Metatranscriptomic Insights into Microbial Dynamics Prior to Disease Onset in Atlantic Salmon Aquaculture

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], Arturo Vera Ponce De Leon[1], Marie Saitou[1]

[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

Abstract

Aquaculture plays a key role in global food security, yet disease outbreaks remain a major challenge. While host transcriptomic response have been widely studied, the role of microbial community in disease risk in aquaculture remains understudied.

Here, we conducted a metatranscriptomic analysis of gill microbiomes in Atlantic salmon (*Salmo salar*) by repurposing publicly available RNA-seq data originally generated for host transcriptomics. These poly(A)-selected RNA-seq datasets were collected prior to a disease outbreak, allowing us to examine functional microbial patterns retrospectively.

Taxonomic profiles generated by Bracken and MATAM showed strong concordance across dominant genera, confirming that host RNA-seq datasets retain sufficient microbial signal for downstream classification. In addition to tank-dependent microbial variation, our presence-absence analysis revealed survival-associated microbiome shifts. Several taxa, including *Sedimentibacter*, *Fusarium*, and *Demequina* were more observed in non-survived individuals, whereas *Micrococcaceae, Pseudomonas, and Burkholderiaceae* were enriched in survivors. were enriched in survivors. Functional gene analysis revealed distinct transcriptional responses within the same microbial taxa between survivors and non-survivors, particularly in metabolic adaptation, stress response, and host-microbe interactions. These results suggest that host-derived RNA-seq datasets, though not originally designed for microbial analysis, offer new opportunities to investigate early microbial signals and environmental factors associated with disease susceptibility in aquaculture.

Keywords: Metatranscriptomics, Aquaculture, Gill microbiome, Functional profiling, *Salmo salar*

Short Summary: This study shows that host RNA-seq data can be repurposed to reveal early microbial changes and functional shifts associated with disease outcomes in Atlantic salmon aquaculture.

Introduction

Aquaculture is a promising solution for meeting the growing global demand for food and nutrition (Little, Newton and Beveridge 2016; Boyd, McNevin and Davis 2022). Despite advancements in intensive breeding systems that have improved production efficiency, infectious diseases remain a major challenge, because of high fish densities in captive environments (Leung and Bates 2013; Assefa and Abunna 2018; Bouwmeester *et al.* 2021; Ferreira *et al.* 2021; Mougin and Joyce 2022). Disease outbreaks not only cause substantial economic losses and food waste (Stentiford *et al.* 2017) but also affect wild fish populations and surrounding ecosystems as aquaculture environments can serve as reservoirs for pathogens (Diana 2009; Colombo *et al.* 2016; Hamza *et al.* 2017; Bouwmeester *et al.* 2021; Glidden *et al.* 2021).

The frequent occurrence of infectious diseases in aquaculture has led to the widespread use of antibiotics. While this approach offers short-term efficacy, it contributes to the emergence of antibiotic-resistant bacteria, posing challenges to long-term disease management and ecosystem health (Defoirdt, Sorgeloos and Bossier 2011; Muziasari *et al.* 2016; Santos and Ramos 2018; Topp *et al.* 2018; Haro-Moreno *et al.* 2020; Lulijwa, Rupia and Alfaro 2020). Vaccination programs also have been implemented for several pathogens (Park 2009; Rodger 2016; Mkulo *et al.* 2024). However, a comprehensive understanding of pathogen diversity and microbial interactions in aquaculture remains elusive, particularly regarding the complexity and functional dynamics of the aquatic microbiome, which plays a crucial role in host health, immune regulation, and pathogen resistance (Rajeev *et al.* 2021; Canellas *et al.* 2022; Chen *et al.* 2022; Samsing and Barnes 2024).

Recent advances in high-throughput sequencing, particularly metatranscriptomics, have enabled the functional characterization of microbiomes in real-time, providing deeper insights into microbial activity

and host-microbe interactions (Jiang *et al.* 2016; Ojala *et al.* 2023). Unlike traditional microbiome analyses that rely on taxonomic profiling, metatranscriptomics allows for the identification of microbial functional signatures, potentially serving as early warning indicators before clinical symptoms emerge. Integrating this approach into aquaculture disease monitoring could facilitate a shift from reactive disease management to proactive risk assessment and control, reducing dependence on antibiotics and improving biosecurity (Jorth *et al.* 2014; Hua *et al.* 2015; Welti-Chanes *et al.* 2020; Massimino *et al.* 2021)

In the winter of 2021, a severe outbreak of infectious skin disease occurred in an Atlantic salmon (*Salmo salar*) breeding facility in Western Norway, leading to high mortality rates and revealing the need for early disease detection strategies. Current disease investigations in aquaculture are largely conducted post-outbreak, yet microbial and immunological alterations may occur before clinical symptoms appear (de Bruijn *et al.* 2018). While most transcriptomic studies in aquaculture focus on host immune responses, an alternative and complementary approach is to examine microbial community function as a predictor of disease susceptibility (Yilmaz *et al.* 2022). The gill microbiome, in particular, plays a key role in pathogen defense and host homeostasis (Morshed and Lee 2023), yet their functional activity prior to disease outbreaks remains largely unexplored.

In this study, we repurposed a publicly available RNA-seq dataset originally generated for host transcriptomic analyses of Atlantic salmon gill biopsies collected prior to a disease outbreak. At the time of sampling, fish exhibited no visible disease symptoms, allowing retrospective investigation of microbial signatures preceding disease emergence. Despite the use of poly(A)-selected libraries, we successfully retrieved both rRNA and mRNA transcripts from microbial taxa. By comparing microbial taxonomic profiles and transcriptional patterns between survivors and non-survivors, we identified taxa- and transcript-level differences that may reflect early microbial responses associated with disease susceptibility. Our findings demonstrate the potential of host-derived RNA-seq data as an underutilized resource for microbial ecological analysis in aquaculture systems.

Materials and Methods

Metatranscriptome dataset

The metatranscriptome dataset analyzed in this study was derived from Atlantic salmon gill samples and is publicly available at the European Nucleotide Archive (ENA) under the project number PRJEB47441. These data originated from 109 full-sib families of the Mowi (Mowi Genetics AS) strain, sampled in November 2019. This study did not involve direct handling of fish; instead, it retrospectively analyzed metatranscriptome data to investigate microbial dynamics prior to disease onset. A comprehensive description of the fish population is provided in Gjerde et al. (Gjerde *et al.* 2024) while details on metatranscriptome data generation are outlined in Grønvold et al. (Grønvold *et al.* 2024). The original experiment was performed according to EU regulations concerning the protection of experimental animals (Directive 2010/63/EU). Appropriate measures were taken to minimize pain and discomfort. The experiment was approved by the Norwegian Food and Safety Authority (FOTS id. number 25658) (Gjerde *et al.* 2024).

Fish Rearing and Sampling Timeline

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

In early February 2021, fish were vaccinated with Alpha Ject micro 6 (PHARMAQ®) against *Aeromonas salonicida, Listonella anguillarum O1* and *O2a, Vibrio salmonicida, Moritella viscosa, and IPN virus Sp*.

In June 2021, one week before seawater transfer, gill biopsies were collected for RNA sequencing at a time when fish showed no visible disease symptoms.

On June 23, 2021, fish were transported to the ARST seawater facility. Approximately eight weeks after transfer, signs of winter ulcers began to appear, and mortality increased throughout the autumn. Despite the removal of affected individuals during trait recordings in early November, the prevalence of winter ulcers persisted into winter. Over the following 10 months, fish were monitored for survival and growth.

Data Selection

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

Metatranscriptome data preparation

Publicly available RNA-seq data were retrieved in FastQ format from the European Nucleotide Archive (ENA, project PRJEB47441). The dataset consists of gill metatranscriptome from Atlantic salmon, with sequencing performed using a paired-end (PE) approach (2×150 bp) generating approximately 20 million reads per sample. Quality assessment of raw reads was conducted using FastQC (v0.12.0). Adapter

sequences, low-quality bases, and short reads were removed using fastp (Chen *et al.* 2018)(v0.23.4) with default parameters. Filtered reads were then mapped to the Atlantic salmon reference genome (Stenløkk 2023) (Ssal_v3.1, GCA_905237065.2) using bwa-mem2 (v2.2.1). Reads that failed to align with the host genome were extracted and retained for microbiome analysis.

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

For preprocessing, adapter trimming and quality filtering were performed using fastp (v0.23.4), followed by host genome mapping against the Atlantic salmon reference genome (Ssal_v3.1, GCA_905237065.2). Unmapped reads were retained for microbiome analysis, under the assumption that they predominantly represent microbial sequences. However, given the poly(A) selection strategy, it is possible that a proportion of these unmapped reads include artifacts from host transcripts or residual polyadenylated bacterial transcripts.

Although this potential artifact is present, it is expected to be uniformly distributed across all samples rather than selectively biased toward a specific group (e.g., survived vs. non-survived fish). Therefore, while this factor must be considered when interpreting metatranscriptomic results, the relative comparisons between survival groups remain valid, as any systematic bias introduced by poly(A) selection would be consistent across conditions.

Taxonomic Classification and Analysis

For taxonomic classification, Kraken2 (Wood, Lu and Langmead 2019) (v2.14) and Bracken (v3.0) were utilized. Kraken2 was executed using a custom-built database (GTDB_plus) (Parks *et al.* 2022). Following

Kraken2 classification, Bracken was used to estimate species taxonomic read count at different taxonomic levels. Each sample was processed iteratively using paired-end FASTQ files. The GTDB_plus database was employed for Kraken2 classification. This extensive database integrates GTDB representative genomes (Release 2.20) and NCBI RefSeq standard libraries, including viral, fungi, plant, protozoa, and human sequences for contamination screening. The database utilizes a hybrid taxonomy, with GTDB taxonomy for prokaryotes and NCBI taxonomy for other taxa. Entries with tax_id greater than 4,000,000 originate from GTDB, while lower taxonomic IDs correspond to NCBI classifications. For rRNA-based microbial community analysis, MATAM (1.6.1) was employed to assemble 16S/23S rRNA reads. Non-host sequence reads were aligned against the SILVA 138.1 database.

