MOLECULAR CHARACTERIZATION OF THE INTESTINAL MICROBIOME OF *NASO TONGANUS* AND ITS UNUSUALLY LARGE POLYPLOID SYMBIONT, *EPULOPISCIUM* TYPE B

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Jennifer Elizabeth Mendell

January 2012

MOLECULAR CHARACTERIZATION OF THE INTESTINAL MICROBIOME OF *NASO TONGANUS* AND ITS UNUSUALLY LARGE POLYPLOID SYMBIONT, *EPULOPISCIUM* TYPE B

Jennifer Elizabeth Mendell, PhD

Cornell University 2012

Intestinal symbionts provide their host with digestive enzymes that convert a substantial portion of the dietary components to fermentation products, which are more readily absorbed by the intestinal epithelia. The objectives of this study were to 1) characterize the abundant DNA found in the unusually large intestinal symbiont, *Epulopiscium*, 2) use 16S rRNA gene clone libraries and T-RFLP to describe the microbial diversity in the gastrointestinal tract of *Naso tonganus*, the host fish of *Epulopiscium*, and 3) to understand how *Epulopiscium* potentially influences the overall structure of this microbial community.

Quantitative PCR targeting single-copy genes was utilized on two populations of *Epulopiscium* cells as well as genomic DNA. Results from this study provided evidence that *Epulopiscium* contains tens of thousands of copies of an approximately 4Mb genome. In addition, copy number is positively correlated with *Epulopiscium* cell size. This magnitude of polyploidy has never been reported in a bacterium.

A combination of 16S rRNA gene clone libraries and T-RFLP was utilized to describe the total microbial community in the intestinal tract of four individual *Naso tonganus* that did not harbor populations of *Epulopiscium*. Clones isolated belonged to the 10 different bacterial phyla, with the majority of the clones affiliating with known fermentative phyla. The bacterial community composition differed in different segments of the gastrointestinal tract. These results are in stark contrast to other studies looking at the microbial community of omnivorous fishes. There were no

statistically significant differences in the T-RFLP profiles generated from segment IV and V, however multi-dimensional scaling (MDS) plots did show clustering of the same segment from different fish.

In order to assess how *Epulopiscium* potentially influences this microbial community, these same T-RFLP analyses were performed using an individual fish that had dense populations of these cells. R- values support differences in the community profiles indicating that the extreme polyploidy in *Epulopiscium* may allow it to replace some of the numerically dominant groups in this gut ecosystem.

BIOLGRAPHICAL SKETCH

Jennifer Elizabeth Mendell was born on August 22, 1969 in El Paso, TX. She attended Silver Lake Regional High School in Kingston, MA. In 1992 she graduated from Wheaton College in Norton, MA with a Bachelor of Arts in Psychology. She returned to school in 1997 and in 1999 graduated magna cum laude from Massasoit Community College with an Associate of Arts and Sciences and an Associate of Science in Nursing. She completed her Registered Nurse licensure in 1999. She continued her studies as a non-matriculated student at Bridgewater State College before entering the graduate program at Cornell University in the Department of Microbiology in 2001. She has accepted a full-time tenure track position at Bridgewater State University upon completion of her PhD.

ACKNOWLEDGMENTS

I would like to express thanks and gratitude to my committee chairperson, Dr. Esther Angert. None of this work would have been possible with out her patience, guidance and support. I would also like to thank my other committee members, Dr. Anthony Hay and Dr. John Helmann for their guidance and advice over the years. Angert Lab members, Joe Flint, Julie Frey, Bekah Ward, Dave Miller, Magnus Bergkvist and especially John Rose have provided me with sound advice, skills and laughter. Thanks should also be given to all of the undergraduates in the Angert Lab who have shown me time and again why I am so passionate about undergraduate education. A special thank you to Roxanne Beinart and Mimi Yan for their tremendous efforts on several of the projects described herein.

Additionally, I would like to thank our collaborators, Kendall Clements, J. Howard Choat, Will Robbins and David Raubenheimer. The work of the Angert Lab could not happen without their support.

I would also like to thank the Department of Microbiology for the opportunity to pursue graduate studies, and for their support over the years. In particular the wonderful support staff: Cathy Shappell, Patricia Lisk, Doreen Dineen and Shirley Crammer deserve special recognition. A heartfelt thank you to Sue Merkle who supported, cultivated and championed my development as an educator. She is a true inspiration.

Nancy Trautmann and Ron Reed also need to be thanked for all of their support, their encouragement, understanding and friendship. They are truly remarkable people and I have learned so much about what it means to be a teacher from each of them.

I would be remiss if I did not thank those at Bridgewater State University who have believed in me, supported me and provided me with such valuable input through this process. Jeff Bowen, Chris Bloch, Kevin Curry and Meri Krevosky, you lifted me up when I didn't think I could go on, you made me laugh when I didn't think I could smile and you made me believe in myself. Meri deserves special recognition as someone who took me into the fold so willingly not even knowing who I was when I came to teach an A&P lab at BSU. She shared her resources, time and knowledge. But she shared something much greater. She balances life so beautifully, and as a woman in science, she is such an inspiration not only to me, but to all female scientists. Her kindness and thoughtfulness and honesty are so cherished.

Finally to my students, especially Yasmin Baig and Jennifer Conway, who have inspired me, supported me, made me laugh, reminded me that this is the path I was meant to follow and always encouraged me to follow my dream, I thank you.

TABLE OF CONTENTS

CHAPTER 3

BACTERIAL DIVERSITY IN THE GASTROINTESTINAL TRACT OF THE OMNIVOROUS MARINE FISH, *NASO TONGANUS* AS ASSESSED BY 16S rRNA GENE CLONE LIBRARIES AND TERMINAL- RESTRICTION FRAGMENT LENGTH POLYMORPHISM

CHAPTER 4

WRATH OF THE ROTTING ROOT: STUDENT INVESTIGATIONS INTO THE MICROBIOLOGY OF DECAY

CHAPTER 5

SUMMARY AND FUTURE WORK

APPENDIX A

CHAPTER 2 SUPPLEMENTAL INFORMATION

APPENDIX B

CORNELL SCIENCE INQUIRY PARTNERSHIP ACTIVITY

APPENDIX C

16S rRNA GENE ANALYSIS OF *BALANTIDIUM JOCULARUM*

APPENDIX D

COMMONLY USED STATISTICAL COMPUTATIONS FOR T-RFLP AND 16S rRNA GENE CLONE LIBRARIES

LIST OF FIGURES

APPENDIX C

LIST OF TABLES

APPENDIX C

CHAPTER 1

POLYPLOIDY IN BACTERIA: A LITERATURE REVIEW

Introduction

Microbial Ecology is the study of microorganisms and their relationship with each other and the surrounding environment. Over the past several decades, there has been an increase in interest in the microbial ecology of gut ecosystems. While the adult human is predicted to be comprised of 10^{13} eukaryotic cells it is estimated that the entire individual is colonized by 10^{14} prokaryotic cells resulting in a combined microbial genome in excess of the human genome (94, 118, 162). The impact of these indigenous microbial communities is most likely greatest in the intestine, which harbors the vast majority of these organisms with densities as high as 10^{11} - 10^{12} cells per gram of colon contents (118).

These mammalian gut microbes carry out a wide variety of biochemical reactions and as such can be viewed as a "metabolically active organ" (69, 118). Perhaps their most important functional role is in nutrition, as they provide their host with enzymes to break down otherwise nondigestible dietary substrates and synthesize vitamins (92, 118, 136). Harboring a resident population of microbes able to breakdown a variety of different compounds confers several advantages on the host. First, the host is relieved of having to evolve the functions to carry out these processes. Secondly, this provides the host with a certain amount of dietary flexibility, which allows for adaptation to changing nutrient availability (69, 163).

This symbiotic relationship between the host and its resident gut microbiota can be characterized as commensal (one partner benefits and the other is neither helped nor harmed) or mutualistic (where the relationship is beneficial to both partners) (118). There are many examples of these relationships in nature involving the exploitation of normally inaccessible nutritive substrates, including nitrogen fixation by *Rhizobium* spp. in the nodules of legumes (104) and cellulose degradation by a variety of cellulolytic bacteria in ruminants (34, 116). Ultimately, gut microbiota aid the host in obtaining maximum nutritional value from their diet as evidenced by the fact that mice with well developed gut microbial consortia require 30% less caloric intake than germ-free mice to maintain the same body weight (158, 159).

The true diversity of these gut ecosystems is not fully understood, as most organisms cannot be cultured in the lab. However, advances in molecular biology have allowed us to expand our understanding of these complex systems by employing culture independent techniques including 16S rRNA gene clone libraries (3, 119), terminal-restriction fragment length polymorphisms (90) and real-time PCR (91). Through the application of these approaches, the microbial diversity in the human intestine has begun to be characterized. Hundreds of species of bacteria have been identified, however only a few phyla dominate (55, 136). Considerable diversity has been reported in other animals including horses (34), pigs (85) cows (44) and other ruminants (116). This diversity was quantified by identifying operational taxonomic units or OTUs. An OTU is a group of 16S rRNA gene sequences, which show sequence similarity. For bacteria, 16S rRNA gene sequences that show $\geq$ 97% similarity are often considered the same species (145) and are therefore treated as a

single OTU in phylogenetic analyses. These studies found the majority of OTUs recovered belonged to the phyla Firmicutes and Bacteroidetes: phyla previously shown to dominate the human gut (86 - 89). Perhaps the most comprehensive work done to date has been an analysis of the fecal microbiota of 106 individual mammals representing 60 species from 13 taxonomic orders (87, 88). A total of 19,548 classified 16S rRNA gene sequences were analyzed, and found to belong to 17 different phyla of bacteria. The majority of these sequences belonged to the Firmicutes (65.7%), while 16.3% of the sequences affiliated with the Bacteroidetes supporting the idea that these organisms dominate the mammalian gut microbiome. While Firmicutes were identified in all of the mammals sampled, each host harbored OTUs not identified in any other sample showing great diversity at the bacterial species level. Generally, the community was more similar between hosts of the same species (conspecific hosts) than between different species of the same order. This held true for conspecific hosts living separately from each other (for instance in the wild versus a zoo), though for some this trend was not observed indicating that other factors, such as diet and environment can influence the gut microbial community (87). Perhaps the most significant factor influencing the overall microbial community is diet: herbivore, omnivore or carnivore. Both tree-based and network-based analyses showed significant clustering by diet. Principal Component Analysis supported and validated these trends in a bias free manner, as hosts did not need to be assigned to one of the three diets (88). In addition to diet influencing the overall gut microbiome, it appears to play a role in the diversity of the gut microbial community. Herbivores had both the

largest number of phyla as well as the highest genus-level richness, followed by omnivores and carnivores (87, 88).

In humans, several factors appear to contribute to the composition of the intestinal microbiome. Initially sterile, colonization of the infant gut begins immediately after birth. This is influenced by several diverse factors including medication, mode of delivery, infant diet (breast milk versus formula) and hygiene levels (57, 69, 112). In adults, the intestinal flora is modulated by diet, lifestyle and age (70, 86, 89) as well as host genotype (163). Many lines of evidence, including culture based studies, 16S rRNA gene libraries, denaturing gradient gel electrophoresis and quantitative PCR support the idea that diet influences the microbial community in ruminants (149). When cattle were switched from a forage-based to a high-grain based diet, there was a shift in the overall community structure, but also a shift in the physical parameters found in the rumen. This included a change in pH and overall Short Chain Fatty Acid (SCFA) composition.

Yet another factor that modulates gut bacteria composition is photoperiod, or the light-dark cycle an animal is exposed to (102). In sheep fed identical diets, but housed under different day-length conditions, differences in bacterial diversity as well as total SCFA concentration were noted between the two populations. The researchers concluded that this difference was due to an increase in food consumption by the population exposed to longer day-length. Photoperiod also modulated the gut bacterial composition in male Siberian hamsters. Specifically, those exposed to short day

lengths experienced a reduction in bacteria belonging to the phylum *Proteobacteria* (11).

While much work has been done looking at the microbial community in the gastrointestinal tract of mammals, there is relatively little information on these communities in fish (30). There is, however, growing evidence supporting the role of intestinal symbionts in the digestion of algae and seagrass consumed by marine fishes. Direct microscopic examination of gut contents from several herbivorous fishes has revealed complex and abundant microbial consortia (28, 29, 48, 112). SCFA analysis demonstrated microbial fermentation products in the hindgut of fishes at levels comparable to those found in the intestines of reptiles and small mammals (27, 111). In the gastrointestinal tract of one group of fishes, the surgeonfish (Family *Acanthuridae),* an unusually large bacterial symbiont, *Epulopiscium,* has been identified (29). It is believed that like many gut associated bacteria, *Epulopiscium* plays a functional role in the nutrition of its surgeonfish host.

Initially classified as a protist due to its large cell size, 16S rRNA gene analysis revealed that *Epulopiscium* belonged to the phylum Firmicutes, with some of its closest relatives belonging to the genus *Clostridium* (9). In the larger *Epulopiscium* spp, binary fission has not been observed. Instead, reproduction is achieved through the production of internal vegetative offspring (8). In addition to their unusually large cell size, and unusual mode of reproduction, *Epulopiscium* cells contain an abundant amount of DNA. This DNA is tightly associated with the cell membrane (51, 128) and appears to be arranged in a mesh-like network surrounding the central cytoplasm of

the cell (128). The abundant DNA in these cells along with the observation that only a small amount of the mother cell DNA is partitioned into the newly forming offspring (8) suggests that *Epulopiscium* are highly polyploid. The remainder of this review will consider polyploidy in several different bacteria.

Polyploidy

Polyploidy, possessing multiple copies of the normal set of chromosomes, is common in eukaryotic organisms (84). There are both advantages and disadvantages associated with being polyploid. Advantages that have been identified include heterosis, a condition where the hybrid offspring, or heterozygous population exhibits greater growth, survival and fertility (26), the ability to reproduce asexually through loss of self -incompatibility and lastly gene redundancy. Gene redundancy confers advantages in two ways: it allows the organism to survive deleterious mutations and it allows for gene diversification leading to the acquisition of new functions and potential niche expansion (1, 31). Disadvantages of being polyploid include epigenetic instability and an increase in aberrant meiotic or mitotic events leading to aneuploidy, or an abnormal number of chromosomes (1). Epigenetics refers to a change in the expression of a gene due to something other than a change in the sequence of the DNA. Epigenetic events can include covalent modifications of the DNA or modifications of histones, the chromatin binding proteins (31). When an organism becomes polyploid, it has been show to undergo epigenetic remodeling (31). This remodeling can result in chromosomal rearrangements and gene expression changes of the allopolyploids. Epigenetic remodeling may also be caused by aneuploidy, which

can lead to a dosage imbalance of chromatin regulatory factors or expose unpaired chromatin to epigenetic remodeling mechanisms. While some of these changes may be advantageous, others could also be disruptive and result in instability of the neopolyploids.

Polyploids are common among plants, amphibians, and insects and while rare in mammals, a polyploid rat, the red viscacha rat was identified in Argentina in 1999 (53). In plants, genetic amplification has been an important driving force in the evolution of both ancient and nascent lineages (1, 125, 146). During polyploidization events, duplicated genes often diverge in function leading to novel phenotypes. It has been proposed that this enhances the evolutionary success of the polyploid species by allowing them to adapt to and colonize new ecological niches (65, 146). In insects, polyploidy changes are not usually associated with population divergence. However, one group, the weevils (Coleoptera: Curculionidae) are particularly proficient at utilizing this mechanism for evolutionary change (93). More frequently, insects use polyploidization for resource provisioning during early development. This is well studied in *Drosophila melanogaster* follicle and nurse cells (84) and in the secretory cells of the salivary glands (6). These cells initiate multiple rounds of DNA replication without subsequent division. This process is referred to as endoreduplication. The result is the giant polytene chromosomes that allow the increased metabolic needs of these cells to be met (6). In amphibians, rapid diversification of species is attributed to genome duplication events particularly in the anurans (12). Finally, endoreduplication events occur in eukaryotic hepatocytes. It is hypothesized that this example of

protective endoreduplication occurs in response to stress, and allows for the regeneration of liver function (84).

Polyploidy is also observed in several organelles of eukaryotic cells. Mitochondria and chloroplasts are the cytoplasmic energy-generating organelles of eukaryotic cells that descended from free-living bacterial ancestors (43). Phylogenetic and biochemical analyses suggest that mitochondria arose from an α-proteobacterium-like ancestor which was engulfed by a pre-eukaryotic cell, while chloroplasts arose when a cyanobacterium-like cell was engulfed by a mitochondrion containing cell (43). Both of these organelles have retained a reduced genome that is found in multiple copies and associated with proteins in structures referred to as nucleoids (133). Mitochondrial DNA (mtDNA) contains on average 40-50 genes with protein encoding genes involved in at least one of the following processes: oxidative phosphorylation, ATP synthesis, translation, transcription, RNA maturation and protein import (19). In addition, genes for tRNA and for rRNA may also be present. The DNA found in chloroplasts (cpDNA) is comprised of between 50 and 200 genes encoding for a limited number of tRNAs, rRNAs as well as a subset of the proteins involved in photosynthesis (126).

Early investigations into the polyploidy of chloroplasts utilized electron microscopy of serial sections of chloroplasts to examine the distribution of DNA in these organelles (67). These studies revealed that the DNA was distributed into spatially distinct regions of the chloroplast and that each region appeared to contain between 4 and 8 individual DNA molecules referred to as a nucleoid. The

development of DNA specific dyes such as 4' 6-diamidino-2-phenylindole (DAPI) allowed for the visualization of the DNA found in the chloroplasts of tobacco (*Nicotiana tabacum*), spinach (*Spinacea oleracea*) and sugar beet (*Beta vulgaris alba*) (75). These analyses revealed distinct fluorescent foci throughout the chloroplast with intensities equal to or greater than the amount of DNA in a single plastome (plastid genome) inferring that there were multiple cpDNA molecules in a single organelle. In the marine alga *Olisthodiscus luteus* DNA was extracted and quantified and chloroplasts per cell determined to calculate the DNA content per chloroplast (22). Contour length, restriction endonuclease digestions and reassociation analysis was then utilized to determine the size of the *O. luteus* chloroplast genome (47). Using these two pieces of information, the number of cpDNA molecules per organelle was shown to be as low as 13 and as high as 44. More recent analysis of cpDNA molecules in the model organism *Arabidopsis thaliana* (Columbia) combined flow cytometry and real-time quantitative PCR of three unlinked gene markers to determine the number of genomes per chloroplast in leaves of 43-day-old plants (129). The results from this analysis indicated that genome copy number per chloroplast varied from about 10 to approximately 35. However, as recently as 2010 chloroplast polyploidy during different stages of leaf development in the sugar beet (*B. vulgaris* L. var. Kleinwanzlebener Marta) was determined using colorimetric cpDNA quantification and DAPI staining of highly purified chloroplasts (126). In these studies the chloroplast genome copy number ranged from a dozen in the chloroplasts of post-primordial leaflet cells to greater than 100 copies in chloroplasts of mature leaf tissue.

The first studies assessing DNA content in individual mitochondria utilized electron microscopy of ultrathin sections of mitochondria as well as the relationship between quantified isolated mitochondrial DNA and protein (115). The microscopic surveys revealed that mitochondria from chick embryo muscle cells contained up to six-DNA containing regions referred to as nucleoids and mitochondria from L cells (mouse fibroblast cells) contained on average between two and four nucleoids. Studies assessing DNA and protein content in mitochondria indicated a ratio of 0.2- 1.8 mg of mtDNA per mg of protein (17, 114). This ratio is equivalent to 2 to 10 mtDNA molecules per mitochondria and varied by cell type examined. Using DAPI the nucleoids in yeast mitochondria appeared as punctate fluorescent structures indicating multiple mtDNA molecules per organelle (106, 156). DAPI staining combined with video-intensified photon counting microscopy (VIM) was utilized to assess the organization of mitochondrial nucleoids in human cells (135). These results demonstrated that individual mitochondria possessed between 1 and 15 copies of mtDNA arranged in one to > 10 nucleoids. Using several mtDNA labeling (immunogold and Anti-DNA antibodies) and microscopy (electron and fluorescence) approaches it was shown that in mammalian mitochondria these nucleoids consist of several mtDNA genomes complexed with proteins (73, 154). Additional studies have employed real time quantitative PCR using extracted DNA (99) or a combined approach of flow cytometry to sort individual mitochondria and real-time PCR to quantify the mtDNA copy number (23). These studies obtained similar results to those obtained using DAPI staining and VIM.

Many studies have been done attempting to quantify the multiple genomes of the energy producing organelles, chloroplasts and mitochondria. However, the reason why these organelles possess many copies of their genome is still not well understood. The multiplicity of the chloroplast genome may accommodate an increase in protein synthesis capacity during active growth by providing an elevated gene dosage of ribosomal RNA genes (13, 129). This idea is supported by the fact that meristematic cells initiate multiple rounds of cpDNA replication with out subsequent cell division leading to an increase in the amount of DNA found in an individual chloroplast (129, 130). As the leaves mature and the need for rapid protein synthesis decreases this content declines due to a dilution effect caused by cell expansion.

In a variety of organisms, mitochondria may possess multiple copies of their genome to satisfy the need for an increased number of mitochondrial ribosomes to synthesize the proteins needed to carry out respiration (13). Evidence in support of this includes the observation that mtDNA copy number in rat liver cells increases when they switch their metabolic program from glycolysis to aerobic respiration after birth (20). This switch would result in an increased demand for the energy producing proteins in the mitochondria.

While there are many examples of polyploidy in eukaryotic cells, historically, prokaryotes have been described as having a single, circular chromosome. Most of the early work on the conformation of the prokaryotic genome was with the model organism, *Escherichia coli* (21). However, as early as the 1960's researchers acknowledged that *E. coli* cells could possess more than one chromosome, and this

condition is intimately related to the bacterial cell cycle (66). In *E. coli,* cell cycle events are initiated immediately after the daughter cells separate during the D phase (66). This first stage is referred to as the B phase, and is marked by an increase in cell size. Once the cells reach a critical mass, they enter the C phase where DNA replication commences. Once replication of the chromosome is complete, the cell enters the D phase, which includes all events through division of the daughter cells. Much like eukaryotic cells, prokaryotic cells will proceed through each of these stages sequentially as long as two conditions are met: all events of the previous phase and all check point events have been completed (39). One difference between the eukaryotic and bacterial cell cycle is that unlike in eukaryotic cells, phases of the bacterial cell cycle, specifically the B and C phases, can overlap (66). This is a necessity given that it takes a minimum of 40 minutes for complete replication of the *E. coli* chromosome, but rapidly growing cells are able to divide in 20 minutes. When the cells are dividing faster than the time necessary for completion of chromosome replication, the daughter cell will inherit a chromosome that contains a replication fork (21, 66). As this chromosome completes replication another round of replication is initiated. This results in cells that have three to four genome equivalents.

More recent work has demonstrated that *E. coli* cells in stationary phase can have as many as 8 copies of their chromosome (2). Using restriction fragment length polymorphism (RFLP) researchers have demonstrated an increase in the amount of genetic diversity present in long-term stationary phase cultures of *E. coli* (50). This increase is due to a higher mutation frequency during stationary phase. One mechanism that has been implicated in this increase of mutation frequency is the

methyl-directed mismatch repair (MMR) system (50, 62, 63). The MMR system, a highly conserved system in both eukaryotes and prokaryotes, recognizes and removes mismatches in newly synthesized double stranded DNA (137). This particular repair system is not specific, repairing any damage that causes a distortion in the DNA helix. This distortion can be caused by mismatches, frame shifts or incorporation of base analogs. In *E. coli*, the enzymes involved in the system recognize the newly synthesized strand based on its methylation state, with newly synthesized DNA of the hemimethylated duplex DNA being unmethylated (62, 63, 105). There are three proteins involved in the MMR system in *E. coli*: MutH, MutL and MutS. The MutS protein binds to the distorted region of the DNA. This allows the MutH and MutL proteins to bind forming a complex that translocates the distorted DNA through the complex until it reaches a hemimethylated GATC sequence. MutH nicks the DNA and then an exonuclease degrades the DNA back to the mismatch. DNA polymerase III, the major replicative polymerase in prokaryotic cells, is then able to synthesize new DNA incorporating the correct bases thus repairing the lesion. During stationary phase, the DNA is fully methylated, so when mismatches occur due to DNA damage the MMR system will repair the incorrect strand half of the time resulting in a mutation (15). Another model suggests that the increase in mutation rate seen during stationary phase is due to the overall suppression of biosynthetic pathways, including those that synthesize the proteins needed for the MMR system (62, 63). The polyploidy observed in *E. coli* cells in stationary phase is transient and it is believed to serve as a survival mechanism for the cell (49). If a mutation occurs in an essential gene, there are additional copies to overcome any lethal effects. While polyploidy is

less common in prokaryotic organisms, there are several characterized exceptions to the rule that bacteria are monoploid. What follows is a short review of the studies of polyploidy in specific bacteria, including *Buchnera aphidicola*, *Azotobacter vinelandii*, *Deinococcus radiodurans*, *Neisseria gonorrhoeae*, *Rhizobium* spp., *Streptomycetes coelicolor*, *Thiomargarita namibiensis*, *Metabacterium polyspora* and *Epulopiscium* type B, with particular emphasis on how polyploidy contributes to the biology of each of these organisms.

Buchnera aphidicola

Buchnera aphidicola are symbionts of aphids (Homoptera, Aphididae). Paul Bauman and colleagues used 16S rRNA gene sequence analysis to identify this organism as allied to the Enterobacteriaceae belonging to the class Gammaproteobacteria of the phylum *Proteobacteria* (113). It was assigned to the novel genus *Buchnera* in recognition of the tremendous efforts of Paul Buchner in advancing the understanding of the microbial-insect relationship (42, 113). *B. aphidicola* have a genome size only one seventh of that of *E. coli*, with sizes ranging from 450 to 657 kb in different strains of *B. aphidicola* from various aphid hosts (74, 80). This type of genome reduction is not unusual in endosymbiotic bacteria and is believed to be an adaptation to intracellular life (38, 54, 120).

Despite having an extremely reduced genome, one class of genes that is not drastically reduced in *B. aphidicola* is those involved in the biosynthesis of essential amino acids. This is remarkably different from the gene collection seen in parasitic bacteria, which obtain nutrients from their hosts (74). Recent genomic and

physiological research has demonstrated that *Buchnera* is responsible for synthesizing essential amino acids for its host (18, 60, 144, 158) thus enabling the insect to utilize a nutritionally inadequate diet of phloem sap (42, 107). *B. aphidicola* have lost all of the genes for biosynthesis of *non*-essential amino acids, suggesting that *B. aphidicola* depends on its aphid host for these nutrients (42, 58). This complementarity between host and symbiont has lead to mutualism between *B. aphidicola* and its aphid host.

Within their aphid host, *Buchnera* are confined to specialized somatic cells called bacteriocytes. It is hypothesized that the initial association between bacteria and their insect hosts was one of parasitism (64). Evolutionarily bacteriocytes could have been selected to help the bacteria evade the host immune system, and to protect the host from bacterial pathogenicity during the transition from parasite to mutualist (64). There have been several studies looking at the molecular evolution of *Buchnera* and their aphid host (42, 64, 108) which suggest that this symbiotic relationship was established 160 -250 million years ago. Repeated transmission of the symbiont to its aphid host resulted in parallel divergence and cospeciation, culminating in the symbiosis known today.

The reduced genome of *B. aphidicola* lacks several classes of genes that would seem to be essential for it to maintain its identity as a cell. These include genes for ABC transport systems, two-component regulatory systems and transcriptional regulators (74). Other suites of genes missing from the *B. aphidicola* genome include those required to produce the enzymes needed for the citric acid cycle, despite *Buchnera* being an aerobic organism (54). It was initially hypothesized that *Buchnera*

compensates for this void in its genetic repertoire by importing enzymes from its aphid host in a manner similar to that of mitochondria and plastids (74, 78). If this mechanism is utilized it would suggest that *Buchnera* transferred the genes for these enzymes to their host (43, 74, 78). There are several lines of evidence that refute this hypothesis.

Analysis of the genome of the pea aphid, *Acyrthosiphon pisum* revealed 12 potential lateral gene transfer (LGT) candidates that appeared to be genes from bacteria (117). To verify that these genes were part of the *A. pisum* genome, and not recovered from contaminants (bacterial symbionts or pathogens within the host), quantitative PCR was used to compare the gene copy number of each LGT candidate with single copy aphid genes. Results from these assays strongly supported the idea that these genes are part of the aphid nuclear genome and were acquired via LGT. To further understand the origin of these genes, detailed structural and molecular phylogenetic analyses were carried out (117). The authors concluded that the *A. pisum* genome harbors eight transcribed genes of bacterial origin, including genes coding for carboxypeptidases, an amidase, lysozymes and rare lipoproteins, the first three of which are involved in peptidoglycan synthesis. None of these appear to be transferred from *Buchnera*, instead these genes appear to be from Alphaproteobacteria, specifically *Wolbachia* spp. and *Orientia tsutsugamushi*, both intracellular symbionts of arthropods and the cause of sporadic infections in aphids. Collectively these results suggest that while LGT to the aphid host genome did not drive genome reduction in *Buchnera*, and the lack of LGT from *Buchnera* demonstrates a clear difference between *Buchnera* and organelles (117). Proteome analysis provided further support

refuting the hypotheses that *Buchnera* acts like an organelle (122). When the proteome of whole aphid cells was compared to the proteome of the bacteriocyte as well as *Buchnera* cells there was no evidence for the selective transfer of proteins between either partners. This suggests that *Buchnera* has a greater genetic autonomy than organelles.

Buchnera, despite possessing a reduced genome, have a cellular volume approximately 15 times that of *E. coli* (80, 81). This, along with the observation that *Buchnera* cells divide less frequently than *E. coli* led researchers to hypothesize that this aphid symbiont is polyploid. Quantification of DNA content in *Buchnera* sp. APS, *E. coli* and *Saccharomyces cerevisiae*, as well as the copy number of two unlinked genes (*groE* and the 16S rRNA gene) in *Buchnera* demonstrated that *Buchnera* contained, on average, 120 copies of its 657 kb genome (80). Aphid colonies are made up of two different morphs, which differ from each other both morphologically and physiologically (81). Apterae (wingless morphs) usually dominate the colony and possess *Buchnera* with fewer genome copies. Alatae (winged morphs), which are fewer in number possess *Buchnera* with 100-400 genome copies.

Aphids also have a well-defined life cycle (107) and it has been shown that polyploidy in *Buchnera* varies with the developmental stage of the aphid host (80, 81). Bacteriocytes of the aphid increase in size during larval development (80, 81) corresponding to an increase in the number and size of *Buchnera* cells. It is believed that the small size of the bacteriocyte prevents the *Buchnera* from dividing, but *Buchnera* continue to replicate their DNA. Only when the bacteriocyte increases in

size can the *Buchnera* cell then divide. It appears this obligate symbiosis and spatial restriction results in repeated divisions of the genome without cell division resulting in the extreme polyploidy observed in *Buchnera* (80).

Insects other than aphids engage in obligate symbioses with polyploid nutrient-provisioning bacteria (160). Sharpshooters (Family: Cicadellidae) are xylem-feeders that possess two bacterial symbionts: *Baumannia cicadellinicola* belongs to the phylum *Proteobacteria* and *Sulcia muelleri* belongs to the phylum *Bacteroidetes* (161). *Sulcia* has been show to possess hundreds of copies of their reduced (150 kb) genome (160). Xylem, which is used in plants primarily to transport water and salts, is low in nitrogen, carbon and organic compounds (5). Analysis of the *Sulcia* genome has revealed these insect symbionts also synthesize essential amino acids for their insect host (101, 161). This would suggest that polyploidy in the obligate symbiotic bacteria of some insects is a common feature that ensures synthesis of an adequate amount of essential amino acids and other nutrients for their hosts.

Azotobacter vinelandii

This aerobic, soil-dwelling organism belongs to the class Gammaproteobacteria of the phylum *Proteobacteria*. *A. vinelandii* possess a wide variety of metabolic capabilities (76, 77). Perhaps the most notable is its ability to fix nitrogen in the free-living state. This is in contrast to more typical soil dwelling, nitrogen fixing bacteria that assume symbiotic relationships with the root nodules of leguminous plants (121, 123). *A. vinelandii* is of particular interest to scientists studying nitrogen fixation because in addition to the molybdenum-containing

nitrogenase enzyme, *A. vinelandii* is capable of synthesizing two alternative nitrogenases (52). In one, the molybdenum is replaced by vanadium and in the other there is no transition metal, but only iron. The synthesis of these alternative nitrogenases is regulated by the availability of the metals, and controlled by the regulatory proteins NifA, VnfA and AnfA (152).

Nitrogen fixation in *A. vinelandii* is able to occur under aerobic conditions. Nitrogenases are inherently sensitive to oxygen, however *A. vinelandii* has evolved several mechanisms for nitrogenase protection (110). An incredibly high respiratory rate allows this organisms to consume oxygen much faster than it can enter into the cytoplasm of the cell ensuring an essentially anoxic environment for the nitrogenase, despite *A. vinelandii* obtaining its energy from a highly aerobic process (143). *A. vinelandii* also synthesizes a protective Iron Sulfur protein (FeSII), which can bind to the nitrogenase in times of oxygen stress thus stabilizing and protecting the protein (110, 143, 147).

Early work investigating the organization of the *A. vinelandii* genome utilized thermal denaturation and DNA renaturation kinetic experiments (121, 132). The results of these experiments revealed *A. vinelandii* and *E. coli* to have similar chromosome sizes, however, *A. vinelandii* harvested during mid-exponential phase contained at least 40-times more DNA (121, 132). This suggested that *A. vinelandii* possessed approximately 40 chromosomes per cell. Copy number of the *leuB, nifH, nifD* and *nifK* genes in each cell was assayed using a quantitative hybridization procedure, and shown to be present in approximately 80 copies per cell (124). These

results suggested that *A. vinelandii* had 80 genomic copies per cell. In order to assess this further, gene dosage experiments were done, but these results seemed to contradict the aforementioned work that suggested that *A. vinelandii* is polyploid (97).

In order to reconcile these disparate results, researchers used flow cytometry to assess DNA content in cells grown in rich media and minimal media during several stages of the bacterial cell cycle (98). The results from this work demonstrated that *A. vinelandii* only achieves significant polyploidy during late exponential phase (~ 40 copies of their chromosome) and stationary phase (> 100 copies of their chromosome) when grown in rich media. Polyploid cells also assume a larger cell size, which is believed to accommodate an increase in DNA content (97). Increase in cell volume was confirmed independently using scanning electron microscopy to measure cell volumes (45). The results of this study demonstrated that the volume of *A. vinelandii* cells was approximately 16 times that of *E. coli* cells. During growth in minimal media, however, *A. vinelandii* has DNA content similar to other enterobacterial species such as *E. coli* and *Salmonella typhimurium* (97).

More recently a chromosome counting technique was utilized to attempt to resolve the ploidy status of *A. vinelandii* (123). This method involved constructing an *A. vinelandii* strain with a selectable genetic marker on its chromosome. This strain was then transformed with a plasmid of known copy number, carrying the same selectable marker. Total genomic DNA was extracted from these cells, digested with the appropriate restriction endonuclease and the DNA fragments separated on an agarose gel. The selectable markers were then located through Southern blotting and

hybridization, and the intensity of the bands determined through densitometric scanning. The results from these studies indicated that *A. vinelandii* does not achieve significant polyploidy. Clearly additional studies need to be carried out in order to verify the genome copy number in *A. vinelandii*, however taken together these data would suggest that polyploidy in this organism represents a phenomenon that occurs only in very specific lab conditions and not in its natural environment, the soil. As such, the biological significance is not known (98). One could hypothesize, however, that if this type of polyploidy was achieved in nature, it could confer some protective adaptation to the organism, by providing the genetic resources to produce enough of the protective FeSII protein allowing *A. vinelandii* to carry out nitrogen fixation in a strictly aerobic environment.

Deinococcus radiodurans

Perhaps best known for its extreme radiation resistance, *D. radiodurans* is another example of a polyploid bacterium. The first *D. radiodurans* strain was isolated in 1956 from a can of ground beef that had been gamma irradiated at 4000 Gy (4): a dose 250 times the amount required to kill *E. coli*. In 1958, R. G. E. Murray and C. F. Robinow described another strain of *D. radiodurans* isolated as an air contaminant from a hospital in Ontario (96). Since then, eleven additional radiation resistant species that are closely related to *D. radiodurans* have been identified from a variety of different environments, including soil, animal feces and tissue and hot springs. To date, however, researchers claim that no systematic exploration of the natural distribution of these unusual organisms has been conducted (16, 35). Originally

classified in the genus *Micrococcus*, 16S rRNA gene analysis has since placed *D. radiodurans* in its own unique group, the Thermococcus-Deinococcus Phylum (36, 96). Perhaps the most unique feature of *D. radiodurans* is its ability to survive massive DNA damage by a wide range of physical and chemical agents, including desiccation, oxidizing agents, UV irradiation and ionizing radiation (16). One factor that contributes to this ability to overcome double strand DNA breaks is the polyploidy observed in *D. radiodurans* (32).

Ionizing radiation is any electromagnetic or particulate radiation, which is strong enough to remove electrons from atoms. When this occurs, highly reactive radicals may form, which are able to modify biologically important molecules in the cell (16, 32). Of all of the effects, DNA damage may have the greatest impact on cell viability (32). This is due to the fact that DNA occupies a large percentage of the overall cell volume, making it more likely to be hit by these damaging molecules. In addition, the genome in most bacteria is in low copy number, so any loss of functionality can be devastating to the organism. Evolutionarily, radiation resistance in bacteria is difficult to explain: there simply are no known naturally occurring environments where organisms would be exposed to 400 Gy a year, thus selecting for this trait (96).

The damage induced by gamma -irradiation (double-strand breaks in the DNA) is also induced during desiccation (40, 41). One study reported a minimum of 60 double-strand DNA breaks during prolonged desiccation in the chromosomes of both *Bacillus subtilis* spores and *D. radiodurans* vegetative cells (40, 100). This is similar

in number to those sustained by *D. radiodurans* during exposure to radiation. Since desiccation is far more likely to occur in natural environments, the cellular mechanisms to deal with this stress could be selected for. This could mean that the ability to survive high doses of gamma-irradiation is simply a consequence of the organisms' ability to survive desiccation (32, 100).

D. radiodurans has been shown to have multiple copies of its genome (61). By determining the complexity of the DNA and the amount of DNA present in an individual cell of a diplococcal pair, it was determined that the ploidy of *D. radiodurans* did not go below four genome equivalents per cell during stationary phase and as many as ten genome equivalents were present during exponential growth. Subsequent sequence analyses of *D. radiodurans* showed that the genome of this organism is comprised of two chromosomes, a megaplasmid and a smaller plasmid (96). In addition to the usual DNA repair mechanisms, *D. radiodurans* also has a novel suite of genes that are upregulated in response to either desiccation or ionizing radiation, however further work will have to be done to elucidate the function of these genes in ionizing radiation resistance (16).

While it is likely that two or more genomic copies are necessary for radioresistance, they are not sufficient to confer this trait on *D. radiodurans* (32). As such, polyploidy in *D. radiodurans* is considered a passive contribution to its radioresistance. This extra genetic material is believed to assist the cell in several ways. First, the redundant genetic material can be used in the process of DNA repair. It has been demonstrated that *E. coli* is only able to repair double strand DNA breaks

during exponential growth in rich media when it possess multiple chromosomal copies per cell (35, 36). Secondly, it is unlikely that all copies of a gene will be inactivated by irradiation, which ensures at least one functioning copy of the gene (16, 32). While polyploidy does not ensure radioresistance, it is required for the repair of double strand DNA breaks (32) so is no doubt important in the survival of *D. radiodurans*.

Polyploidy in Other Bacteria

The above mentioned bacteria are unique in that much research has gone into the elucidation of the level of polyploidy as well as the functional significance of possessing multiple genomes. However, several other bacteria are worth mentioning with regards to polyploidy.

Neisseria gonorrhoeae is a Gram-negative bacteria belonging to the class β – proteobacteria (150). This intracellular pathogen is the causative agent of the sexually transmitted disease gonorrhea. Gonorrhea has persisted in the human population since 5 B.C (109) with approximately 300,000 new cases reported in the US annually (164). One of the reasons this obligate human pathogen has persisted is its ability to evade the immune system by varying surface associated virulence structures (antigenic variation); in this case the type IV pili (33). Pili are structures that are used for attachment to epithelial cells contributing to the establishment of infection. The ability of these organisms to exhibit antigenic variation has also hampered efforts to develop a vaccination for gonorrhea (164). Pilin antigenic variation in *N. gonorrhoeae* is mediated by the homologous recombination of *pilE* (expressed) and *pilS* (silent) genes (59). Expression of the gene variant results in a

functional pilin protein containing formerly “silent” antigenic determinants. This model of pilin antigenic variation requires two copies of the chromosome for the gene conversion event to take place suggesting that *N. gonorrhoeae* cells are polyploid (79). Microarray and qPCR assays demonstrated that replication is only initiated once per round of cell division supporting the idea of polyploidy in *N. gonorrhoeae* (150). Flow cytometry and fluorescent microscopy confirmed these results showing that the DNA content in an individual *N. gonorrhoeae* is greater than that found in a single genome copy (150). Finally, thin section electron microscopy revealed multiple nucleoid regions, with an average of three regions per cell (151) supporting the idea that *N. gonorrhoeae* is yet another example of a polyploid bacterium.

A recent study assessed the ploidy levels in two other species of *Neisseria*, *Neisseria meningitidis* and *Neisseria lactamica* (151). *N. meningitidis* is a pathogenic *Neisseria* species and the most common cause of bacterial meningitis in adolescents and young adults. *N. lactamica* is a commensal of the human nasopharynx (95). Utilizing flow cytometry it was demonstrated that exponentially growing *N. meningitidis* cells maintain three to five genome equivalents, while *N. lactamica* possessed between one and four genome equivalents. When cells were treated with the antibiotic tetracycline (which allows for the completion of replication, but inhibits the initiation of new round of replication) the majority of *N. meningitidis* cells contained 4 genome equivalents (4N), with a smaller proportion of cells containing 8 genome equivalents (8N). This is in contrast to *N. lactamica*, where 1N, 2N and 4N populations of cells were observed. The 1N population represents monococcal (single) cells that have not initiated replication, while the 2N and 4N populations represent

monococcal and diplococcal (paired) cells, respectively, that had initiated and completed replication resulting in two genome copies per monococcal unit. These results suggest that *N. lactamica* is not polyploid. In the case of *Neisseria* spp. polyploidy appears to be directly related to the biology of this organism as a human pathogen. Both *N. gonorrhoeae* and *N. meningitidis* undergo pilin antigenic variation while the commensal *N. lactamica* does not (151). Having multiple chromosomal copies facilitates these high frequency gene conversion events (150, 151).

Rhizobium bacteria associate in a symbiotic relationship with legumes resulting in the formation of root nodules (104). These plant root organs house the bacteroids: intracellular, membrane encapsulated rhizobia capable of fixing nitrogen. This relationship is considered mutualistic with both partners benefiting. The plant supplies a microaerobic environment for the oxygen sensitive nitrogenase enzyme of *Rhizobium* and photosynthates for energy. The plants provide an environment in which nitrogen fixation can occur and they obtain reduced nitrogen, which can subsequently be converted into plant biomass during growth. During nodule formation, growth and differentiation of the infected plant cells results in cell enlargement mediated by repeated endoreduplication events leading to high ploidy levels (104). These events however are not limited to the plant cells, and cytological studies have demonstrated cellular enlargement in the bacterial symbionts. This raised the possibility that cellular differentiation and growth in the bacterial symbiont could also involve genome amplification.

Flow cytometry and comparative genomic hybridization (CGH) demonstrated that the bacteroids in nodules of galegoid legumes (legumes which form indeterminate nodules) undergo a profound transformation including cell elongation and amplification of their tripartite genome (104), which is comprised of a chromosome and two megaplasmids (68). In addition a correlation between increase in cell size and genome copy number was noted (104). This increase in ploidy may result in a decrease in viability of bacteroids, though additional research needs to be done to address this question. Regardless, this increase in genome copy number is believed to confer a better symbiotic advantage on the differentiated nodules by providing the genomic resources for higher rates of nitrogen fixation (104).

The streptomycetes are a diverse group of high G + C Gram-positive bacteria belonging to the phylum Actinobacteria. As early as the 1960's it was recognized that these unusual organisms contained multiple copies of their genome during different stages of development (72). Among the most ubiquitous and numerous of the soil bacteria, streptomycetes are well known for their diverse metabolic capabilities, including the degradation of the abundant biopolymers lignocellulose and chitin, making them vital to carbon recycling (14, 71). In addition to their importance in biotransformations, streptomycetes are preeminent antibiotic producers, synthesizing over two-thirds of the antibiotics currently used worldwide as well as other chemotherapeutic agents, such as immunosuppressants and anti-tumor agents (71). From the time of their discovery in the 1870's through the late 1950's these organisms were believed to be an intermediate between the fungi and bacteria (71). This was due in part to their unusual colony morphology, which is a direct result of their complex

developmental program. *Streptomyces coelicolor* is considered the model streptomycete and the one in which this developmental program as been best studied (56, 131).

Germination and outgrowth of a *S. coelicolor* spore results in a complex vegetative mycelium that is comprised of hyphae that grow through tip extension. Septum formation leads to vegetative hyphal compartments that contain several uncondensed copies of the *S. coelicolor* linear chromosome (14). Environmental signals, such as nutrient depletion and increased cell density result in a change from vegetative hyphae formation to aerial hyphae formation (24, 71). The aerial hyphae may contain up to 50 copies of uncondensed chromosomes in one long tip compartment (131). Once growth of the aerial hyphae ceases, many FtsZ rings form along the aerial hyphae (56, 142). Chromosomes then condense and align along the hyphal tip compartment ensuring that each spore has a single chromosome (71, 131). For streptomycetes polyploidy is a part of their complex developmental cycle that allows for the production of large numbers of spores, which may then be dispersed in the soil or air (71).

The polyploid bacteria described thus far are unremarkable when it comes to their cell size. There are, however, unusually large bacteria that also exhibit polyploidy. *Thiomargarita namibiensis* is a large sulfur bacteria belonging to the class Gammaproteobacteria, within cluster I of the family *Beggiatoacea* (134). These organisms are closely related to the large vacuolated sulfur bacteria species *Beggiatoa* and *Thioploca* (138, 139). One of the last sulfur bacteria to be discovered to date

(139), these large, spherical cells are present in large numbers in surface sediments off of the coast of Namibia (139, 141). In this anoxic environment they obtain energy through the oxidation of internal sulfur or sulfide, which is produced by the degradation of organic material by sulfate-reducing bacteria (140, 141).

T. namibiensis is characterized by very large cells averaging 100 – 300 μm in diameter, though cells as large as 750 μm in diameter have been described (83). Transmission electron microscopy revealed that approximately 98% of the cell volume is taken up by a liquid vacuole that stores nitrate (140). Nitrate is used as the terminal electron acceptor during the oxidation of hydrogen sulfide to elemental sulfur. The vacuole displaces the cytoplasm to a very thin layer (approximately 0.5 – 2 μm) just beneath the cell envelope. In order to assess DNA distribution in *T. namibiensis*, cells were stained with DAPI and observed with confocal scanning laser microscopy (138). Through these observations it was noted that DNA occurs in small clumps and is not evenly distributed throughout the entire cytoplasm. Quantification of this DNA indicates that individual *T. margarita* cells contain between 6,000 to 17,000 genomes (83). It is hypothesized that this allows the cell to maintain membrane potential homeostasis by ensuring a unit area of membrane has associated with it an appropriate number of genes encoding for the protein products involved in oxidative phosphorylation. Ultimately this allows the cells to produce enough ATP to drive cellular processes (83).

Metabacterium polyspora is an intestinal symbiont of guinea pigs with cells ranging in size from 12 μm to 35 μm (25). In *M. polyspora*, binary fission is restricted

to a short time during its life cycle, if it occurs at all (7, 10). Instead, these cells regularly produce up to nine endospores as a means of propagation (7, 127). The production of endospores by *M. polyspora* is tightly coordinated with its passage through the guinea pig gastrointestinal (GI) tract and helps this organism maintain its symbiotic relationship with its guinea pig host (7, 10). Guinea pigs consume the mature endospores found in their feces. Endospores ensure the survival of this anaerobic bacterium outside of the host as well as survival from the chemical, mechanical and enzymatic assaults of the upper GI tract. Once the endospores pass into the ileum, they germinate. DAPI stained cells of *M. polyspora* showed a pattern of DNA distribution in the cell similar to that seen in the surgeonfish symbiont *Epulopiscium* (128) suggesting that these cells may also be polyploid (10). Interestingly, endospores of *M. polyspora* also appear to maintain some degree of polyploidy (154). This is in stark contrast to the model endospore former *Bacillus subtilis,* which forms a single endospore as a means of survival (46).

DNA dynamics during progressive stages of the *M. polyspora* lifecycle were examined using immunolocalization of the nucleotide analog bromodeoxyuridine or BrdU (154). The results of this work support the following model for DNA replication in *M. polyspora*. Shortly after germination of the endospore, DNA replication is initiated. Some DNA is segregated to the poles of the cells and asymmetric division takes place. After engulfment, the genome(s) within the forespores initiate DNA replication with subsequent segregation and division of the forespore resulting in multiple endospores. It is believed that DNA replication continues in these forespores even after the number of forespores is set resulting in polyploid mature endospores.

This accumulation of genomes in the endospores ensures the progeny cell has an adequate number to immediately initiate sporulation after germination, even if the cell is confronted with a nutrient deplete environment.

Epulopiscium type B is another usually large bacterium that is found as a symbiont of certain types of surgeonfish (Family: *Acanthuridae*). However, unlike *T. namibiensis*, *Epulopiscium* type B does not contain a large vacuole, but instead the cell appears to be filled with metabolically active cytoplasm. Attaining lengths of 250μm, reproduction in *Epulopiscium* type B is carried out through the production of intracellular offspring (7). This offspring development initiates at the poles, and the newly formed daughter cells elongate until they fill the mother cell cytoplasm. Images of DAPI stained *Epulopiscium* type B show abundant DNA arranged around the periphery of the cell suggesting that these cells are highly polyploid (128). The significance of having DNA arranged in this manner is twofold. First, this arrangement may be an evolutionary adaptation that has allowed *Epulopiscium* type B to overcome the constraints of diffusion, thus allowing it to attain its unusually large cell size. Secondly, this cellular modification may support the unusual manner in which *Epulopiscium* type B reproduces. Analysis of DNA replication in *Epulopiscium* type B cells at various stages of development has shown that DNA replication in the mother cell continues even after engulfment of the daughter cells (155). This abundant amount of peripherally arranged DNA would allow the cell to accommodate the enlarging daughter cells, while maintaining the metabolic capacity of the mother cells during daughter cell development (155).

In 2010 Nick Lane and William Martin published a paper entitled, "The Energetics of Genome Complexity" (83). Lane and Martin put forth the hypothesis that the evolution of complex life in the form of the eukaryotic cell was dependent upon acquisition of mitochondria to organize the production of useful energy. In prokaryotes, energy is generated across the plasma membrane while in eukaryotes energy is generated in specialized organelles, the mitochondria. Because of an extensive internal membrane system these organelles provide an expanded surface area for energy-producing reactions to occur. Even more important than the expanded surface area is the fact that these organelles possess their own reduced and highly specialized genome. As part of their analyses, the authors consider these factors in relationship to large polyploid bacteria.

They begin their argument with several calculations assessing the energy budget of the eukaryotic and bacterial cells. DNA replication consumes a mere 2% of the overall energy budget of the cell. In contrast, protein synthesis accounts for approximately 75% of the cell's energy budget. Therefore increasing the genome size of a prokaryotic cell tenfold means ten times as many proteins to synthesize, resulting in an energy crisis for the cell. Alternatively, one can calculate the metabolic power of a single cell. Based on average metabolic rates and cell size, Lane and Martin estimated that a typical proteobacterium is producing approximately 0.49 picowatts (pW) per cell while a protist is producing 2,286 pW. The authors speculate that it is this discrepancy in power that has allowed eukaryotes to expand their genomes while the genome sizes of prokaryotes have remained constrained. These calculations can be even further refined by looking at the metabolic power per megabase (Mb) of DNA.

An average bacterium with a 6Mb genome has about 0.08 pW Mb^{-1} while a in a large amoeba (taking into consideration mtDNA) the genomic power is 0.25 pW Mb^{-1}. These numbers are very similar to each other, however for the eukaryotic cell this genomic power sustains a 3,000 Mb nuclear genome. Ultimately this means that an average bacterium would have 0.03 femtowatt (fW) per gene, while an average eukaryotic organisms would have 57.15 fW per gene: 1,900 times more power. It is this power, Lane and Martin propose, generated by mitochondria with their streamlined genomes that allowed the expansion of the size and complexity of the nuclear genome, the development of novel protein families and ultimately new complex traits.

Beyond insufficient power, Lane and Martin posit that the need for a close association between the bioenergetic membrane and the genes that encode for the production of ATP constrains not only the complexity of prokaryotic organisms but the number and location of their genome. This is due to the need to maintain membrane-potential homeostasis across the entire membrane; an idea supported by the DNA arrangement of the large, polyploid bacteria *Epulopiscium* and *Thiomargarita namibiensis*. In *Epulopiscium* the DNA is arranged around the periphery of the cell closely associated with the cell membrane (128). A similar pattern is observed in *T. namibiensis*, with DNA associating with the plasma membrane around the periphery of the cell (138). It may be tempting to view these arrangements as functionally equivalent to mitochondria, however neither organism has attained eukaryotic complexity.

