

## ORIGINAL ARTICLE

# Full-length 16S rRNA gene classification of Atlantic salmon bacteria and effects of using different 16S variable regions on community structure analysis

Terje Klemetsen<sup>1</sup>  | Nils Peder Willassen<sup>1</sup> | Christian René Karlsen<sup>2</sup>

<sup>1</sup>Department of Chemistry, Center for Bioinformatics, UiT The Arctic University of Norway, Tromsø, Norway

<sup>2</sup>Department of fish health, Nofima, Norway

## Correspondence

Terje Klemetsen, Department of Chemistry, Center for Bioinformatics, UiT The Arctic University of Norway, Tromsø, Norway.  
Email: Terje.klemetsen@uit.no

## Funding information

Norges Forskningsråd, Grant/Award Number: 194050/F40; Research Council of Norway, Grant/Award Number: 194050 and F40; UiT The Arctic University of Norway

## Abstract

Understanding fish-microbial relationships may be of great value for fish producers as fish growth, development and welfare are influenced by the microbial community associated with the rearing systems and fish surfaces. Accurate methods to generate and analyze these microbial communities would be an important tool to help improve understanding of microbial effects in the industry. In this study, we performed taxonomic classification and determination of operational taxonomic units on Atlantic salmon microbiota by taking advantage of full-length 16S rRNA gene sequences. Skin mucus was dominated by the genera *Flavobacterium* and *Psychrobacter*. Intestinal samples were dominated by the genera *Carnobacterium*, *Aeromonas*, *Mycoplasma* and by sequences assigned to the order Clostridiales. Applying Sanger sequencing on the full-length bacterial 16S rRNA gene from the pool of 46 isolates obtained in this study showed a clear assignment of the PacBio full-length bacterial 16S rRNA gene sequences down to the genus level. One of the bottlenecks in comparing microbial profiles is that different studies use different 16S rRNA gene regions. Comparisons of sequence assignments between full-length and in silico derived variable 16S rRNA gene regions showed different microbial profiles with variable effects between phylogenetic groups and taxonomic ranks.

## KEYWORDS

Atlantic salmon, Full-length 16S rRNA gene sequence, microbiota

## 1 | BACKGROUND

Second-generation sequencing is widely used to assess the composition of the microbial community through partial sequence analysis of the 16S rRNA gene. Different bacterial 16S rRNA gene regions are used by different researchers, which makes it difficult to perform global comparisons of microbiome studies. Discrepancy in bacterial diversity may also be observed between full-length 16S rRNA gene and variable region 16S rRNA gene datasets (Sun, Jiang, Wu,

& Zhou, 2013; Wagner et al., 2016). Sequence regions but also PCR primer choice used for short-read amplicon sequencing of different 16S rRNA gene hypervariable regions can affect the accuracy of the inferred community profiles and sensitivity to certain bacterial taxa (Chen et al., 2019; Walker et al., 2015). Therefore, sequences of full-length 16S rRNA genes that cover all the variable regions should potentially increase the accuracy and the resolution of closely related taxa. However, full-length sequences compared to shorter sequences generated by other platforms favored Proteobacteria and

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2019 The Authors. *MicrobiologyOpen* published by John Wiley & Sons Ltd.

provided a lower taxonomic profiling of the human feces, partly due to sequence accuracy and low coverage of terminal regions in the 16S rRNA databases (Whon et al., 2018). Still, long-read sequencing such as PacBio circular consensus sequencing (CCS) applied on the 16S rRNA gene provides a promising approach to increase the taxonomic resolution of microbial communities. The CCS technology generates a consensus sequence from a single molecule by reading a ligated circular DNA template multiple times, achieving high read accuracy (Travers, Chin, Rank, Eid, & Turner, 2010). However, few studies have investigated advantages and disadvantages of PacBio CCS for such analyses. A recent study showed that PacBio sequencing error rates were in the same range as Roche 454 and MiSeq platforms (Wagner et al., 2016). The authors reported inconsistencies in species-level analysis between full-length 16S rRNA gene sequences obtained from Sanger and PacBio sequencing comparisons and that more sample types were needed to determine whether partial or full-length 16S rRNA gene sequences was superior in terms of taxonomic profiling effectiveness. The first metagenomic marine environmental samples from PacBio CCS sequences provided a superior taxonomic resolution to the species level compared to in silico derived partial regions of the gene (Pootakham et al., 2017).

In fish, both the skin epithelial surface and the gastrointestinal tract are covered by a mucosal layer. The main components of this mucus layer are secreted glycoproteins called mucins, which are differentially regulated between tissue types in Atlantic salmon, *Salmo salar* (Sveen, Grammes, Ytteborg, Takle, & Jørgensen, 2017). These glycosylated proteins might be a highly attractive substrate for the attachment and settlement of microorganisms. Interaction studies between cutaneous microbiota and the fish surface are scarce, but mutualistic relationships (Beklioglu, Telli, & Gozen, 2006) and the existence of a resilient microbiome (Larsen, Bullard, Womble, & Arias, 2015) has been suggested. Salinity acclimation results in turnover of dominant bacterial taxa within the host microbiome that are unrelated to changes in water microbiota (Schmidt, Smith, Melvin, & Amaral-Zettler, 2015), which is also reported for Atlantic salmon (Karlsen et al., 2017; Lokesh & Kiron, 2016). Microorganisms populating the gastrointestinal tract are believed to take part in digestive functions and contribute to fish health (Nayak, 2010). Many fish gut bacteria are not detected from water samples (Sullam et al., 2012), and gut microbiota profiles in Atlantic salmon change between intestinal compartments (Gajardo, Rodiles, et al., 2016), rearing environments (Dehler, Secombes, & Martin, 2017), and diets (Schmidt, Amaral-Zettler, Davidson, Summerfelt, & Good, 2016). Sequencing the 16S rRNA genes is a powerful tool that provides a comprehensive picture of the phylogenetic diversity and composition of the microorganisms present in a sample as many of the microbial groups are absent or difficult to cultivate. Traditionally, diagnosis of suspected Atlantic salmon bacterial infections has relied on clinical signs and symptoms, and microbiological culture-dependent methods. Many bacteria associated with salmonids have been considered difficult to cultivate and more accurate culture-independent diagnostic procedures are developed (Grove, Reitan, Lunder, & Colquhoun,

2008; Sepúlveda, Bohle, Labra, Grothusen, & Marshall, 2013). However, culturing is an important step to better understand effects of microorganisms. Recovering isolates of microbial symbionts is increasing in focus as they can be used to study activity and functional relationships to a host (Esteves, Amer, Nguyen, & Thomas, 2016; KleinJan, Jeanthon, Boyen, & Dittami, 2017). The present study was also designed to recover and identify members of the Atlantic salmon microbial communities, characterized by 16S rRNA gene sequencing, for future studies.

The changing environment in the salmon production, utilization of different feeds, use of closed or semi-closed recirculation systems and transfer to unprotected seawater environment at the final stage of production affects skin and gut of complex microbial communities. Taxonomic profiling that can reveal alterations or deviations from "normal" microbial communities may advance our understanding of any functional effects of detected microbiota. Appropriate methods that accurately generate and analyze the microbial communities would be an important tool to help improve understanding of microbial effects in the aquaculture industry. Here, we applied long-read technology to demonstrate its utility, as a proof of concept, in characterizing the microbial profiles of the skin surface and the bulk intestinal content of Atlantic salmon.

## 2 | MATERIALS AND METHODS

### 2.1 | Sample collection and DNA extraction

This study utilized fish from an industry research study conducted at 12 ppt salinity with recirculated water with temperature of 13°C at the Research Station for Sustainable Aquaculture (Sunndalsøra, Norway) in accordance with regulations of the Norwegian Food Safety Authority. As proof of concept, six Atlantic salmon were selected for sampling, anesthetized with benzocaine and killed by a blow to the head. The six sampled individuals were split between two dietary groups. Three had been fed fish meal (mean  $\pm$  SEM body weight was 193  $\pm$  41 g), and three were from a diet where fish meal was substituted with krill meal (mean  $\pm$  SEM body weight was 181  $\pm$  20 g). Fish were not fed for 48 hr prior to sampling. Skin mucus samples were obtained by swiping a cell swiper across the left side of the fish and concentrating mucus, which was aspirated by pipetting. The abdominal cavity was then opened and an incision was made to open the distal intestine. The intestinal content was collected by gently scraping the intestine to collect bulk feces including the intestinal mucus layer. All tissue material was stored in 96% EtOH. DNA was extracted from 200 mg skin mucus samples and 100 mg intestinal samples. The protocol was performed using the PowerLyzer® PowerSoil® DNA Isolation Kit (MoBio) according to the manufacturer's specification with the following amendments: samples after adding Solution C1 were heated at 70°C for 10 min. Samples were homogenized with the mechanical bead beater device Precellys®24 (Bertin Technologies) for 1  $\times$  20 s at 5,000 rpm. The DNA was resuspended in 30  $\mu$ l of DNase/RNase free molecular water and concentration determined using a Thermo Scientific Nanodrop 2000c.

## 2.2 | Bacterial isolation and identification

Bacteriology was performed on the same sampled individuals for the total bacterial DNA sequence analysis. Skin mucus (200 mg) and bulk intestinal feces/mucus (100 mg) were separately vortexed and suspended in total of 1 ml 0.9% NaCl saline solution. In addition, gill mucus material and water were similarly suspended in saline solution to retrieve bacterial isolates, included in the Appendix A. A 100  $\mu$ l aliquot of the content was serially diluted up to  $10^{-6}$  and plated in duplicate onto R2A (BD Difco), representing low-nutrient conditions and MacConkey agar (CM007 Oxoid), representing high-nutrient conditions, under aerobic incubation at 12°C for 9 days. Plates were inspected and colony numbers were counted based on morphological characteristics, that is, pigmentation, colony form, elevation, surface appearance, and texture. The relative distribution in percentage between colony morphologies is provided as an average of each sample type. Representative colonies were selected according to dominant morphologies and then identified by 16S rRNA gene sequence analysis. Briefly, representative colonies were selected for purity plating onto R2A or MacConkey plates. Pure colonies determined for freeze stocks were further expanded in Luria-Bertani (LB) broth (Bertani, 1951) with 3.5% NaCl at 12°C before supplementation with 10% glycerol and stored at -80°C. Genomic DNA isolation was performed using PureLink® Genomic DNA Mini Kit (Invitrogen). PCR amplification of the 16S rRNA gene with primers 27F (5'-AGAGTTTGATCMTGGCTCAG) and 1492R (5'-TACCTGTTCAGACTT) was identical to previous descriptions (Karlsen et al., 2014). Products were visualized following agarose (1.0%) gel electrophoresis and RedSafe (ChemBio), before being purified using a QIAquick Gel Extraction Kit (Qiagen) followed by Sanger sequencing (sequenced at GATC Biotech, DNA sequencing services and bioinformatics, Germany). The forward and reverse sequences were assembled in Bioedit (Hall, 1999) and consensus sequences deposited in GenBank (submission: SUB3162162), with accession numbers MG263463-MG263508 (Table A1, Appendix). Sequences were aligned with type strain reference sequences using Sequence Match software from The Ribosomal Database Project II (RDP II) web site (Cole et al., 2014). The phylogenetic relationships between sequences were constructed utilizing selected 16S rRNA sequences of type strains in each genus (Figure A1). Sequences were aligned using the ClustalW algorithm in BioEdit (Hall, 1999). The phylogenetic relationships were determined using maximum likelihood (ML) based on the Tamura 3-parameter model including all coding positions (total of 1,425) with 1,000 bootstrap trials (neighbor-joining tree) in MEGA6 (Tamura, Stecher, Peterson, Filipski, & Kumar, 2013).

