Microbiome Composition and Evolutionary Genomics During Rapid Adaptation to Freshwater in the Copepod *Eurytemora affinis*

By

Martin Bontrager

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Genetics)

at the

UNIVERSITY OF WISCONSIN-MADISON

2016

Date of final oral examination: 12/9/2016

The dissertation is approved by the following members of the Final Oral Committee:

Carol E. Lee, Professor, Zoology
Chris T. Hittinger, Assistant Professor, Genetics
Nicole Perna, Associate Professor, Genetics
John E. Pool, Assistant Professor, Genetics
Katherine D. McMahon, Professor, Bacteriology and Civil and Environmental Engineering

Table of Contents

Abstract

The saline-to-freshwater biogeographic barrier is among the most dramatic physiological challenges that species face, and is among the greatest factor that structures their distributions. Given the scope of global climate change and anthropogenic activities, including the rapid decline of ocean salinity in many coastal waters, this barrier is one that many populations will be forced to overcome. I use the natural experiments provided by the invasive copepod *Eurytemora affinis* to study shifts in the microbiome (Chapter 1) and in the genome (Chapter 3) of species that cross the saline-to-freshwater barrier. I also focus on the microbiome of the zooplankton community as a reservoir for fish pathogens (Chapter 2). In the first chapter, I examine the host-associated microbiome of *E. affinis* in a comparative context across the saline-to-freshwater barrier in independent invasive clades and contrast the host microbiome with that of the surrounding water. I report that the aquatic microbiomes are structured primarily by host-association *versus* presence in the bacterioplankton community, and that copepod microbiomes are structured by salinity rather than geography or host clade. In the second chapter, I give empirical evidence for the presence of fish pathogens in the copepod microbiome, both from 16S rRNA gene sequencing data and from the isolation and pathogenicity trials of a novel copepod-associated strain of *Flavobacterium*. This is of enormous relevance to the fisheries and aquaculture industries, as the ecology and reservoirs of fish pathogens are poorly understood. In the third chapter, I characterize population genomic signatures of selection, particularly focusing on ion-transport genes, associated with saline to freshwater invasions in a clade of the copepod *E. affinis*. My results make significant contributions to the understanding of the bacterial community composition and genomic changes during rapid adaptation to freshwater in *E. affinis*.

List of Tables and Figures

Chapter 1

Chapter 1

The *Eurytemora affinis* microbiome and shifts in composition during multiple independent habitat invasions

Authors: Martin Bontrager[1], Jane Louise Remfert[2†], Hanzel T. Gotia[3], James B. Munro[3], Joana C. Silva[3, 4], Carol Eunmi Lee[2*]

[1] **Laboratory of Genetics, University of Wisconsin-Madison,** 425 Henry Mall, Madison, WI 53706
[2] Center for Rapid Evolution (CORE), University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706
[3] Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street, Baltimore, MD, 21201
[4] Department of Microbiology and Immunology, University of Maryland School of Medicine, 685 W. Baltimore Street, Baltimore, MD, 21201

Abstract

Bacterial communities associated with animal hosts (i.e., their "microbiomes") are often distinct from free-living bacterial assemblages and can perform unique functions that are central to host health. Yet little is known about the composition and characteristics of these microbial communities, particularly for aquatic zooplankton. Within the past century, the copepod *Eurytemora affinis* has invaded multiple freshwater habitats independently from coastal estuarine habitats throughout the Northern Hemisphere. We examined the *E. affinis* microbiome across multiple independent saline-to-freshwater invasions, and compared the composition of *E. affinis* microbiomes to the surrounding bacterioplankton community in the water column. Salinity was the dominant environmental variable structuring the compositions of copepod and water column microbiomes. We found that the copepod microbiomes were highly distinct from the surrounding bacterioplankton and contained members of several lineages known to be obligately or facultatively anaerobic. Copepod microbiomes were also significantly less diverse than the bacterioplankton. Copepod microbiomes were dominated by members of the Proteobacteria and Bacteroidetes, while water microbiomes were dominated by Actinobacteria. Although copepod microbiomes were highly distinct from one-another and composed of many taxa unique to each population, we found evidence for parallel shifts in the occurrence and abundance of bacterial taxa across independent saline-to-freshwater invasions. This study is novel in exploring the microbiomes of an aquatic animal host, and of its surrounding environment, across multiple independent habitat invasions.

Introduction

In recent years, we have undergone a revolution that is revealing the critical roles played by host-associated microbiomes. A remarkable diversity of microbes has been found in association with nearly every multicellular eukaryote, and the functions performed by these host-associated microbes have in many cases been proven essential for host survival and function (Ley *et al.* 2008; Engel & Moran 2013). In insects, for example, researchers have found that host microbiomes can be involved in critical host processes, including digestion (Brune & Ohkuma 2010; Flint *et al.* 2012), food detoxification (Cardoza *et al.* 2006; Adams *et al.* 2009), pathogen resistance (Currie *et al.* 2003; Kaltenpoth 2009; Lee & Mazmanian 2010), and thermal tolerance (Dunbar *et al.* 2007). Other studies, in mice and flies, have shown that host microbiomes may even influence host behavior (Bravo *et al.* 2011; Ezenwa *et al.* 2012; Shropshire & Bordenstein 2016). Given these findings, it is perhaps most appropriate to conceptualize multicellular eukaryotes as complex biological networks, or holobionts, in which biochemical and physiological functions are encoded by a hologenome which includes both the genomic material of the host and the numerous genomes of its associated microbiome (Zilber-Rosenberg & Rosenberg 2008; Bordenstein & Theis 2015). Copepods are one such host in which the microbiome is largely unexplored, and yet of potentially profound importance.

Copepods are the most abundant metazoans in the world's oceans (Verity & Smetacek 1996) and arguably on the planet (Hardy 1970). Due to their enormous biomass, copepods play critical roles in aquatic food webs and provide the food source for economically important fisheries worldwide. Given their high abundance and biomass, copepod populations have the capacity to harbor an enormous bacterial community. Copepod-associated microbes are likely to be distinct from the surrounding water bacterioplankton due to the unique niche provided by the copepod

cuticle (Keyhani & Roseman 1999), hemolymph (Brandin & Pistole 1985) and anaerobic and nutrient rich gut (Harris 1993; Hansen & Bech 1996; Tang *et al.* 2011). We know that copepods can harbor orders of magnitude more bacteria than the surrounding water (Tang 2005; Møller *et al.* 2007) and in some cases copepod-associated bacteria account for up to 40% of total bacteria counts in nearshore marine ecosystems (Heidelberg *et al.* 2002a). Yet we know little about the composition of copepod microbiomes, how they differ from the bacterial community in the surrounding water, and how they might shift in response to changing environmental factors.

The calanoid copepod *Eurytemora affinis* is a dominant grazer in many coastal ecosystems throughout the Northern Hemisphere (Peitsch *et al.* 2000; Winkler *et al.* 2005). This species is a major food source for important fisheries, including herring, smelt, striped bass, and flounder (Viitasalo *et al.* 2001; Dauvin & Dodson 1990; Shaheen *et al.* 2001; Kimmel *et al.* 2006). Within the past century, *E. affinis* has invaded freshwater habitats from estuarine habitats worldwide in multiple independent events (Lee 1999). Newly established freshwater populations of *E. affinis* have undergone rapid evolution in salinity tolerance, ionic regulation, nutritional requirements, life history, and gene expression (Lee *et al.* 2003, 2007, 2011, 2012, 2013; Lee 2016; Gerber *et al.* 2016). *E. affinis* is remarkable in its ability to invade from saline to freshwater environments, as most species are unable to cross this formidable biogeographic barrier (Hutchinson 1957; Khlebovich & Abramova 2000)

Upon introduction into fresh water, *E. affinis* encounters a novel environment with a suite of potential selective pressures, including sparse ionic concentrations in the water, new food sources, and novel pathogens. This rapid environmental shift has the potential to radically alter the copepod microbiome. During a freshwater invasion, members of the saline copepod microbiome may preserve their associations with the copepod host or become lost from the

copepod microbiome. Conversely, those bacterial species in the freshwater environment that are absent in the saline copepod microbiome might form new associations with the invading copepod host. The recent, parallel independent invasions of *E. affinis* make this species a unique and powerful model for studying how host-associated microbes cross the saline to freshwater barrier, and more generally how host microbiomes respond to rapid environmental change.

Thus, the goals of this study were to determine (1) the shifts in the *E. affinis* microbiome during multiple independent saline-to-freshwater habitat invasions and (2) the degree of similarity between the copepod microbiome and the bacterial community in the surrounding water column. To accomplish these goals, we sampled copepod microbiomes and surrounding bacterioplankton from multiple pairs of saline and freshwater habitats representing three independent saline-to-freshwater invasions. We then characterized the microbiome composition of these samples using high-throughput 16S rRNA amplicon sequencing.

We were able to make statistically powerful inferences regarding the impact of environmental shifts on microbiome composition by exploiting the natural experiment provided by replicate invasions of *E. affinis* into freshwater. Previous studies had described the composition of copepod microbiomes from single locations (Grossart *et al.* 2009; Gerdts *et al.* 2013; Bickel & Tang 2014; De Corte *et al.* 2014; Shoemaker & Moisander 2015), or copepod microbiomes across multiple salinities in a saline environment (Dziallas *et al.* 2013). This study is the first to analyze the microbiome of replicate populations of an invasive species across a fundamental biogeographic barrier. This is also the first study to contrast replicate aquatic host microbiomes with the free-living bacterioplankton community across such a barrier. As such, we developed unique insights regarding shifts in microbiome composition during rapid environmental change.

Materials and Methods

Sampling Design

We obtained microbiome samples by collecting paired samples of the copepod *Eurytemora affinis* and the surrounding water from 14 locations in both saline and freshwater environments (Fig. 1). We sampled three sets of ancestral saline (S) and derived freshwater (F) populations, representing three independent invasions, from: (A) the St. Lawrence estuary into the Great Lakes, North America, (B) the Gulf of Mexico into reservoirs along the Mississippi river, USA, and (C) the North and Wadden Sea into freshwater reservoirs in the Netherlands, in Europe. At each location, we performed plankton tows, using a 150 μm mesh net, to collect zooplankton from a depth of 1-2 meters and then isolated *E. affinis* from other members of the plankton community. We gathered 100 individual *E. affinis* copepods from each location, with approximately equal numbers of adult males and females, and rinsed the animals in a mesh filter with a total of 500 mL of sterile water. We then flushed the animals into 1 mL of sterile water. We froze the samples immediately at -20°C, and transferred them to -80°C within 24 hours.

At locations from which we collected copepod samples, we also collected water in sterilized 25 L carboys from a depth of approximately 1.5 meters. Upon collection, we first filtered the water through a 100 μm filter to remove large particulate matter, and subsequently through a 3.0 μm filter to exclude algae and other large particles. Finally, we passed the water through a 0.1 μm filter to collect free-living bacteria. The volumes of water filtered varied with water turbidity, but in all locations ranged between 3-7 liters. We immediately froze the 0.1 μm filters with collected bacteria at -20°C, and subsequently at -80°C, within 24 hours. As DNA extraction failed in some cases, two copepod samples were not paired with matching water samples from the same time

point but instead with water from the subsequent season (Fig. 1, samples AS-1 and AS-2). We collected metadata from each location, including latitude, salinity, and water temperature (Table 1).

DNA Extraction and Sequencing Library Prep

We extracted DNA *via* mechanical lysis (bead-beating) to faithfully represent the microbial community in as unbiased a manner as possible (Yuan *et al.* 2012). We extracted DNA from copepod microbiome samples using the UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) per manufacturer recommendations and with the inclusion of a 10-minute incubation at 65°C prior to mechanical lysis to increase extraction efficiency. We extracted DNA from filtered water samples with the PowerWater DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) according the manufacturers recommendations, again with the additional step of 10 min incubation at 65°C.

The bacterial 16S rRNA gene is highly conserved across all known bacterial taxa. This highly conserved gene contains hypervariable regions, the sequence of which can be used to cluster bacterial taxa based on similarity (Stackebrandt & Goebel 1994). We amplified the V3/V4 hypervariable region of the 16S rRNA gene using the 319F/806R PCR primers, including Illumina adapters and staggered barcodes as implemented by Fadrosh et al (2014). After PCR amplification, we sequenced amplicon libraries using 300bp Illumina paired-end MiSeq sequencing. During sequencing, each sample was split into 6 technical replicates, which were later combined into one large set of reads per sample.

DNA Sequence Processing and Data Analysis

We overlapped paired-end reads using USEARCH (Edgar 2013) "-fastq_mergepairs," resulting in approximately 430bp sequences of the V3/V4 hypervariable region from the 16S gene (Fadrosh *et al.* 2014). After quality filtering and chimeric read removal (Edgar *et al.* 2011; Schloss *et al.* 2011; Bokulich *et al.* 2012), sequence reads were binned into observational taxonomic units (OTUs) at a 97% sequence identity level, a level which serves as a conventional proxy for bacterial "species" (Hagström *et al.* 2000). We assigned taxonomy to OTUs *via* methods implemented in MOTHUR (Schloss *et al.* 2009) and using the greengenes database (version 05-2013) modified to incorporate additional freshwater bacterial 16S sequences (DeSantis *et al.* 2006; Newton *et al.* 2011).

Binned and classified sequences and sequence counts were used to characterize alpha and beta diversity in microbial communities. Alpha diversity is a measure of species richness within samples, and can be quantified simply by OTU count or metrics such as the Abundance-based Coverage Estimator, Chao richness, Simpson Diversity, and Shannon Diversity (Chao 1984; Chao *et al.* 2004; Shannon 1948; Simpson 1949). OTU counts were not rarefied to account for uneven depth of coverage prior to alpha diversity estimation, as doing so has been shown to be inappropriate (McMurdie & Holmes 2014). We quantified species richness with alpha diversity metrics in all samples and compared richness among copepod microbiomes, as well as between copepod and water microbiomes.

In contrast to alpha diversity, beta diversity evaluates differences in taxonomic composition among samples. Using beta diversity measurements, microbiomes can be compared to one another to assess whether, for example, copepod microbiome A is more like copepod microbiome B or to bacterioplankton from environment A. To robustly examine microbiome similarity, we generated multiple distance matrices to analyze beta diversity among samples

including UniFrac (Lozupone & Knight 2005; Lozupone *et al.* 2011), Bray-Curtis (Bray & Curtis 1957), Chao (Chao *et al.* 2004), Horn-Morisita (Horn 1966), Jaccard (Jaccard 1912), Jensen-Shannon (Endres & Schindelin 2003), and Mountford (Mountford 1962). We visualized microbiome similarity using classical multidimensional scaling (MDS). Additionally, we performed statistically robust analysis of similarity among microbiomes using UniFrac distance measurements, which account for the length of unshared branches on a phylogenetic tree to calculate distance between two microbial communities (Lozupone *et al.* 2011). We constructed a phylogeny using maximum likelihood, as implemented in FastTree2 (Price *et al.* 2010), and used UniFrac (Lozupone & Knight 2005) to calculate pairwise distances between microbiomes. We formally tested differences between microbiome types *via* permANOVA (Anderson 2001).

We also examined similarity among microbiomes by assessing which OTUs were shared among multiple samples. We first normalized OTU counts using variance stabilization with DESeq2 (McMurdie & Holmes 2014; Love *et al.* 2014) prior to testing. We then performed the Wald test on variance-stabilized OTU counts to identify bacterial OTUs that differed significantly between environmental conditions such as salinity or latitude. This variance stabilization and Wald testing approach allows for significance testing of taxon abundance between conditions with correction for multiple inferences. Graphical visualizations were constructed using the Phyloseq package (McMurdie & Holmes 2013) in the statistical package R (R Core Team 2014). Full scripts used for pipeline analysis and description of the process used for demultiplexing, quality filtration, OTU binning, and taxonomic classification are available at https://github.com/mbontrager/16S_processing.

Results

Microbiomes of the Copepod Host and of the Surrounding Water are Highly Distinct

Across the broad geographic range of our microbiome sampling (Fig. 1), bacterial community composition differed most greatly between microbiomes of copepod hosts *versus* those from free-living bacterioplankton. To examine differences among microbiomes (beta diversity) and determine the environmental factors that structure these bacterial communities, we clustered the microbiome samples using six distance measures. Classical multidimensional scaling (MDS) using the six distance metrics measured showed clear separation between copepod samples *versus* water samples along the first axis (Fig. 2). When we used a phylogenetically informed UniFrac distance matrix for MDS clustering, we obtained the same result (Fig. 3). This separate clustering of copepod *versus* water samples was highly significant *via* permutational multivariate analysis of variance testing (permANOVA on unweighted UniFrac distances, $P \leq 0.001$).

In terms of community composition, copepod and water microbiomes were distinct even at the level of phyla (Fig. 4). Water microbiomes were dominated by Actinobacteria, which were significantly less common in copepod microbiomes (Wald test, adjusted $P \sim 0$). Indeed, members of the phylum Actinobacteria comprised 32.14% - 90.82% of the total water bacterial community (median = 45.72%), while comprising only 0.08% - 4.28% of the copepod community (median = 0.46%). Copepod microbiomes, in turn, were significantly enriched for members of the phylum Proteobacteria (Wald test, adjusted $P = 5.1 \times 10^{-5}$), the phylum Firmicutes and the relatively new phylum Gracilibacteria ($P \sim 0$ in each case). Proteobacteria were dominant in copepod samples and accounted for 36.66% to 98.41% of the composition of copepod microbiomes (median = 77.23%), but only 5.27% to 44.19% of the water microbiomes (median = 35.82%). Within the phylum Proteobacteria, the Gammaproteobacteria was the most dominant class in copepod microbiomes, comprising 0.38% to 59.66% of the total taxa in copepod-associated samples

(median = 9.86%, Fig. 5). In contrast, in water samples the class Betaproteobacteria within Proteobacteria was the most dominant, comprising 2.05% to 41.71% of the community in water samples (Fig 6).

In general, at the higher taxonomic levels of class and phylum, the copepod microbiomes were more variable with respect to the consistency of classes and phyla abundance than the water samples across locations. Water bacterial communities were invariably dominated by the class Actinobacteria, for example, while different copepod microbiomes were often dominated by different classes including Alphaproteobacteria, Betaproteobacteria, and Saprospirae (Figs. 5 & 6). Variance was significantly higher in the copepod microbiomes when we compared the relative abundance of the top 10 bacterial classes between water and copepod microbiome samples (alternative hypothesis = copepod microbiome class variance > water microbiome class variance, *t*-test, $P = 0.02$). This pattern persisted irrespective of salinity.

In addition to compositional differences between copepod and water microbiomes, copepod communities were also less diverse and contained fewer bacterial taxa than the free-living bacterioplankton communities (Fig. 7). In fact, microbiome diversity was always lower in copepod samples than in water samples from the same location (Fig. 8), except in one case where the diversity between water and copepod samples was equivalent. Alpha diversity indices showed that the copepod microbiomes contained fewer total taxa and had lower overall diversity than the surrounding water (paired *t*-test on Chao diversity, $P = 2.5\text{x}10^{-4}$). The pattern of lower diversity in the copepod microbiomes remained constant regardless of the metric used to measure diversity (i.e. Simpson, Shannon, Chao, and observed OTUs). We found that there was lower alpha diversity in the copepod microbiomes whether we considered all water and copepod samples independently (*t*-test on Chao diversity, $P = 2.4\text{x}10^{-4}$) or we tested paired copepod/water

samples from the same location (paired *t*-test, $P = 2.5x10^{-4}$). Additionally, this pattern of diversity pervaded all hierarchical taxonomic categories above the OTU level, where copepod microbiomes contained fewer genera (paired *t* test, $P = 8.2x10^{-3}$), families ($P = 5.3x10^{-3}$), orders ($P = 2.8x10^{-3}$), classes ($P = 2.5x10^{-3}$), and phyla ($P = 9.7x10^{-4}$) than water samples from the same locations.

The number of bacterial OTUs within microbiomes was quite variable, especially among water microbiomes in different locations (Fig. 7, "Observed"). This effect disappeared when bacterial alpha diversity was weighted by abundance as opposed to simply presence/absence, as it was when calculated with the Shannon and Simpson diversity indices (Fig. 7), suggesting that in water bacterial communities with many observed taxa, most of those taxa are rare and only represented by a few read counts in our data.

Salinity Structures Copepod Host and Surrounding Water Microbiomes

Aside from host-association *versus* free-living, presence in saline *versus* freshwater was the most dominant factor structuring the composition of all microbiomes. When we excluded water samples, salinity was the most significant factor influencing similarity among copepod microbiomes (permANOVA, $P = 0.007$ for UniFrac unweighted distance, Fig. 9). Salinity was also the dominant factor structuring the water microbiomes ($P = 0.006$). In fact, even when all copepod and water samples were analyzed together, salinity was the only significant factor structuring community similarity aside from whether the microbiome was host-associated or free-living ($P = 0.038$, Fig. 3).

Despite the importance of salinity in structuring copepod microbiome composition among populations, there was no difference in bacterial community richness or evenness between saline

and freshwater copepod microbiomes (Fig. 10A, *t*-test $P = 0.40$). We also found no significant difference in microbiome diversity between saline and fresh water microbiome samples (Fig. 10B, $P = 0.6$). Copepod microbiomes were not significantly differentially diverse across continents, latitude, water temperature, and clades (Figs. 11 & 12) These diversity relationships persisted whether we used original unaltered sequence counts or normalized sample sequence counts to account for differential sequencing coverage.

Parallel Shifts in Copepod Host Microbiomes during Saline to Freshwater Invasions

Our finding that salinity structured copepod microbiomes also suggests that copepod bacterial communities converge during independent habitat invasions. Not only did salinity structure the copepod microbiomes, but failed to observe an invasive clade affect in structuring copepod microbiomes (permANOVA, $P = 0.091$ for UniFrac unweighted distance). This suggests that not only does salinity structure copepod microbiome composition wherein only all saline samples cluster together, but that freshwater copepod microbiomes are more like other freshwater copepod microbiomes. Additionally, we found evidence for parallel shifts in the presence and abundance of individual OTUs during the saline-to-freshwater transition. This result persisted in the independent invasions in genetically distinct clades (see Fig. 1).

Our testing of significant differences (cutoff multiple-adjusted $P = 0.05$) among saline and freshwater copepod microbiomes from independent clades identified three OTUs that were significantly more abundant in copepods in saline water than copepods in freshwater across clades: all were unclassified members of the family Rhodobacteraceae (Wald test, adjusted $P < 0.001$ in each case). We also found one bacterial OTU that was significantly more abundant in freshwater copepod microbiomes than in their saline counterparts across clades: an unclassified

member of the family Saprospiraceae (Wald test, adjusted $P = 0.02$). Further testing upon higher taxonomic levels revealed that OTUs from one order of the class Alphaproteobacteria, an uncharacterized order identified as alfVIII in prior studies (Newton *et. al* 2011), were more abundant in freshwater copepod microbiomes than saline copepod microbiomes (Wald test, adjusted $P = 1.5 \times 10^{-3}$).

Bacterial Taxa Enriched in the Copepod Microbiome

Even given the high variability in microbiome composition among copepod microbiomes, several bacterial OTUs were unambiguously associated with copepods in multiple populations and diverse environments (Table 2, adjusted *P* cutoff = 0.01). An unclassified member of the *Leadbetterella* genus (Family *Cytophagaceae*) was found in every copepod microbiome, albeit at low abundance (between 0.014% and 3.19% of the total population). This unclassified *Leadbetterella* was present in all copepod microbiomes across salinity, clade, continent, and season, and was entirely absent from the free-living bacterioplankton. Members of the genera *Flavobacterium*, *Rubrivivax*, and *Aeromonas* were also significantly enriched across all copepod samples. In general, the copepod microbiome was significantly enriched for bacterial lineages that included pathogens of a wide variety of hosts, including those of humans and fish. Putative pathogenic taxa enriched in at least one copepod microbiome included members of the genera *Vibrio*, *Flavobacterium*, *Chryseobacterium, Aeromonas*, *Pseudomonas*, and *Rickettsia.*

Clade-Specific Taxa in Copepod Microbiomes

The results of permANOVA testing did not identify the clade of the sample (see Fig. 1) as a significant factor in copepod microbiome similarity ($P = 0.09$). However, we did see evidence for clade-specific taxa. All copepod microbiome samples from the Atlantic clade (Fig. 1, A, red)

contained a member of the genus *Rickettsia* that was absent from all microbiomes of other copepod populations. We also identified one OTU, *Brevundimonas intermedia,* that was present in most copepod microbiomes from the Atlantic and European clades (Fig. 1, A and C, red and purple, respectively), but absent from the Gulf clade (Fig. 1, B, green). We found three OTUs that were significantly more abundant in the European clade relative to the other two North American clades. Two were members of the family Neisseriaceae and of the phylum Cyanobacteria, respectively. The other was a member of the phylum Proteobacteria, but its sequence was distinct enough from the reference sequences in the Greengenes database to preclude classification even to the level of class. The closest hits to this sequence in the NCBI database were bacteria of the genus *Holospora.*

Discussion

This study makes a major and unique contribution to the study of microbiomes our exploration of changes in an animal host microbiome across replicated habitat transitions, in this case of an invasive species across multiple independent saline to freshwater habitat transitions. This study is the first to compare the microbiome of a host species *versus* its surrounding free-living bacterioplankton community across a biogeographic barrier. We uncovered a diverse microbiome associated with *E. affinis* that was highly distinct from the bacterial community in the surrounding water. Several bacterial OTUs (<3% sequence divergence at the 16S rRNA gene) were significantly enriched in the copepod microbiome, including one which was present in every copepod sample. Our permANOVA testing found that salinity was the most significant factor structuring similarity between copepod microbiomes. This, in addition to our observation of bacterial OTUs that shift from saline to freshwater microbiomes across several independent invasive clades, provides compelling evidence for parallel shifts in the composition of copepod

microbiomes across saline-to-freshwater invasions. Such parallelism is unexpected, as it suggests that the *E. affinis* microbiome experienced dramatic and convergent changes in composition during independent shifts from saline to freshwater habitats.

Salinity Structures Animal Host Microbiomes, in addition to Water Microbiomes

Our study is the first to discover that salinity structures animal host communities. Our initial hypothesis was that copepod microbiomes would be most similar within genetically-distinct clades, and that environmental factors such as salinity and temperature would be of secondary importance. Instead, we found that salinity was the dominant factor structuring copepod bacterial community composition. Both MDS analysis (Fig. 9) and permANOVA testing provided statistically significant support for this pattern. While free-living microbial communities are known to be structured by salinity (Lozupone & Knight 2007, see next paragraph), it is notable that the saline to freshwater boundary imposes a substantial barrier even for host-associated microbes. The host environment might be expected to mitigate the challenges imposed by the salinity barrier by providing a protective buffer, such as constant hemolymph or gut conditions, of the copepod host (Tang *et al.* 2011; Lee *et al.* 2012).

This is a noteworthy finding; it suggests that copepod microbiome compositions converged independently in freshwater habitats across all three invasive lineages. In addition to structuring host communities, we found that salinity was the predominant environmental factor structuring free-living bacterioplankton communities (Fig. 3). These results were consistent with prior studies, which also had found that free-living microbial communities are predominantly structured by salinity (Lozupone & Knight 2007) and that microbial saline to freshwater transitions are rare (Logares *et al.* 2009). The influence of salinity on the composition of these

microbiomes highlights the magnitude of the saline/freshwater biogeographic barrier and the difficulty for bacterial taxa to cross this boundary.

Parallel Shifts in Host Microbiomes during Saline to Freshwater Invasions

Our finding that salinity structured the copepod host microbiomes suggests that copepod microbiome compositions converged independently in freshwater habitats across all three invasive lineages. Support for convergence in copepod microbiome composition during the saline to freshwater transition was further bolstered by the parallel shifts in abundance of specific bacterial taxa. For instance, four bacterial OTUs were significantly different in abundance between copepods in saline and those in fresh water across all invasive clades. Three of these taxa, all members of the family Rhodobacteraceae, were more abundant in saline copepod microbiomes than copepod microbiomes in fresh water. We also found an unclassified OTU of the family Saprospiraceae and several members of an Alphaproteobacterial order (alfVII) that were more abundant in freshwater copepods. In each of these four cases the OTUs in question were significantly more abundant in copepods *versus* the surrounding water, and in the case of the Saprospiraceae sp. the OTU was completely absent from all water microbiomes. This parallel acquisition of an OTU in freshwater leads us to the question of how freshwater copepod microbiomes acquired this same taxon across independent invasions when it was absent from the water column.

Other than these four OTUs, there were relatively few bacterial OTUs that were shared among only saline or among only freshwater microbiomes across multiple copepod populations. This result was puzzling, given that our MDS and permANOVA results clearly showed that the copepod microbiomes clustered by salinity. The phenomenon of copepod microbiome

convergence in freshwater did not seem to be driven by a subset of bacterial OTUs that were shared across most freshwater or most saline microbiomes. Rather, the clustering of copepod microbiomes by salinity seems to have been driven by the cumulative effect of many OTUs that were shared between only a few freshwater or a few saline microbiomes.

It has been shown conclusively that *E. affinis* has invaded freshwater habitats independently in multiple geographical locations from divergent host clades, and that freshwater populations are derived independently from different ancestral populations (Lee 1999). Convergence among the freshwater microbiomes, then, is not driven by evolutionary relationships among copepod populations.

One possible explanation for the similarity among freshwater microbiomes of different *E. affinis* populations is that members of *E. affinis* freshwater populations and their associated microbiomes are transported between freshwater reservoirs in ship ballast water. However, we find no evidence for admixture or geographic mixing of invasive clades in our *E. affinis* population sampling (Carol E. Lee, unpublished data). A much more plausible, explanation for the similarity among freshwater copepod microbiomes is that upon introduction to freshwater, invasive *E. affinis* populations acquire bacteria which are associated with freshwater zooplankton hosts in their native environment. In that case, what we observe in our sampling may be bacteria from what could be considered a global freshwater zooplankton bacterial community, distinct from the saline zooplankton bacterial community.

