Introducing the starlet sea anemone *Nematostella vectensis* as a model for investigating microbial mediation of health and disease in hexacorals

by

Jia Yi Har

B. Sc. Life Sciences (2005)

National University of Singapore

Submitted to the Department of Civil and Environmental Engineering in Partial Fulfillment of the Requirements for the
Degree of Master of Science in Civil and Environmental Engineering

at the

Massachusetts Institute of Technology

June 2009

Signature of author _____
Department of Civil and Environmental Engineering
May 8, 2009

Certified by _____
Janelle R. Thompson
Gilbert W. Winslow Assistant Professor
Department of Civil and Environmental Engineering
Thesis Supervisor

Accepted by _____
Daniele Veneziano
Chair of the Departmental Committee on Graduate Students

Introducing the starlet sea anemone *Nematostella vectensis* as a model for investigating microbial mediation of health and disease in hexacorals

by

Jia Yi Har

Submitted to the Department of Civil and Environmental Engineering
on May 8, 2009 in Partial Fulfillment of the
Requirements for the Degree of Master of Science in
Civil and Environmental Engineering

ABSTRACT

All animals in their natural state harbor complex communities of microbes including those that are beneficial (symbionts), neutral, or harmful (pathogens). The dynamic interactions between animals and their microbiota often dictate the balance between health and disease. Environmental perturbations may disrupt this delicate balance and allow proliferation of pathogens. A global example of this is the case of reef-building corals (class Anthozoa, subclass Hexacorallia) where rising seawater temperatures and coastal eutrophication have been correlated to increased prevalence of microbial diseases and coral bleaching. Despite such correlations, factors that govern whether animal-associated microbes support health or cause disease are largely unknown. Here I employ an emerging Cnidarian model, the starlet sea anemone *Nematostella vectensis,* to investigate whether hexacorals maintain communities of beneficial microbes. I hypothesize that loss of this beneficial community by environmental perturbations will negatively impact the physiology of the host. As a first step to testing my hypothesis I have characterized the microbial communities associated with apparently healthy *N. vectensis* under natural and laboratory conditions using 16S rRNA gene clone libraries, terminal restriction fragment length polymorphism analysis (T-RFLP), fluorescent in situ hybridization (FISH) and cultivation-based analysis.

N. vectensis microbiomes revealed the presence of a novel lineage of epsilon-Proteobacteria associated with animals collected from salt marshes in Sippewissett (Massachusetts), Clinton (Connecticut) and Mahone Bay (Nova Scotia). In addition, ribotypes of *Pseudomonas pseudoalcaligenes, Endozoicimonas elysicola* and Spirochaetes were found associated with multiple environments. Most of the microbial populations in the *N. vectensis* microbiome were closely related to microbial associates of other hexacorallians. Observation of shared ribotypes among different sampling locations suggests a specific association with *N. vectensis*, consistent with a hypothesized role as symbionts. We also observed significant microbial community shifts in reduced-flora *N. vectensis* raised under sterile laboratory conditions and the persistence of a naturally occurring *P. pseudoalcaligenes* ribotype in the microbiome under different antibiotic perturbations, supporting the hypothesis that this population maintains association with *N. vectensis* across changing environmental conditions. Finally FISH analysis using eubacteria and ribotype-specific probes revealed that microbes form cell aggregates in tight association with animal mesentery tissues, suggesting that host-microbe interactions may be tissue-specific.

Thesis Supervisor: Janelle R. Thompson
Title: Gilbert W. Winslow Assistant Professor,
Department of Civil and Environmental Engineering

Acknowledgements

I would like to thank my thesis advisor Janelle Thompson for the honor of being the first graduate student of Thompson lab. I am grateful to Janelle for constantly challenging me with intellectually stimulating ideas and providing guidance and support over the past two years. Her critical thinking and mentorship will continue to help me grow as a scientist.

This work will not be possible without the support and assistance rendered by numerous colleagues from MIT and beyond. I would like to specially thank Ann Tarrant and Adam Reitzel from Woods Hole Oceanographic Institution for generously providing *N. ventensis* samples from the Great Sippewissett Marsh, Clinton and Mahone Bay, and for sharing expertise on maintaining animals in the lab. I am indebted to post-doctoral fellows Samodha Fernando for preparing fluorescent in situ hybridization samples, providing useful figures in this thesis and sharing practical scientific and graduate school advice; Hector Hernandez for his helpfulness in field work and insightful discussions, and lab mate Kyle Peet for occasional assistance in keeping lab "pet" *Nematostella* alive. The technical support given by MIT laboratories has been invaluable. Special thanks to Tsultrim Palden for assistance in gene sequencing, Daniel Sher for lending his expertise in histology techniques and the Delong lab for allowing the use of their sequencing and imaging facilities, Yan Feng and the Fox lab for T-RFLP assistance, and the members the Polz and Alm labs for occasional use of their equipments.

I am also fortunate to have mentored several motivated undergraduate research students - Michael Meyer, Sara Barnowski and Stephanie Yoon – who have contributed their efforts in this work and asked insightful questions during their stints in Thompson lab. Life in the lab would be different without the company of other Parsonites who have made Parsons lab a conducive and vibrant environment to work in.

Last but not least I am appreciative of my fiancé Wesley Weng for his emotional support and intellectual company, and my family for unconditional support for my endeavors.

Table of Contents

List of Figures

List of Tables

Overview

Complex communities of microbes live in symbiotic relationships with animals ranging from sponges to humans. A globally significant example of such interactions is that between corals and the microorganisms that comprise the coral holobiont including photosynthetic zooxanthellae (*Symbiodinium*), bacteria, archaea, protozoa, fungi and viruses. *Symbiodinium* provide fixed carbon (Fallowski *et al.*, 1984) to their coral hosts while nitrogen-fixing bacteria can fulfill up to 50% of the nitrogen requirements of the hosts (Shashar *et al.*, 1994; Lesser, 2004). Other members of the coral microbiota have been shown to produce antimicrobial peptides which may protect corals from pathogen infections (Riley, Gordon, 1999).

Coral reefs worldwide are facing elevated risks of extinction due to the rapid rise in coral diseases. Several diseases have been linked to the destabilization of coral-microbe symbioses caused by rising ocean temperatures, coastal eutrophication and overfishing (Bellwood *et al.*, 2004). In the case of coral bleaching, increased seawater temperatures leads to the loss of carbon-fixing *Symbiodinium* symbionts. Affected corals suffer from reduced skeletal growth, reproductive activities and capacity to resist invasions of competing species and pathogens (Glynn, 2006), leading to large scale deaths (Rosenberg *et al.*, 2007). In other cases, normally non-pathogenic bacteria become virulent (by upregulating virulence genes) under elevated temperatures, killing mutualistic algal symbionts and eventually the corals (Rosenberg, Ben-

Haim, 2002). It is apparent that the delicate balance between different coral-associated microbial populations holds the key to maintaining host health.

In this study, we will approach this complex problem through answering a series of more tractable questions using an emerging Cnidarian model, the starlet sea anemone *Nematostella vectensis*. Previous molecular phylogenetic studies have found that *N. vectensis* (phylum Cnidaria, class Anthozoa, subclass Hexacorallia, order Actiniaria) is monophyletic with respect to other extant hexacorallians (Martindale *et al.*, 2002). *N. vectensis* and other hexacorals, including reef-building corals, display 6-fold radial symmetry in their body plan and only have the benthic polyp stage in their lifecycle (i.e. they do not undergo metamorphosis to form medusa as do other groups of cnidarians). However, as opposed to many reef-building corals, *N. vectensis* does not maintain an algal symbiont or carry out aragonite precipitation.

In order to investigate the roles microbes play in mediating hexacoral health and disease, we start with the characterization of the *N. vectensis* microbiota composition through culture-independent methods - construction of 16S rRNA clone libraries and terminal restriction fragment length polymorphism (T-RFLP) analysis. Once we identify the key players in the microbiome, we will then go on to establish the specificity of these interactions. This can be inferred from the location of the microbes in the host using fluorescent in situ hybridization (FISH) under the hypothesis that if the microbes are intracellular or tightly associated with host tissue, as compared to living freely in the lumen of the host, the interactions are likely to be specific and are possibly mediated by receptors.

Furthermore, if these microbes are truly beneficial, they should be actively selected for by the host and be stably associated with the same host species regardless of geographical location. We test this hypothesis by comparing microbial communities of *N. vectensis* raised in reduce-flora laboratory conditions with those of animals collected from the Great Sippewissett Marsh (Massachusetts), a salt marsh in Clinton, Connecticut, and Mahone Bay (Nova Scotia). We will focus on the presence and enrichment of any ribotypes shared across these samples, as compared to sediment samples where the animals were collected – this will lend further support to the presence of beneficial microbial populations within the host. Taking this hypothesis further, we will also artificially perturb the environment of reduced-flora animals with different combinations of antibiotics to screen for any stably maintained populations. Lastly, we will use this as an opportunity to evaluate the suitability of *N. vectensis* as a laboratory model for hexacorals. If the relationship between cnidarians and microbes are evolutionarily conserved, *N. vectensis*-associated microbes will be most similar to those associated with other hexacorals or anemones.

Chapter 1

Introduction

1.1 Introducing the phylum Cnidaria

The Animal kingdom is divided into Porifera (sponges) and eumetazoans. The phylum Cnidaria was among the first eumetazoans having evolved a defined body plan with an axis, a nerve net and a tissue layer construction (Bosch, 2003). The early-branching phylum Cnidaria – which has evolved separately from the phylum Bilateria that diverged into the majority of animal taxa we see today - includes Anthozoa (sea anemones, corals), Hydrozoa (Hydra, Hydractinia), Cubozoa (box jellies) and Scyphozoa (jellyfishes), and is characterized by a sac-like body with two epithelial tissue layers lined with a nerve net, a single oral opening with numerous tentacles, and the stinging cells (cnidocytes) that gave rise to its name (Nielsen, 2001).

In general, Cnidarians live in a wide range of habitats containing numerous potential pathogens that are capable of tissue destruction and functional impairment (Bosch, 2008). Lacking physical barriers like cuticles or exoskeletons and highly evolved adaptive immunity defenses, the soft-

body cnidarians employ efficient innate immunity mechanisms to discriminate between self and non-self, defend against pathogens and allow mutualistic symbionts to colonize their epithelia.

Moreover, despite being placed at the base of the evolutionary tree and commonly perceived as primitive, recent studies have shown that Cnidarians are genetically complex (Steele, 2002; Putnam *et al.*, 2007) and have preserved much of the genetic complexity of the common metazoan ancestor that is present in higher Bilaterians but lost in invertebrates like *Drosophila* and *Caenorhabditis elegans*.

1.2 Coral diseases can arise due to holobiont destabilization through environmental perturbations

The survival of hexacorals (class Anthozoa, subclass Hexacorallia) is increasingly threatened due to alarming rise in diseases worldwide caused by climate change, water pollution and overfishing (Bellwood *et al.*, 2004). To date, most diseases are only categorized based on their outward symptoms (e.g. 'white band disease') due to ignorance of their causative agents. While many coral-associated microbes are beneficial to the host, coral health is critically dependent on the delicate balance between microbial populations in the holobiont – the host organism and the diverse assemblage of zooxanthellae (algae), bacteria, archaea, fungi and viruses. These symbiotic microbial populations are found in association with the surface mucus layer, polyp tissue (including the gastrodermal cavity) and the calcium carbonate skeleton (Bourne, Munn, 2005).

Microbial populations in the coral holobiont have been hypothesized to perform a variety of symbiotic functions. In reef-building corals, zooxanthellae are a critical component of the holobiont as they fulfill substantial energy requirements of their hosts by transferring photosynthetically-fixed carbon to the corals (Fallowski *et al.*, 1984). Studies on coral mucus demonstrated that this layer supports a diverse and abundant beneficial bacterial community including nitrogen fixers (Shashar *et al.*, 1994; Chimmetto, 2008) and chitin decomposers (Ducklow *et al.*, 1979). Moreover, porous coral skeletons harbor bacteria that satisfy up to 50% of the total nitrogen requirements of the hosts (Ferrer *et al.*, 1988), while some mutualistic symbionts protect the hosts from pathogen invasions through antibiotics secretion (Riley, Gordon, 1999).

The beneficial coral-microbe symbiosis can be destabilized rapidly by environmental stressors, and this often leads to coral diseases. Coral bleaching - where the coral loses its zooxanthellae (*Symbiodinium*) and its color - is the most threatening coral disease globally (Rosenberg *et al.*, 2007). Bleaching episodes generally coincide with the hottest period of the year and are most severe at times when seawater temperatures are warmer than normal (Hoegh-Guldberg, 1999) as high temperatures have been reported to stress the host and increase their susceptibility to infection (Lafferty, Holt, 2003).

Moreover, normally non-pathogenic bacteria become virulent under elevated temperatures. In the bleaching of Mediterranean Sea coral *O. patagonica*, *Vibrio shiloi* expresses a 11-mer peptide Toxin P (PYPVYAPPPWP) that inhibits photosynthesis of the *Symbiodinium* and leads to algae and pigment loss under raised temperatures (Rosenberg, Ben-Haim, 2002). If the algal symbiont

is not re-established in the coral within a certain period, the host is likely to die since it requires substantial fixed carbon from the symbiont. Therefore, microbial partners play critical roles in host survival.

1.2.1 Coral Probiotic Hypothesis: Microbial community shift may allow adaption to external challenges

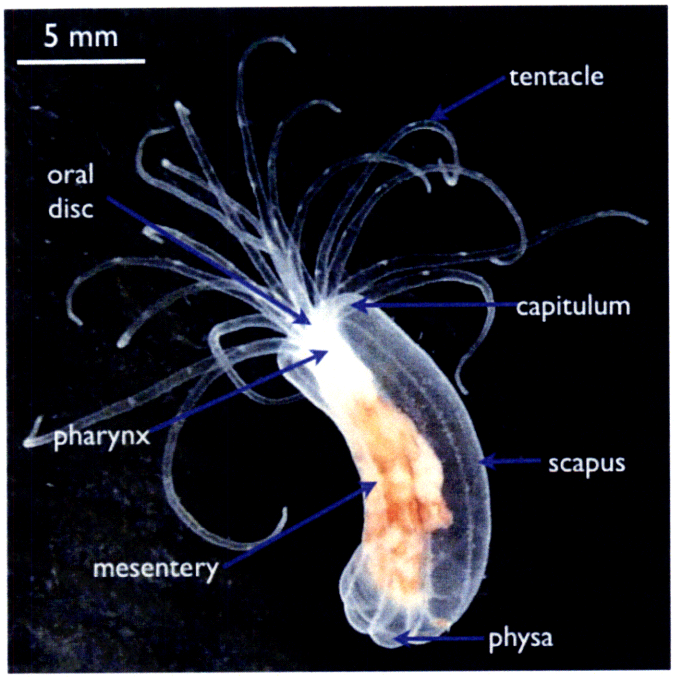
Unlike higher organisms that have adaptive immunity capabilities, hexacorals and other cnidarians do not produce new antibodies and are not expected to develop immunity towards new pathogens. Therefore, it was surprising that *O. patagonica* previously susceptible to *V. shiloi* was able to recover from infections from artificially-introduced *V. shiloi*. The resistant corals seemed to be able to lyse the intracellular *V. shiloi* via unknown mechanisms and avoid the disease (Reshef *et al.*, 2006). The ability of coral to adapt to pathogen infections has led to the proposal of the Coral Probiotic Hypothesis (Reshef *et al.*, 2006). It was suggested that natural pathogen exposure caused a shift in the coral microbial community and at least one of the new bacteria populations was able to secrete antibacterial products to clear the infection. However this remains an open question as direct evidence of microbiota community shifts were not provided. Alternative explanations of infection recovery include the loss of *V. shiloi* virulence due to spontaneous mutations and changes in host defense expression regulation pathways that allowed them to overcome pathogen infections. Corals are difficult to work with in laboratory environments due to challenges associated with their collection, slow growth rates, high-maintenance growth requirements. Therefore, tractable laboratory model systems for corals need to be developed to examine and test hypotheses such as the Coral Probiotic Hypothesis, and to investigate the factors that trigger shifts in microbiome structure and host immune responses.

1.3 Experimental model for studying hexacoral microbiomes: *Nematostella vectensis*

Cnidarian model systems – including *Acropora* (Miller, Ball, 2000), *Eleutheria* (Kuhn *et al.*, 1996), *Hydra* (Steele, 2002), *Hydractinia* (Frank *et al.*, 2001) and *Podocoryne* (Schmid, 1998) - have been previously used to answer developmental evolution questions including pattern formation (Bode, Bode, 1980), self-organization (Technau *et al.*, 2000), coloniality and alloimmunity (Frank *et al.*, 2001). Among Anthozoans, the starlet sea anemone *N. vectensis* is an emerging model system with substantial background research (74 journal articles on *Nematostella* were found from 1920 to 2007) and is continuing to gain attention from the research community as its genome sequences is publicly available (Darling *et al.*, 2005).

Nematostella vectensis, first described as a new species from the southern coast of England by Stephenson in 1935, is a small, burrowing estuarine sea anemone widely distributed in salt marshes along the Atlantic coast of North America from Nova Scotia to Georgia, the Pacific coast of North America from Washington to central California and the southeast coast of England (Hand, Uhlinger, 1994). Salt marshes are challenging habitats due to the large fluctuations in environmental factors including temperature, salinity, ultraviolet light intensity and concentrations of sulfide, oxygen and reactive oxygen species at different seasons of the year. *N. vectensis* is a euryhaline, eurythermal species that has been recorded to survive in salinities of 2 to 52 parts per thousand and temperatures of 1.5 to 32.5 °C and is one of the few animal generalists that can thrive in these challenging environments in the long term (Darling *et al.*, 2004).

The *N. vectensis* adult is a sedentary polyp (Fig. 1) that feeds using its tentacles. Food enters the gut through the pharynx and is digested externally in the gastrointestinal cavity (coelenterons), which is lined by mesenteries that increase surface area for food digestion. The longitudinal retractor muscles of the mesenteries enable the polyp to retract. Along the free edge of each mesentery are cilia which may circulate fluid that otherwise might stagnate in the compartments defined by the mesenteries (Fautin, Mariscal, 1991). It mainly preys on small free-living organisms occur in salt marshes, including copepods, midge larvae, worms (nematodes, polychaetes, and oligochaetes) and rotifers (Frank, Bleakney, 1978; personal observation).



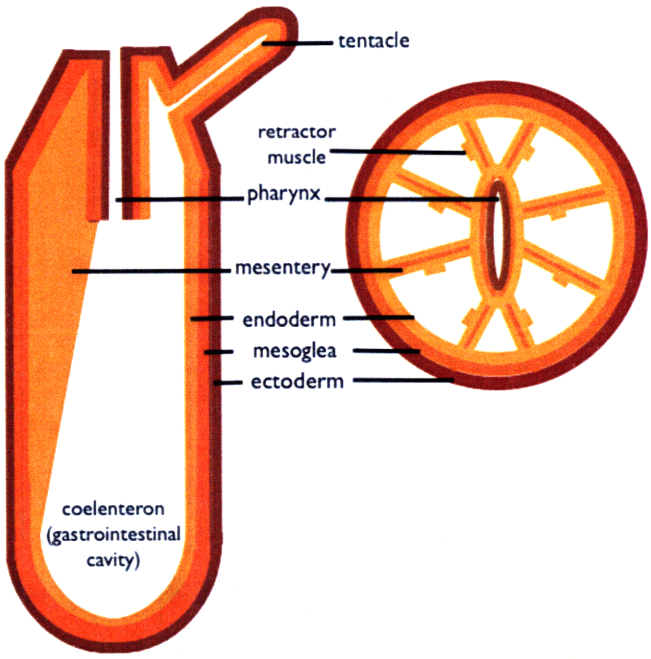


Fig. 1. *N. vectensis* polyp morphology. Top: Labeled photograph of adult animal showing location of body parts. Photo credit: Michael J. Meyer. Bottom: Longitudinal section diagram of *N. vectensis* depicting ectoderm, mesoglea and endoderm tissue layer organization and internal muscles. Adapted from Finnerty *et al.,* 2004.

1.3.1 Genome databases and molecular tools for *N. vectensis* are available

One compelling reason to use *Nematostella vectensis* as a model is the completion of its genome - a first for cnidarians. The earliest gene sequences from *N. vectensis* were published in 1997 (Finnerty, Martindale, 1997) while the first molecular analyses of its developmental lifecycle were published in 2003 (Scholz, Technau, 2003). More recently, the Joint Genome Institute produced the draft genome sequence through random shotgun strategy. The starlet sea anemone genome has been found to be complex and contains gene repertoire, exon-intron structures and large-scale gene linkage patterns more similar to vertebrates than other invertebrates, implying the presence of a complex genome within the eumetazoan ancestor. While only 95% of the known protein coding content is currently captured, the total genome size is predicted to be about 450 Mb (Putnam *et al.*, 2007). Among 18000 protein-coding genes, about 8000 were also present in bilaterians, suggesting that these are ancestral genes (Putnam *et al.*, 2007).

StellaBase (Sullivan *et al.*, 2006; Sullivan *et al.*, 2008) allows querying of the assembled *N. vectensis* genome using both keyword and homology-based search functions and offers a confirmed gene library and a predicted genome. The availability of *N. vectensis* genome and genetic tools such as BAC, EST, cDNA libraries and tested protocols for functional analysis of gene expression (Darling *et al.*, 2005) will continue to facilitate biological and medical research using *N. vectensis*.

1.3.2 *N. vectensis* is relatively easy to maintain under laboratory conditions

N. vectensis individuals can be collected from their land-accessible estuarine habitats and are readily cultured in the laboratory. Adult anemones grow well even at high density in unfiltered,

non-circulating artificial seawater (about 10 parts per thousand) at room temperature. Maintenance includes weekly water changes and feedings with *Artemia* (brine shrimp) nauplii available from pet suppliers (Hand, Uhlinger, 1994).

1.3.3 *N. vectensis* can be artificially induced to reproduce sexually or asexually

N. vectensis individuals are capable of sexual and asexual reproduction and can go through complete developmental cycles – from embryos to adults - under laboratory conditions within 2-3 months. As they have high capability of regeneration after injury, periodic splitting of individuals has been exploited as a way to rapidly produce clonal lines to eliminate experiment variability due to host genotype and to increase population size in the laboratory. On the other hand, when population genetic variation is desired, both sexes can be coaxed to reproduce sexually at predictable frequencies through feeding of mussel tissue and subjecting the animals to fluctuating temperatures (Tarrant *et al.*, 2009).

1.3.4 *N. vectensis* has undisturbed and accessible habitats

Long-term ecological studies are possible with *N. vectensis* due to the relatively undisturbed habitats and ease of *in situ* monitoring (Darling *et al.* 2005). The high accessibility of animals allows continued tracking of host population changes and subsequent microbial community shifts under different temperatures, salinity, pH, oxygen levels over long period. The ease of re-sampling animals from the same location is especially important for molecular and evolutionary ecological studies where stable symbiotic relationships between host and microbes are of interest.

Taken together, *N. vectensis* fulfills multiple criteria of an excellent laboratory model and is especially suitable for answering questions in microbiome mediations of hexacoral health and disease.

1.4 Innate immunity strategies of *Nematostella vectensis* and other cnidarians

Despite the lack of highly evolved adaptive immunity strategies and physical barriers like cuticles or exoskeletons and defenses, cnidarians employ multiple efficient innate immunity mechanisms to allow colonization of mutualistic symbionts and protect against pathogens.

1.4.1 Cnidarians eliminate pathogens via TLR pathways

In general, cnidarians uses an ancient pathogen defense system based on receptor-mediated recognition of pathogen surface components that are not present in the host. These pathogen-associated molecular patterns (PAMPs) are targeted by the host Toll-like receptors (TLR), leucine-rich repeat (LRR) motifs or Toll/IL-1 receptor (TIR) domains. Ligand binding initiates downstream signal transduction networks that activate concerted defense response (Bosch, 2008).

HMM-based search methods revealed 37 innate immunity genes in *N. vectensis* that code for components in TLR, interferon (IFN), evolutionarily-conserved signaling intermediate in Toll (ECSIT) and complement pathways. These include a Toll/TLR family MyD88 homolog (NvMyD88) and a protein (NvTLR-1) with both carboxy- and amino-terminal-flanking cysteine-

rich motifs in the extracellular part of the protein, suggesting that *N. vectensis* TLRs are closely related the ancestral domain structure (Miller *et al.*, 2007).

1.4.2 Some cnidarians secrete antimicrobial peptides (AMP) against pathogens

Despite the lack of migratory phagocytic cells and a conventional TLR and membrane-attack complex/perforin (MAC/PF) system, *Hydra magnipapillate* is protected from pathogens through expressing and secreting antimicrobial peptides from their epithelium. Transcriptome and biochemical analysis have shown that Hydramycin-1, a cationic 60-aa peptide, is expressed exclusively in the endodermal epithelium and has high antimicrobial activity against *Bacillus megaterium*. It is interesting that the expression of host defense genes is affected by nerve cells, suggesting that the close communication between nervous and immune systems are essential for the survival of these ancient organisms (Kasahara, Bosch, 2003).

The Cnidarian *Aurelia aurita* (i.e. a "jellyfish") produces the antimicrobial peptide aurelin (Ovchinnikova *et al.*, 2006) that acts against Gram-positive (*Listeria monocytogenes*) and Gram-negative (*Escherichia coli*) bacteria. While it is still unclear how aurelin is induced, the alignment of aurelin with sea anemone toxin sequences suggested that it could be functionally related to the ShK-like toxin (Ovchinnikova *et al.*, 2006), which resembles a potassium channel-blocking protein (ShK) produced by the sea anemone *Stichodactyla helianthus* (Castaneda *et al.*, 1995).

While there is no literature to date that reports the production of similar antimicrobial peptides by *N. vectensis*, TBLASTN-based searches of the *N. vectensis* genome against the Genbank

database identified a gene matching strongly to the human macrophage expressed protein 1 (MPEG1) (Miller *et al.*, 2007), indicating that it has components of the MAC/PF pathway responsible for cytotoxicity and generations of trans-membrane pores in target cells (Shida *et al.*, 2003). Moreover, in a neurotoxin screening study in *N. vectensis*, the only candidate toxin found was a conserved domain resembling a ShK toxin co-coded by four genes (Moran, Gurevitz, 2006). This hints that the species has a very limited arsenal of neurotoxins, possibly due to the low number of prey in its specific habitat. While it is unclear whether *N. vectensis* has a similarly low repertoire of AMPs, querying *Hydra* antimicrobial peptides (Bosch *et al.*, 2009) against StellaBase (Sullivan *et al.*, 2006) did not yield any matches in the *N. vectensis* genome, suggesting the species does not produce peptides with similar sequences as those in Hydra.

1.4.3 Symbiosis initiation can occur via horizontal or vertical transmission

The difference between microbial composition of coral-associated and bacterioplankton populations suggests that host selection pressure play a role in determining the structure of the holobiont (Rosenberg *et al.*, 2007). Moreover, bacterial communities associated with the same coral species found in geographically separated locations are structurally similar, whereas communities found on different coral species share less resemblance (reviewed in Rosenberg *et al.*, 2007). These mutualistic microbial communities may be acquired through horizontal or vertical transmission.

Vertical transmission of symbionts has been demonstrated for the Cnidarian *Hydra viridis* which forms stable symbiotic relationship with the intracellular green algae Chlorella in their endodermal epithelial cells through vertical transmission (Muscatine, Lenhoff, 1963). The

Chlorella symbiont is transferred to the next generation through translocation into the oocyte, ensuring the presence of the symbiont in newly formed embryos. Screening for symbiosis-related genes revealed that a gene encoding an ascorbate peroxidase is expressed only during oogenesis, and that five other genes with significantly higher expression levels in the presence of Chlorella are vital for oocyte development (Habetha, Bosch, 2005). In fact, *H. viridis* suffers from ovary deformities without these symbionts, indicating that microbial partners can play a role in development and form tight association with their hosts via highly specific recognition processes. This concurs with an early study with bacteria-free cultures of *Hydra* where the organism could not propagate without symbionts, demonstrating that the associated microbial community may have a role in development (Rahat, Dimentman, 1982). The requirement of certain microbes for development would create a very strong selection pressure to evolve mechanisms for vertical transmission of symbionts. In addition, recent studies on freshwater *Hydra* have revealed different core microbial communities associated with different host species (Fraune, Bosch, 2007), leading to the hypothesis that these communities are forced by differences in the innate immune repertoires of the two different host species.

On the other hand, symbioses between photosynthetic zooxanthellae (*Symbiodinium*) and corals occur through horizontal transmission. Most coral species obtain new symbionts from the environment at the start of each generation through complex recognition and specificity processes. Interestingly, in the symbiosis between the scyphozoan *Cassiopea* and *Symbiodinium*, the shift from vertical symbiont acquisition to horizontal acquisition caused a change from mutualism to parasitism that is reflected in host fitness reduction (Sachs, Wilcox, 2006).

Such observations motivate hypotheses regarding the role of microbial symbionts in cnidarian development. How *N. vectensis* acquires mutualistic partners (e.g. by vertical or horizontal transmission) and whether microbes play a role in its lifecycle (e.g. influencing health or development) is currently unknown and will form the basis of future work.

1.4.4 Cnidarians deploy amoebocytes to facilitate phagocytosis of foreign particles and wound healing

Another Cnidarian defense strategy is the deployment of phagocytic amoebocytes. Gorgonian and scleractinian corals recruit amoebocytes to sites of tissue injury to facilitate wound healing and tissue reorganization (Olano, Bigger, 2000). The sea anemone *Actinia equina* possesses amoebocytes in their mesenteric filaments that are capable of killing the gram-negative bacterium *Psychrobacter immobilis*. Stimulation with pathogenic lipopolysaccharide shows that the amoebocytes produce reactive oxygen species and other soluble bacteriocidal factors to degrade the endotoxin (Hutton, Smith, 1996).

Despite the remarkable ability of *N. vectensis* to recover from wounds and regenerate whole body parts after splitting, there is currently no report of amoebocytes in the species in the literature, or based on personal observation of histological preparations.

1.5 Testing hypotheses on *N. vectensis* holobiont community structure

Based upon the analysis presented in this introduction, we believe that *N. vectensis* will be useful as a model for unraveling the molecular nature of hexacoral-microbe interactions and the dynamics of microbial community assembly. To characterize *N. vectensis* as a model system we have undertaken a study of the *N. vectensis* microbiome and potentially symbiotic populations. We hypothesize that if populations are beneficial, they will be selected for by the host and are likely to associate stably with the same host species across different geographical locations. Also, if host-microbe relationships are evolutionary conserved, we would expect *N. vectensis* symbionts to share high sequence similarity to symbionts in other hexacorallians. To investigate these hypotheses, we screened animals collected from three geographically-separated salt marshes – Sippewissett (Massachusetts), Clinton (Connecticut) and Mahone Bay (Nova Scotia) - and sediment sampled at Sippewissett where the animals were collected, and reduced-flora lab-raised animals perturbed with different antibiotics for microbial 16S rRNA. Since only 0.01 to 0.1% of bacterial cells form colonies by standard plating techniques (Kogure *et al.*, 1979), we employed culture-independent methods like clone library construction and terminal restriction fragment length polymorphism (T-RFLP) to sample the microbiome. To evaluate the specificity of host-microbe interactions, we visualized the exact location of symbionts in the host using fluorescent in-situ hybridization (FISH) methods with general eubacteria and group-specific fluorescent-labeled probes.

