Fish microbiomics: Strengths and limitations of MinION sequencing of gilthead sea bream (S*parus aurata*) intestinal microbiota

Socorro Toxqui-Rodríguez[a,b], Fernando Naya-Català[b], Ariadna Sitjà-Bobadilla[a], M. Carla Piazzon[a], Jaume Pérez-Sánchez[b,*].

[a]Fish Pathology Group, Institute of Aquaculture Torre de la Sal (IATS, CSIC), 12595 Ribera de Cabanes, Castellón, Spain.

[b]Nutrigenomics and Fish Growth Endocrinology Group, Institute of Aquaculture Torre de la Sal (IATS, CSIC), 12595 Ribera de Cabanes, Castellón, Spain.

*Corresponding author: Jaume Pérez Sánchez, e-mail: jaime.perez.sanchez@csic.es, phone number: +34 964319500

Abstract

There are several affordable methods involving different sequencing technologies for microbial characterization. However, the choice of the sequencing platform and the downstream analysis can yield somewhat different results. Here we aimed to examine the strengths and limitations of different sequencing platforms for their use in aquacultured fish gut mucosal samples, using gilthead sea bream (*Sparus aurata*) as a model. The attention was focused on the portable Oxford Nanopore Technologies (ONT) MinION device, which offers the possibility of fast profiling of mucosal microbial samples. The standard PCR protocol provided for the MinION full-length sequencing of the 16S rRNA was optimized (input material, annealing temperature and number of cycles) for fish mucosal samples and the optimized protocol was validated using a standard mock community with known bacterial taxa and abundances. The optimized PCR (annealing temperature 52ºC, 30 cycles) yielded highly accurate results when sequencing mock community samples ($R^2 = 0.95$) and was chosen for the subsequent analyses. Finally, the sequencing results were compared to those from Illumina MiSeq sequencing of the V3-V4 region of the 16S rRNA gene to determine strengths and weaknesses from both platforms. Our results showed that MinION is a reliable and accurate tool for the assessment of intestinal bacteria communities, yielding similar results to Illumina with correlation coefficients > 0.75. However, biologically important but less abundant taxa (e.g. Actinobacteriota and Bacteroidota) were apparently masked when comparisons were made with the Illumina MiSeq results. We conclude that the choice of sequencing platform depends on the type of sample and scientific question. Thus, when evaluating fish gut mucosal samples, where the biological interpretation focuses on taxa related to gut function and metabolism, Illumina MiSeq allows a broader amplification of taxa of interest, while MinION provides good results in terms of abundance and fast profiling of

microbial communities, making it very attractive for studies focused on environmentally-

related samples (e.g. gills and skin samples).

Keywords

Microbiota, Aquaculture, 16S rRNA, MinION, Illumina MiSeq, sequencing platform.

Highlights

1. The 16S MinION protocol has been optimized for fish gut mucosal samples

2. MinION sequencing yields an accurate classification of bacterial communities in fish

samples

3. Illumina MiSeq allows the assessment of biologically relevant less abundant

populations

4. The choice of sequencing platform will depend on the type of sample and the

experimental question

1. Introduction

Microbiota research is a fast-growing field in aquaculture with important implications for fish nutrition, health, and welfare (Brugman et al., 2018). In particular, gilthead sea bream (*Sparus aurata*) is one of the main cultured fish in the Mediterranean (APROMAR, 2020), and recently, many studies have been conducted to unravel its intestinal microbial dynamics (Naya-Català et al., 2021a, 2022b; Piazzon et al., 2020; Solé-Jiménez et al., 2021), as the gut microbiome plays a key role in maintaining the health status in fish, including the immune system response and digestion functions (Reda et al., 2022). However, there is still a long way to go in establishing the baseline parameters that guide the mucosal microbial manipulation of most farmed fish, including gilthead sea bream (Terova et al., 2022). Thus, we aimed to critically examine the core practice of common methodological approaches to provide practical guidelines that might serve to facilitate comparisons among different laboratories and technological platforms. Certainly, with the advent of next-generation sequencing (NGS) technologies, the amplification and sequencing of the 16S ribosomal RNA (16S rRNA) is the most widely used technique for the analysis of bacterial communities (Perry et al., 2020). However, conventional short-read sequencer platforms cannot yield reads covering the full length of the 16S rRNA gene, which might cause ambiguity in taxonomic classification (Kuczynski et al., 2011). This technical constrain can be overcome with the use of third-generation sequencing (TGS) platforms, that are capable of producing long sequences with no theoretical read length limit. A prime example is the Oxford Nanopore Technology (ONT) MinION, a portable device that is able to produce low-cost sequencing data (100 kb reads) in an immediacy context (Sevim et al., 2019), unlike other TGS systems (PacBio sequencing). Nevertheless, the accuracy of MinION is lower than other platforms, and its reliable use for quantitative/qualitative determinations could be compromised depending on the type

4

of sample and the threshold accuracy level. Otherwise, differences in primer sequences and PCR-amplification conditions for the 16S rRNA-based multitemplate PCR could bias microbial assessments (Fujiyoshi et al., 2020; Laursen et al., 2017). To challenge the best way in which these methodological constraints can be solved, we aimed to: 1) optimize the experimental 16S protocol when sequencing gilthead sea bream mucosal microbiota samples with the ONT MinION device, 2) evaluate the feasibility of MinION sequencing by comparing 16S amplicon data results from a defined mock community, and 3) evaluate the results from gilthead sea bream intestinal microbiota amplicons using the Illumina MiSeq (performed with standard procedures by sequencing companies), and MinION (on-site sequencing) platforms.

2. Materials and methods

2.1. Fish and mock community samples

The samples used in this study were randomly selected from a previous experiment using a mild hypoxia challenge. Two selected fish were from the HNN group (samples 1 and 2) and one from the NNN group (sample 3) (Naya-Català et al., 2021b). Mucus from the anterior intestinal portion was sampled following a previously optimized protocol (Piazzon et al., 2019). The same samples were used in parallel to compare sequencing platforms. Concerning the ZymoBIOMICS™ mock community, it is composed of 10 microbial species: eight bacteria and two yeasts with cell counts distributed on a log scale from 89.1% (*Listeria monocytogenes*) to 0.000089% (*Staphylococcus aureus*) (Table 1). For both types of samples, DNA was extracted using a High Pure PCR Template Preparation Kit (Roche), including a lysozyme lysis step for optimized DNA extraction (Piazzon et al., 2019). The quantity of DNA was assessed using the Nanodrop 2000c (Thermo Scientific).

