POPULATION GENETIC DIVERGENCE AND ENVIRONMENT INFLUENCE THE GUT MICROBIOME IN OREGON THREESPINE STICKLEBACK

by

ROBERT A. STEURY

A THESIS

Presented to the Department of Biology
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

March 2019

THESIS APPROVAL PAGE

Student: Robert A. Steury

Title: Population Genetic Divergence and Environment Influence the Gut Microbiome in Oregon Threespine Stickleback

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Biology by:

Brendan Bohannan	Chairperson
William Cresko	Co-Chairperson

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2019

THESIS ABSTRACT

Robert A. Steury

Master of Science

Department of Biology

March 2019

Title: Population Genetic Divergence and Environment Influence the Gut Microbiome in Oregon Threespine Stickleback

Studying the microbiome in natural populations could improve our understanding of the biological factors that influence microbiome variation. If host genetic variation is important in microbiota assembly, then understanding genetic divergence among natural populations could be informative. Despite advances in sequencing technology, we have not yet taken full advantage of this technology in natural populations. Here we integrate genome-wide population genomic and microbiome analyses in wild threespine stickleback (*Gasterosteus aculeatus*) fish distributed throughout western Oregon, USA. We found that gut microbiome varied in diversity and composition more among than within wild host populations. Furthermore, this among population variation was better explained by host population genetic divergence than by environment and geography. We also identified a subset of gut microbial taxa that were most strongly sorted both across environments and across genetically divergent populations. We believe this study contributes generalizable methods and findings in host systems. This thesis includes supplemental materials.

CURRICULUM VITAE

NAME OF AUTHOR: Robert A. Steury

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Lane Community College, Eugene

DEGREES AWARDED:

Master of Science, Biology, 2019, University of Oregon
Bachelor of Science, Biology, 2013, University of Oregon

AREAS OF SPECIAL INTEREST:

Microbiome
  Host-associated
Ecology

PROFESSIONAL EXPERIENCE:

Research Assistant, University of Oregon, 2012-2018

ACKNOWLEDGMENTS

I wish to express sincere appreciation to Professors Bohannan and Cresko for their generosity throughout this work and for their assistance in the preparation of this manuscript. In addition, special thanks are due to Mark Currey, whose expertise in conducting research in wild threespine stickleback populations in Oregon were vital in this undertaking. I also thank the members of the Bohannan and Cresko lab groups for their valuable input. Lastly, I thank my family for their support in this endeavor. The investigation was supported National Institute of General Medical Sciences and the National Center for Research Resources of the National Institutes of Health under award numbers P50GM098911 (B.J.B and W.A.C.) and R24RR032670 (W.A.C).

For my children, Elijah and Elliana, and for future generations.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figures Page

LIST OF TABLES

LIST OF SUPPLEMENTAL FIGURES

Supplemental Figures

CHAPTER I: INTRODUCTION

Microbiome research is now a core area of research in ecology and evolution[1], which is partly due to an expansion in our understanding of the importance of the microbiome in animal and plant fitness[2]. Recent microbiome research has focused in the vertebrate gut for several reasons. The densely colonized and metabolically diverse gut microbiome can enhance host nutrition[3] and signal host development[4], including the onset of host immunity[5]. Furthermore, a normal gut microbiome is important in preventing pathogen infection and host diseases, such as IBD and colorectal cancer[6,7,8]. These features of the gut microbiome make it appealing to modulate in order to benefit host fitness.

Despite this work, we still do not fully understand how host and non-host variables relate to the structure and function of an individual's microbiome[9]. The relationship between host genetic and microbiome variation, for example, is poorly understood[10]. During the earliest stages of development, most vertebrate hosts are essentially sterile. Hosts enter a world dominated by microbes[11]. They then can be colonized by microbes that are present in the local source pool[12]. This initial colonization can influence the organism's microbiome throughout its life[13]. Variation in environmental factors can predict microbial community structure across space[14,15]. For instance, salinity gradients and fluctuations in temperature and light are influential factors in aquatic habitats[16,17]. Therefore, a common hypothesis is that the primary factor influencing microbiome variation across hosts is the environment. Indeed, studies have documented that microbiome can co-vary with the spatial patterns of environmental factors[18,19]. However, in some cases microbial taxa found associated with the host are not found its environment, suggesting that the host and stochastic processes must also play roles.

Given the extensive effects the microbiota can have on the host, it is not surprising that the host has evolved to modulate the microbiota. The innate and adaptive immune systems evolve to monitor and alter the abundances of microbes[20]. Like other organismal traits these can be genetically determined, and therefore can be underpinned by genetic variation. For instance, this is inferred from observations of divergence of major gut bacterial phyla between human and animals[21] and microbiome similarities among genetically related humans[10,22–24]. In addition, induced mutations in laboratory animal

populations has resulted in evidence that host immunity genes, such as myD88(-/-), NOD2(-/-), ob/ob, and Rag1(-/-), can directly influence the presence and abundances of microbes. Moreover, microbial quantitative-trait loci mapping (mbQTL) and genome-wide microbial association studies conducted in mice[25,26] and humans[27] revealed that important underlying host genetics are complex and possibly heritable. We do not yet understand the full complexity of these traits, nor how environmental variation and segregating host genetics (as compared to induced mutations) interact to modulate the microbiome in wild hosts.

Advances in understanding how host genetic variation and environmental variation produce microbiome differences among hosts could require applying modern genetic tools and analyses in wild hosts in their native environments. In these environments the processes of ecology and evolution that have taken millions of years maintain their influence in the genomes and microbiomes of wild hosts. In contrast, most work in this area has been conducted in laboratory settings. We have narrowed our study of the host microbiome by studying domesticated hosts. However, this may have come at the cost of missing potentially important variation that is lost due bottlenecks and genetic drift within facilities. We have seemingly skipped the importance of studying the microbiome in the wild, in order to generate hypotheses which to study experimentally in facilities. To achieve this aim, we can start by taking full advantage of modern genome sequencing technology (e.g. RADseq)[28]. These tools allow for the fine scale determination of the relationship among individuals and populations in the wild through measurements of genome-wide genetic diversity using next generation sequencing approaches[29]. These could then be used, for example, to precisely quantify the genetic distance among populations that can be compared to microbiome differences, geographic differences, and differences among microbiomes. This should allow us to determine the relative importance of these variables in microbiome variation among hosts.

An excellent model for studying the combined influences of environment and genetics on microbiome variation in the wild is the threespine stickleback fish (Gasterosteus aculeatus)[30]. This small teleost fish has a thoroughly studied evolutionary and ecological history in numerous locations throughout the Holarctic[30]. Stickleback is abundant and

nearly ubiquitous in coastal regions of the Northern Hemisphere where it is found in marine, anadromous, and freshwater forms that themselves are phenotypically and genetically diversified. A fully sequenced and annotated genome and population genomic tools permit broad genomic sequence resolution and powerful inference of relatedness among populations and individuals[31]. Moreover, this species is amenable to laboratory manipulations and is tractable for microbiome research. For instance, differences in gene transcription have been compared between "gnotobiotic" stickleback (with a known microbiome) raised in either conventional or germ-free lab environments[32]. Gnotobiotic oceanic and freshwater stickleback have also been tested for varying immune response to microbiome exposure[33]. In addition, wild stickleback populations have been used to study how environment and host diet, sex, immune genes, and population level genetic relatedness and diversity relate to the gut microbiome[34–37]. Here we built upon this foundational work by deeply sequencing both the host genome and the gut bacterial communities associated with genetically divergent wild populations of stickleback drawn from different natural environments (estuary and freshwater). We found that microbiome variation was most evident among individual fish rather than among fish populations, across sampling periods, and between environments. Despite this overwhelming interindividual variation, microbiome composition was better predicted by fish population genetic divergence and microbiome diversity was better predicted by among population average genetic distance, rather than geographic distance and environment. Furthermore, we found evidence that these differences among populations could be the result of differences in the relative abundance of a fraction of the whole gut microbiome in these fish.

CHAPTER II: RESULTS

Stickleback population genetic structure in Oregon was partitioned in concordance with geography and environment.

We selected six threespine stickleback populations for their diversity of environments, population genetics, and geographic distribution throughout Oregon (Table 1 and Figure 1). In order to confirm and quantify the previously observed genetic structure of these populations we used PCA and STRUCTURE analyses of genetic variation in a subset of 391 random RADseq SNPs. This dataset included RADseq data generated from four populations that were previously analyzed[31], and two for which collections were made and new data generated in the present study (Lynx Hollow and Eel Creek).

