

Quantifying dominant bacterial genera detected in metagenomic data from fish eggs and larvae using genus-specific primers

Babak Najafpour, Patricia I.S. Pinto, A.V.M. Canario and Deborah M. Power*

Centro de Ciências do Mar (CCMAR/CIMAR), Universidade do Algarve, 8005-139 Faro, Portugal

*** Correspondence:**

Deborah M. Power

dpower@ualg.pt

Abstract

The goal of this study was to design genus-specific primers for rapid evaluation of the most abundant bacterial genera identified using amplicon-based sequencing of the 16S rRNA gene in fish-related samples and surrounding water. Efficient genus-specific primers were designed for eleven bacterial genera including *Alkalimarinus*, *Colwellia*, *Enterovibrio*, *Marinomonas*, *Massilia*, *Oleispira*, *Phaeobacter*, *Photobacterium*, *Polarbacterium*, *Pseudomonas*, and *Psychrobium*. The specificity of the primers was confirmed by the phylogeny of the sequenced polymerase chain reaction (PCR) amplicons that indicated primers were genus-specific except in the case of *Colwellia* and *Phaeobacter*. Copy number of the 16S rRNA gene obtained by quantitative PCR using genus-specific primers and the relative abundance obtained by 16S rRNA gene sequencing using universal primers were well correlated for the five analyzed abundant bacterial genera. Low correlations between quantitative PCR and 16S rRNA gene sequencing for *Pseudomonas* were explained by the higher coverage of known *Pseudomonas* species by the designed genus-specific primers than the universal primers used in 16S rRNA gene sequencing. The designed genus-specific primers are proposed as rapid and cost-effective tools to evaluate the most abundant bacterial genera in fish-related or potentially other metagenomics samples.

Keywords: abundant microbiota, genus-specific primers, aquaculture site, eggs, larvae, qPCR, 16S rRNA gene

1- Introduction

The influence of symbiotic and pathogenic interactions of bacterial microbiota on terrestrial and aquatic vertebrates is of high interest (Sharpton, 2018). The advent of cost-effective next-generation sequencing (NGS) has opened up new avenues of culture-free and high throughput analysis of the entire microbiota in any ecosystem and has significantly modified understanding of their role in animal health and disease (Cao *et al.*, 2017). Since the 16S rRNA gene is present in all bacteria, it is the most common reference gene for studies of bacterial phylogeny and taxonomy, and also for studies of the composition and the relative proportion of microorganisms in a given habitat (Janda and Abbott, 2007; Simon and Daniel, 2011). The structure of the 16S rRNA gene explains its versatility for metagenomics since universal primers can be designed in regions that are highly conserved across species and the intervening hypervariable regions can be used to assign operational taxonomic units (OTU) to the genus taxon (Baker *et al.*, 2003; Wang and Qian, 2009). Furthermore, the hypervariable regions offer the opportunity for the development of genus and even species targeted quantitative PCR (qPCR).

Aquatic organisms are exposed to a ubiquitous and abundant microbiota and the intensification of aquaculture has increased interest in characterizing the microbiota of fish. Initially traditional microbiological approaches based on *in vitro* culture were used, but more recently metagenomics approaches have been deployed (Martínez-Porchas and Vargas-Albores, 2017). Aquaculture differs from terrestrial farming systems as there is a much larger number of species exploited and this is coupled to a wide variety of environmental conditions (e.g. temperature, salinity) and geographical locations (Casa-Resino *et al.*, 2021). The variability of fish microbiota has been linked to geography, species, and environmental conditions, and indicates that fish microbiomes have a degree of farm site-specificity (Najafpour *et al.*, 2021). Metagenomics studies targeting fish, indicate that *Vibrio* and *Pseudomonas* are the dominant bacterial genera (reviewed by Egerton *et al.*, 2018). However, the exclusive dependence on relative abundance data from NGS can lead to misinterpretation of microbial community structure (Jian *et al.*, 2020). For this reason, it has been proposed that the use of genus-specific primers can provide complementary, quantitative data, to corroborate NGS results and contribute to better understanding microbial diversity and population structure (Zhou *et al.*, 2014). A general literature review of microbiome studies in fish reveals that genus-specific primers for the most representative bacterial genera with high relative abundance are unavailable. Although genus-specific primers exist for the detection of pathogen-containing genera such as *Aeromonas*, *Vibrio*, *Edwardsiella*, and *Streptococcus*, their use has not been correlated with 16S rRNA metagenomic profiles in fish (Zhang *et al.*, 2014).

We previously generated 16S rRNA metagenomics datasets for Gilthead seabream (*Sparus aurata*) and European seabass (*Dicentrarchus labrax*) eggs from several commercial production sites in Europe and identified the profile of the main bacterial genera (Najafpour *et al.*, 2021). The objective of the present study was to develop a quick, cost-effective, and practical approach for large-scale screening of the core microbiome during aquaculture production cycles. We report the design of genus-specific primers, exploiting the hypervariable characteristics of the 16S rRNA gene, for the dominant bacterial genera represented in our in-house metagenomic 16S rRNA gene datasets from eggs, larvae, live feed, and tank water samples from seabream and seabass aquaculture sites in Europe (Najafpour *et al.*, 2021). The efficiency and specificity of the genus-specific primers were confirmed by quantitative PCR (qPCR) and sequencing of the PCR amplicons, and coverage of each genus was validated by comparison of qPCR and metagenomics data (Najafpour *et al.*, 2021).

2- Experimental procedures

2.1 Selection of target genera

The summarized workflow for genus choice and primer design is presented in Figure 1. The most represented bacterial genera in seabream and seabass hatcheries in Europe were identified using in-house metagenomic datasets of 16S rRNA gene sequences obtained from eggs, larvae, live feed, and hatchery water (sequenced by Lifesequencing S.L.-ADM, Spain and Stab Vida, Lda, Portugal). The corresponding 16S rRNA gene sequences of target genera were obtained from the LPSN and Silva (SSU r138.1) databases (Quast *et al.*, 2013; Parte *et al.*, 2020).

2.2 Sequence alignment and primer design

Multiple sequence alignments of the retrieved 16S rRNA gene sequences from the LPSN and Silva databases were performed using the MUSCLE algorithm (Edgar, 2004) in the Aliview platform v 1.27 (Larsson, 2014). Genus-specific forward (Fw) and reverse (Rv) primers were designed manually to obtain primers that amplified the maximum number of species in each of the target bacterial genera. The size of the amplicon was set at between 85 to 250 bp. Criteria used for PCR primer selection included the melting temperature (T_m), percentage of GC content, GC clamp, secondary structure, and the tendency to form primer-dimers. Primers were analyzed and optimized using the oligonucleotide sequence calculator, OligoEvaluator™ (Sigma-Aldrich, Germany, <http://www.oligoevaluator.com>) and OligoAnalyzer™ Tool (<https://www.idtdna.com/calc/analyzer>). In general, the threshold for primer selection included $\Delta G > -9$ to minimize the likelihood of self-dimers, hairpins, and heterodimers and 50 - 55 % GC content for both the Fw and Rv primers to favor specific annealing to the targeted templates (Table A1). Primers that did not comply with the selection criteria were rejected, apart from the primers for *Enterovibrio* for which it was not possible to meet all the criteria (Table A1).

To further confirm that the designed genus-specific primer pairs would anneal to the maximal number of species in a given genus, in-silico PCR simulations using TestPrime 1.0, available in the SILVA platform, was used (Klindworth *et al.*, 2013). The primer pairs with the maximum specificity for each taxonomic group and with the best match to the selection criteria were selected and synthesized (Table A2, Specanalitica, Carcavelos, Portugal). On arrival, primers were resuspended in sterile, nuclease-free water to prepare 100 μ M stocks (Table A2).

2.3 Evaluation of primer specificity by PCR amplification

The performance of the genus-specific primers was initially evaluated by running conventional PCR (Bio-Rad, T100 Thermal Cycler). Genomic DNA was extracted from seabream and seabass eggs, whole larvae, and rotifers (feed) using a DNeasy Blood & Tissue Kit (Qiagen, Germany) as previously described (Najafpour *et al.*, 2021). The PCR was carried out using 2 μ l (approximately 40 - 80 ng) of genomic DNA, 2.5 μ l of 10X Dream Taq Green Buffer containing 20 mM $MgCl_2$ (Thermo Scientific, Lithuania), 0.5 μ l of dNTPs (10 μ M stock), 1.25 μ l of the Fw primer (10 μ M), 1.25 μ l of the Rv primer (10 μ M), 0.2 μ l of Dream Taq DNA polymerase (5 U/ μ l, Thermo Scientific, Lithuania), and 17.3 μ l of sterile, nuclease-free water to give a final reaction volume of 25 μ l. The PCR thermocycle consisted of 1 cycle of 95°C for 3 min followed by 34 cycles of 95°C for 10 s, a gradient of melting temperatures tested for each primer pair (57 - 64°C) for 10 s and 72°C for 10 s and a final cycle at 72 °C for 5 min. The genus-specific primers were tested in PCR amplification of genomic DNA extracted from the larval intestine, whole larvae, rotifers, and egg samples of seabream and seabass that had a high relative abundance of a given bacterial genus in 16S rRNA metagenomics.

The 16S rRNA gene PCR reaction products were run on a 2% agarose gel in TAE buffer and the specific amplicons generated by each primer pair were purified using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare, UK) following the manufacturer's instructions. The purified PCR products were ligated into the pGEM-T Easy cloning vector (Promega, Madison, USA) which permits colony selection based on white/blue color and resistance to the antibiotic ampicillin. The ligation reactions were performed overnight at 4 °C in a 10 µl reaction mix containing 5 µl 2X Rapid Ligation Buffer, 50 ng/µl pGEM®-T Easy Vector, 1.5 U of T4 DNA Ligase (Promega, Madrid, Spain), the optimal concentration of each of the purified PCR amplicons (between 19.8 - 80.5 ng) and sterile water, to give a final reaction volume of 10 µl. DH5α competent cells (*E. coli*) were transformed with 5 µl of the ligation reaction, plated on LB agar containing 75 µg/ml ampicillin, 0.5 mM IPTG and 80 µg/ml X-Gal, and incubated at 37 °C overnight. Minipreps of plasmid DNA were prepared from white colonies and at least two colonies per genus were sequenced using the Sanger method.

