from the epithelia (Salyers and O'Brien, 1980.

5.1.3 Xenobiotic metabolism

Glucuronides of drugs and xenobiotic molecules have been a primary focus of research because of their potential importance to therapeutic efficacy and tolerance.

Xenobiotic metabolism in the human host is regulated mainly by cytochrome P450s which carry out Phase I metabolism, whereas Phase II metabolism generates excretable hydrophilic metabolites, with glucuronidation being the major detoxification pathway. Phase III transporters remove xenobiotics and metabolites from cells, excreting the more polar compounds in the gut. Here glucuronides encounter the intestinal microbiota. The majority of human microbiota-xenobiotic interactions occur thus within the GI tract, which contains distinct regions with different epithelial cell physiology, pH, oxygen levels, and nutrient content, thus providing distinct habitats for microorganisms and a vast range of possible metabolic processes (Sender et al., 2016; Aron-wisnewsky et al., 2017; Noecker et al., 2017). Unsurprisingly then the gut microbiota plays a major role in drug metabolism (Koppel et al., 2016; Choi et al., 2018). Crucially, glucuronidated molecules can be hydrolyzed by β-glucuronidase enzymes of resident bacteria (McIntosh et al., 2012; Dabek et al., 2008); GUS can remove the glucuronic acid from the metabolite, releasing the original molecules into the gut lumen (Spanogiannopoulos et al., 2016). These enzymes can then affect the fate of drugs and xenobiotics increasing their lifespan (Carmody et al., 2014), and reactivation of the released xenobiotics can result in toxicity (diarrhea or epithelial injury) in the gut (Dashnyam et al., 2018).

5.1.4 Deglucuronidation side effects

GUS-related toxic side effects are well known for cancer drug irinotecan (Nagar et al., 2006) and Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) (LoGuidice et al., 2012; Boelsterli et al., 2013). Irinotecan is a carbamate-linked dipiperidino prodrug with improved bioavailability and solubility of SN-38, a cancer drug whose mode of action is inhibition of topoisomerase activity (Pizzolato et al., 2003). SN-38 is glucuronidated in the liver to an inactive conjugate (SN-38G), which enters the gut via biliary excretion. GUS enzymes can hydrolyze SN-38G in the large intestine to regenerate the active chemotherapeutic agent. SN-38 then enters colonic epithelial cells, causing intestinal damage and severe diarrhea; these side effects limit the clinical efficacy of irinotecan (Wallace et al., 2010).

NSAIDs can also cause gastrointestinal toxicity in susceptible patients, with consequences such as jejunal/ileal mucosal ulceration, bleeding, and perforation; NSAID-

associated enteropathy is linked to high morbidity and mortality rates and is a significant clinical challenge (Scarpignato et al., 2010). The mechanism for the toxicity of NSAIDs such as diclofenac, indomethacin, or naproxen is based on the initial hepatobiliary excretion and delivery of glucuronidated or others oxidative metabolite conjugates of NSAIDs to the gut, where bacterial GUS hydrolyze them to the corresponding aglycones. These aglycones are then reabsorbed by enterocytes and further metabolized by intestinal cytochrome P450s to potentially reactive intermediates. These oxidative metabolites induce severe endoplasmic reticulum stress or mitochondrial stress and lead to cell death, causing a subsequent inflammatory response that follows the initial injury (Boelsterli et al., 2013). Interestingly, both NSAID (LoGuidice et al., 2012; Saitta et al., 2014) and irinotecan (Wallace et al., 2010) toxicity in the gut can be alleviated by selective inhibition of bacterial GUS with a small-molecule inhibitor (Wallace et al., 2015). Specific inhibition of bacterial GUS is then an intriguing approach for preventing drug reactivation and the relative side effects; as these enzymes are broadly distributed in the microbiome and also are present in mammals, these inhibitors need to be selective for bacterial GUS and nontoxic to both host cells and other gut microbes (Wallace et al., 2015). Inhibitors with these characteristics were identified by in vitro high-throughput screening and show activity against distantly related gut bacteria but did not target the human b-glucuronidase (Wallace et al., 2010). Structural studies have shown that the selectivity of these bacterial GUS inhibitors is due to a specific interaction with a distinct bacterial active-site loop, and specifically with a conserved Asn-Lys motif that interacts with the carboxylic acid of the glucuronic acid sugar but is absent in glycosidases that hydrolyse different substrates (Wallace et al., 2015).

5.1.5 The GUSome

Beta-glucuronidases (GUS) are the key enzyme that remove glucuronic acid residue from conjugated compounds. These proteins were identified first in 1934 (Masamune, 1934; Oshima, 1934) in *Escherichia coli* and some Enterobacteriaceae, but later GUS activity was detected in bacterial species belonging to the most abundant phyla found in gut microbiota: Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria (Gloux et al., 2011; Kim et al., 2000; McBain and Macfarlane, 1998; Nakamura et al., 2002; Russell and Klaenhammer, 2001). Next Generation Sequencing also made possible to identify different GUS types in diverse gut bacteria, providing in the advancement in the understanding of orthologues present in the human gut (Gloux et al., 2011).

Pollet et al. (2017) realized an atlas of the beta-glucuronidases present in the human gut microbiota. Using metagenomes from the "Human Microbiome Project" (HMP) (Turnbaugh et al., 2007) all the bacterial GUS potentially produced in the colon were categorized, assembling the so-called GUSome. Two datasets have been used, HMGI (Human Microbiome Gene Indices), containing 26.5 million translated protein sequences, and HMGC (HM Clustered Gene), harboring 4.8 million protein sequences. The sequences were clustered together based on an identity percentage greater than 95% and, following a multi-step protocol, they were analyzed to identify the GUS-encoding sequences using known GUS as reference. The first step screened the sequences with at least 25% identity to the references. Then the identification was based on the conserved residues in the active sites, due to their importance in substrate recognition and reaction catalysis. At the end of the procedure, two databases were produced: HMGI3013, containing 3013 GUS representing the 0.011% of HMGI dataset, and HMC279, encompassing 279 non-redundant GUS sequences.

The GUS enzymes identified in HMC279 have been classified into 7 categories based on two structural elements, Loop 1 and Loop 2, adjacent to the active site: Loop 1 (L1), Mini-Loop 1 (mL1), Loop 2 (L2), Mini-Loop 2 (mL2), Mini-Loop 1,2 (mL1,2), No Loop (NL) e No Coverage (NC). Loop 1 and Loop 2 present variable length and aminoacids composition, in contrast to the conserved residue in the catalytic site. The class NC encompass the GUS lacking information about one or both loops, and they represent a small fraction of the database (2.2% in HMGC279).

As a whole, human gut GUSome encompasses enzymes with a conserved folding and two different loops, that probably allow to recognize different substrates. Most of GUS sequences included in the HMGC279 database belong to NL class (57.3%), and the others to classes mL1 (15%), L2 (14%), L1 (5.7%), mL2 (4%), NC (2.2%), and mL1,2 (2%).

The taxonomic analysis of the database allowed to identify the phylum of 93.5% of sequences, distributed among Bacteroidetes (52%), Firmicutes (43%), Verrucomicrobia (1.5%), and Proteobacteria (0.5%).

Differences in the cellular localization related to the class and the phylum have been also identified. For example, L1 enzymes lack signal peptide, thus they are intracellular, whereas L2, mL2 and mL1L2 are extracellular. For GUS belonging to classes mL1 and NL the presence of signal peptide is linked to their phylum: in Firmicutes it is absent, while in Bacteroidetes is present.

The analysis also supported the hypothesis that a different biocatalytic activity is correlated to function and GUS structure. In particular, the loops dimensions are pivotal to

substrate recognition. Classes L1, mL1 and L2 are capable to catalyze more efficiently the deglucuronidation of substrate in comparison to classes mL2, mL1,2 and NL.

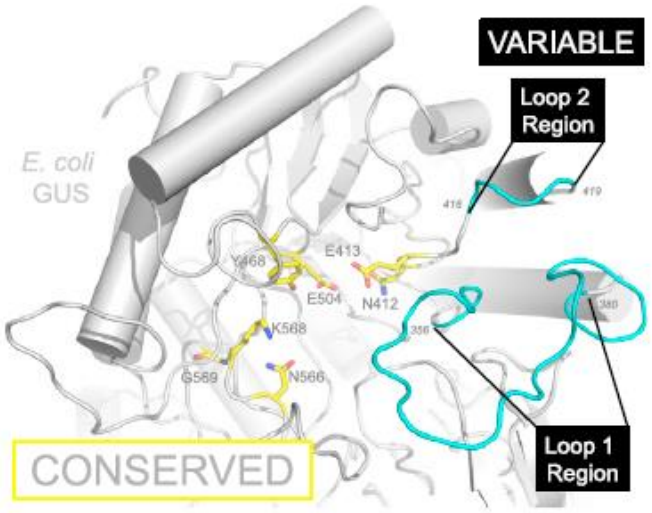


Figure 5.2 Structure of an E.coli GUS. Conserved active site is highlighted in yellow, while the variable Loop1 and Loop2 in cyan. From Pollet et al., 2017.

5.2 AIM OF THE PROJECT

Considering the results of Pollet et al. (2017), it is evident that microbial composition of intestinal microbiota could condition the metabolism of drugs and xenobiotics and could be responsible of different individual responses. The aim of this work was to study the qualitative and quantitative profile of the GUS encoded by the gut microbiota of 60 healthy subjects, in order to identify the extent of differences among subjects in the GUS profile and to assess how these differences can be related to the microbial composition.

We analyzed publicly available metagenomes from five geographically different cohort (China, Ethiopia, Spain, Sweden, United States of America), to circumvent the bias arising from diverse nutritional and life-style habits. We applied two parallel bioinformatic pipeline, one for taxonomic analysis and one for GUS functional analysis.

5.3 MATERIALS AND METHODS

5.2.1 Metagenomic samples

60 publicly available metagenomes of gut microbiota from healthy humans were collected from NCBI Sequence Read Archive (SRA). The selected samples were sequenced through whole-genome shotgun sequencing on Illumina paired-end platforms and produced reads ranging between 100 and 150 bp in length. The samples can be retrieved at the following BioProject codes: PRJNA557323, PRJNA504891, PRJEB33098, PRJEB27308, PRJEB7369, and correspond to five different cohorts from China (CHN), Ethiopia (ETH), Spain (ESP), United States of America (USA), and Sweden (SWE).

5.3.2 Assembly and binning

The raw sequences were downloaded from NCBI SRA. The FASTQ files were checked for quality and primers presence with FastQC v0.11.8 (Andrews, 2010). Spanish cohort samples required primer removal, that was carried out through Trimmomatic (Bolger et al., 2014) with ILLUMINACLIP setting. The reads were assembled in contigs using metaSPAdes v 3.9 (Nurk et al., 2017) with default parameters. The contigs were binned with MaxBin2 v2.2.7 (Wu et al., 2016) to obtain metagenome-assembled genomes (MAGs), that were taxonomically identified with CAT/BAT tool (von Meijenfeldt et al., 2019). Each bin was mapped against the raw reads using Bowtie2 (Langmead and Salzberg, 2013) to assess the relative abundance. Except for CAT/BAT that was run locally, the steps were conducted on Galaxy platform (https://usegalaxy.eu; Afgan et al., 2018).

5.3.3 Bacterial composition and beta-diversity

MetaPhLan2 (Truong et al., 2015; Segata et al., 2012), run on Galaxy platform, was used to define the abundance profile of bacterial species in each sample. The raw reads were analyzed against the MetaPhLan2 clade-specific marker genes database with default parameter, ignoring viruses and archaea profiling. The BIOM files produced by Metaphlan2 were imported into Qiime2 (Bolyen et al., 2019), used locally, to compute beta-diversity

based on Bray-Curtis dissimilarity. Beta distance matrix was utilized for Principal Coordinate Analysis (PCoA).

5.3.4 β-Glucuronidases identification and profiling

The 279 sequences of β-glucuronidases (GUS) identified and classified in Pollet et al. (2017) were searched in the metagenomes. The 279 sequences were blasted to the assembled metagenomes through tBLASTn (Altschul et al., 1990) with e-value 10^{-100}. The results were filtered to keep only hits with a high identity percentage (≥ 98.5%) and redundant hits mapping on the same position of the same contig were discarded based on the e-value.

The abundance of each GUS was correlated to the abundance of the bin containing the contig where the GUS sequence mapped to. In particular, the abundance of each GUS was calculated on the basis of the number of reads mapping on the bin containing the GUS sequence. For sequences found on unbinned contigs, we arbitrarily assigned a reads-count equals to 1. For GUS abundance graphs, the possibility that more than one GUS sequences could be present in the same bin was considered, and the abundance was adjusted accordingly to the number of total binned reads. Abundance table based on the GUS type was produced and imported into Qiime2 to calculate beta-diversity using Jaccard similarity. Beta distance matrix was used for Principal Coordinate Analysis (PCoA). BLAST and Qiime2 ran locally.

5.3.5 Statistical analysis

Statistical analysis ANOVA ($P < 0.05$) followed by Tukey's post hoc test was conducted to compare cohorts in terms of GUS profiles, abundance of bacteria harbouring GUS genes, and relative abundance of each GUS structural category.

For beta-diversity studies related to microbiome composition and GUS profile, the statistical significance among the cohort was analyzed with PERMANOVA statistical test ($P<0.05$) using Bray-Curtis and Jaccard distance matrices.

5.4 RESULTS AND DISCUSSION

5.4.1 General metagenomes features

60 metagenomes of gut microbiota of healthy subjects, sequenced with Illumina paired-end technology, were retrieved and analyzed according to the flowsheet reported in Fig 5.1 in order to search the 279 GUS genes reported by Pollet et al. (2017) and to establish whether the subjects may differ in terms of GUS profile. Attention was particularly focused on interindividual difference in the abundance of the 7 structural classes to which the GUS enzymes belong, which are expected to exert different biocatalytic properties (Pollet et al., 2017). Such interindividual difference may be responsible for a different response to glucuronide-conjugated forms of xenobiotics, such as drugs, thus impacting on absorption, activity and/or toxicity.

The samples covered diverse geographical regions, including China (CHN), Sweden (SWE), Spain (ESP), Ethiopia (ETH), United States of America (USA), in order to circumvent the bias arising from diverse nutritional and life-style habits. After quality filtering and primer trimming, the bacterial composition of the samples was obtained by two diverse approaches. MetaPhlAn2 assessed the relative abundance of microorganisms mapping reads against marker sequences of specific bacterial clades. On the other hand, metaSPAdes and MaxBin2 performed the assembly in contigs and the reference-free binning, respectively, then the bins were quantified with Bowtie2 and eventually assigned a taxonomic designation with CAT/BAT.

The analysis through metaSPAdes was a key step also of the pipeline leading to GUS profiling, since it generated the set of contigs that was mined by tBLASTn to search the 279 GUS genes (Fig. 5.1). Each GUS gene was attributed the abundance of the bin wherein the contig was encompassed.

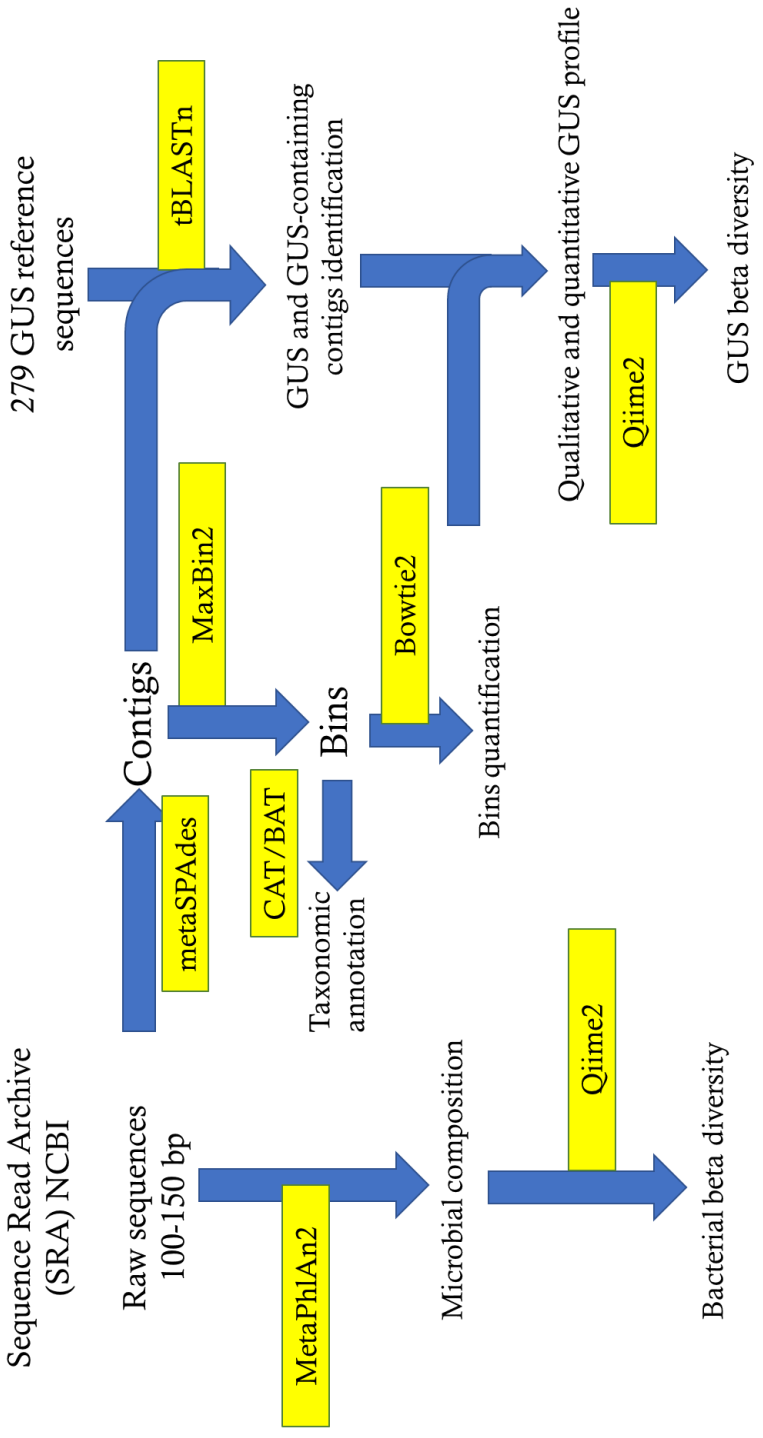


Figure 5.1 Pipeline applied for metagenomes analysis and GUS search.

On average, the metagenomes encompassed 44,214,165 reads ± s.d. 3,50E+07 (median = 2,75E+07) (Fig. 5.2A), with lengths ranging between 100 and 150 bp (Table 5.1).

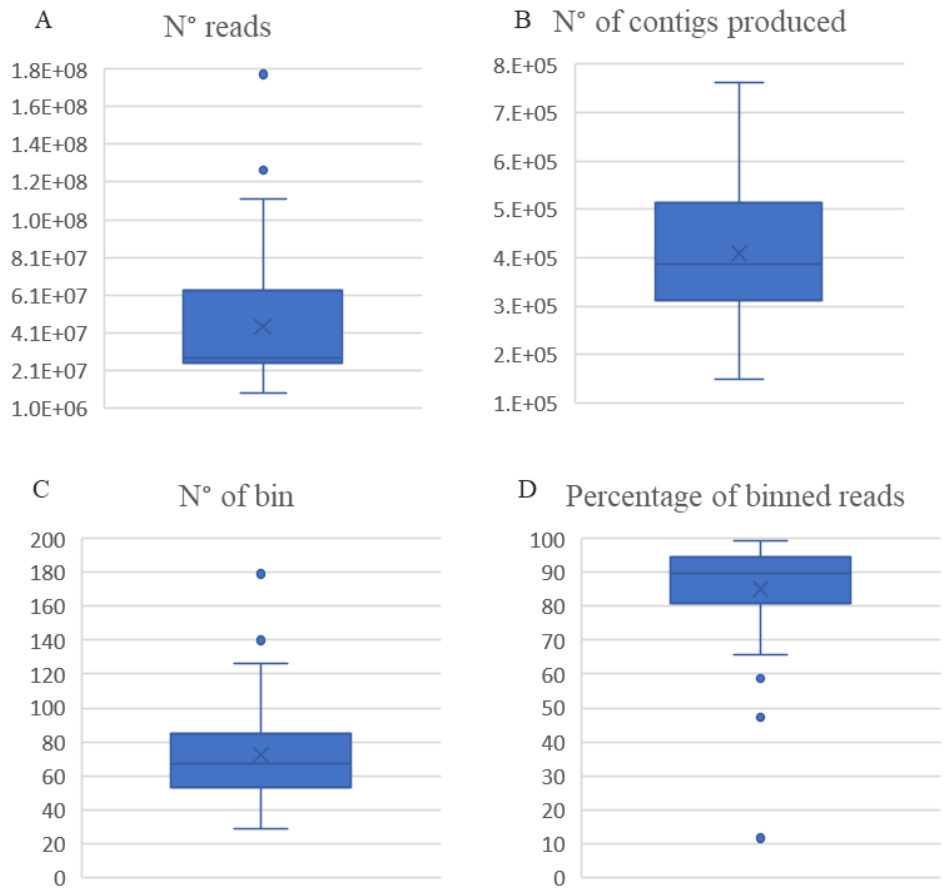


Figure 5.2 Boxplots of the samples parameters. A: number of reads in the samples of the dataset; B: number of contigs produced by metaSPAdes assembly; C: number of bins reconstructed by MaxBin2; D: percentage of reads mapped on the binned contigs with Bowtie2.

Table 5.1 Features of the 60 metagenomes used in this work

Sample	Accession numbers	Reads length (bp)	N° reads
CHN-02	SRR10680552	150	2.55E+07
CHN-04	SRR10680551	150	2.59E+07
CHN-05	SRR10680550	150	2.47E+07
CHN-06	SRR10680549	150	2.74E+07
CHN-07	SRR10680548	150	2.66E+07
CHN-08	SRR10680547	150	3.03E+07
CHN-09	SRR10680545	150	2.90E+07
CHN-31	SRR10680544	150	2.51E+07
CHN-32	SRR10680543	150	2.93E+07
CHN-33	SRR10680542	150	2.76E+07
CHN-34	SRR10680541	150	2.73E+07
CHN-35	SRR10680540	150	2.49E+07
CHN-36	SRR10680445	150	2.50E+07
CHN-37	SRR10680443	150	2.73E+07
CHN-38	SRR10680442	150	2.73E+07
CHN-39	SRR10680441	150	2.49E+07
CHN-40	SRR10680439	150	2.50E+07
ESP-42	ERR3452699	150	2.31E+07
ESP-43	ERR3452574	150	3.12E+07
ESP-44	ERR3452529	150	3.37E+07
ESP-45	ERR3452318	150	3.14E+07
ESP-46	ERR3451635	150	3.60E+07
ESP-47	ERR3450606	150	3.45E+07
ESP-48	ERR3450296	150	2.54E+07
ESP-49	ERR3450229	150	4.47E+07
ESP-50	ERR3450203	150	2.72E+07
ETH-03	SRR8784372	100	3.43E+07
ETH-10	SRR8784387	100	1.94E+07
ETH-11	SRR8784374	100	2.35E+07

ETH-12	SRR8784376	100	2.99E+07
ETH-13	SRR8784385	100	3.02E+07
ETH-14	SRR8784383	100	4.57E+07
ETH-15	SRR8784379	100	2.50E+07
ETH-41	SRR8784390	100	2.50E+07
ETH-51	SRR8784391	100	2.28E+07
ETH-52	SRR8784395	100	2.42E+07
ETH-53	SRR8784394	100	1.72E+07
SWE-01	ERS554193	100	1.27E+08
SWE-16	ERR636369	100	9.12E+07
SWE-17	ERR636383	100	7.10E+07
SWE-18	ERR636385	100	1.03E+08
SWE-19	ERR636375	100	1.78E+08
SWE-20	ERR636405	100	8.48E+07
SWE-21	ERR636371	100	9.41E+07
SWE-22	ERR636411	100	8.67E+07
SWE-23	ERR636351	100	5.67E+07
SWE-24	ERR636373	100	7.78E+07
SWE-25	ERR636353	100	1.12E+08
SWE-26	ERR636391	100	6.57E+07
SWE-27	ERR636389	100	9.99E+07
SWE-28	ERR636355	100	1.01E+08
SWE-29	ERR636359	100	8.42E+07
SWE-30	ERR636363	100	9.76E+07
USA-54	ERR2641799	150	1.37E+07
USA-55	ERR2641792	150	9.17E+06
USA-56	ERR2641793	150	9.40E+06
USA-57	ERR2641795	150	1.22E+07
USA-58	ERR2641798	150	1.92E+07
USA-59	ERR2641800	150	1.11E+07
USA-60	ERR2641801	150	1.04E+07

The mean number of contigs assembled by metaSPAdes was 408,745 (median 386,008) (Fig. 5.2B), whereas MaxBin2 produced 29 to 179 binned contigs per sample, with a mean value of 72 (Fig 5.2C). For most of the samples (> 75%), the reads that were associated to a bin, retrieved through mapping with Bowtie2, accounted for more than 80%, with a mean of 85.18% (median 89.68%) (Fig. 5.2D).

The taxonomic assignment of bins performed with CAT/BAT met difficulties. The high fragmentation of metagenomic assemblies likely hindered the accuracy of binning, resulting in a limited taxonomic diversity that impeded a deep analysis of GUS (Fig. 5.3A). To overcome the limited resolution of microbiota composition, a more accurate microbial profiling was obtained using MetaPhlAn2. Comparison of metagenome mapped by the two different computational tools revealed that MetaPhlAn2 identified more taxa than the binning approach (Fig. 5.3B), whereas through MaxBin2 a higher number of reads merged into the corresponding taxonomic groups (Fig. 5.3C). Thus, MetaPhlAn2 was used for metagenome profiling of bacteria and metaSPAdes associated to MaxBin2 was exploited to trace and quantify GUS types.

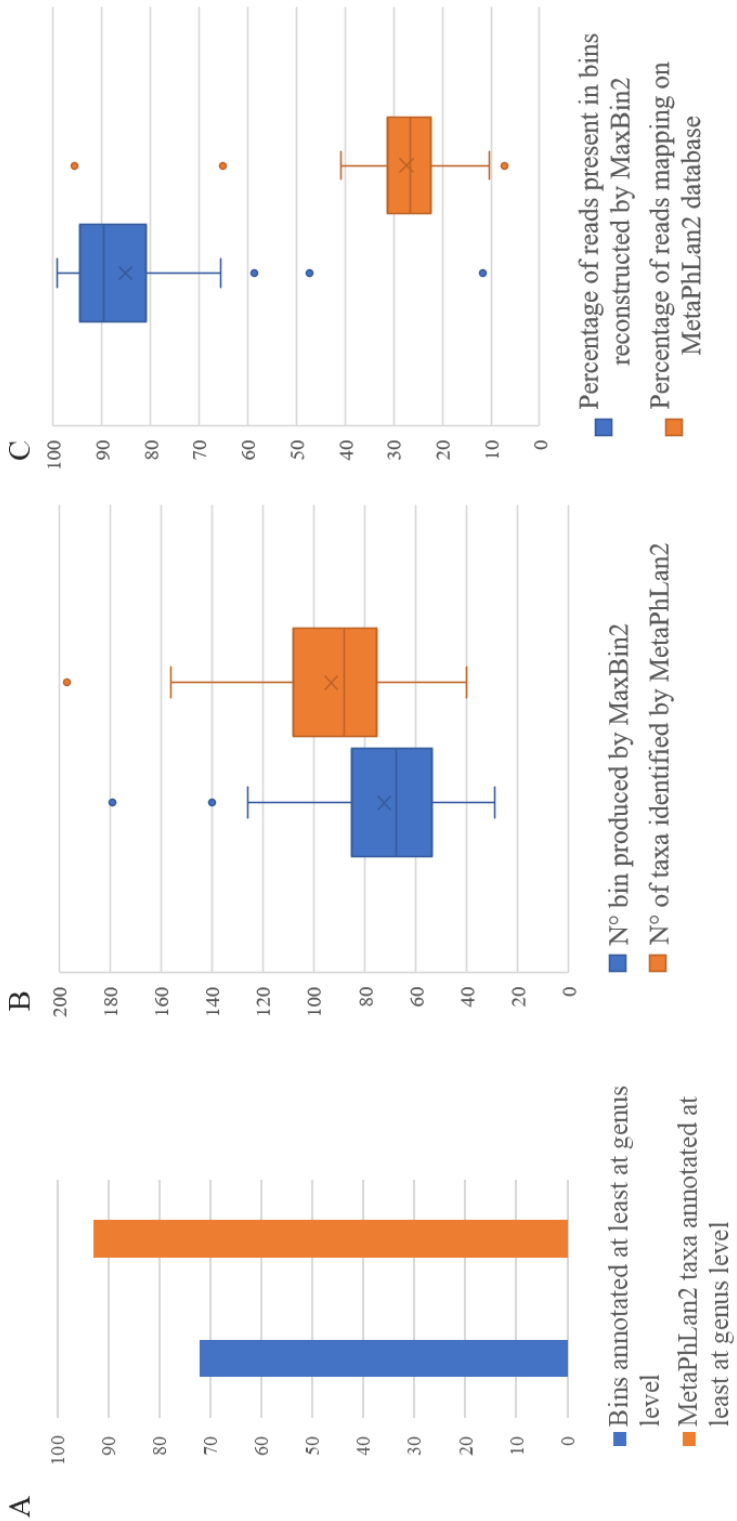


Figure 5.3 Comparison between metagenome analysis by binning approach (metaSPAdes+ MaxBin2+CAT/BAT) (blue) and MetaPhlan2 profiling (orange). A: number of unique bins (MaxBin2+CAT/BAT) or taxa (MetaPhLan2) annotated at least at genus level; B: number of bins produced by MaxBin2 and number of taxa identified by MetaPhLan2; C: percentage of reads mapped on the bins' contigs by Bowtie2 and by MetaPhLan2.

5.3.1 Microbial composition and beta-diversity

According to MetaPhlAn2 analysis (Fig. 5.4), the dominant phyla were Firmicutes and Bacteroidetes. In 4 out of 5 cohorts, Firmicutes outnumbered the other phyla, and only in Chinese metagenomes Bacteroidetes heavily prevailed. The relative amounts of Actinobacteria and Proteobacteria were quite different among samples, ranging from 0.98 to 14.67 % and from 0.35 to 9.69%, respectively. Verrucomicrobia accounted for 0.21 to 5.21 % of the whole bacterial population.

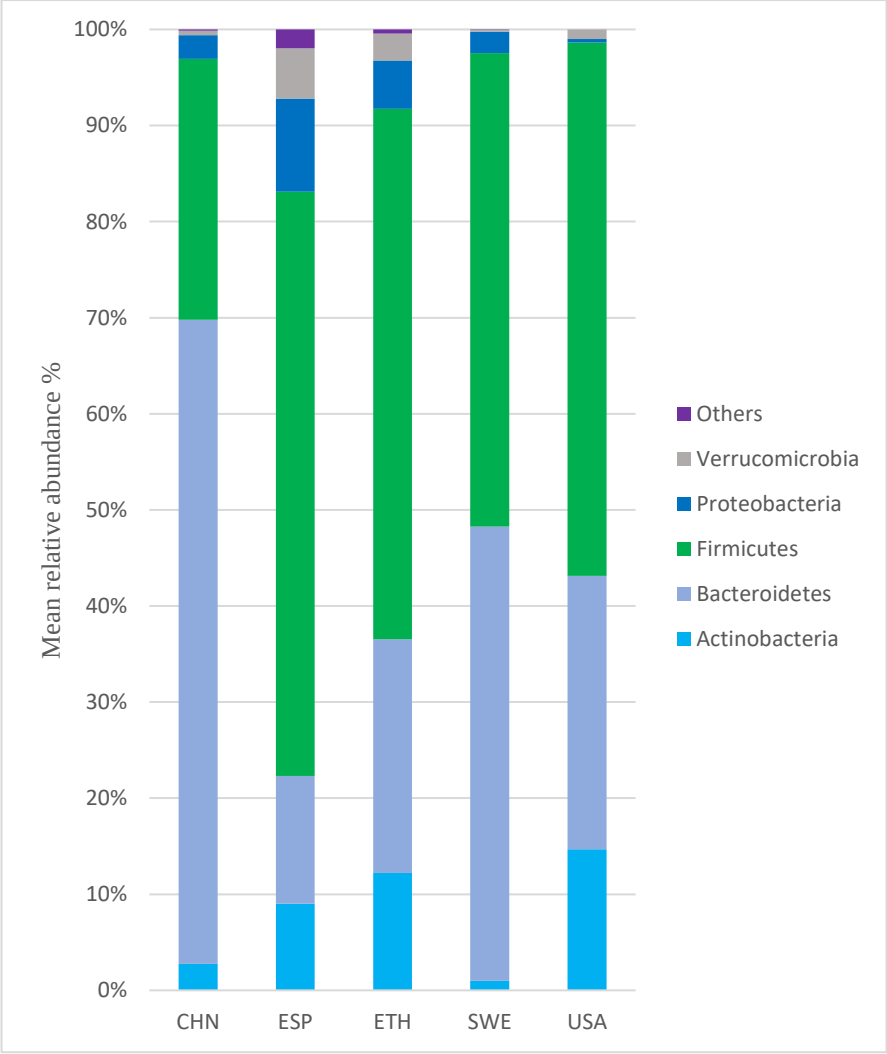


Figure 5.4 Mean relative abundances of the main phyla identified in the five cohorts profiling metagenomes by MetaPhlAn2. "Others" groups phyla lower than 1%.

The beta-diversity among samples was assessed by the Principal Coordinate Analysis (PCoA), calculated from the distance matrix of the Bray-Curtis dissimilarity index (Fig. 5.5). The autovector PCo1 mainly described the diversity in the dataset (29.89%), while PCo2 and PCo3 explained 12.06 and 10.35%, respectively. According to PERMANOVA statistical analysis, grouping in cohorts was significant ($P < 0.05$). ETH and CHN cohort were clearly separated along PCo1, with the Ethiopian samples being located at negative PCo1 values and the Chinese at the positive ones. Samples belonging to the other cohorts where dispersed over the plot, with relevant overlapping of metagenomes from diverse proveniences.

The genus *Bacteroides* mainly contributed to PCo1 positive values that characterized the CHN samples, in agreement with the prevalence of Bacteroidetes over Firmicutes in Chinese metagenomes (Fig. 5.6). Also *Alistipes* positively contributed to PCo1, whereas *Prevotella*, *Bifidobacterium*, and *Faecalibacterium* were biomarkers negatively contributing to PCo1 autovector. Along autovector PCo2, the samples were mainly diversified positively by the genera *Eubacterium*, *Subdoligranulum*, *Ruminococcus*, *Alistipes*, and negatively by *Bacteroides* and *Prevotella*.

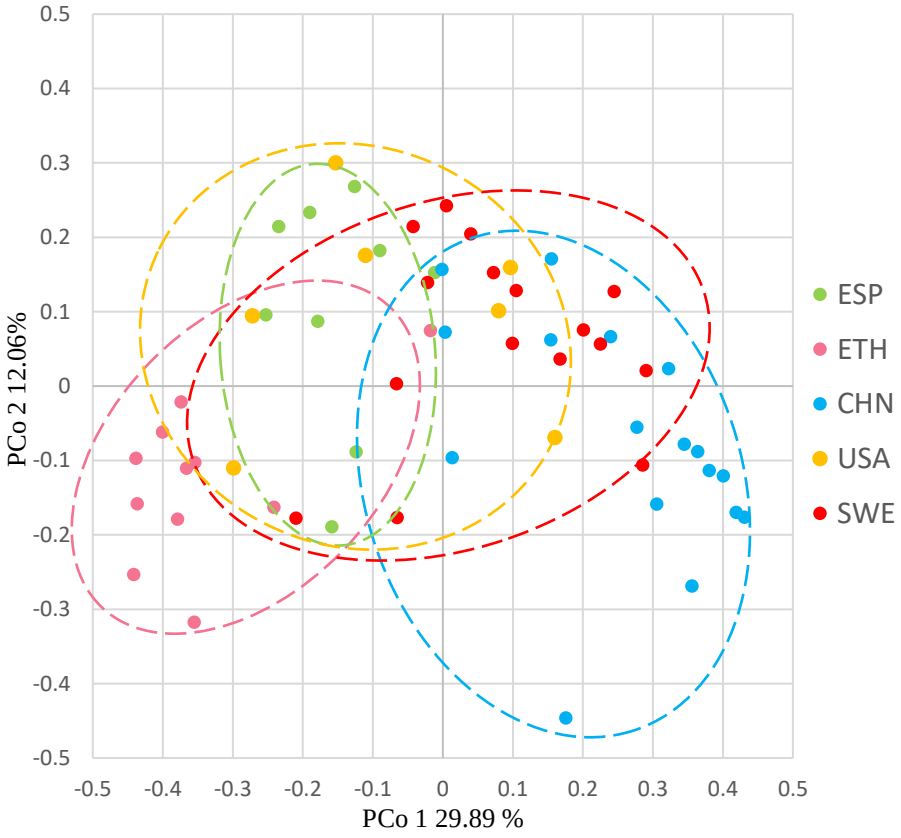


Figure 5.5 PCoA plot of beta-diversity based on Bray-Curtis dissimilarity index of the microbial composition of the 60 metagenomes

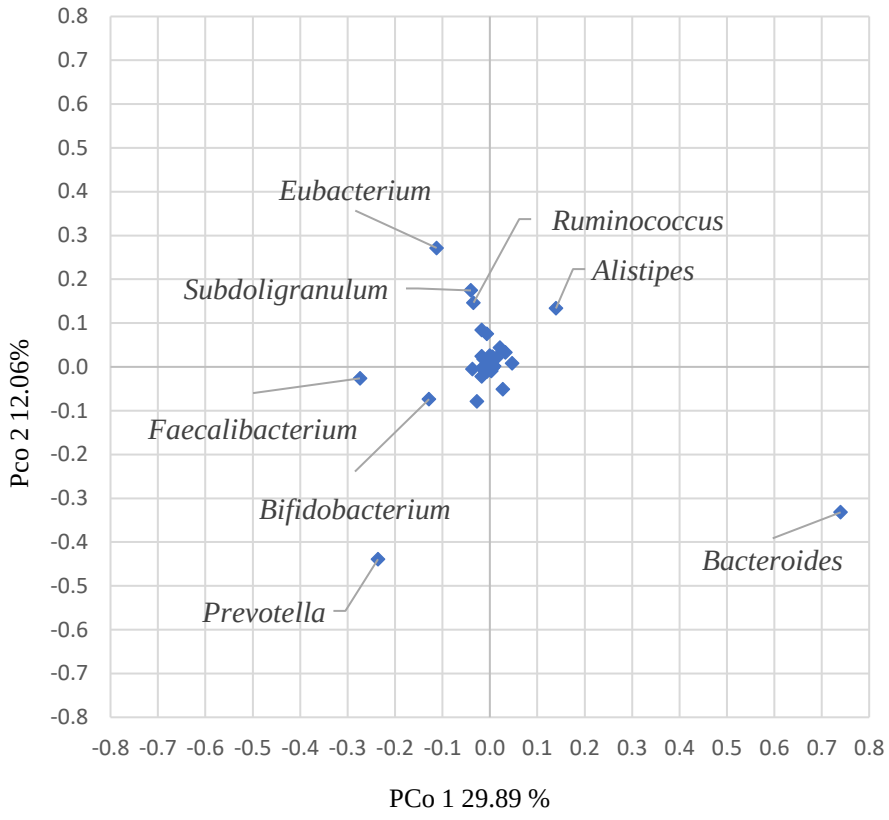


Figure 5.6 PCoA plot of the species contribution to sample differentiation in beta-diversity.

5.3.2 β-glucuronidases composition and beta-diversity

In order to detect in the metagenomes the genes encoding the different sequences types, the 279 GUS proteins were queried against the whole sets of contigs sorted by metaSPAdes assembly. 218 out of 279 different GUS were detected in the 60 samples. The samples encompassed from 4 (ETH-10) to 82 (SWE-28) contigs containing at least a GUS sequence, with a mean of 46 (Fig. 5.7). In most of the samples (58/60), at least one GUS sequence was redundant, being detected on different contigs of the same bin.

Alpha diversity of GUS types calculated by Shannon index highlighted the presence of some outliers with an extremely low index, dominated by a single or a few GUS types (Fig. 5.8A). Alpha diversity was similar among the cohorts, apart from ETH samples, where it was lower ($P < 0.05$) (Fig. 5.8B).

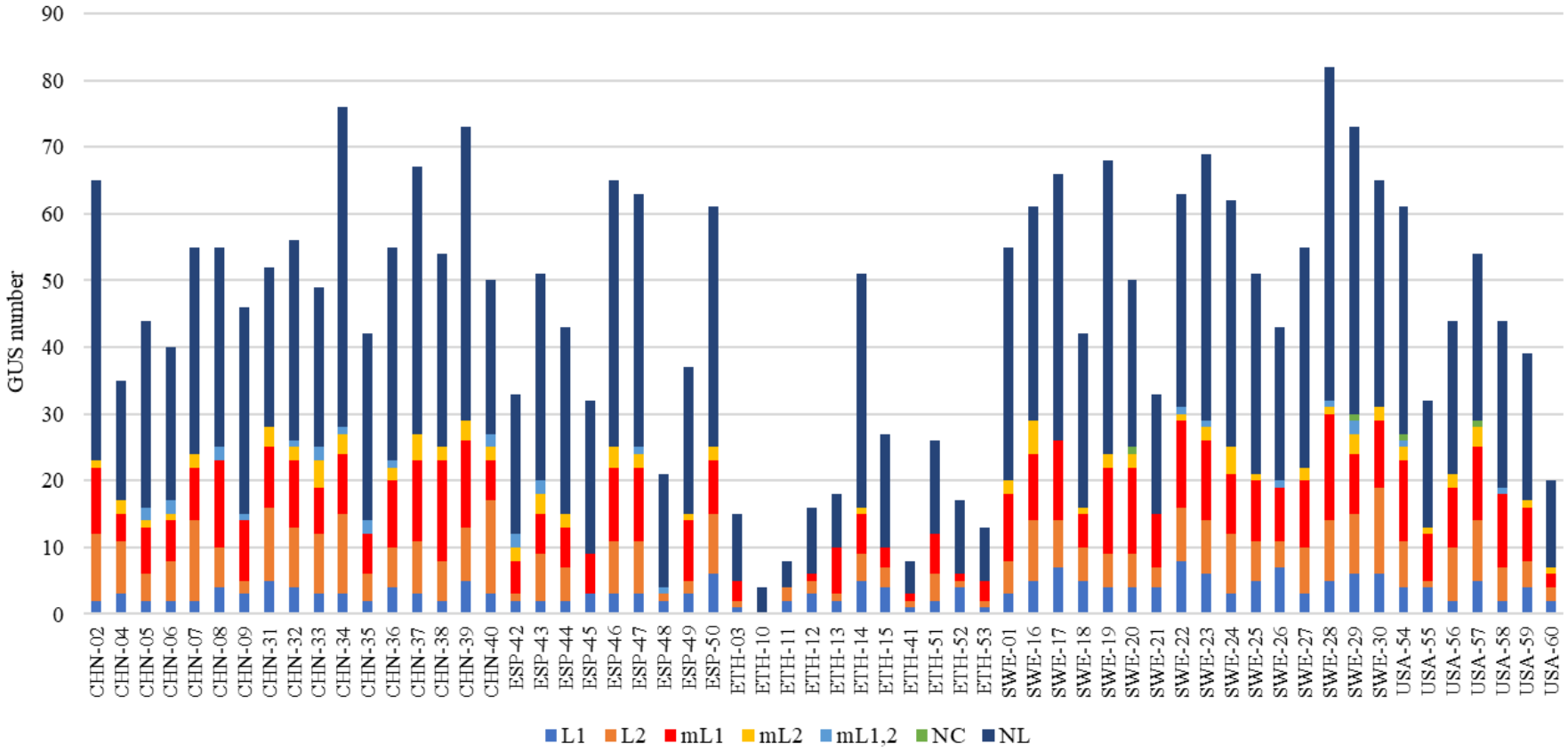


Figure 5.7 Number of GUS identified in each sample. Colors indicate different GUS classes.

The most represented class of GUS was NL, accounting for up to 50 diverse sequences per sample, followed by mL1 (up to 16), L2 (up to 14), and L1 (up to 8). Classes mL2, ml1,2, and NC were less represented or absent in some genomes (Fig. 5.7).

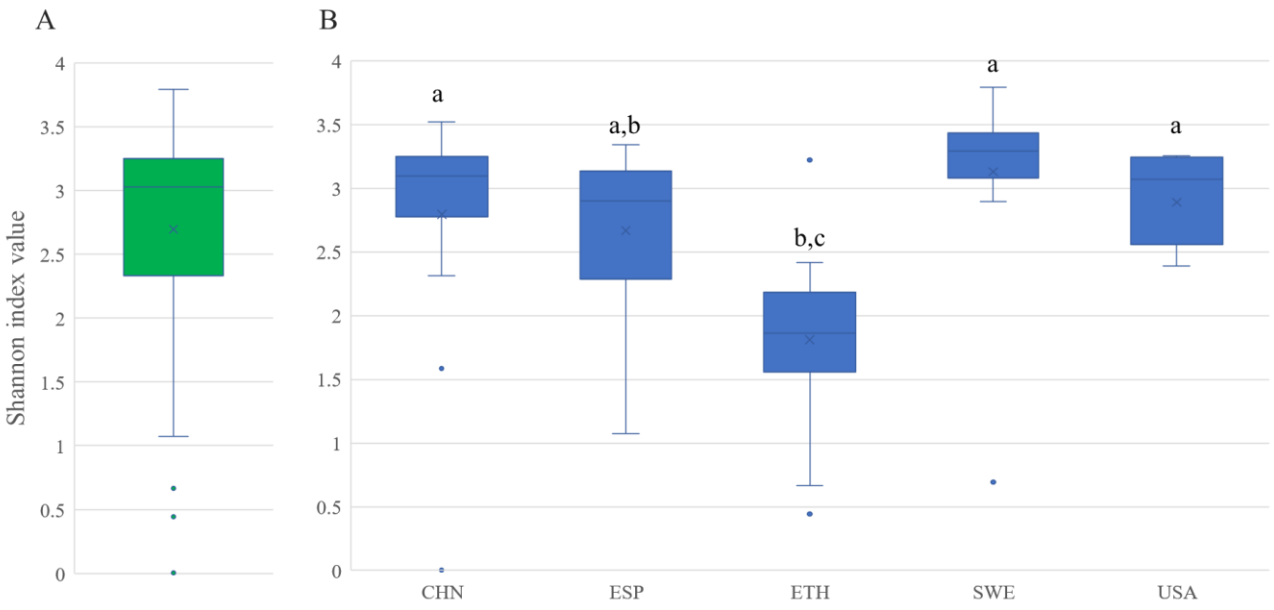


Figure 5.8 Alpha diversity of the GUS abundance in the complete dataset (A) and in the five cohorts (B) based on Shannon's index. Different letters indicate significant differences among cohorts (P<0.05, ANOVA, Tukey)

The class NL was identified in all the metagenomes, accounting for 129/218 of the detected sequences. The class L1 was found in most of the samples (58/60), being represented by 13 sequences. The classes mL1and L2 were both identified in 57/60 metagenomes, with 33 and 30/218 sequences, respectively. GUS genes belonging to the class mL2 were found in 43 samples and 7 different sequences were identified. Classes ml1,2 and NC were the least represented classes in all the samples. NC was detected in 5 metagenomes with 2 sequences; ml1,2 was identified in 21 samples and encompassed 4 different sequences.

Despite the different distribution of structural classes among the subjects, there was no statistical significance in their distribution among the cohorts (P> 0.05, ANOVA, Tukey). In the database of 279 GUS sequences elaborated by Pollet et al. (2017), the classes NL, mL1, L2, L1, mL2, ml1,2 and NC accounted for 160, 41, 39, 16, 11, 6, 6 sequences. respectively. For most of the classes, more than 60% of the database GUS sequences were found in our dataset, except for class NC, where only 2 out of the 6 known GUS were detected.

The GUS abundance representing the relative abundance of intestinal bacteria encoding at least a GUS gene, was calculated associating each GUS with the relative abundance of the corresponding bin, in its turn obtained by the number of reads mapping in each bin. The dataset presented a high variability, even for samples belonging to the same cohort (Figure 5.9). For instance, in sample CHN-9 bacteria harbouring GUS genes accounted only for 0.7%, whereas in sample CHN-05 69.9% of bacteria encoded GUS genes. The mean abundance of GUS encoding bacteria was 27.0% (Fig 5.10A). The amount of bacteria harbouring GUS genes was significantly lower in ETH cohort than in the others ($P < 0.05$) (Fig. 5.10B).

Abundances, frequencies, and taxa mostly contributing to GUSome were explored. The taxonomic attribution of GUS genes was based on taxa associated to each GUS type by Pollet etl al (2017). For each sample, the abundance of a GUS type was calculated summing the relative abundance of each bin harbouring at least a GUS gene, normalized among the whole set of bins, encompassing or not the GUS genes. 37% of the bins identified in the 60 metagenomes encoded GUS genes. The three most abundant GUS were sequences 11, 176, and 36, ascribed to the classes NL, mL1, and NL, respectively. They were also very frequent, being identified in 51, 40, and 35/60 samples, respectively. In particular, sequence 11 was both the most abundant and the most frequently reported. Sequences 11, 176, and 36 all belonged to *Bacteroides.* Among the 20 most abundant GUS sequences, 13 belonged to the genus *Bacteroides*, and one to *Parabacteroides*. The sole genus *Bacteroides* contributed for the 56% to the GUS abundance.

Sequence 220 (L1), ascribed to *Escherichia coli,* even though present only in 12/60 samples, reached the highest relative abundance in a single sample (24.7% in ESP-48) and was the fourth most abundant one. In samples ESP-48, ESP-45, and ETH-13, the corresponding bin of *E. coli* accounted for 24.7, 16.4, and 9.8% of the bacterial populations, respectively. Sequences 223 (L1) and 67 (NL), both from *Faecalibacterium prausnitzii,* were among the most frequently encountered, being present in 45 and 41/60 samples, respectively. The highest relative abundance reached by the corresponding bin was 3.7 % in sample ESP-42 where both sequences have been identified (Table 5.2).

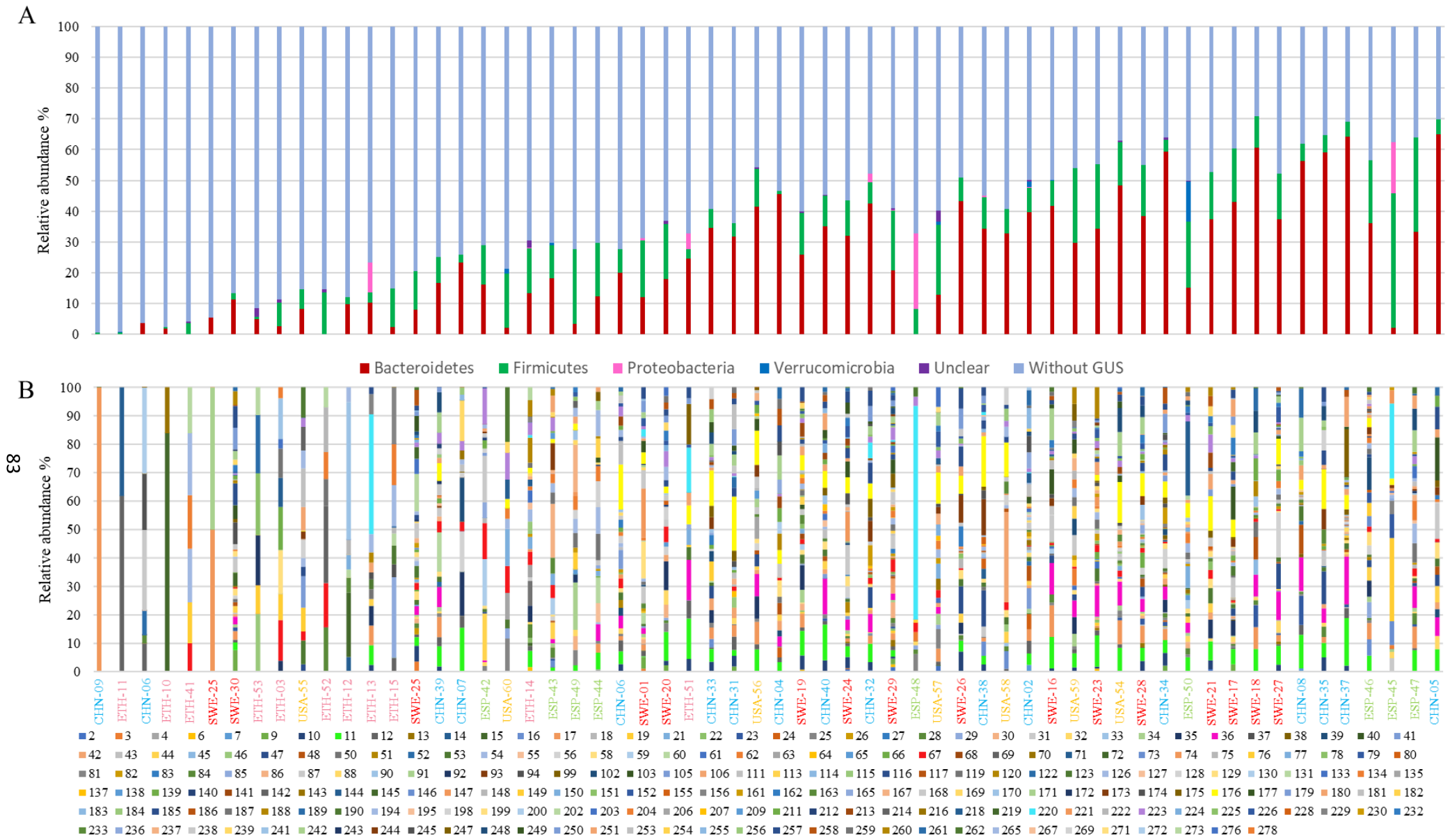


Figure 5.9 GUS abundance profile in each sample. Panel A shows the distribution of bins containing at least one GUS sequence, colored according to their taxonomic designation at phylum level. Panel B shows the distribution of GUS sequences.

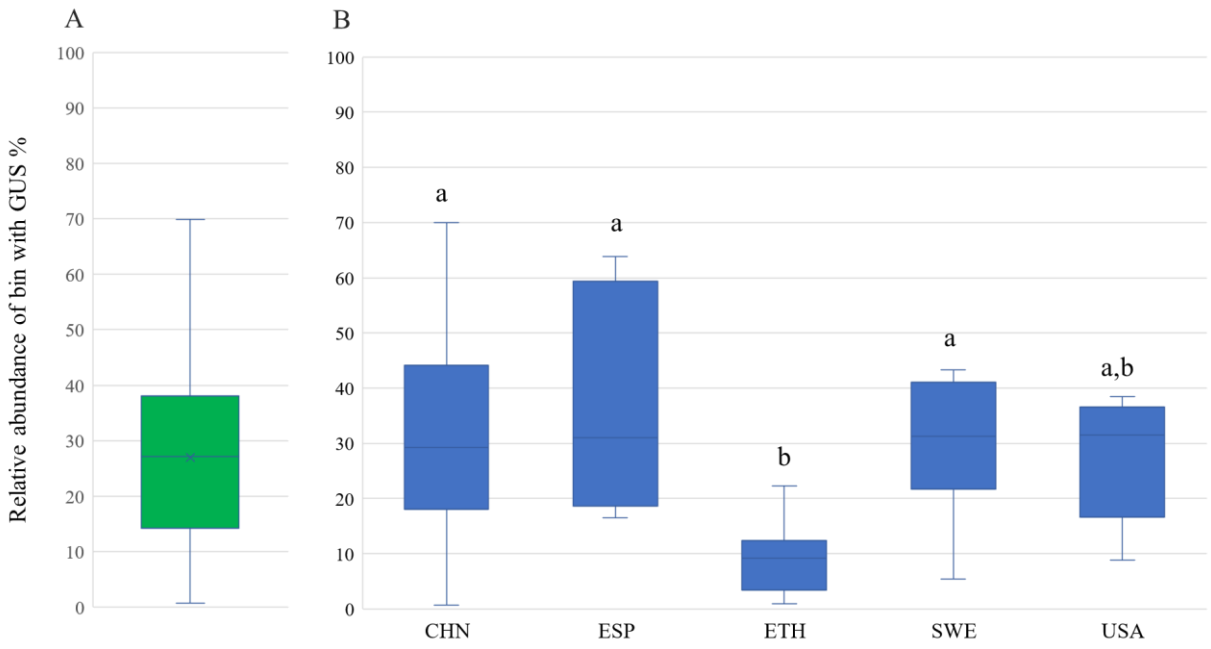


Figure 5.10 Boxplot of the abundance of GUS encoding bacteria for the whole dataset (A) and for the different cohorts (B). Different letters indicate significant differences among cohorts (P<0.05, ANOVA, Tukey).

Table 5.2 List of first 20 most represented GUS sequences. Frequency, relative abundance in the whole dataset, and highest value reached in single sample are shown.

GUS sequence	**GUS type**	**Sequence origin**	**Frequency (N° of samples)**	**Percentage in the dataset (%)**	**Highest value in single sample (%)**	**Sample with highest value**
11	NL	*Bacteroides* multispecies	51	5.8	11.6	CHN-37
176	mL1	*Bacteroides vulgatus*	40	5.2	9.1	CHN-35
36	NL	*Bacteroides uniformis*	35	4.5	11.6	CHN-37
220	L1	*Escherichia coli*	12	2.7	24.6	ESP-48
17	NL	*Bacteroides uniformis*	27	2.6	5.6	SWE-16
242	L2	*Bacteroides uniformis*	23	2.4	7.3	CHN-08
87	NL	*Bacteroides dorei*	28	2.3	8.8	CHN-05
47	NL	*Bacteroides massilensis*	22	1.9	7.4	CHN-35
35	NL	*Bacteroides uniformis*	16	1.8	4.3	USA-56
67	NL	*Faecalibacterium prausnitzii*	41	1.7	3.7	ESP-42
177	mL1	*Bacteroides dorei*	16	1.6	10.6	CHN-05
173	mL1	*Bacteroides massiliensis*	25	1.6	5.7	CHN-38
223	L1	*Faecalibacterium prausnitzii*	45	1.6	3.7	ESP-42

185	mL1	*Bacteroides ovatus*	32	1.6	5.6	CHN-05
257	L2	*Bacteroides ovatus*	41	1.5	3.6	SWE-26
10	NL	*Bacteroides ovatus*	39	1.5	3.6	SWE-26
76	NL	Eubacterium sp.CAG:180	15	1.4	18.2	ESP-45
53	NL	*Parabacteroides merdae*	42	1.4	3.2	CHN-08
126	NL	Firmicutes	40	1.4	4.4	ESP-47
134	NL	Firmicutes	42	1.3	4.4	ESP-47

A qualitative measure of beta-diversity of GUS encoding taxa was calculated with Jaccard similarity, to avoid that abundance of *Bacteroides* could obscure significant patterns of variation among the cohorts. The output was processed through Principal Coordinate Analysis (Figure 5.11). Components PCo1 and PCo2 described 11.8% and 6.4% of the variance, respectively. Cohort grouping based on GUS profile was statistically relevant according to statistical analysis PERMANOVA ($P < 0.05$). CHN and SWE samples were characterized by negative PCo1, ETH by positive one. Most of the taxa that negatively contributed to PCo1 were *Bacteroides*: a *Bacteroides* sp., a *Bacteroides vulgatus* and three *B. uniformis*. On the other side, main positive contributors to PCo1 were *Prevotella*, *Prevotella copri*, *Faecalibacterium prausnitzii*, *Eubacterium*, and *Escherichia coli*, suggesting that GUS profile can be due to the presence of diverse species and genera, also on the basis of geographic provenience.

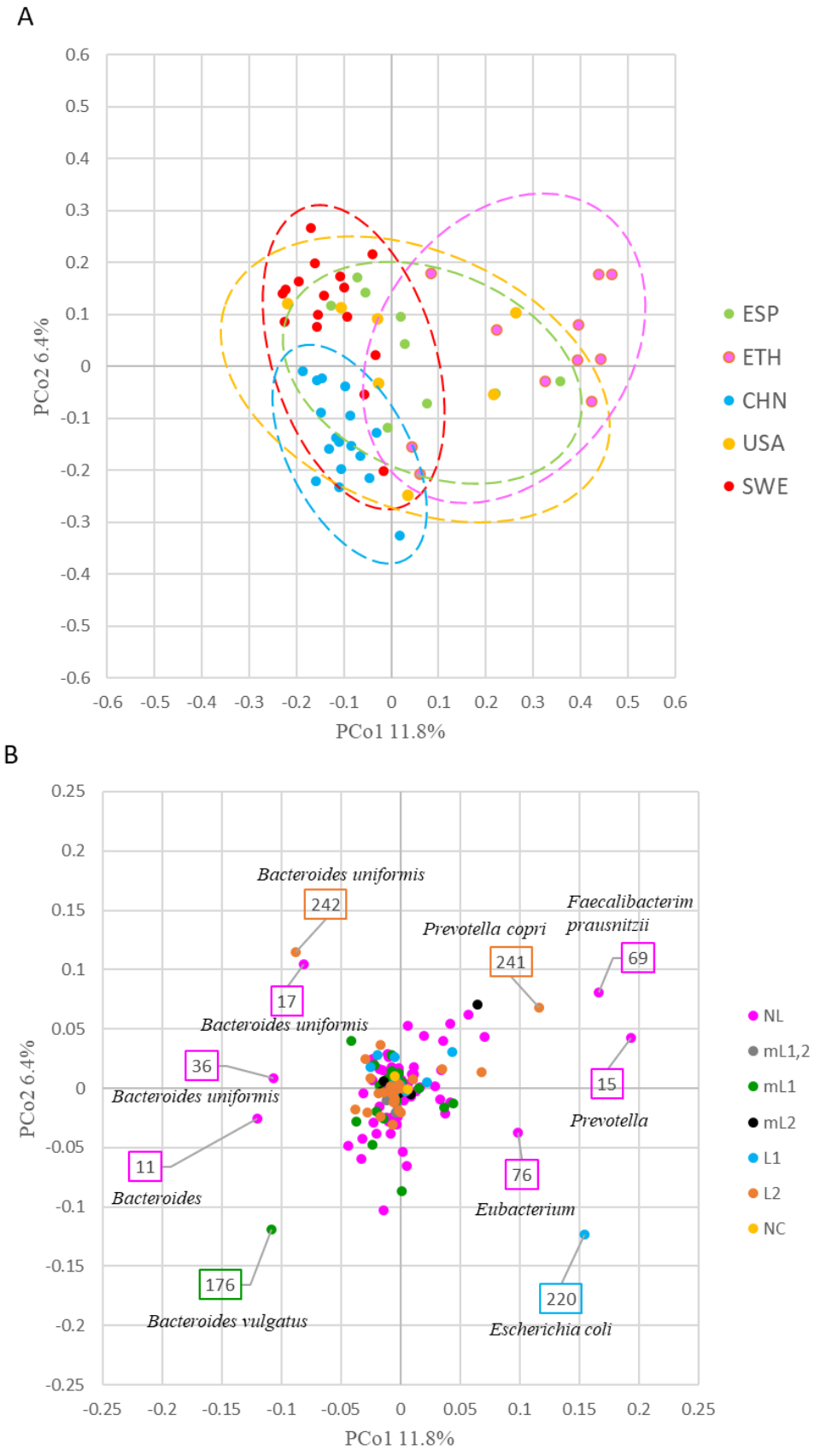


Figure 5.11 (A) PCoA plot of beta-diversity based on Jaccard dissimilarity index of the GUS profiles of the 60 metagenomes. (B) PCoA plot of the GUS contribution to sample differentiation

5.4 CONCLUSIONS

60 metagenomes from fecal samples of healthy subjects coming from different geographical regions were analyzed to identify differences in GUS profile and to correlate them to the bacterial composition. Two combined pipelines were applied to retrieve microbial composition and GUS profile of each sample. The taxonomic analysis conducted with MetaPhLan2 was opted for, due to the more accurate microbial profiling and to higher number of taxa identified, compared to bins reconstruction. According to MetaPhlan2, most of the cohorts were dominated by Firmicutes, except for CHN where Bacteroidetes were prevalent. The differences in terms of microbial composition resulted statistically significant based on the diverse provenance.

