

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=21BAF05CC12C65CDFD2AB7BAE9D813AE74BA2D35A10338AA95909E62468959B96A13FE143C8765D74A7B92D881474161>]

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

1
2
3
4 **1 A bioinformatics pipeline integrating predictive metagenomics profiling for the**
5 **2 analysis of 16S rDNA/rRNA sequencing data originated from foods**
6
7
8
9

10 4 Marios Mataragas^{a*}, Valentina Alessandria^b, Ilario Ferrocino^b, Kalliopi Rantsiou^b and
11 Luca Cocolin^b
12
13
14
15

16 7 ^a Hellenic Agricultural Organization “DEMETER”, Institute of Technology of
17
18 8 Agricultural Products, Department of Dairy Research, Ethnikis Antistaseos 3, 45221,
19
20 9 Ioannina, Greece
21

22 10 ^b University of Turin, Department of Agricultural, Forest and Food Sciences, Laboratory
23
24
25 11 of Food Microbiology, Largo P. Braccini 2, 10095, Grugliasco, Turin, Italy
26
27
28
29
30
31
32

33 15 *Corresponding author: mmatster@gmail.com, Hellenic Agricultural Organization
34
35 16 “DEMETER”, Institute of Technology of Agricultural Products, Department of Dairy
36
37 17 Research, Ethnikis Antistaseos 3, 45221, Ioannina, Greece
38
39
40
41
42
43
44
45

46 21 Keywords: Amplicon Sequencing; Food Microbiology; Metabolic Network; Meta-omics;
47
48 22 Microbial Risk Assessment; Molecular Microbiology; Predictive Functional Profiling
49
50
51
52
53
54
55
56

57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112

24 **Abstract**

25 The recent advances in molecular biology, such as the advent of next-generation
26 sequencing (NGS) platforms, have paved the way to new exciting tools which rapidly
27 transform food microbiology. Nowadays, NGS methods such as 16S rDNA/rRNA
28 metagenomics or amplicon sequencing are used for the taxonomic profiling of the food
29 microbial communities. Although 16S rDNA/rRNA NGS-based microbial data are not
30 suited for the investigation of the functional potential of the identified operational
31 taxonomic units as compared to shotgun metagenomics, advances in the bioinformatics
32 discipline allow now the performance of such studies. In this paper, a bioinformatics
33 workflow is described integrating predictive metagenomics profiling with specific
34 application to food microbiology data. Bioinformatics tools pertinent to each sub-module
35 of the pipeline are suggested as well. The published 16S rDNA/rRNA amplicon data
36 originated from an Italian Grana-type cheese, using an NGS platform, was employed to
37 demonstrate the predictive metagenomics profiling approach. The pipeline identified the
38 microbial community and the changes that occurred in the microbial profile during
39 manufacture of the food product studied (taxonomic profiling). The workflow also
40 indicated significant changes in the functional profiling of the community. The tool may
41 help to investigate the functional potential, alterations, and interactions of a microbial
42 community. The proposed workflow may also find an application in the investigation of
43 the ecology of foodborne pathogens encountered in various food products.

44

113
114
115 **45 1. Introduction**
116

117 46 The objective of this work was to suggest and describe a bioinformatics workflow
118
119 47 for the analysis of metagenomic data based on the 16S rDNA/rRNA amplicon sequencing
120
121 48 originated from the application of next-generation sequencing (NGS) platforms. The
122
123 49 pipeline integrates functional metagenomics, which is an emerging technique with
124
125 50 potential industrial interest (Coughlan et al., 2015). Usually, papers dealing with the
126
127 51 investigation of microbial ecology in food products using NGS methods end up with the
128
129 52 taxonomic profiling of the microbial community after the preprocessing of the obtained
130
131 53 16S rDNA/rRNA data (Alessandria et al., 2016; Delcenserie et al., 2014; Ercolini et al.,
132
133 54 2012; Liu et al., 2015; Parlapani and Boziaris, 2016; Parlapani et al., 2013; Parlapani et
134
135 55 al., 2015; Połka et al., 2015; Sattin et al., 2016). However, data derived from 16S
136
137 56 rDNA/rRNA amplicon sequencing can be exploited to investigate the functional potential
138
139 57 of the identified operational taxonomic units (OTUs). Only recently, a few studies have
140
141 58 performed functional profiling (Ferrocino et al., 2016; Pothakos et al., 2015; Stellato et
142
143 59 al., 2016), but in general, this is not a common practice. The 16S rDNA/rRNA amplicon
144
145 60 sequencing is a form of metagenomics and not metatranscriptomics, and therefore, the
146
147 61 analysis is known as predictive functional profiling (Langille et al., 2013) or predictive
148
149 62 metagenomics profiling (Wood, 2016). Other authors have suggested the integration of
150
151 63 functional metagenomics into 16S rDNA/rRNA studies (Coughlan et al., 2015; Keller et
152
153 64 al., 2014), but a key difference between those studies and the currently proposed
154
155 65 bioinformatics pipeline is the inclusion of an additional step for the prediction of
156
157 66 metabolic interactions between the microbial species found in a community (Mendes-
158
159 67 Soares et al., 2016), an analysis not previously suggested or performed in food
160
161
162
163
164
165
166
167
168

169
170
171 68 metagenomes. In addition, the proposed food-focused pipeline involves a selection of
172
173 69 tools and their specific sequential use along with the statistical tests, describing a step-
174
175 70 wise use of each program and statistical test in each submodule. This will provide a quick
176
177
178 71 and easy reference for the user who would like to use the programs in correct order. The
179
180 72 16S rRNA amplicon data originated from a Grana-type Italian cheese using an NGS
181
182 73 platform (Alessandria et al., 2016) were used to demonstrate the predictive metagenomics
183
184 74 profiling approach.
185
186
187
188

75

76 **2. Bioinformatics workflow**

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

225
226
227 91 Scholz et al., 2012; Sharpton, 2014). On the contrary, the number of available tools for
228
229 92 PMP of 16S rDNA/rRNA amplicon data is limited (Aßhauer et al., 2015; Iwai et al.,
230
231
232 93 2016; Langille et al., 2013).
233
234 94

235
236 95 **3. Case study: taxonomic and functional profiling of the microbial community of a**
237
238 96 **hard, slow-ripened cheese**

239
240 97 The data used were from the study of Alessandria et al. (2016). The Sequence
241
242 98 Read Archive (SRA) website of the National Center for Biotechnology Information
243
244 99 (NCBI) was accessed to download all the deposited sequences in FASTA format
245
246 100 (<https://trace.ncbi.nlm.nih.gov/Traces/sra/>). Three different batches (D, E, and F) of a
247
248 101 Grana-type Italian cheese were used to get food metagenomics data by pyrosequencing
249
250 102 (Roche 454 GS Junior platform) of the amplified V1 to V3 region of the 16S rRNA
251
252 103 marker gene. The authors collected thirty-nine samples in total ($n = 39$; 13 samples per
253
254 104 batch) during manufacture and ripening of the cheese (Whey Starter, WS; Raw Milk,
255
256 105 RM; Raw Milk and Whey Starter, MS; Curd after Cutting, CAC; Curd after Heating,
257
258 106 CAH; Curd after Pressing, CAP; Curd after Storage Room, CASR; Cheese after Salting,
259
260 107 CHAS; Second Ripening Month, CH2RM; Fourth Ripening Month, CH4RM; Sixth
261
262 108 Ripening Month, CH6RM; Eighth Ripening Month, CH8RM; Tenth Ripening Month,
263
264 109 CH10RM) for pyrosequencing purposes. Preprocessing (stage 1 of the proposed
265
266 110 bioinformatics workflow of Fig. 1) of the downloaded sequences had already being
267
268 111 performed by Alessandria et al. (2016) with QIIME v1.9.0 (Caporaso et al., 2010), and
269
270 112 therefore in this case study only the quantification step (submodule 1 and 2 of the
271
272 113 proposed bioinformatics pipeline of Fig. 1) was carried out.
273
274
275
276
277
278
279
280

281
282
283 114
284

285 115 *3.1. Submodule 1: Taxonomic profiling*
286

287
288 116 Taxonomic profiling was performed using the SILVAngs 1.3 pipeline (Quast et
289
290 117 al., 2013). Each downloaded sequence (264826 sequences in total) was aligned using the
291
292 118 SILVA Incremental Aligner (SINA v1.2.10 for ARB SVN, revision 21008) (Pruesse et
293
294 119 al., 2012) against the SILVA SSU rRNA SEED and quality controlled (Quast et al.,
295
296 120 2013). Quality control of the submitted sequences, using the standard settings of the
297
298 121 pipeline, rejected 89660 sequences (number of classified sequences equal to 173225 and
299
300 122 number of “No Relative” equal to 1941). Afterward, identical reads were identified
301
302 123 (dereplication), unique reads were clustered (OTUs), on a per sample basis, and reference
303
304 124 read of each OTU was classified. Dereplication and clustering were made using cd-hit-est
305
306 125 (version 3.1.2) (Li and Godzik, 2006) running in “accurate mode”, ignoring overhangs,
307
308 126 and applying identity criteria of 1.00 and 0.98, respectively. The classification was done
309
310 127 by a local nucleotide BLAST search against the non-redundant version of the SILVA
311
312 128 SSU Ref dataset (release 128) using blastn (version 2.2.30+) with standard settings
313
314 129 (Camacho et al., 2009). Reads without any BLAST hits or reads with weak BLAST hits
315
316 130 (Similarity \leq 93%) remained unclassified (“No Relative”). The output of the pipeline,
317
318 131 among others, was an OTU table containing the OTU abundances per sample at the genus
319
320 132 and species level. The taxonomy at the species level was not possible for all the OTUs.
321
322 133 The matrix was filtered further by applying the same filtering criteria with Alessandria et
323
324 134 al. (2016), i.e. including only those OTUs with abundance \geq 0.5% in at least two samples.
325
326 135 The filtered table was the final output kept for all the subsequent steps.
327
328
329
330
331
332
333
334
335
336