Taxonomic read count Analysis

Bracken output files were processed in R to obtain family-level taxonomic read count data. The data were merged with sample metadata and cleaned to remove unwanted characters and taxa absent in all samples. A threshold-based filtering approach was applied to merge low taxonomic read count taxa (<1% max taxonomic read count) into an "Other" category. Hominidae sequences were excluded. Read counts were normalized by dividing by the total read count per sample and multiplying by 200,000. This serves as a scaling method to allow comparability across samples while preserving relative taxonomic read count within each sample. Alpha diversity indices (Shannon and Simpson) were computed using the vegan package. Statistical differences in diversity indices were tested using the Kruskal-Wallis test. Beta diversity was assessed using Bray-Curtis and Jaccard dissimilarity matrices, and statistical significance was determined using PERMANOVA (Adonis test) (Tang, Chen and Alekseyenko 2016). SIMPER analysis was conducted to identify taxa contributing to beta diversity differences. Barnard's exact test was applied to binary presence-absence data to evaluate the association between taxa and fish health status.

Gene Expression Quantification

Metatranscriptome assembly was performed using MEGAHIT (v1.2.9). For transcript quantification, Salmon (v0.8.1) was used. The analysis was conducted using an index built for the dataset. Paired-end reads were processed with automatic library type detection, validation of mappings. A p-adjusted value < 0.05 was considered statistically significant for the differential expression analysis using DESeq2 (Love, Huber and Anders 2014). Functional annotation was conducted on DRAM (Shaffer *et al.* 2020) on K-base (Arkin *et al.* 2018).

FASTQ Statistics Analysis

Contig lengths were computed using a custom script. The N50 statistic was determined by sorting contig lengths in descending order and identifying the length where the cumulative sum reached half of the total length. Mean and standard error was also calculated.

The scripts used in this study are available at https://github.com/mariesaitou/paper_2025-/tree/main/metaT.

Results

Assessment of Poly(A) selection bias and validation of taxonomic classification

We obtained an average of 537,995 ± 289,590 unmapped reads from a total of 11,328,436 ± 1,190,386 reads from the 28 samples. The observed distribution of poly(A) and poly(T) tail lengths in unmapped reads indicates a selective retention of polyadenylated sequences (**Figure S1**). Shorter poly(A) tails (e.g., 4–8 nucleotides) were detected at high frequencies, while longer tails (≥12 nucleotides) were less

common but consistently present above theoretical expectations. This suggests that bacterial transcripts with short poly(A) tails were retained, aligning with the known role of bacterial polyadenylation in mRNA degradation and turnover rather than transcript stabilization.

To assess the effectiveness of microbial taxonomic assignments in the poly(A)-selected RNA-seq dataset, we first compared results from Bracken (the top-observed taxa are shown in **Table 2AB**),

We evaluated the effectiveness of microbial taxonomic assignments in the poly(A)-selected RNA-seq dataset by comparing Bracken, which classifies reads based on total transcript counts , with MATAM, which reconstructs 16S/23S rRNA reads for taxonomic profiling. MATAM identified 54 genera in total, while its coverage was limited, suggesting that a small fraction of microbial rRNA transcripts may have escaped the poly(A) selection process. The detected rRNA sequences were primarily corresponded to the most transcriptionally dominant bacterial taxa identified by Bracken (**Table 2**). **Figure S2** compares Bracken and MATAM estimates for genera ranked highest in transcript quantification. The strong correlation observed (Spearman's ρ ranging from 0.48 to 0.96; *e.g., Brevundimonas*, $\rho = 0.94$, $p = 2.4e-12$; *Sphingomonas,* $\rho = 0.96$, $p = 0$) suggests that Bracken provides reliable taxonomic resolution despite potential biases from poly(A) selection. Given Bracken's broader taxonomic coverage and its ability to classify both rRNA and mRNA-derived sequences, we opted to use it for downstream analyses. Given that bacterial transcripts were successfully detected despite poly(A) selection, we proceeded with taxonomic classification to profile microbial communities and assess their association with fish health.

Overview of Microbial Community Composition and Environmental Influence

Taxonomic classification was performed using Kraken and Bracken. At the species level, taxonomic read counts were relatively low, leading to unstable classification results. Therefore, initial quantitative analyses were conducted at the family level, followed by a deeper examination of notable taxa at the genus level. The composition of metatranscriptomic taxa exhibited individual variation among fish samples; however, when grouped by conditions, the overall community structure remained relatively stable (**Figure 2A**).

Across all samples, the top ten most transcriptionally represented families accounted for 79.4% of the total taxonomically assigned reads (**Table 2A**). Among these, Micrococcaceae was the dominant family, contributing 20.34% of the total reads, followed by f__UBA9655 (17.18%) and f__UKL13-3 (12.89%). Other notable families included Vampirovibrionaceae (7.23%) and Burkholderiaceae (6.12%), indicating their prevalence in the microbial community. Micrococcaceae includes some opportunistic pathogens that may be involved in fish infections. Their presence in aquaculture environments suggests a possible role in disease dynamics. Families f__UBA9655 and f__UKL13-3 remain insufficiently characterized, with unknown physiological functions, but they have been frequently detected in aquaculture environments, suggesting a potential ecological role. Since they are metagenome-only classifications, their functional significance might be underestimated. Burkholderiaceae comprises highly adaptable environmental bacteria with multifunctional roles. Some species exhibit pathogenic traits, while others show antimicrobial properties (Morya, Salvachúa and Thakur 2020). Their frequent presence in aquatic environments highlights their potential role in host-pathogen interactions and microbiome stability. Enterobacteriaceae includes many pathogenic species known to cause infections in humans and animals, and their presence in fish samples could indicate potential health risks (Janda and Abbott 2021). Pseudomonadaceae is associated with fish diseases, including Aeromonas infections, and is linked to water quality degradation and host immune suppression, potentially serving as an indicator of disease risk. Moraxellaceae, frequently detected in aquatic environments, may serve as an indicator of environmental pollution and water quality stress factors. These families will be further examined at the genus level in subsequent analyses.

Alpha diversity was assessed using the Shannon and Simpson indices, and their relationships with sex, tank, and survival status were examined using Kruskal-Wallis tests. The results showed that neither sex nor survival status significantly influenced Shannon diversity ($p = 0.39$ and $p = 0.39$, respectively). However, tank identity exhibited a highly significant effect on Shannon diversity ($p = 0.00049$), suggesting that environmental conditions played a dominant role in shaping microbial diversity. Similar trends were observed for the Simpson index, with tank identity again showing a significant effect ($p = 0.00041$), while sex and survival status had no significant impact ($p = 0.24$ and $p = 0.27$, respectively) (**Figure 2B**). Tank C1 exhibited higher overall diversity and a more balanced community structure compared to C2, where a smaller number of dominant taxa contributed disproportionately to the community composition. This pattern aligns with the observations in **Figure 2A**, further supporting the notion that tank-specific environmental factors shaped microbial diversity.

Despite the observed differences in microbial diversity between tanks, survival analysis suggested that these environmental differences did not directly influence mortality rates. Chi-squared tests using the total fish record (637 fish, recorded in November 2021 with all the information) for survival status across tanks and sex categories yielded non-significant results ($p = 0.51$ for tank, $p = 0.38$ for sex), indicating that while tank identity influenced microbial community composition, it did not translate into a detectable effect on survival outcomes.

To identify taxa contributing to the observed differences in diversity between tanks, SIMPER analysis highlighted a subset of taxa with the lowest adjusted p-values ($padj < 0.05$). A total of 207 families were identified as contributing to tank-specific microbial differentiation. Notably, 85 of these belonged to functionally uncharacterized metagenomic lineages, including f__UBA2385, f__UBA2392, and f__UBA2428. The prevalence of these uncharacterized taxa suggests that they may play a critical but currently unknown role in shaping microbial community structure and ecological dynamics across tanks.

Beta diversity analysis was conducted using Bray-Curtis and Jaccard distance matrices, followed by PERMANOVA tests to assess the effects of sex, tank, and survival status. The Bray-Curtis-based analysis indicated that sex had no significant effect on microbial composition ($p = 0.49$), while tank identity exerted a strong influence ($p = 0.001$), confirming the results from alpha diversity analysis. Survival status had no significant effect ($p = 0.53$).

The Jaccard distance-based analysis further supported these findings, with tank identity showing no significant impact ($p = 0.59$) and sex remaining non-significant ($p = 0.14$). However, survival status exhibited a marginally significant effect ($p = 0.025$), suggesting that while overall microbial diversity was not strongly linked to survival, specific microbial taxa might still contribute to differences between survivors and non-survivors. However, the effect size was relatively small, as indicated by the low R^2 value (0.059) from the PERMANOVA analysis.

To visualize these patterns, non-metric multidimensional scaling (NMDS) was initially attempted, but even with $k = 10$ dimensions, the stress value remained extremely high (2.3285), suggesting that the underlying data structure was too complex for effective dimensionality reduction. Instead, a principal component analysis (PCA) was performed, which revealed clustering patterns by survival status. The PCA results showed that PC1 accounted for 39.57% of the variance, while PC2 accounted for 8.3%, providing a more informative representation compared to NMDS (MDS1: 18.03%, MDS2: 9.83%) (**Figure 3A**). Nonetheless, PCA did not fully capture the variation in microbial composition, highlighting the complexity of the dataset.

Survival-Associated Taxa Based on Presence-Absence Analysis

Given that the Jaccard-based analysis indicated a significant survival effect, we hypothesized that presence-absence patterns of certain taxa, rather than their relative taxonomic read count, may be key determinants of disease outcomes. To test this, we calculated the proportion of individuals within each survival group that tested positive for each taxon and visualized the differences using histograms (**Figure 3B**). Wilcoxon rank-sum tests were conducted to assess differences in taxon presence between groups. A significant effect was observed for tank identity ($p = 4.89e-10$), reinforcing its role in shaping microbial community structure. However, sex showed no significant effect ($p = 0.077$), while survival status exhibited a strong effect ($p = 4.53e-05$), further supporting the idea that specific taxa are disproportionately associated with mortality.