## 2.3 | PCR amplification, barcoding, and PacBio sequencing

To analyze the microbial population associated with the skin mucus and intestine, sequencing of the 16S rRNA gene was performed using the PacBio sequencing technology (Pacific Biosciences). The full-length 16S rRNA gene was amplified using

degenerated versions of the universal bacterial 16S rRNA gene primers 27F (5'-AGRGTGTTGATYMTGGCTCAG) and 1492R (5'-GGYTACCTGTTCAGACTT). In accordance with "Guidelines for Using PacBio® Barcodes for SMRT® Sequencing" guide, a 5 nt (5'-GGTAG) padding sequence was added to each unique 16 nt barcode to allow all barcodes to ligate to the SMRTbell™ adapter with equal efficiency. The utilization of barcodes allowed the multiplex sequencing of amplicons from several samples in one library using SMRT®. Primers were synthesized and HPLC-purified as recommended in PacBio's SMRT guidelines by Invitrogen Custom DNA Oligos (Thermo Fisher Scientific). Barcoded 16S rRNA amplicons were obtained by a two-step amplification protocol using Phusion® High-Fidelity PCR Master Mix (Thermo Fisher Scientific). The first PCR was performed in triplicate with the 27 F and 1492 R universal bacterial 16S rRNA gene primers using 100 ng of extracted total DNA, 1  $\times$  Phusion Master Mix, 0.5  $\mu$ mol/L 16S-F forward primer, 0.5  $\mu$ mol/L 16S-R reverse primer, in a 50  $\mu$ l reaction volume. Samples were prepared on ice and amplified in the thermocycler with the block preheated to 98°C. The reactions were performed using the following cycling conditions: preincubation at 98°C for 2 min, followed by 25 cycles of denaturation at 98°C for 10 s, annealing at 55°C for 15 s, elongation at 72°C for 60 s, and a final extension step at 72°C for 3 min. Triplicate samples were pooled, and amplification was verified by 1% agarose gel electrophoresis before the reaction products were purified with an Invitrogen™ PureLink™ PCR Purification Kit. The purified PCR products were diluted and triplicates of 1 ng DNA was used as template for the second amplification reactions to generate padded barcoded products, with reagent concentrations as described above. Product amplification was as above with the following changes: 14 cycles with annealing at 60°C for 15 s. Triplicates were pooled, and products were verified by agarose gel electrophoresis and padded barcoded 16S rRNA gene amplicons from the reaction were purified using PureLink™ PCR Purification Kit and quantified using a NanoDrop spectrophotometer. Purified barcoded amplicons from the 12 samples (skin and intestine from 6 fish) were then pooled in equimolar concentrations, and 250 ng of DNA was used for library preparation at the Norwegian Sequencing Centre ([www.sequencing.uio.no](http://www.sequencing.uio.no)). Briefly, library was prepared using PacBio 2 kb library preparation protocol. Size selection was performed using Ampure beads. Adapters were ligated onto the barcoded amplicons, and the library was sequenced on a PacBio RSII system using the P6-C4 polymerase and chemistry with a 360-min movie time, using one SMRT cell.

## 2.4 | Sequence data analysis

Raw reads were filtered and demultiplexing using RS\_subreads.1 pipeline on SMRT Portal (software version 2.3) with the following settings: minimum number of passes = 1, minimum predicted accuracy = 0.90, and minimum barcode score = 30. Read sequences were then prepared by filtering to a window length between 1,000 and 1,600 nt using PRINSEQ v.0.20.4 (Schmieder & Edwards, 2011) and reoriented in accordance with SILVA v.132 SSU (Pruesse et al., 2007)

with the USEARCH v.10 (Edgar, 2010) orient function. The hyper-variable regions v3-v4 and v5-v6 in the filtered data were extracted after using BLASTN v.2.6.0 (Altschul, Gish, Miller, Myers, & Lipman, 1990) (word size 4, gap opening penalty 0, *E*-value 0.01) to mark the flanking regions as described elsewhere (Pootakham et al., 2017). Sequences positively identified with both v-regions were advanced to generate two new trimmed data subsets for the v3-v4 and v5-v6 regions. Sequences were discarded if a flanking region could not be determined, suggesting a missing or partial v-region for a given sequence.

LCAClassifier (Lanzén et al., 2012) was applied to obtain the taxonomic mapping of the datasets (sample acronyms I1-I6 and SM1-SM6) and their respective data subsets. These were individually aligned against the SilvaMod database by MEGABLAST (Morgulis et al., 2008) (identity cutoff = 75.0, *E*-value cutoff = 0.001) as recommended prior to applying LCAClassifier. Finally, the analysis was carried out using default settings in the LCAClassifier program. Numerical data output from taxonomic classification was ordered based on ranks of taxonomy and the assignments obtained for the 12 datasets and data subsets. This was used to calculate the variation in taxon mapping between full-length PacBio CCS 16S rRNA gene sequences and their respective v-regions. To detail taxonomic variations as principal components linked to the sampling sites of intestine and skin mucus only the full-length 16S rRNA gene sequences were considered. The class level was used due to a 98.92% or greater successful assignment of the filtered full-length sequences of any sample. The sample dataset sizes were downscaled to sample SM2, which had the lowest number of successfully assigned sequences. The scaled values of class data, treating zero as "NA," were uploaded to ClustViz (Metsalu & Vilo, ) with parameters set to not perform row scaling and the method set to SVD with imputation. Rarefaction analysis on the obtained operational taxonomic units (OTUs) was conducted using MicrobiomeAnalyst (Dhariwal et al., 2017). Unscaled OTU counts from mapping of reads with USEARCH were provided alongside their taxonomy lineage to genus level. Filtering of data was set using default parameters. Data were not rarefied or transformed, but total sum scaling was applied. The rarefaction curve was obtained with the filtered data using 20 steps.

Cultivability was determined using the 43,910 pooled full-length sequences combined with the 46 plate-isolated strains. Next, the USEARCH function cluster\_fast was applied to cluster all sequences within an identity threshold of 97%. Sequences in clusters containing at least one of the cultivated strains were included when computing the cultivability percentage. OTUs were also determined for the pooled dataset of 43,910 full-length 16S rRNA gene sequences positive for v3v4 and v5v6 regions. USEARCH was applied following dereplication of the dataset, clustering OTUs at a 97% identity threshold, keeping parameters at default. 31,634 (~72%) of the full-length sequences were successfully reassigned to the 10 OTU representatives found and counted according to sample origin. The sum of sample specific assignments was scaled to fit the smallest sample size of SM2 and incorporated as pie charts in the phylogenetic inference described below.

The OTU centroid sequences were further used to infer phylogenetic relationship between these representatives along with the 46 plate-isolated strains and 10 reference type strains. The 66 sequences were aligned with Clustal Omega v.1.2.1 (Sievers & Higgins, 2014) using default nucleotide parameters. The complete alignment was inferred using NeighborNet network with Uncorrected P distances in SplitsTree v.4.13.1 (Huson & Bryant, 2006).

## 3 | RESULTS

### 3.1 | Characteristics of full-length 16S rRNA gene sequencing

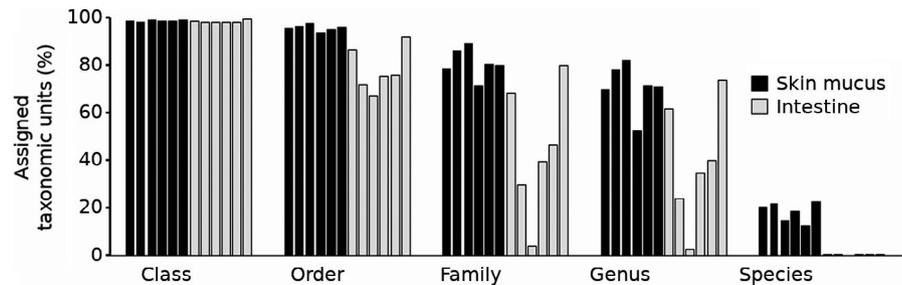
Full-length 16S rRNA gene sequences were amplified from DNA extracted from sampled bulk intestinal content and skin mucus of six Atlantic salmon. Amplicons were generated using a two-step PCR approach with asymmetrical primers during the second round of amplification for a more cost-effective way to multiplex amplicons from several samples. A total of 110,818 reads were obtained with an average read length of 23,881 nt. Of these were 76,723 assembled and demultiplexed into CCS reads with an average accuracy of 98.56% and a mean length of 1,252 nt. The average number of full passes of the CCS reads was 15.6. The number of processed full-length 16S rRNA sequences per sample ranged from 1,483 to 7,634 reads (Table 1). The microbiota composition was determined by a phylotyping approach directly allocating sequences into taxonomic groups based on bitscore and identity threshold. Of the trimmed reads used for assignment, an average of 96.9% of the skin mucus and 78.8% of the intestine were aligned to the taxonomic level order. According to the taxonomic resolution from order to genus, this method assigned more reads to taxonomic references for skin mucus samples compared to intestinal samples (Figure 1). The discriminant taxon, based on the generated reads from the intestine samples was the prominent Clostridiales that became unassigned at higher resolution ranks. The rarefaction analysis (Figure A2) of clustered OTUs showed that the bacterial communities of the intestines are less diverse compared to the skin mucus samples. Convergence were reached (sample I4, I5, I6, SM1, and SM3) but seven out of twelve samples, including both intestine and skin mucus, did not converge properly in the analysis.

