Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

SCUOLA DI DOTTORATO DI RICERCA IN: Bioscienze e Biotecnologie

INDIRIZZO: Biotecnologie

CICLO XXVIII

16S amplicon Next Generation Sequencing approach evaluation and its application to food microbial communities characterization

Direttore della Scuola: Ch.mo Prof. Paolo Bernardi

Coordinatore d'indirizzo: Ch.ma Prof.ssa Fiorella Lo Schiavo

Supervisore: Ch.mo Prof. Andrea Squartini

Co-supervisore: Dr.ssa Barbara Simionati

Dottoranda: Eleonora Sattin

BMR Genomics srl, Padova

sede del laboratorio di ricerca ed analisi dove

è stata svolta gran parte delle attività descritte nella tesi.

ABSTRACT

Foods are complex ecosystems, due to the influence of either intrinsic and extrinsic factors such as the chemical composition, microbial communities and the environment that can have an effect on the final product. In cheese microbes are the most influencing factor because they can effect both shelf-life and quality flavor/taste of the food. In addition microbiome establishes complex interactions within the bacterial community itself and between the food environment, such that nowadays these dynamics have been only partially highlighted with traditional culture-dependent technique. Thus, culture-independent methods such as Next Generation Sequencing (NGS) approach have become a powerful tool to study in depth microbial communities contained in a variety of ecological niches, including food. In fact they should be increasingly considered as a complementary technique to culture-dependent methods in microbiology.

In this comprehensive study we have evaluated and applied a 16S Next Generation Sequencing approach to the study of Ricotta cheese shelf life, its raw materials and to food technology studies in soft or ripened cheeses using adjunct cultures. In all these studies NGS method have always been supported with culture-dependent approaches, other molecular methods and biochemical analysis. The different studies are separated in the subsequent chapters of this dissertation.

First of all using a mock community we evaluated the impact of 16S NGS pipeline on the final community structure and identification, showing that bioinformatic analysis, and above all the OTU-picking strategy, can affect the *alpha*-diversity (within the sample) of the community but not its *beta*-diversity (between samples). To this extent both strategies can be used in order to compare community structures. Moreover the protocol was showed as robust and repeatable.

Secondly we evaluated the raw materials (wheys and cream) used for ricotta production showing that the microbiological quality of these materials must be improved. As they showed a very biodiverse micro flora (mainly *Streptococcaceae* and *Pseudomonadaceae*) and a high concentration of aerobic spores (10 to 10^4 CFU/ml, not present in cream) their usage for Ricotta-making could damage the final product. Moreover NGS combined with Gas Cromatography approach on organoleptic non satisfactory frozen wheys showed that under-hygienic storage of the raw materials drove to the production of fermentation or lipid oxidation compounds influenced by particular bacteria such as *Acinetobacter, Lactococcus* and *Pseudomonas*.

In the third chapter we evaluated Ricotta shelf life during winter and summer season, showing that Ricotta microbiome was mainly composed of spore-forming bacteria,

Bacillus, Paenibacillus and *Clostridium* (more than 80% of relative abundance). Moreover we found a summer ricotta with a pink discoloration and, applying a molecular Multi-locus Sequence Typing approach, we demonstrated that the bacterium causing the defect belonged to the *Bacillus cereus* group, in particular it was a *B. mycoides/wehienstephanensis* strain. This work has been published in Food Microbiology journal.

In the last chapter we applied NGS analysis to the technological use of adjunct cultures on soft or ripened cheeses in order to evaluate their anti-spoilers effect. We evaluated the anti-*gammaproteobacteria* effect of some strains such as *L. rhamnosus, L. sakei* and *Carnobacterium maltoaromaticum* in fresh industrial cheeses prone to package swelling and premature spoiling. From a NGS point of view we found that *C. maltoaromaticum* managed to eliminate almost all contaminants at different storage temperatures.

In conclusion with NGS approach we could improve food microbiological quality assessment and support food technology research. In fact the results obtained in the present studies demonstrate how the application of NGS technologies, describing in-depth microbial communities, could in the next future become a scalable, cost-effective and suitable tool in the food industry for improving the quality and the safety of products.

RIASSUNTO

Gli alimenti sono ecosistemi complessi in quanto vengono influenzati sia da fattori intrinseci che da fattori estrinseci, come la composizione chimica, il profilo microbiologico e l'ambiente, che possono avere forte impatto sul prodotto finale. In particolare nei formaggi i microbi sono il fattore più influente perché possono modificare sia la vita commerciale che il sapore ed il gusto del prodotto. Inoltre il microbiota stabilisce delle complesse interazioni internamente alla comunità batterica e con l'ambiente, a tal punto che al giorno d'oggi queste dinamiche sono state solo in parte evidenziate con metodi tradizionali coltura-dipendenti. Di conseguenza i metodi coltura-indipendenti, come il *Sequenziamento di Nuova Generazione* (NGS), stanno diventando un potente strumento per studiare in profondità le comunità microbiche presenti in una varietà di nicchie ecologiche, compresi gli alimenti. È probabile che ormai tali tecniche debbano essere considerate complementari ai metodi di microbiologia coltura-dipendenti.

In questo ampio progetto abbiamo valutato e applicato un approccio NGS sul gene 16S rRNA allo studio della vita commerciale della Ricotta, delle sue materie prime e ad alcuni studi di tecnologia degli alimenti su formaggi molli o stagionati addizionati con colture batteriche antagoniste. In tutti questi studi i metodi NGS sono sempre stati supportati da approcci cultura-dipendenti, altri metodi molecolari e analisi biochimiche. I vari studi sono suddivisi tra i successivi capitoli di questa tesi.

Per prima cosa utilizzando una comunità batterica artificiale abbiamo valutato l'impatto del *workflow* di processo del protocollo 16S-NGS sulla struttura e sull'identificazione della comunità reale. Con questo approccio abbiamo dimostrato che l'analisi bioinformatica, e principalmente la strategia di raggruppamento delle sequenze in OTU (*Operational Taxonomic Unit*), può influenzare la diversità della comunità di tipo *alfa* (intra-campione), ma non la sua *beta*-diversità (inter-campioni). In questo senso entrambe le strategie possono essere utilizzate al fine di confrontare le strutture delle comunità microbiche. Inoltre il protocollo si è dimostrato robusto e ripetibile.

In secondo luogo abbiamo valutato le materie prime (siero e panna) utilizzate per la produzione di ricotta, dimostrando che la qualità microbiologica di queste materie deve essere migliorata. Poiché queste hanno mostrato una forte biodiversità (principalmente di *Streptococcaceae* e *Pseudomonadaceae*) e un'alta concentrazione di spore aerobie (da 10 a 10^4 UFC/ml, non presenti nella panna), il loro utilizzo per la produzione della ricotta potrebbe danneggiare il prodotto finale. Non di meno l'approccio combinato NGS con Gas Cromatografia su sieri congelati con difetti organolettici ha mostrato che lo stoccaggio non igienico delle materie prime conduce alla produzione di composti di fermentazione o di

ossidazione lipidica, guidati anche da particolari batteri come *Acinetobacter, Lactococcus* e *Pseudomonas*.

Nel terzo capitolo abbiamo valutato la vita commerciale della Ricotta durante l'inverno e l'estate, dimostrando che il microbioma della Ricotta era composto principalmente da batteri sporigeni, *Bacillus*, *Paenibacillus* e *Clostridium* (oltre l'80% di abbondanza relativa). Inoltre abbiamo trovato una ricotta estiva con colorazione rosa e, applicando l'approccio molecolare Multi-locus Sequence Typing, abbiamo dimostrato che il batterio che causava il difetto apparteneva al gruppo del *Bacillus cereus*, in particolare si trattava di un ceppo di *B. mycoides/wehienstephanensis*. Tale lavoro è stato pubblicato nella rivista Food Microbiology.

Nell'ultimo capitolo abbiamo applicato l'analisi NGS a studi di culture ausiliari aggiunte a formaggi a pasta molle o stagionati, al fine di evidenziare il loro effetto antibatterico. Abbiamo così valutato l'effetto anti-*Gammaproteobacteria* di alcuni ceppi come *L. rhamnosus, L. sakei* e *Carnobacterium maltoaromaticum* in formaggi freschi industriali con problematiche di rigonfiamento della confezione e deterioramento prematuro. Dal punto di vista NGS abbiamo evidenziato che *C. maltoaromaticum* riusciva ad eliminare quasi tutti i contaminanti a temperature di conservazione differenti.

In conclusione, con l'approccio NGS, è stato possibile migliorare la valutazione della qualità microbiologica alimentare e sostenere la ricerca tecnologica alimentare. Infatti i risultati ottenuti in questo studio dimostrano come l'applicazione di tecnologie NGS, descrivendo approfonditamente le comunità microbiche, potrebbe diventare nel prossimo futuro uno strumento scalabile ed efficace, adatto per migliorare la qualità e la sicurezza alimentare, nonché favorire la produzione di nuovi alimenti.

Index

1. Chapter I

16S Next Generation Sequencing evaluation for microbial communities characterization

1.1. Introduction

1.1.1. Prokaryotes: the unseen majority

Microorganisms are the most diverse group of living being, divided into *Prokaryotes* (Bacteria and Archaea) and microbial *Eukaryotes* (Fig. 1.1.1). Among them the number of Prokaryotes on Earth is estimated to be $4–6x10^{30}$ cells, distributed in almost every imaginable environment on Earth. They catalyze unique and indispensable transformations in the biogeochemical cycles of the biosphere, produce important components of the Earth's atmosphere and represent a large portion of life's genetic diversity [28].

Historically microbes have been analyzed by isolating them in pure cultures on artificial media. However only in 1990, studies of DNA-DNA reassociation kinetics of soil DNA by Torsvik and colleagues [32] provided the compelling evidence that culturing did not capture the complete spectrum of microorganism. They showed that the majority of microbial cells that could be seen in a microscope could not be induced to produce colonies on Petri plates or cultures in test tubes.

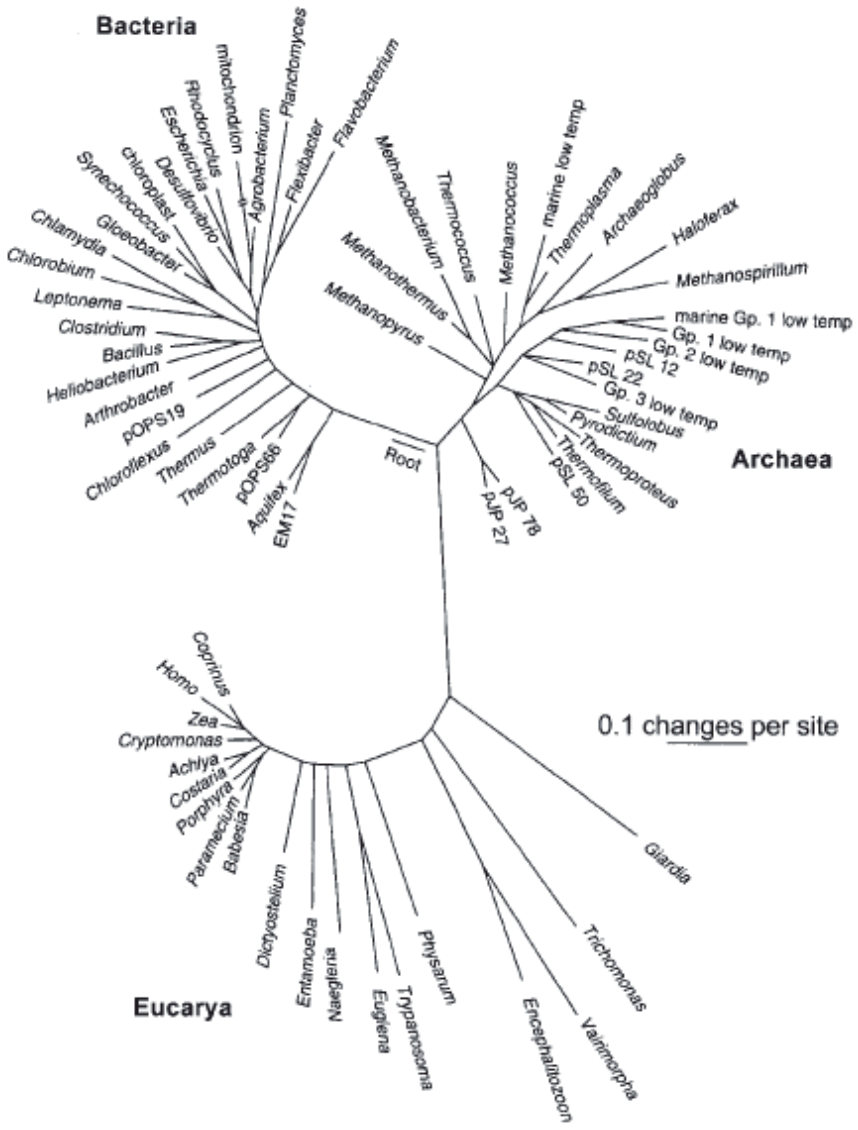

Fig. 1.1.1. Tree of life from [97].

Thus it was clear that, as only an estimated 5 % or less of the microbial diversity in the biosphere is thought to be cultivable with standard culturing techniques, this approach had led to a narrow picture of the diversity of an ecosystem. In fact Bacteria, Archaea, and microbial Eukaryotes do not exist as isolated taxa and do not live in single species ecosystems. Thus the usage of culture-dependent methods has limited our ability to examine the interactions between microbial species, and to understand the species-habitat and community-habitat interactions [27]. To this extent molecular approaches such as nucleic acids amplification and sequencing have become more and more interesting for microbiology studies on communities.

1.1.2. Evolution of microbial communities analysis

Initially, most of the studies carried out on community diversity analysis were based on traditional approaches, such as denaturing gradient gel electrophoresis (DGGE) on marker amplicons [47], terminal restriction fragment length polymorphism (T-RFLP) analysis [48], or Sanger sequencing of 16S rRNA gene clone libraries [49]. Sanger sequencing of 16S rRNA gene was the dominant approach from 1990 onwards and has been used extensively to access microbial community from almost every environment. In fact microbial ecology study was revolutionized by Schmidt and colleagues (1991) with the method of DNA cloning directly from the environment. They characterized 16S rRNA sequences from a Pacific Ocean picoplankton population by cloning environmental DNA into a phage genome and screening for clones that contained 16S rRNA genes [31].

Some years later, in 1998, this technique was called “metagenomics” [50], meaning that researchers were now exploring the genomic DNA from all the organisms in an environmental community, going beyond the single genome. Before arrival of Next Generation Sequencing (NGS) platforms, Venter and colleagues [52] in 2004 generated high magnitude metagenomics sequence data to the tone of 1.66 million reads with an average read length of 818 bp from samples collected from Sargasso Sea. That was a first example of metagenomics big data and from that moment to nowadays, with the advent of high-throughput sequencing strategies, the production of large amounts of data has become much easier and widely used. Moreover later this technique has divided into two branches: amplicon metagenomics (sequencing of libraries of a PCR-amplified gene of interest), and shotgun metagenomics (screening or sequencing of libraries of randomly isolated DNA fragments) [51]. In this chapter we are discussing the amplicon metagenomics on 16S rRNA gene approach.

1.1.3. 16S rRNA: a phylogenetic and taxonomic marker gene

In a comprehensive work published in 1987, Woese and colleagues explained that the 16S rRNA gene is a reliable evolutionary chronometer for microbiology studies [33]. At that time the development of PCR technology and primers designed to amplify the complete 16S rRNA gene had a spiking effect towards this direction, thus allowing 16S rRNA gene to become the phylogenetic marker of choice. In fact rRNA gene is now widely used for estimation of evolutionary history and taxonomic assignment of individual organisms due to its universal presence in all bacteria, its multi-gene nature, and its size (about 1500 bp) large enough for informatics purpose [33].

Moreover the assumption of its conservation is widely accepted: rRNA genes are essential components of the ribosome, which consists of more than 50 proteins and three classes of RNA molecules (16S, 5S and 23S); precise spatial relationships may be essential for assembly of functional ribosomes, so that drastic changes in rRNA gene sequence are prevented [33, 34]. In bacteria, the three rRNA genes are organized into a gene cluster which is expressed as single operon, which may be present in multiple copies in the genome. In organisms with multiple rRNA gene operons, the gene sequences tend to evolve in concert [35]. Copies of rRNA genes within an organism are subject to a homogenization process through homologous recombination that maintains their fit within the ribosome [35]. In fact in the work from Pei A. et al. (2010) just 1% among 425 bacteria species was found to show a variable intra-genic (between 6,38% and 20.38%) diversity that could be explained with the presence of 2° structure constrains of the rRNA [36].

Hypervariable regions

Alignments of bacterial 16S rRNA gene sequences have revealed nine separate hypervariable regions, V1 – V9 [37]. Chakravorty et al. (2007) described the borders of each hypervariable region for many bacterial species based on *E. coli* system of nomenclature. The nine regions spanned nucleotides 69-99, 137-242, 433-497, 576-682, 822-879, 986-1043, 1117-1173, 1243-1294 and 1435-1465 for V1 through V9 respectively [38], fig. 1.1.2. These hypervariable regions are flanked by conserved stretches in most bacteria, enabling PCR amplification of target sequences using universal primers [40]. However every region has a diverse identity percentage and a different taxonomy assignment strength: a number of reports had revealed that the selected primer sets targeting partial 16S rRNA gene suffered from biases among taxa due to their coverage [41, 42]. Consequently it would be important to study which regions may be more suitable for identification of functional bacterial groups as precise as possible, depending on the target community. So far, a number of studies had evaluated the coverage of primers for different hypervariable regions of 16S rRNA gene. Few of them analyzed the taxonomic precision for various regions [43]. During the years many different V regions combination (V1-V3; V3-V4; V5-V6) or alone (V4; V6) have been suggested as the most informative 16S region for microbiome studies [11–13, 41-44]. However, as the fully universal primer pair doesn't exist, the right primer choice depends on the aim of the study and on the target community [6].

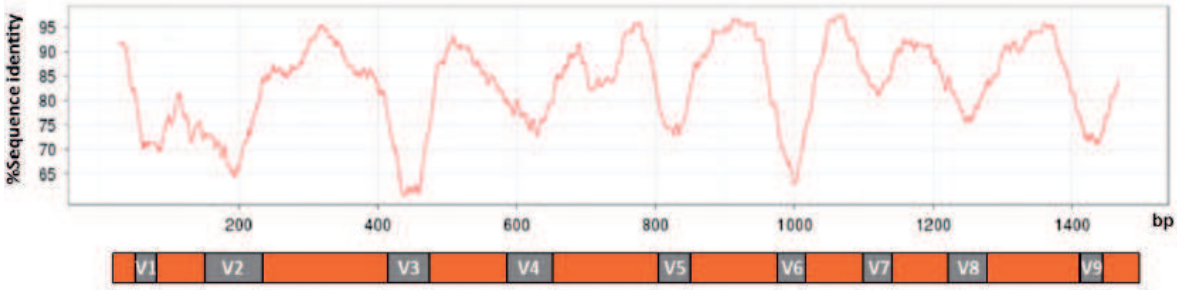

Fig. 1.1.2. Schematic overview of the 16S rRNA gene. The sequence identity of the 16S rRNA gene of more than 6,000 bacteria compared to the consensus sequence is shown. Hypervariable regions (V1-V9) are shown in grey and the conserved regions in orange.

16S databases

Widespread sequencing of ribosomal RNA genes has resulted in the generation of large reference databases, such as the ribosomal database project (RDP) II [39], Greengenes [22], and SILVA [45]. These comprehensive databases allow classification and comparison of environmental 16S rRNA gene sequences. All of these repositories present strengths and weaknesses relative to GreenGenes and can be considered for some analysis. SILVA includes microbial Eukaryotes and has invested substantial effort in cleaning up marine taxa [15]; RDP has close links to formally recognized names in taxonomy, which can be especially useful for medical microbiology [39]. All projects provide web-based software tools for the alignment and classification of sequences as well as probe match functionalities. Downloading of sequences is provided in various formats including the commonly used FASTA and GenBank file formats. Greengenes provides four features: a standardized set of descriptive fields, taxonomic assignment, chimera screening and ARB compatibility, but only for nearly full length sequences (>1250 bases) of *Bacteria* and *Archaea* [22].

1.1.4. Next Generation Sequencing approach

Next Generation Sequencing evolution can be divided into tree-generations steps [58]:

1. First generation: Sanger approach with ddNTP-terminated, dye-labeled products of PCR, which are subjected to high-resolution electrophoretic separation within one of 96 or 384 capillaries [53];

2. Second generation: massive parallel sequencing mostly based on *cyclic-array sequencing* [53] and imaging; clonal amplification of DNA molecules is required; the highest output is obtained by this platforms;

3. Third generation: high parallel sequencing where PCR is not needed before sequencing and the signal is captured in real time on a single molecule [58].

In the following paragraphs second-generation technologies are focused.

Platforms and Technology

NGS platforms allow massive parallel sequencing. This means that hundreds of thousands to hundreds of millions of sequencing reactions are performed and detected simultaneously, resulting in very high throughput. Through this approach, metagenomic DNA can be directly sequenced without spending time in cloning and bacteria-based library preparation, hugely increasing sequencing cost-effectiveness [53]. Even if different sequencing chemistry and platforms are present on the market, they are based on various implementations of *cyclic-array sequencing* [53]. The concept of cyclic-array sequencing can be summarized as the reading of a dense array of DNA molecules by iterative cycles of enzymatic interventions and imaging-based data collection [53-56].

The second generation sequencing platform comprehend the 454 (Roche) the first introduced on the market in 2005, Genome Analyzer from Illumina (earlier Solexa), the SOLiD platform (Applied Biosystems), the Polonator (Dover/Harvard) and the HeliScope Single Molecule Sequencer technology (Helicos; Cambridge, MA, USA). During the years, every sequencing company drove the evolution of different platforms towards increasing of throughput, read length and quality. I.e. in 2007 the Genome Analyzer (GA) from Illumina produced 1 Gb output. Nowadays HiSeq X series form Illumina produce more than 1800 Gb. Moreover read length has been increased from 36 bp (GA) to 300 bp (Miseq). A part from the highest-throughput platforms, sequencing companies have put on the market some smaller benchtop sequencers as *Personal Genome Machine* (PGM, Life Tecnhologies) and *MiSeq* (Illumina) with a lower output (from 2 to 15 Gb) but more versatile for different applications [59]. As Illumina MiSeq has been the selected sequencing platform for our studies, this technology will be focused and discussed in the following paragraphs.

Illumina technology

Clonal amplification

Illumina sequencing technology relies on a first step, in common with the other platforms, of library construction in which genomic DNA is manipulated to become a mixture of adaptor-flanked fragments up to several hundred base-pairs in length [53]. Second generation instruments cannot detect single molecule fluorescence, thus DNA fragments need to be pre-amplified by means PCR, in order to generate stronger fluorescence signals, to be detected by the instrument [62]. Amplified sequencing features are generated by bridge PCR [60,61], fig. 1.1.3. In this approach, both forward and reverse PCR primers are tethered to a solid substrate by a flexible linker, such that all amplicons arising from any single template molecule during the amplification remain immobilized and clustered to a

single physical location on an array. The resulting ‘clusters’ each consist of ~1,000 clonal amplicons [62].

After cluster generation, the amplicons are single stranded (linearization) and a sequencing primer is hybridized to a universal sequence flanking the region of interest. Each cycle of sequence interrogation consists of single-base extension with a modified DNA polymerase and a mixture of four nucleotides. These nucleotides are modified in two ways. They are ‘reversible terminators’, in that a chemically cleavable moiety at the 3′ hydroxyl position allows only a single-base incorporation to occur in each cycle; and one of four fluorescent labels, also chemically cleavable, corresponds to the identity of each nucleotide [63].

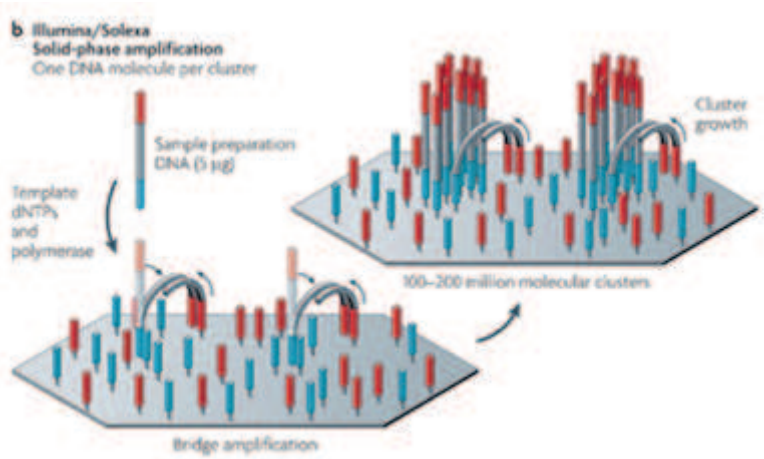

Fig. 1.1.3. Clonal amplification scheme, from [62].

Sequencing and Imaging

Each sequencing cycle includes the simultaneous addition of a mixture of four modified de-oxynucleotide species, each bearing one of four fluorescent dyes and a reversibly terminating moiety at the 3′ hydroxyl position. A modified DNA polymerase drives one-base per time extension of primed sequencing features. This is followed by imaging in four channels, nucleotides clearance and then cleavage of both the fluorescent labels and the terminating moiety. Illumina instruments use two lasers and four filters to detect four types of nucleotide (A, T, G, and C), fig. 1.1.4. The emission spectra of these four kinds of nucleotides have cross-talk, so the images of four nucleotides are not independent and the distribution of bases would affect the quality of sequencing [58]. To this extent Illumina library preparation need to enrich the low-diversity samples (such as 16S amplicons) with high PhiX percentage or a genome library, in order to improve data quality [64]. PhiX is a virus genome that is used as internal control in sequencing runs. It is a small genome, which enables quick alignment and estimation of error rates; it is diverse and contains approximately 45% GC and 55% AT; finally it is a well defined genome sequence that is discarded during post-run processing [64]. The imaging and the fluorescence mapping

from the array is performed by means RTA software (Real Time Analysis software). It is able to maintain focus on fluorescent array, register images to the cluster map, and make proper base calls to deliver high-quality sequencing data. New versions of this software now allow the usage of at least 5% of PhiX [65] while maintaining high quality data mapping in low-diversity samples.

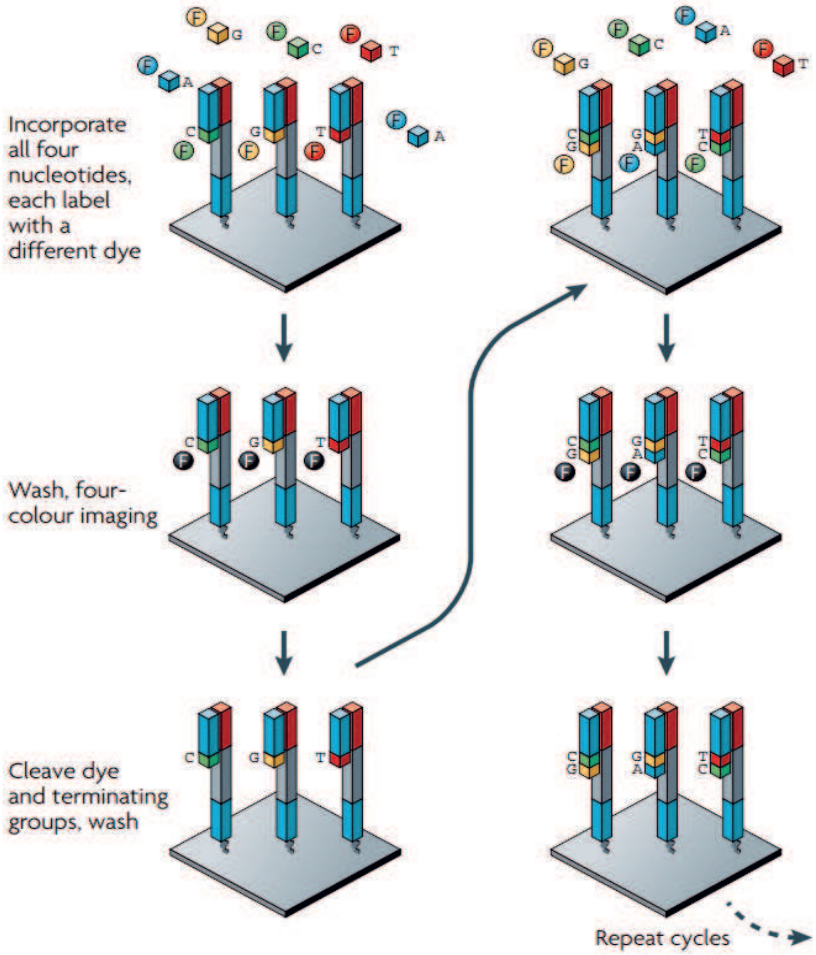

Fig. 1.1.4. Illumina Sequencing By Syntesis cycle from [62].

Illumina Miseq sequencer

MiSeq system integrates cluster-generation, amplification, sequencing and data analysis into a single instrument (fig. 1.1.5). It was released in 2011 and it is considered one of the most versatile benchtop sequencer. In fact it can provide all sequencing formats (from 36 bp to 300bp, either single or paired-end) with different throughputs, making it exploitable in many analysis, such as small (meta)genome, amplicon, exome, (meta)transcriptome sequencing [86]. The latest chemistry V3 has enabled the longest read sequencing for the Illumina sequencers, reaching 300 bp in length in a fast turnaround time. Miseq features are reported in tab. 1.1.1.

Fig. 1.1.5. Illumina MiSeq benchtop sequencer.

MISEQ REAGENT KIT V3			
READ LENGTH	TOTAL TIME	OUTPUT	%> Q30
2 × 75 bp	~21 hrs	3.3-3.8 Gb	>85%
2 × 300 bp	~56 hrs	13.2-15 Gb	>70%
PASSING FILTER READS			
Single-end reads		22-25 M	
Paired-end reads		44-50 M	

Tab. 1.1.1. Miseq reported features.

In comparison with other platforms such as PGM (LifeTechnology) and PacBio, Miseq shows the highest accuracy and the lowest error rate [1, 12, 87, 88]. Moreover compared to 454 pyrosequencing, the Illumina sequencing-bysynthesis methodology has a lower per-base error rate and is not as susceptible to indel errors in homopolymer stretches [66, 67, 89]. The significantly higher sequence quality of Illumina generated sequences, combined with a much lower cost per sequence compared to 454 pyrosequencing, has induced researchers to develop strategies to sequence 16S rRNA gene amplicons using Illumina systems [2, 3, 12, 90].

1.1.5. Bioinformatic analysis: QIIME

QIIME (Quantitative Insights Into Microbial Ecology) is an opensource bioinformatics software package designed for microbial community analysis based on DNA sequence data, which provides a single analysis framework for diverse analysis of raw sequence data [68]. In fact QIIME takes NGS platforms output and generates useful information about the community represented in each sample [15]. QIIME supports a wide range of microbial community analysis and visualizations that are useful for network analysis, histograms of within- or between-sample diversity and analysis of whether ‘core’ sets of organisms are consistently represented in certain habitats.

QIIME also provides graphical displays that allow users to interact with the data. The structure of this wrapper is highly modular and makes extensive use of unit testing to ensure the accuracy of results [68]. Valeria D'argenio and colleagues (2014) compared QIIME with another common pipeline (MG-RAST) and found that taxonomic assignment was more accurate with QIIME which, at family level, assigned a significantly higher number of reads. QIIME generated a more accurate BIOM file, which in turn improved the diversity analysis output. Moreover QIIME offered a wide range of metrics that are useful for downstream applications [16].

Workflow and community analysis

QIIME process can be divided into “upstream” and “downstream” steps, fig. 1.1.6. The upstream stage includes all the processing of the raw data (sequencing output) and generating the fundamental files (OTU table and phylogenetic tree) for microbial analysis [15]. The pre-processing step combines sample demultiplexing, primer removal and quality-filtering, as the latter is known to effect final community structure [20]. Then the reads are grouped into OTUs and representative sequences for each OTU are chosen, aligned and used to build the phylogenetic tree. The downstream step uses the OTU table and phylogenetic tree generated in the upstream step to perform diversity analysis, statistics, and interactive visualizations of the data [15].

The concept of OTU (Operational Taxonomic Unit)

In Biology the definition of species is ambiguous because the “species” does not exist as a natural entity. In fact taxonomists have always tried to work with different definitions, such as “genospecies,” “nomenspecies,” and “taxospecies” to organize life taxonomy, with many difficulties in finding a consensus. These definitions were mainly based on phenotipic and biochemical activity similarity [70]. Finally the arrival of molecular methods and sequencing allowed the usage of DNA to define objectively the relationships within and between species [70]. To this extent 16S rDNA-based method has quickly become a useful identification alternative when phenotypic characterization methods failed [69]. In fact the latter was biased by errors and the variability of character expression while 16S rDNA sequencing provided unambiguous data even for rare isolates, which are reproducible in and between laboratories [69]. On the other hand molecular methods introduced new issues in the taxonomic usage of sequence identity to infer taxa identification [70].

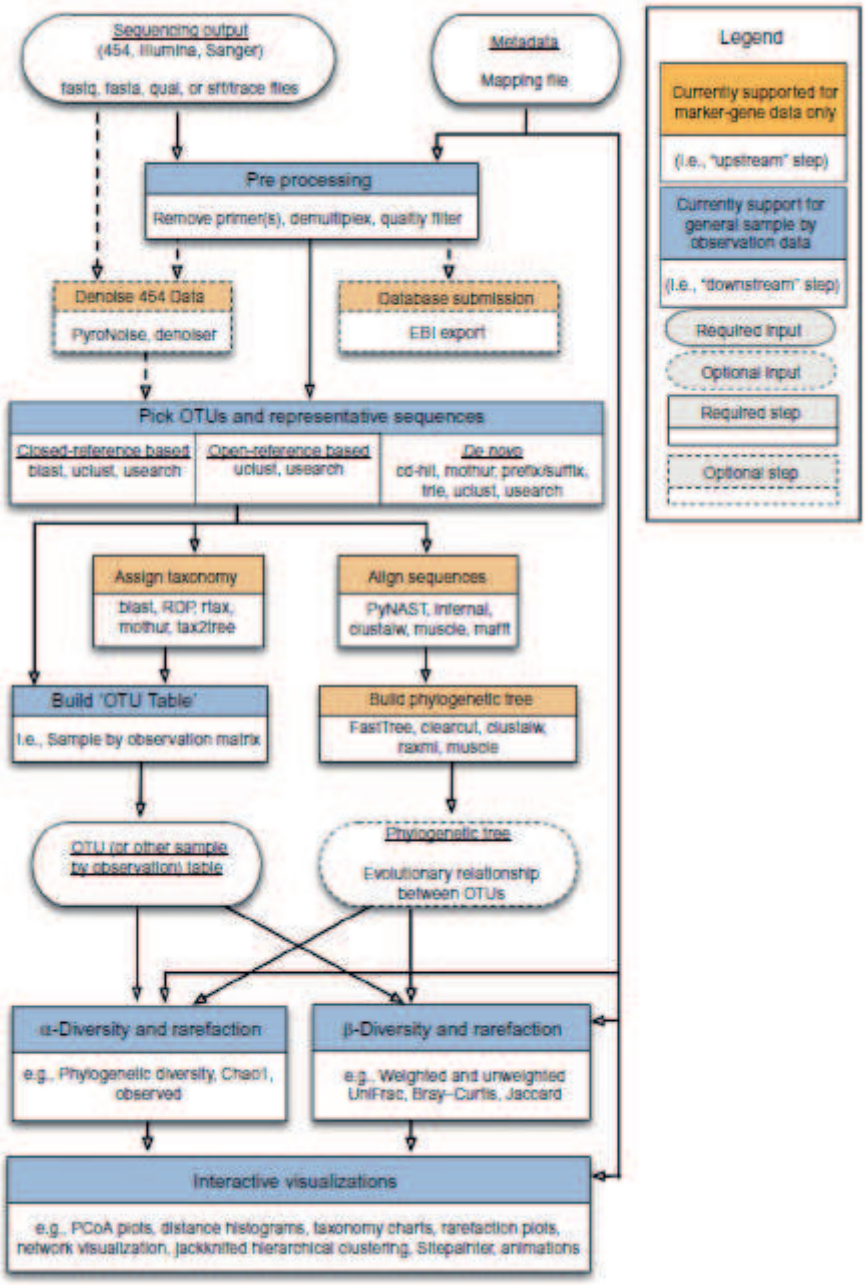

Fig. 1.1.6. Workflow of QIIME analysis, representing the different steps and the various software implemented in the wrapper. [15]

In fact the right threshold of 16S rRNA sequence identity for species definition is a debatable question because there is not always direct *phenotype-genome* and *16S rDNA-genome* identity correlation. Using the 16S rRNA gene sequence, a 97% similarity level has been proposed for the bacterial species delineation. This was because at sequence homology values below about 97%, it is unlikely that two organisms have more than 60 to 70% genomic DNA similarity and hence that they are related at the species level [71].

On the other hand Fox and colleagues showed that effective identity of 16s rRNA sequences does not always guarantee species identity as determined by the DNA-DNA hybridization criterion but strains that are well resolved by 16s rRNA methods are also distinguishable on the basis of DNA-DNA hybridization results. To this extent 16S rRNA gene is useful to exclude species identity [72]. Finally Drancourt and colleagues (2000), performed a work on about 300 bacterial isolates, managing to identify almost 90% of them by using 16S rRNA gene sequencing and 97% of threshold identity for bacterial genus [69]. This work has become the inspiring method for QIIME and other software default clustering algorithm. In this way in NGS field a microbial community is no more clustered into groups of organisms defined by intrinsic phenotypic similarity but clustered into groups of reads that are more similar than 97% identity threshold, defining the OTU [15].

OTU table

The OTU table is a sample-by-observation matrix that includes the OTUs for each sample and also the taxonomic prediction for each OTU. For the OTU table representation, QIIME uses the Genomics Standards Consortium candidate standard Biological Observation Matrix (BIOM) format [73].

Phylogenetic tree

The phylogenetic tree is created from the multiple sequence alignment on the representative sequences from each OTU, made with software like PyNAST or ClustalW. It represents the relationships among sequences in terms of the amount of sequence evolution from a common ancestor. One of the most used method for tree computation is *FastTree* that is based on Neighbor-Joining algorithm [74]. This phylogenetic tree is used in many downstream analyses.

Alpha and Beta diversity concepts

Understanding the principles that govern ecological communities requires rigorous metrics for describing community properties. Communities can be described using the concept of diversity or that of community structure. Whittaker [75] first proposed that diversity could be thought of as occurring on three levels: local diversity (alpha), regional (gamma) and change in community composition across local communities (beta). Beta diversity has emerged as a particularly powerful tool for understanding the principles governing ecological communities [76].

Whittaker used the definition of alpha-diversity addressing species diversity both in a single sample and the mean species diversity in a collection of samples [75]. Usually

alpha-diversity is described by one or more diversity indices, such as *species richness* (number of species present), the *Shannon* index or the *Simpson* index that take into account species abundance and evennes [77,78]. QIIME implements many of the most widely used alpha-diversity indices, including both phylogenetic indices (which require a phylogenetic tree) and non-phylogenetic indices [15]. Among the first group, Phylogenetic Diversity is one of the most used [78]: it is a quantitative measure of phylogenetic diversity, “PD”, and is measured as the minimum total length of all the phylogenetic branches required to span a given set of taxa on the phylogenetic tree [82].

One of the most used metrics in beta-diversity analysis is UniFrac that, coupled with standard multivariate statistical techniques including principal coordinates analysis (PCoA), identifies factors explaining differences among microbial communities [81]. UniFrac measures the amount of unique evolution within each community with respect to another by calculating the fraction of branch length of the phylogenetic tree that is unique to either one of a pair of communities [81].

QIIME implements several variants of UniFrac, including weighted and unweighted UniFrac. The weighted UniFrac metric is weighted by the difference in probability mass of OTUs from each community for each branch, whereas unweighted UniFrac only considers the absence/presence of the OTUs [15]. Thus weighted measures discover if the difference between two communities arises from variable species abundance while unweighted from different species [15, 80]. UniFrac can calculate whether two communities differ significantly using i.e. the UniFrac significance test [81]. The UniFrac significance test measures similarity between communities as the fraction of branch length in the tree that is unique, meaning that it leads to descendants in one environment or the other but not both. The *P*-value is the fraction of trials where the true tree has more unique branch length than trees in which the environment assignments have been randomized [84].

UPGMA hyerarchical clustering

UniFrac can compare many communities simultaneously using hierarchical clustering with the UPGMA (Unweighted Pair Group Method with Arithmetic Mean) algorithm [85]. Using the UniFrac metric, a distance matrix between environments is made calculating values for all possible pairs of environments in the tree [81]. The distance matrix can be visualized either by means a tree or a PCoA (Principal Coordinates Analysis plot).

1.1.6. Tree of decisions for 16S analysis

16S rRNA sequencing is economical and therefore scales to larger projects. However choices made at every step, from study design to analysis, can impact results. The following fig. 1.1.7 highlights the main choices we evaluated and tested for this pipeline.

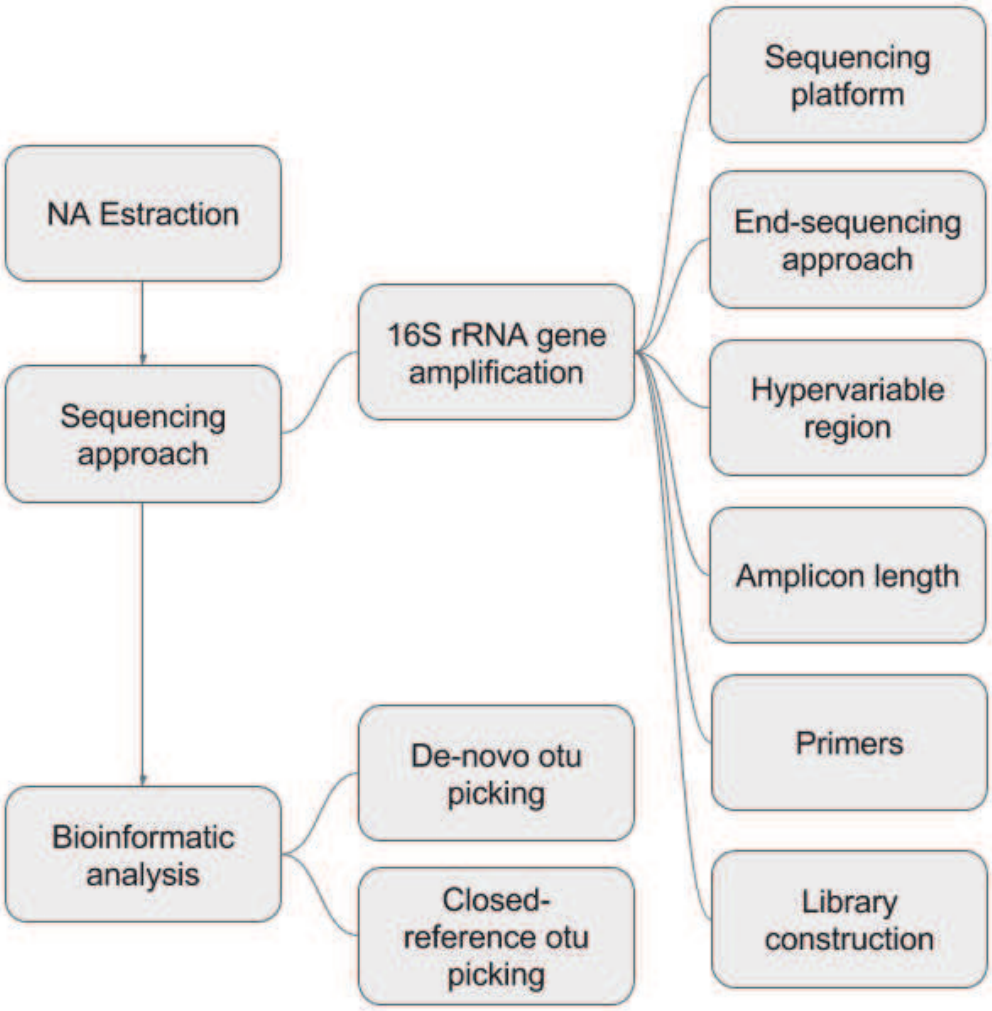

Fig. 1.1.7. Tree of decisions for 16S workflow evaluation.

1.2. Purpose

The primary goal of this study is to develop and evaluate the different protocol steps of 16S rRNA Next Generation Sequencing, in order to create a workflow to be applied in subsequent studies. As every step is strongly linked with the others and can affect final results, we want to evaluate the fundamental nodes in 16S rRNA sequencing pipeline:

- primers, library construction protocol and bioinformatic pipeline efficiency;
- the repeatability of the results in the same run and across different runs using an entire pipeline and a mock community;
- the bias on the results given by a different amount of target DNA.

1.3. Materials and methods

1.3.1. End-sequencing approach

Single-read sequencing involves DNA reading from only one end while paired-end sequencing allows to sequence both 5' and 3' ends of a fragment. The second approach can double the number of bp per read for the Illumina platform. However, before the release of long-reads Illumina sequencers such as Hiseq 2000 and Miseq, the usage of paired end approach for 16S analysis was tricky and uncommon. In fact initial Illumina-based methods for sequencing 16S rRNA genes have been limited by ≤101-base sequence reads [1, 2, 3] and/or an inability to leverage the paired-end approach that would allow for assembly of reads and reduced sequencing error [1, 2].

2x300 bp paired end sequencing approach was applied at the first run on 16S amplicon libraries on DNA extracted from bacteria and dairy products. The output R1 and R2 reads were downloaded from the Basespace and analyzed with FastQC software v.0.11.2, a quality control tool for high throughput sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). R1 and R2 reads were then merged with FLASH [4] with an overlapping windows of 200 bp and analyzed again with FastQC software.

The average quality for R1 and R2 reads was compared with the average quality of the merged reads and best quality match was chosen for further investigations.

1.3.2. Primer evaluation

Primer choice has a strong effect on the final results, shaping the community and the diversity indexes [5, 6]. Based on the maximum acceptable length of the amplicon for an adequate read accuracy, we performed an *in silico* analysis on different sets of primers. We analyzed primer pairs for 4 different hypervariable regions using the RDP Probe Match online tool (https://rdp.cme.msu.edu/probematch/) allowing one mismatch per primer.

