

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Comparison Between Full-Length 16S rRNA Metabarcoding and Whole Metagenome Sequencing Suggests the Use of Either Is Suitable for Large-Scale Microbiome Studies

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1869264> since 2024-03-26T11:12:06Z

Published version:

DOI:10.1089/fpd.2022.0027

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

1 Comparison between full-length 16S metabarcoding and whole metagenome sequencing suggests the
2 use of either is suitable for large-scale microbiome studies.

3 Selene Rubiola^{1†}, Guerrino Macori^{2†}, Tiziana Civera¹, Séamus Fanning², Molly Mitchell² and Francesco
4 Chiesa^{1*}

5 ¹Department of Veterinary Sciences, University of Turin, Grugliasco, TO, Italy

6 ² University College Dublin-Centre for Food Safety School of Public Health, Physiotherapy & Sports
7 Science, Dublin, Ireland

8 [†] Selene Rubiola and Guerrino Macori contributed equally to this work.

9 *Corresponding author

10 **Running title**

11 Milk filters microbiome comparison by FL-16S and WMS

12 **Keywords**

13 full-length 16S rRNA; whole metagenome sequencing; bulk tank milk filters; microbiome; shotgun
14 metagenomic sequencing; metabarcoding

15

16

17

18

19

20

21

22

23 Abstract

24 Since the number of studies of the microbial communities related to food and food-associated
25 matrices almost completely reliant on Next Generation Sequencing techniques is rising, evaluations
26 of these high-throughput methods are critical. Currently, the two most used sequencing methods to
27 profile the microbiota of complex samples, including food and food-related matrices, are the 16S
28 rRNA metabarcoding and the whole metagenome sequencing, both of which are powerful tools for
29 the monitoring of food-borne pathogens and the investigation of the microbiome. Herein, the
30 microbial profiles of 20 bulk tank milk filters from different dairy farms were investigated using both
31 the full-length 16S rRNA metabarcoding, a third-generation sequencing method whose application in
32 food and food-related matrices is yet in its infancy, and the whole metagenome sequencing, in order
33 to evaluate the correlation and the reliability of these two methods to explore the microbiome of
34 food-related matrices. Metabarcoding and metagenomic data were generated on a MinION platform
35 (Oxford Nanopore Technologies, UK) and on a Illumina NovaSeq 6000 platform, respectively. Our
36 findings support the greater resolution of whole metagenome sequencing in terms of both increased
37 detection of bacterial taxa and enhanced detection of diversity; on the other hand, full-length 16S
38 rRNA metabarcoding has proven to be a promising, less expensive and more practical tool to profile
39 most abundant taxa. The significant correlation of the two technologies both in terms of taxa diversity
40 and richness, together with the similar profiles defined for both highly abundant taxa and core
41 microbiomes, including *Acinetobacter*, *Bacillus* and *Escherichia* genera, highlights the possible
42 application of both methods for different purposes.

43 The present study allowed the first comparison of full-length 16S rRNA sequencing and whole
44 metagenome sequencing to investigate the microbial composition of a food-related matrix, pointing
45 out the advantageous use of full-length 16S rRNA to identify dominant microorganisms and the

46 superior power of whole metagenome sequencing for the taxonomic detection of low abundant
47 microorganisms and to perform functional analysis of the microbial communities.

48

49 **1. Introduction**

50 Recent developments in next-generation sequencing (NGS) technologies, together with the
51 reduction in costs and the rise in efficiency, have led to an increase in the number of metabarcoding
52 and metagenomic investigations in different matrices and niches. Two main strategies can be used
53 for the analysis of microbial communities with NGS techniques: whole metagenome sequencing
54 (WMS), also referred to as shotgun metagenomic sequencing, and high-throughput 16S rRNA
55 metabarcoding. Several studies reported on the bovine milk microbiota arising from its association
56 with the quality and safety of dairy products (F. Addis et al., 2016; Rubiola et al., 2020) and more
57 often than not 16S rRNA metabarcoding was applied. The 16S rRNA gene is around 1,600 bp and
58 includes nine hypervariable loci (denoted V1-V9) (Bukin et al., 2019). The 16S rRNA metabarcoding
59 relies on a combination of amplification followed by sequencing of the 16S rRNA gene variable
60 regions, thereby allowing the taxonomic classification and determination of the relative abundance
61 of the bacterial component within a sample. This targeted approach is considered a robust and well-
62 characterized method and has some advantages over shotgun metagenomic sequencing; indeed, it
63 is less expensive than WMS and it does not require the same level of sequencing depth to obtain a
64 proper characterization of the microbiota. Besides, as it is based on a targeted amplification, this
65 technique is not affected by the presence of host (bovine) DNA which characterises milk and dairy
66 products, and data analysis does not require intensive computational power; a wide range of
67 commonly used bioinformatics tools and pipelines for taxonomy and functional analysis are available
68 to facilitate reproducible and modular analysis of 16S rRNA sequencing data in free software
69 platforms, such as QIIME2 (Bolyen et al., 2019) and Mothur (Schloss et al., 2009). Nonetheless, some

