Author

Yin Xun Tan

Date

30.08.2020

Research group

GELIFES — Groningen Institute for Evolutionary Life Sciences

University of Groningen

IBAHCM — Institute of Biodiversity Animal Health & Comparative Medicine

University of Glasgow

Supervisor

Prof. Dr. Bregje Wertheim

University of Groningen

Dr. Martin Llewellyn

University of Glasgow

Title

Something Fishy Going On: Microbiome Development in SalmoSim's Salmon Gut Simulator

Abstract

Fishmeal is the primary protein source of commercial farm feeds for Atlantic salmon. Due to the high price and low availability of fishmeal, there has been an increasing use of alternative feeds with plant-based ingredients, jeopardising the gut health and growth rate of salmon. To further understand the role of salmon gut microbiome in digestion and gut health, an *in vitro* salmon gut simulator was developed and set up by the start-up company SalmoSim, to carry out *in vitro* analyses of gut microbial diversity and composition. Here, we utilise the microbiome data collected and pre-processed by the SalmoSim team to address four study objectives, 1) estimate the time required for microbial communities to stabilise in SalmoSim, 2) compare the microbial communities of *in vitro* SalmoSim to *in vivo* real salmon under certain feed types, 3) compare the microbial communities of *in vitro* SalmoSim under different feed types, 4) compare taxonomic assignment of microbial communities using a PacBio database as opposed to a public database. From the DNA extraction of our study samples, the V1 region and full-length V1-V9 regions of the 16S rRNa gene was amplified and sequenced. For objective 1, alpha diversity over 40 days found overall decreasing diversity and microbiome oscillations before stabilisation in the SalmoSim *in vitro* system. The stomach required more time before stabilisation in SalmoSim, followed by other compartments like the pyloric caeca then the midgut. With the stomach in real salmon, the longer time required may be due to the lower pH conditions acting as an environmental filter for the assembly of microbial communities. For objective 2, the sample type of either SalmoSim or real salmon may have been a factor leading to their difference in microbiome composition. Compared to real salmon guts, the slight difference in SalmoSim environmental conditions may have had an impact on its microbiome composition and core microbes identified in its samples. For objective 3, the feed types of either fishmeal or fishmeal-free showed little effect on the resulting microbiome composition in SalmoSim samples. Instead, the difference in microbiome composition might have been explained better by stochastic factors, whereby the colonising microbes between samples determined the introduction of subsequent microbes. For objective 4, full-length PacBio sequences of the study samples may have enhanced the precision of taxonomic assignment of microbial OTUs from Illumina short read length sequences. As the SalmoSim system is the first *in vitro* fish gut simulator to explore the microbiome aspect, to our knowledge, our study provides the aquaculture industry an innovative system to carry out successful feed trials, and identify feeds to farm salmon commercially and more sustainably than *in vivo* feed trials with sea cages.

Table of contents

Introduction

Fish consumption has been an energy source for humans, bringing high-quality animal proteins and essential micronutrients to the human diet (FAO & UNICEF, 2018). Due to these advantages, the yearly increase in fish consumption at 3.2 percent has exceeded human population growth at 1.6 percent since 1961 (FAO & UNICEF, 2018). To feed the growing human population, the aquaculture industry has been farming Atlantic salmon (*Salmo salar*) as seafood for its high nutrient density and production yield (Torrissen et al., 2011), whilst such farming purportedly reduces environmental impact on agricultural lands too (Olsen, 2011). As a result, the Atlantic salmon comprises over 90 percent of the farmed salmon market (FAO & UNICEF, 2018).

As Atlantic salmon are carnivorous, fishmeal is the primary protein source of commercial farm feeds. Large amounts of wild fishmeal are used to feed farmed salmon (Tacon & Metian 2008), which has led to the decrease of resources for fishmeal production and tremendous rise in the price of fishmeal (Kiron et al., 2012). Thus, the reliance on fishmeal has been unsustainable and cost-ineffective for the aquaculture industry. Overfishing for fishmeal production also carries environmental consequences as it depletes marine ecosystems (Coll et al., 2008). The continuous loss of wild fish species jeopardises the ocean's opportunity for recovery, water quality and food provision (Ellingsen et al., 2009; Pelletier et al., 2009).

Given the unsustainability of wild fishmeal, there has been increasing use of plant-based protein and fat sources more recently by manufacturers of salmon farm feeds (Torrissen et al., 2011; Kiron et al., 2012). Soybean is a commonly used protein in plant-based feeds because of its low cost and wide availability (Green et al., 2013). However, marine carnivores like the Atlantic salmon did not evolve a digestive system to breakdown carbohydrates in plant-based meals (Naylor et al., 2000). Introduction of plant-based alternatives causes intestinal inflammation with negative impact on the growth rate (Beheshti Foroutani et al., 2018) and altered gut health of salmon (Gajardo et al., 2017). To elucidate the role of salmon gut microbiome in digestion and gut health, gut microbial diversity and composition analyses are often carried out with salmon feed trials, providing salmon with different amounts of animal-based ingredients (Green et al., 2013; Gajardo et al., 2017).

Despite the effectiveness of salmon feed trials in studying gut microbial digestion, the usual methods employed in these trials have some drawbacks. Feed and research trials with a sea cage can cost up to £150,000 or 168,000€. Besides the staggering amount of money required, the opportunity for these trials are not widely available as there are only around ten testing sites across Europe for thousands of salmon farms. The amount of resources required for single trials in testing sites is also detrimental to the environment. From a research perspective, longitudinal studies cannot be carried out with usual *in vivo* feed trials to observe the assembly of the gut microbiome with varying treatments over time. Furthermore, when *in vivo* feed trials are carried out, the sampling of salmon gut contents involves the death of the salmon. Feed trials involving the death and manipulation of live salmon may pose ethical issues. As the sentience of fish and their ability to perceive pain are being reviewed and debated, further discussion may justify a similar level of protection for fish as other vertebrates in scientific research (Brown, 2014). Hence, a viable alternative to *in vivo* feed trials could be *in vitro* feed trials with gut model systems.

Gut model systems are highly reproducible and could control for host factors interacting with these systems. Recently, a system was built to simulate the gut of another teleost species (Drieschner et al., 2019). Drieschner et al. (2019)'s system is known as the fish-gut-on-chip model which is a reconstruction of the rainbow trout's intestinal barrier consisting of two intestinal cell lines. To our knowledge, this fish-gut-on-chip model does not incorporate or place emphasis on the gut microbiome of the fish. To explore the microbiome aspect of such model, our study utilises data collected from an ongoing project by SalmoSim (https://www.salmosim.co.uk/) which is a start-up company founded in the University of Glasgow. SalmoSim has recently set up an artificial, comprehensive *in vitro* simulator of the salmon gastrointestinal tract (Fig. 1). More specifically, the SalmoSim system simulates the stomach, pyloric caeca, and midgut regions of the Atlantic salmon. Experimental and sequence data were collected and pre-processed by the SalmoSim team for the analyses carried out in this study.

Before being utilised to carry out applied research for salmon aquaculture, the suitability of *in vitro* gut simulators needs to be tested. As the SalmoSim system is operated over time, the stability of the system is crucial to generate reliable experimental results. During the

activation of the system's bioreactors, a stable microbial community should be most representative of the *in vivo* salmon scenario. This stability ensures that changes in the microbial community are related to the treatments, instead of possible adaptations to the environment of an *in vitro* system (Possemiers et al., 2004; Van de Wiele et al., 2015).

Moreover, to have an *in vitro* gut simulator most representative of the *in vivo* salmon scenario, the microbial communities of SalmoSim need to be somewhat specific to that of live salmon (Van Den Abbeele et al., 2010). The *in vitro* gut simulator needs to undergo a validation process by comparing it to *in vivo* experiments. Validation allows us to assess the degree of similarity between the outcomes of *in vitro* experiments of the simulator and *in vivo* scenario of live salmon (Molly et al., 1994; Kopf-Bolanz et al., 2012; Ménard et al., 2014).

Here, we carry out diversity analyses to monitor the development of the *in vitro* salmon gut simulator (SalmoSim), and compare the gut microbial composition of *in vitro* SalmoSim to the *in vivo* scenario of actual salmon guts with different feed types. Before diversity analyses, microbial taxa are identified as individual operational taxonomic units (OTUs) from sample reads, mostly using pairwise sequence similarity searches against a public 16S database as reference (de Vargas et al., 2015; Mahé et al., 2017). With Illumina HiSeq, the V1 region of the 16S rRNa gene from gut samples provide short amplicons. When taxonomy is assigned with public taxonomy references, short amplicons constrain the taxonomic resolution to mostly the family or genus level (Schloss, 2010). Such low-resolution assignment reduces the precision to identify closely related microbial strains for diversity analyses.

To increase the resolution of taxonomic assignment, full-length Pacific Biosciences (PacBio) sequencing of pooled study samples could be utilised as a reference taxonomy database. Circular consensus sequencing (CCS) reads from PacBio involve the sequencing of 16S rRNA genes numerous times to provide for consensus-sequence error correction (Wagner, 2016). With PacBio CCS reads, more informative sites are sequenced (Mosher et al., 2013) which may be missed by Illumina short read length sequencing. Hence, when paired with Illumina HiSeq amplicons, a taxonomy annotated PacBio database could yield a higher resolution assignment of taxa than public reference databases. We endeavour to assess the

potential of PacBio CCS reads to enhance species-level assignment of sample OTUs for diversity analyses.

Overall, the specific objectives of this study are 1) to estimate the time required for microbial communities to stabilise in SalmoSim, 2) to compare the microbial communities of *in vitro* SalmoSim to *in vivo* real salmon under certain feed types, 3) to compare the microbial communities of *in vitro* SalmoSim under different feed types, 4) to compare taxonomic assignment of microbial communities using a PacBio database as opposed to a public database.

Materials and methods

Prior salmon sample collection

Before our study experiments, farmed Atlantic salmon were initially prepared and fed by Raminta Kazlauskaite (SalmoSim) in collaboration with MOWI Ltd at their research site in Averøy, Norway. Salmon individuals were fed with fishmeal until they weighed approximately 750 grams. The salmon were placed in 5x5 metre marine pens with 150 salmon randomly selected for each pen. Four pens were then randomly selected for each feed type. The two feed types in this study were fishmeal and fishmeal-free (Table 1). This initial feeding step was carried out over five months, then two salmon were randomly collected at the end of the feeding step from random pens for each feed type. The gut compartments collected from each salmon individual were the stomach, pyloric caecum, and midgut. The treatment and collection of samples were carried out as previously done in Heys et al. (2020).

Table 1. Composition of different feeds provided to farmed Atlantic salmon.

Ingredient (% of the feed)	**Fishmeal**	**Fishmeal-free**
Fish meal	17.50	0.00
Soya protein concentrate	12.00	27.80
Corn gluten	7.00	7.35
Wheat gluten	10.00	14.34
Sunflower expeller	3.41	0.00
Wheat	4.81	11.22
Beans dehulled	10.00	0.00
Fish oil	15.68	16.99
Rapeseed oil	11.78	11.79
Linseed oil	3.05	3.20
Mannooligosaccharide	0.40	0.40
Astaxanthin	0.04	0.04
Crystalline amino acids	1.35	1.99
Mineral premixes	1.83	2.66
Vitamin premixes	0.60	0.73

SalmoSim system set-up

Similar to the prior salmon sample collection, the SalmoSim system was previously set up by Martin Llewellyn and Raminta Kazlauskaite (SalmoSim). In setting up the SalmoSim equipment, three 500 ml Applikon Mini Bioreactors were filled with four 1cm^3 cubes of aquarium sponge filters to provide a surface for biofilm formation. The equipment was assembled by connecting tubes and probes to adjust for the redox reactions, temperature, and dissolved oxygen. All equipment was autoclaved and reactor vessels to the Applikon electronic control module were attached for feed connection and acid and base bottles. To maintain anaerobic conditions, nitrogen gas was introduced periodically through each vessel. The reactors were then filled with 400 ml of feed media. Once the set-up was completed, the pH, temperature, and oxygen concentration were stabilised according to measurements obtained from live adult salmon. Figure 1 shows an image of the completed SalmoSim setup, with its schematic representation visualised in Figure 2.

Fig. 1. Image of the completed SalmoSim set-up. Image was obtained from https://www.llewellynlab.com/salmosim.

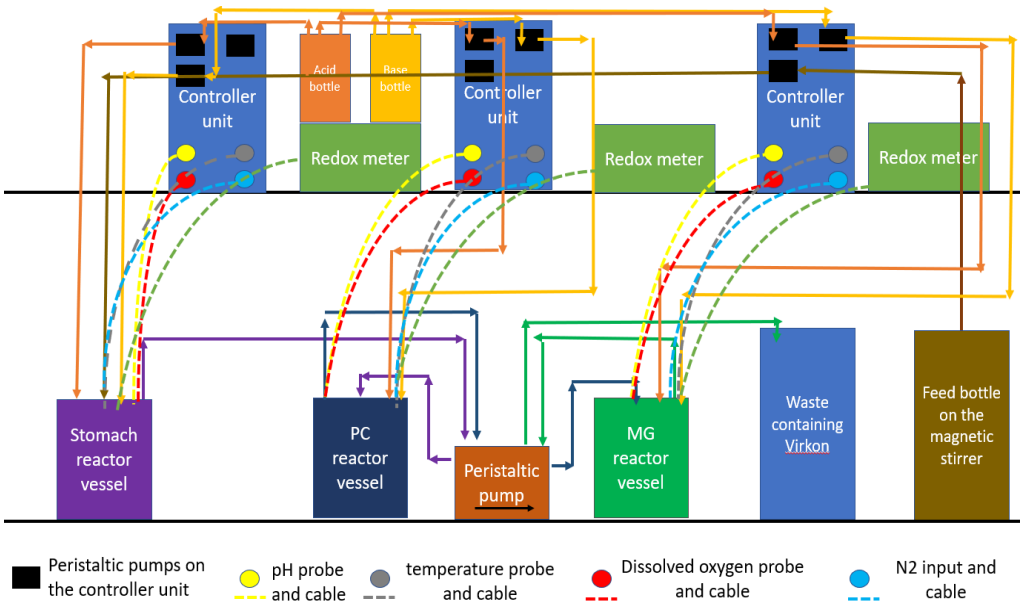


Fig. 2. Schematic representation of the SalmoSim system, illustrated by Raminta Kazlauskaite (SalmoSim).

