BACTERIAL REGULATION OF HOST PANCREATIC BETA CELL DEVELOPMENT

by

JENNIFER HAMPTON HILL

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 2017

DISSERTATION APPROVAL PAGE

Student: Jennifer Hampton Hill

Title: Bacterial Regulation of Host Pancreatic Beta Cell Development

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:

Victoria Herman	Chairperson
Karen Guillemin	Advisor
Judith Eisen	Core Member
Brendan Bohannan	Core Member
Raghuveer Parthasarathy	Institutional Representative

and

Sara D. Hodges	Interim Vice Provost and Dean of the Graduate School

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

Degree awarded September 2017

DISSERTATION ABSTRACT

Jennifer Hampton Hill

Doctor of Philosophy

Department of Biology

September 2017

Title: Bacterial Regulation of Host Pancreatic Beta Cell Development

Diabetes is a metabolic disease characterized by the loss of functional pancreatic beta cells. The incidence of diabetes has risen rapidly in recent decades, which has been attributed at least partially to alterations in host-associated microbial communities, or microbiota. It is hypothesized that the loss of important microbial functions from the microbiota of affected host populations plays a role in the mechanism of disease onset. Because the immune system also plays a causative role in diabetes progression, and it is well documented that immune cell development and function are regulated by the microbiota, most diabetes microbiota research has focused on the immune system. However, microbial regulation is also required for the development of many other important tissues, including stimulating differentiation and proliferation. We therefore explored the possibility that the microbiota plays a role in host beta cell development. Using the larval zebrafish as a model, we discovered that sterile or germ free (GF) larvae have a depleted beta cell mass compared to their siblings raised in the presence of bacteria and other microbes. This dissertation describes the discovery and characterization of a rare and novel bacterial gene, whose protein product is sufficient to rescue this beta cell developmental defect in the GF larvae. Importantly, these findings suggest a possible role for the microbiota in preventing or prolonging the eventual onset

of diabetes through induction of robust beta cell development. Furthermore, the loss of rare bacterial products such as the one described herein could help to explain why low diversity microbial communities are correlated with diabetes.

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

CURRICULUM VITAE

NAME OF AUTHOR: Jennifer Hampton Hill

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Humboldt State University, Arcata

DEGREES AWARDED:

Doctor of Philosophy, Biology, 2017, University of Oregon
Bachelor of Science, Biology, 2010, Humboldt State University

AREAS OF SPECIAL INTEREST:

Microbiology
Animal Physiology and Development

PROFESSIONAL EXPERIENCE:

Graduate Teaching Assistant, University of Oregon, September 2011 - June 2012

Research Technician, Humboldt State University, August 2009 – May 2010

GRANTS, AWARDS, AND HONORS:

Pete von Hippel Graduate Scholar Award, University of Oregon, Institute of Molecular Biology, 2015

Best Flash Talk, University of Oregon, Institute of Molecular Biology Retreat, 2015

Best Poster, Zebrafish for Personalized/Precision Medicine Conference, St. Michael's Hospital, Toronto, CA, 2015

National Institutes of Health (NIH) Genetics Training Grant Appointment, University of Oregon, 2012-2015

California Institute for Regenerative Medicine (CIRM) Bridges Fellow, University of California San Francisco, 2010-2011

PUBLICATIONS:

Hill, JH, Franzosa, EA, Huttenhower, C, and Guillemin, K. (2016). A conserved bacterial protein induces pancreatic beta cell expansion during zebrafish development. eLife, 5, e20145.

Jemielita, M, Taormina, MJ, Burns, AR, Hampton, JS, Rolig, AS, Guillemin, K, and Parthasarathy, R. (2014). Spatial and temporal features of the growth of a bacterial species colonizing the zebrafish gut. mBio vol. 5no. 6 e01751-14.

Arroyo, FA, Siering, PL, Hampton, JS, McCartney, A, Hurst, MP, Wolfe, GV, and Wilson, MS. (2014). Isolation and characterization of novel iron-oxidizing autotrophic and mixotrophic bacteria from boiling springs lake, an oligotrophic, acidic geothermal habitat. Geomicrobiology vol. 32 no. 2. 140-157.

ACKNOWLEDGMENTS

This research would not exist without the profound enthusiasm and guidance of my mentor, Dr. Karen Guillemin, who has always encouraged the development of imaginative scientific questions. I would also like to thank the members of my advisory committee who not only helped to shape my development as a scientist, but also supported me as I explored post-graduate career options. Furthermore, I have had the pleasure of working alongside incredibly creative post-docs, technicians and fellow graduate students who provided training and critical insight for this work. Specifically, fellow graduate student Michelle Sconce has provided assistance and intellectual contribution for experiments in Chapter III and her continued work on this project will surely see fruition on several ideas presented in Chapter IV. Undergraduate, Daniel Derrick also provided technical assistance for many preliminary experiments that laid the groundwork for ideas presented in Chapter IV. Dr. James Remington, Dr. Emily Sweeney, and Karen Kallio were integral in the process of solving the BefA crystal structure presented in Chapter III. Rose Sockol is our zebrafish technician, who's work, along with the entire University of Oregon Zebrafish Facility, was essential for this project.

Lastly, I am grateful to members of my family, who have each been an inspiration to me in my graduate career.

This work is dedicated to my mother, father, and sister. My mother's relentless compassion for helping people, my father's strong love for the natural world, and my sister's limitless creativity have all manifested in this project.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

CHAPTER I

MICROBES, MODELS, AND MALADIES: IDEAS AND TOOLS FOR THE STUDY OF METABOLIC DISEASE

INTRODUCTION

Modern humans predominantly associate microbial organisms with contagious disease. Since the formulation of the germ theory of disease by Pasteur and Koch in the late 1800s, we have taken greater and greater measures to combat the threat of "germs" and prevent infections. Cleanliness of both the body and the home are kept through the regular use of commercial antibacterial products. Thanks to the use of antibiotics and vaccines, the spread of microbial caused illness is now much less prevalent in societies with access to these medical advances, and many life-threatening bacterial infections can thankfully be prevented. However, there seems to be a trade off to these benefits. As the threat of contagion has become low, the incidence of autoimmune disease is unfortunately rising (Bach, 2002). Several theories involving our contemporary practices of cleanliness and limited exposure to microbial organisms have arisen to explain this phenomenon. The Hygiene Hypothesis and the subsequent Old Friend's Hypothesis predict that childhood exposure to diverse microbial antigens is required for the development and function of a healthy immune system (Rook et al., 2004). Similarly, the Disappearing Microbiota Hypothesis suggests that rare microbial species, which were commonly associated with ancient humans, play specific and essential roles in our wellbeing (Blaser and Falkow, 2009). Unfortunately, due to reductions in both vertical

and horizontal transmission over many generations, we have slowly lost our associations with these important organisms (Blaser and Falkow, 2009). All of these theories suggest that exposure to microbes is beneficial, and that most bacteria are not agents of disease. In fact, by removing these "germs", we may have lost an essential component of our overall health.

THE MICROBIOTA: PROVIDING NEW INSIGHT ON HUMAN DISEASE

Our association with microbes is intimate. Unlike the picture that germ theory paints, bacteria and viruses aren't something that we get on us every once in a while from touching something dirty. Rather, we are steeped in microbes from the day of our birth (Mueller et al., 2015), and this assemblage of microscopic beings that inhabit our bodies and those of other animals is collectively called the microbiota. It is comprised of a diverse array of taxonomic groups including species of bacteria, fungi, single celled eukaryotes, and viruses (Morgan et al., 2013). This complex collection of organisms is often referred to as a community, as different members perform different functions that contribute to its overall ecological health (Huttenhower et al., 2012). In the case of the microbiota, the ecological environment is the host animal, which is in turn reliant upon the functions of this microbial community for its own health (McFall-Ngai et al., 2013). So far, the majority or our understanding about the functions of the microbiota comes from studying bacteria. In 2008, the National Institutes of Health (NIH) launched the human microbiome project (HMP). The initial phases of this project involved massive bacterial sampling and sequencing efforts focused on identifying the species associated with different sites of the human body in both health and disease (Integrative HMP

(iHMP) Research Network Consortium, 2014). These and many other metagenomic studies revealed that the community membership of host-associated bacteria often differs significantly between healthy and sick people, with decreased taxonomic diversity being a common feature of disease states (DeGruttola et al., 2016). Animal modeling of disease-associated microbiota provides evidence that the microbial community is capable of causing or perpetuating the disease (Gülden et al., 2015; Ley et al., 2006; McFall-Ngai et al., 2013). The transformation from a healthy to a disease-causing community is called dysbiosis (Levy et al., 2017), and identifying the bacterial changes that occur in dysbiotic communities is a current focus of microbiota research that holds promise for developing many new disease treatments (Petersen and Round, 2014).

The revelation that the microbiota is an understudied and critical component to animal health has resulted in a wave of publications that collectively implicate an almost unimaginable diversity of requirements for bacteria or other microbes in animal health. However, these discoveries become less surprising when we consider the fact that animals evolved in a world dominated by microbial organisms (McFall-Ngai et al., 2013). These tiny creatures ruled and shaped our environment for billions of years before the evolution of multicellular life (Grosberg and Strathmann, 2007). To survive in this microbial environment, eukaryotes must have evolved mechanisms to "eavesdrop" on microbes for important information about their surrounding ecosystem, perhaps through the detection of secreted microbial molecules. For example, nutritional availability can be interpreted through the metabolic by-products of bacteria breaking down various carbon sources. One possible scenario is that multicellular life greatly benefited by "listening in" on the activities of microbes, and as a result, became dependent upon them for important

processes. Today, we know that animals rely on bacteria for regulating their own metabolic homeostasis (Nicholson et al., 2012). When these important bacterial functions go missing from the microbiota, the host is thrown out of balance and oftentimes succumbs to disease (DeGruttola et al., 2016).

DIABETES: A CASE STUDY FOR THE DISAPEARING MICROBIOTA HYPOTHESIS

Diabetes is a group of metabolic diseases classified by the loss of functional insulin, which most commonly manifests through autoimmune destruction of beta cells or insulin resistance in either type 1 or type 2 diabetes respectively (Tai et al., 2015). These diseases are prominent human ailments that are strongly correlated with decreased microbiota diversity (Brown et al., 2011; Giongo et al., 2011; Kostic et al., 2015). In Type 1 diabetes (T1D), monozygotic twins have a discordance rate of about 50%, implicating a prominent environmental cause for the disease in addition to genetic predisposition (Akerblom et al., 2002). Intriguingly, the environmental risk factors correlated with diabetes are largely modern practices of developed countries, which include a high use of antibiotics, anti-bacterial cleansers and soaps, vaccines, cesarean sections, and a high fat diet (Gülden et al., 2015). These are all factors predicted to contribute to our declining interaction with important microbes as described by the Disappearing Microbiota Hypothesis (Blaser and Falkow, 2009). In the case of type 1 diabetes (T1D), the loss of microbial-associated diversity occurs prior to disease onset, often before the age of three years, and is believed to be at least partially causative (Kostic et al., 2015). Animal models have suggested that early alterations to the gut

microbiota, during important developmental events, can impact the health of the host later in life (Cox et al., 2014). In rodent models of T1D, treatment with antibiotics during early life, results in greater incidence and earlier onset of disease than in genetically identical siblings not treated with antibiotics (Brown et al., 2016; Candon et al., 2015). During this early age, beta cell mass is expanding through avenues of both differentiation and proliferation in order to establish a robust insulin production center to efficiently regulate glucose homeostasis into adulthood (Gregg et al., 2012; Hesselson et al., 2009). Insufficient beta cell mass expansion can contribute to diabetes onset (Berger et al., 2015; Figliuzzi et al., 2010) in rodents, and we should investigate whether there is a connection between this process and the microbiota in order to develop new therapeutic approaches to diabetes treatment and prevention.

Unfortunately, rates of diagnoses for both types of diabetes have been rising substantially over the last few decades (Tai et al., 2015), making the need for effective therapeutics greater than ever. Because of the loss of insulin, T1D and late stage T2D have no cure and are much more difficult to control. Currently, the only effective treatments are exogenous insulin injection and islet transplantation (Farney et al., 2016). Insulin injection is only effective at managing symptoms, and requires vigilant monitoring of blood glucose levels. Islet transplantation is only moderately effective, with roughly 50% of recipients remaining insulin independent up to 5 years post transplant (Farney et al., 2016). Furthermore, cadaver islet tissue available for transplantation is rare, and once transplanted, the foreign beta cells are still at the mercy of the patient's malfunctioning immune system in the case of T1D (Farney et al., 2016; Johannesson et al., 2015). Alternatively, many avenues of research are focused on

expanding a patient's endogenous beta cell population in order to restore insulin supply (Johannesson et al., 2015). There are two primary mechanisms of normal beta cell expansion: neogenesis (differentiation from progenitors) and self-proliferation (Johannesson et al., 2015). Researchers hope that a better understanding of these regular developmental events can lead to treatments that will induce the expansion of the functional beta cell mass *in vivo*. Because of the strong correlation between early beta cell expansion and dysbiosis, research into whether or not there is a microbial role in beta cell development is worthwhile and could provide new avenues for diabetes treatments. Furthermore, in a less diverse community, loss of bacterial functions important for host beta cell development could result in reduced beta cell mass and render some individuals more susceptible to diabetes.

ZEBRAFISH: MODELING MICROBIOTA AND METABOLIC DISEASE

In order to understand how dysbiosis contributes to disease, we need to study the molecular mechanisms that contribute to bacterial mediated changes in host physiology. Metagenomic-sequencing efforts inform us about the major bacterial lineages present in a given community, but they can not be used to dissect the complex molecular interactions that occur between the microbiota and its host environment. Techniques developed for the derivation of completely sterile or "germ free" model organisms have allowed researchers to make substantial headway in learning about important host-microbe interactions (McFall-Ngai et al., 2013). Powerful gnotobiotic techniques have been developed in several animal models, including mice, zebrafish, stickleback, flies, hydra, and even the ancestor of multicellular organisms, Choanoflagellates. (McFall-Ngai et al.,

2013; Milligan-Myhre et al., 2016). Studies in these organisms have revealed surprising roles for bacteria in host biology, most notably including in development, immune system function, and metabolic homeostasis (McFall-Ngai et al., 2013). Each of these different animal models brings its own unique advantages and tools to microbiota research.

The zebrafish offers specific strengths for modeling human metabolic diseases of host-microbiota interactions. Historically, it has served is an excellent model for observing and imaging vertebrate development thanks to the strong collection of genetic tools that have been developed in this platform. Zebrafish also have extremely high fecundity, relatively low maintenance costs, and are simple to derive germ free in large numbers (2017). Furthermore, the intestinal microbiota of the zebrafish is well described, and culture isolates are available that are representative of most of the major genera identified by sequencing analyses (Stephens et al., 2015).

Metabolic homeostasis is perhaps influenced by the microbiota in more ways than any other host process. Conveniently, many of the same attributes that make zebrafish a strong developmental model, such as larval transparency and conserved organ function, also present the opportunity to study metabolic function (Gut et al., 2017). For instance, a variety of fluorescent lipophilic dyes are available that allow for the tracking of fatty acids and lipid accumulation in vivo (Santoro, 2014). Several mutant lines provide models that recapitulate aspects of human metabolic diseases (Seth et al., 2013), and established transgenic lines allow for the observation of metabolically important cell types (Gut et al., 2017). Glucose metabolism and gluconeogensis can be tracked to give an accurate readout of endocrine pancreas and liver function respectively (Andersson et al., 2012; Gut et al., 2013). Even overall metabolic rate and oxidative stress can also be

determined in the zebrafish using commercial kits that function on both live and fixed cells (Santoro, 2014). Since the knowledge of the vast significance of microbes for host health is relatively new, only a handful of important studies have so far utilized these zebrafish tools to study the interplay between bacteria and host metabolic function. However, the following early observations illustrate that the zebrafish is a powerful model that can be utilized for furthering our understanding of the mechanisms that give rise to altered host physiology and subsequently disease.

Fat absorption and storage is one area of study where the strengths of the zebrafish have been shown to be particularly amenable to interrogate the interactions between bacteria and host physiology. Not only can researchers visualize lipid accumulation defects (Santoro, 2014), they can track adipose tissue development (Minchin and Rawls, 2017). The Rawls lab has studied the role of the microbiota in regulating intestinal lipid absorption from the diet. They showed that bacteria induce the formation of lipid droplets within the intestinal epithelia through at least two separate mechanisms (Semova et al., 2012a). First, the size of lipid droplets was regulated by the presence of microbes, and second, during feeding Firmicutes species were sufficient to cause an increase in the formation of the number of lipid droplets within the epithelia (Semova et al., 2012a). The authors also showed that this increased lipid accumulation resulted in increased export of lipids to systemic tissues, suggesting that microbial composition can regulate the efficiency and rate of lipid absorption (Semova et al., 2012b). It will be valuable to continue studies like this one to see if changes in the microbiota result in altered lipid transport to various regional ATs, in order to eventually understand how we can utilize bacterial species to modulate fat absorption to attenuate

metabolic diseases such as obesity and cachexia. To aid in this goal, Minchin and Rawls recently released a detailed characterization of zebrafish adipose tissue (AT) that includes a comprehensive description of the anatomical location, development, and size of 34 adipose tissue types in the zebrafish (Minchin and Rawls, 2017). The authors go on to compare adipose tissues between common lab-reared fish strains and evaluate the general morphological changes of AT to a high fat diet (Minchin and Rawls, 2017). Alterations to lipid accumulation in regional AT types is a significant risk factor leading to obesity and diabetes and this new characterization will allow researchers to explore the specific ways in which adipose tissue is influenced by microbial factors (Minchin and Rawls, 2017).