The mitochondrial genome is reduced to "bioenergetically specialized genomes" (83). In protozoa the mitochondrial genome ranges from 6 to 77 kilobases (kb) with an average genome size of around 30 Kb. These genomes are present in high copy number per cell, as an individual cell possesses many mitochondria and mitochondria have been shown to possess multiple copies of their genome (17, 135). These mitochondrial genomic resources ultimately provide the energy needed to support nuclear genomic expansion and allowed for the increase in complexity observed in eukaryotic cells. It is hypothesized that the lack of these specialized genomes focused on energy production is why large polyploid bacteria, such as *Epulopiscium* do not have the complexity observed in eukaryotic cells. Specifically, because *Epulopiscium* must replicate 200,000 copies of its 3.8 Mb genome which uses energy. Using this genome complexity, Lane and Martin again attempt to support the lack of complexity seen in bacteria with numbers, calculating the metabolic power per megabase at 0.075pW Mb^{-1}: enough to maintain the massive amount of DNA but not enough to allow *Epulopiscium* from attaining true eukaryotic complexity. This is in contrast to a large amoeba (taking into consideration both nuclear DNA and mtDNA) where the genomic power is 0.25 pW Mb^{-1}.

There are several assumptions that the authors make that influence the validity of their calculations. Perhaps the biggest error is using a sphere as a model to extrapolate information about large bacteria. Using this model, they calculated the surface area to volume ratio and used this to determine the energy available to a protist and a large bacterium. *Epulopiscium* is far from spherical. *Thiomargarita namibiensis* is spherical, but not when you consider just the cytoplasm. *Epulopiscium* also appears

to possess a highly invaginated cell membrane (128), which would influence the overall surface area to volume ratio and potentially the available energy. Additionally, the authors used a eukaryotic organism, *Amoeba proteus*, to calculate the metabolic rate per cell of a prokaryotic organism, *Epulopiscium*. While quantitative real-time PCR results suggest that *Epulopiscium* contains tens of thousands of copies of its genome (103) only three single copy genes were assayed, none of which are involved in energy production. As such, one cannot rule out the possibility of amplification of specific genomic resources, such as those involved in ATP production. Finally, the authors do not appear to take into consideration that at certain times during its life cycle the cytoplasm of *Epulopiscium* is filled with daughter cells, with their own DNA and energetic membranes which would ultimately influence the bioenergetics of this large cell. However, concerning *Epulopiscium,* their point seems to be that this organism is not more complex because it spends more energy maintaining multiple copies of a complete genome than they (the authors) feel is necessary. However, they make the argument that these cells are polyploid because they are large, which I am in agreement with. This polyploidy would allow these cells to make proteins only when and where they are needed, an energy conserving feature.

While they never come out and say it, behind their arguments seems to be the idea that evolution has a "goal" to create highly complex organisms and prokaryotes must therefore be failures. Evolution however is a process, not a goal-oriented endeavor. Additionally, prokaryotes are prokaryotes because they are highly successful in their niches. Their conclusion that most prokaryotes are limited by energetics is true. An exception to this may be the cyanobacteria, which are

sufficiently small and possess expanded bioenergetic membranes in the form of internal thylakoid membranes (82). In general as cell size increases, the surface-to-volume area decreases so the cell can only produce so much energy. However, their argument that prokaryotes could only be more complex if they had internal, membrane-bound, energy producing structures is somewhat circular: they would then cease to possess prokaryotic cell structure and by definition become a eukaryote. Ultimately their conclusion that mitochondria are essential for the evolution of complexity is sound, as more energy is needed than can be supplied by the plasma membrane.

Conclusions

For many years a defining characteristic of organisms belonging to the domain Bacteria was the arrangement of their genetic material in a single circular chromosome. Advances in the field of molecular biology have expanded this definition to include bacteria that have linear chromosomes, plasmids or more than one of these genetic elements as part of their genome, as well as those that are polyploid. This literature review has examined the better studied of these polyploid organisms, relating polyploidy to the cell biology of the organism. For several of these microorganisms, it would seem that polyploidy has developed as an advantageous adaptation. For example, the polyploidy in *D. radiodurans* allows for the repair of double-strand DNA breaks in response to desiccation or radiation (32). For some, it may be an adaptation to its niche. In *B. aphidicola*, the aphid symbiont, polyploidy appears to be a response to their obligate intracellular lifestyles and the physical

constraints put on cell division, but not chromosome replication (80, 81). By having multiple copies of their chromosome they are able to supply their host with essential amino acids, thus maintaining this obligate symbiotic relationship (60). In others, such as the extremely polyploid surgeonfish symbiont, *Epulopiscium* type B, polyploidy seems to be an evolutionary adaption, which has allowed it to retain its unusually large cell size (9) and unusual mode of reproduction (8, 155). Regardless, true understanding of this state, and those of other polyploid organisms will only be answered by continued study of the organism in its natural environment through molecular microbial ecology based studies.

References

1. **Adams, K. L. & Wendel, J. F.** 2005. Polyploidy and genome evolution in plants. Curr. Opin. Plant Biol. **8:** 135-141.

2. **Akerlund, T., Norddstrom, K. & Bernander, R.** 1995. Analysis of cell size and DNA content in exponentially growing and stationary-phase batch cultures of *Escherichia coli*. J. Bacteriol. **177**: 6791-6797.

3. **Amann, R. I., Ludwig, W. & Schleifer, K. H.** 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. **59:** 143-169.

4. **Anderson, A. W., Norden, H., Cain, R. F., Parrish, G. & Duggan, D.** 1956. Studies on the radio-resistant micrococcus. I. The isolation, morphology, cultural characteristics and resistance to gamma radiation. Food Technol. **10:** 575-577.

5. **Anderson, P. G., Brodbeck, B. V. & Mizell, R. F.** 1992. Feeding of the leafhopper, *Homalodisca coagulata*, in relation to xylem fluid chemistry and tension. J. Insect. Physiol. **38:** 611-622.

6. **Andrew, D., J., Henderson, K. D. & Seshaiah, P.** 2000. Salivary gland development in *Drosophila melanogaster*. Mech. Dev. **92:** 5-17.

7. **Angert, E. R.** 2005. Alternatives to binary fission. Nat. Rev. Microbiol. **3:** 221-224.

8. **Angert, E. R & Clements, K. D.** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:** 827-835.

9. **Angert, E. R., Clements, K. D. & Pace, N. R.** 1993. The largest bacterium. Nature. **362:** 239-241.

10. **Angert, E. R. & Losick, R. M.** 1998. Propogation by sporulation in the guinea pig symbiont *Metabacterium polyspora.* Proc. Natl. Sci. U S A. **95:** 1451-1456.

11. **Bailey, M. T., Walton, J. C., Dowd, S. E., Weil, Z. M., & Nelson, R. J.** 2010. Photoperiod modulates gut bacteria composition in male Siberian hamsters (*Phodopus sungorus*). Brain Behav. Immun. **24:** 577-584.

12. **Becak, M. L. & Kobashi, L. S.** 2004. Evolution of polyploidy and gene regulation in Anura. Genet. Mol. Res. **3:** 195-212.

13. **Bendich, A. J.** 1987. Why do chloroplasts and mitochondria contain so many copies of their genome? BioEssays **6:** 279-282.

14. **Bentley, S. D., Chater, K. F, Cerdeño-Tárraga, A.-M., Challis, L. G., Thomson, N. R., James, K. D., Harris, D. E., Quail, M. A., Kieser, H., Harper, D., Bateman, A., Brown, S., Chandra, G., Chen, C.W., Collins, M., Cronin, A., Fraser, A., Goble, A., Hidalgo, J., Hornsby, T., Howarth, S., Huang, C.-H., Kieser, T., Larke L., Murphy, L., Oliver, K., O'Neil, S., Rabbinowitsch, E., Rajandream, M.-A., Rutherford, K., Rutter, S., Seeger, K., Saunders, D., Sharp, S., Squares, R., Squares, S., Taylor, K., Warren, A., Wietzorrek, J., Woodward, B. G., Barrell, J., Parkhill, T. & Hopwood, D. A.** 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). Nature. **417:** 141-147.

15. **Bhagwat, A. S. & Lieb, M**. 2002. Cooperation and competition in mismatch repair: very short-patch repair and methyl-directed mismatch repair in *Escherichia coli*. Mol. Microbiol. **44:** 1421-1428.

16. **Blasius, M., Sommer, S. & Hubscher, U.** 2008. *Deinococcus radiodurans*: what belongs to the survival kit? Crit. Rev. Biochem. Mol. **43:** 221-238.

17. **Borst, P. & Kroon, A. M.** 1969. Mitochondrial DNA: Physicochemical properties, replication and genetic function. Int. Rev. Cytol. **26:** 107-190.

18. **Brinza, L., Viñuelas, J., Cottret, L., Calevro, F., Rahbé, Y., Febvay, G., Duport, G., Colella, S., Rabatel, A., Gautier, C., Fayard, J.-M., Sagot, M.-F. & Charles, H.** 2009. Systematic analysis of the symbiotic function of *Buchnera aphidicola*, the primary endosymbiont of the pea aphid *Acyrthosiphon pisum*. C.R. Biologies. **332:** 1034-1049.

19. **Burger, G., Gray, M. W. & Lang, F. B.** 2003. Mitochondrial genomes: anything goes. Trends Genet. **19:** 709-716.

20. **Cantfore, P., Polsa, P. L., Francasso, F., Flagella, Z. & Gadalta, M. N.** 1986. Quantitation of mitochondrial RNA species during rat liver development: the concentration of cytochrome oxidase subunit I (CoI) mRNA increases at birth. Cell Different. **19:** 1215-132.

21. **Casjens, S**. 1998. The diverse and dynamic structure of bacterial genomes. Annu. Rev. Genet. **32:** 339-377.

22. **Cattolico, R. A.** 1978. Variation in plastid number. Plany Physiol. **62:** 558-562.

23. **Cavelier, L., Johannisson, A. & Gyllensten, U.** 2000. Analysis of mtDNA copy number and composition of single mitochondrial particles using flow cytometery and PCR. Exp. Cell Res. **259:** 79-85.

24. **Chater, K. F. & Horinouchi, S.** 2003. Signaling early developmental events in two highly diverged *Streptomyces* species. Mol. Microbiol. **48:** 9-15.

25. **Chatton, E. & Perard, C.** 1913. Schizophytes du caecum du cobaye. *Metabacterium polyspora* n. g., n. s. C. R. Hebd. Soc. Biol. (Paris). **74:** 1232-1234.

26. **Chen, Z. J.** 2010. Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci. **15:** 57-71.

27. **Choat, J. H. & Clements, K. D.** 1998. Vertebrate herbivores in marine and terrestrial environments: A nutritional ecology perspective. Annu. Rev. Ecol. Syst. **29:** 375-403.

28. **Clements, K. D.** 1991. Endosymbiotic communities of two herbivorous labroid fishes *Odax cyanomelas* and *O. pullus*. Mar. Biol. **109:** 223-230.

29. **Clements, K. D. & Bullivant, S.** 1991 An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. J. Bacteriol. **173:** 5359-5362.

30. **Clements, K. D. & Choat, J. H.** 1995. Fermentation in Tropical Marine Herbivorous Fishes. Physiol. Zoo. **68:**355 -378.

31. **Comai, L.** 2005. The advantages and disadvantages of being a polyploid. Nat. Rev. Gen. **6:** 836-846.

32. **Cox, M.M. & Battista, J. R.** 2005. *Deinococcus radiodurans* – the consummate survivor. Nat. Rev. Microbiol. **3:** 882-892.

33. **Criss, A. K., Kilne, K. A. & Seifart, H. S.** 2005. The frequency and rate of pilin antigenic variation in *Neisseria gonorrhoeae*. Mol. Microbiol. **58:** 510-519.

34. **Daly K., Stewart, S. C., Flint, H. J. & Shirazi-Beechey, S. P.** 2001. Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. FEMS Microbiol. Ecol. **38:** 141-151.

35. **Daly, M.** 2009. A new perspective on radiation resistance based on *Deinococcus radiodurans*. Nat. Rev. Microbiol. **7:** 237-245.

36. **Daly, M. J. & Minton, K. W.** 1997. Recombination between a resident plasmid and the chromosome following irradiation of the radioresistant bacterium *Deinococcus radiodurans*. Gene. **187:** 225-259.

37. **Daly, M. J., Ouyang, L., Fuchs, P. & Minton, K. W.** 1994. In vivo damage and RecA-dependent repair of plasmid and chromosomal DNA in the radiation-resistant bacterium *Deinococcus radiodurans*. J. Bacteriol. **176:** 3508-3517.

38. **Degnan, P. H., Leonardo, T. E., Cass, B. N., Hurwitz, B., Stern, D., Gibbs, R. A., Richards, S., & Moran, N. A.** 2010. Dynamics of genome evolution in facultative symbionts of aphids. Environ. Microbiol. **12:** 2060-2069.

39. **Donachie, W. D. & Robinson, A. C.** 1987. Cell division: Parameters, values and the process. ASM, Washington, D.C.

40. **Dose, K. A., Bieger – Dose, A., Labusch, M. & Gill, M.** 1992. Survival in extreme dryness and DNA single strand breaks. Adv. Space Res. **12:** 221-229.

41. **Dose, K. Al, Bieger-Dose, A., Kerz, O. & Gill, M.** 1991. DNA-strand breaks limit survival in extreme dryness. Origins Life Evol. Biosphere. **21:** 177-187.

42. **Douglas, A. E.** 1998. Nutritional interactions in insect-microbial symbioses: aphids and their symbiotic bacteria *Buchnera*. Ann. Rev. Entomol. **43:** 17-38.

43. **Dyall, S. D., Brown, M. T. & Johnson, P. J.** 2004. Ancient invasions: from endosymbionts to organelles. Science. **304:** 253-257.

44. **Edwards J. E., Mcewan N. R., Travis A. J. & Wallace R. J.** 2004. 16S rDNA library-based analysis of ruminal bacterial diversity. Antonie van Leeuwenhoek **86:** 263-281.

45. **Efuet, E. T., Pulakat, L. & Gavini, N.** 1996. Investigations on the cell volumes of *Azotobacter vinelandii* by scanning electron microscopy. J. Basic Microbiol. **36:** 229-234.

46. **Errington, J.** 2003. Regulation of endospore formation in *Bacillus subtilis*. Nat. Rev. Microbiol. **1:** 117-126.

47. **Ersland, D. R., Aldrich, J. & Cattolico, R. A.** 1981. Kinetic complexity, homogeneity and copy number of chloroplast DNA from the marine algae *Olisthodiscus luteus.* Plant Physiol. **68:** 1468-1473.

48. **Fidopiastis, P. M., Miyamato, C. M., Jobling, M. G., Meighen, E. A., & Ruby, E. G.** 2006. Characterizing the resident, fermentative microbial consortium in the hindgut of the temperate-zone herbivorous fish, *Hermosilla azurea* (Teleostei : Kyphosidae). Mar. Biol. **148:** 631-642.

49. **Finkel, S. E.** 2006. Long-term survival during stationary phase: Evolution and the GASP phenotype. Nat. Rev. Microbiol. **4:** 113-120.

50. **Finkel, S. E. & Kolter, R.** 1999. Evolution of microbial diversity during prolonged starvation. Proc. Natl. Acad. Sci. USA **96**: 4023-4027.

51. **Fishelson, L., Montgomery, W. L. & Myrberg, A. A.** 1985. A unique

symbiosis in the gut of a tropical herbivorous surgeonfish (Acanthuridae: Teleostei) from the Red Sea. Science. **229:** 49-51.

52. **Frazzon, J. & Dean, D. R.** 2002. Biosynthesis of the nitrogenase iron-molybdenum-cofactor from *Azotobacter vinelandii*. Met. Ions Biol. Syst. **39:** 163-186.

53. **Gallardo, M. H., Bickham, J. W., Honeycutt, R. L, Ojeda, R. A & Kohler, N.** 1999. Discovery of tetraploidy in a mammal. Nature **401:** 341.

54. **Gómez-Valero, L., Silva, F. J., Christophe Simon, J. & Latorre A.** 2007. Genome reduction of the aphid endosymbiont *Buchnera aphidicola* in a recent evolutionary time scale. Gene. **389:** 87-95.

55. **Gordon, J. I., Hooper, L. V., McNevin, M. S., Wong, M. , & Bry, L.** 1997. Epithelial cell growth and differentiation. III. Promoting diversity in the intestine: conversations between the microflora, epithelium, and diffuse GALT. Am. J. Physiol. Gastrointest. Liver Physiol. **273:** G565-G70.

56. **Grantcharova, N., Lustig, U. & Flärdh, K.** 2005. Dynamics of FtsZ assembly during sporulation in *Streptomyces coelicolor* A3(2). J. Bacteriol. **187:** 3227-3237.

57. **Gronlund, M. M., Lehtonen, O. P., Eerola, E., & Kero, P.** 1999. Fecal microflora in healthy infants born by different methods of delivery: permanent changes in intestinal flora after cesarean delivery. J. Pediatr. Gastroenterol. Nutr. **28:** 19-25.

58. **Gunduz, E. A. & Douglas, A. E**. 2009. Symbiotic bacteria enable insect to use a nutritionally inadequate diet. Proc. R. Soc. B. **276:** 987-991.

59. **Hamrick, T. S., Dempsey, J. F., Cohen, M. S. & Cannon, J. G.** 2001. Antigenic variation of gonococcal pilin expression *in vivo*: analysis of the strain FA1090 pilin repertoire and identification of the *pilS* gene copies recombining with *pilE* during experimental human infection. Microbiology. **147:** 839-849.

60. **Hansen, A. K. & Moran, N. A.** 2011. Aphid genome expression reveals host-symbiont cooperation in the production of amino acids. Proc. Natl. Acad. Sci. U S A **108:** 2849-2854.

61. **Hansen, M. T.** 1978. Multiplicity of genome equivalents in the radiation-resistant bacterium *Micrococcus radiodurans.* J. Bacteriol. **134:** 71-75.

62. **Harris, R. S.,** 1997 Mismatch repair protein MutL becomes limiting during stationary-phase mutation. Genes Dev. **11:** 2426-2437.

63. **Harris, R. S**., 1999. Mismatch repair is diminished during stationary phase mutation. Mutat. Res. **437**: 51-60.

64. **Heddi, A., Vallier, A., Anselme, C., Xin, H., Rahbe, Y., & Wäckers, F.** 2005. Molecular and cellular profiles of insect bacteriocytes: mutualism and harm at the initial evolutionary step of symbiogenesis. Cell Microbiol. **7:** 293-305.

65. **Hegarty, M. J. & Hiscock, S. J.** 2008. Genomic clues to the evolutionary success of polyploid plants. Cur. Biol. **18:** R435-R444.

66. **Helmsetter, C. E.** 1987. Timing of synthetic activities in the cell cycle. *In* F. C. Neidhardt et al. (ed.), *Escherichia coli* and *Salmonella*, 2nd ed., vol. 2. ASM Press, Washington DC.

67. **Herrmann, R.G.** 1970. Multiple amounts of DNA related to the size of chloroplasts. Planta. **90:** 80-96.

68. **Honeycutt, R. J., McClelland, M. & Sobral, B. W. S.** 1993. Physical map of the genome of *Rhizobium meliloti* 1021. J. Bacteriol. **175:** 6945-6952.

69. **Hooper L. V., Midtvedt T. & Gordon J.I.** 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. **22:** 283-307.

70. **Hopkins, M. J., Sharp R., & Macfarlane, G. T.** 2001. Age and disease related changes in intestinal bacterial populations assessed by cell culture, 16S rRNA abundance, and community cellular fatty acid profiles. Gut **48**: 198-205.

71. **Hopwood, D. A.** 2006. Soil to genomics: The *Streptomyces* chromosome. Annu. Rev. Genet. **40:** 1 -23.

72. **Hopwood, D. A. & Glauert, A. M.** 1960. Observations of the chromatinic bodies of *Streptomyces coelicolor*. J. Biophysic. and Biochem. Cytol. **8:** 257-265.

73. **Iborra, F. J., Kimura, H. & Cook, P. R.** 2004. The functional organization of mitochondrial genomes in human cells. BMC Biol. **2:** 1-14.

74. **Ishikawa, H.** 2004. Genome of *Buchnera* sp. APS, an intracellular bacterium of the pea aphid *Acyrthosiphon pisum*. Springer. Netherlands.

75. **James, T. W. & Jope, C.** 1978. Visualization by fluorescence of chloroplast DNA in higher plats by means of the DNA-specific probe 4'6-diamidino-2-phenylindole. J. Cell Biol. **79:** 623-630.

76. **Kennedy, C. P., Rudnick, P., Macdonald, T. & Melton, T.** 2005. Genus *Azotobacter* in Bergey's Manual of Systematic Bacteriology. **2:** 384-401.

77. **Kennedy, C. P. & Toukdarian, A.** 1987. Genetics of azotobacters: applications to nitrogen fixation and related aspects of metabolism. Annu. Rev. Microbiol. **41:** 227-258.

78. **Khachane, A. N., Timmis, K. N., & Martins dos Santos, V. A. P.** 2007. Dynamics of reductive genome evolution in mitochondria and obligate intracellular microbes. Mol. Biol. Evol. **24:** 449-456.

79. **Kline, K. A., Sechman, E. V., Skaar, E. P & Seifert, H. S.** 2003. Recombination, repair and replication in the pathogenic *Neisseriae*: the 3 R's of molecular genetics of two human specific bacterial pathogens. Mol. Microbiol. **50:** 3-13.

80. **Komaki, K. & Ishikawa, H.** 1999. Intracellular bacterial symbionts of aphids possess many genomic copies per bacterium. J. Mol. Evol. **48:** 717-722.

81. **Komaki, K. & Ishikawa, H.** 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. **30:** 253-258.

82. **Lane, N.** 2011: Energetics and genetics across the prokaryote-eukaryote divide. Biol. Dir. **6:** 1-35.

83. **Lane, N. & Martin, W.** 2010. The energetics of genome complexity. Nature. **467:** 929-934.

84. **Lee, H. O., Davidson, J. M. & Duronio, R. J**. 2009. Endoreplication: polyploidy with a purpose. Genes Dev. **23:** 2461-2477.

85. **Leser, T. D., Amenuvor, T. K., Jensen T. K., Lindecrona R. H., Boye M. & Møller, K.** 2002. Culture independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Appl. Environ. Microbiol. **68:** 673-690.

86. **Ley, R. E., Backhëd, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I.** 2005. Obesity alters gut microbial ecology. Proc. Natl. Acad. Sci. USA. **102:** 11070-11075.

87. **Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R. R., Bircher, J. S., Schlegel, M. L., Tucker, T. A., Schrenzel, M. D., Knight, R. & Gordon, J. I.** 2008. Evolution of mammals and their gut microbes. Science. **320:** 1647-1651.

88. **Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I.** 2008. Worlds within worlds: evolution of the vertebrate gut microbiota. Nat. Rev. Microbiol. **6:** 776-788.

89. **Ley, R. E., Peterson, D. A. & Gordon, J. I.** 2006. Ecological and evolutionary forces shaping microbial diverstiy in the human intestine. Cell. **124:** 837-848.

90. **Liu, W.T., Marsh, T.L., Cheng, H. & Forney, L. J.** 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. **63:** 4516-4522.

91. **Livak, K.J., Flood, S.J.A, Giusti, W. & Deetz, K.** 1995. Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. PCR Methods Appl. **4:** 357-362.

92. **Louis, P., Scott, K. P., Duncan, S. H. & Flint, H. J.** 2007. Understanding the effects of diet on bacterial metabolism in the large intestine. J. Appl. Micorbiol. **102:** 1197-1208.

93. **Loxdale, H. D.** 2010. Rapid genetic changes in natural insect populations. 4 Ecol. Entomol. **35:**155-164.

94. **Luckey, T.D.** 1972. Introduction to intestinal microecology. Am. J. Clin. Nutr. **25:** 1292-95.

95. **Maiden, M. C.** 2008. Population genomics: diversity and virulence in the *Neisseria*. Curr. Opin. Microbiol. **11:** 467-471.

96. **Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V., & Daly, M. J.** 2001. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. Microbiol. Mol. Biol. Rev. **65:** 44-79.

97. **Maldonado, R., Garzon, A., Dean, D. & Casadesus, J.** 1992. Gene dosage analysis in *Azotobacter vinelandii.* Genetics. **132:** 869-878.

98. **Maldonado, R., Jimenez, J. & Casadesus, J.** 1994. Changes of ploidy during the *Azotobacter vinelandii* growth cycle. J. Bacteriol. **176:** 3911-3919.

99. **Malik, A. N., Shahni, R., Rodriguez-de-Ledesma, A., Laftah, A & Cunningham, P.** 2011. Mitochondrial DNA as a non-invasive biomarker: Accurate quantification using real-time quantitative PCR without co-amplification of pseudogenes and dilution bias. Biochem. Bioph. Res. Co. **412:** 1-7.

100. **Mattimore, V. & Battista, J. R.** 1996. Radioresistance of *Deinococcus radiodurans*: functions necessary to survive ionizing radiation are also necessary to survive prolonged desiccation. J. Bacteriol. **178:** 633-637.

101. **McCutcheon, J. P. & Moran, N. A.** 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbioses. Proc. Natl. Acad. Sci. USA **104:** 19392-19397.

102. **McEwan, N. R., Abecia, L., Regensbogenova, M., Adam, C. L, Findlay, P. A & Newbold, C. J.** 2005. Rumen microbial population dynamics in response to photoperiod. Lett. Appl. Microbiol. **41:** 97-101.

103. **Mendell, J. E., Clements, K. D., Choat, J. H. and Angert, E.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U S A **105:** 6730-6734.

104. **Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A., Barloy-Huber, F., Gailbert, F., Kondorosi, A. & Kondorosi, E.** 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium* – legume symbiosis. Proc. Natl. Acad. Sci. USA **103:** 5230-5235.

105. **Miller, J. H.** 1996. Spontaneous mutators in bacteria: Insights into pathways of mutagenesis repair. Annu. Rev. Microbiol. **50:** 625-643.

106. **Miyakawa, I., Aoi, H. & Sando, N.** 1984. Fluorescence microscopic studies of mitochondrial nucleoids during meiosis and sporulation in the yeast *Saccharomyces cerevisiae*. J. Cell Sci. **66:** 21-38.

107. **Moran, N. A. & Degnan, P. H.** 2006. Functional genomics of *Buchnera* and the ecology of aphid hosts. Mol. Ecol. **15:** 1251-1261.

108. **Moran, N. A., Munson, M. A., Baumann, P. & Ishikawa, H.** 1993. A molecular clock in the endosymbiotic bacteria is calibrated using the insect host. Proc. R. Soc. Lond. B. **253:** 161-171.

109. **Morton, R.S.** 1977. Gonorrhoea. WW Saunders Co. London. 1-3.

110. **Moshiri, F., Kim, J. W., & Maier, R. J.** 1994. The FeSII protein of *Azotobacter vinelandii* is not essential for aerobic nitrogen fixation but confers significant protection to oxygen-mediated inactivation of nitrogenase *in vitro* and *in vivo.* Mol. Microbiol. **14:** 101-114.

111. **Mountfort, D. O., Campbell, J. & Clements, K. D.** 2002. Hindgut fermentation in three species of marine herbivorous fish. Appl. Environ. Microbiol. **68:**1374-1380.

112. **Mountzouris, K. C., McCartney, A. L., & Gibson, G. R.** 2002**.** Intestinal microflora of human infants and current trends for its nutritional modulation. Br. J. Nutr. **87:** 405-420.

113. **Munson, M. A., Baumann, P. & Kinsey, M. G.** 1991. *Buchnera* gen. nov. and *Buchnera aphidicola* sp. nov., a taxon consisting of the mycetocyte-associated, primary endosymbionts of aphids. Int. J. Syst. Bacteriol. **41:** 566-568.

114. **Nass, M. M., Nass, S. & Afzelius, B. A.** 1965. The general occurrence of mitochondrial DNA. Exp. Cell Res. **37:** 516-539**.**

115. **Nass, M. M.** 1969. Mitochondrial DNA: Advances, problems, and goals. Science. **165:** 25-35.

116. **Nelson K. E., Zinder, S., Hance, I., Burr, P., Odongo, D., Wasawo, D., Odenyo, A. & Bishop, R.** 2003. Phylogenetic analysis of the microbial populations in the wild herbivore gastrointestinal tract: Insights into an unexplored niche. Environ. Microbiol. **5:** 1212–1220.

117. **Nikoh, N., McCutcheon, J. P., Kudo, T., Miyagishima, S., Moran, N. A. & Nakabachi, A.** 2010. Bacterial genes in the aphid genome: Absence of functional gene transfer from *Buchnera* to its host. PLoS Genet. **6:** 1-21.

118. **O'Hara, A. M. & Shanahan, F.** 2006. The gut flora as a forgotten organ. EMBO reports. **7:** 688-693.

119. **Pace, N. R., Stahl, D. A., Lane, D. J. & Olsen, G. J.** 1985. Analyzing natural microbial populations by rRNA sequences. ASM News **51:** 4-12.

120. **Pérez-Brocal V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J., Silva, F., Moya, A. & Latorre, A.** 2006. A small microbial genome: the end of a long symbiotic relationship? Science. **314:** 312-3.

121. **Phadnis, S. H., Dimri, G. P. & Das, H. K.** 1988. Segregation characteristics of multiple chromosomes of *Azotobacter vinelandii.* J. Genet. **67:** 37-42.

122. **Poliakov A., Russell C. W., Ponnala L., Hoops H. J., Sun Q., Douglas A. E. & van Wijk, K. J.** 2011. Large-scale label-free quantitative proteomics of the pea aphid-*Buchnera* symbiosis. Mol. Cell Proteomics. **10:** M110.007039**.**

123. **Pulakat, L., Lee, S. & Gavini, N.** 2002. Genome of *Azotobacter vinelandii:* counting of chromosomes by utilizing copy number of a selectable genetic marker. Genetica. **115:** 147-158.

124. **Punita, S. J., Reddy, M. A. & Das, H. K.** 1989. Multiple chromosomes of *Azotobacter vinelandii.* J. Bac. **171:** 3133-3188.

125. **Ranney, T. G.** 2006. Polyploidy: from evolution to new plant development. International Plant Propagators Society **56:**137-142.

126. **Rauwolf, U., Golczyk, H., Greiner, S. & Herrmann, R. G.** 2010. Variable amounts of DNA related to the size of chloroplasts III. Biochemical determinations of DNA amounts per organelle. Mol. Genet. Genomics. **283:** 35-47.

127. **Robinow, C.** 1951. Observations on the structure of *Bacillus* spores. J. Gen. Microbiol. **5:** 439-457.

128. **Robinow, C. & Angert, E. R.** 1998. Nucleoids and coated vesicles of *Epulopiscium* spp. Arch. Microbiol. **170:** 227-235.

129. **Rowen, B. A., Oldenburg, D. J. & Bendich, A.** 2007. A high-throughput method for detection of DNA in chloroplasts using flow cytometry. Plant Methods. **3:** 1-13.

130. **Rowen, B. A. & Bendich, A. J.** 2009. The loss of DNA from chloroplasts as leaves mature: fact or artefact? J Exp. Bot. **60:** 3005-3010.

131. **Ruban-Ośmialowska, B., Jakimowicz, D., Smulczyk-Krawczyszyn, A., Chater, K. & Zakrzewska-Czerwiński.** 2006. Replisome localization in vegetative and aerial hyphae of *Streptomyces coelicolor*. J. Bacteriol. **188:** 7311-7316.

132. **Sadoff, H. L., Shimel, B. & Ellis, S.** 1979. Characterization of *Azotobacter vinelandii* deoxyribonucleic acid and folding chromosomes. J. Bacteriol. **138:** 871-877.

133. **Sakai, A., Takano, H. & Kuroiwa, T.** 2004. Organelle nuclei in higher plants: Structure, composition function and evolution. Int. Rev. Cyt. **238:** 59-118.

134. **Salman, V., Amann, R., Girnth, A-C., Polerecky, L., Bailey, J. V., 4 Høgslund, S., Jessen, G., Pantoja, S. & Schulz-Vogt, H. N.** 2010. A single-cell sequencing approach to the classification of large, vacuolated sulfur bacteria. Syst. Appl. Microbiol. **34:** 243-259.

135. **Satoh, M. & Kuroiwa, T.** 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Exp. Cell Res. **196:** 137-140.

136. **Savage, D. C.** 1986. Gastrointestinal microflora in mammalian nutrition. Ann. Rev. Nutr. **6:** 155-78.

137. **Schofield, M. J. & Hsieh, P.** 2003. DNA mismatch repair: molecular mechanisms and biological function. Annu. Rev. Microbiol. **57:** 579-608.

138. **Schulz, H. N.** 2006. The genus *Thiomargarita* in The Prokaryotes. **6:** 1156-1163.

139. **Schulz, H. N., Brinkhoff, T., Ferdelman, T. G., Mariné, M. H., Teske, A. & Jørgensen, B. B.** 1999. Dense populations of a giant sulfur bacterium in Namibiam shelf sediments. Science. **284:** 493-495.

140. **Schulz, H. N. & de Beer, D.** 2002. Uptake rates of oxygen and sulfide measured with individual *Thiomargarita namibiensis* cells by using microelectrodes. Appl. Environ. Microbiol. **68:** 5746-5749.

141. **Schulz, H. N. & Schulz, H. D.** 2005. Large sulfur bacteria and the formation of phosphorite. Science. **307:** 416-418.

142. **Schwedock, J., McCormick, J. R., Angert, E. R., Nodwell, J.R & Losick, R.** 1997. Assembly of the cell division protein FtsZ into ladder-like structures in the aerial hyphae of *Streptomyces coelicolor*. Mol. Microbiol. **25:** 847-858.

143. **Setubal, J., dos Santos, P., Goldman, B. S., Ertesvag, H., Espin, G., Rubio, L. M., Valla, S., Almeida, N. F., Balasubramanian, D., Cromes, L., Curatti, L., Du, Z., Godsy, E., Goodner, B., Hellner-Burris, K., Hernandez, J. A., Houmiel, K., Imperial, J., Kennedy, C., Larson, T. J., Latreille, P., Ligon, L.S., Lu, J., Maerk, M., Miller, N. M., Norton, S., O'Carroll, I. P., Paulsen, I., Raulfs, E. C., Roemer, R., Rosser, J., Segura, D., Slater, S., Stricklin, S. L., Studholme, D. J., Sun, J., Viana, C.J., Wallin, E., Wang, B., Wheeler, C., Zhu, H., Dean, D., Dixon, R. & Wood, D.** 2009. Genome sequence of *Azotobacter vinelandii*, an obligate aerobe specialized to support diverse anaerobic metabolic processes. J. Bacteriol. **191:** 4534-4545.

144. **Shigenobu, S., Watanabe, H., Hattori, M., Sakaki, Y. & Ishikawa, H.** 2000. Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. APS. Nature. **407:** 81-86.

145. **Stackebrandt E. & B.M. Goebel** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:** 846–849.

146. **Sugimoto-Shirasu, K. & Roberts, K**. 2003. "Big it up": endoreduplication and cell-size control in plants. Curr. Opin. Plant Biol. **6:** 544-553.

147. **Suh, M., Pulakat, L & Gavini, N.** 2001. Isolation and characterization of *nif DK*::kanamycin and nitrogen fixation proficient *Azotobacter vinelandii* strain, and its implication on the status of multiple chromosomes in *Azotobacter.* Genetica. **110:** 101-107.

148. **Taylor, R. W. & Turnbull, D. M.** 2005. Mitochondrial DNA mutations in human disease. Nat. Rev. Genet. **6:** 389-402.

149. **Tijima, K., Arai, S., Ogata, K., Nagamine, T., Matsui, H., Nakamura, R., Aminov, I. and Benno, Y.** 2000. Rumen bacterial community transition during adaptation to high-grain diet. Anaerobe **6:** 273-284.

150. **Tobiason, D. M. & Seifert, H. S.** 2006. The obligate human pathogen, *Neisseria gonorrhoeae* is polyploid. PLoS Biol. **4:** 1069-1076.

151. **Tobiason, D. M. & Seifert, H. S.** 2010. Genomic content of *Neisseria* species. J. Bacteriol. **192:** 2160-2168.

152. **Walmsley, J., Toudarian, A. & Kennedy, C.** 1994. The role of regulatory genes *nifA, vnfA, anfA, nfrA, ntrC* and *rpoN* in expression of genes encoding the three nitrogenases in *Azotobacter vinelandii*. Arch. Microbiol. **162:** 422-429.

153. **Wang, Y. & Bogenhagen, D. F.** 2006. Human mitochondrial DNA nucleoids are linked to protein folding machinery and metabolic enzymes at the mitochondrial inner membrane. J. Biol. Chem. **281:** 25791-25802.

154. **Ward, R. J. & Angert, E. R.** 2008. DNA replication during endospore development in *Metabacterium polyspora.* Mol. Microbiol. **67:** 1360-1370.

155. **Ward, R. J., Clements, K. D., Choat, J. H. & Angert, E. R.** 2009. Cytology of terminally differentiated *Epulopiscium* mother cells. DNA Cell Biol. **28:** 57-64.

156. **Williamson, D. H. & Fennell, D. J.** 1975. The use of fluorescent DNA-binding agent for detecting and separating yeast mitochondrial DNA. In Methods in Cell Biology. **12:** 335-351.

157. **Wilson, A. C., Ashton, P. D., Calevro, F. Charles, H. Colella, S., Febvay, G., Jander, G., Kushlan, P. F., Macdonald, S. J., Schwartz, J. F., Thomas, G. H. & Douglas, A. E.** 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. Insect Mol. Biol. **19 (Suppl. 2):** 249-258.

158. **Wostmann, B. S.** 1981. The germfree animal in nutritional studies. Annu. Rev. Nutr. **1:** 257-79.

159. **Wostmann, B. S., Larkin, C., Moriarty, A., & Bruckner-Kardoss, E.** 1983. Dietary intake, energy metabolism, and excretory losses of adult germfree Wistar rats. Lab Anim. Sci. **33:** 46-50.

160. **Woyke, T., Tighe, D., Mavromatis, K., Clum, A., Copeland, A., Schackwitz, W., Lapidus, A., Wu, D., McCutcheon, J. P., McDonald, B.**

R., Moran, N. A., Bristow, J. & Cheng, J. F. 2010. One bacterial cell, one complete genome. PLoS Biol. **5:** 1-8.

161. **Wu, D., Daugherty, S., Van Eken, S., Pai, G., Watkins, K., Khouri, H., Tallon, L., Zaborsky, J., Dunbar, H., Tran, P. L., Morn, N. A. & Eisen, J. A.** 2006. Metabolic complimentarity and genomics of the dual bacterial symbiosis of sharpshooters. PLoS Biol. **4:** 1079-1092.

162. **Xu, J., & Gordon, J. I.** 2003. Inaugural article: Honor thy symbionts. Proc. Natl. Acad. Sci. USA. **100:** 10452-10459.

163. **Zoetendal, E. G., Akkermans, A. D. L., Akkermans-van Vilet, W. M., DeVisser, J. A. G. M., & DeVos, W. M.** 2001. The host genotype affects the bacterial community in the human gastrointestinal tract. Microb. Ecol. Health Dis. **13:** 129-134.

164. **Zuh, W., Chen, C-J. Thomas, C. E., Anderson, J. E., Jerse, A. & Sparling, P. F.** 2011. Vaccines for gonorrhea: can we rise to the challenge? Front. Microbiol. **2:** 1-13.

CHAPTER 2

EXTREME POLYPLOIDY IN A LARGE BACTERIUM[1]

[1] **Mendell, J. E., Clements, K. D., Choat, J. H. and Angert, E.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U S A **105:** 6730-6734.

Abstract

Cells rely on diffusion to move metabolites and biomolecules. Diffusion is highly efficient but only over short distances. While eukaryotic cells have broken free of diffusion-dictated constraints on cell size, most bacteria and archaea are forced to remain small. Exceptions to this rule are found among the bacterial symbionts of surgeonfish; *Epulopiscium* spp. are cigar-shaped cells that reach lengths in excess of 600 µm. A large *Epulopiscium* contains thousands of times more DNA than a bacterium such as *Escherichia coli*, but the composition of this DNA is not well understood. Here we present evidence that *Epulopiscium* contains tens of thousands of copies of its genome. Using quantitative, single-cell PCR assays targeting single-copy genes, we have determined that copy number is positively correlated with *Epulopiscium* cell size. While other bacteria are known to possess multiple genomes, polyploidy of the magnitude observed in *Epulopiscium* is unprecedented. The arrangement of genomes around the cell periphery may permit regional responses to local stimuli, thus allowing *Epulopiscium* to maintain its unusually large size. Surveys of the sequences of single-copy genes (*dnaA*, *recA* and *ftsZ*) revealed genetic homogeneity within a cell consistent with only a small amount (~1%) of the parental DNA being transferred to the next generation. The results also suggest that the

abundance of genome copies in *Epulopiscium* may allow for an unstable genetic feature, a long mononucleotide tract, in an essential gene. With the evolution of extreme polyploidy and large cell size, *Epulopiscium* has acquired some of the advantages of eukaryotic cells.

Introduction

It is well appreciated that many eukaryotes are orders of magnitude larger than all known members of the Bacterial and Archaeal domains. Eukaryotes have broken free of constraints on cell size by the development of sophisticated nutrient uptake systems, subcellular compartmentalization, and the use of a cytoskeleton and motor proteins to transport vesicles. With the further advance of multicellularity, cell and tissue specialization have allowed eukaryotes to attain tremendous dimensions (10, 65). Until recently (37, 59), bacterial (and archaeal) cells were considered simple, displaying little subcellular organization. While we now know that bacterial cells are also highly organized, possessing motor and cytoskeletal proteins, and even extensive intracellular membranes in some instances (62), these cells are believed to rely on diffusion to access nutrients and other metabolically important chemicals. Diffusion coefficients of small molecules may impose time constraints on metabolite flux (8) that require bacterial cells to maintain very high surface-to-volume ratios. As a result, no part of the cytoplasm is very far from the external environment, and so exchange is rapid enough to sustain metabolic processes. Most large bacteria fit this paradigm and maintain a very thin cytoplasm; they are extremely long and slender, or if spherical they contain an intracellular vacuole to press the cytoplasm into a thin layer just under the cytoplasmic membrane (9, 17, 28, 57). In addition, many large bacteria contain

intracellular mineral inclusions, which further reduce the volume of active cytoplasm (30). Possible exceptions to this standard are found within the Firmicutes (6).

Our model for studying the cell biology of large bacteria is *Epulopiscium* sp. type B, which occurs in the intestinal tract of the unicornfish *Naso tonganus* (2, 15). These type B cells attain lengths of 200 to 300 µm, widths of 50 to 60 µm (4), and reproduce solely by the formation of multiple internal offspring (Figure 2.1). This reproductive strategy likely evolved from endospore formation (5). A large *Epulopiscium* contains a substantial amount of DNA arranged around the periphery of the cytoplasm (51). This unusual feature may be key in the ability of these large cells to maintain an active metabolism despite having a low-surface-to-volume ratio. To characterize the size and conformation of the *Epulopiscium* sp. type B genome, we used quantitative PCR to enumerate the copy number of genes in individuals and in DNA extracted from populations of cells. The results of these surveys suggest that *Epulopiscium* is highly polyploid throughout its life cycle, and an individual contains tens of thousands of copies of its genome.

Materials and Methods

***Epulopiscium* collection.** *Epulopiscium* sp. morphotype B cells were obtained from *Naso tonganus*, collected by spearfishing on reefs around Lizard Island, Australia. Sections of the gut were removed, and intestinal contents were fixed in 80% ethanol and stored at -20°C. Individual *Epulopiscium* cells were collected from gut contents using a standard Gilson micropipettor and a dissecting microscope (Nikon SMZ-U). Cells were transferred five times through sterile ethanol wash buffer (80% ethanol; 145 mM NaCl; 50 mM Tris-HCl, pH 8.0; 0.05% Igepal) and rinsed in sterile deionized

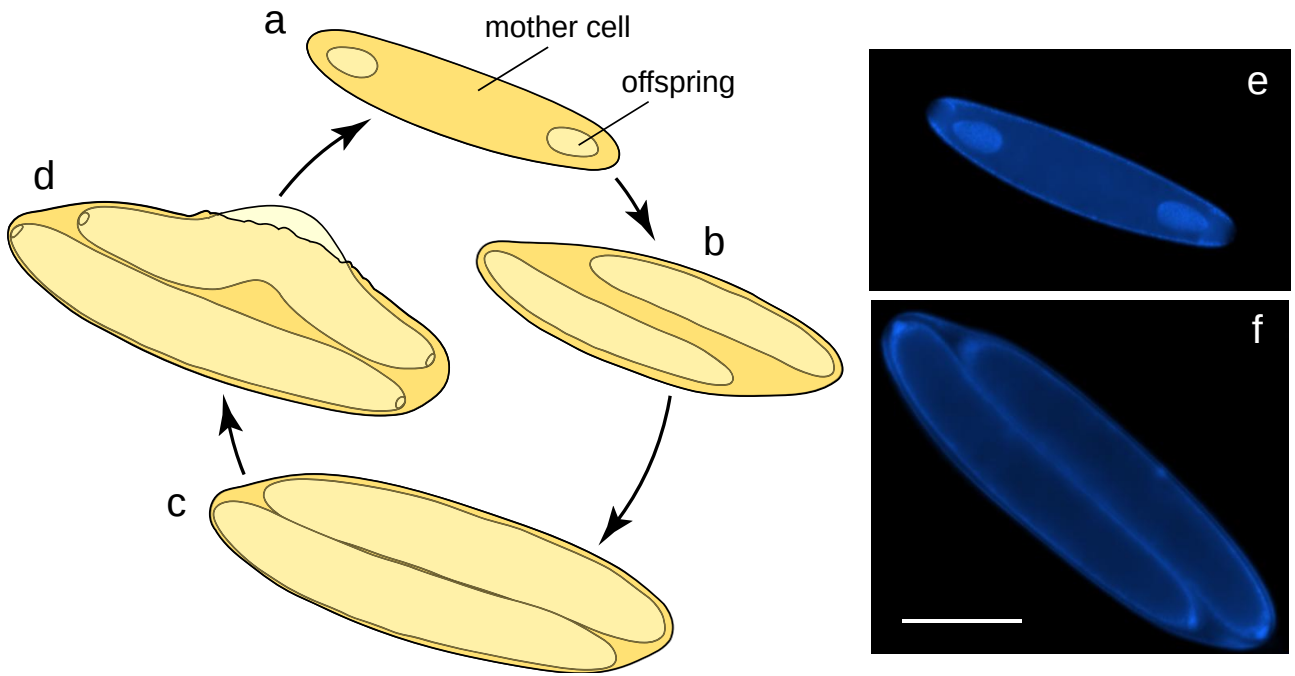


Figure 2.1. *Epulopiscium* sp. type B life cycle. Offspring production follows a circadian cycle. (a) Early in the day, a mother cell possesses small, internal offspring. (b and c) Offspring size increases throughout the day (b) until they fill the mother-cell cytoplasm (c). (d) Finally, ‘‘mature’’ offspring cells emerge from the mother-cell envelope. Note that before emergence, these cells begin to develop the next generation of offspring. (e and f) Images of DAPI-stained cells representing the populations of small (e) and large (f) Epulopiscium cells used in these studies. (Scale bar: 50 μm)

water.

DNA extraction and quantification. DNA was extracted from 5000 handpicked, washed cells. DNA extraction and quantification protocols were based on standard procedures (54). *Epulopiscium* cells were lysed by incubation in proteinase K (100 µg/ml in 10 mM Tris, pH 8.0) at 50°C for 1 h followed by six rounds of freeze-thaw. Cell lysate was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and the nucleic acids were precipitated with 0.3 M sodium acetate and ethanol. The pellet was rinsed with 70% ethanol and air-dried. The pellet was dissolved in sterile wateand treated with RNase A (10 µg/ml) at 37°C for 30 min. PicoGreen assays to quantify genomic DNA was performed as follows. DNA from bacteriophage lambda was diluted in TE (10 mM Tris, pH 8.0; 1 mM EDTA) to generate a dilution series ranging from 500 ng/ml to 10 ng/ml. Genomic DNA from *Epulopiscium* was diluted 1:5000 in TE. Aliquots (50 µl) of the standards and genomic DNA were dispensed in triplicate into wells of a microtiter plate. PicoGreen (Molecular Probes) was prepared according to manufacturer's instructions and 50 µl aliquots were mixed with the DNA solutions in each microtiter plate well. Relative fluorescence was determined using a Perkin Elmer LS50B fluorometer. Genomic DNA was also quantified using an ethidium bromide spot test. For the spot test, lambda DNA was used to generate a serial dilution ranging from 50 µg/ml to 1 µg/ml. Equal volumes of DNA (unknowns or standards) and a 2 µg/ml solution of ethidium bromide were mixed and 10 µl of each mixture was spotted on a petri dish. The spots were illuminated with UV light, and photographed. The fluorescence intensity of each unknown was compared to the fluorescence intensities of the DNA standards.

Primer and probe design for quantitative PCR. *Epulopiscium* type B *dnaA* (GenBank Accession Number EF127641) and *recA* (EF127642) genes were cloned using an approach previously used to clone *ftsZ* (3). TaqMan probes and primer sequences for *Epulopiscium dnaA* (probe: 5′-TTCTTTCTTTTCCGGCGATAAATTGAATATCATCTATTAG, forward primer: 5′-GACC AACCTCCTGCCTTCAGAAATAAA, reverse primer: 5′ -GTGTTAAATGTATGGAAAAATTCTTCTTGT), *ftsZ* (probe: 5′-CACAGGTGACTCGACACTTGCAATC, forward primer: 5′-ATTAAAGGTGCAGGTGGCGTACT, reverse primer: 5′-GCTAGCTCCCGCAT TCAACT), *recA* (probe: 5′-TTAAGAATAAAATTGCTCCTCCATTTAAACAAGCAG, forward primer: 5′-TGCGACAGAAATAATCGGCAGCAAAA, reverse primer: 5′-AGAAGAAATTCCTTCGCCATAGATAA) and the 16S rRNA gene (probe: 5′-CCATGCCGCC TACACACCCTTTACA, forward primer: 5′-ATTAAAGGTGCAGGTGGCGTACT, reverse primer: 5′-GCTAGCTCCCGCATTCAACT) used in real-time PCR were designed using Primer Express (Applied Biosystems Inc., ABI). TaqMan probes used in the real-time PCR assays were 5′-end-labeled with the reporter dye 6-carboxyfluorescein (FAM) and 3′-labeled with the quencher dye carboxytetramethylrhodamine (TAMRA).

TaqMan quantitative PCR assays. Primer sets were tested in standard PCR amplifications using *Epulopiscium* cells, microbial community DNA extracted from surgeonfish intestinal contents and genomic DNA from 5 other Firmicutes. A

proteinase K solution (as above) was irradiated for 3 min using a UV transilluminator (FisherBiotech) and then aliquoted into each tube in a 96-well PCR plate. One washed *Epulopiscium* was added to each tube and incubated at 50°C for 1 h. The plate was heated, 95°C for 15 min, to inactivate the proteinase K. TaqMan assay reaction mixtures were prepared on ice and contained 1X TaqMan Universal Master Mix (ABI), 900 nM of the appropriate forward and reverse primers, and 200 nM of the appropriate fluorogenic probe. To determine the copy number of genes in *Epulopiscium* genomic DNA, the DNA extracts from large- and small-cell populations were diluted 1:10 and 1:100 in TE. The concentrations of all dilutions were determined using the PicoGreen assay described above, except 1 µl of diluted *Epulopiscium* DNA was used in the assays. The small-cell DNA concentration was adjusted to 156 pg/µl, to match the concentration of the large-cell DNA sample. For gene quantification, 1 µl of DNA was added to PCR plate well containing the TaqMan reaction mixture. Each genomic DNA sample was run in triplicate for each gene assay. Quantification standards were generated from serial dilutions of plasmid DNA (2.0×10^7 copies to 2.0×10^2 copies per µl) containing *dnaA*, *ftsZ*, *recA* or the 16S rRNA gene cloned from *Epulopiscium* type B. Standards were run in duplicate and no template controls were run in triplicate for each assay. Thermal cycling conditions were as follows: 2 min at 50°C, 10 min at 95°C followed by 40 cycles of 15 s at 95°C and 1 min at 60°C. Data were compiled and analyzed using Sequence Detection Software, v1.3 (ABI).

Cell volume and gene copy number comparisons. Images of *Epulopiscium* cells were acquired using a Cooke SensiCam CCD Camera and an Olympus BX61

microscope equipped with a LCPlanF1 40X objective. Using Slidebook Software (calibrated with a stage micrometer), the cell length and width were determined. The cell volume was calculated using the formula for a prolate ellipsoid. After image acquisition, each cell was transferred into a well of a 96-well plate and processed for quantitative PCR as described above. Statistical analyses (*t*-test and ANOVA) were performed using SAS software, Version 9.1 of the SAS System for Windows.

