ASSEMBLY OF MICROBIAL COMMUNITIES ASSOCIATED WITH THE DEVELOPING ZEBRAFISH INTESTINE

by

ADAM RAYMOND BURNS

A DISSERTATION

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

September 2016

DISSERTATION APPROVAL PAGE

Student: Adam Raymond Burns

Title: Assembly of Microbial Communities Associated with the Developing Zebrafish Intestine

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Biology by:

Jessica L. Green	Chairperson
Brendan J. M. Bohannan	Advisor
Karen Guillemin	Advisor
William A. Cresko	Core Member
Raghuveer Parthasarathy	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

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

Degree awarded September 2016

DISSERTATION ABSTRACT

Adam Raymond Burns

Doctor of Philosophy

Department of Biology

September 2016

Title: Assembly of Microbial Communities Associated with the Developing Zebrafish Intestine

The communities of microorganisms associated with humans and other animals are characterized by a large degree of diversity and unexplained variation across individual hosts. While efforts to explain this variation in host-associated systems have focused heavily on the effects of host selection, community assembly theory emphasizes the role of dispersal and stochastic demographic processes, otherwise known as ecological drift. In this dissertation, I characterize the communities of microorganisms associated with the zebrafish, *Danio rerio*, intestine, and assess the importance of microbial dispersal and drift to their assembly. First, I describe changes in the composition and diversity of the zebrafish intestinal microbiome over zebrafish development and show that while host development is a major driver of community composition over time, there remains a large amount of unexplained variation among similar hosts of the same age. I go on to show that random dispersal and ecological drift alone in the absence of host selection are sufficient to explain a substantial amount of this variation, but the ability of these processes to predict the distribution of microorganisms across hosts decreases over host development. Finally, I present an experimental test of dispersal in host-associated systems, and show that not only does dispersal among

individual zebrafish hosts have a large impact on the composition and diversity of associated microbial communities, but it can also overwhelm the effects of important host factors, such as the innate immune system. As a whole, this work demonstrates that the composition and diversity of microbial communities associated with animal hosts are not solely the result of selection by the host environment, but rather dispersal and stochastic processes have important and often overwhelming effects on their assembly. To fully understand the assembly of host-microbe systems, we must broaden our focus to include scales beyond that of an individual host and their associated microorganisms.

This dissertation includes previously published and unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Adam Raymond Burns

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
Fullerton College, Fullerton, California

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2016, University of Oregon
Bachelor of Science, Biology, 2010, University of Oregon
Associate of Science, Biology, 2004, Fullerton College

AREAS OF SPECIAL INTEREST:

Community Ecology
Microbial Ecology

PROFESSIONAL EXPERIENCE:

Graduate Research and Teaching Assistant, University of Oregon, 2010-2016

Undergraduate Research Assistant, University of Oregon, 2008-2010

GRANTS, AWARDS, AND HONORS:

Graduate Teaching Fellowship, University of Oregon, 2010-2016

PUBLICATIONS:

Burns AR, Stephens WZ, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJM. (2016). Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *The ISME Journal* **10**: 655-664.

Stephens WZ, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K, Bohannan BJM. (2016). The composition of the zebrafish intestinal microbial community varies across development. *The ISME Journal* **10**: 644-654.

Rolig AS, Parthasarathy R, Burns AR, Bohannan BJM, Guillemin K. (2015). Individual members of the microbiota disproportionately modulate host innate immune responses. *Cell Host & Microbe* **18**: 613-620.

Stephens WZ, Wiles TJ, Martinez ES, Jemielita M, Burns AR, Parthasarathy R, Bohannan BJM, Guillemin K. (2015). Identification of population bottlenecks and colonization factors during assembly of bacterial communities within the zebrafish intestine. *mBio* **6**: e01163-15.

Wong S, Stephens WZ, Burns AR, Stagaman K, David LA, Bohannan BJM, Guillemin K, Rawls JF. (2015). Ontogenetic differences in dietary fat influence microbiota assembly in the zebrafish gut. *mBio* **6**: e00687-15.

Jemielita M, Taromina M, Burns AR, Hampton J, Rolig AS, Guillemin K, Parthasarathy R. (2014). Spatial and temporal features of the growth of a bacterial species colonizing the zebrafish gut. *mBio* **5**: e01751-14.

Taromina M, Jemielita M, Stephens WZ, Burns AR, Hampton J, Guillemin K, Parthasarathy R. (2013). Imaging colonization dynamics and rheological properties of a host and its developing microbiome by light sheet microscopy. *Biophysical Journal* **104**: 639a.

Taormina M, Jemielita M, Stephens WZ, Burns AR, Troll JV, Parthasarathy R, Guillemin K. (2012). Investigating bacterial-animal symbioses with light sheet microscopy. *The Biological Bulletin* **223**: 7-20.

ACKNOWLEDGMENTS

It is perhaps a bit cliché to say that a body of work such as this dissertation would not have been possible without the help of supportive mentors and colleagues, but I certainly would not have persisted as a graduate student without the support of my advisors, Brendan Bohannan and Karen Guillemin. In addition to being great scientists, Brendan and Karen were incredible leaders and mentors and provided excellent examples of how to balance and live a life as academics. In that capacity, they both constantly surprised me by their continual support, encouragement, and confidence in my abilities. I am also grateful to them for granting me the independence to pursue ideas and questions I was passionate about, even when I sensed that they did not always believe those ideas to be good ones, at least at the time. I hope that this work serves to justify their trust in me.

I also thank the members of my dissertation advisory committee, Jessica Green, Raghuveer Parthasarathy, and William Cresko, not only for their support and genuine enthusiasm for my work, but also for continually challenging me to become a more careful and rigorous scientist. In addition to serving as my committee chair, Jessica also took a chance on me as an undergraduate researcher to work in her lab. This experience provided me with my first hands on exposure to microbial community ecology and encouraged me to pursue graduate research.

I was fortunate enough to be able to work in a highly interdisciplinary environment with a number of amazing colleagues and collaborators. Due to our shared interests, I worked frequently with W. Zac Stephens and Keaton Stagaman on a number of projects forming large elements of this dissertation, and the collaborations with the two of them was always smooth and enjoyable. I thank them both for their contributions to

this work. In addition I am very grateful to Zac who, as a more senior graduate student at the time, acted as an additional mentor and patiently taught me many of the skills I needed to succeed in our field. I also thank Meghna Agarwal who also contributed significantly to helping me perform my dissertation research. Though technically she was an undergraduate at the time, working with her often felt more like working with another graduate student due to her abilities as a researcher. Outside of my dissertation research, I was able to engage in a great diversity of projects and work closely with some great collaborators, including, in no particular order, Sandi Wong, Matt Jemielita, Robert Steury, Annah Rolig, Travis Wiles, Tiffani Jones, Sophie Sichel, John Rawls, and Michael Kent.

Finally, I cannot thank Rachel, Ada, and Roland enough for the sacrifices they've made, whether they knew it or not, in supporting me in this endeavor. They have served as my primary source of motivation and have never flinched at any of the challenges a life in graduate school has brought. I look forward to continuing forward with them.

For my father, Stephen Burns, who didn't get to see me finish, but who inspired me to start, and for my children, Ada and Roland Burns, who didn't get to see me start, but inspired me to finish.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Overview of the vertebrate microbiota

The vast majority of life on earth is microbial, and animals and other multi-cellular organisms have always existed in the context of microbial life. It is therefore unsurprising that some of the first microorganisms to be observed were ones associated with the human body. Microorganisms have shaped our evolution and biology throughout our entire history, both as antagonistic pathogens, but also, as is being increasingly understood, as commensals and even mutualists that form complex communities both on and within their hosts. These communities are important mediators of animal health and development. Unfortunately, the study of naturally occurring microbial communities, including those colonizing animal hosts, is hindered by both technical and conceptual limitations. Their small size, abundance, and recalcitrance to culturing means we often must rely on technological advances, such as high-throughput DNA sequencing, to be able to observe and characterize them. However, even once we are able to do so, the vast diversity, both in numbers and types of species but also in terms of evolutionary history and functional potential, of microbial communities is difficult to make sense of without a clearly conceptualized and tested understanding of the processes responsible for generating and maintaining the diversity of these systems.

This last point has become increasingly evident through the exploration of microbial communities associated with vertebrate animals, including humans, mice, and fish, which is often referred to as the host's "microbiome" or "microbiota". Over the past

decade, many of the historical technical limitations of studying microbial communities have been overcome through high-throughput DNA sequencing of genetic "barcodes" used to identify microbial taxa. This has led to large scale sampling of the vertebrate microbiome, including the aptly named Human Microbiome Project (The Human Microbiome Project Consortium *et al.*, 2012), with the goal of fully mapping the composition and function of host-associated communities. These projects quickly revealed what earlier studies had previously hinted at: the vertebrate microbiome is very diverse, both in terms of species composition as well as functional potential, and there is a large amount of variation in the composition of the microbiome even among genetically similar individuals within the same population. Much of this diversity occurs within fine phylogenetic scales, suggesting there may be a large number of similar microbial taxa in these systems. Furthermore, despite continual efforts through additional large scale sampling projects, few consistent host factors have emerged that can explain the variation in microbial community composition among hosts. These trends in diversity both within and among hosts are common for natural human and animal populations, but also even for model animals raised under controlled laboratory conditions, suggesting that it is a fundamental characteristic of these systems.

On its own, this diversity is incredibly interesting and is a part of what makes ecological systems, including microbial ones, so compelling to study and experience. The vertebrate microbiota is certainly not unique in this regard. However, what is so striking in animal associated communities is that it seems at odds with the importance of these communities to the health and development of their hosts. If variations in the composition and dynamics of an individual's microbiota can lead to abnormal development, disease,

and even death of the host, then one would expect the microbiota to be tightly controlled and regulated by the host. However, the variable nature of the microbiota and the lack of identifiable host factors with a strong effect on its composition suggest this may not be the case, at least not to the extent that we might expect. These discrepancies between the expected behaviors of the systems with observed patterns do not disprove the hypothesis that hosts strongly select their microbiota, but they do suggest that we are not accounting for some key processes, and that our conventional conceptual synthesis of the assembly of host-associated microbial communities is currently lacking.

Ecological theory in host-microbe systems

The diversity of the vertebrate microbiome is not unique to that system, but is a common theme across many ecological communities. Community ecology is a science largely devoted to understanding the processes that generate and maintain biological diversity at the level of communities of multiple, potentially interacting species. It is therefore perfectly suited to study the assembly of host-associated communities (Costello *et al.*, 2012). Ecology provides a suite of analytical tools for dealing with complex, highly multidimensional, biological data, and more importantly, provides theoretical frameworks to both generate and test predictions about the processes driving community assembly. Many conceptual models exist that categorize these processes in different ways, however they typically differentiate the effects of ecological selection, in which the abundance and distribution of species in a community is determined by non-symmetrical interactions among species and between species and their environment, dispersal, or the movement of organisms within and among communities, and random demographic

processes, which are referred to as ecological drift (Vellend, 2010). The study of host-associated microbial communities, particularly those associated with vertebrates, has focused heavily on selective processes, while largely ignoring the effects of dispersal and drift. This is in contrast with the field of community ecology as a whole, in which dispersal and chance have historically formed key components of many important theories of assembly (Gleason, 1926; Hubbell, 2001; Leibold *et al.*, 2004; MacArthur and Wilson, 1967). This includes neutral theory, which focuses on the sole effects of random dispersal and drift, and metacommunity theory, which focuses on how dispersal interacts and mediates other factors. These theories have had large impacts on our understanding of how natural communities assemble and are formed.

This raises the question as to whether these assembly models can be applied to host-associated communities and what the relative contribution of dispersal and drift are to their assembly. In this context, the host can be conceptualized as an environment within which a community of microorganisms assembles. The vertebrate microbiome has a number of attributes that make it an ideal system in which to study ecological phenomena and test ecological theory. The communities associated with each individual host have relatively clearly delineated boundaries (e.g. the inside of an animal's gut compared to the outside) that are nevertheless open to migration. Since the microbiota of most vertebrates doesn't form until after birth or hatching, each new individual animal along with their associated microbiota is a kind of natural experiment, making it easy to observe the assembly of many replicate communities.

Animal hosts have several features which potentially make them unique from many other environments in nature, and thus interesting models with which to test the

robustness of ecological theory. Compared to abiotic environments, which are in general relatively static over short timescales, animals undergo rapid change during development, or ontogeny, especially early in life, and thus the environmental conditions that microorganisms experience when associated then can be highly dynamic. While some of these changes are independent of the microbiota, the ontogeny of the host is often influenced by interactions with associated microbes (Sommer and Bäckhed, 2013). Animals use microbial signals as cues to trigger a number of developmental changes, including proliferation and expansion of cells, morphogenesis of organs, and activation and priming of the immune system (Bates *et al.*, 2007, 2006; Hooper *et al.*, 2012). The immune system is a particularly unique feature which allows for the host environment to interact directly and reactively to the microorganisms in the community. Thus the host is able to simultaneously act as both an environment housing a community but also as an interacting member, adding an additional level of complexity to the system that isn't present in most other environments or models of community assembly.

The zebrafish, *Danio rerio*, microbiota

The complexity and diversity of host-microbe systems is both a feature and an obstacle. The use of model systems can help in controlling this complexity and in isolating the influence of different assembly processes. Zebrafish, *Danio rerio*, and the microbial communities colonizing their intestines are an excellent model of the vertebrate microbiome to help accomplish this task for a number of reasons. Large numbers of genetically similar individuals can be raised relatively inexpensively, providing a degree of replication unattainable in many other model vertebrates. Zebrafish are genetically

tractable, and a wide variety of mutant genotypes are available, including fish lacking innate or adaptive immune responses. Zebrafish have an intestine that is analogous in structure and organization to the human gastrointestinal tract. Of particular interest to studies of microbe-host interactions, there exist well established protocols for raising "germ-free" zebrafish in completely sterile environments and inoculating them with defined microbial taxa to create "gnotobiotic" zebrafish. As a natural microcosm, the zebrafish intestinal community combines the experimental tractability of artificial systems while still remaining biologically relevant, thus providing an excellent model system for testing ecological theory concerning the assembly of host associated microbial communities (Srivastava *et al.*, 2004).

Assembly of the zebrafish microbiota

The main goal of the research described here is to determine the role dispersal and ecological drift play in the assembly of the microbial communities associated with the zebrafish intestine, with a particular focus on how these processes interact with two factors that potentially make vertebrate-microbe systems unique: the development of the host and the host's immune system. Chapter II characterizes changes in the composition and diversity of the zebrafish intestinal microbiome over zebrafish development and show that host development is a major driver of community composition. This work helps establish the zebrafish intestinal microbiome as an ideal model for studying the assembly of host-microbe systems throughout host development and demonstrates that inter-individual variation is a key feature of the vertebrate microbiota across development. Chapter III explores these communities in more depth by using a neutral

model to determine how the relative importance of random dispersal and ecological drift to the assembly of the microbiota changes over the course of host development. This analysis also provides a framework for identifying features of these communities that depart from neutral predictions in ecologically informative ways. Chapter IV then explicitly tests the hypothesis that dispersal of microorganisms among hosts can have a significant influence on the structure and diversity of associated communities. Specifically, it describes an experiment showing how inter-host dispersal mediates the effects of another factor unique to animal hosts, the innate immune system, on the zebrafish microbiota. Finally, Chapter V synthesizes these findings together in the context of moving towards a conceptual model of host-microbe systems which incorporates factors and processes occurring beyond the scale of individual hosts.

This dissertation includes previously published and unpublished co-authored material. Chapters II and III were previously published as two separate articles in *The ISME Journal* with W. Zac Stephens, Keaton Stagaman, Sandi Wong, John Rawls, Karen Guillemin, and Brendan Bohannan as co-authors. Chapter IV was prepared with Meghna Agarwal, Karen Guillemin, and Brendan Bohannan as co-authors.

CHAPTER II

THE COMPOSITION OF THE ZEBRAFISH INTESTINAL MICROBIOTA VARIES ACROSS DEVELOPMENT

This chapter was published as an article in the *International Society of Microbial Ecology Journal (ISME J)* in 2016, and was co-authored by myself along with W. Zac Stephens, Keaton Stagaman, Sandi Wong, John F. Rawls, Karen Guillemin, and Brendan J. M. Bohannan. W. Zac Stephens and I contributed equally to this work as lead authors. The experimental work in this study was designed and performed collectively by myself, W. Zac Stephens, and Keaton Stagaman. Data analysis and writing of the manuscript was primarily performed by both myself and W. Zac Stephens with help from Keaton Stagaman. Keaton Stagaman, Sandi Wong, and John Rawls provided editorial assistance, while Brendan Bohannan and Karen Guillemin filled advisory roles throughout the project and acted as principal investigators. Supplemental materials and methods and results for this chapter can be found in Appendix A, and supplemental figures and tables can be found in Appendix B.

Introduction

Animal development occurs in a dynamic microbial world. The resulting associations between animals and microbes profoundly influence the maturation of their tissues and the function of adult organs. In particular, the development of the vertebrate digestive tract, which harbors the vast majority of microbial cells in the body, is strongly influenced by the presence and composition of the gut microbiota (Bates *et al.*, 2006;

Olszak *et al.*, 2012; Semova *et al.*, 2012; Sommer and Bäckhed, 2013). A comprehensive description of animal development must, therefore, include not only a catalog of the birth, specification and differentiation of the animal cells that comprise the body but also the associated microbial cells (McFall-Ngai *et al.*, 2013). Here we present a comprehensive survey of the intestinal microbiota of a single large sibling group (sibship) of the model vertebrate zebrafish (*Danio rerio*) throughout development.

Our study spanned major milestones in zebrafish development under common laboratory rearing conditions. Zebrafish are fertilized externally; therefore, developmental stages are often referenced as time post fertilization, with larval stages and beyond referred to in terms of days post fertilization (dpf). Zebrafish embryos initially develop in essentially sterile chorions and the larval stage begins when the organism hatches from its chorion and first encounters microbes in its external environment (between 2 and 3 dpf). By the time of hatching, most of the larva's organs have been specified but will continue to grow and mature into the adult structures in interaction with associated microbes. This includes the maturation of the intestine, which is open to the surrounding environment between 3 and 4 dpf, allowing exposure to microbial colonists (Bates *et al.*, 2006). At ~5 dpf, the yolk becomes depleted and larval zebrafish begin ingesting food. The development and differentiation of zebrafish continues into adulthood. Although juveniles of both sexes have ovary-like gonads, they differentiate into sex-specific gonads by ~4 weeks post fertilization and continue to develop secondary sex characteristics well into adulthood (~10-12 weeks post fertilization depending on the rearing conditions; Uchida et al., 2002). Initially, the ability of the host to defend against microorganisms is limited to innate immune activities, with the adaptive immune system

reaching functional maturity around 4 weeks post fertilization (Lam *et al.*, 2004). Many of the attributes that make zebrafish an excellent model for studying vertebrate development, such as its early optical transparency, small size, high-fecundity and availability of genetic and genomic resources (Howe *et al.*, 2013; Phillips *et al.*, 2014), also lend it to studies of vertebrate host-microbiota interactions. Large numbers of zebrafish can be maintained in a shared and easily sampled aquatic environment, allowing a high degree of biological replication along with information from associated environmental microbial communities.

The large degree of biological replication that is possible with zebrafish is an important advantage in understanding the extensive interindividual variation observed in vertebrate-associated microbiota (Friswell *et al.*, 2010; The Human Microbiome Project Consortium *et al.*, 2012; Rogers *et al.*, 2014). Interindividual variation in humans is greatest during early stages of infant colonization and decreases with age, while bacterial diversity within individuals generally increases from initial colonization at birth, stabilizing around 2-3 years of age (Avershina *et al.*, 2014; Palmer *et al.*, 2007; Yatsunenko *et al.*, 2012). During this period, weaning marks a dramatic transition for the developing infant microbiota as dietary change, the removal of maternally provided immunologic factors and loss of breastfeeding-derived microbes begins to shift the intestinal microbiota towards an adult-like composition (Bergstrom *et al.*, 2014). Thus, changes in diet and physiology over animal development are closely intertwined and likely interact to shape developmental changes in the associated microbiota.

In the present study, we exploit the advantages of the zebrafish model system to determine how associated microbial communities change along with key developmental,

environmental and dietary transitions of the host. Ours is the largest study to date of vertebrate intestinal microbiota from a single sibship of animals throughout development. We observed stage-specific changes in microbiota composition over development. Within each developmental stage, there remained extensive interindividual variation, despite the fact that the hosts belonged to a single sibship and shared the same rearing conditions and environments. Across development, we observed that the intestinal bacterial communities became increasingly different among individual hosts and distinct from the surrounding environment.

Materials and Methods

Experimental design and sample collection

We surveyed the gut microbiota of a pair of adult zebrafish parents and 135 of their offspring reared concurrently under identical environmental conditions at multiple stages in their development, using high throughput sequencing of the 16S rRNA gene. To reduce potential effects of host genotypic variation, this population consisted solely of offspring from a single mating pair. These siblings were split evenly among four replicate tanks, resulting in 70 fish per tank, and were raised in a manner intended to generally reflect commonly used zebrafish husbandry practices, including diet and water type, flow rate and frequency of changes (Figure 1a, lower portion; details in Supplementary Information). We sampled zebrafish and their surrounding tank environment at multiple time points meant to capture important developmental transitions: when the entire intestinal tract is first open and microbial colonization of the lumen first occurs (4 dpf), once fish must rely on ingesting food for nutrition (10 dpf), the maturation of the adaptive

immune system (21, 28 and 35 dpf), sexual maturity and dimorphism (75 dpf), and senescence (380 dpf; Figure 1a, upper portion). At each time point, we sampled the dissected intestines of multiple fish sampled evenly across each of the four replicate tanks, resulting in 20 fish (five per tank) per time point for ages 4 through 35 dpf, 24 fish (six per tank, three male and three female) at 75 dpf, and 18 fish (six per each of three replicate tanks) at 380 dpf (Figure 1a; some samples were later removed owing to poor sequencing depth). We also measured the standard length (SL) of each fish as a metric of zebrafish staging and growth (Parichy *et al.*, 2009). To examine the maturation of the adaptive immune system, we measured transcript levels of secreted immunoglobulin M (sIgM; Supplementary Information) from the carcasses of the 10, 21, 28, 35 and 75 dpf fish (time points spanning the course of immune maturation). Both SL and sIgM transcript levels increased with development (Figure 1b). Within a given age, there was much greater variation in sIgM transcript levels than in SL, which showed little variation until the last time point.

Intestinal and environmental samples were collected and prepared in a manner that minimizes cross-contamination of samples, tanks and time points. Sampled animals from each time point were collected from the fish facility before they were fed, at approximately the same time of day (between 0930 and 1000 h). Animals were then transported to dissection stations in their own tank water and euthanized by the addition of tricaine (2.1 ml of 0.4% tricaine per 50 ml fish water; 0.22 μm filtered) before dissection. Each animal was dissected on a separate, sterile glass slide (larva) or Petri dish cover (juvenile, adults) under a dissecting microscope as previously described (Milligan-Myhre *et al.*, 2011). Larval and juvenile fish dissections were performed using

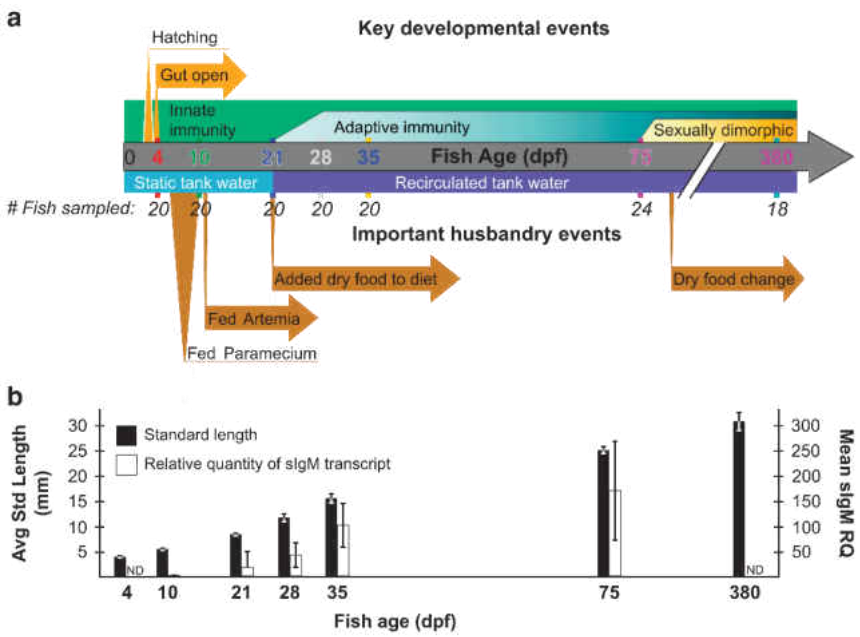


Figure 1: Experimental design and zebrafish development. (**a**) Experimental design showing important developmental events (top) and husbandry events (bottom) during the course of the study. The number of fish initially sampled among four tanks at each time point is shown, although post-sequencing rarefaction in some cases reduced this number for analyses (see Materials and Methods). Artemia are commonly called brine shrimp. Four and 10 dpf, fish are considered larvae, 21-35 dpf fish are juveniles and 75 and 380 dpf fish are adults. (**b**) The mean standard length and secreted IgM (sIgM) transcript levels (a proxy for adaptive immune development) of fish at sampled time points are shown with standard deviations. dpf, days post fertilization; ND, not determined.

individual-use insect pins, while sterile, individual-use scalpels were used for adult fish. The entire intestine from immediately posterior to the esophagus to the vent was removed intact. The swim bladder and liver were explicitly removed from the intestine, whereas no effort was made to remove the pancreas (if attached). The intestines were then placed in 2ml screw cap tubes containing 0.1mm zirconia–silica beads (Biospec Products, Bartlesville, OK, USA) and 200 (4, 10, 21 dpf) or 400 µl (28 dpf and older) of Enzymatic lysis buffer (ELB; Tris-EDTA pH 8.0 with 0.1% v/v Triton X-100; 0.22 µm filter sterilized) before freezing in liquid N2 and subsequent DNA extraction (as detailed in the Supplementary Information). The remaining carcass (without intestine, swim bladder,

liver and likely the pancreas) of each fish was stored in TRIzol (Life Technologies, Carlsbad, CA, USA) at -80 °C for further host RNA extraction and quantification of sIgM transcripts. For each sample age group, DNA extractions were performed on the same day as dissection.

Environmental samples collected at each time point included scrapings from two glass slides each (75 mm × 25 mm) that were affixed to the bottom (all time points) and sides (28, 35 and 75 dpf time points only) of tanks at the beginning of the study, as well as food samples and 500 ml of water per tank. Water samples were filtered through a 0.2 μm cellulose nitrate filter, which was then exposed to bead beating and DNA extraction from the filter using the same method used for the other samples. Poor DNA extraction efficiency or low number of sequences obtained from several of the environmental samples led to the retention of environmental samples from 4, 10 and 75 dpf time points only in this study. We also measured multiple environmental parameters from each tank, including temperature, pH, and ammonia, nitrite and nitrate concentrations, but these were all found to vary too little over the course of the study to provide explanatory power. Thus, discrete diet changes and a single environmental change from static water in a nursery facility to recirculating water in an adjacent main facility were the only measured environmental variables that appreciably varied between time points.