### 3.2 | Comparisons of microbial compositions

Hierarchical clustering strengthened by principal component analyses (Figure 2) revealed that bacteria communities clustered to tissue types sampled from the Atlantic salmon in both the taxonomic rank of order to genus. Skin mucus microbiota profiles are tightly clustered. The intestinal microbiota profiles are dispersed but separates into krill meal diet (acronym I1 – I3) with predominate *Aeromonas* and *Mycoplasma* at the genus level, and fish meal diet (acronym I4 – I6) with predominant *Carnobacterium* at the genus level. The most dominant

**TABLE 1** Sample names, acronyms, and PacBio CCS sequence characteristics

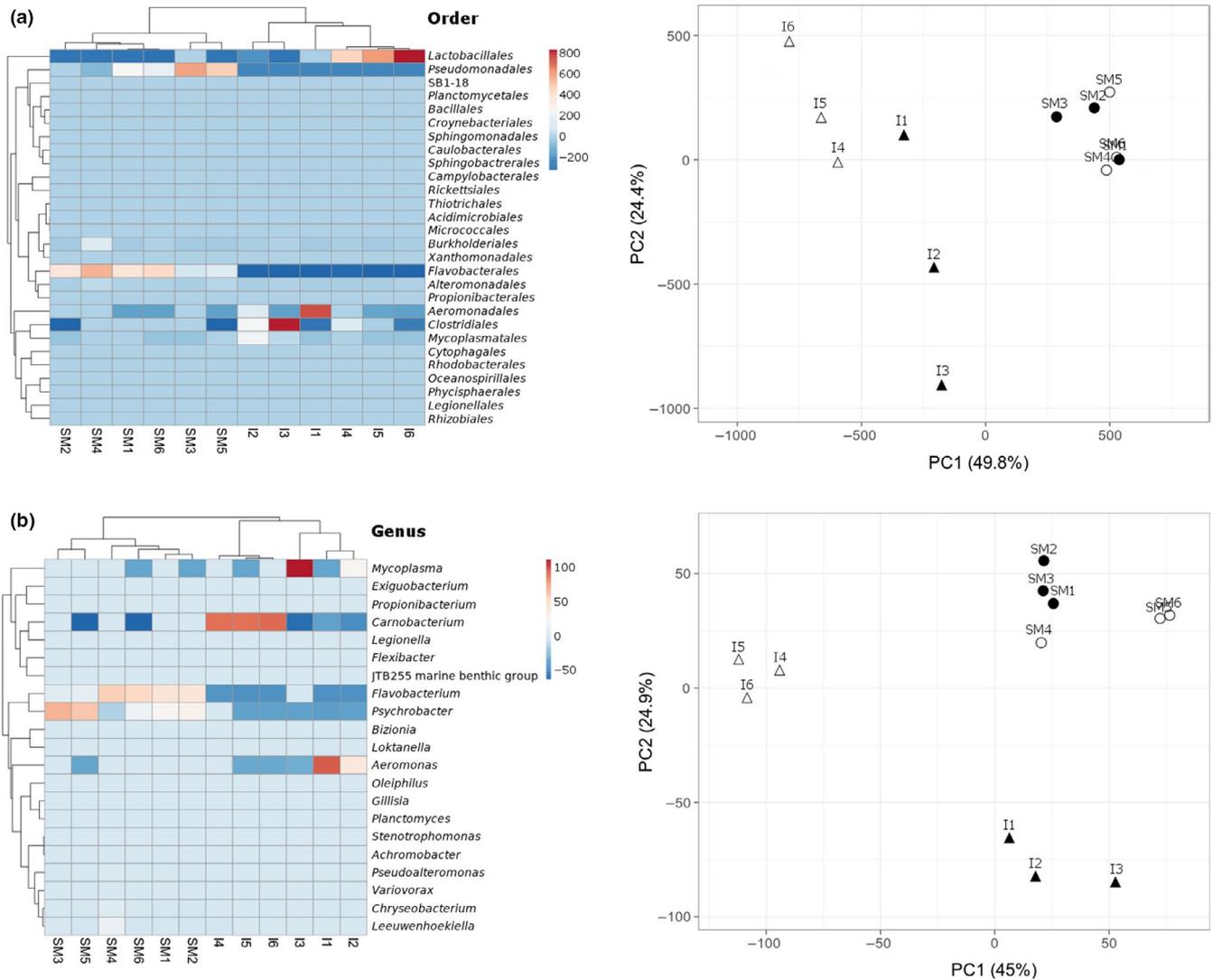
	Sample	Acronym	PacBio CCS reads	Total assignments
Krill meal	Fish 1 Skin Mucus	SM1	4,481	3,675
	Fish 1 Intestine	I1	4,681	4,025
	Fish 2 Skin Mucus	SM2	1,483	1,226
	Fish 2 Intestine	I2	4,810	4,195
	Fish 3 Skin Mucus	SM3	2,816	2,424
	Fish 3 Intestine	I3	7,634	6,664
Fish meal	Fish 4 Skin Mucus	SM4	4,138	3,440
	Fish 4 Intestine	I4	4,971	4,300
	Fish 5 Skin Mucus	SM5	4,022	3,155
	Fish 5 Intestine	I5	5,375	4,581
	Fish 6 Skin Mucus	SM6	4,945	3,960
	Fish 6 Intestine	I6	2,633	2,265

**FIGURE 1** Percentage of taxonomic units assigned as reads at the class, order, family, genus, and species levels. Black bars = sampled fish skin mucus (fish no. 1 – 6). Gray bars = sampled intestine (fish no. 1 – 6)

bacteria orders of the skin mucus samples were Flavobacteriales (51.2%), Pseudomonadales (40.7%), Burkholderiales (2.3%), and Alteromonadales (1.3%). Members of the Flavobacteriales assigned to the genus level were dominated by *Flavobacterium* > *Chryseobacterium* > *Leeuwenhoekiella* > *Bizionia* > *Gillisia*. Genus members within the order of Pseudomonadales were dominated by *Psychrobacter* > *Pseudoalteromonas*. Sequences within remaining orders were not assigned down to genus level. The most dominant bacterial orders of the intestinal samples were Lactobacillales (35.0%) with *Carnobacterium* at the genus level, Clostridiales (26.1%), Aeromonadales (16.0%) with *Aeromonas* at the genus level, Mycoplasmatales (4.1%) with *Mycoplasma* at the genus level. Sequences within Clostridiales were not assigned down to the genus level.

To determine whether longer sequences of the 16S rRNA gene would be advantageous to assign sequences to the lower ranks of taxonomic affiliation, partial sequences spanning the v3v4 and v5v6 regions were extracted in silico from the full-length 16S rRNA gene sequence dataset. The proportion of assigned sequences between the datasets at the class, order, family, genus, and species levels were evaluated (Figure 3a). Using the SilvaMod database showed that overall sequence assignments varied between the full-length 16S rRNA sequences and partial 16S rRNA sequences in addition to differences in the proportions of the assigned sequences at the different taxonomic ranks (Table 2). The v3v4 sequences had the highest proportion of assignment at the family and genus levels. At the species level, the full-length dataset had the highest proportion of assigned sequences (mean 8.9%).

Figure 3b gives insight into the distribution of genus diversity in the different data subsets. The number of genus in full-length, v3v4, and v5v6 datasets were 44, 60, and 63, respectively. The v3v4/v5v6 sequences included 47 genera, while the full-length/v3v4 sequences included 39 genera and the full-length/v5v6 sequences included 39 genera. Thirty-six genera were present in all datasets. The relative proportion of total sequences assigned to the 36 jointly shared genera was 49.04%, 59.62%, and 43.44% for the full, v3v4 and v5v6 datasets, respectively. In contrast, the proportion of sequences assigned to the 42 remaining ancillary genera ranged between 0.01% and 0.05% (Figure 3b). To further compare taxonomic profiling, the number of sequences assigned to each bacterial taxon was identified (Table 2). This revealed differences in efficiency of read assembly in the different datasets. The discriminant taxon at the order level was Clostridiales with 64.5% and 63.2% higher assignments in the v3v4 and v5v6 data subsets, respectively. Discrepancy is also seen between short sequence length data subsets. An apparent example, which is observed down to the genus level, is sequences assigned to *Carnobacterium* where 23.1% more v3v4 sequences are assigned compared to the full-length, while the v5v6 data-subset had only 0.3% assigned compared to the full-length dataset. The data further suggest differences in the proportion of assigned sequences between sequence length and the taxonomic rank genus and species. Both v3v4 and v5v6 region sequences have a higher proportion assigned to the genus *Flavobacterium* compared to the full-length dataset. This is opposite to species level where no v3v4 or v5v6 sequences are assigned while 8.3% of the full-length sequences are assigned to *Flavobacterium frigidarium*.



**FIGURE 2** A hierarchically clustered heatmap of the microbial profiles of Atlantic skin mucus and intestine based on SILVA database taxonomy assigned to (a) order and (b) genus levels. Dendrogram at the top of the heatmap shows the clustering of the sample types, skin mucus (SM1–SM6), and intestine (I1–I6). The dendrogram at the left side shows the distribution of bacteria. The color scale depicts the normalized relative abundance of each rank level. Principal component analysis (PCA) plot of scaled microbiota profiles representing both skin mucus (circles) and intestinal samples (triangles) of Atlantic salmon is depicted in the far right for both order and genus levels. Filled (black) symbols for fish fed krill meal diet and unfilled (white) symbols for fish fed fish meal based diet