Single-cell amplification of the rRNA operon ITS, *ftsZ*, *recA* and *dnaA*. A single, washed *Epulopiscium* type B cell was deposited in a PCR tube and lysed as described (*Materials and Methods*), except after proteinase K inactivation, tubes were filled with 1X HotStarTaq Master Mix (Qiagen) and the appropriate gene primer pair. The following primers were used for PCR amplification of *Epulopiscium* rRNA operon ITSs (1435F: 5′- TCAGTGACCTAACCG, 242R: 5′- TTCGCTCGCCRCTACT) (1) or single-copy genes *ftsZ* (ftsZF: 5′-GGCGGAGGAAATAATGCCGTTGAT, ftsZR: 5′-AGAGTTGAATGCGGGAGCTAGCTT), *dnaA* (dnaAF: 5′-TTTCTCTATGGAGGGGTTGG, dnaAR: 5′-AGGCAAATAGCAATGTATCTA) and *recA* (recAF: 5′-AGGCTCCATTATGAAACTTGGAGA, recAR: 5′-GGACAAGGCCGAGAGAATGCTAAA). Thermal cycling conditions were as follows: 15 min at 95°C (only used with enzymes that required heat activation) followed by 28 cycles of 30 s at 94°C, 30 s at 54°C, 90 s at 72°C, with a final incubation for 10 min at 72°C. Amplifications were performed using either HotStarTaq Master Mix (Qiagen) or PCR SuperMix High Fidelity (Invitrogen) at the concentrations recommended by the manufacturers. The PCR products from amplification of a single-copy gene from a single cell were cloned using the TOPO

TA Cloning kit (Invitrogen) as per manufacturer's instructions. Clones were randomly selected and screened for the presence of a PCR product of the expected size. For each single-copy gene, the sequence of 30 individual clones was determined using Big Dye chemistry and an Applied Biosystems Automated 3730 DNA Analyzer. Alternatively, PCR products were purified (QIAquick PCR Purification Kit, Qiagen) and sequences were determined as described above.

Results and Discussion

DNA content of large and small *Epulopiscium* cells.

Currently, no *Epulopiscium* sp. is available in culture, which prevents the use of standard methods (31) for assessing the composition or conformation of the genome. Based on 16S rRNA gene sequence surveys, *Epulopiscium* type B populations are the most homogeneous of the characterized *Epulopiscium* morphotypes (2), and therefore well suited for the gene-based studies presented here. By quantifying the amount of DNA extracted from 5000 purified large cells with mature offspring (Figure 2.1 f), and 5000 cells with small, immature offspring (Figure 2.1 e), we found that large cells contain upwards of 250 pg of DNA, while small cells contain about 85 pg of DNA. In contrast, a human diploid cell contains 6 pg of DNA. The tremendous amount of DNA in *Epulopiscium* is most likely in one of three conformations: 1) a few copies of an enormous genome (12), 2) thousands of copies of a small genome (51, 11), or 3) many copies of the complete genome but with portions that are highly amplified. Since only a small proportion of the mother-cell DNA is partitioned into newly formed offspring, it is unlikely that *Epulopiscium* cells possess only a few copies of an enormous genome (4). Instead, we hypothesized that

Epulopiscium has a genome comparable to the size of other bacterial genomes (36), and that it is present in very high numbers.

Ploidy of large and small *Epulopiscium* cells.

The composition of DNA in individual *Epulopiscium* cells was assessed using real-time quantitative PCR. Four genes were assayed: *ftsZ, dnaA, recA* and the 16S rRNA gene. The first three of these are generally unlinked, single-copy genes (7, 22, 33) and thus were used to represent the unit genome of *Epulopiscium.* Large *Epulopiscium* cells on average possess 50 to 120 thousand copies of each of these markers (Table 2.1). The genomes of many Firmicutes (low G+C Gram-positive bacteria) have multiple (as many as 15) rRNA operons (49). The ribosomal RNA operon is preferentially amplified in some eukaryotes (42, 64, 67). Ribosomal RNA synthesis is the rate-limiting step in assembly of ribosomes in many bacteria (56). For these reasons, we considered the rRNA operon a good candidate for gene amplification in *Epulopiscium*. Real-time PCR assays of the 16S rRNA gene showed that large *Epulopiscium* cells have 240 to 740 thousand copies of this gene (Table 2.1). Single-cell PCR amplification of the internal transcribed spacer (ITS) between the 16S and 23S rRNA genes showed that these cells have at least four unique ITSs, indicative of multiple rRNA operons (Appendix A, Figure A.1). These results suggest that *Epulopiscium* type B also has multiple rRNA operons per genome, however the rRNA gene copy number in individuals is not large enough, relative to single-copy markers, to indicate substantial amplification of this operon.

It is difficult to determine from single-cell data if all single-copy markers are

Table 2.1. Gene Copy Number in Individual *Epulopiscium*

Gene	Copy Number*	Range†	Control‡
ftsZ	80,600	35,800 - 198,000	240
dnaA	48,700	29,800 - 153,000	240
recA	120,000	60,300 - 20,5000	70
16S	368,000	241,000 - 737,000	310

*Median gene copy number, n = 10

†Copy number range (minimum – maximum)

‡Mean negative controls, surgeonfish gut contents without *Epulopiscium* cells, n = 3

equally present in *Epulopiscium* type B because cell size and gene copy number varied greatly in individuals from a given population. Additionally, using cells instead of purified genomic DNA in these real-time PCR assays could introduce factors that may alter amplification efficiency. We therefore assayed relative gene copy numbers using purified *Epulopiscium* genomic DNA. This approach also allowed for a rough estimation of genome size. Gene copy numbers of the three single-copy markers were statistically similar in genomic DNA extracted from large cells (one-way ANOVA, $F = 1.25$, $P = 0.301$), with markers averaging 40,900 copies in 156 pg of DNA (Table 2.2). These results support the idea that large *Epulopiscium* cells contain tens of thousands of copies of a fully replicated, approximately 3.8 Mb genome. However, the marker numbers obtained from small-cell genomic DNA varied ($F = 11.82$, $P = 0.000$). These small cells were taken at an early stage of their growth cycle, which is presumably a time of active DNA replication. This supposition is supported by the observation that the replication origin-linked marker *dnaA* (47) was more numerous than *ftsZ* or *recA*.

Cell volume per genome.

Bacillus subtilis (also a member of the Firmicutes) maintains a fairly constant cell-volume-to-genomic-DNA ratio over a variety of exponential growth rates (60). We took advantage of the natural variation in cell size in *Epulopiscium* populations to investigate whether genome copy number varied with respect to cytoplasmic volume, and whether this ratio was similar to that of *B. subtilis*. The genome copy number proxy *ftsZ* was assayed in individuals taken from two natural *Epulopiscium* type B populations that represent the extremes of offspring development and cell size

Table 2.2. Gene Copy Number in 156 pg of *Epulopiscium* Genomic DNA

Gene	Large-cell DNA *	Small-cell DNA†	Control DNA‡
ftsZ	42,600 ± 5,820	31,900 ± 1,640	290
dnaA	38,200 ± 7,940	36,500 ± 7,000	120
recA	41,900 ± 8,330	26,300 ± 4,300	30
16S	285,000 ± 36,000	177,000 ± 11,200	40

*Mean ± SD, n = 10

†Mean negative controls, 150 pg of DNA extracted from surgeonfish gut contents with no *Epulopiscium* cells, n = 3

(examples shown in Figure 2.1e and Figure 2.1f). A linear relationship between cell size and copy number was observed in the two populations (Figure 2.2). Cells maintained an average ratio of 1 genome for every 1.9 μm^3 of cytoplasm over a range of volumes from 28,200 to 436,000 μm^3. When analyzed separately, the means of ratios for cells of the two *Epulopiscium* populations were not significantly different (*t*-test, $P = 0.481$). In comparison, *B. subtilis* in exponential growth in rich media harbors 1 chromosome per 0.7 μm^3 of cytoplasm (based on data from Sharpe *et al.* 1998). *Epulopiscium* cells appear to follow the same rules as other bacteria that link cell growth and DNA replication (60, 18, 19), however at least in the populations we compared, *Epulopiscium* maintains a larger cytoplasmic volume to genome ratio than *B. subtilis*. Further studies are needed to determine whether all cytoplasmic regions of these large bacterial cells are functionally equivalent.

Factors leading to extreme polyploidy.

The significance of the substantial polyploidy and genetic expansion recorded for *Epulopiscium* may be considered with respect to the two dominant attributes of this bacterium, its nutritional biology as a symbiont and its extraordinarily large size. Genetic amplification is seen in diverse organisms: bacteria, archaea, protists, as well as fungi, plant, and animal cells, and provides resources to support rapid growth and division, cell specialization and adaptation, and may enhance repair of genetic lesions (13, 14, 25, 38, 50 - 53). For some bacteria, polyploidy is allied with metabolic adaptation and symbiont differentiation. *Buchnera aphidicola*, the proteobacterial symbiont of aphids, contains on average 120 copies of its chromosome, and expansion of genetic resources correlates with host development and presumably increased host

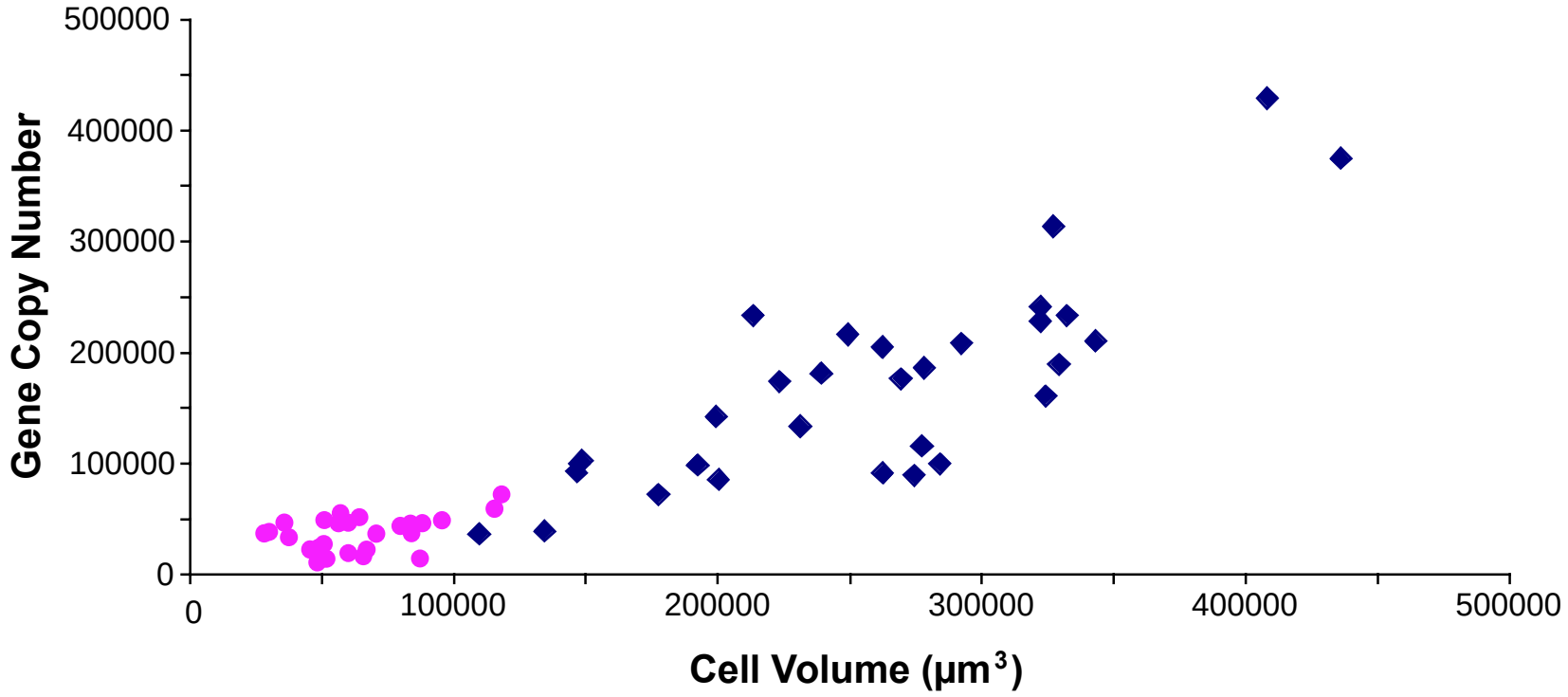


Figure 2.2. The relationship between gene copy number and cytoplasmic volume in *Epulopiscium* sp. type B cells. The pink circles represent data from individuals of the small-cell population (n = 26), and the blue diamonds represent data from the large-cell population (n = 31)

demand for the essential biomolecules provided by *Buchnera* (35). During differentiation to symbiotic bacteroids, some species of rhizobia undergo modest genome proliferation (43). Therefore, one factor that may have led to polyploidy in *Epulopiscium* is selection pressure for the evolution of a symbiotic relationship that contributes to host metabolism.

Alternatively, the extreme polyploidy in *Epulopiscium* may be more tightly linked with cell size. In insect and plant cells, an expansion of genomic resources is accompanied with an increase in size (43, 29, 39, 55), though size is not always proportionate to ploidy (63). Although the advantage of large cell size in *Epulopiscium* spp. has yet to be determined, two features of the symbiont's environment should be considered. The host feeding behavior varies substantially on a daily basis, with concentrations of food occurring at different points in the relatively large alimentary tract. Populations of large *Epulopiscium* cells migrate to distinct regions of the intestinal tract at different times of day (24) presumably in response to host digestive processes. We speculate that large size allows these motile cells better control over their position along the length of the intestinal tract of the host. In addition, the alimentary microbial community of the host is complex and supports very high numbers of ciliate bacterial predators (26). *Epulopiscium* type B cells appear to avoid predation by all but the largest ciliates that inhabit the *N. tonganus* intestinal tract (27).

The biased distribution of DNA within the cytoplasm of *Epulopiscium* permits functional compartmentalization and regional specialization within these large cells. A similar peripheral arrangement of nucleoids has been reported in another large

bacterium, *Thiomargarita namibiensis* (58), although the composition of these nucleoids has yet to be determined. Single spherical cells of *Thiomargarita* are generally 100 – 300 µm in diameter but cells as large as 800 µm occur. Each *Thiomargarita* cell contains a large, fluid-filled vacuole, which takes up approximately 98% of the cell volume. This confines the active cytoplasm to a shallow, 0.5 – 2 µm layer just under the cytoplasmic membrane. As with all bacteria, the close association of DNA with the cell membrane accommodates transertion (46), the linked processes of transcription, translation and insertion of proteins into the cytoplasmic membrane. In large bacteria, genomic copies arrayed around the cellular periphery would permit transcription of any gene at disparate locations within the cell, thus reducing transit time of proteins and metabolites from site of synthesis or entry to site of action. In this way, a large bacterium could function like a microcolony, with different regions of the cell independently responding to local stimuli, which would alleviate some of the pressure to remain small for the sake of rapid intracellular diffusive transport. In stark contrast to *Thiomargarita*, the central cytoplasm of *Epulopiscium* is relatively free of DNA but apparently active. This subcellular arrangement allows for the rapid growth of internal offspring, as seen in the related, but smaller (12 -35 µm long), endospore-forming bacterium *Metabacterium polyspora* (5). It still remains to be seen if other mechanisms in large *Epulopiscium* cells enhance movement of molecules throughout the bulk of the central cytoplasm.

Polyploidy and intracellular genetic diversity.

It is coming to light that genome duplication and subsequent divergence of orthologs has been an important driving force in genome evolution and the generation

of morphological complexity (39, 32). Polyploidy in *Epulopiscium* could allow for diversification of genome copies while supporting an increase in cell size. Such a simple cellular modification may have been an important advance toward the development of the contemporary eukaryotic cell. In a previous study, we estimated that approximately 1% of the DNA in an *Epulopiscium* sp. type B cell is inherited (4). For an average type B cell, this amount of DNA could comprise 230 genome equivalents. Although reproduction imposes a significant genomic population bottleneck, some genetic diversity could be passed on to the offspring.

To evaluate gene diversity in *Epulopiscium*, we followed an approach used by Mark Welch and Meselson to study bdelloid rotifers (40, 41). There is mounting evidence for functional divergence of genes that were once alleles in these asexual, but highly successful, rotifers (48). For *Epulopiscium* gene surveys, we cloned the PCR products from single-cell amplification reactions targeting single-copy genes *ftsZ*, *dnaA* or *recA* (described in Appendix A). As mentioned above, multiple ITS sequences have been recovered from an *Epulopiscium* single-cell PCR amplification and so this method should allow for the recovery of abundant variants in single-copy genes, if present. Alignments of single-copy gene clones revealed that a consensus gene sequence was predominant in each library, however we did detect sequence variants (Appendix A Tables A.1-A.3). For *ftsZ* and *recA*, clones varied from the consensus by one or two nucleotides and all variants were unique. Single nucleotide changes appeared to be transitions and the vast majority coded for non-synonymous amino acid substitutions. The overall frequency of nucleotide differences was comparable to published *Taq* DNA polymerase PCR error rates (20, 21, 23). For *dnaA*, we observed similar trends.

During our analysis of *dnaA*, however, we were surprised to find a common, single-nucleotide deletion in 12 of the 30 *dnaA* clones. The deletion occurred within a mononucleotide tract of ten adenines. Further investigations indicate that at least some of these deletion variants are produced during amplification with thermostable polymerases (Appendix A, Figure A.2). A high frequency of slippage in mononucleotide tracts of this length during PCR amplification has been observed by others (16, 61). Long, mononucleotide repeats are common in eukaryotic genomes, but rare in bacterial or archaeal genomes (1, 44, 45). In bacteria, these highly mutable motifs tend to be found within genes that encode variable surface proteins. This is the first instance in which a long (10 bp) mononucleotide repeat has been found within an essential bacterial gene. The functional significance of this *dnaA* adenine-deletion variant and frequency of its expression have yet to be determined.

Evolutionary implications of extreme polyploidy in bacteria.

Together, the findings support the idea that *Epulopiscium* sp. type B cells are highly polyploid throughout their life cycle. Gene sequence surveys suggest that the bottleneck imposed on *Epulopiscium* genomes during reproduction may restrict diversification of orthologs in an individual. Nevertheless, extreme polyploidy may allow *Epulopiscium* to harbor unstable genetic features, such as mononucleotide tracts, within essential genes, without detriment.

The functional dichotomy of "somatic" and "germline" genomes within an enormous and highly polyploid bacterium represents a novel intermediate between the "typical" asexual, single-celled microorganism and a multicellular organism. The genetic material of a successful, solitary bacterium is replicated and faithfully passed

on to its offspring. For multicellular organisms, only the germline may be inherited; the waste of genetic resources in somatic cells is offset by the diversification of cellular function, which commonly leads to increased size, enhanced access to resources, and improved metabolic capacity. Compared to a solitary existence, colonial microbes (e.g. some actinobacteria) or populations of cooperative microorganisms (e.g. cellular slime molds, myxobacteria) benefit from improved metabolic potential and perhaps better dispersal at the cost of the genomes of cells that play a supporting role (66, 34). The enormous, polyploid *Epulopiscium* cell has converged on the advantages of social microbes but with additional benefits (exceptional motility, enhanced resistance to predation) normally found in large eukaryotic microbes or multicellular organisms.

Acknowledgements

We thank the staff and directors of the Lizard Island Research Station for their support and Will Robbins and David Raubenheimer for their help in the field. We also thank Rebekah Ward, David Miller, John Helmann and two anonymous reviewers for comments on the manuscript. Finally, we thank Simona Despa for statistical consultation. Funding was provided by the National Science Foundation Molecular and Cellular Biology Program Grant MCB 0721583.

References

1. **Ackermann, M. & Chao, L.** 2006. DNA sequences shaped by selection for stability. PLoS Genet. **2**be22.

2. **Angert, E. R., Clements, K. D. & Pace, N. R.** 1993. The largest bacterium. Nature. **362:** 239-241.

3. **Angert, E. R. & Losick, R. M.** 1998.Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. Proc. Natl. Acad. Sci. U S A. **95:** 10218-10223.

4. **Angert, E. R. & Clements, K. D.** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:** 827-835.

5. **Angert, E. R.** 2005. Alternatives to binary fission in bacteria. Nat. Rev. Microbiol. **3:** 214-224.

6. **Angert, E. R.** 2006. The enigmatic cytoarchitechture of *Epulopiscium* spp. in Microbiology Monographs, ed. Shively, J. M. (Springer-Verlag, Berlin Heidelberg). **2:** 285-301.

7. **Baldo, L., Bordenstein, S., Wernegreen, J. J. & Werren, J. H.** 2006. Widespread recombination throughout *Wolbachia* genomes. Mol. Biol. Evol. **23:** 437-449.

8. **Berg, H. C.** 1993. Random Walks in Biology, Princeton University Press, Princeton, NJ.

9. **Blakemore, R. P. & Canale-Parola, E.** 1973. Morphological and ecological characteristics of *Spirochaeta plicatilis*. Arch. Mikrobiol. **89:** 273-289.

10. **Bonner, J. T.** 2004. Perspective: the size-complexity rule. Evolution Int. J. Org. Evolution. **58:**1883-1890.

11. **Bresler, V. & Fishelson, L.** 2003. Polyploidy and polyteny in the gigantic eubacterium *Epulopiscium fishelsoni*. Mar. Biol. **143:** 17-21.

12. **Bresler, V., Montgomery, W. L., Fishelson, L. & Pollak, P. E.** 1998. Gigantism in a bacterium, *Epulopiscium fishelsoni*, correlates with complex patterns in arrangement, quantity, and segregation of DNA. J. Bacteriol. **180:** 5601-5611

13. **Brown, D. D.** 1981. Gene expression in eukaryotes. Science. **211:** 667-674.

14. **Calvi, B. R., Lilly, M. A. & Spradling, A. C.** 1998. Cell cycle control of chorion gene amplification. Genes. Dev. **12:** 734-744.

15. **Clements, K. D., Sutton, D. C. & Choat, J. H.** 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfishes *Acanthuridae* of the Great Barrier Reef Australia. Marine Biol. **102:** 403-412.

16. **Clarke, L. A., Rebelo, C. S., Goncalves, J., Boavida, M. G. & Jordan, P.** 2001. PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences. Mol. Pathol. **54:** 351-353.

17. **Delaporta, B.** 1964. Etude descriptive de bacteriaies de tres grandes dimensions. Annales de l'Institut Pasteur. **107:** 845-862.

18. **Donachie, W. D.** 1968. Relationship between cell size and time of initiation of DNA replication. Nature. **219:** 1077-1079.

19. **Donachie, W. D. & Blakely, G. W.** 2003. Coupling the initiation of chromosome replication to cell size in *Escherichia coli*. Curr. Opin. Microbiol. **6:** 146-150.

20. **Dunning, A. M., Talmud, P. & Humphries, S. E.** 1988. Errors in the polymerase chain reaction. Nucleic Acids Res. **16:** 10393.

21. **Eckert, K. A. & Kunkel, T. A.** 1991. DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl. **1:** 17-24.

22. **Eisen, J. A.** 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J. Mol. Evol. **41:** 1105-1123.

23. **Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P.** 1990. Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc Natl Acad Sci U S A **87:** 2833-2837.

24. **Fishelson, L., Montgomery, W. L. & Myrberg, A. A.** 1985. A unique symbiosis in the gut of a tropical herbivorous surgeonfish (*Acanthuridae*: Teleostei) from the Red Sea. Science. **229:** 49-51.

25. **Foster, P. L.** 2004. Adaptive mutation in *Escherichia coli*. J. Bacteriol. **186:** 4846-4852.

26. **Grim, J. N.** 1993. Description of somatic kineties and vestibular organization

of *Balantidium jocularum* sp. n., and possible implications for the class Litostomatea and genus *Balantidium.* Acta. Protozool. **32:** 37-45.

27. **Grim, J. N.** 2006 Food vacuole contents in the ciliate, *Balantidium jocularum* (Balantididae), a symbiont in the intestine of the surgeonfish, *Naso tonganus (Acanthuridae*). J Eukaryot. Microbiol. **53:** 269-274.

28. **Gundersen, J. K., Jorgensen, B. B., Larsen, E. & Jannasch, H. W.** 1992. Mats of giant sulphur bacteria on deep-sea sediments due to fluctuating hydrothermal flow. Nature. **360:** 454-456.

29. **Hanson, K. K., Kelley, A. C. & Bienz, M.** 2005. Loss of *Drosophila borealin* causes polyploidy, delayed apoptosis and abnormal tissue development. Development. **132:** 4777-4787.

30. **Head, I. M., Gray, N. D., Babenzien, H. & Glöckner, F. O.** 2000. Uncultured giant sulfur bacteria of the genus *Achromatium*. FEMS Microbiol.Ecol. **33:** 171-180.

31. **Herschleb, J., Ananiev, G. & Schwartz, D. C.** 2007. Pulsed-field gel electrophoresis. Nat. Protoc. **2:** 677-684.

32. **Infante, J. J., Dombek, K. M., Rebordinos, L., Cantoral, J. M. & Young, E. T.** 2003. Genome-wide amplifications caused by chromosomal rearrangements play a major role in the adaptive evolution of natural yeast. Genetics. **165:** 1745-1759.

33. **Karlin, S., Weinstock, G. M. & Brendel, V.** 1995. Bacterial classifications derived from recA protein sequence comparisons. J. Bacteriol. **177:** 6881-6893.

34. **Keller, L. & Surette, M. G.** 2006. Communication in bacteria: an ecological and evolutionary perspective. Nat. Rev. Microbiol. **4:** 249-258.

35. **Komaki, K. & Ishikawa, H.** 2000. Genomic copy number of intracellular bacterial symbionts of aphids varies in response to developmental stage and morph of their host. Insect Biochem. Mol. Biol. **30:** 253-258.

36. **Liolios, K., Tavernarakis, N., Hugenholtz, P. & Kyrpides, N. C.** 2006. The Genomes On Line Database (GOLD) v.2: a monitor of genome projects worldwide. Nucleic Acids Res. **34:** D332-334.

37. **Losick, R. & Shapiro, L.** 1999. Changing views on the nature of the bacterial cell: from biochemistry to cytology. J. Bacteriol. **181:** 4143-4145.

38. **Makarova, K. S., Aravind, L., Wolf, Y. I., Tatusov, R. L., Minton, K. W., Koonin, E. V. & Daly, M. J.** 2001. Genome of the extremely radiation-resistant bacterium *Deinococcus radiodurans* viewed from the perspective of comparative genomics. Microbiol. Mol. Biol. Rev. **65:** 44-79.

39. **Maere, S., De Bodt, S., Raes, J., Casneuf, T., Van Montagu, M., Kuiper, M. & Van de Peer, Y.** 2005. Modeling gene and genome duplications in eukaryotes. Proc. Natl. Acad. Sci. U S A .**102:** 5454-5459.

40. **Mark Welch, D. & Meselson, M.** 2000. Evidence for the evolution of bdelloid rotifers without sexual reproduction or genetic exchange. Science. **288:** 1211-1215.

41. **Mark Welch, D. & Meselson, M. S.** 2001. Rates of nucleotide substitution in sexual and anciently asexual rotifers. Proc. Natl. Acad. Sci. U S A. **98:** 6720-6724.

42. **McGrath, C. L. & Katz, L. A.** 2004. Genome diversity in microbial eukaryotes. Trends Ecol. Evol. **19:** 32-38.

43. **Mergaert, P., Uchiumi, T., Alunni, B., Evanno, G., Cheron, A., Catrice, O., Mausset, A. E., Barloy-Hubler, F., Galibert, F., Kondorosi, A. & Kondorosi, E.** 2006. Eukaryotic control on bacterial cell cycle and differentiation in the *Rhizobium*-legume symbiosis. Proc. Natl. Acad. Sci. U S A. **103**: 5230-5235.

44. **Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E.** 1994. Adaptive evolution of highly mutable loci in pathogenic bacteria. Curr. Biol. **4:** 24-33.

45. **Mrázek, J., Guo, X. & Shah, A.** 2007. Simple sequence repeats in prokaryotic genomes. Proc. Natl. Acad. Sci. U S A **104:** 8472-8477.

46. **Norris, V., Turnock, G. & Sigee, D.** 1996. The *Escherichia coli* enzoskeleton. Mol. Microbiol. **19:** 197-204.

47. **Ogasawara, N. & Yoshikawa, H.** 1992. Genes and their organization in the replication origin region of the bacterial chromosome. Mol. Microbiol. **6:** 629-634.

48. **Pouchkina-Stantcheva, N. N., McGee, B. M., Boschetti, C., Tolleter, D., Chakrabortee, S., Popova, A. V., Meersman, F., Macherel, D., Hincha, D. K. & Tunnacliffe, A.** 2007. Functional divergence of former alleles in an ancient asexual invertebrate. Science **318:** 268-271.

49. **Rainey, F. A., Ward-Rainey, N. L., Janssen, P. H., Hippe, H. & Stackebrandt, E.** 1996. *Clostridium paradoxum* DSM 7308T contains multiple 16S rRNA genes with heterogeneous intervening sequences. Microbiology **142:** 2087-2095.

50. **Reams, A. B. & Neidle, E. L.** 2003. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Mol. Microbiol. **47:** 1291-1304.

51. **Robinow, C. & Angert, E. R.** 1998. Nucleoids and coated vesicles of "*Epulopiscium*" spp. Arch. Microbiol. **170:** 227-235.

52. **Romero, D. & Palacios, R.** 1997. Gene amplification and genomic plasticity in prokaryotes. Annu. Rev. Genet. **31:** 91-111.

53. **Rosenberg, S. M. & Hastings, P. J.** 2004. Adaptive point mutation and adaptive amplification pathways in the *Escherichia coli* Lac system: stress responses producing genetic change. J. Bacteriol. **186:** 4838-4843.

54. **Sambrook, J., Fritsch, E. F. & Maniatis, T.** 1989. Molecular Cloning: A Laboratory Manual (Cold Spring Harbor Laboratory Press).

55. **Saucedo, L. J. & Edgar, B. A.** 2002. Why size matters: altering cell size. Curr. Opin. Genet. Dev. **12:** 565-571.

56. **Schneider, D. A., Ross, W. & Gourse, R. L.** 2003. Control of rRNA expression in *Escherichia coli*. Curr. Opin. Microbiol. **6:** 151-156.

57. **Schulz, H. N., Brinkhoff, T., Ferdelman, T. G., Marine, M. H., Teske, A. & Jorgensen, B. B.** 1999. Dense populations of a giant sulfur bacterium in Namibian shelf sediments. Science. **284:** 493-495.

58. **Schulz, H. N.** 2006. The genus *Thiomargarita* in The Prokaryotes, **6:** 1156-1163.

59. **Shapiro, L. & Losick, R.** 1997. Protein localization and cell fate in bacteria. Science **276:** 712-718.

60. **Sharpe, M. E., Hauser, P. M., Sharpe, R. G. & Errington, J.** 1998. *Bacillus subtilis* cell cycle as studied by fluorescence microscopy: constancy of cell length at initiation of DNA replication and evidence for active nucleoid partitioning. J. Bacteriol. **180:** 547-555.

61. **Shinde, D., Lai, Y., Sun, F. & Arnheim, N.** 2003. Taq DNA polymerase slippage mutation rates measured by PCR and quasi-likelihood analysis: (CA/GT)n and (A/T)n microsatellites. Nucleic Acids Res **31:** 974-980.

62. **Shively, J. M.** 2006. Prokaryote complex intracellular structures: descriptions and discoveries. in Microbiology Monographs (Springer-Verlag, Berlin Heidelberg). **2:** 3-22.

63. **Sugimoto-Shirasu, K. & Roberts, K**. 2003. "Big it up": endoreduplication and cell-size control in plants. Curr. Opin. Plant Biol. **6:** 544-553.

64. **Tower, J.** 2004. Developmental gene amplification and origin regulation. Annu. Rev. Genet. **38:** 273-304.

65. **West, G. B. & Brown, J. H.** 2005. The origin of allometric scaling laws in biology from genomes to ecosystems: towards a quantitative unifying theory of biological structure and organization. J. Exp. Biol. **208:** 1575-1592.

66. **West, S. A., Griffin, A. S., Gardner, A. & Diggle, S. P.** 2006. Social evolution theory for microorganisms. Nat. Rev. Microbiol. **4:** 597-607.

67. **Zhang, Z., Macalpine, D. M. & Kapler, G. M.** 1997. Developmental regulation of DNA replication: replication fork barriers and programmed gene amplification in Tetrahymena thermophila. Mol. Cell. Biol. **17:** 6147-6156.

CHAPTER 3

BACTERIAL DIVERSITY IN THE GASTROINTESTINAL TRACT OF THE OMNIVOROUS MARINE FISH, *NASO TONGANUS*, AS ASSESSED BY 16S rRNA GENE CLONE LIBRARIES AND TERMINAL –RESTRICTION FRAGMENT LENGTH POLYMORPHISM

Abstract

The intestinal microbial communities of tropical marine fishes are largely uncharacterized despite their apparent importance in the digestion of complex polysaccharides found in algae. We utilized a 16S ribosomal RNA (rRNA) gene clone library to describe the bacterial diversity in the intestinal tract of the marine fish, *Naso tonganus*. A total of 104 clones were sequenced from four segments of the intestinal tract from four individuals. The segment most proximal to the stomach (segment II) was dominated by organisms most closely related to the *Cyanobacteria*. The more distal segments (segments IV and V) were dominated by phyla known for specialization in fermentation (*Firmicutes* and *Fusobacteria*). T-RFLP analyses of the 16S rRNA gene were used to assess between-fish and between-segment differences in the bacterial communities in segments IV and V. Statistical differences were not observed between segments and SIMPER analysis showed a general trend of low similarity between fish. However, multi-dimensional scaling (MDS) plots did show clustering of bacterial populations of the same gut segment from different individuals. In order to assess how the unusually large bacterial symbiont *Epulopiscium* influences the microbial community, similar analyses were carried out using an individual fish harboring this organism. Global R statistics indicate that the communities in fish with

and without *Epulopiscium* are well separated from each other. This result suggests that *Epulopiscium* may influence the community composition, specifically replacing some of the numerically dominant, fermentative groups.

Introduction

Intestinal microbiota play an important role in digestion for a wide variety of terrestrial animals (6, 20, 28, 37, 40, 41, 42, 47, 58, 62, 66). Among their many contributions to host physiology, these microbial consortia produce enzymes responsible for the breakdown of macromolecules, including complex carbohydrates (49). The resulting molecules then serve as substrates for microbial fermentation. This process provides the host with fermentation end products, such as short-chain fatty acids (SCFA), that are easily absorbed and are useful sources of energy for the host (33). Much work has been done to understand the functional relationship of these complex communities in mammals, including humans (41, 42, 61). However, this understanding is still limited particularly in marine fishes (8, 13, 14, 17).

Herbivorous and omnivorous marine fishes consume a wide variety of marine plants, including algae and sea grasses (64). Marine ecosystems differ from terrestrial ecosystems in the abundance of cellulose. While cellulose is not absent in the cell walls of marine plants, it is only one of a diverse group of structural and matrix polysaccharides, simple sugars and sugar alcohols found in these plants (35). Because of this, the marine fishes that consume these plants are confronted with a wide variety of potentially fermentable carbohydrates (33).

There are several lines of evidence that indicate the importance of intestinal symbionts in the digestion of algae consumed by fishes. Direct microscopic

examination has revealed abundant and complex microbial communities in the hindgut of several species of herbivorous fishes (10, 12, 25, 51). Short-chain fatty acid (SCFA) analysis of intestinal contents has demonstrated nutritionally significant levels of microbial fermentation end products, such as acetate, occurring in the hindgut of a variety of marine fish species (7, 51). Rates of SCFA production in three species of temperate marine herbivorous fish were shown to be comparable to those found in the intestines of herbivorous reptiles and small mammals (51). More recently, molecular characterization of the microbial consortia in the hindgut of three temperate marine herbivorous fishes has demonstrated dominance by bacteria related to the genera *Clostridium* and *Eubacterium* (15). Both of these genera are prevalent in the gastrointestinal (GI) tracts of ruminants and other animals where they may play an important role in the degradation and detoxification of foodstuffs (1, 22, 46). Similarly, the intestinal microbiota of fishes may be responsible for the detoxification of algal secondary metabolites such as phlorotannins and alkaloids thereby allowing the host to diversify their nutritional niche and permitting dietary flexibility (17).

A description of the GI microbiota and the distribution of microbes within the GI tract of fishes are central to understanding their role in the digestive strategies of their host. To date, only indirect identification of fermentative microbial communities in tropical marine reef fish (via SCFA analysis) has been demonstrated (13, 51). Direct characterization of intestinal microbiota has been limited to temperate and subtropical species (15, 25, 50). Other studies have focused on the phylogenetic characterization of individual species, such as the surgeonfish symbionts *Epulopiscium* spp., which are

a conspicuous component of the intestinal microbiota of certain surgeonfish (4, 12, 48).

The characterization of these rich communities is complicated by the inability to culture a comprehensive selection of most complex microbial communities using standard methods (3). The goal of this study was to employ culture-independent techniques to describe the bacterial diversity found in the intestinal tract of four individuals of the tropical reef fish *Naso tonganus*. This omnivorous species of fish feeds primarily on algae but also regularly consumes small amounts of detritus and invertebrates (8). It often hosts populations of the enormous bacterium *Epulopiscium* sp. type B (4, 12). In terms of cell volume and genomic resources, an individual *Epulopiscium* cell is equivalent to tens of thousands of bacteria the size of *Escherichia coli* (4, 48). Because of these vast genomic resources, and the potential for a single *Epulopiscium* cell to synthesize an abundant amount of degradative enzymes, we predict that in regions of the intestinal tract where *Epulopiscium* is located, it is a prevailing force in the community.

In this study, reference libraries of 16S rRNA gene clones were developed from total genomic DNA extracted from *N. tonganus* intestinal samples. Through the use of community 16S rRNA gene terminal restriction fragment length polymorphism (T-RFLP) analyses we describe spatial differences in the bacterial communities in two posterior segments of the intestinal tract of these fish. Finally, using T-RFLP analyses we compared bacterial communities of a single fish harboring *Epulopiscium* spp. in its lower GI tract with those that did not, to assess the potential impact of *Epulopiscium* on the composition of resident bacteria.

Materials and Methods

Sample Collection. *Naso tonganus* were collected in October 2003 by spear fishing on reefs around Lizard Island, Australia. The intestinal tract of each fish was removed and divided into 5 sections as described in Clements and Choat (1995). The stomach was designated as segment I. The intestinal tract of *N. tonganus* has no distinct morphological features, so the remaining intestine was partitioned into four equal-length segments referred to as segments II – V. Contents of intestinal segments II - V were removed, frozen and stored at –80°C.

DNA extraction. The gut content samples from four *N. tonganus* were thawed and heated at 65°C for 10 minutes. For each sample, approximately 3 ml was transferred into sterile 15 ml conical tubes, fixed with 80% ethanol and stored at -20°C. Community genomic DNA was extracted, in triplicate, from approximately 350 μl of the ethanol-fixed gut contents. Gut contents were pelleted by centrifugation and washed twice with filter sterilized 10 mM Tris, pH 8. DNA was extracted using the PowerSoil DNA Isolation Kit (Mo Bio Laboratories, Inc.) with the following modifications. Gut contents were pelleted and the pellet was transferred to the Power Bead Tube. Proteinase K in 10 mM Tris, pH 8.0, was added to a final concentration of 100 μg/ml and the samples were incubated at 50°C for one hour. The suspension was subjected to three rounds of freeze-thaw (transferring the tubes between a dry ice and ethanol bath and a 65°C water bath). Beads and buffer were transferred back into the Power Bead Tube and the manufacturer's extraction protocol was completed. Triplicate extractions were pooled and stored at -20°C.

PCR amplification and 16S rRNA gene library construction. The 16S rRNA genes were amplified from genomic DNA from each segment sample using the bacterial specific primer 8F (5´-AGAGTTTGATCCTGGCTCAG-3´) and the universal primer 1492R (5´-GGTTACCTTGTTACGACTT-3´) (Integrated DNA Technologies) (38). Amplifications were performed in triplicate and carried out in 50 μl reactions containing 1X HotStarTaq PCR buffer (Qiagen), 0.2 μmol of each primer, 100 μg/ml bovine serum albumin (BSA) (New England Biolabs) and 2.5 U of *Taq* DNA polymerase. Genomic DNA (1 μl) was added to each amplification reaction. Reaction mixtures were subjected to the following temperature cycling profile on a PTC-200 Thermal Cycler (MJ Research): 15 min at 95°C followed by 29 cycles of denaturation (94°C for 30 sec), annealing (50°C for 30 seconds) and extension (72°C for 2 min). A final extension step of 72°C for 10 minutes was included. A positive control (25 ng of *Bacillus subtilis* DNA) and a no template control were included. Amplification products were analyzed using a 1.2% agarose gel in 1X Tris-Acetic Acid-EDTA stained with ethidium bromide (56). Replicate amplification products (15 μl each) were pooled and cloned into the plasmid vector pCR 2.1-TOPO. Plasmids were transformed into *Escherichia coli* TOP10 competent cells according to manufacturer's instructions (Invitrogen). White colonies were randomly selected and screened by PCR for the appropriate insert using the vector specific primers M13F (5'-GTAAAACGACGGCCAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3'). Plasmid DNA was recovered using the QIAprep Miniprep Kit (Qiagen).

Clone Library Analysis. DNA sequence of each clone was determined with an ABI 3730 DNA Analyzer (Applied Biosystems, Inc.) using Big Dye Terminator chemistry

and AmpliTaq-FS DNA polymerase. Initially, the 8F primer was used in sequencing reactions. Full-length 16S rRNA gene sequence was determined for unique clones using the 1492R primer. Sequence reads were assembled using the Sequencher software package (Gene Codes) and compared to sequences in GenBank by using BLAST (Basic Local Alignment Search Tool) (2) and to sequences in the Classifier Program available at the Ribosomal Database Project II, Release 9.0 (RDP) (18, 45) to identify their closest relatives and approximate taxonomic position. Putative chimeras were identified using the Bellerophon server and ChimeraCheck programs (18, 32) and excluded from further analysis. Full-length 16S rRNA gene sequences were aligned using ClustalW (39) and sequences with 97% or greater similarity to each other were considered as an operational taxonomic unit (OTU). One clone representing each OTU was used in phylogenetic constructions. Diversity indices and richness estimators (Chao1, ACE, Shannon and Simpson) as well as rarefaction curves to estimate coverage were computed on $OTU_{0.03}$, $OTU_{0.05}$, $OTU_{0.2}$ outputs using the open-source software package mothur (57). These analyses were performed on each clone library generated from intestinal segments II, III, IV and V.

Phylogenetic Analysis. Comparative sequence analyses were carried out using the ARB program package (44). Sequences were aligned to reference rRNA gene sequences from the RDP running locally in the ARB sequence environment. The cloned sequences were automatically aligned using the ARB sequence editor, the alignment inspected and corrected manually based on known secondary structures. To avoid potential treeing artifacts caused by hypervariable regions of the 16S rRNA gene, the Lane bacterial mask was applied to exclude these regions in tree

construction. Evolutionary distance, maximum parsimony and maximum-likelihood analysis were performed using ARB. Trees were constructed using the neighbor-joining algorithm. Tree topologies were tested with 1000 bootstrapping replicates.

Terminal Restriction Fragment Length Polymorphism (T-RFLP) Analysis. 16S rRNA genes were amplified from community DNA samples with the bacterial primers 5´ 6-FAM (carboxyfluorescein) 8F and 926R (5´-ACCGCTTGTGCGGGCCC-3´) using the following program: 15 min at 95°C followed by 29 cycles of denaturation (94°C for 30 sec), annealing (50°C for 30 seconds) and extension (72°C for 2 min). A final extension step of 72°C for 10 minutes was included. Amplification reactions for all samples were run in triplicate. A positive control (25 ng of *Bacillus subtilis* DNA) and a no template control were included. A 5 μl aliquot of each 100 μl reaction was analyzed on a 1.2% agarose gel. The remaining 95 μl portions of triplicate reactions were pooled and purified using the QIAquick PCR purification kit (Qiagen). Purified PCR products were quantified with PicoGreen (Invitrogen) using a Perkin Elmer LS50B fluorometer. Restriction digests of 50 ng of PCR product were set up according to manufacturers instructions (New England Biolabs). DNA was digested over night with the restriction enzymes *Rsa*I and *Msp*I. Only two restriction enzymes were used, because the goal was to generate a comparative analysis for assessing changes in community composition between samples and we assumed microbial diversity was too high to resolve through the use of additional enzymes. A total of 2 μl of each sample was mixed with 0.2 μl of internal size standard (GeneScan 500 Liz, Applied Biosystems) and 17.8 μl of deionized formamide. Samples were run in

triplicate, denatured at 94°C for 3 minutes, and immediately placed on ice. The reactions were fractionated on an ABI 3730 DNA Analyzer. PeakScanner software (Applied Biosystems) was used to determine the sizes of the labeled terminal restriction fragments (T-RFs) by comparison with the internal size standard using the Local Southern size calling method and light smoothing. Peaks that were not represented in all three replicates, were less than 50 bases in length, or greater than 500 bases, had a peak amplitude of less than 50 fluorescent units, and comprised less than 1% of the total fluorescence for that sample were excluded from further analysis.

Statistical analysis. T-RFLP data were statistically analyzed with Primer v 5.2.9 software (PRIMER-E Ltd., Plymouth, UK) as described by Rees *et al* (2004). The Bray-Curtis similarity matrix was calculated for area and presence/absence of T-RFLP data, transforming them with the fourth root. Differences in community composition, represented by presence or absence of T-RFs, were visualized using non-metric multi-dimensional scaling (MDS) of the similarity matrix to produce a two-dimensional ordination figure of the samples. Points that cluster together in an MDS plot represent samples that are similar to each other in community composition. The stress of the plot, which measures the goodness of fit, was calculated. A stress value lower than 0.1 indicates an ideal ordination with no prospect of misinterpretation. Non-parametric analysis of similarity (ANOSIM) was used to test whether there were significant differences in the community composition between segments. SIMPER analysis was used to assess which T-RFs were contributing to the average dissimilarity between samples.

Clone Library – T-RF Associations: Predominate phylotypes observed in the T-RFLP electropherograms were identified for Segments IV and V. These T-RFs were linked back to the predicted T-RFs generated through *in silico* digestion of sequences in our clone library from the same segment.

Nucleotide sequence accession numbers: Sequence data have been deposited in GenBank and assigned the following accession numbers: HM630156 - HM630259.

Results

***Naso tonganus* intestinal tract bacteria**

The bacterial communities in segments II, III, IV and V of four individual *N. tonganus* were assessed using 16S rRNA gene clone libraries. Partial rRNA gene sequences were obtained from 68, 37, 58 and 55 clones from the segment II, III, IV, and V libraries, respectively. For this study, an operational taxonomic unit or OTU was defined as sequences from an individual library that were at least 97% identical. Nearly complete gene sequences (~1400 nucleotides) were determined for one representative of each OTU in each library. Two chimeras were identified and excluded from further analysis. Of the 104 full-length sequences remaining, 33 of these came from segment II, 10 from segment III, 32 from segment IV and 29 from segment V (Tables 3.1-3.4).

The one hundred and four bacterial OTUs affiliated with one of eleven bacterial phyla. In segment II, the *Firmicutes* and *Cyanobacteria* dominated the clone library at 40% and 31%, respectively. *Proteobacteria* (13%), *Planctomycetes* (4%), *Bacteroidetes* (4%), *Tenericutes* (4%), *Deferribacteres* 2% and *Verrucomicrobia* (2%) were also present (Figure 3.1a). *Fusobacteria* and *Verrucomicrobia* dominated

Table 3.1: Percent identity values for *Naso tonganus* segment II 16S rRNA gene sequences[a]

OTU	#[b]	Closest related bacterial sequence	% similarity	Closest related bacterial or environmental sequence	% similarity
Bacteroidetes					
NT2_C68	1	*Flavobacteriaceae* bacterium YMS-2 (EF017801)	99.4	Vailulu'u Seamount uncultured bacterium clone VS_CL-411 (FJ497663)	99.7
NT2_C72	1	*Candidatus Amoebophilus asiaticus* (AB506780)	90.3	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SHFH667 (FJ203585)	90.4
NT2_C97	1	*Riemerella anatipestifer* train 8755 (AY871828)	95.3	Human skin Flavobacteriaceae bacterium clone 4P2-105 (EF419395)	99.6
Cyanobacteria/ Chloroplasts					
NT2_C26	12	*Trichodesmium contortum* (AF013028)	97.2	*Trichodesmium contortum* (AF013028)	97.2
NT2_C35	1	*Synechococcus* sp. P1 (AF132774)	80.0	Sandy carbonate sediment clone CI5cm.G09 (EF208693)	99.5
NT2_C79	2	*Gloeobacter violaceus* PCC 7421 (BA000045)	83.0	Uncultured rumen bacterium 4C0d-2 (AB034016)	95.5
NT2_C80	1	*Gloeobacter violaceus* PCC 7421 (BA000045)	83.5	Uncultured rumen bacterium 4C0d-2 (AB034016)	94.6
NT2_C103	1	*Hydrocoleum lyngbyaceum HBC7* (EU249124)	90.6	Black band diseased (BBD) coral tissue cyanobacterium clone BBD_216_24 (DQ446126)	92.3
NT2_C111	4	*Lyngbya aestuarii* PCC 7419 (AJ000714)	97.5	*Lyngbya aestuarii* PCC 7419 (AJ000714)	97.5
Deferribacteres					
NT2_C45	1	*Cloacibacillus evryensis* (CU463952)	93.3	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone A2G6 (EU885044)	99.0
Firmicutes					
NT2_C11	2	*Epulopiscium fishelsoni* (AF067413)	94.7	*Epulopiscium fishelsoni* (AF067413)	94.7
NT2_C13	13	*Clostridium bifermentans* (AF320283)	97.1	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone A3E11 (EU885014)	97.5
NT2_C22	1	*Epulopiscium* sp. morphotype B (M99574)	94.4	Yellow catfish mucus and GI contents uncultured bacterium clone IC76 (GQ359998)	98.2
NT2_C23	1	*Epulopiscium* sp. morphotype B (M99574)	93.8	Yellow catfish mucus and GI contents uncultured bacterium clone IC76 (GQ359998)	97.0
NT2_C70	2	*Butyricicoccus pullicaecorum* strain 25-3 (EU410376)	90.9	Human fecal clone SJTU_B_15_94 (EF403042)	92.7
NT2_C83	2	*Anaerovorax odorimutans* strain NorPut (AJ251215)	91.9	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone A2A4 (EU885005)	96.7
NT2_C88	2	*Anaerovorax odorimutans* strain NorPut (AJ251215)	92.2	Oil sands tailings pond Uncultured Clostridiales clone 12-17A (EF420223)	93.0
NT2_C95	2	*Turicibacter sanguinis* strain MOL361 (NR_028816)	96.1	Rat fecal material uncultured bacterium clone R-9107 (FJ881096)	96.6
NT2_C102	2	*Catabacter* sp. YIT 12065 (AB490809)	85.9	pre-adolescent turkey cecum clone CFT214E5 (DQ456434)	88.7
Planctomycetes					
NT2_C15	1	*Pirellula* sp.strain ACM 3181 (X86388)	87.2	Uncultured marine environmental genomic DNA *Pirellula* clone 5H12 (AF029076)	95.1
NT2_C39	1	*Rhodopirellula* sp. CS14 (FJ624322)	89.6	Mangrove soil uncultured bacterium clone MSB-2F2 (EF125449)	94.6
NT2_C73	1	*Planctomyces* sp. Schlesner 664 (X81955)	91.0	Heavy metal polluted soil uncultured bacterium clone T1-42 (GQ487907)	90.3
Proteobacteria					
NT2_C14	1	*Roseovarius* sp. B108 (EU742628)	97.2	Eutrophic, saline Salton Sea uncultured bacterium clone SSW59Au (EU592378)	97.2
NT2_C24	1	*Ochrobactrum* sp. LJJS1-2 (DQ133574)	93.0	*Acropora cervicornis* coral uncultured bacterium clone Acer_O09 (GU118008)	98.0
NT2_C25	1	*Acinetobacter junii* (AB101444)	99.7	Fracture water uncultured bacterium clone TT118ant14d08 (GQ921377)	99.7
NT2_C33	1	*Legionella* sp. OUB41 (AB058918)	90.0	Hydrothermal precipitate uncultured bacterium clone ELSC-TVG13-B17 (GU220733)	95.0
NT2_C38	2	*Nitratireductor* sp. C115 (GU447302)	94.6	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SGUS601 (FJ202647)	99.1
NT2_C61	1	*Nitratireductor* sp. CC-MHSW-5 (EU564843)	93.1	*Montastraea faveolata* uncultured bacterium clone Mfav_F16 (GU118597)	95.3
NT2_C76	1	*Diaphorobacter* sp. GS-1 (FJ158841)	99.9	*Diaphorobacter* sp. GS-1 (FJ158841)	99.9
NT2_C78	1	*Anderseniella baltica* (AM712634)	97.4	*Anderseniella baltica* (AM712634)	97.4
Tenericutes					
NT2_C104	1	*Mycoplasma pirum* strain 70-159 (NR_029165)	87.2	*Mycoplasma pirum* strain 70-159 (NR_029165)	87.2
NT2_C212	2	*Mycoplasma moatsii* (AF412984)	93.0	Intestinal tract of *Notothenia coriiceps* uncultured bacterium clone A252_NCI (FJ456770)	92.8
Verrucomicrobia					
NT2_C41`	1	*Haloferula sargassicola* strain MN1-1047 (AB372857)	85.2	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SHFH449 (FJ203414)	88.8

[a] The percent identity towards each database entry is indicated. Accession numbers are provided parenthetically.
[b] The number of clones assigned to each OTU.