To quantify the asymmetry in taxon prevalence between groups, we examined the distribution of presence-absence differences. A skewness analysis indicated a strong right-tailed distribution (skew = 11.615, $p < 2.2e-16$), confirming that taxa more commonly observed in deceased individuals were nearly twice as numerous as those enriched in survivors (698 vs. 360 taxa). This was further supported by a Kolmogorov-Smirnov test ($D = 1$, $p < 2.2e-16$), indicating a significant divergence in distribution.

To further investigate taxa associated with survival status, a presence-absence analysis was conducted using Barnard's exact test. However, traditional multiple testing correction methods, such as the Benjamini-Hochberg procedure, were overly stringent ($\pi 0$ estimate = 0.90), as the method assumes a high proportion of null hypotheses. This conservative approach led to no significant taxa remaining after adjustment. Specifically, the high $\pi 0$ estimate indicated that the majority of tests were expected to reflect true null hypotheses, causing the correction to be excessively stringent and excluding potential true positives. To address this, an Empirical Bayes approach implemented in the ashR package (Stephens 2017) was applied, refining the detection of survival-associated taxa and reducing the number of candidate taxa from 32 to 16 for further, genus-level downstream analysis (**Table 3A**).

Among the taxa with survival associations, several groups emerged, including a substantial proportion of uncharacterized lineages (e.g., f__UBA4486, f__UBA9219). Additionally, some well-characterized taxa were identified, including Aerococcaceae and Sedimentibacteraceae, a potential indicator of water quality

conditions (Fan *et al.* 2023). Nectriaceae, a fungal family associated with infection risks (Ulaszewski *et al.* 2025), and Peduoviridae, a bacterial viruses family whose genome incorporates into the bacterial genome (Wang, Xie and Huang 2024), potentially influencing microbiome stability (**Figure 4**).

Building on the family-level findings, we next examined microbial composition at the genus level to further characterize taxa with potential functional or ecological relevance. At the genus level, some highly represented taxa were identified. *Arthrobacter_D* (Micrococcaceae) was the dominant genus, accounting for 19.87% of the total reads (**Table 2B**). This was followed by several uncharacterized genera, including *g__UKL13-3, g__UBA9655*, and another unidentified *g__UBA9655*, which remain insufficiently understood but were consistently present at high taxonomic read count. Among the well-characterized genera, some stood out due to their potential roles in fish health and aquaculture environments. *Pseudomonas_E* (Pseudomonadaceae), for instance, is known to be associated with waterborne infections, making it a potential risk factor for fish disease (Rahman *et al.* 2024). *Serratia* (Enterobacteriaceae), an opportunistic pathogen in humans and animals, was also notable due to its presence in the aquatic environment (Mahlen 2011). Similarly, *Escherichia* (Enterobacteriaceae), while often a commensal gut bacterium, includes pathogenic strains such as *E. coli*, which could potentially impact fish health (Mumbo *et al.* 2023).

To further investigate the microbial differences between tanks, Mann-Whitney U tests were performed at the genus level, comparing Tank C1 and C2. Genera that showed significant differences (adjusted $p < 0.05$ using BH correction) included taxa associated with environmental and metabolic functions (**Table 4**). Methylophilaceae (*CABHOJ01*) was one such genus, known for its role in one-carbon metabolism, particularly in methanol or methylated compound cycling (Taubert *et al.* 2019). Ktedonobacteraceae (*JALYTT01)*, generally a soil-associated bacterium (Gschwend *et al.* 2022), was also detected, suggesting a possible role in organic matter degradation within the aquatic environment. Another particularly interesting genus was Burkholderiaceae_A (*Kinetoplastibacterium*), which can be found as an endosymbiont of protozoa (Alves *et al.* 2013). The detection of this genus in fish-associated microbial

communities raises interesting questions about potential host-protozoan interactions and whether certain protozoa might be influencing microbial community structures in the tanks. Among the host-associated taxa, Microbacteriaceae (*Microbacterium*) is able to thrive in both environmental and host-associated niches, potentially playing a role in organic matter processing or host-microbe interactions (Wippel *et al.* 2021). Acetobacteraceae (*BOG-930*), generally associated with wine and vinegar fermentation, was also present (Komagata, Iino and Yamada 2014), which might indicate shifts in fermentation-based metabolic processes or pH fluctuations in the tank environment.

Next, genera associated with survival outcomes were examined. Barnard's exact test was applied, but due to the stringent nature of multiple testing corrections, no taxa remained statistically significant after Benjamini-Hochberg adjustment. However, using nominal *p*-values, several genera were observed to differ between survivors and non-survivors. *Demequina*, which is often identified in aquaculture species (Pedrosa-Gerasmio *et al.* 2020), was more detected in the non-surviving group, suggesting that its presence may be linked to disease susceptibility. Similarly, *Fusarium*, a fungal genus that includes opportunistic pathogens, was more frequently observed in non-surviving fish, indicating its potential involvement in disease progression or secondary infections. *Sedimentibacter*, an anaerobic bacterium associated with organic matter degradation, was also enriched in non-surviving fish, which could reflect differences in microbial community structure that associated with host health deterioration (Huang *et al.* 2016) (**Table 3B, Figure 4**).

Transcriptional Signals Reveal Functional Differences Between Survival Groups

Sequence reads from all samples were merged, and sequences corresponding to vertebrates (Atlantic salmon and human) were removed before contig assembly. The assembled contigs had the following characteristics: a total of 16,037 contigs with 7,255,640 bases, an N50 value of 434, and a mean contig length of 452.43 (SE= 4.34) base pairs.

In line with the earlier taxonomic analysis, standard differential gene expression analysis using DESeq2 did not identify any genes that survived Benjamini-Hochberg (BH) correction for significance between the survived and non-survived groups. This is likely due to the stringent multiple testing threshold for such a dataset. To address this, we reanalyzed the data by converting gene presence/absence into a binary format and applied Barnard's test. This approach identified 31 differentially observed genes with nominal $P < 0.05$, which were functionally annotated using DRAM (especially Pfam returned the best comprehensive annotation to this dataset). To determine which microorganisms were responsible for expressing these genes, we performed a BLAST search, revealing that *Achromobacter, Escherichia coli, Pseudomonas, Stenotrophomonas,* and *Serratia* exhibited highly polarized expression patterns across the survived and non-survived groups, where a transcript was frequently absent in one group while being widely present in the other. As previously observed, these bacterial taxa were commonly observed across all groups (**Table 2B**).

Figure 5 illustrates transcript read counts (TPM) for a subset of differentially observed microbial genes, while the complete list of genes is provided in **Table 5.** In *Achromobacter,* the non-survived group exhibited increased expression of genes related to sugar transporters, aldehyde dehydrogenase, and bacterial extracellular solute-binding proteins, whereas the survived group showed increased expression of genes associated with *TonB-dependent receptors, AcrB/AcrD/AcrF efflux systems*, and proline dehydrogenase, suggesting a shift in metabolic strategies. Similarly, in *Escherichia coli*, the non-survived group showed an upregulation of peroxidase genes, indicative of a potential response to oxidative stress, whereas in the survived group, genes encoding *phage minor tail protein U* were more abundant, implying active bacteriophage interactions. For Pseudomonas, the non-survived group showed increased expression of DNA topoisomerase, a gene involved in DNA replication and repair, while the survived group exhibited higher expression of RadC-like JAB domain, another gene potentially linked to DNA repair, though its precise functional role remains unclear. In *Stenotrophomonas*, the non-survived group expressed genes belonging to the POT family (peptide transporters), while the survived group exhibited

increased expression of transketolase and pyrimidine binding proteins, suggesting a potential shift in metabolic preferences (Patil *et al.* 2021). For Serratia, the survived group displayed increased expression of multiple genes, including *SecY (membrane protein transport), His Kinase A domain (signal transduction), Flavodoxin-like proteins (redox reactions), and Holliday junction DNA helicase (DNA repair and recombination mechanisms)*, indicating possible functional advantages under the fish-survival-associated conditions (Iguchi *et al.* 2014).

Notably, *Achromobacter phage* was more frequently observed in the survived group, suggesting a potential association between phage activity and bacterial dynamics under different host conditions (Hoyle *et al.* 2018). These findings indicate that bacterial communities exhibit distinct transcriptional responses depending on host status, with potential implications for disease progression, microbial adaptation, and host-microbe interactions.

Discussion

Metatranscriptome as a Tool for Microbial Ecology and Disease Risk Assessment in Aquaculture

This study conducted a metatranscriptomic analysis using RNA-seq data collected from Atlantic salmon gills before a disease outbreak in aquaculture. The goal was to determine whether microbial transcriptional profiles could provide insights into environmental conditions and potential health risks, as well as to understand the microbial dynamics before disease emergence. Through this retrospective metatranscriptomic analysis, we observed microbial changes before disease outbreaks in aquaculture. By leveraging RNA-seq data collected before clinical disease emergence, microbial shifts associated with survival outcomes were detected. In particular, presence-absence analysis revealed taxa that were associated with mortality risk, suggesting that community shifts are not solely driven by statistical changes in relative taxonomic read count. The strong right-tailed distribution of presence-absence differences, with taxa more frequently found in non-surviving fish nearly doubling those enriched in survivors, suggests the potential contribution of specific microbial members to disease susceptibility.

Furthermore, numerous characterized bacterial taxa (e.g., f__UBA9655, f__UKL13-3) were detected, emphasizing the need for further biological investigation of these insufficiently understood microbial lineages(Zamkovaya *et al.* 2021). While tracking individual microbial dynamics remains complex, our findings suggest that monitoring functionally relevant taxa could provide insights into disease susceptibility and aquaculture ecosystem(Leray *et al.* 2021). In the following sections, we discuss these findings, focusing on the relationship between microbial composition and survival outcomes, functional shifts in microbial transcriptional activity, and the role of environmental factors in shaping microbial diversity.