Chapter 2

Methods and Materials

2.1 Animal Culture and Sample Collection

Lab-raised *Nematostella vectensis* individuals were originally obtained from the laboratory of Ann Tarrant at Woods Hole Oceanographic Institution (WHOI) and maintained in the current lab for six months before DNA extraction. *N. vectensis* were kept in 10 parts per thousand artificial seawater (Instant Ocean, Spectrum Brands, Inc.) at room temperatures (21 to 25 °C) and fed *Artemia* nauplii thrice a week over six months. To prepare "reduced-flora" animals, they were transferred to autoclaved saline and given sterilized *Artemia* (pretreated in 10% bleach) for 4 weeks to reduce the effects of lab microbial contaminants on host microbiome. Prior to genomic DNA extraction, lab animals were cultured in autoclaved 10 parts per thousand Instant Ocean for 2 days without feeding to eliminate digested food particles and rinsed thrice in deionized water to remove loosely attached microbes and debris.

Field animals were collected from Sippewissett Marsh (MA), Clinton (CT) and Mahone Bay (Nova Scotia) between July 2008 to October 2008 by Adam Reitzel, preserved in RNAlater (Ambion, Inc.) and stored at 4 °C until DNA analysis. Before genomic DNA extraction, field animals were directly removed from RNAlater and rinsed thrice in deionized water. Sediment

samples from the site of *N. vectensis* collection were retrieved from Sippewissett Marsh in November 2008 (Table 1). The five samples analyzed in this study were: Reduced-flora animals (LAB), Sippewissett, Massachusetts animals (MA), Clinton, Connecticut animals (CT), Mahone Bay, Nova Scotia animals (MB) and Sippewissett sediment (SED). DNA extractions were done separately at different times.

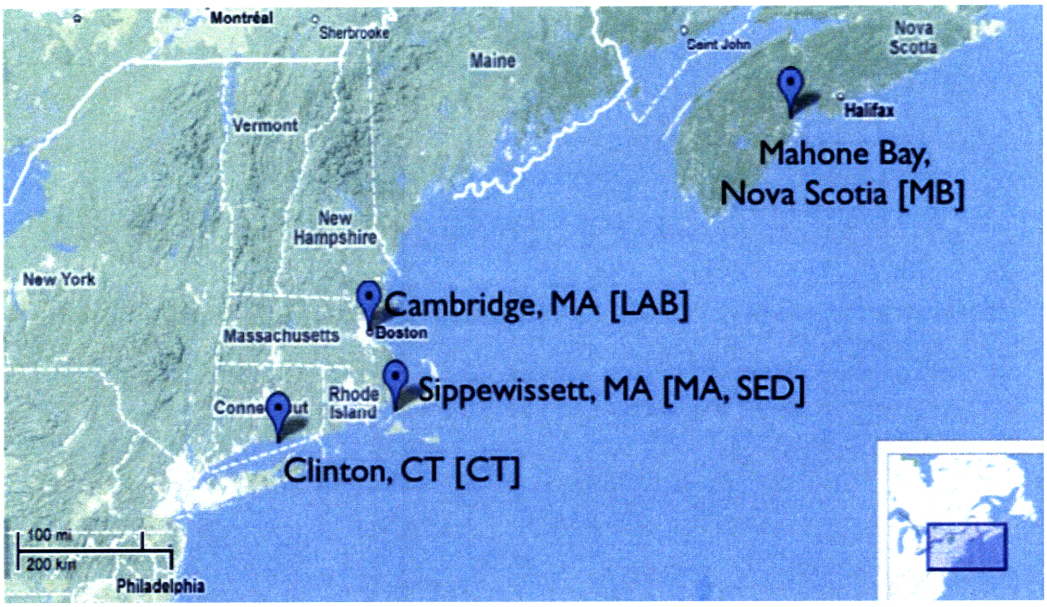


Fig. 2. Map of locations of sample collection. *N. vectensis* individuals were collected from Sippewissett, Clinton and Mahone Bay, while sediment samples were obtained from Sippewissett. Lab animals were raised in Cambridge. Clone libraries corresponding to sites are given in square brackets.

Table 1. Genomic DNA extraction time frame for LAB, MA, CT, MB and SED.

Sample	Collection location (date)	Air temperature (°C)	Water Temperature (°C)	Preservation of tissue sample in formaldehyde (for FISH)	DNA extraction (for clone libraries)
LAB	MIT Laboratory (varies)	21 to 25	21 to 25	varies	February 2008
MA	Sippewissett Marsh, MA (July, 2008)	15 to 37	19 to 43	-	July 2008
MA-II	Sippewissett Marsh, MA (Nov. 2008)	-7 to 23	6	*In situ* November 2008	November 2008
CT	Cliton, CT (July, 2008)	17 to 31	ND	-	July 2008
MB	Mahone Bay, NS (Sept, 2008)	1 to19	ND	-	October 2008
SED	Sippewissett Marsh, MA (Nov. 2008)	-7 to 23	6	-	November 2008

All *N. vectensis* individuals for maintenance in the lab were inundated in marsh water during collection and kept in artificial seawater in the laboratory. Individuals were preserved in the field using RNAlater (Qiagen) for DNA extraction and 10% formaldehyde for FISH. ND: Not determined.

2.2 Molecular analysis

2.2.1 DNA extraction

Genomic DNA was extracted from whole animals (n=3 to 5 per extraction depending on animal size) or sediment using the DNeasy® Blood and Tissue kit (Qiagen Sciences) and UltraClean™

Soil DNA Isolation Kit (Mo Bio Laboratories, Inc.), respectively, according to manufacturer instructions.

2.2.2 Terminal Restriction Fragment Length Polymorphism (T-RFLP)

For T-RELP analysis, FAM-labeled-27F (5'-FAM-AGA GTT TGA TCM TGG CTC AG-3') and 805R (5'-GGA CTA CCA GGG TAT CTA ATC CC-3') were used to amplify positions 8-805 of the *Escherichia coli* 16S rRNA. PCR conditions were as follows: 1 cycle of 94°C for 3 mins; 35 cycles of 94°C for 45 secs, 52°C for 45 secs, 72°C for 1 min 30 secs; 1 cycle of 72°C for 5 mins, using Taq polymerase (New England Biolabs, NEB). The 800bp bands were cut, gel-purified and subjected to restriction digests using HaeIII and HpaII (NEB). Samples were separated on an ABI 310 Genetic Analyzer (Applied Biosystems, ABI) along with ROX 500 size standard (ABI). Peaks with more than 90 intensity units were called using GeneScan version 3.1 (ABI).

2.2.3 PCR screening for Archaea

Archaeal-specific primers Arch20F (5'-TC CGG TTG ATC CYG CCR G-3') and Arch958R (5'-YCC GGC GTT GAM TCC AAT T-3') (Orphan *et al.,* 2000) were used to screen for the presence of Archaea symbionts in LAB, MA and CT animals.

2.3 Clone library construction

Blunt end 16S rRNA products were amplified using 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 805R (5'-GGA CTA CCA GGG TAT CTA ATC CC-3') primers from all five samples using Phusion (Finnzymes) under the following conditions: 1 cycle of 98°C for 30 secs,

35 cycles of 98°C for 10 secs, 52°C for 30s, 72°C for 30s and 1 cycle of 72°C for 10 mins. Quadruplicate PCR were carried out to reduce domination of PCR-generated mutations. PCR products were pooled, gel-purified, cloned into pCR-Blunt vector (Invitrogen) and transformed into chemically competent TOP10 cells (Invitrogen) according to manufacturer instructions. About 90 clones were selected for each library and the cloned sequences were amplified with M13F (5'TGT AAA ACG ACG GCC AGT) and M13R (5'-AGG AAA CAG CTA TGA CCA T-3') primers. PCR products were sequenced from the 27F primer using BigDye® Terminator v3.1 Cycle Sequencing Kit according to manufacturer recommendations on a 3130 Genetic Analyzer (ABI). Sequences were trimmed using Sequencher 4.5 (Gene Codes Corporation) for downstream analyses.

2.4 Sequence analysis

Sequences were edited and grouped into 99% contigs in Sequencher 4.5, and clones were identified by querying the Genbank database using NCBI BLAST (Altschul *et al.*, 1997). Representative clones were aligned using MUSCLE (MUltiple Sequence Comparison by Log-Expectation, (Edgar, 2004) and phylogenies were inferred using the maximum likelihood program PhyML (Guindon *et al.*, 2005). The diversity of all five microbial communities were compared using UniFrac (Lozupone *et al.*, 2006). The maximum number of bacterial phylotypes in each sample was assessed using the Chao 1 non-parametric richness estimator in EstimateS 8.0 (Colwell, Coddington, 1994) and their respective rarefraction curves (Colwell *et al.*, 2004) were plotted based on the calculated Cole Rarefraction values from EstimateS.

2.5 Fluorescent In-situ Hybridization (FISH)

2.5.1 Sample dehydration and sectioning

Animal samples were fixed in 4% paraformaldehyde for 10-16 hours, rinsed twice with PBS for 30 mins and incubated for 1 hour each in 30%, 40%, 50%, 60% and subsequently 70% ethanol. Samples were then stored in fresh 70% ethanol. Animals were embedded in paraffin for sectioning into 4 μm slices and mounted on slides by the MIT Histology facility.

2.5.2 FISH probes

FISH probes were designed using the primer design software Primrose (Ashelford *et al.*, 2002). The principle of probe design is to pick an oligonucleotide sequence completely specific to the target sequences but has at least one mismatch to all other non-target sequences, preferably in the central region to maximize the destabilizing effect of the mismatch (Stahl, Amann, 1991). Probe specificities were confirmed by querying probe sequences against Genbank using the BLASTN program and obtaining the returned max scores (Altschul *et al.*, 1997).

Table 2. Sequences of probes used in FISH analysis.

Probe specificity	Probe sequences
General eubacteria (EUB338)	Cy3-GCTGCCTCCCGTAGGAGT
Pseudomonas	FAM-TCCAGAAGTAGCTAGTCTAACCTT
Epsilon-Proteobacteria clone	FAM-TTAACCATAGAACTGCAAGACAAA

2.5.3 In situ hybridization

Tissue sections of *N. vectensis* were incubated in xylene twice for 5 mins and washed twice with 100 % ethanol to remove paraffin. Rehydration was done by incubation in 100%, 80%, 60%,

40% ethanol, deionized water and finally in PBS for 5 mins each. Proteins were removed by digestion with Proteinase K (10 μg/ml) at 37C for 10 mins and lysozyme (100 μg/ml) at 37 °C for 10 mins. Negative controls using antisense-EUB338 has been performed to optimize hybridization conditions and eliminate false negatives. For probe hybridization the slides were pre-incubated in 20μl hybridization buffer for 30 min (900 mM NaCl, 20 mM Tris/HCl, 30% formamide, and 0.01% SDS). The probes were applied at a concentration of 15 ng/μl and used for hybridization at 60°C for 3 hrs. For probe hybridization, the slides were washed in buffer and rinsed in distilled water. The air-dried fluorescently labeled slides were visualized in dark field with a Zeiss Axioskop 2 (Carl Zeiss MicroImaging Inc.). Bacteria cells labeled positive with Cy3 and FAM were enumerated at 1000 x magnification.

2.6 Isolation of microbial strains from reduced–flora and antibiotic-treated incubations

Reduced-flora animals (n=2) were hand-crushed using an autoclaved tissue homogenizer and with 1ml of sterile saline. 50μl of homogenate was plated on separate LB agar plates with (i) 10 μg/ml chloramphenicol (Cm), (ii) 100 μg/ml kanamycin (Kn), (iii) 100 μg/ml streptomycin (Sm), (iv) 4 μg/ml nalidixic acid (NA), (v) combination of above antibiotics and (vi) no antibiotic to isolate strains and to examine antibiotic resistant bacteria. Colonies that grew on these antibiotic plates were categorized by resistance profile, colony-purified and archived in 25% glycerol and -80 °C. Their 16S rRNA were amplified as in section 2.2 using primers 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 805R (5'-GGA CTA CCA GGG TAT CTA ATC CC-3') and sequenced directly. The isolates were identified by BLAST searches against the Genbank database.

Chapter 3

Results

3.1 Clone library analyses

3.1.1 Overall features of the sequences

A total of 117, 116, 107, 88 and 87 clones were sequenced for MA, CT, MB, SED and LAB 16S rRNA libraries, respectively. For analysis purposes, identified clones were categorized at the phylum level, except for Proteobacteria, which were grouped at the class level. For the Sippewissett sediment library (SED), approximately 42% were from the phylum Cyanobacteria/chloroplasts and 20% from Bacteroidetes, followed by 8% each for delta- and gamma-Proteobacteria, Planctomycetes, and some Chloroflexi (6%), Spirochaetes (2%) alpha- (2%) and beta-Proteobacteria (1%) (Fig. 3-SED). The phylogenetic relationships between all 99% operational taxonomic units (OTUs) found in this study are shown in Figs. 4 (Proteobacteria) and 5 (non-Proteobacteria). An OTU was defined as a group of sequences sharing > 99% nucleotide identity.

The Sippewissett *N. vectensis* (MA) clone library were devoid of alpha- and delta-Proteobacteria and contained mainly Cyanobacteria/chloroplasts (39%), epsilon- (33%), gamma-Proteobacteria

(15%), Bacteroidetes (3%), Planctomycetes (3%), Verrucomicrobia (3%) and about 1% each of beta-Proteobacteria, Chloroflexi and Spirochaetes (Fig. 3-MA). Clone libraries of field animals collected from Clinton, Connecticut (CT) and Mahone Bay, Nova Scotia (MB) were clearly dominated by an epsilon-Proteobacteria ribotype that made up 95% and 98% of the libraries, respectively. The CT library also contained three gamma-Proteobacteria, one Spirochaete and one Verrucomicrobia clones (Fig. 3-CT). The rest of the clones in MB were one gamma-Proteobacteria (98% similarity to *Pseudomonas pseudoalcaligenes*) and one Bacteroidetes (Flavobacteria) (Fig. 3-MB).

Reduced-flora animals (LAB) had a more diverse microbiome than field animals: gamma-Proteobacteria dominated the clone library (75%) with representatives from genera *Thalassomonas, Alteromonas, Pseudoalteromonas* and *Pseudomonas*, followed by Spirochaetes (12%), alpha-Proteobacteria (7%), Bacteroidetes (5%) and a few clones from Planctomycetes (2%), delta-Proteobacteria (1%) and Cyanobacteria/chloroplasts (1%) (Fig. 3-LAB).

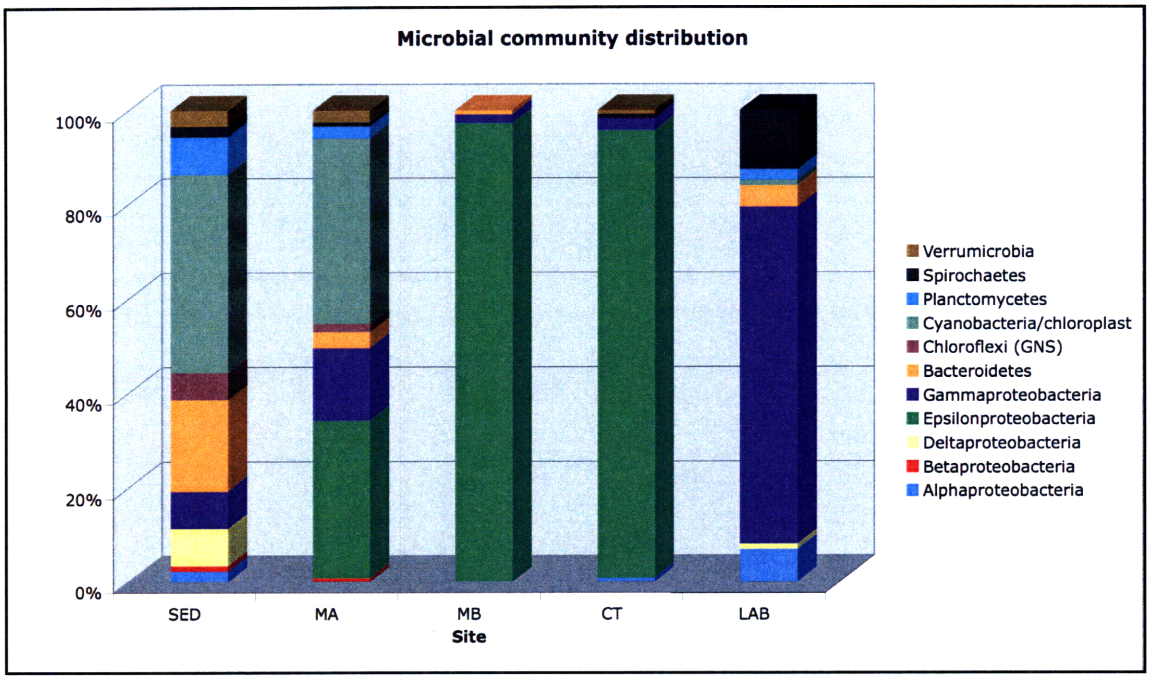


Fig. 3. Microbial community structure for each sample grouped at the phylum level, except for Proteobacteria, which were categorized at the class level.

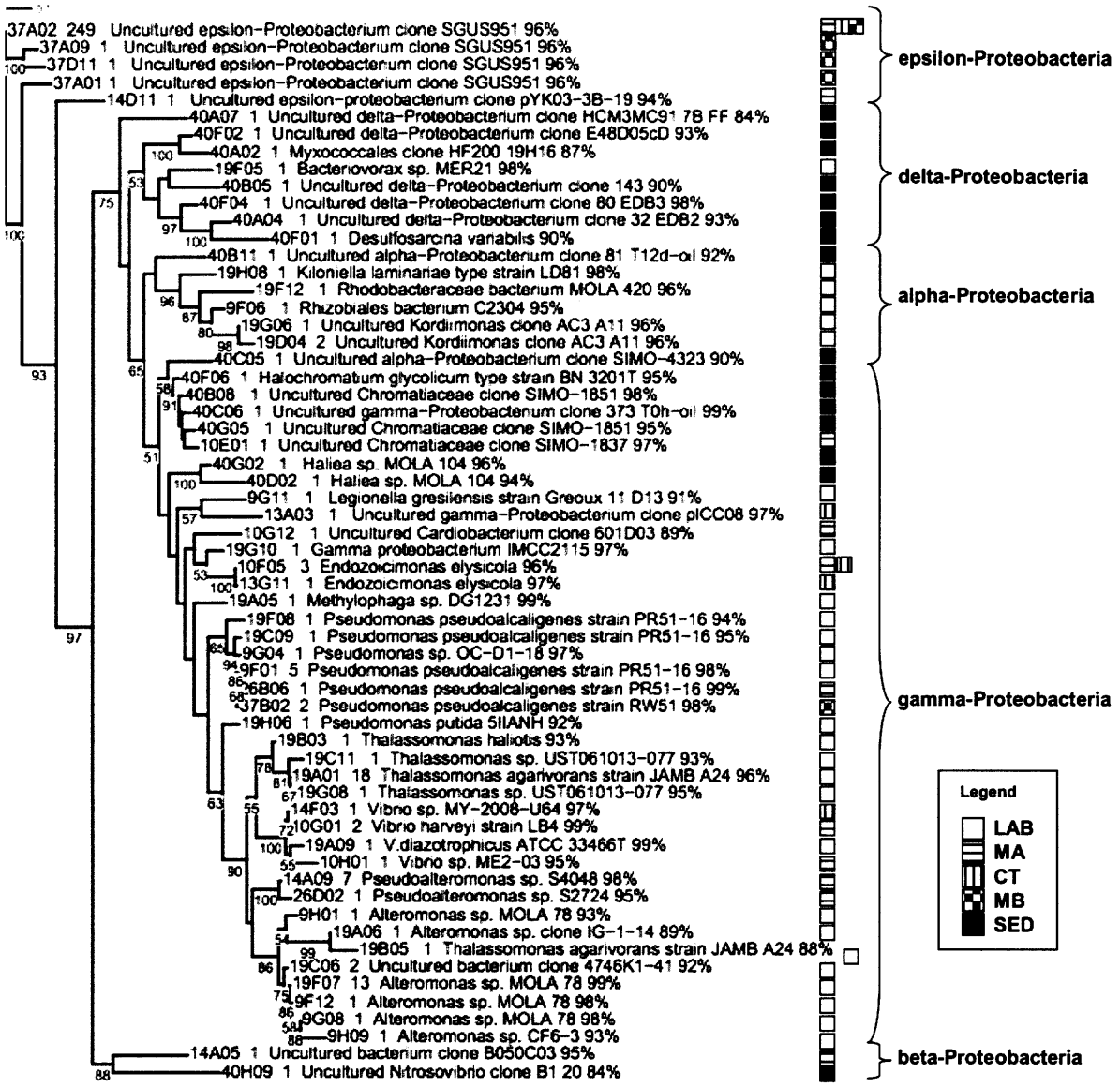


Fig. 4. Maximum likelihood tree of all Proteobacteria clones based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

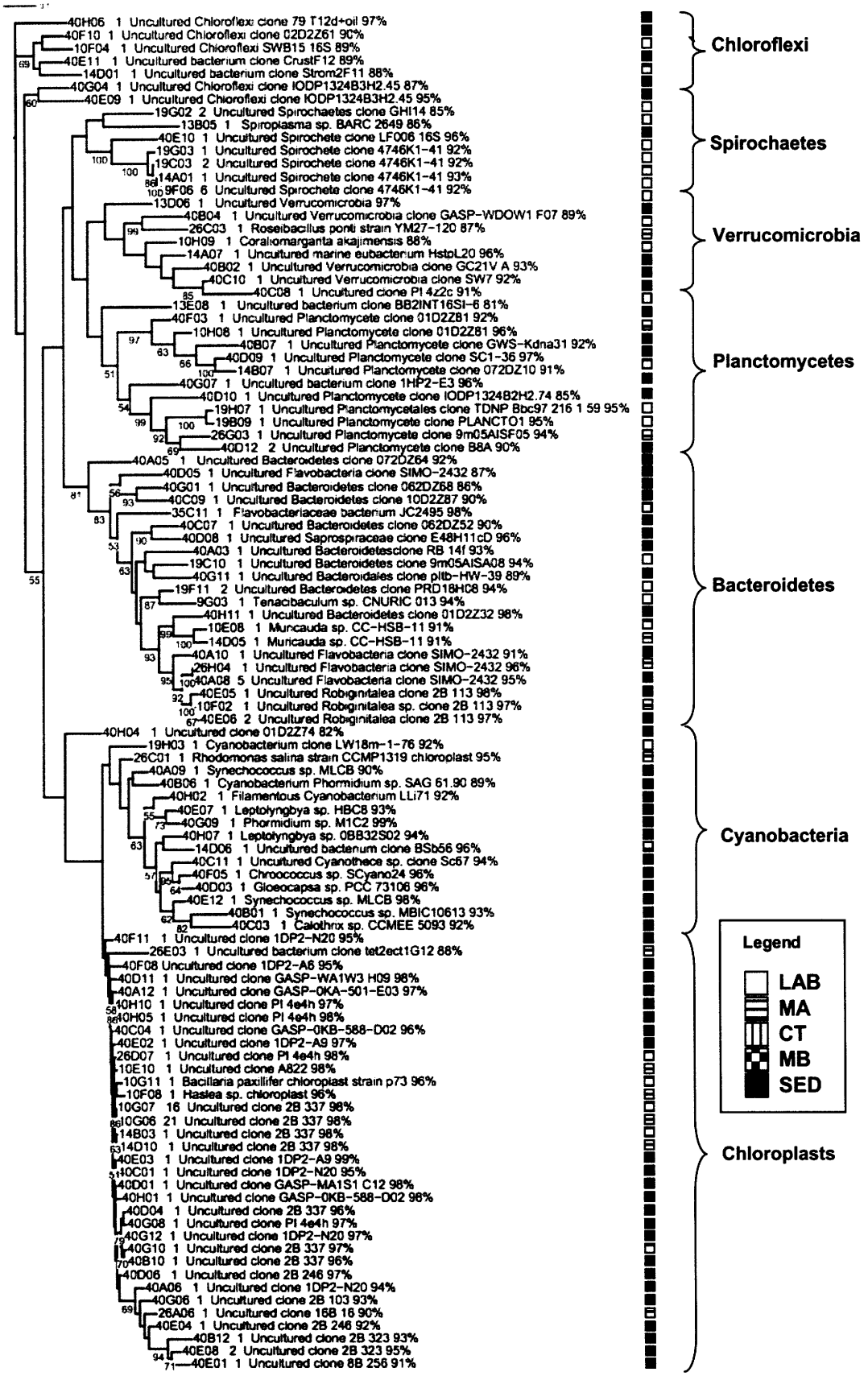

40H06 1 Uncultured Chloroflexi clone 79 T12d+oil 97%
40F10 1 Uncultured Chloroflexi clone 02D2Z61 90%
10F04 1 Uncultured Chloroflexi SWB15 16S 89%
40E11 1 Uncultured bacterium clone CrustF12 89%
14D01 1 Uncultured bacterium clone Strom2F11 88%
40G04 1 Uncultured Chloroflexi clone IODP1324B3H2.45 87%
40E09 1 Uncultured Chloroflexi clone IODP1324B3H2.45 95%
19G02 2 Uncultured Spirochaetes clone GHI14 85%
13B05 1 Spiroplasma sp. BARC 2649 86%
40E10 1 Uncultured Spirochete clone LF006 16S 96%
19G03 1 Uncultured Spirochete clone 4746K1-41 92%
19C03 2 Uncultured Spirochete clone 4746K1-41 92%
14A01 1 Uncultured Spirochete clone 4746K1-41 93%
9F06 6 Uncultured Spirochete clone 4746K1-41 92%
13D06 1 Uncultured Verrucomicrobia 97%
40B04 1 Uncultured Verrucomicrobia clone GASP-WDOW1 F07 89%
26C03 1 Roseibacillus ponti strain YM27-120 87%
10H09 1 Coraliomargarita akajimensis 88%
14A07 1 Uncultured marine eubacterium HstpL20 96%
40B02 1 Uncultured Verrucomicrobia clone GC21V A 93%
40C10 1 Uncultured Verrucomicrobia clone SW7 92%
40C08 1 Uncultured clone PI 4z2c 91%
13E08 1 Uncultured bacterium clone BB2INT16SI-6 81%
40F03 1 Uncultured Planctomycete clone 01D2Z81 92%
10H08 1 Uncultured Planctomycete clone 01D2Z81 96%
40B07 1 Uncultured Planctomycete clone GWS-Kdna31 92%
40D09 1 Uncultured Planctomycete clone SC1-36 97%
14B07 1 Uncultured Planctomycete clone 072DZ10 91%
40G07 1 Uncultured bacterium clone 1HP2-E3 96%
40D10 1 Uncultured Planctomycete clone IODP1324B2H2.74 85%
19H07 1 Uncultured Planctomycetales clone TDNP Bbc97 216 1 59 95%
19B09 1 Uncultured Planctomycete clone PLANCTO1 95%
26G03 1 Uncultured Planctomycete clone 9m05AISF05 94%
40D12 2 Uncultured Planctomycete clone B8A 90%
40A05 1 Uncultured Bacteroidetes clone 072DZ64 92%
40D05 1 Uncultured Flavobacteria clone SIMO-2432 87%
40G01 1 Uncultured Bacteroidetes clone 062DZ68 86%
40C09 1 Uncultured Bacteroidetes clone 10D2Z87 90%
35C11 1 Flavobacteriaceae bacterium JC2495 98%
40C07 1 Uncultured Bacteroidetes clone 062DZ52 90%
40D08 1 Uncultured Saprospiraceae clone E48H11cD 96%
40A03 1 Uncultured Bacteroidetesclone RB 14f 93%
19C10 1 Uncultured Bacteroidetes clone 9m05AISA08 94%
40G11 1 Uncultured Bacteroidales clone pltb-HW-39 89%
19F11 2 Uncultured Bacteroidetes clone PRD18HC8 94%
9G03 1 Tenacibaculum sp. CNURIC 013 94%
40H11 1 Uncultured Bacteroidetes clone 01D2Z32 98%
10E08 1 Muricauda sp. CC-HSB-11 91%
14D05 1 Muricauda sp. CC-HSB-11 91%
40A10 1 Uncultured Flavobacteria clone SIMO-2432 91%
26H04 1 Uncultured Flavobacteria clone SIMO-2432 96%
40A08 5 Uncultured Flavobacteria clone SIMO-2432 95%
40E05 1 Uncultured Robiginitalea clone 2B 113 98%
10F02 1 Uncultured Robiginitalea sp. clone 2B 113 97%
40E06 2 Uncultured Robiginitalea clone 2B 113 97%
40H04 1 Uncultured clone 01D2Z74 82%
19H03 1 Cyanobacterium clone LW18m-1-76 92%
26C01 1 Rhodomonas salina strain CCMP1319 chloroplast 95%
40A09 1 Synechococcus sp. MLCB 90%
40B06 1 Cyanobacterium Phormidium sp. SAG 61.90 89%
40H02 1 Filamentous Cyanobacterium LLi71 92%
40E07 1 Leptolyngbya sp. HBC8 93%
40G09 1 Phormidium sp. M1C2 99%
40H07 1 Leptolyngbya sp. 0BB32S02 94%
14D06 1 Uncultured bacterium clone BSb56 96%
40C11 1 Uncultured Cyanothece sp. clone Sc67 94%
40F05 1 Chroococcus sp. SCyano24 96%
40D03 1 Gloeocapsa sp. PCC 73106 96%
40E12 1 Synechococcus sp. MLCB 98%
40B01 1 Synechococcus sp. MBIC10613 93%
40C03 1 Calothrix sp. CCMEE 5093 92%
40F11 1 Uncultured clone 1DP2-N20 95%
26E03 1 Uncultured bacterium clone tet2ect1G12 88%
40F08 Uncultured clone 1DP2-A6 95%
40D11 1 Uncultured clone GASP-WA1W3 H09 98%
40A12 1 Uncultured clone GASP-0KA-501-E03 97%
40H10 1 Uncultured clone PI 4e4h 97%
40H05 1 Uncultured clone PI 4e4h 98%
40C04 1 Uncultured clone GASP-0KB-588-D02 96%
40E02 1 Uncultured clone 1DP2-A9 97%
26D07 1 Uncultured clone PI 4e4h 98%
10E10 1 Uncultured clone A822 98%
10G11 1 Bacillaria paxillifer chloroplast strain p73 96%
10F08 1 Haslea sp. chloroplast 96%
10G07 16 Uncultured clone 2B 337 98%
10G06 21 Uncultured clone 2B 337 98%
14B03 1 Uncultured clone 2B 337 98%
14D10 1 Uncultured clone 2B 337 98%
40E03 1 Uncultured clone 1DP2-A9 99%
40C01 1 Uncultured clone 1DP2-N20 95%
40D01 1 Uncultured clone GASP-MA1S1 C12 98%
40H01 1 Uncultured clone GASP-0KB-588-D02 98%
40D04 1 Uncultured clone 2B 337 96%
40G08 1 Uncultured clone PI 4e4h 97%
40G12 1 Uncultured clone 1DP2-N20 97%
40G10 1 Uncultured clone 2B 337 97%
40B10 1 Uncultured clone 2B 337 96%
40D06 1 Uncultured clone 2B 246 97%
40A06 1 Uncultured clone 1DP2-N20 94%
40G06 1 Uncultured clone 2B 103 93%
26A06 1 Uncultured clone 16B 16 90%
40E04 1 Uncultured clone 2B 246 92%
40B12 1 Uncultured clone 2B 323 93%
40E08 2 Uncultured clone 2B 323 95%
40E01 1 Uncultured clone 8B 256 91%
Chloroflexi
Spirochaetes
Verrucomicrobia
Planctomycetes
Bacteroidetes
Cyanobacteria
Chloroplasts
Legend
LAB
MA
CT
MB
SED

Fig. 5.

