Metagenomic and metaproteomic analysis of the microbial communities of marine sponges

Author:
Liu, Michael Yizhe

Publication Date:
2012

DOI:
https://doi.org/10.26190/unsworks/15450

Metagenomic and metaproteomic analysis of the microbial communities in marine sponges

Michael Yizhe Liu

A thesis submitted to the University of New South Wales
for the degree of Doctor of Philosophy

February 2012

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **LIU**

First name: **MICHAEL YIZHE** Other name/s:

Abbreviation for degree as given in the University calendar: **Ph.D**

School: **BIOTECHNOLOGY AND BIOMOLECULAR SCIENCES** Faculty: **SCIENCE**

Title:
METAGENOMIC AND METAPROTEOMIC ANALYSIS OF THE MICROBIAL COMMUNITIES IN MARINE SPONGES

Abstract 350 words maximum: (PLEASE TYPE)

Marine sponges harbour complex communities of diverse microorganisms, which have been postulated to form intimate symbiotic relationship with their host. Despite the importance of this symbiosis - from both the ecological and biotechnological perspectives - little is known about the functional properties of the bacterial symbionts and their interactions with the host. This thesis presents the characterisation of the microbial communities associated with two sponges, *Cymbastela concentrica* and *Rhopaloeides odorabile*, from the eastern coast of Australia.

A functional genomic analysis of an uncultured Deltaproteobacterium in the sponge *C. concentrica* showed that this organism represents a novel phylogenetic clade and lives in association with a cyanobacterium within the sponge tissue. The result also provide an overview of the predicted functional and ecological properties as well as the complex interactions of this bacterium with its surrounding.

An integrated metaproteogenomic approach was used to characterise the microbial community associated with *C. concentrica*. The result unveiled the expression of specific proteins involved in the transport of typical sponge metabolites, respiration, stress protection and molecular interactions (e.g. eukaryotic-like proteins). This analysis highlighted the requirement for the microbial community to respond to variable environmental conditions.

A subsequent study investigated the genomic and functional properties of an uncultured *Phyllobacteriaceae* phylotype, This organism was predicted to be capable of using host-derived compounds and to carry out anaerobic respiration through nitrate reduction. Functional signatures associated with survival using defence and host-symbiont interactions were also identified.

The impact of elevated water temperature on the microbial communities within *R. odorabile* was investigated with a metaproteogenomic approach. Changes in taxonomic composition, species richness, genomic content and expressed functional profile of sponge microbial communities were observed during thermal stress. In particular, putative symbiotic functions, such as metabolite transport, metabolism of sponge-specific substrates and maintenance of cellular structure, were identified in unstressed sponge communities, but were absent in stress samples. The decline in expression of symbiont functions in stressed sponge sample demonstrated that sponge necrosis is likely to be caused by the disturbance of interactions within the sponge holobiont, rather than as a result of pathogen infection.

Declaration relating to disposition of project thesis/dissertation

I hereby grant to the University of New South Wales or its agents the right to archive and to make available my thesis or dissertation in whole or in part in the University libraries in all forms of media, now or here after known, subject to the provisions of the Copyright Act 1968. I retain all property rights, such as patent rights. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

I also authorise University Microfilms to use the 350 word abstract of my thesis in Dissertation Abstracts International (this is applicable to doctoral theses only).

26-06-2012

…… …… ……

Signature Witness Date

The University recognises that there may be exceptional circumstances requiring restrictions on copying or conditions on use. Requests for restriction for a period of up to 2 years must be made in writing. Requests for a longer period of restriction may be considered in exceptional circumstances and require the approval of the Dean of Graduate Research.

FOR OFFICE USE ONLY Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

COPYRIGHT STATEMENT

Signed …….....

26th June 2012

Date …….....

AUTHENTICITY STATEMENT

‘I certify that the Library deposit digital copy is a direct equivalent of the final officially approved version of my thesis. No emendation of content has occurred and if there are any minor variations in formatting, they are the result of the conversion to digital format.’

Signed …….....

26th June 2012

Date …….....

ORIGINALITY STATEMENT

'I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.'

Signed ……

26/02/2012

Date ……

Abstract

Marine sponges harbour complex communities of diverse microorganisms, which have been postulated to form intimate symbiotic relationship with their host. Despite the importance of this symbiosis - from both the ecological and biotechnological perspectives - little is known about the functional properties of the bacterial symbionts and their interactions with the host. This thesis presents the characterisation of the microbial communities associated with two sponges, *Cymbastela concentrica* and *Rhopaloeides odorabile*, from the eastern coast of Australia.

A functional genomic analysis of an uncultured Deltaproteobacterium in the sponge *C. concentrica* showed that this organism represents a novel phylogenetic clade and lives in association with a cyanobacterium within the sponge tissue. The result also provide an overview of the predicted functional and ecological properties as well as the complex interactions of this bacterium with its surrounding.

An integrated metaproteogenomic approach was used to characterise the microbial community associated with *C. concentrica*. The result unveiled the expression of specific proteins involved in the transport of typical sponge metabolites, respiration, stress protection and molecular interactions (e.g. eukaryotic-like proteins). This analysis highlighted the requirement for the microbial community to respond to variable environmental conditions.

A subsequent study investigated the genomic and functional properties of an uncultured *Phyllobacteriaceae* phylotype, This organism was predicted to be capable of using host-derived compounds and to carry out anaerobic respiration through nitrate reduction. Functional signatures associated with survival using defence and host-symbiont interactions were also identified.

The impact of elevated water temperature on the microbial communities within *R. odorabile* was investigated with a metaproteogenomic approach. Changes in taxonomic composition, species richness, genomic content and expressed functional profile of sponge microbial communities were observed during thermal stress. In particular, putative symbiotic functions, such as metabolite transport, metabolism of sponge-

specific substrates and maintenance of cellular structure, were identified in unstressed sponge communities, but were absent in stress samples. The decline in expression of symbiont functions in stressed sponge sample demonstrated that sponge necrosis is likely to be caused by the disturbance of interactions within the sponge holobiont, rather than as a result of pathogen infection.

Acknowledgements

My first vote of thanks goes to my primary supervisors, Staffan Kjelleberg and Torsten Thomas. I count myself fortunate to have had the opportunity to work on this stimulating and important project with you. Staffan, thank you for your guidance and support of my project. Your honest, thorough and thoughtful approach to studying science, and your ability to see the big picture has been influential. I hope that I will still be able to turn to you for your advice throughout my career in science. Torsten, you are amazing! Thank you for your encouragement, enthusiasm and efforts in explaining things clearly and simply to me. I have learnt a lot from you, in particular your intellectual curiosity, endless quest of scientific observations and attention to details. I appreciate the countless number of meetings that we had in your office, which is always inspiring and refreshing. I would have been lost without you.

To Chris Fan Lu, I thank you for the extraordinary amount of time that you invested in getting me up to speed with this project and all the laughter we had. I wish you the best of luck in your future career and hope that we will continue to work together on future projects. Chris designed, and performed the metagenomic sequencing of the microbial communities in *C.concentrica* and *R.odorabile*. I would also like to thank Maria Yung, for the metagenomic sequence data of *C.concentrica*. The work presented in Chapter 2 was based on her metagenome sequencing data generated using Sanger sequencing.

I would also like to thank Dr. Nicole Webster and her talented group at AIMS. Nicole didi a great job in coordinating the heat-stress experiment described in Chapter 5 and provide valuable advices to my thesis. Nicole has also helped me to learn FISH technique that was used in my PhD study. I am thankful that we have worked together on such an exciting project!

I am very grateful for having the opportunity to work with Dr Mark Raftery and Dr Zhong Ling from the Bioanalytical Mass Spectrometry Facility at UNSW, thank you for your training in mass spectrometry and proteomics. Thanks to Gavin and Hayley from the Histology and Microscopy unit at UNSW, thank you for processing my frozen sponge samples. I would also like to acknowledge our collaborator, the J. Craig Venter Institute for all the sequencing work.

Thanks to Bill O'Sullivan and Sharon Longford for proofreading various chapters of this thesis and invaluable advises on thesis writing.

I am so proud to have been part of CMB for the past 6 years and this has left me with so many people to thank. I thank all members of lab 304 that came and went: Cath, Sharon, Vickey, Ani, Maria, Neil, Melanie, Chris, Raymond, Shaun, Scott, Wilson, Sherwin, Dennis, Miriam, Gee, MelG, Becca, and Adam. Thank you for your support, advice and great fun! Many thanks to Adam and Kirsty for their administrative support. I am convinced that without both of you, the centre would have struggled to operate.

On a personal note, I need to thank my parents for all of their constant support and advice. Thanks to them for their belief in my abilities, with constant words of encouragement, and countless sacrifices that they made for me over the years. They mean a great deal to me! To my best friends outside of science: Michael, Sandra, Frank and Danni, thank you for putting up with me in the past year while I was writing this thesis. I am finally free, and can start returning the friendship a little better!

Finally, I am exceedingly grateful to my gorgeous wife Cindy for her endless reserve of patience and support. Thank you for your understanding, love and friendship. You have made my PhD candidature more enjoyable and less difficult.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

1D	One-dimensional
2D	Two-dimensional
3D	Three-dimensional
AMD	Acid mine drainage
ATP	Adenosine-5'-triphosphate
BAC	Bacterial artificial chromosome
Bp	base pair
cDNA	complementary Deoxyribonucleic acid
COG	Clusters of Orthologous Group
Da	Daltons
DMSO	dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	dithiothreitol
EDTA	ethylene diamine tetraacetate
FISH	Fluorescence *in situ* Hybridisation
FITC	Fluorescein isothiocyanate
Gbp	giga base pair
HPLC	High Performance Liquid Chromatography
Kb	kilo base pair
L	litre
LC	Liquid Chromatography
LTQ	Linear trap quadrupole
m/z	mass-to-charge ratio
MALDI	Matrix Assisted Laser Desorption/Ionization
mL	millilitre
mRNA	messenger Ribonucleic acid
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
nt	nucleotide
^{o}C	degree celsius
ORF	Open Reading Frame
OTU	Operational Taxonomical Unit

PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
ppm	parts per million
rRNA	ribosomal Ribonucleic acid
SCUBA	Self contained underwater breathing apparatus
SDS	sodium dodecyl sulfate
Tris	Tris (hydroxymethyl) aminomethane
µL	micro litre
µmol	micro molar
V	Volts
x g	times gravity

Chapter One
General Introduction

1.1 Microbial Ecology: the study of microorganisms in the environment

Microorganisms, including bacteria, archaea, viruses, and fungi, are the most diverse and abundant group of organisms in our environment. Ever since the invention of the microscope over 350 years ago, there have been questions about the microbial world, its role in global biochemical processes, its interaction with other organisms and its importance to humans. However, progress in understanding these issues remained slow and difficult despite many microorganisms having already been cultivated, maintained and studied in the laboratory. Challenges associated with the fact that a tiny fraction (<1-5%) of bacteria that can be observed *in situ* could be grown in culture (Handelsman 2004, Staley and Konopka 1985) have long been recognised, and have to date prevented the analysis of the metabolism and the activity of many yet-to-be-cultured lineages. Only recently has the development of powerful new molecular techniques (for example, PCR, automated DNA sequencing, and community finger-printing) overcome the need for cultivation to characterise microorganisms and leading to significant discoveries of the abundance and diversity of microbial life, and its role in global ecology.

Microbial ecology is aimed at connecting the structure and function of microbial communities in natural environments and has been the subject of much interest in the past decade (Kowalchuk *et al.,* 2007). The study of diverse regulatory mechanisms in microbial interactions has provided intriguing insights into the phenomena of food webs, symbiosis and pathogenicity. The key role of the microbial cumulative mass, its distribution and coordinated activity in shaping the environment on Earth, and its activities in biogeochemical reactions (Whitman *et al.,* 1998) has now been recognised. Current rapid advances in microbial ecology will help provide a greater understanding of the complex ecosystem function, and help to potentially predict the Earth's response

to global changes, such as an increased human population, exploitation of natural resources, climate changes and a polluted environment.