The primer specificity and performance were evaluated by assigning taxonomy to each amplicon sequence using the SILVA 16S rRNA gene database as a reference and phylogenetic tree generation, based on the SILVA workflow (de novo including neighbors). For the phylogenetic tree, the sequences of the selected colonies were aligned using the Randomized Axelerated Maximum Likelihood (RAxML) tool with a GTR model and Gamma rate model for likelihoods (Stamatakis, 2014).

2.4 Real-time qPCR optimization and primer efficiency

After confirming primer specificity by amplicon sequencing, quantitative PCR (qPCR) reactions were optimized using a Bio-Rad CFX96 qPCR Instrument (Bio-Rad Laboratories, Hercules, CA). The primer performance was initially evaluated using a gradient of melting temperatures (T_m) in a reaction volume of 10 µl containing 200 nM of each primer, 2 µl of DNA (80 ng DNA/2 µl) for the samples, or 2 µl of serial dilutions (corresponding to 10^2 to 10^7 template copies in the reaction) for the standard curve, 5 µl of 2X Forget-Me-Not™ EvaGreen® qPCR Master Mix (Biotium) and 2.4 µl of sterile nuclease-free water. Thermocycling conditions were 95 °C for 2 min, followed by 40 cycles of 95 °C for 5 s, the optimized melting temperature for each primer pair for 10 s (between 58 - 61 °C) and 72°C for 10 s. A final melting curve was generated by increasing the temperature up to 95 °C in increments of 0.5 °C every 10 s to confirm single reaction products were obtained. Control reactions included substitution of genomic DNA by water to confirm the absence of contamination.

2.5 Bacterial genus quantification and correlation with 16S rRNA microbiome profiling

To compare qPCR and 16S rRNA gene abundance estimates, Spearman correlations were used together with scatter plots generated using the R package ggplot2 v 3.3.5 (Wickham, 2016). *Colwellia*, *Oleispira*, *Phaeobacter*, *Pseudomonas*, and *Psychrobium* were quantified in nine egg samples (Najafpour et al., 2021, Table A3). *Massilia*, *Phaeobacter*, and *Pseudomonas* were quantified in seabream (n = 9; 3 larvae samples in the age range of 5-15 days post-hatch (dph) and 6 samples in the age range of 43-58 dph) and seabass (n = 15; 7 larvae samples in the age range of 5-7 dph and 8 samples in the age range of 42-46 dph) larvae (Table A3). In the case of *Pseudomonas* where the correlation between 16S rRNA gene sequencing and copy number determined by qPCR was low, further in silico analysis was done to assess genus-specific primer performance in comparison to the universal primers used for 16S rRNA metagenomics studies. In-silico PCR v 0.5.1 implemented in Ubuntu (20.04.2 LTS) and unique *Pseudomonas* 16S rRNA gene sequences from the Silva database with a minimum length of 900 bp were used.

3- Results

3.1 Primer specificity

In total, 11 bacterial genera were targeted based on their high relative abundance in fish-related samples as determined by 16S rRNA gene sequencing (Table 1). The species for which the 16S rRNA genes were readily amplified by each primer pair are indicated in Table A4 and the supplemental figure at <https://doi.org/10.5281/zenodo.6301068>. Primer specificity was confirmed by the presence of a single band in agarose gel electrophoresis (Figure A1) and the sequences of the PCR amplicon. In the phylogenetic tree, most of the 16S rRNA amplicon sequences generated clustered with the corresponding taxa in the SILVA 16S rRNA gene database, confirming the specificity of the genus-specific primers (Figure 2). In the *Pseudomonas* clade, the sequence of the 16S rRNA gene amplicon from two samples clustered with several *Pseudomonas* species and included an uncultured *Pseudomonas* sp. (SILVA ID KP398534) and *Pseudomonas azotifigens* (AB189452; Figure 2a). In the *Massilia* clade, the sequence of the 16S rRNA amplicons from 2 samples clustered with several *Massilia* species including *Massilia niastensis* (EU808005), and an uncultured *Massilia* sp. (AB636963). In the *Psychrobium* clade, the sequence of the 16S rRNA gene amplicons from 2 samples clustered with *Psychrobium* species, including an uncultured *Psychrobium* sp. (KT318702), *Psychrobium conchae* (AB930131; Figure 2a). In the *Phaeobacter* clade, the sequenced 16S rRNA gene amplicons clustered with *Phaeobacter* species, including *Phaeobacter inhibens* (CP010668), and *Sedimentitalea* sp. (JN018499; Figure 2a). In the *Colwellia* cluster, one of the 16S rRNA gene amplicons clustered with unidentified *Colwellia* spp. (L10950 and JX569143; Figure 2a) and the other amplicon clustered with *Thalassotalea* sp. (Figure 2b).

3.2 Primer efficiency and correlation between qPCR and 16S RNA gene abundance

The efficiency of bacterial genus-specific primers was evaluated using qPCR with the optimized annealing temperature of each primer pair (Table 2). All the genus-specific primers had an acceptable efficiency within the range of 92 - 105.5 %.

The five most abundant bacterial genera in seabream and seabass eggs were quantified. In general, the relative abundance profiles of the 16S rRNA gene sequencing and qPCR amplification of different bacterial genera were matched with some exceptions (Figure 3). A significant positive correlation was obtained for the relative abundance (%) of bacterial genera detected by both methods (Figure 4). In egg samples ($n = 9$) highly significant positive correlations were found for *Colwellia* ($r = 0.82$), *Oleispira* ($r = 0.86$), and *Psychrobium* ($r = 0.86$, Figure 4). In seabream ($n = 9$) and seabass ($n = 15$) larvae, *Massilia* ($r = 0.32$) and *Phaeobacter* ($r = 0.83$) abundance also had significant positive correlations (Figure 4). For *Pseudomonas*, no correlation was found between the qPCR results for larvae ($r = -0.063$, $n = 24$) and egg ($r = -0.22$, $n = 9$) samples and the 16S rRNA gene sequencing (Figure 4). In-silico PCR analysis comparing the *Pseudomonas*-specific primers and the universal primers generally used for metagenomics (Klindworth *et al.*, 2013) identified 27,734 and 26,780 unique sequences of the *Pseudomonas* 16S rRNA gene (total number in Silva = 60869, filtered sequences number with the minimum size 900 bp = 41607), respectively.

4- Discussion

Genus-specific primers for rapid and cost-effective high throughput monitoring of core microbial genera of seabream and seabass aquaculture were successfully developed. The potential functional importance of the abundant bacterial genera selected makes further studies about their turnover important due to their high potential impact on aquaculture production.

Several studies have focused on *Pseudomonas* spp. due to their widespread distribution and the presence of species that are human, animal, and plant pathogens (Palleroni, 2015). For example, *P. baltica* is a pathogen of marine fish, and primers targeting the *gyrB* gene (c390-F1 and c390-R1) and *rpoD* gene have been developed for rapid diagnosis (López *et al.*, 2016). A multiplex PCR based on *oprI* and *oprL* genes was developed for the detection of *Pseudomonas* strains from a bacterial collection isolated from water (Matthijs *et al.*, 2013). The qPCR comparisons, in the present study, of the abundance of five bacterial genera (*Colwellia*, *Oleispira*, *Pseudomonas*, *Psychrobium*, *Phaeobacter*) in the egg samples revealed a higher than expected copy number of the *Pseudomonas* genus relative to the results of the 16S rRNA metagenomic sequencing. The *Pseudomonas* genus-specific primer pair designed in the present study were of broad scope and had the potential to anneal to 226 out of the 254 16S rRNA gene sequences represented in the LPSN and SILVA databases. Insight into why a higher than expected copy number of *Pseudomonas* was detected by qPCR came from in-silico PCR comparisons of the *Pseudomonas* genus-specific and the 16S rRNA gene universal primers (Silva database, SSU r138.1) since it revealed the universal primers (Klindworth *et al.*, 2013) had lower efficiency and annealed to fewer of the *Pseudomonas* sequences represented in the database. Furthermore, the relatively short amplification product generated by the *Pseudomonas* genus-specific primers favors high qPCR efficiencies compared to previous primers designed to amplify a larger DNA sequence encompassing the ITS1 region (Locatelli *et al.*, 2002). Bergmark *et al.* (2012) reported *Pseudomonas* and *Burkholderia* genus-specific primers for their detection in soil and proposed the use of genus-specific primers for both qPCR and 16S rRNA gene sequencing approaches to overcome issues related to the resolution of the 16S rRNA gene databases, although the exponential increase in available bacterial 16S rRNA gene sequences has now ameliorated this problem (Glöckner, 2019).

A high relative abundance of *Oleispira* and *Colwellia* genera was observed in our in-house database of 16S rRNA gene sequences obtained for egg and water samples from aquaculture installations. The copy number of *Oleispira* and *Colwellia* in the same samples determined by qPCR using genus-specific primers and their relative abundance using 16S rRNA gene sequencing gave a strong positive correlation. However, amplicon sequencing of products generated by the *Colwellia* primers yielded a *Colwellia*-specific amplicon and another amplicon that matched the *Thalassotalea* genera of the Colwelliaceae family. The similarity between the sequence of the *Colwellia* genus-specific primers and the 16S rRNA gene sequence of *Thalassotalea* (LPSN database) should be considered when using the primers. *Oleispira* and *Colwellia* are marine hydrocarbon-degrading bacteria (Mason *et al.*, 2014) and obligate hydrocarbonoclastic bacteria (OHCB) and are usually present in the environment in very low numbers (Golyshin *et al.*, 2010). Pollution or the addition of hydrocarbons to water induces a rapid bloom of OHCB (Kasai *et al.*, 2002) and indicates that variations in *Oleispira* and *Colwellia* abundance may be a useful indicator of excess hydrocarbons in aquaculture systems due, for example, to addition of lipid enriched feeds.

In seabream and seabass eggs, *Pseudophaeobacter* was relatively more abundant than *Phaeobacter*, which was not detected by 16S rRNA gene sequencing in eight out of nine egg samples (Najafpour *et al.*, 2021). In contrast, *Phaeobacter* was more common in 16S rRNA gene sequences of seabream and seabass larvae, and the *Pseudophaeobacter* genus was not detected (in-house database). *Phaeobacter* spp. are common in marine organisms and the environment (Martens *et al.*, 2006; Yoon *et al.*, 2007; Zhang *et al.*, 2008) and have been proposed as a probiotic against pathogenic *Vibrio* spp. in seabass and cod, *Gadus morhua* (D'Alvise *et al.*, 2013; Grotkjær *et al.*, 2016), although in juvenile squid they were associated with mortality (Won, 2009). The abundance-dependent detection of *Phaeobacter* or *Pseudophaeobacter* in seabream and seabass samples highlights one of the challenges when designing primers for short amplicon targets of 16S rRNA genes that show high similarity across bacterial genera. The amplification of *Phaeobacter* or *Pseudophaeobacter* by the genus-specific qPCR primers is unsurprising considering the recent reclassification of the species in *Leisingera-Phaeobacter* to

Sedimentitalea and *Pseudophaeobacter* (Breider *et al.*, 2014). Therefore, it is proposed that the *Phaeobacter* genus-specific primers designed in our study be designated group-specific primers.

