fermentation

MDPI

Article

Microbial Community Profiling from Natural Whey Starter to Mozzarella among Different Artisanal Dairy Factories in Apulia Region (Italy)

Stefano Castellana, Angelica Bianco, Loredana Capozzi, Laura Del Sambro *, Domenico Simone, Marco Iammarino, Valeria Nardelli, Annamaria Caffò, Carmelinda Trisolini, Antonella Castellana, Elisabetta Catalano, Angelica Milano, Giulia Schino, Roldano Sottili and Antonio Parisi

Istituto Zooprofilattico Sperimentale di Puglia e Basilicata, Viale Manfredonia 2, 71121 Foggia, Italy; stefano.castellana@izspb.it (S.C.); angelica.bianco@izspb.it (A.B.); loredana.capozzi@izspb.it (L.C.); dome.simone@gmail.com (D.S.); marco.iammarino@izspb.it (M.I.); valeria.nardelli@izpb.it (V.N.); annamaria.caffo@izspb.it (A.C.); carmelinda.trisolini@izspb.it (C.T.); antonella.castellana@izspb.it (A.C.); elisabetta.catalano@izspb.it (E.C.); angelica.milano@izspb.it (A.M.); giulia.schino@izspb.it (G.S.); info@acv.it (R.S.); antonio.parisi@izspb.it (A.P.)

* Correspondence: laura.delsambro@izspb.it

Abstract: Mozzarella is one of the most popular unripened Apulian cheeses. Knowledge about microbial composition and variability of artisanal mozzarella and its production chain is increasingly growing. In this study, microbial communities from natural whey starters to end products from four renowned Apulian artisanal dairy factories have been explored by means of 16S metagenomics. The chemical properties of mozzarella samples were also detected and analyzed. *Lactobacillus* is the core acidifying component of the used starters, while some psychrophilic or contaminants bacteria appear in site-specific products. Biodiversity was found to be quite similar between the whey and mozzarella sample pools, while a significant variability among production sites (factories) has been detected. Furthermore, mozzarella microbial diversity seems to be in positive correlation with its lactic acid content. Targeted metagenomics would then be a powerful and relatively quick technique to characterize the microbiological variability of traditional milk-based foods.

Keywords: mozzarella; dairy factory; microbiome; 16S rDNA; next-generation sequencing

check for updates

Citation: Castellana, S.; Bianco, A.; Capozzi, L.; Del Sambro, L.; Simone, D.; Iammarino, M.; Nardelli, V.; Caffò, A.; Trisolini, C.; Castellana, A.; et al. Microbial Community Profiling from Natural Whey Starter to Mozzarella among Different Artisanal Dairy Factories in Apulia Region (Italy). *Fermentation* **2023**, *9*, 911. https://doi.org/10.3390/fermentation9100911

Academic Editor: Kai Zhong

Received: 19 September 2023
Revised: 9 October 2023
Accepted: 11 October 2023
Published: 16 October 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

High-moisture mozzarella is one of the most popular unripened Italian cheeses, belonging to the "Pasta Filata" class. Given the differences in milk's intrinsic features, product intermediates, production, and distribution steps, a large variety of "Pasta filata" cheeses exist today, mainly produced in Italy and the United States [1,2]. Within Italy, the most famous "Pasta filata" dairy products include mozzarella, "caciocavallo", and "provolone", with different chemical and nutritional properties [3]. Mozzarella di Bufala Campana has a protected designation of origin ("PDO") and derives only from water buffalo milk (*Bubalus bubalis*), farmed in Southern Italy (mainly Campania region), while different types of "caciocavallo" and "provolone" are manufactured across many central–southern Italy regions. The typical Apulian mozzarella derives from cow's raw milk: its production is interspersed from small family-run factories to large dairy products companies.

Generally, mozzarella production begins with raw milk collection and filtration; then, coagulation is achieved through the usage of calf rennet: the forming filamentous and fibrous "cagliata" (curd) is cut and acidified by different methods, such as the addition of commercial or natural bacterial cultures. The subsequent steps include curd cutting, salting, heating, forming, and cooling, while the final product is mainly stored at 4 °C.

The acidification step is critical for the microbiological, biosafety, and organo-lectic properties of the products [4,5]. Numerous acidification methods can be applied to the curd,

e.g., the usage of citric acid or natural acidification by defined commercial starters ('CS') or natural whey cultures ('NWC') that are retained from preceding productions. Commercial starters consist of lyophilized lactic acid bacteria ('LAB') strains, such as *Streptococcus salivarius subsp. Thermophilus* or *Lactobacillus delbrueckii subsp. bulgaricus*.

On the other hand, as the artisanal production process is microbiologically complex, reliable molecular methods for microbiome identification, contamination control, and process standardization are required. The first experimental approach consisted of direct polymerase chain reactions on the target DNA of specific strains from uncultured6 or cultured [6,7] samples. However, during the last decade, food microbiology studies have taken advantage of next-generation sequencing ("NGS") technologies, especially 16S-gene-based metagenomics. It allows a quick, culture-independent, high-throughput analysis of microbial diversity.

The first deep sequencing approach [8] for mozzarella microbiome study regarded raw milk, natural whey culture, curd, and finite products from two production sites of buffalo mozzarella (total sample number: 10). V1-V3 16S rDNA pyrosequencing revealed higher biodiversity of raw milk with respect to downstream products. *Acinetobacter* and *Pseudomonas* spp. were predominant within milk, while microbial composition for NWC, curd, and mozzarella was quite stable, with *Streptococcus thermophilus* and *Lactobacillus delbrueckii* as the most represented species.

High-throughput sequencing was also used to study microbial communities across 20 cow's milk mozzarella samples, even collected from two production sites at different times [9]. Predominant taxa were the ones typically used within starters, i.e., *S. thermophilus*, *L. delbrueckii subsp. bulgaricus*, and *L. helveticus*, while many other less abundant microbes (e.g., spoilage or raw-milk-associated) were retrieved. Artisanal mozzarella was found to be more microbiologically complex than industrial-made ones, while acidification techniques did not affect microbiome diversity.

