AperTO - Archivio Istituzionale Open Access dell'Università di Torino

A bioinformatics pipeline integrating predictive metagenomics profiling for the analysis of 16S rDNA/rRNA sequencing data originated from foods

This is the author's manuscript

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1680011> since 2018-10-31T15:23:38Z

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)

This is the author's final version of the contribution published as:

Marios Mataragas, Valentina Alessandria, Ilario Ferrocino, Kalliopi Rantsiou and
Luca Cocolin

Food Microbiology, Volume 76, December 2018, Pages 279-286

The publisher's version is available at:

<https://www.ncbi.nlm.nih.gov/pubmed/30166151>

When citing, please refer to the published version.

Link to this full text:

[<https://reader.elsevier.com/reader/sd/pii/S0740002017308900?token=21BAF05CC12C65CDFS2AB7BAE9D813AE74BA2D35A10338AA95909E62468959B96A13FE143C8765D74A7B92D881474161>]

This full text was downloaded from iris-AperTO: <https://iris.unito.it/>

A bioinformatics pipeline integrating predictive metagenomics profiling for the analysis of 16S rDNA/rRNA sequencing data originated from foods

Marios Mataragas^{a*}, Valentina Alessandria^b, Ilario Ferrocino^b, Kalliopi Rantsiou^b and Luca Cocolin^b

^a Hellenic Agricultural Organization “DEMETER”, Institute of Technology of Agricultural Products, Department of Dairy Research, Ethnikis Antistaseos 3, 45221, Ioannina, Greece

^b University of Turin, Department of Agricultural, Forest and Food Sciences, Laboratory of Food Microbiology, Largo P. Braccini 2, 10095, Grugliasco, Turin, Italy

*Corresponding author: mmatster@gmail.com, Hellenic Agricultural Organization “DEMETER”, Institute of Technology of Agricultural Products, Department of Dairy Research, Ethnikis Antistaseos 3, 45221, Ioannina, Greece

Keywords: Amplicon Sequencing; Food Microbiology; Metabolic Network; Meta-omics; Microbial Risk Assessment; Molecular Microbiology; Predictive Functional Profiling

57
58
59 **24 Abstract**
60

61 25 The recent advances in molecular biology, such as the advent of next-generation
62 26 sequencing (NGS) platforms, have paved the way to new exciting tools which rapidly
63 27 transform food microbiology. Nowadays, NGS methods such as 16S rDNA/rRNA
64 28 metagenomics or amplicon sequencing are used for the taxonomic profiling of the food
65 29 microbial communities. Although 16S rDNA/rRNA NGS-based microbial data are not
66 30 suited for the investigation of the functional potential of the identified operational
67 31 taxonomic units as compared to shotgun metagenomics, advances in the bioinformatics
68 32 discipline allow now the performance of such studies. In this paper, a bioinformatics
69 33 workflow is described integrating predictive metagenomics profiling with specific
70 34 application to food microbiology data. Bioinformatics tools pertinent to each sub-module
71 35 of the pipeline are suggested as well. The published 16S rDNA/rRNA amplicon data
72 36 originated from an Italian Grana-type cheese, using an NGS platform, was employed to
73 37 demonstrate the predictive metagenomics profiling approach. The pipeline identified the
74 38 microbial community and the changes that occurred in the microbial profile during
75 39 manufacture of the food product studied (taxonomic profiling). The workflow also
76 40 indicated significant changes in the functional profiling of the community. The tool may
77 41 help to investigate the functional potential, alterations, and interactions of a microbial
78 42 community. The proposed workflow may also find an application in the investigation of
79 43 the ecology of foodborne pathogens encountered in various food products.
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99

100
101
102 44
103
104
105
106
107
108
109
110
111
112

1. Introduction

The objective of this work was to suggest and describe a bioinformatics workflow for the analysis of metagenomic data based on the 16S rDNA/rRNA amplicon sequencing originated from the application of next-generation sequencing (NGS) platforms. The pipeline integrates functional metagenomics, which is an emerging technique with potential industrial interest (Coughlan et al., 2015). Usually, papers dealing with the investigation of microbial ecology in food products using NGS methods end up with the taxonomic profiling of the microbial community after the preprocessing of the obtained 16S rDNA/rRNA data (Alessandria et al., 2016; Delcenserie et al., 2014; Ercolini et al., 2012; Liu et al., 2015; Parlapani and Boziaris, 2016; Parlapani et al., 2013; Parlapani et al., 2015; Polka et al., 2015; Sattin et al., 2016). However, data derived from 16S rDNA/rRNA amplicon sequencing can be exploited to investigate the functional potential of the identified operational taxonomic units (OTUs). Only recently, a few studies have performed functional profiling (Ferrocino et al., 2016; Pothakos et al., 2015; Stellato et al., 2016), but in general, this is not a common practice. The 16S rDNA/rRNA amplicon sequencing is a form of metagenomics and not metatranscriptomics, and therefore, the analysis is known as predictive functional profiling (Langille et al., 2013) or predictive metagenomics profiling (Wood, 2016). Other authors have suggested the integration of functional metagenomics into 16S rDNA/rRNA studies (Coughlan et al., 2015; Keller et al., 2014), but a key difference between those studies and the currently proposed bioinformatics pipeline is the inclusion of an additional step for the prediction of metabolic interactions between the microbial species found in a community (Mendes-Soares et al., 2016), an analysis not previously suggested or performed in food

metagenomes. In addition, the proposed food-focused pipeline involves a selection of tools and their specific sequential use along with the statistical tests, describing a step-wise use of each program and statistical test in each submodule. This will provide a quick and easy reference for the user who would like to use the programs in correct order. The 16S rRNA amplicon data originated from a Grana-type Italian cheese using an NGS platform (Alessandria et al., 2016) were used to demonstrate the predictive metagenomics profiling approach.