The primers for *Psychrobium* were highly specific and this genus was present in all analyzed samples of eggs, larvae, and environmental water and a positive correlation existed between the results of qPCR and 16S rRNA gene sequencing. The lowest number of species were detected for the *Psychrobium* genus compared to the other targeted genera and it was easier to design primers specific for this genus. The high specificity of the *Psychrobium* primers was assigned to the high variability of the *Psychrobium* 16S rRNA gene compared to other bacteria.

Six of the bacterial genera, *Massilia*, *Marinomonas*, *Polaribacter*, *Alkalimarinus*, *Enterovibrio*, and *Photobacterium*, for which genus-specific primers were designed, had low relative abundance in the 16S rRNA metagenome of eggs samples (Najafpour *et al.*, 2021). Nonetheless, the identified genera were well represented in seabream and seabass larvae, rotifers, and tank water, and therefore genus-specific primers were validated in qPCR. The designed genus-specific primers revealed a high relative abundance of the *Massilia* genus in seabream and seabass larvae, *Marinomonas* in rotifer, and *Alkalimarinus*, *Polaribacter*, *Photobacterium*, and *Enterovibrio* in larval seabream intestine. These genera have previously been reported in a wide diversity of samples and experiments. For example, *Massilia* spp. are widespread in soil (Zhang *et al.*, 2006), drinking water (Gallego *et al.*, 2006), and plants (Ofek *et al.*, 2012) and was a dominant genus in the gastrointestinal microbiota of the herbivorous grass carp, *Ctenopharyngodon idellus* (Li *et al.*, 2014). *Marinomonas* species are abundant in seawater (Ivanova *et al.*, 2005; Yoon *et al.*, 2005) and were highly abundant in the microbial community of rotifer cultures prepared as a feed for fish larvae (Rombaut *et al.*, 2001). *Polaribacter* species have been isolated from Antarctic soil and in biofilms on stones from the North Sea (Kim *et al.*, 2013; Choo *et al.*, 2020), and the genus was detected in the intestine of marine organisms and algae (Nedashkovskaya *et al.*, 2013; Hyun *et al.*, 2014; Wei *et al.*, 2018). *Alkalimarinus sediminis* was isolated from marine sediment in Shandong Province, China (Zhao *et al.*, 2015), and *Alkalimarinus* species were found in bone-eating worms (*Osedax mediterranea*) from the Mediterranean Sea (Hewitt *et al.*, 2020). Healthy and diseased *Dentex* and *Sparus aurata* (bony fishes) cultured in Spanish Mediterranean aquaculture contained *Enterovibrio coralii* strains and possibly *Enterovibrio nigricans* (Pascual *et al.*, 2009). *Photobacterium* has been isolated from seawater, mussel, eggs of spiny lobster and fish intestine and the bioluminescence and pathogenicity of some species (e.g. *Photobacterium damsela*) has made studies of them a priority (Osorio *et al.*, 1999; Labella *et al.*, 2017; Egerton *et al.*, 2018).

Conclusion

Genus-specific primers were developed for *Pseudomonas*, *Massilia*, *Psychrobium*, *Phaeobacter*, *Marinomonas*, *Polaribacter*, *Alkalimarinus*, *Enterovibrio*, *Photobacterium*, *Oleispira*, and *Colwellia*. Most of the designed primers were highly efficient in qPCR and were genus-specific as shown by amplicon sequencing and phylogeny. The difficulty was encountered in the design of genus-specific primers for *Phaeobacter* and *Pseudophaeobacter* since the short amplicons of the 16S rRNA gene encompassed regions with 93-98% identity and so these primers were designated group-specific. The genus-specific primers designed in this study will be useful for rapid evaluation and quantification by qPCR of target bacterial genera in fish-related samples and potentially other metagenomic samples since the genus targeted are ubiquitous in a diversity of environments. An interesting observation was that the 16S rRNA gene universal primers used to profile microbiomes do not cover all identified species of some bacterial genera such as *Pseudomonas* and this has the potential to create bias in metagenomics studies and reinforces the value of complimentary qPCR studies. Overall, the designed genus-specific primers provided a rapid and cost-effective evaluation of abundant bacterial genera in

samples and can therefore contribute to the understanding of the modulation of abundant microbiota in complex microbial communities with a potentially high impact on host biology.

Author Contributions

Babak Najafpour: Formal analysis-Lead, Investigation-Lead, Methodology-Lead, Writing – original draft-Lead; **Patricia Pinto:** Formal analysis-Supporting, Methodology-Supporting, Writing – review & editing-Supporting; **Adelino V. M. Canario:** Methodology-Supporting, Project administration-Supporting, Resources-Supporting, Supervision-Supporting, Writing – review & editing-Supporting; **Deborah M. Power:** Conceptualization-Lead, Formal analysis-Supporting, Funding acquisition-Lead, Investigation-Lead, Methodology-Supporting, Project administration-Lead, Resources-Equal, Supervision-Lead, Writing – original draft-Supporting, Writing – review & editing-Lead.

Acknowledgments

The authors acknowledge all the aquaculture hatcheries for supplying the experiment samples. The authors also thank João C. R. Cardoso of the Comparative Endocrinology and Integrative Biology group, CCMAR, for his helpful comments. European Union Horizon 2020 Programme (PerformFISH, grant n° 727610) funded this work and the Ph.D. grant to B.N. This output reflects only the authors' views and the European Union cannot be held responsible for any use that may be made of the information contained herein. Additional funding was obtained from the Portuguese Foundation for Science and Technology (FCT) through project UID/Multi/04326/2020 and researcher contract DL57/2016/CP1361/CT001 to P.I.S.P. and from CRESC Algarve 2020 and COMPETE 2020 through the project EMBRC.PT ALG-01-0145-FEDER-022121. AQUAEXCEL3.0 funding from the European Union's Horizon 2020 research and innovation program (grant No. 871108).

Conflict of Interests

None declared.

Ethics statement

None required

Data Availability Statement

All data are provided in full in this paper apart from amplicon sequences of 16S rRNA genes which are available at www.ncbi.nlm.nih.gov under accession numbers OM685062-OM685080. The supplemental figure is available in Zenodo at <https://doi.org/10.5281/zenodo.6301068> (The alignment and identity of 16S rRNA genes and genus-specific primer sequences for different bacterial species of each target bacterial genus).

Reference

- Baker, G.C., Smith, J.J., and Cowan, D.A. (2003) Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.
- Bergmark, L., Poulsen, P.H.B., Al-Soud, W.A., Norman, A., Hansen, L.H., and Sørensen, S.J. (2012) Assessment of the specificity of *Burkholderia* and *Pseudomonas* qPCR assays for detection of these genera in soil using 454 pyrosequencing. *FEMS Microbiol Lett* **333**: 77–84.
- Breider, S., Scheuner, C., Schumann, P., Fiebig, A., Petersen, J., Pradella, S., et al. (2014) Genome-scale data suggest reclassifications in the *Leisingera*-*Phaeobacter* cluster including proposals for *Sedimentitalea* gen. nov. and *Pseudophaeobacter* gen. nov. *Front Microbiol* **5**: 416.
- Cao, Y., Fanning, S., Proos, S., Jordan, K., and Srikumar, S. (2017) A review on the applications of next generation sequencing technologies as applied to food-related microbiome studies. *Front Microbiol* **8**.
- Casa-Resino, I.D.L., Empl, M.T., Villa, S., Kolar, B., Fabrega, J., Lillicrap, A.D., et al. (2021) Environmental risk assessment of veterinary medicinal products intended for use in aquaculture in Europe: the need for developing a harmonised approach. *Environ Sci Eur* **2021 331** **33**: 1–17.
- Choo, S., Borchert, E., Wiese, J., Saha, M., Künzel, S., Weinberger, F., and Hentschel, U. (2020) *Polaribacter septentrionalilitoris* sp. nov., isolated from the biofilm of a stone from the North Sea. *Int J Syst Evol Microbiol* **70**: 4305–4314.
- D’Alvise, P.W., Lillebø, S., Wergeland, H.I., Gram, L., and Bergh, Ø. (2013) Protection of cod larvae from vibriosis by *Phaeobacter* spp.: A comparison of strains and introduction times. *Aquaculture* **384–387**: 82–86.
- Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.
- Egerton, S., Culloty, S., Whooley, J., Stanton, C., and Ross, R.P. (2018) The gut microbiota of marine fish. *Front Microbiol* **9**: 873.
- Gallego, V., Sánchez-Porro, C., García, M., and Ventosa, A. (2006) *Massilia aurea* sp. nov., isolated from drinking water. *Int J Syst Evol Microbiol* **56**: 2449–2453.
- Glöckner, F.O. (2019) The SILVA Database Project: An ELIXIR core data resource for high-quality ribosomal RNA sequences. *undefined* **3**.
- Golyshin, P.N., Ferrer, M., Chernikova, T.N., Golyshina, O. V., and Yakimov, M.M. (2010) *Oleispira*. *Handb Hydrocarb Lipid Microbiol* 1755–1763.
- Grotkjær, T., Bentzon-Tilia, M., D’Alvise, P., Dourala, N., Nielsen, K.F., and Gram, L. (2016) Isolation of TDA-producing *Phaeobacter* strains from sea bass larval rearing units and their probiotic effect against pathogenic *Vibrio* spp. in *Artemia* cultures. *Syst Appl Microbiol* **39**: 180–188.
- Hewitt, O.H., Díez-Vives, C., and Taboada, S. (2020) Microbial insights from Antarctic and

Mediterranean shallow-water bone-eating worms. *Polar Biol* 2020 4310 **43**: 1605–1621.