Diet is an important factor in nutrient availability and variation in host diet also plays an important role in how microbes can influence host metabolic processes (Nicholson et al., 2012). For instance, a recent study by Gou and colleagues using a larval zebrafish model found that a diet of nucleotides led to decreased metabolic rate through two microbiota-mediated mechanisms. First, microbial suppression of *fiaf* gene expression led to decreased fatty acid (FA) oxidation in the muscle and liver (Guo et al., 2017). The authors also saw decreased inflammation in the head kidney, suspected to further reduce metabolic rate through decreased immune cell activity (Guo et al., 2017). They showed that these phenotypes corresponded to a higher percentage of obligate anaerobes within the microbiota (Guo et al., 2017). Further studies should focus on identifying the microbial mechanism required for this altered host gene expression, and determining whether these changes alter early development of the larval muscle or liver.

Several studies from the Chen lab have also described an effect of diet on zebrafish metabolism. They originally observed that a variety or rich carbohydrate

sources, including glucose and egg yolk, in the larval diet resulted in altered glucose regulation (Maddison and Chen, 2012). They later found that greater glucose availability from these high carbohydrate diets caused an increased demand for insulin, which was alleviated by an eventual increase in total pancreatic beta cells, the source of insulin in the body (Li et al., 2015). Subsequently they implicated FGF1 signaling in the induction of beta cell differentiation, showing that diet and development of this important cell population are linked (Li et al., 2015). The authors did not look at the role of the microbiota in this process, however this would be an easy question to investigate given the ability to study both microbes and beta cell development in the larval zebrafish (Bates et al., 2006; Kinkel and Prince, 2009). Furthermore, as explained earlier, because beta cell function is lost in both T1 and T2D, uncovering novel mechanisms of beta cell renewal is a primary goal for the development of diabetes treatments in order to restore normal-glycaemia without the use of exogenous insulin (Johannesson et al., 2015).

We have outlined the work that has laid a strong foundation for the use of zebrafish to study the interface between metabolism and associated microbes. Although there are few studies to cover thus far, it is clear that the potential for continued research in these areas, as well as new research into other categories of metabolic function would benefit from this model system. Toward this end, this dissertation will utilize the larval zebrafish system in order to investigate a role for host-associated bacteria in beta cell regulated glucose metabolism.

BRIDGE

Research using gnotobiotic zebrafish has helped to illustrate the important role that associated bacteria play in host developmental and metabolic processes (Guo et al., 2017; Semova et al., 2012a). In Chapter 1, we reviewed the studies that have utilized zebrafish to investigate the interactions between metabolism and microbes. In the field of development, our lab has shown that bacteria are required for enterocyte proliferation (Cheesman et al., 2011) and goblet cell differentiation (Bates et al., 2006). However, little work has been done in any model organism to look at the influence of bacteria on the development of other systemic tissues that play a large role in metabolic homeostasis, such as the pancreas. Given the connections between early beta cell expansion, loss of microbial diversity, and diabetes onset in humans, this cell population is also an important area of study for human health. In the next chapter, we use the tools available in the zebrafish system to investigate whether bacteria play a role in beta cell development and glucose regulation.

Chapter II of this dissertation contains previously published, co-authored material reproduced herein with the permission from Jennifer Hill and Karen Guillemin.

Chapter III contains unpublished co-authored material presented with permission from S. James Remington, Emily Sweeney, and Karen Kallio.

CHAPTER II

A CONSERVED BACTERIAL PROTEIN INDUCES PANCREATIC BETA CELL EXPANSION DURING ZEBRAFISH DEVELOPMENT

 I was the primary contributor to this work, carrying out the experiments, analysis and writing. My mentor and principal investigator on the paper, K. Guillemin, contributed to experimental design, analysis and writing. Authors E. Franzosa and C. Huttenhower provided unique expertise for Figure 6, panels A and B.

INTRODUCTION

Host-associated microbes play important roles in the development of animal digestive tracts (Bates et al., 2006; Semova et al., 2012a; Sommer and Bäckhed, 2013). Using the gnotobiotic zebrafish model, our group has shown previously that resident microbes promote host processes in the developing intestine such as epithelial differentiation (Bates et al., 2006) and proliferation (Cheesman et al., 2011). The role of microbes in the development of other digestive organs remains underexplored, despite the fact that many diseases in peripheral digestive organs are correlated with microbial dysbiosis (Chang and Lin, 2016; Gülden et al., 2015). The ability to manipulate resident microbes in the larval zebrafish (Milligan-Myhre et al., 2011), combined with the optical transparency and sophisticated genetic tools of the zebrafish model, make it a powerful

platform to investigate this question. Here, we use gnotobiotic zebrafish to demonstrate a role for resident microbes in promoting pancreatic beta cell development.

The zebrafish has a well-characterized program of beta cell development, which is highly conserved with that of mammals (Kinkel and Prince, 2009). In the zebrafish embryo, initial beta cells arise from precursors within the dorsal and ventral pancreatic buds (Biemar et al., 2001; Field et al., 2003; Wang et al., 2011). The two buds fuse by 52 hours post fertilization (hpf), and give rise to the fully fated pancreas with only a single islet of hormone-secreting endocrine cells, by 3 days post fertilization (dpf) (Biemar et al., 2001; Field et al., 2003; Kumar, 2003). Coinciding with the approximate time of larval emergence from the chorion by 3 dpf, these newly fated beta cells begin to expand (Chung et al., 2010; Dong et al., 2007; Hesselson et al., 2009; Kimmel et al., 2011; Moro et al., 2009). beta cells derived from the dorsal bud become quiescent, while ventral bud derived beta cells begin to undergo expansion via mechanisms of both proliferation and neogenesis (Hesselson et al., 2009). Between 3 and 6 dpf, the number of beta cells within the primary islet will almost double (Moro et al., 2009). Intestinal colonization with microbes occurs concurrently with this early larval period of beta cell expansion. Following development of the gut tube within the sterile embryo, the intestine of the emergent larva becomes open to the environment at both the mouth and the vent by 3.5 dpf, allowing for inoculation by environmental microbes (Bates et al., 2006). Within the larval gut, bacteria proliferate rapidly, such that a single species in mono-association can reach the luminal carrying capacity within several hours (Jemielita et al., 2014).

Human post-natal beta cell expansion also occurs concurrently with intestinal tract colonization by commensal microbes. *In utero*, beta cells are produced via

differentiation from progenitors (Georgia et al., 2006; Stanger et al., 2007) and at birth this newly fated cell population begins to expand by self-proliferation (Georgia and Bhushan, 2004; Gregg et al., 2012; Kassem et al., 2000; Teta et al., 2007). beta cell proliferation rates peak at 2 years of age and then steadily decline (Gregg et al., 2012). By 5 years of age, most of the beta cell mass has become slow cycling and will not expand significantly again unless stimulated by elevated metabolic demands, such as obesity or pregnancy. At birth, infants are exposed to their mothers' vaginal, fecal and skin associated microbes, which immediately begin to colonize the neonatal intestine (Biasucci et al., 2010; Dominguez-Bello et al., 2010; Palmer et al., 2007). By 3 years of age, the composition and complexity of the microbiota typically resembles that of an adult associated community (Murgas Torrazza and Neu, 2011; Palmer et al., 2007; Yatsunenko et al., 2012). However, factors such as diet, birth mode and antibiotic exposure can result in reduced microbial taxonomic diversity during these early years of life (Mueller et al., 2015). Notably, factors that reduce microbiota diversity are also associated with increased risk for diabetes mellitus (Knip et al., 2005). Loss of beta cell function through autoimmunity results in abnormal glucose homeostasis and is the cause of type 1 diabetes (T1D) in humans. Recent studies have shown that decreased taxonomic diversity of the intestinal microbiota is correlated with T1D (Brown et al., 2011; Giongo et al., 2011). Indeed, loss of bacterial diversity precedes the onset of T1D in children, and may play a causative role in disease (Kostic et al., 2015).

To our knowledge, no one has yet investigated a role for the gut microbiota in the development of pancreatic beta cells. Communication between the intestine and the pancreas is critical for overall homeostasis. The two organs are therefore connected

physically, metabolically, and developmentally in order to carry out their essential functions. We propose that this established and important connection might also mediate the influence of resident microbes on developmental processes in the pancreas. Here we examine the effects of microbial colonization on initial expansion of zebrafish primary islet beta cells. We find that beta cell mass expansion, up to at least 6 dpf, is promoted by the presence of the microbiota. Using a culture collection of zebrafish intestinal bacteria, we show that certain strains can restore beta cell expansion in germ free (GF) fish. We report the discovery of a secreted protein, shared among these strains and named herein beta cell expansion factor A (BefA) that is sufficient to recapitulate this effect. Homologs of the *befA* gene are present in the genomes of a subset of human intestinal bacteria, and we show that two of the corresponding proteins share BefA's capacity to induce beta cell expansion in zebrafish.

RESULTS

The microbiota is required for normal expansion of the larval beta cell mass

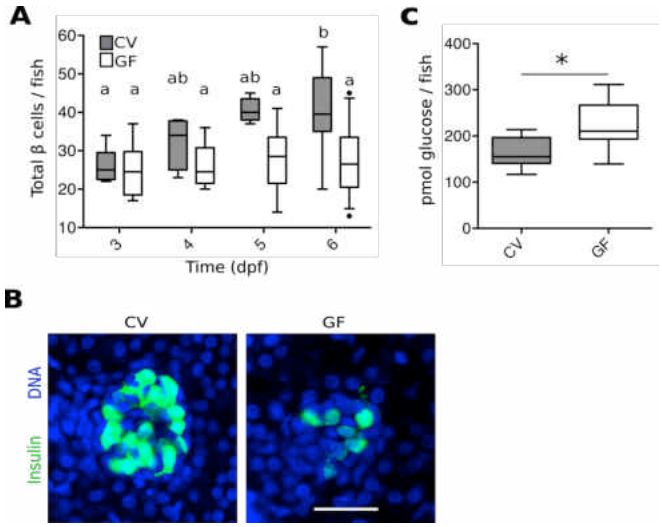
To investigate a possible role for the microbiota in pancreas development and specifically in beta cell expansion, we quantified total beta cells in GF and conventionally reared (CV) *Tg(-1.0insulin:eGFP)* fish (diIorio et al., 2002) at 3, 4, 5 and 6 dpf (Figure 1A, Figure 1 – source data 1). The number of beta cells in CV fish increased steadily from 3 to 6 dpf (Figure 1A). However, the average number of beta cells in GF fish remained static over this time (Figure 1A). Furthermore, at 6 dpf, the overall structure of beta cells within the primary islet also appeared much less densely packed in GF than in

CV fish (Figure 1B). This effect is not likely to be due to changes in initial differentiation of the beta cell population since the total number of beta cells is not different between GF and CV fish at 3 dpf (Figure 1A), a time at which exposure to bacteria is also limited.

Because insulin from beta cells functions to reduce levels of circulating glucose, we tested whether the beta cell deficiency in GF larvae at 6 dpf affected the metabolic function of the fish by measuring free glucose levels. The amount of glucose detected in GF fish was significantly higher than in CV fish (Figure 1C, Figure 1 – source data 2). These data suggest that GF fish, with a paucity of beta cells, are less efficient at importing and processing glucose from the blood due to lower levels of circulating insulin. This is consistent with previous studies showing free glucose levels in zebrafish larvae to be correlated with beta cell numbers (Andersson et al., 2012).

Figure 1. The microbiota is required for normal expansion of the larval beta cell mass (next page)

(A) Total number of beta (β) cells per larva in GF (white box plots) and CV (grey box plots) fish at 3, 4, 5 and 6 dpf. In this, and in all subsequent figures, CV data are shown in grey box plots, and GF data, or statistically similar treatment groups, are shown in white box plots. In all relevant panels and remaining figures, box plot whiskers represent the 95% confidence interval of the data set. Single factor ANOVA indicates that gnotobiology of the fish was significant in determining the number of beta cells present (F_7=9.01, p=1.45e^{-8}). Labels a, ab and b indicate results of post hoc means testing (Tukey). The difference between GF and CV cell counts became significant at 6 dpf (t=-5.91, p<0.001). **(B)** Representative 2D slices from confocal scans through the primary islets of 6 dpf CV and GF *Tg(-1.0insulin:eGFP)* larvae. Each slice is taken from the approximate center of the islet structure. Insulin promoter expressing beta cells are in green and nuclei are blue. Scale bar = 40μM. **(C)** Average amount of glucose (pmol) per larva aged 6 dpf (* t_{17}=-3.65, p<0.01).



Only specific bacterial members of the zebrafish microbiota are sufficient to rescue normal expansion of the GF beta cell mass

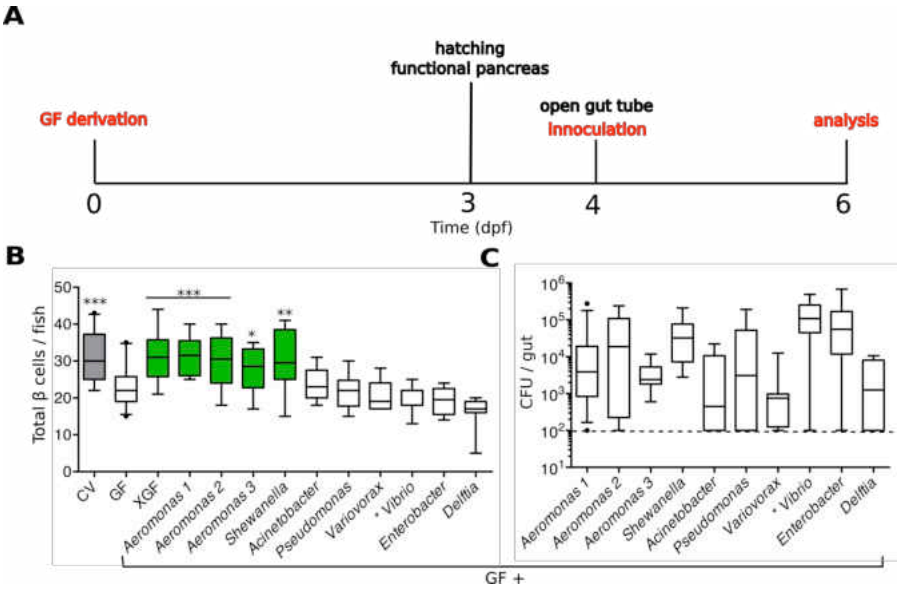
We developed an experimental timeline, depicted in Figure 2A, to test the capacity of individual zebrafish bacterial isolates to induce beta cell expansion. We derived embryos GF at 0 dpf and allowed them to develop in this environment until after hatching. At 4 dpf, when the GF larvae have a patent gut tube, we inoculated them with defined microbes and/or microbial derived products by adding these directly to the embryo media. The fish were incubated with the treatment of interest for 48 hours before analysis of the beta cell mass at 6 dpf.

We found that we could rescue beta cell numbers to CV levels by the addition of non-sterile, normal fish tank water to GF larvae at 4 dpf (Figure 2B, Figure 2 – source

data 1), suggesting that development of the normal number beta cells is dependent upon microbes or microbial-derived products present in the water. We next inoculated 4 dpf GF larvae with a selection of bacterial isolates from the zebrafish intestine (Stephens et al., 2015) as well as one other related strain (Bomar et al., 2013). We prioritized bacterial strains that were capable of forming robust mono-associations with larvae between 4 and 6 dpf, as measured by the number of bacteria found within the gut at 6 dpf (Figure 2C). We found that mono-association with three different species of the genus *Aeromonas* and one species of the genus *Shewanella* was sufficient to rescue GF beta cell numbers to levels observed in CV fish (Figure 2B, Figure 2 – source data 1). Importantly, other isolates such as *Vibrio sp.* and *Delftia sp.* were not sufficient to rescue this phenotype (Figure 2B, Figure 2 – source data 1), indicating that only specific members of the microbiota are capable of inducing expansion of the beta cell mass.

Figure 2. Specific bacterial members of the zebrafish microbiota are sufficient to rescue normal expansion of the GF beta cell mass. (next page)

(A) Experimental timeline for all subsequent zebrafish experiments, unless stated otherwise. Experimental manipulations are denoted by red text. Important zebrafish developmental events are denoted by black text. **(B)** Quantification of beta cells in CV, GF and GF larvae treated at 4 dpf with either non-sterile tank water (XGF) or mono-associated with a specific bacterial strain. Bacterial mono-associations are labeled by genus. Different *Aeromonas sp* are labeled with a number (1, 2 or 3). *$p<0.05$, **$p<0.01$, ***$p<0.001$: Denotes treatment that is significantly different than GF by Tukey analysis. Additionally, here and in all subsequent figures, significant data sets ($p<0.05$ when compared to GF) are also highlighted as green box plots. **(C)** Bacterial isolates of the zebrafish gut and related strains are capable of forming mono-associations with larvae from 4 to 6 dpf. Quantification of the colony forming units (CFUs) per gut for each bacterial strain, assayed after 48-hour exposure to GF larvae. Dashed line denotes the limit of detection.