2. Bioinformatics workflow

The workflow integrates two main stages: the preprocessing (quality control of the sequences) and quantification (identification of the operational taxonomic units – OTUs, their potential interactions, and functional potential). The latter includes two sub-modules: the taxonomic profiling and the predictive metagenomics profiling (PMP) (Fig. 1). To accomplish the objectives of each step of the pipeline there are available various open-source programs which are free for academic use. The available software for the preprocessing and taxonomic profiling of the amplicon sequencing data are numerous. Table 1 presents the use of a specific program and statistical test in each stage and submodule of the pipeline. Alternative software that can be employed is also proposed, to enhance the step-wise description of the analysis workflow. Therefore, this list is not exhaustive but there are several relevant programs which the interested readers can seek in other excellent reviews regarding the existing software tools for bioinformatics analysis of metagenomic data (De Filippo et al., 2012; Dudhagara et al., 2015; Escobar-Zepeda et al., 2015; Ladoukakis et al., 2014; Oulas et al., 2015; Roumpeka et al., 2017;

91 Scholz et al., 2012; Sharpton, 2014). On the contrary, the number of available tools for
92 PMP of 16S rDNA/rRNA amplicon data is limited (Aßhauer et al., 2015; Iwai et al.,
93 2016; Langille et al., 2013).

95 **3. Case study: taxonomic and functional profiling of the microbial community of a** 96 **hard, slow-ripened cheese**

97 The data used were from the study of Alessandria et al. (2016). The Sequence
98 Read Archive (SRA) website of the National Center for Biotechnology Information
99 (NCBI) was accessed to download all the deposited sequences in FASTA format
100 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/>). Three different batches (D, E, and F) of a
101 Grana-type Italian cheese were used to get food metagenomics data by pyrosequencing
102 (Roche 454 GS Junior platform) of the amplified V1 to V3 region of the 16S rRNA
103 marker gene. The authors collected thirty-nine samples in total ($n = 39$; 13 samples per
104 batch) during manufacture and ripening of the cheese (Whey Starter, WS; Raw Milk,
105 RM; Raw Milk and Whey Starter, MS; Curd after Cutting, CAC; Curd after Heating,
106 CAH; Curd after Pressing, CAP; Curd after Storage Room, CASR; Cheese after Salting,
107 CHAS; Second Ripening Month, CH2RM; Fourth Ripening Month, CH4RM; Sixth
108 Ripening Month, CH6RM; Eighth Ripening Month, CH8RM; Tenth Ripening Month,
109 CH10RM) for pyrosequencing purposes. Preprocessing (stage 1 of the proposed
110 bioinformatics workflow of Fig. 1) of the downloaded sequences had already being
111 performed by Alessandria et al. (2016) with QIIME v1.9.0 (Caporaso et al., 2010), and
112 therefore in this case study only the quantification step (submodule 1 and 2 of the
113 proposed bioinformatics pipeline of Fig. 1) was carried out.

114

115 *3.1. Submodule 1: Taxonomic profiling*

116 Taxonomic profiling was performed using the SILVAngs 1.3 pipeline (Quast et
117 al., 2013). Each downloaded sequence (264826 sequences in total) was aligned using the
118 SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN, revision 21008) (Pruesse et
119 al., 2012) against the SILVA SSU rRNA SEED and quality controlled (Quast et al.,
120 2013). Quality control of the submitted sequences, using the standard settings of the
121 pipeline, rejected 89660 sequences (number of classified sequences equal to 173225 and
122 number of “No Relative” equal to 1941). Afterward, identical reads were identified
123 (dereplication), unique reads were clustered (OTUs), on a per sample basis, and reference
124 read of each OTU was classified. Dereplication and clustering were made using cd-hit-est
125 (version 3.1.2) (Li and Godzik, 2006) running in “accurate mode”, ignoring overhangs,
126 and applying identity criteria of 1.00 and 0.98, respectively. The classification was done
127 by a local nucleotide BLAST search against the non-redundant version of the SILVA
128 SSU Ref dataset (release 128) using blastn (version 2.2.30+) with standard settings
129 (Camacho et al., 2009). Reads without any BLAST hits or reads with weak BLAST hits
130 (Similarity $\leq$ 93%) remained unclassified (“No Relative”). The output of the pipeline,
131 among others, was an OTU table containing the OTU abundances per sample at the genus
132 and species level. The taxonomy at the species level was not possible for all the OTUs.
133 The matrix was filtered further by applying the same filtering criteria with Alessandria et
134 al. (2016), i.e. including only those OTUs with abundance $\geq$ 0.5% in at least two samples.
135 The filtered table was the final output kept for all the subsequent steps.

After removal of the sample identified as outliers, no significant differences were observed between the three D, E and F batches regarding the microbial community profile using the ANOSIM (Analysis of Similarity) statistical test ($P = 0.352$; $P_{D-E} = 0.311$; $P_{D-F} = 0.370$; $P_{E-F} = 0.376$) (Fig. 2) of the Past v3.15 software (Hammer et al., 2001). Fig. 3 displays an overview of the microbial community profile at the genus level during manufacture of the Grana-type cheese using the Community-Analyzer program (Kuntal et al., 2013). The arrows show the presence of an OTU in a particular metagenomic sample. Raw milk, for example, was characterized by the presence of microbial taxa with industrial interest and contaminants indicative of the quality of milk used for the manufacture of the product. Taxonomic groups located at the same horizontal level indicates symbiotic relationships amongst them. On the contrary, OTUs placed at a different location across the vertical axis indicate mutually inhibitory relationships, e.g. *Lactobacillus* vs. other contaminants. Grouping of the samples is made based on the similarities in the abundance profile of the OTUs and the relative location of these taxonomic groups. Therefore, the taxonomic abundance profile of the metagenomic sample raw milk, located far away from the other samples, was different in comparison with the rest. The two metagenomic samples “whey starter” and “raw milk plus whey starter” were grouped displaying similar taxonomic abundance patterns. Finally, a third distinct group containing only the samples originated from curd and ripening was formed.

For investigating in more detail the identified taxa within and between the samples, the data of the OTU table obtained with SILVAngs pipeline were introduced to GraphPad Prism v6.07 (GraphPad Software, Inc., San Diego, CA, USA) to construct a stacked bars chart (Fig. 4). The figure presents the main microorganisms found in the

