Undergraduate Journal of Experimental Microbiology and Immunology (UJEMI) Vol. 28:1-11

Ocean spatial geography drives the midgut microbial composition of marine fish

Joean Lu, Kathy Wong, Catherine Yu, Angie Zhou

Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, Canada

SUMMARY Having emerged 600 million years ago, fish represent the largest diversity of vertebrates and have adapted exceptionally to continuous changes in the aquatic ecosystems. Despite their importance, fish have been significantly underrepresented in microbial ecology analyses and little is known about the environmental drivers of the fish gut microbial composition. To investigate the relationship between habitat depth and distance from shore in driving fish gut microbial composition, we analysed and compared the microbiota of the midgut and hindgut of marine fishes. Our study found that water depth and distance from shore do interact to drive fish microbial composition in midgut samples, but not in hindgut samples. We also found that genera associated with nutrient cycles (Rhodopiruella, Desulfotalea, Sva0081, Milano WF1B-44, Woeseia, Sulfurovum, and Massilia) and water pollution (Fluviicola, Ulvibacter, Fluviicola, and Rhizobium) can be reflected in the midgut microbial composition of marine fish. These findings provide a reference for future studies of the gut microbiome of fish as well as insights into the key role of spatial geography of the host habitat in influencing gut microbial composition.

INTRODUCTION

Despite the growing body of literature investigating the relationship between microbial communities and the environmental biomes in different animal hosts, past studies have mainly focused on the gut microbiota of mammals [1]. Fish, on the other hand, are relatively underrepresented in microbial ecology studies, and little is known regarding how anthropogenic (human-originating) activities influence the gut microbiota of fish in aquatic environments [2]. Throughout their evolution, fish have developed various physiological adaptations and have co-evolved with symbiotic gut microbes to cope with the continuously changing conditions in different environments [1]. Therefore, as environmental stressors rise due to anthropogenic pollutants, understanding the role of the aquatic habitat on the gut microbiota of fish is important for maintaining fish health and biodiversity, both of which have significant impacts on crucial industries like aquaculture and commercial fisheries. Considering the sensitivity of fish to the presence of contaminants and their ability to uptake toxins from within the water body [3], evaluating the environmental drivers of the gut microbiota in fish populations may also represent an important step in defining microbial biomarkers for water pollution.

Currently, the gut microbiota in fish is known to be influenced by various biological factors such as host phylogeny, age, and diet, as well as environmental factors, such as climate, habitat, and geography [4]. For instance, trophic level is known to play a role in shaping the gut microbial communities of fish as different symbiotic bacteria aid in the acquisition of nutrients from different diets [5]. Previous studies have also demonstrated that pelagic zone and water depth can independently affect microbiome diversity [4]. Furthermore, the midgut and hindgut of fish are known to have different dominant species compared to other body sites (mucus, skin, and gills) [6] in addition to differences in microbial communities between the gut sites themselves [7, 8]. Despite these recent advances in fish microbial ecology, there appears to be very few studies that have performed a comprehensive analysis on the role of host habitat in driving microbial diversity in the fish gut. As such, understanding how spatial factors like water depth and distance from shore

Published Online: September 2023

Citation: Lu, Wong, Yu, Zhou. 2023. Ocean spatial geography drives the midgut microbial composition of marine fish. UJEMI 28:1-11

Editor: Lauren Pugsley and Shruti Sandilya, University of British Columbia

Address correspondence to: https://jemi.microbiology.ubc.ca/

interact to drive gut microbial composition, especially in fish of a particular trophic level, remains a considerably understudied subject in the field of fish microbiology.

Using the Fish Microbiome Project (FMP) dataset by Minich *et al.* [4], we sought to determine whether habitat depth and distance from shore can interact to drive gut microbial diversity of trophic level three (TL3) (secondary consumer) fishes. Previous research on this dataset aimed to investigate the primary factors that influence fish microbial communities and identified body site to be the strongest predictor of diversity [4]. Habitat depth and distance from shore were also shown to independently drive changes in alpha and beta diversity, although some of these analyses were selectively performed on smaller subsets of the data and focused only on gill samples or only on fishes in the neritic zone [4]. By comparing fish samples from three representative habitat regions defined by spatial geography (Figure 1) and

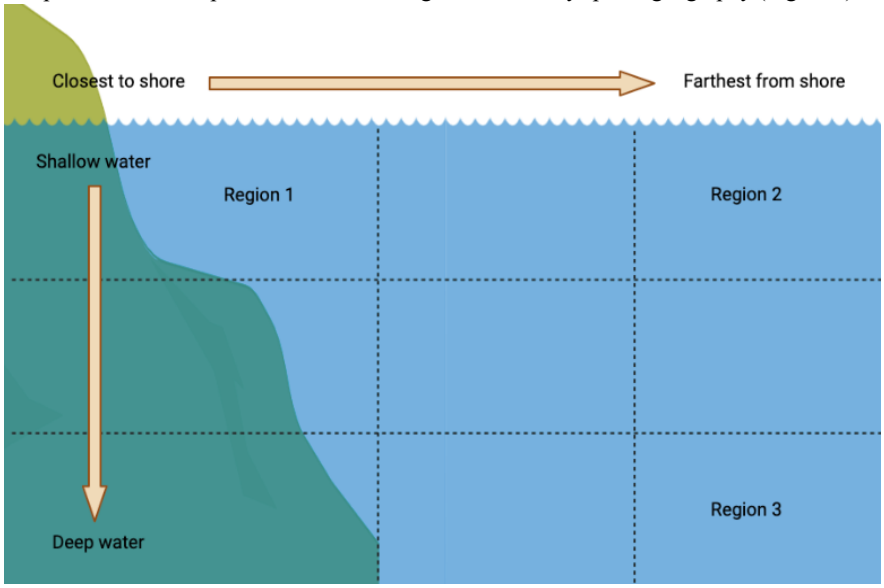

FIG. 1 Fish microbial samples were categorized by ocean spatial geography based on water depth and distance from shore. Microbial diversity was compared between Regions 1, 2, and 3. Region 1 consists of samples collected from shallow waters closer to shore, Region 2 (control) consists of samples collected from shallow waters farther from shore, and Region 3 consists of samples collected from deeper waters farther from shore. This figure was generated using Biorender.

assessing only TL3 fish, we can novelly investigate if there is a relationship between habitat depth and distance from shore in driving gut microbial composition, and whether one of these factors has a stronger impact on diversity than the other. To address our research questions, we conducted diversity and differential abundance analyses and compared results between the midgut and hindgut.