98 All procedures involving experimental animals were approved by the Ethics and Animal

99 Welfare Committee of IATS and CSIC. They were carried out in a registered installation

100 facility (code ES120330001055) in accordance with the principles published in the

101 European Animal Directive (2010/63/EU) and Spanish laws (Royal Decree RD53/2013)

102 for the protection of animals used in scientific experiments.

103 **2.2 DNA input and PCR optimization**

104 The MinION 16S Barcoding Kit 1-24 (SQK-16S024) protocol version

105 16S_9086_v1_revR_14 Aug 2019 was optimized for assessing the composition of

106 bacterial communities in gilthead sea bream mucosal samples using the on-site MinION

107 device platform. The standard protocol contemplates an initial DNA input of 10 ng/µl for

108 the PCR amplification, but our sea bream mucosal samples had a large proportion of host

109 genetic material (Naya-Català et al., 2022a), and several mucosal DNA inputs (10, 100,

110 500 and 1000 ng/µl, Nanodrop measurements) were tested until an optimum

111 concentration was reached for sequencing. After PCR, amplicons were quantified using

112 PicoGreen™ (Thermo Fisher Scientific), and run in an agarose gel electrophoresis (1%

113 w/v in Tris-EDTA buffer) to determine the presence of a specific band of ~1500 bp. For

114 the PCR optimization procedure, five different conditions (PCR1-PCR5), including

115 changes in the number of cycles (from 25 to 30), annealing temperature (from 55ºC to

116 52ºC) and nesting were considered (Table 2). These protocol modifications were

117 performed using as starting point the PCR procedure recommended by the kit's

118 manufacturer (PCR1: 95°C for 1 min, 25 cycles of 95°C for 20 s, 55°C for 30 s, and 65°C

119 for 2 min, and a final extension step of 65°C for 5 min). In the case of the mock

120 community samples, the PCR testing procedure was reduced to the standard (PCR1) and

121 PCR4 with DNA inputs of 10 ng/µl, in order to compare the conditions recommended in

122 the kit and the optimized conditions. The used Taq polymerase (LongAmp Hot Start Taq

123 2X Master Mix (NEB, M0533S), PCR primers (F: 5' - ATCGCCTACCGTGAC - barcode

124 - AGAGTTTGATCMTGGCTCAG - 3' and R: 5' - ATCGCCTACCGTGAC - barcode -

125 CGGTTACCTTGTTACGACTT - 3') and reaction volume (50 μl) were those

126 recommended by the kit.

127 **2.3. ONT MinION sequencing and bioinformatics pipeline**

128 The complete 16S rRNA gene (V1-V9) was sequenced using the ONT MinION device

129 and the 16S Barcoding Kit 1-24 (SQK-16S024) version

130 16S_9086_v1_revR_14Aug2019, according to the manufacturer's protocol with the

131 previously mentioned modifications of input DNA and PCR conditions. The amplified

132 DNA was quantified using PicoGreen™ and 100 fmol were loaded into the MinION

133 device. Libraries were sequenced using an R9.4/FLO-MIN106 flow cell and

134 demultiplexed using the MinKNOW v21.11.17. The sequencing was stopped when

135 approximately 100,000 reads were obtained for each sample, which constituted a

136 sequencing run time of 21-23 h for the gilthead sea bream gut mucosal samples and ~2 h

137 for the mock community samples. Between runs, the MinION flow cell was washed

138 according to the ONT Flow Cell Wash Kit (EXP-WSH004) instructions. After

139 sequencing, the basecalling of the samples was performed with Guppy v5.1.12, using the

140 default parameters. The resulting FASTQ reads were pre-processed using Porechop

141 v0.2.4 (https://github.com/rrwick/Porechop) for removing sequencing adapters from

142 reads, NanoFilt v2.8.0 (De Coster et al., 2018) for filtering reads below 1,200 base pairs

143 (bp), and above 1,800 bp and yacrd v0.6.2 (Marijon et al., 2020) for chimera detection

144 and removal. Sequences were assigned as distinct amplicon sequence variants (ASVs)

145 and subsequently mapped for taxonomy assignment with Minimap2 v2.17-r941 (Li,

146 2021), using SILVA v138.1 (Yilmaz et al., 2014) as the reference database.

147

2.4. Illumina MiSeq and bioinformatics pipeline

149 The Illumina MiSeq platform at the Unidad de Genómica del Parque Científico de Madrid

150 (FPCM) was used for Sequencing the V3-V4 region of the 16S rRNA gene of fish

151 mucosal samples following the same protocol, primers and conditions described

152 elsewhere (Piazzon et al., 2019). FASTQ paired-reads were merged using VSEARCH

153 v2.15.1 (Rognes et al., 2016), and then pre-processed using Prinseq v0.20.4 (Schmieder

154 and Edwards, 2011). Finally, sequences were clustered at a 100% identity using

155 VSEARCH v2.15.1, identified as distinct ASVs, and subsequently aligned for taxonomy

156 assignment with Minimap2 v2.17-r941 (Li, 2021) with SILVA v138.1 (Yilmaz et al.,

157 2014) as the reference database.

158 **2.5. Statistics and data availability**

159 Mann-Whitney U test was used to compare differences in mean abundances and the

160 correlation coefficient was used to determine differences between individual profiles. All

161 calculations and visualizations were performed using the R statistical software.

162 Raw sequence data from all experiments was uploaded to the Sequence Read Archive

163 (SRA) under Bioproject accession number PRJNA891255 (BioSample accession

164 numbers: SAMN31314095-11114).