Grana-type cheese samples. *Lactobacillus* species dominated all metagenomic samples. *Lb. helveticus* was in high abundance followed by *Lb. delbrueckii*. In the cured and early ripening samples, *Lb. helveticus* and *Lb. delbrueckii* dominated the manufacturing process. On the contrary, in the middle and late ripening metagenomic samples these two *Lactobacillus* species displayed a decrease in their abundance compared to the other samples. At the same time, *Lb. rhamnosus*, *Lb. casei* group and *Lb. fermentum* occurred during ripening. A similar trend, i.e. presence in curd and ripening samples, was also observed for *Propionibacterium* sp. Finally, *Lb. gallinarum* although in a relatively small amount was detected in all metagenomic specimens. The latter together with *Lb. helveticus*, *Lb. delbrueckii*, *Lactobacillus* sp. and *Streptococcus* sp. comprised the core microbiota. *Lb. brevis* and *Lb. plantarum* as well as *Lactococcus lactis*, recovered from whey starter, curd or ripening samples, were incorporated into *Lactobacillus* sp. and *Lactococcus* sp., respectively, because they were not visible alone in Fig. 4. *Streptococcus thermophilus* also was detected in most of the samples. Statistical comparison of the metagenomic samples with the web-based program METAGENassist (Arndt et al., 2012) revealed the significance of the species *Lb. helveticus*, *Lb. delbrueckii*, *Lb. rhamnosus*, *Lb. casei* group, *Lb. fermentum*, *Streptococcus* sp. and *Str. thermophilus*. These observations highlight the specific role of the *Lactobacillus* species as well as the role of the non-starter lactic acid bacteria (NSLAB) and other species during the Grana-type cheese production (Lazzi et al., 2004; Parente and Cogan, 2004; Rossetti et al., 2008; Rossi et al., 2012). The heat map in Fig. 5 shows the symbiotic (between species with industrial interest) and antagonistic (between contaminants and species with industrial interest) interactions that occurred. The web-based program

METAGENassist for comparative metagenomics was used to construct the heat map. Substantial differences between the two taxonomic profiles, the current (with SILVA as reference database) and the one from Alessandria et al. (2016) (with Greengenes as reference database), were not observed, yet some discrepancies do exist. In the present study, *Lb. gallinarum* was found to belong to the core microbiota; and *Lb. rhamnosus* along with *Propionibacterium* species (other than the contaminant *Propionibacterium acnes* present in the study of Alessandria et al., 2016) were recovered from samples during ripening. Such differences were expected since different databases (SILVA vs. Greengenes) were used to perform the taxonomic profiling (Yilmaz et al., 2014; Balvočiūtė and Huson, 2017).

In the raw milk samples, several contaminants were detected such as *Acidovorax* sp., *Acinetobacter* sp., *Acinetobacter baumannii/calcoaceticus* group, *Anoxybacillus* sp., *Clostridium* sp., *Sphingomonas* sp. and *Staphylococcus* sp. The category “other” of Fig. 4 included other contaminants such as *Pseudomonas* sp., *Enterobacter* sp., *Escherichia-Shigella*, *Rubrobacter* sp., *Bacillus* sp or *Listeria monocytogenes*. The recovery of such microorganisms from raw milk using NGS platforms has been reported elsewhere as well (Quigley et al., 2013). Despite the occurrence of several contaminants, these were decreased gradually due to the antagonistic activity experienced by the rest of microbiota, especially the one originated from the *Lactobacillus* species (Fig. 5) supporting the observation made in Fig. 3.

3.2. Submodule 2: Predictive Metagenomics Profiling

3.2.1. Statistical analysis

205 The OTU abundance table, obtained from the 16S rRNA data, was used as input
 206 for the submodule 2 to presume for metabolic functions. Currently, there are three tools
 207 available for PMP: PICRUSt (Langille et al., 2013), Tax4Fun (Aßhauer et al., 2015) and
 208 Piphillin (Iwai et al., 2016). In the present study, the Tax4Fun program performed the
 209 PMP, which works with the SILVA database. The PICRUSt requires the Greengenes
 210 database whereas the Piphillin tool is not obliged to any unique data pre-processing
 211 protocol supporting KEGG and BioCyc as a reference database. The output of the
 212 Tax4Fun is a table with a similar layout to the OTU abundance containing the functional
 213 predictions of KEGG Orthology (KO) or Pathways (ko). Statistical analysis (Kruskal-
 214 Wallis H-test with Tukey-Kramer), using the STAMP v2.1.3 software (Parks and Beiko,
 215 2010; Parks et al., 2014) showed that 1629 KO and 121 ko displayed substantial changes.
 216 A *P*-value lower than 0.05, corrected for multiple tests according to the Benjamini-
 217 Hochberg FDR (False Discovery Rate) procedure, indicated significant differences. PCA
 218 (Principal Component Analysis) plots, made with Past v3.15 software, display the
 219 orientation of the metagenomic samples and the most abundant KEGG Pathways (ko)
 220 (Fig. 6).

221 The curd and early ripening metagenomic samples were dominated by pathways
 222 associated with carbohydrate metabolism (Fig. 6a). Cheese making (curd) and early
 223 ripening samples were mainly located in the right part of the graph (Fig. 6b). The ko
 224 02060 (phosphotransferase system – PTS; membrane transport), 00564
 225 (glycerophospholipid metabolism; lipid metabolism) and 00260 (glycine, serine,
 226 threonine metabolism; amino acid metabolism) also appeared on the right of the vertical
 227 axis (Fig. 6a). The PTS is a mechanism of the bacteria with which they uptake

228 carbohydrates (Kotrba et al., 2001). Lactobacilli consume sugars such as galactose and
229 lactose, and the glucose can be converted to pyruvate through glycolysis (Hemme et al.,
230 1981; Premi et al., 1972). Pyruvate is an important precursor of many metabolites such as
231 lactic acid, formic acid, acetic acid, acetaldehyde, ethanol, acetoin, diacetyl, and butane-
232 2,3-diol (Hickey et al., 1983). Moreover, thermophilic lactobacilli such as *Lb. helveticus*
233 and *Lb. delbrueckii* can produce peptides, amino acids and other metabolites that
234 stimulate the growth of *Str. thermophilus* (Courtin and Rul, 2004; Hemme et al., 1981)
235 and propionibacteria (Baer, 1995; Kerjean et al., 2000; Piveteau et al., 1995).
236 Metatranscriptomics revealed that genes associated with carbohydrate metabolism
237 (pentose phosphate pathway and glycolysis) were enriched during the cheese making
238 process of the traditional Italian cheese Caciocavallo Silano PDO (Protected Designation
239 of Origin) (De Filippis et al., 2016).