For the SalmoSim system, the microbial inoculums were obtained from the three gut compartments (stomach, pyloric caecum, and midgut) of each individual salmon. The inoculums were stored in falcon tubes containing 30% glycerol solution at -80 °C, then dissolved in 1 ml of autoclaved 35 g/L Instant Ocean® Sea Salt solution. The SalmoSim system's feed media was prepared in 2 litres with the combination of 35 g/L of Instant Ocean® Sea Salt, 10 g/L of the fishmeal or fishmeal-free diet (Table 1), 1 g/L freeze-dried mucous collected from the pyloric caecum, and 2 litres of deionised water. For the pre-growth phase, microbial communities from prepared inoculums were pre-grown in SalmoSim for four days to allow the establishment of these microbial communities.

Objective 1 - Longitudinal experiment

After the pre-growth phase, the longitudinal test with SalmoSim was run in a total of 40 days consisting 20 days of fishmeal diet then 20 days of fishmeal-free diet. Throughout this experiment and the following experiments for objectives 2 and 3, all physiochemical conditions were maintained with similar values measured in live adult salmon for the three separate gut compartments within the SalmoSim.

Objective 2 - Validation experiment

As part of the validation experiment, real salmon were also provided fishmeal and fishmeal-free diets in the 40 days feed cycle carried out with SalmoSim. The three gut compartments of these real salmon were collected and combined to be analysed and compared with the SalmoSim compartments.

Objective 3 - Feed trial experiment

In regard to the feed trial, only gut contents from SalmoSim were used in this experiment. SalmoSim compartments on day 20 were collected and combined to be analysed and compared with those on day 40. Days 20 and 40 were selected to represent the end of the fishmeal and fishmeal-free diet treatments respectively.

DNA extraction of SalmoSim 'gut' content samples and sequencing

As the SalmoSim experiment began, sampling from each bioreactor vessel was carried out every two days. For each gut compartment, a total of 21 samples were obtained over the 40 days experimental phase. DNA extraction of the study samples was carried out using the QIAamp® DNA Stool kit (Qiagen, Valencia, CA, USA) according to the manufacturer's protocol (Claassen et al., 2013).

From the eluted DNA, the V1 region of the 16S rRNa gene was amplified. Amplicons were sequenced on the Illumina HiSeq 2500 instrument. The full-length of V1-V9 regions of the 16S rRNA gene were also amplified with pooled samples. The amplicons from the pooled samples were then sequenced on the PacBio SMRT sequencing instrument, and the sequencing output was assembled into CCS reads.

Data processing

For the Illumina HiSeq reads, quality filtering and trimming (>Q30 quality score) were performed on the 16S rRNA V1 hypervariable region by using the Sickle software (Joshi & Fass, 2011). Read error correction was made using the BayesHammer module within the

SPAdes software to obtain high-quality assemblies (Nikolenko et al., 2013). Paired-end reads were overlapped at 50 bp with the PANDAseq software with simple_bayesian read merging algorithm (Masella et al., 2012; Schirmer et al., 2016). The resulting overlapped reads were dereplicated, sorted, and had their chimeras and singletons removed using the VSEARCH tool (Rognes et al., 2016).

After initial processing, the overlapped Illumina HiSeq reads were clustered into OTUs using the VSEARCH software at 97% identity. From these reads, an OTU table was generated in QIIME 2, which is a next-generation microbiome bioinformatics platform (Bolyen et al., 2019). This QIIME 2 OTU table was then converted to the biological observation matrix (BIOM) format to carry out diversity analyses. Microbial taxonomy of the OTUs was assigned using the QIIME 2 Naive Bayes feature classifier trained against the the Ribosomal Database Project (RDP) reference database (Cole et al., 2014; Bolyen et al., 2019). As an estimate of the microbial phylogeny, a MAFFT tree was generated in QIIME 2 (Katoh & Standley, 2013).

Alpha diversity analyses

To estimate the time required for microbial communities to stabilise in SalmoSim, longitudinal visualisations were performed on alpha diversity measures such as effective richness, Shannon diversity, and Simpson diversity. The changing alpha diversities amongst SalmoSim gut compartments were plotted over the 21 time points from days 0 to 40 in QIIME 2, with three replicates for each time point and compartment. With the OTU BIOM table, alpha diversity measures were calculated amongst sample types (real salmon, and SalmoSim) with the same diet for the validation experiment, and amongst different diets within SalmoSim for the feed trial experiment. With a custom script, these alpha diversity measures were carried out in the R programme (R Core Team, 2019) and included effective richness, Shannon diversity, and Simpson diversity. Alpha diversity differences by categorical variables in the metadata were tested using ANOVA in R.

Beta diversity analyses and microbial specificity

In R, beta diversity measures were calculated as weighted UniFrac distances for the validation amongst sample types, and feed trial amongst different diets. Principle-coordinate analyses (PCoA) were carried out on the weighted UniFrac distances and visualised with ggplot2 (Wickham, 2016) in R using a custom script. Tests for categorical differences in beta diversity were carried out using a permutational multivariance analysis of variance (PERMANOVA) (Anderson, 2014) as implemented in the vegan R package (Oksanen et al., 2013).

To observe the relative frequency of specific OTUs amongst sample types for the validation experiment, a taxa bar plot was constructed in R with a custom script. The same taxa bar plot was also constructed across samples with different feed types for the feed trial experiment. The OTUs in both taxa bar plots were focused on the genus level representing the top twenty OTUs.

Objective 4 - Comparison of database for taxonomic assignment

In this study, the public database for taxonomic assignment was the Ribosomal Database Project (RDP) reference database for the validation and feed trial experiments. To our knowledge, the RDP database is updated most regularly and allows taxonomic assignment at the species level (Cole et al., 2014). To construct the alternative long read sequence database, full-length Pacific Biosciences (PacBio) sequencing of our pooled study samples was utilised. Using VSEARCH (Rognes et al., 2016), the sequences of both RDP and alternative PacBio reads were dereplicated, sorted, followed by the removal of chimeras and singletons. OTUs were clustered at 100% identity to achieve a higher taxonomic resolution and assignment of OTUs at the species level. The Illumina HiSeq reads were BLAST-ed (Altshcul et al., 1990) against both RDP and PacBio reads as databases, and the outputs were parsed with a custom perl script to obtain the resulting E-values and bit-scores.

The BLAST E-value represents the number of expected hits of similar quality found by chance. Only database matches with E-values smaller than 1e-50 were counted, as this low

number of hits includes matches of high quality with very similar sequences. Whereas the bit-score is the required size of the sequence database for the individual matches to be found by chance. The higher the bit-score, the larger the database size required, which indicates lower precision of a database to the Illumina HiSeq reads.

As the RDP database contains taxonomy information, the SINTAX algorithm (Edgar, 2016) was used to predict taxonomy for queried Illumina HiSeq sequences. On the other hand, as our PacBio reads did not contain taxonomy information, they were first subjected to taxonomy annotation from the RDP database. After taxonomy annotation, the SINTAX algorithm was also utilised with the taxonomy annotated PacBio database to query Illumina HiSeq sequences. The workflow of this database comparison is visualised in Figure 3.

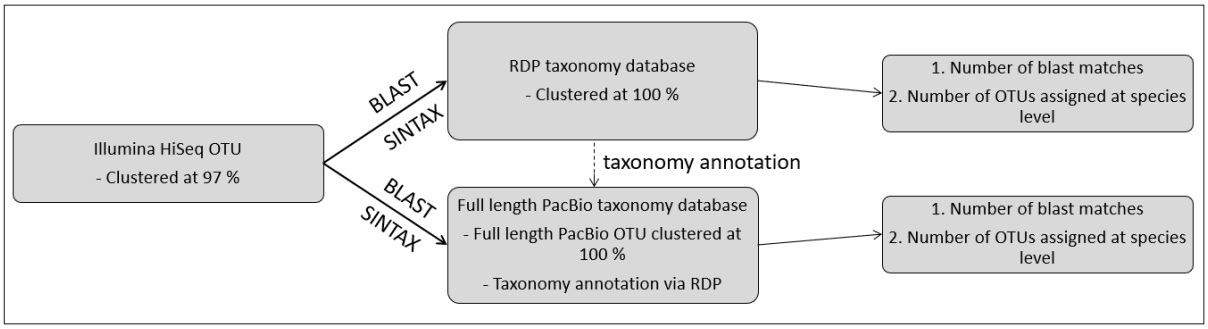
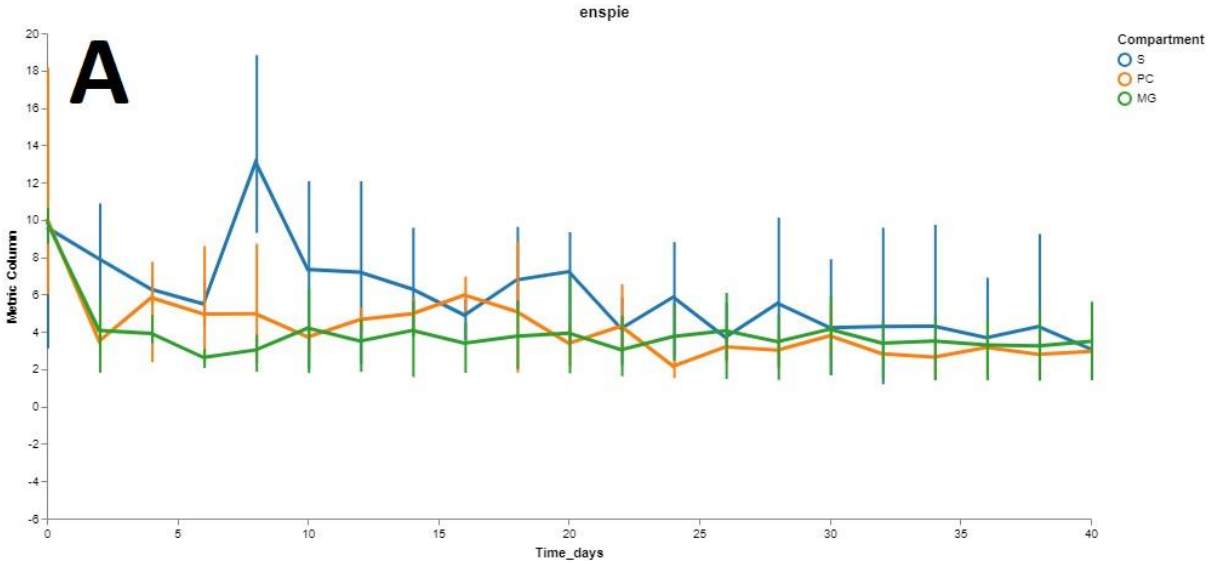


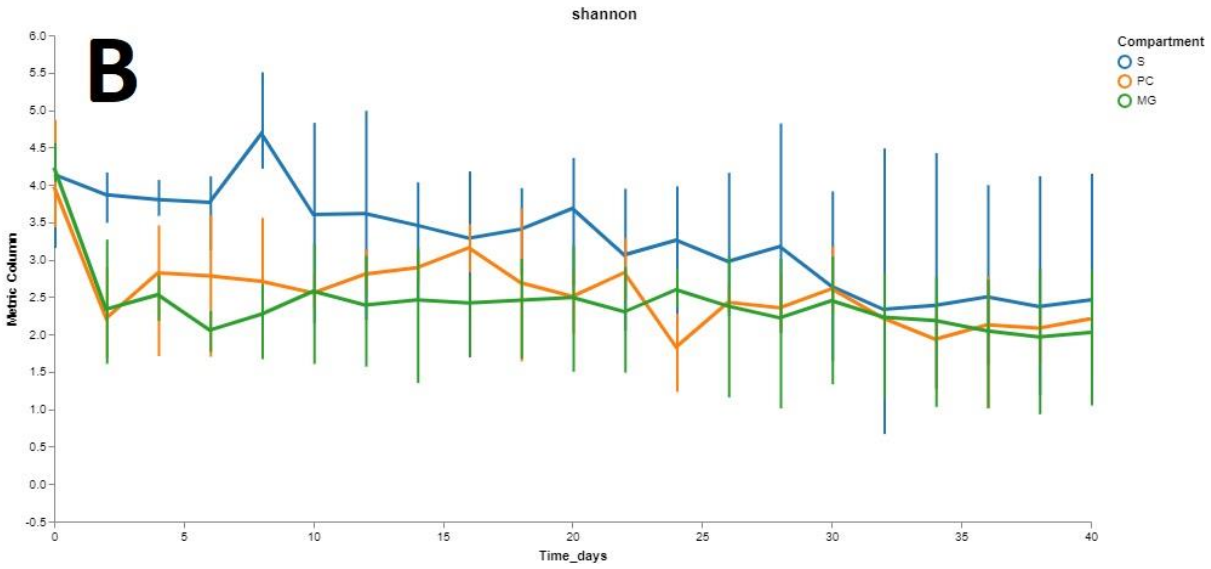
Fig. 3. Workflow for obtaining BLAST matches and number of operational taxonomic units (OTUs) assigned at species level from Illumina HiSeq reads. BLAST and SINTAX were carried out against two reference databases: the public Ribosomal Database Project (RDP) reference database, and custom full-length Pacific Biosciences (PacBio) reference database. The PacBio database was constructed with taxonomy annotation of the full-length PacBio OTU reads via the RDP database.