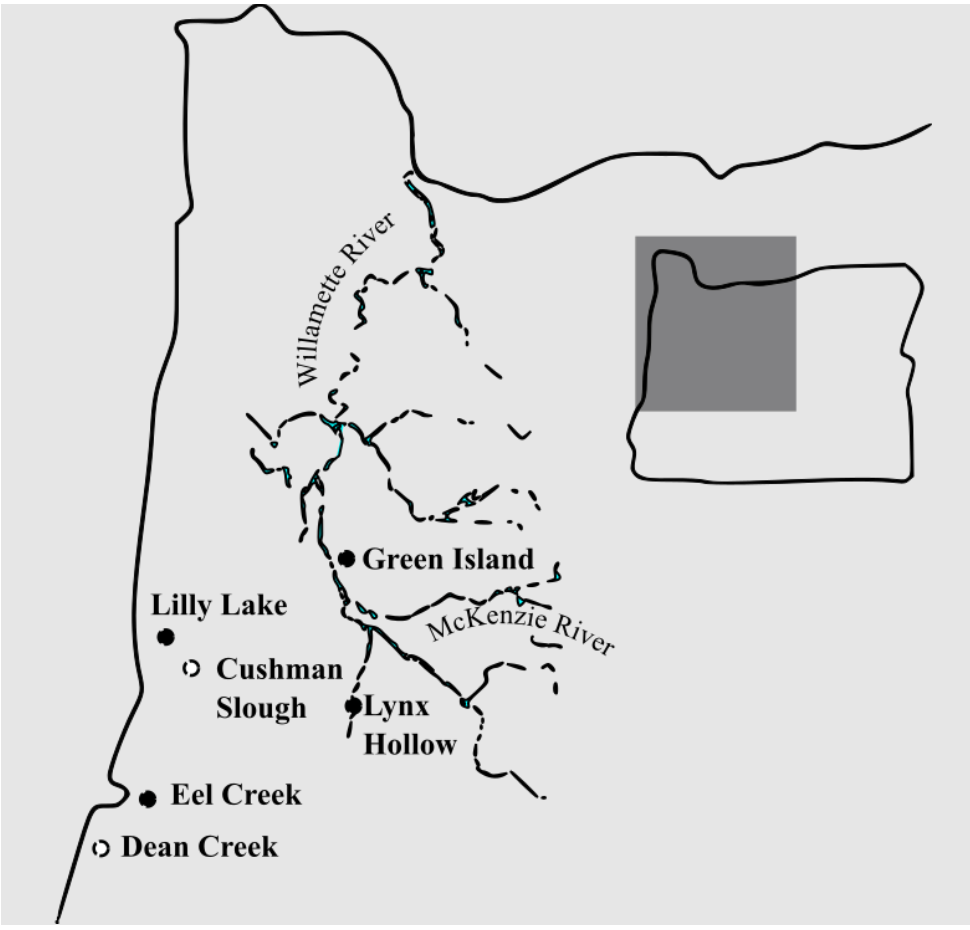

Figure 1. Site map in Oregon. Filled and open points represent freshwater and estuary; respectively. Two inland freshwater populations and four coastal populations were sampled.

Site	Region	Environment	Location	Trap Date	# Guts
Dean Creek	Coast	Estuary	N 43.6949, W 124.0432	06/13/17	20
Eel Creek	Coast	Freshwater	N 45.5869, W124.1861	06/13/17	20
Cushman Slough	Coast	Estuary	N 43.9881, W124.0395	07/29/15	22
Lilly Lake	Coast	Freshwater	N 44.0886, W124.1140	08/04/15	22
Green Island	Inland	Freshwater	N 44.1582, W 123.1189	07/24/15	22
Lynx Hollow	Inland	Freshwater	N 43.8613, W 123.0249	07/03/17	20

Table 1. Oregon fish collection details.

We found a major partition in population genetic structure between fish found inland in the Willamette Basin and fish found in coastal basins, confirming previous findings[31]. Inland and coastal populations grouped separately along the first principal component (PC1), which accounted for 46.33% of the overall genetic variation (Figure 2). We also confirmed that coastal populations grouped separately by environment (estuary vs. freshwater) along PC2, which accounted for 8.56% of overall genetic variation (Figure 2). In congruence with these PCA results, average population genetic divergence, as measured by F_{ST} in pairwise comparisons between coastal and inland fish populations, was on average two-fold greater (~0.242) than between inland (~0.125) and four-fold greater than among coastal populations (~0.063) (Table 2).

	Eel Creek	Cushman Slough	Lilly Lake	Green Island	Lynx Hollow
Dean Creek	0.1293	0.0212	0.0698	0.3117	0.3942
Eel Creek		0.0919	0.0793	0.4709	0.5748
Cushman Slough			0.0653	0.2392	0.2789
Lilly Lake				0.3028	0.3661
Green Island					0.1488

Table 2. Population genetic divergence (F_{ST}) among six threespine stickleback populations in Oregon based on RADseq SNPs.

We next wanted to determine the degree to which these populations shared ancestry, or represented admixture of genotypes, by assigning posterior probabilities of assignment to different groups. We suspected that populations sharing ancestry might also share gut microbiome features in common. We utilized Structure to perform analyses across a range of hypothesized groupings. Using *Ln P(D)* and the deltaK method[38] we found that a model with *K* =2 best fits the full genetic dataset which further confirmed the inland-coastal genetic structure found along PC1 (Supplementary Figure 1). Subgrouping both inland and coastal populations at *K*=6 also corroborated the population groupings found along PC2 (Supplementary Figure 1).

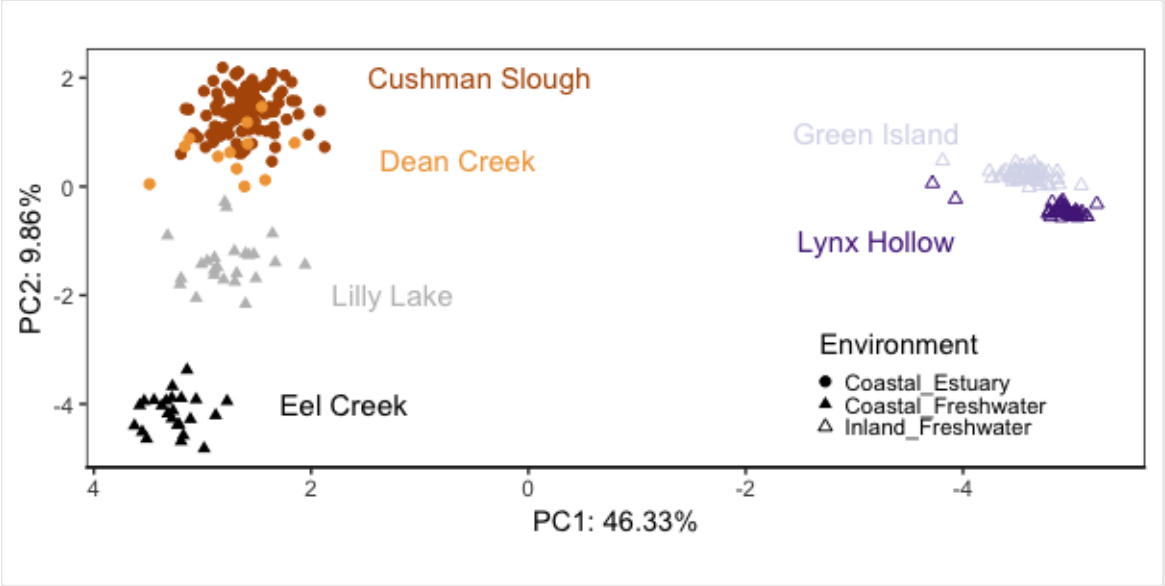

Figure 2. Threespine stickleback population genetic structure partitioned between inland and coastal sites, and between estuary and freshwater sites on the coast. This biplot includes the first two PC axes which accounted for over half (56.19%) of the overall variation in 391 randomly selected genome-wide SNPs for the six populations in this study. Each point represents a fish. Shape fill represents whether sites were in the Willamette Basin ("Inland") or in watershed along the coast ("Coastal"). Colors represent collection sites ("populations") in Oregon.

Stickleback gut microbiome diversity and composition was better predicted by population genetic divergence than by environmental and geographic differences.

Using 16s rDNA amplicon sequences variants (ASVs) collected in these populations, we next documented microbiome variation and compared it to population genetic and environmental variation. We found that gut communities varied in terms of both alpha and beta diversity among individual fish, as well among populations on average (Figure 3 and 4). Coastal populations tended to have greater inter-individual and lower individual gut community diversity on average.