A larger survey was performed by Marino et al. 2018 [10]. rDNA-based metagenomics was also applied to 39 high-moisture mozzarella samples (from cow's and water buffalo milk) produced with different acidification methods by 14 industrial firms. Biodiversity in cow's milk mozzarella was found to be higher than the water buffalo one. Indeed, microbial abundance profiles from these two sources are clearly distinct, with Lactobacillus predominating buffalo's milk samples while psychrophilic genera prevail within the cow's milk.

Microbiome from whey starters for cow and water buffalo mozzarella production was studied by means of several techniques on bacterial isolates and metagenomic samples [11]. Differences in lactic bacteria abundance emerged between the two starter types, with water buffalo starter being characterized by an even lactobacilli distribution. Starters for cow's milk mozzarella had a reduced quote of *lactobacilli* in favor of *streptococci* and other groups. Microbial variability within raw milk intermediates and end products was recently explored by the next-generation sequencing approach on water buffalo mozzarella within two production sites. *Lactobacillus* spp. and *Streptococcus thermophilus* were confirmed to be the main components of natural whey starters [12].

V.A.L.O.Re Puglia ("Valorizzazione dell'Agroalimentare Lattiero-caseario di Origine della Regione PUGLIA") is an Apulian region research project, aimed at the characterization, valorization, and protection of the Apulian traditional mozzarella and its production chain. The "Mozzarella di Gioia del Colle" PDO defines the intrinsic requirements of the traditional mozzarella. Among them, raw milk must originate from cow farms located in the Murgia geographical area, while curd acidification has to be carried out through the usage of autochthonous (factory-specific) whey starter. Furthermore, this PDO implies other specific aspects, from farming methods to consistency, shape, color, and dimension of the final products.

Whey starters and finite products for this research have been collected during 2021 from four different dairy plants located in the southeast part of Bari province and involved in "Mozzarella di Gioia del Colle" PDO production. In this work, we have tried to define

the microbiological profile of their traditional whey starters from selected factories and their evolution within the final product. Secondarily, the possible relationship between mozzarella's chemical features and microbial profile variability was also investigated.

2. Materials and Methods

2.1. Sample Collection

A total of 57 samples (18 mozzarella, indicated as "M"; 39 natural whey starters, indicated as "S") were collected from January to December 2021 from four factories located in the Apulia region (Supplementary Table S1). Conventionally, each dairy company has been assigned a numeric code: "f1" to "f4". All the samples were collected and stored at +4 °C until arrival at the laboratory and frozen at −20 °C if it was not possible to proceed immediately with DNA extraction.

2.2. DNA Extraction

Total DNA extraction was performed using two kits, depending on the matrices. For M samples, we used DNeasy mericon Food Kit (Qiagen, Hilden, Germany) following the manufacturer's protocol, whereas for the whey starters, we used DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) with a slight initial modification. One mL of samples was taken and centrifuged for 5 min at 13,000 rpm. The fat layer was removed with a micropipette. The step was repeated until all impurities were (approximately 4–5 times). Then, the supernatant was removed, and the cell was resuspended in 180µL of ATL buffer and incubated at 56 °C for 10 min. After incubation, the manufacturer's instruction for DNA extraction for cells was followed. The concentration of the extracted nucleic acid was determined using a Qubit® 3.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). The DNA was stored at −20 °C.

2.3. Sequencing

The 16S rRNA libraries were prepared using the protocol recommended in the Illumina 16S metagenomic library preparation instruction (Illumina, San Diego, CA, USA). Briefly, the hypervariable V3-V4 region of the 16S rRNA gene was amplified using Illumina primers, which include the Illumina sequencing adapter (forward primer: 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGNGGCWGCAG-3′; reverse primer: 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVG GGTATCTAATCC-3′). An amplicon of ~460 bp was visualized by electrophoresis performed by the QIAxcel Advanced System using a QIAxcel DNA Screening Kit (Qiagen, Hilden, Germany). Then, dual-index barcodes were applied to the amplicons PCR, following the manufacturer's instructions. The quality of purified libraries was visualized by electrophoresis performed by QIAxcel Advanced System (Qiagen, Hilden, Germany), and fluorometric quantification was measured by Qubit® 3.0 Fluorometer (ThermoFisher Scientific, Waltham, MA, USA). According to the Illumina protocol, the libraries were normalized and pooled in an equimolar and then paired-end sequenced (2 × 300) on the MiSeq platform (Illumina, San Diego, CA, USA) in the Molecular Biology Laboratory of Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata (Putignano section).

2.4. Computational and Statistical Analyses

Raw reads from sequencing run were initially quality checked by the FastQC [13] tool, while DADA2 [14] (version 1.26) was used for read demultiplexing, filtering, forward/reverse merging, and ASV ("Amplicon Sequence Variants") inference and count table generation. The taxonomic classification for sample-specific sequences was performed through a naïve Bayes classifier trained on SILVA v138.1 [15]. Alpha and beta diversity analyses were carried out in R v4.2.2 environment [16] and RStudio 2023.1.0 [17], using the following packages: Phyloseq v1.40.0 [18], Microbiome v1.20.0 [19], Vegan v2.6-4 [20]. Sampling completeness was tested through the computation of estimated sample coverage [21], while statistical comparison for alpha index distributions (Chao1 [22], Shannon [23], inverse

Simpson [24], Pielou's evenness [25]) was performed through Wilcoxon rank sum exact test, for both "Mozzarella (M)/Natural whey starter (S)" and "Cheese Factory" sample classes. While the first abovementioned indexes evaluate how much species diversity within a community (higher values meaning greater diversity), the latter index is one of the most popular measures of species evenness, i.e., how the different species are evenly (or not) abundant in such community (Pielou's evenness ranging from 0, no evenness, to 1, perfect evenness where all species have equal relative abundance). Beta diversity ("M vs. S" and across cheese factories) was assessed through the PERMANOVA test (9999 permutations) on two pairwise dissimilarity matrices (Bray–Curtis [26] and Jaccard [27] distances).

The identification of taxa with relevant differential abundance profiles across mozzarella and starter sample groups was performed as follows: ultra-rare (less than five counts in at least the 5% of samples), mitochondrial-, and chloroplast-associated ASVs were discarded, while surviving ASVs were collapsed at the genus level. Then, classical statistical methods (based on fold-change) were considered using Phyloseq ad hoc functions. Secondarily, given the limited group size and asymmetry, we adopted a completely different statistical approach by considering three "feature selection" methods, namely Coda-lasso [28], Clr-lasso [29], and Selbal [30].