We assessed whether the original correlation holds true when a combination of spatial factors are considered within TL3 fish. If these abiotic environmental factors do interact to determine TL3 fish microbiota, we would expect to find between habitat regions close to shore and in shallow waters and habitat regions far from shore and in deeper waters. We further assessed whether these differences in microbial composition are consistent between the midgut and hindgut, as previous studies have found that microbial abundance and composition differs with body site [4]. Thus, our study will provide evidence on the effects of spatial geography on the gut microbial composition of marine fish, as well as identify differentially-abundant genera in each host habitat.

METHODS AND MATERIALS

Datasets and metadata. The FMP dataset was generated by Minich *et al.*, wherein the microbiota from four primary fish mucosal body sites (gill, skin, midgut, and hindgut) were sampled and analyzed for 101 species (28 orders, 55 families, and 83 genera) of marine fishes from the Eastern Pacific Ocean in Southern California [4]. DNA extractions were processed using the Qiagen PowerMag kit and PCR was performed following the standard Earth Microbiome Project (EMP) protocols [9] for the V4 region of the 16S rRNA gene 515F/806 Rb [10]. Samples were then sequenced using Illumina MiSeq [11] and NovaSeq protocols [12]. The associated metadata contained information about several life history metrics such

as diet and trophic level, biometrics, and habitat classifications. Our study focused on fish from TL3 and the metadata categories distance from shore and habitat depth.

Metadata manipulation. Metadata manipulation was performed in R (version 4.2.2) [13] using the packages Dplyr (version 1.1.0) and Tidyverse (version 1.3.2) [14, 15]. All data manipulation and analyses performed are detailed in the supplemental R script (RScript). Our investigation focused on the microbial communities of the gut, therefore samples were filtered for TL3 to control for variation associated with diet. Trophic level was assigned by Minich *et al.* using previously documented diet data, in which TL3 was defined as secondary consumers [4]. The dataset was then binned by pelagic zone into low depth "intertidal or neritic" (1–50m), medium depth "mesopelagic or bathypelagic" (150–500m), and high depth "abyssopelagic or benthopelagic" (>500m) categories. The dataset was further binned by distance from shore with samples falling into either "close to shore," "medium distance to shore," or "farthest from shore." The new depth and distance classifications were then inputted into a new column in the dataset. Using these metrics, the samples were assigned to one of nine habitat regions, defined by their spatial geography. This study compares microbial diversity between Regions 1, 2, and 3 host habitats, defined using the water depth and distance from shore categories described above (Figure 1). Region 1 contains 67 samples, Region 2 contains 27 samples, and Region 3 contains 40 samples.

Data processing and phylogeny using the QIIME2 pipeline. All data analyses were performed using the Quantitative Insights Into Microbial Ecology (QIIME2) bioinformatics pipeline [16] and are detailed in the supplemental QIIME2 script (QIIME2Script). Filtered manifest files for the midgut and the hindgut were generated using R [13] and imported back into QIIME2 (version 2021.11.0) to generate demultiplexed 16S rRNA sequences [13, 16]. The following processing steps were performed in parallel for the midgut and the hindgut datasets. Sequence quality control was performed using the QIIME2 plugin Divisive Amplicon Denoising Algorithm 2 (q2-DADA2) which detects and corrects sequencing errors to allow researchers to resolve microbial communities at the ASV level [17]. A read length of 180 nucleotides was retained to maintain a median Phred quality score of 25. Given that the median Phred score dropped to 11 at truncation lengths greater than 180 (maximum retained bases), 180 nucleotides was the max sequence length retaining sufficient sequence quality. After denoising using the DADA2 method, the generated features tables for both midgut and hindgut were exported into R for further filtering [17]. To facilitate downstream filtering and diversity analyses, the QIIME2 phylogeny tool Multiple Alignment using Fast Fourier Transform (MAFFT) [18, 19] was used to perform multiple sequence alignment and the FastTree q2-phylogeny plugin [20] was applied to create the corresponding rooted phylogenetic tree which was then exported into R.

Taxonomic classification. The q2-feature-classifier plugin was used to train a Naive Bayes classifier [21] on the 99% Silva 138 reference database [22] using the same 515F/806R primer pair that was used for Illumina sequencing of the original dataset samples. Sequences were then truncated to 180 nucleotides to match the length of ASVs generated using DADA2 [17]. After the classifier was trained with the new ref-seq file, a taxonomy artifact was generated and visualised by the taxa barplot function in QIIME2. Lastly, the taxonomy artifact containing taxonomic assignments of the representative sequences was exported for analysis in R. This process was performed in parallel for the midgut and hindgut samples.

Features table filtering and ASV rarefaction. The filtered metadata and the imported feature table, taxonomy, and rooted phylogenetic tree artifacts for both the midgut and hindgut were integrated into their respective phyloseq objects in R [13] using the Phyloseq package [23]. The phyloseq object was filtered to exclude low-abundance taxa (<5 reads) and samples (<100 reads), and subsetted to include only bacterial ASVs (removed mitochondrial and chloroplast DNA). The resulting filtered phyloseq object was used to generate an alpha rarefaction curve. Using the rarefaction curve, a sequencing depth of 2000 reads per sample was selected to maximize the number of samples and ASVs retained. After rarefaction, 22 midgut and 18 hindgut samples remained. For the midgut and hindgut, Region 1 contains 11

samples and Region 2 contains two samples. Region 3 contains nine samples and 5 samples in the midgut and hindgut respectively.