- Hyun, D.-W., Shin, N.-R., Kim, M.-S., Kim, P.S., Jung, M.-J., Kim, J.Y., et al. (2014) *Polaribacter atrinae* sp. nov., isolated from the intestine of a comb pen shell, *Atrina pectinata*. *Int J Syst Evol Microbiol* **64**: 1654–1661.
- Ivanova, E.P., Onyshchenko, O.M., Christen, R., Lysenko, A.M., Zhukova, N. V., Shevchenko, L.S., and Kiprianova, E.A. (2005) *Marinomonas pontica* sp. nov., isolated from the Black Sea. *Int J Syst Evol Microbiol* **55**: 275–279.
- Janda, J.M. and Abbott, S.L. (2007) 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *J Clin Microbiol* **45**: 2761–2764.
- Jian, C., Luukkonen, P., Yki-Järvinen, H., Salonen, A., and Korpela, K. (2020) Quantitative PCR provides a simple and accessible method for quantitative microbiota profiling. *PLoS One* **15**: e0227285.
- Kasai, Y., Kishira, H., Sasaki, T., Syutsubo, K., Watanabe, K., and Harayama, S. (2002) Predominant growth of *Alcanivorax* strains in oil-contaminated and nutrient-supplemented sea water. *Environ Microbiol* **4**: 141–147.
- Kim, B.C., Oh, H.W., Kim, H., Park, D.S., Hong, S.G., Lee, H.K., and Bae, K.S. (2013) *Polaribacter sejongensis* sp. nov., isolated from Antarctic soil, and emended descriptions of the genus *Polaribacter*, *Polaribacter butkevichii* and *Polaribacter irgensii*. *Int J Syst Evol Microbiol* **63**: 4000–4005.
- Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., and Glöckner, F.O. (2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* **41**: e1.
- Labella, A., Arahal, D., Castro, D., Lemos, M., and Borrego, J. (2017) Revisiting the genus *Photobacterium*: taxonomy, ecology and pathogenesis. *Int Microbiol* **20**: 1–10.
- Larsson, A. (2014) AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**: 3276–3278.
- Li, J., Ni, J., Li, J., Wang, C., Li, X., Wu, S., et al. (2014) Comparative study on gastrointestinal microbiota of eight fish species with different feeding habits. *J Appl Microbiol* **117**: 1750–1760.
- Locatelli, L., Tarnawski, S., Hamelin, J., Rossi, P., Aragno, M., and Fromin, N. (2002) Specific PCR amplification for the genus *Pseudomonas* targeting the 3' half of 16S rDNA and the whole 16S-23S rDNA spacer. *Syst Appl Microbiol* **25**: 220–227.
- López, J.R., Lorenzo, L., Marcelino-Pozuelo, C., Marin-Arjona, M.C., and Navas, J.I. (2016) *Pseudomonas baetica*: pathogenicity for marine fish and development of protocols for rapid diagnosis. *FEMS Microbiol Lett* **364**: fnw286.
- Martens, T., Heidorn, R., Pukall, M., Simon, B., Tindall T., and Brinkhoff (2006) Reclassification of *Roseobacter gallaeciensis* Ruiz-Ponte et al. 1998 as *Phaeobacter gallaeciensis* gen. nov., comb. nov., description of *Phaeobacter inhibens* sp. nov., reclassification of *Ruegeria algicola* (Lafay

- et al. 1995) Uchino et al. 1999 as Marinovu. *Int J Syst Evol Microbiol* **56**: 1293–1304.
- Martínez-Porchas, M. and Vargas-Albores, F. (2017) Microbial metagenomics in aquaculture: a potential tool for a deeper insight into the activity. *Rev Aquac* **9**: 42–56.
- Mason, O.U., Han, J., Woyke, T., and Jansson, J.K. (2014) Single-cell genomics reveals features of a *Colwellia* species that was dominant during the Deepwater Horizon oil spill. *Front Microbiol* **5**.
- Matthijs, S., Coorevits, A., Gebrekidan, T.T., Tricot, C., Wauven, C. Vander, Pirnay, J.P., et al. (2013) Evaluation of *oprI* and *oprL* genes as molecular markers for the genus *Pseudomonas* and their use in studying the biodiversity of a small Belgian River. *Res Microbiol* **164**: 254–261.
- Najafpour, B., Pinto, P.I.S., Moutou, K.A., Canario, A.V.M., and Power, D.M. (2021) Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream. *Microorg 2021, Vol 9, Page 2275* **9**: 2275.
- Nedashkovskaya, O.I., Kukhlevskiy, A.D., and Zhukova, N. V. (2013) *Polaribacter reichenbachii* sp. nov.: A new marine bacterium associated with the green alga *ulva fenestrata*. *Curr Microbiol* **66**: 16–21.
- Ofek, M., Hadar, Y., and Minz, D. (2012) Ecology of root colonizing *Massilia* (Oxalobacteraceae). *PLoS One* **7**: e40117.
- Osorio, C.R., Collins, M.D., Toranzo, A.E., Barja, J.L., and Romalde, J.L. (1999) 16S rRNA gene sequence analysis of *Photobacterium damsela* and nested PCR method for rapid detection of the causative agent of fish Pasteurellosis. *Appl Environ Microbiol* **65**: 2942.
- Palleroni, N.J. (2015) *Pseudomonas*. *Bergey's Man Syst Archaea Bact* 1–1.
- Parte, A.C., Carbasse, J.S., Meier-Kolthoff, J.P., Reimer, L.C., and Göker, M. (2020) List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* **70**: 5607–5612.
- Pascual, J., Macián, M.C., Arahal, D.R., Garay, E., and Pujalte, M.J. (2009) Description of *Enterovibrio nigricans* sp. nov., reclassification of *Vibrio calviensis* as *Enterovibrio calviensis* comb. nov. and emended description of the genus *Enterovibrio* Thompson et al. 2002. *Int J Syst Evol Microbiol* **59**: 698–704.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590.
- Rombaut, G., Suantika, G., Boon, N., Maertens, S., Dhert, P., Top, E., et al. (2001) Monitoring of the evolving diversity of the microbial community present in rotifer cultures. *Aquaculture* **198**: 237–252.
- Sharpton, T.J. (2018) Role of the gut microbiome in vertebrate evolution. *mSystems* **3**: e00174-17.
- Simon, C. and Daniel, R. (2011) Metagenomic analyses: past and future trends. *Appl Environ Microbiol* **77**: 1153–1161.

- Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.
- Wang, Y. and Qian, P.Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PLoS One* **4**: e7401.
- Wei, N., Wang, C., Xiao, S., Huang, W., Lin, M., Yan, Q., and Ma, Y. (2018) Intestinal microbiota in large yellow croaker, *Larimichthys crocea*, at different ages. *J World Aquac Soc* **49**: 256–267.
- Wickham, H. (2016) ggplot2: elegant graphics for data analysis. Springer-Verlag New York.
- Won, B.W. (2009) Characterizing the role of *Phaeobacter* in the mortality of the squid, *Euprymna scolopes*. *Honor Sch Theses*.
- Yoon, J.H., Kang, S.J., Lee, S.Y., and Oh, T.K. (2007) *Phaeobacter daeponensis* sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. *Int J Syst Evol Microbiol* **57**: 856–861.
- Yoon, J.H., Kang, S.J., and Oh, T.K. (2005) *Marinomonas dokdonensis* sp. nov., isolated from sea water. *Int J Syst Evol Microbiol* **55**: 2303–2307.
- Zhang, D., Zhang, Q., and Li, A. (2014) Development of a multiplex PCR assay for rapid and simultaneous detection of four genera of fish pathogenic bacteria. *Lett Appl Microbiol* **59**: 471–478.
- Zhang, D.C., Li, H.R., Xin, Y.H., Liu, H.C., Chi, Z.M., Zhou, P.J., and Yu, Y. (2008) *Phaeobacter arcticus* sp. nov., a psychrophilic bacterium isolated from the Arctic. *Int J Syst Evol Microbiol* **58**: 1384–1387.
- Zhang, Y.Q., Li, W.J., Zhang, K.Y., Tian, X.P., Jiang, Y., Xu, L.H., et al. (2006) *Massilia dura* sp. nov., *Massilia albidiflava* sp. nov., *Massilia plicata* sp. nov. and *Massilia lutea* sp. nov., isolated from soils in China. *Int J Syst Evol Microbiol* **56**: 459–463.
- Zhao, J.X., Liu, Q.Q., Zhou, Y.X., Chen, G.J., and Du, Z.J. (2015) *Alkalimarinus sediminis* gen. nov., sp. nov., isolated from marine sediment. *Int J Syst Evol Microbiol* **65**: 3511–3516.
- Zhou, L., Li, H., Zhang, Y., Han, S., and Xu, H. (2014) Development of genus-specific primers for better understanding the diversity and population structure of *Sphingomonas* in soils. *J Basic Microbiol* **54**: 880–888.

Tables

Table 1- Most abundant bacterial genera detected by *16S rRNA* metagenomics across different sample types

Bacterial genus	Maximum relative abundance of each bacterial genus in different samples (%) [□]						
	Fish egg	Fish larvae	Fish intestine	rotifer	artemia	algae	water
<i>Pseudomonas</i>	3.8	5.2	5.4	2.1	14.8	1.3	-
<i>Massilia</i>	-	36.1	1.2	-	-	2.2	89.6
<i>Psychrobium</i>	31.3	8.6	3.7	-	-	-	33.4
<i>Phaeobacter</i>	-	3.4	-	-	-	-	-
<i>Marinomonas</i>	1.4	9	2.7	24.2	6.4	-	14.6
<i>Polaribacter</i>	11.8	13.6	94.7	-	-	4.7	19.2
<i>Alkalimarinus</i>	1.9	3.2	29.4	-	-	-	-
<i>Enterovibrio</i>	-	-	94.4	-	-	-	-
<i>Photobacterium</i>	1.5	3.2	21.4	-	-	-	40.8
<i>Oleispira</i>	6.6	2.2	1.7	-	0.3	-	8.3
<i>Colwellia</i>	9.5	4.8	1	-	-	-	20.6

□ = Identified in the 10 top bacterial genera in each sample type, (-) signifies not detected within the most abundant bacterial genera of the specific sample type.