165 # 3. Results

166 **3.1. Optimization of ONT 16S rRNA protocol for fish microbiota sequencing**

167 The testing of different DNA concentrations for ONT MinION sequencing of intestinal

168 adherent microbiota resulted in the selection of an initial input of 500-1000 ng/µl. Indeed,

169 DNA concentrations of 10 and 100 ng/µl did not reach the minimum input level (50 fmol)

170 required for library preparation in the 16S rRNA ONT MinION protocol (SQK-16S024),

and did not show a clear band when run in an agarose gel. Regarding PCR results (Table 2), PCR2 and PCR3 procedures yielded the lowest number of bacteria phyla (3), decreasing the eukaryotic sequence abundance from 4.93% in PCR2 to 0.24% in PCR3. A highest number of phyla (5) was found in PCR1 and PCR5, while the abundance of eukaryotic sequences decreased from 5.17% in PCR1 to 0.01% in PCR5. The PCR4 conditions also yielded a low abundance of eukaryotic sequences (0.04%) with reads assigned to 4 microbial phyla, but the number of reads assigned to the Proteobacteria phylum decreased in PCR4 (85.30%) when compared to PCR5 (96.88%), allowing to detect more Firmicutes, taxa of relevance from a functional point of view (Table 3). Therefore, PCR4 was chosen as the best conditions for gilthead sea bream mucosal samples. PCR1 (Mock1) and PCR4 (Mock2) conditions were used for sequencing the bacterial mock community samples to compare the results obtained with the conditions specified by the commercial kit and the newly customized protocol, in order to determine that no biases were being introduced. After the ONT MinION sequencing a total of 280,844 and 238,441 raw reads were obtained, respectively, and a high taxonomic assignation percentage (>91%) was reached in both cases. The correlation between the theoretical abundance of each genus and the observed abundance in the ONT MinION sequencing (Figure 1) resulted in significant coefficients between these two values ($R^2 =$ 0.8803 for Mock1; $R^2 = 0.9464$ for Mock2).

3.2. Illumina MiSeq and ONT MinION sequencing results comparison for gut microbiota samples

The same gilthead sea bream gut mucosal samples were sequenced with the ONT MinION sequencer (using the optimized conditions of PCR4) and Illumina MiSeq. Using the ONT MinION device, an average of 94,164 raw reads/sample were generated, with a mean of 97.42% (91,738) assigned reads. Taxonomic classification yielded a mean of 3

phyla, which were present with an abundance of more than 0.25% per sample. Using

Illumina MiSeq, an average of 316,490 raw reads were obtained per sample and a mean

of 74.38% (235,428) of the raw reads was classified into a mean of 6 phyla, which were

present with an abundance of more than 0.25% per sample. Although at a first glance the

sequencing results for both platforms yielded somewhat different results (Figure 2A), in

both cases, when applying the abundance threshold, three phyla: Proteobacteria,

Firmicutes, and Bacteroidota were shared, ranging from 85.3 to 63.6%, from 14.2 to

21.3% and from 0.38 to 2.6% (ONT MinION to Illumina MiSeq), respectively. To

determine whether there was a significant difference between the shared phyla, a Mann-

Whitney U test was conducted resulting in non-significant differences with a p-value of

0.4 for Proteobacteria, a p-value of 0.7 for Firmicutes, and a p-value of 0.1 for

Bacteroidota. Three additional phyla (Actinobacteriota, Patescibacteria, and

Cyanobacteria) were detected with both sequencing platforms, although in the ONT

MinION results, their mean abundance percentage was below 0.25%, being 0.024, 0.001,

and 0.0001%, respectively. For these phyla, the Mann-Whitney U test also resulted in

non-significant differences, with a p-value of 0.07 for Actinobacteriota and

Patescibacteria and a p-value of 1 for Cyanobacteria. Considering that the lack of

statistical significance could be due to the large variability among samples, the individual

results were also compared (Figure 2B-D). Again, although the results were apparently

different, when calculating the correlation coefficients, high R^2 values were obtained

(0.77, 0.99 and 0.89 for samples 1, 2 and 3, respectively).

4. Discussion

Nowadays, metagenomic approaches, such as the 16S rRNA gene amplicon sequencing,

provide new insights into microbial compositions and several methodologies have been

optimized to measure this. However, it is important to consider that the optimized

221 protocols have been constructed using specific sets of samples like mock communities,

222 which contain pure and defined microorganisms, or human or environmental samples,

223 with very different characteristics and distributions than the ones that can be found in fish

224 mucosal samples. In fact, we have recently described that gilthead sea bream gut mucosal

225 samples contain approximately 5-10% of bacterial transcripts, being the majority of

226 transcripts from host origin (Naya-Català et al., 2022a). Hence, it is not unexpected that

227 the initial DNA input required in order to achieve a sufficient amount of template for 16S

228 rRNA amplification needs to be > 500 ng, which would roughly correspond to the 10

229 ng/µl of microbial DNA commonly specified in standardized protocols. The input

230 required will depend on the quantification method used and the percentage of host

231 material yielded by the extraction protocol. Thus, the template amount should be

232 optimized for each specific protocol. In fact, Kennedy et al. (2014) reported that higher

233 overall template concentration should be used for amplicon-based 16S rRNA sequencing

234 in the case of low nucleic acid recoveries in samples with limited microbial biomass (e.g.

235 surfaces, skin, and tissues).

236 Among the previously indicated factors that affect the accuracy of the results using the

237 ONT MinION sequencer (D'Amore et al., 2016; Gołębiewski and Tretyn, 2020; Hongoh

238 et al., 2003; Sipos et al., 2007), we focused on the number of PCR cycles and the

239 annealing temperature to optimize the best conditions for our gilthead sea bream gut

240 mucosal samples. Concerning the number of PCR cycles, we added five extra cycles to

241 the standard protocol (25 cycles), and even though several studies showed that greater

242 PCR cycle numbers can cause chimera generation and/or interfere with bacterial

243 community structure analysis (Gołębiewski and Tretyn, 2020), under our conditions, we

244 did not detect significant biases or problems with the quality of the reads, including

245 chimera formation, as evidenced by the high number of retained and assigned reads (>

246 90%). Similar studies, dealing with the optimization of ONT technologies for non-

247 standard samples, reported that the increase in the number of cycles is necessary when

248 working with samples with low amounts of microbial cells (Fujiyoshi et al., 2020).

249 Regarding the annealing temperature, its significant effects on PCR bias have been

250 characterized before (Peng et al., 2018; Sipos et al., 2007). In these types of multitemplate

251 PCRs, primers have different binding energies depending on the template which will have

252 different GC content and different mismatches (Laursen et al., 2017). Thus, it was

253 suggested that the annealing temperature used in multitemplate PCRs should be the

254 lowest possible where the reaction is still specific (Peng et al., 2018; Sipos et al., 2007).

255 Our results show that with our primers and conditions, higher temperatures tend to have

256 a bias towards Proteobacteria. An annealing temperature of 52ºC reduced this bias and

257 demonstrated to be more specific when amplifying known bacterial communities,

258 illustrated by the higher correlation coefficient in the Mock2 sample. Thus, considering

259 the number of phyla, unspecific amplification products, and the distribution of the

260 bacterial community, 30 cycles and 52ºC annealing temperature were the most suitable

261 conditions for our type of samples.