Fig. 5, cont. Maximum likelihood tree of all non-Proteobacteria clones based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

3.1.2 Sequence diversity estimation

In order to estimate how completely the libraries have been sampled and to predict total sequence diversity, sequences were grouped into 99% OTUs for the purpose of Chao 1 non-parametric diversity estimator calculations. Because the product of coverage and number of species (CV) was > 0.5, EstimateS recommended the use of Chao 1 Classic algorithm. Also, the Abundance-based Coverage Estimator of species richness (ACE) was taken as the best estimate for abundance-based richness if its value is higher that of than Chao 1.

Since chloroplasts are not part of the microbial diversity, they were removed from the species diversity analysis. The dataset indicated that estimated microbial diversity in the sediment was high with an ACE mean value of over 980. On the other hand, the Chao 1 mean value of animals obtained from the same site as the sediment was only 35% of that value at 340. CT and MB animals had even lower microbiome diversity, with ACE means of 23 and 47, respectively. The predicted microbiome diversity of reduced-flora LAB animals was higher than MB and CT animals but lower than MA animals at 150 (Table 3).

Rarefraction curves which plot the number of ribotypes detected vs. the number of clones screened concur with trends in Chao 1 analysis – SED, MA and LAB animal libraries did not reach an asymptote whereas CT and MB did - indicating that diversity in these environments was well-sampled while that in Massachusetts animals and sediments and lab animals was under-sampled (Fig. 6).

Table 3. Species richness estimation

Sample	ACE Mean	Chao 1 Mean	Chao 1 95% CI Lower Bound	Chao 1 95% CI Upper Bound
CT	**23**	23	11.31	75.92
MB	**46.65**	34.5	13.59	177.22
MA	261.06	**342.5**	85.87	1778.08
LAB	**150.16**	109.9	61.53	253.67
SED	**981.52**	814.24	259.99	2889.34

ACE and Chao 1 means, 95% confidence level lower bound and upper bound values as computed in EstimateS version 8. SED had the highest predicted diversity, followed by MA, LAB, MB and CT. Values taken as richness estimates (refer to section 3.1.3) are in boldface.

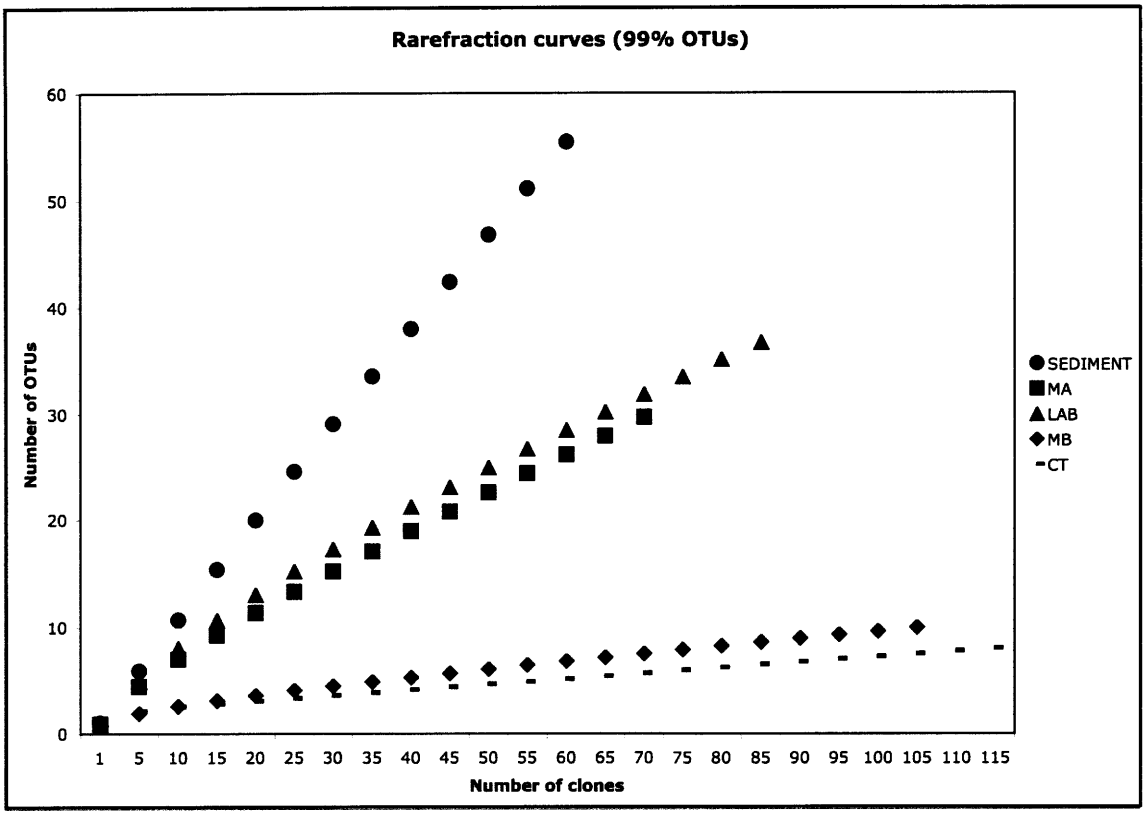


Fig. 6. Rarefraction curves for OTUs from SED, LAB, MA, MB and CT libraries (top to bottom) grouped at 99% similarity based on EstimateS version 8. The SED, LAB and MA libraries were under-sampled as the rarefraction curves showed that an asymptote has yet to be reached. On the other hand, the MB and CT libraries appeared to be well-sampled.

3.1.3 Phylogeny of 16S rRNA sequences in clone libraries

Clone identification was done by querying sequence from clone libraries against the Genbank database using the BLASTN program and taking the highest-scoring, closest related cultured isolate hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. For all libraries, maximum likelihood trees were calculated using the nucleotide substitution model GTR and evaluated with bootstrap value of 100 replicates.

The Sippewissett sediment sample (SED) was dominated by a wide range of Cyanobacteria/chloroplasts clones including sequences with highest identity to the well-described genera like *Synechococcus, Phormidium, Leptolyngbya, Chroococcus, Gloeocapsa, Calothrix*. Clusters of potentially novel Chloroflexi, Bacteroidetes, Planctomycetes that have 87% to 97%, 86% to 96% and 85% to 97% similarities to their respective closest BLAST hits made up about 33% of the clone library. The rest of the clones were gamma-Proteobacteria from the genera *Haliea* and *Halochromatium*, and other unknown delta-and alpha-Proteobacteria and Verrucomicrobia (Fig. 7).

The MA clone library (Fig. 8) was dominated by sequences with 96% to 98% identity to uncultured Cyanobacteria/chloroplasts and two known chloroplast isolates from the genera *Haslea* and *Bacillaria*. A surprising discovery is a previously unknown epsilon-Proteobacteria ribotype with 96% similarity score to the uncultured epsilon-Proteobacteria clone SGUS951 associated with White Plague Disease in coral *Montastraea faveolata* (Genbank ID: FJ202415, Sunagawa *et al.*, 2009) that made up a third of the clones. It is noteworthy that the same 99% OTU was found to constitute 95% and 98% of the CT (Fig. 9) and MB (Fig. 10) clone libraries. In addition, a gamma-Proteobacteria clone with 95% to 97% similarity to *Endozoicimonas elysicola* was present, though in much lower abundance, in both MA and CT libraries.

Clones from the CT library also included one novel *Spiroplasma* with 86% identity to the closest known isolate, and one of each of uncultured *Verrucomicrobia*, *Vibrio* sp. and unknown gamma-Proteobacteria clones 97% similar to their respective closest BLAST hits. The MB clone library was even more homogenous with only two additional clones other than the dominant epsilon-

Proteobacteria – one Flavobacteriaceae and two clones 98% similar to *Pseudomonas pseudoalcaligenes*. It is noted that clones with 97% to 99% similarity to *P. pseudoalcaligenes* were also present in the LAB (n=5) and MA (n=1) libraries.

Significantly different from field animals, reduced-flora animals (LAB) were dominated by 72% gamma-Proteobacteria from genera *Alteromonas, Pseudoalteromonas, Pseudomonas, Thalassomonas* and a few *Legionella, Vibrio* and *Methylophaga* clones. The rest of the clones consisted of alpha-Proteobacteria from genera *Kiloniella, Rhodobacteraceae, Rhizobiales* and *Kordiimonas*, novel Spirochetes clones with 92% or less similarity to their highest BLAST hits, Bacteroidetes from genus *Tenacibaculum*, and a delta-Proteobacterium from genus *Bacteriovorax* (Fig. 11).

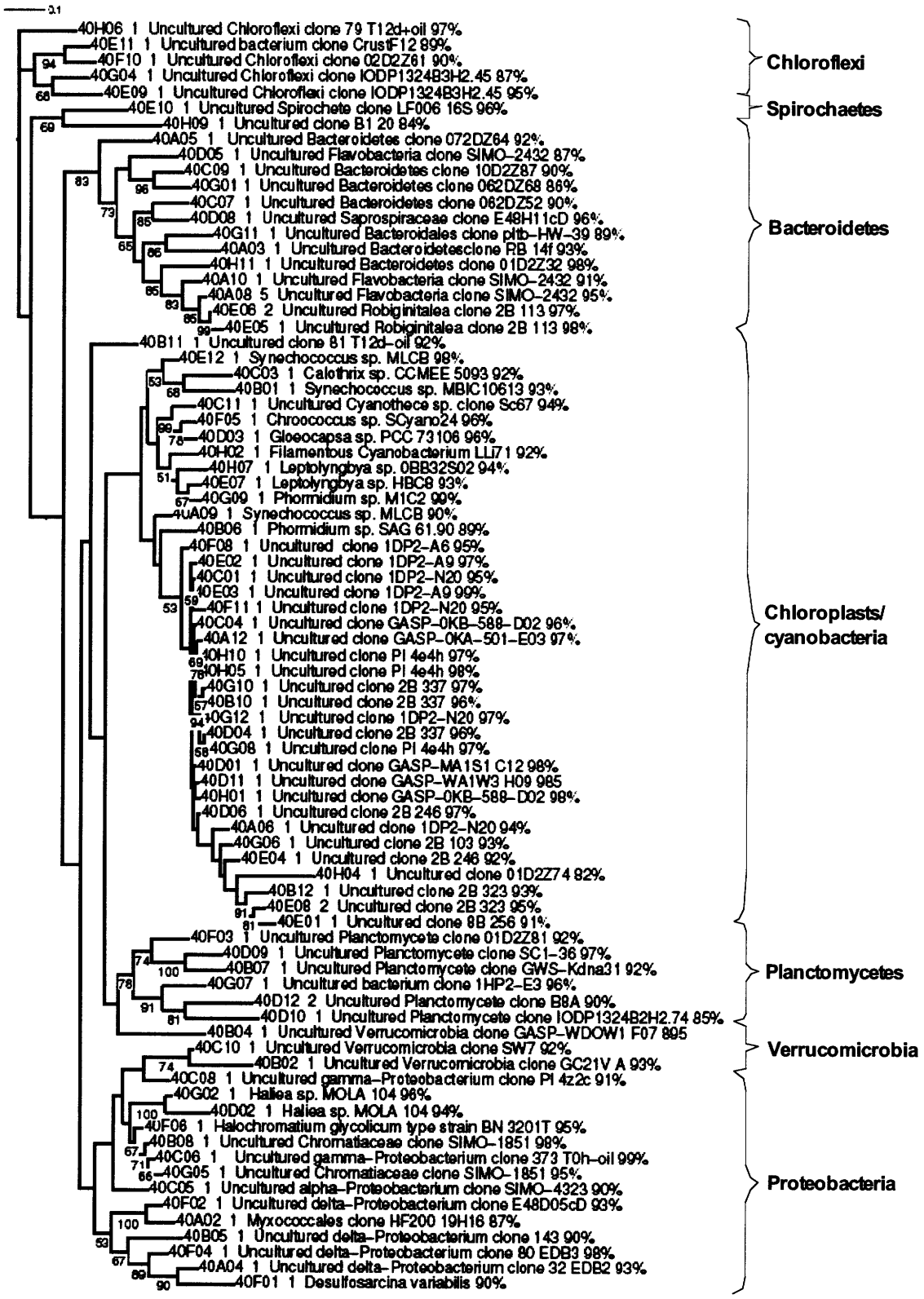

0.1
40H06 1 Uncultured Chloroflexi clone 79 T12d+oil 97%
40E11 1 Uncultured bacterium clone CrustF12 89%
40F10 1 Uncultured Chloroflexi clone 02D2Z61 90%
40G04 1 Uncultured Chloroflexi clone IODP1324B3H2.45 87%
40E09 1 Uncultured Chloroflexi clone IODP1324B3H2.45 95%
Chloroflexi
40E10 1 Uncultured Spirochete clone LF006 16S 96%
Spirochaetes
40H09 1 Uncultured clone B1 20 84%
40A05 1 Uncultured Bacteroidetes clone 072DZ64 92%
40D05 1 Uncultured Flavobacteria clone SIMO-2432 87%
40C09 1 Uncultured Bacteroidetes clone 10D2Z87 90%
40G01 1 Uncultured Bacteroidetes clone 062DZ68 86%
40C07 1 Uncultured Bacteroidetes clone 062DZ52 90%
40D08 1 Uncultured Saprospiraceae clone E48H11cD 96%
40G11 1 Uncultured Bacteroidales clone pltb-HW-39 89%
40A03 1 Uncultured Bacteroidetesclone RB 14f 93%
40H11 1 Uncultured Bacteroidetes clone 01D2Z32 98%
40A10 1 Uncultured Flavobacteria clone SIMO-2432 91%
40A08 5 Uncultured Flavobacteria clone SIMO-2432 95%
40E06 2 Uncultured Robiginitalea clone 2B 113 97%
40E05 1 Uncultured Robiginitalea clone 2B 113 98%
Bacteroidetes
40B11 1 Uncultured clone 81 T12d-oil 92%
40E12 1 Synechococcus sp. MLCB 98%
40C03 1 Calothrix sp. CCMEE 5093 92%
40B01 1 Synechococcus sp. MBIC10613 93%
40C11 1 Uncultured Cyanothece sp. clone Sc67 94%
40F05 1 Chroococcus sp. SCyano24 96%
40D03 1 Gloeocapsa sp. PCC 73106 96%
40H02 1 Filamentous Cyanobacterium LLi71 92%
40H07 1 Leptolyngbya sp. 0BB32S02 94%
40E07 1 Leptolyngbya sp. HBC8 93%
40G09 1 Phormidium sp. M1C2 99%
40A09 1 Synechococcus sp. MLCB 90%
40B06 1 Phormidium sp. SAG 61.90 89%
40F08 1 Uncultured clone 1DP2-A6 95%
40E02 1 Uncultured clone 1DP2-A9 97%
40C01 1 Uncultured clone 1DP2-N20 95%
40E03 1 Uncultured clone 1DP2-A9 99%
40F11 1 Uncultured clone 1DP2-N20 95%
40C04 1 Uncultured clone GASP-0KB-588-D02 96%
40A12 1 Uncultured clone GASP-0KA-501-E03 97%
40H10 1 Uncultured clone PI 4e4h 97%
40H05 1 Uncultured clone PI 4e4h 98%
40G10 1 Uncultured clone 2B 337 97%
40B10 1 Uncultured clone 2B 337 96%
40G12 1 Uncultured clone 1DP2-N20 97%
40D04 1 Uncultured clone 2B 337 96%
40G08 1 Uncultured clone PI 4e4h 97%
40D01 1 Uncultured clone GASP-MA1S1 C12 98%
40D11 1 Uncultured clone GASP-WA1W3 H09 985
40H01 1 Uncultured clone GASP-0KB-588-D02 98%
40D06 1 Uncultured clone 2B 246 97%
40A06 1 Uncultured clone 1DP2-N20 94%
40G06 1 Uncultured clone 2B 103 93%
40E04 1 Uncultured clone 2B 246 92%
40H04 1 Uncultured clone 01D2Z74 82%
40B12 1 Uncultured clone 2B 323 93%
40E08 2 Uncultured clone 2B 323 95%
40E01 1 Uncultured clone 8B 256 91%
Chloroplasts/ cyanobacteria
40F03 1 Uncultured Planctomycete clone 01D2Z81 92%
40D09 1 Uncultured Planctomycete clone SC1-36 97%
40B07 1 Uncultured Planctomycete clone GWS-Kdna31 92%
40G07 1 Uncultured bacterium clone 1HP2-E3 96%
40D12 2 Uncultured Planctomycete clone B8A 90%
40D10 1 Uncultured Planctomycete clone IODP1324B2H2.74 85%
Planctomycetes
40B04 1 Uncultured Verrucomicrobia clone GASP-WDOW1 F07 895
40C10 1 Uncultured Verrucomicrobia clone SW7 92%
40B02 1 Uncultured Verrucomicrobia clone GC21V A 93%
Verrucomicrobia
40C08 1 Uncultured gamma-Proteobacterium clone PI 4z2c 91%
40G02 1 Haliea sp. MOLA 104 96%
40D02 1 Haliea sp. MOLA 104 94%
40F06 1 Halochromatium glycolicum type strain BN 3201T 95%
40B08 1 Uncultured Chromatiaceae clone SIMO-1851 98%
40C06 1 Uncultured gamma-Proteobacterium clone 373 T0h-oil 99%
40G05 1 Uncultured Chromatiaceae clone SIMO-1851 95%
40C05 1 Uncultured alpha-Proteobacterium clone SIMO-4323 90%
40F02 1 Uncultured delta-Proteobacterium clone E48D05cD 93%
40A02 1 Myxococcales clone HF200 19H16 87%
40B05 1 Uncultured delta-Proteobacterium clone 143 90%
40F04 1 Uncultured delta-Proteobacterium clone 80 EDB3 98%
40A04 1 Uncultured delta-Proteobacterium clone 32 EDB2 93%
40F01 1 Desulfosarcina variabilis 90%
Proteobacteria

Fig. 7

Fig. 7. Maximum likelihood tree of SED library based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

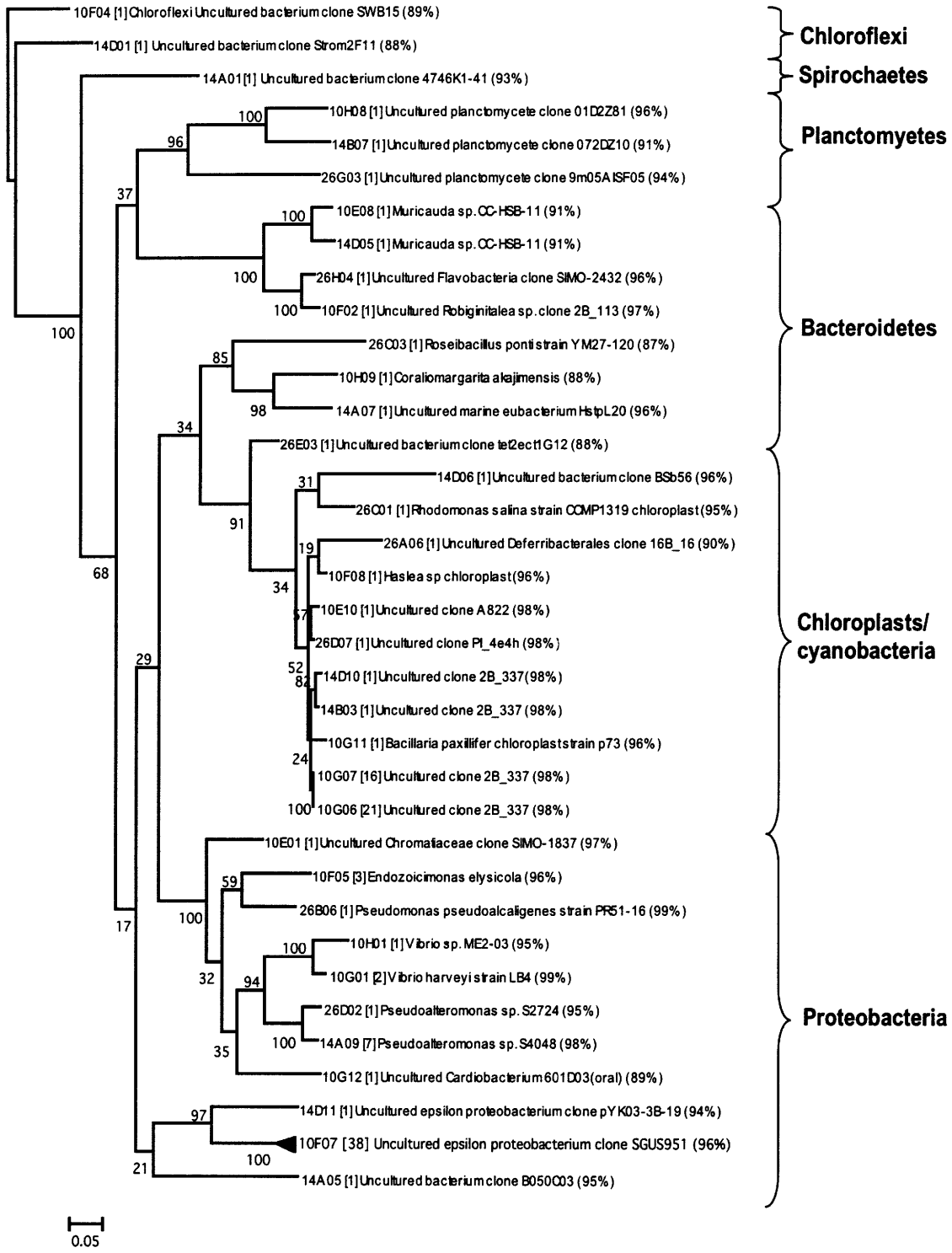

10F04 [1] Chloroflexi Uncultured bacterium clone SWB15 (89%)
14D01 [1] Uncultured bacterium clone Strom2F11 (88%)
14A01 [1] Uncultured bacterium clone 4746K1-41 (93%)
10H08 [1] Uncultured planctomycete clone 01D2Z81 (96%)
14B07 [1] Uncultured planctomycete clone 072DZ10 (91%)
26G03 [1] Uncultured planctomycete clone 9m05AISF05 (94%)
10E08 [1] Muricauda sp. CC-HSB-11 (91%)
14D05 [1] Muricauda sp. CC-HSB-11 (91%)
26H04 [1] Uncultured Flavobacteria clone SIMO-2432 (96%)
10F02 [1] Uncultured Robiginitalea sp. clone 2B_113 (97%)
26C03 [1] Roseibacillus ponti strain YM27-120 (87%)
10H09 [1] Coraliomargarita akajimensis (88%)
14A07 [1] Uncultured marine eubacterium HstpL20 (96%)
26E03 [1] Uncultured bacterium clone tet2ect1G12 (88%)
14D06 [1] Uncultured bacterium clone BSb56 (96%)
26C01 [1] Rhodomonas salina strain CCMP1319 chloroplast (95%)
26A06 [1] Uncultured Deferribacterales clone 16B_16 (90%)
10F08 [1] Haslea sp chloroplast (96%)
10E10 [1] Uncultured clone A822 (98%)
26D07 [1] Uncultured clone PI_4e4h (98%)
14D10 [1] Uncultured clone 2B_337 (98%)
14B03 [1] Uncultured clone 2B_337 (98%)
10G11 [1] Bacillaria paxillifer chloroplast strain p73 (96%)
10G07 [16] Uncultured clone 2B_337 (98%)
10G06 [21] Uncultured clone 2B_337 (98%)
10E01 [1] Uncultured Chromatiaceae clone SIMO-1837 (97%)
10F05 [3] Endozoicimonas elysicola (96%)
26B06 [1] Pseudomonas pseudoalcaligenes strain PR51-16 (99%)
10H01 [1] Vibrio sp. ME2-03 (95%)
10G01 [2] Vibrio harveyi strain LB4 (99%)
26D02 [1] Pseudoalteromonas sp. S2724 (95%)
14A09 [7] Pseudoalteromonas sp. S4048 (98%)
10G12 [1] Uncultured Cardiobacterium 601D03(oral) (89%)
14D11 [1] Uncultured epsilon proteobacterium clone pYK03-3B-19 (94%)
10F07 [38] Uncultured epsilon proteobacterium clone SGUS951 (96%)
14A05 [1] Uncultured bacterium clone B050C03 (95%)
0.05
Chloroflexi
Spirochaetes
Planctomyetes
Bacteroidetes
Chloroplasts/ cyanobacteria
Proteobacteria

Fig. 8

Fig. 8. Maximum likelihood tree of MA library based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

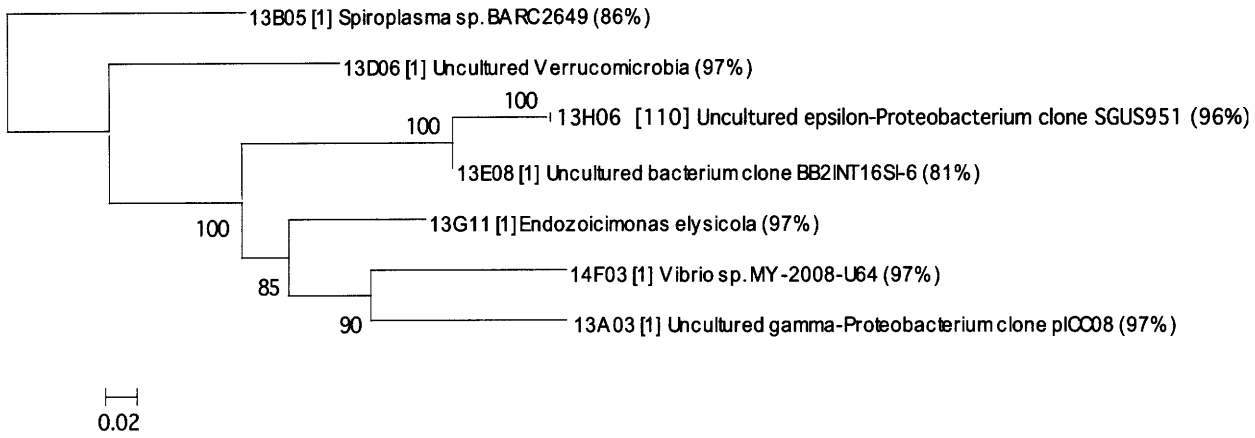


Fig. 9. Maximum likelihood tree of CT library based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

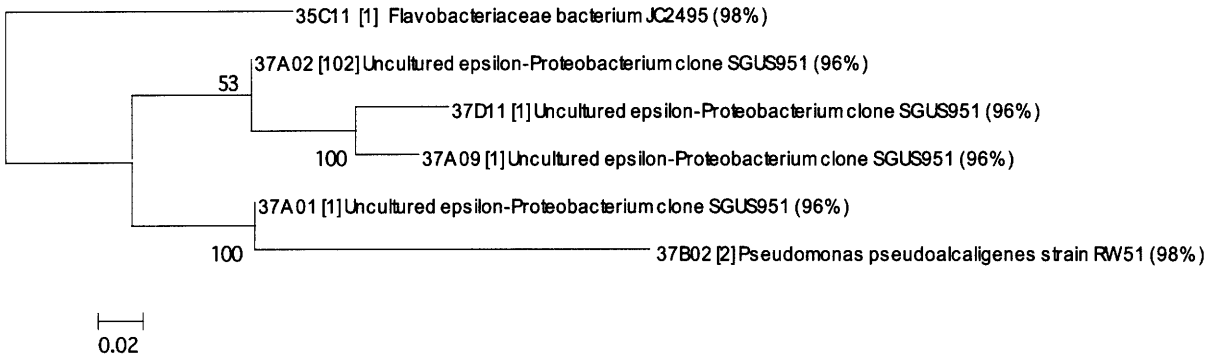


Fig. 10. Maximum likelihood tree of MB library based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