1.2 Methods for studying microbial diversity and function

Many methods have been developed and used to assess microbial diversity and function. Early studies in microbial ecology relied heavily upon culturing and for many years remained the only way to examine community structure, based on the representative bacteria that are amenable to growth on laboratory media. Despite recent advances in culturing methodology (Joseph *et al.,* 2003, Schut *et al.,* 1993, Stingl *et al.,* 2007) (and reviewed by Giovannoni and Stingl (2007)), most microbes still elude cultivation, making studies of their functional traits extremely difficult. To overcome this problem, an array of culture-independent methods have been developed, thus enabling scientists to make considerable progress in better understanding the phylogenetic and functional diversity of microorganisms in the environment.

1.2.1 Molecular approach using 16S rRNA gene amplification and sequencing

Microbial phylogeny has mainly been based on amplifying or sequencing the DNA of small subunit rRNAs (e.g. 16S rRNA gene) to examine the richness and diversity of microbial life in the environment (Pace *et al.,* 1986, Woese 1987). The implementations of such methods have revolutionised how scientists interpret microbial diversity and evolution (Pace 1997). rRNA gene sequences have a number of useful characteristics. They are universally found in prokaryotes and contain mixtures of conserved and variable regions, which reflected evolutionary relationships. These features have been exploited to develop a system of taxonomy for bacteria, and the 16S rRNA gene sequence databases became the largest repository of bacterial gene sequences (Ludwig and Schleifer 1999, Woese 1987).

The use of the polymerase chain reaction (PCR) to enrich the 16S rRNA gene has been applied in direct culture-independent phylogeny based surveys of environmental microbial communities (e.g. (Britschgi and Giovannoni 1991, Giovannoni *et al.,* 1990, Schmidt *et al.,* 1991, Ward *et al.,* 1990)). More recent developments in automated

sequencing technologies have accelerated this research. For example, a greater level of diversity was demonstrated through massively parallel tag 454 pyrosequencing of deep sea water samples (Sogin *et al.,* 2006) and demonstrate an unexpected level of phylogenetic diversity in natural environment (Head *et al.,* 1998, Pace 1997, Riesenfeld *et al.,* 2004). The extensive sequencing of 16S rRNA gene sequences across a wide range of communities and isolates has resulted in the generation of several large reference databases, such as the ribosomal database project (RDP) II (Cole *et al.,* 2009), Greengenes (DeSantis *et al.,* 2006) and SILVA database (Pruesse *et al.,* 2007). These comprehensive databases allow for the classification of 16S rRNA gene sequences and continue to contribute to microbial ecology by revealing valuable taxonomical information. Based on the 16S rRNA gene marker and relevant databases, a number of other culture-independent approaches have been developed, including community fingerprinting methods and fluorescence *in situ* hybridisation (FISH).

1.2.2 Community (DNA) fingerprinting technique and fluorescence in situ hybridisation (FISH)

Community (DNA) fingerprint techniques, such as denaturing gradient gel electrophoresis (DGGE) and terminal restriction fragment length polymorphism (T-RFLP), are widely applied methods to examine shifts in microbial communities across temporal and spatial scales. For a T-RFLP analysis, the 16S rRNA genes in the community DNA are first amplified by PCR with a fluorescently labelled primer, then digested and finally separated by capillary electrophoresis. The size distribution of the terminal fragments presents a fingerprinting of the community structure (Liu *et al.,* 1997).

Fluorescent *in situ* hybridisation (FISH) is another commonly used culture-independent method which has been developed based on the targeting of phylogenetic marker genes, such as the 16S rRNA gene (Amann *et al.,* 2001, Amann *et al.,* 1995). Using the taxonomical information of the marker gene, a specific group of microorganisms can be detected *in situ*, and provide information on abundance, localisation and distribution. The method involves the hybridisation of fluorescently labeled synthetic DNA fragments (probes) to complementary regions within the target nucleic acids (e.g.

ribosomal RNA) of fixed cells or samples (Amann *et al.,* 2001, Amann *et al.,* 1995). Samples can be examined using fluorescence or confocal microscopy. This approach takes advantage of the high copy number of ribosomal RNA genes within a cell and their conserved nature among microorganisms (Amann and Fuchs 2008). Up to seven different fluorescently labeled probes can be used simultaneously in a single experiment (Amann *et al.,* 1996). However, like other ribosomal RNA gene based approaches, FISH analysis does not reveal any functional information on the target microorganism(s).

1.3 Metagenomics and Metatranscriptomics

Metagenomics was developed to link the taxonomic based structure of an environmental sample to its functional profile (Figure 1.1). Although the term 'metagenomics' was coined by the Handelsman group in 1998 (Handelsman *et al.,* 1998), referring to the function-based analysis of mixed environmental DNA species, the first experiments using "metagenomic" approaches had already occurred twenty years earlier, when bacteriophage-cloning techniques were used to produce genomic libraries (Schmidt *et al.,* 1991). A number of strategies have since been developed and the field can be broadly categorised into two parts, the function-based and the sequence-based approach. The two experimental approaches are not mutually exclusive, or dependent upon one another.

1.3.1 Functional metagenomics

Function-based metagenomics relies on the cloning of environmental DNA into expression vectors and propagating them in appropriate hosts, followed by the relevant activity screening (Chistoserdova 2010). The screening can be performed in two different ways: 1) function-then-phylogeny and 2) phylogeny-then-function.

The function-then-phylogeny approach screens for clones with specific functions (phenotypes) in a metagenomic library and then attempts to determine the phylogenetic origin of the cloned DNA, by sequencing the insert and searching for its phylogenetic markers (e.g. 16S rRNA) or to perform a taxonomic classification based on sequence or

compositional signatures. This approach does not require prior knowledge of the gene sequences and has led to the discovery of novel genes or pathways. For example, a long-chain *N*-acyltyrosine synthase encoding antibacterial activity was identified through functional screening of a soil metagenomic library (Brady *et al.,* 2004). Sequence analysis showed it was 14 to 37% identical to a predicted protein of previously unknown function from *Nitrosomonas europaea*, a gram-negative nitrifying *Betaproteobacterium*. A list of biocatalysts derived from functional-screening of metagenomes was summarised by Simon and Daniel (Simon and Daniel 2009) in their recent review. Such examples validate the use of functional screens as one means to characterise metagenomic libraries.

The phylogeny-then-function approach first screens the cloned DNA for specific phylogenetic markers. After a phylogentic marker is determined, the flanking regions of the marker are screened for functional genes. One of the first studies to link taxonomy and function was through the development and sequencing of large-insert clone libraries used a fosmid library (~40 kb) of environmental archaeal DNA cloned in an *Escherichia coli* vector (Stein *et al.,* 1996). The library was screened for clones carrying the phylogenetic marker (rRNA gene) and the full insert within these target clones was sequenced and searched for functional genes flanking the marker (Stein *et al.,* 1996). Other phylogeny-then-function studies have reported a number of surprising and important discoveries. For example, a 130 kb BAC clone was isolated from an uncultivated SAR86 bacterium and, through sequencing, a new class of genes of the rhodopsin family (proteorhopsin) was discovered (Beja *et al.,* 2000). These light-driven proton pumps had not been previously recognised in bacteria, nor in the ocean, and this led to the discovery of a new type of energy generation in oceanic bacteria. The use of larger-insert libraries has also been applied to ecological studies of high diversity samples (Rondon *et al.,* 2000) and for natural product screening (Handelsman *et al.,* 1998, Schleper *et al.,* 1998). Those culture-independent studies have provided useful insights into the taxonomical and physiological potential of abundant, yet uncultured, microbes in our environment.

Another very important use of functional metagenomics is the high-throughput screening for bioactive compounds and natural products. Functional screening is achieved by screening the clone libraries for phenotypes traditionally associated with

the production of small molecules. The recent developments for discovery of bioactive compounds have been extensively reviewed in the literatures (Banik and Brady 2010, Simon and Daniel 2009).

1.3.2 Sequencing based metagenomics

Sequence-based metagenomic analysis uses a random shotgun sequencing approach, which was developed for whole-genome sequencing. The direct sequencing approach represented a big step beyond the then prevailing marker gene based surveys, to offer a relatively unbiased view of not only the community structure, but also of the community's functional potential (Hugenholtz and Tyson 2008). The application of metagenomic sequencing to an acid mine drainage (AMD) biofilm community by Tyson and colleagues was one of the first applications to demonstrate the viability of such a strategy (Tyson *et al.,* 2004). In their study, community DNA from a low-diversity microbial biofilm was sequenced and it was possible to reconstruct almost complete genomes for two bacterial groups, *Leptospirillum* group II and *Ferroplasma* type II, in addition to reporting on three other partial genomes (Tyson *et al.,* 2004). The low diversity in the system allowed for the assignment of specific functions to individual members; for example nitrogen fixation was attributed specifically to *Leptospirillum* group II. This low-diversity community provided an excellent example to test the application of high-throughput community sequencing and delivered unsurpassed information on ecosystem function.

Venter and colleagues (Venter *et al.,* 2004) also used a similar metagenomic approach, and obtained approximately 1.6 Gbp of non-redundant microbial sequence information from seven seawater samples, from various locations within the Sargasso Sea, near Bermuda. The authors were able to conclude that there were at least 1800 different genomes within 148 unknown bacterial phylotypes, and 1.2 million previously unknown genes were discovered through the assembly of large scaffolds and the use of complex bioinformatic approaches (Venter *et al.,* 2004). Subsequently, the Global Ocean Survey (GOS) extended the analysis across a global transect, covering several thousand kilometers (Rusch *et al.,* 2007). The massive dataset indicated the presence of

many subtypes within dominant planktonic clades, raising the question of how such similar organisms can co-exist in the ocean.

With decreasing sequencing costs and initial success demonstrated by such studies, many large metagenomic sequence analyses from a diverse range of environments have followed, with particular focus in the marine environment (Table 1.1). Recently, a metagenomic approach was also applied to host-associated microbial communities in organisms such as corals (Dinsdale *et al.,* 2008b, Vega Thurber *et al.,* 2009) and sponges (Thomas *et al.,* 2010). This metagenomic approach proved to be useful for comparative analysis. For example, Tringe and colleagues compared the metagenomic profile of marine plankton, whale fall and farm soil and demonstrated that gene complements vary distinctly between ecosystems (Tringe *et al.,* 2005). Similar observations were made for the comparison of deep and surface planktonic ocean communities (DeLong *et al.,* 2006, Konstantinidis *et al.,* 2009, Sharon *et al.,* 2011, Xie *et al.,* 2011). The studies here represent just a few examples of many metagenomic sequencing approaches that have been reported or are underway. There are 4,335 metagenomes listed in the Metagenomics RAST (MG-RAST) server at the time (Oct-2011) of writing (Meyer *et al.,* 2008).

Table 1.1 Examples of recent sequence based metagenomic studies focused on the marine environment.