Results

SalmoSim alpha diversity over time

A total of 978 OTUs were identified from all 207 samples in this study (Data Set S1 metadata_all.csv). From all gut samples within the SalmoSim system, selected alpha diversity measures were recorded longitudinally from days 0 to 20 with fishmeal diet and days 20 to 40 with fishmeal-free diet. From days 0 to 40, there was very little change in Simpson diversity over time in different gut compartments (Fig. 4C). There were overall decreases for the measures of the effective richness (Fig. 4A) and Shannon diversity (Fig. 4B) from days 0 to 40, and initial changes in alpha diversity with all three compartments before stabilisation from different days onwards. The stomach experienced a sharp increase in alpha diversity from days 6 to 8, then decrease towards day 10 before stabilising. The pyloric caeca experienced a sharp decrease from days 0 to 2, then increase towards day 4 before stabilisation. Lastly, the alpha diversity of the midgut decreased sharply from days 0 to 2 then stabilised from day 2 onwards.


enspie
A
Compartment
S
PC
MG
Metric Column
Time_days


shannon
B
Compartment
S
PC
MG
Metric Column
Time_days

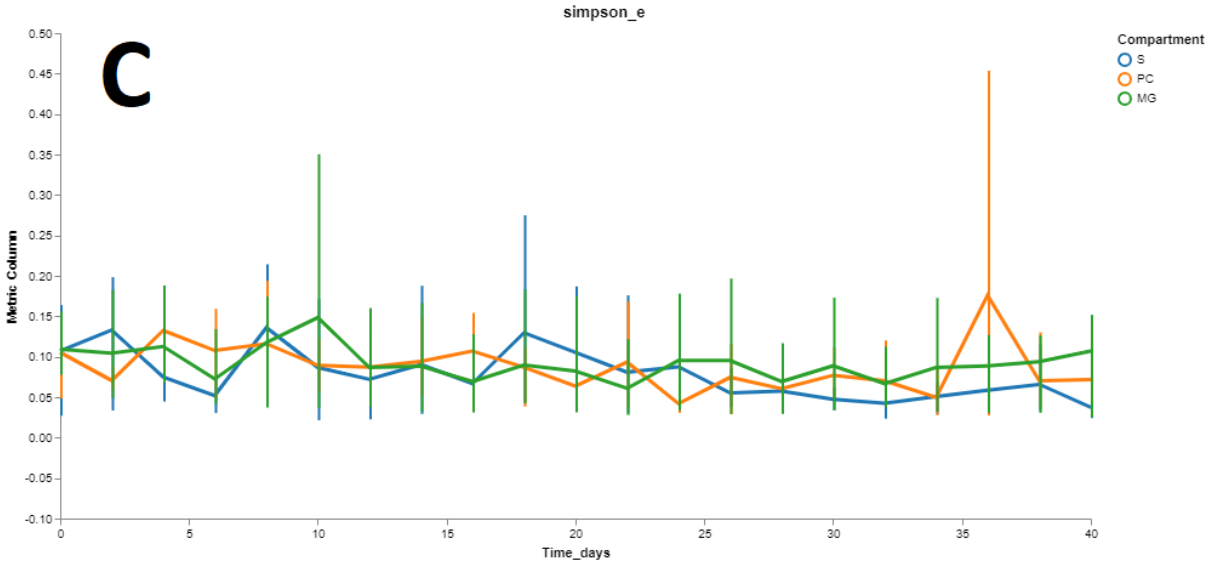

simpson_e
C
Compartment
S
PC
MG
Metric Column
Time_days

Fig. 4. Variance in alpha diversity of operational taxonomic units (OTUs) from days 0 to 40 amongst different SalmoSim compartments. The alpha diversity measures are (A) effective richness, (B) Shannon diversity, followed by (C) Simpson diversity representing the y-axes. Fishmeal diet was provided from days 0 to 20 and fishmeal-free from days 20 to 40 with sampling carried out every two days. Colours represent the three gut compartments: stomach (blue), pyloric caeca (yellow), midgut (green).

Validation experiment with SalmoSim

From the 18 fishmeal diet samples of real salmon and SalmoSim (day 20), the real salmon showed higher values than SalmoSim across all alpha diversity measures on average (Fig. 5). Amongst all the alpha diversity measures, only the Simpson diversity showed a significant difference at the 5% level (ANOVA, $p = 0.02$) between real salmon and SalmoSim samples (Fig. 5C). Effective richness (ANOVA, $p = 0.62$) and Shannon diversity (ANOVA, $p = 0.19$) did not show any significant difference at the 5% level between the sample types (Fig 5A; Fig. 5B).

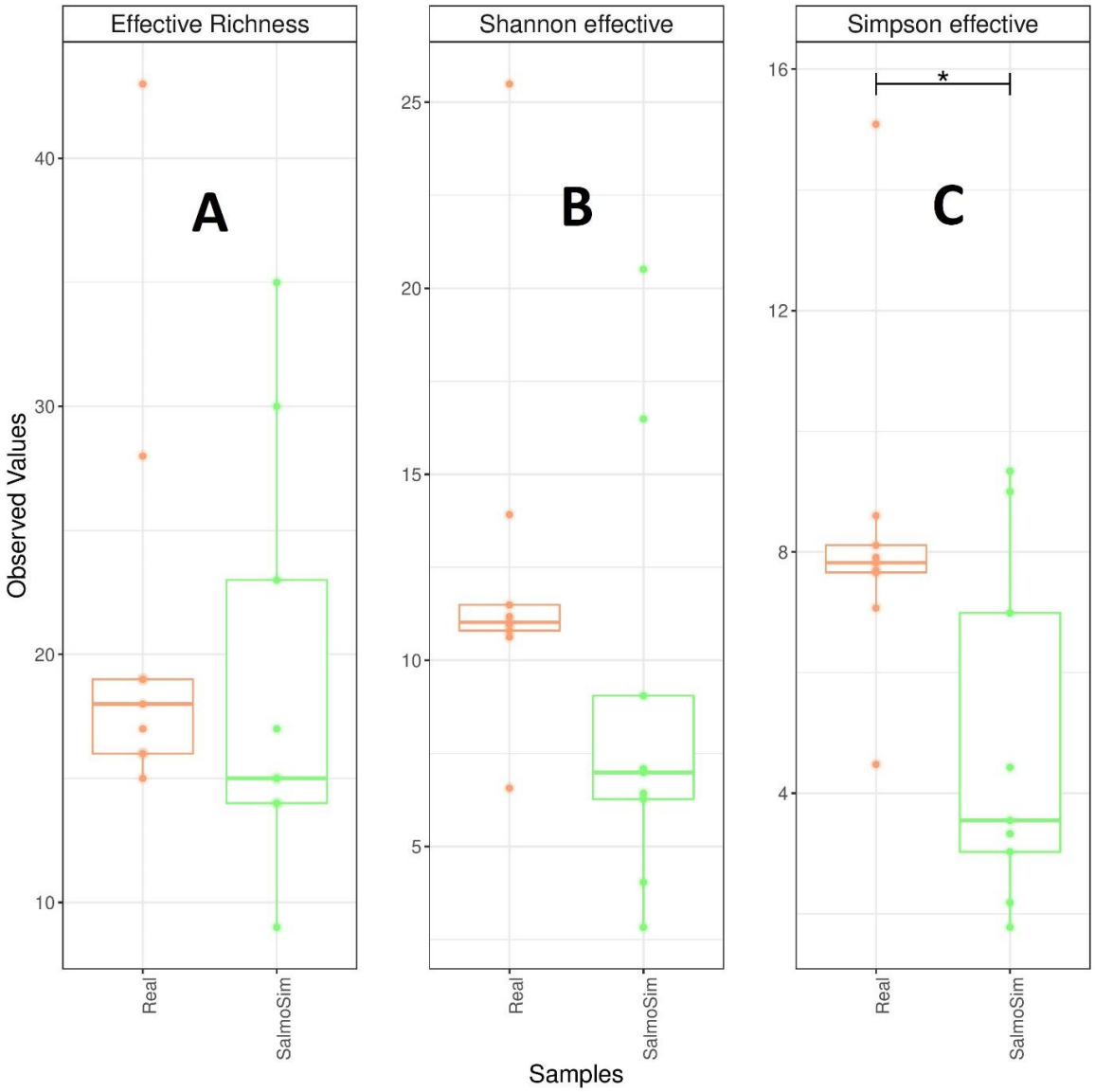


Fig. 5. Alpha diversity of operational taxonomic units (OTUs) amongst sample types (real salmon, and SalmoSim) under fishmeal diet. The alpha diversity measures are (A) effective richness, (B) Shannon diversity, followed by (C) Simpson diversity. Asterisks show significant differences between treatments with different fractions, with p-value < 0.05 (*).

With the remaining 18 fishmeal-free diet samples of real salmon and SalmoSim (day 40), the real salmon also showed higher alpha diversity than SalmoSim across all measures on average (Fig. 6). All alpha diversity measures showed significant differences at the 5% level with effective richness (ANOVA, p = 0.02), Shannon diversity (ANOVA, p = 1.60e-03), and Simpson diversity (ANOVA, p = 3.70e-07) between real salmon and SalmoSim samples (Fig. 6).

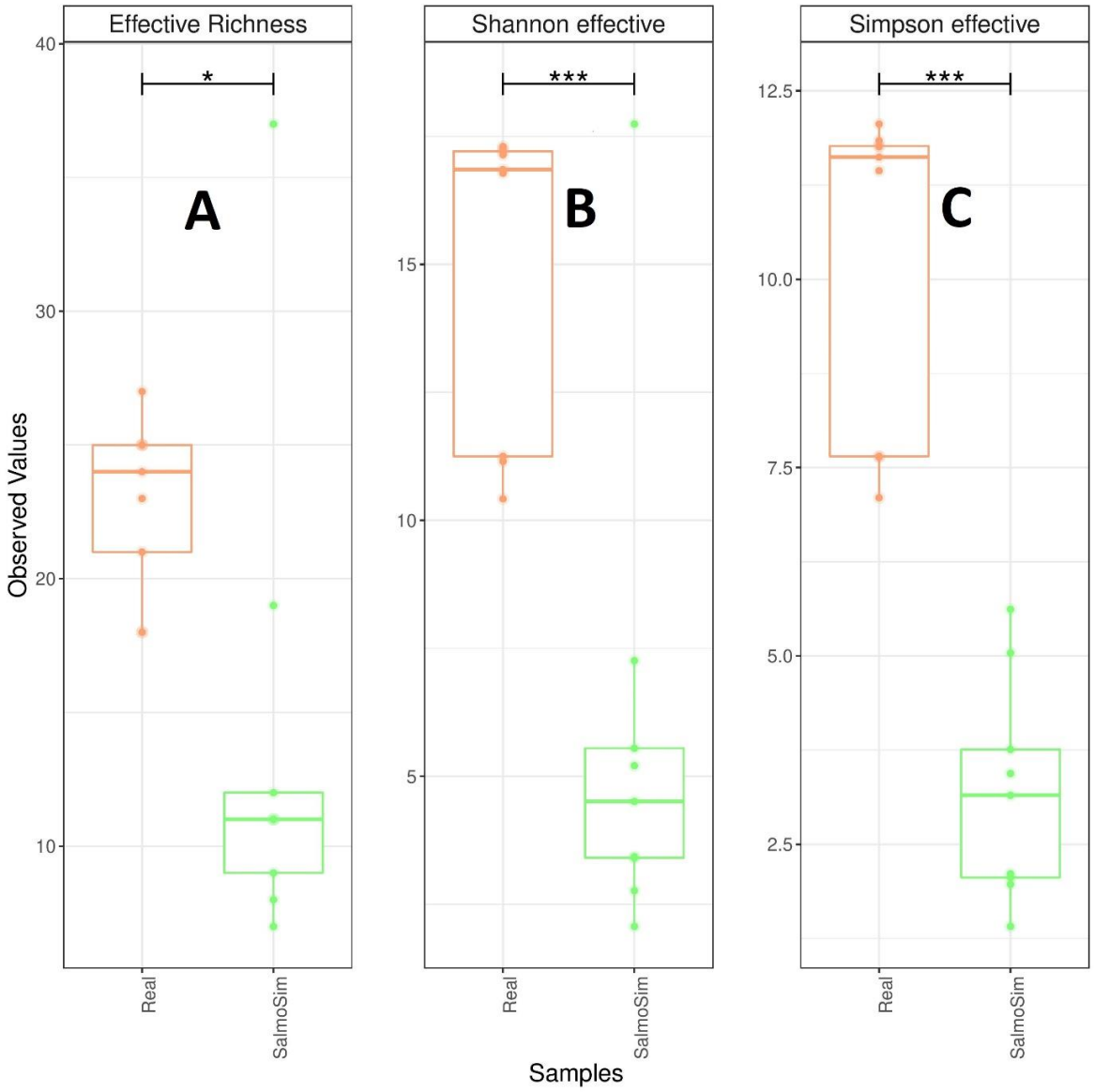


Fig. 6. Alpha diversity of operational taxonomic units (OTUs) amongst sample types (real salmon, and SalmoSim) under fishmeal-free diet. The alpha diversity measures are (A) effective richness, (B) Shannon diversity, followed by (C) Simpson diversity. Asterisks show significant differences between treatments with different fractions, with p-value < 0.001 (***), or p-value < 0.05 (*).

Several PERMANOVA analyses of the weighted UniFrac microbial beta diversity showed recorded variables contributing to the variation (Table 2). Unlike the variable of compartment (R^2 = 0.089, p = 0.743; R^2 = 0.071, p = 0.950), sample type was a significant variable at the 5% level and explained more variation at approximately 19% (R^2 = 0.185, p = 0.003) and 22% (R^2 = 0.215, p = 0.002) under fishmeal and fishmeal-free diets respectively.

Table 2. Permutational multivariate analysis of variance (PERMANOVA) of sample variables for the validation experiment, based on weighted UniFrac distance metrics. For the sample type variable, the two levels were either SalmoSim or real salmon. The R2 value represents the degree of variation in beta diversity explained by a variable.

Experiment	**Feed type**	**Variable**	**SumsofSqs**	**F**	**R2**	**Pr(>F)**
Validation: SalmoSim vs real salmon	Fishmeal	Sample type	0.675	3.624	0.185	0.003
		Compartment	0.327	0.736	0.089	0.743
	Fishmeal-free	Sample type	0.829	4.384	0.215	0.002
		Compartment	0.273	0.572	0.071	0.950

For the PCoA of the validation experiment, all samples under both fishmeal and fishmeal-free diets showed that real salmon formed a clear distinct cluster at the centre of the PCoA grid (Fig. 7). Unlike real salmon, SalmoSim samples did not display any clear distinct clusters regardless of diet, but loose clusters at corners of the PCoA grid with most points being quite dispersed (Fig. 7).