Table 3.2 Percent identity values for *Naso tonganus* segment III 16S rRNA gene sequences [a]

OTU	#[b]	Closest related bactrial sequence	% similarity	Closest related bacterial or environmental sequence	% similarity
Bacteroidetes					
NT3_C11	1	*Alistipes sp.* RMA 9912 (GQ140629)	92.8	Human feces clone RL310_aam39d08 (DQ795036)	93.6
Cyanobacteria/ Chloroplasts					
NT3_C1	1	*Musa acuminata* (EU017026)	99.3	*Musa acuminata* (EU017026)	99.3
NT3_C23	2	*Trichodesmium contortum* (AF013028)	97.1	*Trichodesmium contortum* (AF013028)	97.1
Firmicutes					
NT3_C19	1	*Blautia sp.* Ser8 (GU124472)	89.7	Uncultured Firmicutes bacteriumof black band disease microbial mats clone RB_13f (EF123527)	91.2
Fusobacteria					
NT3_C48	5	*Cetobacterium somerae* strain WAL 14325 (NR_025533)	96.4	Ex-germfree adult mouse cecum clone aaa94g05 (DQ817023)	96.5
NT3_C54	14	*Cetobacterium somerae* strain WAL 14325 (NR_025533)	97.6	Zebrafish digestive tract clone aab56g04 (DQ815008)	97.7
Proteobacteria					
NT3_C26	1	*Mesorhizobium sp.* DG943 (AY258089)	95.3	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SGUS601 (FJ202647)	99.2
NT3_C51	4	*Trojanella thessalonices* (AF069496)	85.4	Springbok feces uncultured bacterium clone SBSD_aaa02h07 (EU475489)	95.1
NT3_C53	1	*Legionella dresdeniensis* (AM747393)	88.1	Marine sediment Uncultured bacterium clone BD72BR164 (GU362977)	89.3
Verrucomicrobia					
NT3_C32	7	*Akkermansia muciniphila* ATCC BAA-835 (AY271254)	89.8	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone AC9 (EU885079)	94.9

[a] The percent identity towards each database entry is indicated. Accession numbers are provided parenthetically.
[b] The number of clones assigned to each OTU.

Table 3.3 Percent identity values for *Naso tonganus* segment IV 16S rRNA gene sequences [a]

OTU	#[b]	Closest related bactrial sequence	% similarity	Closest related bacterial or environmental sequence	% similarity
Bacteroidetes					
NT4_C8	1	*Alistipes massiliensis* (AY547271)	88.0	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone A2C6 (EU884961)	88.8
NT4_C75	1	*Alistipes onderdonkii* WAL 8169 (AY974071.1)	93.8	Human fecal clone SJTU_C_05_06 (EF404122)	94.0
Cyanobacteria/ Chloroplasts					
NT4_C1	1	*Fucus vesiculosus* (DQ307678)	94.7	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SHFH627 (FJ203555)	99.5
NT4_C32	1	*Gloeobacter violaceus* PCC 8105 (AF132791)	82.1	Rat fecal material uncultured bacterium clone C-23 (EU622705)	93.6
NT4_C34	1	*Synechococcus* sp. IR11 (AF448079)	82.8	Uncultured rumen bacterium 4C0d-2 (AB034016)	95.4
Firmicutes					
NT4_C2	2	*Clostridium lituseburense* (M59107.1)	97.4	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone A2E11 (EU885014)	98.4
NT4_C5	1	*Ruminococcus bromii* strain YE282 (DQ882649)	90.0	Steer rumen fluid clone 1103200823104 (EU842424)	92.8
NT4_C7	3	*Clostridium sp.* CS2 (FJ638499)	97.4	*Clostridium sp.* CS2 (FJ638499)	97.4
NT4_C11	1	*Catabacter sp.* YIT 12065 (AB490809)	85.8	Pre-adolescent turkey cecum clone CFT214E5 (DQ456434.1)	89.0
NT4_C17	1	*Holdemania filiformis* strain J1-31B-1 1 (NR_029335)	84.8	Uncultured rumen bacterium clone CF235 (EU871388)	92.4
NT4_C24	6	*Bacillus funiculus* (NR_028624)	88.2	Rat fecal material uncultured bacterium clone R-9107 (FJ881096)	96.6
NT4_C26	2	*Clostridium* sp. 14505 (AJ318890)	92.6	Human fecal SJTU_E_08_59 (EF400077)	93.6
NT4_C31	1	*Clostridium* sp. FG4 (AB207248.1)	85.0	Uncultured bacterium clone TU1_aaa01f02 (EU470107)	94.9
NT4_C33	1	*Streptococcus thermophilus* strain IMAU40151 (FJ915689)	99.9	Fermented Yak Milk uncultured bacterium clone IMAU 528 (GQ267990)	99.8
NT4_C39	1	*Eubacterium oxidoreducens* strain G2-2 (AF202259.1)	93.7	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone AG6 (EU885021)	95.4
NT4_C42	1	*Oscillibacter valericigenes* (AB238598)	94.5	Human fecal clone SJTU_C_03_84 (EF404039)	94.5
NT4_C48	1	*Eubacterium* sp. Pei061 (AJ629069)	91.4	*Eubacterium* sp. Pei061 (AJ629069)	91.4
NT4_C58	1	*Clostridium* sp. 14505 (AJ318890.1)	93.4	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone AG6 (EU885021)	99.4
NT4_C60	2	*Clostridium methylpentosum* (NR_029355.1)	90.6	Human fecal clone clone SJTU_F_07_20 (EF399590)	90.9
NT4_C83	1	*Butyricicoccus pullicaecorum* strain 25-3 (EU410376)	89.9	Steer rumen fluid clone 1103200823528 (EU843243)	90.8
NT4_C87	1	*Catabacter* sp. YIT 12065 (AB490809)	86.2	Biogas slurry uncultured Clostridiales bacterium clone 182 (GU1 12196)	92.4
NT4_C109	1	*Clostridium* sp. YIT 12069 (AB491207)	89.5	Human fecal uncultured bacterium clone A4-137 (GQ898015)	95.6
NT4_C114	1	*Clostridium* sp. pandaD (AY957603)	93.8	Yellow catfish mucus and GI contents uncultured bacterium clone IC76 (GQ359998)	96.6
Fusobacteria					
NT4_C22	7	*Cetobacterium somerae* strain WAL 14325 (NR_025533)	97.6	Rock oysters Saccostrea glomerata digestive gland uncultured Cetobacterium sp. isolate SRODG069 (FM995174)	97.7
Proteobacteria					
NT4_C6	1	*Roseovarius crassostreae* isolate CV919-312 (AF114484)	95.7	Caribbean coral Montastraea faveolata uncultured bacterium clone SGUS1237 (FJ202179)	98.1
NT4_C9	2	*Ruegeria pomeroyi* strain LS80 (FJ937909)	98.8	Uncultured Rhodobacterales bacterium (AB294975)	99.0
NT4_C13	1	*Rickettsia conorii* strain Malish 7 (AE006914)	88.3	Lake water bacterium clone ELB16-030 (DQ015802)	91.0
NT4_C15	1	*Acinetobacter junii* (AB101444)	99.9	Fracture water uncultured bacterium clone TT118ant14d08 (GQ921377)	99.9
NT4_C43	3	*Trojanella thessalonicensis* (AF069496)	85.7	Uncultured bacterium clone SBSD_aaa02h07 (EU475489)	95.1
Spirochaetes					
NT4_C27	2	*Brevinema andersonii* strain ATCC 43811 (GU993264.1)	90.9	Mudsucker gut spirochete clone TP-1 (DQ340184)	96.6
Tenericutes					
NT4_C21	1	*Mycoplasma pirum* (M23940.1)	86.9	*Mycoplasma pirum* (M23940.1)	86.9
Verrucomicrobia					
NT4_C3	7	*Akkermansia muciniphila* strain Muc (AY271254.1)	89.9	Hindgut chamber of *Pomacanthus sexstriatus* uncultured bacterium clone AC9 (EU885079)	94.5

[a] The percent identity towards each database entry is indicated. Accession numbers are provided parenthetically.
[b] The number of clones assigned to each OTU.

Table 3.4 Percent identity values for *Naso tonganus* segment V 16S rRNA gene sequences[a]

OTU	#[b]	Closest related bactrial sequence	% similarity	Closest related bacterial or environmental sequence	% similarity
Bacteroidetes					
NT5_C8	1	*Alistipes finegoldii* strain 4401054 (AY643082.1)	87.5	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone A2E3 (EU884969)	93.3
NT5_C10	1	*Paludibacter propionicigenes* (AB078842)	83.2	Marine sediment uncultured bacterial clone BFE-P13 (DQ080148)	91.6
NT5_C12	1	*Rikenella microfusus* strain Q-1 (NR_025910)	87.6	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone AF10 (EU884965)	93.5
NT5_C13	1	*Alistipes finegoldii* strain 4401054 (AY643082.1)	89.0	Medicinal leech digestive tract Rikenella sp. clone PW8 (DQ355177)	88.8
NT5_C16	1	*Alistipes sp.* RMA 9912 (GQ140629)	93.2	Human fecal uncultured bacterium clone SJTU_C_05_06 (EF404122)	93.6
NT5_C48	1	*Rikenella microfusus* ATCC 29728 (L16498)	85.2	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone AD9 (EU884992)	92.0
NT5_C76	2	*Alistipes onderdonkii* strain WAL 8169 (AY974071)	94.8	Uncultured human fecal bacterium clone S2-126 (GQ898387)	94.9
NT5_C78	1	*Alistipes massiliensis* strain 3302398 (AY547271	88.0	Medicinal leech digestive tract Rikenella sp. clone PW8 (DQ355177)	89.8
Firmicutes					
NT5_C1	4	*Hespellia stercorisuis* strain PC18 (NR_025207.1)	93.6	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone AG6 (EU885021)	99.0
NT5_C15	2	*Clostridium ramosum* (M23731)	93.3	*Pachnoda ephippiata* larva midgut clone PeM37 (AJ576396)	94.4
NT5_C17	2	*Clostridium* sp. strain Z6 (AY949859)	90.8	Steer rumen fluid clone 1103200823104 (EU842424)	93.2
NT5_C49	2	*Anaerovorax odorimutans* strain NorPut (AJ251215)	93.5	Oil sands tailings pond clone 12-17A (EF420223)	94.7
NT5_C50	1	*Butyricicoccus pullicaecorum* strain 25-3 (EU410376)	93.5	Artic seal colon content uncultured bacterium clone CSG66 (GQ867494)	93.8
NT5_C53	1	*Clostridium alkalicellum* (AY959944)	84.0	Yunnan snub-nosed monkey uncultured bacterium clone J276 (GQ451281)	92.6
NT5_C68	1	*Clostridium leptum* (AJ305238)	93.0	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone A2B6 (EU885026)	96.0
NT5_C72	1	*Eubacterium* sp. Pei061 (AJ629069)	91.8	*Eubacterium* sp. Pei061 (AJ629069)	91.8
NT5_C80	1	*Anaerofilum agile* strain F (NR_029315)	89.9	Steer fecal clone EMP_X13 (EU794292)	92.8
NT5_C87	1	*Turicibacter sanguinis* strain MOL361 (NR_028816.1)	96.1	Rat feces uncultured bacterium clone R-9107 (FJ881096)	96.7
Fusobacteria					
NT5_C52	17	*Cetobacterium somerae* (NR_0255533.1)	97.6	Zebrafish digestive tract clone aab56g04 (DQ815008)	97.6
Lentisphaerae					
NT5_C34	1	*Cytophaga* sp. Dex80-43 (AJ431234)	87.7	Caribbean coral *Montastraea faveolata* uncultured bacterium clone SGUS1201 (FJ202698)	92.4
Planctomycetes					
NT5_C18	1	*Rhodopirellula* sp. SM50 (FJ624356)	96.4	Brackish water floc assosiated uncultured planctomycete clone Bac28_Flocs (AB491829)	98.0
Proteobacteria					
NT5_C5	1	*Desulfovibrio* sp. LNB2 (AY554146)	87.8	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone A3G10 (EU885107)	94.9
NT5_C39	1	*Trojanella thessalonices* (AF069496.1)	85.6	Springbok feces uncultured bacterium clone SBSD_aaa02h07 (EU475489)	95.4
NT5_C70	1	*Maritimibacter alkaliphilus* strain HTCC2654 (DQ915443.3)	94.3	Uncultured bacterium treated coral clone Dstr_I12 (GU1 18282)	98.3
NT5_C74	2	*Shewanella algae* strain MAS2758 (GQ372878)	95.4	Rock oysters *Saccostrea glomerata* digestive gland cultured bacterium isolate SRODG022 (FM995171)	95.8
Tenericutes					
NT5_C2	1	*Acholeplasma brassicae* (AY538163)	90.7	Marine herbivorous fish, *Kyphosus sydneyanus* , hindgut clone DMI (AY590186)	95.7
NT5_C56	1	*Anaeroplasma abactoclasticum* (M25050)	89.4	Uncultured human fecal bacterium clone C4-35 (GQ897174)	91.8
Verrucomicrobia					
NT5_C6	3	*Akkermansia muciniphila* ATCC BAA-835 (CP001071)	90.3	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone AC9 (EU885079)	95.2
NT5_C84	1	*Akkermansia muciniphila* ATCC BAA-835 (CP001071)	90.5	Hindgut chamber of Pomacanthus sexstriatus uncultured bacterium clone AG4 (EU885080)	95.0

[a] The percent identity towards each database entry is indicated. Accession numbers are provided parenthetically.
[b] The number of clones assigned to each OTU.

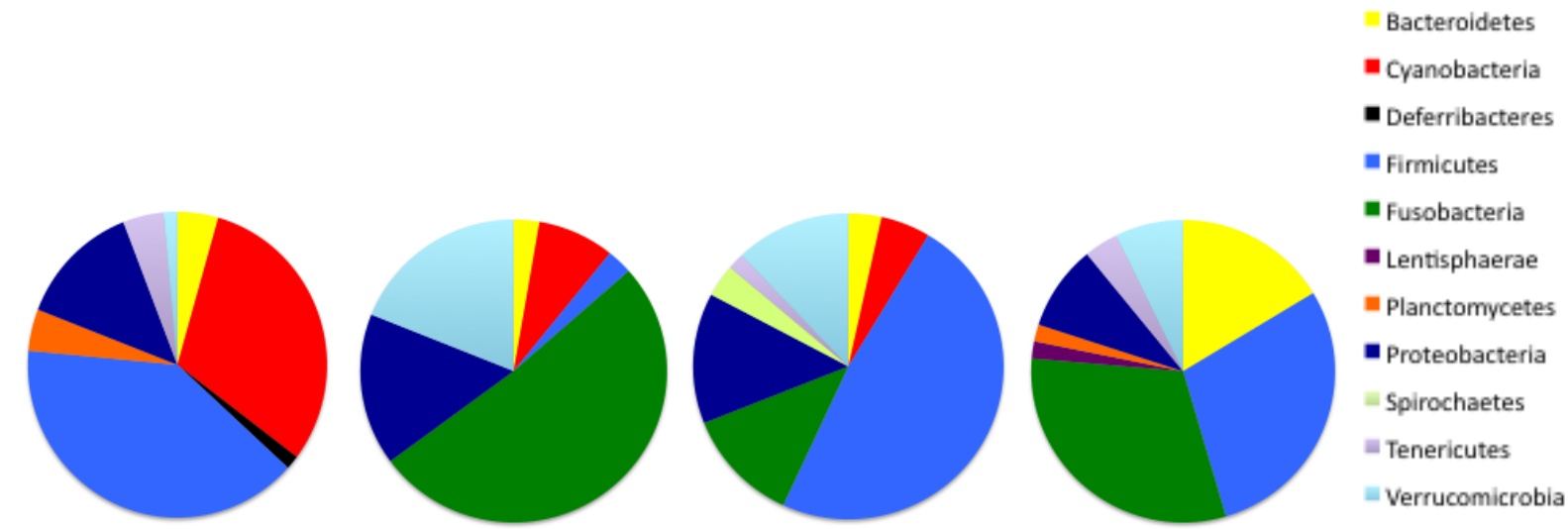


Figure 3.1. Taxonomic distribution of 16S rRNA gene sequences from bacteria isolated from the four segments of the *Naso tonganus* intestinal tract. Segment II: *Firmicutes* 40% *Cyanobacteria* 31%, *Proteobacteria* (13%), *Planctomycetes* (4%), *Bacteroidetes* (4%), *Tenericutes* (4%), *Deferribacteres* (2%) and *Verrucomicrobia* (2%). Segment III: *Fusobacteria* (51%), *Verrucomicrobia* (19%), *Proteobacteria* (16%), *Cyanobacteria* (8%), *Bacteroidetes* (3%) and *Firmicutes* (3%) Segment IV: *Firmicutes* (49%), *Proteobacteria* (14%), *Fusobacteria* (12%), *Verrucomicrobia* (12%), *Cyanobacteria* (5%), *Bacteroidetes* (3%), *Spirochaetes* (3%) and *Tenericutes* (2%). Segment V: *Fusobacteria* (31%), *Firmicutes* (29%), *Bacteroidetes* (16%), *Proteobacteria* (9%), *Verrucomicrobia* (7%), *Tenericutes* (4%), *Lentisphaerae* (2%) and *Planctomycetes* (2%).

segment III at 51% and 19%, respectively. *Proteobacteria* (16%), *Cyanobacteria* (8%), *Bacteroidetes* (3%) and *Firmicutes* (3%) were also present (Figure 3.1b). In segment IV, the dominant lineage was again *Firmicutes* (49%) followed by *Proteobacteria* (14%), *Fusobacteria* (12%), *Verrucomicrobia* (12%), *Cyanobacteria* (5%), *Bacteroidetes* and *Spirochaetes* both at 3% and *Tenericutes* at 2% (Figure 3.1c). Finally, the *Fusobacteria* dominated Segment V at 31% and the *Firmicutes* at 29%. *Bacteroidetes* (16%), *Proteobacteria* (9%), *Verrucomicrobia* (7%), *Tenericutes* (4%) and the *Lentisphaerae* and *Planctomycetes* (2%) were also present (Figure 3.1d).

Rarefaction curves generated from all four segments failed to level off at the species level distance of 0.03 (97% similarity) or at the genus level of 0.05 (95% similarity) except segment III where the beginning of leveling was observed at 0.05 (Figure 3.2). Chao diversity estimates based on 97% similarity were 47.55 (Segment II), 14.00 (Segment III), 72.00 (Segment IV) and 55.50 (Segment V) OTUs, indicating that samples represent between 44% and 71% of the species present in the environment. The abundance based coverage estimator, ACE, also supported these trends. At a distance of 0.3, Shannon values of greater than 2.7 in all but segment III indicate large unevenness in the distribution of the OTUs in the samples. The Simpson numbers also confirm high diversity in all of our samples except for segment III (Table 3.5).

Phylogenetic trees were constructed with *N. tonganus* bacterial clones from phyla that are important in fermentative processes: the *Firmicutes, Bacteroidetes* and *Fusobacteria.* Three of the five classes of *Firmicutes* were represented (Figure 3.3), with the majority of the clones belonging to the class *Clostridia* (76.9 %). The

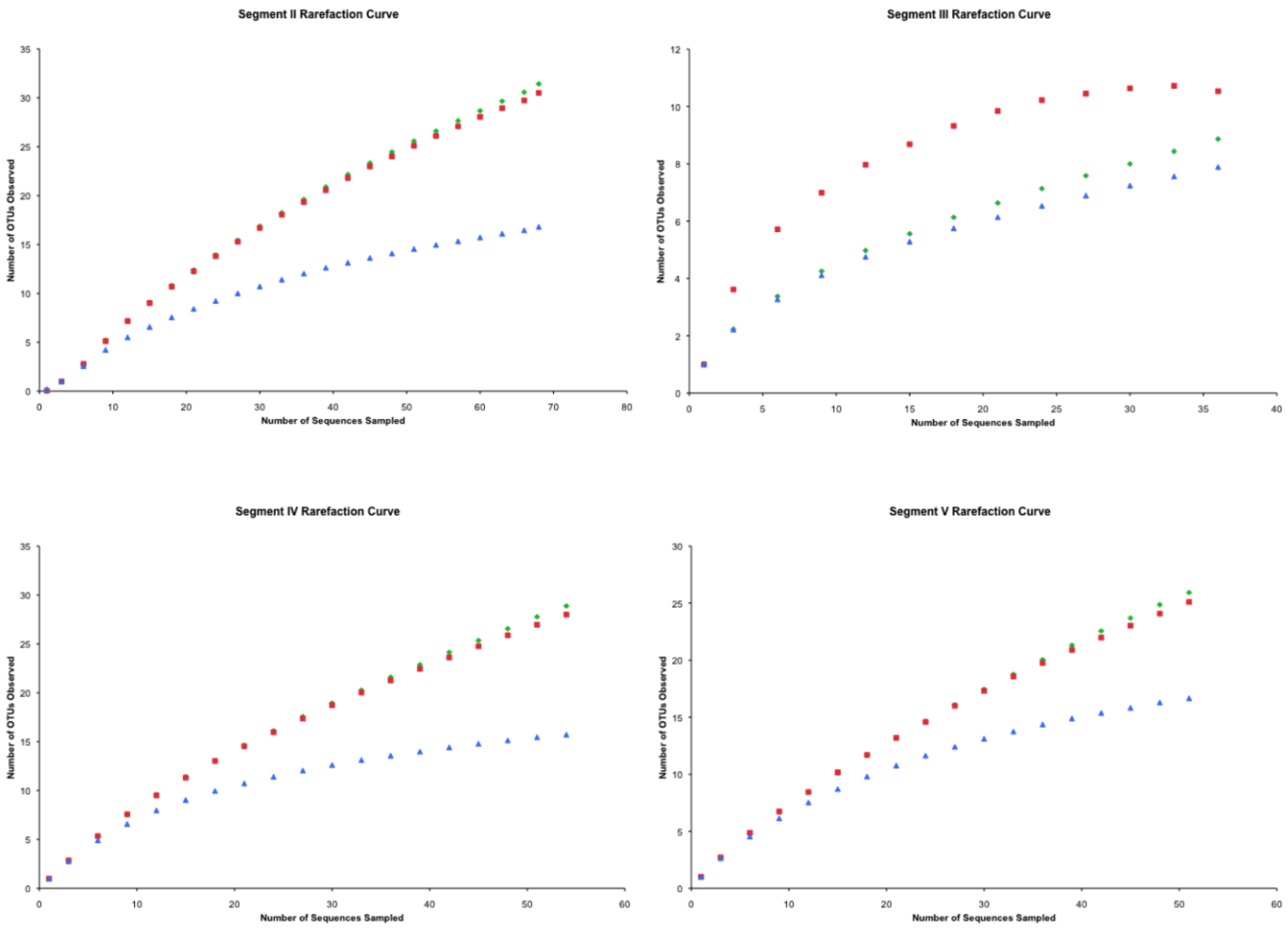
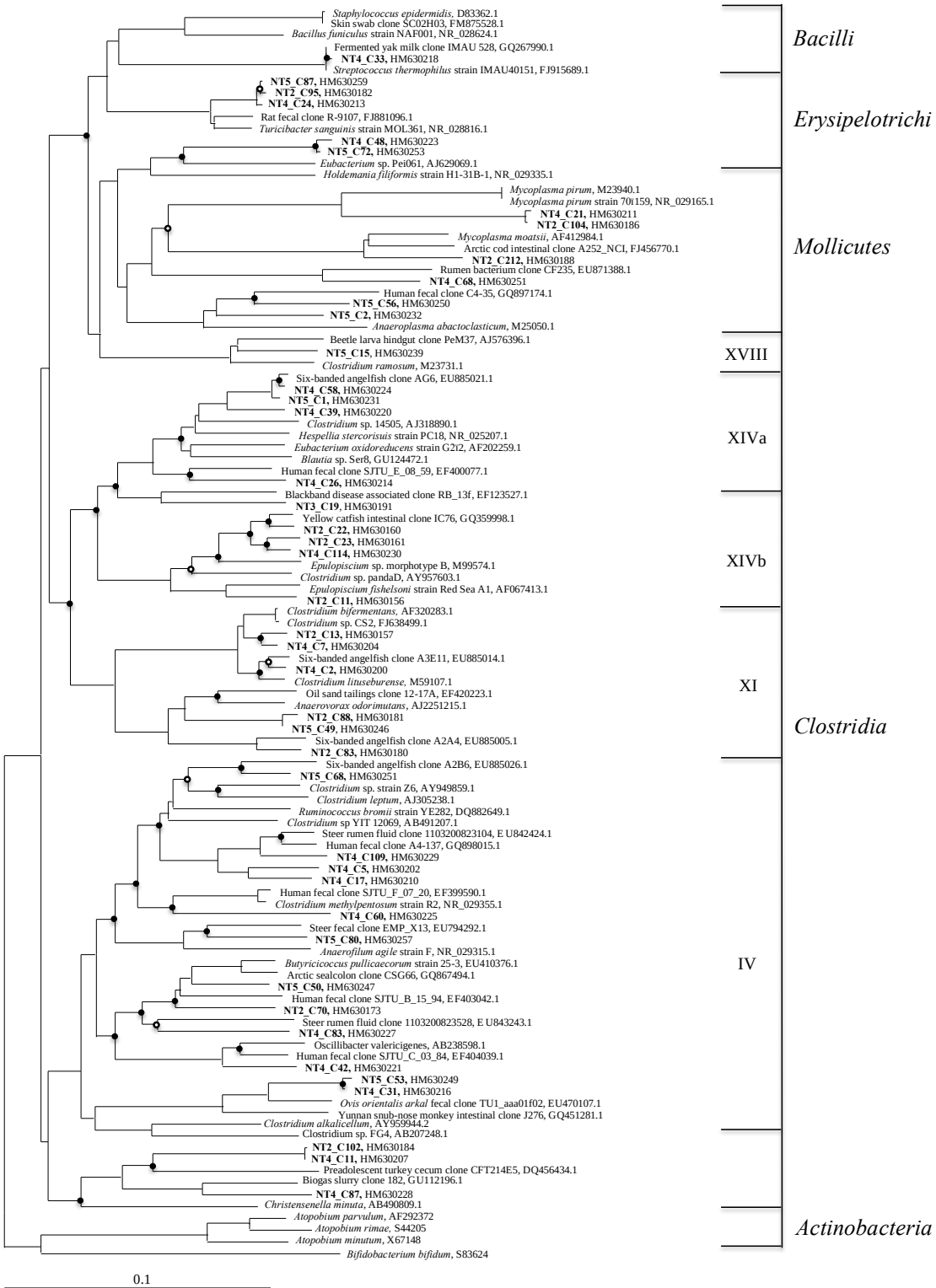


Figure 3.2. Rarefaction curves for segments II, III, IV and V comparing observed OTUs to the total number of sequences sampled at a distance level of 0.03 (green diamonds) 0.05 (red squares) and 0.2 (blue triangles).

Table 3.5. Estimated numbers of operational taxonomic units (OTUs) for each segment using nonparametric estimators (Chao1 and ACE), and diversity indices (Shannon and Simpson).

	Distance	ACE	Chao1	Shannon	Simpson
Segment II	0.03	55.75	47.55	3.01	0.07
	0.05	43.64	40.13	2.76	0.10
	0.20	24.34	20.50	2.18	0.16
Segment III	0.03	28.20	14.00	1.54	0.30
	0.05	24.87	14.00	1.51	0.30
	0.20	17.90	14.00	1.49	0.30
Segment IV	0.03	149.11	72.00	3.08	0.05
	0.05	59.60	39.00	2.97	0.05
	0.20	22.35	20.00	2.29	0.16
Segment V	0.03	68.67	55.50	2.76	0.11
	0.05	55.11	45.43	2.74	0.11
	0.20	26.26	21.00	2.36	0.17

Figure 3.3. Neighbor joining dendrogram demonstrating the distribution of *Naso tonganus* intestinal microbiota 16S rRNA gene sequences (bold) within the *Firmicutes*. Branching points supported by ≥ 90% of the bootstrap replicates are indicated with closed circles and those supported by 70% to 90% are indicated by open circles. Class names are listed to the right of the bracket. *Clostridium* clusters as identified by Collins *et al* (1994) are indicated by roman numerals to the left of the bracket. The unnamed group lacks cultured representatives with high similarity to the clones therein, such that a clear family affiliation is uncertain. The scale bar represents 10% sequence divergence


Staphylococcus epidermidis, D83362.1
Skin swab clone SC02H03, FM875528.1
Bacillus funiculus strain NAF001, NR_028624.1
Fermented yak milk clone IMAU 528, GQ267990.1
NT4_C33, HM630218
Streptococcus thermophilus strain IMAU40151, FJ915689.1
NT5_C87, HM630259
NT2_C95, HM630182
NT4_C24, HM630213
Rat fecal clone R-9107, FJ881096.1
Turicibacter sanguinis strain MOL361, NR_028816.1
NT4_C48, HM630223
NT5_C72, HM630253
Eubacterium sp. Pei061, AJ629069.1
Holdemania filiformis strain H1-31B-1, NR_029335.1
Mycoplasma pirum, M23940.1
Mycoplasma pirum strain 70i159, NR_029165.1
NT4_C21, HM630211
NT2_C104, HM630186
Mycoplasma moatsii, AF412984.1
Arctic cod intestinal clone A252_NCI, FJ456770.1
NT2_C112, HM630188
Rumen bacterium clone CF235, EU871388.1
NT4_C68, HM630251
Human fecal clone C4-35, GQ897174.1
NT5_C56, HM630250
NT5_C2, HM630232
Anaeroplasma abactoclasticum, M25050.1
Beetle larva hindgut clone PeM37, AJ576396.1
NT5_C15, HM630239
Clostridium ramosum, M23731.1
Six-banded angelfish clone AG6, EU885021.1
NT4_C58, HM630224
NT5_C1, HM630231
NT4_C39, HM630220
Clostridium sp. 14505, AJ318890.1
Hespellia stercorisuis strain PC18, NR_025207.1
Eubacterium oxidoreducens strain G2i2, AF202259.1
Blautia sp. Ser8, GU124472.1
Human fecal clone SJTU_E_08_59, EF400077.1
NT4_C26, HM630214
Blackband disease associated clone RB_13f, EF123527.1
NT3_C19, HM630191
Yellow catfish intestinal clone IC76, GQ359998.1
NT2_C22, HM630160
NT2_C23, HM630161
NT4_C114, HM630230
Epulopiscium sp. morphotype B, M99574.1
Clostridium sp. pandaD, AY957603.1
Epulopiscium fishelsoni strain Red Sea A1, AF067413.1
NT2_C11, HM630156
Clostridium bifermentans, AF320283.1
Clostridium sp. CS2, FJ638499.1
NT2_C13, HM630157
NT4_C7, HM630204
Six-banded angelfish clone A3E11, EU885014.1
NT4_C2, HM630200
Clostridium lituseburense, M59107.1
Oil sand tailings clone 12-17A, EF420223.1
Anaerovorax odorimutans, AJ2251215.1
NT2_C88, HM630181
NT5_C49, HM630246
Six-banded angelfish clone A2A4, EU885005.1
NT2_C83, HM630180
Six-banded angelfish clone A2B6, EU885026.1
NT5_C68, HM630251
Clostridium sp. strain Z6, AY949859.1
Clostridium leptum, AJ305238.1
Ruminococcus bromii strain YE282, DQ882649.1
Clostridium sp YIT 12069, AB491207.1
Steer rumen fluid clone 1103200823104, E U842424.1
Human fecal clone A4-137, GQ898015.1
NT4_C109, HM630229
NT4_C5, HM630202
NT4_C17, HM630210
Human fecal clone SJTU_F_07_20, EF399590.1
Clostridium methylpentosum strain R2, NR_029355.1
NT4_C60, HM630225
Steer fecal clone EMP_X13, EU794292.1
NT5_C80, HM630257
Anaerofilum agile strain F, NR_029315.1
Butyricicoccus pullicaecorum strain 25-3, EU410376.1
Arctic sealcolon clone CSG66, GQ867494.1
NT5_C50, HM630247
Human fecal clone SJTU_B_15_94, EF403042.1
NT2_C70, HM630173
Steer rumen fluid clone 1103200823528, E U843243.1
NT4_C83, HM630227
Oscillibacter valericigenes, AB238598.1
Human fecal clone SJTU_C_03_84, EF404039.1
NT4_C42, HM630221
NT5_C53, HM630249
NT4_C31, HM630216
Ovis orientalis arkal fecal clone TU1_aaa01f02, EU470107.1
Yunnan snub-nose monkey intestinal clone J276, GQ451281.1
Clostridium alkalicellum, AY959944.2
Clostridium sp. FG4, AB207248.1
NT2_C102, HM630184
NT4_C11, HM630207
Preadolescent turkey cecum clone CFT214E5, DQ456434.1
Biogas slurry clone 182, GU112196.1
NT4_C87, HM630228
Christensenella minuta, AB490809.1
Atopobium parvulum, AF292372
Atopobium rimae, S44205
Atopobium minutum, X67148
Bifidobacterium bifidum, S83624
0.1
Bacilli
Erysipelotrichi
Mollicutes
XVIII
XIVa
XIVb
XI
Clostridia
IV
Actinobacteria

Erysipelotrichi comprised 15.4 % of the clones, while the *Bacilli* made up 7.7 % of the library. Clones recovered from the fish fell into two classes of *Bacteroidetes* (Figure 3.4). The majority of these clones (78.6%) belong to the class *Bacteroidia*, while 14.3% belong to the class *Flavobacteria*. One clone, NT2_C72, affiliated with the *Bacteroidetes* but could not be accurately placed with either class. All of the *Fusobacterial* clones showed high identity to the previously described anaerobic bacterium *Cetobacterium somerae* (26).

T-RFLP analysis of bacterial community structure

T-RFLP analyses of bacterial 16S rRNA gene amplicons were used to explore between-fish and between-segment differences in the bacterial communities. *In silico* digests of the full-length clone sequences obtained from segments IV and V were performed. The clone restriction patterns of all commercially available 4-base recognition endonucleases from New England Biolabs were compared. Based on this analysis, *Rsa*I and *Msp*I were chosen for T-RFLP studies as these enzymes resolved the greatest number of dissimilar T-RFs in the 50 – 500 base pair range, with *Rsa*I and *Msp*I able to theoretically resolve 56 and 61 T-RFs, respectively.

In segment IV, *Rsa*I generated 18, 22, 15 and 13 distinct T-RFs in amplification products from fish 1-4, respectively (Figure 3.5) while in segment V of the same fish, 19, 17, 17 and 12 distinct T-RFs were observed (Figure 3.6). Samples digested with the restriction enzyme *Msp*I resulted in 17, 20, 12 and 18 unique T-RFs in segment IV of fish 1-4 (Figure 3.7) and 8, 19, 8 and 8 unique T-RFs in segment V of the same fish (Figure 3.8). In samples treated with *Rsa*I, the T-RF sizes 115, 317 and 414 were found in segment IV of all fish, while T-RF sizes 162, 451 and 468 were

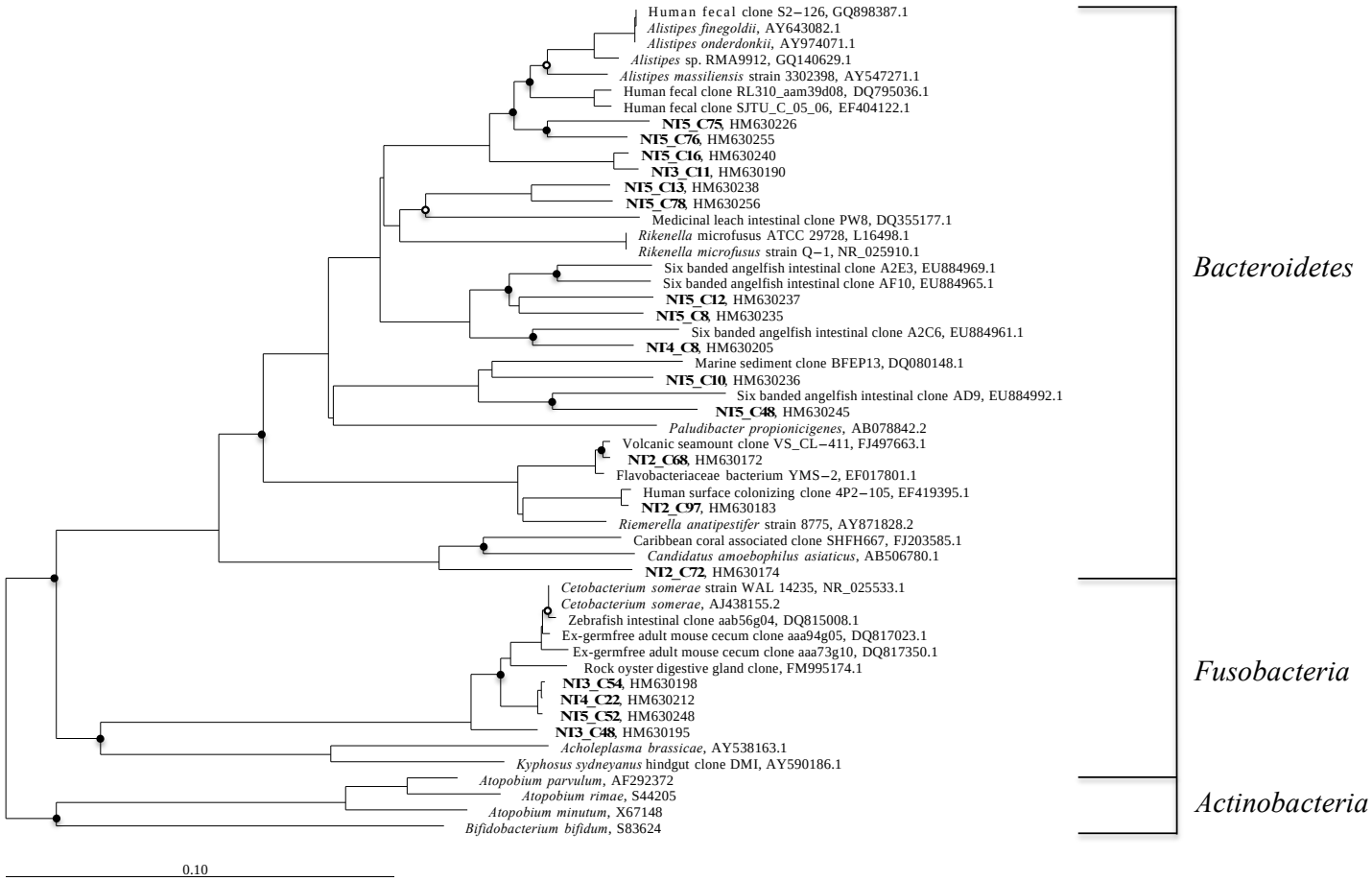


Figure 3.4. Neighbor joining dendrogram demonstrating the distribution of *Naso tonganus* intestinal microbiota 16S rRNA gene sequences (bold) within the *Bacteroidetes* and *Fusobacteria*. Branching points supported by ≥ 90% of the bootstrap replicates are indicated with closed circles and those supported by 70% to 90% are indicated by open circles. The scale bar represents 10% sequence divergence.

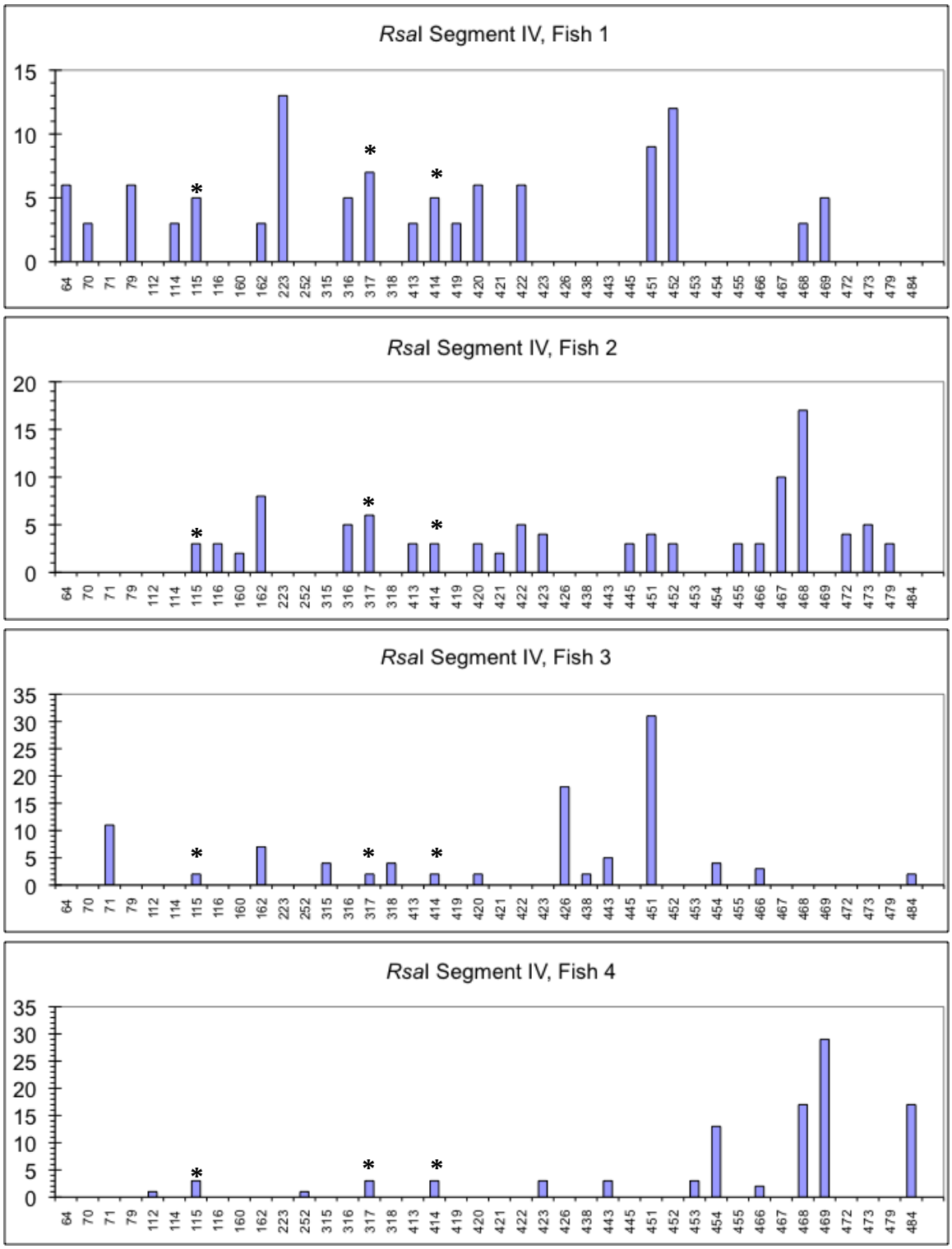


Figure 3.5. Terminal-restriction fragment length polymorphism profiles representing microbial community diversity from Segment IV, fish 1 - 4. T-RFs were generated with restriction enzyme *Rsa*I. The x axis is the T-RF size (bases) and the y axis is the fluorescence intensity (arbitrary units). T-RFs common in all fish are indicated by an asterisk.

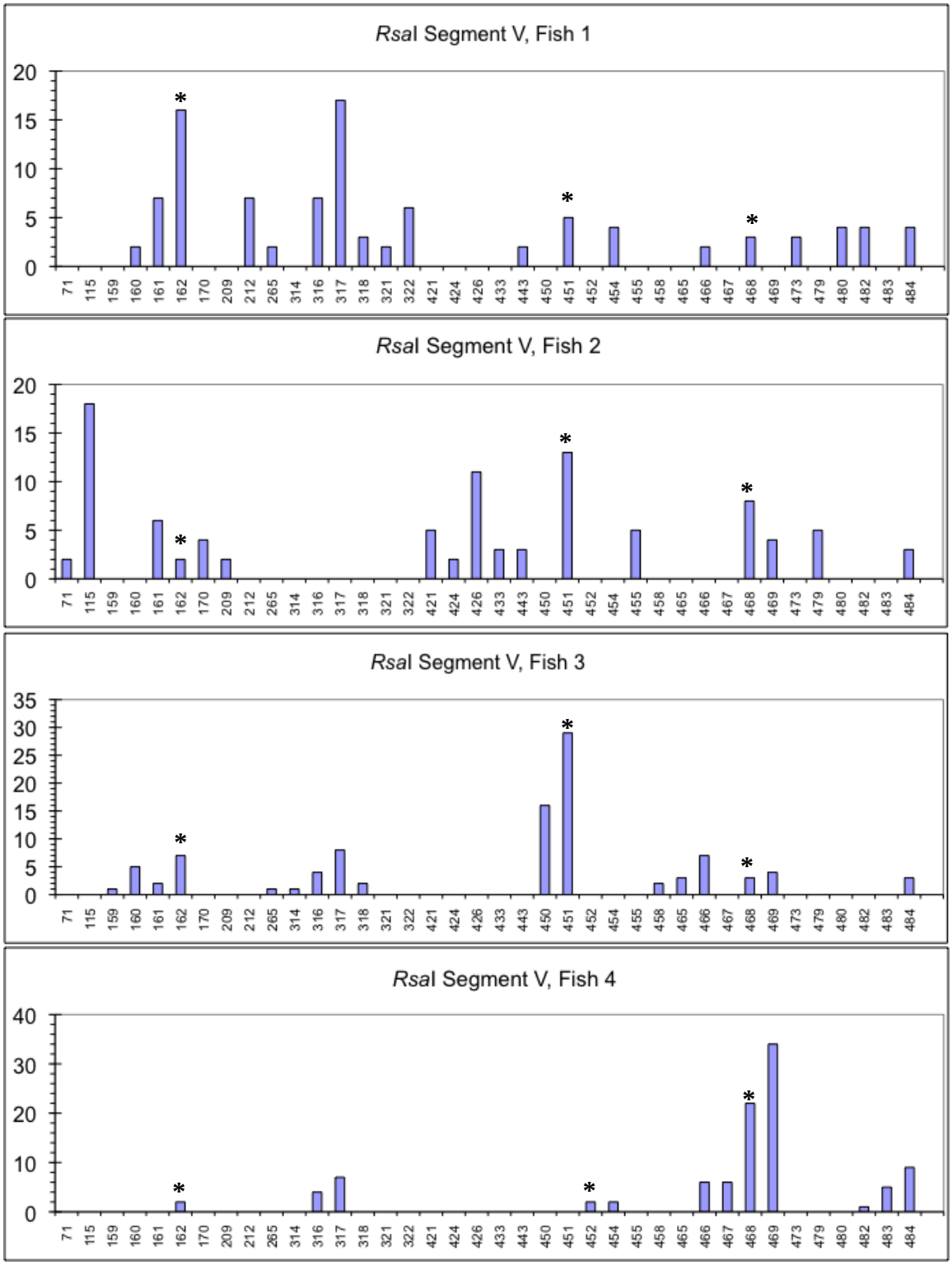


Figure 3.6. Terminal-restriction fragment length polymorphism profiles representing microbial community diversity from Segment V, fish 1 - 4. T-RFs were generated with restriction enzyme *Rsa*I. The x axis is the T-RF size (bases) and the y axis is the fluorescence intensity (arbitrary units). T-RFs common in all fish are indicated by an asterisk.

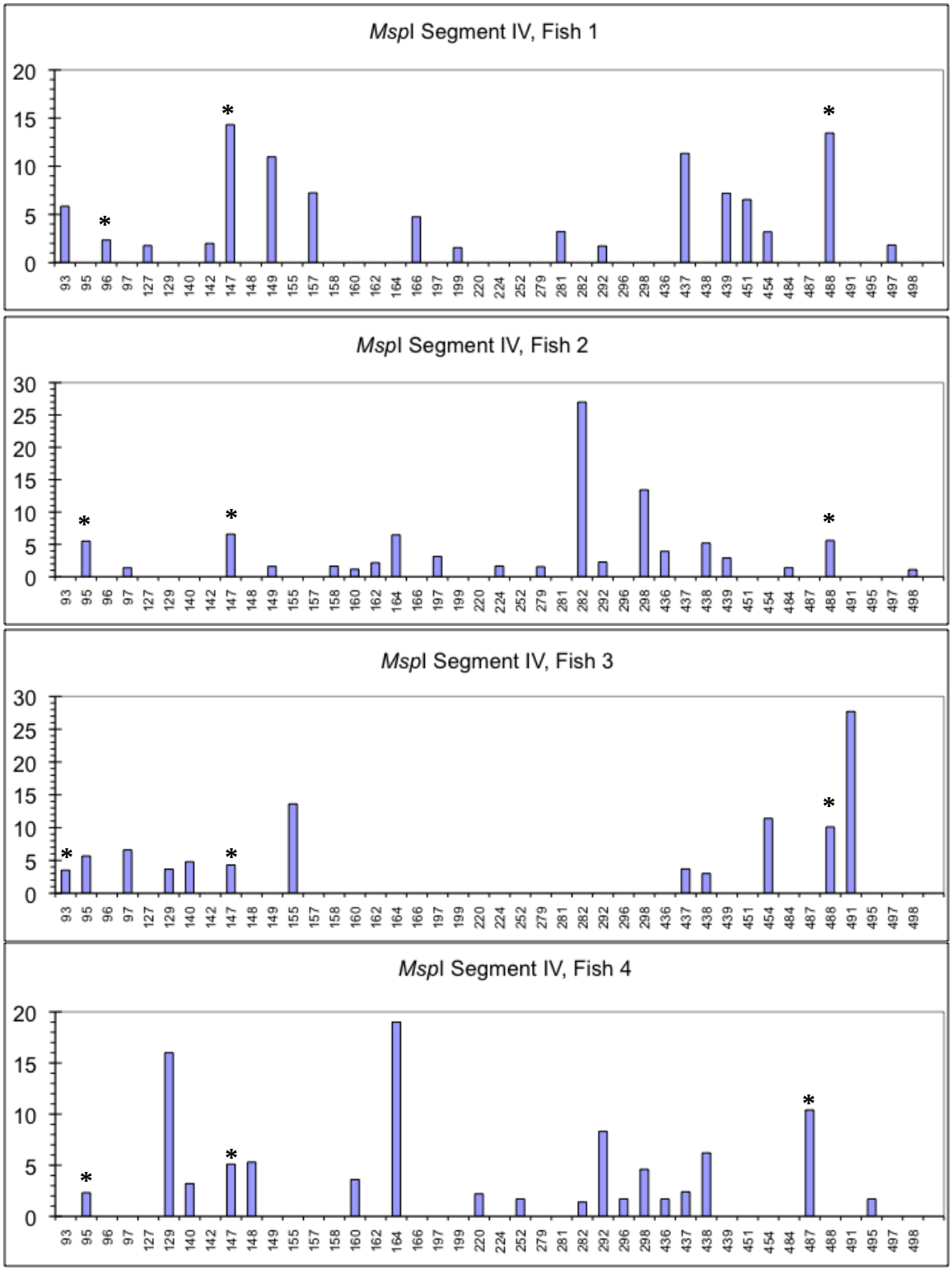


Figure 3.7. Terminal-restriction fragment length polymorphism profiles representing microbial community diversity from Segment IV, fish 1 - 4. T-RFs were generated with restriction enzyme *Msp*I. The x axis is the T-RF size (bases) and the y axis is the fluorescence intensity (arbitrary units). T-RFs common in all fish are indicated by an asterisk.

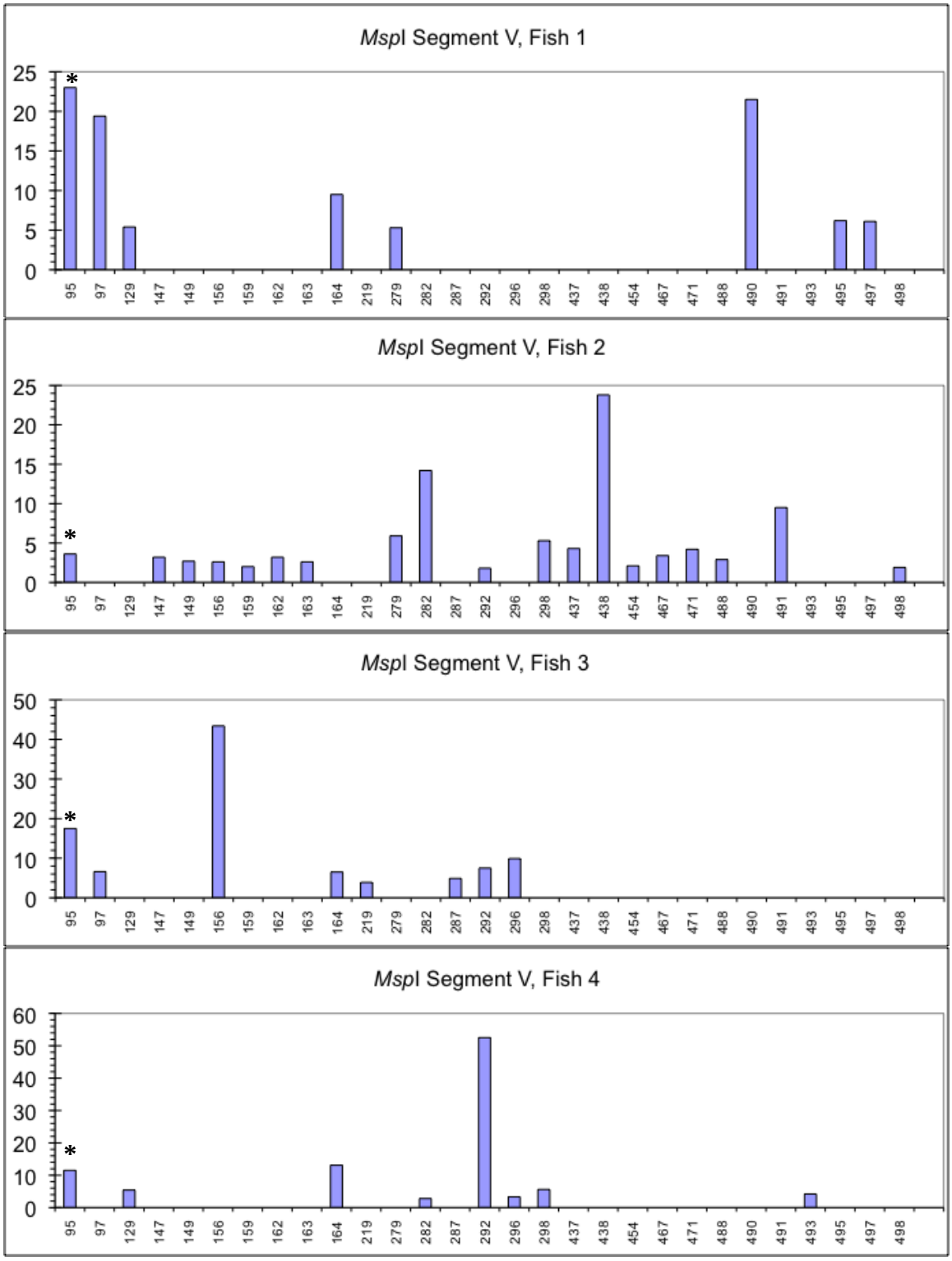


Figure 3.8. Terminal-restriction fragment length polymorphism profiles representing microbial community diversity from Segment V, fish 1 - 4. T-RFs were generated with restriction enzyme *Msp*I. The x axis is the T-RF size (bases) and the y axis is the fluorescence intensity (arbitrary units). T-RFs common in all fish are indicated by an asterisk.