They were used to select the most relevant differences in genus abundance profiles between the "Mozzarella" and "Starter" sample groups. These algorithms were applied as described in https://malucalle.github.io/Microbiome-Variable-Selection/ (accessed on 20 April 2023) and were chosen for a number of reasons: they were adequately conceived for sparse and compositional data (like microbiomes); their results could easily be compared each other; they are not computationally intensive and are easily be implemented in the R environment.

Correlation analyses between chemical and microbiological data were performed in the R environment by Spearman's rank correlation rho test after checking for distribution normality. We performed correlations among chemical data (%lactic acid, %lactose, %humidity, %fat) and alpha diversity distributions. Furthermore, we put in relationship chemical content with microbial abundance profiles within the "M" sample group.

2.5. Lactic Acid Determination

A total of 18 mozzarella samples, deriving from the same production lot and factory site as the sequenced ones, were investigated for the analysis of some relevant chemical components at the Chemistry Laboratory of Istituto Zooprofilattico Sperimentale della Puglia e della Basilicata. The procedure proposed [31] was adopted for the determination of lactic acid as a lactate ion. Briefly, 4 g of homogenized mozzarella samples were placed in a Falcon tube and mixed with 40 mL of a NaOH 8.5×10^{-3} M solution. The extraction was obtained by placing the tube in an ultrasonic bath, using a frequency of 100 Hz, heating at 40 °C for 10 min, and then vortexing. Sample purification was obtained by centrifugation (1500× *g*, 10 min at room temperature) and consequent microfiltration of supernatant using Minisart® GF syringe filters (0.2 µm, Sartorius AG, Goettingen, Germany). Prior to chromatographic analysis (injection volume: 25 µL), the high amount of chloride in the sample was removed by purifying ~1 mL of filtrate using OnGuard II Ag chromatography filters (Thermo Fisher Scientific Inc., Waltham, MA, USA), previously activated with 1 mL of ultrapure water. All the chromatographic determinations were performed using a high-performance ion chromatography system (Thermo Scientific™ Dionex™ ICS-6000 HPIC™ System, Thermo Fisher Scientific Inc., Waltham, MA, USA). This system was composed of an SP single pump (ICS-6000), a Dionex anion self-regenerating suppressor, a gradient mixer (Dionex GM-4, 2 mm), and a DC detector set to conductivity mode. The temperature of the column compartment was set at 20 °C. The electrochemical suppression was set at the recommended voltage. The chromatographic separations were accomplished using an anion-exchange column IonPac® AS9-HC (250 mm × 4 mm i.d., particle size: 9 µm) eluted in gradient mode at a flow rate of 1.0 mL min^{-1}, of 2 mobile phases composed of different

concentrations (0.9 and 28.5 mM) of Na_2CO_3. The limit of quantification of lactic acid in the matrix was equal to 1.9 mg kg^{-1}.

2.6. Other Chemical Analyses

The determination of moisture and fat percentage was accomplished according to AOAC official methods [32] within the abovementioned laboratory. Briefly, moisture determination was carried out by weighing 5 g of homogenized sample in a quartz crucible and then placing it in an oven at 80 °C for 3 h until constant weight. The determination of total fat was obtained by weighing 4 g of the sample together with 4 g of Extrelut® powder in a vessel tube and placing the sample in an oven at 80 °C for 15 min. Afterward, 60 mL of solvent composed of (chloroform/methanol/petroleum ether, 2:1:1, *v*/*v*/*v*) was added, and the sample was placed in an ASE (accelerated solvent extraction) system (Dionex ASE 350, Thermo Fisher Scientific Inc., Waltham, MA, USA). The extraction was completed after 15 min, the solvent was dried under nitrogen flow, and the final residue was weighed. Finally, the determination of lactose was obtained using an EnzytecTM Liquid Lactose/D-Glucose kit (R-Biopharm Italia Srl, Milan, Italy), with a limit of quantification equal to 50 mg kg^{-1} of the sample.

3. Results

A total of 57 samples (39 for "S" class, 18 for "M") were analyzed: they were subdivided into 17 (Factory site 1, 6 out of them for "M" class), 13 for Factory 2 (4 "M"), 13 for Factory 3 ("4" M), and 14 for the fourth site (4 "M") and collected along the 2021 (details in Supplementary Table S1).

Sequencing throughput ranged from 64,000 to 195,000 raw reads (Supplementary Table S3), although the proportion of high-quality DADA2-filtered reads ranged from 8200 to 169,000. Such proportion was remarkably variable for the "Mozzarella" sample class (7 to 66%), while a more homogeneous yield was obtained from the "Starter" one (54–89% high-quality reads, usable for downstream analyses). Globally, the high-quality read proportion for class "M" was lower than the one obtained from class "S" (Wilcoxon rank sum exact test, p-value = $4.479e^{-13}$).

From raw reads, DADA2 implementation allowed the identification of a total of 4272 Amplicon Sequence Variants (Supplementary Table S2) that have been taxonomically collapsed at 501, 392, and 201 groups, respectively, species, genera, and bacterial families. Within samples, the number of identified ASVs ranged from 9 to 250.

ASV taxonomical assignment evidenced *Lactobacillus helveticus* as the predominant constituent of starters and mozzarella sample groups, with a relative abundance ranging from 60 to 99% across every sample. Differences in composition emerge among factories and samples, as shown in Figure 1 (*Lactobacillus* abundance data removed). Factory 1 shows a moderate proportion of *Acinetobacter*, with an "S" sample with a relevant presence of genera *Prevotella* and *Porphyromonas*. Acinetobacter was relatively abundant also in Factories 2 and 4, while Factory 3 was characterized by a more reduced presence of L. helveticus and a major relative abundance of *Streptococcus*, *Thermus*, *Pseudomonas*. *Acinetobacter*, and *Streptococcus* were among the most represented genera throughout factories and samples, with a relative abundance ranging from 0.05 to 30% (Figure 1, Supplementary Table S7).