Regions	Primer name	Reference	Sequence (5'-3')	Amplicon length (bp)
V1-2	27 F (8 F)	[8]	AGAGTTTGATYMTGGCTCAG	330
	338R	[9]	GCTGCCTCCCGTAGGAGT	
V1-3	27 F (8 F)	[8]	AGAGTTTGATYMTGGCTCAG	530
	519R (536R)	[10]	GTATTACCGCGGCKGCTG	
V3-4	331F	[11]	TCCTACGGGAGGCAGCAGT	476
	797R	[11]	GGACTACCAGGGTATCTAATCCTGTT	
V4-5	517F	[13]	GCCAGCAGCCGCGGTAA	410
	926R	[13]	CCGTCAATTYYTTTRAGTTT	

Tab. 1.3.1. Primers evaluated in this dissertation.

Primer pair with best hit in the database was chosen for further investigations.

1.3.3. Library preparation

16S hypervariable regions have to be adequately prepared in order to be sequenced on Illumina Miseq. In fact they need to be isolated from bacterial genomes, amplified and tagged with molecular barcodes, specific for every sample. Moreover every amplicon must be added with sequencing primer for SBS initiation. For this reason most 16S high-throughput sequencing protocols are based on PCR amplification [1, 2, 14] and usage of long-tailed primers, adding specific nucleotides necessary for Illumina sequencing technology.

Illumina Two-step PCR protocol

Berry et al. (2011) [14] describe better results reliability by using a double step of PCR rather than a single step. Illumina provides a well-tested protocol for library preparation that consists of two subsequent steps of PCR: the first step amplifies the target region with modified primers while the second step integrates Illumina indices, for library recognition, and sequencing primers. In particular, primers for the first step are modified with a 5' overhang tail that is complementary to the 3' portion of Illumina index primers.

Forward overhang: 5’ TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG- [locus specific sequence]-3'

Reverse overhang: 5’ GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus specific sequence]-3'

Primer construction

Primers chosen for microbial community analysis [11] were added with overhang tails and ordered from Invitrogen with HPLC purification.

Forward primer: Univ16S_BactF

5’-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGTCCTACGGGAGGCAGCAGT-3'

Reverse primer: Univ16S_BactR

5’-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGACTACCAGGGTATCTAATCCTGTT-3'

Primers were evaluated with OligoAnalyzer 3.1 for hetero-dimers formation (https://eu.idtdna.com/calc/analyzer).

PCR Step 1

This step uses PCR to amplify template out of a DNA sample using region of interest-specific primers with overhang adapters attached.

KAPA HiFi HotStart Readymix (2X) (Kapa biosystems) containing DNA polymerase, reaction buffer, dNTPs (0.3 mM each) and MgCl2 (2.5 mM) was used for this step. In each sample the volume of the reaction mixture was equal to 25 ul. The reagents with the relative final concentrations and the cycle used are shown in the tab. 1.3.2 below, and 1.3.3.

Reagents	**Concentration**	**Quantity (µl)**
KAPA *ReadyMix* 2X	1X	12.5
Primer forward 1µM	0.2 µM	5
Primer reverse 1µM	0.2 µM	5
DNA (5 ng/µl)	12.5 ng	2.5

Tab. 1.3.2. Reagents used for PCR1 step.

Temperature	Time	Num. cycles
95°	3 min	1 cycle
95°	30 sec	
55°	30 sec	25 cycles
72°	30 sec	
72°	5 min	1 cycle

Tab. 1.3.3. Cycle for PCR1.

PCR – cleanup

This step uses AMPure XP beads (Beckman Coulter) to purify the 16S amplicon away from free primers and primer dimer species.

20 ul of magnetic beads were aliquoted in each well containing 25 ul of PCR. The content was gently mixed for 10 times and incubated at room temperature for 5 minutes. After incubation the plate was placed on a magnetic stand for 2 minutes after which the supernatant was eliminated, while keeping the plate on the magnetic support, and taking care not to withdraw the beads. 200 ul of 80% ethanol was added and incubated on the magnetic support for 30 seconds and finally removed. A second washing step with 80% ethanol was performed, removing any residual ethanol. The plate was then incubated for about 10 minutes on the magnetic support to allow the beads to dry. After incubation, the plate was removed from the magnetic stand and 52,5 ul of 10 mM Tris pH 8.5 each well was added. The content was mixed gently for 10 times or until when the magnetic beads were fully resuspended. The plate was incubated at room temperature for 2 minutes and then placed on the magnetic support for at least 2 minutes. After this step, 50 ul of supernatant from each well were transferred into a new plate.

PCR Step 2

This step attaches to the 16S amplicon different combinations of dual indices and Illumina sequencing adapters using the Nextera XT Index Kit. Illumina indices are listed in Tab. 1.3.4.

Index 1 (i7)	Sequence	Index 2 (i5)	Sequence
N701	TAAGGCGA	S501	TAGATCGC
N702	CGTACTAG	S502	CTCTCTAT
N703	AGGCAGAA	S503	TATCCTCT
N704	TCCTGAGC	S504	AGAGTAGA
N705	GGACTCCT	S505	GTAAGGAG
N706	TAGGCATG	S506	ACTGCATA
N707	CTCTCTAC	S507	AAGGAGTA
N708	CAGAGAGG	S508	CTAAGCCT
N709	GCTACGCT		
N710	CGAGGCTG		
N711	AAGAGGCA		
N712	GTAGAGGA		

Tab. 1.3.4. List of Illumina indices used during the second step of PCR.

For every sample, the final reaction volume was 50 µl. The amount of reagents is shown in the Tab. 1.3.5.

Reagents	Quantity (µl)
KAPA ReadyMix 2X	25
Nextera XT Index Primer 1 (N7xx)	5
Nextera XT Index Primer 2 (S5xx)	5
DNA	5
H_2O	10

Tab. 1.3.5. List of reagents used for the PCR 2.

This amplification step was performed on the termocycler (*Mastercycler, Eppendorf*) with the following cycle, in tab. 1.3.6:

Time	Num. cycles
3 min.	1 cycle
30 sec. 30 sec. 30 sec.	8 cycles
5 min.	1 cycle

Tab. 1.3.6. Cycle for PCR2 step.

This amplification was followed by another purification step with *AMPure Beads XP* and 25 ul of the eluted sample were transferred in a new plate.

Quantification, normalization and library pooling

Qubit (see paragraph ch. 2.3.4) was performed to quantify the amplification products and the measured concentrations were used to properly dilute the libraries using 10 mM Tris pH 8.5. Finally, 5 ul of each sample were pooled and loaded on MiSeq sequencer with 5% of PhiX.

Library protocol checking: Agilent Bioanalyzer 7500

Nucleic acids samples are introduced in the chip wells, from whom micro-channels depart. During chip preparation, the micro-channels are filled with a total of 27 ul of polymer and a fluorescence dye (5 µl each well), making the chip an electrical circuit. After loading, the chip is shaken 1 min at 1000 rpm and loaded in the cartridge. Charged biomolecules like DNA or RNA are electrophoretically driven by a voltage gradient, similar to slab gel electrophoresis. The molecules are separated by size due to a constant mass-to-charge ratio and to the presence of a sieving polymer matrix. Dye molecules intercalate into DNA strands to be detected by laser-induced fluorescence. Data are translated into gel-like images (bands) and electropherograms (peaks). With the help of a ladder that contains fragments of known sizes and concentrations, a standard curve of migration time versus fragments size is plotted. From the migration times measured for each fragment in the sample, the size is calculated. Two marker fragments (for RNA only one marker fragment) are run with each of the samples bracketing the overall sizing range. The "lower" and "upper" markers are internal standards used to align the ladder data with data from the sample wells [91]. Agilent Bioanalyzer 7500 was used for amplicon length and amplification completeness checking.

1.3.4. Mock community construction

Strain culture

Bacteria strains furnished by BCA – Legnaro labs, were cultured for two days in 5 ml of Tryptic Soy Broth (TSB – Merck Millipore) at 30°C. The different strains are listed in the Tab 1.3.7.

Original taxonomy	Gram
Staphylococcus epidermidis	+
Lysinibacillus fusiformis	+
Paenibacillus sp	+
Salmonella ATCC 14028	-
E.coli K12	-
Pseudomonas 753229	-
Pseudomonas 76_3301	-
Aeromonas 4254	-
Vibrio 150	-
Vibrio 151	-

Tab 1.3.7. Bacteria pooled for mock community construction.

DNA extraction

5 ml of bacteria broth culture were centrifuged at 13500 rpm for 10 minutes. Supernatant was discarded and bacterial pellet was resuspended with 40 ul of proteinase K and 400 ul of *Lysis buffer.* The tubes were then incubated at 56°C shaking for one hour. DNA extraction was performed using *Spin Tissue Mini Kit Invisorb* (Invitek, Berlin, Germany). 200 ul of *Binding Buffer* were added to each tube; after vortexing, the solution was transferred into the silica membrane columns provided by the kit. Centrifugation at 11000 rpm for 2 minutes was performed and the eluted was discarded. 500 ul of Wash buffer were added and columns were centrifuged at 11000 rpm for 1 minute after which the filtrate was again eliminated. The washing step was performed twice. Columns were then centrifuged at maximum speed for 4 minutes in order to remove any residual ethanol. Finally 50 ul of *Elution buffer,* previously heated to 52°C, were added directly on the membrane. The columns were incubated at room temperature for 3 minutes and subsequently centrifuged at 11000 rpm for 2 minutes. This elution step was repeated a second time with an equal volume of *Elution Buffer.*

PCR amplification (Polimerase-Chain Reaction)

Molecular methods are based on the amplification of specific genomic sequences. PCR is an *in vitro* technique through which a DNA molecule is amplified exponentially by means of a DNA-dependent DNA-polymerase. DNA to be amplified is added to a mix of reagents (primers, dNTPs, DNA polymerase, buffer, magnesium salt) that, under appropriate conditions, trigger the polymerization reaction. The reaction is enhanced by temperature changes that allow DNA to denature, primer to anneal and strands to elongate. These three

steps have different characteristics:

- denaturation: DNA solution is brought to a temperature between 94°C and 99 °C to allow the DNA strands to separate;
- annealing: 40-60°C in order to allow annealing of primers to their complementary regions in the DNA strands;
- elongation: the temperature is raised up to 65-72°C in order to maximize the action of Taq polymerase. Using the single-stranded DNA as template the enzyme elongates the primers. In this phase new DNA strands are produced.

Some conditions are variable:

- annealing temperature: it varies depending on the type of primer used (length, GC content) and the desired stringency;
- elongation time: it depends on the length of the fragment and the throughput of the enzyme.

These steps are commonly repeated for about 20-35 times. Generally 50 cycles are not exceeded because amplicon production reaches the *plateau* for lack of primer or dNTPs. PCR can be used as a method of identification (i.e. Species-specific PCR) or typing (i.e. Multi-locus Sequence-Typing).

In this study we used PCR for V3-V4 16S amplification and sequencing of mock community bacteria using the primers from Nadkarni et al. 2002 [11] listed in the Tab. The amplification step was performed at *BMR Genomics* with internal protocols.

DNA quality control: electrophoresis on agarose gel

Electrophoresis is based on the migration of nucleic acids in an agarose or acrylamide matrix, under the impulse of an electric field. In fact nucleic acids that are negatively charged, due to the presence of phosphate groups, migrate towards the cathode if subjected to a potential difference. The choice of agarose concentration is due to the size of the nucleic acid to be separated, thus genomes were usually run in agarose 1% while PCR products in agarose 1,5%. Gel was prepared using a buffer of low ionic strength, such as the TAE, also used as the electrophoresis running buffer and is composed of 2M Tris, glacial acetic acid 1M and 0.1M EDTA. The use of SYBR® Safe (Invitrogen), intercalating nucleic acids, allows DNA visualization. Agarose (Euroclone) was dissolved in 1X TAE, added with SYBR® Safe (Invitrogen) at a concentration of 0.1 mg/ml. After solidification, the gel was introduced into the electrophoresis tray and covered with TAE 1X. The loading of the samples was carried out by combining 3 ul of PCR product and 3 ul

of Loading Dye Buffer 2X (Thermo Scientific). 100 bp or 1 Kb molecular weight marker was run with all samples (GeneRuler 1kb DNA Ladder Plus or GeneRuler 100bp DNA Ladder Plus – Codisan). The agarose gel was run at 100V for about 20 minutes, then it was exposed to ultraviolet radiation on transilluminator (Explera) in order to display nucleic acids and record the photo. Agarose gel electrophoresis was exploited to assess successful extraction or amplification of genomic DNA or PCR products.

Sanger sequencing

The sequencing reaction is an application of the modified PCR. In the reaction mixture just one primer and two different types of nucleotides are present (normal and fluorescently marked di-deoxynucleotides). The taq polymerase catalyzes the attach of both normal nucleotides and modified, randomly. The incorporation of the di-deoxynucleotides blocks the elongation of the DNA: in this way the PCR reaction products a large number of amplicons, differentiated one each other by length. This type of analysis has been conducted on PCR from extracted genomes of pure bacteria cultures. The amplification of the 16S rRNA region and sequencing reactions were carried out at *BMR genomics* with the ABI 3730XL sequencer (Applied Biosystems).

Real-Time PCR

This kind of PCR allows to amplify DNA and quantify its copy number at the same time. DNA amount is directly linked with fluorescence of an intercalating agent (SYBR Green) that absorbs light at blue wavelength and emits light at green one. SYBR Green is inserted in dsDNA helix during amplification, so that through every cycle the increase of fluorescence is measured. Finally for every sample a Ct value is assigned: it indicates the cycle number in which the fluorescence exceeded a specified threshold.

In microbiology Real-Time PCR is commonly used for bacteria quantification or for marker-gene expression quantification. In this study real-time PCR was used for amplification of V3-V4 hypervariable regions of every bacterium to be inserted in the mock community with primer from Nadkarni et al., (2002) [11] listed in the Tab 3.1. Bacterial DNA was diluted 500 to 1500 folds depending on estimated concentration on agarose. *Lysinibacillus* DNA was serially diluted 1:10 folds to construct the calibration line for Ct comparison. All reactions were performed in a volume of 10 ul with Sybr Fast ABI Prism Readymix Kit and volumes of all reagents are listed in Tab. 1.3.8.

Reagents	Quantity (ul)
Mastermix	5 ul
Forward Primer (10 uM)	0,5 ul
Reverse Primer (10 uM)	0,5 ul
Water	1,5 ul
DNA	2,5

Tab. 1.3.8. Reagents for RealTime PCR for mock community construction.

Amplification step was performed on 7900HT Fast (Applied Biosystems) with cycle in tab. 1.3.9.:

Temperature	Time	Cycle
50°	2 min	1 cycle
95°	10 min	1 cycle
95°	10 sec	45 cycles
60°	60 sec	

Tab. 1.3.9. Cycle for RealTime PCR amplification.

1.3.5. Bioinformatic analysis: Qiime

QIIME is a comprehensive wrapper that provides many different tools for microbial communities analysis. Many of the tools and scripts that are described in the following paragraphs are used with default settings, modifications are described.

1.3.6. Upstream analysis

Preprocessed sequences were joined into a single file in *.fna* format, suitable for QIIME pipeline, and a *mapping file* with metadata info, such as sample ID (name), barcode (indices), and grouping was created. The mapping file in tab-delimited format is reported in the tab 1.3.10 below.

#SampleID	BarcodeSequence	LinkerPrimerSequence	Treatment	Reverseprimer	Description
ID30741	CGTACTAGAGAGTAGA	TCCTACGGGAGGCAGCAGT	TQ	TCCTACGGGAGGCAGCAGT	MITQ1
ID30742	AGGCAGAAAGAGTAGA	TCCTACGGGAGGCAGCAGT	TQ	TCCTACGGGAGGCAGCAGT	MITQ2
ID30743	TCCTGAGCAGAGTAGA	TCCTACGGGAGGCAGCAGT	TQ	TCCTACGGGAGGCAGCAGT	MITQ3
...	...	...	...	...	...

Tab. 1.3.10. Typical mapping file architecture.

Within QIIME 1.8.0 version we applied two different OTU-picking strategies (*de novo* and closed-reference) in order to evaluate them on our mock community for subsequent analysis. As reference database we used Greengenes database version 13_8 [22, 26].

The *de-novo* approach groups sequences based on sequence identity. We used the default wrapper-command:

```
pick_de_novo_otus.py -i $PWD/seqs.fna -o $PWD/uclust_otus/
```

to start the analysis on the preprocessed *seqs.fna* (-i). With this strategy we:

1. clustered sequences in OTUs based on 97% of similarity using *uclust* algoritm [17];
2. picked a representative sequence for each OTU (the OTU centroid sequence);
3. aligned the representative set with PyNAST [18];
4. assigned taxonomy with the *uclust* consensus taxonomy assigner;
5. filter the alignment prior to tree building, removing positions which are all gaps, and specified as 0 in the lanemask; build a phylogenetic tree with FastTree [19];
6. build an OTU table.

All output files were written to the directory specified by -o. $PWD code represent absolute path.

The *closed-reference* approach clusters reads against a reference database. We used the command:

```
pick_closed_reference_otus.py -i $PWD/seqs.fna -r
$PWD/refseqs.fna -o $PWD/otus/
```

to start the analysis on the preprocessed *seqs.fna* (-i). With this strategy we clustered sequences in OTUs using *uclust* algoritm [17] based on 97% of similarity with the reference database. As there is not a wrapper algorithm for this strategy, we used the commands below for subsequent analysis following *de-novo* order list:

2. `pick_rep_set.py`,
3. `parallel_align_seqs_pynast.py`,
4. `assign_taxonomy.py`,
5. `make_phylogeny.py`,
6. `make_otu_table.py`

Chimera detection and filtering

During the PCR amplification process, some of the amplified sequences can be produced from multiple parent DNA molecules, generating sequences known as chimeras. These sequences are technical artifacts and could influence final community composition. QIIME supports different algorithms but the recommended method for identifying chimeric sequences is UCHIME [21]. We performed the command:

```
identify_chimeric_seqs.py -m usearch61 -i seqs.fna -r
ref_sequences.fasta -o usearch61_chimera_checking/
```

on sequences undergoing the *de-novo* approach, as the *closed-reference* approach is not biased by this problem [22]. Chimera checking results a *chimeras.txt* file that contains the ID of the reads that must be discarded: this file is used in the following script to filter the total reads:

```
filter_fasta.py -f seqs.fna -o seqs_chimeras_filtered.fna -s
usearch61_chimera_checking/chimeras.txt -n
```

where the -n option specifies that chimeras in the .txt file should be removed.

1.3.7. Downstream analysis

OTU abundance quality filtering

Once OTU table and the phylogenetic tree are generated, the downstream analysis can start. At this point, Bokulich et al. (2013) recommend performing a second level of quality-filtering based on OTU abundance [20]. The suggested procedure is to discard those OTUs with a number of sequences <0.005% of the total number of sequences. With the script:

```
filter_otus_from_otu_table.py --min_count_fraction 0.00005
```

it is possible to reduce the problem of spurious OTUs, most of which are present at very low abundance.

Taxa summary

To visualize OTUs `summarize_taxa.py` command was used, which summarizes in a spreadsheet the relative abundance of the *taxa* present in a set of samples on multiple taxonomic levels (e.g., phylum, order, etc.) [15].

Alpha-diversity and Beta-diversity analysis

Alpha-diversity is defined as the diversity of organisms in one sample or environment while Beta-diversity is the difference in diversities across samples or environments. QIIME implements dozens of the most widely used alpha-diversity indices, including both

phylogenetic indices (which require a phylogenetic tree) and non-phylogenetic indices [15]. We used `alpha_rarefaction.py` and `alpha_diversity.py` commands respectively for sample rarefaction plotting and diversity indices evaluation (observed_species, PD_whole_tree, chao1, goods_coverage, shannon).

QIIME calculates the beta-diversities between each pairs of input samples, forming a distance matrix. The distance matrix then can be visualized with methods such as principal coordinates analysis (PCoA) [23] and hierarchical clustering [24]. We used `beta_diversity_through_plots.py` wrapper for:

1. Randomly subsample otu_table.biom to even number of sequences per sample;

2. Compute a weighted and unweighted unifrac distance matrices;

3. Peform a PCoA on the result of Step 2;

4. Generate a 2D and 3D plots for all mapping fields.

Statistical analysis

QIIME implements several scripts that perform a broad range of statistical tests between samples and groups of samples using both alpha- and beta-diversity measurements. We used these analysis for comparisons of mock samples within the same run and across different runs, based on beta-diversity testing.

For beta-diversity comparisons we used `beta_significance.py` command on weighted unifrac: this script returns a tab delimited text file with each pair of samples and a p value representing the probability that a random sample/sequence assignment will result in more dissimilar samples than the actual pair of samples. The command `make_distance_boxplots.py` was used to generate a box-plot graph and to perform a t-test over sample categories. Another useful tool is the Procrustes analysis that compares spatial maps by transforming them and maximizing a measure of the similarity of the transformed maps [98]. Procrustes analysis is performed in QIIME with the script `transform_coordinate_matrices.py` and is followed by a script that creates the plot, `make_emperor.py`. Finally, the code `compare_categories.py` can be used for multivariate analysis between sample grouping because it provides additional powerful tools for exploring significant relationships between the beta-diversity distance matrix and factors or covariates where ANOSIM and adonis are usually employed [15].

1.4. Results

1.4.1. Library construction checking

First and second PCR steps were checked for amplicon length and amplification effectiveness on Bioanalyzer chip 7500. The first step amplified about 550 bp, fig. 1.4.1. The second PCR produced an amplicon about 620 bp long: increase in amplicon length of about 60 bp was considered acceptable, fig. 14.2, while 30 bp or few bases shift would not be accepted as the amplicon would be only partially amplified.

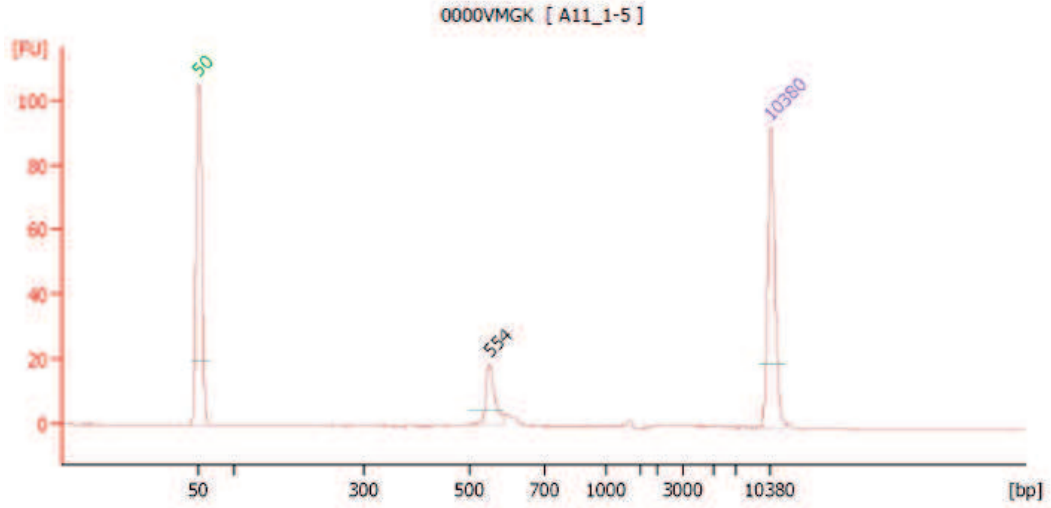

Fig. 1.4.1. Agilent Bioanalyzer 7500 gel representing the length of the amplicon produced during the PCR1 step. About 60 bp are added at the original amplicon length due to the primer tails.

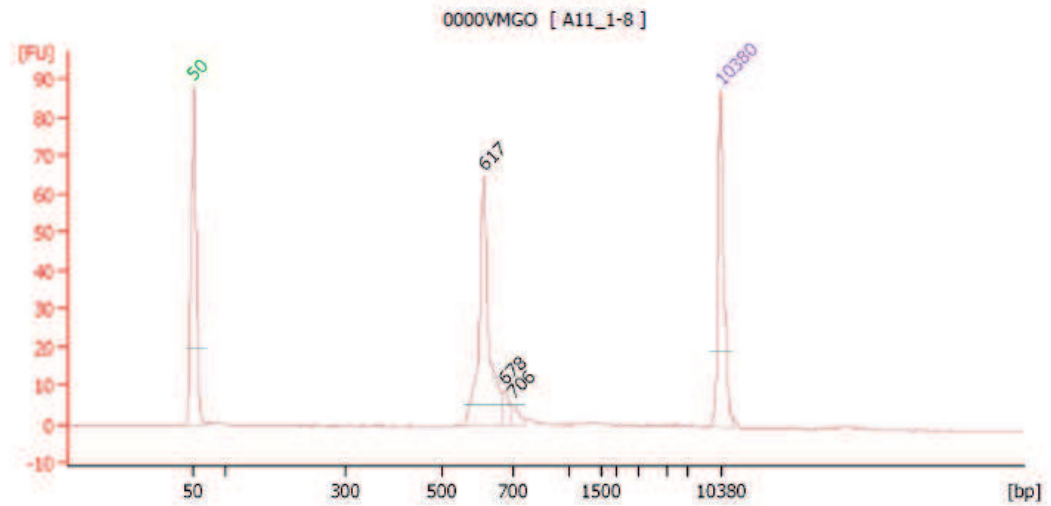

Fig. 1.4.2. Agilent Bioanalyzer 7500 gel representing the length of the amplicon produced during the PCR 2 step. About 60 bp are added at the original amplicon length due to the index tails.

Primer "universality" assessment

Each primer pair showed different effectiveness for matching bacteria sequences in RDP database. Allowing a mismatch, we saw that the fewest number of strains was recognized by V1-V2 (458443 sequences) and V1-V3 (489112) primers while the highest number was found by V3-V4 primers (1812035), fig. 1.4.3. Following these results we chose V3-V4 primers for further investigations.

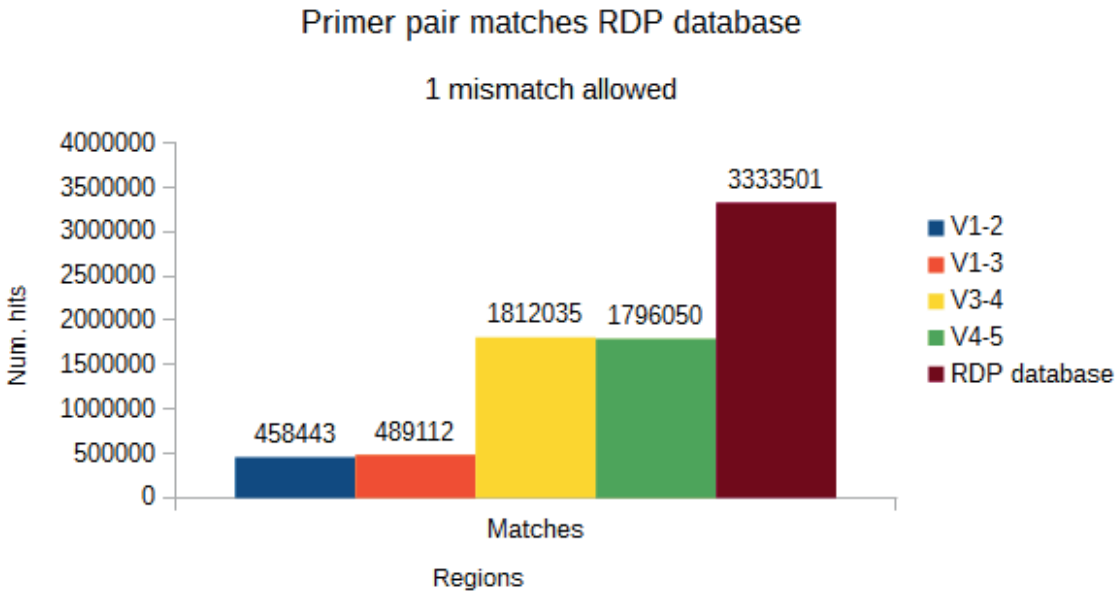

Fig. 1.4.3. Matching efficency of the primers against RDP database.

The chosen V3-V4 primers were fused with overhang tails for Illumina adapters and showed a ΔG = -7,58 kcal/mole for hetero-dimers formation. On the other hand the other fused primers showed even lower ΔG (-9 to -11 kcal/mole), suggesting a more difficult usage in PCR.

1.4.2. Sequencing approach evaluation

In the first Miseq runs we obtained different percentages of reads with Q>30 of Phred scores. Run values are reported in tab. 1.4.1:

	% >= Q30	
Level	**Run (19-02-14)**	**Run (19-05-14)**
R1	73.7	76.99
Index 1	76.03	74.01
Index 2	86.50	81.05
R2	58.23	45.89

Tab. 1.4.1. Features of the first two runs on Miseq.

The average quality of single end reads per sample was different between R1 and R2 reads, usually with R1 reads showing higher average quality (Fig. 1.4.3 e 1.4.4).

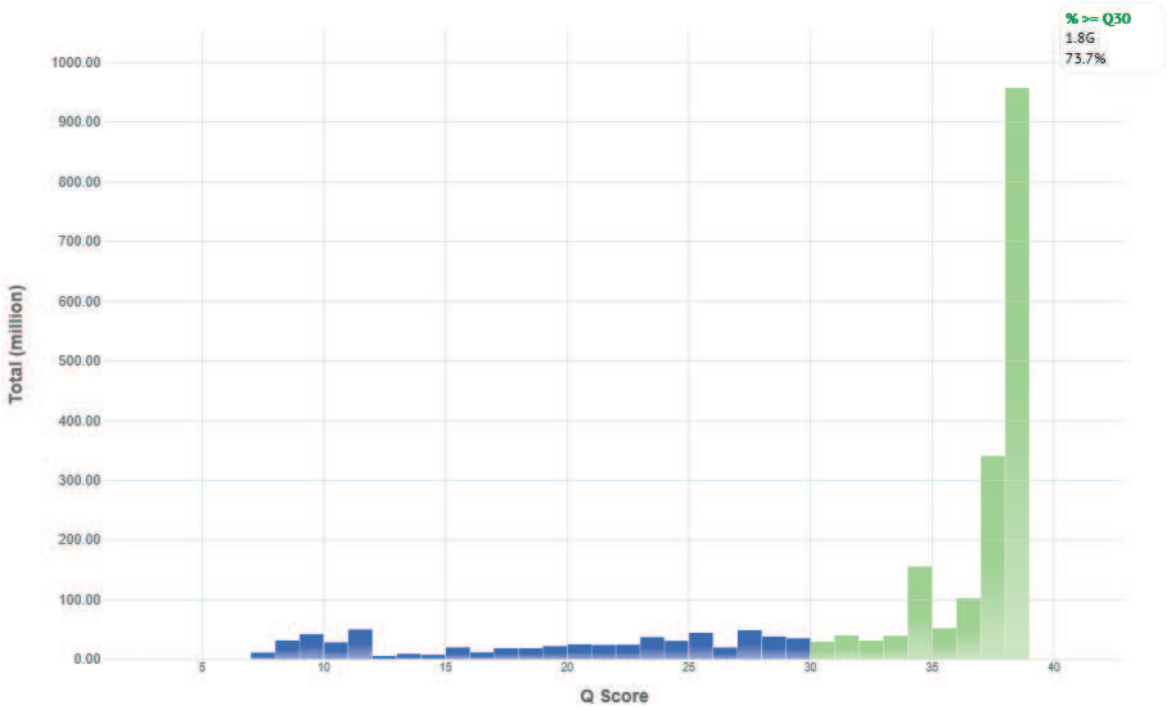

Fig. 1.4.3. Quality score map of the run 19-02-14 on Miseq. 73,7% of the forward reads are above the Q30.

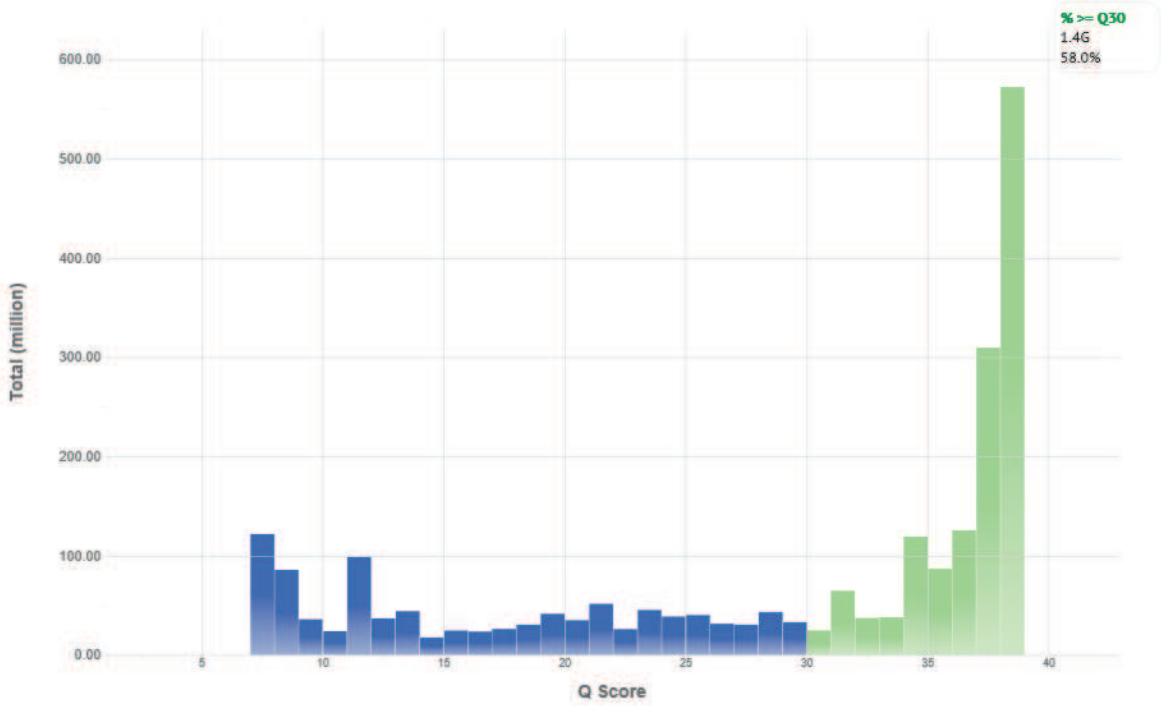

Fig. 1.4.4. Quality score map of the run 19-02-14 on Miseq. 58% of the reverse reads are above the Q30.

Moreover the per base quality decreased along the read, spanning from 38 to 28 of Phred score in about the first 195-200 bp, falling quickly to an average value of 14 after this threshold (fig. 1.4.5).

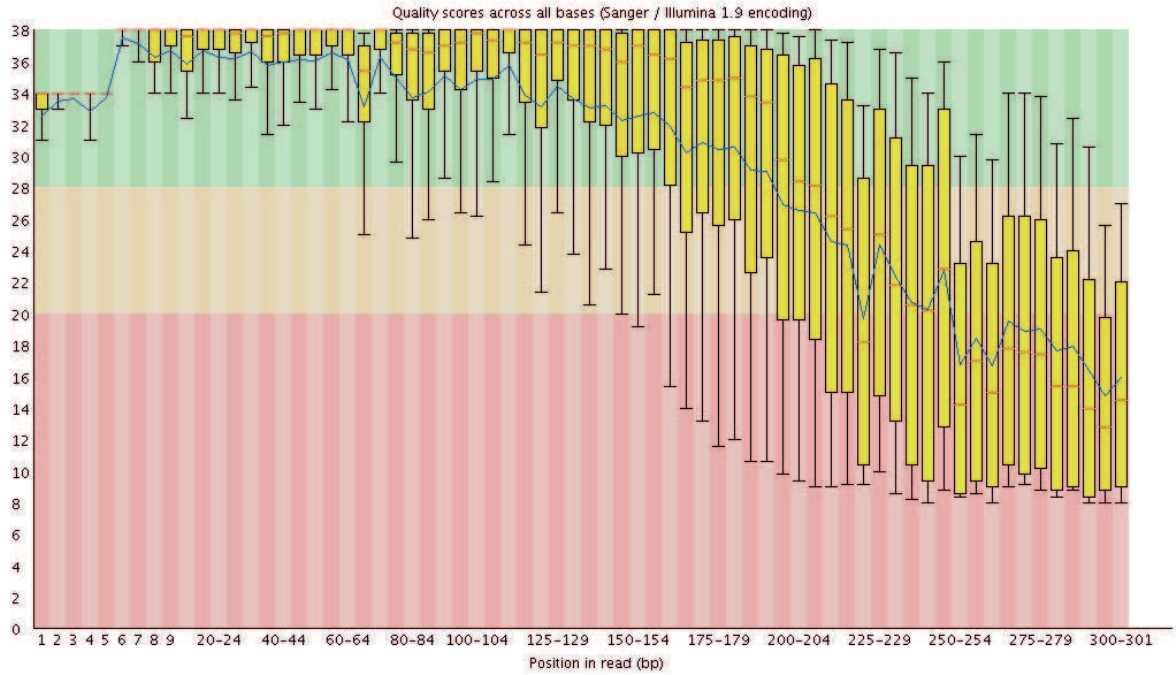

Fig. 1.4.5. FastQC report of the per-base quality of the forward reads of MiSeq run. X axis values are the position of the base (grouped by 5 positions) while y axis shows the Phred score.

After the merge of R1 and R2 reads, FastQC statistics on the reads showed that average per-base quality was increased (fig. 1.4.6) and the middle part of the contig (from 170 bp to 320bp) had an average higher score, around 28. For this reason we chose to follow the paired end sequencing approach with a minimum overlap of 100 bp.

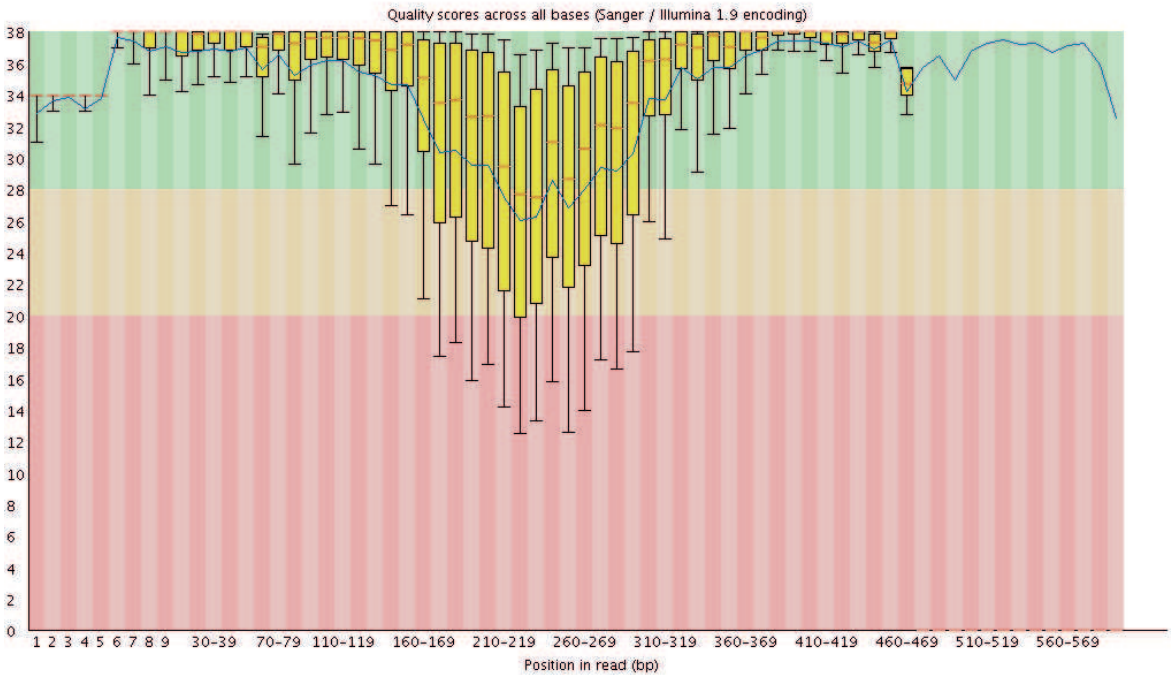

Fig. 1.4.6. FastQC report of the per-base quality of the merged reads. X axis values are the position of the base (grouped by 5 positions) while y axis shows the Phred score.

1.4.3. Mock community construction

The DNA extracted from cultured pure bacteria were verified on agarose 1% and sequenced at *BMR Genomics* with 16S Sanger protocol. Resulting reads were compared with RDP and Greengenes databases. Identifications are listed in tab. 1.4.2.

Original typing	**Species Greengenes database**	**Species RDP database**
Staphylococcus epidermidis	*Staphylococcus epidermidis*	*Staphylococcus epidermidis*
Lysinibacillus fusiformis	*Lysinibacillus fusiformis (Bacillus sp.)**	*Lysinibacillus fusiformis*
Paenibacillus sp	*Paenibacillus sp. (Bacillus sp.)**	*Paenibacillus sp*
Salmonella ATCC 14028	*Salmonella sp.*	*Salmonella enterica*
E.coli K12	*Escherichia coli (Shigella)**	*Escherichia coli*
Pseudomonas 753229	*Pseudomonas sp.*	*Pseudomonas sp.*
Pseudomonas 76_3301	*Pseudomonas sp.*	*Pseudomonas fluorescens*
Aeromonas 4254	*Aeromonas sp.*	*Aeromonas sp*
Vibrio 150	*Vibrio sp.*	*Vibrio sp.*
Vibrio 151	*Vibrio sp.*	*Vibrio sp.*

Tab. 1.4.2. identification of mock bacteria with Sanger approach and comparison against Greengenes Database and RDP database. *Genera in brackets are identified at the same level of identity of the accepted genus.

Measured with the *Qubit*, the concentration of *Lysinibacillus* DNA was 3 ng/ul and it was used to construct the calibration line by 10 fold serial dilutions. Results from Real-Time PCR are listed in the table 1.4.3 below.

Bacterium (Genus)	**Dilution**	**Ct mean**	**Final conc. ng/ul**	**Ratio for pooling**
Staphylococcus	1:500	15,79	14,35	3,73
Paenibacillus	1:500	14,23	43,41	1,23
Salmonella	1:1500	17,39	15,14	3,54
Escherichia	1:1000	16,56	17,72	3,02
Pseudomonas 75	1:1000	16,57	17,97	2,98
Pseudomonas 76	1:1000	16,63	17,31	3,1
Aeromonas	1:1000	18,65	4,45	12,05
Vibrio 150	1:1000	14,94	53,6	1
Vibrio 151	1:1000	17,25	13,16	4,07
Lysinibacillus	STD	-	3	17,87

Tab. 1.4.3. Results from RealTime PCR performed on different dilution of bacteria of the mock community.

Concentration ratio was used to measure the adequate volume for each bacterium DNA to be mixed for mock community construction. The final pool was amplified with primers for V3-V4 hypervariable regions. The mock community was then processed with Two-step PCR Illumina protocol for library preparation.

This Mock was used to assess:

1. OTU-picking strategy effect on community composition and structure;
2. even community composition after sequencing and taxonomy assignment correctness;
3. repeatability and reliability of the library protocol within the same run and across different runs;
4. DNA amount effect on final community structure.

Thus we loaded 6 Mock communities on the same Miseq run. Three of them had a PCR concentration of 5 ng (q29944, q29945, q29946) and 3 had a concentration of 20 ng

(q29947, q29948, q29949). We analyzed them with different approaches. One more mock (q27946) was then loaded in another Miseq run and joined for the final analysis about method reproducibility across different runs.

We obtained 784,674 row reads for the first run and 90,796 row reads for the second. After FLASH merge and quality filtering we obtained 407,658 reads.

1.4.4. QIIME OTU-picking strategy effects on the Mock community

In order to evaluate the effect of the OTU-picking algorithm on the final results and to understand when to use one or the other approach, we analyzed 7 mock communities replicates with *de-novo* clustering method and *closed-reference* method.

Community structure with closed-reference strategy

Exploiting the closed-reference approach for OTU-picking strategy, we obtained 218 OTUs with a failure on database matching of 15712 putative OTUs. After OTU filtering step [20] we obtained 78 OTUs, better representing the original mock. We assigned taxonomy to the OTUs at the genus level using the Greengenes database 13_8 version. After taxonomy summary we obtained the following 8 families and 9 genera.

k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae;g__Bacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Lysinibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__Plesiomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Pseudoalteromonadaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Pseudoalteromonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio

The table 1.4.4 below shows the relative abundance of families across the samples while the tab. 1.4.5 shows the relative abundance of genera across samples. The relative abundance of the different samples are represented in the fig. 1.4.7.

Tab. 1.4.4. Families relative abundance with closed-reference approach.

Taxon (Family)	**q29949**	**q29948**	**q29945**	**q29947**	**q29946**	**q29944**	**q27946**	**Exp**
Staphylococcaceae	12,49%	12,28%	13,48%	13,76%	14,91%	13,00%	12,33%	10%
Planococcaceae	13,97%	13,84%	14,70%	14,19%	14,31%	17,23%	16,75%	10%
Paenibacillaceae	4,65%	4,85%	4,75%	4,36%	3,57%	4,52%	6,02%	10%
Enterobacteriaceae	19,14%	19,15%	18,05%	17,84%	18,39%	19,48%	18,46%	20%
Pseudomonadaceae	14,52%	15,30%	15,67%	15,64%	14,54%	9,49%	11,69%	20%
Aeromonadaceae	7,83%	8,21%	8,16%	8,00%	8,38%	5,37%	7,53%	10%
Vibrionaceae	19,09%	18,87%	18,35%	18,18%	17,81%	21,62%	19,65%	20%
Other	8,32%	7,50%	6,84%	8,03%	8,08%	9,29%	7,57%	0%

Tab. 1.4.5. Genera relative abundance with closed-reference approach.*Genus is shown for those OTUs for which it was present.