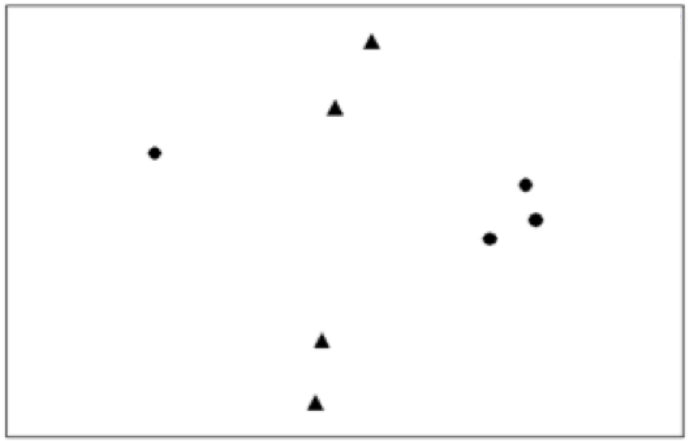
present in all segment V samples. When the same analysis was carried out with PCR products digested with the restriction enzyme *Msp*I, T-RF sizes 95, 147 and 488 were found in segment IV of all fish, while only one universal T-RF, size 95, was present in segment V of all fish.

Analysis of T-RFs using the Bray-Curtis coefficient yielded the same statistical conclusions for presence/absence data and for relative abundance data, therefore the presence/absence data set was used for MDS analysis. The MDS plots for segment IV and V are shown in Figure 3.9. The stress of these plots is 0.12 and 0.09, respectively. Differences in the T-RF community structure in Segment IV and V were investigated using ANOSIM and global R statistics (Table 3.6). The overall significance level of the sample statistic was set at $P < 0.05$. No differences were observed between segment IV and segment V when either *Rsa*I ($R = 0.172$, $P = 0.257$) or *Msp*I ($R = 0.120$, $P = 0.314$) was used, inferring that the microbial communities are not statistically different from each other.

SIMPER analysis revealed a general trend of low similarity between fish for all segments regardless of restriction enzyme ($\leq 52\%$), however some fish did show higher similarity (Table 3.7). These results are consistent with those obtained by the MDS plots (Figure 3.9). The highest average dissimilarities between fish were contributed by T-RF sizes 223, 426 484, 451 and 469 with *Rsa*I for segment IV and 115, 450, 451 and 469 for segment V and 282, 491, 164, 487and 129 for segment IV and 438, 156, 490, 292 and 493 for segment V with *Msp*I.

The fish used in the studies described above were unusual in that they harbored few *Epulopiscium* cells. To investigate the influence of *Epulopiscium* populations on

A.



B.

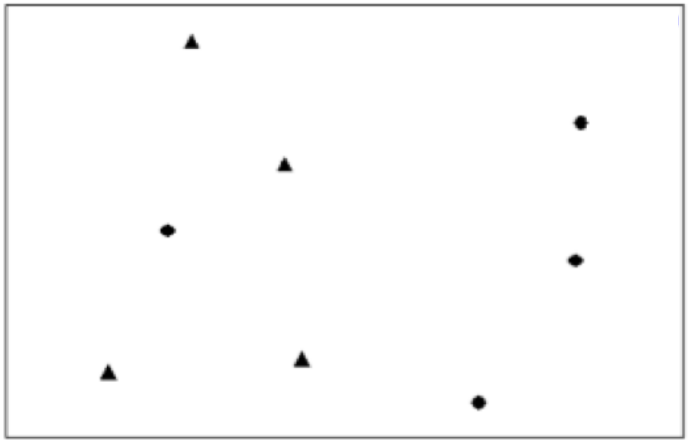


Figure 3.9. Multi-dimensional scaling plots of samples for T- RFLP community profiles generated with the bacterial specific primer set and restriction enzymes *Rsa*I (panel A) and *Msp*I (panel B). Samples with out *Epulopiscium, g*astrointestinal segment IV (▲), samples with out *Epulopiscium, g*astrointestinal segment V(●). The stress of the plots generated with *Rsa*I and *Msp*I were 0.12 and 0.09, respectively.

Table 3.6. Pairwise comparisons between community profiles in segment IV and segment V generated with ANOSIM analysis [a]

Restriction Enzyme	Sample Statistic	Significance Level
*Rsa*I	0.172	0.257
*Msp*I	0.12	0.314

[a] R sample statistic and corresponding significance level were computed to assess pairwise differences between segment IV and segment V

Table 3.7. SIMPER analysis of Segment IV and Segment V T-RFs

% Similarity of T-RFs between fish (presence/absence):

	1 and 2	1 and 3	1 and 4	2 and 3	2 and 4	3 and 4
Segment IV *Rsa*I	52	32	29	34	33	42
Segment V *Rsa*I	31	57	45	34	26	46
Segment IV *Msp*I	36	34	20	25	45	35
Segment V *Msp*I	14	37	37	18	26	46

the *N. tonganus* intestinal microbial community, the same analyses were performed on PCR products obtained from an individual fish that possessed dense populations of *Epulopiscium* sp. type B. The abundant DNA found in *Epulopiscium* (48) would likely skew these analyses, therefore *Epulopiscium* cells were individually picked out of the sample prior to DNA extraction and PCR amplification of the 16S rRNA gene. Eleven distinct T-RFs were observed in this fish after digestion with either the restriction enzyme *Rsa*I or *Msp*I (Figure 3.10).

ANOSIM was again used to assess if these three T-RF community profiles (segment IV and V without *Epulopiscium* and the mixed segment with *Epulopiscium*) were different from each other. There were no differences observed between segment IV and segment V when either *Rsa*I or *Msp*I was used to generate the community profile. When either segment without *Epulopiscium* was compared with the intestinal sample from fish that had dense populations of *Epulopiscium*, pairwise differences were not observed with either enzyme (R = 0.667, P= 0.200 and R = 0.917, P = 0.200 for *Rsa*I and R = 0.417, P = 0.200 for both pairings with *Msp*I, respectively) (Table 3.8).

Assignment of Clones to T-RFs

Several T-RFs could be linked to specific OTUs identified in the clone libraries. In both segments IV and V, a single *Fusobacterial* OTU related to *C. somerae*, was numerically dominant. The *in silico* digest of this clone yielded a predicted T-RF size of 451 (±2) bases when digested with *Rsa*I. These were the only clones that yielded this size T-RF and this result is consistent with the frequency of OTU recovery in the libraries. In segment IV, fish 1 and 3 and in segment V, fish 2

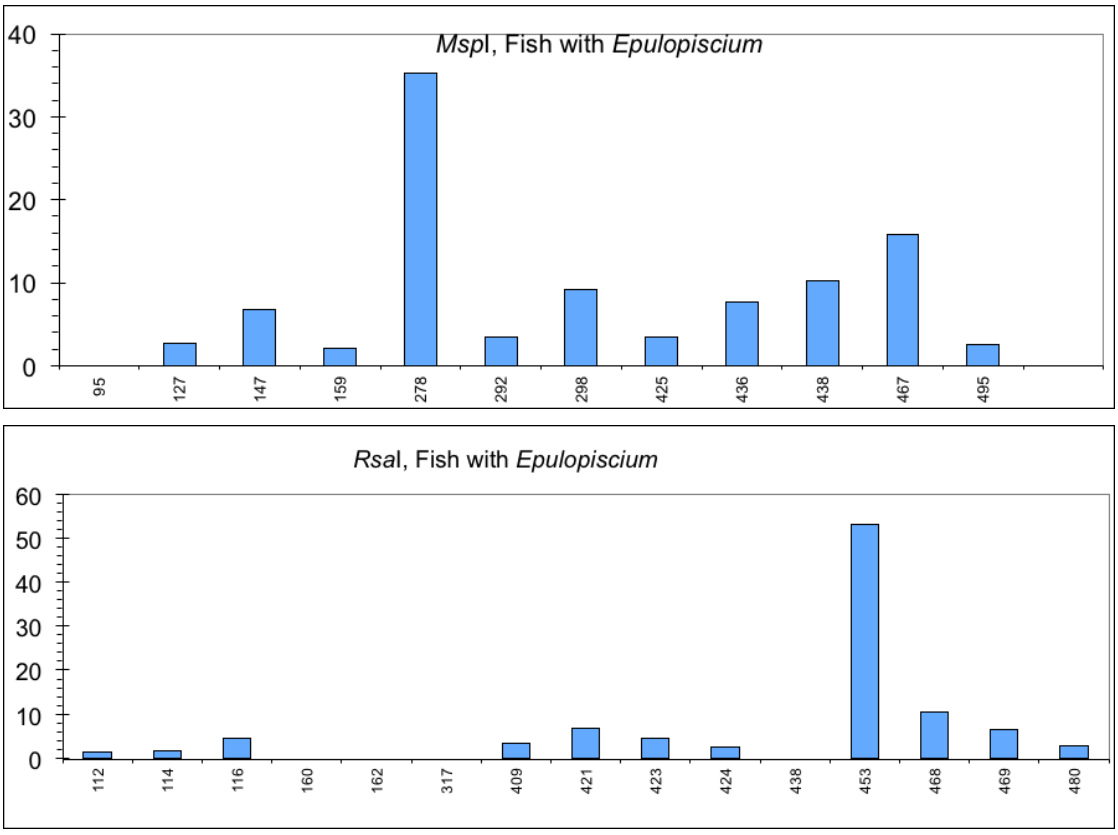


Figure 3.10. Terminal-restriction fragment length polymorphism profiles representing microbial community diversity from the fish harboring *Epulopiscium* cells. T-RFs in the top panel were generated with restriction enzyme *Msp*I. T-RFs in the bottom panel were generated with the restriction enzyme *Rsa*I. The x axis is the T-RF size (bases) and the y axis is the fluorescence intensity (arbitrary units).

Table 3.8. Global R sample statistics generated with ANOSIM analysis for segments IV and V without *Epulopiscium* (group 4 and group 5) and the mixed segment with *Epulopiscium* (group 6).

Restriction Enzyme	Groups	Sample Statistic	Significance Level
*Rsa*I	4 & 6	0.667	0.200
	5 & 6	0.917	0.200
*Msp*I	4 & 6	0.417	0.200
	5 & 6	0.417	0.200

and 3 this was one of the T-RFs with the highest relative abundance. It was also present in segment IV, fish 2 and segment V, fish 1. When this clone was digested with *Msp*I, a predicted T-RF size of 158 (±2) bases was obtained. A T-RF of an equivalent size was observed in segment IV, fish 1, 2 and 4 and in segment V, fish 2 and 3. One of the common T-RFs (317, ±2 bases) identified in segment IV (*Rsa*I digest) could be linked to several clones within the phyla *Bacteroidetes* and *Proteobacteria.* In segment V (*Rsa*I digest), one of the common T-RFs (468, ±2 bases) could be linked to several clones in the phylum *Firmicutes.* In addition, while this fragment was not one of the common T-RFs identified in segment IV, it was present at high relative abundance in three of the four fish, and could be putatively linked to another suite of clones belonging to the *Firmicutes*. When PCR products were digested with *Msp*I, a common T-RF of 95 (±2 bases) was observed in all segments of all fish. In both segments clones belonging to the phyla *Bacteroidetes* possessed T-RFs of this size.

T-RFLP profiles of the sample that had *Epulopiscium* removed, revealed a T-RF of 451 (±2 bases) with digestion by *Rsa*I and 278 (±2 bases) with *Msp*I. These correspond to T-RFs obtained by *in silico* digest of the 16S rRNA gene sequence available for *Epulopiscium* type B in the NCBI GenBank database (4).

Discussion

Historically, the role of microbial symbionts as important contributors to digestion in marine herbivorous or omnivorous fishes has been dismissed (31). More emphasis has been placed on the release of algal cellular contents through mechanical or chemical means (i.e. by the grinding action of a muscular stomach or pharyngeal

mill, and lysis of cells by stomach acid) and the utilization of endogenous enzymes produced by the fish (31). However several studies have suggested that microbes present in the digestive tract of fishes may play an important role in the breakdown and fermentation of algae and seagrasses (10, 13, 14, 15, 17, 25, 35, 49). By taking a culture-independent approach, this study provides insights into the complex intestinal microbiome of the omnivorous fish *N. tonganus*. Adult *N. tonganus* (formerly identified as *N. tuberosus*) are usually found on outer coral reefs in the tropical southern Pacific Ocean (36). They feed primarily on green and red algae, but also consume brown algae, detritus and invertebrates (8). These marine algal foodstuffs contain a diverse array of polysaccharides, such as carrageen, agarose, alginate, a modified starch polymer and laminarin (64). Accessing these polysaccharides would require the fish to evolve an assortment of carbohydrases or alternatively develop associations with microbes with these degradative capacities (59, 61).

Clone Libraries Reveal Novel Bacteria from Phyla Normally Found in Intestinal Ecosystems

The bacterial sequences recovered from all intestinal segments were diverse; many of the clones showed a low level of sequence identity to previously described bacterial sequences (Tables 3.1 – 3.4). In all segments, a greater percentage of the sequences recovered (10.3% to 42.2%) shared a higher identity to environmental sequences when compared to the percentage (3.5% to 30.0%) that were more closely affiliated to cultured sequences. The highest level of divergence from previously described sequences was found in the more posterior segments, which also possessed the highest OTU diversity. This level of community divergence from previously

described organisms demonstrates that we still have much to learn about the structure and function of this microbiome.

A significant number of the 16S rRNA genes recovered from the intestinal tract of *N. tonganus* belonged to the phyla *Firmicutes* and *Bacteroidetes. Firmicutes* are represented in libraries from many different gastrointestinal systems (41, 42) including humans (23), horses (20), pigs (40), ruminants (24, 53) and other herbivorous fishes (15, 50). The Gram-negative *Bacteroidetes* have been identified as a dominant group in 16S rRNA gene clone libraries generated from human, porcine, equine and rumen sources (27, 41, 42). Interestingly, a clone library generated from contents of the distal gut in the surgeonfish *Acanthurus nigricans* was dominated by these phyla as well (60).

Several of the most closely related clones available from Genbank were derived from other intestinal ecosystems including the intestinal tract of other fish. Clones belonging to the phylum *Firmicutes*, showed a high degree of similarity to clones recovered from the intestinal tract and intestinal mucus of the Yellow Catfish (67). One clone recovered from segment V, belonging to the phylum *Tenericutes*, showed 95.2% similarity to a clone recovered from the temperate marine herbivorous fish *Kyphosus sydneyanus* (50). Of particular interest is the large number of clones, 17 OTUs total, showing similarity to sequences obtained from the hindgut chamber of the omnivorous tropical reef fish *Pomacanthus sexstriatus* (R. Ward, unpublished). The majority of these clones (76.5%) were recovered from segments IV and V in *N. tonganus*. Unlike *N. tonganus*, *P. sexstriatus* has a defined hindgut chamber, an anatomical feature typically associated with high levels of fermentative activity (16).

While these fish have different feeding strategies and anatomical features, they appear to share related suites of microbial populations.

Bacterial Community Composition Differs in Different Regions of the Intestinal Tract

In *N. tonganus*, the composition and diversity of bacteria differ from one section of the intestinal tract to another. Proteobacteria were found throughout the intestinal tract but do not appear numerically dominant as is found in most other fish systems studied (52, 60). In general, bacterial biomass was lowest in the anterior region (segment II) and most of the sequences recovered aligned with *Cyanobacteria* or *Firmicutes*. The relative proportion of cyanobacterial sequences waned in the more posterior segments. We interpret this trend as representative of residual algae ingested by the fish being digested or refractory algal cells becoming a smaller proportion of the bacteria present. A large percentage of clones (19%) recovered in segment II are related to *Clostridium bifermentans* but similar sequences were not recovered in more posterior regions of the intestinal tract. As *C. bifermentans* ferments amino acids, this population of *C. bifermentans*-like bacteria in *N. tonganus* may contribute to the fermentation end products identified in this intestinal segment.

The majority of clones (>50%) recovered from segment III showed high sequence similarity (> 96%) to *C. somerae*, a Gram-negative bacterium originally isolated from human feces (26) and later found to be predominant isolates from both carp and tilapia (65). This genus appears to be prevalent in gut communities. In pure culture, members of this genus have the ability to ferment a variety of carbohydrates

and some produce large amounts of vitamin B_{12} but their role in the *N. tonganus* microbiome and other intestinal ecosystems has yet to be determined.

In the more posterior segments (IV and V) OTU diversity and the proportion of organisms belonging to the *Firmicutes* and *Bacteroidetes* increased. These phyla also dominated the distal gut of another surgeonfish species, *A. nigricans* (60). Several of the *Firmicutes* clones recovered are related to fermentative bacteria from genera known to possess carbohydrases including *Clostridium*, *Ruminococcus* and *Eubacterium* (27). Members of the *Bacteroidetes* produce a variety of glucosidases, which breakdown polysaccharides found in algae (25) as well as host derived polysaccharides (27). Some *Bacteroidetes* are then able to ferment these substrates producing succinate and propionate (27). The shifts in community composition toward diverse populations of fermentative bacteria in the posterior gut correlate with the short-chain fatty acid data, which show acetate concentration increasing from 3.46 mM and 5.06 mM in segments II and III, to 11.73 mM and 12.69 mM in segments IV and V, respectively (13).

This observed change in bacterial diversity along the length of the *N. tonganus* intestinal tract is in stark contrast to the bacterial diversity observed in the digestive tract of juvenile Atlantic salmon (*Salmo salar*). In these fish, there is little change in diversity among the different gastrointestinal segments, which appear to be dominated by *Pseudomonas* spp. (52). This lack of diversity and dominance by a single bacterial group has been observed in other carnivorous and omnivorous fishes (30, 34, 54, 60). The fact that *N. tonganus* primarily consume algae may explain these differences. This is in agreement with what has been seen in terrestrial mammals and marine fishes

where fecal bacterial diversity was found to increase from carnivory to ominivory to herbivory (42, 60).

T-RFLP Assessment of Community Structure Reveals Some Common T-RFs Shared Between Fish

Comparisons of bacterial community T-RFLP profiles were performed to explore differences between segments in addition to fish-to-fish variation in the two posterior intestinal segments where fermentation end products are highest. There were no statistically significant differences between the segment IV and segment V community structures when all were compared using ANOSIM (Table 3.6). However, when MDS analysis was carried out, the resulting plots did show clustering of segment IV and segment V community profiles. We interpret these results as illustrating some T-RF commonalities within a segment but these similarities are muted by other factors influencing the T-RF profiles. Some of these differences may simply reflect microbiome components that are unique to a particular individual, a feature that has been reported in other gastrointestinal systems (41, 42). The intestinal microbiome of an adult may be influenced by life history, condition and diet of an individual (21, 42, 43). In addition, there are no anatomical features (significant changes in intestinal diameter or presence of sphincters) delineating intestinal segments in *N. tonganus*. Our arbitrary designation of segments does not take into account regional differences (e.g. chemical gradients) that would influence the location of particular populations of symbionts. In addition, the entire contents of segment IV and segment V were sampled for this study. All of these factors could potentially mute community signatures dictated by gut location.

To the best of our knowledge this is the first report using T-RFLP to characterize and compare the microbial community in individual fish. A previous study used this methodology to compare community composition between different segments and size classes of fish, but used pooled DNA from several individuals to generate these data (50). In mammalian systems, gut microbial communities are reported to be relatively stable within an individual, however differences between individuals have been observed (63). In order to assess this in *N. tonganus*, SIMPER analysis was carried out on segments IV and V in all fish. The results indicated low similarity (< 50%) in all but two of the pairwise comparisons, indicating that the overall community structure varies from individual to individual (Table 3.7).

Epulopiscium Influences Community Dynamics

Epulopiscium is typically a conspicuous component of the gastrointestinal microbiota of *N. tonganus*, where it is believed to assist in the breakdown of complex polysaccharides and carry out fermentative metabolism. Using counts of DAPI stained cells, the ratio of other bacteria to of *Epulopiscium* cells in the fish intestinal tract is on average ~17.5: 1 (8012 "other" bacterial cells : 458 *Epulopiscium* cells). Based on previous work, we estimate that a mid-size *Epulopiscium* would have approximately 122,000 copies of a 3.8 Mb genome (48). Therefore in terms of potential genetic impact, a single *Epulopiscium* cell is equivalent to approximately 122,000 normal-sized bacteria. Based on this measure, *Epulopiscium* biomass outweighs other bacteria in the gut by approximately 7000: 1, which suggests that these organisms could be a dominant force in shaping the total community structure and function in *N. tonganus*. To test this, we carried out the same T-RFLP-based analyses on a single sample that

had *Epulopiscium* cells present. *Epulopiscium* cells were manually removed from the sample prior to DNA extraction to reduce the influence that the abundant DNA found in these cells would have on the T-RFLP profile. Since significance levels in ANOSIM analyses are strongly influenced by sample size, we followed the recommendations of Clarke and Warwick (2001) and based our analysis on R-values. The meaning of R-values have been classified as followed: $R > 0.75$ groups are well separated, $R > 0.50$ groups are overlapping but clearly different, and $R < 0.25$ groups are barely separable (9). The R-values indicate that while communities with or without *Epulopiscium* share some common T-RFs, they differ from each other. These results suggest that *Epulopiscium* may displace some of the numerically dominant groups, such as the *Firmicutes* and *Bacteroidetes*. Because *Epulopiscium* cells contain tens of thousands of copies of their genome (48), they may be able to synthesize an adequate number of enzymes, compensating for the lack of those phyla typically associated with fermentative processes.

Several studies looking at fermentation in the gastrointestinal track of marine fishes have utilized culture independent methods, and while the number of these types of studies is increasing, there is still a lack of information on the total microbial community found in tropical marine fishes. This study imparts a better understanding of this community and identifies similarities to those communities found in terrestrial herbivores. Of particular interest was how microbial community composition appears to change along the length of the intestinal tract, an observation that has not been reported in other fish microbiome studies. Using comparative studies of T-RFLP fingerprints we found that while communities were not statistically distinct in the two

posterior segments, multidimensional scaling plots did show clustering of the same segment from different individuals. Finally, it does appear that *Epulopiscium* influences total community dynamics based on the observation that communities found in either segment in fish without these cells differ from the community found in the intestinal tract of an individual that harbored large numbers of these organisms. However this intriguing result needs to be further refined with functional studies with a larger number of fish.

Acknowledgements

We thank the following groups for their contributions to this research: the staff and directors of the Lizard Island Research Station for their support and Will Robbins and David Raubenheimer for their help in the field, Roxanne Beinart for assistance with cloning and sequence analysis, John Rose for assistance with ARB, Bekah Ward for her assistance with statistical analysis, and Christopher Bloch for statistical consultation.

References

1. **Allison, M. J. & Cook, H. M.** 1981. Oxalate degradation by microbes of the large bowel of herbivores: the effect of dietary oxalate. Science. **212:** 675-676.

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

3. **Amann , R. I., Ludwig, W and Schleifer, K.H** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143-169.

4. **Angert, E. R., Clements, K. D. & Pace, N. R.** 1993. The largest bacterium. Nature **362:**239-241.

5. **Benson, S., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D. L** 2004. GenBank: update. Nucleic Acids Res. **32:**D23-36.

6. **Bjerrum L., Engberg, R.M., Leser T.D., Jensen B., Finster K. & Pedersen K.** 2006. Microbial community composition of the ileum and cecum of broiler chicken as revealed by molecular and culture-based techniques. Poult. Sci. **85:** 1151–1164.

7. **Choat, J. H. & Clements, K.D.** 1998. Vertebrate herbivores in marine and terrestrial environments: A nutritional ecology perspective. Annu. Rev. Ecol. Syst. **29:**375-403.

8. **Choat, J. H., Clements, K.D. & Robbins, W.D.** 2002. The trophic status of herbivorous fishes on coral reefs 1: Dietary analyses. Mar. Biol. **140:**613-623.

9. **Clarke, K.R & Gorley, R.N.** 2001. PRIMER v5: User Manual/Tutorial. Plymouth, UK, PRIMER-E.

10. **Clements, K. D.** 1991. Endosymbiotic communities of two herbivorous labroid fishes *Odax cyanomelas* and *Odax pullus*. Mar. Biol. **109:** 223-230.

11. **Clements, K. D.** (1997). Fermentation and gastrointestinal microorganisms in fishes. Gastrointestinal microbiology: Gastrointestinal ecosystems and fermentations. **1:** 156-198.

12. **Clements, K. D. & Bullivant, S.** 1991 An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. J. Bacteriol. **173:**5359-5362.

13. **Clements, K. D. & Choat, J. H.** 1995. Fermentation in Tropical Marine Herbivorous Fishes. Physiol. Zoo. **68:**355 -378.

14. **Clements, K. D. & Choat, J. H.** 1997. Comparison of herbivory in the closely-related marine fish genera *Girella* and *Kyphosus*. Mar. Biol. **127:**579-586.

15. **Clements, K. D., Pasch, I. B. Y., Moran, D. and Turner, S. J.** 2007. Clostridia dominate 16S rRNA gene libraries prepared form the hindgut of temperate marine herbivorous fishes. Mar. Biol. **150:** 431-1440.

16. **Clements, K. D. & Raubenheimer, D.** 2006. Feeding and nutrition. The physiology of fishes. Gainesville, FL. CRC Press.

17. **Clements, K. D., Raubenheimer, D., Choat, J. H.** 2009. Nutritional ecology of marine herbivorous fishes: ten years on. Funct. Ecol. **23:**79-92.

18. **Cole J.R., Chai, B., Marsh, T.L., Farris, R.J., Wang, Q., Kulam, S.A., Chandra, S., McGarrell, D.M., Schmidt, T. M., Garrity, G. M. & Tiedje, J. M.** 2003. The ribosomal database project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. Nuclei Acids Res. **3:**442-443.

19. **Collins, M. D., Lawson, P. A., Willems, A., Cordoba, J. J., Fernandez-Garayzabal, J., Garcia, P., Cai, J., Hippe, H. & Farrow, J. A. E.** 1994. The phylogeny of the Genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int. J. Syst. Bacteriol. **44:** 812-826.

20. **Daly K., Stewart, S. C., Flint, H. J.& Shirazi-Beechey, S.P.** 2001. Bacterial diversity within the equine large intestine as revealed by molecular analysis of cloned 16S rRNA genes. FEMS Microbiol. Ecol. **38:**141-151.

21. **Dethlefsen L., Mcfall-Ngai M. & Relman D.A.** 2007. An ecological and evolutionary perspective on human-microbe mutualism and disease. Nature **449:** 811–818

22. **Dominguez-Bello, M. G.** 1996. Detoxification in the rumen. Ann. Zootech. **45:** 323-327.

23. **Eckburg, P. B., Bik, E. M., Bernstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S. R., Nelson, K. E., & Relman, D. A.** 2005. Diversity of

the human intestinal microbial flora. Science **308:** 1635-1638.

24. **Edwards J. E., Mcewan N. R., Travis A. J. & Wallace R. J.** 2004 16S rDNA library-based analysis of ruminal bacterial diversity. Antonie van Leeuwenhoek **86:** 263–281.

25. **Fidopiastis, P. M., Miyamato, C.M., Jobling, M. G., Meighen, E. A., & Ruby, E. G.** 2006. Characterizing the resident, fermentative microbial consortium in the hindgut of the temperate-zone herbivorous fish, *Hermosilla azurea* (Teleostei : Kyphosidae). Mar. Biol. **148:** 631-642.

26. **Finegold, S., Vaisanen, M. L., Molitoris, D. R., Tomzynski, T. J., Song, Y., Liu, C., Collins, M. D. & Lawson, P. A.** 2003. *Cetobacterium somerae* sp. nov. from human feces and emended description of the genus *Cetobacterium*. Syst. Appl. Microbiol. **26:** 177-181.

27. **Flint, H. J., Bayer, E. A., Rincon, T. M., Lamed, R. and White, B.A.** 2008. Polysaccharide utilization by gut bacteria: potential for new insights from genomic analysis. Nat. Rev. Microbiol. **6:** 121-131.

28. **Frey J. C. , Rothman, J., Pell A. N., Nizeyi J. B., Cranfield M. R. , Angert E. R.** 2006. Fecal bacterial diversity in a wild gorilla. Appl. Environ. Microbiol. **72:** 3788-3792.

29. **Hansen, G. H. & Olafsen, J.A.** 1999. Bacterial interactions in early life stages of marine cold water fish. Microb. Ecol. **38:** 1-26.

30. **Holben, W. E., Williams, P. , Gilbert, M. A., Saarinen, M., Sarkilahti, L.K. & Apajalahti, J. H.** 2002. Phylogenetic analysis of intestinal microflora indicates a novel Mycoplasma phylotype in farmed and wild salmon. Microb. Ecol. **44:** 175-185

31. **Horn, M. M.** 1992. Herbivorous fishes: feeding and digestive mechanisms. Oxford, Clarendon Press.

32. **Huber, T., Faulkner, G. & Hugenholtz, P.** 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. Bioinformatics **20:** 2317-2319.

33. **Karasov, W. H. & Martinez del-Rio, C.** 2007. Physiological Ecology: How animals process energy, nutrients and toxins. Princeton, NJ, Princeton University Press.

34. **Kim, D. H., Brunt, J., & Austin, B.** 2007. Microbial diversity of intestinal contents and mucus in rainbow trout (*Oncorhynchus mykiss*). J. Appl. Microbiol. **102:** 1654-1664.

35. **Kloareg, B. & Quatrano, R. S.** 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. Oceanogr. Mar. Biol. Annu. Rev. **26:** 259-315.

36. **Kuiter, R. H. & Debelius, H.** 2001. Surgeonfishes, Rabbitfishes and their relatives. Chorleywood, TMC Publishing.

37. **Lan, P. T. N. , Hayashi, H., Sakamoto M. & Benno Y.** 2002. Phylogeneic analysis of cecal microbiota in chicken by the use of 16S rDNA clone libraries. Microbiol Immunol **46:** 372-382.

38. **Lane, D. J.** 1991. 16S/23S rRNA seequencing. Chichester, Unkited Kindgom, John Wiley and Sons.

39. **Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R. McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G.** 2007. Clustal W and Clustal X Version 2.0. Bioinformatics **23:** 2947-2948.

40. **Leser, T. D., Amenuvor, T. K., Jensen T. K., Lindecrona R. H., Boye M. & Møller, K.** 2002. Culture independent analysis of gut bacteria: the pig gastrointestinal tract microbiota revisited. Appl. Environ. Microbiol. **68:** 673-690.

41. **Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R.R., Bircher, J. S., Schiegel, M. L., Tucker, T. A, Schrenzel, M. D., Knight, R. & Gordon, J. I.** 2008. Evolution of mammals and their gut microbes. Science **320:** 47-51.

42. **Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I.** 2008. Worlds within worlds: Evolution of the vertebrate gut microbiota. Nat. Rev. Micro. **6:** 776–788.

43. **Ley, R. E., Turnbaugh, P. J., Klein, S. & , Gordon J. I.** 2006. Microbial ecology: human gut microbes associated with obesity. Nature **444:** 1022-1023.

44. **Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, Buchner, A., Lai, T., Steppi, S., Jobb, G., Förster, W., Brettske, I., Gerber, S., Ginhart, A. W., Gross, O., Grumann, S., Hermann, S., Jost, R., König, A., Liss, T., Lüßmann, R., May, M., Nonhoff, B., Reichel, B., Strehlow, R., Stamatakis, A., Stuckmann, N.,**

Vilbig, A., Lenke, M., Ludwig, T., Bode, A. & Schleifer, K. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32**:1363-1371.

45. **Maidak, B. L., Cole, J. R., Lilburn, T. G., Parker, C. T.,Jr, Saxman, P. R., Farris, R. J., Garrity, G. M., Olsen, G. J., Schmidt, T. M. and Tiedje, J. M.** 2001. The RDP-II (Ribosomal Database Project). Nucleic Acids Res. **29:** 173-174.

46. **Majak, W. C. & Cheng, K. J.** 1981. Identification of rumen bacteria that anaerobically degrade aliphatic nitrotoxins. Can. J. Microbiol. **27:** 646-650.

47. **Matsui, H., Kato, Y., Chikaraishi, T., Moritani, M., Ban-Tokuda, T., Wakita, M.** 2009. Microbial diversity in ostrich ceca as revealed by 16S ribosomal RNA gene clone library and detection of novel *Fibrobacter* species. Anaerobe. **16:** 1 - 11.

48. **Mendell, J. E., Clements, K. D., Choat, J. H. and Angert, E.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U S A **105:** 6730-6734.

49. **Moran, D. & Clements, K. D.** 2002. Diet and endogenous carbohydrases in the temperate marine herbivorous fish *Kyphosus sydneyanus*. J. Fish Biol. **60:** 1190-1203.

50. **Moran, D., Turner, S. J. & Clements, K.D.** 2005. Ontogenetic development of the gastrointestinal microbiota in the marine herbivorous fish *Kyphosus sydneyanus*. Microb. Ecol. **49:** 590-597.

51. **Mountfort, D. O., Campbell, J. & Clements, K. D.** 2002. Hindgut fermentation in three species of marine herbivorous fish. Appl. Environ. Microbiol. **68:** 1374-1380.

52. **Navarrete, P., Espejo, R. T. and Romero, J.** 2009. Molecular analysis of microbiota along the digestive tract of juvenile atlantic salmon (*Salmo salar* L.). Microb. Ecol. **57:** 550 - 561.

53. **Nelson K.E., Zinder, S., Hance, I., Burr, P., Odongo, D., Wasawo, D., Odenyo, A. & Bishop, R.** 2003 Phylogenetic analysis of the microbial populations in the wild herbivore gastrointestinal tract: Insights into an unexplored niche. Environ. Microbiol. **5:** 1212–1220.

54. **Pond, M. J., Stone, D. M. & Alderman D. J.** 2006. Comparison of conventional and molecular techniques to investigate the intestinal microflora of rainbow trout (*Oncorhynchus mykiss*). Aquaculture. **261:** 194–203.

55. **Rees, G. N., Baldwin, D. S., Watson, G. O., Perryman, S. & Neilson, D. L.** 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. Antonie van Leeuwenhoek. **86:** 339-347.

56. **Sambrook, J., Fritsch, E. F. and Maniatis, T.** 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

57. **Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J. & Weber, C. F.** 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl. Environ. Microbiol. **75:** 7537-7541.

58. **Scupham, A. J., Patton, T., Bent, E. & Bayles, D. O.** 2008. Comparison of the cecal microbiota of domestic and wild turkeys. Microb. Ecol. **56:** 322- 331.

59. **Skea G., Mountfort, D. & Clements, K. D.** 2005. Gut carbohydrases from the New Zealand marine herbivorous fishes *Kyphosus sydneyanus* (Kyphosidae), *Aplodactylus arctidens* (Aplodactylidae) and *Odax pullus* (Labridae). Comp. Biochem. Physiol. B **140:** 259-269.

60. **Smirga, S., Sandin, S. A. & Azam, F.** 2010 Abundance, diversity and activity of microbial assemblages associated with coral reef fish guts and feces. FEMS Microbiol. Ecol. **73:** 31-42.

61. **Stevens C. E., & Hume, I. D.** 1998. Contributions of microbes in vertebrate gastrointestinal tract production and conservation of nutrients. Physiol. Rev. **78:** 393-427.

62. **Suau, A, Bonnet, R., Sutren, M., Godon, J. J., Gibson, G. R., Collins, M.D. & Doré, J.** 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Appl. Environ. Microbiol. **65:** 4799-4807.

63. **Thompson, C. L., Hofer, M. J., Campbell, I. L. & Holmes, A. J.** 2010. Community dynamics in the mouse gut microbiota: A possible role for IRF9-regulated genes in community homeostasis. PLoS ONE **5:** e10335.

64. **Tolentino-Pablico, G., Bailly, N., Froese, R. & Elloran, C.** 2008. Seaweeds preferred by herbivorous fishes. J. Appl. Phycol. **20:** 933-938.

65. **Tsuchiya, C., Sakata, T. & Sugita, H.** 2008. Novel Ecological niche of *Cetobacterium somerae*, an anaerobic bacterium in the intestinal tracts of freshwater fish. Letters in Appl. Microbiol. **46:** 43-43.

66. **Tsukinowa, E., Karita, S., Asano, S., Wakai, Y., Oka, Y., Furuta, M. & Goto, M.** 2008. Fecal microbiota of a dugong (*Dugong dugong*) in captivity at Toba Aquarium. J. Gen. Appl. Microbiol. **54:** 25 - 38.

67. **Wu, S., Gao, T., Zheng, Y., Wang W., Cheng, Y. & Wang, G.** 2010. Microbial diversity of intestinal contents and mucus in yellow catfish (*Pelteobagrus fulvidraco*). Aquaculture. **303:** 1-7.

CHAPTER 4

WRATH OF THE ROTTING ROOT: STUDENT INVESTIGATIONS INTO THE MICROBIOLOGY OF DECAY

Abstract

Currently, most high schools do not effectively utilize inquiry-based teaching methods to promote learning about the nature of science. Nor do they include explicit lessons in microbiology. Yet, microbiology impacts many aspects of our lives, from disease, to nutrient cycling, to the food industry. Developing effective inquiry-based teaching modules can facilitate student understanding in the area of microbiology and microbiology-related topics can provide a model system for developing inquiry-based activities. With support from the NSF funded Cornell Science Inquiry Project (CSIP) a graduate researcher from Cornell University in Ithaca, NY and a high school teacher from Cortland Junior Senior High School in Cortland New York collaboratively developed and implemented an inquiry based project that looked at the decomposition of carrots, and subsequent isolation of the organisms potentially responsible for the decomposition. The activity engaged students in an open inquiry project to learn about various concepts in microbiology. Students who participated in the project scored statistically higher on a set of questions from the New York State Regents Exam related to the activity, and on the exam overall than those who did not participate in the activity. In addition these students showed a modest gain in retaining information from their biology class as compared to those who did not. These results support the idea that similar inquiry based units should be more commonly used in science

classrooms.

Introduction

It is widely recognized that historically, one of the components lacking in the traditional high school science classroom is inquiry-based and investigative laboratory modules (4, 5, 23, 37). Instead, too often, students participate in "cookbook" laboratory exercises with known outcomes. In this type of lab experience students follow step-by-step instructions, often recording data without understanding the larger question or having opportunities to reflect on their observations (12, 13, 21, 26, 28). In this teacher-centered instruction little emphasis is placed on developing the skills needed to carry out scientific inquiry. Instead, the focus is on a lecture format of teaching that delivers a large number of facts and vocabulary words, which may be memorized by the students (23). Learning and comprehension are then assessed though exams, which often are composed of questions with a single correct answer. This archaic approach, which presents scientific knowledge in terms of proven facts and absolute truths communicated through texts and lectures, encourages students to view science as a static body of knowledge (31). The fact that many educators at the high school level still employ this pedagogical practice has created an environment where students are simply mastering a set of disconnected facts without gaining a broader understanding of the nature of science, or building their critical reasoning and problem solving skills (25, 26, 28). If, on the other hand, students are allowed to actively engage in the process of science, they will understand that scientific knowledge is based on the outcomes of experiments, and these results are not absolute. This inquiry-based approach encourages not only the understanding of known facts

and theories of a specific scientific discipline, but the experimental methods and practices of that science as well (37, 42).

The idea of student-centered, inquiry-based learning in science is not new. A movement to encourage teachers to engage students in this method of instruction dates back to John Dewey in the late 1930s (16). This approach to teaching was again promoted in the 1960s by the Biological Sciences Curriculum Study in its "Invitation to Inquiry". More recent reform efforts by the National Research Council (NRC) resulted in the *National Science Education Standards* (25). These *Standards* suggest that grades kindergarten through twelfth (GK-12) classrooms should create a learning environment where students are able to develop the "abilities necessary to do scientific inquiry" and an "understanding of scientific inquiry" (25). Specifically, inquiry based activities in the science classroom should involve exploring nature to identify meaningful questions, developing and conducting investigations to answer those questions, analyzing and interpreting data from those investigations, and communicating the results of their analyses. These types of learning activities will ultimately allow students to understand how scientists carry out research and foster a deeper and more meaningful understanding of the nature, process and content of scientific inquiry (24, 26, 27, 34).

Despite the vast amount of literature (reviewed in 3) identifying inquiry as a fundamental approach to promote a deeper knowledge of and appreciation for science, many teachers are reluctant to engage in this style of teaching (37). Several studies assessing educator reluctance have identified common barriers to incorporating

inquiry-based activities into their existing curriculum. These include: time required to develop and implement these activities, availability of materials, student abilities and willingness to carry out activities and safety concerns (6, 20, 43). One way to help GK-12 educators overcome these obstacles to inquiry in the classroom is to pair them with a researcher who engages in outreach activities. This "scientist in the classroom" model is frequently used to bring expertise and enthusiasm into the classroom and raise student awareness about careers in science (3, 22, 39, 40). Additional benefits to these programs are an increase in teachers' understanding of science topics, effective teaching methodologies and increased confidence in conducting "real" science (22). The Cornell Science Inquiry Partnership (CSIP) at Cornell University was one such program. CSIP provided an opportunity for graduate students in several different science disciplines to partner with local high school and middle school teachers and collaboratively bring inquiry-based projects into their classrooms.

In most high school biology classes, the curriculum contains very few lessons that explicitly address topics in microbiology. Historically, this discipline has often been overlooked in secondary education (17, 18), yet the field of microbiology interfaces readily with many other disciplines providing the students with a cross-curricular experience. For example, by studying a particular epidemic, the student not only gains an intimate understanding of the disease process and biology of the causative organism, but they are also able to conduct a historical study of the progression of the disease, including advances in treatment as well as gain a sociological perspective on disease, disease prevention and health promotion. Ewald (17) identified several interdisciplinary connections that can be achieved by

incorporating microbiology into the high school science curriculum, from Art (observations, enlarging and drawing to scale) to Earth Science (decomposition and soil health, mineral cycles and mining) to Mathematics (serial dilutions, colony counts and exponential numbers).

In addition to assisting students in gaining knowledge in other disciplinary areas, incorporation of microbiology into the classroom assists students in attainment of general educational goals of the high school curriculum (35). Investigative lab exercises focused on microbiology can help students develop their observational, analytical and critical thinking skills, as well as provide a direct means for implementation of the scientific method. Employing debates on ethically controversial topics can improve student's abilities to formulate and support an argument as well as refine public speaking skills. Finally, case study-based activities provide a conduit for increasing writing in the sciences (32, 35).

Many high school students can readily see the relevance of microbiology in their everyday lives. Food-borne illnesses are frequently in the news, and many students will most likely be able to recall that "*E. coli"* was responsible for the outbreak. Teens are also fascinated by shows like *CSI*, *ER* and *House, MD*, which frequently draw on the principals and topics of microbiology. These medical and genetic topics engage teenagers and could serve to increase the number of students who ultimately end up pursuing a career in science, which is currently below the number needed to meet future employment demands (17). Despite their student's interest in the subject there are several challenges that high school teachers face when

introducing microbiology into their science class curriculum. Microbiology is frequently perceived as a specialized discipline that requires specialized, often costly equipment that high schools do not possess. Often, high school teachers feel ill prepared to teach the subject of microbiology having never taken a microbiology course during their undergraduate careers. Finally, some may continue to believe that all microbes are "bad". In fact, an article on teaching microbiology at the high school level published 50 years ago stated "…it is dangerous to work with microorganisms because we all know most microorganisms are deadly devastating demons of destruction" (18).

While great strides have been made to overcome these barriers, such as kits that allow for the implementation of activities without specialized equipment, detailed protocols and background information and an increase in public awareness of the beneficial activities of microorganisms, there is still a long way to go. Fortunately, there are many college and university summer or year-long outreach programs that enable GK-12 teachers to gain more knowledge and confidence in understanding microbiology. In central New York, for example, Cornell University (10, 11) and Hobart and William Smith Colleges (19) offer opportunities for high school educator professional development. In Northern New York, Clarkson University and St. Lawrence University (9) place undergraduate science majors with tenured middle and high school science teachers and a faculty member at the university. Together the triad develops meaningful short-term, inquiry-based activities for implementation in the classroom.

Despite the abovementioned concerns, developing and implementing active learning modules around topics in microbiology will not only better secondary education, it may also assist in the successful transition of students into science majors at the undergraduate level. Even those students who do not pursue a career in science still need to become 'scientifically literate': learn how to reason scientifically, think critically, evaluate and validate multi-media claims and use evidence in a logical manner to draw conclusions and make decisions (1, 2, 7, 25, 26, 33). These skills will allow these citizens to understand and make informed decisions around critical modern issues such as emerging disease, vaccinations, GMOs and bioremediation.

The goal of this project was to simultaneously address both a lack of inquiry-based, investigative activities and the absence of curriculum that relates to microbiology in the high school setting. Supported by the NSF GK-12 funded Cornell Science Inquiry Partnerships (CSIP) program, this project allowed the partnering of a graduate student in the Department of Microbiology at Cornell University in Ithaca, NY with a high school biology teacher at Cortland Junior-Senior High School in Cortland, NY to develop and implement a scientific inquiry unit which allowed students to explore various factors involved in decomposition of organic matter (a carrot), the potential role of microbes in this process and assess various biochemical abilities of these organisms.

Materials and Methods

Four different biology classes consisting of 18, 21, 18 and 19 students participated in this 6-week long-guided inquiry process. Table 4.1 summarizes the

Table 4.1. Sequence of lessons

Day 1	(preferably a double period): Introduce the process of decomposition, brainstorm about factors that could influence the rate of decomposition, develop data collection/recording sheet, and complete experimental design.
Day 2:	Set up microcosms, take initial masses of carrots and microcosms (in case carrots decompose too quickly, students can mass the microcosm and still collect data over several weeks), make baseline observations, review data collection procedure.
Bi-weekly:	Have students mass their carrots and microcosms. Students should be making physical observations of the carrot (smell, texture, appearance) and also thinking about *why* they are seeing a change in mass over time (whether this is a gain or loss of mass).
After carrot decomposes:	Have students isolate bacteria and fungi on the carrot surface by swabbing it with a sterile swab. Have the students set up a serial dilution to isolate the bacteria and fungi from the soil.
2-3 days after bacterial and fungal isolation:	Have students describe the colony morphology (shape) of the different bacteria and fungi they have isolated. They should talk about the color, size, shape, edge, elevation and texture of the colonies. Based on unique colony morphology, have the students isolate two different bacteria from their original bacterial isolation plates.
2-3 days after isolating 2 separate bacteria:	Have a brainstorming session about what a carrot is made up of and how that could influence the types of enzymes the decomposing bacteria and the advantage of being able to exploit a wide a wide variety of resources, such as dead bacteria. Have the students come up with hypotheses about the different biochemical tests. Have the students inoculate the biochemical tests.
2-3 days after inoculating the biochemical tests:	Have the students read the tests, record the results and see if their hypotheses were correct. If they were not, have them speculate *why*.

sequence of lessons in this unit. Throughout the unit, students completed a series of worksheets to guide their inquiry, some individually and some as a group (Appendix B). These worksheets addressed both the experiments performed as well as assisting the students in making connections to real world issues.

The teacher's roles included giving short lectures introducing the information required to carry out the days activities, summarizing learning, facilitating brain storming sessions and assisting students in drawing connections to real-world events and other scientific processes. In addition, the teachers acted as mentors as well as collaborators on the project to assist students in assuming the role of a scientist engaging in inquiry (12 - 14).

Class Meeting 1

During the first meeting, the students were introduced to the process of decomposition, and the various factors (physical, chemical and biological, abiotic and biotic) that can influence the rate of decomposition. Particular emphasis was put on the role of bacteria and fungi in this process. Ways of assessing the rate of decomposition in the classroom were discussed (such as utilizing the change in the mass as an indicator of decomposition) and the scientific method, including the idea of controls and replicates in scientific experimentation was reviewed. The basic experimental set up was described. Students would be given dirt and clear plastic zip-top bags to create two types of soil microcosms (one experimental and one control and a carrot, which would serve as the organic material being decomposed. A variety of "conditions" were made available to the students. These included several different soil

types (potting soil, compost and sterile dirt), varying levels of temperature, light and oxygenation and slightly acidic (diluted vinegar) and basic (baking soda) solutions. Students were then separated into groups and asked to come up with a hypothesis about factors that affect the rate of decomposition and develop an experiment to test this hypothesis. As a group, with the teacher and graduate student as facilitators, a data collection sheet was then developed (Table 4.2).

Class Meeting 2

During the second class meeting, the scientific method and importance of controls were reviewed. The students set up their microcosms and took the initial mass of their carrot and microcosm (in case the carrot decomposes too quickly, this allows the students to continue to collect data over the course of several weeks). Students were encouraged to make baseline observations of the carrot (color, texture, smell) and the soil (color, texture, moisture, smell, presence of invertebrates and organic matter, etc.), reinforcing the importance of observation and careful data collection in science.

Bi-weekly

Students began the process of assessing for decomposition of their carrot by massing the carrot. After carefully removing the carrot with sterile forceps, any dirt was brushed off the surface. The carrot was then placed on a sterile piece of weigh paper on a triple beam balance and the mass was recorded. The change in mass for that day, and the total change in mass was calculated and recorded. Observations of the

Table 4.2. Example Data Collection Sheet

	Experimental	**Experimental**	**Experimental**	**Experimental**	**Experimental**	**Experimental**
Date	Mass Carrot 1	D Mass Carrot 1	Mass Carrot 2	D Mass Carrot 2	Mass Carrot 3	D Mass Carrot 3
	Overall D Mass Carrot 1		Overall D Mass Carrot 2		Overall D Mass Carrot 3	

carrot were made during each of these massing events and recorded. Students were encouraged to take pictures as an additional way to record the change in appearance of their carrot.

Once decomposition began

To investigate the microorganisms potentially responsible for decomposition of the carrot, students isolated bacteria and/or fungi from the surface of the carrot. Sterile swabs were used to carefully swab the surface of the carrot. These swabs were then placed in sterile buffer. Serial dilutions (1:10 and 1:100) were prepared, and swabbed onto the surface of Luria Bertania (LB) agar, which selected for bacteria and Sabaroud Dextrose Agar (SDA) agar, which selected for fungi. The plates were inverted and incubated at room temperature for 3 days.

Once growth appeared on the plate

After three days, students made observations of their fungal and bacterial isolation plates. At the beginning of class, a brain storming session was conducted to generate a list of descriptive terms to describe the growth observed on their isolation plates. Using these and standard colony morphology descriptions used in any microbiology class, students described the color, size, shape, edge, elevation and texture of the colonies. Students continued to observe the plates and record changes over the next week.

Once the students had isolated colonies, they selected two morphologically unique colonies and streaked them on plates to obtain a pure culture. This pure culture plate was then used for inoculating media for biochemical testing.

Once growth appeared on the pure culture plates

Teachers facilitated a brain storming session to identify the biologically important macromolecules that are present in a carrot and how these could influence the types of enzymes the decomposing bacteria could potentially possess. Other topics addressed were the type of environment these organisms are found in, and the advantages of being able to exploit a wide variety of resources, such as other cells in the environment. After this brainstorming session, students developed hypotheses about the four different biochemical tests available (Glucose Fermentation, Sucrose Fermentation, Starch Hydrolysis and DNA Hydrolysis). Students then inoculated the biochemical tests with their isolates.

Once growth appeared in/on biochemical tests

Teachers reviewed the biochemistry behind the tests, and explained how to interpret the tests. Students then worked in groups to assess their tests, and determine if their hypotheses were correct or not. If they were not, then they speculated why they weren't correct.

Student assessment and survey

Education research design employed to assess this unit utilized a mixed-methods approach, including quantitative post science content tests and responses to a

partially open ended questionnaire designed to elicit student reaction to the unit. Additional data included the authors' journal entries written after the lessons, teacher feedback about student learning in both inquiry based and traditional classroom settings and student's final projects. To specifically assess how this guided inquiry unit affected student learning, questions in the 2006 New York State Living Environment Regents Exam (29) were identified that directly related to the content covered in this module. The exam in its entirety can be accessed at: http://www.nysedregents.org/LivingEnvironment/20060621exam.pdf. Since this exam was given to all biology students at Cortland Junior-Senior High School, we were able to assess how students that participated in this inquiry-based module scored as compared to those who did not. We also administered a post-test two years after the activity to assess retention. This post-test was administered to students who were enrolled in biology during the year the activity was implemented. The questions we designed related to concepts covered in this activity, but also more broadly to New York State and National Science Education Standards. Feedback from the teacher was also solicited and included observations of student's attitudes, engagement with the activity and abilities post activity.