Alpha and beta diversity analysis. R [13] was used to calculate both α-diversity (Chao1 and Shannon diversity index) [24, 25] and β-diversity (Bray-Curtis diversity index) [26] using the Vegan package [27] for both the midgut and hindgut samples. All data analyses performed are detailed in the supplemental R script (RScript). Statistical testing of habitat regions against alpha diversity measures was performed using non-parametric Wilcoxon rank-sum test [28]. Statistical testing of habitat regions against beta diversity measures was performed using PerMANOVA [29]. Alpha and beta diversity plots were generated using the packages ggplot2 [30] and ggpubr [31] in R.

Differential abundance analysis. Differential abundance analysis was performed in R [13] for both the midgut and the hindgut using the raw data and the packages Phyloseq, Tidyverse, Vegan, Ape, and DESeq2 [23, 15, 27, 32, 33]. All data analyses performed are detailed in the supplemental R script (RScript) and includes analysis of differentially abundant taxa in Region 2 compared to Region 1, Region 3 compared to Region 2, and Region 3 compared to Region 1. Significance was defined as differentially abundant genera with an adjusted Wald Chi-Squared Test p-value of <0.01 [34]. Significant differentially abundant genera were plotted using ggplot2 [30].

Relative abundance analysis. Relative abundance analysis was performed in R [13] for both the midgut and hindgut samples using the packages Phyloseq and Tidyverse [23, 15]. Significance was defined as differentially abundant genera with an adjusted Wald Chi-Squared Test p-value of <0.01 [34]. Significant relatively abundant genera were plotted using ggplot2 [30].

Figure formatting. All figures generated by R [13] were reformatted and annotated using Microsoft Powerpoint and Word.

Data availability. All samples from the FMP dataset are publicly available at the European Nucleotide Archive under project number PRJEB54736 and at Qiita under study ID 13414 [4].

RESULTS

Marine fish exhibit differences in richness and abundance in the midgut based on habitat region. To assess the microbial composition of samples across different habitat regions, beta diversity analyses were performed on the FMP dataset using the Bray-Curtis dissimilarity metric. Midgut samples showed significant separation between Region 1 and Region 3 clusters, with samples in the control, Region 2, falling in between those clusters (Figure 2A). Region 1 and Region 3 clusters were far apart, indicating that samples collected from habitats closest to shore and in shallow waters and habitats farthest from shore and in deep water have high levels of dissimilarity (Figure 2A). The samples from Region 2 were collected from habitats farthest from shore but in shallow waters, thus the clustering of these samples with Region 1 and 3 samples is consistent with Region 2 sharing characteristics of the other two habitats (Figure 1). Combined, these trends highlight the robustness of Region 2 as a control and allows us to parse out the differences in the midgut microbial composition that may be driven independently by depth or distance from shore. However, these results are not consistent across gut sites as no clustering by region is seen in the hindgut (Figure 2B). As the samples significantly differ by habitat region in the midgut (Figure 2A), these results suggest that microbial richness and abundance of the midgut, but not the hindgut, is driven by habitat region.

Alpha diversity analyses were also performed using the Chao1 and Shannon diversity metrics. No significant differences ($p < 0.05$) in diversity between the habitat regions or across either gut site were observed (Supplemental Figure S2A, S2B).

Differentially abundant taxa in the midgut are observed in different habitat regions. Differential abundance analysis was performed using DESeq2 to assess differentially abundant taxa in the habitat regions. When assessing taxa driven by distance from shore in the midgut, Region 1 samples collected from habitats closest to shore showed increased abundance in 14 taxa compared to the control, Region 2 samples which were collected from habitats farthest from shore (Figure 3A). Moreover, when assessing taxa driven by water depth, Region 3 samples collected from habitats in deep waters showed increased abundance in 12 genera compared to Region 2 samples collected from habitats in shallow waters (Figure 3B). Lastly, when assessing taxa driven by both distance from shore and water depth, Region 3 samples collected from habitats farthest from shore and in deep waters showed decreased abundance in 9 genera and increased abundance in 31 genera compared to Region 1 samples collected from habitats closest to shore and in shallow waters (Figure 3C). Interestingly, the three regions showed stark contrast between their differentially abundant genera, with all DESeq2 microbes decreased in Region 2 vs Region 1, and all microbes increased in Region 3 vs Region 2 (Figure 3A, 3B). Taken together, these findings suggest that differential abundance in the midgut is driven by both water depth and distance from shore, leading to overall differences in microbial diversity between habitat regions.

DISCUSSION

There are limited microbial ecology studies that investigate how environmental factors influence the gut microbial composition of fish in aquatic environments [4]. This literature gap has consequences as the rise of environmental stressors due to anthropogenic activities may have adverse effects on the gut microbial composition of fish [2]. Using beta diversity analyses and differential abundance testing, we found that water depth and distance from shore interact to drive fish microbial composition in midgut samples. Additionally, differentially abundant genera associated with nutrient cycling and components of water pollution were found in the midgut of marine fish between habitat regions.