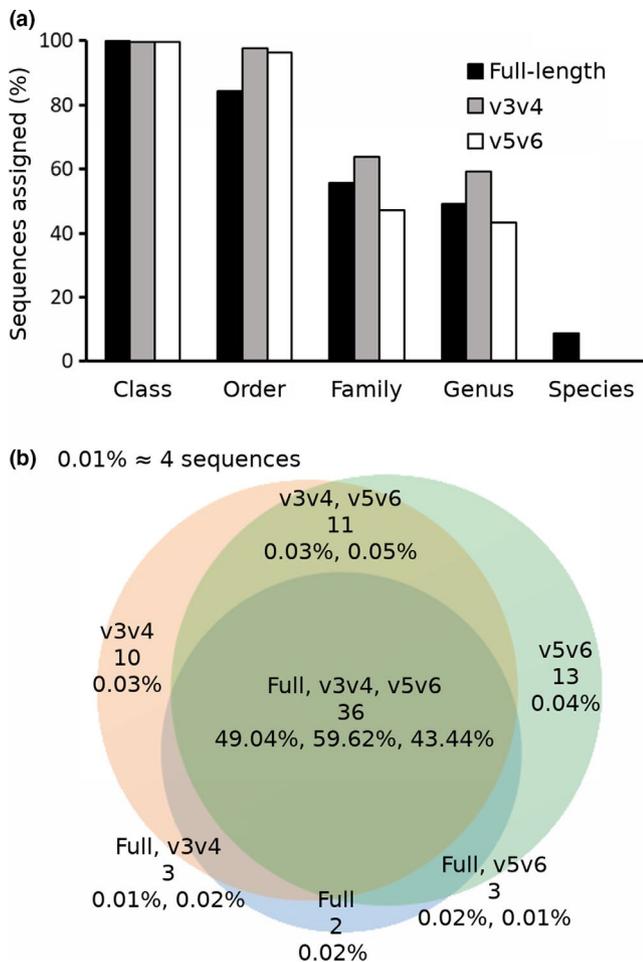
### 3.3 | Bacteria recovered by culture-dependent methods

The culture-based method aimed to provide representative isolates within anticipated genera to compare sequence information against the recovered CCS pool of sequences. Bacterial colonies were phenotypically categorized, and a total of 46 representative colonies were further identified and allocated to seven different genera based on the comparative 16S rRNA gene sequence alignment (Table A1). The relative distribution between colony morphologies on plates retrospectively identified by phylogeny is provided as an average of each sample type in Table A2, Appendix. Isolate information, sample origin, diet group, growth medium used and the most closely related type strains to each genus group is shown in Figure A1, Appendix.

Skin mucus samples resulted in *Flavobacterium*, *Psychrobacter*, and *Exigubacterium* isolates on R2A plates, and *Pseudomonas* and *Shewanella* on MacConkey plates. *Carnobacterium* dominated intestinal samples on both media plates. Representative colonies aligning to either genus *Flavobacterium*, *Carnobacterium*, or *Exigubacterium* were identical at the 16S rRNA gene level. Variants within the 16S rRNA gene level were found for isolates within each genus *Chryseobacterium*, *Psychrobacter*, *Pseudomonas*, and *Shewanella*.

### 3.4 | Assignment of CCS reads to the bacterial isolates

The distribution of the 46 identified bacteria isolates among the pool of sequences is shown in Figure 4, where the displayed OTUs



**FIGURE 3** Percentage of assigned sequences at the class, order, family, genus, and species levels from full-length (black bars), v3v4 (gray bars), and v5v6 (white bars) datasets (a). Venn diagram of genus level assignments illustrating the number of taxa shared in and between each dataset (b). Top line shows dataset(s), middle line shows the number of unique genus taxa shared among dataset(s), bottom line shows the relative abundance of sequences assigned to the involved taxa and the given dataset(s)

in tree branches are representative CCS within each allocated taxon. As expected, not all taxa detected by sequencing are cultivable under the conditions used. The proportion of cultivable bacteria was investigated by assigning 16S rRNA gene sequences from the obtained isolates to the full-length 16S rRNA CCS generated reads. Sequences (CCS reads) in cluster with 97% sequence identity to isolates accounted for 3.99% of the CCS reads, that is, cultivability of 3.99%. Comparing the relative abundance of OTUs assigned to the genus level shows domination of *Flavobacterium* and *Psychrobacter* in skin mucus and *Carnobacterium* in the intestinal samples. Two of the most abundant genera in the skin mucus, *Psychrobacter* and *Flavobacterium*, are represented by eight isolates, each (Figure 4). They are clearly distinctive to their respective branch, although the *Psychrobacter* appear to include subgroups that cannot be discriminated based on their 16S rRNA gene sequences (Figure A1). An exception within the tree is created by OTU 10 *Psychrobacter* where the CCS is placed between

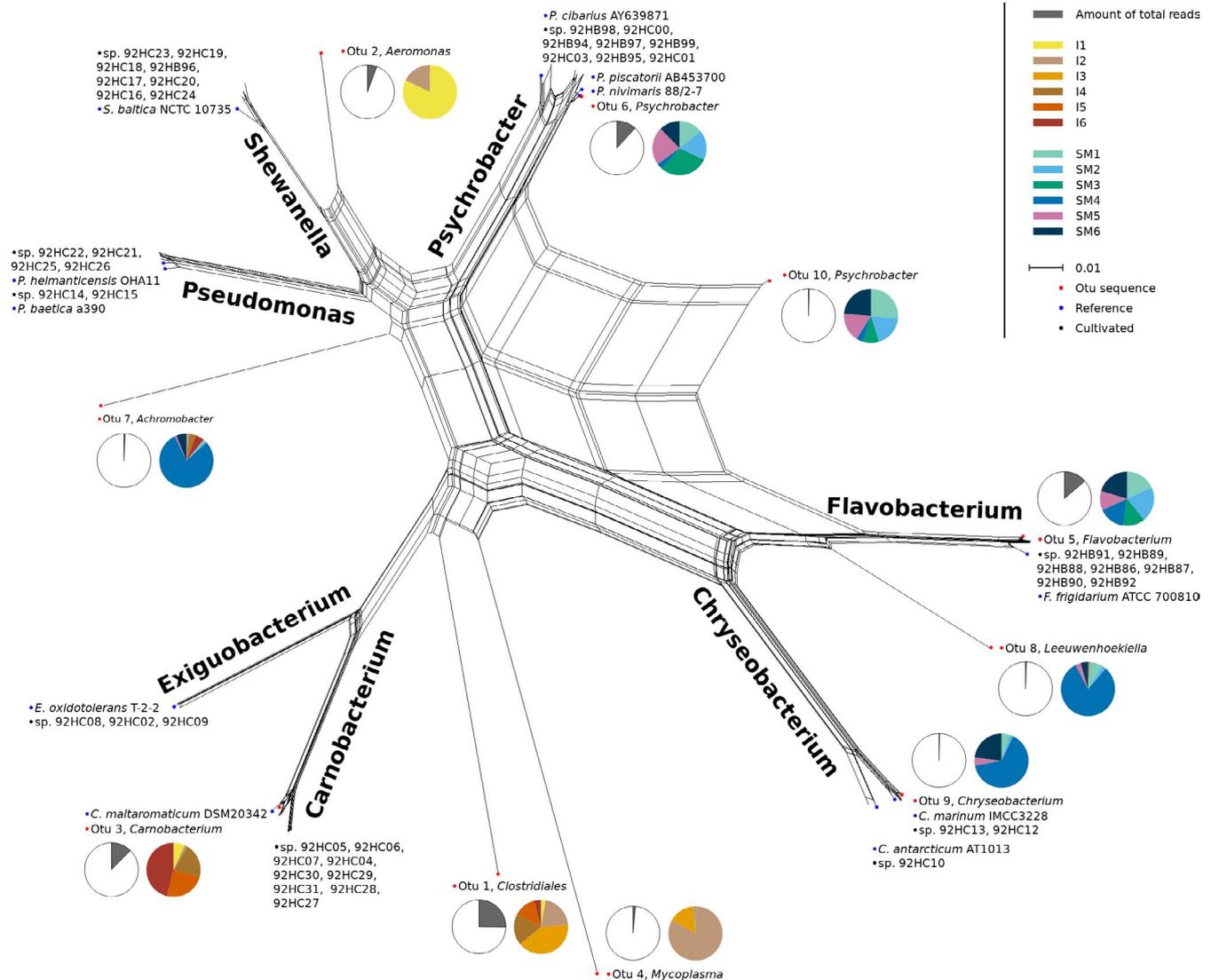
**TABLE 2** Assignments for pooled data of full-length 16S rRNA, v3v4 and v5v6 regions

Or.	Fa.	Ge.	Sp.	Full	v3v4	v5v6
Aeromonadales				3,242	+14	+5
Aeromonadaceae				2,801	+369	+302
<i>Aeromonas</i>				2,598	+420	+407
Alteromonadales				260	+13	+22
Pseudoalteromonadaceae				126	+103	+93
<i>Pseudoalteromonas</i>				94	+118	+117
Bacillales				34	+170	+7
Burkholderiales				591	+35	+28
Alcaligenaceae				387	+171	+142
Clostridiales				8,163	+5,264	+5,157
Ruminococcaceae				0	+11	+176
Enterobacteriales				0	0	+72
Flavobacteriales				9,391	+119	-3
Flavobacteriaceae				7,877	+1,219	+1,117
<i>Chryseobacterium</i>				154	+155	+141
<i>Chryseobacterium marinum</i>				84	-84	-84
<i>Flavobacterium</i>				6,500	+1,281	+1,058
<i>Flavobacterium frigidarium</i>				3,802	-3,802	-3,802
<i>Leeuwenhoekella</i>				256	+235	+187
<i>Bizionia</i>				19	+19	+24
Others				13	+31	+40
Lactobacillales				7,420	-83	-217
Carnobacteriaceae				6,306	+873	-6,283
<i>Carnobacterium</i>				5,630	+1,303	-5,615
Mycoplasmatales				955	+70	+61
Mycoplasmataceae				756	+217	+201
<i>Mycoplasma</i>				571	+251	+343
Pseudomonadales				6,953	+66	+74
Moraxellaceae				6,114	+619	+523
<i>Psychrobacter</i>				5,645	+756	+790
Rhodobacterales				44	+20	+7
Rhodobacteraceae				23	+32	+19
<i>Sphingobacteriales</i>				56	+16	+10
Saprospiraceae				23	+25	+26

Note: Sum of assignments where the sub-datasets of v-regions are displayed relative to the full dataset. Only assignments above 0.1% (>44 no. of sequences) of the total 43,910 sequences represented in each of the three datasets is shown.

Abbreviations: Fa., family; Ge., Genus; Or., order; Sp., Species.

the genera *Psychrobacter* and *Flavobacterium*. Reads belonging to the genus *Exiguobacterium* was not detected in the intestine and is represented with  $\leq 4$  CCS reads in the skin mucus samples. *Shewanella* CCS reads are represented by  $\leq 2$  sequences in two samples. No CCS reads belong to the genus *Pseudomonas*.