Data analysis

Identical questions were used to calculate the test grade (%) for students who had participated in the inquiry-based activity (n = 37) and those who had not (n = 107) for the total Regents test score and for the subset of questions specifically related to the unit. In addition, final averages for all students in these classes were also calculated.

These data were then subjected to independent sample one-tailed t-tests to determine whether there was a statistically significant difference between the scores obtained by students who had participated in the unit and those who had not. Overall scores of the post-test were determined for students who participated in the inquiry-based unit and those who had not. An independent sample one-tailed t-test was used to determine whether overall test scores differed between the two groups. In order to further explore the relationship between the post-test scores obtained by students who had participated in the inquiry based unit, and those who had not effect size statistics (Cohen's *d*) were calculated.

Results and Discussion

The Activity Addresses Key Science Education Standards

Despite the vast amount of literature supporting curricular changes to incorporate inquiry-learning modules in the high school classroom, many teachers are still reluctant to do so. In addition to the previously mentioned barriers, another reason for reluctance is the fear of not addressing key education standards, and preparing students to succeed on state administered standardized tests (42). Test preparation is a major concern of teachers in New York State, where passing the Regents Living Environment Exam has become a requirement for graduation (41). Despite the fact that this activity was not designed with this criterion in mind, it did address several NYS Science Education Standards required by the Regents exam, as well as many of key National Science Education Standards (Table 4.3). Aside from addressing the most obvious standard of being able to understand the process of science and scientific

Table 4.3. New York State and National Science Education Standards

NYS Science Education Standards Addressed

Standard 1: Students will use mathematical analysis, scientific inquiry, and engineering design, as appropriate, to pose questions, seek answers, and develop solutions.

Key Idea 1: The central purpose of scientific inquiry is to develop explanations of natural phenomena in a continuing, creative process.

Key Idea 2: Beyond the use of reasoning and consensus, scientific inquiry involves the testing of proposed explanations involving the use of conventional techniques and procedures and usually requiring considerable ingenuity.

Key Idea 3: The observations made while testing proposed explanations, when analyzed using conventional and invented methods, provide new insights into phenomena.

Standard 4: Students will understand and apply scientific concepts, principles, and theories pertaining to the physical setting and living environment and recognize the historical development of ideas in science.

Key Idea 1 Living things are both similar to and different from each other and from nonliving things.

Key Idea 5: Organisms maintain a dynamic equilibrium that sustains life.

Key Idea 6: Plants and animals depend on each other and their physical environment.

National Science Education Standards Addressed

Science as Inquiry

Abilities necessary to do scientific inquiry
Understandings about scientific inquiry

Life Science

Interdependence of organisms
Matter, energy, and organization in living systems

Earth and Space Science

Geochemical cycles

Science and Technology

Abilities of technological design and understandings about science and technology

inquiry, several activities addressed additional standards. For example, the biochemical tests addressed NYS Science Education Standard 4, Key Idea 1, Performance Indicator 1.2, Major Understandings f, g and h, as well as Performance Indicator 1.3, Major Understanding a. These Standards address cell structure and function, metabolic pathways and nutrient processing. This single activity engaged students in understanding that a single cell organism is able to utilize several substrates and energy sources. Additional curriculum alignment revealed that this activity also met 15 of the 17 required lab technical skills from Appendix A of the NYS Living Environment Curriculum (29).

Students Who Participated in the Inquiry-Based Activity Performed Significantly Better on the High Stakes Regents Exam

To assess how this activity impacted learning, two out of the four classes (one morning class and one afternoon class) were randomly selected for each of the four biology teachers who taught the Living Environment curriculum during the 2005-2006 academic year. Overall Regents Exam scores, as well as scores for the targeted Regents Exam questions were calculated for all students in the selected classes. Three of these teachers did not participate in the inquiry-based activity and their classes served as the control group. To determine if the three teachers classes could be treated as a single control group, a One-way ANOVA was conducted. There were no statistically significant differences between the groups ($F(2,104) = 0.07$, $P = 0.934$) allowing them to be treated as a single group in subsequent statistical analysis. Eleven questions from the 2006 Living Environment Regents Exam were chosen based on their overall relevance to both the content of and process of science

utilized in the activity. Questions 19 and 28 addressed metabolic processes, the assimilation of nutrients into cellular components and energy transfers in ecosystems. These themes were carried through the entire activity, from exploring the process of decomposition and the assimilation of nutrients into biomass to assessing for specific metabolic pathways by the decomposers. Question 42 asked students to identify organisms that acted as decomposers in the environment. In the activity, students not only identified the types of organisms that acted as decomposers in the environment, but also isolated them from their soil microcosms. The last set of questions, 39, 43-47, 57 and 74 asked students to either design an experiment, develop a hypothesis or interpret data from an experiment. All of these skill sets were reinforced continually throughout the activity as students made observations about decomposition, asked questions about this process, developed a testable hypothesis, identified controls, carried out the experiments and recorded, interpreted and graphed the data they collected.

The total average for the eleven Regents questions identified as aligned with the activity was calculated for each student in both the control and experimental groups and subjected to a one-tailed two sample t-test. There was a significant gain in the scores for students who participated in ($M = 0.754$, $SD = 0.181$) as compare to those who did not participate in ($M = 0.639$, $SD = 0.198$) the inquiry-based activity; $t(68) = 3.26$, $p = 0.001$. These results suggest that the use of inquiry-based activities positively influences students overall comprehension of the information presented in their courses. In addition, this type of pedagogical approach may facilitate retention of course material (38). The average for each group (control and experimental) was

calculated based on the results of the entire 2006 Living Environment Regents Exam. Again, a significant gain was noted in the scores for the students who participated in ($M = 0.772$, $SD = 0.118$) as compared to those who did not participate in ($M = 0.724$, $SD = 0.125$) the inquiry based activity; $t(66) = 2.10$, $p = 0.020$). These results indicate that students that participate in these activities are able to apply not only the content, but also the concepts learned to other topics not directly related to the inquiry-based project. This ability is due to the inherent nature of the inquiry model, which engages students in tangible, hands-on experiences where they have to utilize critical thinking skills to develop questions and solve problems often applying information to distantly related concepts. Ultimately, this type of learning promotes the transfer of concepts to new problem questions. Other studies that have looked at student gains in inquiry-based classrooms versus traditional classrooms indicated that inquiry- based science instruction better prepares students to succeed, as measured by standardized tests in Earth Science (8) Regents Biology (21) and eighth grade general science (36). One study assessing students in middle school science classes reported statistically significant increases on high stakes standardized test scores in cohorts that participated in inquiry-based learning science classrooms when compared with students who were taught in traditional classrooms (24).

Retention of Specific Information Showed an Effect Size

In order to assess retention of the materials covered in the inquiry-based unit, as well as the Living Environment curriculum in general, a post-test was administered approximately two years after students had completed the unit in their biology class

(Table 4.4). The average for each group (control and experimental) was calculated based on the results of the post-test and subjected to a One-tailed two sample t-Test. A gain approaching significant was noted in the scores for the students who participated in ($M = 0.602$, $SD = 0.166$) and those who did not participate in ($M = 0.554$, $SD = 0.160$) the inquiry based activity; $t(81) = 1.61$, $p = 0.055$). To better understand these data, effect size statistics (Cohen's *d*) were calculated. When the averages of the students were used in this analysis, an effect size of 0.29 was obtained. Cohen labeled an effect size of 0.20 as small, one of 0.50 as medium and an effect size of 0.80 as large. However, it has been noted that effect sizes in education tend to be lower, and as such using Cohen's benchmarks may be misleading (41). The results we obtained are similar to other studies that looked at the effect size between students that had participated in inquiry-based learning and those who did not.

One of the reasons why we may not have seen a larger effect size with this assessment relates to the assessment itself. Several of the questions were highly specific in nature. The nature of inquiry emphasizes understanding the process of science, not the memorization of facts. As such, the skills and concepts learned in inquiry based activities are not easily tested by multiple-choice or true/false format. Regardless, inquiry-based instruction had other positive impacts as evidenced by student comments and reflections of the teacher.

Inquiry-based Learning Positively Impacted Both the Students and Teacher

Several of the students that participated in the inquiry-based activity had the opportunity to present the results of the project as a poster at the Student Science

Table 4.4. Post-test to assess retention

Gender Male _____ Female _____
Was Mr. Reed your biology teacher? Yes _____ No _____
Did you present a poster on carrot degradation at Cornell University for the CSIP Student Congress? Yes _____ No _____
Read each question and mark an "x" under the answer that best describes your knowledge of the question:

	True	False	Don't Know
1. All bacteria cause disease.			
2. Bacteria and fungi are NOT important in the decomposition of organic matter.			
3. Decomposition occurs under aerobic (oxygenic) conditions, only.			
4. Measuring the change in mass of an object is an effective way to determine if that object is being degraded.			
5. Biochemical processes and metabolic activities carried out by bacteria and fungi help to maintain the atmosphere of our planet.			
6. All decomposers must ingest their food before digesting it.			
7. Some bacteria are able to break down DNA and use it as an energy source.			
8. Fermentation is performed only by yeast cells during the production of ethyl alcohol (ethanol) and carbon dioxide.			
9. When bacteria breakdown glucose or sucrose in media to an acidic end product, then phenol red, also present in the media will turn it yellow.			
10. In an experiment, the dependent variable is the one that you are measuring.			
11. When conducting an experiment, it is always important to include controls in the experimental design.			
12. *Epulopiscium* is a small bacterium that lives in the digestive system of coral polyps in Australia's Great Barrier Reef			

Conference held at Cornell University. To assess how the activity impacted these individuals, interviews were conducted two years post-activity. The results revealed that these students enjoyed the activity, found the learning skills they acquired useful in other courses and felt that they learned and retained more during this activity than other lab modules.

Students strongly agreed that (1) they enjoyed working in groups, (2) more labs should be "design your own experiment based" and (3) designing their own experiments helped them to better understand controls, variables and the process of science. One student, when asked if there was anything else she wanted to share said, "All science teachers should do this". These results, as well as student progress and observations by their teacher and the graduate student indicate that students were engaged with the activity (30).

In the classroom there was visible excitement when students collected their data for the rate of decomposition. Students would discuss their data, and try to explain their results. Some even asked to repeat the experiment with different conditions. This desire to repeat a lab without explicit procedures or requirement clearly indicates student engagement (5). Other studies have highlighted the importance of allowing students to choose their own question, a characteristic of inquiry (12, 15). All of the students interviewed stated that designing their own experiment helped them to better understand the science behind the activity.

Not only did the students benefit from this activity, but the teacher did as well. Several lines of evidence support the positive effects of teacher – graduate student

pairings (22, 39, 40). During this collaboration, we were able to transform a traditional classroom, where lecture was the norm, into one where student inquiry was the driving force. During these lessons, students held themselves and one another accountable for their performance. Frequently, students could be heard explaining difficult concepts to their peers, and according to the teacher, true leadership skills arose out of some, cooperation became the norm rather than the exception and they began to think about what they were doing. Beyond allowing his students to experience true scientific inquiry, the teacher's desires to return to research were also reignited. After this partnership, he has spent a year studying geospatial information technologies, received a teacher-research fellowship through the Environmental Science Education Partnership and spent six weeks conducting biomedical research at Cornell University. Most importantly there is a greater understanding that inquiry is absolutely essential to the success of his students in becoming literate science citizens.

Conclusions

By providing students with the opportunity to conduct genuine investigations through discovery, they are more likely to take ownership of their learning and be engaged in the process. The findings from this educational study strongly suggest that employing inquiry-based teaching pedagogies in the classroom is an effective approach to engaging students, and allowing a wide range of students to gain a deeper understanding of the subject matter being taught, as well as the process of science.

The activities described herein gave the students an opportunity to engage in inquiry, make connections between seemingly disparate topics, and develop an in-

depth understanding of those topics. This was supported by the gains seen in the Regents Exam by students who participated in the activity, compared to those who did not. In summary, these finding suggest that inquiry-based education is an effective approach to engage students in critical thinking, and problem solving while maintaining their interest in science.

Acknowledgements

I would like to thank the following people for their contributions to this research: Ron Reed who allowed me to partner with him in his classroom, Nancy Trauttman for her suggestions for analysis, Esther Angert for allowing me to use equipment in her lab and Christopher Bloch for statistical consultation. This work was supported by the National Science Foundation (NSF) under the GK-12: Cornell Science Inquiry Partnerships (CSIP) Grant DGE 0231913 and by Cornell University.

References

1. **AAAS**. 1989. Science for all Americans: Project 2061 Report American Association for the Advancement of Science. Washington, DC.

2. **AAAS**. 1993. Benchmarks for science literacy. Oxford University Press. New York, NY.

3. **Alberts, B. M.** 1993. Science Education Partnerships. Manual for Scientists and K-12 Teachers. San Fransisco, CA.

4. **Anderson, R. D.** 2002. Reforming science teaching: What research says about inquiry. J. Sci. Teach. Ed. **13:** 1-12.

5. **Backus, L.** 2005. A year without procedures. The Sci. Teach. **72:** 54-58.

6. **Baker, W. P., Lang, M. & Lawson, A. E.** 2002. Classroom management for successful student inquiry. The Clearinghouse. **75:** 248-252.

7. **Brickman, P., Gormally, C., Armstrong, N. & Hallar, B.** 2009. Effects of inquiry-based learning on students' science literacy skills and confidence. Int. J. Scholar. Teach. Learn. **3:** 1-22.

8. **Chang, C. Y., & Mao, S. L.** 1998. The effects of an inquiry-based instructional method on earth science students' achievement. Annual Meeting of the National Association for Research in Science Teaching. San Diego, CA.

9. **Clarkson Univeristy.** 2010. Project Based Learning Partnership Program. Online. http://www.clarkson.edu/highschool/k12/.

10. **Cornell University**. 2010. Cornell's Learning Initiative in Medicine and Bioengineering. Online. http://climb.bme.cornell.edu/.

11. **Cornell University.** 2011. Cornell Institute for Biology Teachers. Online http://cibt.bio.cornell.edu/.

12. **Crawford, B. A.** 2000. Embracing the essence of inquiry: New roles for science teachers. J. Res. Sci. Teach. **37:** 916-937.

13. **Crawford, B. A.** 2007. Learning to teach science as inquiry in the rough and tumble of practice. J. Res. Sci. Teach. **44:** 613-642.

14. **Crawford, B. A.** 2008. Situating learning science in authentic settings: Moving students to the inner circle. American Educational Research Association Conference. New York, NY.

15. **Crawford, B. A., Krajcik, J. S. & Marx, R. W**. 1999. Elements of a community of learners in a middle school science classroom. Sci. Ed. **83:** 701-723.

16. **Dewey, J.** 1933. How We Think. A restatement of the relation of reflective thinking to the educative process (Revised). D. C. Heath. Boston, MA.

17. **Ewald, H.** 1997. A Microbiology Curriculum for K through 12 Grades. Annual Meeting of the American Society for Microbiology. Miami, FL.

18. **Fagle, D. L.** 1961. Teaching Microbiology in High School Biology. Am. Biol. Teach. **23:** 42-45.

19. **Hobart and William Smith Colleges.** 2011. Crossing Boundaries. Online. http://www.crossingboundaries.org/index.php.

20. **Jackson, D. & Boboc, M.** 2008. Facilitating an inquiry-based science classroom. Sci. Scope. **32:** 64-67.

21. **Kaiser, B.** 1996. Teachers, Research and Reform: Improving Teaching and Learning in High School Science Courses. Global Summit on Science and Science Education. San Francisco, CA.

22. **Laursen, S., Liston, C., Thiry, H. & Graf, J.** What good is a scientist in the classroom? Participant outcomes and program design features for a short-duration science outreach intervention in K -12 classrooms. CBE Life Sci. Ed.. **6:** 49-64.

23. **Leonard, W. H. & Chandler, P. M.** 2003. Where is the inquiry in biology textbooks? The Biol. Teach. **65:** 485-487.

24. **Marx, R. W., Blumenfeld, P. C., Krajcik, J. S., Fishman, B., Soloway, E. Geier, R. & Tal, R. T.** 2004. Inquiry-based science in the middle grades: Assessment of learning in urban systemic reform. J. Res. Sci. Teach. **41:** 1063-1080.

25. **National Research Council.** 1996. National Science Education Standards. National Academy Press. Washington, D.C.

26. **National Research Council**. 2000. Inquiry and the National Science Education Standards. National Academy Press. Washington, D.C.

27. **National Research Council.** 2003. BIO 2010. Transforming Undergraduate Education for Future Research Biologists. National Academy Press. Washington, D.C.

28. **National Research Council.** 2005. How Students Learn: Science in the Classroom. National Academy Press. Washington, D.C.

29. **NY State Education Department.** 2006. State Assessment – Science Regents Examinations [Online] http://www.nysedregents.org/testing/scire/regentearth.html

30. **Nystrand, M., & Gamoran, A.** 1991. Instructional discourse, student engagement, and literature achievement. Res. in the Teach. of Engl. **25:** 261-290.

31 **Roth W. & Roychoudhury, A.** 2003. Physics students' epistemologies and views about knowing and learning. J. Res. Sci. Teach. **40(supp):** S114-S139.

32. **Schneider, R.M., Krajcik, J., Marx, R.W. & Soloway, E.** 2002. Performance of students in project-based science classrooms on a national measure of science achievement. J. Res. Sci. Teach. **39:** 410-422.

33. **Shipman, H.L.** 2004. Inquiry Learning in College Classrooms, p. 357-387, *In* L. B. Flick and N. G. Lederman, eds. Scientific Inquiry and Nature of Science: Implications for Teaching, Learning, and Teacher Education. Kluwer Academic Publishers. Boston, MA.

34. **Singer, J., Marx, R.W., Krajcik, J. & Chambers, J.C.** 2000. Constructing extended inquiry projects: Curriculum materials for science education. Edu. Psychol. **35:** 1-35.

35. **Smith, H.** 2007. A case for microbiology in secondary education and implication for undergraduate educators. FOME. **14:** 8-9.

36. **Stohr-Hunt, P.** 1996. An analysis of frequency of hands-on experience and science achievement. J. Res. Sci. Teach. **33:**101-109.

37. **Timmerman, B.E., Srickland, D.C. & Carstensen, S. M.** 2008. Curricular reform and inquiry teaching in biology: where are our efforts most fruitfully invested? Integr. Comp. Biol. **48:** 226-240.

38. **Tretter, T.R.** & **Jones, M.G.** 2003. Relationship between inquiry based teaching and physical science standardized test scores. Sch. Sci. Math.. **103:** 345-350.

39. **Trautmann, N. M. & Kransy, M. E.** Integrating Teaching and research: A new model for graduate education. BioScience. **56:** 159-165.

40. **Trautmann, N. M & MaKinster, J. G.** 2005. Teacher/Scientist Partnerships as Professional Development: Understanding How Collaboration Can Lead to Inquiry. AETS International Conference. Colorado Springs, CO.

41. **Valentine, J. & Cooper, H.** 2006. Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. What Works Clearinghouse. Washington, DC.

42. **Veronesi, P., & Voorst, C. V.** 2000 Science Teacher Beliefs: Toward an Understanding of State Science Exams and Their Influence on Teacher Belief. Annual International Conference of the Association for the Education of Teachers in Science. Akron, OH.

43. **Wheeler, G.F.** 2000. Inquiring into inquiry learning and teaching in science. Association for the Advancement of Science. Washington D.C.

CHAPTER 5

SUMMARY AND FUTURE WORK

Summary

The studies presented herein have provided an increased understanding into the cell biology and ecological niche of the unique bacterium, *Epulopiscium* Type B. By quantifying the amount of DNA in a known number of cells from two different cell populations (small cells containing immature offspring and large cells containing fully developed offspring, Chapter 2, figure 2.1) it was shown that large *Epulopiscium* Type B cells contain approximately 250 pg of DNA, while small cells contain approximately 85 pg of DNA. Single-cell, quantitative PCR assays of the genome proxy genes *ftsZ, dnaA* and *recA* demonstrated that individual *Epulopiscium* Type B cells contain tens of thousands of copies of their genome. A similar quantitative PCR approach was utilized with quantified genomic DNA. These analyses revealed that the genome of *Epulopiscium* Type B is approximately 3.8 Mb. When the unit genome proxy *ftsZ* was assayed in cells of known volume, a positive correlation between genome copy number and cytoplasmic volume was observed over a wide range of cell sizes. However, *Epulopiscium* Type B maintains a larger cytoplasmic volume to genome ratio than more typical rod shaped bacteria, such as *Bacillus subtilis*. The abundant amount of DNA seen in *Epulopiscium* Type B would appear to be an adaptation, which allows these organisms to maintain an unusually large cell size (29).

Gene diversity in *Epulopiscium* Type B was assessed through single-cell PCR amplification of the single copy genes *ftsZ, dnaA* and *recA.* For both *ftsZ* and *recA* a consensus sequence was identified and the majority of clone variants possessed single

nucleotide changes that appeared to be transitions encoding for nonsynonymous amino acid substitutions. The overall frequency of these differences was comparable to published *Taq* DNA polymerase error rates. *dnaA* exhibited similar trends, however in 12 of the 30 *dnaA* clones a single-nucleotide deletion in a mononucleotide tract of 10 adenines was observed. In order to determine if these deletion mutations were generated during PCR amplification, a plasmid borne copy of this mononucleotide tract was amplified with a high-fidelity mix of thermostable polymerases and the sequence of the PCR product determined. The electropherograms showed minor peaks after the poly-A tract from the subpopulation of genes carrying the deletion. When the sequence of the plasmid-borne *dnaA* gene was determined directly, this subpopulation of variants was not observed. This indicates that at least some of the single-nucleotide deletions observed in the mononucleotide tract were produced during amplification with thermostable polymerases. In bacteria, homopolymeric (HP) tracts have been associated with genes that encode variable surface proteins allowing for phase variation in pathogens (5, 23). Frameshifts in HP tracts result from slipped-strand mispairing during DNA replication and can lead to reversible gene inactivation (23). In essential and highly expressed genes, however, HP tracts are rare (4). The DnaA protein is required for the initiation of DNA replication and as such is considered an essential protein (19). In *Epulopiscium* the deletion variant of the *dnaA* gene, if translated, would result in a truncated peptide that lacked the DNA binding domain. This ultimately could affect replication initiation in the cell.

While there have been many studies done on the diversity of the mammalian gut microbiome, the diversity in fish remains largely unexplored (9, 13, 14,15). To

better understand the intestinal microbiome in the host fish of *Epulopiscium,* a 16S rRNA gene clone library was generated from the contents of 4 segments of the intestinal tract of 4 individual *N. tonganus*. The clones recovered were highly diverse with the majority associating with the phyla Firmicutes, Bacteroidetes and Fusobacteria, phyla associated with intestinal clone libraries from a wide range of animals (27, 28). A comparison of the community structure in the most posterior two segments of the *N. tonganus* intestinal tract using T-RFLP revealed no statistically significant differences between fish or segments. However when these data were subjected to non-metric Multi Dimensional Scaling (MDS) analysis, the resulting plots showed clustering of segments indicating some common T-RFs within the segments (11). These data were generated from fish that did not harbor populations of *Epulopiscium* Type B. To investigate the potential role of this large symbiont in influencing community composition, an individual fish that had a dense population of *Epulopiscium* cells was used in T-RFLP analyses. The data generated from these analyses were subjected to Analysis of Similarities (ANOSIM) testing. Statistical differences were observed between the communities found in the fish that harbored *Epulopiscium* Type B and those that did not, suggesting that *Epulopiscium* Type B and its vast genomic resources may replace numerically dominant bacteria. This study was by no means an exhaustive analysis of the community found in *N. tonganus*, however it does begin to increase our understanding of the intestinal microbiome in marine fish. To better understand the influence of *Epulopiscium* on the total microbial community, additional fish harboring theses unique and exciting organisms need to be analyzed.

Future Work

Transcriptional Analysis of Carbohydrase Genes

Epulopiscium sp. type B are frequently a conspicuous component of the intestinal microbiota of their surgeonfish host, *N. tonganus* (2, 12, 16). *N. tonganus* are considered omnivores, with a diet dominated by algae (9, 10) which includes fleshy algae, thallose chlorophyte and rhodophyte algae and filamentous rhodophyte algae as well as benthic invertebrates. Marine algae are comprised of a chemically diverse group of polysaccharides, differing from their terrestrial counterparts both in complexity and composition (25). These complex carbohydrates include agar polymers, alginic acid, glucomannans, pectinic compounds, xylans, α- linked glucans, laminarin and modified cellulose. It has been hypothesized that *Epulopiscium* plays a role in the degradation of algal polysaccharides and it may ferment the sugars released (9, 10, 12). These fermentation products, short chain fatty acids (SCFAs), may be taken up and utilized as an energy source by their host fish (24).

Currently there is only indirect evidence to support this hypothesis. First, direct microscopic analysis reveals an abundant microbial consortium in the intestinal tract of *N. tonganus*, which includes *Epulopiscium* Type B cells (12). In addition, previous work assessing the concentration of SCFAs in the surgeonfish host of *Epulopiscium* Type B demonstrated that SCFAs increase in the more posterior segments of the intestinal tract, where this microbial cell biomass is greatest (8). These observations lead to the following question: Does *Epulopiscium* type B express the carbohydrase genes required to breakdown these complex polysaccharides found in algae and does

the type of algae consumed by the host fish influence the expression of specific carbohydrases? Based on the data from sequencing the *Epulopiscium* genome, I hypothesize that *Epulopiscium* will express a variety of carbohydrase genes and the expression level of specific carbohydrase genes will correlate with the types of algae present in the gut of the host fish. RT-qPCR assays have been used successfully in natural populations of *Epulopiscium* Type B cells to quantify expression of the *Epulopiscium spoIIE* and *rpoB* genes from extracted community RNA (30). A similar approach can be utilized to quantify expression of carbohydrase genes. In order to understand if the types of algae present in the host fish intestinal tract influence the expression of these carbohydrase genes, dietary analysis of the fish from which cells are collected can be conducted.

In several species of Clostridia it has been shown that the expression of cellulase and hemicellulase genes is influenced by the presence of specific extracytoplasmic polysaccharides (20, 33, 35, 39). In *Clostridium cellulovorans*, cellulase and hemicellulase genes exhibited significant expression when grown with polysaccharide substrates such as cellulose, xylan and pectin, and low levels of expression when grown with a variety of mono- or di-saccharides (20). In *Clostridium thermocellum*, cellulase synthesis in cells grown in media with Avicel (Microcrystalline cellulose) was nine-fold greater than those grown in cellobiose (39). Based on this one could predict that a similar pattern of expression, one dependent upon substrate availability, would be observed in *Epulopiscium* Type B. For example, agar is comprised of agarose polymers and can comprise up to 70% of the cell wall of red algae (Rhodophyta) (25). Samples with a high red algae content as determined by

dietary analysis could be predicted to have an increased level of expression of the agarase genes as compared to other carbohydrase genes.

Using the protocol of Choat and Clements (1992), field specimens of frequently consumed algae can be collected and used as reference material to assist in the identification of algal fragments found in formalin fixed gastrointestinal samples of *N. tonganus*. Quantification of the proportions of the different dietary components can be calculated by spreading the intestinal contents over a gridded tray, using a dissecting microscope randomly counting four transects, summing the amount of each food category and transforming these values into percentages.

Using the *Epulopiscium* Type B genome annotation data available in the NCBI database, primers will be designed for all putative carbohydrases. Primer sets will be tested in standard PCR amplifications using *Epulopiscium* cells and microbial community DNA extracted from surgeonfish intestinal contents. Specificity of products generated will be confirmed by sequence analysis. Optimized primer sets will be used for RT-qPCR assays with RNA extracted from *N. tonganus* intestinal contents as described in Miller et al. (2011). This data will be normalized against the housekeeping gene, *rpoB* to account for cell density differences in the samples. For each assay, the expression level of all identified carbohydrase genes will be measured. This will allow for a comparison between the expression levels of each gene in the presence of various substrates.

The benefits of this transcriptional analysis of the carbohydrase systems of this organism will be two fold. Ultimately the substrates utilized by *Epulopiscium* Type B will depend upon the regulatory response of the cell in the presence of a potentially

vast array of complex polysaccharides (25, 36). By understanding the expression levels of the genes involved in the utilization of the carbohydrates found in the natural environment of *Epulopiscium* Type B, one can build upon what is already known about this complex intestinal ecosystem to further refine the development of culture media to maintain these cells in the lab. Secondly this type of analysis will provide direct evidence of the functional role that *Epulopiscium* Type B plays in the intestinal tract of their host fish.

RNA Polymerase Localization in Whole *Epulopiscium* Type B cells

Epulopiscium Type B cells have an abundant amount of DNA, which appears to be arrayed around the periphery of the cell (34). It is hypothesized that this arrangement allows these cells to transcribe genes at the location in the cell where the gene products are needed, thus overcoming the constraints of diffusion on the movement of biologically important molecules from site of synthesis to site of action. As such, this cellular modification could be a driving force in allowing *Epulopiscium* Type B to attain such a large cell size. However, this raises an interesting question: Are chromosomes in all regions of the cell transcriptionally active or, are areas of the cell transcriptionally quiescent? Given the large cell size and cytoplasmic volume of *Epulopiscium*, I hypothesize that a large number of the chromosomes will be transcriptionally active to support the metabolic needs of these large cells. To begin to assess this, RNA polymerase (RNAP) could be immunolocalized in the cell. If all of the chromosomes are transcriptionally active, then one would expect to see numerous fluorescent RNAP foci over the periphery of the cytoplasm, associated with the cellular DNA. However, if some chromosomes are transcriptionally quiescent, then we

would expect to only see foci in more confined areas of the cell. If none of the chromosomes were transcriptionally active, then there would be no discrete fluorescent RNAP foci. Instead a homogenous fluorescent signal associated with the DNA would be observed (6).

Subcellular localization of RNAP has been determined in both *Bacillus subtilis* and *Escherichia coli* through GFP-tagging of the β' subunit C terminus of RNAP (6, 17, 26). In both of these organisms, RNAP was found to associate with the nucleoid of the cells resulting in "transcription foci" (6). When cells were treated with rifampicin, an antibiotic that inhibits RNAP from initiating or reinitiating transcription, these foci disappeared. Unfortunately, as *Epulopiscium* Type B is unavailable in culture and no genetic transformation system currently exists, the use of GFP- tagging is not feasible. However, protein immunolocalization has been successfully utilized in these cells (1) as well as in their close relative *Metabacterium polyspora* (3), to visualize the cell division protein FtsZ. As such this method could be implemented to visualize RNAP in *Epulopiscium* Type B cells.

The sequence of the β' subunit of RNAP in *Epulopiscium* Type B has been identified in the partial genome sequence data available in GenBank (Accession: ZP_02693020.1). Following the protocol in Angert and Losick (1998), a region of the RNAP β' subunit gene from *Epulopiscium* Type B cells will be amplified, cloned into the appropriate pRSET expression vector to preserve the correct reading frame, and transformed into *E. coli* strain BL21. *E. coli* will be grown in Luria Bertani broth with the appropriate antibiotic to an OD_{600} of 0.9 and harvested 2 hours after induction with IPTG. The 6-Histidine (His)-tagged partial RNAP polypeptide will be isolated using a

Ni-nitrilotriacetate (NTA) column, purified via dialysis to reduce the amount of urea in the solution and the purified peptide will be used to generate antibodies in a rabbit. Specificity and the optimal concentration of antiserum to use against *Epulopiscium* Type B RNAP β' subunit in the immunolocalization studies will be determined through Western blot analysis. Briefly, molecular weight marker loaded in lanes 1, 3, 5, 7, 9 and 11 and purified RNAP protein loaded in lanes 2, 4, 6, 8, 10 and 12 will be separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). The separated proteins will be electrophoretically transferred to a membrane and the membrane blocked and washed. After washing, the membrane will be cut into six strips, each containing molecular weight marker and purified RNAP protein. Each strip will be incubated with one of the following dilutions of antiserum: 1:500, 1:1,000, 1:5,000, 1: 10,000, 1:50,000 and 1:100,000. The membrane will be washed, and incubated in HRP-conjugated anti-rabbit secondary antibody at room temperature for 1 hour. The membrane will be washed and protein bands will be detected using Pierce ECL Western Blotting Substrate, followed by exposure of membrane to autoradiography film and development using a film autoprocessor. As a control, a second gel will be run in parallel, and processed as described above except it will be incubated with the same dilutions of preimmune serum. The RNAP antiserum dilution that results in the most robust signal will be used in the immunolocalization studies.

Naso tonganus will be collected by spearfishing from the outer reefs around Lizard Island, Great Barrier Reef, Australia. Intestinal contents will be removed and fixed for one hour in formaldehyde at a final concentration of 3.7% (6). This particular fixative forms intermolecular bridges stabilizing the RNAP localization pattern. Cells

will be centrifuged and resuspended in 1X Phosphate Buffered Saline (PBS). To assess the RNAP localization pattern in transcriptionally inactive cells, a method similar to that used by Ward *et al.* (2009) for *in situ* BrdU labeling of *Epulopiscium* cells will be used. Working in an anaerobic chamber, the intestinal tracts of individual fish will be removed and tied off into segments with cotton string. Rifampicin will be injected into some segments at a final concentration of approximately 50 μg/ml and the segments incubated for an hour at approximately 24°C (comparable to local water temperature). To ensure that this incubation does not perturb the native RNAP localization pattern in the cells, for each fish an intestinal tract segment without rifampicin will be used as a control. Samples will be recovered by cutting open and removing the intestinal contents of the cordoned off segments and processed as described above.

Immunolocalization of RNAP in transcriptionally active and transcriptionally inactive *Epulopiscium* Type B cells will be performed as previously described (1) with the following modifications. The appropriate concentration of antiserum against *Epulopiscium* Type B RNAP β' subunit, as determined above, will be applied to the samples on microscope slides and the slides incubated in a humid chamber overnight at 4°C. Slides will be washed and incubated with Texas Red labeled anti-rabbit Fc antibodies according to manufacturers instructions while simultaneously stained with the DNA-specific dye 4', 6-diamidino-2-phenylindole (DAPI). The cells will then be mounted in an antifade reagent. To ensure that the RNAP localization pattern observed is due to specific binding of the antibodies, controls without primary or secondary

antibody will be prepared as above, omitting either the anti-RNAP antibodies or the labeled anti-rabbit Fc antibodies respectively.

Previous immunolocalization studies of proteins in *Epulopiscium* Type B used epifluorescence microscopy with deconvolution and three-dimensional reconstruction (1). Given the thickness of these cells, up to 80 μm, an alternative approach would be the use of confocal laser scanning microscopy (CLSM) (32). CLSM allows for serial optical sections from thick specimens while reducing the amount of fluorescent signal from structures outside of the focal plane leading to a much more detailed image. This technique has been used to successfully co-localize RNA polymerase I and DNA in human cancer cells (18). Using a Confocal system equipped with a filter block that allows for simultaneous observation of Texas Red and DAPI, one image of RNAP and one of DNA will be simultaneously obtained at each position of the z-stepping motor and specialized computer software will be used in 3-D reconstruction of the series of images.

Epulopiscium Type B cells can attain lengths of 250 μm with cell volumes a million times that of *E. coli*. Other large bacteria possess inclusions reducing the volume of active cytoplasm or vacuoles that hold the cytoplasm in a very thin layer near the cell surface allowing these organisms to maintain a high surface area-to-volume ratio. While *Epulopiscium* Type B cells increase their effective surface area by having a highly convoluted cytoplasmic membrane (34), in comparison to other large bacteria they appear to maintain a low-surface-to-volume ratio. The studies presented above would provide insights into how these cells are able to maintain an active metabolism. If all chromosomes were transcriptionally active this would allow these

unusually large cells to overcome the constraints of diffusion by transcribing genes where their products are needed.

dnaA gene variant protein expression in *Epulopiscium* Type B populations

We investigated the potential for genetic diversification in *Epulopiscium* by sequencing clones from PCR products of single copy genes amplified from single *Epulopiscium* cells (29). For *dnaA*, a high proportion of clones (12 of 30) contained a single-nucleotide deletion within a mononucleotide tract of ten adenines, which if transcribed results in a premature stop codon downstream of the deletion. These observations raise the following question: Are both variants of the *dnaA* gene expressed and translated into protein by *Epulopiscium* Type B or do these cells only synthesize the functional protein product? I hypothesize that *Epulopiscium* cells harbor these two gene variants and synthesize both the truncated and full-length proteins.

DnaA functions as a transcription factor, but is perhaps better known as the DNA replication initiator protein (7, 38). During replication initiation, DnaA oligomers bind to conserved sequences called DnaA boxes, which are arrayed around the origin of replication (oriC) in bacterial chromosomes. Once bound, DnaA facilitates melting of the DNA and the recruitment of additional proteins required for replication. In *E. coli* cells the expression of DnaA is tightly coordinated with the number of DnaA boxes. As replication proceeds and the number of origins (therefore origin proximal DnaA boxes) increases, so does the amount of DnaA (21).

Elegant work done in our lab by Dr. Rebekah Ward investigating DNA replication dynamics in *Epulopiscium* Type B demonstrated that DNA replication

occurs in both the mother cell and the developing offspring (37). Assuming that these cells are initiating multiple rounds of replication, it is likely that DnaA is expressed throughout the developmental cycle. The differences in size of the two DnaA protein forms, 445 amino acids (~51 kDa) versus 299 amino acids (~35 kDa), would allow us to employ Western Blotting to assess if these cells are synthesizing both forms.

DNA coding for the first 260 amino acids (~30 kDa) of the DnaA protein (upstream of the polyA tract and stop codon in the gene variant) from individual *Epulopiscium* Type B cells will be PCR amplified and cloned into the appropriate pRSET vector. Only this domain of the protein will be used in the generation of antibodies to ensure that the antibodies target regions present in both the full length and the truncated proteins. The His-tagged partial DnaA polypeptide will be isolated as described above and used to generate antibodies in a rabbit. This polyclonal antiserum will be affinity purified and used in immunoprecipitation and Western blotting of the DnaA proteins from *Epulopiscium* cells. Affinity purified DnaA antiserum will be prepared as previously described (31). Briefly, polyclonal serum will be applied to a column of immobilized DnaA protein crosslinked to CNBr-activated Sepharose (Pharmacia). The column will be washed to remove any unbound antibody, bound antibody will be eluted with phosphoric acid (0.2 M, pH 2.4) and the eluate neutralized and stored in sodium phosphate buffer (0.1 M, pH 6.8) at -20°C until further use. The specificity of the affinity purified DnaA antiserum will be assessed via Western blot with the purified *Epulopiscium* DnaA fragment serving as a positive control.

In preparation for immunoaffinity purification of DnaA and subsequent Western analysis, the purified anti-DnaA antibodies will be bound to protein A beads in the following manner. A loose slurry of antibodies and protein A beads will be incubated with gentle rocking at room temperature for one hour. Beads will be washed twice with and resuspended in sodium borate (0.2 M, pH 9.0), mixed with the bifunctional coupling reagent dimethyl pimelimidate (DMP) to a final concentration of 20 mM and incubated at room temperature for 30 minutes with gentle mixing. Beads will be washed in ethanolamine (0.2M, pH 8.0) and then incubated in the same solution at room temperature for 2 hours with gentle mixing. Supernatant will be removed and antibody-coupled beads resuspended in PBS/merthiolate (0.01%) and stored at 4°C until used in immunoaffinity purification. As a control, preimmune serum will be coupled to protein A beads following the same protocol. To assess the efficiency of coupling, 10 μl of beads will be removed before the addition of DMP and after incubation with DMP, boiled in an equal volume of 2X Laemmli sample buffer (4% SDS, 10% 2-mercaptoethanol, 20% glycerol, 0.004% bromophenol blue, 0.125 M Tris-HCl), run on a SDS- polyacrylamide gel and stained with Coomassie Blue. If adequate coupling conditions have been achieved, heavy chains will be visualized in the "before addition of DMP" sample but not the "after incubation with DMP" sample (22).

A large number of *Epulopiscium* cells will need to be collected to purify enough DnaA for the Western analysis. *Naso tonganus* will be collected as described above. To limit autolysis and protease activity in the cells, working at 4°C, the fish intestinal tract will be removed placed on ice and immediately moved into a portable

anaerobic bag that has been purged of oxygen and filled with carbon dioxide several hours in advance. The bag will be resealed and a GasPack activated to purge any introduced oxygen. Working quickly, intestinal contents will be removed and *Epulopiscium* enriched for by low speed centrifugation to remove the smaller bacterial cells and filtration through loosely packed glass wool to remove the larger vegetal matter. Cells will be stored at -70°C until further processed.

Total protein will be extracted from samples enriched for *Epulopiscium* as follows. Cell pellets will be resuspended in 10 cell volumes of buffer (50 mM glucose, 10 mM Ethylenediaminetetraacetic acid (EDTA), 25 mM Tris-HCl (pH 8.0) containing 4 mg/ml of lysozyme and incubated at room temperature for 5 minutes. Tubes will be transferred to ice and RIPA lysis buffer (50 mM Tris- HCl (pH 8.0), 150 mM NaCL, 1% nonyl phenoxypolyethoxylethanol (NP-40), 0.5% deoxycholate (DOC) and 0.1% SDS) will be added in the presence of protease inhibitors (Minicocktail Protease Inhibitor, Roche) and PMSF (50 mg/ml final concentration) (22). Protein concentration will be determined using Bio-Rad Detergent Compatible protein assay (Bio-Rad Laboratories, Hercules, CA). To further ensure that efficient protein extraction has been achieved 20 μg of total extracted protein will be heated at 95°C with an equal amount of 2X Laemmli sample buffer, run on a SDS-polyacrylamide gel and stained with Coomassie Blue. Purified DnaA protein will be run as a positive control.

Affinity purified DnaA antiserum bound to protein A beads will be used to isolate DnaA from *Epulopiscium* lysates using immunoaffinity purification (IP). Three separate columns will be generated: two with affinity purified DnaA antiserum to

which either total *Epulopiscium* cell lysate or purified DnaA protein will be added, respectively, and one with the total polyclonal preimmune serum to which *Epulopiscium* cell lysate will be added. The antibody coupled beads will be transferred to a column, and washed with 20 bed-volumes of PBS. The specific antigen solution will be applied to the appropriate column and passed through the column at a flow rate of 1 ml per hour. The columns will be washed with 20 bed-volumes of binding buffer (20 mM sodium phosphate, 500 mM sodium chloride (NaCl)). The optimal elution conditions will need to be determined experimentally, following the suggestions of Harlow and Lane (1988), employing the gentlest elution conditions first (22). The column will be washed with 20 volumes of pre-elution buffer. 0.5 bed volumes of elution buffer will be applied to the column and the fraction collected. This will be repeated, and individual fractions assessed for antigen via Western Blot analysis. The fractions with high antigen concentration will be combined. Eluted samples will be resuspended in an equal volume of 2X Laemmli buffer and heated at 95°C for five minutes. In order to test that only bound protein is being eluted from the Protein A-antibody complexes, 2X Laemmli buffer will be added to the resins from the three columns, heated at 95°C for five minutes and resolved along with the immunoprecipitated samples.

Ten lanes on the gel will be run as follows: Lane 1, Molecular Weight Marker. Lane 2, IP from total lysate (using affinity purified DnaA antiserum). Lane 3, IP with purified DnaA protein (affinity purified DnaA antiserum). Lane 4, IP from total cell lysate (using pre-immune serum). Lanes 5 through 7, boiled resin from each of the three columns. Lane 8, purified DnaA protein. Lane 9, affinity purified DnaA

antiserum and lane 10 preimmune serum. Samples (20 μl per well) will be resolved by SDS-PAGE, and electrophoretically transferred onto a polyvinylidene difluoride (PVDF) membrane (Invitrogen, Carlsbad, CA). The membrane will be blocked and probed with affinity purified anti-DnaA antibodies over night at 4°C. The membrane will be washed, and incubated in HRP-conjugated anti-rabbit secondary antibody at room temperature for 1 hour and protein bands detected as described above.

Expected results will yield two bands of approximately 35 and 51 kDa (truncated and full length DnaA proteins, respectively) in lane 2, and a band of approximately 30 kDa (His-tagged purified DnaA protein) in lane 3. The lack of an ~50 kDa band in this lane will confirm that the IP samples are not contaminated with the affinity purified DnaA antiserum and the ~51 kDa band observed in lane 2 is the full length DnaA protein. No bands will be expected in lane 4. In lanes 5 through 7 (boiled resin from the three columns) bands of approximately 25 kDa and 50 kDa will be expected, corresponding to the light and heavy chains of the liberated antibodies, respectively. In lane 8, purified His-tagged DnaA, a band of approximately 30 kDa will be expected, similar to that in lane 3. In lanes 9 and 10, one would expect bands corresponding to the heavy and light chains due to secondary antibody binding.

Should no bands be identified in the IP of DnaA from the *Epulopiscium* cell lysate, while a band is present in the positive control (IP purified DnaA), it is conceivable that the cells used were not expressing DnaA at levels detectable by this approach. In order to overcome this, protein will be extracted from cells at different stages in the *Epulopiscium* developmental cycle (1). Alternatively, if a ~50 kDa band (full length DnaA) is present but a ~35 kDa band (truncated DnaA) is not present in

the IP of DnaA from the *Epulopiscium* cell lysate then perhaps this protein is not as abundant as the full-length, functional protein. Another possibility is that this *dnaA* deletion variant, which is devoid of its DNA binding domain may be unstable and therefore targeted for proteolysis.

These analyses could provide a greater understanding into the ability of *Epulopiscium* Type B cells to harbor unstable genetic elements. However, additional work would need to be done to fully understand the functional significance of these gene variants. The amount of DnaA has been tightly linked to the initiation of DNA replication (38). In these highly polyploid cells, possessing chromosomes with *dnaA* genes that encode for nonfunctional protein products could serve as a way to attenuate the amount of DnaA in the cell, ultimately regulating DNA replication.

Many insights have been gained into the biology of this unusually large intestinal symbiont. The studies presented above would expand upon this knowledge. Transcriptional analyses of carbohydrase genes would provide a deeper understanding about the nutritional requirements of *Epulopiscium* Type B, thus facilitating its growth in culture, which would allow for the development of *in vivo* studies and development of a genetic transformation system. Understanding the transcriptional activity of the thousands of genomes present in an individual *Epulopiscium* cell will further our understanding of the forces that have allowed this organism to attain its large cell size as well as provide insights into transcriptional regulation in this unique and exciting organism. Finally, an understanding of the protein expression of the *dnaA* gene variants in *Epulopiscium* will provide insights into potentially novel modes of

regulation as well as how extreme polyploidy allows for some amount of genetic diversity in these cells.

References

1. **Angert, E. R. & Clements, K. D** 2004. Initiation of intracellular offspring in *Epulopiscium*. Mol. Microbiol. **51:** 827-835.

2. **Angert, E. R., Clements, K. D. & Pace, N. R.** 1993. The largest bacterium. Nature. **362:** 239-241.

3. **Angert, E. R. & Losick, R. M.** 1998. Propagation by sporulation in the guinea pig symbiont *Metabacterium polyspora*. Proc. Natl. Acad. Sci. **95:** 10218-10223.

4. **Baranov, P. V., Hammer, A. W., Zhou, J., Gesteland, R. F. & Atkins J. F.** 2005. Transcriptional slippage in bacteria: distribution in sequenced genomes and utilization in IS element gene expression. Genome Biol. **6:** R25.

5. **Bayliss, C. D.** 2009. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. FEMS Microbiol. Rev. **33:** 504-520.

6. **Cabrera, J. E. & Jin, D. J.** 2003. The distribution of RNA polymerase in *Escherichia coli* is dynamic and sensitive to environmental cues. Mol. Microbiol. **50:** 1493-1505.

7. **Cheng, L. & Keiler, K. C.** 2009. Correct timing of *dnaA* transcription and initiation of DNA replication requires *trans* translation. J. Bacteriol. **191:** 4268-4275.

8. **Choat, J. H. & Clements, K.D.** 1998. Vertebrate herbivores in marine and terrestrial environments: A nutritional ecology perspective. Annu. Rev. Ecol. Syst. **29:** 375-403

9. **Choat, J. H., Clements, K.D. & Robbins, W.D.** 2002. The trophic status of herbivorous fishes on coral reefs 1: Dietary analyses. Mar. Biol. **140:** 613-623.

10. **Choat, J. H., Clements, K.D. & Robbins, W.D.** 2004. The trophic status of herbivorous fishes on coral reefs II:Food processing modes and trophodynamics. Mar. Biol. **145:** 445-454.

11. **Clarke, K.R & Gorley, R.N.** 2001. PRIMER v5: User Manual/Tutorial. Plymouth, UK, PRIMER-E.

12. **Clements, K. D. & Bullivant, S.** 1991 An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. J. Bacteriol. **173:** 5359-5362.

13. **Clements, K. D. & Choat, J. H.** 1995. Fermentation in Tropical Marine Herbivorous Fishes. Physiol. Zoo. **68:** 355 -378.

14. **Clements, K. D. & Choat, J. H.** 1997. Comparison of herbivory in the closely-related marine fish genera *Girella* and *Kyphosus*. Mar. Biol. **127:** 579-586.

15. **Clements, K. D., Raubenheimer, D., Choat, J. H.** 2009. Nutritional ecology of marine herbivorous fishes: ten years on. Funct. Ecol. **23:** 79-92.

16. **Clements, K. D., Sutton, D. C. & Choat, J. H.** 1989. Occurrence and characteristics of unusual protistan symbionts from surgeonfish (Acanthuridae) of the Great Barrier Reef, Australia. Marine Biol. **102:** 403-412.

17. **Doherty, G. P., Fogg, M. J., Wilkinson, A. J. & Lewis, P. J.** 2010. Small subunits of RNA polymerase: localization, levels and implications for core enzyme composition. Microbiology. **156:** 3532-3543.

18. **Gilbert, N., Lucas, L., Klein, C., Menager, M., Bonnet, N. & Ploton, D.** 1995. Three-dimensional co-location of RNA polymerase I and DNA during interphase and mitosis by confocal microscopy. J. Cell Sci. **108:** 115-125.

19. **Georgopoulus, C.** 1989. The *E. coli* dnaA initiation protein: a protein for all seasons. Trends Genet. **5:** 319-321.

20. **Han, S. O., Yukawa, H., Inui, M. & Roy, H. D.** 2003. Regulation of expression of cellulosomal cellulase and hemicellulase genes in *Clostridium cellulovorans.* J. Bacteriol. **185:** 6067-6075.

21. **Hansen, F. G., Koefoed, S. Sorensein, L. & Atlung, T.** 1987. Titration of DnaA protein by *oriC* DnaA-boxes increases *dnaA* gene expression *in Escherichia coli*. **66:** 255-258.

22. **Harlow E. & Lane, D.** 1988. Antibodies: a laboratory manual. Cold Spring Harbor Laboratory Cold Spring Harbor, NY

23. **Henderson, I. R., Owen, P. & Nataro, J. P.** 1999. Molecular switches- the ON and OFF of bacterial phase variation. Mol. Microbiol. **33:** 919-932.

24. **Karasov, W. H. & Martinez del-Rio, C.** 2007. Physiological Ecology: How animals process energy, nutrients and toxins. Princeton, NJ, Princeton University Press.

25. **Kloareg, B. & Quatrano, R. S.** 1988. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. Oceanogr. Mar. Biol. Annu. Rev. **26:** 259-315.

26. **Lewis, P. J, Doherty, G. P & Clarke, J.** 2008. Transcription factor dynamics. Microbiology.**154:** 1837-1844.

27. **Ley, R. E., Hamady, M., Lozupone, C., Turnbaugh, P. J., Ramey, R.R., Bircher, J. S., Schiegel, M. L., Tucker, T. A, Schrenzel, M. D., Knight, R. & Gordon, J. I.** 2008. Evolution of mammals and their gut microbes. Science **320:** 47-51.

28. **Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I.** 2008. Worlds within worlds: Evolution of the vertebrate gut microbiota. Nat. Rev. Micro. **6:** 776–788.

29. **Mendell, J. E., Clements, K. D., Choat, J. H. and Angert, E.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U S A **105:** 6730-6734.

30. **Miller, D. Choat, J. H., Clements, K. D. & Angert, E. R.** 2011. The *spoIIE* homolog of *Epulopiscium* sp. Type B is expressed early in intracellular offspring development. J. Bacteriol. **193:** 2642-2646.

31. **Newman, G. & Crooke, E.** 2000. DnaA, the initiator of *Escherichia coli* chromosomal replication, is located at the cell membrane. J. Bacteriol. **182:** 2604-2610.

32. **Pawley, J. B.** 2006. Handbook of Biological Confocal Microscopy, 3 ed. Springer Science and Business Media, LLC. New York, NY.

33. **Raman, B., McKewon, C. K., Rodriguez, M., Brown, S. D. & Mielenz, J.3 R.** 2011. Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. BMC Microbiol. **11:** 134-148.

34. **Robinow, C. & Angert, E. R.** 1998. Nucleoids and coated vesicles of "*Epulopiscium*" spp. Arch. Microbiol. **170:** 227-235.

35. **Servinsky, M. D., Kiel, J. T. Dupuy, N. F. & Sund, C. J.** 2010. Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. Microbiology. **156:** 3478-3491.

36. **Tolentino-Pablico, G., Bailly, N., Froese, R. & Elloran, C.** 2008. Seaweeds preferred by herbivorous fishes. J. Appl. Phycol. **20:** 933-938.

37. **Ward, R. J., Clements, K. D., Choat, J. H. & Angert, E. R.** 2009. Cytology of terminally differentiated *Epulopiscium* mother cells. DNA Cell Biol. **28:** 57-64.