Water depth and distance from shore are demonstrated to be environmental drivers of microbial composition in the midgut, but not the hindgut. We investigated the role of spatial geography, specifically the interplay of water depth and distance from shore in affecting fish gut microbiome diversity. Using beta diversity analyses, we found that only midgut samples clustered by habitat regions, suggesting that the microbial composition of the midgut is driven by both water depth and distance from shore (Figure 2). This finding is

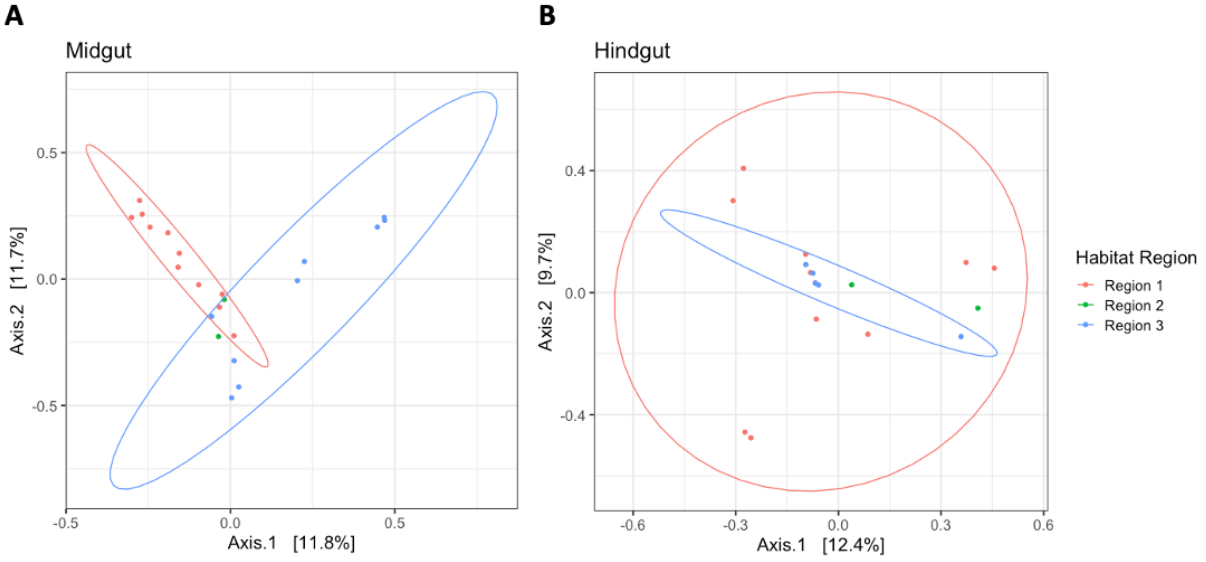

FIG. 2 Fishes in varying habitat regions exhibit differences in richness and abundance of microbial communities in the midgut. Principal coordinate analysis (PCoA) plots of the midgut **(A)** and hindgut **(B)** microbiota based on Bray-Curtis dissimilarity shows significant differences in beta diversity between Regions 1 and 3 in the midgut ($p = 0.001$), but not in the hindgut ($p = 0.168$). Statistical analysis was determined by permutational multivariate analysis of variance (PERMANOVA). Ellipses represent 95% confidence level; Region 2 contains too few samples to be recognized as a group.

consistent with the literature as it has been previously documented that the fish gut microbiota is influenced by various environmental factors, such as habitat and geography [2]. Depending on the level of water depth, habitats vary in the availability of organic matter, nutrients, and sunlight exposure, as well as oxygen concentration, and water temperature, which subsequently affects microbiota diversity [3,4]. Consequently, deep-sea fishes, particularly those with limited vertical migration, would experience different microbial exposures at varying depths [2]. Another environmental factor, distance from shore, has also been found to determine the nutrient and terrestrial composition of different aquatic habitats. It is likely that these changes in the host habitat may correlate with changes in fish microbiomes, as fish too are affected by nutrient and terrestrial input [6].

Interestingly, these patterns of microbial diversity were not seen in the hindgut as our alpha and beta diversity analyses both did not show any significant differences in microbial abundance or composition according to habitat regions (Figure 2B, Supplementary Figure 1B). This was not unexpected as the midgut is where the majority of digestion occurs, and the hindgut is only for waste excretion [12]. The midgut receives water and food that are populated with microorganisms, which subsequently affects the makeup of the resident microbiota [11]. The lack of significance in the hindgut may be explained by the low overall gamma diversity associated with hindgut samples in fish species and the fact that the hindgut was found to have the lowest site specific ASVs compared to other body sites [2]. Furthermore, the significant results found in the midgut coincide with the fact that the fish gut greatly varies in physicochemical conditions such as oxygen, pH, and organic substrate levels, compared to other body sites (mucus, skin, and gills) [11]. Our findings propose that the midgut drives variance and suggest that the microbiome of the midgut may be a more useful reflection of environmental factors than other gut sites. Our results support our hypotheses that gut microbial composition could be influenced by both water depth and distance from shore, and that the midgut and hindgut demonstrate differences in microbial diversity.

Differentially abundant genera between regions are associated with their respective habitats. To identify differentially abundant taxa between regions, DESeq2 was performed, and the genera identified in the midgut of fish were in agreement with the regions they were sampled from. For instance, Ulvibacter, Halioblogus, and the NS5 marine group genera were enriched in coastal waters and have also been associated with coastal habitats [35–37] (Figure 3A, C). These groups were less prevalent in Regions 2 or 3 compared to Region 1, indicating that they are enriched in coastal habitats. Complementary to this finding, we found that genera enriched in deeper waters, Marixanthomonas, Brevundimonas, Alcanivorax, and Salinimicrobium, are more associated with deep oceans or seas [38–41]. These expected results help to validate our DESeq2 approach.