Taxon (Genus)*	**q29949**	**q29948**	**q29945**	**q29947**	**q29946**	**q29944**	**q27946**	**Exp**
Staphylococcus	12,49%	12,28%	13,48%	13,76%	14,91%	13,00%	12,33%	10%
Lysinibacillus	13,97%	13,84%	14,69%	14,19%	14,30%	17,23%	16,74%	10%
Paenibacillus	4,65%	4,85%	4,75%	4,36%	3,57%	4,52%	6,02%	10%
Enterobacteriaceae	18,97%	19,00%	17,89%	17,68%	18,21%	19,32%	18,33%	20%
Pseudomonas	14,52%	15,30%	15,67%	15,64%	14,54%	9,49%	11,69%	20%
Aeromonadaceae	7,83%	8,21%	8,16%	8,00%	8,38%	5,37%	7,53%	10%
Vibrio	19,09%	18,87%	18,35%	18,18%	17,81%	21,62%	19,65%	20%
Other	8,49%	7,66%	7,01%	8,19%	8,27%	9,46%	7,70%	0%

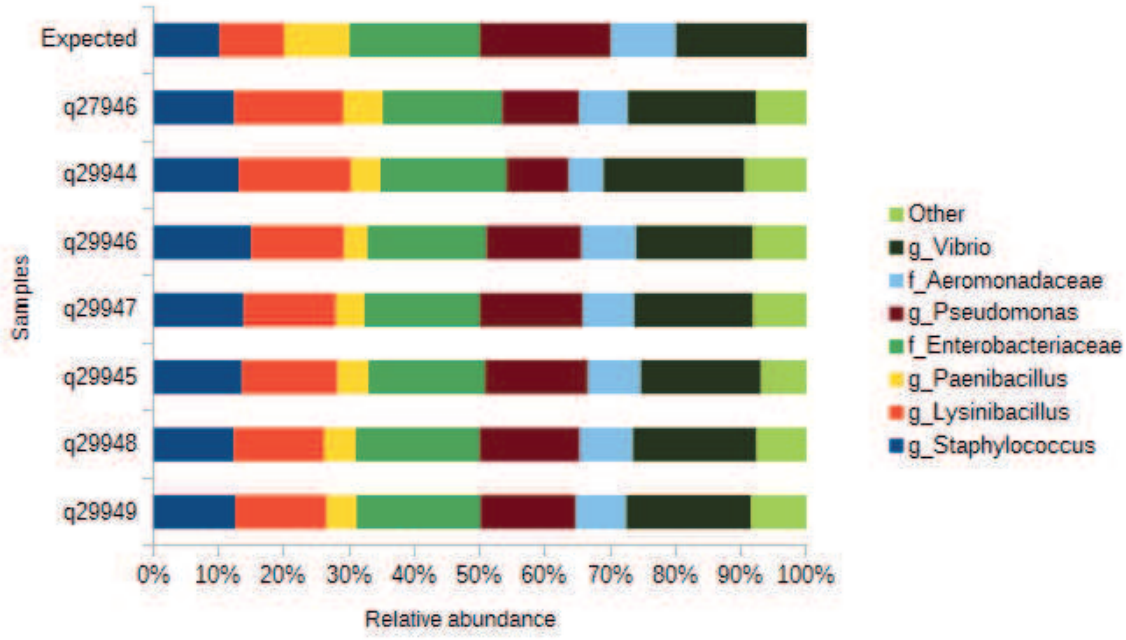

Fig. 1.4.7. Comparison between expected and observed relative abundances and taxa identification across samples.

Every family was represented consistently across the samples, with a standard deviation from mean value spanning from a minimum of 0,62% to a maximum of 2,35%, fig. 1.4.8. The mean imbalance from the expected abundance was comprised between 0,92% (*Vibrio*) and 6,62% (*Pseudomonas*).

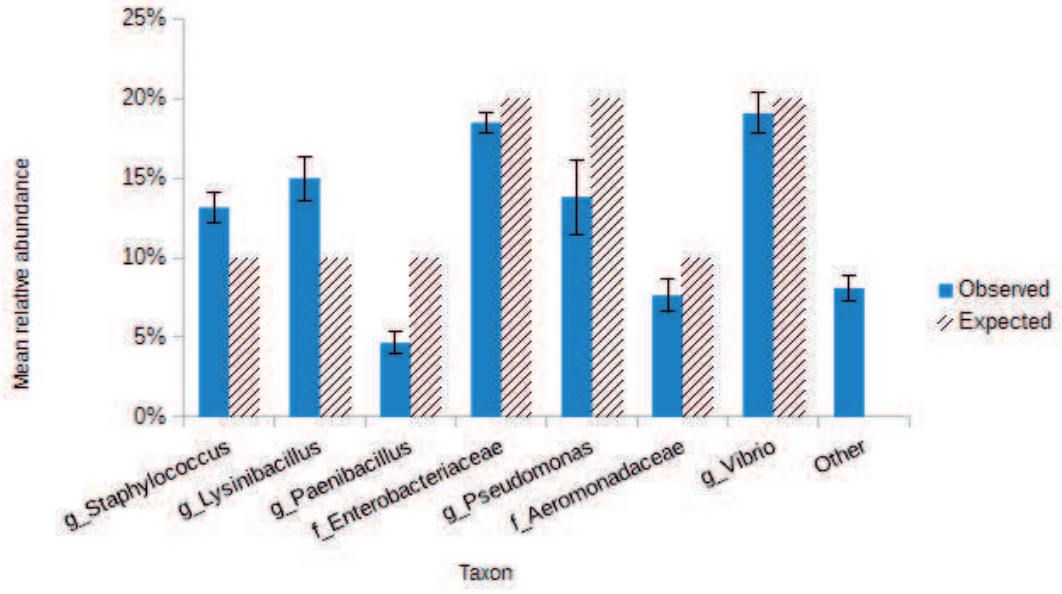

Fig. 1.4.8. Mean relative abundance for every taxon in comparison with its expected value. Vertical lines represent standard deviation.

Mislabelling or spurious OTUs were clustered into the “Other” group, showed in tab. 1.4.6. This group showed a high relative abundance of the *Pseudoalteromonadaceae* family (*Alteromonadales* order, *Gammaproteobacteria* class) that in Greengenes is wrongly assigned to the *Vibrionales* order. The other spurious families correctly belonged to *Gammaproteobacteria* and to *Bacilli*, the two classes dividing the chosen mock bacteria.

Taxon (Family)	q29949	q29948	q29945	q29947	q29946	q29944	q27946
Pseudoalteromonadaceae	8,24%	7,43%	6,70%	7,96%	7,97%	9,23%	7,49%
Moraxellaceae	0,07%	0,06%	0,12%	0,05%	0,08%	0,05%	0,08%
Streptococcaceae	0,01%	0,00%	0,01%	0,02%	0,01%	0,00%	0,00%
Bacillaceae	0,00%	0,01%	0,01%	0,00%	0,01%	0,01%	0,00%

Tab. 1.4.6. Unexpected taxa grouped into the “Other” cluster.

Alpha-diversity indices for the different replicates at a rarefaction depth of 25000 reads are reported in tab. 1.4.7.

Sample	observed_species	PD_whole_tree	shannon
q29949	72	3,02	4,22
q29944	71	2,92	4,14
q29948	75	2,96	4,24
q29945	73	2,95	4,22
q29947	70	3,01	4,22
q29946	76	3,10	4,24
q27946	73	2,97	4,26

Tab. 1.4.7. Alpha-diversity indices: observed_species (OTU), Phylogenetic diversity (PD) and Shannon index.

Community structure with de-novo strategy

After chimera checking step we obtained 7092 ref_chimeras (reference-based chimeras) and 2076 denovo_chimeras (abundance-based chimeras). Thus, after chimeras elimination from total reads, we performed the *de-novo* OTU-picking step with which we obtained 3356 OTUs. With OTU filtering step [20] we obtained 196 OTUs, better representing the original mock. After taxonomy assignment with the Greengenes database 13_8 version and taxonomy summary, we obtained 11 families and 7 genera. With this approach the unassigned reads appeared.

Unassigned;Other;Other;Other;Other;Other
k__Bacteria;p__Bacteroidetes;c__Bacteroidia;o__Bacteroidales;f__Prevotellaceae;g__Prevotella
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Paenibacillaceae;g__Paenibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;Other
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Planococcaceae;g__Lysinibacillus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Staphylococcaceae;g__Staphylococcus
k__Bacteria;p__Firmicutes;c__Bacilli;o__Lactobacillales;f__Streptococcaceae;g__Streptococcus
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;Other;Other;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Aeromonadales;f__Aeromonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Enterobacteriales;f__Enterobacteriaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Pseudomonadaceae;g__Pseudomonas
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;Other;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Pseudoalteromonadaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Pseudoalteromonadaceae;g__
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;Other
k__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Vibrionales;f__Vibrionaceae;g__Vibrio

The table 1.4.8 below shows the relative abundance of families across the samples while the tab. 1.4.9 shows the relative abundance of genera across samples. The relative abundance of the different samples are represented in the fig. 1.4.9.

Taxon (Family)	**q29948**	**q29947**	**q29946**	**q29945**	**q29944**	**q29949**	**q27946**	**Exp**
Staphylococcaceae	12,15%	13,58%	14,71%	13,36%	12,88%	12,33%	12,08%	10%
Planococcaceae	13,69%	14,01%	14,16%	14,57%	17,06%	13,82%	16,52%	10%
Paenibacillaceae	4,83%	4,32%	3,58%	4,69%	4,47%	4,62%	5,94%	10%
Enterobacteriaceae	18,88%	17,58%	18,25%	17,83%	19,20%	18,91%	18,22%	20%
Pseudomonadaceae	15,25%	15,61%	14,40%	15,69%	9,50%	14,48%	11,65%	20%
Aeromonadaceae	8,31%	8,15%	8,46%	8,15%	5,41%	7,88%	7,43%	10%
Vibrionaceae	15,46%	15,61%	15,29%	14,63%	17,54%	16,09%	16,19%	20%
Other	11,44%	11,14%	11,16%	11,09%	13,94%	11,87%	11,97%	0%

Tab. 1.4.8. Families relative abundance with *de-novo* approach.

Taxon (Genus)*	q29948	q29947	q29946	q29945	q29944	q29949	q27946	Exp
Staphylococcus	12,15%	13,58%	14,71%	13,36%	12,88%	12,33%	12,08%	10%
Lysinibacillus	13,69%	14,00%	14,15%	14,55%	17,05%	13,81%	16,50%	10%
Paenibacillus	4,83%	4,32%	3,58%	4,69%	4,47%	4,62%	5,94%	10%
Enterobacteriaceae	18,88%	17,58%	18,25%	17,83%	19,20%	18,91%	18,22%	20%
Pseudomonas	15,23%	15,57%	14,36%	15,62%	9,47%	14,46%	11,57%	20%
Aeromonadaceae	8,31%	8,15%	8,46%	8,15%	5,41%	7,88%	7,43%	10%
Vibrionaceae	15,46%	15,61%	15,29%	14,63%	17,54%	16,09%	16,19%	20%
Other	11,46%	11,18%	11,22%	11,18%	13,97%	11,90%	12,07%	0%

Tab. 1.4.9. Genera relative abundance with *de-novo* approach.*Genus is shown for those OTUs for which it was present.

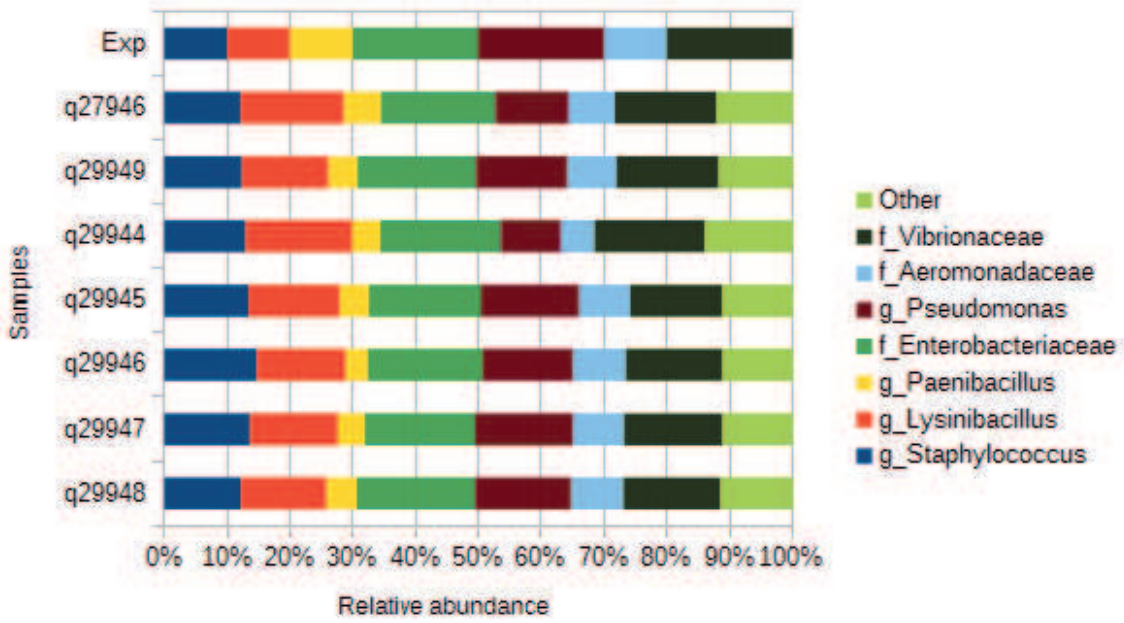

Fig. 1.4.9. Comparison between expected and observed relative abundances and taxa identification across samples.

Every family was represented consistently across the samples, with a standard deviation from mean value spanning from a minimum of 0,60% to a maximum of 2,34%, fig. 1.4.10. The mean imbalance from the expected abundance was comprised between 1,59% (*Enterobacteriaceae*) and 6,25% (*Pseudomonas*). Generally shift from expected values was higher for *de-novo* approach.

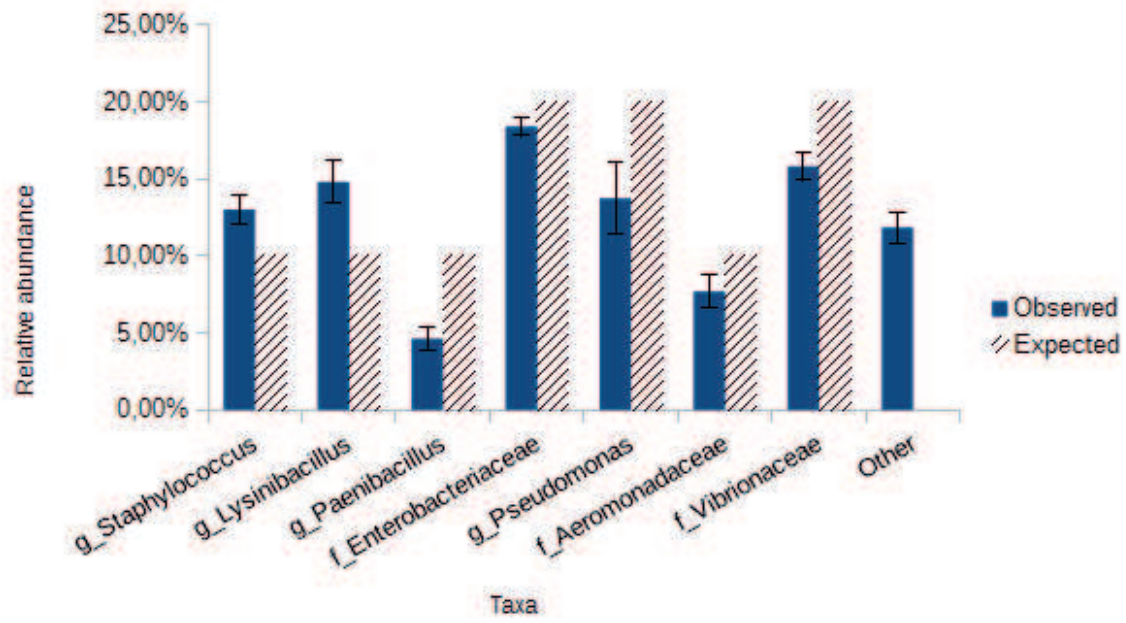

Fig. 1.4.10. Mean relative abundance for every taxon in comparison with its expected value. Vertical lines represent standard deviation.

Unassigned or spurious OTUs were clustered into the "Other" group, showed in tab. 1.4.10. This group showed again a high relative abundance of the *Pseudoalteromonadaceae* family. The other spurious families correctly belonged to *Gammaproteobacteria* and to *Bacilli*, the two classes dividing the chosen mock bacteria. However with this approach another class is found, *Bacteroidia* (for *Prevotellaceae* family) and 0,01% to 0,05% of reads fail to be identified (unassigned).

Taxon	**q29948**	**q29947**	**q29946**	**q29945**	**q29944**	**q29949**	**q27946**
Unassigned	0,01%	0,01%	0,02%	0,02%	0,02%	0,01%	0,05%
f__Pseudoalteromonadaceae	11,29%	10,99%	10,96%	10,90%	13,79%	11,68%	11,72%
o__Vibrionales	0,06%	0,06%	0,06%	0,03%	0,07%	0,07%	0,06%
o__Aeromonadales	0,05%	0,03%	0,04%	0,08%	0,04%	0,04%	0,11%
c__Gammaproteobacteria	0,01%	0,00%	0,03%	0,03%	0,01%	0,01%	0,02%
f__Prevotellaceae	0,01%	0,02%	0,01%	0,00%	0,00%	0,04%	0,00%
f__Streptococcaceae	0,00%	0,02%	0,01%	0,01%	0,00%	0,01%	0,00%
f__Moraxellaceae	0,00%	0,00%	0,01%	0,02%	0,01%	0,00%	0,02%

Tab. 1.4.10. Unexpected taxa grouped into the "Other" cluster.

Alpha-diversity indices for all replicates at a rarefaction depth of 25000 reads are reported in tab 1.4.11.

Sample	observed_species	PD_whole_tree	shannon
q29948	168	6,11	3,80
q29947	168	6,13	3,80
q29946	187	6,42	3,86
q29945	179	5,59	3,84
q29944	174	5,26	3,78
q29949	167	6,10	3,82
q27946	184	5,49	3,92

Tab. 1.4.11. Alpha-diversity indices: observed_species (OTU), Phylogenetic diversity (PD) and Shannon index.

De-novo and closed-reference approaches

De-novo approach shows an inter-samples variability in taxa abundance that spans from 0,6% (for f_*Enterobacteriaceae*) to 2,34% (for g_*Pseudomonas*) from mean value, and this pattern is similarly repeated for closed-reference approach (0,62%-2,35%) fig. 1.4.11. However, for the latter approach, the minimum difference between observed and expected values is 0,92% (g_*Vibrio*) while for *de-novo* method is 1,59% (f_*Enterobacteriaceae*). *g_Vibrio* (or *f_Vibrionaceae*) is the most differently represented taxon between the two approaches: 15,83% mean for *de-novo* and 19,08% mean for closed-reference. Moreover the "Other" group is more abundant with *de-novo* (mean 11,86%) than with closed-reference approach (mean 8,11%).

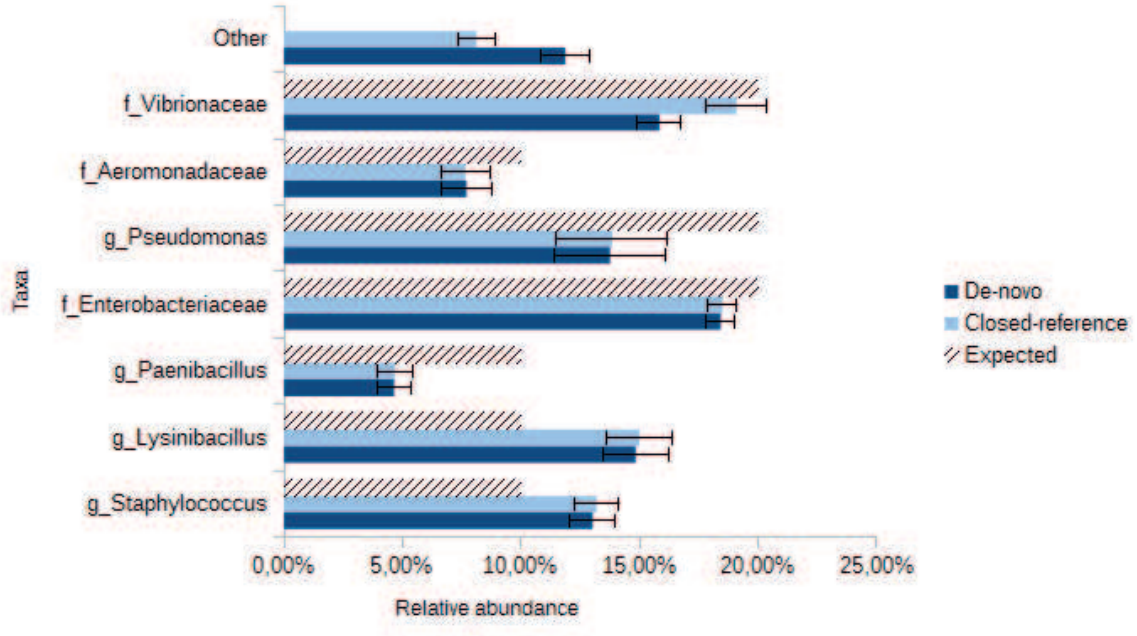

Fig. 1.4.11. Comparison between closed-reference and *de-novo* approaches in taxa identification and relative abundance results.

To better visualize the differences in the community structures driven by the methods, we performed a Procrustes analysis on the weighted UniFrac distances, fig. 1.4.12. As the *p*-value<0.0001 and $M^2 = 0.044$, we can draw the same beta diversity conclusions from either OTU-picking methods. Moreover the cluster correlation is quite good, meaning that the same samples are linked to their picking-method counterpart even if they are not completely overlapped.

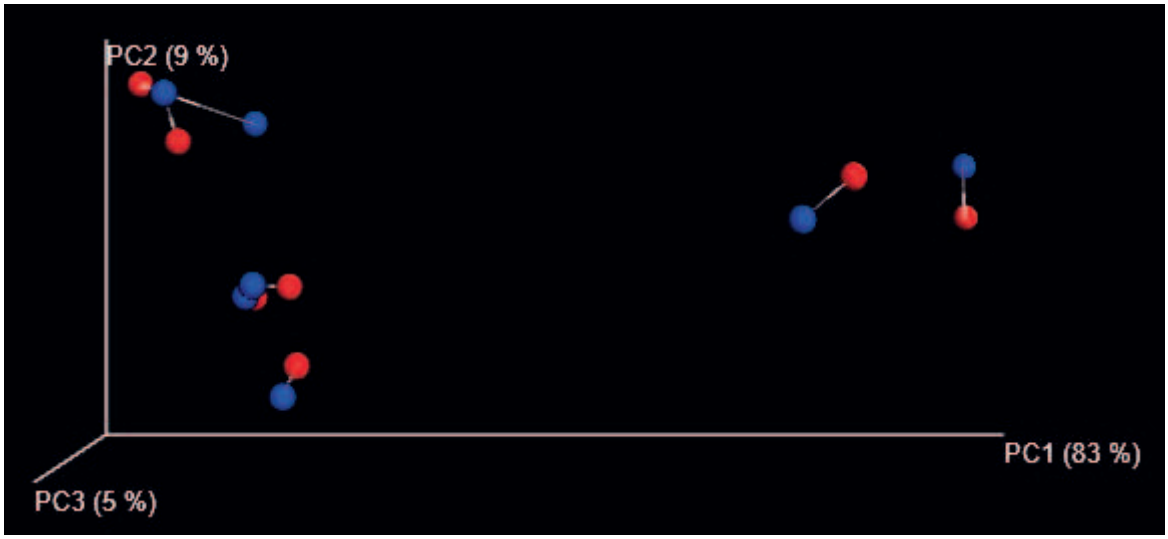

Fig. 1.4.12. PCoA Procrustes analysis on weighted UniFrac distance metrics, visualized with Emperor. Red spheres are *de-novo* samples while blue spheres are closed-reference samples.

Both methods showed an unusual abundance in *Pseudoalteromonadaceae* family, that has not been introduced into the mock. Some reads identified with this taxonomy were isolated and analyzed with the online tools of Greengenes database (version 11_10, http://greengenes.lbl.gov/) and RDP (RDP Naive Bayesian rRNA Classifier Version 2.10, training set 14). The first database resulted that the reads belonged to *Vibrio sp.* or *harveyi*. genus with more than 99% of identity, the second to *unclassified_Vibrionaceae* family (100% identity), *Lucibacterium* (78% identity), that is an old taxonomy for *Vibrio harveyi*. Thus incorporating the wrong *Pseudoalteromonadaceae* taxon to the *Vibrionaceae* taxon, the mock community is more represented without the unexpected taxon, fig. 1.4.12.

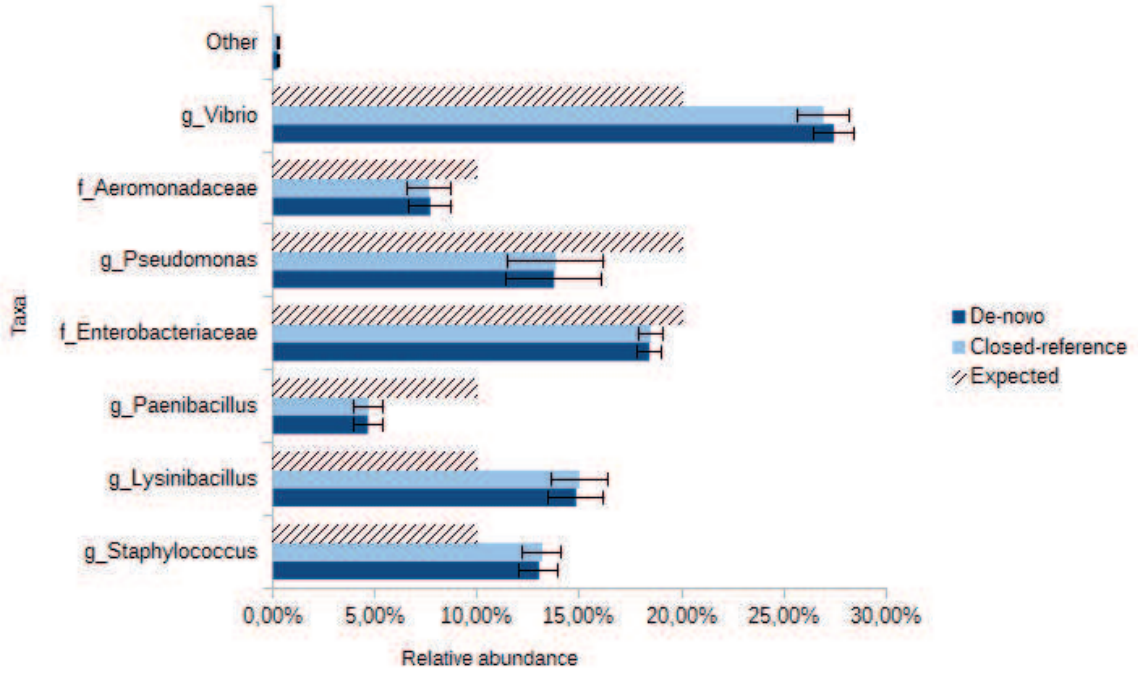

Fig. 1.4.12. Comparison between closed-reference and *de-novo* approaches in taxa identification and relative abundance results, after elimination of the unexpected taxon.

Repeatability of the protocol across samples and runs

In order to assess protocol reproducibility, we analyzed statistical difference between mock replicates by means the beta-diversity analysis protocol on *closed-reference OTUs*, that gave us better results in therms of relative abundance and taxonomy assignments. We tested the Weighted Unifrac significance test for each pair of samples (with Bonferroni correction): it returned the probability that a random sample/sequence assignment would result in more dissimilar samples than the actual pair of samples. As the *p*-value we obtained was equal to 1 for all comparisons, we could assess that all samples were phylogenetically not significantly different. We compared also differences across groups and runs. We divided mock replicates into not diluted (TQ), diluted (DIL) and second run (RUN2) and performed a Two-sided Student's two-sample t-test, (Bonferroni corrected). Resulting *p*-value was equal to 1 for all comparisons, all within, all between, TQ vs. DIL, DIL vs. RUN2 and TQ vs. RUN2, with a low distance value, fig. 1.4.13.

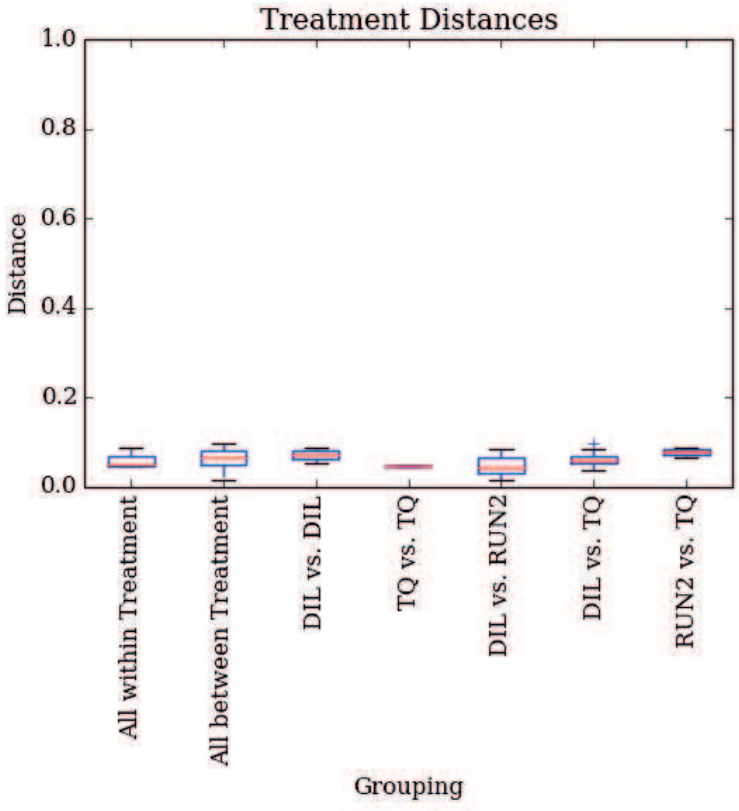

Fig. 1.4.13. Box-plot representing the distance between groups of samples, TQ, DIL and RUN2.

DNA amount effect

To test weather 5 or 20 ng of gDNA could provoke a bias in mock community structure, we analyzed the replicates with Weighted upgma tree on each sample and Two-sided Student's two-sample t-test, (Bonferroni corrected) on groups. If the different amount of DNA could bias the final result, we could expect that diluted replicates were more similar to diluted replicates and viceversa. On a tree based on distance metric, we could expect that samples belonging to the same group would cluster on the same node. We found that replicates of the same group branched on different nodes of the weighted Upgma tree (fig. 1.4.14). Moreover as the TQ vs. DIL group resulted not significantly different with Student's t-test (fig. 1.4.13), we could assess that those amounts of starting DNA did not affect the final mock community structure.

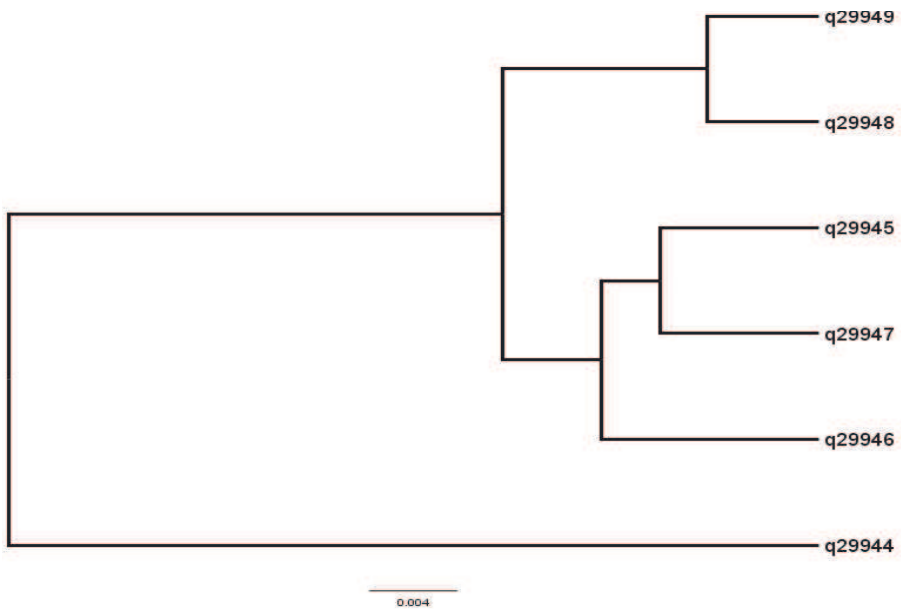

Fig. 1.4.14. Weighted Upgma tree not dividing the replicates in different subgroups based on the amount of DNA. The replicates are: TQ (29944, 29945, 29946) and DIL (29947, 29948, 29949).

1.5. Discussion

The primary goal of this study was to develop and evaluate the different protocol steps of 16S rRNA Next Generation Sequencing, in order to create a workflow to be applied in subsequent studies. As every step is basically linked with the others and can affect final results, we started by evaluating the fundamental nodes in 16S rRNA sequencing: among the others, hypervariable regions, primers and sequencing chemistry are the most deeply interconnected and influencing.

First of all the choice of Illumina Miseq with V3 chemistry (up to 300 bp reads) as reliable and accurate sequencing platform [58, 66, 67] allowed us to decide whether to sequence in single or in paired-end direction and which was the acceptable amplicon length to be sequenced. It is already known that the two approaches give different results on community structure for error rate increasing at the end of the read [92]. We showed that by overlapping at least 100 bp of forward and reverse reads, the average quality increased 2 folds, reaching acceptable values of 28.

The second fundamental choice dealt with primers as it effects the length of the amplicon, the hypervariable regions studied and the target match effectiveness, thus influencing taxonomy and phylogenetic coverage of the microbial community [3, 6, 11, 13, 14]. Our V3-V4 primers [11] produced an amplicon of 476 bp (with 133 bp overlapping), managed to match the highest number of bacteria in RDP database among the others, showed the higest ΔG and the lowest rate of dimer and hairpins production (even after tail coupling).

Many mock communities are produced by equimolar pooling of well characterized and sequenced bacterial 16S rRNA genes [88, 94]. In order to have an internal control with non characterized bacterial strains collected from food and environmental samples, we based the pooling ratio on amplification efficiency in RealTime PCR. The mock community was then sequenced and analyzed with different bioinformatc pipelines.

As reported in a recent work, the mock community never shows the exact structure as the expected [88]: neither for the number of OTUs or relative abundance. The number of OTUs found with both *de-novo* and closed-reference OTU-picking approach was higher than the expected; the first method showed about 3 fold OTUs more in comparison with the other (170 vs. 60 mean per sample). On the other hand the closed reference method resulted the most realistic number of expected families and genera: 8 families and 9 genera represented the correct composition of the community. Besides the numbers, the taxonomy assignments were almost correct, presenting all the expected families and genera plus 4 contaminant or mislabelling taxa more (such as *Moraxellace*, *Bacillaceae, Streptococcaceaae* and *Pseudoalteromonadaceae*). The presence of *Bacillaceae* family could also be linked to the double identification of the same reads, belonging to *Paenibacillaceae*, in Greengenes as reported in tab. 1.4.2. Nevertheless the *de-novo* approach showed more OTUs per sample and increased number in families and genera: 11 families and 7 genera of which 4 taxa were mislabelling or contaminants (*Pseudoalteromonadaceae, Prevotellaceae, Streptococcaceae, Moraxellaceae*). *De-novo* approach showed less profound taxonomic assignment efficacy for some taxa that were correctly identified with closed-reference, such as *Vibrio* genus; moreover only the first approach failed to assign some reads. Generically the overall structure of the community was similar with both approaches, even if some taxa abundance were better represented with closed-reference approach (fig. 1.4.11). In this way alpha-diversity showed a slight shift across the two methods. On the other hand with Procrustes analysis we showed that the beta-diversity was significantly similar with both methods, because *p*-value and M^2 were low (<0.0001 and <0.05 respectively). To this extent *de-novo* and closed-reference result in a similar community structure. As the differencies from expected abundance of some taxa such as *Pseudomonas* or *Paenibacillus* are mirrored in both approaches, the discrepancy could be due to a bias during the pooling phase or a preferencial target bias during PCR steps.

Another interesting point arose from the appearance of an unexpected taxon: *Pseudoalteromonadaceae* family. First of all we noticed that the family was wrongly assigned to *Vibrionales* order while it should be under *Alteromonadales* order. As the reads passed the UCHIME filter, we could exclude that the reads were artifacts. Thus a blast

against RDP database or the previous version of Greengenes, revealed that those reads belonged to *Vibrio* genus. In this way the *Vibrio* or *Vibrionaceae* taxon increased of about 10%. The last version of Greengenes 13_8 had put much work on improvement of taxonomic assignments of *Aeromonas* taxon [88] as in the previous version that taxon wasn't present [95]. In fact *Aeromonas* taxon in our mock was correctly detected. Moreover we couldn't characterize two genera: *Escherichia* and *Salmonella*, as they were grouped under *Enterobacteriaceae* family without any deeper specification. This is an issue coming from the database [88] because in the previous version of Greengenes (12_11) *Escherichia* was detectable.

In order to evaluate the reproducibility of the protocol, we sequenced 7 replicates of the mock community and we observed that with statistical comparisons the structure of the communities were not significantly different (Unifrac significance test *p*-value=1). This reproducibility was mirrored either with different starting amounts of DNA and across different runs. In fact we couldn't see any clustering difference in microbial community structure given by different amounts target DNA, as showed by weighted Upgma tree. As reported in a recent work [93], replicates of the same samples are more different when low DNA amounts (i.e. 0,1 ng) are used as starting template because PCR biases increase and selection of target bias becomes more common.

1.6. Conclusions and future perspectives

With this work we assessed a workflow for 16S rRNA analysis with Next Generation Sequencing approach on Illumina Miseq. We showed that this technique is reliable and reproducible, becoming a very powerful tool for microbial communities characterization in deep. On the other hand it shows some biases that have to be taken into account, both on wet-lab and bioinformatic side. As reported in scientific reports, the primer choice can heavily affect the results, thus comparison between different studies should be done carefully. Moreover more attention should be put on the databases, in which only known diversity is reported and some mislabelling or misassigned taxa could be present. To this extent, relying on a reference dataset such as for OTU clustering or taxonomy assignment should take caution in the interpretation of their results. Finally we compared the *de-novo* and closed-reference OTU-picking strategy on the same mock community and obtained slightly better performance with the closed-reference approach in terms of taxonomy assignment depth and precision. However this method can work finely when the target community is known and represented in the database, otherwise many reads are lost. In order to discover new variability or study more in depth a microbial community, *de-novo* approach could be the favorite method.

The analysis of contaminants in this kind of studies is a crucial point to understand how the environment (lab instruments, lab areas, materials, reagents) could influence the final result. For this reason we are running a set of experiments to assess the DNA-free quality of our reagents (mastermix, plastic materials). Contaminant sequences of the negative controls can be used to filter final sequences. For this kind of approach software as SourceTracker in QIIME [96] can be used to erase contaminant sequences from the row reads using a model template. Moreover during wet-lab phase some methods can be developed in order to allow error correction on errors raising during sequencing step [1, 12, 87, 88].

In conclusion 16S rRNA gene sequencing on MiSeq, keeping in mind the shown issues, can become a gold standard, cost-effective, useful and versatile tool for microbial communities characterization and, above all, comparison.

2. Chapter II

NGS application to evaluate the microbiological quality of raw materials for Ricotta cheese production.

2.1. Introduction

2.1.1. Definition of Ricotta cheese

Ricotta is a typical Italian cheese (from Latin “re-coctus”, that means cooked twice) produced with milk whey coming from cheese-making process. A heat treatment at 85°-90° C and the addition of acidification agents, such as lemon or acetic acid make whey protein coagulate [98]. Ricotta is not considered as a cheese because it is made from whey milk and coagulation of the proteins is not determined by the action of rennet. For this reason it is not protected by a specific legislation but hygienic regulations [99]. Moreover Ricotta comprehends plenty of products obtained from raw materials that are different for origin and composition, produced with variable techniques. Generally it is a fresh product but can be characterized by more or less prolonged shelf-life. Besides the different production techniques the main steps for Ricotta production are:

- selection and preparation of raw materials;
- thermal denaturation and aggregation of denatured proteins;
- separation of Ricotta from “scotta” (liquid by-product of ricotta-making process);
- cooling and packaging.

The fundamental raw material of Ricotta is whey (cow, sheep, or buffalo) but can also be a

mixture of wheys and milk and/or cream. In addition to milk or cream the most widespread ingredient is the salt (0.1-1% indicatively). Other methods consist in the addition of whey powder, casein or milk protein obtained by ultra-filtration [128].

Whey

Whey is a by-product which is obtained from the cheese-making process in general and the workings of the dairy industry. It is the liquid fraction coming from the coagulation of the milk and the subsequent separation of the curd. The separation of the whey from the clot is driven by a complex mechanism of "internal factors", the characteristics elasticity and permeability of the curd, and "external factors", represented by mechanical actions (mainly breaking), thermal (cooking) and physic-chemical (acidification-salting) [99].

Whey shows yellowish/greenish, sometimes with hints of blue, color depending on the type and quality of original milk [100]. For every Kg of cheese product on average about 9 liters of whey are obtained; thus dairies produce huge quantities of this by-product (about 100 liters per day) [101], fig. 2.1.1.

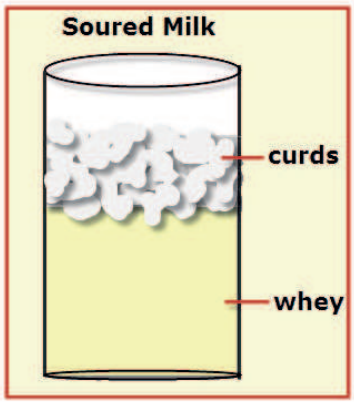

Fig. 2.1.1. representation of the separation of whey from curds during casein coagulation.

Until a few years ago, whey was used mainly for the production of cottage cheese and for animal feed but in recent years market proposals of drinks made with this compound have increased [102]. In fact whey contains high nutritional value and today is used in many commercial uses. Furthermore, whey proteins are widely used in the food industry: because of their excellent nutritional and functional properties they are often used as nutritional supplements. Moreover whey is able to form a gel capable of containing water, lipids and other components, properties which make it indispensable in the formulation of many foods such as processed meat, dairy and bakery products [103]. The exact composition of whey depends on the cheese-making process, however the main components are proteins, peptides (i.e. α-lactalbumin, β-lactoglobulin, whey albumin,

lactoferrin and lactoperoxidase, glycopeptides, phosphopeptides, different enzymes and hydrolysis products), lactose, vitamins and minerals along with traces of fats [103].

Cream

Separation process of the fats from milk followed by their pasteurization or sterilization allows to obtain different derivatives such as cream and butter.

There are different types of cream [104]:

- Surfacing Cream: natural surfacing happens during the processing of milk to produce cheeses like Parmesan. It is the most contaminated by microorganisms that are present in the milk because, adhering to the fat globules, they rise to the surface;
- Spinned Cream: it is obtained by centrifuging the fresh milk. It is the best from the microbiological point of view; it is not acidic and shows a fat concentration of 30-35%;
- Spinned Whey Cream: it is a product of whey centrifugation. It is acidic, rich in lactic acid bacteria and shows a highly variable fat concentration;
- Re-made Cream: it is produced when different butters are fused. After the fusion, whey is removed and the fat is re-suspended in milk.

2.1.2. Whey contamination source: raw milk

Microorganisms present in the products of milk processing are related to the sum of those initially present in the raw milk, of those added by man as starter or of those that are derived from the environment. The latter source can include equipment, ingredients (such as rennet), workers' contaminants. These microorganisms may be good, pathogenic or spoilers. The final product often doesn't show all the microorganisms that have developed at different stages of transformation, but only those bacteria that have withstood the sharp selection imposed by the fermentation and technological conditions of processing and ripening [99].

The composition of the microbiota in the milk can have a negative impact on its own quality and shelf-life; for example, psychrophilic bacteria can proliferate during milk cooling and spoil it through the production of extracellular lipase and protease [105, 106]. Moreover it can affect the subsequent ripening of dairy products. In fact, microorganisms can have different impacts on the texture, the taste and the organoleptic properties of the different products [107]. Microbial composition of the milk may also have health implications, because raw milk drinking, if contaminated by pathogenic micro-organisms

can lead, in some cases, to serious illness [108].

Microbial contaminants

The spread of microorganisms in the livestock environment (air, water, soil), either of human or animal sources, promote the opportunities for bacterial contamination of milk [109]. Moreover there are multiple events that can influence milk production chain, from farm to table, conditioning the hygienic quality of the product [110]. Starting from the milking step, possible sources of contamination increase: contamination by direct or indirect contact with the skin and mucous membranes of sick or healthy carriers (humans or animals); from the environment, through direct or indirect contact with surfaces, tools or contaminated equipment or by other means such as air, powders or water used to wash [111].

Verdier-Metz and colleagues (2012) analyzed the microbial communities present on the surface of the breasts of the cows through the creation of libraries, followed by sequencing and then compared them to the principal microorganisms found in milk and milk products. Bacteria identified corresponded to many food-technologically relevant microorganisms such as *Lactobacillus*, *Leuconostoc* and *Enterococcus* spp. They detected also bacteria that can contribute to the flavor and aroma of dairy products, as well as the occurrence of particular colors in cheese like, coagulase-negative staphylococci, *Arthrobacter*, *Brevibacterium* and *Corynebacterium* spp. However, some of the microorganisms detected on the surface of the nipple, for example, *Solobacterium*, and *Arcanobacterium Clavibacter* spp., have not been identified in the milk [112].

The cross-contamination in the milk can be attributed to multiple factors: for example, microorganisms present in the litter can contaminate the animal surface and therefore potentially could be transferred into the milk [113]. Similarly, milking machines may constitute a reservoir of microorganisms, thus the differences between the machines and the relative processing practices may affect the microbial population of the milk collected [114].

It is noteworthy that the application of hygienic processing led to a reduction of the microbial load of milk, either spoilers and good bacteria. This creates new space for additional contamination or bacteria spreading, negatively affecting the cheese produced with traditional processes and craft [115]. Furthermore refrigeration caused a diminution of the presence of *Lc. Lactis, Klebsiella pneumoniae, Lactobacillus plantarum*, and has made possible the identification, as a result of limited growth, of psychrophilic species such as *Listeria spp.* (fig. 2.1.2) and *Aeromonas hydrophila* [116].