The Copepod Microbiome is Distinct from the Surrounding Water

By far the most dominant factor driving bacterial community similarity was whether microbiomes were copepod-associated or free-living (Fig. 2). In addition to our study, several

previous studies have also shown that host-associated microbial communities in aquatic environments are highly distinct from free-living bacterioplankton (Grossart 2010; Bickel & Tang 2014; De Corte *et al.* 2014; Shoemaker & Moisander 2015). Our finding that the copepod microbiome is distinct from the water and composed of taxa that persist across locations and salinities (see section "Bacterial Taxa Enriched in the Copepod Microbiomes" below), demonstrates that the copepod is a unique microhabitat in aquatic environments (Tang *et al.* 2010). The functions performed by bacteria in such a microhabitat, especially in the anaerobic copepod gut, might be expected to be significantly different from functions performed by the free-living bacterioplankton community. Given the enormity of the copepod biomass, uncovering the functions performed by the copepod microbiome will provide invaluable insights into aquatic ecosystems.

We found that copepod-associated bacterial communities were significantly less diverse than water bacterial communities, with fewer OTUs (<3% sequence divergence in the 16S rRNA gene) within copepod microbiomes as well as significantly fewer genera, families, orders, classes, and phyla than in water. The higher diversity in water was not only the result of a larger number of OTUs from a single overrepresented group (phylum Actinomycetes, for example), but was the result of an increase in OTUs present across a large diversity of bacterial lineages. These findings were consistent with previous studies showing that copepod-associated microbiomes tend to be lower diversity overall than their surrounding environment (De Corte *et al.* 2014; Shoemaker & Moisander 2015). Such host-associated communities tend to be less diverse than their surrounding water even though host-associated microbiomes can harbor orders of magnitude more bacterial cell counts than comparable volumes of water (Tang 2005; Møller *et al.* 2007).

Given that the water and copepod-associated microbiome samples in each location shared a common temperature, season, and salinity, we hypothesized that locations with more diverse water bacterial communities might also have more diverse copepod-associated communities. This was not the case, as we found no correlation between diversity of the free-living bacterioplankton community and that of the copepod microbiome (Fig. 13). This result was somewhat surprising, and shows that factors influencing the diversity of the water microbiome are likely not the same factors that influence copepod microbiome diversity. While some studies have shown that bacterial community alpha diversity is correlated with latitude and temperature (Fuhrman *et al.* 2008; Sul *et al.* 2013; Sunagawa *et al.* 2015), we did not observe significant effects of either factor in our data (Fig. 12).

Bacterial Taxa Enriched in, or Exclusive to, the Copepod Microbiome

OTUs that were more common in the copepod microbiome across all populations represent potential core members of the microbiome that might have been transported along with *E. affinis* during invasion events—or picked up multiple times independently. We observed several such OTUs, including one from the genus *Leadbetterella* and several from the genus *Flavobacterium,* among others (Table 2). Given their ubiquity in copepod microbiomes and their low abundance (or complete absence) in the water column, it is likely that these taxa exploit the unique microhabitat provided by the copepod and its anaerobic gut. This raises the possibility that certain host-microbe interactions are persistent and that some of the bacteria in the *E. affinis* microbiome have formed long-term associations with their copepod host.

The only bacterial OTU that was present in every copepod microbiome and entirely absent from the water column was a member of the genus *Leadbetterella*. The other characterized member of

the genus, *Leadbetterella byssophila,* was isolated from soil waste from mushroom cultivation and appears to be quite isolated from any related bacterial strains (Abt *et al.* 2011). *L. byssophila*, and most of the *Cytophagaceae* family to which it belongs, are aerobic gram-negative rods. Other members of the *Cytophagaceae* are routinely isolated from marine and aquatic environments (McBride *et al.* 2014). Comparatively little is known about this genus, although the presence of this taxon in all copepod microbiomes suggests a potential close association with *E. affinis*.

In addition to *Leadbetterella*, we found several other taxa that were primarily associated with copepods, and absent from the surrounding water, although none were present in all samples from all locations. *Rubrivivax gelatinosus* was highly enriched in the copepod microbiome. This bacterium is a facultative photoheterotroph, preferring to generate energy through photosynthesis, and is often isolated from anaerobic environments (Imhoff 2015). This species is also able to fix nitrogen (Hu *et al.* 2012), which lends some support to the hypothesis of an important role for the zooplankton microbiome to nitrogen-fixation in the aquatic environment (Mahaffey *et al.* 2005). The reduction of nitrogen is a fundamentally important ecosystem process and requires an oxygen-depleted environment. Given the enormous biomass of copepods in aquatic environments and the anaerobic copepod gut, the copepod microbiome might contribute greatly to large-scale ecosystem processes *via* the action of bacterial constituents that can fix nitrogen.

There were several other potentially interesting OTUs enriched in copepod microbiomes. One shared significant sequence identity with a member of the family Saprospiraceae, *Phaeodactylibacter luteus.* Our data showed that this OTU was present in many copepod microbiomes and absent from the water. *P. luteus* was originally isolated from a *Picochlorum*

algal culture from the Indian Ocean (Chen *et al.* 2015). The presence of this taxon could suggest that we are observing bacteria that are associated with the algal community upon which the copepod feeds, or that this species is associated with copepods in addition to algae. We also found a member of the phylum Proteobacteria associated with several copepod populations across clades and salinities. The closest sequences in nucleotide identity to this OTU in the NCBI database belong to symbionts of the terrestrial isopod *Porcellio scaber* (common wood louse) of the order *Rickettsiales* (Anton-Erxleben *et al.* 2004). This result is intriguing, given that most members of the *Rickettsiales* are obligate intracellular symbionts of their hosts. In fact, we found another example of potential obligate intracellular symbionts in the copepod microbiome (see section below "Members of the Copepod Microbiome are Specific to Host Clades").

There were several prospective anaerobes that were enriched in copepod microbiomes. OTUs within the families Comomondaceae, Flavobacteriaceae, and Weeksellaceae were also found only in copepod microbiomes, although not in every population. Several of these taxa are members of anaerobic or facultatively anaerobic clades (Collins *et al.* 1983; Heidelberg *et al.* 2002b). However, many of these genera are relatively newly classified and not well-described, and the functions performed by these strains are unknown. Members of the phylum Firmicutes were significantly more abundant in copepod microbiomes than in the surrounding water. This result supports the findings of several prior studies, which also found an overrepresentation of Firmicutes in copepod microbiomes (Grossart *et al.* 2009; Homonnay *et al.* 2012; Shoemaker & Moisander 2015), and lends credence to the hypothesis that members of the Firmicutes form close associations with copepod hosts and take advantage of the hypoxic or anoxic conditions in and around the copepod gut (Shoemaker & Moisander 2015).

Interestingly, taxa from the genera *Pseudonocardia, Streptomyces,* and *Klebsiella,* which were present in all copepod-associated microbiomes in an initial round of 454 16S rRNA sequencing, were not present in our data (Gelembiuk 2015). This may be due to seasonal fluctuations in the copepod microbiome, or the use of different 16S primer sets for amplicon generation, as different 16S primers are known to bias the taxonomic representation of a sample due to variation in 16S sequence specificity (Suzuki & Giovannoni 1996; Sipos *et al.* 2007). In either case, our use of specific primers and samples from individual time-points almost certainly fails to capture the full picture of bacterial diversity within the *E. affinis* microbiome. Future studies might sample *E. affinis* microbiomes from a time series across seasons and sequence with multiple primer sets to provide a more comprehensive picture of microbiome diversity.

Bacterial Taxa Enriched in, or Exclusive to, the Water Microbiome

In addition to OTUs that were overrepresented in copepod microbiomes, there were an even greater number of OTUs that were significantly overrepresented in water samples (Table 3). Such OTUs are dominated by members of the phylum Actinobacteria. Members of the Actinobacteria are major constituents of freshwater (Glockner *et al.* 2000; Sekar *et al.* 2003; Newton *et al.* 2011) and estuarine environments (Riemann *et al.* 2008), and could change in abundance across a salinity gradient (Kirchman *et al.* 2005; Stevens *et al.* 2007), although we did not see a salinity response in our data. Members of the Actinobacteria are by no means the only taxa that were more abundant in the water than in the copepod, as OTUs from nearly all phyla that were present in our sampling were overrepresented in the water samples (Table 3). The finding of an excess of OTUs enriched in the water was consistent with our observation of overall higher diversity of microbial taxa in the bacterioplankton (Fig. 7), as a greater number of bacterial taxa in the water would result in greater numbers of taxa that are unique to water. The

pattern was also due to the high variability in taxonomic composition among copepod microbiomes (e.g. Figs. 5 and 6). Since microbial communities vary greatly in composition among copepod populations, there were fewer bacterial OTUs that could be identified as enriched across all copepod microbiomes.

Members of the Copepod Microbiome are Specific to Host Clades

A particularly intriguing bacterial taxon, found in only one *E. affinis* clade, was a member of the *Rickettsia* genus of endosymbionts. This taxon was present in all copepod microbiomes from the Atlantic clade (Fig. 1, clade A) but was absent from the surrounding water. The most closely related bacterial species in the Greengenes database was a taxon identified as an endosymbiont of the water beetle *Deronectes platynotus* and other members of *Deronectes* (Küchler *et al.* 2009). Members of the genus *Rickettsia* are obligate intracellular endosymbionts of a diverse array of taxa, including many members of the Arthropoda. They are often parasites, sometimes mutualists, and can be horizontally transmitted, although vertical transmission is the more common means of acquisition (Perlman *et al.* 2006; Weinert 2015). Members of this genus are best known as the causative agents of typhus and Rocky Mountain spotted fever in humans, diseases that are acquired via transmission of *Rickettsia* endosymbionts from haematophagous arthropod hosts such as ticks, lice, and fleas to humans (Raoult & Roux 1997; Parola *et al.* 2005). Even in their natural hosts, members of the genus are often pathogenic.

Rickettsia are inherited vertically through the cytoplasm of the host's eggs, but not through male gametes. Thus, members of the genus are propagated through females only and can spread though the host population by increasing the frequency of females in the population, regardless of their overall effect on host fitness (O'Neill *et al.* 1997). In many cases, this has led *Rickettsia*

to manipulate their hosts' reproduction through a variety of means including skewing sex ratio in favor of females (Takahashi *et al.* 1997; Himler *et al.* 2011), inducing parthenogenesis (Giorgini *et al.* 2010), or killing male hosts (Werren *et al.* 1994; Majerus *et al.* 2000; Lawson *et al.* 2001). While many members of the *Rickettsia* genus are pathogenic, the genus also contains many examples of mutualists as well as members whose relationship with their host is unknown (Weinert 2015). The uneven distribution of these endosymbionts across *E. affinis* clades is puzzling and may suggest a relatively recent acquisition of the *Rickettsia* endosymbiont in the Atlantic clade subsequent to the divergence of *E. affinis* populations. The potential for host reproductive manipulation and especially the presence of this taxon in only one lineage of *E. affinis* warrant further investigation.

A few bacterial taxa were significantly more abundant in the European *E. affinis* microbiome (Fig. 1, clade C), the most dominant of which was a member of the genus *Holospora*. Other characterized members of this genus are intranuclear parasites of protozoan genus *Paramecium* (Fokin & Gortz 2009). These bacteria appear to have species specificity and differ in nuclear localization within their hosts, where some species preferentially infect the macronucleus and some infect the micronucleus (Gromov & Ossipov 1981). The occurrence of this bacterial OTU might reflect the presence of members of *Paramecium* in the copepod microflora, or perhaps that *Paramecium* are a food source, as copepods are known to consume ciliated protists (Williamson 1980; Wiackowski *et al.* 1994).

The Copepod as a Potential Waterborne Disease Vector

Previous studies have identified potential pathogens in the *E. affinis* microbiome (Gelembiuk 2015) and have shown that copepods in general may serve as vectors for agents of human

disease, specifically cholera (Huq *et al.* 2005; Turner *et al.* 2014). In addition to the members of *Rickettsia* and *Clostridium* that we found in copepod microbiomes, members of the genus *Flavobacterium* and *Chryseobacterium* were especially abundant in the copepod microbiome and were significantly enriched in copepod-associated samples *vs*. water samples. *Flavobacterium columnare* is a major pathogen of freshwater fish, including salmon and trout (Austin & Austin 2007; Loch *et al.* 2013; Loch & Faisal 2014) and was highly abundant in some of our copepod microbiomes. *Chryseobacterium piscicola* has also been implicated in the diseases of farmed fish (Ilardi *et al.* 2009). This raises the interesting question of whether zooplankton, and copepods in particular, serve as common vectors for fish diseases (See Chapter 2).

This is the first study to examine the microbiome of an individual zooplankton in a comparative context across a major biogeographic barrier, and to contrast host-associated microbiomes with free-living bacterioplankton in such a context. We have found that the *Eurytemora affinis* microbiome is highly distinct from the surrounding water, and that salinity structures this host-associated bacterial community more than any other factor. The question remains as to whether individual members of the microbiome are inherited vertically or acquired environmentally. In order to address this question, future studies might examine the evolutionary history of individual members of the bacterial community and to compare the *E. affinis* microbiome with the microbiome associated with other members of the zooplankton community.

Acknowledgments

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256259. Funds for this project were provided by National Science Foundation OCE- 1046372 to Carol E. Lee and OCE-1046371 to Joana C.

Silva. The authors would like to acknowledge the undergraduate research assistants in the Carol Lee lab at the University of Wisconsin-Madison, whose work in the field and in the lab was invaluable for this project. Kristin Lee, Mackenzie Carlson, and Joseph Connolly assisted with collections in the field. Former members of the Lee lab, including Quinn Langdon, Soo Hyung Kim, and Marijan Posavi contributed to this project. Gregory Gelembiuk's previous results on the copepod microbiome and methods for copepod sampling formed the basis for this work.

Literature Cited

Abt B, Teshima H, Lucas S *et al.* (2011) Complete genome sequence of Leadbetterella byssophila type strain (4M15). *Standards in Genomic Sciences*, **4**, 2–12.

Adams AS, Currie CR, Cardoza Y, Klepzig KD, Raffa KF (2009) Effects of symbiotic bacteria and tree chemistry on the growth and reproduction of bark beetle fungal symbionts. *Canadian Journal of Forest Research*, 39, 1133–1147.

Anderson MJ (2001) A new method for non parametric multivariate analysis of variance. *Austral Ecology*, 26, 32–46.

Anton-Erxleben F, Zimmer M, Brune A, Wang Y, Stingl U (2004) Candidatus Hepatincola porcellionum gen. nov., sp. nov., a new, stalk-forming lineage of Rickettsiales colonizing the midgut glands of a terrestrial isopod. *Archives of Microbiology*, 181, 299–304.

Austin B, Austin DA (2007) *Bacterial Fish Pathogens: Disease of Farmed and Wild Fish*. Springer Science & Business Media.

Bickel SL, Tang KW (2014) Zooplankton-associated and free-living bacteria in the York River, Chesapeake Bay: comparison of seasonal variations and controlling factors. *Hydrobiologia*, 722, 305–318.

Bokulich NA, Subramanian S, Faith JJ *et al.* (2012) Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10, 57–59.

Bordenstein SR, Theis KR (2015) Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLOS Biol*, 13, e1002226.

Brandin E, Pistole T (1985) Presence of microorganisms in hemolymph of the horseshoe crab Limulus polyphemus. *Applied and Environmental Microbiology*, 49, 718–720.

Bravo J a, Forsythe P, Chew M V *et al.* (2011) Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proceedings of the National Academy of Sciences*, 108, 16050–16055.

Bray JR, Curtis JT (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27, 325–349.

Brune A, Ohkuma M (2010) Role of the Termite Gut Microbiota in Symbiotic Digestion. In: *Biology of Termites: a Modern Synthesis* (eds Bignell ED, Roisin Y, Lo N), pp. 439–475. Springer Netherlands, Dordrecht.

Cardoza YJ, Klepzig KD, Raffa KF (2006) Bacteria in oral secretions of an endophytic insect inhibit antagonistic fungi. *Ecological Entomology*, 31, 636–645.

Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, 11, 265–270.

Chao A, Chazdon RL, Colwell RK, Shen T-J (2004) A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecology Letters*, 8, 148–159.

Chen Z, Chen Y, Zheng T *et al.* (2015) Phaeodactylibacter luteus sp. nov., isolated from the oleaginous microalga Picochlorum sp. *International Journal of Systematic and Evolutionary Microbiology*, 65, 2666–2670.

Collins M, Lund B, Farrow J, Schleifer K (1983) Chemotaxonomic Study of an Alkalophilic Bacterium, *Exiguobacterium aurantiacum* gen. nov., sp. nov. *Journal of General Microbiology*, 129, 2037–2042.

De Corte D, Lekunberri I, Sintes E *et al.* (2014) Linkage between copepods and bacteria in the North Atlantic Ocean. *Aquatic Microbial Ecology*, 72, 215–225.

Currie CR, Wong B, Stuart AE *et al.* (2003) Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science*, 299, 386–8.

Dauvin J, Dodson J (1990) Relationship between feeding incidence and vertical and longitudinal distribution of rainbow smelt larvae(Osmerus mordax) in a turbid well-mixed estuary. *Marine Ecology Progress Series*, 60, 1–12.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, 72, 5069–5072.

Dunbar HE, Wilson ACC, Ferguson NR, Moran NA (2007) Aphid Thermal Tolerance Is Governed by a Point Mutation in Bacterial Symbionts (C Godfray, Ed,). *PLoS Biology*, 5, 1006–1015.

Dziallas C, Grossart H-P, Tang KW, Nielsen TG (2013) Distinct Communities of Free-Living and Copepod-Associated Microorganisms along a Salinity Gradient in Godthåbsfjord, West Greenland. *Arctic, Antarctic, and Alpine Research*, 45, 471–480.

Edgar RC (2013) UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nature Methods*, 10, 996–998.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200.

Endres D, Schindelin J (2003) A new metric for probability distributions. *IEEE Transactions on Information Theory*, 49, 1858–1860.

Engel P, Moran NA (2013) The gut microbiota of insects – diversity in structure and function. *FEMS Microbiology Reviews*, 37, 699–735.

Ezenwa VO, Gerardo NM, Inouye DW, Medina M, Xavier JB (2012) Animal Behavior and the Microbiome. *Science*, 338, 198–199.

Fadrosh DW, Ma B, Gajer P *et al.* (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, 2, 6.

Flint HJ, Scott KP, Duncan SH, Louis P, Forano E (2012) Microbial degradation of complex carbohydrates in the gut. *Gut Microbes*, 3, 289–306.

Fokin SI, Gortz H-D (2009) Diversity of Holospora Bacteria in Paramecium and Their Characterization. In: *Endosymbionts in Paramecium* Microbiology Monographs. (ed Fujishima M). Springer, Berlin, Heidelberg.

Fuhrman JA, Steele JA, Hewson I *et al.* (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proceedings of the National Academy of Sciences*, 105, 7774–7778.

Gelembiuk, G. W. (2015). Departing the Sea: Molecular Ecology of Freshwater Invasions by Saltwater Animals. Department of Zoology. University of Wisconsin, Madison, WI.

Gerber L, Lee CE, Grousset E *et al.* (2016) The Legs Have It: In Situ Expression of Ion Transporters V-Type H + -ATPase and Na + /K + -ATPase in the Osmoregulatory Leg

Organs of the Invading Copepod Eurytemora affinis. *Physiological and Biochemical Zoology*, 89, 233–250.

Gerdts G, Brandt P, Kreisel K *et al.* (2013) The microbiome of North Sea copepods. *Helgoland Marine Research*, 67, 757–773.

Giorgini M, Bernardo U, Monti MM, Nappo AG, Gebiola M (2010) Rickettsia Symbionts Cause Parthenogenetic Reproduction in the Parasitoid Wasp Pnigalio soemius (Hymenoptera: Eulophidae). *Applied and Environmental Microbiology*, 76, 2589–2599.

Glockner FO, Zaichikov E, Belkova N *et al.* (2000) Comparative 16S rRNA Analysis of Lake Bacterioplankton Reveals Globally Distributed Phylogenetic Clusters Including an Abundant Group of Actinobacteria. *Applied and Environmental Microbiology*, 66, 5053–5065.

Gromov BV, Ossipov DV (1981) Holospora (ex Hafkine 1890) nom. rev., a Genus of Bacteria Inhabiting the Nuclei of Paramecia. *International Journal of Systematic Bacteriology*, 31, 348–352.

Grossart H-P (2010) Ecological consequences of bacterioplankton lifestyles: changes in concepts are needed. *Environmental Microbiology Reports*, 2, 706–714.

Grossart H-P, Dziallas C, Tang KW (2009) Bacterial diversity associated with freshwater zooplankton. *Environmental Microbiology Reports*, 1, 50–55.

Hagström Ä, Pinhassi J, Zweifel UL (2000) Biogeographical diversity among marine bacterioplankton. *Aquatic Microbial Ecology*, 21, 231–244.

Hansen B, Bech G (1996) Bacteria associated with a marine planktonic copepod in culture. I. Bacterial genera in seawater, body surface, intestines and fecal pellets and succession during fecal pellet degradation. *Journal of Plankton Research*, 18, 257–273.

Hardy A (1970) The Open Sea. In: *The World of Plankton*. Collins.

Harris JM (1993) The presence, nature, and role of gut microflora in aquatic invertebrates: a synthesis. *Microbial ecology*, 25, 195–231.

Heidelberg J, Heidelberg K, Colwell R (2002a) Bacteria of the gamma-subclass Proteobacteria associated with zooplankton in Chesapeake Bay. *Applied and environmental microbiology*, 68, 5498–5507.

Heidelberg JF, Paulsen IT, Nelson KE *et al.* (2002b) Genome sequence of the dissimilatory metal ion–reducing bacterium Shewanella oneidensis. *Nature Biotechnology*, 20, 1118–1123.

Himler AG, Adachi-Hagimori T, Bergen JE *et al.* (2011) Rapid Spread of a Bacterial Symbiont in an Invasive Whitefly Is Driven by Fitness Benefits and Female Bias. *Science*, 332, 254–256.

Homonnay ZG, Kéki Z, Márialigeti K, Tóth EM (2012) Bacterial communities in the gut of the freshwater copepod Eudiaptomus gracilis. *Journal of Basic Microbiology*, 52, 86–90.

Horn HS (1966) Measurement of“ overlap” in comparative ecological studies. *American Naturalist*, 100, 419–424.

Hu P, Lang J, Wawrousek K *et al.* (2012) Draft Genome Sequence of Rubrivivax gelatinosus CBS. *Journal of Bacteriology*, 194, 3262–3262.

Huq A, Sack RB, Nizam A *et al.* (2005) Critical Factors Influencing the Occurrence of Vibrio cholerae in the Environment of Bangladesh. *Applied and Environmental Microbiology*, 71, 4645–4654.

Ilardi P, Fernandez J, Avendano-Herrera R (2009) Chryseobacterium piscicola sp. nov., isolated from diseased salmonid fish. *International Journal of Systematic and Evolutionary Microbiology,* 59, 3001–3005.

Imhoff JF (2015) Incertae Sedis V. Rubrivivax. In: *Bergey's Manual of Systematics of Archaea and Bacteria*. John Wiley & Sons, Ltd, Chichester, UK.

Jaccard P (1912) The distribution of the flora in the alpine zone. *New phytologist*, 11, 37–50.

Kaltenpoth M (2009) Actinobacteria as mutualists: general healthcare for insects? *Trends in Microbiology*, 17, 529–535.

Keyhani NO, Roseman S (1999) Physiological aspects of chitin catabolism in marine bacteria. *Biochimica et Biophysica Acta*, 1473, 108–22.

Khlebovich VV, Abramova EN (2000) Some problems of crustacean taxonomy related to the phenomenon of Horohalinicum. *Hydrobiologia*, 417, 109–113.

Kimmel DG, Miller WD, Roman MR (2006) Regional scale climate forcing of mesozooplankton dynamics in Chesapeake Bay. *Estuaries and Coasts*, 29, 375–387.

Kirchman DL, Dittel AI, Malmstrom RR, Cottrell MT (2005) Biogeography of major bacterial groups in the Delaware Estuary. *Limnology and Oceanography*, 50, 1697–1706.

Küchler SM, Kehl S, Dettner K (2009) Characterization and localization of Rickettsia sp. in water beetles of genus Deronectes (Coleoptera: Dytiscidae). *FEMS Microbiology Ecology*, 68, 201–11.

Lawson ET, Mousseau TA, Klaper R, Hunter MD, Werren JH (2001) Rickettsia associated with male-killing in a buprestid beetle. *Heredity*, 86, 497.

Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod Eurytemora affinis. *Evolution*, 53, 1423–1434.

Lee CE (2016) Evolutionary mechanisms of habitat invasions, using the copepod Eurytemora affinis as a model system. *Evolutionary Applications*, 9, 248–270.

Lee CE, Kiergaard M, Gelembiuk GW, Eads BD, Posavi M (2011) Pumping Ions: Rapid Parallel Evolution of Ionic Regulation Following Habitat Invasions. *Evolution*, 65, 2229–2244.

Lee YK, Mazmanian SK (2010) Has the Microbiota Played a Critical Role in the Evolution of the Adaptive Immune System? *Science*, 330, 1768–1773.

Lee CE, Moss WE, Olson N *et al.* (2013) Feasting in fresh water: impacts of food concentration on freshwater tolerance and the evolution of food × salinity response during the expansion from saline into fresh water habitats. *Evolutionary Applications*, 6, 673–689.

Lee CE, Posavi M, Charmantier G (2012) Rapid evolution of body fluid regulation following independent invasions into freshwater habitats. *Journal of Evolutionary Biology*, 25, 625–633.

Lee CE, Remfert JL, Chang Y-M (2007) Response to selection and evolvability of invasive populations. *Genetica*, 129, 179–192.

Lee CE, Remfert JL, Gelembiuk GW (2003) Evolution of physiological tolerance and performance during freshwater invasions. *Integrative and Comparative Biology*, 43, 439.

Lee YK, Mazmanian SK (2010) Has the Microbiota Played a Critical Role in the Evolution of the Adaptive Immune System? *Science*, 330, 1768–1773.

Ley RE, Hamady M, Lozupone C *et al.* (2008) Evolution of Mammals and Their Gut Microbes. *Science*, 320, 1647–1651.

Loch T, Faisal M (2014) Deciphering the biodiversity of fish-pathogenic Flavobacterium spp. recovered from the Great Lakes basin. *Diseases of Aquatic Organisms*, 112, 45–57.

Loch TP, Fujimoto M, Woodiga S a. *et al.* (2013) Diversity of fish-associated flavobacteria of Michigan. *Journal of Aquatic Animal Health*, 25, 149–164.

Logares R, Bråte J, Bertilsson S *et al.* (2009) Infrequent marine–freshwater transitions in the microbial world. *Trends in Microbiology*, 17, 414–422.

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

Lozupone C, Knight R (2005) UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, 71, 8228–8235.

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

Lozupone C, Lladser ME, Knights D, Stombaugh J, Knight R (2011) UniFrac: an effective distance metric for microbial community comparison. *The ISME Journal*, 5, 169–172.

Mahaffey C, Michaels A, Capone D (2005) The conundrum of marine N2 fixation. *American Journal of Science*, 305, 546–595.

Majerus MEN, Hinrich J, Schulenburg GVD, Zakharov IA (2000) Multiple causes of male-killing in a single sample of the two-spot ladybird, Adalia bipunctata (Coleoptera: Coccinellidae) from Moscow. *Heredity*, 84, 605–609.

McBride MJ, Liu W, Lu X, Zhu Y, Zhang W (2014) The Family Cytophagaceae. In: *The Prokaryotes*, pp. 577–593. Springer Berlin Heidelberg, Berlin, Heidelberg.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data (M Watson, Ed,). *PLoS ONE*, 8, e61217.

McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible (AC McHardy, Ed,). *PLoS Computational Biology*, 10, e1003531.

Møller EF, Riemann L, Søndergaard M (2007) Bacteria associated with copepods: abundance, activity and community composition. *Aquatic Microbial Ecology*, 47, 99–106.

Mountford M (1962) An index of similarity and its application to classificatory problems. *Progress in Soil Zoology*, 43, 50.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews*, 75, 14–49.

O'Neill SL, Hoffman AA, Werren JH (1997) *Influential passengers: inherited microorganisms and arthropod reproduction.* Oxford University Press.

Parola P, Paddock CD, Raoult D (2005) Tick-Borne Rickettsioses around the World: Emerging Diseases Challenging Old Concepts. *Clinical Microbiology Reviews*, 18, 719–756.

Peitsch A, Köpcke B, Bernát N (2000) Long-term investigation of the distribution of Eurytemora affinis (Calanoida; Copepoda) in the Elbe Estuary. *Limnologica - Ecology and Management of Inland Waters*, 30, 175–182.

Perlman SJ, Hunter MS, Zchori-Fein E (2006) The emerging diversity of Rickettsia. *Proceedings of the Royal Society B*, 273, 2097–2106.

Price MN, Dehal PS, Arkin AP (2010) FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments (AFY Poon, Ed,). *PLoS ONE*, 5, e9490.

R Core Team (2014) R: A Language and Environment for Statistical Computing.

Raoult D, Roux V (1997) Rickettsioses as paradigms of new or emerging infectious diseases. *Clinical Microbiology Reviews*, 10, 694–719.

Riemann L, Leitet C, Pommier T *et al.* (2008) The Native Bacterioplankton Community in the Central Baltic Sea Is Influenced by Freshwater Bacterial Species. *Applied and Environmental Microbiology*, 74, 503–515.