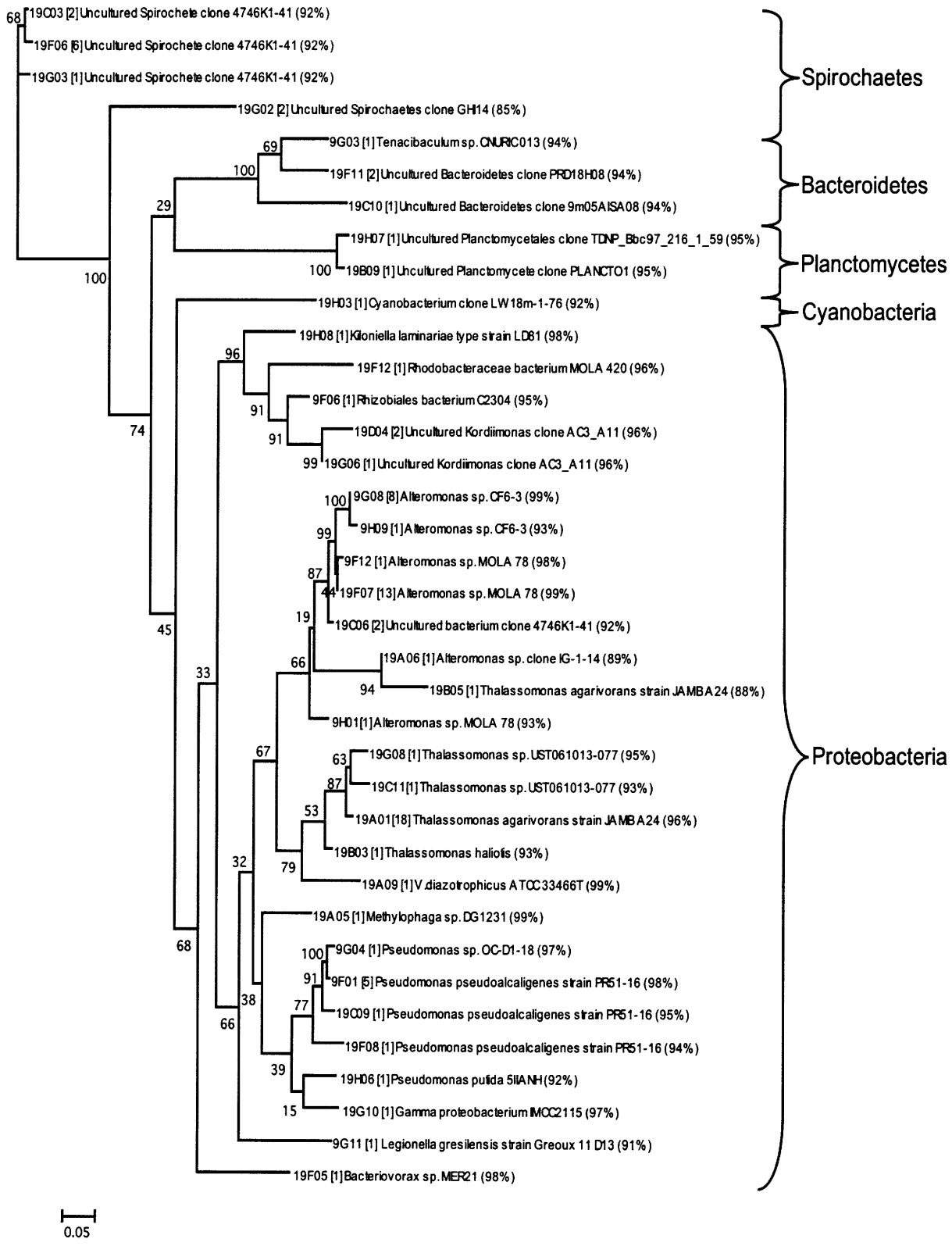

19C03 [2] Uncultured Spirochete clone 4746K1-41 (92%)
19F06 [6] Uncultured Spirochete clone 4746K1-41 (92%)
19G03 [1] Uncultured Spirochete clone 4746K1-41 (92%)
19G02 [2] Uncultured Spirochaetes clone GH14 (85%)
9G03 [1] Tenacibaculum sp. CNURIC013 (94%)
19F11 [2] Uncultured Bacteroidetes clone PRD18H08 (94%)
19C10 [1] Uncultured Bacteroidetes clone 9m05AISA08 (94%)
19H07 [1] Uncultured Planctomycetales clone TDNP_Bbc97_216_1_59 (95%)
19B09 [1] Uncultured Planctomycete clone PLANCTO1 (95%)
19H03 [1] Cyanobacterium clone LW18m-1-76 (92%)
19H08 [1] Kloniella laminariae type strain LD81 (98%)
19F12 [1] Rhodobacteraceae bacterium MOLA 420 (96%)
9F06 [1] Rhizobiales bacterium C2304 (95%)
19D04 [2] Uncultured Kordiimonas clone AC3_A11 (96%)
19G06 [1] Uncultured Kordiimonas clone AC3_A11 (96%)
9G08 [8] Alteromonas sp. CF6-3 (99%)
9H09 [1] Alteromonas sp. CF6-3 (93%)
9F12 [1] Alteromonas sp. MOLA 78 (98%)
19F07 [13] Alteromonas sp. MOLA 78 (99%)
19C06 [2] Uncultured bacterium clone 4746K1-41 (92%)
19A06 [1] Alteromonas sp. clone IG-1-14 (89%)
19B05 [1] Thalassomonas agarivorans strain JAMBA24 (88%)
9H01 [1] Alteromonas sp. MOLA 78 (93%)
19G08 [1] Thalassomonas sp. UST061013-077 (95%)
19C11 [1] Thalassomonas sp. UST061013-077 (93%)
19A01 [18] Thalassomonas agarivorans strain JAMBA24 (96%)
19B03 [1] Thalassomonas haliotis (93%)
19A09 [1] V. diazotrophicus ATCC33466T (99%)
19A05 [1] Methylophaga sp. DG1231 (99%)
9G04 [1] Pseudomonas sp. OC-D1-18 (97%)
9F01 [5] Pseudomonas pseudoalcaligenes strain PR51-16 (98%)
19C09 [1] Pseudomonas pseudoalcaligenes strain PR51-16 (95%)
19F08 [1] Pseudomonas pseudoalcaligenes strain PR51-16 (94%)
19H06 [1] Pseudomonas putida 5IIANH (92%)
19G10 [1] Gamma proteobacterium IMCC2115 (97%)
9G11 [1] Legionella gresilensis strain Greoux 11 D13 (91%)
19F05 [1] Bacteriovorax sp. MER21 (98%)
Spirochaetes
Bacteroidetes
Planctomycetes
Cyanobacteria
Proteobacteria
0.05

Fig. 11

Fig. 11. Maximum likelihood tree of LAB library based on 16S rRNA partial sequences *E. coli* positions 57-784. Every leaf represents a 99% OTU and is labeled by a unique identifier, number of sequences contained in the OTU and the identity of the top known cultured BLAST hit. If a cultured representative was not available in the top 20 hits, the uncultured hit with the highest similarity score was chosen. Bootstrap values were shown for branches with >50% support.

3.1.4 Comparison between microbial communities

Principal component analysis within the Unifrac suite of programs was used to evaluate the similarity between clone libraries from different environments (Fig. 12). As expected from the similarity in the clone libraries, the two geographically-isolated samples MB and CT were tightly clustered, whereas MA and SED formed another loose cluster, indicating the two samples obtained from the same location share some similarity as suggested by the clone libraries. The LAB animal sample was furthest away from any other clusters, confirming sequencing results that lab animal microbiome was significantly different from field animals. As anticipated, the MA sample was placed in between the MB/CT cluster and SED library due to presence of both epsilon-Proteobacteria (present in MB/CT but not in SED) and Cyanobacteria/chloroplasts (present in SED but not MB/CT) in the MA clone library.

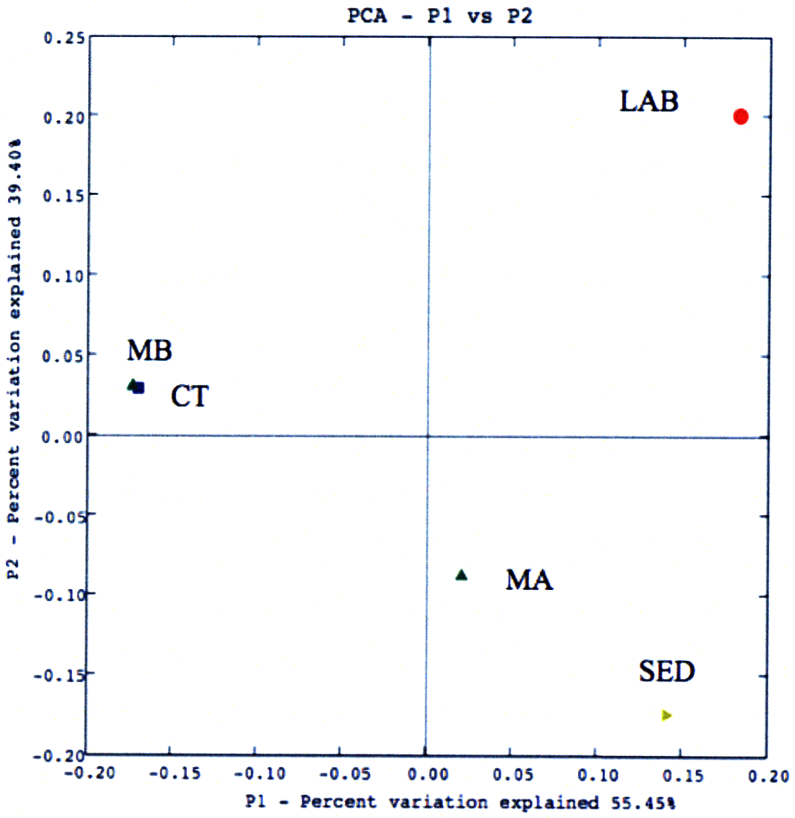


Fig. 12. Principal component analysis scatter plot. The percentages in the axis labels represent the percentages of variation explained by the principal coordinates. MB and CT clone libraries were the most similar to one another, MA and SED formed another loose cluster, while the LAB library was significantly different from the field samples.

3.1.5 Screening of Archaea in *N. vectensis* microbiome

Amplification of lab, Sippiwissett and Clinton *N. vectensis* genomic DNA with Archaea-specific primers (Arch20F and Arch958F) did not yield any PCR product. We conclude that Archaea are not a major component of the lab-raised or field animal holobiont.

3.2 Termination restriction fragment length polymorphism (T-RFLP)

Epsilon-Proteobacteria, *Pseudomonas pseudocaligenes* and Flavobacteria (Bacteroidetes) clones were digested with HpaII and HaeIII and used as T-RFLP size standards. T-RFLP profiles (Tables 4 and 5) of reduced-flora animals (LAB), Sippewissett (MA), Clinton (CT) and Mahone Bay (MB) animals and Sippewissett sediment (SED) revealed that the epsilon-Proteobacteria peak (168 bp in HpaII and 212 bp in HaeIII digestions) was present in all MA, MB and CT animals. The Flavobacteria (200 bp in HpaII and 33 bp in HaeIII digestions) and *Pseudomonas* (480bp in HpaII and 33bp in HaeIII digestions) peaks were detected in reduced-flora animals but not in the field samples. This may be due to the relatively lower proportion of these groups in the field animals as compared to epsilon-Proteobacteria. In reduced-flora lab animals, a total of 6 peaks were observed.

3.2.1 Digestion with HpaII

Table 4. Common T-RFLP fragments observed across multiple samples from HpaII digestion of 27F-FAM-805R PCR product of each sample.

Sample	Peak size (probable ID*)			Other peaks (bp)
	168 bp (E-Proteobacteria)	200 bp (Flavobacteria)	480 bp (*Pseudomonas*)	
LAB	-	X	X	90, 139, 434, 490
MB	X	-	-	-
CT	X	-	-	-
MA	X	-	-	500
SEDIMENT	-	-	-	-

"X" denotes presence of peak, "-" denotes absence of peak. *Peak identifications were based on in silico digestion prediction of known sequence types and confirmation with T-RFLP fragment sizes of corresponding clones.

3.2.2 Digestion with HaeIII

Table 5. Common T-RFLP fragments observed across multiple samples from HaeIII digestion of 27F-FAM-805R PCR product of each sample.

Sample	Peak size (probable ID*)		Other peaks (bp)
	33 bp (*Pseudomonas*/Flavobacteria)	212 bp (E-Proteobacteria)	
LAB	X	-	192, 215, 260, 324
MB	-	X	-
CT	-	X	-
MA	-	X	-
SEDIMENT	X	-	-

"X" denotes presence of peak, "-" denotes absence of peak. *Peak identifications were based on in silico digestion prediction of known sequence type and confirmation with T-RFLP fragment sizes of corresponding clones.

3.3 Fluorescent in-situ hybridization (FISH) analysis

3.3.1 Bacteria form cell aggregates on animal mesenteries

Clone libraries sequence analyses have shown that the microbiome of field *Nematostella vectensis* taken from more than one site contained substantial representatives of a previously unknown epsilon-Proteobacteria clone, *Pseudomonas pseudoalcaligenes* and *Endozoicimonas elysicola*. To visualize on which tissues these endosymbionts reside within *N. vectensis*, we prepared paraffin-embedded 4 μm sections of analyzed Sippewissett and lab animals and performed fluorescent in situ hybridization (FISH) with Cy3-labeled general eubacteria probe (EU338) and FAM-labeled probes specific to the epsilon-Proteobacteria ribotype found in this

study and *Pseudomonas* in separate reactions. Lab animals were in general 1 cm long and field animals about 5 mm long, probably due to more regular feeding under lab conditions.

N. vectensis tissue structure was examined using haematoxylin and eosin (H&E) stains. Figs. 13 and 14 show the relative position of the mesenteries within an adult animal with the three tissue layers – ectoderm, mesoglea and endoderm - and location of cell nuclei (dyed blue by haematoxylin). Bacteria occurred in aggregates associated with mesenteries. Fig. 15 shows the position of bacteria aggregates with respect to the mesenteries. Like other bacteria, epsilon-Proteobacteria (Fig. 16) and *Pseudomonas* (Fig. 17) occur primarily in cell aggregates on the mesentery walls that divide the animal tubular column, but were not found on other parts of the animal like the tentacles or the mouth. These aggregates ranged from small aggregates of less than 10 cells each to more than 100 cells each. The average number of aggregates on a single animal longitudinal section was 144 ± 35 and 139 ± 17 for lab and field animals, respectively (Fig. 18). In both lab and field animals, most aggregates (~ 65% of total aggregates) contained 10-30 cells each, while those with <10 cells and 30-60 cells constituted another 15% each. Very large aggregates with more than 60 cells were uncommon and constituted only <5% of total aggregates observed (Figs. 19 and 20).

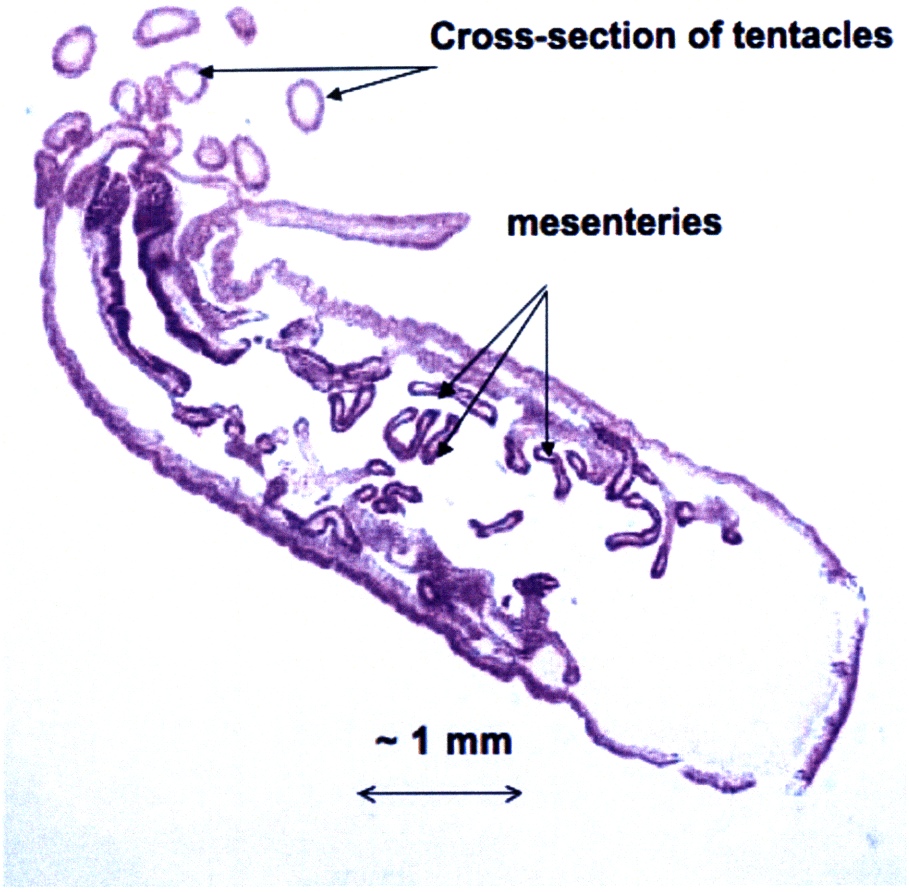


Fig 13. Haematoxylin and eosin (H&E) staining of longitudinal section of whole *N. vectensis* at 40 x magnification showing position of mesenteries in the context of the whole animal. Mesenteries that increase surface area for absorption of food in the mid-section of the coelenteron and are not present in the posterior end. Cell nuclei were stained blue by haematoxylin, while cytoplasm, connective tissue and other extracellular substances pink or red by eosin. Specimen preparation and photo by Samodha C. Fernando.

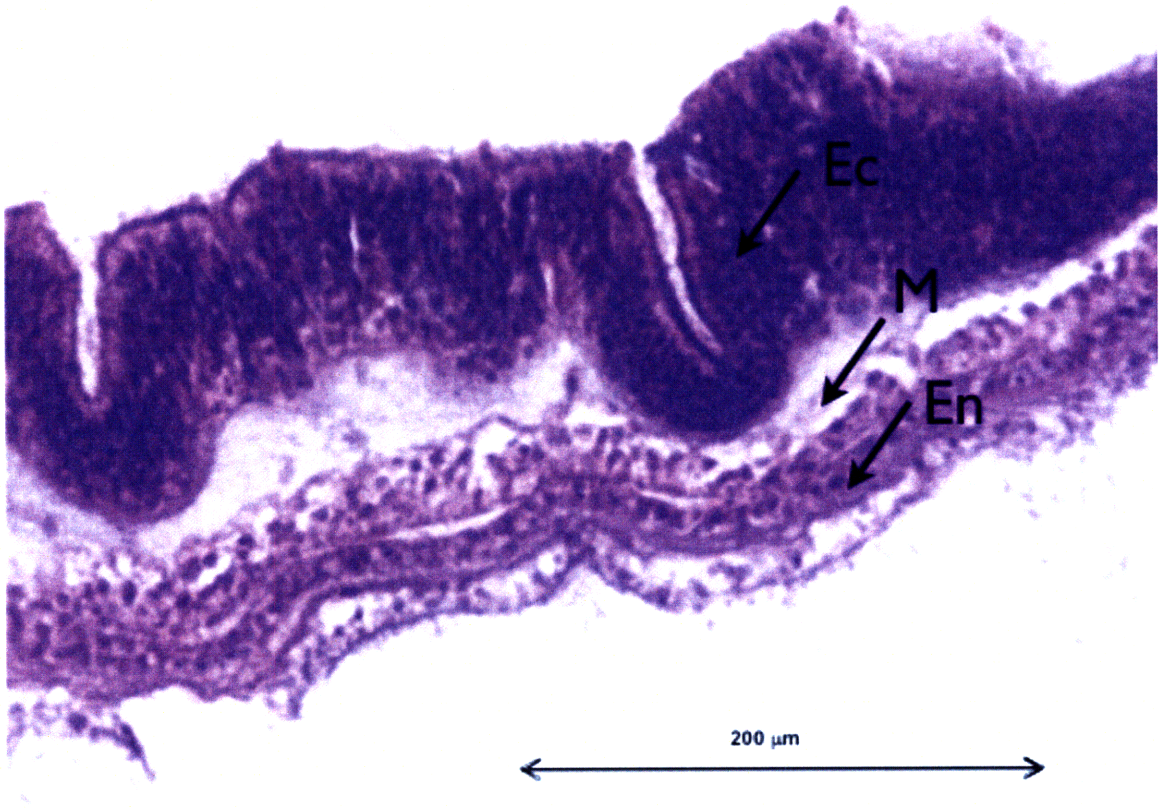


Fig. 14. Haematoxylin and eosin (H&E) staining of one of the mesenteries at 1000 x magnification, showing three distinct tissue layers from top to bottom: ectoderm, mesoglea and endoderm. Cell nuclei were stained blue by haematoxylin, while cytoplasm, connective tissue and other extracellular substances pink or red by eosin. Specimen preparation and photo by Samodha C. Fernando.

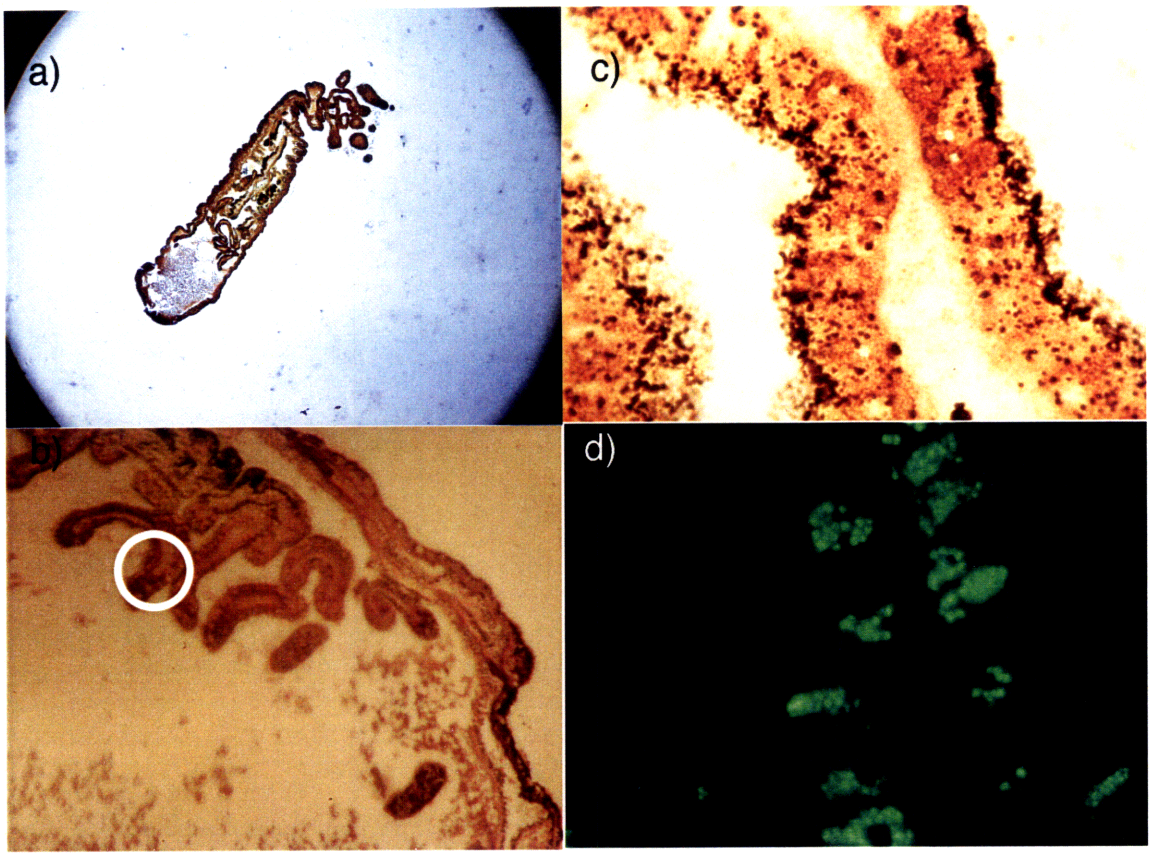


Fig 15. *N. vectensis* longitudinal sections using silver stain under a) 4 x magnification b) 400 x magnification showing the relative location of mesenteries. c) and d) are 1000 x magnification of the same region with silver staining and FAM-EUB338 probe (labeling bacteria cells green), respectively, showing the locations and arrangements of bacteria cell aggregates. Specimen preparation and photos by Samodha C. Fernando.

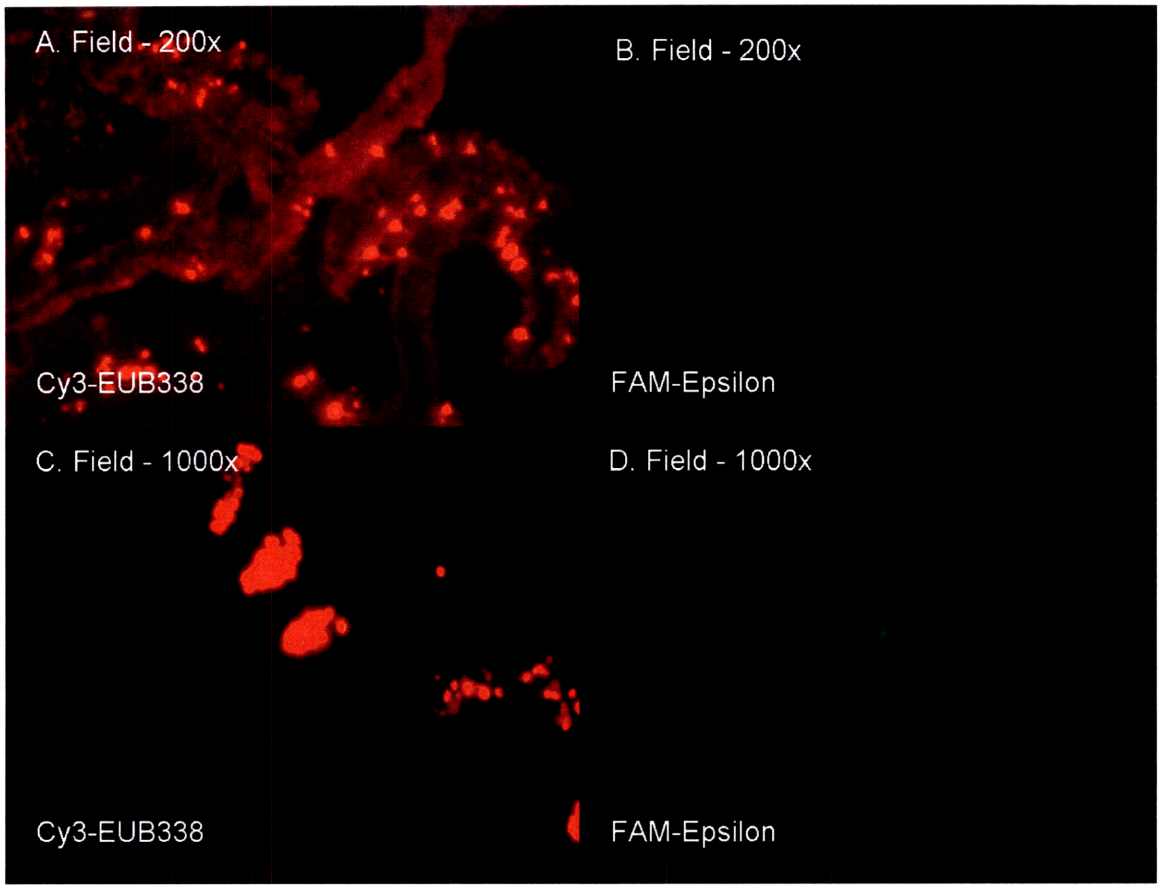


Fig. 16. FISH analysis of epsilon-Proteobacteria in field *N. vectensis* mensenteries under 200 x (A, B) and 1000 x (C, D) magnifications. Longitudinal sections of adult animals were hybridized with Cy3-EUB338 (Red: A, C) and FAM-labeled probes specific for epsilon-Proteobacteria (Green: B, D). Bacteria cells form cell aggregates ranging from less than 10 cells per aggregate to around 100 cells per aggregate, and are closely associated with animal mesentery tissues. Specimen preparation by Samodha C. Fernando, photos by Jia Yi Har.

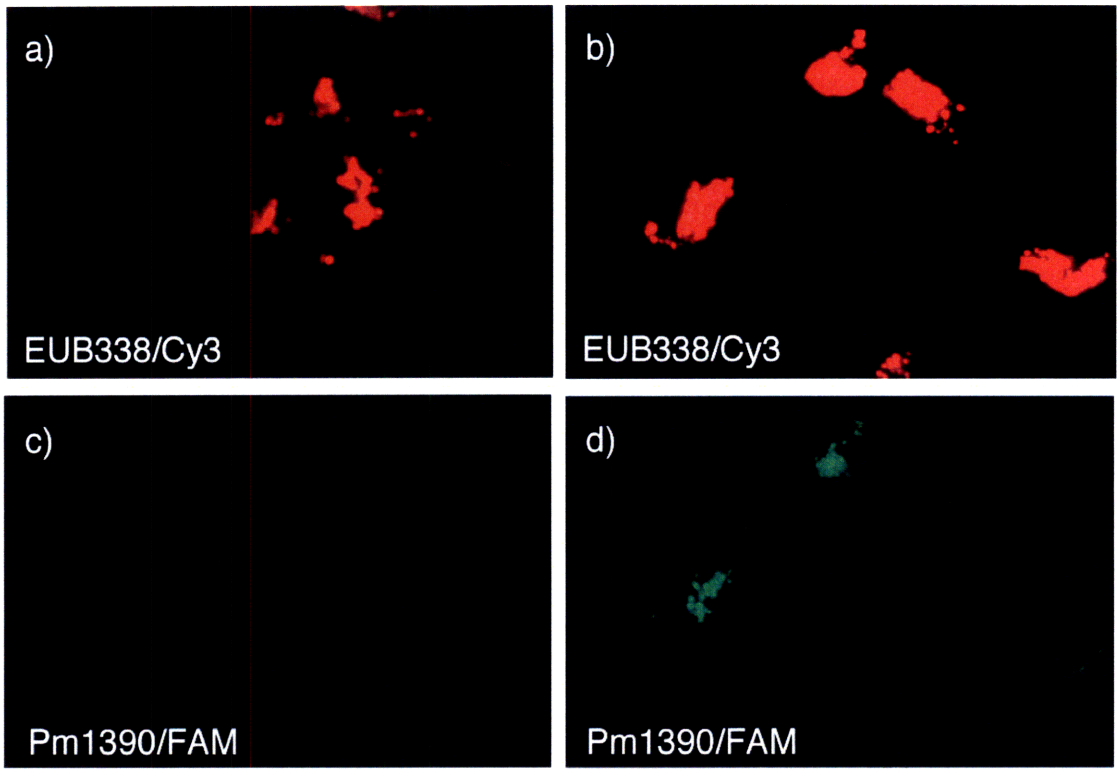


Fig. 17. FISH analysis of *Pseudomonas* in field (A, C) and lab (B, D) *N. vectensis* mensenteries under 1000 x magnification. Longitudinal sections of adult animals were hybridized with Cy3-EUB338 (Red: A, B) and FAM-labeled probes specific for *Pseudomonas* (Green: C, D). In this particular section, *Pseudomonas*-positive cells were only present in the lab but not in the field. Bacteria cells form cell aggregates and are closely associated with animal mesentery tissues. Specimen preparation and photos by Samodha C. Fernando.