337
338
339 136 After removal of the sample identified as outliers, no significant differences were
340
341 137 observed between the three D, E and F batches regarding the microbial community
342
343 138 profile using the ANOSIM (Analysis of Similarity) statistical test ($P = 0.352$; $P_{D-E} =$
344
345 139 0.311 ; $P_{D-F} = 0.370$; $P_{E-F} = 0.376$) (Fig. 2) of the Past v3.15 software (Hammer et al.,
346
347 140 2001). Fig. 3 displays an overview of the microbial community profile at the genus level
348
349 141 during manufacture of the Grana-type cheese using the Community-Analyzer program
350
351 142 (Kuntal et al., 2013). The arrows show the presence of an OTU in a particular
352
353 143 metagenomic sample. Raw milk, for example, was characterized by the presence of
354
355 144 microbial taxa with industrial interest and contaminants indicative of the quality of milk
356
357 145 used for the manufacture of the product. Taxonomic groups located at the same
358
359 146 horizontal level indicates symbiotic relationships amongst them. On the contrary, OTUs
360
361 147 placed at a different location across the vertical axis indicate mutually inhibitory
362
363 148 relationships, e.g. *Lactobacillus* vs. other contaminants. Grouping of the samples is made
364
365 149 based on the similarities in the abundance profile of the OTUs and the relative location of
366
367 150 these taxonomic groups. Therefore, the taxonomic abundance profile of the metagenomic
368
369 151 sample raw milk, located far away from the other samples, was different in comparison
370
371 152 with the rest. The two metagenomic samples “whey starter” and “raw milk plus whey
372
373 153 starter” were grouped displaying similar taxonomic abundance patterns. Finally, a third
374
375 154 distinct group containing only the samples originated from curd and ripening was formed.
376
377

378
379 155 For investigating in more detail the identified taxa within and between the
380
381 156 samples, the data of the OTU table obtained with SILVAngs pipeline were introduced to
382
383 157 GraphPad Prism v6.07 (GraphPad Software, Inc., San Diego, CA, USA) to construct a
384
385 158 stacked bars chart (Fig. 4). The figure presents the main microorganisms found in the
386
387
388
389
390
391
392

393
394
395 159 Grana-type cheese samples. *Lactobacillus* species dominated all metagenomic samples.
396
397 160 *Lb. helveticus* was in high abundance followed by *Lb. delbrueckii*. In the cured and early
398
399 161 ripening samples, *Lb. helveticus* and *Lb. delbrueckii* dominated the manufacturing
400
401 162 process. On the contrary, in the middle and late ripening metagenomic samples these two
402
403 163 *Lactobacillus* species displayed a decrease in their abundance compared to the other
404
405 164 samples. At the same time, *Lb. rhamnosus*, *Lb. casei* group and *Lb. fermentum* occurred
406
407 165 during ripening. A similar trend, i.e. presence in curd and ripening samples, was also
408
409 166 observed for *Propionibacterium* sp. Finally, *Lb. gallinarum* although in a relatively small
410
411 167 amount was detected in all metagenomic specimens. The latter together with *Lb.*
412
413 168 *helveticus*, *Lb. delbrueckii*, *Lactobacillus* sp. and *Streptococcus* sp. comprised the core
414
415 169 microbiota. *Lb. brevis* and *Lb. plantarum* as well as *Lactococcus lactis*, recovered from
416
417 170 whey starter, curd or ripening samples, were incorporating into *Lactobacillus* sp. and
418
419 171 *Lactococcus* sp., respectively, because they were not visible alone in Fig. 4.
420
421 172 *Streptococcus thermophilus* also was detected in most of the samples. Statistical
422
423 173 comparison of the metagenomic samples with the web-based program METAGENassist
424
425 174 (Arndt et al., 2012) revealed the significance of the species *Lb. helveticus*, *Lb.*
426
427 175 *delbrueckii*, *Lb. rhamnosus*, *Lb. casei* group, *Lb. fermentum*, *Streptococcus* sp. and *Str.*
428
429 176 *thermophilus*. These observations highlight the specific role of the *Lactobacillus* species
430
431 177 as well as the role of the non-starter lactic acid bacteria (NSLAB) and other species
432
433 178 during the Grana-type cheese production (Lazzi et al., 2004; Parente and Cogan, 2004;
434
435 179 Rossetti et al., 2008; Rossi et al., 2012). The heat map in Fig. 5 shows the symbiotic
436
437 180 (between species with industrial interest) and antagonistic (between contaminants and
438
439 181 species with industrial interest) interactions that occurred. The web-based program
440
441
442
443
444
445
446
447
448

449
450
451 182 METAGENassist for comparative metagenomics was used to construct the heat map.
452
453 183 Substantial differences between the two taxonomic profiles, the current (with SILVA as
454
455 184 reference database) and the one from Alessandria et al. (2016) (with Greengenes as
456
457 185 reference database), were not observed, yet some discrepancies do exist. In the present
458
459 186 study, *Lb. gallinarum* was found to belong to the core microbiota; and *Lb. rhamnosus*
460
461 187 along with *Propionibacterium* species (other than the contaminant *Propionibacterium*
462
463 188 *acnes* present in the study of Alessandria et al., 2016) were recovered from samples
464
465 189 during ripening. Such differences were expected since different databases (SILVA vs.
466
467 190 Greengenes) were used to perform the taxonomic profiling (Yilmaz et al., 2014;
468
469 191 Balvočiūtė and Huson, 2017).

472
473 192 In the raw milk samples, several contaminants were detected such as *Acidovorax*
474
475 193 sp., *Acinetobacter* sp., *Acinetobacter baumannii/calcoaceticus* group, *Anoxybacillus* sp.,
476
477 194 *Clostridium* sp., *Sphingomonas* sp. and *Staphylococcus* sp. The category “other” of Fig. 4
478
479 195 included other contaminants such as *Pseudomonas* sp., *Enterobacter* sp., *Escherichia-*
480
481 196 *Shigella*, *Rubrobacter* sp., *Bacillus* sp or *Listeria monocytogenes*. The recovery of such
482
483 197 microorganisms from raw milk using NGS platforms has been reported elsewhere as well
484
485 198 (Quigley et al., 2013). Despite the occurrence of several contaminants, these were
486
487 199 decreased gradually due to the antagonistic activity experienced by the rest of microbiota,
488
489 200 especially the one originated from the *Lactobacillus* species (Fig. 5) supporting the
490
491 201 observation made in Fig. 3.

492
493
494 202

495 203 3.2. Submodule 2: Predictive Metagenomics Profiling

496 204 3.2.1. Statistical analysis

497
498
499
500
501
502
503
504

505
506
507
508 205 The OTU abundance table, obtained from the 16S rRNA data, was used as input
509
510 206 for the submodule 2 to presume for metabolic functions. Currently, there are three tools
511
512 207 available for PMP: PICRUSt (Langille et al., 2013), Tax4Fun (Abhauer et al., 2015) and
513
514 208 Piphillin (Iwai et al., 2016). In the present study, the Tax4Fun program performed the
515
516 209 PMP, which works with the SILVA database. The PICRUSt requires the Greengenes
517
518 210 database whereas the Piphillin tool is not obliged to any unique data pre-processing
519
520 211 protocol supporting KEGG and BioCyc as a reference database. The output of the
521
522 212 Tax4Fun is a table with a similar layout to the OTU abundance containing the functional
523
524 213 predictions of KEGG Orthology (KO) or Pathways (ko). Statistical analysis (Kruskal-
525
526 214 Wallis H-test with Tukey-Kramer), using the STAMP v2.1.3 software (Parks and Beiko,
527
528 215 2010; Parks et al., 2014) showed that 1629 KO and 121 ko displayed substantial changes.
529
530 216 A *P*-value lower than 0.05, corrected for multiple tests according to the Benjamini-
531
532 217 Hochberg FDR (False Discovery Rate) procedure, indicated significant differences. PCA
533
534 218 (Principal Component Analysis) plots, made with Past v3.15 software, display the
535
536 219 orientation of the metagenomic samples and the most abundant KEGG Pathways (ko)
537
538 220 (Fig. 6).

541
542 221 The curd and early ripening metagenomic samples were dominated by pathways
543
544 222 associated with carbohydrate metabolism (Fig. 6a). Cheese making (curd) and early
545
546 223 ripening samples were mainly located in the right part of the graph (Fig. 6b). The ko
547
548 224 02060 (phosphotransferase system – PTS; membrane transport), 00564
549
550 225 (glycerophospholipid metabolism; lipid metabolism) and 00260 (glycine, serine,
551
552 226 threonine metabolism; amino acid metabolism) also appeared on the right of the vertical
553
554 227 axis (Fig. 6a). The PTS is a mechanism of the bacteria with which they uptake
555
556
557
558
559
560

561
562
563 228 carbohydrates (Kotrba et al., 2001). Lactobacilli consume sugars such as galactose and
564
565 229 lactose, and the glucose can be converted to pyruvate through glycolysis (Hemme et al.,
566
567 230 1981; Premi et al., 1972). Pyruvate is an important precursor of many metabolites such as
568
569 231 lactic acid, formic acid, acetic acid, acetaldehyde, ethanol, acetoin, diacetyl, and butane-
570
571 232 2,3-diol (Hickey et al., 1983). Moreover, thermophilic lactobacilli such as *Lb. helveticus*
572
573 233 and *Lb. delbrueckii* can produce peptides, amino acids and other metabolites that
574
575 234 stimulate the growth of *Str. thermophilus* (Courtin and Rul, 2004; Hemme et al., 1981)
576
577 235 and propionibacteria (Baer, 1995; Kerjean et al., 2000; Piveteau et al., 1995).
578
579 236 Metatranscriptomics revealed that genes associated with carbohydrate metabolism
580
581 237 (pentose phosphate pathway and glycolysis) were enriched during the cheese making
582
583 238 process of the traditional Italian cheese Caciocavallo Silano PDO (Protected Designation
584
585 239 of Origin) (De Filippis et al., 2016).