Schloss PD, Gevers D, Westcott SL (2011) Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies (JA Gilbert, Ed,). *PLoS ONE*, 6, e27310.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, 75, 7537–7541.

Sekar R, Pernthaler A, Pernthaler J *et al.* (2003) An Improved Protocol for Quantification of Freshwater Actinobacteria by Fluorescence In Situ Hybridization. *Applied and Environmental Microbiology*, 69, 2928–2935.

Shaheen P a., Stehlik LL, Meise CJ *et al.* (2001) Feeding behavior of newly settled winter flounder (Pseudopleuronectes americanus) on calanoid copepods. *Journal of Experimental Marine Biology and Ecology*, 257, 37–51.

Shannon, C.E., Weaver, W. (1948). The mathematical theory of communication. The University of Illinois Press, Urbana, IL. pp. 117.

Shoemaker KM, Moisander PH (2015) Microbial diversity associated with copepods in the North Atlantic subtropical gyre. *FEMS Microbiology Ecology*, 91, 1–11.

Shropshire JD, Bordenstein SR (2016) Speciation by Symbiosis: the Microbiome and Behavior. *mBio*, 7, e01785–15.

Simpson E.H. 1949. Measurement of diversity. *Nature,*163:688.

Sipos R, Székely AJ, Palatinszky Márton *et al.* (2007) Effect of primer mismatch, annealing temperature and PCR cycle number on 16S rRNA gene-targetting bacterial community analysis. *FEMS Microbiology Ecology*, 60, 341–350.

Stackebrandt E, Goebel BM (1994) Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, 44, 846–849.

Stevens H, Brinkhoff T, Rink B, Vollmers J, Simon M (2007) Diversity and abundance of Gram positive bacteria in a tidal flat ecosystem. *Environmental Microbiology*, 9, 1810–1822.

Sul WJ, Oliver TA, Ducklow HW, Amaral-Zettler LA, Sogin ML (2013) Marine bacteria exhibit a bipolar distribution. *Proceedings of the National Academy of Sciences*, 110, 2342–2347.

Sunagawa S, Coelho LP, Chaffron S *et al.* (2015) Structure and function of the global ocean microbiome. *Science*, 348, 1261359–1261359.

Suzuki M, Giovannoni S (1996) Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Applied and Environmental Microbiology*, 62, 625–630.

Takahashi M, Urakami H, Yoshida Y *et al.* (1997) No Title. *European Journal of Epidemiology*, 13, 19–23.

Tang K (2005) Copepods as microbial hotspots in the ocean: effects of host feeding activities on attached bacteria. *Aquatic Microbial Ecology*, 38, 31–40.

Tang KW, Glud RN, Glud A, Rysgaard S, Nielsen TG (2011) Copepod guts as biogeochemical hotspots in the sea: Evidence from microelectrode profiling of Calanus spp. *Limnology and Oceanography*, 56, 666–672.

Tang K, Turk V, Grossart H (2010) Linkage between crustacean zooplankton and aquatic bacteria. *Aquatic Microbial Ecology*, 61, 261–277.

Turner JW, Malayil L, Guadagnoli D, Cole D, Lipp EK (2014) Detection of V ibrio parahaemolyticus , V ibrio vulnificus and V ibrio cholerae with respect to seasonal fluctuations in temperature and plankton abundance. *Environmental Microbiology*, 16, 1019–1028.

Verity P, Smetacek V (1996) Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Marine Ecology Progress Series*, 130, 277–293.

Viitasalo M, Flinkman J, Viherluoto M (2001) Zooplanktivory in the Baltic Sea: a comparison of prey selectivity by Clupea harengus and Mysis mixta, with reference to prey escape reactions. *Marine Ecology Progress Series*, 216, 191–200.

Weinert L (2015) The Diversity and Phylogeny of Rickettsia. In: *Parasite Diversity and Diversification* (eds Morand S, Krasnov BR, Littlewood TJ), pp. 150–181. Cambridge University Press.

Werren JH, Hurst GD, Zhang W *et al.* (1994) Rickettsial relative associated with male killing in the ladybird beetle (Adalia bipunctata). *Journal of Bacteriology*, 176, 388–394.

Wiackowski K, Brett MT, Goldman CR (1994) Differential effects of zooplankton species on ciliate community structure. *Limnology and Oceanography*, 39, 486–492.

Williamson CE (1980) The predatory behavior of Mesocyclops edax: Predator preferences, prey defenses, and starvation-induced changes. *Limnology and Oceanography*, 25, 902–909.

Winkler G, Sirois P, Johnson LE, Dodson JJ (2005) Invasion of an estuarine transition zone by Dreissena polymorpha veligers had no detectable effect on zooplankton community structure. *Canadian Journal of Fisheries and Aquatic Sciences*, 62, 578–592.

Yuan S, Cohen DB, Ravel J, Abdo Z, Forney LJ (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PloS one*, 7, e33865.

Zilber-Rosenberg I, Rosenberg E (2008) Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. *FEMS Microbiology Reviews*, 32, 723–735.

Tables and Figures

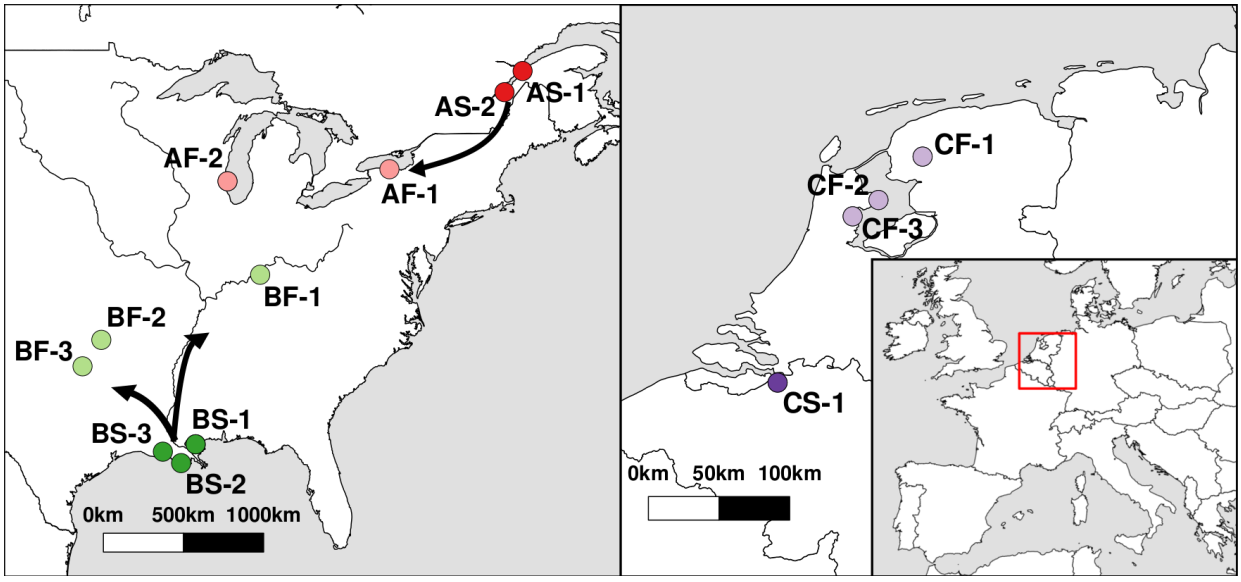

Figure 1 – Population sampling of *Eurytemora affinis* for this study – Samples include populations from three independent invasions (A, B, and C) of *E. affinis* originating from the St. Lawrence (Atlantic clade, A, red), Gulf of Mexico (Gulf Clade, B, green), and Baltic/Wadden Sea (European clade, C, purple) with multiple saltwater (S, dark) and freshwater (F, light) populations sampled from each invasion. From each location we sampled both copepod microbiomes (C) and water bacterial communities (W). See Table S1 for sample metadata.

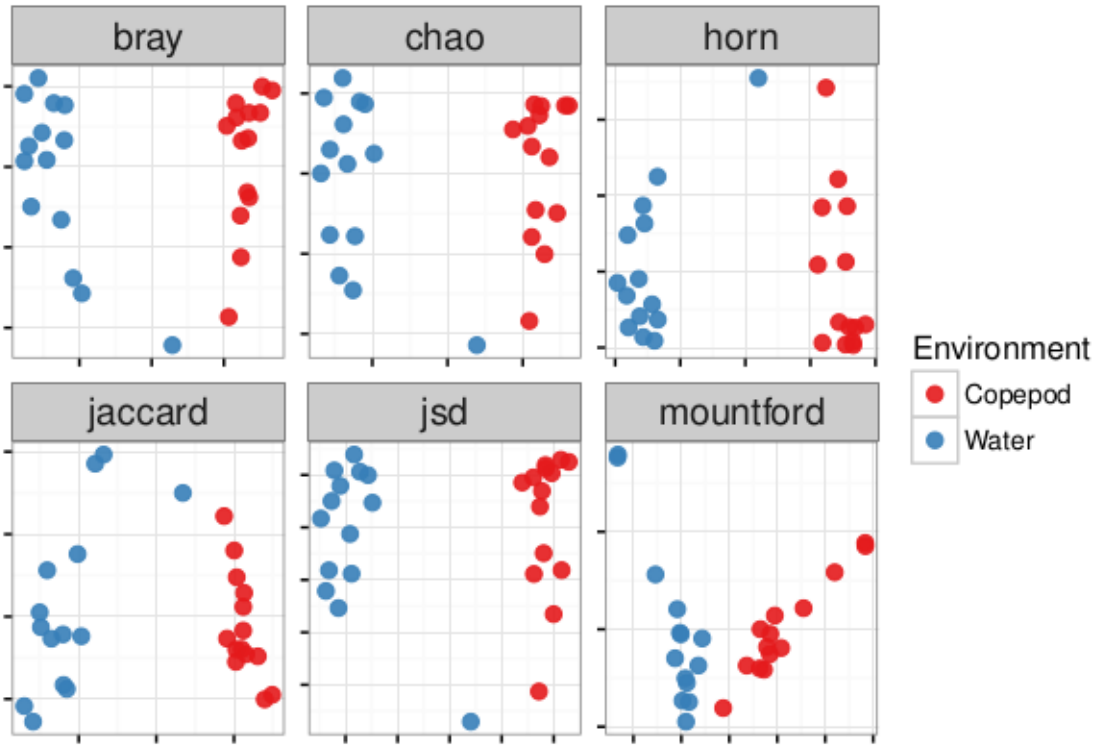

Figure 2 – Copepod-associated and water bacterial communities are distinct – Classical Multidimensional Scaling (MDS) clustering of copepod microbiome samples (red) and free-living water bacterial communties (blue) by six different distance measures (Bray-Curtis dissimilarity, Chao index, Horn-Morisita, Jaccard distance, Jensen-Shannon divergence, and Mountford index). Copepod-associated and water bacterial communities cluster independently into two distinct clusters no matter the distance metric.

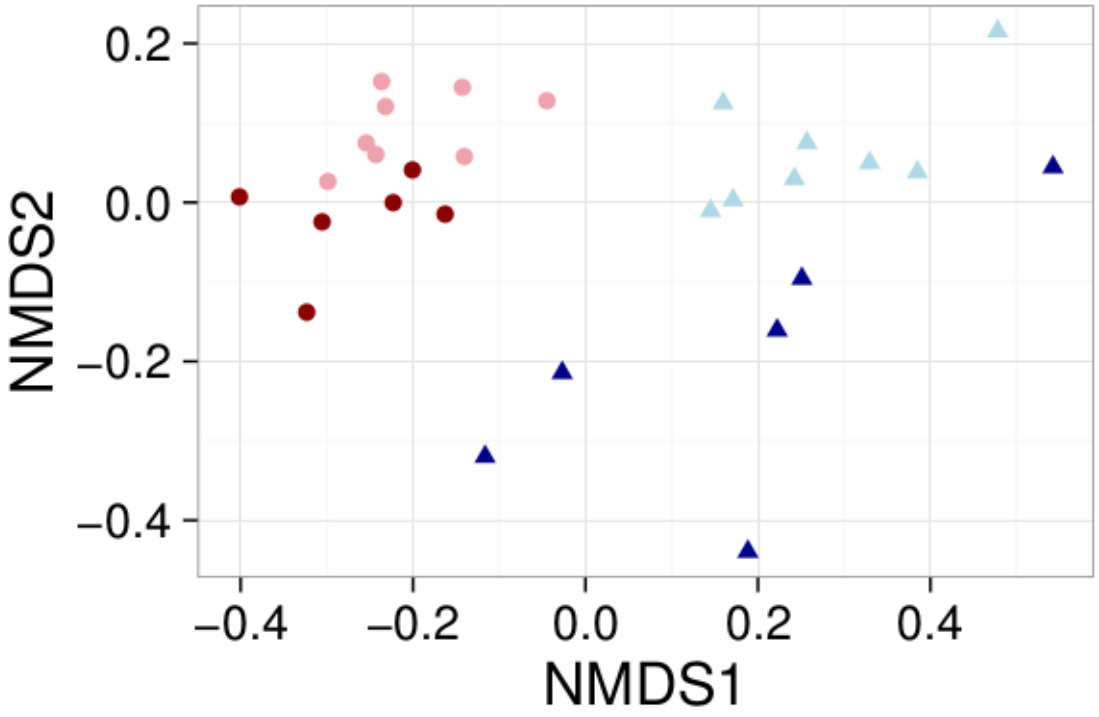

Figure 3 - Copepod and water microbiomes are both structured by salinity – Non-metric multidimensional scaling (NMDS) visualization of unweighted UniFrac distances between microbiomes. Both copepod (red) and water (blue) microbiomes cluster into groups of freshwater (light) and saline water (dark). This clustering of all microbiomes by salinity is significant via permANOVA testing (P = 0.038).

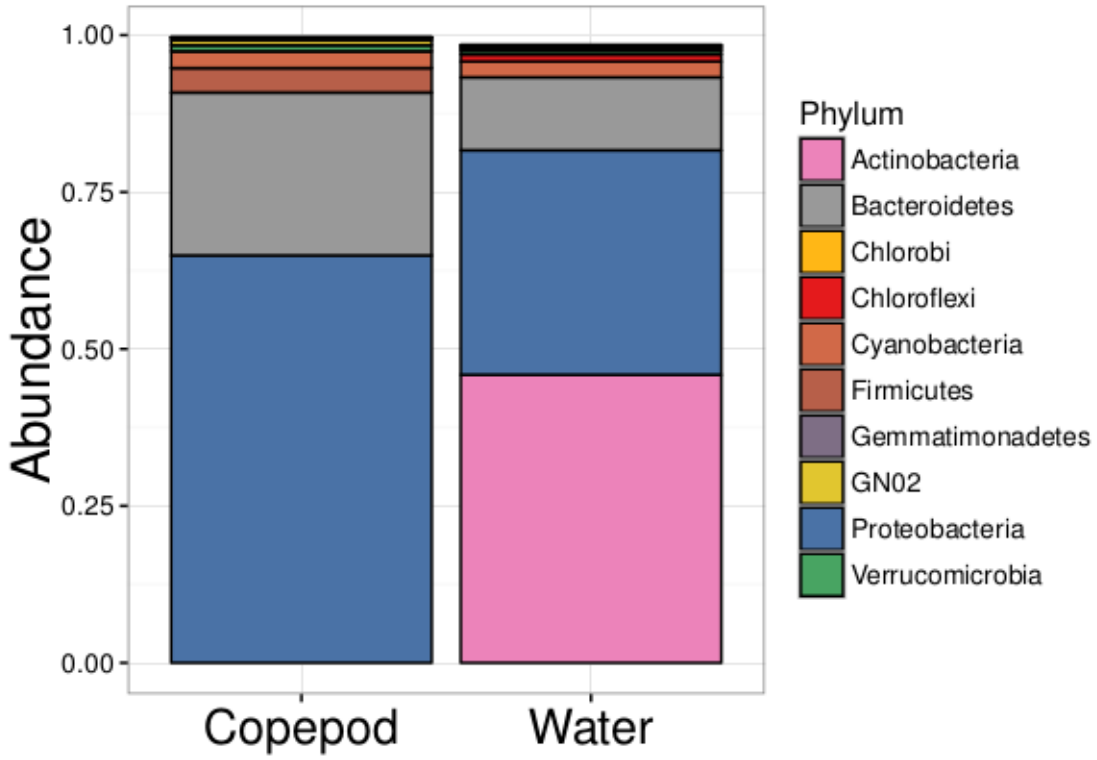

Figure 4 – Copepod and water microbiomes contain different phyla – Comparison of the abundance of the ten most common phyla between all copepod associated samples and all water bacterioplankton samples. Water bacterial communities are dominated by Actinobacteria. Copepod communities are dominated by Proteobacteria.

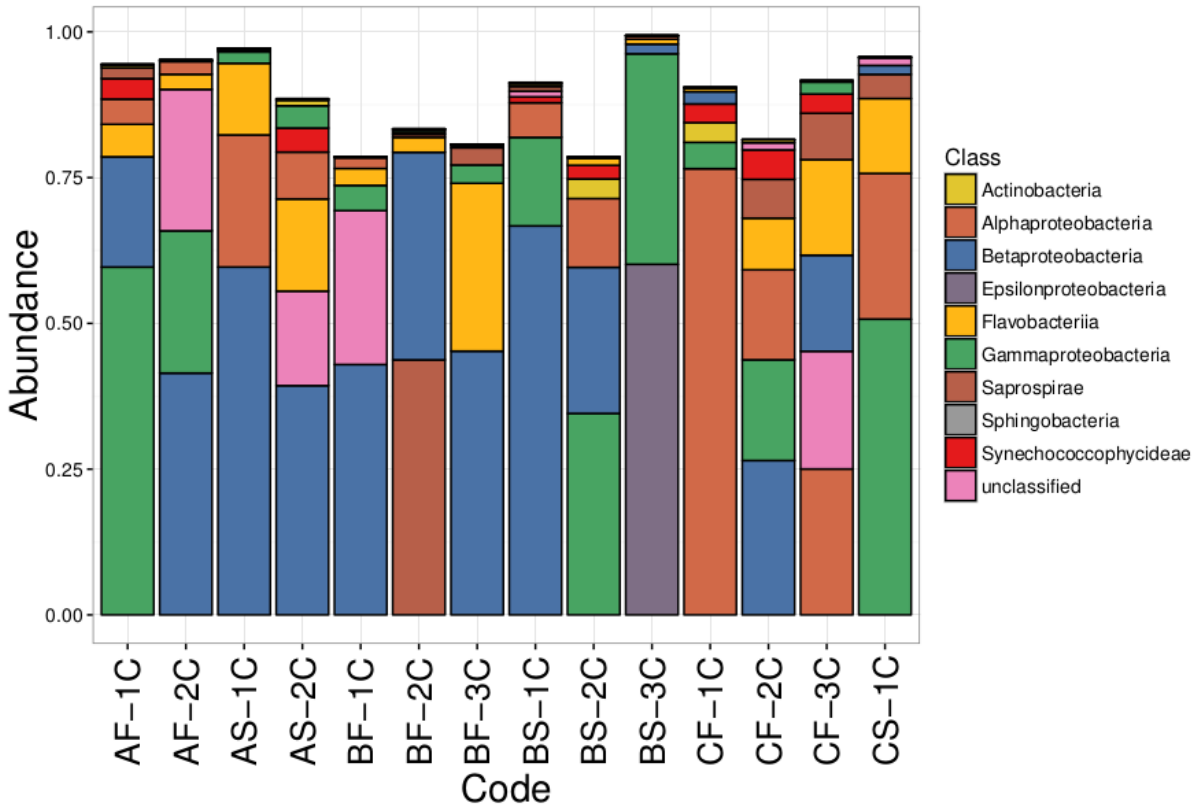

Figure 5 – Bacterial class abundance in copepods – The ten most common bacterial classes by relative abundance in copepod microbiomes. Bacterial classes vary greatly among copepod microbiome samples.

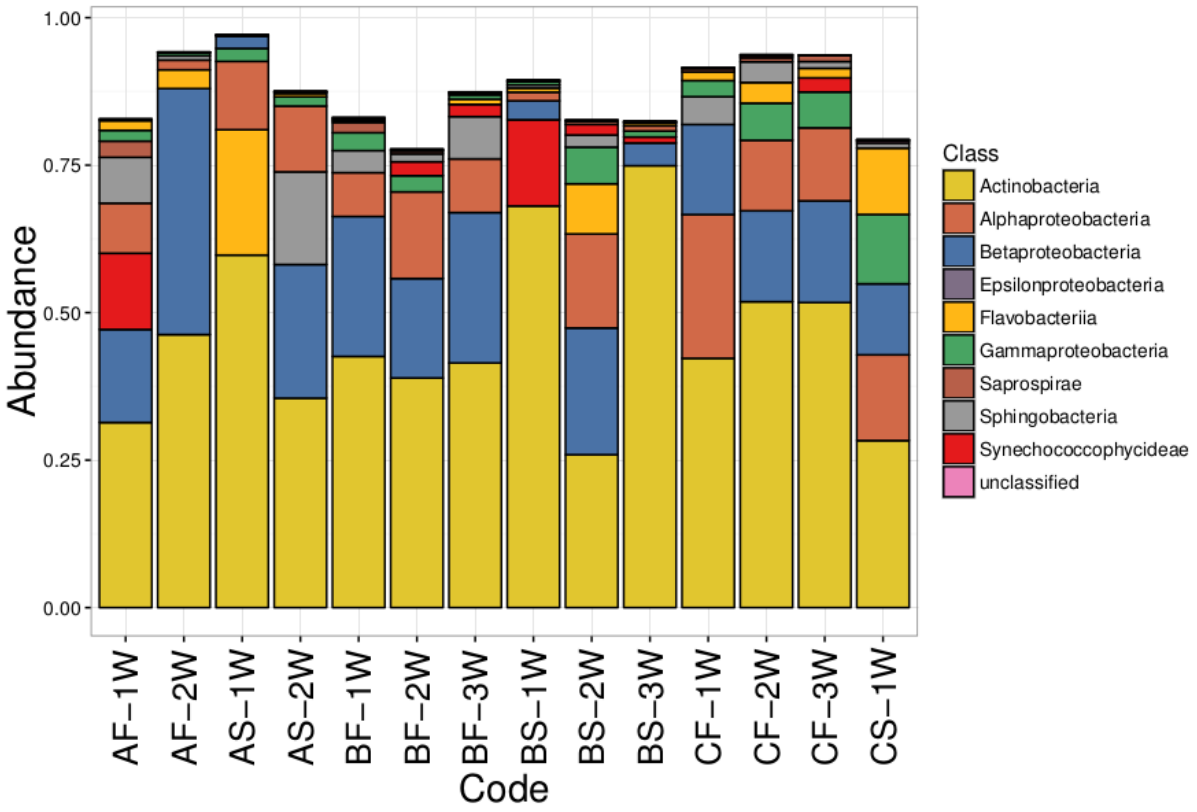

Figure 6 – Bacterial class abundance in water – The ten most common bacterial classes by relative abundance in water microbiomes. Water microbiomes vary less than copepod microbiomes in terms of relative class abundance (t-test, $P = 0.02$)

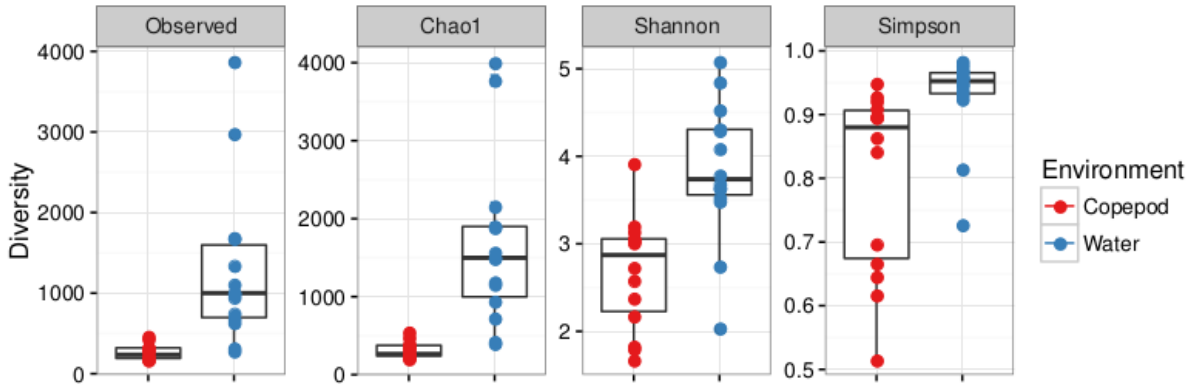

Figure 7 – Copepod-associated microbiomes are less diverse than water bacterial communities – Each panel represents a different measure of diversity (Observed OTUs at 97% 16S identity, Chao diversity, Shannon diversity, and Simpson diversity). A paired t-test between diversity in water and copepod associated samples shows that water bacterial communities (blue) are significantly more species-rich than copepod-associated microbial communities (red) (Chao diversity, $P = 2.5\text{x}10^{-4}$)

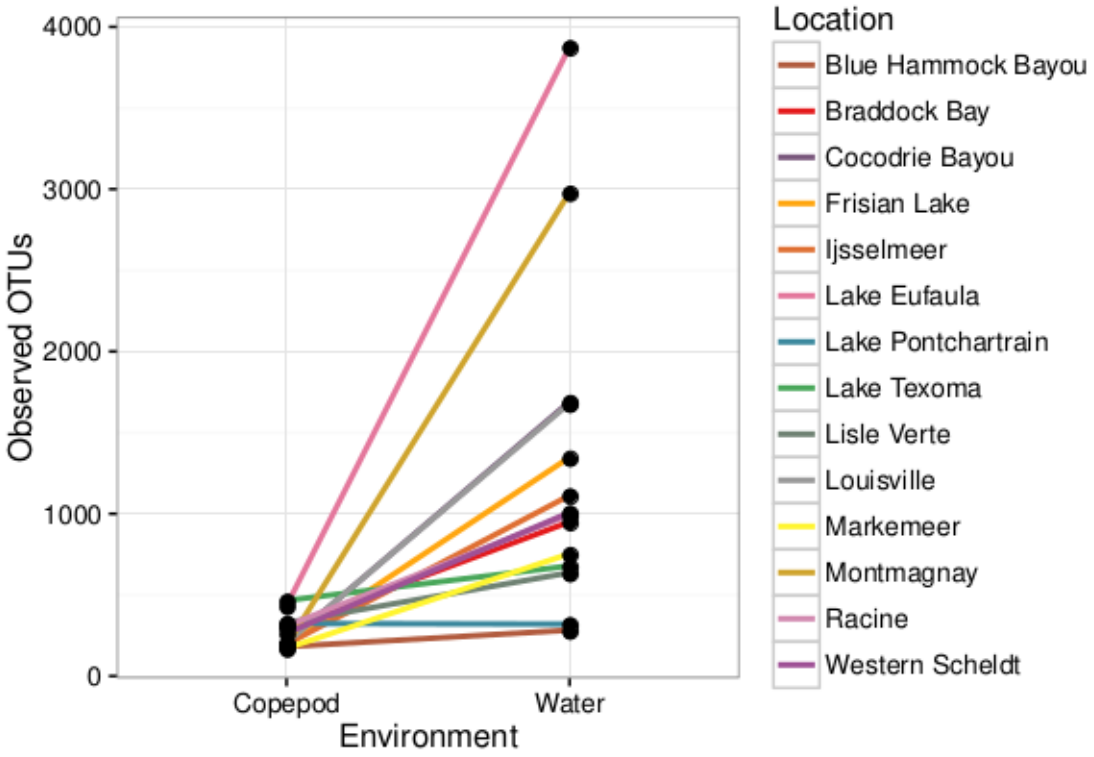

Figure 8 – Water microbiomes are more diverse than copepod microbiomes from the same location – The observed number of bacterial OTUs in paired copepod and water samples. Paired samples are connected by lines.

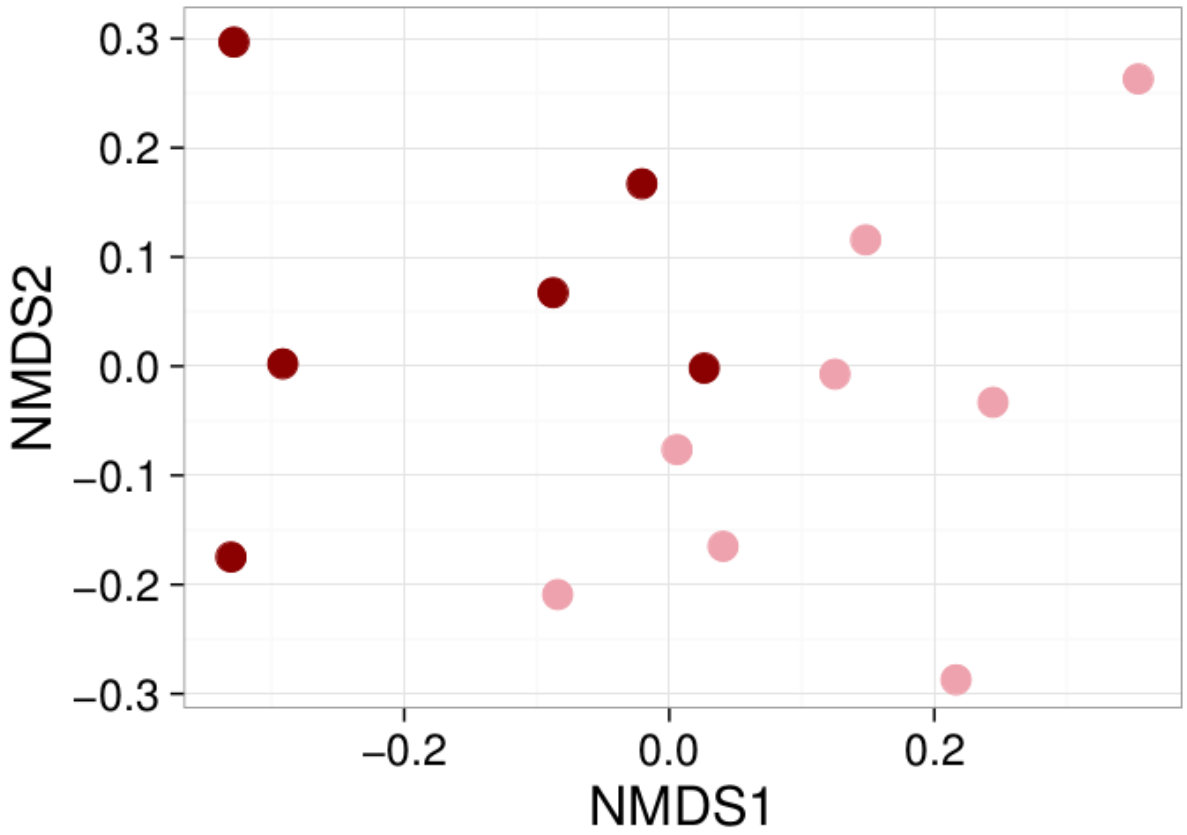

Figure 9 – Salinity is the primary driver of copepod microbiome similarity – Clustering of copepod microbiomes via non-metric multidimensional scaling based on UniFrac distances, unweighted by abundance. Copepod microbiomes in freshwater (light red) and those in saline water (dark red) form two distinct clusters. This clustering by salinity is significant via permANOVA testing ($P = 0.007$).