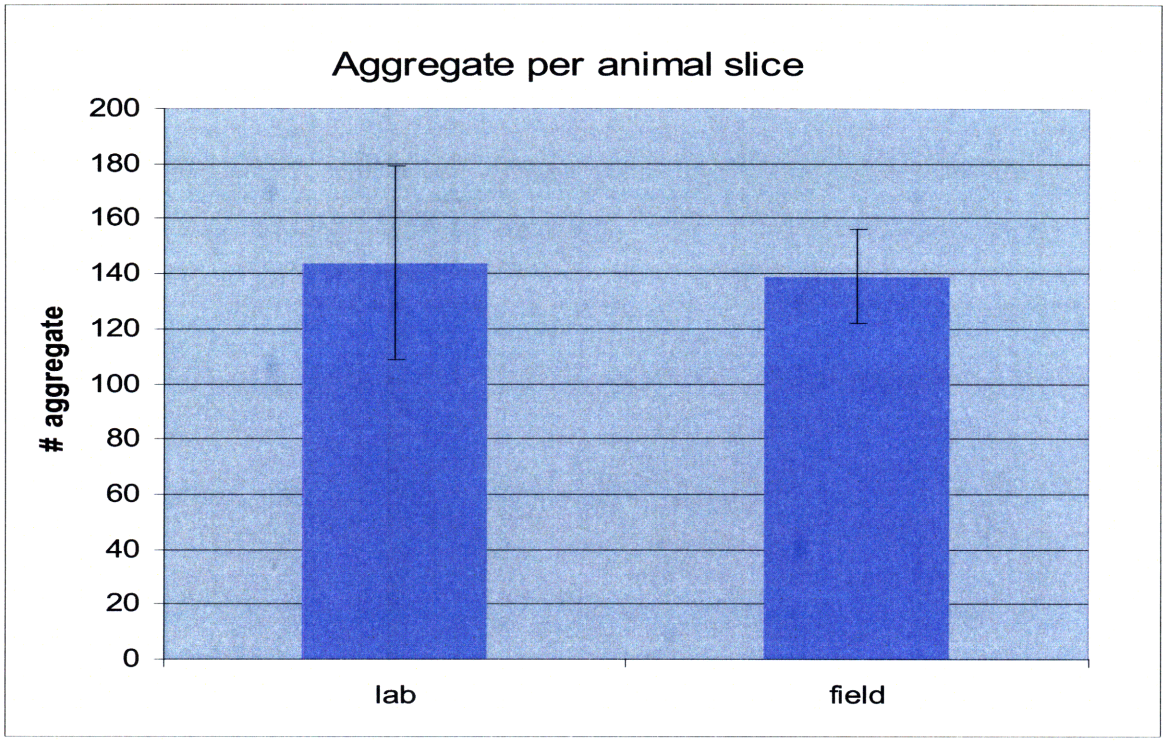


Fig. 18. Histogram of total number of aggregates per animal thin-section in lab (n=3) and field (n=3) *N. vectensis*.

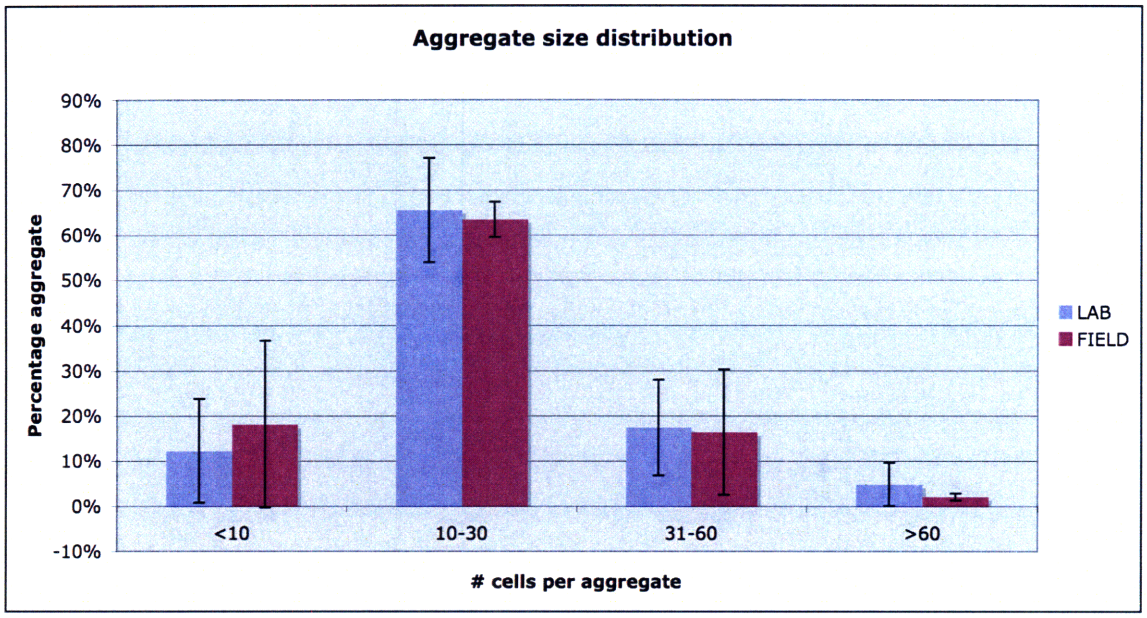


Fig.19. Histogram showing the distribution of aggregate sizes in lab (n=3) and field (n=3) *N. vectensis*.

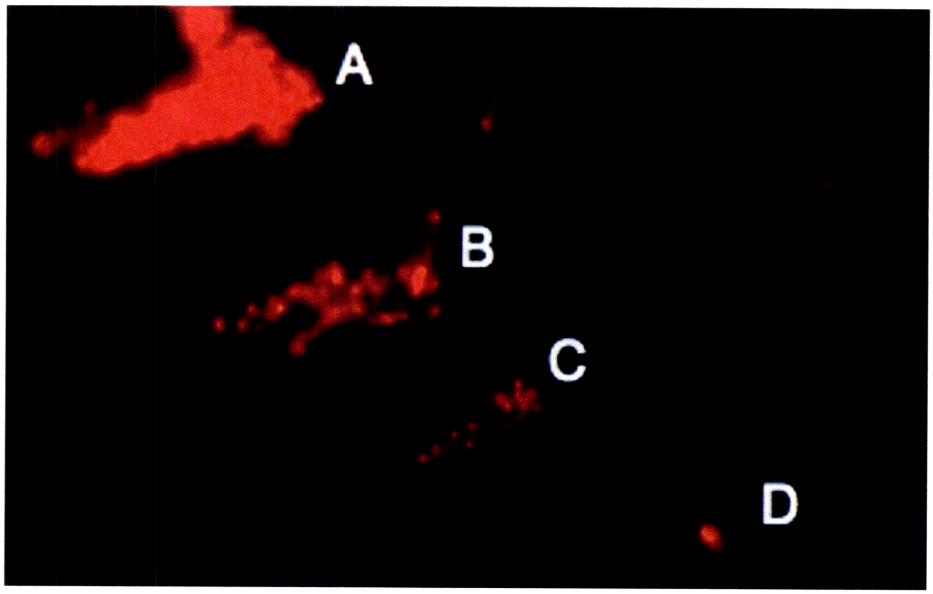


Fig.20. FISH image of Cy3-EUB338 labeled cells on a *N. vectensis* thin section at 1000 x magnification. A) 'very large' aggregate with > 60 cells, B) 'large' aggregate with ~ 40 cells, C) 'medium' aggregate with ~ 20 cells, D) 'small' aggregate with ~5 cells.

3.3.2 Estimation of number of bacteria cells in one animal

As a way to increase animal population size, *N. vectensis* adults raised in our laboratory were routinely cut into halves once they grew to ~ 2 cm. Therefore, the length of animals in the lab (including the ones used in FISH analysis) was on average ~ 1 cm. Since the diameter of the animal is about a quarter of its length, we assumed the diameter to be 0.25 cm.

From cell enumeration in FISH analysis,

$$n_{\text{aggregate per animal slice}} = 140 \frac{aggregate}{slice}$$

$$\text{Weighted average of } n_{\text{cells per aggregate}} = 25 \frac{cells}{aggregate}$$

$$\therefore n_{\text{cells per slice}} = 25 \frac{cells}{aggregate} * 140 \frac{aggregate}{slice} = 3500 \frac{cells}{slice}$$

Assuming animal diameter = 0.25 cm, animal length = 1 cm, thickness of thin-section = 4 μm (Fig. 21),

$$v_{slice} = dlt = 2.5 \times 10^{-3} \text{ m} * 10^{-2} \text{ m} * 4 \times 10^{-6} \text{ m} = 10^{-10} \text{ m}^3$$

Assuming the animal is a cylinder and that the animal is sectioned longitudinally through its midpoint for aggregate enumeration,

$$v_{animal} = \pi r^2 t = \pi (0.125 \text{ cm})^2 * 1 \text{ cm} = 5 \times 10^{-8} \text{ m}^3$$

$$n_{\text{cells per animal}} = v_{animal} / v_{slice} * n_{\text{cells per slice}}$$

$$= 5 \times 10^{-8} \text{ m}^3 / 10^{-10} \text{ m}^3 * 3500$$

$$= \mathbf{1.8 \times 10^6 \text{ cells / animal}}$$

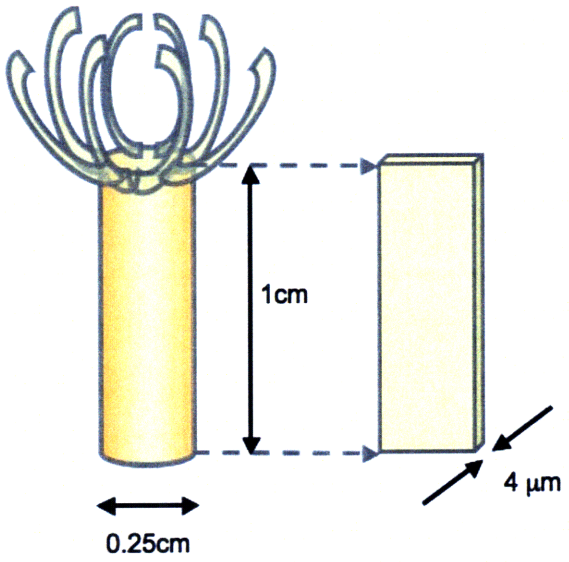


Fig. 21. Diagram of *N. vectensis* showing dimension of animal and animal slices used for FISH analysis and cell number estimation. Animals were typically cylindrical with diameter of 0.25 cm and length of 1 cm. Animal longitudinal sections were 4 µm thick.

3.3.3 Specificity of Epsilon-Proteobacteria and *Pseudomonas* probes

Querying the 24-mer Epsilon-Proteobacteria probe (Query) against the Genbank database (Subject) revealed that it hybridizes to positions 565 to 588 of its closest match in database - the uncultured epsilon-Proteobacteria clone SGUS951 - and had one mismatch in the 14th position, i.e. in the middle of the probe (Fig. 22). The match had a max score of 40.1 in BLASTN. The weighted mismatch score calculated according to ARB-Probe Match Online (Ludwig *et al.*, 2004) that indicates the level of destabilization due to mismatched bases between probe and target was 0.9, as opposed to 0 for perfect matches.

```
Query  1    TTAACCATAGAACTGCAAGACAAA  24
            ||||||||||||| ||||||||||
Sbjct  565  TTAACCATAGAACGGCAAGACAAA  588
```

Fig. 22. Alignment between epsilon-Proteobacteria (Query) probe and epsilon-proteobacteria clone SGUS951 (Sbjct), showing a T-G mismatch at position 14 of the probe.

The rest of the top 100 BLAST hits were other epsilon-Proteobacteria that have complementary sequences to the first 17 bases and last 3 bases of the probe, with 4 mismatches from positions 18-21 of the probe and had max scores of 34.2 and a weighted mismatch score of 4.5 (Fig. 23). Therefore we could assume that the epsilon-Proteobacteria probe is specific to the epsilon-Proteobacteria clone found in this study and are sensitive to mismatches in nearest non-targets.

```
Query  1    TTAACCATAGAACTGCAAGACAAA  24
            ||||||||||||||||    |||
Sbjct  568  TTAACCATAGAACTGCATTTGAAA  584
```

Fig. 23. Alignment between epsilon-Proteobacteria probe (Query) and *Helicobacter suis* 16S rRNA (representative of all top hits other than SGUS951, Sbjct), showing 4 mismatches at positions 18-21 of the probe.

The 24-mer *Pseudomonas* probe was designed to hit *Pseudomonas* species including *P. pseudoalcaligenes, P. mendocina and P. stutzeri* with perfect match.

3.4 Microdiversity of Epsilon-Proteobacteria across different field sites

In order to examine single polynucleotide polymorphisms (SNPs), we have grouped all 251 epsilon-Proteobacteria clones into 100% OTUs and found 17 unique types with either SNPs, single-, di-, or multimer indels. The dominant OTU (represented by sequence ID 13A08)

contained 226 sequences, while the other 5 OTUs only had less than 6 sequences each. The rest of the 11 were singletons. Positions with insertions, deletions and SNPs are presented in Table 6. The full alignment is shown in Fig. 24.

Table 6. Single nucleotide polymorphisms and indels in epsilon-Proteobacteria.

E. coli positions	Sequence ID	Nucleotide change	Sample origin
63	37A01	C→ G	MB
64	37A01	INS: CGAGCGGATGAAGGG	MB
86	37A01	INS: TCT	MB
87	37A01	A →G	MB
88-94	37A01	DEL: GTGTCAG	MB
95	37A01	C → T	MB
97	37A01	A→ C	MB
104-109 105	37A01 10F07	GGACGG → CCTATA G → C	MB MA
121	37A01	A→ C	MB
126-130	37A01	ACGTA → CTC::	MB
141-144	37A01	CT → GG	MB
149-154	37A01	CAGTTG→CGTTCC	MB
160	37A01	C → G	MB
164	37A01	T → C	MB
175-176	37A01	GA → CG	MB
179-181	37A01	TTC → CGT	MB
155-156	37A01	CA→ AC	MB
194-207	37A01	INS: GCAGGGGACCTTCG	MB
211-216	37A01	TTTATT→ CCTTGC	MB
220-227	37A01	CAAGATCG → TCAGATGA	MB
237-241	37A01	CTATC → GGATT	MB
272-275	37A01	CAAT → CGAC	MB
278-284	37A01	CGGGTAG → TCCGTAA	MB
399	**13E11** **13H06**	**C→ T** **C→ T**	**CT** **CT**
566	35B09	INS: G	MB
587	37E05	C → G	MB
598-599	37A06	INS: GT	MB
674-675	**37A09** **37D11**	**TG→CT** **TG→CT**	**MB** **MB**
654	35E12	INS: G	MB
681	35E12	INS: G	MB

681-699	37A09 37D11	DEL GGTGGTGTAGGGGTAAAAT → :CCAGCACACTGGCGG:::	MB MB
711	37A09	A → C	MB
712-744	37A09	DEL	MB
714	35A07	A → G	MB
717-721	37D11	GAATG → TGGAT	MB
725-731	37D11	CGGCGAA → GCTCGGT	MB
739-745	37D11	DEL	MB
748	37A09 37D11	A→ C A→ G	MB MB
750	37A09	T → C	MB
751	37D11	A→ G	MB
753-754	37D11	TG → AT	MB
755-761	37A09	ACGCTGAG → GCGGCC:	MB
762	**13B12** **13C11**	**G → A** **G → A**	**CT** **CT**
763-764	37D11	:G→ TA	MB
770-779	37A09	AAGCGTGGGG→ GCATGCATCT	MB
776-785	37D11	DEL	MB
787	37A09 14G11	A →C A→ G	MB CT
788	**37A09** **37D11**	**G → C** **G → C**	**MB** **MB**
790-791	37D11	AT → CC	MB
793	37A09	A → C	MB
794	**37D11** **13E03** **13G06**	**G→A** **G→A** **G→A**	**MB** **CT** **CT**
795-797	37A09	DEL	MB
798	37D11	C → G	MB
802-804	37A09	GGT → ATA	MB

"→" denotes substitution, "INS" denotes insertion, "DEL" denotes deletion of all positions, ":" denotes gap. Positions with more than one sequence sharing the same substitution patterns are marked in boldface.

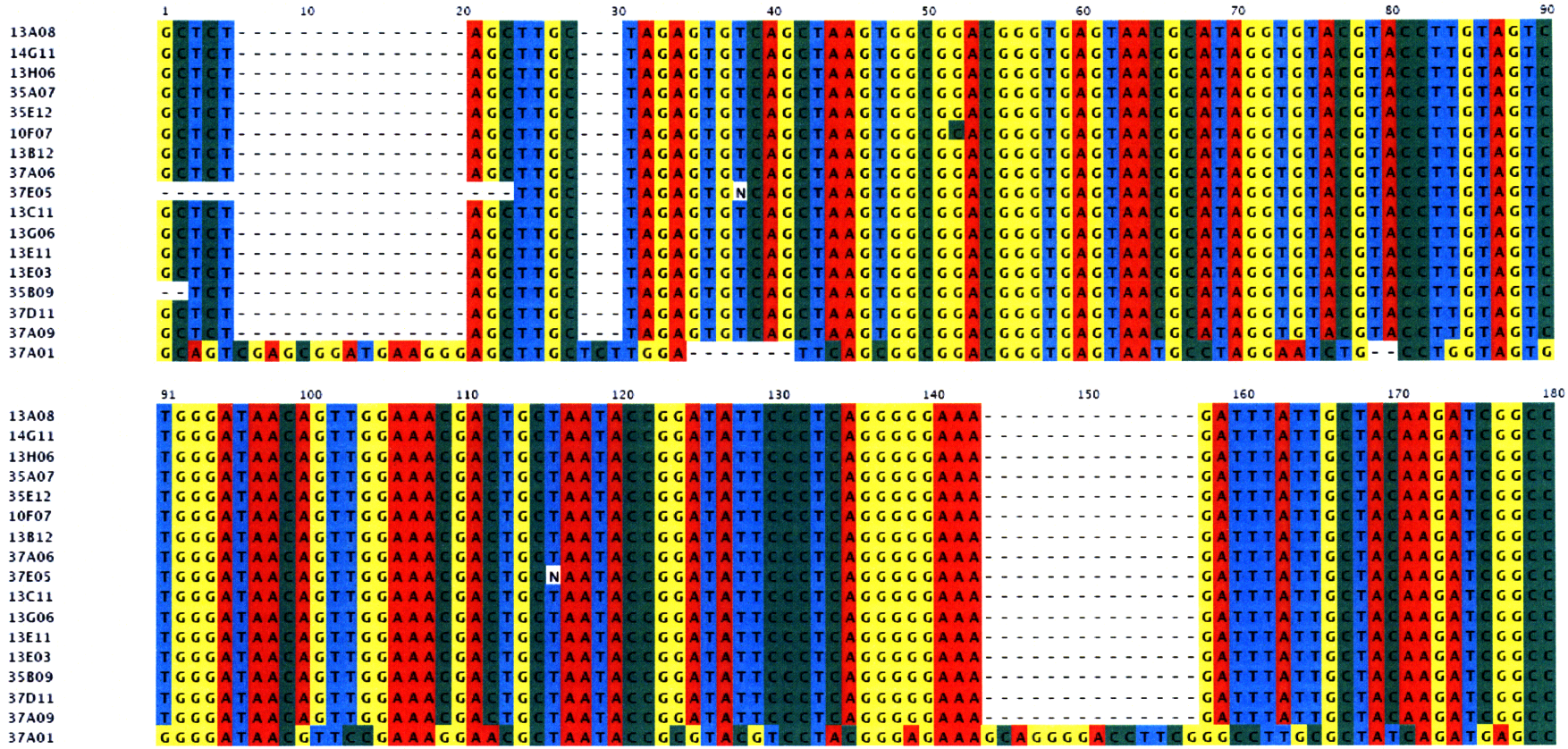


Fig. 24. Multiple sequence alignment showing microdiversity of epsilon-Proteobacteria clones. 17 unique sequence types were found after grouping all 251 epsilon-Proteobacteria clones into 100% OTUs. The dominant OTU (represented by 13A08) contained 226 sequences. Positions with insertions, deletions and SNPs are summarized in Table 6. Bases are presented in different colors to show nucleotide polymorphisms. Positions are consensus alignment positions with gaps and should be used as reference only.

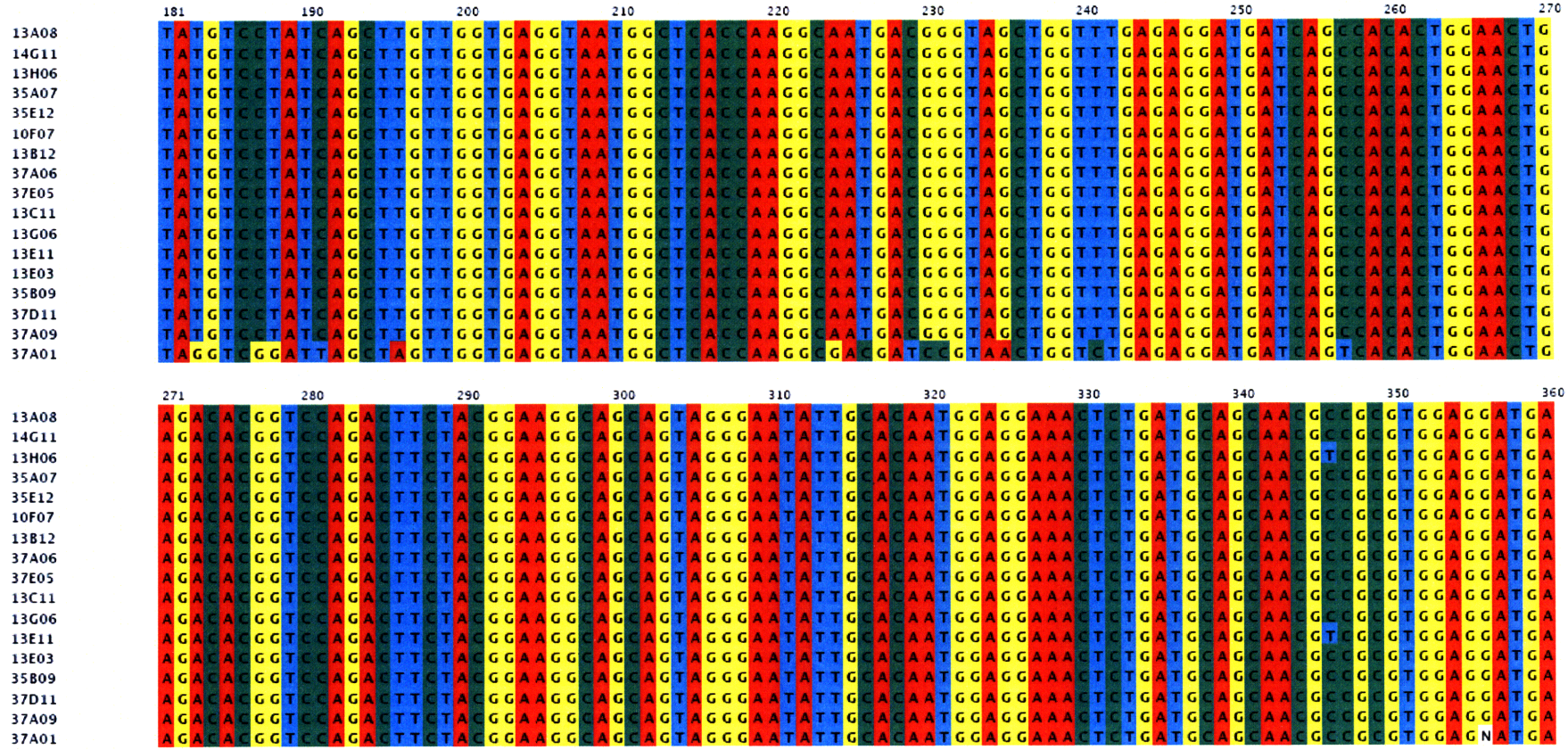


Fig. 24 (cont'd). Multiple sequence alignment showing microdiversity of epsilon-Proteobacteria clones. 17 unique sequence types were found after grouping all 251 epsilon-Proteobacteria clones into 100% OTUs. The dominant OTU (represented by 13A08) contained 226 sequences. Positions with insertions, deletions and SNPs are summarized in Table 6. Bases are presented in different colors to show nucleotide polymorphisms. Positions are consensus alignment positions with gaps and should be used as reference only.

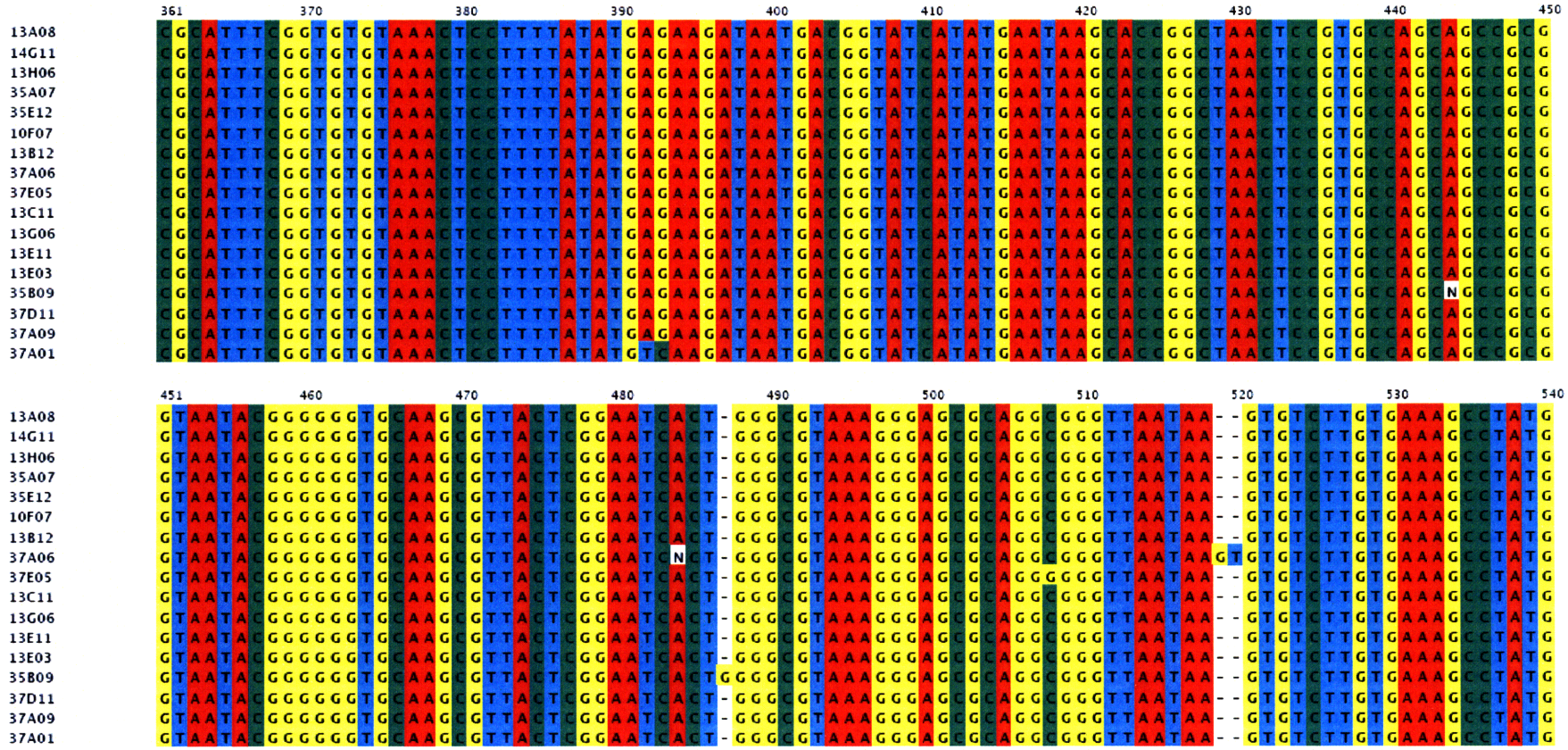


Fig. 24 (cont'd). Multiple sequence alignment showing microdiversity of epsilon-Proteobacteria clones. 17 unique sequence types were found after grouping all 251 epsilon-Proteobacteria clones into 100% OTUs. The dominant OTU (represented by 13A08) contained 226 sequences. Positions with insertions, deletions and SNPs are summarized in Table 6. Bases are presented in different colors to show nucleotide polymorphisms. Positions are consensus alignment positions with gaps and should be used as reference only.

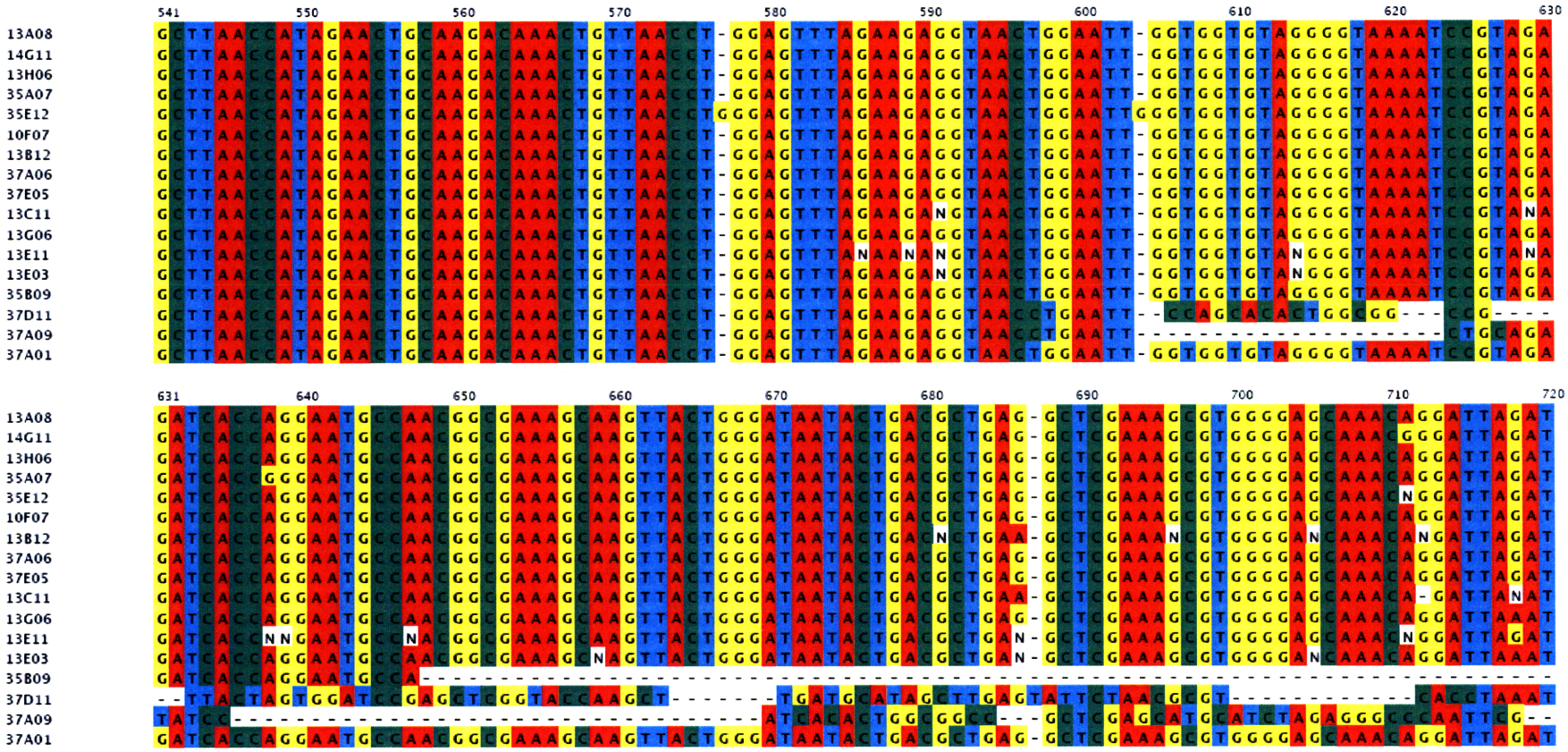


Fig. 24 (cont'd). Multiple sequence alignment showing microdiversity of epsilon-Proteobacteria clones. 17 unique sequence types were found after grouping all 251 epsilon-Proteobacteria clones into 100% OTUs. The dominant OTU (represented by 13A08) contained 226 sequences. Positions with insertions, deletions and SNPs are summarized in Table 6. Bases are presented in different colors to show nucleotide polymorphisms. Positions are consensus alignment positions with gaps and should be used as reference only.