240 On the top left corner of Fig. 6b, the metagenomic samples of the middle and late
241 ripening formed a separate group, compared to the other samples. Accordingly, the
242 KEGG Pathways located in the top left area of the PCA graph (Fig. 6a) were related to
243 amino acid (ko00280, valine, leucine, isoleucine degradation; ko00360, phenylalanine
244 metabolism) and lipid (ko00061, fatty acid biosynthesis; ko00071, fatty acid metabolism)
245 metabolism. Also, pathways referred to carbohydrate metabolism (ko00020, TCA cycle;
246 ko00640, propanoate metabolism; ko00630 glyoxylate and dicarboxylate metabolism)
247 were also observed, which may participate in the production of aroma compounds. Flavor
248 formation in cheeses is a complex process involving proteolytic and lipolytic activities in
249 which key players are NSLAB and other non lactic acid bacteria (Smit et al., 2005).
250 Interestingly, samples taken during middle and late ripening of the cheese were

characterized by the gradual increase of *Lb. rhamnosus*, *Lb. casei*, *Lb. fermentum*, *Str. thermophilus* and *Propionibacterium* sp. (Fig. 4). These microorganisms are known for their proteolytic and/or lipolytic activity as well as for their ability to produce aroma compounds (González-Olivares et al., 2014; Hong-Xin et al., 2015; Smit et al., 2005; Thierry et al., 2011). The above results showed good correlation with the observations made during ripening of the traditional Italian cheese Caciocavallo Silano PDO using metatranscriptomics (De Filippis et al., 2016).

3.2.2. Metabolic interactions

Usually, thermophilic lactic starters, propionibacteria, and NSLAB follow one another during ripening of Swiss-type cheeses (Gagnaire et al., 2001). A similar trend was observed in the present study for an Italian Grana-type cheese. Propionibacteria growth is dependent on the availability of lactate which is produced by *Lb. helveticus*, *Lb. delbrueckii* and *St. thermophilus* (Kurtz et al., 1959). Propionibacteria preferably utilize lactate as the energy source (Brendehaug and Langsrud, 1985; Fröhlich-Wyder et al., 2002). Despite the fundamental role of NSLAB in cheese flavor, propionibacteria should be present as well, but not in excess, to allow Grana-type cheeses such as Grana Padano and Parmigiano Reggiano develop their typical organoleptic characteristics (Carcano et al., 1995). The uncontrolled growth of propionibacteria may lead to an undesirable situation known as “late blowing” or “late fermentation” (Carcano et al., 1995; Fröhlich-Wyder et al., 2002).

So, both micro-flora NSLAB and propionibacteria have a role to play during the development of the organoleptic characteristics of the Grana-type cheeses. But how do

propionibacteria interact with NSLAB? Facultatively heterofermentative lactobacilli (FHL) such as *Lb. casei* and *Lb. rhamnosus* may compromise propionibacteria growth (Fröhlich-Wyder et al., 2002; Jimeno et al., 1995), especially when FHL are added as supplemental cultures. Fröhlich-Wyder et al. (2002) have showed that the addition of NSLAB in Swiss-type cheeses inhibited lactate fermentation by the propionibacteria. If, however, NSLAB are naturally occurring during cheese ripening, do they have the same effect on propionibacteria growth or not? Most NSLAB do not affect propionibacteria levels in cheese. The influence of *Lactobacillus* spp. on propionibacteria growth is likely to be less important than the impact of technological parameters such as pH and salt in cheeses (Carcano et al., 1995; Noël, 1999).

Consequently, the question above was explored using the microbial metabolic interactions (MMinte) tool (Mendes-Soares et al., 2016) for investigating the interplay between the naturally occurring flora of NSLAB and propionibacteria. Within the microbial community, the MMinte indicates the nature of that interplay (positive, negative or no interaction) based on the comparison of growth rates between the pairs of the microorganisms by constructing predictive genome-scale metabolic models. The sequence data and the subset of correlations between the considered OTUs, as estimated by the METAGENassist program (Fig. 5), were introduced to the MMinte tool and were run through the six widgets available: widget 1, only the representative 16S rDNA/rRNA sequence data were kept for further analysis based on the provided subset of the microorganisms pairs; widget 2, a genome ID is assigned to each OTU using BLAST (the 16S rDNA/rRNA sequences are compared with reference sequences available in NCBI) (Altschul et al., 1990); widget 3, a predictive metabolic model is constructed using

ModelSEED (Henry et al., 2010) for each genome ID; widget 4, a predictive two-species community metabolic model is created by the mean of COBRApy (Ebrahim et al., 2013; Klitgord and Segrè, 2010) for each pair of microbes provided in the first widget; widget 5, predictions on the growth rates are made using flux balance analysis (Heinken and Thiele, 2015; Varma and Palsson, 1994) for each two-species community; widget 6, the metabolic network of the microbial community is drawn and the nature of interactions is indicated using the D3.js visualization tool (Bostock et al., 2011).

The network in Fig. 7 depicts the predicted interplays between the propionibacteria and NSLAB. When the availability of the metabolites is high (“Complete_100”), propionibacteria is predicted to grow thus take a benefit from the presence of the NSLAB. NSLAB are heterofermentative microorganisms producing lactate among others, which can be used by propionibacteria. *Lb. casei* converts glucose, especially when glucose is limited, to lactate (predominantly), acetate, formate, and ethanol (Liu, 2003). Interestingly, when the availability of the metabolites was reduced by ten times (“Complete_10”) there was an increase in the number of negative interactions predicted to occur between propionibacteria and NSLAB, without a positive interplay between them, meaning that NSLAB impairs the growth of propionibacteria when there is higher competition for nutrients. Consequently, the nature of the interaction is altered based on the metabolites availability. This probably explains partially the observation that NSLAB inhibit the propionibacteria growth when added as supplement cultures. This generates a much greater competition by the NSLAB for nutrients, resulting in the accumulation of elevated quantities (excess) of metabolic products such

319 as acetate, formate, and diacetyl that suspend the increase of propionibacteria (Jimeno et
320 al., 1995).

322 **4. Conclusion**

323 The bioinformatics pipeline described in the present study may also find an
324 application to foodborne pathogens occurring in foodstuffs. As predictive microbiology
325 enters a new era of the integration of meta- and multi- omics in predictive modeling and
326 quantitative risk assessment in foods (Brul et al., 2012; Cocolin et al., 2017; Rantsiou et
327 al., 2011), the workflow proposed here may constitute a useful tool. For instance, it can
328 respond to questions that concern risk assessors, food microbiologists and others dealing
329 with microbiological risk assessment studies: How the foodborne pathogens found in
330 food interact with the rest of microbiota? Why strains of the same species behave
331 differently? How environmental conditions influence important features of the foodborne
332 pathogens such as virulence?

