A compositional reanalysis of poly(A)-selected RNA-seq reveals tank effects but no survival-associated differences

Marie Saitou[1], Arturo Vera Ponce De Leon[1], Akira Harding*[1-2], Celian Diblasi*[1], Domniki Manousi[1], Jun Soung Kwak[1-3], David G. Hazlerigg[4] T, Turid Mørkøre[1], Lars-Gustav Snipen[5],

[1] Department of Animal and Aquacultural Sciences, Section for Genome Biology, Faculty of Biosciences, Norwegian University of Life Sciences, 1433 Ås, Norway

[2] Joint Master's Programme in Marine Biotechnology, Valencia Catholic University Saint Vincent Martyr, 46001 Valencia, Spain

[3] Department of Aquatic Life Medicine, Kongju National University, Chungcheongnam-do, 32588, South Korea

[4] Department of Arctic and Marine Biology, UiT The Arctic University of Norway, N-9037, Tromsø, Norway

[5] Faculty of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, Ås, Norway

*These authors contributed equally to this work.

†Corresponding author: marie.saitou@nmbu.no, arturo.vera.ponce.de.leon@nmbu.no, lars.snipen@nmbu.no

email/ORCID:

Marie Saitou: marie.saitou@nmbu.no / https://orcid.org/0000-0002-8794-3217

Arturo Vera Ponce de Leon: arturo.vera.ponce.de.leon@nmbu.no / https://orcid.org/0000-0003-4085-5700

Akira Harding: akira.harding@mail.ucv.es / https://orcid.org/0009-0007-8930-157X

Célian Diblasi: celian.diblasi@nmbu.no / https://orcid.org/0009-0001-6548-0661

Domniki Manousi: domniki.manousi@nmbu.no / https://orcid.org/0009-0006-9428-9512

28 Jun Soung Kwak : jskwak@kongju.ac.kr / https://orcid.org/0000-0002-6830-5455

29 David Hazlerigg: david.hazlerigg@uit.no

30 Turid Mørkøre: turid.morkore@nmbu.no

31 Lars-Gustav Snipen: lars.snipen@nmbu.no

32

33 Author contributions (CRediT)

34 Akira Harding: Conceptualization, Validation, Formal analysis, Investigation

35 Celian Diblasi: Conceptualization, Methodology, Investigation, Writing – Original Draft

36 Domniki Manousi: Data curation

37 Jun Soung Kwak: Conceptualization, Methodology

38 David Hazlerigg: Resources, Funding acquisition

39 Turid Mørkøre: Resources, Visualization

40 Arturo Vera Ponce De Leon: Conceptualization, Methodology, Supervision

41 Lars-Gustav Snipen: Methodology, Supervision

42 Marie Saitou: Conceptualization, Investigation, Writing – Review & Editing, Supervision

43

44

45

46 **Abstract**

47 Poly(A)-selected RNA-seq datasets are routinely generated in aquaculture research, yet the microbial

48 information contained in unmapped reads is seldom explored due to the low abundance of nonhost

49 transcripts and concerns about contamination. In this study, we repurposed Atlantic salmon gill RNA-

50 seq data to assess whether meaningful microbial signals can be recovered using a contamination-

51 aware and compositionally appropriate framework. Unmapped reads were analyzed with a custom

52 Kraken2 database composed exclusively of complete, circularized salmon-associated bacterial

53 genomes together with all available Atlantic salmon assemblies and the human genome. Although

microbial sequences represented only a small fraction of total reads, 21 genera were detectable across samples. Genus-level profiles, Jaccard-based ordination, and ANCOM-BC analyses consistently revealed clear differences between tanks, whereas no associations were observed for sex or survival status. Three species exhibited significant tank-specific effects, indicating that environmental factors contributed the strongest detectable structure in the data. The limited microbial diversity recovered here reflects the expected constraints of poly(A)-enriched libraries, yet the results demonstrate that unmapped reads from host-derived RNA-seq can still provide informative environmental signatures when analyzed with curated reference databases and compositional statistical approaches. This strategy offers a practical means to extract exploratory microbiome information from existing transcriptomic datasets.

Keywords: poly(A)-selected RNA-seq; unmapped reads; contamination-aware microbiome profiling; Kraken2; compositional analysis; Atlantic salmon.