262 Within the many sources of variability for 16S amplicon sequencing results, the choice

263 of primers has been repeatedly reported as one of the most determinant (Fouhy et al.,

264 2016; Hongoh et al., 2003; Peng et al., 2018; Suzuki and Giovannoni, 1996; Walker et

265 al., 2015). Up to date, due to its accuracy, the most popular platform to study microbiota

266 by 16S amplicon sequencing is Illumina MiSeq. However, this platform only allows

267 sequencing of fragments of the 16S rRNA gene (100-300 bp), and several studies report

268 that different primer pairs targeting different regions of the gene, provide different

269 taxonomic results (Cai et al., 2013; Guo et al., 2013; Klindworth et al., 2013; Wang et al.,

270 2018). Besides, short-read sequencing allows lower resolution during taxonomic

271 assignment (Wommack et al., 2008). In addition, there is evidence of multiple sources of

272 bias of Illumina-based bacterial 16S rRNA sequencing, including GC content, DNA

273 extraction protocol, sample storage conditions and library preparation (Aird et al., 2011;

274 Cardona et al., 2012; Kennedy et al., 2014; Sato et al., 2019; Schirmer et al., 2015).

275 Concerning the PCR conditions, Aird et al. (2011) found that factors such as the

276 thermocycler and temperature ramp rate, play a surprisingly big role in introducing bias.

277 Thus, it is clear that comparisons between two different platforms (Illumina MiSeq and

278 ONT MinION) using different sets of primers targeting different regions (V3-V4 and full

279 length, respectively), library preparation protocols, PCR conditions, and posterior

280 bioinformatic analysis will hardly yield fully comparable results. In our case, the

281 comparison of the data generated using both platforms showed considerable overlap in

282 the detected taxa, at least at phylum level, but the higher abundance of Proteobacteria

283 (85.3%) detected by Nanopore full-length sequencing seems to mask the detection of less

284 abundant taxa as Actinobacteriota or Bacteroidota, whose functional relevance in gut

285 functions is well known (Estruch et al., 2015; Kormas et al., 2014; Magne et al., 2020;

286 Piazzon et al., 2019). The dominance of Proteobacteria is probably not an artifact, in fact,

287 Proteobacteria is the most abundant bacteria phylum in samples for marine origin

288 (Najafpour et al., 2021; Rosado et al., 2019; Salgueiro et al., 2020; Thomas et al., 2021).

289 Short-read and long-read sequencing approaches allow rapid, high-throughput, and

290 accurate classification of bacterial communities with identification of 16S rRNA genes

291 (Wei et al., 2020). Nevertheless, when it comes to the selection of sequencing platforms,

292 besides evaluating their characteristics and library preparation protocols available, the

293 expected output should be considered. In our case, we evaluated the microbial

294 composition of gut mucosal gilthead sea bream samples observing that although no

295 significant differences were observed in the most abundant phyla of bacteria, some taxa

296 with lower abundance such as Actinobacteriota or Bacteroidota were not detected, being

297 masked by highly abundant phyla. These taxa, present in lower proportions, have been

298 previously defined as important in gut physiology (Estruch et al., 2015; Kormas et al.,

299 2014; Magne et al., 2020; Piazzon et al., 2019) and the V3-V4 primers designed for

300 Illumina 16S sequencing seem to amplify their presence allowing for a most accurate

301 biological interpretation concerning gut function and metabolism. Thus, the choice of

302 primers and sequencing platform needs to consider the aim of the study to obtain the

303 required results. In our hands, to have an enhanced view of gut metabolic function-related

304 taxa we would recommend the use of Illumina V3-V4 sequencing to zoom up in these

305 relevant, although less abundant, populations. Samples more related to the environment,

306 such as water samples or gill and skin mucosal samples, would benefit from the use of

307 ONT technologies, which provide accurate relative abundances in a low-cost and time-

308 efficient manner. The current study was focused on establishing an optimized protocol

309 and assess the differences between the two sequencing platform results, thus, a limited

310 number of samples was used. Future studies should be directed to evaluate optimal

311 number of samples, subsamples and replicates to maximize efficiency when using ONT

312 MinION sequencing.

5. Conclusions

314 The microbial composition of gilthead sea bream gut mucosal samples differed depending

315 on the PCR conditions and the sequencing platform used. Concerning the PCR results,

316 besides the primer choice, changes in the annealing temperature and number of cycles

317 had an effect on gilthead sea bream gut microbial samples. In addition, the results

318 demonstrate that it is not fully possible to compare amplicon data from different

319 sequencing platforms, especially when it comes to terms of abundance and not in terms

320 of the functionality of the taxa of interest, and the approach of the planned experiment.

321 Finally, the ONT MinION device was validated with the ZymoBIOMICS™ mock

322 community amplicon data, and the 16S (SQK-16S024) ONT protocol was optimized in

323 order to work with gilthead sea bream mucosal samples, making this device suitable for

324 our future experiments, although the choice of platform will depend on the type of

325 information required from the experiment, the type of sample, the expected results, and

326 the time.

Funding

Author's contribution

337 **S. Toxqui-Rodriguez:** Methodology, Validation, Formal analysis, Investigation, Writing

338 – original draft, Visualization. **F. Naya-Català:** Validation, Formal analysis,

339 Investigation, Data curation, Writing – review & editing, Visualization. **A. Sitjà-**

340 **Bobadilla:** Investigation, Project administration, Funding acquisition, Writing – review

341 & editing. **M.C. Piazzon:** Methodology, Validation, Formal analysis, Investigation,

342 Supervision, Writing – original draft, Visualization. **J. Pérez-Sánchez:**

343 Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Writing –

344 review & editing, Visualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that the research was conducted in the absence of any commercial or

financial relationships that could be construed as a potential conflict of interest.

References

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12, R18. https://doi.org/10.1186/gb-2011-12-2-r18

APROMAR, 2020. Aquaculture in Spain 2020 [WWW Document]. URL http://www.apromar.es/content/informes-anuales (accessed 3.10.21).