333 As any novel method, PMP also has constraints. In order, the method to make
334 reliable and accurate predictions about the gene content of an OTU, the genome of
335 reference or at least closely related microorganisms should be sequenced and available.
336 Despite this limitation, PMP is a cost-effective and straightforward method to start with
337 when 16S rDNA/rRNA data are available (Wood, 2016). The tool may help to investigate
338 the functional potential, alterations, and interactions of a microbial community. Thus, it
339 will provide evidence for further exploration of the community and guide future
340 experiments based on the genes or gene groups predicted to change (Wood, 2016).

341 Finally, PMP is in line with the multi-omics approach in food (safety) microbiology
342 (Ferrocino and Cocolin, 2017).

343

897
898
899 **References**
900

- 901 345 Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: Predicting
902 346 functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882-
903 347 2884.
- 904 348 Alessandria, V., Ferrocino, I., De Filippis, F., Fontana, M., Rantsiou, K., Ercolini, D.,
905 349 Cocolin, L., 2016. Microbiota of an Italian Grana-like cheese during manufacture
906 350 and ripening, unraveled by 16S rRNA-based approaches. *Appl. Environ.*
907 351 *Microbiol.* 82, 3988-3995.
- 908 352 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local
909 353 alignment search tool. *J. Mol. Biol.* 215, 403-410.
- 910 354 Arndt, D., Xia, J., Liu, Y., Zhou, Y., Guo, A.C., Cruz, J.A., Sinelnikov, I., Budwill, K.,
911 355 Nesbø, C.L., Wishart, D.S., 2012. METAGENassist: a comprehensive web server
912 356 for comparative metagenomics. *Nucleic Acids Res.* 40, W88-W95.
- 913 357 Baer, A., 1995. Influence of casein proteolysis by starter bacteria, rennet and plasmin on
914 358 the growth of propionibacteria in Swiss-type cheese. *Lait* 75, 391-400.
- 915 359 Balvočiūtė, M., Huson, D.H., 2017. SILVA, RDP, Greengenes, NCBI and OTT – how do
916 360 these taxonomies compare? *BMC Genomics* 18 (Suppl 2), 114.
- 917 361 Bostock, M., Ogievetsky, V., Heer, J. 2011. D³: Data-Driven Documents. *IEEE Trans.*
918 362 *Vis. Comput. Graph.* 17, 2301-2309.
- 919 363 Brendehaug, J., Langsrud, T., 1985. Amino acid metabolism in propionibacteria: resting
920 364 cells experiments with four strains. *J. Dairy Sci.* 68, 281–289.

- 365 Brul, S., Bassett, J., Cook, P., Kathariou, S., McClure, P., Jasti, P.R., Betts, R., 2012.
- 366 'Omics' technologies in quantitative microbial risk assessment. Trends Food Sci.
- 367 Technol. 27, 12-24.
- 368 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden,
- 369 T., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.
- 370 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.
- 371 K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley,
- 372 S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D.,
- 373 Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters,
- 374 W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME
- 375 allows analysis of high-throughput community sequencing data. Nat. Methods 7,
- 376 335-336.
- 377 Carcano, M., Todesco, R., Lodi, R., Brasca, M., 1995. Propionibacteria in Italian hard
- 378 cheeses. Lait 75, 415-426.
- 379 Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.-F., Jagadeesan, B.,
- 380 Rantsiou, K., Phister, T., 2017. Meta-Omics and Microbial Risk Assessment: the
- 381 next need for integration. Int. J. Food Microbiol. (under revision).
- 382 Coughlan, L.M., Cotter, P.D., Hill, C., Alvarez-Ordóñez, A., 2015. Biotechnological
- 383 applications of functional metagenomics in the food and pharmaceutical
- 384 industries. Front. Microbiol. 6, 672.
- 385 Courtin, P., Rul, F., 2004. Interactions between microorganisms in a simple ecosystem:
- 386 yogurt bacteria as a study model. Lait 84, 125-134.

387 De Filippis, F., Genovese, A., Ferranti, P., Gilbert, J.A., Ercolini, D., 2016.

388 Metatranscriptomics reveals temperature-driven functional changes in

389 microbiome impacting cheese maturation rate. *Sci. Rep.* 6, 21871.

390 De Filippo, C., Ramazzotti, M., Fontana, P., Cavalieri, D., 2012. Bioinformatics

391 approaches and pathway inference in metagenomics data. *Brief. Bioinformatics*

392 13, 696-710.

393 Delcenserie, V., Taminiau, B., Delhalle, L., Nezer, C., Doyen, P., Crevecoeur, S.,

394 Roussey, D., Korsak, N., Daube, G., 2014. Microbiota characterization of a

395 Belgian protected designation of origin cheese, Herve cheese, using metagenomic

396 analysis. *J. Dairy Sci.* 97, 6046–6056.

397 Dhariwal, A., Chong, J., Habib, S., King, I.L., Agellon, L.B., Xia, J., 2017.

398 MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and

399 meta-analysis of microbiome data. *Nucleic Acids Res.* 45, W180-W188.

400 Dudhagara, P., Bhavsar, S., Bhagat, C., Ghelani, A., Bhatt, S., Patel, R., 2015. Web

401 Resources for Metagenomics Studies. *Genomics Proteomics Bioinformatics* 13,

402 296-303.

403 Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. COBRApy: COntstraints-

404 Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* 7, 74.

405 Ercolini, D., De Filippis, F., La Stora, A., Iacono, M., 2012. “Remake” by High-

406 Throughput Sequencing of the Microbiota Involved in the Production of Water

407 Buffalo Mozzarella Cheese. *Appl. Environ. Microbiol.* 78, 8142-8145.

- Escobar-Zepeda, A., Vera-Ponce de Leon, A., Sanchez-Flores, A., 2015. The Road to Metagenomics: From Microbiology to DNA Sequencing Technologies and Bioinformatics. *Front. Genet.* 6, 348.
- Ferrocino, I., Cocolin, L., 2017. Current perspectives in food-based studies exploiting multi-omics approaches. *Curr. Opin. Food Sci.* 13, 10-15.
- Ferrocino, I., Greppi, G., La Stora, A., Rantsiou, K., Ercolini, D., Cocolin, L., 2016. Impact of Nisin-Activated Packaging on Microbiota of Beef Burgers during Storage. *Appl. Environ. Microbiol.* 82, 549-559.
- Fröhlich-Wyder, M.-T., Bachmann, H.-P., Casey, M.G., 2002. Interaction between propionibacteria and starter / non-starter lactic acid bacteria in Swiss-type cheeses. *Lait* 82, 1-15.
- Gagnaire, V., Molle, D., Herrouin, M., Leonil, J., 2001. Peptides identified during Emmental cheese ripening: Origin and proteolytic systems involved. *J. Agric. Food Chem.* 49, 4402-4413.
- González-Olivares, L.G., López-Cuellar, Z.L., Añorve-Morga, J., Franco-Fernández, M.J., Castañeda-Ovando, A., Contreras-López, E., Jaimez-Ordaz, J., Rodríguez-Serrano, G.M., 2014. Viability and Proteolytic Capacity of *Lactobacillus bulgaricus* 2772 and *Lactobacillus rhamnosus* GG during Cheese Ripening. *J. Biosci. Med.* 2, 7-12.
- Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* 4, 9.
- Heinken, A., Thiele, I., 2015. Anoxic conditions promote species-specific mutualism between gut microbes *in silico*. *Appl. Environ. Microbiol.* 81, 4049-4061.