67

68 **Introduction**

69

70 Microbial dynamics in aquaculture systems strongly influence fish health, environmental stability, and
71 pathogen transmission (1). Although high-throughput sequencing has greatly advanced the characteri-
72 zation of aquatic microbiomes, most studies rely on dedicated 16S rRNA gene or metagenomic ap-
73 proaches, which are not always available for large production facilities or retrospective analyses (2). In
74 contrast, host-focused RNA-seq datasets are routinely generated in aquaculture research and breeding
75 programs yet the microbial information contained in unmapped reads is typically overlooked.

76 Poly(A)-selected RNA-seq is not designed for microbial profiling because bacterial transcripts generally
77 lack polyadenylation and because microbial RNA represents only a small fraction of total RNA in host
78 tissues (2, 3). Even so, unmapped reads in such datasets may retain detectable microbial signatures
79 when analyzed with contamination-aware reference databases and compositional statistical methods.
80 Recovery of such signals is particularly relevant in aquaculture, where operational and environmental
81 conditions, including water management and tank configuration, can generate distinct microbial envi-
82 ronments that influence fish health and pathogen exposure (4, 5).

83 Given that dedicated microbial sequencing is not always feasible, especially in commercial or high-
84 throughput production settings, repurposing existing host-derived RNA-seq datasets offers a practical
85 and cost-effective means to explore microbial information (6). Such an approach may help identify envi-
86 ronmental signatures without requiring additional sampling.

87 Here, we repurposed poly(A)-selected RNA-seq data from Atlantic salmon gill tissue to evaluate the ex-
88 tent to which microbial information can be recovered and whether such information reflects biological or
89 environmental structure. Using a contamination-aware Kraken2 database and compositional analyses,
90 we characterized the microbial taxa detectable in unmapped reads and assessed their variation across

91 sex, survival status, and tank identity. While microbial signals were limited in quantity, they were suffi-
92 cient to reveal consistent and statistically supported differences between tanks, whereas no associa-
93 tions were observed for sex or survival. These findings illustrate both the constraints and the utility of
94 host-derived poly(A)-selected RNA-seq data for exploratory microbial analyses in aquaculture systems
95 (7–9).

96

97

98

99

100 **Materials and Methods**

101 **Metatranscriptome dataset**

102 The metatranscriptome dataset analyzed in this study was derived from Atlantic salmon gill samples
103 collected in June 2021, one week prior to seawater transfer, at which time the fish showed no visible
104 signs of disease. These fish originated from 109 full-sib families of the Mowi (Mowi Genetics AS) strain,
105 which had been generated through selective breeding in November 2019. The metatranscriptome dataset
106 is publicly available at the European Nucleotide Archive (ENA) under project number PRJEB47441. This
107 study did not involve direct handling of fish; instead, it retrospectively analyzed transcriptomic data to
108 investigate microbial dynamics prior to disease onset. A comprehensive description of the fish population
109 is provided in (10) while details on metatranscriptome data generation are outlined in (11). The original
110 experiment was performed according to EU regulations concerning the protection of experimental
111 animals (Directive 2010/63/EU). Appropriate measures were taken to minimize pain and discomfort. The
112 experiment was approved by the Norwegian Food and Safety Authority (FOTS id. number 25658) (10).

113

114 **Fish Rearing and Sampling Timeline**

115 Fish were initially maintained under continuous light (LL) from hatching until they reached approximately
116 50g. They were then exposed to a 6-week 12:12 photoperiod (12 hours light, 12 hours dark) before
117 returning to LL for 8 weeks, reaching 100g, the standard size for seawater transfer.

118 In early February 2021, fish were vaccinated with Alpha Ject micro 6 (PHARMAQ®) against *Aeromonas*
119 *salonicida, Listonella anguillarum O1* and *O2a, Vibrio salmonicida, Moritella viscosa, and IPN virus Sp*.
120 In June 2021, one week before seawater transfer, gill biopsies were collected for RNA sequencing at a
121 time when fish showed no visible disease symptoms.

122 On June 23, 2021, fish were transported to the ARST seawater facility. Approximately eight weeks after
123 transfer, signs of winter ulcers began to appear, and mortality increased throughout the autumn. Despite
124 the removal of affected individuals during trait recordings in early November, the prevalence of winter
125 ulcers persisted into the winter. At the end of November, phenotypic data were recorded, marking the
126 conclusion of the experimental phase.