An estimated sample coverage ranging from 0.982 to 1 was achieved for all samples, meaning that the proportion of singleton (taxonomic groups identified by a single read count) was very small or null. Inter-group statistical comparison of alpha diversity measures revealed a mildly significant greater complexity for the "Mozzarella" group, as exemplified by Wilcoxon tests for Chao, Shannon, and inverse Simpson index distributions (p-values = 0.039, 0.026, and 0.039, respectively) (Figure 2a, Supplementary Tables S4 and S6). Pielou's evenness results are statistically comparable between the two sample groups (Figure 2b, Supplementary Tables S4 and S6), even if taxa abundance profiles for "Mozzarella" samples were more evenly distributed (i.e., higher evenness values with respect to class "S" samples).

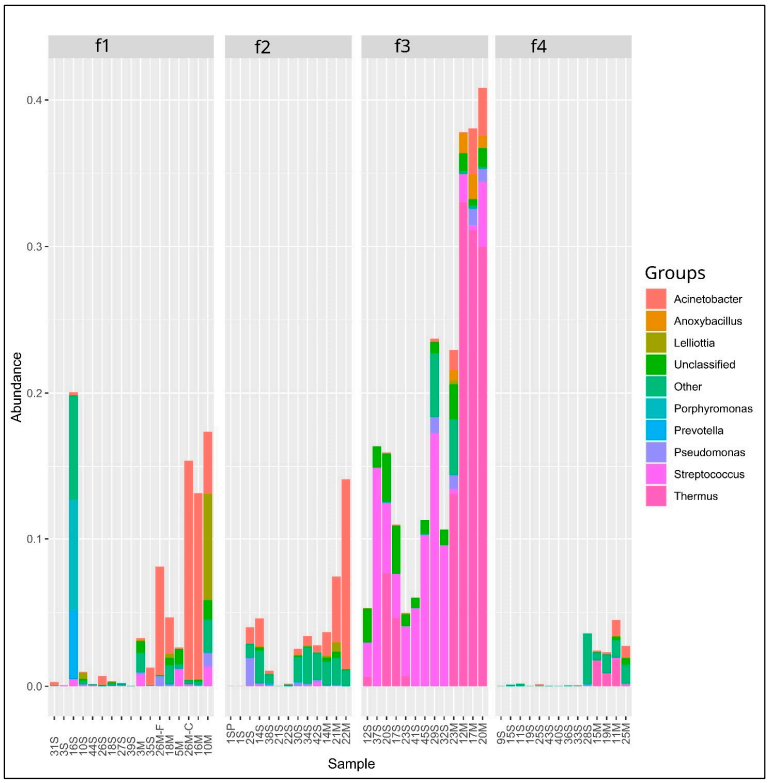
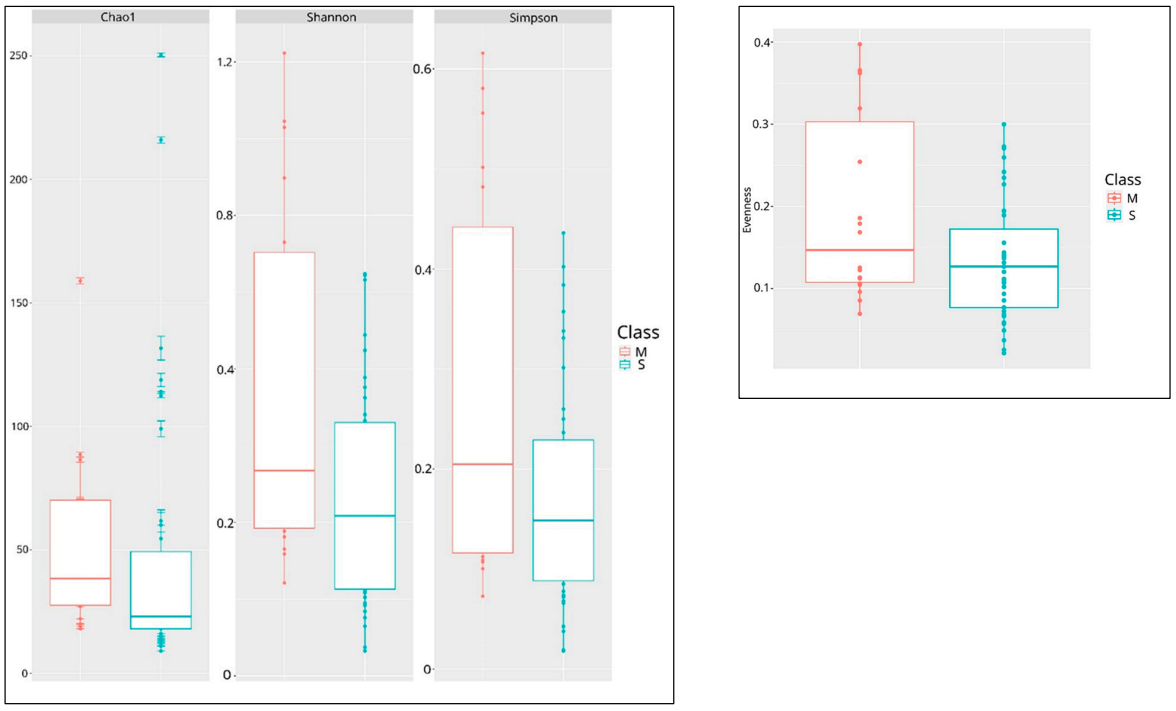


Figure 1. Relative abundance of the top 10 represented group across samples (*x*-axis, sample identifiers are shown) and factories (f1, f2, f3, f4), with the exclusion of "Lactobacillus"; "Unclassified": amplicon sequence variants with no taxonomical classification; "Other": taxonomic groups other than the listed ones.



(**a**) (**b**)

Figure 2. (**a**) Boxplot for class-specific ("M": mozzarella group, "S": "Starter" group) distributions of alpha diversity measures Chao1, Shannon, and inverse Simpson. Dots represent sample-specific values. (**b**) Boxplot for group-specific distributions for Pielou's evenness measure.

At the "factory-specific" level, an increase in microbial diversity was evident for Factories 1 and 3 mozzarella samples (p-values = 0.0067, 0.009 for Shannon and inverse Simpson measures, respectively; p-values = 0.0028 for analogs tests for Factory 3). The evenness index results were significantly higher for "M" samples from the two factories (p-values = 0.033 and 0.0028, Supplementary Table S6). However, it has to be noted that sample sizes at the factory level are limited (number of samples = 17, 13, 13, and 14 for, respectively, Factories 1, 2, 3, and 4). We also evaluated the "Collection month" as a possible influencing factor on sample variability, but no significant effect was found.