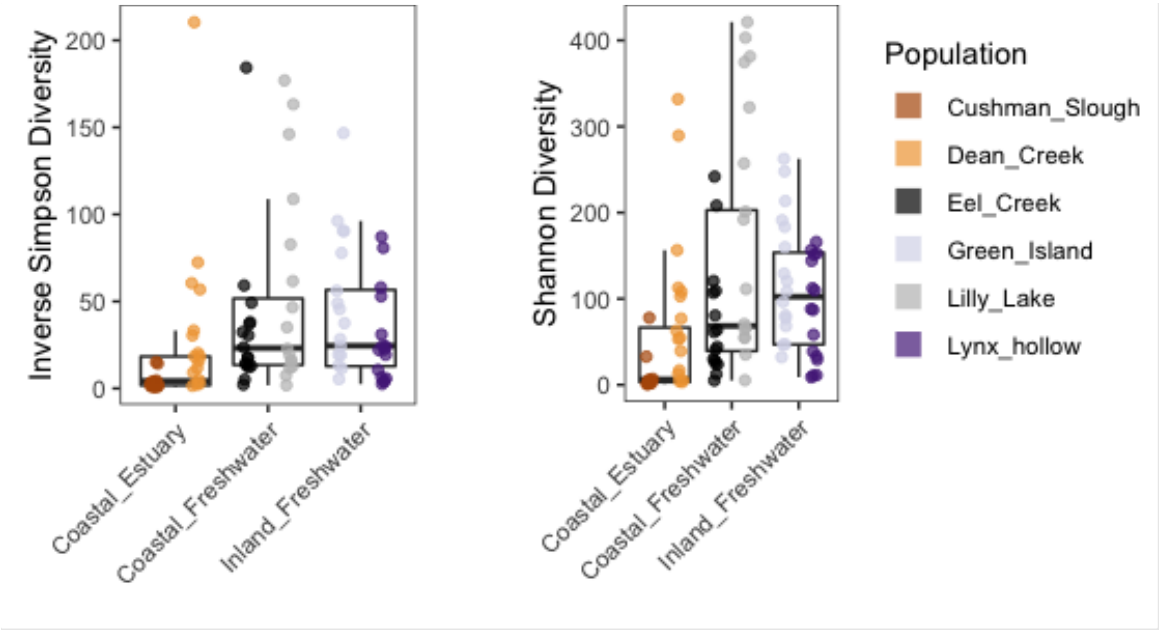

Figure 3. Gut microbiome alpha diversity among threespine stickleback populations in Oregon in terms of inverse Simpson diversity (left) and Shannon diversity (right) measures using rarefied taxa abundance (1000/sample). Colors represent collection sites ("populations"). Each point is a fish gut. Mid-box lines are pooled means of major habitats (e.g. Coastal Freshwater"). Box whiskers are pooled standard deviation of major habitats as well.

We compared microbiome variation to genetic divergence (F_{ST}) predicted patterns and environment using a linear mixed model (LMM) approach. We included sample period in order to account for potential effects caused by collecting samples at different times, in this case two years apart. To test how relatively well these factors predicted gut ASV diversity among fish populations we used inverse Simpson and effective Shannon diversities (alpha), and both weighted and unweighted UniFrac distance-to-centroid metrics (beta), as response variables in this LMM.

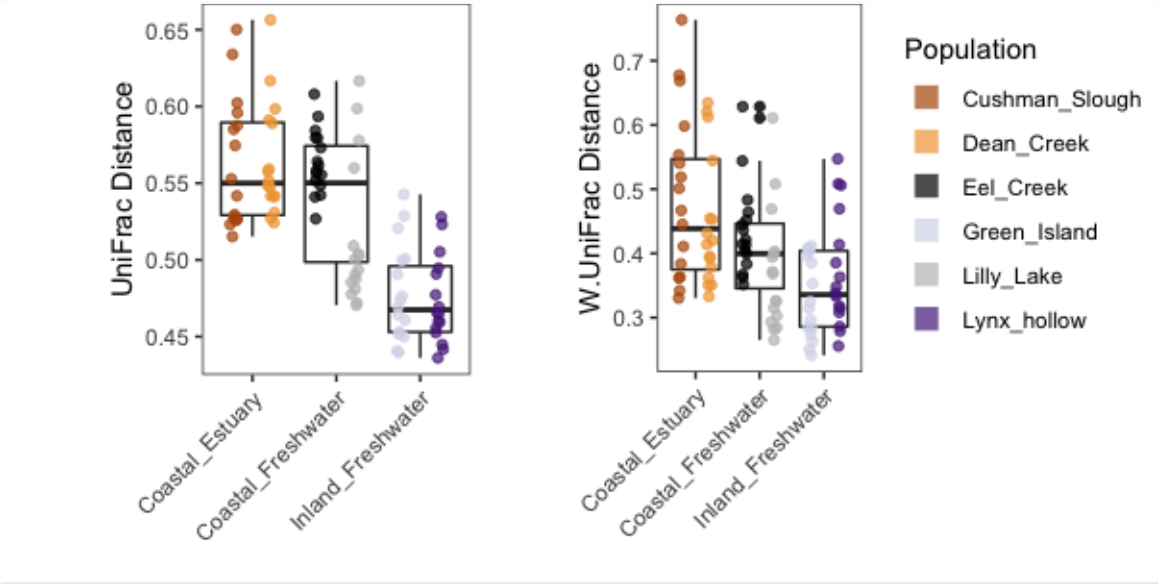

Figure 4. Gut microbiome beta diversity among threespine stickleback populations in terms of distance-to-centroid transformed weighted (left) and unweighted UniFrac (right) distance of rarefied ASV counts (1000/sample). Colors represent collection sites ("populations"). Each point is a fish gut. Mid-box lines are pooled means of major habitats (e.g. Coastal Freshwater"). Box whiskers are pooled standard deviation of major habitats as well.

We found that gut microbiome Simpson diversity was better predicted by population F_{ST} ($\chi 2 = 19.07$, $df = 1$, p <1e-4) than by environment ($F = 4.65$, $df = 1$, $p = 0.10$) and sample period ($\chi 2 = 0$, $df = 1$, $p = 1$), with neither environment nor sample period statistically

significant. Similarly, Shannon diversity was better predicted by F_{ST} ($\chi 2 = 5.30$, $df = 1$, $p = 0.02$) than by environment ($F = 3.48$, $df = 1$, $p = 0.16$) and sample period ($F > 0.99$, $df = 1$, $p = 0.93$) (Figure 3 and Table 4). In addition, we found that F_{ST} both in terms of unweighted ($\chi 2 = 10.20$, $df = 1$, $p < 0.0014$) and weighted ($\chi 2 = 3.16$, $df = 1$, $p = 0.07$) UniFrac distance to centroid better predicted gut beta diversity than environment (w.UniFrac: $F = 3.90$, $df = 2$, $p = 0.15$ and UniFrac: $F = 1.48$, $df = 2$, $p = 0.40$) and sample period t (w.UniFrac: $F <$ 1e-13, $df = 1$, $p > 0.99$ and UniFrac: $F = 0.39$, $df = 2$, $p = 0.53$) (Figure 4 and Table 3). The random effects of sample period were completely negligible for all diversity metrics we tested (Table 3).

populations in Oregon based on RADseq SNPs.

PERMANOVA	R^2	*df*	*p* **value**
Population F_{ST}	0.068	1	0.001
Environment	0.061	1	0.001
Sample Period	0.061	1	0.001
River Miles	0.049	1	0.001
Geographic Distance	0.046	1	0.001

LMM	**Simpson**			**1/Shannon**			**UniFrac**			**W. UniFrac**		
Fixed	*F*	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*	*F*	*df*	*p*
Environ.	4.65	1	0.10	3.48	1	0.16	1.23	1	0.347	6.29	1	0.07
Random	$\chi 2$			$\chi 2$			$\chi 2$			$\chi 2$		
Pop. F_{ST}	16.93	1	<1e-04	5.30	1	0.02	17.59	1	<1e-04	4.61	1	0.03
Sample	<1e-13	1	1	>0.99	1	0.93	0.393	1	0.531	<1e-13	1	>0.99

Table 3. PERMANOVA (in order of proportion explained) and LMM statistics results. Significant results in bold font where $p < 0.05$. Gut microbiome composition (top) and diversity (bottom).

After finding that microbiome diversity among populations was best explained by population genetic divergence, we wanted to know if overall gut community composition varied among fish and what relative degree it was predicted by several factors. We first needed to determine if ASVs covaried among fish gut communities in order to choose the

appropriate statistical approach. We ordinated relative abundance and square root transformed Bray-Curtis dissimilarity in a PCoA and found that ASVs did not covary strongly (Figure 5 and Supplementary Figure 2). Similar results were found using both weighted and unweighted UniFrac distance. These results contrasted with the robust explanatory power of the first two axes in the PCA of population genetic structure. Instead of using the first PCoA axes as response variables in the same LMMs for hypothesis testing, we instead used a PERMANOVA to test the relative ability of population genetic divergence, environment, geography, and sample period to predict the overall variation in the Bray-Curtis dissimilarity matrix. Nonetheless, the first PCo axis scores appeared to somewhat correlate on average with both environment and geography when considered together (Figure 6).