Differentially abundant taxa are related to nutrient cycling. Many of the differentially abundant genera found in the midgut of fish suggest a role in nutrient cycles, such as sulfur oxidation and reduction, and necromass degradation (the recycling of dead microbes). Genera associated with sulfate/sulfur reduction, Rhodopiruella, Desulfotalea, and the Sva0081 sediment group [42–44], were enriched in deeper and farther waters whereas the coastal waters were enriched in sulfur oxidizing genera, such as Milano WF1B-44, Woeseia, and Sulfurovum [45–47] (Figure 3A–C). This activity alludes to potential roles in the sulfur cycle which describes how sulfur is transferred between rocks, sediments, aquatic phases, and the atmosphere through various forms [48]. This cycling is mitigated by microbes and helps connect different parts of an ecosystem that require different forms of sulfur [48]. Sulfur reduction is when oxidized sulfur acts as an electron acceptor to yield H_2 or H_2S [48]. This is generally observed on the ocean floor when organic material settles, which is further supported by its increased abundance in Regions 2 and 3 [49]. Conversely, sulfur oxidation is when the sulfur compounds are acting as electron donors rather than acceptors [48]. These sulfates are usually introduced through run-off from land and support the increased abundance of sulfur oxidizing genera in coastal waters. This ties in with another genus that was enriched in deeper, farther waters – Massilia (Figure 3C) [50]. This genus is associated with necromass degradation and may be interacting with the sulfur reducers as a source of oxidized sulfur. As coastal waters may be more affected by run-off based on sheer proximity, it would follow that sulfur oxidizers capable of processing sulfates would localize in regions

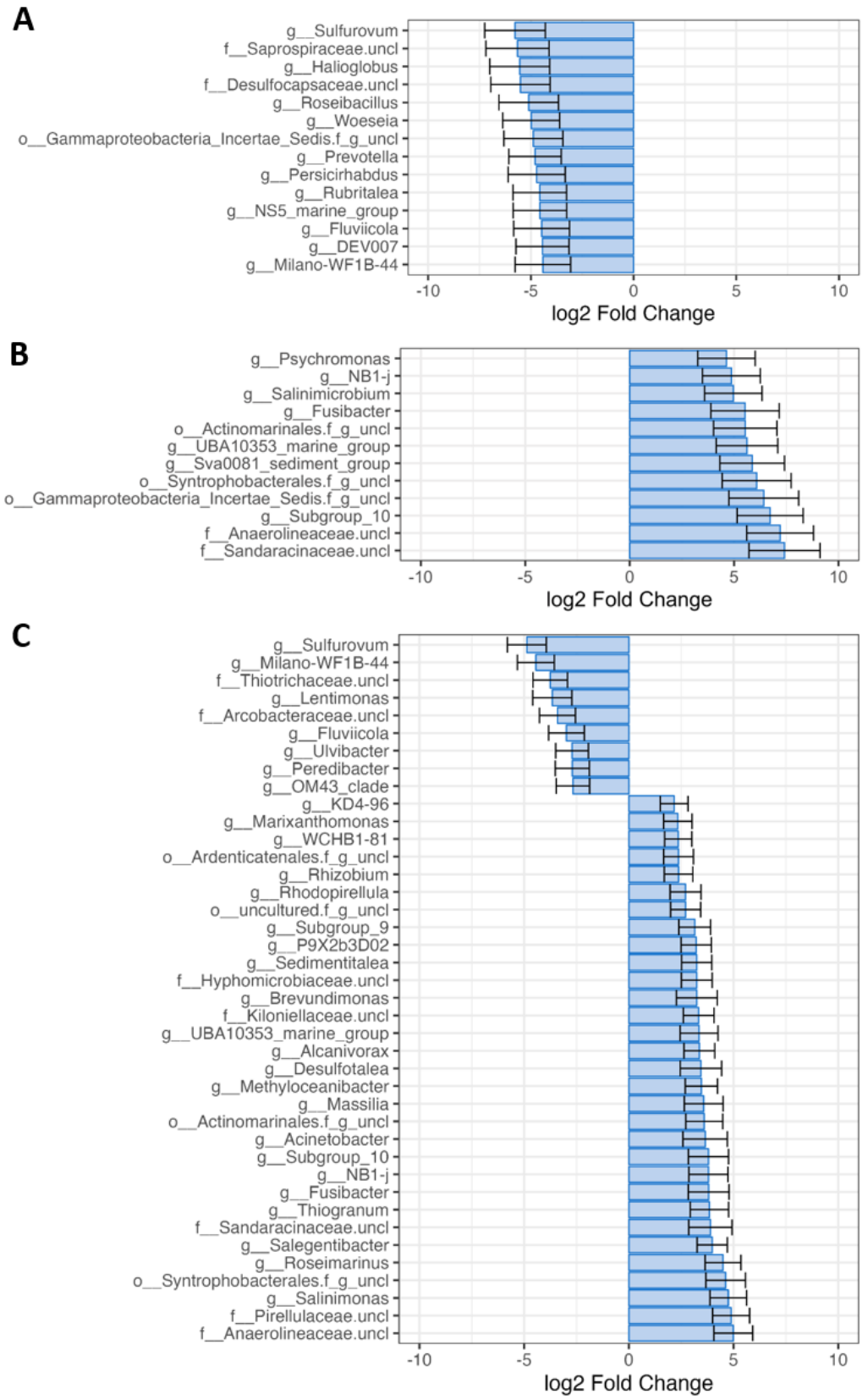

FIG. 3 Differentially abundant taxa are observed in different habitat regions in the midgut. Differentially abundant taxa in Region 2 compared to Region 1 **(A)**, Region 3 compared to Region 2 **(B)**, and Region 3 compared to Region 1 **(C)** were assessed by differential abundance analysis using DESeq2/Phyloseq. Significance was defined as differentially abundant genera with an adjusted Wald Chi-Squared Test p-value of <0.01.

with more sulfate. Together, being able to see a visual representation of the sulfur cycle through the fish midgut microbiota with prior understanding of the sulfur cycle, yields implications for future studies on nutrient cycling potentially using midgut microbial diversity as a readout. For instance, recent research has suggested that sulfur is a major evolutionary and ecological factor in determining microbial life on marine seafloors, and thus, understanding these trends can provide insight on how sulfur sustains specific niches and how these may be affected with changes in nutrient cycling.