Aeromonas secretes a factor that rescues normal expansion of the GF beta cell mass

Bacterial interactions with host organisms often involve secreted molecules. To test whether a secreted bacterial factor(s) could influence beta cell expansion, we harvested cell free supernatant (CFS) from overnight cultures of each *Aeromonas* strain shown to rescue beta cell expansion (Figure 2B) and added these to GF larvae at 4 dpf. For each of the three strains of *Aeromonas* tested, the CFS alone was able to restore beta cell numbers in GF fish (Figure 3A, Figure 3 – source data 1), indicating that a secreted factor (or factors) produced by these bacteria is (are) sufficient to induce beta cell expansion. As a control, we also treated GF fish with CFS from a *Vibrio sp*. isolate, which colonized the zebrafish gut (Figure 2C, *), but did not induce beta cell expansion (Figure 2B, *). We found the number of beta cells in fish receiving *Vibrio* CFS was not

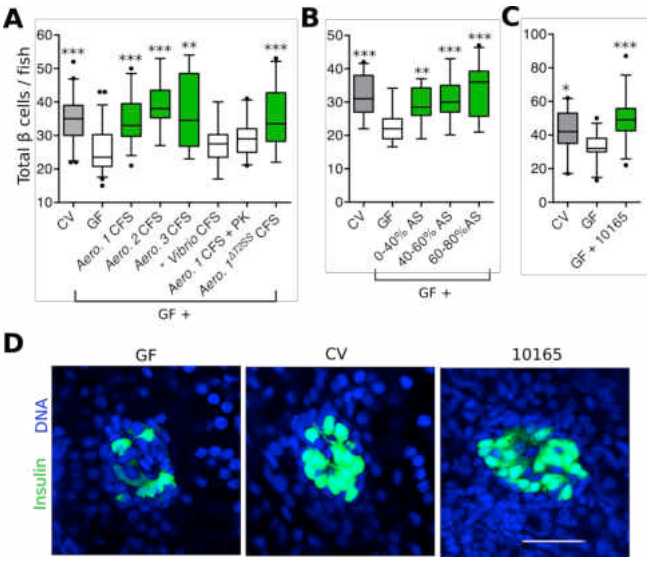
significantly different from that of GF fish (Figure 3A, Figure 3 – source data 1). Furthermore, the capacity to induce increased beta cell numbers was lost when the *Aeromonas 1* (*A. veronii*) CFS sample was treated with proteinase K (Figure 3A, Figure 3 – source data 1), indicating that our secreted factor(s) of interest was likely to be a protein. Because of existing genetic reagents available for the *A. veronii* strain (Bomar et al., 2013), and its capacity to modulate traits of gnotobiotic zebrafish and other hosts (Bates et al., 2006; Cheesman et al., 2011; Graf, 1999; Rolig et al., 2015), we focused on this strain for the remainder of our analysis.

To narrow down the list of candidate proteins secreted by *A. veronii*, we tested whether the activity was present in the CFS of an *A. veronii*$^{\Delta T2SS}$ mutant strain (Maltz and Graf, 2011) lacking a functional type 2 secretion system (T2SS), one of the major protein secretion pathways of Gram-negative bacteria. Despite the fact that it has a reduced secretome, CFS harvested from the *A. veronii*$^{\Delta T2SS}$ strain was sufficient to rescue GF beta cell numbers (Figure 3A, Figure 3 – source data 1). Conveniently, this finding significantly decreased the number of candidate secreted *A. veronii* proteins with beta cell expansion capacity. This result also suggested that our protein(s) of interest was secreted through an alternative mechanism.

We next used ammonium sulfate precipitation to further separate proteins within the *A. veronii*$^{\Delta T2SS}$ CFS. Each of the fractions was able to increase beta cells in GF fish (Figure 3B, Figure 3 – source data 2), suggesting that either *A. veronii*$^{\Delta T2SS}$ produces multiple proteins with this activity, or that the effector was present to some extent within each fraction. Since the 60-80% fraction was able to induce the greatest increase in beta cell numbers (Figure 3B), we used mass spectrometry to analyze the content of this

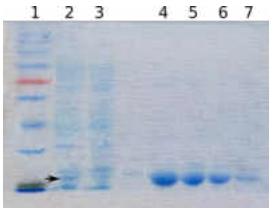
fraction, which led to identification of 163 proteins. To identify promising candidates from this list, we took advantage of the fact that our zebrafish-associated bacterial isolates, for which we have draft genome sequences (Stephens et al., 2015), differed in their capacity to induce beta cells (Figure 2B). Using basic local alignment search tool (BLAST) we identified those proteins from our candidate list that were, first, predicted to be encoded by the genomes of the four bacterial strains with beta cell expansion capacity, and second, absent from the strains lacking this capacity. Our analysis identified one single candidate gene, denoted by the locus tag, *M001_10165 (10165),* predicted to encode a putative protein of 261 amino acids. Consistent with the candidate protein being found in the CFS, the putative protein contained a predicted N-terminal secretion sequence.

Figure 3. *Aeromonas* secretes a factor that rescues normal expansion of the GF beta cell mass. (next page)
(A) Total beta cell numbers in GF, CV and GF fish treated at 4 dpf with different cell free supernatant (CFS) samples. "*Aero."* refers to bacteria of the genus *Aeromonas*, with each number (1, 2, 3) denoting a separate species. *"*+ PK" indicates proteinase K addition to the CFS sample prior to treatment. *p<0.05, **p<0.01, ***p<0.001: Denotes treatment that is significantly different than GF by Tukey analysis. **(B)** Total beta cell numbers in CV, GF and GF fish treated at 4 dpf with separate ammonium sulfate fractions (% AS) prepared from the *Aeromonas* $1^{\Delta T2SS}$ CFS. Note that the 60-80% ammonium sulfate fraction resulted in the greatest increase in beta cell numbers. **(C)** Total beta cells in GF, CV and GF fish treated with purified protein. 10165 represents purified protein from the *M001_10165* locus. **(D)** Representative 2D slices from confocal scans through the primary islets of GF, CV and 10165 protein treated *Tg(-1.0insulin:eGFP)* 6 dpf larvae. Insulin promoter expressing beta cells are shown in green and nuclei are blue. Scale bar = 40μM.



To test whether *10165* encoded the secreted protein responsible for inducing beta cell expansion, we cloned the gene into an inducible expression vector in *E. coli* strain BL21, which contains no *10165* homologues in its genome. We expressed and purified the 10165 protein to homogeneity, as confirmed by SDS-page gel electrophoresis (Figure 3 – figure supplement 1). Purified protein was added to flasks of 4 dpf GF zebrafish larvae. This treatment was sufficient to rescue beta cell numbers to CV levels by 6 dpf (Figure 3C, Figure 3 – source data 3). The islets of larvae treated with the purified protein were visibly expanded compared to those of GF animals (Figure 3D). Therefore, we have named this protein beta cell expansion factor A (BefA) after its observed activity in zebrafish.

Figure 3S – figure supplement 1.10165 (BefA) protein purification. (next page)
SDS-page gel image showing subsequent steps in the purification of BefA (black arrowhead) from *E. coli* cell lysate; lane 1: ladder, lane 2: cell lysate after IPTG induction, lane 3: supernatant from cell lysate after addition of nickel beads, lanes 4-7: elutions of BefA from beads.



BefA is required for Aeromonas *to induce GF beta cell expansion*

To determine whether the *befA (10165)* locus is necessary for *A. veronii* to induce an increase in beta cell numbers, we generated an *A. veronii*$^{\Delta befA}$ mutant strain by replacing the coding region of *befA* with a chloramphenicol resistance gene. To ensure that the loss of the *befA* gene would not affect the ability of *A. veronii* to form mono-associations with larvae, we performed growth and colonization assays and saw no deficiency in either the *in vitro* growth rate (Figure 4 – figure supplement 1A) or the ability of *A. veronii*$^{\Delta befA}$ to colonize the GF intestine compared to the wild-type (WT) strain (Figure 4A). However, when inoculated in a 1:1 ratio together with *A. veronii*^{WT}, the *A. veronii*$^{\Delta befA}$ strain showed a small yet reproducible fitness disadvantage as measured by colonization level and competition index after 48 hours (Figure 4 – figure supplement 1B, C). This result indicates that BefA confers some colonization benefit for *A. veronii* within the larval gut.

GF fish were mono-associated with the *A. veronii*$^{\Delta befA}$ strain, or treated with its CFS from 4 to 6 dpf. Neither treatment was sufficient to rescue beta cell numbers to CV levels (Figure 4B, Figure 4 – source data 1). However, mono-associations of *A. veronii*$^{\Delta befA}$ could be complemented in trans with either CFS from *A. veronii*^{WT} or purified BefA protein, which resulted in restoration of the beta cell population (Figure 4B, Figure

4 – source data 1). Taken together, these data demonstrate that the BefA protein is necessary in an *A. veronii* mono-association for early beta cell expansion and suggests that *A. veronii* only produces a single effector of host beta cell expansion.

Figure 4. BefA is required for *Aeromonas* to induce GF beta cell expansion.
(A) Quantification of the colony forming units (CFUs) per gut in GF fish mono-associated (MA) with either wild type (WT) or mutant (*ΔbefA*) *A. veronii* strains for 48 hours. Dashed line denotes the limit of detection **(B)** Total beta cells in GF fish that have been mono-associated with *ΔbefA*, treated with CFS from either WT or *ΔbefA*, treated with purified BefA, or have been inoculated with a combination of these. **p<0.01, ***p<0.001: Denotes treatment that is significantly different than GF by Tukey analysis.

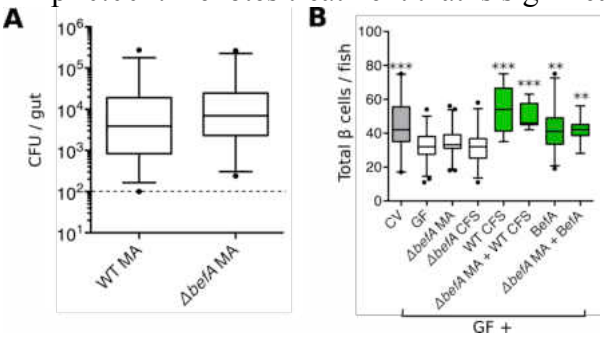
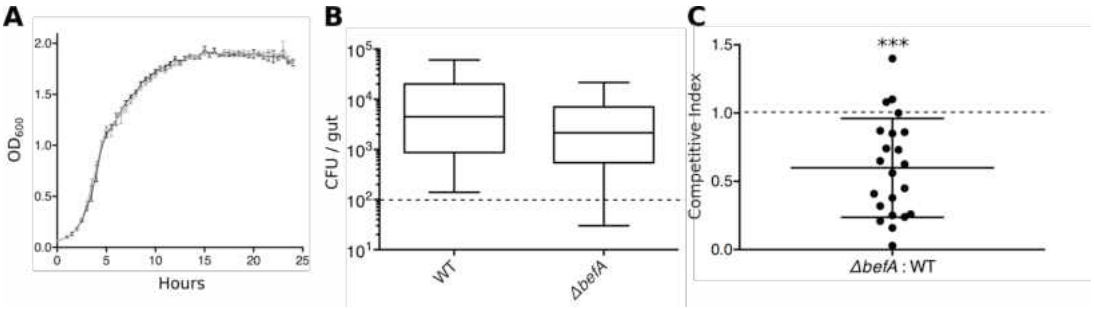


Figure 4S – figure supplement 1. *BefA* confers a colonization advantage in the larval zebrafish gut. (next page)

(A) Growth rates of *A. veronii*^{WT} (black trace) and *A. veronii*$^{\Delta befA}$ (grey trace) *in vitro.* Density measurements (OD_{600}) were taken every half hour for 25 hours on three replicate cultures grown in Lauria broth. **(B)** Resulting CFU's of *A. veronii*^{WT} (WT) and *A. veronii*$^{\Delta befA}$ (*ΔbefA*) within the 6dpf larval gut after inoculation with a 1:1 ratio of each strain at 4 dpf. Dashed line denotes the limit of detection. **(C)** Competitive index (CI) calculation for data within panel B. CI value was calculated for each fish (n=22) by dividing the ratio of mutant to WT bacteria within each gut by 6 dpf, divided by the ratio of mutant to WT bacteria used to inoculate the fish at 4 dpf. A one-sample t-test indicates that the mean CI value is significantly less than 1 (dashed line) (***t_{21}=-3.21, p<0.0001.) A CI value of 1 is expected if no competition exists.



BefA facilitates beta cell expansion by inducing proliferation

Proliferation is the primary mode of human neonatal beta cell expansion (Gregg et al., 2012; Kassem et al., 2000; Teta et al., 2007). In 4-6 dpf zebrafish larvae, proliferation also contributes to beta cell expansion (Field et al., 2003; Hesselson et al., 2009; Moro et al., 2009). Therefore, we investigated whether CV larvae had higher levels of beta cell proliferation than GF larvae. 4 dpf larvae were treated with the thymadine analog, 5-ethynyl-2'-deoxyuridine (EdU) for 48 hours to mark cells that underwent proliferation during this time window. We found that, by 6 dpf, CV larvae had significantly more EdU labeled insulin-expressing cells than GF larvae (Figure 5A, B, Figure 5 – source data 1). Next we asked whether treatment of GF larvae with BefA was sufficient to restore beta cell proliferation to CV levels. We found that BefA-treated GF larvae had EdU incorporation similar to CV fish and significantly greater than GF (Figure 5A, B, Figure 5 – source data 1). CFS from our *A. veronii*$^{\Delta befA}$ strain was not sufficient to increase proliferation rates in GF fish (Figure 5B, Figure 5 – source data 1). Our results show that BefA is sufficient to increase cell proliferation that gives rise to an expanded beta cell population during early larval development. Furthermore, BefA seems to be the only product of the *A. veronii* CFS that is capable of inducing this cell proliferation.

In zebrafish larvae, both the proliferation of existing beta cells as well as the proliferation of progenitors contribute to the expansion of beta cells that occurs between 4 and 6 dpf (Dong et al., 2007; Field et al., 2003). Because our 48-hour EdU pulse labeled β cells born from both events, our experiment did not distinguish the exact cell population undergoing proliferation in response to BefA. Due to their low rates of proliferation, dividing β cells are difficult to detect without pulse labeling. Neogenesis of β cells from progenitors is also rare, but can be detected as the appearance of insulin positive cells in the extra-pancreatic duct (EPD) (Dong et al., 2007; Hesselson et al., 2009). We quantified insulin expressing cells in the EPD in 6 dpf CV and GF larvae. In a survey of over 500 *Tg(-1.0insulin:eGFP)* larvae, we found a slight but significant increase in EPD-localized insulin expressing cells in CV versus GF fish (Figure 5 – figure supplement 1), suggesting that the microbiota increases endocrine progenitor proliferation. Whether the microbiota also promote proliferation of mature beta cells in the islet and whether BefA promotes the proliferation of one or both of these cell populations remains to be determined.

To test whether BefA activity was specific to endocrine tissue, or whether it acts as a nonspecific pro-proliferative stimulant in the pancreas, we analyzed its ability to induce proliferation in exocrine pancreatic tissue by treating *Tg(ptf1a:eGFP)* larvae (Thisse et al., 2004) with EdU and BefA from 4 to 6 dpf and quantifying proliferative eGFP positive cells. We found no difference in the level of exocrine cell proliferation across GF, CV and BefA treatments (Figure 5C). To test whether beta cells were the only endocrine cell type in the islet to be responsive to BefA, we also quantified the total number of glucagon-expressing α (Figure 5D) and somatostatin-expressing δ (Figure 5E)

cells in GF, CV and BefA treated fish. We again found no difference in the total numbers of these cells across treatments (Figure 5F, G). These results suggest that in the pancreas, beta cells alone are responsive to the presence of BefA.

Figure 5. BefA facilitates beta cell mass expansion through proliferation.
(A, D & E) Representative 2D slices from confocal scans through the primary islets of GF, CV and BefA (10165) protein treated 6 dpf larvae. Scale bars = 40 μM. **(A)** Insulin promoter expressing beta cells are shown in green, all nuclei are blue, and EdU containing nuclei are magenta. Left hand panels are a merge of all three markers. For ease of resolving cells that are double positive for both insulin and EdU, the right hand panels show location of insulin outlined by white dashed lines. **(B)** Percentage of EdU positive beta cells in CV, GF or GF treated with either purified BefA or CFS from *A. veronii*$^{\Delta befA}$ cultures (*ΔbefA* CFS). ***p<0.001: Denotes treatment that is significantly different than GF by Tukey analysis. **(C)** Total EdU positive exocrine cells quantified from the approximate central longitudinal plane of the pancreas in each fish. **(D)** Insulin promoter expressing beta cells are shown in green, all nuclei are blue, and α cells, stained with anti-glucagon antibody are magenta. **(E)** Somatostatin promoter expressing δ cells are shown in white, all nuclei are blue, and beta cells stained with anti-insulin antibody are outlined in green. **(F)** Total α cells in GF, CV and GF fish treated with BefA. **(G)** Total δ cells in GF, CV and GF fish treated with BefA.