All zebrafish experiments were conducted in conformity with the Public Health Service Policy on Humane Care and Use of Laboratory Animals using standard protocols approved by the Institutional Animal Care and Use Committees of the University of Oregon and the University of North Carolina at Chapel Hill.

Illumina library preparation and 16S rRNA gene sequence analysis

The microbial communities of individual samples were characterized by Illumina (San Diego, CA, USA) sequencing of 16S rRNA gene amplicons. To obtain Illumina compatible amplicons that were amenable to a high degree of multiplexing, we used a two-step PCR method to add dual indices and Illumina adapter sequences to the V4 region of the bacterial 16S rRNA gene (see Supplementary Information for details and Supplementary Table 1 for oligonucleotide sequences) and obtain pairedend 150 nucleotide reads on the Illumina HiSeq 2000 platform. Illumina sequence reads have been deposited under the NCBI SRA accession number SRP047327.

The 16S rRNA gene Illumina reads were processed using methods implemented by mothur 1.28.0 (Schloss *et al.*, 2009) and QIIME 1.6.0 (Caporaso *et al.*, 2010) as detailed in the Supplementary Information. The final operational taxonomic unit (OTU) table was rarefied to a depth of 4250 sequences per sample, allowing us to retain all but seven (one each from 21, 28, 35 and 75 dpf groups and three from 380 dpf group discarded owing to low sequence depth) of the originally collected fish intestinal samples in subsequent analyses. Rarefaction curves showed that at this high depth of sampling, we were able to sample a large portion of the OTUs (defined using 97% sequence similarity) and diversity present while still retaining a large number of samples within fish of a given age (Supplementary Figure 1).

Diversity measures and statistical tests

All measures of community diversity and similarity, including OTU richness, phylogenetic diversity, Simpson's index and unweighted UniFrac distances, were

calculated in R (R Core Team, 2014) using the *vegan* (Oksanen *et al.*, 2016), *picante* (Kembel *et al.*, 2010). Permutational multivariate analysis of variance tests were performed using the *adonis* function from the *vegan* package. Phylogenetic diversity was measured as the total shared branch length of OTUs within each community (Faith, 1992). Tests for unimodality were done using Hartigans' dip test for unimodality (Hartigan and Hartigan, 1985). Identification of significant differences in relative abundances in bacterial classes or KEGG (Kyoto Encyclopedia of Genes and Genomes) functional groups among age groups was accomplished using the Kruskal–Wallis test with the Benjamini–Hochberg FDR correction. Discriminatory analysis of taxonomic groups among zebrafish ages was performed with LEfSe (Segata *et al.*, 2011).

Results

Zebrafish development is marked by major shifts in the dominant bacterial taxa of the intestinal microbiota

We set out to characterize the zebrafish intestinal microbiota over key developmental time points under standard laboratory rearing conditions, including diet and environment changes during larval and early juvenile stages. Diet and environment were held constant during late juvenile and an early adult stage (75 dpf), whereas a late adult stage (380 dpf), after a facility diet change, was added to compare adult microbiota of aged fish (Figure 1). Over the course of zebrafish development, the diversity of observed intestinal microbiota decreased significantly, both in terms of the number of OTUs (Figure 2a; $r^2 = 0.19$, $P < 1\times10^{-7}$), and phylogenetic diversity (Figure 2b; $r^2 = 0.15$, $P < 1\times10^{-5}$), with the largest changes occurring between 35 and 75 dpf, during which

time diet and environment were held constant but the fish experienced major developmental changes, such as sexual differentiation. The evenness of communities, however, remained relatively constant over host development (Figure 2c; $P = 0.288$). We did not observe significant differences in diversity between the stages of adult fish. We noticed that the 10 dpf samples appeared to be bimodally distributed with respect to taxa richness, with half of the samples having high richness and half low. We explicitly tested the unimodality of each distribution, and found that only the 10 dpf distribution was significantly non-unimodal (Hartigans' dip statistic $D = 0.1336$, $P = 0.002$ for 10 dpf samples and $P >> 0.1$ for all the other age groups). Although there was no significant difference between the means of the SLs of the two distributions (two tailed *t*-test: $P = 0.4159$), we did observe that the community composition of samples belonging to the high richness distribution were significantly more similar to 4 dpf samples than were those belonging to the low richness distribution (measured by comparing pairwise UniFrac distances: $P < 0.001$). This suggests that at ~10 dpf, the zebrafish transition from a larval to a juvenile microbiota, and that the fish we sampled were at different stages of this process despite being the same age. One potential explanation for this pattern would be a difference between fish that had begun consuming exogenous food earlier or later.

These changes in community diversity were accompanied by significant changes in the phylum level composition of larval (4 and 10 dpf), juvenile (21, 28 and 35 dpf) and

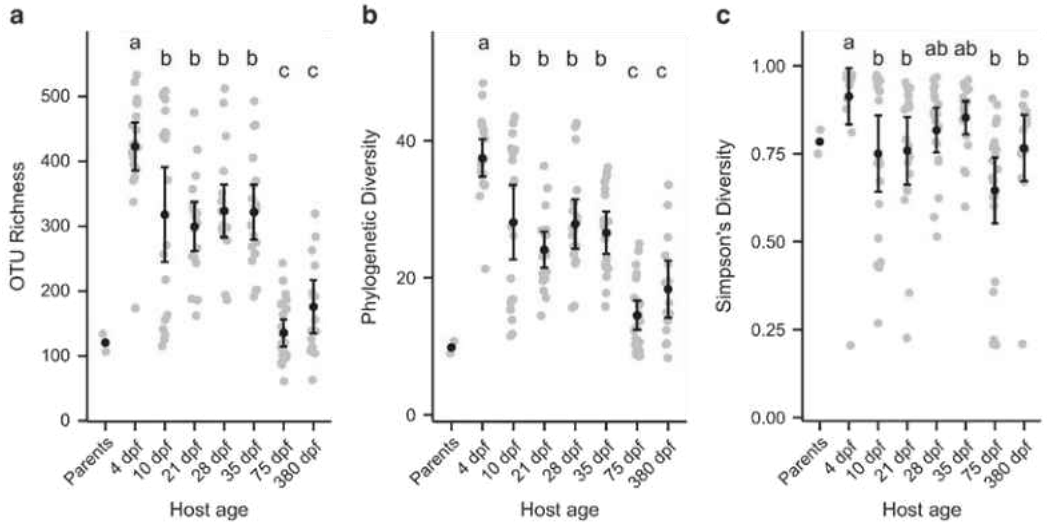


Figure 2: Significant changes in diversity of individual zebrafish intestinal communities throughout development. (**a**) Number of observed taxa. (**b**) Faith's phylogenetic diversity. (**c**) Simpson's diversity index. Black circles and error bars represent the means and 95% confidence intervals, respectively. Letters above age groups indicate significant differences in the means.

adult (75 and 380 dpf) fish, with particularly large differences in the taxonomic class composition of the Proteobacteria (Figure 3a, Supplementary Figure 3). The γ-proteobacteria were the most abundant class of bacteria in the study, and were especially abundant in larval intestines as well as environmental samples. Concurrent with the change in food and environment at 21 dpf, a marked increase in the abundance of α-proteobacteria was observed in the intestines and was followed by a decrease in abundance during the 28, 35 and 75 dpf age classes, during which time diet and environment were held constant. There was a decrease in the abundance of β-proteobacteria during these same stages, from a peak at 28 dpf Interestingly, the β-proteobacteria were particularly abundant in all food and environmental samples collected from 10 dpf fish and before (71% average; Supplementary Figure 3) yet were not consistently as abundant in intestinal samples until 35 dpf, suggesting a time lag

between initial exposure and detection of abundant colonization by this class. Although their relative abundances differed in environmental samples, the most abundant β-proteobacteria OTUs in 35 dpf fish were also detected in environmental samples, with the notable exception of an OTU belonging to the family Neisseriaceae that contributed to an average of 5.4% of the reads from 35 dpf fish. The proportion of these Proteobacteria classes decreased strongly in 75 dpf fish, despite being fed the same diet as the juveniles. Consistent with previous studies of the adult zebrafish intestine (Rawls *et al.*, 2004, 2006; Roeselers *et al.*, 2011), we found Fusobacteria to be abundant in the adult stage (75 and 380 dpf) intestinal samples (30% and 12%, respectively), although they accounted for <1% of the total community in the 21–35 dpf fish. Also in agreement with our previous study (Roeselers *et al.*, 2011), we found a large diversity of Fusobacteria OTUs within intestines (168), with the majority of these OTUs (90%) being classified in the genus *Cetobacterium*, which was found in all 38 adult intestines. OTUs belonging to the Aeromonadaceae family (γ-proteobacteria class) that could not be further classified to genus were the only grouping found in all of the 137 intestinal samples analyzed (Table 1). When broken down by developmental group, 10 more genera were found in all the larval intestines (4 and 10 dpf) and 11 more in all the juvenile (21, 28 and 35 dpf) intestines, whereas only the *Plesiomonas* and *Cetobacterium* genera were also found in all the adult intestines. Many of these genera were previously identified as part of a core microbiota of the adult zebrafish intestine (Roeselers *et al.*, 2011). Additional core genera found in >90% of intestines from a given developmental stage group included *Shewanella*, *Vibrio*, *Pseudomonas* and *Streptococcus*. The overall abundance of these core taxa varied from nearly 14% for the OTUs within the Aeromonadaceae to < 1% for

the *Streptococcus*, with considerable variation in abundance observed among age groups. An abundant class of uncultured Firmicutes (placed as a separate phyla in some reference taxonomies) referred to as 'CK-1C4-19' was found in 89% of all intestinal samples and represented nearly 14% of the total reads in adult intestines. We detected 87 different OTUs from this uncultured class, which represented 3.8% of the total OTUs in the adult zebrafish intestines (Supplementary Figure 3). We next asked whether any of these taxonomic groups were strongly associated with particular stages of zebrafish development. For this analysis, we combined the 75 and 380 dpf fish intestinal communities together into a single class ('adult'), and analyzed all the age classes using a nonparametric test of significance and linear discriminant analysis with the defaults implemented by LEfSe (Kruskal–Wallis; $P < 0.05$ and log 10 linear discriminant analysis score > 2.0). These analyses identified 184 discriminatory taxa, of which the majority (95) distinguished the youngest, nonfeeding (4 dpf) age class from all others. The discriminating taxa for the 4 dpf fish largely belonged to the Proteobacteria despite this phylum's abundance in the entire data set. To determine the most highly discriminatory bacterial taxa for each age class, we implemented stricter cutoffs for LEfSe ($P < 0.01$, log 10 linear discriminant analysis score > 3.5; Supplementary Figure 4). The 10 dpf age class, which had begun feeding on *Paramecium*, were distinguished by the consistent presence of the genus *Mycobacterium*. Notably, this genus contains the fish pathogens *M. marinum* and *M. chelonae*, which were known to be present in our facility during the time of the experiment, and which have recently been shown to be efficiently transmitted to zebrafish via ingestion of infected paramecia (Peterson *et al.*, 2013). Although our sequences did not allow us to resolve the species-level identification of these

Mycobacterium OTUs, we detected *Mycobacterium* sequences in every environmental sample associated with *Paramecium* feeding (eight samples, average 4% abundance), including all three replicates of the *Paramecium* food samples, suggesting a possible transmission route. The juvenile age classes were largely discriminated by the presence of β-proteobacteria lineages, while the adult class was distinguished by the low abundance Bacteroidaceae family, the prevalent Fusobacteria (specifically the Cetobacterium genus) and by the CK-1C4-19 candidate class of Firmicutes. Differences in taxonomic composition throughout development were reflected by differences in the predicted functional capacity of these taxa (Supplementary Information), which included predicted differences in the representation of genes involved in cell motility and carbohydrate metabolism between adult and younger fish (Supplementary Figure 6c).

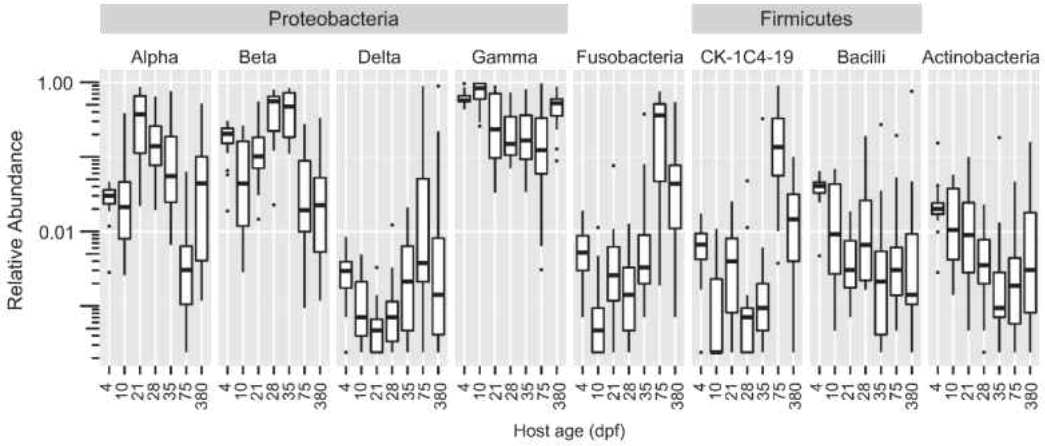


Figure 3: Major shifts in bacterial taxa throughout development. Bacterial classes with 41% average relative abundance across all ages, plotted on a log scale (All taxonomic classes $P < 0.0001$, Kruskal–Wallis).

Table 1 Genera found in greater than 90% of larval, juvenile or adult intestines

Genus or best (taxa level) classification	Phylum	All intestinal samples		Larval		Juvenile		Adult		Relative core genera from Roeselers et al., 2011[a]	Zebrafish isolated strain representatives with genome sequences[b,c]
		% Presence in all intestines (n = 137)[c]	% Total intestinal reads[c]	% Presence in larval intestines (n = 40)	% Total larval reads[c]	% Presence in juvenile intestines (n = 57)	% Total juvenile reads[c]	% Presence in adult intestines (n = 38)	% Total adult reads[c]		
Unclassified Aeromonadaceae (family)	Proteobacteria	100.00	13.92	100.00	14.22	100.00	16.92	100.00	9.84	Aeromonas	ZOR0001 ZOR0002
Shewanella	Proteobacteria	97.08	2.65	100.00	6.77	100.00	1.07	89.47	0.76	Shewanella	ZOR0012
Unclassified Enterobacteriaceae (family)	Proteobacteria	95.62	2.82	100.00	7.32	100.00	0.77	84.21	1.32	Unclassified Enterobacteriaceae	ZOR0011 ZOR0014
Other Enterobacteriaceae (family)	Proteobacteria	95.62	1.35	100.00	3.46	94.74	0.65	92.11	0.24		
Unclassified Comamonadaceae (family)	Proteobacteria	95.62	11.31	97.50	6.03	100.00	21.39	88.47	2.35	Diaphorobacter	ZNO0006 ZNO0007 ZNO0008 ZOR0011
Plesiomonas	Proteobacteria	95.62	6.81	97.50	8.95	91.23	0.95	100.00	12.35		
Other Gammaproteobacteria (class)	Proteobacteria	94.89	0.88	100.00	2.34	96.49	0.35	86.84	0.18		
Catobacterium	Fusobacteria	94.89	6.92	85.00	0.38	98.25	1.31	100.00	22.39	Catobacterium	ZWU0022 ZOR0034
Unclassified Halieaceae (family)	Proteobacteria	94.16	3.76	95.00	0.40	100.00	8.01	89.47	1.14		ZOR0017
Pseudomonas	Proteobacteria	90.51	3.49	100.00	4.13	100.00	4.01	71.05	2.22	Pseudomonas	ZWU0006
Rhodobacter	Proteobacteria	89.78	1.10	87.50	0.20	100.00	1.51	81.58	1.47		
Other Pseudomonadaceae (family)	Proteobacteria	89.05	0.62	97.50	0.80	94.74	0.60	73.68	0.48		
Unclassified CK-1C4-19 (class)	Firmicutes	89.05	4.76	80.00	0.46	91.23	0.94	94.74	13.97		ZOR0006
Other Comamonadaceae (family)	Proteobacteria	86.86	1.10	95.00	0.92	98.25	1.95	65.79	0.07		
Unclassified Xanthomonadaceae (family)	Proteobacteria	84.67	0.38	100.00	0.79	78.95	0.21	78.95	0.23	Stenotrophomonas	
Vibrio	Proteobacteria	83.21	1.83	100.00	5.39	89.47	0.55	57.89	0.10	Vibrio	ZWU0020 ZOR0035
Unclassified Betaproteobacteria (class)	Proteobacteria	81.75	0.22	97.50	0.39	91.23	0.23	55.26	0.04		
Unclassified Rhizobiales (order)	Proteobacteria	81.75	1.44	75.00	0.15	100.00	3.21	63.16	0.23		ZNO0028 ZNO0032
Pseudoalteromonas	Proteobacteria	81.02	2.01	100.00	5.48	96.49	0.89	39.47	0.13		
Unclassified Legionellales (order)	Proteobacteria	80.29	0.74	87.50	0.87	98.25	0.85	50.00	0.49		
Unclassified Aeromonadales (order)	Proteobacteria	78.83	0.18	97.50	0.19	71.93	0.13	73.68	0.23		ZOR0001 ZOR0002
Unclassified Hyphomicrobiaceae (family)	Proteobacteria	78.83	0.30	90.00	0.60	87.72	0.23	57.89	0.11		
Other Rhizobiales (order)	Proteobacteria	78.83	3.18	85.00	1.47	100.00	6.55	42.11	0.07		
Other Betaproteobacteria (class)	Proteobacteria	77.37	0.16	85.00	0.15	100.00	0.26	38.47	0.02		
Unclassified Rhodospirillaceae (family)	Proteobacteria	75.91	0.22	92.50	0.36	85.96	0.18	47.37	0.12		
Streptococcus	Firmicutes	72.99	0.63	100.00	1.73	77.19	0.26	42.11	0.08	Streptococcus	
Delftia	Proteobacteria	72.26	0.30	92.50	0.33	66.67	0.13	63.16	0.54		ZNO0008
Unclassified Pseudomonadaceae (family)	Proteobacteria	70.80	0.41	90.00	0.37	87.72	0.69	28.95	0.07		ZWU0006
Unclassified Rhodobacteraceae (family)	Proteobacteria	70.07	0.89	40.00	0.03	98.25	1.91	63.16	0.29		
Other Rhodobacteraceae (family)	Proteobacteria	68.61	1.80	20.00	0.02	100.00	3.99	73.68	0.48		
Unclassified Vibrionaceae (family)	Proteobacteria	67.88	1.34	72.50	0.23	45.61	0.02	94.74	4.54	Vibrio	ZWU0020 ZOR0018 ZOR0035
Pelomonas	Proteobacteria	66.42	0.81	97.50	1.52	68.42	0.84	34.21	0.04		
Marinomonas	Proteobacteria	65.69	0.25	92.50	0.64	82.46	0.14	13.79	0.01		
Unclassified Rhizobiaceae (family)	Proteobacteria	64.23	1.83	47.50	0.05	96.49	4.08	34.21	0.43		ZNO0028
Corynebacterium	Actinobacteria	62.77	0.43	95.00	1.24	56.14	0.12	39.47	0.05		
Mycobacterium	Actinobacteria	62.77	0.26	92.50	0.48	56.14	0.07	44.74	0.33		
Zoogloea	Proteobacteria	61.31	0.48	90.00	1.40	54.39	0.13	42.11	0.08		
Halomonas	Proteobacteria	60.58	1.19	95.00	3.54	70.18	0.35	13.16	0.04		
Erwinia	Proteobacteria	58.39	0.14	95.00	0.38	50.88	0.06	34.21	0.02		
Staphylococcus	Firmicutes	56.93	0.41	92.50	0.64	43.86	0.49	39.47	0.07		ZWU0021
Candidatus Rhabdochlamydia	Chlamydiae	50.36	0.21	92.50	0.68	42.11	0.02	21.05	0.03		

[a]Genera shared in all adults from Roeselers et al. are listed with corresponding genera (or family where genus was ambiguous) from this study. No attempt is made to relate the single core operational taxonomic unit (OTU) from Roeselers et al. with only a phylum-level assignment. [b]Isolated strains with newly sequenced genomes that belong within a listed taxonomic group. Some strains are listed more than once where they may fall within more than one taxonomic grouping that could not be confidently classified to finer levels with short Illumina reads. [c]Isolated strains which were identified as discriminatory for a given age class are highlighted in bold (larval), italics (juvenile) or underlined (adult). [d]"All intestines" includes the two parent samples as well as the larval, juvenile and adult intestines from their offspring, listed to the right. [e]Percentages are derived from samples rarefied to 4250 sequences per sample.

Table 1: Genera found in greater than 90% of larval, juvenile or adult intestines.

Variation in microbial community composition changes over zebrafish development

Despite the genetic similarity and shared environment of their hosts, the overall composition of microbial intestinal communities showed a substantial amount of variation among fish, as measured by the UniFrac distance (Figure 4). Communities associated with individual fish were more similar to communities associated with fish of the same age than they were to those associated with fish of different ages (Figure 4a; permutational multivariate analysis of variance; $r^2 = 0.18$; $P < 0.001$). Over time, variation among hosts significantly increased, but the effect was small and non-monotonic in the juvenile stages (Figure 4b; $r^2 = 0.10$; $P < 0.001$). To determine whether there were possible tank effects, we performed a permutational multivariate analysis of variance with age, tank and age by tank as factors. Neither tank nor the interaction of age by tank produced significant effects ($P = 0.930$ and $P = 0.363$, respectively), suggesting there was little to no tank effect that would influence the interpretation of our results. Before 75 dpf, we were unable to assign sex to each zebrafish using external traits; however, we did not see a significant effect of sex on community similarity at 75 or 380 dpf ($P = 0.11$ and $P = 0.12$, respectively). Given these results, we grouped samples from fish raised in different tanks and fish of both sexes together for the remainder of the analyses.

We next attempted to explain variation in community composition using measures of host age, SL, and sIgM transcript levels to disentangle the relative influence of time, development, and immune maturation. We used multiple regression analysis (Lichstein, 2007) to partition the variation in pairwise UniFrac distances among hosts into the total amount of variation explained by the above host variables (that is, 'total'), the variation

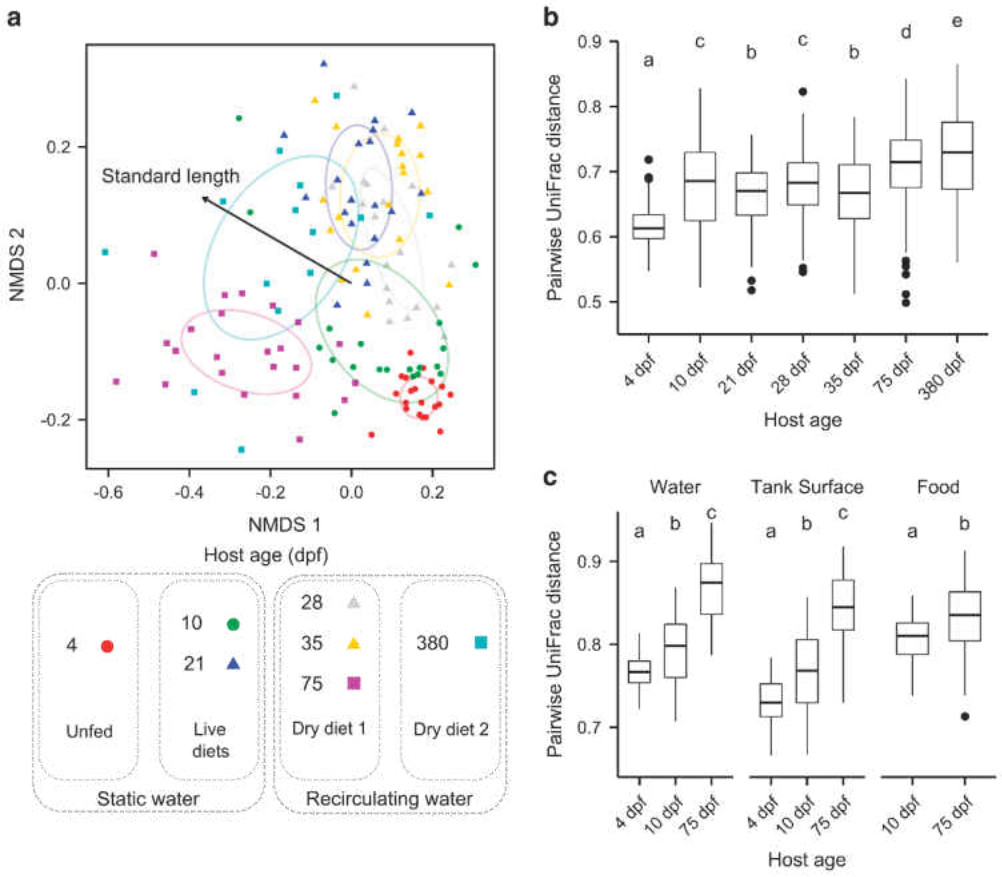


Figure 4: Phylogenetic dissimilarity of microbiota from fish and environmental samples. (**a**) A non-metric multidimensional scaling (NMDS) ordination of (unweighted) UniFrac distances among zebrafish intestinal communities across development. Host age is differentiated by the color of points, while husbandry conditions (diet and water flow) are described in the legend. The age class of the host is indicated by the shape of points: circles indicate larvae, triangles indicate juveniles and squares indicate adults. The effect of standard length on the spread of points is shown by a vector. (**b**) Pairwise Unifrac distances among fish within each age group. (**c**) Pairwise Unifrac distances between fish and environmental communities at each age group. For each boxplot, letters above age groups indicate significant differences in the means.

explained simultaneously by multiple variables (that is, 'shared') and the variation uniquely explained by each variable independent of the others (that is, 'unique'; Table 2). Across the data set, differences in the SL of the zebrafish explained more variation in among-host UniFrac distances than did differences in host age, despite host age and SL themselves being strongly correlated ($r^2 = 0.60$, $P < 0.001$; Figure 1b). To determine the

potential role of adaptive immune function in structuring communities, as well as isolate the effects of development from changes in diet and housing, we next compared the explanatory power of differences in SL and differences in sIgM transcript levels (for those samples with measurable sIgM transcription levels and that shared common husbandry conditions; that is, 28-75 dpf samples). We found that SL was a much stronger predictor than sIgM transcript abundance, which explained relatively little variation in UniFrac distances (Table 2). It is also worth noting that the explanatory power of SL was much higher for these age groups, possibly the result of housing conditions being constant for these time points, thus enhancing the relative contribution of host development.

Table 2: Results of multiple regressions comparing community dissimilarity with differences in host age, standard length and sIgM concentrations.

Host Variable	R^2	P-value*
Across all ages:		
DPF + Standard Length (total)	0.2818	0.001
DPF + Standard Length (shared)	0.0775	0.001
DPF (unique)	0.0057	0.029
Standard Length (unique)	0.1986	0.001
For 10, 21, 28, 35, and 75 dpf zebrafish†:		
Standard Length + [sIgM] (total)	0.5084	0.001
Standard Length + [sIgM] (shared)	0.0495	0.001
Standard Length (unique)	0.4427	0.001
[sIgM] (unique)	0.0162	0.037

Abbreviations: days post fertilization, dpf; secreted immunoglobulin M, sIgM.
* P-values were calculated from a distribution of 1000 random permutations.
† These ages had measurable sIgM levels and shared common husbandry conditions.