Table 2- List of bacterial genus-specific primers designed and optimized for amplification of 16S rRNA genes in fish

Bacterial genus		Primer	Size (bp)	T _m (°C)	Eff. (%)	R ²
<i>Pseudomonas</i>	Pseu-F	ACCGCATACGTCCTACGG	250	61	99.9	0.99
	Pseu-R	CGAAGACCTTCTTCACACACG				
<i>Massilia</i>	Mass-F	GCGTAGAGATGTGGAGGAAC	142	61	96.4	1.0
	Mass-R	RACCCRACAACCTAGTAGACATCG				
<i>Psychrobium</i>	Psyc-F	GGAGGAACTCTGATGCAGC	128	61	97.7	0.99
	Psyc-R	GTCCTTCTTCTGCGAGTAACG				
<i>Phaeobacter</i>	Phae-F	CACGTAGGCGGATCAGAAAG	149	61	97.6	0.99
	Phae-R	GCCACTGGTGTTCTCCG				
<i>Marinomonas</i>	Mari-F	GAAGCACCGGCTAACTCTG	140	58.6	95.8	0.99
	Mari-R	GTGCMATTCCAAGGTTGAG				
<i>Polaribacter</i>	Pola-F	CTGGTTGACTTGAGTCATATGG	85	58	98.1	1.0
	Pola-R	CGCAATCGGTATTCTGTG				
<i>Alkalimarinus</i>	Alka-F	CGTAGGTGGTTTGTTAAGCGAG	165	61	95.9	1.0
	Alka-R	GTCCAGTAAGTCGCCTTCG				
<i>Enterovibrio</i>	Entv-F	GTGAGTAATGGCTGGGAACC	163	58	105.5	0.97
	Entv-R	CTTGGTGAGCCATTACCTCAC				
<i>Photobacterium</i>	Phot-F	TRGCCAGGTGRGATTAG	167	58	92	1
	Phot-R	GGCTGCATCAGGGTTTCC				
<i>Oleispira</i>	Olei-F	CGGCTAATTTAGTGCCAG	171	58.2	95.2	1.0
	Olei-R	CCACTAACCTCTCTCGTACTC				
<i>Colwellia</i>	Colw-F	ATACGAGGGGTGCAAGCG	188	61	98.3	1.0
	Colw-R	GATGTTCCCTCCAATCTCTACGC				

Figure legends

Figure 1- The workflow followed for the design of the bacterial *16S rRNA* genus-specific primers for the most abundant genera in the microbiome of fish larvae, fish eggs, zooplankton-, phytoplankton and water. The most abundant genus detected in each dataset was ranked using an in-house 16S rRNA gene database. The egg microbiome profile obtained from 16S rRNA gene sequencing has previously been reported (Najafpour et al., 2021).

Figure 2. Analysis of the specificity of the eleven designed genus-specific primer pairs by building a phylogenetic tree using the 16S rRNA gene amplicon sequences and sequences retrieved from the SILVA database. **a)** Phylogenetic clustering of the qPCR amplified and sequenced 16S rRNA genes, Phao.C1, Phao.C2, Phao.C3, Pseu.C1, Pseu.C2, Psyc.C1, Psyc.C2, Mari.C1, Mari.C2, Mari.C3, Pola.C1, Colw.C1, Colw.C2, Colw.C3, Entv.C1, Entv.C2, Mass.C1, Mass.C2, Olei.C1, Olei.C2, Phot.C1, Phot.C2, Alka.C1, Alka.C2 (NCBI accession number OM685062- OM685080) and the 16S rRNA gene sequences retrieved from the SILVA database. The phylogeny was performed using the Randomized Axelerated Maximum Likelihood (RAxML) tool with a GTR model and Gamma rate model for likelihoods (Stamatakis, 2014). **b)** Least-common-ancestor (LCA) sequence classification, which classified each amplicon based on the last level of classified taxa using the available sequences in the SILVA database. The alignment quality score and identity (%) for each amplicon alignment using the LCA method are presented. The number of neighbors per query sequence was set at 10.

Figure 3. Analysis of the efficiency of five genus-specific primer sets by comparing the relative abundance of the genus detected in seabream and seabass eggs using 16S rRNA gene sequencing and the copy number determined by qPCR. **a)** the relative abundance (%) of five bacterial genera determined by 16S rRNA gene sequencing in nine egg samples from seabream (SA, n = 6) and seabass (DL, n = 3) from three different aquaculture sites (S1, S2, and S3). **b)** Quantitative analysis of five bacterial genera in nine egg samples from seabream (n = 6) and seabass (n = 3) using qPCR with the designed genus-specific primers. Full details of the experimental design, DNA extraction method and the microbiome profile obtained using 16S rRNA gene sequencing of eggs samples are available in the Najafpour *et al.* (2021). The graphical plots were generated in the R package ggplot2 v 3.3.5.

Figure 4. The correlation between genus-specific copy number established by qPCR quantification and relative abundance of six bacterial genera determined by 16S rRNA gene sequencing of seabream and seabass egg samples. The scatter plots were generated using the “Spearman” method in an R environment. Two sets of samples and databases were used for the correlation analysis. 1) for the genera, *Colwellia*, *Oleispira*, and *Psychrobium*, the correlation analysis was established by comparing the relative abundance of the genera in nine egg samples of seabass (n = 3) and seabream (n = 6) obtained from 16S rRNA gene sequencing (Najafpour *et al.*, 2021) and the copy number determined using genus-specific primers and qPCR. 2) for the genera, *Massilia*, *Phaeobacter*, and *Pseudomonas*, the correlation analysis was established by comparing the relative abundance of the genera in 24 larval samples obtained from 16S rRNA gene sequencing (in-house database) and copy number determined using genus-specific primers and qPCR. The list of samples used in the analysis is available in Table A3.

Appendix

Figure A1. Analysis of PCR products (5 ul) on 2% agarose gel electrophoresis after staining with GreenSafe. The first lane (L) represents the DNA ladder and the size (base pair, bp) of some markers is shown in red on the left-hand side. Each lane contains the amplicon generated using the genus-specific primer pairs (P) and sample type (S): P = *Massilia* (*Mass*), S = 49 days post hatch (dph) seabream larvae; P = *Phaeobacter* (*Phae*), S = 9 dph seabass larvae; P = *Pseudomonas* (*Pseu*), S = 46 dph seabass larvae; P = *Psychrobium* (*Psyc*), S = 39 dph seabream larvae; P = *Alkalimarinus* (*Alka*), S = intestine of seabream larvae; P = *Polaribacter* (*Pola*), S = intestine of seabream larvae; P = *Photobacterium* (*Phot*), S = intestine of seabream larvae; P = *Marinomonas* (*Mari*), S = rotifer; P = *Oleispira* (*Olei*), S = seabream eggs; P = *Enterovibrio* (*Entv*), S = intestine of seabream larvae; P = *Colwellia* (*Colw*), S = seabream eggs. The size of the PCR products amplified with each primer pair is presented in red (e.g., *Mass*, product size 142 bp).

Table A1. Evaluation of primer GC content, dimer and secondary structure formation, and length using OligoAnalyzer™ Tool

Bacterial genus primer pair		GC (%)	Self-dimer (ΔG kcal.mole ⁻¹)	Hetero-dimer (ΔG kcal.mole ⁻¹)	Hairpin (ΔG kcal.mole ⁻¹)	Length (bp)
<i>Pseudomonas</i>	Fw	61.1	-6.68, -0.96	-4.95, -1.34	-2.18, -1.97	18
	Rv	52.4	-6.69, -1.57		-3.24, -3.13	21
<i>Massilia</i>	Fw	55	-3.61, -0.96	-6.35, -1.34	0.39, 1.24	20
	Rv	47.8	-6.84, -0.96		0.01, 0.28	23
<i>Psychrobium</i>	Fw	55	-7.05, -1.47	-6.69, -1.34	-2.86, -1.87	20
	Rv	52.4	-3.61, -0.96		-1.17	21
<i>Phaeobacter</i>	Fw	52.4	-3.61, -0.96	-3.61, -1.34	-1.17	21
	Rv	66.7	-5.02, -1.34		-0.97, -0.41	18
<i>Marinomonas</i>	Fw	57.9	-9.75, -0.96	-6.44, -1.34	-2.04	19
	Rv	50	-7.6, -1.57		-0.73	19
<i>Polaribacter</i>	Fw	45.5	-7.82, -1.34	-3.9, -1.34	-2.42	22
	Rv	50	-3.61, -1.47		-0.27, 0.54	18
<i>Alkalimarinus</i>	Fw	50	-4.85, -0.96	-8.33, -0.96	-0.11, 0.18	22
	Rv	57.9	-3.61, -0.96		-0.52, -0.4	19
<i>Enterovibrio</i>	Fw	55	-3.14, -0.96	-15.48, -1.34	-1.25	20
	Rv	52.4	-6.47, -0.96		-3.04	21
<i>Photobacterium</i>	Fw	55.6	-8.54, -0.96	-8.54, -1.46	0.9, 1.57	18
	Rv	61.1	-7.05, -1.47		-0.87, -0.09	18
<i>Oleispira</i>	Fw	61.1	-7.05, -1.47	-6.36, -1.34	-0.87, -0.09	18
	Rv	52.4	-3.65, -0.96		1.37, 1.41	21

<i>Colwellia</i>	Fw	61.1	-7.05, -0.96	-6.75, -0.96	-0.55, 0.15	18
	Rv	47.8	-3.61, -0.96		0.49, 1.46	23

ΔG = free energy required to break the structure, OligoAnalyzer™ Tool (<https://www.idtdna.com/calc/analyzer>)

Table A2. Details of the genus-specific primers and their preparation for qPCR

Primer target		Amount received (nmol)	Primer Melting temp. (°C)	Optical density (260nm)	Sterile water* (μl)
<i>Pseudomonas</i>	Fw	9.7	57.2	2	97
	Rv	10.2	65.1	2	102
<i>Massilia</i>	Fw	9.5	55.4	2	95
	Rv	7.8	55.5	2	78
<i>Psychrobium</i>	Fw	10.1	56.2	2	101
	Rv	10.5	55.6	2	105
<i>Phaeobacter</i>	Fw	9.7	55.9	2	97
	Rv	12.8	59.7	2	128
<i>Marinomonas</i>	Fw	11.1	56.3	2	111
	Rv	9.9	52.8	2	99
<i>Polaribacter</i>	Fw	9.5	52.8	2	95
	Rv	11.9	50.8	2	119
<i>Alkalimarinus</i>	Fw	9.2	56.6	2	92
	Rv	11.3	55.9	2	113
<i>Enterovibrio</i>	Fw	10	56.0	2	100
	Rv	10.5	55.9	2	105
<i>Photobacterium</i>	Fw	9.7	53.7	2	97
	Rv	12.2	57.2	2	122
<i>Oleispira</i>	Fw	11.6	50.6	2	116
	Rv	11	54.0	2	110
<i>Colwellia</i>	Fw	10.8	58.4	2	108
	Rv	9.7	55.5	2	97