**FIGURE 4** Phylogenetic relationships as NeighborNet network of representative OTUs (red), cultivated bacteria sequences (black), and reference strains (blue). Pie charts detail the given OTU by indicating the proportion of reassigned full-length 16S reads (gray) and how these are distributed from the samples (colored). The sample values are shown scaled to the minimal sample SM2. Scale bar shows distance as number of nucleotide substitutions per site

## 4 | DISCUSSION

The skin mucus microbial community was dominated by Proteobacteria within the genus *Flavobacterium* and *Psychrobacter*. This corroborates previous findings reporting the dominance of a few Proteobacteria-affiliated phylotypes in Atlantic salmon skin mucus from both controlled experiments (Lokesh & Kiron, 2016; Minniti et al., 2017) and commercial production systems (Karlsen et al., 2017). The *Flavobacterium* genus has several important fish pathogens, primarily found in freshwater hatchery-reared fish (Starliper, 2011), that may adversely affect Atlantic salmon (Loch & Faisal, 2015). *Psychrobacter* spp. are also associated with marine fish species (Ramírez & Romero, 2017a, 2017b; Småge, Frisch, Brevik, Watanabe, & Nylund, 2016). Prominent intestinal bacteria were of the order Clostridiales, Aeromonadales, and Lactobacillales, which are repeatedly reported from the Atlantic

salmon intestine (Catalán, Villasante, Wacyk, Ramírez, & Romero, 2017; Gajardo, Jaramillo-Torres, et al., 2016; Llewellyn et al., 2015; Zarkasi et al., 2014). The number of fish in this trial makes it difficult to provide any conclusive insight into any dietary effect. Of note is the indication of more predominant *Carnobacterium* in the fish fed fish meal. This corroborates a previous study that also substituted fish meal with krill meal followed by culture-dependent techniques on gut microbiota (Ringø et al., 2006) where *Carnobacterium* was present in non-krillmeal-fed fish. A large proportion of the Clostridiales CCS 16S rRNA gene sequences could not be taxonomically assigned above the order rank, indicating the presence of so far uncharacterized bacteria.

Bacteria in the seawater column often inhabit a nutrient poor environment and a large number of the marine bacteria that occur in this environment are suggested to be nonculturable using standard culture-based techniques (Giovannoni & Stingl, 2005). Bacteria

associated with the Atlantic salmon integument are also considered difficult to cultivate. This includes pathogens such as *Tenacibaculum* (Olsen et al., 2011) and *Moritella* (Grove et al., 2008), and culture-based diagnostics of Atlantic salmon are considered unreliable (Grove et al., 2008). Early studies reported relative high cultivability of bacteria from salmonid intestines (Huber et al., 2004; Spanggaard et al., 2000), but also inconsistencies between culture-dependent and molecular-based methods (Hovda, Lunestad, Fontanillas, & Rosnes, 2007). Our approach did not aim to assess culturability, but was applied to recover Atlantic salmon isolates for future studies. Dominant bacteria in the intestine within the order Clostridiales and OTUs assigned to *Aeromonas* and *Mycoplasma* were not isolated, likely due to the growth media and aerobic condition used. Still, an overall ratio of culturability of 3.99% was demonstrated based on the presence/absence of all assigned CCS 16S rRNA gene sequences corresponding to cultivated isolates with 97% sequence identity. This dropped to 0.13% at 99% sequence identity. Both the dominating genera *Flavobacterium* and *Psychrobacter* are studied for their degrading properties (Lasa & Romalde, 2017; Loch & Faisal, 2015), and *Carnobacterium* is isolated from a range of environments (Leisner, Laursen, Prévost, Drider, & Dalgaard, 2007). It is possible that these genera are well-adapted to grow on the laboratory cultivation media used in this study.

Like many genera associated with the Atlantic salmon, the phylogenetic assignment is often based on the 16S rRNA gene. However, there is a lack of properly defined bacterial species within the aquatic environment that in general hamper our ability to understand and organize bacterial diversity to these environments. There are also technological limitations that may impact on the ecological data description of these analyses (Schmidt, Matias, & Mering, 2015). Related to Atlantic salmon, many of the dominant groups or recovered isolates cannot be discriminated at the species level by their 16S rRNA gene sequences (Grove et al., 2010; Småge et al., 2015). Furthermore, our data suggest that the salmon contains more than one species within observed genera such as *Psychrobacter*. Covering the full-length sequences of 16S rRNA genes is expected to be advantageous for inferring phylogenetic affiliations and provide a more precise microbial community profiling of Atlantic salmon. However, our partial gene sequences were assigned in a higher abundance compared to the full-length sequences. Only at the species level was full-length sequences assigned in a higher proportion (mean 8.9% compared to 0.01% and 0.005% for v3v4 and v5v6 data subsets, respectively). Effects on abundance profiling and discrepancy with an overestimation in bacterial diversity between full-length 16S rRNA gene and partial variable datasets are in similar accordance with other studies (Singer et al., 2016; Sun et al., 2013; Wagner et al., 2016; Whon et al., 2018). The taxonomic resolution in amplicon sequencing might be affected by several factors. In our study, where we generated different sequence lengths in silico one possibility is differences of the intravariability within each targeted hypervariable region (v3v4 and v5v6) in comparison with the full-length sequence (Kumar, Brooker, Dowd, & Camerlengo, 2011). Another influencing factor could be the reference sequences in the choice of database used

(Werner et al., 2012). Evaluating the species richness, we could not fully describe the community in all samples. The proportion of pooled sequences and the sample sizes obtained can be contributing factors to this lack of convergence where potentially important taxa have not been identified. Using the pooled Atlantic salmon community, we observed discrepancy in community structure and phylogenetic resolution across multiple taxonomic levels between full-length and partial sequences. Differences were revealed more clearly in some of the phylogenetic lineages. In our data, especially Clostridiales at the order level but also genera within Flavobacteriaceae and Carnobacteriaceae. This highlights that taxonomic affiliation using the bacterial 16S rRNA gene should be concluded with care. The 36 genera shared among the full-length and partial data subsets compose most of the assigned sequences (49.04%, 59.62%, and 43.44% of the full-length, and v3v4 and v5v6 data subsets, respectively) and greatly outnumber the proportion of assigned sequences in the 42 ancillary genera identified from one or two of the combined datasets. This high number of genera that is derived from a small part of the sequence sets might indicate that some genera represent false positives, caused by the shorter stretches of the 16S rRNA linked to a lack of resolution in the sequence regions. Collectively, these data suggest that sequence length and part of variable region used on the 16S rRNA gene will influence the outcome of the obtained microbial profile, in a way that is different between taxonomic rank and phylogenetic group. Although this study suggests confounding effects when using different variable regions of the 16S rRNA gene to characterize microbial profiles, it also has key limitations. The sample size should have been increased to better allow robust assessment of experimental effects. The microbiota in context to associations and interactions in animal trials are likely to be complex. Increasing the number of samples may better take in considerations concerning factors such as individual effects caused by husbandry or disease status (Moore & Stanley, 2016). Methodological errors or inadequacy to lyse and extract DNA from bacterial cells between tissue types have also a potential for introducing biases into the results. In addition, the taxonomic resolution in amplicon sequencing might be affected by several additional factors such as PCR conditions used to amplify the product (Lorenz, 2012), primer specificity (Beckers et al., 2016; Tremblay et al., 2015) and possible contamination in low microbial biomass samples (Eisenhofer et al., 2019). To further narrow the gaps and identify the best approach for microbial profiling of Atlantic salmon, future studies should compare primer sets targeting different sequence regions combined with sequencing technology platforms using several tissue types and DNA from a mock community containing a known number of species.

By comparing the distribution of representative OTUs to 16S rRNA gene sequences from cultivated isolates, we aimed to assess any anomaly or difference in the taxonomic classification. The sequences were clearly assigned down to the genus level, except OTU 10 which was placed between the genera *Psychrobacter* and *Flavobacterium*. An interesting observation is that *Exiguobacterium* and *Shewanella* are observed in relative high abundance by the cultivation method

compared to the extraction of DNA. *Pseudomonas* was at most represented by five sequences in the pooled v3v4 data subset of the generated 16S rRNA CCS reads. Methodological errors or inadequacy to lyse and extract DNA from bacterial cells, primer specificity, or PCR conditions could be possible explanations. However, because all cultivable taxa were detectable with the nonbarcoded version of the primers it is unlikely to account for an almost complete absence of a genus. Another explanation is that the sequencing depth may have been too low as indicated by the rarefaction analysis. Bacteria not detected by molecular methods may also be site or human-derived plate contaminants or cultivation procedures may have facilitated growth of these bacteria. The phenomenon of cultivable isolates not being detected in corresponding gene libraries are reported from a wide distribution of marine sample types such as seawater (Eilers, Pernthaler, Glöckner, & Amann, 2000), sponges (Esteves et al., 2016), and algae (KleinJan et al., 2017).

In this study, both skin mucus and intestinal samples were successfully utilized to generate full-length 16S rRNA gene sequences by the PacBio CCS technology. A high proportion of reads from skin surface samples were allocated down to the level genus. In contrast, intestinal samples dominated by reads assigned to the order Clostridiales were lost at higher resolutions. This highlights the need to further expand the microbial 16S rRNA gene catalogue from underrepresented marine taxa into reference databases. Our data also suggest that different variable regions and sequence length of the 16S rRNA gene will influence the microbial profile differently by taxonomic rank and phylogenetic group. To identify the best taxonomic profiling effectiveness between different variable regions and full-length 16S rRNA gene would need further validation. However, at present, one of the bottlenecks in comparing microbial profiles is due to the different 16S rRNA gene regions used in different studies. Using the full-length 16S rRNA gene sequence has the potential to become a tool for more precise microbial community profiling that better allows global comparisons of microbiome studies.

## ACKNOWLEDGMENTS

This study was supported by UiT The Arctic University of Norway and through the Research Council of Norway (RCN) by the Nofima basic research grant (194050/F40). The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

## CONFLICT OF INTERESTS

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.

## AUTHOR CONTRIBUTIONS

CK initiated the study and performed the laboratory work. TK conducted and performed the bioinformatic analyses. NW coordinated

the work. CK and TK drafted the manuscript and all authors provided critical feedback and helped shape the research, analysis, and manuscript.

## ETHICS STATEMENT

Fish in this study was based on post mortem sampling of material from fish harvested from a different industry research study for other purposes. Fish at the Research Station for Sustainable Aquaculture (Sunndalsøra, Norway) was reared by trained professionals in accordance with guidelines and regulations of the Norwegian Food Safety Authority.

## DATA AVAILABILITY STATEMENT

The datasets generated and analyzed in this study are available in the BioProject repository at <https://www.ncbi.nlm.nih.gov/bioproject/PRJEB28410>.