We computed the pairwise dissimilarity between fish intestinal communities and each of the three communities associated with the external environment: the tank water, tank surfaces and food fed to the fish. Fish intestinal communities were more similar to other fish intestinal communities than they were to any environmental communities (Figure 4c; $P < 0.001$ for all comparisons post Bonferroni correction). The dissimilarity between intestinal communities and environmental communities increased over time, ($P < 0.0001$; $r^2 = 0.59$, 0.52 and 0.13 for comparisons to tank water, surfaces and food samples, respectively). As a result, the intestinal communities associated with young 4 and 10 dpf fish were significantly more similar to surrounding environmental communities than were older 75 dpf fish ($P < 0.001$ for tank water, surface and food environments). This pattern was further manifested by increased differentiation of predicted fish associated metagenomes from predicted environmental metagenomes (Supplementary Figure 6e).

Discussion

The microbial community of the animal gut has been described as an additional host 'organ'; however, its assembly, analogous to the process of organ development, is poorly understood. Here we have characterized the development of the intestinal microbiota from a single sibship of zebrafish and show that while distinct bacterial communities assemble at different stages, the cellular composition is much less stereotyped than developing host tissue. Instead, we observe extensive interindividual variation in intestinal microbiota composition at each developmental stage, despite our ability to control host genotype and environment, that mirrors the interindividual

variation routinely observed in other vertebrate hosts, including humans (Caporaso *et al.*, 2011; The Human Microbiome Project Consortium *et al.*, 2012) and mice (Benson *et al.*, 2010; Rogers *et al.*, 2014). We conclude that interindividual variation in microbiota is a characteristic of vertebrates across development.

This study was designed to understand developmental stage-specific intestinal microbiota composition and diversity under standard laboratory rearing conditions of the zebrafish, and provides a reference for future studies investigating the crosstalk between developing zebrafish hosts and their microbiota. We observed major compositional shifts both during periods of development when diet and environment were also changing (that is, from larval to juvenile stages) as well as when diet and environment remained constant (that is, from late juvenile to adult stages) suggesting that host physiological development likely has significant effects on the microbiota independent of the other factors. The appearance of an adult-like microbiota in mammals begins shortly after weaning, when the introduction of solid foods and the removal of maternally provided immune factors (in particular breast-milk-derived immunoglobulins) impact the composition of the infant gut microbiota (Bergstrom *et al.*, 2014; Rogier *et al.*, 2014). In this study, we detected bacterial taxa characteristic of adult zebrafish (such as Fusobacteria and the CK-1C4-19 class) early in development but they remained low until the adult stages. While we observed increased sIgM levels during the transition from juvenile to adulthood that could influence these compositional shifts, differences in sIgM levels explained far less variation among microbiota than did SL. This further supports the notion that morphological changes during development are likely the dominant drivers of changes in the microbiota, at least during periods when diet and environment

are constant.

We found that the communities associated with larval fish were more similar to communities associated with the surrounding environment than were adult fish, indicating a greater role of environmental exposure early in development. This is consistent with observed correlations in humans between birth delivery mode and the composition of the intestinal microbiota (Dominguez-Bello *et al.*, 2010), and the relative instability of the intestinal microbiota between early stages of postnatal development in human newborns (Mackie *et al.*, 1999; Palmer *et al.*, 2007; Koenig *et al.*, 2011; Bäckhed *et al.*, 2015) and in mice (Pantoja-Feliciano *et al.*, 2013). These similarities suggest that the intestinal environments of mammals and fish may be similar in some key aspects (for example, relative changes in oxygen concentration through development), and that environmental exposures including diet can have significant impacts on the observed composition of intestinal microbiota early in development.

The changes we observed in community composition during the development highlight the need for careful consideration of developmental context in studies of host-microbe interactions. At the very least, comparisons across studies should strive to use consistent ages and development stages of the sampled hosts. It is well established that the presence and composition of the microbiota influences a wide array of host developmental and physiologic processes in zebrafish and other animal hosts (Rawls *et al.*, 2004; Bates *et al.*, 2006; McFall-Ngai *et al.*, 2013). Our results highlight observations that some of the observed phenotypic variation in animal studies is, in part, owing to variations in the microbiota. For example, it was recently shown that differences in microbial community composition in wild-type mice alter the intestinal IgA levels,

thereby differentially influencing susceptibility to a chemically induced model of colitis (Moon *et al.*, 2015). Likewise, it is possible that our observation of greater variation in sIgM transcript abundance in older fish, with little variation in size, is simply a reflection of increasing microbiota variation between individuals in response to widely varying microbial communities. In the future, it may be helpful to develop and deploy engineered communities of cultured microbes to provide reproducible microbiotas for broad use, or use experimental design strategies that control for microbial variation between individuals, stages, clutches, tanks, pedigrees and facilities. Our characterization of intestinal microbiota dynamics across zebrafish development and the genomes of representative members of these communities provide a useful resource and framework for such future studies.

BRIDGE

In Chapter II, I characterized the microbial communities associated with zebrafish (*Danio rerio*) intestines and described how the composition and diversity of these communities changes over host development. This work is largely exploratory, but important in that it provides an ideal model system in which to study the assembly of microbial communities associated with vertebrate hosts. Of note is the observation that unlike the typical development of host tissues, these communities are characterized by high degrees of inter-host variation, and the magnitude of this variation changes over host development. Furthermore, while host factors, particularly standard length, could explain a large amount of this variation across age groups, none of the measured factors could explain a significant amount of variation among hosts of the same age. This raises the question of what the source of this variation is. In Chapter III, I attempt to answer this question by using a neutral model to determine whether random processes alone, specifically passive dispersal and ecological drift, are sufficient to explain variation among hosts in the distribution of microbial taxa. In doing so, I also test the hypothesis that the importance of these neutral assembly processes decreases as hosts undergo development.

CHAPTER III

CONTRIBUTION OF NEUTRAL PROCESSES TO THE ASSEMBLY OF GUT MICROBIAL COMMUNITIES IN THE ZEBRAFISH OVER HOST DEVELOPMENT

This chapter was published as an article in the *International Society of Microbial Ecology Journal (ISME J)* in 2016, and was co-authored along with W. Zac Stephens, Keaton Stagaman, Sandi Wong, John F. Rawls, Karen Guillemin, and Brendan J. M. Bohannan, with myself filling the role of lead author. The study described in this chapter uses the same dataset described in Chapter II above that was collected jointly by myself, W. Zac Stephens, and Keaton Stagaman. However, the analysis and writing of the manuscript was performed by myself with the other co-authors providing editorial assistance. Once again, Brendan Bohannan and Karen Guillemin filled advisory roles and Brendan Bohannan was the principal investigator of the study. Supplemental figures and tables can be found in Appendix B.

Introduction

The microorganisms that reside on and inside humans and other vertebrate animals are remarkable, not only because of their importance to host health and development (Bates *et al.*, 2006; Fraune and Bosch, 2010; Sommer and Bäckhed, 2013), but also because they assemble into complex communities *de novo* in every new hatchling or infant host. The processes responsible for structuring these complex systems, often referred to as a host's *microbiota*, are not well understood despite a strong interest

in manipulating them to improve human health. Recent advances in sequencing allow us to observe and describe microbial communities with unprecedented depth and accuracy, but using these data to make inferences about how they assemble remains challenging. One approach to addressing this issue is to adopt a conceptual framework where animal hosts are viewed as ecosystems and their associated microbiota are treated as ecological communities (Costello *et al.*, 2012; Dethlefsen *et al.*, 2006; Robinson *et al.*, 2010). This approach is attractive because it allows researchers to borrow concepts and tools developed over decades of research in ecology.

In host-associated systems there are a large number of specific factors that may contribute to community assembly. Many host specific factors have been studied, including host species, genotype, diet, and health (Benson *et al.*, 2010; Rawls *et al.*, 2006; Goodrich *et al.*, 2014; Turnbaugh *et al.*, 2006), as well as microbe specific factors, including mutualistic and competitive (Levy and Borenstein, 2013; de Muinck *et al.*, 2013). While the list of potential factors is long, they can be divided into two major categories: selective processes, in which microbes establish and thrive in an environment (in this case the host itself) due to differences in their relative ecological fitness; and neutral processes, which include the dynamics of passive dispersal (e.g. sampling individuals from a source pool of available colonists) and the effects of ecological drift (the stochastic loss and replacement of individuals; Chase and Myers, 2011). While considerable progress has been made to investigate the roles of specific interactions between microbes and their hosts, the relative roles of dispersal and ecological drift in shaping host-associated microbial communities has largely been ignored (but see Jeraldo *et al.*, 2012; Lankau *et al.*, 2012; Venkataraman *et al.*, 2015). In contrast, these processes

have been studied in the general field of ecology for decades, with a renewed surge of interest in recent years (Caswell, 1976; Hubbel, 2001; Rosindell *et al.*, 2011).

Neutral and other sampling based theories provide an ideal starting place for investigating assembly patterns because of their relative simplicity. Neutral theory derives its name from its defining assumption of equivalent per-capita growth, death, and dispersal rates of species, thus assuming species are "neutral" in their ecological fitness. In the absence of such differences, community assembly is the result of the stochastic processes of dispersal and drift; organisms in the community are randomly lost, and are replaced at random by individuals from within the community or by dispersal of individuals from outside the community. While these assumptions of ecological equivalence may seem over-simplified, neutral models have successfully predicted the structures of many communities, including microbial communities (Östman *et al.*, 2010; Ofiteru *et al.*, 2010; Woodcock *et al.*, 2007; Venkataraman *et al.*, 2015). Such models are particularly useful in modeling microbial systems, where the immense diversity of communities makes characterizing the specific ecological traits of each individual taxon difficult. They also allow researchers to quantify the importance of processes which are difficult to observe directly, such as dispersal, but can nevertheless have large impacts on microbial communities (Kerr *et al.*, 2002; Lindström and Östman, 2011).

Given the variable nature of host-associated microbial communities, a comprehensive investigation of the role of neutral processes in structuring these communities requires a high degree of replication and control. In this regard the intestinal microbiota of the zebrafish (*Danio rerio*) is an ideal experimental system. Zebrafish have historically been used to study vertebrate development, but have also

recently emerged as an ideal model for studying interactions between vertebrate hosts and their associated microbial communities (Rawls *et al.*, 2006; Roeselers *et al.*, 2011; Yan *et al.*, 2012; Zac Stephens *et al.*, 2016). This is in large part due to the feasibility of raising a large number of individuals from a single crossing and co-housing them throughout their lifespan in highly controlled environments, thereby minimizing the effects of inter-host variation and ensuring that all individuals are exposed to a shared source pool of microorganisms.

In the present study we assess the ability of neutral models to explain the distribution of microorganisms among a population of zebrafish, and then determine the conditions leading to departures from neutral behavior. In doing so, we adopt a conceptual framework in which we consider the microorganisms associated with individual zebrafish hosts to be *local communities* that are a part of a broader *metacommunity* consisting of the microorganisms associated with all of the hosts in the population (Costello *et al.*, 2012; Leibold *et al.*, 2004). We hypothesized that the ability of hosts to differentially select their microbial inhabitants increases with developmental age, thereby decreasing the relative importance of neutral processes. Assuming that decreases in the fit of the neutral model are indicative of increased selection pressures, we expected that deviations from the neutral prediction should be compositionally and phylogenetically distinct, to the extent that ecological traits are phylogenetically conserved. In addressing these hypotheses, we also provide a framework for identifying communities and taxa of potential interest based on the degree to which they diverge from the predictions of neutral theory.

Materials and Methods

Zebrafish Microbiota Longitudinal Study

For the present study, we used a 16S rRNA gene sequence dataset from a previously reported longitudinal study of the developing zebrafish intestinal microbiome (Zac Stephens *et al.*, 2016). A brief description of the study design and sample collection follows, and readers are referred to the previous chapter for additional details. A population of zebrafish resulting solely from a single mating pair was raised under identical conditions, in order to minimize both genetic and environmental heterogeneity, and sampled at multiple ages, conventionally measured by days post fertilization (dpf). Zebrafish embryos develop within sterile chorions and are not exposed to microorganisms in their environment until they hatch (between 2 and 3 dpf). This population was divided evenly among four replicate tanks (resulting in 70 fish per tank) prior to hatching in order to ensure a shared initial exposure and account for potential tank effects. These fish were then raised under standard laboratory rearing conditions. The intestines of individual fish from this population were aseptically removed (as per Milligan-Myhre *et al.*, 2011), and the associated microbial communities were characterized by 16S rRNA gene amplicon sequencing at important developmental milestones: 4 dpf (complete opening of the digestive tract), 10 dpf (after feeding began), 21, 28, and 35 dpf (activation of the adaptive immune system, Lam *et al.*, 2004), 75 dpf (sexual maturity), and finally 380 dpf (onset of senescence). At each time point, twenty fish (five from each tank) were randomly selected for sampling, with the exception of the 75 dpf time point at which time twenty-four fish (three male and three female from each tank; prior to 75 dpf the sex of fish could not be confidently determined), and the 380 dpf

time point, at which time six fish from each of three remaining tanks were sampled. In addition, the microbial communities of the surrounding water, tank surfaces, and food for 4, 10, and 75 dpf time points were also sampled and characterized.

The zebrafish used in this study were raised under conventional laboratory conditions. This involved a number of husbandry changes. Prior to 21 dpf, the zebrafish were raised in a nursery tank with uncirculated water that was exchanged manually on a daily basis. Just prior to sampling at 21 dpf, the fish were transferred to a main facility system where water was continuously recirculated at a fixed rate through a sand and UV filter. The diet also changed over the course of the study: prior to 6 dpf fish were not fed and subsisted off their yolks alone, after which time fish were feed live *Paramecia* from 6 dpf to 10 dpf, live *Artemia* (brine shrimp) just after fish were sampled at 10 dpf to 21 dpf, and a standard dry fish food mixture from 21 dpf onward. Between 75 and 380 dpf, the manufacturer of this standard diet changed, but the feeding schedule remained the same. All zebrafish experiments were conducted in conformity with the Public Health Service Policy on Humane Care and Use of Laboratory Animals using standard protocols approved by the Institutional Animal Care and Use Committees of the University of Oregon and the University of North Carolina at Chapel Hill.

The microbial communities sampled in this study were characterized by Illumina sequencing of the 16S rRNA gene. 16S rDNA sequences from the V4 region of the 16S rRNA gene, subsequent 97% similarity operation taxonomic unit (OTU) tables rarefied to 4,250 sequences per sample, and taxonomic classifications were taken directly from (Zac Stephens *et al.*, 2016).

Sloan neutral community model for prokaryotes

To determine the potential importance of neutral processes to community assembly, we assessed the fit of the Sloan Neutral Community Model for Prokaryotes to the distributions of microbial taxa in our data (Sloan *et al.*, 2006). This neutral model predicts the relationship between the frequency with which taxa occur in a set of local communities (in this case individual zebrafish intestinal communities) and their abundance across the wider metacommunity (the intestinal communities of all zebrafish sampled at a given time point). In general, the model predicts that taxa that are abundant in the metacommunity will also be widespread, since they are more likely to disperse by chance and be randomly sampled by an individual host, while rare taxa are more likely to be lost from individual hosts due to ecological drift. In contrast to many other contemporary neutral models, namely the unified neutral theory of biodiversity (Hubbel, 2001), the neutral model used here does not incorporate the process of speciation. However, while microbial speciation and diversification are no doubt important in generating the diversity of microorganisms in this system at the broad, regional level, our explicit focus is on the assembly of host-associated communities over the course of host development. As such, it is highly unlikely that microbial diversification will occur over that time span to an extent that it would impact diversity among communities at the resolution we observe them (i.e. 97% similarity in 16S gene sequences).

The Sloan neutral model is fit to the observed frequency of occurrence of OTUs (i.e. the proportion of local communities in which each OTU is detected) and their abundance in the metacommunity (estimated in this case by the mean relative abundance across all local communities) by a single free parameter describing the migration rate, m.

This estimated migration rate is the probability that a random loss of an individual in a local community will be replaced by dispersal from the metacommunity, as opposed to reproduction within the local community, and can thus be interpreted as a measure of dispersal limitation. The fitting of this parameter was performed in R using non-linear least-squares fitting and the *minpack.lm* package (R Core Team, 2016). Binomial proportion 95% confidence intervals around the model predictions were calculated using the Wilson score interval (Brown *et al.*, 2001). We assessed the overall fit of the model to observed data by comparing the sum of squares of residuals, SS_{err}, to the total sum of squares, SS_{total}: model fit = 1 - SS_{err}/SS_{total} (generalized R-squared; Östman *et al.*, 2010). The fit of the neutral model was also compared to the fit of a binomial distribution model in order to determine whether incorporating drift and dispersal limitations improve the fit of a model beyond just random sampling of the source metacommunity (Sloan *et al.*, 2007). Sampling from a binomial distribution represents the case where local communities are random subsets of the metacommunity in the absence of processes of drift and dispersal limitations. While generalized r-squared is a useful measure for comparing the fit of multiple datasets to a single model, it is a poor choice for comparing the fit of multiple models to a single dataset (Spiess and Neumeyer, 2010). Therefore, to compare the fit of the neutral and binomial model, we compared the Akaike information criterion (AIC) of each model. Computation of the Akaike information criterion was done in R. Calculation of 95% confidence intervals around all fitting statistics were done by bootstrapping with 1000 bootstrap replicates.

To analyze deviations from the neutral model predictions, we compared the composition and diversity of neutrally and non-neutrally distributed OTUs. To

accomplish this, samples belonging to the same age group were first pooled, and OTUs from this pool were subsequently sorted into three partitions depending on whether they occurred more frequently than ("above" partition), less frequently than ("below" partition), or within ("neutral" partition) the 95% confidence interval of the neutral model predictions. Each partition was then treated as a distinct community sample for further analysis, resulting in 21 total partitions (3 per each of the 7 age groups). To facilitate comparisons among partitions, each partition was rarefied to an equal number of OTUs corresponding to the number of OTUs in the smallest partition, unless otherwise noted.

Diversity and taxonomic analysis

To quantify the variation in phylogenetic composition we calculated pairwise unweighted UniFrac distances among neutral and non-neutral metacommunity partitions (Lozupone and Knight, 2005). Differences in this distance among groups was assessed by permutational multivariate analysis of variance (PerMANOVA) using 1000 random permutations, while differences in the degree of variation within groups was assessed by an analysis of variance (ANOVA) of average distance to centroid within groups (multivariate homogeneity of groups dispersions test; Anderson *et al.*, 2006). Non-metric multidimensional scaling (NMDS) of UniFrac distances was performed to visualize difference among neutral and non-neutral partitions. Calculation of PerMANOVA, multivariate homogeneity of groups dispersion test, and NMDS were performed in R using the *vegan* package (Oksanen *et al.*, 2016).

In order to identify microbial taxonomic groups that distinguish neutral from non-neutral partitions of the metacommunity, we performed logistic regression with partition

type (above, below, or within the neutral prediction) as a predictor and the presence or absence of each taxon as a binary response variable. To determine the significance of this relationship, we compared the deviation of the fitted regression model with that of an empty null model (chi-square test).

An indicator taxa analysis was performed to identify OTUs associated with either fish or tank environment (i.e. water, surface, and food) samples. Each OTU was assigned an indicator value based on their abundance and occurrence frequency in either intestinal or environmental samples; OTUs found frequently in high abundance in one sample type but not the other would have a high indicator value for that sample type (Dufrêne and Legendre, 1997). Significance of this association with sample type was determined by comparing the observed value with the values from 1000 random permutations. Calculation of the indicator values and probability for OTUs was performed in R. Since a full set of environmental data was only available for the 4, 10, and 75 dpf age groups, this analysis was performed only for those time points.

Phylogenetic sampling theory

To further examine and compare the phylogenetic structure of neutral and non-neutral partitions of the observed communities, we employed a phylogenetic sampling theory that analytically predicts the phylogenetic diversity in a local community assuming random sampling from the phylogenetic tree of the metacommunity (O'Dwyer *et al.*, 2012). Observed measures of phylogenetic diversity for individual samples can be compared to these predictions to determine the degree to which communities appear random with respect to phylogeny as opposed to over-dispersed or clustered. If the

observed phylogenetic diversity is greater than the expected diversity, then we consider the community to be phylogenetically over dispersed, meaning that distantly related taxa were more likely to be sampled than closely related taxa. When the observed phylogenetic diversity is less than the expected diversity, then we consider the community to be phylogenetically clustered, meaning that closely related taxa were more likely to be sampled (Horner-Devine and Bohannan, 2006).

Implementation of the phylogenetic sampling theory was performed in R using methods described in O'Dwyer *et al.* (2012) and the *picante* package (Kembel *et al.*, 2010). Phylogenetic diversity was defined as the sum of the total phylogenetic branch length for a sample (Faith, 1992). Random sampling of the regional phylogenetic tree was modeled by binomial sampling. A strength of this approach is that it can be used to compare samples of unequal sizes. As such, this analysis was applied to un-rarefied data, and differences between observed and expected phylogenetic diversity were compared by calculating and comparing standardized deviations, or z-scores, for each partition.

Results

Relative importance of neutral processes decreases over host development

Overall, the frequency with which microbial taxa occurred in individual communities was well described by the neutral model (Figure 5; Supplemental Figure 7). However, the fit of the model varied over host development, and was negatively correlated with host age (Spearman's *rho* = -0.93, $p = 0.007$; Figure 6a). In all cases the neutral model outperformed a binomial distribution model, suggesting that the processes of passive dispersal and ecological drift have an impact above and beyond just random sampling of the source community (Figure 6b). Overall, estimated migration rates tended to be higher in younger than older fish, suggesting that communities become increasingly dispersal limited with age (Spearman's *rho* = -0.86, $p = 0.02$; Figure 6c).

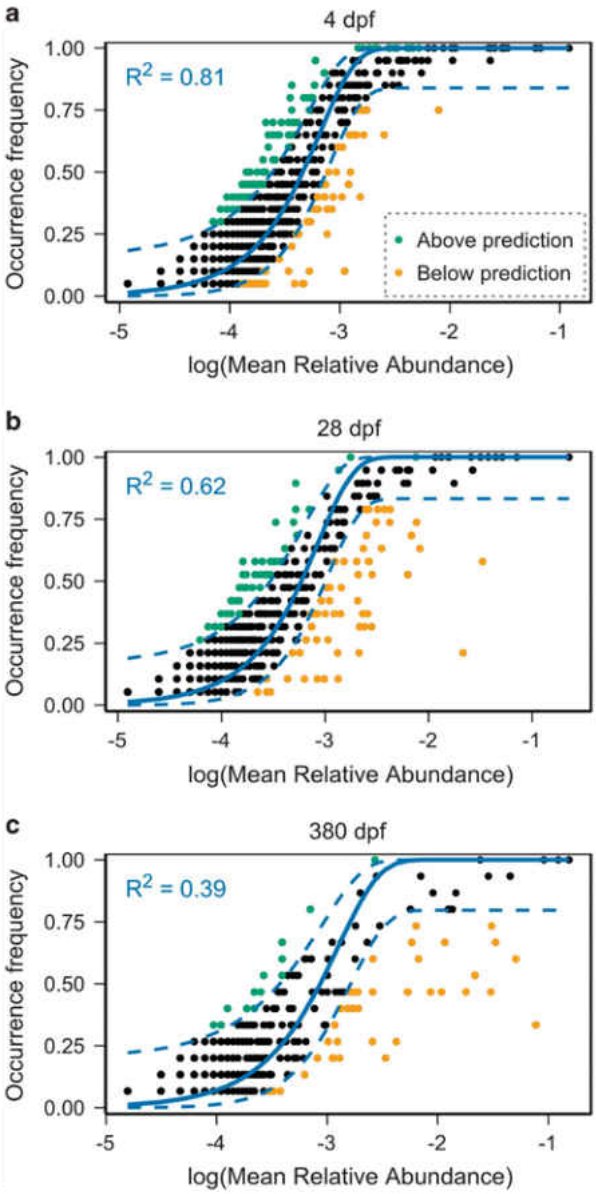


Figure 5: Neutral model fit. The predicted occurrence frequencies for 4 (**a**), 28 (**b**), and 380 (**c**) dpf zebrafish communities representing larval, juvenile, and adult developmental stages respectively. OTUs that occur more frequently than predicted by the model are shown in green while those that occur less frequently than predicted are shown in orange. Dashed lines represent 95% confidence intervals around the model prediction (blue line).

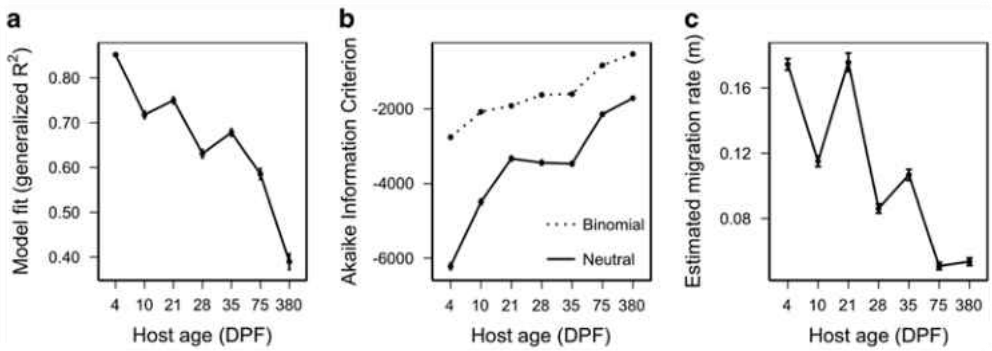


Figure 6: Neutral model fit decreases over host development. The goodness-of-fit of the Sloan neutral (**a**), comparison of the maximum likelihood fit of the neutral and binomial models (**b**), and the estimated migration rate (**c**) for zebrafish associated communities.

Deviations from neutral predictions are ecologically distinct

For any age group of fish, there were a number of microbial taxa that occurred more or less frequently than predicted by the model given their overall abundance in the metacommunity (points above, in green, or below, in orange, the line in Figure 5). We would expect points that differ significantly from the neutral prediction to be indicative of taxa that are actively being selected for or against by the host. Specifically, points above the prediction represent taxa that are found more frequently than expected, suggesting that they are actively being maintained and selected for by the host, while points found below the prediction represent taxa found less frequently than expected, suggesting that they are either selected against by the host or are especially dispersal limited. We expect that these selective processes should be reflected in the taxonomic and phylogenetic composition of taxa that deviate from the neutral prediction. We tested this hypothesis explicitly and examined how these differences may be informative of the overall ecology of the intestinal community.