Target Microbiome	Sampling site	Experimental approach	Sequencing technique	Reference
Coastal seawater (Microcosms)	Sapelo Island, Georgia, USA	Immunocapture targeting new synthesised DNA in response to the presence of dissolved organic carbon substrates.	Pyrosequencing 454 Life Science Total sequences 306,679	(Mou *et al.,* 2008)
Coastal surface seawater	La Jolla, California, USA	Population enrichment using flow cytometric sorting.	Pyrosequencing 454 GS20 Total sequences 369,811	(Palenik *et al.,* 2009)
Coastal seawater	Boothbay Harbor, Maine, USA	Population enrichment using flow cytometric sorting. Multiple displacement amplification focusing on *Flavobacteria*	Combined Sanger and 454 pyrosequencing Total sequences 3,228	(Woyke *et al.,* 2009)
Coastal mesocosm on a flotilla	Bergen, Norway	Combined metagenomics and metatranscriptomics to examine the genetic diversity and gene expression of coastal mesocosm.	Pyrosequencing 454 GS-FLX Total sequences 1,498,577	(Gilbert *et al.,* 2008)
Coastal seawater & mesocosm	Western English Channel (WEC) & Bergen Norway	Combination of fosmid libraries from WEC water sample and metagenomic data from mesocosm in Bergen, Norway to demonstrate the ubiquity and importance of phosphonate-degradation bacteria in coastal marine ecosystems.	Fosmid library construction and sequencing Pyrosequencing for transcript analysis	(Gilbert *et al.,* 2009)
Oxygen-deficient deep seawater	Saanich Inlet, British Columbia, Canada	Sequencing of a fosmid library from oxygen minimum water indicated an extensive metabolic repertoire from the uncultured SUP05 gammaproteobacterial lineage. (e.g. reduction of nitrate to fuel autotrophic carbon assimilation via the Calvin-Benson-Bassham cycle)	Fosmid library construction and sequencing Total fosmid end sequences 243,264	(Walsh *et al.,* 2009) (Zaikova *et al.,* 2010)
Hydrothermal chimney biofilm	Lost City Hyderotherm-al Field, mid-Atlanic ridge.	DNA from the biofilm was cloned into the pUC18 vector and randomly end sequenced. Considerable abundance of transposases was determined from the community gene profile.	Vector cloning and random end sequencing Total end sequences 46,316	(Brazelton and Baross 2009)
Marine sediments	South China Sea	Construction of fosmid library and screening for ability to alter the phenotype of the host organism, *E. coli.* One clone producing melanin was fully sequenced.	Sequencing of 4-kb insert DNA	(Huang *et al.,* 2009)

Intertidal flat sediment of coast region	Saemankum, South Korea	Construction of fosmid library to identify lipase encoding genes. Directed study to identify biotechnologically revelent genes.	Random shotgun sequencing of 30-kb insert DNA	(Kim *et al.,* 2009)
Open ocean	Hawaii Ocean Time Series (HOT) station ALOHA in North Pacific Subtropical Gyre	Metagenomic sequence database complementary to a metatranscriptomic database. Data demonstrated that *Cyanobacteria* and unknown bacterial taxa contributed the largest fraction of gene transcripts.	Pyrosequencing 454 GS20 Total sequences 414,323 (DNA) 128,324 (cDNA)	(Frias-Lopez *et al.,* 2008)
Open ocean Surface water	North Pacific and South Pacific subtropical gyres	Pyrosequencing approach produced 1.1 million reads and demonstrated highly significant differences in gene profiles between samples in gyre ecosystems.	Pyrosequencing 454 GS FLX Total sequences 11,096 (cDNA)	(Hewson *et al.,* 2009)
Open ocean (15 meter)	Station ALOHA North Pacific	Pair-end shotgun sequencing to assemble a complete genome of a unicellular nitrogen-fixing *Cyanobacterium* from the UCYN-A clade.	Pyrosequencing 454 GS FLX	(Tripp *et al.,* 2010)

1.3.3 Sequencing technology and functional annotation of metagenomic sequence data

Apart from the well-known Sanger sequencing technology, new sequencing technologies are having a substantial impact on the field of metagenomics. These next-generation sequencing (NGS) technologies are represented by three commercial platforms, namely the 454 pyrosequencer (Roche Diagnostic Corporation), the GAII/ HiSeq sequencer (Illumina) and the SOLiD instrument (Life Technologies Corporation). The technologies behind each platform were reviewed in detail by Metzker (2010). The shifts from Sanger to NGS technologies has greatly improved metagenomic studies (reviewed in (Gilbert and Dupont 2011, Heidelberg *et al.,* 2010)). First, the lower cost per base of NGS technologies meant that many previous cost-prohibitive studies could be carried out. Second, the high throughput of NGS allowed for deep sequencing of environmental DNA and this has revealed an unprecedented amount of information associated with the genetic structure of particular environmental samples. A large number of publications on metagenomics and metatranscriptomics (Table 1.1) have used 454-pyrosequencing, particularly so for samples from the marine environment (e.g (Dinsdale *et al.,* 2008a, Frias-Lopez *et al.,* 2008, Gilbert *et al.,* 2008, Gilbert *et al.,* 2009, Palenik *et al.,* 2009, Tripp *et al.,* 2010). Illumina sequencing has also been used in

metagenome studies of the microbiota of the human oral (Lazarevic *et al.*, 2009) and the gastro-intestinal tract (Qin *et al.,* 2010). More recently, SOLiD sequencing has been used in metatranscriptomic studies of the phototrophic microbial mat community of Mushroom Spring (Klatt *et al.,* 2011, Liu *et al.,* 2011).

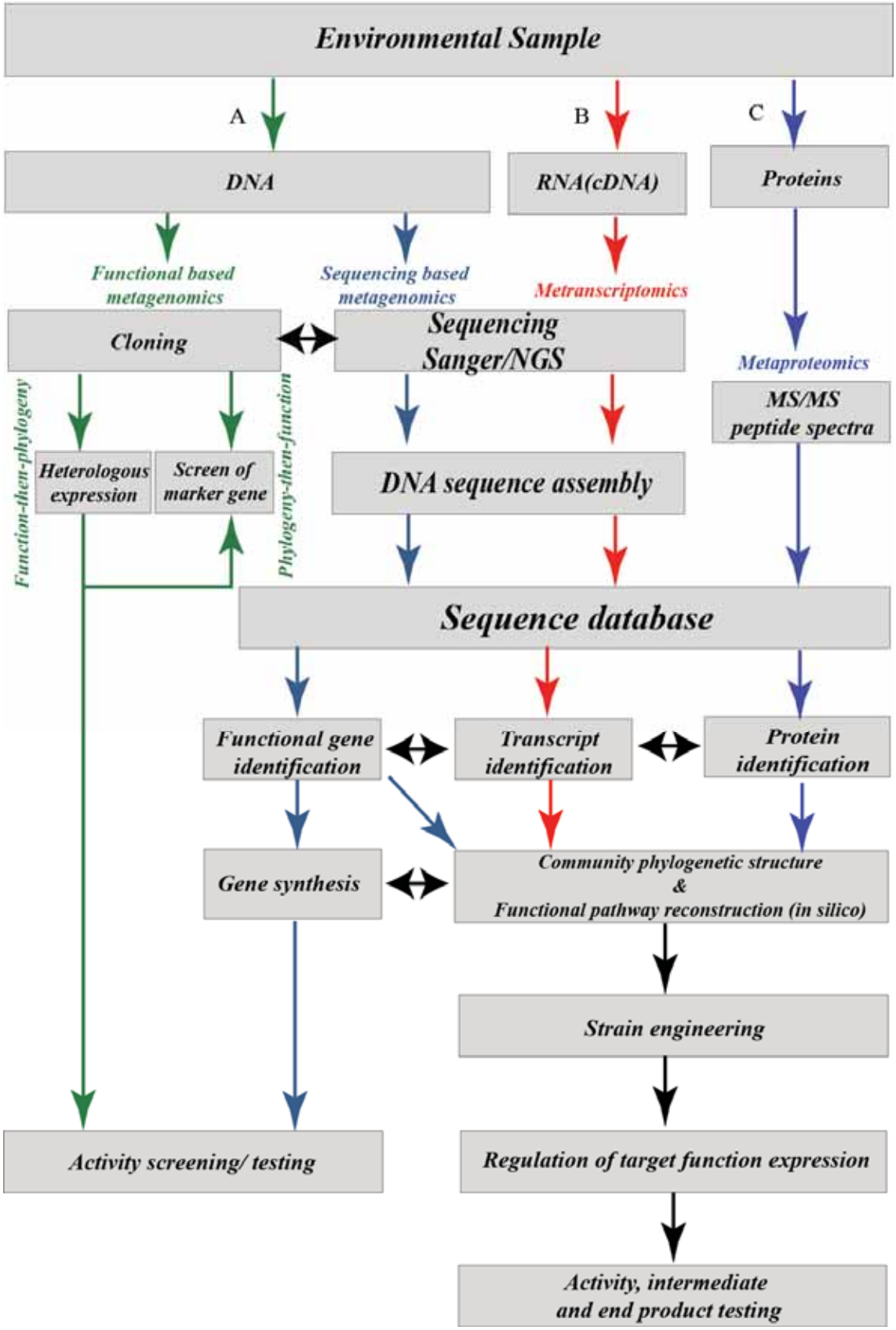


Figure 1.1 Work flow of metagenome based microbial community function analysis. Metagenomics (A), metatranscriptics (B) and metaproteomics (C). Modified from Chistoserdova (2010).

1.3.4 Metatranscriptomics

Because of the developments in sequencing technologies, metatranscriptomics by intensive cDNA sequencing has emerged as a powerful tool to study gene transcription in complex microbial communities. In combination with high throughput sequencing technology, this technique allows a deeper insight into those genes of the community that are actively transcribed and thus most probably belonging to the gene pool that is functional at the time of sampling. Several studies from a variety of environments (e.g. Gilbert *et al.,* 2008, Hewson *et al.,* 2009, Klatt *et al.,* 2011, Liu *et al.,* 2011, Poretsky *et al.,* 2009, Shi *et al.,* 2009, Stewart *et al.,* 2010, Urich *et al.,* 2008) have not only shown its usefulness, but also the advantages in defining community structure (Urich *et al.,* 2008) and identifying mRNAs of novel genes (Gilbert *et al.,* 2008). The use of reverse transcribed total RNA allows the determination simultaneously of both the structure of the active community and the gene expression (the "double RNA approach", (Urich *et al.,* 2008)). With this approach, the actively transcribing genes of a community can be studied in a single preparatory and sequencing procedure at both levels (taxonomy and function), while it allows for validation of taxonomic assignments of mRNA. Furthermore, it makes it possible to study the composition of even relatively unexplored microbial communities, with respect to all three domains of life simultaneously (*Bacteria*, *Archaea* and *Eukarya*), and it is independent of PCR-based methods that potentially introduce biases.

Despite the promise of this approach, major limitations also exist. First, extraction of RNA directly from an environmental sample is often problematic and the concentration is often low. For this reason, previous studies have used additional amplification steps to increase the concentration of initial transcripts (Frias-Lopez *et al.,* 2008, Gilbert *et al.,* 2008). Second, separation of mRNA from the abundant non-mRNA (e.g. ribosomal or transfer RNA) is also problematic and, as a result, the gene expression profile of the sample often remains limited. Consequently, the low gene expression profile is insufficient to provide satisfactory coverage or to infer statistically meaningful transcription patterns under different conditions, for most genes in a complex community. Because of this problem, previous studies have only focused on the expression by the most dominant members of the community.

1.3.5 Bioinformatics of metagenomic and metaproteomic data sets

In parallel to the development of sequencing technologies, bioinformatics tools for sequence assembly, gene prediction, functional annotation and statistical comparison have also been developed. These bioinformatics tools have been extensively reviewed in Kunin *et al*. (2008). Typically, the raw metagenomic sequences can be annotated by directly searching against databases of annotated functional proteins, like the National Center for Biotechnology Information's nucleotide database (nr, http://www.ncbi.nlm.nih.gov/nuccore). Alternatively, the raw metagenomic sequences can be clustered using CD-hit (Li and Godzik 2006), allowing sequences with similar homology to be grouped together. A representative from each clustered group can then be annotated. The raw metagenomic sequences can also be subjected to gene prediction using software such as orf-finder (open reading frame finder, http://www.ncbi.nlm.nih.gov.) and MetaGene (Noguchi *et al.,* 2008). Once the proteins have been established, they can again be subjected to clustering and annotation against different databases like TIGRFAM (Haft *et al.,* 2003), PFAM (Finn *et al.,* 2010) or Cluster of Orthologous Group (COG) (Tatusov *et al.,* 2003). In addition, several online tools and software including the MG-RAST annotation platform (Meyer *et al.,* 2008), the CAMERA 2.0 portal (http://www.camera.calit2.net) and the newly updated version of software MEGAN 4.0 (MEtaGenome Analyzer)(Huson *et al.,* 2007) have also been made available for rapid analysis of metagenomic datasets.

1.4 Metaproteomics

Metaproteomics is the analysis concerned with the identification and assessment of function of the proteins from environmental samples. It directly addresses the microbial functional profile. In addition to the rapid development witnessed in the field of mass spectrometry, the enormous increases of genomic and metagenomic data and improvement in computing power and bioinformatics provide a more solid basis for protein identification (Schneider and Riedel 2010). Because of the important interdependency between metagenome and metaproteome data, the term proteogenomics is also used to describe metaproteomic studies (e.g. Delmotte *et al.,*

2009, Denef *et al.*, 2010, Ng *et al.*, 2010). The development and application of mass spectrometry in proteomic analysis has been extensively reviewed by Aebersold and Mann (2003).