38. **Zakrzewska-Czerwinska, J., Jakimowicz, D., Zawilak-Pawlik, A. M. & Messer, W.** 2007. Regulation of the initiation of chromosomal replication in bacteria. FEMS Microbiol. Rev. **31:** 378-387.

39. **Zhang, Y-H. P. & Lynd, L. R.** 2005. Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum*. J. Bacteriol. **187:** 99-106.

APPENDIX A

CHAPTER 2: SUPPLEMENTAL INFORMATION

Materials and Methods

Single-cell amplification of the rRNA operon ITS, *ftsZ*, *recA* and *dnaA*. A single, washed *Epulopiscium* type B cell was deposited in a PCR tube and lysed as described (Materials and Methods Chapter 2), except after proteinase K inactivation, tubes were filled with 1X HotStarTaq Master Mix (Qiagen) and the appropriate gene primer pair. The following primers were used for PCR amplification of *Epulopiscium* rRNA operon ITSs [1435F: 5′- TCAGTGACCTAACCG, 242R: 5′-TTCGCTCGCCRCTACT (7)] or single-copy genes *ftsZ* (ftsZF: 5′-GGCGGAGGAAATAATGCCGTTGAT, ftsZR: 5′-AGAGTTGAATGCGGGAGCTAGCTT), *dnaA* (dnaAF: 5′-TTTCTCTATGGAGGGGTTGG, dnaAR: 5′-AGGCAAATAGCAATGTATCTA) and *recA* (recAF: 5′-AGGCTCCATTATGAAACTTGGAGA, recAR: 5′-GGACAAGGCCGAGAGAATGCTAAA). Thermal cycling conditions were as follows: 15 min at 95°C (only used with enzymes that required heat activation) followed by 28 cycles of 30 s at 94°C, 30 s at 54°C, 90 s at 72°C, with a final incubation for 10 min at 72°C. Amplifications were performed using either HotStarTaq Master Mix (Qiagen) or PCR SuperMix High Fidelity (Invitrogen) at the concentrations recommended by the manufacturers. The PCR products from amplification of a single-copy gene from a single cell were cloned using the TOPO

TA Cloning kit (Invitrogen) as per manufacturer's instructions. Clones were randomly selected and screened for the presence of a PCR product of the expected size. For each single-copy gene, the sequence of 30 individual clones was determined using Big Dye chemistry and an Applied Biosystems Automated 3730 DNA Analyzer. Alternatively, PCR products were purified (QIAquick PCR Purification Kit, Qiagen) and sequences were determined as described above.

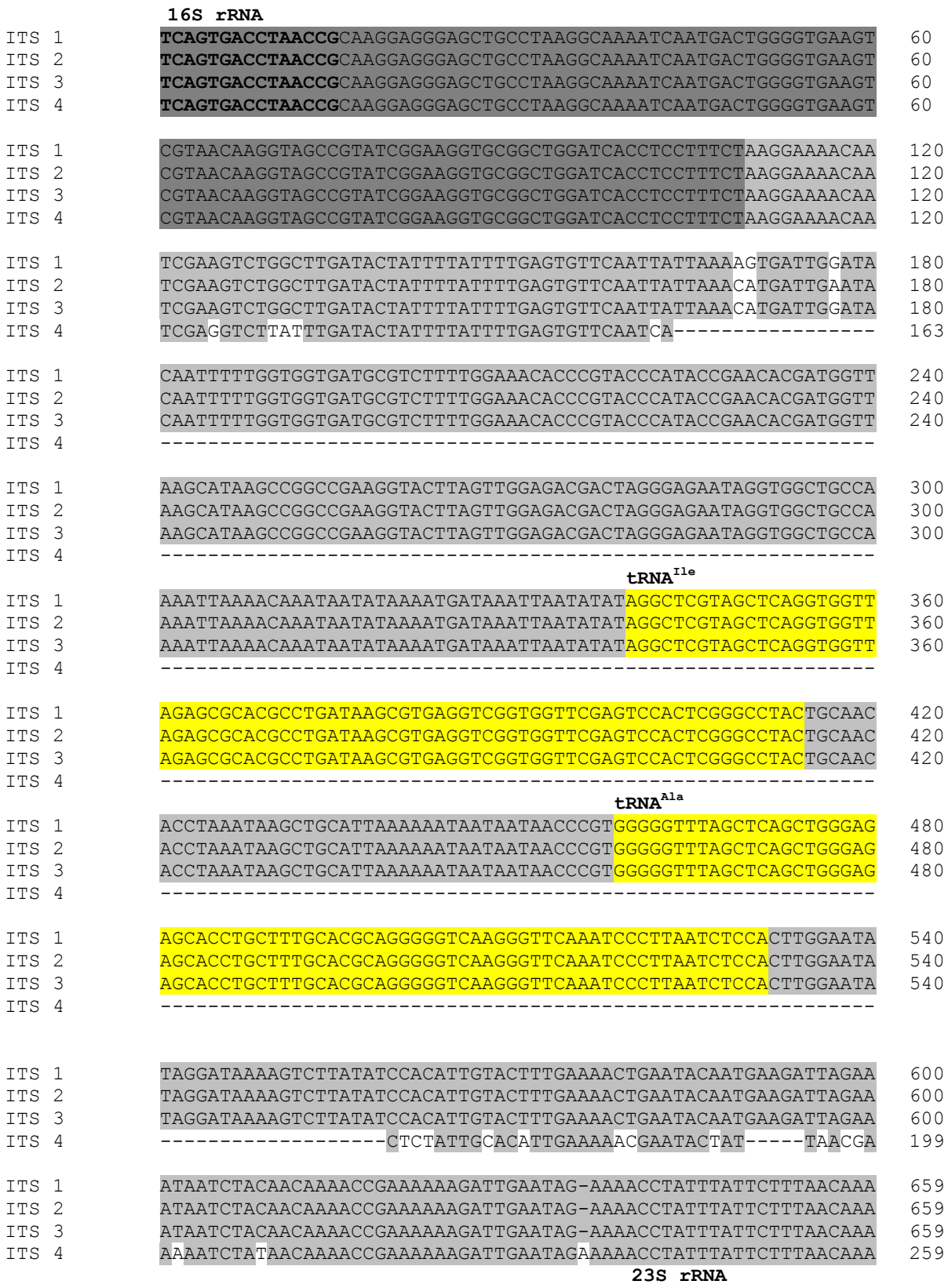
Sequence analysis of *ftsZ*, *recA* and *dnaA* PCR products from single *Epulopiscium* cells.

For *ftsZ* and *recA*, some clones varied from a consensus sequence by one or two nucleotides and all variants were unique (Tables A.1 & A.2). Single nucleotide changes appeared to be transitions and the vast majority (~71%) coded for non-synonymous amino acid substitutions. The overall frequency of nucleotide differences was approximately 0.0006 per nucleotide (38 changes in 50,160 positions), which is comparable to published *Taq* DNA polymerase PCR error rates (3-5). For *dnaA*, we observed a similar trend in substitution frequency, with 18 unique substitutions in 20,880 positions, mostly transitions (Table A.3).

In the analysis of *dnaA*, we recovered a high proportion of clones that contained a common single nucleotide deletion in a stretch of 10 adenines. We used a series of experiments to help determine if these deletions were generated during PCR amplification. A fragment of the *dnaA* gene was amplified from 12 individual *Epulopiscium* cells using either *Taq* DNA polymerase or a mixture of thermostable DNA polymerases with higher fidelity than *Taq* alone. The sequence of each PCR product was determined. All electropherograms showed a consensus sequence

upstream of the poly-A region, but after, all showed minor peaks from the subpopulation of genes carrying the deletion (Figure A.2). When the *Epulopiscium dnaA* gene was amplified from plasmid DNA, similar electropherograms were generated from the PCR product. It appeared that the amplitude of the minor peaks from plasmid DNA amplicons were not as prevalent as those in the DNA amplified from single cells. When the sequences of plasmid DNAs were determined directly, evidence for a large subpopulation of variants was not observed. These results indicate that at least some deletion variants are produced during amplification with thermostable polymerases.

If the *dnaA* deletion variant were translated *in vivo*, a truncated peptide, devoid of its DNA binding domain (6, 9, 11), would be produced. This peptide may be unstable and targeted for proteolysis (1, 8). Alternatively, it could interact with full-length DnaA and alter the cooperative binding of DnaA to *dnaA* boxes near the replication origin, which in turn may affect initiation. It is worth noting that for *E. coli* and *B. subtilis*, lower than normal DnaA concentrations in the cell lead to an increase in cell mass (2, 10).


16S rRNA
ITS 1 TCAGTGACCTAACCGCAAGGAGGGAGCTGCCTAAGGCAAAATCAATGACTGGGGTGAAGT 60
ITS 2 TCAGTGACCTAACCGCAAGGAGGGAGCTGCCTAAGGCAAAATCAATGACTGGGGTGAAGT 60
ITS 3 TCAGTGACCTAACCGCAAGGAGGGAGCTGCCTAAGGCAAAATCAATGACTGGGGTGAAGT 60
ITS 4 TCAGTGACCTAACCGCAAGGAGGGAGCTGCCTAAGGCAAAATCAATGACTGGGGTGAAGT 60
ITS 1 CGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAAAACAA 120
ITS 2 CGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAAAACAA 120
ITS 3 CGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAAAACAA 120
ITS 4 CGTAACAAGGTAGCCGTATCGGAAGGTGCGGCTGGATCACCTCCTTTCTAAGGAAAACAA 120
ITS 1 TCGAAGTCTGGCTTGATACTATTTTATTTTGAGTGTTCAATTATTAAAAGTGATTGGATA 180
ITS 2 TCGAAGTCTGGCTTGATACTATTTTATTTTGAGTGTTCAATTATTAAACATGATTGAATA 180
ITS 3 TCGAAGTCTGGCTTGATACTATTTTATTTTGAGTGTTCAATTATTAAACATGATTGGATA 180
ITS 4 TCGAGGTCTTATTTGATACTATTTTATTTTGAGTGTTCAATCA---------- 163
ITS 1 CAATTTTTGGTGGTGATGCGTCTTTTGGAAACACCCGTACCCATACCGAACACGATGGTT 240
ITS 2 CAATTTTTGGTGGTGATGCGTCTTTTGGAAACACCCGTACCCATACCGAACACGATGGTT 240
ITS 3 CAATTTTTGGTGGTGATGCGTCTTTTGGAAACACCCGTACCCATACCGAACACGATGGTT 240
ITS 4 ----------
ITS 1 AAGCATAAGCCGGCCGAAGGTACTTAGTTGGAGACGACTAGGGAGAATAGGTGGCTGCCA 300
ITS 2 AAGCATAAGCCGGCCGAAGGTACTTAGTTGGAGACGACTAGGGAGAATAGGTGGCTGCCA 300
ITS 3 AAGCATAAGCCGGCCGAAGGTACTTAGTTGGAGACGACTAGGGAGAATAGGTGGCTGCCA 300
ITS 4 ----------
tRNA Ile
ITS 1 AAATTAAAACAAATAATATAAAATGATAAATTAATATATAGGCTCGTAGCTCAGGTGGTT 360
ITS 2 AAATTAAAACAAATAATATAAAATGATAAATTAATATATAGGCTCGTAGCTCAGGTGGTT 360
ITS 3 AAATTAAAACAAATAATATAAAATGATAAATTAATATATAGGCTCGTAGCTCAGGTGGTT 360
ITS 4 ----------
ITS 1 AGAGCGCACGCCTGATAAGCGTGAGGTCGGTGGTTCGAGTCCACTCGGGCCTACTGCAAC 420
ITS 2 AGAGCGCACGCCTGATAAGCGTGAGGTCGGTGGTTCGAGTCCACTCGGGCCTACTGCAAC 420
ITS 3 AGAGCGCACGCCTGATAAGCGTGAGGTCGGTGGTTCGAGTCCACTCGGGCCTACTGCAAC 420
ITS 4 ----------
tRNA Ala
ITS 1 ACCTAAATAAGCTGCATTAAAAAATAATAATAACCCGTGGGGGTTTAGCTCAGCTGGGAG 480
ITS 2 ACCTAAATAAGCTGCATTAAAAAATAATAATAACCCGTGGGGGTTTAGCTCAGCTGGGAG 480
ITS 3 ACCTAAATAAGCTGCATTAAAAAATAATAATAACCCGTGGGGGTTTAGCTCAGCTGGGAG 480
ITS 4 ----------
ITS 1 AGCACCTGCTTTGCACGCAGGGGGTCAAGGGTTCAAATCCCTTAATCTCCACTTGGAATA 540
ITS 2 AGCACCTGCTTTGCACGCAGGGGGTCAAGGGTTCAAATCCCTTAATCTCCACTTGGAATA 540
ITS 3 AGCACCTGCTTTGCACGCAGGGGGTCAAGGGTTCAAATCCCTTAATCTCCACTTGGAATA 540
ITS 4 ----------
ITS 1 TAGGATAAAAGTCTTATATCCACATTGTACTTTGAAAACTGAATACAATGAAGATTAGAA 600
ITS 2 TAGGATAAAAGTCTTATATCCACATTGTACTTTGAAAACTGAATACAATGAAGATTAGAA 600
ITS 3 TAGGATAAAAGTCTTATATCCACATTGTACTTTGAAAACTGAATACAATGAAGATTAGAA 600
ITS 4 ----------CTCTATTGCACATTGAAAAACGAATACTAT-----TAACGA 199
ITS 1 ATAATCTACAACAAAACCGAAAAAAGATTGAATAG-AAAACCTATTTATTCTTTAACAAA 659
ITS 2 ATAATCTACAACAAAACCGAAAAAAGATTGAATAG-AAAACCTATTTATTCTTTAACAAA 659
ITS 3 ATAATCTACAACAAAACCGAAAAAAGATTGAATAG-AAAACCTATTTATTCTTTAACAAA 659
ITS 4 AAAATCTATAACAAAACCGAAAAAAGATTGAATAGAAAAACCTATTTATTCTTTAACAAA 259
23S rRNA

```
ITS 1       TTCTTTAGTCAAGAAAAGACACAAAGACCAAAATATAGGTCAAGCTACTAAGAGCGTAGA   719
ITS 2       TTCTTTAGTCAAGAAAAGACACAAAGACCAAAATATAGGTCAAGCTACTAAGAGCGTAGA   719
ITS 3       TTCTTTAGTCAAGAAAAGACACAAAGACCAAAATATAGGTCAAGCTACTAAGAGCGTAGA   719
ITS 4       TTCTTTAGTCAAGAAAAGACACAAAGACCAAAATATAGGTCAAGCTACTAAGAGCGTAGA   319

ITS 1       GTGGATGCCTTGGCACCGAGAGCCGATGAAGGACGTGGTAAGCTGCGAAAAGCTTGGGTA   779
ITS 2       GTGGATGCCTTGGCACCGAGAGCCGATGAAGGACGTGGTAAGCTGCGAAAAGCTTGGGTA   779
ITS 3       GTGGATGCCTTGGCACCGAGAGCCGATGAAGGACGTGGTAAGCTGCGAAAAGCTTGGGTA   779
ITS 4       GTGGATGCCTTGGCACCGAGAGCCGATGAAGGACGTGGTAAGCTGCGAAAAGCTTGGGTA   379

ITS 1       AGTTGCAAACAAACGCTATAACCCAAGATGTCCGAATGGGGAAACCTAGCTGAGTAAACC   839
ITS 2       AGTTGCAAACAAACGCTATAACCCAAGATGTCCGAATGGGGAAACCTAGCTGAGTAAACC   839
ITS 3       AGTTGCAAACAAACGCTATAACCCAAGATGTCCGAATGGGGAAACCTAGCTGAGTAAACC   839
ITS 4       AGTTGCAAACAAACGCTATAGCCCGAGATATCCGAATGGGGAAACCTAGCTGAGTAAACC   439

ITS 1       TCAGTTGTTATATGGTAAATTCATAGCCATATAAAGGGAACGTTGGGAACTGAAACATCT   899
ITS 2       TCAGTTGTTATATGGTAAATTCATAGCCATATAAAGGGAACGTTGGGAACTGAAACATCT   899
ITS 3       TCAGTTGTTATATGGTAAATTCATAGCCATATAAAGGGAACGTTGGGAACTGAAACATCT   899
ITS 4       TCAGTTGTTATATGGTAAATTCATAGCCATATAAAGGGAACGTTGGGAACTGAAACATCT   499

ITS 1       AAGTACCAACAGGAGGAGAAAGAAAAATCGATTTCCTAAGTAGYGGCGAGCGAA         953

ITS 2       AAGTACCAACAGGAGGAGAAAGAAAAATCGATTTCCTAAGTAGYGGCGAGCGAA         953
ITS 3       AAGTACCAACAGGAGGAGAAAGAAAAATCGATTTCCTAAGTAGYGGCGAGCGAA         953
ITS 4       AAGTACCAACAGGAGGAGAAAGAAAAATCGATTTCCTAAGTAGYGGCGAGCGAA         553
```

Figure A.1. Alignment of four unique 16S-23S rRNA ITS sequences recovered from an *Epulopiscium* sp. type B single-cell amplification clone library. Two different ITS lengths were recovered. Longer ITSs contain putative coding sequences for tRNAs (highlighted in yellow). Sequences conserved between the ITSs are highlighted in gray. Regions coding for 16S and 23S rRNAs are shown in dark gray. Amplification primer sequences are in bold. The sequences of 41 clones were determined from which four consensus sequences were recovered. Each ITS consensus sequence was represented by 8-13 clones, other clones deviated from a consensus by 1 to 3 unique nucleotides. GenBank Accession numbers for ITS1, ITS2, ITS3 and ITS4 are EU500235, EU500236, EU500237 and EU500234, respectively.

Table A.1. Sequence variation in *Epulopiscium ftsZ* clones

Clone*	Position in partial sequence†	Difference from consensus	Resulting codon‡
25	133	A → G	AA**G** (syn)
11	146	T → C	**C**TG (syn)
11	152	G → A	**A**CA
5	185	G → A	**A**CC
15	207	T → C	**A**CC
27	253	A → G	GG**G** (syn)
1	287	G → A	**A**TG
8	306	A → G	A**G**G
20	435	C → T	C**T**G
20	449	T → C	**C**TA (syn)
19	478	A → G	AT**G**
5	481	A → G	GA**G** (syn)
1	552	C → T	C**T**G
6	573	T → C	T**C**T
3	578	G → A	**A**AT
28	600	A → G	G**G**C
29	613	T → C	GC**C** (syn)
19	656	G → A	**A**AG

21	687	T → C	C**C**T
24	700	A → G	AC**G** (syn)
28	755	A → G	**G**TG
10	756	T → C	A**C**G
23	465	deletion	frameshift

* Of the 30 clones analyzed, 18 deviated from the consensus sequence. Of these, 13 had a single nucleotide difference and 5 (in bold) had two differences.

† Position number is based on nucleotide numbering of partial gene sequence available in GenBank (accession number AF067821).

‡Position change is highlighted. (syn) indicates synonymous.

Table A.2. Sequence variation in *Epulopiscium recA* clones

Clone*	Position in partial sequence†	Difference from consensus	Resulting codon‡
21	121	A → G	**G**AA
1	153	A → G	G**G**G
29	165	G → A	A**A**C
23	183	T → C	T**C**C
34	187	A → G	AT**G**
4	244	C → T	TC**T** (syn)
17	247	C → T	GA**T** (syn)
4	331	T → C	GT**C** (syn)
27	417	T → C	C**C**T
2	421	G → A	AT**A**
1	544	A → G	GG**G** (syn)
31	563	T → C	**C**CT
27	567	C → T	T**T**T
7	635	C → T	**T**GT
33	776	T → C	**C**GG

* Of the 30 clones analyzed, 13 deviated from the consensus sequence. Of these, 11 had a single nucleotide difference and 2 (in bold) had two differences.

† Position number is based on nucleotide numbering of partial gene sequence available in GenBank (accession number EF127642).

‡Position change is highlighted. (syn) indicates synonymous.

Table A.3. Sequence variation in *Epulopiscium dnaA* clones

Clone*	Position in partial sequence†	Difference from consensus	Resulting codon‡
54	14	A → G	C**G**T
6	50	A → G	C**G**T
63	124	G → A	**A**AC
38	204	A → G	AA**G** (syn)
52	206	A → G	G**G**A
10	254	C → A	G**A**C
10	321	T → C	CT**C** (syn)
10	360	A → G	CA**G** (syn)
17	398	A → G	A**G**A
22, **19**, 8, **17**, **56**, **39**, 60, 37, 20, **11**, 14, **6**	393-402	A deletion	frameshift
19	413	A → G	**G**AA
56	440	T → C	**C**GC
56	486	A → G	after STOP
26	497	A → G	A**G**T
11	506	T → C	after STOP
17	554	A → G	after STOP

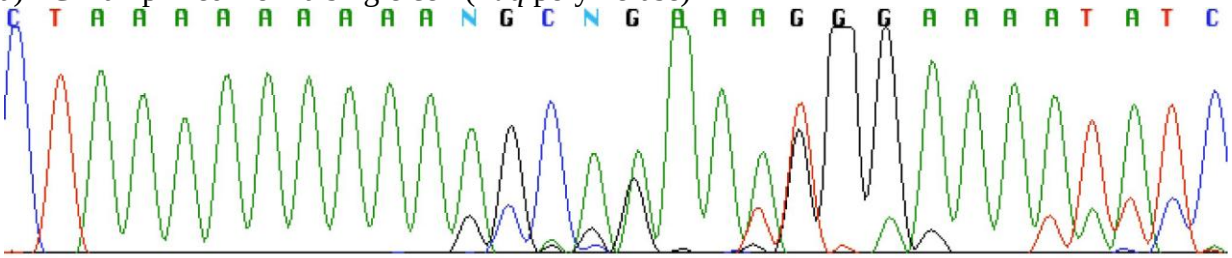
6	566	A → G	after STOP
39	621	A → G	after STOP
53	677	G → A	C**A**A

* Of the 30 clones analyzed, 21 deviated from the consensus sequence. Of these, 7 had more than one difference from the consensus. 12 had an A deletion within a homopolymeric tract, leading to a frameshift and premature STOP codon downstream.

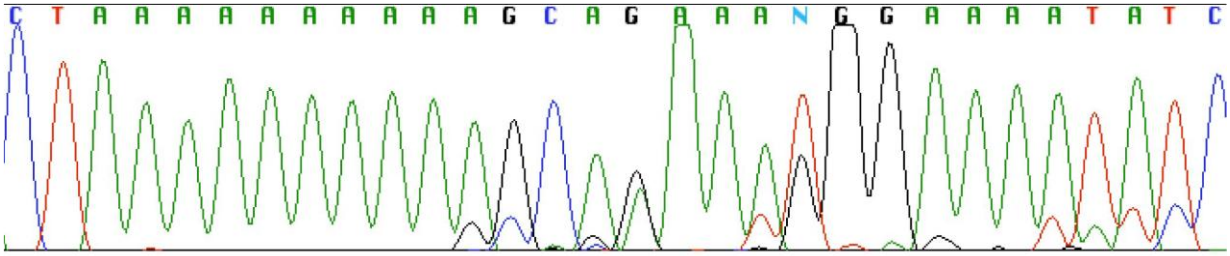
† Position number is based on nucleotide numbering of partial gene sequence available in GenBank (accession number EF127641).

‡Position change is highlighted. (syn) indicates synonymous.

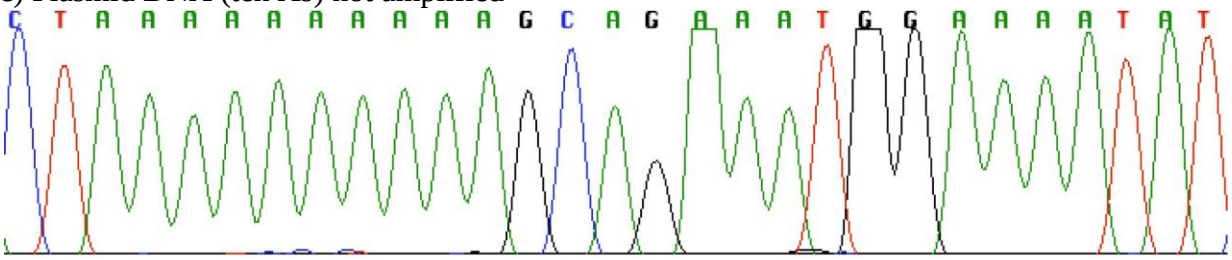
a) PCR-amplified from a single cell (*Taq* polymerase)



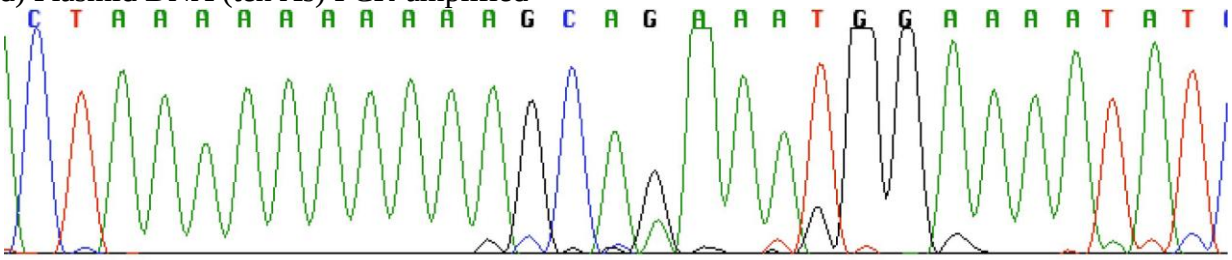
b) PCR-amplified from a single cell (HiFi polymerase mix)



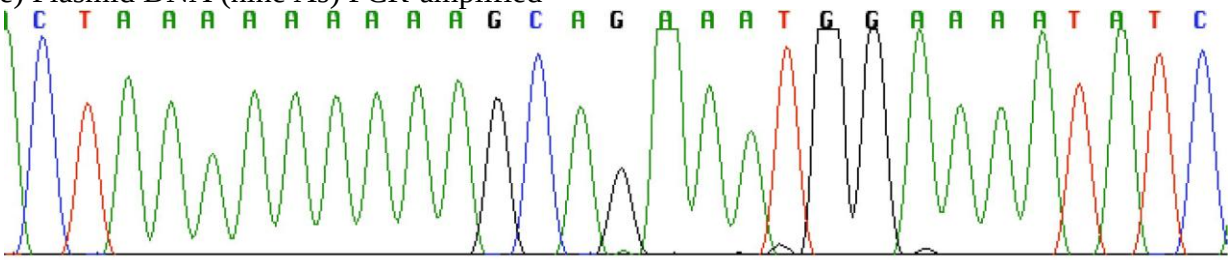
c) Plasmid DNA (ten As) not amplified



d) Plasmid DNA (ten As) PCR-amplified



e) Plasmid DNA (nine As) PCR-amplified



f) Plasmid DNA (nine As) not amplified

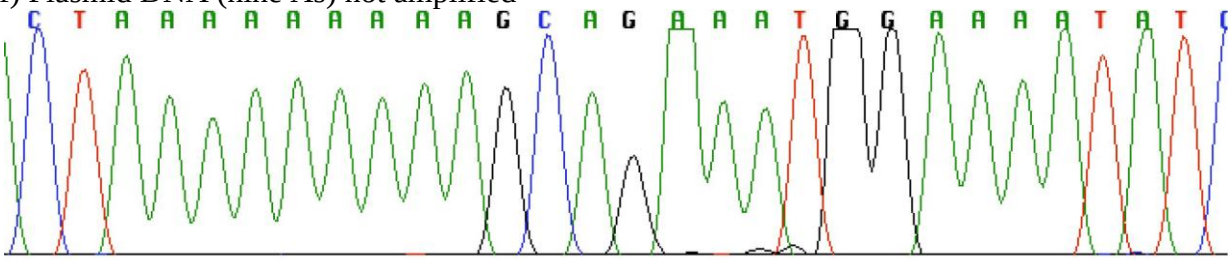


Figure A.2. Electropherograms of a region of the *Epulopiscium dnaA* gene that contains a long, mononucleotide repeat. (**a** – **b**) *dnaA* amplified from single *Epulopiscium* cells using either *Taq* DNA polymerase or a high-fidelity mixture of polymerases, respectively. (**c**) Sequence from a plasmid borne copy of *dnaA* with 10 As. (**d**) Sequence read from *dnaA* amplified from plasmid DNA using a high-fidelity mix of thermostable polymerases. (**e**) Sequence from the deletion variant of *dnaA*, amplified from a plasmid using a high-fidelity mix of thermostable polymerases. (**f**) Plasmid borne *dnaA* variant with 9 As.

References

1. **Blaesing, F., Weigel, C., Welzeck, M. & Messer, W.** 2000. Analysis of the DNA-binding domain of *Escherichia coli* DnaA protein. Mol. Microbiol. **36:** 557-569.

2. **Donachie, W. D. & Blakely, G. W.** 2003. Coupling the initiation of chromosome replication to cell size in *Escherichia coli*. Curr. Opin. Microbiol. **6:** 146-150.

3. **Dunning, A. M., Talmud, P. & Humphries, S. E.** 1988. Errors in the polymerase chain reaction. Nucleic Acids Res. **16:** 10393.

4. **Eckert, K. A. & Kunkel, T. A.** 1991. DNA polymerase fidelity and the polymerase chain reaction. PCR Methods Appl. **1:** 17-24.

5. **Ennis, P. D., Zemmour, J., Salter, R. D. & Parham, P.** 1990. Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc. Natl. Acad. Sci. U S A. **87:** 2833-2837.

6. **Kaguni, J. M.** 2006. DnaA: controlling the initiation of bacterial DNA replication and more. Annu. Rev. Microbiol. **60:** 351-375.

7. **Lane, D. J.** 1991. 16S/23S rRNA Sequencing in Nucleic Acid Techniques in Bacterial Systematics, eds. Stackebrandt, E. & Goodfellow, M. (Wiley, New York), 115-175.

8. **Moriya, S., Kato, K., Yoshikawa, H. & Ogasawara, N.** (1990) Isolation of a *dnaA* mutant of *Bacillus subtilis* defective in initiation of replication: amount of DnaA protein determines cells' initiation potential. Embo. J. **9:** 2905-2910.

9. **Mott, M. L. & Berger, J. M.** 2007. DNA replication initiation: mechanisms and regulation in bacteria. Nat. Rev. Microbiol. **5:** 343-354.

10. **Ogura, Y., Imai, Y., Ogasawara, N. & Moriya, S.** 2001. Autoregulation of the *dnaA-dnaN* operon and effects of DnaA protein levels on replication initiation in *Bacillus subtilis*. J. Bacteriol. **183:** 3833-3841.

11. **Zakrzewska-Czerwinska, J., Jakimowicz, D., Zawilak-Pawlik, A. & Messer, W.** 2007. Regulation of the initiation of chromosomal replication in bacteria. FEMS Microbiol. Rev. **31:** 378-387.

APPENDIX B

The Wrath of the Rotting Root
Teacher's Guide

Overview
Decomposition is defined as the breaking down of a substance into a less complicated chemical structure. The decomposition of organic material is critical in maintaining the planet, and its atmosphere, as we know it. This activity will allow students to design an experiment to look at how different factors affect the rate of decomposition of organic material (a carrot), isolate the bacteria and fungi responsible for its decomposition, and assess various biochemical abilities of these organisms.

Subject
Biology, Microbiology, Ecology

Audience
High School or Middle School

Time Required
Dependent upon how many parts of the activity you choose to do.

Day 1 (preferably a double period): Introduce the process of decomposition, brainstorm about factors that could influence the rate of decomposition, develop data collection/recording sheet, and complete experimental design.

Day 2: Set up microcosms, take initial masses of carrots and microcosms (in case carrots decompose too quickly, students can mass the microcosm and still collect data over several weeks), make baseline observations, review data collection procedure.

Bi-weekly or weekly as time allows: Have students mass their carrots and microcosms. Students should be making physical observations of the carrot (smell, texture, appearance) and also thinking about *why* they are seeing a change in mass over time (whether this is a gain or loss of mass).

After the carrot has started to decompose: Have students isolate bacteria and fungi on the carrot surface by swabbing it with a sterile swab. Have the students set up a serial dilution to isolate the bacteria and fungi from the soil.

2-3 days after bacterial and fungal isolation: Have students describe the colony morphology (shape) of the different bacteria and fungi they have isolated. They should talk about the color, size, shape, edge, elevation and texture of the colonies. Based on unique colony morphology, have the students isolate two different bacteria from their original bacterial isolation plates.

2-3 days after isolating 2 separate bacteria: Have a brainstorming session about what a carrot is made up of and how that could influence the types of enzymes the decomposing bacteria could possess. Discuss the environment these organisms are found in, and the advantage of being able to exploit a wide variety of resources, such as dead bacteria. Have the students come up with

hypotheses about the different biochemical tests. Have the students inoculate the biochemical tests.

2-3 days after inoculating the biochemical tests: Have the students read the tests, record the results and see if their hypotheses were correct. If they were not, have them speculate *why*.

Background

A leaf falls from a tree in the forest. It settles on the ground, but it doesn't remain there forever. It eventually breaks down, or *decomposes* into something that you no longer recognize as a leaf. There are many things that can influence the rate at which the leaf decomposes. In a tropical rainforest where it is hot and moist, leaves and other organic matter decompose at a much faster rate than in a dry temperate region. Things like temperature, moisture, wind, and other environmental conditions are considered abiotic factors. Biotic factors can also influence the rate of decomposition. The visible changes to the leaf can be the result of insects, worms, birds, or other animals feeding on it. However, the final decomposition of the leaf is carried out by bacteria and fungi, which feed on dead tissue. Without these organisms, the planet would become littered with dead and decaying material. The animals, birds, insects, and worms simply would not be able to keep up with the amount of material that needs to be decomposed. In addition, the fungi and bacteria convert the organic matter into material that is usable by plants and other primary producers, helping to maintain the atmosphere of our planet as we know it. This process, called nutrient cycling, is essential to preserving life on earth!

Learning Objectives

1. Students will be able to develop a research question and design an experiment to address this question.

2. Students will be able to identify the independent and dependent variable of their experiment.

3. Students will be able to accurately collect and record quantitative data, and make and record qualitative observations.

4. Students will be able to interpret the data they have collected to determine if their hypothesis is correct.

5. Students will be able to come up with some reasons why their hypothesis was or was not correct.

6. Students will be able to grow bacteria and fungi responsible for decomposition of their carrot.

7. Students will identify an individual bacterium based on colony morphology and other observations. Students will isolate this bacterium, and use it to inoculate biochemical tests to assess the metabolic capabilities of the bacterium.

8. Students will interpret their biochemical tests to identify what enzymes their bacterial isolates possess. The students will relate this to the composition of the carrot and the environment in which the bacteria came from.

NYS Science Education Standards Addressed

Standard 1: (High School and Middle School):
Key Idea 1: The central purpose of scientific inquiry is to develop explanations of natural phenomena in a continuing, creative process.
Key Idea 2: Beyond the use of reasoning and consensus, scientific inquiry involves the testing of proposed explanations involving the use of conventional techniques and procedures and usually requiring considerable ingenuity.
Key Idea 3: The observations made while testing proposed explanations, when analyzed using conventional and invented methods, provide new insights into phenomena.
Performance Indicator 5.1 (Middle school):
Compare the way a variety of living specimens carry out basic life functions and maintain dynamic equilibrium
Key Idea 5 (High School)**:** Organisms maintain a dynamic equilibrium that sustains life.
Performance Indicator 7.1 (Middle School)**:** Describe how living things, including humans, depend upon the living and nonliving environment for their survival.

National Science Education Standards Addressed
Science as Inquiry
- Abilities necessary to do scientific inquiry
- Understandings about scientific inquiry

Life Science
- Interdependence of organisms
- Matter, energy, and organization in living systems

Earth and Space Science
- Geochemical cycles

Science and Technology
- Abilities of technological design
- Understandings about science and technology

Science in Personal and Social Perspectives
- Environmental quality
- Science and technology in local, national, and global challenges

Assessment Strategy

1. Given the appropriate background information on experimental design, and factors that could affect the rate of decomposition, students will complete the experimental design sheet. Students will be assessed by their ability to collect and analyze the data they obtain over time. Students should be able to identify whether their hypothesis was correct, and provide some explanation of this outcome.

2. Given the appropriate background information on bacterial growth and colony morphology, students will successfully identify and isolate an individual bacterium. Students will be assessed by having a single bacterial species on each of their agar plates.

3. Given the appropriate background information on the biochemical tests, students will successfully inoculate these tests. Students will be assessed by their ability to correctly interpret the tests and accurately describe what kind of biochemical processes their bacteria are able to carry out.

For other specific assessment strategies of learning objectives, please refer to the **Learning Objectives** section.

Teaching Tips

Day 1: (*Handout #1*) Give a short presentation on decomposition (make sure that the students know what decomposition is!). Explain the difference between abiotic and biotic forces. During the brainstorming session, have students come up with different forces that could influence the rate of decomposition. Break them up into groups and have them come up with several hypotheses. Once they have decided on the one that they as a group want to test, have them come up with an experimental design to test that hypothesis. You may want to have them formally write this up, to reinforce writing and communication in science. This will also give you a chance to review the experimental design and make sure that they have a clear question to answer and have identified the independent and dependent variable. You may want them to think about data collection while they are coming up with their experimental design, and come up with some sort of data collection sheet. Make sure that everyone in the group records the data! Inevitably, if only one person has the sheet, it will get lost. If this happens for some reason, make sure to have "dummy data" so that the students have something for analysis.

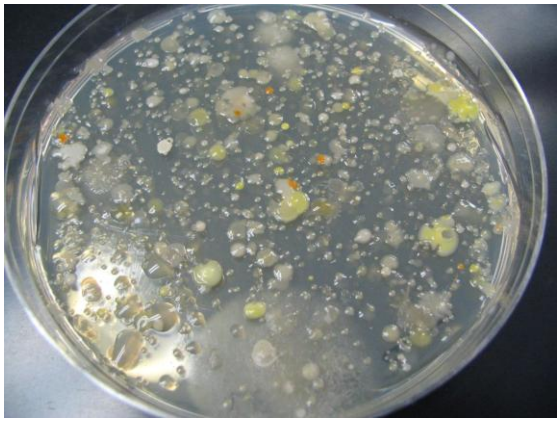
Day 2: If the students haven't already done so, have them design a data-recording sheet (See *Sample Data Collection Sheet*). Have the students set up their microcosms (a chunk of some hard root vegetable, such as carrot, parsnip, or potato, buried in 1 ¾ cup dirt in zip type plastic baggie), making sure that they weigh both the carrot, and the microcosm in its entirety. They may also want to make some baseline observations of the texture, color and smell of both the carrot and the soil. Additional observations may include the moisture content of the soil and presence of organic matter or macroscopic organisms (if using soil from a garden or other outdoor site). If possible, have the studnets set up their

experimental and control microcosms in triplicate. Ask them to write a short explanation of why it is useful to run experiments in triplicate.

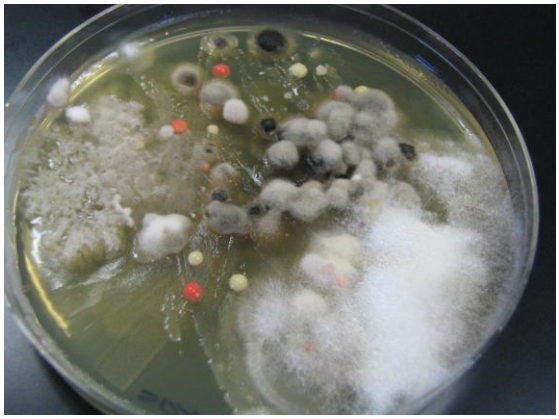
Bi-weekly or weekly as time allows: (*Handout #2*) Have the students weigh their microcosms, record the weight, and calculate the change in mass. After they have weighed the microcosm in its entirety, have the students carefully remove the carrot from the bag, gently remove as much dirt as possible, weigh the carrot, record the weight of the carrot and calculate the change in mass. Have them make qualitative observations. Have them think about what is happening to their carrot. If they see an initial gain in weight, why could this be happening? Why could they see different results when they weigh the entire microcosm vs. the carrot? As time progresses and the carrots decompose even further, ask them why carrots don't decompose in the ground where they grow. If you would like, you could end the activity here and have them write up a formal lab report on their findings.

After the carrot has started to decompose: (*Handout #3*) If you are able to get the supplies, then you can have the students isolate bacteria and/or fungi from the carrot surface, and/or the soil on solid agar media. Bacteria are not picky and will grow on many types of media including LB (Luria-Bertani) and NA (Nutrient Agar). To isolate the fungi, there is a special media called SDA (Sabouraud Dextrose Agar) that has a low pH, which will select for the fungi.

2-3 days after bacterial and fungal isolation: (*Handout #4*) Since these are soil organisms, you can incubate these plates at room temperature for 2-3 days and have beautiful growth. You can also have the students note changes over time as the slower growing could take 7-10 days to appear on the plate. The organisms that you isolate will be affected by the experimental conditions of the microcosm, but the variety is astounding as you can see in the following two pictures.



Bacteria isolated from carrot surface using LB media.



Fungi isolated from carrot surface using SDA media.

Have the students describe what the colonies look like, smell like, etc. (see *Handout #4* for colony morphology (shape) descriptions). Encourage them to be creative. Tell

them to pretend that they are describing this to a friend on the phone. What are some of the adjectives they would use?? You could stop the activity here, if you wish.

NOTE: The vast majority of bacteria and fungi that are isolated are harmless to humans, and are organisms that we come into contact with on a daily basis. However, to minimize exposure it is best to tape the plates shut, and make sure that the students wash their hands after handling them.

If you chose to continue on, have the students isolate an individual bacterium based on colony morphology. To ensure student safety, you may want them to wear gloves. Have them chose a bacteria that is separate from others. See example below:

Either of these colonies would be good examples of colonies to use to isolate an individual bacterium.

2-3 days after isolating 2 separate bacteria: (*Handout #5* and *#6*) Have another brainstorming session about the different macromolecules that make up a carrot. You may need to guide them with this activity. You want them to come up with sugar, starch and anything else (since you will be testing for glucose and sucrose fermentation and starch hydrolysis). Also discuss the environment that the carrot is found in, and how it can be very nutrient poor. Although many people think of the soil as nutrient rich, when in reality, it isn't. I use the example of what happens when a heavy rain falls and percolates down through the soil – it washes the nutrients away. Introduce the concept of DNase – they should recognize the *–ase* ending as belonging to an enzyme, and facilitate a discussion about having different enzymes to exploit resources in the environment, like the DNA from other bacteria that may die.

Safety

Make sure that students wash their hands with soap and water after handling the plates and inoculating the media. After students have swabbed their plates, or inoculated the

biochemical tests, have them place their swabs in a zipper type plastic bag. Make sure that they know not to touch their skin, or their lab mates with the swabs. Swabs can be disposed of in the trash. Agar media plates should be autoclaved or incinerated after use.

Extensions

- Have students develop and implement an "in-house" composting program. Based on the results of their decomposition experiment, ask them to come up with the optimum conditions for composting.

- Have students design a regulation pathway of the dnase gene. Let them be creative with possible regulation strategies of this gene (ie- positive vs. negative regulation).

- Give each group a large piece of paper, and ask them to design a concept map at the end of the activity. Ask them to apply the concepts they leaned in this activity to other activities……you may be amazed at the connections that they make!

Handout #1

Name: _____
Date: _____
Period:_____

Wrath of the Rotting Root

A leaf falls from a tree in the forest. It settles on the ground, but it doesn't remain there forever. It eventually breaks down, or *decomposes* into something that you no longer recognize as a leaf. There are many things that can influence the rate at which the leaf decomposes. In a tropical rainforest where it is hot and moist, leaves and other organic matter decompose at a much faster rate than in a dry temperate region. Things like temperature, moisture, wind and other environmental conditions are considered abiotic factors. Biotic factors can also influence the rate of decomposition.

The visible changes to leaf can be the result of insects, worms, birds, or other animals feeding on it. However, the final decomposition of the leaf is carried out by bacteria and fungi, which feed on dead tissue. Without these organisms, the planet would become littered with dead and decaying material. The animals, birds, insects and worms simply could not keep up with the amount of material that needs to be decomposed. In addition, the fungi and bacteria convert the organic matter into material that is usable by plants and other primary producers, helping to maintain the atmosphere of our planet as we know it. This process, called nutrient cycling, is essential to preserving life on earth!

The digestion of organic material by bacteria and fungi is accomplished by specialized proteins called enzymes. Each enzyme has a specific substrate, or target material which it is able to break down. Consider the leaf that has fallen to the ground. The organic material that makes up the leaf has been around for a very long time. The bacteria and fungi have *evolved* to break down this organic material. The same is not true for "newer" substances, such as plastics. This is why landfills become full, and a plastic container left in the woods by careless hikers won't break down like the leaf. The bacteria and fungi simply don't possess the enzymes to break down the bonds in the plastic. In order to help over come this problem of increasing waste, scientists are working to develop plastics which bacteria will be able to degrade!

Your job is to design an experiment, which will test how some factor affects the rate of decomposition. In your group, list some of the things that are likely to affect decomposition, and then chose one of these ideas to test in a scientific experiment. You will use a chunk of carrot as the material to be decomposed, and a zipper-type plastic bag filled with dirt will be your "environment". By measuring the mass of the carrot over time, you will be able to determine the rate of decomposition.

Designing the Experiment

As a group, prepare a **detailed** written proposal describing your experiment. Make sure to have this approved by Mr. Reed or Jenna BEFORE setting up your experiment.

The following suggestions may be helpful to you as you design your experiment:

1. What is the question you are trying to answer?
2. How will you try and answer it?
3. What is your independent variable (the one you vary)?
4. What is your dependent variable (the one you measure)?
5. What will be your control?
6. How will you record your data?

Handout #2

Name: _____
Date: _____
Period:_____

Take your carrot out of the bag and GENTLY brush the dirt off. Weigh your carrot, and record the mass and the change (D) in mass. Did your carrot lose mass, or gain mass? Why do you think this happened?

Describe what your carrot looks like, how it feels, and the smell or any other characteristics that you may think are important. Feel free to draw your carrot, or if you'd like, take a picture of it!

Handout #3

Name: _____
Date: _____
Period:_____

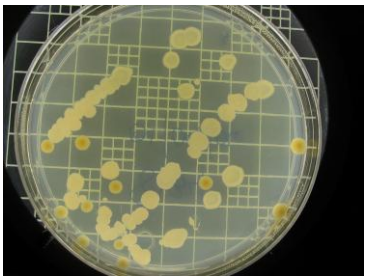
WHO IS DOING THE DECOMPOSING???

1. You have four plates. Two are labeled LB, which you will use to culture or grow bacteria. The other two are labeled "SDA" which you will use to culture fungi. Label each plate with your initials, and the date. On one LB plate write "dilution" and on the other LB plate write "carrot". Do the same for the SDA plates.
2. Using the spatula, transfer a SMALL amount of dirt – about the size of a pea - from one of your bags to the test tube labeled "1". Record what bag you removed the dirt from (experimental or control).
3. Mix well by swirling. DO NOT SHAKE! The test tube covers are not tight and you will end up with stinky rotten carrot water all over you and your friends (who will quickly become NOT your friends!).
4. Put a tip onto the pipetman, and transfer 100 mls (this is 0.1 mls) of the dirt water from the test tube labeled "1" into the test tube labeled "2". CONGRATULATIONS! You have just created a dilution!
5. Carefully open the small tinfoil package at the end with the stripped tape and carefully remove one sterile swab.
6. Remove the cover from the test tube labeled "2", and carefully, without touching the outside of the test tube, soak your applicator in the dirt water.
7. Streak the LB plate you labeled "dilution". (See example below) DO NOT COMPLETELY REMOVE THE LID OF THE PETRI DISH! JUST TILT IT UP ON ONE SIDE.
8. Place your swab on a paper towel for disposal.
9. Repeat this process for the SDA plate labeled "dilution".
10. If you have carrot left in your decomposition microcosm, open the bag, and extract the carrot with the tweezers (remember, you want to look at the bacteria and fungi on the carrot, NOT on your hands).
11. Carefully remove another sterile swab from the tinfoil package, and streak the surface of the carrot. If you don't have any carrot left, touch the swab to carrot mush, or the dirt in the bag.
12. Use this same swab to streak the surface of the LB plate labeled "carrot".
13. Place your swab on a paper towel for disposal.
14. Repeat this process for the SDA plate labeled "carrot"
15. Seal your plates with a strip of parafilm.
16. Incubate agar side up (upside down).

Handout #4
PLATE DESCRIPTION, COLONY MORPHOLOGY
AND BACTERIAL ISOLATION

Imagine you are a bacterium on the surface of the carrot. Someone comes in with a swab and snatches you up (OH NO!). But wait……it isn't so bad! This nice person just deposited you on a very cushy surface, full of yummy nutrients which you can use for all of your life processes, including reproduction.

Bacteria reproduce via a process called *binary fission*. In this process the bacteria replicates its chromosome, and then the cell divides in two. Then those cells divide in two and so on and so on until you have many bacteria!!!!!!!

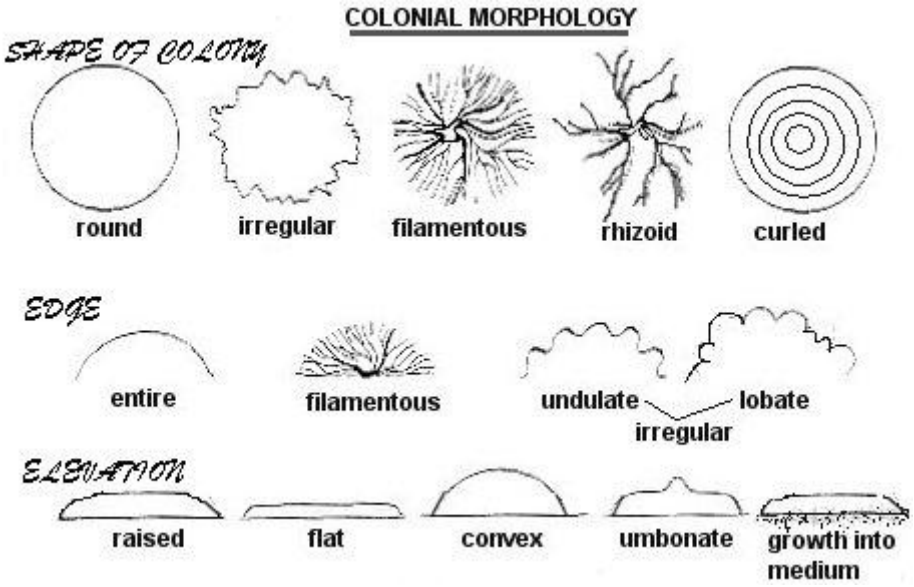


After many, many divisions and when grown on a solid surface such as the LB media in your Petri dishes, the bacteria are visible to the naked eye in a structure called a *colony*. The colonies that you see on your plates all arose from one bacterium! Its pretty amazing that these are visible to the naked eye, since an average bacteria is between 1 -3 microns in length (remember that a micron is 1/1000 of a millimeter.

Colonies can be described based on their *morphology*, or shape. On the next page are pictures of different types of colony morphologies. Do any of your colonies on the LB plates have these morphologies? Which types of morphologies are represented on your LB plates? What colors are your colonies? Are they dry looking, or wet looking. Some microbiologists use the term "mucousy" to describe colonies because they look like snot. In the space below describe the colonies on your LB plates:

DILUTION PLATE:

CARROT PLATE:



Fungal reproduction is much more complex, but many fungi still form colonies. Look at your SDA plates and describe the colony characteristics.
DILUTION PLATE:

CARROT PLATE:

On the last page you will find an area to sketch each of your plates.
BACTERIAL ISOLATION: Now that you have described the colonies on your LB plate, you may notice that they look very different from one another. Chances are that each unique colony type represents a unique bacterium. In order to better understand who was doing the decomposing in your microcosms, we are asking you to *isolate* or separate an individual type of bacteria on another LB plate.

1. Identify two (2) unique colony morphologies on your LB plates. One can come from the dilution plate, and one from the carrot plate, or they can both come from the dilution or carrot plate. Just make sure that they each have some distinguishing characteristic like different colors, or different colony morphologies.
2. Circle each colony on the bottom of your Petri dish. Label one “Colony #1” and the other “Colony #2”.

3. Next, label the two (2) LB plates with your initials, and the date. On one write "Colony #1" and on the other write "Colony #2". Remember to write on the edge of the plate so you can visualize the colonies after they grow.
4. Take your tinfoil package containing your sterile swabs and carefully open it up on the end with the tape.
5. Carefully remove a sterile swab, and crack the cover of the Petri dish that has your "Colony #1" on it.
6. Gently touch the swab to "Colony #1" being careful to avoid touching other colonies. REMEMBER WE WANT TO *ISOLATE* ONE TYPE OF BACTERIA!
7. Streak the fresh LB Plate that you labeled with "Colony #1".
8. Place the swab in the plastic bag provided at your station being careful not to touch ANYTHING ELSE.
9. Repeat these steps for "Colony #2".
10. IMMEDIATELY WASH YOUR HANDS WITH SOAP AND WATER.
11. Let your teacher know when you are done so she or he can collect the plates and used swabs for safe disposal.