Fw – forward primer, Rv – reverse primer; * μl of sterile nuclease-free water added to the lyophilized primers to give a 100 μM stock

Table A3. List of samples used to correlate qPCR and 16s rRNA gene abundance estimates

Sample [□]	Sample type	Site	Species	Sample code	Sequenced reads*
E1	Egg	site1	seabream	S1.SA.E.AD3	370906
E2	Egg	site1	seabream	S1.SA.E.AD4	513514
E3	Egg	site1	seabream	S1.SA.E.BD3	409804
E4	Egg	site1	seabream	S1.SA.E.BD4	312384

E5	Egg	site2	seabream	S2.SA.E.AD1	417468
E6	Egg	site2	seabream	S2.SA.E.BD1	587762
E7	Egg	site3	seabass	S3.DL.E.BD2	628184
E8	Egg	site3	seabass	S3.SA.E.AD2	616148
E9	Egg	site3	seabass	S3.SA.E.BD2	522126
L1-L3	Larvae 5-15 dph	F, G, I	seabream	SA.1.JM	150859*
L4-L10	Larvae 5-7 dph	A, F, H, E, D, G	seabass	DL.1.JM	72736*
L11-L16	Larvae 39-58 dph	A, B, C, F, G, H	seabream	SA.3.JM	105275*
L17-24	Larvae 42-46 dph	A, C, D, E, F, G, H	seabass	DL.3.JM	55062*

□ Nine egg samples (E1-E9) used for qPCR reactions of five bacterial genera (*Pseudomonas*, *Oleispira*, *Colwellia*, *Psychrobium*, and *Phaeobacter*), and the details of their 16S rRNA gene sequencing profiles are available in Najafpour et al. 2021. “Factors driving bacterial microbiota of eggs from commercial hatcheries of European seabass and gilthead seabream” and the metagenomics raw data were deposited at NCBI SRA (sequence read archive) under project number PRJNA727018. Twenty-four larvae samples (L1-L24) and their corresponding qPCR and 16S rRNA gene sequencing data (available as in-house datasets) were also used to correlate qPCR and 16S rRNA gene abundance estimates of three genera (*Massilia*, *Phaeobacter*, and *Pseudomonas*). * For larvae samples, the average number of the sequenced reads is presented for each age range. dph = days post-hatch.

Table A4. The numbers of potentially amplified and un-amplified bacterial species with each of the designed genus-specific primer set

Primer	Total number of species ^a	Number of amplified species	Un-amplified species
<i>Pseu</i>	254	226	<i>P. luteola</i> (AB681955), <i>P. kirkiae</i> (MK159379), <i>P. ianjinensis</i> (MF083697), <i>P. asuensis</i> (AJ871471), <i>P. caeni</i> (EU620679), <i>P. oryzihabitans</i> (AB681726), <i>P. zeshuii</i> (JN411093), <i>P. duriflava</i> (EU046271), <i>P. litoralis</i> (FN908483), <i>P. psychrotolerans</i> (AJ575816), <i>P. rhizoryzae</i> (MK759856).
<i>Mass</i>	49	48	<i>M. brevitalea</i> (EF546777)
<i>Psyc</i>	1	1	-
<i>Phae</i>	6	5	<i>P. italicus</i> (AM904562)
<i>Mari</i>	33	31	<i>M. ushuaiensis</i> (AJ627909), <i>M. algicida</i> (KR003453)
<i>Pola</i>	26	20	<i>P. marinivivus</i> (KM017972), <i>P. huanghezhanensis</i> (MH791326), <i>P. aestuariivivens</i> (MH791326), <i>P. porphyrae</i> (AB695286), <i>P. lacunae</i> (KJ728849), <i>P. pacificus</i> (KU315476)

<i>Alka</i>	1	1	-
<i>Photo</i>	38	35	<i>P. galathea</i> (FJ457476), <i>P. salinisoli</i> (KP054474), <i>P. halotolerans</i> (AY551089)
<i>Olei</i>	2	2	-
<i>Colw</i>	21	19	<i>C. demingiae</i> (U85845), <i>C. hornerae</i> (JN175346)
<i>Entv</i>	6	5	<i>E. pacificus</i> (JN118548)

Δ = 16S rRNA genes of bacterial species present in the LPSN db (linked to the Silva db); a full list of the sequences used in the alignments is presented in the supplemental figure at <https://doi.org/10.5281/zenodo.6301068>

Bacterial genus

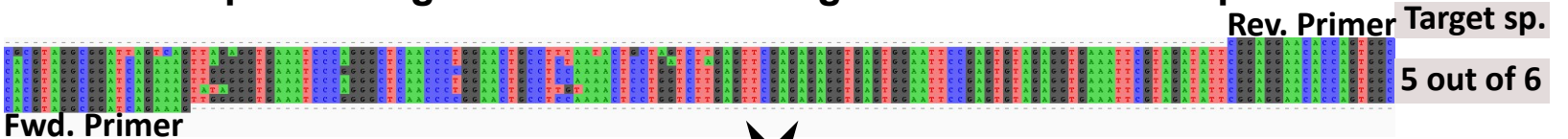
<i>Pseudomonas</i> (pseu)	<i>Massilia</i> (Mass)	<i>Psychrobium</i> (Psych)	<i>Alkalimarinus</i> (Alka)	<i>Marinomonas</i> (Mari)
<i>Polaribacter</i> (Pola)	<i>Oleispira</i> (Olei)	<i>Enterovibrio</i> (Entv)	<i>Photobacterium</i> (Phot)	<i>Phaeobacter</i> (Phae)
		<i>Colwellia</i> (Colw)		



16S rRNA genes of bacteria species searched in LPSN and SILVA db



Sequence alignment of 16S rRNA genes in the Aliview platform



Specificity of the designed primers tested by sequence alignment of different genera and in-silico PCR using TestPrime in SILVA db



The identity of the primers to species-related 16S rRNA genes of each bacterial genus specified in GeneDoc

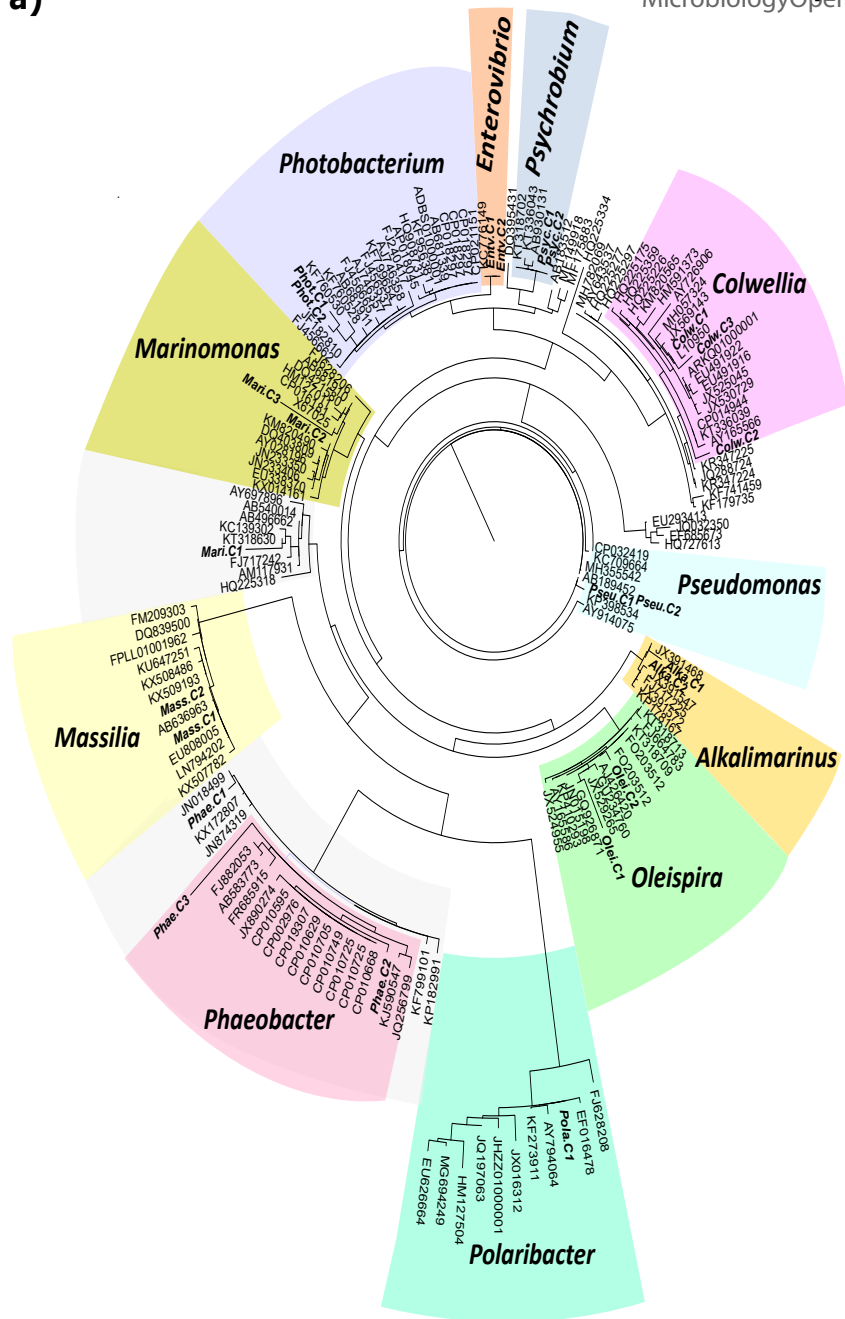


Final evaluation of primer specificity and efficiency carried out by cloning and qPCR on the purified bands of each bacterial genus

a)