Typical metaproteomic analyses comprise four basic steps: i) sample preparation including protein extraction, purification and concentration, ii) protein denaturation and reduction, iii) protein or peptide separation, digestion and MS analysis, iv) protein identification based on the obtained MS and or MS/MS data (Schneider and Riedel 2010). It is crucial to metaproteomic analysis that the quality and quantity of proteins extracted are representative of the sample collected (Maron *et al.*, 2007). Therefore, careful development in sample preparation, protein separation, MS analysis and protein identifications have been developed, as reviewed by Schneider and Riedel (2010).

1.4.1 Metaproteomics of microbial communities

Recently, several important studies to characterise proteins extracted from the natural environments have been published (Table 1.2). The metaproteomic analysis of the AMD biofilm system was one of the first conducted (Ram *et al.*, 2005). Taking advantage of the relative simple community structure and the near complete genomes of its dominant members, the analysis identified over 2000 proteins, from the five abundant organisms within the community. Proteins involved in refolding and oxidative stress were identified to be important for the adaptation of the microbial community to an acidic, metal-rich environment (Ram *et al.*, 2005). Half of the proteome of the *Leptospirillum* group II, the most abundant organism in the sample, was identified. A significant finding of this study was the possibility of directly localizing the *Leptospirillum* group II proteins (based on detection of proteins in the extracellular or membrane fraction) and inferred their abundance back into the genomic context (VerBerkmoes *et al.*, 2009a). By combining abundance, localization and genomic context information, it was possible to formulate hypotheses regarding the function of some highly abundant proteins of unknown function. Later studies identified and purified two novel cytochromes (Cyt572 and Cyt579), which were central to iron oxidation and AMD formation (Jeans *et al.*, 2008, Singer *et al.*, 2008). Recently, a strain-resolved community study was published, where community genomic data was

used for the identification of proteins from dominant community members with strain specificity (Lo *et al.*, 2007). This study provided important evidence that gene exchange occurred during adaptation to specific ecological niches.

Table 1.2 Overview of proteomics approaches to study structure and function of microbial communities.

Microbiome	Number of peptides identified	Number of proteins identified	Protein/peptide separation method and MS platform	Peptide identification method	Reference
Sludge	NA	46	2D-PAGE, MALDI-ToF, QToF, MS/MS	Spectral matching	(Wilmes and Bond 2004)
Ocean	184	NA	2D-PAGE, 2D Nano-LC LTQ, MS/MS	Spectral matching + *De novo*	(Powell *et al.,* 2005)
Acid mine drainage biofilm (AMD)	6188	2033	2D Nano-LC LTQ, MS/MS	Spectral matching	(Ram *et al.,* 2005)
Lake and Soil	NA	513	2D Nano-LC QToF, MS/MS	Spectral matching	(Schulze *et al.,* 2005)
Estuary	7	3	2D-PAGE, LC QToF, MS/MS	*De novo*	(Kan *et al.,* 2005)
Ocean (SAR11)	3	1	1D-PAGE, MALDI-TOF	Spectral matching	(Giovannoni *et al.,* 2005)
Trophosome of *Riftia pachyptila*	NA	220 (493)	2D- PAGE, 1D-PAGE, 2D Nano-LC, MALDI-ToF, QToF, MS/MS	Spectral matching	(Markert *et al.,* 2007, Markert *et al.,* 2011)
Infant gastrointestinal tract	11	1	2D-PAGE, MALDI-ToF, MS	*De novo*	(Klaassens *et al.,* 2007)
AMD (strain-resolved)	8137	3234	2D Nano-LC LTQ, MS/MS	Spectral matching	(Lo *et al.,* 2007)
Waste water treatment reactor	NA	109	2D-PAGE, MALDI-ToF, MS	Spectral matching + *De novo*	(Lacerda *et al.,* 2007)
Contaminated soil/ground water	NA	59	1D-PAGE, 2D-PAGE, LC, MS/MS	Spectral matching	(Benndorf *et al.,* 2007)
Sludge	4472	2378	2D Nano-LC LTQ, Orbitrap, MS/MS	Spectral matching	(Wilmes *et al.,* 2008a, Wilmes *et al.,* 2008b)
Sludge (EPS)	50	10	1D-PAGE, LC 4000 Qtrap, MS/MS	Spectral matching	(Park *et al.,* 2008)

Ocean (low nutrient)	6533	1042	2D Nano-LC LTQ, MS/MS	Spectral matching	(Sowell *et al.,* 2009)
Acid mine drainage biofilm (AMD)	NA	2752	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Denef *et al.,* 2009)
AMD (strain-specific)	NA	2382	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Goltsman *et al.,* 2009)
Human gut microbiome	NA	2214	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Verberkmoes *et al.,* 2009b)
Leaf phyllosphere	12345	2883	1D-PAGE, LC, LTQ-Orbitrap, MS/MS	Spectral matching	(Delmotte *et al.,* 2009)
Sheep rumen	NA	4	1D PAGE, LC-ESI, LCQ, MS/MS	Spectral matching	(Toyoda *et al.,* 2009)
Contaminated aquifer sediment	57	23	2D-PAGE, 2D Nano-LC LCQ, MS/MS	Spectral matching + *De novo*	(Benndorf *et al.,* 2009)
Ground water	NA	> 2500	2D LC, Orbitrap, MS/MS	Spectral matching	(Wilkins *et al.,* 2009)
Ace lake, Antarctica	3970	504	1D-PAGE, LC, LTQ, MS/MS	Spectral matching	(Ng *et al.,* 2010)
Ocean	5389	2273	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Morris *et al.,* 2010)
Ground water (Toulene-degrading community)	NA	202	2D-PAGE, LC/MSD TRAP, MS/MS	Spectral matching	(Jehmlich *et al.,* 2010)
Fungal-bacterial consortium	NA	369	2D-PAGE, MALDI-ToF, MS + 1D-PAGE, LC_ESI, MS/MS	Spectral matching	(Moretti *et al.,* 2010)
Human salivary microbiome	257	NA	3D-peptide fractionation*, LC LTQ, MS/MS	Spectral matching	(Rudney *et al.,* 2010)
Ocean (costal upwelling system)	7151	13469^	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Sowell *et al.,* 2011)
AMD (Lab cultivated)	NA	2148	2D Nano-LC LTQ-Orbitrap, MS/MS	Spectral matching	(Belnap *et al.,* 2011)

Acid mine drainage contaminated with Arsenic	NA	>500	1D-PAGE, LC, Q-ToF, MS/MS	Spectral matching	(Bertin *et al.,* 2011)
Hindgut of terminte (*Nasutitermes corniger)*	NA	886	SCX fractionation, LTQ-Orbitrap, MS/MS	Spectral matching	(Burnum *et al.,* 2011)
Crop rhizospheric soil	NA	122	2D-PAGE, MALDI-ToF, MS	Spectral matching	(Wang *et al.,* 2011)
Waste water treatment reactor	NA	33	2D-PAGE, nLC-ESI, MS/MS	Spectral matching	(Abram *et al.,* 2011)

* 3D-peptide fractionation: a three-stage process, where the first step was preparative isoelectric focusing using a free-flow electrophoresis system (FFE). All preparative isoelectric fractions were subjected to a preliminary MS/MS analysis to determine which fractions contained the most complex peptide mixtures. Peptides within each of the complex fractions were then further separated according to their absolute number of basic amino acid residues, using strong cation exchange step gradient chromatography (SCX). ^ 7151 distinct peptide sequences (typically 5–25 amino acids in length) were mapped to 13 469 eCDSs (full- or nearly full-length translated environmental protein-coding sequences). Many peptides matched to multiple eCDSs and many eCDSs had similar functions.

Table modified from VerBerkmoes *et al.* (2009) and Schneider and Riedel (2010).

Metaproteomic analysis has also been applied to estuary and ocean water (Table 1.2). Kan and colleagues compared protein profiles from various origins of the Chesapeake Bay using 2D PAGE and identified protein spots excised from the gel by LC-MS/MS. The information gained was rather limited, because of the relatively poor resolution of 2D PAGE and the small sequence database available at the time (Kan *et al.*, 2005). A recent study of the dominant population (SAR11, *Prochlorococcus* and *Synechococcus*) in the surface of the Sargasso Sea showed a high number of periplasmic substrate-binding proteins from SAR11 (Sowell *et al.,* 2009). The highest abundance of proteins was observed for periplasmic substrate-binding proteins, for phosphate, amino acids, phosphonate, sugars and spermidine (Sowell *et al.*, 2009). This high abundance of SAR11 transporters was proposed to reflect the means to maximise nutrient acquisition in the highly nutrient-depleted environment. The high abundance of transporter proteins was also observed in a later study from a highly productive coastal upwelling system (Sowell *et al.*, 2011). In contrast to the nutrient-depleted water, this analysis focused on the survey of a metaproteome from nutrient-rich summer surface waters from the Oregon shelf (Sowell *et al.,* 2011). The different transporters for the amino acids, taurine, polyamine and glutamine synthetase indicated that carbon and nitrogen are more limiting compared to phosphate in nutrient rich environment, which supported the developed ecological theory for the region that Oregon shelf water is replete with

phosphorus (Sowell *et al.,* 2011). One of the most important characteristics of metaproteomic analysis by these recent studies was the provision of taxon-specific confirmation of the expression of the genomic potential. For example, this was the case for the confirmation of SAR11 proteorhodopsin expression in seawater and the obligate oxidation of C1 compounds by the *Betaprotebacterium* (OM43 clade MDH) (Sowell *et al.,* 2009, Sowell *et al.,* 2011). Using protein-centric comparative metaproteomic approach focused on membrane protein, Morris and colleagues (2010) studied plankton in South Atlantic surface water from a low-nutrient gyre to highly productive coastal upwelling region. They reported variation in the proteins for nutrient utilization and energy transduction along the environmental gradient. The importance of such a finding is that the relative abundance of the proteins or peptides detected, for a specific function, can help to reveal the nutrient status of cells in marine waters (Morris *et al.,* 2010, Sowell *et al.,* 2011).

A number of metaproteomic analyses of other complex systems, like the human microbiome, wastewater sludge, hindgut microbiome and phyllosphere bacteria (Table 1.2) have also been reported. Such studies further demonstrated the usefulness and power of a metaprotomics approach to study diverse microbial ecosystems.

1.5 Sponge-associated bacteria-host interactions

All living multicellular organisms, plants, insects and animals have associated microorganisms. The co-existence of a host-microbial community can have beneficial or deleterious effects on the host and the interactions between the microorganisms and the host spans the spectrum from parasitic, to commensal, to mutualistic. The terms ‘symbiosis’ and ‘symbiont’ are used throughout this thesis with their broadest possible definition, to describe any close, permanent and long-term relationship between two or more biological species. Symbiotic microbial communities can thrive on the surfaces and in the tissue of diverse multicellular organisms, for example on the surface of plants (Andrews and Harris 2000, Whipps *et al.,* 2008), on human skin (Grice *et al.,* 2009) and in the human gut (Eckburg *et al.,* 2005, Turnbaugh and Gordon 2009). Marine plants and invertebrates like corals, sponges and algae also provide a wide range of potential habitats for microbes in the ocean (Hentschel *et al.,* 2002, Olson and Kellogg 2010,

Rohwer *et al.,* 2002, Taylor *et al.,* 2007). Long term association between the microbial communities and their host normally indicate a positive relationship. One of the best characterised host-microbe associations in the marine environment is that of the luminescent bacterium *Vibrio fischeri* and its host, the squid *Euprymna scolopes*, where the relationship is beneficial for both (Visick and McFall-Ngai 2000). *V. fischeri* can colonise juvenile squid within hours of their hatching and utilise the rich mixture of amino acids provided by the host (Graf and Ruby 1998). The colonizing bacteria can cause major morphological changes that lead to the development of a light organ. The luminescent symbiont within the light organ then provides light during darkness, which is believed to mimic moonlight and to prevent detection of the squid by a predator below (Visick and McFall-Ngai 2000).