589 240 On the top left corner of Fig. 6b, the metagenomic samples of the middle and late
590
591 241 ripening formed a separate group, compared to the other samples. Accordingly, the
592
593 242 KEGG Pathways located in the top left area of the PCA graph (Fig. 6a) were related to
594
595 243 amino acid (ko00280, valine, leucine, isoleucine degradation; ko00360, phenylalanine
596
597 244 metabolism) and lipid (ko00061, fatty acid biosynthesis; ko00071, fatty acid metabolism)
598
599 245 metabolism. Also, pathways referred to carbohydrate metabolism (ko00020, TCA cycle;
600
601 246 ko00640, propanoate metabolism; ko00630 glyoxylate and dicarboxylate metabolism)
602
603 247 were also observed, which may participate in the production of aroma compounds. Flavor
604
605 248 formation in cheeses is a complex process involving proteolytic and lipolytic activities in
606
607 249 which key players are NSLAB and other non lactic acid bacteria (Smit et al., 2005).
608
609 250 Interestingly, samples taken during middle and late ripening of the cheese were
610
611
612
613
614
615
616

617
618
619 251 characterized by the gradual increase of *Lb. rhamnosus*, *Lb. casei*, *Lb. fermentum*, *Str.*
620
621 252 *thermophilus* and *Propionibacterium* sp. (Fig. 4). These microorganisms are known for
622
623 253 their proteolytic and/or lipolytic activity as well as for their ability to produce aroma
624
625 254 compounds (González-Olivares et al., 2014; Hong-Xin et al., 2015; Smit et al., 2005;
626
627 255 Thierry et al., 2011). The above results showed good correlation with the observations
628
629 256 made during ripening of the traditional Italian cheese Caciocavallo Silano PDO using
630
631 257 metatranscriptomics (De Filippis et al., 2016).
632
633
634 258

636 259 3.2.2. Metabolic interactions

638 260 Usually, thermophilic lactic starters, propionibacteria, and NSLAB follow one
639
640 261 another during ripening of Swiss-type cheeses (Gagnaire et al., 2001). A similar trend
641
642 262 was observed in the present study for an Italian Grana-type cheese. Propionibacteria
643
644 263 growth is dependent on the availability of lactate which is produced by *Lb. helveticus*, *Lb.*
645
646 264 *delbrueckii* and *St. thermophilus* (Kurtz et al., 1959). Propionibacteria preferably utilize
647
648 265 lactate as the energy source (Brendehaug and Langsrud, 1985; Fröhlich-Wyder et al.,
649
650 266 2002). Despite the fundamental role of NSLAB in cheese flavor, propionibacteria should
651
652 267 be present as well, but not in excess, to allow Grana-type cheeses such as Grana Padano
653
654 268 and Parmigiano Reggiano develop their typical organoleptic characteristics (Carcano et
655
656 269 al., 1995). The uncontrolled growth of propionibacteria may lead to an undesirable
657
658 270 situation known as “late blowing” or “late fermentation” (Carcano et al., 1995; Fröhlich-
659
660 271 Wyder et al., 2002).
661
662

664 272 So, both micro-flora NSLAB and propionibacteria have a role to play during the
665
666 273 development of the organoleptic characteristics of the Grana-type cheeses. But how do
667
668
669
670
671
672

673
674
675 274 propionibacteria interact with NSLAB? Facultatively heterofermentative lactobacilli
676
677 275 (FHL) such as *Lb. casei* and *Lb. rhamnosus* may compromise propionibacteria growth
678
679 276 (Fröhlich-Wyder et al., 2002; Jimeno et al., 1995), especially when FHL are added as
680
681 277 supplemental cultures. Fröhlich-Wyder et al. (2002) have showed that the addition of
682
683 278 NSLAB in Swiss-type cheeses inhibited lactate fermentation by the propionibacteria. If,
684
685 279 however, NSLAB are naturally occurring during cheese ripening, do they have the same
686
687 280 effect on propionibacteria growth or not? Most NSLAB do not affect propionibacteria
688
689 281 levels in cheese. The influence of *Lactobacillus* spp. on propionibacteria growth is likely
690
691 282 to be less important than the impact of technological parameters such as pH and salt in
692
693 283 cheeses (Carcano et al., 1995; Noël, 1999).

694
695
696 284 Consequently, the question above was explored using the microbial metabolic
697
698 285 interactions (MMinte) tool (Mendes-Soares et al., 2016) for investigating the interplay
699
700 286 between the naturally occurring flora of NSLAB and propionibacteria. Within the
701
702 287 microbial community, the MMinte indicates the nature of that interplay (positive,
703
704 288 negative or no interaction) based on the comparison of growth rates between the pairs of
705
706 289 the microorganisms by constructing predictive genome-scale metabolic models. The
707
708 290 sequence data and the subset of correlations between the considered OTUs, as estimated
709
710 291 by the METAGENassist program (Fig. 5), were introduced to the MMinte tool and were
711
712 292 run through the six widgets available: widget 1, only the representative 16S rDNA/rRNA
713
714 293 sequence data were kept for further analysis based on the provided subset of the
715
716 294 microorganisms pairs; widget 2, a genome ID is assigned to each OTU using BLAST (the
717
718 295 16S rDNA/rRNA sequences are compared with reference sequences available in NCBI)
719
720 296 (Altschul et al., 1990); widget 3, a predictive metabolic model is constructed using
721
722
723
724
725
726
727
728

729
730
731 297 ModelSEED (Henry et al., 2010) for each genome ID; widget 4, a predictive two-species
732
733 298 community metabolic model is created by the mean of COBRAPy (Ebrahim et al., 2013;
734
735 299 Klitgord and Segrè, 2010) for each pair of microbes provided in the first widget; widget
736
737 300 5, predictions on the growth rates are made using flux balance analysis (Heinken and
738
739 301 Thiele, 2015; Varma and Palsson, 1994) for each two-species community; widget 6, the
740
741 302 metabolic network of the microbial community is drawn and the nature of interactions is
742
743 303 indicated using the D3.js visualization tool (Bostock et al., 2011).
744
745

746 304 The network in Fig. 7 depicts the predicted interplays between the
747
748 305 propionibacteria and NSLAB. When the availability of the metabolites is high
749
750 306 (“Complete_100”), propionibacteria is predicted to grow thus take a benefit from the
751
752 307 presence of the NSLAB. NSLAB are heterofermentative microorganisms producing
753
754 308 lactate among others, which can be used by propionibacteria. *Lb. casei* converts glucose,
755
756 309 especially when glucose is limited, to lactate (predominantly), acetate, formate, and
757
758 310 ethanol (Liu, 2003). Interestingly, when the availability of the metabolites was reduced
759
760 311 by ten times (“Complete_10”) there was an increase in the number of negative
761
762 312 interactions predicted to occur between propionibacteria and NSLAB, without a positive
763
764 313 interplay between them, meaning that NSLAB impairs the growth of propionibacteria
765
766 314 when there is higher competition for nutrients. Consequently, the nature of the interaction
767
768 315 is altered based on the metabolites availability. This probably explains partially the
769
770 316 observation that NSLAB inhibit the propionibacteria growth when added as supplement
771
772 317 cultures. This generates a much greater competition by the NSLAB for nutrients,
773
774 318 resulting in the accumulation of elevated quantities (excess) of metabolic products such
775
776
777
778
779
780
781
782
783
784

785
786
787
788 319 as acetate, formate, and diacetyl that suspend the increase of propionibacteria (Jimeno et
789
790 320 al., 1995).

791
792 321

793 322 **4. Conclusion**

796 323 The bioinformatics pipeline described in the present study may also find an
797
798 324 application to foodborne pathogens occurring in foodstuffs. As predictive microbiology
799
800 325 enters a new era of the integration of meta- and multi- omics in predictive modeling and
801
802 326 quantitative risk assessment in foods (Brul et al., 2012; Cocolin et al., 2017; Rantsiou et
803
804 327 al., 2011), the workflow proposed here may constitute a useful tool. For instance, it can
805
806 328 respond to questions that concern risk assessors, food microbiologists and others dealing
807
808 329 with microbiological risk assessment studies: How the foodborne pathogens found in
809
810 330 food interact with the rest of microbiota? Why strains of the same species behave
811
812 331 differently? How environmental conditions influence important features of the foodborne
813
814 332 pathogens such as virulence?

817 333 As any novel method, PMP also has constraints. In order, the method to make
818
819 334 reliable and accurate predictions about the gene content of an OTU, the genome of
820
821 335 reference or at least closely related microorganisms should be sequenced and available.
822
823 336 Despite this limitation, PMP is a cost-effective and straightforward method to start with
824
825 337 when 16S rDNA/rRNA data are available (Wood, 2016). The tool may help to investigate
826
827 338 the functional potential, alterations, and interactions of a microbial community. Thus, it
828
829 339 will provide evidence for further exploration of the community and guide future
830
831 340 experiments based on the genes or gene groups predicted to change (Wood, 2016).
832
833
834
835
836
837
838
839
840

841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896

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

343

897
898
899 **344 References**
900

901 345 Aßhauer, K.P., Wemheuer, B., Daniel, R., Meinicke, P., 2015. Tax4Fun: Predicting
902
903 346 functional profiles from metagenomic 16S rRNA data. *Bioinformatics* 31, 2882-
904
905
906 347 2884.

907
908 348 Alessandria, V., Ferrocino, I., De Filippis, F., Fontana, M., Rantsiou, K., Ercolini, D.,
909
910 349 Cocolin, L., 2016. Microbiota of an Italian Grana-like cheese during manufacture
911
912 350 and ripening, unraveled by 16S rRNA-based approaches. *Appl. Environ.*
913
914 351 *Microbiol.* 82, 3988-3995.

915
916 352 Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local
917
918 353 alignment search tool. *J. Mol. Biol.* 215, 403-410.

919
920 354 Arndt, D., Xia, J., Liu, Y., Zhou, Y., Guo, A.C., Cruz, J.A., Sinelnikov, I., Budwill, K.,
921
922 355 Nesbø, C.L., Wishart, D.S., 2012. METAGENassist: a comprehensive web server
923
924 356 for comparative metagenomics. *Nucleic Acids Res.* 40, W88-W95.

925
926 357 Baer, A., 1995. Influence of casein proteolysis by starter bacteria, rennet and plasmin on
927
928 358 the growth of propionibacteria in Swiss-type cheese. *Lait* 75, 391-400.

929
930 359 Balvočiūtė, M., Huson, D.H., 2017. SILVA, RDP, Greengenes, NCBI and OTT – how do
931
932 360 these taxonomies compare? *BMC Genomics* 18 (Suppl 2), 114.

933
934 361 Bostock, M., Ogievetsky, V., Heer, J. 2011. D³: Data-Driven Documents. *IEEE Trans.*
935
936 362 *Vis. Comput. Graph.* 17, 2301-2309.