127

128 **Data Selection**

129 To allow for a balanced investigation of microbial communities in the context of fish health, sample
130 selection was based on multiple biological and experimental constraints. The dataset included 15
131 surviving fish, alive at five months post-transfer, and 13 non-surviving fish, which died within 40 days
132 post-transfer (**Table 1**). Sex was evenly distributed, with 13 males and 15 females, to account for potential
133 sex-specific microbial differences. Fish were reared in two separate tanks, C1 and C2, allowing the
134 assessment of environmental effects. The died group consisted exclusively of fish with visible wounds,
135 ensuring a consistent disease phenotype, while the survived group included only wound-free individuals,
136 providing a clear contrast (**Figure 1**). While the selection aimed for a balanced representation across
137 conditions, final sample availability was constrained by the accessibility of metatranscriptome data and
138 the biological conditions of the cohort.

139

140 **Metatranscriptome data preparation**

141 The RNA-seq data used in this study are publicly available in FastQ format from the European Nucleotide
142 Archive (ENA; project PRJEB47441). The dataset consists of gill metatranscriptome from Atlantic salmon,
143 with sequencing performed using a paired-end (PE) approach (2×150 bp) generating approximately 20
144 million reads per sample. Quality assessment of raw reads was conducted using FastQC (v0.12.0).
145 Adapter sequences, low-quality bases, and short reads were removed using fastp (12)(v0.23.4) with
146 default parameters. Filtered reads were then mapped to the Atlantic salmon reference genome (13)
147 (Ssal_v3.1, GCF_905237065.1) using bwa-mem2 (v2.2.1). Reads that failed to align with the host
148 genome were extracted and retained for microbiome analysis.

149 The RNA-seq libraries were prepared using a poly(A) selection method to enrich for mRNA, rather than
150 actively depleting ribosomal RNA (rRNA). While poly(A) selection primarily targets eukaryotic mRNA,
151 certain bacteria and some prokaryotic organisms possess polyadenylated mRNA at low frequencies,
152 which may result in the retention of microbial transcripts. Additionally, host-derived poly(A) transcripts
153 from Atlantic salmon may be present at a consistent proportion across samples.

154 For preprocessing, adapter trimming and quality filtering were performed using fastp (v0.23.4), followed
155 by host genome mapping against the Atlantic salmon reference genome (Ssal_v3.1, GCF_905237065.1).
156 Unmapped reads were retained for microbiome analysis, under the assumption that they predominantly
157 represent microbial sequences. However, given the poly(A) selection strategy, it is possible that a
158 proportion of these unmapped reads include artifacts from host transcripts or residual polyadenylated
159 bacterial transcripts. Although this potential artifact is present, it is expected to be uniformly distributed
160 across all samples rather than selectively biased toward a specific group (e.g., survived vs. died fish).
161 Therefore, while this factor must be considered when interpreting metatranscriptomic results, the relative
162 comparisons between survival groups remain valid, as any systematic bias introduced by poly(A)
163 selection would be consistent across conditions.

164

165 **Construction of a custom Kraken2 reference database**

166 A custom Kraken2 database was constructed to ensure contamination-free taxonomic classification and
167 to minimize false-positive assignments arising from incomplete or contaminated reference genomes. The
168 database consisted exclusively of: (1) bacterial genomes included in the Salmon Microbiome Genome
169 Atlas (SMGA(14)); (2) the Atlantic salmon genome Ssal_v3.1 (Ref Seq assembly GCF_905237065.1);
170 and (3) the complete human genome assembly GRCh38 (Ref Seq assembly GCF_000001405.26). All
171 bacterial reference genomes were retrieved from
172 https://ns9864k.web.sigma2.no/TheMEMOLab/projects/SMGA_v1.0/ and only assemblies annotated as
173 "complete genome (higher than 75 % complete and less than 5 % contamination according to ChekM2
174 scores (15))" were included for the bacterial component to reduce the likelihood of cross-species
175 contamination. The database was built in October 2025, following the standard Kraken2 build procedure.
176 The final database contained 211 bacterial genomes, the *S. salar* genome, and the human genome.
177 Kraken2 (v2.1.4) was used for taxonomic assignment of unmapped reads (16).

178

179 **Taxonomic assignment of unmapped reads**

180 Unmapped reads were classified using Kraken2. For each sample, species-level counts were extracted
181 from the Kraken2 report files by retaining entries assigned at rank "S" (species). Counts assigned to
182 *Salmo salar* and *Homo sapiens* were removed prior to analysis. The resulting species-by-sample table
183 was assembled by merging all Kraken2 report outputs. When species were absent from a given sam-
184 ple, counts were recorded as zero.

185 **Genus-level aggregation**

186 Species names were parsed to extract the genus (the first element of the species name). Genus-level
187 counts were obtained by summing species-level counts within each genus. For compositional summar-

188 ies, counts were further aggregated within each tank and converted to relative abundances. To visual-
189 ize dominant taxa, the eight genera with the highest total abundance across both tanks were retained
190 as individual categories, with all remaining genera grouped as “Others.”