Brugman, S., Ikeda-Ohtsubo, W., Braber, S., Folkerts, G., Pieterse, C.M.J., Bakker, P.A.H.M., 2018. A comparative review on microbiota manipulation: Lessons from fish, plants, livestock, and human research. Front. Nutr. 5, 80. https://doi.org/10.3389/fnut.2018.00080

Cai, L., Ye, L., Tong, A.H.Y., Lok, S., Zhang, T., 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One 8, e53649. https://doi.org/10.1371/journal.pone.0053649

Cardona, S., Eck, A., Cassellas, M., Gallart, M., Alastrue, C., Dore, J., Azpiroz, F., Roca, J., Guarner, F., Manichanh, C., 2012. Storage conditions of intestinal microbiota matter in metagenomic analysis. BMC Microbiol. 12, 158. https://doi.org/10.1186/1471-2180-12-158

D'Amore, R., Ijaz, U.Z., Schirmer, M., Kenny, J.G., Gregory, R., Darby, A.C., Shakya, M., Podar, M., Quince, C., Hall, N., 2016. A comprehensive benchmarking study of protocols and sequencing platforms for 16S rRNA community profiling. BMC Genomics 17, 55. https://doi.org/10.1186/s12864-015-2194-9

De Coster, W., D'Hert, S., Schultz, D.T., Cruts, M., Van Broeckhoven, C., 2018. NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics 34, 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

Estruch, G., Collado, M.C., Penaranda, D.S., Tomas Vidal, A., Jover Cerda, M., Perez Martinez, G., Martinez-Llorens, S., 2015. Impact of fishmeal replacement in diets for gilthead sea bream (*Sparus aurata*) on the gastrointestinal microbiota determined by pyrosequencing the 16S rRNA gene. PLoS One 10, e0136389. https://doi.org/10.1371/journal.pone.0136389

Fouhy, F., Clooney, A.G., Stanton, C., Claesson, M.J., Cotter, P.D., 2016. 16S rRNA gene sequencing of mock microbial populations- impact of DNA extraction method, primer choice and sequencing platform. BMC Microbiol. 16, 123. https://doi.org/10.1186/s12866-016-0738-z

Fujiyoshi, S., Muto-Fujita, A., Maruyama, F., 2020. Evaluation of PCR conditions for characterizing bacterial communities with full-length 16S rRNA genes using a

384   portable nanopore sequencer. Sci. Rep. 10, 12580. https://doi.org/10.1038/s41598-
385   020-69450-9

386 Gołębiewski, M., Tretyn, A., 2020. Generating amplicon reads for microbial community
387   assessment with next-generation sequencing. J. Appl. Microbiol. 128, 330–354.
388   https://doi.org/10.1111/jam.14380

389 Guo, F., Ju, F., Cai, L., Zhang, T., 2013. Taxonomic precision of different hypervariable
390   regions of 16S rRNA gene and annotation methods for functional bacterial groups
391   in biological wastewater treatment. PLoS One 8, e76185.
392   https://doi.org/10.1371/journal.pone.0076185

393 Hongoh, Y., Yuzawa, H., Ohkuma, M., Kudo, T., 2003. Evaluation of primers and PCR
394   conditions for the analysis of 16S rRNA genes from a natural environment. FEMS
395   Microbiol. Lett. 221, 299–304. https://doi.org/10.1016/S0378-1097(03)00218-0

396 Kennedy, K., Hall, M.W., Lynch, M.D.J., Moreno-Hagelsieb, G., Neufeld, J.D., 2014.
397   Evaluating bias of illumina-based bacterial 16S rRNA gene profiles. Appl.
398   Environ. Microbiol. 80, 5717–5722. https://doi.org/10.1128/AEM.01451-14

399 Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner,
400   F.O., 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for
401   classical and next-generation sequencing-based diversity studies. Nucleic Acids
402   Res. 41, e1. https://doi.org/10.1093/nar/gks808

403 Kormas, K.A., Meziti, A., Mente, E., Frentzos, A., 2014. Dietary differences are
404   reflected on the gut prokaryotic community structure of wild and commercially
405   reared sea bream (*Sparus aurata*). Microbiol. Open 3, 718–728.
406   https://doi.org/10.1002/mbo3.202

407 Kuczynski, J., Lauber, C.L., Walters, W.A., Parfrey, L.W., Clemente, J.C., Gevers, D.,
408   Knight, R., 2011. Experimental and analytical tools for studying the human
409   microbiome. Nat. Rev. Genet. 13, 47–58. https://doi.org/10.1038/nrg3129

410 Laursen, M.F., Dalgaard, M.D., Bahl, M.I., 2017. Genomic GC-content affects the
411   accuracy of 16S rRNA gene sequencing based microbial profiling due to PCR bias.
412   Front. Microbiol. 8, 1934. https://doi.org/10.3389/fmicb.2017.01934

413 Li, H., 2021. New strategies to improve minimap2 alignment accuracy. Bioinformatics
414   37, 4572–4574. https://doi.org/10.1093/bioinformatics/btab705

415 Magne, F., Gotteland, M., Gauthier, L., Zazueta, A., Pesoa, S., Navarrete, P.,
416   Balamurugan, R., 2020. The Firmicutes/Bacteroidetes ratio: A relevant marker of
417   gut dysbiosis in obese patients? Nutrients 12, 1474.
418   https://doi.org/10.3390/nu12051474

419 Marijon, P., Chikhi, R., Varré, J.-S., 2020. yacrd and fpa: Upstream tools for long-read
420   genome assembly. Bioinformatics 36, 3894–3896.
421   https://doi.org/10.1093/bioinformatics/btaa262

422 Najafpour, B., Pinto, P.I.S., Moutou, K.A., Canario, A.V.M., Power, D.M., 2021.
423   Factors driving bacterial microbiota of eggs from commercial hatcheries of
424   European seabass and gilthead seabream. Microorganisms 9, 2275.
425   https://doi.org/10.3390/microorganisms9112275

426 Naya-Català, F., do Vale Pereira, G., Piazzon, M.C., Fernandes, A.M., Calduch-Giner,

J.A., Sitjà-Bobadilla, A., Conceição, L.E.C., Pérez-Sánchez, J., 2021a. The cross-talk between intestinal microbiota and host gene expression in juveniles of gilthead sea bream (Sparus aurata). Insights in fish feeds for increased circularity and resource utilization. Front. Physiol. 12, 784265. https://doi.org/10.3389/fphys.2021.748265

Naya-Català, F., Piazzon, M.C., Calduch-Giner, J.A., Sitjà-Bobadilla, A., Pérez-Sánchez, J., 2022a. Diet and host genetics drive the bacterial and fungal intestinal metatranscriptome of gilthead sea bream. Front. Microbiol. 13, 883738. https://doi.org/10.3389/fmicb.2022.883738