937
938 363 Brendehaug, J., Langsrud, T., 1985. Amino acid metabolism in propionibacteria: resting
939
940 364 cells experiments with four strains. *J. Dairy Sci.* 68, 281–289.
941
942
943
944
945
946
947
948
949
950
951
952

- 953
954
955 365 Brul, S., Bassett, J., Cook, P., Kathariou, S., McClure, P., Jasti, P.R., Betts, R., 2012.
956
957 366 'Omics' technologies in quantitative microbial risk assessment. Trends Food Sci.
958
959 367 Technol. 27, 12-24.
960
961 368 Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden,
962
963 369 T., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421.
964
965 370 Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.
966
967 371 K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley,
968
969 372 S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D.,
970
971 373 Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters,
972
973 374 W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R., 2010. QIIME
974
975 375 allows analysis of high-throughput community sequencing data. Nat. Methods 7,
976
977 376 335-336.
978
979 377 Carcano, M., Todesco, R., Lodi, R., Brasca, M., 1995. Propionibacteria in Italian hard
980
981 378 cheeses. Lait 75, 415-426.
982
983 379 Cocolin, L., Mataragas, M., Bourdichon, F., Doulgeraki, A., Pilet, M.-F., Jagadeesan, B.,
984
985 380 Rantsiou, K., Phister, T., 2017. Meta-Omics and Microbial Risk Assessment: the
986
987 381 next need for integration. Int. J. Food Microbiol. (under revision).
988
989 382 Coughlan, L.M., Cotter, P.D., Hill, C., Alvarez-Ordóñez, A., 2015. Biotechnological
990
991 383 applications of functional metagenomics in the food and pharmaceutical
992
993 384 industries. Front. Microbiol. 6, 672.
994
995 385 Courtin, P., Rul, F., 2004. Interactions between microorganisms in a simple ecosystem:
996
997 386 yogurt bacteria as a study model. Lait 84, 125-134.
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008

- 1009
1010
1011 387 De Filippis, F., Genovese, A., Ferranti, P., Gilbert, J.A., Ercolini, D., 2016.
1012
1013 388 Metatranscriptomics reveals temperature-driven functional changes in
1014
1015 389 microbiome impacting cheese maturation rate. *Sci. Rep.* 6, 21871.
1016
1017
1018 390 De Filippo, C., Ramazzotti, M., Fontana, P., Cavalieri, D., 2012. *Bioinformatics*
1019
1020 391 approaches and pathway inference in metagenomics data. *Brief. Bioinformatics*
1021
1022 392 13, 696-710.
1023
1024 393 Delcenserie, V., Taminiau, B., Delhalle, L., Nezer, C., Doyen, P., Crevecoeur, S.,
1025
1026 394 Roussey, D., Korsak, N., Daube, G., 2014. Microbiota characterization of a
1027
1028 395 Belgian protected designation of origin cheese, Herve cheese, using metagenomic
1029
1030 396 analysis. *J. Dairy Sci.* 97, 6046–6056.
1031
1032
1033 397 Dhariwal, A., Chong, J., Habib, S., King, I.L., Agellon, L.B., Xia, J., 2017.
1034
1035 398 MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and
1036
1037 399 meta-analysis of microbiome data. *Nucleic Acids Res.* 45, W180-W188.
1038
1039 400 Dudhagara, P., Bhavsar, S., Bhagat, C., Ghelani, A., Bhatt, S., Patel, R., 2015. *Web*
1040
1041 401 *Resources for Metagenomics Studies. Genomics Proteomics Bioinformatics* 13,
1042
1043 402 296-303.
1044
1045 403 Ebrahim, A., Lerman, J.A., Palsson, B.O., Hyduke, D.R., 2013. *COBRAPy: CONstraints-*
1046
1047 404 *Based Reconstruction and Analysis for Python. BMC Syst. Biol.* 7, 74.
1048
1049 405 Ercolini, D., De Filippis, F., La Stora, A., Iacono, M., 2012. “Remake” by High-
1050
1051 406 Throughput Sequencing of the Microbiota Involved in the Production of Water
1052
1053 407 Buffalo Mozzarella Cheese. *Appl. Environ. Microbiol.* 78, 8142-8145.
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064

- 1065
1066
1067
1068 408 Escobar-Zepeda, A., Vera-Ponce de Leon, A., Sanchez-Flores, A., 2015. The Road to
1069
1070 409 Metagenomics: From Microbiology to DNA Sequencing Technologies and
1071
1072 410 Bioinformatics. *Front. Genet.* 6, 348.
- 1073
1074 411 Ferrocino, I., Cocolin, L., 2017. Current perspectives in food-based studies exploiting
1075
1076 412 multi-omics approaches. *Curr. Opin. Food Sci.* 13, 10-15.
- 1077
1078 413 Ferrocino, I., Greppi, G., La Stora, A., Rantsiou, K., Ercolini, D., Cocolin, L., 2016.
1079
1080 414 Impact of Nisin-Activated Packaging on Microbiota of Beef Burgers during
1081
1082 415 Storage. *Appl. Environ. Microbiol.* 82, 549-559.
- 1083
1084 416 Fröhlich-Wyder, M.-T., Bachmann, H.-P., Casey, M.G., 2002. Interaction between
1085
1086 417 propionibacteria and starter / non-starter lactic acid bacteria in Swiss-type
1087
1088 418 cheeses. *Lait* 82, 1-15.
- 1089
1090 419 Gagnaire, V., Molle, D., Herrouin, M., Leonil, J., 2001. Peptides identified during
1091
1092 420 Emmental cheese ripening: Origin and proteolytic systems involved. *J. Agric.*
1093
1094 421 *Food Chem.* 49, 4402-4413.
- 1095
1096 422 González-Olivares, L.G., López-Cuellar, Z.L., Añorve-Morga, J., Franco-Fernández,
1097
1098 423 M.J., Castañeda-Ovando, A., Contreras-López, E., Jaimez-Ordaz, J., Rodríguez-
1099
1100 424 Serrano, G.M., 2014. Viability and Proteolytic Capacity of *Lactobacillus*
1101
1102 425 *bulgaricus* 2772 and *Lactobacillus rhamnosus* GG during Cheese Ripening. *J.*
1103
1104 426 *Biosci. Med.* 2, 7-12.
- 1105
1106 427 Hammer, Ø., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological statistics
1107
1108 428 software package for education and data analysis. *Palaeontol. Electron.* 4, 9.
- 1109
1110 429 Heinken, A., Thiele, I., 2015. Anoxic conditions promote species-specific mutualism
1111
1112 430 between gut microbes *in silico*. *Appl. Environ. Microbiol.* 81, 4049-4061.
- 1113
1114
1115
1116
1117
1118
1119
1120

- 1121
1122
1123 431 Hemme, D.H., Schmal, V., Auclair, J.E., 1981. Effect of the addition of extracts of
1124
1125 432 thermophilic lactobacilli on acid production by *Streptococcus thermophilus* in
1126
1127 433 milk. J. Dairy Res. 48, 139-148.
1128
1129 434 Henry, C.S., De Jongh, M., Best, A.A., Frybarger, P.M., Linsay, B., Stevens, R.L., 2010.
1130
1131 435 High-throughput generation, optimization and analysis of genome-scale metabolic
1132
1133 436 models. Nat. Biotechnol. 28, 977-982.
1134
1135 437 Hickey, M.W., Hillier, A.J., Jago, G.R., 1983. Metabolism of Pyruvate and Citrate in
1136
1137 438 Lactobacilli. Aust. J. Biol. Sci. 36, 487-496.
1138
1139 439 Hong-Xin, J., Mi-Ya, S., Guang-Yu, G., 2015. Influence of *Lactobacillus casei* LC2W on
1140
1141 440 the proteolysis and aroma compounds of Cheddar cheese during ripening period.
1142
1143 441 CyTA – J. Food 13, 464-471.
1144
1145 442 Iwai, S., Weinmaier, T., Schmidt, B.L., Albertson, D.G., Poloso, N.J., Dabbagh, K.,
1146
1147 443 DeSantis, T.Z., 2016. Piphillin: Improved Prediction of Metagenomic Content by
1148
1149 444 Direct Inference from Human Microbiomes. PLoS One 11, e0166104.
1150
1151 445 Jimeno, J., Lazaro, M.J., Sollberger, H., 1995. Antagonistic interactions between
1152
1153 446 propionic acid bacteria and non-starter lactic acid bacteria. Lait 75, 401-413.
1154
1155 447 Keller, A., Horn, H., Forster, F., Schultz, J., 2014. Computational integration of genomic
1156
1157 448 traits into 16S rDNA microbiota sequencing studies. Gene 549, 186–191.
1158
1159 449 Kerjean, J.R., Condon, S., Lodi, R., Kalantzopoulos, G., Chamba, J.F., Suomalainen, T.,
1160
1161 450 Cogan, T., Moreau, D., 2000. Improving the quality of European hard-cheeses by
1162
1163 451 controlling of interactions between lactic acid bacteria and propionibacteria. Food
1164
1165 452 Res. Int. 33, 281–287.
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176

- 1177
1178
1179 453 Klitgord, N., Segrè D., 2010. Environments that induce synthetic microbial ecosystems.
1180
1181 454 PLoS Comput. Biol. 6, e1001002.
1182
1183 455 Kotrba, P., Inui, M., Yukawa, H., 2001. Bacterial phosphotransferase system (PTS) in
1184
1185 carbohydrate uptake and control of carbon metabolism. J. Biosci. Bioeng. 92,
1186 456
1187 502-517.
1188 457
1189
1190 458 Kuntal, B.K., Ghosh, T.S., Mande, S.S., 2013. Community-Analyzer: A platform for
1191
1192 visualizing and comparing microbial community structure across microbiomes.
1193
1194 460 Genomics 102, 409-418.
1195
1196 461 Kurtz, F.E., Hupfer, J.A., Corbin, E.A., Hargrove, R.E., Walter, H.E., 1959.
1197
1198 462 Interrelationships between pH, populations of *Propionibacterium shermanii*,
1199
1200 levels of free fatty acids, and the flavor ratings of Swiss cheeses. J. Dairy Sci. 42,
1201 463
1202 1008-1019.
1203 464
1204
1205 465 Ladoukakis, E., Kolisis, F.N., Chatziioannou, A.A., 2014. Integrative workflows for
1206
1207 metagenomic analysis. Front. Cell Dev. Biol. 2, 70.
1208
1209 467 Langille, M.G.I., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J.A.,
1210
1211 Clemente, J.C., Burkepille, D.E., Vega Thurber, R.L., Knight, R., Beiko, R.G.,
1212
1213 469 Huttenhower, C., 2013. Predictive functional profiling of microbial communities
1214
1215 using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814-821.
1216 470
1217
1218 471 Lazzi, C., Rossetti, L., Zago, M., Neviani, E., Giraffa, G., 2004. Evaluation of bacterial
1219
1220 472 communities belonging to natural whey starters for Grana Padano cheese by
1221
1222 473 length heterogeneity-PCR. J. Appl. Microbiol. 96, 481-490.
1223
1224 474 Li, W., Godzik, A., 2006. Cd-hit: a fast program for clustering and comparing large sets
1225
1226 475 of protein or nucleotide sequences. Bioinformatics 22, 1658-1659.
1227
1228
1229
1230
1231
1232