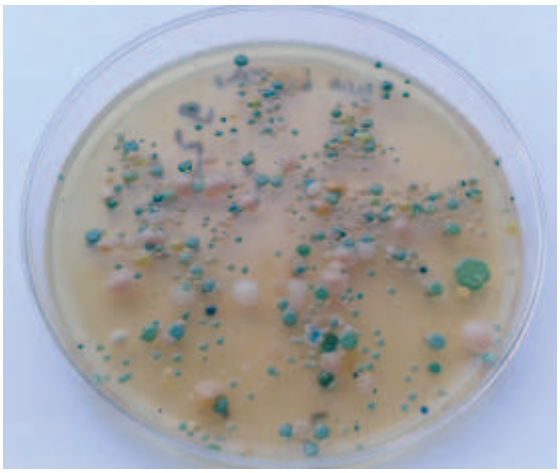

Fig. 2.1.2. ALOA medium for *Listeria spp.* isolation.

Bacterial spoilage can affect also food undergone heat treatment. Usually food spoilage by psychrophilic non spore-forming bacteria occurs due to inadequate heating or contamination after pasteurization, which can be eliminated by corrections in the protocols of pasteurization and proper sanitation [117]. In contrast, Gram-positive psychrophilic spore-forming bacteria have the ability to survive the conventional pasteurization and also to grow during storage at refrigeration temperatures; in addition some of them produce protease, resulting in occurrence of unpleasant flavors in the finished products [118]. The spores of these bacteria are widespread in the environment (especially soil) and can contaminate milk and derivatives, either directly or by ingestion of contaminated fodder and the consequent elimination in the feces. In fact raw and pasteurised milk exposed to post-pasteurisation contamination is most often spoiled by Gram-negative bacteria, predominantly *Pseudomonas*. while uncontaminated pasteurised milk is mainly spoiled by the Gram-positive thermoduric sporeformers, *Bacillus* and *Paenibacillus* [121, 122, 123].

The pathogenic species of interest in the dairy sector is *B. cereus*. The spores of this aerobic microorganism can be easily found in milk but they have to develop to a greater charge to 10^6 cells/g to produce a sufficient amount of toxin to cause poisoning [104]. *Bacillus* and *Paenibacillus* were identified as the main genera of spore-forming bacteria involved in the dairy industry. *Bacillus spp*. is found primarily in early shelf-life of pasteurized milk, while *Paenibacillus* predominates during product storage [119].

In addition, the analysis of degrading enzyme showed that mesophilic strains of *B. cereus* and *B. subtilis*, are strongly proteolytic, while the thermophilic strains showed a lower enzyme activity and thus a low deteriorating power [120].

2.1.3. Milk adjunct starter cultures

Until the beginning of the XX century all cheeses were produced without the addition of graft so cheese making and ripening was driven by the micro flora of raw milk. The unacceptable number of defects due to "wild" fermentation has led to the introduction of the graft or starter. The starter culture is a culture of lactic bacteria naturally present in milk, while the whey starter culture is a culture of lactic bacteria that grows in cheese making by-product. Depending on the type of cheese, the lactic acid bacteria used as starter belong to the group of mesophilic (growth temperature, 20-40 ° C) or thermophilic (growth temperature, 30-55 ° C). The former are represented mainly by *Lactococcus lactis subsp. cremoris* and subsp. *lactis* and *Leuconostoc mesenteroides subsp. cremoris*. The thermophilic starter, mainly omofermenting metabolism, are represented by *Streptococcus thermophilus, Lactobacillus helveticus* and *Lactobacillus delbrueckii subsp. Bulgaricus.*

Influence of starters on organoleptic characteristics of whey

The production of volatile compounds by the microbiota present in dairy products, is a fundamental feature for the quality of the product itself. However, when certain limits are exceeded or when there is an imbalance of aromatic compounds can be observed defects in the product [124].

Free fatty acids, in particular unsaturated fatty acids, are oxidation-prone, as they are known to be precursors of lipid oxidation products. The presence of these compounds constitutes a risk factor for the formation of unpleasant flavors in whey-derived products. A significant increase in volatile compounds from lipid oxidation, in particular as regards the hexanal, can be detected during storage along with a decrease in free fatty acids. Tomaino and colleagues (2004) on one hand showed that lipid oxidation begins during the production of the whey and continues during storage and, on the other hand, that the starter cultures significantly influence the level of lipid oxidation products detected [125].

Furthermore the usage of thermophilic starter cultures produce flavors that are different from those obtained with mesophilic starter cultures [126]; in particular the lipid oxidation products are more concentrated in dried whey protein from mesophilic starter cultures if compared with whey protein derived from whey obtained with other microbial cultures [127].

Liaw and colleagues (2010) documented with descriptive sensory analysis the flavor profiles of the liquid whey coming from the production of Mozzarella and Cheddar cheese. Volatile components were extracted and characterized by gas chromatography-mass spectrometry and gas chromatography-olfactometry. They showed that the whey from the Cheddar cheese production has sweeter aromatic flavors if compared to whey from the

production of Mozzarella. In particular the most abundant aromatic compounds detected in the wheys derived from Cheddar cheese were diacetyl, 1-octen-3-one, 2-phenethanol, butyric acid, and (E) -2-nonenal, while in the whey resulting from the production of Mozzarella there were mainly diacetyl, octanal, (E) -2-nonenal, and 2-phenethanol [126].

2.1.4. Food microbiology

Food microbiology is a fundamental tool for food safety and hygienic condition monitoring. In fact bacteria in food can be pro-technological and probiotic microorganisms but also pathogens, anti-technological or spoilers. Some of them are considered marker of health risk or un-hygiene. Food microbiology analysis can assess the presence/absence of these microorganisms, the overall presence of any bacteria present and allow to quantify them. Moreover they can provide any information about the origin, biology, ecology, metabolic activities and the systematic of these bacteria evaluating their role and their behavior in the different food ecosystems. A basic point is the preparation of the sample: it has to be representative of the entire analyzed food, minimizing the contamination of exogenous microbes [99]. Usually these analysis are used in combination with quality certification systems such as HACCP. Currently, besides these classical approaches, there are new methods to identify and quantify that developed in response to the need for more sensitive, faster, automated and cheap techniques [99]. Thus actually food microbiology is divided into two main branches: culture-dependent and culture-independent techniques.

Culture-dependent methods

Culture-dependent methods are based on traditional culturing methods on lab-media (fig. 2.1.3) in combination with phenotype (morphological, physiological and biochemical) and genotype analysis (species-specific PCR or RAPD) necessary for correct identification and typing. Culture-dependent methods usually manage to define bacterial groups/genera as their sensitivity rarely get to correctly describe the species. Thus the isolation of the colonies from the plate and their identification, biochemical typing and/or molecular identification are needed to increase the sensitivity of the analysis [129]. These investigation systems show limitations due mainly to the problem of non-culturability of many microorganisms with classical culture methods [32, 33] (see Chapter I). In complex microbial ecosystems the ability of some bacteria to grow can be stimulated by the presence and interaction with other microorganisms that are able to supply for these deficits [11]. Thus on an artificial medium many bacteria cannot be counted.

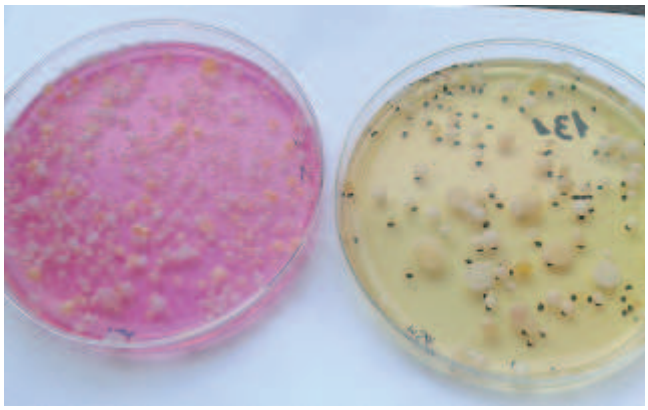

Fig. 2.1.3. Lab media (MSA) for culture-dependent analysis.

Culture-independent methods

Microbial ecology has deeply changed in recent decades with regard to the methods used in the analysis of microbial communities. The difference between the diversity that emerges using classic culturing methods and the real diversity has increased the importance of the usage of culture-independent molecular approaches [32, 33, 130]. By means these methods, microorganisms are no longer cultured and isolated on Petri dishes, but they are detected directly from the sample through the analysis of their DNA and RNA. The main novelty of these techniques is the direct extraction of nucleic acids from the complex matrix in study, which then are analyzed with specific tools capable of defining their diversity. Thus with these methods the complexity of the microbial system is studied comprehensively [131].

Studying DNA it is possible to understand how many species are present in the sample, the biodiversity, while analyzing the RNA it is possible to understand which are the metabolically active species [132]. These methods are based both on cloning and/or sequencing of the DNA fragments or often on the amplification of target sequences using PCR. The most widely used culture-independent methods are: *Denaturing Gradient Gel Electrophoresis* (DGGE), *Temperature Gradient Gel Electrophoresis* (TGGE), *Single-Strand Conformation Polymorphism* (SSCP), *Restriction Fragment Length Polymorphism* (RFLP), *Terminal Restriction Fragment Length Polymorphism* (T-RFLP), *Length Heterogeneity Polymerase Chain Reaction* (LH-PCR), *Automated Ribosomal Intergenic Spacer Analysis* (ARISA), cloning and sequencing of 16S amplicons, DNA microarrays, Next Generation Sequencing (NGS) libraries of whole genome shotgun or 16S rRNA [133]. For a comprehensive discussion on Next Generation approaches see Chapter I introduction.

2.1.5. Next Generation Sequencing applied to food microbial communities

Traditional culture-dependent methods are considered partial and labouring. Molecular methods, however, allow for rapid detection spoilage microorganisms, for quality of dairy products improvement and provide a better ratio of cost-effective results [134]. Furthermore Next Generation Sequencing analysis represents a huge improvement in food microbiology over the traditional Sanger sequencing [135]. In fact as the latter has a lower throughput, for very complex matrices it can lead to a significant underestimation of the present microbial communities [136].

Next Generation Sequencing can be applied to specific target genes, such as 16S rRNA, as well as to all genomes (metagenomics) and transcripts (metatrascriptomic) of a bacterial community. This makes it possible to determine the relative proportions or active metabolic pathways within complex populations [137]. These approaches are used more commonly in the characterization of the community of food, such as cheese. In fact cheese quality depends mainly on its microbial profile, because microorganisms contribute not only aroma, but also the defects of the product itself [134]. The whole genome sequencing or metagenome sequencing allow to better understand the relationships of organisms within the food matrix and to highlight those microorganisms that can cause defects in dairy products or bacteria that improve or protect the product [137].

2.2. Purpose

The purpose of this study was to evaluate the microbial composition of the raw materials used by Dairy Elda (VR) for the industrial production of cottage cheese, in order to monitor the microbiological quality and eventually implement it.

In particular, this study aimed to:

- integrate culture-dependent with culture-independent methods on liquid whey, concentrated whey and cream, to establish a background of microbiological knowledge on these raw materials and to understand how storage period may change their quality;
- apply Next Generation Sequencing (NGS) and Gas Chromatography analysis (GS) to a business issue concerning the alteration of the organoleptic characteristics in some liquid wheys, in order to study whey microbial composition and to understand which bacteria were implicated in the spoilage process.

2.3. Materials and Methods

2.3.1. Samples collection

In this study the analysis were performed on raw materials used by the dairy "ELDA" (Vestenanova, VR) for the industrial production of Ricotta: liquid whey (PLW), concentrated whey (PCW), and cream (CR). Raw materials used for the production of two lots of Ricotta were analyzed by means culture-dependent and -independent methods.

Finally 9 batches of frozen wheys were analyzed by means culture-independent method. These materials were classified by the Company as "compliant" (4) and "non-compliant" (5) wheys, based on organoleptic properties. Sampling summary and analysis are reported in tab. 2.3.1.

Sampling Number	Sample		State	Season of sampling	Analysis
1°	Liquid whey (1_PLW)		Fresh	Winter (12/2013)	M+S
	Concentrated whey (1_PCW)				
	Cream (1_CR)				
2°	Liquid whey (2_PLW)		Fresh	Winter (02/2014)	M+S+NGS
	Concentrated whey (2_PCW)				
	Cream (2_CR)				
3°	Liquid whey (3_PLW)		Fresh	Summer (07/2014)	M+NGS
	Concentrated whey (3_PCW)				
	Cream (3_CR)				
4°	Liquid whey (PLF)	Compliant	Frozen	Summer (2013)	NGS + GC
		Non compliant			

Tab. 2.3.1. Sampling summary, period, analysis and type of sample. M= classic microbiology; S= Sanger sequencing; NGS=Next Generation Sequencing; GC= Gas Chromatography.

2.3.2. Microbiological analysis

In the first sampling we evaluated the shelf-life of the different raw materials following the shelf life adopted by the Dairy, while the other sampling were more focused on microbiological and molecular analysis of cream and whey at the moment of their

processing for Ricotta-making. In tab. 2.3.2 sample analysis and media used are shown.

Microbiology analysis		
Sampling Number	***Shelf life* analysis**	**Lab media**
1°	Yes	Milk PCA, Milk PCA + starch, MYP
2°	No	Milk PCA, Milk PCA + starch, SPS, PAB/CFC
3°	No	Milk PCA, Milk PCA + starch, PAB/CFC

Tab. 2.3.2. Sampling, analysis and media performed on the raw materials.

MRD (Maximum Recovery Diluent; 1 g /l universal pepton and 0.75% NaCl) was used to produce serial dilutions 1:10 from the different samples. The plated dilution was chosen according to the microbial load observed by the internal Company laboratory during routine microbiological analysis. Before the dilution we performed a heat-shock treatment on sub-samples of raw materials in order to evaluate the spores concentration, by killing the vegetative bacteria and inducing spore germination. The treatment was carried out on 4 ml of the product at 80 ° C for 10 minutes and subsequently cooled in ice water.
The number of CFU was evaluated by means the formula:

$$N= \sum C/ [V (n1 + 0,1\ n2)\ d]$$

ΣC = Sum of colonies counted in the two plates in a row;

V = Volume used for inoculum;

n1 and n2 = number of plates employed in the first and second dilution taken into account;

d = first dilution considered in the counts (the most concentrated).

Medium	Purpose	Heat- shock	Culturing	Volume	Temperature of incubation	Characteristics of incubation
Plate Count Agar with skimmed Milk (Milk PCA; *Biokar diagnostics*)	Total mesophylic count (TMC)		Inclusion	1 ml	30°C	O_2
Plate Count Agar with skimmed Milk + starch 0,2% (Milk PCA; *Biokar diagnostics*)	Aerobic spores	Yes	Inclusion	1 ml	30°C	O_2
Mannitol Egg Yolk Polymyxin Agar (MYP; Sacco)	*Bacillus* genus		Spreading	100 µl	30°C	O_2
Sulfite Polymyxin Sulfadizine Agar (SPS, Sacco)	*Clostridium* genus		Spreading	100 µl	37°C	No-O_2
CFC Pseudomonas Agar Base (PAB/ CFC); *Oxoid Microbiology Products, Thermo Scientific*	*Pseudomonas* genus		Spreading	100 µl	22°C	O_2

Tab. 2.3.3. Media used for bacterial count and isolation.

2.3.3. Colonies storage

During microbiological analysis, some colonies that showed different morphologies were selected and stored, for subsequent molecular analysis. The selected colonies were stored at -80 ° C in 750 L of TSB and 750 µl glycerol.

2.3.4. Molecular analysis

The exponential development of molecular methods for the analysis of DNA/RNA and the impact of bioinformatics have revolutionized microbiological analysis and have allowed a better understanding of the microbial diversity and phylogenetic relationships [138]. These methods can be adjoined to classical methods and are used in the identification, classification or typing of bacteria, contamination sources, pathogens and microbial communities profiling. RNA analysis is very important in the field of food microbiology as it allows to discriminate the active species at the time of sampling [130].

DNA extraction

Single colony DNA extraction

The extraction of DNA from single colonies was performed by means boiling method. The bacterial colony grown on Petri dish is picked up with a sterile loop and put in 100 µl of DNA-free BDH water. Thus bacterial suspension was subsequently boiled in the heating block for 10 minutes, after which it was centrifuged at 14000 rpm for 2 minutes. The supernatant containing the DNA was collected, diluted 1: 100 in BDH water and stored at -20 ° C.

DNA extraction from complex matrices

This kind of extraction was performed on frozen whey samples, after slowly thawing. 2 ml of whey were transferred in DNA-free Eppendorf and centrifuged at 13500 rpm for 10 minutes. The supernatant obtained was eliminated by inversion of the tube. The pellet was resuspended with 1 ml of PBS (Phosphate Buffered Saline) and fat residuals from the eppendorf walls were removed before performing a second centrifugation at 5000 rpm for 5 minutes. The protocol of extraction was performed as described in paragraph Ch. 1.3.4.

RNA extraction

5 ml of fresh samples were collected, diluted 1:10 in MRD and centrifuged at 10,000 g for 5 minutes. The pellet was resuspended in 700 µl of RLT Buffer, 3 µl of β-mercaptoethanol and about 250 µl of zirconium beads. The whole was homogenized with Ribolyzer bead-beater (Hybaid) for 20 seconds at speed 4. A subsequent centrifugation for 10 seconds at

maximum speed was performed at the end of which about 700 L of supernatant were collected and transferred to a new Eppendorf. The lysate thus obtained could be stored at -80 ° C until extraction.

The RNA extraction was performed with RNeasy® Tissue Mini Kit (Qiagen, Hilden, Germany). The tubes with the lysate were added with 700 µl of 70% ethanol; the content was then mixed by gentle inversion and 700 µl of the resulting solution were transferred into a new tube containing silica columns provided by the kit. A centrifugation at maximum speed for 15 seconds with subsequent discard of the flow-through was performed. The step was repeated twice with the remaining volume. 350 µl of Buffer RW1 were added into each column and centrifuged at maximum speed for 15 seconds in order to discard the flow-through. 80 µl of DNase solution diluted in Buffer RDD were added to every column and incubated at room temperature for 15 minutes. Again 350 µl of Buffer RW1 were added into each column and centrifuged at maximum speed for 15 seconds in order to discard the flow-through. 500 µl of Buffer RPE (containing ethanol) were added to the columns and centrifuged at maximum speed for 15 seconds. This washing step was repeated a second time with a 2 minutes spin. An additional centrifugation step was performed for one minute at maximum speed in order to remove residual ethanol. Subsequently, the filters were transferred into new tubes and added with 30 µl of RNase-free water. A centrifugation at 10000 rpm for 1 minute was repeated twice as the first elution volume was loaded back on the filter and spinned again.

Reverse transcription

Reverse transcription relies on the Reverse Transcriptase enzyme that manages to synthesize a complementary DNA molecule (cDNA) from the RNA. Reverse transcription of RNA was performed using the kit SuperScript® II Reverse Trascriptase (Invitrogen). The protocol was applied to each sample in duplicate: RT (reversely transcribed) and Nort (not reversely transcribed) The latter was produced in order to monitor the presence of contaminating DNA in the extracted RNA. 10 µl of RNA for each sample were added with 2 µl of mix n°1 containing:

- 1 µl of *Random examers* (EuroClone)
- 1 µl of dNTP mix (10 mM)

Tubes were heated to 65° C for 5 minutes and then placed in ice for 2 minutes. Each sample was added with 7 µl of mix n°2 containing:

- 4 µl *First-Strand Buffer* (5X)
- 2 µl DTT (0.1M)

- 1 µl (40 U/µl) *RNaseOUT™*

Samples were incubated for 2 minutes at 25° C. 1 µl of *SuperScript™* II RT (200 U/µl) was added to each sample. Tubes were incubated at 25° C for 10 minutes and subsequently at 42° C for 50 minutes. The reaction was inactivated by heating at 70° C for 15 minutes.

Extraction quality controls

RealTime PCR

This technique has been extensively described previously (Ch. 1.3.4). In this study the method of RealTime-PCR was used to assess the quality of the extracted DNA form frozen whey samples and to completely exclude the presence of genomic DNA in extracted RNA from fresh samples. The latter was done by comparing the amplification of RNA transcripts with not reverse transcribed ones. For the reaction of RealTime-PCR we used the *Platinum™ SYBR® Green qPCR Supermix-UDG* mastermix (Invitrogen). The reaction mixture contained *SYBR Green, Platinum Taq DNA polymerase (60 U/µl)*, $MgCl_2$ 6 mM, *Uracil-DNA glycosidase* (UDG, 40 U/µl), dNTPs 40 mM, Tris-HCl (pH 8.4) and KCl (100 mM).

Primers used were those reported in tab 1.3.1 from Nadkarni et al. (2002) [11]. 2,5 µl of diluted sample were added with 7,5 µl of reaction mixture that is reported in tab. 2.3.4.

Reagents	**Concentration**	**Quantity (µl)**
Mastermix 2X	1X	5
Primer forward 10 µM	0.25 µM	0.25
Primer reverse 10 µM	0.25 µM	0.25
H_2O	-	2

Tab. 2.3.4. RealTime-PCR mastermix reagents concentration.

Amplification step was performed on *LightCycler 480, Roche®* with the cycle reported in tab. 2.3.5.

Temperature	**Time**	**N° cycles**
50 °C	2 min	1 cycle
95 °C	10 min	1 cycle
95 °C 60 °C	10 sec 1 min	45 cycles

Tab. 2.3.5. RealTime-PCR cycle used in this study.

NanoDrop spectrophotometer

The purity of DNA and RNA extracted was valued at *Nanodrop ND-1000* (NanoDrop Technologies, Wilmington, DE, USA). NanoDrop is a broad-spectrum spectrophotometer (220-750 nm), able to quantify nucleic acids evaluating their purity. It takes advantage of a particular technology based on surface tension that is created when a small volume of liquid is located between two neighboring surfaces. Quantification is possible by measuring absorbance level at 260 nm while purity evaluation is based on the ratio of absorbance at both 260/280 nm (to detect the presence of proteins or other contaminants) and both 260/230 nm (to evaluate the presence of residual solvent). A good quality ratio for nucleic acid is comprised between 1.8 and 2.

Qubit® fluorometer

The Qubit® 2.0 Fluorometer (Invitrogen, Life Technologies) is a fluorometer used for assay quantification of nucleic acids and proteins. The technology is based on the usage of fluorescent probes capable of emitting light only when bound to specific target molecules. This makes the quantification sensitive and specific, even at low concentrations. The assay was performed using the Qubit® Assay Kit (Invitrogen, Life Technologies). Standards and samples were prepared following manufacturer instructions. Working Solution I (*Quant-IT*^{TM} reagent in *Quant-IT*^{TM} buffer) was added to the correct sample dilution, with a volume comprised between 180 and 199 µl in order to achieve the final volume of 200 µl. Samples and standards were vortexed for 2-3 seconds and incubated at room temperature for 2 minutes, finally quantification was performed.

Polymerase Chain Reaction (PCR)

This technique has been discussed extensively in Chapter 1.3.4.

In this study the following PCR were performed on the samples described in tab. 2.3.6. while primers used are shown in tab 2.3.7.

Sampling	**PCR L_1G_1**	**PCR *rpoD* (*Pseudomonas fluorescens*)**	**PCR *pycA* (*Bacillus cereus*)**
1°	Yes	No	Yes
2°	No	Yes	No

Tab. 2.3.6. Type of locus-specific PCR performed on different samplings (but 3rd and 4th samplings).

Primer	**Sequence**	**Tm (°C)**
L_1	CAAGGCATCCACCGT	50.6
G_1	GAAGTCGTAACAAGG	45.1
rpoD-F	CTGATCCAGGAAGGCAACATCGG	64.5
rpoD-R	ACTCGTCGAGGAAGGAGCG	65.6
pycA-F	GCGTTAGGTAACGAAAG	57
pycA-R	CGCGTCCAAGTTTATGGAAT	57

Tab. 2.3.7. Primers used for locus-specific PCR.

L_1G_1 PCR

In bacteria belonging to the same genus the intergenic region called Spacer of Ribosomal genes 16S-23S has the same length. The primers used to amplify this intergenic region are taken from the work of Coppola and colleagues (2001) [132], and were used for the first time in 1993 by Jensen et al. [139]. They were selected on conserved regions at the 3' end (primers L_1) of 16S rDNA and 5' end of the 23S rDNA (primer G_1) [139].

In this study, we applied this analysis to highlight the variability of the species present in the samples in order to sequence only the representative colonies with the Sanger method. In each sample the volume of the reaction mixture was 20 µl, the mastermix used was *GoTaq® Flexi DNA Polymerase* (Promega). Reagents used are shown in Table (Tab. 2.3.8), with the appropriate final concentrations.

Reagents	**Concentration**	**Quantity (µl)**
Buffer 5X	1X	4
$MgCl_2$ 25mM	1.9 mM	1.5
dNTPs 25 µM each	0.1 µM	0.2
Primer forward 10 µM	0.25 µM	0.5
Primer reverse 10 µM	0.25 µM	0.5
Taq Polimerase 5U/µl	1U	0.2
DNA 2 ng/µl	10 ng	5
H_2O BDH	Up to 20 µl	8.1

Tab. 2.3.8. Reagents used for PCR L_1G_1 in a volume of 20 µl.

PCR was performed with the touch-down cycle (decreasing annealing temperature by cycle) shown in tab. 2.3.9, on thermocycler (*EuroClone*, Celbio).

Temperature	**Time**	**N° of cycles**
94 °C	2 min	1 cycle
94 °C 48 °C (-0,5°C each cycle) 72 °C	30 sec 30 sec 45 sec	16 cycles
94 °C 40 °C 72 °C	30 sec 30 sec 45 sec	20 cycles
72 °C	7 min	1 cycle

Tab. 2.3.9. L_1G_1 PCR cycle conditions.

RpoD PCR

In order to confirm that bacteria grown on medium *CFC Pseudomonas Agar base* belonged to *Pseudomonas fluorescens* group we performed a molecular screening on the gene *RpoD*. This locus-specific PCR allows to screen directly micro-lysed colonies picken up from the medium: the amplification is successful only if bacterium belongs to the specific genus. Primers used in this analysis belong to a group of 7 primers targeting different

housekeeping genes, as part of the so-called MLST (Multi-Locus Sequence Typing assay). This approach allows to recognize and subtype bacterial strains that are phylogenetically linked to the *Pseudomonas fluorescens* group, the most deleterious bacterial group for food industries [140]. In each sample the volume of the reaction mixture was equal to 20 µl. The reagents used are shown in the table 2.3.9 below, with the appropriate final concentrations, the used mastermix was *GoTaq® Flexi DNA Polymerase* (Promega).

Reagents	**Concentration**	**Quantity (µl)**
Buffer 5X	1X	4
$MgCl_2$ 25mM	2.5 mM	2
dNTPs 25 µM each	0.1 µM	0.112
Primer forward 10 µM	0.1 µM	0.25
Primer reverse 10 µM	0.1 µM	0.25
Taq Polimerase 5U/µl	0.5 U	0.1
DNA 2 ng/µl	10 ng	5
H_2O BDH	Up to 20 µl	8.1

Tab. 2.3.10. Reagents used for *RpoD* PCR in a volume of 20 µl.

PCR was performed with the cycle shown in tab. 2.3.11, on thermocycler (*EuroClone*, Celbio).

Temperature	**Time**	**N° of cycles**
94 °C	2 min	1 cycle
94 °C 60 °C 72 °C	20 sec 30 sec 1 min	35 cycles
72 °C	7 min	1 cycle

Tab. 2.3.11. *RpoD* PCR cycle conditions.

PycA PCR

We performed a molecular screening for the presence of the bacterium *Bacillus cereus* complex in the raw materials using specific primers for the housekeeping gene *PycA* of this species [141]. We selected different colonies grown on defined medium for *Bacillus cereus*

(MYP), micro-lysated and amplified with primers for *PycA*. The the gene amplification allows to understand if the bacterium belongs to the *Bacillus cereus* complex. These primers are part of a group of 7 primers that, within the MLST assay, allow to correctly determine the phylogenetic distance of a bacterial strain with *Bacillus cereus* complex [141].

Reaction was performed using the same reaction mixture as the *RpoD* PCR, reported in tab. 2.3.10, with the following cycle (tab. 2.3.12).

Temperature	**Time**	**N° of cycles**
94 °C	2 min	1 cycle
94 °C 55 °C 70 °C	20 sec 30 sec 30 sec	35 cycles
70 °C	5 min	1 cycle

Tab. 2.3.12. *PycA* PCR cycle conditions.

All these amplifications were checked on agarose gel as described in paragraph ch. 1.3.4,.

2.3.5. Gas chromatography assay

Gas chromatography is a method of separating the components of mixtures of gases and vapors; it works as a molecular filter which selectively retards the passage of molecules according to certain physical characteristics such as shape, size, weight and boiling point. A mixture of gases, in the course of its passage through the chromatograph, will divide into its components and emerge from the distal end of the column in regular order, to be measured by a detector whose output may be recorded graphically or electronically [142].

For more information on the molecules present in the two types of frozen wheys and to understand which bacteria may have been responsible for the chemical reactions and the resulting changes in organoleptic features, 3 compliant wheys and 3 wheys particularly altered in olfactory characteristics were analyzed by Veneto Agricoltura (Thiene -VI).

2.3.6. Canonical Correspondence Analysis and Correlation Analysis

Statistical analysis were performed with R-studio software 3.2.3 using *cca* function of the {*vegan*} package, showing the first 5 driving vectors, while CCA plot with function *plot.cca* from package {*vegan*}. *cor* function of the {stats} package was used to produce correlation analysis data while analysis plot was produced with *heatmap.2* {*gplots*}.

2.4. Results

2.4.1. Preliminary analysis on raw materials shelf-life

The first analysis were performed on different lots of raw materials from Dairy Elda, following their shelf life in order to establish a background knowledge on microbiological composition and to understand how the storage time could change the quality of raw materials. The analysis of the shelf life of raw materials was performed on a batch of cream (1_CR), concentrated whey (1_PCW) and liquid whey (1_PLW), sampled from the Company.

The samples were analyzed for a period of time compatible to the shelf life used by the company, which provided the following storage conditions:

- PCW up to 4 days at 10° C
- PLW up to 2 days at 12 ° C.
- CR up to 15 days at 4 ° C.

All materials were stored at 4°C.

Shelf-life counts

Microbiological analysis focused on three media: *Milk PCA*, *Milk PCA starch* and *MYP*, following the instructions given in table 2.3.2 of the section Materials and Methods. Raw materials were quite different from each other for bacterial loads and composition: 1_CR showed low mesophilic total counts which became higher only in end of the shelf life and showed no spores (Fig. 2.4.1 and fig. 2.4.2); nevertheless 1_ PLW and 1_PCW had a total bacterial load around 10^7 CFU/ml at the end of shelf life. Both showed a high abundance of aerobic spores. Moreover, as the TMC counts and MYP counts are very similar, it seems that bacteria in wheys belong mainly to *Bacillaceae* (fig. 2.4.3).

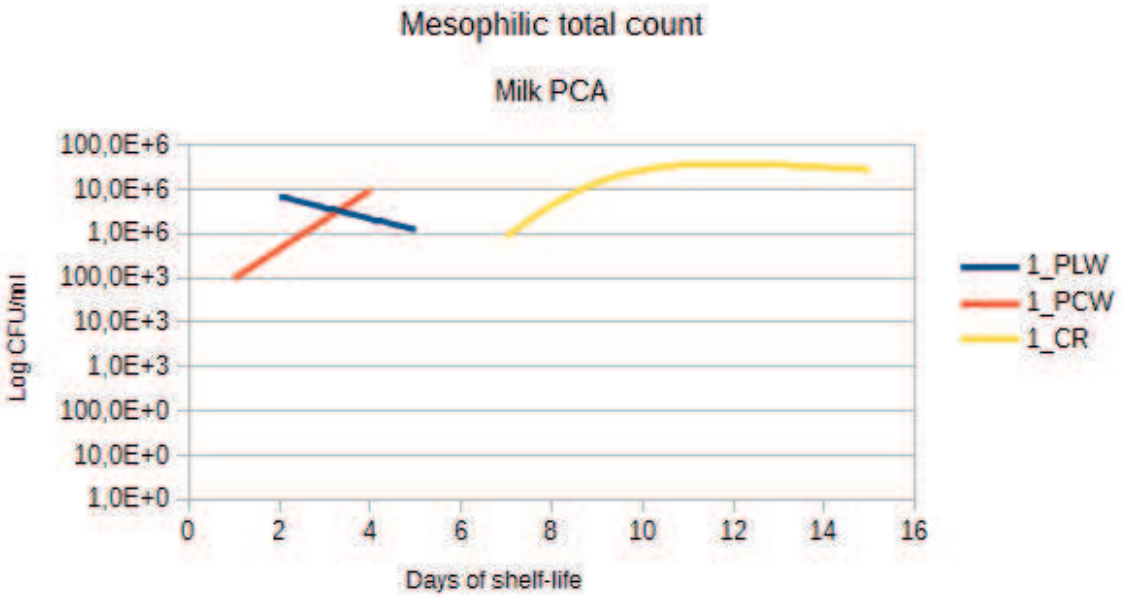

Fig. 2.4.1. Variation of TMC for the different raw materials during shelf-life.

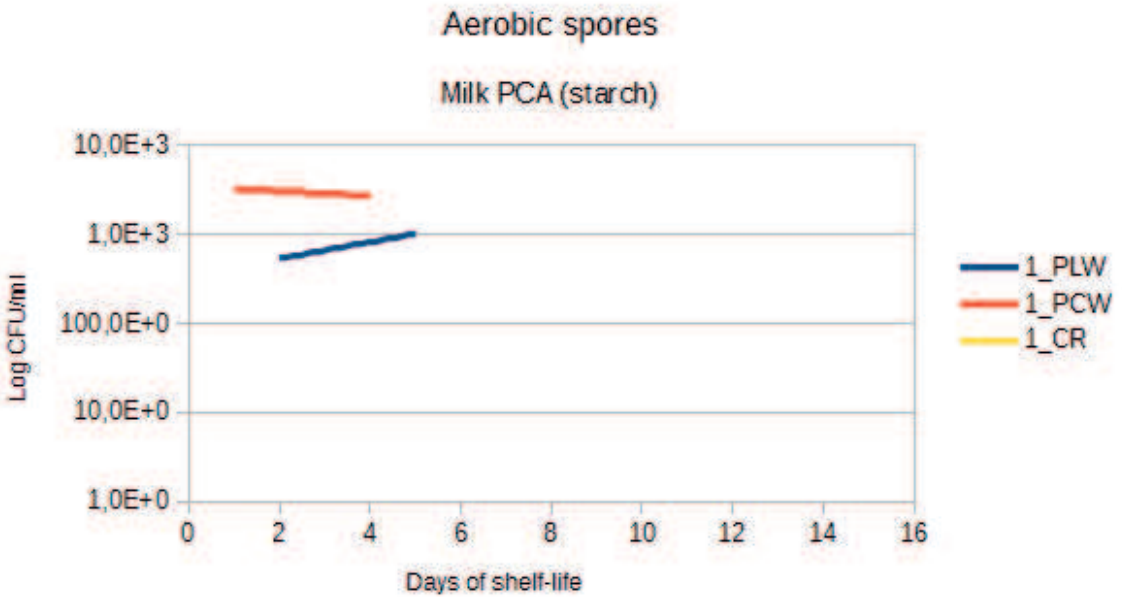

Fig. 2.4.2. Variation of aerobic spores count for the different raw materials during shelf-life.

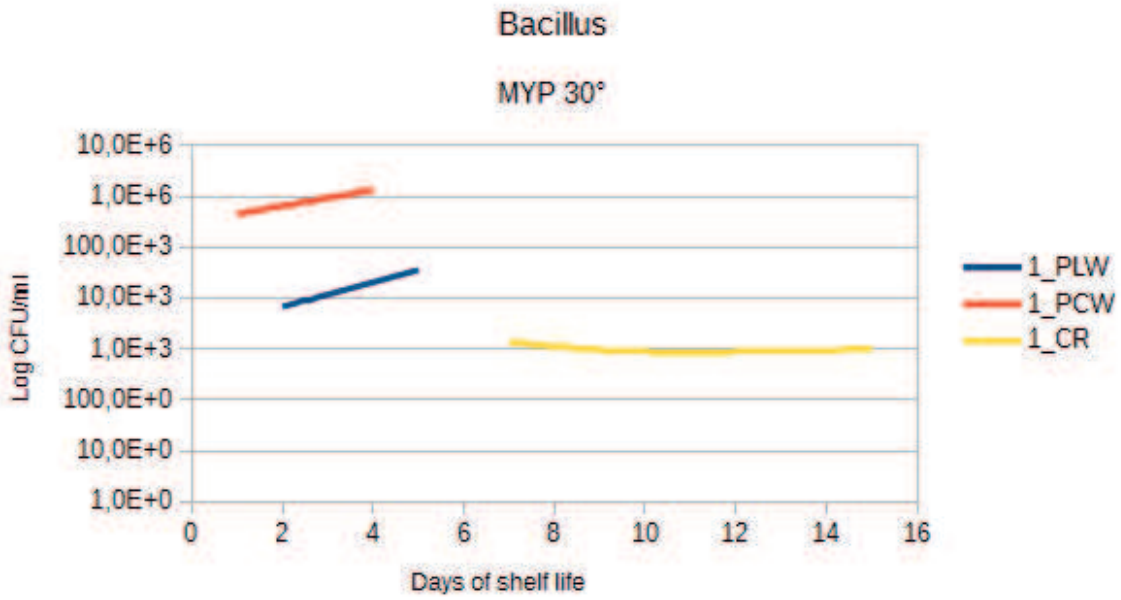

Fig. 2.4.3. Variation of *Bacillaceae* counts for the different raw materials during shelf-life.

Amplification of PycA for Bacillus cereus detection

As described in paragraph ch. 2.3.3, we screened many colonies grown on MYP medium using the *PycA* amplification approach. A strain of *Bacillus cereus*, already identified in the lab, was used as positive control. The gel is showed in fig. 2.4.4 resulting no amplification for any colony but the positive control.

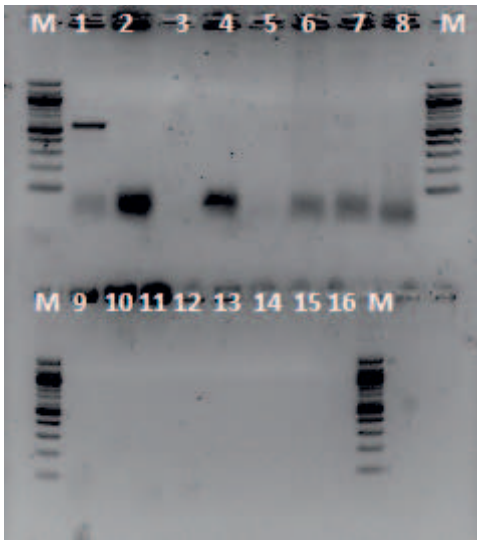

Fig. 2.4.4. Agarose gel 1,5%, M= 100 bp ladder. Well 1 contains the positive control, while wells 2-16 contain samples.

2.4.2. Evaluation of microbial composition using culture-dependent and -independent methods

Given the heterogeneity detected during the evaluation of the raw material shelf life, we decided to analyze in parallel raw materials by means microbiological analysis and classical NGS system with Illumina sequencing. In this way we could achieve more detailed information about the bacterial communities present in wheys and cream at the moment I which they were used for the production of ricotta cheese.

Culture-dependent methods results

Microbial load was evaluated through the analysis of *Milk PCA*, *Milk PCA* added with starch and PAB/*CFC*. The latter is a selective medium for the *Pseudomonas* genus, and it was included in the analysis as from previous sequencing analysis the presence of these common spoilage bacteria in the cream was detected.

As shown in fig. 2.4.5 a heavy mesophilic bacteria load is present, especially in wheys (fig. 2.4.5 A, B), with charges ranging from 10^6 and 10^8 CFU/ml as already highlighted in the previous analysis. Again aerobic spore load was found only in the wheys, spanning from 10^3 to 10^4 CFU/ml. Furthermore the analysis on selective medium PAB/CFC has revealed a high load of bacteria belonging to *Pseudomonas* genus, in the range of 10^6 CFU/ml for the CR (fig. 2.4.5 C) and between 10^4 and 10^7 CFU/ml, respectively, for PLW and PCW.

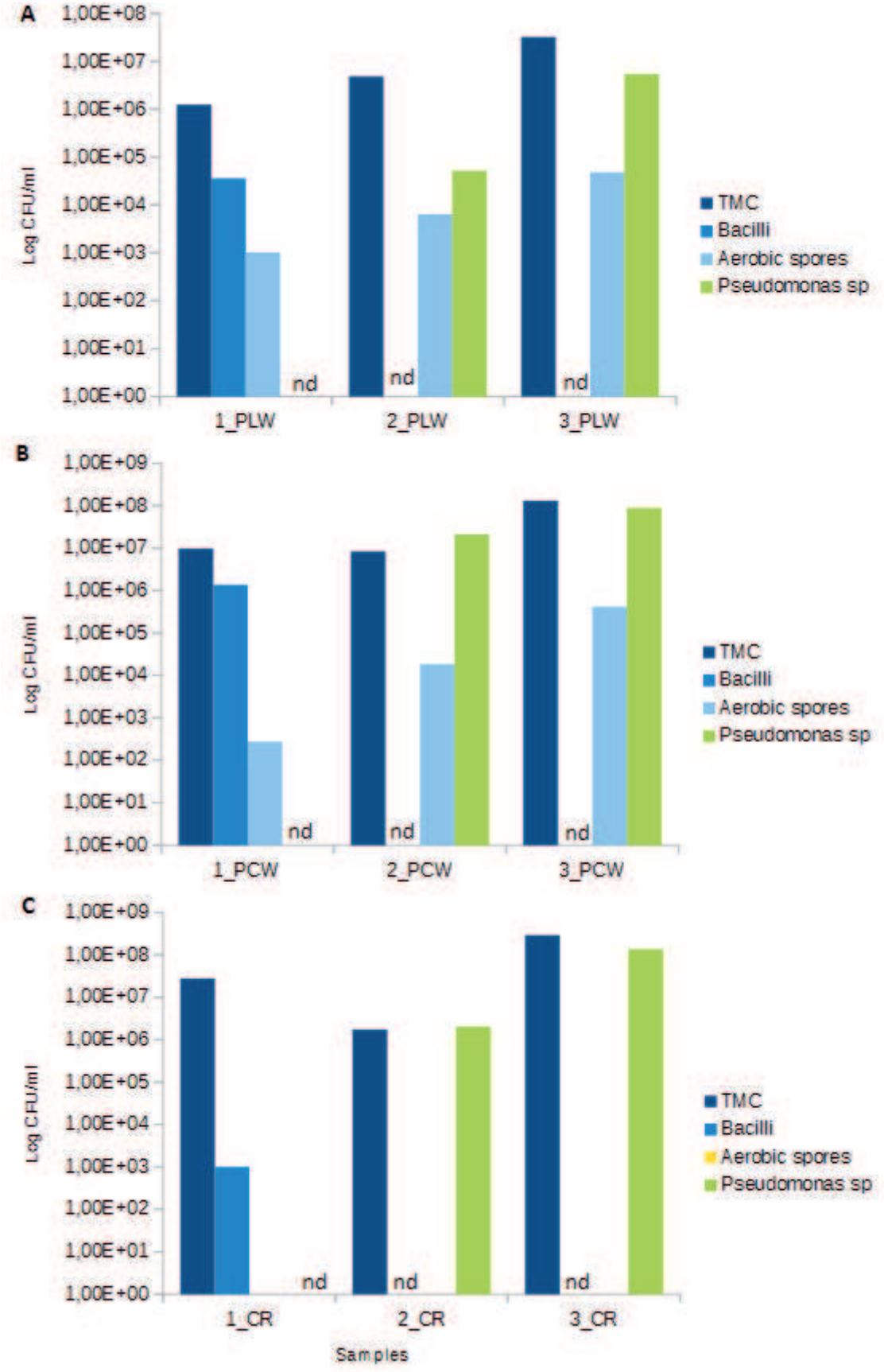
A
1,00E+08
1,00E+07
1,00E+06
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00
Log CFU/ml
nd
nd
nd
1_PLW
2_PLW
3_PLW
TMC
Bacilli
Aerobic spores
Pseudomonas sp
B
1,00E+09
1,00E+08
1,00E+07
1,00E+06
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00
Log CFU/ml
nd
nd
nd
1_PCW
2_PCW
3_PCW
TMC
Bacilli
Aerobic spores
Pseudomonas sp
C
1,00E+09
1,00E+08
1,00E+07
1,00E+06
1,00E+05
1,00E+04
1,00E+03
1,00E+02
1,00E+01
1,00E+00
Log CFU/ml
nd
nd
nd
1_CR
2_CR
3_CR
Samples
TMC
Bacilli
Aerobic spores
Pseudomonas sp

Fig. 2.4.5. Bacterial counts are plotted for every sample, PLW (A), PCW (B)and CR (C) and for the different samplings. Nd represents the absence of data.

Amplification of RpoD gene for Pseudomonas fluorescens species detection

Many colonies grown on PAB/CFC were visualized on the transilluminator to highlight their possible fluorescence, a typical feature of *P. fluorescens* (fig. 2.4.6). As a result we could find that many of the colonies were actually fluorescent.

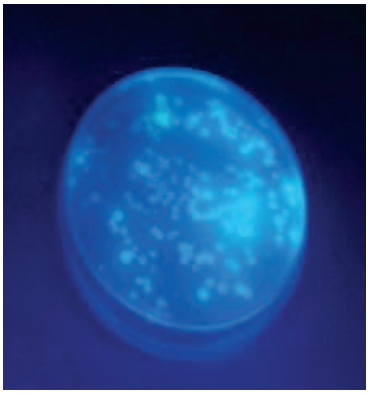

Fig. 2.4.6. Fluorescent colonies of *Pseudomonas* grown on PAB/CFC medium.