Differentially abundant taxa have implications for water pollution. Water pollution has many adverse effects ranging from disrupting aquatic ecosystems, to spreading water-borne diseases for both humans and fish [51, 52]. Therefore, it is important to understand the spread of water pollution and the sources of water pollution. Differentially abundant genera between different regions found in the midgut of fish are associated with different components of water pollution, such as microplastics, wastewater, sewage, and minerals (Figure 3A–C). Fluviicola and Ulvibacter are genera that have been isolated from sewage and wastewater and were enriched in coastal waters (Figure 3A, C) [37,53]. Wastewater and sewage can be introduced indirectly through run-off, when groundwater flow carries waste that has leached into the soil into the water [54]. This is further supported by the NS5 marine group [36] which is enriched with Fluviicola [53] and is associated with eutrophication (Figure 3B, C). Similarly to wastewater contamination, eutrophication is excessive richness of minerals and nutrients in water and is introduced by run-off. Together, these results support that run-off water pollution is indeed present in the ocean, but more significantly, these results suggest that the effects of run-off become diluted with distance from shore. Therefore, when trying to mitigate run-off water pollution, efforts may be better spent on cleaning coastal waters.

However, this may not be extrapolated to other forms of water pollution. Rhizobium were found enriched in deeper and farther waters (Figure 3C) and have been associated with microplastics [55]. This further exemplifies the impact of microplastics on the environment as these results suggest that they are present even in the deepest parts of the ocean. This trend is opposite to the run-off water pollution suggesting that different forms of water pollution need to be addressed differently as they do not behave the same, and, at large, water pollution is found in the ocean regardless of distance and depth from shore. Water pollution has many adverse effects on ecosystems and human health, and these results further support that efforts for mitigating water pollution should be prioritised. These efforts could be supported by monitoring the relative abundance of species associated with different types of pollution, as it may be difficult to isolate microplastics from the seafloor in a representative manner.

Limitations A main limitation of this study was that the samples were low biomass samples, which resulted in many samples returning too few reads for useful analysis after denoising. After filtering and rarifying the data, only 22 midgut samples remained for beta diversity analysis. This significant loss in the amount of reads reduces representation within our data. Moreover, this study only focuses on TL3 fish. Since trophic level alone can highly influence the diversity in fish microbial communities [2], the results from this study may not be representative of all marine fish. The impact of other confounding variables, such as habitat substrates, effects of evolutionary distance, and host phylogeny were also not accounted for in this study.

Conclusions This study aimed to assess the influence of ocean spatial geography on the gut microbial composition of marine fish. Marine fish exhibited differences in richness and abundance in the midgut based on habitat region, suggesting that water depth and distance from shore interacts to drive fish microbial composition in the midgut. Additionally, marine fish from different habitat regions showed differentially abundant genera associated with nutrient cycling and components of water pollution in the midgut. Observed farther from shore include more sulfur-reducing bacteria, less sulfur-oxidizing bacteria, and less bacteria associated with wastewater, sewage, and excessive mineral deposition. This suggests that differentially abundant genera associated with nutrient cycling and water pollution are reflected in the midgut microbial composition of marine fish. In contrast, the hindgut showed

no significant differences in beta diversity and few differentially abundant taxa between habitat regions. While our study provides evidence on the influence of spatial geography on the fish gut microbial composition, further studies need to be conducted to explore the potential impacts of environmental pollutants on the fish microbiome.

Future Directions To address the limitations in this study, future research could replicate this study but with a larger, more complete dataset with higher quality reads. This would result in a larger sample size for producing more statistically significant and accurate results. Additionally, future studies could investigate the dominant differentially abundant taxa at the species level rather than the genus and family level. This could improve the specificity of the results and possibly bring to light some important species-specific information that was not uncovered in our study. Furthermore, future directions could also evaluate other trophic levels to assess the similarities and differences in the midgut microbial composition between different trophic level fish. Since different trophic levels are associated with variations in diet and the gut microbiomes aid in the digestion of food, fish of different trophic levels could presumably have varying microbial compositions [4]. Performing further analyses could also provide us with more valuable information to support our findings. For example, indicator taxa analysis could determine unique indicator taxa associated with the different habitat regions, such that the presence or absence of certain taxa reflects a specific environmental condition. Additionally, Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) could be performed to predict metagenome functional abundance.

Another future direction could be to monitor the fish midgut microbial composition long-term to investigate whether the differentially abundant genera associated with pollution in our study would continue to increase or decrease in abundance over time. In particular, it would be interesting to assess the change in abundance of these unique taxa post-exposure to certain pollutants, such as an oil spill, to investigate microbiome recovery after a polluting event. Although fish are generally one of the most important aquatic communities concerning humans, the overexploitation of water resources for human developmental activities has led to increasing levels of pollutants resulting in harmful effects on fish [3]. Monitoring the fish midgut microbiome over time would be important in evaluating the impact of water pollution on fish, identifying the presence of toxins in aquatic ecosystems, and investigating recovery of the gut microbiome after a polluting event.

ACKNOWLEDGEMENTS

We would like to thank Dr. Evelyn Sun, Dr. Melissa Chen, Avril Metcalfe-Roach, and the rest of the MICB 475 teaching team for their guidance and support throughout this project. We would also like to thank the UBC Department of Microbiology and Immunology for providing the facilities, funding, and resources for this project. Lastly, we would like to acknowledge Minich et al. for providing the study data materials.

CONTRIBUTIONS

Joean Lu (JL), Kathy Wong (KW), Catherine Yu (CY), and Angie Zhou (AZ) contributed equally to the writing of all scripts in QIIME2 and R, and the writing of the manuscript was a collective effort of all authors. Generally, the QIIME2 processing and alpha diversity analyses were performed by JL and CY while the beta diversity analyses, differential abundance and relative abundance testing were performed by KW and AZ. All authors contributed to editing of the manuscript.