70 limitations of this approach are recognised including, 16S rRNA metabarcoding does not provide
71 functional information about the genes encoded by those microbial communities being investigated
72 (Biegert et al., 2021) and it has a low taxonomic resolution, usually limited to family or genus level.
73 Additionally, different reference databases (e.g. GreenGenes (DeSantis et al., 2006), SILVA (Quast et
74 al., 2013), the Ribosomal Database Project (Cole et al., 2014)) can influence the sample taxonomy
75 outcomes of the 16S rRNA metabarcoding (Abellan-Schneyder et al., 2021), which is furthermore
76 affected by a loss of diversity due to PCR bias (F. Addis et al., 2016). Indeed, different 16S rRNA
77 hypervariable regions exhibit differences in their ability to resolve taxa, and the choice of primer
78 designs used is crucial, as the amplification of some regions has been shown to exhibit a bias resulting
79 in over- or under-representation of specific taxa (Laudadio et al., 2018). Among commonly targeted
80 16S rRNA loci, the V3 – V4 and V4 – V5 are the most widely used and their different outcomes in
81 terms of bacterial taxa distribution and alpha diversity have been recognised in different matrices,
82 including biological and environmental samples (Cuccato et al., 2021; Rintala et al., 2017; Soriano-
83 Lerma et al., 2020), as well food matrices, dairy products and fermented foods (Choi et al., 2020;
84 Ferrocino et al., 2017; Liu et al., 2019; Macori and Cotter, 2018). As the short length of the targeted
85 16S rRNA loci represents one of the limitations for taxa identification below the family level, in recent
86 years third-generation sequencing technologies facilitating long-read sequencing has been
87 developed, enabling full-length 16S (FL-16S) gene sequencing (Catozzi et al., 2020); although
88 platforms supporting these techniques, including Pacific Biosciences (PacBio) sequencers and Oxford
89 Nanopore Technologies (ONT) devices, generate read data with lower nucleotide accuracy when
90 compared to the Illumina platforms, reading the FL-16S gene sequence can have better classification
91 resolution (Jeong et al., 2021), as confirmed in recent studies applying this sequencing technique on
92 mock communities and complex matrices such as wastewater samples (Numberger et al., 2019),
93 human faeces (Leggett et al., 2017; Matsuo et al., 2021) and water buffalo milk (Catozzi et al., 2020).

94 In contrast, shotgun metagenomic sequencing confers several advantages over 16S rRNA
95 metabarcoding. First and foremost this strategy can provide functional information about the
96 investigated microorganisms; further, it provides an improved profile of the diversity of the sample
97 and can reach taxa resolution at the species level (Biegert et al., 2021). In this case, whole
98 metagenomic DNA is first extracted, fragmented and then sequenced, independent of the
99 amplification of targeted genes (F. Addis et al., 2016). Thus, a large amount of data is generated to
100 be interrogated for features, including the taxonomic profile of the microbial community, its
101 metabolic pathways and functions. Despite these advantages, some limitations are also recognised,
102 including the computational power required, tools and expertise necessary to properly analyse the
103 data generated; partial sequencing of those less represented microorganisms, whilst background
104 host DNA can be present in significant amounts, especially in host-derived samples including milk and
105 dairy products, requiring the use of different molecular and bioinformatic tools to mask these
106 features, such as pre-extraction methods applying commercially available kits or chemicals to lyse
107 mammals cells, and post-extraction methods enriching microbial DNA by selectively binding and
108 removing CpG-methylated host DNA (Rubiola et al., 2020; Yap et al., 2020). Finally, the shotgun
109 metagenomic sequencing technique is usually more expensive when compared to 16S rRNA
110 metabarcoding and requires a higher coverage (Catozzi et al., 2020). Comparison between WMS and
111 short-read 16S rRNA metabarcoding has been recently explored in different matrices, especially soil
112 and stool samples targeted to investigate the gut microbiome (Brumfield et al., 2020; Durazzi et al.,
113 2021; Jovel et al., 2016; Laudadio et al., 2018; Shah et al., 2010; Tessler et al., 2017); indeed, the
114 extent to which these two sequencing technologies correlate with each other is a crucial assumption,
115 which should be investigated in depth. However, food and food-related matrices have been poorly
116 investigated using both these sequencing techniques; further, the comparison between WMS and
117 FL-16S sequencing is still unexplored. In this context, several studies have suggested the use of milk

118 filters as useful tools to investigate the microbiome of bulk tank milk and to identify the presence of
119 foodborne pathogens (Murphy et al., 2005; Sonnier et al., 2018)

120 To fill the aforementioned knowledge gap, in the present study milk filters sampled in the context of
121 a previous work aiming to evaluate the milk production environment resistome were reanalysed
122 using both the FL-16S rRNA metabarcoding and WMS in order to compare the microbial community
123 profiles and evaluate the reliability of these two methods to explore the microbial communities of
124 food-related matrices.

125

126

127 **2. Materials and methods**

128 **2.1 Farms selection, samples collection and DNA extraction**

129 The samples were collected in May 2020 from the bulk tank of 10 dairy farms located in Piedmont,
130 North-West Italy, with the support of ARAP (Associazione Regionale Allevatori del Piemonte). The
131 sampling procedure included the use of disposable in-line milk filters that were taken from the bulk
132 tank of each farm under aseptic conditions, then inserted in sterile plastic sampling bags (Whirl-Pack,
133 NASCO) and transported in controlled temperature to the Laboratory of Food Inspection -
134 Department of Veterinary Science, University of Turin - where DNA extraction was performed
135 immediately. The sampling was repeated in May 2021, for a total of 20 milk filters.