The functional analysis allowed to identify 218 of the 279 GUS sequences, with a high variability per samples, that encompassed 4 to 82 different GUS genes. The cohorts showed similar richness, except for ETH samples that presented lower values. The most abundant structural class was NL, followed by mL1, L2, L1, mL2, ml1,2, and NC. The abundance of bacteria encoding GUS genes was also very variable (0.7 to 69.9%), so it is likely that the gut microbiota of different subjects possesses a quite different deglucuronidation capacity. GUS sequences 11, 176 and 36, all belonging to genus *Bacteroides*, were the most abundant and among the most frequently detected among the samples. Sequences identified in *Bacteroides* (13 in total) were among the 20 most abundant GUS in the dataset, and some of them heavily contributed to sample differentiation in PCoA plot based on the Jaccard similarity distance of the GUS types. This data highlights the relevance of GUS sequences encoded by Bacteroides, as already revealed in Pollet et al. (2017).

The PCoA analysis also suggested that the differences in GUS profiles can be due to both the presence of diverse species and genera and the geographical provenience. These differences, in terms of both GUS types and quantity of GUS-encoding bacteria, could explain a diverse biocatalytic activity that can contribute to a different response against glucuronidated drugs and xenobiotics. A different contribution of the gut microbiota to deglucuronidation could explain the anomalous responses to certain drugs documented for some subjects and could be took into consideration for drug design. New studies focused on the correlation between GUS classes and microbial composition also based on model substrates are highly recommended.

The manuscript reporting the data of this project is in draft.

6. Project 3

16S rRNA metagenomic profiling: mucin degraders in human gut microbiota

6.1 INTRODUCTION

6.1.1 Mucus and mucins

Mucus is a complex gel barrier that covers the wet epithelial surfaces throughout the body, including the gastrointestinal tract, offering protection against exogenous and endogenous aggressive agents (Bansil and Turner, 2018; Alemao et al., 2020). It exerts a variety of functions such as lubrication, hydration, chemical protection, sensing, nutrient reservoir, and barrier against pathogen invasion. A continuous turnover, consisting in a dynamic balance between biosynthesis, secretion, and degradation of the structural components, is crucial for these functions (Wagner et al., 2018).

Mucins are major component of mucus, made up by glycoproteins with high level of glycosylation. Galactose, N-acetylglucosamine, N-acetylgalactosamine, fucose, and sialic acid, with relatively small amounts of mannose, form the oligosaccharides which represent approx. 80% of mucin mass (Tailford et al. 2015). The protein core is mainly organized in tandem repeated regions rich in serine, threonine, and proline. Serine and threonine are the site where an O-glycosid bound links the protein core to GalNAc, the first moiety of the glycan chain (Bergstrom and Xia, 2013; Corfield, 2018).

Among all the sites of human body inhabited by a resident microbial community, the colon hosts the most complex and concentrated microbiota (Sender et al., 2016). In this site, mucins behave as decoy molecules that avoid the interaction of bacterial adhesins with receptors of the colonic epithelium. The gut lumen is coated by a two-layered mucus system, with an inner dense stratus firmly attached to the epithelium and an outer one, looser and unattached (Johansson et al., 2011). The inner layer is thick and stratified and prevents gut bacteria from reaching the epithelial cell surface. It is progressively converted into a looser and expanded coating through the lytic action of proteases and glycosidases of both the host and the commensal bacteria. The outer mucus layer is the colonization site of the resident commensal microbiota, which differs from that of the digesta-associated and fecal content in terms of relative abundance of the different taxa (Donaldson et al., 2016).

Mucus layers are the frontline of the interaction between the host and the gut bacteria, thus a balanced and symbiotic relationship which benefits both the former and latter greatly relies on mucus structure (Alemao et al., 2020**)**. Mucin plays a pivotal role in the selection of bacteria colonizing the mucus by supplying carbon and nitrogen substrates and exposing O-glycan chains that serve as attachment sites for colonization (Johansson et al., 2011; Tailford et al. 2015). In turn, commensal bacteria feeding on and adhering to mucin limit the penetration of pathogens in the outer layer of the mucus (Linden et al., 2008). These bacteria tightly interact with the host and modulate mucin gene expression, glycosylation, and secretion, thus affecting mucus and epithelial homeostasis (Cornick et al. 2015; Wagner et al. 2018).

6.1.2 The metabolism of mucins

Saccharolytic metabolism

Approximately 20-30 g of diet-derived carbohydrates reach the colon every day, escaping digestion by the enzymes of the small intestine. Carbohydrates are the carbon sources for the saccharolytic bacteria, that represent the vast majority of the intestinal microbiota. The carbohydrates reaching the colon are mainly resistant starches, non-starch polysaccharides (NSP) including cell wall polysaccharides, oligosaccharides, and some mono- and di- saccharides (Macfarlane, 1997).

In the proximal colon and in the caecum, the fermentation of carbohydrate represents the major activity of the gut microbiota. The saccharolytic metabolism of bacteria consists of

the breakdown of polysaccharides to smaller oligomers and their component sugars that are further channeled into their fermentative routes. This metabolism yields short chain fatty acids (SCFA), such as acetate, propionate, and butyrate, and lowers the pH in the large intestine. Others minor metabolites such as ethanol, lactate, and succinate do not accumulate in the colon because they are metabolized by cross-feeding bacteria. SCFA concentrations are the highest in the proximal colon because of the availability of substrates, the microbiota composition, the pH and the hydrogen partial pressure (Louis et al., 2014).

The production of oxidized substances as acetate is directly associated with ATP generation, therefore acetate is the most abundant SCFA released in the colon. It is produced by numerous intestinal bacteria as a fermentation product of carbohydrates and also by the specific group of acetogenic bacteria (e.g. *Blautia hydrogenotrophica*) from H_2, CO_2 and from formate.

The amount of propionate released and present among fecal SCFA depends on the relative abundance of *Bacteroidetes* and some *Firmicutes*. Bacteria belonging to the genus *Bacteroides* yield propionate predominantly via the succinate pathway, but also the acrylate or propanediol pathways can occur. On the other hand, some *Firmicutes* produce butyrate from butyrate, lactate, and other organic acids derived from saccharolytic fermentation. Eventual butyrate production is catalyzed by either butyrate kinase or butyryl-CoA:acetate CoA transferase. The acetate CoA transferase pathway involves numerous of the most abundant species that colonize a healthy gut microbiota: *Fecalibacterium prausnitzii*, *Roseburia spp*., *Eubacterium rectale*, *Eubacterium halii* and *Anaerostipes spp* (Anand et al., 2016).

SCFA are rapidly adsorbed from the gut lumen and serve as nutrients for the colonocites. They also modulate colonic regulatory T-cells, contribute to down regulate pro-inflammatory cytokines, such as IL-6 ad IL-12, in colonic macrophages (Smith et al., 2013, Chung et al., 2012), and improve the solubility and the absorption of minerals such as calcium. The metabolism of carbohydrates also decreases the concentrations of potentially harmful metabolites derived from proteolytic activity in the colon (Hamer et al., 2012).

Proteolytic metabolism

Proteins are another major source of carbon and energy for colonic bacteria. Most of dietary proteins are digested and absorbed in the small intestine, with unabsorbed proteins and peptides that reach the colon where are fermented by the residing bacteria. Carbon dioxide, ammonia, and organic acids are primary products of amino acid fermentation (Barker, 1981; Riedel et al., 2017). Fermentation products are linear and branched organic acids (acetate, butyrate, propionate, valerate, isobutyrate, 2-methylbutyrate, isolavalerate, etc.), phenols and indole derivatives, nitrogen or sulphur-containing compounds (Smith and Macfarlane, 1996; Davila et al., 2013; Russell et al., 2013a).

Many of these products are detrimental to health (in particular ammonia, phenols, indoles, amines, sulfides and N-nitroso compounds), being responsible for a variety of harmful effects, including systemic toxicity, nephrotoxicity, and carcinogenesis (Barrios et al., 2015; Russel et al., 2013a, 2013b; Blachier et al., 2010; Kobayashi, 2017).

The first steps in proteins breakdown in the colon are carried out by a variety of hosts endopeptidases and bacterial proteases (MacFarlane and Cummings, 1991). The proteolysis yields peptides and/or amino acids that can be subjected to Stickland reactions or other fermentation pathways by true amino acid-fermenting bacteria (Smith and Macfarlane, 1998; Kim et al., 2004).

Stickland reactions (Fig. 6.1), consisting in the coupled oxidation and reduction of pairs of amino acids, are common examples of amino acid fermentation routes, yielding organic acids and ammonia (Smith and Macfarlane, 1998). However, aminoacids and their derivatives can undergo a variety of bacterial transformations, which yield a particularly diversified spectrum of products. In fact, they can serve as electron acceptors or donors in oxidations and reductions of other compounds (cheto-acids, H_2, unsaturated fatty acids), and can be further subjected to decarboxylation and transamination reactions.

Amino acids like valine, isoleucine and leucine are mostly employed by bacteria belonging to the *Clostridium* genus. These amino acids are utilized in Stickland fermentation (Larsson, 2006) to yield SCFA as acetate, butyrate, propionate, and branched chain fatty Acids (BCFA) as isobutyrate, 2-methylbutyrate and isolavalerate.

A recent paper (Amaretti et al., 2019) aimed to identify intestinal proteolytic bacteria present in the colon. The experimental design was similar to that we used for this project focused on mucin degraders. Fecal enrichment cultures were performed inoculating fecal

slurries of healthy individuals on a medium containing only proteins and peptones as carbon source. Bacterial composition of fecal cultures was compared during the enrichment, in order to identify these microbial groups that significantly get enriched during the enrichment process. The study was the first approach exploiting metataxonomic analysis to identify the proteolytic bacteria of the intestinal gut. 16S rRNA gene profiling permitted the identification of unculturable bacteria taking advantage of only proteins and peptones as fermentable substrates, expanding the previous knowledge, based on traditional culture approaches.

Functional information on the bacteria involved in gut protein catabolism date to culture-based studies dependent on isolation and physiological characterization of cultivable species (Cummings and MacFarlane, 1991; Smith and Macfarlane, 1998; Russell et al., 2013a). Bacteria able to exert a proteolytic metabolism in the human colon belong to the genera *Bacteroides*, *Propionibacterium*, *Clostridium*, *Fusobacterium*, *Streptococcus*, *Eubacterium* and *Lactobacillus* (MacFarlane and Cummings, 1991). These genera encompass different proteolytic species: *Bacteroides thetaiotaomicron, Bacteroides eggerthii, Bacteroides ovatus, Bacteroides fragilis, Parabacteroides distasonis, Clostridium bartlettii,* e *Eubacterium hallii*.

16SrRNA profiling of the enrichment cultures and analysis of differential abundances revealed that the families Enterobacteriaceae, Burkholderiaceae, and Desulfovibrionaceae, and the genera *Esherichia-Shigella*, *Sutterella*, *Parasutterella*, and *Bilophil*a proliferated in the protein based medium. The families Lachnospiraceae and Ruminococcaceae encompassed many taxa that significantly expanded, showed the strongest correlation with products of amino acid fermentations such as ammonium, indole, and p-cresol. In particular, the genera *Anaerotruncus, Dorea, Oscillibacter, Eubacterium oxidoreducens, Lachnoclostridium, Paeniclostridium, Rombutsia* significantly increased in terms of relative amounts.

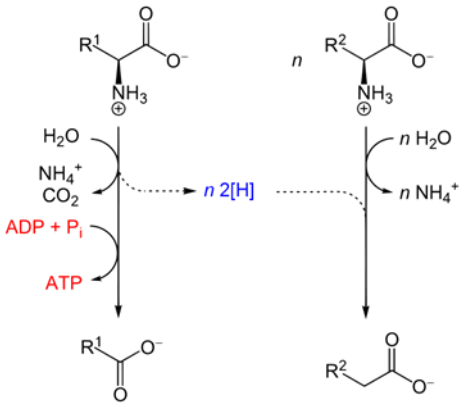


Figure 6.1 General scheme of Stickland fermentation of aminoacids.

Fermentation of aromatic amino acids such as phenylalanine, tyrosine, and tryptophan by intestinal anaerobes yields phenolic compounds (De Loor et al., 2005). These aromatic compounds can be absorbed by the host and, in normal conditions, they are detoxified and excreted. However, in cases of not adequate functionality of kidney and liver, they can accumulate in tissues and became toxic. The main products of the bacterial degradation of aromatic amino acids are p-cresol and phenylpropionate (from tyrosine), phenylacetate (from phenylalanine), and indole propionate and indole acetate (from tryptophane).

Clostridia, *Bacteroides*, and *Enterobacteria* are involved in the production of phenolic compound that are absorbed in the colon, detoxified by the liver and excreted in the urine as p-cresols (Tamm and Villako, 1971, Amaretti et al., 2019). The phenolic compounds are also referred to as Uremic Retention Molecules (URMs) or uremic toxins and their accumulation over time leads to the appearance of several diseases as Uremic Syndrome or cancer or hepatic encephalopathy. Phenolic compounds can alter and inhibit the endothelial proliferation and reduce the regeneration ability of the epithelium (Dou et al. 2004). Moreover, p-cresol could have a harmful role in patients affected by chronic kidney disease worsening the dysfunction of the vascular endothelial smooth muscle. p-cresylsulfate has a pro-inflammatory effect on the leucocyte (Schepers et al., 2007).

Starting from the metabolism of the tryptophan, colonic bacteria such as *E. coli* can release indolic compounds through the action of triptophanase. Inside the intestine, indole is normally absorbed into the blood flow and while it reaches the liver it is metabolized into indoxylsulfate and excreted into the urine. Indoxyl sulfate and its precursor indole are related

to a decreased renal functionality, an increased possibility of glomerular sclerosis and to a major transcription of the TGF-β1gene that cause nephrotoxicity and renal fibrosis (Adijiang et al., 2008).

It was demonstrated that in presence of readily fermentable carbohydrates, the net production of phenolic compounds decrease (Smith and MacFarlane, 1996); moreover, it seems probable that, like ammonia, these sources of nitrogen are utilized for bacterial growth when carbohydrate sources are depleted, and the longer transient times promote a more efficient proteolytic metabolism increasing amino acids fermentation.

6.1.3 Fermentation of mucins

Mucins are glycoproteins and can potentially be fermented by both saccharolytic and proteolytic bacteria. Fermentation of mucins firstly requires hydrolysis and degradation of the complex array of oligosaccharides linked to the protein skeleton. In turn, this hydrolysis involves a number of different enzymatic activities, according to the diverse bonds linking the different monomers. The protein backbone can be attacked and fermented only when glucidic chains have been utilized as an energy source by some commensal microorganisms, allowing a more limited proteolytic metabolism, on the ratio of carbohydrate : protein in mucins.

6.2Aim of the project

This study aimed to investigate mucin degraders of the gut, in order to unveil how mucins influence selection of commensal bacteria and how specific bacteria shape the mucus layer. In strictly anaerobic conditions, fresh feces of 5 healthy adults were subjected to three consecutive steps of enrichment in a medium containing only mucins as carbon and nitrogen sources. To identify the taxa that grew taking advantage of mucins as nitrogen and carbon source, the composition of the bacterial community was compared before and after the enrichment steps by metataxonomic analysis.

6.3 MATERIALS AND METHODS

6.3.1 Enrichment cultures

Five healthy adults (3 men and 2 women, aged 25–50 years) were enrolled for the study after having obtained written informed consent, according to the experimental protocol approved with ref. no. 125-15 by the local research ethics committee (Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy). Subjects had not taken prebiotics and/or probiotics in the previous 2 weeks or antibiotics for at least 3 months prior to sample collection. Fecal samples were collected fresh, sealed in anaerobic plastic bags (AnaeroGen, Oxoid, Basingstoke, UK), transferred in an anaerobic cabinet (Concept Plus, Ruskinn Technology, Ltd., Bridgend, UK) with an atmosphere of 85% N_2, 10% CO_2, and 5% H_2, and homogenized (10%, w/v) in sterile mucin-based medium (MM). MM contained 3.0 g/L mucin from porcine stomach type II, 2.0 g/L KH_2PO_4, 4.5 g/L NaCl, 0.5 g/L $MgSO_4 \cdot 7H_2O$, 0.045 g/L $CaCl_2 \cdot 2H_2O$, 0.005 g/L $FeSO_4 \cdot 7H_2O$, 0.01 g/L hemin, 0.05 g/L bile salts (Oxgall, BD Difco, Sparks, MD, United States), 0.6 mg/L resazurin, 2.0 mL/L minerals solution (0.5 g/L EDTA, 0.010 g/L $ZnSO_4 \cdot 7H_2O$, 0.003 g/L $MnCl_2 \cdot 7H_2O$, 0.03 g/L H_3BO_3, 0.02 g/L $CoCl_2 \cdot 6H_20$, 0.001 g/L $CuCl_2 \cdot 2H_2O$, 0.002 g/L $NiCl_2 \cdot 6H_2O$, and 0.003 g/L $NaMoO_4 \cdot 2H_2O$), 1.4 mL/L vitamins solution (1.0 g/L menadione, 2.0 g/L biotin, 2.0 g/L calcium pantothenate, 10 g/L nicotinamide, 0.5 g/L cyanocobalamin, 0.5 g/L folic acid, 4 g/L thiamine, and 5 g/L PABA), and 40 mL/L reducing solution (12.5 g/L L-cysteine • HCl and 80 g/L $NaHCO_3$). The fecal slurries (hereinafter referred to as FS) were inoculated with a syringe (10% v/v) into butyl-rubber stoppered bottles containing 50 mL of sterile anaerobic MM. After an incubation of 72 h at 37 °C, these cultures were utilized to seed (10% v/v) the next passage in MM, repeated to accomplish three enrichment steps. Once grown, the third enrichment culture (hereinafter referred to as EC) was utilized for 16S rRNA metagenomic profiling.

6.3.2 DNA extraction

Total DNA was extracted from 2 mL of FS and EC samples with the QiAmp PowerFecal DNA kit (Qiagen, Hilden, Germany), according to the manufacturer's protocol. The DNA was normalized to 5 ng/µL after quantification with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA).

6.3.3 Sequencing

Partial 16S rRNA gene sequences were amplified using Probio_Uni / Probio_Rev primers, which targeted the V3 region of the 16S rRNA gene. Amplicons were sequenced using a MiSeq platform (Illumina Inc., San Diego, CA, USA) according to Milani et al. (2013).

6.3.4 16S rRNA gene profiling

Raw sequences were analyzed with the QIIME2 pipeline, version 2019.10 (Bolyen et al. 2019), with appropriate plugins for trimming (CUTADAPT) and denoising (DADA2) into amplicon sequence variants (ASVs) (Martin 2011; Callahan et al. 2016). Taxonomy assignment was carried out with the feature classifier VSEARCH (Rognes et al. 2016) utilizing as reference SILVA SSU database release 138 (https://www.arb-silva.de/download/arb-files/) with the similarity threshold set at 0.97. The appropriate QIIME2 plugins were utilized to compute the alpha- (observed taxa, Chao1, Shannon, and Pielou's evenness) and beta-diversity (Unweighted UniFrac, and Weighted UniFrac) and to compare them within and between FS and EC samples (i.e., the Kruskal-Wallis test for alpha diversity; ANOSIM and PERMANOVA for beta-diversity). Principal Coordinate Analysis (PCoA) was computed with QIIME2, based on the beta-diversity distance matrices.

Linear discriminant analysis Effect Size (LEfSe, http://huttenhower.sph.harvard.edu/galaxy) algorithm was utilized to discover distinctive taxonomic features characterizing FS and EC samples (Segata et al. 2011).

The 16S rRNA gene sequences datasets generated and analyzed during the current study are available in the NCBI repository with the BioProject ID: PRJNA649741 (https://www.ncbi.nlm.nih.gov/).

6.3 RESULTS

The metataxonomic survey of 16S rRNA gene in FS and EC samples yielded a total 415,051 sequences (31,153 – 55,162 *per* sample). The reads were dereplicated into 522 ASVs hitting a reference sequence in Silva database, and collapsed at the 7th level of taxonomic annotation into 150 OTUs.

The analysis of microbiota alpha-diversity revealed that the richness (Chao1) and the evenness (Shannon and Pielou) tended to decrease with the enrichment, but did not reach statistical significance (P > 0.05) (Fig. 6.2).

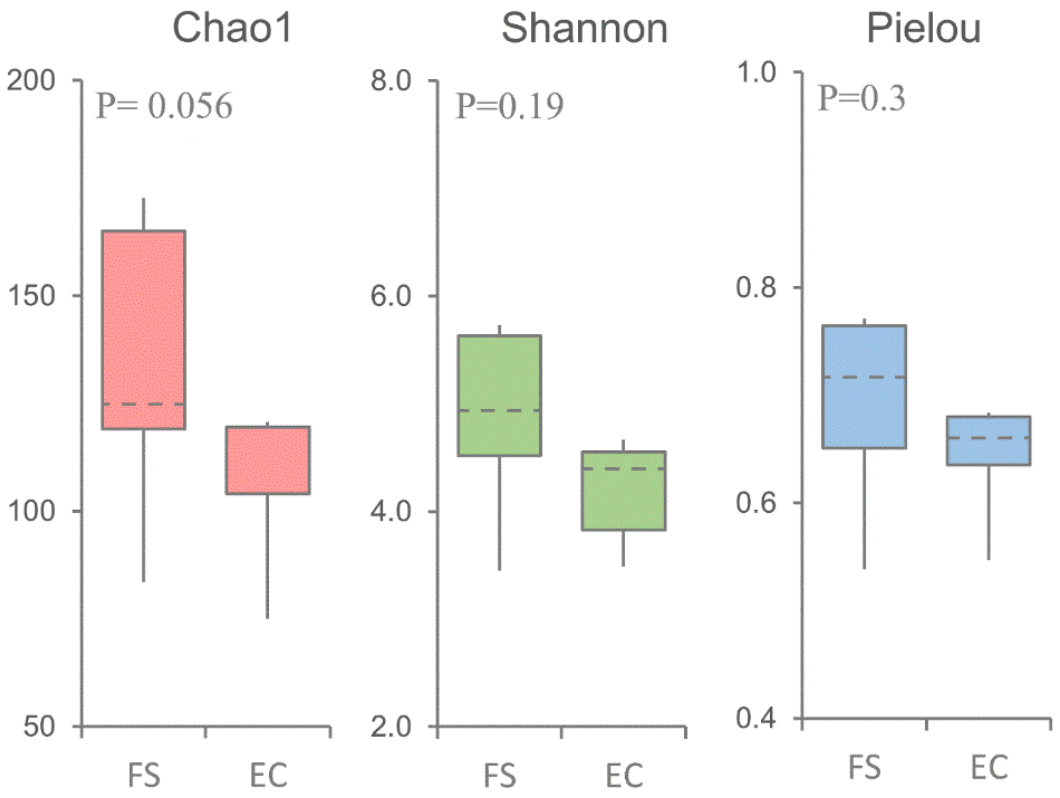


Figure 6.2 Alpha diversity metrics, comprising Chao1, Shannon, and Pielou indexes, of the microbiota in FS and EC samples. The median (dashed line), the 25th and 75th percentiles (colored box), the 10th and 90th percentiles (whiskers), are indicated. Means were compared between paired samples with Kruskal-Wallis test.

According to Weighted UniFrac computation of beta-diversity, FS and EC samples clustered in distinct groups (ANOSIM, PERMANOVA, $P < 0.05$) in a PCoA plot, with the former laying at lower values of PCo1 than the latter (Fig. 6.3A). *Escherichia-Shigella* presented the most positive contribution to PCo1, followed by ASVs of *Bacteroides* and of Lachnospiraceae, while other ASVs of *Bacteroides*, Prevotellaceae NK3B31, and *Roseburia* negatively weighted (Fig. 6.3B). Unweighted UniFrac yielded more dispersed groups (ANOSIM, PERMANOVA, $P > 0.05$), with EC samples laying at lower PCo2 compared to the corresponding FS ones (Fig. 6.4A). The ASVs that negatively contributed to PCo2 included *Clostridium sensu stricto I* and other *Bacteroides*, while Prevotellaceae NK3B31 and *Eubacterium coprostanoligenes* were positive contributors (Fig. 6.4B). Weighted UniFrac computation takes into account abundances of taxa, unlike Unweighted Unifrac that considers the mere presence or absence. Thus, the taxa that emerged from the former presented relevant changes in terms of relative abundance, while those that emerged from the latter likely appeared or disappeared over the enrichment steps.

The microbiota of FS samples encompassed 125 of the 150 identified OTUs. It was dominated by Bacteroidota (59.5%), with remarkably high amounts of *Bacteroides* (39.3%) and Prevotellaceae (8.4%), followed by Firmicutes (35.3%), especially Lachnospiraceae (15.4%), and Proteobacteria (2.2%; Fig. 6.5). The microbiota of EC samples encompassed 101 OTUs, 25 of which became detectable as a result of the enrichment steps, as they lay below the limit of detection in FS samples. EC were dominated by Bacteroidota (39.3%) and Proteobacteria (30.8%). *Escherichia-Shigella* (with Enterobacteriaceae, Enterobacteriales, Gammaproteobacteria, and Proteobacteria) presented a significant positive differential abundance in EC compared to FS (Fig. 6.6) and represented one of the dominant genera in EC samples (mean = 30.5%), with remarkably high abundance in sample EC-5 (56.7%). Several taxa of Firmicutes, such *Enterococcus* and many Clostridia (e.g. Oscillospiraceae, Peptococcaceae, Ruminococcaceae UBA1819, *Soleaferrea*, *Lachnoclostridium*, and *Anaerotruncus*), presented a significant increase in EC samples, but generally remained below a mean abundance of 1%.

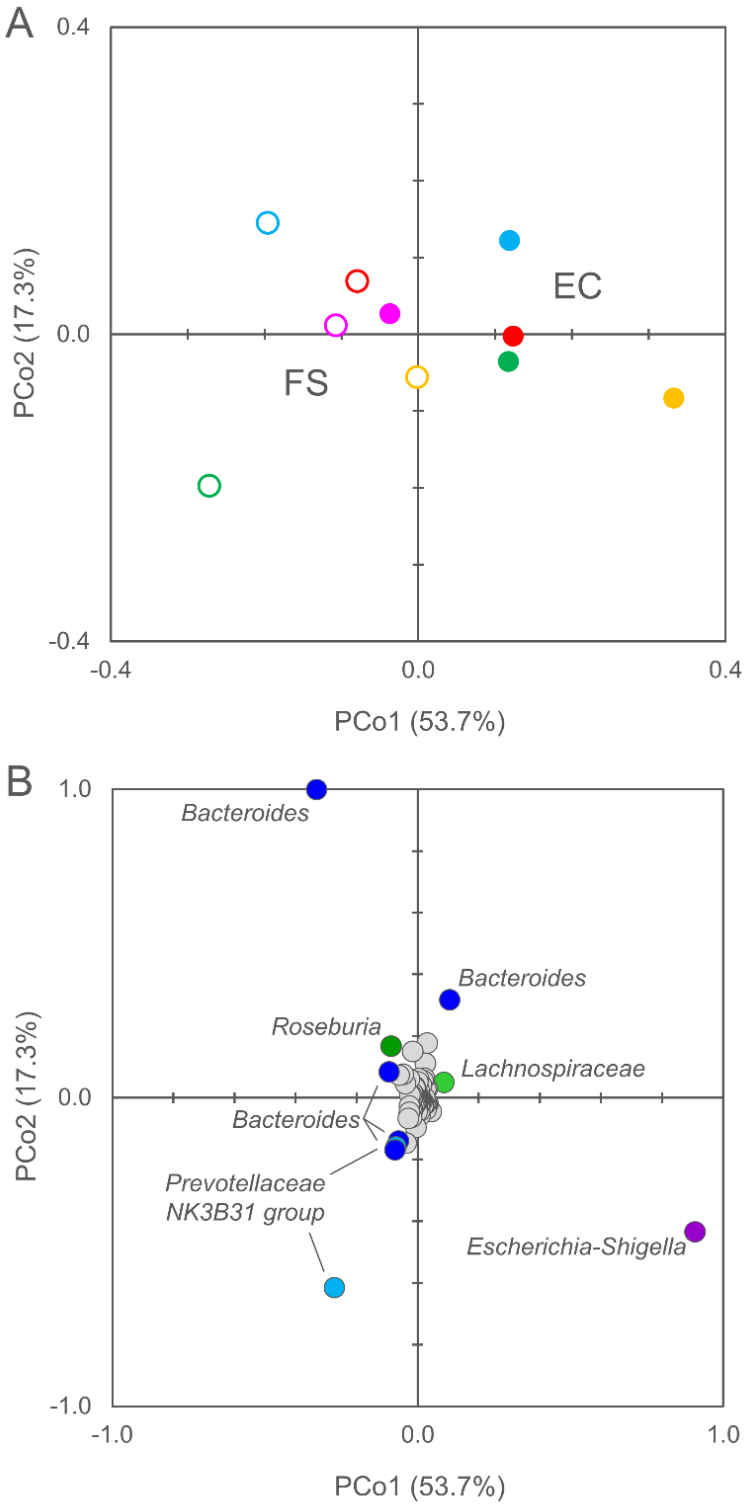


Figure 6.3 Beta-diversity analysis of the microbiota in FS and EC samples. Panel A, PCo1-PCo2 visualization of distances computed with Weighted Unifrac, that takes into account phylogenetic relationships among bacteria. Symbols: FS, empty circle; EC, full circle; different colors correspond to different subjects (1, fuchsia; 2, cyan; 3, green; 4, red; 5, yellow). Panel B, PCo1-PCo2 visualization of the contribution of each ASV. The 10 ASVs exerting the highest effect on PCo1 are labelled and colored according to their phylum (Firmicutes, green; Bacteroidetes, blue; Preteobacteria, purple).

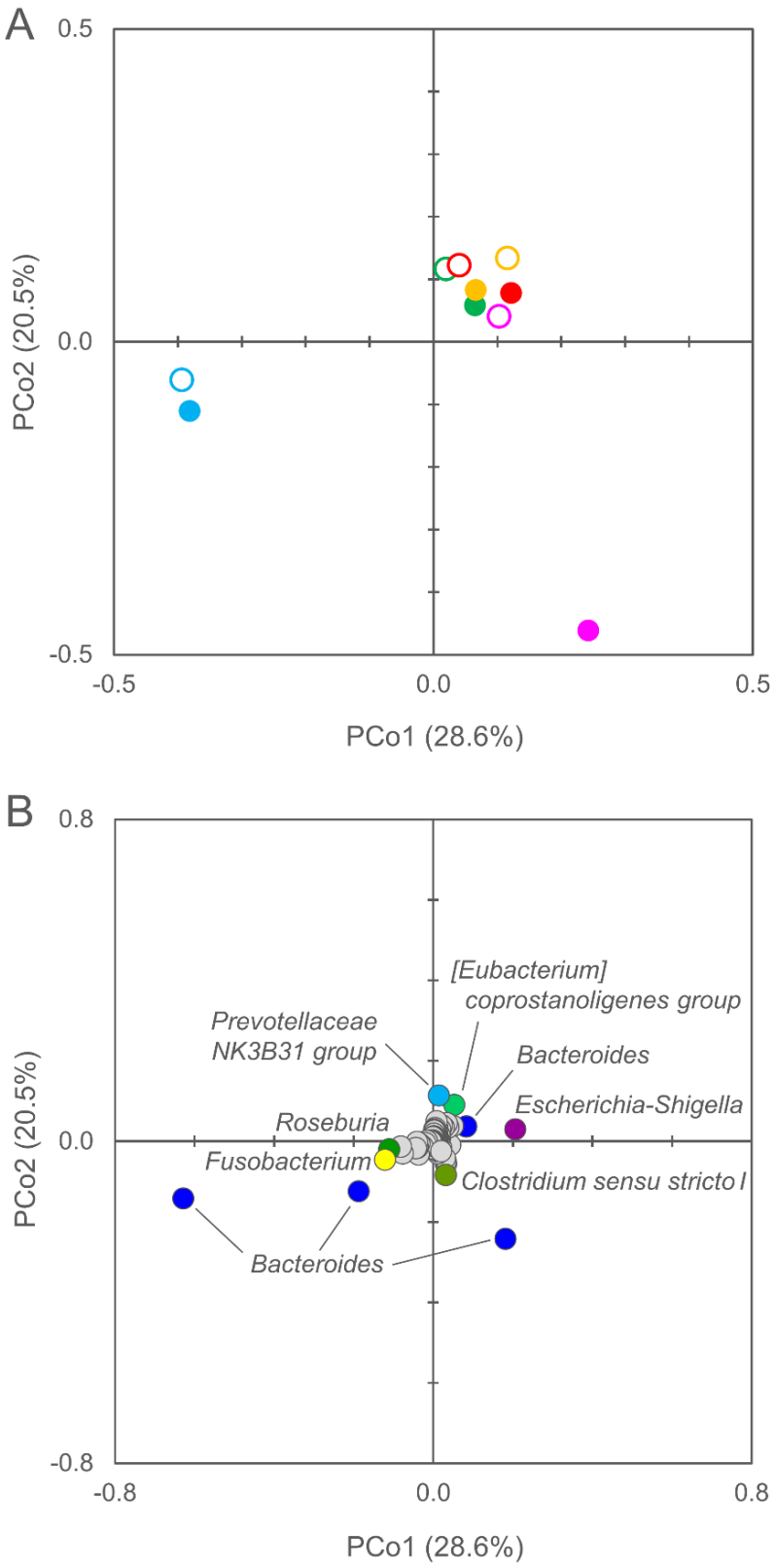


Figure 6.4 Beta-diversity analysis of the microbiota in FS and EC samples. Panel A, PCo1-PCo2 visualization of distances computed with Unweighted Unifrac. Symbols: FS, empty circle; EC, full circle; different colors correspond to different subjects (1, fuchsia; 2, cyan; 3, green; 4, red; 5, yellow). Panel B, PCo1-PCo2 visualization of the contribution of each ASV. The 10 ASVs exerting the highest effect on PCo1 are labelled and colored according to their phylum (Firmicutes, green; Bacteroidetes, blue; Preteobacteria, purple; Fusobacteria, yellow).

A number of taxa were not detected in any FS sample, but appeared after the enrichment: *Eggerthella* occurred in 4 EC samples; *Clostridium innocuum*, *Enterococcus durans*, *Erysipelatoclostridium*, and *Soleaferrea* in 3; *Angelakisiella* and *Eubacterium nodatum* in 2 samples. These taxa were generally a minor portion of the microbiota (< 1%), with the exception of *Clostridium innocuum* that reached the 7.3% in EC-2. *Enterorhabdus*, *Gordonibacter*, *Dielma*, *Enterococcus faecalis* and other *Enterococcus* spp., *Catabacter*, *Ruminiclostridium*, Lachnospiraceae FCS020 group, *Hydrogenoanaerobacterium*, *Caproiciproducens*, Ruminococcaceae DTU089, *Peptostreptococcus*, *Peptoniphilus*, *Tissierella*, *Sedimentibacter*, *Citrobacter*, and *Cloacibacillus* were detected in a single EC sample, generally in very low concentration, with the exception of *Dielma* (1%) and *Cloacibacillus* (1.5%).

Bacteroidota, Bacteroidia, and Bacteroidales, significantly decreased as consequence of mucin enrichment, albeit the genus *Bacteroides* remained the most abundant in 4 out of 5 EC samples. Also, the genus *Streptococcus* diminished, leading to significantly lower abundances also of Streptococcaceae. Numerous taxa of Clostridia decreased during the enrichment steps, such as *Eubacterium ventriosum*, the genera *Faecalibacterium*, *Fusicatenibacter*, *Blautia*, *Monoglobus* (plus Monoglobaceae and Monoglobales), and *Butyricicoccus* (and Butyricicoccaceae). Furthermore, the abundance of a clade of Lachnospiraceae (NK4A136) and Ruminococcaceae *incertae sedis* significantly diminished. Other taxa detected only in some FS samples decreased below the limit of detection in the EC samples, such as Coriobacteriales *incertae sedis*, other Eggerthellaceae, *Paraprevotella*, Defluviitaleaceae_UCG-011, *Eubacterium siraeum* group, *Veillonella*, *Butyricimonas*, Erysipelotrichaceae_UCG-003, *Turicibacter*, *Coprococcus*, *Lachnospira*, *Roseburia*, Anaerovoracaceae Family XIII UCG-001, and *Haemophilus*.

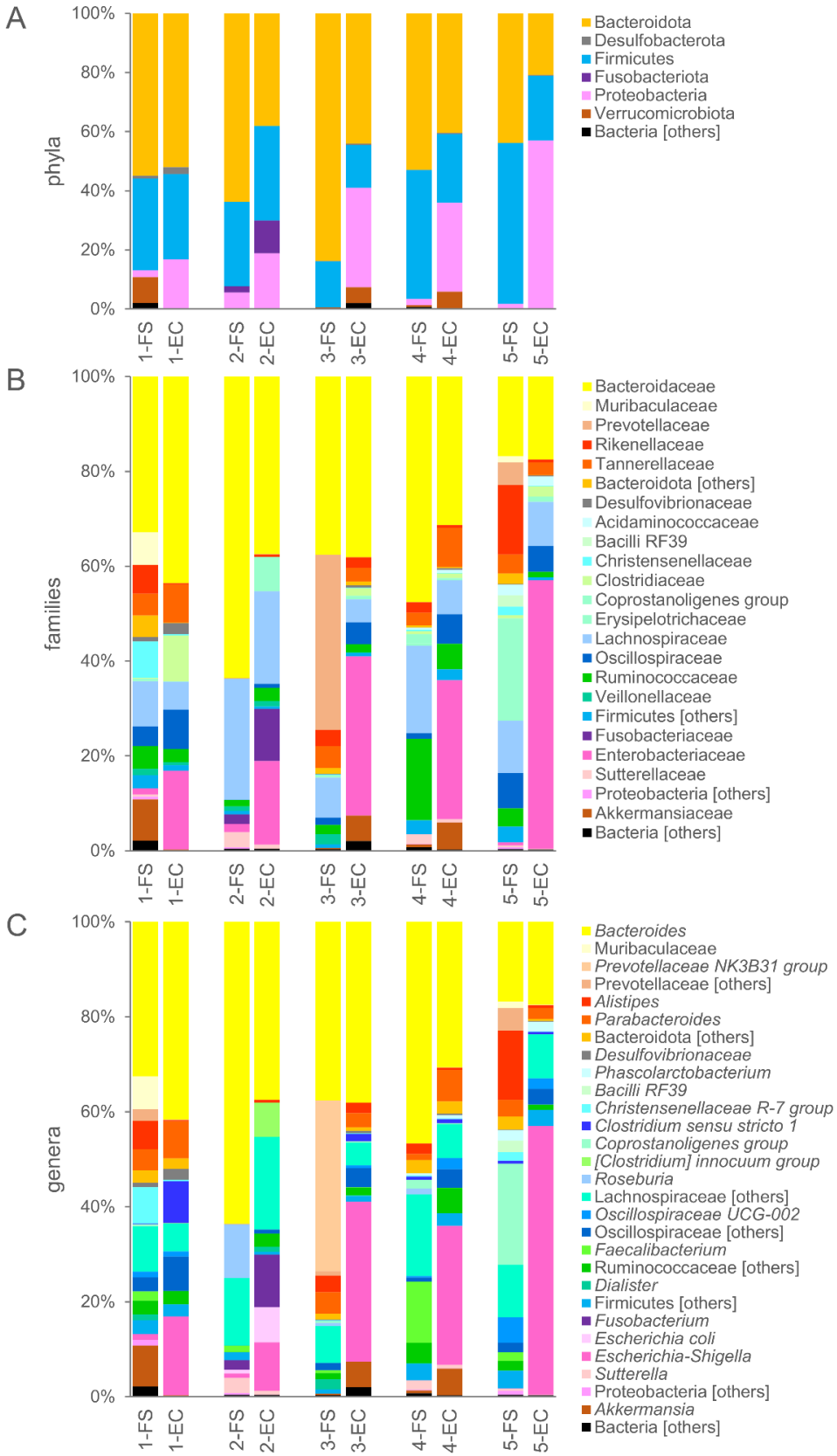


Figure 6.5 Stacked bar-plot representation of microbiota composition before and after the enrichment in MM, with taxonomic features collapsed at the level of phyla (panel A), families (panel B), and genera (panel C). The taxa that remained unclassified at the deeper level or that never occurred with abundance higher than 0.5% are grouped and marked with "[others]".

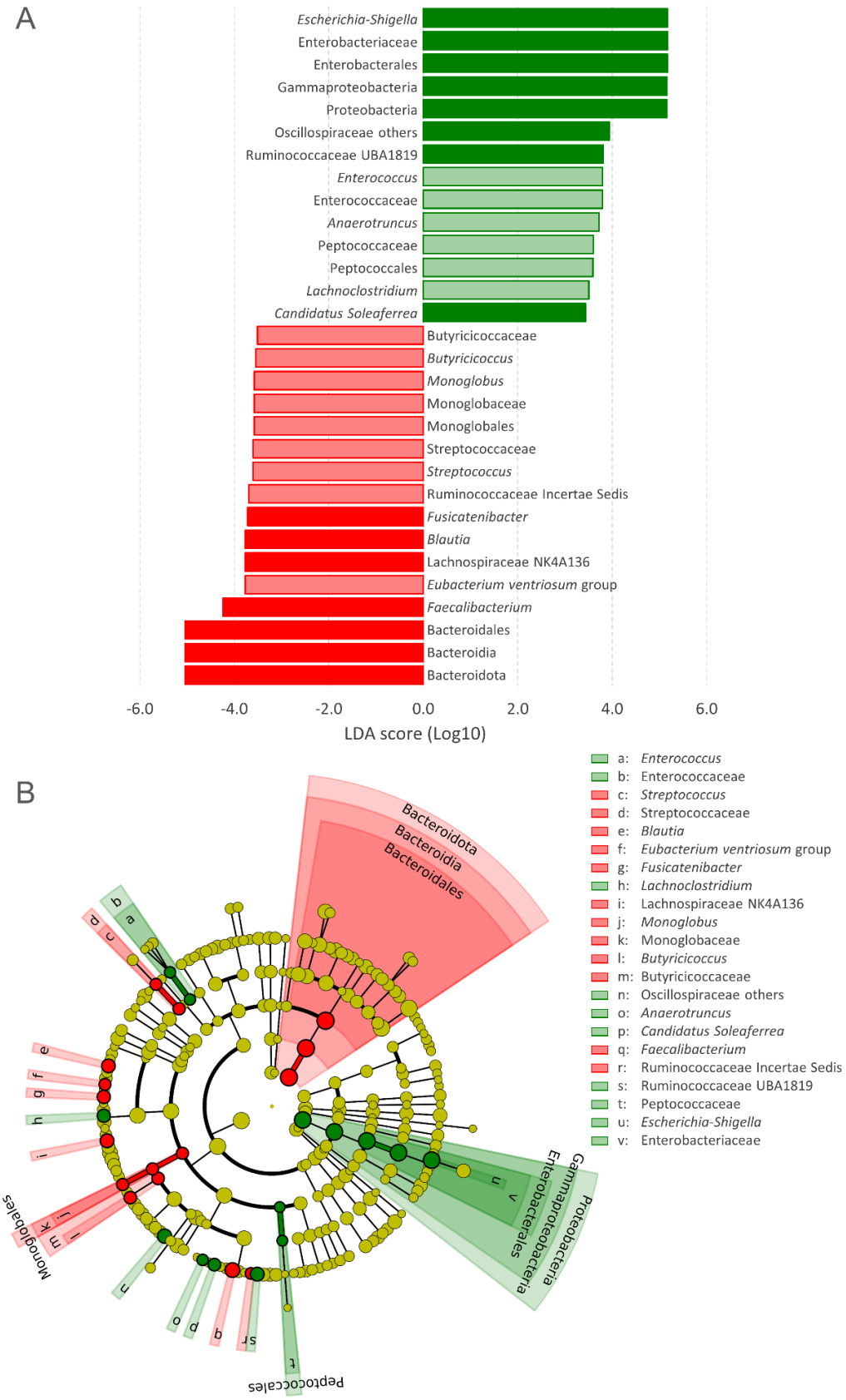


Figure 6.6 Effect of mucin enrichment on fecal microbiota. The taxonomic biomarkers characterizing FS and EC according to LEfSe are colored in red and green, respectively. Panel A, logarithmic LDA scores of the taxa exhibiting significant differential abundance ($P < 0.05$, logarithmic LDA score ≥ 2.0) between FS and EC; taxa that were > 1% in at least one sample are reported with bars in dark shades. Panel B, Cladogram visualization of the taxonomic biomarkers.

6.4 DISCUSSION

This study aimed to identify and isolate the intestinal bacteria that can have a role in mucin degradation, utilizing three steps of enrichement where the intestinal microbiota of healthy subjects was inoculated in a medium containing mucin as the sole carbon and nitrogen source. Mucin can support the growth of commensal microbes with different metabolic capabilities. Both saccharolytic and proteolytic activities are involved in its breakdown, providing the bacterial community with carbon and nitrogen supply and sustaining growth of both sugar and amino acid fermenters.

The 16S rRNA gene profiling highlighted the taxa that proliferated in mucin, likely as a result of cooperation and cross-feeding. Proteobacteria and taxa related to Enterobacteriaceae encompassed the main biomarkers of growth, in agreement with their major role as protein degraders in gut microbiota. They are opportunistic commensals that may persist in the gut contributing to the mucus homeostasis without interacting with epithelial cells (Troge et al. 2012; Raimondi et al. 2019b) or may degrade mucins, reach the epithelium, and initiate disease (Shawki ei al. 2016). The high affinity of Enterobacteriaceae for mucins confirms the possibility that they may get in contact with the intestinal epithelial barrier and exert a pathogenic action. In fact, these bacteria exhibit an intrinsic resistance to the oxygen released from the host tissue and may concur to disrupt the integrity of the epithelium, promoting a leaky gut and triggering systemic inflammatory response (Ugalde-Silva et al. 2016).

The protein backbone of mucin was likely degraded by Enterobacteriaceae, Peptococcales / Peptococcaceae, a variety of Clostridia (such as Oscillospiraceae, *Lachnoclostridium*, and *Anaerotruncus*), and the Coriobacteriale *Eggerthella*. These taxa are all known protein degraders that have been identified among those that got most abundantly enriched when the gut microbiota was cultured with proteins and peptides as sole fermentable susbtrates (Amaretti et al. 2019). *Lachnoclostridium* is positively associated to colorectal cancer, being significantly enriched in adenoma, and it has been proposed as a novel bacterial marker for the non-invasive diagnosis of colorectal adenoma and cancer (Liang et al 2020), while *Eggerthella* is emerging as an uncommon human pathogen (Bo et al. 2020). The genus *Anaerotruncus* may participate to the degradation of both the protein and saccharidic portions

of mucins. In fact, the species *Anaerotruncus colihominis* and *Anaerotruncus massiliensis* can ferment a small set of carbohydrates, including N-acetyl-glucosamine, a major component of mucin, and some amino acid (Lau et al. 2006; Togo et al. 2019). Also *Holdemania*, a Firmicutes belonging to the family Erysipelotrichaceae, and Enterococcaceae (Lactobacillales) significantly increased, likely fermenting the carbohydrate chains. Enterococcaceae are saccharolytic facultative anaerobes and can behave as virulent and pathogenic bacteria (García-Solache and Rice 2019**)**. Several other taxa, which are still poorly characterized at the physiological level, emerged as participating in mucin catabolism, for some species representing one of the first sources of functional information.

In the gut, a complex community of saccharolytic and proteolytic bacteria is expected to participate to mucin breakdown, with linkage-specific glycosidases and proteases of primary degraders providing other bacteria with simpler oligosaccharides and peptides, and eventually with their monomeric moieties. It is plausible that several taxa detected by the 16S rRNA gene analysis could have grown on mucins only when co-cultured within a bacterial community where each species contributed with different hydrolytic activities, while they could not grow in pure cultures where they did not benefit from cross-feeding relationships.

To date, *Bacteroides* and *Akkermansia muciniphila* have been recognized as main mucin degraders, with novel insights on endo-acting O-glycanases involved in the first steps of glycan breakdown (Png et al. 2010; Crouch et al. 2020). In this study, the relative abundance of *Bacteroides* did not increase significantly, even if some ASVs of *Bacteroides* heavily contributed to separate samples along both PCo1 or PCo2 in the Weighted UniFrac analysis, confirming the presence of some species greatly associated to mucin degradation. It is plausible that different populations of *Bacteroides*, which could not be discriminated into species by the partial 16S rRNA gene sequencing, behaved differently with respect to mucin breakdown, exerting contrary effects and balancing the counts of the genus. Indeed, a wide number of ASVs attributed to *Bacteroides* presented opposite contributions along the principal coordinates, in agreement with the fact that the genus *Bacteroides* comprises tens of species that supply different and complementary functions to the intestinal community (Wexler 2007).

One of the key members of the colonic mucus-associated microbiota is *Akkermansia muciniphila*. It releases monosaccharides and amino acids during mucin degradation, providing nutrients to other bacteria of the gut, and exerts beneficial effects in various metabolic disorders, being generally negatively correlated with inflammation and metabolic

disorders (Zhou and Zhang 2019). In this study, a sole ASV ascribed to *Akkermansia* was detected in 3 out of 5 set of samples. In subject 1, it was remarkably abundant in the founding microbiota (8.7%) but was negligible (0.2%) after the enrichment steps. In samples 3 and 4, *Akkermansia* was initially < 0.6% and increased by one magnitude over the enrichment. It is plausible that in sample 1 the high initial load of *Akkermansia* prevented a net increase of the abundance, that was further squeezed by major increase in some other groups.

The absence of *Bacteroides* and *Akkermansia* also within the isolates is noteworthy. These bacteria likely encountered some substrate limitation that hampered growth. In fact, the medium did not contain nitrogen sources other than mucin in order to be as much stringent as possible. Thus, all the amino acids had to be obtained from the protein backbone of mucin, and this might have been too challenging, in the absence of a cooperative crossfeeding. The composition of the medium and also the choice of pH are major drivers of microbial selection and a mucin based medium with different composition would likely result in the isolation of different bacterial species. Moreover, in the present study, a possible drawback of the experimental design was the use of mucin from porcine stomach, that differs from the colonic one in terms of low purity and glycosylation profile (Holmén Larsson et al. 2013; Corfield et al. 2018).

6.5 CONCLUSIONS

This study shed light on the bacterial taxa thriving in cultures of human microbiota with mucin as primary substrate for growth. Studies on the isolation and whole genome sequencing of strains from mucin-based medium are in progress, as well as the formulation of a new mucin-based medium allowing a faster growth with higher yields, in order to better address physiology and metabolism studies. Proteobacteria, including Enterobacteriaceae and *E. coli*, are among the microbial group that most benefit from mucin as substrate, increasing the risk of the flourishing of opportunistic pathogens or bacteria associated to gut inflammation. Furthermore, metataxonomic analysis identified *Lachnoclostridium*, *Eggerthella*, *Anaerotruncus*, *Enterococcaceae*, and *Peptococcaceae* as taxa that took advantage, likely in a cross-feeding relationship, in the mucin-based medium.

These results can be found in the paper “Mucin degraders of human gut microbiota:

16S rRNA gene profiling and isolation of mucinolytic strains”, currently under review at *Gut Microbes*

Stefano Raimondi[1, †], Eliana Musmeci[1, †], **Francesco Candeliere**[1], Alberto Amaretti[1,2], Maddalena Rossi[1,2]

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy

[2] Biogest-Siteia, University of Modena and Reggio Emilia, Modena, Reggio Emilia, Italy

† The authors contributed equally to this work

7. Project 4

16S rRNA metagenomic profiling: the microbiota of *Hermetia illucens* larvae

7.1 INTRODUCTION

Insects are promising protein sources for livestock and human diet and may represent a sustainable contribution for feeding the growing world population. Larvae and prepupae of *Hermetia illucens* (i.e. black soldier fly, BSF) are receiving increasing interest as potential food and feed source because of the high nutritional value and the low environmental impact of the rearing process (Wang and Shelomi, 2017). In particular, insect meal of BSF was successfully utilized as alternative protein source for poultry and aquaculture (Cutrignelli et al., 2018; Terova et al., 2019; Belghit et al., 2019; Zarantoniello et al., 2019).

The utilization of mature larvae and prepupae in insect-based ingredients (e.g. powders, flours, protein bars, pasta, burgers, and nuggets) requires the development of cheap practices to produce and stabilize the insect biomass, providing a safe starting bulk product (Truzzi et al., 2020; Bruni et al., 2018). The ability of *H. illucens* to grow on a variety of solid organic matrices can be exploited to transform and valorize organic streams such as by-products of the agroindustry, livestock manures, or urban solid wastes, reducing environmental pollution and converting organic wastes into biomass rich in protein and fat (Barragan-Fonseca et al., 2017; Lalander et al., 2015; Nguyen et al., 2015; Bortolini et al., 2020).

The possibility to rear *H. illucens* on different waste streams in a biorefinery approach requires attention to safety issues, both chemical and microbiological. The European Food Safety Authority reported the lack of data regarding microbiology, virology, parasitology, and toxicology of insects reared, addressing the potential hazards of non-processed insects in comparison with other non-processed sources of protein of animal origin (EFSA, 2015). Hygiene issues of edible insects can originate from the substrate but also from the microbial community of the insect gut and may be affected by processing steps linking farming and consumption.

7.2 AIM OF THE PROJECT

Recent works reported microbiota composition of BSF larvae and prepupae (Terova et al., 2019; De Smet et al., 2018; Osimani et al., 2018; Wynants et al., 2018). However, little attention has been paid to microbial dynamics associated to extrinsic parameters, such as the rearing temperature. In a bioeconomy perspective, the biotransformation of wastes toward new valuable compounds should be a low-energy process that can be exploited in different districts with a good tolerance to temperature changes, minimizing the requirement of heating or cooling.

From this standpoint, BSF larvae were reared at different temperatures, and since this parameter may be a main driver of microbial population composition, the microbiota of prepupae reared at 20, 27, and 33 °C was characterized. To avoid biases due to the intrinsic different microbial communities of feed, the larvae were fed a basal substrate (Bruno et al., 2019; Jeon et al., 2011; Zheng et al., 2013). A survey of the 16S rRNA gene was utilized to determine the microbiota composition in samples of substrate, initial larvae, grown prepupae, and their frass (hereinafter referred to as S, L, PP, and F samples). The impact of rearing temperature on microbiota is expected to offer knowledge and tools to prevent microbiological hazards, protect the health and the welfare of the consumers, and promote the development of safe trade in food and feed products.

7.3 MATERIALS AND METHODS

7.3.1 *Hermetia illucens* rearing

The BSF larvae used in this study were obtained from a colony kept at the Applied Entomology Laboratory, BIOGEST-SITEIA (Reggio Emilia, Italy), originally established with prepupae purchased from CIMI srl (Cuneo, Italy). BSF larvae were routinely reared by the BIOGEST-SITEIA staff in a standard vegetable substrate (S), composed of 25% zootechnical use corn flour (mill waste), 15% wheat bran, 10% alfa-alfa flour, and 50% water (Hogsett, 1992). The experiments were set up using second to third instar larvae that were obtained from eggs hatched in S and incubated in climatic chambers under controlled conditions of 27 ± 0.5 °C, 65% humidity, and 16:8 light:dark cycles.

To compare the rearing temperatures, 100 small larvae (L) were collected and reared in glass containers containing 400 g of S within climatic chambers at 20 ± 0.5, 27 ± 0.5, or 33 ± 0.5 °C. Two independent experiments, starting with different batches of L, were carried out, where each rearing temperature was tested in triplicate. At the end of the development period, when a minimum of 90% of the initial larvae became prepupae (PP), the experiment was considered concluded and the insects were collected from the frass (F).

7.3.2 DNA extraction

For each rearing batch, triplicate samples of L, S, $PP_{20°C}$, $PP_{27°C}$, $PP_{33°C}$, and $F_{20°C}$, $F_{27°C}$, and $F_{33°C}$ were pooled in equal weights. Approximately 2 g of material were 10-fold diluted in PBS and were homogenized by Ultra Turrax. Total DNA was isolated from the suspensions using the DNeasy Mericon Food Kit (Qiagen, Hilden, Germany), following the manufacturer's standard protocol. The DNA was normalized to 5 ng/µL after quantification with a Qubit 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA).

7.3.3 Sequencing

Partial 16S rRNA gene sequences were amplified using Probio_Uni / Probio_Rev primers, which targeted the V3 region of the 16S rRNA gene. Amplicons were sequenced using a MiSeq (Illumina) platform according to Milani et al., 2013. The 16S rRNA gene sequences are available at NCBI repository with the BioProject ID: PRJNA573042.

7.3.4 16S rRNA gene profiling

Quality filtered sequences were processed with QIIME2 pipeline (Version 2019.1) for closed-reference picking of amplicon sequence variants (ASVs), taxonomy assignment, collapsing into operational taxonomic units (OTUs) (Bokulich et al., 2018; Bolyen et al., 2018). Closed-reference picking and taxonomy assignation were carried out utilizing as reference SILVA SSU database release 132 (https://www.arb-silva.de/download/arb-files/) with the similarity threshold set at 0.97. The appropriate QIIME2 plugins were utilized to compute the alpha (observed taxa, Chao1, Shannon, and Pielou's evenness) and beta-diversity (Jaccard, Bray-Curtis, Canberra, Unweighted UniFrac, and Weighted UniFrac) and to compare them within and between groups of samples (i.e. the Kruskal-Wallis test for alpha diversity; ANOSIM and PERMANOVA for beta-diversity). Beta-diversity distance matrices were utilized for the Principal Coordinate Analysis (PCoA), using QIIME2.

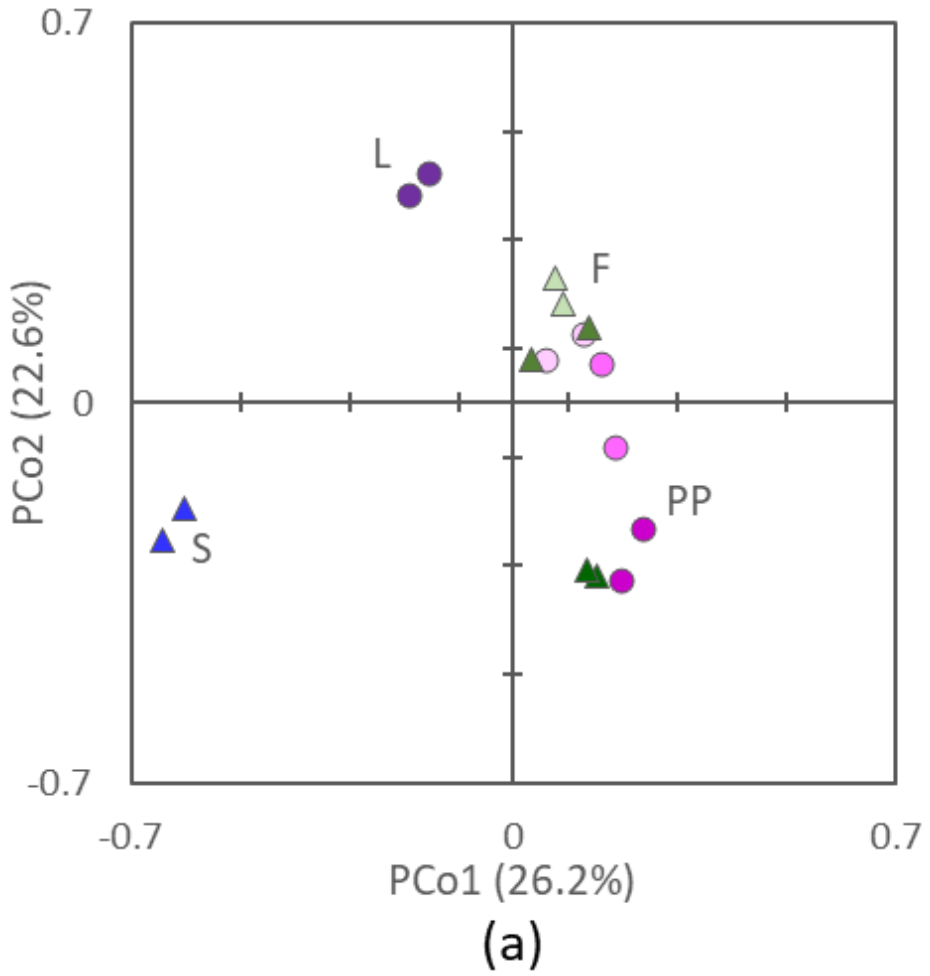
Linear discriminant analysis Effect Size (LEfSe, http://huttenhower.sph.harvard.edu/galaxy) algorithm was applied to discover distinctive taxonomic features characterizing the groups of samples (Segata et al., 2011). In the analysis of the taxa characterizing the development stage, L and PP samples were entered as 'classes'. In the comparison between frass and insects, F and PP were entered as 'classes'. Spearman's rank correlation was applied to identify the taxa in PP that positively or negatively correlated with the three temperatures. In both LEfSe and Spearman's analysis the alpha value for statistical significance was set at 0.05.

7.4 RESULTS

The 16S rRNA gene profiling of L, PP, S, and F yielded a total of 755,313 sequences, that were dereplicated into 3974 ASVs hitting a reference sequence in SILVA database, and collapsed at the 7th level of taxonomic annotation (i.e. the species, if available) into 536 OTUs. Based on Bray-Curtis distance, the microbiota of L, PP, S, and F clustered in distinct groups (Figure 7.1A) that had different centroids (PERMANOVA, $P = 0.001$) and intra-group permutational similarity significantly greater than the inter-group one (ANOSIM, $P = 0.001$). F represented an exception, the similarity within the group being comparable to that between

F and PP (P > 0.05). Grouping was evident in the PCo1-PCo2 plot (describing the 49% of total variance), where F and PP were characterized by positive PCo1, L by positive PCo2, and S by negative PCo1. Both F and PP lay at progressively lower PCo2 with the increase of the rearing temperature. Main taxa positively contributing to PCo1 were *Myroides*, *Alcaligenes faecalis*, *Bacillus*, and *Morganella*, while Enterobacteriaceae weighted negatively. Positive PCo2 was strongly owed to *Providencia*, followed by *Klebsiella* and *Myroides*, and negative to *Bacillus*, *Pseudomonas*, and *A. faecalis* (Fig. 7.1B).

The richness of the microbiota (evaluated as total no. of OTUs and Chao1 index) was similar in all the groups of samples, regardless of the rearing temperature (Figure 7.2). The evenness (estimated in Shannon and Pielou metrics) was the highest in S samples, followed by PP and F, grouped regardless of the rearing temperature, and the lowest in L (Figure 7.2).