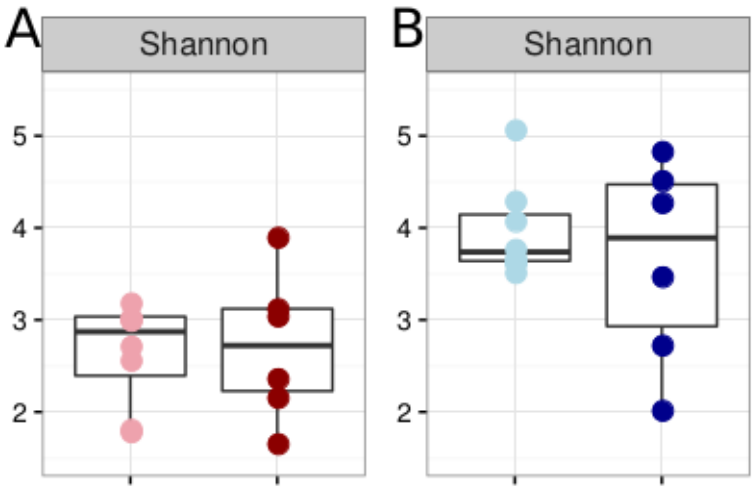

Figure 10 – Salinity does not affect measures of within-sample (alpha) diversity – Neither copepod microbiomes (A) nor water microbiomes (B) differ in diversity between freshwater (light colored) and saltwater (dark colored) sampling locations.

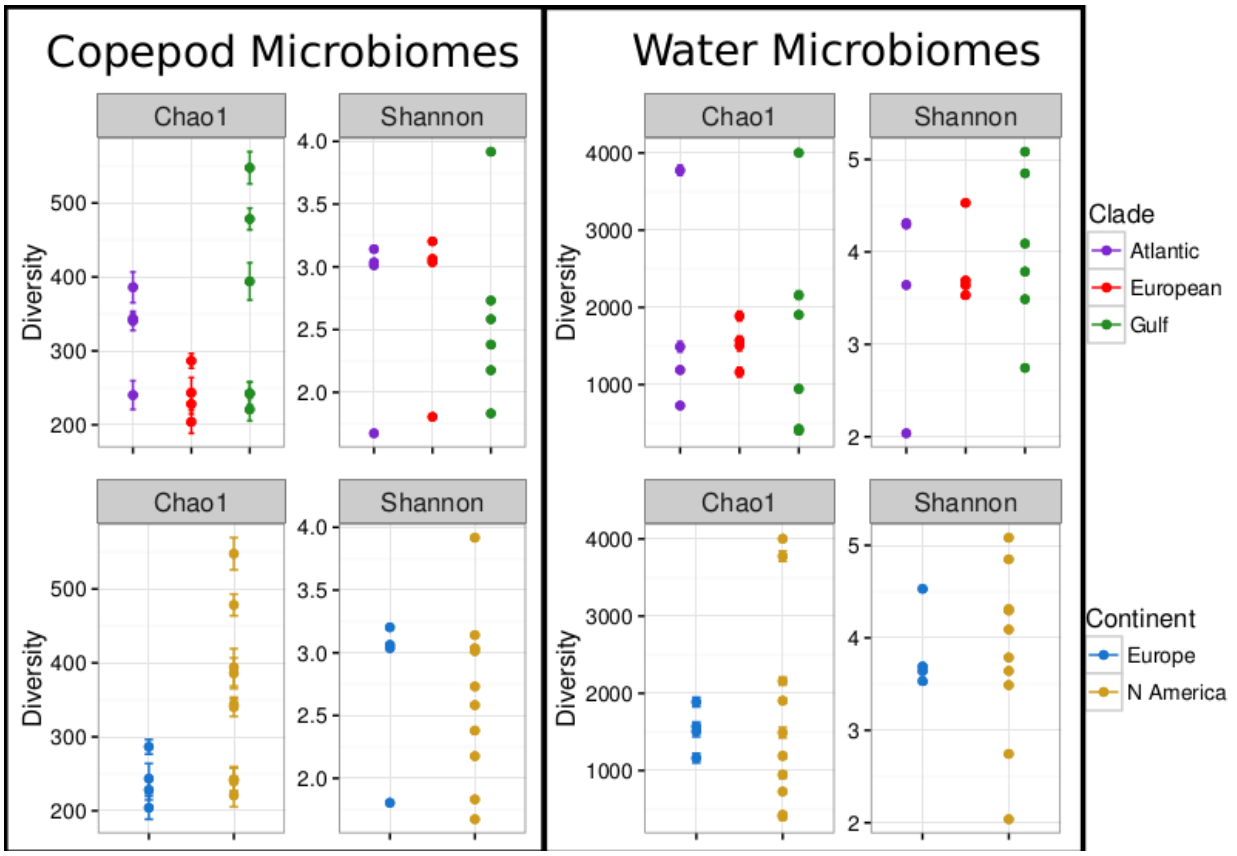

Figure 11 – Microbiome diversity is equivalent between clades and continents – Copepod microbiome diversity (left panel) and water microbiome diversity (right panel) among clades (top) and between continents (bottom). There is no significant difference in either Chao or Shannon diversity between clades or continents.

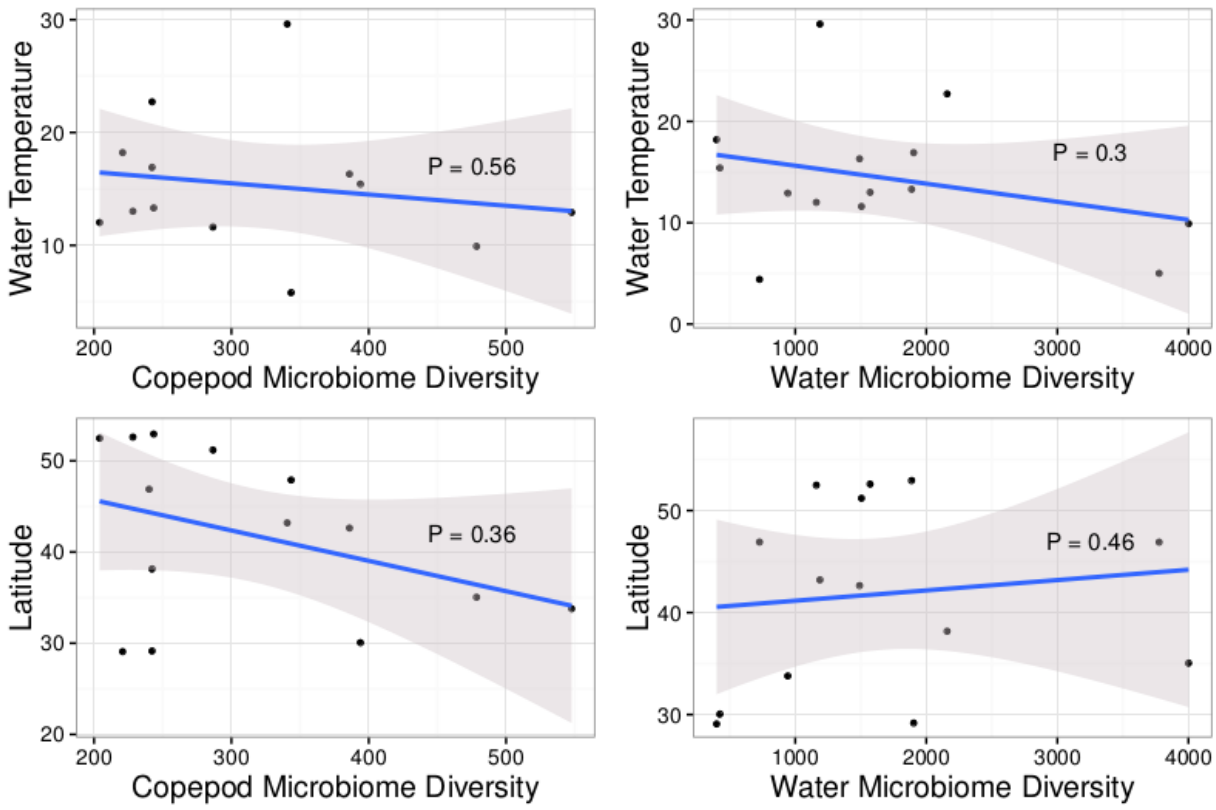

Figure 12 – Microbiome diversity is uncorrelated with latitude or water temperature – Chao diversity in copepod (left) and water (right) microbiomes versus water temperature (top) and latitude (bottom). There is no correlation between either water temperature or latitude and microbiome diversity.

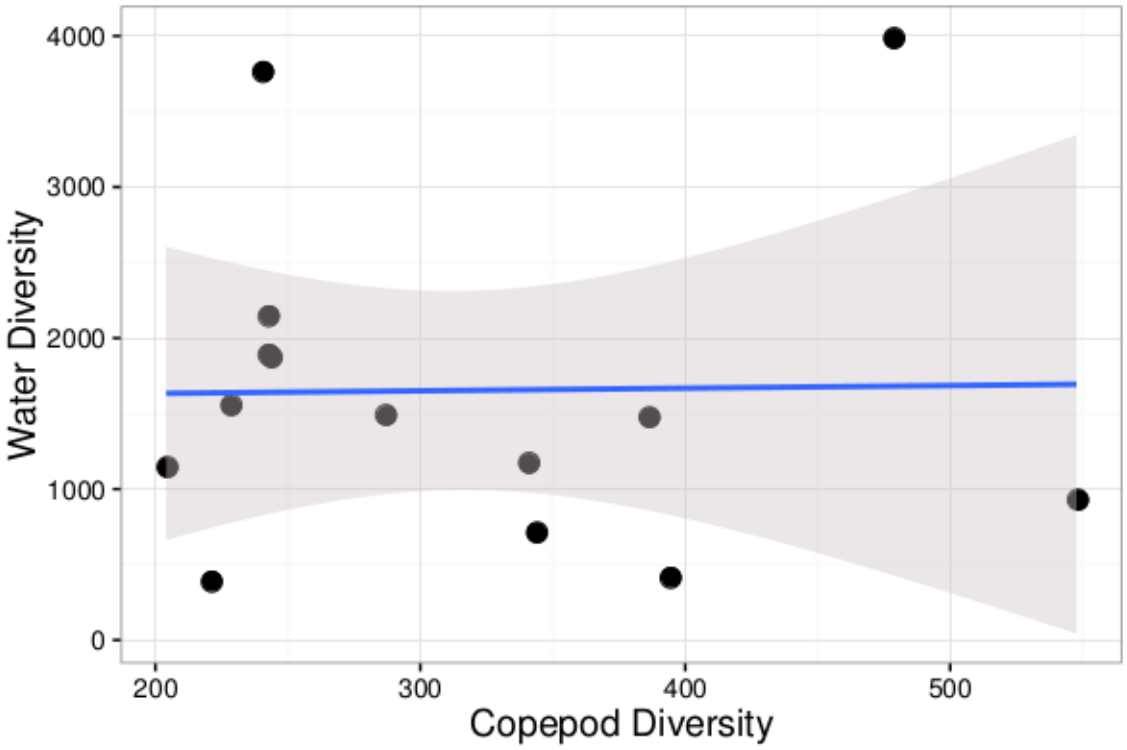

Figure 13 – Copepod microbiome diversity and water diversity from the same location are uncorrelated – Alpha diversity (as measured by Chao diversity) in paired water and copepod samples taken from the same location are uncorrelated. The gray shaded region is the 95% confidence interval for the regression line.

Table 1 – Sample Metadata

Sample	Read Count	Environment	Water Type	Salinity	Temp	Location	Clade	Lat	Long	Continent
BRE	41721	Copepod	Fresh	0.0	29.7	Braddock Bay	Northeast	43.307	-77.706	N America
BRW	88311	Water	Fresh	0.0	29.7	Braddock Bay	Northeast	43.307	-77.706	N America
CBE	4369	Copepod	Salt	4.9	17.0	Cocodrie Bayou	South	29.254	-90.664	N America
CBW	135913	Water	Salt	4.9	17.0	Cocodrie Bayou	South	29.254	-90.664	N America
EUE	124069	Copepod	Fresh	0.0	10.0	Lake Eufaula	South	35.146	-95.627	N America
EUW	107594	Water	Fresh	0.0	10.0	Lake Eufaula	South	35.146	-95.627	N America
FRE	14567	Copepod	Fresh	0.0	13.4	Frisian Lake	Europe	53.031	5.729	Europe
FRW	166748	Water	Fresh	0.0	13.4	Frisian Lake	Europe	53.031	5.729	Europe
IJE	22427	Copepod	Fresh	0.0	13.1	Ijsselmeer	Europe	52.699	5.290	Europe
IJW	116445	Water	Fresh	0.0	13.1	Ijsselmeer	Europe	52.699	5.290	Europe
LOE	27736	Copepod	Fresh	0.0	22.8	Louisville	South	38.260	-85.747	N America
LOW	153142	Water	Fresh	0.0	22.8	Louisville	South	38.260	-85.747	N America
MAE	6805	Copepod	Salt	5.0	NA	Montmagnay	Northeast	46.990	-70.550	N America
MIE	7586	Copepod	Fresh	0.1	15.7	Milwaukee	Northeast	43.051	-87.882	N America
MIW	39624	Water	Fresh	0.0	15.7	Milwaukee	Northeast	43.051	-87.882	N America
MME	23016	Copepod	Fresh	0.0	12.1	Markemeer	Europe	52.574	5.033	Europe
MMW	39880	Water	Fresh	0.0	12.1	Markemeer	Europe	52.574	5.033	Europe
SCE	55403	Copepod	Salt	7.0	11.7	Western Scheldt	Europe	51.302	4.286	Europe
SCW	39856	Water	Salt	7.0	11.7	Western Scheldt	Europe	51.302	4.286	Europe
SJE	145107	Copepod	Salt	3.1	NA	St Jean	Northeast	47.219	-70.281	N America
TBE	14029	Copepod	Salt	4.0	16.0	Taylor Bayou	South	29.883	-94.051	N America
TBW	121406	Water	Salt	4.0	16.0	Taylor Bayou	South	29.883	-94.051	N America
TXE	194123	Copepod	Fresh	0.9	13.0	Lake Texoma	South	33.882	-96.797	N America
TXW	40812	Water	Fresh	0.9	13.0	Lake Texoma	South	33.882	-96.797	N America
V1E	108551	Copepod	Salt	13.7	5.9	Lisle Verte	Northeast	48.002	-69.423	N America

Table 2 – Bacterial OTUs enriched in copepod microbiomes *versus* water microbiomes – Greengenes taxonomic classification, log fold change, and Wald test multiple-adjusted *P*-value associated with bacterial OTUs which are enriched in the copepod microbiome across samples.

Phylum	Class	Order	Family	Genus	Species	padj
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Leadbetterella	unclassified	0.00
Bacteroidetes	Saprospirae	Saprospirales	Saprospiraceae	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Rubrivivax	gelatinosus	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	Methylobacterium	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	unclassified	0.00
Bacteroidetes	Saprospirae	Saprospirales	Saprospiraceae	unclassified	unclassified	0.00
GN02	BD1-5	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Weeksellaceae	Chryseobacterium	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Propionibacteriaceae	Propionibacterium	acnes	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Renibacterium	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	unclassified	0.01
Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	unclassified	0.01
Bacteroidetes	Flavobacteriia	Flavobacteriales	Weeksellaceae	Cloacibacterium	unclassified	0.01

Table 3 – Bacterial OTUs enriched in water microbiomes *versus* copepod microbiomes – Greengenes taxonomic classification, log fold change, and Wald test multiple-adjusted *P*-value associated with bacterial OTUs which are enriched in water microbiomes across samples.

Phylum	Class	Order	Family	Genus	Species	padj
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-Bd	acI-B1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A6	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A4	0.00
Proteobacteria	Alphaproteobacteria	Rickettsiales	alfV	alfV-A	LD12	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacI	bacI-A	bacI-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acTH1	acTH1-A	acTH1-A1	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacIII	bacIII-B	Algor	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betII	Pnec	PnecB	0.00
Actinobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	MWH-UniP1	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-A	Iluma-A1	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betII	Pnec	PnecD	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacIII	bacIII-A	unclassified	0.00
Chloroflexi	SL56	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acSTL	acSTL-A	acSTL-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Candidatus_Rhodoluna	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A5	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Luna1	Luna1-A	Luna1-A2	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-C	acI-C2	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Proteobacteria	Alphaproteobacteria	Sphingomonadales	alfIV	alfIV-B	M-L-85	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betIII	betIII-A	betIII-A1	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B2	0.00
Proteobacteria	Alphaproteobacteria	Sphingomonadales	alfIV	alfIV-B	Pyxis	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Luna1	Luna1-A	Luna1-A2	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	Phila	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00

Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	unclassified	unclassified	0.00
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	unclassified	unclassified	0.00
Chloroflexi	SL56	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Thermoleophilia	Gaiellales	Gaiellaceae	unclassified	unclassified	0.00
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	211ds20	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	HTCC2188	HTCC	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-D	Iamia	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Luna1	Luna1-A	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacVI	bacVI-B	Pedo	0.00
Proteobacteria	Betaproteobacteria	Methylophilales	betIV	betIV-A	LD28	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	HTCC2188	HTCC	unclassified	0.00
Verrucomicrobia	Pedosphaerae	Pedosphaerales	R4-41B	unclassified	unclassified	0.00
Spirochaetes	Leptospirae	Leptospirales	Leptospiraceae	Leptospira	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacVI	bacVI-B	Pedo	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Mycobacteriaceae	Mycobacterium	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A4	0.00
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae	Rickettsiella	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B1	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A4	0.00
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	MWH-UniP1	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betIII	betIII-A	betIII-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	Phila	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Luna3	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A4	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betVII	betVII-B	betVII-B1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-B	acI-B4	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Luna1	Luna1-A	Luna1-A4	0.00
Proteobacteria	Betaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	Sediminibacterium	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A3	0.00

Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acSTL	acSTL-A	acSTL-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-C	acI-C1	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A7	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-C	Iluma-C1	0.00
Proteobacteria	Alphaproteobacteria	Rhodospirillales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-C	acI-C1	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter	starrii	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Planctomycetes	Planctomycetia	Gemmatales	Gemmataceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	unclassified	unclassified	0.00
Actinobacteria	Thermoleophilia	Solirubrobacterales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acIII	acIII-A	Luna2	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylocystaceae	Methylosinus	unclassified	0.00
Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Sulfuricurvum	kujiense	0.00
Proteobacteria	Gammaproteobacteria	LiUU-3-334	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A3	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVII	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	MWH-UniP1	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A4	0.00
Chlorobi	OPB56	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acI	acI-A	acI-A4	0.00

Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Chloroflexi	SL56	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betII	Pnec	PnecC	0.00
Planctomycetes	Phycisphaerae	Phycisphaerales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Ellin329	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Gemmatimonadaceae	Gemmatimonas	unclassified	0.00
Verrucomicrobia	Opitutae	Cerasicoccales	Cerasicoccaceae	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	HTCC2188	HTCC	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Cyanobacteria	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	unclassified	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
Cyanobacteria	Chloroplast	Stramenopiles	unclassified	unclassified	unclassified	0.00
Firmicutes	Erysipelotrichi	Erysipelotrichales	Erysipelotrichaceae	unclassified	unclassified	0.00
Cyanobacteria	Chloroplast	Cryptophyta	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	211ds20	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylotenera	mobilis	0.00
Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae	Rickettsiella	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acSTL	acSTL-A	acSTL-A2	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Actinobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acSTL	acSTL-A	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	Methylocella	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A2	0.00
Proteobacteria	Betaproteobacteria	MWH-UniP1	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Spartobacteria	Chthoniobacterales	Chthoniobacteraceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Ellin6067	unclassified	unclassified	unclassified	0.00

Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter	starrii	0.00
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	Reyranella	massiliensis	0.00
Chloroflexi	Chloroflexi	Roseiflexales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	ACK-M1	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Thermoleophilia	Solirubrobacterales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Phenylobacterium	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Legionellales	Coxiellaceae	Rickettsiella	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Spartobacteria	Chthoniobacterales	Chthoniobacteraceae	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A2	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Mycoplana	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	unclassified	unclassified	0.00
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B2	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	acTH2	Myco	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVII	unclassified	unclassified	unclassified	0.00
TM7	TM7-1	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Arcobacter	cryaerophilus	0.00
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	unclassified	unclassified	0.00
Actinobacteria	Thermoleophilia	Solirubrobacterales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betII	Pnec	PnecC	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Methylacidiphilae	Methylacidiphilales	LD19	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Proteobacteria	Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	Sulfuricurvum	kujiense	0.00
Cyanobacteria	Chloroplast	Cryptophyta	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae	CandidatusXiphinematobacter	verI-B	0.00
unclassified	unclassified	unclassified	unclassified	unclassified	unclassified	0.00

Fusobacteria	Fusobacteriia	Fusobacteriales	unclassified	unclassified	unclassified	0.00
TM7	TM7-3	EW055	unclassified	unclassified	unclassified	0.00
Acidobacteria	Acidobacteria-6	iii1-15	RB40	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Sphingomonadales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	alfI	alfI-A	alfI-A1	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Hydrogenophaga	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacII	bacII-A	unclassified	0.00
Gemmatimonadetes	Gemmatimonadetes	KD8-87	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacI	bacI-B	bacI-B1	0.00
OD1	SM2F11	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	Methylotenera	mobilis	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betII	Pnec	PnecC	0.00
OD1	ZB2	unclassified	unclassified	unclassified	unclassified	0.00
unclassified	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Gemmatimonadetes	Gemmatimonadetes	KD8-87	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
TM7	SC3	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Vibrionales	unclassified	unclassified	unclassified	0.00
Cyanobacteria	Chloroplast	Chlorophyta	Trebouxiophyceae	unclassified	unclassified	0.00
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter	starrii	0.00
Proteobacteria	Deltaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Deltaproteobacteria	Myxococcales	0319-6G20	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Cyanobacteria	Chloroplast	Cryptophyta	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	RS62	unclassified	0.00
Actinobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Sulfuritalea	unclassified	0.00
Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	unclassified	0.00
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A2	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00

Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhodospirillales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	Acinetobacter	unclassified	0.00
Chlorobi	OPB56	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteria	Sphingobacteriales	bacVI	unclassified	unclassified	0.00
WS6	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
OD1	Mb-NB09	unclassified	unclassified	unclassified	unclassified	0.00
Cyanobacteria	Chloroplast	Stramenopiles	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	Caulobacter	vibrioides	0.00
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	C39	unclassified	0.00
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingobium	xenophagum	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Candidatus_Aquiluna	rubra	0.00
Cyanobacteria	Chloroplast	Chlorophyta	Trebouxiophyceae	unclassified	unclassified	0.00
Actinobacteria	Thermoleophilia	Solirubrobacterales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	unclassified	unclassified	0.00
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Leadbetterella	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
Chlamydiae	Chlamydiia	Chlamydiales	Rhabdochlamydiaceae	Rhabdochlamydia	unclassified	0.00
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	C39	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	Agrococcus	jenensis	0.00
Proteobacteria	Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	Sulfurospirillum	unclassified	0.00
Bacteroidetes	Bacteroidia	Bacteroidales	Prevotellaceae	Prevotella	unclassified	0.00
Gemmatimonadetes	Gemmatimonadetes	KD8-87	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
WS5	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B2	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	unclassified	0.00

Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Rhodocyclales	Rhodocyclaceae	Sulfuritalea	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.00
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.00
unclassified	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVII	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Cyanobacteria	Chloroplast	Stramenopiles	unclassified	unclassified	unclassified	0.00
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bdellovibrionaceae	Bdellovibrio	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola	unclassified	0.00
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Proteobacteria	Gammaproteobacteria	Alteromonadales	211ds20	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Rhizobiales	Beijerinckiaceae	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	alfVI	unclassified	unclassified	unclassified	0.00
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	Sphingomonas	wittichii	0.00
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.00
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B2	0.00
Bacteroidetes	Flavobacteria	Flavobacteriales	bacV	unclassified	unclassified	0.00
OD1	ZB2	unclassified	unclassified	unclassified	unclassified	0.00
TM7	SC3	unclassified	unclassified	unclassified	unclassified	0.00
Cyanobacteria	Synechococcophycideae	Synechococcales	Synechococcaceae	Synechococcus	unclassified	0.00
Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	unclassified	unclassified	0.01
Bacteroidetes	Flavobacteriia	Flavobacteriales	Cryomorphaceae	Fluviicola	unclassified	0.01
Actinobacteria	Actinobacteria	Actinomycetales	Micromonosporaceae	Micromonospora	chokoriensis	0.01
TM7	SC3	unclassified	unclassified	unclassified	unclassified	0.01
Acidobacteria	Acidobacteria-6	iii1-15	mb2424	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.01
Proteobacteria	Gammaproteobacteria	Alteromonadales	211ds20	unclassified	unclassified	0.01
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.01
Verrucomicrobia	Opitutae	Opitutales	Opitutaceae	Opitutus	unclassified	0.01
Verrucomicrobia	Opitutae	Cerasicoccales	Cerasicoccaceae	unclassified	unclassified	0.01
Bacteroidetes	Saprospirae	Saprospirales	Chitinophagaceae	Sediminibacterium	unclassified	0.01
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.01
OD1	ZB2	unclassified	unclassified	unclassified	unclassified	0.01
Chloroflexi	SL56	unclassified	unclassified	unclassified	unclassified	0.01

Proteobacteria	Deltaproteobacteria	Spirobacillales	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Methylophilales	Methylophilaceae	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	betV	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	unclassified	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.01
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	Peredibacter	starrii	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A1	0.01
Firmicutes	Bacilli	Bacillales	Bacillaceae	Bacillus	horikoshii	0.01
Proteobacteria	Alphaproteobacteria	Rhizobiales	Hyphomicrobiaceae	Hyphomicrobium	sulfonivorans	0.01
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Actinomycetales	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	betI	betI-A	Lhab-A4	0.01
Proteobacteria	Betaproteobacteria	MND1	unclassified	unclassified	unclassified	0.01
Proteobacteria	unclassified	unclassified	unclassified	unclassified	unclassified	0.01
Bacteroidetes	unclassified	unclassified	unclassified	unclassified	unclassified	0.01
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Rhodospirillaceae	unclassified	unclassified	0.01
Chloroflexi	Chloroflexi	Roseiflexales	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Methylibium	unclassified	0.01
Proteobacteria	Gammaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.01
Verrucomicrobia	unclassified	unclassified	unclassified	unclassified	unclassified	0.01
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Erythrobacteraceae	Porphyrobacter	dokdonensis	0.01
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.01
Verrucomicrobia	Spartobacteria	Spartobacteriales	Spartobacteriaceae	CandidatusXiphinematobacter	verI-B	0.01
TM7	TM7-3	EW055	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	unclassified	unclassified	0.01
Actinobacteria	Acidimicrobiia	Acidimicrobiales	C111	unclassified	unclassified	0.01
OD1	ZB2	unclassified	unclassified	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Actinomycetales	Streptomycetaceae	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-B	Iluma-B2	0.01
Cyanobacteria	Chloroplast	Stramenopiles	unclassified	unclassified	unclassified	0.01
Proteobacteria	Betaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.01
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Actinomycetales	acTH2	Myco	unclassified	0.01
Proteobacteria	Alphaproteobacteria	Rhodospirillales	unclassified	unclassified	unclassified	0.01
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae	unclassified	unclassified	0.01
Proteobacteria	Alphaproteobacteria	unclassified	unclassified	unclassified	unclassified	0.01
Actinobacteria	Actinobacteria	Acidimicrobiales	acIV	acIV-C	Iluma-C1	0.01
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Hyphomonadaceae	Hyphomonas	unclassified	0.01

Chapter 2

The copepod microbiome as a reservoir for fish pathogens, including a novel pathogenic strain of *Flavobacterium*

Authors: Martin Bontrager[1], Joseph Connolly, James B. Munro[3], Joana C. Silva[3, 4], David Hunnicutt[5], Carol Eunmi Lee[2]

[1] **Laboratory of Genetics, University of Wisconsin-Madison,** 425 Henry Mall, Madison, WI 53706
[2] Center for Rapid Evolution (CORE), University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706
[3] Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street, Baltimore, MD, 21201
[4] Department of Microbiology and Immunology, University of Maryland School of Medicine,685 W. Baltimore Street, Baltimore, MD, 21201
[5] Department of Biology, St. Norbert College, 100 Grant Street GMS, 3094, De Pere, WI 54115

Abstract

Pathogens of wild and farm-raised fish cause millions of dollars of damage to the fisheries and aquaculture industries each year. Yet little is known about the ecology of fish pathogens, or their reservoirs outside of fish disease outbreaks. The zooplankton community represents an enormous niche within aquatic ecosystems in which microbes could potentially associate. Despite the abundance and ubiquity of zooplankton in both aquatic habitats and aquaculture facilities, no studies have explored the potential of zooplankton to harbor fish pathogens. In this study, we compared the microbial communities associated with the copepod *Eurytemora affinis* to the bacterioplankton communities in the ambient water from 14 saline and freshwater locations in North America and Europe. We found that bacterial genera known to contain fish pathogens, most notably *Flavobacterium*, *Chryseobacterium*, and *Aeromonas,* were significantly enriched in copepod microbiomes relative to the surrounding water. At one site, we found that the fish pathogen *Flavobacterium columnare* comprised approximately 5% of sequence reads from the copepod microbiome but was entirely absent from the surrounding water. To examine the potential of the zooplankton community to harbor fish pathogens, we tested the virulence of a novel strain of *Flavobacterium* isolated from the *E. affinis* microbiome from Lake Michigan. We found that this strain, Fc Col1, was pathogenic to zebrafish in infection trials. Fish exposed to Fc Col1 were significantly more likely to die within 72 hours than controls ($P = 5.4\text{x}10^{-6}$). These results are the first to demonstrate that a microbe associated with a zooplankton host is pathogenic in fish. As such, our results highlight the potential importance of zooplankton as reservoirs for fish pathogens.