191 **Phyloseq object construction**

192 A *phyloseq* object was created using sample metadata and the species-level count matrix. All down-
193 stream diversity and differential abundance analyses were performed within this framework.

194 **Differential abundance analysis**

195 Differential abundance testing was conducted using ANCOM-BC2(17). The model included Tank, Sex,
196 and Survival status as fixed effects. Benjamini–Hochberg correction was applied to control the false dis-
197 covery rate. Taxa were considered statistically significant when the q-value was less than 0.05 and the
198 differential abundance indicator for the corresponding coefficient was TRUE. Significant species associ-
199 ated with tank identity were extracted for reporting in Table 1.

200 **Alpha diversity**

201 Shannon and Simpson diversity indices were computed from species-level counts after removal of host
202 and human reads. Diversity values were compared across Tank, Sex, and Survival categories using
203 pairwise Wilcoxon rank-sum tests within each index, followed by Benjamini–Hochberg adjustment.

204 **Beta diversity**

205 Presence–absence community dissimilarity was quantified using Jaccard distances calculated from the
206 species-level count table. Classical multidimensional scaling (principal coordinates analysis) was ap-
207 plied to the Jaccard distance matrix. The first two principal coordinates were plotted to visualize com-
208 munity structure by tank identity, sex, and survival status.

209 **Software**

210 All analyses were performed in R using the packages *tidyverse*, *phyloseq*, *ANCOMBC*, *vegan*, and
211 *rstatix*(18–22). Scripts used for all processing steps, including taxonomic classification and statistical
212 analysis, are provided as supplementary material to ensure full reproducibility.

213 We used ChatGPT (version 5.2, OpenAI) to assist with code refactoring, language editing, and
214 proofreading. All scientific interpretations, analyses, and conclusions are the responsibility of the authors.
215
216
217 The scripts are available in the Figshare repository: https://doi.org/10.6084/m9.figshare.29482994
218
219
220

221 **Results**

222 We obtained an average of 537,995 ± 289,590 unmapped reads from a total of 11,328,436 ± 1,190,386
223 reads from the 28 samples. To characterize the microbial taxa detectable in the poly(A)-selected RNA-
224 seq data, species-level Kraken2 assignments were aggregated to the genus level after exclusion of reads
225 assigned to Salmo salar and Homo sapiens. Across all samples, 21 genera were observed. When genera
226 were ranked by their total counts across both tanks, the most frequently detected genera included
227 *Pseudomonas, Brucella, Paeniglutamicibacter, Micavibrio, Sphingomonas, Sphingobium, Williamsia,*
228 *and Kocuria.* These eight genera were retained as individual categories, and all remaining genera were
229 grouped as "Others." The resulting aggregated genus-level relative abundances for Tanks C1 and C2
230 are shown in **Figure 1,** illustrating the proportional contribution of these genera within each tank. To
231 summarize within-sample diversity, Shannon and Simpson indices were calculated from the species-
232 level microbial count table after removing reads assigned to *Salmo salar* and *Homo sapiens*. Diversity
233 values were compared across Sex, Survival status, and Tank categories. No differences were observed
234 for either Shannon or Simpson diversity between Sex groups (female versus male) or between Survival
235 groups (not survived versus survived), with Benjamini–Hochberg–adjusted P-values exceeding 0.33 for

236 all comparisons. In contrast, both Shannon and Simpson indices differed between Tank C1 and Tank C2,
237 with adjusted P-values of 0.012 for Shannon and 0.0049 for Simpson (**Figure 1B**). To examine sample-
238 to-sample differences based on taxonomic presence or absence, Jaccard dissimilarity was calculated
239 from the species-level read assignments. The first two principal coordinates are presented in **Figure 1C**.
240 Samples from each tank are displayed according to their tank identity, allowing the spatial arrangement
241 of samples to be viewed in ordination space. This representation indicates how samples relate to one
242 another based on shared and non-shared microbial species.

243 To identify species exhibiting statistically detectable differences across sample variables, ANCOM-BC2
244 was applied to the species-level count table with Tank, Sex, and Survival status included as covariates.
245 Three species showed a significant coefficient for the Tank contrast after Benjamini–Hochberg correction:
246 Brucella rhizosphaerae (lfc = −1.306, SE = 0.322, q = 0.00713), Pseudomonas qingdaonensis (lfc =
247 −1.321, SE = 0.322, q = 0.00713), and Pseudomonas iridis (lfc = −1.225, SE = 0.319, q = 0.00801). No
248 species exhibited significant associations with Sex or Survival status. These taxa are listed in **Table 1**.
249 The species identified as significant contribute to genera that appear within the genus-level profiles in
250 **Figure 1,** although genus-level aggregation does not yield a direct one-to-one correspondence between
251 the two representations.