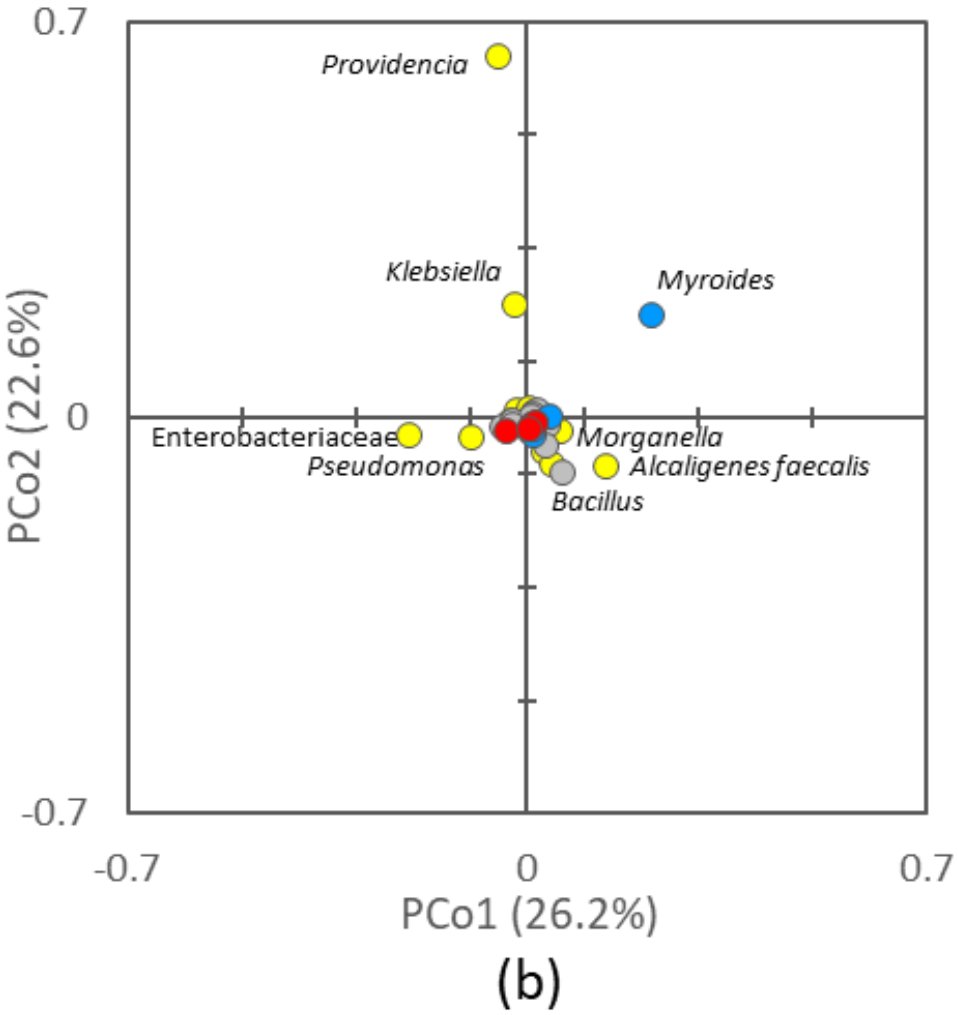


Figure 7.1 PCoA visualization of microbiota beta-diversity (Bray Curtis) among L, S, PP, and F: (a) scores of L (purple circle), S (blue triangles), PP (pink circles, with 20, 27, and 33°C from the lightest to the darkest), and F (pink triangles, with 20, 27, and 33°C from the lightest to the darkest), and F (pink triangles, with 20, 27, and 33°C from the lightest to the darkest), (b) contribution of the single bacterial taxa (Proteobacteria, yellow; Bacteroidetes, blue, Firmicutes, grey; Actinobacteria, red). Labels indicate taxa with the greatest contribution along PCo1 and PCo2.

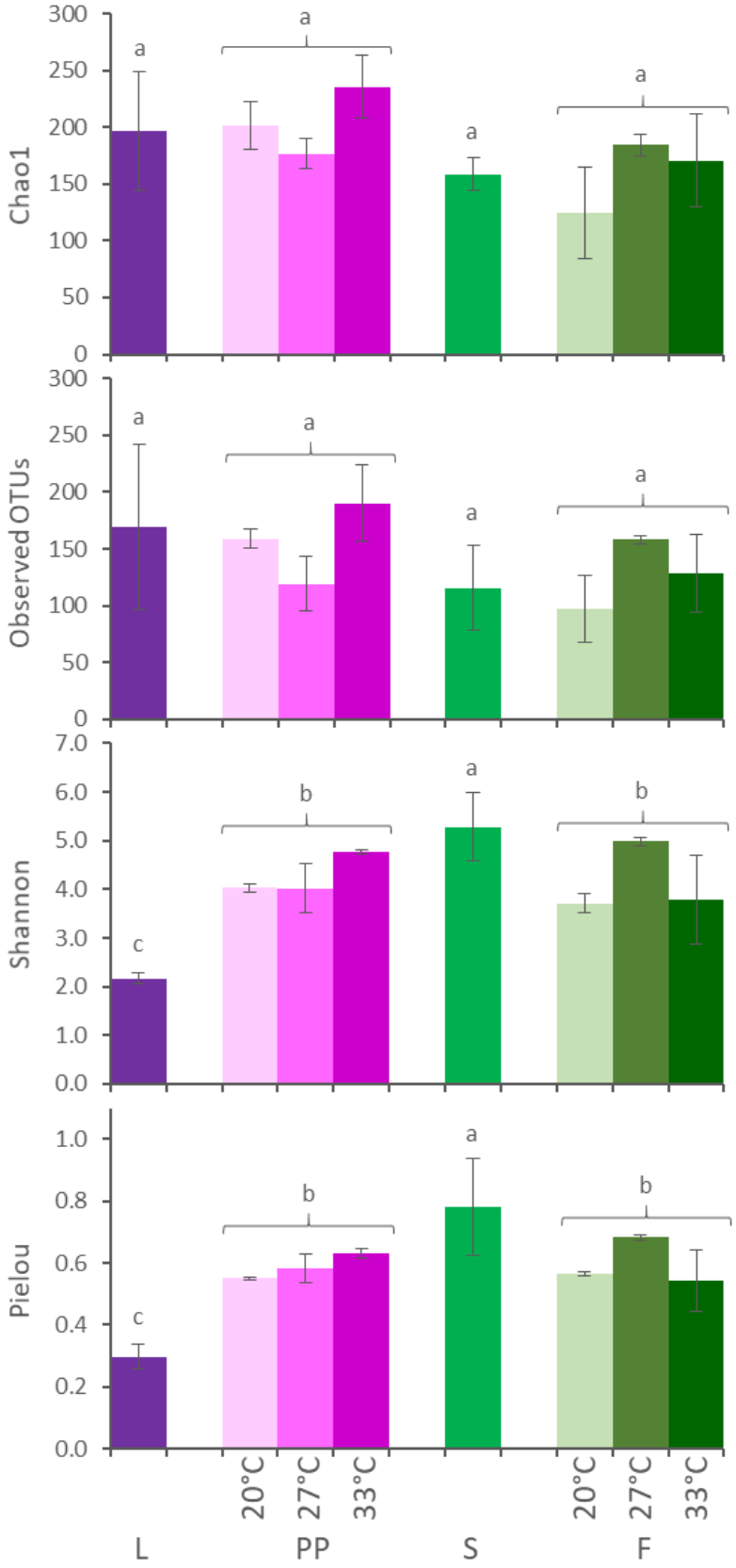


Figure 7.23 Alpha diversity metrics of the microbiota in L, PP, S, and F. Groups significance: within each panel, values sharing the same letter do not significantly differ (Kruskal-Wallis, P > 0.05); PP and F were grouped regardless of the rearing temperature.

A reduced dataset of 181 OTUs, occurring for more than 0.2% in at least one sample, still represented 99% of the reads. The distribution of the OTUs in the groups of samples is reported in the Venn diagram of Fig. 7.3. The bacterial composition of L, PP, S, and F is reported in Figure 7.4, showing the mean abundance of the main bacterial groups in the two experiments. The two experiments were not averaged for the analysis of the differential abundance of bacterial groups and the effect of temperature.

The microbiota of L encompassed 98 of the 181 OTUs. It was dominated by Proteobacteria (87.4%), with remarkably high amounts of *Providencia* (64.1%) and *Klebsiella* (16.4%), followed by Firmicutes (9.3%), especially *Bacillus* (4.3%). The microbiota of PP included 156 OTUs, 95 of which shared with L. PP were poorer than L in Proteobacteria and richer in Bacteroidia, Actinobacteria, and Bacilli (Figure 7.4, 7.5). *Providencia* remained abundant in PP (from 6.4 to 26.1%, depending on the temperature), although it significantly decreased compared to L, likewise *Klebsiella* and other minor Enterobacteriaceae. The genera *Morganella*, *Alcaligenes*, *Bordetella*, and *Kerstersia* behaved in contrast to most Proteobacteria and were significantly enriched in PP. *Morganella* and *Alcaligenes*, in particular, reached remarkable levels (8.9 and 16.9%, respectively). The significant increase of Bacteroidia in PP was mainly due to *Myroides*, a Flavobacteriaceae that reached 30%. Bacilli also included several biomarkers characterizing PP, such as Planococcaceae, Paenibacillaceae, *Pseudogracilibacillus*, *Oceanobacillus*, and unclassified members of the genus *Bacillus*. Planococcaceae, *Pseudogracilibacillus*, and unclassified *Bacillus* reached 8.5, 5.9, and 13.1%, respectively. Within Actinobacteria, the main biomarkers characterizing PP were Micrococcales belonging to *Brevibacterium*, which increased up to 3.3%, and several minor genera.

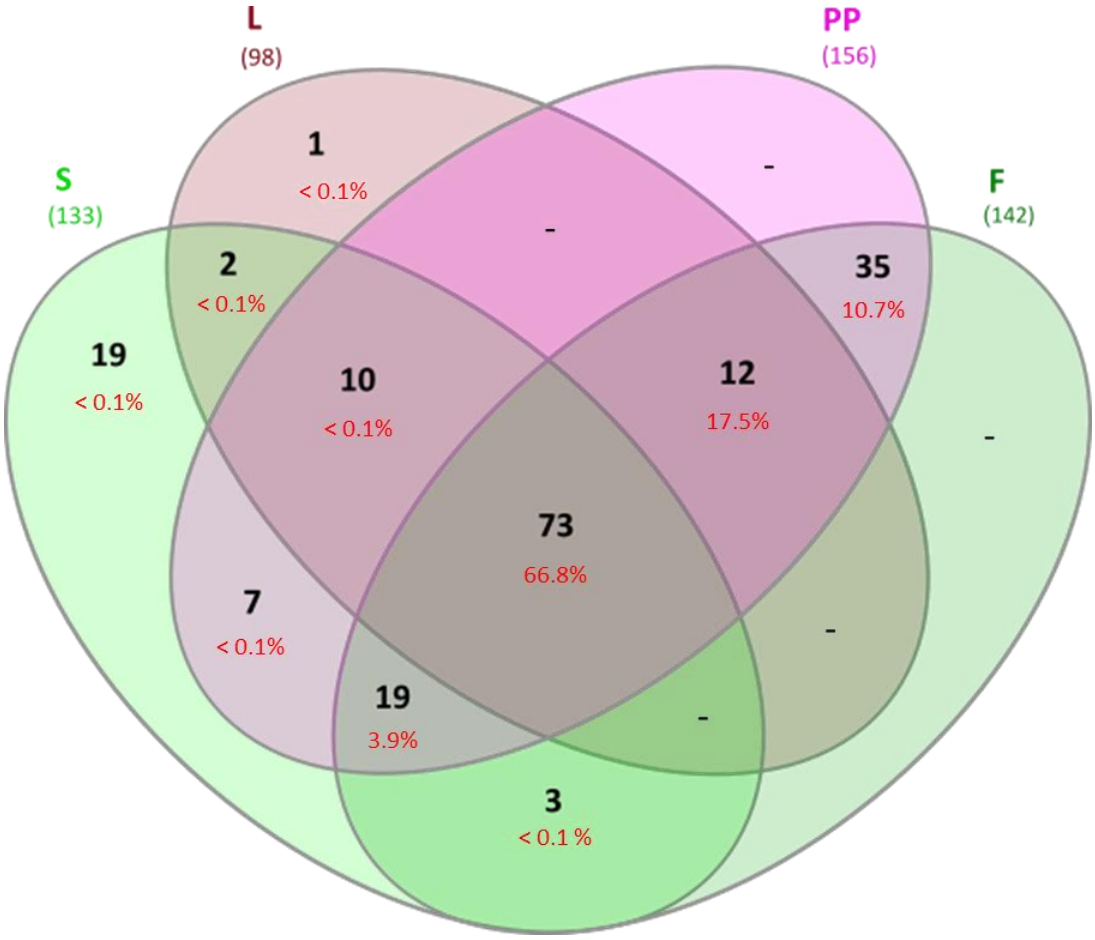


Figure 7.3 Distribution in the main 181 OTUs (accounting for more than 0.2% in at least one sample) in L, PP, S, and F samples. For each intersection, the number of OTUs and the percentage relative to the total reads are reported in black and red, respectively

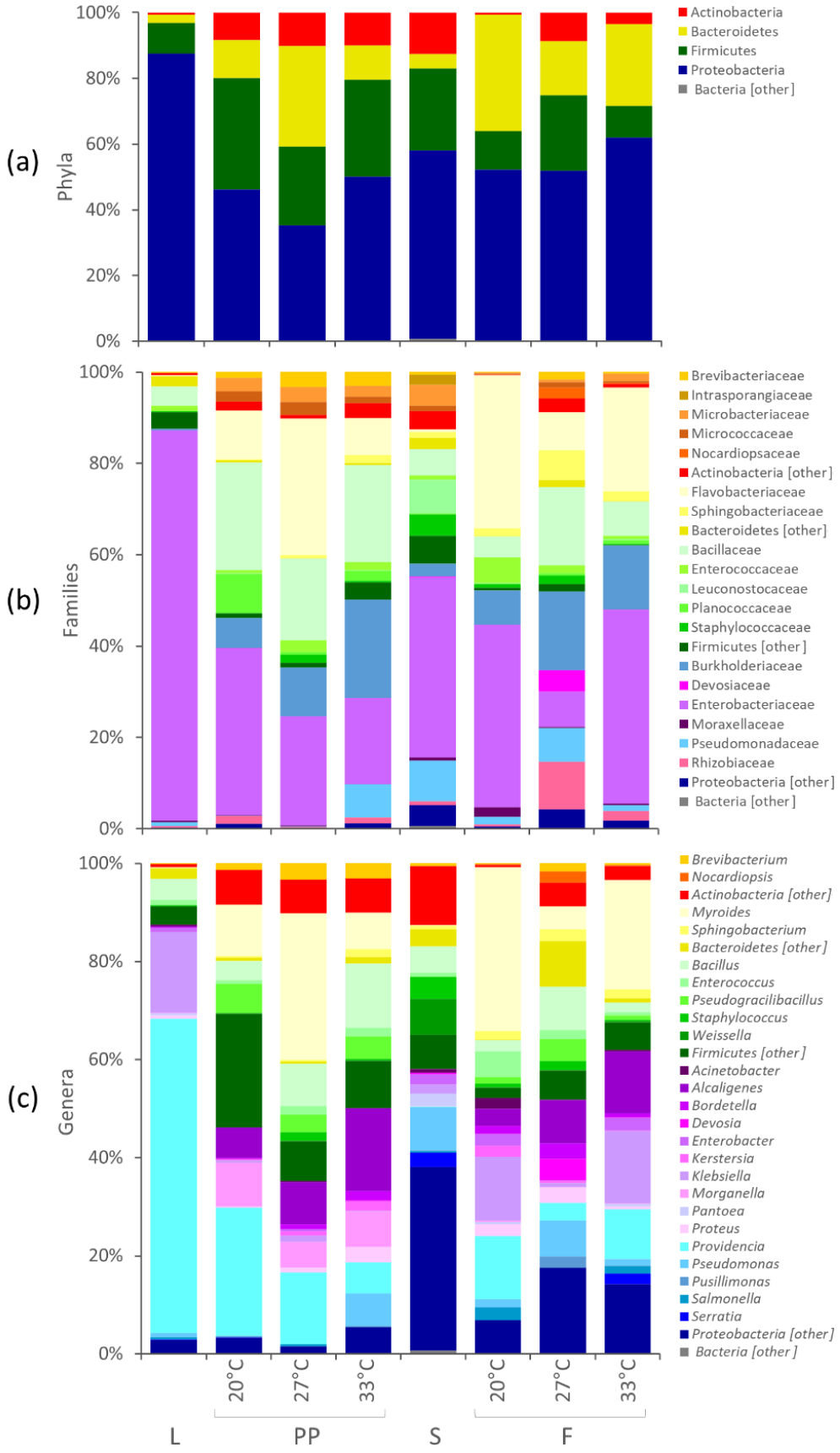
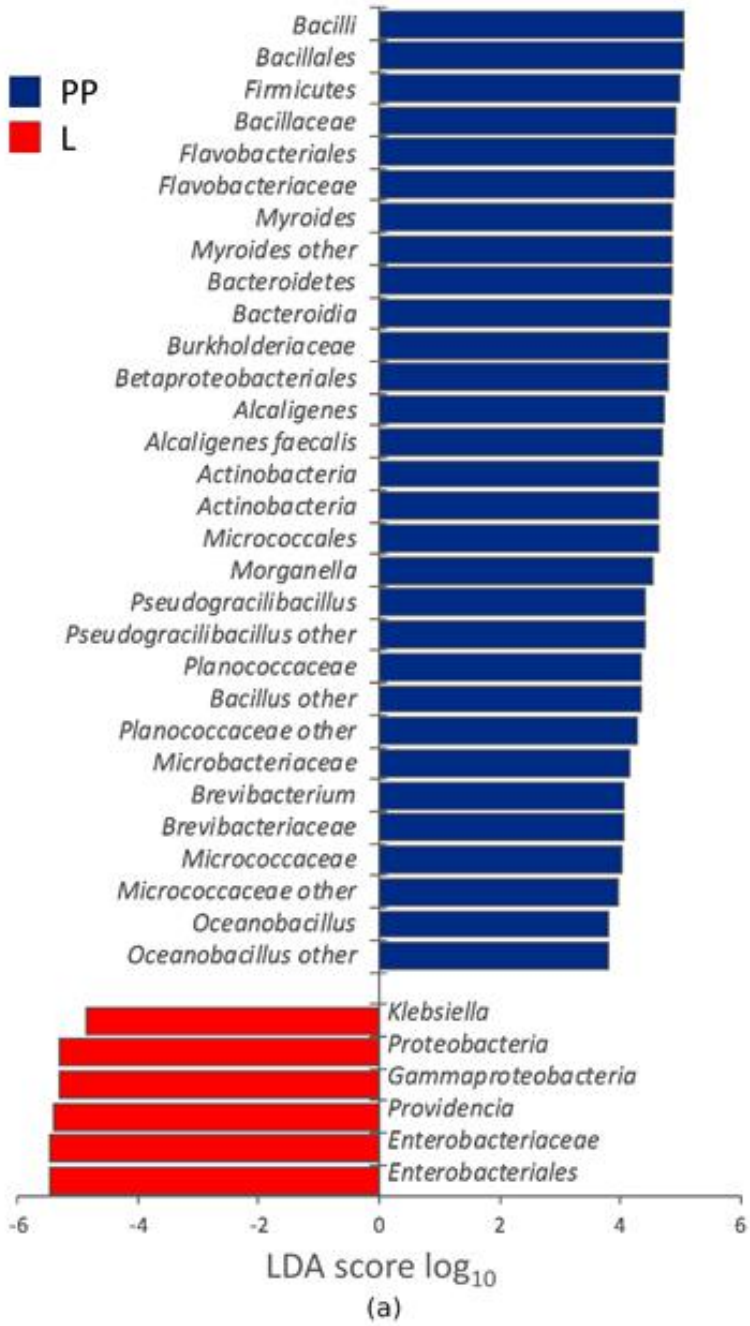
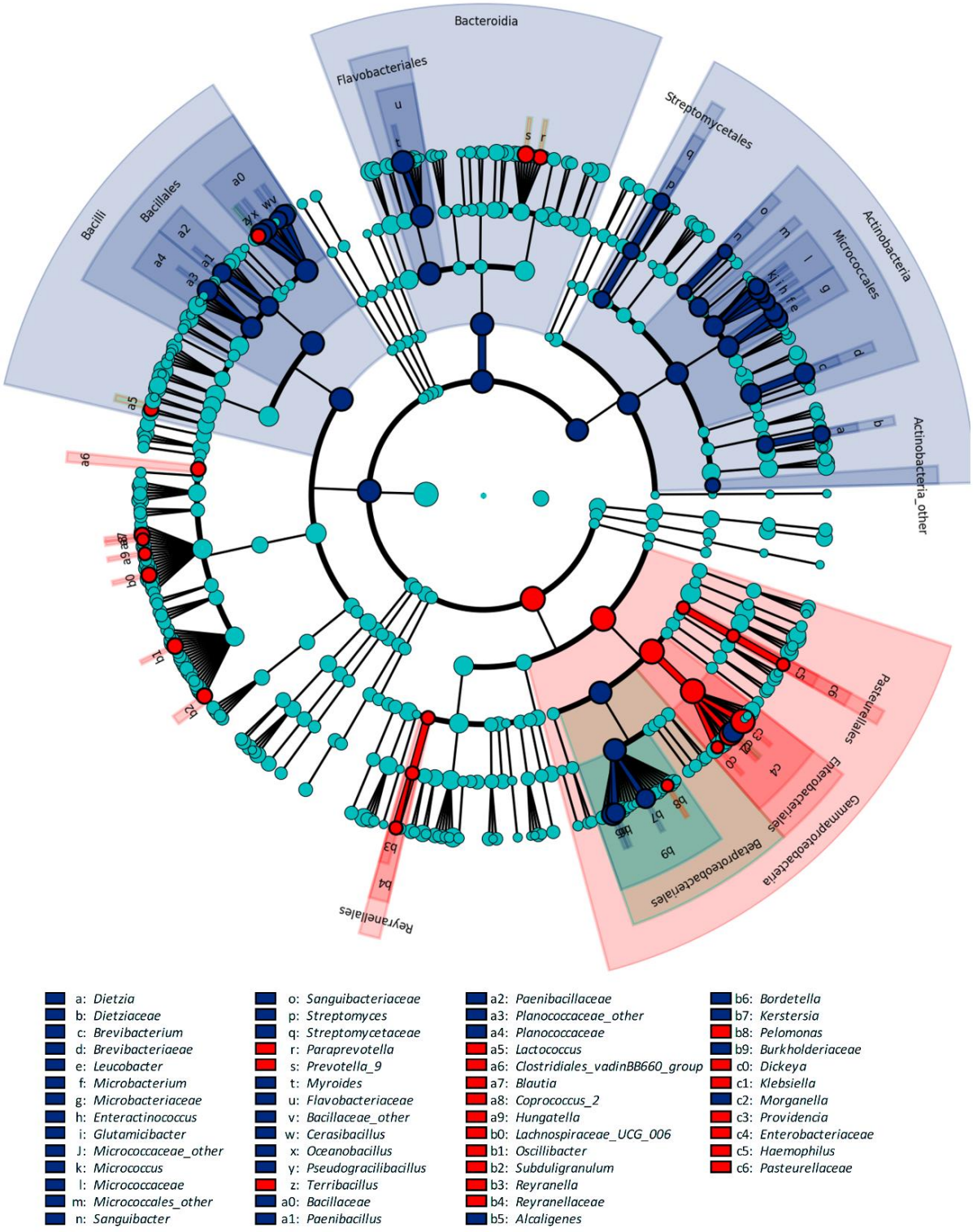


Figure 7.4 Stacked bar-plot representation of microbiota composition in L, S, PP, and F with taxonomic features collapsed at the level of phyla (a), families (b), and genera (c). The mean values of the two independent experiments are reported. The phyla, families, and genera that remained unclassified or never occurred with abundance > 2.0% are grouped as others.


PP
L
Bacilli
Bacillales
Firmicutes
Bacillaceae
Flavobacteriales
Flavobacteriaceae
Myroides
Myroides other
Bacteroidetes
Bacteroidia
Burkholderiaceae
Betaproteobacteriales
Alcaligenes
Alcaligenes faecalis
Actinobacteria
Actinobacteria
Micrococcales
Morganella
Pseudogracilibacillus
Pseudogracilibacillus other
Planococcaceae
Bacillus other
Planococcaceae other
Microbacteriaceae
Brevibacterium
Brevibacteriaceae
Micrococcaceae
Micrococcaceae other
Oceanobacillus
Oceanobacillus other
Klebsiella
Proteobacteria
Gammaproteobacteria
Providencia
Enterobacteriaceae
Enterobacteriales
-6
-4
-2
0
2
4
6
LDA score log10
(a)



(b)

Figure 7.5 4 LEfSe analysis of taxonomic features differentiating L (n = 2) and PP (n = 6), regardless of the growth temperature: (a) LDA logarithmic scores of taxonomic biomarkers exhibiting significant differential abundance ($P < 0.05$, LDA logarithmic score ≥ 2.0) and appearing at least once with abundance > 2.0%; (b) cladogram of all the taxonomic biomarkers.

The microbiota of S was characterized by 133 OTUs, 85 of which shared with L and 109 shared with PP. The remaining OTUs represented 6.3% of S microbiota, with only *Rahnella* being > 1%. The microbiota of F encompassed 142 OTUs, 139 of which shared with PP. The composition of F differed from that of PP for being significantly richer in Proteobacteria (such as *Klebsiella* and *Acinetobacter*) and poorer in Actinobacteria (mainly Micrococcales, such as *Brevibacterium*), Firmicutes (Bacillales) and the Enterobacteriaceae *Morganella* (Figure 7.6). The increase of incubation temperature negatively correlated with the level of Enterobacteriaceae in PP, mostly due to the decrease of *Providencia* (Figure 7.7).

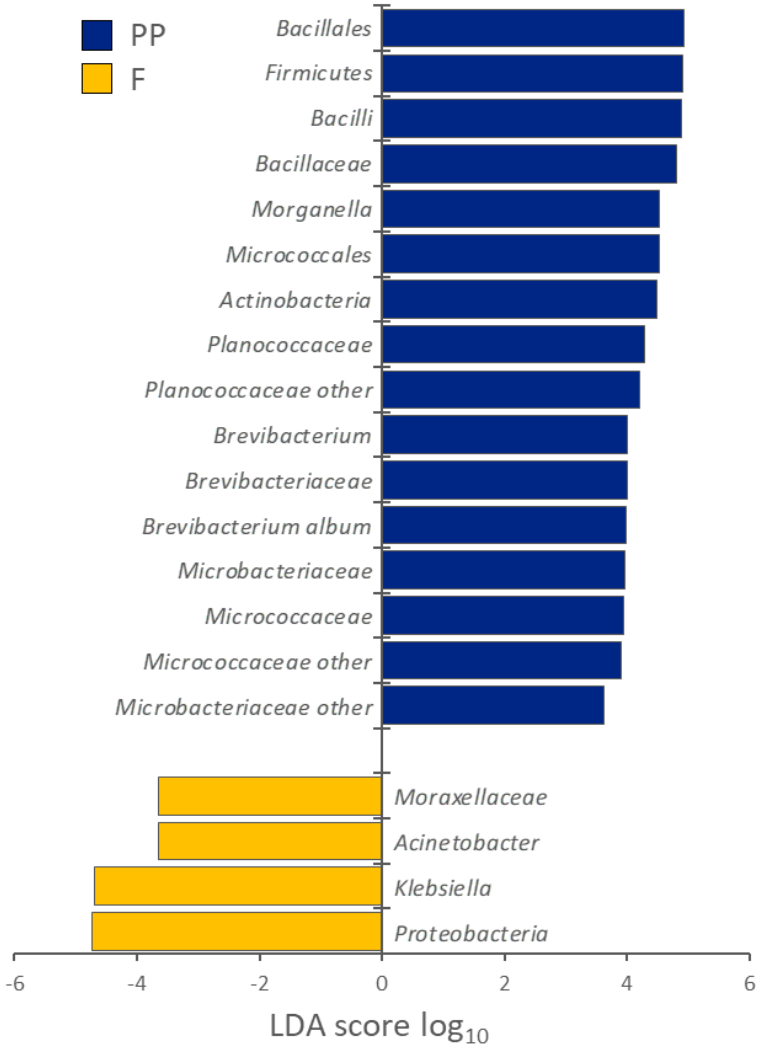


Figure 7.6 LEfSe analysis of taxonomic features differentiating F and PP (both n =6), regardless of the growth temperature. The plot reports LDA logarithmic scores of taxonomic biomarkers exhibiting significant differential abundance ($P < 0.05$, LDA logarithmic score ≥ 2.0) and appearing at least once with abundance > 2.0%.

Unlike *Providencia*, other Proteobacteria positively correlated with temperature, such as *Kertersia*, *Proteus*, *Pseudomonas*, *Alcaligenes*, and *Bordetella*. *Pseudomonas* was abundant only in PP reared at 33 °C (6.7%), whereas at the lower temperatures it accounted for less than 0.1%. *Bacillus* was also positively associated with increasing temperature and reached 13.1% at 33 °C. Increasing temperatures also favored the populations of other Actinobacteria (*Brevibacterium*), Bacilli (e.g. *Enterococcus*, *Pediococcus*, *Paenibacillus*, *Pseudogracilibacillus*), and Bacteroidetes (*Sphingobacterium*), although occurring in low amounts. The correlation of taxa with temperature is in agreement with their contribution in the PCoA biplot, resulting in PP samples being located at lower PCo1 values with the increase of the rearing temperature (Figure 7.1).

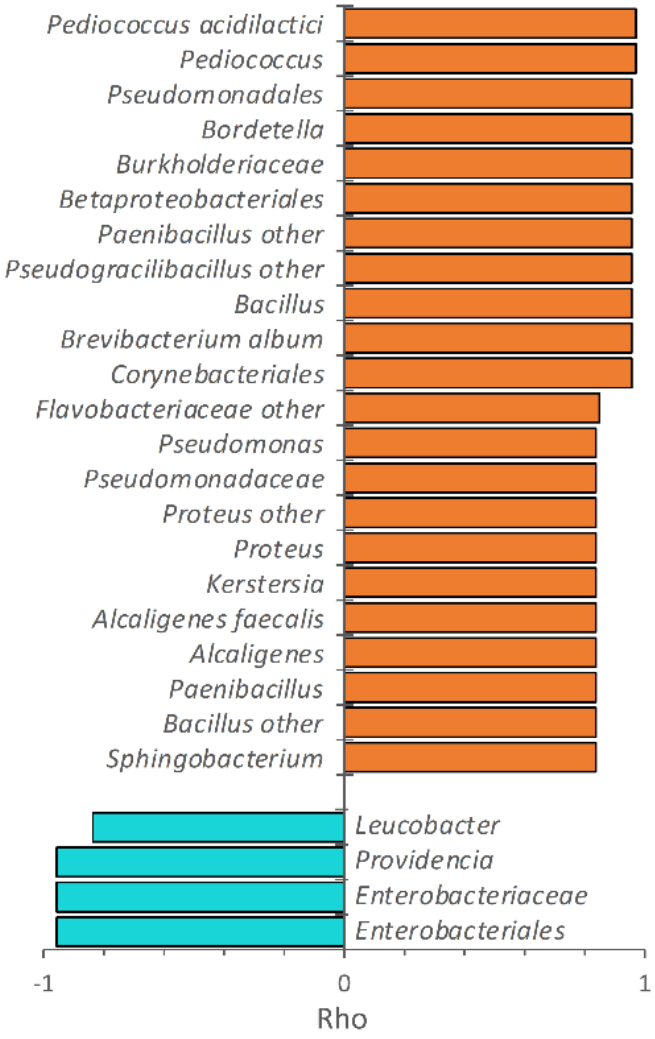


Figure 7.75 Spearman's rank correlation of the bacterial taxa in PP and the growth temperature. The correlation coefficient rho is reported only for the taxa exhibiting a significant correlation with temperature and appearing at least once with abundance > 1.0% (n = 2, P <0.05).

7.4 DISCUSSION

In the present study, the larvae of *H. illucens* were reared at three different temperatures (20, 27, and 33 °C), utilizing a standard vegetal substrate, until reaching the PP stadium.

Metagenome analysis provided comprehensive information on ecological aspects related to microbiological risk assessment in food or feed. Previous studies determined the microbiota composition of *H. illucens* by 16S rRNA gene profiling, focusing on entire specimens, with or without any surface sterilization, or on the dissected gut, in some cases distinguishing specific gut sections (De Smet et al., 2018; Wynants et al., 2018; Bruno et al., 2019; Jeon et al., 2011; Zheng et al., 2013; Shelomi et al., 2020). These studies pointed toward a great variability of *H. illucens* microbiota, with the substrate, the stage of development of the insect, and the gut section being the major players shaping the structure and the diversity of the bacterial community. In some cases, the microbiota of *H. illucens* was dominated by Bacteroidetes, in others by Proteobacteria, accompanied by very variable amounts of Actinobacteria and Firmicutes (De Smet et al., 2018). However, the existence of a core composition, independent of the substrate and transmitted through the development stages, seems plausible. In particular, *Providencia* was reported as one of the most recurrent and abundant genera, occurring in insects reared on different substrates.

Consistently, in the present study, the microbiota of L was dominated by *Providencia* (> 60%) and other Proteobateria (mainly *Klebsiella*) and evolved to a more complex composition with insect development. A lower amount of *Providencia* was still present in the microbiota of PP, while Actinobacteria, Bacteroidetes, and Bacilli bloomed (Figure 7.4). Many taxa dominating the microbiota of PP (e.g. *Providencia*, *Alcaligenes faecalis*, *Myroides*, *Morganella*, *Bordetella*, and *Kerstersia*) occurred also in L and were not found or were negligible in S. These taxa took the most advantage from the interaction with the insect, exhibiting the greatest fitness relative to *H. illucens*. Other taxa, such as *Pseudogracilibacillus* and other unclassified Bacillaceae, that colonized PP were not found in either L or S. It is not clear whether they were contaminants of the rearing environment or they initially lay below the limit of detection in L and S.

For the first time, a significant effect of the rearing temperature on the microbiota composition of *H. illucens* is reported. The temperature was negatively associated with the

amount of *Providencia*, which was the lowest at 33°C. On the contrary, the temperature was positively associated with a variety of other genera, such as *Alcaligenes*, *Pseudogracilibacillus, Bacillus, Proteus, Enterococcus*, *Pediococcus, Bordetella*, *Pseudomonas*, and *Kerstersia*. Thus, the temperature has to be included among the main factors affecting the microbiota and may have contributed, together with the rearing substrate, to the wide differences in the structure of the microbial community associated with *H. illucens* reported in literature.

The present study also indicates that *H. illucens* largely shares the microbiota with the medium where it is reared, except for a few lowly represented taxa. The main taxa constituting the microbiota of F and PP were the same, although some of them presented a differential abundance between the two environments, confirming that the insects and the medium shape the microbiota of each other (Bruno et al., 2019).

In addition to the classical microbiological risk assessment required in food regulation as stated in Advisory Reports 2014/2372 and 2019/6200 (NVWA, 2014; NVWA 2019), attention should be paid to the relevant abundance of some genera, such as *Myroides*, *Proteus, Providencia*, and *Morganella*, revealed by the 16S rRNA gene profiling. Species within these genera are not considered primary pathogens, although there is a vast literature describing them as opportunistic pathogens that may bear drug resistances and may cause severe morbidity (Armbruster et al., 2019; LaVergne et al., 2019; Minnullina et al., 2019; O'Hara et al., 2000; Sharma et al., 2017). Characterizing the species of *Myroides, Proteus, Providencia,* and *Morganella* harbored by *H. illucens* is recommended to establish whether they could be capable of causing adverse effects if present in the final product or throughout the production process.

These results of this studies were presented in the paper:

microorganisms

MDPI

Article

Effect of Rearing Temperature on Growth and Microbiota Composition of *Hermetia illucens*

Stefano Raimondi [1], Gloria Spampinato [1], Laura Ioana Macavei [1], Linda Lugli [1], Francesco Candeliere [1], Maddalena Rossi [1,2], Lara Maistrello [1,2] and Alberto Amaretti [1,2,*]

[1] Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy; stefano.raimondi@unimore.it (S.R.); gloria.spampinato@unimore.it (G.S.); lauraioana.macavei@unimore.it (L.I.M.); lugli.linda@gmail.com (L.L.); francesco.candeliere@unimore.it (F.C.); maddalena.rossi@unimore.it (M.R.); lara.maistrello@unimore.it (L.M.)

[2] BIOGEST–SITEIA, University of Modena and Reggio Emilia, 42124 Reggio Emilia, Italy

* Correspondence: alberto.amaretti@unimore.it; Tel.: +39-059-205-8588

Received: 1 June 2020; Accepted: 12 June 2020; Published: 15 June 2020

check for updates

8. REFERENCES

- Aagaard et al., (2014). The placenta harbors a unique microbiome. Sci Transl Med. May 2014, 21;6(237):237. doi.org/10.1126/scitranslmed.3008599
- Adijiang, Goto, Uramoto. (2008). Indoxyl sulphate promotes aortic calcification with expression of osteoblast-specific proteins in hypertensive rats. doi.org/10.1093/ndt/gfm861
- Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Cech, M. et al., (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic acids Res 46, W537-W544
- Alcock, B.P., Raphenya, A.R., Lau, T.T.Y., Tsang, K.K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48:D517-525. doi: 10.1093/nar/gkz935.
- Alemao, C.A., Budden, K.F., Gomez H.M., Rehman, S. F., Marshall, J.E., Shukla, S.D., et al. (2020). Impact of diet and the bacterial microbiome on the mucous barrier and immune disorders. Allergy. doi: 10.1111/all.14548
- Allard, M.W., Bell, R., Ferreira, C.M., Gonzalez-Escalona, N., Hoffmann, M., Muruvanda, T.et al. (2017). Genomics of foodborne pathogens for microbial food safety. Curr. Opin. Biotechnol. 49, 224–229. https://doi.org/10.1016/j.copbio.2017.11.002.
- Allard, M.W., Strain, E., Melka, D., Bunning, K., Musser, S.M., Brown, E.W. et al. (2016). Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J. Clin. Microbiol. 54 (8), 1975–1983. https://doi.org/10.1128/JCM.00081-16.
- Allison, C., MacFarlane, G.T. (1989). Influence of pH, nutrient availability and growth rate on amine production by *Bacteroides fragilis* and *Clostridium perfringens*. Applied Environmental Microbiology, 55: 2894-2898.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990). Basic local

alignment search tool. J Mol Biol. 215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

- Amaretti, A., Gozzoli, C., Simone, M., Raimondi, S., Righini, L., Perez-Brocal, V., et al. (2019). Profiling of Protein Degraders in Cultures of Human Gut Microbiota. Front Microbiol. 10:2614. doi:10.3389/fmicb.2019.02614
- Anand, S., Kaur, H., & Mande, S. S. (2016). Comparative In silico Analysis of Butyrate Production Pathways in Gut Commensals and Pathogens. Frontiers in microbiology, 7, 1945. https://doi.org/10.3389/fmicb.2016.01945
- Andreevskaya M, Hultman J, Johansson P, Laine P, Paulin L, Auvinen P, Björkroth J. (2016) Complete genome sequence of *Leuconostoc gelidum* subsp. *gasicomitatum* KG16-1, isolated from vacuum-packaged vegetable sausages. Stand Genomic Sci. 2016 Jun 7;11:40. doi: 10.1186/s40793-016-0164-8.
- Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc
- Armbruster, C.E.; Mobley, H.L.T.; Pearson M.M. (2019). Pathogenesis of *Proteus mirabilis* infection. EcoSal Plus. 8. doi: 10.1128/ecosalplus.ESP-0009-2017
- Arndt, D., Grant, J.R., Marcu, A., Sajed, T., Pon, A., Liang, Y. et al. (2016). PHASTER: a better, faster version of the PHAST phage search tool. Nucleic Acids Res. 44:W16-21. doi: 10.1093/nar/gkw387
- Aron-Wisnewsky, J., Doré, J. & Clement, K. (2012).The importance of the gut microbiota after bariatric surgery. Nat. Rev. Gastroenterol. Hepatol. **9**, 590–598 doi:10.1038/nrgastro.2012.161
- Asano, Y., Hiramoto, T., Nishino, R., Aiba, Y., Kimura, T., Yoshihara, K., Koga, Y., & Sudo, N. (2012). Critical role of gut microbiota in the production of biologically active, free catecholamines in the gut lumen of mice. American journal of physiology. Gastrointestinal and liver physiology, 303(11), G1288–G1295. https://doi.org/10.1152/ajpgi.00341.2012
- Asatoor AM, Simenhoff ML. (1965). The origin of urinary dimethylamine. Biochimica and Biophysica Acta Journal, III: 384-392.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R,

Vassieva O, Vonstein V, Wilke A, Zagnitko O. (2008). The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.

- Bäckhed F, Ley RE, Sonnenburg JL, Peterson DA, Gordon JI. (2005) Host-bacterial mutualism in the human intestine. Science. 2005;307(5717):1915–1920. doi: 10.1126/science.1104816. PMID: 15790844.
- Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al., (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19, 455–477
- Bansil, R., Turner, B.S. (2018). The biology of mucus: Composition, synthesis and organization. Adv Drug Deliv Rev. 124:3-15. doi:10.1016/j.addr.2017.09.023
- Barragan-Fonseca, K.B.; Dicke, M.; van Loon, J.J.A. (2017). Nutritional value of the black soldier fly (Hermetia illucens L.) and its suitability as animal feed – a review. J. Ins. Food Feed. 3, 105–120. doi: 10.3920/JIFF2016.0055
- Bekal, S., Van Beeumen, J., Samyn, B., Garmyn, D., Henini, S., Diviès, C., Prévost, H. (1998). Purification of *Leuconostoc mesenteroides* citrate lyase and cloning and characterization of the citCDEFG gene cluster. J Bacteriol 180(3):647-654. doi: 10.1128/JB.180.3.647-654.1998.
- Belghit, I., Liland, N. S., Gjesdal, P., Biancarosa, I., Menchetti, E., Li, Y., et al. (2019). Black soldier fly larvae meal can replace fish meal in diets of sea-water phase Atlantic salmon (*Salmo salar*). Aquaculture 503, 609-619. doi: 10.1016/j.aquaculture.2018.12.032
- Bentley, D., Balasubramanian, S., Swerdlow, H. et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. Nature 456, 53–59. https://doi.org/10.1038/nature07517
- Bergstrom, K.S., Xia, L. (2013). Mucin-type O-glycans and their roles in intestinal homeostasis. Glycobiology. 23:1026-1037. doi: 10.1093/glycob/cwt045.
- Björkroth, K.J., Vandamme, P., Korkeala, H.J. (1998). Identification and characterization of *Leuconostoc carnosum*, associated with production and spoilage of vacuum-packaged, sliced, cooked ham. Appl Environ Microbiol 64:3313–3319. https://doi.org/10.1128/AEM.64.9.3313-3319.1998

- Bo, J., Wang, S., Bi, Y., Ma, S., Wang, M., Du, Z. (2020). *Eggerthella lenta* bloodstream infections: two cases and review of the literature. Future Microbiol. 15:981-985. doi:10.2217/fmb-2019-0338
- Boelsterli, U. A., Redinbo, M. R. & Saitta, K. S. (2013). Multiple NSAID-induced hits injure the small intestine: Underlying mechanisms and novel strategies. Toxicol. Sci. **131**, 654–667 doi: 10.1093/toxsci/kfs310.
- Boisvert S, Laviolette F, Corbeil J. (2010). Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17:1519 –1533. https://doi.org/10.1089/cmb.2009.0238.
- Bokulich, N.A., Kaehler, B.D., Rideout, J.R., Dillon, M., Bolyen E., Knigh,t R., et al. (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome. 6, 90. doi: 10.1186/s40168-018-0470-z
- Bolger, A.M., Lohse, M., Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 30(15):2114-2120. doi:10.1093/bioinformatics/btu170
- Bolyen, E., Rideout, J.R., Chase, J., Pitman, T.A., Shiffer, A., Mercurio, W. et al. (2018) An introduction to applied bioinformatics: a free, open, and interactive text. J Open Source Educ. 1, 27. doi: 10.21105/jose.00027
- Bolyen, E., Rideout, J.R., Dillon, M.R., et al. (2019).Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol. 37(8):852-857. doi:10.1038/s41587-019-0209-9
- Bortolini, S., Macavei, L.I., Hadj Saadoun, J., Foca, G., Ulrici, A., Bernini, F. et al. (2020). *Hermetia illucens* (L.) larvae as chicken manure management tool for circular economy. J. Clean. Prod. 262, 121289. doi: 10.1016/j.jclepro.2020.121289
- Bray & Curtis, (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs 27: 325–349
- Bruni, L., Pastorelli, R., Viti, C., Gasco, L., Parisi, G. (2018) Characterization of the intestinal microbial communities of rainbow trout (*Oncorhynchus mykiss*) fed with *Hermetia illucens* (black soldier fly) partially defatted larva meal as partial dietary protein source. Aquaculture 487, 56–63 doi: 10.1016/j.aquaculture.2018.01.006

- Bruno, D., Bonelli, M., De Filippis, F., Di Lelio, I., Tettamanti, G., Casartelli, M. et al. (2019). The intestinal microbiota of *Hermetia illucens* larvae is affected by diet and shows a diverse domposition in the different midgut regions. Appl. Environ. Microbiol. 85, e01864-18. doi: 10.1128/AEM.01864-18
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. Nature methods, 12(1), 59–60. https://doi.org/10.1038/nmeth.3176
- Budde, B.B., Hornbaek, T., Jacobsen, T., Barkholt, V., Koch, A.G. (2003). *Leuconostoc carnosum* 4010 has the potential for use as a protective culture for vacuum-packed meats: culture isolation, bacteriocin identification, and meat application experiments. Int J Food Microbiol 83:171–184. https:// doi.org/10.1016/S0168-1605(02)00364-1.
- Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J., Holmes, S.P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods. 13(7):581-583. doi:10.1038/nmeth.3869
- Calmels S, Ohshima H, Henry Y, Bartsch H. (1996). Characterization of bacterial cytochrome cd1-nitrite reductase as one enzyme responsible for catalysis of secondary amines. Carcinogenesis, 17: 533-536.
- Campedelli, I., Mathur, H., Salvetti, E., Clarke, S., Rea, M.C., Torriani, S. et al. (2019). Genus-wide assessment of antibiotic resistance in *Lactobacillus* spp. Appl Environ Microbiol 85:e01738-18. https://doi.org/10.1128/AEM.01738-18
- Candeliere, F., Raimondi, S., Spampinato, G., Tay, M.Y.F., Amaretti, A., Schlundt, J. et al. (2020). Draft genome sequences of 12 *Leuconostoc carnosum* strains isolated from cooked ham packaged in a modified atmosphere and from fresh sausages. Microbiol Resour Announc. 9(2). pii: e01247-19. doi: 10.1128/MRA.01247-19.
- Cao, Y., Fanning, S., Proos, S., Jordan, K., & Srikumar, S. (2017). A Review on the Applications of Next Generation Sequencing Technologies as Applied to Food-Related Microbiome Studies. Frontiers in microbiology, 8, 1829. https://doi.org/10.3389/fmicb.2017.01829
- Caplice E, Fitzgerald GF. (1999). Food fermentations: role of microorganisms in food production and preservation. Int J Food Microbiol. Sep 15;50(1-2):131-49. doi: 10.1016/s0168-1605(99)00082-3.

- Carleton H.A., Gerner-Smidt P., (2016). Whole-genome sequencing is taking over foodborne disease surveillance. Microbe Mag 11:311–317. 7. https:// doi.org/10.1128/microbe.11.311.1.
- Carmody, R. N. & Turnbaugh, P. J. (2014). Host-microbial interactions in the metabolism of therapeutic and diet-derived xenobiotics. J. Clin. Invest. **124**, 4173–4181 doi: 10.1172/JCI72335.
- Chao, (1984). Nonparametric Estimation of the Number of Classes in a Population. Scand J Statist 11: 265-270, 1984.
- Chen, Y.S., Wang, L.T., Wu, Y.C., Mori, K., Tamura, T., Chang, C.H et al. (2020). *Leuconostoc litchii* sp. nov., a novel lactic acid bacterium isolated from lychee. Int J Syst Evol Microbiol. 70:1585–1590. doi: 10.1099/ijsem.0.003938
- Choi, M. S., Yu, J. S., Yoo, H. H. & Kim, D. H. (2018). The role of gut microbiota in the pharmacokinetics of antihypertensive drugs. Pharmacol. Res. **130**, 164–171 doi: 10.1016/j.phrs.2018.01.019
- Chung H et al. (2012). Gut immune maturation depends on the colonization with a host specific microbiota. Ell, 149. 1578-1593. doi: 10.1016/j.cell.2012.04.037.
- Claesson et al., (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. Nucleic Acids Res. 2010 Dec;38(22):e200 https://doi.org/10.1093/nar/gkq873
- Cobos, Á., Díaz, O. (2015). Chemical Composition of Meat and Meat Products. In: Cheung P, Mehta B (eds.) Handbook of Food Chemistry. Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/978-3-642-36605-5_6
- Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed- Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M., Tiedje, J.M., (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. Nucleic Acids Res. 37, 141-145. doi: 10.1093/nar/gkn879.
- Collins, M.D., Samelis, J., Metaxopoulus, J., Wallbanks, S., (1993). Taxonomic studies on some *Leuconostoc*-like organisms from fermented sausages: description of a new genus *Weissella* for the *Leuconostoc paramesenteroides* group of species. J. Appl. Bacteriol. 75, 595–603.

- Corfield, A.P. (2018). The Interaction of the Gut Microbiota with the Mucus Barrier in Health and Disease in Human. Microorganisms. 6(3):78. doi:10.3390/microorganisms6030078
- Cornick, S., Tawiah, A., Chadee, K. (2015). Roles and regulation of the mucus barrier in the gut. Tissue Barriers. 3(1-2):e982426. doi:10.4161/21688370.2014.982426
- Couvin, D., Bernheim, A., Toffano-Nioche, C., Touchon, M., Michalik, J., Néron, B. et al. (2018). CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 46:W246-251. doi: 10.1093/nar/gky425
- Cummings JH, Southgate DAT, Branch WJ, Wiggins HS. (1979). The digestion of pectin in the human gut and its effect on calcium absorption and large bowel function. British Journal of Nutrition, 41, 477.
- Cutrignelli, M.. I., Messina, M., Tulli, F., Randazzo, B., Olivotto, I., Gasco, L. et al. (2018). Evaluation of an insect meal of the Black Soldier Fly (*Hermetia illucens*) as soybean substitute: Intestinal morphometry, enzymatic and microbial activity in laying hens. Res. Vet Sci. 117, 209-215. doi: 10.1016/j.rvsc.2017.12.020
- Dabek, M., McCrae, S. I., Stevens, V. J., Duncan, S. H. & Louis, P. (2008). Distribution of β-glucosidase and β-glucuronidase activity and of β-glucuronidase gene gus in human colonic bacteria. FEMS Microbiol. Ecol. **66**, 487–495 doi: 10.1111/j.1574-6941.2008.00520.x.
- Darling, A.C., Mau, B., Blattner, F.R., Perna, N.T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res, 14, 1394-1403
- Dashnyam, P. et al. (2018).β-Glucuronidases of opportunistic bacteria are the major contributors to xenobiotic-induced toxicity in the gut. Sci. Rep. **8**, 1–12 doi: 10.1038/s41598-018-34678-z.
- David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling AV, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. (2014). Diet rapidly and reproducibly alters the human gut microbiome. Nature, 505(7484):559-63. doi: 10.1038/nature12820

- De Filippis, F., La Storia, A., Villani, F., and Ercolini, D. (2013). Exploring the sources of bacterial spoilers in beefsteaks by culture-independent high-throughput sequencing. PLoS ONE 8:e70222. doi: 10.1371/journal.pone.0070222
- De Filippis, F., Parente, E., & Ercolini, D. (2017). Metagenomics insights into food fermentations. Microbial biotechnology, 10(1), 91–102. https://doi.org/10.1111/1751-7915.12421
- De Loor H, Bammens B, Evenepoel P, de Preter V, Verbeke K. (2005). Gas chromatographic-mass spectrometric analysis for measurement of p-cresol and its conjugated metabolites in uremic and normal serum. Clinical Chemistry, 51(8):1535-8. doi: 10.1373/clinchem.2005.050781
- De Smet, J., Wynants, E., Cos, P., Van Campenhout, L. (2018). Microbial community dynamics during rearing of black soldier fly larvae (*Hermetia illucens*) and impact on exploitation potential. Appl. Environ. Microbiol. 84, e02722-17. doi: 10.1128/AEM.02722-17
- De Vuyst L, Leroy F. (2007) Bacteriocins from lactic acid bacteria: production, purification, and food applications. J Mol Microbiol Biotechnol. (2007);13(4):194-9. doi: 10.1159/000104752. PMID: 17827969.
- Delcher, A. L., Bratke, K. A., Powers, E. C., & Salzberg, S. L. (2007). Identifying bacterial genes and endosymbiont DNA with Glimmer. Bioinformatics (Oxford, England), 23(6), 673–679. https://doi.org/10.1093/bioinformatics/btm009
- Dethlefsen et al., (2008). The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. PLoS Biol. 2008 Nov; 6(11). doi: 10.1371/journal.pbio.0060280
- Di Martino ML, Campilongo R, Casalino M, Micheli G, Colonna B, Prosseda G. (2013). Polyamines: emerging players in bacteria-host interactions. International Journal of Medical Microbiology, 303 484-491. doi: 10.1016/j.ijmm.2013.06.008
- Dicks, L.M.T., Dellaglio, F., Collins, M.D., (1995). Proposal to reclassify *Leuconostoc oenos* as *Oenococcus oeni* [corrig.] gen. nov., comb. nov. Int. J. Syst. Bacteriol. 45, 395–397.
- DiStefano, J. J., 3rd, de Luze, A., & Nguyen, T. T. (1993). Binding and degradation of 3,5,3'-triiodothyronine and thyroxine by rat intestinal bacteria. The American journal of physiology, 264(6 Pt 1), E966–E972.

https://doi.org/10.1152/ajpendo.1993.264.6.E966

- Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, et al. (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. Proc. Natl. Acad. Sci. USA 107: 11971–75. doi: 10.1073/pnas.1002601107
- Donaldson, G.P., Lee, S.M., Mazmanian, S.K. (2016). Gut biogeography of the bacterial microbiota. Nat Rev Microbiol. 14(1):20-32. doi:10.1038/nrmicro3552
- Dou L, Bertrand E, Cerini C, Faure V, Sampol J, Vanholder R, Berland Y, Brunet P. (2004). The uremic solutes p-cresol and indoxyl sulfate inhibit endothelial proliferation and wound repair. Kidney International, 65(2):442-51. doi.org/10.1111/j.1523-1755.2004.00399.x
- Douillard, F.P., and de Vos, W.M. (2014) Functional genomics of lactic acid bacteria: from food to health. Microb Cell Fact 13: S8. https://doi.org/10.1186/1475-2859-13-S1-S8
- Doyle, R. M., Alber, D. G., Jones, H. E., Harris, K., Fitzgerald, F., Peebles, D., & Klein, N. (2014). Term and preterm labour are associated with distinct microbial community structures in placental membranes which are independent of mode of delivery. Placenta, 35(12), 1099–1101. https://doi.org/10.1016/j.placenta.2014.10.007
- Dutton, G. (Ed.). (2012). Glucuronic acid free and combined: chemistry, biochemistry, pharmacology, and medicine. Elsevier.
- E.C. Pielou, (1966) The measurement of diversity in different types of biological collections, Journal of Theoretical Biology, Volume 13, 1966, Pages 131-144, ISSN 0022-5193, https://doi.org/10.1016/0022-5193(66)90013-0.
- Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M. et al. (2005). Diversity of the human intestinal microbial flora. Science (New York, N.Y.), 308(5728), 1635–1638. https://doi.org/10.1126/science.1110591
- EFSA Scientific Committee (2015). Risk profile related to production and consumption of insects as food and feed. EFSA J. 13, 4257. doi: 10.2903/j.efsa.2015.4257
- Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-time DNA

sequencing from single polymerase molecules. Exch. Organ. Behav. Teach. J. 323, 133–138. doi:10.1126/science.1162986

- Eisenhofer, G., Aneman, A., Friberg, P., Hooper, D., Fåndriks, L., Lonroth, H., Hunyady, B., & Mezey, E. (1997). Substantial production of dopamine in the human gastrointestinal tract. The Journal of clinical endocrinology and metabolism, 82(11), 3864–3871. https://doi.org/10.1210/jcem.82.11.4339
- Endo, A., Okada, S., (2008). Reclassification of the genus *Leuconostoc* and proposals of *Fructobacillus fructosus* gen. nov., comb. nov., *Fructobacillus durionis* comb. nov., *Fructobacillus ficulneus* comb. nov. and *Fructobacillus pseudoficulneus* comb. nov. Int. J. Syst. Evol. Microbiol. 58, 2195–2205. doi: 10.1099/ijs.0.65609-0
- Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L. et al. (2015). Anvi'o: an advanced analysis and visualization platform for 'omics data. PeerJ 3, e1319. doi: 10.7717/peerj.1319.
- Eren, A.M., Maignien, L., Sul, W.J., Murphy, L.G., Grim, S.L., Morrison, H.G., Sogin, M.L., (2013). Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol. Evol. 4, 1111–1119. https://doi.org/10.1111/2041-210X.12114.
- Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM et al. (2012). A pilot study of rapid benchtop sequencing of *Staphylococcus aureus* and *Clostridium difficile* for outbreak detection and surveillance. BMJ Open 2:e001124. https://doi.org/10.1136/bmjopen-2012-001124.
- Fanaro S., Chierici R., Guerrini P., Vigi V. (2003). Intestinal microflora in early infancy: composition and development. Acta Paediatr. Suppl. 91, 48–55. doi: 10.1111/j.1651-2227.2003.tb00646.x.
- Felix, J.V., Papathanasopoulos, M.A., Smith, A.A., von Holy, A., Hastings, J.W. (1994). Characterization of Leucocin B-Talla: A Bacteriocin from *Leuconostoc carnosum* Talla Isolated from Meat. Curr Microbiol 29:207–212. https://doi.org/10.1007/BF01570155
- Ferrocino, I., Greppi, A., Lastoria, A., Rantsiou, K., and Ercolini, D. (2016). Impact of nisin-activated packaging on microbiota of beef burgers during storage. Appl. Environ. Microbiol. 82, 1–13. doi: 10.1128/AEM.03093-15
- Fevery, J., Blanckaert, N., Leroy, P., Michiels, R., & Heirwegh, K. P. (1983).

Analysis of bilirubins in biological fluids by extraction and thin-layer chromatography of the intact tetrapyrroles: application to bile of patients with Gilbert's syndrome, hemolysis, or cholelithiasis. Hepatology (Baltimore, Md.), 3(2), 177–183. https://doi.org/10.1002/hep.1840030207

- Fierer N. (2017). Embracing the unknown: disentangling the complexities of the soil microbiome. Nature reviews. Microbiology, 15(10), 579–590. https://doi.org/10.1038/nrmicro.2017.87
- Flemström, G., Säfsten, B., & Jedstedt, G. (1993). Stimulation of mucosal alkaline secretion in rat duodenum by dopamine and dopaminergic compounds. Gastroenterology, 104(3), 825–833. https://doi.org/10.1016/0016-5085(93)91019-e
- Flores et al., 2014. Temporal variability is a personalized feature of the human microbiome. Genome Biology 2014, 15:531. doi: 10.1186/s13059-014-0531-y.
- Florin, T. H., Neale, G., Goretski, S., & Cummings, J. H. (1993). The sulfate content of foods and beverages. Journal of food composition and analysis, 6(2), 140-151. doi.org/10.1006/jfca.1993.1016
- Gao, Z., Tseng, C. H., Pei, Z., & Blaser, M. J. (2007). Molecular analysis of human forearm superficial skin bacterial biota. Proceedings of the National Academy of Sciences of the United States of America, 104(8), 2927–2932. https://doi.org/10.1073/pnas.0607077104
- García Quintáns, N., Blancato, V. S., Repizo, G. D., Magni, C., López, P. (2008). “Citrate metabolism and aroma compound production in lactic acid bacteria” in Molecular Aspects of Lactic Acid Bacteria for Traditional and New Applications. Eds. Mayo, B., López, P., Pérez-Martínez, G. (Thiruvananthapuram, Kerala, India; Research Signpost), Chapter 3:65-88.
- García-Solache, M., Rice, L.B. (2019). The *Enterococcus*: a model of adaptability to its environment. Clin Microbiol Rev. 32(2):e00058-18. doi:10.1128/CMR.00058-18
- Garvie, E.I. (1986). Genus *Leuconostoc* van Tieghem 1878, 198AL emend mut.char. Hucker and Pederson 1930, 66AL. In: Sneath, P.H.A., Mair, N.S., Sharpe, M.E. & Holt, J.G. (eds). Bergey’s Manual of Systematic Bacteriology, vol. 2. The Williams & Wilkins Co., Baltimore. Pp. 1071-1075.
- Geeraerts, W., Pothakos, V., De Vuyst, L., Leroy, F. (2018). Variability within the dominant microbiota of sliced cooked poultry products at expiration date in the

Belgian retail. Food Microbiol 73:209 –215. https://doi.org/ 10.1016/j.fm.2018.01.019

- Geypens B, Claus D, Evenepoel P, Hiele M, Maes B, Peeters M, Rutgeerts P, Ghoos Y. (1997). Influence of dietary protein supplements on the formation of bacterial metabolites in the colon. Gut, 41: 70-76. doi: 10.1136/gut.41.1.70.
- Gill, C. I., & Rowland, I. R. (2002). Diet and cancer: assessing the risk. The British journal of nutrition, 88 Suppl 1, S73–S87. https://doi.org/10.1079/BJN2002632
- Gloux, K., Berteau, O., El Oumami, H., Béguet, F., Leclerc, M., & Doré, J. (2011). A metagenomic β-glucuronidase uncovers a core adaptive function of the human intestinal microbiome. Proceedings of the National Academy of Sciences of the United States of America, 108 Suppl 1(Suppl 1), 4539–4546. https://doi.org/10.1073/pnas.1000066107
- Goodwin, S., McPherson, J.D., McCombine, W.R. (2016). Coming of age: ten years of next generation sequencing technologies. Nat. Rev. Genet. 17, 333–351. https://doi.org/10.1038/nrg.2016.49.
- Gotelli, N. J. and R. K. Colwell. (2011). Estimating species richness. Frontiers in measuring biodiversity. Oxford University Press, New York
- Graef, V., Furuya, E., & Nishikaze, O. (1977). Hydrolysis of steroid glucuronides with beta-glucuronidase preparations from bovine liver, Helix pomatia, and E. coli. Clinical chemistry, 23(3), 532–535.
- Gurevich A., Saveliev V., Vyahh, N., Tesler G. (2013). QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072-1075
- H.B. Ghoddusi. (2011). Biotechnology of lactic acid bacteria: novel applications. Int. J. Dairy Technol., 64 (3), pp. 456-458.
- Hamer, H. M., De Preter, V., Windey, K., & Verbeke, K. (2012). Functional analysis of colonic bacterial metabolism: relevant to health?. American journal of physiology. Gastrointestinal and liver physiology, 302(1), G1–G9. https://doi.org/10.1152/ajpgi.00048.2011
- Hammes, W.P. and Hertel, C. (2006). The Genera *Lactobacillus* and *Carnobacterium*. In The Prokaryotes, 3rd edn. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackenbrandt, E. (eds). New York: Springer, vol. 4, pp. 319-402.

- Haskel, Y., & Hanani, M. (1994). Inhibition of gastrointestinal motility by MPTP via adrenergic and dopaminergic mechanisms. Digestive diseases and sciences, 39(11), 2364–2367. https://doi.org/10.1007/BF02087652
- Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Moller N, Aarestrup FM. (2014). Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol 52:139 –146. doi.org/10.1128/JCM.02452-13.
- Hastings, J.W., Stiles, M.E., von Holy, A. (1994). Bacteriocins of *Leuconostocs* isolated from meat. Int J Food Microbiol. 24:75-81. doi: 10.1016/0168-1605(94)90107-4
- Hazenberg, M. P., de Herder, W. W., & Visser, T. J. (1988). Hydrolysis of iodothyronine conjugates by intestinal bacteria. FEMS microbiology reviews, 4(1), 9–16. https://doi.org/10.1111/j.1574-6968.1988.tb02709.x-i1
- Hemme, D. and Foucaud-Scheunemann, C. (2004) *Leuconostoc*, Characteristics, Use in Dairy Technology and Prospects in Functional Foods. International Dairy Journal, 14, 467-494. http://dx.doi.org/10.1016/j.idairyj.2003.10.005
- Herbert J. Buckenhüskes, (1993) Selection criteria for lactic acid bacteria to be used as starter cultures for various food commodities, FEMS Microbiology Reviews, Volume 12, Issue 1-3, September 1993, Pages 253–271, https://doi.org/10.1111/j.1574-6976.1993.tb00022.x.
- Hold, G. L., Pryde, S. E., Russell, V. J., Furrie, E., & Flint, H. J. (2002). Assessment of microbial diversity in human colonic samples by 16S rDNA sequence analysis. FEMS microbiology ecology, 39(1), 33–39. https://doi.org/10.1111/j.1574-6941.2002.tb00904.x
- Holmén Larsson, J.M., Thomsson, K.A., Rodríguez-Piñeiro, A.M., Karlsson, H., Hansson, G.C. (2013). Studies of mucus in mouse stomach, small intestine, and colon. III. Gastrointestinal Muc5ac and Muc2 mucin O-glycan patterns reveal a regiospecific distribution. Am J Physiol Gastrointest Liver Physiol. 305:G357-363. doi: 10.1152/ajpgi.00048.2013
- Hong, Y., Yang, H.S., Li, J., Han, S.K., Chang, H.C., Kim, H.Y. (2017). Identification of lactic acid bacteria in salted Chinese cabbage by SDS-PAGE and PCR-DGGE. J Sci Food Agric. 94:296–300. doi:10.1002/jsfa.6257
- Hughes R. (1999). The effects of diet on colonic N-nitrosation and biomarkers of

DNA damage. PhD Thesis, University of Cambridge.