Recent phylogenetic studies have suggested that marine sponges harbour specific, stable microbial communities, that are distinct in composition from those of the surrounding seawater (for example: (Hentschel *et al.,* 2002, Santavy *et al.,* 1990, Taylor *et al.,* 2004, Webster *et al.,* 2001b, Wilkinson 1978b). The existence of such distinct communities is remarkable given that the sponge bacterial community is constantly exposed to large numbers of filtered, planktonic bacteria (Vogel 1977). Microbial communities can comprise as much as 40% of sponge biomass (Friedrich *et al.,* 1999, Hentschel *et al.,* 2006, Taylor *et al.,* 2007), with densities of between 10^8 and 10^{10} bacteria per gram of wet weight sponge (Hentschel *et al.,* 2006). The associated microbial communities in sponges have been recognised to contribute to the host sponge biology in a variety of ways, including for the provision of nutrients (Wilkinson and Garrone 1980), enhancement of chemical defenses (Unson *et al.,* 1994), contribution to structural rigidity (Wilkinson 1978a), and transport of waste compounds or active metabolites (Wilkinson 1978a, Wilkinson 1983).

Sponges are also a rich source of biologically active natural products with promising industrial and medical applications (Taylor *et al.,* 2007). It has been established that more novel bioactive metabolites are obtained from sponges each year than from any other marine taxa (Munro *et al.,* 1999). Many of the natural products so far identified bear structural similarities to bacterial compounds and therefore have been proposed to be potentially produced by associated microbial symbionts (Newman and Hill 2006). Examples of those natural products include jaspamide from the sponge *Jaspis* spp.

(Zabriskie *et al.*, 1986) and the cyclodepsipeptide chondramide D isolated from the myxobacterium *Chondromyces crocatus* (Erickson *et al.*, 1997) and the myxobacterial metabolite apicularen A, which is almost identical to salicylihalamide A from *Haliclona* spp. (Kunze *et al.*, 1995).

Despite the recent increased interest in sponge associated microbial communities, and in particular their diversity, the microbial function and factors that influence the host and the surrounding environment is still not well understood. The following sections, provide an overview of the biology of marine sponges, sponge-microbe interactions and current knowledge on the model organisms used in this study, *Cymbastela concentrica* and *Rhopaloeides odorabile* (see Figure 1.2).

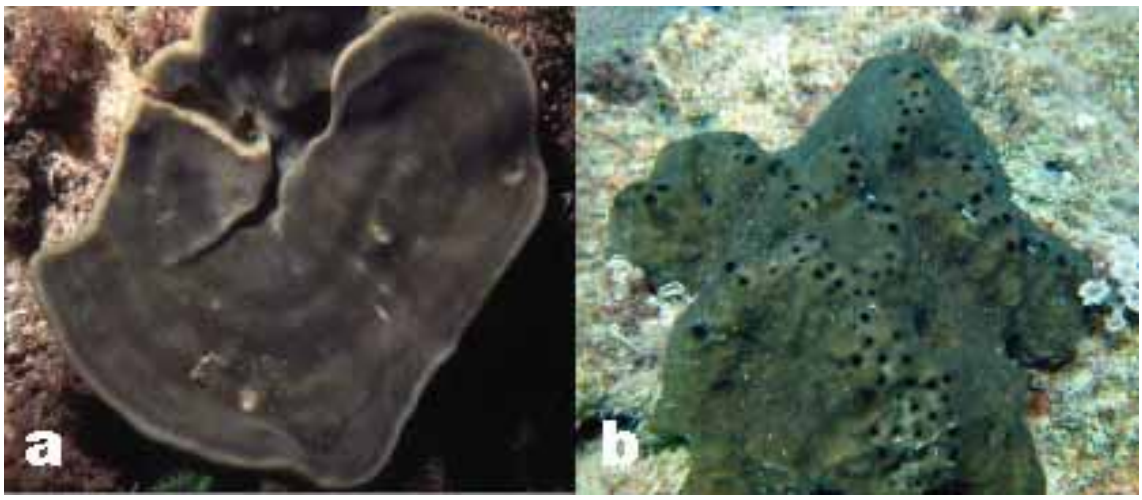


Figure 1.2 *Cymbastela concentrica* (a) at Bare Island, La Perouse, Sydney, NSW and *Rhopaloeides odorabile* (b) at the Great Barrier Reef, QLD, Australia

1.5.1 General background and basic biology of sponges

Sponges (phylum *Porifera*) are sessile filter feeders that remove bacteria from surrounding sea water by pumping large volumes of water (e.g. up to 24 cubic meter of water per kilogram of sponge per day (Kennedy *et al.*, 2007, Vogel 1977) through their aquiferous system. They are among the oldest of the multicellular animals (*Metazoa*) with fossil records from as early as 600 million years ago (Borchiellini *et al.*, 2001). Today, marine sponges represent a significant component of benthic communities throughout the world, in both biomass and their potential to influence benthic or pelagic processes (Taylor *et al.*, 2007). More than 6000 species of sponges have been identified

from both shallow and deep water populations, occupying as much as 80% of available surfaces in some areas (Dayton 1989, Taylor *et al.*, 2007).

Sponge architecture is unlike that of any other taxa, and sponge morphology greatly affects many aspects of their biology, including interactions with microorganisms (Taylor *et al.,* 2007). The basic structure of sponges comprises several different cell layers (Figure 1.3). The outer layer (pinacoderm) is formed by epithelial cells known as pinacocytes. The pores on the surface of the sponges are called ostia and these cells extend along the interior canals that permeate the sponge. Inside the sponge, a series of chambers (choanoderms) are formed by the flagellated cell, choanocytes, and they are responsible for pumping the water into the sponge through the ostia. The feeding process also takes place within these chambers, where choanocyts filter out food particles (including bacteria and microalgae) from the water and transfer them to the mesohyl. The transferred particles in the mesohyl are eventually digested via phagocytosis by archaeocytes (Reiswig 1971), a group of totipotent cells capable of differentiating into any other sponge cell type. In the feeding process, some bacteria pumped into the sponge and transferred into the mesohyl may survive and can establish themselves as part of the sponge-specific microbiota (Kennedy *et al.,* 2007, Taylor *et al.,* 2007). At the end of the feeding process, the water is expelled from the sponge via the exhalant opening or osculum.

Despite their simple body plan, sponge morphology is diverse. They come in many different shapes, sizes and colors (Bergquist 1978). Many of these morphologies directly reflect their ecological function. For example, photosynthetic cyanobacterium-containing sponges are often flat, thus optimizing light reception (Sara *et al.,* 1998, Wilkinson 1983). The totipotent nature of the sponge cell means that they are capable of re-aggregation following dissociation, making them capable of regeneration or remodeling after partial mortality. The structural integrity and mineral skeleton of the sponge is either siliceous or calcareous, usually in the form of discrete elements, spicules, which have a bewildering range of shapes, sizes and patterns of organization (Bergquist 2001). Collagenous tissues, such as spongin, also have a role in providing structural support and together with spicules, allow the development of very large individuals, such as those found among many tropical species (Hooper and van Soest 2002)

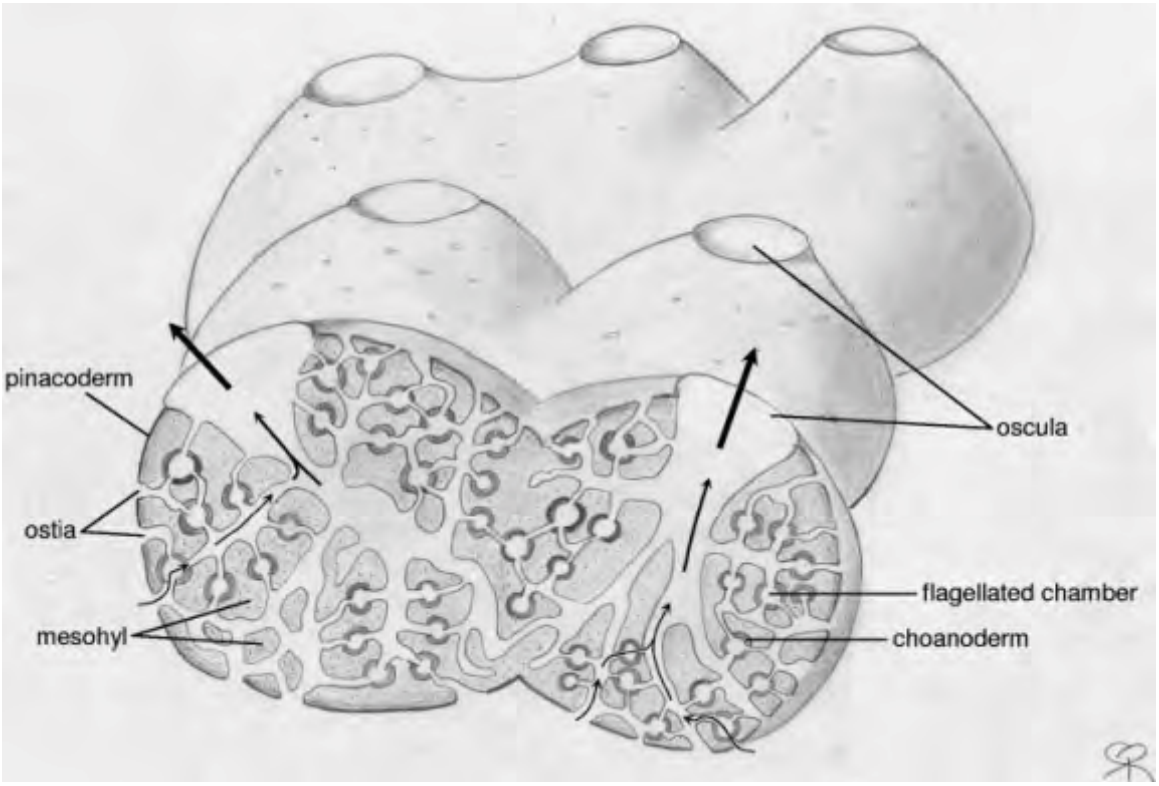


Figure 1.3 Schematic representation of a sponge.
Arrows indicate the direction of water flow through the sponge. Adopted from Taylor *et al*. (2007).

1.5.2 Biodiversity of sponge-associated microorganisms

Thirty bacterial phyla and two Archaea phyla have been reported from sponges (Lee *et al.,* 2011, Taylor *et al.,* 2007, Webster and Blackall 2008, Zhu *et al.,* 2008). The extensive information available on the biodiversity of the microbial communities within sponges was comprehensively reviewed by Taylor *et al*. (2007). In general, past studies have shown that many of the known sponge-associated microbes are unique to sponges (Hentschel *et al.,* 2002, Taylor *et al.,* 2007). Among these sponge-specific microbes, some are common in different types of sponges from different geographical locations, while some are only specific to certain sponges. Furthermore, stable bacterial-sponge association has also been demonstrated over temporal and spatial scales in certain sponges (Hentschel *et al.,* 2002, Hill *et al.,* 2006, Taylor *et al.,* 2005).

Both culture-dependent and culture-independent approaches have been used to study the diverse sponge microbial community. Morphologically diverse bacteria have been

isolated from various marine sponges through different cultivation approaches (Hentschel *et al.,* 2001, Olson *et al.,* 2000, Santavy *et al.,* 1990, Webster and Hill 2001). Development in molecular tools, such as 16S rRNA gene based community fingerprinting, has enabled the discovery of exceptional diversity in sponge associated microbial communities and demonstrated that the species composition of the cultured bacterial communities differ from the uncultured community. For example, only approximately 0.1% of the total community in the sponge *R. odorabile* was estimated as culturable (Webster and Hill 2001). A later study using additional media and supplements improved the culturable proportion to 5% for two deep-water *Scleritoderma* spp. sponges (Olson and McCarthy 2005), similar to the 3.4 to 11% bacterial recoverability found by Santavy *et al*. (1990), which also employed a variety of growth media to cultivate bacteria from a *Certoporella nicholsoni* sponge (Santavy *et al.,* 1990). These findings further support the notion that the majority of microorganisms are not easily cultured using standard microbiological techniques (Rappe and Giovannoni 2003).