Taxa found above, below, or within the prediction of the neutral model formed phylogenetically distinct partitions of the total metacommunity. For each age group, we

separated the metacommunity into three partitions comprised of those OTUs found above, below, or not significantly different from the neutral prediction and calculated the phylogenetic dissimilarity among partitions (Supplemental Figure 8). We found that partitions clustered strongly based on whether and how they deviated from the neutral prediction (i.e. above, below, or within the neutral prediction) across host age (PerMANOVA $r^2 = 0.19$, $p < 0.001$; Figure 7a). Thus, the phylogenetic composition of the sub-groups that diverge from neutral patterns remain relatively similar across host development, despite the composition of communities as a whole changing (Stephens *et al.* submitted). Across age groups, non-neutral partitions of the metacommunity were also much more homogeneous than the neutral partitions, (ANOVA, $p < 0.01$; illustrated by the spread of points in Figure 7a). A possible consequence of the heterogeneity is that we identified very few taxonomic groups that strongly distinguished neutral partitions. Partitions above the neutral prediction were most strongly distinguished by the presence of Fusobacteria ($p < 0.001$) and γ-Proteobacteria ($p = 0.022$), in particular the families *Enterobacteriaceae* ($p = 0.003$) and *Aeromonadaceae* ($p < 0.001$), while partitions below the neutral prediction were distinguished by the presence of Actinobacteria ($p = 0.004$), Bacilli ($p = 0.004$), and Clostridia ($p = 0.031$) and the genera *Lactobacillus* ($p = 0.004$), *Staphylococcus* ($p = 0.037$), and *Stenotrophomonas* ($p = 0.012$).

The taxa comprising neutral and non-neutral partitions of the metacommunity also tended to be associated with different environments. We first performed an indicator taxa analysis to identify OTUs that were significantly associated with either intestinal or environmental (tank water, surfaces, or food) samples in our full dataset. We then compared the proportion of OTUs significantly associated with fish to those significantly

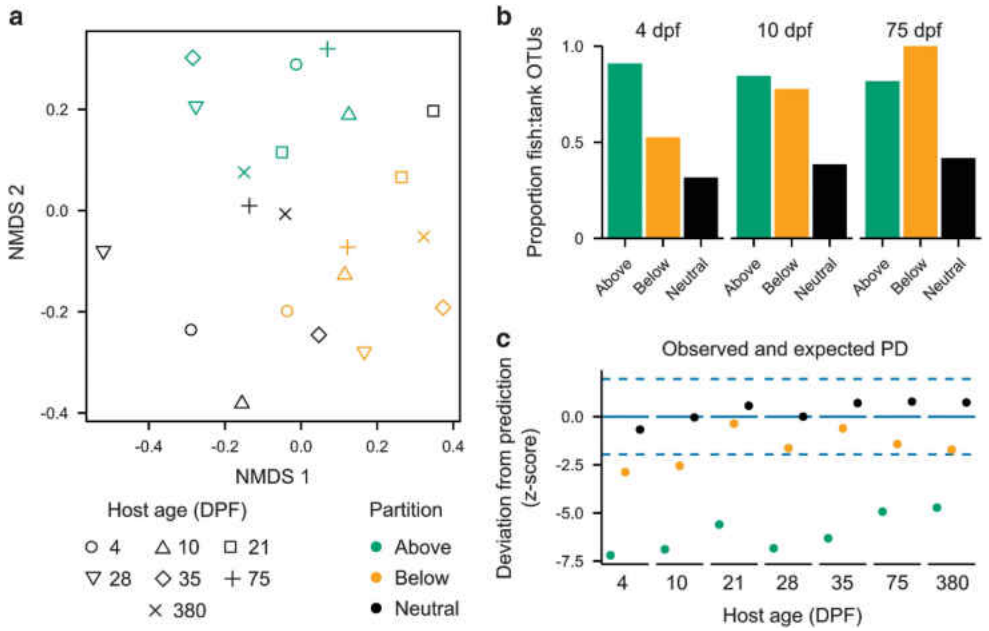


Figure 7: Neutral and non-neutral partitions of the metacommunity are compositionally and phylogenetically distinct. For each age group, communities were pooled and OTUs were then divided into separate partitions based on whether they were consistent with (in black) or deviated above (in green) or below (in orange) the neutral prediction (color coding is consistent for all panels). (**a**) Non-metric multidimensional scaling ordination based on UniFrac distances. (**b**) The proportion of fish associated to tank associated OTUs in each partition following an indicator taxa analysis. Results are shown for 4, 10, and 75 dpf fish only as environmental samples were not available for the other time points. (**c**) The standardized difference, in units of standard deviations (z-score), between observed and expected phylogenetic diversity assuming random sampling for each partition. Solid blue lines represent the expected phylogenetic diversity for each age group while dashed lines represent 95% confidence intervals. The distance points are either above or below the line represent the degree to which those partitions are phylogenetically over-dispersed or clustered respectively.

associated with the tank environment in each partition and found that this proportion was much higher above and below than it was within the model's prediction (Figure 7b). This pattern was consistent across host development. In other words, non-neutral partitions of the metacommunities were more likely to be comprised of microbial taxa that were associated with zebrafish, while taxa largely associated with the tank environment were more likely to be neutrally distributed across fish intestinal communities.

Finally, we found that non-neutral partitions were phylogenetically clustered with respect to the metacommunity as a whole. For this we calculated the phylogenetic diversity of each partition and compared these observed values to the phylogenetic diversity expected if taxa were sampled randomly with respect to phylogeny (O'Dwyer *et al.*, 2012). As expected, the observed phylogenetic diversity for neutrally distributed groups was consistent with the random prediction across host development. In contrast, the phylogenetic diversity of the partitions that deviated above the neutral prediction was consistently less than expected, indicative of phylogenetic clustering, while those partitions below the neutral prediction were phylogenetically clustered in larval fish but became less so as the hosts aged (Figure 7c). Assuming that more closely related microorganisms are on average more ecologically or functionally similar than more distantly related ones (Burns and Strauss, 2011), this result reinforces the conclusion that taxa which deviate from the neutral prediction, particularly those more widespread than expected, are portions of the microbiota that are more likely to be actively selected (for or against) by the host.

Discussion

The neutral model used in this study was able to predict the microbial distributions across communities by incorporating only the effects of random dispersal and demographic processes. Even in adult zebrafish, where the fit of the model was relatively poor compared to the younger fish, the distribution of OTUs in the metacommunity still followed the same basic trend of abundant taxa being widespread, consistent with neutral theory (Supplemental Figure 7). These findings illustrate an

important point which is often ignored: not all of the variation among host-associated microbial communities need be the result of differences among hosts or associated microorganisms. On the contrary, neutral processes of drift and dispersal are powerful enough on their own to generate a large amount of diversity both within and among hosts, and these processes can explain a significant portion of the structure of communities observed in this study. This is not to say that neutral processes are the only important factors, but they can act alongside and may even swamp the effects of non-neutral forces. These results also indicate that in addition to local ecological factors (e.g. the environment of a zebrafish intestine, differential competitive fitness among microorganisms, etc.), host-associated microbial communities are heavily influenced by ecological dynamics occurring outside of an individual host at a broader scale.

While the model's general success highlights the potential importance of neutral processes, it is also useful as a null model to identify the conditions under which the model's predictions fail, which can lead to a better understanding of specific additional factors structuring these communities. Within each age group there were a number of microbial taxa whose distributions deviated from neutral predictions. These taxa were not randomly distributed throughout the total metacommunity, implying they are distinct in ways that are ecologically informative. The taxa whose deviations from the neutral pattern led them to be more widespread than expected are likely taxa that are specifically adapted to, and selected by, the host environment. This is supported by the dominance of intestinal associated OTUs within non-neutral partitions and is consistent with these partitions being phylogenetically clustered, suggesting the host habitat selects microbial taxa based on a specific set of phylogenetically conserved traits (Figures 7b and c).

Likewise, abundant taxa that occurred less frequently than expected may be characteristic of "invasive" microorganisms and potential pathogens that are selected against by the zebrafish hosts overall, but are nevertheless able to proliferate in a few susceptible individuals. If true, this would explain why these taxa were more likely to be significantly associated with fish despite having distributions suggesting that they are being selected against. This is in contrast to neutrally distributed taxa, which were more likely to be associated with exogenous environmental tank samples as well as exhibiting greater phylogenetic variation and diversity across age groups. Such patterns suggest these neutrally distributed taxa are less likely to be specifically adapted to the host and their presence in any given community is largely the result of their abundance in the surrounding metacommunity and source pool. It is worth emphasizing, however, that this does not mean that these taxa are functionally unimportant or even that they are not interacting intimately with their hosts. Rather the host environment is not differentially selecting them, and consequently their distributions are the result of neutral dispersal and drift.

As hosts aged the ability of the neutral model to predict the distribution of associated microbial taxa decreased, indicating that neutral processes become relatively less important as the host ages (Figure 6a). We suspect this pattern is largely the result of the development of the host. The 4 dpf time point, for example, occurs shortly after the intestinal tract of the zebrafish is fully opened and colonized by bacteria but before the fish develops an active adaptive immune response (between 21 and 35 dpf) and reaches sexual maturity (between 35 and 75 dpf). It is also probable that husbandry changes over the course of the experiment had an impact on this pattern. The strongest evidence of

neutral dynamics occurred before the fish began eating (4 dpf) and while the fish were housed in nursery tanks unconnected to the main facility water system (from 4 to 21 dpf). At 21 dpf fish were not only moved from nursery tanks to the main facility system tanks, but also had their diet significantly changed (see methods). These changes to the host's physiology and environment gradually accumulate over development, and likely differentiate the ability of microbial taxa to establish and thrive within them. While it is difficult to disentangle whether the observed patterns are driven mostly by developmental or husbandry changes, we note that the decrease in the fit of the model continues between 28 and 380 dpf, during which time the zebrafish continue to develop (see above), but their housing conditions remain unchanged. The decrease in the fit of the model was accompanied by a decrease in the estimated migration rate, which suggests that these changes in the hosts may also decrease the ability of microorganisms to disperse into and among hosts. This is further supported by our previous observation that communities associated with 4 dpf and 10 dpf fish were more similar to environmental communities than those associated with the older 75 dpf fish (Zac Stephens *et al.*, 2016), as well as the observation that within-host diversity decreased over the same time span, which is a predicted consequence of decreased dispersal rates (Cadotte, 2006).

The patterns of neutral assembly in the zebrafish intestinal microbiota described here are consistent with and provide possible explanations for observed patterns in human associated microbial communities. In general, large-scale studies of human intestinal microbiota have revealed high variation in community composition, both across individuals and within individuals over time (Costello *et al.*, 2009; The Human Microbiome Project Consortium *et al.*, 2012; Yatsunenko *et al.*, 2012). Often, this

variation is not easily explained by measured host factors, suggesting that much of it might be explained by neutral assembly processes. Our observation that communities associated with young fish were the most neutrally assembled could also explain observations that variation is greater among communities associated with young humans (Kurokawa *et al.*, 2007; Palmer *et al.*, 2007; Yatsunenko *et al.*, 2012), the general variable nature of infant microbiota over time (Koenig *et al.*, 2011), and observations that the infant microbiota is heavily influenced by exogenous microbial communities, specifically those of the mother (Dominguez-Bello *et al.*, 2010; Funkhouser and Bordenstein, 2013).

These results may also be extended more broadly beyond animal associated communities. Using a similar conceptual framework to the one used here (multiple local communities sampling from a broader metacommunity), Jabot *et al.* (2008) found that the distribution of young saplings in a tropical forest was better fit by a neutral model than that of older trees. Likewise, Dini-Andreote *et al.* (2015) found that the relative importance of stochastic processes decreased over the succession of microbial salt marsh communities. The consistency of this pattern across these different systems may be indicative of more fundamental ecological processes. Even if the structure of communities is ultimately determined by differential selection, many communities in nature may exist in transitory, non-equilibrium states such that these selective processes do not have the opportunity to fully play out and manifest their effects (Manceau *et al.*, 2015).

Despite extensive research on microbial communities associated with animal hosts, it has remained difficult to explain the high levels of variation in these systems. We

addressed this question by adopting a framework that recognizes that these questions are ecological in nature and can be addressed through the use of established ecological theory. Because this framework is grounded in ecological theory, it provides hypotheses that can be tested in an explicit manner. For example, it would be interesting to see whether the neutral and non-neutral partitions of the metacommunity are physically delineated in the intestine, wherein we might expect neutral taxa to be found in the lumen while deviations from neutral patterns might be more intimately attached to the epithelial layer where interactions between host and bacterial cells may be more likely to occur. Additionally, it is possible that non-neutral behavior in these communities is driven by differences among taxa in dispersal rates, in which case partitioning the communities on the basis of differences in immigration rates would likely improve neutral predictions (Janzen *et al.*, 2015). It might also be fruitful to compare the neutral patterns seen in healthy hosts to those seen in diseased, infected, or diet-altered individuals which we predict will be characterized by deviations from neutral predictions. Similarly, we predict that infectious or pathogenic microorganisms could be identified by their deviations from neutral predictions, occurring much less frequently than expected given their relative abundance in a metacommunity.

Ultimately, one of the goals of studying host-associated communities is to better understand how they might be altered or manipulated to improve health and prevent disease. At their core, our results demonstrate a relationship between the abundance of a microorganism and how widespread that microorganism is in a population of hosts. In other words, the distribution of microorganisms in these systems is the result of both local factors specific to individual hosts and those processes occurring at a broader

metacommunity scale linking multiple hosts. Attempts to manipulate a host's microbiota must therefore focus on understanding not only the communities within an individual host, but also on the communities of microorganisms present around them.

BRIDGE

Neutral theory has had a controversial, yet nevertheless substantial, impact on the field of community ecology. The assumption of ecological equivalence of species intuitively seems unrealistic, yet the strength of neutral theory is that it highlights the importance of processes that are otherwise difficult to study in natural communities, random dispersal and ecological drift, and it can serve as a useful null model to infer the relative importance of ecological selection. In Chapter III, I showed that neutral processes were sufficient to explain a large amount of the variation in the distribution of microbial taxa across zebrafish hosts, but that the relative importance of neutral processes decreased as hosts accumulated developmental changes, suggesting that the impact of selection by the host environment increases. Thus, the assembly of these communities is a balance between neutral and non-neutral processes, and this balance shifts as host factors change. In Chapter IV, I attempt to explore the balance between dispersal and host factors further by experimentally testing the hypothesis that inter-host dispersal can mediate and override the effects of host factors, namely the innate immune system.

CHAPTER IV

INTER-HOST DISPERSAL MEDIATES THE ASSEMBLY OF MICROBIAL COMMUNITIES ASSOCIATED WITH WILD-TYPE AND IMMUNE DEFICIENT ZEBRAFISH

This experiment was designed by me and carried out by both myself and Meghna Agarwal under my direction. The analysis of data and writing of the manuscript was performed by me. Brendan Bohannan and Karen Guillemin filled advisory roles throughout the entire process and Brendan Bohannan was the principal investigator of the study.

Introduction

The microbial communities associated with animal hosts are highly dynamic, variable, and malleable, leading to the perception that the rules governing their assembly are idiosyncratic (Bashan *et al.*, 2016). However, unlike many other attributes of an animal's biology that are important to its health and fitness, the microbiota is subject to migration both from external abiotic sources and transmission from other hosts. If the influence of this dispersal among hosts is substantial, then a comprehensive model of determinates of community structure must include consideration of not just the local factors associated with individual hosts, but also the composition and structure of the population of hosts to which they belong. Such a model would ultimately allow us to better predict and manipulate animal associated communities to an extent that has thus far

eluded researchers, but building one will first require a careful and explicit test of the effects of dispersal that is guided by a broader predictive framework.

Research on animal associated communities has increasingly utilized frameworks from general ecological theory to guide experiments and interpret patterns (Costello *et al.*, 2012). Much of the modern conceptual synthesis of community ecology emphasizes the role of dispersal in the assembly of communities and the maintenance of diversity (Vellend, 2010), from island biogeography theory (MacArthur and Wilson, 1967), to neutral theory (Hubbell, 2001), and metacommunity theory (Leibold *et al.*, 2004). Metacommunity theory is especially focused on dispersal among multiple discrete "local" communities, and is thus particularly well suited to describing host-microbe systems, where hosts act as environments that are home to and select local communities of microorganisms embedded in a broader "metacommunity" associated with a population of hosts that are linked by inter-host dispersal. In such a system, dispersal can allow for the persistence of species in hosts in which they would otherwise go extinct by immigrating from hosts or environments in which they are abundant (i.e. "mass effects"), or through tradeoffs between dispersal and competitive ability (Livingston *et al.*, 2012). Empirically, dispersal has been shown to be an important determinate of the structure and functioning of bacterial communities both experimentally (Lindström and Östman, 2011; Declerck *et al.*, 2013; Zha *et al.*, 2016) as well as in naturally occurring communities (Lear *et al.*, 2014; Martiny *et al.*, 2011; Yeh *et al.*, 2015).

The incorporation of dispersal into studies of microbial communities associated with animal hosts has been limited, however there are several emerging observations and phenomena that suggest it is likely important. At the broad scale, for example,

biogeographic patterns in the distribution of commensal microorganisms have begun to be observed for communities associated with natural populations of humans and mice (Linnenbrink *et al.*, 2013; Martínez *et al.*, 2015). Furthermore, for pathogenic microorganisms at least, there is a long history of work showing how dispersal and transmission among hosts influences not only the spread but the diversity and function of pathogen, though this work has largely been limited to individual species and surprisingly little of these advances have been applied to the study of communities of commensal (i.e. non-pathogenic) microorganisms (Faith *et al.*, 2015). Despite these preliminary but converging lines of evidence, experimental studies explicitly testing the importance of dispersal in microbial communities associated with animal hosts remain rare.

Many manipulative experiments, however, have been indirectly testing the effects of dispersal on the microbiota, where so called "cage effects" routinely explain significant amounts of variation in the composition of communities associated with laboratory animals housed in separate units, as well as in phenotypes known or suspected to be mediated by the microbiota (Arthur *et al.*, 2012; Deloris Alexander *et al.*, 2006; Hildebrand *et al.*, 2013; McCafferty *et al.*, 2013; Campbell *et al.*, 2012). Interestingly, experiments studying the innate immune system have often shown that cohousing of healthy and immune-deficient animals can transfer phenotypes associated with immune pathway mutants, including increased inflammation, colitis, and obesity-related metabolic abnormalities (Brinkman *et al.*, 2013; Vijay-Kumar *et al.*, 2010; Wen *et al.*, 2008; Zenewicz *et al.*, 2013). Similar investigations of the link between innate immunity and associated microbial communities have had conflicting or inconclusive results, with some finding little to no effect of innate immune pathways on community composition or

diversity, especially in cases were both wild type and immune-deficient animals are cohoused or from the same litter (Dimitriu *et al.*, 2013; Elinav *et al.*, 2011; Loh *et al.*, 2008; Albert *et al.*, 2009; Ubeda *et al.*, 2012). Taken together, these patterns not only reinforce the notion that dispersal has an important impact on the assembly of host-associated microbial communities, but further that an explicit accounting of dispersal in these systems could resolve their inconsistencies and idiosyncratic behavior.

The lack of studies on dispersal in host-microbe systems can in part be attributed to the difficulties in both measuring and manipulating dispersal among animal hosts. Studies involving humans and other natural populations of animals cannot control inter-host transmission for extended periods of time and are thus limited to natural experiments. Even in the popular mouse model system, it can be impractical to separately house large numbers of individuals and identifying the medium or vectors through which inter-host dispersal occurs can be difficult. To overcome these challenges, zebrafish, *Danio rerio*, together with the microorganisms inhabiting their intestines provide an ideal model system in which to study how inter-host dispersal interact with individual host factors to shape the composition of associated microbial communities. Doing so allows us to raise large numbers (>100) of individuals in housing conditions that control the transmission of microbes among fish as well as sample and characterize the microbial communities associated with their environment (i.e. tank water and food). The genetic malleability of zebrafish also enables us to alter host characteristics in order to set up populations consisting of varying levels of host heterogeneity, specifically for factors which we suspect interact directly with microorganisms such as innate immune function.

Finally, the ease with which zebrafish embryos can be made "germ-free" means we can easily control the initial microbial exposure and source pool (Pham *et al.*, 2008).

We performed an experiment using zebrafish to determine how inter-host dispersal interacts with individual host factors to shape the communities of microorganisms associated with the zebrafish intestine. We were particularly interested in such interactions with the host innate immune system, given its ability to directly sense and interact with commensal microorganisms, its overall importance to host health, and the previously described inconsistent effects on associated microbial communities. Therefore, our first goal was to assess the impact of innate immunity on the composition and diversity of communities in the zebrafish intestine by comparing communities associated with wild type (WT) zebrafish that possess a fully functional immune system to those associated with immune-deficient *myd88*$^{-}$ mutant zebrafish. The myeloid differentiation primary response gene 88, or MyD88, encodes a universal adapter protein in the Toll-Like Receptor (TLR) pathway and is responsible for activating several immune responses in response to signaling from the microbiota including the production of pro-inflammatory cytokines and antimicrobial peptides and the detoxification of the bacterial product lipopolysaccharide (Bates *et al.*, 2006; Karmarkar and Rock, 2013; Janssens *et al.*, 2002). Thus, while mutants lacking MyD88 still have some innate immune functionality through other pathways, overall the ability of zebrafish to mount a normal innate immune response is compromised. However, given the inconsistencies among previously reported studies of the effects of host immunity, including the TLR/MyD88 pathways on the microbiota, our ultimate goal was to ask whether dispersal among hosts was a powerful enough process to mediate, or even overwhelm, the effects

of the host innate immune system. Here we show that not only does the presence or absence of dispersal among zebrafish have a substantial impact on microbial community composition and diversity, but inter-host dispersal also overwhelms the effects of the MyD88 pathway such that the impact of *myd88* deletion on intestinal microbial communities is only apparent in the absence of inter-host dispersal.

Materials and Methods

Zebrafish husbandry

To study the effects of innate immunity on the zebrafish intestinal microbiota, we employed wild type (WT) AB/Tübingen heterozygous fish with a fully functional immune system and isogenic immunocompromised mutant *myd88*$^{-}$ zebrafish. Because there is no practical or non-invasive method to reliably distinguish *myd88*$^{-}$ from WT zebrafish embryos, the embryos for each genotype were generated from two crosses of homozygous parents. To eliminate potential maternity effects, standard gnotobiotic zebrafish protocols were used to make the embryos "germ-free" (free of microorganisms) prior to beginning the experiment.

Without a complete understanding of the modes and vectors by which microorganisms disperse from one zebrafish host intestine to another, it is difficult to know how to manipulate inter-host dispersal. We therefore took an implicit approach, and either allowed dispersal to occur among hosts or prevented all inter-host dispersal. This was done by cohousing or isolating zebrafish in glass Erlenmeyer flasks, such that microorganisms could disperse among hosts in the same flasks, but not among hosts in different flasks. Beginning as germ-free embryos, zebrafish were raised in flasks alone

("solitary"), or cohoused with ten total zebrafish of the same genotype ("segregated"), or five of each genotype ("mixed"), per flask (Figure 8a). Initially, 20 fish of each genotype were raised alone in solitary flasks, 40 fish of each genotypes were raised across four replicate segregated cohoused flasks, and 20 fish of each genotype were raised across four replicate mixed cohoused flasks, resulting in a total of 160 zebrafish at the beginning of the study. The volume of embryo media and size of flask was scaled to the number of fish, i.e. 50mL of embryo media in a 125mL Erlenmeyer flask for solitary conditions and 500mL of embryo media in a 1L Erlenmeyer flask for cohoused conditions. Doing so allowed the density of fish and the surface area to volume ratio of liquid in the flask to be equivalent across housing conditions. In order to maintain water quality, 75-90% of the embryo media in each flask was removed and replaced with fresh, but not sterilized, embryo media every day. During this time, the majority of food debris and zebrafish feces, as well as any dead fish carcasses, were removed as well. Once the zebrafish fully hatched from their chorions (by 4 days post fertilization or "dpf"), fish were fed live rotifers to a concentration of 20 individuals per mL, followed by the addition of live brine shrimp beginning at 10 dpf once per day. The zebrafish were raised in this manner until 21 dpf, at which point the experiment was ended.

Sampling and DNA/RNA extractions

At 21 dpf, the juvenile zebrafish were euthanized and dissected in order to sample their intestinal communities by 16S rRNA gene sequencing, as well as to characterize their innate immune response by qPCR of two genes encoding innate immune cytokines: *il-1β* and *c3*. Each individual intestine was aseptically removed and placed in a sterile

2mL screw cap tube with 200mL nuclease-free water while the remainder of the zebrafish carcass was placed in a 2mL screw cap tube with 1mL of TRIzol (Life Technologies, Carlsbad, CA, USA). Both sample types were then immediately frozen in liquid nitrogen and stored at -80°C until DNA/RNA extractions were performed. In order to identify the genotype of mixed-cohoused zebrafish, mixed-cohoused samples were genotyped by PCR of the *myd88* gene.

DNA and RNA was co-extracted from intestinal samples using the MoBio PowerMicrobiome RNA Isolation kit (product number: 26000-50) with the addition of β-Mercaptoethanol (Sigma product number: M3148-25ML) using the manufacturer's suggestions for co-extraction of both DNA and RNA. Unfortunately, the resulting concentrations of RNA from intestinal samples were too low for reliable measurements of *il-1β* and *c3* expression by qPCR. Therefore we estimated systemic responses by measuring gene expression of *il-1β* and *c3* in RNA extracted from the remaining zebrafish carcasses. RNA was extracted from carcasses using a standard laboratory TRIzol extraction protocol.

The water from each flask was collected and individually passed through 2 μm cellulose nitrate filters in order to collect the microbial biomass for DNA extraction and subsequent community profiling. DNA was then extracted off of the filters using the MoBio PowerWater DNA Isolation kit (product number: 14900-50-NF) as per the manufacturer's instructions.

cDNA conversion and qPCR

Extracted zebrafish RNA was converted into cDNA using the Superscript IV reverse transcriptase kit (Invitrogen) following the manufacturer's instructions. qPCR assays were performed in 20 ul reactions with 20 ng of cDNA, and 400 nM gene-specific or control primers. Gene-specific primers were ordered from Eurofins Genomics with the following sequences: IL-1B: F: 5'CATCAAACCCCA ATCCACAG-3', R: 5'-CACCACGTTCACTTCACGCT-3'; C3: F: 5'-CGGACGCTG ACATCTACCAA-3', R: 5'-TCCAGGTCTGCTCT CCCAAG-3'. Primers for the housekeeping genes SDHA and ElF-1B (used to normalize the results) were ordered from PrimerDesign. All reactions were performed using a Bio-Rad CFX96 Real-Time PCR (qPCR) Thermocycler.

16S rRNA gene sequencing and processing

We characterized the microbial communities of individual samples via Illumina (San Diego, CA, USA) sequencing of 16S rRNA gene amplicons. The V4 region of the 16S rRNA gene was targeted using the following primers: 515F (5'-TATGGTAATTGTGTGCCAGCMGCCGCGGTAA-3') and 806R (5'-AGTCAGTCAGCCGGACTACHVGGGTWTCTAAT-3'). To generate amplicons that could be used for Illumina sequencing, we used a single-step PCR to add dual indices and adapter sequences to the V4 region of the bacterial 16S rRNA gene and generate paired-end 150 nucleotide reads on the Illumina HiSeq 2500 platform. The resulting 16S rRNA gene sequences were quality filtered and processed using methods implemented by FLASH (Magoc and Salzberg, 2011) and the FASTX Toolkit (Hannon Lab, 2010). Operational taxonomic units (OTUs) were defined *de novo* using 97% sequence

similarity in the USEARCH pipeline (Edgar, 2010). Read assembly, quality control, and OTU table building were done on the University of Oregon ACISS cluster.