136 Upon arrival at the laboratory, 10 g of each milk filter were added to 90 ml of sterile buffered
137 saline solution (Ringer's solution, Oxoid, Basingstoke, UK) in a sterile stomacher bag and
138 homogenized for 2 min at 230 rpm in a stomacher (Seward Stomacher Blender 400, London, UK).

139 Total DNA was then extracted from filter homogenates using the DNeasy Blood and Tissue Kit

140 (QIAGEN, Hilden, Germany), with minor adjustments. Samples were centrifuged for 10 min at $100 \times g$
141 to pellet and discard eukaryotic cells; milk serum was then centrifuged at $13,000 \times g$ for 15 min at
142 4°C to pellet prokaryotic cells and pellets recovered resuspended in phosphate-buffered saline [PBS]
143 (Oxoid Basingstoke, UK). Isolation of genomic DNA was then performed following the manufacturer's
144 protocol, including the recommended modification for Gram-positive bacteria (Schwenker et al.,
145 2022); DNA was eluted in $50 \mu\text{l}$ 10 mM Tris-HCl buffer (pH 8.5) and frozen at -20°C until analyzed.
146 Template DNA of each sample was quantified using a Qubit 2.0 Fluorometer (Life Technologies,
147 Carlsbad, CA, USA) with the Qubit double-stranded DNA (dsDNA) high-sensitivity assay kit. DNA
148 integrity and purity were verified by conventional 2% agarose gel electrophoresis and also using a
149 NanoDrop spectrophotometer (ThermoFisher Scientific, Belgium). Samples meeting quality criteria
150 were submitted for FL-16S rRNA metabarcoding and WMS.

151 **2.2 DNA sequencing**

152 Purified DNA was submitted to both FL-16S gene sequencing and WMS. Library preparation for FL-
153 16S was carried out starting from 10 ng of purified DNA from each sample using the 16S Barcoding
154 Kit 1–12 (SQK-RAB204, ONT, UK), following the manufacturer's instruction which includes the
155 generation of FL-16S rRNA genes amplicons using primers 27F (5'-AGAGTTTGATCMTGGCTCAG-3')
156 and 1492R (5'-TACGGYTACCTTGTTACGACTT-3') starting with $10 \mu\text{l}$ input DNA (10 ng), $1 \mu\text{l}$ 16S
157 Barcode, at $10 \mu\text{M}$, $25 \mu\text{l}$ LongAmp Taq 2X master mix (NEB, UK). The amplification was conducted
158 using the following cycling conditions: initial denaturation 1 min at 95°C (1 cycle); denaturation 20
159 secs at 95°C (25 cycles); annealing 30 secs at 55°C (25 cycles); extension 2 mins at 65°C (25 cycles);
160 final extension 5 mins at 65°C (1 cycle). The samples were processed following the manufacturer's
161 instruction with no modifications. Pooled libraries were then sequenced on a MinION platform (ONT,
162 UK) using Flongle (FLO-FLG001) flow cells (ONT, UK) for 24 h.

163 WMS DNA library preparation was carried out according to the NEBNext Ultra II DNA Library Prep Kit
164 for Illumina (New England Biolabs, Ipswich, MA); four PCR cycles were used to amplify the library.
165 Libraries quality and fragment lengths were determined using the Agilent Bioanalyzer 2100 and the
166 High-Sensitivity DNA kit (Agilent Technologies, Santa Clara, CA, USA).
167 The samples were sequenced on the Illumina NovaSeq 6000 platform using an S2 flow cell (Illumina,
168 San Diego, USA) with a 150-cycles paired-end (PE) chemistry, generating 50 million PE reads for each
169 sample.

170

171 2.3 Bioinformatic and statistical analyses

172 FL-16S base-calling was performed using Guppy (version 5.0.15) and Flye (version 2.9) was used as
173 *de novo* assembler.

174 The Fastp tool (Chen et al., 2018) was used to remove reads shorter than 1,000 bp and those reads
175 retained thereafter filtered on a minimum average read quality score of 9, according to the
176 recommendations from Nygaard et al (Nygaard et al., 2020). Processed sequencing data quality was
177 assessed with MultiQC v1.11 (Ewels et al., 2016). Taxonomic classification was performed using
178 Kraken2 v2.1.2 (Wood et al., 2019) and Bracken v2.5.0 (Lu et al., 2017) (threshold=10) with the NCBI
179 NT database.

180 Raw reads generated by WMS were quality assessed using FastQC v.0.11.9
181 (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>) and MultiQC v1.11 with default
182 parameters. Raw reads were quality-trimmed using Trimmomatic version 0.39 (leading, 3; trailing, 3;
183 slidingwindow, 4:20; minlen, 36), removing low-quality regions, adaptor sequences and sequencing
184 primers. After the quality filtering step, clean reads were aligned using Bowtie2 v.2.4.4 against the
185 *Bos taurus* ARS-UCD1.2 bovine reference genome (NCBI Genome ID: 82), to remove host DNA

186 sequences. Taxonomic classification of host-filtered reads was carried out using Kraken2 (Wood et
187 al., 2019); the package Bracken (Lu et al., 2017) was then used on Kraken reports to re-estimate
188 species abundance (threshold=10). Microbial taxonomic assignments of both amplicon and shotgun
189 metagenomic sequence data were performed using the NCBI NT database.