- Huson, D. H., Albrecht, B., Bağcı, C., Bessarab, I., Górska, A., Jolic, D., & Williams, R. (2018). MEGAN-LR: new algorithms allow accurate binning and easy interactive exploration of metagenomic long reads and contigs. Biology direct, 13(1), 6. https://doi.org/10.1186/s13062-018-0208-7
- Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. Genome research, 17(3), 377–386. https://doi.org/10.1101/gr.5969107
- Huttenhower C, Morgan XC 2012. Chapter 12: Human Microbiome Analysis. PLoS Comput Biol 8(12). doi: 10.1371/journal.pcbi.1002808
- Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
- Hyman, R. W., Fukushima, M., Diamond, L., Kumm, J., Giudice, L. C., & Davis, R. W. (2005). Microbes on the human vaginal epithelium. Proceedings of the National Academy of Sciences of the United States of America, 102(22), 7952–7957. https://doi.org/10.1073/pnas.0503236102
- Jackson, B.R., Tarr, C., Strain, E., Jackson, K.A., Conrad, A., Carleton, H., et al., (2016). Implementation of nationwide real-time whole-genome sequencing to enhance Listeriosis outbreak detection and investigation. Clin. Infect. Dis. 63, 380–386. https://doi.org/10.1093/cid/ciw242.
- Jacobsen, T., Budde, B.B., Koch, A.G. (2003). Application of *Leuconostoc carnosum* for biopreservation of cooked meat products. J Appl Microbiol 95:242–249. https://doi.org/10.1046/j.1365-2672.2003.01970.x.
- Jain, S., Drendel, W. B., Chen, Z. W., Mathews, F. S., Sly, W. S., & Grubb, J. H. (1996). Structure of human beta-glucuronidase reveals candidate lysosomal targeting and active-site motifs. Nature structural biology, 3(4), 375–381. https://doi.org/10.1038/nsb0496-375
- Jeon, H., Park, S., Choi, J., Jeong, G., Lee, S.B., Choi, Y. et al.(2011). The intestinal bacterial community in the food waste-reducing larvae of *Hermetia illucens*. Curr. Microbiol. 62, 1390–1399. doi: 10.1007/s00284-011-9874-8

- Jeon, H.H., Kim, K.H., Chun, B.H., Ryu, B.H., Han, N.S., Jeon, C.O. (2017). A proposal of *Leuconostoc mesenteroides* subsp. *jonggajibkimchii* subsp. nov. and reclassification of *Leuconostoc mesenteroides* subsp. *suionicum* (GU et al., 2012) as *Leuconostoc suionicum* sp. nov. based on complete genome sequences. Int J Syst Evol Microbiol 67:2225–2230. doi: 10.1099/ijsem.0.001930
- Jernberg et al., (2007). Long-term impacts of antibiotic exposure on the human intestinal microbiota. ISME J. 2007 May;1(1):56-66. doi: 10.1038/ismej.2007.3
- Johansson, M.E., Larsson, J.M., Hansson, G.C. (2011). The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci U S A. 108 Suppl 1(Suppl 1):4659-4665. doi:10.1073/pnas.1006451107
- Jolliffe, I.T., (1986). Principal Component Analysis. New York: Springer.
- Jung, J.Y., Lee, S.H., Jeon, C.O. (2012). Complete genome sequence of *Leuconostoc carnosum* strain JB16, isolated from kimchi. J Bacteriol. 194:6672-6673. doi: 10.1128/JB.01805-12
- Jung, J.Y., Lee, S.H., Jeon, C.O. (2014) Kimchi microflora: history, current status, and perspectives for industrial kimchi production. Appl Microbiol Biotechnol. 98(6):2385–2393. doi:10.1007/s00253-014-5513-1
- K.H. Schleifer, W. Ludwig (1995) Phylogeny of the Genus *Lactobacillus* and Related Genera, Systematic and Applied Microbiology, Volume 18, Issue 4, 1995, Pages 461-467,ISSN 0723-2020, https://doi.org/10.1016/S0723-2020(11)80404-2.
- Kable, M. E., Srisengfa, Y., Laird, M., Zaragoza, J., Mcleod, J., Heidenreich, J., et al. (2016). The core and seasonal microbiota of raw bovine milk in tanker trucks and the impact of transfer to a milk processing facility. MBio 7, 1–13. doi: 10.1128/mBio.00836-16
- Kandler, O. and Weiss, N. (1986) Regular, Non-Sporing Gram-Positive Rods. In: Sneath, H.A., Mair, N.S., Sharpe, M.E. and Holt, J.G., Eds., Bergey's Manual of Systematic Bacteriology, Williams and Wilkins, Baltimore, 1208-1234.
- Kanehisa, M., Goto, S., (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 28 (1), 27–30. doi:10.1093/nar/28.1.27
- Kanehisa, M., Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci. 29:28-35. doi: 10.1002/pro.3711.

- Kanehisa, M., Sato, Y., Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. J Mol Biol. Feb 428:726-731. doi: 10.1016/j.jmb.2015.11.006.
- Kang, D. D., Froula, J., Egan, R. & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3, e1165
- Kappes, R.M., Kempf, B., Kneip, S., Boch, J., Gade, J., Meier-Wagner, J. et al. (1999). Two evolutionarily closely related ABC transporters mediate the uptake of choline for synthesis of the osmoprotectant glycine betaine in Bacillus subtilis. Mol Microbiol. 32:203-216. doi: 10.1046/j.1365-2958.1999.01354.x
- Kim DH, Hong SW, Kim BT, Bae EA, Park HY, Han MJ. (2000) Biotransformation of glycyrrhizin by human intestinal bacteria and its relation to biological activities. Arch Pharm Res. 2000 Apr;23(2):172-7. doi: 10.1007/BF02975509.
- Kim, J., Hetzel, M., Boiangiu, C. D., and Buckel, W. (2004). Dehydration of (R)-2-hydroxyacyl-CoA to enoyl-CoA in the fermentation of alpha-amino acids by anaerobic bacteria. FEMS Microbiol. Rev. 28, 455–468. doi: 10.1016/j.femsre.2004.03.001
- Kobayashi, N., Nishino, K., Yamaguchi, A. (2001). Novel macrolide-specific ABC-type efflux transporter in *Escherichia coli*. J Bacteriol. 183 5639-5644. doi: 10.1128/JB.183.19.5639-5644.2001
- Koh, A., De Vadder, F., Kovatcheva-Datchary, P. & Bäckhed, F. (2016).From Dietary Fiber to Host Physiology: Short-Chain Fatty Acids as Key Bacterial Metabolites. Cell **165**, 1332–1345 doi: 10.1016/j.cell.2016.05.041
- Konstantinidis, K.T., Tiedje, J.M. (2005). Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A 102 (7), 2567–2572. doi.org/10.1073/pnas.0409727102.
- Koppel, N., Rekdal, V. M. & Balskus, E. P. (2017). Chemical transformation of xenobiotics by the human gut microbiota. Science (80-.). **356**, 1246–1257 doi: 10.1126/science.aag2770.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

- Kreek, M. J., Guggenheim, F. G., Ross, J. E., & Tapley, D. F. (1963). Glucuronide formation in the transport of testosterone and androstenedione by rat intestine. Biochimica et biophysica acta, 74, 418–427. https://doi.org/10.1016/0006-3002(63)91385-4
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. Genome biology, 5(2), R12. https://doi.org/10.1186/gb-2004-5-2-r12.
- Kvistholm Jensen, A., Nielsen, E.M., Bjorkman, J.T., Jensen, T., Muller, L., Persson, S. et al. (2016). Whole-genome sequencing used to investigate a nationwide outbreak of listeriosis caused by ready-to-eat delicatessen meat, Denmark, 2014. Clin. Infect. Dis. 63, 64–70. https://doi.org/10.1093/cid/ciw192.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100 –3108. https://doi.org/10.1093/nar/gkm160.
- Lagier, J. C., Dubourg, G., Million, M., Cadoret, F., Bilen, M., Fenollar, F., Levasseur, A., Rolain, J. M., Fournier, P. E., & Raoult, D. (2018). Culturing the human microbiota and culturomics. Nature reviews. Microbiology, 16, 540–550. https://doi.org/10.1038/s41579-018-0041-0
- Lagier, J. C., Khelaifia, S., Alou, M. T., Ndongo, S., Dione, N., Hugon, P. et al. (2016). Culture of previously uncultured members of the human gut microbiota by culturomics. Nature microbiology, 1, 16203. https://doi.org/10.1038/nmicrobiol.2016.203
- Lagier, J. C., Million, M., Hugon, P., Armougom, F., & Raoult, D. (2012). Human gut microbiota: repertoire and variations. Frontiers in cellular and infection microbiology, 2, 136. https://doi.org/10.3389/fcimb.2012.00136
- Lalander, C.H., Fidjeland, J., Diener, S., Eriksson, S., Vinnerås, B. (2015). High waste-to-biomass conversion and efficient *Salmonella* spp. reduction using black soldier fly for waste recycling. Agronomy Sust. Develop. 35, 261–271. doi.org/10.1007/s13593-014-0235-4
- Langmead, B., Salzberg, S., (2013). Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357-359. doi:10.1038/nmeth.1923
- Larsen MV, Cosentino S, Lukjancenko O, Saputra D, Rasmussen S, Hasman H,

Sicheritz-Ponten T, Aarestrup FM, Ussery DW, Lund O. (2014). Benchmarking of methods for genomic taxonomy. J Clin Microbiol 52:1529 –1539. https://doi.org/10.1128/JCM.02981-13.

- Larsson W. (2006). Meat consumption and risk of colorectal cancer: A meta-analysis of prospective studies. International Journal of Cancer, 119(11):2657-64; doi: 10.1002/ijc.22170
- Laslett D, Canback B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
- Lau, S.K., Woo, P.C., Woo, G.K, Fung, A.M., Ngan, A.H.Y., Song, Y., et al. (2006). Bacteraemia caused by *Anaerotruncus colihominis* and emended description of the species. J Clin Pathol. 59(7):748-752. doi:10.1136/jcp.2005.031773
- LaVergne, S., Gaufin, T., Richman, D. (2019). Myroides injenensis Bacteremia and severe cellulitis. Open Forum Infect. Dis. 6, ofz282. doi: 10.1093/ofid/ofz282
- Lawton, M.R., Jencarelli, K.G., Kozak, S.M., Alcaine, S.D. (2020). Short communication: Evaluation of commercial meat cultures to inhibit Listeria monocytogenes in a fresh cheese laboratory model. J Dairy Sci 103:1269–1275. doi:10.3168/jds.2019-17203
- Legendre, P., Legendre, L., (1998). Numerical Ecology, Elsevier.
- Letunic, I., Peer Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments, Nucleic Acids Res, Volume 47, Issue W1, 02 July 2019, Pages W256–W259, https://doi.org/10.1093/nar/gkz239
- Ley R, Walter J, (2011). The human gut microbiome: ecology and recent evolutionary changes. Annual Review of Microbiology, 65:411-29. doi: 10.1146/annurev-micro-090110-102830
- Li Q, Wang C, Tang C, Li N, Li J. (2012). Molecular-phylogenetic characterization of the microbiota in ulcerated and non-ulcerated regions in the patients with Crohn's disease. PLoS One, 7: e34939. doi: 10.1371/journal.pone.0034939.
- Li, D., Liu, C.-M., Luo, R., Sadakane, K. & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31, 1674–1676

- Li, L., Wen, X., Wen, Z., Chen, S., Wang, L., Wei, X. (2018). Evaluation of the biogenic amines formation and degradation abilities of Lactobacillus curvatus from Chinese bacon. Front Microbiol. 9:1015. doi: 10.3389/fmicb.2018.01015.
- Li, X., Li, C., Ye, H., Wang, Z., Wu, X., Han, Y. et al. (2019). Changes in the microbial communities in vacuum-packaged smoked bacon during storage. Food Microbiol. 77:26–37. doi: 10.1016/j.fm.2018.08.007.
- Liang, J.Q., Wong, S.H., Szeto, C.H., Chu, E.S., Lau, H.C., Chen, Y., et al. (2020). Fecal microbial DNA markers serve for screening colorectal neoplasm in asymptomatic subjects. J Gastroenterol Hepatol. Advance online publication. 10.1111/jgh.15171. doi:10.1111/jgh.15171
- Linden, S.K., Sutton, P., Karlsson, N.G., Korolik, V., McGuckin, M.A. (2008). Mucins in the mucosal barrier to infection. Mucosal Immunol. 1(3):183-197. doi:10.1038/mi.2008.5
- Liston, H.L., Markowitz, J.S., DeVane, C.L., (2001). Drug glucuronidation in clinical psychopharmacology. J Clin Psychopharmacol. 21(5):500-515. doi:10.1097/00004714-200110000-00008
- Lo, R., Turner, M. S., Weeks, M., and Bansal, N. (2016). Culture-independent bacterial community profiling of carbon dioxide treated raw milk. Int. J. Food Microbiol. 233, 81–89. doi: 10.1016/j.ijfoodmicro.2016.06.015
- LoGuidice, A., Wallace, B. D., Bendel, L., Redinbo, M. R. & Boelsterli, U. A. (2012). Pharmacologic targeting of bacterial β-glucuronidase alleviates nonsteroidal anti-inflammatory drug-induced enteropathy in mice. J. Pharmacol. Exp. Ther. **341**, 447–454
- Lombardi, P., Goldin, B., Boutin, E., & Gorbach, S. L. (1978). Metabolism of androgens and estrogens by human fecal microorganisms. Journal of steroid biochemistry, 9(8), 795–801. https://doi.org/10.1016/0022-4731(78)90203-0
- Louis P, Hold LG, Flint JH. (2014). The gut microbiota, bacterial metabolites and colorectal cancer. Nature Reviews Microbiology, 12, 661–672. doi: 10.1038/nrmicro3344
- Lozupone, C., Knight, R., (2007). Global patterns in bacterial diversity, PNAS 104, 11436-11440. doi: 10.1073/pnas.0611525104
- Lu J, Breitwieser FP, Thielen P, Salzberg SL. (2017). Bracken: estimating species

abundance in metagenomics data. PeerJ Computer Science 3:e104 https://doi.org/10.7717/peerj-cs.104

- Lubelski, J., de Jong, A., van Merkerk, R., Agustiandari, H., Kuipers, O.P., Kok, J. et al. (2006). LmrCD is a major multidrug resistance transporter in *Lactococcus lactis*. Mol Microbiol. 61:771-781. doi:10.1111/j.1365-2958.2006.05267.x
- MacFarlane GT and Cummings JH. (1991). The colonic flora, fermentation and large bowel digestive function. The Large Intestine: Physiology, Pathophysiology and Disease, 51-92. Raven Press, New York, USA.
- Macfarlane GT, Macfarlane S. (1997). Human colonic microbiota: ecology, physiology and metabolic potential of intestinal bacteria. Scandinavian Journal of Gastroenterology Suppl, 222:3-9.
- Mandoli, C. et al. (2010). Stem Cell Aligned Growth Induced by CeO2 Nanoparticles in PLGA Scaffolds with Improved Bioactivity for Regenerative Medicine. Adv. Funct. Mater. **20**, 1617–1624. doi:10.1002/adfm.200902363
- Margulies,M. (2006). Corrigendum: Genome sequencing in microfabricated high density picolitre reactors. Nature 441:502. doi: 10.1038/nature04726
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17: 10. doi:10.14806/ej.17.1.200.
- Martinez et al., (2013). Long-term temporal analysis of the human fecal microbiota revealed a stable core of dominant bacterial species. PLoS One. 2013 Jul 16;8(7):e69621.
- Masamune, H. (1934). Biochemical studies on carbohydrates. IV. On an enzyme which catalyses the hydrolysis of biosynthetic osides of glucuronic acid. J. Biochem. 19, 353–375.
- Matern, S., Matern, H., Farthmann, E. H., & Gerok, W. (1984). Hepatic and extrahepatic glucuronidation of bile acids in man. Characterization of bile acid uridine 5'-diphosphate-glucuronosyltransferase in hepatic, renal, and intestinal microsomes. The Journal of clinical investigation, 74(2), 402–410. https://doi.org/10.1172/JCI111435
- Maynard et al., (2012). Reciprocal interactions of the intestinal microbiota and immune system. Nature. 2012 Sep 13; 489(7415): 231–241. doi:

10.1038/nature11551.

- McBain, A. J., & Macfarlane, G. T. (1998). Ecological and physiological studies on large intestinal bacteria in relation to production of hydrolytic and reductive enzymes involved in formation of genotoxic metabolites. Journal of medical microbiology, 47(5), 407–416. https://doi.org/10.1099/00222615-47-5-407
- McCance and Widdowson. (2002). The composition of Foods, edition 6. Food Standards Agency.
- McDonald, D., Price, M.N., Goodrich, J., Nawrocki, E.P., DeSantis, T.Z., Probst, A., Andersen, G.L., Knight, R., Hugenholtz, P., (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6, 610-618. doi: 10.1038/ismej.2011.139
- McIntosh, F. M. et al. (2012). Phylogenetic distribution of genes encoding β-glucuronidase activity in human colonic bacteria and the impact of diet on faecal glycosidase activities. Environ. Microbiol. **14**, 1876–1887 doi: 10.1111/j.1462-2920.2012.02711.x.
- McIntyre A, Gibson PR, Young GP. (1993). Butyrate producing form dietary fibre and protection against large bowel cancer in a rat model. Gut, 34; 386-91. doi: 10.1136/gut.34.3.386.
- Meech, R., Miners, J.O., Lewis, B.C., MacKenzie, P.I., (2012). The glycosidation of xenobiotics and endogenous compounds: Versatility and redundancy in the UDP glycosyltransferase superfamily. Pharmacol Ther. 134(2):200-218. doi:10.1016/j.pharmthera.2012.01.009
- Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.P., Göker, M. (2013). Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 14:60. https://doi.org/10.1186/1471-2105-14-60
- Menzel, P., Ng, K. & Krogh, A. (2016). Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7, 11257. https://doi.org/10.1038/ncomms11257

- Milani, C., Hevia, A., Foroni, E., Duranti, S., Turroni, F., Lugli, G.A., et al. (2013). Assessing the fecal microbiota: an optimized ion torrent 16S rRNA gene-based analysis protocol. PLoS One. 8(7):e68739. doi:10.1371/journal.pone.0068739
- Miller, R.S., Hoskins, L.C. (1981). Mucin degradation in human colon ecosystems. Fecal population densities of mucin-degrading bacteria estimated by a "most probable number" method. Gastroenterology. 81(4):759-65.
- Minnullina, L.F., Pudova, D., Shagimardanova, E., Shigapova, L., Sharipova, M., Mardanova, A. (2019). Comparative genome analysis of uropathogenic Morganella morganii strains. Front. Cell. Infect. Microbiol. 9, 167. doi: 10.3389/fcimb.2019.00167
- Moura, A., Criscuolo, A., Pouseele, H., Maury, M.M., Leclercq, A., Tarr, C. et al. (2016). Whole genome-based population biology and epidemiological surveillance of Listeria monocytogenes. Nat. Microbiol. 2, 16185. https://doi.org/10.1038/nmicrobiol.2016.185.
- Murray, K. E., Shaw, K. J., Adams, R. F., & Conway, P. L. (1993). Presence of N-acyl and acetoxy derivatives of putrescine and cadaverine in the human gut. Gut, 34(4), 489–493. https://doi.org/10.1136/gut.34.4.489
- Nagar, S. & Blanchard, R. L. (2006). Pharmacogenetics of uridine diphosphoglucuronosyltransferase (UGT) 1A family members and its role in patient response to irinotecan. Drug Metab. Rev. **38**, 393–409 doi: 10.1080/03602530600739835
- Nakamura, J., Kubota, Y., Miyaoka, M., Saitoh, T., Mizuno, F., & Benno, Y. (2002). Comparison of four microbial enzymes in *Clostridia* and *Bacteroides* isolated from human feces. Microbiology and immunology, 46(7), 487–490. https://doi.org/10.1111/j.1348-0421.2002.tb02723.x
- Nawrocki EP, Eddy SR. (2013). Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics 29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509.
- Nguyen, N.P., Warnow, T., Pop, M., White, B., (2016). A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. NPJ Biofilms Microbiomes 2, 16004. https://doi.org/10.1038/npjbiofilms.2016.4.

- Nguyen, T.T., Tomberlin, J.K., Vanlaerhoven, S. (2015). Ability of Black Soldier Fly (Diptera: Stratiomyidae) Larvae to Recycle Food Waste. Environ Entomol. 44, 406–410. doi: 10.1093/ee/nvv002
- Nieminen T.T., Säde E., Endo A., Johansson P., Björkroth J. (2014) The Family Leuconostocaceae. In: Rosenberg E., DeLong E.F., Lory S., Stackebrandt E., Thompson F. (eds) The Prokaryotes. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-30120-9_208.
- Noecker, C., McNally, C. P., Eng, A. & Borenstein, E. (2017). High-resolution characterization of the human microbiome. Transl. Res. **179**, 7–23 doi: 10.1016/j.trsl.2016.07.012
- Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A. (2017). metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27:824-834. doi:10.1101/gr.213959.116
- NVWA. (2014). Advisory Report on the risks associated with the consumption of mass-reared insects. Netherlands Food and Consumer Product Safety Authority, Utrecht, The Netherlands; NVWA/BuRO/2014/2372. doi: 10.5281/zenodo.439001
- NVWA. (2019). Advice on animal and public health risks of insects reared on former foodstuffs as raw material for animal feed. Netherlands Food and Consumer Product Safety Authority, Utrecht, The Netherlands; TRCNVWA/2019/6200/EN
- O'Hara, C. M., Brenner, F. W., Miller, J. M. (2000). Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. Clin Microbiol. Rev. 13, 534–546. doi: 10.1128/CMR.13.4.534
- Orrhage and Nord, (1999). Factors controlling the bacterial colonization of the intestine in breastfed infants. Acta Paediatr Suppl. 1999 Aug;88(430):47-57. doi: 10.1111/j.1651-2227.1999.tb01300.x.
- Oshima, G. (1934). Biochemical studies on carbohydrates. XII. On beta-glucuronidase, 2nd communication. J. Biochem. 20, 361–370.
- Osimani, A., Milanović, V., Garofalo, C., Cardinali, F., Roncolini, A., Sabbatini, R. et al. (2018). Revealing the microbiota of marketed edible insects through PCR-DGGE, metagenomic sequencing and real-time PCR. Int J Food Microbiol. 276, 54–62. doi: 10.1016/j.ijfoodmicro.2018.04.013

- Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics. 31:3691-3693. https://doi.org/10.1093/bioinformatics/btv421
- Parente, E., Moles, M., Ricciardi, A. (1996). Leucocin F10, a bacteriocin from *Leuconostoc carnosum*. Int J Food Microbiol 33:231–243. https://doi.org/ 10.1016/0168-1605(96)01159-2.
- Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25, 1043-1055 doi:10.1101/gr.186072.114
- Pasolli, E., Asnicar, F., Manara, S., Zolfo, M., Karcher, N., Armanini, F. et al. (2019). Extensive Unexplored Human Microbiome Diversity Revealed by Over 150,000 Genomes from Metagenomes Spanning Age, Geography, and Lifestyle. Cell, 176(3), 649–662.e20. https://doi.org/10.1016/j.cell.2019.01.001
- Peekhaus, N., & Conway, T. (1998). What's for dinner?: Entner-Doudoroff metabolism in *Escherichia coli*. Journal of bacteriology, 180(14), 3495–3502. https://doi.org/10.1128/JB.180.14.3495-3502.1998
- Peet RK, (1974). The Measurement of Species Diversity. Annual Review of Ecology and Systematics. Vol. 5: 285-307
- Pellock, S. J., & Redinbo, M. R. (2017). Glucuronides in the gut: Sugar-driven symbioses between microbe and host. The Journal of biological chemistry, 292(21), 8569–8576. https://doi.org/10.1074/jbc.R116.767434
- Pereira, P.M., Vicente, A.F. (2013). Meat nutritional composition and nutritive role in the human diet. Meat Sci. 93:586-592. doi:10.1016/j.meatsci.2012.09.018
- Petersen TN, Brunak S, von Heijne G, Nielsen H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. https://doi.org/10.1038/nmeth.1701.
- Pinholt M, Gumpert H, Bayliss S, Nielsen JB, Vorobieva V, Pedersen M, Feil E, Worning P, Westh H. (2017). Genomic analysis of 495 vancomycin-resistant Enterococcus faecium reveals broad dissemination of a vanA plasmid in more than 19 clones from Copenhagen, Denmark. J Antimicrob Chemother 72:40–47. https://doi.org/10.1093/jac/dkw360.
- Pizzolato, J. F. & Saltz, L. B. (2003). The camptothecins. Lancet (London, England)

361, 2235–2242 doi: 10.1016/S0140-6736(03)13780-4.

- Png, C.W., Lindén, S.K., Gilshenan, K.S., Zoetendal, E.G., McSweeney, C.S., Sly, L.I., et al. (2010). Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am J Gastroenterol. 105(11):2420-2428. doi:10.1038/ajg.2010.281
- Pollet, R.M., D'Agostino, E.H., Walton, W.G., et al., 2017. An Atlas of β-Glucuronidases in the Human Intestinal Microbiome. Structure. 5(7):967-977.e5. doi:10.1016/j.str.2017.05.003
- Posthuma, C.C., Bader, R., Engelmann, R., Postma, P.W., Hengstenberg, W., Pouwels, P.H. (2002). Expression of the xylulose 5-phosphate phosphoketolase gene, xpkA, from *Lactobacillus pentosus* MD363 is induced by sugars that are fermented via the phosphoketolase pathway and is repressed by glucose mediated by CcpA and the mannose phosphoenolpyruvate phosphotransferase system. Appl Environ Microbiol. 68:831-837. doi: 10.1128/aem.68.2.831-837.2002.
- Pothakos, V., Stellato, G., Ercolini, D., and Devlieghere, F. (2015). Processing environment and ingredients are both sources of *Leuconostoc gelidum*, which emerges as a major spoiler in ready-to-eat meals. Appl. Environ. Microbiol. 81, 3529–3541. doi: 10.1128/AEM.03941-14
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2009). FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. Mol Biol Evol. 26:1641-1650, doi:10.1093/molbev/msp077.
- Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. Nature, 464, 59-65. doi: 10.1038/nature08821
- Quainoo, S., Coolen, J.P.M., van Hijum, S.A.F.T., Huynen, M.A., Melchers, W.J.G., van Schaik, W. et al. (2017). Whole-genome sequencing of bacterial pathogens: the future of nosocomial outbreak analysis. Clin. Microbiol. Rev. 30, 1015–1063. https://doi.org/10.1128/CMR.00016-17
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., Glöckner, F.O., (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 41,590-596 doi: 10.1093/nar/gks1219.
- Raftogianis, R., Creveling, C., Weinshilboum, R., & Weisz, J. (2000). Estrogen

metabolism by conjugation. Journal of the National Cancer Institute. Monographs, (27), 113–124. https://doi.org/10.1093/oxfordjournals.jncimonographs.a024234

- Raimondi, S., Luciani, R., Sirangelo, T.M., Amaretti, A., Leonardi, A., Ulrici, A., et al. (2019). Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: microbiota of sliced cooked ham in MAP. Int J Food Microbiol 289:200–208. https://doi.org/10.1016/j.ijfoodmicro.2018.09.017.
- Raimondi, S., Nappi, M.R., Sirangelo, T.M., Leonardi, A., Amaretti, A., Ulrici, A., et al. (2018). Bacterial community of industrial raw sausage packaged in modified atmosphere throughout the shelf life. Int J Food Microbiol 280:78–86. https://doi .org/10.1016/j.ijfoodmicro.2018.04.041.
- Raimondi, S., Righini, L., Candeliere, F., Musmeci, E., Bonvicini, F., Gentilomi, G., et al. (2019). Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of E. coli Isolated from the Feces of Healthy Subjects. Microorganisms. 7(8):251. doi:10.3390/microorganisms7080251
- Rajilic-Stojanovic et al., (2009). Development and application of the human intestinal tract chip, a phylogenetic microarray: analysis of universally conserved phylotypes in the abundant microbiota of young and elderly adults. Environ Microbiol. Jul. 2009;11(7):1736-51. https://doi.org/10.1111/j.1462-2920.2009.01900.x
- Rajilić-Stojanović, M., & de Vos, W. M. (2014). The first 1000 cultured species of the human gastrointestinal microbiota. FEMS microbiology reviews, 38(5), 996–1047. https://doi.org/10.1111/1574-6976.12075
- Rantsiou, K., Kathariou, S., Winkler, A., Skandamis, P., Saint-Cyr, M.J., Rouzeau-Szynalski, K. et al. (2017). Next generation microbiological risk assessment: opportunities of whole genome sequencing (WGS) for foodborne pathogen surveillance, source tracking and risk assessment. Int. J. Food Microbiol. (in press). doi:10.1016/j.ijfoodmicro.2017.11.007
- Richter, M., Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci USA. 106:19126-19131. doi: 10.1073/pnas.0906412106.
- Rideout, J. R., He, Y., Navas-Molina, J. A., Walters, W. A., Ursell, L. K., Gibbons, S. M., Chase, J., McDonald, D., Gonzalez, A., Robbins-Pianka, A., Clemente, J. C., Gilbert, J. A., Huse, S. M., Zhou, H. W., Knight, R., & Caporaso, J. G. (2014). Subsampled open-reference clustering creates consistent, comprehensive OTU

definitions and scales to billions of sequences. PeerJ, 2, e545. https://doi.org/10.7717/peerj.545

- Roberts, M. S., Magnusson, B. M., Burczynski, F. J., & Weiss, M. (2002). Enterohepatic circulation: physiological, pharmacokinetic and clinical implications. Clinical pharmacokinetics, 41(10), 751–790. https://doi.org/10.2165/00003088-200241100-00005
- Rodrigo-Torres, L., Yépez, A., Aznar, R., Arahal, D.R. (2019). Genomic insights into five strains of *Lactobacillus plantarum* with biotechnological potential isolated from chicha, a traditional maize-based fermented beverage from north-western Argentina. Front Microbiol. 10:2232. doi: 10.3389/fmicb.2019.02232.
- Rodrigues, M. X., Lima, S. F., Canniatti-Brazaca, S. G., and Bicalho, R. C. (2017). The microbiome of bulk tank milk: characterization and associations with somatic cell count and bacterial count. J. Dairy Sci. 100, 2536–2552. doi: 10.3168/jds.2016-11540
- Rodriguez-R, L.M., Konstantinidis, K.T. (2016). The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Preprints 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. PeerJ. 4:e2584. doi:10.7717/peerj.2584
- Ronaghi,M., Karamohamed, S., Pettersson, B., Uhlén,M., and Nyrén, P. L. (1996). Real-time DNA sequencing using detection of pyrophosphate release. Anal. Biochem. 242, 84–89. doi:10.1006/abio.1996.0432
- Rosen, M.J., Callahan, B.J., Fisher, D.S., Holmes, S.P., 2012. Denoising PCR-amplified metagenome data. BMC Bioinf. 13, 283. https://doi.org/10.1186/1471-2105-13-283.
- Rossi M, Martínez-Martínez D, Amaretti A, Ulrici A, Raimondi S, Moya A. (2016). Mining metagenomic whole genome sequences revealed subdominant but constant *Lactobacillus* population in the human gut microbiota. Environ Microbiol Rep. 2016 Jun;8(3):399-406. doi: 10.1111/1758-2229.12405.
- Russell, W. M., & Klaenhammer, T. R. (2001). Identification and cloning of gusA, encoding a new beta-glucuronidase from Lactobacillus gasseri ADH. Applied and environmental microbiology, 67(3), 1253–1261. https://doi.org/10.1128/AEM.67.3.1253-1261.2001

- Saitta, K. S. et al. (2014). Bacterial β-glucuronidase inhibition protects mice against enteropathy induced by indomethacin, ketoprofen or diclofenac: mode of action and pharmacokinetics. Xenobiotica. **44**, 28–35 doi: 10.3109/00498254.2013.811314.
- Salonen A, de Vos WM. (2014). Impact of diet on human intestinal microbiota and health. Annu Rev Food Sci Technol. 2014;5:239-62. doi: 10.1146/annurev-food-030212-182554
- Salyers, A. A., & O'Brien, M. (1980). Cellular location of enzymes involved in chondroitin sulfate breakdown by *Bacteroides thetaiotaomicron*. Journal of bacteriology, 143(2), 772–780. https://doi.org/10.1128/JB.143.2.772-780.1980
- Samelis, J., Björkroth, J., Kakouri, A., Rementzis, J. (2006). *Leuconostoc carnosum* associated with spoilage of refrigerated whole cooked hams in Greece. J Food Prot 69:2268 –2273. https://doi.org/10.4315/0362-028x -69.9.2268.
- Scarpignato, C. & Hunt, R. H. 2010). Nonsteroidal antiinflammatory drug-related injury to the gastrointestinal tract: clinical picture, pathogenesis, and prevention. Gastroenterol. Clin. North Am. **39**, 433–464 doi: 10.1016/j.gtc.2010.08.010.
- Schepers E, Meert N, Glorieux G, Goeman J, Van der Eycken J, Vanholder R. (2007). P-cresylsulphate, the main in vivo metabolite of p-cresol, activates leucocyte free radical production. Nephrology Dialysis Transplantation, 22(2):592-6. doi: 10.1093/ndt/gfl584.
- Schleifer KH (2009) Family V. Leuconostocaceae fam. nov. In: De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer KH, Whitman WB (eds) Bergey's manual of systematic bacteriology (The Firmicutes), vol 3, 2nd edn. Springer, Dordrecht/Heidelberg/London/New York, p 62.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B. et al.(2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. 2009, 75, 7537–7541. doi: 10.1128/AEM.01541-09
- Scott, K. P., Gratz, S. W., Sheridan, P. O., Flint, H. J., and Duncan, S. H. (2013). The influence of diet on the gut microbiota. Pharmacol. Res. 69, 52–60. doi: 10.1016/j.phrs.2012.10.020

- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics. 30:2068-2069. https://doi.org/10.1093/bioinformatics/btu153
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., et al. (2011). Metagenomic biomarker discovery and explanation. Genome Biol. 12(6):R60. doi:10.1186/gb-2011-12-6-r60
- Segata, N., Waldron, L., Ballarin,i A., Narasimhan, V., Jousson, O., Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat Methods. 9(8):811-814. doi:10.1038/nmeth.2066
- Seidel ER, Haddox MK, Johnson LR. (1984). Polyamines in the response to intestinal obstruction. American Journal of Physiology, 246: G649.
- Sekelja et al., (2011). Unveiling an abundant core microbiota in the human adult colon by a phylogroup-independent searching approach. ISME J. 2011 Mar; 5(3): 519–531. doi: 10.1038/ismej.2010.129
- Sender, R., Fuchs, S. & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. PLoS One 1–14 doi:10.1371/journal.pbio.1002533
- Sharma, D., Sharma, P., Soni, P. (2017). First case report of *Providencia rettgeri* neonatal sepsis. BMC Res. Notes, 10, 536. doi: 10.1186/s13104-017-2866-4
- Sharon, G. et al. (2014). Specialized metabolites from the microbiome in health and disease. Cell Metab. **20**, 719–730 doi: 10.1016/j.cmet.2014.10.016
- Shaw, B.G., Harding, C.D. (1989). *Leuconostoc gelidum* sp. nov. and *Leuconostoc carnosum* sp. nov. from chill-stored meats. Int J Syst Bacteriol 39:217–223. doi.org/10.1099/00207713-39-3-217
- Shawki, A., and McCole, D.F. (2016). Mechanisms of intestinal epithelial barrier dysfunction by adherent-invasive Escherichia coli. Cell Mol Gastroenterol Hepatol. 3(1):41-50. doi:10.1016/j.jcmgh.2016.10.004
- Shelomi, M., Wu, M.K., Chen, S.M., Huang, J.J., Burke, C.G. (2020). Microbes associated with black soldier fly (Diptera: Stratiomiidae) degradation of food waste. Environ. Entomol. 49, 405-411. doi: 10.1093/ee/nvz164
- Shuker DEG, Margison GP. (1997). Nitrosated glycine derivatives as a potential source of O6-methylguanine in DNA Cancer Res. 57: 366-369.

- Siguier, P., Perochon, J., Lestrad,e L., Mahillon, J., Chandler, M. (2006). ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 34:D32-6. doi: 10.1093/nar/gkj014.
- Silvester KR, Bingham SA, Pollock JRA, Cummings JH, O'Neill IK. (1997). Effect of meat and resistant starch on fecal excretion of apparent N-nitroso compounds and ammonia from the human large bowel. Nutrition and Cancer, Pages 13-23.
- Smith EA, Macfarlane GT. (1996). Enumeration of human colonic bacteria producing phenolic and indolic compounds: effects of pH, carbohydrate availability and retention time on dissimilatory aromatic amino acid metabolism. Journal of Applied Bacteriology, 81(3):288-302
- Smith TA. (1980). Amines in food. Food Chemistry, 6: 169.
- Smith, E. A., and Macfarlane, G. T. (1998). Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. FEMS Microbiol. Ecol. 25, 355–368. doi: 10.1111/j.1574-6941.1998.tb00487.x
- Smith, P. M., Howitt, M. R., Panikov, N., Michaud, M., Gallini, C. A., Bohlooly-Y, M., Glickman, J. N., & Garrett, W. S. (2013). The microbial metabolites, short-chain fatty acids, regulate colonic Treg cell homeostasis. Science (New York, N.Y.), 341(6145), 569–573. https://doi.org/10.1126/science.1241165
- Spanogiannopoulos, P., Bess, E. N., Carmody, R. N. & Turnbaugh, P. J. (2016). The microbial pharmacists within us: a metagenomic view of xenobiotic metabolism. Nat. Rev. Microbiol. **14**, 273–287 https://doi.org/10.1038/nrmicro.2016.17
- Stiles, M.E. (1994). Bacteriocins produced by *Leuconostoc* species. J Dairy Sci. 77:2718-2724. https://doi.org/10.3168/jds.S0022-0302(94)77214-3
- Suau, A., Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M. D. et al. (1999). Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Applied and environmental microbiology, 65(11), 4799–4807. https://doi.org/10.1128/AEM.65.11.4799-4807.1999
- Tailford, L.E., Crost, E.H., Kavanaugh, D., Juge, N. (2015). Mucin glycan foraging in the human gut microbiome. Front Genet. 6:81. doi:10.3389/fgene.2015.00081
- Tamm A, Villako K. (1971). Urinary volatile phenols in patients with intestinal

obstruction. Scandinavian Journal of Gastroenterology, 6(1):5-8.

- Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J. P., Ugarte, E., Muñoz-Tamayo, R., Paslier, D. L., Nalin, R., Dore, J., & Leclerc, M. (2009). Towards the human intestinal microbiota phylogenetic core. Environmental microbiology, 11(10), 2574–2584. https://doi.org/10.1111/j.1462-2920.2009.01982.x
- Tatusov, R. L., Koonin, E. V., & Lipman, D. J. (1997). A genomic perspective on protein families. Science (New York, N.Y.), 278(5338), 631–637. https://doi.org/10.1126/science.278.5338.631
- Terova, G., Rimoldi, S., Ascione, C., Gini, E., Ceccotti, C., Gasco, L. (2019). Rainbow trout (*Oncorhynchus mykiss*) gut microbiota is modulated by insect meal from Hermetia illucens prepupae in the diet. Rev. Fish Biol. Fisher. 29, 465–486. doi: 10.1007/s11160-019-09558-y
- Tettelin, H., Riley, D., Cattuto, C., Medini, D. (2008). Comparative genomics: the bacterial pan-genome. Curr Opin Microbiol. 11:472-477. doi: 10.1016/j.mib.2008.09.006
- Togo, A.H., Diop, A., Dubourg, G., Khelaifia, S., Richez, M., Armstrong, N., et al. (2019). *Anaerotruncus massiliensis* sp. nov., a succinate-producing bacterium isolated from human stool from an obese patient after bariatric surgery. New Microbes New Infect. 29:100508. doi:10.1016/j.nmni.2019.01.004
- Topisirovic L, Kojic M, Fira D, Golic N, Strahinic I, Lozo J. (2006). Potential of lactic acid bacteria isolated from specific natural niches in food production and preservation. Int J Food Microbiol. 2006 Dec 1;112(3):230-5. doi: 10.1016/j.ijfoodmicro.2006.04.009.
- Troge, A., Scheppach, W., Schroeder, B.O., Rund, S.A., Heuner, K., Wehkamp, J., et al. (2012). More than a marine propeller--the flagellum of the probiotic Escherichia coli strain Nissle 1917 is the major adhesin mediating binding to human mucus. Int J Med Microbiol. 302(7-8):304-314. doi:10.1016/j.ijmm.2012.09.004
- Truong, D.T., Franzosa, E.A., Tickle, T.L., et al. (2015). MetaPhlAn2 for enhanced metagenomic taxonomic profiling. Nat Methods. 12(10):902-903. doi:10.1038/nmeth.3589

- Truzzi, C., Giorgini, E., Annibaldi, A., Antonucci, M., Illuminati, S., Scarponi, G., et al. (2020). Fatty acids profile of black soldier fly (Hermetia illucens): Influence of feeding substrate based on coffee-waste silverskin enriched with microalgae. Anim. Feed Sci. Technol. 2020, 259, 114309. doi: 10.1016/j.anifeedsci.2019.114309
- Turnbaugh et al., (2009). A core gut microbiome in obese and lean twins. Nature. 2009 Jan 22;457(7228):480-4.
- Turnbaugh, P., Ley, R., Hamady, M. et al. (2007). The Human Microbiome Project. Nature 449, 804–810 https://doi.org/10.1038/nature06244
- Ugalde-Silva, P., Gonzalez-Lugo, O., Navarro-Garcia, F. (2016). Tight junction disruption induced by type 3 secretion system effectors injected by enteropathogenic and enterohemorrhagic *Escherichia coli*. Front Cell Infect Microbiol. 6:87. doi:10.3389/fcimb.2016.00087
- van Heel, A.J., de Jong, A., Song, C., Viel, J.H., Kok, J., Kuipers, O.P. (2018). BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res. 46:W278-281. doi: 10.1093/nar/gky383.
- Van Hoorde, K., Butler, F. (2018). Use of next-generation sequencing in microbial risk assessment. EFSA Journal 16 (S1). https://doi.org/10.2903/j.efsa.2018.e16086
- van Laack, R.L.J.M., Schillinger, U., Holzapfel, W.H. (1992). Characterization and partial purification of a bacteriocin produced by *Leuconostoc carnosum* LA44A. Int J Food Microbiol 16:183–195. https://doi.org/10.1016/0168 -1605(92)90079-i.
- Vasilopoulos, C., De Mey, E., Dewulf, L., Paelinck, H., De Smedt, A., Vandendriessche, F. et al. (2010). Interactions between bacterial isolates from modified-atmosphere-packaged artisan-type cooked ham in view of the development of a bioprotective culture. Food Microbiol. 27:1086–1094. doi:10.1016/j.fm.2010.07.013
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A. et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. Science (New York, N.Y.), 304(5667), 66–74. https://doi.org/10.1126/science.1093857
- Visek WK. (1978). Diet and cell growth modulation by ammonia. American Journal of Clinical Nutrition, 31: S216-S220.
- Vítek, L., & Carey, M. C. (2003). Enterohepatic cycling of bilirubin as a cause of 'black' pigment gallstones in adult life. European journal of clinical

investigation, 33(9), 799–810. https://doi.org/10.1046/j.1365-2362.2003.01214.x

- Vítek, L., & Ostrow, J. D. (2009). Bilirubin chemistry and metabolism; harmful and protective aspects. Current pharmaceutical design, 15(25), 2869–2883. https://doi.org/10.2174/138161209789058237
- von Meijenfeldt, F.A.B., Arkhipova, K., Cambuy, D.D. et al., (2019). Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. Genome Biol 20, 217. https://doi.org/10.1186/s13059-019-1817-x
- Vrieze, A., de Groot, P. F., Kootte, R. S., Knaapen, M., van Nood, E., & Nieuwdorp, M. (2013). Fecal transplant: a safe and sustainable clinical therapy for restoring intestinal microbial balance in human disease? Best practice & research. Clinical gastroenterology, 27(1), 127–137. https://doi.org/10.1016/j.bpg.2013.03.003
- Wagner, C.E., Wheeler, K.M., Ribbeck, K. (2018). Mucins and Their Role in Shaping the Functions of Mucus Barriers. Annu Rev Cell Dev Biol. 34:189-215. doi:10.1146/annurev-cellbio-100617-062818
- Wallace, B. D. et al. (2010).Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. Science (80-.). **330**, 831–836 doi: 10.1126/science.1191175
- Wallace, B. D. et al. (2015). Structure and Inhibition of Microbiome β-Glucuronidases Essential to the Alleviation of Cancer Drug Toxicity. Chem. Biol. **22**, 1238–1249 doi: 10.1016/j.chembiol.2015.08.005
- Wan, X., Saris, P.E.J., Takala, T.M. (2015). Genetic characterization and expression of leucocin B, a class IId bacteriocin from Leuconostoc carnosum 4010. Res Microbiol 166(6):494-503. doi:10.1016/j.resmic.2015.04.003.
- Wang, Y. S., Shelomi, M. (2017). Review of Black Soldier Fly (*Hermetia illucens*) as Animal Feed and Human Food. Foods. 6, 91. doi: 10.3390/foods6100091
- Weber FL, Banwell JG, Fresard KM, Cummings JH. (1987). Nitrogen in faecal bacteria, fiber and soluble fractions of patients with cirrhosis: effects of lactulose and lactulose plus neomycin. Journal of Laboratory and Clinical Medicine, 110: 259.
- Wexler, H.M. (2007). Bacteroides: the good, the bad, and the nitty-gritty. Clin Microbiol Rev. 20(4):593-621. doi:10.1128/CMR.00008-07
- Whittaker, R. H. (1960) Vegetation of the Siskiyou Mountains, Oregon and California. Ecological Monographs, 30, 279–338. doi:10.2307/1943563

- Wilke, A., Bischof, J., Gerlach, W., Glass, E., Harrison, T., Keegan, K.P. et al. (2016). The MG-RAST metagenomics database and portal in 2015. Nucleic Acids Res. 44 (D1), D590–D594. https://doi.org/10.1093/nar/gkv1322
- Williams, T. R., Moyne, A. L., Harris, L. J., and Marco, M. L. (2013). Season, irrigation, leaf age, and Escherichia coli inoculation influence the bacterial diversity in the lettuce phyllosphere. PLoS ONE 8:e68642. doi: 10.1371/journal.pone.0068642
- Willing et al., (2010). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes. Gastroenterology Dec 2010;139(6):1844-1854. https://doi.org/10.1053/j.gastro.2010.08.049
- Wilson, K. J., Hughes, S. G., & Jefferson, R. A. (1992). The *Escherichia coli* gus operon: induction and expression of the gus operon in *E. coli* and the occurrence and use of GUS in other bacteria. In GUS protocols. Using the GUS gene as reporter of gene expression (pp. 7-22). Academic Press Inc.
- Winter, J., & Bokkenheuser, V. D. (1987). Bacterial metabolism of natural and synthetic sex hormones undergoing enterohepatic circulation. Journal of steroid biochemistry, 27(4-6), 1145–1149. https://doi.org/10.1016/0022-4731(87)90201-9
- Wood, D.E., Salzberg, S.L. (2014). Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 15, R46 https://doi.org/10.1186/gb-2014-15-3-r46
- Wu, S., Zhu, Z., Fu, L., Niu, B., Li W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. BMC Genomics. 12:444. doi: 10.1186/1471-2164-12-444.
- Wu, Y., Simmons, B.A., Singer, S.W., (2016). MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 32(4):605-607. doi:10.1093/bioinformatics/btv638
- Wynants, E., Frooninckx, L., Crauwels, S., Verreth, C., De Sme,t J., Sandrock, C. et al. (2018). Assessing the microbiota of black soldier fly larvae (*Hermetia illucens*) reared on organic waste streams on four different locations at laboratory and large scale. Microbial Ecology, 77, 913–930. doi: 10.1007/s00248-018-1286-x
- Yamanaka, H., Nakajima, M., Katoh, M., & Yokoi, T. (2007). Glucuronidation of thyroxine in human liver, jejunum, and kidney microsomes. Drug metabolism and

disposition: the biological fate of chemicals, 35(9), 1642–1648. https://doi.org/10.1124/dmd.107.016097

- Zarantoniello, M., Randazzo, B., Truzzi, C., Giorgini, E., Marcellucci, C., Vargas-Abúndez, J. A., et al. (2019). A six-months study on Black Soldier Fly (*Hermetia illucens*) based diets in zebrafish. Sci. Rep., 9, 8598 doi: 10.1038/s41598-019-45172-5

- Zerbino, D.R., Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821– 829. https://doi.org/10.1101/gr.074492.107.

- Zheng J, Wittouck S, Salvetti E, Franz CMAP, Harris HMB, Mattarelli P, O'Toole PW, Pot B, Vandamme P, Walter J, Watanabe K, Wuyts S, Felis GE, Gänzle MG, Lebeer S. A taxonomic note on the genus *Lactobacillus*: Description of 23 novel genera, emended description of the genus *Lactobacillus* Beijerinck 1901, and union of Lactobacillaceae and Leuconostocaceae. Int J Syst Evol Microbiol. 2020 Apr;70(4):2782-2858. doi: 10.1099/ijsem.0.004107.

- Zheng, L., Crippen, T.L., Singh, B., Tarone, A.M., Dowd, S., Yu, Z. et al. (2013). A survey of bacterial diversity from successive life stages of black soldier fly (Diptera: Stratiomyidae) by using 16S rDNA pyrosequencing. J. Med. Entomol. 2013, 50, 647–658. doi: 10.1603/ME12199

- Zhou, J.C., Zhang, X.W. (2019). *Akkermansia muciniphila*: a promising target for the therapy of metabolic syndrome and related diseases. Chin J Nat Med. 17(11):835-841. doi:10.1016/S1875-5364(19)30101-3

- Zikmanis P, Brants K, Kolesovs S, Semjonovs P. (2020). Extracellular polysaccharides produced by bacteria of the Leuconostoc genus. World J Microbiol Biotechnol. 2020 Sep 29;36(11):161. doi: 10.1007/s11274-020-02937-9.

PUBLICATIONS

1. Raimondi, S., Amaretti, A., Gozzoli, C., Simone, M., Righini, L., **Candeliere, F**., Brun P., Ardizzoni A., Colombari B., Paulone S., Castagliuolo I., Cavalieri D., Blasi E., Rossi M., Peppoloni S. (**2019**). Longitudinal survey of fungi in the human gut: ITS profiling, phenotyping and colonization. ***FRONTIERS IN MICROBIOLOGY***, *10*, 1575. doi.org/10.3389/fmicb.2019.01575

2. Raimondi, S., Righini, L., **Candeliere, F**., Musmeci, E., Bonvicini, F., Gentilomi, G., Starčič Erjavec M., Amaretti A., Rossi, M. (**2019**). Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of E. coli Isolated from the Feces of Healthy Subjects. ***MICROORGANISMS***, *7*(8), 251. doi.org/10.3390/microorganisms7080251

3. **Candeliere, F**., Raimondi, S., Spampinato, G., Tay, M. Y. F., Amaretti, A., Schlundt, J., Rossi, M. (**2020**). Draft Genome Sequences of 12 Leuconostoc carnosum Strains Isolated from Cooked Ham Packaged in a Modified Atmosphere and from Fresh Sausages. ***MICROBIOLOGY RESOURCE ANNOUNCEMENTS***, *9*(2). doi.org/10.1128/MRA.01247-19

4. Amaretti, A., Righini, L., **Candeliere, F**., Musmeci, E., Bonvicini, F., Gentilomi, G. A., Rossi, M., Raimondi, S. (**2020**). Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of Non-Escherichia coli Enterobacterales from the Gut Microbiota of Healthy Subjects. ***INTERNATIONAL JOURNAL OF MOLECULAR SCIENCES***, 21(5), 1847. doi.org/10.3390/ijms21051847

5. Raimondi, S., Spampinato, G., Macavei, L. I., Lugli, L., **Candeliere, F**., Rossi, M., Maistrello, L., Amaretti, A. (**2020**). Effect of Rearing Temperature on Growth and Microbiota Composition of Hermetia illucens. ***MICROORGANISMS,*** *8*(6), 902 doi.org/10.3390/microorganisms8060902

6. **Candeliere, F**., Raimondi, S., Spampinato, G., Tay, M. Y. F., Amaretti, A., Schlundt, J., Rossi, M. (**2020**). Comparative genomics of *Leuconostoc carnosum.* Submitted to ***FRONTIERS IN MICROBIOLOGY,*** **accepted the 04/12/2020. In press.**

REPORT OF ACTIVITIES

1st Year

PARTICIPATION TO THE SCHOOL COURSES

Date: 16/01/18; 17/01/18; 18/01/18; 06/02/18; 07/02/18; 08/02/18

Title/subject: Scientific English

Date: 29/05/18

Title/subject: The model organism *Saccharomyces cerevisiae* mitochondrial inheritance as case study

Date: 05/07/18

Title/subject: Microbial Collection

Date: 12/09/18; 13/09/18

Title/subject: Colour and chemical imaging: RGB and hyperspectral image analysis for food monitoring

Date: 21/09/18

Title/subject: Model Plants I

Date: 02/10/18

Title/subject: Model Plants II

Date: 03/10/18

Title/subject: Innovation in food yeast starter cultures: current state, perspectives and limits

Date: 09/10/18; 10/10/18

Title/subject: Infrared spectroscopy in food analysis

PARTICIPATION TO SYMPOSIA

Date: 17-19/05/18

Location: Cortona

Title: Cortona Procarioti 2018

2nd Year

PARTICIPATION TO THE SCHOOL COURSES

Date: 12/11/2018

Title/subject: Medicinal plants as sources of potential pharmaceutical products

Date: 13/11/2018

Title/subject: Cameroonian wild mushrooms from the Bamoun Region: Nutritional and Medicinal Properties

Date: 18/09/2019

Title/subject: Microbial Biotechnologies for Biorefineries

Date: 18-19/09/2019

Title/subject: Food bioactive compounds

Date: 04/10/2019

Title/subject: Animal models

PARTICIPATION TO SYMPOSIA

Date: 12-14/06/2019

Location: Nanyang Technological University, Singapore

Title: 12th Global Microbial Identifier Meeting

Date: 18/01/2019

Location: Nanyang Technological University, Singapore

Title: iFood 2019

PERIODS OF STUDY ABROAD

Where: Nanyang Technological University, Singapore

How long: 14/01/2019 - 30/06/2019

3rd Year

PARTICIPATION TO THE SCHOOL COURSES

Date: 10/06/20; 12/06/20; 16/06/20

Title/subject: Applications of the multivariate analysis in the agri-food context

Date: 16/09/20; 23/09/20; 30/09/20

Title/subject: Introduction to MATLAB environment

frontiers in Microbiology

ORIGINAL RESEARCH
published: 10 July 2019
doi: 10.3389/fmicb.2019.01575

Longitudinal Survey of Fungi in the Human Gut: ITS Profiling, Phenotyping, and Colonization

Stefano Raimondi[1†], Alberto Amaretti[1†], Caterina Gozzoli[1], Marta Simone[1], Lucia Righini[1], Francesco Candeliere[1], Paola Brun[2], Andrea Ardizzoni[3], Bruna Colombari[3], Simona Paulone[3], Ignazio Castagliuolo[2], Duccio Cavalieri[4], Elisabetta Blasi[3], Maddalena Rossi[1] and Samuele Peppoloni[3]*

[1] Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy, [2] Department of Molecular Medicine, University of Padua, Padua, Italy, [3] Department of Surgery, Medicine, Dentistry and Morphological Sciences with Interest in Transplant, Oncology and Regenerative Medicine, University of Modena and Reggio Emilia, Modena, Italy, [4] Department of Biology, University of Florence, Firenze, Italy

OPEN ACCESS

Edited by:
Jessy L. Labbé,
Oak Ridge National Laboratory (DOE), United States

Reviewed by:
Mira Edgerton,
University at Buffalo, United States
Daniel Prieto,
Complutense University of Madrid, Spain

***Correspondence:**
Maddalena Rossi
maddalena.rossi@unimore.it

[†] These authors have contributed equally to this work

Specialty section:
This article was submitted to Systems Microbiology, a section of the journal Frontiers in Microbiology

Received: 25 February 2019
Accepted: 25 June 2019
Published: 10 July 2019

Citation:
Raimondi S, Amaretti A, Gozzoli C, Simone M, Righini L, Candeliere F, Brun P, Ardizzoni A, Colombari B, Paulone S, Castagliuolo I, Cavalieri D, Blasi E, Rossi M and Peppoloni S (2019) Longitudinal Survey of Fungi in the Human Gut: ITS Profiling, Phenotyping, and Colonization. Front. Microbiol. 10:1575. doi: 10.3389/fmicb.2019.01575

The fungal component of the intestinal microbiota of eight healthy subjects was studied over 12 months using metagenome survey and culture-based approaches. *Aspergillus*, *Candida*, *Debaryomyces*, *Malassezia*, *Penicillium*, *Pichia*, and *Saccharomyces* were the most recurrent and/or dominant fungal genera, according to metagenomic analysis. The biodiversity of fungal communities was lower and characterized by greater unevenness, when compared to bacterial microbiome. The dissimilarities both among subjects and over the time within the same subject suggested that most of the fungi passed through the gastro-intestinal tract (GIT) without becoming stable colonizers. Certain genera, such as *Aspergillus* and *Penicillium*, were isolated in a minority of cases, although they recurred abundantly and frequently in the metagenomics survey, likely being environmental or food-borne fungi that do not inhabit the GIT. *Candida* genus was recurrently detected. *Candida albicans* isolates dominated among the cultivable mycobiota and longitudinally persisted, likely as commensals inhabiting the intestine or regularly reaching it from *Candida*-colonized districts, such as the oral cavity. Other putative colonizers belonged to *Candida zeylanoides*, *Geotrichum candidum*, and *Rhodotorula mucilaginosa*, with persisting biotypes being identified. Phenotyping of fungal isolates indicated that *C. albicans* adhered to human epithelial cells more efficiently and produced greater amounts of biofilm *in vitro* than non-*albicans Candida* (NAC) and non-*Candida* fungi (NCF). The *C. albicans* isolates also induced the highest release of HBD-2 by human epithelial cells, further differing from NAC and NCF. Nine representative isolates were administered to mice to evaluate the ability to colonize the intestine. Only two out of three *C. albicans* strains persisted in stools of animals 2 weeks after the end of the oral administration, whereas NAC and NCF did not. These results confirm the allochthonous nature of most the intestinal fungi, while *C. albicans* appears to be commonly involved in stable colonization. A combination of specific genetic

features in the microbe and in the host likely allow colonization from fungi normally present solely as passengers. It remains to be established if other species identified as potential colonizers, in addition to *Candida*, are true inhabitants of the GIT or rather reach the intestine spreading from other body districts.

Keywords: gut microbiota, fungi, colonization, metagenomics, *Candida*

INTRODUCTION

The complex microbial community hosted in the gastrointestinal tract (GIT) of humans and animals is composed of bacteria, archaea, fungi, protozoa, and viruses (Qin et al., 2010; Huseyin et al., 2017). The complexity of the GI bacterial population, accounting up to 10^{12} microorganisms per gram of content, has been widely investigated, while abundance, role, and diversity of fungi in the human gut are still understudied (Huseyin et al., 2017). Cultivable fungi in feces range between 10^2 and 10^6 cfu/g (Simon and Gorbach, 1984), representing a minor component of gut microbiota, and their genes cover less than 0.1% of the whole microbial metagenome (Qin et al., 2010). Although several fungal species are harbored in the normal GIT their potential role in host health status have been only partially investigated (Seed, 2014; Suhr and Hallen-Adams, 2015; Strati et al., 2016; Nash et al., 2017; Auchtung et al., 2018; Borges et al., 2018).

Most studies on the human mycobiome have been focused on opportunistic pathogens, such as *Candida* species, and in particular *Candida albicans* (Mayer et al., 2013; Merseguel et al., 2015; Cavalieri et al., 2018). This species colonizes oropharynx, genital, and gastrointestinal mucosa of 30–70% of healthy individuals, and colonization can evolve into infection under certain circumstances, such as decay of host immune defense or structural impairment of muco-cutaneous barriers (Brandt and Lockhart, 2012). Disruption of mucosal barriers and host immune defenses impairment may lead to endogenous candidiasis arising from commensal strains (Yan et al., 2013). Indeed, *C. albicans* is the most virulent species of the genus *Candida*, due to several virulence factors and fitness attributes that contribute to the pathogenic potential (Cavalieri et al., 2018). The first step in the infection onset is the interaction between *Candida* and host cells (Moyes et al., 2015). *C. albicans*, as well as other *Candida* spp., expresses several adhesins, which bind to extracellular proteins of epithelial cells. Adhesion of fungi to the mucosal cell membrane may stimulate biofilm formation, that increases the resistance to host immune defenses and antifungal drugs. As many opportunistic pathogens, *C. albicans* takes advantage from biofilm formation to escape host immune defenses and antifungal therapy as well (Tsui et al., 2016). Other *Candida* species, such as *C. dubliniensis*, *C. glabrata*, *C. guilliermondii*, *C. kefyr*, *C. krusei, C. tropicalis*, and *C. parapsilosis* can reside as commensal in diverse niches of the human body and may be responsible of biofilm-associated severe infections (Muadcheingka and Tantivitayakul, 2015). Currently, the significant increase of infections caused by other *Candida* spp. is matter of clinical concern, because of their resistance to conventional antifungal agents (Merseguel et al., 2015; Wu et al., 2017).

Nowadays, high throughput next generation sequencing (NGS) has become a common technology for microbiome studies and has overtaken traditional microbiological methods for the identification of members of the microbial communities. The knowledge on the bacterial component of the gut microbiota has received a major impulse from 16S rRNA gene profiling. Likewise, NGS targeting the Internal Transcribed Spacers (ITS) within eukaryotic rRNA genes allows to characterize the intestinal fungal population (Tang et al., 2015). Studies investigating the composition of gut mycobiota identified tens of fungal operational taxonomic units (OTUs), most of which ascribed to *Aspergillus*, *Candida*, *Cladosporium*, *Clavispora*, *Cyberlindnera*, *Debaryomyces*, *Galactomyces*, *Malassezia*, *Penicillium*, *Pichia*, *Rhodotorula*, and *Saccharomyces* (Hoffmann et al., 2013; Seed, 2014; Suhr and Hallen-Adams, 2015; Strati et al., 2016; Suhr et al., 2016; Nash et al., 2017; Auchtung et al., 2018; Borges et al., 2018). Recently, a strain specific analysis using whole genome sequencing of single isolates was applied to study strain specific features associated with the ability of *Saccharomyces cerevisiae* to colonize the human gut, showing that strains from different origin differ in the ability to adapt to the gut environment (Ramazzotti et al., 2019).

Several factors affect the description of the mycobiota by culture-independent sequencing, including non-standardized methods of DNA preparation, the target of primers for amplification, the consistency of taxonomic information in reference databases, and the characteristics of the cohort (e.g., geographical location, diet, and climate) (Nash et al., 2017). In particular, the selection of the best markers for profiling fungi is debated, since ITS1 has been regarded as the best target for profiling fungi, while the most recent studies indicate ITS2 as more accurate (Wang et al., 2015; Yang et al., 2018). However, the clustering and taxonomic capacities did not differ between ITS1 and ITS2, that generate similar patterns of community structure (Yang et al., 2018).

The ecological and functional role of intestinal fungi, the source (allochthones or autochtones), the interaction with the host, and the capability to permanently colonize or to transit through the GIT are topics that still deserve deep investigation (Huseyin et al., 2017). The aim of the present study was to investigate the presence and persistence of fungi in the feces of eight healthy subjects using a combination of culture-based approaches and metagenomics. The fungal isolates were phenotypically characterized, assessing their ability to colonize the murine GIT, adhere to human Caco-2 cells and induce the release of human β-defensin 2 (HBD2), and to form biofilm *in vitro*. In particular, the behavior of *C. albicans*, non-*albicans Candida* (NAC), and non-*Candida* fungi (NCF) was compared.