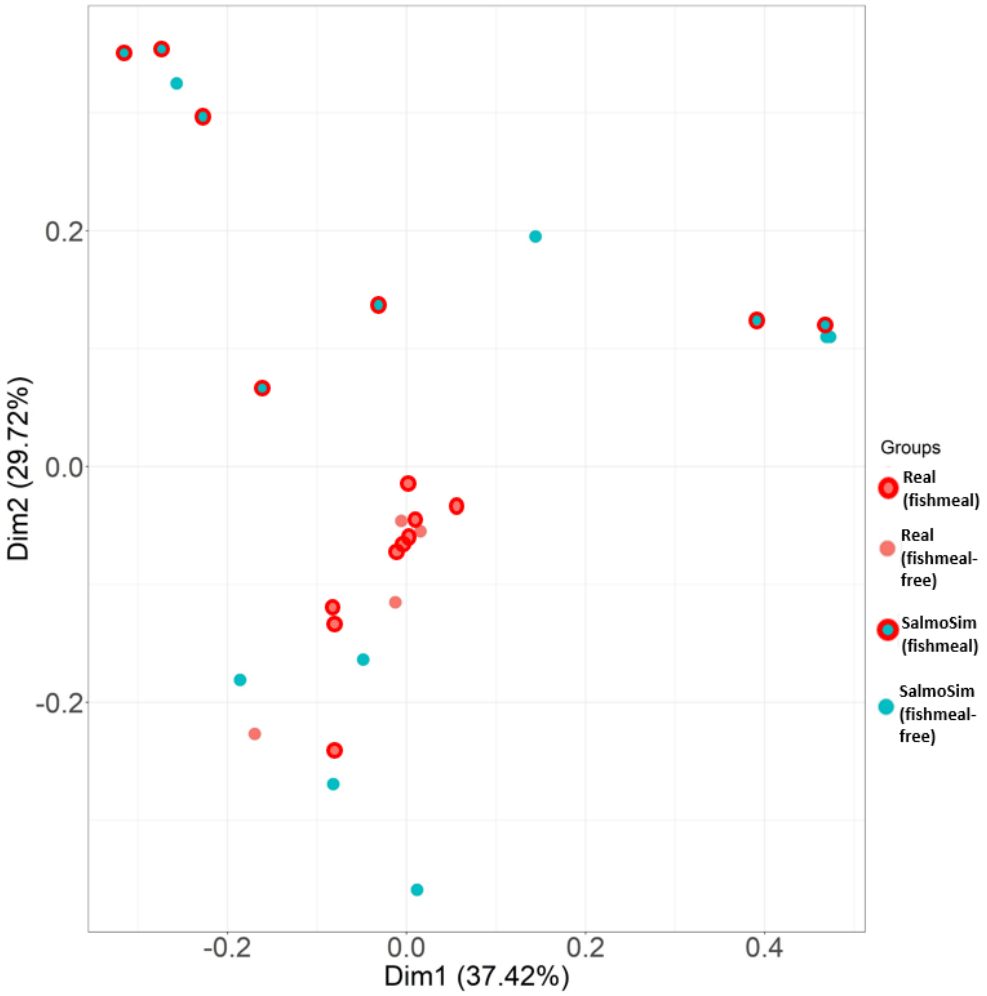


Fig. 7. Principal-coordinate analysis (PCoA) of weighted UniFrac distances amongst sample types (real salmon, and SalmoSim) with 36 samples altogether. Colours represent the two different sample types: real salmon (red), SalmoSim (blue). Red circles around the colours represent sample types under fishmeal diet, whilst colours without the red circle represent sample types under fishmeal-free diet.

A taxa bar plot was constructed to visualise the relative frequency of the top twenty OTUs at the genus level amongst sample types. In the plot across both sample types (Fig. 8), real salmon and SalmoSim samples were in general dominated by the *Psychrobacter* (OTU_27, OTU_16, OTU_18, OTU_22) and *Pseudomonas* genera (OTU_1, OTU_32, OTU_44), with a relatively moderate proportion of *Staphylococcus* (OTU_14) amongst the identified genera. The *Psychrobacter* and *Pseudomonas* genera were more equally present in individual real salmon samples, compared to SalmoSim samples which comprised a majority of either *Psychrobacter*, *Pseudomonas*, or *Staphylococcus* (Fig. 8).

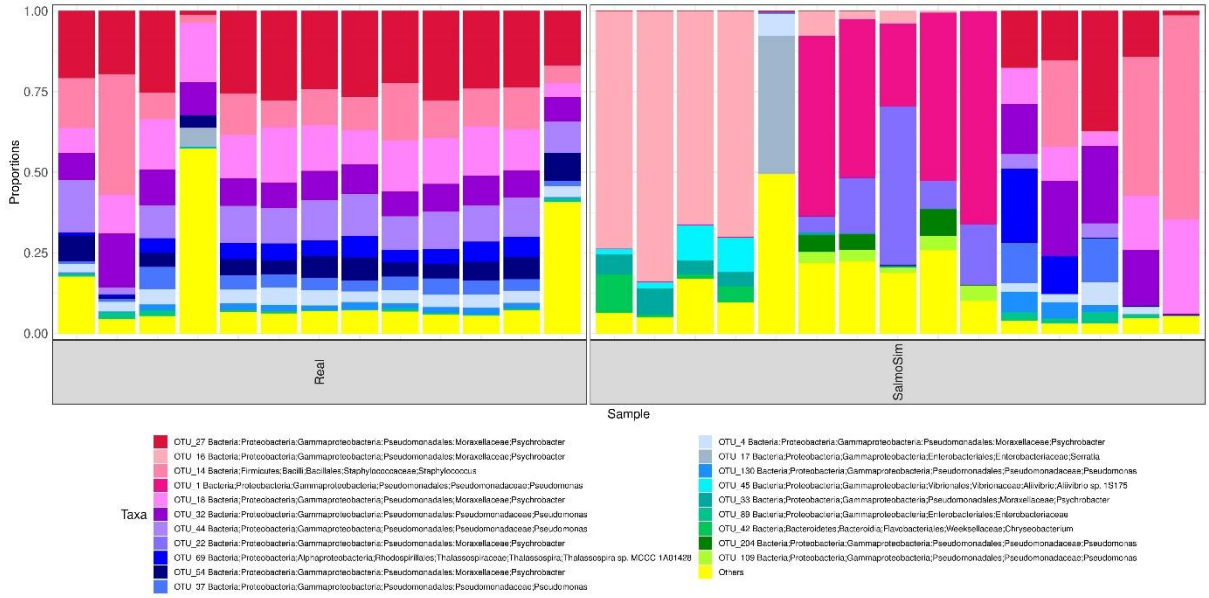


Fig. 8. Composition of the top twenty microbial operational taxonomic units (OTUs) at the genus level amongst sample types. The different sample types are represented by the labels on the x-axis: real salmon, SalmoSim.

Feed trial experiment with SalmoSim

From the 18 SalmoSim samples under fishmeal (day 20) and fishmeal-free (day 40) diets, fishmeal displayed higher alpha diversity than fishmeal-free across all measures on average (Fig. 9). However, none of these differences were significant at the 5% level whether with effective richness (ANOVA, $p = 0.23$), Shannon diversity (ANOVA, $p = 0.24$), or Simpson diversity (ANOVA, $p = 0.14$) between samples under fishmeal or fishmeal-free diets (Fig. 9).

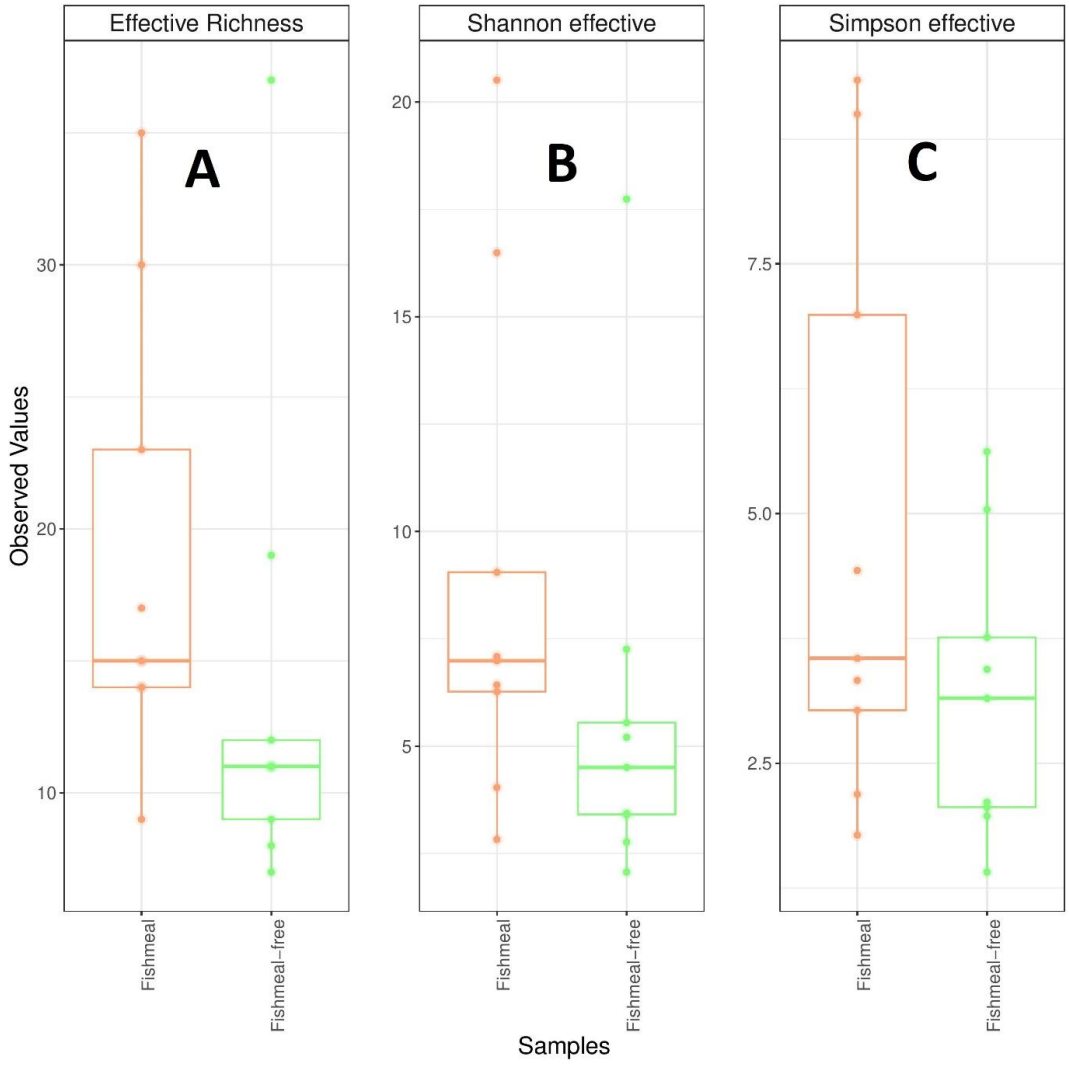


Fig. 9. Alpha diversity of operational taxonomic units (OTUs) amongst samples under different diets within SalmoSim. The alpha diversity measures are (A) effective richness, (B) Shannon diversity, followed by (C) Simpson diversity.

Based on the PERMANOVA analyses between the variables, compartment explained approximately 15% of the variation (R2 = 0.151, p = 0.175) which was more than the 3.5% variation explained by feed type (R2 = 0.035, p = 0.784). However, none of these focus variables were significant at the 5% level in contributing to the variation (Table 3).

Table 3. Permutational multivariate analysis of variance (PERMANOVA) of sample variables for the feed trial experiment, based on weighted UniFrac distance metrics. The R2 value represents the degree of variation in beta diversity explained by a variable.

Experiment	**Feed type**	**Variable**	**SumsofSqs**	**F**	**R2**	**Pr(>F)**
Feed type on SalmoSim	Fishmeal, fishmeal-free	Feed	0.148	0.588	0.035	0.784
		Compartment	0.630	1.333	0.151	0.175

For the PCoA, all samples under SalmoSim did not display any clear distinct clusters when grouped by the alternative variable of compartment (Fig. S1). Focusing on the diet for the feed trial experiment, no clear clustering was observed when samples were also grouped by feed type (Fig. 10).

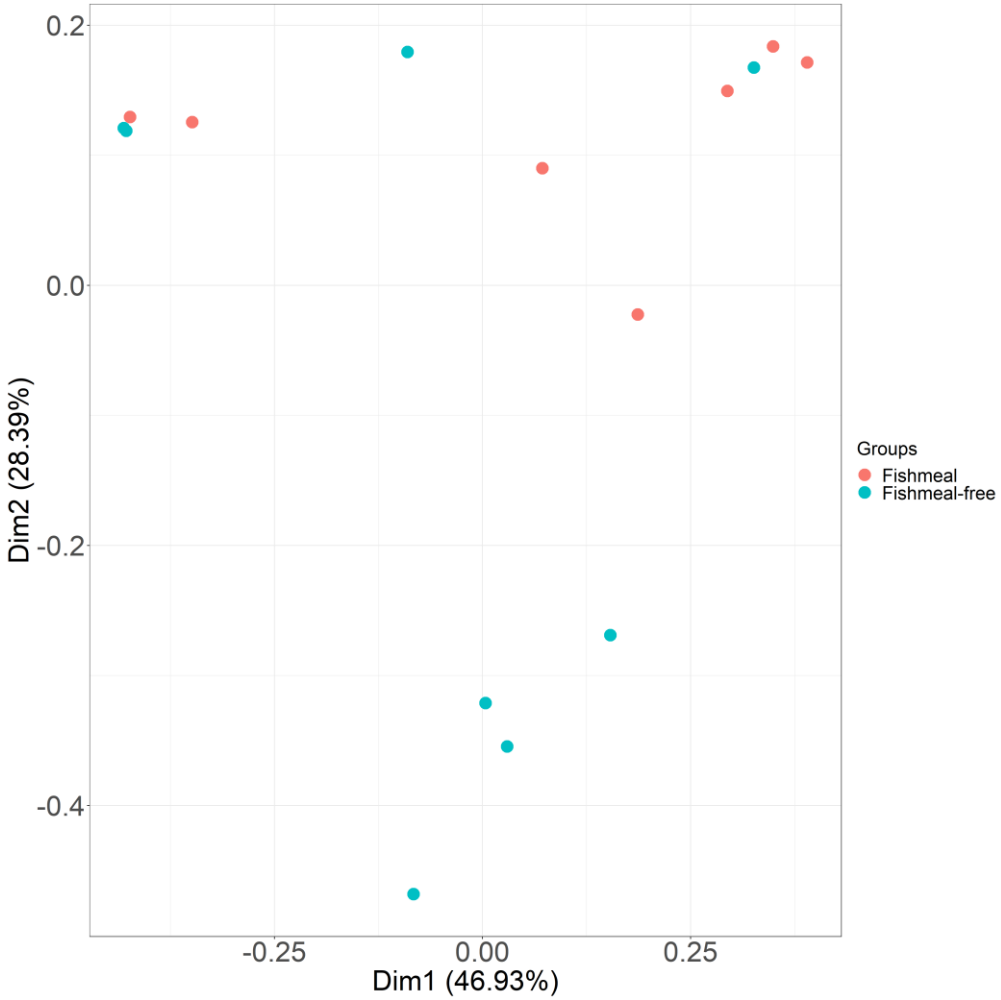


Fig. 10. Principal-coordinate analysis (PCoA) of weighted UniFrac distances amongst samples under different diets within SalmoSim. Colours represent the two different diets: fishmeal (red), fishmeal-free (blue).