Consistently with alpha diversity comparative analyses, "Mozzarella" samples' microbial composition was almost indistinguishable from the analog in "S" samples (p-values for PERMANOVA test around 0.05, for Jaccard- and Bray–Curtis-based inter-samples dissimilarity, Supplementary Table S6). While differences between the two matrices are minimal, microbial composition results are associated with each production environment. Indeed, samples deriving from the same factory tend to significantly cluster together, with few highly dissimilar samples for f1 and f2 (PERMANOVA test p-values = $6.01e^{-4}$ and 0.0007, for Jaccard- and Bray–Curtis-based distances, Figure 3, Supplementary Table S6).

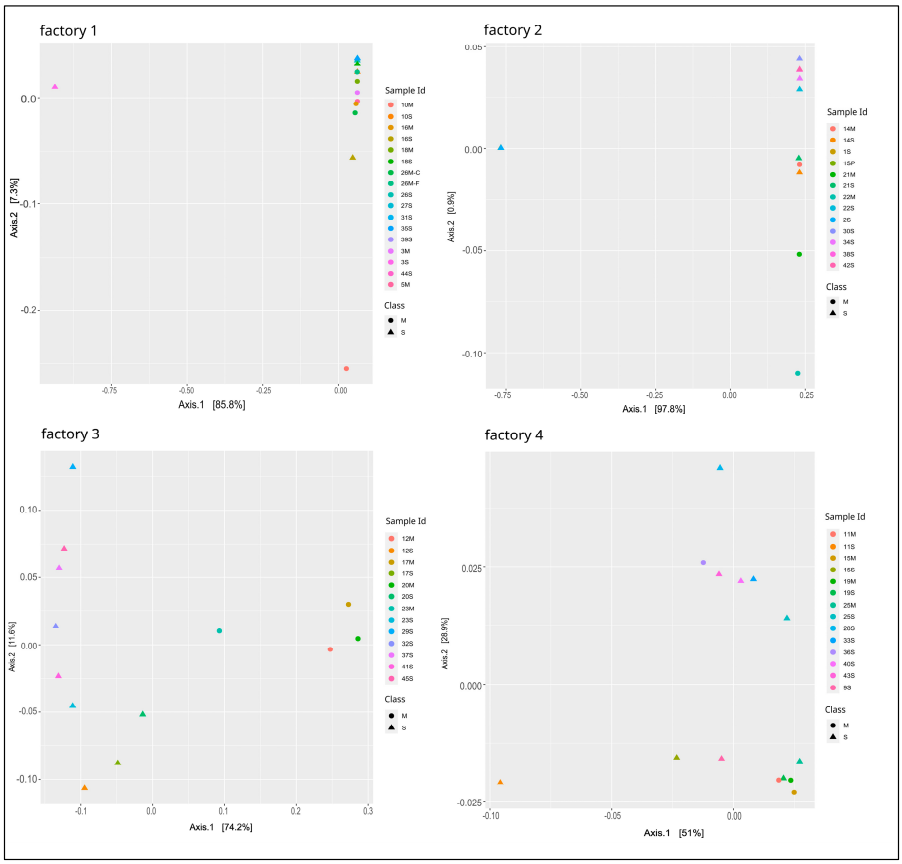


Figure 3. Principal component analysis plots, divided by factory. Samples and classes are evidenced by color and shapes, respectively. Beta diversity distances across samples: Bray–Curtis dissimilarity.

Beyond the predominance of Lactobacillus across samples and factories, three alternative compositional data analysis methods were implemented in order to identify even subtle variations in genus-level (43 total genera, Supplementary Table S6) quantities from "S" to "M" matrices. Such analysis confirmed the decrease in *Lactobacillus* within "Mozzarella" samples, while two of three methods (coda_lasso, clr_lasso) also put in evidence the down-representation of *Brachybacterium* (Figure 4a). The *Acinetobacter* and the *Burkholderia-Caballeronia-Paraburkholderia* genera appeared to be over-abundant in the "Mozzarella" rather than the "Starter" group (Figure 4b, Supplementary Table S8). Other taxonomical groups, evidenced by one (or two) implemented methods, were *Streptococcus* (mildly over-

abundant in the "S" class), *Thermus*, *Enterobacteriaceae* (specifically, an un-classified genus within this family), *Raoultella*, *Candidatus hemobacterium*, *Shewanella*, and *Deinococcus* (mildly over-abundant in the "M" class, Figure 4b, Supplementary Table S8).

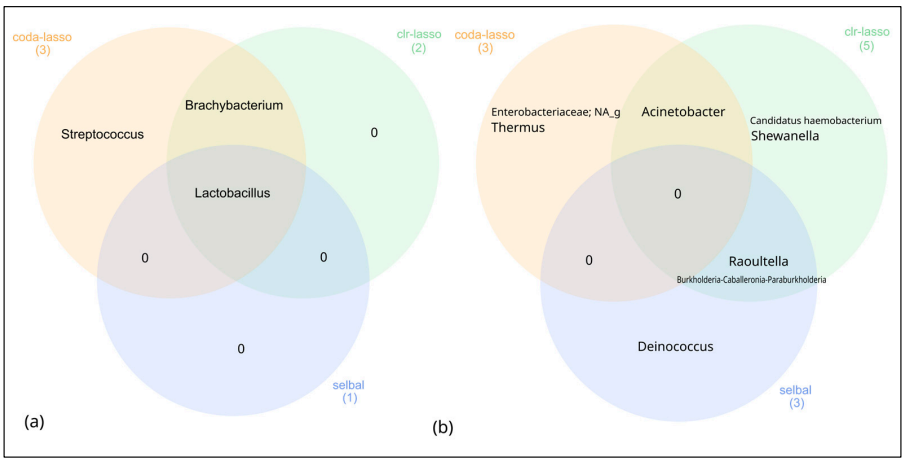


Figure 4. (**a**) Venn diagram for over-represented taxa in the "S" group for the three differential abundance analysis methods. (**b**) Venn diagram for over-represented taxa in the "M" group.