190 Relative abundance tables for all samples were merged and imported into MicrobiomeAnalyst (Chong
191 et al., 2020) for statistical and diversity analysis. Data from both WMS and FL-16S were analysed using
192 alpha diversity metrics to assess the divergence of the microbial communities within each filter
193 sample. Shannon Diversity (Mouillot and Leprêtre, 1999) and Simpson Diversity indexes were
194 calculated from the observed operational taxonomic unit (OTU) counts for FL-16S and WMS data
195 after centered log-ratio (clr) normalization. Rarefaction curves were generated to assess the
196 saturation of samples analyzed using the WMS and FL-16S sequencing.

197 In order to perform a comparative statistical analysis of FL-16S and WMS data, each sample value
198 from each dataset was paired with its corresponding value for the same sample in the other dataset.
199 The pairwise Spearman's correlation test was applied to investigate the amount of agreement
200 between the two datasets, including alpha diversity measures, richness (observed OTUs) and indexes
201 of Shannon's and Simpson's diversity. The composition of the core microbiome was assessed at genus
202 and family levels for FL-16S and WMS datasets using 50% and 1% cut-off values for occurrence and
203 relative abundance of the OTUs, respectively (Neu et al., 2021); the abundance of shared OTUs was
204 visualized using heatmaps and Venn diagrams.

205

206

207 **3. Results**

208 Shotgun metagenomic sequencing yielded 1.06 billion reads, with an average of 53.1 million reads
209 per sample (range 44.8-76.8 million); out of 1.06 billion reads, a total of 6.2 million were identified at
210 the bacterial and archaeal phyla level. FL-16S sequencing resulted in 166.928 reads, with an average
211 of 8.346 reads per sample (range 2.759-30.168).

212 The number of observed OTUs gained from WMS was found to be higher in comparison with FL-16S
213 sequencing datasets at each taxonomic level. In particular, at the family level, the number of families
214 detected by WMS was significantly greater compared to the number of families detected by FL-16S
215 sequencing (p-value <0.001, T-test), ranging from 7 to 37 OTUs per sample for FL-16S (mean 24.9)
216 and from 227 to 301 OTUs per sample for WMS (mean 278,9); similarly, at the genus level, the
217 number of annotated genera observed by WMS was significantly greater compared to the number
218 of genera detected by FL-16S sequencing (p-value <0.001, T-test), ranging from 12 to 64 OTUs per
219 sample for 16S sequencing (mean 41,2) and from 614 to 850 OTUs per sample for WMS (mean 779.7).

220 The number of genera and families identified in each sample by the two sequencing techniques are
221 reported as boxplots in **Figure 1**. Alpha diversity patterns were calculated at the family and genus
222 level using Shannon's and Simpson's indexes (**Figure 1**). At the family level, across samples analysed
223 by WMS, both Shannon index alpha diversity and Simpson index alpha diversity were significantly
224 greater than alpha diversity values of samples analysed by FL-16S sequencing (difference between
225 means = $1,138 \pm 0,1926$, 95% CI 0,7483-1,528, p-value <0.0001; difference between means = $0,1324$
226 $\pm 0,04367$, 95% CI 0,04396 to 0,2208, p-value <0.005, T-test). Similarly, at the genus level, across
227 samples analysed by WMS, both Shannon index alpha diversity and Simpson index alpha diversity
228 were significantly greater than alpha diversity values of samples analysed by FL-16S sequencing
229 (difference between means = $1,238 \pm 0,2132$, 95% CI 0,8069-1,670, p-value <0.0001; difference
230 between means = $0,07861 \pm 0,03808$, 95% CI 0,001535-0,1557, p-value <0.05, T-test). Thus, both the
231 observed Shannon index alpha diversity values and the Simpson index alpha diversity values were

232 greater for samples analysed by WMS compared to samples analysed by FL-16S sequencing at each
233 taxonomic level. Rarefaction curves showed that almost all samples reached the asymptote or started
234 to plateau despite the different technique applied (Supplementary File S1).

235 The top 10 most abundant genera profiled across the 20 samples by FL-16S sequencing and WMS
236 corresponded to *Acinetobacter*, *Lactococcus*, *Escherichia*, *Streptococcus*, *Staphylococcus*, *Bacillus*,
237 *Corynebacterium*, *Pseudomonas*, *Lactobacillus* and *Clostridium* (**Figure 2**). Most of the highly
238 abundant genera detected per farm were detected by both FL-16S and WMS; however, different
239 relative abundances were observed, mainly due to the overall lower number of OTUs annotated by
240 full-length sequencing, consistently with the results of richness and diversity indexes.

241 All the diversity and richness measures, including observed OTUs, Shannon and Simpson diversity
242 indexes, were tightly correlated between FL-16S sequencing and WMS, at both the family (Observed
243 OTUs Spearman R = 0.6, p-value = 0.005; Shannon Spearman R = 0.75, p-value = 0.0002; Simpson
244 Spearman R = 0.6, p-value = 0.006) and genus level (Observed OTUs Spearman R = 0.68, p-value =
245 0.0008 ; Shannon Spearman R = 0.66, p-value = 0.001; Simpson Spearman R = 0.52, p-value = 0.01)
246 (**Figure 3**).