252 In sum, C1 contained higher relative abundances of several dominant genera, whereas C2 showed a
253 more evenly distributed genus composition. Consistent with this, alpha diversity values were higher in C2
254 than in C1. ANCOM-BC2 identified three species belonging to the genera observed to be enriched in C1,
255 further reflecting the concentration of specific taxa within that tank.

256

257 **Discussion**

258 In this study, we examined microbial signatures detectable in poly(A)-selected RNA-seq data from
259 Atlantic salmon by reanalyzing unmapped reads using a contamination-aware taxonomic approach. After
260 removing reads assigned to the host and human genomes, only a modest number of microbial reads

261 remained, consistent with the expectation that poly(A) enrichment does not target prokaryotic transcripts.
262 While poly(A) selection bias is expected to be systematic across samples, it may still differentially affect
263 taxa depending on the degree of transcript polyadenylation (23, 24). Thus, despite inherent biases,
264 meaningful comparative analyses remain possible within the detectable subset of microbial transcripts.
265 Despite these constraints, the detectable microbial features revealed clear differences between tanks,
266 whereas neither sex nor survival status showed any evidence of association.
267
268 The genus-level community profiles and Jaccard-based ordination both indicated that samples clustered
269 according to tank identity. This pattern was further supported by differential abundance testing using
270 ANCOM-BC2, which identified three species with significant tank-associated effects. No species were
271 associated with sex or survival status. These results suggest that, within the limits of a poly(A)-selected
272 dataset, environmental factors related to tank conditions exert a stronger influence on detectable
273 microbial composition than fish-level biological traits (25).
274
275 The taxonomic composition observed here differed in part from profiles typically reported in 16S rRNA
276 gene sequencing or shotgun metagenomic studies of Norwegian aquaculture environments. While some
277 genera detected in this study, such as *Pseudomonas*, are also commonly reported components of marine
278 and aquaculture-associated microbiota, their relative abundances differed from those observed in
279 previous studies(26). In contrast, other taxa identified here, including *Brucella*, are not typically reported
280 as abundant members of aquaculture microbiomes and are rarely detected in comparable surveys. These
281 discrepancies likely reflect methodological differences, including the use of poly(A)-selected RNA-seq
282 data and stringent contamination-aware filtering, which together constrain the subset of microbial
283 transcripts that can be recovered. Combined with the use of a contamination-controlled database, these
284 factors likely restrict detectable diversity and shift relative abundances away from patterns observed with
285 less biased methods. Accordingly, the results presented here should be interpreted as reflecting the
286 subset of microbial signals that remain after host-dominated transcriptional data have been filtered rather
287 than as comprehensive estimates of microbial community structure. The importance of using

288 contamination-controlled reference databases has been underscored in other RNA-seq–based
289 microbiome studies. In particular, incomplete or contaminated bacterial genomes can generate extensive
290 false-positive assignments in human tumor RNA-seq data (27), and that restricting the database to
291 validated complete genomes markedly reduces such artefacts. This methodological consideration is
292 directly relevant here and provides context for the limited diversity and shifted relative abundances
293 observed in our dataset
294
295 Overall, our findings illustrate that microbial information present in poly(A)-selected RNA-seq data,
296 although limited, can still capture robust environmental signatures when analyzed with appropriate
297 contamination-aware and compositional methods. Microbiome count data are inherently compositional,
298 such that analyses based on unadjusted relative abundances can lead to misleading inferences (28).
299 This consideration further supports the use of ANCOM-BC2 in the present reanalysis and contextualizes
300 the tank-associated differences detected here. The tank-specific differences detected here demonstrate
301 that unmapped reads can provide biologically meaningful insights into environmental influences on
302 salmon-associated microbial communities. Future work comparing poly(A)-selected datasets with rRNA-
303 depleted metatranscriptomes or metagenomics may help to clarify how methodological choices shape
304 the detectable microbiome and to establish best practices for repurposing host-focused RNA-seq data
305 for microbial investigations.
306