Correlation analyses between microbial communities and chemical composition for the "Mozzarella" matrix evidenced some slightly significant patterns for the "lactic acid" component (chemical data within Supplementary Table S9). A mild positive correlation was detected between "%lactic acid" and "Shannon diversity" (Spearman's rank correlation rho test, rho = 0.524, S = 460.69, *p*-value = 0.024, Figure 5), "Inverse Simpson diversity" (rho = 0.528, S = 456.68, *p*-value = 0.024), and "Pielou's evenness" (rho = 0.527, S = 456.59, *p*-value = 0.023). No other significant correlations emerged across genus-specific abundances and chemical measurements within mozzarella samples.

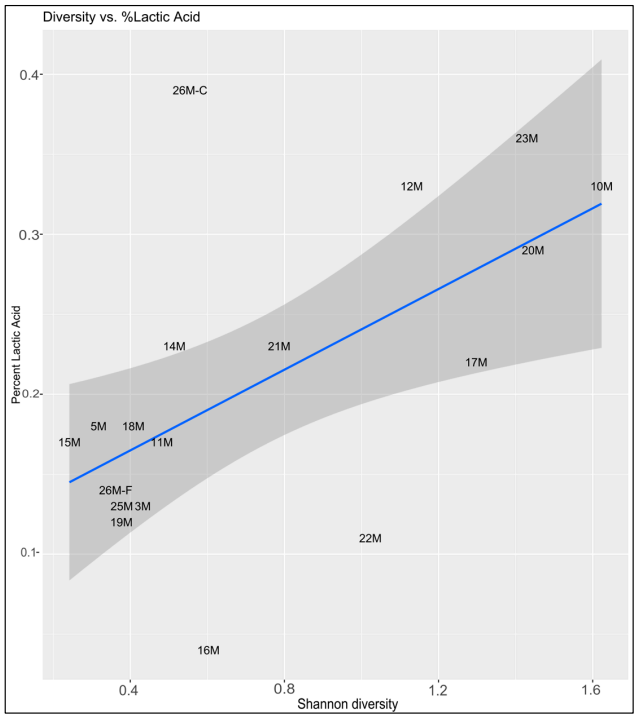


Figure 5. Scatterplot between Shannon diversity (*X*-axis) and lactic acid proportion (*Y*-axis); data derived from independent samples within the same mozzarella lot; labels indicate identifiers of sequenced samples. Blue line: smooth regression line with confidence interval (gray).

4. Discussion

In this work, we investigated the variability of microbial communities within natural whey starters and mozzarella samples across four different dairy factories involved in the Apulian region V.A.L.O.Re Project. Around 15 samples from each production site were received and processed over one year: globally, 39 whey starters and 18 mozzarella samples were suitable for downstream computational analyses. A larger and more symmetrical sampling size would have surely allowed a more accurate investigation, especially for mozzarella.

Nonetheless, the considered sample set was in line with the experimental size reported in analog publications [9,10,12], in which a few tens of different matrices (milk, starters, dairy finite products) were processed. However, results from microbiological studies are hardly comparable among each other, given the different adopted experimental techniques, computational methods for species detection, classification and quantification, study design and sample size, intrinsic properties of raw milk (e.g., buffalo's or cow's milk, farming modality, etc.), differences in production, and distribution chains. Indeed, the literature shows the variability of the total number of identified OTUs ("operational taxonomy units") across investigated samples, typically from a few tens to around 1000, even from the same study [10]. We also highlight this aspect, with classified amplicon sequence variants oscillating between tens and hundreds.

Among microbial sequences, the highest amount was represented by *Lactobacillus helveticus*, which constitutes the main acidifying bacterium for all four investigated production sites. Thus, the four distinct autochthonous starters used for the "Mozzarella di Gioia del Colle PDO" production are mainly composed of the same microorganism.

Together with *L. delbrueckii* and *Streptococcus thermophilus*, *L. helveticus* are common NWS components for various types of mozzarella [7–9,11,12]. Genus *Streptococcus*, frequently found as the key regulator of acidification [8–10], is also a relevant component within one of the investigated NWCs (used in Factory 3).

Microbiome analyses revealed the impact of environmental contaminants on the finite products within all production sites. For example, the *Acinetobacter* genus was found at relatively high frequency in mozzarella from Factory 1 and 2: it has been commonly recognized as a spoilage organism in several works [9,10]. Heat-resistant *Thermus* genus has been identified in mozzarella samples from Factory 3 and also in samples from commercial brands [10]. Advanced feature-selection techniques (coda-lasso, clr-lasso, and selbal) strengthened the fact that microbial profile is almost invariable from starter to mozzarella, although a decrease in the core NWC component, *L. helveticus* can be observed across the four investigated factories. Such reduction, together with the presence of spoilage microorganisms (*Acinetobacter*, *Thermus*), would reflect a mildly significant larger biodiversity within the mozzarella group (Supplementary Table S6). Indeed, we confirmed that the production site plays a relevant role in the variation in the microbiological profile of end products [9]. Two of the investigated sites (Factories 1 and 3) show a statistically significant increase in species biodiversity (Shannon, inverse Simpson indexes); furthermore, in these factories, species abundance tends to be less asymmetric within the mozzarella group (higher Pielou's evenness index with respect to group S).

In addition to microbiological profiling, mozzarella samples from the same production lot and dairy factory sites were inspected for their organo-lectic features. Nutritional and chemical properties of dairy products have been deeply explored by comparing, for example, a series of traditional and industrial Italian dairy products [33,34]. We found a mildly significant positive correlation between alpha diversity indexes and the percentage of lactic acid within mozzarella samples. This can be due to the decrease in lactose, which would influence the metabolism and growth of starter bacteria (mainly *Lactobacillus helveticus*) and favor the increase in non-lactic bacteria [35]. However, more bacterial diversity has often been found across "M" samples from only one site (Factory 3, samples: 20M, 23M, 17M, 12M). To our knowledge, correlation analyses among microbiological and chemical properties of PDO Apulian mozzarella are quite novel, although the detected

significant correlation should be considered with extreme caution, given the limited sample size together and the significant "environmental" (factory site) effect.