Introduction

Fisheries have been a primary source of protein and vital nutrients for human populations for thousands of years. Human reliance on seafood has led to a remarkable expansion of the seafood production industry. Between the capture fishing and aquaculture sectors, the seafood industry is worth over $200 billion worldwide annually in terms of production alone (Sumaila *et al.* 2007; FAO 2016; *The Sunken Billions Revisited* 2016). These enormous economic engines bring jobs, stability, and food to hundreds of millions of people. In addition to their economic significance, commercial seafood species in both freshwater and saline environments are vital components of aquatic ecosystems and are integral to aquatic food webs. Unfortunately these food webs are becoming increasingly fragile, as aquatic and marine species struggle to survive the challenges of overfishing, climate change, pollution, and the rise of emerging infectious diseases (Jackson *et al.* 2001; Mullon *et al.* 2005; Murray & Peeler 2005; Allison *et al.* 2009; Cheung *et al.* 2013).

Fish diseases are an enormous problem in the aquaculture industry (Pillay *et al.* 2005). Pathogens spread rapidly in aquatic environments (McCallum *et al.* 2003), and this is especially true in close-quarters aquaculture facilities. The worldwide economic loss due to diseases in farmed seafood alone reaches billions of dollars each year and limits the capacity of aquaculture operations (Lafferty *et al.* 2015). Much of this cost is borne by the farmed-fish industry, especially Atlantic salmon and channel catfish, among others (Wagner *et al.* 2002; Costello 2009). One of the most dominant classes of fish pathogens, especially in the Great Lakes and in aquaculture facilities in the U.S., are members of the bacterial family Flavobacteraceae, particularly the genera *Flavobacterium* and *Chryseobacterium* (Bernardet *et al.* 2006; Bernardet & Bowman 2006; Austin & Austin 2007). Other fish pathogens include members of the genus

Aeromonas, which can infect humans from diseased fish (Castro-Escarpulli *et al.* 2003; Novotny *et al.* 2004). Analyzing the transmission and reservoirs of these fish diseases is of fundamental importance for understanding patterns disease outbreaks in both fish and humans.

Despite the enormous economic costs, ecological disruption, and human health impacts associated with fish diseases, we know little regarding the reservoirs, vectors, or transmission pathways of fish pathogens outside of disease outbreak (Olivares-Fuster *et al.* 2007). Once disease outbreaks become established it is clear that pathogens are passed from one infected animal to another (Austin & Austin 2007), but fish diseases tend to outbreak in unpredictable ways in both natural and aquaculture environments. Sources of infection are often difficult to identify and future outbreaks are difficult to predict. (Leung & Bates 2013). The zooplankton community in general, and copepods in particular, have been hypothesized to play a critical role as reservoirs and vectors of waterborne disease (Colwell 1996; Vezzulli *et al.* 2010). However, this hypothesis has not yet been explicitly tested. Constituents of copepod microbiomes have not been tested for their capacity to cause disease in fish or humans.

Copepods form the largest metazoan biomass in the world's oceans (Hardy 1970; Verity & Smetacek 1996) and often dominate zooplankton communities. They are primary consumers of algae and a critical link in aquatic food chains. Copepods are the primary source of food for many fish species, especially for larval and juvenile fish. The copepod bacterial community (its "microbiome") is highly distinct from free-living bacterial community in the surrounding water environment and contains many bacterial members which are also found in fish-associated microbial communities (see Chapter 1, Kim *et al.* 2007; Roeselers *et al.* 2011; Loch *et al.* 2013; Shoemaker & Moisander 2015). Some evidence implicates copepods as disease vectors of human pathogens, particularly for members of the genus *Vibrio,* including *Vibrio cholerae* and to

a lesser extent *V. parahaemolyticus*. But despite such examples of zooplankton communities as potential reservoirs for human pathogens, there are almost no studies of zooplankton communities as reservoirs for diseases which affect fish or other metazoans.

The copepod *Eurytemora affinis* is widespread throughout the Northern Hemisphere (Lee 1999, 2000), has independently invaded freshwater habitats from saline environments multiple times (Lee 1999), and is the primary source of food for many economically important fisheries (Viitasalo *et al.* 2001; Shaheen *et al.* 2001; Kimmel *et al.* 2006). The ubiquity and huge biomass of *E. affinis* makes it an ideal model with which to identify prospective fish pathogens in zooplankton communities.

Therefore, the goal of study, was to examine the microbiome of *Eurytemora affinis* across salinity environments, latitude, and continents to assess the copepod microbiome as a potential reservoir for fish pathogens. Along with copepod host-associated microbiomes, we characterized the microbiome of surrounding water environments to examine the capacity of copepods to maintain unique and specialized associations with fish pathogens, which might not readily persist free-living in water and assess the degree to which such fish pathogens dominate host-associated *versus* free-living bacterial communities. We further isolated a novel strain of bacteria of the genus *Flavobacterium* from the microbiome of a population of *E. affinis* in Lake Michigan, and examined the pathogenicity of this strain in zebrafish.

Materials and Methods

Sampling Strategy and Detection of Fish Pathogens in the Copepod Microbiome

For complete details on sample collection see Chapter 1 – "Materials and Methods". Briefly, we obtained microbiome samples from simultaneously collected paired copepod

Eurytemora affinis and surrounding water samples from 12 locations, and temporally unpaired copepod and water samples from 2 locations (Fig. 1). We sampled copepod populations from three clades in both saline and fresh water. At each location, we performed plankton tows, to collect members of the zooplankton, and collected water in 25 L sterilized carboys.

We extracted DNA was extracted from copepod microbiome samples using the UltraClean Microbial DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) and from filtered water samples with the PowerWater DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) according the manufacturer's recommendations with the additional step of a 10-minute incubation at 65°C. We amplified the V3/V4 region of the bacterial 16S rRNA gene using 319F/806R PCR primers and created amplicon libraries for paired-end sequencing. Amplicon libraries were sequenced on the Illumina MiSeq sequencer using 300bp sequencing.

We overlapped paired-end reads, filtered for quality, and removed chimeric reads. We then binned sequences together into observational taxonomic units (OTUs) in groups of 97% sequence identity at the V3/V4 region. We then assigned taxonomy to clustered OTUs with MOTHUR (Schloss *et al.* 2009) using the greengenes database (version 05-2013), modified to incorporate additional freshwater bacterial 16S sequences (DeSantis *et al.* 2006; Newton *et al.* 2011). Read counts were normalized and graphical visualizations were constructed using the DESeq2 and Phyloseq packages (McMurdie & Holmes 2013, 2014; Love *et al.* 2014) in the statistical package R (R Core Team 2014). We performed the Wald test on normalized OTU counts to identify bacterial genera that differed significantly between copepod-associated and water bacterial communities. The taxonomy of sequences identified as the genera *Flavobacterium*, *Chryseobacterium*, and *Aeromonas* using this approach were verified to the best of our ability using NCBI BLAST and *via* classification with the RDP classifier (Cole *et al.*

2014). Scripts used for 16S sequence processing and analysis are available at https://github.com/mbontrager/flavobacterium.

Culturing and Isolation of *Flavobacterium* from wild *Eurytemora affinis*

We sampled the *E. affinis* population from which the bacterial strains were isolated in August 2015 from Lake Michigan in Racine, WI, USA (42°43 46N, 87°46 44W, Fig. 1). Several individual *E. affinis* were isolated from the rest of the zooplankton community and rinsed three times with approximately 50 mL of sterile Millipore water to remove any unattached microbes from the copepod. We then homogenized the animals with a sterile 1.5ml Pestle (USA Scientific) in a 1.5 mL Eppendorf tube. The homogenate was diluted 1:10, 1:100 and 1:1000 in 0.2 micron filter sterilized water and the dilutions were plated on casein peptone, yeast extract, beef extract, sodium acetate (cytophaga) agar (Hardy Diagnostics, Santa Maria, CA). Plates were stored both at room temperature and at 17°C. Among strains of bacteria that grew up on these plates, we isolated four colonies showing the characteristic yellow color of *Flavobacterium*, and individual colonies were picked and grown up for DNA extraction.

DNA was extracted from bacterial cultures using the QIAGEN DNeasy Blood & Tissue Kit (Qiagen, Valencia, CA). We pretreated samples with an enzymatic lysis buffer (20 mM Tris*Cl, 2 mM sodium EDTA, 1.2% Triton X-100, Lysozyme 20 mg/ml) for 30 min at 37°C to ensure complete cell lysis. Extracted DNA was then used for PCR reactions using the Roche FastStart™ PCR kit (Roche Life Sciences, Indianapolis, IN) with 16S rRNA 27F forward (5` - AGAGTTTGATCCTGGCTCAG – 3`) and 1492R reverse (5` - ACGGCTACCTTGTTACGACTT – 3`) primers. The presence of PCR products was verified with gel electrophoresis and PCR products were sequenced *via* Sanger sequencing at the University of Wisconsin Biotechnology Center, Madison, WI, USA. Chromatograms were

viewed in UGENE (Okonechnikov *et al.* 2012) and the taxonomy of sequences was analyzed using the Seqmatch tool on the Ribosome Database Project (Cole *et al.* 2014, Center for Microbial Ecology-Michigan State University). One strain that we identified as a member of the genus *Flavobacterium*, which we have identified as Fc Col1, was used for subsequent pathogenicity trials.

Pathogenicity trials of *Flavobacterium* Isolate Fc Col1

We obtained zebrafish (*Brachydanio rerio*) from Ekk-Will Waterlife Resources (Ruskin, FL) and housed under standard conditions. *Flavobacterium* strain Fc Col1 that we isolated from cultures and then froze at -80^{o}C was grown in Modified Shieh media overnight at 28^{o}C shaken at 150 rpm. The intraperitoneal injections of strain Fc Col1 into zebrafish were carried out essentially as described by Neely et al. (2002). Zebra fish were anesthetized by immersion in MS-222 (3.84 μg/ml; Sigma Chemical) and placed supine on moistened cheesecloth. A 29-gauge needle was positioned parallel to the fish's spine and inserted cephalad into the abdomen just posterior to the pectoral fins. The needle was then inserted to the end of the bevel, and 50 μl of overnight cultures of the Fc Col1 was injected into the fish. Control fish (N = 6 per trial for three trials) were injected with 50 μl of sterile Modified Shieh media. Injected fish (N = 6 per trial for three trials) were transferred to 1 liter beakers with tank water for observation, with mortalities being recorded every 12 hours. We analyzed survival data and performed a log-rank (Mantel-Cox) test using the Survival package in R (Therneau 2015).

Results

Potential fish pathogens in the copepod microbiome

We found that, relative to the surrounding water, the copepod microbiome was highly enriched with bacterial genera containing many prospective fish pathogens (Table 1). The genus *Flavobacterium,* for example, which contains several major fish pathogens, was highly enriched in the copepod microbiome relative to the surrounding water (Wald test, adjusted $P < 0.001$). In the sample taken from copepods in Lake Ontario from Braddock Bay, NY, we saw a significant enrichment of a bacterial OTU, which was identified as *Flavobacterium columnare* by greengenes taxonomic classification. In this sample, *F. columnare* made up just under 5% of the total bacterial 16S reads from the copepod population. We also found *F. columnare* in three other copepod samples, albeit at significantly lower abundance, in the Ohio river in Louisville, KY, the St. Lawrence river in Montmagny, QC, Canada, and in Lake Michigan from Racine harbor in Racine, WI. Several other strains of *Flavobacterium* were present in other samples, including one copepod sample from Lake Texoma, TX in which members of the *Flavobacterium* composed over 25% of the total bacterial community by 16S read count. Most of these strains were unclassifiable to the strain or species levels. In almost all cases, *Flavobacterium* was more common in the copepod microbiome than in the surrounding water, often by orders of magnitude.

Members of the genera *Chryseobacterium* and *Aeromonas* were not present in every copepod microbiome, as in the case of *Flavobacterium*, yet they were still significantly enriched in the copepod microbiome overall relative to the surrounding water (Wald test, adjusted $P < 0.001$ in both cases). Abundance of members of these two genera reached as high as 13% of the overall bacterial community for the *Chryseobacterium*, and 8.5% for *Aeromonas.* Overall, these three genera represented approximately 8% of the copepod microbiome averaged across all samples, compared to only 0.2% of the composition of the water bacterial community (Fig. 2).

Pathogenicity of *Flavobacterium* Fc Col1 from Wild Copepods

All four colonies of *Flavobacterium* that we recovered from copepods in Lake Michigan were identical in sequence at the 16S rRNA gene. This strain, which we named Fc Col1, was used for subsequent analyses. The 16S gene of Fc Col1 sequence was classified to 100% confidence within the genus *Flavobacterium* using the Ribosomal Database Project classifier (Cole *et al.* 2014). Fc Col1 could not, however, be assigned to any well-characterized *Flavobacterium* species or strain with confidence.

Infection and pathogenicity trials showed that Fc Col1 is pathogenic to zebrafish (Fig. 3, Tables S1-S3). Fish exposed to the *Flavobacterium* strain were significantly more likely to die than control fish injected with sterile media (Mantel-Cox log-rank test, χ^2 t.s. = 20.7, d.f. = 1, P = 5.4x10-6, Fig. 3). Mortality was greatest in the first 24-hour period post infection, although fish continued to die up to 72 hours after infection. There was no mortality in control fish injected with sterile Modified Shieh broth.

Discussion

Our analysis of the microbiome of *E. affinis* from multiple locations across environmental conditions showed that fish pathogens were widespread in the copepod microbiome. Furthermore, we found that potential fish pathogens were orders of magnitude more abundant in copepod microbiomes than in the surrounding water and represented approximately 8% of the copepod microbiome versus only 0.7% of the water microbiome on average. Members of the prospective fish pathogenic genera *Flavobacterium*, *Chryseobacterium*, and *Aeromonas* were all highly enriched in copepod microbiomes *versus* the surrounding water (Fig. 2). In some

copepod populations, bacteria in these three genera were represented by over 25% of 16S gene sequence reads from the copepod microbiome. We found that the fish pathogen *Flavobacterium columnare* comprised almost 5% of the total copepod bacterial community in one population from the Great Lakes, representing a striking example of the capacity of the copepod microbiome to serve as a reservoir for fish pathogens.

As further evidence for the potential of the copepod microbiome to harbor fish pathogens, we isolated a strain of *Flavobacterium* from the bacterial community of a population of *E. affinis* from Lake Michigan. We found this strain to be pathogenic to zebrafish in infection trials. To our knowledge, this is the first time that a bacterial strain isolated from zooplankton-associated microbiota has been proven to be pathogenic in fish. Survival probability in fish exposed to Fc Col1 was significantly lower than in control fish (Mantel-Cox log-rank test, $P = 5.4x10\text{-}6$, Fig. 3). This is important evidence for the role that zooplankton may play in the pathways and transmission of waterborne disease, as has been previously hypothesized (Colwell 1996; Vezzulli *et al.* 2010).

Most putative fish pathogens in our survey of the copepod microbiome belong to the family Flavobacteraceae, and more specifically from the genus *Flavobacterium*. Members of this genus are common environmental isolates, especially in aquatic environments. Many are psychrotolerant, and several are psychrophilic and are commonly found in polar environments (Bernardet & Bowman 2006). Several freshwater species of *Flavobacterium*, as well as some saltwater species, are the causative agents of a variety of fish diseases and have been studied extensively (Madsen *et al.* 2005; Declercq *et al.* 2013). In the Great Lakes, members of the genus are estimated to account for more fish mortality than all other pathogens combined (Faisal & Hnath 2005; Loch *et al.* 2013). The most prominent of these pathogenic strains are *F.*

psychrophilum, *F. columnare*, and *F. branchiophilum*, although numerous other strains have been implicated in fish diseases (Austin & Austin 2007). Several species of *Flavobacterium* were present in our sampling of the *E. affinis* microbiome, including *F. columnare* and *F. succicans*, both of which are known fish pathogens.

Sequence reads from *F. columnare* comprised almost 5% of the total 16S reads in the copepod microbiome from one of our samples from Lake Ontario. Across all our sampling, *F. columnare* was entirely absent from bacterioplankton communities in the surrounding water column despite its presence in some of the copepod microbiome samples. *F. columnare* is widespread in freshwater environments worldwide and can infect many freshwater fish species (Austin & Austin 2007; Olivares-Fuster *et al.* 2007). Columnaris, the disease caused by *F. columnare*, is a particularly destructive pathogen in aquaculture environments. In farmed channel catfish ponds, columnaris disease is the second-leading cause of fish mortality in the U.S. southeast, causing an estimated $30 million annual loss (Wagner *et al.* 2002). Outbreaks in aquaculture facilities also cause mortality in salmon, trout, and tilapia (Figueiredo *et al.* 2005; Austin & Austin 2007; Avendaño-Herrera *et al.* 2011; Dong *et al.* 2015). The high abundance of *F. columnare* in the copepod population in Lake Ontario, along with the presence of *F. columnare* in three other copepod populations, provides compelling evidence that copepods, and perhaps the zooplankton community in general, may serve as reservoirs for this destructive fish pathogen. Our inability to detect this species in the water column indicates that that the composition of the bacterial community in the water column would not accurately gauge the presence of fish pathogens in an aquatic environment.

Additionally, our detection of members of the *Chryseobacterium* clade in the copepod microbiome provided further evidence of the importance of the copepod microbiome as

reservoirs and vectors of putative fish pathogens. Originally classified within the genus *Flavobacterium, Chryseobacterium* spp. also inhabit diverse habitats and are associated with a host of animals (Bernardet *et al.* 2006). Members of the genus have been recovered from diseased animals, from invertebrates (Burešová *et al.* 2006) to frogs (Mauel *et al.* 2002) and humans (Green & Nolan 2001), but it is in fish that they have most often been causative agents of disease (Loch & Faisal 2015). *Chryseobacterium spp*. Are the causative agents of diseases of turbot (Mudarris & Austin 1989), salmon (Ilardi *et al.* 2010), rainbow, lake, and brown trout (Ilardi *et al.* 2009; Loch & Faisal 2014), among others. *Chryseobacterium spp.* in our sampling were 97.9% - 99.8% similar to *Chryseobacterium* strains isolated from Chinook salmon fry and eggs in hatcheries along Lake Michigan (Thomas Loch, personal communication). *Chryseobacterium spp*. are also often implicated in human infections, especially in immunocompromised individuals (Bernardet *et al.* 2006). Interestingly, a previous member of the genus *Chryseobacterium,* now classified as *Elizabethkingia meningoseptica*, made up almost 2.5% of the sequence reads from the population of *E. affinis* in the Ohio River in Louisville, KY. *E. meningoseptica* is the most well-known human pathogen in the family Flavobacteraceae, and while rare, is the cause of several cases of infant mortality and increasing reports of clinical infections (Bernardet *et al.* 2006; Jean *et al.* 2014).

Our results together provide strong support for the presence of a pathogenic strain of *Flavobacterium*. Taxonomic classification *via* 16S rRNA gene sequence reads alone does not provide conclusive proof of the presence of fish pathogens in the copepod microbiome, as a locus such as 16S rRNA provides systematic information, but does not indicate pathogenicity. Our isolation of a novel strain of *Flavobacterium* from the copepod microbiome and the results of our infection trials on zebrafish does provide strong support. We found in the copepod

microbiome a novel strain Fc Col1, which was previously uncharacterized. In pathogenicity trials, Fc Col1 was responsible for approximately 75% mortality in zebrafish exposed to the strain after 72 hours. This result suggests that Fc Col1 is less virulent than *F. columnare*, but more virulent than common *Flavobacterium* environmental isolates, such as *F. johnsoniae* (David Hunnicutt, personal obs.).

The closest 16S rRNA gene sequence matches in the NCBI database belong to a *Flavobacterium* that was found in the 16S sequencing of the skin microbiome of amphibians (Walke *et al.* 2015). It appears that a strain of *Flavobacterium* that is 99.3% identical to Fc Col1 at the 16S rRNA gene was also isolated from the waters of trout hatcheries along the Great Lakes (Thomas Loch, personal communication). Given the potential for pathogenicity in fish and the similarity of this strain to *Flavobacterium* strains detected in trout hatcheries, further characterization of this isolate is necessary. Subsequent studies should perform further isolation and pathogenicity testing of other *Flavobacterium* and putative pathogens from the copepod microbiome given its potential to harbor pathogens.

Fish disease prevalence may increase in the near future in the face of global climate change and the associated impact on pathogen-host dynamics and range-shifts in pathogens worldwide in response to increasing temperatures (Harvell *et al.* 2002; Karvonen *et al.* 2010). It is difficult to assess the economic and ecological damages caused by fish pathogens in wild environments, but the potential for fish pathogens to rapidly evolve greater virulence in aquaculture environments (Mennerat *et al.* 2010; Sundberg *et al.* 2016) and the potential for those pathogens to spread to wild populations, raises a cause for concern. Altogether, our findings implicate the zooplankton community as a reservoir for fish diseases and for the first time provide strong support for the presence of fish pathogens in zooplankton microbiomes. Our

results have significant potential to affect aquaculture practices, as zooplankton and copepods are often used as feed for fish fry in aquaculture settings. The environmental reservoirs for fish diseases are the subject of much speculation and are of immense interest to fisheries biologists and aquaculturists. Copepods have long been thought to be a major reservoir and vector of waterborne diseases, albeit for human pathogens, and our study provides explicit support for this hypothesis. The finding that copepods, and potentially other zooplankton, harbor fish pathogens in their host-associated bacterial communities has the potential to transform our understanding of the ecology of fish pathogens and may affect the ways in which fisheries professionals manage disease.

Acknowledgments

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256259. Funds for this project were provided by National Science Foundation OCE- 1046372 to Carol E. Lee and OCE-1046371 to Joana C. Silva. The authors would like to acknowledge the contribution of the Hunnicutt laboratory at St. Norbert College for the efforts in performing zebrafish injections and pathogenicity trials. Drs. Mohammed Faisal and Thomas Loch at Michigan State University provided invaluable insights into the distribution of *Flavobacterium* in the Great Lakes. Joseph Connolly contributed significantly to the culturing of *Flavobacterium* from wild copepods and to the extraction and sequencing of DNA from bacterial colonies. Jane Remfert, Kristin Lee, and Mackenzie Carlson assisted with collections in the field.

Literature Cited

Allison EH, Perry AL, Badjeck M-C *et al.* (2009) Vulnerability of national economies to the impacts of climate change on fisheries. *Fish and fisheries*, **10**, 173–196.

Austin B, Austin DA (2007) *Bacterial Fish Pathogens: Disease of Farmed and Wild Fish*. Springer Science & Business Media.

Avendaño-Herrera R, Gherardelli V, Olmos P *et al.* (2011) Flavobacterium columnare associated with mortality of salmonids farmed in Chile: a case report of two outbreaks. *Bulletin of the European Association of Fish Pathologists*, **31**, 36–44.

Bernardet J, Bowman JP (2006) The Genus Flavobacterium. In: *Prokaryotes*, pp. 481–531.

Bernardet J-F, Hugo C, Bruun B (2006) The genera Chryseobacterium and Elizabethkingia. In: *The prokaryotes*, pp. 638–676. Springer.

Burešová V, Franta Z, Kopáček P (2006) A comparison of Chryseobacterium indologenes pathogenicity to the soft tick Ornithodoros moubata and hard tick Ixodes ricinus. *Journal of Invertebrate Pathology*, **93**, 96–104.

Castro-Escarpulli G, Figueras MJ, Aguilera-Arreola G *et al.* (2003) Characterisation of Aeromonas spp. isolated from frozen fish intended for human consumption in Mexico. *International journal of food microbiology*, **84**, 41–49.

Cheung WWL, Watson R, Pauly D (2013) Signature of ocean warming in global fisheries catch. *Nature*, **497**, 365–368.

Cole JR, Wang Q, Fish JA *et al.* (2014) Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, **42**, D633–D642.

Colwell RR (1996) Global climate and infectious disease: the cholera paradigm. *Science*, **274**, 2025–2031.

Costello MJ (2009) The global economic cost of sea lice to the salmonid farming industry. *Journal of fish diseases*, **32**, 115–118.

Declercq AM, Haesebrouck F, Van Den Broeck W, Bossier P, Decostere A (2013) Columnaris disease in fish: A review with emphasis on bacterium-host interactions. *Veterinary Research*, **44**, 1–17.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* (2006) Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Applied and Environmental Microbiology*, **72**, 5069–5072.

Dong HT, Nguyen VV, Le HD *et al.* (2015) Naturally concurrent infections of bacterial and viral pathogens in disease outbreaks in cultured Nile tilapia (Oreochromis niloticus) farms. *Aquaculture*, **448**, 427–435.

Faisal M, Hnath JG (2005) Fish health and diseases issues in the Laurentian Great Lakes. In: *Health and diseases of aquatic organisms: bilateral perspectives*, pp. 331–349. Michigan State University Press, East Lansing, MI.

FAO (2016) *The State of World Fisheries and Aquaculture 2016. Contributing to food security and nutrition for all*. Rome.

Figueiredo HCP, Klesius PH, Arias CR *et al.* (2005) Isolation and characterization of strains of Flavobacterium columnare from Brazil. *Journal of Fish Diseases*, **28**, 199–204.

Green BT, Nolan PE (2001) Cellulitis and Bacteraemia due to Chryseobacterium indologenes. *Journal of Infection*, **42**, 219–220.

Hardy A (1970) The Open Sea. In: *The World of Plankton*, p. . Collins.

Harvell CD, Mitchell CE, Ward JR *et al.* (2002) Climate Warming and Disease Risks for Terrestrial and Marine Biota. *Science*, **296**, 2158–2162.

Ilardi P, Abad J, Rintamäki P, Bernardet J-F, Avendaño-Herrera R (2010) Phenotypic, serological and molecular evidence of Chryseobacterium piscicola in farmed Atlantic salmon, Salmo salar L., in Finland. *Journal of Fish Diseases*, **33**, 179–181.

Ilardi P, Fernandez J, Avendano-Herrera R (2009) Chryseobacterium piscicola sp. nov., isolated from diseased salmonid fish. *INTERNATIONAL JOURNAL OF SYSTEMATIC AND EVOLUTIONARY MICROBIOLOGY*, **59**, 3001–3005.

Jackson JBC, Kirby MX, Berger WH *et al.* (2001) Historical Overfishing and the Recent Collapse of Coastal Ecosystems. *Science*, **293**, 629–637.

Jean SS, Lee WS, Chen FL, Ou TY, Hsueh PR (2014) Elizabethkingia meningoseptica: an important emerging pathogen causing healthcare-associated infections. *Journal of Hospital Infection*, **86**, 244–249.

Karvonen A, Rintamäki P, Jokela J, Valtonen ET (2010) Increasing water temperature and disease risks in aquatic systems: Climate change increases the risk of some, but not all, diseases. *International Journal for Parasitology*, **40**, 1483–1488.

Kim D-H, Brunt J, Austin B (2007) Microbial diversity of intestinal contents and mucus in rainbow trout (Oncorhynchus mykiss). *Journal of Applied Microbiology*, **102**, 1654–1664.

Kimmel DG, Miller WD, Roman MR (2006) Regional scale climate forcing of mesozooplankton dynamics in Chesapeake Bay. *Estuaries and Coasts*, **29**, 375–387.

Lafferty KD, Harvell CD, Conrad JM *et al.* (2015) Infectious Diseases Affect Marine Fisheries and Aquaculture Economics. *Annual Review of Marine Science*, **7**, 471–496.

Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod Eurytemora affinis. *Evolution*, **53**, 1423–1434.

Lee CE (2000) Global Phylogeography of a Cryptic Copepod Species Complex and Reproductive Isolation between Genetically Proximate "Populations." *Evolution*, **54**, 2014–2027.

Leung TLF, Bates AE (2013) More rapid and severe disease outbreaks for aquaculture at the tropics: implications for food security. *Journal of Applied Ecology*, **50**, 215–222.

Loch TP, Faisal M (2014) Chryseobacteriumaahli sp. nov., isolated from lake trout (Salvelinus namaycush) and brown trout (Salmo trutta), and emended descriptions of Chryseobacterium ginsenosidimutans and Chryseobacterium gregarium. *International Journal of Systematic and Evolutionary Microbiology*, **64**, 1573–1579.