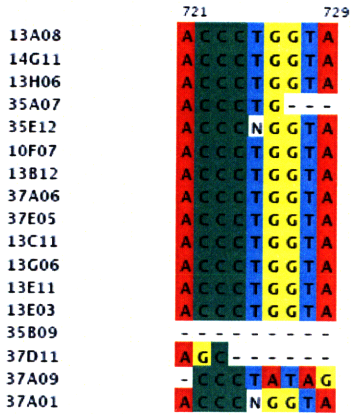


Fig. 24 (cont'd). Multiple sequence alignment showing microdiversity of epsilon-Proteobacteria clones. 17 unique sequence types were found after grouping all 251 epsilon-Proteobacteria clones into 100% OTUs. The dominant OTU (represented by 13A08) contained 226 sequences. Positions with insertions, deletions and SNPs are summarized in Table 6. Bases are presented in different colors to show nucleotide polymorphisms. Positions are consensus alignment positions with gaps and should be used as reference only.

3.5 Phylogenetic relationships of *N. vectensis*-associated bacteria shared in multiple samples to highest homology sequences identified by BLAST

3.5.1 Epsilon-Proteobacteria

Clone libraries from all three sites – Sippewissett (MA), Clinton (CT) and Mahone Bay (Nova Scotia) - have revealed the presence of a novel lineage of epsilon-Proteobacteria ribotype. Its closest uncultured clone in Genbank (Table 7) was an epsilon-Proteobacteria clone (with 96% 16S rRNA identity to our epsilon-Proteobacteria clone) found in Caribbean coral *Montastraea faveolata* with White Plague Disease (Genbank ID: FJ202415, Sunagawa *et al.*, 2009) while the second closest relative was a clone recovered from a Georgia salt marsh (with 93% 16S rRNA identity). The closest cultured relatives (albeit at 85% 16S rRNA identities) are *Helicobacter* and *Arcobacter*, both of which have representative pathogenic species in human gastrointestinal tract (Engberg *et al.*, 2000). The rest of the top 50 BLAST hits consisted of epsilon-Proteobacteria recovered from hydrothermal vent, vent gastropods, mangrove sediments, polluted harbor sediments and sulfide microbial incubator (Fig. 25).

3.5.2 *Pseudomonas*

One *Pseudomonas putida* (19H06) and seven *Pseudomonas pseudoalcaligenes* clones have been found in lab (9G04, 8E06, 19D09, 9F03), Sippewissett (26B06) and Mahone Bay (37B02, 35B03) samples. *P. pseudoalcaligenes* are closely related (with 96% 16S rRNA identity) to the *P. aeruginosa*, an important human pathogen (Yamamoto *et al.*, 2000) (Fig. 26).

3.5.3 *Endozoicimonas*-like gamma-Proteobacteria

This study has found 3 clones from MA (10F05, 14C06, 14D02) and 1 clone from CT (13G11) which have 97% identity at 99% coverage to the 16S rRNA gene of a gamma-Proteobacteria, *Endozoicimonas elysicola*, a symbiont of the sea anemone *Metridium senile* (Table 7). The next closest relative was a gamma-Proteobacteria clone associated with the oyster *Crassostrea ariakensis* (98% identity at 79% coverage). The rest of the top 50 BLAST hits were cultured species from the genera *Oceanospirillum*, *Microbulbifer* and *Sphingobacter* (associated with marine sponges) and uncultured gamma-Proteobacteria associated with invertebrates like cauliflower coral *Pocillopora*, sponge Porifera and sea squirts (colonial ascidian) (Fig. 27).

3.5.4 Spirochaetes

A total of nine Spirochaete clones sharing > 98% 16S rRNA identity, and with only 92% similarity to the closest top hit in Genbank have been found in MA (14A01) and LAB (19F06 [n=6], 19C03 [n=2], 19G03) clone libraries. Their closest hit in Genbank (Table 7) was the uncultured Spirochaete clone 4746K1-41 discovered recently in association with the cold-water Scleractinian coral *Lophelia pertusa* in the Northeastern Gulf of Mexico (Genbank ID: FJ041427, Kellogg *et al.*, 2009). Belonging to the same clade is an uncultured Spirochaete clone retrieved from empty intestine of a shrimp *Pestarella tyrrhena* (Genbank ID: DQ890423, Kormas *et al.* unpublished) (Fig. 28).

3.5.5 Bacteroidetes

Even though specific Bacteroidetes ribotypes with high percentage identities were not found, Bacteroidetes are included in this phylogenetic analysis due to their presence in the *N. vectensis* genome and their association with the sea anemone (*Aiptasia pulchella*). A few members of the phylum Bacteroidetes have been found in association with LAB (9G03, 7H05, 19F11, 19C10), MA (26H04, 14D05, 10E08) and MB (35C11) clone libraries in this study (Fig. 29). Also included in the tree are *Polaribacter (Tenacibaculum)* sp. MED152 found within the *N. vectensis* genome (Starcevic *et al.*, 2008) and *Tenacibaculum aiptasiae*, a novel species of Flavobacteria isolated from a sea anemone (*Aiptasia pulchella*) (Wang *et al.*, 2008).

Table 7. Ribotypes with closest Genbank clones associated with other Hexacorallians

Ribotypes	Top BLAST hit (% identity)	Host	Reference
Epsilon-Proteobacteria	FJ202415: uncultured epsilon-Proteobacteria clone SGUS951 (96%)	Coral *Montastraea faveolata* (with White Plague Disease)	Sunagawa *et al.*, 2009
Endozoicimonas elysicola-like	AB196667: *Endozoicimonas elysicola* (97%)	Sea anemone *Metridium senile* (tentacle lumens)	Schuett *et al.*, 2007
Unknown Spirochaete	FJ041427: Spirochete clone 4746K1-41 (92%)	Cold-water Scleractinian coral *Lophelia pertusa*	Kellogg *et al.*, 2009

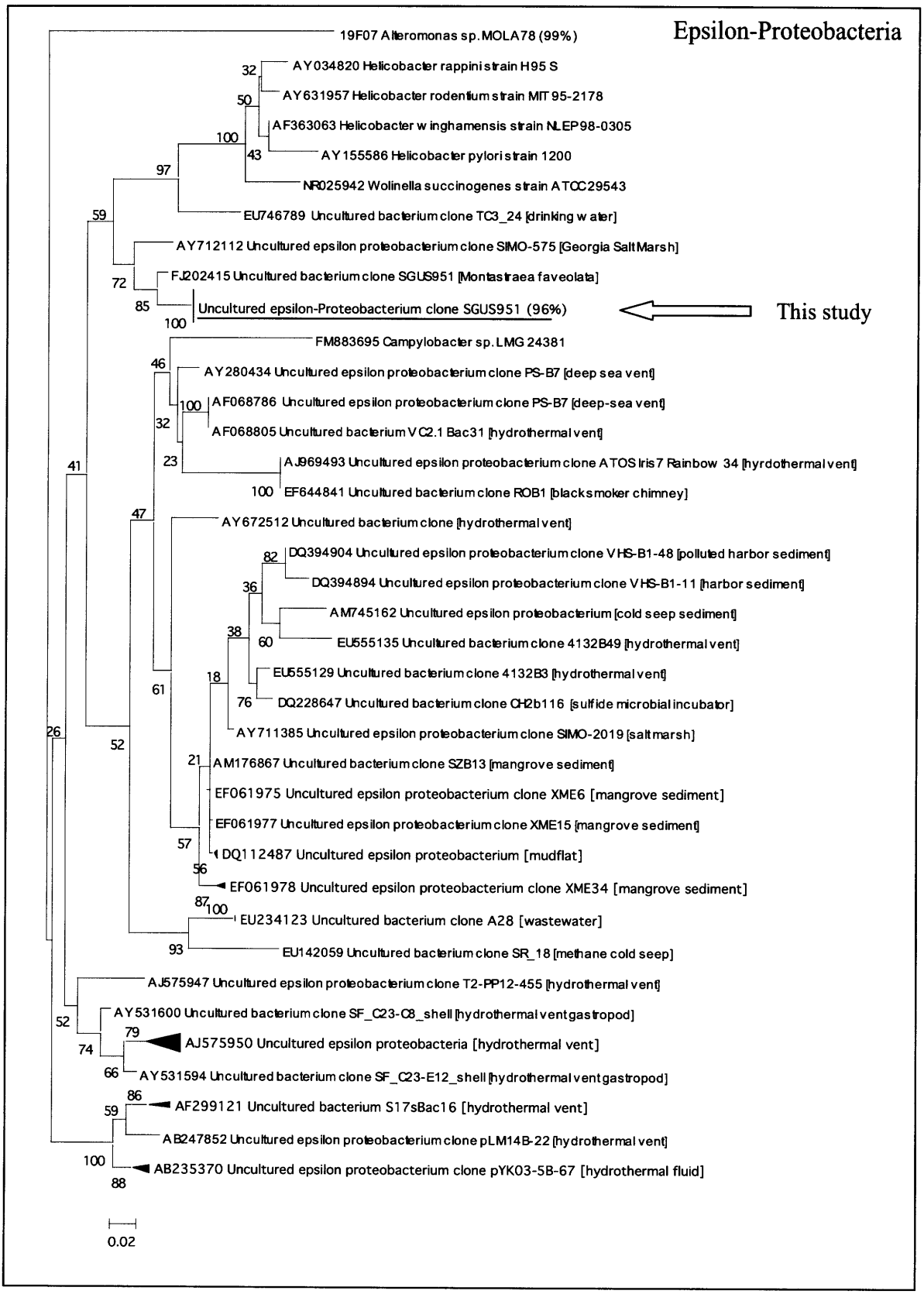

Epsilon-Proteobacteria
19F07 Alteromonas sp. MOLA78 (99%)
AY034820 Helicobacter rappini strain H95 S
AY631957 Helicobacter rodentium strain MIT 95-2178
AF363063 Helicobacter winghamensis strain NLEP98-0305
AY155586 Helicobacter pylori strain 1200
NR025942 Wolinella succinogenes strain ATCC29543
EU746789 Uncultured bacterium clone TC3_24 [drinking water]
AY712112 Uncultured epsilon proteobacterium clone SIMO-575 [Georgia Salt Marsh]
FJ202415 Uncultured bacterium clone SGUS951 [Montastraea faveolata]
Uncultured epsilon-Proteobacterium clone SGUS951 (96%)
This study
FM883695 Campylobacter sp. LMG 24381
AY280434 Uncultured epsilon proteobacterium clone PS-B7 [deep sea vent]
AF068786 Uncultured epsilon proteobacterium clone PS-B7 [deep-sea vent]
AF068805 Uncultured bacterium VC2.1 Bac31 [hydrothermal vent]
AJ969493 Uncultured epsilon proteobacterium clone ATOS Iris7 Rainbow 34 [hyrdothermal vent]
EF644841 Uncultured bacterium clone ROB1 [blacksmoker chimney]
AY672512 Uncultured bacterium clone [hydrothermal vent]
DQ394904 Uncultured epsilon proteobacterium clone VHS-B1-48 [polluted harbor sediment]
DQ394894 Uncultured epsilon proteobacterium clone VHS-B1-11 [harbor sediment]
AM745162 Uncultured epsilon proteobacterium [cold seep sediment]
EU555135 Uncultured bacterium clone 4132B49 [hydrothermal vent]
EU555129 Uncultured bacterium clone 4132B3 [hydrothermal vent]
DQ228647 Uncultured bacterium clone CH2b116 [sulfide microbial incubator]
AY711385 Uncultured epsilon proteobacterium clone SIMO-2019 [salt marsh]
AM176867 Uncultured bacterium clone SZB13 [mangrove sediment]
EF061975 Uncultured epsilon proteobacterium clone XME6 [mangrove sediment]
EF061977 Uncultured epsilon proteobacterium clone XME15 [mangrove sediment]
DQ112487 Uncultured epsilon proteobacterium [mudflat]
EF061978 Uncultured epsilon proteobacterium clone XME34 [mangrove sediment]
EU234123 Uncultured bacterium clone A28 [wastewater]
EU142059 Uncultured bacterium clone SR_18 [methane cold seep]
AJ575947 Uncultured epsilon proteobacterium clone T2-PP12-455 [hydrothermal vent]
AY531600 Uncultured bacterium clone SF_C23-C8_shell [hydrothermal vent gastropod]
AJ575950 Uncultured epsilon proteobacteria [hydrothermal vent]
AY531594 Uncultured bacterium clone SF_C23-E12_shell [hydrothermal vent gastropod]
AF299121 Uncultured bacterium S17sBac16 [hydrothermal vent]
AB247852 Uncultured epsilon proteobacterium clone pLM14B-22 [hydrothermal vent]
AB235370 Uncultured epsilon proteobacterium clone pYK03-5B-67 [hydrothermal fluid]
0.02

Fig. 25

Fig. 25. Maximum likelihood tree of epsilon-Proteobacteria clone found in *N. vectensis* (underlined), its top 50 BLAST hits and cultured characterized epsilon representatives from genera *Helicobacter, Campylobacter* and *Wolinella*, based on 16S rRNA partial sequences *E. coli* positions 104-645. The gamma-proteobacterial sequence Alteromonas sp. MOLA78 is included as an outgroup. Every sequence is labeled by a unique identifier and the identity of the top known cultured BLAST hit. Bootstrap values were shown for branches with >50% support.

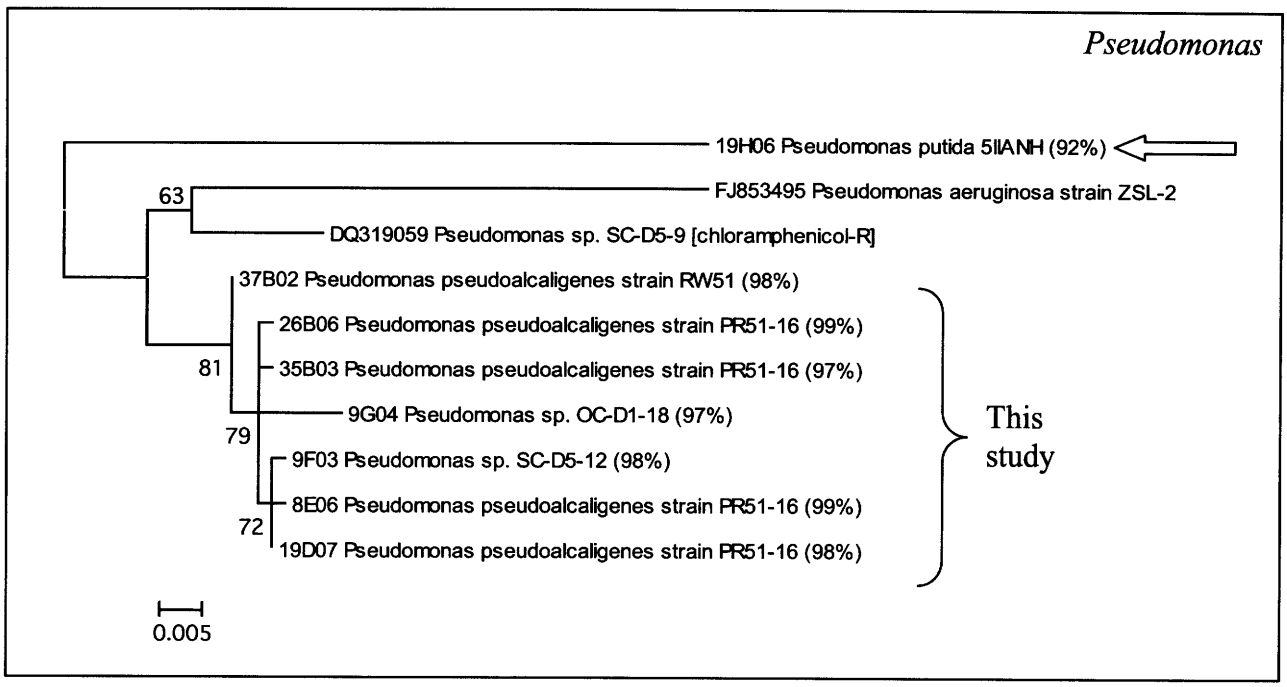


Fig. 26. Maximum likelihood tree of *Pseudomonas* found in this study and from selected literature based on 16S rRNA partial sequences *E. coli* positions 63-695. Every sequence is labeled by a unique identifier and the identity of the top known cultured BLAST hit. Bootstrap values were shown for branches with >50% support.

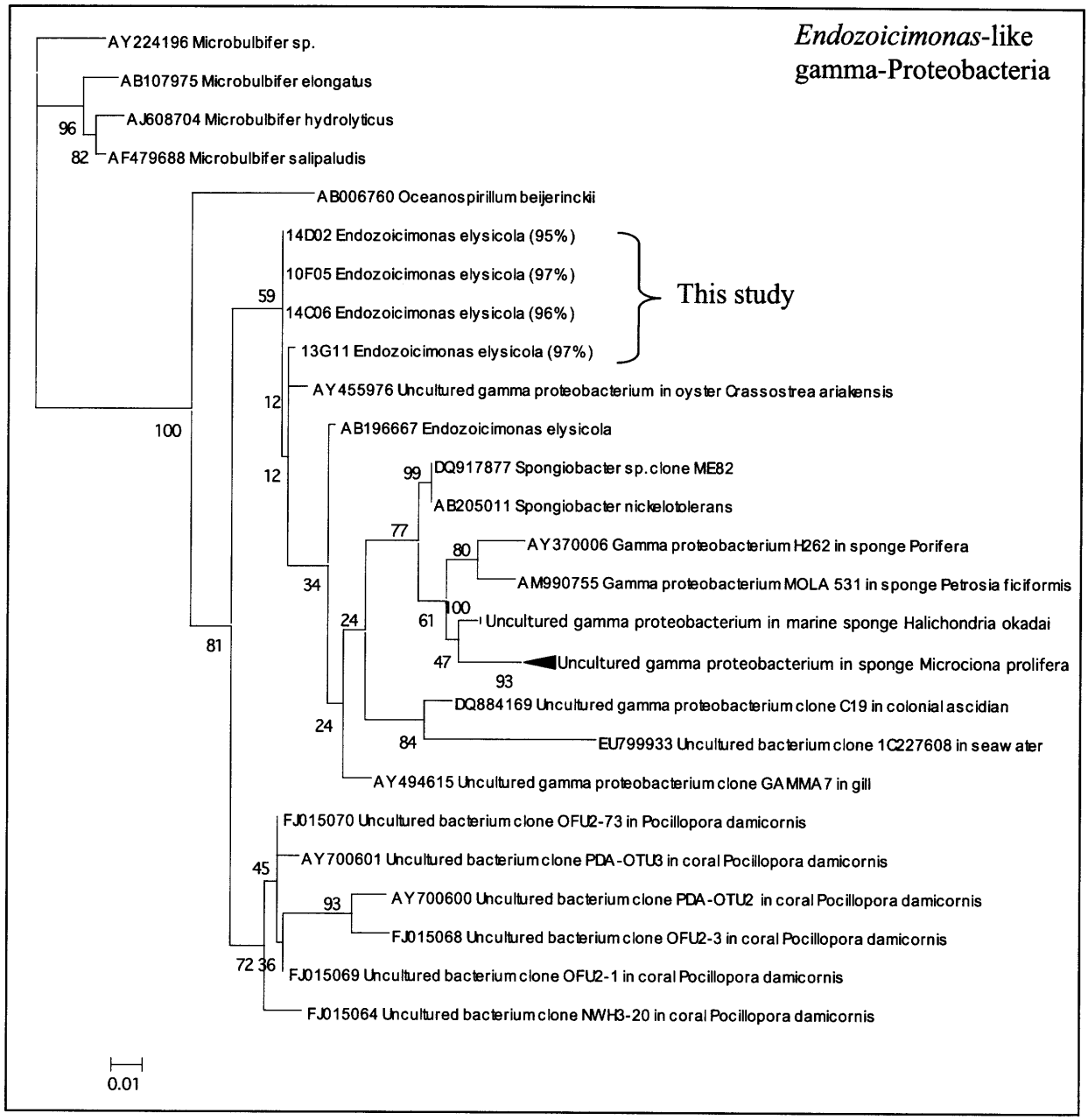


Fig. 27. Maximum likelihood tree of *Endozoicimonas*-like gamma-Proteobacteria found in this study and the top 50 BLAST hits of these clones based on 16S rRNA partial sequences *E. coli* positions 67-588. Every sequence is labeled by a unique identifier and the identity of the top known cultured BLAST hit. Bootstrap values were shown for branches with >50% support.

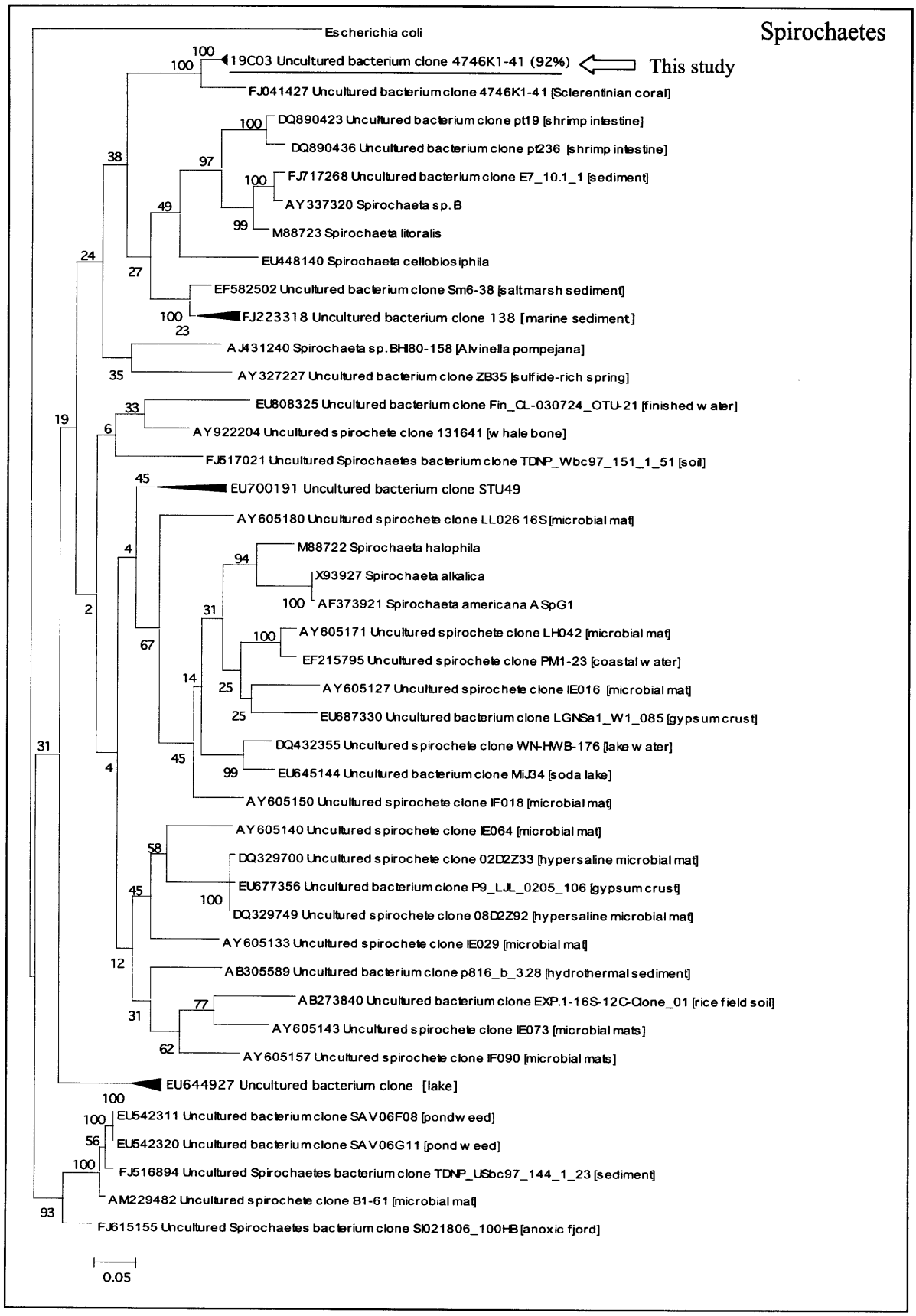

Spirochaetes
Escherichia coli
19C03 Uncultured bacterium clone 4746K1-41 (92%)
This study
FJ041427 Uncultured bacterium clone 4746K1-41 [Sclerentinian coral]
DQ890423 Uncultured bacterium clone pt19 [shrimp intestine]
DQ890436 Uncultured bacterium clone pt236 [shrimp intestine]
FJ717268 Uncultured bacterium clone E7_10.1_1 [sediment]
AY337320 Spirochaeta sp. B
M88723 Spirochaeta litoralis
EU448140 Spirochaeta cellobiosiphila
EF582502 Uncultured bacterium clone Sm6-38 [saltmarsh sediment]
FJ223318 Uncultured bacterium clone 138 [marine sediment]
AJ431240 Spirochaeta sp. BH80-158 [Alvinella pompejana]
AY327227 Uncultured bacterium clone ZB35 [sulfide-rich spring]
EU808325 Uncultured bacterium clone Fin_CL-030724_OTU-21 [finished water]
AY922204 Uncultured spirochete clone 131641 [whale bone]
FJ517021 Uncultured Spirochaetes bacterium clone TDNP_Wbc97_151_1_51 [soil]
EU700191 Uncultured bacterium clone STU49
AY605180 Uncultured spirochete clone LL026 16S [microbial mat]
M88722 Spirochaeta halophila
X93927 Spirochaeta alkalica
AF373921 Spirochaeta americana ASpG1
AY605171 Uncultured spirochete clone LH042 [microbial mat]
EF215795 Uncultured spirochete clone PM1-23 [coastal water]
AY605127 Uncultured spirochete clone IE016 [microbial mat]
EU687330 Uncultured bacterium clone LGNSa1_W1_085 [gypsum crust]
DQ432355 Uncultured spirochete clone WN-HWB-176 [lake water]
EU645144 Uncultured bacterium clone MiJ34 [soda lake]
AY605150 Uncultured spirochete clone IF018 [microbial mat]
AY605140 Uncultured spirochete clone IE064 [microbial mat]
DQ329700 Uncultured spirochete clone 02D2Z33 [hypersaline microbial mat]
EU677356 Uncultured bacterium clone P9_LJL_0205_106 [gypsum crust]
DQ329749 Uncultured spirochete clone 08D2Z92 [hypersaline microbial mat]
AY605133 Uncultured spirochete clone IE029 [microbial mat]
AB305589 Uncultured bacterium clone p816_b_3.28 [hydrothermal sediment]
AB273840 Uncultured bacterium clone EXP.1-16S-12C-Clone_01 [rice field soil]
AY605143 Uncultured spirochete clone IE073 [microbial mats]
AY605157 Uncultured spirochete clone IF090 [microbial mats]
EU644927 Uncultured bacterium clone [lake]
EU542311 Uncultured bacterium clone SAV06F08 [pondweed]
EU542320 Uncultured bacterium clone SAV06G11 [pond weed]
FJ516894 Uncultured Spirochaetes bacterium clone TDNP_USbc97_144_1_23 [sediment]
AM229482 Uncultured spirochete clone B1-61 [microbial mat]
FJ615155 Uncultured Spirochaetes bacterium clone SI021806_100HB [anoxic fjord]
0.05

Fig. 28

Fig. 28. Maximum likelihood tree of Spirochaetes found in this study (underlined) and the top 50 BLAST hits of these clones based on 16S rRNA partial sequences *E. coli* positions 75-805. The *E. coli* sequence is included as an outgroup. Every sequence is labeled by a unique identifier and the identity of the top known cultured BLAST hit. Bootstrap values were shown for branches with >50% support.

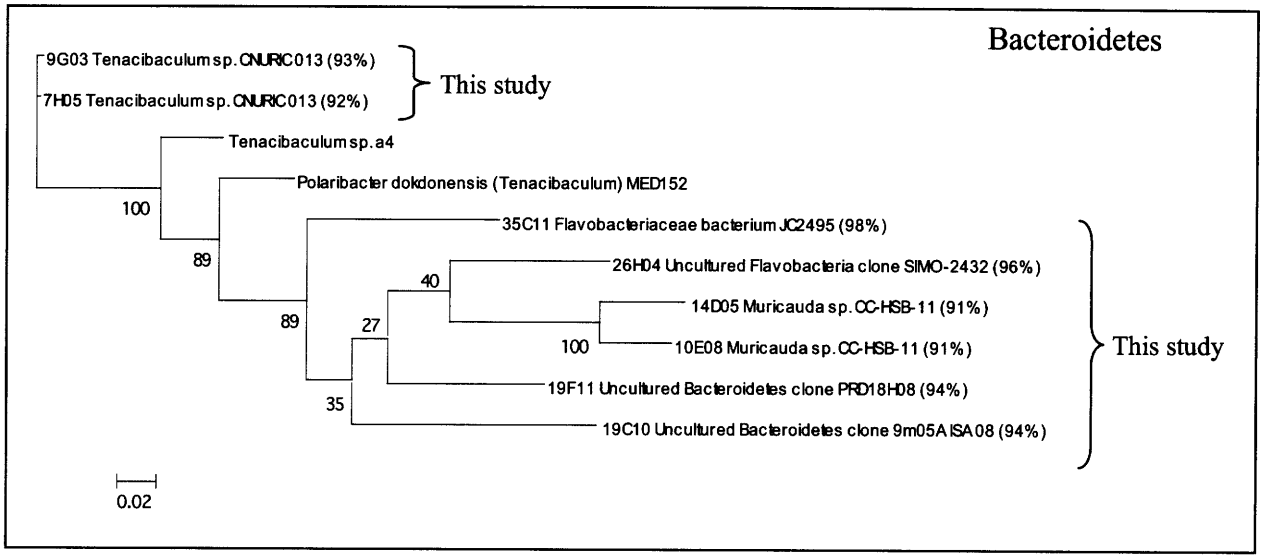


Fig. 29. Maximum likelihood tree of Bacteroidetes found in this study and selected literature based on 16S rRNA partial sequences *E. coli* positions 81-602. Every sequence is labeled by a unique identifier and the identity of the top known cultured BLAST hit. Bootstrap values were shown for branches with >50% support.