The marker-based next-generation sequencing approach results in a quite efficient method for the microbial investigation of intermediates and final dairy products. However, the selection of which gene marker and sequencing strategy to consider in microbial taxonomical identification is largely debated [36,37]. In our approach, investigation of microorganisms by means of 16S V3-V4 amplicon sequencing is limited to the bacterial kingdom until the "genus" level. Furthermore, it has to be noted that NGS-based microbial profiling does not return conclusive and accurate results. In a diagnostic framework, a more precise and reliable species detection can be achieved by implementing other molecular methods (e.g., culture-based, polymerase chain reaction-based, etc.).

In addition, amplicon-based sequencing does not provide the possibility to explore all genomic properties of the identified microorganisms. For example, investigating the presence of prophage, antibiotic resistance, and/or virulence factor genes, contaminating microorganisms would be of great interest, especially for what concerns starter viability [38] or final product deterioration [39].

In perspective, we plan to extend this preliminary research work by focusing on microbial variability on a deeper longitudinal scale. One factory site will be considered for the collection of raw milk, natural whey starters, and mozzarella samples during a 6-month-long period, an interval time of one sampling per week and, if possible, five samples for each sample type. Secondarily, microbiological–chemical preliminary research, analog to the one presented, will involve 3–4 dairy sites from other areas of the Apulia region that produce dairies using chemical starters. Finally, the creation of a sequence databank for the collected whey starters should be of interest to both the microbiological research community and agri-food industry stakeholders.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/fermentation9100911/s1, Table S1: Sample metadata (identifier, class, dairy factory site, and collection date); Table S2: Amplicon sequence variant counts by sample and taxonomical classification; Table S3: Sequencing statistics per sample, as computed by RDADA2 tool; Table S4: Per sample alpha biodiversity indexes (columns D-H); Table S5: Per sample ASV counts, collapsed at the "genus" level; Table S6: Alpha and beta diversity statistical tests: comparison type, statistics, measures, and adjusted p-values are indicated; Table S7: Per sample relative abundance for top 10 highly represented genera; Table S8: Complete results of feature-selection methods applied to "S vs. M" sample comparison. Note that two out of three methods (coda-lasso and clr-lasso) calculate "coefficients" for significantly selected features (i.e., genera); Table S9: Chemical determinations (percentage) for lactose, lactic acid, fat, and moisture for mozzarella samples within the same factory site and lots (of sequenced ones).

Author Contributions: Conceptualization, S.C., L.D.S., D.S. and A.P.; methodology, A.B., L.C., L.D.S., M.I., V.N., A.C. (Annamaria Caffò), A.C. (Antonella Castellana), C.T., E.C., A.M. and G.S.; software, S.C. and D.S.; writing—original draft preparation, S.C.; writing—review and editing, D.S., A.B., L.D.S. and A.P.; supervision, A.P.; project administration, R.S. and A.P.; funding acquisition, R.S. and A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by "Valorizzazione dell'Agroalimentare Lattiero-Caseario di Origine della Regione Puglia (V.A.L.O.Re Puglia)" Research Project. Grant number (CUP): F34I20000100002.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data (raw sequence files) presented in this study are openly available at: https://www.ncbi.nlm.nih.gov/sra/PRJNA1015551 (13 September 2023)