Community analysis and statistics

Prior to analysis, OTU abundance tables were rarefied to 13700 sequences per sample and then Hellinger transformed in order to standardize counts (Legendre and Gallagher, 2001). We measured differences in community composition using the Canberra distance. In order to assess whether different treatments or host factors had a significant effect on community composition, we performed permutational multivariate analysis of variance (PerMANOVA) with 1000 random permutations using these distances. To measure the overall variation in community composition within groups, we performed a multivariate homogeneity of group dispersions test, which briefly measures the average distance of each community from the group centroid in multivariate space followed by analysis of variance (ANOVA) to assess the significance of differences among groups (Anderson *et al.*, 2006). Variance partitioning on community composition by host factors (standard length and *il-1β* and *c3* gene expression) was done by canonical redundancy analysis to measure both the unique and shared contributions of each host factor (Legendre, 2008). Calculation of the Canberra distance, PerMANOVA, Shannon diversity, multivariate dispersions test, and redundancy analysis were all performed in R (R Core Team, 2016) using the vegan package (Oksanen *et al.*, 2016). All other statistical analyses were performed in base R, while plots were made with the help of the ggplot2 package (Wickham, 2009).

Results

We raised two types of zebrafish, immunocompetent wild type (WT) zebrafish and immune deficient *myd88*$^{-}$ zebrafish, under three distinct housing conditions designed to either allow or restrict dispersal among hosts: solitary conditions in which each individual zebrafish was raise alone with no exposure to other individuals (i.e. no inter-host dispersal), or cohoused, either with other members of the same genotype only (segregated) or with members of both genotypes (mixed; Figure 8a). Like all jawed vertebrates, zebrafish also have an adaptive immune system, however the adaptive immune system in zebrafish does not become active until the fish are approximately 28 dpf (Lam *et al.*, 2004). Therefore, to isolate the effects of the innate immune system while still allowing the fish to grow and develop into juveniles, we raised fish to 21 dpf. At this point we euthanized the fish and characterized both the microbial communities associated with their intestines and those associated with their food and flask water. As one might expect given the importance of host immunity to defense against pathogens, immune-deficient *myd88*$^{-}$ fish had higher mortality rates, and notably, their mortality rates were higher in the cohoused treatments compared to the solitarily treatment (Figure 8b). Interestingly, mortality rates for WT fish were also higher when cohoused, especially when cohoused with *myd88*$^{-}$ fish. Because of this, by the end of the experiment the number of fish in each flask were no longer equal. However we did not observe a significant effect of the ultimate number of fish per flask on community composition within or across treatments (PerMANOVA: $p > 0.05$).

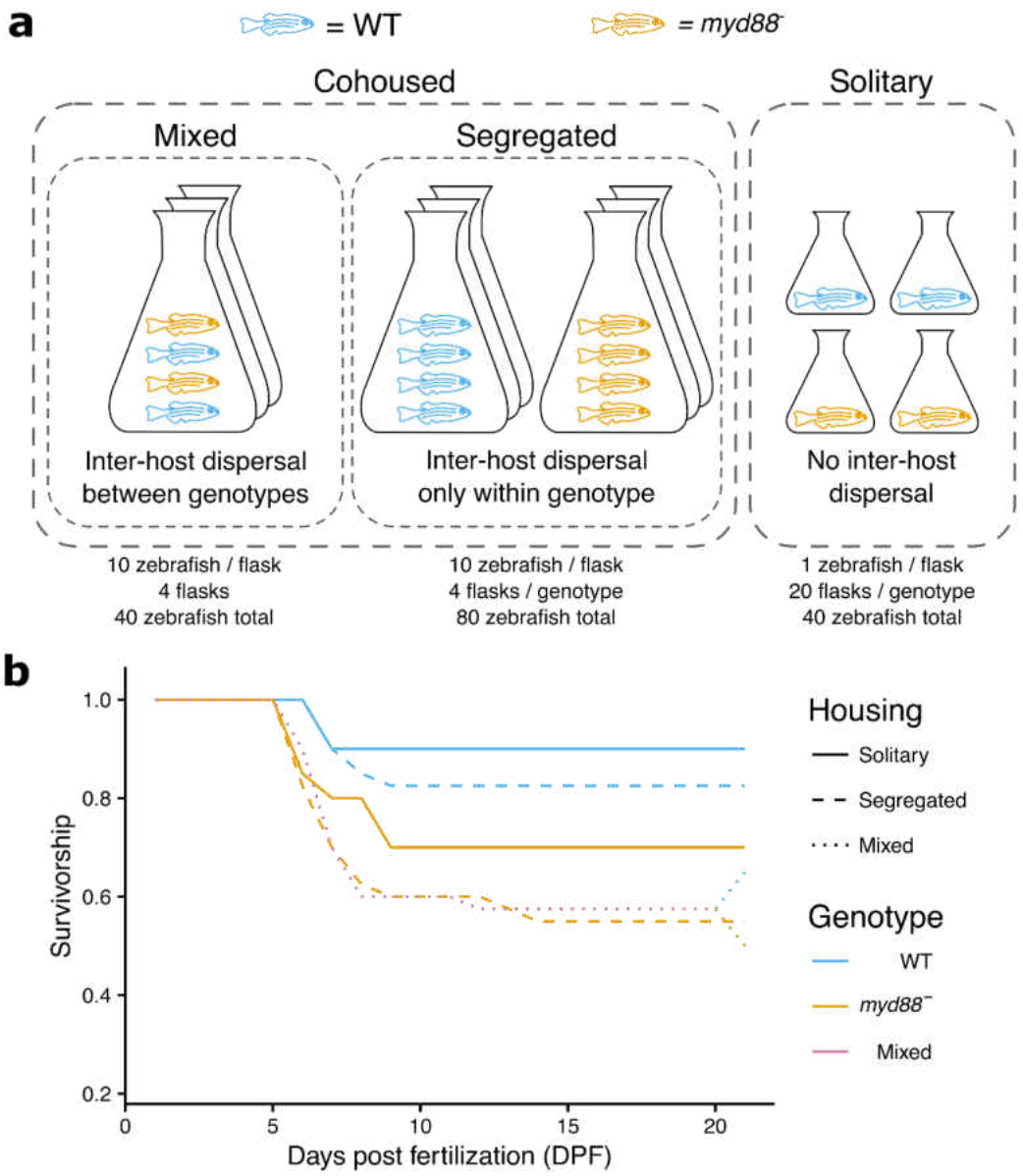


Figure 8: (**a**) Experimental design. WT and *myd88*- zebrafish were raised in one of three housing conditions in order to manipulate the degree of inter-host dispersal: housed alone with no inter-host dispersal (solitary), cohoused with only individuals of the same genotype (segregated), or cohoused with individuals of both genotypes (mixed). (**b**) Proportion of surviving individuals in each genotype by housing treatment over time (days post fertilization, d.p.f). For the mixed housing treatment, the genotype of individuals could not be determined until the end of the experiment, so the survivorship curve represents both genotypes until the last time point.

Overall, there was a significant but weak difference in the composition of communities associated with WT and *myd88*$^{-}$ zebrafish across the entire dataset (Table 3; Figure 9a). However, overall this effect was subtle, and there was no significant difference between genotypes in terms of either within-host diversity (ANOVA of the effect of genotype on Shannon diversity index: $p > 0.05$; Figure 9b) or the overall similarity of communities (Multivariate dispersion test; $p > 0.05$; Figure 9c). This changed once dispersal limitations among hosts were considered. Not only did the housing condition alone explain a greater amount of variation in community composition, but there was also a strong interaction between housing and genotype such that there was a much greater difference in microbial community composition between genotypes when hosts were raised solitarily compared to cohoused, either within or across genotypes (Table 3). This was true both in terms of overall composition as well as differences in diversity (ANOVA of effect of housing conditions on Shannon diversity index: F-statistic = 36.6, $p < 0.001$). Communities associated with solitary fish of both genotypes were significantly less diverse than cohoused communities and bore less of a similarity to communities in the surrounding flask water (Figure 9d).

The strong effect of housing conditions on community composition and diversity is consistent with our predictions of how inter-host dispersal would mediate the assembly of host-associated communities. To further investigate, we measured the relationship between individual host factors and community composition within each housing treatment. We hypothesized that dispersal among heterogeneous hosts would dilute the effects of local host factors, and therefore the relationship would be strongest for solitary hosts and weakest for mixed-genotype cohoused hosts. We first measured the standard

Table 3: Results of a PerMANOVA analysis on the effects of genotype and housing conditions on community composition. P-value calculated from a distribution of 1000 random permutations.

Factor	df	SS	MS	F-stat	R^2	P-value
Across housing treatments						
Genotype	1	0.007	0.007	7.192	0.053	< 0.001
Housing	2	0.018	0.009	9.569	0.140	< 0.001
Genotype × Housing	2	0.017	0.008	8.629	0.126	< 0.001
Within housing treatments						
Genotype-*Mixed*	1	0.004	0.004	2.790	0.054	0.001
Genotype-*Segregated*	1	0.014	0.014	3.391	0.151	0.010
Genotype-*Solitary*	1	0.067	0.067	34.108	0.577	< 0.001

length of each zebrafish (Figure 10a), which is known to be an overall indicator of fish development and health (Parichy *et al.*, 2009), and which we had previously shown was a strong predictor of intestinal microbial communities across zebrafish development (Stephens et al. 2016). To characterize the level of innate immune activity of each host, we also measured the transcriptional levels of two immune genes: one, *c3*, in the MyD88-independent complement pathway, and one, *il-1β*, in the MyD88-dependent pathway. As expected given their genotype, we found expression of *il-1β* to be lower in *myd88*$^{-}$ compared to WT hosts, while expression of *c3* was similar between the two genotypes (ANOVA: F-statistic = 13.9, $p < 0.001$ for *il-1β* and F-statistic = 0.11, $p = 0.74$ for *c3;* Figure 10b and c). It is noteworthy that despite having a strong effect on the microbiota, housing conditions had no clear effect on host innate immune response (ANOVA: $p > 0.05$ for both *il-1β* and *c3*). This led us to believe that inter-host dispersal mainly altered

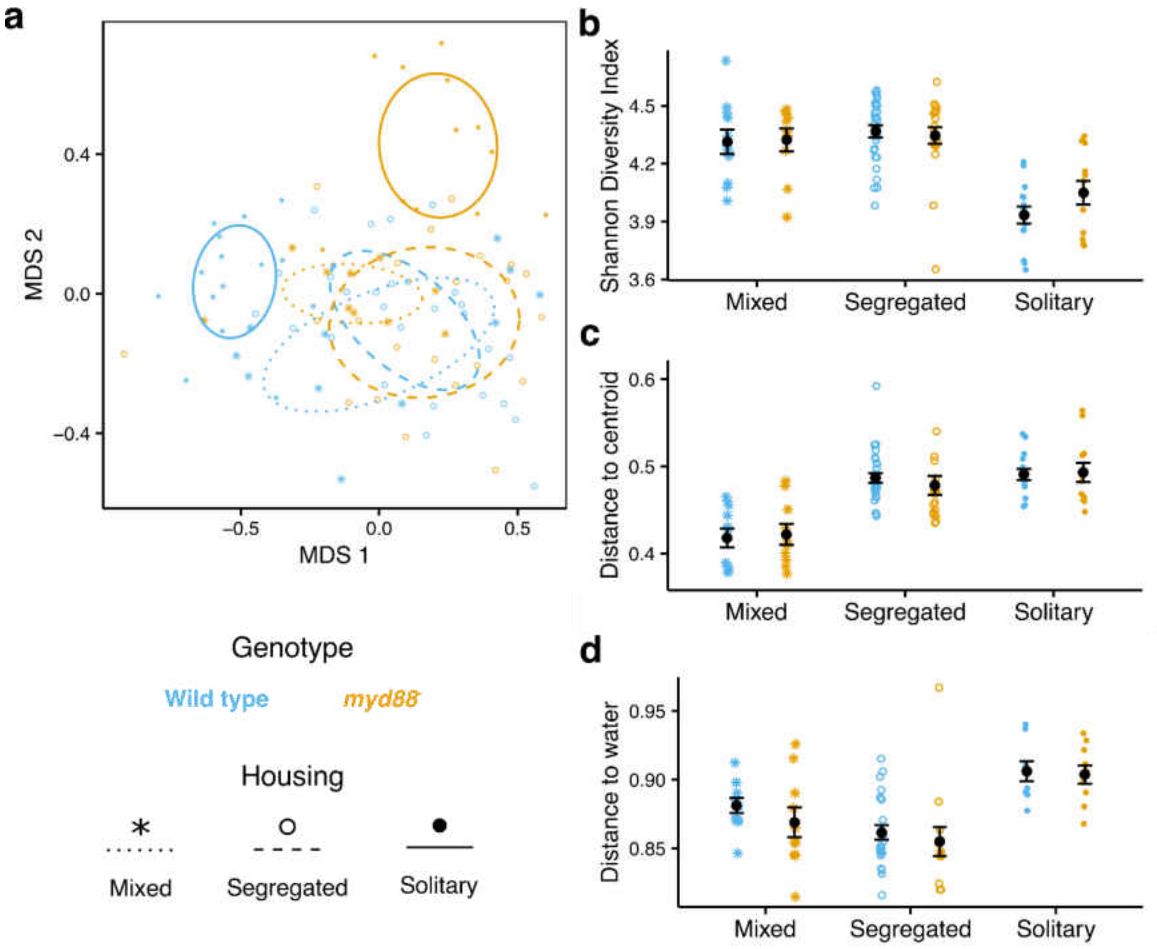


Figure 9: Effects of host genotype and housing conditions on the composition and diversity of associated microbial communities. **(a)** NMDS ordination of Canberra distances among individual intestinal communities. **(b)** Shannon diversity, **(c)** beta-dispersion, and **(d)** similarity to water communities. For **b**-**d**, blue and orange points represent values for individual WT and *myd88*$^{-}$ samples, black points represent group means with standard error bars.

the ability of microorganisms to track the local host environment rather than changing the host environment itself. To test this, we performed a redundancy analysis to determine the unique and shared contribution of each host factor to explaining the variance in community composition. Consistent with our hypothesis, a greater amount of variance in community composition was explained by host factors in solitary, but not segregated- or mixed- cohoused, WT and *myd88*$^{-}$ hosts (Figure 10d).

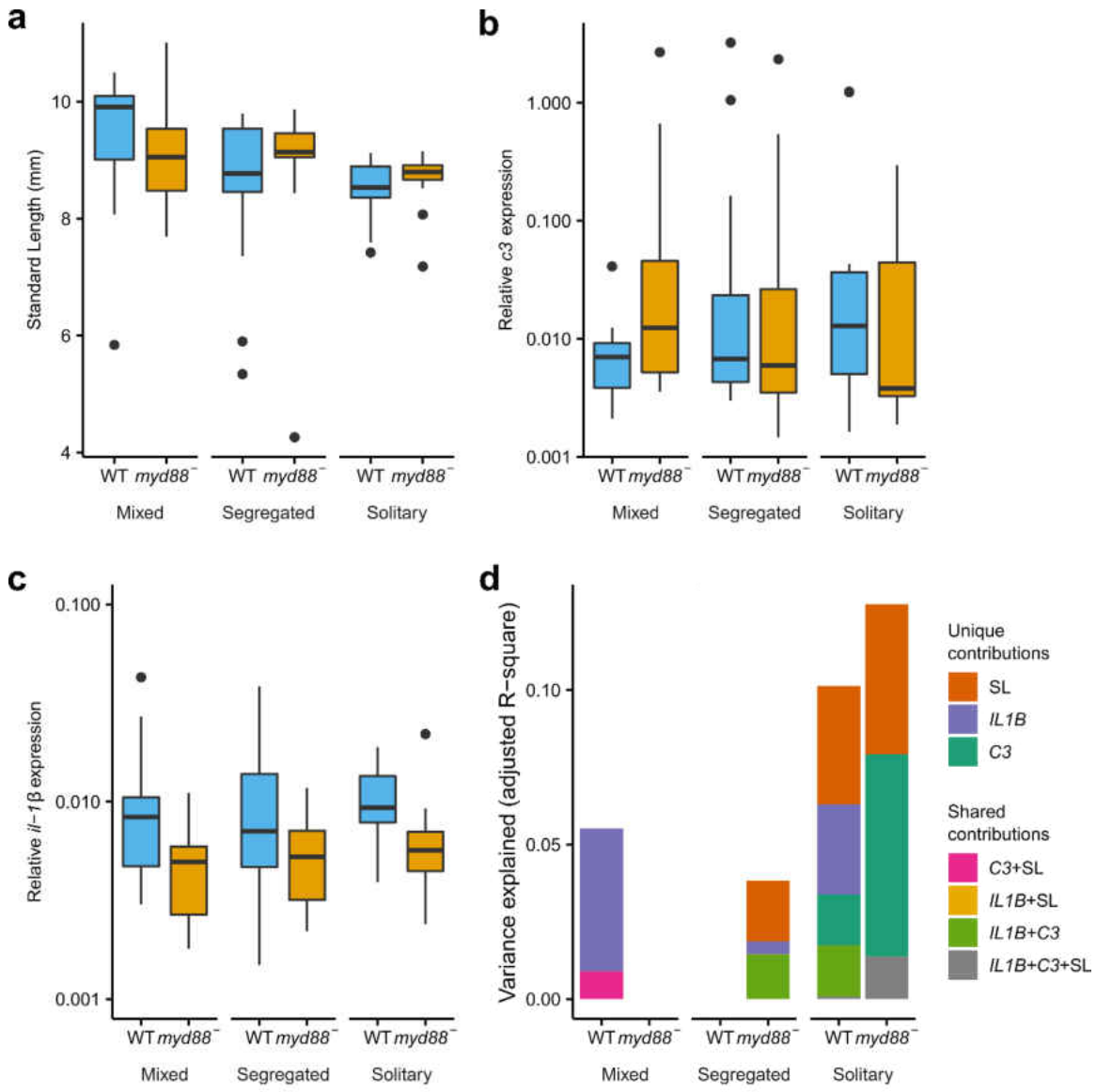


Figure 10: Relationship between host factors and microbial community composition. Observed standard length (SL) of zebrafish hosts (**a**), the relative expression of host *c3* (**b**) and *il-1β* (**c**) genes, and partitioned variance explained following a redundancy analysis (**d**). Shown are the adjusted R-square values for the unique and shared contribution of multiple host factors: *il-1β* expression (*IL1B*), *c3* expression (*C3)*, and standard length (SL). Negative adjusted R-square values are not shown (considered as null).

Interestingly, cohousing hosts had a similar effect on communities regardless of whether hosts were cohoused with only members of the same genotype or with members of both genotypes, such that there was no difference between segregated-cohoused and

mixed-cohoused hosts either within or across genotypes (PerMANOVA: $p > 0.05$). This suggests that the homogenizing effect of cohousing isn't just due to exchange among hosts diluting the effects of immune activity, but it is also likely that inter-host dispersal alone allows for the persistence of taxa that would otherwise be excluded from a system with only a single host. In support of this explanation, we found that the majority of taxa that occurred in solitary communities were also detected in their cohoused counterparts (86% and 88% for solitary WT and *myd88*$^{-}$ hosts respectively), while a much smaller proportion of taxa that occurred in cohoused communities were also detected in their solitary counterparts (59% and 64% for cohoused WT and *myd88*$^{-}$ hosts respectively).

Discussion

Dispersal in ecological communities is a notoriously difficult process to study, and this is no less true for the communities of microorganisms that colonize human and other animal hosts. This is in large part due to the technical and logistical difficulties in observing, measuring, and manipulating the movement of organisms in nature. It is perhaps for this reason that there are few empirical studies on the effects of dispersal on host-associated microbial communities, despite the fact that inter-host transmission is a key feature defining the fitness of pathogenic organisms as well as a large body of ecological theory and empirical work predicting its importance in communities in general. Our approach was thus to utilize a model system, the zebrafish gut microbiota, wherein we could manipulate inter-host dispersal to assess its relative importance compared to individual host factors. Doing so, we showed that not only is inter-host dispersal a strong determinate of the structure of the gut microbial communities, but it

can also mediate and even negate the effects of host factors, such as innate immunity, that directly interact with microorganisms. Specifically, inter-host dispersal weakened differences between immune-deficient and wild type hosts such that host innate immune activity was only able to predict community composition in the absence of inter-host dispersal.

The majority of the patterns in microbial diversity we observed were consistent with both predictions from metacommunity theory and empirical observations in other systems: namely lower diversity within individual solitary hosts, as well as dispersal limitations increasing the difference in community composition between host genotypes (Cadotte, 2006; Declerck *et al.*, 2013; Lear *et al.*, 2014). Taken together, these results provide some of the first experimental evidence that metacommunity theory provides an appropriate framework for the study of host-associated microbial communities, as has previously been suggested but not tested (Costello *et al.*, 2012). However, contrary to our hypothesis and general predictions from metacommunity theory, there was no difference in the variation in community composition among solitary hosts compared to cohoused hosts (Figure 9c). We suspect that this discrepancy might be due to the seemingly subtle but important distinction that our experiment was designed to only manipulate the degree of dispersal among hosts, but explicitly not the rate of dispersal into and out of each individual host, whereas conventional metacommunity models from which our predictions were derived generally assume dispersal only occurs among local communities. Thus, each individual solitary host is experiencing equal migration from a shared source (in this case, communities associated with exchanged embryo media and rotifer/brine shrimp cultures). This could lead to their homogenization compared to

cohoused hosts, which despite also sharing the same source pool, have additional metacommunity dynamics at play to introduce heterogeneity (e.g. colonization-competition tradeoffs). This would also explain why we observed a stronger relationship between local host factors (standard length and expression of inflammatory cytokine genes *il-1β* and *c3*) and community composition in solitary compared to cohoused hosts (Figure 3d), whereas the strength of environmental filtering often increases with dispersal rates in other metacommunity systems, presumably because of the increased probability that each species in the system will have the opportunity to colonize their preferred niche (Heino, 2013; Yeh *et al.*, 2015). This suggests that when cohoused, dispersal rates among hosts are high enough that the metacommunity dynamics are dominated by "mass effects", which refers to the persistence of species in local communities through immigration from communities in which they are abundant, rather than by dispersal limitations (Ng *et al.*, 2009).

At a practical level, these results are important in that they can inform both the interpretation and design of manipulative experiments on host-microbe interactions. At the very least, when reporting experimental results, researchers should explicitly describe the housing conditions used, which unfortunately is currently not always the case. Often cohousing is considered a control to ensure all individuals are exposed to the same source pool of microorganisms. While likely true, our results demonstrate that doing so can also have an impact on the dynamics of associated microorganisms and therefore potentially alter host-microbe interactions and ultimately the inferences that are made about the importance of various host factors. We also anticipate that there will be variation in the degree to which microbially mediated host phenotypes are impacted by inter-host

dispersal. If true, it will be interesting to see what factors best predict whether such factors are influenced by microbial dispersal.

These results have wider implications for the assembly of host-associated communities as well. Ecological communities assemble at multiple scales with different processes acting at each (Lindström and Langenheder, 2012; Cadotte and Fukami, 2005; Münkemüller *et al.*, 2014). To date, most research on host-microbe systems has focused on the local scale and the interactions between an individual host and the microorganisms that reside within them, or at the even finer scale of specific host cell types (Donaldson *et al.*, 2015; Spor *et al.*, 2011). These studies have and continue to discover novel mechanisms that shape the animal microbiota at fine scales, but there remains the almost frustrating inability to identify clear and strong factors shaping it at broad scales (Falony *et al.*, 2016; Zhernakova *et al.*, 2016). Our results demonstrate that this approach alone is insufficient to fully understand the assembly and dynamics of these systems. Instead, such an understanding will require us to expand our focus to the processes shaping these systems at the scales of populations of hosts or even communities of multiple host species.

CHAPTER V

CONCLUSION

Synthesis

Some of the oldest debates in community ecology concern the relative importance of chance and stochastic processes compared to deterministic forces in the formation of communities. Is the assemblage of species occurring at a site the "best" possible combination of species that could occur in that environment, or is the composition of that assemblage a function of varying dispersal rates and probabilities, stochastic birth-death processes, and the unique history of that particular site? Perhaps one reason these debates have been featured so prominently over the past century is that one of the ultimate goals of community ecology is to be able to accurately predict, manipulate, and engineer communities with specific compositions, behavior, and functions. This has certainly been a main goal of animal-microbiome studies: to understand the mechanisms of their assembly well enough that they can be manipulated to improve the health of the animal host.

The aim of the research presented in this dissertation has been to further this understanding by testing the importance of two assembly mechanisms that have often been overlooked in animal-microbiome studies, namely microbial dispersal and ecological drift. Furthermore these mechanisms were studied in the context of factors common to animal hosts but relatively rare in other environments, such as ontogenetic development and the immune system. The results of this research have reflected some of the historical tension between deterministic and individualistic models of community

assembly. It is interesting, for example, that early community ecologists compared the assembly of communities to the development of an organism, while more recently the assembly of the animal-microbiome was proposed to be analogous to the development of a host's organ. In Chapter II, I showed that while the composition and diversity of zebrafish intestinal communities changed in a predictable manner as a function of host development, these communities were characterized by a high degree of inter-individual variation and their assembly was far less stereotyped than the development of a host organ. In Chapter III, I explored the assembly of these communities in more depth and showed that a neutral model incorporating only the effects of random dispersal and drift could explain a large amount of this variation. However, the ability of the model to do so decreased over host development, and microbial taxa whose distributions deviated from neutral predictions were ecologically and phylogenetically distinct, showing signs of selection by the host environment. Finally, in Chapter IV, I showed that dispersal of microorganisms among zebrafish hosts can not only heavily alter the effects of the host immune system, but even override it, such that the effect of the immune system on community composition was much stronger in the absence of migration from other hosts.

No man is an island, a microbiome unto itself…

A common theme of this dissertation is that while host-specific factors certainly help shape the composition of host-associated microbial communities, they are insufficient to fully explain the assembly of these communities. Instead, dispersal and chance can have surprising large effects on animal associated microbial communities, and can even heavily modify the effects of host-specific factors. This seems like a simple

conclusion, but it is important because animals in nature typically do not exist in a vacuum. Instead, they are members of populations and those populations form complex communities of multiple interacting species. Thus the assembly and dynamics of communities associated with any given animal host are not isolated to that individual, but are instead influenced by the assembly and dynamics occurring in the communities associated with all the other animal hosts and abiotic environments with which the individual interacts. A complete model of the assembly of animal associated microbial communities will therefore require understanding both the ecology of associated microorganisms as well as the ecology of their hosts, and an explicit consideration of dispersal in order to link the two. Building such a model will likely be an enormous task, and will likely require some conceptual shifts that can be informed by the history of community ecology. Hopefully this dissertation will help us move closer towards that goal.