247 The presence of a core microbiome common to the sampled milk filters was confirmed in both
248 samples analysed by FL-16S sequencing and WMS. Out of 361 families detected across all samples,
249 thirteen families were found in the core microbiome associated with milk filters analysed by FL-16S
250 sequencing and thirteen families were found in the core microbiome associated with milk filters
251 analysed by WMS; four of them were shared between the two core microbiomes, namely
252 *Moraxellaceae*, *Enterobacteriaceae*, *Bacillaceae* and *Streptococcaceae*. Consistently, of the 1,078
253 genera identified across all samples, thirteen were found in the core microbiome associated with milk
254 filters analysed by FL-16S sequencing, namely *Acinetobacter*, *Escherichia*, *Staphylococcus*,
255 *Lactococcus*, *Bacillus*, *Streptococcus*, *Aerococcus*, *Clostridioides*, *Lactobacillus*, *Clostridium*,

256 *Oscillibacter* and *Paeniclostridium*, eight were found in the core microbiome associated with milk
257 filters analysed by WMS, namely *Acinetobacter*, *Corynebacterium*, *Bifidobacterium*, *Actinoalloteichus*,
258 *Pseudomonas*, *Bradyrhizobium*, *Escherichia* and *Bacillus*, and three were shared between the two
259 core microbiomes, that is *Acinetobacter*, *Escherichia* and *Bacillus* (**Figure 4**).

260 **4. Discussion**

261 The two most used sequencing methods to profile the microbiota of complex samples, including food
262 and food-related matrices, are the 16S metabarcoding and shotgun metagenomic sequencing. Both
263 these NGS techniques offer different advantages over culture-based methods; the 16S
264 metabarcoding has been used more frequently mainly due to its low cost, low computational power
265 requirements and standardized analysis methods, WMS is becoming more attractive for in-depth
266 studies of microbial populations due to the large amount of information provided by this untargeted
267 sequencing technique, which facilitates study of the functional profile of complex microbiomes.
268 Recently, comparisons between high-throughput 16S rRNA sequencing and WMS have been
269 performed in selected matrices, including gut, soil and water samples (Brumfield et al., 2020; Ranjan
270 et al., 2016; Tessler et al., 2017). However food and food-related matrices are poorly investigated for
271 several reasons including, the large amount of host DNA that characterizes these samples might
272 greatly interfere with different sequencing techniques; those comparative studies performed have
273 been based on selected hypervariable loci within the 16S rRNA gene, while the FL-16S sequencing
274 has proved to allow a less biased study of different microbial ecosystems (Catozzi et al., 2020). This
275 study reports on the comparison of FL-16S and WMS to investigate the microbial population of bulk
276 tank milk filters, both of which are powerful tools for the monitoring of food-borne pathogens and
277 the investigation of the microbiome of bulk tank milk.

278 Although 16S metabarcoding is a promising, less expensive and more practical tool to investigate the
279 microbiome when compared to WMS, in the present study it allowed the identification of only most

280 abundant microorganisms in the biological samples investigated. Consistently, some previous studies
281 highlighted a significant amount of agreement between 16S metabarcoding and WMS methods at a
282 higher order of taxa, with a high degree of correlation found between 16S and WMS (Biegert et al.,
283 2021; Vogtmann et al., 2016). Our findings support the greater resolution of WMS in terms of both
284 increased detection of bacterial taxa and enhanced detection of diversity; the superior richness in
285 the profiles of microbes obtained and their diversity must also be weighted with the already known
286 advantages related to the possibility of investigating the function of predicted genes. Our results are
287 in accordance with studies analysing human faecal (Ranjan et al., 2016) and soil (Brumfield et al.,
288 2020) microbiomes, which, despite investigating targeted hypervariable regions of the 16S rRNA
289 gene, revealed a greater diversity of microorganisms through the use of WMS. In this context, it must
290 be stated that the actual composition of the microbiome of analysed milk filter samples was
291 unknown; thereby, our approach, while enabling us to draw some conclusions on sensitivity, does
292 not enable the evaluation of the specificity of each sequencing technique. This issue goes beyond the
293 aims of the present study and can be addressed using simulated NGS data.

294 The present investigation of the milk filters' core microbiome through the application of both
295 techniques has allowed the definition of a group of bacterial genera common to all the selected
296 samples; in particular, while different sequencing methods defined different core microbiomes,
297 *Acinetobacter*, *Bacillus* and *Escherichia* genera were shared between the FL-16S and the WMS cores.
298 Although the microbiota profiles of distinct bulk tank milk filters were different, the presence of a
299 well-defined core microbiome, characterized by both the sequencing technique applied, highlights
300 the possibility to integrate multiple techniques to confirm the consistency of the achieved outcomes.
301 The overall high occurrence and relative abundance of members of the *Moraxellaceae*,
302 *Enterococcaceae*, *Bacillaceae* and *Streptococcaceae* families in milk filters are consistent with the
303 profiled core microbiome of recent studies focusing on raw bovine milk collected in tankers (Kable et

304 al., 2016; McHugh et al., 2020), thereby highlighting the deep correlation of microbial communities
305 of bulk tank milk and microbial communities of in-line milk filters; most of the taxa belonging to the
306 core microbiomes profiled by FL-16S and WMS are known to be associated with dairy processing
307 environments.