3.6 Antibiotic screening of *P. pseudoalcaligenes* isolates

We isolated strains from lab "reduced-flora" *N. vectensis* on antibiotic-containing and regular LB agar. We sequenced the 16S rRNA gene for representative isolates and determined their antibiotic sensitivity profiles. We observed that *Pseudomonas pseudoalcaligenes* was the most abundant colony forming units (cfu) on all antibiotic plates. Of the 32 strains tested, we observed four antibiotic phenotypes - I: $Cm^R/Kn^S/Sm^S/NA^R$ (n = 25), II: $Cm^R/Kn^R/Sm^R/NA^R$ (n = 5), III: $Cm^R/NA^S/Kn^S/Sm^S$ (n = 1) and IV: $Cm^S/NA^S/Kn^S/Sm^S$ (n = 1) (Table 8). To examine whether the variable Kn^R/Sm^R phenotype was plasmid encoded, we screened the low molecular weight plasmid profiles of representative strains. These profiles revealed that *P. pseudoalcaligenes* were associated with plasmids 2 kb, 4 kb and 10 kb but did not show any differences in low molecular weight plasmid profiles correlation to antibiotic resistance phenotype, suggesting this phenotype may be chromosomal or encoded by a high molecular weight plasmid. Since these bacteria were also present in field animals, we now have in collection a powerful tool to test further hypotheses in host selection.

Table 8. *Pseudomonas pseudoalcaligenes* strains isolated from N. *vectensis* homogenate with multiple antibiotic resistance.

Strain number	Strain Origin (antibiotic plate)	Antibiotic phenotype	Cm	Kn	Sm	NA
B04	Cm	I	+	-	-	+
B05	Cm	I	+	-	-	+
B11	Cm	I	+	-	-	+
B12	Cm	I	+	-	-	+
C01	Cm	I	+	-	-	+
C02	Cm	I	+	-	-	+

C05	Kn	I	+	-	-	+
C07	Kn	II	+	+	+	+
C08	Kn	I	+	-	-	+
C09	NA	I	+	-	-	+
C10	NA	I	+	-	-	+
C11	NA	I	+	-	-	+
D09	Sm	I	+	-	-	+
A01	Combo	II	+	+	+	+
A02	Combo	II	+	+	+	+
A03	No Ab	I	+	-	-	+
A05	No Ab	II	+	+	+	+
A06	No Ab	I	+	-	-	+
A07	No Ab	I	+	-	-	+
A09	No Ab	I	+	-	-	+
B01	No Ab	II	+	+	+	+
E06	No Ab	I	+	-	-	+
E08	No Ab	I	+	-	-	+
E10	No Ab	I	+	-	-	+
E11	No Ab	I	+	-	-	+
E12	No Ab	I	+	-	-	+
F02	No Ab	IV	-	-	-	-
F04	No Ab	III	+	-	-	-
F07	No Ab	I	+	-	-	+
F09	No Ab	I	+	-	-	+
F11	No Ab	I	+	-	-	+
F12	No Ab	I	+	-	-	+

"+" denotes presence of colony, i.e. resistance to a certain antibiotic, "-" denotes absence of colony, i.e. sensitivity to the antibiotic. 25 strains were $Cm^R/Kn^S/Sm^S/NA^R$, (Antibiotic phenotype I), 5 were $Cm^R/Kn^R/Sm^R/NA^R$ (Antibiotic phenotype II), and 1 strain each were $Cm^R/NA^S/Kn^S/Sm^S$ (Antibiotic phenotype III) and $Cm^S/NA^S/Kn^S/Sm^S$ (Antibiotic phenotype IV), respectively. Cm: 10 μg/ml chloramphenicol, Kn: 100 μg/ml kanamycin, Sm: 100 μg/ml streptomycin, NA: 4 μg/ml nalidixic acid, Combo: combination of above antibiotics, No Ab: no antibiotic control.

Chapter 4

Discussions

4.1 Comparison between Sippewissett (MA), Clinton Bay (CT) and Mahone Bay (MB) animals reveals shared ribotypes including a novel epsilon-Proteobacteria

16S rRNA clone library analysis of *N. vectensis* collected from Sippewissett Marsh, MA, Clinton Bay, CT and Mahone Bay, Nova Scotia revealed a highly similar microbiome structure dominated by a novel epsilon-Proteobacterial ribotype (at 99% identity threshold, n = 248). The discovery of this epsilon-Proteobacterial population associated with *N. vectensis* at sites separated by large geographical distances (Fig. 2) leads us to hypothesize that it has a specific relationship with the anemone that may be symbiotic. In addition, three more distantly related epsilon-Proteobacteria ribotypes (which returned the same top hit in BLAST search) were also contained in the MB library (Fig. 10). These clones dominated all the field animal clone libraries, making up 33%, 95% and 98% of the MA, CT, and MB clones, respectively.

Epsilon-Proteobacteria are important as primary colonizers, primary producers and symbiotic partners of animals. In many sulfidic habitats with oxic-anoxic interfaces (such as a salt marsh)

they may be the dominant microbes involved in the redox of these compounds (Madrid *et al.*, 2001). Epsilon-Proteobacteria have been found in symbiotic associations with hydrothermal vent metazoans, in salt marshes and other habitats with oxygen and sulfur gradients (Campbell *et al.* 2006). The epsilon-Proteobacteria clone found in field animals may play a similar metabolic role and provide energy and nutrients to their *N. vectensis* hosts. This will be further studied using stable isotope tracing methods. FISH revealed that cells hybridizing to the epsilon-Proteobacteria ribotype-specific probe were associated with the mesentery tissues inside the animal (Fig. 16), supporting the hypothesis that it is an endosymbiont rather than a marsh population associated with the exterior of the animal (Fig. 16). Furthermore, the closest known relative with 96% 16S rRNA similarity is an epsilon-Proteobacteria associated with the Caribbean coral *Montastraea faveolata* (Sunagawa *et al.*, 2009), indicating that this bacterium may have an evolutionarily conserved mechanism of association with hexacorals.

The second closest relative of the dominant epsilon proteobacterial ribotype is a clone recovered from a salt marsh at the Sapelo Island Microbial Observatory in Georgia (93% 16S rRNA identity) – raising the possibility that this lineage may have a wider distribution in other marine or marsh environments. The closest cultured relatives with 85% 16S rRNA identity are the notable human and animal pathogens of the genus *Helicobacter* which contain *H. pylori* that causes gastric ulcers in humans. It is of note that several *Helicobacter* colonize the mammalian gastric epithelium - similar to the localization of our epsilon-Proteobacteria ribotype on the *N. vectensis* mesenteries within the coelenteron (i.e. gastro-vascular cavity).

Comparison of clone libraries of lab and field animals also revealed three additional ribotypes detected in more than one environment. Four clones (> 98% 16S rRNA identity) with 95-97% 16S rRNA identity to *Endozoicimonas elysicola* were discovered in MA and CT animals (Figs. 4 and 27). In addition, Spirochaete sequences sharing > 92% 16S rRNA identity with the uncultured coral-associated Spirochete clone 4746K1-41 (Genbank ID: FJ041427, Kellogg *et al.*, 2009) were found in both MA (n = 1) and LAB clone library (n = 9) (Figs. 4 and 28). A ribotype with 97-99% 16S rRNA identity to *Pseudomonas pseudoalcaligenes* was also present in LAB, MB and MA samples (Figs. 4 and 26). *P. pseudoalcaligenes* has 96% 16S rRNA identity *to P. aeroginosa* - an important human pathogen - and belongs to the *P. aeruginosa* "cluster" in *Pseudomonas* clades (Yamamoto *et al.*, 2000). FISH analysis showed that the distribution of *Pseudomonas* was similar to that of the epsilon-Proteobacteria in the field. Like the epsilon-Proteobacteria, *Pseudomonas* form cell aggregates in lab and field organisms along the animal mesentery and were not found in the outer epithelia (Fig. 17).

AFLP and mtDNA analysis of *N. vectensis* collected along the Atlantic Coast has revealed that Sippewissett, Clinton and Mahone Bay *N. vectensis* populations form distinct geographical clades (Reitzel *et al.* 2008). Thus, differences in the salt marsh microbiome compositions between MA, CT and MB samples may be influenced by the variability in host genetic makeup, or abiotic factors like temperature, salinity and chemical gradients, or they may be random/neutral.

4.2 Comparison between Sippewissett sediment (SED) and animal (MA, MB, CT) clone libraries suggests that *Nematostella vectensis* enriches for epsilon- and gamma-Proteobacteria

16S rRNA clone library analysis of Sippewissett sediment and *N. vectensis* samples from Sippewissett, Clinton and Mahone Bay revealed that the sediment and animal microbial community structures were vastly different from one another. It is important to note that sediment samples from Sippewissett Marsh were collected four months after the animal samples, and thus were exposed to a colder climate (-7° to 23° C) than the *N. vectensis* obtained from Sippewissett Marsh during July 2008 (15° to 37° C). However, the range of daily temperatures during the month of sediment collection (November 2008) was comparable to the range of temperatures during collection of *N. vectensis* from Mahone Bay, Nova Scotia in October 2008 (1° to 19° C) (Table 1).

A wide range of bacteria and plastids were present in the sediment library, including Cyanobacteria/chloroplast (42%), Bacteroidetes (20%), delta-Proteobacteria (8%), gamma-Proteobacteria (8%) closely related to *Chromatiaceae* and *Haliea*, Planctomycetes (8%), Chloroflexi (6%) and a lower abundance of other groups (Fig. 7). The high proportion of Cyanobacteria and chloroplast sequences in sediments is not surprising as salt marshes are among the most productive environments with primary production rates ranging from 460 to 3700g cm^{-2} $year^{-1}$ (Gallagher *et al.,* 1980) due to the abundance of Spartina marsh grass, photosynthesizing cyanobacteria, eukaryotic algal chloroplast and high diversity of anoxygenic

phototrophic bacteria including green-sulfur, green non-sulfur (e.g. Chloroflexi), purple-sulfur, purple non-sulfur bacteria.

Sulfate reduction is the primary respiratory pathway in the Sippewissett marsh, accounting for the consumption of 1.8 kg organic C m^{-2} $year^{-1}$ (Howarth, Teal, 1979). Sulfate-reducing bacteria (SRB) of the family Desulfobacteriaceae including *Desulfosarcina variabilis* (clone 40F01 found in SED library) which can oxidize acetate and other fatty acids (Devereux, Stahl, 1993) are the most abundant SRB in salt marshes (Devereux *et al.*, 1996) and can comprise up to 40% of bacterial rRNA associated with Spartina roots (Hines *et al.*, 1999). The relatively high proportion of other delta-Proteobacteria (clones 40A02, 40A04, 40B05, 40F02, 40F04) and Chromatiaceae (clones 40B08, 40G05) - a purple sulfur gamma-Proteobacteria that can oxidize sulfide to sulfate - are also expected to be part of the sulfur biogeochemical cycle of the marsh.

On the other hand, the microbial community structure of *N. vectensis* collected from the different marsh environments was much less diverse than that of the sediment, as indicated by Chao 1 mean values that were <35% that of the sediment. The proportion of Cyanobacteria/chloroplasts sequences in the clone library of *N. vectensis* collected from the Sippewissett marsh was similar at about 40%. Bright field microscopy of live specimens revealed that Sippewissett Marsh *N. vectensis* (collected November 2008) were associated with detritus containing several morphotypes of algal cells including some diatoms (J. Thompson, personal communication). These detrital algae may be the source of chloroplast sequences observed in the clone libraries. Diatom or other algal cells were not observed to be associated with the lumens or mesenteries of

field *N. vectensis* – supporting the localization of chloroplast sequences to the exterior of the organism

Clone library analyses revealed that the previously unknown epsilon-Proteobacteria ribotype associated with Sippewissett, Clinton and Mahone Bay *N. vectensis* was not found in the sediment clone library. It was surprising that despite the similar sequencing coverage of all libraries, this dominant epsilon-Proteobacteria ribotype in the *N. vectensis* libraries was completely absent from the sediment library. PCR primers designed to be specific for this epsilon-Proteobacteria ribotype were used to screen the microbiome of in *Tubifex* worms (a prey of *N. vectensis*) obtained from Sippewissett sediments. This screen only yielded non-specific amplification products that were identified as Vibrios by sequencing (data not shown), suggesting that this lineage of epsilon-Proteobacteria is not ubiquitous in sediment invertebrates.

In addition, the animal-associated gamma-Proteobacteria members were significantly different from those in the sediment, with the exception of one Chromatiaceae clone. Like the epsilon-Proteobacteria ribotype, the gamma-Proteobacteria ribotypes (including *Endozoicimonas elysicola* and *Pseudomonas pseudoalcaligenes*) found in Sippewissett, Clinton and Mahone Bay animal libraries were also absent from the sediment library. These groups have been previously reported to be associated with marine invertebrates: *E. elysicola* was reported to form bacteria aggregates within the tentacle lumens of the sea anemone *Metridium senile* (Schuett *et al.*, 2007), *V. harveyi* is commonly associated with cnidarians (Chimetto *et al.*, 2009) and *Pseudoalteromonas* are generally found in association with marine eukaryotes and can display anti-bacterial, bacteriolytic, agarolytic and algicidal activities compounds promoting the survival

of other marine organisms (Holmstrom, Kjelleberg, 1999). The unknown Spirochete with 92% 16S rRNA identity to clone 4746K1-41 found in MA library was also missing from the sediment library.

Interestingly, delta-Proteobacteria which formed 8% of the sediment library (including the SRBs) were not found among the animal-associated clones. Other groups that were under-represented in animal tissue compared to the sediment library were alpha-, beta-Proteobacteria, Chloroflexi, Planctomycetes and Verrucomicrobia, suggesting that *N. vectensis* does not enrich these groups.

Comparison between sediment and animal clone libraries provided the first evidence that ecological niches associated with the host – such as resource supply or host immune interactions - may play a role in shaping the *N. vectensis* microbiome. The difference in microbial community between Sippewissett sediment and Mahone Bay *N. vectensis* may be explained to some extent by geographic separation. The time difference and corresponding seasonal shifts in the marsh may explain some differences between the Sippewissett sediment and Sippewissett *N. vectensis* clone libraries. This will be addressed by construction of a clone library (MA-II) with Sippewissett animals obtained at the same time and place (November 2008) as sediment collection. This was not done within the scope of this thesis due to time constraints.

4.3 Comparison between laboratory-raised *N. vectensis* and field animals suggests the laboratory environment favors alpha-, gamma-Proteobacteria and Spirochaetes but not epsilon-Proteobacteria

Clone library analyses revealed substantial differences between lab-raised and field *N. vectensis* microbiomes that may be related to their environment conditions. Compared to field animals, which are subjected to huge fluctuations in temperature, salinity, and food availability over day-night cycles and different seasons, lab animals live in a relatively stable environment and were at least twice bigger in size (~1 cm) than their wild counterparts. One remarkable observation was that the epsilon-Proteobacteria, which constituted the dominant clones in field animals, were missing from the lab library. Cyanobacteria/chloroplast clones that made up 40% of clones in Sippewissett animals were also almost absent from lab animals. The exception was a ribotype with homology to Cyanobacteria clone LW18m-1-76 (Genbank ID: EU642239). FISH hybridization suggested that the epsilon-Proteobacteria may still occur at low abundance in the mesenteries of lab animals, so the absence from lab clone library may be due to incomplete sampling of sequence richness. Further analysis using quantitative FISH will help to elucidate the relationship between epsilon-Proteobacteria and lab-raised *N. vectensis*. Nonetheless, we can conclude that epsilon-Proteobacteria was at least greatly decreased in lab animals.

The possible reasons for epsilon-Proteobacteria decline under lab conditions are multifold. Many epsilon-Proteobacteria are chemoliautothotrophic and harness energy from redox gradients that form where hydrogen sulfide from the anoxic mud meets oxygenated seawater (e.g. in salt marshes) and may be the dominant microbes involved in the recycling of these compounds

(Madrid *et al.*, 2001). However, when animals were transferred to the lab environment where the saline solution was always well aerated the absence of a redox gradient may place the epsilon-Proteobacteria at a competitive disadvantage. Moreover, the lab animals were fed at regular intervals so there was no lack of nutrient and energy sources, thereby negating the need for a potential chemoautotrophic endosymbiont to provide these resources. Another possible reason for the greatly diminished epsilon-Proteobacteria populations is that they were out-competed for space or resources by the gamma-Proteobacteria which rose to dominance in the lab animals - a surge in gamma-Proteobacteria from 15% to 72% was observed.

The gamma-Proteobacteria in the lab clone library were largely different from those in the field with the exception of 7 *Pseudomonas pseudoalcaligenes* clones that were also found in Sippewissett and Mahone Bay samples and one *Vibrio* with 95% 16S rRNA identity to those found in Sippewissett (Fig. 11). 44% of the gamma-Proteobacteria in the lab *N. vectensis* clone library were Alteromonas clones with > 93% similarity to Alteromonas clones obtained from deep sea sediment (strain MOLA 78, Genbank ID: AM990853, Larcher *et al.* unpublished) and coastal area (Strain CF6-3, Genbank ID: FJ169998, Zhang *et al.* unpublished), and 35% were *Thalassomonas* sp. clones with > 93% similarity to a *T. agarivorans* strain obtained from deep sea sediment (strain JAMB A24, Genbank ID: AB304803) (Miyazaki *et al.*, 2008). Experimental perturbations of the *N. vectensis* microbiome with the antibiotics chloramphenicol, kanamycin, streptomycin and/or nalidixic acid revealed that these *P. pseudoalcaligenes* isolates were stably associated with the *N. vectensis* microbiome across changing environmental conditions, surviving both direct effects of the various antibiotics or indirect effects caused by changes in other co-occurring microbial populations. Moreover, *P. pseudoalcaligenes* displayed multiple

antibiotic phenotypes (Table 8) suggesting that there was some strain variation within the population, and this variation may help the population adapt to changing environments. The dominant antibiotic resistance phenotype was $Cm^R/Kn^S/Sm^S/NA^R$ (n = 25/31), followed by $Cm^R/Kn^R/Sm^R/NA^R$ (n = 5/31) and $Cm^R/Kn^S/Sm^S/NA^S$(n = 1/31) (Table 8). These are consistent with a previous study that *P. pseudoalcaligenes* were the dominant chloramphenicol-resistant bacteria in Jiaozhou Bay, China, even though the *P. pseudoalcaligenes* did not contain chloramphenicol acetyltransferase genes (Dang, *et al.* 2008).

Closely-related Spirochaetes with > 98% 16S rRNA identity were found in both lab animals (n = 9) and Sippewissett field animals (n = 1) representing 10% and 0.9% of the LAB and MA libraries, respectively. Besides, six alpha-Proteobacteria clones from the families Kiloniellaceae, Rhodobacteraceae, Rhizobiaceae and Kordiimonadaceae were found in the lab clone library but were absent from field clone libraries.

The rise of these gamma-, alpha-Proteobacteria and Spirochaetes may be related to the decline of epsilon-Proteobacteria. It is hard to ascertain at this point whether competition posed by the former groups or the decline of the latter group triggered the shift in microbial community structure. The observed community shift between field and reduced-flora lab animal microbiomes may have been affected by microbial populations introduced into the microbiome from the laboratory environment during their initial maintenance under non-sterilized conditions. One way to test how the microbiome responds to this shift from the field to the laboratory environment will be to transplant *N. vectensis* individuals directly from field into sterile lab conditions (i.e. incubation in sterilized saline and feeding with sterilized *Artemia* nauplii) and to

use cultivation independent analysis to monitor the shift in community structure over time as the microbiome adapts to the new environment.

4.4 Fluorescent in situ hybridization (FISH)

4.4.1 Physical location of bacteria in *Nematostella vectensis*

FISH analysis with Cy3-EUB338 and FAM-labeled probes specific to the epsilon-Proteobacteria ribotype found in this study and the *Pseudomonas* genus revealed that bacteria in both lab-raised and wild *N. vectensis* formed aggregates of <10 to >100 cells each in close association with *N. vectensis* mesenteries (Figs.19 and 20). That no cells were seen in other parts of the animals (e.g. mouth, lumen, tentacles and external tissue) suggests that the association is specific to the mesentery tissues and may be mediated by receptors present on the host mesentery. Interestingly, this is different from a previous study done on a sea anemone *Aiptasia pallida* (Palincsar *et al.*, 1989) where *Vibrio*-like bacteria aggregates were most common at the tentacles and oral disc, and at lower density throughout the animal column; and from another study (Schuett *et al.*, 2007) of sea anemone *Metridium senile* where *Endozoicimonas elysicola*- and *Pseudomonas saccherophilia*-like bacteria were found in bacteria aggregates in the tentacles. One possible reason may be different recognition mechanisms employed by these sea anemone species in their host-microbe interaction. Further investigation is needed to provide an explanation to this intriguing observation.

We also noted that both epsilon-Proteobacteria and *Pseudomonas* cells in aggregates within the animal mesenteries seemed to be cocci (i.e. spherical in shape) while *Pseudomonas* cells grown

in liquid culture are rod-shaped. The spherical geometry may be a way to minimize surface area required per cell as the symbionts compete for limited space on the mesenteries. It may also be a response to host signaling factors. Investigating the recognition mechanisms and cross-talk involved in animal-bacteria interaction will be the subject of future studies.

Quantitative FISH studies of *N. vectensis* microsections have also been initiated and are currently in progress. Preliminary data suggests that the ratio of epsilon-Proteobacteria probe-positive cells to total bacteria cells (i.e. the proportion of epsilon-Proteobacteria) was 54% in MA field animals, while that of lab animals was 20% (n=1 for each). The proportion of epsilon-Proteobacteria cells obtained through FISH (54%) was similar to the proportion of ribotypes in the MA clone library (55%) after removing chloroplasts clones suspected to be associated with detritus on the outside of the animal. In contrast the number of epsilon-Proteobacteria probe-positive cells seems to be higher in the laboratory *N. vectensis* than we expected as we did not find any epsilon-Proteobacteria sequences in the LAB clone library. Under-sampling of lab diversity may help explain this disparity. Alternatively the FISH probe may be binding non-specifically to non-target populations in the lab, although *in silico* analysis (Figs. 22 and 23) suggests it is highly specific to our lab clone and whole cell hybridization indicates it does not bind to *Pseudomonas* under the hybridization conditions used. Whether epsilon-Proteobacteria are also present in the lab-raised animals remains an open question under investigation. However, it is clear that there is less epsilon-Proteobacteria in the lab than in the field.

Likewise from FISH cell enumeration, *Pseudomonas*-positive cells made up 46% and 53% of MA and lab bacteria cells, respectively. These figures were also higher than expected,

suggesting there may be a higher prevalence of *Pseudomonas* associated with *N. vectensis* than suggested by the relative abundance of sequence types in the clone libraries. Additional biological replications and probe-specificity checks are needed to confirm these numbers.

4.4.2 Estimations of number of bacteria in *N. vectensis*

The number of bacteria cells in *N. vectensis* was estimated to be 1.8×10^6 per animal (section 3.3.2). Considering that the wet weight of an adult animal from the lab is ~ 10 μg (Harter, Matthews, 2005), we can estimate that there are about 2×10^{11} bacteria cells per gram animal tissue. This is about 100 times the density of bacteria in soil (10^9 bacteria/gram soil). As seen from FISH analysis, bacteria cells are tightly packed in the animal mesentery and this may allow the high density of bacteria to persist within the host.

Since eukaryotic cells are 10 times bigger in diameter than prokaryotic cells, the volume – and therefore the mass, assuming they have similar density - of a eukaryotic cell will be 1000 times higher than that of a typical prokaryotic cell. It has been found that the mass of an *E. coli* cell is about 100 femtogram (10^{-13}g) (Burg *et al.*, 2007). The mass of a eukaryotic cell (e.g. a cnidarian cell) will be about 10^{-10}g. Given the mass of an adult *N. vectensis* of about 10 μg, a typical individual organism would have 10^5 body cells. As we have estimated that each animal has ~1.8×10^6 bacteria cells, the cnidarian cell-to-microbe ratio will be roughly 10:1, the same as the ratio of cells in the human microbiome to cells of human tissue.

4.5 Microdiversity of Epsilon-Proteobacteria across different field sites

The epsilon-Proteobacteria clones found in all field animal libraries were examined in closer detail. After manually correcting the sequences for miscalled bases and excluding all ambiguities, it was found that there were four distinct 99% OTUs – the dominant OTU with 248 out of 251 sequences was present across Sippewissett, Clinton and Mahone Bay sites, while the other three (37A01, 37A09, 37D11) were only found once each in Mahone Bay. These Mahone Bay clones were significantly different from the dominant cluster as shown in the alignment in Fig. 10.

While the A → G change may be attributed to amplification error by Taq polymerase (Bracho *et al.*, 1998), the rest of the SNPs and indels may represent real microdiversity. It is interesting to see certain location-specific SNPs, such as the G → C substitution found only in MB at position 788 and C→ T in CT at position 399. The observed SNPs may represent geographically isolated populations that arose due to accumulation of random mutation during clonal diversification. On the other hand, they may represent distinct ecological populations occupying hosts from different niches. Population genetic analysis of *N. vectensis* reveals location-specific clades of mtDNA haplotypes (Reitzel *et al.*, 2008). Moreover, since the 16S rRNA is a nearly neutral and slowly evolving gene, selective sweeps due to adaptive mutation in the epsilon-Proteobacteria genome would have to occur before variation is seen at the 16S rRNA level (Palys *et al.*, 1997). The observation of location specific microdiversity supports the proposal that the epsilon-Proteobacteria is an endosymbiont of *N. vectensis* and that the association may be ancient

enough to allow for geographically differentiated *N. vectensis*-specific epsilon-Proteobacterial populations.

4.6 Common bacteria ribotypes found in *N. vectensis* from different locations

The presence of common bacteria ribotypes in *N. vectensis* individuals collected from sites that are separated by large geographical distances is suggestive of a role for host selection pressure in maintaining these groups in the microbiome. Alternative explanations for the finding of similar bacteria in multiple sites may include ingestion of bacteria associated with common prey across habitats and the presence of common populations that colonize the *N. vectensis* microbiome in different salt marshes. However, the former explanation does not hold as *P. pseudoalcaligenes* found in association with field *N. vectensis* persisted in lab-raised animals after the feeding regime shifted to an *Artemia* diet even after antibiotic perturbations. The latter hypothesis was not supported by the comparison between sediment and field *N. vectensis* clone libraries. However to fully address this we will construct clone library (MA-II) of animals obtained at the same time and location as sediment collection. We will also screen for ribotypes discovered in this study – the epsilon-Proteobacteria, *P. pseudoalcaligenes*, the Spirochaetes and *Endozoicimonas elysicola* - in the sediment using ribotype-specific PCR primers.

4.7 Symbiont acquisition mechanism of *N. vectensis*

It is noted that the sediment clone library does not contain the *N. vectensis*-associated ribotypes discovered in this study. Despite the presence of oxygen and sulfide gradients that select for epsilon-Proteobacteria, the dominant symbiotic epsilon-Proteobacteria ribotype in field animals in Sippewissett, Clinton and Mahone Bay – or any of its distantly related relatives found in other salt marshes and hydrothermal vents - were absent in the sediment clone library. Likewise, no member of the genus *Pseudomonas* was found the sediment clone library. The *P. pseudoalcaligenes* ribotype was only observed in the Sippiwessett (and Mahone Bay) and lab animals. The Spirochaetes ribotypes with >98% identity too were absent from the sediment library, although a Spirochaete clone with 90% 16S rRNA identity to the *N. vectensis*-associated ribotype was present. The *Endozoicimonas elysicola*-like ribotype associated with Sippewissett and Clinton *N. vectensis* was also not found in the sediment library.

Taken together, these suggest that the *N. vectensis*-associated ribotypes were not acquired horizontally from abundant populations in the environment. The symbiont transmission mode (i.e. vertical vs. horizontal) can be further investigated by histological observation and 16S rRNA screening of surface-sterilized gametes and larvae. If gametes are preloaded with certain microbial symbionts by the parents (as in the case of *Hydra*), we can conclude that these microbes are vertically-transmitted. The developmental progress and microbial community of fertilized gametes can also be tracked over time to detect any acquisition of new microbial partners through horizontal transmission.

Chapter 5

Conclusions

This study represents the first investigation into the microbiome of the starlet anemone *Nematostella vectensis*, a basal eumetazoan fast becoming the foremost Cnidarian laboratory model. Comparison between 16S ribosomal RNA gene clone libraries of Sippewissett sediment and field animals collected from Sippewissett Marsh (Massachusetts), Clinton Bay (Connecticut) and Mahone Bay (Nova Scotia) revealed that the microbiome of field animals is largely different from that of the sediment. In all three field sites, *N. vectensis* microbiome clone libraries were dominated by a previously unknown ribotype of epsilon-Proteobacteria. 248/251 of these epsilon-Proteobacteria clones share > 99% 16S rRNA identity and display location-specific nucleotide polymorphisms, suggesting that they may be stably associated endosymbionts that have lived with the host over evolutionary timescales. Gamma-Proteobacteria such as *Pseudomonas pseudoalcaligenes*, *Endozoicimonas elysicola,* an uncultured Spirochaete clone and some Bacteroidetes distantly related to each other also had representatives in multiple samples but were absent in the sediment, suggesting that these too may be specifically-associated with the host. The majority of *N. vectensis*-associated 16S rRNA clones share the highest level of sequence identity to Hexacorallia-associated clones in the NCBI Genbank database, further supporting the hypothesis that these are specific associates of hexacorals and affirming the use of *N. vectensis* as a model to study hexacoral-microbe interactions.

Moreover, this work has shown that microbial communities conserved in the wild can undergo significant change under laboratory conditions. The clone library of the reduced-flora animals was significantly different from those of the field animals, mainly due to the loss of the epsilon-Proteobacteria and the abundance of gamma-Proteobacteria from genera *Thalassomonas, Alteromonas* and *Pseudomonas*, including the *P. pseudoalcaligenes* clones. Perturbations of the *N. vectensis* laboratory habitat with different antibiotics revealed persistent *P. pseudoalcaligenes* strains suggesting that these populations can adapt to remain in association with the host over a wide range of environmental conditions.

Lastly, fluorescent in situ hybridization (FISH) analysis showed that bacteria including epsilon-Proteobacteria and *Pseudomonas* form cell aggregates that are tightly associated with *N. vectensis* mesentery tissues, suggesting that the host-microbe interaction is tissue-specific and may be receptor-mediated, as opposed to neutral or random.