307 **Author contributions (CRediT)**

308 Akira Harding: Conceptualization, Validation, Formal analysis, Investigation

309 Celian Diblasi: Conceptualization, Methodology, Investigation, Writing – Original Draft

310 Domniki Manousi: Data curation

311 Jun Soung Kwak: Conceptualization, Methodology

312 David Hazlerigg: Resources, Funding acquisition

313 Turid Mørkøre: Resources, Visualization

314 Arturo Vera Ponce De Leon: Conceptualization, Methodology, Supervision

315 Lars-Gustav Snipen: Methodology, Supervision

316 Marie Saitou: Conceptualization, Investigation, Writing – Review & Editing, Supervision

317

318 **Data Availability Statement**

319 The metatranscriptomic sequence data analyzed in this study are publicly available at the European

320 Nucleotide Archive (ENA) under project number PRJEB47441.

321 The scripts and annotation files used for taxonomic and functional profiling are available in the Figshare

322 repository: https://doi.org/10.6084/m9.figshare.29482994

323 No new sequencing data were generated as part of this study.

324

325 **Conflicts of interest**

326 The authors declare no competing interests.

327 **Funding information**

328 This study was supported by the following funding bodies: The Research Council of Norway (project

329 numbers: 325874). The Norwegian Seafood Research Fund (FHF), through the SynchroSmolt project:

330 “Smolt production protocols and breeding strategies for synchronized smoltification”. The EU Erasmus+

331 Programme. BIOVIT, Norwegian University of Life Sciences (PhD funding support).

332

Acknowledgements

The authors thank the Galaxy.no and Orion high-performance computing infrastructures for providing computational resources. We are also grateful to Côme Morel (CIRAD-BIOS, UMR ASTRE) for his valuable advice on alpha and beta diversity analyses.

338
339
340

341 Tables and Figures

342 **Table 1. Most transcriptionally represented microbial taxa in Atlantic salmon gill**

343 **metatranscriptomes.**

taxon	lfc_TankC2	se_TankC2	q_TankC2
Brucella rhizosphaerae	-1.306	0.322	0.00713
Pseudomonas qingdaonensis	-1.321	0.322	0.00713
Pseudomonas iridis	-1.225	0.319	0.00801

344

345 Species exhibiting significant differential abundance between Tanks C1 and C2 according to ANCOM-

346 BC2. Taxa were considered significant when the differential abundance indicator for the tank contrast

347 was TRUE and the FDR-adjusted q-value < 0.05. Negative log-fold change (lfc) values indicate higher

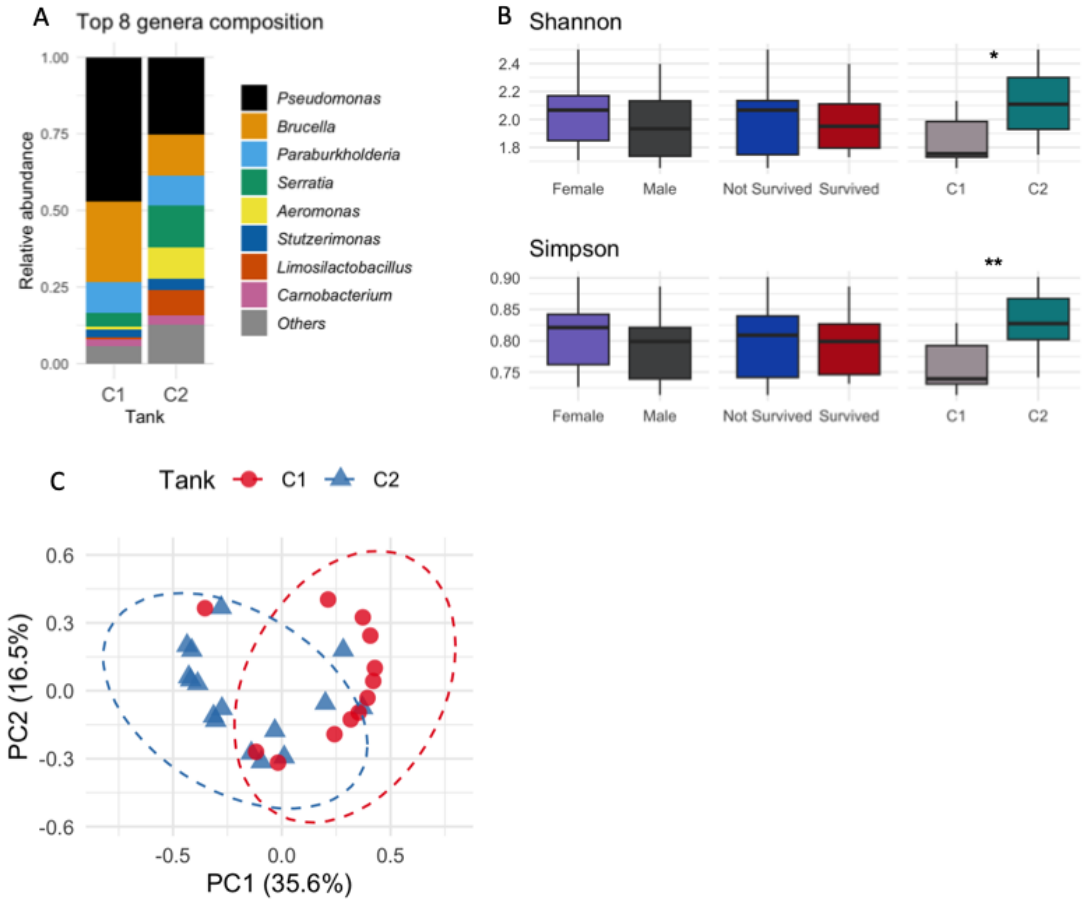
348 abundance in Tank C1.