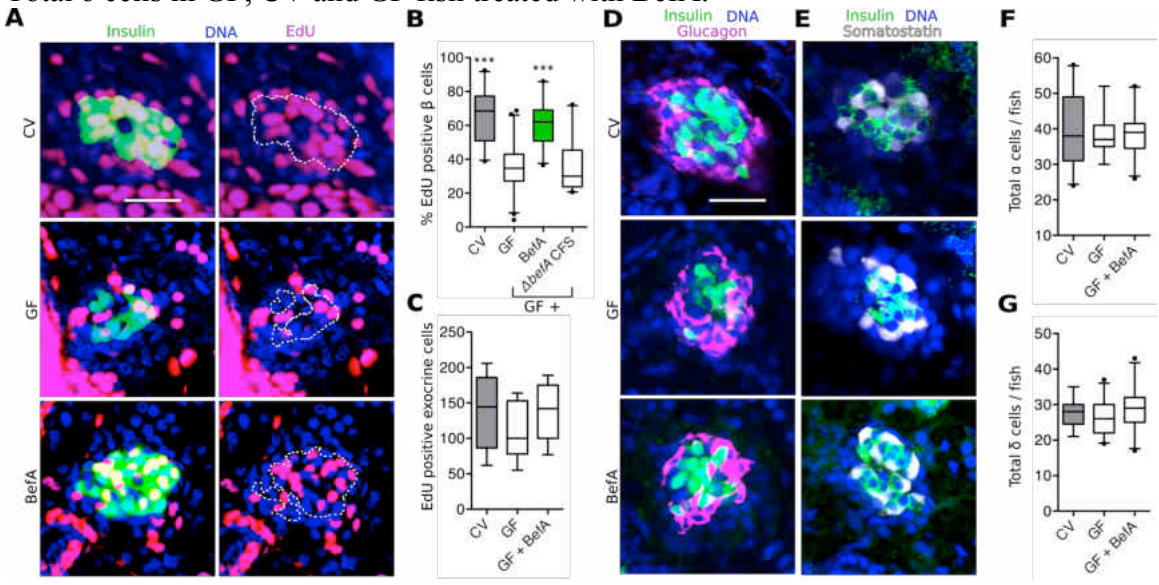
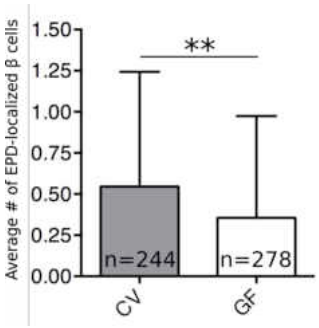


Figure 5S – figure supplement 1. The microbiota increase beta cell neogenesis from the EPD.

Quantification of EDP localized insulin expressing cells per animal in 6 dpf CV and GF larvae. Error bars represent the standard deviation. **t_{520}=3.28, p=0.0011.



BefA homologs are produced by members of the human gut microbiota and have conserved function.

We wondered if BefA-like proteins are produced by the human microbiota. Phylogenetic analysis of related sequences in bacterial genomes uncovered close homologs (at least 82% amino acid sequence identity) in many, but not all, species of the *Aeromonas*, *Vibrio*, and *Photobacterium* genera. We also found an example of a highly related sequence in the human-associated species *Enterococcus gallinarum*, which was likely acquired through a horizontal gene transfer event (Figure 6A). Widening the search to include more distant homologs identified potentially related genes in three additional human-associated genera: *Enterobacter*, *Escherichia*, and *Klebsiella* (Figure 6B).

We tested whether representative BefA-like proteins from human-associated bacteria had the capacity to induce beta cell expansion in our gnotobiotic zebrafish model. We cloned into BL21 *E. coli* two *befA*-like genes: the more closely related homologue from *Enterococcus gallinarum* and a more distantly related homologue from

Enterobacter aerogenes. The amino acid sequence alignment of these two homologs against the *Aeromonas* BefA sequence is shown in Figure 6 – figure supplement 1. Both the *Aeromonas* and *Enterococcus* sequences contain a short N-terminal hydrophobic secretion signal, which is not predicted in the more distant *Enterobacter* sequence. The most conserved region of these proteins is the C-terminal portion, which contains a putative SYLF domain of unknown function. Induction of expression of each gene in *E. coli* yielded CFS that were dominated by each of the respective homologous proteins, in contrast to the CFS from control *E. coli* expressing an empty vector (Figure 6C). Upon addition of these supernatants to GF larval zebrafish, we observed rescue of beta cell numbers to the CV level with both the *Enterococcus gallinarum* and *Enterobacter aerogenes* proteins, but not the empty vector control (Figure 6D, Figure 6 – source data 1). These results indicate that members of human-associated microbiota produce secreted proteins capable of inducing beta cell expansion.

Figure 6. Homologs of BefA encoded in the human microbiome have conserved function in zebrafish.

(A) Close homologs of BefA across microbial species. Each species is represented by its closest BefA homolog, with a minimum allowed amino acid sequence identity of 50% (relative to the query sequence). Notably, the *Enterococcus gallinarum* homolog clusters among homologs from the *Aeromonas* genus, which is evidence of a possible lateral gene transfer event. (**B**) A view of the BefA phylogeny including more distant homologs (sequence identity >20%) and grouped by genus. The portion of the tree represented in A is contained in the light gray box. In both panels, red numbers indicate branch support (values closer to 1 are better supported); branches with support values <0.5 have been collapsed. Blue clades indicate genera that were associated with humans in metagenomes produced during the Human Microbiome Project (HMP). Black arrowheads indicate genera tested for functional conservation in panel D. Scale bars indicate amino acid substitutions per amino acid site. **(C)** SDS-page gel: 1 = ladder, 2 = CFS from induction of *E. coli* BL21 carrying an empty vector, 3 = CFS from induction of *E. coli* BL21 carrying vector with *Enterococcus gallinarum* homolog, estimated size of 29 kDa, lane 4

= CFS from induction of *E. coli* BL21 carrying vector with *Enterobacter aerogenes* homolog, estimated size of 21 kDa. White arrows indicate induced proteins. **(D)** Total beta cells in CV, GF and GF fish that have been treated with either induced BL21 *E. coli* supernatant dominated by the homologous BefA protein encoded from *Enterococcus gallinarum* (*E. gal.* homolog) and *Enterobacter aerogenes* (*E. aero.* homolog), or induced supernatant from an empty vector control. *p<0.05, **p<0.01, ***p<0.001: Denotes treatment that is significantly different than GF by Tukey analysis.

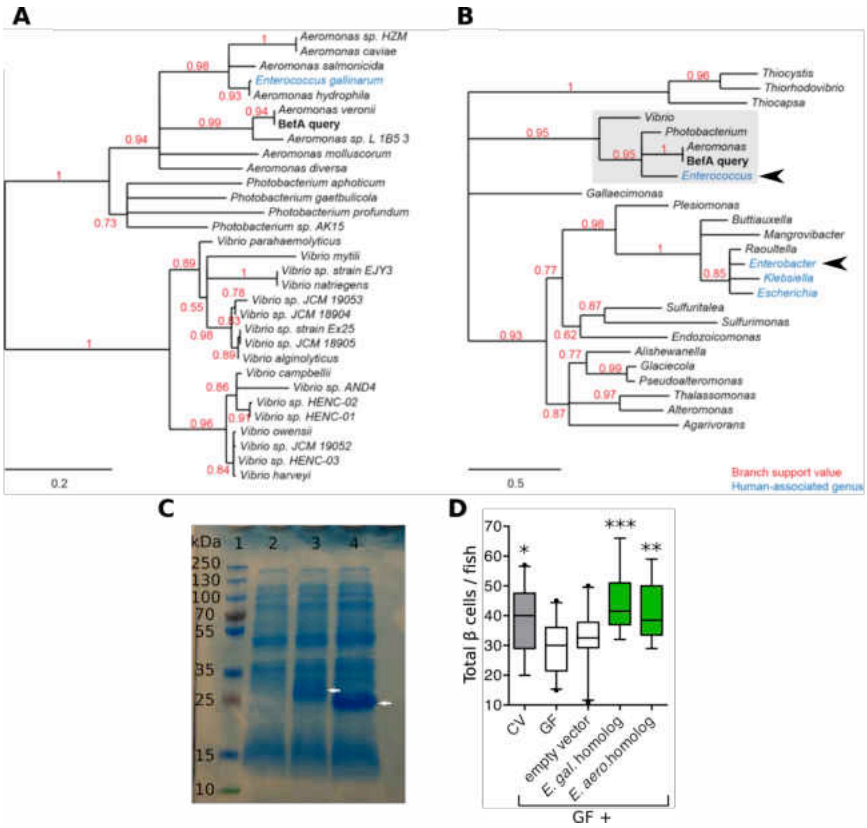
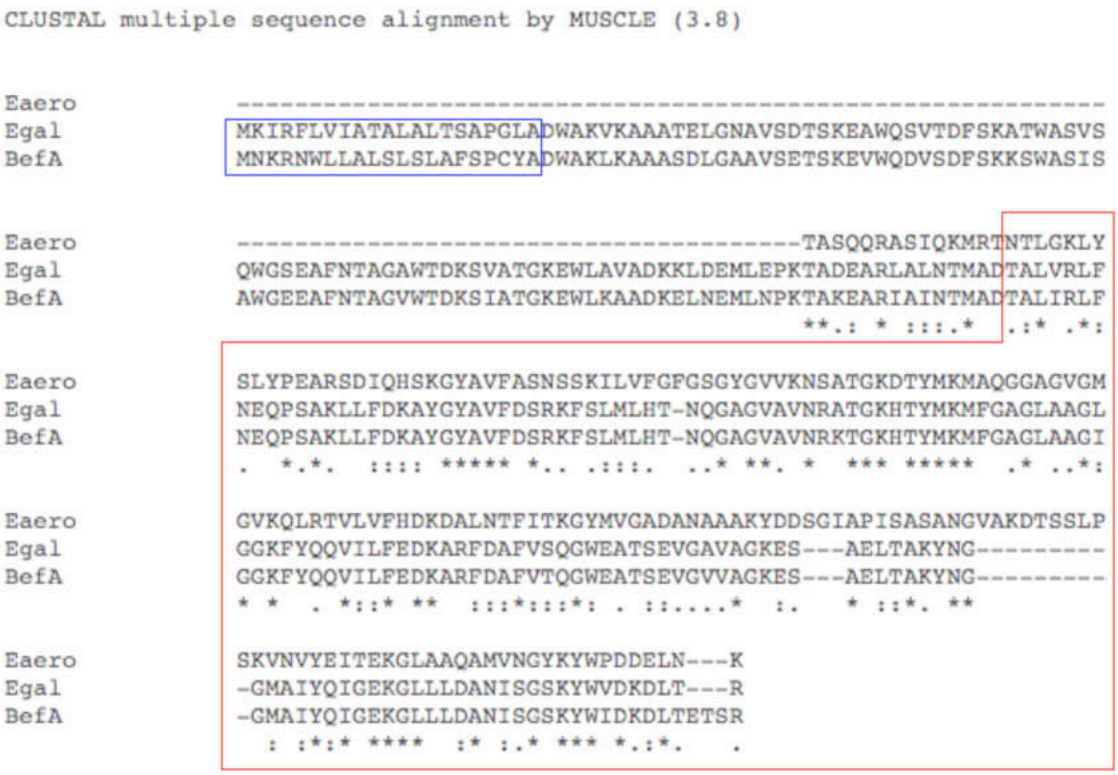


Figure 6S – figure supplement 1. Amino acid sequence alignment of BefA and functionally conserved homologs. (next page)

Amino acid sequence alignment by MUSCLE. Egal = *Enterococcus gallinarum* homolog sequence, Eaero = *Enterobacter aerogenes* homolog sequence, and BefA = original *Aeromonas veronii* HM21 BefA sequence. Red box contains predicted SYLF domain. Blue box indicates predicted secretion peptides.



DISCUSSION

Using a gnotobiotic zebrafish model, we have discovered a class of proteins produced by resident gut bacteria that have the capacity to increase expansion of pancreatic beta cells during early zebrafish development. BefA and related homologues are predicted to contain a C-terminal SYLF domain, which has been described in proteins from organisms in all kingdoms of life, including humans, but for which little is known functionally beyond a possible role in lipid binding (Hasegawa et al., 2011). Genes encoding BefA and related proteins are found in a small subset of all bacteria genera, with a predominance in genera of host-associated bacteria, but *befA* homologues are not ubiquitously present in any of these genera.

Our finding of a role for specific secreted bacterial proteins in beta cell development raises the possibility of a new link between the resident microbiota and diseases of beta cell paucity, such as diabetes mellitus. Type 1 diabetes (T1D), is caused by both genetic and environmental factors, as indicated by the 50% disease discordance among monozygotic twins (Akerblom et al., 2002). One environmental factor associated with T1D is microbiota composition (Gülden et al., 2015). Mechanistic models for the role of the microbiota in T1D etiology have focused on the capacity of the microbiota to modulate the development and function of the immune system, and thus influence the propensity of genetically susceptible individuals to develop autoimmunity to beta cell antigens (Gülden et al., 2015). Multiple aspects of host immune cell development and function known to play a role in T1D are altered by the loss of microbes, including development of lymphoid tissue (Macpherson and Harris, 2004) and T cell differentiation and function (Alam et al., 2011; Farkas et al., 2015; Ivanov et al., 2008). We hypothesize an additional role for the early microbiota in establishing the beta cell population size that would either buffer against, or render individuals susceptible to, beta cell depletion by autoimmune destruction.

In humans, beta cells undergo a period of postnatal expansion, before becoming quiescent around age two (Gregg et al., 2012). Differences in beta cell growth during this time are thought to account for the wide variation in beta cell mass observed in adults (Wang et al., 2015). The idea that early life beta cell census could influence diabetes risk is supported by studies in both rodents and humans, and has been theorized as an important risk factor for type 2 diabetes (Kaijser et al., 2009), a disease which is also influenced strongly by microbiota composition (Cox and Blaser, 2014). Compromised

beta cell development in rats results in an insufficient number of cells to adequately control glucose metabolism (Figliuzzi et al., 2010). In mice, perinatal beta cell proliferation rates can be tuned via the modulation of Gi-GPCR signaling (Berger et al., 2015). Changes to early beta cell proliferation capacity in these mice correlates directly with adult beta cell mass, which subsequently impacts glucose regulation (Berger et al., 2015). Furthermore, meta analysis of human data has revealed a correlation between early age of beta cell loss and more rapid onset of T1D (Klinke, 2008), consistent with the model that failure to generate a reserve of beta cells early in development increases disease risk.

We hypothesize that neonatal microbiomes with a low abundance of BefA equivalents would result in reduced beta cell proliferation, lower beta cell mass, and increased diabetes risk. We do not know how many different microbiota-derived molecules can stimulate beta cell proliferation, but for the case of *befA* homologues, we know these to be sparsely distributed in bacterial genomes, such that microbiomes of low taxonomic diversity could lack these genes. The idea that microbiota-derived factors capable of protecting against diabetes are not widely conserved is consistent with human microbiota profiling data (Morgan et al., 2013), our own functional assays of bacteria in gnotobiotic zebrafish, and other gnotobiotic rodent experiments. For example, specific bacterial lineages have been shown to attenuate disease in diabetes models, including *Segmented Filamentous Bacteria* (SFB) in the non-obese diabetic (NOD) mouse (Kriegel et al., 2011; Yurkovetskiy et al., 2013) and *Lactobacillus johnsonii* in the Biobreeding rat model (Valladares et al., 2010). Furthermore, Wen and colleagues have shown that certain microbial assemblages, but not others, confer disease protection in neonatal NOD

mice (Peng et al., 2014). Additional recent work by Wen and colleagues demonstrates early development as a critical window for microbiota modulation of disease risk in NOD mice (Hu et al., 2015). We have shown that BefA acts during early developmental stages in zebrafish, and we hypothesize that beta cell expansion during this developmental window is important for disease prevention, and may be a critical period for clinical intervention for infants at risk for T1D development. Further work will be required to determine whether BefA is capable of inducing proliferation of adult beta cells in zebrafish or other animals.

Why certain bacteria produce BefA is unclear. In the context of the zebrafish intestinal environment, BefA confers a slight colonization advantage to *A. veronii*, however this is unlikely to be related to its capacity to induce beta cell mass, because the colonization requirement is only apparent in the context of co-colonization with wild type *A. veronii* that induce normal beta cell numbers. It is possible that bacterial modulation of host beta cell number serves a purpose for the bacteria not measured in our assay. Alternatively, bacteria may produce BefA for a purpose independent of beta cell expansion and the host simply uses this bacterial molecule as a cue for its own developmental program. Learning the molecular basis for BefA sensing by the host, and whether it interacts directly or indirectly with beta cells, will help shed light on the nature and evolutionary conservation of this interspecies signaling. It will also be important to understand the bacterial function of BefA in order to be able to manipulate its abundance for potential therapeutic purposes.