- 1233
1234
1235 476 Liu, S.Q., 2003. Practical implications of lactate and pyruvate metabolism by lactic acid
1236
1237 bacteria in food and beverage fermentations. *Int. J. Food Microbiol.* 83, 115-131.
1238 477
1239 478 Liu, W., Zheng, Y., Kwok, L.-Y., Sun, Z., Zhang, J., Guo, Z., Hou, Q., Menhe, B.,
1240
1241 Zhang, H., 2015. High-throughput sequencing for the detection of the bacterial
1242 479
1243 and fungal diversity in Mongolian naturally fermented cow's milk in Russia.
1244 480
1245 *BMC Microbiol.* 15, 45.
1246 481
1247
1248 482 Mendes-Soares, H., Mundy, M., Mendes-Soares, L., Chia, N., 2016. MMinte: an
1249
1250 483 application for predicting metabolic interactions among the microbial species in a
1251
1252 484 community. *BMC Bioinformatics* 17, 343.
1253
1254 485 Minot, S.S., Krumm, N., Greenfield, N.B., 2015. One Codex: A Sensitive and Accurate
1255
1256 486 Data Platform for Genomic Microbial Identification. *bioRxiv*, doi:
1257
1258 487 <http://dx.doi.org/10.1101/027607>.
1259
1260
1261 488 Noël, Y., Poyoval P., Thierry A., Gagnaire V., Grappin, R., 1999. Eye formation and
1262
1263 489 Swiss-type cheeses, In: Law, B.A. (Ed.), *Technology of cheese making*. CRC
1264
1265 490 Press, Boca Raton, pp. 222-250.
1266
1267 491 Oulas, A., Pavludi, C., Polymenakou, P., Pavlopoulos, G.A., Papanikolaou, N.,
1268
1269 492 Kotoulas, G., Arvanitidis, C., Iliopoulos, I., 2015. Metagenomics: Tools and
1270
1271 493 Insights for Analyzing Next-Generation Sequencing Data Derived from
1272
1273 494 Biodiversity Studies. *Bioinform. Biol. Insights* 9, 75-88.
1274
1275
1276 495 Parente, E., Cogan, T.M., 2004. Starter cultures: general aspects. In: Fox, P.F., et al.
1277
1278 496 (Eds.), *Cheese – Chemistry, Physics, and Microbiology*. Elsevier Academic Press,
1279
1280 497 London, pp. 123-148.
1281
1282
1283
1284
1285
1286
1287
1288

- 1289
1290
1291 498 Parks, D.H., Beiko, R.G., 2010. Identifying biologically relevant differences between
1292
1293 499 metagenomic communities. *Bioinformatics* 26, 715-721.
1294
1295 500 Parks, D.H., Tyson, G.W., Hugenholtz, P., Beiko, R.G., 2014. STAMP: statistical
1296
1297 501 analysis of taxonomic and functional profiles. *Bioinformatics* 30, 3123-3124.
1298
1299 502 Parlapani, F.F., Meziti, A., Kormas, K.Ar., Boziaris, I.S., 2013. Indigenous and spoilage
1300
1301 503 microbiota of farmed sea bream stored in ice identified by phenotypic and 16S
1302
1303 504 rRNA gene analysis. *Food Microbiol.* 33, 85-89.
1304
1305 505 Parlapani, F.F., Kormas, K.Ar., Boziaris, I.S., 2015. Microbiological changes, shelf life
1306
1307 506 and identification of initial and spoilage microbiota of sea bream fillets stored
1308
1309 507 under various conditions using 16S rRNA gene analysis. *J. Sci. Food Agric.* 95,
1310
1311 508 2386–2394.
1312
1313 509 Parlapani, F.F., Boziaris, I.S., 2016. Monitoring of spoilage and determination of
1314
1315 510 microbial communities based on 16S rRNA gene sequence analysis of whole sea
1316
1317 511 bream stored at various temperatures. *LWT - Food Sci. Technol.* 66, 553-559.
1318
1319 512 Piveteau, P.G., Condon S., Cogan T.M., 1995. Interactions between lactic and propionic
1320
1321 513 acid bacteria. *Lait* 75, 331-343.
1322
1323 514 Połka, J., Rebecchi, A., Pisacane, V., Morelli, L., Puglisi, E., 2015. Bacterial diversity in
1324
1325 515 typical Italian salami at different ripening stages as revealed by high-throughput
1326
1327 516 sequencing of 16S rRNA amplicons. *Food Microbiol.* 46, 342-356.
1328
1329 517 Pothakos, V., Stellato, G., Ercolini, D., Devlieghere, F., 2015. Processing Environment
1330
1331 518 and Ingredients Are Both Sources of *Leuconostoc gelidum*, Which Emerges as a
1332
1333 519 Major Spoiler in Ready-To-Eat Meals. *Appl. Environ. Microbiol.* 81, 3529-3541.
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344

- 1345
1346
1347 520 Premi, L., Sandine, W.E., Elliker, P.R., 1972. Lactose-hydrolyzing enzymes of
1348
1349 521 *Lactobacillus* species. Appl. Microbiol. 24, 51-57.
1350
1351 522 Pruesse, E., Peplies, J., Glöckner, F.O., 2012. SINA: accurate high-throughput multiple
1352
1353 523 sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823-1829.
1354
1355 524 Pylro, V.S., Morais, D.K., de Oliveira, F.S., dos Santos, F.G., Lemos, L.N., Oliveira, G.,
1356
1357 Roesch, L.F.W., 2016. BMPOS: a Flexible and User-Friendly Tool Sets for
1358 525
1359 Microbiome Studies. Microb. Ecol. 72, 443-447.
1360 526
1361 527 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J.,
1362
1363 528 Glöckner, F.O., 2013. The SILVA ribosomal RNA gene database project:
1364
1365 529 improved data processing and web-based tools. Nucleic Acids Res. 41, D590-
1366
1367 530 D596.
1368
1369 531 Quigley, L., O'Sullivan, O., Stanton, C., Beresford, T.P., Ross, R.P., Fitzgerald, G.F.,
1370
1371 532 Cotter, P.D., 2013. The complex microbiota of raw milk. FEMS Microbiol. Rev.
1372
1373 533 37, 664-698.
1374
1375 534 Rantsiou, K., Mataragas, M., Jespersen, L., Cocolin, L., 2011. Understanding the
1376
1377 535 behavior of foodborne pathogens in the food chain: New information for risk
1378
1379 536 assessment analysis. Trends Food Sci. Technol. 22, S21-S29.
1380
1381 537 Robertson, C.E., Harris, J.K., Wagner, B.D, Granger, D., Browne, K., Tatem, B., Feazel,
1382
1383 538 L.M., Park, K., Pace1, N.R., Frank, D.N., 2013. Explicet: graphical user interface
1384
1385 539 software for metadata-driven management, analysis and visualization of
1386
1387 540 microbiome data. Bioinformatics 29, 3100-3101.
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400

- 1401
1402
1403 541 Rossetti, L., Fornasari, M.E., Gatti, M., Lazzi, C., Neviani, E., Giragga, G., 2008. Grana
1404
1405 542 Padano cheese whey starters: Microbial composition and strain distribution. *Int. J.*
1406
1407
1408 543 *Food Microbiol.* 127, 168-171.
- 1409
1410 544 Rossi, F., Gatto, V., Sabattini, G., Torriani, S., 2012. An assessment of factors
1411
1412 545 characterising the microbiology of Grana Trentino cheese, a Grana-type cheese.
1413
1414 546 *Int. J. Dairy Technol.* 65, 1-9.
- 1415
1416 547 Roumpeka, D.D., Wallace, R.J., Escalettes, F., Fotheringham, I., Watson, M., 2017. A
1417
1418 548 review of bioinformatics tools for bio-prospecting from metagenomic sequence
1419
1420 549 data. *Front. Genet.* 8, 23.
- 1421
1422 550 Sattin, E., Andreani, N.A., Carraro, L., Lucchini, R., Fasolato, L., Telatin, A., Balzan, S.,
1423
1424 551 Novelli, E., Simionati, B., Cardazzo, B., 2016. A Multi-Omics Approach to
1425
1426 552 Evaluate the Quality of Milk Whey Used in Ricotta Cheese Production. *Front.*
1427
1428 553 *Microbiol.* 7, 1272.
- 1429
1430
1431 554 Scholz, M.B., Lo, C.-C., Chain, P.S.G., 2012. Next generation sequencing and
1432
1433 555 bioinformatics bottlenecks: the current state of metagenomic data analysis. *Curr.*
1434
1435 556 *Opin. Biotechnol.* 23, 9-15.
- 1436
1437 557 Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. *Front.*
1438
1439 558 *Plant Sci.* 5, 209.
- 1440
1441 559 Smit, G., Smit, B.A., Engels, W.J.M., 2005. Flavour formation by lactic acid bacteria and
1442
1443 560 biochemical flavour profiling of cheese products. *FEMS Microbiol. Rev.* 29, 591-
1444
1445 561 610.
- 1446
1447
1448 562 Stellato, G., La Storia, A., De Filippis, F., Borriello, G., Villani, F., Ercolini, D., 2016.
1449
1450 563 Overlap of Spoilage-Associated Microbiota between Meat and the Meat

- 1457
1458
1459 564 Processing Environment in Small-Scale and Large-Scale Retail Distributions.
1460
1461 565 Appl. Environ. Microbiol. 82, 4045-4054.
1462
1463 566 Thierry, A., Deutsch, S.-M., Falentin, H., Dalmasso, M., Fabien J. Cousin, F.J., Jan, G.,
1464
1465 567 2011. New insights into physiology and metabolism of *Propionibacterium*
1466
1467 568 *freudenreichii*. Int. J. Food Microbiol. 149, 19-27.
1468
1469
1470 569 Varma, A., Palsson, B., 1994. Stoichiometric flux balance models quantitatively predict
1471
1472 570 growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110.
1473
1474 571 Appl. Environ. Microbiol. 60, 3724–3731.
1475
1476 572 Wood, J., 2016. Predictive metagenomics profiling: why, what and how? Bioinformatics
1477
1478 573 Rev. 2, 1-4. The article is available at the URL:
1479
1480 574 [https://bioinformaticsreview.com/20160320/predictive-metagenomics-profiling-](https://bioinformaticsreview.com/20160320/predictive-metagenomics-profiling-why-what-and-how/)
1481
1482 575 [why-what-and-how/](https://bioinformaticsreview.com/20160320/predictive-metagenomics-profiling-why-what-and-how/)
1483
1484
1485 576 Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T.,
1486
1487 577 Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and “All-species
1488
1489 578 Living Tree Project (LTP)” taxonomic frameworks. Nucleic Acids Res. 42,
1490
1491 579 D643-D648.
1492
1493 580 Yoon, S.-H., Ha, S.M., Kwon, S., Lim, J., Kim, Y., Seo, H., Chun, J., 2017. Introducing
1494
1495 581 EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and
1496
1497 582 whole-genome assemblies. Int. J. Syst. Evol. Microbiol. 67, 1613-1617.
1498
1499
1500 583

1513
1514
1515 **584 Figure legends**
1516

1517 **585 Fig. 1.** Proposed bioinformatics pipeline for analysis of NGS-based 16S rDNA/rRNA
1518
1519 **586** sequencing data derived from food metagenomics integrating both taxonomic and
1520
1521 **587** functional profiling. Solid lines show the workflow of the analysis pipeline. Dashed lines
1522
1523 **588** indicate the two steps interfering with the analysis workflow. The quantification step
1524
1525 **589** includes two submodules.