Some of the fluorescent colonies grown on *Pseudomonas* CFC agar base were screened for the amplification of RpoD gene, to confirm that they belonged to *Pseudomonas fluorescens* group (Fig. 2.4.7).

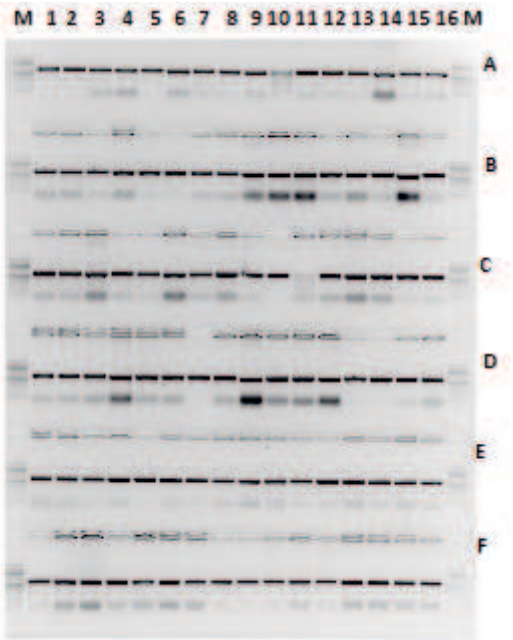

Fig. 2.4.7. Agarose gel 1.5%, M=100 bp ladder. Wells A 1-6, B1-6, C 1-16, D 1-16, E 1-16, F 1-16 are loaded with locus-specific PCR performed on colonies grown on PAB/CFC medium.

Bacteria identification with Sanger Sequencing

After the counting step, many colonies grown on the different media were sampled randomly, according to the different morphological characteristics. This analysis was performed on colonies of both the first and the second sampling (Table 2.3.2).

First we amplified the intergenic spacer of ribosomal genes 16S-23S (paragraph 2.3.3), which has the same length in bacteria of the same genus. In this way it was possible to highlight the species variability of the samples in order to select only the representative colonies to be sequenced by means Sanger approach. Figure 2.4.8 shows an example of gel electrophoresis in which amplification products of PCR L_1G_1 were loaded.

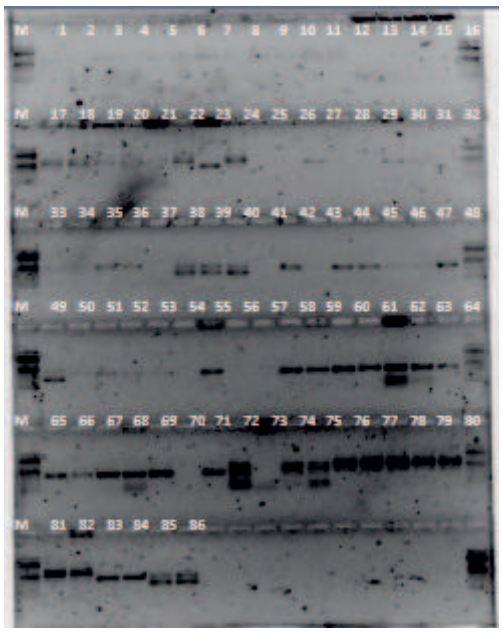

Fig. 2.4.8. Agarose gel 1.5%, M=100 bp ladder. Wells are loaded with L_1G_1 PCR performed on colonies grown on different media from 1° and 2° samplings.

43 representative colonies were selected and sequenced with the Sanger method on the 16S rDNA gene at BMR Genomics and the obtained sequences were compared with 16S database to identify bacteria.

	Species	Raw material	Sampling	Medium	Morphology
1	*Aeromonas* sp.	CR	1	Milk PCA	translucent
2	*Aeromonas* sp.	CR	1	Milk PCA	translucent
3	*Aeromonas* sp.	CR	1	Milk PCA	?
4	?	CR	1	Milk PCA	?
5	*Aeromonas* sp.	CR	1	Milk PCA	?

6	*Hafnia alvei*	PLW	1	Milk PCA	?
7	*Lactococcus lactis subsp. cremoris*	PLW	1	Milk PCA	?
8	*Hafnia alvei strain*	PLW	1	Milk PCA	?
9	*Microbacterium* sp.	PLW	1	Milk PCA	?
10	*Bacillus* sp.	PLW	1	Milk PCA	swarming
11	Streptococcus thermophilus	PLW	1	Milk PCA	?
12	*Bacillus cereus*	PLW	1	Milk PCA+starch	?
13	*Uncultured Streptococcus*	PLW	1	Milk PCA+starch	?
14	*Lactococcus lactis*	PCW	1	Milk PCA	big
15	*Hafnia alvei*	PCW	1	Milk PCA	swarming
16	*Hafnia alvei*	PCW	1	Milk PCA	?
17	*Lactococcus lactis*	PCW	1	Milk PCA	?
18	*Bacillus licheniformis*	PLW	1	Milk PCA	swarming
19	*Kocuria rhizophila*	PLW	1	MYP	Yellow with pink halo
20	*Macrococcus caseolyticus*	PLW	1	MYP	Irregular white-pink
21	*?*	PLW	1	MYP	Orange with white center
22	*Macrococcus caseolyticus*	PLW	1	MYP	Pinkish with halo
23	*Lactococcus lactis*	PLW	1	MYP	transparent
	Species	**Raw material**	**Sampling**	**Medium**	**Morphology**
24	*Kocuria rhizophila*	PLW	1	MYP	Pink-orange with white halo
25	*Lactococcus lactis*	PLW	1	MYP	transparent
26	*?*	PLW	1	MYP	white
27	*?*	PLW	1	MYP	Pink with white center

28	*Bacillus cereus*	PCW	1	MYP	big pink, white halo plus patina
29	*?*	PCW	1	*Milk PCA*+starch	big
30	*Bacillus licheniformis*	PCW	1	*Milk PCA*+starch	irregular
31	*Streptococcus thermophilus*	PLW	1	Milk PCA	Little white
32	*?*	CR	1	Milk PCA	Round transparent
33	*Pseudomonas psychrophila*	CR	1	MYP	transparent
34	*Pseudomonas sp.*	CR	1	Milk PCA	Round, radiant, dark center
35	*Bacillus*	CR	2	Milk PCA	Round, inibitory halo
36	*Leuconostoc lactis*	CR	2	SPS	small
37	*Streptococcus*	PCW	2	SPS	Flat, radiant
38	*Streptococcus macedonicus*	PLW	2	SPS	Raadiant, snow flake
39	*Bacillus cereus*	PCW	2	*Milk PCA*+starch	swarming with halo
40	*Bacillus cereus*	PCW	2	*Milk PCA*+starch	swarming irregular
41	*Leuconostoc mesenteroides*	PCW	2	*Milk PCA*+starch	Little, white, round
42	?	PLW	2	*Milk PCA*+starch	Little, white, round, dark center
43	*Shewanella*	PLW	2	*Milk PCA*+starch	Big light orange

Tab. 2.4.1. Summary table of the colonies sequenced in the first and second sample, indicating the observed phenotypic characteristics, the type of medium and the raw material from which they were isolated.

This analysis showed that the microbial flora of the cream was composed mainly of *Aeromonas sp.* and *Pseudomonas* (Fig. 2.4.9), typical environmental contaminants in this type of samples, which are erased through the production processes.

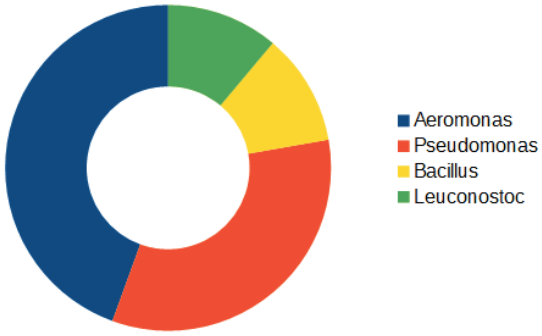

Fig. 2.4.9. Bacterial composition of CR samples, with Sanger approach.

The micro flora of the wheys was, however, much more variable and contained both lactic bacteria (which derive from milk and is resistant to pasteurization, such as *Streptococcus* and *Lactococcus*) and contaminants (*Enterococcacee*, *Hafnia, Kocuria, Macrococcaceae*) and spore-forming bacteria (*Bacillus*) (Fig. 2.4.10).

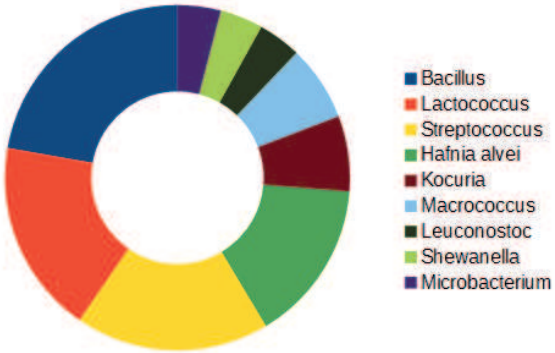

Fig. 2.4.10. Plot representing microbial composition of PCW and PLW samples with Sanger approach.

Community analysis with Next Generation Sequencing approach

Quantity and quality evaluation of extracted RNA

RNA quality was evaluated by means Nanodrop spectrophotometer. Nanodrop showed good absorbance ratios at 260/280 (absence of contamination of nucleic acids) and 260/230 (reduced contamination by solvents for some samples) between 1,7 and 2,2. All concentration values obtained ranged between 1 and 20 ng/µl, thus we decided to dilute the samples 1: 200 for RealTime – PCR analysis. After the reverse transcription step, we performed a Real-Time PCR on cDNA (rt) and RNA (*Nort*): as the *Nort* samples weren't amplified, RNA was not contaminated with genomic DNA.

Next generation sequencing results

The diluted cDNA underwent library preparation as described in paragraph Ch. 1.3.3. Resulting libraries were quantified, normalized and loaded on Illumina Miseq.

Data resulting from the run were then analyzed with QIIME as described in paragraph Ch1.3.5 with closed reference approach for OTU picking step (Ch. 1.3.6).

We obtained 2,601,176 raw reads and, after merging and quality filtering, reads were 1,037,630. Finally the number of OTUs spanned from 124 (2_CR) to 634 (3_PCW).

The most abundant phyla belonged to *Firmicutes*, *Proteobacteria*, *Bacteroidetes* and *Actinobacteria* (fig. 2.4.11).

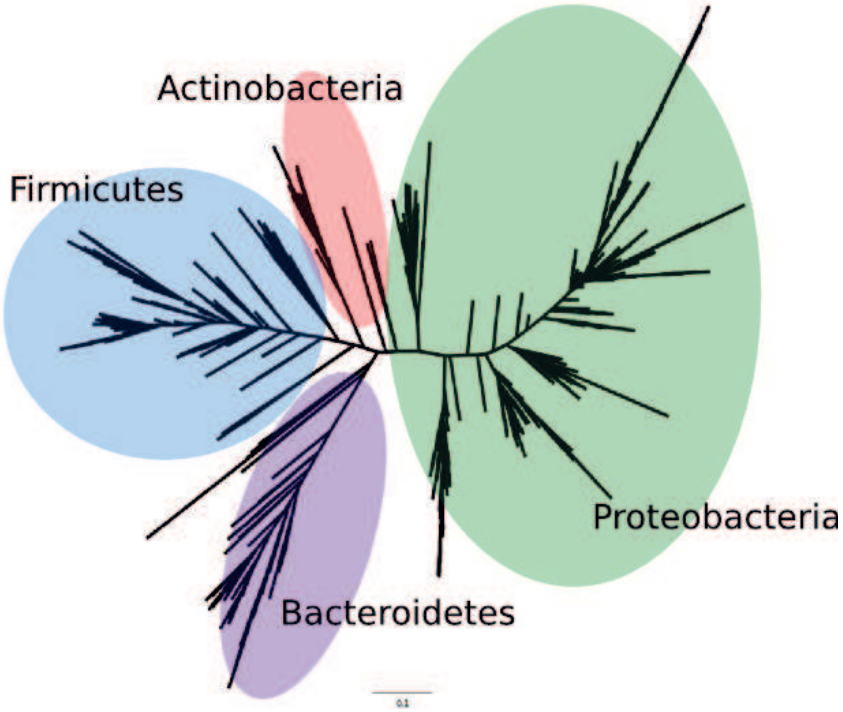

Fig. 2.4.11. Phylogenetic tree of the OTUs found in raw material samples. Colored cycles group the branches belonging to the same *Phylum*.

On average the most abundant bacteria in wheys (fig. 2.4.12) belonged to *Pseudomonas* (range 11% to 28%) and *Lactococcus* (range 0,6% to 49%), while *Lactobacillus* become the most abundant just in 3_PLW (74%), replacing *Lactococcus*. Streptococcus was abundant (29,2%) only in 2_PLW. Moreover *Anoxybacillus* become the most abundant sporeforming bacterium in 3_PCW (19%). Generally wheys show a biodiverse community, without reflecting any division based on sampling period. Finally a big chunk of viable bacteria is made by the class of *Gammaproteobacteria*, contaminants to which *Shewanellaceae*, *Moraxellaceae* and *Enterobacteriaceae* belong.

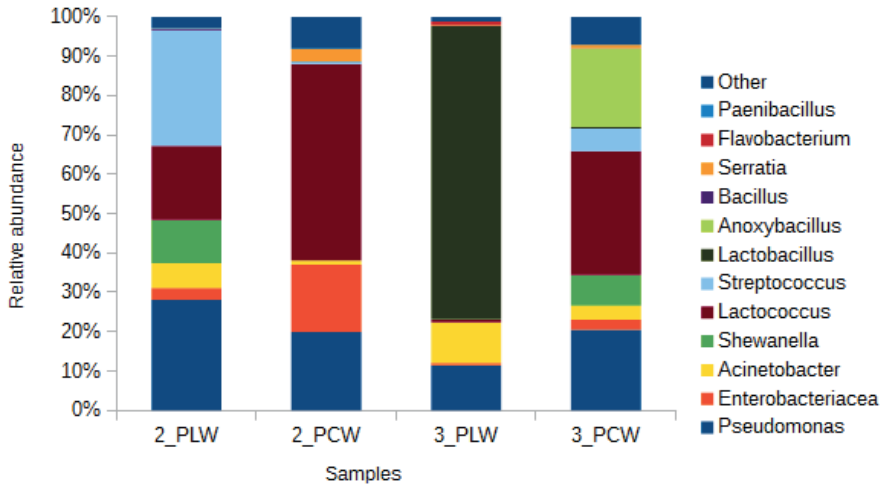

Fig. 2.4.12. Bar plot representing the relative abundance of the most bacteria present in wheys. Groups below 1% are clustered into the “Other” group.

On the other hand the most abundant bacteria in cream (fig. 2.4.13) was *Pseudomonas* (ranging from 52% to 86%) and *Lactococcus* (from 3% to 20%). Only in summer cream *Enterobacteriaceae* family becomes more abundant (from 0.3% in 2_CR to 12% in 3_CR) and *Leuconostoc* appears (1%) as well as *Shewanella* (7,5%). As shown with culture-dependent method, cream didn't bring any spores.

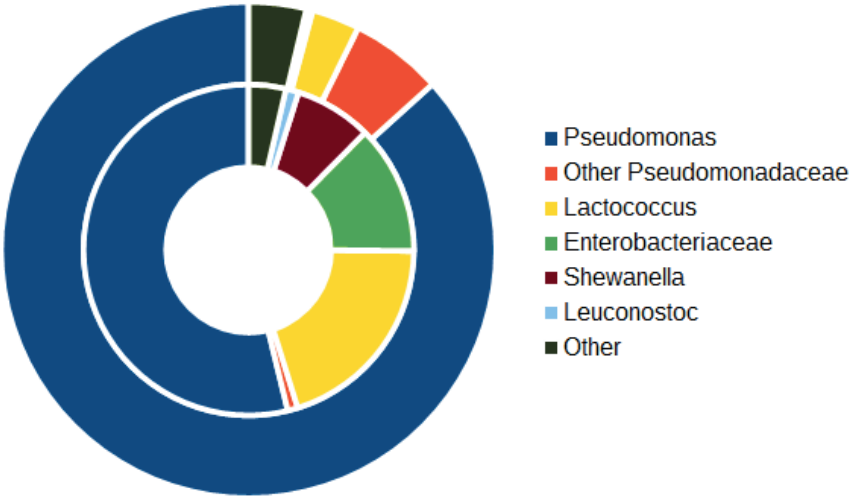

Fig. 2.4.13. Plot showing bacterial composition of CR samples: the external ring represents 2_CR, the internal one represents 3_CR.

2.4.3. Application of NGS and GS approaches to a Dairy business issue

NGS approach was applied to analyze microbial communities profiling 4 frozen wheys that complied with Company standards organoleptic qualities (1PWF, 2PWF, 3PWF, 4PWF) and 5 frozen wheys that did not comply with those standards (5PWFnc, 6PWFnc, 7PWFnc,

8PWFnc, 9PWFnc). The latter wheys showed altered organoleptic qualities: they presented smell and flavor described as taste of "grass" and "cooked". Whey 7PWFnc was the only to present a smell of “rancid/putrid”. As these samples were collected during summer 2013 and frozen, it was not possible to perform traditional microbiological analysis. Thus by means NGS and GS analysis we aimed to show which microbial community profile could provoke these kind of defects.

Evaluation of extracted DNA quality

DNA was extracted from wheys as described in paragraph Ch2-3.3.1 and in Ch1-3.4.2, in triplicate for each analyzed whey. Concentrations measured at the Nanodrop varied widely between 7.8 and 79.3 ng/ul. Quality absorbance ratios at 260/280 and 260/230 were not particularly good, suggesting RNA or other organic molecules contamination (Tab. 2.4.2).

		NANODROP			**REAL TIME**		
Sample	**Replicate**	**Conc.**	**260/280**	**260/230**	**Ct1**	**Ct2**	**Ct mean**
1PWF	1A	7,8	1,69	0,16			
	1B	15,1	2,11	0,48	12,49	12,79	12,64
	1C	9,2	3,09	0,11			
2PWF	2A	32	2,06	6,86			
	2B	24,2	1,67	1,68	12,35	12,08	12,22
	2C	32,8	2,02	0,14			
3PWF	3A	19	2,38	0,1			
	3B	25,1	1,66	2,22	18,17	18,21	18,19
	3C	24,6	1,95	6,88			
4PWF	4A	28	1,95	0,28			
	4B	18,6	2,06	0,24	17,47	17,52	17,5
	4C	23,7	1,81	0,29			
5PWFnc	5A	7,7	2,32	0,05			
	5B	13,6	1,77	2,13	15,63	15,72	15,68

	5C	12	2,35	0,06			
6PWFnc	6A	17,8	2,01	9,29			
	6B	13,8	1,77	6,91			
	6C	14,3	1,59	0,44	12,74	13,48	13,11
7PWFnc	7A	15,2	1,61	0,26	11,51	12,57	12,04
	7B	79,3	1,36	0,19			
	7C	66	1,35	0,19	10,79	10,45	10,62
8PWFnc	8A	7,8	1,55	2,16	19,1	21,2	20,15
	8B	18,7	1,98	0,06			
	8C	10,9	1,43	0,24			
9PWFnc	9A	12,6	1,55	0,34			
	9B	15,6	1,67	0,34			
	9C	35,6	1,7	2	13,15	11,89	12,52

Tab. 2.4.2. Table summary of measure made with Nanodrop and RealTime-PCR.

NGS library construction and bioinformatic analysis

DNAs were processed as described in paragraph 1.3.3. Resulting libraries were quantified, normalized and loaded on Illumina Miseq.

Data resulting from the run were then analyzed with QIIME as described in paragraph Ch1.3.5 with closed reference approach for OTU picking step (Ch1.3.6).

We obtained 7,925,033 raw reads and replicate reads were joined into one sample each. After merging and quality filtering, reads were 3,450,251. Finally the number of OTUs were 1154.

The most abundant genera in frozen wheys (Fig. 2.4.14) were *Streptococcus* (12%-82%), *Lactococcus* (11%-43%) and *Pseudomonas* (0,02%-56%). As they were distributed equally either in compliant and not compliant wheys, this suggests that they could not be responsible for the measured defects. However this does not exclude that on one hand, within those more abundant genera there may be some harmful species; on the other hand within less abundant genera there could be strong spoilers.

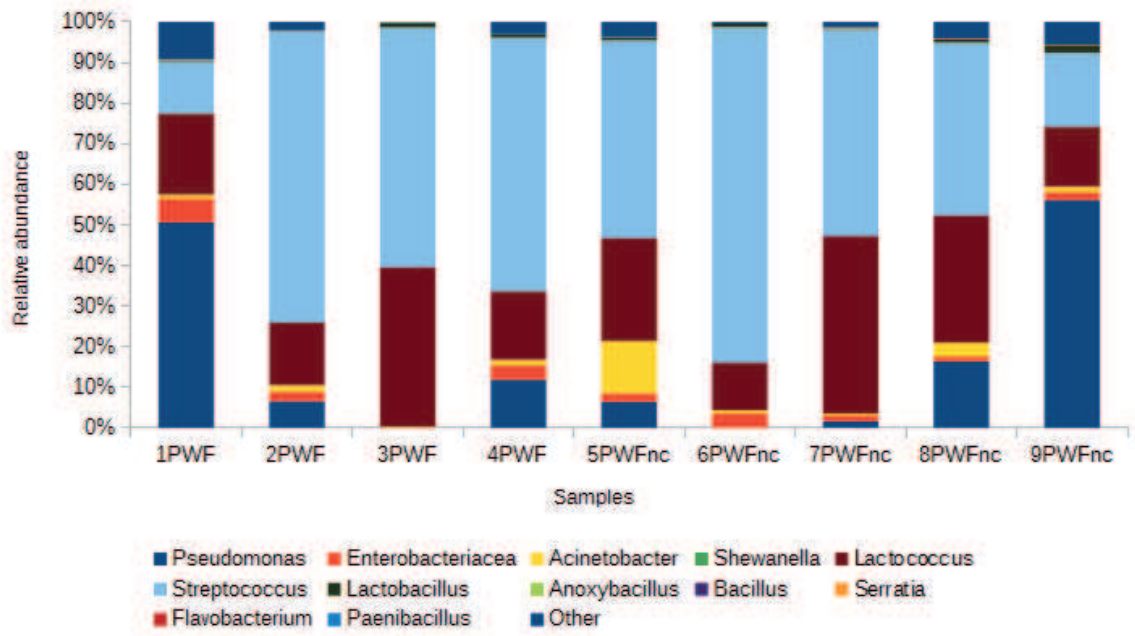

Fig. 2.4.14. Bar plot representing bacterial relative abundance in frozen wheys. Taxa below 1% was grouped into the "Other" cluster.

Figure 2.4.15 shows how there is not complete division between compliant and Not Compliant wheys. In fact 8PWFnc and 5PWFnc show a very similar microbial composition but are part of a group that also contains the compliant whey 4PWF. The same goes for samples 2PWF, 3PWF, 6PWF and 7PWF. The two batches 9PWFnc and 1PWF, however, are very close one each other and different from all the other.

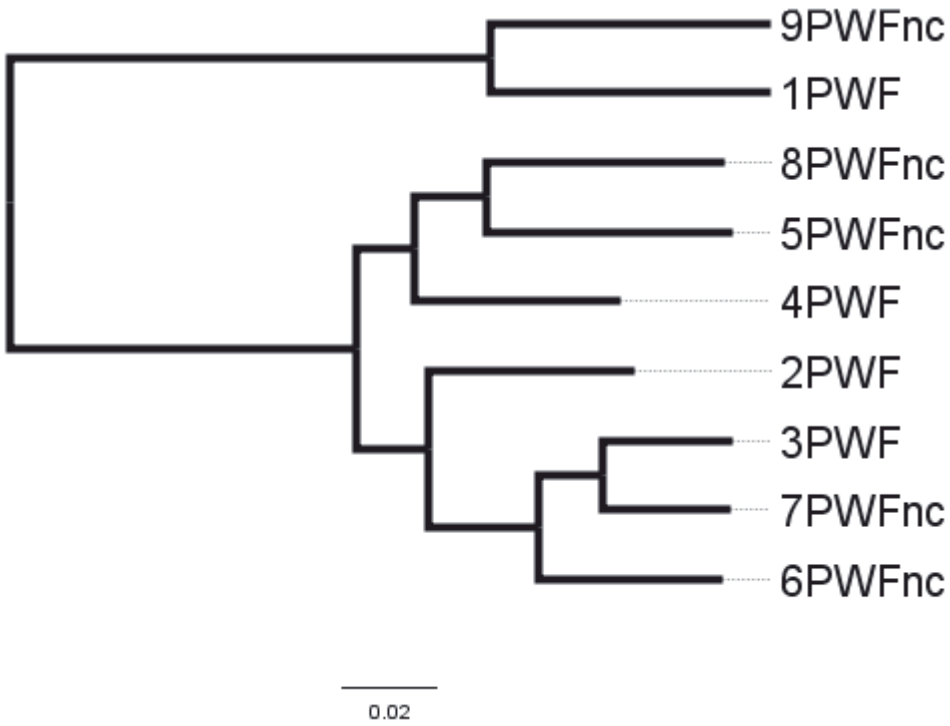

Fig. 2.4.15. Weighted Upgma tree showing microbial structure similarities between frozen whey samples.

Analyzing bacterial populations present in low percentage, we found that many of them were more abundant in not compliant wheys: in particular bacteria such as *Flavobacterium*, *Acinetobacter* and *Acetobacter* were 2 folds to 10 folds more abundant in bad ways. These genera are known for their ability to produce particular enzymes (proteases or lipases) that can modify the organoleptic characteristics of cheese products (Fig. 2.4.16).

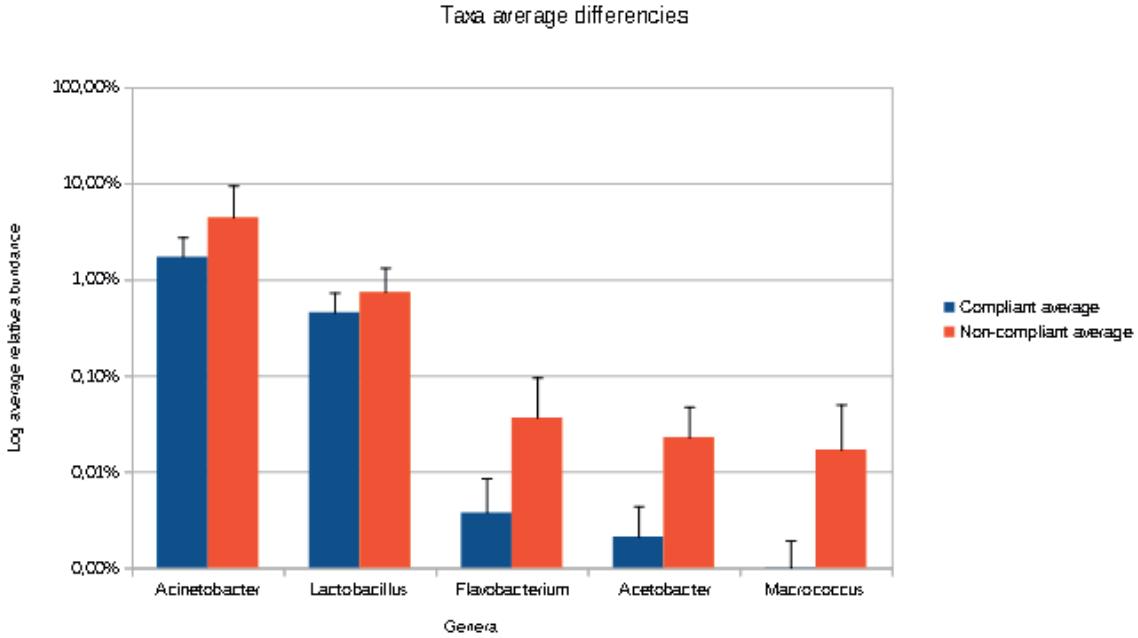

Fig. 2.4.16. Bar plot representing low abundance taxa differences between compliant and non compliant wheys. Y axis is in logarithmic scale and vertical bars represent standard deviation.

Gas chromatography analysis of frozen wheys

In order to obtain more information about the different molecules present in the two types of wheys, we performed a GS analysis. Our aim was to understand which bacteria may have been responsible for the chemical reactions and the resulting changes in organoleptic characteristics of the wheys. GS analysis was performed on all samples but 2PWF, 8PWFnc and 9PWFnc, because the company had delivered us just one sample for those wheys. Thus samples present once were used only for molecular microbiology analysis.

Figure 2.4.17 shows the most differently abundant compounds in 3 representative wheys 1PWF, 5PWFnc (grass/cooked defect) and 7PWFnc (rancid defect).

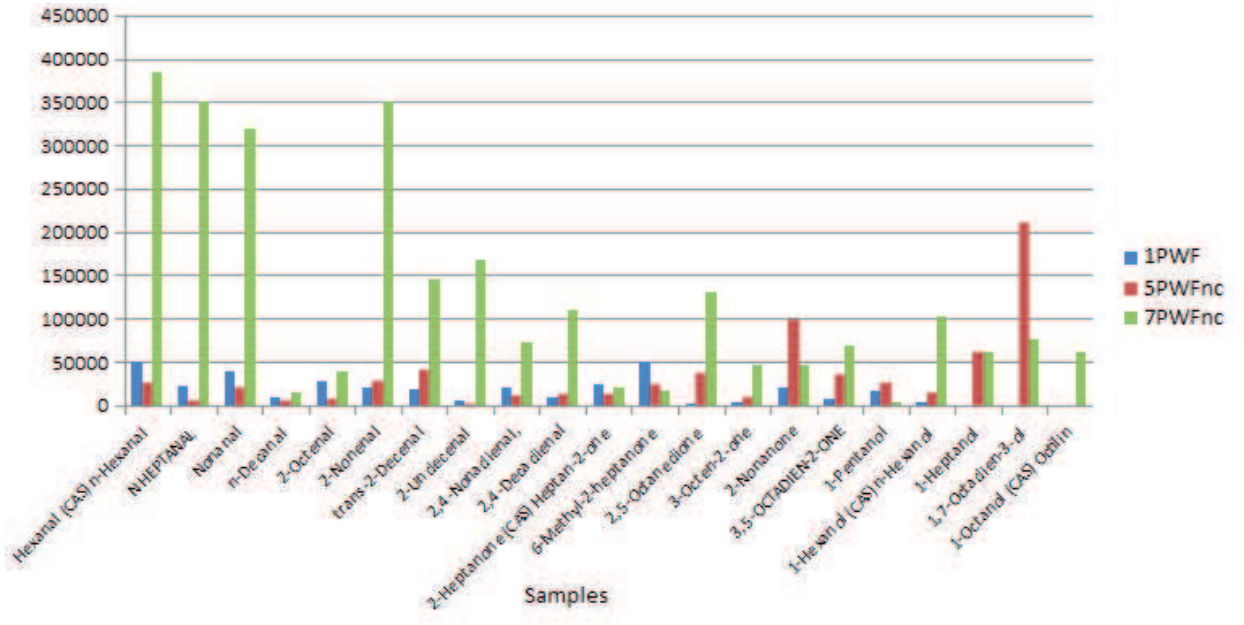

Fig. 2.4.17. Bar plot with the most abundant compounds found in three representative wheys. Y axis is the area of the compound peak.

The saturated and unsaturated aldehydes, ketones and alcohols (such as 2,4-nonadienal, 2-eptanon, optanol) are considered as an autoxidation product of the of the lipid fraction of the wheys. In fact they are present in low abundance in 1PWF while 7PWFnc whey presents a high abundance of these compounds, allowing to ascribe the non-compliance to a phenomenon of rancidity process.

On the other hand 5PWFnc whey is distinguished from other samples for the presence of two characteristic fermentation products, diacetyl (2,4-butanedione) and acetoin (3-hydroxy-2-butanone) fig. 2.4.18, which were not detected neither in compliant whey and in 7PWFnc. Thus this defect is mostly linked to fermentation processes.

In order to understand which could be the most driving bacteria in this processes we performed a Canonical Correspondence Analysis (cca) analysis and a correlation analysis on the taxonomy OTU table and the GC compounds matrices.

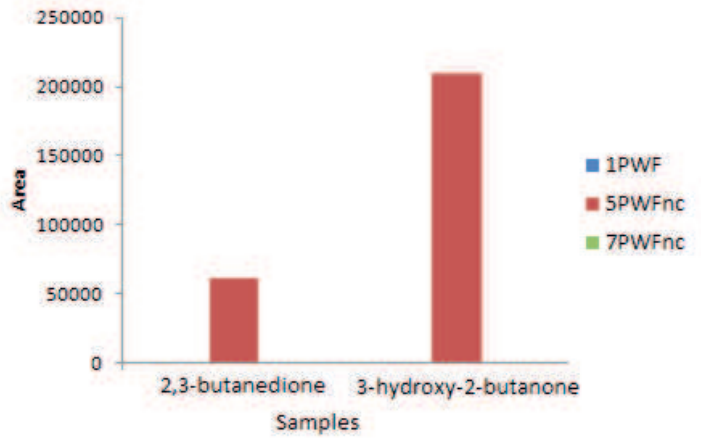

Fig. 2.4.18. Bar plot with the most particular compound found in 5PWFnc whey.

Statistical analysis CCA and correlation between chemical compounds and bacterial abundance

Through the correlation analysis plot (fig. 2.4.19) we can notice that compounds are divided into 3 main clusters driven by alcohols, aldehydes with ketons, and fermentation products. Above all 3-hydroxy-2-butanone and 2,3-butanedione are positively correlated with many *Gammaproteobacteria* such as *Aeromonas, Enterococcus, Acetobacter, Acinetobacter* and other *Pseudomonadaceae*. Moreover these compounds are found when starters as *Leuconostoc* and *Carnobacterium* are present. On the other hand most part of aldehydes and ketons are positively correlated with many spoilers bacteria such as *Macrococcus, Flavobacterium, Weissella, Staphylococcus* but also with starters as *Leuconostoc, Carnobacterium* and *Lactococcus*. Finally *Streptococcus, Salmonella, Ruminococcus* and *Corynebacterium* are positively associated with the presence of alcohols and negatively correlated with all the rest of compounds.

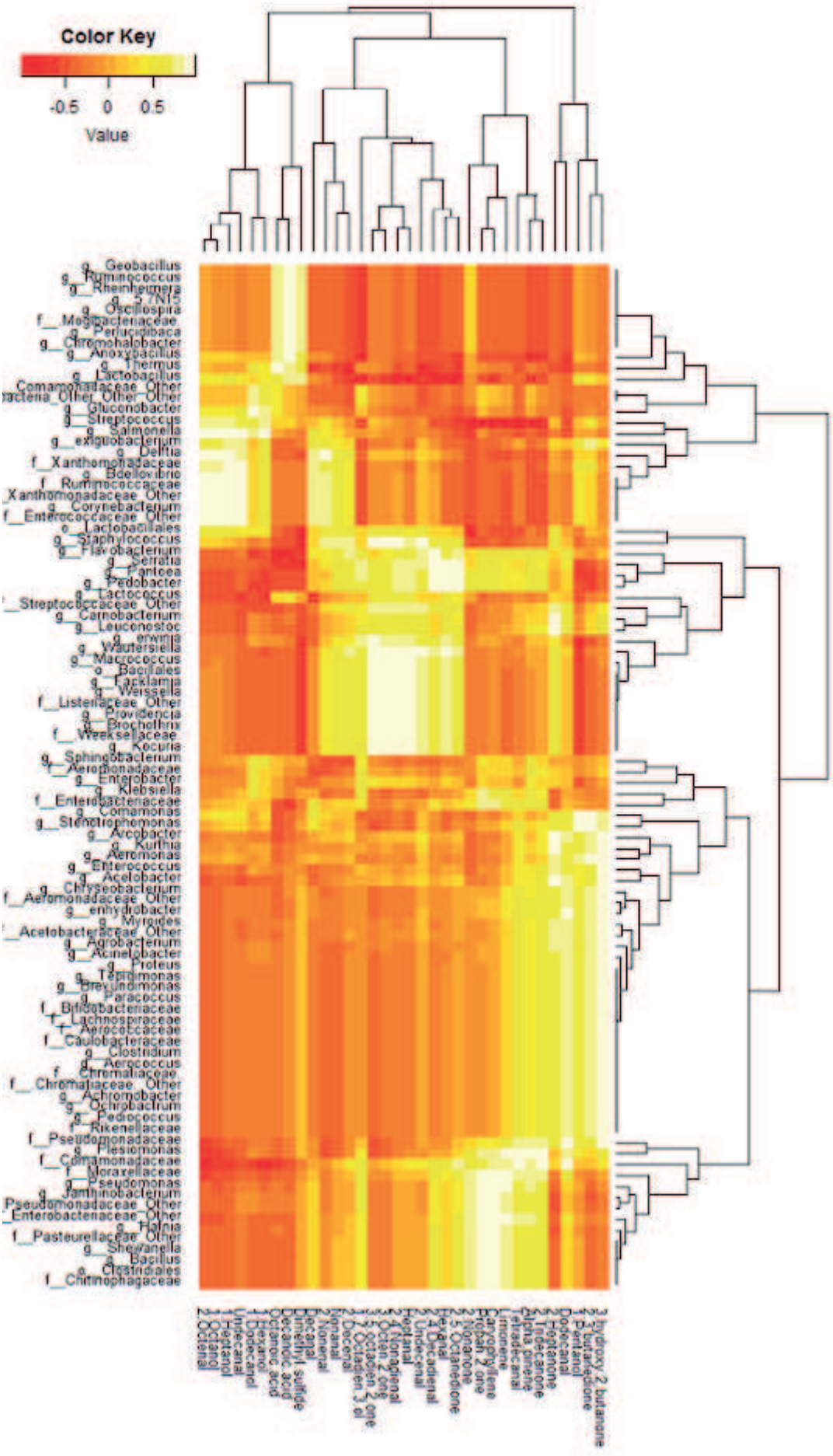
Color Key
-0.5
0
0.5
Value
g__Geobacillus
g__Ruminococcus
g__Rheinheimera
g__5.7N15
g__Oscillospira
f__Moraxellaceae
g__Perlucidibaca
g__Chromohalobacter
g__Anoxybacillus
g__Thermus
g__Lactobacillus
Comamonadaceae_Other
g__Gluconobacter
g__Streptococcus
g__Salmonella
g__exiguobacterium
g__Delftia
f__Xanthomonadaceae
g__Bdellovibrio
f__Ruminococcaceae
_Xanthomonadaceae_Other
g__Corynebacterium
f__Enterococcaceae_Other
o__Lactobacillales
g__Staphylococcus
g__Flavobacterium
g__Serratia
g__Pantoea
g__Pedobacter
g__Lactococcus
_Streptococcaceae_Other
g__Carnobacterium
g__Leuconostoc
g__erwinia
g__Wautersiella
g__Macrococcus
o__Bacillales
g__Lactalmia
g__Weissella
f__Listeriaceae_Other
g__Providencia
g__Brochothrix
f__Weeksellaceae
g__Kocuria
g__Sphingobacterium
f__Aeromonadaceae
g__Enterobacter
g__Klebsiella
f__Enterobacteriaceae
g__Comamonas
g__Stenotrophomonas
g__Arcobacter
g__Kurthia
g__Aeromonas
g__Enterococcus
g__Acelobacter
g__Chryseobacterium
f__Aeromonadaceae_Other
g__enhydrobacter
g__Myroides
_Acetobacteraceae_Other
g__Agrobacterium
g__Acinetobacter
g__Proteus
g__Terromonas
g__Brevundimonas
g__Paracoccus
f__Bifidobacteriaceae
f__Lachnospiraceae
f__Aerococcaceae
f__Caulobacteraceae
g__Clostridium
g__Aerococcus
f__Chromatiaceae
f__Chromatiaceae_Other
g__Achromobacter
g__Ochrobactrum
g__Pediococcus
f__Rikenellaceae
f__Pseudomonadaceae
g__Plesiomonas
f__Comamonadaceae
f__Moraxellaceae
g__Pseudomonas
g__Janthinobacterium
_Pseudomonadaceae_Other
_Enterobacteriaceae_Other
g__Hafnia
f__Pasteurellaceae_Other
g__Shewanella
g__Bacillus
o__Clostridiales
f__Chitinophagaceae
2-Octenal
2-Octanol
1-Heptanol
Undecanal
1-Dodecanol
1-Hexanol
Octanoic acid
Decanoic acid
Dimethyl sulfide
Decanal
2-Nonenal
Nonanal
2-Decenal
3-Octadecen-2-one
3-Octen-2-one
4-Nonadienal
Heptanal
2-Undecenal
2,4-Decadienal
Hexanal
2,5-Octanedione
2-Nonanone
Propan-2-one
Caryophyllene
Limonene
Tetradecanal
Alpha-pinene
2-Tridecanone
2-Heptanone
Dodecanal
1-Pentanol
3-hydroxy-2-butanone

Fig. 2.4.19. Heatmap showing the correlation of GC compounds with bacterial taxa of the frozen wheys.

CCA plot (fig. 2.4.20) shows that the explanatory variables (GC compounds) and qualitative variables (wheys) aren't divided into sharp groups thus either fermentation products and rancidity process products are found in compliant and non-compliant wheys equally. Octanoic acid and Decanoic acid are more likely to be found in 3PWF (sit2). As expected, many lipid oxidation products such as 2-Nonenal and 3-Octene2-one (mainly aldeides and ketons) can be found in particular in 7PWFnc (sit6) while many alcohols are found more likely in 4PWF (sit3), above all when they are coupled with a high abundance of *Streptococcus*. Interestingly Tetradecanal and Dodecanal seem to reach the maximum value in 1PWF (sit1), in couple with *Shewanella* and *Pseudomonas* genus. Finally in particular 3-hydroxy-2-butanone can be found mainly in 5PWFnc (sit4), and this separation is driven mainly by *Acinetobacter* genus.

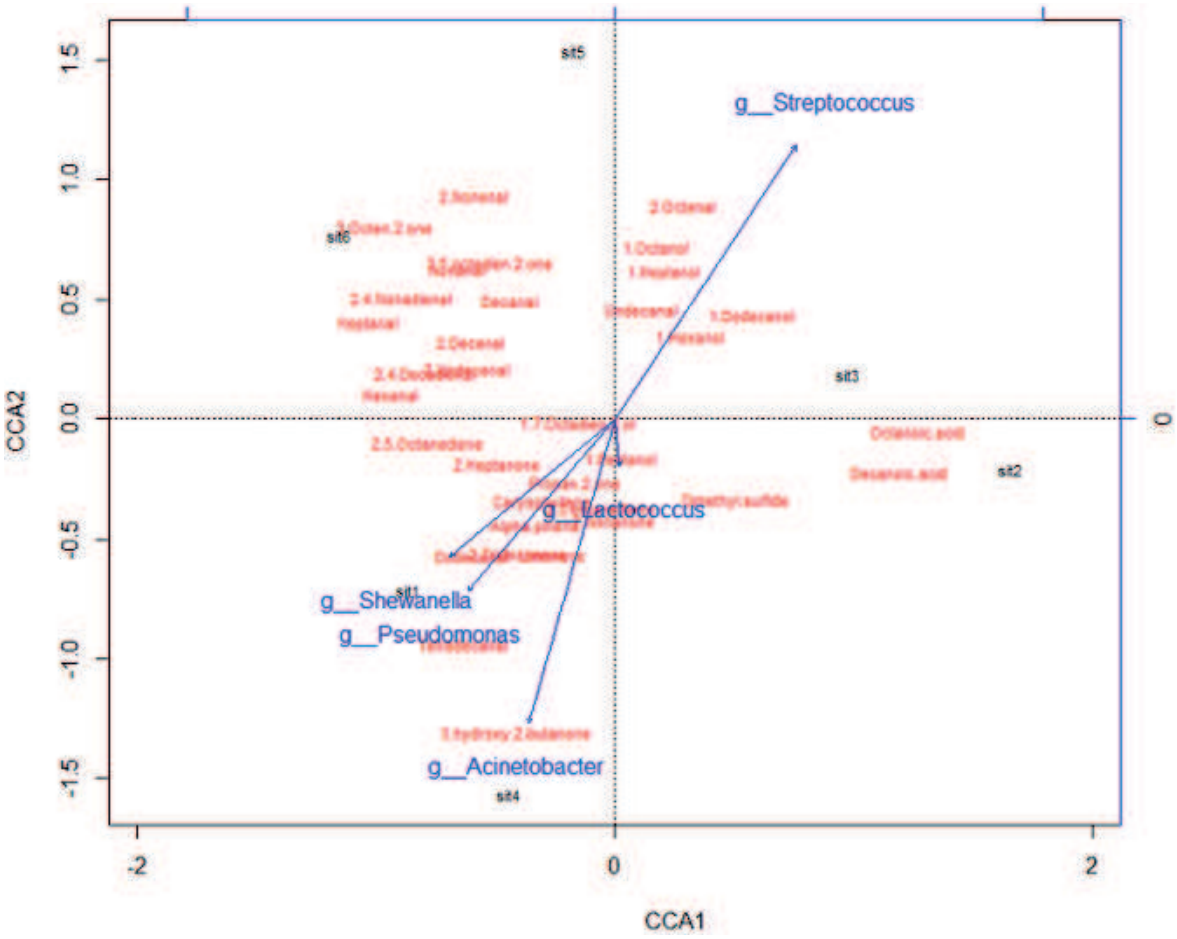

Fig. 2.4.20. Plot showing *cca* analysis data on GC compounds (explanatory variables, in red), wheys (qualitative variables in black, where 1PWF=sit1, 3PWF=sit2, 4PWF=sit3, 5PWFnc=sit4, 6PWFnc=sit5 and 7PWFnc=sit6) and OTU txonomy table (5 most important driving vectors in blue).

2.5. Discussion

We performed the analysis of raw material shelf-life on liquid whey (PLW), concentrated whey (PCW) and cream (CR) in order to establish a background of knowledge on the microbiological quality of these products before the next generation sequencing community characterization. All raw materials showed a very high mesophilic total charge (TMC), with values around 10^6 to 10^7 CFU/ml. Such high concentrations could derive from milk bacterial loads at the time of its use for dairy production and partly could due to downstream contamination of the production chain. In fact, several studies conducted on different types of milk showed that mesophilic total load ranged between 10^2 and 10^7 CFU/ml. Bacteria present belonged mainly to lactic acid bacteria, including *Lactobacillus* and *Lactcoccus* as well as *Enterobacteriaceae* and *Micrococcaceae* which can resist heat treatments [136]. Generally bacterial loads are 1 or 2 log CFU/ml higher in summer samples.