A taxa bar plot was constructed to visualise the relative frequency of the top twenty OTUs at the genus level amongst feed types. In the plot across both fishmeal and fishmeal-free feed types (Fig. 11), most samples were dominated either by the *Psychrobacter* (OTU_16, OTU_22, OTU_27, OTU_18) or *Pseudomonas* genera (OTU_1, OTU_32, OTU_37). However, some fishmeal-free samples comprised a majority of the *Staphylococcus* genus (OTU_14) which was not found in fishmeal samples (Fig. 11).

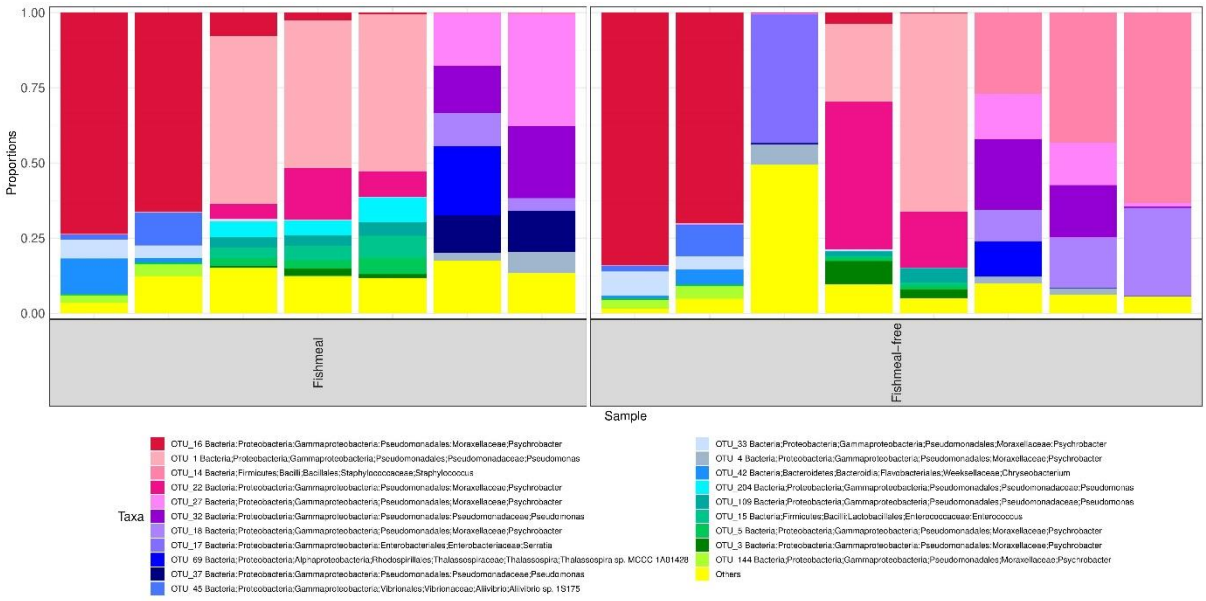


Fig. 11. Composition of the top twenty microbial operational taxonomic units (OTUs) at the genus level amongst samples under different diets within SalmoSim. The different diets are represented by the labels on the x-axis: fishmeal, fishmeal-free.

Database comparison for taxonomic assignment

The Illumina HiSeq reads were BLAST-ed against RDP and PacBio reads as databases, and only matches with BLAST E-values smaller than 1e-50 were counted (Data Set S2 blast_hiseq_rdp.csv; Data Set S3 blast_hiseq_pacbio.csv). With both the RDP and PacBio databases, the percentage of Illumina HiSeq OTUs matched were very similar (Fig. S2). However, the public RDP database displayed a higher average bit-score than the PacBio database (Fig. 12). When compared to the PacBio database, a larger sequence database is required of the public RDP database for individual OTU matches to be found by chance.

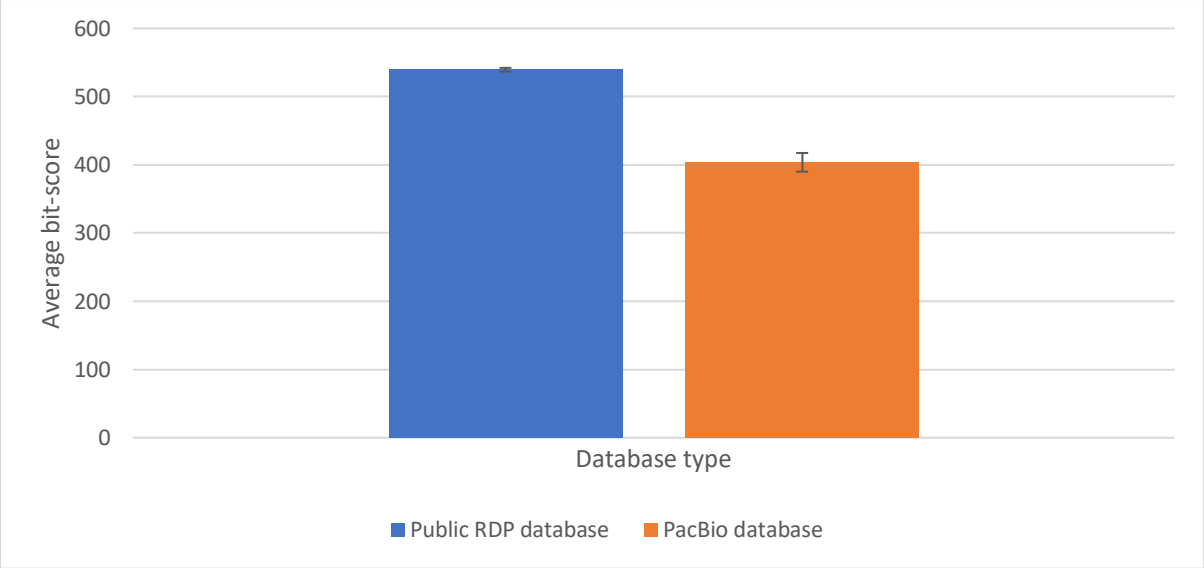


Fig. 12. The average bit-score of operational taxonomic units (OTUs) matched when Illumina HiSeq reads were BLAST-ed against the public Ribosomal Database Project (RDP) reference database, and full-length Pacific Biosciences (PacBio) reference database. For both databases, 16S rRNA sequences were clustered at 100% similarity into OTUs. Only matches with BLAST E-values smaller than 1e-50 were counted.

With taxonomic assignment via the SINTAX algorithm, the RDP database assigned 112 of the 978 Illumina HiSeq OTUs at the species level (Data Set S4 sintax_hiseq_rdp.csv). On the other hand, the PacBio database assigned 71 of the 978 Illumina HiSeq OTUs at the species level (Data Set S5 sintax_hiseq_pacbio.csv). Hence, the PacBio database assigned a lower percentage of Illumina HiSeq OTUs at the species level compared to RDP (Fig. 13). Both RDP and PacBio databases shared only ten OTUs assigned at the species level, with three of the ten OTUs assigned to the same species (Data Set S6 sintax_compare_rdp_pacbio.csv). Of the three OTUs assigned to the same species with both databases, two species were *Psychrobacter nivimaris* (OTU_5, OTU_18) and one species was *Stenotrophomonas rhizophila* (OTU_60) based on the SINTAX algorithm (Data Set S6 sintax_compare_rdp_pacbio.csv).

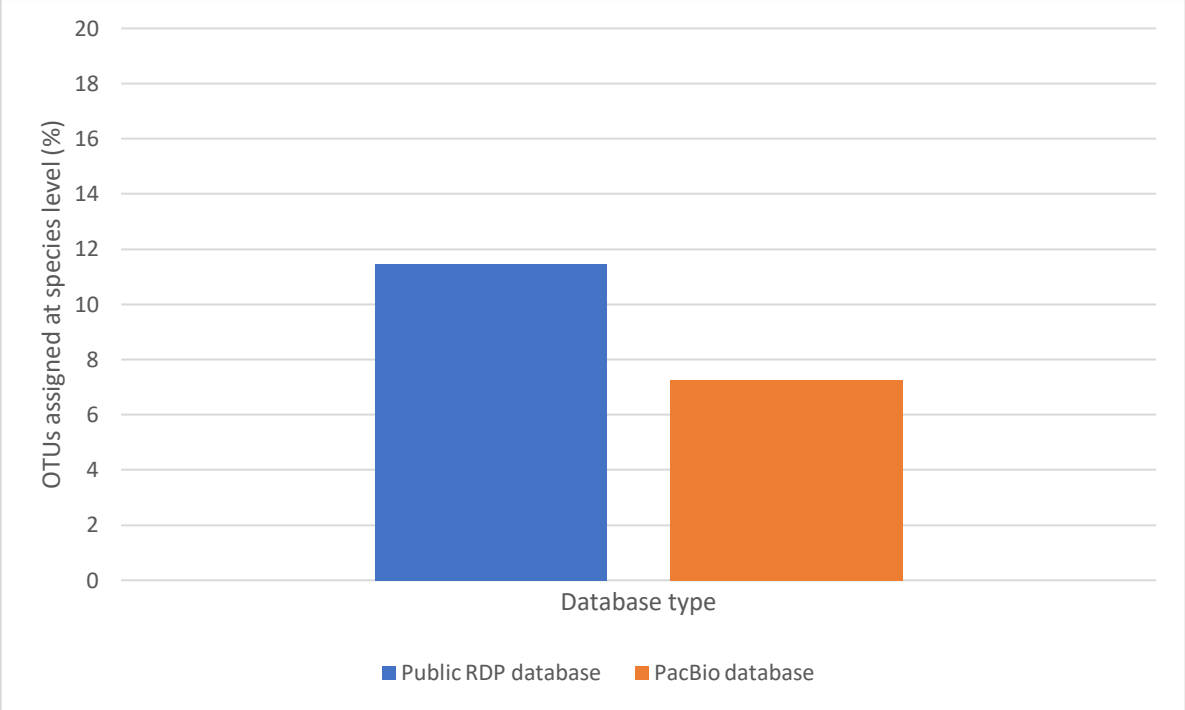


Fig. 13. The proportion of operational taxonomic units (OTUs) taxonomically assigned at the species level when Illumina HiSeq reads were queried against the public Ribosomal Database Project (RDP) reference database and full-length Pacific Biosciences (PacBio) reference database. For both databases, 16S rRNA sequences were clustered at 100% similarity into OTUs.

Conclusions and discussion

The four objectives of this study were 1) to estimate the time required for microbial communities to stabilise in SalmoSim, 2) to compare the microbial communities of *in vitro* SalmoSim to *in vivo* real salmon under certain feed types, 3) to compare the microbial communities of *in vitro* SalmoSim under different feed types, 4) to compare taxonomic assignment of microbial communities using a PacBio database as opposed to a public database.

For objective 1, we found overall decreases in alpha diversity over 40 days with the SalmoSim *in vitro* system, and microbiome oscillations before stabilisation. The stomach required more time before stabilisation in SalmoSim, followed by other compartments like the pyloric caeca then the midgut. In regard to the validation experiment for objective 2, the sample type of either SalmoSim or real salmon may have been a factor leading to their difference in microbiome composition. SalmoSim samples had a lower alpha diversity compared to real salmon. Also, the microbiome composition of SalmoSim was highly dissimilar from one another and did not indicate any core microbial communities which were observed with real salmon samples. With the feed trial experiment for objective 3, the feed types of either fishmeal or fishmeal-free did not show a significant effect on the resulting microbiome composition in SalmoSim samples. Without any significant factors explaining the different in microbiome composition of the samples, other factors might have played a crucial role but was not taken into account in this experiment. Lastly for objective 4, compared to the public RDP database, a smaller sequence database is required of the PacBio database for Illumina HiSeq OTU matches to be found by chance. The PacBio database assigned a lower percentage of Illumina HiSeq OTUs at the species level compared to RDP.

Stabilisation of microbial communities in SalmoSim

There were overall decreases in alpha diversity over 40 days with the SalmoSim *in vitro* system, with the exception of the Simpson diversity measure (Fig. 4). Besides counting the number of microbial species present in SalmoSim over time, the Simpson diversity measure takes into account the evenness of these microbial species as well (Simpson, 1949). In SalmoSim, changing alpha diversity may be more influenced by the presence or absence of

rare microbial species compared to common species, which corresponds to the minimal changes in Simpson diversity that takes into account the evenness of all species over time.

The SalmoSim system acts as a host gut for its inhabiting microbiomes introduced by the microbial inoculums. Microbiome oscillations often occur in a host gut environment to adjust to available metabolites from dietary introductions and physical environmental changes (El Aidy et al., 2013). Such oscillations were observed when SalmoSim was provided with fishmeal diet after the pre-growth phase (Fig. 4). During the microbiome oscillations, the stomach required the highest number of days before the microbial communities stabilised in SalmoSim, followed by other compartments like the pyloric caeca then the midgut (Fig. 4).

The time required for microbial communities to stabilise may be related to the pH conditions of the compartments in real salmon. The pH ranges are known to be between 3.0 and 4.5 in the stomach, between 7.0 and 7.5 in the pyloric caeca, and approximately 9.0 in the midgut compartment (Ransom et al., 1984; Ringø et al., 2003). The lower pH conditions of the SalmoSim stomach may have acted as an environmental filter, exerting stronger selective pressures on the reassembly of its microbial communities (Shaani et al., 2018). The stomach may be more resilient to microbial colonisation which makes it require more microbiome oscillations before the establishment of new microbial communities. On the other hand, increasing pH conditions in the SalmoSim pyloric caeca followed by the midgut may have exerted a weaker selective pressure on the reassembly of microbial communities, requiring fewer microbiome oscillations before the stabilisation in these SalmoSim compartments.

SalmoSim validation microbiome compositions

When constructing an *in vitro* system like SalmoSim to study the microbiome response to external food sources, it is crucial to ensure that SalmoSim's microbiome composition resembles that of real salmon guts. This study analysed gut microbiomes of SalmoSim and real salmon across different gut compartments. There were two potential main factors explaining the microbiome compositions amongst the two sample types: SalmoSim, real salmon. Compared to compartment, sample type was the factor that explained significantly

more variation for the gut microbiome (Table 2), representing greater influence of sample type on the salmon gut microbiome composition.

Amongst the relevant environmental factors, SalmoSim would have deviated slightly from the guts of real salmon in terms of osmotic pressure, affecting the composition of microbial communities in SalmoSim. The feed media of SalmoSim was prepared with sea salt solution which exposes some strains of bacterial taxa to high osmotic pressure, reducing the overall survival rate of SalmoSim bacterial taxa (Wood, 2015) relative to taxa in real salmon samples. Although *in vivo* live salmon are also exposed to high osmotic pressure as they 'drink' salt water (Kerstetter & White, 1994), they possess mechanisms like gills and kidneys to regulate their internal osmotic pressure (Perry et al., 2013) which may influence the gut compartments and its microbiome composition. Furthermore, in the epithelial cells of real salmon gut, there may be *Mycoplasma* which is a bacterial genus possibly enhancing the intestine's tolerance against osmotic stress (Cheaib et al., 2020). Despite its absence in the top twenty OTUs, *Mycoplasma* is known to be a dominant genus in the salmon gut environment (Holben et al., 2002; Llewellyn et al., 2016). The *Mycoplasma* genus may even be highly adapted to maintain its presence in the Atlantic salmon gut (Cheaib et al., 2020; Heys et al., 2020), and this may not be the case for the *in vitro* SalmoSim environment which indicated a disruption in its core microbial communities.