REFERENCES

1. **Kim PS, Shin N-R, Lee J-B, Kim M-S, Whon TW, Hyun D-W, Yun J-H, Jung M-J, Kim JY, Bae J-W.** 2021. Host habitat is the major determinant of the gut microbiome of fish. Microbiome **9**:166.
2. **Nolorbe-Payahua CD, de Freitas AS, Roesch LFW, Zanette J.** 2020. Environmental contamination alters the intestinal microbial community of the livebearer killifish Phalloceros caudimaculatus. Heliyon **6**:e04190.
3. **Malik, DS., Sharma, AK., Sharma, AK., Thakur, R. and Sharma, M.** 2020. A review on impact of water pollution on freshwater fish species and their aquatic environment. In: Advances in

Environmental Pollution Management: Wastewater Impacts and Treatment Technologies, Volume 1, Eds. Kumar, V., Kamboj, N., Payum, T., Singh, J. and Kumar, P., pp. 10-28, https://doi.org/10.26832/aesa-2020-aepm-02.

4. **Minich JJ, Härer A, Vechinski J, Frable BW, Skelton ZR, Kunselman E, Shane MA, Perry DS, Gonzalez A, McDonald D, Knight R, Michael TP, Allen EE.** 2022. Host biology, ecology and the environment influence microbial biomass and diversity in 101 marine fish species. 1. Nat Commun **13**:6978.
5. **Sullam KE, Essinger SD, Lozupone CA, O'connor MP, Rosen GL, Knight R, Kilham SS, Russell JA.** 2012. Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. Molecular Ecology **21**:3363–3378.
6. **Ruiz-Rodríguez M, Scheifler M, Sanchez-Brosseau S, Magnanou E, West N, Suzuki M, Duperron S, Desdevises Y.** 2020. Host Species and Body Site Explain the Variation in the Microbiota Associated to Wild Sympatric Mediterranean Teleost Fishes. Microb Ecol **80**:212–222.
7. **Hovda MB, Lunestad BT, Fontanillas R, Rosnes JT.** 2007. Molecular characterisation of the intestinal microbiota of farmed Atlantic salmon (Salmo salar L.). Aquaculture **272**:581–588.
8. **Ye L, Amberg J, Chapman D, Gaikowski M, Liu W-T.** 2014. Fish gut microbiota analysis differentiates physiology and behavior of invasive Asian carp and indigenous American fish. 3. ISME J **8**:541–551.
9. **Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, Prill RJ, Tripathi A, Gibbons SM, Ackermann G, Navas-Molina JA, Janssen S, Kopylova E, Vázquez-Baeza Y, González A, Morton JT, Mirarab S, Zech Xu Z, Jiang L, Haroon MF, Kanbar J, Zhu Q, Jin Song S, Kosciolek T, Bokulich NA, Lefler J, Brislawn CJ, Humphrey G, Owens SM, Hampton-Marcell J, Berg-Lyons D, McKenzie V, Fierer N, Fuhrman JA, Clauset A, Stevens RL, Shade A, Pollard KS, Goodwin KD, Jansson JK, Gilbert JA, Knight R.** 2017. A communal catalogue reveals Earth's multiscale microbial diversity. 7681. Nature **551**:457–463.
10. **MO BIO's PowerSoil DNA Isolation Kit Handbook - QIAGEN. https://www.qiagen.com/ca/resources/resourcedetail?id=5c00f8e4-c9f5-4544-94fa-653a5b2a6373&lang=en.** Retrieved 15 April 2023.
11. **MiSeq System | Focused power for targeted gene and small genome sequencing. https://www.illumina.com/systems/sequencing-platforms/miseq.html.** Retrieved 16 April 2023.
12. **NovaSeq 6000 System. https://www.illumina.com/systems/sequencing-platforms/novaseq.html.** Retrieved 16 April 2023.
13. **R: The R Project for Statistical Computing. https://www.r-project.org/.** Retrieved 15 April 2023.
14. **Wickham H, François R, Henry L, Müller K, Vaughan D** (2023). _dplyr: A Grammar of Data Manipulation_. R package version 1.1.0
15. **Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H.** 2019. Welcome to the Tidyverse. J Open Source Softw **4**:1686
16. **Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J.** 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome **6**:90.
17. **Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP.** 2016. DADA2: High-resolution sample inference from Illumina amplicon data. 7. Nat Methods **13**:581–583.
18. **Katoh K, Standley DM.** 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution **30**:772–780.
19. **Katoh K, Misawa K, Kuma K, Miyata T.** 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res **30**:3059–3066.
20. **Price MN, Dehal PS, Arkin AP.** 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE **5**:e9490.
21. **Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J.** 2018. Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome **6**:90.
22. **Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res **35**:7188–7196.
23. **McMurdie PJ, Holmes S.** 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE **8**:e61217.
24. **Chao A.** 1984. Nonparametric Estimation of the Number of Classes in a Population. Scandinavian Journal of Statistics **11**:265–270.
25. **Shannon C, Weaver W.** The Mathematical Theory of Communication.
26. **Bray JR, Curtis JT.** 1957. An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs **27**:325–349.
27. **Oksanen J, Blanchet FG, Kindt R, Legendre P, Minchin P, O'Hara B, Simpson G, Solymos P, Stevens H, Wagner H.** 2015. Vegan: Community Ecology Package. R Package Version 22-1 **2**:1–2
28. **Wilcoxon F.** 1945. Individual Comparisons by Ranking Methods. Biometrics Bulletin **1**:80–83.
29. **Anderson MJ.** 2017. Permutational Multivariate Analysis of Variance (PERMANOVA), p. 1–15. *In* Wiley StatsRef: Statistics Reference Online. John Wiley & Sons, Ltd.
30. **Wickham H.** 2016. ggplot2: Elegant Graphics for Data Analysis. Springer.