The use of culture-independent tools has greatly accelerated the understanding of the phylogeny of sponge-associated microbes. Underpinning this development is the application of next-generation sequencing technologies. A recent study using 454 amplicon pyrosequencing revealed extraordinary species richness in Australia's Great Barrier Reef sponges (e.g. *Ianthella basta, Ircina ramose* and *R. odorabile*) (Webster *et al.*, 2009). Despite the enormous diversity demonstrated in those sponges, the result, at least at the phyla level, did not alter the established understanding of the major microbial community members, i.e. the study found that the dominant bacterial taxa were *Chloroflexi*, *Acidobacteria*, *Actinobacteria*, and *Proteobacteria* (*Alpha, Delta, Gamma*), which are the same phyla that featured in conventional 16S rRNA gene libraries (Taylor *et al.,* 2007). Newly discovered phyla by 454 sequencing (e.g. *Deferribacters, Tenericutes*) were also present, but only at low abundance (Webster *et al.,* 2009).

1.5.3 Symbiont Acquisition and Transmission

A further aspect of microorganisms associated with sponges is the acquisition and transmission of bacterial symbionts. For example, how do sponges acquire the majority of their symbionts (vertically or environmentally) and how can distantly related species from geographically isolated regions acquire shared bacterial symbionts? A review by Taylor *et al.* (2007) has suggested a number of scenarios including ancient symbiosis maintained by vertical transmission, parental and environmental symbiont transmission and environmental acquisition. The vertical transmission of sponge symbionts was proposed fifty years ago through microscopic observations (Levi and Porte 1962) and more recently has been documented in studies using different molecular techniques including DGGE, 16S rRNA gene sequencing and FISH (Enticknap *et al.*, 2006, Lee *et al.*, 2009, Schmitt *et al.*, 2007, Schmitt *et al.*, 2008, Sharp *et al.*, 2007, Steger *et al.*, 2008). A recent study using 16S rRNA gene based amplicon pyrosequencing demonstrated that nearly 50% of the sponge-specific sequence clusters (SSSC) can be found in both the adult and larvae of the same sponge, implying vertical transmission of these groups (Mukherjee *et al.*, 2009). Besides these shared taxa, some lineages including '*Poribacteria*' were also found with low abundance among the tags retrieved from surrounding seawater. It suggests that the rare seawater biosphere can act as a seed bank for sponge-specific microbes and that environmental transmission may also play a significant role in symbiont acquisition by sponges, in particular for juveniles (Webster *et al.*, 2009).

1.5.4 Sponge symbiotic function

The current understanding on the function of sponge-associated microorganisms is lagging behind the understanding on their taxonomic affiliation. However, symbiont function has become a major focus of recent studies, in particular for nitrogen metabolism. For example, stable isotope experiments have demonstrated denitrification in sponges *Geodia barretti* (Hoffmann *et al.*, 2009), *Dysidea avara* and *Chondrosia reniformis* (Schlappy *et al.*, 2010). Further analysis with *G. barretti* recovered two genes from *Betaproteobacteria* and *Gammaproteobacteria* potentially encoding cytochrome cd1-type nitrite reductases (*nirS*) that are responsible for denitrification in this sponge (Hoffmann *et al.*, 2009). Genes encoding enzymes involved in

denitrification, such as nitrite reductase and nitric oxide reductase, were also identified in the genome of uncultivated '*Poribacteria*' from the sponge *Aplysina aerophoba* (Siegl *et al.,* 2011). Anammox activity was also demonstrated in the sponge *G.barretti* (Hoffmann *et al.,* 2009) and the presence of 16S rRNA gene sequences affiliated with *Anammox planctomycete* was also identified in the sponge *Mycale laxissima* (Mohamed *et al.,* 2010). The latter study provided the first molecular evidence for the presence of potential anammox bacteria in sponges, suggesting that sponges are a source from which to enrich and cultivate novel anammox bacteria (Mohamed *et al.,* 2010).

Photosynthesis is another well-studied symbiotic function in sponges. The host benefits from the provision of photosynthetically fixed carbon from its symbionts, mostly from *cyanobacteria* (Cheshire and Wilkinson 1991, Cheshire *et al.,* 1997, Steindler *et al.,* 2002, Wilkinson 1979b, Wilkinson 1983, Wilkinson 1987). Photosynthesis is particularly important for "phototrophic" sponges as photosynthetically fixed carbon from cyanobacterial symbionts can provide more than 50% of the energy requirements in certain tropical sponges (Wilkinson 1983). Translocation of photosynthates (e.g. glycerol) from *cyanobacteria* to the host has also been shown for marine sponges (Wilkinson 1979b), while glucose produced by a chlorella-like green alga has also been shown to be passed to its freshwater sponge host, *Ephydatia fluviatilis* (Wilkinson 1980). The importance of photosynthesis as a symbiotic function is also reflected through the geographical distribution of sponges. Phototrophic sponges comprise nearly half of the total sponge biomass on the outer reefs of the Great Barrier Reef, where the water is clearer (Wilkinson 1987, Wilkinson and Cheshire 1989). Great Barrier Reef sponges located at depths of 15 to 30 meters may still derive much of their nutrients from photosynthetic symbionts, because of the reduced light attenuation in clearer water (Cheshire and Wilkinson 1991). *R. odorabile* is a common, abundant reef-associated sponge that colonises both inshore and offshore reefs at the Great Barrier Reef (Bannister *et al.,* 2010, Wilkinson and Cheshire 1989), such that the habitat distribution of this coral reef sponge correlates with light availability. Furthermore, the culturable microbial community of *R. odorabile* has been isolated and includes a unique cyanobacterium within the order *Oscillatoriales* (Webster and Hill 2001), suggestive of photosynthesis.

In contrast to the hypothesis on photosynthesis regulated localization, a recent study which conducted photorespirometry trials, showed no evidence of photosynthesis in *R. odorabile* (Bannister *et al.,* 2011). Furthermore, no photopigments were present in these sponges and no cyanobacteria could be detected within the tissue. The results did not vary between sponges collected from nutrient rich inner- and mid-shelf reefs, or from oligotrophic outershelf reefs. The findings demonstrated that *R. odorabile* is not a phototrophic sponge and that habitat distribution that clearly correlates with light is likely to be due to factors other than photosynthesis. This agrees with earlier observations that phototrophic sponges are largely absent from Caribbean reefs, where only low numbers of sponge-associated *cyanobacteria* are present (Wilkinson 1987).

Sponges have also been recognised to be capable of differentiating food bacteria from symbionts. Early studies by Wilkinson and colleagues, using the shallow-water sponge *Aplysina aerophoba* suggest that the chemical composition of the bacterial outer layer may play a role in sponge symbiont recognition (Wilkinson *et al.,* 1984). However, a later study by Wehrl *et al.* (2007), using the same sponge demonstrated that the retention rates of different food bacteria were similar, despite the difference in cell surface properties. Therefore, the long-standing question of how sponge discriminate between food and symbionts remains unsolved. Recent analysis on the metagenome of *C.concentrica* has revealed a number of important genomic factors including tetratricopeptide repeat domain encoding proteins (TPR) and ankyrin repeat proteins (AR). The presence of the TPR and ARP proteins in the *C. concentrica* metagenome, with signal peptides for extracellular secretion in Gram-negative bacteria indicating that they may interact with surrounding cells and proteins (Thomas *et al.*, 2010). Thus these proteins could represent potential mechanisms that allow the host sponge to discriminate between food and symbiont bacteria (Thomas *et al.,* 2010).

The use of whole-genome amplification (WGA) or metagenomics of sponge-associated microbes has revealed unprecedented insights into sponge symbiotic functions. Recently, the microbial community of sponge *A. aerophoba* was sorted by fluorescence-activated cell sorting (FACS) and then subjected to WGA (Siegl and Hentschel 2010). A cosmid library was constructed using the WGA products. The functional data revealed a sponge specific polyketide synthase (PKS) from the '*Poribacteria*' clade and a non-ribosomal peptide synthetase (NRPS) from the *Chloroflexi* clade (Siegl and

Hentschel 2010), both of which represent prominent enzyme classes that are responsible for the synthesis of bioactive compounds (Kennedy *et al.*, 2007). A more recent follow up study demonstrated that the '*Poribacteria*' are mixotrophs that can undertake autotrophic carbon fixation via the Wood-Ljungdahl pathway through the detection of carbon monoxide dehydrogenase gene (Siegl *et al.*, 2011). The carbon monoxide dehydrogenase was also identified in the metagenome of *C. concentrica* (Thomas *et al.*, 2010).

From the same metagenomic study of *C. concentrica*, a number of other important observations were made. First, a surprisingly large number of insertion elements were observed within the metagenome. Large numbers of mobile elements were previously seen within intracellular symbionts in other hosts (Wu *et al.*, 2004), and proposed to have important roles in the evolution of bacterial genomes for symbiotic relationship with their hosts. Second, a high frequency of clustered regularly interspaced short palindromic repeats and CRISPR-associated proteins (CAS) were observed in the metagenome data set. The elevated abundance of these CRISPRs and CAS were proposed to form a viral-specific defense system (Barrangou *et al.*, 2007). The defense system may provide proteins effective against bacteriolytic viral infection for the sponge symbionts, as sponge bacteria are expected to be exposed to as many as 1000 viral particles on a single day. Therefore, the system would be essential for the survival of high cell-density, non-mobile sponge microbial communities and to maintain a barrier against phage-mediated gene transfer from the surrounding planktonic community, where a high phage load could be expected. Third, several shared metabolic interactions between bacteria and host, including vitamin production, nutrient transport and utilization, and redox sensing and response were found. These new previously unrecognised genomic signatures and functions of sponge bacteria provide an insight into the potential metabolic interdependency between host and microorganisms, and into the specific mechanisms used by symbionts to remain in sponge tissue. Also, these novel markers could be used for monitoring the status of the symbiotic relationships between sponge and bacteria, by assessing the abundance, diversity and expression of these functional genes (Thomas *et al.*, 2010).

1.6 The marine sponges Cymbastela concentrica *and* Rhopaloeides odorabile

Cymbastela concentrica is an erect, cup-shaped, abundant lamellate demosponge found in shallow waters along the Australian east coast. The sponge has been previously studied with respect to its bacterial community composition (Taylor *et al.,* 2004, Taylor *et al.,* 2005). Seven divisions of bacterial groups have been identified from *C. concentrica,* including *Actinobacteria, Bacterioidates, Chloroflexi, Cyanobacteria, Nitrospira, Proteobacteria* (*Alpha, Beta, Gamma* and *Delta*) and *Verrucomicrobia. C. concentrica* is also known to harbor photosynthetic diatoms/micro-algae within its peripheral skeleton, suggesting that these symbionts may contribute to the phototrophic nature of the sponge. Previous investigations using both community fingerprinting analysis, like DGGE, and metagenomic analysis revealed that the sponge harbors a stable bacterial community over time and space and that the community structure was significantly different from that of the surrounding seawater (Taylor *et al.,* 2004, Thomas *et al.,* 2010). The recent metagenomic study on the microbial community associated with *C. concentrica* not only revealed an abundance of sponge bacteria with specific functional fingerprints (Thomas *et al.,* 2010), it also provided new valuable insight into the evolution of symbiotic diversity, microbial metabolism and host-microbe interactions in sponges.

The other model organism studied in this thesis was the Great Barrier Reef dictyoceratid sponge *Rhopaloeides odorabile*. It produces a variety of biologically active compounds like spongiatriols, tri-acetates and spongiadiols, which have a potential for pharmaceutical applications. Since the isolation and immediate supply of bioactives is often difficult (Osinga *et al.,* 1998), the isolation of bioactive from symbiotic sponge bacteria may be a more viable and sustainable option for the continued supply of these compounds. For this reason, the microbial community associated with *R. odorabile* was also extensively studied and appeared to be highly stable over time and space (Webster and Hill 2001, Webster *et al.,* 2001b, Webster *et al.,* 2008a). The microbial diversity within *R. odorabile* is extremely high and includes members of both major archaeal lineages (Webster *et al.,* 2001b) and 19 bacterial phyla including the *Proteobacteria* (*Alpha, Beta, Gamma, Delta* and *Epsilon*), *Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Nitrospira, Cyanobacteria, Planctomycetes, Deferribacteres,*

Gemmatimonadetes, *Spirochaetes*, *Tenericutes*, *Verrucomicrobia* and the candidate phyla TM7, OP1, OP11, WS3 and *Poribacteria* (Webster and Hill 2001, Webster *et al.*, 2001b, Webster *et al.*, 2008a, Webster *et al.*, 2011). A recent study assessing the community dynamic of *R. odorabile* indicated that whether the sponge was cultivated in the wild or in small flow-through aquaria, the community remained stable. However the community showed a large shift when cultivated in large-scale mesocosms for 12 months (Webster *et al.*, 2011). Furthermore, past studies have shown that environmental stressors such as heavy metal pollution (Webster *et al.*, 2001a), elevated seawater temperature (Webster *et al.*, 2008a) and disease (Webster *et al.*, 2002) could cause a disruption to the stable symbiotic microbial community, a disruption that is correlated with a decline in sponge health.