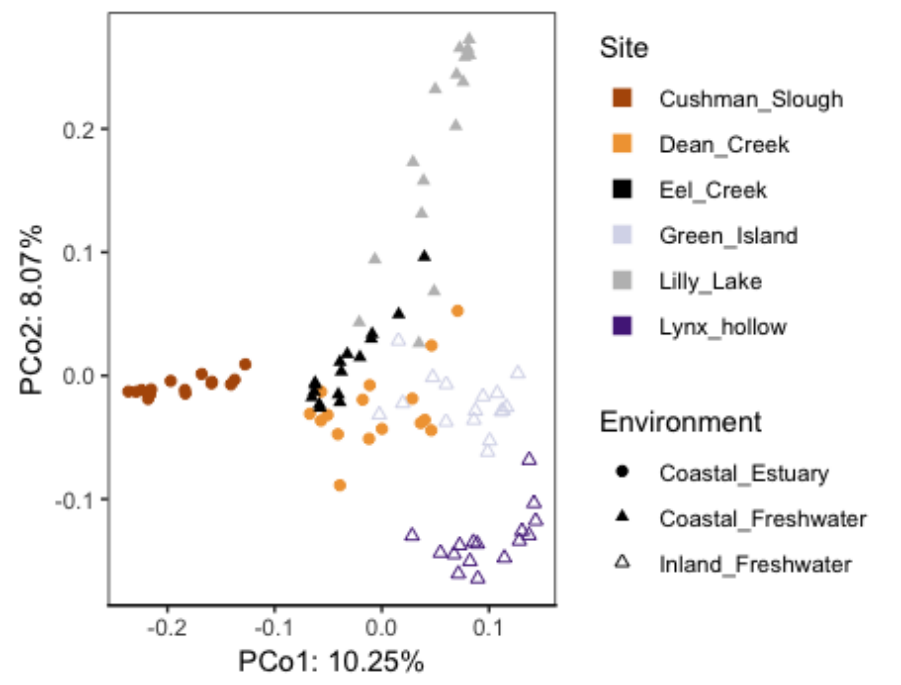

Figure 5. PCoA ordination of a Bray-Curtis dissimilarity matrix using square root and relative abundance transformed ASV counts. PCoA axes 1-2 are plotted here. Each point represents a fish gut microbial community. Shape fill represents whether fish were in the Willamette Basin ("Inland") or in watershed along the coast ("Coastal"). Colors represent collection sites ("populations") in Oregon.

We found that population genetic divergence explained the most variation relative to the other factors (Table 4). Furthermore, population F_{ST} (R^2 =.068, $p < 0.001$), environment (R^2 =.061, $p < 0.001$), sample period (R^2 =.061, $p < 0.001$), and both river mile (R^2 =.049, $p < 0.001$) and geographic distance (R^2 =.046, $p < 0.001$), explained near equal portions of variation in gut microbiome in terms of Bray-Curtis dissimilarity among fish gut communities. F_{ST} covaried with both geographic distance (Pearson's product-moment correlation = 0.97, t =39.31, df = 94, p <1e-16) and river miles (Pearson's product-moment correlation = 0.94, t =26.0, df = 94, p <1e-16). The remaining majority of gut microbiome variation among these six wild fish populations remained unexplained by the factors we accounted for in this study (~76%).

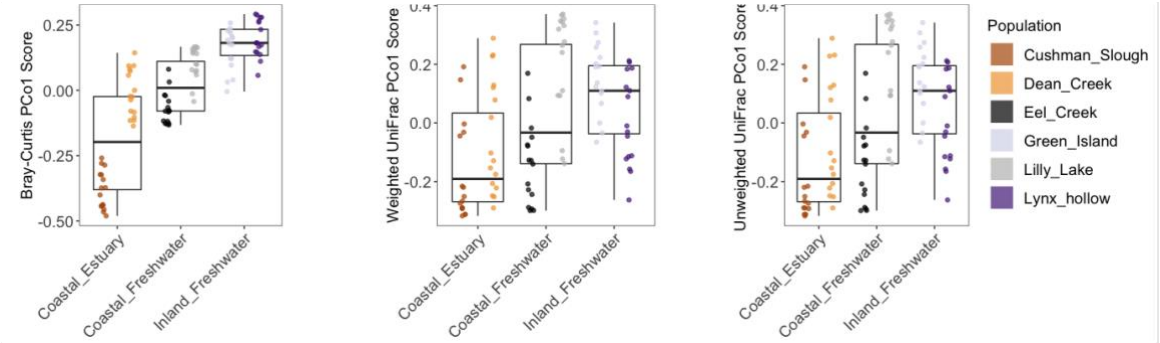

Figure 6. Gut microbiome composition among threespine stickleback in Oregon varied in terms of PCoA Bray-Curtis dissimilarity (left), as well as both Weighted (middle) and Unweighted UniFrac (right) distance. Only PCoA axis 1 is shown here for all three approaches. Colors represent collection sites ("populations"). Mid-box lines are pooled means of major habitats (e.g. Coastal Freshwater"). Box whiskers are pooled standard deviation of major habitats as well.

Differential ASV abundance among gut microbiomes was due to just a small fraction of total bacteria present

We next wanted to determine how gut microbiome compositional variation among threespine stickleback in Oregon was influenced by host population genetic structure and environment. We first looked broadly at what bacterial groups were present in our gut samples. We found that the stickleback microbiome in Oregon consisted of predominately bacteria in the phyla Proteobacteria, Fermicutes, Bacteroides, Actinobacteria, Acidobacteria, Planctomycetes, Verrucomicrobia, Chloroflexi, Tenericutes, and Fusobacteria (ordered by most to least abundant) based on relative abundance in the metacommunity (Supplemental Figure 3). These groups are found in the gut in other wild fish surveys[37,39–41].

We next looked at microbiome composition at a finer taxonomic resolution by first quantifying ASV abundance differentiation across the major geographic partition in host population structure (Figure 2 and Supplemental Figure 1). We tested ASVs for significant differential abundance between only coastal and inland freshwater populations to reduce the effects of environment (see methods). Significant differential abundance was found in a subset of 1,100 of the total 16,530 bacterial ASVs and 16 of the total 23 phyla we documented (Figure 7 and Table 5). Nearly half of these ASVs (46%) comprised genera in the phylum Proteobacteria. Single ASVs were exceptionally enriched (log2fold > 20) in fish guts in the Proteobacteria (inland), and on the Fermicutes (coast). Nine phyla were enriched in fish guts both in inland and coastal populations. In contrast, Tenericutes were enriched only in inland freshwater fish, and Fusobacteria, Spirochaetes, Nitrospirae, Ignavibacteriae, Lentisphaerae, and Cloacimonetes were enriched only in coastal freshwater fish (Figure 7). The Tenericutes in inland fish guts comprised one ASV whose sequence most closely matched (95%) a human pathogen species, *Mycoplasma penetrans*[39,42,43]. Each phylum enriched exclusively in coastal freshwater fish guts was represented by a single ASV as well, the exception being the Spirochaetes, which was represented by two ASVs. A couple of these match most closely with bacteria species associated with host pathogenicity (*Fusobacterium varium* and

Treponema spps.), but the rest were taxa typically associated with lake and sludge communities.