- 431 Hemme, D.H., Schmal, V., Auclair, J.E., 1981. Effect of the addition of extracts of
432 thermophilic lactobacilli on acid production by *Streptococcus thermophilus* in
433 milk. J. Dairy Res. 48, 139-148.
- 434 Henry, C.S., De Jongh, M., Best, A.A., Frybarger, P.M., Lindsay, B., Stevens, R.L., 2010.
435 High-throughput generation, optimization and analysis of genome-scale metabolic
436 models. Nat. Biotechnol. 28, 977-982.
- 437 Hickey, M.W., Hillier, A.J., Jago, G.R., 1983. Metabolism of Pyruvate and Citrate in
438 Lactobacilli. Aust. J. Biol. Sci. 36, 487-496.
- 439 Hong-Xin, J., Mi-Ya, S., Guang-Yu, G., 2015. Influence of *Lactobacillus casei* LC2W on
440 the proteolysis and aroma compounds of Cheddar cheese during ripening period.
441 CyTA – J. Food 13, 464-471.
- 442 Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K.,
443 DeSantis, T.Z., 2016. Piphillin: Improved Prediction of Metagenomic Content by
444 Direct Inference from Human Microbiomes. PLoS One 11, e0166104.
- 445 Jimeno, J., Lazaro, M.J., Sollberger, H., 1995. Antagonistic interactions between
446 propionic acid bacteria and non-starter lactic acid bacteria. Lait 75, 401-413.
- 447 Keller, A., Horn, H., Forster, F., Schultz, J., 2014. Computational integration of genomic
448 traits into 16S rDNA microbiota sequencing studies. Gene 549, 186–191.
- 449 Kerjean, J.R., Condon, S., Lodi, R., Kalantzopoulos, G., Chamba, J.F., Suomalainen, T.,
450 Cogan, T., Moreau, D., 2000. Improving the quality of European hard-cheeses by
451 controlling of interactions between lactic acid bacteria and propionibacteria. Food
452 Res. Int. 33, 281–287.

- 453 Klitgord, N., Segrè D., 2010. Environments that induce synthetic microbial ecosystems.
454 PLoS Comput. Biol. 6, e1001002.
- 455 Kotrba, P., Inui, M., Yukawa, H., 2001. Bacterial phosphotransferase system (PTS) in
456 carbohydrate uptake and control of carbon metabolism. J. Biosci. Bioeng. 92,
457 502-517.
- 458 Kuntal, B.K., Ghosh, T.S., Mande, S.S., 2013. Community-Analyzer: A platform for
459 visualizing and comparing microbial community structure across microbiomes.
460 Genomics 102, 409-418.
- 461 Kurtz, F.E., Hupfer, J.A., Corbin, E.A., Hargrove, R.E., Walter, H.E., 1959.
462 Interrelationships between pH, populations of *Propionibacterium shermanii*,
463 levels of free fatty acids, and the flavor ratings of Swiss cheeses. J. Dairy Sci. 42,
464 1008-1019.
- 465 Ladoukakis, E., Kolisis, F.N., Chatziioannou, A.A., 2014. Integrative workflows for
466 metagenomic analysis. Front. Cell Dev. Biol. 2, 70.
- 467 Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A.,
468 Clemente, J.C., Burkepille, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G.,
469 Huttenhower, C., 2013. Predictive functional profiling of microbial communities
470 using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814-821.
- 471 Lazzi, C., Rossetti, L., Zago, M., Neviani, E., Giraffa, G., 2004. Evaluation of bacterial
472 communities belonging to natural whey starters for Grana Padano cheese by
473 length heterogeneity-PCR. J. Appl. Microbiol. 96, 481-490.
- 474 Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets
475 of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.

- 476 Liu, S.Q., 2003. Practical implications of lactate and pyruvate metabolism by lactic acid
477 bacteria in food and beverage fermentations. *Int. J. Food Microbiol.* 83, 115-131.
- 478 Liu, W., Zheng, Y., Kwok, L.-Y., Sun, Z., Zhang, J., Guo, Z., Hou, Q., Menhe, B.,
479 Zhang, H., 2015. High-throughput sequencing for the detection of the bacterial
480 and fungal diversity in Mongolian naturally fermented cow's milk in Russia.
481 *BMC Microbiol.* 15, 45.
- 482 Mendes-Soares, H., Mundy, M., Mendes-Soares, L., Chia, N., 2016. MMinte: an
483 application for predicting metabolic interactions among the microbial species in a
484 community. *BMC Bioinformatics* 17, 343.
- 485 Minot, S.S., Krumm, N., Greenfield, N.B., 2015. One Codex: A Sensitive and Accurate
486 Data Platform for Genomic Microbial Identification. *bioRxiv*, doi:
487 <http://dx.doi.org/10.1101/027607>.
- 488 Noël, Y., Poyoval P., Thierry A., Gagnaire V., Grappin, R., 1999. Eye formation and
489 Swiss-type cheeses, In: Law, B.A. (Ed.), *Technology of cheese making*. CRC
490 Press, Boca Raton, pp. 222-250.
- 491 Oulas, A., Pavloudi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N.,
492 Kotoulas, G., Arvanitidis, C., Iliopoulos, I., 2015. Metagenomics: Tools and
493 Insights for Analyzing Next-Generation Sequencing Data Derived from
494 Biodiversity Studies. *Bioinform. Biol. Insights* 9, 75-88.
- 495 Parente, E., Cogan, T.M., 2004. Starter cultures: general aspects. In: Fox, P.F., et al.
496 (Eds.), *Cheese – Chemistry, Physics, and Microbiology*. Elsevier Academic Press,
497 London, pp. 123-148.