Naya-Català, F., Piazzon, M.C., Torrecillas, S., Toxqui-Rodríguez, S., Calduch-Giner, J.À., Fontanillas, R., Sitjà-Bobadilla, A., Montero, D., Pérez-Sánchez, J., 2022b. Genetics and nutrition drive the gut microbiota succession and host-transcriptome interactions through the gilthead sea bream (Sparus aurata) production cycle. Biology (Basel). 11, 1744. https://doi.org/10.3390/biology11121744

Naya-Català, F., Simó-Mirabet, P., Calduch-Giner, J., Pérez-Sánchez, J., 2021b. Transcriptomic profiling of Gh/Igf system reveals a prompted tissue-specific differentiation and novel hypoxia responsive genes in gilthead sea bream. Sci. Rep. 11, 16466. https://doi.org/10.1038/s41598-021-95408-6

Peng, W., Li, X., Wang, C., Cao, H., Cui, Z., 2018. Metagenome complexity and template length are the main causes of bias in PCR-based bacteria community analysis. J. Basic Microbiol. 58, 987–997. https://doi.org/10.1002/jobm.201800265

Perry, W.B., Lindsay, E., Payne, C.J., Brodie, C., Kazlauskaite, R., 2020. The role of the gut microbiome in sustainable teleost aquaculture. Proceedings. Biol. Sci. 287, 20200184. https://doi.org/10.1098/rspb.2020.0184

Piazzon, M.C., Naya-Català, F., Perera, E., Palenzuela, O., Sitjà-Bobadilla, A., Pérez-Sánchez, J., 2020. Genetic selection for growth drives differences in intestinal microbiota composition and parasite disease resistance in gilthead sea bream. Microbiome 8, 168. https://doi.org/10.1186/s40168-020-00922-w

Piazzon, M.C., Naya-Català, F., Simó-Mirabet, P., Picard-Sánchez, A., Roig, F.J., Calduch-Giner, J.A., Sitjà-Bobadilla, A., Pérez-Sánchez, J., 2019. Sex, age, and bacteria: How the intestinal microbiota is modulated in a protandrous hermaphrodite fish. Front. Microbiol. 10, 2512. https://doi.org/10.3389/fmicb.2019.02512

Reda, R.M., Maricchiolo, G., Quero, G.M., Basili, M., Aarestrup, F.M., Pansera, L., Mirto, S., Abd El-Fattah, A.H., Alagawany, M., Abdel Rahman, A.N., 2022. Rice protein concentrate as a fish meal substitute in *Oreochromis niloticus*: Effects on immune response, intestinal cytokines, Aeromonas veronii resistance, and gut microbiota composition. Fish Shellfish Immunol. 126, 237–250. https://doi.org/10.1016/j.fsi.2022.05.048

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahe, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584. https://doi.org/10.7717/peerj.2584

Rosado, D., Pérez-Losada, M., Severino, R., Cable, J., Xavier, R., 2019. Characterization of the skin and gill microbiomes of the farmed seabass

471   (*Dicentrarchus labrax*) and seabream (*Sparus aurata*). Aquaculture 500, 57–64.
472   https://doi.org/10.1016/j.aquaculture.2018.09.063

473 Salgueiro, V., Manageiro, V., Bandarra, N.M., Reis, L., Ferreira, E., Caniça, M., 2020.
474   Bacterial diversity and antibiotic susceptibility of *Sparus aurata* from aquaculture.
475   Microorganisms 8, 1343. https://doi.org/10.3390/microorganisms8091343

476 Sato, M.P., Ogura, Y., Nakamura, K., Nishida, R., Gotoh, Y., Hayashi, M., Hisatsune,
477   J., Sugai, M., Takehiko, I., Hayashi, T., 2019. Comparison of the sequencing bias
478   of currently available library preparation kits for Illumina sequencing of bacterial
479   genomes and metagenomes. DNA Res. 26, 391–398.
480   https://doi.org/10.1093/dnares/dsz017

481 Schirmer, M., Ijaz, U.Z., D'Amore, R., Hall, N., Sloan, W.T., Quince, C., 2015. Insight
482   into biases and sequencing errors for amplicon sequencing with the Illumina
483   MiSeq platform. Nucleic Acids Res. 43, e37. https://doi.org/10.1093/nar/gku1341

484 Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic
485   datasets. Bioinformatics 27, 863–864.
486   https://doi.org/10.1093/bioinformatics/btr026

487 Sevim, V., Lee, Juna, Egan, R., Clum, A., Hundley, H., Lee, Janey, Everroad, R.C.,
488   Detweiler, A.M., Bebout, B.M., Pett-Ridge, J., Göker, M., Murray, A.E.,
489   Lindemann, S.R., Klenk, H.-P., O'Malley, R., Zane, M., Cheng, J.-F., Copeland,
490   A., Daum, C., Singer, E., Woyke, T., 2019. Shotgun metagenome data of a defined
491   mock community using Oxford Nanopore, PacBio and Illumina technologies. Sci.
492   data 6, 285. https://doi.org/10.1038/s41597-019-0287-z

493 Sipos, R., Szekely, A.J., Palatinszky, M., Revesz, S., Marialigeti, K., Nikolausz, M.,
494   2007. Effect of primer mismatch, annealing temperature and PCR cycle number on
495   16S rRNA gene-targetting bacterial community analysis. FEMS Microbiol. Ecol.
496   60, 341–350. https://doi.org/10.1111/j.1574-6941.2007.00283.x

497 Solé-Jiménez, P., Naya-Català, F., Piazzon, M.C., Estensoro, I., Calduch-Giner, J.A.,
498   Sitjà-Bobadilla, A., Van Mullem, D., Pérez-Sánchez, J., 2021. Reshaping of gut
499   microbiota in gilthead sea bream fed microbial and processed animal proteins as
500   the main dietary protein source. Front. Mar. Sci. 8, 705041.
501   https://doi.org/10.3389/fmars.2021.705041

502 Suzuki, M.T., Giovannoni, S.J., 1996. Bias caused by template annealing in the
503   amplification of mixtures of 16S rRNA genes by PCR. Appl. Environ. Microbiol.
504   62, 625–630. https://doi.org/10.1128/aem.62.2.625-630.1996

505 Terova, G., Naya-Català, F., Rimoldi, S., Piazzon, M.C., Torrecillas, S., Toxqui, M.S.,
506   Fontanillas, R., Calduch-Giner, J., Hostins, B., Sitjà-Bobadilla, A., Montero, D.,
507   Pérez-Sánchez, J., 2022. Highlights from gut microbiota survey in farmed fish -
508   European sea bass and gilthead sea bream case studies. Aquac. Eur. 47, 5–10.