1.7 Aims

Previous investigations have greatly improved our understanding of the phylogenetic and functional diversity of microbial communities associated with the marine sponge *C. concentrica*. However, there are still questions unanswered concerning the specific phylotypes associated with the sponge and their potential function. The overall aim of this project was to further assess the phylogenetic and functional diversity of the microbial community of *C. concentrica* using a combined metagenomic and metaproteomic approach. The combined approach would help further characterise specific members within the sponge community and their potential function with their host or the environment. The metaproteomic approach developed in this project could be further used to assess the functional dynamics of the microbial communities associated with the Great Barrier Reef sponge *R. odorabile* when subjected to elevated temperatures.

1.8 Chapter synopsis

Chapter 2: To establish the genomic and functional characterization of an uncultured Deltaproteobacterium associated with the sponge *C. concentrica*. By examining the 16S rRNA gene sequence of this novel bacterium, it was demonstrated that this organism represents a novel phylogenetic clade and probably lives in close association with a

cyanobacterium. Functional analysis based on the annotated partial genome information provided an overview of the predicted functional and ecological properties of this Deltaproteobacterium, including its complex interactions with the surrounding cells and milieu, traits of cell attachment, nutrient transport and protein-protein interactions.

Chapter 3: The result of a combined metagenomic and metaproteomic approach was used to characterise the functional features of the microbial community of *C. concentrica.* The expression of specific transport functions for typical sponge metabolites such as halogenated aromatics and dipeptides were detected, which demonstrated the metabolic interactions between the microbial community and the host. Simultaneous performance of aerobic nitrification and anaerobic denitrification, which would aid to remove ammonium secreted by the sponge were detected. The analysis also highlighted the requirement for the microbial community to respond to variable environmental conditions based on the detection of an array of stress protection proteins. There was also evidence that molecular interactions between symbionts and their host could be mediated by a set of expressed eukaryotic-like proteins and cell-cell mediators. Finally, some sponge-associated bacteria (e.g. sponge *Phyllobacteriaceae* phylotype) appeared to still undergo an evolutionary adaptation process to the sponge environment as indicated by active mobile genetic elements.

Chapter 4 describes the functional genomic characterization of the *Phyllobacteriaceae* phylotype, which was identified as one of the dominant members through metagenomic analysis and was also functionally active through the metaproteomic analysis. The result from the genomic analysis of the re-constructed partial genome predicted that this organism is capable of degrading a number of aromatic compounds and carries out anaerobic respiration through nitrate reduction. The specific functional signatures identified provided unique insights into the potential mechanisms which the *Phyllobacteriaceae* phylotype use to survive in sponge and suggested a potential metabolic interdependency between the symbiont and host (mutualism). An isolation attempt was also made based on a dilution to extinction approach and using specific media designed based on the unique metabolic functions identified.

In Chapter 5, the metaproteogenomic approached developed using *C. concentrica* was further applied to the investigation of the changes in the microbial communities of the

Great Barrier Reef sponge *Rhopaloeides odorabile* during elevated seawater temperature. Dramatic changes in taxonomic composition, species richness, genomic content and expressed protein profiles of the sponge microbial communities were detected between the healthy, intermediate and necrotic sponge samples during thermal stress. Putative symbiotic functions like active transport, metabolism of sponge-specific substrates and functions related to maintaining cellular structure were identified from native microbial community from the health sponge. However, changed protein expression profiles were observed from the intermediate and necrotic samples. The results indicated that environmental stress factors such as elevated water temperature could cause a breakdown in symbiosis between sponge and microbes. This breakdown could further lead to the changes in the microbial community and lead to a decline in sponge health. It demonstrated that the sponge necrosis is not caused by pathogen infection, but rather by the disturbance and collapse of the sponge-symbiont interactions.

A general discussion of this project is provided in Chapter 6. The overall conclusion of this project is discussed and potential future work is proposed to further our understanding of microbial communities associated with marine sponges.

Chapter Two
Functional genomic analysis of an uncultured Deltaproteobacterium in the sponge *Cymbastela concentrica*

2.1 Introduction

Sponges are ancient, sessile filter-feeding metazoans, which represent a significant component of marine, benthic communities throughout the world. Despite their simple body structure, they are remarkably efficient in obtaining food by pumping surrounding seawater through a specialised canal system. It has been estimated that as much as 24000 litres of water can be pumped though a kilogram of sponge, on a daily basis and that the filtering system used by the sponge leaves the expelled water essentially sterile (Reiswig 1971, Reiswig 1974, Turon *et al.,* 1997). Despite this process, sponges also permanently host a large number of microorganisms in their body, which can contribute up to 60% of the total biomass (Hentschel *et al.,* 2006, Wilkinson 1978a). These sponge-associated bacteria represent a remarkable diversity across major bacterial and archaeal phyla; some clades found in sponges represent host-specific groups of bacteria (Taylor *et al.,* 2007, Webster and Blackall 2008, Zhu *et al.,* 2008). It has been postulated that the interactions between sponges and microorganisms range from mutualism to commensalism and parasitism, however little is understood about the exact nature of the symbiosis or the functional role played by many of the groups of microbes within the sponge (Webster and Blackall 2008). A major limitation for an understanding of sponge-microbe interactions is the inherent difficulty of culturing the relevant symbiotic microorganisms. Thus for most sponges, the dominant or representative members are not available for further analysis.

A whole-environment shotgun sequencing to characterise the bacterial community of the sponge *Cymbastela concentrica*, an abundant sponge species in shallow, temperate waters along the Australian east coast was recently carried out (Thomas *et al.,* 2010). This metagenomic study revealed previously unrecognised genomic signatures and functions of sponge bacteria including defense mechanisms against the introduction of

foreign DNA, metabolic interactions with the host through vitamin biosynthesis, nutrient transport and utilisation, redox sensing and response, as well as protein-protein interactions mediated through ankyrin and tetratricopeptide repeat proteins. A divergent 16S rRNA gene sequence that belongs to the *Deltaproteobacteria* class was also discovered in the metagenomic dataset. In order to understand the functional characteristics of this novel organism and how these relate to sponge-bacteria interactions, a detailed genomic analysis and determination of cellular localization were carried out. It was found that this organism represents a novel clade and that it lives within the sponge in association with a putative cyanobacterium. The predicted properties of the partial Deltaproteobacterium genome, indicate complex interactions with surrounding cells and their environment, as reflected in the traits for cell attachment, nutrient transport and protein-protein interactions.

2.2 Materials and Methods

2.2.1 Phylogenetic analysis

The initial taxonomic classification of the Deltaproteobacteria 16S rRNA gene sequence was carried out using the Bayesian classifier algorithm with default parameters at the Ribosomal Database Project (RDP) (Cole *et al.,* 2009, Wang *et al.,* 2007). The nearest-neighbour sequences (both for uncultured and isolated representatives) of the Deltaproteobacterial 16S rRNA gene sequence were extracted from RDP version 10, using the seqmatch function and then aligned using the SINA web aligner (Pruesse *et al.,* 2007). A phylogenetic tree was constructed using the maximum likelihood algorithm with near full-length 16S rRNA gene sequences (>1200 nt) in the ARB software package (Ludwig *et al.,* 2004). To avoid the use of highly variable regions for comparisons, a positional variability-by-parsimony mask (pos_var_Bacteria_94) and a user-specific end mask were used, resulting in the comparison of 1178 nucleotides in the tree construction. An uncultured *euryarchaeote*.sp (Genbank Accession: AB077227) was used as an outgroup in the tree construction.

2.2.2 Genomic reconstruction and comparative genomics

For scaffolds with more than 20Kb sequence described in Thomas *et al*. (2010), the tetranucleotide patterns were determined using TETRA (Teeling *et al.,* 2004) and exported as normalised Z-scores. Clustering was performed with Euclidian distance and complete linkage using the software Cluster 3.0 and visualised with JavaTreeView (Eisen *et al.,* 1998). A hand-curated sub-set of scaffolds was defined by one scaffold (ID 1108814257562), which contained the Deltaproteobacterial 16S rRNA gene and several scaffolds with marker genes that could be unambiguously assigned to an appropriate taxonomic level (i.e. *Deltaproteobacteria* or lower). These scaffolds showed no evidence of mis-assemblies and were used to define the border and depth of a cluster for the Deltaproteobacterial species in the hierarchical tree. The clusters were subsequently checked for robustness using k-means clustering (Euclidian distance with Cluster 3.0) with the number of seeds set at 14, corresponding to the number of major clusters identified in the sponge's bacterial community, (Thomas *et al.,* 2010). The hand-curated scaffolds were again confined to a distinct k-means cluster, implying that alternative clustering to the one used in the hierarchical tree was unlikely. This cluster of hand-curated sequence was subsequently used to define the Deltaproteobacterium's genome bin by including all scaffolds contained in the corresponding sub-tree (Figure 2.1). This bin was subsequently expanded by also including scaffolds with more than 5Kb sequence and further validated as follows. First, each scaffold was compared to all proteins in the non-redundant sequence database (NR) from the National Center for Biotechnology Information (NCBI) and best matching positions were assigned to the taxon associated with the protein. The most abundant taxon at any level (species to kingdom), was then used as a taxonomic assignment of the scaffolds at a particular level. This homology-based taxon assignment in no case showed a conflict with the composition-based taxonomic assignment. Second, the Deltaproteobacterium's genome bin had only one representative from a set of 31 conserved, single-copy marker proteins (Ciccarelli *et al.,* 2006), indicating that no "hybrid" or "chimeric" genome bin was created. Scaffolds of the Deltaproteobacterium's genome bin were then processed with Metagene (Noguchi *et al.,* 2006), to identify open reading frames (ORFs). For functional annotation, each ORF was searched against the Clusters of Orthologous Group (COG) (Tatusov *et al.,* 2003) and TIGRFAM (Haft *et al.,* 2003) databases using the hmmer version 2.3.2 (Eddy 1998) and the curated section of SwissProt (Boeckmann

et al., 2003) using blastP (Altschul *et al.,* 1990). Analysis of functional production was supported by extensive manual curation. Transport proteins were further characterised using the Transporter Classification Database (TCDB; http://www.tcdb.org).

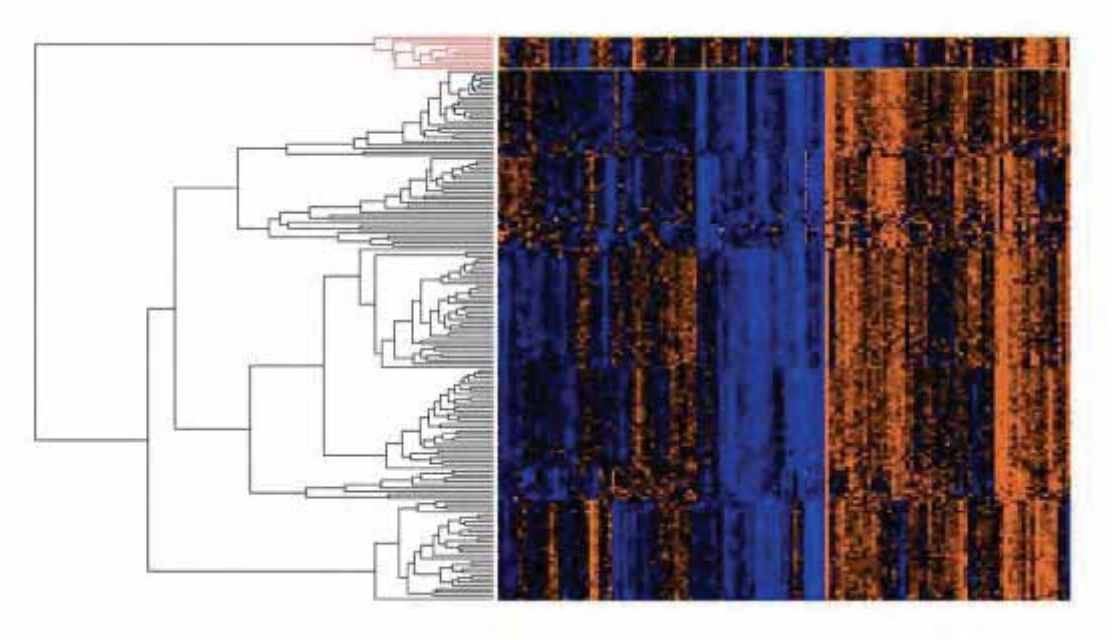


Figure 2.1 Tetra-nucleotide-based clustering of scaffolds with more than 20Kb sequence for the sponge-bacteria derived metagenome.
The left tree shows the clustering of the scaffolds. The heat map shows over- and under-represented tetra-nucleotides in red and blue, respectively. Scaffold clusters that could be unambiguously linked to the sponge Deltaproteobacterium is boxed in yellow and highlighted in red in the tree.