APPENDIX A

SUPPLEMENTAL INFORMATION FOR CHAPTER II

SUPPLEMENTAL MATERIALS AND METHODS

Animal husbandry

Multiple pairs from a cohort of 320 dpf Tü strain zebrafish maintained at the University of Oregon zebrafish facility were set up overnight in separate, autoclaved spawning cages containing system water at the University of Oregon's zebrafish facility with dividers separating male and female in order to obtain enough embryos from a single pair for the duration of the study. The following day, dividers were removed from each of the separate spawning cages and fish were allowed to spawn for 2.5 hours, at which point embryos were removed from spawning cages (where adults remained, each pair in a separate cage) and allowed to develop until 8 hours post-fertilization in embryo medium (EM). A single sibship was identified which contained 280 fertilized, developing embryos and was divided evenly among four previously autoclaved tanks containing standard, non-sterile EM as well as sterile glass slides for collecting surface environmental samples. The parents of the sibship used were euthanized 8 hours after spawning, their intestines dissected and frozen in liquid nitrogen and stored at -80ºC for later DNA extraction.

From 8 hours post-fertilization until 21 dpf, fish were maintained in tanks with uncirculated system water that was manually changed to remove debris and uneaten food on a daily basis by siphoning and replacing approximately one third of the tank volume. At 21 dpf, the tanks were transferred to the main facility that housed other zebrafish lines

and utilized a recirculating water system, where they remained for the remainder of the study. The 21 dpf sampled fish were briefly exposed to the recirculating water system before being sampled, but had otherwise been on the closed system of the nursery for the entirety of their prior development. Prior to 6 dpf, larval zebrafish were not fed and subsisted off yolk alone. At 6 dpf fish were fed with live *Paramecium* (prepared according to methods described by ZIRC, Zebrafish International Resource Center) only until 10 dpf. After the 10 dpf sampling until 21 dpf, fish were fed *Artemia* (brine shrimp; Artemia International LLC; cultured according to methods described by ZIRC) supplemented with Larval AP100 (Aquatic Eco-Systems, now Pentair Aquatic Eco-Systems). After the 21 dpf sampling, fish were fed a facility-standard dry food mixture twice per day (once in the morning and once in the late afternoon; ~5 hours apart) in the main fish facility. The dry diet consisted of a mix of 50% Tetramin Flake (Tetra, Blacksburg, VA), 25% Nelson Silver Cup Trout Pellet (Nelson and Sons, now Skretting, Tooele, UT) and 25% Zeigler Adult Zebrafish Diet (Zeigler Brothers Inc., Gardners, PA), supplemented with a small amount of Cyclop-Eeze (Argent Chemical Laboratories, Redmond, WA). Between the 75 dpf sampling and 380 dpf, the standard dry food diet fed to fish in the facility was changed to New Life Spectrum Small Fish Formula while fed according to the same schedule. While on the main recirculating water system, tanks were not cleaned or changed until after 75 dpf to reduce the introduction of bacteria from the facility staff.

DNA extraction, 16S rRNA gene amplification and sequencing

Dissected intestinal samples were flash frozen in liquid nitrogen then thawed at 65ºC and DNA extracted using a protocol modified from the Qiagen DNeasy kit to include a bead-beating step. After thawing, the samples were bead-beat on high on a Mini-Beadbeater-16 instrument (Biospec, Bartlesville, OK) for 1 minute once (4, 10, 21 dpf), or twice (28 dpf and older), with chilling on ice in between beatings. Subsequently, lysozyme was added to a concentration of 20mg/ml and the samples incubated for 45 minutes at 30ºC. Then, 220 µl (4, 10, 21 dpf) or 440 µl (28 dpf and older) buffer AL (Qiagen) was added in addition to 10 µl (4, 10, 21 dpf) or 20 µl (28dpf and older) proteinase K (supplied with Qiagen kit), samples vortexed well and incubated for 30 minutes at 56ºC. After enzymatic digestions, 220 µl (4, 10, 21dpf) or 440 µl (28dpf and older) of 100% ethanol was added, samples vortexed, and beads allowed to settle to prevent addition to the DNA-binding columns. Then, up to 700 µl of digestion mix was added to a Qiamp DNA micro (4, 10, 21 dpf) or Qiagen DNeasy (28 dpf and older) column and spun 30 seconds at 6,000 x *g*. Flow-through was discarded and remaining digestion mix was added to columns and spun down again if necessary. In order to ensure maximum recovery of DNA, we then rinsed the beads with a 1:1:1 (v/v/v) mix of ELB, buffer AL and ethanol and allowed the beads to settle again. The supernatant was then gently removed to avoid sucking up beads, and was applied to the column and spun again. Following binding of DNA to column the manufacturer's guidelines for washing were followed and DNA was eluted in 34 µl (4, 10, 21 dpf) or 100 µl buffer AE.

DNA templates were used in a two-step PCR method to sequence the V4 region of the bacterial 16S rRNA gene. The first round of PCR employed primers that were comprised of (in 5' to 3' direction) partial Illumina adapter sequences, a 6 nucleotide

index (each index contained 2 or more mismatches from one another) and the V4 targeting forward or reverse primer sequence. In order to target the V4 region and obtain as much useful sequence as possible from our paired-end 150 nucleotide sequencing runs, we used the "V4-reverse" primer identified by (Claesson *et al.*, 2010) along with a slightly modified (to reduce anneal temperature range closer to the reverse primer) version of the V4 forward primer used by (Caporaso *et al.*, 2011) (see Supplementary Table 2 for full oligonucleotide sequences). The first round of PCR was performed in triplicate for each sample with approximately equal amounts of DNA template (up to 250 ng per reaction) for each sample. The reactions were carried out with a 2 minute denature step at 98ºC, followed by 22 cycles of denature at 98ºC for 20 seconds, annealing at 50ºC for 30 seconds and extension at 72ºC for 20 seconds, with a final extension at 72ºC for 2 minutes. Triplicate reactions were pooled and cleaned using Zymogen (Irvine, CA) 96-well format spin columns and eluted in 30 µl. Six microliters of this eluate was used as template in the second round of PCR which utilized primers complementary to the partial Illumina adapters added in the first round, and added the rest of the Illumina adapter sequences. The cycling conditions for the second round of PCR were as before except a 66ºC anneal step was employed with only 12 cycles. Both PCR reactions used Phusion HotStart II polymerase (Thermo Scientific), GC buffer, 200 nM each primer (primers were HPLC purified) and Mo Bio (Carlsbad, CA) certified DNA-free water in a total reaction volume of 25 µl. The product from the second round PCRs were run out on a 1.5% agarose gel in 0.5X TBE to separate out low molecular weight primer-dimers as well as a smaller than expected band that was cloned and identified as containing zebrafish mitochondrial sequences. DNA product in the range of approximately 320 –

600 base-pairs was excised from the gel and cleaned up using Zymogen's ZR-96 Zymoclean Gel DNA Recovery Kit. Cleaned products were quantified using a Qubit (Life Technologies) flourometer then mixed in equal amounts and submitted for sequencing on two lanes of the Illumina HiSeq 2000 platform at the University of Oregon's genomics core facility with paired-end 150 nucleotide reads. Samples that returned poor numbers of sequences despite good amplification were remixed and sequenced on part of a third lane.

Isolate genome DNA extraction and preparation for sequencing

For sequencing of zebrafish isolated bacterial genomes, genomic DNA was isolated from cultures using either the Qiagen DNeasy or Mo Bio Ultraclean Microbial DNA Isolation kits. Genomic DNA was prepared for paired-end 150 cycle Illumina sequencing using standard paired-end shotgun sequencing library methods. Briefly, 2 µg of cleaned genomic DNA in a 100 µl volume of elution buffer was sheared to a median size of approximately 300 base-pairs using a Bioruptor (Diagenode; Denville, NJ), then cleaned up and concentrated to 35 µl with a Qiagen MinElute column. The entire eluted DNA volume was then end-repaired for 30 minutes at room temperature in a 50 µl reaction using the NEB (Ipswich, MA) Quick Blunting Kit as per the manufacturer's recommendations. The end-repair reaction was then cleaned up and eluted in 32 µl using a Qiagen MinElute column. In order to add A-overhangs, the entire elution volume was mixed with 5 µl of NEB buffer 2, 10 µl of 1mM dATP, and 3 µl of NEB Klenow (exo-) enzyme and incubated for 30 minutes at 37ºC. This reaction was again cleaned up with a Qiagen MinElute column, eluted in 20 µl elution buffer and quantified with a Qubit

fluorometer. Next, in a 60 µl total volume reaction, Illumina adapters containing 5 base-pair indices and T-overhangs were ligated to the cleaned genomic DNA fragments with A-overhangs by adding a 10-fold molar excess of adapters to genomic DNA along with 1 µl T4 DNA ligase (NEB), 6 µl of NEB Buffer 2 and 6 µl of 10mM rATP and then incubated 2 – 4 hours at room-temperature. The entire reaction was cleaned and concentrated on a Qiagen MinElute column, eluting in 20 µl. From the cleaned ligation, 5 µl was used as template for PCR with Phusion polymerase and primers "PE_Amp1/2" (Supplementary Table 2) to generate the asymmetrical adapters. Cycling conditions were as follows: 2 minute denature step at 98ºC, followed by 16 cycles of denature at 98ºC for 10 seconds, annealing at 65ºC for 30 seconds and extension at 72ºC for 30 seconds, with a final extension at 72ºC for 2 minutes. Reactions were subsequently cleaned and run on a 2% (w/v) agarose gel with 0.5X TBE and a gel fragment in the range of 150 – 600 base-pairs excised to avoid adapter-dimers, then DNA was extracted using the Qiagen gel extraction kit and eluted in 35 µl volume.

16S rRNA gene sequence processing and analysis

Raw reads were end-trimmed at their first nucleotide below quality score 5 in order to remove very low-quality nucleotides at the ends that are characteristic of Illumina reads. After trimming barcodes and primer sequences, a sliding window quality filtering was used with a cutoff of an average of 20 quality score over a 15-nucleotide window to trim low quality regions, and the remaining read pairs were subsequently trimmed to 100 nucleotides on each end. The high-quality, 100-nucleotide reads were then aligned to the zebrafish genome using Bowtie (Langmead *et al.*, 2009) to remove

host-derived reads, then to the Greengenes core bacterial alignment using mothur to filter reads that did not align to the 16S rRNA gene region targeted by our primers. Pairs of 100-nucleotide sequences that passed all quality and alignment filters were then selected and concatenated prior to demultiplexing using the combined 12 nucleotide index (6 from each end). For the analysis of microbial community composition, an open-reference OTU picking approach using UCLUST (Edgar, 2010) was used to bin sequences into 97% similar OTUs against the Greengenes October 2012 reference set (McDonald *et al.*, 2012), followed by *de novo* OTU clustering of the remaining reads that failed to cluster against the reference. A minimum OTU cluster size cutoff of four was used based on simulations of expected error rates, in which a minimum cutoff of two sequences per OTU was required to minimize the creation of spurious OTUs in a dataset the size of the one used in this study. The most abundant sequence in each OTU cluster was picked as the representative sequence. Representative sequences were aligned against the trimmed Greengenes core set, filtered with the Lane mask and used to build a phylogeny with FastTree 2 (Price *et al.*, 2010). Taxonomy of representative sequences was assigned using Rtax (Soergel *et al.*, 2012) with the Greengenes reference set trimmed to the sequenced region to improve taxonomic calls (Werner *et al.*, 2012). After taxonomic assignment OTUs that were assigned as chloroplast were removed in addition to those that were not assigned as at least "k__Bacteria". The latter set was determined by BLAST against NCBI's nr/nt database to consist mainly of host-derived sequences that escaped our previous Bowtie alignment filter, as well as some fungal mitochondrial sequences.

Genomic ancestral state reconstruction sets and metagenomic predictions with PICRUSt

The accuracy of metagenome inferences by PICRUSt (Langille *et al.*, 2013) can be affected by the relatedness of observed OTUs to the genomes used as references. This relationship can be measured by the phylogenetic distance between each OTU (weighted by its abundance) in the sample and its closest relative among the reference genomes, resulting in the nearest sequence taxon index (NSTI). In order to decrease this index, and thus increase the likely predictive value of PICRUSt inferences, we selected a subset of our culture collection of zebrafish gut-associated bacteria and sequenced their genomes (deposited under NCBI umbrella BioProject number PRNA202207; see Supplementary Table 3 for individual NCBI and IMG accession numbers) for incorporation into the reference database (Table 1). We chose 26 isolates that represented the diversity of taxa present at different developmental stages (Supplementary Table 3) including a unique strain from the previously uncultured class of Firmicutes, CK-1C4-19.

In order to rebuild genomic predictions for PICRUSt (Langille *et al.*, 2013) we first obtained KEGG Orthology (KO) term counts and 16S rRNA gene counts from 12 899 bacterial genomes on the Integrated Microbial Genomes (IMG) database version 4.2 (including our zebrafish-isolated strain genomes and publicly available genomes), as well as the 16S rRNA gene sequences from the May 2013 Greengenes release. We first used the Greengenes provided IMG to Greengenes ID map to extract all the 16S rRNA gene sequences in Greengenes that had a corresponding IMG genome and place them at the top of sequence files containing the rest of the Greengenes 99% OTU representative sequences so that sequenced genomes would preferentially form the cluster seeds. These

sequences were then reclustered at 99% similarity, and representative sequences were used to build a phylogeny (now preferentially represented by sequenced genomes) for *de novo* genome predictions by ancestral state reconstruction. To determine which zebrafish isolated bacterial genomes already had closely related organisms with available genome sequences, we first obtained full-length 16S rRNA gene Sanger sequences directly from each of our isolate strains (GenBank accession numbers listed in Supplementary Table 3). We then used open-reference clustering of the zebrafish isolated bacterial 16S rRNA gene sequences against our newly created reference set at 99% similarity against the 99% Greengenes OTUs representative sequences which had a corresponding IMG genome sequence, to identify the zebrafish isolates strains that were unique as well as those that clustered with already sequenced genomes. This resulted in 3 unique strains (ZOR0006, ZWU0021 and ZOR0020) that were added to the zebrafish isolate genome incorporated dataset and 23 that clustered with an already sequenced genome, of which 22 (1 pair of isolates clustered together) were used to replace the corresponding already sequenced IMG genome in the zebrafish isolate genome incorporated dataset (Supplementary Table 3).

For each of the four different sets of NSTI values from metagenomic inferences shown in Supplementary Figure 2, a separate phylogeny was created from 16S rRNA gene sequences (accession numbers in Supplementary Table 3) and unique genome predictions made with or without different zebrafish isolated bacterial genomes included replacing the IMG genome closest relative. Metagenomic predictions were then made for the microbial community OTU tables from closed-reference OTU picking. The inclusion of genomes from zebrafish isolated strains significantly reduced the NSTI values at 21,

35, 75 and 380 dpf time points ($p < 0.01$, Wilcoxon rank sum test) and had the greatest impact on the adult fish where the previously uncultured CK-1C4-19 class of bacteria was most abundant. This reduction was not simply due to the inclusion of the novel CK-1C4-19 genome, as when we replaced only the 23 IMG reference genomes with closely related zebrafish isolate genomes ("ZIG included – no novel" in Supplementary Figure 2) there were still significant decreases in the NSTI values at late-juvenile and adult time points. As the inclusion of all zebrafish isolated bacterial genome sequences provided the lowest NSTI values, all subsequent analyses of predicted metagenomes were based off these PICRUSt inferences. Analysis of differences in metagenomic profiles was performed using STAMP v2.0.5 (Parks and Beiko, 2010). Analysis of discriminatory genes among genomes of zebrafish isolated strains was performed with LEfSe (Segata *et al.*, 2011).

Isolation of and naming convention of zebrafish bacterial strains

All bacterial isolates were initially cultured directly from zebrafish intestinal samples from a variety of developmental stages (Supplementary Table 3). Strains were isolated from fish at the University of Oregon (prefixed ZOR) and the University of North Carolina at Chapel Hill (prefixed ZNC) or obtained from a previously published collection at Washington University in St. Louis (prefixed ZWU; Rawls *et al.*, 2006). All strains were able to grow on general-purpose brain heart infusion (BHI) media under aerobic conditions at 30°C with the exception of ZOR0034 and ZWU0022 which required anaerobic conditions.

Genome sequencing, assembly and annotation of zebrafish bacterial strains

Raw Illumina reads were first quality trimmed and filtered then screened for remaining adapter sequences with Cutadapt (Martin, 2011). The quality filtered reads were then put through FLASH (Magoč and Salzberg, 2011) to assemble any overlapping pairs of reads into longer contigs. Overlapping FLASH assembled contigs, non-overlapping paired reads and remaining quality-filtered, unpaired reads were assembled using velvet (Zerbino and Birney, 2008). All gene calling and annotation of the velvet assembled contigs was done by IMG except for ZWU0020 which was previously annotated on RAST and gene calls on contigs were imported into IMG and further annotation was performed by IMG.

Taxonomic placement of novel zebrafish bacterial strains

In order to confidently assign the taxonomic relationship of the 26 isolate genomes we determined their phylogenetic relationships among sequenced genomes based on the alignment of 400 proteins as implemented by PhyloPhlAn (Segata *et al.*, 2013). The resulting taxonomies were in agreement with assignments from full length 16S rRNA gene sequences derived from each of the strains, which were used to extend the taxonomic assignments where high confidence results were obtained (Supplementary Table 3) using SINA (Pruesse *et al.*, 2012) alignment and SILVA taxonomy inference and RDP classification (Wang *et al.*, 2007). The CK-1C4-19 strain full taxonomic string was not confidently inferred by PhyloPhlAn due to an absence of closely related genomes, but its phylogenetic placement was in agreement with it being part of the Firmicutes and its nearest neighbor being the class Erysipelotrichia. Given the abundance

of the CK-1C4-19 class in our study, its candidate status, and the lack of previous description of it outside of high-throughput sequencing datasets, we also investigated if these 16S rRNA gene sequences were present in the high-quality full length clone library derived from our previous zebrafish study (Roeselers *et al.*, 2011). We found a single clone isolated from the University of North Carolina at Chapel Hill facility that clustered at 97% similarity (NCBI Accession: HM780377.1) with the ZOR0006 16S rRNA gene sequence and was also classified only as a Firmicutes, with other known Firmicutes clones clustering together with ZOR0006 at the shallower 82% similarity. Taken together, these independent pieces of evidence suggest that the uncultured and previously unsequenced CK-1C4-19 bacteria are commonly found in zebrafish intestines and are most appropriately placed as a class within the Firmicutes, in line with the Greengenes taxonomy.

Quantification of secreted IgM (sIgM) expression

Three fish from each tank from timepoints 10, 21, 28, 35, and 75 dpf (60 fish total) were analyzed for sIgM expression via qRT-PCR. Total RNA extraction was performed on the fish carcasses (after gut dissection described above) using a TRIzol/chloroform extraction protocol. cDNA was synthesized using an oligo-dT primer and SuperScript® III Reverse Transcriptase (Life Technologies) following the accompanying protocol, including an addition of RNaseH (Life Technologies) at the end.

Each sample was run in duplicate (including the two controls). This required spreading the samples across 3 different runs, divided such that run was not covarying with tank or timepoint. The controls were housekeeping genes SDHA and ElF1β (primers

purchased from PrimerDesign). The master mixed used was KAPA SYBR® FAST ABI Prism® 2X qPCR Master Mix (KK4605). Primers targeting sIgM were as identified in (Danilova and Steiner, 2002), and are listed in Supplementary Table 2.

SUPPLEMENTAL RESULTS

Discriminatory genes in zebrafish isolated strain genomes

Amongst our culture collection of 26 zebrafish isolated bacterial strains with sequenced genomes we identified 8 strains that belonged to genera that were highly discriminatory of the larval (4 and 10 dpf), 35 dpf and adult (75 and 380 dpf) age classes (Supplementary Figure 4), as well as a representative of the previously uncultured CK-1C4-19 candidate class that discriminated adult fish microbiotas. We asked what genes distinguished these isolates representative of each age class from the others within our culture collection. We identified a few genes overrepresented among isolates discriminating each age class, including two KO (KEGG Orthology) groups involved in chemotaxis among isolates discriminating larval fish (K03406; methyl-accepting chemotaxis protein, K03408; purine-binding chemotaxis protein CheW) and two ferrous iron transport proteins (K04758, K04759) in isolates discriminating adult fish. Because the strength of this analysis is limited due to the few isolate genome comparisons, we asked if these differentially abundant genes within isolates were also changing throughout development when all the predicted genomes of reference OTUs identified by 16S rRNA gene sequencing (the predicted metagenome) were considered. PICRUSt derived metagenomic predictions for each age class showed a corresponding increase in abundance of KO groups identified as enriched in isolates discriminating a given age

class (Supplementary Figure 6d), showing the ability of PICRUSt to reproduce expected changes in gene abundances on the scale of the entire predicted metagenome.

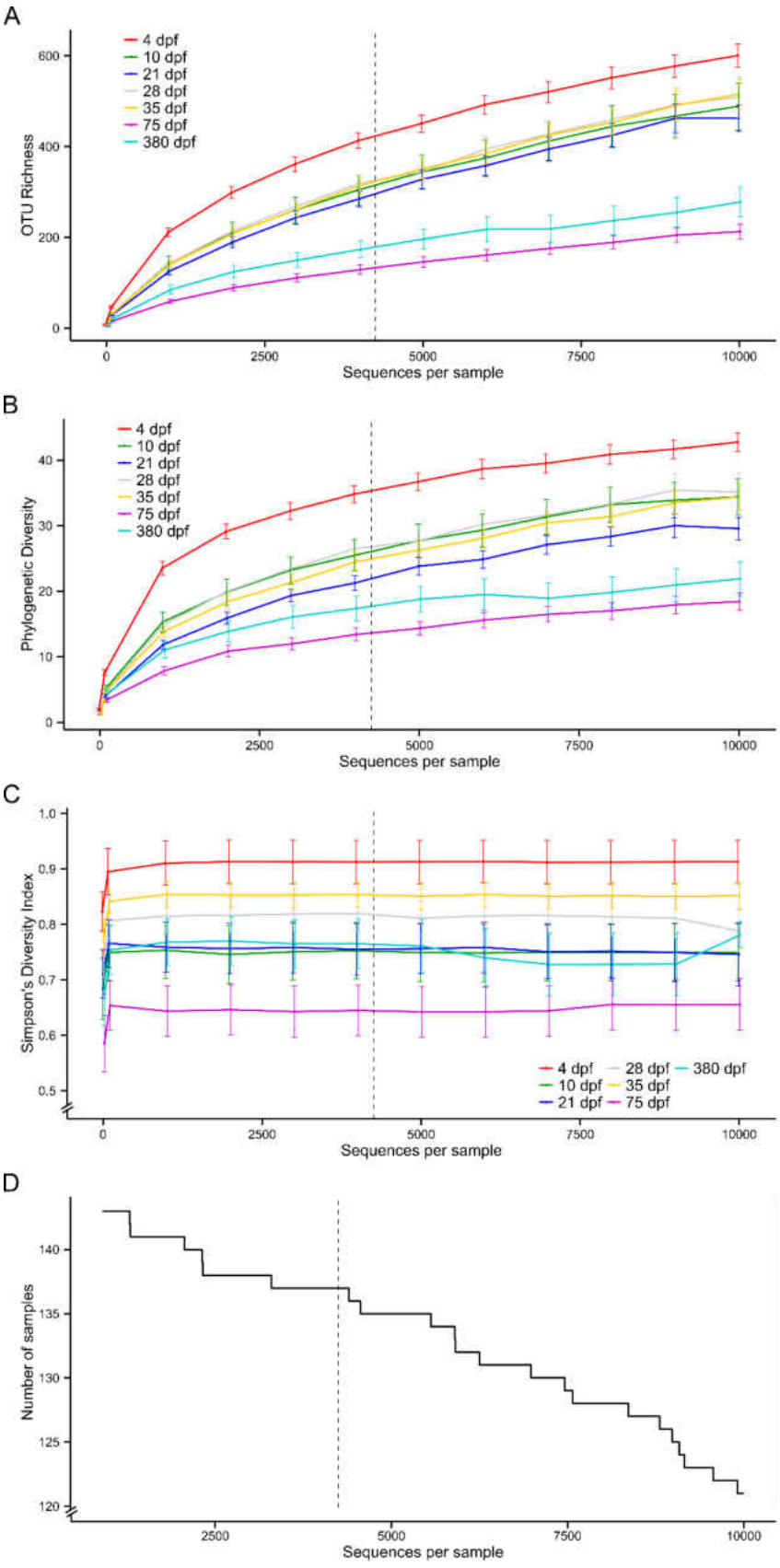
Predicted metagenomes suggest functional shifts by the microbiota during development.

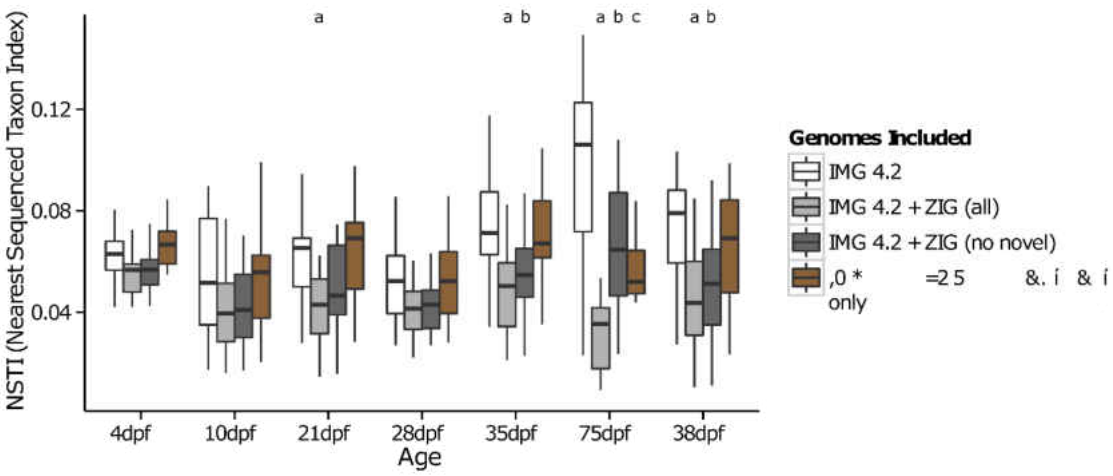
While most broad predicted functional categories exhibited relatively little variation (Supplementary Figure 5) compared to variations in taxonomic abundances, we found that larval, juvenile and adult metagenomes had distinct abundances of predicted KO groups that clustered these age classes together (Supplementary Figure 6a), and some significant differences in functional categories between larval and adult microbiota were apparent (Supplementary Figure 6b). In particular, we noted that cell motility, including the aforementioned bacterial chemotaxis genes, was predicted to be enriched in larval microbiotas and carbohydrate metabolism enriched in adult microbiotas (Supplementary Figure 6c). While KO terms summarized by functional categories showed little differences between environmental and intestinal samples (data not shown), there was a significant increase in the overall dissimilarity in abundance of predicted KO terms in the environment and intestines throughout development (Supplementary Figure 6e; Pearson's $r = 0.255$; $p < 0.0001$). Additionally, there was an increase in predicted KO term dissimilarity between individuals throughout development (Supplementary Figure 6e; Pearson's $r = 0.457$; $p < 0.0001$).