31. **Kassambara A** (2023). _ggpubr: 'ggplot2' Based Publication Ready Plots_. R package version 0.6.0
32. **Paradis E, Schliep K.** 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**:526–528.
33. **Love MI, Huber W, Anders S.** 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology **15**:550.
34. **Wald A.** 1943. Tests of Statistical Hypotheses Concerning Several Parameters When the Number of Observations is Large. Trans Am Math Soc **54**:426–482.
35. **She C-X, Lin H, Su Y-P, Zhang Y, Lin W-Z, Lan R-F.** 2021. [Diversity and community characteristics of organic phosphate-mineralizing bacteria in the sea area between Minjiang Estuary to Pingtan.]. Ying Yong Sheng Tai Xue Bao **32**:1863–1872.
36. **Priest T, Heins A, Harder J, Amann R, Fuchs BM.** 2022. Niche partitioning of the ubiquitous and ecologically relevant NS5 marine group. ISME J **16**:1570–1582.
37. **Kawano K, Ushijima N, Kihara M, Itoh H.** 2020. Patiriisocius marinistellae gen. nov., sp. nov., isolated from the starfish Patiria pectinifera, and reclassification of Ulvibacter marinus as a member of the genus Patiriisocius comb. nov. International Journal of Systematic and Evolutionary Microbiology **70**:4119–4129.
38. **Romanenko LA, Uchino M, Frolova GM, Mikhailov VV.** 2007. Marixanthomonas ophiurae gen. nov., sp. nov., a marine bacterium of the family Flavobacteriaceae isolated from a deep-sea brittle star. International Journal of Systematic and Evolutionary Microbiology **57**:457–462.
39. **Tsubouchi T, Koyama S, Mori K, Shimane Y, Usui K, Tokuda M, Tame A, Uematsu K, Maruyama T, Hatada Y.** 2014. Brevundimonas denitrificans sp. nov., a denitrifying bacterium isolated from deep subseafloor sediment. Int J Syst Evol Microbiol **64**:3709–3716.
40. **Golyshin PN, Martins Dos Santos VAP, Kaiser O, Ferrer M, Sabirova YS, Lünsdorf H, Chernikova TN, Golyshina OV, Yakimov MM, Pühler A, Timmis KN.** 2003. Genome sequence completed of Alcanivorax borkumensis, a hydrocarbon-degrading bacterium that plays a global role in oil removal from marine systems. J Biotechnol **106**:215–220.
41. **Nedashkovskaya OI, Vancanneyt M, Kim SB, Han J, Zhukova NV, Shevchenko LS.** 2010. Salinimicrobium marinum sp. nov., a halophilic bacterium of the family Flavobacteriaceae, and emended descriptions of the genus Salinimicrobium and Salinimicrobium catena. International Journal of Systematic and Evolutionary Microbiology **60**:2303–2306.
42. **Wegner C-E, Richter-Heitmann T, Klindworth A, Klockow C, Richter M, Achstetter T, Glöckner FO, Harder J.** 2013. Expression of sulfatases in Rhodopirellula baltica and the diversity of sulfatases in the genus Rhodopirellula. Marine Genomics **9**:51–61.
43. **Fonseca A, Espinoza C, Nielsen LP, Marshall IPG, Gallardo VA.** 2022. Bacterial community of sediments under the Eastern Boundary Current System shows high microdiversity and a latitudinal spatial pattern. Front Microbiol **13**:1016418.
44. **Rabus R, Ruepp A, Frickey T, Rattei T, Fartmann B, Stark M, Bauer M, Zibat A, Lombardot T, Becker I, Amann J, Gellner K, Teeling H, Leuschner WD, Glöckner F-O, Lupas AN, Amann R, Klenk H-P.** 2004. The genome of Desulfotalea psychrophila, a sulfate-reducing bacterium from permanently cold Arctic sediments. Environ Microbiol **6**:887–902.
45. **Nascimento JR, Easson CG, Jurelevicius D de A, Lopez JV, Bidone ED, Sabadini-Santos E.** 2020. Microbial community shift under exposure of dredged sediments from a eutrophic bay. Environ Monit Assess **192**:539.
46. **Mußmann M, Pjevac P, Krüger K, Dyksma S.** 2017. Genomic repertoire of the Woeseiaceae/JTB255, cosmopolitan and abundant core members of microbial communities in marine sediments. ISME J **11**:1276–1281.
47. **Xie S, Wang S, Li D, Shao Z, Lai Q, Wang Y, Wei M, Han X, Jiang L.** 2019. Sulfurovum indicum sp. nov., a novel hydrogen- and sulfur-oxidizing chemolithoautotroph isolated from a deep-sea hydrothermal plume in the Northwestern Indian Ocean. Int J Syst Evol Microbiol **71**.
48. **Sanz JL.** 2011. Sulfur Cycle, p. 1618–1619. In Gargaud, M, Amils, R, Quintanilla, JC, Cleaves, HJ (Jim), Irvine, WM, Pinti, DL, Viso, M (eds.), Encyclopedia of Astrobiology. Springer, Berlin, Heidelberg.
49. **Sievert S, Kiene R, Schulz-Vogt H.** 2007. The Sulfur Cycle. Oceanography **20**.
50. **Geesink P, Taubert M, Jehmlich N, von Bergen M, Küsel K.** 2022. Bacterial Necromass Is Rapidly Metabolized by Heterotrophic Bacteria and Supports Multiple Trophic Levels of the Groundwater Microbiome. Microbiol Spectr **10**:e0043722.
51. **Lin L, Yang H, Xu X.** 2022. Effects of Water Pollution on Human Health and Disease Heterogeneity: A Review. Frontiers in Environmental Science 10.
52. **Austin B.** 1998. The effects of pollution on fish health. J Appl Microbiol 85 Suppl **1**:234S-242S.
53. **Yang H-X, Wang X, Liu X-W, Zhang J, Yang G-Q, Lau KWK, Li S-P, Jiang J-D.** 2014. Fluviicola hefeinensis sp. nov., isolated from the wastewater of a chemical factory. International Journal of Systematic and Evolutionary Microbiology **64**:700–704.
54. **Stefanakis A, Zouzias D, Marsellos A.** 2015. Groundwater Pollution: Human and Natural Sources and Risks, p. 82–102.
55. **Wen B, Liu J-H, Zhang Y, Zhang H-R, Gao J-Z, Chen Z-Z.** 2020. Community structure and functional diversity of the plastisphere in aquaculture waters: Does plastic color matter? Science of The Total Environment **740**:140082.