The incidence of autoimmune diseases such as T1D has been increasing markedly in developed nations over the past several decades. One theory to explain this

phenomenon is the disappearing microbiota hypothesis, which proposes that over time, as our modern lifestyles have become increasingly sterile, we have lost ancestral microbial symbionts important for specific aspects of our health (Blaser and Falkow, 2009). Our discovery of a specific class of bacterial proteins that promote beta cell expansion in early development is consistent with the hypothesis that loss of specific microbial taxa from gut microbiota could underlie increased diabetes risk. Specifically, we suggest that BefA-like proteins promote the establishment of a robust beta cell population that is more resilient to subsequent beta cell loss. Because *befA* is a relatively rare component of the microbiome, we cannot measure it directly from available metagenomic sequence data to test our hypothesis that *befA* abundance correlates with reduced diabetes risk. The low abundance of *befA* in metagenomes also highlights the challenge of discovering disease determinants from metagenomic data, and emphasizes the importance of functional screening approaches. The larval zebrafish has served as a valuable high-throughput vertebrate model for the identification of new compounds and pathways that can increase beta cell numbers exogenously (Andersson et al., 2012; Wang et al., 2015). We have employed the gnotobiotic zebrafish to explore how microbial cues modulate beta cell development. Our discovery of BefA highlights the importance of the microbiota in shaping the development of an extra-intestinal tissue and influencing the overall metabolic state of the host. We postulate that resident bacteria are a rich and underexplored source of functionally conserved molecules that shape early host development in ways that impact disease risk in later life.

MATERIALS AND METHODS

Gnotobiotic zebrafish

All zebrafish experiments were performed using protocols approved by the University of Oregon Institutional Care and Use Committee and followed standard protocols. Zebrafish embryos were derived germ-free (GF) as previously described (Bates et al., 2006). XGF and mono-associated larvae were also generated as previously described (Bates et al., 2006), except that all bacterial inoculate were added to GF flasks at 4 dpf at a final concentration of 10^6 CFUs/mL. In experiments quantifying the colonization levels of bacterial isolates, each strain was added to the embryo media (EM) and incubated with the larvae for 48 hours at 27° C. Larvae were sacrificed at 6 dpf, immediately before the gut was removed and homogenized in a small sample of sterile EM. Dilutions of this gut slurry were plated onto tryptic soy agar and allowed to incubate overnight at 30° C. Colonies from each gut were quantified. A minimum of 10 guts per mono-association or di-association were analyzed.

Free Glucose Assay

To measure beta cell function in GF and CV zebrafish larvae, levels of free glucose were measured at 6 dpf using a free glucose assay kit (BioVision, Milpitas, CA) as described previously (Andersson et al., 2012; Gut et al., 2013) except that only 10 larvae were combined per tube. Three to five biological replicates (sets of 10 larvae) were completed for both GF and CV treatments each time the assay was conducted. Data shown here were combined from 3 separate experimental assays or technical replicates.

Cell Free Supernatant

GF fish were inoculated with secreted bacterial products at 4 dpf by adding cell free supernatant (CFS) at a final concentration of 500 ng/mL to the water of the sterile flasks. CFS was harvested from a 50 mL overnight culture of the specified bacterial strain. The cultures were centrifuged at 7000 g for 10 minutes at 4° C. The supernatant was then filtered through a 0.22-μm sterile tube top filter (Corning Inc., Corning, NY); sterile supernatant was concentrated at 4° C for 1 hour at 3000 g with a centrifugal device that has a 10 kda weight cut off (Pall Life Sciences, Port Washington, NY).

For experiments utilizing proteinase K (Qiagen, Hilden, Germany), the enzyme was added to samples of CFS at a final concentration of 100 μg/mL and allowed to incubate at 55° C for 1 hour before inactivating the enzyme at 90° C for 10 minutes.

Ammonium Sulfate Fractionation

Ammonium sulfate fractionation was performed on un-concentrated, sterile CFS from a 50 mL overnight culture by slowly adding 100% ammonium sulfate until solutions of 20%, 40%, 60% and 80% ammonium sulfate were achieved. These solutions were prepared at 4° C. Precipitated proteins were collected from each fraction by centrifugation at 4° C and 14000 g for 15 minutes. The proteins were resuspended in cold EM and dialyzed for 2-3 hours at 4° C before adding them to 4 dpf GF larvae at a final concentration of 500 ng/mL.

Mass Spectrometry

The 60-80% ammonium sulfate fraction of the *A. veronii*$^{\Delta T2SS}$ CFS was sent to the

Proteomics Lab at Oregon Health and Science University in Portland, OR for protein identification (partial sequencing) analysis.

Protein expression and purification

The nucleotide sequence for the *befA* gene from was amplified from *A. veronii* using the following forward and reverse PCR primers respectively: 5'-GCCCATATGatgaacaagcgtaactggttgctg-3' and 5'-GGCCTCGAGgcggctcgtttcagtcaagtc-3'. The nucleotide sequences for both the *Enterococcus gallinarum* and *Enterobacter aerogenes befA* gene homologs were obtained from NCBI and subsequently synthesized by GenScript, Piscataway, NJ. Each of these two genes was then cloned separately into the pET-21b plasmid (Novagen, Darmstadt, Germany), which contains an IPTG inducible promoter. A His□Tag® was added to the C-terminal of the original BefA protein sequence for subsequent purification. As a control, a second version was also constructed lacking the tag. These vectors were then transformed into BL21 *Escheria coli* (RRID:WB_HT115(DE3)), treated with 0.5 – 1.0 mM IPTG during exponential growth phase (OD_{600} = 0.4 - 0.6) and allowed to grow for 3-4 more hours at 30° C. This resulted in both a CFS and cell lysate dominated by our proteins of interest, as confirmed via SDS-page gel electrophoresis by the presence of dark bands of the expected sizes for each protein. These bands were absent from BL21 cultures carrying an empty pET-21b vector. The CFS from these inductions was added to GF zebrafish at 4 dpf at a final concentration of 500 ng/mL.

For purification of BefA, IPTG induced BL21 cells were sonicated at 32,000 g in a 50 nM Tris, 150 mM NaCl buffer (buffer A). The supernatant was then added to a

solution of nickel beads (Thermo Scientific HisPur^{TM} Ni-NTA Resin, Waltham, MA) to capture the His□tag. The beads were washed several times in a 30 mM imidazol solution in buffer A and subsequently eluted in 300 mM imidazole solution in buffer A. The isolation of pure BefA was confirmed with SDS-page gel electrophoresis by the presence of a single band of about 29 kDa in size. Purified BefA was added to 4 dpf GF fish at a final concentration of 500 ng/mL.

Experimental bacterial strains

To create the *A. veronii*$^{\Delta befA}$ mutant strain, a vector containing a chloramphenicol resistance cassette was transformed into SM10 *E. coli*. Conjugation between wild-type *Aeromonas veronii* HM21 and the vector carrying SM10 *E. coli* strain was carried out, allowing the chloramphenicol resistance gene to replace the *befA* locus in *A. veronii* via allelic exchange. Candidate mutants were selected for loss of the plasmid and maintenance of chloramphenicol resistance. Insertion of the chloramphenicol cassette into the *befA* locus was verified in these candidates by PCR.

Joerg Graf graciously provided us with the *A. veronii*$^{\Delta T2SS}$ strain (Maltz and Graf, 2011).

Primary islet cell type quantifications and EdU staining in larvae

*Tg(-1.0insulin:eGFP) (*RRID:ZFIN_ZDB-GENO-100513-10, ZIRC, Eugene, OR) (*diIorio et al., 2002)* zebrafish embryos were used to visualize and quantify the total number of beta cells in developing larvae. *Tg(insulin:PhiYFP-2a-nsfB, sst2:mCherry)* (RRID:ZFIN_ZDB-GENO-120217-6) (Wang et al., 2015) were obtained from Jeff

Mumm and were used to visualize and quantify δ cells. All experiments were analyzed at 6 dpf unless otherwise specified. At all time points in all experiments, larvae were fixed with 4% paraformaldehyde supplemented with 0.01% Triton® X-100 (Thermo Fisher Scientific, Waltham, MA) at 4° C overnight, or at room temperature for 2-3 hours, and then washed with PBS. The following antibodies were used to distinguish α and beta cells: guinea-pig anti-insulin (Dako Cat# A0564, RRID:AB_10013624 , Carpinteria, CA), mouse anti-glucagon (Sigma-Aldrich Cat# G2654, RRID:AB_259852), St. Louis, MO), rabbit anti-GFP (Molecular Probes Cat# A-11122, RRID:AB_221569), mouse anti-mCherry (Abcam Cat# ab125096, RRID:AB_11133266, Cambridge, MA), Alexa Fluor® 488 goat anti-rabbit (Thermo Fisher Scientific, Waltham, MA), anti-mouse Cy3 (Jackson ImmunoResearch Laboratories Inc., West Grove ,PA), Alexa Fluor® 488 goat anti-guinea-pig (Thermo Fisher Scientific, Waltham, MA), and TO-PRO®-3-Iodide (642/661) (Thermo Fisher Scientific, Waltham, MA).

For experiments quantifying proliferation, EdU was added at 4 dpf directly to the EM at a final concentration of 0.1 mg/mL. The Click-iT® EdU Imaging Kit (Thermo Fisher Scientific, Waltham, MA) was used to process the EdU label in whole fixed zebrafish prior to antibody staining, according to the manufacturer's protocols. Whole, antibody-stained larvae were mounted for confocal microscopy (BioRad Radiance 2100) with their right side facing up against the cover slip, which was flattened sufficiently to spread our the cells within the islet for optimal quantification of individual cells. For quantification of beta cells and other primary islet cells, the entire endocrine portion of the pancreas was scanned using a 60X objective (Nikon Eclipse E600FN), and Fiji (RRID:SCR_002285) (Schindelin et al., 2012) software was used to analyze each image

stack. For quantification of pancreatic exocrine tissue proliferation, *Tg(ptf1a:eGFP)* (RRID:ZFIN_ZDB-GENO-080111-1, ZIRC, Eugene, OR) (Thisse et al., 2004) zebrafish were scanned through the entire pancreas with a 20X objective (Nikon Eclipse E600FN) and Fiji was used to analyze the percentage of proliferative cells in single sections from the center of the organ. Images were prepared for publication using the open source Inkscape software (RRID:SCR_014479).

For experiments quantifying insulin-expressing cells in the region of the EPD, zebrafish were processed as described above, and analyzed on a Leica fluorescent microscope using a 2x objective.

BefA phylogenetic analysis

We screened for BefA homologs across microbial species using a *blastp*-based (Altschul et al., 1997) search of the UniProt Knowledgebase (The UniProt Consortium, 2015) (version 6/2015); default search parameters were changed to allow (i) a maximum *E*-value of 1.0 and (ii) an arbitrarily large number of database hits. We classified database hits as "close homologs" if amino acid sequence identity exceeded 50% (relative to the query length) and "distant homologs" if their percent identity exceeded 20%. For phylogenetic analysis at the species level, each species was represented by the hit of highest percent identity to BefA among isolates of that species (if any); an analogous procedure was used for genus-level analysis. Aligned portions of database sequences were isolated and multiply aligned with *MUSCLE* (RRID:SCR_011812) (Edgar, 2004). Phylogenetic trees were constructed from these multiple sequence alignments using *PhyML* (RRID:SCR_014629) (Guindon and Gascuel, 2003) and visualized within the

Phylogeny.fr webserver (Dereeper et al., 2008). Microbial genera were classified as "human-associated" if they occurred with relative abundance >0.01% in at least 5 metagenomes from the Human Microbiome Project (Huttenhower et al., 2012) as profiled by MetaPhlAn (RRID:SCR_004915) (Segata et al., 2012). Secretion signal peptides were predicted from amino acid sequences using SignalP (Petersen et al., 2011).

Statistical Analysis

Appropriate sample sizes for all experiments were estimated *a priori* using a power of 80% and a significance level of 0.05. From preliminary experiments we estimated variance and effect. For larval beta cell quantification, these parameters suggested using a sample size of 30 in order to detect significant changes between treatment groups. Therefore, each experiment contained about 10-15 biological replicates or individual fish per treatment group, although some larger experiments had fewer biological replicates due to limited material. Entire experiments or technical replicates were repeated multiple times, resulting in pooled data sets of about 20-50 biological replicates. These data are represented in the figures as box and whisker plots, which display the data median (line within the box), first and third quartiles (top and bottom of the box), and 95% confidence interval (whiskers). Any data point falling outside the 95% confidence interval is represented as a solid dot. These pooled data were analyzed through the statistical software RStudio®. For experiments comparing just two differentially treated populations, a Student's t-test with equal variance assumptions was used. For experiments measuring a single variable with multiple treatment groups, a single factor ANOVA with post hoc means testing (Tukey) was utilized. A p-value of

less than 0.05 was required to reject the null hypothesis that no difference existed between groups of data.

BRIDGE

In Chapter II, we showed that the bacterial protein BefA, is sufficient to induce zebrafish pancreatic beta cell expansion, and that members of both the zebrafish and human intestinal microbiota produce and secrete BefA into the host environment. However, we do not yet know how the host senses and responds to this protein. To further our understanding, in the next chapter we present the atomic structure of BefA, which represents a new protein fold that includes the first structural description of an SYLF domain, which is the most conserved region of the protein amongst homologs. We show that the SYLF domain is sufficient to induce larval beta cell expansion in GF zebrafish larvae. Using zebrafish islet explants, we present evidence to support the hypothesis that this effect is mediated through a direct interaction between the SYLF domain and pancreatic cells, suggesting that BefA acts directly on the pancreas, possibly transported to this tissue from the gut lumen.

CHAPTER III

THE ATOMIC STRUCTURE OF BEFA REVEALS A NEW PROTEIN FOLD SUFFICIENT FOR BETA CELL EXPANSION

The procedures involved in solving the BefA crystal structure described in this chapter were completed thanks to large contributions from Dr. Emily Sweeney, who helped to derive selenium methionine substituted BefA, to solve the x-ray diffraction pattern of the resulting crystals, and to design truncation constructs, Dr. S. James Remington who solved the crystal structure of BefA, and Karen Kallio, who guided me in the protocol to find appropriate crystallization conditions for BefA.

INTRODUCTION

Beta cell expansion factor A (BefA) is a small 29 kDa protein produced by some members of the intestinal microbiota of various hosts including zebrafish and humans (Hill et al., 2016). Recently, we showed that the microbiota is required for zebrafish beta cell expansion during the larval period after the animals hatch from their chorions and first become colonized by environmental microbes. We discovered BefA in a screen for bacterial products that could rescue germ free (GF) larval beta cell development. Because mechanisms of beta cell renewal are sought after as potential treatments for diabetes, we became interested in whether or not human associated bacteria produce BefA or similar proteins. Upon searching metagenomic databases from the Human Microbiome Project, we found several species that had genes encoding predicted homologs of BefA. Notably, these genes were not common across any bacterial genus, but seemed to be present in a

small handful of strains or species. We examined two of these homologs, one with only 34% amino acid sequence similarity to BefA, and found that they were both sufficient to rescue GF larval beta cell development.

To begin to understand how these proteins could induce this effect, we used amino acid alignments and found that the C-terminus was the most conserved region across BefA homologs. This conserved region is relatively large compared to the size of BefA, encompassing roughly two thirds of the entire amino acid sequence, and is predicted to encode a SYLF/YAB domain, also called DUF (domain of unknown function) 500. Interestingly, some of the homologs of BefA are smaller proteins and this predicted domain constitutes their entire amino acid sequence. Little is known about these protein domains, despite their prevalence across the kingdoms of life, from bacteria to fungi and animals (Hasegawa et al., 2011). However, their role in two distinct proteins, one found in yeast and the other in mammals, has been investigated, providing hints about their biological functions (Hasegawa et al., 2011; Robertson et al., 2009; Tonikian et al., 2009).

The YAB domain was described first in 2009 in the yeast protein Ysc84, and is named after its observed function for **Y**sc84 **a**ctin **b**inding (Robertson et al., 2009). The YAB domain is located at the N-terminus of Ysc84 and is attached by a linker to a SH3 domain in the C-terminus (Robertson et al., 2009). Robertson et al. showed that the YAB domain of Ysc84 is sufficient to bind to F-actin, a function necessary for its role in endocytosis. In 2011, Hasegawa and colleagues described a homolog of the YAB domain in the mammalian SH3YL1 protein, and they re-named the region SYLF, after each of the proteins known to contain it at that time (**S**H3YL1, **Y**sc84p/Lsb4p, **L**sb3p, and

FYVE). SH3YL1 was first discovered in mice, and is widely expressed in many organ systems (Urbanek et al., 2015). Like Ysc84, the SYLF domain of SH3YL1 is also located in the N-terminus with an attached C- terminus SH3 domain (Hasegawa et al., 2011). However, the SYLF domain of SH3YL1 can't bind to F-actin (Hasegawa et al., 2011). Instead it interacts with D5-phosphoinositides and has a role in regulating membrane ruffle formation (Hasegawa et al., 2011). Plant proteins containing a C-terminus SYLF domain attached to an N-terminus FYVE domain can be found in the genomes of *Arabidopsis thaliana* and *Oryza sativa*, however their specific SYLF domain functions have not been studied (Urbanek et al., 2015). Similar to the mammalian SH3YL1 SYLF domain, FYVE domains in *Arabidopsis sp.* are known to interact with phosphatidylinositol-3-phosphate at the plasma membrane, and generally play a role in endocytosis and/or vesicular trafficking (Wywial and Singh, 2010). Although research on SYLF/YAB domains is limited, it seems that an emerging theme surrounding the role of this newly discovered domain is the organization of protein interactions to carry out cellular functions at the plasma membrane. So far our description of BefA is the only insight into the function of a bacterial SYLF/YAB domain containing protein (Hill et al., 2016). Here we attempt to better understand the role of the SYLF domain in the function of BefA in host physiology.