Besides the influence of its external environment, the SalmoSim system lacks mechanisms and potential symbiotic bacteria found in real salmon to maintain its internal osmotic pressure. Without the maintenance functions found in real salmon, the survival rate of SalmoSim bacterial taxa would decrease, explaining SalmoSim's lower alpha diversity compared to real salmon regardless of the feed type provided (Fig. 5; Fig. 6). On the contrary, the higher survival rate of real salmon bacterial taxa could be related to the clear distinct cluster of its samples in the PCoA grid, which was also far apart from most of the SalmoSim samples (Fig. 7). As they were able to regulate their internal osmotic pressure more effectively, the core microbial communities of real salmon were not disrupted, unlike SalmoSim samples which deviated from real salmon samples and did not share any core communities within its own group (Fig. 7). Overall, the lower survival rate of SalmoSim

bacterial taxa could also be related to the reducing SalmoSim diversity in the longitudinal experiment (Fig. 4).

Some of the core bacterial genera known to be in the salmon gut are Gram-negative *Pseudomonas* and *Psychrobacter* (Navarrete et al., 2008; Ringø et al., 2008; Hatje et al., 2014), as they were in large proportions across samples of our validation experiment (Fig. 8). However, in only some of the SalmoSim samples, the Gram-positive *Staphylococcus* was found to be a dominant genus (Fig. 8). Gram-positive bacteria is more resistant to high osmotic pressures than Gram-negative bacteria (Hill and White, 1929), which may by chance have allowed *Staphylococcus* to outcompete core *Pseudomonas* and *Psychrobacter* bacteria under the relatively suboptimal osmotic pressure of the SalmoSim system. Besides the factor of the osmotic pressure, the different microbiome composition in SalmoSim may have also been affected by other relevant environmental factors (de Bruijn et al., 2018) which deviated slightly from the levels found in real salmon guts. These environmental factors may include pH, oxygen level, temperature, and nutrients levels. Regarding the difference between real salmon and SalmoSim microbiomes, another explanation could be the effect of inter-individual variation as the SalmoSim inoculum was not collected from the same individual to the real salmon during sampling.

SalmoSim feed trial microbiome compositions

For the feed trial experiment, the main comparison was between the gut microbiomes of SalmoSim samples under fishmeal and fishmeal-free diets. However, the diet was not a significant factor explaining the variation for the gut microbiome at the 5% level, nor was the salmon gut compartment as the alternative factor (Table 3). The lack of significant factors may reflect an inter-sample variation in microbiome composition instead, even when samples are from the same salmon but different gut compartments. In agreement with this, there were no clear distinct clusters in the separate PCoA grids whether samples were grouped by diet or compartment (Fig. 10; Fig. S1). It is often challenging to determine the exact mechanisms in the assembly of the gut microbiome of fish (Roeselers et al., 2011). In some microbial environments, inter-sample variation may be influenced by stochastic processes whereby the first microbial coloniser determines the resulting introduction of subsequent microbes (Star et al., 2013). Indeed, a recent analysis indicates that stochastic processes are the primary drivers

underpinning the microbial colonisation of salmon gut by transient microbes (Heys et al., 2020). Instead of internal host factors, external environmental factors play a pivotal role in the assembly of microbial communities, based on neutral community models across gut compartments and life cycle stages of Atlantic salmon (Heys et al., 2020). Nevertheless, although none of the factors were significant at the 5% level, salmon gut compartment seemed to explain more variation for the gut microbiome relative to the factor of diet in SalmoSim (Table 3). The factor of gut compartment was also previously emphasised by a bacterial analysis across the Atlantic salmon gut, showing that bacterial communities differ across the gut compartments and even sections of the intestine itself (Gajardo et al., 2016).

Compared to fishmeal diets in the context of the feed trial, a recent analysis (Egerton et al., 2020) has also shown that plant-protein (fishmeal-free) diets decrease the alpha diversity of Atlantic salmon gut microbiome as seen with our samples (Fig. 9). Plant-based diets are linked to significant amounts of indigestible fiber, carbohydrates, and 'antinutrients' which negatively affects the gut microbiome composition and its digestive functions (Bakke-McKellep et al., 2007; Gajardo et al., 2017). Regarding the microbes present in the feed trial samples, most samples in both feed types were either dominated by the *Psychrobacter* or *Pseudomonas* genera, with the exception of some plant-based fishmeal-free samples containing high proportions of the *Staphylococcus* genus (Fig. 11). Plant-based diets have also previously been linked to shifts in higher *Firmicutes*:*Proteobacteria* ratio of phyla in salmon guts (Desai et al., 2012; Gajardo et al., 2017), potentially representing the increase of *Staphylococcus* to *Psychrobacter* and *Pseudomonas* within these phyla. With the fishmeal samples, the absence of *Staphylococcus* may be related to the presence of lactic acid bacteria (LAB) which is known to produce antimicrobial metabolites that hinder the growth of *Staphylococcus* (Hor and Liong, 2014). This explanation is supported by the small proportions of LAB in the *Lactobacillales* order (OTU_15) identified in our fishmeal samples when their microbiome compositions were visualised (Fig. 11).

Precision of PacBio long reads for taxonomic assignment

Throughout this study's validation and feed trial experiments, taxonomic assignment of Illumina HiSeq OTUs was carried out utilising the RDP reference database. To construct an alternative long read sequence database, full-length PacBio sequencing of our pooled study

samples was processed, then compared to the Illumina HiSeq database in terms of taxonomic assignment precision. At the species level, the alternative PacBio reads required a smaller sequence database for Illumina HiSeq OTU matches and was more specific to the Illumina HiSeq OTUs. Regarding the RDP database, it contains partial 16S rRNA gene sequences (Wang et al., 2007) and probably had lower coverage of the Illumina HiSeq V1 region than the PacBio database. The larger proportions of Illumina HiSeq OTUs matched with RDP may have contained imprecise assignment of closely related taxa, causing the identification of false positive microbial species (Berger et al., 2011). Hence, using the RDP database could display an inaccurate lack of diversity or inflated diversity of species.

On the other hand, the full-length PacBio database had a more complete coverage of the Illumina HiSeq reads (Franzén et al., 2015). The incorporation of full-length sequencing has previously shown the ability to enhance the classification of species correctly for ecological diversity analyses (Whon et al., 2018; Johnson et al., 2019). With the clustering of PacBio OTUs at a stringent 100% identity threshold, it enabled the discrimination between closely related taxa to study microbial diversity at the species level accurately.

As we compared the RDP with PacBio database in taxonomic assignment of Illumina HiSeq OTUs, both databases only shared ten OTUs assigned at the species level (Data Set S6 sintax_compare_rdp_pacbio.csv), which is a small overlap between both databases. Out of the ten overlapping OTUs, only three OTUs were assigned to the same species by both databases (Data Set S6 sintax_compare_rdp_pacbio.csv), which indicates an even smaller overlap when we focus on the particular species assigned. We speculate that these small overlaps between both databases could be due to differences in the pre-processing of the separate sequence datasets which were used to construct the databases. Otherwise, the small overlaps could also be due to some technical differences between full-length sequencing of PacBio reads and partial sequencing of queried Illumina HiSeq reads (Rhoads and Au, 2015), although this line of reasoning is also highly speculative and requires further understanding of the different sequencing platforms.

Of the three overlapping OTU species, two species were *Psychrobacter nivimaris* bacterium which is part of the predominant *Psychrobacter* genus across all our samples (Fig. 8; Fig. 11). In terms of 16S rDNA sequence, *P. nivimaris* is most similar to the *Psychrobacter proteolyticus* species commonly found in the Antarctic region (Heuchert et al., 2004), which could be the environment sourcing the salmon in MOWI Ltd's products (https://mowi.com/products/). The *P. proteolyticus* species was also found in the stomach of Antarctic krill (Denner et al., 2001) known to be consumed by Atlantic salmon. Given the molecular similarity between *P. nivimaris* and *P. proteolyticus* and diet of Atlantic salmon, both taxonomy databases in our analysis may have identified *P. nivimaris* originating from the natural external environment of MOWI Ltd's salmon.

The one remaining species identified by both RDP and PacBio databases was the *Stenotrophomonas rhizophila* species. The bacterium *S. rhizophila* is associated with plants (Wolf et al., 2002) which may explain why its genus (*Stenotrophomonas*) was not identified in the top twenty OTUs in our experiments. However, *S. rhizophila* was unanimously identified by both taxonomy databases (Data Set S6 sintax_compare_rdp_pacbio.csv) as it is purported to be resistant to a variety of antibiotics (Wolf et al., 2002), allowing *S. rhizophila* to be preserved in our study samples. Interestingly, *S. rhizophila* has also been isolated from rapeseed oil (Wolf et al., 2002) which constitutes nearly 12% of the ingredients in fishmeal and fishmeal-free feeds respectively. Hence, *S. rhizophila* in our study samples might have originated from the feeds provided, containing some percentage of rapeseed oil or other plant-based ingredients.

Limitations of methodology and analyses, and further avenues

Throughout this study, there were some limitations in the sample collection, study design, and analyses. Firstly, although there were a total of 207 samples across all experiments in this study, there were only 36 samples for the validation and feed trial experiments. With more sequence data in the SalmoSim project and the incorporation of other available datasets, similar experiments could have much larger sample sizes and statistical power. It would also be important to maintain equal sample numbers of compartments along the salmon gut, as these compartments vary in bacterial populations (Gajardo et al., 2016) and may display significant differences between compartments during *in vitro* trials.

Furthermore, the beta diversity analysis of the feed trial suggests that stochastic factors influenced the microbiome composition between different SalmoSim samples. To control for the stochastic factors influencing the microbial community assembly, *Mycoplasma* could be cultured and standardised in the communities of our study samples, as this bacterium is dominant and adapted to the gut environment of salmon under various conditions (Holben et al., 2002; Llewellyn et al., 2016; Cheaib et al., 2020; Heys et al., 2020). In our validation experiment, the SalmoSim inoculum was also not collected from the same individual to the real salmon during sampling. Future validation studies could ensure that both *in vivo* and *in vitro* samples are from the same salmon individual and compartment, as this procedure controls for potential inter-individual variation and allows us to conduct a more representative comparison between the *in vivo* and *in vitro* samples.

As different compartments required different number of days before stabilisation, the 40 days feed trial cycle could have been administered after the stabilisation in their respective compartments, instead of starting the cycle at the same time for all compartments as carried out in this study. Alternatively, before the collection of salmon gut compartments in aquaculture facilities, individual salmon may undergo antibiotic administration (Navarrete et al., 2008; Higuera-Llantén et al., 2018) or gut cleansing procedures (Jalanka et al., 2015) to produce a stable SalmoSim baseline community. Stabilising the SalmoSim microbial community to a baseline could inform us if this community responded consistently to different feed types and allow the reproducibility of certain outcomes.

In the feed trial experiment, only two levels of feed type were provided: fishmeal and fishmeal-free. Perhaps with varying percentages of fish or soymeal feed types, follow up studies may observe intermediate responses in bacterial diversity, or possibly change in patterns or trends in the microbiome. Another issue regarding the feed types may be the source of these feeds having an impact on the microbiome responses. Variation has been found in nutritional value between batches of the same plant protein from different sources (Urán et al., 2009). The difference between protein sources include factors like genetics, growing facilities, harvesting, processing, storage and so on (Urán et al., 2009). To control for such inter-source variations in the feeds, each feed type level could be represented by several sources of feeds with closely similar percentages of fish or soymeal content.

Although the taxonomic profiling provided the composition of sample microbiomes, the phenotypic characteristics of individual microbial taxa cannot be assessed by such compositional analyses. It is essential to elucidate the biological processes and functions carried out by the identified microbial taxa. With advancing sequencing technologies, targeted metatranscriptomics could be utilised to deepen our understanding of the functions provided by microbial taxa. The understanding of taxa functions may also be crucial as the alteration of the microbiome's functions could occur even without any changes in microbiome composition (Li et al., 2019). We could identify keystone taxa, microbial communities, and functions to be integrated with indigenous microbial communities via feeds with probiotics (Bentzon-Tilia et al., 2016; Stentiford et al., 2017). Such integrations may enhance the overall microbial functional potential, using probiotics to protect salmon from gut inflammation (de Bruijn et al., 2018). Moreover, the addition of bacterial secondary metabolites was shown to be a factor in the shifting of the gut microbiome (Li et al., 2019), which may also affect the microbiome functionality of *in vitro* gut systems. To take this factor into account, further studies could test and quantify the microbiome metabolites produced during *in vitro* microbial development.

Lastly, with the improved precision of the long read PacBio sequences as a reference database, further studies could employ phylogenetic methods to determine microbial diversity and composition more effectively (Matsen et al., 2010; Barbera et al., 2019). A reference phylogenetic tree could be constructed by long reads to bridge the phylogenetic gaps between references often left by the placement of short reads (Jamy et al., 2020). Ultimately, by observing whether long queries cluster with known sequences or create new clades on a phylogenetic tree, this resulting tree of prokaryotes may clarify the relationships between the queries to enhance the study of microbial diversity and composition (Jamy et al., 2020).

Summary

Our alpha diversity analyses over 40 days found overall decreasing diversity and microbiome oscillations before stabilisation in the SalmoSim *in vitro* system. The stomach required more time before stabilisation in SalmoSim, followed by other compartments like the pyloric caeca then the midgut. With the respective gut compartments in real salmon, the longer time required may be affected by the pH conditions in the compartment. Compared to the pyloric

caeca and midgut, the lower pH conditions in the stomach act as an environmental filter, requiring more microbiome oscillations during the establishment of its microbial communities.

In regard to the validation experiment, the sample type which were either SalmoSim or real salmon may have led to their difference in microbiome composition. Although SalmoSim was set up to have similar environmental conditions as real salmon guts, the slight difference in these conditions may have had an impact on the microbiome composition and the core microbes identified in the samples. Future *in vitro* experiments could control relevant environmental conditions to a greater degree, for example, by taking into account the mechanisms and potential symbiotic bacteria found in real salmon to maintain its internal osmotic pressure.