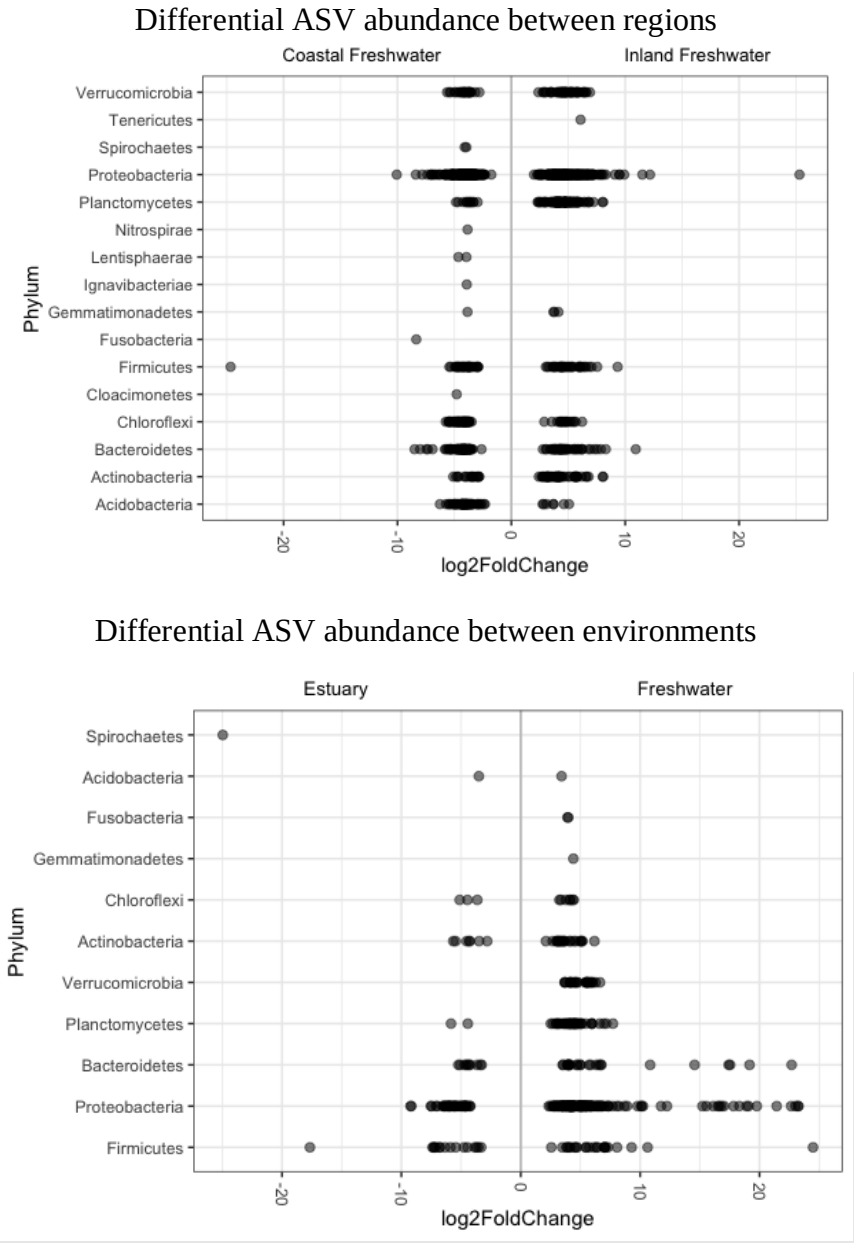

Figure 7. Significant differential ASV abundance in fish gut microbiome was determined between inland and coast (top) and freshwater and estuary (bottom) environments in Oregon. Each point is an ASV that is enriched in a phylum (left axis). Negative and positive log2Fold values are enrichments in estuary and freshwater fish guts; respectively.

We similarly tested how gut microbiome taxa were differentially enriched among coastal and inland stickleback populations using the full dataset. Significant differential abundance was found in a subset of 404 ASVs comprising 11 phyla (Figure 7). Over half (53%) of these enrichments comprised genera in the phylum Proteobacteria. There were 8 ASVs that were exceptionally enriched (log2fold > 20) in estuary fish (1 Spirochaetes) and freshwater (5 Proteobacteria, 1 Bacteroides, and 1 Firmicutes). Seven phyla were enriched in gut microbiomes in both environments. In contrast, the Spirochaetes were enriched in only estuary fish and the Fusobacteria, Gemmatimonadetes, and Verrucomicrobia, were enriched in only freshwater fish (Figure 7). The Spirochaetes enrichment in estuary fish comprised a single ASV that didn't closely match any known species, but classified to the Brevinema, a genus which consists of bacteria detected in fish guts in other studies[44,45]. Like we found in inland freshwater fish, the phyla that were enriched in only freshwater fish comprised a couple of ASVs that matched most closely with bacteria species associated with host pathogenicity (*Fusobacterium varium* and *Treponema* spps.), the rest were not well matched with bacteria that had known origins.

CHAPTER III: METHODS

Sample collections

Between June and August in 2015 and 2017 we collected 126 wild, male and female, adult threespine stickleback fish in western Oregon, USA. Collections included ≥18 fish from each of six sites varying in distance apart from one another among four major watersheds (Figure 1 and Table 1). Two freshwater sites, Lilly Lake and Eel Creek, were located over 80 river miles apart on the coast. Cushman Slough and Dean Creek estuaries were over ~50 river miles apart, and each adjacent to the freshwater sites (within ~25 river miles) on the coast. These coastal sites comprised two different watersheds, the Umpqua (included Dean Creek and Eel Creek) and the Siuslaw (included Cushman Slough and Lilly Lake). We also collected stickleback at two freshwater sites, Green Island and Lynx Hollow, located over 430 river miles inland comprising the Middle Fork Willamette and Coastal Willamette watersheds, respectively. All collections were carried out as in Catchen et al. (2013), according to approved University of Oregon IACUC protocols, and under the auspices appropriate state and federal collecting permits.

Gut isolation and soma preservation for genomic analysis

Fish were euthanized, gut-dissected, and guts were stored in liquid nitrogen in the field immediately following removal from traps, until transfer to -80˚C lab storage. In the 2017 collection, whole fish were preserved in 100% ethanol following dissections. Salinity (via refractometer) and temperature at each site were measured at the time of sampling.

Fish DNA extraction

Fin clips (~20 mg) from ethanol-stored whole fish were DNA extracted following the Qiagen Blood and Tissue Kit protocol. Quantification of DNA concentration was performed on a Qubit fluorometer using the dsDNA broad range assay kit. DNA dilution plates for each population were created by diluting to a concentration of 25 ng/ul using EB in 96 deep-well format.

Host DNA extraction and sequencing

RADseq libraries were created as in Catchen et al. (2013). Briefly, for each population sequenced, endonuclease *SbfI-HF* (New England Biolabs) was used to cut DNA at specific locations spread throughout the genome. P1 adaptors with unique barcodes were ligated to the digested DNA of each individual fish within each population so that sequences from all that could be bioinformatically recovered. Individual samples were then pooled and sheared. Sheared fragments of 300 – 500 bp in length were size selected and a second adapter, P2, was ligated to the fragments. Primers specific to sequences located on the P1 and P2 adaptors were used to perform a PCR reaction to enrich only for fragments with both a P1 and P2 adaptor in 12 cycles of amplification. The resulting amplified library was size selected by gel electrophoresis for fragments of 388 – 520 bps and then these fragments were sequenced using an Illumina HiSeq4000 at the University of Oregon Genomics Core Facility.

Stacks software package was utilized for processing and analyzing the RAD-seq data. We de-multiplexing the libraries by barcode and filtered RAD reads with < 90% probability of being correct (based on phred score) using the process_radtags function in the *Stacks* software pipeline[46]. Processed RAD tags were then aligned against the stickleback genome using GSNAP[47]. Aligned tags were then run through the Stacks pipeline using rxstacks to identify and call SNPs throughout the genome and create a catalog of SNPs from all populations included. Genomic reads that aligned 100% to stickleback mitochondrial DNA were removed using bowtie/2.2.9[48].

Stickleback population genomic analyses

Genome-wide estimates of average relatedness (F_{ST}) were calculated by averaging overall variable and non-variable restriction-site SNPs using the populations program within the *Stacks* software framework[46]. 1000 randomly sampled SNPs were generated using a Perl script as described in the *Stacks* manual to reduce the data to computationally manageable size (391 after filtering). To visualize population structure and substructure

in this subsample, broad-scale analyses were performed with all six of the Oregon populations using principal component analyses (PCA) in **R** version 3.4.2 [49] and STRUCTURE analysis[50]. STRUCTURE analysis was performed using a burn-in of 20,000 steps, and with 20,000 replicates for each value of *K*, where *K* = number of groupings. The number of *K*s tested was equal to the number of populations (*K*=6). Each value of *K* was replicated 10 times and the Evanno method using Harvester[38]. Visual inspection of the change in Ln P(D), and biological relevance was used to determine the level of *K* that best fit the data.

Microbiome DNA extraction and 16S sequencing

Frozen guts were weighed, rapidly thawed, and samples were immediately processed for DNA extraction using a modified Dneasy Blood and Tissue Kit protocol. Guts were transferred to preheated Qiagen PowerBead Solution in QIAmp PowerFecal DNA Kit garnet bead lysis tubes and homogenized on a FastPrep120. A volume of each homogenate equivalent to 20 mg of frozen gut was transferred to preheated Qiagen ATL Buffer in 1.5ml RINO screw-cap tubes containing 100 uL ZrOB015 beads (Next Advance) and homogenized again. We added both a ThermoFisher proteinase K and a Rnase step to remove reduce DNA degradation and degrade RNA in samples. Dneasy Blood and Tissue Kit protocol (Qiagen #69504). DNA was quantified using Invitrogen Qubit Broad Range Assay and stored at -20˚C until sequencing. Total yields of gDNA were ~2 ng on average consistent with other fish gut microbiome reports (sample average = 176.88 ng/µl gDNA).

DNA libraries were fully processed and sequenced in the Genomics and Cell Characterization Core Facility (GC3F) at the University of Oregon. PCR was performed with barcoded V4 bacterial 16S ribosomal RNA primers (515F *GTGCCAGCMGCCGCGGTAA*, 806R *GGACTACHVGGGTWTCTAAT*). Our PCR ran 27 cycles with an annealing temperature of 61˚C using Q5 Hot Start HiFi PCR master mix (NEBNext). Abundant genomic template and primers that remained in the PCR product after PCR were reduced using standard magnetic bead PCR cleanup methods.

Sequencing and adonis multiplexing followed default Illumina paired-end library protocols.