1526
1527
1528 **590 Fig. 2.** Box-plots of the ANOSIM statistical test for the microbial communities of the
1529
1530 **591** batch D (Group 1), E (Group 2) and F (Group 3).
1531

1532 **592 Fig. 3.** Overview of the microbial community at the genus level (green boxes) found in
1533
1534 **593** the Grana-type cheese samples (blue boxes). WS, whey starter; RM, raw milk; MS, raw
1535
1536 **594** milk and whey starter; CAC, curd after cutting; CAH, curd after heating; CAP, curd after
1537
1538 **595** pressing; CASR, curd after storage room; CHAS, cheese after salting; CH2RM, cheese
1539
1540 **596** after two months of ripening; CH4RM, cheese after four months of ripening; CH6RM,
1541
1542 **597** cheese after six months of ripening; CH8RM, cheese after eight months of ripening; and
1543
1544 **598** CH10RM, cheese after ten months of ripening.

1545
1546
1547 **599 Fig. 4.** Stacked bars chart depicting the percentage of taxa in each metagenomic sample
1548
1549 **600** and the changes occurred between samples. WS, whey starter; RM, raw milk; MS, raw
1550
1551 **601** milk and whey starter; Curd, all curd samples (curd after cutting, curd after heating, curd
1552
1553 **602** after pressing and curd after storage room); ER, early ripening samples (cheese after
1554
1555 **603** salting and cheese after two months of ripening); MR, middle ripening samples (cheese
1556
1557 **604** after four and six months of ripening); and LR, late ripening samples (cheese after eight
1558
1559 **605** and ten months of ripening).
1560

1561
1562 **606 Fig. 5.** Heat map of the correlation matrix between the taxa.
1563
1564
1565
1566
1567
1568

1569
1570
1571
1572 607 **Fig. 6.** The orientation of the a) most abundant KEGG Pathways (ko) and b)
1573
1574 608 metagenomic samples using Principal Component Analysis (PCA). Two principal
1575
1576 609 components (1 and 2) were extracted based on the total variance explained. The
1577
1578 610 percentage shows the variance explained by each particular linear component. Upper-
1579
1580 611 right quadrant has higher readings than points in the lower-left quadrant. Colors indicate
1581
1582 612 KEGG Pathways (ko) related with specific metabolism or function such as carbohydrate
1583
1584 613 metabolism (green), lipid metabolism (red), amino acid metabolism (blue), metabolism of
1585
1586 614 cofactors and vitamins (gray), xenobiotics biodegradation and metabolism (salmon), and
1587
1588 615 membrane transport (orange). Phosphotransferase system – PTS (ko02060),
1589
1590 616 Glycolysis/Gluconeogenesis (ko00010), Galactose metabolism (ko00052), Starch and
1591
1592 617 sucrose metabolism (ko00500), Pentose phosphate pathway (ko00030),
1593
1594 618 Glycerophospholipid metabolism (ko00564), Glycine, serine, threonine metabolism
1595
1596 619 (ko00260), Pyruvate metabolism (ko00620), Valine, leucine, isoleucine degradation
1597
1598 620 (ko00280), Phenylalanine metabolism (ko00360), Fatty acid biosynthesis (ko00061),
1600
1601 621 Fatty acid metabolism (ko00071), Benzoate metabolism (ko00362), Aminobenzoate
1602
1603 622 metabolism (ko00627), Folate biosynthesis (ko00790), Ascorbate metabolism (ko00053),
1604
1605 623 Glyoxylate and dicarboxylate metabolism (ko00630), TCA (citrate) cycle (ko00020) and
1606
1607 624 Propanoate metabolism (ko00640). WS (whey starter – khaki), RM (raw milk – gold),
1608
1609 625 MW (raw milk plus whey starter – salmon), Curd (cheese making – green), ER (early
1610
1611 626 ripening – blue), MR (middle ripening – orange), LR (late ripening – red) are the group
1612
1613 627 names of the metagenomic samples.

1616 628 **Fig. 7.** Metabolic interaction network between propionibacteria and NSLAB;
1617
1618 629 “Complete_100” and “Complete_10” indicate the availability of the metabolites to the
1619
1620
1621
1622
1623
1624

1625
1626
1627 630 microbial community; “Complete_100”, over 400 metabolites are available, with a flux
1628
1629 631 for the import reactions of 100 mmol/gDW/h (high availability); “Complete_10”, the
1630
1631 632 same metabolites but with reaction fluxes of 10 mmol/gDW/h (ten-times lower
1632
1633 633 availability); The percentage next to the microorganisms indicates the similarity between
1634
1635 634 the OTU sequence provided and the reference sequence (genome ID assigned); The
1636
1637 635 thickness and the length of the links inside the network imitate the correlation values
1638
1639 636 provided. When the correlation value increases, the line is becoming thicker and shorter.
1640
1641
1642 637
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680

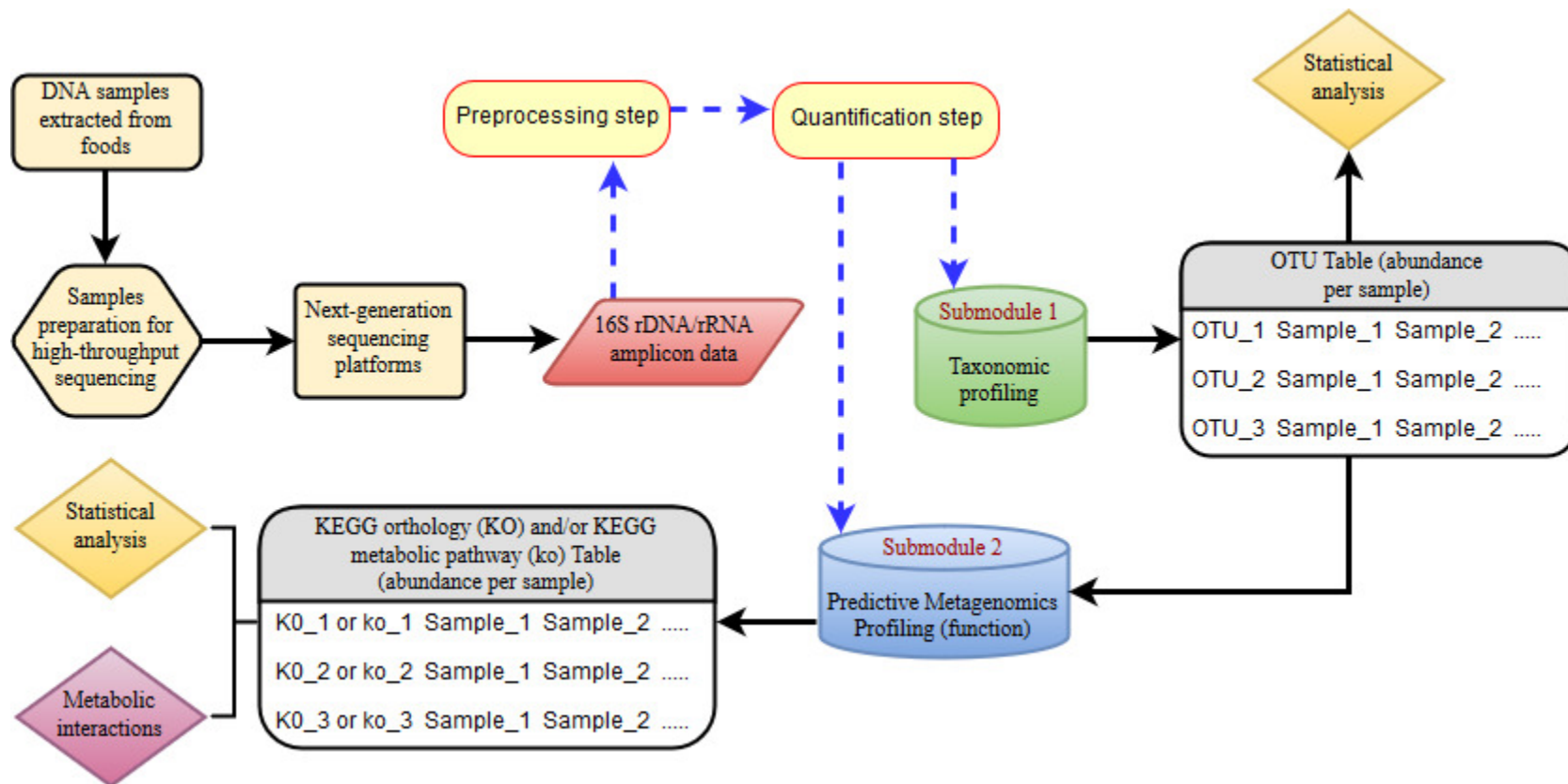
638 **Table 1.** Software and statistical tests used in each stage of the pipeline. Alternative
 639 software is also proposed.