Loch TP, Faisal M (2015) Emerging flavobacterial infections in fish: A review. *Journal of Advanced Research*, **6**, 283–300.

Loch TP, Fujimoto M, Woodiga S a. *et al.* (2013) Diversity of fish-associated flavobacteria of Michigan. *Journal of Aquatic Animal Health*, **25**, 149–164.

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

Madsen L, Møller JD, Dalsgaard I (2005) Flavobacterium psychrophilum in rainbow trout, Oncorhynchus mykiss (Walbaum), hatcheries: studies on broodstock, eggs, fry and environment. *Journal of Fish Diseases*, **28**, 39–47.

Mauel MJ, Miller DL, Frazier KS, Hines ME (2002) Bacterial Pathogens Isolated from Cultured Bullfrogs (Rana Castesbeiana). *Journal of Veterinary Diagnostic Investigation*, **14**, 431–433.

McCallum H, Harvell D, Dobson A (2003) Rates of spread of marine pathogens. *Ecology Letters*, **6**, 1062–1067.

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data (M Watson, Ed,). *PLoS ONE*, **8**, e61217–e61217.

McMurdie PJ, Holmes S (2014) Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible (AC McHardy, Ed,). *PLoS Computational Biology*, **10**, e1003531–e1003531.

Mennerat A, Nilsen F, Ebert D, Skorping A (2010) Intensive Farming: Evolutionary Implications for Parasites and Pathogens. *Evolutionary Biology*, **37**, 59–67.

Mudarris M, Austin B (1989) Systemic disease in turbot Scophthalmus maximus caused by a previously unrecognised Cytophaga-like bacterium. *Diseases of Aquatic Organisms*, **6**, 161–166.

Mullon C, Fréon P, Cury P (2005) The dynamics of collapse in world fisheries. *Fish and Fisheries*, **6**, 111–120.

Murray AG, Peeler EJ (2005) A framework for understanding the potential for emerging diseases in aquaculture. *Preventive Veterinary Medicine*, **67**, 223–235.

Neely MN, Pfeifer JD, Caparon M (2002) Streptococcus-zebrafish model of bacterial pathogenesis. *Infection and immunity*, **70**, 3904–3914.

Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011) A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews*, **75**, 14–49.

Novotny L, Dvorska L, Lorencova A, Beran V, Pavlik I (2004) Fish: a potential source of bacterial pathogens for human beings. A review. *Veterinarni Medicina-UZPI (Czech Republic)*.

Okonechnikov K, Golosova O, Fursov M, others (2012) Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*, **28**, 1166–1167.

Olivares-Fuster O, Baker JL, Terhune JS *et al.* (2007) Host-specific association between Flavobacterium columnare genomovars and fish species. *Systematic and Applied Microbiology*, **30**, 624–633.

Pillay TVR, Kutty MN, others (2005) *Aquaculture: principles and practices.* Blackwell publishing.

R Core Team (2014) R: A Language and Environment for Statistical Computing.

Roeselers G, Mittge EK, Stephens WZ *et al.* (2011) Evidence for a core gut microbiota in the zebrafish. *The ISME Journal*, **5**, 1595–1608.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*, **75**, 7537–7541.

Shaheen P a., Stehlik LL, Meise CJ *et al.* (2001) Feeding behavior of newly settled winter flounder (Pseudopleuronectes americanus) on calanoid copepods. *Journal of Experimental Marine Biology and Ecology*, **257**, 37–51.

Shoemaker KM, Moisander PH (2015) Microbial diversity associated with copepods in the North Atlantic subtropical gyre. *FEMS Microbiology Ecology*, **91**, 1–11.

Sumaila UR, Marsden AD, Watson R, Pauly D (2007) A Global Ex-vessel Fish Price Database: Construction and Applications. *Journal of Bioeconomics*, **9**, 39–51.

Sundberg L-R, Ketola T, Laanto E *et al.* (2016) Intensive aquaculture selects for increased virulence and interference competition in bacteria. *Proc. R. Soc. B*, **283**, 20153069.

The Sunken Billions Revisited: Progress and Challenges in Global Marine Fisheries (2016) World Bank, Washington, DC.

Therneau T (2015) *A package for survival analysis in S.*

Verity P, Smetacek V (1996) Organism life cycles, predation, and the structure of marine pelagic ecosystems. *Marine Ecology Progress Series*, **130**, 277–293.

Vezzulli L, Pruzzo C, Huq A, Colwell RR (2010) Environmental reservoirs of Vibrio cholerae and their role in cholera. *Environmental Microbiology Reports*, **2**, 27–33.

Viitasalo M, Flinkman J, Viherluoto M (2001) Zooplanktivory in the Baltic Sea: a comparison of prey selectivity by Clupea harengus and Mysis mixta, with reference to prey escape reactions. *Marine Ecology Progress Series*, **216**, 191–200.

Wagner BA, Wise DJ, Khoo LH, Terhune JS (2002) The Epidemiology of Bacterial Diseases in Food-Size Channel Catfish. *Journal of Aquatic Animal Health*, **14**, 263–272.

Walke JB, Becker MH, Hughey MC *et al.* (2015) Most of the Dominant Members of Amphibian Skin Bacterial Communities Can Be Readily Cultured. *Applied and Environmental Microbiology*, **81**, 6589–6600.

Tables and Figures

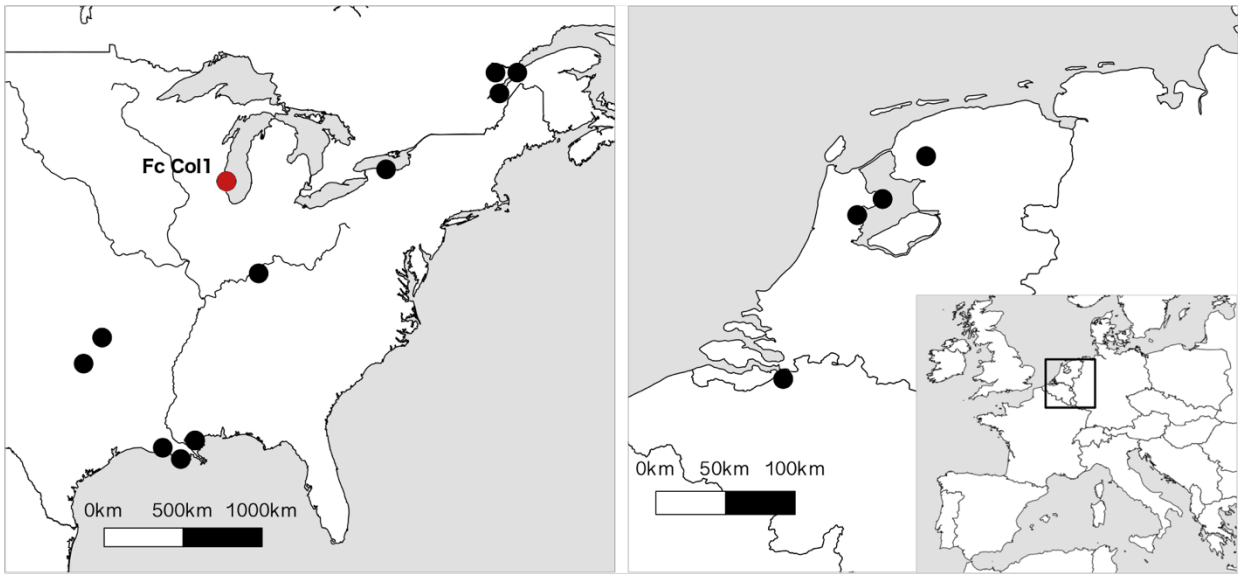

Figure 1. Population sampling of *Eurytemora affinis* for this study. Samples include populations from three independent invasions of *E. affinis*; two originating from North America from the St. Lawrence and the Gulf of Mexico, and one from the Baltic/Wadden Sea in Europe. Populations were sampled across salinity, season, and temperature. From each location, we sampled both copepod microbiomes and water bacterial communities. See Table S1 for sample metadata. Fc Col1, the *Flavobacterium* strain which was isolated from the copepod microbiome and used for pathogenicity analysis, was isolate from a population of copepods in Lake Michigan (red dot).

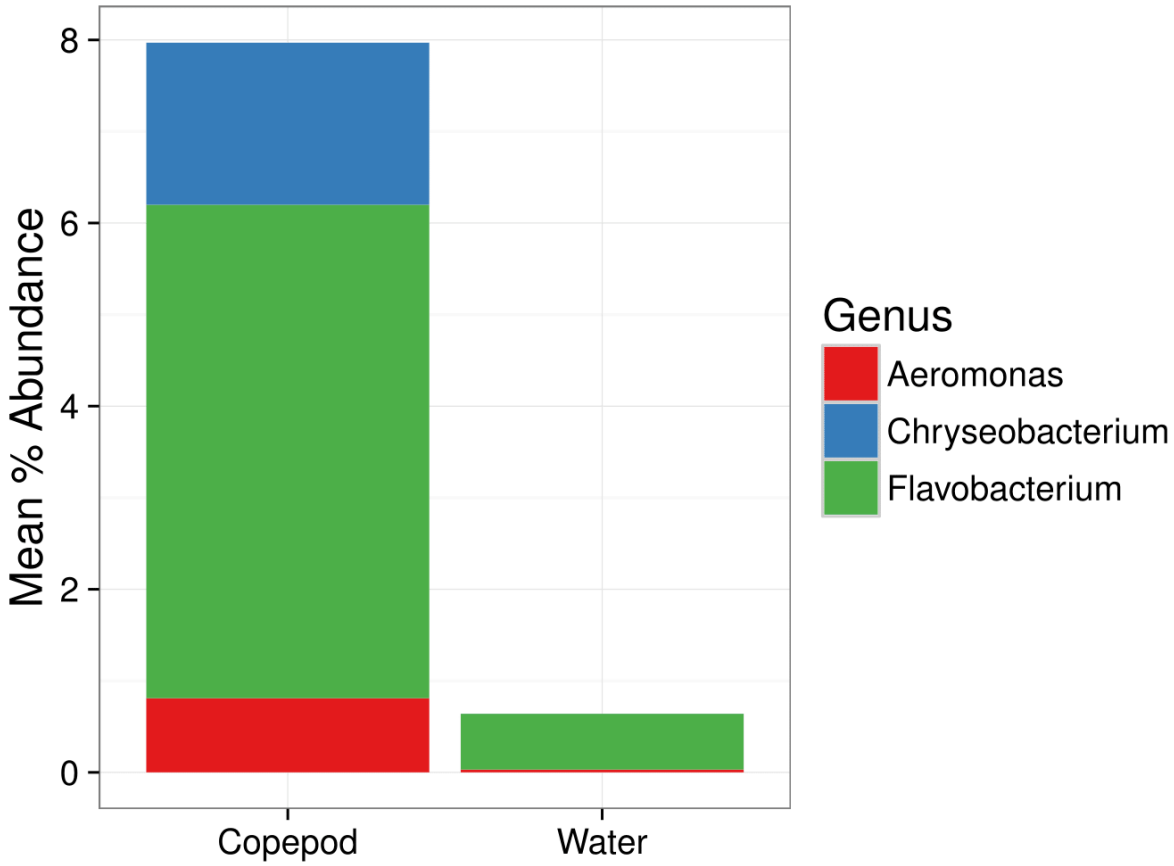

Figure 2. **Abundance of pathogenic fish genera associated with the copepod host *versus* water.** Mean percent abundance of the genera *Aeromonas*, *Chryseobacterium*, and *Flavobacterium* was averaged across all copepod (n=14) and all water (n=14) samples in terms of percentage of 16S rRNA reads classified to each genus *via* taxonomic classification with the greengenes database. These three genera, which contain prospective fish pathogens, were significantly enriched in copepod-associated bacterial communities (Wald test, adjusted *P* value < 0.001 for all three comparisons).

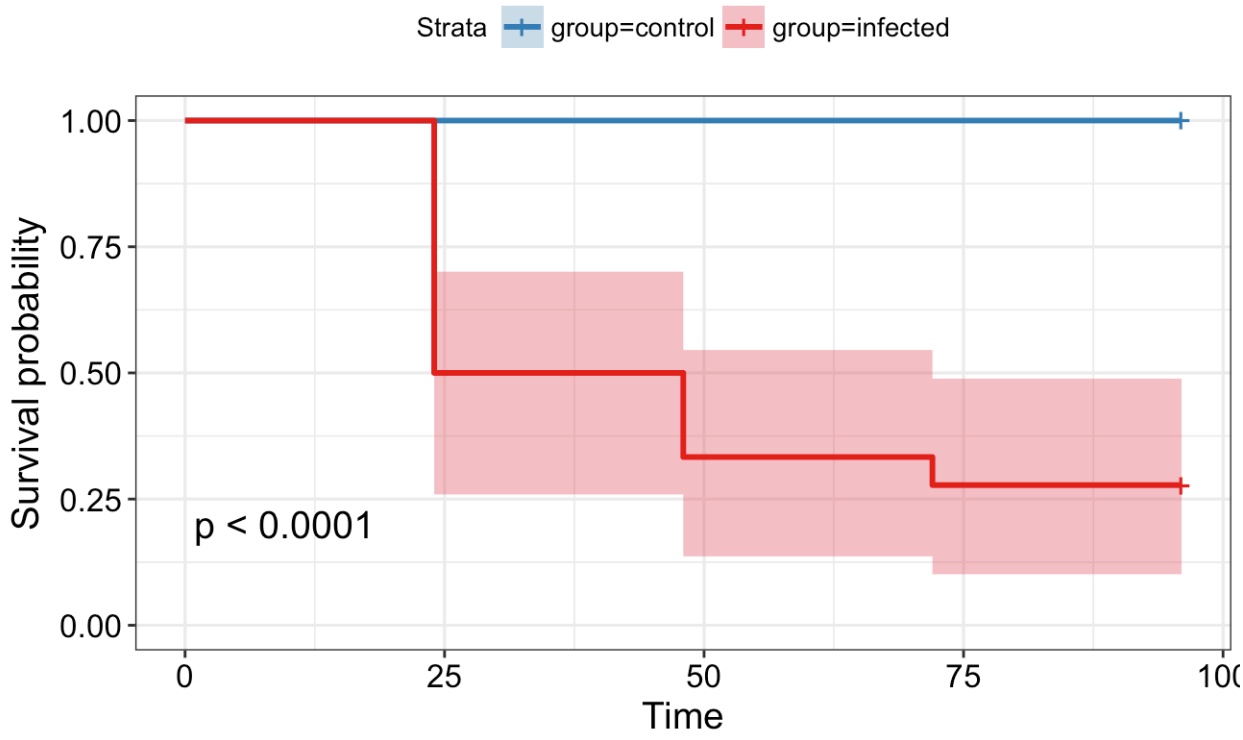

Figure 3: Survival probability of zebrafish after exposure to Fc Col1 – Infection data from three independent trials of zebrafish infection. In each of the three trials six fish in each group were injected with one of two treatments: 50µL sterile Modified Shieh (control) or 50µL Fc Col1 grown in Modified Shieh (Fc Col1). Survival probability indicate the rate of survival at the indicated 24-hour period. 95% confidence intervals are shown as light shading surrounding lines. Modified Shieh Optical Density was 270 for trial 1, 282 for trial 2, and 318 for trial 3.

Table 1 – Sample Metadata

Location	Environment	Water Type	Salinity (PSU)	Temperature (°C)	Invasive Clade	Latitude	Longitude	Continent
Lake Ontario	Copepod	Fresh	0.0	29.7	Northeast	43.307	-77.706	N America
Lake Ontario	Water	Fresh	0.0	29.7	Northeast	43.307	-77.706	N America
Cocodrie Bayou	Copepod	Salt	4.9	17.0	South	29.254	-90.664	N America
Cocodrie Bayou	Water	Salt	4.9	17.0	South	29.254	-90.664	N America
Lake Eufaula	Copepod	Fresh	0.0	10.0	South	35.146	-95.627	N America
Lake Eufaula	Water	Fresh	0.0	10.0	South	35.146	-95.627	N America
Frisian Lake	Copepod	Fresh	0.0	13.4	Europe	53.031	5.729	Europe
Frisian Lake	Water	Fresh	0.0	13.4	Europe	53.031	5.729	Europe
Ijsselmeer	Copepod	Fresh	0.0	13.1	Europe	52.699	5.290	Europe
Ijsselmeer	Water	Fresh	0.0	13.1	Europe	52.699	5.290	Europe
Ohio River	Copepod	Fresh	0.0	22.8	South	38.260	-85.747	N America
Ohio River	Water	Fresh	0.0	22.8	South	38.260	-85.747	N America
Montmagny	Copepod	Salt	5.0	NA	Northeast	46.990	-70.550	N America
Montmagny	Water	Salt	5.0	NA	Northeast	46.990	-70.550	N America
Lake Michigan	Copepod	Fresh	0.1	15.7	Northeast	43.051	-87.882	N America
Lake Michigan	Water	Fresh	0.0	15.7	Northeast	43.051	-87.882	N America
Markemeer	Copepod	Fresh	0.0	12.1	Europe	52.574	5.033	Europe
Markemeer	Water	Fresh	0.0	12.1	Europe	52.574	5.033	Europe
Western Scheldt	Copepod	Salt	7.0	11.7	Europe	51.302	4.286	Europe
Western Scheldt	Water	Salt	7.0	11.7	Europe	51.302	4.286	Europe
Taylor Bayou	Copepod	Salt	4.0	16.0	South	29.883	-94.051	N America
Taylor Bayou	Water	Salt	4.0	16.0	South	29.883	-94.051	N America
Lake Texoma	Copepod	Fresh	0.9	13.0	South	33.882	-96.797	N America
Lake Texoma	Water	Fresh	0.9	13.0	South	33.882	-96.797	N America
L'Isle Verte	Copepod	Salt	13.7	5.9	Northeast	48.002	-69.423	N America
L'Isle Verte	Water	Salt	13.7	5.9	Northeast	48.002	-69.423	N America

Table 2 – Bacterial genera which were significantly enriched in the Copepod Microbiome versus the bacterioplankton from the surrounding water. Classification by genus was done by alignment to the Greengenes taxonomy. *P*-values are from Wald tests on variance-stabilized OTU counts and were adjusted for multiple testing.

Phylum	Class	Order	Family	Genus	Log Fold Change	Adjusted *P* value
Bacteroidetes	Saprospirae	Saprospirales	Saprospiraceae	unclassified	-8.636575	0
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Rubrivivax	-8.09993	0
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	Leadbetterella	-5.896205	0
Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	Aeromonas	-5.275551	0.0000004
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	Flavobacterium	-4.759589	0.0000015
Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	Shewanella	-4.386033	0.0000055
Actinobacteria	Actinobacteria	Actinomycetales	Corynebacteriaceae	Corynebacterium	-5.22193	0.0000445
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	Pseudomonas	-4.297981	0.0000513
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Delftia	-3.750081	0.0000579
Proteobacteria	Gammaproteobacteria	Alteromonadales	Chromatiaceae	Rheinheimera	-4.646792	0.0000643
Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	Renibacterium	-4.426645	0.0001411
Bacteroidetes	Flavobacteriia	Flavobacteriales	Weeksellaceae	Chryseobacterium	-6.437576	0.0001548
Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	unclassified	-5.056017	0.0002596
GN02	BD1-5	unclassified	unclassified	unclassified	-3.042396	0.0002645
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	Stenotrophomonas	-3.234051	0.0003624
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	Polaromonas	-4.480607	0.000407
Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae	unclassified	-4.323869	0.0005665
Proteobacteria	unclassified	unclassified	unclassified	unclassified	-3.842931	0.0006584
Cyanobacteria	Chloroplast	Chlorophyta	Chlamydomonadaceae	unclassified	-5.205313	0.0008513
Proteobacteria	Gammaproteobacteria	Oceanospirillales	Halomonadaceae	Halomonas	-3.334231	0.0014559
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	unclassified	-4.49808	0.0023209
Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	Enterovibrio	-5.342666	0.0073836
Proteobacteria	Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	unclassified	-2.794012	0.0077864

Table 3. Infection data from trial run on August 27, 2016. Six fish in each group were injected with one of the three treatments: 50µL sterile Modified Shieh, 50µL of Fc Col1 grown in 1/10 CYE, or 50µL Fc Col1 grown in Modified Shieh. Numbers indicate the number fish deaths occurring during the indicated 24-hour period. Modified Shieh Optical Density was 270, 1/10 CYE Optical Density was 44.

Hours	Control	Modified Shieh	1/10 CYE
24	0	4	1
48	0	1	1
72	0	0	0

Table 4. Infection data from trial run on September 5,2016. Six fish in each group were injected with 50µL of either sterile modified Shieh (served as control), Fc Col1 grown in Modified Shieh, or Fc Col 1 grown in TYES. Numbers indicate the number of fish deaths occurring during the indicated 24-hour period. Modified Shieh Optical Density was 282, TYES Optical Density was 181.

Hours	Control	Modified Shieh	TYES
24	0	2	0
48	0	0	1
72	0	1	1

Table 5. Infection data from trial run on September 19, 2016. Six fish in each group were injected with one of the five different treatments: Fc Col1 injection of 50µL grown in Modified Shieh, Fc Col1 injection of 10µL grown in Modified Shieh, Fc Col1 injection of 50µL grown in TYES, Fc Col1 injection of 10µL grown in TYES, or a control group injected with 50µL of sterile Modified Shieh. Numbers indicate the number of fish deaths occurring during the indicated 24-hour period. Modified Shieh Optical Density was 318, TYES Optical Density was 164.

Hours	Control	Modified Shieh (10µL)	Modified Shieh (50µL)	TYES (10µL)	TYES (50µL)
24	0	2	3	2	0
48	0	2	2	1	0
72	0	0	0	0	0

Chapter 3

Population Genomic Signatures of Selection in Ion-Transport Genes during Saline-to-Freshwater Invasions of the Copepod *Eurytemora affinis*

Authors: Martin Bontrager[1], Tiago Ribeiro[2], Joana C. Silva[3, 4], Carol Eunmi Lee[5]

[1] **Laboratory of Genetics, University of Wisconsin-Madison,** 425 Henry Mall, Madison, WI 53706
[2] Department of Zoology, University of Wisconsin-Madison, 145 Noland Hall, 250 North Mills St, Madison, WI 53706
[3] Institute for Genome Sciences, University of Maryland School of Medicine, 801 W. Baltimore Street, Baltimore, MD, 21201
[4] Department of Microbiology and Immunology, University of Maryland School of Medicine,685 W. Baltimore Street, Baltimore, MD, 21201
[5] Center for Rapid Evolution (CORE), University of Wisconsin-Madison, 430 Lincoln Drive, Madison, WI 53706

Abstract

Climate change is imposing swift and dramatic environmental changes on many species, and will continue to do so in the coming years. Species must either adjust *via* a plastic response, migrate, evolve, or perish in the face of novel selective pressures. One such pressure that is expected to occur is the freshening of waters in coastal and estuarine habitats. Most saline adapted organisms cannot survive in freshwater habitats, and are threatened by global freshening of waters. However, the saline copepod *Eurytemora affinis* has invaded freshwater habitats multiple times from independent source populations in the past century. These invasions were accompanied by rapid evolution of protein activity and gene expression in freshwater populations, making the species a powerful model in which to investigate the genetics underlying freshwater adaptation in ecological timescales. In this study, we took a population genomics approach to investigate genome-wide patterns of genetic differentiation among four populations of *E. affinis* in the St. Lawrence drainage across the saline-freshwater boundary and to identify candidate genes under selection during adaptation to freshwater. We found that the highest genomic differentiation among copepod populations was between the most saline population and all other populations, while the population in brackish water was more closely related to invasive freshwater populations. We also found homologs of ion transport genes including Na^+/K^+-ATPase, V-type H^+ -ATPase, Na^+/H^+ antiporter, and $Na^+/K^+/2Cl^-$ cotransporter in genomic regions of high differentiation between saline and freshwater populations. Not only were ion transport genes enriched in differentiated regions between saline and freshwater populations, but some of the same genes were present in highly differentiated regions between high and low-salinity populations. In addition to illuminating some of the genes which may be responsible for freshwater adaptation, these results suggest that adaptation to freshwater in *E. affinis* proceeds in

a stepwise fashion from high salinity to intermediate salinity and finally to freshwater. The evolution of ion uptake is likely to be a major constraint on the survival of saline species in freshening environments, and populations with a history of exposure to low-salinity conditions are likely to be those that survive.

Introduction

Rapid climate change will affect nearly every population on the planet within the next century, and represents a significant, potentially catastrophic, threat to global biodiversity and integrity in both terrestrial and aquatic ecosystems (Cheung *et al.* 2009; Bellard *et al.* 2012; Pimm *et al.* 2014). The ecological disruption caused by climate change is projected to impose severe selective pressures on populations, with best estimates suggesting that up to one in six species may risk extinction (Thomas *et al.* 2004; Urban 2015). Species must respond to increased temperatures, changing weather patterns, seasonal shifts, and a host of other conditions imposed by climate change or march towards extinction.

Range shifts and phenotypic plasticity may not be sufficient mechanisms for species to cope with major environmental shifts like those associated with climate change. In such cases, rapid adaptation via evolutionary processes is the only way for species to escape extinction. Local and global freshening of brackish and saline water is a major environmental shift brought on by climate change that is expected to impact many regions worldwide (Helm *et al.* 2010; Durack *et al.* 2012; Hegerl *et al.* 2015). However, there are significant gaps in our understanding of the evolutionary response of species to such rapid shifts in environmental salinity. Some models suggest that local populations might undergo steep decline (Thøgersen *et al.* 2015), but

we have very little knowledge about how such freshening might be expected impact species within the affected areas and how species might adapt to freshwater environments.

Historically, the colonization of dilute freshwater environments is a major biogeographic barrier to species range expansion (Hutchinson 1957; Little 1983; Miller & Labandeira 2002). Despite this hurdle, saline to freshwater species invasions are not especially rare, and invasions from estuarine or brackish-water habitats into freshwater are particularly common (Lee & Bell 1999; Ricciardi & MacIsaac 2000; May *et al.* 2006b). In many of these cases of invasion from saline or brackish water, establishment in freshwater is accompanied by rapid evolution of the population in response to novel conditions in the freshwater environment (May *et al.* 2006a; Lee *et al.* 2011, 2012). It has been hypothesized that a history of disturbances in ancestral habitats might promote the evolvability of certain populations, lending them the raw genetic variation necessary for rapid adaptation to novel conditions (Di Castri 1989; Lee 2011). The environmental changes encountered during the introduction of species to new environments make invasive species a valuable model for evaluating fundamental questions about rapid evolution and adaptation to novel habitats (Waddington 1965; Lee 2011).

A fundamental area of research regarding adaptation to freshwater involves identifying the genes that are responsible for adaptation to freshwater and the physiological mechanisms that permit such adaptation to occur. One clear set of targets which may be under selection during freshwater adaptations are genes involved in ion-transport. Upon removal from the saline ionic environment species must find a way to maintain body fluid homeostasis in the face of ion paucity in the freshwater environment. Previous studies have shown fixed evolutionary differences in physiological regulation of hemolymph osmolality and increased ion uptake in *E. affinis* in freshwater adapted populations relative to their saline ancestors (Lee *et al*. 2011, 2012). This

maintenance of homeostasis requires evolutionary changes in gene expression and/or function in ion-transport genes, including V-type H^+ ATPase (VHA), Na^+,K^+-ATPase (NKA), and the Na^+/H^+ antiporter (NHA). These genes are involved in ionic regulation in other invertebrates (Tsai & Lin 2007; Xiang *et al*. 2012), and shift in expression in *E. affinis* in laboratory freshwater-adapted animals during shifts in salinity (Lee *et al*. 2011).

In this study, we set out to identify genes which underlie adaptation to freshwater in an invasive lineage of the calanoid copepod *Eurytemora affinis* and to determine whether those genes which show signatures of selection in the wild populations during freshwater invasions agree with a set of candidate genes underlying freshwater adaptation from previous studies (Lee *et al.* 2011; Posavi 2015). *Eurytemora affinis* has a census size in the billions in many coastal environments; especially in estuaries and salt marshes in the Northern Hemisphere (Morgan *et al.* 1997; Lee 2000; Peitsch *et al.* 2000). *E. affinis* has a recent history of multiple independent invasions from saline environments into freshwater reservoirs worldwide; likely due to transport along with ship ballast water (Lee 1999). Despite the saline-to-freshwater biogeographic barrier, the *E. affinis* species complex has invaded freshwater multiple times from independent source populations. These natural traits of the *E. affinis* species complex make it an excellent model in which to examine rapid adaptation to freshwater environments.

Much previous work to understand rapid evolution has focused on population genetics of single genes in populations, comparative analysis of genomes from only a few individuals, or population genomics across a single habitat shift. We sequenced the full genome of hundreds of individuals from several populations across the saline-to-freshwater barrier. By taking a pooled population genomics approach to address evolutionary questions across this biogeographic barrier, we were able to identify genes that showed signatures of selection in *E. affinis* during the

saline-to-freshwater transition and to further illuminate the genomic differences between copepod populations.

Materials and Methods

Sampling Wild Copepod Populations

We sampled wild *Eurytemora affinis* populations from one invasive lineage, consisting of two ancestral saline and two derived freshwater populations along the St. Lawrence estuary and in the Great Lakes of North America (Fig. 1). Within the past century, this lineage of copepods has invaded from estuarine saline-water populations in the St. Lawrence into the Great Lakes (Lee 1999). Our sampling included one site from the saline tidal pools near L'Isle Verte, QC Canada, one population in slightly less saline, brackish, water near Montmagny, QC, Canada, and two freshwater populations, one from Lake Ontario, NY, USA and one from Lake Michigan, WI, USA (Table 1). From each sampling location, we isolated 100 adult *E. affinis* in an approximately 1:1 ratio of males to females and preserved the animals at -80°C. The strength of this design is its reliance on the natural evolutionary experiment provided by the freshwater invasion of *E. affinis* into the Great Lakes and the statistical power derived from sampling hundreds of individuals from the enormous populations of *E. affinis* at each location.