349

350

351

352



353

354 **Figure 1. Variation in detectable microbial communities across conditions.**

355 **A. Genus-Level Microbial Composition in Tanks C1 and C2** Species-level Kraken2 counts were
356 aggregated to the genus level after removing reads assigned to *Salmo salar* and *Homo sapiens*. The
357 eight genera with the highest total counts across both tanks were retained individually, and all remaining
358 genera were grouped as "Others." Bars represent relative abundances within each tank. **B. Shannon**
359 **and Simpson alpha diversity per group.** Shannon and Simpson diversity indices calculated from
360 species-level microbial counts. Diversity values were compared across sex, survival status, and tank
361 using Wilcoxon rank-sum tests with Benjamini–Hochberg correction. An asterisk (*) indicates $q < 0.05$,
362 and a double asterisk (**) indicates $q < 0.01$. **C. Principal coordinates analysis (PCoA) of microbial**
363 **communities based on Jaccard dissimilarity.** Each point represents one sample and is colored and
364 shaped by tank identity. Ellipses indicate 95% confidence intervals. Samples separate according to tank,
365 indicating that detectable microbial composition differed between tanks.

References

1. Mougin,J. and Joyce,A. (2022) Fish disease prevention via microbial dysbiosis‐associated biomarkers in aquaculture. *Rev. Aquac.*, 10.1111/raq.12745.

2. Ojala,T., Häkkinen,A.-E., Kankuri,E. and Kankainen,M. (2023) Current concepts, advances, and challenges in deciphering the human microbiota with metatranscriptomics. *Trends Genet.*, **39**, 686–702.

3. Jiang,Y., Xiong,X., Danska,J. and Parkinson,J. (2016) Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome*, **4**, 2.

4. Bouwmeester,M.M., Goedknegt,M.A., Poulin,R. and Thieltges,D.W. (2021) Collateral diseases: Aquaculture impacts on wildlife infections. *J. Appl. Ecol.*, **58**, 453–464.

5. Glidden,C.K., Nova,N., Kain,M.P., Lagerstrom,K.M., Skinner,E.B., Mandle,L., Sokolow,S.H., Plowright,R.K., Dirzo,R., Leo,G.A.D., *et al.* (2021) Human-mediated impacts on biodiversity and the consequences for zoonotic disease spillover. 10.22541/au.162333361.10264935/v4.

6. Massimino,L., Lamparelli,L.A., Houshyar,Y., D'Alessio,S., Peyrin-Biroulet,L., Vetrano,S., Danese,S. and Ungaro,F. (2021) The Inflammatory Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nature Computational Science*, **1**, 511–515.

7. Chen,C.-Z., Li,P., Liu,L. and Li,Z.-H. (2022) Exploring the interactions between the gut microbiome and the shifting surrounding aquatic environment in fisheries and aquaculture: A review. *Environ. Res.*, **214**, 114202.

8. Canellas,A.L.B., Costa,W.F., Freitas-Silva,J., Lopes,I.R., de Oliveira,B.F.R. and Laport,M.S. (2022) In sickness and in health: Insights into the application of omics in aquaculture settings under a microbiological perspective. *Aquaculture*, **554**, 738132.

9. Samsing,F. and Barnes,A.C. (2024) The rise of the opportunists: What are the drivers of the increase in infectious diseases caused by environmental and commensal bacteria? *Rev. Aquac.*, **16**, 1787–1797.