## ORCID

Terje Klemetsen  <https://orcid.org/0000-0002-2024-1798>

## REFERENCES

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Beckers, B., Op De Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., & Vangronsveld, J. (2016). Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial Microbiomes in Metabarcoding Studies. *Frontiers in Microbiology*, 7, 650. <https://doi.org/10.3389/fmicb.2016.00650>
- Beklioglu, M., Telli, M., & Gozen, A. G. (2006). Fish and mucus-dwelling bacteria interact to produce a kairomone that induces diel vertical migration in *Daphnia*. *Freshwater Biology*, 51(12), 2200–2206. <https://doi.org/10.1111/j.1365-2427.2006.01642.x>
- Bertani, G. (1951). Studies on lysogenesis I. : The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, 62(3), 293–300.
- Catalán, N., Villasante, A., Wacyk, J., Ramírez, C., & Romero, J. (2017). Fermented Soybean Meal Increases Lactic Acid Bacteria in Gut Microbiota of Atlantic Salmon (*Salmo salar*). *Probiotics and Antimicrobial Proteins*, 10(3), 566–576. <https://doi.org/10.1007/s12602-017-9366-7>
- Chen, Z., Hui, P. C., Hui, M., Yeoh, Y. K., Wong, P. Y., Chan, M. C. W., ... Chan, P. K. S. (2019). Impact of preservation method and 16S rRNA hypervariable region on gut microbiota profiling. *mSystems*, 4(1), e00271–e00318. <https://doi.org/10.1128/mSystems.00271-18>
- Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., ... Tiedje, J. M. (2014). Ribosomal database project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, 42(D1), D633–D642. <https://doi.org/10.1093/nar/gkt1244>
- Dehler, C. E., Secombes, C. J., & Martin, S. A. M. (2017). Environmental and physiological factors shape the gut microbiota of Atlantic salmon parr (*Salmo salar* L.). *Aquaculture (Amsterdam, Netherlands)*, 467, 149–157. <https://doi.org/10.1016/j.aquaculture.2016.07.017>
- Dhariwal, A., Xia, J., Chong, J., Agellon, L. B., Habib, S., & King, I. L. (2017). MicrobiomeAnalyst: A web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic*

- Acids Research*, 45(W1), W180–W188. <https://doi.org/10.1093/nar/gkx295>
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26(19), 2460–2461.
- Eilers, H., Pernthaler, J., Glöckner, F. O., & Amann, R. (2000). Culturability and *in situ* abundance of pelagic bacteria from the North sea. *Applied and Environmental Microbiology*, 66(7), 3044–3051. <https://doi.org/10.1128/AEM.66.7.3044-3051.2000>
- Eisenhofer, R., Minich, J. J., Marotz, C., Cooper, A., Knight, R., & Weyrich, L. S. (2019). Contamination in low microbial biomass microbiome studies: Issues and recommendations. *Trends in Microbiology*, 27(2), 105–117. <https://doi.org/10.1016/j.tim.2018.11.003>
- Esteves, A. I. S., Amer, N., Nguyen, M., & Thomas, T. (2016). Sample processing impacts the viability and cultivability of the sponge microbiome. *Frontiers in Microbiology*, 7, 499. <https://doi.org/10.3389/fmicb.2016.00499>
- Gajardo, K., Jaramillo-Torres, A., Kortner, T. M., Merrifield, D. L., Tinsley, J., Bakke, A. M., & Krogdahl, Å. (2016). Alternative protein sources in the diet modulate microbiota and functionality in the distal intestine of Atlantic salmon (*Salmo salar*). *Applied and Environmental Microbiology*, 83(5), e02615–e02616. <https://doi.org/10.1128/aem.02615-16>
- Gajardo, K., Rodiles, A., Kortner, T. M., Krogdahl, Å., Bakke, A. M., Merrifield, D. L., & Sørum, H. (2016). A high-resolution map of the gut microbiota in Atlantic salmon (*Salmo salar*): A basis for comparative gut microbial research. *Scientific Reports*, 6, 30893. <https://doi.org/10.1038/srep30893>
- Giovannoni, S. J., & Stingl, U. (2005). Molecular diversity and ecology of microbial plankton. *Nature*, 437, 343. <https://doi.org/10.1038/nature04158>
- Grove, S., Reitan, L. J., Lunder, T., & Colquhoun, D. (2008). Real-time PCR detection of *Moritella viscosa*, the likely causal agent of winter ulcer in Atlantic salmon *Salmo salar* and rainbow trout *Oncorhynchus mykiss*. *Diseases of Aquatic Organisms*, 82(2), 105–109. <https://doi.org/10.3354/dao01972>
- Grove, S., Wiik-Nielsen, C. R., Lunder, T., Tunsjø, H. S., Tandstad, N. M., Reitan, L. J., ... Colquhoun, D. J. (2010). Previously unrecognised division within *Moritella viscosa* isolated from fish farmed in the North Atlantic. *Diseases of Aquatic Organisms*, 93(1), 51–61. <https://doi.org/10.3354/dao02271>
- Hall, T. A. (1999). BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, 95–98.
- Hovda, M. B., Lunestad, B. T., Fontanillas, R., & Rosnes, J. T. (2007). Molecular characterisation of the intestinal microbiota of farmed Atlantic salmon (*Salmo salar* L.). *Aquaculture*, 272, 581–588. <https://doi.org/10.1016/j.aquaculture.2007.08.045>
- Huber, I., Spanggaard, B., Appel, K. F., Rossen, L., Nielsen, T., & Gram, L. (2004). Phylogenetic analysis and *in situ* identification of the intestinal microbial community of rainbow trout (*Oncorhynchus mykiss*, Walbaum). *Journal of Applied Microbiology*, 96(1), 117–132. <https://doi.org/10.1046/j.1365-2672.2003.02109.x>
- Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), 254–267. <https://doi.org/10.1093/molbev/msj030>
- Karlsen, C., Ellingsen, A. B., Wiik-Nielsen, C., Winther-Larsen, H. C., Colquhoun, D., & Sørum, H. (2014). Host specificity and clade dependent distribution of putative virulence genes in *Moritella viscosa*. *Microbial Pathogenesis*, 77, 53–65. <https://doi.org/10.1016/j.micpath.2014.09.014>
- Karlsen, C., Ottem, K. F., Brevik, Ø. J., Davey, M., Sørum, H., & Winther-Larsen, H. C. (2017). The environmental and host-associated bacterial microbiota of Arctic seawater-farmed Atlantic salmon with ulcerative disorders. *Journal of Fish Diseases*, 40(11), 1645–1663. <https://doi.org/10.1111/jfd.12632>
- KleinJan, H., Jeanthon, C., Boyen, C., & Dittami, S. M. (2017). Exploring the cultivable *Ectocarpus* microbiome. *Frontiers in Microbiology*, 8, 2456. <https://doi.org/10.3389/fmicb.2017.02456>
- Kumar, P. S., Brooker, M. R., Dowd, S. E., & Camerlengo, T. (2011). Target region selection is a critical determinant of community fingerprints generated by 16S pyrosequencing. *PLoS ONE*, 6(6), e20956. <https://doi.org/10.1371/journal.pone.0020956>
- Lanzén, A., Jørgensen, S. L., Huson, D. H., Gorfer, M., Grindhaug, S. H., Jonassen, I., ... Urich, T. (2012). CREST – Classification Resources for Environmental Sequence Tags. *PLoS ONE*, 7(11), e49334. <https://doi.org/10.1371/journal.pone.0049334>
- Larsen, A., Bullard, S., Womble, M., & Arias, C. (2015). Community structure of skin microbiome of Gulf killifish, *Fundulus grandis*, is driven by seasonality and not exposure to oiled sediments in a Louisiana salt marsh. *Microbial Ecology*, 70(2), 534–544. <https://doi.org/10.1007/s00248-015-0578-7>
- Lasa, A., & Romalde, J. L. (2017). Genome sequence of three *Psychrobacter* sp. strains with potential applications in bioremediation. *Genomics Data*, 12, 7–10. <https://doi.org/10.1016/j.gdata.2017.01.005>
- Leisner, J. J., Laursen, B. G., Prévost, H., Drider, D., & Dalggaard, P. (2007). *Carnobacterium*: Positive and negative effects in the environment and in foods. *Fems Microbiology Reviews*, 31(5), 592–613. <https://doi.org/10.1111/j.1574-6976.2007.00080.x>
- Llewellyn, M. S., McGinnity, P., Dionne, M., Letourneau, J., Thonier, F., Carvalho, G. R., ... Derome, N. (2015). The biogeography of the atlantic salmon (*Salmo salar*) gut microbiome. *ISME Journal*, 10(5), 1280–1284. <https://doi.org/10.1038/ismej.2015.189>
- Loch, T. P., & Faisal, M. (2015). Emerging flavobacterial infections in fish: A review. *Journal of Advanced Research*, 6(3), 283–300. <https://doi.org/10.1016/j.jare.2014.10.009>
- Lokesh, J., & Kiron, V. (2016). Transition from freshwater to seawater reshapes the skin-associated microbiota of Atlantic salmon. *Scientific Reports*, 6, 19707. <https://doi.org/10.1038/srep19707>
- Lorenz, T. C. (2012). Polymerase chain reaction: Basic protocol plus troubleshooting and optimization strategies. *Journal of Visualized Experiments*, 63, e3998–e3998. <https://doi.org/10.3791/3998>
- Metsalu, T., & Vilo, J. (2015). ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Research*, 43(W1), W566–W570. <https://doi.org/10.1093/nar/gkv468>
- Minniti, G., Hagen, L. H., Porcellato, D., Jørgensen, S. M., Pope, P. B., & Vaaje-Kolstad, G. (2017). The skin-mucus microbial community of farmed Atlantic salmon (*Salmo salar*). *Frontiers in Microbiology*, 8, 2043. <https://doi.org/10.3389/fmicb.2017.02043>
- Moore, R. J., & Stanley, D. (2016). Experimental design considerations in microbiota/inflammation studies. *Clinical & Translational Immunology*, 5(7), e92–e92. <https://doi.org/10.1038/cti.2016.41>
- Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T. L., Agarwala, R., & Schäffer, A. A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics*, 24(16), 1757–1764. <https://doi.org/10.1093/bioinformatics/btn322>
- Nayak, S. K. (2010). Role of gastrointestinal microbiota in fish. *Aquaculture Research*, 41(11), 1553–1573. <https://doi.org/10.1111/j.1365-2109.2010.02546.x>
- Olsen, A. B., Nilsen, H., Sandlund, N., Mikkelsen, H., Sørum, H., & Colquhoun, D. J. (2011). *Tenacibaculum* sp. associated with winter ulcers in sea-reared Atlantic salmon *Salmo salar*. *Diseases of Aquatic Organisms*, 94(3), 189–199.
- Pootakham, W., Mhuantong, W., Yoocha, T., Putchim, L., Sonthirod, C., Naktang, C., ... Tangphatsornruang, S. (2017). High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio SMRT sequencing system. *Scientific Reports*, 7(1), 2774. <https://doi.org/10.1038/s41598-017-03139-4>
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: A comprehensive online resource for

- quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, 35(21), 7188–7196. <https://doi.org/10.1093/nar/gkm864>
- Ramírez, C., & Romero, J. (2017a). Fine flounder (*Paralichthys adspersus*) microbiome showed important differences between wild and reared specimens. *Frontiers in Microbiology*, 8, 271. <https://doi.org/10.3389/fmicb.2017.00271>
- Ramírez, C., & Romero, J. (2017b). The Microbiome of *Seriola lalandi* of Wild and Aquaculture Origin Reveals Differences in Composition and Potential Function. *Frontiers in Microbiology*, 8, 1844. <https://doi.org/10.3389/fmicb.2017.01844>
- Ringø, E., Sperstad, S., Myklebust, R., Mayhew, T. M., Mjelde, A., Melle, W., & Olsen, R. E. (2006). The effect of dietary krill supplementation on epithelium-associated bacteria in the hindgut of Atlantic salmon (*Salmo salar* L.): A microbial and electron microscopical study. *Aquaculture Research*, 37(16), 1644–1653. <https://doi.org/10.1111/j.1365-2109.2006.01611.x>
- Schmidt, T. S. B., Matias, R. J. F., & Mering, C. (2015). Limits to robustness and reproducibility in the demarcation of operational taxonomic units. *Environmental Microbiology*, 17(5), 1689–1706. <https://doi.org/10.1111/1462-2920.12610>
- Schmidt, V., Amaral-Zettler, L., Davidson, J., Summerfelt, S., & Good, C. (2016). Influence of fishmeal-free diets on microbial communities in Atlantic salmon (*Salmo salar*) recirculation aquaculture systems. *Applied and Environmental Microbiology*, 82(15), 4470–4481. <https://doi.org/10.1128/aem.00902-16>
- Schmidt, V. T., Smith, K. F., Melvin, D. W., & Amaral-Zettler, L. A. (2015). Community assembly of a euryhaline fish microbiome during salinity acclimation. *Molecular Ecology*, 24(10), 2537–2550. <https://doi.org/10.1111/mec.13177>
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863–864. <https://doi.org/10.1093/bioinformatics/btr026>
- Sepúlveda, D., Bohle, H., Labra, Á., Grothusen, H., & Marshall, S. H. (2013). Design and evaluation of a unique RT-qPCR assay for diagnostic quality control assessment that is applicable to pathogen detection in three species of salmonid fish. *BMC Veterinary Research*, 9, 183–183. <https://doi.org/10.1186/1746-6148-9-183>
- Sievers, F., & Higgins, D. G. (2014). Clustal omega, accurate alignment of very large numbers of sequences. In D. J. Russell (Ed.), *Multiple Sequence Alignment Methods* (pp. 105–116). Totowa, NJ: Humana Press.
- Singer, E., Bushnell, B., Coleman-Derr, D., Bowman, B., Bowers, R. M., Levy, A., ... Woyke, T. (2016). High-resolution phylogenetic microbial community profiling. *The ISME Journal*, 10, 2020. <https://doi.org/10.1038/ismej.2015.249>
- Småge, S. B., Brevik, Ø. J., Duesund, H., Ottem, K. F., Watanabe, K., & Nylund, A. (2015). *Tenacibaculum finnmarkense* sp. nov., a fish pathogenic bacterium of the family Flavobacteriaceae isolated from Atlantic salmon. *Antonie Van Leeuwenhoek*, 109(2), 273–285. <https://doi.org/10.1007/s10482-015-0630-0>
- Småge, S. B., Frisch, K., Brevik, Ø. J., Watanabe, K., & Nylund, A. (2016). First isolation, identification and characterisation of *Tenacibaculum maritimum* in Norway, isolated from diseased farmed sea lice cleaner fish *Cyclopterus lumpus* L. *Aquaculture*, 464, 178–184. <https://doi.org/10.1016/j.aquaculture.2016.06.030>
- Spanggaard, B., Huber, I., Nielsen, J., Nielsen, T., Appel, K. F., & Gram, L. (2000). The microflora of rainbow trout intestine: A comparison of traditional and molecular identification. *Aquaculture*, 182(1), 1–15. [https://doi.org/10.1016/S0044-8486\(99\)00250-1](https://doi.org/10.1016/S0044-8486(99)00250-1)
- Starliper, C. E. (2011). Bacterial coldwater disease of fishes caused by *Flavobacterium psychrophilum*. *Journal of Advanced Research*, 2(2), 97–108. <https://doi.org/10.1016/j.jare.2010.04.001>
- Sullam, K. E., Essinger, S. D., Lozupone, C. A., O'Connor, M. P., Rosen, G. L., Knight, R., & Kilham, S. S. (2012). Environmental and ecological factors that shape the gut bacterial communities of fish: A meta-analysis. *Molecular Ecology*, 21(13), 3363–3378. <https://doi.org/10.1111/j.1365-294X.2012.05552.x>
- Sun, D.-L., Jiang, X., Wu, Q. L., & Zhou, N.-Y. (2013). Intragenomic heterogeneity of 16S rRNA genes causes overestimation of prokaryotic diversity. *Applied and Environmental Microbiology*, 79(19), 5962–5969. <https://doi.org/10.1128/aem.01282-13>
- Sveen, L. R., Grammes, F. T., Ytteborg, E., Takle, H., & Jørgensen, S. M. (2017). Genome-wide analysis of Atlantic salmon (*Salmo salar*) mucin genes and their role as biomarkers. *PLoS ONE*, 12(12), e0189103. <https://doi.org/10.1371/journal.pone.0189103>
- Tamura, K., Stecher, G., Peterson, D., Filipowski, A., & Kumar, S. (2013). MEGA6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, 30(12), 2725–2729. <https://doi.org/10.1093/molbev/mst197>
- Travers, K. J., Chin, C.-S., Rank, D. R., Eid, J. S., & Turner, S. W. (2010). A flexible and efficient template format for circular consensus sequencing and SNP detection. *Nucleic Acids Research*, 38(15), e159–e159. <https://doi.org/10.1093/nar/gkq543>
- Tremblay, J., Singh, K., Fern, A., Kirton, E. S., He, S., Woyke, T., ... Tringe, S. G. (2015). Primer and platform effects on 16S rRNA tag sequencing. *Frontiers in Microbiology*, 6, 771–771. <https://doi.org/10.3389/fmicb.2015.00771>
- Wagner, J., Coupland, P., Browne, H. P., Lawley, T. D., Francis, S. C., & Parkhill, J. (2016). Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC Microbiology*, 16(1), 274. <https://doi.org/10.1186/s12866-016-0891-4>
- Walker, A. W., Martin, J. C., Scott, P., Parkhill, J., Flint, H. J., & Scott, K. P. (2015). 16S rRNA gene-based profiling of the human infant gut microbiota is strongly influenced by sample processing and PCR primer choice. *Microbiome*, 3(1), 26. <https://doi.org/10.1186/s40168-015-0087-4>
- Werner, J. J., Koren, O., Hugenholtz, P., DeSantis, T. Z., Walters, W. A., Caporaso, J. G., ... Ley, R. E. (2012). Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *The ISME Journal*, 6(1), 94–103. <https://doi.org/10.1038/ismej.2011.82>
- Whon, T. W., Chung, W.-H., Lim, M. Y., Song, E.-J., Kim, P. S., Hyun, D.-W., ... Nam, Y.-D. (2018). The effects of sequencing platforms on phylogenetic resolution in 16 S rRNA gene profiling of human feces. *Scientific Data*, 5, 180068–180068. <https://doi.org/10.1038/sdata.2018.68> <https://doi.org/10.1038/sdata.2018.68>
- Zarkasi, K. Z., Abell, G. C. J., Taylor, R. S., Neuman, C., Hatje, E., Tamplin, M. L., ... Bowman J. P. (2014). Pyrosequencing-based characterization of gastrointestinal bacteria of Atlantic salmon (*Salmo salar* L.) within a commercial mariculture system. *Journal of Applied Microbiology*, 117(1), 18–27.