Microbiome DNA sequence processing using DADA2 pipeline

A total of 38,29,364 high quality raw sequence reads were imported into **R** version 3.4.2. Processing using **DADA2** version 1.6[51] with default parameters except where we specify "()," resulted in 21,962,537 non-chimeric merged reads, or 198,835 reads per gut on average (full details in Table 2). Paired-end 150bp reads were trimmed (f=145 and r=140), denoised, merged, and truncated (reads between 250-257 bp were kept). We constructed a table of 100% unique ASVs from these processed reads, removed chimeras, and assigned taxonomy to the resulting 26,506 ASVs using the Ribosomal Database Project classifier (doi: http://dx.doi.org/10.1128/AEM.00062-07).

Site	Input	Filtered	Denoised	Merged	Tabled	Nonchimeric
Cushman Slough	9,256,421	4,888,393	4,888,393	4,298,407	4,298,407	4,264,096
Dean Creek	3,359,145	3,069,584	3,069,584	2,521,985	2,521,985	2,204,735
Eel Creek	2,287,323	2,055,175	2,055,175	1,905,651	1,905,651	1,663,629
Lilly Lake	10,541,264	7,072,180	7,072,180	6,777,026	6,777,026	6,721,944
Green Island	8,616,378	4,260,381	4,260,381	3,699,754	3,699,754	3,681,706
Lynx Hollow	4,232,833	3,939,422	3,939,422	3,624,234	3,624,234	3,426,427

Table 5. Step-wise number of reads in each population during DADA2 processing of 16s bacterial sequence reads.

Microbiome ASV table processing and merge into Phyloseq dataset

To reduce potential confounding from variables that we were not interested in measuring in this study, we removed sequences from our ASV table that classified either as Archaea or chloroplast DNA, as well as sequences found in our negative controls (8,097 and 688 of 26,506 ASVs, respectively). Using supervised filtering we also removed the

Synergistetes and Chlorobi phyla that we suspected might be the result of sequencing error based on their extremely poor representation in our data set. ASVs without taxonomic assignment were also filtered out. We lastly executed prevalence threshold filtering at a level of 1% of the total samples. Sixteen samples were then randomly resampled from each population resulting in the inclusion of 96 samples and 16,401 bacterial ASVs in all downstream data analyses. We made a random tree using the **rtree** function in the **ape** package[52] in R. A phyloseq dataset was made using the **phlyoseq** function in the **phyloseq** R package v1.22.3[53]. We merged ASV table, sample data, and rtree into a phyloseq dataset object for downstream analyses.

Microbiome diversity

We found that variance grew with mean ASV abundance across gut samples, and thus corrected this lack of homoscedasticity by square root transforming our ASV abundance data. We were concerned that differences in number of sample and low sequencing depth in a few of the samples would influence diversity measurements. To correct any potential related errors, samples with <1000 reads were removed from the dataset (resulted in the loss of 1 sample from the dataset) and then sixteen samples from each population were randomly selected without replacement to generate a new dataset (resulted in a subset of 16,397 ASVs). Random resampling at a variety of depths (15-100 samples) didn't change the interpretation of LMM statistical testing results. Using this subset of these data we assessed gut microbiome alpha diversity in terms of Simpson and Shannon diversity indices using the **diversity** function in the **vegan v2.4.2** R package[54]. We transformed the Shannon index (1/D) to calculate the effective species number. Beta diversity was calculated using both weighted and unweighted UniFrac phylogenetic distances among fish gut microbiome using the **distance** function in the **phyloseq v1.22.3**[53] R package. These matrices were transformed into distance to centroid values for each gut community using the **betadisper** function in the **vegan v2.4.2**[54] R package.

Linear mixed models

We tested how well variables predicted gut diversity metrics using the following linear mixed model:

Diversity ~ Environment + (1 | F_{ST}) + (1 | Sample Period);

here "environment" is a fixed term (freshwater or estuary), "sample period" is also a fixed term (2015 or 2017), and F_{ST} (population average genetic distance) is a random variable.

Models were fit using restricted maximum likelihood methodology using lmer function in the lme4 package[55] in R. Using the step function in the lmerTest[56] R package, non-significant effects in LMMs were eliminated in a backward manner starting with random predictor variables, followed by fixed variables. Population and difference of means were calculated for the fixed part of the model and a final model provided. For hypothesis tests, the p-values for the fixed effect (estuary or freshwater) were calculated from F test based on Sattethwaite's approximation, rather than the pooled standard error formula, in order to deal with differences in standard deviation among samples. The p-values for random effects (population genetic structure, sample period, and starting DNA concentration) in our models were based on likelihood ratio tests ($\chi 2$). Distance among sites in terms of both river miles and land were tested for collinearity with host genetic distance using a Pearson's product-moment correlation. Distance was strongly colinear with population distance, thus distance was not included in LMMs to avoid inflation of model estimates.

Microbiome composition and PERMANOVA

We relative abundance transformed ASV counts by using their proportion of total ASV counts in each sample, in order to reduce the strong effects of potentially over-sampled taxa. PCoA ordination was carried out with a Bray-Curtis dissimilarity metric using the function **ordinate** in R package **phyloseq v1.22.3**[53]. This ordination was used to survey taxa covariance in the fish gut microbial metacommunity. A permutational multivariate

analysis of variance was used instead of a linear mixed model approach. We made this decision based on a lack of collinearity structure among ASVs observed along any given axis plotted in multivariate spatial models. We included fish population average genetic distance (F_{ST}), geographic distance (both river miles and land miles), sample period (2015 or 2017), and environment as a fixed binary variable (estuary or freshwater) in our full model. We tested the relative degree to which these variables explained gut microbiome composition in terms of Bray-Curtis dissimilarity among fish guts using the **adonis** function in **vegan v2.4.2** R package[54].

ASV abundance enrichment analysis

We converted the phyloseq to a DESeq dataset using the **phyloseq_to_deseq2** function in the **DESeq2 v1.18.1** R package[57]. The function **estimateSizeFactors** does not handle ASV count values=0. In order to handle zero values, we applied **gm_mean** = function(x, na.rm=TRUE){exp(sum(log(x[x > 0]), na.rm = na.rm)/length(x))}; where x is ASV counts. The **estimateDispersions** function was used to estimate dispersions. We tested gut ASVs for significant differential abundance both between inland and coastal freshwater populations and between estuary and freshwater populations using the **DESeq** function with the default Benjamini-Hochberg multiple-inference correction and with the **fitType** parameter set to 'local'. Only ASV differential abundances that were statistically significant ($p \leq 0.05$) were considered in our biological interpretation of these statistics results.

CHAPTER IV: DISCUSSION

Host population genomic divergence correlated with gut microbiome divergence

We have come to appreciate the importance of the host-associated microbiome in vertebrate health and disease. This current understanding is due to recent rapid advances in this field, primarily using experimental vertebrate models (e.g. mice and fish). More is constantly being revealed regarding the roles that the microbiome plays, intensifying our urgency to determine what influences the assemblages found among hosts. However, despite our awareness and efforts, we still lack a complete understanding of the relative roles of environment and host genetic variation in host microbiome assembly. In order to comprehensively study these interactions, we need to address a shortfall in this field. Wild hosts can possess incredible genetic variation and are found among diverse environments, yet little work in this field has taken advantage of this. In addition, studies that have been done using wild populations have not broadly considered the host genome. Rather they have focused narrowly on specific genes or small regions of the host genome deemed important by experimental work. This is an important issue to address in order to understand how experimental research should be interpreted and to advance the field.

Here we address this problem by collecting both host population genomic and microbiome data from wild populations of threespine stickleback fish in Oregon that inhabit very different habitats. We hypothesized that if host genetics played an important role in gut microbiome assembly, then with a robust level of population genetic resolution we would find a pattern in the gut metacommunity that reflected the genetic relationship among populations of wild hosts. We selected six populations of stickleback situated in Oregon among three water sheds over a range of geographic distances (~12-80 km apart as a crow flies), and across two distinct environments (estuary and freshwater). This allowed us to measure the relative degree to which each factor explained variation in gut microbiome among hosts.

Using several hundred single nucleotide polymorphisms randomly selected from thousands of genome-wide restriction site DNA markers (RADseq), we first confirmed a strong genetic differentiation between inland and coastal populations, as previously seen

in Oregon[31]. In addition, on the coast, estuary populations further differentiated, although to less an extent, from freshwater populations. Using metabarcoding data (16s rRNA) from the gut communities of these fish, we found that gut microbiome varied among both individual fish and fish populations in terms of composition and diversity in threespine stickleback in Oregon. We were therefore able to produce corresponding host genetic and microbiome data from natural populations of a vertebrate that is also amenable to manipulative microbiome studies in the laboratory[30]. In contrast to most model organisms (e.g. mice or zebrafish), in stickleback we can now integrate a much better understanding of the relative patterning of host genomic and microbiome diversity in the wild affiliated with the power of subsequent manipulative studies.