MicrobiologyOpen

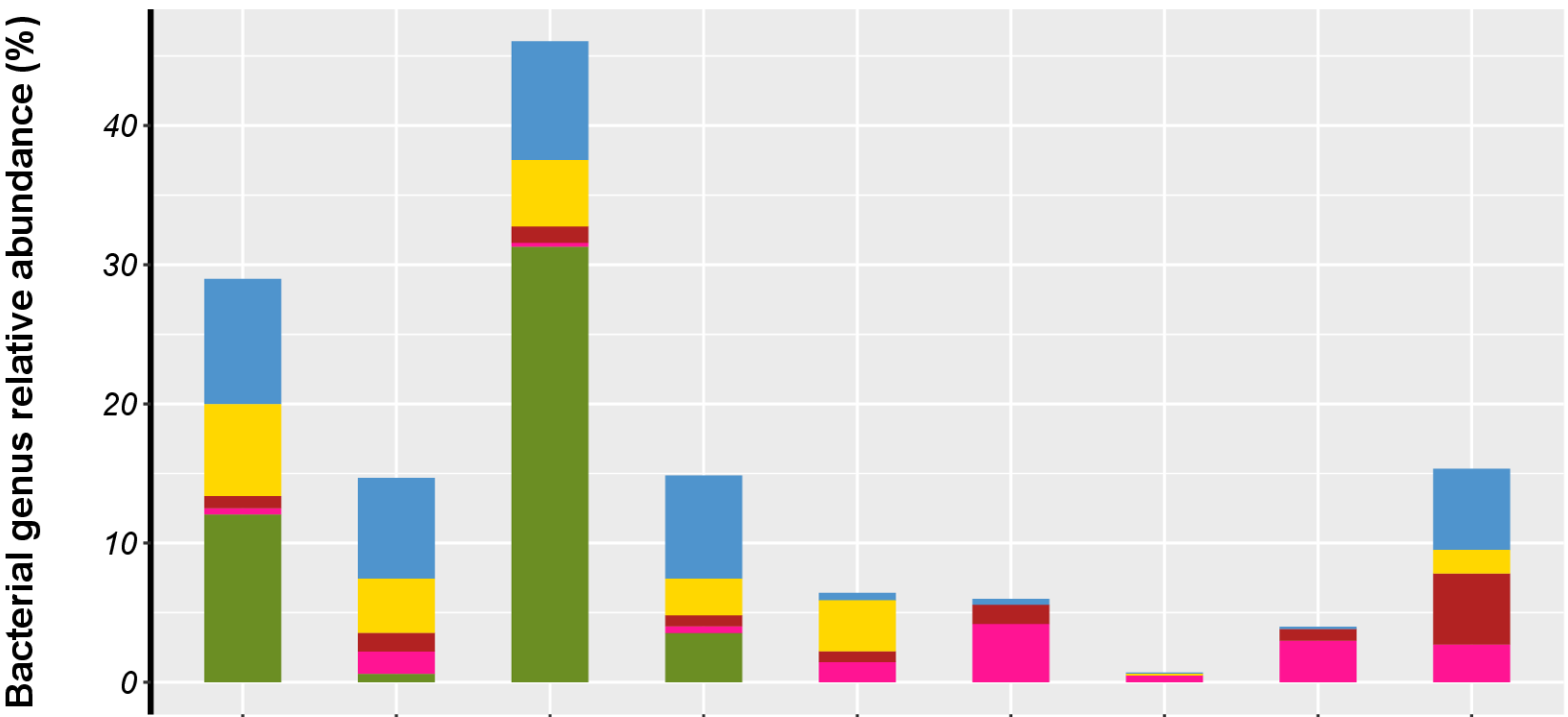
Page 20 of 24



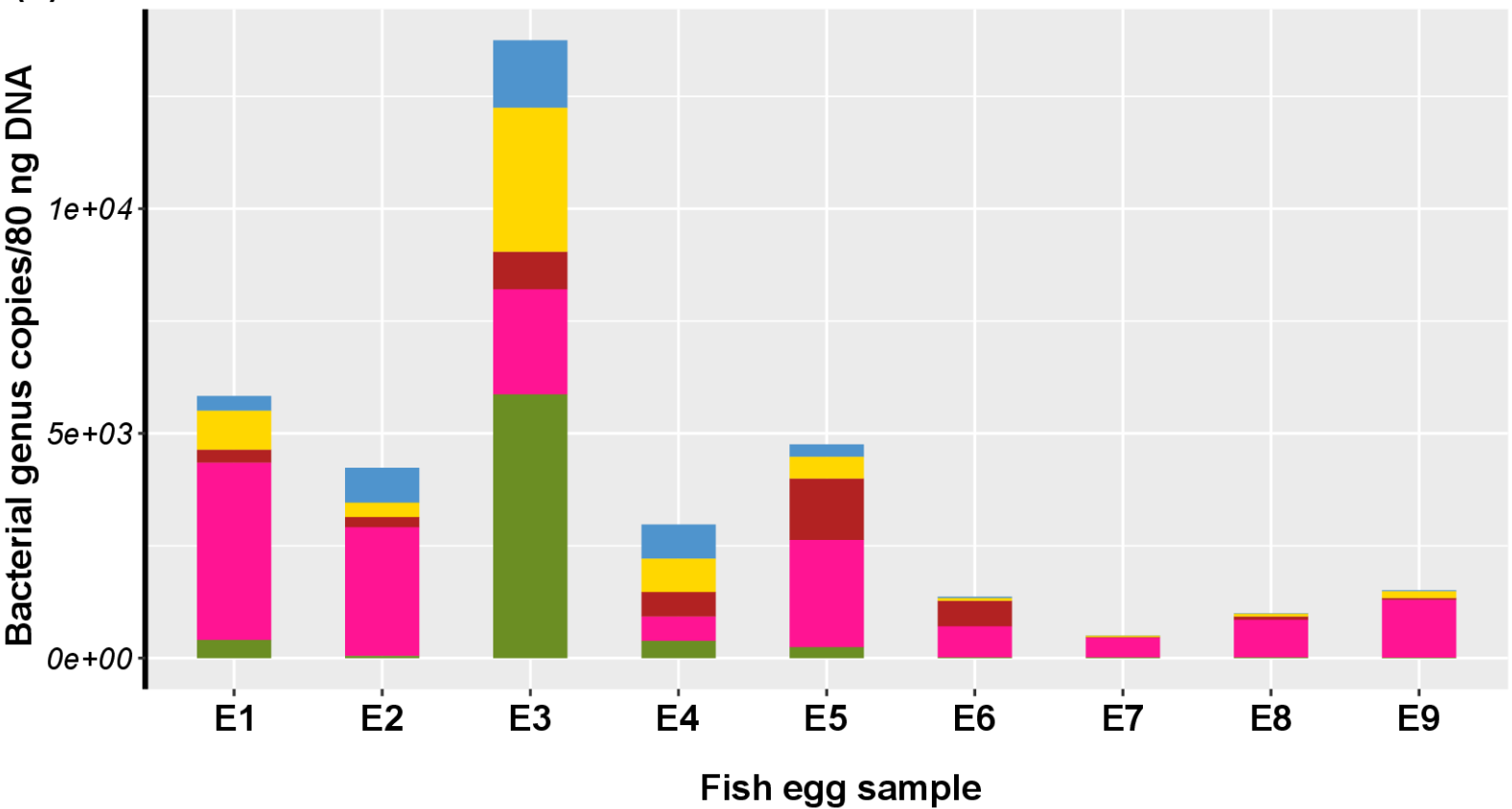
b)