- 498 Parks, D.H., Beiko, R.G., 2010. Identifying biologically relevant differences between
499 metagenomic communities. *Bioinformatics* 26, 715-721.
- 500 Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G., 2014. STAMP: statistical
501 analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123-3124.
- 502 Parlapani, F.F., Meziti, A., Kormas, K.Ar., Boziaris, I.S., 2013. Indigenous and spoilage
503 microbiota of farmed sea bream stored in ice identified by phenotypic and 16S
504 rRNA gene analysis. *Food Microbiol.* 33, 85-89.
- 505 Parlapani, F.F., Kormas, K.Ar., Boziaris, I.S., 2015. Microbiological changes, shelf life
506 and identification of initial and spoilage microbiota of sea bream fillets stored
507 under various conditions using 16S rRNA gene analysis. *J. Sci. Food Agric.* 95,
508 2386–2394.
- 509 Parlapani, F.F., Boziaris, I.S., 2016. Monitoring of spoilage and determination of
510 microbial communities based on 16S rRNA gene sequence analysis of whole sea
511 bream stored at various temperatures. *LWT - Food Sci. Technol.* 66, 553-559.
- 512 Piveteau, P.G., Condon S., Cogan T.M., 1995. Interactions between lactic and propionic
513 acid bacteria. *Lait* 75, 331-343.
- 514 Połka, J., Rebecchi, A., Pisacane, V., Morelli, L., Puglisi, E., 2015. Bacterial diversity in
515 typical Italian salami at different ripening stages as revealed by high-throughput
516 sequencing of 16S rRNA amplicons. *Food Microbiol.* 46, 342-356.
- 517 Pothakos, V., Stellato, G., Ercolini, D., Devlieghere, F., 2015. Processing Environment
518 and Ingredients Are Both Sources of *Leuconostoc gelidum*, Which Emerges as a
519 Major Spoiler in Ready-To-Eat Meals. *Appl. Environ. Microbiol.* 81, 3529-3541.

- 520 Premi, L., Sandine, W.E., Elliker, P.R., 1972. Lactose-hydrolyzing enzymes of
521 *Lactobacillus* species. Appl. Microbiol. 24, 51-57.
- 522 Pruesse, E., Peplies, J., Glöckner, F.O., 2012. SINA: accurate high-throughput multiple
523 sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823-1829.
- 524 Pylro, V.S., Morais, D.K., de Oliveira, F.S., dos Santos, F.G., Lemos, L.N., Oliveira, G.,
525 Roesch, L.F.W., 2016. BMPOS: a Flexible and User-Friendly Tool Sets for
526 Microbiome Studies. Microb. Ecol. 72, 443-447.
- 527 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J.,
528 Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project:
529 improved data processing and web-based tools. Nucleic Acids Res. 41, D590-
530 D596.
- 531 Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T.P., Ross, R.P., Fitzgerald, G.F.,
532 Cotter, P.D., 2013. The complex microbiota of raw milk. FEMS Microbiol. Rev.
533 37, 664-698.
- 534 Rantsiou, K., Mataragas, M., Jespersen, L., Cocolin, L., 2011. Understanding the
535 behavior of foodborne pathogens in the food chain: New information for risk
536 assessment analysis. Trends Food Sci. Technol. 22, S21-S29.
- 537 Robertson, C.E., Harris, J.K., Wagner, B.D, Granger, D., Browne, K., Tatem, B., Feazel,
538 L.M., Park, K., Pace1, N.R., Frank, D.N., 2013. Explicet: graphical user interface
539 software for metadata-driven management, analysis and visualization of
540 microbiome data. Bioinformatics 29, 3100-3101.

- 541 Rossetti, L., Fornasari, M.E., Gatti, M., Lazzi, C., Neviani, E., Giragga, G., 2008. Grana
542 Padano cheese whey starters: Microbial composition and strain distribution. *Int. J.*
543 *Food Microbiol.* 127, 168-171.
- 544 Rossi, F., Gatto, V., Sabattini, G., Torriani, S., 2012. An assessment of factors
545 characterising the microbiology of Grana Trentino cheese, a Grana-type cheese.
546 *Int. J. Dairy Technol.* 65, 1-9.
- 547 Roumpeka, D.D., Wallace, R.J., Escalettes, F., Fotheringham, I., Watson, M., 2017. A
548 review of bioinformatics tools for bio-prospecting from metagenomic sequence
549 data. *Front. Genet.* 8, 23.
- 550 Sattin, E., Andreani, N.A., Carraro, L., Lucchini, R., Fasolato, L., Telatin, A., Balzan, S.,
551 Novelli, E., Simionati, B., Cardazzo, B., 2016. A Multi-Omics Approach to
552 Evaluate the Quality of Milk Whey Used in Ricotta Cheese Production. *Front.*
553 *Microbiol.* 7, 1272.
- 554 Scholz, M.B., Lo, C.-C., Chain, P.S.G., 2012. Next generation sequencing and
555 bioinformatics bottlenecks: the current state of metagenomic data analysis. *Curr.*
556 *Opin. Biotechnol.* 23, 9-15.
- 557 Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. *Front.*
558 *Plant Sci.* 5, 209.
- 559 Smit, G., Smit, B.A., Engels, W.J.M., 2005. Flavour formation by lactic acid bacteria and
560 biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591-
561 610.
- 562 Stellato, G., La Storia, A., De Filippis, F., Borriello, G., Villani, F., Ercolini, D., 2016.
563 Overlap of Spoilage-Associated Microbiota between Meat and the Meat

Processing Environment in Small-Scale and Large-Scale Retail Distributions.
 Appl. Environ. Microbiol. 82, 4045-4054.

Thierry, A., Deutsch, S.-M., Falentin, H., Dalmasso, M., Fabien J. Cousin, F.J., Jan, G.,
 2011. New insights into physiology and metabolism of *Propionibacterium*
freudenreichii. Int. J. Food Microbiol. 149, 19-27.

Varma, A., Palsson, B., 1994. Stoichiometric flux balance models quantitatively predict
 growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.
 Appl. Environ. Microbiol. 60, 3724–3731.