MATERIALS AND METHODS

ITS and 16S Amplicon Libraries, Sequencing, and Analysis

DNA was extracted from feces using the FastDNA SPIN Kit for Feces (MP-Biomedicals, United States). DNA quality was checked on 1% agarose gel TAE 1X and quantified with a NanoDrop spectrophotometer. Fungal ITS1 rDNA region was amplified with a fusion primer set, consisting of 18SF and 5.8S1R primers (5′-GTAAAAGTCGTAACAAGGTTTC-3′ and 5′-GTTCAAAGAYTCGATGATTCAC-3′, Findley et al., 2013) coupled with adaptors, key sequence and barcode, according to 454 Sequencing System Guidelines for Amplicon Experimental Design (Roche, Switzerland). Bacterial V1-V3 hypervariable regions of the 16S rRNA gene were generated with universal 27-Forward and 533-Reverse primers (5′-AGAGTTTGATCMTGGCTCAG-3′ and 5′-TTACCGCGGCTGCTGGCAC-3′, Barelli et al., 2015) coupled with adaptors, key sequence and barcode (Roche, Switzerland). ITS1 and V1-V3 regions were amplified according to Strati et al. (2016) and Barelli et al. (2015), respectively. The PCR products were electrophoresed, cleaned using the AMPure XP beads kit (Beckman Coulter, United States), and quantified via quantitative PCR using 454 Kapa Library Quantification Kit (KAPA Biosystems, United States).

Equimolar amounts of purified amplicons were pooled into ITS1 and 16S amplicon libraries, that were subjected to pyrosequencing on the Roche GS FLX+ system using the XL+ chemistry (Roche, Switzerland). Pyrosequencing reads were processed using the MICCA pipeline for trimming and quality filtering (Albanese et al., 2015). Quality filtered sequences were processed with Qiime 2 pipeline (Bolyen et al., 2018) for closed-reference OTU picking, taxonomy assignation, and diversity metrics computation. For the fungal microbiota, the OTU picking and taxonomy assignation were carried out utilizing UNITE ver. 7.2 Dynamic Classifier[1] as reference database with the similarity threshold set at 0.97. For the bacterial microbiota, the closed-reference OTU picking and taxonomy assignation were carried out utilizing as reference SILVA SSU database release 132[2] with the similarity threshold set at 0.97.

ITS1 and 16S rRNA gene sequences have been submitted to NCBI repository with the BioProject ID: PRJNA545913.

Culture-Dependent Enumeration of Fecal Yeast-Like Fungi

Fresh feces were collected from 8 healthy adult volunteers (hereinafter referred to as V1–V8), 4 males and 4 females aged 25–55, enrolled among the employees of the University of Modena and Reggio Emilia and their relatives. The subjects were not in relationship with the researchers, followed a western omnivore diet, and had not been treated with prebiotics and/or probiotics for 1 month and antibiotics for at least 3 months before being enrolled and throughout the study. Written informed consent was obtained from all the enrolled subjects, in accordance with the protocol approved by the local research ethics committee (reference number 225-15, Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy).

Feces were collected and analyzed at the beginning of the study (sample I), 3 (sample II), and 12 months (sample III) later. Stool samples were collected in sterile containers and forwarded to the laboratory within 3 h. The fecal samples were serially diluted and 100 μL dispersed onto Potates Dextrose Agar (PDA, BD Difco, Franklin Lakes, NJ, United States) supplemented with 100 mg/L chloramphenicol and 100 mg/L ampicillin. The plates were aerobically incubated at 30°C for 72 h, then fungal colonies were counted and isolated.

RAPD-PCR Analysis of Fungal Isolates and Taxonomic Attribution

For each fecal sample, 48 colonies from PDA plates were subjected to RAPD-PCR clustering. Genomic DNA was extracted and subjected to RAPD-PCR amplification according to Raimondi et al. (2017). The reaction was performed in 15 μL of a mixture consisting of Dream Taq Buffer (Thermo Fisher Scientific, Waltham, MA), 0.5 μM of M13-RAPD primer (5′-GAGGGTGGCGGTTCT-3′), 100 μM of each dNTP, 0.75 U Taq polymerase (DreamTaq, Thermo Fisher Scientific), and 50 ng of gDNA from the isolates. The thermocycle was the following: 94°C for 4 min; 45 cycles of 94°C for 1 min, 34°C for 1 min, and 72°C for 2 min; 72°C for 7 min. The PCR products were electrophoresed for 4 h at 160 V in a 25 × 25 cm 2% (w/v) agarose gel in TAE buffer. RAPD-PCR fingerprints were digitally captured, then analyzed with Gene Directory 2.0 software (Syngene, United Kingdom), which calculated similarities and derived a dendrogram with an unweighted pair group method with arithmetic means (UPGMA).

To attribute each biotype to a species, the ITS1 and ITS2 spacer regions were amplified, sequenced, and compared with GenBank database. Amplification was performed with ITS1 (5′-TCCGTAGGTGAACCTGCGG-3′) and ITS4 (5′-TCCTCCGCTTATTGATATGC-3′) primers within 150 μL of PCR Master Mix (Thermo Fisher Scientific), containing 0.2 μM of each primer, and 50 ng of gDNA. The thermocycle was the following: 95°C for 5 min; 30 cycles of 95°C for 1 min, 60°C for 1 min, and 72°C for 1 min; and 72°C for 7 min.

Strain Typing of *C. albicans*

For microsatellite length polymorphism (MLP) typing, the loci CAI and CAIV were amplified by PCR with 5′ fluorescently labeled forward primer (6-carboxyfluorescein) according to Sampaio et al. (2005). The PCR products were injected in an ABI 370 genetic analyzer (Applied Biosystems, Foster City, CA, United States) and the fragment sizes were determined using the GeneScan 3.7 software. For single-strand conformation polymorphism (SSCP) electrophoresis analysis, approximately 100 ng of unlabeled PCR products were loaded on an acrylamide

[1] https://unite.ut.ee/repository.php

[2] https://www.arb-silva.de/download/arb-files/

gel and separated with a DCode Universal Mutation Detection System (Bio-Rad, Hercules, CA, United States). Electrophoresis and silver staining were performed according to Li and Bai (2007).

Biofilm Formation Assay

Candida isolates were grown overnight in SDA plates at 37°C. Yeast cells were washed with PBS and standardized to 1×10^5 or 1×10^6 yeast cells/mL in RPMI-1640 medium (Gibco, Grand Island, NY, United States) supplemented with 10% heat-inactivated Fetal Bovine Serum (hiFBS; Defined Hyclone, Logan, Utah, United States), 2 mM L-glutamine, and 50 mg/mL gentamycin, hereafter referred to as cRPMI. The yeast cell suspension was placed (0.1 mL/well) in a 96-well polystyrene microplate and incubated at 37°C for 24 and 48 h (Orsi et al., 2014). After incubation, the biofilm was quantified by the crystal violet (CV) assay. The cells were washed 3 times with PBS at room temperature, fixed with methanol, stained with 1% CV for 15 min, washed three times with distilled water, and treated with 33% acetic acid for 10 min. The color development, read at 540 nm (OD_{540}), was proportional to biofilm. The consensus threshold for biofilm formation was established at $OD_{540} = 1.5$.

Human Epithelial Cells

The human epithelial colorectal adenocarcinoma cell line Caco-2 was grown in 75-cm^2 tissue-culture flasks (Nalge Nunc International, Naperville, IL, United States). When reaching 80–90% confluence, the cells were split 1:3 by standard methods. The cells were cultured at 37°C under 5% CO_2 in Dulbecco's Modified Eagle Medium (EuroClone, Milan, Italy) supplemented with 10% hiFBS, 2 mM L-glutamine, and 50 mg/mL gentamycin, hereafter referred to as cDMEM.

Adhesion Assay

Caco-2 cells (6.5×10^5/mL, 200 μL/well) were seeded in Lab-Tek II chamber slides, grown to confluence for 24 h at 37°C, then supplemented with a suspension of yeasts (2.6×10^6/mL, 200 μL/well) in cRPMI at E:T ratio 1:2. The cells were incubated for 1.5 or 3 h at 37°C under 5% CO_2. Fifteen minutes before the end of incubation, 40 μL/well of 1% Uvitex 2B (Polysciences Europe GmbH, Germany) were added. The wells were washed three times with PBS to remove the unbound yeast cells. The remaining yeasts were fixed for 30 min with PBS-buffered 4% paraformaldehyde. The Caco-2 cells were washed twice with PBS and then treated with PLGAR (Thermo Fisher Scientific, Waltham, MA, United States). The blue fluorescence of adhered yeasts was visualized by epifluorescence microscopy Nikon Eclipse 90i imaging system, equipped with Nomarski DIC optics (Nikon Instruments Inc., Melville, NY, United States). Samples were photographed with a DS-2Mv Nikon digital camera. The percentage of adhesion was defined as the ratio between Caco-2 cells with at least one adhering fungal cell and the total number of Caco-2 cells observed. At least 200 epithelial cells per sample were examined.

Secretory Activity of Human Epithelial Cell Line Caco-2

Caco-2 cells suspended at 8.0×10^5/mL in cDMEM, were seeded (200 μL/well) in 24 well-plates and allowed to reach confluence for 24 h at 37°C. Yeast cells (1.6×10^8/mL, 1 mL/well, E:T ratio 1: 100), 500 μg/mL β-D-glucan, or 10 μg/mL LPS were added, and the plates were incubated for additional 24 h. The supernatants were collected and the level of HBD-2 was assessed by the enzymatic immunoassay β-2-Defensin Human (Phoenix Pharmaceuticals, Karlsruhe, Germany).

Animal Trial

Nine intestinal yeast strains (*C. albicans* 02-10, *C. albicans* 04-10, *C. albicans* 08-06, *C. parapsilosis* 01-18, *Candida pararugosa* 04-14, *Candida zeylanoides* 01-03, *Geotrichum candidum* 02-01, *S. cerevisiae* 03-01, and *Rhodotorula mucilaginosa* 04-01) were grown for 24 h in 200 mL of Potato Dextrose Broth (BD Difco). The biomass was harvested by centrifugation, suspended in PBS containing 15% v/v glycerol to obtain single-dose aliquots of 10^7 cells in 150 μL and stored at −80°C. Aliquots were thawed one by one and utilized in the animal trial. Yeast viability was confirmed by enumeration onto PDA plates.

Six-weeks-old CD1 male mice were purchased from Envigo Laboratories (Oderzo, Italy) and housed in groups of 2 per individually ventilated cage, in a temperature-controlled environment (22 ± 2°C) under a 12-h light-dark cycle. Chow food and water were provided *ad libitum*. Mice were allowed to acclimate to the laboratory for 1 week before entering experimentation. Nine groups of 10 mice were established, each group being assigned a yeast strain. The mice received for 14 days a daily dose of yeast biomass via gastric gavage. Controls were fed with 150 μL of PBS containing 15% v/v glycerol. At the end of the treatment and after 2, 4, and 8 weeks of washout, fecal pellets were collected from each mouse and immediately stored at -80°C in pre-weighted tubes containing 50% glycerol. Yeasts were counted onto PDA plates supplemented with 100 mg/L chloramphenicol and 100 mg/L ampicillin. Isolates were subjected to RAPD-PCR profiling, in order to check the identity with the supplemented strain. This study was carried out in compliance with the appropriate laws and institutional guidelines (DL n. 116/92 art. 5). The protocol was approved by the Animal Care and Use Ethics Committee of the University of Padua, in compliance with the national and European guidelines for handling and use of experimental animals.

Statistics

In biofilm formation, adherence, and HBD-2 assays, the data are presented as means ± SD of at least three independent experiments. Means for *C. albicans*, NAC, and NCF groups were compared with one-way ANOVA followed by Bonferroni's *post hoc* test and were considered significantly different for $P < 0.05$. In the colonization assay, the viable counts at 2, 4, and 8 weeks were compared to that at 0 weeks using Wilcoxon

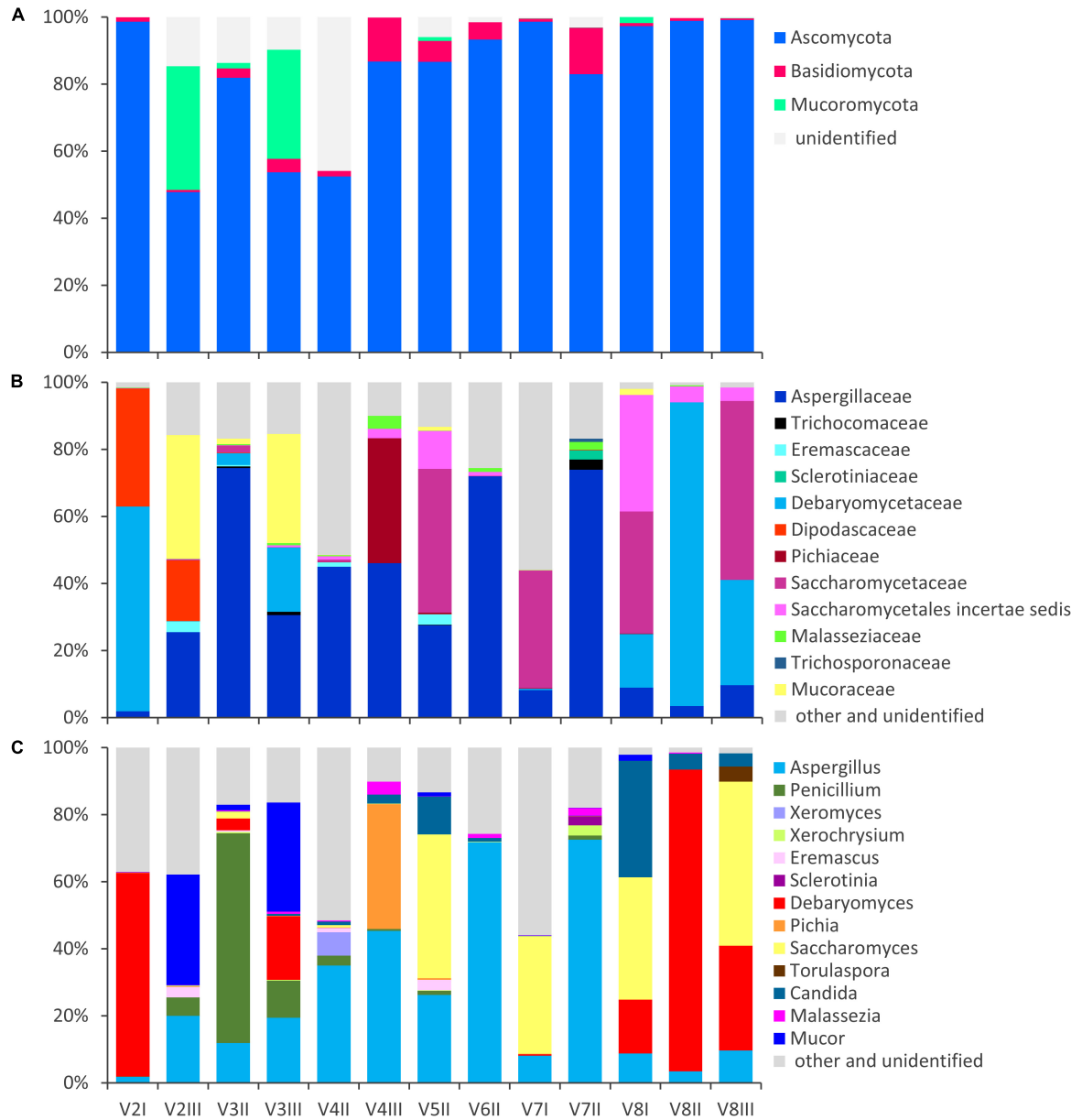


FIGURE 1 | Stacked bar-plot representation of the relative abundances of the main fungal taxa in the intestinal mycobiota, based on metagenomic ITS1 sequences. Only the taxa with a taxonomic attribution at the level of phylum **(A)**, family **(B)**, and genus **(C)** and appearing at least once with abundance higher than 1% are reported, all the others being grouped as "other and unclassified."

signed-rank test T and were considered significantly different for $P < 0.05$.

RESULTS

Fungal Communities Based on Metagenomic ITS1 Sequencing

Fecal samples of eight healthy adults were collected at three time points across a period of 1 year, i.e., 0, 3, and 12 months, hereinafter referred to as samples I, II, and III, respectively, and subjected to the profiling fungal and bacterial communities using metagenomic ITS1 and 16S sequencing, respectively. Only 13 out of 24 samples provided a valuable number of fungal reads, that resulted in a total of 140,225 quality-trimmed ITS1 sequences (on average 10,787 reads per sample) and 103 OTUs net of singletons (on average 28.2 OTUs per sample, ranging from 11 to 40) (**Figure 1** and **Supplementary Datasheet S1**). The unsuccessful sequencing was attributed to the too low concentration of fungal DNA, likely due to high concentrations of non-target DNA and to contaminants that may have been co-extracted with DNA. The 16S sequencing of the whole set of samples yielded 231,082 reads,

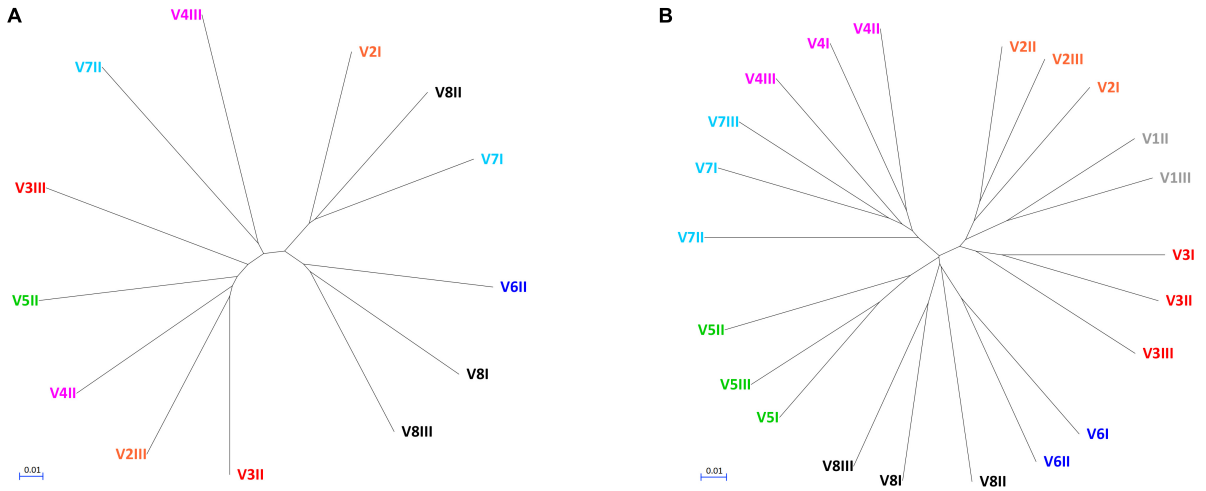


FIGURE 2 | UPGMA cladograms of fungal **(A)** and bacterial **(B)** communities, constructed using the corresponding Jaccard distance matrices.

attributed to 343 OTUs net of singletons (on average 130 OTUs per sample ranging between 81 and 178) (**Supplementary Figure S1** and **Supplementary Datasheet S2**).

The alpha diversity indices (Chao1, Shannon, and Pielou), all lower for fungi compared to bacteria, indicated that fungal communities presented lower richness and evenness than the bacterial ones (**Supplementary Figure S2**). Fungal communities were also more dissimilar/distant from each other, while the bacterial ones presented evident subject-based groups with the Jaccard beta diversity metric (**Figure 2**). PERMANOVA analysis of beta diversity highlighted significant differences among bacterial communities from different subjects ($P < 0.01$), with the distance between the subjects being generally greater than that within each subject. With regards to the fungal community, no significant difference among subjects was observed ($P > 0.05$), resulting from great distances both among the subjects and in the same subject over the time.

The fungal OTUs were ascribed to Ascomycota (62), Basidiomycota (34), and Mucorimycota (5). Ascomycota and Basidiomycota were ubiquitously present across the samples, with the former, always being the most abundant, ranging from 47.8 to 99.5% and the latter from 0.5 to 13.9% (**Figure 1**). Mucoromycota were generally negligible (not detected or <2%) with the exception of the samples V2III and V3III (36.9 and 32.6%, respectively).

Within Ascomycota, the family Aspergillaceae was the most frequent and on average the most abundant. The genus *Aspergillus* was the most widespread and copious, accounting for up to 72.5% (mean = 25.7%), while *Penicillium* was less than 1% in most cases, albeit it dominated in sample V3II (62.6%). A number of *Aspergillus* and *Penicillium* species occurred sporadically and were negligible in terms of abundance, with the exception of *Aspergillus niger* that was highly recurrent (11/13), although being always less than 1%.

Other families of Ascomycota that reached remarkably high amounts in few samples were Debaryomycetaceae (mostly *Debaryomyces*), Dipodascaceae (unidentified genus), Saccharomycetaceae (mostly *Saccharomyces* and *Torulaspora*), Pichiaceae (*Pichia*), and the Saccharomycetales *incertae sedis* encompassing the genus *Candida*. *Debaryomyces udenii* and *S. cerevisiae* were the most frequent and abundant species. *D. udenii* was identified in 11 out of 13 samples, with relative amounts up to 89.5% (19.9% on average), while *S. cerevisiae* was detected in 12 of 13 samples, with relative amounts up to 48.9% (13.9% on average). High levels of *Pichia terricola* and *Torulaspora delbrueckii* characterized the samples V4III and V8III, respectively. The genus *Candida* occasionally occurred at remarkably high level, with only a few detected species (e.g., *C. ethanolica*, *C. inconspicua*, *C. parapsilosis*, *C. tropicalis*, and *C. zeylanoides*), that never accounted for more than 0.5%. *C. albicans* was never identified.

Basidiomycota mostly belonged to unclassified members of Tremellomycetes (up to 9.7%) and to family Malasseziaceae (genus *Malassezia*). *Malassezia restricta* was common (12/13), reaching up to 3.8% of abundance. *Mucor* encompassed most of the OTUs of Mucoromycota. *Mucor piriformis* and *Mucor circinelloides* were recurrent (8/13 and 7/13, respectively), reaching the remarkably high level of approx. 32% in some samples.

In order to establish whether a relationship could occur between bacterial and fungal genera, Spearman rank correlations between the genera were calculated on the basis of 16S and ITS1 metagenome analysis (data not shown). No significant correlations between bacterial and fungal genera were found, excluding the correlations resulting from bacterial and fungal genera occurring only in a sample.

Enumeration, Isolation, and Taxonomic Identification of Cultivable Fungi

In parallel to NGS profiling, the fresh fecal samples were spread onto selective plates in order to determine the viable counts of fungi. In five samples, the fungal load was below the limit

of detection ($<10^2$ cfu/g), while it ranged from 1.4×10^2 to 4.5×10^5 cfu/g in the others (**Table 1**). A wide quantitative variability was detected over the time within a same subject, with a minority of cases presenting comparable fungal loads ($P > 0.05$) between different time points.

Forty-eight colonies from each sample were randomly selected and clustered into a total of 27 biotypes based on their RAPD-PCR profile, referred to 11 genera and 17 species (**Table 2**). Some RAPD-PCR profiles of *C. albicans*, *C. parapsilosis*, *C. pararugosa*, *R. mucilaginosa*, and *I. terricola* were similar across different hosts, indicating that this technique was unsuitable for typing such isolates at the strain level. SSCP and MLP fingerprinting differentiated *C. albicans* isolates in 7 biotypes, each occurring in a single subject (**Table 2**).

Candida albicans was never isolated in samples from V1 and V3, while it generally represented the sole or the dominant species among cultivable fungi in samples from the other six subjects. In subject V5, the *C. albicans* 8 D recurred at all the time points as the most abundant among cultivable fungi (72–100%). *C. zeylanoides* was found only in V1, with one of the two biotypes present at all the time-points as one of the most abundant fungal isolates (35–61%). The NAC isolates belonging to the species *C. parapsilosis*, *C. pararugosa*, *C. guillermondi*, and *C. lusitaniae* never occurred more than once in the same individual and accounted for 4–23% of the isolates.

Geotrichum candidum was isolated only from V1 and V2. In the latter subject, it presented five biotypes, one of which recurring across different time points and being always dominant among the isolates. *R. mucilaginosa* was isolated from five subjects. The species was particularly abundant in some samples (up to 47% of the isolates inV1-II). *S. cerevisiae*, *Debaryomyces hansenii*, and *Aspergillus candidus* were found only once in one or in a few subjects, but were generally abundant. *D. hansenii* was the dominant species in V3-III (58%).

Biofilm Formation

The ability of the fungal isolates to form biofilm *in vitro* was investigated (**Figure 3A**; **Supplementary Table S1**). All but one of *C. albicans* isolates produced large amounts of biofilm after 48 h of incubation, yielding OD_{540} values higher than 1.5. Conversely, both the NAC and NCF did not form biofilm. On average, the OD_{540} was 2.39 ± 0.79 for *C. albicans*, 0.21 ± 0.13 for NAC, and 0.20 ± 0.14 for NCF. The ability of *C. albicans* to produce biofilms was significantly greater ($P < 0.05$) than that of NAC and NCF.

Adherence of Fungal Isolates to Human Epithelial Cells Caco-2

An *in vitro* adhesion assay employing the human colorectal carcinoma cell line Caco-2 was used to evaluate the ability of the yeast isolates to adhere to intestinal epithelial cells (**Figure 3B**, **Supplementary Table S1**). The *C. albicans* isolates adhered more efficiently ($P < 0.05$) than NAC and NCF ones (13.4 ± 3.8, 8.1 ± 3.0, and 7.4 ± 3.9%, respectively). The percentage of adherence to Caco-2 cells was >11% for most *C. albicans*. Only two *C. albicans* isolates (i.e., 07-02 and 08-06) showed a quite low adherence capability. Conversely, the majority of NAC adhered less than 9%, with only few strains (*C. parapsilosis* 01-18, *C. pararugosa* 04-14, *C. zeylanoides* 01-03) showing values comparable to those of *C. albicans*. This feature was not species-specific, since other strains did not present similar behavior.

Human β-Defensin 2 Production by Caco-2 Cells Exposed to Fungal Isolates

The secretion of HBD-2 by Caco-2 epithelial cells was analyzed after exposure to the fungal isolates (**Figure 3C** and **Supplementary Table S1**). The cells challenged with *C. albicans* produced higher amounts of HBD-2 than those treated with NAC or NCF ($P < 0.05$). The response of Caco-2 cells to the diverse isolates of *C. albicans* was quite different, with HBD-2 levels ranging between 2.6 and 592.7 pg/mL. In contrast, most of the NAC did not induce any HBD-2 production, with the exception of five strains that prompted the release of 2.0–47.5 pg/mL HBD-2. NCF caused a variable accumulation of HBD-2 (51.7 ± 40.6 pg/mL). *Issatchenkia terricola* and *Pichia manshurica* did not induce any HBD-2 production. *G. candidum* caused the highest HBD-2 release within NCF isolates, ranging between 71.4 and

TABLE 1 | Counts of cultivable fungi in the feces of eight healthy volunteers, collected at 0 (I), 3 (II), and 12 months (III).

Subject	Cultivable fungi (cfu/g)		
	I	II	III
V1	$(8.1 \pm 1.8) \times 10^{2*}$	$(9.9 \pm 1.2) \times 10^{2*}$	$(3.6 \pm 0.2) \times 10^3$
V2	$(2.9 \pm 0.7) \times 10^3$	$(6.3 \pm 0.8) \times 10^2$	$(4.5 \pm 0.2) \times 10^5$
V3	$<10^2$	$(2.6 \pm 0.6) \times 10^{4*}$	$(2.1 \pm 0.3) \times 10^{4*}$
V4	$(2.2 \pm 0.9) \times 10^{3*}$	$(3.6 \pm 0.6) \times 10^2$	$(1.4 \pm 0.1) \times 10^{3*}$
V5	$(4.5 \pm 1.7) \times 10^2$	$(3.3 \pm 1.0) \times 10^3$	$(1.8 \pm 0.2) \times 10^3$
V6	$(2.7 \pm 0.6) \times 10^2$	$<10^2$	$<10^2$
V7	$<10^2$	$<10^2$	$(5.9 \pm 0.6) \times 10^2$
V8	$(1.4 \pm 0.1) \times 10^4$	$(4.7 \pm 0.7) \times 10^2$	$(6.2 \pm 0.3) \times 10^4$

*Values are means ± SD, n = 3. Within each row, for samples presenting detectable loads, * indicate means that did not significantly differ (P > 0.05).*

TABLE 2 | RAPD-PCR, SSCP, and MLP analysis of fungal isolates.

Subject	I			II			III		
	Isolates	**Biotype***	**r.a. (%)**	**Isolates**	**Biotype***	**r.a. (%)**	**Isolates**	**Biotype***	**r.a. (%)**
V1	*Candida zeylanoides 01-03*	2	61	*Rhodotorula mucilaginosa 01-08*	13	38	*Candida zeylanoides 01-15*	2	47
	Candida zeylanoides 01-04	3	23	*Candida zeylanoides 01-11*	2	35	*Candida lusitaniae 01-17*	19	20
	Candida pararugosa 01-07	1	8	*Rhodotorula mucilaginosa 01-09*	10	9	*Rhodotorula mucilaginosa 01-20*	10	14
	Exsofiala dermatitidis 01-06	4	8	*Issatchenkia terricola 01-10*	11	9	*Candida parapsilosis 01-18*	9	7
				Penicillium crustosum 01-12	12	9	*Geotrichum candidum 01-19*	20	7
							Pichia manshurica 01-14	21	5
V2	*Geotrichum candidum 02-01*	5	86	*Geotrichum candidum 02-09*	14	86	*Geotrichum candidum 02-11*	14	100
	Geotrichum candidum 02-02	6	7	*Candida albicans 02-10*	8 A	14			
	Geotrichum candidum 02-03	7	7						
V3	–			*Saccharomyces cerevisiae 03-01*	15	100	*Debaryomyces hansenii 03-10*	23	58
							Candida guilliermondii 03-03	22	16
							Candida parapsilosis 03-04	9	12
							Talaromyces purpureogenus 03-02	25	7
							Penicillium diversum 03-09	24	5
							Rhodotorula mucilaginosa 03-05	10	2
V4	*Candida albicans 04-02*	8 B	88	*Aspergillus candidus 04-08*	16	50	*Candida albicans 04-10*	8 C	73
	Rhodotorula mucilaginosa 04-01	10	8	*Aspergillus candidus 04-21*	17	25	*Candida pararugosa 04-14*	1	23
	Candida parapsilosis 04-04	9	4	*Aspergillus candidus 04-22*	18	25	*Saccharomyces cerevisiae 04-11*	26	4
V5	*Candida albicans 05-01*	8 D	100	*Candida albicans 05-05*	8 D	84	*Candida albicans 05-09*	8 D	72
				Rhodotorula mucilaginosa 05-06	10	16	*Rhodotorula mucilaginosa 05-07*	10	28
V6	*Candida albicans 06-01*	8 E	100	–			–		
V7	–			–			*Candida albicans 07-02*	8 F	100
V8	*Candida albicans 08-06*	8 G	98	*Candida albicans 08-08*	8 G	98	*Saccharomyces cerevisiae 08-11*	27	100
	Issatchenkia terricola 08-03	11	2	*Rhodotorula mucilaginosa 08-07*	10	2			

*Taxonomic characterization through ITS sequencing and relative abundance are reported for each cluster identified in the eight volunteers at 0 (I), 3 (II), and 12 months (III). – Indicates that no fungal colonies were isolated. *Numbers indicate RAPD-PCR profiles; letters represent SSCP and MLP fingerprinting profiles of Candida albicans isolates. r.a., relative abundance expressed as %.*

110.9 pg/mL, followed by *R. mucilaginosa* (36.1–78.6 pg/mL), and by *S. cerevisiae* (8.5–18.5 pg/mL).

Yeast Colonization of the Murine GIT

Nine yeast isolates, i.e., 3 *C. albicans* (*C. albicans* 02-10, *C. albicans* 04-10, and *C. albicans* 08-06), 3 NAC (*C. parapsilosis* 01-18, *C. pararugosa* 04-14, and *C. zeylanoides* 01-03), and 3 NCF (*G. candidum* 02-01, *R. mucilaginosa* 04-01, and *S. cerevisiae* 03-01), were tested for the ability to colonize the GIT of mice. Yeasts were daily given to mice for 14 days and the persistence in feces was evaluated at the end of the treatment (0), then after 2, 4, and 8 weeks of washout (**Figure 4**). All the strains were recovered at the end of the supplementation period. The most abundant were *C. albicans* 02-10, *C. parapsilosis* 01-18,

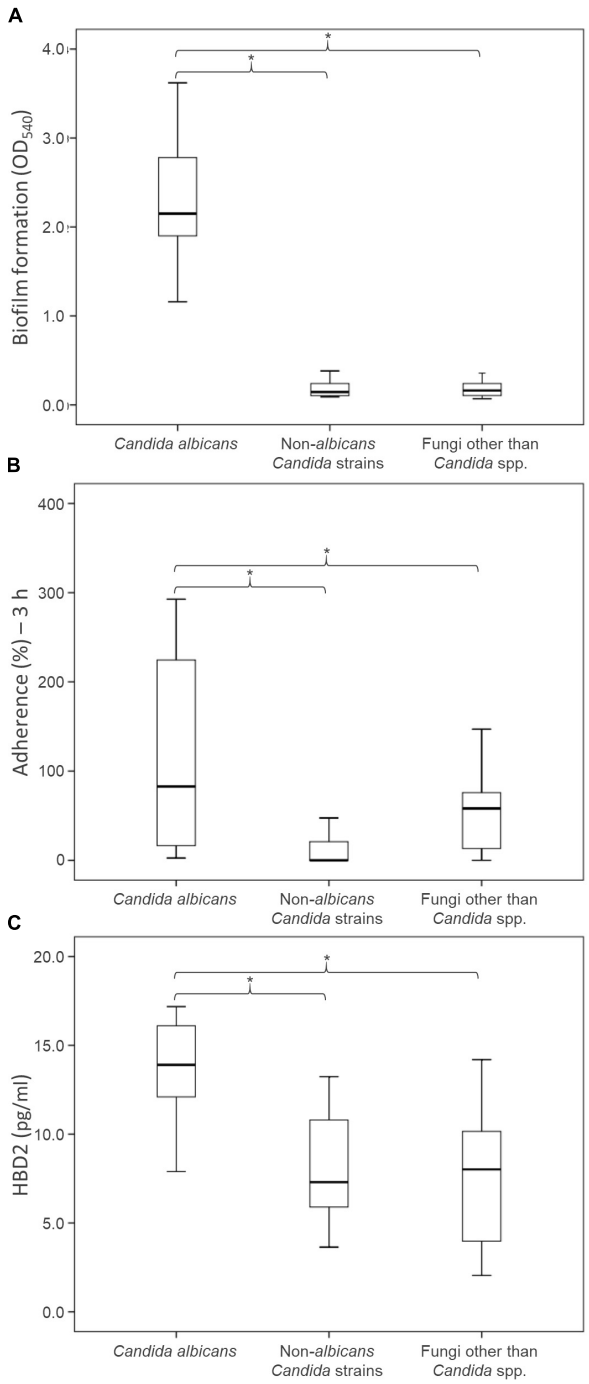


FIGURE 3 | Health relevant properties of the fungi isolated from fecal specimens of healthy donors: biofilm production **(A)**, adhesion to epithelial Caco-2 cells **(B)**, and induction of HBD-2 production in CaCo-2 cells **(C)**. Boxes indicate the median and 25th and 75th percentiles; whiskers indicate 10th and 90th percentiles; *indicate groups with significantly different means ($P < 0.05$).

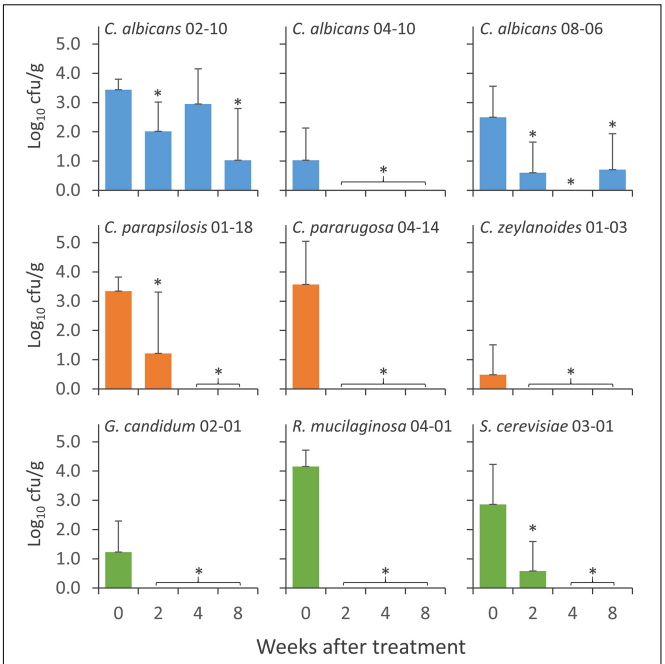


FIGURE 4 | Concentration of the strains of *Candida albicans* (cyan), NAC (orange), and NCF (green) in the feces of mice groups during the wash period that followed the treatment with *C. albicans* 02-10, *C. albicans* 08-06, *C. albicans* 04-10, *C. parapsilosis* 01-18, *C. pararugosa* 04-14, *C. zeylanoides* 01-03, *G. candidum* 02-01 (G), *S. cerevisiae* 03-01 (H), and *R. mucilaginosa* 04-01 (I). Values are means ($n = 10$), the error bars representing the SD. For each strain, *indicate statistically significant difference compared to 0 weeks values (Wilcoxon signed-rank test, $P < 0.05$).

C. pararugosa 04-14, and *R. mucilaginosa* 04-01, all presenting a mean concentration >3.0 Log_{10} cfu/g. *C. albicans* 04-10 and *C. zeylanoides* 01-03 were the less abundant, with a mean concentration <1 Log_{10} cfu/g. *C. albicans* 04-10, the NAC, and the other yeasts dropped below the limit of detection after 2 or 4 weeks of washout. On the other hand, *C. albicans* 02-10 and *C. albicans* 08-06 were found for up to 8 weeks after the end of treatment. The former strain remained in the range of 1–3 Log_{10} cfu/g throughout the whole washout period, whereas the latter drastically decreased in the first 2 weeks, remaining present, although sporadically, in the feces of mice after 8 weeks.

DISCUSSION

This study investigated the fecal mycobiome in healthy adults over a time span of 12 months, using ITS1 deep sequencing combined with a culture-based method. The ITS profiling provided a comprehensive description of the fungi harbored in the intestine, whereas isolation of viable fungi allowed the study of the properties involved in gut colonization. Compared to the bacterial microbiome, the biodiversity of fungal communities was lower and characterized by greater unevenness, in agreement with previous studies (Qin et al., 2010; Suhr and Hallen-Adams, 2015; Nash et al., 2017). Most samples were dominated by one or two fungal genera. The unevenness in fungal composition resulted in great distances both among subjects and over the time within the same subject. This was consistent with the

occurrence of fungal cells transiting through the GIT without being stable colonizers, or with the possibility of a high variability in the individual mycobiome. On the contrary, the bacterial microbiome yielded subject-based clusters coherent with a more stable resident community inhabiting the intestine of each donor.

Previous studies and the data herein presented consistently indicated *Aspergillus*, *Candida*, *Debaryomyces*, *Malassezia*, *Penicillium*, *Pichia*, and *Saccharomyces* as the most recurrent and/or dominant fungal genera. The major discrepancies with literature were associated with the detection of *Mucor* and to the absence or scarcity of some genera such as *Cladosporium*, *Clavispora*, *Cyberlindnera*, and *Galactomyces* (Hoffmann et al., 2013; Strati et al., 2016; Nash et al., 2017; Auchtung et al., 2018; Borges et al., 2018). Geographic distribution of the volunteers and diet habits probably accounted for the differences observed, since the present study and Strati's investigation, both enrolling volunteers in Italy, come to similar conclusions. Many species of *Aspergillus* and *Penicillium* detected in this trial, each occurring sporadically and/or negligibly in terms of abundance, are regarded as health concerns, being potential sources of mycotoxins or opportunistic infections targeting mucosal tissues (Hedayati et al., 2007; Perrone and Gallo, 2017; Perrone and Susca, 2017). This supported the hypothesis of their transient presence in the gut as a result of the consumption of exogenous sources, such as contaminated food and beverages. In the GIT, fungal population can be made by true inhabitants or by microbes occasionally ingested with fermented or spoiled foods, likewise lactic acid bacteria abundantly occurring in food (Rossi et al., 2016). Despite *Aspergillus* and *Penicillium* were found in nearly all the samples, isolates ascribable to these genera were obtained only in a minority of cases. A similar observation was reported also by Strati et al. (2016) and suggests that fecal *Aspergillus* and *Penicillium* may be allochthons that do not survive the gastrointestinal transit. Likewise, other frequent and abundant taxa detected in the metagenomic survey (e.g., *Mucor*, *Malassezia*) were not identified by culture-based analysis, raising the question of whether they were actually viable inhabitants of the GIT.

The genus *Candida* was identified by metagenomics survey in most samples, occasionally occurring at remarkably high amounts. Unlike previous studies, *C. albicans* was not detected (Hoffmann et al., 2013; Seed, 2014; Suhr and Hallen-Adams, 2015; Strati et al., 2016; Suhr et al., 2016; Nash et al., 2017; Auchtung et al., 2018; Borges et al., 2018). Most of the *Candida* genus-related sequences received the "*Candida* unidentified" taxonomy using the last available version of the UNITE database (release 2017-12-01), but were given the designation "*Candida albicans*" if a previous version was utilized (release 2015-08-01) (**Supplementary Figure S3**). This underlines that the reliability of reference databases is pivotal for a consistent taxonomic attribution and for the comparison with other metagenomic studies and culture dependent approaches. As a matter of fact, isolates of *C. albicans* were frequently obtained and dominated the cultivable mycobiota of different samples. The molecular fingerprinting of the isolates lead to the identification of strains that longitudinally persisted within the same subject, likely as commensals inhabiting the intestine or systematically reaching it from the oral cavity.

Only few NAC species (*C. ethanolica*, *C. inconspicua*, *C. parapsilosis*, *C. tropicalis*, and *C. zeylanoides*) were detected by metagenome analysis. Evidence of a possible stable colonization was provided for *C. zeylanoides* by culture-dependent experiments, with a biotype isolated in the feces of a subject over a time span of one year. Based on the present metagenomics analysis, *S. cerevisiae* resulted as one of the most frequent and abundant species. Albeit cultivation of *S. cerevisiae* is generally not an issue, isolates were obtained only from two samples. This result was in agreement with a previous study, where the isolation succeeded only in a minority of samples (4/111) (Strati et al., 2016). Therefore, conclusive evidence was not obtained establishing whether *S. cerevisiae* is an occasional or a permanent colonizer of the human GIT. A recent study conducted a phylogenetic analysis on 12 microsatellite loci and 1715 combined CDSs from whole-genome sequencing on a set of *S. cerevisiae* strains isolated from the gut of normal and inflammatory bowel disease patients (Ramazzotti et al., 2019). The analysis revealed evidence of clonal colonization within the host's gut. The phenotypic analysis of the isolates combined with immunological profiling indicated that both genetic and environmental factors involved in cell wall remodeling and sporulation are the main drivers of adaptation in *S. cerevisiae* populations in the human gut. The most likely hypothesis is that a combination of specific genetic features in the microbe and in the host allow colonization from microorganisms normally present solely as passengers, often when alterations in the gastrointestinal barrier occur (Sokol et al., 2017).

Fingerprinting of the NCF isolates pointed out that also the *G. candidum* and *R. mucilaginosa* may stably inhabit the GIT, even if genera *Geotrichum* and *Rhodotorula* were not detected by metagenome survey. However, the family of Dipodascaceae which includes the genus *Geotrichum* was found to be abundant in the corresponding samples. Interestingly, metagenomic revealed that also the ecologically understudied *D. udenii* is a putative permanent colonizer that, when detected, is present in all the samples of the same subject.

Functional investigation focused on comparison of *C. albicans*, NAC and NCF isolates. *C. albicans* adhered to human epithelial cells more efficiently than NAC and NCF and produced greater amounts of biofilm *in vitro*. In contrast, NAC and NCF were not biofilm producers. This evidence is in agreement with previous studies biofilm formation by of *C. albicans* and/or other fungi, such as *S. cerevisiae* (Chandra et al., 2001). Although the fungal population investigated in this study has been isolated from the stool of healthy volunteers, the ability to adhere and to form biofilm appears to be a trait peculiar of the opportunistic fungus *C. albicans*, the behavior of which supports the potential virulence and the impact on clinical infections.

Increasing evidence indicates that fungi may stimulate host intestinal immune system, beneficially affecting host homeostasis (Yan et al., 2013). The *C. albicans* isolates also induced the highest release of HBD-2 by human epithelial cells, further differing from NAC and NCF. On these bases, our results suggest that the release

of HBD-2 by *C. albicans* may have an important role in regulating the size and the activity of the fungal population harbored in the gut of healthy individuals. The intestinal epithelial cells play a crucial role in local antimicrobial defense, since they not only represent a physiologic barrier for pathogens, but they also function as a part of innate immune system via production of antimicrobial peptides and cytokines, such as HBD-2, endowed with a wide spectrum of antibacterial and antifungal activity (Ganz, 2003; Selsted and Ouellette, 2005).

Three representative fungal isolates from each group have been administered to mice to evaluate the ability to colonize the intestine. Only two out of three *C. albicans* isolates were recovered from the stools of animals, 8 weeks after the end of the treatment. In contrast, the other tested fungi including NAC and NCF were not isolated yet after 2 weeks of wash-out. These results provide expand initial evidence on the allochthonous nature of most fungi in the gut, confirming that stability of mycobiota in the intestine over the time is quite low (Erturk-Hasdemir and Kasper, 2013).

CONCLUSION

NGS approach combined with a culture-based methods revealed that most of the fungi transit through the GIT without being stable colonizers. The panel of fungi detected by metagenomic surveys is dominated by environmental or food-borne taxa that are rather washed out by the gastrointestinal transit, while only a minority of viable fungi behave as stable colonizers, with certain biotypes being longitudinally persistent within a subject. Notoriously, *C. albicans* exhibits features involved in stable colonization (e.g., transition to hyphal form, biofilm formation, adhesion to intestinal epithelial cells). Accordingly, it is the sole species that successfully colonized the murine intestine. It remains to be established if other species, herein identified as putative colonizers, such as *D. udenii*, *C. zeylanoides*, *G. candidum*, *S. cerevisiae*, and *R. mucilaginosa,* are true inhabitants of the intestine or are delocalized from other body districts. A combination of specific features in the microbe and in the host seems the driver of colonization of microorganisms normally present solely as passengers. As whole, the fungal microbiome is much more variable at the individual level than the bacterial one.

DATA AVAILABILITY

The datasets generated for this study can be found in NCBI, PRJNA545913.

ETHICS STATEMENT

The animal study was carried out in compliance with the appropriate laws and institutional guidelines (DL n. 116/92 art. 5). The protocol was approved by the Animal Care and Use Ethics Committee of the University of Padua, in compliance with the national and European guidelines for handling and use of experimental animals. For the collection of fecal samples from healthy human subject, written informed consent was obtained from all the enrolled subjects, in accordance with the protocol approved by the local research ethics committee (reference number 225-15, Comitato Etico Provinciale, Azienda Policlinico di Modena, Italy).

AUTHOR CONTRIBUTIONS

MR, DC, IC, SaP, and EB conceived and designed the experiments. AlA, FC, and DC performed the metagenomic study and the bioinformatics. SR, CG, MS, and LR isolated and characterized the cultivable fungi. AnA, SiP, and BC performed the phenotypical characterization of yeast isolates. PB and IC performed the animal trial. AlA and MR wrote the manuscript with contributions from all other authors.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2019.01575/full#supplementary-material

REFERENCES

Albanese, D., Fontana, P., De Filippo, C., Cavalieri, D., and Donati, C. (2015). MICCA: a complete and accurate software for taxonomic profiling of metagenomic data. *Sci. Rep.* 5:9743. doi: 10.1038/srep09743

Auchtung, T. A., Fofanova, T. Y., Stewart, C. J., Nash, A. K., Wong, M. C., Gesell, J. R., et al. (2018). Investigating colonization of the healthy adult gastrointestinal tract by fungi. *mSphere* 3:e00092-18. doi: 10.1128/mSphere.00092-18

Barelli, C., Albanese, D., Donati, C., Pindo, M., Dallago, C., Rovero, F., et al. (2015). Habitat fragmentation is associated to gut microbiota diversity of an endangered primate: implications for conservation. *Sci. Rep.* 5:14862. doi: 10.1038/srep14862

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., et al. (2018). QIIME 2: reproducible, interactive, scalable, and extensible microbiome data science. *PeerJ* 6:e27295v2.

Borges, F. M., de Paula, T. O., Sarmiento, M. R. A., de Oliveira, M. G., Pereira, M. L. M., Toledo, I. V., et al. (2018). Fungal diversity of human gut microbiota among eutrophic, overweight, and obese individuals based on aerobic culture-dependent approach. *Curr. Microbiol.* 75, 726–735. doi: 10.1007/s00284-018-1438-8

Brandt, M. E., and Lockhart, S. R. (2012). Recent taxonomic developments with *Candida* and other opportunistic yeasts. *Curr. Fungal Infect. Rep.* 6, 170–177. doi: 10.1007/s12281-012-0094-x

Cavalieri, D., Di Paola, M., Rizzetto, L., Tocci, N., De Filippo, C., Lionetti, P., et al. (2018). Genomic and phenotypic variation in morphogenetic networks of two *Candida albicans* isolates subtends their different pathogenic potential. *Front. Immunol.* 8:1997. doi: 10.3389/fimmu.2017.01997

Chandra, J., Kuhn, D. M., Mukherjee, P. K., Hoyer, L. L., McCormick, T., and Ghannoum, M. A. (2001). Biofilm formation by the fungal pathogen *Candida albicans*: development, architecture, and drug resistance. *J. Bacteriol.* 183, 5385–5394. doi: 10.1128/JB.183.18.5385-5394.2001

Erturk-Hasdemir, D., and Kasper, D. L. (2013). Resident commensals shaping immunity. *Curr. Opin. Immunol.* 25, 450–455. doi: 10.1016/j.coi.2013.06.001

Findley, K., Oh, J., Yang, J., Conlan, S., Deming, C., Meyer, J. A., et al. (2013). Topographic diversity of fungal and bacterial communities in human skin. *Nature* 498, 367–370. doi: 10.1038/nature12171

Ganz, T. (2003). Defensins: antimicrobial peptides of innate immunity. *Nat. Rev. Immunol.* 3, 710–720. doi: 10.1038/nri1180

Hedayati, M. T., Pasqualotto, A. C., Warn, P. A., Bowyer, P., and Denning, D. W. (2007). *Aspergillus flavus*: human pathogen, allergen and mycotoxin producer. *Microbiology* 153, 1677–1692. doi: 10.1099/mic.0.2007/007641-0

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., et al. (2013). Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS One* 8:66019. doi: 10.1371/journal.pone.0066019

Huseyin, C. E., O'Toole, P. W., Cotter, P. D., and Scanlan, P. D. (2017). Forgotten fungi-the gut mycobiome in human health and disease. *FEMS Microbiol. Rev.* 41, 479–511. doi: 10.1093/femsre/fuw047

Li, J., and Bai, F. Y. (2007). Single-strand conformation polymorphism of microsatellite for rapid strain typing of *Candida albicans*. *Med. Mycol.* 45, 629–635. doi: 10.1080/13693780701530950

Mayer, F. L., Wilson, D., and Hube, B. (2013). *Candida albicans* pathogenicity mechanisms. *Virulence* 4, 119–128. doi: 10.4161/viru.22913

Merseguel, K. B., Nishikaku, A. S., Rodrigues, A. M., Padovan, A., Ferreira, R. C., de Azevedo Melo, A. S., et al. (2015). Genetic diversity of medically important and emerging *Candida* species causing invasive infection. *BMC Infect. Dis.* 15:57. doi: 10.1186/s12879-015-0793-3

Moyes, D. L., Richardson, J. P., and Naglik, J. R. (2015). *Candida albicans*-epithelial interactions and pathogenicity mechanisms: scratching the surface. *Virulence* 6, 338–346. doi: 10.1080/21505594.2015.1012981

Muadcheingka, T., and Tantivitayakul, P. (2015). Distribution of *Candida albicans* and non-*albicans Candida* species in oral candidiasis patients: correlation between cell surface hydrophobicity and biofilm forming activities. *Arch. Oral Biol.* 60, 894–901. doi: 10.1016/j.archoralbio.2015.03.002

Nash, A. K., Auchtung, T. A., Wong, M. C., Smith, D. P., Gesell, J. R., Ross, M. C., et al. (2017). The gut mycobiome of the human microbiome project healthy cohort. *Microbiome* 5:153. doi: 10.1186/s40168-017-0373-4

Orsi, C. F., Borghi, E., Colombari, B., Neglia, R. G., Quaglino, D., Ardizzoni, A., et al. (2014). Impact of *Candida albicans* hyphal wall protein 1 (HWP1) genotype on biofilm production and fungal susceptibility to microglial cells. *Microb. Pathog.* 69-70, 20–27. doi: 10.1016/j.micpath.2014.03.003

Perrone, G., and Gallo, A. (2017). *Aspergillus* species and their associated mycotoxins. *Methods Mol. Biol.* 1542, 33–49. doi: 10.1007/978-1-4939-6707-0_3

Perrone, G., and Susca, A. (2017). *Penicillium* species and their associated mycotoxins. *Methods Mol. Biol.* 1542, 107–119. doi: 10.1007/978-1-4939-6707-0_5

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464, 59–65. doi: 10.1038/nature08821

Raimondi, S., Amaretti, A., Rossi, M., Fall, P. A., Tabanelli, G., Gardini, F., et al. (2017). Evolution of microbial community and chemical properties of a sourdough during the production of Colomba, an Italian sweet leavened baked product. *LWT Food Sci. Technol.* 86, 31–39. doi: 10.1016/j.lwt.2017.07.042

Ramazzotti, M., Stefanini, I., Di Paola, M., De Filippo, C., Rizzetto, L., Berná, L., et al. (2019). Population genomics reveals evolution and variation of *Saccharomyces cerevisiae* in the human and insects gut. *Environ. Microbiol.* 21, 50–71. doi: 10.1111/1462-2920.14422

Rossi, M., Martínez-Martínez, D., Amaretti, A., Ulrici, A., Raimondi, S., and Moya, A. (2016). Mining metagenomic whole genome sequences revealed subdominant but constant Lactobacillus population in the human gut microbiota. *Environ. Microbiol. Rep.* 8, 399–406. doi: 10.1111/1758-2229.12405

Sampaio, P., Gusmão, L., Correia, A., Alves, C., Rodrigues, A. G., Pina-Vaz, C., et al. (2005). New microsatellite multiplex PCR for *Candida albicans* strain typing reveals microevolutionary changes. *J. Clin. Microbiol.* 43, 3869–3876. doi: 10.1128/JCM.43.8.3869-3876.2005

Seed, P. C. (2014). The human mycobiome. *Cold Spring Harb. Perspect. Med.* 5:a019810. doi: 10.1101/cshperspect.a019810

Selsted, M. E., and Ouellette, A. J. (2005). Mammalian defensins in the antimicrobial immune response. *Nat. Immunol.* 6, 551–557. doi: 10.1038/ni1206

Simon, G. L., and Gorbach, S. L. (1984). Intestinal flora in health and disease. *Gastroenterology* 86, 174–193. doi: 10.1016/0016-5085(84)90606-1

Sokol, H., Leducq, V., Aschard, H., Pham, H. P., Jegou, S., Landman, C., et al. (2017). Fungal microbiota dysbiosis in IBD. *Gut* 66, 1039–1048. doi: 10.1136/gutjnl-2015-310746

Strati, F., Di Paola, M., Stefanini, I., Albanese, D., Rizzetto, L., Lionetti, P., et al. (2016). Age and gender affect the composition of fungal population of the human gastrointestinal tract. *Front. Microbiol.* 7:1227. doi: 10.3389/fmicb.2016.01227

Suhr, M. J., Banjara, N., and Hallen-Adams, H. E. (2016). Sequence-based methods for detecting and evaluating the human gut mycobiome. *Lett. Appl. Microbiol.* 62, 209–215. doi: 10.1111/lam.12539

Suhr, M. J., and Hallen-Adams, H. E. (2015). The human gut mycobiome: pitfalls and potentials–a mycologist's perspective. *Mycologia* 107, 1057–1073. doi: 10.3852/15-147

Tang, J., Iliev, I. D., Brown, J., Underhill, D. M., and Funari, V. A. (2015). Mycobiome: approaches to analysis of intestinal fungi. *J. Immunol. Methods* 421, 112–121. doi: 10.1016/j.jim.2015.04.004

Tsui, C., Kong, E. F., and Jabra-Rizk, M. A. (2016). Pathogenesis of *Candida albicans* biofilm. *Pathog. Dis.* 74: ftw018. doi: 10.1093/femspd/ftw018

Wang, X., Liu, C., Huang, L., Bengtssonpalme, J., Chen, H., Zhang, J. H., et al. (2015). ITS1: a DNA barcode better than ITS2 in eukaryotes? *Mol. Ecol. Resour.* 15, 573–586. doi: 10.1111/1755-0998.12325

Wu, P. F., Liu, W. L., Hsieh, M. H., Hii, I. M., Lee, Y. L., Lin, Y. T., et al. (2017). Epidemiology and antifungal susceptibility of candidemia isolates of non-albicans *Candida* species from cancer patients. *Emerg. Microbes Infect.* 6:e87. doi: 10.1038/emi.2017.74

Yan, L., Yang, C., and Tang, J. (2013). Disruption of the intestinal mucosal barrier in *Candida albicans* infections. *Microbiol. Res.* 168, 389–395. doi: 10.1016/j.micres.2013.02.008

Yang, R. H., Su, J. H., Shang, J. J., Wu, Y. Y., Li, Y., Bao, D. P., et al. (2018). Evaluation of the ribosomal DNA internal transcribed spacer (ITS), specifically ITS1 and ITS2, for the analysis of fungal diversity by deep sequencing. *PLoS One* 13:e0206428. doi: 10.1371/journal.pone.0206428

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Raimondi, Amaretti, Gozzoli, Simone, Righini, Candeliere, Brun, Ardizzoni, Colombari, Paulone, Castagliuolo, Cavalieri, Blasi, Rossi and Peppoloni. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

microorganisms

MDPI

Article

Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of *E. coli* Isolated from the Feces of Healthy Subjects

Stefano Raimondi [1], Lucia Righini [1], Francesco Candeliere [1], Eliana Musmeci [1], Francesca Bonvicini [2], Giovanna Gentilomi [2,3], Marjanca Starčič Erjavec [4], Alberto Amaretti [1] and Maddalena Rossi [1,*]

1 Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125, Modena, Italy
2 Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, Via Massarenti 9, 40138 Bologna, Italy
3 Unit of Microbiology, Alma Mater Studiorum-University of Bologna, S. Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy
4 Department of Biology, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia
* Correspondence: maddalena.rossi@unimore.it; Tel.: +39-059-205-8589

Received: 19 July 2019; Accepted: 7 August 2019; Published: 10 August 2019

Abstract: *Escherichia coli* may innocuously colonize the intestine of healthy subjects or may instigate infections in the gut or in other districts. This study investigated intestinal *E. coli* isolated from 20 healthy adults. Fifty-one strains were genotyped by molecular fingerprinting and analyzed for genetic and phenotypic traits, encompassing the profile of antibiotic resistance, biofilm production, the presence of surface structures (such as curli and cellulose), and their performance as recipients in conjugation experiments. A phylogroup classification and analysis of 34 virulence determinants, together with genes associated to the *pks* island (polyketide-peptide genotoxin colibactin) and conjugative elements, was performed. Most of the strains belonged to the phylogroups B1 and B2. The different phylogroups were separated in a principal coordinate space, considering both genetic and functional features, but not considering pulsed-field gel electrophoresis. Within the B2 and F strains, 12 shared the pattern of virulence genes with potential uropathogens. Forty-nine strains were sensitive to all the tested antibiotics. Strains similar to the potential pathogens innocuously inhabited the gut of healthy subjects. However, they may potentially act as etiologic agents of extra-intestinal infections and are susceptible to a wide range of antibiotics. Nevertheless, there is still the possibility to control infections with antibiotic therapy.

Keywords: *Escherichia coli*; typing; gut microbiota; PFGE (pulsed-field gel electrophoresis); virulence; antibiotic resistance; conjugation; curli; co-occurrence

1. Introduction

E. coli are permanent colonizers of the human microbiota that may persist as gut commensals without inducing any intestinal or extraintestinal infections. On the other hand, certain *E. coli* strains may instigate infections not only in the gut but also in other districts, such as those caused by the extraintestinal pathogenic *E. coli* (ExPEC) [1]. The commensalism or virulence of *E. coli* derives from a complex balance between the whole status of the host and the presence and expression of virulence determinants. Commonly, ExPEC strains reside as harmless commensals in the gut of healthy subjects. However, these strains can cause an infection in compromised patients, in case they reach a usually sterile extraintestinal site, such as the urinary tract [2]. The gut microbiota, therefore, act as a powerful

reservoir of ExPEC strains potentially responsible for infections, with pathogenic and commensal *E. coli* generally differing in terms of their phylogenetic backgrounds and virulence attributes [3].

The strains that innocuously colonize the intestine of healthy subjects may differ from those that are prone to cause diseases, especially those in possession of accessory traits that confer fitness and competitiveness and shape a specific relationship with the host. The virulence capability of *E. coli* depends on adhesion, biofilm formation, attachment, acquirement of nutrients, competition with other bacteria, toxin production, and avoidance or subversion of host defense mechanisms [4].

Adherence to the epithelial cells is mediated by surface structures or molecules, like fimbrial and afimbrial adhesins, curli, and outer membrane proteins encoded by the *pap* cluster and other genes (*afa/draBC*, *fimH*, *focG*, *gafD*, *hra*, *iha*, *sfa/focDE*, *sfaS*, and *yfcV*) [5]. Furthermore, *E. coli* exploits several mechanisms of iron uptake that are associated with siderophores and other binding proteins encoded by *chuA*, *fyuA*, *ire*, *iroN*, and *iutA* [5,6]. Indeed, like other pathogens, needing iron for metabolism, *E. coli* must face the host's response to infection, which involves a reduction in the amount of iron available via a decrease of intestinal iron absorption, the synthesis of additional iron proteins, and shifting iron from the plasma pool into intracellular storage. *E. coli* virulence is also enhanced by the production of toxins (e.g., cytotoxic necrotizing factor 1, autotransporter toxins, and alpha hemolysin) that target the cell's skeleton, metabolism, or cytoplasmic membrane [5,7].

Genetic exchange increases the success of commensals in invasion, intracellular survival, and spread, providing them with increased fitness and versatility. Hence, the boundary between commensals and pathogens is made fainter by horizontal gene transfer. Mobile genetic elements, such as transposons, plasmids, and insertion sequences, contribute to the plasticity of the *E. coli* genome, resulting in an extremely large pangenome of more than 16,000 genes [8]. Moreover, horizontal gene transfer favors the diffusion of antimicrobial resistance (AR) among both *E. coli* and other commensals, thereby enlarging the spectrum of resistance and promoting epidemiological success, with a bloom in worldwide public health concern associated with the misuse of antibiotics. In particular, conjugation is one of the most important ways for genes to exchange in prokaryotes, leading to genetic variation within a species and the acquisition of new traits. This process requires complex circuits that regulate the transcription of conjugation genes, the assembly of conjugative pili, the formation of the mating pore connecting donor and recipient cells, and the enzymatic processing of plasmid DNA to be transferred [9].

Research interest has been manly focused on characterizing virulent clinical *E. coli.* [10], whereas strains isolated from healthy subjects have been mainly investigated in comparative studies with patients affected by specific diseases [11,12]. A few studies specifically describing the intestinal *E. coli* of healthy subjects mainly focused on antibiotic resistance, without performing a thorough genetic and phenotypic analysis of the strains [13–15]. The present study aimed to deeply characterize the population of *E. coli* isolated from the feces of 20 healthy adults in order to determine whether the relationship between PFGE genotyping, phylogroups, genetic determinants, and functional features can be established. A set of 51 strains was analyzed and characterized according to a generally accepted phylotype classification of *E. coli* that includes seven phylogroups of *E. coli sensu stricto* (A, B1, B2, C, D, E, and F) [16,17]. Thirty-four virulence determinants were searched, together with the *pks* island claimed as a concurrent cause in the development of human colorectal cancer [18], the gene *traD*, as an indicator of the presence of conjugative elements [19], the genes encoding E7 colicin (*colE7*), and the immunity protein (*immE7*) [20]. Phenotypical characterization of the isolates included the AR profile, the capability to form biofilms, and the presence of the surface structures involved in adherence.