509 Thomas, A., Konteles, S.J., Ouzounis, S., Papatheodorou, S., Tsakni, A., Houhoula, D.,
510   Tsironi, T., 2021. Bacterial community in response to packaging conditions in
511   farmed gilthead seabream. Aquac. Fish. In Press.
512   https://doi.org/10.1016/j.aaf.2021.09.002

513 Walker, A.W., Martin, J.C., Scott, P., Parkhill, J., Flint, H.J., Scott, K.P., 2015. 16S
514   rRNA gene-based profiling of the human infant gut microbiota is strongly

515   influenced by sample processing and PCR primer choice. Microbiome 3, 26.
516   https://doi.org/10.1186/s40168-015-0087-4

517   Wang, F., Men, X., Zhang, G., Liang, K., Xin, Y., Wang, J., Li, A., Zhang, H., Liu, H.,
518   Wu, L., 2018. Assessment of 16S rRNA gene primers for studying bacterial
519   community structure and function of aging flue-cured tobaccos. AMB Express 8,
520   182. https://doi.org/10.1186/s13568-018-0713-1

521   Wei, P.-L., Hung, C.-S., Kao, Y.-W., Lin, Y.-C., Lee, C.-Y., Chang, T.-H., Shia, B.-C.,
522   Lin, J.-C., 2020. Characterization of fecal microbiota with clinical specimen using
523   long-read and short-read sequencing platform. Int. J. Mol. Sci. 21, 7110.
524   https://doi.org/10.3390/ijms21197110

525   Wommack, K.E., Bhavsar, J., Ravel, J., 2008. Metagenomics: Read length matters.
526   Appl. Environ. Microbiol. 74, 1453–1463. https://doi.org/10.1128/AEM.02181-07

527   Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C., Schweer, T.,
528   Peplies, J., Ludwig, W., Glöckner, F.O., 2014. The SILVA and "All-species Living
529   Tree Project (LTP)" taxonomic frameworks. Nucleic Acids Res. 42, D643-8.
530   https://doi.org/10.1093/nar/gkt1209

531

Tables

532

533   **Table 1.** Description of the ZymoBIOMICS™ Microbial Community Standard II (Log

534   distribution), showing the theoretical composition in percentages, followed by the results

535   obtained by sequencing two samples with different PCR conditions (Mock1: PCR1,

536   Mock2: PCR4, Table 2).

Genus	Phylum	Theoretical genomic DNA (%)	Results Mock1 (%)	Results Mock2 (%)
Listeria	Firmicutes	89.1	70.6253953	72.09623
Pseudomonas	Proteobacteria	8.9	17.9869543	20.08756
Bacillus	Firmicutes	0.89	6.94775891	6.039736
Escherichia	Proteobacteria	0.089	2.01739171	0.844545
Salmonella	Proteobacteria	0.089	2.31229798	0.811664
Lactobacillus	Firmicutes	0.0089	0.00504445	0.011261
Enterococcus	Firmicutes	0.00089	0.02056583	0.020269
Staphylococcus	Firmicutes	0.000089	0.00232821	0.00045
Saccharomyces	Ascomycota	0.89	N/D	N/D
Cryptococcus	Basidiomycota	0.00089	N/D	N/D
Other	-	N/A	0.08226333	0.08828

537   N/D: not detected. N/A: not applicable.

538

539   **Table 2**. PCR conditions tested to prepare the MinION input and results. Annealing

540   temperatures and the number of cycles were modified using the protocol recommended

541 PCR conditions as a base (95°C for 1 min, 25 cycles of 95°C for 20 s, 55°C for 30 s, and

542 65°C for 2 min, and a final extension step of 65°C for 5 min), equivalent to PCR1.

| PCR	Annealing temperature (ºC)	Number of cycles	Number of bacteria phyla >0.5% abundance	Eukaryotic sequence abundance (%)
PCR1	55	25	5	5.17
PCR2	52	25	3	4.93
PCR3	55	30	3	0.24
PCR4	52	30	4	0.06
PCR5	55	25+10	5	0.01

543

544 **Table 3**. Relative abundance in percentage of gilthead sea bream gut mucosal samples

545 16S rRNA amplicons sequenced by MinION when using two different PCR conditions.

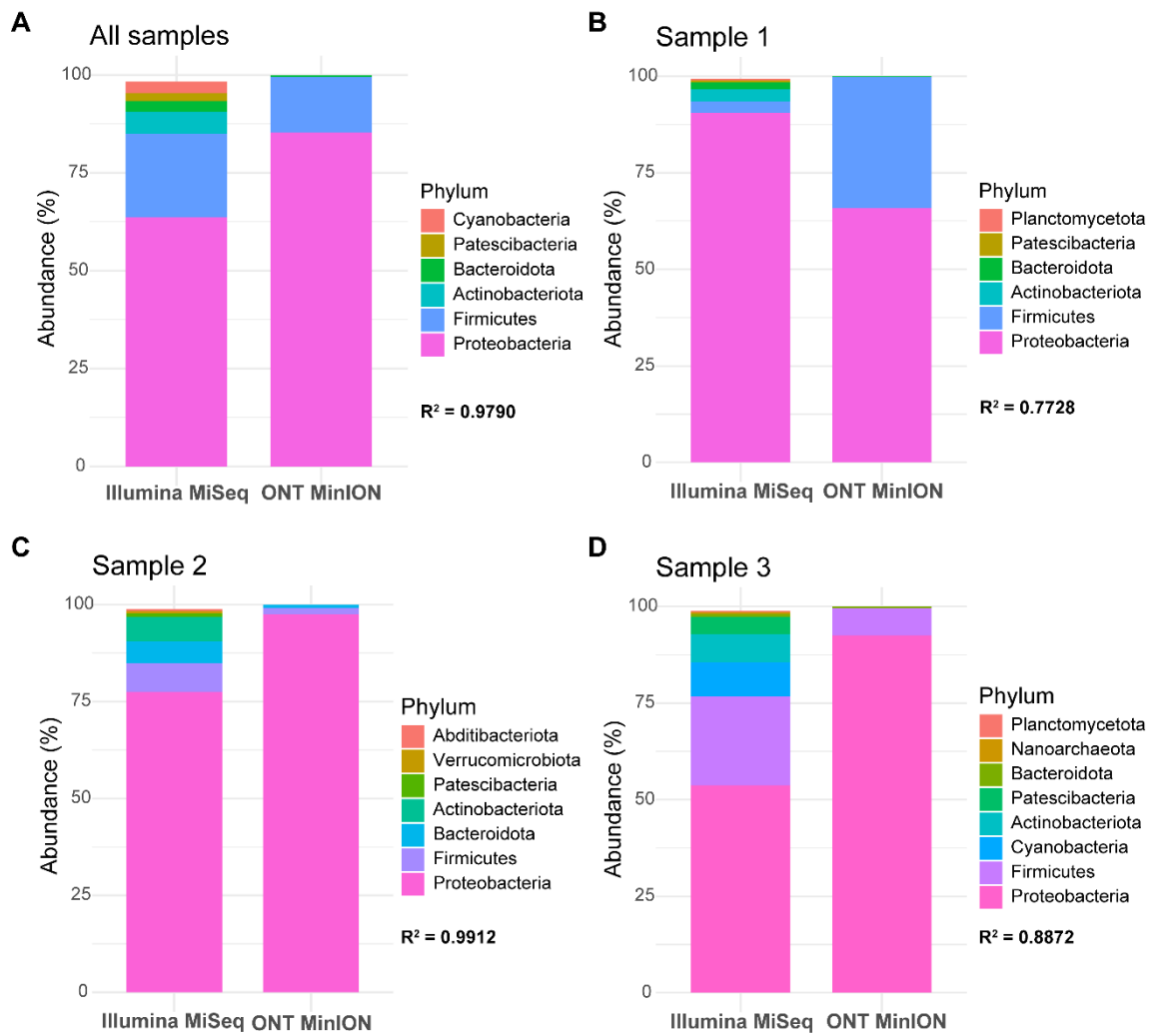| Phylum	PCR4	PCR5
Proteobacteria	85.30042051	96.88892776
Firmicutes	14.22259342	2.048680748
Actinobacteriota	0.024211079	0.384833875
Bacteroidota	0.381837708	0.559126165
Cyanobacteria	0	0.101555222
Others	0.006886186	0.016876231