APPENDIX B

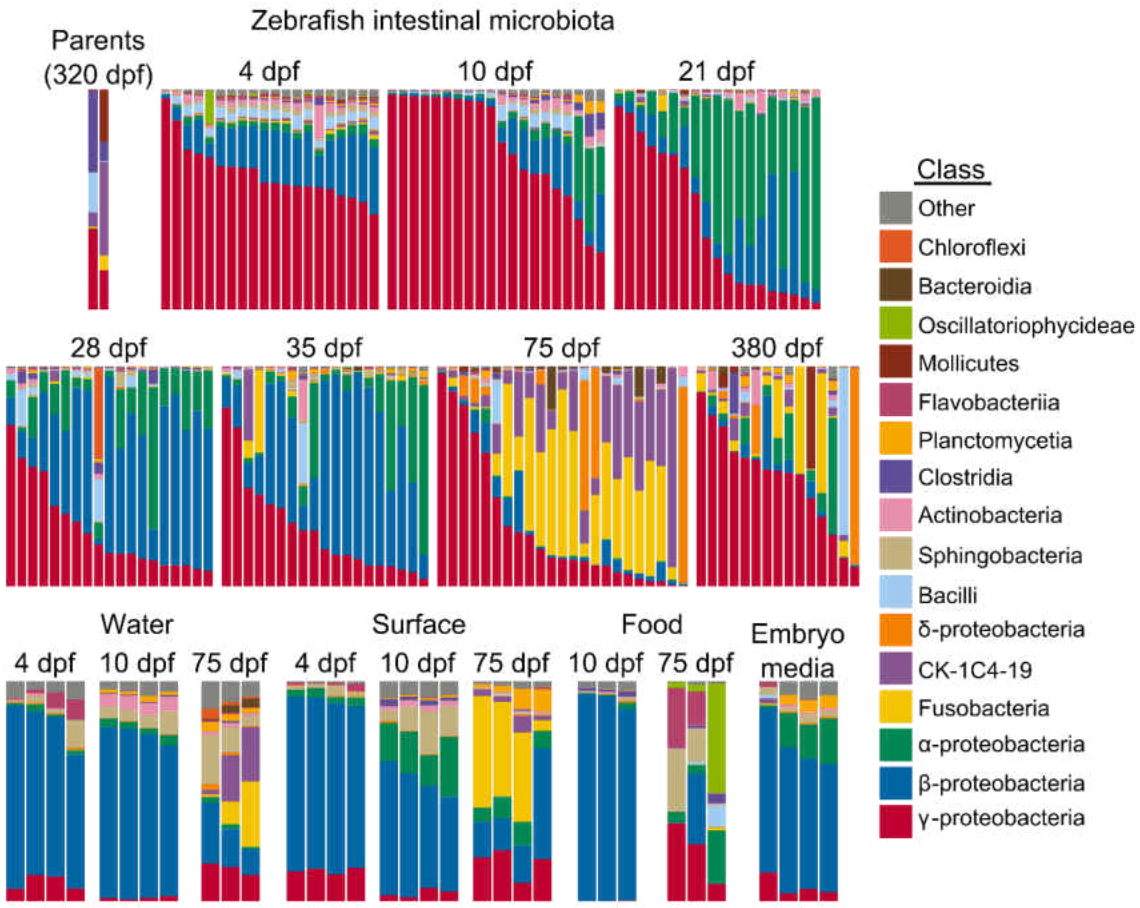
SUPPLEMENTAL FIGURES AND TABLES

Supplementary Figure 1 (following page): Rarefaction curves showing the average number of OTUs (**a**), phylogenetic diversity (**b**), and Simpson's diversity index (**c**) for intestinal samples from each age group as the number of sequences used per sample increases. Slight, non-monotonic behavior in the curves at high rarefaction depths are the result of the number of samples decreasing with depth (**d**). Dotted vertical lines indicate the rarefaction depth used for the majority of the analysis in this study (4250 sequences per sample). Error bars indicate 95% confidence intervals.

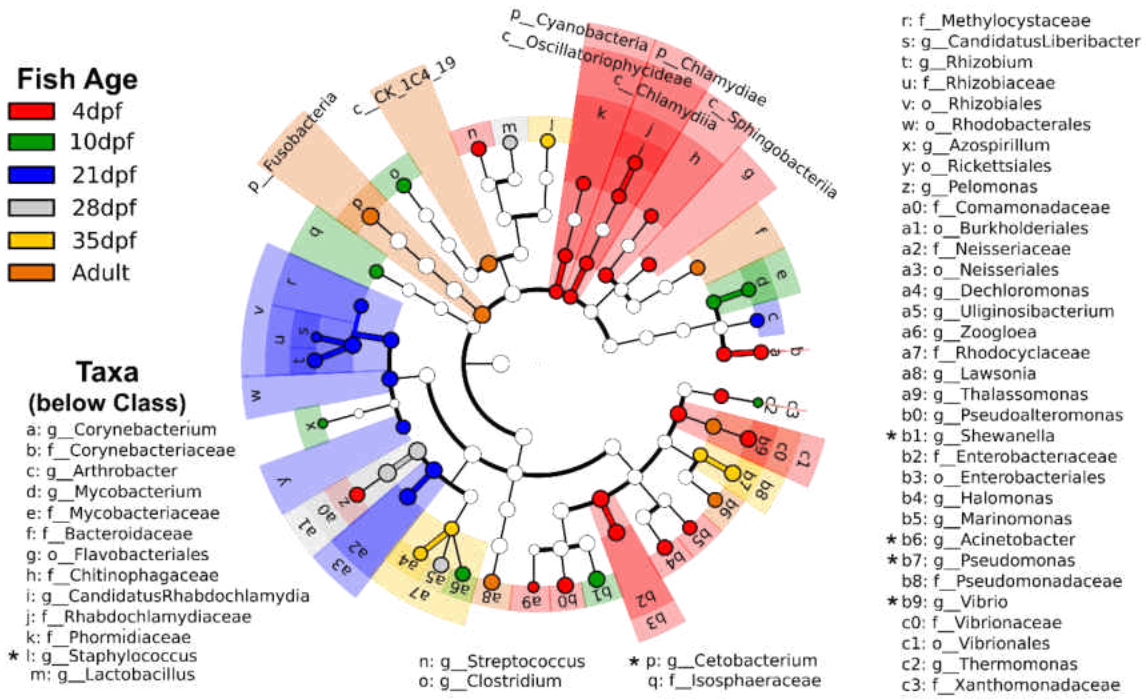

A
4 dpf
10 dpf
21 dpf
28 dpf
35 dpf
75 dpf
380 dpf
OTU Richness
0
200
400
600
0
2500
5000
7500
10000
Sequences per sample
B
4 dpf
10 dpf
21 dpf
28 dpf
35 dpf
75 dpf
380 dpf
Phylogenetic Diversity
0
10
20
30
40
0
2500
5000
7500
10000
Sequences per sample
C
Simpson's Diversity Index
0.5
0.6
0.7
0.8
0.9
1.0
4 dpf
28 dpf
380 dpf
10 dpf
35 dpf
21 dpf
75 dpf
0
2500
5000
7500
10000
Sequences per sample
D
Number of samples
120
125
130
135
140
2500
5000
7500
10000
Sequences per sample



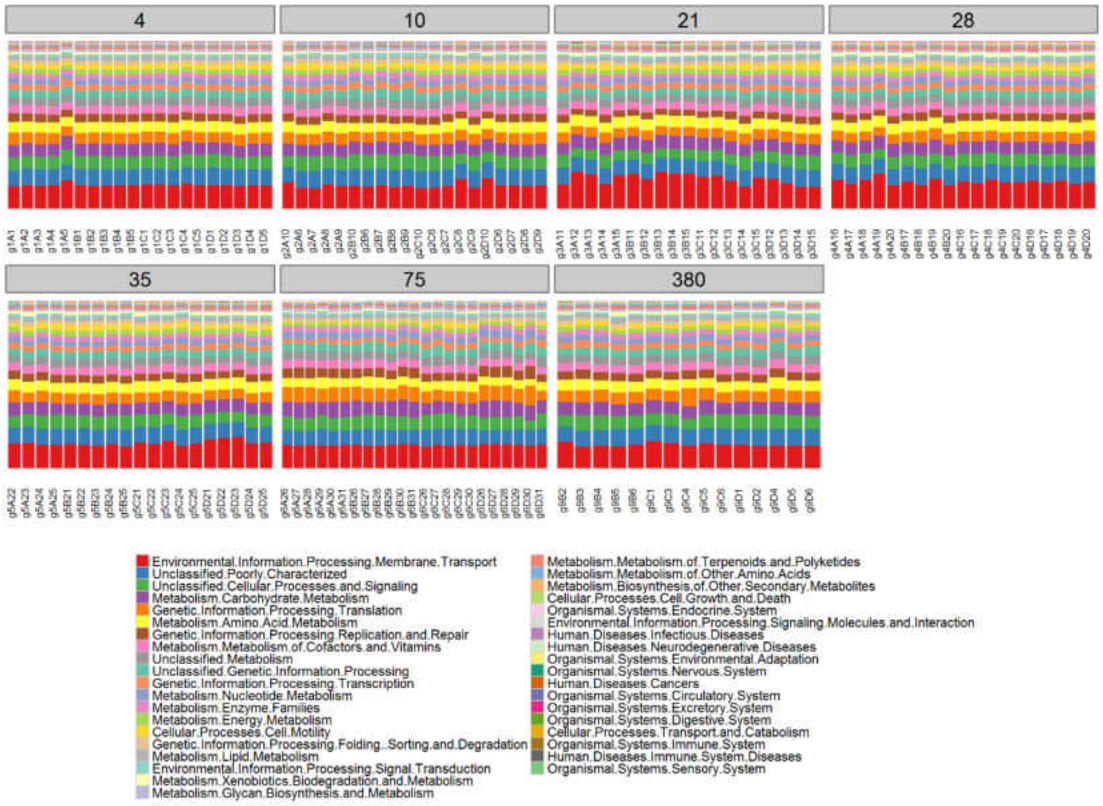
Supplementary Figure 2: Inclusion of zebrafish-specific genomes is predicted to increase inferred metagenome accuracy. NSTI values from PICRUSt derived metagenomic predictions with and without zebrafish isolate genomes (ZIG). Letters denote significant differences ($P < 0.01$; Wilcoxon rank sum test) in NSTI values between one of the three groups including ZIG data and the original IMG 4.2 data without ZIGs.



Supplementary Figure 3: Class level composition of microbiotas from individual samples. The composition of microbiotas from individual fish and environmental samples. Individual Samples under "Embryo Media" represent the initial environments of embryos. Sample w0MF is from the water the embryos were spawned into while samples w0EM1, 2 and 3 represent replicate samples from the embryo media before the embryos were added immediately after collection from spawning tanks (see Supplementary Table 3 for sample name descriptions).

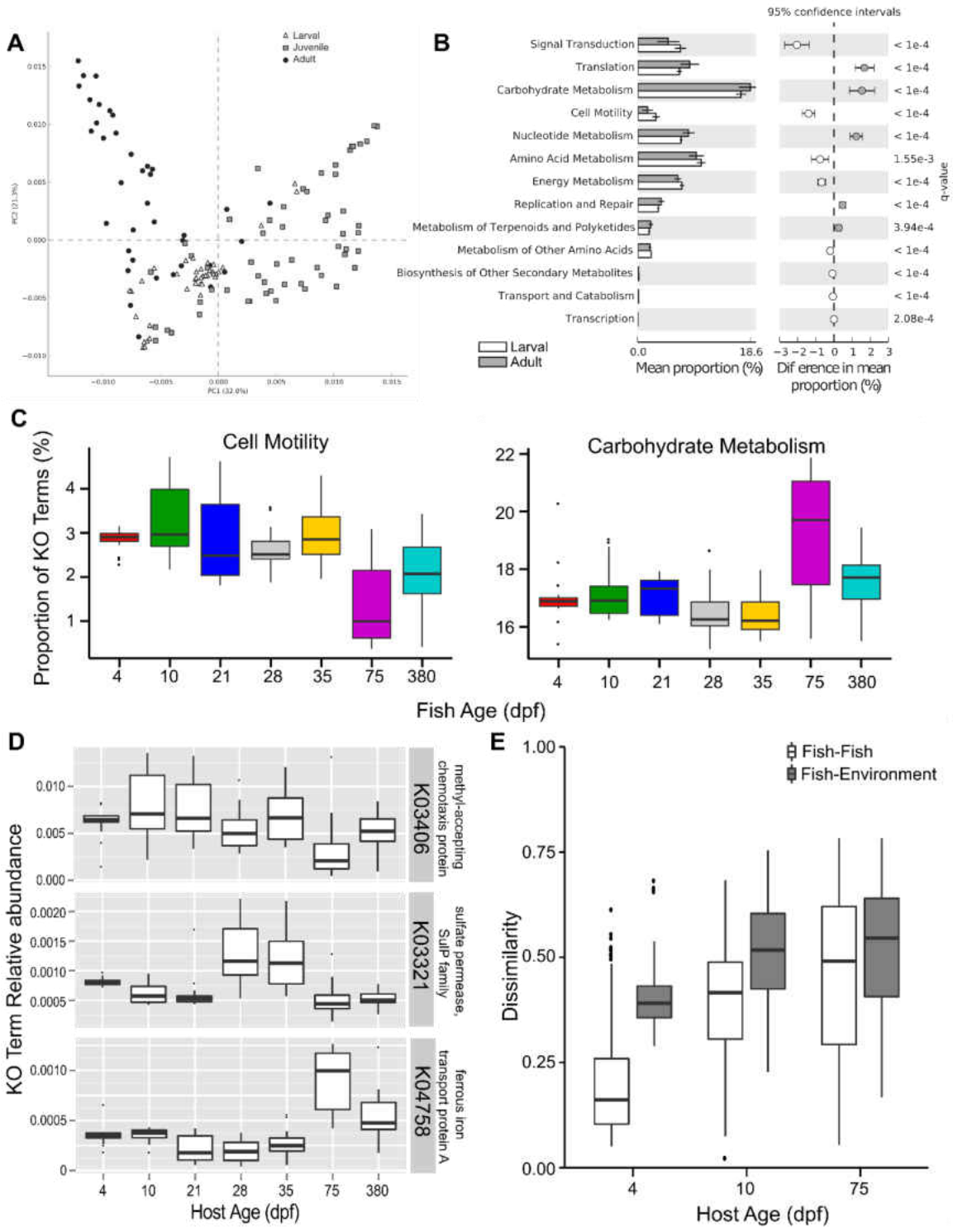


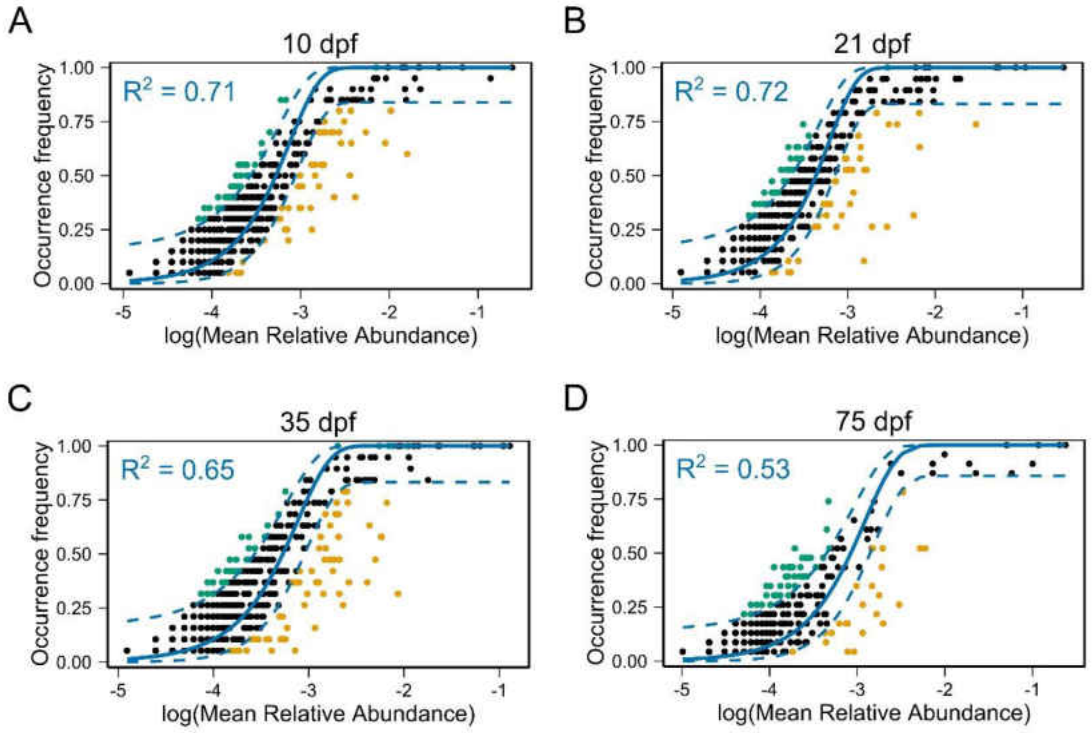
Supplementary Figure 4: Highly discriminatory taxa among developmental time points. Adult class includes 75 and 380 d.p.f. fish. Circles represent discriminatory taxa colored by their respective fish age with color underlay extended to outer circle where taxonomic alphanumeric key is listed. Non-significant lineages have been trimmed for clarity (except parent taxa, open-white circle). Bacterial orders, families and genera are denoted by alphanumerics, while phylum and class names are shown directly on the cladogram. Asterisks (*) denote discriminatory genera that have a zebrafish isolated strain representative with a sequenced genome.



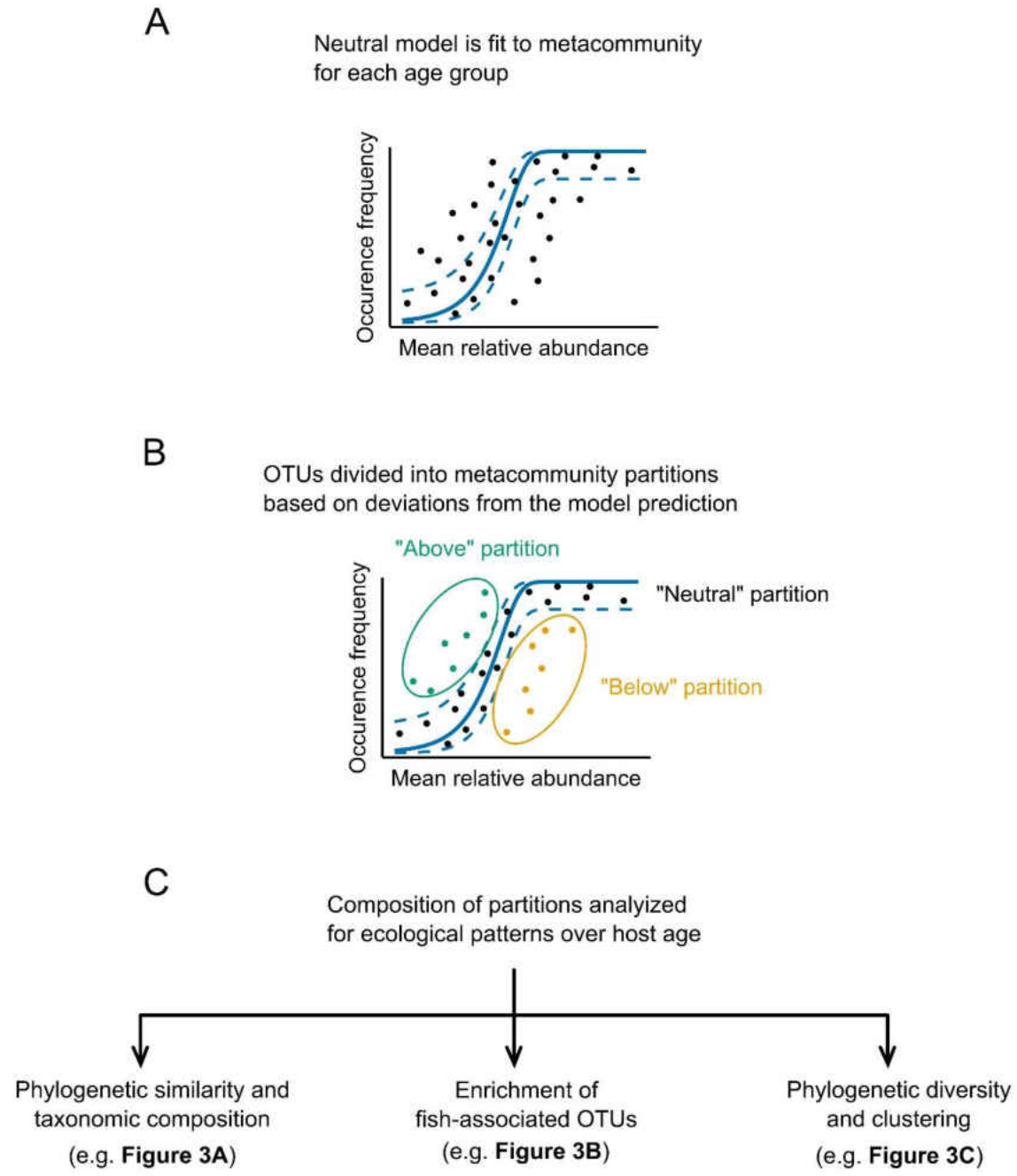
Supplementary Figure 5: Functional categories inferred in individual fish microbiotas. KO term categories summarized at level 2 within the KO hierarchy show few major changes throughout development. Shown for zebrafish gut samples only.

Supplementary Figure 6 (following page): PICRUSt predicted metagenomes reflect unique functions of discriminatory isolate genomes and show increasing dissimilarity among predicted metagenomes throughout development. (**a**) PCoA plot from the abundance of all individual KO terms from PICRUSt predictions of intestinal communities with all ZIG included shows clustering of larval, juvenile and adult metagenomes. (**b**) Significantly differentially abundant (those with $P < 0.01$) level 2 summarized KO groups between larval and adult intestines. Average abundances of KO terms belonging to each functional group are shown along with Benjamini-Hochberg corrected q-values from Welch's t-test. (**c**) Boxplots showing the change in the relative abundance of KO terms belonging to two of the most significantly changed functional groups between larval and adult intestines (Kruskal-Wallis, $P < 0.0001$; cell motility and carbohydrate metabolism). (**d**) Abundances of three KO terms in PICRUSt predicted metagenomes throughout development. K03406 was found to be enriched in isolate genomes discriminatory of larval (4 and 10 d.p.f.) fish, K03321 in genomes discriminatory of 35 d.p.f. fish and K04758 in genomes discriminating adults (75 and 380 d.p.f.). (**e**) Predicted metagenome's Bray-Curtis dissimilarity between fish and between fish and environment increases as the host develops.


A
Larval
Juvenile
Adult
B
95% confidence intervals
Signal Transduction
Translation
Carbohydrate Metabolism
Cell Motility
Nucleotide Metabolism
Amino Acid Metabolism
Energy Metabolism
Replication and Repair
Metabolism of Terpenoids and Polyketides
Metabolism of Other Amino Acids
Biosynthesis of Other Secondary Metabolites
Transport and Catabolism
Transcription
< 1e-4
1.55e-3
3.94e-4
2.08e-4
q-value
Larval
Adult
Mean proportion (%)
Dif erence in mean proportion (%)
C
Cell Motility
Carbohydrate Metabolism
Proportion of KO Terms (%)
Fish Age (dpf)
D
KO Term Relative abundance
K03406
methyl-accepting chemotaxis protein
K03321
sulfate permease, SulP family
K04758
ferrous iron transport protein A
Host Age (dpf)
E
Fish-Fish
Fish-Environment
Dissimilarity
Host Age (dpf)



Supplemental Figure 7: Fit of neutral model across zebrafish development. The predicted occurrence frequencies for 10 (**A**), 21 (**B**), 35 (**C**), and 75 (**D**) dpf zebrafish communities. OTUs that occur more frequently than predicted by the model are shown in green while those that occur less frequently than predicted are shown in orange. Dashed lines represent 95% confidence intervals around the model prediction (blue line).



Supplemental Figure 8: Partitioning of metacommunities into neutral and non-neutral groups. A) The neutral model was fit to the metacommunity of each age group. **B**) OTUs were then divided into separate partitions based on whether they were consistent with (in black) or deviated above (in green) or below (in orange) the neutral prediction. This resulted in 3 partitions per each of 7 age groups (21 total). **C**) Partitions were analyzed to test the hypothesis that deviations from neutral predictions are ecologically informative.

Primer Name	Oligonucleotide Sequence [1]	Use
V4F_b	CGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATXXXXXX**GTGTGCCAGCMGCCGCGG**	16S PCR, round 1
V4R_b	ACACTCTTTCCCTACACGACGCTCTTCCGATCTXXXXXX**TACNVGGGTATCTAATCC**	16S PCR, round 1
16S_ill_step2_P2	AAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGC	16S PCR, round 2
16S_ill_step2_P1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG	16S PCR, round 2
sIgM_f	ATGGAGCAATGGCACTGTG	sIgM qPCR
sIgM_r	CCAAGTCACAAACACCTCCTTGGGC	sIgM qPCR
PE_Amp1	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC	Genome sequencing
PE_Amp2	CAAGCAGAAGACGGCATACGAGATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATC	Genome sequencing

1. Bold characters denote the sequences targeting the 16S rRNA gene.
   Underlined Xs illustrate the 6 base-pair index positions.
   Corresponding colored characters indicate overlapping bases in round 1 and round 2 PCR primers.

Supplemental Table 1: Primers used in the study.

Supplemental Table 2 (following page): Identity and taxonomic placement of 26 zebrafish isolated strains used in this study.