2. Results

2.1. Quantitation, Phyloptyping, and Genotyping

E. coli were quantified in the feces of 20 healthy subjects by enumeration onto selective plates and through qPCR. These techniques yielded consistent values ($p = 0.34$), ranging from 3.4×10^5 to

9.3×10^7 cfu, or cells per g of feces (Figure 1; Table S1). The ratio between the *E. coli* and total bacteria ranged from 5.6×10^{-6} to 1.9×10^{-3}, with a mean of 2.8×10^{-5}.

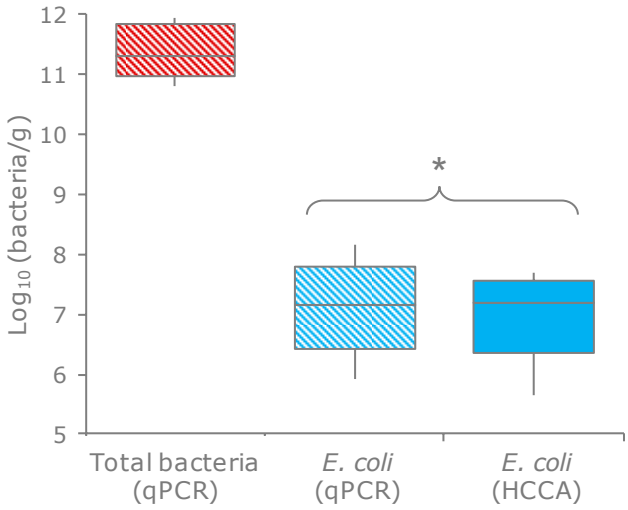


Figure 1. Counts of *E. coli* and total bacteria in the feces of 20 subjects. Boxes indicate the median and 25th and 75th percentiles; whiskers indicate the 10th and 90th percentiles. * indicates the means that did not significantly differ ($p > 0.05$, paired samples t-test). HCCA, HiCrome Coliform Agar (the selective medium for *E. coli*).

Ninety-six colonies per sample were randomly selected and clustered based on their ERIC-PCR and RAPD-PCR profiles (Table 1). When profiles were similar across different hosts, an isolate was picked from each of the subjects. A collection of 51 *E. coli* strains was obtained, also encompassing 10 β-glucuronidase-negative clones isolated from the same samples during a parallel study that aimed to characterize Enterobacteriaceae other than *E. coli* (which were ascribed to this species by a MALDI Biotyper matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) system (Bruker Daltonik GmbH, Bremen, Germany) [21]). Within each sample, one to five biotypes were recognized, with most of the samples (14 out of 20) characterized by a dominant biotype (>80%). According to the XbaI PFGE analysis, the 51 strains were ascribed to 44 different pulsotypes (Figure 2).

The strains were assigned to a phylogroup according to Clermont et al. [17] (Table 1). The most prevalent phylogroup was B2 (14 strains), followed by B1, F, A, D, E, and C (12, 7, 7, 5, 4, and 2 strains, respectively). The strains of phylogroup B2 presented the highest absolute charge and were the most abundant within the *E. coli* population, accounting for the all colonies recovered from several samples. B1 phylogroups were detected at lower concentrations and represented a minor subpopulation (Table 1).

Table 1. *E. coli* isolates from 20 healthy subjects. The strains were genotyped by ERIC-PCR, RAPD-PCR, and PFGE analysis. The relative abundance represents the ratio of each strain within the set of *E. coli* clones analyzed. The charge of the viable cells has been calculated by multiplying the total *E. coli* count and the relative amount. Phlylotyping was done according to Clermont et al. 2013 [17]. The presence of β-glucuronidase was confirmed by growth of blue colonies on the HCCA medium. The performance of isolates as recipients in conjugation experiments was assessed only in β-glucuronidase negative strains. Biofilm formation was tested on LBWS and M9glu media. Curli and cellulose production was determined.

E. coli Strain	Genotyping and Phyloptyping						Phenotype Assays					
	ERIC-PCR			RAPD-PCR Profile	*Xba*I-PFGE Profile	Phylo-Group	β-glu	Conju-Gation	Biofilm		Curli	Cellulose
	Profile	Relative Abundance	Log (cfu/g)						LBWS	M9		
01.01	E01	100%	7.56	R01	P01	B1	+		−	−	−	−
02.03	E02	8%	5.22	R02	P02	B2	+		−	−	−	−
02.16	E03	92%	6.28	R03	P03	F	+		+	−	+	+
03.25	E04	99%	7.52	R04	P04	B2	+		+	−	−	−
03.73	E05	1%	5.53	R03	P05	A	+		+	+	−	−
04.05	E06	5%	4.22	R05	P06	A	+		+	−	+	−
04.06	E07	70%	5.37	R06	P07	B2	+		−	−	−	−
04.10	E08	25%	4.92	R07	P08	E	+		−	−	+	+
05.20	E07	96%	7.78	R02	P09	B2	+		−	−	−	+
05.47	E09	4%	6.40	R08	P10	B1	−	+	−	−	−	−
06.16	E07	100%	7.28	R07	P11	B2	+		−	−	−	−
07.09	E10	100%	6.45	R04	P12	D	+		+	−	+	−
08.01	E11	18%	4.91	R09	P12	D	+		+	−	+	−
08.02	E12	5%	4.37	R10	P13	A	+		−	+	−	−
08.10	E13	5%	4.37	R11	P14	A	+		+	+	+	−
08.27	E14	36%	5.20	R11	P15	B1	−	+	−	+	−	−
08.57	E15	34%	5.18	R12	P16	F	−	+	−	−	−	−
08.109	E16	2%	3.86	R13	P17	F	−	+	−	−	−	−
09.13	E01	92%	7.09	R14	P18	C	+		+	−	+	+
09.21	E17	4%	5.73	R12	P19	F	+		−	−	−	−
09.42	E07	4%	5.73	R12	P09	B2	+		−	−	−	−
10.17	E04	100%	7.62	R12	P04	B2	+		+	−	−	−

Table 1. *Cont.*

E. coli Strain	Genotyping and Phyloptyping						Phenotype Assays				
	ERIC-PCR			RAPD-PCR Profile	*Xba*I-PFGE Profile	Phylo-Group	β-glu	Conju-Gation	Biofilm		Curli	Cellulose
	Profile	Relative Abundance	Log (cfu/g)						LBWS	M9		
11.02	E18	50%	7.67	R15	P20	A	+		−	−	−	−
11.04	E19	9%	6.91	R10	P21	A	+		−	−	−	−
11.06	E07	37%	7.54	R16	P22	B2	+		−	−	−	−
11.16	E20	4%	6.57	R17	P23	B2	+		−	−	−	−
12.04	E13	71%	7.43	R18	P24	B2	+		−	−	+	−
12.09	E11	29%	7.04	R18	P25	D	+		−	−	−	−
13.01	E01	3%	4.98	R19	P26	B1	+		−	−	+	−
13.03	E11	1%	4.47	R11	P27	D	+		−	−	+	+
13.20	E21	97%	6.46	R13	P28	F	−	−	−	−	−	−
14.03	E22	100%	6.85	R13	P29	E	+		−	−	+	−
15.01	E07	9%	6.45	R13	P30	B2	+		−	−	−	+
15.02	E01	87%	7.45	R11	P31	B1	+		+	−	+	+
15.13	E11	4%	6.15	R20	P32	D	+		−	−	+	−
16.04	E21	1%	5.74	R21	P33	F	+		−	−	−	−
16.15	E01	1%	5.74	R22	P34	B1	+		+	−	+	−
16.27	E21	98%	7.73	R23	P33	F	−	−	−	−	−	−
17.10	E23	87%	7.45	R23	P35	A	+		−	−	−	−
17.18	E07	13%	6.63	R24	P36	B2	+		+	−	−	−
18.03	E24	39%	6.25	R25	P37	B1	+		−	+	−	−
18.09	E13	41%	6.28	R25	P24	B2	+		−	−	+	−
18.18	E25	19%	5.93	R13	P38	E	+		−	−	−	+
18.50	E26	1%	4.70	R21	P39	B1	−	+	−	−	−	−
19.02	E07	100%	7.19	R14	P40	B2	+		+	−	−	−
20.04	E27	12%	6.39	R26	P41	B1	+		−	−	+	−
20.17	E28	5%	6.01	R12	P42	C	+		−	−	+	−
20.27	E29	53%	7.03	R27	P43	E	+		+	−	+	+
20.50	E30	22%	6.65	R04	P44	B1	−	+	−	−	−	−
20.51	E31	4%	5.91	R06	P44	B1	−	+	−	−	−	−
20.71	E09	4%	5.91	R06	P44	B1	−	+	−	−	−	−

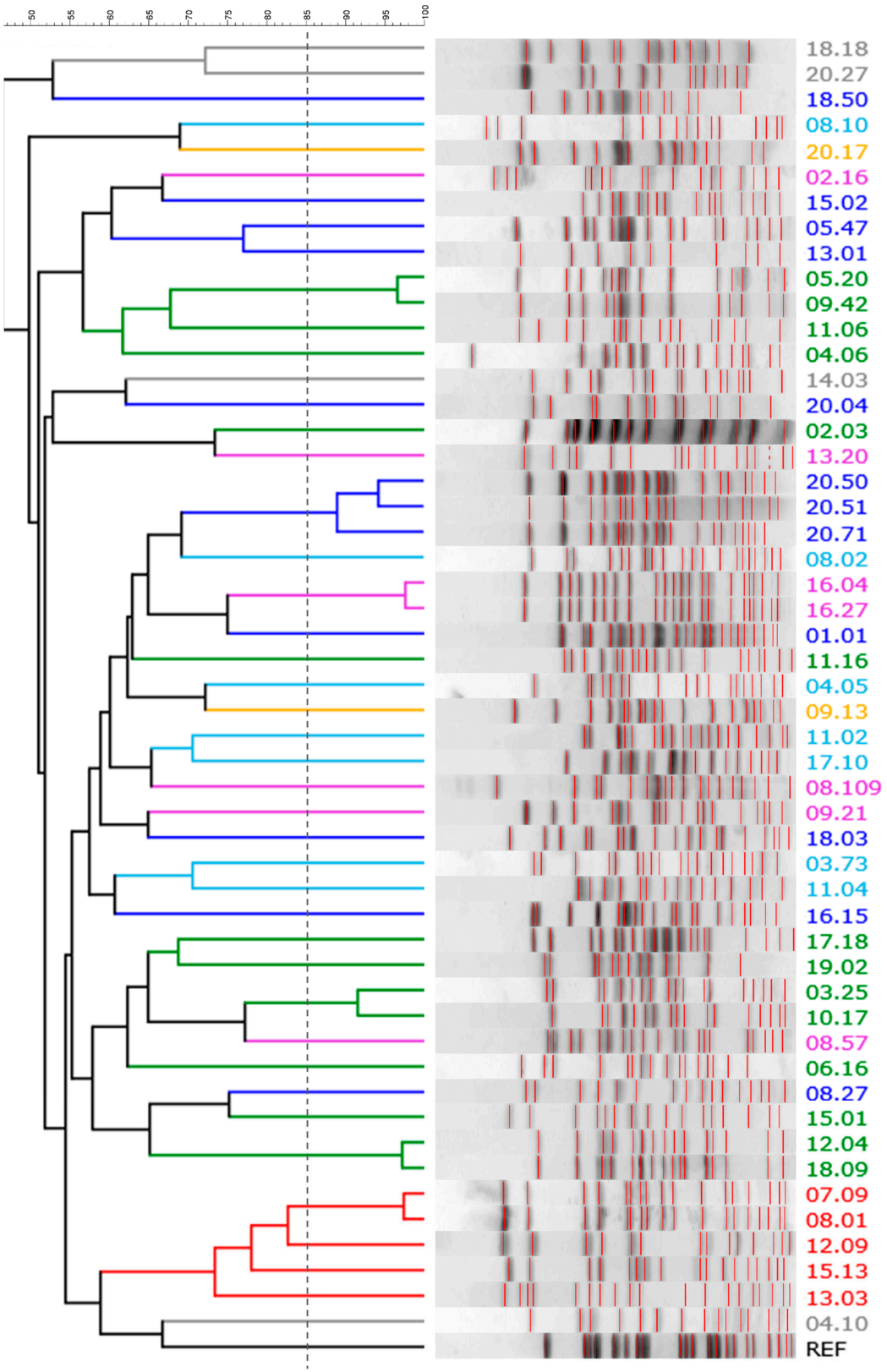


Figure 2. The *Xba*I-PFGE pattern of *E. coli* strains: a UPGMA dendrogram derived from Dice's coefficients, calculated based on their band profiles. Patterns with similarities >85% (dashed line) were ascribed to the same pulsotype. Strains are colored based on their phylogroup: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

2.2. Virulence Factors

Thirty-four genetic determinants encoding virulence factors that potentially enhance the risk of infections were screened by PCR. Most of the strains were positive for *fimH*, *fyuA*, *ompT*, *traT*, *chuA*, and *kpsMTII* (47, 36, 35, 35, 32, and 27 strains, respectively). On the other hand, *afa*/draBC, *cnf1*, *focG*, *hlyD*, *ibeA*, *rfc*, *sfa*/*focDE*, and *sfaS* recurred rarely (≤ 5 strains), while *cdtB*, *gafD*, *pic*, and *vat* were absent in all the strains (Figure 3). Based on virulence factors, all the strains belonging to the phylotypes B2, D (with an exception), E, and F clustered together (Figure 4). Strains ascribed to the phylotypes C were grouped in another cluster that also included the majority of B1 and most of the A isolates. A single B1 strain lay outside both clusters.

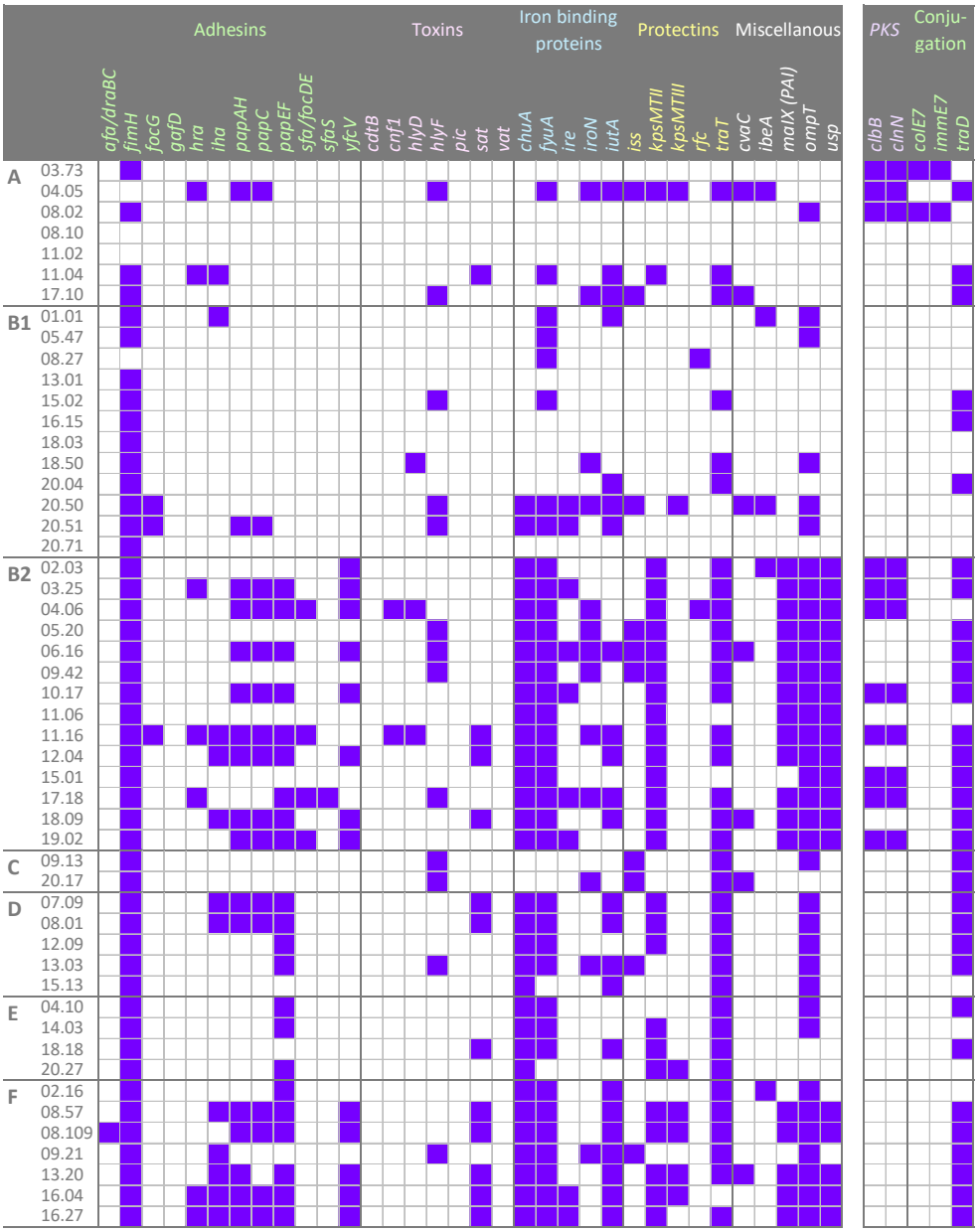


Figure 3. The pattern of genetic determinants in the 51 *E. coli* strains, ordered on the basis of their phylotypes. Virulence genes coding for adhesins, toxins, iron binding proteins, and protectins were screened. The presence of the *pks* pathogenicity island and other genes involved in the conjugation machinery were searched. Purple squares, presence; white squares, absence.

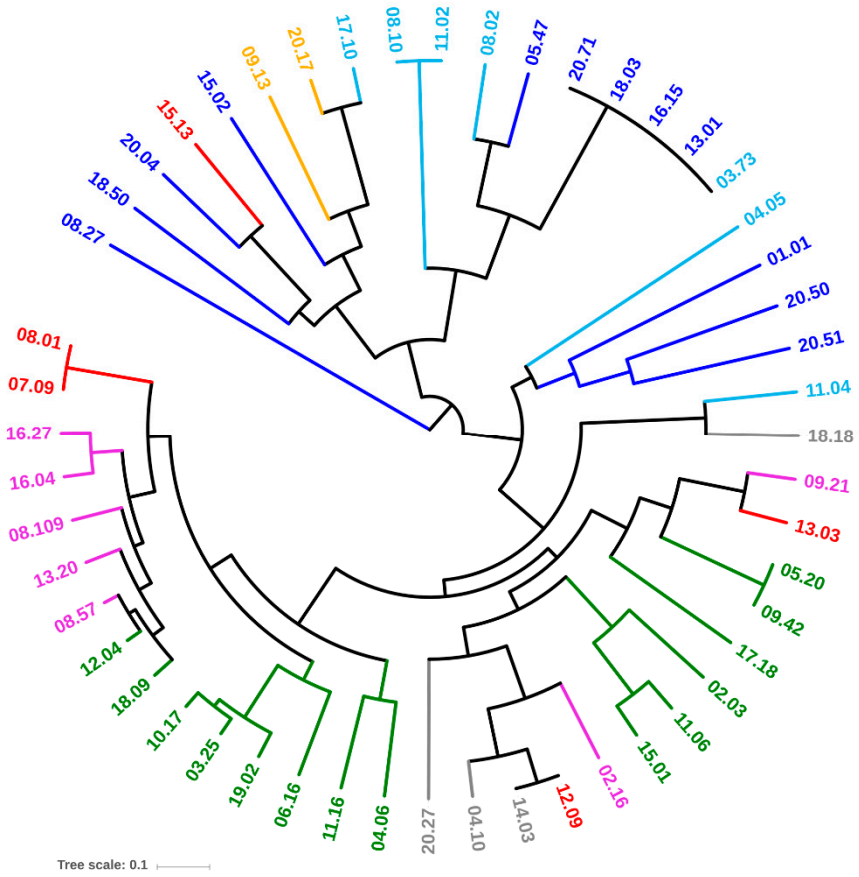


Figure 4. UPGMA dendrogram of *E. coli* strains, computed from Dice's distance matrix of the virulence determinants. Strains are colored based on their phylogroups: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

2.3. Other Genetic Determinants Affecting Fitness and Pathogenicity

E. coli strains were screened for the presence of the *pks* pathogenicity island (Figure 3). The genes *clbB* and *clbN*, utilized as markers, were found in 11 out of 51 strains, 3 ascribed to phylotype A and 8 to B2.

The presence of *traD* was investigated to establish the potential in conjugative DNA exchange. Most isolates were positive to *traD* (32/51), including the the majority of pks-positive strains (Figure 3). *traD* was distributed differently among the phylogroups: most B2, C, D, and F strains harbored the gene, while only a minority within A and B1 did the same. When the same subject harbored more biotypes, some isolates bore *traD* while others did not.

The genes *colE7* and *immE7*, respectively encoding the bactericidal nuclease Colicin E7 and the corresponding immunity protein, were searched. Two strains harbored both the genes.

2.4. Antibiotic Susceptibility

Phenotypic antibiotic susceptibility tests were carried out on the 51 *E. coli* strains (Table S1). Nearly all the isolates were extensively sensitive to the whole set of tested antibiotics (49/51). Only 2 strains recovered from the same fecal sample, belonging to the phylogroups B2 (17.18) and A (17.10), were resistant to gentamycin. *E. coli* 17.10 was also resistant to ciprofloxacin.

2.5. Conjugation

The 10 β-glucuronidase-negative *E. coli* strains, belonging to phylotypes B1 (6) and F (4), were challenged as conjugation recipients for receiving the pOX38: Cm plasmid from *E. coli* N4i (Table 1).

The plasmid transfer to most of the strains succeeded with high efficiency: all the B1 strains and 2 out of the the 4 F strains acquired the plasmid.

2.6. Production of Curli and Cellulose and Biofilm Formation

The production of curli was assessed by observing the colonies grown in CR-containing LB agar plates. B2 and F strains presented the lowest ability to bind CR dye (2/14 and 1/7, respectively), whereas C, D, and E strains seemed to be inclined to produce curli (Table 1). The occurrence of cellulose-like extracellular components was analyzed by growing the isolates on LB plates supplemented with CF. The ability of the strains to bind CF was less frequent (9/51) than their ability to bind CR.

E. coli strains were assayed for biofilm formation in LBWS and M9glu. Most of the strains did not form a biofilm (Figure 5). Formation occurred more frequently in LBWS than in M9glu (13/51 and 5/51 strains, respectively). Only two strains, both ascribed to phylotype A (03.73 and 08.10), produced a biofilm in LBWS and M9glu. For five other strains, the score of the biofilm formation ranged between 0.6 and 0.8. All but one of these strains presented a very high standard deviation (>0.5; Figure 5), suggesting that biofilm formation was sensitive to environmental conditions that were hardly controlled by the operator.

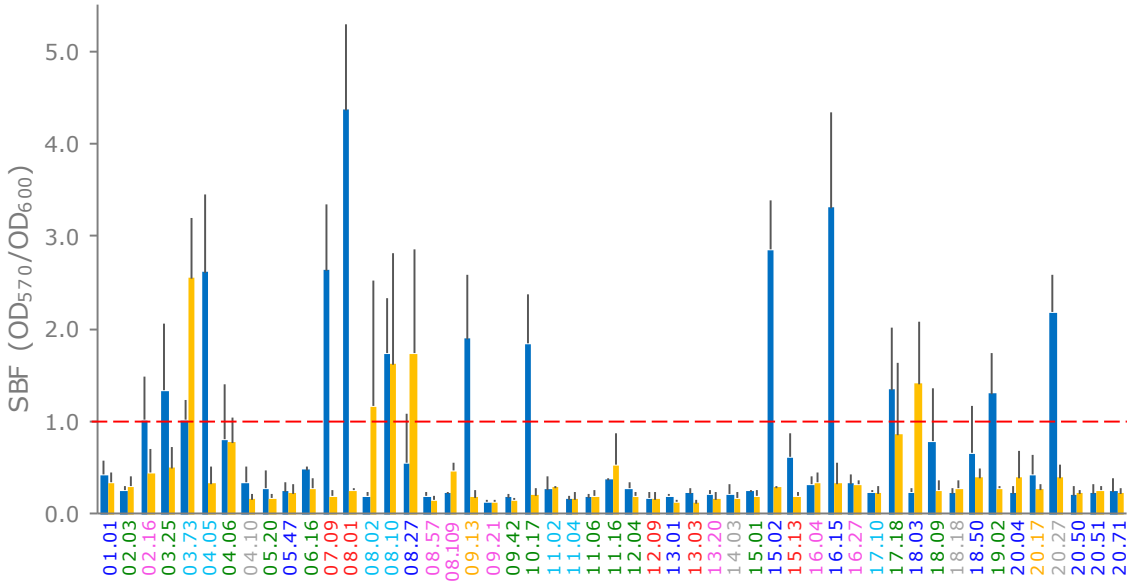


Figure 5. formation of *E. coli* strains on LBWS (blue bars) and M9glu (yellow bars). The specific biofilm formation index (SBF) is calculated as the ratio between the crystal violet absorbance at 570 nm and the culture turbidity at 600 nm, setting a threshold of 1 for biofilm producers (red dashed line). The reported data are means ± standard deviations of at least three independent experiments, each carried out in triplicate. The names of the strains are colored based on their phylogroups: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; and F, pink (hereafter referred to as LBWS and M9glu, respectively).

2.7. Co-Occurrence and Principal Coordinate Analysis (PCoA) Analysis of Genetic and Functional Features

A co-occurrence analysis was applied to the whole set of genetic and function observations (Figure 6). The presence of pks islands and several virulence determinants was significantly associated ($p < 0.05$ or < 0.01) to certain genotypic classifications (such as the ERIC profile, the pulsotype, and the phylogroup). In particular, the genes encoding some adhesins (*iha*, *papAH*, *papC*, *papF*, and *sat*), iron binding proteins (*chuA*, *fyuA*, and *iutA*), protectins (*kpsMTII* and *traT*), and *malX*, *ompT*, *usp*, and *traD* were significantly associated to phylogroups and presented a significantly positive tendency to co-occur within the same strains ($p < 0.05$ or 0.01).

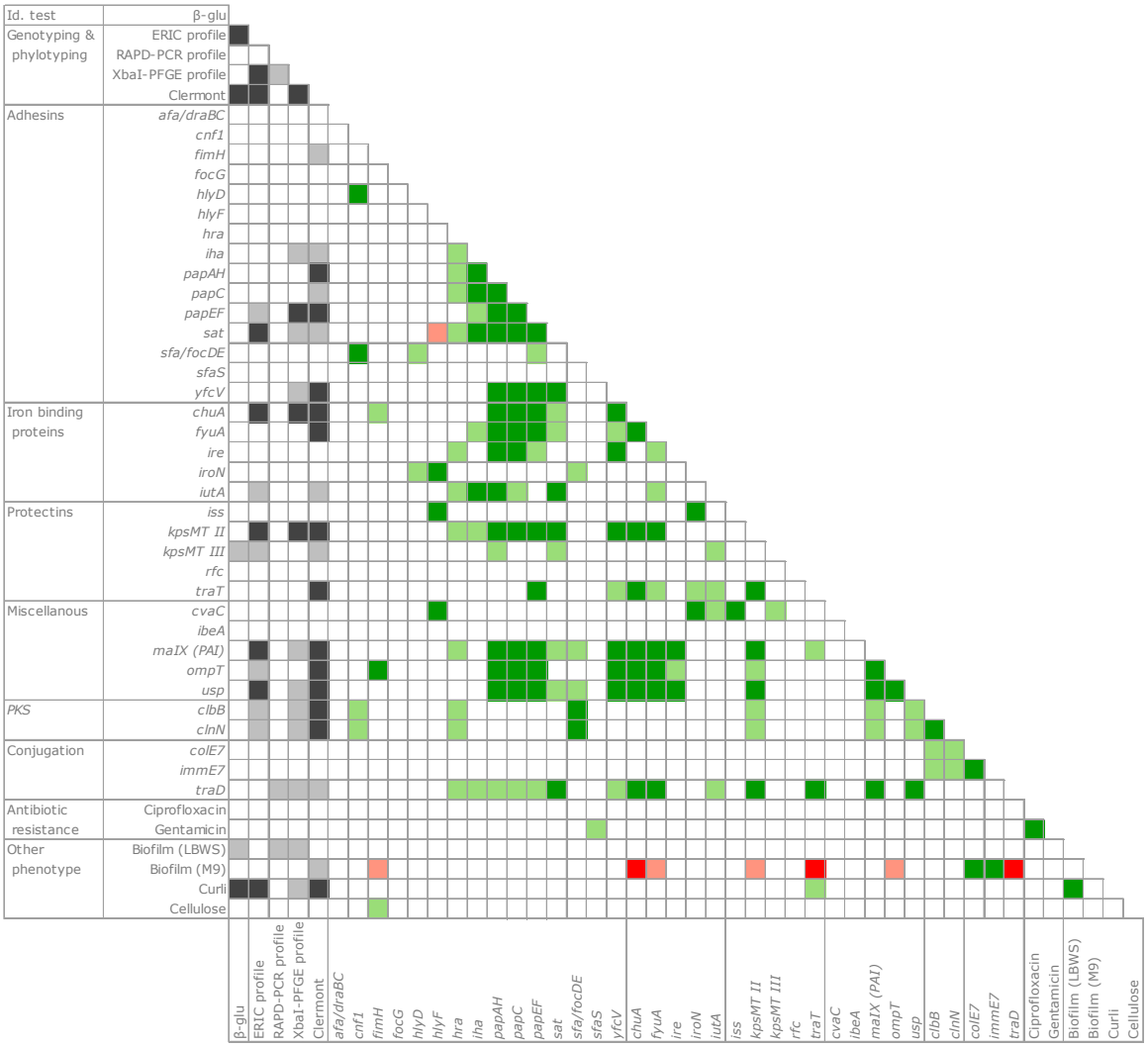


Figure 6. Genetic determinants and phenotypic properties evaluated by Cramér's V metrics. Association with genotype/phylotype clusters, grey; positive co-occurrence, green; negative co-occurrence, red. Dark and light shades indicate the levels of statistical significance ($p < 0.01$ and $p < 0.05$, respectively).

With regards to the phenotypic features involved in bacterial adhesion, the production curli were significantly associated with the phylogroup ($p < 0.01$). Although biofilm formation was not frequent, curli presented a significant co-occurrence with the biofilm in the LBWS medium.

A Principal Coordinate Analysis (PCoA) was applied to assist in the identification of clusters, taking into account both their genetic and functional features (Figure 7). The first two coordinates were the most informative, accounting for approximately 28.9% and 10.2% of the total observed variances. In the PCoA plot, PCo1 separated B2 and F strains from B1, C, and E ones (negative and positive PCo1, respectively). F strains were grouped in two clusters. With respect to PCo2, most of the A, C, D, and E strains lie in the positive quadrants.

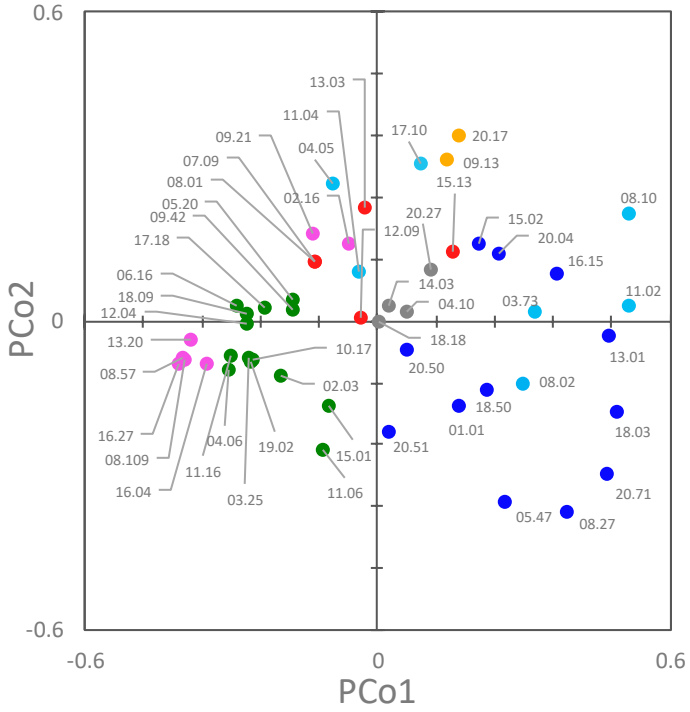


Figure 7. 2D Principal Coordinate Analysis (PCoA) visualization of *E. coli* strains, computed from the Dice's distance matrix of genetic and phenotypic determinants. Strains are colored based on their phylogroup: A, cyan; B1, blue; B2, green; C, yellow; D, red; E, grey; F, pink.

3. Discussion

Fifty-one *E. coli* strains were isolated from the feces of 20 healthy adults and genotypically and phenotypically characterized. Most of the strains belonged to phylogroups B1 (12) and B2 (14). Several B2 strains presented a greater attitude to colonization, with higher charges and an overall dominance in the strains belonging to other phylotypes. This result is consistent with the fact that B2 strains showed an enhanced ability to persist in the intestinal microbiota, as demonstrated in infants [22,23]. B1 strains, on the contrary, showed lower performance in colonization, generally representing minor *E. coli* subpopulations within each sample.

Phylogroup classification and PFGE pulsotypes were consistent, but the strains belonging to the same phylogroups were spread over different clades of the PFGE dendrogram, with only D isolates clustered together. PFGE analysis targeted the whole genome, most of which is composed of accessory genes that can be acquired by horizontal gene transfer. The PFGE fingerprint of *E. coli* results from the digestion of a genome, the size of which, including plasmids and prophages, ranges approximately from 4.6 to 5.9 Mbp, encoding from 4200 to 5500 genes [3]. This huge difference is mainly due to the genes associated with bacteriophage elements and involved in virulence or resistance to antimicrobials. The phylotypes recognized in this study were shuffled in PFGE, likely because of the genome evolution, which was mainly due to the recombination and horizontal transfer that broke the old phylotype relationships, thereby emphasizing more recent changes. On the other hand, MLST typing, the gold-standard approach for *E. coli* classification, allows a phylotyping-consistent analysis, because MLST is based on the conserved nature of the housekeeping genes of the core genome [24]. However, when both the genetic and functional features of each strain were elaborated together in this study (Figure 7), the different phylogroups clustered separately in a PCoA space.

The majority of the strains were sensitive to all the tested antibiotics, including amoxicillin plus clavulanic acid and cephalosporins, the most diffused antibiotics for the first management of infections. Two sole exceptions, resistant to gentamycin and ciprofloxacin, were isolated from the same subject. The absence of the AR phenotype does not exclude the presence of AR genes, which may be expressed

in vivo or can be involved in the diffusion and spread of AR genes [25]. However, the rapid emergence of resistant bacteria occurring worldwide attributed to the overuse and misuse of antibiotics, healthy subjects harbor an endogenous *E. coli* population still sensitive to a wide range of antibiotics, contrary to the pathogenic strains isolated from different clinical specimens [26–28]. The scarce relevance of commensal *E. coli* as an AR carrier disagrees with other studies [29–32] but is consistent with a recent analysis of metagenomes that investigated AR genes in the gut microbiota of healthy people. As a whole, the results show that commensal *E. coli* may not pose a threat by itself in terms of AR, though major differences are registered among countries [33,34].

The great variability of pathogenicity-associated features suggests that the genetic determinants of virulence had a role in shaping the genome of the isolates. A significant co-occurrence was observed in a set of B2 and F strains for *pap* genes (*papAH*, *papC*, *papEF*), encoding proteins of the fimbria favoring the ascension of the urinary tract and promoting colonization and infection, *sat*, which is associated with a cytopathic secreted autotransported toxin exerting effect, *kpsMTII*, the marker of K2 capsular polysaccharides, which plays a main role in pathogenesis, *chuA* (heme receptor), *yfcV* (Yfc fimbria), along with *fyuA*, *malX*, *usp*, *ompT* (Figure 3; Figure 6). Among them, twelve strains, belonging to the phylogroups B2 (03.25, 04.06, 06.16, 10.17, 12.04, 18.09, and 19.02) and F (08.57, 08.109, 13.20, 16.04, and 16.27), presented a pattern of virulence genes similar to that of the strains responsible for urinary tract infections, on the basis of the presence of the genes *chuA* (heme receptor), *yfcV* (Yfc fimbria), along with *fyuA* (*Yersinia* siderophore receptor) and *ompT* (aspartyl protease) [35]. This reinforces the idea that it is possible that pathogenic *E. coli* potentially inhabit the gut of healthy subjects without instigating infections, although they can act as etiologic agents of extra-intestinal infections, without a clear distinction between commensal *E. coli* and pathogens.

Capsular polysaccharides represent a class of macromolecules contributing to the surface properties. They are involved in important biological processes including adhesion and resistance to the host's immune responses, such as complement-mediated killing and phagocytosis [5]. Many strains (27/51) carried *kpsMTII*, among which six strains also carried *kpsMTIII*, encoding another capsular antigen. Macromolecules on the surface of bacteria confer ultrastructural stability are important for recognition by, and interaction with, the environment and in pathogenic bacteria form a defensive barrier against the host's immune system. All but one strain was negative to the *afa/draBC* gene, excluding the presence of afimbrial adhesins encoded by the *afa* operons [36]. Likewise, for capsular polysaccharides, curli are involved in adhesion and strengthen *E. coli* defenses, particularly against complement-mediated killing by the host, thereby contributing to the pathogenic potential of strains [37]. The positive relationship between curli and biofilm formation was confirmed in the present study, although the latter feature was not frequent among the observed strains. Interestingly, the phylogroups B2 and F, which include the strains with the largest set of potential pathogenic determinants, are less prone to produce curli, a condition that may increase their susceptibility to host defenses.

E. coli pathogenicity islands serve as integration sites for genetic elements. Thus, virulence genes and AR determinants can be rapidly and simultaneously acquired, changing commensal strains into major risk agents [38]. Nineteen stains, representing most of the B2 or F isolates, harbored a PAI island, which produced a positive result for the maltose- and glucose-specific component IIa of the phosphoenolpyruvate dependent phosphotransferase system (*malX*). Furthermore, all the isolates also carried the gene *usp*, encoding a putative bacteriocin generally located on the PAI [31]. Eleven out of 51 strains, belonging to the phylogroups A and B2, harbored the 54-kb polyketide synthases *pks* pathogenicity island, encoding the multi-enzymatic machinery for the synthesis of a peptidepolyketide hybrid genotoxin called colibactin [39]. *hlyF*, encoding for the toxin α-emolysin [40], occurred in other strains, which were spread over most of the phylotypes.

The potentiality of the isolates to exchange genetic material was assessed by searching the gene *traD* and challenging the β-glucuronidase negative *E. coli* strains as recipients in conjugation experiments. *traD* is a key gene in the conjugation process, encoding a main component in the transferosome of type IV secretion systems [19]. The presence of *traD* has been chosen to screen the presence of the

conjugation apparatus and the potentiality of the strains to exchange DNA. The majority of the strains (32/51) carried *traD*, regardless of the phylotype group. When more than a single *E. coli* strain was identified in the same subject (Table 1), some isolates encoded *traD* and others did not. The 9 strains harboring the *iss* gene (responsible for the increased survival of bacteria in the serum and generally carried by the big virulence plasmid, ColV) were all positive to *traD*, according to the fact that ColV plasmids are usually conjugative [41]. The majority of β-glucuronidase-negative strains, assayed as recipients in the conjugation experiment, received the plasmid (8/10), confirming their receptive aptitude for the genetic exchange and DNA shuffling of commensal *E. coli*. Interestingly, 2 of the 8 transconjugant strains harbored *traD*, a putative marker of a conjugative plasmid that is expected to exclude the acquirement of another plasmid through conjugation.

The fitness and competitiveness of the bacteria can be improved by bacteriocin production, which is generally associated with the counterpart immunity protein, thereby protecting the producing host cell from the lethal action of the bacteriocin. Colicins are plasmid encoded toxins produced by *E. coli* under conditions of stress and able to kill related bacteria competing for niches and nutrients. Colicin E7, and its immunity counterpart, ImmE7, have been used to develop bacterial conjugation-based antimicrobial agents [42,43]. The antibacterial activity of the colicin ColE7, bacterial "kill"–"anti-kill" antimicrobial system has been determined, thereby offering new perspectives in the development of *E. coli* targeted antimicrobials. The strains isolated in this study were screened for the presence of the *colE7* and *immE7* genes, in order to investigate the susceptibility of commensal strains to the action of recombinant antimicrobials. Most of the strains were negative for both the genes, suggesting that, if necessary, they could be potentially vulnerable to this new approach.

4. Materials and Methods

4.1. Isolation and Enumeration and of E. coli

Fresh fecal samples were collected from 20 healthy adult subjects, who were all caucasians: 10 males and 10 females aged 35 to 45, following a western omnivore diet, who had not been treated with prebiotics and/or probiotics for 1 month and antibiotics for 3 months. They did not have any relevant diseases or drug treatments in their medical histories. These subjects were enrolled as volunteers among the employees of the University of Modena and Reggio Emilia and their relatives and were not in a relationship with the researchers. All human clinical samples were collected after the subjects gave written informed consent for their participation in the study. All personal data used in the study were anonymized.

Fresh fecal samples were homogenized (10% *w/v*) and serially diluted in isotonic Buffered Peptone Water (Sigma, Steinheim, Germany). Then, they were spread onto HiCrome Coliform Agar (HCCA, Sigma) and incubated. After 24 h at 37 °C, the putative *E. coli* colonies were checked for indole production with Kovac's reagent (Sigma). Colonies that exhibited consistent reactions were ascribed to the species *E. coli*. Isolates were maintained and propagated in a Luria Bertani (LB) medium aerobically at 37 °C.

A total of 96 *E. coli* clones per sample were subjected to ERIC-PCR [44] and RAPD-PCR [45] and clustered into biotypes with a similarity level of 75% using the Pearson correlation coefficient. For each strain, the concentration was determined by multiplying the relative abundance and total plate count of *E. coli*.

For qPCR quantification, the feces were suspended in PBS pH 7.8 (10% *w/w*). Bacterial gDNA was extracted with the QIAmp DNA Stool Mini Kit (Qiagen, Hilden, Germany), quantified with a NanoPhotometer *p*-Class (Implen GmbH, Munchen, Germany), diluted to 2.5 ng/μL in a TE buffer (pH 8), and subjected to qPCR analysis, with primers targeting eubacteria and *E. coli*. [46]. The mixture contained 10 μL of SsoFast EvaGreen Supermix, 4 μL of each 2 μM primer, and 2 μL of the template. qPCR reactions were carried out with the CFX96 Real-Time System (Bio-Rad Laboratories, Redmond,

WA, USA) according to the following program: 98 °C for 2 min; 45 cycles at 98 °C for 0.05 min, 60 °C for 0.05 min, and 95 °C for 1 min; 65 °C for 1 min.

4.2. Phyloptyping and PFGE Genotyping

Standard multiplex PCR was used for all the isolates, according to the revisited phylotyping method published by Clermont et al. [17]. PFGE was performed according to the PulseNet protocol [47]. The genomic DNA was digested with 50 U of *Xba*I at 37 °C for 3 h. Macrorestriction fragments were resolved by counter-clamped homogeneous electric field electrophoresis in a CHEF-DRIII apparatus (Bio-Rad, USA). The run time was 24 h at 6.0 V/cm, with initial and final switch times of 2.2 s and 54.2 s, respectively. PFGE images were digitally captured and analyzed with the GelCompare II 6.0 software (Applied Maths NV, Belgium). Dice's coefficients were calculate based on the band profiles, setting the position tolerance at 1% and the optimization at 1%. A similarity dendrogram was derived using the unweighted pair group method with arithmetic means (UPGMA). Strains were ascribed to the same pulsotype if the PFGE profile possessed >85% similarity.

4.3. Virulence Genotyping

All the isolates were screened by multiplex-PCR for a carriage of 34 genes associated to virulence factors: *afa/draBC*, *cdtB*, *chuA*, *cnf1*, *cvaC*, *fimH*, *focG*, *fyuA*, *gafD*, *hlyD*, *hlyF*, *hra*, *ibeA*, *iha*, *ire*, *iroN*, *iss*, *iutA*, *KpsMTII*, *kpsMTIII*, *malX* (*PAI*), *ompT*, *papAH*, *papC*, *papEF*, *pic*, *rfc*, *sat*, *sfa*, *sfaS*, *traT*, *usp*, *vat*, and *yfcV*. Primer sequences, the pool of primers, and amplification reactions were set up according to Johnson et al. [48].

4.4. Detection of the Genes traD, colE7, immE7, and of pks Island

A triplex-PCR was utilized to screen the isolates for the presence of the genes *colE7*, *immE7*, and t*raD*, utilizing the primers, colE7-F/colE7-R2 (5′-AAGTCAGATGCTGATGTTGC-3′/ 5′-ATAGACACCACCATTTTGGC-3′), immE7-F/immE7-R2 (5′-GCAACTGATGATGTGTTAGATG-3′/ 5′-TGTTTAAATCCTGGCTTACCG-3′), and traD-F/traD-R2 (5′-GATAAATACATGATCTGGTGCGG-3′/ 5′-TTGATGGAAAGCAGTGTGTC-3′), respectively, specifically developed in this study. The amplification reaction was carried out in a final volume of 50 µL containing 5 µL of 10X Dream Taq Green Buffer, 5 µL dNTP Mix (2 mM each), 0.25 µL Dream Taq, 9 µL primer mix (2 µM each), and 50 ng of template DNA. The thermocycle was: 95 °C for 5 min, 30 cycles at 95 °C for 30 sec, 58 °C for 30 sec, and 72 °C for 1 min; and 72 °C for 7 min.

The polyketide synthases *pks* pathogenicity island was searched by duplex-PCR, amplifying the *clbB* and *clbN* genes located at the 5′ and 3′ regions of the island, according to Sarshar et al. [49].

4.5. Antibiotic Susceptibility

Antimicrobial susceptibility was tested with a Vitek2 semi-automated system (bioMerieux, France). Minimum inhibitory concentrations (MICs) were interpreted according to EUCAST (European Committee on Antimicrobial Susceptibility Testing—www.eucast.org). Clinical breakpoints for the tested antibiotics were expressed as mg/L (S, susceptible; R, resistant): for amikacin, S ≤ 8 and R > 16; for amoxicillin/clavulanic acid, S ≤ 8 and R > 8; for cefotaxime, 1 ≤ 8 and R > 2; for ceftazidime, S ≤ 8 and R > 8; for ciprofloxacin, S ≤ 0.5 and R > 1; for gentamicin, S ≤ 2 and R > 4; for piperacillin + tazobactam, S ≤ 8 and R > 16; and for trimethoprim/sulfamethoxazole, S ≤ 40 and R > 80.

4.6. Biofilm and Phenotype Assays

Biofilm formation was assayed by crystal violet (CV) staining, as described in [50]. *E. coli* strains were cultured in LB without NaCl and in an M9 medium (BD Difco, Sparks, Maryland, USA) containing 4 g/L glucose and 0.25 g/L yeast extract (hereinafter referred to as LBWS and M9glu, respectively). The specific biofilm formation (SBF) index was calculated as the ratio between CV absorbance at 570 nm

and the culture's turbidity at 600 nm, setting a threshold of 1. The data reported here are the means of 3 independent experiments, each carried out in triplicate.

Curli and cellulose formation were assayed in LBWS agar plates supplemented with congo red (CR) or calcofluor white (CF), respectively, according to [51]. Red colonies were considered positive to CR. Emission of fluorescence as result of UV exposure (315–400 nm) was considered positive to CF.

4.7. Solid Mating Conjugation Experiments

The donor strain *E. coli* N4i pOX38: Cm (N4i: EcN immE7 Gmr; pOX38: Cm: Tra+ RepFIA+ Cmr) was derived from *E. coli* Nissle 1917 [42,43]. The donor was cultured aerobically at 37 °C in LB supplemented with 15 μg/mL gentamycin and 20 μg/mL chloramphenicol, while the recipients were cultured in LB. Overnight donor and recipient cultures were seeded in their respective media (1% v/v) and aerobically incubated for 2 h. One mL of the recipient exponential culture was centrifuged (10,000× *g* for 5 min at 4 °C) and resuspended in 400 μL of the donor exponential culture. The whole volume was spread onto an LB–agar plate and incubated at 37 °C for 4 h, and then the cells were collected from the surface with 1 mL of PBS, serially diluted, and spread onto HCCA to differentiate the recipient and transconjugant from the donor colonies and HCCA supplemented with 20 μg/mL of chloramphenicol to select and differentiate the transconjugant from the donor colonies.

4.8. Statistical Analysis

Dice's coefficient was utilized to gauge the distance between strains, based on the sets of binary data, i.e., the presence/absence of genes or phenotypic properties. Distance matrices between strains were utilized to compute and display UPGMA dendrograms and Principal Coordinates Analysis plots. The co-occurrence of genetic determinants and phenotypic properties were evaluated by means of contingency tables and Cramér's V metrics, considering the limit of significance as $p < 0.01$ and $p < 0.05$. Statistical analyses were performed with the SPSS Statistics 21 software (Armonk, New York, USA).

5. Conclusions

A characterization of *E. coli* isolated from fecal samples of healthy subjects was performed. Several strains had the potential to cause extraintestinal infections because of the presence of genes associated with adhesins, siderophores, and toxins. The recurrence of strains sharing a pattern of virulence genes similar to that of potentially pathogenic strains may pose a health threat. However, a main outcome of this study shows the great sensitivity of the isolates to antibiotics, which are susceptible to the common antimicrobial therapy used for Gram-negative bacteria, such as amoxicillin/clavulanate. This suggests that the *E. coli* populations inhabiting the gut of healthy subjects and of patients are more likely differentiated on the basis of AR than virulence factors, emphasizing the importance of the prudent use of antibiotics in the general population.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/7/8/251/s1. Table S1: Dataset of all the genetic determinants and the phenotypic features of *E. coli* isolates.

Author Contributions: Conceptualization, M.R; data curation, S.R., F.C. and A.A.; investigation, L.R., E.M. and F.B.; methodology, S.R., L.R. and G.G.; supervision, S.R.; writing—original draft, S.R., A.A. and M.R.; writing—review & editing, G.G. and M.S.E.

Funding: No dedicated funds.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Russo, T.A.; Johnson, J.R. A proposal for an inclusive designation for extraintestinal pathogenic *Escherichia coli*: ExPEC. *J. Infect. Dis.* **2000**, *181*, 1753–1754. [CrossRef] [PubMed]
2. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* **2019**, *32*. [CrossRef] [PubMed]

3. Leimbach, A.; Hacker, J.; Dobrindt, U. *E. coli* as an all-rounder: The thin line between commensalism and pathogenicity. *Curr. Top Microbiol. Immunol.* **2013**, *358*, 3–32. [CrossRef] [PubMed]
4. Croxen, M.A.; Law, R.J.; Scholz, R.; Keeney, K.M.; Wlodarska, M.; Finlay, B.B. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26*, 822–880. [CrossRef] [PubMed]
5. Sarowska, J.; Futoma-Koloch, B.; Jama-Kmiecik, A.; Frej-Madrzak, M.; Ksiazczyk, M.; Bugla-Ploskonska, G.; Choroszy-Krol, I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathog.* **2019**, *11*, 10. [CrossRef]
6. Searle, L.J.; Méric, G.; Porcelli, I.; Sheppard, S.K.; Lucchini, S. Variation in siderophore biosynthetic gene distribution and production across environmental and faecal populations of *Escherichia coli*. *PLoS ONE* **2015**, *10*. [CrossRef]
7. Welch, R.A. Uropathogenic *Escherichia coli*-Associated Exotoxins. *Microbiol. Spectr.* **2016**, *4*. [CrossRef]
8. Kaas, R.S.; Friis, C.; Ussery, D.W.; Aarestrup, F.M. Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse *Escherichia coli* genomes. *BMC Genom.* **2012**, *13*, 577. [CrossRef]
9. Koraimann, G. Spread and persistence of virulence and antibiotic resistance genes: A ride on the F plasmid conjugation module. *EcoSal Plus* **2018**, *8*, 1. [CrossRef]
10. Johnson, J.R.; Stell, A.L. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **2000**, *181*, 261–272. [CrossRef]
11. Kudinha, T.; Johnson, J.R.; Andrew, S.D.; Kong, F.; Anderson, P.; Gilbert, G.L. Distribution of phylogenetic groups, sequence type ST131, and virulence-associated traits among *Escherichia coli* isolates from men with pyelonephritis or cystitis and healthy controls. *Clin. Microbiol. Infect.* **2013**, *19*. [CrossRef] [PubMed]
12. Lee, J.G.; Han, D.S.; Jo, S.V.; Lee, A.R.; Park, C.H.; Eun, C.S.; Lee, Y. Characteristics and pathogenic role of adherent-invasive *Escherichia coli* in inflammatory bowel disease: Potential impact on clinical outcomes. *PLoS ONE* **2019**, *14*. [CrossRef]
13. Bailey, J.K.; Pinyon, J.L.; Anantham, S.; Hall, R.M. Commensal Escherichia coli of healthy humans: A reservoir for antibiotic-resistance determinants. *J. Med. Microbiol.* **2010**, *59*, 1331–1339. [CrossRef] [PubMed]
14. Mathai, D.; Kumar, V.A.; Paul, B.; Sugumar, M.; John, K.R.; Manoharan, A.; Kesavan, L.M. Fecal carriage rates of extended-spectrum β-lactamase-producing *Escherichia coli* among antibiotic naive healthy human volunteers. *Microb. Drug Resist.* **2015**, *21*, 59–64. [CrossRef] [PubMed]
15. González, D.; Gallagher, E.; Zúñiga, T.; Leiva, J.; Vitas, A.I. Prevalence and characterization of β-lactamase-producing Enterobacteriaceae in healthy human carriers. *Int. Microbio.* **2019**, in press. [CrossRef]
16. Clermont, O.; Bonacorsi, S.; Bingen, E. Rapid and simple determination of the *Escherichia coli* phylogenetic group. *Appl. Environ. Microbiol.* **2000**, *66*, 4555–4558. [CrossRef]
17. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [CrossRef]
18. Wassenaar, T.M. *E. coli* and colorectal cancer: A complex relationship that deserves a critical mindset. *Crit. Rev. Microbiol.* **2018**, *44*, 619–632. [CrossRef]
19. Frost, L.S.; Ippen-Ihler, K.; Skurray, R.A. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiol. Rev.* **1994**, *58*, 162–210.
20. Chak, K.F.; Kuo, W.S.; Lu, F.M.; James, R. Cloning and characterization of the ColE7 plasmid. *J. Gen. Microbiol.* **1991**, *137*, 91–100. [CrossRef]
21. Raimondi, S.; Amaretti, A.; Gentilomi, G.; Bonvicini, F.; Rossi, M. Antibiotic resistance, virulence factors, phenotyping, and genotyping of Enterobacteriaceae (other than *E. coli*) isolated from the feces of healthy subjects. Unpublished; manuscript in preparation.
22. Nowrouzian, F.L.; Wold, A.E.; Adlerberth, I. *Escherichia coli* strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. *J. Infect. Dis.* **2005**, *191*, 1078–1083. [CrossRef]
23. Ostblom, A.; Adlerberth, I.; Wold, A.E.; Nowrouzian, F.L. Pathogenicity island markers, virulence determinants *malX* and *usp*, and the capacity of *Escherichia coli* to persist in infants' commensal microbiotas. *Appl. Environ. Microbiol.* **2011**, *77*, 2303–2308. [CrossRef]

24. Robins-Browne, R.M.; Holt, K.E.; Ingle, D.J.; Hocking, D.M.; Yang, J.; Tauschek, M. Are *Escherichia coli* pathotypes still relevant in the era of whole-genome sequencing? *Front. Cell. Infect. Microbiol.* **2016**, *6*, 141. [CrossRef]
25. Suzuki, S.; Horinouchi, T.; Furusawa, C. Prediction of antibiotic resistance by gene expression profiles. *Nat. Commun.* **2014**, *5*, 5792. [CrossRef]
26. Sahuquillo-Arce, J.M.; Selva, M.; Perpiñán, H.; Gobernado, M.; Armero, C.; López-Quílez, A.; González, F.; Vanaclocha, H. Antimicrobial resistance in more than 100,000 *Escherichia coli* isolates according to culture site and patient age, gender, and location. *Antimicrob. Agents Chemother.* **2011**, *55*, 1222–1228. [CrossRef]
27. Cole, B.K.; Ilikj, M.; McCloskey, C.B.; Chavez-Bueno, S. Antibiotic resistance and molecular characterization of bacteremia *Escherichia coli* isolates from newborns in the United States. *PLoS ONE* **2019**, *14*. [CrossRef]
28. Morley, V.J.; Woods, R.J.; Read, A.F. Bystander selection for antimicrobial resistance: Implications for patient health. *Trends in Microbiol.* **2019**, in press. [CrossRef]
29. Dyar, O.J.; Hoa, N.Q.; Trung, N.V.; Phuc, H.D.; Larsson, M.; Chuc, N.T.; Lundborg, C.S. High prevalence of antibiotic resistance in commensal *Escherichia coli* among children in rural Vietnam. *BMC Infect. Dis.* **2012**, *12*, 92. [CrossRef]
30. Gurnee, E.A.; Ndao, I.M.; Johnson, J.R.; Johnston, B.D.; Gonzalez, M.D.; Burnham, C.A.; Hall-Moore, C.M.; McGhee, J.E.; Mellmann, A.; Warner, B.B.; et al. Gut colonization of healthy children and their mothers with pathogenic ciprofloxacin-resistant *Escherichia coli*. *J. Infect. Dis.* **2015**, *212*, 1862–1868. [CrossRef]
31. Bag, S.; Ghosh, T.S.; Banerjee, S.; Mehta, O.; Verma, J.; Dayal, M.; Desigamani, A.; Kumar, P.; Saha, B.; Kedia, S.; et al. insights into antimicrobial resistance traits of commensal human gut microbiota. *Microb. Ecol.* **2019**, *77*, 546–557. [CrossRef]
32. Purohit, M.R.; Lindahl, L.F.; Diwan, V.; Marrone, G.; Lundborg, C.S. High levels of drug resistance in commensal *E. coli* in a cohort of children from rural central India. *Sci. Rep.* **2019**, *9*, 6682. [CrossRef]
33. Feng, J.; Li, B.; Jiang, X.; Yang, Y.; Wells, G.F.; Zhang, T.; Li, X. Antibiotic resistome in a large-scale healthy human gut microbiota deciphered by metagenomic and network analyses. *Environ. Microbiol.* **2018**, *20*, 355–368. [CrossRef]
34. Pormohammad, A.; Nasiri, M.J.; Azimi, T. Prevalence of antibiotic resistance in *Escherichia coli* strains simultaneously isolated from humans, animals, food, and the environment: A systematic review and meta-analysis. *Infect. Drug Resist.* **2019**, *12*, 1181–1197. [CrossRef]
35. Spurbeck, R.R.; Dinh, P.C.; Walk, S.T.; Stapleton, A.E.; Hooton, T.M.; Nolan, L.K. *Escherichia coli* isolates that carry *vat*, *fyuA*, *chuA*, and *yfcV* efficiently colonize the urinary tract. *Infect. Immun.* **2012**, *80*, 4115–4122. [CrossRef]
36. Servin, A.L. Pathogenesis of *Afa/Dr.* diffusely adhering *Escherichia coli*. *Clin. Microbiol. Rev.* **2005**, *18*, 264–292. [CrossRef]
37. Biesecker, S.G.; Nicastro, L.K.; Wilson, R.P.; Tükel, Ç. The functional amyloid curli protects *Escherichia coli* against complement-mediated bactericidal activity. *Biomolecules* **2018**, *8*, 5. [CrossRef]
38. Kaushik, M.; Kumar, S.; Kapoor, R.K.; Virdi, J.S.; Gulati, P. Integrons in Enterobacteriaceae: Diversity, distribution and epidemiology. *Int. J. Antimicrob. Agents* **2018**, *51*, 167–176. [CrossRef]
39. Johnson, J.R.; Johnston, B.; Kuskowski, M.A.; Nougayrede, J.P.; Oswald, E. Molecular epidemiology and phylogenetic distribution of the *Escherichia coli pks* genomic island. *J. Clin. Microbiol.* **2008**, *46*, 3906–3911. [CrossRef]
40. Murase, K.; Martin, P.; Porcheron, G.; Houle, S.; Helloin, E.; Pénary, M.; Nougayrède, J.P.; Dozois, C.M.; Hayashi, T.; Oswald, E. *HlyF* produced by extraintestinal pathogenic *Escherichia coli* is a virulence factor that regulates outer membrane vesicle biogenesis. *J. Infect. Dis.* **2016**, *213*, 856–865. [CrossRef]
41. Gophna, U.; Parket, A.; Hacker, J.; Ron, E.Z. A novel *ColV* plasmid encoding type IV pili. *Microbiol.* **2003**, *149*, 177–184. [CrossRef]
42. Starčič Erjavec, M.; Petkovšek, Ž.; Kuznetsova, M.V.; Maslennikova, I.L.; Žgur-Bertok, D. Strain ŽP—The first bacterial conjugation-based "kill"-" anti-kill" antimicrobial system. *Plasmid* **2015**, *82*, 28–34. [CrossRef]
43. Maslennikova, I.L.; Kuznetsova, M.V.; Toplak, N.; Nekrasova, I.V.; Žgur Bertok, D.; Starčič Erjavec, M. Estimation of the bacteriocin ColE7 conjugation-based "kill" - "anti-kill" antimicrobial system by real-time PCR, fluorescence staining and bioluminescence assays. *Lett. Appl. Microbiol.* **2018**, *67*, 47–53. [CrossRef]
44. Wilson, L.A.; Sharp, P.M. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: Evolution and implications for ERIC-PCR. *Mol. Biol. Evol.* **2006**, *23*, 1156–1168. [CrossRef]

45. Quartieri, A.; Simone, M.; Gozzoli, C.; Popovic, M.; D'Auria, G.; Amaretti, A.; Raimondi, S.; Rossi, M. Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **2016**, *38*, 130–137. [CrossRef]
46. Simone, M.; Gozzoli, C.; Quartieri, A.; Mazzola, G.; Di Gioia, D.; Amaretti, A.; Raimondi, S.; Rossi, M. The probiotic *Bifidobacterium breve* B632 inhibited the growth of Enterobacteriaceae within colicky infant microbiota cultures. *Biomed. Res. Int.* **2014**. [CrossRef]
47. Standard Operating Procedure for PulseNet PFGE of Escherichia coli O157:H7, Escherichia coli non-O157 (STEC), Salmonella serotypes, Shigella sonnei and Shigella flexneri. Available online: http://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf (accessed on 19 July 2019).
48. Johnson, J.R.; Porter, S.; Johnston, B.; Kuskowski, M.A.; Spurbeck, R.R.; Mobley, H.L.; Williamson, D.A. Host characteristics and bacterial traits predict experimental virulence for *Escherichia coli* bloodstream isolates from patients with urosepsis. *Open Forum Infect. Dis.* **2015**, 2. [CrossRef]
49. Sarshar, M.; Scribano, D.; Marazzato, M.; Ambrosi, C.; Aprea, M.R.; Aleandri, M.; Pronio, A.; Longhi, C.; Nicoletti, M.; Zagaglia, C.; et al. Genetic diversity, phylogroup distribution and virulence gene profile of *pks* positive *Escherichia coli* colonizing human intestinal polyps. *Microb. Pathog.* **2017**, *112*, 274–278. [CrossRef]
50. Sicard, J.F.; Vogeleer, P.; Le Bihan, G.; Rodriguez Olivera, Y.; Beaudry, F.; Jacques, M.; Harel, J. N-Acetyl-glucosamine influences the biofilm formation of *Escherichia coli*. *Gut Pathog.* **2018**, *22*, 10–26. [CrossRef]
51. Vogeleer, P.; Tremblay, Y.D.N.; Jubelin, G.; Jacques, M.; Harel, J. Biofilm-forming abilities of shiga toxin-producing *Escherichia coli* isolates associated with human infections. *Appl. Environ. Microbiol.* **2015**, *28*, 1448–1458. [CrossRef]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of 12 *Leuconostoc carnosum* Strains Isolated from Cooked Ham Packaged in a Modified Atmosphere and from Fresh Sausages

Francesco Candeliere,[a] Stefano Raimondi,[a] Gloria Spampinato,[a] Moon Yue Feng Tay,[c,d] Alberto Amaretti,[a,b] Joergen Schlundt,[c,d] Maddalena Rossi[a,b]

[a]Department of Life Sciences, University of Modena and Reggio Emilia, Modena, Italy
[b]BIOGEST—SITEIA, University of Modena and Reggio Emilia, Modena, Italy
[c]Nanyang Technological University Food Technology Centre (NAFTEC), Singapore
[d]School of Chemical and Biomedical Engineering, Nanyang Technological University, Singapore

ABSTRACT *Leuconostoc carnosum* is a lactic acid bacterium that preferentially colonizes meat. In this work, we present the draft genome sequences of 12 *Leuconostoc carnosum* strains isolated from modified-atmosphere-packaged cooked ham and fresh sausages. Three strains harbor bacteriocin genes.