The number of aerobic spores was high either in PLW and PCW with bacterial loads ranging between 10^3-10^4 CFU/ml, whereas spores weren't detected in CR. Aerobic spores, produced primarily by *Bacillaceae* and *Paenibacillacee*, are commonly found in silage and can withstand the heat treatment and persist in dairy raw materials [145]. Contamination with these spores can lead to abnormal bacterial growth and to a premature spoilage of raw materials and derived food products. *B. cereus,* one of the most common spoilers in this field, can produce enzyme such as protease, lipase and phospholipase, that can cause changes in the composition of products and typical alterations in taste and in flavor [122].

The introduction of a specific medium for *Pseudomonadaceae* (CFC), from second sampling, also revealed a high charge of bacteria belonging to the genus *Pseudomonas.* This common spoiler was present at about 10^6 CFU/ml for the CR and between 10^4 and 10^7 CFU/ml respectively for PLW and PCW. Microorganisms belonging to this genus could be responsible for the release of heat-resistant proteolytic enzymes involved in the degradation of the fresh product or the raw materials [146, 147]. Nevertheless this bacterium could be involved in the production of blue pigments visible on the finished product [148].

Next Generation Sequencing (NGS) technology is a powerful tool to study in depth microbial communities of foods and is now considered a complementary technique to microbiology culture-dependent methods. As many microorganisms are difficult to be isolated with common cultural methods, the latter analysis could lead to significant underestimation of some bacteria of the microbial community present in the food matrix (Ch. 1.1) thus, overcoming this step, NGS approach can give better representation of the real communities [136]. Furthermore, in these last years with the new available platforms

this culture-independent method is producing larger volume of data at a decreasing price. For this reason on one hand this opens the possibility in research food microbiology to increase the sample size and the techniques or biological replicas for each sample as well as to increase the predictive ability of development of microbial populations in foods. On the other hand NGS could become the future of the diagnostic routine in application to food microbiology [62]. Thus in this study microbiological analysis of raw materials was supported with NGS analysis in order to provide a deeper and global overview of the microbial communities. NGS approach was also applied to evaluate the composition of the microbial community in frozen wheys with organoleptic defects.

2.5.1. Sanger and NGS comparison

Sanger and NGS approach were both exploited on 2_CR, 2_PLW and 2_PCW samples. The two methods showed high variability of microbial communities in raw materials: among them the community of both wheys was the most biodiverse and contained lactic acid bacteria (such as *Streptococcus* and *Lactococcus*), contaminating bacteria (*Enterococcacee*, *Hafnia*, *Kocuria*, *Macrococcaceae*) and spore-forming bacteria (*Bacillus*). The two approaches confirmed that the most abundant bacteria in wheys belonged to *Streptococcaceae family* (*Streptococcus* and *Lactococcus*) and *Gammaproteobacteria* (*Pseudomonadaceae, Shewanellaceae, Moraxellaceae and Enterobacteriaceae*). Moreover both approaches confirmed that CR community was primarily composed of *Aeromonas sp.* and *Pseudomonas*.

As expected the two approaches provided some different results in therms of bacterial abundance and taxa, above all *Bacillaceae* taxon. This difference is mainly due to the fact that i.e. aerobic spores were differentially detected by NGS approach in comparison with Sanger. On one hand spores were enriched and isolated on specific media so it was more probable to pick them up and sequence them. On the other hand one of the biases introduced by NGS approach lies on the incomplete DNA/RNA extraction method. However the metabolic activity of the spore is still uncertain and controversial, although it has been shown that bacterial spores contain some enzymatic activities [149]; thus RNA extraction and detection could fail due to the absence of active transcription. This was clear with the third sampling analyzed with NGS approach on RNA, in which 3_PCW showed a high abundance of *Anoxybacillus*, spore-former bacterium. Anyway it was not possible to understand if the bacterium was in the active state or the extraction kit performed better on it. Nevertheless NGS method remains more advantageous than Sanger because it can detect many bacteria present in very low abundance.

2.5.2. NGS and GC analysis on frozen wheys

NGS approach was applied to analyze microbial communities of 9 wheys collected during summer 2013 in the factory and stored frozen. The Dairy company eliminated 5 of these whey lots from the production line because they were not-compliant with the Company standards, on the basis of organoleptic properties. The results show that three particular microbial profiles were equally distributed in compliant and not-compliant wheys. However, some low abundance bacterial groups were more concentrated in the altered wheys. In particular, *Acinetobacter, Flavobacterium* and *Acetobacter* showed an average relative abundance, from 3 to 10 times higher in non-compliant wheys. These kinds of bacteria can be involved in fermentation processes that trigger the release of odorous compounds. Furthermore, these bacteria can express exogenous lipase [150] or heat-resistant protease [146], which can alter the organoleptic characteristics of food matrices. However, their low abundance is unlikely to be the cause of the defect, thus they could be a consequence rather than the cause of the alteration.

For a more precise defect identification we performed a gas chromatography analysis on 3 compliant and 3 non-compliant whey samples. As expected non-compliant wheys showed a different profile of volatile gases within them and between the complian wheys. In both 5PLWnc and 6PLWnc the analysis demonstrated the presence of compounds such as diacetyl and acetoin, that are responsible for the typical aroma of butter and suggest the presence of fermentative activity. This process could be provoked by some population of lactic acid bacteria used as starter during cheese-making and contained in large quantities in the whey. However in the case of 7PLWnc diacetyl and acetoin were not present: we detected some compounds, such as alcohols, aldehydes and ketones that could be attributable to processes of rancidity and lipid oxidation and can rise unpleasant flavors.

CCA and correlation analysis showed that many of the fermentation compounds were more associated with the presence of *Acinetobacter* and *Lactococcus*, along with a low abundance of *Streptococcus*. This latter genus seemed to drive the presence of alcohols in both compliant and non-compliant wheys. Moreover other lipase-producer-spoilers could provoke some degradation of fats, as some oxidation products are more likely to be present when *Shewanella* and *Pseudomonas* are abundant. An important focus on this analysis involves the usage of DNA as target molecule: as it allows to detect viable and non-viable cells, it is not possible to understand if all bacteria found were actually active. On the other hand high abundance in DNA mirrors high number of bacterial cells, supporting the hypotesis of bacterial elevate metabolic activity and replication. To this extent the most abundant bacteria such as *Streptococcus, Acinetobacter* and the others, were confidentially also metabolically active. Finally, many different aldehydes and ketones weren't driven by

the presence of any particular bacterium, supporting the hypotesis of spontaneous lipid oxidation.

Thus from this data in the case of 5PLWnc, it could be hypothesize that storage at inadequate temperature could allow bacteria to start fermentation. In the case of 7PLWnc an excessive pasteurization was applied to sanitize that already spoiled the batch and a subsequent inadequate storage could spur the rancidity process. In conclusion it appear clear that the different composition of volatile products can be influenced by the starter cultures used in the production of cheese and that the under-hygienic storage conditions can lead whey to spoil [125].

2.6. Conclusions and future perspectives

Traditional microbiology is certainly a starting point to define certain microbiological conditions and is still a relatively low-cost method with which it is possible to analyze multiple samples simultaneously. Molecular methods often support classic methods to define the characteristics of the isolated colonies in less time and with greater accuracy than biochemical methods.

Supporting microbiological analysis of raw materials for the production of Ricotta with a thorough NGS analysis has allowed us to introduce new and more complete information about microbial communities present in these food matrices. We showed that NGS approach is an effective tool for the study of food microbial communities: it is complementary to culture-dependent methods, providing most data that are under-detectable with common analysis methods.

In particular the application of this tool to a business issue on organoleptic alterations of wheys allowed to evaluate the defects found and to deduce the role of specific microbial populations in spoilage processes. The results obtained in previous and this studies demonstrate the applicability of this approach also to other foods or to entire food chain production. Moreover 16S approach can be integrated with other NGS analysis such as metagenomic and metatranscriptomic methods that could be very useful tools for a better understanding of food environment and microbial communities.

3. Chapter III

Microbial dynamics during shelf-life of industrial Ricotta cheese and identification of a *Bacillus* strain as a cause of a pink discoloration.

Sattin E.[1], Andreani N.A.[2], Carraro L.[2], Fasolato L.[2], Balzan S. [2], Novelli E.[2], Squartini A.[3], Telatin A. [1], Simionati B.[1], Cardazzo B.[2]

[1] BMR genomics, Padova, Italy

[2] Department of Comparative Biomedicine and Food Science, University of Padova, Italy

[3] Department of Agronomy, Food, Natural Resources, Animals, Environment (DAFNAE), University of Padova, Italy

Adapted from the original manuscript published in **Food Microbiology**

Article reference: YFMIC2509

Journal title: Food Microbiology

Corresponding author: Dr. B. Cardazzo

First author: Dr. E. Sattin

Final version published online: 12-JAN-2016

Full bibliographic details: Food Microbiology (2016), pp. 8-15

DOI information: 10.1016/j.fm.2015.12.009

3.1. Abstract

Dairy products are perishable and have to be preserved from spoilage during the food chain to achieve the desired shelf-life. Ricotta is a typical Italian soft dairy food produced by heat coagulation of whey proteins and is considered to be a light and healthy product. The shelf-life of Ricotta could be extended, as required by the international food trade market; however, heat resistant microflora causes spoilage and poses issues regarding the safety of the product. Next-generation sequencing (NGS) applied to the Ricotta samples defined the composition of the microbial community in-depth during the shelf-life. The analysis demonstrated the predominance of spore-forming bacteria throughout the shelf-life, mostly belonging to *Bacillus, Paenibacillus* and *Clostridium* genera. A strain involved in spoilage and causing a pink discolouration of Ricotta was isolated and characterized as *Bacillus mycoides/weihenstephanensis*. This is the first report of a food discoloration caused by a toxigenic strain belonging to the *Bacillus cereus* group that resulted the predominant strain in the community of the defective ricotta. These results suggest that the processing of raw materials to eliminate spores and residual micro flora could be essential for improving the quality and the safety of the product and to extend the shelf-life of industrial Ricotta.

Keywords: Ricotta; shelf-life; microbial community; spore forming bacteria; pink discoloration

3.2. Introduction

Ricotta is a soft, white, fresh dairy product with a slightly sweet flavor typical of Italy and Ibero-American countries, manufactured from bovine, sheep, buffalo or goat whey milk by heat-coagulation of the whey proteins. The production of Ricotta cheese is a system for enhancing value of a cheese making residue, largely applied in Italy both at the artisan or industrial level. Fifteen percent of the whey produced in Italy is intended for the production of Ricotta. The amount of Ricotta consumed can be estimated at about 55,000 tons (retail), roughly 7% of cheeses purchased. Moreover, sales data show that Italian consumers seem to prefer Ricotta prepackaged instead of as a bulk product sold over a cheese counter with attendant [151]. Due to its low fat and salt content, high protein content and easy digestibility, Ricotta could respond to the demands of consumers and the market for light and healthy products. Ricotta can be eaten as a soft cheese even if it is more frequently used as an ingredient in dishes and desserts.

The name is derived from the Latin word *re-coctus*, literally re-cooked or cooked twice. Ricotta production technology uses the principle of coagulation and precipitation of the whey protein (mainly globulin and albumin) favored by whey acidification (pH <4.6) and heating. Ricotta is manufactured by thermally treating the whey at 80–90 °C followed by

the addition of lactic or citric acid (1.5–2.5%). Afterward, the surfacing curd is collected in moulds to partially remove whey and cooling. Whey could be enriched with whole raw milk or cream (1-5%) and salt (0.5–1.5%) to increase the yield and improve the organoleptic characteristics [152, 153].

Dairy products are characterized by a reduced shelf-life because they are an excellent growth medium for a wide range of microorganisms [154]. The spoilage process is produced as a consequence of food contamination by bacteria and fungi in raw materials or during production steps. Many of these microorganisms can produce undesirable reactions that deteriorate flavor, odor, color, and the sensory and textural properties of foods. In addition, some microorganisms can potentially cause food-borne illness [155, 156].

Ricotta has a high moisture content, high concentration of residual sugars, an initial pH above 6.0, and starters are not added during the production. As a consequence of these properties, Ricotta has a limited shelf-life even under refrigeration [157, 158]. The Ricotta industrial production process, that includes heat treatment applied during manufacture and a final pasteurization step, inactivates most of the resident micro flora and limits the post-processing contaminations. For this reason, the shelf-life of industrial Ricotta is considered to be between 20 and 40 days [99].

However, with the globalization of the food trade and the distribution from centralized processing, the request for an extended shelf-life of food products is becoming essential and a pressing issue. On the other hand, spores and thermophilic bacteria resist the production process and, with the extension of shelf-life, enhance spoilage events and pose issues regarding the safety of the product. Thus the extension of shelf-life remains a challenging goal [160, 161]. The chance of spoilage and pathogenic microorganism growth and survival depends on extrinsic factors associated with production and storage conditions, but also on intrinsic factors such as the composition of the microbial community [162].

Knowing the composition of the food microbiome is very important for defining the safety and the quality of a food product. In fact next-generation sequencing (NGS) platforms is an interesting approach for food microbiology, allowing deep microbial community definition directly on food samples [163, 159]. The analysis is based on millions of sequence reads obtained in a single run of 16S rRNA gene amplicons. 16S rRNA sequences are clustered into similarity groups, defined as Operational Taxonomic Unit (OTU), and classified by comparison against 16S rRNA sequence database. The strong competition between manufacturers has resulted in sustained technical improvements and cost reduction of almost all NGS platforms, allowing a wider usage of these technologies and providing a more complete description of the microbial community and its interaction and evolution

[66, 164].

In the present study, the NGS approach was applied to evaluate the composition and evolution of the microbial community during the shelf-life of an industrial Ricotta. The NGS approach provided a very comprehensive view of the microbial population composition and demonstrated that the improvement and application of such techniques on food microbiology could be an excellent method to evaluate food microbiological quality in-depth. Moreover the identification of the critical steps of the production process could be essential to control microbial load and to suggest solutions for safe food production. During the shelf-life study, a package of Ricotta revealing a pink discoloration was analyzed leading to the identification of the bacterial strains involved in the spoilage.

3.3. Materials and methods

3.3.1. Samples

Two lots of bovine Ricotta were supplied from the Ricotta factory “Elda” (Vestenanova, VR, Italy) in January 2014 (winter samples named W) and July 2014 (summer samples named S). Ricotta is produced at 90 °C and sealed in plastic food packages of 100 g just after a pasteurisation step (1 min at 80 °C). The recipe includes pasteurized whey, cream at final concentration of 20%, lactic acid 0.1% and salt 0.1%. Twenty packages of Ricotta W and twenty-six of Ricotta S were collected in the factory the day after the production, transported to the laboratory in refrigerated containers and stored at 8 °C in a refrigerated incubator (MPM Instruments, Bernareggio, Italy) for 60 days. The temperature of 8 °C was chosen as mild thermal abuse to simulate the condition frequently occurring to the product during its commercial life.

After 4, 11, 14, 21, 25, 32, 39 and 60 days of storage, three packages of Ricotta were analyzed (except for 39_S for which four samples were analyzed). Each Ricotta sample was named with the day from production, the lot (W and S) and a progressive number. Sample 14_S3, presenting a pink discoloration, was processed with other samples, but an additional Ricotta package, 14_S4, was sampled and analyzed at the same day.

3.3.2. Microbiological analysis

Twenty grams of Ricotta were added to 180 mL of Buffered Peptone Water (BPW Biokar Diagnostics, Beauvais Cedex, France) and serially diluted in the same solution. Samples were analyzed for total aerobic mesophilic microorganisms (Total Mesophilic Count, TMC) plating on Plate Count Agar with skimmed milk (milkPCA, Biokar Diagnostics). The plates were incubated at 30 °C for 24–48 h. For aerobic spore count (ASC), 10 ml of 1:10 diluted samples were treated for 10 min at 80 °C and plated in Plate Count Agar

added with 0.2% Starch (sPCA, Biokar Diagnostics) and incubated at 30 °C for 2–5 days.

The Ricotta sample 14_S3 (presenting pink discolouration) was analyzed also for yeast and mould counts on Oxytetracycline Glucose Yeast Extract Agar (OGYE; Oxoid Microbiology Products, Thermo Scientific, Waltham MA, USA) and incubated for 3–6 days at 25 °C; for *Micrococcaceae* on Mannitol Salt Agar (MSA Oxoid Microbiology Products, Thermo Scientific); for *Lactobacilli* on MRS agar (MRS agar Oxoid Microbiology Products, Thermo Scientific); and for *Pseudomonas* on *Pseudomonas* Agar Base (PAB; Oxoid Microbiology Products, Thermo Scientific) incubated at 22 °C for 24–72 h. Moreover, with the aim to recover pink colonies, the 14_S3 sample diluted in BPW was plated in Minimal Bacterial Medium Agar (MBM 0.7% K2HPO4, 0.3% KH2PO4, 0.05% tri-sodium citrate, 0.01% MgSO4, 0.1% (NH4)2SO4, 0.2% glucose and 1.5% agar) and a Tryptic Soy Agar (TSA Oxoid Microbiology Products, Thermo Scientific) medium. Two pink colonies were resuspended, each in 100 µl of buffered peptone water and inoculated with a sterile syringe in two packages of Ricotta. The packages were opened after 14 days of incubation at 8 °C.

3.3.3. pH and organic acid determination

The pH of Ricotta samples was determined using a Portamess pH-meter (Knick 910, Berlin, Germany) equipped with an INLAB 427 electrode (Mettler Toledo, Urdof, Switzerland). With the purpose of identification and quantification of organic acids, the samples of Ricotta were finely ground and diluted in a ratio 1:5 with the mobile phase (Sulfuric acid 0.01 N), held at 40 °C for 20 minutes, homogenised for 5 minutes with the Stomacher, centrifuged at 10,000 rpm at 4 °C for 10 minutes, and filtered first through a Whatman 4 and then through a syringe filter of 0.45 microns. The organic acids were determined by liquid chromatography with a Bio-Rad HPLC system equipped with a titanium pump mod.1350T and UV detector with variable wavelength mod. 1706. The column was an Aminex HPX-87H 300 x 7.8 mm (Bio-Rad) kept at a temperature of 60 °C, and mobile phase 0.01 N sulfuric acid flowing at 0.6 ml/min. A sample volume of 20 ul was injected by an autosampler plus 717 WISP (Waters) and the data were acquired and processed with the Empower 2 software (Waters). The working absorbance was at a wavelength of 210 nm and the quantification of the different organic acids was carried out by external calibration.

3.3.4. RNA extraction and reverse transcription

For RNA extraction, 100 mg of Ricotta suspended in 700 µl Buffer RLT (Qiagen, Hilden, Germany) and extracted using RNeasy Mini Kit (Qiagen) following the manufacturer's protocol. The concentration and purity of RNA were analyzed using a NanoDrop ND-1000

(Thermo Scientific). 10 µl of the extracted RNA (20 to 100 ng, depending on the yield of the samples) were reverse transcribed into cDNA using *Superscript II* (Invitrogen, Carlsbad, CA, USA).

To determine the exact aliquot to be used for the NGS library construction, an aliquot (2.5 µl) of diluted (1:200) cDNA template was amplified in a final volume of 10 µl containing 5 µl of Platinum SYBR Green qPCR Supermix UDG (Invitrogen), and 0.25 µl of each primer 331F and 797R (10 µM) (Nadkarni et al 2002). ROX fluorochrome was used as an internal check. The amplification protocol consisted of an initial step of 2 min at 50 °C and 2 min at 95 °C, followed by 40 cycles of 15 s at 95 °C and 30 s at 60 °C. All experiments were performed on an MX3000P machine (Stratagene, La Jolla CA, USA).

3.3.5. Dual index 16S amplicon libraries preparation and bioinformatic analysis

Sequencing protocol was performed at BMR Genomics srl (Padova, Italy). For every sample, V3-V4 regions of 16S rRNA gene were amplified using 331F primer and 797R [11]. Primers were modified with forward overhang:

5’-TCGTCGGCAGCGTCAGATGTGTATAAGAGAAG -[locus-specific sequence]-3' and with reverse overhang:

5’-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-[locus-specific sequence]-3', necessary for dual index library preparation. After indexing and pooling steps, libraries were loaded on two different runs on Illumina MiSeq with 2x300 bp paired end approach.

Sequencing reads were filtered for average quality (Q>30), and R1 and R2 were merged by means FLASH with default parameters [4]. QIIME [68] version 1.8 was used to perform the full analysis, using the *pick_de_novo_otus* wrapper for OTU picking (*uclust* clustering method, [17]), assigning Greengenes (version 13_8) taxonomy to OTUs, creating the OTU table and phylogenetic tree. The following wrappers with default parameters were used for the other steps: *summarize_taxa_through_plots* (to produce the taxonomy files and charts), alpha_rarefaction and *beta_diversity_through_plots* (to assess, respectively, the alpha- and beta- diversity indices).

A two-sided Student's two-sample statistical t-test on the Euclidean distance between all samples was performed to evaluate β-diversity significance between seasonal composition of ricotta cheeses.

All data on microbial composition were deposited in SRA database (SRP060430).

3.3.6. Species identification, MLST and toxin analysis of pink colonies

Molecular identification of pink isolates was carried out by 16S rRNA gene sequencing. For DNA extraction, a single colony from a fresh culture was resuspended in 100 µl nuclease-free water, vortexed at high speed for 5 s, and incubated at 94 °C for 10 min. The tube was vortexed again and centrifuged for 2 min at 14,000 rpm. The supernatant was transferred to a fresh tube and stored at -20°C. The concentration and quality of extracted DNA were determined using NanoDrop ND-1000 (Thermo Scientific). About 25 ng of DNA was subjected to partial 16S rRNA gene amplification with primers 16S rRNA F and 16S rRNA R [165]. The amplified fragments were sequenced and the sequences obtained were aligned with the closest sequences available in the GenBank database (98% of homology; http://www.ncbi.nml.nih.gov/BLAST). The MLST approach was applied to the DNA extracted from pink colonies using primers and conditions reported in the primers section of the *B. cereus* MLST database (http://pubmlst.org/bcereus/) and in Cardazzo et al. (2008) [141].

The visualization, analysis and editing of the chromatograms obtained for the seven genes were performed with FinchTV 1.4.0 software (Geospiza, Seattle, USA). The sequence of each allele was compared to those available in the *B. cereus* MLST database by using the BLAST program and the Sequence Type (ST) was defined. The new STs were submitted to the *B. cereus* MLST database. The concatenated sequences (*glp-gmk-ilv-pta-pur-pyc-tpi*) of ST1008 and ST985 were aligned with the concatenated sequences obtained from a selection of complete genome sequences of the *Bacillus cereus* group species deposited in GenBank. The complete list with accession numbers is reported in Table 2S. The phylogenetic analysis was conducted with the software MEGA 5.0, using the neighbor-joining algorithm.

Species	Strain	Accession number
Bacillus anthracis	Han	CP008854
Bacillus cereus	G9842	NC_011772
Bacillus cereus	FM1	CP009369
Bacillus cereus	G9241	CP009590
Bacillus cereus	03BB87	CP009318
Bacillus mycoides	ATCC 6462	CP009692
Bacillus thuringiensis	HD1002	CP009720
Bacillus thuringiensis	HD789	NC_018508
Bacillus thuringiensis	BGSC4AA1	CP018500
Bacillus thuringiensis	HD771	NC_018500
Bacillus thuringiensis	HD1	CP004870
Bacillus thuringiensis	MC28	NC_018693
Bacillus thuringiensis	HD1011	CP010089
Bacillus toyonensis	BCT7112	NC_022781
Bacillus weihenstephanensis	WSBC10204	CP009746
Bacillus weihenstephanensis	KBAB4	NC_010184

Table 2S. Complete genome sequence accession numbers of Bacillus strains included in the phylogenetic analysis

For toxin gene analysis, portions of the *hblA, hblC, hblD, nheA, nheB, nheC, entFM, bceT* genes and the emetic-specific sequence (em) were amplified using the set of primers listed in Yang et al (2005) [166], with the 16S-23S rRNA Internal Transcribed Sequence (ITS) used as an internal control. The amplification was performed as described in Cardazzo et al. (2008). DNA extracted from *Bacillus cereus* group strains, positive to toxins [141], was used as positive control for PCR amplifications.

3.4. Results and discussion

3.4.1. Microbiological counts and pH values of Ricotta samples

The microbiological and chemical analysis on W and S lots were conducted to give a general view of the products. The Ricotta samples of lot W and S were analysed by culture-dependent methods to describe the general microbiological condition (TMC and ASC) of samples. As shown in Figure 1A, the TMC are between 10^4 and 10^8 CFU/ml and the ASC between 10 and 10^5, demonstrating small variability due to the seasonality with a slightly anticipated increase of TMC in S samples, probably due to the higher temperature during summer season.

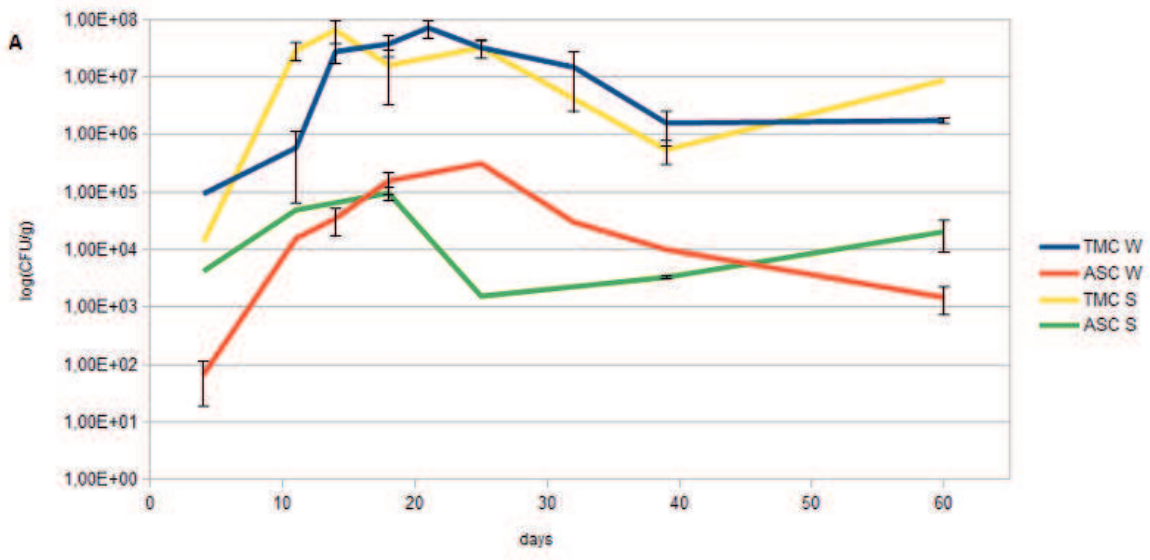

Figure 1A. TMC, aerobic spore counts in ricotta cheese samples during shelf life (from 4 to 60 days, the medium value and standard deviation were calculated among the three ricotta samples 1, 2 and 3 of the same time point).

The pH values are presented in Figure 1B as average values of the three samples collected at each time. The Ricotta samples started with a pH values of 6.3–6.4 (as expected for this dairy product, [99]) but after 14 days it decreased to under 6. Such a decrease resulted similarly in W and S samples and concurrently to the increase of the TMC in both lots, with an anticipated decrease in S samples.

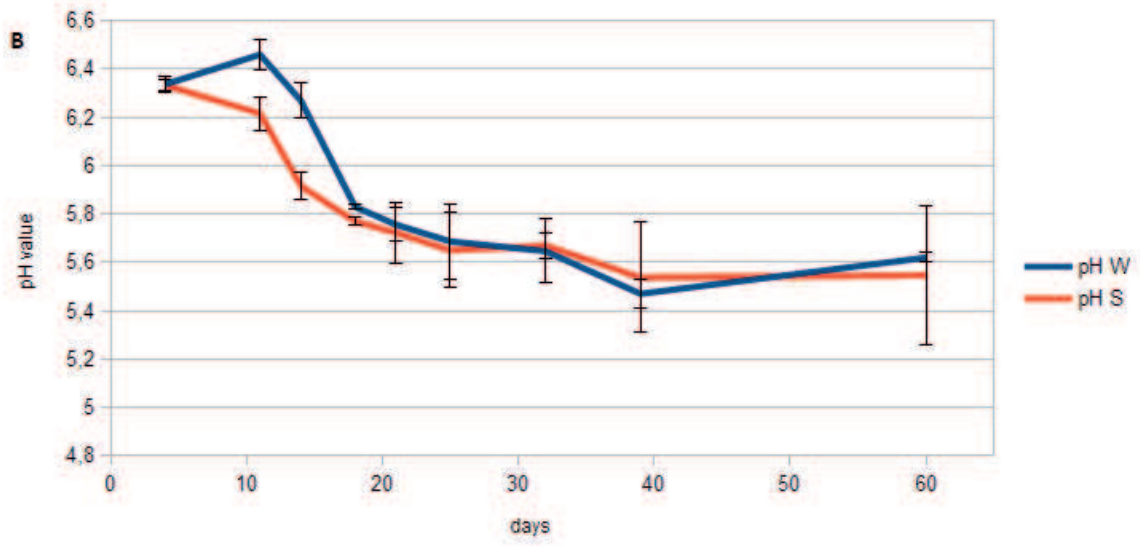

Figure 1B. PH values in ricotta cheese samples during shelf life (from 4 to 60 days, the average value and standard deviation were calculated among the three ricotta samples 1, 2 and 3 of the same time point).

All the counts and pH values with average and standard deviation are reported in Table 3S. The parameters TMC and pH are routinely evaluated by the factory for each lot produced, and the two lots investigated resulted as the standard production of the factory. The increase of the TMC after 14–18 days must be ascribed mostly to the development of heat resistant micro flora (thermophilic bacteria and spores) that survived to the heat processing during production and pasteurization [167].

	TMC			ASC			pH		
sample	counts	average/day	St dev	counts	average/day	St dev	values	average/day	St dev
4_W1	2,73E+03	9,32E+04	1,56E+05	9,09E+01	6,67E+01	5,84E+01	6,380	6,337	0,038
4_W2	2,73E+05			1,09E+02			6,320		
4_W3	4,00E+03			0,00E+00			6,310		
11_W1	1,59E+04	5,87E+05	6,40E+05	1,56E+04	1,56E+04		6,450	6,460	0,009
11_W2	1,28E+06			-			6,465		
11_w3	4,68E+05			-			6,465		
14_W1	2,90E+07	2,77E+07	1,33E+07	4,45E+04	3,49E+04	2,17E+04	6,320	6,270	0,161
14_W2	1,38E+07			1,00E+04			6,090		
14_W3	4,04E+07			5,01E+04			6,400		
18_W1	3,00E+07	3,76E+07	1,83E+07	2,50E+05	1,56E+05	8,11E+04	5,770	5,830	0,053
18_W2	2,44E+07			1,09E+05			5,870		
18_W3	5,85E+07			1,10E+05			5,850		
21_W1	1,03E+08	7,15E+07	2,97E+07	-			5,700	5,757	0,060
21_W2	6,82E+07			-			5,750		
21_W3	4,36E+07			-			5,820		
25_W1	3,73E+07	3,23E+07	1,33E+07	-	3,14E+05		5,700	5,687	0,042
25_W2	1,73E+07			-			5,720		
25_W3	4,25E+07			3,14E+05			5,640		
32_W1	3,15E+07	1,48E+07	1,52E+07	3,00E+04	3,00E+04		5,650	5,647	0,055
32_W2	1,12E+07			-			5,590		
32_W3	1,82E+06			-			5,700		
39_W1	6,09E+05	1,58E+06	1,16E+06	1,00E+004	1,00E+04		5,530	5,470	0,072
39_W2	1,26E+06			-			5,490		
39_W3	2,87E+06			-			5,390		
60_W1	1,55E+06	1,75E+06	2,83E+05	2,20E+03	1,46E+03	1,04E+03	5,600	5,620	0,028
60_W2	1,95E+06			7,27E+02			5,640		

Table 3S. Counts and pH values of ricotta samples with average and standard deviation.

The analysis of organic acid content was carried out in the S samples at day 4 (one sample as control), 14 and 18. These times were selected on the basis of the occurrence of a pH decrease in W samples (see Figure 1B). The result is reported in Table 1. The content of

organic acid is very similar to that reported in Mucchetti and Neviani (2006) for industrial Ricotta [99]. The acetic acid is the only acid that increased from the day 4 to day 18 due to microbial metabolism. Lactic acid was added for the coagulation of whey proteins, whereas citric acid from the fourth sampling day is probably due to the use of whey from dairy plants that make mozzarella cheese.

Samples	pH value	Citric acid	Pyruvic acid	Lactic acid	Formic acid	Acetic acid
4_S1	6,43	405	0,18	184	76	0,80
14_S1	5,96	477	1,89	134	97	15,2
14_S2	5,84	459	0,80	134	115	23,9
14_S3	5,75	484	0,20	130	106	22,6
14_S4	5,95	448	1,19	127	92	11,3
18_S1	5,76	456	1,03	135	126	30,7
18_S2	5,80	465	0,74	136	111	25,6
18_S3	5,76	566	1,02	165	141	35,5

Table 1. Organic acid content in Ricotta samples. The acid quantity is mg/100gr.

3.4.2. Microbial community during shelf-life

The microbiological and chemical analysis allowed for the classification of the lots selected as representative of standard Ricotta products ([99], ELDA factory, personal communication) and suitable for the definition of the microbial communities.

Microbial community profiles of the Ricotta samples were assessed by 16S rRNA gene sequencing using Illumina technology. The retrotranscribed RNA was used as a template for the library construction as better representative of the living micro flora. High-throughput 16S rRNA gene library sequencing robustly determines the diversity and abundance of microbial communities in a quantitative and qualitative form.

For W and S samples, MiSeq runs produced 2,361,140 and 1,380,586 raw reads, respectively, with an average amplicon length of 466 bp, including V3 and V4 regions of the 16S rRNA gene. After quality trimming and merging of raw data, filtered sequences were used for subsequent analysis.

A total of 2,447 and 2,302 OTUs for W and S samples, respectively, were detected considering the communities from all samples. Only OTUs represented by more than five reads were considered. The α-diversity metrics estimated during rarefaction, reported in Figure 1S, demonstrated that the OTU are distributed among samples from a minimum of

1000 to a maximum of 4000 for W samples, and from a minimum of 400 to a maximum of 3500 in summer samples. A higher homogeneity in the number of reads was obtained for W samples, 31,000-33,000 reads for each sample, while S sample reads were from 9,000 to 70,000. However, that number of reads might be enough to describe the micro flora of a pasteurized food [163].

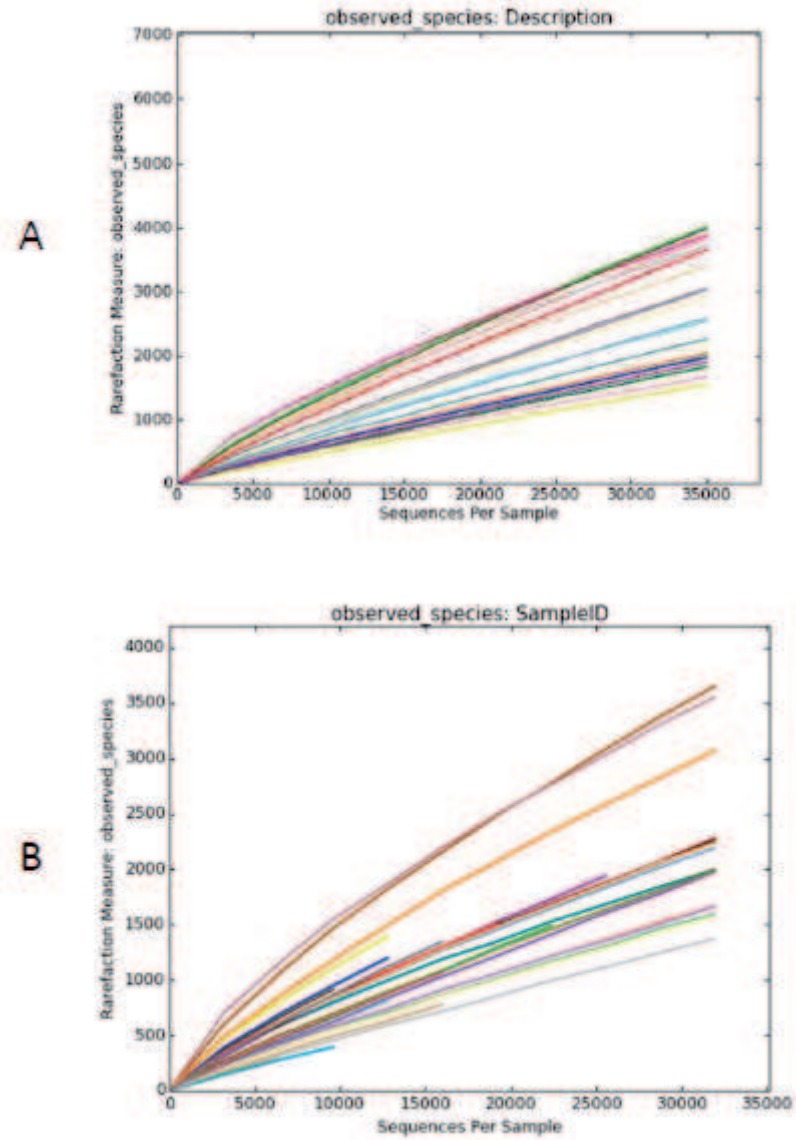

Figure 1S. Rarefaction curves using observed_species metric of 16S rRNA gene amplicon library sequencing from winter (A) and summer (B) Ricotta samples.

The sample 11_S3 presented a very unusually large number of OTUs, and the community resulted completely different compared to the two ricotta packages sampled at the same time (11_S1 and 11_S2). The microbial communities of the samples 11_S3 and 14_S3 (presenting pink discoloration) were excluded by the general analysis.

Four different phyla (*Proteobacteria, Firmicutes, Actinobacteria, and Bacteroides*) represented more than 99% of the W and S population, and the percent of representation of

each one is reported in Table 2. The majority of the microflora in all samples belonged to *Firmicutes*, as expected in dairy products [154, 177]. *Firmicutes* represented more than 99% of the population in all samples excluding the W day 14 samples where the mean value among the three samples was about 98% (the three samples presented W1 99.3%, W2 98.3% and W3 96.2%). The 14_W2 and 14_W3 samples had a percentage of *Proteobacteria* (α- and ɣ-) about ten-fold higher than the 14_W1 sample and all the other samples of both W and S.

Phyla	**11**	**14**	**18**	**21**	**25**	**32**	**39**	**60**
Actinobacteria								
Winter	n.a.	0,148 ±0,117	0,030 ±0,027	0,046 ±0,035	0,003 ±0,003	0,002 ±0,003	0	0,001 ±0,002
Summer	0,012 ±0,006	0	0,023 ±0,035	0,005 ±0,004	0,001 ±0,001	0,006 ±0,01	0	0
Bacteroidetes								
Winter	n.a.	0,073 ±0,052	0,015 ±0,012	0,030 ±0,046	0,003 ±0	0,001 ±0,002	0,001 ±0,002	0
Summer	0,004 ±0,006	0	0,011 ±0,015	0,006 ±0,007	0,011 ±0,013	0,004 ±0,007	0	0,001 ±0,002
Firmicutes								
Winter	n.a.	97,970 ±1,585	99,603 ±0,197	99,481 ±0,560	99,813 ±0,067	99,894 ±0,037	99,882 ±0,002	99,910 ±0,006
Summer	99,544 ±0,201	99,680 ±0,098	99,380 ±0,358	99,806 ±0,051	99,788 ±0,068	99,854 ±0,034	99,459 ±0,452	99,112 ±0,679
Proteobacteria								
Winter	n.a.	1,543 ±1,444	0,247 ±0,126	0,304 ±0,384	0,108 ±0,039	0,073 ±0,024	0,057 ±0,017	0,047 ±0,018
Summer	0,214 ±0,159	0,096 ±0,061	0,132 ±0,165	0,046 ±0,057	0,036 ±0,045	0,019 ±0,009	0,005 ±0,007	0,009 ±0,004

Table 2. Percentage of *Phyla* in Ricotta cheese at each time of shelf-life. For each time media and standard deviation among ricotta samples are reported. (n.a. Not available data)

Analyzing the microbial diversity for a deeper taxonomic assignment, the number of OTUs defining genera were 1927 and 2028 (for W and S, respectively), demonstrating that most of the OTUs define the genus, as expected using a V3-V4 region of the 16S rRNA gene [159], and the genera distribution could be very descriptive for the microbial population of these samples.

Genera				Species			
W Samples	n. OTU	S Samples	n. OTU	W Samples	n. OTU	S Samples	n. OTU
Actinobacteria		***Actinobacteria***		***Actinobacteria***		***Actinobacteria***	
Corynebacterium	4	-	1	*Rothia mucilaginosa*		-	
Micrococcus	1						
Mycetocola	1						
Parascardovia	1						
Propionibacterium	1						
Rothia	2						
Bacteroidetes		***Bacteroidetes***		***Bacteroidetes***		***Bacteroidetes***	
Chryseobacterium	2	-	1	-		-	
Flavobacterium	1						
Porphyromonas	1						
Firmicutes		***Firmicutes***		***Firmicutes***		***Firmicutes***	
Abiotrophia	1	*Anoxybacillus*	1	*Bacillus muralis*	18	*Anoxybacillus kestanbolensis*	1
Bacillus	40	*Bacillus*	152	*Clostridium butyricum*	6	*Bacillus anthracis*	1
Carnobacterium	1	*Caloramator*	24	*Lactobacillus helveticus*	1	*Bacillus cereus*	22
Clostridium	9	*Clostridium*	467	*Lactobacillus zeae*	1	*Bacillus flexus*	1
Finegoldia	1	*Geobacillus*	1	*Paenibacillus amylolyticus*	29	*Bacillus horikoshii*	1
Lactobacillus	3	*Granulicatella*	1	*Staphylococcus equorum*	1	*Bacillus muralis*	20
Lactococcus	3	*Lactobacillus*	1	*Streptococcus infantis*	1	*Clostridium bowmanii*	11
Paenibacillus	1811	*Lactococcus*	3			*Clostridium butyricum*	5
Staphylococcus	2	*Leuconostoc*	1			*Clostridium neonatale*	3
Streptococcus	3	*Lysinibacillus*	7			*Clostridium pasteurianum*	38
Weissella	1	*Oxobacter*	10			*Clostridium tyrobutiricum*	1
		Paenibacillus	1344			*Lysinibacillus boronitolerans*	6
		Staphylococcus	1			*Paenibacillus lentimorbus*	2
		Streptococcus	4			*Paenibacillus stellifer*	1
		Trichococcus	1				

Genera				Species			
W Samples	n. OTU	S Samples	n. OTU	W Samples	n. OTU	S Samples	n. OTU
Proteobacteria		***Proteobacteria***		***Proteobacteria***		***Proteobacteria***	
Acinetobacter	7	*Acinetobacter*	2	*Acinetobacter johnsonii*	2	*Acinetobacter johnsonii*	1
Agrobacterium	1	*Hemophilus*	1	*Acinetobacter iwoffii*	1	*Hemophilus parainfluenzae*	1
Azomonas	1	*Neisseria*	2	*Azomonas insignis*	1	*Neisseria subflava*	1
Bradyrhizobium	2	*Pseudomonas*	2	*Enhydrobacter aerosaccus*	1	*Pseudomonas fragi*	1
Chromohalobacter	2	*Sphingomonas*	1	*Hemophilus parainfluenzae*	1		
Delftia	1			*Propionibacterium acnes*	1		
Devosia	1			*Pseudomonas fragi*	2		
Enhydrobacter	2			*Pseudomonas veronii*	2		
Haemophilus	1			*Sphingomonas echinoides*	1		
Idiomarina	1						
Janthinobacterium	1						
Neisseria	1						
Novosphingobium	1						
Paracoccus	1						
Pseudomonas	7						
Roseomonas	1						
Skermanella	1						
Sphingomonas	3						
Stemotrophomonas	1						
Other		***Other***		***Other***		***Other***	
Fusobacterium	1	*Fusobacterium*	2	-		-	
Thermus	1						

Table 3. Genera and species in W and S Ricotta samples.

These OTUs identified 41 and 20 different genera for W and S samples respectively (Table 3). The large number of genera identified in W samples could possibly be due to the lower number of reads obtained for S samples that have as a consequence a less deep description of the community with loss of the minority population (*Actinobacteria, Bacterioidetes and Proteobacteria*). On the contrary, more genera of *Firmicutes* were identified in S samples, which could be due to a higher biodiversity in the predominant micro flora.

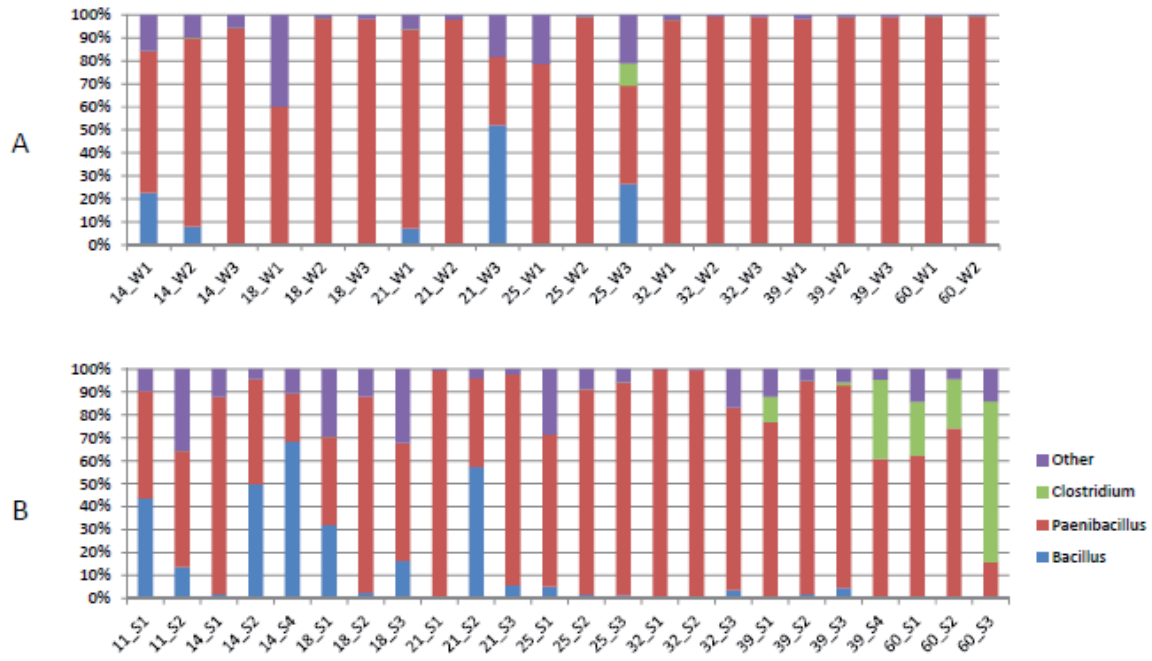

Figure 2. Percentage of the most represented genera in microbial community of W (A) and S (B) Ricotta samples.