Toward this goal, we have solved the crystal structure of BefA. This new protein structure not only exhibits a novel-folding pattern, it also provides the first known structure for any SYLF domain, which will be helpful in furthering our understanding of this widely distributed protein domain's function. We also present evidence that the SYLF domain of BefA is required for its ability to induce host beta cell proliferation, and

that this effect is mediated through direct interactions between the SYLF domain and pancreatic cells. This observation raises the question of whether or not the SYLF domain of BefA, like other described SYLF/YAB domains, is important for organizing interactions at the cell membrane. Importantly, our data provide a new understanding of how microbial products can impact host cell biology.

RESULTS

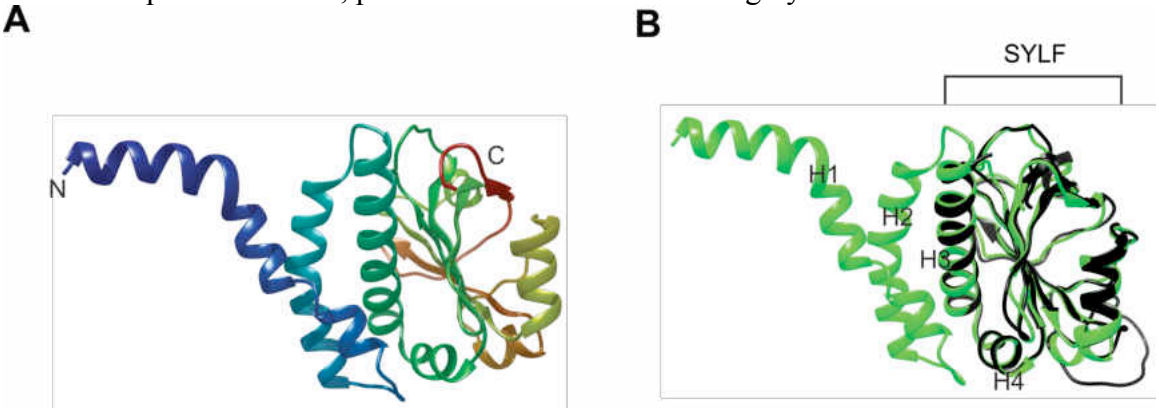
The structure of BefA reveals a novel protein fold containing a SYLF domain

To better understand how BefA functions, we set out to determine its 3D atomic structure. Purified BefA protein crystallized readily and the structure was determined to 1.6 Å resolution using a selenomethionine derivative and single-wavelength anomalous dispersion (SAD) phasing. The atomic resolution structure revealed a compact partial beta-barrel carboxy-terminal domain with three flanking α-helices and four loosely packed α-helices in the amino terminal region (Figure 3.1A&B). From sequence comparison and analysis using the Conserved Domain Databank (Marchler-Bauer et al., 2017), we conclude that BefA contains a SYLF domain. Surprisingly, the BefA structure does not have homology to any structures deposited in the Protein Data Bank (PDB), using PDBeFold (Krissinel and Henrick, 2004, 2005), indicating that we have determined a novel protein fold.

To better define the structural features that constitute a functional SYLF domain, we performed a detailed structural comparison between the *Aeromonas veronii* BefA and a distant BefA homolog identified in *Enterobacter aerogenes.* Although the two sequences have only 34% identity, the *E. aerogenes* homolog is fully functional in

increasing beta cell numbers in larval zebrafish (Hill et al., 2016). We obtained a model structure of *E. aerogenes* using the Iterative Threading ASSEmbly Refinement (I-TASSER) server and BefA as a template (Roy et al., 2010; Yang et al., 2014; Zhang, 2008). Overlaying the structure of BefA and the model of the BefA homolog (Figure 3.1B, black) revealed that the two amino terminal α-helices (H1, H2) are dispensable for function (Figure 3.1B). We therefore have putatively defined the functional SYLF domain to be a curved beta-sheet containing seven antiparallel beta-strands interspersed with five α-helices (Figure 3.1B).

Figure 3.1. The structure of BefA reveals a novel protein fold containing a SYLF domain
(A) Cartoon ribbon structure of BefA, N and C designate the amine and carboxyl terminus respectively. **(B)** BefA ribbon structure in green, overlaid with predicted structure for the BefA homolog produced by *E. aerogenes* in black, H1-H4 label N-terminus alpha helices 1-4, predicted SYLF domain is roughly between black brackets.



One of the best-characterized behaviors of SYLF domains is lipid binding (Urbanek et al., 2015). Preliminary work to determine if BefA binds lipids has been inconclusive. Lipid binding strip assays revealed specific binding to phosphatidylserine (PS) and cardiolipin. However, subsequent lipid binding assays using more

physiologically relevant lipid bilayers, including tryptophan fluorescence, vesicle pelleting and molecular dynamic simulations, have shown little to no binding to phosphatidylserine (PS), despite success using annexin V, which binds stongly to PS, as a positive control (data not shown). More thorough lipid binding characterization is required to determine whether BefA binds lipids in vivo and whether BefA's pro-proliferative molecular mechanism involves lipid membrane or vesicle binding.

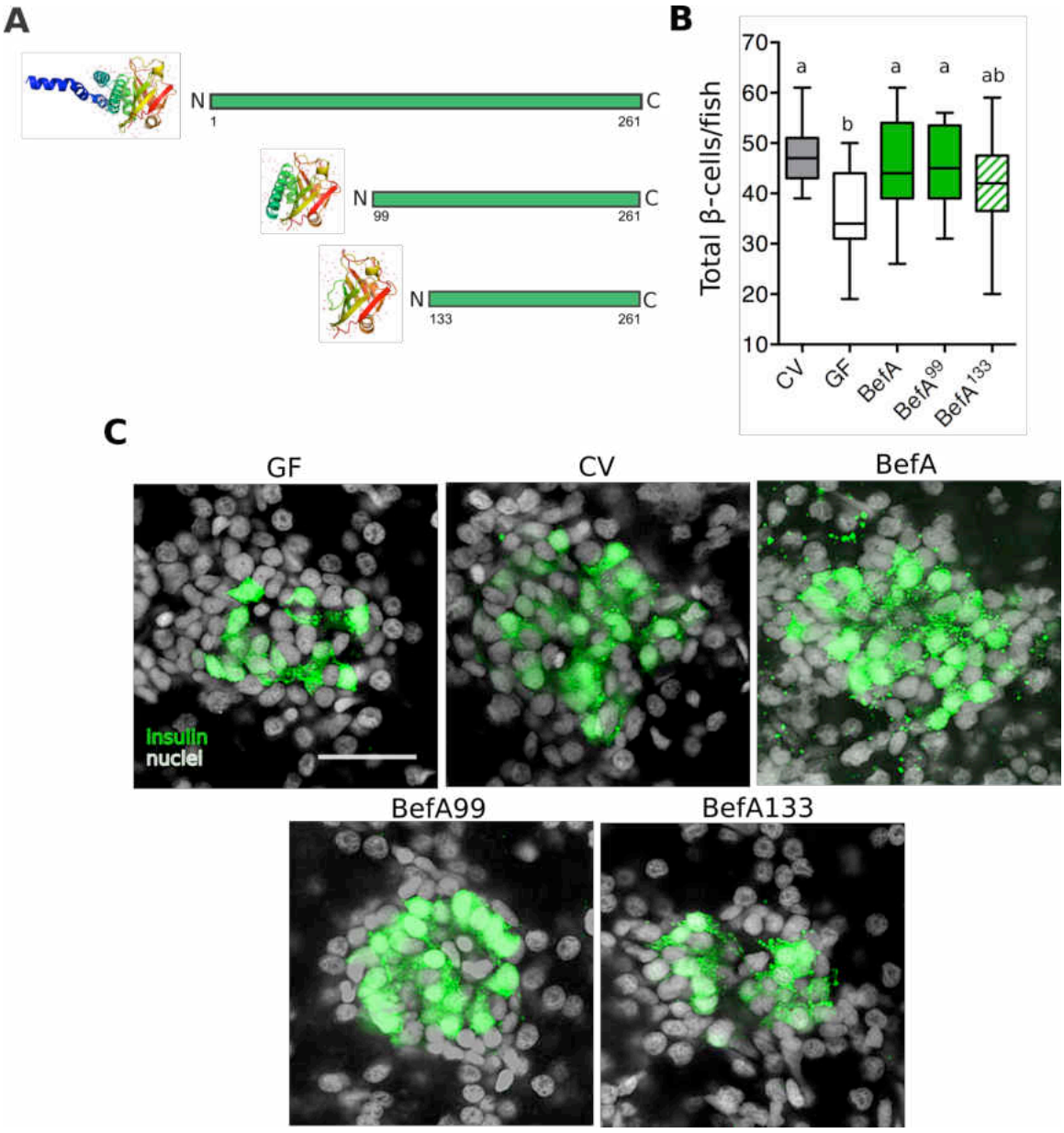
Because we have evidence that BefA may interact with PS, which is most known for its role in apoptosis, we performed Tunnel staining for apoptotic cells on 6 dpf larvae that were either GF, CV or GF treated with BefA. We did not detect any Tunnel positive beta cells within any of the samples (data not shown), indicating that apoptosis does not normally occur in developing larval beta cells and suggesting that BefA does not modulate apoptotic signaling to regulate the beta cell population.

The SYLF domain of BefA is sufficient to rescue normal expansion of the GF beta cell mass

We cloned two truncated BefA proteins to determine which region of BefA was necessary and sufficient to mediate the pro-proliferative effect on larval beta cells. One truncation comprised amino acid numbers 99 – 261, which encode the entire predicted C terminal SYLF region ($BefA^{99}$) (Figure 3.1B, Figure 3.2A). The second shorter truncation incorporated only amino acids 133-261, which corresponds to the C terminus compact partial beta-barrel with the small flanking α-helices ($BefA^{133}$) (Figure 3.2A). We tested each purified truncated BefA protein, in comparison to the full length BefA, on 4 dpf GF larvae by adding them to the water at a final concentration of 500 ng/mL. By 6

dpf, following 48 hours of treatment, BefA^{99} was equally effective at rescuing GF beta cell expansion as full length BefA, suggesting that the SYLF domain is sufficient to induce larval beta cell proliferation (Figure 3.2B). BefA^{133} was only able to partially rescue GF beta cell numbers (Figure 3.2B). Confocal images of the center of the primary zebrafish islet show many more GFP-expressing cells in fish treated with either full length BefA or BefA^{99} than in GF counterparts (Figure 3.2C). The BefA^{133} treated islets often appear similar to GF (Figure 3.2C), however they sometimes had more robust beta cell numbers approaching levels within CV fish (data not shown). These data suggest that the two most N terminal alpha helices (H1 & H2) are dispensable for the full function of BefA and that both H3 and H4, or important residues therein, are required for complete function.

Figure 3.2. The SYLF domain of BefA is sufficient to rescue normal expansion of the GF beta cell mass (next page)
(A) Left are cartoon representations of the structures predicted to result from the corresponding truncation scheme illustrated on the right, each amino acid sequence is represented by a green rectangle with numbers beneath corresponding to the amino acid number on BefA where the truncation begins. **(B)** Quantification of total beta cells per fish, in these and all subsequent box plots the whiskers denote the 95% confidence interval of the data and lowercase letters denote results of post host means testing where $p<0.05$. **(C)** Representative 2D slices from confocal scans through the primary islets of 6 dpf CV and GF *Tg(-1.0insulin:eGFP)* larvae. Each slice is taken from the approximate center of the islet structure. Insulin promoter expressing beta cells are in green and nuclei are white. Scale bar = 40μM.

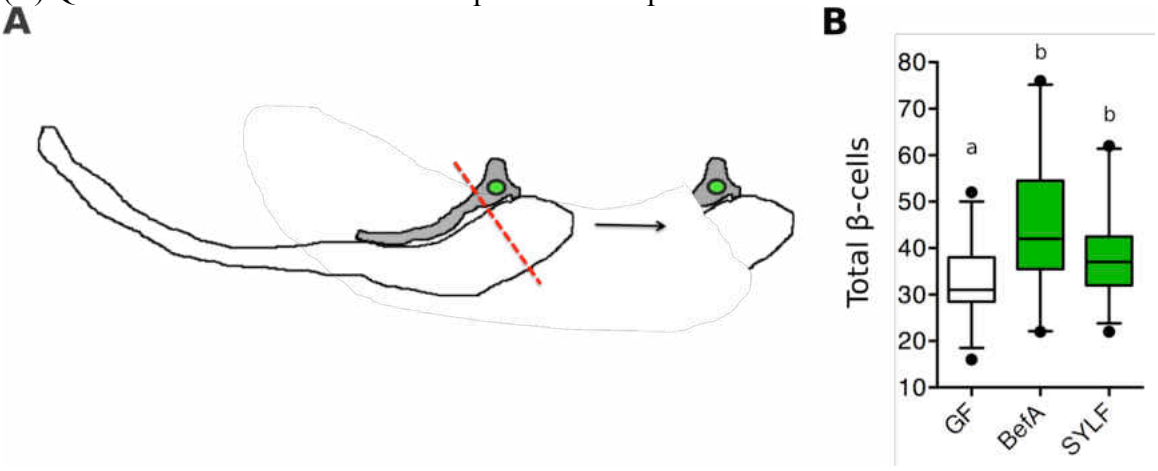


BefA increases beta cell number of GF ex vivo larval zebrafish islets

A critical question about BefA's mechanism of action is whether it acts directly on pancreatic cells or alternatively via an indirect route such as eliciting expression of a secondary messenger in intestinal cells. To test whether BefA could act directly on

pancreatic cells, we dissected pancreatic heads containing the primary islet from GF 4 dpf Tg(*-1.0insulin:eGFP*) larvae (Figure 3.3A) and maintained them *ex vivo* in culture for 48 hours untreated or in the presence of either BefA or BefA^{99}. *Ex vivo* larval beta cells appeared healthy after 48 hours as indicated by robust insulin promoter driven *gfp* transgene expression, and an otherwise normal appearance. After the treatment period, we immediately imaged and quantified total GFP expressing cells in each islet in each treatment and saw significantly more beta cells in explants treated with BefA or BefA^{99} than in those that received only rich cell culture media (Figure 3.3B), suggesting that BefA elicits beta cells expansion through a direct interaction with pancreatic cells and that the SYLF domain of BefA is sufficient for this activity. From these data, we can not distinguish between a direct interaction between BefA and the beta cells in the explants or an interaction between BefA an some other cell type in the explants, such as other endocrine cells, cells in the pancreatic ductal system, or exocrine cells, that produces a signal to elicit proliferation of beta cells.

Figure 3.3. BefA increases beta cell number of GF *ex vivo* larval zebrafish islets
(A) Cartoon depicting dissection of pancreas head from a 4 dpf *Tg(-1.0insulin:eGFP)* larvae, gut in white, pancreas in grey, primary islet in green, dashed red line = cut site.
(B) Quantification of total beta cells per dissected pancreatic head.



DISCUSSION

Solving the crystal structure of BefA revealed a novel protein fold, which includes the first description of an SYLF/YAB domain. The fold is comprised of a series of four alpha helices followed by seven beta strands, which form a curved beta sheet in the C terminus of the BefA protein. Based on our mutational analysis of the functional domain of the *A. veronii* BefA and comparisons with the smaller functional *E. aerogenes* homologue, we propose that a functional SYLF domain constitutes the third and fourth N-terminal alpha helices and the curved beta sheet. This atomic structure of the SYLF domain is consistent with previous predictions about the SH3YL1 SYLF domain structure. Hasegawa and colleagues predicted a N-terminal alpha helix in the SYLF domain, and showed that loss of this region corresponded to loss of lipid binding, suggesting the amphipathic region within this helix is necessary to mediate lipid binding in SH3YL1 (Hasegawa et al., 2011). Similarly, we found partial loss of function with the $BefA^{133}$ truncation mutant, which had no N-terminal alpha helices, suggesting that in our model, one or both of these helices may also be important for function. However, more work will be required to determine whether lipid binding or actin binding by this region is involved. Importantly, in both yeast Ysc84 and mammalian SH3YL1, their functions at the membrane are dependent upon the C-terminal SH3 domain (Urbanek et al., 2015). Likewise, in plants, SYLF domains are attached to a second functional domain with membrane lipid binding capabilities (Hasegawa et al., 2011; Wywial and Singh, 2010). In BefA and other bacterial homologs, the SYLF domain comprises the majority of the protein and is not attached to other functional regions making predictions about its mechanism in host beta cells difficult. Further work to identify potential binding partners

of BefA, will hopefully shed more light on how it elicits beta cell proliferation. Furthermore, this information would be valuable in learning more about the function BefA provides to bacteria.