Handout #5

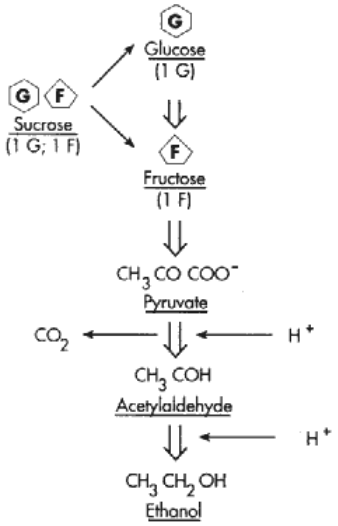
Name: _____
Date: _____
Period:_____

You have successfully isolated two different types of bacteria! CONGRATULATIONS! Now that we know something about *who* is there, lets find out what they do, and relate this back to the environment that they were isolated from (your degradation soil microcosm, or carrot surface). Microbiologists determine the biochemical or metabolic capabilities of an organism by inoculating or introducing these bacteria to growth media with specific substrates and looking at the results. You have four different biochemical tests available. It is your job to come up with hypotheses about the outcome of these test. When forming you hypotheses, keep in mind the things we talked about during the brain storming session about what a carrot is made of, and they type of environmental conditions that these bacteria may face in a natural setting.

Phenol Red Broth is a medium that contains a protein called peptone, a Durham tube (the little upside down tube in the larger test tube), one type of carbohydrate (sugar) and phenol red. Phenol red is an ***acid-base indicator***. Acid-base indicators are substances, which change colors depending upon the pH of the solution. Phenol red will change from red to yellow when the medium has a pH < 6.8 (more acidic) and from red to fuchsia when the medium has a pH >7.4 (less acidic).

We will use two different kinds of phenol red broths. One contains glucose and one contains sucrose. If your bacterium **is** able to metabolize (eat) the sugar through the process of ***fermentation*** (see pathway below) an acid by-product is produced, and the medium will turn **yellow**. If the organism **is not** able to metabolize (eat) the sugar but **is** able to metabolize the protein, then the by-product of this reaction is ammonia, which raises the pH of the media and turns it **fuchsia**.

When the organism is able to use the carbohydrate, a gas by-product ***may*** be produced. If it is, an air bubble will be trapped inside the Durham tube. What is this gas by-product?



(or lactic acid)

Do you think that your bacteria are able to ferment glucose or sucrose? Do you think that they will produce gas as a by-product?

GLUCOSE

H1:

H2:

Which hypothesis have you and your group decided on?
Bacterium 1: _____

Bacterium 2: _____

SUCROSE

H1:

H2:

Which hypothesis have you and your group decided on?
Bacterium 1: _____

Bacterium 2: _____

Starch agar is a medium that tests the ability of an organism to produce certain *exoenzymes* (enzymes that are produced inside of the cell, and transported out of the cell to break down certain substances) that hydrolyze starch. What does *hydrolyze* mean? Starch molecules are too big to enter the bacterial cell, so some bacteria secrete these exoenzymes to degrade or break down the starch into subunits that are SMALL and SOLUBLE that can then be used by the bacteria.

Starch agar medium has nutrients with starch added. Since no color change occurs in the medium when organisms hydrolyze starch (unlike the phenol-red broth) we add iodine to the plate after incubation. Iodine turns the plate blue, purple, or black (depending on the concentration of iodine) in the presence of starch. A clear area around the bacterial growth indicates that the organism has hydrolyzed starch.

Do you think that your bacteria are able to hydrolyze starch?

H1:

H2:

Which hypothesis have you and your group decided on?
Bacterium 1: _____

Bacterium 2: _____

Dase agar is a medium that tests the ability of an organism to produce an exoenzyme, called deoxyribonuclease or DNase that hydrolyzes DNA. DNase agar contains nutrients for the bacteria, DNA, and methyl green as an indicator. Methyl green is a positively charged molecule that binds to the negatively charged DNA. Deoxyribonuclease allows the organisms that produce it to break down DNA into smaller fragments. When the DNA is broken down, it no longer binds to the methyl green, and a clear halo will appear around the areas where the DNase-producing organism has grown.

Why would the ability to break down DNA be beneficial for a soil-dwelling organism to break down DNA?

Do you think that your bacteria are able to degrade DNA?

H1:

H2:

Which hypothesis have you and your group decided on?
Bacterium 1: _____

Bacterium 2: _____

Handout #6

INOCULATION OF MEDIA

Phenol red broth

1. Label your plates with the isolated bacteria "1" and "2".

2. Using the tape provided, label one glucose tube "G" with your initials, the date and a "1" and the other glucose tube "G" with your initials, the date and a "2".

3. Label one sucrose tube "S" with your initials, the date and a "1" and the other sucrose tube "S" with your initials, the date and a "2".

4. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "1".

5. Holding the tube labeled "G1" at an angle, carefully remove the cover, and place the swab in the media.

6. Remove the swab and place it in the plastic bag provided.

7. Repeat this process for the tube labeled "S1", the sucrose tube.

8. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "2".

9. Holding the tube labeled "G2" at an angle, carefully remove the cover, and place the swab in the media.

10. Remove the swab, replace the cover on the tube and place the swab in the plastic bag provided.

11. Repeat this process for the tube labeled "S2", the sucrose tube.

Starch Agar

1. Label one starch plate your initials, the date and "1" and the other starch plate your initials, the date and "2".

2. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "1".

3. Lift the lid on the starch plate labeled "1" at an angle and carefully swab a small patch on the plate (see below). DO NOT SWAB THE WHOLE PLATE!!!!!!!!!!!

4. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "2".

5. Lift the lid on the starch plate labeled "2" at an angle and carefully swab a small patch on the plate. DO NOT SWAB THE WHOLE PLATE!!!!!!!!!!!

DNase Agar

1. Label one DNase plate your initials, the date and "1" and the other DNase plate your initials, the date and "2".

2. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "1".

3. Lift the lid on the DNase plate labeled "1" at an angle and carefully swab a small patch on the plate (see above). DO NOT SWAB THE WHOLE PLATE!!!!!!!!!!!

4. Carefully remove one swab from the sterile tinfoil packet, and touch it to the isolated bacteria on the plate that you labeled "2".

5. Lift the lid on the DNase plate labeled "2" at an angle and carefully swab a small patch on the plate. DO NOT SWAB THE WHOLE PLATE!!!!!!!!!!!

Handout #7

Name: _____
Date: _____
Period:_____

1. Phenol Red Broth – Glucose and Sucrose
Remember that you are assessing if your organism is able to ferment glucose or sucrose, and if they are able to produce CO_2 gas as a by-product.

Record **A** for **acid production** (yellow color), **NA** for **no acid production** (red color) ; **G** for **gas production** (bubble in Durham tube), **NG** for **no gas production** (no bubble in Durham tube) for each isolate and each sugar.

Inoculum	Glucose	Sucrose
Isolate 1		
Isolate 2		

2. Starch Plate
Remember that you are assessing if your organism is able to secrete exoenzymes which break down starch, making it SMALL and SOLUBLE. This allows the bacteria to take it up and use it for various metabolic processes.

Apply a small amount of Lugol's iodine around the patch of bacteria. If the bacteria are able to break down the starch in the plate, then you will see a clear zone around the bacteria. If the bacteria are not able to break down the starch in the plate, then you will see the blue-black color of the starch-iodine complex right up to the boarder of the bacterial growth.

If your bacteria **are able** to break down starch, record a + (indicates that your organism is starch positive). If your bacteria **are not able** to break down starch, record a – (indicates that your organism is starch negative).

Inoculum	Starch
Isolate 1	
Isolate 2	

3. DNase Plate
Remember that you are assessing if your organism is able to secrete exoenzymes which break down DNA, making it SMALL and SOLUBLE, which allows the bacteria to take it up and use it to make more DNA.
Look at the area immediately around the bacterial growth. If the bacteria are able to break down the DNA in the plate, then there will be a clear area around the bacterial growth. If the bacteria aren't able to break down the DNA in the plate, then the green color in the plate will go right up to the boarder of the bacterial growth. It may help to look at this plate against a white background, like the back of this paper.

If your bacteria **are able** to break down DNA, label + (indicates that your organism is DNase positive). If your bacteria **are not able** to break down DNA, label – (indicates that your organism is DNase negative).

Inoculum	Dnase
Isolate 1	
Isolate 2	

Now that you have made some observations about your bacteria, write a brief paragraph describing each of your bacterial isolates. Make sure to include the color (yellow, white, cream, etc.) of your bacteria, the texture (dry, gooey, snotty, runny, etc) of your bacteria, the smell of your bacteria (earthy, woodsy, like rotten eggs, etc) and the biochemical properties that you have recorded above.

Bacterial Isolate 1:

Bacterial Isolate 2:

APPENDIX C

Introduction

Advances in the field of microbial ecology have allowed researchers to begin to understand the complex community found in the gastrointestinal tracts of many organisms. Through the use of molecular techniques such PCR amplification and sequencing of the small subunit rRNA gene, we are able to describe and often identify the organisms present in these ecosystems (3). The ciliated protist *Balantidium jocularum* (Family: Balantidiidae) belongs to the same taxonomic family as the human pathogen *Balantidium coli* (14, 22) and has been observed in the gastrointestinal tract of the herbivorous marine surgeonfish *Naso tonganus* (formerly *Naso tuberosus*) (11, 13). This is the same fish species that harbors populations of the unusually large bacterial symbiont *Epulopiscium* type B (4, 6, 20). In this preliminary study, PCR amplification of the 18S rRNA gene of both extracted community DNA and individual *Balantidium* cells (7) was employed to characterize these intestinal ciliates of *N. tonganus*.

Background

Rumen ciliates were discovered 170 years ago (16). The most abundant of the rumen protozoa, ciliates contribute to the digestion and metabolism of plant matter (24, 25). While little is still known about the microbial community in the intestinal tract of marine herbivorous fishes, over the past decade it has become apparent that these fish also harbor ciliated protozoa as intestinal symbionts (10, 12, 13, 15). One of

theses ciliates, *B. jocularum* was described as a new species of the genus *Balantidium* in 1993 (11). Several unique features lead to the identification of *B. jocularum* as a new species, including a distinct curvature of the vestibulum, specialized cilia to the right of the vestibulum referred to as the dextr-oral field, and unusual body dimensions.

B. jocularum is found in the same fish host as the unusually large bacterium *Epulopiscium* sp. type B (13). Due to its size, *Epulopiscium* is able to avoid predation by many of the symbionts found in the intestinal tract of the host fish. However, *B. jocularum* has been shown to ingest *Epulopiscium* and package them in food vacuoles making these balantidia opportunistic feeders, similar to *B. coli* (13). In an attempt to better understand the phylogenetic relationship of the *Balantidium* spp. found in the intestinal tract of *N. tonganus*, PCR amplification of the 18S rRNA gene from both extracted community DNA and individual trophozoite cells of two different populations of *Balantidium* spp. was carried out. Trophozoites, or the metabolically active cells were selected for single-cell PCR analysis as they do not possess a double walled covering like the cyst stage and as such should be more amenable to lysis (22). Trophozoites are easily identifiable by their visible cilia covering the surface of the cell.

The first population of cells (hereafter *B. jocularum*-like, Figure C.1a) possessed characteristics similar to those of the previously described surgeonfish symbiont, *B. jocularum* (11, 13). These cells were smaller in size and more elongate. The second population of cells (hereafter "Big" *Balantidium*, Figure C.1b) were larger

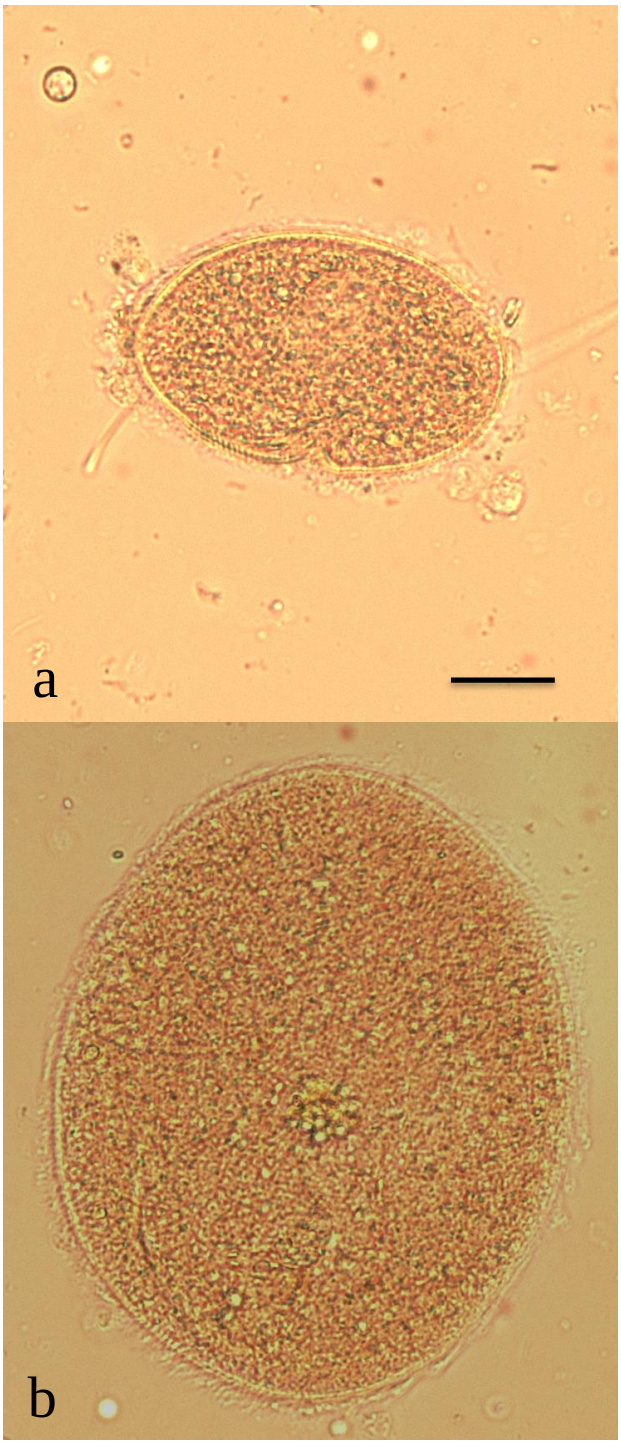


Figure C.1. Bright field microscopy images of cells representing the *B. jocularam*-like (a) and the "Big" *Balantidium* (b) populations used in these studies. (Scale bar: 20 μm)

and rounder. The PCR products generated from these two populations of cells were used in cloning reactions and sequencing of these clones was carried out.

Materials and Methods

Balantidium collection

Balantidium cells were obtained from *Naso tonganus*, collected by spearfishing on reefs around Lizard Island, Australia. Sections of the gut were removed, and intestinal contents were fixed in 80% ethanol and stored at -20°C. Samples were enriched for *Balantidium* cells by aliquoting 1 ml of the sample into the well of a sterile tissue culture plate and removing all visible materials that were not *Balantidium* cells using a standard Gilson micropipettor and a dissecting microscope (Nikon SMZ-U). In addition, individual *Balantidium* cells were collected, transferred five times through sterile ethanol wash buffer (80% ethanol; 145 mM NaCl; 50 mM Tris-HCl, pH 8.0; 0.05% Igepal) and rinsed in sterile deionized water.

Cell length measurements

Sample "fish #5, 18-10-03" was used for the *B. jocularum*-like cell analysis and sample "fish #4, 18-10-03" was used for the "Big" *Balantidium* analysis. Images of the two populations of *Balantidium* cells were acquired using a Zeiss Axiocam CCD Camera and a Zeiss Axiovert 200 microscope equipped with an LCPlanF1 40X objective. Using Zeiss Axiovision Software (calibrated with a stage micrometer), measurements of the long axis of the cells were determined. Statistical analysis (2-sample *t*-test) was performed using SAS software, Version 9.1 of the SAS System for

Windows.

DNA extraction

Community genomic DNA was extracted from samples that had been enriched for *Balantidium* cells, in triplicate, from approximately 3 ml of the ethanol-fixed gut contents. Sample "fish #5, 18-10-03" was used for the *B. jocularum*-like cell analysis and sample "fish #4, 18-10-03" was used for the "Big" *Balantidium* analysis. Gut-contents were pelleted by centrifugation and washed twice with filter sterilized 10 mM Tris, pH 8. The samples were resuspended in 300 µl 10 mM Tris, pH 8. Proteinase K was added to a final concentration of 100 µg/ml and the samples were incubated at 50°C for 1 hour, followed by six rounds of freeze-thaw. Cell lysate was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1) and the nucleic acids were precipitated with 0.3 M sodium acetate and ethanol. The pellet was rinsed with 70% ethanol and air-dried. The pellet was dissolved in sterile water and treated with RNase A (10 µg/ml) at 37°C for 30 minutes. Triplicate extractions were pooled and stored at -20°C.

For single-cell PCR, a proteinase K solution (as above) was irradiated for 3 minutes using a UV transilluminator (FisherBiotech) and then aliquoted into each PCR tube. One washed *Balantidium* cell was added to each tube and incubated at 50°C for 1 hour. The tube was then heated at 95°C for 15 min, to inactivate the proteinase K.

PCR amplification and 18S rRNA gene library construction.

The 18S rRNA genes were amplified from genomic DNA using three different

eukaryotic specific primer sets: EK1F (5'- CTGGTTGATCCTGCCAG-3') and EK1520R (5´-CYGCAGGTTCACCTAC-3´) (19) EK82F (5' – GAAACTGCGAATGGCTC -3') (5) and EK1520R and PP1F (5' - CGTAGGTGAACCTGCGGAAG - 3') and PP1R (5' - GCTCGGGGATCGCTCTACTT -3') (9) (Integrated DNA Technologies). The best amplification was obtained from the EK1F/EK1520R primer set, so these were used in all subsequent amplification reactions. Amplifications were performed in triplicate and carried out in 50 μl reactions containing 1X HotStarTaq PCR buffer (Qiagen), 0.2 μmol of each primer and 2.5 U of *Taq* DNA polymerase. Genomic DNA (1 μl) was added to each amplification reaction. Reaction mixtures were subjected to the following temperature cycling profile on a PTC-200 Thermal Cycler (MJ Research): 15 min at 95°C followed by 29 cycles of denaturation (94°C for 30 sec), annealing (50°C for 30 seconds) and extension (72°C for 2 min). A final extension step of 72°C for 10 minutes was included. A positive control (*B. subtilis*) and a no template control were included. Amplification products were analyzed using a 1.2% agarose gel in 1X Tris-Acetic Acid-EDTA stained with ethidium bromide (21).

Amplification products were cloned into the plasmid vector pCR 2.1-TOPO. Plasmids were transformed into *Escherichia coli* TOP10 competent cells according to manufacturer's instructions (Invitrogen). White colonies were randomly selected and screened by PCR for the appropriate insert using the vector specific primers M13F (5'-GTAAAACGACGGCCAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3'). Plasmid DNA was recovered using the QIAprep Miniprep Kit (Qiagen).

Clone Analysis.

DNA sequence of each clone was determined with an ABI 3730 DNA Analyzer (Applied Biosystems, Inc.) using Big Dye Terminator chemistry and AmpliTaq-FS DNA polymerase. The M13F primer was used in all sequencing reactions. Full-length 18S rRNA gene sequence was determined for unique clones using the M13R primer and 515F (5'-GTGCCAGCMGCCGCGGTAA-3') (23) and 1195RE (5'-GGGCATCACAGACCTG-3') (17). Sequence reads were assembled using the Sequencher software package (Gene Codes) and compared to sequences in GenBank by using the BLAST (Basic Local Alignment Search Tool) (1) to identify their closest relatives. Full-length 18S rRNA gene sequences were aligned using ClustalW (18)

Results

Cell length measurements were obtained for a total of 25 cells from each population. Cell lengths from the smaller, *B. jocularum*-like cells ranged from 40.8 μm to 70.1 μm with an average length of 58.9 μm. "Big" *Balantidium* cells ranged in size from 95.8 μm to 141.1 μm and had an average length of 115.5 μm (Table C.1). When these data were subjected to a 2-sample t-test there was a statistically significant difference between the lengths of the two populations: $t(39) = 16.20$, $p < 0.001$.

The primer set EK1F and EK1520R resulted in the best amplification and were used for all subsequent PCR reactions. Since this initial approach resulted in extracted DNA from all cells in the fixed intestinal contents the resulting PCR products used for

Table C.1. Length measurements of individual cells and cell length averages for the two populations used in this study.

Cell Number	*B. jocularum*-like cell length (μm)	"Big"*Balantidium* cell length (μm)
1	67.6	102.6
2	45.3	115.8
3	53.4	113.9
4	49.2	110.5
5	51.7	103.4
6	62.9	122.9
7	66.9	98.7
8	40.8	95.8
9	57.5	123.3
10	65.8	122.0
11	68.3	105.1
12	60.1	102.9
13	64.7	116.7
14	68.3	141.1
15	63.1	116.1
16	60.4	1140.0
17	66.2	137.4
18	56.1	151.6
19	45.1	111.7
20	55.1	113.3
21	49.7	100.8
22	67.0	106.2
23	48.4	101.4
24	69.8	102.5
25	70.1	131.2
Average (μm)	58.9	155.5

cloning were a mixture of 18S rRNA genes from all eukaryotic organisms present in the sample. Screening of these clones resulted in several clones with the correct sized insert. When those clones were sent for sequencing, only two of the 11 resulted in BLAST hits with high homology to other ciliate sequences. The remaining nine were affiliated with 18S rRNA gene sequences recovered from fish.

Individual *B. jocularum*-like cells pretreated with proteinase K were used in a subsequent PCR amplification and cloning reaction. Of the seven clones sent for sequencing, all but one resulted in BLAST hits with high homology to other ciliates. In order to assess the similarity between a morphologically unique group of *Balantidium* cells ("Big" *Balantidium*) and the *B. jocularum*-like cells, individual "Big" *Balantidium* cells were micromanipulated out of the gut contents, pretreated with proteinase K and used in PCR amplification and subsequent cloning reactions. Of the seven clones sent for sequencing, five resulted in BLAST hits showing high similarity to ciliate sequences. Thirteen full length sequences, eight from *B. jocularum*-like cells and five from "Big" *Balantidium* cells were aligned against the 18S rRNA gene sequence from the human pathogen *B. coli* using ClustalW. All sequences showed 98 – 99% similarity to each other, and 95% similarity to *B. coli* (Figure C.2).

Discussion

Initial cloning attempts using PCR products from community DNA extractions enriched for *B. jocularum*-like cells proved troublesome: a large number of clones contained non-*Balantidium* sequences, specifically sequences affiliated with fish. This

```
B.jocularum-like GTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGACTAAGCCATGCATGTCTAAGTAT 60
"Big"Balantidium GTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGACTAAGCCATGCATGTCTAAGTAT 60
B. coli          GTTGATCCTGCCAGTAGTCATATGCTTGTCTCAAAGACTAAGCCATGCATGTCTAAGTAT 60
                 ************************************************************

B.jocularum-like AAATAACTACACAGTAAAACTGCGAATGGCTCATTAAAACAGTTATAGTTTATTTGATAC 120
"Big"Balantidium AAATAACTACACAGTAAAACTGCGAATGGCTCATTAAAACAGTTATAGTTTATTTGATAC 120
B. coli          AAATAACTACACAGTAAAACTGCGAATGGCTCATTAAAACAGTTATAGTTTATTTGATAC 120
                 ************************************************************

B.jocularum-like ATTAGATGGATAACTGTAGAAAAACTAGAGCTAATACATGCCAAGGYCGAAAGGCCGTAT 180
"Big"Balantidium ATTAGATGGATAACTGTAGAAAAACTAGAGCTAATACATGCCTAGGCCGAAAGGTCGTAT 180
B. coli          ATTAAATGGATAACTGTAGAAAAACTAGAGCTAATACATGCCGAGGCCGTAAGGTCGTAT 180
                 ****.*************************************** *** **:**** *****

B.jocularum-like TTATTAGATATGCCAATTAAGGTGAATCATAATAACTTCGCAAATCGCGATTTCGTCGCG 240
"Big"Balantidium TTATTAGATATGCCAATTAAGGTGAATCATAATAACTTCGCAAATCGCGATTTCGTCGCG 240
B. coli          TTATTAGATATTCCAATTAAGGTGAATCATAATAACTTCGCAAATCGCGATTTTGTCGCG 240
                 *********** ****************************************** ******

B.jocularum-like ATAAATCATYCAAGTTTCTGCCCTATCATGCTTTCGATGGTAGTGTATTGGACTACCATG 300
"Big"Balantidium ATAAATCATCCAAGTTTCTGCCCTATCATGCTTTCGATGGTAGTGTATTGGACTACCATG 300
B. coli          ATAAATCATCCAAGTTTCTGCCCTATCATGCTTTCGATGGTAGTGTATTGGACTACCATG 300
                 ********* ************************************************

B.jocularum-like GCTTTTACGGGTAACGGGGAATTAGGGTTCGATTCCGGAGAAGGAGCCTGAGAAACGGCT 360
"Big"Balantidium GCTTTTACGGGTAACGGGGAATTAGGGTTCGATTCCGGAGAAGGAGCCTGAGAAACGGCT 360
B. coli          GCTTTCACGGGTAACGGGGAATTAGGGTTCGATTCCGGAGAAGGAGCCTGAGAAACGGCT 360
                 ***** ********************************************************

B.jocularum-like ACTACATCTACGGAAGGCAGCAGGCGCGTAAATTACCCAATCCTGACTCAGGGAGGTGGT 420
"Big"Balantidium ACTACATCTACGGAAGGCAGCAGGCGCGTAAATTACCCAATCCTGACTCAGGGAGGTGGT 420
B. coli          ACTACATCTACGGAAGGCAGCAGGCGCGTAAATTACCCAATCCTGACTCAGGGAGGTGGT 420
                 ************************************************************

B.jocularum-like GACAAGATATAACGACGTGATTAAAATCACGATTGTAGTGAGGGTATTCCAAACCGAACC 480
"Big"Balantidium GACAAGATATAACGACGTGATTAAAATCACGATTGTAGTGAGGGTATTCCAAACCGAACC 480
B. coli          GACAAGATATAACGACGCAATTTATTTTGTGATTGTAGTGAGGGTATTCCAAACCGAACC 480
                 ***************** .***:*::* . ******************************

B.jocularum-like ACTAGTACAATTAGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCTAATAG 540
"Big"Balantidium ACTAGTACAATTAGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCTAATAG 540
B. coli          ACTAGTACGATTAGAGGGCAAGTCTGGTGCCAGCAGCCGCGGTAATTCCAGCTCTAATAG 540
                 ********.***************************************************

B.jocularum-like CGTATATTAAAGTTGCTGCAGTTAAAAAGCTCGTAGTTGGATTTCAAGGATTATATGCAC 600
"Big"Balantidium CGTATATTAAAGTTGCTGCAGTTAAAAAGCTCGTAGTTGGATTTCAAGGATTATATGCAC 600
B. coli          CGTATATTAAAGTTGTTGCAGTTAAAAAGCTCGTAGTTGGATTTCAAGGCGTGTATACTC 600
                 *************** ********************************. *.***.*:*

B.jocularum-like TTCG-GTGTATGTACCCTACTAGCCTTCGGGCTGTTACTGTGAGAAAATTAGAGTGTTTC 659
"Big"Balantidium TTCG-GTGTATGTACCCTACTAGCCTTTGGGCTGTTACTGTGAGAAAATTAGAGTGTTTC 659
B. coli          TTTYTGAGTATGCTACCTACTAGTCTCTG-ACTGTTACTGTGAGAAAATTAGAGTGTTTC 659
                 **   *:***** :.******** **  * .******************************

B.jocularum-like AAGCAGGCTATTGCAAGAATACATTAGCATGGAATAACGAATGTYRCTAGAATCTTGGTT 719
"Big"Balantidium AAGCAGGCTATTGCAAGAATACATTAGCATGGAATAACGAATGTTACTAGAATCTTGGTT 719
B. coli          AAGCAGGCTTTTGCAAGAATACATTAGCATGGAATAACGAATGTGTCTAGAATCTTGGTT 719
                 *********:*********************************  ***************

B.jocularum-like TGTTCTAG-TTGCGGTTAATAGGGACAGTTGGGGGCATTAGTATTTAATAGTCAGAGGTG 778
"Big"Balantidium TGTTCTAG-TTGCGGTTAATAGGGACAGTTGGGGGCATTAGTATTTAATAGTCAGAGGTG 778
B. coli          AATTCTAGATTGCGATTAATAGGGACAGTTGGGGGCATTAGTATTTAATTGTCAGAGGTG 779
                 :.****** ******.*********************************:***********

B.jocularum-like AAATTCTTGGATTTGTTAAAGACTAACTTATGCGAAAGCATTTGCCAAGGATGTTTTCAT 838
"Big"Balantidium AAATTCTTGGATTTGTTAAAGACTAACTTATGCGAAAGCATTTGCCAAGGATGTTTTCAT 838
B. coli          AAATTCTTGGATTTGTTAAAGACTAACGTATGCGAAAGCATTTGCCAAGGATGTTTTCAT 839
                 **************************** ********************************

B.jocularum-like TAATCAAGAACGAAAGATAGGGGATCAAAGACAATCAGATACTGTCGTAGTCCTATCTAT 898
"Big"Balantidium TAATCAAGAACGAAAGATAGGGGATCAAAGACAATCAGATACTGTCGTAGTCCTATCTAT 898
B. coli          TAATCAAGAACGAAAGATAGGGGATCAAAGACAATCAGATACTGTCGTAGTCCTATCTAT 899
                 ************************************************************
```

```
B.jocularum-like AAACTATGCCGACTAGGGATTGGAGTG-AGATTGCATCACTTCAGTACCTTATGAGAAAT 957
"Big"Balantidium AAACTATGCCGACTAGGGATTGGAGTG-AAATAGCATCACTTCAGTACCTTATGAGAAAT 957
B. coli          AAACTATGCCGACTAGGGATTGGAATGGTTATAACGCCGTTTCAGTACCTTATGAGAAAT 959
                 ************************.** : **:.*. *. ********************

B.jocularum-like CAAAGTCTTTGGGTTCTGGGGGGAGTATGGTCGCAAGACTGAAACTTAAAGAAATTGACG 1017
"Big"Balantidium CAAAGTCTTTGGGTTCTGGGGGGAGTATGGTCGCAAGACTGAAACTTAAAGAAATTGACG 1017
B. coli          CAAAGTCTTTGGGTTCWGGGGGGAGTATGGTCGCAAGACTGAAACTTAAAGAAATTGACG 1019
                 *************** ******************************************

B.jocularum-like GAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTAC 1077
"Big"Balantidium GAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTAC 1077
B. coli          GAAGGGCACCACCAGGAGTGGAGCCTGCGGCTTAATTTGACTCAACACGGGGAAACTTAC 1079
                 ************************************************************

B.jocularum-like CAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCTCTTTCTTGATTCTATGGGTGGT 1137
"Big"Balantidium CAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCTCTTTCTTGATTCTATGGGTGGT 1137
B. coli          CAGGTCCAGACATAGTAAGGATTGACAGATTGATAGCTCTTTCTTGATTCTATGGGTGGT 1139
                 ************************************************************

B.jocularum-like GGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGA 1197
"Big"Balantidium GGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGA 1197
B. coli          GGTGCATGGCCGTTCTTAGTTGGTGGAGTGATTTGTCTGGTTAATTCCGATAACGAACGA 1199
                 ************************************************************

B.jocularum-like GACCTTAACCTGCTAACTAGTCGCGTCCATTTTATGGTTGCTGACTTCTTAGAGGGACTA 1257
"Big"Balantidium GACCTTAACCTGCTAACTAGTCGCCTCCATTTTATGGATGCTGACTTCTTAGAGGGACTA 1257
B. coli          GACCTTAACCTGCTAACTAGTCTAATCCATTTTATGGAAMATGACTTCTTAGAGGGACTA 1259
                 ********************** . ************:: .********************

B.jocularum-like TGTGATATAAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTATATGTCC 1317
"Big"Balantidium TGTGCTATAAGCACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTATATGTCC 1317
B. coli          TGT-ATTTAAATACATGGAAGTTTGAGGCAATAACAGGTCTGTGATGCCCTTATATGTCC 1318
                 *** .*:***. ************************************************
B.jocularum-like TGGGCTGCACGCGTGCTACACTGATACATACAACAAGTACCTAGCCCGCTAGGGTATGGC 1377
"Big"Balantidium TGGGCTGCACGCGTGCTACACTGATACATACAACAAGTACCTAGCCCGCTAGGGTATGGC 1377
B. coli          TGGGCTGCACGCGTGCTACACTGATGCATACAACAAGTGCCTAGCCCGCCAGGGTATGGC 1378
                 *************************.************.********** **********

B.jocularum-like AATCTGGAATATGTATCGTGATGGGGATTGATCTTTGCAATTATAGATCATGAACGAGGA 1437
"Big"Balantidium AATCTGGAATATGTATCGTGATGGGGATTGATCTTTGCAATTATAGATCATGAACGAGGA 1437
B. coli          AATCTCGAATATGCATCGTGATGGGGATAGATCTTTGCAATTATAGATCTTGAACGAGGA 1438
                 ***** ******* **************:********************:**********

B.jocularum-like ATTCCTAGTAAATGCAAGTCATCATCTTGCGTTGATTATGTCCCTGCCCTTTGTACACAC 1497
"Big"Balantidium ATTCCTAGTAAATGCAAGTCATCATCTTGCGTTGATTATGTCCCTGCCCTTTGTACACAC 1497
B. coli          ATTCCTAGTAAGTGCAAGTCATCATCTTGCATTGATTATGTCCCTGCCCTTTGTACACAC 1498
                 ***********.******************.****************************

B.jocularum-like CGCCCGTCGCTCCTACCGATACCGGGTGATCCGGTGAACCTTTTGGACCGCATTGCGGAA 1557
"Big"Balantidium CGCCCGTCGCTCCTACCGATACCGGGTGATCCGGTGAACCTTTTGGACCGCATTGCGGAA 1557
B. coli          CGCCCGTCGCTCCTACCGATACCGGGTGATCCGGTGAACCTTTTGGACCGCTAYGCGGAA 1558
                 ******************************************************:: ******

B.jocularum-like AAATAAGTAAACCATATCACCTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGA 1617
"Big"Balantidium AAATAAGTAAACCATATCACCTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGA 1617
B. coli          AAATAAGTAAACCATATCACCTAGAGGAAGGAGAAGTCGTAACAAGGTTTCCGTAGGTGA 1618
                 ************************************************************

B.jocularum-like ACCTGCAG 1625
"Big"Balantidium ACCTGCAG 1625
B. coli          ACCTGCGG 1626
                 ******.*
```

Figure C.2. Alignment of *Balantidium*-like consensus sequence, "Big" *Balantidium* consensus sequence and *B. coli* 18S rRNA genes

was most likely due to the fact that epithelial cells of the intestinal wall sloughed off and were lysed during the DNA extraction procedure. To increase the efficiency of our cloning reactions, we decided to use individual *B. jocularum*-like cells in our PCR reactions. Utilizing this approach six of the seven clones sequenced affiliated with other ciliate sequences including *B. coli.* The seventh clone exhibited high sequence homology (98%) to the 18S rRNA gene sequence of fish.

Initial surveys of the gut contents revealed that in addition to the *B. jocularum*-like cells that had previously been described, a larger, morphologically distinct *Balantidium* ("Big" *Balantidium*) was also present in the sample. In order to investigate the phylogenetic relationship between these two populations, the same approach used with the *B. jocularum*-like cells was employed. Five of the six clones recovered showed high homology (95%) to *B. coli* 18S rRNA gene sequences. The sixth clone exhibited 98% sequence homology to the 18S rRNA gene sequence of fish. Surprisingly these sequences also showed very high homology (98 – 99%) to the *B. jocularum*-like cells, even though these two populations were significantly different from each other in overall length, suggesting that they are separate populations.

There have been reports of some ciliate protists changing their cell size and shape in response to increase predation risk or fluctuating resource availability (2). The hypotrich cilate of the genus *Euplotes* appear to significantly increase their width, which provides protection against predation of gape predators such as *Stenostomum* sp. (2). Perhaps this population of unique cells has also undergone some sort of morphological change in response to environmental cues. In a study looking at ciliates

in the tanks of bromeliads, a gigantic tetrahymenid ciliate was discovered (8). However when the 18S rRNA gene sequence from this novel organism was compared to those found in the NCBI database, it was found to be nearly identical (99.3% similarity) to a small, ubiquitous ciliate, *Tetrahymena corlissi*. This result was not supported by the morphological and ecological data, which suggested that this unique ciliate should be placed in a distinct family. While the 16S rRNA gene has become the "gold standard" for classification of bacteria, perhaps another molecular marker will better represent the morphological divergence seen in ciliates more adequately than the 18S rRNA gene.

While the results of this survey are far from conclusive, requiring additional sequence data of additional markers, this work has provided a foundation on which to build a better understanding of the ciliated protists that inhabit the intestinal tract of the herbivorous surgeonfish *Naso tonganus* thus furthering our overall understanding of this complex symbiotic microbial community.

References

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

2. **Altwegg, R., Marchinko, K. B., Duquette, S. and Anholt, B. R**. 2004. Dynamics of an inducible defence in the protist *Euplotes*. Arch. Hydrobiol. **160:** 431-446.

3. **Amann, R. I., Ludwig, W and Schleifer, K. H**. 1995. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. Microbiol. Rev. **59:** 143-169.

4. **Angert, E. R., Clements, K. D. & Pace, N. R.** 1993. The largest bacterium. Nature. **362:** 239-241.

5. **Auinger, B., Pfandl, K. and Boenigk, J.** 2008. Improved methodology for identification of protists and microalgae from plankton samples preserved in Lugol's iodine solution: combining microscopic analysis with single-cell PCR. Appl. Environ. Microbio. **74:** 2505-2510.

6. **Clements, K. D. and S. Bullivant** 1991. An unusual symbiont from the gut of surgeonfishes may be the largest known prokaryote. J. Bacteriol. **173**: 5359-5362.

7. **Engelkirk, P. G. & Duben-Englekirk, J.** 2008. Laboratory Diagnosis of Infectious Disease: Essentials of Diagnostic Microbiology. Philadelphia, PA, Lipincott, Williams and Wilkins.

8. **Foissner, W., Struder-Kypke, M., va der Staay, G. W. M., Moon-van der Staay, S. and Hackstein, J. H. P.** 2003. Endemic ciliates (Protozoa, Ciliophora) from tank bromeliads (Bromeliaceae): a combined morphological, molecular, and ecological study. Europ. J. Protistol. **39:** 365-372.

9. **Gorokhova, E., Lehtiniemi, M., Viitasalo-Fro, S. & Haddockc, S. H. D.** 2009. Molecular Evidence for the occurrence of ctenophore *Metensia ovum* in the northern Baltic Sea and implications for the status of the *Mnemiopsis leidyi* invasion. Limnol. Oceanogr. **54:** 2025-2033.

10. **Grim, N. J.** 1988. A somatic kinetid study of the pycnotrichid ciliate *Vestibulongum corlissi* n. g., n. sp. (class: Litostomatea), symbiont in the intestines of the surgeonfish, *Acanthurus xanthopertus*. J. Protozool. **35:** 227-230.

11. **Grim, N. J.** 1993. Description of Somatic Kineties and Vestibular Organization of *Balantidium jocularum* sp. n. and Possible Taxonoic Implications for the Class Litstomatea and the Genus *Balantidium*. Acta. Protozoologica **32:** 37-45.

12. **Grim, N. J.** 1993b. Endonuclear symbionts within a symbiont: the surgeonfish intestinal symbiont, *Balantidium jocularum* (Ciliophora) is host to a grampositive macronuclear inhabiting bacterium. Enterocyt. Cell Res. **9:** 209-214.

13. **Grim, N. J.** 2006. Food Vacuole Contents in the Ciliate, *Balantidium jocularum* (*Balantididae*), a Symbiont in the Intestine of the Surgeonfish, *Naso tonganus* (*Acanthuridae*). J. Eukaryot. Microbiol. **53:** 269 - 274.

14. **Grim, N. J. & Buonanno, F.** A re-description of the ciliate genus and type species *Balantidium entozoon.* Eur. J. Protistol. **45:** 174-182.

15. **Grim, N. J. & Clements, K. D.** New species of *Balantidium* and *Paracichlidotherus* (Ciliophora) inhabiting the intestines of four surgeonfish species from the Tuvalu Islands, Pacific Ocean. J. Eukaryot. Microbiol. **49:** 146-153.

16. **Gruby, D. & Delafond, H. M.** 1843. Recherches ser des animalcules se dtveloppant en grand nombre dans l'estomac et dans les intestins, pedant la digestion des animaux herbivores et carnivores. Compt. Rend. Acad. Sci. **17:** 1304 - 1308.

17. **Lane, D. J.** 1991. 16S, 23S rRNA sequencing. Nucleic Acid Techniques in Bacterial Systematics. John Wiley, New York. 115-175.

18. **Larkin, M. A., Blackshields, G., Brown, N.P., Chenna, R. McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J. & Higgins, D.G.** 2007. Clustal W and Clustal X Version 2.0. Bioinformatics. **23:** 2947-2948.

19. **Lopez-Garcia, P., Rodriguez-Valera, F., Pedros-Alio, C. & Moreira, D.** 2008. Unexpected diversity of small eukaryotes in deep-sea Antarctic plankton. Nature. **409:** 603-606.

20. **Mendell, J. E., Clements, K.D., Choat, J. H. and Angert, E.** 2008. Extreme polyploidy in a large bacterium. Proc. Natl. Acad. Sci. U S A. **105:** 6730-6734.

21. **Sambrook, J., Fritsch, E. F. and Maniatis, T.** 1989. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

22. **Schuster, F. L. & Ramierz-Avila, L.** 2008. The current world status of *Balantidium coli*. Clin. Microbiol. Rev. **21:** 626-638.

23. **Turner, S., Pryer, K. M., Miao, V. P. W., and Palmer, J. D.** 1999. Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. J. of Eukaryot. Microbiol. **46:** 327-338.

24. **Veira, D. M.** 1986. The role of ciliate protozoa in nutrition of the ruminant. J. Anim. Sci. **63:** 1547-1560.

25. **Wright, A.G, Dehority, B. A. & Lynn, D. H.** 1997. Phylogeny of the Rumen Ciliates *Entodinium*, *Epidinium* and *Polyplastron* (Litostomatea: Entodiniomorphida) Inferred from Small Subunit Ribosomal RNA Sequences. J. Eukaryot. Microbiol. **44:** 61-67.

APPENDIX D

Commonly Used Statistical Computations for T-RFLP and 16S rRNA Gene Clone Libraries

1. **Shannon's Diversity Index (H)** (11): Shannon's index accounts for both the richness and evenness of the species present when estimating the diversity of a community. This index is based on adding up the logs of the proportional abundances of species. As the number of species increases, or the evenness of their abundances increases, the index increases. It usually ranges from 1.5 to 3.5, rarely exceeding 4.

$$H = -\sum_{i=1}^{S} p_i \ln p_i$$

where s is the total number of species in the community (richness) and p_i is the proportion of *s* made up of the i^{th} species.

2. **Simpson's Diversity Index (D)** (8, 12): This index is based on estimating the probability that two individuals randomly selected from the community will be the same species and is weighted more strongly by the abundance of the most dominant species than Shannon's index with a D value of 1 representing no diversity and a D value of 0 representing infinite diversity. Simpson's index is especially good when species richness in the community is low, as the shape of the species abundance distribution doesn't strongly influence the value of the index unless species richness exceeds 10.

$$D = \frac{1}{\sum_{i=1}^{S} p_i^2}$$

or

$$D = 1/\sum_{i=1}^{S} p_i^2$$

where s is the total number of species in the community (richness) and p_i is the proportion of *s* made up of the *i*th species. Simpson's Diversity Index can also be used to calculate the evenness of the sample.

3**. Chao1 (S_1)** (2): A coverage based estimator, this approach uses the number of singletons (the number of species with only a single occurrence in the sample) and doubletons (the number of species with exactly two occurrences in the sample) to estimate the number of species yet to be identified in a given sample.

$$S_1 = S_{obs} + \frac{F_1^2}{2F_2}$$

where S_{obs} is the number of species in the sample, F_1 is the number of singletons and F_2 is the number of doubletons.

4. **Abundance Coverage Estimator (S_{ACE})** (3): ACE is a coverage estimator that separates the observed species into rare and abundant groups. The rare group is then used to estimate the number of undiscovered species.

$$S_{ace} = S_{common} + \frac{S_{rare}}{C_{ace}} + \frac{F_1}{C_{ace}} \gamma^2_{\text{ace}}$$

where S_{common} is the number of species that occur greater than ten times in the sample, S_{rare} is the number of species which occur ten times or less, C_{ace} is the sample abundance coverage estimator and γ^2_{ace} is the estimated coefficient of variation for F_1, the number of singletons.

5. **Rarefaction Curves** (7): Rarefaction curves plot the total number of individuals counted with repeated samplings against the total number of species found in those samplings. The resultant curve increases steeply at first, then gradually levels off. The asymptote leveling point indicates that additional sampling would yield no additional unique species. The total number of species in a community and the number of rare species determines how many individuals must be counted to reach the optimal sample size, assuming that all individuals are equally likely to be sampled.

6. **Bray-Curtis Coefficient** (1, 4, 5): The Bray-Curtis Coefficient is commonly used in ecological studies and measures the degree of similarity between samples. A similarity coefficient of zero means that the two samples are totally dissimilar; while a similarity coefficient of 100 means that the two samples are totally similar. This analysis generates a similarity matrix for a given data set, which can be used in additional analyses (see below).

$$\sigma_{jk} = 100 \left\{ 1 - \frac{\sum_p \left| \Upsilon_{ij} - \Upsilon_{ik} \right|}{\sum_p \left(\Upsilon_{ij} - \Upsilon_{ik} \right)} \right\}$$

Where γ_{ji} = the i^{th} species in sample j and γ_{ki} = the i^{th} species in sample k. Frequently data is transformed using the double root, or fourth root prior to applying the Bray-Curtis Coefficient. This transformation down-weighs the influence of the more abundant species, allowing the mid-range species and more rare species to exert more of an influence on the similarity calculations.

7. Multidimensional Scaling (MDS) (5): MDS is a two-dimensional configuration of the similarity matrix data. Points clustering together in an MDS plot represent samples that are similar in composition, while points that cluster distantly are samples that differ in composition. The stress measure allows for evaluation of how well the MDS configuration reproduces the similarity matrix data, or the goodness of fit, with a stress level less than 0.1 indicating an accurate representation of the data.

8. Species Contributions to Similarity (SIMPER) (4, 5): SIMPER is a statistical testing method, which identifies the individuals responsible for the observed differences between groups of samples. The overall significance of the difference is often assessed by ANOSIM.

9. Analysis of Similarity (ANOSIM) (4, 5): ANOSIM is used for hypothesis testing with regards to spatial and temporal changes between samples. Specifically, it is used to test the null hypothesis that there is no difference between samples. ANOSIM generates an R-value between -1 and +1, with a value of zero supporting the null hypothesis. Pairwise comparisons of R-values allows one to assess how separate groups are, with R-values > 0.75 indicating groups that are well separated, R > 0.5 as overlapping but clearly different and R < 0.25 as barely separable at all.

10. Neighbor joining (NJ) method (9, 10): The Neighbor Joining (NJ) method for inferring phylogenetic trees uses a distance matrix of pairwise genetic distances for step-wise clustering of OTUs in order to compute the most robust tree. When generating a tree based on 16S rRNA gene sequences, distance matrix methods compare aligned sequences to calculate a dissimilarity value between sequence pairs. These values are then used to generate the distance matrix from which phylogenetic relatedness can be estimated. Stepwise clustering in the NJ method involves assessing every sequence pair, and joining those that result in the shortest branch length. This process is repeated, sequentially reducing the complexity of the data set resulting in a tree with the shortest internal branch lengths. NJ has the advantage of being very fast to execute, and recognizes that not all organisms evolve at the same rate. The major disadvantage to generating trees using the Neighbor Joining method is the possibility of incorrect tree topology due to the algorithm selecting for shortest branch lengths at every step of tree construction. However, despite the theoretical shortcomings of NJ, it has been shown to perform well in practice and was used to infer the neighbor joining dendrograms presented in Chapter 3 of this dissertation.

11. Maximum parsimony (MP) method (10): The Maximum Parsimony (MP) method for inferring phylogenetic trees is a non-parametric statistical method that uses discrete character states to build a large number of trees. When using data from 16S rRNA genes, each position in the alignment is a "character" with the specific nucleotide at that position as a "state". These multiple tree topologies are then assessed through exhaustive searches. Finally, the best tree is chosen based upon the specific criteria of the evolutionary model being used. The MP model begins by

assessing a certain tree topology. A site-by-site analysis of the nucleotide sequence data is conducted. The minimum number of nucleotide changes required to explain each site pattern are then calculated and, a parsimony score is generated for that tree topology by summing the number of changes over all sites. This process is repeated for all tree topologies. The tree requiring the smallest number of evolutionary changes/ substitutions for all sequences to derive from a common ancestor is identified as the most parsimonious. Advantages to using the MP method are it is based on characters, so a similarity matrix does not need to be calculated and sequence data is not reduced to a single number. This method also assesses multiple trees. Disadvantages include this method is computationally intensive and slow to execute, does not provide information on the branch lengths and is subject to biases such as long branch attraction.

12. Maximum likelihood (ML) method (10): Like the MP method, the Maximum Likelihood (ML) method generates trees based on discrete character sets and exhaustive searches. However, ML utilizes parametric statistical methods to determine the probability or likelihood of phylogenetic trees. This method calculates the probability that a specific tree, based on a specific evolutionary model would have generated the sequence data presented. The ML method assumes that each nucleotide position evolves independently and uses this to calculate the likelihood of a specific nucleotide at that site. The product of the likelihoods at each site is the overall likelihood for the tree. The maximum likelihood tree is the one with the highest probability. There are several advantages of the ML method. When compared with other methods, the ML method has the lowest variance and is statistically well founded. Like the MP method, the ML method uses all of the sequence information, as well as evaluating many tree topologies for the best fit. Disadvantages of the ML method are it is slower to execute, and the results are dependent upon the evolutionary model used in the analysis.

13. Bootstrap analysis (10): Bootstrap analysis is a sampling technique that statistically analyzes the reliability of inferred trees by resampling from the original dataset. In bootstrap analysis, a new alignment, or bootstrap replicate is generated by randomly selecting columns from the original alignment. In this new alignment, characters in individual columns may be represented once, more than once or not at all. This process is repeated a specified number of times (between 200 and 2000 samplings is recommended). A tree is then constructed from each of these artificial datasets and the bootstrap proportion, or frequency with which a given branch is found on the tree, is calculated.

14. Local Southern size calling with light smoothing (6): Local Southern is a peak size-calling algorithm used in Terminal Restriction Fragment Length Polymorphism (T-RFLP) analysis, which uses the reciprocal relationship between fragment length and fragment mobility to determine fragment size. This method only uses the four standard points closest to the unknown fragment to determine the size of the unknown fragment. Specifically, a standard curve is generated using the two standard points

below and one standard point above the unknown, and this curve is used to calculate the size of the unknown. A second standard curve is generated, this time by the one standard point above and the two standard points below the unknown and this curve is used to calculate the size of the unknown. The two unknown size values are averaged to determine the length of the unknown fragment. This size calling method is less affected by changes in electrophoresis conditions and as such will normalize data across lanes and gels. Smoothing is a type of data processing parameter that can be applied to raw data prior to analysis. Smoothing helps to reduce the number of false peaks as well as smooth over peaks that could be caused by gel inconsistency, while still allowing for single-base differences to be distinguished. Light smoothing provides the best results for typical data (data without many broad peaks).

15. 2-sample *t*-test (13): The *t*-test for independent samples is a statistical test that evaluates differences between the means of two samples. Specifically, it determines if there is a statistically significant difference between the two means, thus allowing rejection of the null hypothesis that the two data sets came from the same statistical population. A *t*-test makes certain assumptions about the data, such as that the variables are normally distributed and that the variances of the two groups are equal.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{{SD_1}^2}{n_1} - \frac{{SD_2}^2}{n_2}\right)}}$$

References

1. **Bray J. R. & Curtis J. T.** 1957. An ordination of the upland forest communities of Southern Wisconsin. Ecol. Monogr. **27:** 325-349.

2. **Chao, A.** 1984. Nonparametric estimation of the number of classes in a population. Scand. J. Stat. **11:** 265-270.

3. **Chao, A. & Lee, S. M.** 1992. Estimating the number of classes via sample coverage. J. Am. Stat. Assoc**. 87**: 210-217.

4. **Clarke, K. R.** 1993. Non-parametric multivariate analysis of changes in community structure. Aust. J. Ecol. **18:** 117-143.

5. **Clarke, K. R. & Gorley, R. N.** 2001. Primer v5: User Manual/Tutorial. Plymouth UK. pp. 25-66.

6. **Elder, J. K. & Southern, E. M.** 1983. Measurement of DNA length by gel electrophoresis. Part II: Comparison of methods for relating mobility to fragment length, Anal. Biochem. **128:** 227-231.

7. **Gotelli, N. J. & Colwell, R. K.** 2001. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol. Lett. **4:** 379-391.

8. May, R.M. 1975. Patterns of species abundance and diversity. In Ecology and Evolution of Communities.ed. M.L. Cody and J.M. Diamond. Harvard University Press, Cambridge, MA. 81-120.

9. **Mihaescu R., Levy D. & Pachter L.** 2006. Why neighbour-joining works. Algorithmica. **54:** 1-24.

10. **Salemi, M. & Vandamme, A. (eds)** 2003. The phylogenetic handbook: A practical approach to DNA and protein phylogeny. Cambridge University Press, Cambridge, UK.

11. **Shannon, C. E.** 1948. A mathematical theory of communication. Bell Sys. Tech. J. **27**: 379–423.

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

13. **Zimmerman, D. W.** 1997. A note on Interpretation of the paired-samples *t*-test. J. educ. Behav. Stat. **22**: 349-360.