After the feed trial experiment, the feed types of either fishmeal or fishmeal-free showed little effect on the resulting microbiome composition in SalmoSim samples. Instead, the difference in microbiome composition might have been explained better by stochastic factors, whereby the colonising microbes between samples determined the introduction of subsequent microbes. To control for the potential stochasticity, the 40 days feed trial cycle could have been administered after the microbiome stabilisation of respective SalmoSim compartments, or at the same time after ensuring a standard baseline community across all compartments.

Lastly, full-length PacBio sequences of the study samples may enhance the precision of taxonomic assignment of microbial OTUs from Illumina short read length sequences. The PacBio sequences were annotated with a public taxonomy database to generate an alternative PacBio taxonomy database. With the difference between OTUs assigned by the public and full-length PacBio databases, both databases could be utilised in parallel to have a more holistic approach to taxonomic assignment. With the precision of taxonomic assignment and diversity analyses of microbial taxa, the functions of particular taxa could be identified with targeted metagenomics to understand their role in salmon gut compartments.

Overall, the present study emphasises the challenge of validating *in vitro* systems and carrying out *in vitro* experiments to simulate the gut microbiome development of salmon. Hence, with the microbiome data and ongoing project by SalmoSim, we provide a framework for future *in vitro* gut simulator designs to understand microbiome dynamics in response to feed trials. As a more innovative system, the aquaculture industry could utilise results from *in vitro* feed trials to identify feeds to farm salmon commercially and more sustainably compared to *in vivo* feed trials with sea cages. With successful *in vitro* feed trials, alternative feeds with probiotic potential could also be developed to ultimately improve the gut health of farmed salmon.

Bibliography

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J., 1990. Basic local alignment search tool. Journal of molecular biology, 215(3), pp.403-410.

Anderson, M.J., 2014. Permutational multivariate analysis of variance (PERMANOVA). Wiley statsref: statistics reference online, pp.1-15.

Bakke-McKellep, A.M., Penn, M.H., Salas, P.M., Refstie, S., Sperstad, S., Landsverk, T., Ringø, E. and Krogdahl, Å., 2007. Effects of dietary soyabean meal, inulin and oxytetracycline on intestinal microbiota and epithelial cell stress, apoptosis and proliferation in the teleost Atlantic salmon (Salmo salar L.). British Journal of Nutrition, 97(4), pp.699-713.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T. and Stamatakis, A., 2019. EPA-ng: massively parallel evolutionary placement of genetic sequences. *Systematic biology*, *68*(2), pp.365-369.

Beheshti Foroutani, M., Parrish, C.C., Wells, J., Taylor, R.G., Rise, M.L. and Shahidi, F., 2018. Minimizing marine ingredients in diets of farmed Atlantic salmon (Salmo salar): effects on growth performance and muscle lipid and fatty acid composition. PloS one, 13(9), p.e0198538.

Bentzon-Tilia, M., Sonnenschein, E.C. and Gram, L., 2016. Monitoring and managing microbes in aquaculture–Towards a sustainable industry. Microbial Biotechnology, 9(5), pp.576-584.

Berger, S.A., Krompass, D. and Stamatakis, A., 2011. Performance, accuracy, and web server for evolutionary placement of short sequence reads under maximum likelihood. Systematic biology, 60(3), pp.291-302.

Bolyen, E., Rideout, J.R., Dillon, M.R., Bokulich, N.A., Abnet, C.C., Al-Ghalith, G.A., Alexander, H., Alm, E.J., Arumugam, M., Asnicar, F. and Bai, Y., 2019. Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nature biotechnology, 37(8), pp.852-857.

Brown, C., 2015. Fish intelligence, sentience and ethics. Animal cognition, 18(1), pp.1-17.

Cheaib, B., Yang, P., Kazlauskaite, R., Lindsay, E., Heys, C., De Noa, M., Schaal, P., Dwyer, T., Sloan, W., Ijaz, U.Z. and Llewellyn, M.S., 2020. Unpicking the mysterious symbiosis of Mycoplasma in salmonids. bioRxiv.

Claassen, S., du Toit, E., Kaba, M., Moodley, C., Zar, H.J. and Nicol, M.P., 2013. A comparison of the efficiency of five different commercial DNA extraction kits for extraction of DNA from faecal samples. Journal of microbiological methods, 94(2), pp.103-110.

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R. and Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic acids research, 42(D1), pp.D633-D642.

Coll, M., Libralato, S., Tudela, S., Palomera, I. and Pranovi, F., 2008. Ecosystem overfishing in the ocean. PLoS one, 3(12), p.e3881.

de Bruijn, I., Liu, Y., Wiegertjes, G.F. and Raaijmakers, J.M., 2018. Exploring fish microbial communities to mitigate emerging diseases in aquaculture. FEMS microbiology ecology, 94(1), p.fix161.

De Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., Lara, E., Berney, C., Le Bescot, N., Probert, I. and Carmichael, M., 2015. Eukaryotic plankton diversity in the sunlit ocean. Science, 348(6237).

Denner, E.B., Mark, B., Busse, H.J., Turkiewicz, M. and Lubitz, W., 2001. Psychrobacter proteolyticus sp. nov., a psychrotrophic, halotolerant bacterium isolated from the Antarctic krill Euphausia superba Dana, excreting a cold-adapted metalloprotease. Systematic and applied microbiology, 24(1), pp.44-53.

Desai, A.R., Links, M.G., Collins, S.A., Mansfield, G.S., Drew, M.D., Van Kessel, A.G. and Hill, J.E., 2012. Effects of plant-based diets on the distal gut microbiome of rainbow trout (Oncorhynchus mykiss). Aquaculture, 350, pp.134-142.

Drieschner, C., Könemann, S., Renaud, P. and Schirmer, K., 2019. Fish-gut-on-chip: development of a microfluidic bioreactor to study the role of the fish intestine in vitro. Lab on a Chip, 19(19), pp.3268-3276.

Edgar, R.C., 2016. SINTAX: a simple non-Bayesian taxonomy classifier for 16S and ITS sequences. BioRxiv, p.074161.

Egerton, S., Wan, A., Murphy, K., Collins, F., Ahern, G., Sugrue, I., Busca, K., Egan, F., Muller, N., Whooley, J. and McGinnity, P., 2020. Replacing fishmeal with plant protein in Atlantic salmon (Salmo salar) diets by supplementation with fish protein hydrolysate. Scientific reports, 10(1), pp.1-16.

El Aidy, S., Derrien, M., Merrifield, C.A., Levenez, F., Doré, J., Boekschoten, M.V., Dekker, J., Holmes, E., Zoetendal, E.G., Van Baarlen, P. and Claus, S.P., 2013. Gut bacteria–host metabolic interplay during conventionalisation of the mouse germfree colon. The ISME journal, 7(4), pp.743-755.

Ellingsen, H., Olaussen, J.O. and Utne, I.B., 2009. Environmental analysis of the Norwegian fishery and aquaculture industry—A preliminary study focusing on farmed salmon. Marine Policy, 33(3), pp.479-488.

FAO, I. and UNICEF, 2018. WFP and WHO: The State of Food Security and Nutrition in the World 2018. Building climate resilience for food security and nutrition, 200.

Franzén, O., Hu, J., Bao, X., Itzkowitz, S.H., Peter, I. and Bashir, A., 2015. Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. Microbiome, 3(1), pp.1-14.

Gajardo, K., Jaramillo-Torres, A., Kortner, T.M., Merrifield, D.L., Tinsley, J., Bakke, A.M. and Krogdahl, Å., 2017. 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).

Gajardo, K., Rodiles, A., Kortner, T.M., Krogdahl, Å., Bakke, A.M., Merrifield, D.L. and 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, p.30893.

Green, T.J., Smullen, R. and Barnes, A.C., 2013. Dietary soybean protein concentrate-induced intestinal disorder in marine farmed Atlantic salmon, Salmo salar is associated with alterations in gut microbiota. Veterinary Microbiology, 166(1-2), pp.286-292.

Hatje, E., Neuman, C., Stevenson, H., Bowman, J.P. and Katouli, M., 2014. Population dynamics of Vibrio and Pseudomonas species isolated from farmed Tasmanian Atlantic salmon (Salmo salar L.): a seasonal study. Microbial ecology, 68(4), pp.679-687.

Heuchert, A., Glöckner, F.O., Amann, R. and Fischer, U., 2004. Psychrobacter nivimaris sp. nov., a heterotrophic bacterium attached to organic particles isolated from the South Atlantic (Antarctica). Systematic and applied microbiology, 27(4), pp.399-406.

Heys, C., Cheaib, B., Busetti, A., Kazlauskaite, R., Maier, L., Sloan, W.T., Ijaz, U.Z., Kaufmann, J., McGinnity, P. and Llewellyn, M.S., 2020. Neutral processes dominate microbial community assembly in Atlantic salmon, Salmo salar. Applied and environmental microbiology, 86(8).

Higuera-Llantén, S., Vásquez-Ponce, F., Barrientos-Espinoza, B., Mardones, F.O., Marshall, S.H. and Olivares-Pacheco, J., 2018. Extended antibiotic treatment in salmon farms select multiresistant gut bacteria with a high prevalence of antibiotic resistance genes. *PLoS One*, *13*(9), p.e0203641.

Hill, J.H. and White, E.C., 1929. Sodium chloride media for the separation of certain Gram-positive cocci from Gram-negative bacilli. Journal of bacteriology, 18(1), p.43.

Holben, W.E., Williams, P., Saarinen, M., Särkilahti, L.K. and Apajalahti, J.H., 2002. Phylogenetic analysis of intestinal microflora indicates a novel Mycoplasma phylotype in farmed and wild salmon. Microbial ecology, 44(2), pp.175-185.

Hor, Y.Y. and Liong, M.T., 2014. Use of extracellular extracts of lactic acid bacteria and bifidobacteria for the inhibition of dermatological pathogen Staphylococcus aureus. Dermatologica Sinica, 32(3), pp.141-147.

Jalanka, J., Salonen, A., Salojärvi, J., Ritari, J., Immonen, O., Marciani, L., Gowland, P., Hoad, C., Garsed, K., Lam, C. and Palva, A., 2015. Effects of bowel cleansing on the intestinal microbiota. *Gut*, *64*(10), pp.1562-1568.

Jamy, M., Foster, R., Barbera, P., Czech, L., Kozlov, A., Stamatakis, A., Bending, G., Hilton, S., Bass, D. and Burki, F., 2020. Long-read metabarcoding of the eukaryotic rDNA operon to phylogenetically and taxonomically resolve environmental diversity. *Molecular Ecology Resources*, *20*(2), pp.429-443.

Johnson, J.S., Spakowicz, D.J., Hong, B.Y., Petersen, L.M., Demkowicz, P., Chen, L., Leopold, S.R., Hanson, B.M., Agresta, H.O., Gerstein, M. and Sodergren, E., 2019. Evaluation of 16S rRNA gene sequencing for species and strain-level microbiome analysis. Nature communications, 10(1), pp.1-11.

Joshi, N.A. and Fass, J.N., 2011. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33)[Software].

Katoh, K. and Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), pp.772-780.

Kerstetter, T.H. and White, R.J., 1994. Changes in intestinal water absorption in coho salmon during short-term seawater adaptation: a developmental study. Aquaculture, 121(1-3), pp.171-180.

Kiron, V., Phromkunthong, W., Huntley, M., Archibald, I. and De Scheemaker, G., 2012. Marine microalgae from biorefinery as a potential feed protein source for Atlantic salmon, common carp and whiteleg shrimp. Aquaculture Nutrition, 18(5), pp.521-531.

Kopf-Bolanz, K.A., Schwander, F., Gijs, M., Vergeres, G., Portmann, R. and Egger, L., 2012. Validation of an in vitro digestive system for studying macronutrient decomposition in humans. The Journal of nutrition, 142(2), pp.245-250.

Li, L., Abou-Samra, E., Ning, Z., Zhang, X., Mayne, J., Wang, J., Cheng, K., Walker, K., Stintzi, A. and Figeys, D., 2019. An in vitro model maintaining taxon-specific functional activities of the gut microbiome. Nature communications, 10(1), pp.1-11.

Llewellyn, M.S., McGinnity, P., Dionne, M., Letourneau, J., Thonier, F., Carvalho, G.R., Creer, S. and Derome, N., 2016. The biogeography of the atlantic salmon (Salmo salar) gut microbiome. The ISME journal, 10(5), pp.1280-1284.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., Singer, D., Mayor, J., Bunge, J., Sernaker, S. and Siemensmeyer, T., 2017. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. Nature Ecology & Evolution, 1(4), p.0091.

Masella, A.P., Bartram, A.K., Truszkowski, J.M., Brown, D.G. and Neufeld, J.D., 2012. PANDAseq: paired-end assembler for illumina sequences. BMC bioinformatics, 13(1), p.31.

Matsen, F.A., Kodner, R.B. and Armbrust, E.V., 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC bioinformatics*, *11*(1), p.538.

Ménard, O., Cattenoz, T., Guillemin, H., Souchon, I., Deglaire, A., Dupont, D. and Picque, D., 2014. Validation of a new in vitro dynamic system to simulate infant digestion. Food chemistry, 145, pp.1039-1045.

Molly, K., Woestyne, M.V., Smet, I.D. and Verstraete, W., 1994. Validation of the simulator of the human intestinal microbial ecosystem (SHIME) reactor using microorganism-associated activities. Microbial ecology in health and disease, 7(4), pp.191-200.

Mosher, J.J., Bernberg, E.L., Shevchenko, O., Kan, J. and Kaplan, L.A., 2013. Efficacy of a 3rd generation high-throughput sequencing platform for analyses of 16S rRNA genes from environmental samples. Journal of microbiological methods, 95(2), pp.175-181.

Navarrete, P., Espejo, R.T. and Romero, J., 2009. Molecular analysis of microbiota along the digestive tract of juvenile Atlantic salmon (Salmo salar L.). Microbial ecology, 57(3), p.550.

Naylor, R.L., Goldburg, R.J., Primavera, J.H., Kautsky, N., Beveridge, M.C., Clay, J., Folke, C., Lubchenco, J., Mooney, H. and Troell, M., 2000. Effect of aquaculture on world fish supplies. Nature, 405(6790), pp.1017-1024.

Nikolenko, S.I., Korobeynikov, A.I. and Alekseyev, M.A., 2013, January. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. In BMC genomics (Vol. 14, No. S1, p. S7). BioMed Central.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. and Oksanen, M.J., 2013. Package 'vegan'. Community ecology package, version, 2(9), pp.1-295.