Our study demonstrates that publicly available poly(A)-selected RNA-seq datasets, originally generated for host transcriptomics, can be repurposed for microbial metatranscriptomic analysis. While poly(A) selection introduces constraints, most notably the exclusion of rRNA-derived taxonomic markers, it also enables the capture of actively transcribed microbial mRNA, providing functional insights that complement traditional 16S rRNA-based profiling. The strong correlation observed between Bracken-based abundance estimates and MATAM-based rRNA profiles suggests that despite the absence of rRNA, meaningful taxonomic classification can still be achieved when analyzing microbial mRNA.

One important consideration is that poly(A) selection may differentially retain microbial transcripts depending on the extent of polyadenylation (Hajnsdorf and Kaberdin 2018). While some bacterial transcripts are naturally polyadenylated as part of RNA degradation pathways, others may be underrepresented due to the lack of poly(A) tails(Mofayezi *et al.* 2024). This potential bias should be acknowledged when interpreting taxonomic and functional profiles derived from poly(A)-selected datasets. However, given that any such bias is likely to be systematic across all samples, comparative analyses, such as those between surviving and non-surviving fish, remain valid.

One important limitation is the lack of 16S rRNA data, which makes it difficult to directly confirm taxonomic identities using standard phylogenetic approaches. Although this does not prevent meaningful microbial profiling, it does require careful attention to the methods used. Future studies aiming to maximize

microbial resolution in aquaculture metatranscriptomics may benefit from hybrid approaches that integrate poly(A)-selected RNA-seq with targeted rRNA sequencing or metagenomic methods.

Despite these limitations, our findings highlight the potential of repurposing host RNA-seq datasets for microbial analysis, particularly in retrospective studies where dedicated metatranscriptomic data may not be available. Given the increasing availability of RNA-seq datasets across diverse aquaculture settings, further exploration of poly(A)-selected metatranscriptomics could provide valuable insights into microbial dynamics and their role in water environment and disease susceptibility.

Environmental Influence on Microbial Community Structure

An unexpected key finding of this study was that microbial diversity and community composition varied between tanks, despite the absence of intentional environmental modifications. This raises a question: what factors might have contributed to the observed differences? One possibility is that unmonitored environmental variations played a role in shaping microbial community structure. In closed aquaculture systems, small-scale differences in water flow, temperature microgradients, microbial seeding effects, or organic matter accumulation can drive shifts in microbial communities (Testerman *et al.* 2022). Previous research has shown that biofilm development on tank surfaces and infrastructure can serve as reservoirs for microbial colonization, influencing the microbial load in the surrounding water and subsequently affecting the fish microbiome (Dang and Lovell 2016; Dahle *et al.* 2023)(Dahle *et al.* 2020) (Dang and Lovell 2016; Dahle *et al.* 2023). Differences in microbial succession during early rearing stages, even if tanks started with the same water source, could have resulted in distinct community structures over time. While it was impossible to pinpoint specific environmental drivers using current datasets, the observed tank-dependent microbial differentiation suggests that environmental heterogeneity, even in controlled aquaculture settings, plays a substantial role in shaping microbial community structure. This highlights the need for longitudinal environmental monitoring, including water physicochemical properties and microbial colonization dynamics, to fully understand the mechanisms underlying these patterns. Also, despite these differences, no direct association between tank identity and fish survival rates was

detected. This suggests that the compositional shifts in microbial communities did not translate into differential mortality rates in this case or that the key drivers of mortality were independent of the observed inter-tank variations.

Taxa and Transcripts Associated With Host Disease Outcome

Beyond environmental influences, our results suggest that the presence or absence of specific microbial taxa was associated with survival outcomes. The enrichment of opportunistic pathogens such as *Pseudomonas* and *Aeromonas* in non-survived fish raises concerns about their potential role in disease progression. These bacteria are known to exploit compromised immune defenses, which may indicate that their increased prevalence reflects a state of physiological stress or immunosuppression in affected fish. Fish wounds serve as entry points for such opportunistic pathogens, facilitating their colonization and potentially exacerbating disease progression (Joseph *et al.* 1979; Hal and El-Barbary 2020; Wang *et al.* 2022). Conversely, taxa such as Burkholderiaceae and Achromobacter-associated phages were more frequently detected in survivors. Burkholderiaceae comprises metabolically versatile bacteria that have been implicated in both pathogenic and beneficial interactions, raising the possibility that some members of this family may contribute to microbial stability or even exert protective effects against pathogens (Carrión *et al.* 2018). The presence of bacteriophages in survivors suggests that phage-mediated microbial regulation may influence disease resistance by modulating bacterial populations within the host-associated microbiome(Ranveer *et al.* 2024). While these associations provide intriguing insights, it remains unclear whether these taxa actively influenced disease outcomes or merely reflected underlying health conditions. Further experimental infection models and functional assays will be necessary to better understand the role of specific microbial taxa.

Future Directions and Ecological Implications

While this study provides valuable insights, several methodological challenges must be considered when applying metatranscriptomics to aquaculture microbiome research. One of the primary limitations is the reliance on host RNA-seq datasets that have undergone rRNA depletion, which removes 16S rRNA, the conventional marker for microbial taxonomic classification. Although previous studies have demonstrated that mRNA-based taxonomic profiling can approximate rRNA-based methods (Stewart, Ottesen and DeLong 2010; Ranjan *et al.* 2021; Zhang *et al.* 2021), differences in transcript taxonomic read count and turnover rates may introduce biases in microbial community representation. To solidify this limitation, future follow-up studies could integrate rRNA-based sequencing with metatranscriptomics to provide a more comprehensive assessment of microbial community dynamics.

Though these sequences were excluded from downstream microbial analysis, we observed major histocompatibility complex (MHC) sequences that were absent in the reference genome but reported in NCBI. Their presence suggests that conventional reference genome-based mapping may fail to detect certain variants, particularly in the MHC region, due to its extreme diversity. This issue is widely recognized across vertebrates, including humans and other model organisms, where highly polymorphic immune-related loci often escape detection in standard genome-wide analyses (Okada *et al.* 2015; He, Minias and Dunn 2021; Luo *et al.* 2021; Heimeier *et al.* 2024). Parallel selection on MHC regions has been observed in aquaculture populations of Atlantic salmon, including this Western Norway aquaculture population (Buso *et al.* 2024), suggesting that such variation may be relevant to disease resistance. Although investigating host genomic variation is beyond the scope of this study, it is likely that additional MHC diversity exists within this population. Future studies focusing on host-pathogen interactions may benefit from targeted sequencing approaches to fully capture non-reference MHC alleles and their potential role in host immune response.

While we identified microbial signatures associated with survival outcomes, it remains unclear whether these shifts occurred as a direct cause of disease susceptibility or as a secondary consequence of host

physiology. To address this, future research should prioritize longitudinal studies that track microbial dynamics over time, from early life stages through disease outbreaks, to identify robust microbial biomarkers for disease prediction. Experimental validation through controlled infection trials also could provide mechanistic insights into the role of specific microbial taxa in disease emergence and transmission.

Tables and Figures

Table 1. Summary of individual-level metadata used in this study.

This table lists sex, tank identity, survival outcome, and skin lesion status for each fish included in the analysis. It also includes raw and unmapped RNA-seq read counts used for downstream metatranscriptomic profiling. The skin condition column indicates whether the fish exhibited visible wounds at the time of sampling. The Unmapped column represents the number of unmapped RNA-seq reads, which potentially contain non-host microbial sequences, while the Raw column indicates the total number of raw sequencing reads per sample. Samples were selected to ensure representation across key biological variables while maximizing available high-quality metatranscriptome data.

				Table1 Sample info				
ID	ENA_ID	Sex	Tank	Status	Skin	Mapped_Reads	Unmapped_Reads	Ratio
ns_1	Plate22H2	Female	C1	not survived	wounds	21,843,548	92,009	0.00421
ns_10	Plate16F11	Female	C2	not survived	wounds	25,731,837	74,096	0.00288
ns_11	Plate18H4	Male	C2	not survived	wounds	26,733,388	51,610	0.00193
ns_12	Plate20D7	Male	C2	not survived	wounds	20,369,628	254,381	0.01249
ns_13	Plate19H1	Female	C2	not survived	wounds	26,574,632	43,839	0.00165
ns_2	Plate22B6	Female	C1	not survived	wounds	20,556,179	92,061	0.00448
ns_3	Plate24C8	Female	C1	not survived	wounds	25,453,351	805,946	0.03166
ns_4	Plate25A11	Male	C1	not survived	wounds	21,969,245	951,660	0.04332
ns_5	Plate19C6	Female	C2	not survived	wounds	21,485,507	387,453	0.01803
ns_6	Plate20D6	Male	C2	not survived	wounds	31,966,933	43,896	0.00137
ns_7	Plate21C2	Male	C2	not survived	wounds	24,339,317	130,984	0.00538
ns_8	Plate17A12	Male	C2	not survived	wounds	28,818,781	64,005	0.00222
ns_9	Plate17F12	Female	C2	not survived	wounds	30,032,409	97,111	0.00323
s_1	Plate21C10	Female	C1	survived	no-wounds	23,609,879	36,054	0.00153
s_10	Plate20H3	Male	C2	survived	no-wounds	21,128,196	104,443	0.00494
s_11	Plate22D4	Male	C1	survived	no-wounds	25,073,829	144,122	0.00575
s_12	Plate21E10	Male	C1	survived	no-wounds	22,723,648	200,530	0.00882
s_13	Plate16D5	Female	C2	survived	no-wounds	23,069,585	66,850	0.00290
s_14	Plate18E4	Female	C2	survived	no-wounds	22,873,120	46,668	0.00204
s_15	Plate17H5	Male	C2	survived	no-wounds	27,028,680	58,394	0.00216
s_2	Plate21E5	Female	C1	survived	no-wounds	28,229,183	75,661	0.00268
s_3	Plate24G6	Male	C1	survived	no-wounds	25,547,504	242,683	0.00950
s_4	Plate22A10	Male	C1	survived	no-wounds	26,715,029	128,125	0.00480
s_5	Plate18E12	Female	C2	survived	no-wounds	23,150,040	84,397	0.00365
s_6	Plate19E12	Female	C2	survived	no-wounds	26,662,780	115,416	0.00433
s_7	Plate25H1	Female	C1	survived	no-wounds	27,517,530	500,383	0.01818
s_8	Plate19F2	Male	C2	survived	no-wounds	24,271,043	280,739	0.01157
s_9	Plate24C3	Female	C1	survived	no-wounds	25,564,010	106,096	0.00415

Table 2. Top transcriptionally represented microbial taxa in Atlantic salmon gill metatranscriptomes. (A) Top 20 most transcriptionally abundant microbial families across all 28 samples. (B) Top 20 most transcriptionally abundant microbial genera across all 28 samples. Taxonomic abundance was estimated using Bracken based on Kraken2 assignments.