The distribution of major genera represented in the different samples is presented in the bar-graph in Figure 2. The three genera covering more than 80% of the micro flora in most of the samples (excluding five samples in which the three genera are comprised between 60% and 80%) were *Bacillus*, *Paenibacillus* and *Clostridium*. The spore forming bacteria resulted in the predominant microflora in all the Ricotta samples as expected, since Ricotta is produced by prolonged heating process and pasteurization [167]. Such processes that inactivate most of the vegetative forms are ineffective against spores and, on the contrary, might contribute to the germination of the spores [168]. The evolution the micro flora demonstrated a variability in genera distribution that could probably be due to the chance of the growth of a genus in the single package. This result was evident in the samples until day 25 in W and day 21 in S. However, a trend of predominance of *Paenibacillus* in the latter part of the shelf-life and an abundance of *Bacillus* during the first part resulted in Ricotta samples as described in several studies on milk storage [145, 169, 170]. That trend could be due to the adaptability to cold storage and to a decrease in the pH for

Paenibacillus. *Clostridia* were present in most of the samples (both W and S) at a very low level due to the aerobic or microaerobic condition of the Ricotta packages. The relevant presence of *Clostridia* was found at the end of the shelf-life in four S samples and in one W sample (25_W3). The chance of growth, representing a relevant portion of the global population, might occur late in the shelf-life as consequence of oxygen consumption by aerobic micro flora, in particular *Bacillus* and *Paenibacillus*. Therefore, the development of *Clostridia* could be lot specific and not dependent on seasonality, but additional summer and winter lots should be analyzed to confirm the results. The identification to the species level was obtained for 70 and 117 OTUs for W and S samples, identifying 17 and 18 different species, respectively (Table 3). The distribution of species confirms the major biodiversity of *Firmicutes* in S samples with five species of *Bacillus*, five of *Clostridium* and two of *Paenibacillus* (one of each in W samples). However, the definition of the OTUs at species level could be not representative of the real situation in the sample, since the genus level is the deepest with an acceptable classification performance for short sequencing reads [25].

β-diversity significance test demonstrated that seasonality does not influence the composition of the community ($p = 1$), while the difference in variability within the same season samples was significant (p-value=$1{,}4 \times 10^{-8}$). The major variability was evident in S samples and could be due to a major biodiversity in the starting population that evolve during shelf-life. Even if the majority of the population is the same in W and S samples, different taxa became predominant for chance in some samples. Moreover the major number of *Firmicutes* genera and species found in the analysis was probably not enough to differentiate S and W samples, eventually due to the representability of each taxa in the global population.

3.4.3. Molecular characterization of pink colonies isolated from sample 14_S3

The Ricotta sample 14_S3 at the beginning presented a pink discoloration. Similar defects have been already reported specifically in the factory (Elda Factory, personal communication) and generally in cheese [171]. Several additional media were used to isolate the bacteria involved in pigment production, as described in the Material and Methods section. On the basis of bibliography reporting on microorganism producing pink pigment, media for yeasts [172], *Lactobacilli* [171], *Micrococcaceae* [173] and *Pseudomonas* [174] were tested. A negative result was obtained for all media, leading to the exclusion of these taxonomic groups. White colonies recovered in the TSA plate after one month of incubation at 8 °C became pink, while in minimal medium the colonies remained white. Two pink colonies (B1 and B3), grown in liquid media, were used to

inoculate two Ricotta packages demonstrating the ability after 14 days at 8 °C to reproduce the pink discoloration observed on the 14_S3 Ricotta sample. The sequencing of portion of the 16S rRNA from the B1 and B3 colonies demonstrated for both an identity of 99% with strains belonging to the *Bacillus cereus* group. For species and strain identification, an MLST approach was applied [141] and the sequences were compared with the *Bacillus cereus* MLST database (http://pubmlst.org/bcereus/). The definition of the ST demonstrated a new allelic combination, identical for B1 and B3 colonies, that was submitted to the *Bacillus cereus* MLST database, getting ST=1008, (*glp* 84, *gmk* 10, *ilv* 22, *pta* 217, *pur* 57 *pyc* 22 *tpi* 11). The most similar ST present in the *Bacillus cereus* MLST database to ST1008 was 985 (the two STs are single locus variant for *glp* 64). The phylogenetic analysis conducted with MLST concatenated sequences of ST1008, ST985 (extracted from MLST database) and a selection of *Bacillus cereus* group strains (extracted from complete genome sequences deposited in GenBank) demonstrated the B1/B3 strain belonged to *B. mycoides/B. wehienstephanensis* species (Figure 3A). The composition of the microbial community of sample 14_S3 is presented in Figure 3B.

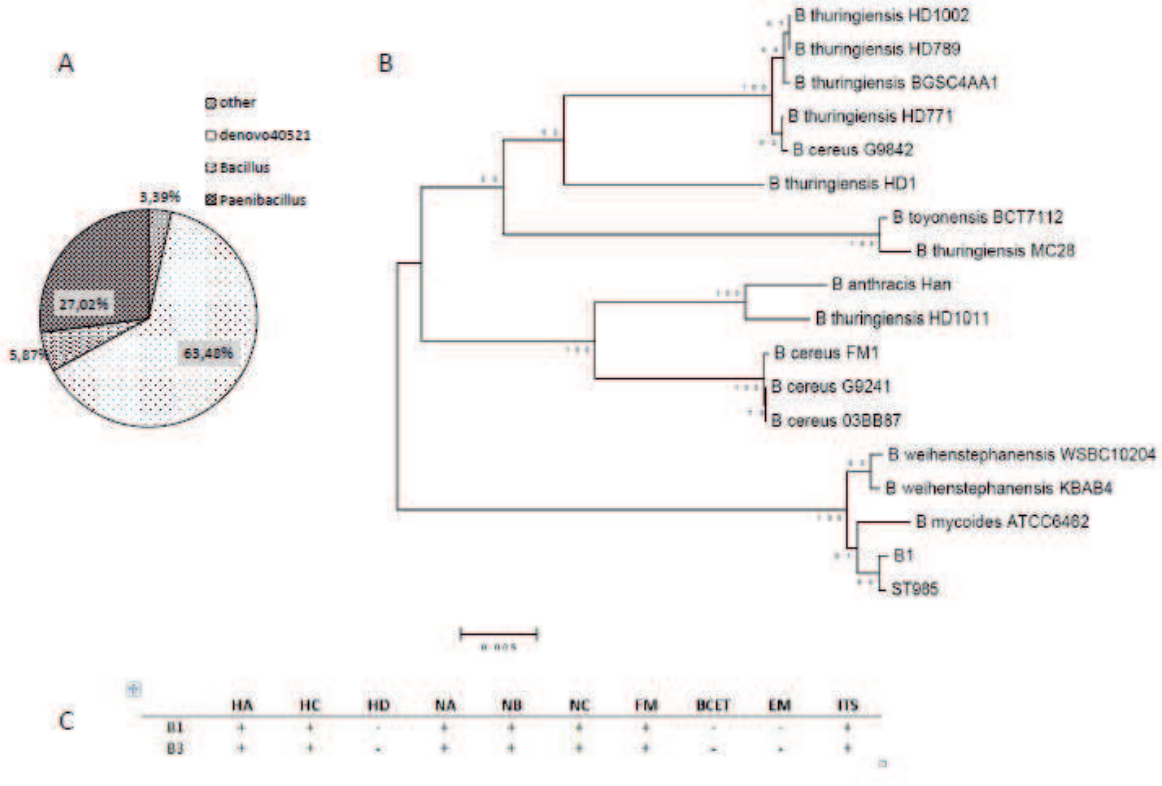

Figure 3. A. Microbial community description of Ricotta sample 14_S3. B. Neighbor-joining tree of concatenated sequences of B1, ST985 and selected *Bacillus cereus* group strains. C. Toxin gene PCR profile of B1 and B3 strains.

The 65.5% of the community was composed by an OTU identified as family *Bacillaceae*, and the genus was not defined (see Table 1S). The corresponding sequence, denovo40521, was extracted from the "pynast_aligned_seqs" file produced by QIIME and the Blast analysis against Nucleotide GenBank identified the *Bacillus cereus* group. The nucleotide alignment of the denovo40521 and the 16S rRNA gene sequences of B1/ B3 strain reported an identity close to 100%. These data demonstrated that the B1/B3 strain is the predominant micro flora in this sample and the cause of the pink discoloration. It is interesting to note that this OTU was predominant only in this sample on the total 46 Ricotta packages analyzed (Table 1S). This is the first report of a food discoloration caused by a strain belonging to the *Bacillus cereus* group. The isolated pink colonies were also analyzed for the presence of toxin genes and the result is reported in Figure 3C. The strains resulted positive to several toxin genes (*Hbl, Nhe* and *entFM*), but negative for the emetic toxin previously identified in *B. weihenstephanesis* strains [175, 1756].

3.5. Conclusions

The composition of the microbial community during the shelf-life of Ricotta demonstrated the presence of a large population of spore forming bacteria that is the main cause of the spoilage of the products before expiry of the suggested or presumed shelf life indicated for industrial and pasteurized Ricotta [99]. No significant differences were reported between S and W samples but the variability reported between samples belonging to the same season could be explained by the chance for different taxa to become predominant in each Ricotta package. In addition, that population poses a serious problem for the safety of the product. Indeed pathogenic bacteria as *Bacillus cereus* are included and in some cases predominant in the community. A strain belonging to *Bacillus cereus* group was identified as cause of the pink discoloration and the community determination of the discolored Ricotta demonstrated the predominance of that strain. The spores are derived from raw material, mostly from the whey and in the Ricotta they might find the ideal conditions to germinate and grow. Heat processing during production and pasteurization could promote the activation of germination. The elimination of spores from raw material appears to be a unique strategy to elongate the shelf-life of industrial Ricotta.

The results obtained in the present study demonstrate how the application of NGS technologies describing in-depth microbial communities could, in the future, become a suitable tool in the food industry for improving the quality and the safety of products.

3.6. Acknowledgments

The study was supported by the Italian Ministry of Economic Development (Industria 2015 Made In Italy project code MI01_00132.) to BMR genomics (Padova, Italy), and the Ricotta factory “Elda” (Vestenanova, VR, Italy). The PhD school of Veterinary Sciences of University of Padova supported the education of N.A.A. The authors are grateful to the Ricotta factory “Elda” (Vestenanova, VR, Italy) for providing samples, and to Dr. Rosaria Lucchini (Istituto Zooprofilattico delle Venezie) for suggestions and scientific support.

4. Chapter IV

NGS approach in food technology applications

4.1. Introduction

The production of fermented foods is one of the oldest food processing technologies known to man. Since the dawn of civilization, different methods for the fermentation of milks, meats and vegetables have been reported, with earliest descriptions dating back to 6000 BC and the civilizations of the fertile crescent in the Middle East [178]. These processes were artisan in nature and obviously the role of microorganisms could have not been appreciated. Nevertheless, traditions were established by which the handling and storage of certain raw materials with a specific process resulted in the development of foods that not only had keeping qualities that were better than those of the original substrate, but that also had desirable and organoleptically pleasing characteristics.

By the middle of the XIX century, two events had occurred which had a very strong impact on the manner in which food fermentation were performed and on our understanding of the process. Firstly, the industrial revolution induced the concentration of large masses of people in towns and cities. Thus the traditional method of supplying foods within local communities could be no longer applied. The ability to satisfy these new markets required products to be made in large quantities necessitating the industrialization of the manufacturing process. Secondly, the spreading of microbiology allowed to understand for the first time the biological basis of fermentation. Thus, the essential role of bacteria, yeasts and moulds in the generation of fermented foods came to be comprehended and this resulted in more controlled and efficient fermentation processes [179].

Thus towards the end of the 19th century for many fermented foods, in particular milk-derived products, the characterization of the microorganisms responsible for the fermentation led to the isolation of starter cultures. These microorganism could be produced on a large scale to supply Dairy factories involved in the manufacture of these products. This significant development had a major impact on the processes used and contributed to ensuring consistency of product and reliability of fermentation.

4.1.1. Role of fermentation

The original and primary purpose of fermenting food substrates was to obtain a preservation effect. However, as particularly in the Western World many common and effective alternative preservation technologies were developed, this is no longer the most pressing requirement and many of these foods are manufactured because their unique flavour, aroma and texture attributes are much appreciated by the consumer. Anyway the conditions generated by the fermentation are essential in ensuring the shelf-life and microbiological safety of the products [179].

Preservation of foods by fermentation depends on the principle of oxidation of carbohydrates and related compounds to generate end-products such as acids, alcohol and carbon dioxide. These final compounds control the growth of food spoilers and because of the partial oxidation, the food maintains sufficient energy potential to be of nutritional benefit for the consumer.

The chemical definition of ‘fermentation’ is applied to describe a strictly anaerobic process; however, the general meaning of the term now encompasses both aerobic and anaerobic carbohydrate breakdown processes. Fermentation may also increase the nutritional quality of food by increasing digestibility as in the fermentation of milk to cheese. Toxicity of foods may also be reduced by fermentation as occurs in the production of “Gari” [179].

The link between fermentation and preservation is biopreservation which refers to the extension of the shelf-life and improvement of the safety of food using microorganisms and/or their metabolites. In this respect, it is well known that starter microorganisms can produce a wide range of antimicrobial compounds and proteinaceous substances which can inhibit or reduce undesirable flora in food products, such as primary and secondary metabolites including organic acids, diacetyl, CO_2 and even antibiotics [180].

4.1.2. Food fermentation by LAB

When considering food fermentations (as distinct from alcoholic fermentations involving yeast), a group of bacteria known collectively as the LABs is primarily responsible for many of the microbial transformations found in the more common fermented food products (Tab. 4.1.1). This group is composed of different genera including *Lactococcus, Lactobacillus, Enterococcus, Streptococcus, Leuconostoc* and *Pediococcus* and generally produces lactic acid as their major end-product. They are strictly fermentative and lack the ability to synthesize *heme* which means that they are catalase-negative and lack a terminal electron transport chain [181].

Product	**Microorganisms**	**Substrate**
Wine, beer	*Saccharomyces cerevisiae*, LAB	grapes, grain, hops
Bread	*Saccharomyces cerevisiae*, LAB	wheat, rye, grains
Cheddar cheese	*Lactococcus* (*cremoris, lactis*) and *leuconostoc*	milk
Swiss-type cheese	*Lactobacillus* (*delbruckii, bulgaricus, helveticus*)	milk
Mould- and smear-ripened cheeses	*Carnobacterium piscicola, Brevibacterium linens*	milk
Yogurts	*St. thermophilus* and *Lb. bulgaricus*	milk
Kefir	Lactococci, yeast, *Lb. kefir* (and others)	milk
Fermented meats	Pediococci, Staphylococci, various LAB	pork, beef
Sauerkraut	*L. lactis, Leuc. Mesent., Lactobacillus* (*brevis, plantarum,curvatus, sake*)	cabbage
Soy sauce	*Aspergillus oryzae*/*soyae*, lactobacilli and*Zygosaccharomyces rouxii*	soy beans and wheat
Vegetables	*Enterococcus* (*mundtii, faecium*), *Lactococcus* (*cremoris,lactis*), *Lactobacillus* (*plantarum, casei*)	vegetables
Fish	*Carnobacterium* (*piscicola, divergens*)	fish

Tab. 4.1.1. Biopreservation by lactic acid bacteria from [179].

Lactic acid bacteria are generally mesophilic but can grow at temperatures as low as 5°C or as high as 45°C. Similarly, while the majority of strains grow at pH 4.0–4.5, some are active at pH 9.6 and others at pH 3.2. They are often weakly proteolytic and lipolytic strains and require preformed amino acids, purine and pyrimidine bases and B vitamins for growth [179].

Members of the LAB can be subdivided into two distinct groups based on their carbohydrate metabolism (fig. 4.1.1). The homofermentative group, comprehending *Lactococcus, Pediococcus, Enterococcus, Streptococcus* and some *lactobacilli,* use the Embden–Meyerhof–Parnas pathway to convert 1 mol of glucose into 2 mol of lactate. In contrast, equimolar amounts of lactate, CO_2 and ethanol from glucose are produced by heterofermentative bacteria, using the hexose monophosphate or pentose pathway. Thus they generate only half the energy of the homofermentative group. Members of this group include *Leuconostoc, Weissella* and some *lactobacilli* [180].

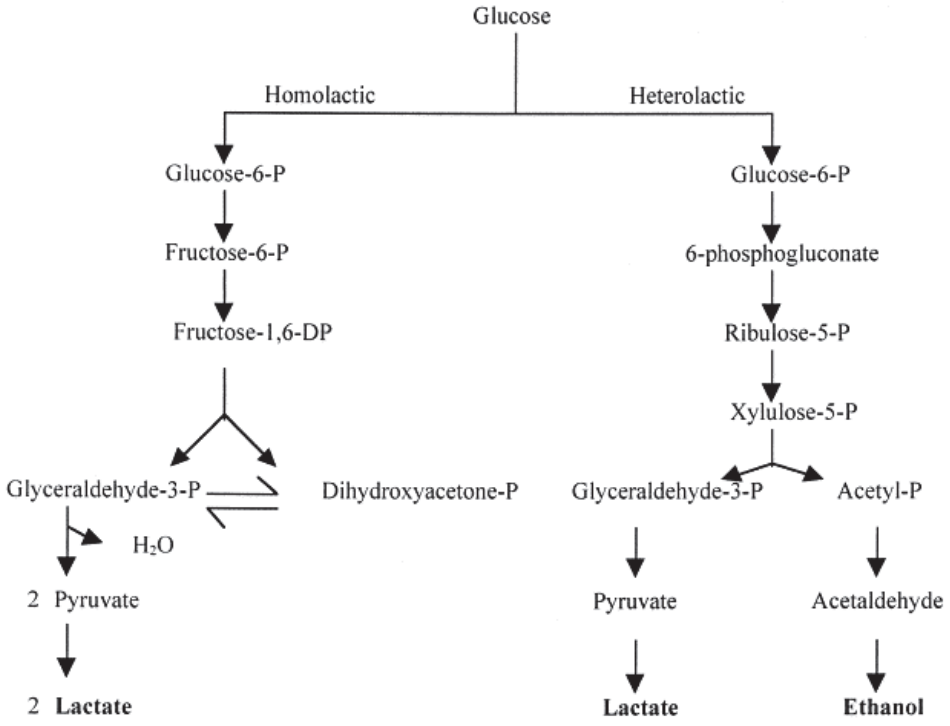

Fig. 4.1.1. Homofermentative and heterofermentative pathways of LAB metabolism.

4.1.3. Antimicrobial compounds produced by LAB

The specific antimicrobial mechanisms of lactic acid bacteria exploited in the biopreservation of foods include the production of organic acids, hydrogen peroxide, carbon dioxide, diacetyl, broad-spectrum antimicrobials such as reuterin and bacteriocins [182]. Each antimicrobial compound produced during fermentation provides an additional obstacle for pathogens and spoilage bacteria to overcome before they can survive and/or proliferate in a food or beverage, during production chain to time of consumption. As any microorganism may produce a number of different inhibitory substances, its antimicrobial potential is defined by the comprehensive action of its metabolic products on undesirable bacteria [180].

Organic acids, acetaldehyde and ethanol

The direct antimicrobial effects of organic acids including lactic, acetic and propionic which may be produced by lactic acid bacterial fermentation of foods are well understood. The antagonism is believed to result from the action of the acids on the bacterial cytoplasmic membrane which interferes with the maintenance of membrane potential and inhibits active transport [182]. Acetic acid is more inhibitory than lactic acid and can inhibit yeasts, moulds and bacteria [195] while propionic acid inhibits fungi and bacteria. The contribution of acetaldehyde to biopreservation is minor since the flavour threshold is much lower than the levels that are considered necessary to achieve inhibition of microorganisms [196]. Similarly, although ethanol may be produced by lactic cultures, again the levels produced in food systems are so low that the anti-microbial contribution is minimal.

Hydrogen peroxide

Lactic acid bacteria lack true catalase to break down the hydrogen peroxide generated in the presence of oxygen. It is argued that the H_2O_2 can accumulate and be inhibitory to some microorganisms [181]. H_2O_2 has a strong oxidizing effect on membrane lipids and cellular proteins and is produced with enzymes as the flavoprotein oxidoreductases NADH peroxidase, NADH oxidase and α-glycerophosphate oxidase [181]. Hydrogen peroxide may also activate the lactoperoxidase system of fresh milk with the formation of hypothiocyanate and other antimicrobials [181, 182, 183].

Carbon dioxide

Carbon dioxide, formed from heterolactic fermentation, can directly create an anaerobic environment and is toxic to some aerobic food microorganisms through its action on cell membranes and its ability to reduce internal and external pH [184, 182]. At low concentration, it may be stimulatory to the growth of some bacteria [185]. Production of CO2 resulting from the use of lactate by propionibacteria in Swiss cheese manufacture is responsible for the characteristic “eyes” of the finished product.

Diacetyl

Diacetyl is a product of citrate metabolism [185] and is responsible for the aroma and flavour of butter and some other fermented milk products. Many lactic acid bacteria including strains of *Leuconostoc, Lactococcus, Pediococcus* and *Lactobacillus* may produce diacetyl although production is repressed by the fermentation of hexoses [186,187]. Gram-negative bacteria, yeasts and moulds are more sensitive to diacetyl than gram-positive bacteria and its mode of action is believed to be due to interference with the

utilisation of arginine [182, 188]. Diacetyl is rarely present in food fermentations at sufficient levels to make a major contribution to antibacterial activity.

Bacteriocins

It has been known for some time that many members of the LAB produce proteinaceous inhibitors called collectively bacteriocins. These inhibitors generally provoke target cell membrane depolarization or inhibit cell wall synthesis [189], and range in specificity from a narrow spectrum of activity (lactococcins which only inhibit *lactococci*) to those which have a broad range of activity such as the lantibiotic nisin [182]. These proteinaceous inhibitors have attracted an intensive research interest over the last three decades, inducing the discovery and characterization of many different types of bacteriocins from LAB. Bacteriocins can be divided into 4 main groups as follows in table 4.1.2 [192].

Class	Subclass	Description
I		Lantibiotics–small, heat stable, containing unusual amino acids
II		Small (30–100 amino acids), heat stable, non-lantibiotic
	IIa	Pediocin-like bacteriocins, with anti-listerial effects
	IIb	Two peptide bacteriocins
	IIc	Sec-dependent secretion of bacteriocins
III		Large (> 30 kDa) heat-labile proteins
IV		Complex bacteriocins with glyco- and/or lipidmoieties

Table 4.1.2. Classes of bacteriocins produced by lactic acid bacteria.

4.1.4. LAB technological application as “adjunct cultures”

The advent of retailing and mass marketing required that products of important quality and safety could be available. The traditional approach of backslopping or even natural fermentation of substrate was not the appropriate approach upon which any large-scale industrial process could be based. Fortunately for many fermented foods, but particularly milk-derived products, the characterization of the microorganisms responsible for the fermentation towards the end of the 19th century led to the isolation of starter cultures which could be produced on a large scale to supply factories involved in the manufacture of these products. This significant development had a major impact on the processes used and contributed to ensuring consistency of product and reliability of fermentation. [180].

Besides in recent years, adequate procedures to produce milk have been developed at the farm and plant levels and obtain cheese with the lowest number of microorganisms became possible. The introduction of a low-temperature pasteurization process coupled with microfiltration led to the production of almost sterile milk [194]. Although these methods provide the consumers with dairy products exhibiting a high degree of microbiological quality and safety, they also lead to the disappearance of or at least to a dramatic reduction in the number of desirable non-starter bacteria. For the cheese industry to offer to the consumers safe and consistent cheeses with high organoleptic properties in a reasonable ripening time, they began to look for new technologies such as “adjunct cultures.” Adjunct cultures can be defined as selected strains of cheese related microorganisms that are added to the cheese milk to improve development of cheese sensory quality. They were also developed to accelerate cheese ripening, which may allow substantial cost savings to the cheese industry. In contrast to naturally occurring NSLAB, adjuncts are specifically selected and intentionally added to supplement the microflora of cheese milk to improve overall quality of finished cheese [191].

The microorganisms involved in cheese making and cheese ripening can be divided into two major groups: 1) microorganisms that are added to the cheese milk after being carefully selected by the starter manufacturer or the cheese-making company, and 2) nonstarter lactic acid bacteria (NSLAB). Group 1 can be further subdivided into two subgroups: the primary starter and the secondary starter [191].

Primary starter

The role of the primary starter culture is to ensure consistent acid development during cheese making. This group is also involved in the degradation of protein and fat during ripening. Cultures in this group also play an important role in the biological protection of the product (e.g., low pH, bacteriocin production) [191].

Secondary starter

This group contains cultures that are added to a limited number of cheese varieties to provide well-defined functions. In addition, through their diverse enzymatic systems, secondary starters can be also heavily involved in the ripening process; their contributions are indispensable for the development of the typical flavor of many cheeses such as Roquefort [191, 192]. Some examples are: *Propionibacterium shermanii* ssp. *freudenreichii*, while *Brevibacterium linens* is one of the major contributors to surface coloration in surface ripened cheese *Penicillium roqueforti* and *Penicillium camemberti.*

Non-starter lactic acid bacteria (NSLAB)

Includes the nonstarter lactic bacteria. This group has been shown to contribute to flavor development in some varieties of cheeses and could therefore be considered a desirable contaminant of either the milk supply or the subsequent cheese. *Lactobacillus* strains are the most common and can be found in relatively high numbers (fig. 4.1.2); *L. casei, L. paracasei, L. plantarum*, and *L. curvatus* are the predominant species. *Pediococci* and *enterococci* are also members of the group but are usually present in smaller numbers [193].

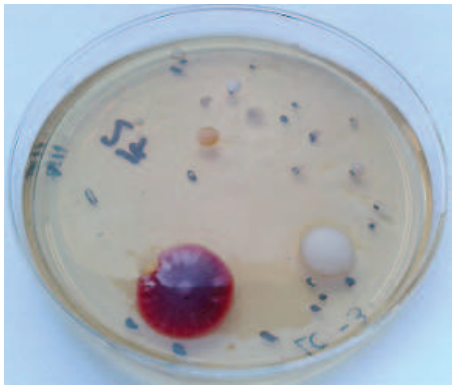

Fig. 4.1.2. Medium M17 for *Lactobacillales* culturing and isolation.

Protective effect of adjunct cultures

Adjunct cultures can improve both cheese taste/flavors but also its microbiological quality. Indeed, there are many examples which report the inhibition of spoilage and pathogenic bacteria by LAB [198, 199]. In a study by Daly et al. (1970), the food-borne pathogens *Pseudomonas fluorescens, Staphylococcus aureus* and *Clostridium perfringens* were inhibited when co-cultured with a *Streptococcus diacetylactis* strain. In this case, the *S. aureus* numbers were reduced by more than 99% in foods such as ham sandwich spread, chicken gravy and ground beef. This inhibition was most likely a direct result of acid production by the starter *L. lactis* subsp. *lactis* biovar. *diacetylactis* strain [197].

In addition, many LAB have shown also a strong anti-fungal activity [200], in table 4.2.3 some examples are reported.

LAB species (number of strains)	**Source**	**Antifungal compounds**	**Mould and yeast activity spectrum**	**Reference**
Lactobacillus plantarum (30)	Flowers, sourdough, grass silage, sorghum, wheat, dairy products, sausages, wheat semolina, *kimchi* (Korean pickles), malted barley, fresh vegetables	Organic acids, PLA, 4-hydroxyphenyllactic acid, cyclo(L-Phe-L-Pro), cyclo(L-Phe-*trans*-4-OH-L-Pro), 3-phenyllactic acid, proteinaceous, ethanol, ethyl acetate, 3-hydroxy fatty acids, cyclo(Leu-Leu), cyclo(L-Leu-L-Pro)	*Penicillium* sp.,*Monilia* sp.,*Aspergillus* sp.,*Fusarium* sp.,*Eurotium sp.*,*Talaromyces* sp.,*Epicoccum* sp.,*Cladosporium* sp.,*Rhizopus stolonifer*,*Sclerotium oryzae*,*Rhizoctonia solani*,*Botrytis cinerea*,*Sclerotinia minor*, *Endomyces fibuliger*,*Rhodotorula* sp.,*Candida albicans*,*Debaryomyces hansenii*,*Kluyveromyces marxians*,*Saccharomyces*sp., *Phichia* sp.	[200-204]
Lactobacillus sakei(2)	Dandelion flour and leaves	ND	*A. fumigatus*, *F. sporotrichioides*	[205]

Tab. 4.2.3. Some lactic acid bacteria studied with antifungal activity, their source, antifungal compounds, and spectral inhibitory activity.

4.1.5. Microbiological methods to study adjunct cultures

Different approaches are used to study the microbiota of fermented food and in particular cheese, traditionally cultured–dependent methods are used to enumerate the starter, wild microbial population and spoilage bacteria present in cheese. Recently, cultured-independent methods, and in particular next generation sequencing is used to evaluate the complete microbiota of complex ecosystem such as foods. Many of these techniques have been discussed in Chapter 2.

4.2. Purpose

We applied NGS analysis to the technological use of adjunct cultures on soft or ripened cheeses in order to evaluate their anti-spoilers effect.

We evaluated the anti-*Gammaproteobacteria* (against *Enterobacteriaceae* and *Pseudomonadaceae*) effect of some strains such as *L. rhamnosus*, *L. sakei* and *Carnobacterium maltoaromaticum* in fresh industrial cheeses prone to package swelling and premature spoiling.

4.3. Materials and Methods

4.3.1. Sample preparation

Three different productions of fresh cheese were collected. Each production was performed in the standard way (STD) and adding different protection strains: plus *Lactobacillus sakei* (SA), plus *Lactobacillus rhamnosus* (RH), plus *Carnobacterium maltaromaticum* (CB) and plus a mix of these cultures (MX). Half part of each production was contaminated with a mixture culture of contaminant microrganisms (STD1, SA1, RH1, CB1, MX1), previously isolated from the same cheese product during this research, in order of 10^2 UFC/ item. Samples were collected in triplicate and stored at 8°C and 14°C until deadline, a summary of the samples is reported in tab. 4.3.1.

Name	Adjunct culture	Spoilers added	Storage temperature
STD8	Standard production	No	8 °C
STD18	Standard production	Yes	8 °C
SA8	*L. sakei*	No	8 °C
SA18	*L. sakei*	Yes	8 °C
RH8	*L. rhamnosus*	No	8 °C
RH18	*L. rhamnosus*	Yes	8 °C
CB8	*C. maltoaromaticum*	No	8 °C
CB18	*C. maltoaromaticum*	Yes	8 °C
MX8	all	No	8 °C
MX18	all	Yes	8 °C
STD14	Standard production	No	14 °C
STD114	Standard production	Yes	14 °C
SA14	*L. sakei*	No	14 °C
SA114	*L. sakei*	Yes	14 °C
RH14	*L. rhamnosus*	No	14 °C
RH1114	*L. rhamnosus*	Yes	14 °C
CB14	*C. maltoaromaticum*	No	14 °C
CB114	*C. maltoaromaticum*	Yes	14 °C
MX14	all	No	14 °C
MX114	all	Yes	14 °C

Tab. 4.3.1. Sample summary with adjunct culture, added spoilers and storage temperature.

4.3.2. DNA extraction

Only one of the tree productions was analyzed by means NGS, thus 20 samples of the lot 1716 were extracted in triple starting from 200 mg of cheese. 1,8 ml of sterile solution of Sodium citrate 2% was added to the cheese, homogenated and incubated 10 minutes at 45°C. After a centrifugation step of 8000 rpm for 8 minutes the fat layer on the surface must be discarded together with the supernatant. The lysis step on bacterial pellet was performed as described in section 1.3.4.

Extracted DNA was loaded on agarose gel 1% to check genome integrity and RNA contaminations. In addition randomly chosen samples were quantified by means Qubit in order to quantify DNA and verify extraction success.

4.3.3. 16S rRNA amplicon library construction and bioinformatic analysis

Library was prepared according to the method described in section 1.3.5, using 5 µl of DNA. Moreover the resulting reads of the replicates were merged, pooled and analyzed with QIIME software as described in section 1.3.6 using closed-reference approach for OTU-picking method. The biplot 3D PCoA was drawn with `make_emperor.py` script using the taxonomy file at species level.

4.4. Results and discussion

All samples were not visible on agarose gel. In fact Qubit quantification demonstrated that DNA concentration spanned from 0.5 to 2 ng/µl. Thus for library preparation at least 5 µl of DNA were used.

4.4.1. Microbiota of standard production

From Miseq run we obtained 13,565,705 raw reads and, after merging with FLASH, replicate pooling and quality filtering, the final reads were 5,614,272, with an average length of 465 bp (± 6,6) bp. We obtained 1911 OTUs that decreased to 181 after conservative filtering step [20].

At the end of the shelf life of the fresh cheese, NGS analysis showed that the microbial community of standard production of soft cheese (STD) was composed mainly by *Proteobacteria* (60%) and *Firmicutes* (38%) while at 14°C *phyla* flip and *Proteobacteria* decrease to the 39% and Firmicutes rise to 60%. At genus and species level (fig. 4.4.1), in STD at 8°C of storage, *Lactococcus* was around 21%; *Moraxellaceae* family was represented mainly by *Acinetobacter johnsonii* (25%), *Enterobacteriaceae* (40%) were represented by *Morganella* (around 9,6%), while *Clostidiaceae* (mainly *Clostridium*) was around 4,7%. IN STD1 at 8 °C storage *Carnobacterium* increases to 6%, as *Streptococcus* to 18,3%, while *Lactococcus* decreses to 10%. *Enterobacteriaceae* increase to 25% while *Acinetobacter johnsonii* decreases to 18%.

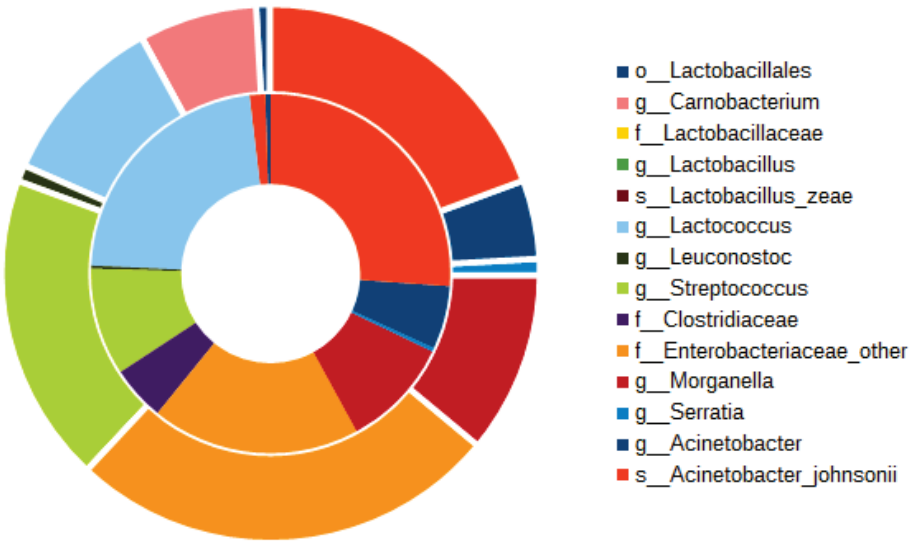

Fig. 4.4.1. Taxa of STD (internal ring) and STD1 (external ring) at 8°C storage.

In STD at 14°C storage, *Streptococcaceae* were represented by *Streptococcus* (38%) and *Lactococcus* (6,6%), fig. 4.4.2; *Leuconostoc* (2%) and *Lactobacillaceae* (around 4%) were represented by *Lactobacillus zeae* (1,9%); *Moraxellaceae* were represented by *Acinetobacter johnsonii* (around 3%), *Enterobacteriaceae* was represented by *Morganella* (5%) and others *Enterobacteriaceae* in amount of 30%. *Clostridiaceae* is composed only by *Clostridium* (6%). Contaminated STD1 at 14° C shows an increased number of *Lactobacillales* (16%) and *Lactobacillus zeae* (4%), Clostridiaceae and Citrobacter (respectively 7,5% and 1,5%).

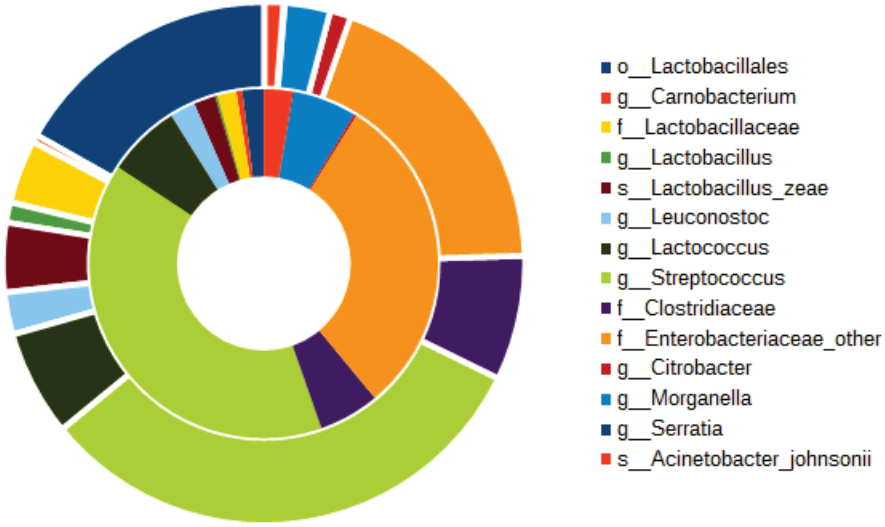

Fig. 4.4.2. Taxa of STD (internal ring) and STD1 (external ring) at 14°C storage.

4.4.2. Microbiota of products with adjunct cultures

Simultaneously to the standard production other different production were made, adding with the starter culture one or more protective cultures and stored samples at 8°C and 14°C. In comparison with STD, *Lactobacillaceae* family abundance increased in SA (49%) and both *Enterobacteriaceae* and *Moraxellaceae* families decreased at 8°C storage, fig. 4.4.3. The artificial contamination of the sample SA induced the exponential growth of *Serratia* (55%) and a strong decrease of *Moraxellaceae* (2%). In RH sample the amount of *Enterobacteriaceae* and *Acinetobacter johnsonii* was slightly decreased while in RH1 it decreased 2 folds. *Clostridiaceae* only appeared in SA and STD samples. Thus at 8°C a massive reduction of spoilers was observed in presence of the adjunct culture *Carnobacterium* where communities of *Enterobacteriaceae, Moraxellaceae and Clostridiacea*e were reduced under 1%. Surprisingly *Morganella* did spread in MX sample (21%).

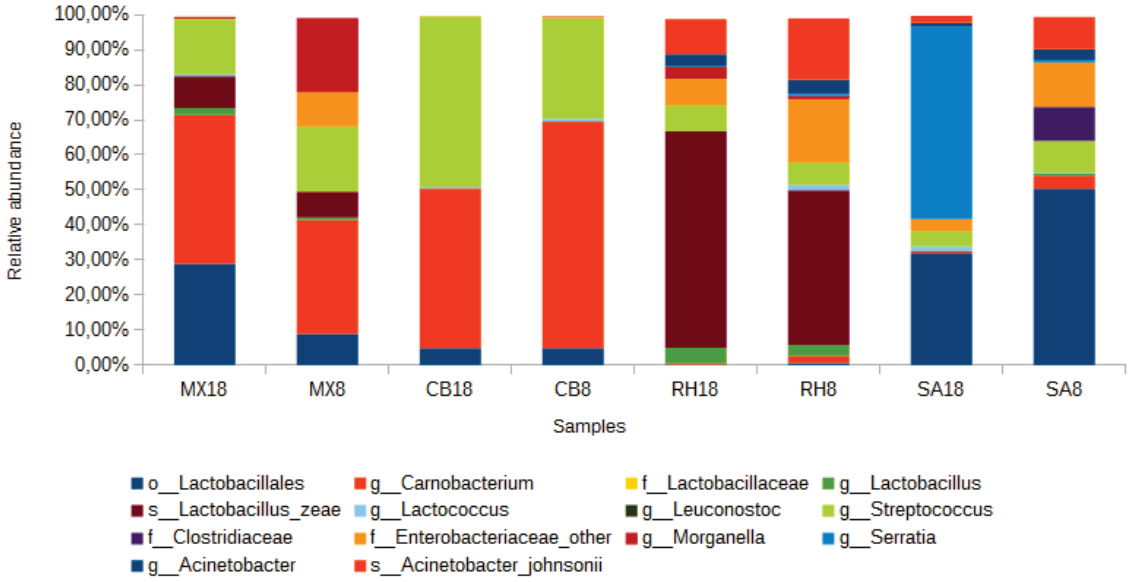

Fig. 4.4.3. Taxa of samples with protective culture and contaminants at 8°C storage.

In the samples stored at 14°C *Streptococcaceae* (8-58%), *Lactobacillaceae* (51%) and *Clostridiaceae* (5-16%) and *Enterobacteriaceae* (11% mean) families were on average more abundant than in those stocked at 8°C while *Moraxellaceae* (5-7%) and *Carnobacteriaceae* (14-20%) families showed difficult growth (Fig. 4.4.4). Finally also at 14°C the spoilage microbiota was mainly reduced by the adjunct culture *Carnobacterium*; *Enterobacteriaceae*, *Morganella* and *Acinetobacter johnsonii* were reduced under 1%, only *Clostridiaceae* had a remarkable increase in comparison to STD production.

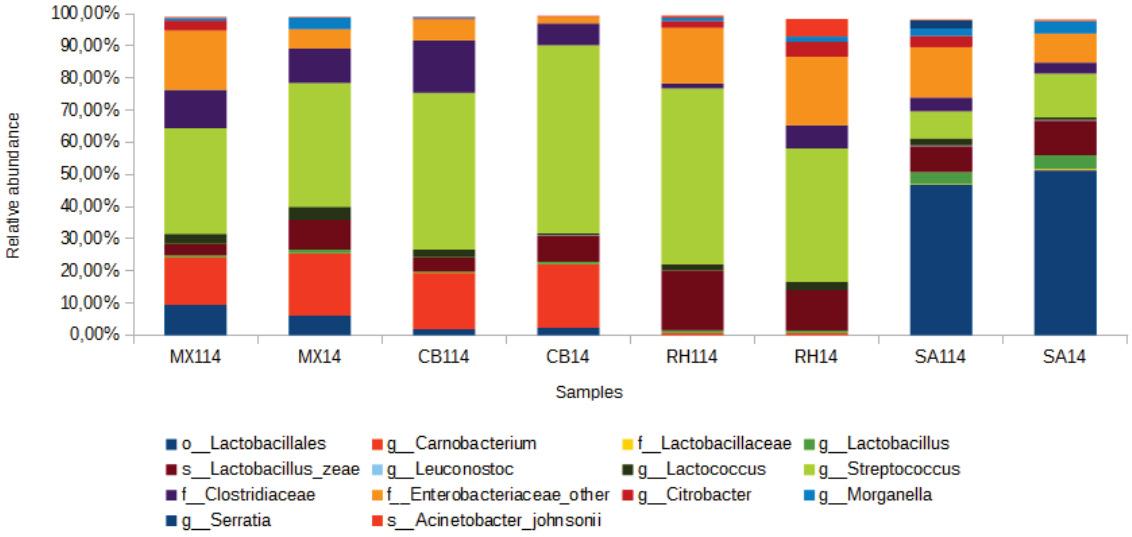

Fig. 4.4.4. Taxa of samples with protective culture and contaminants at 14°C storage.

Nevertheless figure 4.4.5 shows how different bacteria drove the unweighted beta-diversity separation between samples. Generally samples are divided into two group based on storage temperature but it is clear that, both at 8°C and 14°C, *Carnobacterium* allowed division between CB, coupled with MX samples, and the others. In fact CB and MX showed a similar microbiota, suggesting that Carnobacterium induced the *anti-spoiler* effect of the protective mixture. Moreover *Lactobacillales*, *Streptococcus* and *Enterobacteriaceae* drove the SA isolation while *Acinetobacter johnsonii* and *Lactobacillus zeae* induced the separation from STD and RH.