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	–

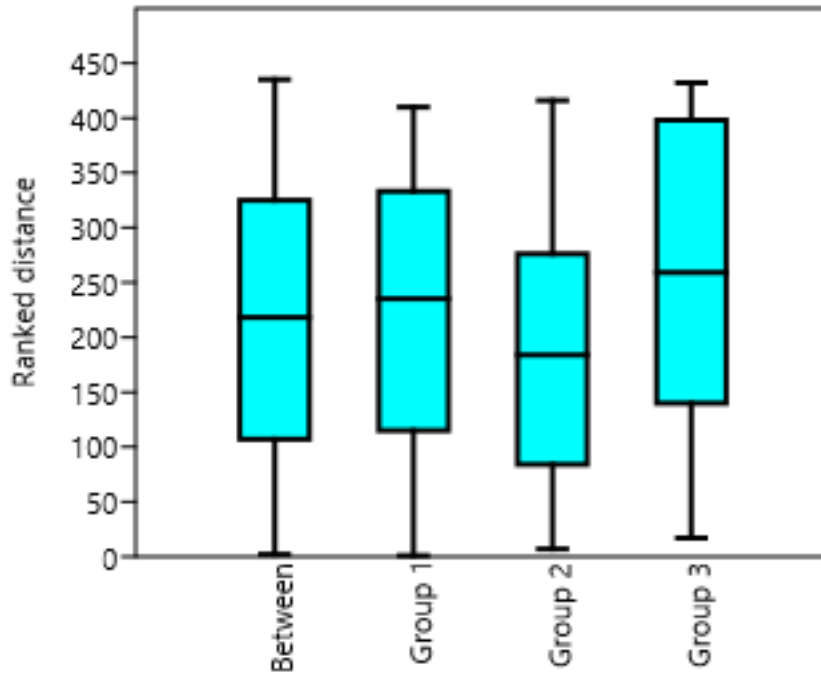
640

1737
 1738
 1739
 1740
 1741
 1742
 1743
 1744
 1745
 1746
 1747
 1748
 1749
 1750
 1751
 1752
 1753
 1754
 1755
 1756
 1757
 1758
 1759
 1760
 1761
 1762
 1763
 1764
 1765
 1766
 1767
 1768
 1769
 1770
 1771
 1772
 1773
 1774
 1775
 1776
 1777
 1778

Fig. 1

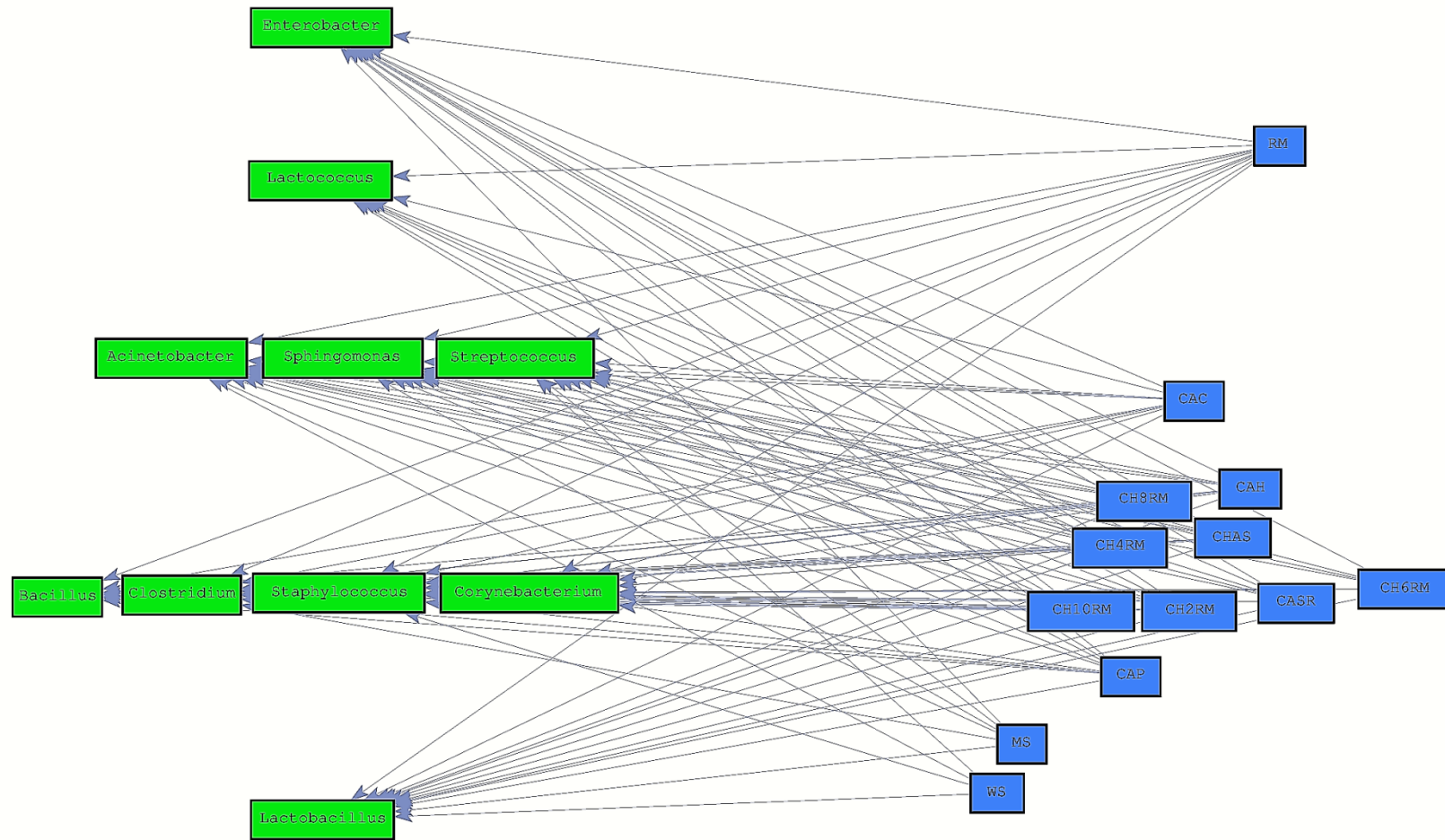


1779
1780
1781 **Fig. 2**
1782
1783
1784
1785
1786
1787
1788



1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876

Fig. 3



1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918

Fig. 4

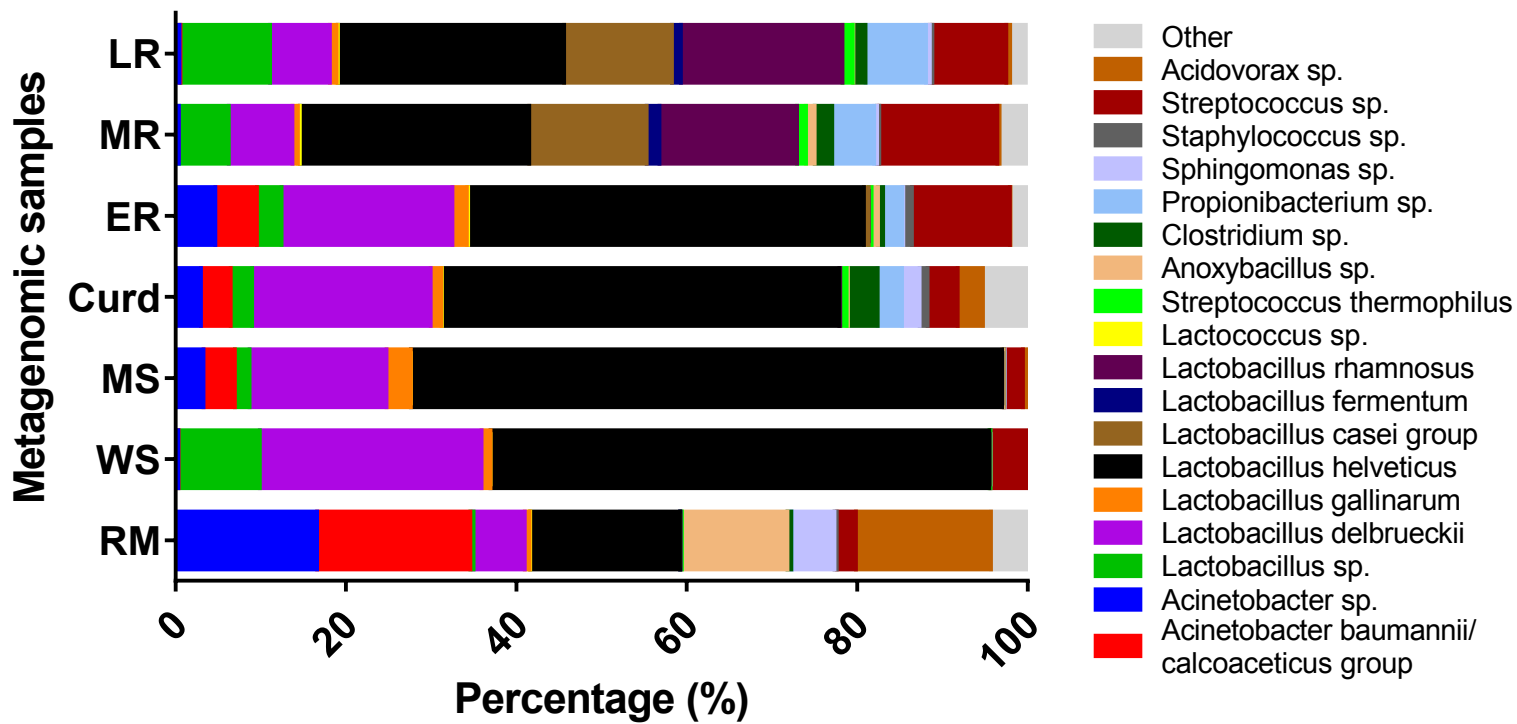
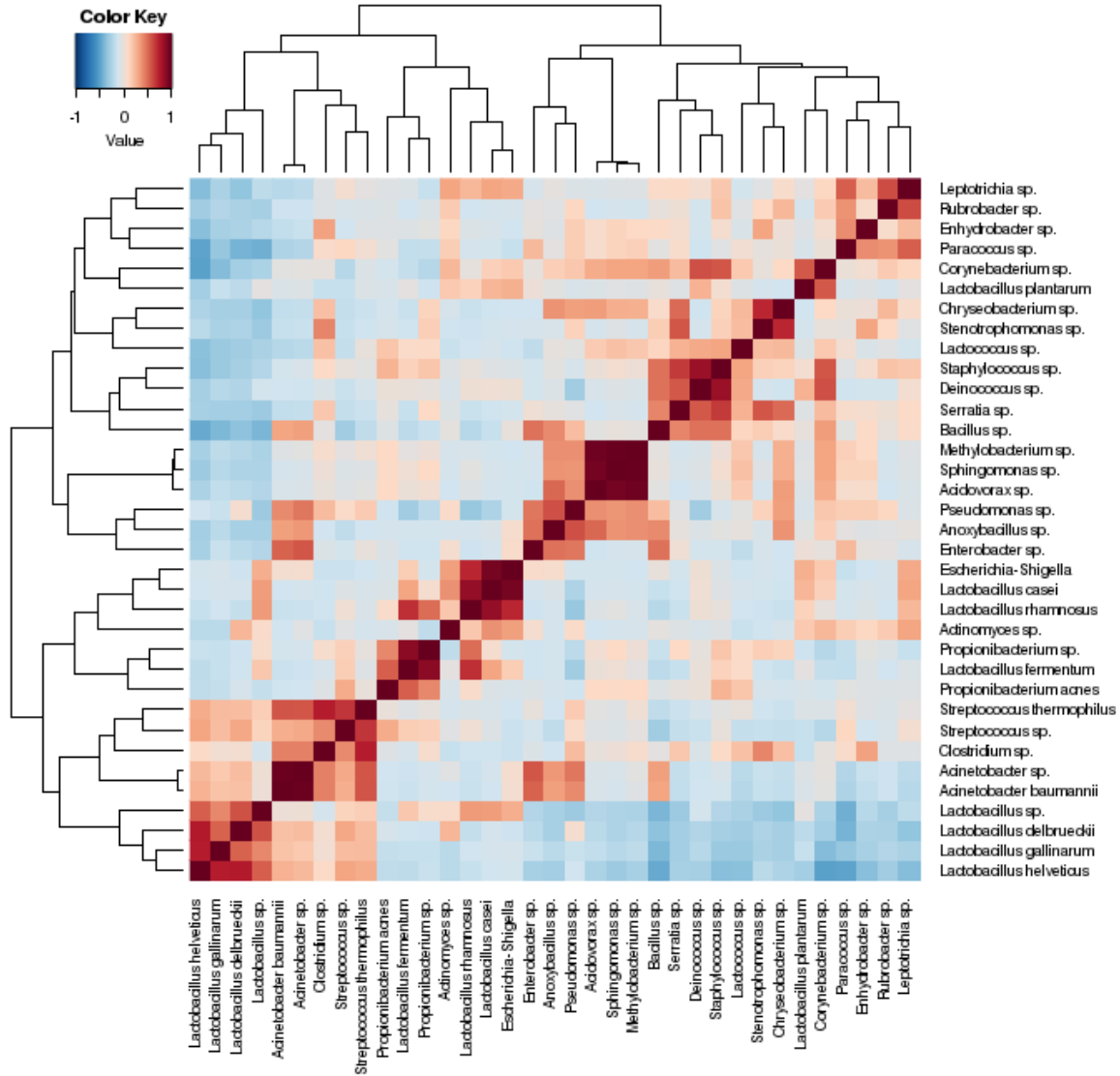