308 This study set out the use of different high-throughput molecular methods to provide an in-depth
309 description of the microbiota of a food processing environment using milk filters as promising tools;
310 however certain limitations must be considered. This research was performed using a small number
311 of samples, although this was sufficient to identify significant differences between the compared
312 methods. Furthermore, a comparison including the most commonly used hypervariable regions of
313 the 16S rRNA gene (e.g. the variable V3 and V4 regions), together with the FL-16S and the WMS
314 approach could provide further data to choose the more suitable method for different scientific
315 purposes. To our knowledge, this is the first study aiming to compare the use of FL-16S and WMS to
316 investigate the microbial composition of a food-related matrix. Although, as anticipated, the
317 resolution power of WMS has proved to be greater than that provided by 16S sequencing, the
318 significant correlation of the two technologies both in terms of taxa diversity and richness, together
319 with the similar profiles defined for both highly abundant taxa and core microbiomes, highlights the
320 possible application of both methods for different purposes. Thus, our findings suggest that the use
321 of FL-16S to perform large-scale microbiome studies can provide rapid and valuable data at a fraction
322 of the cost of WMS, which, on the other hand, is an incomparable tool to perform in-depth studies
323 of the microbiome, including low abundance taxa and functional profiles.

324

325

326

327

328

329

330

331 **Acknowledgments**

332 This study was supported by Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) under
333 the programme "Dipartimenti di Eccellenza ex L.232/2016" to the Department of Veterinary Science,
334 University of Turin."

335 **Authorship contribution**

336 **Selene Rubiola:** Writing – original draft; formal analysis; investigation; conceptualization. **Guerrino**
337 **Macori:** Writing – review & editing; resources; validation; conceptualization. **Tiziana Civera:** Writing –
338 review & editing; funding acquisition. **Séamus Fanning:** Writing – review & editing; resources. **Molly**
339 **Mitchell:** Writing – review & editing; formal analysis. **Francesco Chiesa:** Writing – review & editing;
340 visualization; supervision; conceptualization.

341 **Conflict of Interest**

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

344 **Funding**

345 This work was supported by European Regional Development Funds (FESR 2014-2020 –
346 D24I19000980002)–TECH4MILK

347 **References**

348 Abellan-Schneyder I, Matchado MS, Reitmeier S, et al. Primer, Pipelines, Parameters: Issues in 16S
349 RRNA Gene Sequencing. *mSphere* 2021;6(1):e01202-20; doi: 10.1128/mSphere.01202-20.

350 Biegert G, El Alam MB, Karpinets T, et al. Diversity and Composition of Gut Microbiome of Cervical
351 Cancer Patients: Do Results of 16S RRNA Sequencing and Whole Genome Sequencing
352 Approaches Align? *Journal of Microbiological Methods* 2021;185:106213; doi:
353 10.1016/j.mimet.2021.106213.

354 Bolyen E, Rideout JR, Dillon MR, et al. Reproducible, Interactive, Scalable and Extensible
355 Microbiome Data Science Using QIIME 2. *Nat Biotechnol* 2019;37(8):852–857; doi:
356 10.1038/s41587-019-0209-9.

357 Brumfield KD, Huq A, Colwell RR, et al. Microbial Resolution of Whole Genome Shotgun and 16S
358 Amplicon Metagenomic Sequencing Using Publicly Available NEON Data. *PLOS ONE*
359 2020;15(2):e0228899; doi: 10.1371/journal.pone.0228899.

360 Bukin YS, Galachyants YP, Morozov IV, et al. The Effect of 16S RRNA Region Choice on Bacterial
361 Community Metabarcoding Results. *Sci Data* 2019;6(1):190007; doi: 10.1038/sdata.2019.7.

362 Catozzi C, Ceciliani F, Lecchi C, et al. Short Communication: Milk Microbiota Profiling on Water
363 Buffalo with Full-Length 16S RRNA Using Nanopore Sequencing. *Journal of Dairy Science*
364 2020;103(3):2693–2700; doi: 10.3168/jds.2019-17359.

365 Chen S, Zhou Y, Chen Y, et al. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. 2018;274100;
366 doi: 10.1101/274100.

367 Choi J, In Lee S, Rackerby B, et al. Assessment of Overall Microbial Community Shift during Cheddar
368 Cheese Production from Raw Milk to Aging. *Appl Microbiol Biotechnol* 2020;104(14):6249–6260;
369 doi: 10.1007/s00253-020-10651-7.

370 Chong J, Liu P, Zhou G, et al. Using MicrobiomeAnalyst for Comprehensive Statistical, Functional,
371 and Meta-Analysis of Microbiome Data. *Nat Protoc* 2020;15(3):799–821; doi: 10.1038/s41596-
372 019-0264-1.

373 Cole JR, Wang Q, Fish JA, et al. Ribosomal Database Project: Data and Tools for High Throughput
374 rRNA Analysis. *Nucleic Acids Research* 2014;42(D1):D633–D642; doi: 10.1093/nar/gkt1244.

375 Cuccato M, Rubiola S, Giannuzzi D, et al. 16S rRNA Sequencing Analysis of the Gut Microbiota in
376 Broiler Chickens Prophylactically Administered with Antimicrobial Agents. *Antibiotics*
377 2021;10(2):146; doi: 10.3390/antibiotics10020146.

378 DeSantis TZ, Hugenholtz P, Larsen N, et al. Greengenes, a Chimera-Checked 16S rRNA Gene
379 Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*
380 2006;72(7):5069–5072; doi: 10.1128/AEM.03006-05.