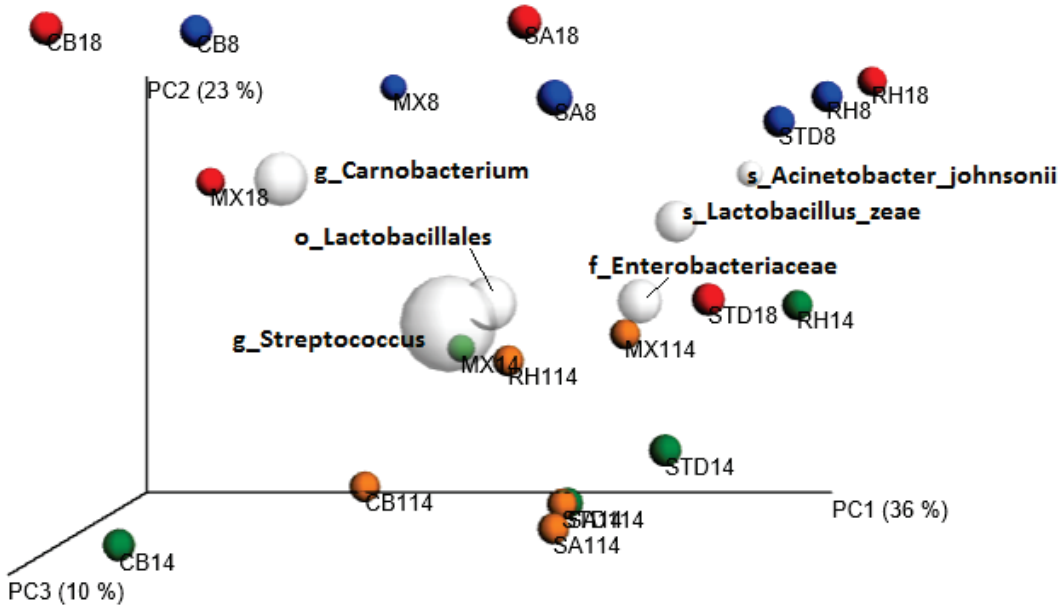

Figure 4.4.5. Biplot showing unweighted unifrac PCoA with taxa spheres. Blue spheres are samples at 8°C, red are 8°C samples, contaminated. Yellow spheres are 14°C samples while green spheres are 14°C samples, contaminated. Grey spheres represent the 6 most driving taxa and their diameter describes the abundance and the importance of the taxa for the beta-diversity structure.

To this extent the alpha-diversity at 120000 reads of rarefaction depth showed that CB performed best at both temperatures (tab. 4.4.1). Moreover either CB, CB1 and MX, MX1 samples had the lowest number of species and the lowest Phylogenetic diversity value at 8°C of storage. On the other hand RH1 and CB, CB1 resulted the lowest number of species at 14°C while MX was one of the most biodiverse.

	8 °C			14 °C		
Sample	**Observed species**	**PD**	**Shannon**	**Observed species**	**PD**	**Shannon**
STD	151	8,19	4,42	165	9,21	4,27
STD1	149	8,33	4,42	161	8,62	4,38
SA	141	8,01	3,38	154	8,57	3,35
SA1	125	7,39	2,77	159	8,61	3,91
RH	144	7,72	3,45	149	8,16	3,70
RH1	124	7,21	2,50	140	7,61	3,05
CB	109	6,32	1,86	120	7,01	2,64
CB1	95	6,17	2,00	140	7,92	3,15
MX	112	6,67	3,07	145	7,75	3,59
MX1	107	6,64	2,68	151	8,18	4,03

Tab. 4.4.1. Table showing alpha-diversity values of soft cheeses microbiota.

These results demonstrated that *Carnobacterium maltoaromaticum* was the most effective anti-spoilers bacterium, managing to inhibit the growth of most *Gammaproteobacteria* either at 8° and at 14°C. Furthermore the same results were obtained with parallel microbiological analysis (data not shown). On the other hand it was not so effective against another *Firmicutes* such as *Clostridium*. Furthermore *L. rhamnosus* showed an improved protective efficiency at 8° above all when cheese was contaminated. It is interesting to note that the MX, that contains all the anti-spoilers, is not as effective as CB itself, demonstrating that the interaction between adjunct cultures with the same effect is not always the sum of the single influences. Moreover every protective bacterium could work differently at diverse temperatures.

In this study the usage of adjunct cultures that were previously characterized with different methods such as RAPD, Sanger sequencing and RDP classification, shows a remarkable issue in bacteria identification at molecular level. At species level the various 16S databases (discussed in section 1.1.3) can present some differences in taxonomy assignment, due to the phylogenetic and specific regulation on which the database itself is built [39, 22, 45, 88]. Thus analyzing the same read with different databases sometimes can drive to different results. i.e. OTU 658224, that is the most common in SA samples, in Greengenes is classified as *k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__; g__; s__*, while in RDP it is attributed to different strains of *k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus;*

s__*sakei*. Another example is the OTU 1132297 that in Greengenes is assigned to *k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__zeae,* while in RDP is assigned to *k__Bacteria; p__Firmicutes; c__Bacilli; o__Lactobacillales; f__Lactobacillaceae; g__Lactobacillus; s__rhamnosus* (and *casei* or *paracasei*). To this extent for some applications it is possible to re-train one database with the other, or compare the same reads with different reference databases [206], in order to make results homogeneous between different approaches. Thus biplot results with RDP database would represent the major separation-driving taxa in some case with the names of the adjunct cultures (fig. 4.4.6).

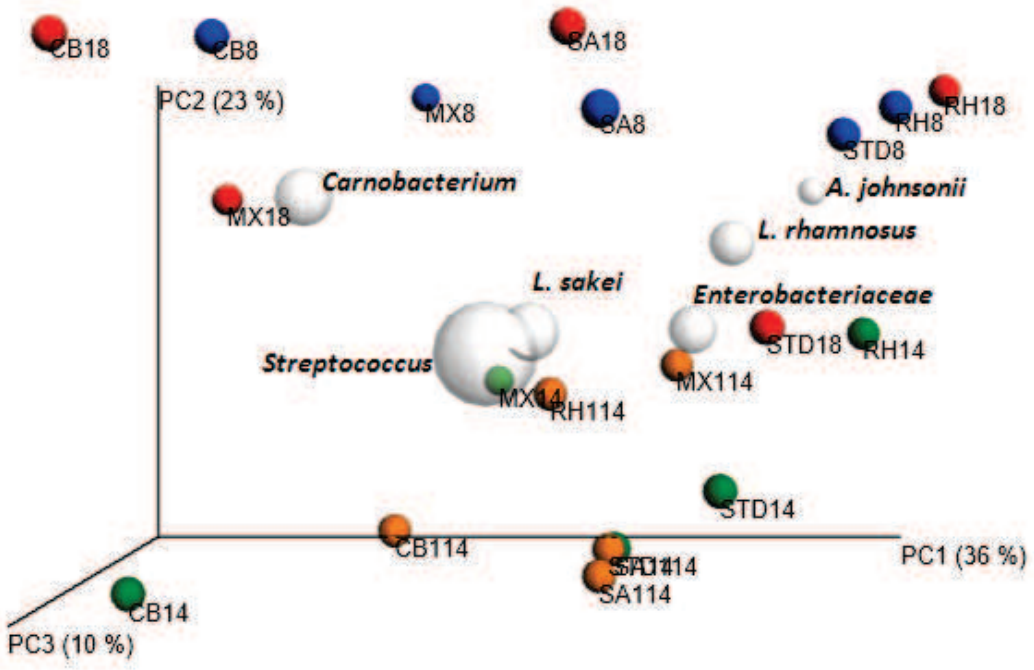

Figure 4.4.6. Biplot showing unweighted unifrac PCoA with taxa spheres (RDP database). Blue spheres are samples at 8°C, red are 8°C samples, contaminated. Yellow spheres are 14°C samples while green spheres are 14°C samples, contaminated. Grey spheres represent the 6 most driving taxa and their diameter describes the abundance and the importance of the taxa for the beta-diversity structure.

4.5. Conclusions

In this work we evaluated the anti-spoiler effect of tree different LABs, *Lactobacillus rhamnosus, Lactobacillus sakei* and *Carnobacterium maltoaromaticum* by means NGS approach. With this technique we characterized in deep the overall microbial community of the standard product and how it changed in presence of protective bacteria and contaminants. NGS approach described the ability of every adjunct culture to interact with the original microflora in therms of overall decrease in microbiota biodiversity and spoilers defeat. We found that among the others *Carnobacterium maltoaromaticum* was the most effective *Gammaproteobacteria* antagonist, but not anti-*Firmicutes* culture. As the same results were obtained on Petri plate, we can conclude that 16S NGS approach can definitely become a suitable tool for food technology studies and industry application, supporting and sometimes replacing the traditional culture-dependent methods.

4.6. Final remarks

All throughout these projects, 16S-NGS was applied with different purposes, showing that this approach is flexible and versatile. It is a tool that can be coupled with many approaches (we saw GC approach, MLST analysis, culture-dependent methods) but above all it can be used in different fields. In this dissertation we applied his approach at food microbiology studies but it can be used in human and animal microbiology, moreover in plant and environmental microbiology. In fact at BMR Genomics this approach has been applied also to human gut microbiology studies, a promising exploration field for medicine, pharmaceutics and food testing. Moreover this genetic approach on a single gene is going to be supported also with metagenomic and meta transcriptomic approaches through which it is possible to visualize no more who is there (16S) but what it can do (global genomes) and what it is doing (global transcriptome).

Finally with this work we can support the thesis that as (food) microbiology research can take great advantages from these powerful approaches, food industries can improve cheese making processes through technological and knowledge transfer. In fact, applying this researches to food quality improvement and safety management, food industries can satisfy new customer needs, extend products shelf-life, open new markets and face modern challenges of the globalization era.

REFERENCES

1. Caporaso J. G., et al. 15 March 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Natl. Acad. Sci. U. S. A. 108:4516–4522.

2. Lazarevic V., et al. 2009. Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. J. Microbiol. Methods 79:266–271.

3. Gloor G. B., et al. 2010. Microbiome profiling by Illumina sequencing of combinatorial sequence-tagged PCR products. PLoS One 5:e15406.

4. Tanja Magoč, Steven L. Salzberg. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011 November 1; 27(21): 2957–2963

5. Group Jumpstart Consortium Human Microbiome Project Data Generation Working Group (2012) Evaluation of 16S rDNA-Based Community Profiling for Human Microbiome Research. PLoS ONE 7(6): e39315. doi:10.1371/journal.pone.0039315

6. Soergel, D.A., Dey, N., Knight, R., and Brenner, S.E. (2012). Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. ISME J. 6, 1440–1444.

7. Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M. and Glöckner, F.O. (2012) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Acids Res. 2013 Jan 1;41(1):e1. doi: 10.1093/nar/gks808

8. Ludwig W, Mittenhuber G, Friedrich CG: Transfer of Thiosphaera pantotropha to Paracoccus denitrificans. Int J Syst Bacteriol 1993, 43:363–367.

9. Suzuki MT, Giovannoni SJ: Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. Appl Environ Microb 1996, 62:625–630."

10. Ruffroberts AL, Kuenen JG, Ward DM: Distribution of cultivated and uncultivated cyanobacteria and Chloroflexus-like bacteria in hot spring microbial mats. Appl Environ Microb 1994, 60:697–704

11. Nadkarni et al., (2002) Determination of bacterial load by real-time PCR using a broad-range (universal) probe and primers set. Microbiology. 2002 Jan;148(Pt 1):257-66.

12. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J.

13. Nossa et al., (2010). Design of 16S rRNA gene primers for 454 pyrosequencing of the human foregut microbiome. World J Gastroenterol. 2010 Sep 7;16(33):4135-44.

14. Berry, D, et al. (2011) Barcoded Primers Used in Multilex Amplicon Pyrosequencing Bias Amplification. Appl Environ Microbiol 77: 7846-7849

15. José A. Navas-Molina et al. (2013). Advancing Our Understanding of the Human Microbiome Using QIIME. Methods in Enzymology, Volume 531, ISSN 0076-6879

16. Valeria D'Argenio et al. (2014). Comparative Metagenomic Analysis of Human Gut Microbiome Composition Using Two Different Bioinformatic Pipeline. Volume 2014: 325340.

17. Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics, 26(19), 2460–2461.

18. Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: A flexible tool for aligning sequences to a template alignment. Bioinformatics, 26(2), 266–267.

19. Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution, 26(7), 1641–1650.

20. Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. Nature Methods, 10(1), 57–59.

21. Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. Bioinformatics, 27(16), 2194–2200.

22. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72: 5069–5072.

23. Mardia, K. V., Kent, J. T., & Bibby, J. (1979). Multivariate analysis. London: Academic Press.

24. Tryon, R. C. (1939). Cluster analysis. Ann Arbor, MI: Edwards Bros.

25. Mizrahi-Man O, Davenport ER, Gilad Y (2013) Taxonomic Classification of Bacterial 16S rRNA Genes Using Short Sequencing Reads: Evaluation of Effective Study Designs. PLoS ONE 8: e53608. doi: 10.1371/journal.pone.0053608 PMID: 23308262

26. McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst, A.,et al. (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME Journal, 6(3), 610–618.

27. Curtis TP1, et al., (2002).Estimating prokaryotic diversity and its limits. Proc Natl Acad Sci U S A. 2002 Aug 6;99(16):10494-9.

28. Whitman et al. (1998). Prokaryotes: The unseen majority. Proc. Natl. Acad. Sci. USA 95

29. Eigen, M., et al. (1985). Pattern analysis of 5S rRNA. Proc. Natl. Acad. Sci. U. S. A. 82:2437–2441.

30. Kuntzel, et al. (1981). Phylogenetic tree derived from bacterial, cytosol and organelle 5S rRNA sequences. Nucleic Acids Res. 9:1451–1461.

31. Schmidt TM et al. (1991).Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J Bacteriol; 173(14):4371-8.

32. Torsvik V. et al. (1990). High diversity in DNA of soil bacteria. Appl Environ Microbiol.; 56(3):782-7.

33. Woese CR (1989). Bacterial evolution. Microbiol Rev. 51(2):221-71

34. Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. Science 284:2124–2129.

35. Hashimoto, J. G., B. S. Stevenson, and T. M. Schmidt. (2003). Rates and consequences of recombination between rRNA operons. J. Bacteriol. 185:966–972.

36. Pei et al., (2010). Diversity of 16S rRNA Genes within Individual Prokaryotic Genomes. Applied and Environmental Microbiology, p. 3886–3897

37. Van de Peer Y et al. (1996). A quantitative map of nucleotide substitution rates in bacterial rRNA. Nucleic Acids Res.; 24(17):3381-91.

38. Soumitesh Chakravorty et al. (2007).A detailed analysis of 16S ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. J Microbiol Methods.;

69(2): 330–339.

39. Cole JR, Chai B, Marsh TL, Farris RJ, Wang Q, Kulam SA, Chandra S, McGarrell DM, Schmidt TM, Garrity GM, Tiedje JM. (2003). The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nucleic Acids Res.; 31(1):442-3.

40. McCabe KM, Zhang YH, Huang BL, Wagar EA, McCabe ER, (1999).Bacterial species identification after DNA amplification with a universal primer pair. Mol Genet Metab.; 66(3):205-11.

41. Kumar PS, Brooker MR, Dowd SE, Camerlengo T (2011) Target region selection is a critical determinant of community fingerprints generated by 16S Pyrosequencing. PLoS ONE 6(6): e20956. doi: 10.1371/journal.pone.0020956

42. Wang Y, Qian PY (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. PLoS ONE 4(10): e7401. doi: 10.1371/journal.pone.0007401

43. Liu ZZ, DeSantis TZ, Andersen GL, Knight R (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. Nucleic Acids Research 36(18): e120. doi: 10.1093/nar/gkn491

44. Cai L, Ye L, Tong AHY, Lok S, Zhang T (2013) Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS ONE 8(1): e54649 doi:10.1371/journal.pone.0053649.

45. Elmar Pruesse,et al., (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007 Dec; 35(21): 7188–7196. 10.1093/nar/gkm864PMCID: PMC2175337

46. Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. Appl Environ Microbiol.; 71(12):7724-36.

47. G. Muyzer, E. C. de Waal, and A. G. Uitterlinden, (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. Applied and Environmental Microbiology, vol. 59, no. 3, pp. 695–700.

48. N. Fierer and R. B. Jackson, (2006).“The diversity and biogeography of soil bacterial communities,” Proceedings of the National Academy of Sciences of the

United States of America, vol. 103, no. 3, pp. 626–631.

49. M. L. Sogin, H. G. Morrison, J. A. Huber et al., (2006). “Microbial diversity in the deep sea and the underexplored ‘rare biosphere’”. Proceedings of the National Academy of Sciences of the United States of America, vol. 103, no. 32, pp. 12115–12120.

50. Handelsman J, Rondon MR, Brady SF et al (1998) Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol 5:R245–R249. Doi: 10.1016/S1074-5521(98)90108-9

51. Gilbert JA, Dupont CL (2011) Microbial metagenomics: beyond the genome. Annu Rev Mar Sci 3:347–371. doi:10.1146/annurev-marine-120709-142811

52. J. C. Venter, K. Remington, J. F. Heidelberg et al., (2004). “Environmental genome shotgun sequencing of the Sargasso Sea,” Science,vol. 304, no. 5667, pp. 66–74.

53. Jay Shendure & Hanlee Ji (2008). Next-generation DNA sequencing. Nat. Biotech. doi:10.1038/nbt1486

54. Shendure, J. et al. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728–1732.

55. Margulies, M. et al. (2005) Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376–380.

56. Mitra, R.D. & Church, G.M. (1999) In situ localized amplification and contact replication of many individual DNA molecules. Nucleic Acids Res. 27, e34 .

57. 16. Mitra, R.D., Shendure, J., Olejnik, J., Edyta Krzymanska, O. & Church, G.M. (2003) Fluorescent in situ sequencing on polymerase colonies. Anal. Biochem. 320, 55–65.

58. Lin Liu,* Yinhu Li, Siliang Li, Ni Hu, Yimin He, Ray Pong, Danni Lin, Lihua Lu, and Maggie Law (2012). Comparison of Next-Generation Sequencing Systems. J Biomed Biotechnol. 2012; 2012: 251364

59. http://www.illumina.com/systems

60. Adessi, C. et al. (2000). Solid phase DNA amplification: characterisation of primer attachment and amplification mechanisms. Nucleic Acids Res. 28, e87.

61. Fedurco, M., Romieu, A., Williams, S., Lawrence, I. & Turcatti, G. BT (2006). A, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA

colonies. Nucleic Acids Res. 34, e22.

62. Metzker, M.L. (2010) Sequencing technologies - the next generation. Nat. Rev. Genet., 11 (1): 31-46.

63. Turcatti, G., Romieu, A., Fedurco, M. & Tairi, A.P. (2008). A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis. Nucleic Acids Res. 36, e25.

64. http://www.illumina.com/content/dam/illumina-support/documents/products/technotes/technote_phixcontrolv3.pdf

65. http://www.illumina.com/documents/products/technotes/technote_low_diversity_rta.pdf

66. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012). Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol 30: 434–439.

67. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, et al. (2013). Updating benchtop sequencing performance comparison. Nat Biotechnol 31: 294–296.

68. Caporaso JG, et al., (2010). QIIME allows analysis of high-throughput community sequencing data. Nat Methods.; 7(5):335-6

69. Drancourt, M., Bollet, C., Carlioz, A., Martelin, R., Gayral, J. P., & Raoult, D. (2000). 16S ribosomal DNA sequence analysis of a large collection of environmental and clinical unidentifiable bacterial isolates. Journal of Clinical Microbiology, 38(10), 3623–3630.

70. Cowan, S. T. (1968). A dictionary of microbial taxonomic usage. Oliver and Boyd, Edinburgh.

71. Stackebrandt E, Goebel B M. (1994). A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int J Syst Bacteriol.;44:846–849

72. Fox GE, Wisotzkey JD, Jurtshuk P Jr (1992). How close is close: 16S rRNA sequence identity may not be sufficient to guarantee species identity. Int J Syst Bacteriol.; 42(1):166-70.

73. McDonald, D., Clemente, J. C., Kuczynski, J., Rideout, J. R., Stombaugh, J., Wendel, D.,et al. (2012). The Biological Observation Matrix (BIOM) format or: How I learned to stop worrying and love the ome-ome. Gigascience, 1(1), 7.

74. Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution, 26(7), 1641–1650.

75. Whittaker, R. H. (1972). Evolution and measurement of species diversity. Taxon, 21(2/3), 213–251.

76. Green, J., and B. J. Bohannan. (2006). Spatial scaling of microbial biodiversity. Trends Ecol. Evol. 21:501-507

77. Lande, R. (1996) Statistics and partitioning of species diversity, and similarity among multiple communities. Oikos, 76, 5-13.

78. Veech, J. A. et al. (2002) The additive partitioning of species diversity: recent revival of an old idea. Oikos, 99, 3-9

79. Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61(1), 1–10.

80. Lozupone, C. A., Hamady, M., Kelley, S. T., & Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. Applied and Environmental Microbiology, 73(5), 1576–1585.

81. Lozupone, C., & Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. Applied and Environmental Microbiology, 71(12), 8228–8235.

82. Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. Biological Conservation, 61(1), 1–10.

83. Martin AP (2002). Phylogenetic approaches for describing and comparing the diversity of microbial communities. Appl Environ Microbiol,68(8):3673–3682

84. Lozupone C. et al. (2006). UniFrac – An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7:371

85. Felsenstein J (2004). Inferring Phylogenies. Sunderland, Massachusetts , Sinauer Associates, Inc.

86. http://www.illumina.com/documents/products/datasheets/datasheet_miseq.pdf

87. Michael A Quail et al. (2012)A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics.; 13: 341.

88. Nelson MC. Et al. (2014). Analysis, optimization and Verification of Illumina-generated 16S rRNA gene amplicon surveys. Plos One. 9(4):e94249.

89. Huse SM, Huber JA, Morrison HG, Sogin ML, Welch DM (2007) Accuracy and quality of massively parallel DNA pyrosequencing. Genome Biol 8: R143.

90. Degnan PH, Ochman H (2011) Illumina-based analysis of microbial community diversity. ISME J 6: 183–194.

91. Agilent 2100 expert software online help. Copyright © Agilent Technologies, Version 25.11.2003.

92. Jeffrey J Werner et al. (2012). Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. The ISME Journal 6, 1273–1276

93. Katherine Kennedy et al (2014). Evaluating Bias of Illumina-Based Bacterial 16S rRNA Gene Profiles. AEM;80;18;p. 5717–5722

94. Haas BJ, Gevers D, Earl A, Feldgarden M, Ward DV, et al. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. Genome Res.

95. Laufer AS, Siddall ME, Graf J (2008) Characterization of the digestive-tract microbiota of Hirudo orientalis, a european medicinal leech. Appl Environ Microbiol 74:6151–6154.

96. Knights, Dan et al. (2011). Bayesian community-wide culture-independent microbial source tracking. Nature Methods

97. Jill E. Clarridge III. (2004). Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. Clinical microbiology reviews, p. 840–862

98. Fuselli et al. (2014). Detection of fraudolent addition of bovine whey in water buffalo ricotta cheese by isoelettric focusing. Journal of the Science of Food Agriculture 10.1002/jsfa.7019

99. Mucchetti, G. and Neviani, E. (2006) “Microbiologia e tecnologia lattiero-casearia qualità e sicurezza”. Milano. Tecniche Nuove, 131, 165-168, 363-382.

100. Smithers, G.W. (2008) “Whey and whey proteins- from ‘gutter-to-gold”. Int. Dairy J., 18:695-704.

101. Jelen, P. (2003) “Whey processing: utilization and products. Encyclopedia

of dairy sciences". Edited by H. Roginski, J.W. Fuquay, and P. F. Fox. 2739-2745. New York: Accademy. Press.

102. Pescuma, M., Hébert, E.M., Mozzi, F., Font de Valdez, G., (2010). Functional fermented whey-based beverage using lactic acid bacteria. International Journal of Food Microbiology 141, 73-81.

103. Outinen, M., Rantamäki, P., Heino, A. (2010) "Effect of milk pretreatment on the whey composition and whey powder functionality". J.Food Sci., 75(1) :1-10.

104. Volonterio, A.G. (2005) "Microbiologia degli alimenti. Milano". Ambrosiana, 212-217, 235,239.

105. Desmasures, N., and Gueguen, M. (1997) "Monitoring the microbiology of high quality milk by monthly sampling over 2 years". J. Dairy Res., 64: 271–280.

106. Hantsis-Zacharov, E. and Halpern, M. (2007) "Chryseobacterium haifense sp. nov., a psychrotolerant bacterium isolated from raw milk". Int. J. Syst. Evol. Microbiol, 57: 2344–2348.

107. Wouters, J.T.M., Ayad, E.H.E., Hugenholtz, J. and Smit, G. (2002) "Microbes from raw milk for fermented dairy products". Int. Dairy J., 12: 91–109.

108. Oliver, S.P., Boor, K.J., Murphy, S.C. and Murinda, S.E. (2009) "Food safety hazards associated with consumption of raw milk". Foodborne Pathog. Dis., 6: 793–806.

109. Deeth, H.C., Khusniati, T., Datta, N., Fallace, R.B., (2002) "Spoilage patterns of skim and whole milks". J. Dairy Res., 69: 227-241.

110. Carminati, D., Bellini, E., Perrone, A., Neviani, E., Mucchetti, G. (2002) "Qualità microbiologica e conservabilità della ricotta vaccina tradizionale". Ind. Alim. 41, 549-555.

111. Jayarao, B.M., Henning D.R.(2001) "Prevalence of foodborne pathogens in bulk tank milk". J. Dairy Sci., 84: 2157-2162.

112. Verdier-Metz, I., Gagne, G., Bornes, S., Monsallier, F., Veisseire, P., Delbes-Paus, C. and Montel, M.C. (2012) "Cowteat skin, a potential source of diverse microbial populations for cheese production". Appl. Environ. Microbiol.,78: 326–333.

113. Vacheyrou, M., Normand, A.C., Guyot, P., Cassagne, C., Piarroux, R. and Bouton, Y. (2011) "Cultivable microbial communities in raw cow milk and potential transfers from stables of sixteen French farms". Int. J. Food Microbiol.,

146: 253–262.

114. Michel, V., Hauwuy, A. and Chamba, J. (2006) “Raw milk microbial composition: differences in links with microbial practices”. Rencontres Rech Rumin, 13: 309–312.

115. Monsallier, F., Verdier-Metz, I., Agabriel, C., Martin, B. and Montel, M.C. (2012) “Variability of microbial teat skin flora in relation to farming practices and individual dairy cow characteristics”. Dairy Sci. Technol., 92: 265–278.

116. Lafarge, V., Ogier, J.C., Girard, V., Maladen, V., Leveau, J.Y., Gruss, A., Delacroix-Buchet, A., (2004) “Raw cow milk bacterial populations shifts attributable to refrigeration”. Appl. Environ. Microbial, 70 (9), 5644-5650

117. Dogan, B., Boor. K.J. (2003) “Genetic diversity and spoilage potentials among Pseudomonas spp. isolated from fluid milk products and dairy processing Plants”. Appl. Environ. Microbiol., 69:130 –138.

118. Dutt, K., Gupta, P., Saran, S., Misra, S., Saxena, R.K. (2009) “Production of milk-clotting protease from Bacillus subtilis”. Appl. Biochem. Biotechnol., 158:761–772.

119. Ranieri, M.L., Boor, K.J. (2010) “Tracking and eliminating sporeformers in dairy systems”. Aust. J. Dairy Technol., 65:74–80.

120. Reid A. I., Ranieri, M.L., Martin N.H., Henk C. den Bakker, Xavier, M.B., Wiedmann, M. and Boor, K.J. (2012) “Identification and Characterization of Psychrotolerant Sporeformers Associated with Fluid Milk Production and Processing”. Applied and enviromental Microbiology, 78(6): 1853-1864.

121. Boor, K., Fromm, H., 2009. Managing microbial spoilage in the dairy industry, in: Blackburn,C.d.W. (Ed.), Food Spoilage Microorganisms. Woodhead Publishing.

122. Heyndrickx, M., Marchand, S., Jonghe, V.D., Smet, K., Coudijzer, K., Block, J.D., 2010. Understanding and preventing consumer milk microbial spoilage and chemical deterioration, in: Griffiths, M.W. (Ed.), Improving the Safety and Quality of Milk, Volume 2 - Improving Quality in Milk Products. Woodhead Publishing, pp. 96-135.

123. Sørhaug, T., Stepaniak, L., 1997. Psychrotrophs and their enzymes in milk and dairy products: Quality aspects. Trends in Food Science and Technology 8, 35-41.

124. Mc Sweeney, P.L.H. (2007) “Cheese Problems Solved”. Cambridge, Wood head Publishing Ltd.

125. Tomaino, R.M., Turner, L.G., Larick, D.K. (2004) “The effect of lactococcus lactis starter cultures on the oxidative stability of liquid whey”. J. Dairy Sci., 87:300.

126. Liaw, I., Eshpari, H., Tong, P.S., Drake, M.A. (2010) “The impact of antioxidant addition on flavor of Cheddar and Mozzarella whey and Cheddar whey protein concentrate”. J. Food Sci., 75:559–69.

127. Campbell, R.E., Miracle, R.E., Drake, M.A. (2011) “Effects of starter culture and storage on the flavor of liquid whey”. Journal of food science, Vol. 76, Nr.5. 354-361.

128. Mucchetti, G., Carminati, D., Parisi, A. (2002) “Ricotta fresca vaccina ed ovina: osservazioni sulle tecniche di produzione e sul prodotto”. Latte, 27 (2): 154-166.

129. Jany, J.L., and Barbier, G. (2008) “Culture-independent methods for identifying microbial communities in cheese”. Food Microbiology. 25:839–848.

130. Abee, T., Van Schaik, W. and Siezen, R.J. (2004) “Impact of genomics on microbial food safety”. Trends Biotechnol. 22:653-60.

131. Giraffa, G., and Neviani. E. (2001) “DNA-based, culture-independent strategies for evaluating microbial communities in food-associated ecosystems”. Int. J. Food Microbiol, 67(1-2):19-34.

132. Coppola, S., Blaiotta, G., Ercolini D. and Moschetti, G. (2001) “Molecular evaluation of microbial diversity occurring in different types of Mozzarella cheese”. J. Appl. Microbiol., 90:414–20.

133. Spiegelman, D., Gavin, W. and Greer W.C. (2005) "A survey of the methods for the characterization of microbial consortia and communities." Can. J. Microbiol., 2005: 355-386.

134. O’Sullivan, D.J., Giblin, L., McSweeney, P.H., Sheehan, J.J. and Cotter, P.D. (2013) “Nucleic-acid based approaches to investigate microbial-related cheese quality defects”. Frontiers in Microbiology.

135. Morozova, O., and Marra, M.A. (2008) “Applications of next- generation sequencing technologies in functional genomics”. Genomics, 92: 255–264.

136. Quigley. L., O’Sullivan, O., Beresford, T.P., Ross, R.P., Fitzgerald, G.F. and

Cotter, P.D. (2011) “Molecular approaches to analysing the microbial composition of raw milk and raw milk cheese”. Int. J. Food. Microbiol., 150: 81–94.

137. O’Flaherty, S. and Klaenhammer, T. R. (2011) “The impact of Omic-technologies on the study of food microbes”. Annu. Rev. Food Sci. Technol., 2: 353–371.

138. Siezen, R.J., Van Enckevort, F.H., Kleerebezem, M. and Teusink, B. (2004) “Genome data mining of lactic acid bacteria: the impact of bioinformatics”. Curr. Opin. Biotechnol., 15(2):105-15.

139. Jensen, M.A., Webster, J.A. and Straus, N. (1993) “Rapid identification of bacteria on the basis of polymerase chain reaction-amplified ribosomal DNA spacer polymorphisms”. Applied and Environmental Microbiology, 59, 945–952.

140. Andreani, N.A., Martino, M.E., Fasolato, L., Carraro, L., Montemurro, F., Mioni, R., Bordin, P. and Cardazzo, B. (2013) “Tracking the blue: a MLST approach to characterise the *Pseudomonas fluorescens* group”. Food Microbiol., 39:116-26.

141. Cardazzo, B., Negrisolo, E., Carraro, L., Alberghini, L., Patarnello, T. and Giaccone V. (2007). “Multiple-locus sequence typing and analysis of toxin genes in Bacillus cereus food-borne isolates”. Appl. Environ. Microbiol., 74:850-60.

142. WT Fothergill , (1968). Gas Chromatography – technique. Proc R Soc Med. 61(5): 525–528.

143. Legendre P, Legendre L. Numerical Ecology. 2nd ed. Amsterdam: Elsevier, 1998. ISBN 978-0444892508

144. Becker, R. A., Chambers, J. M. and Wilks, A. R. (1988) The New S Language. Wadsworth & Brooks/Cole.

145. Ivy, R. A., M. L. Ranieri, N. H. Martin, H. C. den Bakker, B. M. Xavier, M. Wiedmann and K. J. Boor. 2012. Identification and characterization of psychrotolerant sporeformers associated with fluid milk production and processing. Appl. Environ. Microbiol. 78:1853-1864.

146. Marchand, S., Vandriesche, G., Coorevits, A., Coudijzer, K., De Jonghe, V., Dewettink, K., De Vos, P., Devreese, B., Heyndrickx, M. and De Block, J. (2009) “Heterogeneity of heat-resistant proteases from milk Pseudomonas species”. International Journal of Food Microbiology, 133: 68-77.

147. De Jonghe, V., Coorevits, A., Van Hoorde, K., Messens, W., Van

Landschoot, A., De Vos, P. and Heyndrick, M. (2011) “Influence of storage conditions on the growth of Pseudomonas species in refrigerated raw milk”. Applied and environmental microbiology, 77: 460-470.

148. Gennari, M. and Dragotto, F. (1992) “A study of the incidence of different fluorescent Pseudomonas species and biovars in the microflora of fresh and spoiled meat and fish, raw milk, cheese, soil and water”. J. Appl. Bacteriol., 72(4):281-8.

149. Sanchez-Salas, J.L., Setlow, B., Zhang, P., Li, Y.Q. and Setlow, P. (2011) “Maturation of released spores is necessary for acquisition of full spore heat resistance during Bacillus subtilis sporulation”. Appl. Environ. Microbiol., 77: 6746–6754.

150. Snellman, E.A., Colwell, R.R. (2004) “Acinetobacter lipases:molecular biology, biochemical properties and biotechnological potential”. J. ind. Microbial. Biotechnol. 31(9): 391-400.

151. Troiani C., 2015. Il siero di latte è una risorsa per diversi mercati. Il Latte, 1. Retrieved from http://www.lattenews.it/il-siero-di-latte-e-una-risorsa-per-diversi-mercati/

152. Salvadori Del Prato O. (1992). Trattato di tecnologia casearia. Ed. Calderini Edagricole. Cap. 16 “La Ricotta e il Mascarpone”, 564-576. Edizione del 2001.

153. Mancuso I., Cardamone, C., Fiorenza, G., Macaluso, G., Arcuri, L., Miraglia, V., Scatassa, M.L., 2014. Sensory and microbiological evaluation of traditional ovine Ricotta cheese in modified atmosphere packaging. Italian Journal of Food Safety, 3(1725): 122-124.

154. Quigley L., O'Sullivan O., Stanton C., Beresford T.P., Ross R.P., Fitzgerald G.F., Cotter P.D., 2013. The complex microbiota of raw milk. J Dairy Sci. 96(8):4928-37.

155. Lu M., Shiau Y., Wong J., Lin R., Kravis H., Blackmon T., Pakzad T., Jen T., Cheng A., Chang J., Ong E., Sarfaraz N., Wang N.S., 2013. Milk spoilage: methods and practices of detecting milk quality. Food and Nutrition Sciences 4:113-123.

156. Remenant B., Jaffrès E., Dousset X., Pilet M.F., Zagorec M., 2015. Bacterial spoilers of food: behavior, fitness and functional properties. Food Microbiol. 45(Pt A):45-53.

157. Hough G., Puglieso M.L., Sanchez R., Mendes da Silva O., 1999. Sensory

and microbiological Shelf-Life of a commercial Ricotta cheese. J Dairy Sc. 82(3): 454-459.

158. Martins J.T., Cerqueira M.A., Souza B.W., Carmo Avides M.d., Vicente A.A., 2010. Shelf-life extension of Ricotta cheese using coatings of galactomannans from nonconventional sources incorporating nisin against Listeria monocytogenes. J Agric Food Chem. 58(3):1884-91.

159. Mayo B., Rachid C.T., Alegría A., Leite A.M., Peixoto R.S., Delgado S., 2014.Impact of next generation sequencing techniques in food microbiology. Curr Genomics. 15(4):293-309.

160. Lücking G., Stoeckel M., Atamer Z., Hinrichs J., Ehling-Schulz M., 2013. Characterization of aerobic spore-forming bacteria associated with industrial dairy processing environments and product spoilage. Int J Food Microbiol. 166(2):270-9

161. Postollec F., Mathot A.G., Bernard M., Divanach M.L., Pavan S., Sohier D., 2012. Tracking spore-forming bacteria in food: from natural biodiversity to selection by processes. Int J Food Microbiol. 158(1):1-8

162. Ledenbach L.H., Marshall R.T., 2009. Microbiological spoilage of dairy products. In Compendium of the Microbiological Spoilage of Foods and Beverages. Food Microbiology and Food Safety. W.H. Sperber, M.P. Doyle (eds.). Springer Science+Business Media B.V. Dordrecht, Netherlands.

163. Ercolini D., 2013. High-throughput sequencing and metagenomics: moving forward in the culture-independent analysis of food microbial ecology. Appl. Environ. Microbiol. 79:3148–3155.

164. Shokralla S., Spall J.L., Gibson J.F., Hajibabaei M., 2012. Next-generation sequencing technologies for environmental DNA research. Mol. Ecology 21: 1794–1805.

165. Carraro L., Maifreni M., Bartolomeoli I., Martino M.E., Novelli E., Frigo F., Marino M., Cardazzo B., 2011. Comparison of culture-dependent and culture-independent methods for bacterial community monitoring during Montasio cheese manufacturing. Res Microbiol. 162(3):231-9.

166. Yang, I. C., Shih, D. Y., Huang, T. P., Huang, Y. P., Wang, J. Y., Pan, T. M., 2005. Establishment of a novel multiplex PCR assay and detection of toxigenic strains of the species in the Bacillus cereus group. J. Food Protect. 68:2123-2130.

167. Buehner K.P., Anand S., Djira G.D., Garcia A., 2014. Prevalence of

thermoduric bacteria and spores on 10 Midwest dairy farms. J Dairy Sci. (11):6777-84.

168. Setlow P., 2014. Germination of spores of Bacillus species: what we know and do not know. J Bacteriol. 196(7):1297-305.

169. Munsch-Alatossava P., Ghafar A., Alatossava T., 2013. Potential of nitrogen gas (n2) flushing to extend the shelf life of cold stored pasteurised milk. Int J Mol Sci. 14(3):5668-85.

170. Moreno Switt A.I., Andrus A.D., Ranieri M.L., Orsi R.H., Ivy R., den Bakker H.C., Martin N.H., Wiedmann M, Boor K.J., 2014. Genomic comparison of sporeforming bacilli isolated from milk. BMC Genomics. 15:26.

171. Daly D., McSweeney P., Sheehan J., 2012. Pink discolouration defect in commercial cheese: a review. Dairy Science & Technology 92(5):439-453

172. Moliné M., Libkind D., van Broock M., 2012 Production of torularhodin, torulene, and β-carotene by Rhodotorula yeasts. Methods Mol Biol. 898:275-83.

173. Jagannadham M.V., Chattopadhyay M.K., Shivaji S., 1996. The major carotenoid pigment of a psychrotrophic Micrococcus roseus strain: fluorescence properties of the pigment and its binding to membranes. Biochem Biophys Res Commun. 220(3):724-8.

174. Downs J., Harrison D.E., 1974. Studies on the production of pink pigment in Pseudomonas extorquens NCIB 9399 growing in continuous culture. J Appl Bacteriol. 37(1):65-74.

175. Thorsen L., Hansen B.M., Nielsen K.F., Hendriksen N.B., Phipps R.K., Budde B.B., 2006. Characterization of emetic Bacillus weihenstephanensis, a new cereulide-producing bacterium. Appl Environ Microbiol. 72(7):5118-21.

176. Stenfors Arnesen L.P., Fagerlund A., Granum P.E., 2008. From soil to gut: Bacillus cereus and its food poisoning toxins. FEMS Microbiol Rev. 32(4):579-606.

177. Montel MC, Buchin S, Mallet A, Delbes-Paus C, Vuitton DA, Desmasures N, Berthier F., (2014). Traditional cheeses: rich and diverse microbiota with associated benefits. Int J Food Microbiol. 177:136-54.

178. Fox et al., (1996)P.F. Fox, J.M. Wallace, S. Morgan, C.M. Lynch, E.J. Niland, J. Tobin Acceleration of cheese ripening Antoine van Leeuwenhoek, 70 , pp. 175–201

179. E. Caplice, G.F. Fitzgerald, (1999). Food fermentations: role of microorganisms in food production and preservation. International Journal of Food Microbiology, 50 , pp. 131–149

180. R Paul Ross, a, , S Morgana, C Hillb (2002). Preservation and fermentation: past, present and future. International Journal of Food Microbiology.Volume 79, Issues 1–2, 15 Pages 3–16

181. Condon, 1987. Responses of lactic acid bacteria to oxygen. FEMS Microbiological Reviews, 46 (1987), pp. 269–280

182. L. De Vuyst, E.J. Vandamme. (1994). Antimicrobial potential of lactic acid bacteria. Blackie Academic and Professional, London, pp. 91–149

183. K.M. Pruitt, J. Tenovuo, B. Mansson-Rahemtulla, P. Harrington, D.C. Baldone.(1986) Is thiocyanate peroxidation at equilibrium in vivo? Biochim. Biophys. Acta, 870 , pp. 385–391

184. T. Eklund. (1984). The effect of carbon dioxide on bacterial growth and on uptake processes in the bacterial membrane vesicles. Intern. J. Food Microbiol., 1 (1984), pp. 179–185

185. S.E. Lindgren, W.J. Dobrogosz. (1990). Antagonistic activities of lactic acid bacteria in food and feed fermentations. FEMS Microbiol. Rev., 87 (1990), pp. 149–163

186. J.M. Jay (1982). Antimicrobial properties of diacetyl. Appl. Environ. Microbiol., 44 (1982), pp. 525–532

187. Cogan. (1986). The leuconostocs: Milk products. S.E. Gilliland (Ed.), Bacterial Starter Cultures for Foods, CRC Press, Boca Raton, Florida (1986), pp. 25–40

188. A.M. Motlagh, M.C. Johnson, B. Ray (1991). Viability loss of foodborne pathogens by starter culture metabolites. J. Food Protect. (1991), pp. 873–878

189. T. Abee, L. Krockel, C. Hill (1995). Bacteriocins: modes of action and potentials in food preservation and control of food poisoning. International Journal of Food Microbiology, 28 , pp. 169–185

190. I.F. Nes, D.B. Diep, L.S. Havarstein, M.B. Brurberg, V. Eijsink, H. Holo. (1996). Biosynthesis of bacteriocins in lactic acid bacteria. Antoine van Leeuwenhoek, 70 (1996), pp. 113–128

191. M. El Soda, S. A. Madkor, and P. S. Tong (1999). Adjunct Cultures: Recent

Developments and Potential Significance to the Cheese Industry. Marschall rhodia international dairy science award lecture.

192. Johnson, M. E. 1998. Cheese products. Pages 213–249 in Applied Dairy Microbiology. E. H. Marth and J. L Steele, ed. Marcel Dekker, Inc. New York.

193. Fox, P. F., P.L.H. McSweeney, and C. M. Lynch. 1998. Significance on nonstarter lactic acid bacteria in Cheddar cheese. Aust. J. Dairy Technol. 53:83–89.

194. Larsen, P. H. 1996. Microfiltration for pasteurized milk. Pages 232–239 in Heat Treatments and Alternatives Methods. Proc. IDF Symp., Vienna, Austria.

195. H. Blom, C. Mortvedt (1991). Anti-microbial substances produced by food-associated micro-organisms. Biochem. Soc. Trans., 19, pp. 694–698

196. Kulshrestha, E.H. Marth. (1974). Inhibition of bacteria by some volatile and non-volatile compounds associated with milk. I. Escherichia coli. J. Milk Food Technol., 37, pp. 510–516

197. Daly, W.E. Sandine, P.R. Elliker (1970). Interactions of food starter cultures and food-borne pathogens: *Streptococcus diacetylactis* versus food pathogens. Journal of Milk and Food Technology, 35 (1970), pp. 349–357

198. W.H. Holzapfel, R. Geisen, U. Schillinger (1995). Biological preservation of foods with reference to protective cultures, bacteriocins and food-grade enzymes. International Journal of Food Microbiology, 24 (1995), pp. 343–362

199. T.R. Klaenhammer (1993). Genetics of bacteriocins produced by lactic acid bacteria. FEMS Microbiology Reviews, 12 (1993), pp. 39–86

200. Pedro M. Oliveira, Emanuele Zannini, Elke K. Arendt, (2014). Cereal fungal infection, mycotoxins, and lactic acid bacteria mediated bioprotection: From crop farming to cereal products. Food Microbiology Volume 37, February 2014, Pages 78–95

201. R. Coda, A. Cassone, C.G. Rizzello, L. Nionelli, G. Cardinali, M. Gobbetti, (2011) Antifungal activity of *Wickerhamomyces anomalus* and *Lactobacillus plantarum* during sourdough fermentation: identification of novel compounds and long-term effect during storage of wheat bread. Applied and Environmental Microbiology, 77 , pp. 3484–3492

202. F. Dal Bello, C. Clarke, L. Ryan, H. Ulmer, T. Schober, K. Ström, J. Sjögren, D. Van Sinderen, J. Schnürer, E. Arendt. (2007). Improvement of the quality and shelf life of wheat bread by fermentation with the antifungal strain

Lactobacillus plantarum FST 1.7 Journal of Cereal Science, 45 (2007), pp. 309–318

203. C. De Muynck, A.I.J. Leroy, S. De Maeseneire, F. Arnaut, W. Soetaert, E.J. Vandamme, (2004). Potential of selected lactic acid bacteria to produce food compatible antifungal metabolites. Microbiological Research, 159 (2004), pp. 339–346

204. T. Franco, S. Garcia, E. Hirooka, Y. Ono, J. dos Santos, (2011). Lactic acid bacteria in the inhibition of Fusarium graminearum and deoxynivalenol detoxification. Journal of Applied Microbiology, 111 (2011), pp. 739–748

205. J. Magnusson, K. Ström, S. Roos, J. Sjögren, J. Schnürer, (2003). Broad and complex antifungal activity among environmental isolates of lactic acid bacteria. FEMS Microbiology Letters, 219 (2003), pp. 129–135

206. De Filippis F, La Storia A, Stellato G, Gatti M, Ercolini D (2014) A Selected Core Microbiome Drives the Early Stages of Three Popular Italian Cheese. Manufactures. PLoS ONE 9(2): e89680. doi:10.1371/journal.pone.0089680