Leuconostoc carnosum is a heterofermentative, catalase-negative lactic acid bacterium (LAB) comprising the major component of the microbiota in modified-atmosphere-packaged (MAP) cooked ham at the end of the shelf life, where it reaches a concentration of 10^7 to 10^8 CFU/g (1). After a few weeks of shelf life, a high dose of alive *L. carnosum* may be ingested when consuming MAP cooked ham (1). *L. carnosum* easily colonizes several other meat-based food matrices, as suggested by the species name (2–6). It has been associated with ham spoilage (7, 8), but some strains have been proposed as bioprotective starters (4, 9) due to their production of bacteriocins against *Listeria monocytogenes* (5, 6). When we started this project, a sole completed genome sequence was available in GenBank (*L. carnosum* JB16, accession number SAMN02603179), obtained from a strain isolated from kimchi (10). Twelve *L. carnosum* genomes have been sequenced to characterize this species, belonging to 9 strains isolated from MAP cooked ham and 3 from fresh sausages, according to Raimondi et al. (1, 2) (Table 1).

For whole-genome sequencing, each strain was cultivated in brain heart infusion broth (Becton, Dickinson, USA) at 30°C for 48 h under microaerophilic conditions. Biomass was collected by centrifugation for 5 min at 5,000 × *g*. The genomic DNA was extracted with a QIAamp DNA minikit (Qiagen GmbH, Germany). The concentration was determined with a Qubit fluorometer (Invitrogen, USA). The DNA samples were submitted to the Singapore Centre for Environmental Life Sciences Engineering (SCELSE), where they were tagged with Illumina TruSeq high-throughput (HT) DNA dual barcodes for library preparation and sequenced on an Illumina HiSeq 2500 instrument (USA). For each sample, 250-bp paired-end reads were obtained. The raw and postprocessed reads were checked for quality with FastQC v0.11.7 (11). Cutadapt v1.16 was used to trim Illumina adapters, and reads with a quality score lower than 20 (–overlap = 15, –minimum-length = 30, –quality-cutoff = 20) were removed (12). The trimmed reads were assembled with SPAdes v3.12 (–careful –cov-cutoff auto -k auto) (13), and the quality of the assemblies was evaluated using QUAST v5.0.2 (14). Trimming, quality checking, and assembly were performed on the Galaxy platform (https://usegalaxy.eu/) (15). The completeness of the assemblies was determined with CheckM v1.0.8 (16). The taxonomy was confirmed with KmerFinder on the CGE server (https://cge.cbs.dtu.dk/services/KmerFinder/) (17). The assembled genomes were ordered with Mauve v2.4.0

Citation Candeliere F, Raimondi S, Spampinato G, Tay MYF, Amaretti A, Schlundt J, Rossi M. 2020. Draft genome sequences of 12 *Leuconostoc carnosum* strains isolated from cooked ham packaged in a modified atmosphere and from fresh sausages. Microbiol Resour Announc 9:e01247-19. https://doi.org/10.1128/MRA.01247-19.

Editor Catherine Putonti, Loyola University Chicago

Copyright © 2020 Candeliere et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Maddalena Rossi, maddalena.rossi@unimore.it.

Received 3 October 2019
Accepted 9 December 2019
Published 9 January 2020

Downloaded from http://mra.asm.org/ on January 21, 2020 at MOMED CLINIC

TABLE 1 Statistics of assembled genomes

GenBank accession no.	SRA accession no.	No. of raw reads	Strain	Source	Country of origin	Genome size (bp)	No. of contigs[a]	N_{50} (bp)	Coverage (×)[b]	G+C content (%)[a]	No. of CDS[c]	No. of rRNAs[d]	No. of tRNAs[d]	Completeness (%)[e]
VBXG01000000	SRR10389907	2,103,282	WC0318	Cooked ham	Italy	1,745,630	15	1,142,374	638	37.2	1,739	5	51	99.476
VBXF01000000	SRR10389906	1,718,930	WC0319	Cooked ham	Italy	1,700,071	11	1,137,580	521	37.1	1,679	5	49	99.476
VBXE01000000	SRR10389899	1,848,530	WC0320	Cooked ham	Poland	1,812,114	18	1,109,738	561	37.1	1,853	5	51	99.476
VBXD01000000	SRR10389898	1,753,950	WC0321	Cooked ham	France	1,804,293	40	1,106,035	532	37.1	1,854	5	52	99.476
VBXC01000000	SRR10389897	1,696,897	WC0322	Cooked ham	France	1,853,239	16	306,690	515	37	1,898	5	50	99.476
VBXB01000000	SRR10389896	1,847,670	WC0323	Cooked ham	Germany	1,773,698	23	256,026	560	37.2	1,802	5	52	99.476
VBXA01000000	SRR10389895	1,515,136	WC0324	Cooked ham	Germany	1,765,760	13	1,137,276	460	37.1	1,768	5	49	99.476
VBWY01000000	SRR10389893	1,654,963	WC0325	Cooked ham	Italy	1,830,248	27	229,291	502	37.2	1,881	5	52	99.476
VBWW01000000	SRR10389905	1,893,372	WC0326	Sausage	Italy	1,770,048	22	412,621	574	37.2	1,767	5	49	99.476
VBWV01000000	SRR10389904	1,872,273	WC0327	Sausage	Italy	1,769,683	20	412,621	568	37.2	1,768	5	49	99.476
VBWS01000000	SRR10389901	1,774,316	WC0328	Sausage	Italy	1,815,949	36	239,739	538	37.1	1,826	5	51	98.953
VBWR01000000	SRR10389900	1,498,696	WC0329	Cooked ham	France	1,650,966	14	256,123	455	37.2	1,644	5	52	99.476

[a] Determined using QUAST v5.0.2.
[b] Coverage has been calculated as (2 × 250 × number of reads)/reference genome length.
[c] Determined using the RAST server.
[d] Determined using Prokka v1.13, with default parameters.
[e] Determined using CheckM v1.0.8.

Downloaded from http://mra.asm.org/ on January 21, 2020 at MOMED CLINIC

(18), using the sequence of *L. carnosum* JB16 as a reference. The draft genomes were annotated on the RAST server (19) and with Prokka v1.13 (20). BAGEL4 (21) was used to identify bacteriocin genes. The sequence data were deposited in GenBank under the BioProject accession number PRJNA542256. Table 1 shows the GenBank accession number, number of raw reads, assembly statistics, and features of the genomes. The average G+C content (± the standard deviation) was 37.14% ± 0.07%. An average of 1,790 coding sequences (CDS) per genome were annotated, with genome sizes ranging between 1,650,966 and 1,853,239 bp and an average completeness of 99.43%. According to *in silico* analysis, 3 out of 12 strains were identified as potential bacteriocin producers.

Data availability. The sequence data were deposited in GenBank under BioProject accession number PRJNA542256. Table 1 lists, for each sample, the GenBank accession number and the number of raw reads.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

We acknowledge the Singapore Centre for Environmental Life Sciences Engineering (SCELSE) for sequencing the samples. We thank Kelyn L. G. Seow and Yalei Xu for the support in wet lab work.

REFERENCES

1. Raimondi S, Luciani R, Sirangelo TM, Amaretti A, Leonardi A, Ulrici A, Foca G, D'Auria G, Moya A, Zuliani V, Seibert TM, Søltoft-Jensen J, Rossi M. 2019. Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: microbiota of sliced cooked ham in MAP. Int J Food Microbiol 289:200–208. https://doi.org/10.1016/j.ijfoodmicro.2018.09.017.
2. Raimondi S, Nappi MR, Sirangelo TM, Leonardi A, Amaretti A, Ulrici A, Magnani R, Montanari C, Tabanelli G, Gardini F, Rossi M. 2018. Bacterial community of industrial raw sausage packaged in modified atmosphere throughout the shelf life. Int J Food Microbiol 280:78–86. https://doi.org/10.1016/j.ijfoodmicro.2018.04.041.
3. Geeraerts W, Pothakos V, De Vuyst L, Leroy F. 2018. Variability within the dominant microbiota of sliced cooked poultry products at expiration date in the Belgian retail. Food Microbiol 73:209–215. https://doi.org/10.1016/j.fm.2018.01.019.
4. Budde BB, Hornbaek T, Jacobsen T, Barkholt V, Koch AG. 2003. *Leuconostoc carnosum* 4010 has the potential for use as a protective culture for vacuum-packed meats: culture isolation, bacteriocin identification, and meat application experiments. Int J Food Microbiol 83:171–184. https://doi.org/10.1016/S0168-1605(02)00364-1.
5. van Laack RLJM, Schillinger U, Holzapfel WH. 1992. Characterization and partial purification of a bacteriocin produced by *Leuconostoc carnosum* LA44A. Int J Food Microbiol 16:183–195. https://doi.org/10.1016/0168-1605(92)90079-i.
6. Parente E, Moles M, Ricciardi A. 1996. Leucocin F10, a bacteriocin from *Leuconostoc carnosum*. Int J Food Microbiol 33:231–243. https://doi.org/10.1016/0168-1605(96)01159-2.
7. Björkroth KJ, Vandamme P, Korkeala HJ. 1998. Identification and characterization of *Leuconostoc carnosum*, associated with production and spoilage of vacuum-packaged, sliced, cooked ham. Appl Environ Microbiol 64:3313–3319.
8. Samelis J, Björkroth J, Kakouri A, Rementzis J. 2006. *Leuconostoc carnosum* associated with spoilage of refrigerated whole cooked hams in Greece. J Food Prot 69:2268–2273. https://doi.org/10.4315/0362-028x-69.9.2268.
9. Jacobsen T, Budde BB, Koch AG. 2003. Application of *Leuconostoc carnosum* for biopreservation of cooked meat products. J Appl Microbiol 95:242–249. https://doi.org/10.1046/j.1365-2672.2003.01970.x.
10. Jung JY, Lee SH, Jeon CO. 2012. Complete genome sequence of *Leuconostoc carnosum* strain JB16, isolated from kimchi. J Bacteriol 194:6672–6673. https://doi.org/10.1128/JB.01805-12.
11. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
12. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
13. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
15. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Grüning BA, Guerler A, Hillman-Jackson J, Hiltemann S, Jalili V, Rasche H, Soranzo N, Goecks J, Taylor J, Nekrutenko A, Blankenberg D. 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res 46:W537–W544. https://doi.org/10.1093/nar/gky379.
16. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
17. Hasman H, Saputra D, Sicheritz-Ponten T, Lund O, Svendsen CA, Frimodt-Møller N, Aarestrup FM. 2014. Rapid whole-genome sequencing for detection and characterization of microorganisms directly from clinical samples. J Clin Microbiol 52:139–146. https://doi.org/10.1128/JCM.02452-13.
18. Darling AC, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.
19. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
20. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
21. van Heel AJ, de Jong A, Song C, Viel JH, Kok J, Kuipers OP. 2018. BAGEL4: a user-friendly Web server to thoroughly mine RiPPs and bacteriocins. Nucleic Acids Res 46:W278–W281. https://doi.org/10.1093/nar/gky383.

Downloaded from http://mra.asm.org/ on January 21, 2020 at MOMED CLINIC

International Journal of
Molecular Sciences

MDPI

Article

Antibiotic Resistance, Virulence Factors, Phenotyping, and Genotyping of Non-*Escherichia coli* Enterobacterales from the Gut Microbiota of Healthy Subjects

Alberto Amaretti [1,2,†], Lucia Righini [1,†], Francesco Candeliere [1], Eliana Musmeci [1], Francesca Bonvicini [3], Giovanna Angela Gentilomi [3,4], Maddalena Rossi [1,2] and Stefano Raimondi [1,*]

1 Department of Life Sciences, University of Modena and Reggio Emilia, via Campi 103, 41125 Modena, Italy; alberto.amaretti@unimore.it (A.A.); lucia.righini@unimore.it (L.R.); francesco.candeliere@unimore.it (F.C.); eliana.musmeci@unimore.it (E.M.); maddalena.rossi@unimore.it (M.R.)
2 Biogest-Siteia, University of Modena and Reggio Emilia, Modena, Viale Amendola 2, 42122 Reggio Emilia, Italy
3 Department of Pharmacy and Biotechnology, Alma Mater Studiorum-University of Bologna, Via Massarenti 9, 40138 Bologna, Italy; francesca.bonvicini4@unibo.it (F.B.); giovanna.gentilomi@unibo.it (G.A.G.)
4 Unit of Microbiology, Alma Mater Studiorum-University of Bologna, S. Orsola-Malpighi Hospital, Via Massarenti 9, 40138 Bologna, Italy
* Correspondence: stefano.raimondi@unimore.it; Tel.: +39-059-205-8595
† These authors contributed equally to the work.

Received: 2 March 2020; Accepted: 5 March 2020; Published: 7 March 2020

Abstract: Non-*Escherichia coli* Enterobacterales (NECE) can colonize the human gut and may present virulence determinants and phenotypes that represent severe heath concerns. Most information is available for virulent NECE strains, isolated from patients with an ongoing infection, while the commensal NECE population of healthy subjects is understudied. In this study, 32 NECE strains were isolated from the feces of 20 healthy adults. 16S rRNA gene sequencing and mass spectrometry attributed the isolates to *Klebsiella pneumoniae*, *Klebsiella oxytoca*, *Enterobacter cloacae*, *Enterobacter aerogenes*, *Enterobacter kobei*, *Citrobacter freundii*, *Citrobacter amalonaticus*, *Cronobacter* sp., and *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens*. Multiplex PCR revealed that *K. pneumoniae* harbored virulence genes for adhesins (*mrkD*, *ycfM*, and *kpn*) and enterobactin (*entB*) and, in one case, also for yersiniabactin (*ybtS*, *irp1*, *irp2*, and *fyuA*). Virulence genes were less numerous in the other NECE species. Biofilm formation was spread across all the species, while curli and cellulose were mainly produced by *Citrobacter* and *Enterobacter*. Among the most common antibiotics, amoxicillin-clavulanic acid was the sole against which resistance was observed, only *Klebsiella* strains being susceptible. The NECE inhabiting the intestine of healthy subjects have traits that may pose a health threat, taking into account the possibility of horizontal gene transfer.

Keywords: Enterobacterales; *Klebsiella*; *Enterobacter*; *Citrobacter*; virulence; antibiotic resistance; biofilm

1. Introduction

Important enteric pathogens belong to Enterobacterales, a bacterial order within the phylum Proteobacteria. This order encompasses permanent colonizers of the human gut that, in healthy conditions, constitute minor bacterial components of the microbiota. Opportunistic Enterobacterales can persist as gut commensals without inducing any infections, as long as the microbiota is balanced and

the complex and dense bacterial community prevents their overgrowth. A bloom of Enterobacterales may occur as a result of disturbance of the microbiota, yielding pathogen-mediated infections and triggering inflammatory host responses.

Escherichia coli is the most studied among the Enterobacterales with regards to the traits that differentiate commensalism and pathogenicity. It normally colonizes the intestine but comprises both harmless commensals and different pathogenic variants that may instigate infections in the gut or in other tissues [1,2]. Virulent strains of *E. coli* isolated from infected patients attracted most research interest [3,4], but also fecal isolates from healthy subjects and environmental strains are the target of increasing attention, aiming to determine the pathogenic potential of a wider biodiversity reservoir [5,6].

Many non-*E. coli* Enterobacterales (hereinafter referred to as NECE) that can colonize the gut (e.g., *Klebsiella*, *Enterobacter*, *Citrobacter*, and *Serratia*) also present traits that can confer them virulence and pathogenicity or phenotypes that may result in severe heath concern, such as multidrug resistance [7–10]. The greatest efforts have been carried out to describe virulent strains, generally isolated from patients with an ongoing infection, while the pathogenic potential of NECE inhabiting the gut of healthy subjects has not been thoroughly investigated with genetic and phenotypic analysis, except for some genera [9]. The research herein presented aims to fill this gap, providing a genotypic and phenotypic description of the NECE population isolated from the feces of 20 healthy adults, and to complement a previous study that described the *E. coli* population of the same cohort of subjects [5]. A set of 32 NECE strains was isolated, taxonomically classified, subjected to PFGE genotyping, and described in terms of the genotypic determinants and phenotypic traits that may confer on them potential pathogenicity or invasiveness.

Information on the genes associated to virulence is detailed especially for *Klebsiella* spp., where a number of genes associated to harmful traits were identified, such as those encoding adhesins, siderophores (e.g., enterobactin, aerobactin, yersiniabactin), protectines, or invasins (responsible for mucoid phenotype and invasiveness), and involved in allantoin metabolism [7,11]. For other NECE genera, such as *Enterobacter*, *Cronobacter*, and *Citrobacter*, the knowledge of the genetic determinants associated to virulence and invasiveness is less comprehensive and, with few exceptions (e.g., *Citrobacter koseri*), mainly acquired from better characterized pathogens [12–15].

In addition to fimbrial and afimbrial adhesins, the production of surface cellulose structures and curli favors the adhesion of Enterobacterales and can exert a significant role in enteric biofilm-related infections [16,17]. Although not directly involved in pathogenic mechanisms, the acquisition of multiple antibiotic resistances favors the success of opportunistic Enterobacterales pathogens in invasion, survival, and spread, severely complicating the containment and treatment of infections [9,18]. Therefore, the occurrence of drug resistant bacteria within a commensal population and the possibility to exchange genetic material by horizontal gene transfer may represent a major health concern.

In the present study, multiplex PCR assays were utilized to screen the NECE, isolated from the feces of healthy subjects, for the presence of 17 main virulence genes associated to those of *Klebsiella* and *E. coli* [19–21]. From the point of view of the phenotype, the isolates were characterized for the ability to form biofilm and to yield curli and surface cellulose, were screened for the susceptibility to the most common antibiotics and for the ability to act as recipients in conjugation experiments, and biochemical tests were performed to compare the metabolic profile.

2. Results

2.1. Counting and Isolation of NECE

The selective differential medium HiCrome Coliform Agar (HCCA) was utilized to count and isolate *E. coli* [5] and NECE from fecal samples of 20 healthy adults. Total counts in HCCA ranged from 4.6×10^5 to 2.2×10^8 cfu/g (Figure 1). Blue colonies attributed to *E. coli* overcounted the pink ones attributed to NECE in all the samples except V11 (Figure 1; Supplementary Figure S1). NECE ranged between $< 10^4$ and 1.3×10^8 cfu/g and, except in V11, accounted for a minority of total Enterobacterales

(NECE + *E. coli*), the 75th percentile being the 5.7% (Figure 1). In some cases, NECE were not recovered, being outnumbered by *E. coli*. Spearman's rank correlation analysis excluded any significant correlation between NECE and *E. coli* counts.

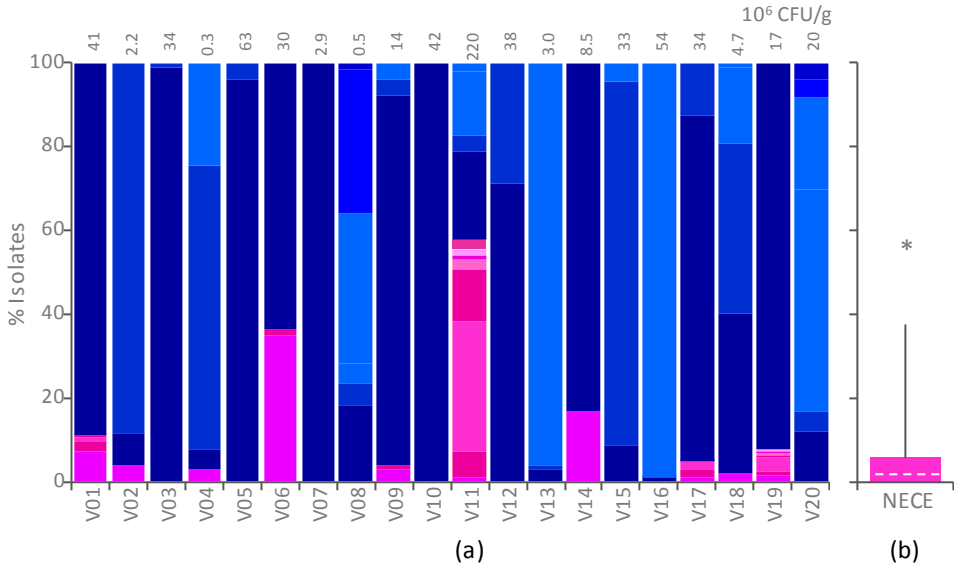


Figure 1. Counts of *Escherichia coli* and non-*Escherichia coli* Enterobacterales (NECE), enumerated onto HiCrome Coliform Agar (HCCA) plates. (**a**) Percentage of colonies attributed to *E. coli* (blue shades) and NECE (pink shades) in the feces of 20 subjects. For each subject, different shades indicate different biotypes according to enterobacterial repetitive intergenic consensus-PCR (ERIC-PCR) and random amplification of polymorphic DNA-PCR (RAPD-PCR) fingerprinting. The total count of Enterobacterales (*E. coli* + NECE) is reported in the top margin. (**b**) Distribution of the percentage of NECE colonies. The median (dashed line), the 25th and 75th percentiles (colored box), the 10th and 90th percentiles (whiskers), and outliers (*) are indicated.

2.2. Taxonomic Attribution and PFGE Genotyping

The isolates putatively attributed to NECE were clustered in 32 different biotypes utilizing ERIC-PCR (enterobacterial repetitive intergenic consensus-PCR) and RAPD-PCR (random amplification of polymorphic DNA-PCR) fingerprinting. A representative isolate of each biotype was assigned a taxonomic designation utilizing 16S rRNA gene sequencing and MALDI-TOF MS (Supplementary Table S1). The genus *Klebsiella* was the most represented (14/32), with the species *K. pneumoniae* (10 strains) and *Klebsiella oxytoca* (4) found in seven and three fecal samples, respectively. The genus *Enterobacter* was represented by eight strains belonging to *Enterobacter cloacae* (6), *Enterobacter aerogenes*, and *Enterobacter kobei* (1 strain each). Other strains belonged to *Citrobacter* (4 to *Citrobacter freundii* and 1 to *Citrobacter amalonaticus*), *Cronobacter* sp. (2), and to *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens* (1 strain each).

PFGE highlighted a wide diversity of the NECE isolates, which did not cluster according to the taxonomic attribution (Figure 2).

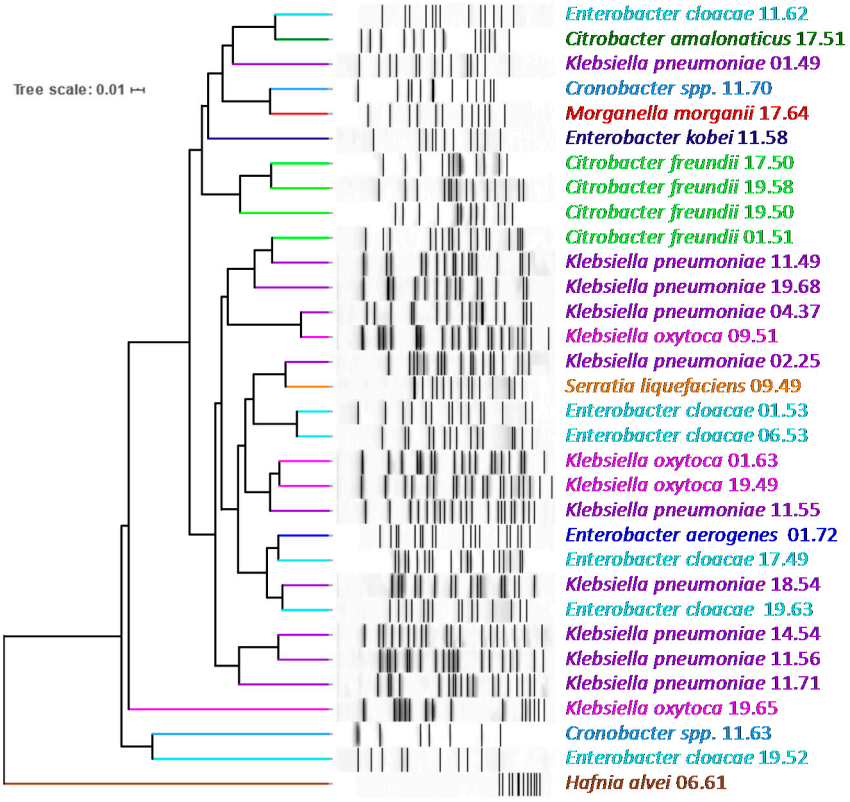


Figure 2. *Xba*I-PFGE pattern of NECE strains: unweighted pair group method with arithmetic means (UPGMA) dendrogram derived from Dice's coefficients, calculated based on the band profile. Strains are colored based on their MALDI-TOF MS taxonomic attribution.

2.3. Virulence Genotyping

PCR was used to investigate 17 virulence genes encoding adhesins (*fimH1*, *mrkD*, *kpn*, and *ycfM*), siderophores (enterobactin, *entB*; aerobactin, *iutA*; yersiniabactin, *irp-1*, *irp2*, *ybtS*, *fyuA*; catechols receptor, *iroN*; and other, *kfu*), protectines or invasins (*K2*, *magA*, *rmpA*, and *traT*), and involved in allantoin metabolism (*allS*).

Most strains of *K. pneumoniae* harbored the *mrkD*, *ycfM*, and *kpn* encoding adhesins and *entB* encoding enterobactin (Figure 3). Only *K. pneumoniae* 11.55 was positive to the main virulence genes involved in the synthesis of *Yersinia* siderophore, including *ybtS* (encoding for the synthase), *irp1* and *irp2* (for regulatory proteins), and *fyuA* (for the siderophore receptor). *irp2* was also detected in most of the other strains of *K. pneumoniae* although they lacked the counterpart *ybtS*. All the *K. pneumoniae* strains were negative to the genes *K2*, *magA*, and *rmpA* associated with hypermucoid phenotype and invasivity, except for *K. pneumoniae* 01.49 that was positive to *K2*. Similarly, other virulence genes, such as *allS*, *kfu*, and *iutA* occurred only once among the tested strains.

The strains of *K. oxytoca* harbored *entB* (three out of four isolates) but were negative to most of the other virulence genes. A sole strain harbored *ytbS*. Most of *Cronobacter* and *Enterobacter* isolates were characterized by the presence of the gene *irp2* but never harbored *ybtS* or other *Yersinia* siderophore genes. A few strains were positive to *mrkD* or *entB*.

The strains ascribed to *Citrobacter*, *H. alvei*, and *M. morganii* were negative to those virulence genes whose presence could not be excluded by primer-blast search. The strain of *S. liquefaciens* was positive to *ytbS*.

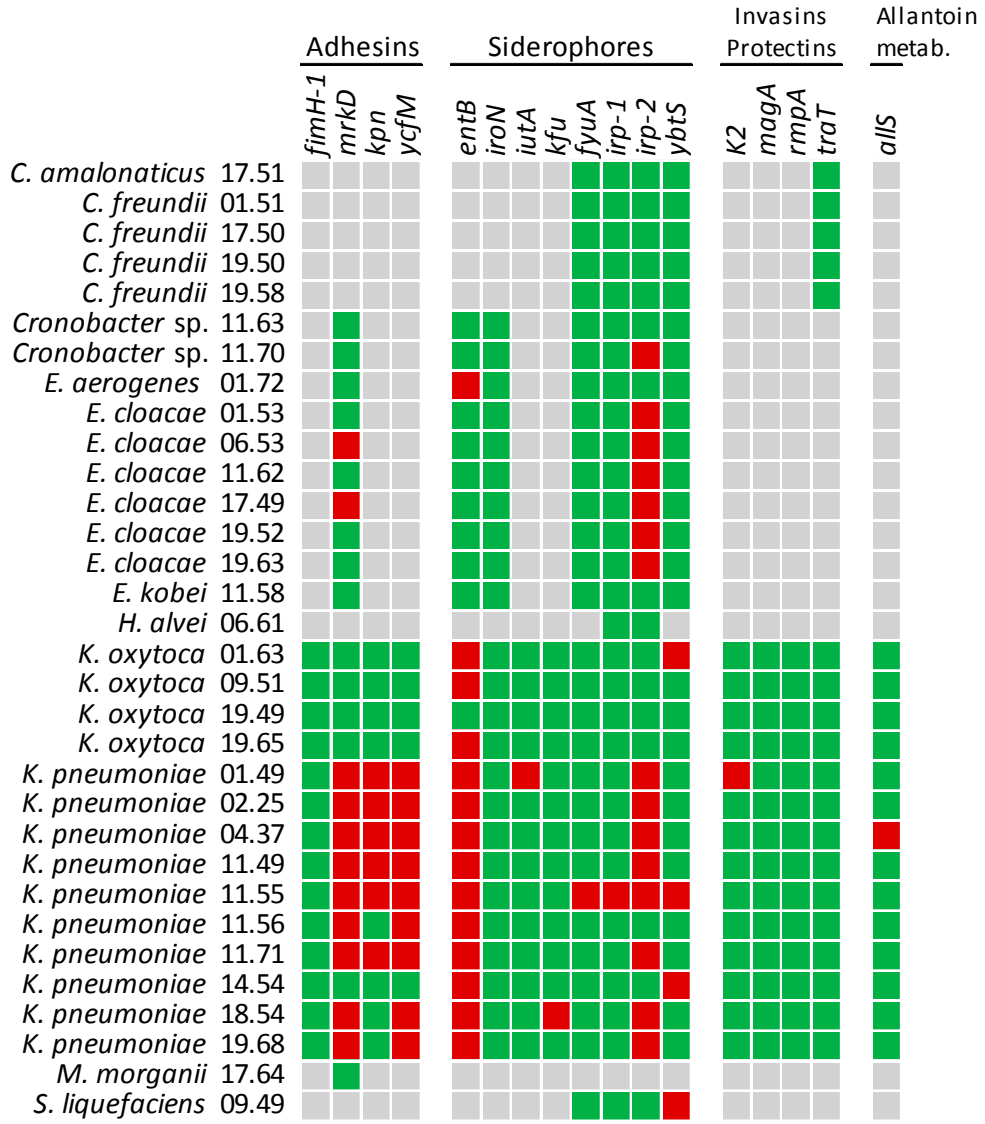


Figure 3. PCR assay of the NECE isolates for the presence of virulence genes. Colors: red, positive amplification; green, negative amplification; grey, PCR analysis not performed since the gene was putatively absent based on a primer-blast search.

2.4. Biofilm Formation and Production of Curli and Cellulose

NECE strains were tested for biofilm formation in minimal and rich media (M9 and LBWS, respectively; Supplementary Figure S2). The vast majority of the strains (26 out of 32), belonging to all the species except *E. aerogenes*, *H. alvei*, and *M. morganii*, formed biofilm in M9 (Figure 4; Supplementary Figure S2). Biofilm formation was less frequent in LBWS, being observed only in 10 strains of *K. oxytoca*, *K. pneumoniae,* and *S. liquefaciens*. The extent of biofilm production was always less abundant in the rich medium compared to M9 ($p < 0.05$).

Extracellular cellulose was detected in most of *Citrobacter* and *Enterobacter* strains, in five strains of *K. pneumoniae* and two out of four of *K. oxytoca*, in *H. alvei* and in *S. liquefaciens*. Curli were produced by nearly all *Citrobacter*, *Cronobacter*, and *Enterobacter* strains and by one isolate belonging to *K. oxytoca*. The isolates of *H. alvei* and *M. morganii* were also positive to curli. The strains belonging to *Citrobacter*, *E. cloacae*, *H. alvei,* and *K. oxytoca* 19.49 produced both cellulose and curli.

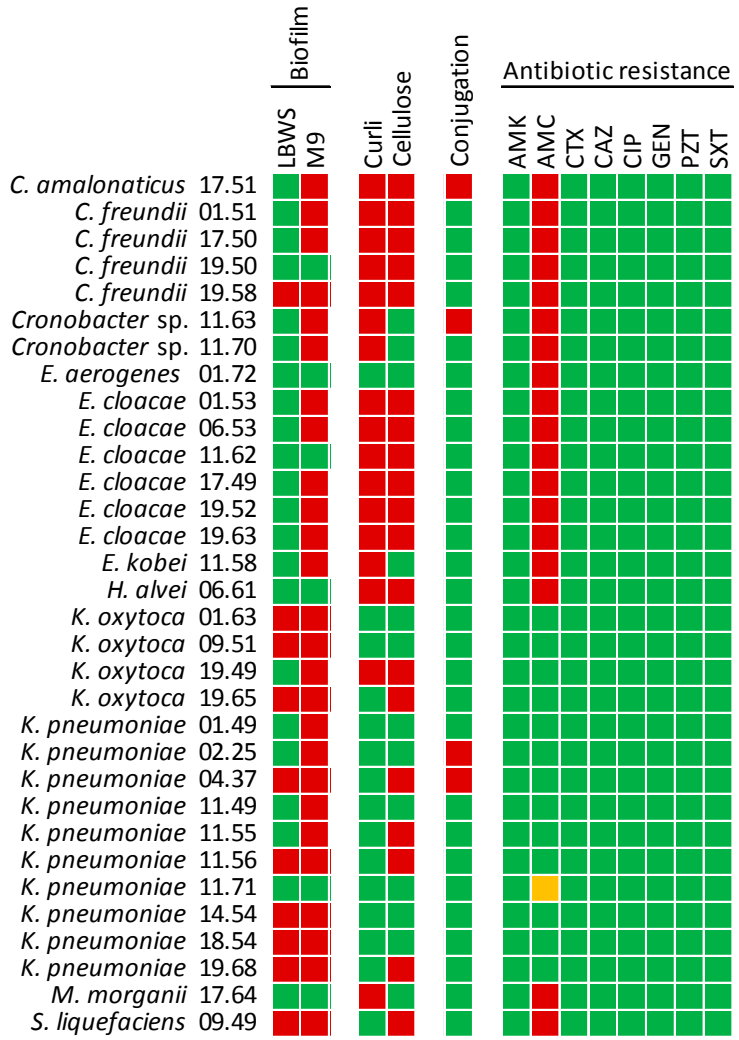


Figure 4. Phenotypic characterization of NECE isolates: biofilm formation in LBWS and M9 media, curli and cellulose production, conjugation, and antibiotic resistance. Colors: red, positive; green, negative. For antibiotics: red, resistant; green, susceptible, yellow, intermediate. Antibiotics: amikacin (AMK), amoxicillin–clavulanic acid (AMC), cefotaxime (CTX), ceftazidime (CAZ), ciprofloxacin (CIP), gentamicin (GEN), piperacillin-tazobactam (PZT), and trimethoprim-sulfamethoxazole (SXT).

2.5. Conjugation

The strains were challenged as conjugation recipients for receiving pOX38: Cm plasmid from *E. coli* N4i. Only two strains of *K. pneumoniae*, and single strains of *Cronobacter* and *Citrobacter amalonaticus* succeeded in plasmid acquisition (Figure 4).

2.6. Antibiotic Resistance

Phenotypic susceptibility to amikacin, amoxicillin–clavulanic acid, cefotaxime, ceftazidime, ciprofloxacin, gentamicin, piperacillin-tazobactam, and trimethoprim-sulfamethoxazole was assayed. Amoxicillin-clavulanic acid was the sole antibiotic against which few isolates presented some resistance, with all the strains of *Enterobacter*, *Citrobacter*, *Cronobacter*, *H. alvei*, *M. morganii*, and *S. liquefaciens* being resistant. All the 14 biotypes of *Klebsiella* spp. were sensitive to the whole set of tested antibiotics, with the exception of *K. pneumoniae* 11.71 that was partially resistant to amoxicillin–clavulanic acid, presenting a minimum inhibitory concentration (MIC) intermediate between resistance and susceptibility thresholds.

2.7. Biochemical Characterization

The fermentation of substrates and some distinctive enzymatic reactions and metabolic routes were assayed utilizing the API 20 E system (Figure 5). Generally, the NECE strains were positive to β-galactosidase. Most strains were capable of utilizing citrate, glucose, mannose, inositol, sorbitol, rhamnose, sucrose, melibiose, amygdalin, and arabinose. The main exceptions were *M. morganii* that could utilize only glucose, *H. alvei* that fermented a restricted number of sugars, and some *Citrobacter*, *Cronobacter*, and *Enterobacter* strains that exhibited specific substrate preferences. The majority of the isolates produced either lysine decarboxylase (*Klebsiella*) or ornithine decarboxylase (*Cronobacter* and *Enterobacter*). Urease was characteristic of *Klebsiella*, while arginine dihydrolase was found in most *Enterobacter*, *Citrobacter*, and *Cronobacter*. Acetoin was produced by *Cronobacter*, *Enterobacter*, *Hafnia*, and *Klebsiella*. Indole was produced by *K. oxytoca* and by few other strains, H_2S by two strains of *Citrobacter freundii*. Only *S. liquefaciens* was positive to gelatinase. All the strains except *M. morganii* and *S. liquefaciens* exhibited denitrifying activity, in most cases yielding nitrite. Nitrate reduction to N_2 was observed in *K. pneumoniae* and few other strains.

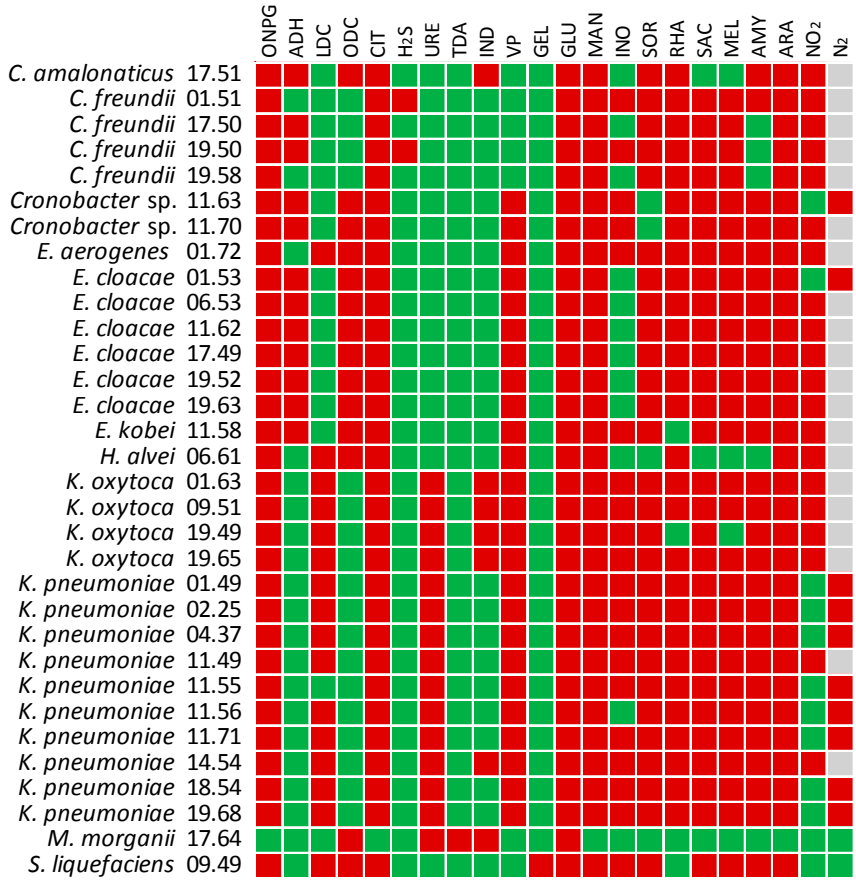


Figure 5. Biochemical reaction profiles of NECE isolates in the API 20 E assay: β-galactosidase (ONPG), arginine dihydrolase (ADH), lysine decarboxylase (LDC), ornithine decarboxylase (ODC), citrate utilization (CIT), production of hydrogen sulfide (H_2S), urease (URE), tryptophan deaminase (TDA), indole (Kovac's test, IND), acetoin (Voges-Proskauer test, VP), gelatinase (GEL), fermentation of glucose (GLU), mannitol (MAN), inositol (INO), sorbitol (SOR), rhamnose (RHA), sucrose (SAC), melibiose (MEL), amygdalin (AMY), arabinose (ARA), and reduction of nitrates to nitrites (N_2O) or nitrogen (N_2, tested only in case of negative N_2O). Colors: red, positive; green, negative.

3. Discussion

Thirty-two NECE strains were isolated from the feces of 20 healthy adults that did not present any dysbiosis, and thus as members of a relatively balanced gut microbiota. The load of Enterobacterales was in the order of millions or tens of millions per gram of feces, with a sole exception where they reached the magnitude of 10^8. NECE represented a small population of Enterobacterales, with *E. coli* being on average 20 times more abundant, and a minor component of the whole microbiota, being less than 0.1%. The genus *Klebsiella*, which is ubiquitous in nature, colonizing humans, animals, and plants, and frequently detected in waters, sewages, and soils, was the most represented, encompassing 14 of the 32 strains. *K. pneumoniae*, the most frequently isolated species (10 strains), was present in 7 out of 20 fecal samples, in accordance to the 35% of healthy adults from which it was isolated in a pioneering study [22].

K. pneumoniae encompasses opportunistic pathogens that can cause human infections in lungs, urinary tract, and bloodstreams, mostly to hospitalized and/or immunocompromised patients [23,24]. Virulence of *K. pneumoniae* is associated to the presence of capsule and pili, to the production of lipopolysaccharides and siderophores, to allantoin utilization, and to iron uptake systems, efflux pumps, and type VI secretion systems [7]. Surface molecules, such as capsular polysaccharides and lipopolysaccharides, are some of the major virulence factors that *Klebsiella* use to protect itself from the host innate immune apparatus. Furthermore, the capability to compete for iron has a pivotal role to the establishment of the infection. The genes involved in iron assimilation are generally clustered in pathogenicity islands, large chromosomal regions that were likely acquired by horizontal transfer. Virulence genetic determinants can be located both in the core or in the accessory genome [7], the former also including metabolic genes required for some species to cause disease.

Among the 10 *K. pneumoniae* isolates, all but one strain harbored *entB*, encoding a common catecholate siderophore located in the core genome, and *mrkD*, involved in the synthesis of type 3 pili that promote adherence to the surfaces [19]. The genes encoding other adhesins, such as *ycfM* and *kpn* were detected in nine and five strains, respectively. Only two *K. pneumoniae* strains were positive to the yersiniabactin gene *ybtS*, a common virulence factor associated with human infections [25,26], and one of the two also harbored *irp1*. These genes are involved in the synthesis of the siderophore yersiniabactin by virulent *Yersinia* strains, which harbor them within the high-pathogenicity island (HPI). HPI is widely distributed among members of the order Enterobacterales, including *E. coli*, *K. pneumoniae*, *Citrobacter* spp., *Salmonella enterica*, *Serratia liquefaciens*, and *Enterobacter* spp. [27,28]. In addition, *irp2*, another marker of HPI, was detected in the majority of the strains of *K. pneumonia* and *E. cloacae*, albeit they lacked the counterpart *ybtS*.

Although Enterobacterales are normally present as a low fraction of commensal bacteria in the healthy gut, their numbers can increase in the inflamed gut, and take advantage over other commensals [29]. Siderophores are major contributors of exploitative competition, since iron is an essential nutrient present in very low amounts in the gut and may play a role in virulence. Overall, *K. pneumoniae* 11.55 was the strain equipped with the broadest set of virulence genes involved in iron metabolism, including those encoding enterobactin (*entB*), yersiniabactin (*irp1*, *irp2*, and *ybtS)*, and, the sole strain among the 32 isolates, also the *Yersinia* siderophore receptor (*fyuA*).

In agreement with their commensal behavior, none of the *K. pneumoniae* strains were positive for the two genes associated with invasive infections, i.e., the mucoviscosity-associated gene *magA* and the regulator of mucoid phenotype *rmpA* [30,31], that are associated with a hypervirulent and hypermucoid phenotype. In our isolates, the presence of other genetic determinants of virulence was sporadic or quite rare. In the gut of the healthy host, the Enterobacterales resides in the outer loose mucus layer, separated from the epithelial mono-cell barrier by an inner dense mucin coat [32]. During the infection, they penetrate the mucus layer, interact with the epithelial cells, and may breach the mucosal barrier. In case these enterobacteria strains reach other tissues, biofilm formation may act as a fitness factor concurring to pathogenesis [33]. Adhesion factors and extracellular matrix components are involved in formation of biofilms [34]. All but one strain produced biofilm in minimal medium M9, whereas this

phenotype was less frequent when growth occurred in the rich medium LBWS. The higher extent of biofilm formation in M9 is consistent with the fact that this mineral medium is more challenging for these bacteria. Extracellular cellulose structures, determined in five strains, were consistent with the capability to form biofilm on both media. Curli were found in a sole strain that presented also cellulose structures (19.58 CA) and produced biofilm on both the substrates.

Based on genetic and functional features, some *K. pneumoniae* isolates present a higher potential to cause infections, albeit they are present at low charge in the microbiota of healthy hosts. The link between colonization and infection by *K. pneumoniae* in hospitalized patients has been demonstrated, with robust evidence that their own microbiota is the main source of the infective strains [35,36]. Thus, potentially more virulent *K. pneumoniae* strains may take advantage of critical conditions, becoming responsible for nosocomial infections [37].

The isolates of *K. oxytoca*, another prominent pathogen that may be involved in diseases and life-threatening infections [38,39], generally encoded the siderophore gene *entB*, but were negative to most of the other virulence genes. A sole strain harbored the gene encoding the yersiniabactin siderophore. Biofilm production was a general feature of the *K. oxytoca* strains, regardless of the presence of curli or cellulose structures.

Similar considerations are valid for the other NECE strains herein described, belonging to genera that may have clinical relevance, such as *Enterobacter*, *Citrobacter*, and *Cronobacter*. The isolates were generally capable of forming biofilm and producing curli and cellulose and were negative for most virulence genes. However, unlike *Klebsiella* that shares many virulence genes with *E. coli*, the genetic determinants of virulence of these genera have not been fully disclosed.

In this study emerged that NECE isolates from feces of healthy subjects are still quite susceptible to most of the antibiotics. This is important, since any treatment of opportunistic outbreaks of NECE (e.g., in case nosocomial infections) requires antibiotics, but resistance developments would seriously curb the therapeutic options [40]. All the isolates of *Klebsiella* were sensitive to the whole set of tested antibiotics. Amoxicillin-clavulanic acid was the sole antibiotic against which was detected some resistance: the strains ascribed to the genus *Enterobacter* and the isolates belonging to *Citrobacter*, *Cronobacter*, *Hafnia alvei*, *Morganella morganii*, and *Serratia liquefaciens* were all resistant to this combination of antibiotics. Interestingly, Enterobacterales presented the highest increase in terms of relative abundance in a short-term amoxicillin-clavulanic acid treatment in healthy adults [41]. Some genera belonging to this family, such as *Enterobacter* and *Citrobacter*, are recognized as intrinsically resistant [42], and may take advantage to this antibiotic treatment. In general, the profile of resistance was independent by the subject of the fecal sample, but two clusters encompassing sensitive or resistant strains were sharply differentiated by taxonomy.

PFGE genotyping was carried out to evaluate the genetic similarity among the bacterial isolates of this study and highlighted a wide dispersion of the strains, regardless their taxonomic attribution and phylogenetic relationships. The spreading of the strains regardless of species or genera may be attributed to the presence of plasmids, the horizontal acquisition of additional genes from diverse species of Enterobacterales, and to the exchange of mobile elements that rapidly integrate and promote DNA shuffling, in agreement with the capability of some strains to accept DNA by conjugation from *E. coli* as a donor. However, the biochemical profiling mostly clustered the strains in species (data not shown), confirming that energy production and conservation, and lipid, amino acid, and nucleotide metabolism are part of the conserved reactome, despite the genome plasticity.

The Enterobacterales encountered in this study are generally recognized as opportunistic pathogens, with some potential capability to cause disease, on the basis of predicted virulence factors. Except for *K. pneumoniae* hypervirulent strains, NECE virulence seems more associated to the host features than to the strain traits. According to the similar results obtained for the *E. coli* isolated from the same cohort of healthy subjects, the absence of antibiotic resistance for most of the tested antibiotics does not pose a serious challenge for infection control. This highlights the stratification

of antibiotic resistance distribution among healthy and hospitalized/diseased subjects, with NECE associated risk increasing with both illness and antibiotic therapy.

4. Materials and Methods

4.1. Isolation and Enumeration of Enterobacterales

Fresh fecal samples were collected from 20 healthy adult subjects who gave written informed consent regarding their participation in the study in accordance with the protocol approved by the local research ethics committee (reference number: 974/2019/SPER-UNIMO-ENTEROPOP; Comitato etico dell'Area Vasta Emilia Nord, Italy). The subjects—10 males and 10 females aged 35 to 45, following a western omnivore diet, and who had not been treated with prebiotics and/or probiotics for 1 month and antibiotics for 3 months—were enrolled among the employees of the University of Modena and Reggio Emilia and their relatives and were not in relationship with the researchers.

Feces were homogenized (10% *w/v*) in isotonic Buffered Peptone Water (Sigma, Steinheim, Germany), then serial dilutions were spread onto plates of HiCrome Coliform Agar (HCCA, Sigma) and incubated at 37 °C. The medium differentiates *E. coli* (blue colonies) from NECE (salmon to red). For each subject, up to 48 colonies of putative NECE were picked and clustered into biotypes with ERIC-PCR [43] and RAPD-PCR [44] fingerprint presenting Pearson's similarity > 75%.

4.2. PFGE Genotyping

PFGE was performed according to PulseNet protocol (http://www.cdc.gov/pulsenet/PDF/ecoli-shigella-salmonella-pfge-protocol-508c.pdf). The genomic DNA was digested with 50 U of *Xba*I at 37 °C for 3 h. Fragments were resolved in a CHEF-DRIII apparatus (Bio-Rad, Hercules, CA, USA) using counter-clamped homogeneous electric field electrophoresis (24 h at 6.0 V/cm; initial switch time, 2.2 s; final switch time, 54.2 s). The run was digitally captured and analyzed with GelCompare II 6.0 software (Applied Maths NV, Sint-Martens-Latem, Belgium). Dice coefficient was computed to evaluate similarity between band profiles (position tolerance, 1%; optimization, 1%) and to derive an UPGMA dendrogram (unweighted pair group method with arithmetic means). Strains were ascribed to the same pulsotype if PFGE profile possessed >85% similarity.

4.3. Taxonomic Attribution

A strain for each biotype was taxonomically characterized by partial sequencing of the 16S rRNA gene sequencing, utilizing primers targeting the V1-V3 portion. Primer sequences and PCR conditions were set up according to Raimondi et al. [45]. The sequences, obtained from a service provider (Eurofins Genomics, Ebersberg, Germany), were compared to those in SILVA SSU database utilizing SINA Aligner v1.2.11 (https://www.arb-silva.de/aligner/).

In addition, the MALDI-TOF MS-based biotyping was carried using the MALDI Biotyper 3.1 system (Bruker Daltonics, Bremen, Germany). Sample preparation for MALDI-TOF MS was performed as previously described with minor modifications [46]. Briefly, colonies of fresh overnight culture were placed on a MALDI sample slide (Bruker Daltonics) and dried at room temperature. The sample was then overlaid with 1 µL of matrix solution (α-cyano-4-hydroxycinnamic acid in 50% acetonitrile and 2.5% trifluoroacetic acid) and dried at room temperature. A MALDI-TOF MS measurement was performed with a Bruker MicroFlex MALDI-TOF MS (Bruker Daltonics) using FlexControl software and a *Escherichia coli* DH5α protein extract (Bruker Daltonics) was placed on the calibration spot of the sample slide for external calibration. Spectra collected in the positive-ion mode within a mass range of 2000–20,000 Da were analyzed using a Bruker Biotyper (Bruker Daltonics) automation control and the Bruker Biotyper 3.1 software and library (a database with 5627 entries). High confidence species identification was accepted, if the log(score) was ≥2.00, low confidence species identification log(score) values (≥1.70 and <2.00) were accepted if the three best matches showed the same species name. Any results with log(score)<1.70 were considered as an unacceptable identification.

4.4. Profiling of Virulence Genes

All the isolates were screened by multiplex-PCR for the genes associated to 17 virulence factors: *allS*, *entB*, *fimH*, *fyuA*, *iroN*, *irp1*, *irp2*, *iutA*, *K2*, *kfu*, *kpn*, *magA*, *mrkD*, *rmpA*, *traT*, *ybtS*, and *ycfM*. Primer sequences and amplification conditions were set up according to El Fertas-Aissani et al. [19], Compain et al. [20], and Johnson et al. [21]. In order to assess the possibility to obtain the amplicon in the different NECEs species, a search with the primer-blast tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was performed for all the set of primers developed for *K. pneumoniae* and *E. coli*. The result of PCR amplification was reported only for the species in which the annealing and the possibility to yield an amplicon were predicted.

4.5. Biofilm and Phenotype Assays

Biofilm formation was quantified with crystal violet in the microtiter assay described in [47]. Two growth media were compared: Luria Bertani without salt (LBWS) and M9 (BD Difco, Sparks, MD, USA) containing 4 g/L glucose and 0.25 g/L yeast extract. Strains exhibiting a specific biofilm formation (i.e., the ratio between crystal violet absorbance at 570 nm and culture turbidity at 600 nm) > 1. The data herein reported are the means of three independent experiments, each carried out in triplicate.

The strains were screened for curli and cellulose production utilizing LBWS agar plates supplemented with the appropriate stain [48]. Red colonies in Congo red-supplemented plates were considered positive to curli. Colonies in calcofluor white-supplemented plates that emitted fluorescence due to UV exposure (315–400 nm) were considered positive to cellulose.

4.6. Solid Mating Conjugation Experiments

The strains were screened as recipients in conjugation experiments with the donor *E. coli* N4i pOX38:Cm (N4i: EcN immE7 Gmr; pOX38:Cm: Tra+ RepFIA+ Cmr) [5]. The donor and the recipient strains were cultured and put in contact onto LB plates under the conditions described in [5]. HCCA and HCCA with 20 μg/mL chloramphenicol were utilized to differentiate recipient, transconjugant, and donor colonies [5].

4.7. Antibiotic Susceptibility

The strains were tested for antimicrobial susceptibility with a Vitek2 semi-automated system (bioMerieux, Marcy-l'Étoile, France). Minimum inhibitory concentrations (MICs) were interpreted according to EUCAST (European Committee on Antimicrobial Susceptibility Testing—www.eucast.org) and susceptibility (S) or resistance (R) were defined based on the following thresholds (mg/L): amikacin, S < 8 and R > 16; amoxicillin/clavulanic acid, S < 8 and R > 8; cefotaxime, S < 8 and R > 2; ceftazidime, S < 8 and R > 8; ciprofloxacin, S < 0.5 and R > 1; gentamicin, S < 2 and R > 4; piperacillin + tazobactam, S < 8 and R > 16; trimethoprim/sulfamethoxazole, S < 40 and R > 80.

4.8. Biochemical Characterization

The strains were tested for distinctive enzymatic reactions and metabolic routes utilizing API 20 E test system (bioMerieux, France), according to the manufacturer's instructions.

Supplementary Materials: Supplementary materials can be found at http://www.mdpi.com/1422-0067/21/5/1847/s1.

Author Contributions: Conceptualization, G.A.G., F.B., M.R, and S.R.; data curation, S.R., F.C., and A.A.; investigation, A.A., F.C., L.R., E.M., S.R., and F.B.; methodology, S.R., L.R., F.B., M.R., and G.A.G.; project administration, M.R.; resources, M.R. and G.A.G.; supervision, M.R.; visualization, A.A.; writing—original draft, F.B., S.R., A.A. and M.R.; writing—review & editing, G.A.G. All authors have read and agreed to the published version of the manuscript.

Funding: Please add: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

NECE	non-*E. coli* Enterobacterales
HCCA	HiCrome Coliform Agar
MALDI-TOF MS	Matrix-assisted laser desorption ionization–time of flight mass spectrometry
ERIC	Enterobacterial repetitive intergenic consensus
RAPD	Random Amplification of Polymorphic DNA
UPGMA	Unweighted pair group method with arithmetic means

References

1. Leimbach, A.; Hacker, J.; Dobrindt, U. *E. coli* as an all-rounder: The thin line between commensalism and pathogenicity. *Curr. Top. Microbiol. Immunol.* **2013**, *358*, 3–32. [CrossRef] [PubMed]
2. Manges, A.R.; Geum, H.M.; Guo, A.; Edens, T.J.; Fibke, C.D.; Pitout, J.D.D. Global extraintestinal pathogenic *Escherichia coli* (ExPEC) lineages. *Clin. Microbiol. Rev.* **2019**, *32*. [CrossRef] [PubMed]
3. Clermont, O.; Christenson, J.K.; Denamur, E.; Gordon, D.M. The Clermont *Escherichia coli* phylo-typing method revisited: Improvement of specificity and detection of new phylo-groups. *Environ. Microbiol. Rep.* **2013**, *5*, 58–65. [CrossRef] [PubMed]
4. Croxen, M.A.; Law, R.J.; Scholz, R.; Keeney, K.M.; Wlodarska, M.; Finlay, B.B. Recent advances in understanding enteric pathogenic *Escherichia coli*. *Clin. Microbiol. Rev.* **2013**, *26*, 822–880. [CrossRef]
5. Raimondi, S.; Righini, L.; Candeliere, F.; Musmeci, E.; Bonvicini, F.; Gentilomi, G.; Starčič Erjavec, M.; Amaretti, A.; Rossi, M. Antibiotic resistance, virulence factors, phenotyping, and genotyping of *E. coli* isolated from the feces of healthy subjects. *Microorganisms* **2019**, *7*, 251. [CrossRef]
6. Sarowska, J.; Futoma-Koloch, B.; Jama-Kmiecik, A.; Frej-Madrzak, M.; Ksiazczyk, M.; Bugla-Ploskonska, G.; Choroszy-Krol, I. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: Recent reports. *Gut Pathog.* **2019**, *11*, 10. [CrossRef]
7. Martin, R.M.; Bachman, M.A. Colonization, Infection, and the accessory genome of *Klebsiella pneumoniae*. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 4. [CrossRef]
8. Davin-Regli, A.; Pagès, J.M. *Enterobacter aerogenes* and *Enterobacter cloacae*; versatile bacterial pathogens confronting antibiotic treatment. *Front. Microbiol.* **2015**, *6*, 392. [CrossRef]
9. Liu, L.; Chen, D.; Liu, L.; Lan, R.; Hao, S.; Jin, W.; Sun, H.; Wang, Y.; Liang, Y.; Xu, J. Genetic diversity, multidrug resistance, and virulence of *Citrobacter freundii* from diarrheal patients and healthy individuals. *Front. Cell. Infect. Microbiol.* **2018**, *8*, 233. [CrossRef]
10. Gajdács, M.; Urbán, E. Resistance Trends and epidemiology of citrobacter-enterobacter-serratia in urinary tract infections of inpatients and outpatients (RECESUTI): A 10-year survey. *Medicina (Kaunas)* **2019**, *55*, 285. [CrossRef]
11. Lee, C.R.; Lee, J.H.; Park, K.S.; Jeon, J.H.; Kim, Y.B.; Cha, C.J.; Jeong, B.C.; Lee, S.H. Antimicrobial resistance of hypervirulent *Klebsiella pneumoniae*: Epidemiology, hypervirulence-associated determinants, and resistance mechanisms. *Front. Cell. Infect. Microbiol.* **2017**, *7*, 483. [CrossRef] [PubMed]
12. Azevedo, P.A.A.; Furlan, J.P.R.; Oliveira-Silva, M.; Nakamura-Silva, R.; Gomes, C.N.; Costa, K.R.C.; Stehling, E.G.; Pitondo-Silva, A. Detection of virulence and β-lactamase encoding genes in *Enterobacter aerogenes* and *Enterobacter cloacae* clinical isolates from Brazil. *Braz. J. Microbiol.* **2018**, *49* (Suppl. 1), 224–228. [CrossRef] [PubMed]
13. Mokracka, J.; Koczura, R.; Kaznowski, A. Yersiniabactin and other siderophores produced by clinical isolates of *Enterobacter* spp. and *Citrobacter* spp. *FEMS Immunol. Med. Microbiol.* **2004**, *40*, 51–55. [CrossRef]
14. Yuan, C.; Yin, Z.; Wang, J.; Qian, C.; Wei, Y.; Zhang, S.; Jiang, L.; Liu, B. Comparative genomic analysis of *Citrobacter* and key genes essential for the pathogenicity of *Citrobacter koseri*. *Front. Microbiol.* **2019**, *10*, 2774. [CrossRef]
15. Cruz, A.; Xicohtencatl-Cortes, J.; González-Pedrajo, B.; Bobadilla, M.; Eslava, C.; Rosas, I. Virulence traits in *Cronobacter* species isolated from different sources. *Can. J. Microbiol.* **2011**, *57*, 735–744. [CrossRef]
16. Zogaj, X.; Bokranz, W.; Nimtz, M.; Römling, U. Production of cellulose and curli fimbriae by members of the family Enterobacteriaceae isolated from the human gastrointestinal tract. *Infect. Immun.* **2003**, *71*, 4151–4158. [CrossRef]

17. Tursi, S.A.; Tükel, Ç. Curli-containing enteric biofilms inside and out: Matrix composition, immune recognition, and disease implications. *Microbiol. Mol. Biol. Rev.* **2018**, *82*. [CrossRef]
18. Partridge, S.R. Resistance mechanisms in Enterobacteriaceae. *Pathology* **2015**, *47*, 276–284. [CrossRef]
19. El Fertas-Aissani, R.; Messai, Y.; Alouache, S.; Bakour, R. Virulence profiles and antibiotic susceptibility patterns of *Klebsiella pneumoniae* strains isolated from different clinical specimens. *Pathol. Biol. (Paris)* **2013**, *61*, 209–216. [CrossRef]
20. Compain, F.; Babosan, A.; Brisse, S.; Genel, N.; Audo, J.; Ailloud, F.; Kassis-Chikhani, N.; Arlet, G.; Decré, D. Multiplex PCR for detection of seven virulence factors and K1/K2 capsular serotypes of *Klebsiella pneumoniae*. *J. Clin. Microbiol.* **2014**, *52*, 4377–4380. [CrossRef]
21. Johnson, J.R.; Porter, S.; Johnston, B.; Kuskowski, M.A.; Spurbeck, R.R.; Mobley, H.L.; Williamson, D.A. Host characteristics and bacterial traits predict experimental virulence for *Escherichia coli* bloodstream isolates from patients with urosepsis. *Open Forum Infect. Dis.* **2015**, *2*. [CrossRef] [PubMed]
22. Davis, T.J.; Matsen, J.M. Prevalence and characteristics of *Klebsiella* species: Relation to association with a hospital environment. *J. Infect. Dis.* **1974**, *130*, 402–405. [CrossRef] [PubMed]
23. Lau, H.Y.; Huffnagle, G.B.; Moore, T.A. Host and microbiota factors that control *Klebsiella pneumoniae* mucosal colonization in mice. *Microbes Infect.* **2008**, *10*, 1283–1290. [CrossRef] [PubMed]
24. Magill, S.S.; Edwards, J.R.; Bamberg, W.; Beldavs, Z.G.; Dumyati, G.; Kainer, M.A.; Lynfield, R.; Maloney, M.; McAllister-Hollod, L.; Nadle, J.; et al. Multistate point-prevalence survey of health care-associated infections. *N. Engl. J. Med.* **2014**, *370*, 1198–1208. [CrossRef]
25. Lan, Y.; Zhou, M.; Jian, Z.; Yan, Q.; Wang, S.; Liu, W. Prevalence of pks gene cluster and characteristics of *Klebsiella pneumoniae*-induced bloodstream infections. *J. Clin. Lab. Anal.* **2019**, *33*, e22838. [CrossRef]
26. Lawlor, M.S.; O'connor, C.; Miller, V.L. Yersiniabactin is a virulence factor for *Klebsiella pneumoniae* during pulmonary infection. *Infect. Immun.* **2007**, *75*, 1463–1472. [CrossRef]
27. Schubert, S.; Cuenca, S.; Fischer, D.; Heesemann, J. High-pathogenicity island of *Yersinia pestis* in Enterobacteriaceae isolated from blood cultures and urine samples: Prevalence and functional expression. *J. Infect. Dis.* **2000**, *182*, 1268–1271. [CrossRef]
28. Schubert, S.; Rakin, A.; Heesemann, J. The *Yersinia* high-pathogenicity island (HPI): Evolutionary and functional aspects. *Int. J. Med. Microbiol.* **2004**, *294*, 83–94. [CrossRef]
29. Raffatellu, M. Learning from bacterial competition in the host to develop antimicrobials. *Nat. Med.* **2018**, *24*, 1097–1103. [CrossRef]
30. Fang, C.T.; Chuang, Y.P.; Shun, C.T.; Chang, S.C.; Wang, J.T. A novel virulence gene in *Klebsiella pneumoniae* strains causing primary liver abscess and septic metastatic complications. *J. Exp. Med.* **2004**, *199*, 697–705. [CrossRef]
31. Yu, W.L.; Ko, W.C.; Cheng, K.C.; Lee, H.C.; Ke, D.S.; Lee, C.C.; Fung, C.P.; Chuang, Y.C. Association between *rmpA* and *magA* genes and clinical syndromes caused by *Klebsiella pneumoniae* in Taiwan. *Clin. Infect. Dis.* **2006**, *42*, 1351–1358. [CrossRef] [PubMed]
32. Johansson, M.E.; Sjovall, H.; Hansson, G.C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **2013**, *10*, 352–361. [CrossRef] [PubMed]
33. Flores-Mireles, A.L.; Walker, J.N.; Caparon, M.; Hultgren, S.J. Urinary tract infections: Epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **2015**, *13*, 269–284. [CrossRef] [PubMed]
34. Hobley, L.; Harkins, C.; MacPhee, C.E.; Stanley-Wall, N.R. Giving structure to the biofilm matrix: An overview of individual strategies and emerging common themes. *FEMS Microbiol. Rev.* **2015**, *39*, 649–669. [CrossRef] [PubMed]
35. Martin, R.M.; Cao, J.; Brisse, S.; Passet, V.; Wu, W.; Zhao, L.; Malani, P.N.; Rao, K.; Bachman, M.A. Molecular epidemiology of colonizing and infecting isolates of *Klebsiella pneumoniae*. *mSphere* **2016**, *1*. [CrossRef] [PubMed]
36. Gorrie, C.L.; Mirceta, M.; Wick, R.R.; Edwards, D.J.; Thomson, N.R.; Strugnell, R.A.; Pratt, N.F.; Garlick, J.S.; Watson, K.M.; Pilcher, D.V.; et al. Gastrointestinal carriage is a major reservoir of *Klebsiella pneumoniae* infection in intensive care patients. *Clin. Infect. Dis.* **2017**, *65*, 208–215. [CrossRef] [PubMed]
37. Gajdács, M.; Bátori, Z.; Ábrók, M.; Lázár, A.; Burián, K. Characterization of resistance in gram-negative urinary isolates using existing and novel indicators of clinical relevance: A 10-year data analysis. *Life (Basel)* **2020**, *10*, 16. [CrossRef]

38. Carrie, C.; Walewski, V.; Levy, C.; Alexandre, C.; Baleine, J.; Charreton, C.; Coche-Monier, B.; Caeymaex, L.; Lageix, F.; Lorrot, M.; et al. *Klebsiella pneumoniae* and *Klebsiella oxytoca* meningitis in infants. Epidemiological and clinical features. *Arch. Pediatr.* **2019**, *26*, 12–15. [CrossRef]
39. Pötgens, S.A.; Brossel, H.; Sboarina, M.; Catry, E.; Cani, P.D.; Neyrinck, A.M.; Delzenne, N.M.; Bindels, L.B. Klebsiella oxytoca expands in cancer cachexia and acts as a gut pathobiont contributing to intestinal dysfunction. *Sci. Rep.* **2018**, *8*, 12321. [CrossRef]
40. Gajdács, M.; Ábrók, M.; Lázár, A.; Burián, K. Comparative epidemiology and resistance trends of common urinary pathogens in a tertiary-care hospital: A 10-year surveillance study. *Medicina (Kaunas)* **2019**, *55*, 356. [CrossRef]
41. MacPherson, C.W.; Mathieu, O.; Tremblay, J.; Champagne, J.; Nantel, A.; Girard, S.A.; Tompkins, T.A. Gut bacterial microbiota and its resistome rapidly recover to basal state levels after short-term amoxicillin-clavulanic acid treatment in healthy adults. *Sci. Rep.* **2018**, *8*, 11192. [CrossRef] [PubMed]
42. Leclercq, R.; Cantón, R.; Brown, D.F.; Giske, C.G.; Heisig, P.; MacGowan, A.P.; Mouton, J.W.; Nordmann, P.; Rodloff, A.C.; Rossolini, G.M.; et al. EUCAST expert rules in antimicrobial susceptibility testing. *Clin. Microbiol. Infect.* **2013**, *19*, 141–160. [CrossRef] [PubMed]
43. Wilson, L.A.; Sharp, P.M. Enterobacterial repetitive intergenic consensus (ERIC) sequences in *Escherichia coli*: Evolution and implications for ERIC-PCR. *Mol. Biol. Evol.* **2006**, *23*, 1156–1168. [CrossRef] [PubMed]
44. Quartieri, A.; Simone, M.; Gozzoli, C.; Popovic, M.; D'Auria, G.; Amaretti, A.; Raimondi, S.; Rossi, M. Comparison of culture-dependent and independent approaches to characterize fecal bifidobacteria and lactobacilli. *Anaerobe* **2016**, *38*, 130–137. [CrossRef]
45. Raimondi, S.; Luciani, R.; Sirangelo, T.M.; Amaretti, A.; Leonardi, A.; Ulrici, A.; Foca, G.; D'Auria, G.; Moya, A.; Zuliani, V.; et al. Microbiota of sliced cooked ham packaged in modified atmosphere throughout the shelf life: Microbiota of sliced cooked ham in MAP. *Int. J. Food Microbiol.* **2019**, *16*, 200–208. [CrossRef]
46. Mencacci, A.; Monari, C.; Leli, C.; Merlini, L.; De Carolis, E.; Vella, A.; Cacioni, M.; Buzi, S.; Nardelli, E.; Bistoni, F.; et al. Typing of nosocomial outbreaks of Acinetobacter baumannii by use of matrix-assisted laser desorption ionization-time of flight mass spectrometry. *J. Clin. Microbiol.* **2013**, *51*, 603–606. [CrossRef]
47. Sicard, J.F.; Vogeleer, P.; Le Bihan, G.; Rodriguez Olivera, Y.; Beaudry, F.; Jacques, M.; Harel, J. N-Acetyl-glucosamine influences the biofilm formation of *Escherichia coli*. *Gut Pathog.* **2018**, *22*, 10–26. [CrossRef]
48. Vogeleer, P.; Tremblay, Y.D.N.; Jubelin, G.; Jacques, M.; Harel, J. Biofilm-forming abilities of shiga toxin-producing *Escherichia coli* isolates associated with human infections. *Appl. Environ. Microbiol.* **2015**, *28*, 1448–1458. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

microorganisms

MDPI

Article

Effect of Rearing Temperature on Growth and Microbiota Composition of *Hermetia illucens*

Stefano Raimondi [1], Gloria Spampinato [1], Laura Ioana Macavei [1], Linda Lugli [1], Francesco Candeliere [1], Maddalena Rossi [1,2], Lara Maistrello [1,2] and Alberto Amaretti [1,2,*]

[1] Department of Life Sciences, University of Modena and Reggio Emilia, 41125 Modena, Italy; stefano.raimondi@unimore.it (S.R.); gloria.spampinato@unimore.it (G.S.); lauraioana.macavei@unimore.it (L.I.M.); lugli.linda@gmail.com (L.L.); francesco.candeliere@unimore.it (F.C.); maddalena.rossi@unimore.it (M.R.); lara.maistrello@unimore.it (L.M.)