We also showed evidence that a direct interaction between the SYLF domain of BefA and pancreatic cells results in larval beta cell mass expansion. This observation raises many interesting questions about the dynamics of BefA's transport from the gut lumen to the pancreas. For instance, maintaining the integrity of the intestinal barrier between bacteria within the lumen and the host cells is important to maintain host health and to control inflammation (Natividad and Verdu, 2013). Breaches of this barrier, by both bacteria and their products, are often associated with disease or pathogen infections (Mu et al., 2017). However, BefA is a bacterial product that promotes host development and healthy glucose homeostasis so we must begin to consider how the host controls and interprets microbial factors that cross this barrier. Additionally, once out of the gut lumen, there are multiple physical pathways that BefA could use to reach its systemic target. In general, physical travel from the gut to the pancreas involves a relatively great distance, across a range of conditions. The fact that BefA is able to interact directly with the distant pancreatic cells, suggests that it can withstand these environmental stresses, although we have no idea whether it diffuses across these landscapes or is actively transported and protected by a host mechanism. One scenario is that epithelial or immune cells take up BefA from the gut lumen and subsequently deposit it into the blood stream where it would be shuttled directly to the highly vascularized pancreatic cells via the hepatic portal vein. Alternatively, the extra pancreatic duct provides a direct luminal pathway from the gut to the pancreas. However, we have not ruled out the possibility that

BefA is also able to utilize indirect signaling mechanisms to act systemically. Future work is required to fully understand the biologically relevant manner in which BefA mediates beta cell proliferation. Nevertheless, the possibility that BefA and beta cells directly interact holds promise for its use in promoting *in vitro* expansion of beta cell tissue. This is an important step for treatments utilizing both human donor tissue and stem cell derived material for transplantation as a method to treat diabetes(Balamurugan et al., 2012).

Novel protein folding patterns have not been described in several years. Our discovery of BefA's structure suggests that host-associated bacterial genomes are an untapped source of protein structural novelty. Further investigation into the biochemical properties of these molecules could help to expand our knowledge of protein functions and their potential uses for human wellbeing.

MATERIALS AND METHODS

Protein expression & purification

The *befA* gene was expressed and purified as previously described (Hill et al. 2016). Additionally, BefA protein containing selenium methionine was produced as described by Van Duyne et al, and purified using the same methods as native BefA (Hill et al. 2016). The nucleotide sequences corresponding to the Befa^{99} and BefA^{133} truncations were amplified using the same reverse PCR primer previously published for amplifying BefA (), which was paired with the following forward PCR primers for each truncation respectively: 5'-GGCCATATGATGaagacggcgaaagaggcgagg-3' and 5'-GGCCATATGATGggttatgcggtgttcgattcgcgc-3'. Each construct was cloned into the pET-

21b plasmid (Novagen, Darmstadt, Germany) and expressed and purified using a C terminal His□Tag® as previously described for the native BefA protein (Hill et al. 2016).

BefA crystallization and structure determination

Purified BefA and selenium methionine substituted BefA (semetBefA) were crystalized from a starting concentration of 10-16 mg/mL in a reservoir solution of: 24-25% PEG 3350 (Hampton Biologicals), 0.1M citric acid pH 3.5, and for semetBefA, 1mM TCEP. Hanging drop vapor diffusion with ratios of protein to reservoir solution of either 1:1 or 2:1, resulted in crystals within 7-10 days. Crystals were cryoprotected in the reservoir solution plus 20% PEG 200 (Hampton Biologicals). Crystals were flash frozen in liquid nitrogen for data collection at the Advanced Light Source in Berkeley, CA beamline 5.0.2 using the Pilatus detector.

Gnotobiotic zebrafish

All zebrafish experiments were performed using protocols approved by the University of Oregon Institutional Care and Use Committee and followed standard protocols. Zebrafish embryos were derived GF as previously described (Bates et al., 2006). For experiments involving the treatment of larvae with purified protein, the protein was added directly to the embryo media at a final concentration of 500 ng/mL. *Tg(-1.0insulin:eGFP)* larvae were incubated with proteins from 4-6 dpf and beta cell mass was determined as previously described (Hill et al., 2016).

Zebrafish primary islet dissection and treatment

Gut dissections were performed on 4 dpf *Tg(-1.0insulin:eGFP)* larval zebrafish as previously described (Bates et al., 2006). Since the pancreas is tightly associated with the gut, it came along readily with dissected gut tissue. The top portion of the gut bulb, and the head of the pancreas just below the islet, was dissected away from the rest of the gastronitestinal organs, so all that remained was the primary islet and a small amount of surrounding tissue, which varied in make-up between dissections, but likely included some amounts of exocrine pancreas, duct, and intestine. We found that leaving this small amount of supporting tissue was necessary to keep the islet in tact for accurate beta cell quantification. Each dissection was transferred into a sterile PCR tube containing 50 uL of Leibovitz's L15 Media with GlutaMax Supplement (Thermo Fischer) supplemented with 100 µg/mL penn/strep. Protein treatments were added at a final concentration of 500 ng/mL and allowed to incubate at RT for 48 hours. Following treatment, single islets were mounted directly onto a slide with 5 uL of Prolong with DAPI (Molecular Probes) and a coverslip. Each islet was imaged immediately after mounting on an Olympus confocal microscope and then analyzed as previously described.

Statistical Analysis

Statistical analysis for beta cell mass determination was conducted in the same manner as was described previously (Hill et al., 2016).

CHAPTER IV

CONCLUDING REMARKS AND FUTURE DIRECTIONS

Using the zebrafish model system, we have described a new role for the microbiota in pancreatic development. By taking advantage of the natural and coincidental timelines of larval pancreas development and initial microbial colonization, we were able to observe and characterize this process. Upon hatching, the larval gut becomes colonized by bacteria within a matter of hours (Jemielita et al., 2014). At the same time, the newly fated pancreatic beta cells begin to expand their numbers to meet the growing nutritional needs of the young fish (Moro et al., 2009). Similarly, humans are limited to bacterial exposures in the womb, but acquire a rich intestinal microbiota after birth, during the same period when beta cell mass is growing (Gregg et al., 2012; Mueller et al., 2015). From this work, we now know that microbial colonization with specific bacteria is required for robust beta cell expansion in larval zebrafish. This powerful model system provided us the high throughput platform required to systematically identify the microbial signal responsible for conveying this important phenotype. The genetic and imaging tools already available for studying zebrafish will be useful for continuing our research into uncovering the molecular mechanism through which the host senses and interprets BefA.

One area of focus that zebrafish will lend a particular advantage is in observing the natural trafficking of BefA. With a combination of tools such as transgenic reporter lines (labeling important tissues like the pancreatic ducts) and fluorescently tagged BefA constructs, we can visualize the *in vivo* movement of BefA. This approach will also be

useful for observing how BefA interacts with beta cells, which could give us clues as to the host sensing machinery that is required. These experiments will be advantageous in designing methods of delivery for terrestrial organisms in order to study the effects of BefA in mammalian models, as well as for planning potential drug delivery mechanisms in humans.

The discovery of functionally conserved BefA homologs in human-associated bacteria is encouraging for the potential of BefA to be developed into a therapeutic some day. It suggests the possibility of a conserved signaling pathway that can be explored and exploited. Experiments that utilize mammalian beta cells will be an important next step to test the translational potential of BefA. For instance the addition of BefA to primary islet cultures from murine as well as primate and even human sources will provide a simple system to test whether the activity of BefA or its homologs is limited to zebrafish. One recent piece of evidence that mammalian beta cells require microbes during development is the finding that juvenile GF mice have significantly reduced beta cell mass compared to their CV counterparts. One obvious experiment would be to administer BefA to GF mice in order to test whether it can boost beta cell mass.

In addition to understanding whether BefA's activity is restricted to specific vertebrate lineages, it will also be important to understand whether or not its activity is restricted developmentally. For instance, beta cells may have an expiration time on their ability to sense BefA. So far, we have shown that BefA is sufficient to induce beta cell proliferation during early larval stages, when these cells are naturally programmed to proliferate and expand. Adult beta cells are biologically different; they are largely quiescent, which may alter their responsiveness to BefA. The effectiveness of BefA at

different stages of life will be important in determining the extent of its therapeutic potential. Both zebrafish and mice would provide suitable platforms for investigating this question. There are good models of diabetes in mice whereby genetic lesions result in spontaneous autoimmune destruction of beta cells, allowing for study of the environmental influences on disease progression (Alam et al., 2011). For example the Blaser lab has found that early exposure to antibiotics in these mouse models results in a higher penetrance of disease (Cox and Blaser, 2014). In zebrafish, beta cell ablation models allow researchers to temporally control the destruction of beta cells and study the regenerative capacity of exogenous molecules at any developmental stage (Curado et al., 2007).

Lastly, as a rare yet important product of the microbiota, BefA may be lost more easily from a less diverse host-associated community, an event that may have consequences for development and subsequently for host health. The *befA* gene, however, is rare in bacterial genomes and our current efforts have failed to detect it in metagenomic shotgun data, which only contain sequences from genomes of the most abundant species and widely conserved genes. To begin to understand whether this is a relevant hypothesis, we should design targeted search efforts for BefA within metagenomic samples taken from a spectrum of healthy, pre-diabetic and diabetic individuals. By correlating BefA abundance, or abundance of species that produce BefA, to host health status we might begin to understand whether it's depletion or loss is a critical step in disease progression.

The discovery of BefA illustrates the vast potential waiting to be uncovered within the microbiota. Future studies across gnotobiotic systems will benefit from

learning more about this newly appreciated source of bioactive molecules, many of which have likely shaped the evolution of the animals containing them. I hope we will continue to be amazed by the discoveries surrounding this field for many years to come.

REFERENCES CITED

Akerblom, H.K., Vaarala, O., Hyöty, H., Ilonen, J., and Knip, M. (2002). Environmental factors in the etiology of type 1 diabetes. Am. J. Med. Genet. *115*, 18–29.

Alam, C., Bittoun, E., Bhagwat, D., Valkonen, S., Saari, A., Jaakkola, U., Eerola, E., Huovinen, P., and Hänninen, A. (2011). Effects of a germ-free environment on gut immune regulation and diabetes progression in non-obese diabetic (NOD) mice. Diabetologia *54*, 1398–1406.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. *25*, 3389–3402.

Andersson, O., Adams, B.A., Yoo, D., Ellis, G.C., Gut, P., Anderson, R.M., German, M.S., and Stainier, D.Y.R. (2012). Adenosine signaling promotes regeneration of pancreatic beta cells in vivo. Cell Metab. *15*, 885–894.

Bach, J.-F. (2002). The Effect of Infections on Susceptibility to Autoimmune and Allergic Diseases. N. Engl. J. Med. *347*, 911–920.

Balamurugan, A.N., Kumaravel, V., Pugazhenthi, S., and Naziruddin, B. (2012). Islet Cell Biology, Regeneration, and Transplantation. Int. J. Endocrinol. *2012*, 1–3.

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

Berger, M., Scheel, D.W., Macias, H., Miyatsuka, T., Kim, H., Hoang, P., Ku, G.M., Honig, G., Liou, A., Tang, Y., et al. (2015). Gαi/o-coupled receptor signaling restricts pancreatic beta-cell expansion. Proc. Natl. Acad. Sci. U. S. A. *112*, 2888–2893.

Biasucci, G., Rubini, M., Riboni, S., Morelli, L., Bessi, E., and Retetangos, C. (2010). Mode of delivery affects the bacterial community in the newborn gut. Early Hum. Dev. *86 Suppl 1*, 13–15.

Biemar, F., Argenton, F., Schmidtke, R., Epperlein, S., Peers, B., and Driever, W. (2001). Pancreas Development in Zebrafish: Early Dispersed Appearance of Endocrine Hormone Expressing Cells and Their Convergence to Form the Definitive Islet. Dev. Biol. *230*, 189–203.

Blaser, M.J., and Falkow, S. (2009). What are the consequences of the disappearing human microbiota? Nat. Rev. Microbiol. *7*, 887–894.

Bomar, L., Stephens, W.Z., Nelson, M.C., Velle, K., Guillemin, K., and Graf, J. (2013). Draft Genome Sequence of Aeromonas veronii Hm21, a Symbiotic Isolate from the Medicinal Leech Digestive Tract. Genome Announc. *1*.

Brown, C.T., Davis-Richardson, A.G., Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Casella, G., Drew, J.C., Ilonen, J., Knip, M., et al. (2011). Gut microbiome metagenomics analysis suggests a functional model for the development of autoimmunity for type 1 diabetes. PloS One *6*, e25792.

Brown, K., Godovannyi, A., Ma, C., Zhang, Y., Ahmadi-Vand, Z., Dai, C., Gorzelak, M.A., Chan, Y., Chan, J.M., Lochner, A., et al. (2016). Prolonged antibiotic treatment induces a diabetogenic intestinal microbiome that accelerates diabetes in NOD mice. ISME J. *10*, 321–332.

Candon, S., Perez-Arroyo, A., Marquet, C., Valette, F., Foray, A.-P., Pelletier, B., Milani, C., Ventura, M., Bach, J.-F., and Chatenoud, L. (2015). Antibiotics in Early Life Alter the Gut Microbiome and Increase Disease Incidence in a Spontaneous Mouse Model of Autoimmune Insulin-Dependent Diabetes. PLOS ONE *10*, e0125448.

Chang, C., and Lin, H. (2016). Dysbiosis in gastrointestinal disorders. Best Pract. Res. Clin. Gastroenterol. *30*, 3–15.

Cheesman, S.E., Neal, J.T., Mittge, E., Seredick, B.M., and Guillemin, K. (2011). Epithelial cell proliferation in the developing zebrafish intestine is regulated by the Wnt pathway and microbial signaling via Myd88. Proc. Natl. Acad. Sci. U. S. A. *108 Suppl 1*, 4570–4577.

Chung, W.-S., Andersson, O., Row, R., Kimelman, D., and Stainier, D.Y.R. (2010). Suppression of Alk8-mediated Bmp signaling cell-autonomously induces pancreatic -cells in zebrafish. Proc. Natl. Acad. Sci. *107*, 1142–1147.

Cox, L.M., and Blaser, M.J. (2014). Antibiotics in early life and obesity. Nat. Rev. Endocrinol. *11*, 182–190.

Cox, L.M., Yamanishi, S., Sohn, J., Alekseyenko, A.V., Leung, J.M., Cho, I., Kim, S.G., Li, H., Gao, Z., Mahana, D., et al. (2014). Altering the Intestinal Microbiota during a Critical Developmental Window Has Lasting Metabolic Consequences. Cell *158*, 705–721.

Curado, S., Anderson, R.M., Jungblut, B., Mumm, J., Schroeter, E., and Stainier, D.Y.R. (2007). Conditional targeted cell ablation in zebrafish: A new tool for regeneration studies. Dev. Dyn. *236*, 1025–1035.

DeGruttola, A.K., Low, D., Mizoguchi, A., and Mizoguchi, E. (2016). Current Understanding of Dysbiosis in Disease in Human and Animal Models: Inflamm. Bowel Dis. *22*, 1137–1150.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V., Lescot, M., et al. (2008). Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. *36*, W465–W469.

diIorio, P.J., Moss, J.B., Sbrogna, J.L., Karlstrom, R.O., and Moss, L.G. (2002). Sonic hedgehog is required early in pancreatic islet development. Dev. Biol. *244*, 75–84.

Dominguez-Bello, M.G., Costello, E.K., Contreras, M., Magris, M., Hidalgo, G., Fierer, N., and Knight, R. (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.

Dong, P.D.S., Munson, C.A., Norton, W., Crosnier, C., Pan, X., Gong, Z., Neumann, C.J., and Stainier, D.Y.R. (2007). Fgf10 regulates hepatopancreatic ductal system patterning and differentiation. Nat. Genet. *39*, 397–402.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113–113.

Farkas, A.M., Panea, C., Goto, Y., Nakato, G., Galan-Diez, M., Narushima, S., Honda, K., and Ivanov, I.I. (2015). Induction of Th17 cells by segmented filamentous bacteria in the murine intestine. J. Immunol. Methods *421*, 104–111.

Farney, A.C., Sutherland, D.E.R., and Opara, E.C. (2016). Evolution of Islet Transplantation for the Last 30 Years. Pancreas *45*, 8–20.

Field, H.A., Dong, P.D.S., Beis, D., and Stainier, D.Y.R. (2003). Formation of the digestive system in zebrafish. II. Pancreas morphogenesis. Dev. Biol. *261*, 197–208.

Figliuzzi, M., Bonandrini, B., Cattaneo, I., Remuzzi, G., and Remuzzi, A. (2010). Effect of inborn pancreatic islet deficit in the Munich Wister Frömter rat. Islets *2*, 318–322.

Georgia, S., and Bhushan, A. (2004). beta cell replication is the primary mechanism for maintaining postnatal beta cell mass. J. Clin. Invest. *114*, 963–968.