Wood, J., 2016. Predictive metagenomics profiling: why, what and how? Bioinformatics
 Rev. 2, 1-4. The article is available at the URL:
[https://bioinformaticsreview.com/20160320/predictive-metagenomics-profiling-
 why-what-and-how/](https://bioinformaticsreview.com/20160320/predictive-metagenomics-profiling-why-what-and-how/)

Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Priesse, E., Quast, C., Schweer, T.,
 Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and “All-species
 Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res. 42,
 D643-D648.

Yoon, S.-H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J., 2017. Introducing
 EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and
 whole-genome assemblies. Int. J. Syst. Evol. Microbiol. 67, 1613-1617.

584 **Figure legends**

585 **Fig. 1.** Proposed bioinformatics pipeline for analysis of NGS-based 16S rDNA/rRNA
586 sequencing data derived from food metagenomics integrating both taxonomic and
587 functional profiling. Solid lines show the workflow of the analysis pipeline. Dashed lines
588 indicate the two steps interfering with the analysis workflow. The quantification step
589 includes two submodules.

590 **Fig. 2.** Box-plots of the ANOSIM statistical test for the microbial communities of the
591 batch D (Group 1), E (Group 2) and F (Group 3).

592 **Fig. 3.** Overview of the microbial community at the genus level (green boxes) found in
593 the Grana-type cheese samples (blue boxes). WS, whey starter; RM, raw milk; MS, raw
594 milk and whey starter; CAC, curd after cutting; CAH, curd after heating; CAP, curd after
595 pressing; CASR, curd after storage room; CHAS, cheese after salting; CH2RM, cheese
596 after two months of ripening; CH4RM, cheese after four months of ripening; CH6RM,
597 cheese after six months of ripening; CH8RM, cheese after eight months of ripening; and
598 CH10RM, cheese after ten months of ripening.

599 **Fig. 4.** Stacked bars chart depicting the percentage of taxa in each metagenomic sample
600 and the changes occurred between samples. WS, whey starter; RM, raw milk; MS, raw
601 milk and whey starter; Curd, all curd samples (curd after cutting, curd after heating, curd
602 after pressing and curd after storage room); ER, early ripening samples (cheese after
603 salting and cheese after two months of ripening); MR, middle ripening samples (cheese
604 after four and six months of ripening); and LR, late ripening samples (cheese after eight
605 and ten months of ripening).

606 **Fig. 5.** Heat map of the correlation matrix between the taxa.

Fig. 6. The orientation of the a) most abundant KEGG Pathways (ko) and b) metagenomic samples using Principal Component Analysis (PCA). Two principal components (1 and 2) were extracted based on the total variance explained. The percentage shows the variance explained by each particular linear component. Upper-right quadrant has higher readings than points in the lower-left quadrant. Colors indicate KEGG Pathways (ko) related with specific metabolism or function such as carbohydrate metabolism (green), lipid metabolism (red), amino acid metabolism (blue), metabolism of cofactors and vitamins (gray), xenobiotics biodegradation and metabolism (salmon), and membrane transport (orange). Phosphotransferase system – PTS (ko02060), Glycolysis/Gluconeogenesis (ko00010), Galactose metabolism (ko00052), Starch and sucrose metabolism (ko00500), Pentose phosphate pathway (ko00030), Glycerophospholipid metabolism (ko00564), Glycine, serine, threonine metabolism (ko00260), Pyruvate metabolism (ko00620), Valine, leucine, isoleucine degradation (ko00280), Phenylalanine metabolism (ko00360), Fatty acid biosynthesis (ko00061), Fatty acid metabolism (ko00071), Benzoate metabolism (ko00362), Aminobenzoate metabolism (ko00627), Folate biosynthesis (ko00790), Ascorbate metabolism (ko00053), Glyoxylate and dicarboxylate metabolism (ko00630), TCA (citrate) cycle (ko00020) and Propanoate metabolism (ko00640). WS (whey starter – khaki), RM (raw milk – gold), MW (raw milk plus whey starter – salmon), Curd (cheese making – green), ER (early ripening – blue), MR (middle ripening – orange), LR (late ripening – red) are the group names of the metagenomic samples.

Fig. 7. Metabolic interaction network between propionibacteria and NSLAB; “Complete_100” and “Complete_10” indicate the availability of the metabolites to the

630 microbial community; “Complete_100”, over 400 metabolites are available, with a flux
631 for the import reactions of 100 mmol/gDW/h (high availability); “Complete_10”, the
632 same metabolites but with reaction fluxes of 10 mmol/gDW/h (ten-times lower
633 availability); The percentage next to the microorganisms indicates the similarity between
634 the OTU sequence provided and the reference sequence (genome ID assigned); The
635 thickness and the length of the links inside the network imitate the correlation values
636 provided. When the correlation value increases, the line is becoming thicker and shorter.
637

Table 1. Software and statistical tests used in each stage of the pipeline. Alternative software is also proposed.

Pipeline step	Statistical test and Software used	Alternative software
Preprocessing	Performed by Alessandria et al. (2016) using Qiime v.1.9.0	SILVAngs pipeline BMPOS pipeline (Pylro et al., 2016)
Taxonomic profiling (Taxonomic Operational Units – OTUs)	SILVAngs pipeline using the SILVA database	BMPOS pipeline (Greengenes database) EzBioCloud database (Yoon et al., 2017) One Codex pipeline (Minot et al., 2015)
Statistical comparison of the metagenomic samples	Analysis of Similarity (ANOSIM) using the Past software	Stamp MicrobiomeAnalyst (Dhariwal et al., 2017) Explicet (Robertson et al., 2013)
Microbial community overview	Community-Analyzer Stacked bars chart using the GraphPad Prism software	MicrobiomeAnalyst Explicet
Statistical significance of the identified OTUs	METAGENassist	MicrobiomeAnalyst Explicet
Symbiotic and antagonistic relationships within the microbial community	Heatmap constructed based on the Pearson correlation using the METAGENassist software	MicrobiomeAnalyst Explicet
Predictive Metagenomics Profiling (PMP)	Tax4Fun	Picrust Piphillin MicrobiomeAnalyst
Statistical analysis of the PMP results	Kruskal-Wallis H-test with Tuckey-Kramer corrected for multiple tests according to Benjamini-Hockberg False Discovery Rate using the Stamp software	MicrobiomeAnalyst
Orientation of the metagenomic samples of the most abundant KEGG pathways	Principal Component Analysis (PCA) using the Past software	MicrobiomeAnalyst Stamp
Metabolic interactions within the microbial community	MMinte	–

Fig. 1

Fig. 2

Fig. 3

Fig. 4

Fig. 5

Fig. 6

a

b

Fig. 7