[2] BIOGEST–SITEIA, University of Modena and Reggio Emilia, 42124 Reggio Emilia, Italy

* Correspondence: alberto.amaretti@unimore.it; Tel.: +39-059-205-8588

Received: 1 June 2020; Accepted: 12 June 2020; Published: 15 June 2020

Abstract: The potential utilization of black soldier fly (*Hermetia illucens*) as food or feed is interesting due to the nutritive value and the sustainability of the rearing process. In the present study, larvae and prepupae of *H. illucens* were reared at 20, 27, and 33 °C, to determine whether temperature affects the whole insect microbiota, described using microbiological risk assessment techniques and 16S rRNA gene survey. The larvae efficiently grew across the tested temperatures. Higher temperatures promoted faster larval development and greater final biomass but also higher mortality. Viable Enterobacteriaceae, *Bacillus cereus*, *Campylobacter*, *Clostridium perfringens*, coagulase-positive staphylococci, *Listeriaceae*, and *Salmonella* were detected in prepupae. *Campylobacter* and *Listeriaceae* counts got higher with the increasing temperature. Based on 16S rRNA gene analysis, the microbiota of larvae was dominated by *Providencia* (>60%) and other *Proteobateria* (mainly *Klebsiella*) and evolved to a more complex composition in prepupae, with a bloom of *Actinobacteria*, *Bacteroidetes*, and *Bacilli*, while *Providencia* was still present as the main component. Prepupae largely shared the microbiota with the frass where it was reared, except for few lowly represented taxa. The rearing temperature was negatively associated with the amount of *Providencia*, and positively associated with a variety of other genera, such as *Alcaligenes*, *Pseudogracilibacillus*, *Bacillus*, *Proteus*, *Enterococcus*, *Pediococcus*, *Bordetella*, *Pseudomonas*, and *Kerstersia*. With respect to the microbiological risk assessment, attention should be paid to abundant genera, such as *Bacillus*, *Myroides*, *Proteus*, *Providencia*, and *Morganella*, which encompass species described as opportunistic pathogens, bearing drug resistances or causing severe morbidity.

Keywords: *Hermetia illucens*; microbiological risk assessment; black soldier fly; microbiota; 16S rRNA gene; metagenome

1. Introduction

Insects are promising protein sources for livestock and human diet and may represent a sustainable contribution for feeding the growing world population. Larvae and prepupae of *Hermetia illucens* (i.e., black soldier fly, BSF) are receiving increasing interest as potential food and feed source because of the high nutritional value and the low environmental impact of the rearing process [1]. In particular, insect meal of BSF was successfully utilized as an alternative protein source for poultry and aquaculture [2–5]. The utilization of mature larvae and prepupae in insect-based ingredients (e.g., powders, flours, protein bars, pasta, burgers, and nuggets) requires the development of cheap practices to produce and stabilize

the insect biomass, providing a safe starting bulk product [6,7]. The ability of *H. illucens* to grow on a variety of solid organic matrices can be exploited to transform and valorize organic streams such as byproducts of the agroindustry, livestock manures, or urban solid wastes, reducing environmental pollution and converting organic wastes into biomass rich in protein and fat [8–11].

The possibility to rear *H. illucens* on different waste streams in a biorefinery approach requires attention to safety issues, both chemical and microbiological. The European Food Safety Authority reported the lack of data regarding microbiology, virology, parasitology, and toxicology of insects reared, addressing the potential hazards of nonprocessed insects in comparison with other nonprocessed sources of protein of animal origin [12]. Hygiene issues of edible insects can originate from the substrate but also from the microbial community of the insect gut and may be affected by processing steps linking farming and consumption.

Recent works reported the microbiota composition of BSF larvae and prepupae [3,13–15]. However, little attention has been paid to microbial dynamics associated with extrinsic parameters, such as the rearing temperature. From a bioeconomy perspective, the biotransformation of wastes toward new valuable compounds should be a low-energy process that can be exploited in different districts with good tolerance to temperature changes, minimizing the requirement of heating or cooling.

From this standpoint, BSF larvae were reared at different temperatures, to determine the effect of this parameter on the growth rate and the biomass yield of prepupae. Since temperature may be a primary driver of microbial population composition, the microbiota of prepupae reared at 20, 27, and 33 °C was characterized. To avoid biases due to the intrinsic different microbial communities of feed, the larvae were fed a basal substrate [16–18]. A survey of the 16S rRNA gene was utilized to determine the microbiota composition in samples of substrate, initial larvae, grown prepupae, and their frass (hereinafter referred to as S, L, PP, and F samples). To assess the risks associated with prepupae utilization as food and feed supplement/ingredient, the metagenomic approach was accompanied by standard microbiology quality control techniques targeting the most relevant food contaminants and pathogens [12,19–21]. The impact of rearing temperature on microbiota is expected to offer knowledge and tools to prevent microbiological hazards, protect the health and the welfare of the consumers, and promote the development of safe trade in food and feed products.

2. Materials and Methods

2.1. Hermetia illucens Rearing

The BSF larvae used in this study were obtained from a colony kept at the Applied Entomology Laboratory, BIOGEST-SITEIA (Reggio Emilia, Italy), originally established with prepupae purchased from CIMI srl (Cuneo, Italy). BSF larvae were routinely reared in a standard vegetable substrate (S), composed of 25% zootechnical use cornflour (mill waste), 15% wheat bran, 10% alfa-alfa flour, and 50% water [22]. The experiments were set up using second to third instar larvae that were obtained from eggs hatched in S and incubated in climatic chambers under controlled conditions of 27 ± 0.5 °C, 65% humidity, and 16:8 light/dark cycles. To compare the rearing temperatures, 100 small larvae (L) were collected and reared in glass containers (20 × 12 × 8 cm) containing 400 g of S within climatic chambers at 20 ± 0.5, 27 ± 0.5, or 33 ± 0.5 °C. Two independent experiments, starting with different batches of L, were carried out, where each rearing temperature was tested in triplicate. Test checks were performed every three days, starting from day seven until the end of the development period, when a minimum of 90% of the initial larvae became prepupae (PP) and the experiment was considered concluded. The time point presenting the highest PP appearance in a single control was registered as the peak of PP. At the end of the experiment, insects were collected from the frass (F). PP and the remaining larvae (L_{final}) were counted and weighed to determine total biomass weight, the mean growth rate of each insect, and larvae mortality, according to the following:

$$Total\ biomass\ (g)\ =\ total\ PP\ weight + L_{final}\ weight \quad (1)$$

$$Mean\ growth\ rate\left(\frac{g}{d}\right) = \frac{\left(mean\ PP\ weight\ +\ mean\ L_{final}\ weight -\ mean\ L\ weight\right)}{development\ period\ (d)} \quad (2)$$

$$\text{Larvae mortality \%} = 100 \times \left(\text{No. L} - \text{No. } L_{final} - No.\ PP\right) \quad (3)$$

Throughout all the rearing phases, standard hygienic conditions were applied, without using surface disinfection of larvae, substrate sterilization or aseptic incubation. Samples of L, S, $PP_{20\,°C}$, $PP_{27\,°C}$, $PP_{33\,°C}$, $F_{20\,°C}$, $F_{27\,°C}$, and $F_{33\,°C}$ from each experiment were frozen and maintained at –80 °C until analyzed.

2.2. Culture Dependent Microbiological Analyses

Samples consisting of 5 g of S, PP, or F were suspended (10% *w/v*) in buffered peptone water (BD Difco, Franklin Lake, NJ, USA) and homogenized by Ultra Turrax (T25 Ika, Staufen, Germany). Proper dilutions were spread on selective media provided by BD Difco (Franklin Lake, NJ, USA). Total mesophilic aerobes were counted on Plate Count Agar (PCA) after incubation at 30 °C for 72 h. Spore-forming aerobic bacteria were grown on plates of PCA supplemented with 2 g/L of starch incubated at 37 °C for 48 h, with samples heated at 80 °C for 10 min to inactivate vegetative bacteria. Lactic acid bacteria were detected on Lactobacilli MRS agar plates after incubation at 30 °C for 72 h under microaerophilic conditions (GasPack EZ, BD Difco). Enterobacteriaceae were enumerated on Violet Red Bile Agar plates incubated at 37 °C for 24 h. Yeasts and molds were counted on Dichloran Rose Bengal Chloramphenicol agar after incubation at 25 °C for 5 d. Standard methodologies were utilized to detect and enumerate the following food pathogens: *Bacillus cereus*, *Campylobacter* spp.; *Clostridium perfringens*, *Listeria monocytogenes* and *Listeria* spp.; *Salmonella* spp.; and *Staphylococcus aureus*, and coagulase-positive staphylococci [23–29]. Triplicate samples from the two rearing batches were analyzed (n = 6). Microbial contamination was compared with recommendations of international authorities [12,19,20].

2.3. 16S rRNA Gene Profiling

For each rearing batch, triplicate samples of L, S, $PP_{20\,°C}$, $PP_{27\,°C}$, $PP_{33\,°C}$, and $F_{20\,°C}$, $F_{27\,°C}$, and $F_{33\,°C}$ were pooled in equal weights. Approximately 2 g of material were 10-fold diluted in PBS and were homogenized by Ultra Turrax. Total DNA was isolated from the suspensions using the DNeasy Mericon Food Kit (Qiagen, Hilden, Germany), following the manufacturer's standard protocol. The DNA was normalized to 5 ng/µL after quantification with a Qubit v. 3.0 fluorimeter (Thermo Fisher Scientific, Waltham, MA, USA). Partial 16S rRNA gene sequences were amplified using Probio_Uni/Probio_Rev primers, which targeted the V3 region of the 16S rRNA gene. Amplicons were sequenced using a MiSeq (Illumina, San Diego, CA, USA) platform according to Milani et al. [30]. The 16S rRNA gene sequences are available at NCBI repository with the BioProject ID: PRJNA573042.

Raw sequences were cleaned and filtered by size and quality using the software MOTHUR v. 1.25.0 [31]. Quality filtered sequences were processed with QIIME2 pipeline (v. 2019.1) for closed-reference picking of amplicon sequence variants (ASVs), taxonomy assignment, collapsing into operational taxonomic units (OTUs) [32,33]. Closed-reference picking and taxonomy assignation were carried out utilizing as reference SILVA SSU database release 132 (https://www.arb-silva.de/download/arb-files/) with the similarity threshold set at 0.97. The appropriate QIIME2 plugins were utilized to compute the alpha- (observed taxa, Chao1, Shannon, and Pielou's evenness) and beta diversity (Jaccard, Bray–Curtis, Canberra, Unweighted Normalized UniFrac, and Weighted Normalized UniFrac) and to compare them within and between groups of samples (i.e., the Kruskal–Wallis test for alpha diversity; ANOSIM and PERMANOVA for beta diversity). Beta diversity distance/dissimilarity matrices were utilized for the hierarchical clustering of samples in UPGMA trees and the Principal Coordinate Analysis (PCoA), using QIIME2.

Linear discriminant analysis Effect Size (LEfSe, http://huttenhower.sph.harvard.edu/galaxy) algorithm was applied to discover distinctive taxonomic features characterizing the groups of

samples [34]. In the analysis of the taxa characterizing the development stage, L and PP samples were entered as "classes". In the comparison between frass and insects, F and PP were entered as "classes". Spearman's rank correlation was applied to identify the taxa in PP that positively or negatively correlated with the three temperatures. In both LEfSe and Spearman's analysis, the alpha value for statistical significance was set at 0.05.

2.4. Statistics

Plate counts and growth parameters are presented as means ± SD (n = 6). ANOVA followed by Tukey's post-hoc test was carried out with the software SPSS Statistics (v. 21, IBM, Armonk, NY, USA). Differences were considered statistically significant for $p < 0.05$. Spearman's rank correlation was utilized to estimate the correlation with temperature.

3. Results

3.1. Growth Performance of Hermetia illucens at Different Temperatures

The growth parameters of *H. illucens* are presented in Table 1. The total development period was the longest at 20 °C (40 d), while it was approximately 24 d at 27 and 33 °C. Rearing at 20 °C resulted in both the lowest total biomass and mean growth rate (15.9 g and 4.1 mg/d). No significant differences were observed between 27 °C and 33 °C (approximately 21 g and 10 mg/d). The highest larval mortality (14.5%) was registered at 33 °C, whereas it was similar at 20 and 27 °C (3.8 and 5.5%, respectively). The temperature significantly affected the time point at which the peak of PP was observed. The higher the temperature, the earlier came the peak (30.0, 19.7, and 16.0 d at 20, 27, and 33 °C, respectively).

Table 1. Growth parameters of *H. illucens* at the three temperatures tested (definitions in materials and methods). Within each line, values marked with different letters indicate significant differences (ANOVA followed by Tukey's post-hoc test, $p < 0.05$).

Growth Parameter	20 °C	27 °C	33 °C
Development period (d)	40.0 ± 0 ^{a}	24 ± 1.1 ^{b}	23.7 ± 1.0 ^{b}
Total biomass (g)	15.9 ± 2.3 ^{b}	21.3 ± 3.8 ^{a}	21.0 ± 1.3 ^{a}
Mean growth rate (mg/d)	4.1 ± 0.6 ^{b}	9.4 ± 1.5 ^{a}	10.4 ± 0.7 ^{a}
Larvae mortality (%)	3.8 ± 2.0 ^{b}	5.5 ± 5.3 ^{b}	14.5 ± 7.4 ^{a}
Peak of PP (d)	30.0 ± 1.5 ^{a}	19.7 ± 2.6 ^{b}	16.0 ± 1.6 ^{c}

3.2. Culture Dependent Microbiological Analysis

The viable count of total mesophilic aerobes in PP positively correlated with temperature ($\varrho = 0.87$, $p < 0.05$), being the lowest at 20 °C (7.4 Log_{10} cfu/g) and approximately one magnitude higher at both 27 and 33 °C (Table 2). Aerobic spore-forming bacteria were less abundant than total aerobes, at all the temperatures ($p < 0.05$). Lactic bacteria and Enterobacteriaceae lay in the range of 6.2–6.9 and 3.7–5.2 Log_{10} cfu/g, respectively, without any significant difference associated with temperature. A low load of cultivable yeasts and molds was observed, between 0.7 and 1.2 Log_{10} cfu/g, without significant influence of temperature.

Bacillus cereus, *Campylobacter* spp.; *Clostridium perfringens*, coagulase-positive staphylococci, and Listeriaceae were detected in PP reared at all the tested temperatures. The amount of *Campylobacter* and Listeriaceae was positively correlated to temperature ($\varrho > 0.87$, $p < 0.05$). The former passed from 3.2 to 4.7 Log_{10} cfu/g and the latter from 4.8 to 5.8 Log_{10} cfu/g increasing the temperature from 20 to 33 °C. Coagulase-positive staphylococci, *B. cereus*, and *C. perfringens* (laying in the ranges of 3.7–4.2, 2.3–3.2, and 0.8–1.6 Log_{10} cfu/g, respectively) were not affected by the temperature. *Salmonella* spp. was found in 7 out of 18 samples. Among the positive samples, only one (PP grown at 20 °C) allowed the recovery of colonies by direct plating, at a concentration of 1.1 Log_{10} cfu/g.

The microbiological analysis of S revealed a mean viable charge of 4.5 Log_{10} CFU/g total mesophilic aerobes, 4.9 Log_{10} CFU/g aerobic spore-forming bacteria, 2.9 Log_{10} CFU/g lactic acid bacteria, and 3.4 Log_{10} CFU/g fungi, while Enterobacteriaceae, *B. cereus*, *Campylobacter* spp. *C. perfringens*, coagulase-positive staphylococci, and *Salmonella* spp. were absent.

Table 2. Cultivable microorganisms observed in substrate (S) and grown prepupae (PP) grown at 20, 27 and 33 °C. Counts are expressed as Log_{10} CFU/g. For *Salmonella* spp. the number of positive samples out of six is reported.

Viable Counts	S	PP			Contamination Limits [1]
		20 °C	27 °C	33 °C	
Total mesophilic aerobes	4.5 ± 0.2	7.4 ± 0.9 ^{b}	8.5 ± 0.4 ^{a}	8.7 ± 0.3 ^{a}	<5.7
Aerobic spore-forming bacteria	4.9 ± 0.8	6.8 ± 1.3 ^{a}	7.9 ± 0.5 ^{a}	7.9 ± 0.2 ^{a}	n.s.
Lactic acid bacteria	2.9 ± 0.7	6.6 ± 0.4 ^{a}	6.9 ± 0.7 ^{a}	6.2 ± 0.5 ^{a}	n.s.
Enterobacteriaceae	-	3.7 ± 0.3 ^{b}	5.2 ± 0.6 ^{a}	4.8 ± 1.3 ^{ab}	<3.0
Yeasts and molds	3.4 ± 0.3	0.7 ± 0.9 ^{a}	1.2 ± 1.0 ^{a}	0.9 ± 0.8 ^{a}	n.s.
Bacillus cereus	-	2.3 ± 0.9 ^{a}	3.2 ± 0.7 ^{a}	2.3 ± 0.7 ^{a}	<5.0
Campylobacter spp.	-	3.2 ± 0.3 ^{c}	4.2 ± 0.4 ^{b}	4.7 ± 0.2 ^{a}	absent in 25 g
Clostridium perfringens	-	0.8 ± 0.7 ^{a}	1.6 ± 1.3 ^{a}	1.0 ± 0.8 ^{a}	<5.0
Coagulase-positive staphylococci	-	3.9 ± 0.7 ^{a}	4.2 ± 0.7 ^{a}	3.7 ± 0.9 ^{a}	<5.0
Listeriaceae	2.6 ± 0.4	4.8 ± 0.4 ^{b}	5.5 ± 0.3 ^{a}	5.8 ± 0.4 ^{a}	<2.0
Salmonella spp.	0/6	2/6	1/6	4/6	absent in 25 g

Values are means ± SD (n = 6). - indicates < 2 Log_{10} CFU/g in the 6 replicates. For PP, within each line, values marked with different letters indicate significant differences (ANOVA followed by Tukey's post-hoc test, $p < 0.05$). [1] Reference values according to NVVWA and EFSA [12,19,20]. n.s., limit not specified.

3.3. Microbiota Composition by Metagenome Analysis

The 16S rRNA gene profiling of L, PP, S, and F yielded a total of 755,313 sequences, that were dereplicated into 3974 ASVs hitting a reference sequence in Silva database, and collapsed at the seventh level of taxonomic annotation (i.e., the species, if available) into 536 OTUs. Based on Bray–Curtis distance, the microbiota of L, PP, S, and F clustered in distinct groups (Figure 1) that had different centroids (PERMANOVA, $p = 0.001$) and intragroup permutational similarity significantly greater than the intergroup one (ANOSIM, $p = 0.001$). F represented an exception, the similarity within the group being comparable to that between F and PP ($p > 0.05$). Grouping was evident in the PCo1-PCo2 plot (describing the 49% of total variance), where F and PP were characterized by positive PCo1, L by positive Pco2, and S by negative PCo1. Both F and PP lay at progressively lower PCo2 with the increase of the rearing temperature. Main taxa positively contributing to PCo1 were *Myroides*, *Alcaligenes faecalis*, *Bacillus*, and *Morganella*, while Enterobacteriaceae weighted negatively. Positive PCo2 was strongly owed to *Providencia*, followed by *Klebsiella* and *Myroides*, and negative to *Bacillus*, *Pseudomonas*, and *A. faecalis*.

The richness of the microbiota (evaluated as total no. of OTUs and Chao1 index) was similar in all the groups of samples, regardless of the rearing temperature (Supplementary Figure S1). The evenness (estimated in Shannon and Pielou metrics) was the highest in S samples, followed by PP and F, grouped regardless of the rearing temperature, and the lowest in L (Supplementary Figure S1).

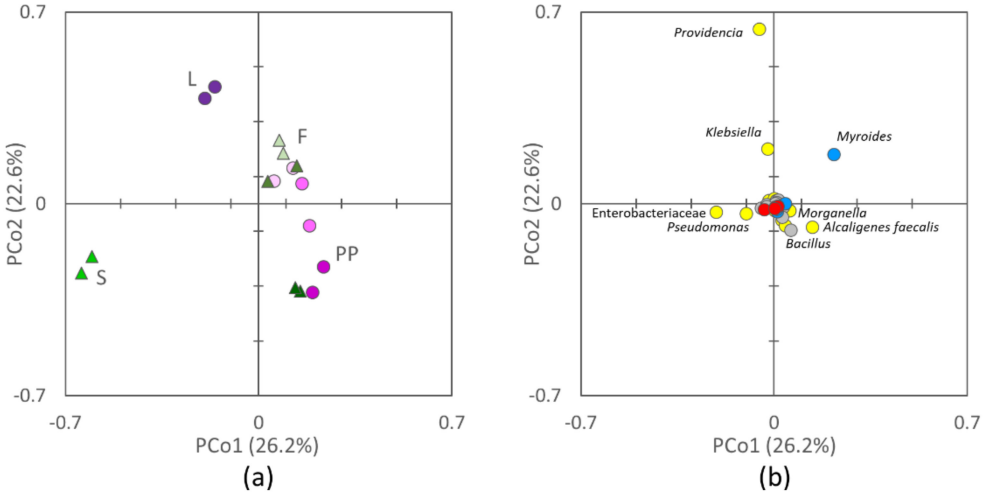


Figure 1. Two-dimensional (2d) Principal Coordinate Analysis (PCoA) visualization of microbiota beta diversity (Bray–Curtis) among initial larvae (L), S, PP, and frass (F): (**a**) scores of L (green circle), S (green triangles), PP (pink circles, with 20, 27, and 33 °C from the lightest to the darkest), and F (pink triangles, with 20, 27, and 33 °C from the lightest to the darkest), (**b**) contribution of the single bacterial taxa (Proteobacteria, yellow; Bacteroidetes, blue, Firmicutes, gray; Actinobacteria, red). Labels indicate taxa with the greatest contribution along PCo1 and PCo2.

A reduced dataset of 181 OTUs, occurring for more than 0.2% in at least one sample, still represented 99% of the reads. The distribution of the OTUs in the groups of samples is reported in the Venn diagram of Supplementary Figure S2. The bacterial composition of L, PP, S, and F is reported in Figure 2, showing the mean abundance of the main bacterial groups in the two experiments. The two experiments were not averaged for the analysis of the differential abundance of bacterial groups and the effect of temperature. The microbiota of L encompassed 98 of the 181 OTUs. It was dominated by Proteobacteria (87.4%), with remarkably high amounts of *Providencia* (64.1%) and *Klebsiella* (16.4%), followed by Firmicutes (9.3%), especially *Bacillus* (4.3%). The microbiota of PP included 156 OTUs, 95 of which shared with L. PP were poorer than L in Proteobacteria and richer in Bacteroidia, Actinobacteria, and Bacilli (Figures 2 and 3). *Providencia* remained abundant in PP (from 6.4 to 26.1%, depending on the temperature), although it significantly decreased compared to L, likewise *Klebsiella* and other minor Enterobacteriaceae. The genera *Morganella*, *Alcaligenes*, *Bordetella*, and *Kerstersia* behaved in contrast to most Proteobacteria and were significantly enriched in PP. *Morganella* and *Alcaligenes*, in particular, reached remarkable levels (8.9 and 16.9%, respectively). The significant increase of Bacteroidia in PP was mainly due to *Myroides*, a Flavobacteriaceae that reached 30%. *Bacilli* also included several biomarkers characterizing PP, such as Planococcaceae, Paenibacillaceae, *Pseudogracilibacillus*, *Oceanobacillus*, and unclassified members of the genus *Bacillus*. Planococcaceae, *Pseudogracilibacillus*, and unclassified *Bacillus* reached 8.5, 5.9, and 13.1%, respectively. Within Actinobacteria, the main biomarkers characterizing PP were Micrococcales belonging to *Brevibacterium*, which increased up to 3.3%, and several minor genera.

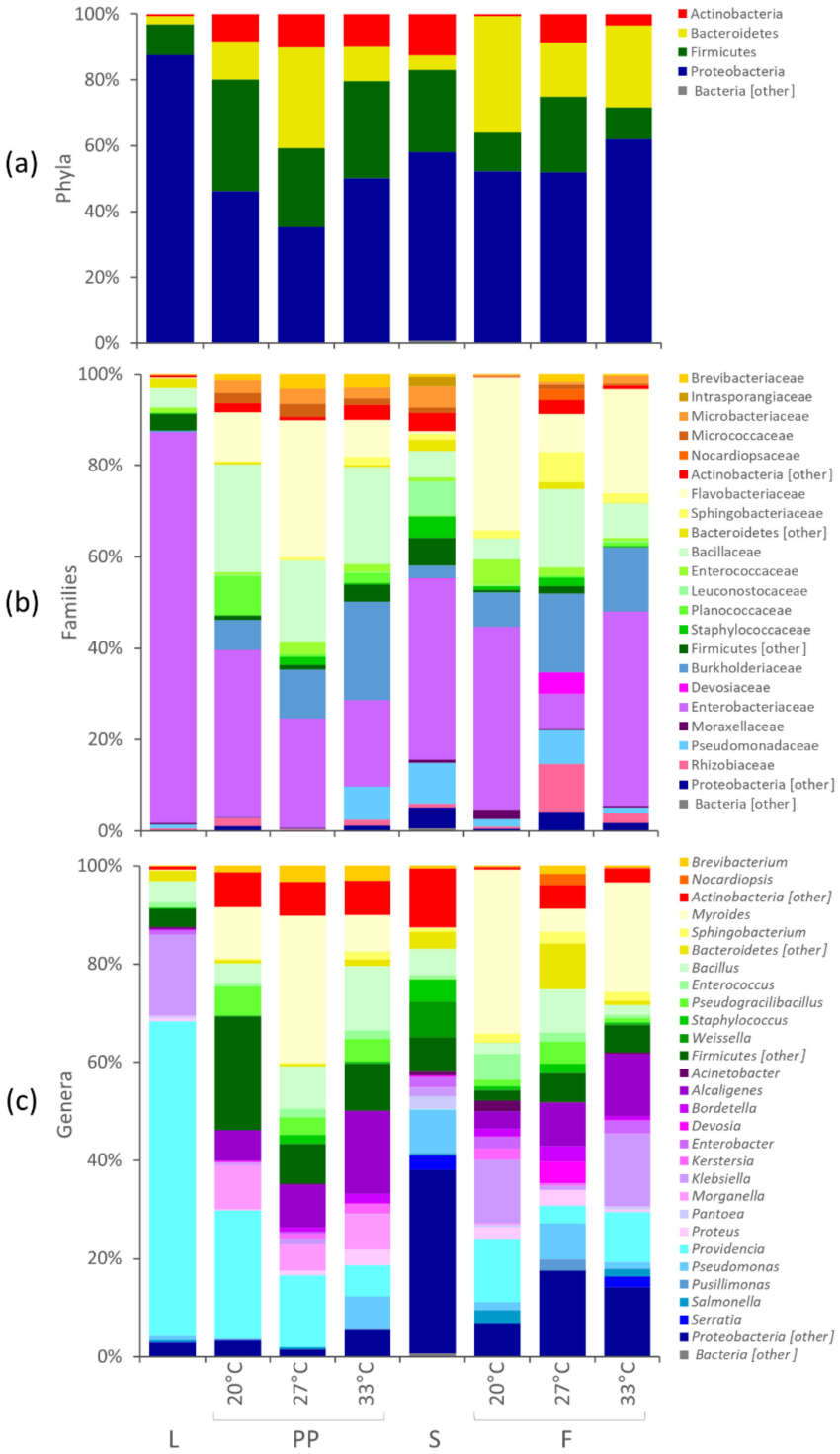


Figure 2. Stacked bar-plot representation of microbiota composition in L, S, PP, and F with taxonomic features collapsed at the level of phyla (**a**), families (**b**), and genera (**c**). The mean values of the two independent experiments are reported. The phyla, families, and genera that remained unclassified or never occurred with abundance >2.0% are grouped as others.

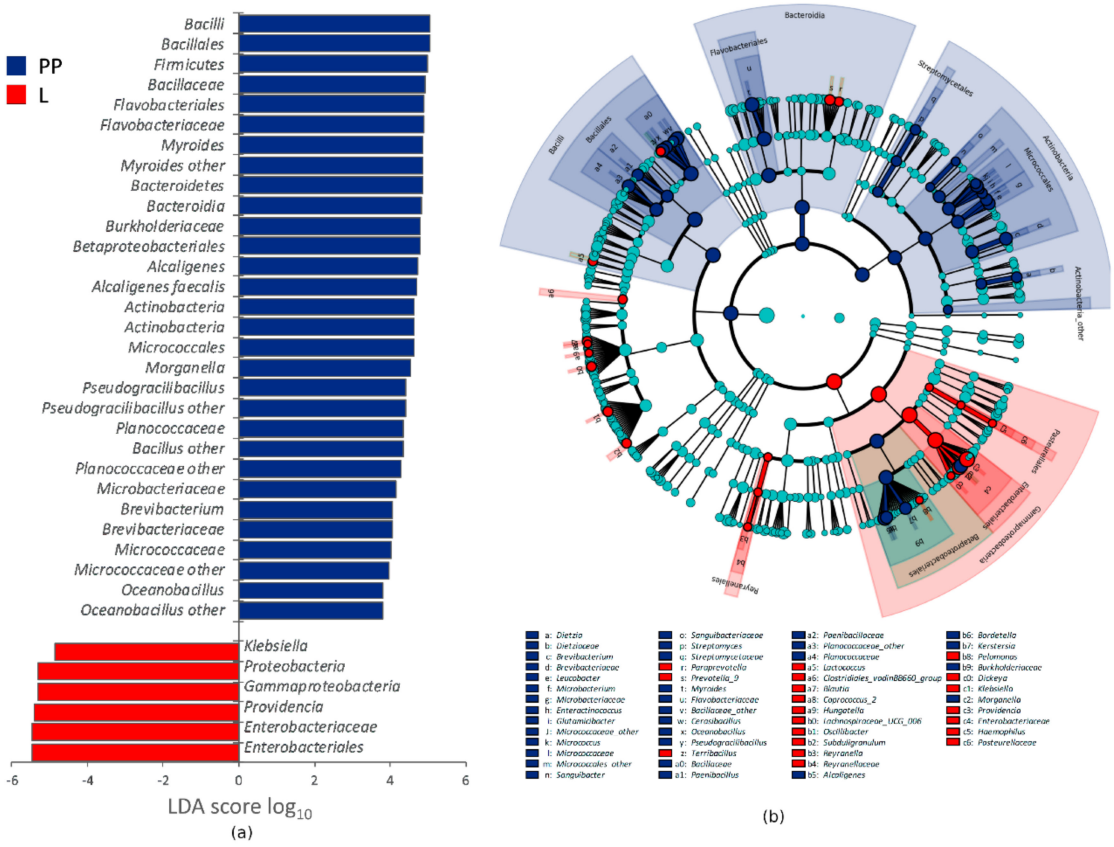


Figure 3. Linear discriminant analysis Effect Size (LEfSe) analysis of taxonomic features differentiating L (n = 2) and PP (n = 6), regardless of the growth temperature: (**a**) LDA logarithmic scores of taxonomic biomarkers exhibiting significant differential abundance ($p < 0.05$, logarithmic LDA logarithmic score ≥ 2.0) and appearing at least once with abundance >2.0%; (**b**) cladogram of all the taxonomic biomarkers.

The microbiota of S was characterized by 133 OTUs, 85 of which shared with L and 109 shared with PP. The remaining OTUs represented 6.3% of S microbiota, with only *Rahnella* being >1%. The microbiota of F encompassed 142 OTUs, 139 of which shared with PP. The composition of F differed from that of PP for being significantly richer in Proteobacteria (such as *Klebsiella* and *Acinetobacter*) and poorer in Actinobacteria (mainly Micrococcales, such as *Brevibacterium*), Firmicutes (Bacillales) and the Enterobacteriaceae *Morganella* (Supplementary Figure S3).

The increase of incubation temperature negatively correlated with the level of Enterobacteriaceae in PP, mostly due to the decrease of *Providencia* (Figure 4). Unlike *Providencia*, other Proteobacteria positively correlated with temperature, such as *Kertersia*, *Proteus*, *Pseudomonas*, *Alcaligenes*, and *Bordetella*. *Pseudomonas* was abundant only in PP reared at 33 °C (6.7%), whereas at the lower temperatures it accounted for less than 0.1%. *Bacillus* was also positively associated with increasing temperature and reached 13.1% at 33 °C. Increasing temperatures also favored the populations of other Actinobacteria (*Brevibacterium*), Bacilli (e.g., *Enterococcus*, *Pediococcus*, *Paenibacillus*, *Pseudogracilibacillus*), and Bacteroidetes (*Sphingobacterium*), although occurring in low amounts. The correlation of taxa with temperature is in agreement with their contribution in the PCoA biplot, resulting in PP samples being located at lower PCo1 values with the increase of the rearing temperature (Figure 1).

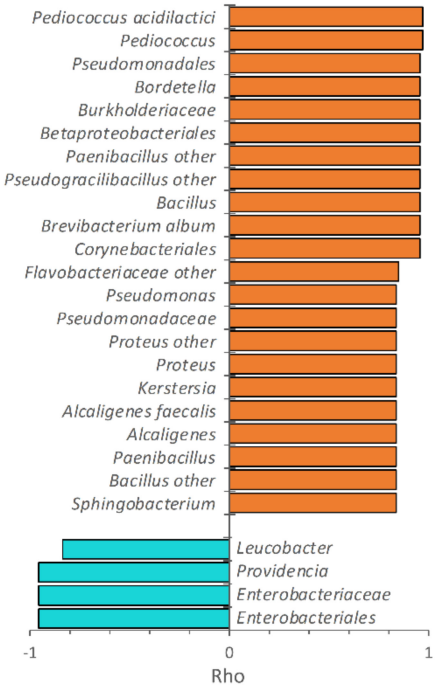


Figure 4. Spearman's rank correlation of the bacterial taxa in PP and the growth temperature. The correlation coefficient rho is reported only for the taxa exhibiting a positive (orange) or negative (cyan) significant correlation with temperature and appearing at least once with abundance >1.0% ($n = 2$, $p < 0.05$).

4. Discussion

In the present study, the larvae of *H. illucens* were reared at three different temperatures (20, 27, and 33 °C), utilizing a standard vegetal substrate, until reaching the PP stadium. The larvae efficiently grew across a range of temperatures, although with differences in the development rate, confirming the robustness of *H. illucens*. In the tropics, the development of *H. illucens* occurs yearlong, while it is restricted to a few generations per year in warm temperate regions [35], generally requiring temperatures above 26 °C [36]. Higher temperatures promoted faster larval development and greater final biomass but also higher mortality, in agreement with previous studies [37–39], suggesting that the best rearing temperature is 27 °C. The flexible rearing temperature enables the use of these insects also in developing countries, with minimal investment in energy-consuming devices for cooling or heating.

The first step of microbiological risk assessment in food or feed involves the collection of all information pertaining to potential pathogens that may exert any adverse effects on human or animal health. In the present study, the effect of rearing larvae at different temperatures on the microbiota of *H. illucens* was evaluated, with a focus on the viable load of the main food pathogens according to the Advisory Reports 2014/2372 and 2019/6200 [19,20]. The rearing temperature slightly affected the microbial load of PP, with only total mesophilic aerobes and Enterobacteriaceae significantly lower at 20 °C than at the higher temperatures. All the searched pathogens were detected in the whole set of PP samples and the counts of *Campylobacter* and Listariaceae got higher with the increasing temperature. Interestingly, the substrate was negative to all the pathogens except Listeriaceae. The larvae were not disinfected and were not reared under aseptic conditions, according to hygienic conditions that may be achieved in a production facility, where the substrate (e.g., a nonsterilized vegetal waste) is expected to be the main source of microbial contamination. It is likely that Enterobacteriaceae, *B. cereus*, *Campylobacter* spp., *C. perfringens*, coagulase-positive staphylococci, and *Salmonella* spp., laying below the limit of detection in S, were environmental contaminants or were introduced with small larvae. In any case, hygienic conditions must be kept under control throughout the multiple manufacturing processing stages, because the rearing of PP favored the blooming of the pathogens. Major attention has

to be paid to microbiological optimization through the processes aimed at stabilizing the insects, such as powdering, heating, drying, UV treating, high-energy microwaving, pasteurizing, and acidifying, which are expected to contain the load of pathogens.

Metagenome analysis was complementary to the outcome of traditional microbiological approaches, providing more comprehensive information on ecological aspects. Previous studies determined the microbiota composition of *H. illucens* by 16S rRNA gene profiling, focusing on entire specimens, with or without any surface sterilization, or on the dissected gut, in some cases distinguishing specific gut sections [13,15–18,40]. These studies pointed toward a great variability of *H. illucens* microbiota, with the substrate, the stage of development of the insect, and the gut section being the major players shaping the structure and the diversity of the bacterial community. In some cases, the microbiota of *H. illucens* was dominated by Bacteroidetes, in others by Proteobacteria, accompanied by very variable amounts of Actinobacteria and Firmicutes [13]. However, the existence of a core composition, independent of the substrate and transmitted through the development stages, seems plausible. In particular, *Providencia* was reported as one of the most recurrent and abundant genera, occurring in insects reared on different substrates. Consistently, in the present study, the microbiota of L was dominated by *Providencia* (>60%) and other Proteobateria (mainly *Klebsiella*) and evolved to a more complex composition with insect development. A lower amount of *Providencia* was still present in the microbiota of PP, while Actinobacteria, Bacteroidetes, and Bacilli bloomed (Figure 2). Many taxa dominating the microbiota of PP (e.g., *Providencia*, *Alcaligenes faecalis*, *Myroides*, *Morganella*, *Bordetella*, and *Kerstersia*) occurred also in L and were not found or were negligible in S. These taxa took the most advantage from the interaction with the insect, exhibiting the greatest fitness relative to *H. illucens*. Other taxa, such as *Pseudogracilibacillus* and other unclassified Bacillaceae, that colonized PP were not found in either L or S. It is not clear whether they were contaminants of the rearing environment or they initially lay below the limit of detection in L and S.

For the first time, a significant effect of the rearing temperature on the microbiota composition of *H. illucens* is reported. The temperature was negatively associated with the amount of *Providencia*, which was the lowest at 33 °C. On the contrary, the temperature was positively associated with a variety of other genera, such as *Alcaligenes*, *Pseudogracilibacillus, Bacillus, Proteus, Enterococcus*, *Pediococcus*, *Bordetella, Pseudomonas*, and *Kerstersia*. Thus, the temperature has to be included among the main factors affecting the microbiota and may have contributed, together with the rearing substrate, to the wide differences in the structure of the microbial community associated with *H. illucens* reported in the literature.

The present study also indicates that *H. illucens* largely shares the microbiota with the medium where it is reared, except for a few lowly represented taxa. The main taxa constituting the microbiota of F and PP were the same, although some of them presented a differential abundance between the two environments, confirming that the insects and the medium shape the microbiota of each other [16].

With respect to the microbiological risk assessment, attention should be paid to the relevant abundance of some genera, such as *Myroides*, *Proteus*, *Providencia*, and *Morganella*, revealed by the 16S rRNA gene profiling. Species within these genera are not considered primary pathogens, although there is a vast literature describing them as opportunistic pathogens that may bear drug resistances and may cause severe morbidity [41–45]. Characterizing the species of *Myroides*, *Proteus*, *Providencia*, and *Morganella* harbored by *H. illucens* is recommended to establish whether they could be capable of causing adverse effects if present in the final product or throughout the production process.

Supplementary Materials: The following are available online at http://www.mdpi.com/2076-2607/8/6/902/s1, Figure S1: Alpha diversity metrics of the microbiota in L, PP, S, and F; Figure S2: Distribution in the main 181 OTUs in L, PP, S, and F samples; Figure S3: LEfSe analysis of taxonomic features differentiating F and PP.

Author Contributions: Conceptualization, S.R., M.R., L.M., and A.A.; software, F.C.; formal analysis, A.A. and F.C.; investigation, S.R., G.S., L.I.M., L.L., and F.C.; resources, L.M.; data curation, A.A.; writing—original draft preparation, S.R., M.R., and A.A.; visualization, A.A.; supervision, S.R., M.R., L.M., and A.A.; project administration, L.M.; funding acquisition, L.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Rural Development Plan 2014-2020 (Regione Emilia-Romagna, Op. 16.1.01—GO EIP-Agri - FA 5C) in the project "BIOECO-FLIES", coordinated by CRPV (Cesena, Italy), and in part by ERDF Emilia-Romagna Regional Operational Programme 2014–2020, Italy, in the project "FLIES4VALUE" (grant number PG/2018/631984), coordinated by L. Maistrello.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Wang, Y.S.; Shelomi, M. Review of Black Soldier Fly (*Hermetia illucens*) as Animal Feed and Human Food. *Foods* **2017**, *6*, 91. [CrossRef]
2. Cutrignelli, M.I.; Messina, M.; Tulli, F.; Randazzo, B.; Olivotto, I.; Gasco, L.; Loponte, R.; Bovera, F. Evaluation of an insect meal of the Black Soldier Fly (*Hermetia illucens*) as soybean substitute: Intestinal morphometry, enzymatic and microbial activity in laying hens. *Res. Vet. Sci.* **2018**, *117*, 209–215. [CrossRef]
3. Terova, G.; Rimoldi, S.; Ascione, C.; Gini, E.; Ceccotti, C.; Gasco, L. Rainbow trout (*Oncorhynchus mykiss*) gut microbiota is modulated by insect meal from *Hermetia illucens* prepupae in the diet. *Rev. Fish Biol. Fisher.* **2019**, *29*, 465–486. [CrossRef]
4. Belghit, I.; Liland, N.S.; Gjesdal, P.; Biancarosa, I.; Menchetti, E.; Li, Y.; Waagbø, R.; Krogdahl, Å.; Lock, E.J. Black soldier fly larvae meal can replace fish meal in diets of sea-water phase Atlantic salmon (Salmo salar). *Aquaculture* **2019**, *503*, 609–619. [CrossRef]
5. Zarantoniello, M.; Randazzo, B.; Truzzi, C.; Giorgini, E.; Marcellucci, C.; Vargas-Abúndez, J.A.; Zimbelli, A.; Annibaldi, A.; Parisi, G.; Tulli, F.; et al. A six-months study on Black Soldier Fly (*Hermetia illucens*) based diets in zebrafish. *Sci. Rep.* **2019**, *9*, 8598. [CrossRef] [PubMed]
6. Truzzi, C.; Giorgini, E.; Annibaldi, A.; Antonucci, M.; Illuminati, S.; Scarponi, G.; Riolo, P.; Isidoro, N.; Conti, C.; Zarantoniello, M.; et al. Fatty acids profile of black soldier fly (*Hermetia illucens*): Influence of feeding substrate based on coffee-waste silverskin enriched with microalgae. *Anim. Feed Sci. Technol.* **2020**, *259*, 114309. [CrossRef]
7. Bruni, L.; Pastorelli, R.; Viti, C.; Gasco, L.; Parisi, G. Characterisation of the intestinal microbial communities of rainbow trout (*Oncorhynchus mykiss*) fed with *Hermetia illucens* (black soldier fly) partially defatted larva meal as partial dietary protein source. *Aquaculture* **2018**, *487*, 56–63. [CrossRef]
8. Barragan-Fonseca, K.B.; Dicke, M.; van Loon, J.J.A. Nutritional value of the black soldier fly (*Hermetia illucens* L.) and its suitability as animal feed—A review. *J. Ins. Food Feed.* **2017**, *3*, 105–120. [CrossRef]
9. Lalander, C.H.; Fidjeland, J.; Diener, S.; Eriksson, S.; Vinnerås, B. High waste-to-biomass conversion and efficient *Salmonella* spp. reduction using black soldier fly for waste recycling. *Agron. Sust. Dev.* **2015**, *35*, 261–271. [CrossRef]
10. Nguyen, T.T.; Tomberlin, J.K.; Vanlaerhoven, S. Ability of Black Soldier Fly (*Diptera*: *Stratiomyidae*) Larvae to Recycle Food Waste. *Environ. Entomol.* **2015**, *44*, 406–410. [CrossRef]
11. Bortolini, S.; Macavei, L.I.; Hadj Saadoun, J.; Foca, G.; Ulrici, A.; Bernini, F.; Malferrari, D.; Setti, L.; Ronga, D.; Maistrello, L. *Hermetia illucens* (L.) larvae as chicken manure management tool for circular economy. *J. Clean. Prod.* **2020**, *262*, 121289. [CrossRef]
12. EFSA Scientific Committee. Risk profile related to production and consumption of insects as food and feed. *EFSA J.* **2015**, *13*, 4257. [CrossRef]
13. De Smet, J.; Wynants, E.; Cos, P.; Van Campenhout, L. Microbial community dynamics during rearing of black soldier fly larvae (*Hermetia illucens*) and impact on exploitation potential. *Appl. Environ. Microbiol.* **2018**, *84*, e02722-17. [CrossRef] [PubMed]
14. Osimani, A.; Milanović, V.; Garofalo, C.; Cardinali, F.; Roncolini, A.; Sabbatini, R.; De Filippis, F.; Ercolini, D.; Gabucci, C.; Petruzzelli, A.; et al. Revealing the microbiota of marketed edible insects through PCR-DGGE, metagenomic sequencing and real-time PCR. *Int. J. Food Microbiol.* **2018**, *276*, 54–62. [CrossRef]
15. Wynants, E.; Frooninckx, L.; Crauwels, S.; Verreth, C.; De Smet, J.; Sandrock, C.; Wohlfahrt, J.; Van Schelt, J.; Depraetere, S.; Lievens, B.; et al. Assessing the microbiota of black soldier fly larvae (*Hermetia illucens*) reared on organic waste streams on four different locations at laboratory and large scale. *Microbial. Ecol.* **2018**, *77*, 913–930. [CrossRef]

16. Bruno, D.; Bonelli, M.; De Filippis, F.; Di Lelio, I.; Tettamanti, G.; Casartelli, M.; Ercolini, D.; Caccia, S. The Intestinal Microbiota of *Hermetia illucens* Larvae Is Affected by Diet and Shows a Diverse Composition in the Different Midgut Regions. *Appl. Environ. Microbiol.* **2019**, *85*, e01864-18. [CrossRef]
17. Jeon, H.; Park, S.; Choi, J.; Jeong, G.; Lee, S.B.; Choi, Y.; Lee, S.J. The intestinal bacterial community in the food waste-reducing larvae of *Hermetia illucens*. *Curr. Microbiol.* **2011**, *62*, 1390–1399. [CrossRef]
18. Zheng, L.; Crippen, T.L.; Singh, B.; Tarone, A.M.; Dowd, S.; Yu, Z.; Wood, T.K.; Tomberlin, J.K. A survey of bacterial diversity from successive life stages of black soldier fly (*Diptera*: *Stratiomyidae*) by using 16S rDNA pyrosequencing. *J. Med. Entomol.* **2013**, *50*, 647–658. [CrossRef]
19. NVWA. *Advisory Report on the Risks Associated with the Consumption of Mass-Reared Insects*; NVWA/BuRO/2014/2372; Netherlands Food and Consumer Product Safety Authority: Utrecht, The Netherlands, 2014. [CrossRef]
20. NVWA. *Advice on Animal and Public Health Risks of Insects Reared on Former Foodstuffs as Raw Material for Animal Feed*; TRCNVWA/2019/6200/EN; Netherlands Food and Consumer Product Safety Authority: Utrecht, The Netherlands, 2019.
21. Grabowski, N.T.; Klein, G. Microbiology of processed edible insect products—Results of a preliminary survey. *Int. J. Food Microbiol.* **2017**, *243*, 103–107. [CrossRef]
22. Hogsette, J.A. New diets for production of house-flies and stable flies (*Diptera*, *Muscidae*) in the laboratory. *J. Econ. Entomol.* **1992**, *85*, 2291–2294. [CrossRef]
23. ISO. *Microbiology of Food and Animal Feeding Stuffs—Horizontal Method for the Enumeration of Presumptive Bacillus Cereus—Colony-Count Technique at 30 Degrees C*; ISO 7932:2004; International Organization for Standardization: Geneva, Switzerland, 2004.
24. ISO. *Microbiology of Food and Animal Feeding Stuffs—Horizontal Method for the Enumeration of Clostridium perfringens—Colony-Count Technique*; ISO 7937:2004; International Organization for Standardization: Geneva, Switzerland, 2004.
25. ISO. *Microbiology of Food and Animal Feed—Horizontal Method for the Detection, Enumeration and Serotyping of Salmonella—Part 2: Enumeration by a Miniaturized Most Probable Number Technique*; ISO/TS 6579-2:2012; International Organization for Standardization: Geneva, Switzerland, 2012.
26. ISO. *Microbiology of the Food Chain—Horizontal Method for Detection and Enumeration of Campylobacter spp.—Part 2: Colony-Count Technique*; ISO 10272-2:2017; International Organization for Standardization: Geneva, Switzerland, 2017.
27. ISO. *Microbiology of the Food Chain—Horizontal Method for the Detection and Enumeration of Listeria monocytogenes and of Listeria spp.—Part 2: Enumeration Method*; ISO 11290-2:2017; International Organization for Standardization: Geneva, Switzerland, 2017.
28. ISO. *Microbiology of the Food Chain—Horizontal Method for the Detection, Enumeration and Serotyping of Salmonella—Part 1: Detection of Salmonella spp.*; ISO 6579-1:2017; International Organization for Standardization: Geneva, Switzerland, 2017.
29. ISO. *Microbiology of Food and Animal Feeding Stuffs—Horizontal Method for the Enumeration of Coagulase-Positive Staphylococci (Staphylococcus aureus and other species)—Part 1: Technique Using Baird-Parker Agar Medium*; ISO 6888-1:1999/AMD 2:2018; International Organization for Standardization: Geneva, Switzerland, 2018.
30. Milani, C.; Hevia, A.; Foroni, E.; Duranti, S.; Turroni, F.; Lugli, G.A.; Sanchez, B.; Martín, R.; Gueimonde, M.; van Sinderen, D.; et al. Assessing the fecal microbiota: An optimized ion torrent 16S rRNA gene-based analysis protocol. *PLoS ONE* **2013**, *8*, e68739. [CrossRef]
31. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [CrossRef]
32. Bokulich, N.A.; Kaehler, B.D.; Rideout, J.R.; Dillon, M.; Bolyen, E.; Knight, R.; Huttley, G.A.; Caporaso, J.G. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **2018**, *6*, 90. [CrossRef] [PubMed]
33. Bolyen, E.; Rideout, J.R.; Chase, J.; Pitman, T.A.; Shiffer, A.; Mercurio, W.; Dillon, M.R.; Caporaso, J.G. An introduction to applied bioinformatics: A free, open, and interactive text. *J. Open Source Educ.* **2018**, *1*, 27. [CrossRef] [PubMed]
34. Segata, N.; Izard, J.; Waldron, L.; Gevers, D.; Miropolsky, L.; Garrett, W.S.; Huttenhower, C. Metagenomic biomarker discovery and explanation. *Genome Biol.* **2011**, *12*, R60. [CrossRef] [PubMed]

35. Tomberlin, J.K.; Sheppard, D.C. Lekking behavior of the black soldier fly (*Diptera*: *Stratiomyidae*). *Fla. Entomol.* **2001**, *84*, 729–730. [CrossRef]
36. Park, K.H.; Kim, W.T.; Lee, S.B.; Choi, Y.C.; Nho, S.K. Seasonal Pupation, Adult Emergence and Mating of Black Soldier Fly, *Hermetia illucens* (Diptera: Stratiomyidae) in Artificial Rearing System. *Int. J. Ind. Entomol.* **2010**, *21*, 189–191.
37. Gligorescu, A.; Toft, S.; Hauggaard-Nielsen, H.; Axelsen, J.A.; Nielsen, S.A. Development, metabolism and nutrient composition of black soldier fly larvae (*Hermetia illucens*; Diptera: *Stratiomyidae*) in relation to temperature and diet. *J. Insects Food Feed* **2018**, *4*, 123–133. [CrossRef]
38. Harnden, L.M.; Tomberlin, J.K. Effects of temperature and diet on black soldier fly, *Hermetia illucens* (L.) (Diptera: *Stratiomyidae*), development. *Forensic Sci. Int.* **2016**, *266*, 109–116. [CrossRef]
39. Tomberlin, J.K.; Adler, P.H.; Myers, H.M. Development of the Black Soldier Fly (*Diptera*: *Stratiomyidae*) in Relation to Temperature. *Environ. Entomol.* **2009**, *38*, 930–934. [CrossRef]
40. Shelomi, M.; Wu, M.K.; Chen, S.M.; Huang, J.J.; Burke, C.G. Microbes associated with black soldier fly (Diptera: Stratiomiidae) degradation of food waste. *Environ. Entomol.* **2020**, *49*, 405–411. [CrossRef] [PubMed]
41. Armbruster, C.E.; Mobley, H.L.T.; Pearson, M.M. Pathogenesis of *Proteus mirabilis* infection. *EcoSal Plus* **2019**, *8*. [CrossRef] [PubMed]
42. LaVergne, S.; Gaufin, T.; Richman, D. *Myroides injenensis* Bacteremia and Severe Cellulitis. *Open Forum Infect. Dis.* **2019**, *6*, ofz282. [CrossRef] [PubMed]
43. Minnullina, L.F.; Pudova, D.; Shagimardanova, E.; Shigapova, L.; Sharipova, M.; Mardanova, A. Comparative genome analysis of uropathogenic *Morganella morganii* strains. *Front. Cell. Infect. Microbiol.* **2019**, *9*, 167. [CrossRef] [PubMed]
44. O'Hara, C.M.; Brenner, F.W.; Miller, J.M. Classification, identification, and clinical significance of *Proteus*, *Providencia*, and *Morganella*. *Clin. Microbiol. Rev.* **2000**, *13*, 534–546. [CrossRef] [PubMed]
45. Sharma, D.; Sharma, P.; Soni, P. First case report of *Providencia rettgeri* neonatal sepsis. *BMC Res. Notes* **2017**, *10*, 536. [CrossRef]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ORIGINAL RESEARCH ARTICLE
Front. Microbiol. | doi: 10.3389/fmicb.2020.605127

Comparative genomics of Leuconostoc carnosum

Provisionally accepted

The final, formatted version of the article will be published soon. Notify me

Francesco Candeliere[1], Stefano Raimondi[1], Gloria Spampinato[1], Moon Y. Tay[2, 3], Alberto Amaretti[1, 4], Jorgen Schlundt[2, 3] and Maddalena Rossi[1, 4*]

[1]University of Modena and Reggio Emilia, Italy
[2]Nanyang Technological University, Singapore
[3]School of Chemical and Biomedical Engineering, College of Engineering, Nanyang Technological University, Singapore
[4]Biogest Siteia, University of Modena and Reggio Emilia, Italy

Leuconostoc carnosum is a known colonizer of meat-related food matrices. It reaches remarkably high loads during the shelf life in packaged meat products and plays a role in spoilage, although preservative effects have been proposed for some strains. In this study, the draft genomes of 17 strains of L. carnosum (i.e. all the strains that have been sequenced so far)were compared to decipher their metabolic and functional potential and to determine their role in food transformations. Genome comparison and pathway reconstruction indicated that L. carnosum is a compact group of closely related heterofermentative bacteria sharing most of the metabolic features. Adaptation to a nitrogen-rich environment such as meat is evidenced by 23 peptidase genes identified in the core genome and by the autotrophy for nitrogen compounds including several amino acids, vitamins, and cofactors. Genes encoding the decarboxylases yielding biogenic amines were not present. All the strains harbored 1 to 4 of 32 different plasmids, bearing functions associated to proteins hydrolysis, transport of amino acids and oligopeptides, exopolysaccharides, and various resistances (e.g. to environmental stresses, bacteriophages, and heavy metals). Functions associated to bacteriocin synthesis, secretion, and immunity were also found in plasmids. While genes for lactococcin were found in most plasmids, only three harbored the genes for leucocin B, a class IIa antilisterial bacteriocin. Determinants of antibiotic resistances were absent in both plasmids and chromosomes.

Keywords: Leuconostoc carnosum, Genomics, Pangenome analysis, bacteriocin, Metabolism

Received: 11 Sep 2020; **Accepted:** 04 Dec 2020.

Copyright: © 2020 Candeliere, Raimondi, Spampinato, Tay, Amaretti, Schlundt and Rossi. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

*** Correspondence:** Prof. Maddalena Rossi, University of Modena and Reggio Emilia, Modena, 41121, Emilia-Romagna, Italy, maddalena.rossi@unimore.it