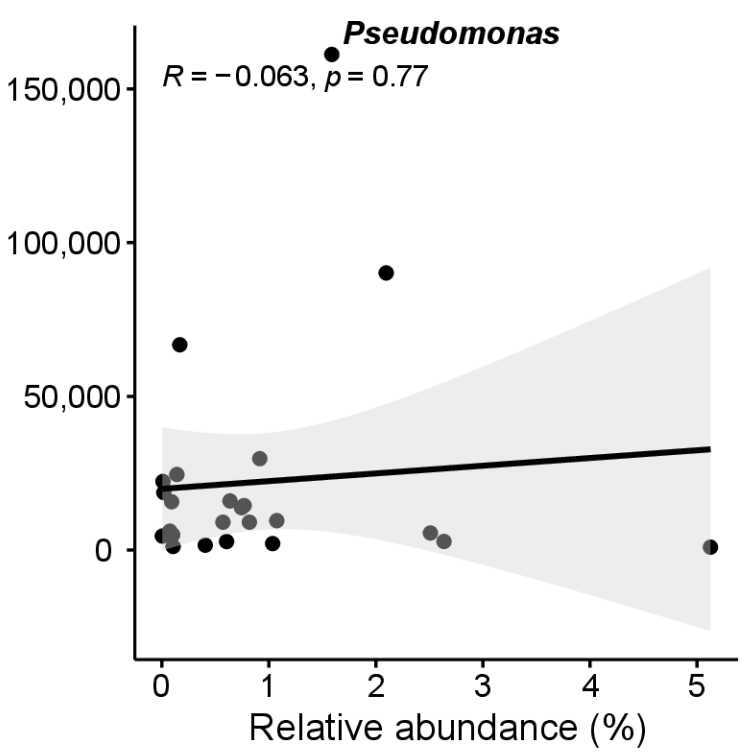
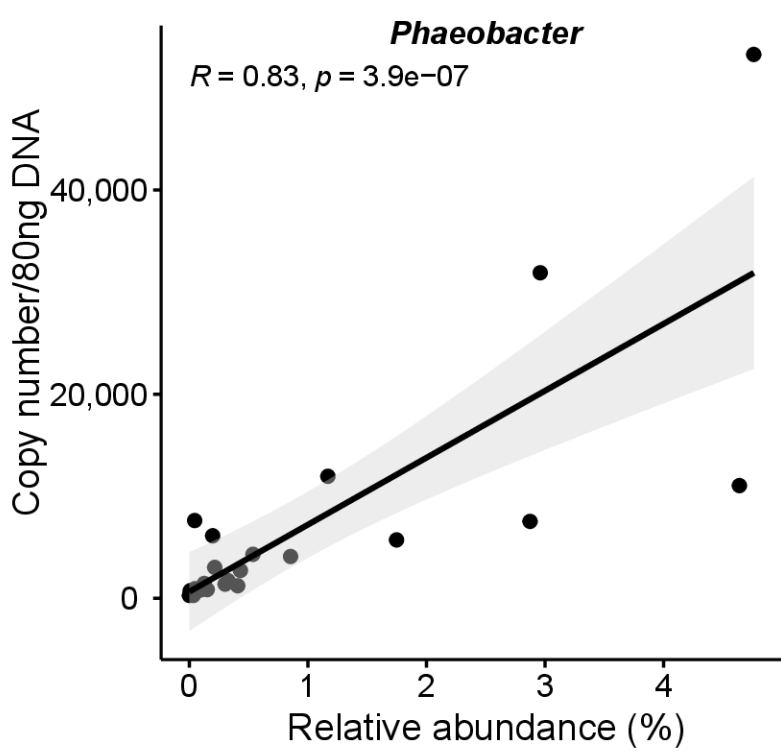
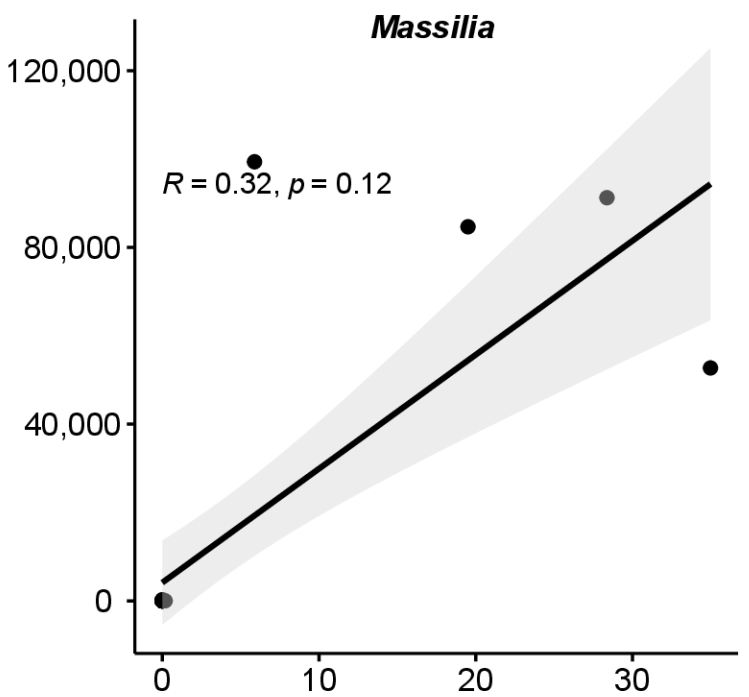
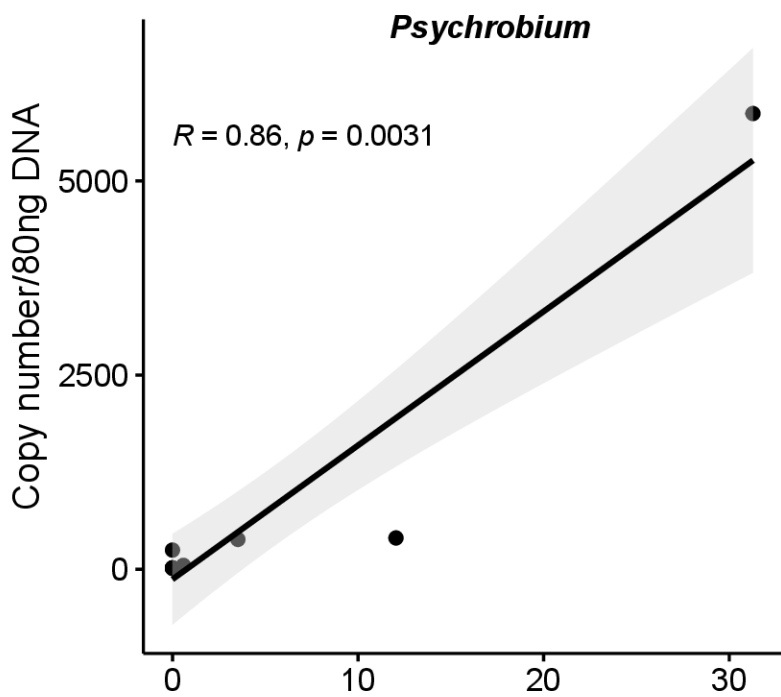
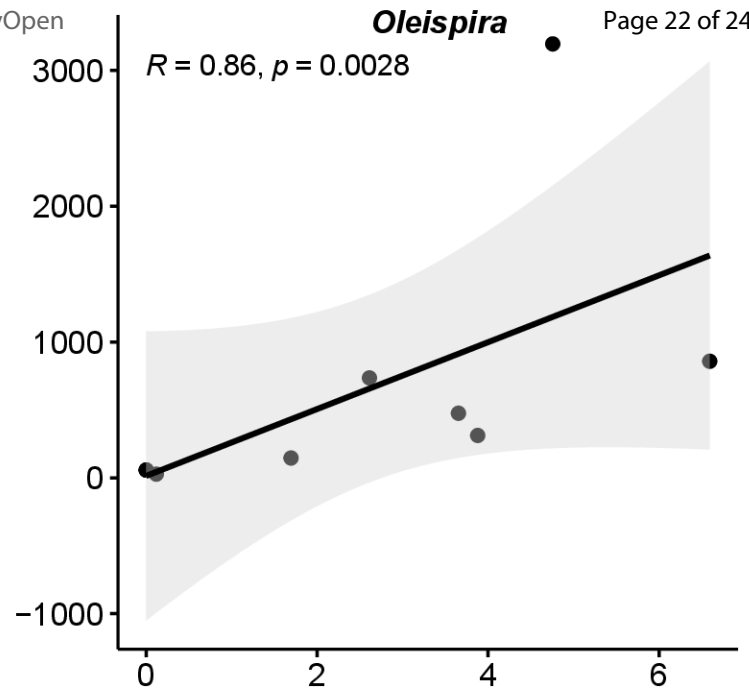
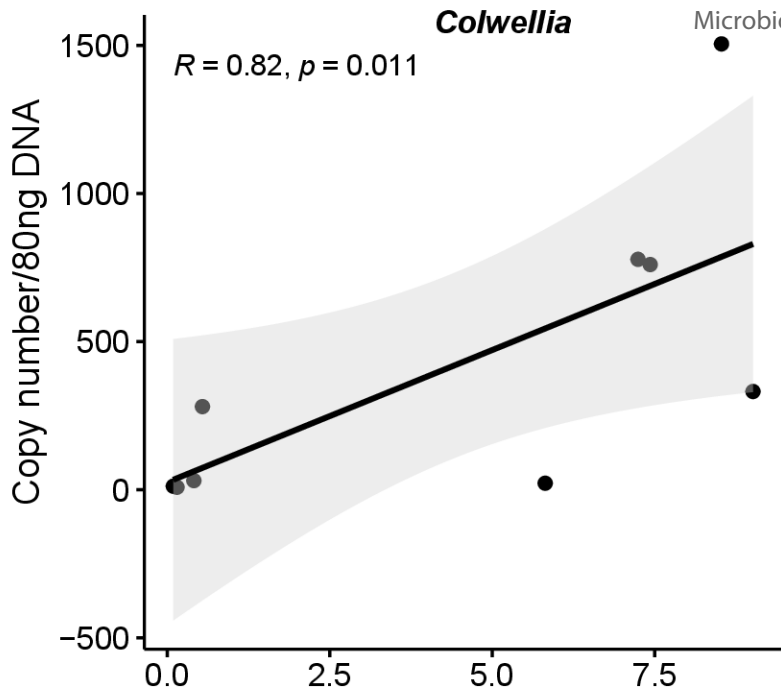
Sequence	Identity	Quality	SILVA taxa
Phae.C1	98.0	97	Rhodobacteraceae
Phae.C2	98.7	98	<i>Phaeobacter</i>
Phae.C3	73.6	57	Unclassified
Pseu.C1	99.6	99	Pseudomonadaceae
Pseu.C2	99.6	99	Pseudomonadaceae
Psyc.C1	99.2	98	Gammaproteobacteria
Psyc.C2	99.2	98	Gammaproteobacteria
Mari.C1	96.4	96	Pseudomonadales
Mari.C2	99.3	99	<i>Marinomonas</i>
Mari.C3	72.3	58	Unclassified
Pola.C1	100.0	100	<i>Polaribacter</i>
Colw.C1	100.0	100	<i>Colwellia</i>
Colw.C2	96.8	99	<i>Thalassotalea</i>
Colw.C3	96.9	93	<i>Colwellia</i>
Entv.C1	85.3	81	<i>Enterovibrio</i>
Entv.C2	84.7	81	<i>Enterovibrio</i>
Mass.C1	100.0	100	<i>Massilia</i>
Mass.C2	100.0	100	<i>Massilia</i>
Olei.C1	98.7	95	<i>Oleispira</i>
Olei.C2	99.4	98	<i>Oleispira</i>
Phot.C1	97.6	96	<i>Photobacterium</i>
Phot.C2	98.8	98	<i>Photobacterium</i>
Alka.C1	98.2	97	<i>Alkalimarinus</i>
Alka.C2	99.4	99	Pseudomonadales

(a)



(b)



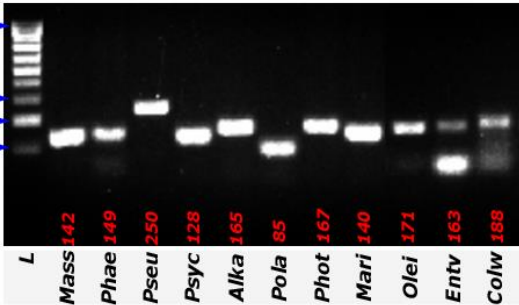


1000

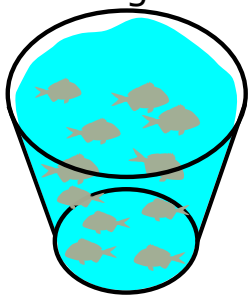
300

200

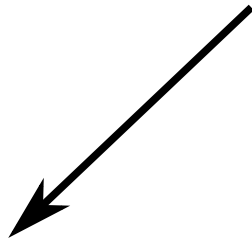
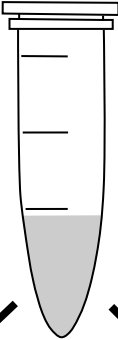
100



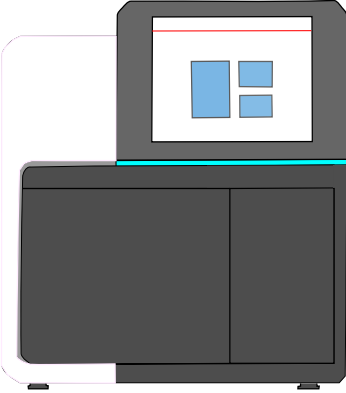
Environmental samples from aquaculture or other ecological niches



DNA extraction



NGS method



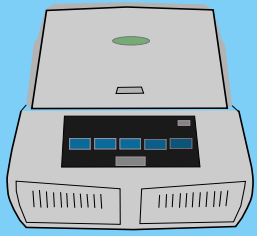
16S rRNA gene sequencing using universal primers



Bioinformatics

Rapid and cost-effective tool for a broader range of samples

PCR method



qPCR on abundant bacteria using genus-specific primer sets

complementary to NGS results