Table 2A

Family	Read Percentage
Micrococcaceae	20.34
f__UBA9655	17.18
f__UKL13-3	12.89
Vampirovibrionaceae	7.23
Burkholderiaceae	6.12
Caulobacteraceae	3.87
Enterobacteriaceae	3.59
Cellulosilyticaceae	3.33
Pseudomonadaceae	2.63
Moraxellaceae	2.23
Xanthomonadaceae	1.56
Lachnospiraceae	1.49
DSM-18226	1.12
Sphingomonadaceae	1.05
Xanthobacteraceae	0.87
Rhodocyclaceae	0.70
Nostocaceae	0.65
Rhizobiaceae	0.63
Beijerinckiaceae	0.58
Nitrincolaceae	0.48

Table 2B

Genus	Family	Read Percentage
Arthrobacter_D	Micrococcaceae	19.9
CAISCU01	f__UKL13-3	13.0
JACPNY01	f__UBA9655	8.8
g__UBA9655	f__UBA9655	8.2
Vampirovibrio	Vampirovibrionaceae	7.3
Brevundimonas	Caulobacteraceae	3.6
Epulopiscium_A	Cellulosilyticaceae	3.4
Pseudomonas_E	Pseudomonadaceae	2.3
Acinetobacter	Moraxellaceae	2.1
Serratia	Enterobacteriaceae	2.0
Comamonas	Burkholderiaceae	1.8
Achromobacter	Burkholderiaceae	1.7
Stenotrophomonas	Xanthomonadaceae	1.4
Escherichia	Enterobacteriaceae	1.2
Neobacillus	DSM-18226	1.1
Wujia	Lachnospiraceae	1.1
Sphingomonas	Sphingomonadaceae	0.8
Nostoc_B	Nostocaceae	0.6
Agrobacterium	Rhizobiaceae	0.5
Methylobacterium	Beijerinckiaceae	0.5

Table 3. Microbial genera showing differences in relative transcript representation between Tank C1 and Tank C2. Genus-level comparisons were conducted using Mann–Whitney U tests. P-values were adjusted with the Benjamini-Hochberg method. Only genera with adjusted $p < 0.05$ are shown.

		Table 3			
Genus	**p_value**	**p_adjusted**	**Family**	**Relative_reads_TankC1**	**Relative_reads_TankC2**
AGPOR5	0.0007304	0.0246	f__WGA-3G	4	0
CABHOJ01	0.00066066	0.0246	Methylophilaceae	6	0
CANMOZ01	0.00051346	0.0246	DSM-6294	11	2
Caedimonas	0.00068318	0.0246	Caedimonadaceae	40	5
JACDDX01	0.00063682	0.0246	GWC2-71-9	53	5
JACPNY01	0.00013181	0.0246	f__UBA9655	74	5
JALYTT01	0.00026971	0.0246	Ktedonobacteraceae	5	0
Kinetoplastibacterium	0.00018121	0.0246	Burkholderiaceae_A	12198	21481
Microbacterium	0.00027474	0.0246	Microbacteriaceae	2	0
Sporocytophaga	0.00050941	0.0246	Cytophagaceae	20	2
UBA1952	0.00038959	0.0246	Chitinophagaceae	1091	143
g__JASGCL01	0.00056538	0.0246	f__JASGCL01	17	3
g__UBA9655	0.00054034	0.0246	f__UBA9655	107	5
g__JAOIVE01	0.0011377	0.0357	f__JAOIVE01	121	14
Bin55	0.00128023	0.0377	Azotimanducaceae	75	6
BOG-930	0.00152158	0.0422	Acetobacteraceae	11292	20107

Table 4. Microbial taxa associated with survival status in Atlantic salmon.

Taxa were identified based on presence-absence analysis using Barnard's exact test. Only taxa with nominal $p < 0.05$ are shown. Includes both characterized and unclassified microbial lineages.

Table 4A

Taxon_name_family		Positive_Not Survived	Positive_Survived	Negative_Not Survived	Negative_Survived	Rate_not_survived	Rate_survived	Difference	Barnard_p	lfdr	svalue	PosteriorMean
Aerococcaceae		8	0	5	15	0.62	0	0.62	0.000	0.023	0.007	0.609
Alcanivoracaceae		6	0	7	15	0.46	0	0.46	0.004	0.037	0.008	0.560
Demequinaceae		1	7	12	8	0.08	0.47	-0.39	0.005	0.038	0.008	0.556
Gorgyraeaceae		9	3	4	12	0.69	0.2	0.49	0.005	0.039	0.008	0.554
Kineococcaceae		8	3	5	12	0.62	0.2	0.42	0.007	0.041	0.009	0.548
Nectriaceae		8	2	5	13	0.62	0.13	0.48	0.007	0.041	0.009	0.548
Peduoviridae		12	7	1	8	0.92	0.47	0.46	0.008	0.043	0.009	0.543
Sedimentibacteraceae		8	15	5	0	0.62	1	-0.38	0.012	0.047	0.010	0.531
f__CAG-698		10	5	3	10	0.77	0.33	0.44	0.013	0.047	0.010	0.529
f__JAAEKA01		6	13	7	2	0.46	0.87	-0.41	0.016	0.049	0.011	0.525
f__JABDET01		7	2	6	13	0.54	0.13	0.41	0.016	0.049	0.011	0.525
f__JAJDZL01		7	2	6	13	0.54	0.13	0.41	0.016	0.049	0.011	0.525
f__JAMCZE01		1	6	12	9	0.08	0.4	-0.32	0.016	0.049	0.012	0.524
f__SZAS-18		4	11	9	4	0.31	0.73	-0.43	0.016	0.049	0.012	0.524
f__UBA4486		9	4	4	11	0.69	0.27	0.43	0.016	0.049	0.013	0.524
f__UBA9219		9	5	4	10	0.69	0.33	0.36	0.016	0.049	0.013	0.524

Table 4B

Taxon_name_family	Taxon_name_genus	Positive_Not Survived	Positive_Survived	Negative_Not Survived	Negative_Survived	Rate_not_survived	Rate_survived	Difference	Barnard_p
Sedimentibacteraceae	*Sedimentibacter*	6	1	7	14	0.46	0.07	0.39	0.01
Demequinaceae	*Demequina*	10	5	3	10	0.77	0.33	0.44	0.01
f__JAAEKA01	*g__JAAEKA01*	0	5	13	10	0.00	0.33	-0.33	0.01
f__JAJDZL01	*g__JAJDZL01*	7	2	6	13	0.54	0.13	0.41	0.01
Nectriaceae	*Fusarium*	9	4	4	11	0.69	0.27	0.43	0.02
Peduoviridae	*uncharacterized*	5	12	8	3	0.38	0.8	-0.42	0.02
Alcanivoracaceae	*Isoalcanivorax*	1	6	12	9	0.08	0.4	-0.32	0.03
Aerococcaceae	*WM01*	3	0	10	15	0.23	0	0.23	0.04
f__UBA4486	*uncharacterized*	8	4	5	11	0.62	0.27	0.35	0.05

Table 5 Differentially observed microbial transcripts between surviving and non-surviving Atlantic salmon. Transcripts were identified using binary presence-absence calls and Barnard's test, with gene annotations provided by DRAM and Pfam. Taxonomic origin was determined using BLAST alignment. Only transcripts with nominal $p < 0.05$ are included.

Sequence	Not Survived_positive	Not Survived_negative	Survived_positive	Survived_negative	Taxa	pfam_hits_description	pfam_ID
k141_2253_1	5	8	1	14	*Achromobacter*	Sugar transporter	PF00083.27
k141_633_1	6	7	1	14	*Achromobacter*	Aldehyde dehydrogenase	PF00171.25
k141_1992_1	6	7	1	14	*Achromobacter*	Bacterial extracellular solute-binding	PF00496.25
k141_13760_1	0	13	8	7	*Achromobacter*	TonB-dependent Receptor	PF07715.18
k141_1120_1	0	13	5	10	*Achromobacter*	AcrB/AcrD/AcrF	PF00873.22
k141_4255_2	0	13	6	9	*Achromobacter*	Proline dehydrogenase	PF01619.21
k141_2562_1	5	8	1	14	*Escherichia coli*	Peroxidase	PF00141.26
k141_3771_1	1	12	7	8	*Escherichia coli*	Phage minor tail protein U	PF06141.14
k141_6116_1	7	6	3	12	*Pseudomonas*	DNA topoisomerase	PF01131.23
k141_10189_1	0	13	6	9	*Pseudomonas*	RadC-like JAB domain	PF04002.18
k141_14202_1	5	8	0	15	*Stenotrophomonas*	POT family	PF00854.24
k141_5394_1	0	13	5	10	*Stenotrophomonas*	Transketolase, pyrimidine binding	PF02779.27
k141_14429_1	0	13	5	10	*Serratia*	SecY	PF00344.23
k141_1963_1	0	13	5	10	*Serratia*	His Kinase A domain	PF00512.28
k141_10334_1	0	13	6	9	*Serratia*	Flavodoxin like	PF02525.20
k141_129_1	0	13	5	10	*Serratia*	Holliday junction DNA helicase	PF05496.15
k141_10700_1	1	12	7	8	*Achromobacter*	CobW/HypB/UreG, nucleotide-binding domain	PF02492.22
k141_12698_1	0	13	5	10	*Achromobacter*	FecR protein	PF04773.16
k141_15084_1	1	12	6	9	*Achromobacter*	AcrB/AcrD/AcrF family	PF00873.22
k141_15128_1	1	12	7	8	*Achromobacter*	Phosphate acetyl/butaryl transferase	PF01515.22
k141_3419_2	4	9	10	5	*Achromobacter*	Oxidoreductase FAD-binding domain	PF00970.27
k141_362_2	5	8	11	4	*Achromobacter*	ABC transporter	PF00005.30
k141_10013_29	3	10	9	6	*Achromobacter phage*	3'-5' exoribonuclease Rv2179c-like domain	PF16473.8
k141_13338_1	0	13	6	9	*Escherichia coli*	UDP-glucose/GDP-mannose dehydrogenase family, central domain	PF00984.22
k141_13126_1	0	13	5	10	*Escherichia coli*	Protein of unknown function	PF13116.9
k141_8152_2	4	9	12	3	*Escherichia coli*	IS1 transposase	PF03400.16
k141_15759_1	1	12	6	9	*Pseudomonas*	LysR substrate binding domain	PF03466.23
k141_265_1	0	13	5	10	*Pseudomonas*	Monomeric isocitrate dehydrogenase	PF03971.17
k141_4916_2	5	8	1	14	*Pseudomonas*	Transmembrane secretion effector	PF05977.16
k141_12843_1	5	8	1	14	*Serratia*	AAA lid domain	PF17871.4
k141_5692_1	0	13	5	10	*Serratia*	FCD domain	PF07729.15