546

Figure legends

548 **Figure 1**. Correlation plots between the expected standard distribution of the mock

549 community (X axis) and the relative abundances detected in our sequencing output (Y

550 axis). Mock1 was sequenced using the standard PCR conditions described in the original

551 protocol (PCR1, Table 2), whereas Mock2 was sequenced using the conditions optimized

552 for gilthead sea bream mucosal samples (PCR4, Table 2).

Mock1

R² = 0.8803

Mock2
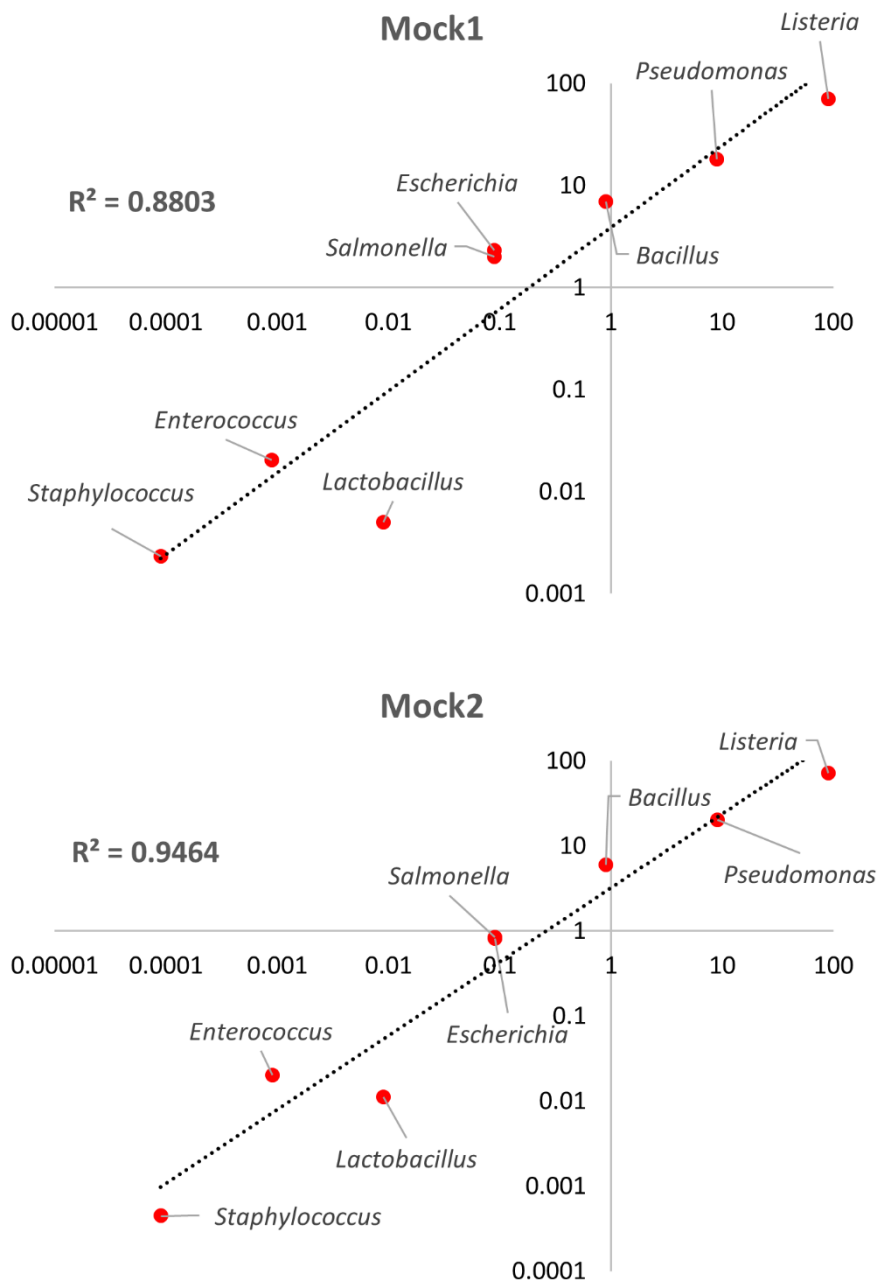
R² = 0.9464

Figure 2. Relative abundance in percentage of gilthead sea bream gut microbiota phyla (> 0.25% in abundance in at least one sample) in samples sequenced with the Illumina Miseq and the ONT-MinION platforms. A) Average of the same three samples sequenced in each platform. No significant differences were found between phyla (Mann-Whitney U test, p > 0.05). B-D) Individual profiles obtained for each of the three samples

sequenced in the two platforms. Correlation coefficients (R^2 values) are shown for each

comparison.

561