Strain ID	Phylum	Class	Order	Family	Genus	Species	Host Age	NCBI BioProject Accession	IMG Genome Taxon ID	Greengenes 99% OTU cluster representative[5]
ZNC0006	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Variovorax*	*paradoxus*	Larval	PRJNA205579	2526164568	57761
ZNC0007	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Comamonas*	*testosteroni*	Larval	PRJNA205594	2528311002	4328104
ZNC0008	Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	*Delftia*		Larval	PRJNA205595	2528311000	4075598
ZNC0028	Proteobacteria	Alphaproteobacteria	Rhizobiales	Rhizobiaceae	*Ensifer*		Larval	PRJNA205580	2526164570	2105739
ZNC0032	Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	*Methylobacterium*		Larval	PRJNA205581	2526164571	262609
ZOR0001	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	*Aeromonas*		Larval	PRJNA205571	2526164560	4327593
ZOR0002	Proteobacteria	Gammaproteobacteria	Aeromonadales	Aeromonadaceae	*Aeromonas*		Larval	PRJNA205572	2526164561	4323512
ZOR0005	Firmicutes	Bacilli	Bacillales	Family XII Incertae Sedis	*Exiguobacterium*		Larval	PRJNA205567	2526164559	718694
ZOR0006[1]	Firmicutes	CK-1C4-19					Juvenile	PRJNA205588	2526164558	NONE
ZOR0008	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Moraxellaceae	*Acinetobacter*	*calcoaceticus*	Adult	PRJNA205573	2526164562	853882
ZOR0009[2]	Bacteroidetes	Bacteroidia	Bacteroidales	Rikenellaceae	*Alistipes*		Juvenile	PRJNA205590	2526164725	4303027
ZOR0011[1]	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Plesiomonas*		Juvenile	PRJNA205574	2526164564	10395
ZOR0012	Proteobacteria	Gammaproteobacteria	Alteromonadales	Shewanellaceae	*Shewanella*		Larval	PRJNA205575	2526164563	831929
ZOR0014	Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	*Enterobacter*		Larval	PRJNA205576	2526164565	788922
ZOR0017	Proteobacteria	Betaproteobacteria	Neisseriales	Neisseriaceae	*Chitinibacter*		Juvenile	PRJNA205591	2526164726	558848
ZOR0018[3]	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	*Vibrio/Listonella*		Juvenile	PRJNA205566	2528311003	9345
ZOR0019	Actinobacteria	Actinobacteria	Actinomycetales	Microbacteriaceae	*Microbacterium*		Juvenile	PRJNA205577	2526164566	3846873
ZOR0020[4]	Actinobacteria	Actinobacteria	Actinomycetales	Micrococcaceae	*Kocuria*/Unclassified		Juvenile	PRJNA205578	2526164567	NONE
ZOR0034	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Cetobacterium*		Adult	PRJNA205592	2528311001	658591
ZOR0035	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	*Vibrio*	*furnissii*	Larval	PRJNA205593	2524614720	836702
ZWU0006	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	*mendocina*	Adult	PRJNA205583	2526164572	194489
ZWU0009	Firmicutes	Bacilli	Bacillales	Family XII Incertae Sedis	*Exiguobacterium*		Adult	PRJNA205584	2526164575	4472037
ZWU0011	Firmicutes	Bacilli	Lactobacillales	Carnobacteriaceae	*Carnobacterium*		Adult	PRJNA205596	2526164727	170136
ZWU0020	Proteobacteria	Gammaproteobacteria	Vibrionales	Vibrionaceae	*Vibrio*		Adult	PRJNA205585	2522572152	4355411
ZWU0021	Firmicutes	Bacilli	Bacillales	Staphylococcaceae	*Staphylococcus*		Adult	PRJNA205586	2526164573	NONE
ZWU0022	Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	*Cetobacterium*		Adult	PRJNA205587	2526164574	658591

Shaded cells represent high confidence assignments derived from Phylophlan

1. Incomplete Phylophlan confidence
2. SILVA assigns genus 'Blvii28 wastewater-sludge group'
3. RDP assigns to Listonella genus (80% conf.) while SILVA suggests Vibrio.
4. RDP assigns to Kocuria genus while SILVA does not assign a genus.
5. Isolates that clustered with a 99% Greengenes 13_5 OTU were used to replace the Greengenes OTU when building zebrafish isolated genome (ZIG) included metagenomic predictions with PICRUSt. "NONE" indicates the strain did not cluster with a Greengenes reference 16S sequence at 99% similarity.

REFERENCES CITED

Albert EJ, Sommerfeld K, Gophna S, Marshall JS, Gophna U. (2009). The gut microbiota of toll-like receptor 2-deficient mice exhibits lineage-specific modifications. *Environ Microbiol Rep* **1**: 65–70.

Anderson MJ, Ellingsen KE, McArdle BH. (2006). Multivariate dispersion as a measure of beta diversity. *Ecol Lett* **9**: 683–693.

Arthur JC, Perez-Chanona E, Mühlbauer M, Tomkovich S, Uronis JM, Fan T-J, *et al.* (2012). Intestinal inflammation targets cancer-inducing activity of the microbiota. *Science* **338**: 120–3.

Avershina E, Storrø O, Øien T, Johnsen R, Pope P, Rudi K, *et al.* (2014). Major faecal microbiota shifts in composition and diversity with age in a geographically restricted cohort of mothers and their children. *FEMS Microbiol Ecol* **87**: 280–90.

Bäckhed F, Roswall J, Peng Y, Feng Q, Jia H, Kovatcheva-Datchary P, *et al.* (2015). Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell Host Microbe* **17**: 690–703.

Bashan A, Gibson TE, Friedman J, Carey VJ, Weiss ST, Hohmann EL, *et al.* (2016). Universality of human microbial dynamics. *Nature* **534**: 259–262.

Bates JM, Akerlund J, Mittge E, Guillemin K. (2007). Intestinal Alkaline Phosphatase Detoxifies Lipopolysaccharide and Prevents Inflammation in Zebrafish in Response to the Gut Microbiota. *Cell Host Microbe* **2**: 371–382.

Bates JM, Mittge E, Kuhlman J, Baden KN, Cheesman SE, Guillemin K. (2006). Distinct signals from the microbiota promote different aspects of zebrafish gut differentiation. *Dev Biol* **297**: 374–386.

Benson AK, Kelly SA, Legge R, Ma F, Low SJ, Kim J, *et al.* (2010). Individuality in gut microbiota composition is a complex polygenic trait shaped by multiple environmental and host genetic factors. *Proc Natl Acad Sci U S A* **107**: 18933–18938.

Bergstrom A, Skov TH, Bahl MI, Roager HM, Christensen LB, Ejlerskov KT, *et al.* (2014). Establishment of Intestinal Microbiota during Early Life: a Longitudinal, Explorative Study of a Large Cohort of Danish Infants. *Appl Environ Microbiol* **80**: 2889–2900.

Brinkman BM, Becker A, Ayiseh RB, Hildebrand F, Raes J, Huys G, *et al.* (2013). Gut Microbiota Affects Sensitivity to Acute DSS-induced Colitis Independently of Host Genotype. *Inflamm Bowel Dis* **19**: 2560–2567.

Brown LD, Cai TT, DasGupta A. (2001). Interval Estimation for a Binomial Proportion. *Stat Sci* **16**: 101–117.

Burns JH, Strauss SY. (2011). More closely related species are more ecologically similar in an experimental test. *Proc Natl Acad Sci U S A* **108**: 5302–5307.

Cadotte MW. (2006). Dispersal and Species Diversity: A Meta-Analysis. *Am Nat* **167**: 913–924.

Cadotte MW, Fukami T. (2005). Dispersal, spatial scale, and species diversity in a hierarchically structured experimental landscape. *Ecol Lett* **8**: 548–557.

Campbell JH, Foster CM, Vishnivetskaya T, Campbell AG, Yang ZK, Wymore A, *et al.* (2012). Host genetic and environmental effects on mouse intestinal microbiota. *ISME J* **6**: 2033–2044.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, *et al.* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335–336.

Caporaso JG, Lauber CL, Walters W a, Berg-Lyons D, Lozupone C a, Turnbaugh PJ, *et al.* (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A* **108 Suppl** : 4516–22.

Caswell H. (1976). Community structure: A neutral model analysis. *Ecol Monogr* **46**: 327–354.

Chase JM, Myers JA. (2011). Disentangling the importance of ecological niches from stochastic processes across scales. *Philos Trans R Soc Lond B Biol Sci* **366**: 2351–2363.

Claesson MJ, Wang Q, O'Sullivan O, Greene-Diniz R, Cole JR, Ross RP, *et al.* (2010). Comparison of two next-generation sequencing technologies for resolving highly complex microbiota composition using tandem variable 16S rRNA gene regions. *Nucleic Acids Res* **38**: e200–e200.

Costello EK, Lauber CL, Hamady M, Fierer N, Gordon JI, Knight R. (2009). Bacterial community variation in human body habitats across space and time. *Science* **326**: 1694–1697.

Costello EK, Stagaman K, Dethlefsen L, Bohannan BJM, Relman D a. (2012). The application of ecological theory toward an understanding of the human microbiome. *Science* **336**: 1255–62.

Danilova N, Steiner LA. (2002). B cells develop in the zebrafish pancreas. *Proc Natl Acad Sci U S A* **99**: 13711–6.

Declerck SAJ, Winter C, Shurin JB, Suttle CA, Matthews B. (2013). Effects of patch connectivity and heterogeneity on metacommunity structure of planktonic bacteria and viruses. *ISME J* **7**: 533–542.

Deloris Alexander A, Orcutt RP, Henry JC, Baker J, Bissahoyo AC, Threadgill DW. (2006). Quantitative PCR assays for mouse enteric flora reveal strain-dependent differences in composition that are influenced by the microenvironment. *Mamm Genome* **17**: 1093–1104.

Dethlefsen L, Eckburg P, Bik E, Relman D. (2006). Assembly of the human intestinal microbiota. *Trends Ecol {&} Evol* **21**: 517–23.

Dimitriu PA, Boyce G, Samarakoon A, Hartmann M, Johnson P, Mohn WW. (2013). Temporal stability of the mouse gut microbiota in relation to innate and adaptive immunity. *Environ Microbiol Rep* **5**: 200–210.

Dini-Andreote F, Stegen JC, van Elsas JD, Salles JF. (2015). Disentangling mechanisms that mediate the balance between stochastic and deterministic processes in microbial succession. *Proc Natl Acad Sci* **112**: E1326–E1332.

Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, *et al.* (2010). Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. *Proc Natl Acad Sci U S A* **107**: 11971–11975.

Donaldson GP, Lee SM, Mazmanian SK. (2015). Gut biogeography of the bacterial microbiota. *Nat Rev Microbiol* **14**: 20–32.

Dufrêne M, Legendre P. (1997). SPECIES ASSEMBLAGES AND INDICATOR SPECIES:THE NEED FOR A FLEXIBLE ASYMMETRICAL APPROACH. *Ecol Monogr* **67**: 345–366.

Edgar RC. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Elinav E, Strowig T, Kau AL, Henao-Mejia J, Thaiss CA, Booth CJ, *et al.* (2011). NLRP6 Inflammasome Regulates Colonic Microbial Ecology and Risk for Colitis. *Cell* **145**: 745–757.

Faith DP. (1992). Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**: 1–10.

Faith JJ, Colombel J-F, Gordon JI. (2015). Identifying strains that contribute to complex diseases through the study of microbial inheritance. *Proc Natl Acad Sci* **112**: 633–640.

Falony G, Joossens M, Vieira-Silva S, Wang J, Darzi Y, Faust K, *et al.* (2016). Population-level analysis of gut microbiome variation. *Science* **352**: 560–4.

Fraune S, Bosch TCG. (2010). Why bacteria matter in animal development and evolution. *BioEssays* **32**: 571–580.

Friswell MK, Gika H, Stratford IJ, Theodoridis G, Telfer B, Wilson ID, *et al.* (2010). Site and Strain-Specific Variation in Gut Microbiota Profiles and Metabolism in Experimental Mice Ahmed N (ed). *PLoS One* **5**: e8584.

Funkhouser LJ, Bordenstein SR. (2013). Mom Knows Best: The Universality of Maternal Microbial Transmission. *PLoS Biol* **11**. e-pub ahead of print, doi: 10.1371/journal.pbio.1001631.

Gleason HA. (1926). The Individualistic Concept of the Plant Association. *Bull Torrey Bot Club* **53**: 7.

Goodrich JK, Waters JL, Poole AC, Sutter JL, Koren O, Blekhman R, *et al.* (2014). Human Genetics Shape the Gut Microbiome. *Cell* **159**: 789–799.

Hannon Lab. (2010). FastX Toolkit. *http://hannonlab.cshl.edu/fastx_toolkit/index.html*. http://hannonlab.cshl.edu/fastx_toolkit/index.html.

Hartigan JA, Hartigan PM. (1985). The Dip Test of Unimodality. *Ann Stat* **13**: 70–84.

Heino J. (2013). Does dispersal ability affect the relative importance of environmental control and spatial structuring of littoral macroinvertebrate communities? *Oecologia* **171**: 971–980.

Hildebrand F, Nguyen TLA, Brinkman B, Yunta RG, Cauwe B, Vandenabeele P, *et al.* (2013). Inflammation-associated enterotypes, host genotype, cage and inter-individual effects drive gut microbiota variation in common laboratory mice. *Genome Biol* **14**: R4.

Hooper L V, Littman DR, Macpherson AJ, Ley RE, Lozupone CA, Hamady M, *et al.* (2012). Interactions between the microbiota and the immune system. *Science* **336**: 1268–73.

Horner-Devine MC, Bohannan BJM. (2006). Phylogenetic clustering and overdispersion in bacterial communities. *Ecology* **87**. e-pub ahead of print, doi: 10.1890/0012-9658(2006)87[100:PCAOIB]2.0.CO;2.

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, *et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498–503.

Hubbell SP. (2001). The unified neutral theory of biodiversity and biogeography (MPB-32). Princeton University Press.

The Human Microbiome Project Consortium. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* **486**: 207–14.

Jabot F, Etienne RS, Chave J. (2008). Reconciling neutral community models and environmental filtering: theory and an empirical test. *Oikos* **117**: 1308–1320.

Janssens S, Beyaert R, Lord K, al. et, Lord K, al. et, *et al.* (2002). A universal role for MyD88 in TLR/IL-1R-mediated signaling. *Trends Biochem Sci* **27**: 474–482.

Janzen T, Haegeman B, Etienne RS. (2015). A sampling formula for ecological communities with multiple dispersal syndromes. *J Theor Biol* **374**: 94–106.

Jeraldo P, Sipos M, Chia N, Brulc JM, Dhillon AS, Konkel ME, *et al.* (2012). Quantification of the relative roles of niche and neutral processes in structuring gastrointestinal microbiomes. *Proc Natl Acad Sci* **109**: 9692–9698.

Karmarkar D, Rock KL. (2013). Microbiota signalling through MyD88 is necessary for a systemic neutrophilic inflammatory response. *Immunology* **140**: 483–492.

Kembel SW, Cowan PD, Helmus MR, Cornwell WK, Morlon H, Ackerly DD, *et al.* (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463–1464.

Kerr B, Riley MA, Feldman MW, Bohannan BJM. (2002). Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* **418**: 171–174.

Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, *et al.* (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proc Natl Acad Sci U S A* **108 Suppl** : 4578–4585.

Kurokawa K, Itoh T, Kuwahara T, Oshima K, Toh H, Toyoda A, *et al.* (2007). Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res* **14**: 169–181.

Lam S., Chua H., Gong Z, Lam T., Sin Y. (2004). Development and maturation of the immune system in zebrafish, Danio rerio: a gene expression profiling, in situ hybridization and immunological study. *Dev Comp Immunol* **28**: 9–28.

Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes J a, *et al.* (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol* **31**: 814–21.

Langmead B, Trapnell C, Pop M, Salzberg SL. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.

Lankau EW, Hong PY, MacKie RI. (2012). Ecological drift and local exposures drive enteric bacterial community differences within species of Gal??pagos iguanas. *Mol Ecol* **21**: 1779–1788.

Lear G, Bellamy J, Case BS, Lee JE, Buckley HL. (2014). Fine-scale spatial patterns in bacterial community composition and function within freshwater ponds. *ISME J* **8**: 1715–1726.

Legendre P. (2008). Studying beta diversity: ecological variation partitioning by multiple regression and canonical analysis. *J Plant Ecol* **1**: 3–8.

Legendre P, Gallagher E. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.

Leibold MA, Holyoak M, Mouquet N, Amarasekare P, Chase JM, Hoopes MF, *et al.* (2004). The metacommunity concept: A framework for multi-scale community ecology. *Ecol Lett* **7**: 601–613.

Levy R, Borenstein E. (2013). Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proc Natl Acad Sci* **110**: 12804–12809.

Lindström ES, Langenheder S. (2012). Local and regional factors influencing bacterial community assembly. *Environ Microbiol Rep* **4**: 1–9.

Lindström ES, Östman Ö. (2011). The importance of dispersal for bacterial community composition and functioning Stal LJ (ed). *PLoS One* **6**: e25883.

Linnenbrink M, Wang J, Hardouin EA, Künzel S, Metzler D, Baines JF. (2013). The role of biogeography in shaping diversity of the intestinal microbiota in house mice. *Mol Ecol* **22**: 1904–1916.

Livingston G, Matias M, Calcagno V, Barbera C, Combe M, Leibold MA, *et al.* (2012). Competition–colonization dynamics in experimental bacterial metacommunities. *Nat Commun* **3**: 1234.

Loh G, Brodziak F, Blaut M. (2008). The Toll-like receptors TLR2 and TLR4 do not affect the intestinal microbiota composition in mice. *Environ Microbiol* **10**: 709–715.

Lozupone C, Knight R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**: 8228–8235.

MacArthur RH, Wilson EO. (1967). Theory of Island Biogeography.(MPB-1). Princeton University Press.

Mackie RI, Sghir A, Gaskins HR. (1999). Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr* **69**: 1035S–1045S.

Magoč T, Salzberg S. (2011). FLASH : Fast Length Adjustment of Short Reads to Improve Genome Assemblies. *Bioinformatics* **27**: 2957–2963.

Magoc T, Salzberg SL. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**: 2957–2963.

Manceau M, Lambert A, Morlon H. (2015). Phylogenies support out-of-equilibrium models of biodiversity Mooers A (ed). *Ecol Lett* **18**: 347–356.

Martin M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**.

Martínez I, Stegen JC, Maldonado-Gómez MX, Eren AM, Siba PM, Greenhill AR, *et al.* (2015). The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Rep* **11**: 527–538.

Martiny JBH, Eisen JA, Penn K, Allison SD, Horner-Devine MC. (2011). Drivers of bacterial -diversity depend on spatial scale. *Proc Natl Acad Sci* **108**: 7850–7854.

McCafferty J, Mühlbauer M, Gharaibeh RZ, Arthur JC, Perez-Chanona E, Sha W, *et al.* (2013). Stochastic changes over time and not founder effects drive cage effects in microbial community assembly in a mouse model. *ISME J* **7**: 2116–2125.

McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, *et al.* (2012). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J* **6**: 610–8.

McFall-Ngai M, Hadfield MG, Bosch TCG, Carey H V., Domazet-Lošo T, Douglas AE, *et al.* (2013). Animals in a bacterial world, a new imperative for the life sciences. *Proc Natl Acad Sci* **110**: 3229–3236.

Milligan-Myhre K, Charette JR, Phennicie RT, Stephens WZ, Rawls JF, Guillemin K, *et al.* (2011). Study of host-microbe interactions in zebrafish. *Methods Cell Biol* **105**: 87–116.

Moon C, Baldridge MT, Wallace MA, Burnham C-AD, Virgin HW, Stappenbeck TS. (2015). Vertically transmitted faecal IgA levels determine extra-chromosomal phenotypic variation. *Nature* **521**: 90–93.

de Muinck EJ, Stenseth NC, Sachse D, vander Roost J, Rønningen KS, Rudi K, *et al.* (2013). Context-Dependent Competition in a Model Gut Bacterial Community. *PLoS One* **8**: e67210.

Münkemüller T, Gallien L, Lavergne S, Renaud J, Roquet C, Abdulhak S, *et al.* (2014). Scale decisions can reverse conclusions on community assembly processes. *Glob Ecol Biogeogr* **23**: 620–632.

Ng ISY, Carr CM, Cottenie K. (2009). Hierarchical zooplankton metacommunities: distinguishing between high and limiting dispersal mechanisms. *Hydrobiologia* **619**: 133–143.

O'Dwyer JP, Kembel SW, Green JL. (2012). Phylogenetic Diversity Theory Sheds Light on the Structure of Microbial Communities. *PLoS Comput Biol* **8**. e-pub ahead of print, doi: 10.1371/journal.pcbi.1002832.

Ofiteru ID, Lunn M, Curtis TP, Wells GF, Criddle CS, Francis CA, *et al.* (2010). Combined niche and neutral effects in a microbial wastewater treatment community. *Proc Natl Acad Sci U S A* **107**: 15345–15350.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, *et al.* (2016). vegan: Community Ecology Package. https://cran.r-project.org/package=vegan.

Olszak T, An D, Zeissig S, Vera MP, Richter J, Franke A, *et al.* (2012). Microbial exposure during early life has persistent effects on natural killer T cell function. *Science* **336**: 489–93.

Östman Ö̈, Drakare S, Kritzberg ES, Langenheder S, Logue JB, Lindström ES. (2010). Regional invariance among microbial communities. *Ecol Lett* **13**: 118–127.

Palmer C, Bik EM, DiGiulio DB, Relman DA, Brown PO. (2007). Development of the human infant intestinal microbiota. *PLoS Biol* **5**: 1556–1573.

Pantoja-Feliciano IG, Clemente JC, Costello EK, Perez ME, Blaser MJ, Knight R, *et al.* (2013). Biphasic assembly of the murine intestinal microbiota during early development. *ISME J* **7**: 1112–1115.

Parichy DM, Elizondo MR, Mills MG, Gordon TN, Engeszer RE. (2009). Normal table of postembryonic zebrafish development: Staging by externally visible anatomy of the living fish. *Dev Dyn* **238**: 2975–3015.

Parks DH, Beiko RG. (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26**: 715–21.

Peterson TS, Ferguson JA, Watral VG, Mutoji KN, Ennis DG, Kent ML. (2013). Paramecium caudatum enhances transmission and infectivity of Mycobacterium marinum and M. chelonae in zebrafish Danio rerio. *Dis Aquat Organ* **106**: 229–39.

Pham LN, Kanther M, Semova I, Rawls JF. (2008). Methods for generating and colonizing gnotobiotic zebrafish. *Nat Protoc* **3**: 1862–1875.

Phillips JB, Westerfield M, Ablain J, Zon LI, Austin-Tse C, Halbritter J, *et al.* (2014). Zebrafish models in translational research: tipping the scales toward advancements in human health. *Dis Model Mech* **7**: 739–43.

Price MN, Dehal PS, Arkin AP. (2010). FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.

Pruesse E, Peplies J, Glöckner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–9.

R Core Team. (2016). R: A Language and Environment for Statistical Computing. https://www.r-project.org/.

Rawls JF, Mahowald MA, Ley RE, Gordon JI. (2006). Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-free Recipients Reveal Host Habitat Selection. *Cell* **127**: 423–433.

Rawls JF, Samuel BS, Gordon JI. (2004). Gnotobiotic zebrafish reveal evolutionarily conserved responses to the gut microbiota. *Proc Natl Acad Sci U S A* **101**: 4596–4601.

Robinson CJ, Bohannan BJM, Young VB. (2010). From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* **74**: 453–476.

Roeselers G, Mittge EK, Stephens WZ, Parichy DM, Cavanaugh CM, Guillemin K, *et al.* (2011). Evidence for a core gut microbiota in the zebrafish. *ISME J* **5**: 1595–1608.

Rogers GB, Kozlowska J, Keeble J, Metcalfe K, Fao M, Dowd SE, *et al.* (2014). Functional divergence in gastrointestinal microbiota in physically-separated genetically identical mice. *Sci Rep* **4**: S56–7.

Rogier EW, Frantz AL, Bruno MEC, Wedlund L, Cohen DA, Stromberg AJ, *et al.* (2014). Secretory antibodies in breast milk promote long-term intestinal homeostasis by regulating the gut microbiota and host gene expression. *Proc Natl Acad Sci* **111**: 3074–3079.

Rosindell J, Hubbell SP, Etienne RS. (2011). The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends Ecol Evol* **26**: 340–348.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, *et al.* (2009). Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl Environ Microbiol* **75**: 7537–7541.

Segata N, Börnigen D, Morgan X, Huttenhower C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nat Commun* **4**: 2304.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, *et al.* (2011). Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Semova I, Carten JD, Stombaugh J, Mackey LC, Knight R, Farber SA, *et al.* (2012). Microbiota Regulate Intestinal Absorption and Metabolism of Fatty Acids in the Zebrafish. *Cell Host Microbe* **12**: 277–288.

Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. (2006). Quantifying the roles of immigration and chance in shaping prokaryote community structure. *Environ Microbiol* **8**: 732–740.

Sloan WT, Woodcock S, Lunn M, Head IM, Curtis TP. (2007). Modeling taxa-abundance distributions in microbial communities using environmental sequence data. In: Vol. 53. *Microbial Ecology*. pp 443–455.

Soergel D a W, Dey N, Knight R, Brenner SE. (2012). Selection of primers for optimal taxonomic classification of environmental 16S rRNA gene sequences. *ISME J* **6**: 1440–4.

Sommer F, Bäckhed F. (2013). The gut microbiota--masters of host development and physiology. *Nat Rev Microbiol* **11**: 227–38.

Spiess A-N, Neumeyer N. (2010). An evaluation of R2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharmacol* **10**: 6.

Spor A, Koren O, Ley R. (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nat Rev Microbiol* **9**: 279–290.

Srivastava DS, Kolasa J, Bengtsson J, Gonzalez A, Lawler SP, Miller TE, *et al.* (2004). Are natural microcosms useful model systems for ecology? *Trends Ecol Evol* **19**: 379–384.

Turnbaugh PJ, Ley RE, Mahowald MA, Magrini V, Mardis ER, Gordon JI. (2006). An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**: 1027–1031.

Ubeda C, Lipuma L, Gobourne A, Viale A, Leiner I, Equinda M, *et al.* (2012). Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J Exp Med* **209**: 1445–1456.

Vellend M. (2010). Conceptual Synthesis in Community Ecology. *Q Rev Biol* **85**: 183–206.

Venkataraman A, Bassis CM, Beck JM, Young VB, Curtis JL, Huffnagle GB, *et al.* (2015). Application of a Neutral Community Model To Assess Structuring of the Human Lung Microbiome. *MBio* **6**: e02284–14.

Vijay-Kumar M, Aitken JD, Carvalho FA, Cullender TC, Mwangi S, Srinivasan S, *et al.* (2010). Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**: 228–31.

Wang Q, Garrity GM, Tiedje JM, Cole JR. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**: 5261–7.

Wen L, Ley RE, Volchkov PY, Stranges PB, Avanesyan L, Stonebraker AC, *et al.* (2008). Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**: 1109–1113.

Werner JJ, Koren O, Hugenholtz P, DeSantis TZ, Walters W a, Caporaso JG, *et al.* (2012). Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *ISME J* **6**: 94–103.

Wickham H. (2009). ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York.

Woodcock S, Van Der Gast CJ, Bell T, Lunn M, Curtis TP, Head IM, *et al.* (2007). Neutral assembly of bacterial communities. *FEMS Microbiol Ecol* **62**: 171–180.

Yan Q, van der Gast CJ, Yu Y. (2012). Bacterial community assembly and turnover within the intestines of developing zebrafish. *PLoS One* **7**: e30603.

Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, *et al.* (2012). Human gut microbiome viewed across age and geography. *Nature*. e-pub ahead of print, doi: 10.1038/nature11053.

Yeh Y-C, Peres-Neto PR, Huang S-W, Lai Y-C, Tu C-Y, Shiah F-K, *et al.* (2015). Determinism of bacterial metacommunity dynamics in the southern East China Sea varies depending on hydrography. *Ecography (Cop)* **38**: 198–212.

Zac Stephens W, Burns AR, Stagaman K, Wong S, Rawls JF, Guillemin K, *et al.* (2016). The composition of the zebrafish intestinal microbial community varies across development. *ISME J* **10**: 644–654.

Zenewicz LA, Yin X, Wang G, Elinav E, Hao L, Zhao L, *et al.* (2013). IL-22 Deficiency Alters Colonic Microbiota To Be Transmissible and Colitogenic. *J Immunol* **190**: 5306–5312.

Zerbino DR, Birney E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–9.

Zha Y, Berga M, Comte J, Langenheder S, Hooper D, Chapin F, *et al.* (2016). Effects of Dispersal and Initial Diversity on the Composition and Functional Performance of Bacterial Communities Li J-T (ed). *PLoS One* **11**: e0155239.

Zhernakova A, Kurilshikov A, Bonder MJ, Tigchelaar EF, Schirmer M, Vatanen T, *et al.* (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science* **352**: 565–9.