Figure 1 Disease representation in Atlantic Salmon A. A Non-survived fish with a severe ulcerative lesion on the lateral body surface (highlighted in yellow) B. A survived fish with no visible external lesions.

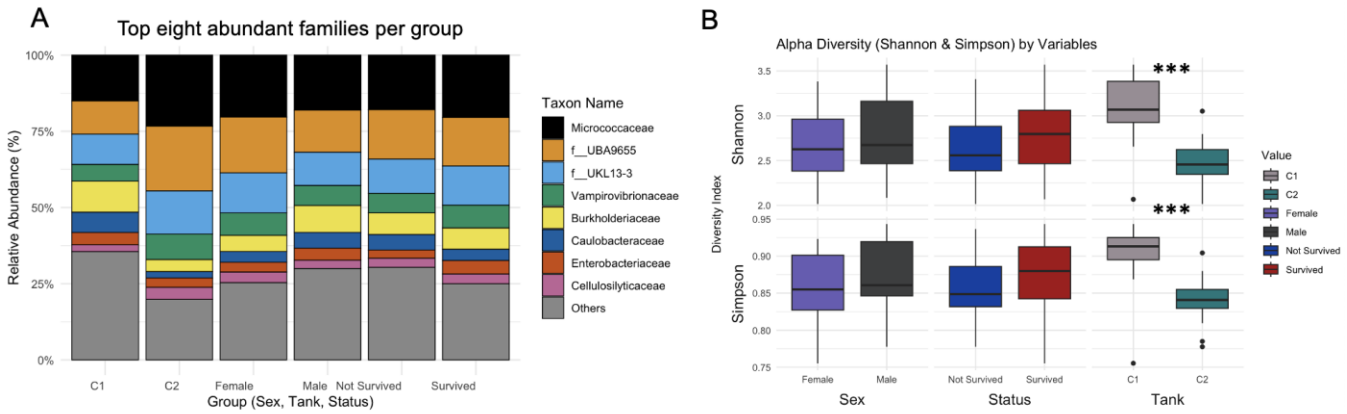

Figure 2: Taxonomic Composition and Alpha Diversity by Tank, Sex, and Survival Status

(A) The relative abundance (taxonomic read count) of microbial families aggregated by tank (C1, C2), sex (female, male), and survival status (not survived, survived). (B) Alpha Diversity Boxplots: Shannon (top) and Simpson (bottom) diversity indices comparing microbial diversity across sex, survival status, and tank conditions.

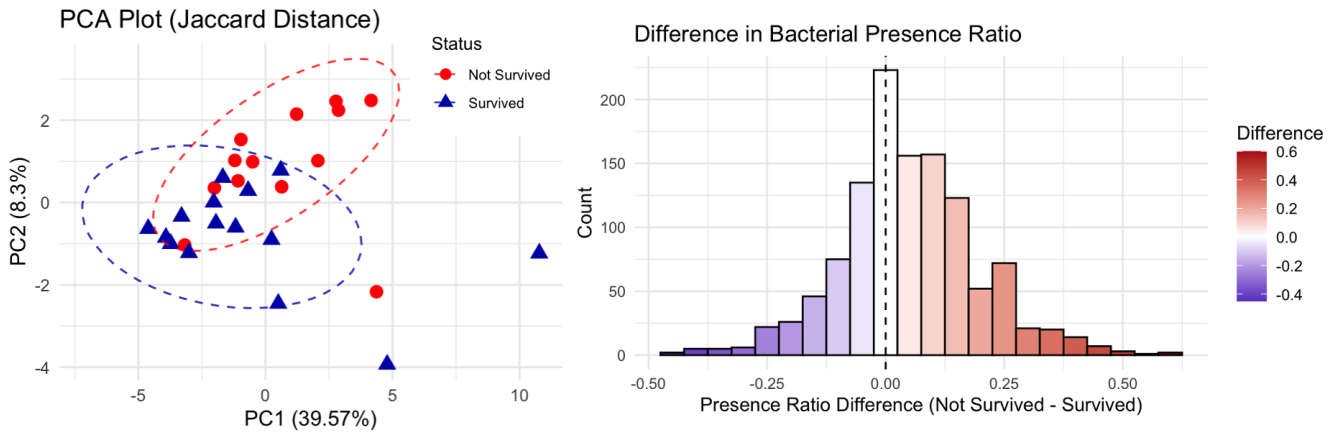

Figure 3 A Principal Component Analysis (PCA) of Microbial Community Structure

This figure presents a PCA based on Jaccard distance, colored by survival status. PC1 (39.57% variance) and PC2 (8.3% variance) capture the major differences in microbial composition.

histogram of Presence-Absence Differences Between Surviving and Non-Surviving Groups

This histogram visualizes the distribution of microbial taxa based on their presence or absence in surviving vs. non-surviving fish. B. A histogram. X-axis: Difference in presence proportion between the two groups (negative values indicate taxa more common in survivors, positive values indicate taxa more common in non-survivors). Y-axis: Number of taxa.

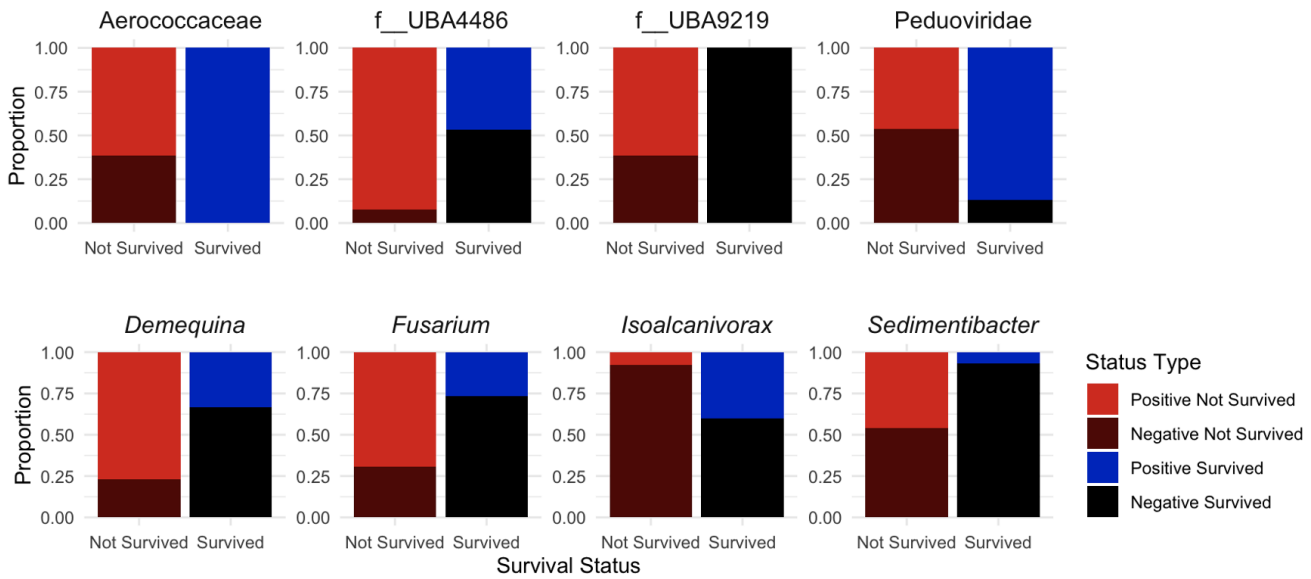

Figure 4: Proportion of fish with and without detection of taxa that differed between survivors and non-survivors

This figure displays the relative proportions of microbial genera and family significantly associated with survival outcomes in Atlantic salmon. Each bar represents the presence (blue/red) or absence (black/dark brown) of specific genera in surviving and non-surviving fish.