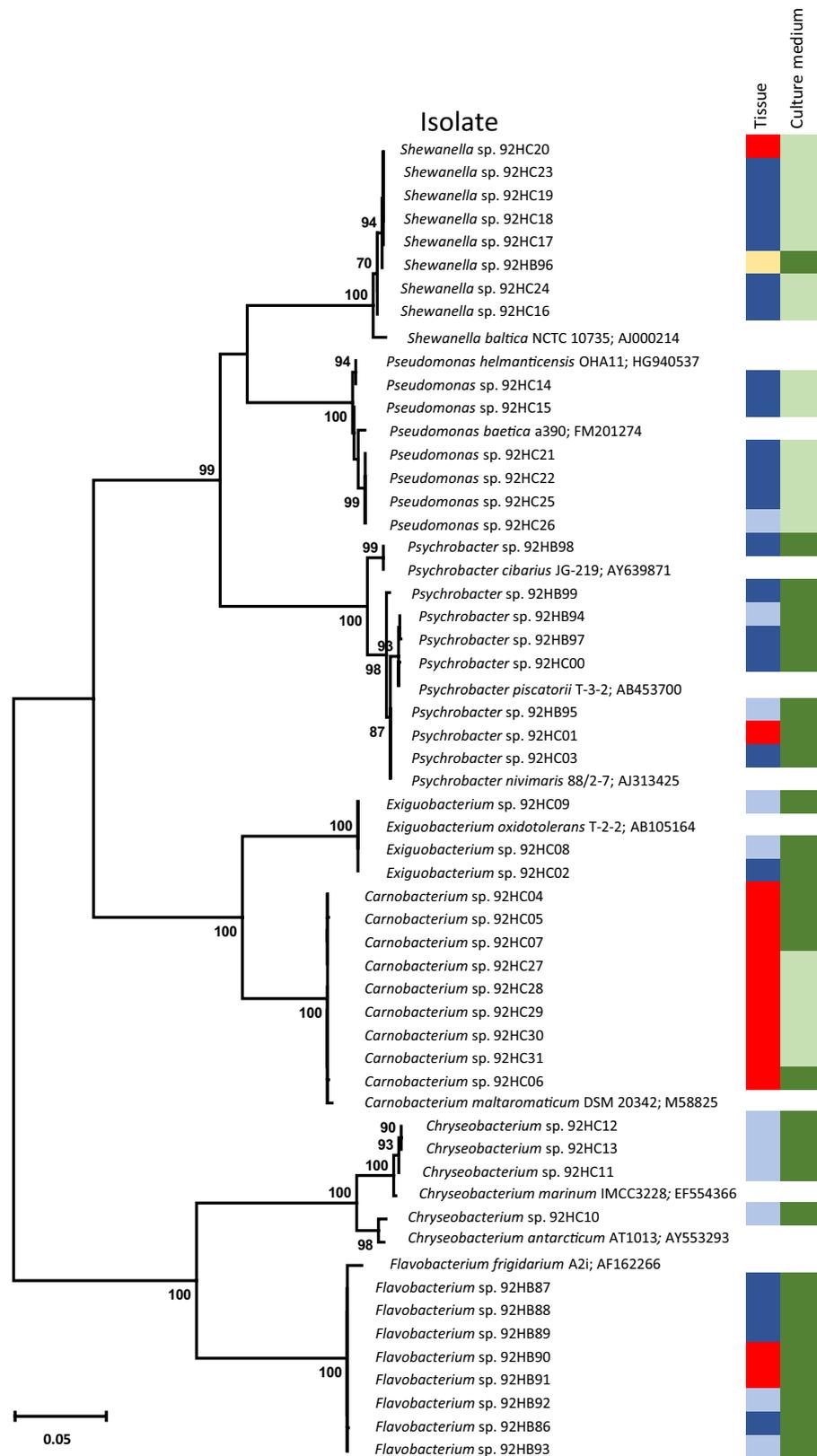
**How to cite this article:** Klemetsen T, Willassen NP, Karlsen CR. Full-length 16S rRNA gene classification of Atlantic salmon bacteria and effects of using different 16S variable regions on community structure analysis. *MicrobiologyOpen*. 2019;e898. <https://doi.org/10.1002/mbo3.898>

## APPENDIX A

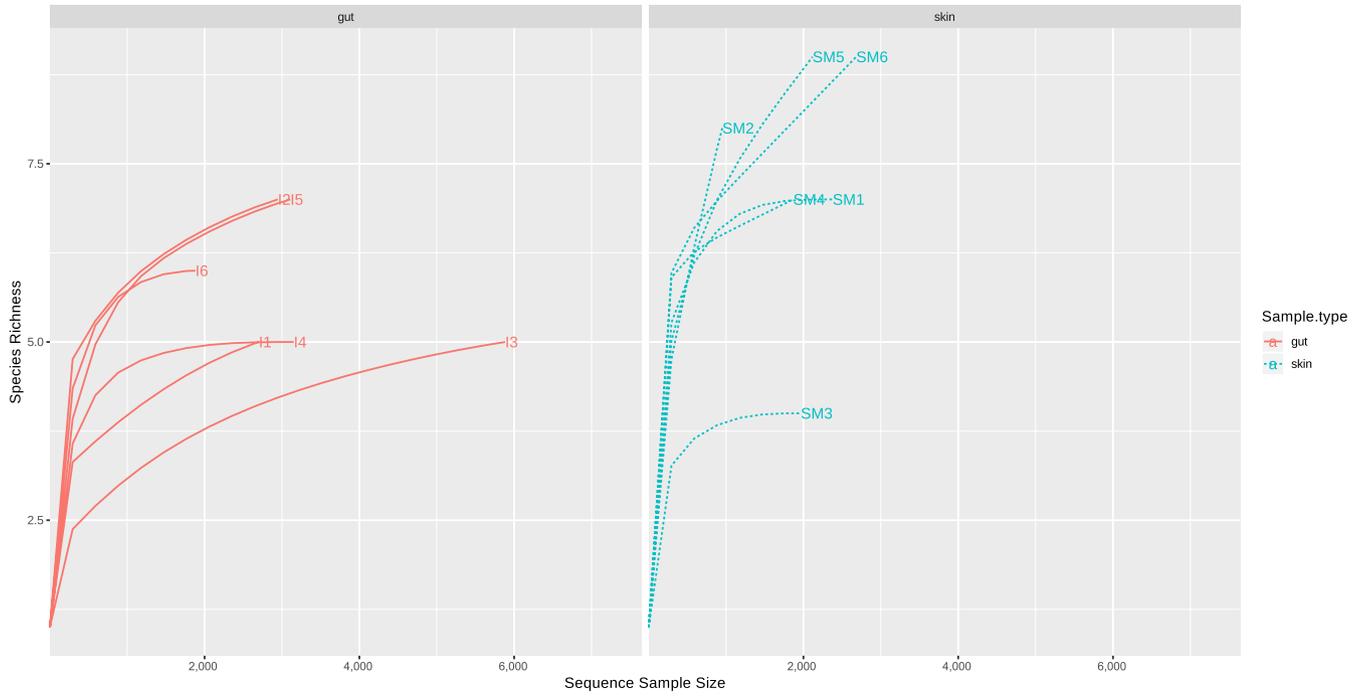
**TABLE A1** Novel bacterial strains isolated in this study presented with reconstructed phylogenetic genus groups, 16S accession number, isolation source, and culture medium

Isolates	GenBank acc. no.	Tissue	Culture medium
<i>Flavobacterium</i> sp. 92HB86	MG263463	Skin	R2A
<i>Flavobacterium</i> sp. 92HB87	MG263464	Skin	R2A
<i>Flavobacterium</i> sp. 92HB88	MG263465	Skin	R2A
<i>Flavobacterium</i> sp. 92HB89	MG263466	Skin	R2A
<i>Flavobacterium</i> sp. 92HB90	MG263467	Intestine	R2A
<i>Flavobacterium</i> sp. 92HB91	MG263468	Intestine	R2A
<i>Flavobacterium</i> sp. 92HB92	MG263469	Gill	R2A
<i>Flavobacterium</i> sp. 92HB93	MG263470	Gill	R2A
<i>Psychrobacter</i> sp. 92HB94	MG263471	Gill	R2A
<i>Psychrobacter</i> sp. 92HB95	MG263472	Gill	R2A
<i>Psychrobacter</i> sp. 92HB97	MG263473	Skin	R2A
<i>Psychrobacter</i> sp. 92HB98	MG263474	Skin	R2A
<i>Psychrobacter</i> sp. 92HB99	MG263475	Skin	R2A
<i>Psychrobacter</i> sp. 92HC00	MG263476	Skin	R2A
<i>Psychrobacter</i> sp. 92HC01	MG263477	Intestine	R2A
<i>Psychrobacter</i> sp. 92HC03	MG263478	Skin	R2A
<i>Exiguobacterium</i> sp. 92HC02	MG263480	Skin	R2A
<i>Exiguobacterium</i> sp. 92HC08	MG263481	Gill	R2A
<i>Exiguobacterium</i> sp. 92HC09	MG263482	Gill	R2A
<i>Chryseobacterium</i> sp. 92HC10	MG263487	Gill	R2A
<i>Chryseobacterium</i> sp. 92HC11	MG263488	Gill	R2A
<i>Chryseobacterium</i> sp. 92HC12	MG263489	Gill	R2A
<i>Chryseobacterium</i> sp. 92HC13	MG263490	Gill	R2A
<i>Pseudomonas</i> sp. 92HC14	MG263491	Skin	MacConkey
<i>Pseudomonas</i> sp. 92HC15	MG263492	Skin	MacConkey
<i>Pseudomonas</i> sp. 92HC21	MG263493	Skin	MacConkey
<i>Pseudomonas</i> sp. 92HC22	MG263494	Skin	MacConkey
<i>Pseudomonas</i> sp. 92HC25	MG263495	Skin	MacConkey
<i>Pseudomonas</i> sp. 92HC26	MG263496	Gill	MacConkey
<i>Shewanella</i> sp. 92HB96	MG263479	Water	R2A
<i>Shewanella</i> sp. 92HC16	MG263497	Skin	MacConkey
<i>Shewanella</i> sp. 92HC24	MG263498	Skin	MacConkey
<i>Shewanella</i> sp. 92HC17	MG263499	Skin	MacConkey
<i>Shewanella</i> sp. 92HC18	MG263500	Skin	MacConkey
<i>Shewanella</i> sp. 92HC19	MG263501	Skin	MacConkey
<i>Shewanella</i> sp. 92HC20	MG263502	Intestine	MacConkey
<i>Shewanella</i> sp. 92HC23	MG263503	Skin	MacConkey
<i>Carnobacterium</i> sp. 92HC27	MG263504	Intestine	MacConkey
<i>Carnobacterium</i> sp. 92HC28	MG263505	Intestine	MacConkey
<i>Carnobacterium</i> sp. 92HC29	MG263506	Intestine	MacConkey
<i>Carnobacterium</i> sp. 92HC30	MG263507	Intestine	MacConkey
<i>Carnobacterium</i> sp. 92HC31	MG263508	Intestine	MacConkey
<i>Carnobacterium</i> sp. 92HC04	MG263483	Intestine	R2A
<i>Carnobacterium</i> sp. 92HC05	MG263484	Intestine	R2A
<i>Carnobacterium</i> sp. 92HC06	MG263485	Intestine	R2A
<i>Carnobacterium</i> sp. 92HC07	MG263486	Intestine	R2A





**FIGURE A1** Unrooted neighbor-joining tree constructed from maximum likelihood distances between 16S rRNA gene sequences obtained from isolates of this study with the most closely aligned type strain obtained from the RDP. Type strains are displayed with strain ID followed by the GenBank 16S rRNA gene accession number. GenBank accession number for isolates obtained in this study is provided in Table A1. Bootstrap values  $\geq 70\%$  based on 1,000 replicates are indicated, and the scale bar represents the number of substitutions per site. The right color-coding columns of the phylogenetic clades show the tissue type each strain was isolated from and culture medium used. Skin is dark blue, gill is light blue, intestine is red, and water is yellow. Culture medium R2A is dark green, and MacConkey is light green



**FIGURE A2** Rarefaction plots based on OTU analysis for intestine and skin samples. Underlying data represent the number of sequences in each dataset mapped to OTUs clustered with a 97% identity threshold. Plot was generated using total sum scaling of filtered data showing 20 steps