10. Gjerde,B., Boison,S.A., Hazlerigg,D., Ytrestøyl,T., Mørkøre,T., Jørgensen,E., Striberny,A. and Sandve,S.R. (2024) Impact of three light smoltification regimes on performance and genetic parameters of traits in Atlantic salmon. *Genetics*.

11. Grønvold,L., van Dalum,M.J., Striberny,A., Manousi,D., Ytrestøyl,T., Mørkøre,T., Boison,S., Gjerde,B., Jørgensen,E., Sandve,S.R., *et al.* (2024) Transcriptomic profiling of gill biopsies to define predictive markers for seawater survival in farmed Atlantic salmon. *J. Fish Biol.*, 10.1111/jfb.16025.

12. Chen,S., Zhou,Y., Chen,Y. and Gu,J. (2018) fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, **34**, i884–i890.

13. Stenløkk,K.S.R. (2023) Genomic structural variations as drivers of adaptation in salmonid fishes.

404 14. Vera-Ponce de León,A., Hensen,T., Hoetzinger,M., Gupta,S., Weston,B., Johnsen,S.M.,
405 Rasmussen,J.A., Clausen,C.G., Pless,L., Veríssimo,A.R.A., *et al.* (2024) Genomic and
406 functional characterization of the Atlantic salmon gut microbiome in relation to nutrition and
407 health. *Nat. Microbiol.*, **9**, 3059–3074.

408 15. Chklovski,A., Parks,D.H., Woodcroft,B.J. and Tyson,G.W. (2023) CheckM2: a rapid, scalable and
409 accurate tool for assessing microbial genome quality using machine learning. *Nat. Methods*, **20**,
410 1203–1212.

411 16. Wood,D.E., Lu,J. and Langmead,B. (2019) Improved metagenomic analysis with Kraken 2.
412 *Genome Biol.*, **20**, 257.

413 17. Lin,H. and Peddada,S.D. (2020) Analysis of compositions of microbiomes with bias correction. *Nat.*
414 *Commun.*, **11**, 3514.

415 18. McMurdie,P.J. and Holmes,S. (2013) phyloseq: an R package for reproducible interactive analysis
416 and graphics of microbiome census data. *PLoS One*, **8**, e61217.

417 19. Lin,H. and Peddada,S.D. (2024) Multigroup analysis of compositions of microbiomes with covariate
418 adjustments and repeated measures. *Nat. Methods*, **21**, 83–91.

419 20. Wagner,H. (2022) vegan: Community Ecology Package CRAN.

420 21. Wickham,H. (2009) Ggplot2: Elegant Graphics for Data Analysis 2nd ed. Springer Publishing
421 Company, Incorporated.

422 22. McMurdie,P.J. and Holmes,S. (2012) Phyloseq: a bioconductor package for handling and analysis
423 of high-throughput phylogenetic sequence data. *Pac. Symp. Biocomput.*

424 23. Hajnsdorf,E. and Kaberdin,V.R. (2018) RNA polyadenylation and its consequences in prokaryotes.
425 *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **373**, 20180166.

426 24. Mofayezi,A., Jadaliha,M., Zangeneh,F.-Z. and Khoddami,V. (2024) Poly(A) tale: From A to A; RNA
427 polyadenylation in prokaryotes and eukaryotes. *Wiley Interdiscip. Rev. RNA*, **15**, e1837.

428 25. Testerman,T., Beka,L., Reichley,S.R., King,S., Welch,T.J., Wiens,G.D. and Graf,J. (2022) A large-
429 scale, multi-year microbial community survey of a freshwater trout aquaculture facility. *FEMS*
430 *Microbiol. Ecol.*, **98**.

431 26. Clinton,M., Wyness,A.J., Martin,S.A.M., Brierley,A.S. and Ferrier,D.E.K. (2024) Association of
432 microbial community structure with gill disease in marine-stage farmed Atlantic salmon (Salmo
433 salar); a yearlong study. *BMC Vet. Res.*, **20**, 340.

434 27. Gihawi,A., Ge,Y., Lu,J., Puiu,D., Xu,A., Cooper,C.S., Brewer,D.S., Pertea,M. and Salzberg,S.L.
435 (2023) Major data analysis errors invalidate cancer microbiome findings. *MBio*, **14**, e0160723.

436 28. Gloor,G.B., Macklaim,J.M., Pawlowsky-Glahn,V. and Egozcue,J.J. (2017) Microbiome datasets are
437 compositional: And this is not optional. *Front. Microbiol.*, **8**.

438