Our results provide evidence that genome-wide genetic variation among host populations can influence divergence in the gut microbiome. These findings were expected given that major mutations of single host immunity genes in gut assembly (e.g. myD88, NOD2, ob/ob, and Rag1). Our findings agree with findings in other systems that relatively more similar gut microbiome among related as compared to unrelated individuals[10,21–24] and that aspects of the gut microbiome can act as complex and heritable host traits[25–27].

Our findings also align with previous work in Canadian stickleback. Smith et al. (2015) found that average gut microbiome phylogenetic distance and host population genetic divergence in six satellite markers positively correlated[37]. Our work expands our understanding of the both the geographic and genetic scales at which this pattern is evident in wild threespine stickleback. We found similar results among populations in Oregon separated by ~12-80 km and used genome-wide markers to quantify host population genetic divergence.

We also found that both alpha and beta diversity varied among fish populations. This diversity was best predicted by population genetic divergence among populations. Stickleback with higher heterozygosity in Major Histocompatibility Complex class II (MHCII) alleles can have reduced gut alpha diversity[36]. Smith et al. (2015) later speculated that if high allelic heterozygosity in stickleback populations was a proxy for high MHCII, then it may have explained why they found lower beta diversity in Canadian stickleback populations with greater heterozygosity. We wondered if this was

true in Oregon populations as well. We found a negative correlation between population allele heterozygosity and gut community alpha diversity using all six populations in Oregon (t = -1.2259, df = 94, p = 0.2233). This weakly corroborated the previous speculation. In addition, stickleback closer to the ocean typically share higher gene flow with diverse marine populations than in freshwater environments. Therefore, they can maintain greater inter-individual genetic diversity than fish inland with limited gene flow with marine populations. We further speculated that this could lead to greater inter-individual variation in gut microbiome in coastal fish if the genetic variation occurs in traits important in gut assembly. Transplant experimentation using controlled environments might be necessary to test this.

Even though host population genetic divergence best predicted gut diversity, environment also explained a similar portion of the overall variation in gut microbiome. In addition, stickleback populations found in estuaries had on average higher beta diversity and lower alpha diversity than freshwater populations. This greater inter-individual variation in estuary fish could have been due to the exposure of estuarian fish to a more diverse pool of microbes. Stickleback acquire their microbes in part through diet, sediment, and water sources[34,37]. These inputs all vary among freshwater, marine, and estuary environments. Furthermore, estuarian environments can be a unique blend of both freshwater and marine inputs both in terms of microbiota and nutrients important in microbial growth[16]. Stickleback fish can traverse and have varying residency times among these environments in Oregon. Therefore, we speculated that this diverse exposure of the host among marine, estuary, and freshwater environments might have contributed to gut microbiome diversity among individual stickleback. Fish found in estuaries known to spend varying amounts of time in both marine and estuary environments could have led to the greater average inter-individual diversity we found in estuary populations.

We also observed that estuary and freshwater fish had different gut communities in terms of the relative abundance of major phyla present in the gut, as was the case in another study in Canada[37]. The Firmicutes in estuary fish and the Protoebacteria in freshwater fish were the most prevalent phyla. Differences in prevalent taxa found in the gut microbiome among environments could in part be due to variation in factors that

differentiate the microbial pool available among environments. Variation in the microbial pool across space can lead to differential colonization among hosts occupying different environments. Salinity is an important variable that globally sorts microbial communities in the environment[58].

A small subset of microbial taxa contributed to microbiome divergence linked to host genetics and environment

Given that population genetic divergence explained a portion of gut metacommunity variation, we wanted to know which microbial taxa contributed to significant differential abundance between inland and coastal freshwater fish. We found that a small subset of the gut metacommunity was significantly differentiated in terms of abundance across the main divergence in host population genetic structure in Oregon. These ASVs represent a group of bacteria that could strongly interact with the host in terms of genome-wide allele variation by being heritable and important in host fitness. This notion seemed plausible considering the history of threespine stickleback. These ASVs, if isolated, could be tested for differential colonization in lab fish raised free of microbes, a benefit of work in a system that exhibits both natural variation and can be manipulated in the laboratory. Further work using experimental manipulations of stickleback lines in the laboratory might reveal the genetic underpinning of host interplay with and the heritability of these isolates.

It is unclear if host-environment interactions can influence the gut microbiome indirectly via the host. There was a subset of phyla comprised of ASVs that were significantly enriched in both or only one of these environments. A smaller subset of these was enriched log2fold > 20 suggesting a strong enrichment due to environment-related processes. We concluded that a strong association of some taxa in the gut can be specific to the environments in which hosts are found. This specificity could be due to the sorting of the microbial pool in the environment. However, it has been found that microbes in the environment don't predict what is found in the stickleback gut[37]. Furthermore, little is understood how the environment influences gut assembly to this end, as few studies have

yet explored these processes in wild populations and in natural environments. Further testing in stickleback could help illuminate these aspects of gut assembly. For instance, parallel work in lab experimentation and using bacterial strains curated from the wild incorporated with environmental gradients could elucidate host-environment-microbe interactions.

CHAPTER V: CONCLUSION

We acknowledge that sampling two years apart explained as much gut community variation as environment (~6%). In dynamic aquatic systems, much can fluctuate over time in terms of both microbiome and environment. In addition, stochastic processes such as dispersal and random replacement can contribute over time. For example, high microbial dispersal rates in lab fish can influence gut assembly to the extent that the effects of fish genotype are negligible[59]. However, microbial dispersal among hosts has not been directly studied in the wild. Instead of direct measurement, dispersal has been inferred in wild systems by observing shifts in gut microbiome diversity[60], and by using the spatial distribution of hosts[37] and their social interactions[61] as proxies for microbial dispersal. Interestingly, we found that geography explained less gut microbiome variation than F_{ST} and that these factors strongly covaried. We conclude that to infer microbial dispersal among stickleback populations using geographic distance as a proxy is confounding. Therefore, we made no such inference. Lastly, random sampling of microbes from the source pool and the stochastic loss and replacement of microbes can also explain variation in gut microbiome[62]. These "neutral processes" in gut microbiome assembly have not yet been studied in earnest in wild populations and have never been explored in threespine stickleback. We acknowledge these caveats, and thus our conclusions should be taken as initial results that warrant subsequent studies.

Fortunately, the stickleback is just the system for this work given advancing tools and the majority of gut microbiome variation remaining unexplained (~76%) after accounting for FST and environment. There is a lot we can do in this system. A lack of a comprehensive explanation for all this remaining variation, which is consistent with other systems, could be due to the lack of our understanding of both host and non-host factors that influence the microbiota. These can be explored in stickleback both in wild and laboratory settings. In addition, as technology advances, we are continually refining how we define and measure the microbiota. For example, we can consider the microbiota in terms of metaproteome[63], metabolome[64], and metatranscriptome[65] functions as alternatives to taxonomic profiling. Myriad environmental variables that were not measured in this study

could also be important, such as dietary inputs of fat[66] and microbes[37] functions as alternatives to taxonomic profiling.

Microbiomes serve important roles in the health and development of animal hosts, including humans. Therefore, it is imperative that we understand how microbiomes influence their hosts and develop methods to manipulate this interplay to alter host fitness. This requires understanding the factors that influence microbiome variation across individual hosts. Most research toward this aim has been conducted in laboratory animal populations with little standing genetic variation and a poor grasp of their natural ecology and evolution (e.g. mice and zebrafish), or through comparative studies of natural populations which lack laboratory tractability and extensive genetic tools (e.g. non-human primates and humans). Threespine stickleback is an ideal system with which to improve our understanding of the factors that influence microbiome variation by circumventing these issues. The ecology and evolution of threespine stickleback is well understood, it is nearly ubiquitous in diverse natural aquatic systems across the holarctic, and is tractable in the laboratory. These features will allow for cross-examination between laboratory and natural systems permitting a comprehensive approach to this area of scientific enquiry.

REFERENCES CITED

1. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320,** 1647–51 (2008).

2. Bäckhed, F. *et al*. Host-Bacterial mutualism in the human intestine. *Science* **307,** 1915–1920 (2005).

3. Hacquard, S. *et al.* Microbiota and host nutrition across plant and animal kingdoms. *Cell Host Microbe* **17,** 603–616 (2015).

4. Sommer, F. & Bäckhed, F. The gut microbiota — masters of host development and physiology. *Nature Review Microbiology* **11,** 227–238 (2013).

5. Kanther, M. & Rawls, J. F. Host-microbe interactions in the developing zebrafish. *Curent Opinion Immunoogy* **22,** 10–9 (2010).

6. Kamada, N. *et al*. Control of pathogens and pathobionts by the gut microbiota. *Nature Immunology* **14,** 685–90 (2013).