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

References

1. Licitra, G.; Carpino, S. The Microfloras and Sensory Profiles of Selected Protected Designation of Origin Italian Cheeses. *Microbiol. Spectr.* **2014**, *2*, CM-0007-2012. [CrossRef] [PubMed]
2. Kindstedt, P. Symposium Review: The Mozzarella/Pasta Filata Years: A Tribute to David M. Barbano. *J. Dairy Sci.* **2019**, *102*, 10670–10676. [CrossRef] [PubMed]
3. Angelis, M.D.; Gobbetti, M. *Pasta-Filata Cheeses: Traditional Pasta-Filata Cheese*; Elsevier: Amsterdam, The Netherlands, 2016.
4. Losito, F.; Arienzo, A.; Bottini, G.; Priolisi, F.R.; Mari, A.; Antonini, G. Microbiological Safety and Quality of Mozzarella Cheese Assessed by the Microbiological Survey Method. *J. Dairy Sci.* **2014**, *97*, 46–55. [CrossRef] [PubMed]
5. Gatti, M.; Bottari, B.; Lazzi, C.; Neviani, E.; Mucchetti, G. Invited Review: Microbial Evolution in Raw-Milk, Long-Ripened Cheeses Produced Using Undefined Natural Whey Starters. *J. Dairy Sci.* **2014**, *97*, 573–591. [CrossRef] [PubMed]
6. Coppola, S.; Blaiotta, G.; Ercolini, D.; Moschetti, G. Molecular Evaluation of Microbial Diversity Occurring in Different Types of Mozzarella Cheese. *J. Appl. Microbiol.* **2001**, *90*, 414–420. [CrossRef]
7. De Candia, S.; De Angelis, M.; Dunlea, E.; Minervini, F.; McSweeney, P.L.H.; Faccia, M.; Gobbetti, M. Molecular Identification and Typing of Natural Whey Starter Cultures and Microbiological and Compositional Properties of Related Traditional Mozzarella Cheeses. *Int. J. Food Microbiol.* **2007**, *119*, 182–191. [CrossRef]
8. Ercolini, D.; De Filippis, F.; La Storia, A.; Iacono, M. "Remake" by High-Throughput Sequencing of the Microbiota Involved in the Production of Water Buffalo Mozzarella Cheese. *Appl. Environ. Microbiol.* **2012**, *78*, 8142–8145. [CrossRef]
9. Guidone, A.; Zotta, T.; Matera, A.; Ricciardi, A.; De Filippis, F.; Ercolini, D.; Parente, E. The Microbiota of High-Moisture Mozzarella Cheese Produced with Different Acidification Methods. *Int. J. Food Microbiol.* **2016**, *216*, 9–17. [CrossRef]
10. Marino, M.; Dubsky de Wittenau, G.; Saccà, E.; Cattonaro, F.; Spadotto, A.; Innocente, N.; Radovic, S.; Piasentier, E.; Marroni, F. Metagenomic Profiles of Different Types of Italian High-Moisture Mozzarella Cheese. *Food Microbiol.* **2019**, *79*, 123–131. [CrossRef]
11. Marasco, R.; Gazzillo, M.; Campolattano, N.; Sacco, M.; Muscariello, L. Isolation and Identification of Lactic Acid Bacteria from Natural Whey Cultures of Buffalo and Cow Milk. *Foods* **2022**, *11*, 233. [CrossRef]
12. Levante, A.; Bertani, G.; Marrella, M.; Mucchetti, G.; Bernini, V.; Lazzi, C.; Neviani, E. The Microbiota of Mozzarella Di Bufala Campana PDO Cheese: A Study across the Manufacturing Process. *Front. Microbiol.* **2023**, *14*, 1196879. [CrossRef] [PubMed]
13. Andrews, S. FASTQC. A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 11 April 2023).
14. Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High-Resolution Sample Inference from Illumina Amplicon Data. *Nat. Methods* **2016**, *13*, 581–583. [CrossRef] [PubMed]
15. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596. [CrossRef] [PubMed]
16. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
17. Posit Team. RStudio: Integrated Development Environment for R. 2023. Available online: http://www.posit.co/ (accessed on 11 April 2023).
18. McMurdie, P.J.; Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE* **2013**, *8*, e61217. [CrossRef]
19. Lahti, L.; Shetty, S. Tools for Microbiome Analysis in R. Version. 2017. Available online: https://github.com/microbiome/microbiome (accessed on 30 May 2023).
20. Oksanen, J.; Simpson, G.; Blanchet, F.; Kindt, R.; Legendre, P.; Minchin, P.; O'Hara, R.; Solymos, P.; Stevens, M.; Szoecs, E.; et al. *Vegan: Community Ecology Package*; R Foundation for Statistical Computing: Vienna, Austria, 2022.
21. Chao, A.; Jost, L. Coverage-Based Rarefaction and Extrapolation: Standardizing Samples by Completeness Rather than Size. *Ecology* **2012**, *93*, 2533–2547. [CrossRef] [PubMed]
22. Chao, A. Non-Parametric Estimation of the Number of Classes in a Population. *Scand J. Stat.* **1984**, *11*, 265–270.
23. Shannon, C.E.; Weaver, W.; Wiener, N. The Mathematical Theory of Communication. *Phys. Today* **1950**, *3*, 31–32. [CrossRef]
24. Simpson, E.H. Measurement of Diversity. *Nature* **1949**, *163*, 688. [CrossRef]
25. Pielou, E.C. The Measurement of Diversity in Different Types of Biological Collections. *J. Theor. Biol.* **1966**, *13*, 131–144. [CrossRef]
26. Bray, J.R.; Curtis, J.T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **1957**, *27*, 325–349. [CrossRef]
27. Jaccard, P. Nouvelles Recherches Sur La Distribution Florale. *Bull. Soc. Vaud. Sci. Nat.* **1908**, *44*, 223–270.
28. Lu, J.; Shi, P.; Li, H. Generalized Linear Models with Linear Constraints for Microbiome Compositional Data. *Biometrics* **2019**, *75*, 235–244. [CrossRef] [PubMed]
29. Zou, H.; Hastie, T. Regularization and Variable Selection via the Elastic Net. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **2005**, *67*, 301–320. [CrossRef]
30. Rivera-Pinto, J.; Egozcue, J.J.; Pawlowsky-Glahn, V.; Paredes, R.; Noguera-Julian, M.; Calle, M.L. Balances: A New Perspective for Microbiome Analysis. *mSystems* **2018**, *3*, e00053-18. [CrossRef]
31. Iammarino, M.; Di Taranto, P. Development and Validation of an Ion Chromatography Method for the Simultaneous Determination of Seven Food Additives in Cheeses. *J. Anal. Sci. Methods Instrum.* **2013**, *3*, 30–37. [CrossRef]
32. Association of Official Analytical Chemists. *Official Methods of Analysys of AOAC International*, 16th ed.; Association of Official Analytical Chemists: Rockville, MD, USA, 1995; Volume 2.

33. Bergamaschi, M.; Bittante, G. From Milk to Cheese: Evolution of Flavor Fingerprint of Milk, Cream, Curd, Whey, Ricotta, Scotta, and Ripened Cheese Obtained during Summer Alpine Pasture. *J. Dairy Sci.* **2018**, *101*, 3918–3934. [CrossRef]
34. Manuelian, C.L.; Pozza, M.; Franzoi, M.; Righi, F.; Schmutz, U.; De Marchi, M. Comparison of Organic and Conventional Italian Cheeses Chemical Composition from Parallel Production. *J. Dairy Sci.* **2023**, *106*, 6646–6654. [CrossRef]
35. Blaya, J.; Barzideh, Z.; LaPointe, G. Symposium Review: Interaction of Starter Cultures and Nonstarter Lactic Acid Bacteria in the Cheese Environment. *J. Dairy Sci.* **2018**, *101*, 3611–3629. [CrossRef]
36. Johnson, J.S.; Spakowicz, D.J.; Hong, B.-Y.; Petersen, L.M.; Demkowicz, P.; Chen, L.; Leopold, S.R.; Hanson, B.M.; Agresta, H.O.; Gerstein, M.; et al. Evaluation of 16S rRNA Gene Sequencing for Species and Strain-Level Microbiome Analysis. *Nat. Commun.* **2019**, *10*, 5029. [CrossRef]
37. Abellan-Schneyder, I.; Matchado, M.S.; Reitmeier, S.; Sommer, A.; Sewald, Z.; Baumbach, J.; List, M.; Neuhaus, K. Primer, Pipelines, Parameters: Issues in 16S rRNA Gene Sequencing. *mSphere* **2021**, *6*, e01202-20. [CrossRef]
38. Wilkinson, M.G.; LaPointe, G. Invited Review: Starter Lactic Acid Bacteria Survival in Cheese: New Perspectives on Cheese Microbiology. *J. Dairy Sci.* **2020**, *103*, 10963–10985. [CrossRef] [PubMed]
39. Baruzzi, F.; Lagonigro, R.; Quintieri, L.; Morea, M.; Caputo, L. Occurrence of Non-Lactic Acid Bacteria Populations Involved in Protein Hydrolysis of Cold-Stored High Moisture Mozzarella Cheese. *Food Microbiol.* **2012**, *30*, 37–44. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.