Georgia, S., Soliz, R., Li, M., Zhang, P., and Bhushan, A. (2006). p57 and Hes1 coordinate cell cycle exit with self-renewal of pancreatic progenitors. Dev. Biol. *298*, 22–31.

Giongo, A., Gano, K.A., Crabb, D.B., Mukherjee, N., Novelo, L.L., Casella, G., Drew, J.C., Ilonen, J., Knip, M., Hyöty, H., et al. (2011). Toward defining the autoimmune microbiome for type 1 diabetes. ISME J. *5*, 82–91.

Graf, J. (1999). Symbiosis of Aeromonas veronii biovar sobria and Hirudo medicinalis, the medicinal leech: a novel model for digestive tract associations. Infect. Immun. *67*, 1–7.

Gregg, B.E., Moore, P.C., Demozay, D., Hall, B.A., Li, M., Husain, A., Wright, A.J., Atkinson, M.A., and Rhodes, C.J. (2012). Formation of a human beta-cell population within pancreatic islets is set early in life. J. Clin. Endocrinol. Metab. *97*, 3197–3206.

Grosberg, R.K., and Strathmann, R.R. (2007). The Evolution of Multicellularity: A Minor Major Transition? Annu. Rev. Ecol. Evol. Syst. *38*, 621–654.

Guindon, S., and Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. Syst. Biol. *52*, 696–704.

Gülden, E., Wong, F.S., and Wen, L. (2015). The gut microbiota and Type 1 Diabetes. Clin. Immunol. Orlando Fla *159*, 143–153.

Guo, X., Ran, C., Zhang, Z., He, S., Jin, M., and Zhou, Z. (2017). The Growth-Promoting Effect of Dietary Nucleotides in Fish Is Associated with an Intestinal Microbiota-Mediated Reduction in Energy Expenditure. J. Nutr. *147*, 781–788.

Gut, P., Baeza-Raja, B., Andersson, O., Hasenkamp, L., Hsiao, J., Hesselson, D., Akassoglou, K., Verdin, E., Hirschey, M.D., and Stainier, D.Y.R. (2013). Whole-organism screening for gluconeogenesis identifies activators of fasting metabolism. Nat. Chem. Biol. *9*, 97–104.

Gut, P., Reischauer, S., Stainier, D.Y.R., and Arnaout, R. (2017). LITTLE FISH, BIG DATA: ZEBRAFISH AS A MODEL FOR CARDIOVASCULAR AND METABOLIC DISEASE. Physiol. Rev. *97*, 889–938.

Hasegawa, J., Tokuda, E., Tenno, T., Tsujita, K., Sawai, H., Hiroaki, H., Takenawa, T., and Itoh, T. (2011). SH3YL1 regulates dorsal ruffle formation by a novel phosphoinositide-binding domain. J. Cell Biol. *193*, 901–916.

Hesselson, D., Anderson, R.M., Beinat, M., and Stainier, D.Y.R. (2009). Distinct populations of quiescent and proliferative pancreatic beta-cells identified by HOTcre mediated labeling. Proc. Natl. Acad. Sci. U. S. A. *106*, 14896–14901.

Hill, J.H., Franzosa, E.A., Huttenhower, C., and Guillemin, K. (2016). A conserved bacterial protein induces pancreatic beta cell expansion during zebrafish development. eLife *5*.

Hu, Y., Peng, J., Tai, N., Hu, C., Zhang, X., Wong, F.S., and Wen, L. (2015). Maternal Antibiotic Treatment Protects Offspring from Diabetes Development in Nonobese Diabetic Mice by Generation of Tolerogenic APCs. J. Immunol. *195*, 4176–4184.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J.H., Chinwalla, A.T., Creasy, H.H., Earl, A.M., FitzGerald, M.G., Fulton, R.S., et al. (2012). Structure, function and diversity of the healthy human microbiome. Nature *486*, 207–214.

Integrative HMP (iHMP) Research Network Consortium (2014). The Integrative Human Microbiome Project: dynamic analysis of microbiome-host omics profiles during periods of human health and disease. Cell Host Microbe *16*, 276–289.

Ivanov, I.I., Frutos, R. de L., Manel, N., Yoshinaga, K., Rifkin, D.B., Sartor, R.B., Finlay, B.B., and Littman, D.R. (2008). Specific microbiota direct the differentiation of IL-17-producing T-helper cells in the mucosa of the small intestine. Cell Host Microbe *4*, 337–349.

Jemielita, M., Taormina, M.J., Burns, A.R., Hampton, J.S., Rolig, A.S., Guillemin, K., and Parthasarathy, R. (2014). Spatial and temporal features of the growth of a bacterial species colonizing the zebrafish gut. mBio *5*.

Johannesson, B., Sui, L., Freytes, D.O., Creusot, R.J., and Egli, D. (2015). Toward beta cell replacement for diabetes. EMBO J. *34*, 841–855.

Kaijser, M., Bonamy, A.-K.E., Akre, O., Cnattingius, S., Granath, F., Norman, M., and Ekbom, A. (2009). Perinatal risk factors for diabetes in later life. Diabetes *58*, 523–526.

Kassem, S.A., Ariel, I., Thornton, P.S., Scheimberg, I., and Glaser, B. (2000). Beta-cell proliferation and apoptosis in the developing normal human pancreas and in hyperinsulinism of infancy. Diabetes *49*, 1325–1333.

Kimmel, R.A., Onder, L., Wilfinger, A., Ellertsdottir, E., and Meyer, D. (2011). Requirement for Pdx1 in specification of latent endocrine progenitors in zebrafish. BMC Biol. *9*, 75.

Kinkel, M.D., and Prince, V.E. (2009). On the diabetic menu: Zebrafish as a model for pancreas development and function. BioEssays *31*, 139–152.

Klinke, D.J. (2008). Extent of Beta Cell Destruction Is Important but Insufficient to Predict the Onset of Type 1 Diabetes Mellitus. PLoS ONE *3*, e1374.

Knip, M., Veijola, R., Virtanen, S.M., Hyöty, H., Vaarala, O., and Akerblom, H.K. (2005). Environmental triggers and determinants of type 1 diabetes. Diabetes *54 Suppl 2*, S125–S136.

Kostic, A.D., Gevers, D., Siljander, H., Vatanen, T., Hyötyläinen, T., Hämäläinen, A.-M., Peet, A., Tillmann, V., Pöhö, P., Mattila, I., et al. (2015). The dynamics of the human infant gut microbiome in development and in progression toward type 1 diabetes. Cell Host Microbe *17*, 260–273.

Kriegel, M.A., Sefik, E., Hill, J.A., Wu, H.-J., Benoist, C., and Mathis, D. (2011). Naturally transmitted segmented filamentous bacteria segregate with diabetes protection in nonobese diabetic mice. Proc. Natl. Acad. Sci. U. S. A. *108*, 11548–11553.

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256–2268.

Krissinel, E., and Henrick, K. (2005). Multiple Alignment of Protein Structures in Three Dimensions. In Computational Life Sciences, M. R. Berthold, R.C. Glen, K. Diederichs, O. Kohlbacher, and I. Fischer, eds. (Berlin, Heidelberg: Springer Berlin Heidelberg), pp. 67–78.

Kumar, M. (2003). Pancreas specification: a budding question. Curr. Opin. Genet. Dev. *13*, 401–407.

Levy, M., Kolodziejczyk, A.A., Thaiss, C.A., and Elinav, E. (2017). Dysbiosis and the immune system. Nat. Rev. Immunol. *17*, 219–232.

Ley, R.E., Turnbaugh, P.J., Klein, S., and Gordon, J.I. (2006). Microbial ecology: human gut microbes associated with obesity. Nature *444*, 1022–1023.

Li, M., Page-McCaw, P., and Chen, W. (2015). FGF1 Mediates Overnutrition-Induced Compensatory ?-Cell Differentiation. Diabetes db150085.

Macpherson, A.J., and Harris, N.L. (2004). Interactions between commensal intestinal bacteria and the immune system. Nat. Rev. Immunol. *4*, 478–485.

Maddison, L.A., and Chen, W. (2012). Nutrient Excess Stimulates ?-Cell Neogenesis in Zebrafish. Diabetes *61*, 2517–2524.

Maltz, M., and Graf, J. (2011). The type II secretion system is essential for erythrocyte lysis and gut colonization by the leech digestive tract symbiont Aeromonas veronii. Appl. Environ. Microbiol. *77*, 597–603.

Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R., et al. (2017). CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. *45*, D200–D203.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., Dubilier, N., Eberl, G., Fukami, T., Gilbert, S.F., et al. (2013). Animals in a bacterial world, a new imperative for the life sciences. Proc. Natl. Acad. Sci. U. S. A. *110*, 3229–3236.

Milligan-Myhre, K., Charette, J.R., Phennicie, R.T., Stephens, W.Z., Rawls, J.F., Guillemin, K., and Kim, C.H. (2011). Study of host-microbe interactions in zebrafish. Methods Cell Biol. *105*, 87–116.

Milligan-Myhre, K., Small, C.M., Mittge, E.K., Agarwal, M., Currey, M., Cresko, W.A., and Guillemin, K. (2016). Innate immune responses to gut microbiota differ between

oceanic and freshwater threespine stickleback populations. Dis. Model. Mech. *9*, 187–198.

Minchin, J.E.N., and Rawls, J.F. (2017). A classification system for zebrafish adipose tissues. Dis. Model. Mech. *10*, 797–809.

Morgan, X.C., Segata, N., and Huttenhower, C. (2013). Biodiversity and functional genomics in the human microbiome. Trends Genet. *29*, 51–58.

Moro, E., Gnügge, L., Braghetta, P., Bortolussi, M., and Argenton, F. (2009). Analysis of beta cell proliferation dynamics in zebrafish. Dev. Biol. *332*, 299–308.

Mu, Q., Kirby, J., Reilly, C.M., and Luo, X.M. (2017). Leaky Gut As a Danger Signal for Autoimmune Diseases. Front. Immunol. *8*, 598.

Mueller, N.T., Bakacs, E., Combellick, J., Grigoryan, Z., and Dominguez-Bello, M.G. (2015). The infant microbiome development: mom matters. Trends Mol. Med. *21*, 109–117.

Murgas Torrazza, R., and Neu, J. (2011). The developing intestinal microbiome and its relationship to health and disease in the neonate. J. Perinatol. Off. J. Calif. Perinat. Assoc. *31 Suppl 1*, S29–S34.

Natividad, J.M.M., and Verdu, E.F. (2013). Modulation of intestinal barrier by intestinal microbiota: Pathological and therapeutic implications. Pharmacol. Res. *69*, 42–51.

Nicholson, J.K., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., Jia, W., and Pettersson, S. (2012). Host-Gut Microbiota Metabolic Interactions. Science *336*, 1262–1267.

Palmer, C., Bik, E.M., DiGiulio, D.B., Relman, D.A., and Brown, P.O. (2007). Development of the human infant intestinal microbiota. PLoS Biol. *5*, e177.

Peng, J., Narasimhan, S., Marchesi, J.R., Benson, A., Wong, F.S., and Wen, L. (2014). Long term effect of gut microbiota transfer on diabetes development. J. Autoimmun. *53*, 85–94.

Petersen, C., and Round, J.L. (2014). Defining dysbiosis and its influence on host immunity and disease. Cell. Microbiol. *16*, 1024–1033.

Petersen, T.N., Brunak, S., von Heijne, G., and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods *8*, 785–786.

Robertson, A.S., Allwood, E.G., Smith, A.P.C., Gardiner, F.C., Costa, R., Winder, S.J., and Ayscough, K.R. (2009). The WASP homologue Las17 activates the novel actin-regulatory activity of Ysc84 to promote endocytosis in yeast. Mol. Biol. Cell *20*, 1618–1628.

Rolig, A.S., Parthasarathy, R., Burns, A.R., Bohannan, B.J.M., and Guillemin, K. (2015). Individual Members of the Microbiota Disproportionately Modulate Host Innate Immune Responses. Cell Host Microbe *18*, 613–620.

Rook, G.A.W., Adams, V., Palmer, R., Brunet, L.R., Hunt, J., and Martinelli, R. (2004). Mycobacteria and other environmental organisms as immunomodulators for immunoregulatory disorders. Springer Semin. Immunopathol. *25*, 237–255.

Roy, A., Kucukural, A., and Zhang, Y. (2010). I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. *5*, 725–738.

Santoro, M.M. (2014). Zebrafish as a model to explore cell metabolism. Trends Endocrinol. Metab. *25*, 546–554.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nat. Methods *9*, 676–682.

Segata, N., Waldron, L., Ballarini, A., Narasimhan, V., Jousson, O., and Huttenhower, C. (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. Nat. Methods *9*, 811–814.

Semova, I., Carten, J.D., Stombaugh, J., Mackey, L.C., Knight, R., Farber, S.A., and Rawls, J.F. (2012a). Microbiota regulate intestinal absorption and metabolism of fatty acids in the zebrafish. Cell Host Microbe *12*, 277–288.

Semova, I., Carten, J., Stombaugh, J., Mackey, L., Knight, R., Farber, S., and Rawls, J. (2012b). Microbiota Regulate Intestinal Absorption and Metabolism of Fatty Acids in the Zebrafish. Cell Host Microbe *12*, 277–288.

Seth, A., Stemple, D.L., and Barroso, I. (2013). The emerging use of zebrafish to model metabolic disease. Dis. Model. Mech. *6*, 1080–1088.

Sommer, F., and Bäckhed, F. (2013). The gut microbiota--masters of host development and physiology. Nat. Rev. Microbiol. *11*, 227–238.

Stanger, B.Z., Tanaka, A.J., and Melton, D.A. (2007). Organ size is limited by the number of embryonic progenitor cells in the pancreas but not the liver. Nature *445*, 886–891.

Stephens, W.Z., Wiles, T.J., Martinez, E.S., Jemielita, M., Burns, A.R., Parthasarathy, R., Bohannan, B.J.M., and Guillemin, K. (2015). Identification of Population Bottlenecks and Colonization Factors during Assembly of Bacterial Communities within the Zebrafish Intestine. mBio *6*, e01163–01115.

Tai, N., Wong, F.S., and Wen, L. (2015). The role of gut microbiota in the development of type 1, type 2 diabetes mellitus and obesity. Rev. Endocr. Metab. Disord. *16*, 55–65.

Teta, M., Rankin, M.M., Long, S.Y., Stein, G.M., and Kushner, J.A. (2007). Growth and regeneration of adult beta cells does not involve specialized progenitors. Dev. Cell *12*, 817–826.

Thisse, B., Heyer, V., Lux, A., Alunni, V., Degrave, A., Seiliez, I., Kirchner, J., Parkhill, J.-P., and Thisse, C. (2004). Spatial and temporal expression of the zebrafish genome by large-scale in situ hybridization screening. Methods Cell Biol. *77*, 505–519.

Tonikian, R., Xin, X., Toret, C.P., Gfeller, D., Landgraf, C., Panni, S., Paoluzi, S., Castagnoli, L., Currell, B., Seshagiri, S., et al. (2009). Bayesian modeling of the yeast SH3 domain interactome predicts spatiotemporal dynamics of endocytosis proteins. PLoS Biol. *7*, e1000218.

Urbanek, A., Chan, R., and Ayscough, K. (2015). Function and interactions of the Ysc84/SH3yl1 family of actin- and lipid-binding proteins. Biochem. Soc. Trans. *43*, 111–116.

Valladares, R., Sankar, D., Li, N., Williams, E., Lai, K.-K., Abdelgeliel, A.S., Gonzalez, C.F., Wasserfall, C.H., Larkin, J., Schatz, D., et al. (2010). Lactobacillus johnsonii N6.2 mitigates the development of type 1 diabetes in BB-DP rats. PloS One *5*, e10507.

Wang, L., Wang, F.-S., and Gershwin, M.E. (2015). Human autoimmune diseases: a comprehensive update. J. Intern. Med. *278*, 369–395.

Wang, Y., Rovira, M., Yusuff, S., and Parsons, M.J. (2011). Genetic inducible fate mapping in larval zebrafish reveals origins of adult insulin-producing -cells. Development *138*, 609–617.

Wywial, E., and Singh, S.M. (2010). Identification and structural characterization of FYVE domain-containing proteins of Arabidopsis thaliana. BMC Plant Biol. *10*, 157.

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., and Zhang, Y. (2014). The I-TASSER Suite: protein structure and function prediction. Nat. Methods *12*, 7–8.

Yatsunenko, T., Rey, F.E., Manary, M.J., Trehan, I., Dominguez-Bello, M.G., Contreras, M., Magris, M., Hidalgo, G., Baldassano, R.N., Anokhin, A.P., et al. (2012). Human gut microbiome viewed across age and geography. Nature *486*, 222–227.

Yurkovetskiy, L., Burrows, M., Khan, A.A., Graham, L., Volchkov, P., Becker, L., Antonopoulos, D., Umesaki, Y., and Chervonsky, A.V. (2013). Gender bias in autoimmunity is influenced by microbiota. Immunity *39*, 400–412.

Zhang, Y. (2008). I-TASSER server for protein 3D structure prediction. BMC Bioinformatics *9*, 40.

(2017). The zebrafish: disease models and chemical screens (Cambridge, MA San Diego, MA London Oxford: Elsevier, Academic Press).