381 Durazzi F, Sala C, Castellani G, et al. Comparison between 16S rRNA and Shotgun Sequencing Data
382 for the Taxonomic Characterization of the Gut Microbiota. *Sci Rep* 2021;11(1):3030; doi:
383 10.1038/s41598-021-82726-y.

384 Ewels P, Magnusson M, Lundin S, et al. MultiQC: Summarize Analysis Results for Multiple Tools and
385 Samples in a Single Report. *Bioinformatics* 2016;32(19):3047–3048; doi:
386 10.1093/bioinformatics/btw354.

387 F. Addis M, Tanca A, Uzzau S, et al. The Bovine Milk Microbiota: Insights and Perspectives from -
388 Omics Studies. *Molecular BioSystems* 2016;12(8):2359–2372; doi: 10.1039/C6MB00217J.

389 Ferrocino I, Bellio A, Romano A, et al. RNA-Based Amplicon Sequencing Reveals Microbiota
390 Development during Ripening of Artisanal versus Industrial Lard d’Arnad. *Applied and*
391 *Environmental Microbiology* 2017; doi: 10.1128/AEM.00983-17.

392 Jeong J, Yun K, Mun S, et al. The Effect of Taxonomic Classification by Full-Length 16S RRNA
393 Sequencing with a Synthetic Long-Read Technology. *Sci Rep* 2021;11(1):1727; doi:
394 10.1038/s41598-020-80826-9.

395 Jovel J, Patterson J, Wang W, et al. Characterization of the Gut Microbiome Using 16S or Shotgun
396 Metagenomics. *Frontiers in Microbiology* 2016;7.

397 Kable ME, Srisengfa Y, Laird M, et al. The Core and Seasonal Microbiota of Raw Bovine Milk in
398 Tanker Trucks and the Impact of Transfer to a Milk Processing Facility. *mBio* 2016; doi:
399 10.1128/mBio.00836-16.

400 Laudadio I, Fulci V, Palone F, et al. Quantitative Assessment of Shotgun Metagenomics and 16S
401 rDNA Amplicon Sequencing in the Study of Human Gut Microbiome. *OMICS: A Journal of*
402 *Integrative Biology* 2018;22(4):248–254; doi: 10.1089/omi.2018.0013.

403 Leggett RM, Alcon-Giner C, Heavens D, et al. Rapid MinION Metagenomic Profiling of the Preterm
404 Infant Gut Microbiota to Aid in Pathogen Diagnostics. 2017;180406; doi: 10.1101/180406.

405 Liu Z, Li J, Wei B, et al. Bacterial Community and Composition in Jiang-Shui and Suan-Cai Revealed
406 by High-Throughput Sequencing of 16S RRNA. *International Journal of Food Microbiology*
407 2019;306:108271; doi: 10.1016/j.ijfoodmicro.2019.108271.

408 Lu J, Breitwieser FP, Thielen P, et al. Bracken: Estimating Species Abundance in Metagenomics Data.
409 *PeerJ Computer Science* 2017;3:e104; doi: 10.7717/peerj-cs.104.

410 Macori G and Cotter PD. Novel Insights into the Microbiology of Fermented Dairy Foods. *Current*
411 *Opinion in Biotechnology* 2018;49:172–178; doi: 10.1016/j.copbio.2017.09.002.

412 Matsuo Y, Komiya S, Yasumizu Y, et al. Full-Length 16S RRNA Gene Amplicon Analysis of Human Gut
413 Microbiota Using MinION™ Nanopore Sequencing Confers Species-Level Resolution. *BMC*
414 *Microbiol* 2021;21(1):35; doi: 10.1186/s12866-021-02094-5.

415 McHugh AJ, Feehily C, Fenelon MA, et al. Tracking the Dairy Microbiota from Farm Bulk Tank to
416 Skimmed Milk Powder. *mSystems* 2020;5(2):e00226-20; doi: 10.1128/mSystems.00226-20.

417 Mouillot D and Leprêtre A. A Comparison of Species Diversity Estimators. *Res Popul Ecol*
418 1999;41(2):203–215; doi: 10.1007/s101440050024.

419 Murphy BP, Murphy M, Buckley JF, et al. In-Line Milk Filter Analysis: Escherichia Coli O157
420 Surveillance of Milk Production Holdings. *International Journal of Hygiene and Environmental*
421 *Health* 2005;208(5):407–413; doi: 10.1016/j.ijheh.2005.03.001.

422 Neu AT, Allen EE and Roy K. Defining and Quantifying the Core Microbiome: Challenges and
423 Prospects. *PNAS* 2021;118(51); doi: 10.1073/pnas.2104429118.

424 Numberger D, Ganzert L, Zoccarato L, et al. Characterization of Bacterial Communities in
425 Wastewater with Enhanced Taxonomic Resolution by Full-Length 16S rRNA Sequencing. *Sci Rep*
426 2019;9(1):9673; doi: 10.1038/s41598-019-46015-z.

427 Nygaard AB, Tunsjø HS, Meisal R, et al. A Preliminary Study on the Potential of Nanopore MinION
428 and Illumina MiSeq 16S rRNA Gene Sequencing to Characterize Building-Dust Microbiomes. *Sci*
429 *Rep* 2020;10(1):3209; doi: 10.1038/s41598-020-59771-0.