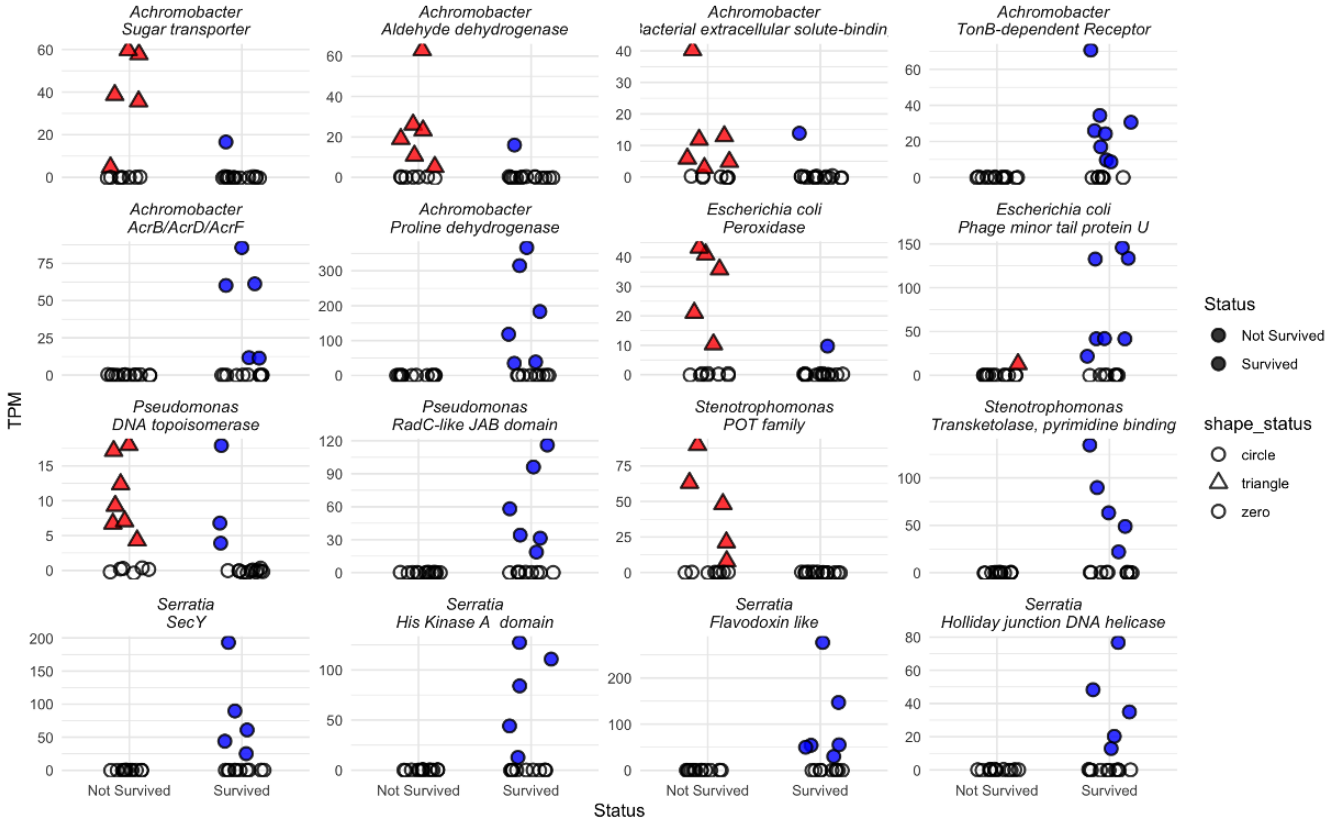

Figure 5: TPM Distribution of Differentially Expressed Genes by Survival Status

This figure illustrates the TPM distributions of differentially observed genes in microbial taxa between surviving and non-surviving Atlantic salmon.X-axis: Survival status (Not Survived vs. Survived). Y-axis: Read count for each gene. Shapes indicate gene expression levels: Triangles: Genes with elevated expression in non-surviving fish. Circles: Genes with elevated expression in surviving fish. Open circles: Zero read count (not detected).

Author contributions (CRediT)

Akira Harding: Conceptualization, Validation, Formal analysis, Investigation

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

Domniki Manousi: Data curation

Jun Soung Kwak: Conceptualization, Methodology

David Hazlerigg: Resources, Funding acquisition

Turid Mørkøre: Resources, Visualization

Arturo Vera Ponce De Leon: Conceptualization, Methodology, Supervision

Lars-Gustav Snipen: Methodology, Supervision

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

Conflicts of interest

The authors declare no competing interests.

Funding information

This study was supported by the following funding bodies: The Research Council of Norway (project numbers: 325874). The Norwegian Seafood Research Fund (FHF), through the SynchroSmolt project: “Smolt production protocols and breeding strategies for synchronized smoltification”. The EU Erasmus+ Programme. BIOVIT, Norwegian University of Life Sciences (PhD funding support).

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.

References

Alves JMP, Serrano MG, Maia da Silva F *et al.* Genome evolution and phylogenomic analysis of Candidatus Kinetoplastibacterium, the betaproteobacterial endosymbionts of Strigomonas and Angomonas. *Genome Biol Evol* 2013;**5**:338–50.

Arkin AP, Cottingham RW, Henry CS *et al.* KBase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* 2018;**36**:566–9.

Assefa A, Abunna F. Maintenance of Fish Health in Aquaculture: Review of Epidemiological Approaches for Prevention and Control of Infectious Disease of Fish. *Vet Med Int* 2018;**2018**:5432497.

Bouwmeester MM, Goedknegt MA, Poulin R *et al.* Collateral diseases: Aquaculture impacts on wildlife infections. *J Appl Ecol* 2021;**58**:453–64.

Boyd CE, McNevin AA, Davis RP. The contribution of fisheries and aquaculture to the global protein supply. *Food Secur* 2022;**14**:805–27.

de Bruijn I, Liu Y, Wiegertjes GF *et al.* Exploring fish microbial communities to mitigate emerging diseases in aquaculture. *FEMS Microbiol Ecol* 2018;**94**, DOI: 10.1093/femsec/fix161.

Buso P, Diblasi C, Manousi D *et al.* Parallel selection in domesticated Atlantic salmon from divergent founders including parallel selection on WGD-derived homeologous regions. *bioRxiv* 2024, DOI: 10.1101/2024.07.22.604627.

Canellas ALB, Costa WF, Freitas-Silva J *et al.* In sickness and in health: Insights into the application of omics in aquaculture settings under a microbiological perspective. *Aquaculture* 2022;**554**:738132.

Carrión VJ, Cordovez V, Tyc O *et al.* Involvement of Burkholderiaceae and sulfurous volatiles in disease-suppressive soils. *ISME J* 2018;**12**:2307–21.

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

Chen S, Zhou Y, Chen Y *et al.* fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;**34**:i884–90.

Colombo S, Arioli S, Guglielmetti S *et al.* Virome-associated antibiotic-resistance genes in an experimental aquaculture facility. *FEMS Microbiol Ecol* 2016;**92**:fiw003.

Dahle SW, Bakke I, Birkeland M *et al.* Production of lumpfish (Cyclopterus lumpus L.) in RAS with distinct water treatments: Effects on fish survival, growth, gill health and microbial communities in rearing water and biofilm. *Aquaculture* 2020;**522**:735097.

Dahle SW, Gaarden SI, Buhaug JF *et al.* Long-term microbial community structures and dynamics in a commercial RAS during seven production batches of Atlantic salmon fry (Salmo salar). *Aquaculture* 2023;**565**:739155.

Dang H, Lovell CR. Microbial surface colonization and biofilm development in marine environments. *Microbiol Mol Biol Rev* 2016;**80**:91–138.

Defoirdt T, Sorgeloos P, Bossier P. Alternatives to antibiotics for the control of bacterial disease in aquaculture. *Curr Opin Microbiol* 2011;**14**:251–8.

Diana JS. Aquaculture Production and Biodiversity Conservation. *Bioscience* 2009;**59**:27–38.

Fan H, Li J, Wu W *et al.* Description of a moderately acidotolerant and aerotolerant anaerobic bacterium Acidilutibacter cellobiosedens gen. nov., sp. nov. within the family Acidilutibacteraceae fam. nov., and proposal of Sporanaerobacteraceae fam. nov. and Tepidimicrobiaceae fam. nov. *Syst Appl Microbiol* 2023;**46**:126376.

Ferreira JG, Taylor NGH, Cubillo A *et al.* An integrated model for aquaculture production, pathogen interaction, and environmental effects. *Aquaculture* 2021;**536**:736438.

Gjerde B, Boison SA, Hazlerigg D *et al.* Impact of three light smoltification regimes on performance and genetic parameters of traits in Atlantic salmon. *Genetics* 2024.

Glidden CK, Nova N, Kain MP *et al.* Human-mediated impacts on biodiversity and the consequences for zoonotic disease spillover. 2021, DOI: 10.22541/au.162333361.10264935/v4.

Grønvold L, van Dalum MJ, Striberny A *et al.* Transcriptomic profiling of gill biopsies to define predictive markers for seawater survival in farmed Atlantic salmon. *J Fish Biol* 2024, DOI: 10.1111/jfb.16025.

Gschwend F, Hartmann M, Mayerhofer J *et al.* Site and land-use associations of soil bacteria and fungi define core and indicative taxa. *FEMS Microbiol Ecol* 2022;**97**, DOI: 10.1093/femsec/fiab165.

Hajnsdorf E, Kaberdin VR. RNA polyadenylation and its consequences in prokaryotes. *Philos Trans R Soc Lond B Biol Sci* 2018;**373**:20180166.

Hal AM, El-Barbary MI. Gene expression and histopathological changes of Nile tilapia (Oreochromis niloticus) infected with Aeromonas hydrophila and Pseudomonas fluorescens. *Aquaculture* 2020;**526**:735392.

Hamza F, Satpute S, Banpurkar A *et al.* Biosurfactant from a marine bacterium disrupts biofilms of pathogenic bacteria in a tropical aquaculture system. *FEMS Microbiol Ecol* 2017;**93**, DOI: 10.1093/femsec/fix140.

Haro-Moreno JM, Coutinho FH, Zaragoza-Solas A *et al.* Dysbiosis in marine aquaculture revealed through microbiome analysis: reverse ecology for environmental sustainability. *FEMS Microbiol Ecol* 2020;**96**, DOI: 10.1093/femsec/fiaa218.

Heimeier D, Garland EC, Eichenberger F *et al.* A pan-cetacean MHC amplicon sequencing panel developed and evaluated in combination with genome assemblies. *Mol Ecol Resour* 2024;**24**:e13955.

He K, Minias P, Dunn PO. Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biol Evol* 2021;**13**, DOI: 10.1093/gbe/evaa270.

Hoyle N, Zhvaniya P, Balarjishvili N *et al.* Phage therapy against Achromobacter xylosoxidans lung infection in a patient with cystic fibrosis: a case report. *Res Microbiol* 2018;**169**:540–2.