Olsen, Y., 2011. Resources for fish feed in future mariculture. Aquaculture Environment Interactions, 1(3), pp.187-200.

Pelletier, N., Tyedmers, P., Sonesson, U., Scholz, A., Ziegler, F., Flysjo, A., Kruse, S., Cancino, B. and Silverman, H., 2009. Not all salmon are created equal: life cycle assessment (LCA) of global salmon farming systems.

Perry, S.F., Shahsavarani, A., Georgalis, T., Bayaa, M., Furimsky, M. and Thomas, S.L.Y., 2003. Channels, pumps, and exchangers in the gill and kidney of freshwater fishes: their role in ionic and acid-base regulation. Journal of Experimental Zoology Part A: Comparative Experimental Biology, 300(1), pp.53-62.

Possemiers, S., Verthé, K., Uyttendaele, S. and Verstraete, W., 2004. PCR-DGGE-based quantification of stability of the microbial community in a simulator of the human intestinal microbial ecosystem. FEMS Microbiology Ecology, 49(3), pp.495-507.

Ransom, D.P., Lannan, C.N., Rohovec, J.S. and Fryer, J.L., 1984. Comparison of histopathology caused by Vibrio anguillarum and Vibrio ordalii in three species of Pacific salmon. Journal of Fish Diseases, 7(2), pp.107-115.

Rhoads, A. and Au, K.F., 2015. PacBio sequencing and its applications. Genomics, proteomics & bioinformatics, 13(5), pp.278-289.

Ringø, E., Olsen, R.E., Mayhew, T.M. and Myklebust, R., 2003. Electron microscopy of the intestinal microflora of fish. Aquaculture, 227(1-4), pp.395-415.

Ringø, E., Sperstad, S., Kraugerud, O.F. and Krogdahl, Å., 2008. Use of 16S rRNA gene sequencing analysis to characterize culturable intestinal bacteria in Atlantic salmon (Salmo salar) fed diets with cellulose or non-starch polysaccharides from soy. Aquaculture Research, 39(10), pp.1087-1100.

Roeselers, G., Mittge, E.K., Stephens, W.Z., Parichy, D.M., Cavanaugh, C.M., Guillemin, K. and Rawls, J.F., 2011. Evidence for a core gut microbiota in the zebrafish. The ISME journal, 5(10), pp.1595-1608.

Rognes, T., Flouri, T., Nichols, B., Quince, C. and Mahé, F., 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ, 4, p.e2584.

Schirmer, M., D'Amore, R., Ijaz, U.Z., Hall, N. and Quince, C., 2016. Illumina error profiles: resolving fine-scale variation in metagenomic sequencing data. BMC bioinformatics, 17(1), p.125.

Schloss, P.D., 2010. The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. PLoS Comput Biol, 6(7), p.e1000844.

Shaani, Y., Zehavi, T., Eyal, S., Miron, J. and Mizrahi, I., 2018. Microbiome niche modification drives diurnal rumen community assembly, overpowering individual variability and diet effects. The ISME journal, 12(10), pp.2446-2457.

Simpson, E.H., 1949. Measurement of diversity. nature, 163(4148), pp.688-688.

Star, B., Haverkamp, T.H., Jentoft, S. and Jakobsen, K.S., 2013. Next generation sequencing shows high variation of the intestinal microbial species composition in Atlantic cod caught at a single location. BMC microbiology, 13(1), p.248.

Stentiford, G.D., Sritunyalucksana, K., Flegel, T.W., Williams, B.A., Withyachumnarnkul, B., Itsathitphaisarn, O. and Bass, D., 2017. New paradigms to help solve the global aquaculture disease crisis. PLoS pathogens, 13(2), p.e1006160.

Tacon, A.G. and Metian, M., 2008. Global overview on the use of fish meal and fish oil in industrially compounded aquafeeds: Trends and future prospects. Aquaculture, 285(1-4), pp.146-158.

Team, R.C., 2016. R Project for Statistical Computing,(2019). Available online: www. R-project. org/(accessed on 8 October 2014).

Torrissen, O., Olsen, R.E., Toresen, R., Hemre, G.I., Tacon, A.G., Asche, F., Hardy, R.W. and Lall, S., 2011. Atlantic salmon (Salmo salar): the "super-chicken" of the sea?. Reviews in Fisheries Science, 19(3), pp.257-278.

Urán, P.A., Schrama, J.W., Jaafari, S., Baardsen, G., Rombout, J.H.W.M., Koppe, W. and Verreth, J.A.J., 2009. Variation in commercial sources of soybean meal influences the severity of enteritis in Atlantic salmon (Salmo salar L.). *Aquaculture Nutrition*, *15*(5), pp.492-499.

Van de Wiele, T., Van den Abbeele, P., Ossieur, W., Possemiers, S. and Marzorati, M., 2015. The simulator of the human intestinal microbial ecosystem (SHIME®). In The impact of food bioactives on health (pp. 305-317). Springer, Cham.

Van den Abbeele, P., Grootaert, C., Marzorati, M., Possemiers, S., Verstraete, W., Gérard, P., Rabot, S., Bruneau, A., El Aidy, S., Derrien, M. and Zoetendal, E., 2010. Microbial community development in a dynamic gut model is reproducible, colon region specific, and selective for Bacteroidetes and Clostridium cluster IX. Applied and environmental microbiology, 76(15), pp.5237-5246.

Wagner, J., Coupland, P., Browne, H.P., Lawley, T.D., Francis, S.C. and Parkhill, J., 2016. Evaluation of PacBio sequencing for full-length bacterial 16S rRNA gene classification. *BMC microbiology*, *16*(1), p.274.

Wang, Q., Garrity, G.M., Tiedje, J.M. and Cole, J.R., 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and environmental microbiology, 73(16), pp.5261-5267.

Whon, T.W., Chung, W.H., Lim, M.Y., Song, E.J., Kim, P.S., Hyun, D.W., Shin, N.R., Bae, J.W. and Nam, Y.D., 2018. The effects of sequencing platforms on phylogenetic resolution in 16 S rRNA gene profiling of human feces. Scientific data, 5, p.180068.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis. springer.

Wolf, A., Fritze, A., Hagemann, M. and Berg, G., 2002. Stenotrophomonas rhizophila sp. nov., a novel plant-associated bacterium with antifungal properties. International journal of systematic and evolutionary microbiology, 52(6), pp.1937-1944.

Wood, J.M., 2015. Bacterial responses to osmotic challenges. Journal of General Physiology, 145(5), pp.381-388.

Appendices

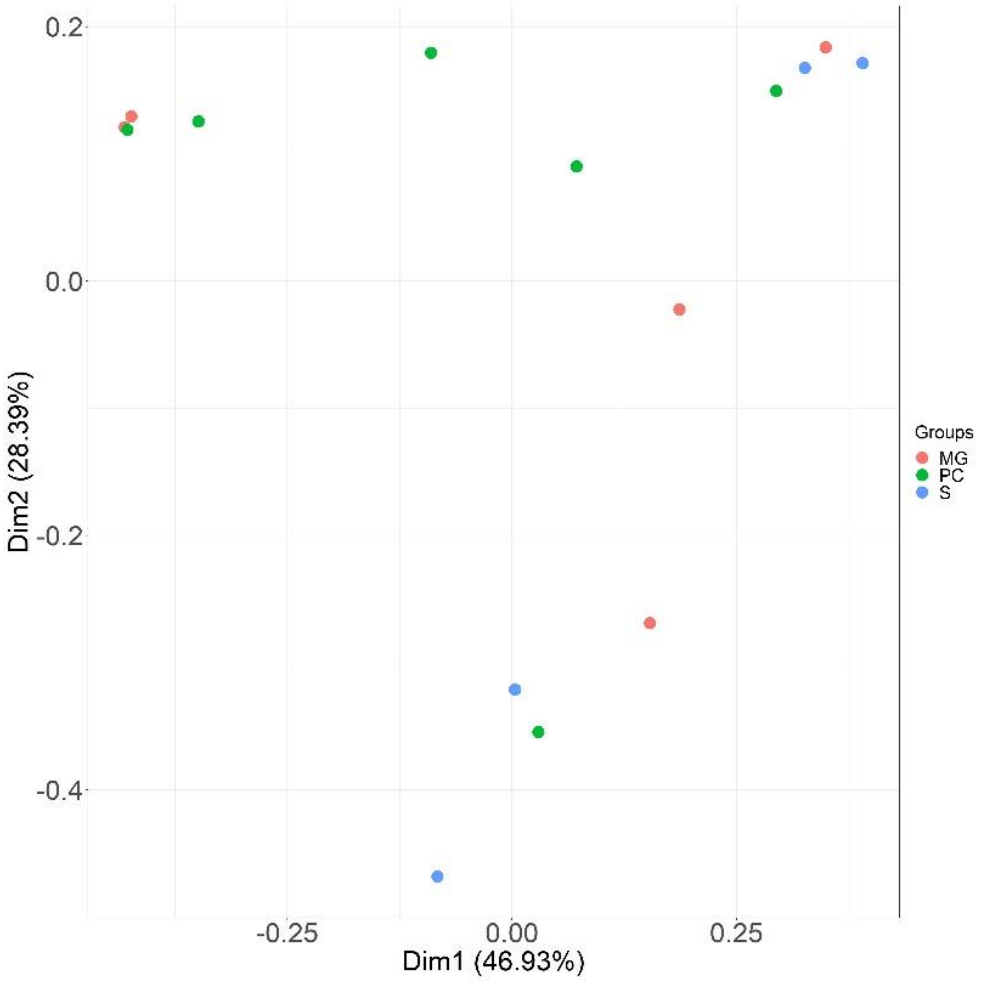


Fig. S1. Principal-coordinate analysis (PCoA) of weighted UniFrac distances amongst sample compartments (midgut, pyloric caeca, midgut) within SalmoSim. Colours represent the three different compartments: midgut (red), pyloric caeca (green), stomach (blue).

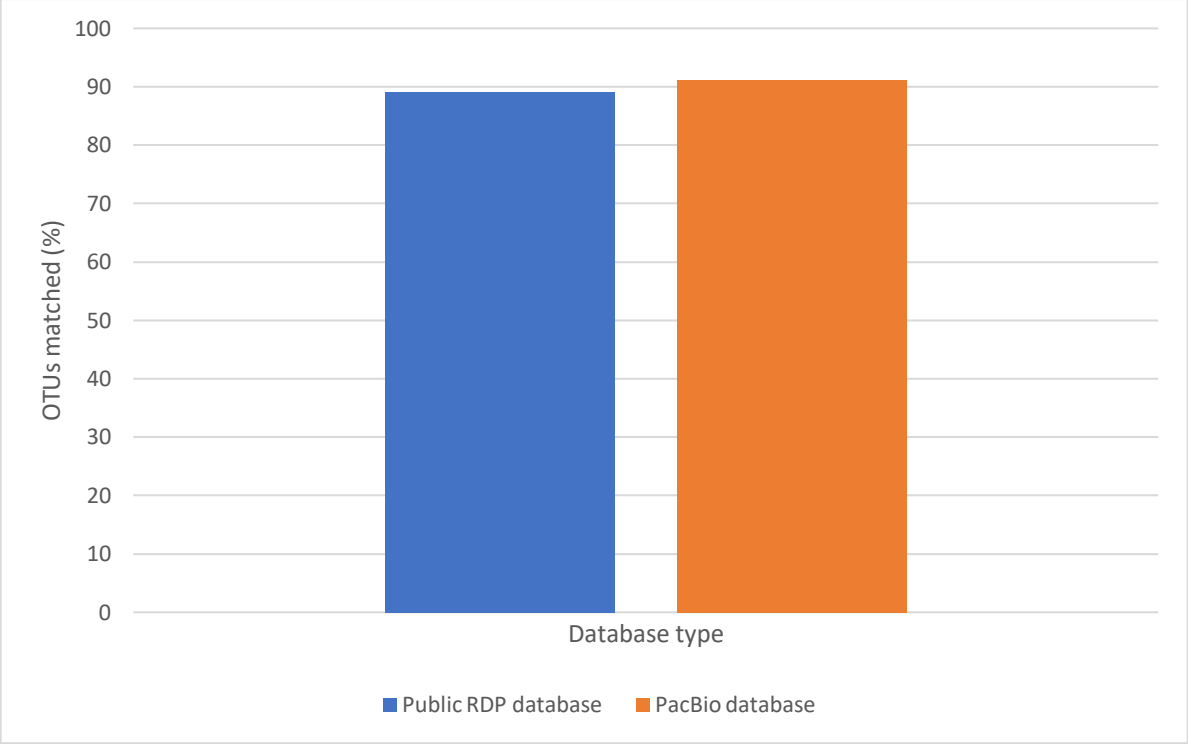


Fig. S2. The proportion of operational taxonomic units (OTUs) matched when Illumina HiSeq reads were BLAST-ed against the public Ribosomal Database Project (RDP) reference database and full-length Pacific Biosciences (PacBio) reference database. For both databases, 16S rRNA sequences were clustered at 100% similarity into OTUs. Only matches with BLAST E-values smaller than 1e-50 were counted.

Data Set S1 (separate file)

metadata_all.csv

Data Set S2 (separate file)

blast_hiseq_rdp.csv

Data Set S3 (separate file)

blast_hiseq_pacbio.csv

Data Set S4 (separate file)

sintax_hiseq_rdp.csv

Data Set S5 (separate file)

sintax_hiseq_pacbio.csv

Data Set S6 (separate file)

sintax_compare_rdp_pacbio.csv

Scripts in provided zipped ‘Appendix’ folder with files named:

1. R
2. QIIME2.txt
3. BLAST_and_SINTAX.txt

Acknowledgements

All data in this project were utilised from an ongoing project by SalmoSim (https://www.salmosim.co.uk/) which is a start-up company founded in the University of Glasgow. The SalmoSim system was set up by the SalmoSim team. The experimental and sequence data were collected and pre-processed by the SalmoSim team for the analyses carried out in this study. I thank my supervisor from the University of Glasgow and SalmoSim, Martin Llewellyn for organising this project with me, carrying it out, and refining it together as much as we could. I thank Martin Llewellyn, and Raminta Kazlauskaite from SalmonSim, for the set-up of the SalmoSim system, collection of the samples for our analyses, organising the samples, and producing and pre-processing of the molecular data for this project. I thank my supervisor from the University of Groningen, Bregje Wertheim, and Martin Llewellyn for providing valuable feedback and guidance throughout this project. I thank Bachar Cheaib, and Umer Ijaz for assisting me with the analyses and programme scripts involved in this project. Lastly, I thank my supervisors, work colleagues, family, and friends for all forms of support I received throughout this entire project.