430 Quast C, Pruesse E, Yilmaz P, et al. The SILVA Ribosomal RNA Gene Database Project: Improved Data
431 Processing and Web-Based Tools. *Nucleic Acids Res* 2013;41(Database issue):D590-596; doi:
432 10.1093/nar/gks1219.

433 Ranjan R, Rani A, Metwally A, et al. Analysis of the Microbiome: Advantages of Whole Genome
434 Shotgun versus 16S Amplicon Sequencing. *Biochem Biophys Res Commun* 2016;469(4):967–977;
435 doi: 10.1016/j.bbrc.2015.12.083.

436 Rintala A, Pietilä S, Munukka E, et al. Gut Microbiota Analysis Results Are Highly Dependent on the
437 16S rRNA Gene Target Region, Whereas the Impact of DNA Extraction Is Minor. *J Biomol Tech*
438 2017;28(1):19–30; doi: 10.7171/jbt.17-2801-003.

439 Rubiola S, Chiesa F, Dalmaso A, et al. Detection of Antimicrobial Resistance Genes in the Milk
440 Production Environment: Impact of Host DNA and Sequencing Depth. *Front Microbiol* 2020;11;
441 doi: 10.3389/fmicb.2020.01983.

442 Schloss PD, Westcott SL, Ryabin T, et al. Introducing Mothur: Open-Source, Platform-Independent,
443 Community-Supported Software for Describing and Comparing Microbial Communities. *Applied*
444 *and Environmental Microbiology* 2009; doi: 10.1128/AEM.01541-09.

445 Schwenker JA, Friedrichsen M, Waschina S, et al. Bovine Milk Microbiota: Evaluation of Different
446 DNA Extraction Protocols for Challenging Samples. *MicrobiologyOpen* 2022;11(2):e1275; doi:
447 10.1002/mbo3.1275.

448 Shah N, Tang H, Doak TG, et al. COMPARING BACTERIAL COMMUNITIES INFERRED FROM 16S RRNA
449 GENE SEQUENCING AND SHOTGUN METAGENOMICS. In: *Biocomputing 2011 WORLD*
450 *SCIENTIFIC*; 2010; pp. 165–176; doi: 10.1142/9789814335058_0018.

451 Sonnier JL, Karns JS, Lombard JE, et al. Prevalence of Salmonella Enterica, Listeria Monocytogenes,
452 and Pathogenic Escherichia Coli in Bulk Tank Milk and Milk Filters from US Dairy Operations in
453 the National Animal Health Monitoring System Dairy 2014 Study. *Journal of Dairy Science*
454 2018;101(3):1943–1956; doi: 10.3168/jds.2017-13546.

455 Soriano-Lerma A, Pérez-Carrasco V, Sánchez-Marañón M, et al. Influence of 16S RRNA Target
456 Region on the Outcome of Microbiome Studies in Soil and Saliva Samples. *Sci Rep*
457 2020;10(1):13637; doi: 10.1038/s41598-020-70141-8.

458 Tessler M, Neumann JS, Afshinnekoo E, et al. Large-Scale Differences in Microbial Biodiversity
459 Discovery between 16S Amplicon and Shotgun Sequencing. *Sci Rep* 2017;7(1):6589; doi:
460 10.1038/s41598-017-06665-3.

461 Vogtmann E, Hua X, Zeller G, et al. Colorectal Cancer and the Human Gut Microbiome:
462 Reproducibility with Whole-Genome Shotgun Sequencing. PLOS ONE 2016;11(5):e0155362; doi:
463 10.1371/journal.pone.0155362.

464 Wood DE, Lu J and Langmead B. Improved Metagenomic Analysis with Kraken 2. Genome Biol
465 2019;20(1):257; doi: 10.1186/s13059-019-1891-0.

466 Yap M, Feehily C, Walsh CJ, et al. Evaluation of Methods for the Reduction of Contaminating Host
467 Reads When Performing Shotgun Metagenomic Sequencing of the Milk Microbiome. Sci Rep
468 2020;10(1):21665; doi: 10.1038/s41598-020-78773-6.

469

470

471

472

473

474

475

476

477

478

479

480

481

482

483

484 **Figure legends**

485 **Figure 1.** Boxplots showing the number of OTUs, the Shannon and Simpson alpha-diversity indexes
486 observed at the family and genus level. All indexes showed a statistically significant difference
487 between the mean measures observed in samples analysed by FL-16S and WMS.

488 **Figure 2.** The top-10 most abundant genera identified across the 20 samples analysed after centered-
489 log-ratio normalization; genera with a lower relative abundance are binned into “others” category.
490 Samples are organized by farm and year of sampling.

491 **Figure 3.** Correlation between WMS and FL-16S in terms of diversity at family and genus level. Each
492 data point represents a single sample. Consensus between both sequencing methods in terms of
493 alpha diversity was calculated by Spearman’s correlation. The slope of the correlation is represented
494 by the grey, continuous line, while the 95% confidence interval is represented by the area delimited
495 by the grey dotted lines. The data derived from FL-16S sequencing correlates well with the diversity
496 assessment values derived from WMS for diversity.

497 **Figure 4.** Core heatmaps and Venn diagrams showing bacterial families and genera detected in more
498 than 50% of samples with more than 1% of relative abundance. Four OTUs at both family and genus
499 levels were detected in all samples by FL-16S sequencing and WMS, thereby representing the shared
500 core microbiome.

501