Huang Y, Hou X, Liu S *et al.* Correspondence analysis of bio-refractory compounds degradation and microbiological community distribution in anaerobic filter for coking wastewater treatment. *Chem Eng J* 2016;**304**:864–72.

Hua Z-S, Han Y-J, Chen L-X *et al.* Ecological roles of dominant and rare prokaryotes in acid mine drainage revealed by metagenomics and metatranscriptomics. *ISME J* 2015;**9**:1280–94.

Iguchi A, Nagaya Y, Pradel E *et al.* Genome evolution and plasticity of Serratia marcescens, an important multidrug-resistant nosocomial pathogen. *Genome Biol Evol* 2014;**6**:2096–110.

Janda JM, Abbott SL. The changing face of the family Enterobacteriaceae (order: "Enterobacterales"): New members, taxonomic issues, geographic expansion, and new diseases and disease syndromes. *Clin Microbiol Rev* 2021;**34**, DOI: 10.1128/CMR.00174-20.

Jiang Y, Xiong X, Danska J *et al.* Metatranscriptomic analysis of diverse microbial communities reveals core metabolic pathways and microbiome-specific functionality. *Microbiome* 2016;**4**:2.

Jorth P, Turner KH, Gumus P *et al.* Metatranscriptomics of the human oral microbiome during health and disease. *MBio* 2014;**5**:e01012–4.

Joseph SW, Daily OP, Hunt WS *et al.* Aeromonas primary wound infection of a diver in polluted waters. *J Clin Microbiol* 1979;**10**:46–9.

Komagata K, Iino T, Yamada Y. The Family Acetobacteraceae. *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, 3–78.

Leray M, Wilkins LGE, Apprill A *et al.* Natural experiments and long-term monitoring are critical to understand and predict marine host-microbe ecology and evolution. *PLoS Biol* 2021;**19**:e3001322.

Leung TLF, Bates AE. More rapid and severe disease outbreaks for aquaculture at the tropics: implications for food security. *J Appl Ecol* 2013;**50**:215–22.

Little DC, Newton RW, Beveridge MCM. Aquaculture: a rapidly growing and significant source of sustainable food? Status, transitions and potential. *Proc Nutr Soc* 2016;**75**:274–86.

Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.

Lulijwa R, Rupia EJ, Alfaro AC. Antibiotic use in aquaculture, policies and regulation, health and environmental risks: a review of the top 15 major producers. *Rev Aquac* 2020;**12**:640–63.

Luo Y, Kanai M, Choi W *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat Genet* 2021;**53**:1504–16.

Mahlen SD. Serratia infections: from military experiments to current practice. *Clin Microbiol Rev* 2011;**24**:755–91.

Massimino L, Lamparelli LA, Houshyar Y *et al.* The Inflammatory Bowel Disease Transcriptome and Metatranscriptome Meta-Analysis (IBD TaMMA) framework. *Nature Computational Science* 2021;**1**:511–5.

Mkulo EM, Wang B, Amoah K *et al.* The current status and development forecasts of vaccines for aquaculture and its effects on bacterial and viral diseases. *Microb Pathog* 2024;**196**:106971.

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

Morshed SM, Lee T-H. The role of the microbiome on fish mucosal immunity under changing environments. *Fish Shellfish Immunol* 2023;**139**:108877.

Morya R, Salvachúa D, Thakur IS. Burkholderia: An untapped but promising bacterial genus for the conversion of aromatic compounds. *Trends Biotechnol* 2020;**38**:963–75.

Mougin J, Joyce A. Fish disease prevention via microbial dysbiosis-associated biomarkers in aquaculture. *Rev Aquac* 2022, DOI: 10.1111/raq.12745.

Mumbo MT, Nyaboga EN, Kinyua JK *et al.* Antimicrobial resistance profiles of salmonella spp. and escherichia coli isolated from fresh nile tilapia (oreochromis niloticus) fish marketed for human consumption. *BMC Microbiol* 2023;**23**:306.

Muziasari WI, Pärnänen K, Johnson TA *et al.* Aquaculture changes the profile of antibiotic resistance and mobile genetic element associated genes in Baltic Sea sediments. *FEMS Microbiol Ecol* 2016;**92**:fiw052.

Ojala T, Häkkinen A-E, Kankuri E *et al.* Current concepts, advances, and challenges in deciphering the human microbiota with metatranscriptomics. *Trends Genet* 2023;**39**:686–702.

Okada Y, Momozawa Y, Ashikawa K *et al.* Construction of a population-specific HLA imputation reference panel and its application to Graves' disease risk in Japanese. *Nat Genet* 2015;**47**:798–802.

Parks DH, Chuvochina M, Rinke C *et al.* GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;**50**:D785–94.

Park SI. Disease Control in Korean Aquaculture. 魚病研究 2009;**44**:19–23.

Patil PP, Kumar S, Kaur A *et al.* Global transcriptome analysis of Stenotrophomonas maltophilia in response to growth at human body temperature. *Microb Genom* 2021;**7**, DOI: 10.1099/mgen.0.000600.

Pedrosa-Gerasmio IR, Nozaki R, Kondo H *et al.* Gut bacterial community profile in Pacific white shrimp *Litopenaeus vannamei* following 5-aminolevulinic acid supplementation. *Aquac Res* 2020;**51**:4075–86.

Rahman ANA, Mahboub HH, Elshopakey GE *et al.* Pseudomonas putida infection induces immune-antioxidant, hepato-renal, ethological, and histopathological/immunohistochemical disruptions in Oreochromis niloticus: the palliative role of titanium dioxide nanogel. *BMC Vet Res* 2024;**20**:127.

Rajeev R, Adithya KK, Kiran GS *et al.* Healthy microbiome: a key to successful and sustainable shrimp aquaculture. *Rev Aquac* 2021;**13**:238–58.

Ranjan K, Bharti MK, Siddique RA *et al.* Metatranscriptomics in microbiome study: A comprehensive approach. *Microbial Metatranscriptomics Belowground*. Singapore: Springer Singapore, 2021, 1–36.

Ranveer SA, Dasriya V, Ahmad MF *et al.* Positive and negative aspects of bacteriophages and their immense role in the food chain. *Npj Sci Food* 2024;**8**:1.

Rodger HD. Fish Disease Causing Economic Impact in Global Aquaculture. In: Adams A (ed.). *Fish Vaccines*. Basel: Springer Basel, 2016, 1–34.

Samsing F, Barnes AC. The rise of the opportunists: What are the drivers of the increase in infectious diseases caused by environmental and commensal bacteria? *Rev Aquac* 2024;**16**:1787–97.

Santos L, Ramos F. Antimicrobial resistance in aquaculture: Current knowledge and alternatives to tackle the problem. *Int J Antimicrob Agents* 2018;**52**:135–43.

Shaffer M, Borton MA, McGivern BB *et al.* DRAM for distilling microbial metabolism to automate the curation of microbiome function. *Nucleic Acids Res* 2020;**48**:8883–900.

Stenløkk KSR. Genomic structural variations as drivers of adaptation in salmonid fishes. 2023.

Stentiford GD, Sritunyalucksana K, Flegel TW *et al.* New Paradigms to Help Solve the Global Aquaculture Disease Crisis. *PLoS Pathog* 2017;**13**:e1006160.

Stephens M. False discovery rates: a new deal. *Biostatistics* 2017;**18**:275–94.

Stewart FJ, Ottesen EA, DeLong EF. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* 2010;**4**:896–907.

Tang Z-Z, Chen G, Alekseyenko AV. PERMANOVA-S: association test for microbial community composition that accommodates confounders and multiple distances. *Bioinformatics* 2016;**32**:2618–25.

Taubert M, Grob C, Crombie A *et al.* Communal metabolism by Methylococcaceae and Methylophilaceae is driving rapid aerobic methane oxidation in sediments of a shallow seep near Elba, Italy. *Environ Microbiol* 2019;**21**:3780–95.

Testerman T, Beka L, Reichley SR *et al.* A large-scale, multi-year microbial community survey of a freshwater trout aquaculture facility. *FEMS Microbiol Ecol* 2022;**98**, DOI: 10.1093/femsec/fiac101.

Topp E, Larsson DGJ, Miller DN *et al.* Antimicrobial resistance and the environment: assessment of advances, gaps and recommendations for agriculture, aquaculture and pharmaceutical manufacturing. *FEMS Microbiol Ecol* 2018;**94**, DOI: 10.1093/femsec/fix185.

Ulaszewski B, Sandoval-Denis M, Groenewald JZ *et al.* Genomic features and evolution of lifestyles support the recognition of distinct genera among fusarioid fungi. *Mycol Prog* 2025;**24**, DOI: 10.1007/s11557-024-02025-4.

Wang H, Xie G, Huang J. Genome-based characterization of a novel prophage of Vibrio parahaemolyticus, VPS05ph1, a novel member of Peduoviridae. *Virology* 2024;**595**:110087.

Wang X, Pan J, Chen L *et al.* Prevalence, virulence-related genes and antimicrobial resistance of Aeromonas spp. from loach Misgurnus anguillicaudatus with skin ulcer and healthy controls in Southern China. *Aquaculture* 2022;**552**:738040.

Welti-Chanes J, Serna-Saldívar SO, Campanella O *et al. Science and Technology of Fibers in Food Systems*. Springer Nature, 2020.

Wippel K, Tao K, Niu Y *et al.* Host preference and invasiveness of commensal bacteria in the Lotus and Arabidopsis root microbiota. *Nat Microbiol* 2021;**6**:1150–62.

Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;**20**:257.

Yilmaz M, Çakir M, Oral O *et al.* Using machine learning technique for disease outbreak prediction in rainbow trout (*Oncorhynchus mykiss*) farms. *Aquac Res* 2022;**53**:6721–32.

Zamkovaya T, Foster JS, de Crécy-Lagard V *et al.* A network approach to elucidate and prioritize microbial dark matter in microbial communities. *ISME J* 2021;**15**:228–44.

Zhang Y, Thompson KN, Branck T *et al.* Metatranscriptomics for the human microbiome and microbial community functional profiling. *Annu Rev Biomed Data Sci* 2021;**4**:279–311.