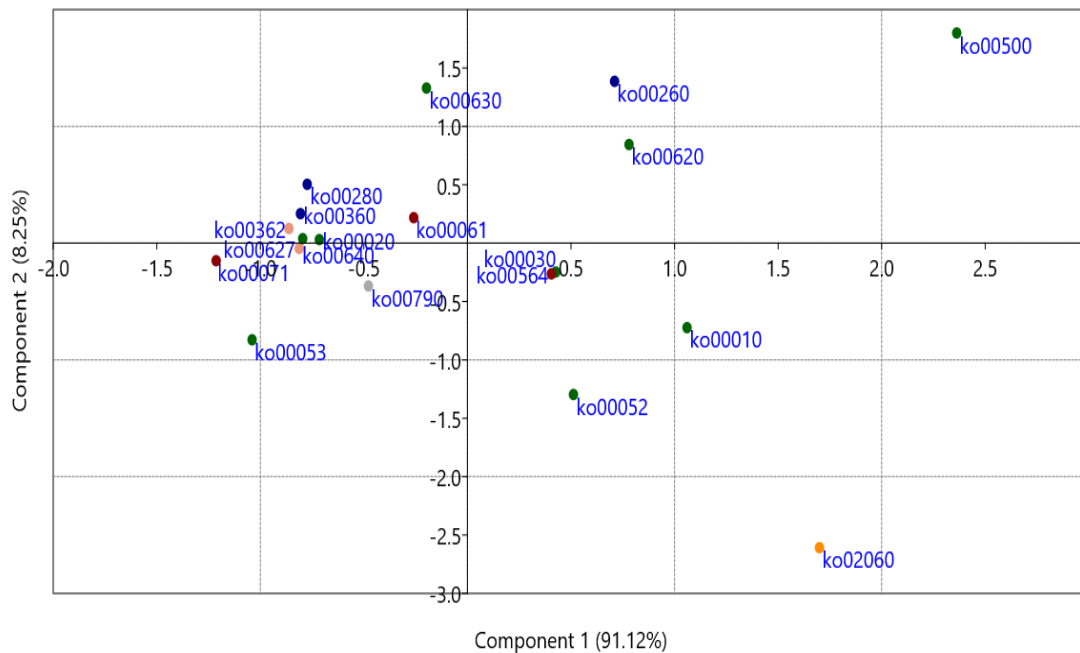
Fig. 5



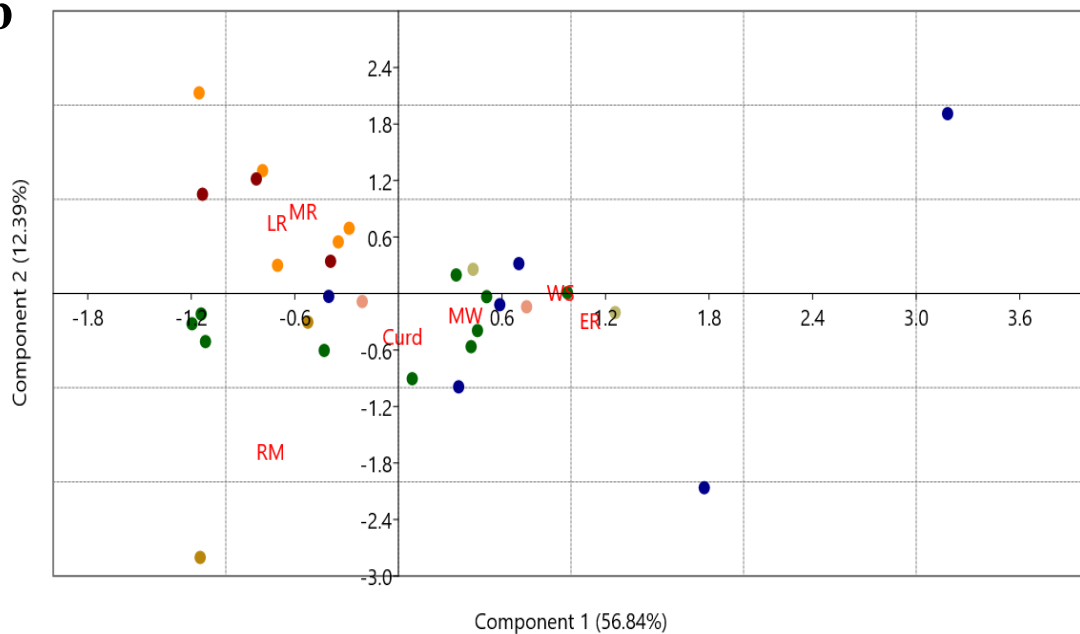
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030

Fig. 6

a



b



2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072

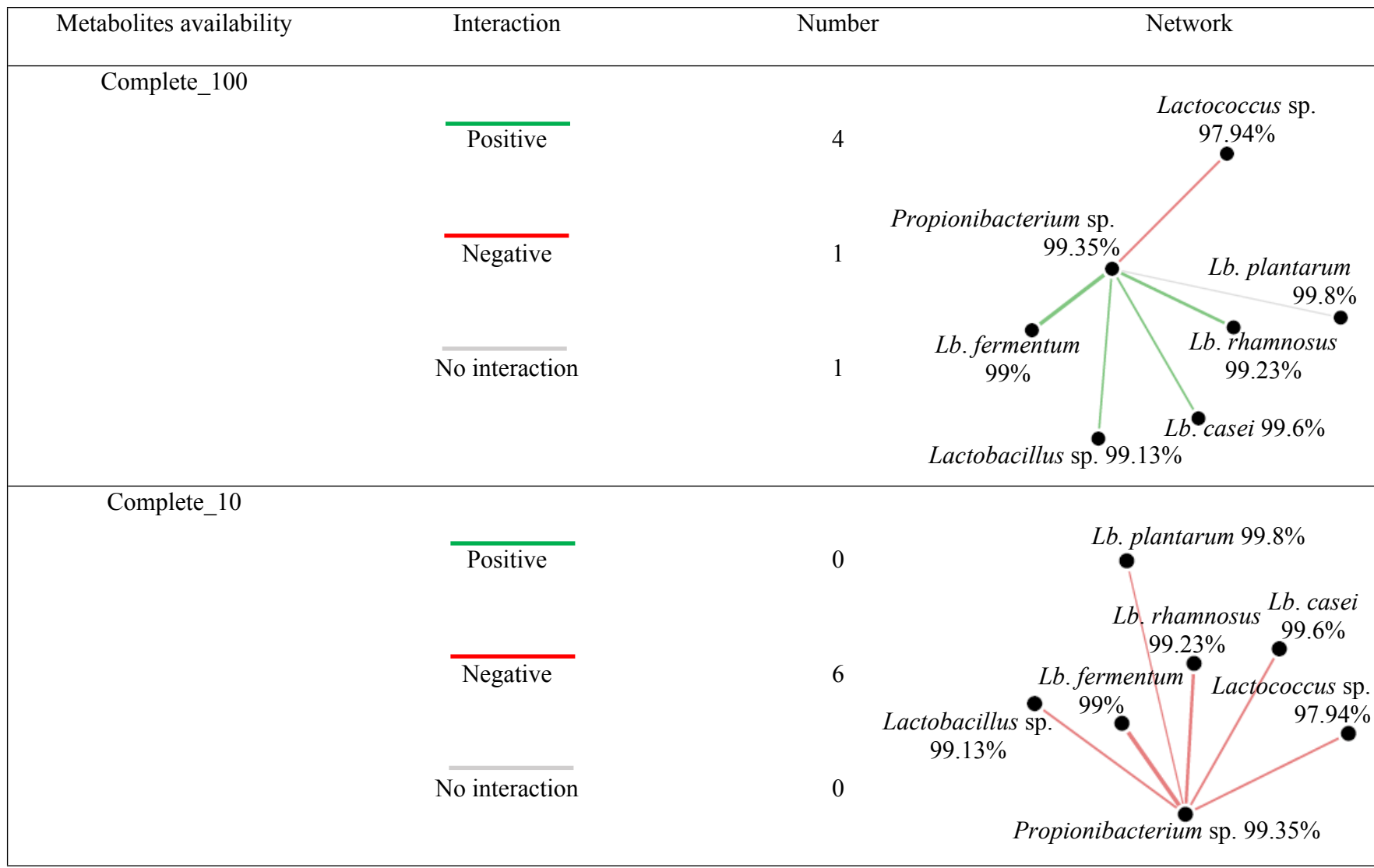


Fig. 7