7. Marchesi, J. R. *et al.* The gut microbiota and host health: a new clinical frontier. *Gut* **65**, 2 (2015). doi:10.1136/gutjnl-2015-309990

8. Honda, K. & Littman, D. R. The microbiota in adaptive immune homeostasis and disease. *Nature* **535,** 75–84 (2016).

9. Spor, A. *et al*. Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Review Microbiology* **9,** 279–290 (2011).

10. Goodrich, J. K. *et al.* Human Genetics Shape the Gut Microbiome. *Cell* **159,** 789–799 (2014).

11. Green, J. & Bohannan, B. J. M. Spatial scaling of microbial biodiversity. *Trends in Ecology and Evolution* **21,** 501–507 (2006).

12. Burns, A. R. *et al.* Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME Journal* **10,** 655–664 (2016).

13. Stephens, W. Z. *et al.* The composition of the zebrafish intestinal microbial community varies across development. *ISME Journal* **10**, 644-654 (2015).

14. Dolan, J. R. An introduction to the biogeography of aquatic microbes. *Aquatic Microbial Ecology* **41**, 39-48 (2005). doi:10.3354/ame041039

15. Martiny, J. B. H. *et al.* Microbial biogeography: putting microorganisms on the map. *Nature Review Microbiology* **4,** 102–112 (2006).

16. Crump, B. C. *et al*. Microbial biogeography along an estuarine salinity gradient: combined influences of bacterial growth and residence time. *Applied Environmental Microbioogy.* **70,** 1494–505 (2004).

17. Fortunato, C. S. *et al*. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME Journal* **6,** 554–563 (2012).

18. Sullam, K. E. *et al.* Environmental and ecological factors that shape the gut bacterial communities of fish: a meta-analysis. *Molecular Ecology* **21,** 3363–3378 (2012).

19. Yildirim, S. *et al.* Characterization of the fecal microbiome from non-human wild primates reveals species specific microbial communities. *PLoS One* **5,** e13963 (2010).

20. Ley, R. E. *et al.* Evolution of mammals and their gut microbes. *Science* **320,** 1647-1651(2008).

21. Rawls, J. F. *et al*. Reciprocal Gut Microbiota Transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* **127,** 423–433 (2006).

22. Zoetendal, E. G. *et al.* The host genotype affects the bacterial community in the human gastronintestinal tract. *Microbial Ecology in Health and Disease* **13,** 129–134 (2001).

23. Turnbaugh, P. J. & Gordon, J. I. The core gut microbiome, energy balance and obesity. *Journal of Physiology.* **587,** 4153–8 (2009).

24. Nelson, A. *et al*. Assessment of sample handling practices on microbial activity in sputum samples from patients with cystic fibrosis. *Letter in Applied Microbiology* **51,** 272–277 (2010).

25. Benson, A. K. *et al.* Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proceedings of The National Acadamy of Science*s **107,** 18933–18938 (2010).

26. Leamy, L. J. *et al.* Host genetics and diet, but not immunoglobulin A expression, converge to shape compositional features of the gut microbiome in an advanced intercross population of mice. *Genome Biology* **15,** 552 (2014).

27. Davenport, E. R. *et al.* Genome-wide association studies of the human gut microbiota. *PLoS One* **10,** e0140301 (2015).

28. Davey, J. W. *et al.* RADSeq: next-generation population genetics. *Oxford Journals Briefings in Functional Genomics* **9,** 416–423 (2010).

29. Hohenlohe, P. A. *et al.* Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics* **6,** e1000862 (2010).

30. Lescak, E. A. & Milligan-Myhre, K. C. Teleosts as Model Organisms to Understand Host-Microbe Interactions. *Journal of Bacteriology* **199**, 15 e00868-16 (2017). doi:10.1128/JB.00868-16

31. Catchen, J. *et al.* The population structure and recent colonization history of Oregon threespine stickleback determined using RAD-seq. *Molecular Ecology* **22**, 2864-2883 (2013).

32. Small, C. M. *et al.* Host genotype and microbiota contribute asymmetrically to transcriptional variation in the threespine stickleback gut. *Genome Biology and Evolution* **9,** 504–520 (2017).

33. Milligan-Myhre, K. *et al.* Innate immune responses to gut microbiota differ between oceanic and freshwater threespine stickleback populations. *Disease Models and Mechinisms* **9,** 187–98 (2016).

34. Bolnick, D. I. *et al.* Individuals' diet diversity influences gut microbial diversity in two freshwater fish (threespine stickleback and Eurasian perch). *Ecology Letters* **17,** 979–987 (2014).

35. Bolnick, D. I. *et al.* Individual diet has sex-dependent effects on vertebrate gut microbiota. *Natre Communications* **5,** 4500 (2014).

36. Bolnick, D. I. *et al.* Major histocompatibility complex class iib polymorphism influences gut microbiota composition and diversity. *Molecular Ecology* **23,** 4831–4845 (2014).

37. Smith, C. C. R. *et al.* Dietary input of microbes and host genetic variation shape among-population differences in stickleback gut microbiota. *ISME J.* (2015). doi:10.1038/ismej.2015.64

38. Earl, D. A. & vonHoldt, B. M. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* **4,** 359–361 (2012).

39. Holben, W. E. *et al.* Phylogenetic analysis of intestinal microflora indicates a novel mycoplasma phylotype in farmed and wild salmon. *Microbial Ecology* **44,** 175–185 (2002).

40. Romero, J., Ringø, E. & Merrifield, D. L. *Aquaculture Nutrition.* 75–100 (John Wiley & Sons, Ltd, 2014). doi:10.1002/9781118897263.ch4

41. Llewellyn, M. S. *et al.* Teleost microbiomes: the state of the art in their characterization, manipulation and importance in aquaculture and fisheries. *Frontiers in Microbiology* **5,** 207 (2014).

42. Sasaki, Y. *et al.* The complete genomic sequence of Mycoplasma penetrans, an intracellular bacterial pathogen in humans. *Nucleic Acids Research* **30,** 5293–300 (2002).

43. Burns, A. R. *et al.* Transmission of a common intestinal neoplasm in zebrafish by cohabitation. *Journal of Fish Disease* **41**, 569-579 (2017). doi:10.1111/jfd.12743

44. Salinas, I. *et al*. Analysis of the gut and gill microbiome of resistant and susceptible lines of rainbow trout (Oncorhynchus mykiss). *Fish Shellfish Immunology* **86**, 497-506 (2019).

45. Lyons, P. P. *et al*. Phylogenetic and functional characterization of the distal intestinal microbiome of rainbow trout *Oncorhynchus mykiss* from both farm and aquarium settings. *Journal of Applied Microbiology* **122,** 347–363 (2017).

46. Catchen, J. *et al*. Stacks: an analysis tool set for population genomics. *Molecular Ecology*. **22,** 3124–40 (2013).

47. Wu, T. D. *et al*. GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Methods in molecular biology* **1418,** 283–334 (2016).

48. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9,** 357–359 (2012).

49. R Core Team. R: A language and environment for statistical computing. (2015). at <https://www.r-project.org/>

50. Pritchard, J. K. *et al.* Documentation for structure software: Version 2.3. (2010). at <http://pritch.bsd.uchicago.edu/structure.html>

51. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods* **13,** 581–583 (2016).

52. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20,** 289–290 (2004).

53. McMurdie, P. J. & Holmes, S. Phyloseq: an r package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8,** e61217 (2013).

54. Oksanen, J. *et al.* Vegan: community ecology package. (2015). at <https://cran.r-project.org/web/packages/vegan/index.html>

55. Bates, D. *et al*. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67** (2015).

56. Kuznetsova, A. *et al*. LmerTest: tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). *Journal of Statistical Software* **82**,13 (2014). doi:http://CRAN.R-project.org/package=lmerTest

57. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15,** 550 (2014).

58. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proceedings of the National Academy of Sciences* **104,** 11436-11440 (2007).

59. Burns, A. R. *et al.* Interhost dispersal alters microbiome assembly and can overwhelm host innate immunity in an experimental zebrafish model. *Proceedings of the Nationa. Acadamy of Sciences* **114,** 11181–11186 (2017).

60. Martínez, I. *et al.* The gut microbiota of rural papua new guineans: composition, diversity patterns, and ecological processes. *Cell Rep.* **11,** 527–538 (2015).

61. Tung, J. *et al.* Social networks predict gut microbiome composition in wild baboons. *Elife* **4,** (2015).

62. Burns, A. R. *et al.* Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J.* **10**, 655-664 (2015). doi:10.1038/ismej.2015.142

63. Grassl, N. *et al.* Ultra-deep and quantitative saliva proteome reveals dynamics of the oral microbiome. *Genome Medicine.* **8,** 44 (2016).

64. McHardy, I. H. *et al.* Integrative analysis of the microbiome and metabolome of the human intestinal mucosal surface reveals exquisite inter-relationships. *Microbiome* **1,** 17 (2013).

65. Giannoukos, G. *et al.* Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology* **13,** r23 (2012).

66. Wong, S. *et al.* Ontogenetic differences in dietary fat influence microbiota assembly in the zebrafish gut. *MBio* **6,** e00687-15 (2015).