Here we conclude that *N. vectensis* is a suitable model for laboratory and field studies of hexacoral-microbe interactions. We anticipate that future work with *N. vectensis* will shed light on the molecular determinants mediating symbiotic community assembly and activity of hexacoral-associated microbial populations. We hope that such work will ultimately help us to understand how beneficial microbes influence the balance between health and disease in hexacorals including the reef-building corals currently threatened by global extinction.

References

Altschul SF, Madden TL, Schaffer AA, *et al.* (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389-3402.

Aragone MR, Maurizi DM, Clara LO, Navarro Estrada JL, Ascione A (1992) *Pseudomonas mendocina*, an environmental bacterium isolated from a patient with human infective endocarditis. J Clin Microbiol 30, 1583-1584.

Ashelford KE, Weightman AJ, Fry JC (2002) PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. Nucleic Acids Res 30, 3481-3489.

Bellwood DR, Hughes TP, Folke C, Nystrom M (2004) Confronting the coral reef crisis. Nature 429, 827-833.

Bode PM, Bode HR (1980) Formation of pattern in regenerating tissue pieces of Hydra *attenuata*. I. Head-body proportion regulation. Dev Biol 78, 484-496.

Bosch TCG (2003) Ancient signals: peptides and the interpretation of positional information in ancestral metazoans. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 136 (2):185-196

Bosch TCG (2008) The Path Less Explored: Innate Immune Reactions in Cnidarian. H. Heine (ed.), Innate Immunity of Plants, Animals, and Humans. Nucleic Acids and Molecular Biology 21. Springer-Verlag Berlin Heidelberg. pp. 27-42

Bosch TCG, Augustin R *et al.* (2009) Uncovering the evolutionary history of innate immunity: The simple metazoan Hydra uses epithelial cells for host defence. Developmental & Comparative Immunology. 33 (4): 559-569

Bourne DG, Munn CB (2005) Diversity of bacteria associated with the coral *Pocillopora damicornis* from the Great Barrier Reef. Environ Microbiol 7, 1162-1174.

Bracho MA, Moya A, Barrio E (1998) Contribution of Taq polymerase-induced errors to the estimation of RNA virus diversity. J Gen Virol 79 (Pt 12), 2921-2928.

Burg TP *et al.* (2007) Weighing of biomolecules, single cells and single nanoparticles in fluid. Nature 446, 1066-1069

Campbell BJ, Engel AS, Porter ML, Takai K (2006) The versatile epsilon-proteobacteria: key players in sulphidic habitats. Nat Rev Microbiol 4, 458-468.

Castaneda O, Sotolongo V, Amor AM, *et al.* (1995) Characterization of a potassium channel toxin from the Caribbean Sea anemone Stichodactyla helianthus. Toxicon 33, 603-613.

Chimetto LA, Brocchi M, Thompson CC, *et al.* (2008) Vibrios dominate as culturable nitrogen-fixing bacteria of the Brazilian coral Mussismilia hispida. Syst Appl Microbiol 31:312-9.

Chimetto LA, Brocchi M, Gondo M, *et al.* (2009) Genomic diversity of vibrios associated with the Brazilian coral *Mussismilia hispida* and its sympatric zoanthids (*Palythoa caribaeorum, Palythoa variabilis* and *Zoanthus solanderi*). J Appl Microbiol.

Colwell RK, Coddington JA (1994) Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Lond B Biol Sci 345, 101-118.

Colwell RK, Rahbek C, Gotelli NJ (2004) The mid-domain effect and species richness patterns:what have we learned so far? Am Nat 163, E1-23.

Cottrell MT, Kirchman DL (2000) Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. Appl Environ Microbiol 66, 1692-1697.

Dang H *et al.* (2008) Dominant chloramphenicol-resistant bacteria and resistance genes in coastal marine waters of Jiaozhou Bay, China. World J Microbiol Biotechnol 24:209–217

Darling JA, Reitzel AM, Finnerty JR (2004) Regional population structure of a widely introduced estuarine invertebrate: *Nematostella vectensis* Stephenson in New England. Mol Ecol 13, 2969-2981.

Darling JA, Reitzel AR, Burton PM, *et al.* (2005) Rising starlet: the starlet sea anemone, *Nematostella vectensis*. Bioessays 27, 211-221.

DeLong EF *et al.* (1993) Phylogenetic Diversity of Aggregate-Attached vs. Free-Living Marine Bacterial Assemblages Limnology and Oceanography, Vol. 38, No. 5 pp. 924-934

Devereux R,Stahl DA (1993) Phylogeny of sulfate-reducing bacteria and a perspective for analyzing their natural communities. - The Sulfate-Reducing Bacteria: Contemporary Perspectives, Springer-Verlag.

Devereux R, Hines ME, Stahl, DA (1996). S cycling: characterization of natural communities of sulfate-reducing bacteria by 16S-rRNA sequence comparisons. Microb. Ecol. 32: 283–292

Ducklow HW, Boyle PJ, Maugel PW, Strong C, Mitchell R (1979) Bacterial flora of the schistosome vector snail Biomphalaria glabrata. Appl Environ Microbiol 38, 667-672.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32, 1792-1797.

Engberg J *et al.* (2000) Prevalence of Campylobacter, Arcobacter, Helicobacter, and Sutterella spp. in Human Fecal Samples as Estimated by a Reevaluation of Isolation Methods for Campylobacters. Journal of Clinical Microbiology 38(1): 286-291

Fallowski PG, Dubinsky Z, Muscatine L. & Porter JW (1984) Light and the bioenergetics of a symbiotic coral. Bioscience 34, 705–709.

Fautin, D. G. and R. N. Mariscal. (1991) Cnidaria: Anthozoa. F. W. Harrison and J. A. Westfall (eds.), Microscopic Anatomy of Invertebrates, volume 2: Placozoa, Porifera, Cnidaria, and Ctenophora. Wiley-Liss, Inc., New York. pp. 267-358

Ferrer, L. M. & Szmant, A. M. (1988) Nutrient regeneration by the endolithic community in coral skeletons. Proc. 6th Int. Coral Reef Symp. Australia 3, 1–4

Finnerty JR *et al.* (2004) Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. Science 304, 1335−1337

Finnerty JR, Martindale MQ (1997) Homeoboxes in sea anemones (Cnidaria:Anthozoa): a PCR-based survey of Nematostella *vectensis* and Metridium senile. Biol Bull 193, 62-76.

Frank PG and Bleakney JS (1978). Asexual reproduction, diet, and anomalies of the anemone Nematostella *vectensis* in Nova Scotia. Canadian Field Naturalist 92, 259-263.

Frank U, Leitz T, Muller WA (2001) The hydroid Hydractinia: a versatile, informative cnidarian representative. Bioessays 23, 963-971.

Fraune S, Bosch TCG (2007) Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. PNAS 104: 13146-13151.

Gallagher JL, Reimold RJ, Linthurst RA and Pfeiffer WJ (1980) Aerial Production, Mortality, and Mineral Accumulation-Export Dynamics in Spartina Alterniflora and Juncus Roemerianus Plant Stands in a Georgia Salt Marsh. Ecology, Vol. 61, No. 2, pp. 303-312

Garcia-Valdes E, Cozar E, Rotger R, Lalucat J, Ursing J (1988) New naphthalene-degrading marine *Pseudomonas* strains. Appl Environ Microbiol 54, 2478-2485.

Glynn PW (2006) Coral reef bleaching: facts, hypotheses and implications Global Change Biology 2(6): 495 - 509

Guerrero R, Berlanga M, Aas JA, Boumenna T, Dewhirst FE and Paster BJ. Phylogenetic diversity and temporal variation in the Spirochetal community in microbial mats of Ebro and Camargue Deltas. Unpublished.

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33, W557-559.

Habetha M, Bosch TC (2005) Symbiotic Hydra express a plant-like peroxidase gene during oogenesis. J Exp Biol 208, 2157-2165.

Hand C, Uhlinger K (1994) The unique, widely distributed sea anemone, *Nematostella vectensis* Stephenson: A review, new facts, and questions. Estuaries 17, 501-508.

Harter VL, Matthews RA. (2005) Acute and Chronic Toxicity Test Methods for *Nematostella vectensis* Stephenson. Bull. Environ. Contam. Toxicol. 74:830–836

Haruta S, Yamaguchi H, Yamamoto ET, *et al.* (2000) Functional analysis of the active site of a metallo-beta-lactamase proliferating in Japan. Antimicrob Agents Chemother 44, 2304-2309.

Harwood CS, Canale-Parola E (1984) Ecology of spirochetes. Annu Rev Microbiol 38, 161-192.

Hilyard EJ, Jones-Meehan JM, Spargo BJ, Hill RT (2008) Enrichment, isolation, and phylogenetic identification of polycyclic aromatic hydrocarbon-degrading bacteria from Elizabeth River sediments. Appl Environ Microbiol 74, 1176-1182.

Hines ME, Evans RS *et al.* (1999) Molecular phylogenetic and biogeochemical studies of sulfate-reducing bacteria in the rhizosphere of Spartina alterniflora. Appl. Environ. Microbiol. 65:2209-2216

Hoegh-Guldberg O (1999) Climate change, coral bleaching and the future of the world's coral reefs. Mar. Freshwater Res. 50, 839–866.

Holmstrom C, Kjelleberg S (1999) Marine Pseudoalteromonas species are associated with higher organisms and produce biologically active extracellular agents. FEMS Microbiol Ecol 30, 285-293.

Hovius JW, van Dam AP, Fikrig E (2007) Tick-host-pathogen interactions in Lyme borreliosis. Trends Parasitol 23, 434-438.

Howarth RW, Teal JM. (1979) Sulfate Reduction in a New England Salt Marsh. Limnology and Oceanography, 24(6): 999-1013

Hutton DMC and Smith VJ (1996) Antibacterial Properties of Isolated Amoebocytes From the Sea Anemone *Actinia equina*. The Biological Bulletin, Vol 191, Issue 3 441-451

Kasahara S, Bosch TC (2003) Enhanced antibacterial activity in Hydra polyps lacking nerve cells. Dev Comp Immunol 27, 79-85.

Kellogg CA, Lisle JT, Galkiewicz JP (2009) Culture-independent characterization of bacterial communities associated with the cold-water coral *Lophelia pertusa* in the northeastern Gulf of Mexico. Appl Environ Microbiol 75, 2294-2303.

Kirchman DL *et al.* (2001) Structure of bacterial communities in aquatic systems as revealed by filter PCR. Aquat Microb Ecol Vol. 26: 13–22

Kogure K, Simidu U, Taga N (1979) A tentative direct microscopic method for counting living marine bacteria. Can J Microbiol 25, 415-420.

Kormas KA, Ntemiri A and Thessalou-Legaki M. Bacterial phylotypes from empty intestine of *Pestarella tyrrhena* (Decapoda: Thalassinidea). Unpublished.

Kuhn K, Streit B, Schierwater B (1996) Homeobox genes in the cnidarian *Eleutheria dichotoma*: evolutionary implications for the origin of Antennapedia-class (HOM/Hox) genes. Mol Phylogenet Evol 6, 30-38.

Kurahashi & Yokota (2007) *Endozoicomonas elysicola* gen. nov., sp. nov., a γ-proteobacterium isolated from the sea slug Elysia ornata Systematic and Applied Microbiology 30(3), 202-206

Lafond RE, Lukehart SA (2006) Biological basis for syphilis. Clin Microbiol Rev 19, 29-49.

Larcher M, Ghiglione J, Intertaglia L and Lebaron P. Structure and functions of bacterial communities in a coastal NW Mediterranean ecosystem. Unpublished.

Lau SCK *et al.* (2002) Bioactivity of bacterial strains isolated from marine biofilms in Hong Kong waters for the induction of larval settlement in the marine polychaete Hydroides elegans Mar Ecol Prog Ser. Vol. 226: 301–310

Lesser, M. P., C. H. Mazel, M. Y. Gorbunov, and P. G. Falkowski. 2004. Discovery of symbiotic nitrogen-fixing cyanobacteria in corals. Science 305:997-1000

Lozupone C, Hamady M, Knight R (2006) UniFrac--an online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7, 371.

Ludwig W, Strunk O, Westram R, *et al.* (2004) ARB: a software environment for sequence data. Nucleic Acids Res 32, 1363-1371.

Luque-Almagro VM, Blasco R, Huertas MJ, *et al.* (2005) Alkaline cyanide biodegradation by *Pseudomonas pseudoalcaligenes* CECT5344. Biochem Soc Trans 33, 168-169.

Madrid VM, Taylor GT, Scranton MI, Chistoserdov AY (2001) Phylogenetic diversity of bacterial and archaeal communities in the anoxic zone of the Cariaco Basin. Appl Environ Microbiol 67, 1663-1674.

Martindale MQ, Finnerty JR, Henry JQ (2002) The Radiata and the evolutionary origins of the bilaterian body plan. Molecular Phylogenetics and Evolution 24 (3): 358-365

Miller DJ, Ball EE (2000) The coral Acropora: what it can contribute to our knowledge of metazoan evolution and the evolution of developmental processes. Bioessays 22, 291-296.

Miller DJ, Hemmrich G, Ball EE, *et al.* (2007) The innate immune repertoire in cnidaria--ancestral complexity and stochastic gene loss. Genome Biol 8, R59.

Miyazaki M, Nogi Y, Ohta Y, *et al.* (2008) *Microbulbifer agarilyticus* sp. nov. and *Microbulbifer thermotolerans* sp. nov., agar-degrading bacteria isolated from deep-sea sediment. Int J Syst Evol Microbiol 58, 1128-1133.

Moran Y, Gurevitz M (2006) When positive selection of neurotoxin genes is missing: The riddle of the sea anemone Nematostella *vectensis*. FEBS Journal 273(17): 3886-3892

Muscatine L, Lenhoff HM (1963) Symbiosis: On the Role of Algae Symbiotic with Hydra. Science 142, 956-958.

Nakagawa S, Takai K, Inagaki F, *et al.* (2005) Distribution, phylogenetic diversity and physiological characteristics of epsilon-Proteobacteria in a deep-sea hydrothermal field. Environ Microbiol 7, 1619-1632.

Nelson KE, Weinel C, Paulsen IT, *et al.* (2002) Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440. Environ Microbiol 4, 799-808.

Nielsen C (2001) Animal evolution: interrelationships of the living phyla. 2nd Ed. Oxford University Press. Pp. 51-67

Nishino SF, Spain JC (1993) Degradation of nitrobenzene by a *Pseudomonas pseudoalcaligenes*. Appl Environ Microbiol 59, 2520-2525.

Olano CT, Bigger CH (2000) Phagocytic activities of the gorgonian coral *Swiftia exserta*. J Invertebr Pathol 76, 176-184.

Orphan VJ, Taylor LT, Hafenbradl D, Delong EF. (2000) Culture-Dependent and Culture-Independent Characterization of Microbial Assemblages Associated with High-Temperature Petroleum Reservoirs. Appl. Environ. Microbiol. 66: 700-711

Ovchinnikova TV, Balandin SV, Aleshina GM, *et al.* (2006) Aurelin, a novel antimicrobial peptide from jellyfish Aurelia aurita with structural features of defensins and channel-blocking toxins. Biochem Biophys Res Commun 348, 514-523.

Palincsar EE *et al.* (1989) Bacterial Aggregates Within the Epidermis of the Sea Anemone *Aiptasia pallida.* Biol.Bull. 177: 130-140

Palys T, Nakamura LK, Cohan FM (1997) Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int J Syst Bacteriol 47, 1145-1156.

Putnam NH, Srivastava M, Hellsten U, *et al.* (2007) Sea anemone genome reveals ancestral eumetazoan gene repertoire and genomic organization. Science 317, 86-94.

Rahat M, Dimentman C (1982) Cultivation of bacteria-free Hydra viridis: missing budding factor in nonsymbiotic hydra. Science 216:67–68.

Reitzel AM, Darling JA, Sullivan JC and Finnerty JR. (2008) Global population genetic structure of the starlet anemone Nematostella *vectensis*: multiple introductions and implications for conservation policy. Journal of Biological Invasions. 10 (8): 1197-1213

Reshef L, Koren O, Loya Y, Zilber-Rosenberg I, Rosenberg E (2006) The coral probiotic hypothesis. Environ Microbiol 8, 2068-2073.

Riley MA, Gordon DM (1999) The ecological role of bacteriocins in bacterial competition. Trends Microbiol 7, 129-133.

Rosenberg E, Ben-Haim Y (2002) Microbial diseases of corals and global warming. Environ Microbiol 4, 318-326.

Rosenberg E, Koren O, Reshef L, Efrony R, Zilber-Rosenberg I (2007) The role of microorganisms in coral health, disease and evolution. Nat Rev Microbiol 5, 355-362.

Sachs JL, Wilcox TP (2006) A shift to parasitism in the jellyfish symbiont *Symbiodinium microadriaticum.* Proc Biol Sci 273, 425-429.

Sakai K, Yamanaka H, Moriyoshi K, Ohmoto T, Ohe T (2007) Biodegradation of bisphenol A and related compounds by Sphingomonas sp. strain BP-7 isolated from seawater. Biosci Biotechnol Biochem 71, 51-57.

Santiago-Vazquez LZ, Bruck TB, Bruck WM, *et al.* (2007) The diversity of the bacterial communities associated with the azooxanthellate hexacoral *Cirrhipathes lutkeni.* ISME J 1, 654-659.

Schmid V, Yanze N, Spring J, Reber-Muller S. (1998) The striated muscle of hydrozoan medusae: Development and stability of the differentiated state. Zoology-Analysis of Complex Systems 101: 365-374

Scholz CB, Technau U (2003) The ancestral role of Brachyury: expression of NemBra1 in the basal cnidarian *Nematostella vectensis* (Anthozoa). Dev Genes Evol 212, 563-570.

Schuett C *et al.* (2007) Bacterial aggregates in the tentacles of the sea anemone *Metridium senile* Helgol Mar Res 61:211–216

Shashar N, Cohen, Y, Loya, Y & Sar, N (1994) Nitrogen fixation (acetylene reduction) in stony corals: evidence for coral–bacteria interactions. Mar. Ecol. Prog. Ser. 111, 259–264.

Shida K, Terajima D, Uchino R, *et al.* (2003) Hemocytes of Ciona intestinalis express multiple genes involved in innate immune host defense. Biochem Biophys Res Commun 302, 207-218.

Stahl, DA & Amann RI (1991) Nucleic acid techniques in bacterial systematics. In Development and Application of Nucleic Acid Probes (Stackebrandt E & Goodfellow M eds.), John Wiley & Sons, New York, New York. Pp. 205—248.

Starcevic A, Akthar S, Dunlap WC, *et al.* (2008) Enzymes of the shikimic acid pathway encoded in the genome of a basal metazoan, Nematostella *vectensis*, have microbial origins. Proc Natl Acad Sci U S A 105, 2533-2537.

Steele RE (2002) Developmental signaling in Hydra: what does it take to build a "simple" animal? Dev Biol 248, 199-219.

Stephens EA, Braissant O, Visscher PT (2008) Spirochetes and salt marsh microbial mat geochemistry:Implications for the fossil record Notebooks on Geology, Brest, Article.

Sullivan JC, Reitzel AM, Finnerty JR (2008) Upgrades to StellaBase facilitate medical and genetic studies on the starlet sea anemone, Nematostella *vectensis*. Nucleic Acids Res 36, D607-611.

Sullivan JC, Ryan JF, Watson JA, *et al.* (2006) StellaBase: the Nematostella *vectensis* Genomics Database. Nucleic Acids Res 34, D495-499.

Sunagawa S, Desantis TZ, Piceno YM, *et al.* (2009) Bacterial diversity and White Plague Disease-associated community changes in the Caribbean coral Montastraea faveolata. ISME J.

Suzuki Y, Kojima S, Sasaki T, *et al.* (2006) Host-symbiont relationships in hydrothermal vent gastropods of the genus Alviniconcha from the Southwest Pacific. Appl Environ Microbiol 72, 1388-1393.

Tarrant AM, Reitzel AM, Blomquist CH, *et al.* (2009) Steroid metabolism in cnidarians: insights from Nematostella *vectensis*. Mol Cell Endocrinol 301, 27-36.

Technau U, Cramer von Laue C, Rentzsch F, *et al.* (2000) Parameters of self-organization in Hydra aggregates. Proc Natl Acad Sci U S A 97, 12127-12131.

Wang JT *et al.* (2008) Tenacibaculum aiptasiae sp. nov., isolated from a sea anemone Aiptasia pulchella. Int J Syst Evol Microbiol 58, 761-766

Webster NS, Smith LD, Heyward AJ, *et al.* (2004) Metamorphosis of a scleractinian coral in response to microbial biofilms. Appl Environ Microbiol 70, 1213-1221.

Yamamoto S, Kasai H *et al.* (2000) Phylogeny of the genus *Pseudomonas*: intrageneric structure reconstructed from the nucleotide sequences of gyrB and rpoD genes. Microbiology 146, 2385-2394

Yen KM, Karl MR, Blatt LM *et al.* (1991) Cloning and characterization of a *Pseudomonas* mendocina KR1 gene cluster encoding toluene-4-monooxygenase. J Bacteriol 173, 5315-5327.

Zhang Y, Chen X and Zhou M. Diversity of cultivable bacteria in deep-sea sediment. Unpublished.

Appendix A

Description of bacteria groups found in multiple samples

The following section includes a more detailed description of different groups of *N. vectensis*-associated populations in the environment. They are discussed at different taxonomic ranks due to the disparities in the extent of current knowledge of the nearest known clone or isolate in Genbank.

A. 1 Epsilon-Proteobacteria

Epsilon-Proteobacteria is the most poorly characterized class within the phylum Proteobacteria. The earlier known epsilon-Proteobacteria were infectious human pathogens belonging to the genera Campylobacter and Helicobacter. More representatives were later found in diverse environments, ranging from shallow marine habitats like salt marshes to deep-sea hydrothermal vents, vent fauna, deep-sea marine subsurfaces, to terrestrial systems like groundwater, caves and springs. Some episymbiotic epsilon-Proteobacteria were discovered in association with deep-sea vent metazoans (Nakagawa *et al.*, 2005).

Epsilon-Proteobacteria are important as primary colonizers, primary producers and symbiotic partners of animals. They play key roles in the biogeochemical cycling of carbon, nitrogen and sulfur. In many sulfidic habitats, especially at oxic-anoxic interfaces such as the chemocline zone where hydrogen sulfide from sediment meets oxygenated sea water, they may be the dominant microbes involved in the cycling of these compounds (Madrid *et al.*, 2001). Epsilon-Proteobacteria are versatile and different species can either reduce or oxidize sulfur and nitrogen compounds, which heavily influence the sulfur/nitrogen concentration gradients within a habitat.

Many of the epsilon-Proteobacteria studied so far are capable of chemolithoautotrophy, which is thought to be the first type of metabolic strategy to have evolved (Campbell *et al.*, 2006).

Some Epsilon-Proteobacteria are also endosymbionts of hydrothermal vent fauna. One single epsilon-Proteobacteria phylotype was found to provide nutrients to their gastropod hosts *Alviniconcha* (Suzuki *et al.*, 2006). Other examples include symbiosis with Pompeii worms (*Alvinella pompejana*) of the East Pacific Rise and the hydrothermal vent shrimp *Rimicaris exoculata* of the Mid-Atlantic Ridge. *A. pompejana* contains two closely related epsilon-Proteobacterial phylotypes that dominate an episymbiont biomass 16S rRNA gene library (Campbell *et al.*, 2006).

Based on comparative analysis of 16S rRNA genes, epsilon-Proteobacterial sequences belong to two main orders, the Nautiliales (families Nautilia, Caminibacter and Lebetimonas) and the Campylobacterales (families Campylobacteraceae, Helicobacteraceae and Hydrogenimonaceae). The family Campylobacteraceae contains genera *Campylobacter, Arcobacter, Sulfurospirillum* and *Thiovulum*, while Helicobacteraceae includes *Helicobacter* and *Wolinella*. Other unclassified sequences retrieved from various marine and terrestrial habitats fall into an "environmental" group (Campbell *et al.* 2006). The closest cultured relatives of the *N. vectensis*-associated epsilon-Proteobacteria ribotype (albeit at 85% 16S rRNA identities) are Helicobacter and Arcobacter, both of which have representative pathogenic genera in human gastrointestinal tract (Engberg *et al.*, 2000).

A.2 *Pseudomonas*: *P. pseudoalcaligenes* & *P. putida*

P. putida is a metabolically versatile soil bacterium that is commonly used as a biosafety host for the cloning of foreign genes and has considerable potential for biotechnological applications (Nelson *et al.*, 2002). *P. pseudoalcaligenes* has been shown capable of degrading nitrobenzene (Nishino, Spain, 1993), degrade naphthalene (Garcia-Valdes *et al.*, 1988), as well as utilizing cyanide as a nitrogen source (Luque-Almagro *et al.*, 2005), while *P. mendocina,* which is closely related to *P. pseudoalcaligenes* has been found to degrade toluene via Toluene-4-monooxygenase (Yen *et al.*, 1991). These bacteria are ubiquitous in soil as well as marine and fresh water environments and have been isolated in enrichments containing aromatic hydrocarbons (Hilyard *et al.*, 2008; Sakai *et al.*, 2007). *P. pseudoalcaligenes* is also able to increase degradation rate of Bisphenol A *Sphingomonas* species (Sakai *et al.*, 2007).

However, despite their possible biotechnological applications in bioremediation, these strains are of concern as potential opportunistic pathogens. *P. mendocina* has been isolated from human ulcers and blood, including a case of human infective endocarditis that was most likely caused by *P. mendocina* (Aragone *et al.*, 1992). Further concerns of *P. pseudoalcaligenes* and *P. mendocina* as potential pathogens are that these species were found to be among the dominant chloramphenicol-resistant bacteria in Jiaozhou Bay, China (Dang, *et al.*, 2008). The most common mechanisms of resistance in marine bacteria are chloramphenicol acetyltransferase genes, but surprisingly neither *P. pseudoalcaligenes* nor *P. mendocina* contained these genes (Dang, *et al.* 2008).

Previous studies have shown that the beta-lactamase gene was found to be widely distributed in antibiotic resistant *P. aeruginosa* clinical isolates (Haruta *et al.*, 2000). That *P. pseudoalcaligenes* shares high 16S rRNA identity with a human pathogen makes it a convenient environmental model for human health as it may have evolutionarily conserved mediations of interactions with *N. vectensis* tissue.

A.3 Gamma-Proteobacteria: *Endozoicimonas elysicola*

This species was first isolated from a mollusk, the sea slug *Elysia ornata* collected in seawater off the coast of Izu-Miyake Island, Japan in 2007. A Gram-negative, strictly aerobic, rod-formed bacterium, it constitutes a novel lineage in gamma-Proteobacteria related to the genera Zooshikella, Oceanospirillum, Microbulbifer, Marinobacter, Saccharospirillum and *Pseudomonas* (Kurahashi, Yokota, 2007).

A later study on the tentacles of the sea anemone *Metridium senile* revealed tightly packed bacteria aggregates. Sequence analysis showed that the dominant band detected in all of the samples was 98% similar to *Endozoicimonas elysicola* (Schuett *et al.*, 2007). The role E*ndozoicimonas* plays in animals is currently unknown.

A.4 Spirochaete

The phylum Spirochaetes consist of only six genera: Borrelia, Brevinema, Cristispira, Spirochaeta, Spironema, Treponema, most of which are pathogenic. Borrelia includes several species that cause relapsing fever (*B. recurrentis*) and Lyme disease (*B. burgdorferi*) in humans

(Hovius *et al.*, 2007), while Treponema contains the agents of syphilis (*T. pallidum*) and yaws (*T. pertenue*) (Lafond, Lukehart, 2006).

The clones found in this study belong to the genus Spirochaeta, which are free-living, anaerobic heterotrophic bacteria ubiquitous in most natural habitats, including the fresh water and anoxic sediments, in human mouth and animal digestive tract (Harwood, Canale-Parola, 1984). In general, Spirochaetes play an important role in microbial mat biogeochemistry through sulfide removal - using oxygen or nitrate as electron acceptors - and degradation of exopolymeric substances (EPS) by stimulating the sulfate-reducing bacteria. It was also proposed that EPS degradation by spirochetes results in calcium liberation and formation of nucleation sites for $CaCO_3$ precipitation, enhancing the long term preservation of microbial mats (Stephens *et al.* 2008). The role of Spirochaetes in marine invertebrate physiology is currently unknown.

A.5 Bacteroidetes

Hints of intimate symbiotic relationship between Flavobacteria and marine animals have surfaced recently. Starcevic *et al.* (2008) reported four shikimic acid pathway genes that are closely related to those of a Flavobacteria, *Polaribacter (*Tenacibaculum) sp. MED152, within the *N. vectensis* genome, suggesting the existence of a previously unknown bacterial symbiont in the basal metazoan, or a lateral gene transfer of a bacterial gene into the *N. vectensis* genome. Moreover, the holobiont contained many *Tenacibaculum*-like gene orthologs, including a Flavobacteriaceae 16S rRNA sequence, suggesting that these genes were transferred from bacteria to the host (Starcevic *et al.*, 2008). In another study, *Tenacibaculum aiptasiae*, a novel

species of Flavobacteria was isolated from a sea anemone (*Aiptasia pulchella*). The role *T. aiptasiae* plays in *A. pulchella* is currently unknown.

Cytophaga-Flavobacteria is the main microbial group in marine environments, making up about half of all bacteria cells countable in FISH. They are sometimes the dominant group in mature freshwater biofilms (Kirchman, 2001), and are important in determining the biofilm community structure. Principal-component analysis of FISH data showed that alpha-Proteobacteria and Cytophaga-Flavobacterium had the largest influence on the overall community composition (Webster *et al.*, 2004). Among other groups present in marine biofilm like gamma-Proteobacteria and Gram-positive bacteria, some Cytophaga-Flavobacteria have been found to induce larval settlement of the marine polychaete *Hydoides elegans* (Lau *et al.* 2002).

Cultured isolates of Cytophaga-Flavobacteria are capable of degrading high molecular mass biopolymers such as cellulose, chitin and pectin (Cottrell, Kirchman, 2000). Previous studies showed that Cytophaga-Flavobacteria are enriched on particulate organic detritus in marine habitats (Delong *et al.* 1993).

Some symbiotic Cytophaga-Flavobacteria have been found in the hexacoral *Cirrhipathes lutkeni*. FISH counts for the whole coral holobiont showed that gamma-Proteobacteria and Actinobacteria makes up 40% of the microbiota, followed by alpha-Proteobacteria (14%), Firmicutes (9%), Cytophaga-Flavobacterium (7%), beta-Proteobacteria (6%) and Chloroflexi (2%) (Santiago-Vazquez *et al.*, 2007).