Sequencing and Alignment

All 100 copepods sampled from each population were pooled for DNA extraction and whole-genome sequencing (Pool-seq). Pool-seq is a cost effective method for variant discovery and the estimation of allele frequencies in populations (Futschik & Schlötterer 2010; Schlötterer

et al. 2014). We extracted DNA from each pool of 100 animals using the UltraClean DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) per manufacturer recommendations and with an added 10-minute incubation at 65°C prior to mechanical lysis to increase extraction efficiency. Sequencing libraries were created with the Illumina Nextera DNA Sample Prep Kit (Illumina Inc., San Diego, CA, USA) and sequencing was performed on the Illumina HiSeq 2000 using one lane per sample for the four total samples.

We used the software package Trimmomatic (Bolger *et al.* 2014) to filter and trim low quality reads, adapter sequences, and unpaired reads from the raw sequencing read data. Repetitive regions of the draft reference *Eurytemora affinis* genome were masked with RepeatMasker 4.0.6 (Smit *et al.* 2013) to prevent alignment to those regions, and the filtered, trimmed reads were aligned to the masked reference genome. Read alignment was performed in a pipeline that began with BWA-MEM (Li 2013) for an initial round of alignment, followed by a second round of alignment of unaligned reads with Stampy (Lunter & Goodson 2011) to improve the mapping of divergent reads. After alignment we removed duplicate reads and re-aligned around indels using Picard and GATK IndelRealigner (McKenna *et al.* 2010). We processed the read alignments with SAMtools to count variant nucleotides at each position in the reference genome (Li *et al.* 2009; Li 2011). We detected and excluded aligned positions within five positions up- and down-stream of indels to improve the quality of single nucleotide polymorphism (SNP) calling. Using a sliding window approach with 5kb overlapping windows, we measured genome-wide nucleotide diversity and identified all SNPs among the populations with the software PoPoolation2 (minimum allele count = 6, minimum coverage = 20, and maximum coverage = 98th percentile) (Kofler *et al.* 2011).

Detecting Signatures of Selection

In order to detect genomic regions with signatures of selection between freshwater and saline populations we calculated pair-wise F_{ST} (Wright 1949), among all samples, and an F_{ST}-based index called Population Branch Excess (PBE, Yassin *et al.* 2016), an extension of the Population Branch Statistic (PBS, Yi *et al.* 2010) that measures the excess of differentiation in a single focal population among three total populations. Pair-wise F_{ST} was calculated for every nucleotide position with allele frequency differences between populations using PoPoolation2 and used to calculate F_{ST} for 5 kb windows at 1 kb intervals. Loci under positive selection in a given population are expected to differentiate more from other populations than neutral loci, and therefore genome-wide measures of F_{ST} are informative to identify candidate loci under selection in different populations (Lewontin & Krakauer 1973; Beaumont 2005; Nielsen 2005). To identify genome regions with signatures of selection between populations we defined regions of high differentiation as regions which were in the 99.5th percentile of F_{ST} or PBE values. These highly differentiated regions between populations were genomic regions which were candidates under selection in either, or both, of the populations in the comparison.

Pair-wise F_{ST} values were transformed and used to calculate the PBE statistics according to Yassin *et al.* (2016). Our population sampling design, given two ancestral saline populations and derived freshwater populations, lends itself well to the implementation of the PBE statistic to test for signatures of selection in focal freshwater populations. We calculated the PBE statistic genome-wide independently using both freshwater populations as focal populations, in each case calculating excess differentiation from the pair of saline populations. In this way, we were able to measure the excess differentiation that was unique to freshwater focal populations. After calculating the PBE statistic in 5kb windows, we used the same threshold as in F_{ST}-based

analyses to determine regions of high differentiation, defining such regions as those in the 99.5th percentile. After identifying regions of high differentiation in both freshwater populations, we took the intersection of highly differentiated PBE regions between both sets of comparisons and used this intersection for subsequent analysis. We assumed that regions of high differentiation that were shared between both freshwater populations were regions that were likely to be involved in adaptation to freshwater as opposed to regions involved in adaptation to local conditions of each lake or noise introduced from alignment errors.

Annotation and Enrichment Analysis

We identified genes overlapping genomic regions showing signatures of selection (i.e. highly differentiated regions) by using the bedtools package to find the intersect of these regions with the location of gene models and annotated genes in the *E. affinis* reference genome (Quinlan & Hall 2010). In some cases, gene models in the reference genome lacked robust validation. To avoid this problem, we also found the intersection of genomic regions showing signatures of selection and assembled transcripts from a full transcriptome sequencing run instead of predicted gene models, but our results were identical. We annotated gene models within highly differentiated regions using a workflow described in (Wit *et al.* 2012). Briefly, we aligned translated CDS nucleotide sequences from gene models against sequences in the NCBI nr, UniProt SwissProt, and TrEMBL databases. Top hits against those databases, if they existed, were used to annotate our gene models with likely functions via Gene Ontology (GO) category terms.

To assess enrichment of functional categories and GO terms, we compared GO terms from genes in regions of high differentiation to GO terms from the full set of *E. affinis* gene

models genome-wide, using the same method as described above to annotated the function of the full set of gene models (Wit *et al.* 2012). We performed a Fisher's exact test of GO term enrichment with the topGO package in R identify GO terms which were significantly enriched in the set of shared highly-differentiated genes (Alexa & Rahnenfuhrer 2016).

Results

DNA Sequencing and Polymorphism Detection

Each of the four *E. affinis* populations was represented by between 308 and 385 million raw population genomic sequencing reads per sample (Table 2). After filtering raw reads to exclude low quality sequences and adapter sequences, and trimming low-quality read ends, we subsequently aligned trimmed and filtered reads to the reference genome. The reference *Eurytemora affinis* genome against which we aligned population contains 495 Mbp on 6,899 scaffolds and was generated from an inbred laboratory population of *Eurytemora affinis* derived from the population of copepods in L'Isle Verte, Quebec, Canada. We aligned between 165 – 229 million reads to the reference per population. That corresponds to an average depth of coverage between ~22x – 32x per population sample (Fig.2). The population sample from L'Isle Verte, Quebec, Canada suffered from a lack of sequencing coverage, being 10x lower coverage than other samples due to lower initial depth of sequencing coverage.

After masking repetitive regions in the reference genome and excluding positions within 5bp of indels, SNP-calling with PoPoolation2 identified 753,142 high-quality polymorphic sites (minimum allele count = 6, minimum coverage = 20, maximum coverage = 98th percentile) across all populations. This represents approximately one high-quality polymorphic site per

657bp across the genome, although this is a very conservative estimate of the overall number of polymorphic sites given the low quality of the reference assembly and the exclusion of low-frequency polymorphic sites through our hard filtration. Median nucleotide diversity (π, Nei & Li 1979) genome-wide ranged from 4.0×10^{-3} in the Montmagny, Quebec population to 6.2×10^{-3} in the population from L'Isle Verte, Quebec (Table 2). Perhaps surprisingly given the recency of their invasion, freshwater populations from Lake Ontario ($\pi = 4.51 \times 10^{-3}$) and Lake Michigan ($\pi = 4.3 \times 10^{-3}$) had slightly higher nucleotide diversity than the ancestral saline population from Montmagny, Quebec ($\pi = 4.03 \times 10^{-3}$).

Genomic Differentiation between Freshwater and Saline Populations

The largest divergence among the populations, as measured by F_{ST}, was found between the population from L'Isle Verte, QC, Canada and all other populations, including the brackish-water population from Montmagny, QC, Canada (Table 3). Median F_{ST} genome-wide between Montmagny, Lake Ontario, and Lake Michigan was approximately 0.02, while median F_{ST} between L'Isle Verte and all other populations was 0.11. Copepods from Lake Ontario and Montmagny had the lowest level of population differentiation ($F_{ST} = 0.019$) despite the salinity difference between their environments. The highest population differentiation was between the populations from Lake Michigan and L'Isle Verte ($F_{ST} = 0.114$), which are the furthest from one-another geographically.

We performed PBE analysis independently with each of the two freshwater populations as the target population and using both saline populations as ancestors, and then took the subset of highly-differentiated regions that were shared between both freshwater populations. There were 258 5kb windows on 106 scaffolds in the top 0.5% of PBE differentiated regions that were

shared between both freshwater populations (Fig. 3). This represents an overlap of approximately 12.5% between the most differentiated regions of both freshwater populations. If highly differentiated regions were randomly distributed between the two comparisons, we would have expected approximately 0.5% overlap. Many of these differentiated windows were clustered together in tandem on individual scaffolds as opposed to being randomly distributed throughout the genome.

When we took the intersection of these regions of high PBE differentiation in freshwater with the locations of annotated genes or gene models generated from *E. affinis* reference genome, there were 109 different annotated genes or gene models which overlapped with windows of high differentiation in freshwater. Several of these gene models had no homologous matches when searched against the UniProt or NCBI non-redundant protein databases, but 32 gene models did have homology to known genes (Table 4). Among these, homologs of Na/H antiporter (NHA), V-Type H^+ ATPase (VHA), Na/K ATPase (NKA), and Na-K-Cl cotransporter (NKCC) genes were represented among the genes in regions of high PBE differentiation between both freshwater populations and their saline ancestors. The results of GO enrichment analysis of these highly differentiated genes showed the functional enrichment of several terms, most of which are involved in ion-transport including hydrogen-ion transmembrane transport activity, solute-proton antiporter activity, and solute-cation antiporter activity (Table 5).

We also compared regions of high differentiation between the two saline populations, in that case using F_{ST} as a metric of population differentiation as opposed to PBE since only two populations were being compared. Overall differentiation between L'Isle Verte and all other populations, including the saline population in Montmagny, Quebec, was high (see Table 3). We used the same criteria for identifying regions of high F_{ST} differentiation, selecting those regions

in the 99.5th percentile. When we compared this list of regions with the 258 regions of overlap between both freshwater populations, we found that 30 regions (~11.5%) were highly differentiated between both the two saline populations and the saline and freshwater populations (Fig. 4). In a random draw of regions, we would expect that perhaps 1-2 (0.05%) regions of high differentiation would overlap between these two sets of comparisons. Annotated gene models in these overlapping regions include a V-type ATPase subunit and an Na-K-Cl co-transporter (NKCC) homolog.

Discussion

Several recent models have shown that climate change is expected to have a significant impact on salinity in aquatic environments via changes in the global water cycle (Helm *et al*. 2010; Durack *et al*. 2012; Hegerl *et al*. 2015). The melting of enormous freshwater reserves now sequestered in the polar ice caps may affect ocean salinity on a large scale (Munk 2003; Rabe *et al*. 2011), while shifting patterns of precipitation will impact salinity on a local scale (Bintanja & Selten 2014; Loder *et al*. 2015). In the Baltic Sea, for example, models suggest that water might freshen *via* a mean decline of 3 PSU over the next century, leaving the entire sea considerably less saline (Meier 2006). Around the St. Lawrence estuary and Atlantic Canada, ocean salinities are expected to decline by anywhere between 1-4 PSU across the region within the next 50 years (Loder *et al*. 2015). Our results suggest that the evolution of ion transport is likely to be critical for species to adapt to freshening conditions in such environments.

Ion Transport Evolution during Freshwater Adaptation

Ion transport is an enormously important and energetically costly activity, so much so that it has been dubbed the "pacemaker of cellular metabolism" (Ismail-Beigi & Edelman 1970;

Asano *et al.* 1976). It is estimated that in some species of teleost fish, for example, osmoregulation may account for up 10-15% of total metabolism (Kirschner 1993; Kidder *et al.* 2006). Thus, the energetic cost alone of maintaining internal ionic homeostasis is an enormous pressure facing saline species introduced to ion-sparse environments. Despite the importance of ion-transport to survival of aquatic species, we know very little about the exact mechanisms of ion-transport across species; and the changes that occur to ion-transport during saline-to-freshwater transitions are mostly unknown. Several models of ion-transport have been proposed that seek to explain the transport of ions from a dilute environment across a steep concentration gradient to maintain internal homeostasis.

Based on data collected through decades of work, a hypothetical model for ion-transport in crustaceans has emerged which includes the action of Na^+/H^+ antiporter (NHA), Na^+/K^+-ATPase (NKA), V-type H^+ -ATPase (VHA), and $Na^+/K^+/2Cl^-$ cotransporter (NKCC), among others (Towle & Weihrauch 2001; Rheault *et al.* 2007; Xiang *et al.* 2012; Posavi 2015; Lee 2016; Gerber *et al.* 2016). We found that homologs of ion transport genes and/or subunits including NHA, NKA, NKCC, and VHA were all present in regions showing signatures of selection between saline and freshwater populations. GO term enrichment analysis verified important role of ion-transport evolution, as terms related to ion transport were significantly enriched in our set of highly-differentiated genes between saline and freshwater populations. Not only were ion transport genes enriched in highly differentiated regions between saline and freshwater populations, but some of the same genes were enriched in comparisons between high- and low-salinity populations. This phenomenon suggests that changes in allele frequency in ion-transport genes in brackish water environments are an intermediate stage in the adaptation to freshwater.

The H^+ V-type ATPase proton pump (VHA) seems to play a critical role in facilitating the uptake of ions at low salinity environments (Wieczorek *et al.* 1991, 1999; Jensen *et al.* 2002; Chambrey *et al.* 2013). We found that a homolog of V-Type ATPase subunit D inhabited a 5kb window was a significant PBE outlier between freshwater and saline populations. This same gene was also a significant F_{ST} outlier between the most saline population in L'Isle Verte, Quebec and a population at lower salinity in Montmagny, Quebec. Several groups have proposed that the model of action of this H^+ pump is that VHA pumps H^+ ions out of the cell into the external environment, and the resulting voltage gradient facilitates the uptake of ions into the cell through associated transport proteins (Ehrenfeld & Klein 1997; Beyenbach & Wieczorek 2006). In several cases, including in freshwater fish, VHA has been proposed as a major player in the uptake of Na^+ ions from the external environment (Ehrenfeld *et al.* 1989; Wieczorek *et al.* 1999). Evolution of VHA has been hypothesized to be critical to the colonization of both fresh water and terrestrial habitats across species (Morris 2001; Tsai & Lin 2007; Lee *et al.* 2011). Ours is not the first study to implicate VHA evolution in response to freshwater. In fact, VHA or its expression is frequently identified as having undergone evolution during freshwater adaptation across several species of fish including killifish and stickleback (Hohenlohe *et al.* 2010; Jones *et al.* 2012; Kozak *et al.* 2014). Previous work on *Eurytemora affinis* revealed that V-Type ATPase exhibits evolutionary shifts in enzyme activity during freshwater invasions, and that such shifts in activity can occur on extremely rapid time-scales and in fact can be induced in a laboratory setting in around 12 generations (Lee *et al.* 2011). Our results strengthen the conclusion that the evolution of H^+ V-type ATPase is critical to rapid adaptation to freshwater and to the establishment of an H+ ion-gradient.

Recent work has uncovered a novel family of electrogenic sodium/hydrogen antiporter (NHA) genes in *Anopheles gambiae* and *Drosophila melanogaster,* and subsequent analysis showed that members of the gene family are widespread across taxa (Rheault *et al*. 2007; Day *et al*. 2008; Xiang *et al*. 2012). The existence of such electrogenic antiporters was foreseen by Weiczorek and colleagues, who had hypothesized an electrogenic antiporter which in cooperation with VHA worked to exchange cations with H^+ ions from the external environment (Wieczorek *et al*. 1991; Beyenbach & Wieczorek 2006). The nature and identity of this "Wieczorek exchanger" was unknown, but the NHA family of proteins seems fulfill the necessary criteria. Analysis of the *Eurytemora affinis* genome shows the presence of 8 paralogous copies of the sodium/hydrogen antiporter (NHA) gene, 7 of which are arranged in tandem in a single region (CE Lee, unpublished data). Highly differentiated PBE regions overlapped 3 copies of the NHA gene in this region of tandem NHA duplicates. In fact, we found regions of high differentiation across this entire 7-gene tandem duplication region, at the exclusion of any other region on the scaffold, albeit not always within annotated genes but sometimes up- or down-stream of coding regions. If the NHA gene family is indeed the missing Wieczorek exchanger, the rapid evolution of this region of NHA paralogs during a saline-to-freshwater invasion lends significant support to the hypothesis that the model of action for cation exchange in coordination with V-type H^+ ATPase is an electrogenic co-transporter like NHA as opposed to an electroneutral co-transporter like NHE.

Another candidate gene of interest was the Na^+/K^+ ATPase (NKA) ion transporter. The classic model of ion uptake from the external environment has NKA as the primary active transporter involved in Na^+ transport into the cell (Lucu & Towle 2003; Kirschner 2004). However evidence suggests that the action of NKA alone is insufficient to account for Na^+ uptake in dilute (freshwater) environments (Larsen *et al*. 1996; Weihrauch *et al*. 2004). In our analyses, an NKA

subunit beta homolog was in a highly differentiated region between freshwater and saline populations of *E. affinis*. We have previously found evolutionary shifts in enzyme activity and expression of NKA between freshwater and saline laboratory populations, with NKA downregulated in freshwater animals relative to saline (Lee *et al*. 2011; Posavi 2015). Taken together with the differentiation of NKA regions between wild freshwater and saline populations, it seems likely that NKA plays a critical role in adaptation to freshwater environments. Other studies in amphipods and fish found stable differences in enzyme activity between animals in fresh or brackish *versus* saline water, further supporting an important functional role that NKA plays between populations in fresh and salt water (Brooks & Mills 2006; Bystriansky *et al*. 2006; Nilsen *et al*. 2007; McCormick *et al*. 2009).

Many other genes showed signatures of selection in our analyses, although they may not have a role in ion-transport. A homolog of the human kidney mitochondrial carrier protein-1 gene (KMCP1; Haguenauer *et al*. 2005) also showed signatures of selection between both saline and fresh water and between the two saline populations. This gene is a member of the SLC25 family of mitochondrial solute carriers and in other systems is responsible for the transport of metabolites across the inner mitochondrial membrane (Palmieri 2004). Another interesting candidate that showed signatures of selection in all comparisons was a homolog of fatty-acid binding protein-9 (FABP9), a member of a family of proteins which binds with high-affinity to hydrophobic ligands (Esteves & Ehrlich 2006). Recent work has shown that this protein is involved in the innate immune response of crustaceans (Cheng *et al*. 2013), which is interesting given the unique microbial environment that *E. affinis* encounters during invasion from saline to fresh water (see Chapter 1).

Population Structure and Genomic Differentiation among Copepod Populations

The two freshwater populations and the population from brackish water near Montmagny had very similar nucleotide diversity and overall F_{ST} differentiation among themselves, while the L'Isle Verte population was more diverse and highly differentiated from all other populations. The invasion of *E. affinis* into the Great Lakes is well-documented and occurred approximately 50-100 years ago (Anderson & Clayton 1959; Faber & Jermolajev 1966; Gannon 1974). Populations of *E. affinis* in the St. Lawrence estuary are significantly older, however, and were probably established after the retreat of the glaciers around 15,000 years ago (Lee 2000). The low nucleotide diversity and high similarity between the Montmagny estuarine population and the two freshwater populations raises the possibility that the transition from the moderate salinity environment (~5 PSU) near Montmagny into freshwater was relatively recent, and that Montmagny was likely the source population for freshwater invasive populations in the Great Lakes. Indeed, copepods from the populations in Lake Ontario and Montmagny show the lowest levels of differentiation between each other than do any other pair of populations. The higher nucleotide diversity in the population from L'Isle Verte and its differentiation from all other populations suggests that the split between copepods in the higher salinity environment of L'Isle Verte and all other populations may be a more ancestral split. Previous work has shown that there is a biogeographic barrier in water at ~5PSU across which many aquatic invertebrates are unable to cross (Khlebovich & Abramova 2000). That the population of *E. affinis* in Montmagny is most similar to other freshwater populations raises the possibility that the transition from saline water to water of intermediate salinity around 5 PSU is indeed presents major physiological challenges, but that the transition from intermediate salinity to freshwater is more readily accomplished.

It has been hypothesized that a history of disturbance, anthropogenic or otherwise, in the

native range may predispose populations towards the colonization of novel habitats (Thebaud *et al.* 1996; Lee & Gelembiuk 2008; Hufbauer *et al.* 2012). Theoretical models show that under certain conditions, genetic variation might be selected for and maintained in a population undergoing fluctuating selection (Gillespie & Turelli 1989; Turelli & Barton 2004; Lee & Gelembiuk 2008; Lee 2016). Estuarine habitats often undergo dramatic environmental fluctuations. Copepod populations in the St. Lawrence estuary that gave rise to freshwater invaders live in habitats that are prone to dramatic seasonal shifts in salinity (Winkler *et al.* 2008). If it is the case that ancestral saline or estuarine populations undergo frequent disturbance, standing genetic variation in ancestral estuarine populations then might provide the raw materials upon which selection can act in the freshwater environment and thus contribute to the adaptive and invasive success of newly-established freshwater populations. This is especially relevant as it pertains to the situation that many species face under scenarios of global climate change, as populations which derive from regions with a history of disturbance might be predisposed to adapt to rapid climate change, while populations from regions of relative stability might lack the genetic variation necessary for rapid adaptation to occur.

Conclusion

Freshening of aquatic habitats *via* global climate change will present major physiological challenges for many saline and brackish-water species in the coming years. Uncovering the mechanisms and genomic patterns underlying freshwater adaptation is of fundamental importance to better understand species' responses to such rapid environmental changes. In this study, we identified genome-wide patterns of population differentiation between saline and freshwater populations of the recent freshwater invader *Eurytemora affinis,* and identified genes

in regions of high differentiation that showed signs of being under selection during the saline-to-freshwater invasion. We found that a subset of the same genes that showed signatures of selection between saline and freshwater populations also showed signatures of selection between the most saline population and an intermediate brackish water population, a pattern which is consistent with a stepwise invasion into freshwater from a mostly saline ancestral population through a brackish-water - though still saline - intermediate, and eventually into freshwater. Many of the genes we found in regions of high differentiation between saline and freshwater populations were involved in ion-transport, and GO analysis confirmed that several molecular functional terms involving ion-transport were enriched in highly differentiated regions between freshwater and saline populations. Several of these ion-transport genes play a significant role in adaptive response to freshwater in other systems, and were previously identified as differentially expressed between salinity conditions in laboratory populations of *E. affinis*. This enrichment of ion transport genes under selection during saline-to-freshwater invasions is to be expected given the environmental shift, but the scope and rapidity of such changes raises interesting questions about the maintenance of genetic diversity in ion transport genes and the mechanisms responsible for maintaining that diversity.

Acknowledgments

This material is based on work supported by the National Science Foundation Graduate Research Fellowship Program under Grant No. DGE-1256259. Funds for this project were provided by National Science Foundation OCE- 1046372 to Carol E. Lee and OCE-1046371 to Joana C. Silva. J. C. Silva and J. Munro directed library preparation and sequencing efforts.

Many people assisted with field collections of copepods, including J. Remfert, G. Winkler, K. Lee, Q. Langdon, and M. Carlson. T. Ribeiro provided valuable comments and discussion.

Literature Cited

Alexa A, Rahnenfuhrer J (2016) *topGO: Enrichment Analysis for Gene Ontology*.

Anderson DV, Clayton D (1959) Plankton in Lake Ontario. *Div. Res., Dept. Lands and Forests, Maple, Ontario*.

Asano Y, Liberman UA, Edelman IS (1976) Thyroid thermogenesis. Relationships between Na+-dependent respiration and Na+ + K+-adenosine triphosphatase activity in rat skeletal muscle. *Journal of Clinical Investigation*, **57**, 368–379.

Beaumont MA (2005) Adaptation and speciation: what can F ST tell us? *Trends in Ecology & Evolution*, **20**, 435–440.

Bellard C, Bertelsmeier C, Leadley P, Thuiller W, Courchamp F (2012) Impacts of climate change on the future of biodiversity. *Ecology Letters*, **15**, 365–377.

Beyenbach KW, Wieczorek H (2006) The V-type H+ ATPase: molecular structure and function, physiological roles and regulation. *Journal of Experimental Biology*, **209**, 577–589.

Bintanja R, Selten FM (2014) Future increases in Arctic precipitation linked to local evaporation and sea-ice retreat. *Nature*, **509**, 479–482.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, btu170.

Brooks SJ, Mills CL (2006) Gill Na+, K+-ATPase in a series of hyper-regulating gammarid amphipods. Enzyme characterisation and the effects of salinity acclimation. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **144**, 24–32.

Bystriansky JS, Richards JG, Schulte PM, Ballantyne JS (2006) Reciprocal expression of gill Na+/K+-ATPaseα-subunit isoforms α1a and α1b during seawater acclimation of three

salmonid fishes that vary in their salinity tolerance. *Journal of Experimental Biology*, **209**, 1848–1858.

Chambrey R, Kurth I, Peti-Peterdi J *et al.* (2013) Renal intercalated cells are rather energized by a proton than a sodium pump. *Proceedings of the National Academy of Sciences*, **110**, 7928–7933.

Cheng L, Jin X-K, Li W-W *et al.* (2013) Fatty Acid Binding Proteins FABP9 and FABP10 Participate in Antibacterial Responses in Chinese Mitten Crab, Eriocheir sinensis. *PLOS ONE*, **8**, e54053.

Cheung WWL, Lam VWY, Sarmiento JL *et al.* (2009) Projecting global marine biodiversity impacts under climate change scenarios. *Fish and Fisheries*, **10**, 235–251.

Day JP, Wan S, Allan AK *et al.* (2008) Identification of two partners from the bacterial Kef exchanger family for the apical plasma membrane V-ATPase of Metazoa. *Journal of Cell Science*, **121**, 2612–2619.

Di Castri F (1989) History of biological invasions with special emphasis on the Old World. *Biological invasions: a global perspective*, 1–30.

Durack PJ, Wijffels SE, Matear RJ (2012) Ocean Salinities Reveal Strong Global Water Cycle Intensification During 1950 to 2000. *Science*, **336**, 455–458.

Ehrenfeld J, Klein U (1997) The key role of the H+ V-ATPase in acid-base balance and Na+ transport processes in frog skin. *Journal of Experimental Biology*, **200**, 247–256.

Ehrenfeld J, Lacoste I, Harvey BJ (1989) The key role of the mitochondria-rich cell in Na+ and H+ transport across the frog skin epithelium. *Pflügers Archiv*, **414**, 59–67.

Esteves A, Ehrlich R (2006) Invertebrate intracellular fatty acid binding proteins. *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, **142**, 262–274.

Faber DJ, Jermolajev EG (1966) A new copepod genus in the plankton of the Great Lakes. *Limnology and Oceanography*, **11**, 301–303.

Futschik A, Schlötterer C (2010) The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. *Genetics*, **186**, 207–218.

Gannon JE (1974) The crustacean zooplankton of Green Bay, Lake Michigan. In: *IN: PROCEEDINGS, SEVENTEENTH CONFERENCE ON GREAT LAKES RESEARCH, PART 1*

Gerber L, Lee CE, Grousset E *et al.* (2016) The Legs Have It: In Situ Expression of Ion Transporters V-Type H + -ATPase and Na + /K + -ATPase in the Osmoregulatory Leg Organs of the Invading Copepod Eurytemora affinis. *Physiological and Biochemical Zoology*, **89**, 233–250.

Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics*, **121**, 129–138.

Haguenauer A, Raimbault S, Masscheleyn S *et al.* (2005) A New Renal Mitochondrial Carrier, KMCP1, Is Up-regulated during Tubular Cell Regeneration and Induction of Antioxidant Enzymes. *Journal of Biological Chemistry*, **280**, 22036–22043.

Hegerl GC, Black E, Allan RP *et al.* (2015) Challenges in quantifying changes in the global water cycle. *Bulletin of the American Meteorological Society*, **96**, 1097–1115.

Helm KP, Bindoff NL, Church JA (2010) Changes in the global hydrological-cycle inferred from ocean salinity. *Geophysical Research Letters*, **37**, L18701.

Hohenlohe P a, Bassham S, Etter PD *et al.* (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS genetics*, **6**, e1000862–e1000862.

Hufbauer RA, Facon B, Ravigne V *et al.* (2012) Anthropogenically induced adaptation to invade (AIAI): contemporary adaptation to human-altered habitats within the native range can promote invasions. *Evolutionary Applications*, **5**, 89–101.

Hutchinson GE (1957) A treatise on Limnology.

Ismail-Beigi F, Edelman IS (1970) Mechanism of Thyroid Calorigenesis: Role of Active Sodium Transport. *Proceedings of the National Academy of Sciences*, **67**, 1071–1078.

Jensen L, Willumsen N, Larsen E (2002) Proton pump activity is required for active uptake of chloride in isolated amphibian skin exposed to freshwater. *Journal of Comparative Physiology B*, **172**, 503–511.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Khlebovich VV, Abramova EN (2000) Some problems of crustacean taxonomy related to the phenomenon of Horohalinicum. *Hydrobiologia*, **417**, 109–113.

Kidder GW, Petersen CW, Preston RL (2006) Energetics of osmoregulation: I. oxygen consumption by Fundulus heteroclitus. *Journal of Experimental Zoology Part A: Comparative Experimental Biology*, **305A**, 309–317.

Kirschner LB (1993) The energetics of osmotic regulation in ureotelic and hypoosmotic fishes. *Journal of Experimental Zoology*, **267**, 19–26.

Kirschner LB (2004) The mechanism of sodium chloride uptake in hyperregulating aquatic animals. *Journal of Experimental Biology*, **207**, 1439–1452.

Kofler R, Pandey RV, Schlötterer C (2011) PoPoolation2: Identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*, **27**, 3435–3436.

Kozak GM, Brennan RS, Berdan EL, Fuller RC, Whitehead A (2014) Functional and population genomic divergence within and between two species of killifish adapted to different osmotic niches. *Evolution*, **68**, 63–80.

Larsen EH, Christoffersen BC, Jensen LJ, Sørensen JB, Willumsen NJ (1996) Role of mitochondria-rich cells in epithelial chloride uptake. *Experimental physiology*, **81**, 525.

Lee CE (1999) Rapid and repeated invasions of fresh water by the copepod Eurytemora affinis. *Evolution*, **53**, 1423–1434.

Lee CE (2000) Global Phylogeography of a Cryptic Copepod Species Complex and Reproductive Isolation between Genetically Proximate "Populations." *Evolution*, **54**, 2014–2027.

Lee CE (2011) Evolution of Invasive Populations. In: *Encyclopedia of Biological Invasions*, pp. 215–22.

Lee CE (2016) Evolutionary mechanisms of habitat invasions, using the copepod Eurytemora affinis as a model system. *Evolutionary Applications*, **9**, 248–270.

Lee CE, Bell M a (1999) Causes and consequences of recent freshwater invasions by saltwater animals. *Trends in Ecology & Evolution*, **14**, 284–8.

Lee CE, Gelembiuk GW (2008) Evolutionary origins of invasive populations. *Evolutionary Applications*, **1**, 427–448.

Lee CE, Kiergaard M, Gelembiuk GW, Eads BD, Posavi M (2011) Pumping Ions: Rapid Parallel Evolution of Ionic Regulation Following Habitat Invasions. *Evolution*, **65**, 2229–2244.

Lee CE, Posavi M, Charmantier G (2012) Rapid evolution of body fluid regulation following independent invasions into freshwater habitats. *Journal of Evolutionary Biology*, **25**, 625–633.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, **27**, 2987–2993.

Li H (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, **25**, 2078–2079.

Little C (1983) *The colonisation of land: origins and adaptations of terrestrial animals*. Cambridge University Press.

Loder JW, van der Baaren A, Yashayaev I (2015) Climate comparisons and change projections for the Northwest Atlantic from six CMIP5 models. *Atmosphere-Ocean*, **53**, 529–555.

Lucu Č, Towle DW (2003) Na++K+-ATPase in gills of aquatic crustacea. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, **135**, 195–214.

Lunter G, Goodson M (2011) Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, **21**, 936–939.

May GE, Gelembiuk GW, Panov VE, Orlova MI, Lee CE (2006a) Molecular ecology of zebra mussel invasions. *Molecular Ecology*, **15**, 1021–1031.

May GE, Gelembiuk GW, Panov VE, Orlova MI, Lee CE (2006b) Molecular ecology of zebra mussel invasions. *Molecular ecology*, **15**, 1021–31.

McCormick SD, Regish AM, Christensen AK (2009) Distinct freshwater and seawater isoforms of Na+/K+-ATPase in gill chloride cells of Atlantic salmon. *Journal of Experimental Biology*, **212**, 3994–4001.

McKenna A, Hanna M, Banks E *et al.* (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*, **20**, 1297–1303.

Meier HEM (2006) Baltic Sea climate in the late twenty-first century: a dynamical downscaling approach using two global models and two emission scenarios. *Climate Dynamics*, **27**, 39–68.

Miller MF, Labandeira CC (2002) Slow crawl across the salinity divide: delayed colonization of freshwater ecosystems by invertebrates. *GSA today*, **12**, 4–10.

Morgan C a., Cordell JR, Simenstad C a. (1997) Sink or swim? Copepod population maintenance in the Columbia River estuarine turbidity-maxima region. *Marine Biology*, **129**, 309–317.

Morris S (2001) Neuroendocrine regulation of osmoregulation and the evolution of air-breathing in decapod crustaceans. *Journal of Experimental Biology*, **204**, 979–989.

Munk W (2003) Ocean freshening, sea level rising. *Science*, **300**, 2041–2043.

Nei M, Li W-H (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proceedings of the National Academy of Sciences*, **76**, 5269–5273.

Nielsen R (2005) Molecular signatures of natural selection. *Annual review of genetics*, **39**, 197–218.

Nilsen TO, Ebbesson LO, Madsen SS *et al.* (2007) Differential expression of gill Na+, K+-ATPaseα-and β-subunits, Na+, K+, 2Cl-cotransporter and CFTR anion channel in

juvenile anadromous and landlocked Atlantic salmon Salmo salar. *Journal of Experimental Biology*, **210**, 2885–2896.

Palmieri F (2004) The mitochondrial transporter family (SLC25): physiological and pathological implications. *Pflügers Archiv*, **447**, 689–709.

Peitsch A, Köpcke B, Bernát N (2000) Long-term investigation of the distribution of Eurytemora affinis (Calanoida; Copepoda) in the Elbe Estuary. *Limnologica - Ecology and Management of Inland Waters*, **30**, 175–182.

Pimm SL, Jenkins CN, Abell R *et al.* (2014) The biodiversity of species and their rates of extinction, distribution, and protection. *Science*, **344**, 1246752.

Posavi M (2015) Evolutionary Mechanisms of Rapid Adaptation During Freshwater Invasions by the Saline Copepod Eurytemora affinis. University of Wisconsin - Madison, Madison, WI.

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**, 841–842.

Rabe B, Karcher M, Schauer U *et al.* (2011) An assessment of Arctic Ocean freshwater content changes from the 1990s to the 2006–2008 period. *Deep Sea Research Part I: Oceanographic Research Papers*, **58**, 173–185.

Rheault MR, Okech BA, Keen SBW *et al.* (2007) Molecular cloning, phylogeny and localization of AgNHA1: the first Na+/H+ antiporter (NHA) from a metazoan, Anopheles gambiae. *Journal of Experimental Biology*, **210**, 3848–3861.

Ricciardi A, MacIsaac HJ (2000) Recent mass invasion of the North American Great Lakes by Ponto–Caspian species. *Trends in Ecology & Evolution*, **15**, 62–65.

Schlötterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. *Nature Publishing Group*, **15**, 749–763.

Smit A, Hubley R, Green P (2013) *RepeatMasker Open-4.0*.

Thebaud C, Finzi AC, Affre L, Debussche M, Escarre J (1996) Assessing why two introduced Conyza differ in their ability to invade Mediterranean old fields. *Ecology*, **77**, 791–804.

Thøgersen T, Hoff A, Frost HS (2015) Fisheries management responses to climate change in the Baltic Sea. *Climate Risk Management*, **10**, 51–62.

Thomas CD, Cameron A, Green RE *et al.* (2004) Extinction risk from climate change. *Nature*, **427**, 145–148.

Towle DW, Weihrauch D (2001) Osmoregulation by gills of euryhaline crabs: molecular analysis of transporters. *American Zoologist*, **41**, 770–780.

Tsai J-R, Lin H-C (2007) V-type H+-ATPase and Na+,K+-ATPase in the gills of 13 euryhaline crabs during salinity acclimation. *Journal of Experimental Biology*, **210**, 620–627.

Turelli M, Barton NH (2004) Polygenic Variation Maintained by Balancing Selection: Pleiotropy, Sex-Dependent Allelic Effects and G × E Interactions. *Genetics*, **166**, 1053–1079.

Urban MC (2015) Accelerating extinction risk from climate change. *Science*, **348**, 571–573.

Waddington C (1965) Introduction to the symposium. In: *The genetics of colonizing species* (eds Baker H, Stebbins G), pp. 1–6. Academic Press, New York.

Weihrauch D, McNamara JC, Towle DW, Onken H (2004) Ion-motive ATPases and active, transbranchial NaCl uptake in the red freshwater crab, Dilocarcinus pagei (Decapoda, Trichodactylidae). *Journal of Experimental Biology*, **207**, 4623–4631.

Wieczorek H, Brown D, Grinstein S, Ehrenfeld J, Harvey WR (1999) Animal plasma membrane energization by proton-motive V-ATPases. *BioEssays*, **21**, 637–648.

Wieczorek H, Putzenlechner M, Zeiske W, Klein U (1991) A vacuolar-type proton pump energizes K+/H+ antiport in an animal plasma membrane. *Journal of Biological Chemistry*, **266**, 15340–15347.

Winkler G, Dodson JJ, Lee CE (2008) Heterogeneity within the native range: population genetic analyses of sympatric invasive and noninvasive clades of the freshwater invading copepod Eurytemora affinis. *Molecular ecology*, **17**, 415–30.

Wit P, Pespeni MH, Ladner JT *et al.* (2012) The simple fool's guide to population genomics via RNA-Seq: an introduction to high-throughput sequencing data analysis. *Molecular ecology resources*, **12**, 1058–1067.

Wright S (1949) The genetical structure of populations. *Annals of eugenics*, **15**, 323–354.

Xiang MA, Linser PJ, Price DA, Harvey WR (2012) Localization of two Na+-or K+-H+ antiporters, AgNHA1 and AgNHA2, in Anopheles gambiae larval Malpighian tubules and the functional expression of AgNHA2 in yeast. *Journal of insect physiology*, **58**, 570–579.

Yassin A, Vincent D, Bastide H *et al.* (2016) Recurrent Specialization on a Toxic Fruit in an Island Drosophila Population. *Proceedings of the National Academy of Sciences of the United States of America, USA*.

Yi X, Liang Y, Huerta-Sanchez E *et al.* (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science*, **329**, 75–78.

Tables and Figures

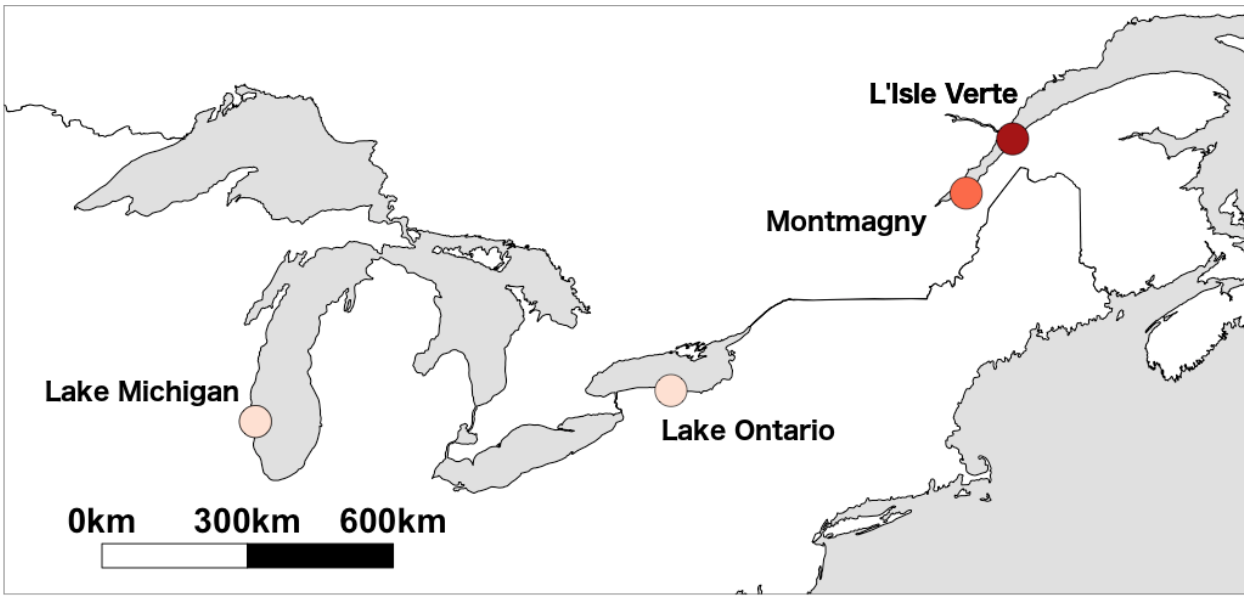

Figure 1 - Population sampling of *Eurytemora affinis* for this study – Samples include four population of *E. affinis*: two from the saline St. Lawrence estuary (L'Isle Verte, Quebec, Canada and Montmagny, Quebec, Canada), and two from the freshwater Great Lakes (Lake Ontario, NY, USA and Lake Michigan, WI, USA). The two light red dots represent freshwater populations the red and dark red dots represent the intermediate salinity of Montmagny and more saline water in L'isle Verte, respectively.

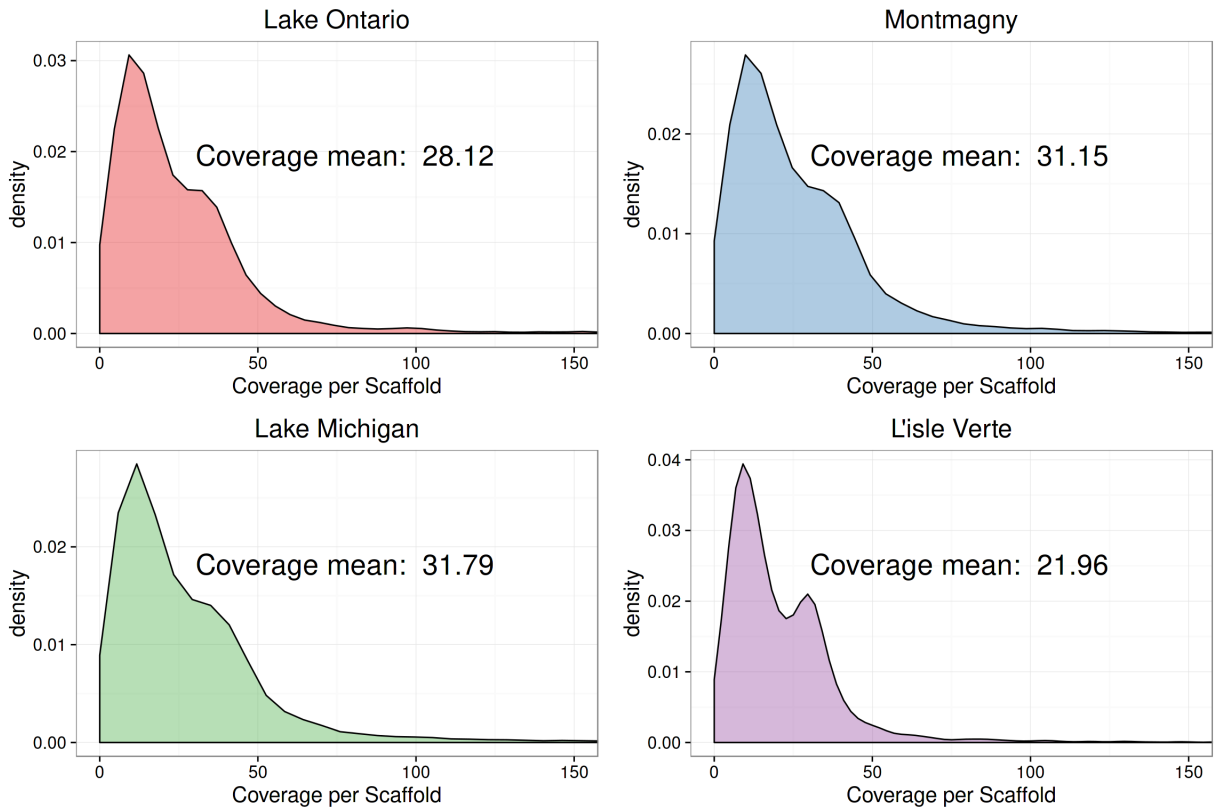

Figure 2 – Mean Coverage per Scaffold in all Sequenced Populations – Sequencing coverage across the reference *E. affinis* genome for all four populations. Lake Ontario, Montmagny, and Lake Michigan populations were all sequenced to a similar depth of coverage. The population from L'Isle Verte was sequenced to a lower depth of coverage, owing mostly to the lower initial sequencing coverage of that population (see Table 2).

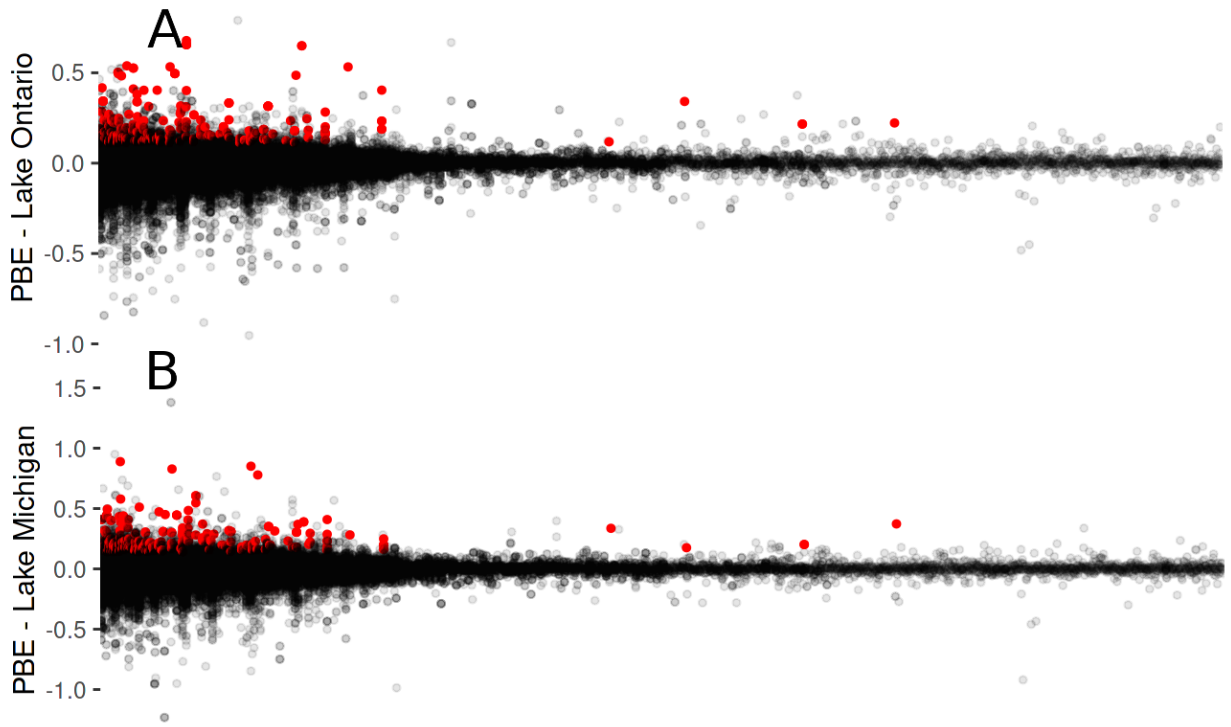

Figure 3 - Genomic regions of high differentiation between freshwater and saline populations in the St. Lawrence drainage – Each point represents the mean PBE value within a 5kb window. All scaffolds in the reference *E. affinis* genome assembly are positioned along the x-axis, in order from longest scaffold to shortest. The reduction in spread in the data is due to the decreasing size of scaffolds moving from left to right. Red dots are genomic regions which are highly differentiated (99.5^{th} quantile) in both comparisons between freshwater populations and saline ancestors (a) Saline St. Lawrence populations (Montmagny, Quebec and L'isle Verte, Quebec) versus freshwater Lake Ontario. (b) Saline St. Lawrence populations versus freshwater Lake Michigan.

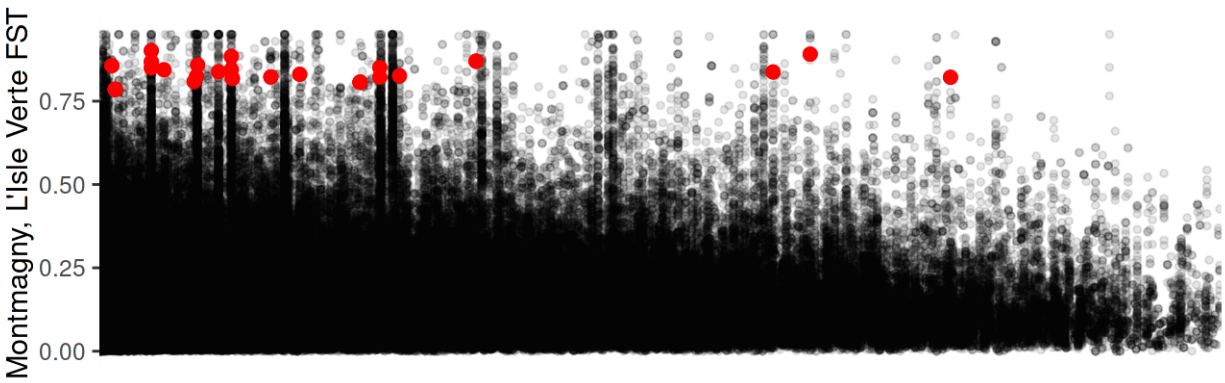

Figure 4 - Genomic regions of high differentiation between saline populations in the St. Lawrence estuary – Each point represents the mean F_{ST} value within a 5kb window. For the sake of clarity, only the longest 1000 scaffolds (out of 6899) are shown. Scaffolds are positioned along the x-axis in order from longest scaffold to shortest. The reduction in spread in the data are due to the decreasing size of scaffolds moving from left to right. Red dots are genomic regions which are highly differentiated (99.5^{th} quantile) in both comparisons between freshwater populations and saline ancestors and also highly differentiated between the two saline populations.

Table 1 – Sample Metadata

Location	Water Type	Salinity (PSU)	Temp (°C)	Latitude	Longitude
L'Isle Verte	Salt	13.7	5.9	48.002	-69.423
Montmagny	Salt	5	6.3	46.99	-70.55
Lake Ontario	Fresh	0	29.7	43.307	-77.706
Lake Michigan	Fresh	0.1	15.7	43.051	-87.882

Table 2 – Sequencing and Population Genomics Summary Statistics– Nucleotide diversity (π) was measured in 5kb windows every 1kb across the *E. affinis* reference genome as described in Nei & Li (1979) genome-wide.

Location	**Raw Reads**	**Filtered Aligned Reads**	**Mean Coverage**	**Median Nucleotide Diversity (π)**
L'Isle Verte, Quebec	308,081,218	165,059,946	21.96x	$6.25x10^{-3}$
Montmagny, Quebec	383,438,428	226,547,166	31.15x	$4.03x10^{-3}$
Lake Ontario, NY	353,435,194	207,149,336	28.12x	$4.51x10^{-3}$
Lake Michigan, WI	385,263,696	229,353,274	31.79x	$4.3x10^{-3}$

Table 3 – Median F_{ST} values between populations - F_{ST} measurements were calculated using PoPoolation2 (Kofler et al. 2009) in 5kb windows every 1kb.

	L'Isle Verte	Montmagny	Lake Ontario
Montmagny	0.1126725		
Lake Ontario	0.1115333	0.01969097	
Lake Michigan	0.1143886	0.0222202	0.02299863

Table 4 – Annotated Genes and Gene Models in Regions of High Differentiation between Populations– List of gene models in regions of high differentiation that were shared between both freshwater populations. Of the 109 gene models that were in shared regions of high differentiation, the 32 genes in this table shared protein sequence homology with proteins in the UniProt database and could be functionally annotated with GO terms. NHA genes (EAFF008240-RA, EAFF008241-RA, and EAFF008244-RA) did share homology with proteins in UniProt, but had been previously manually annotated in the E. affinis reference genome. Gene models in bold were found in both highly differentiated regions in freshwater PBE comparisons and in highly differentiated regions between the two ancestral saline populations.

Gene Model	UniProt ID	% Identity	E-value	Gene	Description
EAFF001602-RA	P56941	30.43	2E-16	NPC1	Niemann-Pick C1 protein
EAFF001761-RA	Q55CN6	70.97	0.0000005	pks3	Probable polyketide synthase 3
EAFF003425-RA*	**Q6GQ22**	**47.27**	**1E-28**	**KMCP1**	**Kidney mitochondrial carrier protein 1**
EAFF003625-RA	Q24048	41.04	7E-25	nrv2	Sodium/potassium-transporting ATPase subunit beta-2
EAFF003666-RA	Q7SYD5	44.26	2E-10	sec31a	Protein transport protein Sec31A
EAFF006258-RA	Q5R1W5	49.38	2E-15	SRSF2	Serine/arginine-rich splicing factor 2
EAFF006839-RA	P25822	41.33	6E-11	pum	Maternal protein pumilio
EAFF008240-RA	-	-	-	NHA	Na/H antiporter (NHA) 1_6
EAFF008241-RA	-	-	-	NHA	Na/H antiporter (NHA) 1_5
EAFF008244-RA	-	-	-	NHA	Na/H antiporter (NHA) 1_3
EAFF008518-RA	P15143	67.98	4E-75	POU2F1	POU domain, class 2, transcription factor 1
EAFF009104-RA*	**Q9W4P5**	**83.91**	**9E-47**	**VhaAC39-1**	**V-type proton ATPase subunit d**
EAFF009308-RA*	**Q0Z7S8**	**44.74**	**0.000001**	**FABP9**	**Fatty acid-binding protein 9**
EAFF009735-RA	Q9P2K8	44.87	3E-31	EIF2AK4	eIF-2-alpha kinase GCN2
EAFF012022-RA	Q7ZY29	60.77	7E-48	esrp1	Epithelial splicing regulatory protein 1
EAFF012630-RA	O55196	33.12	4E-11	Enam	Enamelin
EAFF013773-RA*	**A0A0K2T212**	**72.55**	**2E-19**	**SVEP1**	**Sushi, von Willebrand factor type A, EGF and pentraxin domaincontaining protein 1**
EAFF015738-RA	Q5R5U3	51.43	9E-120	ZNF271	Zinc finger protein 271
EAFF016345-RA	A0A0K2TFG6	58.95	6E-25	-	Protein disulfide-isomerase
EAFF016474-RA	Q05024	51.67	1E-11	TRI1	Protein TRI1
EAFF017103-RA	Q98930	36.96	1E-22	SORL1	Sortilin-related receptor
EAFF018184-RA	R7VRW4	18.87	8E-10	-	Putative ubiquitin thioesterase L96
EAFF021029-RA	Q95KE5	41.98	1E-11	MRPL43	39S ribosomal protein L43, mitochondrial
EAFF022430-RA	P12545	50.67	3E-15	PLG	Plasminogen
EAFF024882-RA	Q64380	54.02	3E-54	Sardh	Sarcosine dehydrogenase, mitochondrial
EAFF026040-RA	Q8TD57	91.11	0.0000002	D-H3	Dynein heavy chain 3, axonemal

EAFF026626-RA	A0A023F5F9	75.61	1E-12	-	Putative signal transducing adaptor protein stam/stam2
EAFF026740-RA*	**P55011**	**39.31**	**0**	**NKCC1**	**Solute carrier family 12 member 2**
EAFF026928-RA	Q91XQ0	34.21	7E-24	Dnah8	Dynein heavy chain 8, axonemal
EAFF027816-RA	P42519	34.62	0.0000003	Sardh	Protein Star
EAFF029699-RA	A0A067RHX8	49.37	3E-16	-	Carbohydrate sulfotransferase 5
EAFF029735-RA	Q09332	48.15	2E-89	Ugt	UDP-glucose:glycoprotein glucosyltransferase

Table 5 – Molecular Function GO Terms Enriched in Highly-Differentiated Genes – Gene models overlapping highly differentiated PBE regions were annotated with GO terms from homologs in the UniProt database, and the enrichment of GO terms was tested with a Fisher's exact test without correction for multiple testing, since the number of genes was so low.

GO.ID	Term	Annotated	Significant	Expected	P-value
GO:0015078	hydrogen ion transmembrane transporter activity	64	3	0.14	0.00036
GO:0015299	solute:proton antiporter activity	27	2	0.06	0.00157
GO:0015298	solute:cation antiporter activity	44	2	0.1	0.00413
GO:0003980	UDP-glucose:glycoprotein glucosyltransferase activity	3	1	0.01	0.00656
GO:0004694	eukaryotic translation initiation factor 2alpha kinase activity	3	1	0.01	0.00656
GO:0008480	sarcosine dehydrogenase activity	3	1	0.01	0.00656
GO:0008511	sodium:potassium:chloride symporter activity	3	1	0.01	0.00656
GO:0046997	oxidoreductase activity	4	1	0.01	0.00874
GO:0015077	monovalent inorganic cation transmembrane activity	385	4	0.84	0.00927
GO:0022804	active transmembrane transporter activity	390	4	0.85	0.00969
GO:0030021	extracellular matrix structural constituent	7	1	0.02	0.01525
GO:0030345	structural constituent of tooth enamel	7	1	0.02	0.01525
GO:0003777	microtubule motor activity	89	2	0.2	0.01609
GO:0015297	antiporter activity	89	2	0.2	0.01609
GO:0000900	translation repressor activity, nucleic acid binding	8	1	0.02	0.0174
GO:0015291	secondary active transmembrane transporter	260	3	0.57	0.01865
GO:0005542	folic acid binding	9	1	0.02	0.01956
GO:0030371	translation repressor activity	10	1	0.02	0.02171
GO:0015377	cation:chloride symporter activity	11	1	0.02	0.02386
GO:0090079	translation regulator activity, nucleic acid binding	11	1	0.02	0.02386
GO:0003729	mRNA binding	111	2	0.24	0.02436
GO:0036442	hydrogen-exporting ATPase activity	14	1	0.03	0.03027
GO:0046961	proton-transporting ATPase activity, rotational mechanism	14	1	0.03	0.03027
GO:0022890	inorganic cation transmembrane transport	561	4	1.23	0.03235
GO:0044769	ATPase activity, coupled to transmembrane movement of ions	15	1	0.03	0.03239
GO:0035251	UDP-glucosyltransferase activity	18	1	0.04	0.03875
GO:0045182	translation regulator activity	18	1	0.04	0.03875
GO:0003735	structural constituent of ribosome	157	2	0.34	0.04597
GO:0022892	substrate-specific transporter activity	1229	6	2.69	0.04602
GO:0051082	unfolded protein binding	22	1	0.05	0.04716