For comparative analysis, the complete and annotated genome sequences of the species *Bdellovibrio bacteriovorus* strain HD100 (3587 ORFs) (Rendulic *et al.,* 2004) and *Desulfurmonas acetoxidans* DSM 684 (3234 ORFs) were downloaded from the NCBI database. The complete, predicted protein sets of these two organisms and the partial protein dataset for the sponge Deltaproteobacterium were searched with blastP (E-value $<10e^{-5}$) against the COG database. Proteins with at least 30% identity to an entry were then classified into specific COGs and COG categories. Statistical, pairwise, comparison of the COG profile with respect to the sponge Deltaproteobacterium, *B. bacteriovorus* and *D. acetoxidans* was performed by resampling (n=1000) of COG subsamples (n=300) as outlined in Lauro *et al.* (2009). MEGAN (Huson *et al.,* 2007) was used to examine the taxonomic relationship of proteins within the Deltaproteobacterial partial genome. MEGAN employs a Lowest Common Ancestor (LCA) algorithm to assign ORFs into taxonomical groups based on BLAST bit-scores.

Parameters were set with "min support" at 1, "min score" at 0 and "top percentage" set at 10%.

2.2.3 Sponge sampling and preparation for FISH analysis

Specimens of the marine sponge *C. concentrica* were collected from Botany Bay near Bare Island, Sydney, Australia (S 33.59.461; E 151.13.946) during August 2009. Following collection, *C. concentrica* tissue was rinsed with calcium and magnesium free seawater (CMFSW; 25 g NaCl, 0.8 g KCl, 1 g Na_2SO_4, 0.04 g $NaHCO_3$ per 1 L) and immediately transferred to 15% sucrose. The specimens were then transferred to the laboratory on ice (15 minutes travel time) and incubated for 3 hours at 4℃. Specimens were transferred into a 15% sucrose/ Optical Coherence Tomography medium (OCT medium) solution at the ratio of 35:10 for fixation overnight (4℃) and exposed to a series of sucrose/OCT medium solutions (15:10 for 2 hrs, 15:30 for 2 hrs and 100% OCT for 2hrs; all at 4℃). Fixed specimens were then snap frozen in liquid nitrogen and stored at -80℃ until use. Specimens were sectioned with a Shandon Cryotome SME (Thermo Electron Corporation, Pittsburg, PA) to 10µm thickness, placed on a microscope slide and air-dried.

2.2.4 Fluoresecnce in situ hybridisation (FISH)

The sections were dehydrated in an aqueous ethanol series (50%, 80% and 96%) for 3 minutes each and air-dried again, prior to probe hybridisation. Hybridisation was performed in a hybrid chamber at 46℃ in hybridisation buffer [0.9M NaCl, 20mM Tris-HCl (pH 7.4), 0.01% sodium dodecyl sulfate (SDS) and variable amounts of formamide (25% and 30%, Table 2.1)] for 2 hours. Fluorescent probes were used at a concentration of 5ng labelled probe μl^{-1} hybridisation buffer [0.9M NaCl, 20mM Tris-HCl (pH 7.4), 0.01% SDS and variable amounts of formamide]. Oligonucleotide probes (Thermo Fisher, Germany) used (Table 2.1) were labelled with either fluorescein-isothiocyanate (FITC) or indocarbocyanine (CY3). Slides were incubated in 50 to 100µl of hybridisation buffer dependent on the size of the sections, then carefully rinsed and further incubated at 43℃ in wash buffer [0.2M Tris-HCl (pH7.4), 5mM EDTA, 0.01% SDS and variable amount of NaCl (Table 2.1)] for 25 minutes. The wash buffer was

then carefully rinsed off and the slides air-dried in the dark. The anti-fading agent Citifluor (Citifluor Ltd., London) was mounted on the slides to prevent fluorochrome bleaching.

Table 2.1 Oligonucleotide probes used for *in situ* hybridisation.
(FITC: fluoresceine-isothiocynanate; CY3: indocarbocyanine; Tm: melting temperature) Cc: Marine sponge *C. concentrica.*

Probe	Sequence 5'-3'	T_m (°C)	Formamide (%)	Specificity	Labeling
EUB338[1]	GCTGCCTCCCGTAGGAGT	51.5	25	Eubacteria	FITC, CY3
Bdello_Cy3	ATTCGCCTCCCGAAGGTT	53.4	30	Deltaproteobacteria (*C. concentrica)*	CY3

[1]: EUB338, Amann *et al.* (1990)

2.2.5 Microscopy and image analysis

Microscopic analysis was carried out with an Olympus FV1000 Laser Scanning Microscope (inverted), with excitation wavelengths of 488nm and 543nm for FITC and Cy3 labelled probes, respectively. Colour microscopy images were acquired using the camera attached to the microscope (Olympus, Japan) and processed with Adobe Photoshop CS. The presence and distribution of bacteria in sponge tissues was determined by investigating multiple regions from at least three replicate specimens. The relative proportion of the Deltaproteobacterium with respect to the total microbiota was estimated by counting fluorescence specific pixels in digital FISH images using the ImageJ software. This software classifies and counts particles on the basis of their relative density (fluorescence) compared to the background, via a threshold process. Proportions were defined as the ratio between the area of the Cy3-fluorescence (i.e. Deltaproteobacterium) and FITC fluorescence (e.g. all bacteria). Counting of individual cells was not possible, as the bacteria were mostly in aggregates. The analysis was performed on five images from different locations of the sponges and averages and standard deviations of proportions were calculated.

2.2.6 Probe design and validation

A specific oligonucleotide probe for the sponge Deltaproteobacterium was designed using the Probe Design program within the ARB software package, according to Hugenholtz *et al.* (2002) and target positions 1434 to 1452 (*E. coli* numbering). The

specificity of the probe was checked against the 1951 16S rRNA gene sequences of the bacterial community from *C. concentrica* described by Thomas *et al*. (2010) and the Greengenes database. It was found that two or more mismatches could give signals against non-target organisms. Therefore a two-probe hybridisation was performed using the Cy3 labelled Deltaproteobacterium specific probe and a FITC labeled bacteria-specific probes to pure cultures of *Colwellia hornerae* str. ACAM 607^{T}(two mismatches, kindly supplied by John Bowman, Tasmanian Institute of Agricultural Science, Australia) and *Alcanivorax dieselolei* st. B5 (three mismatches, kindly supplied by Zongze Shao, Key Lab of Marine Biogenetic Resources, Third Institute of Oceanography, State Oceanic Administration, China). Pure cultures of non-target organisms were fixed with 4% paraformaldehyde and *in situ* hybridisation was performed as described above. Strong signals for bacteria-specific probes and no hybridisation and for the Deltaproteobacterium specific probe were observed (Figure 2.2). The results confirmed the specificity of the probe to less than two mismatches.

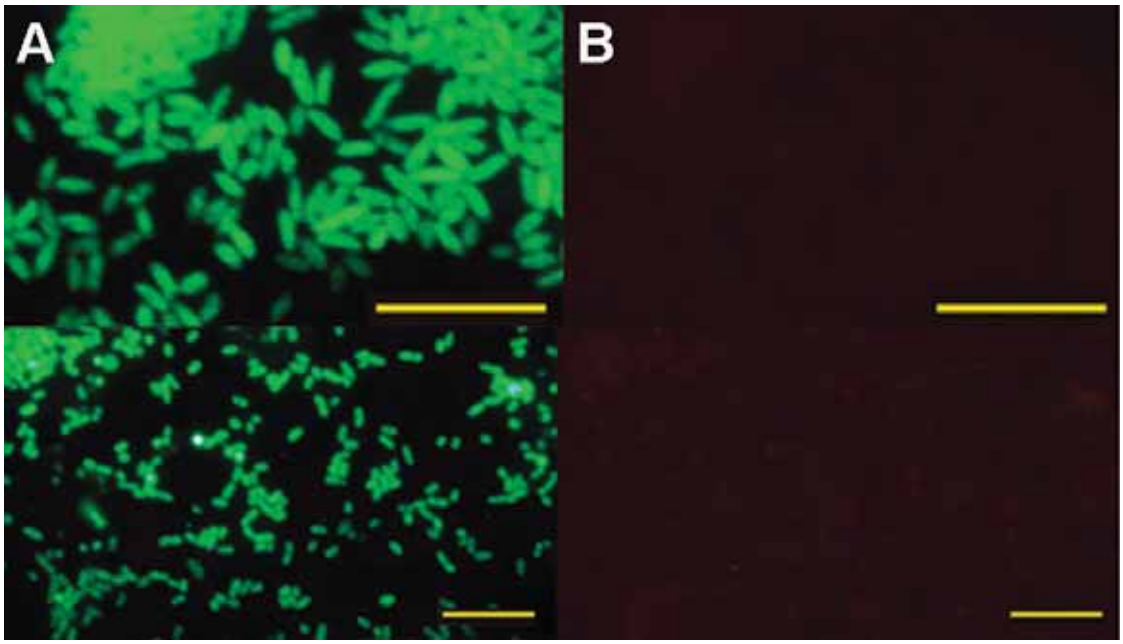


Figure 2.2 Evaluation of probe specificity by fluorescence *in situ* hybridisation (FISH).
Pure cultures of *Colwellia hornerae* str. ACAM 607T (upper panel) and *Alcanivorax dieselolei* st. B5 (lower panel) were hybridised with a FITC-labelled (green) bacteria-specific probe, EUB338 (A), and a Cy3-labelled Deltaproteobacterium-specific probe (B), respectively. Yellow scale bar: 10 μm.

2.3 Results and Discussions

2.3.1 A novel Deltaproteobacterial clade of sponge-associated bacteria

In a phylogenetic survey of the bacterial community associated with the sponge *C. concentrica*, a 16S rRNA sequence (Thomas *et al.*, 2010) that was assigned by the RDP classifier to the class *Deltaproteobacteria* (87% confidence) and the order of *Bdellovibrionales* (81% confidence) was discovered. Phylogenetic analysis and comparison of the sponge Deltaproteobacterial sequence with sequences from closely-related isolates and uncultured organisms (Figure 2.3) indicated that it fell into a cluster of uncultured bacteria that were all found in marine sponges. Specifically, the sponge Deltaproteobacterium sequence was identical to a previously described sequence of an uncultured bacterium (Cc110, AY942772) from *C. concentrica*, indicating a permanent association of this bacterium with the sponge. Other sequences in this cluster were from the reef sponge *Axinella corrugata* (clones B2505_B3 (EF092213) and MBrad_A7 (EF092260)) and the Caribbean sponge *Svenzea zeai* (clone A115 (FJ529308)) (Lee *et al.*, 2009). These five sponge-associated organisms were only distantly related (>15% sequence dissimilarity), to the nearest cluster of cultured organisms that was defined by *Bdellovibrio bacteriovorus* strains. The latter strains were isolated from soil (100T, AF084850), animal gut (HD100, AJ292759 and HD127, AJ292760) and an aqua culture farm (JSF1, EU884925) and had high levels of 16S rRNA gene sequence identity (>99%). In addition, three other *B. bacteriovorus* isolates from soil (MPA, AY294215), root extracts (BEP2, AF148938) and sewage (JSS, EF687743) were found in the same cluster, but with lower levels of similarity.

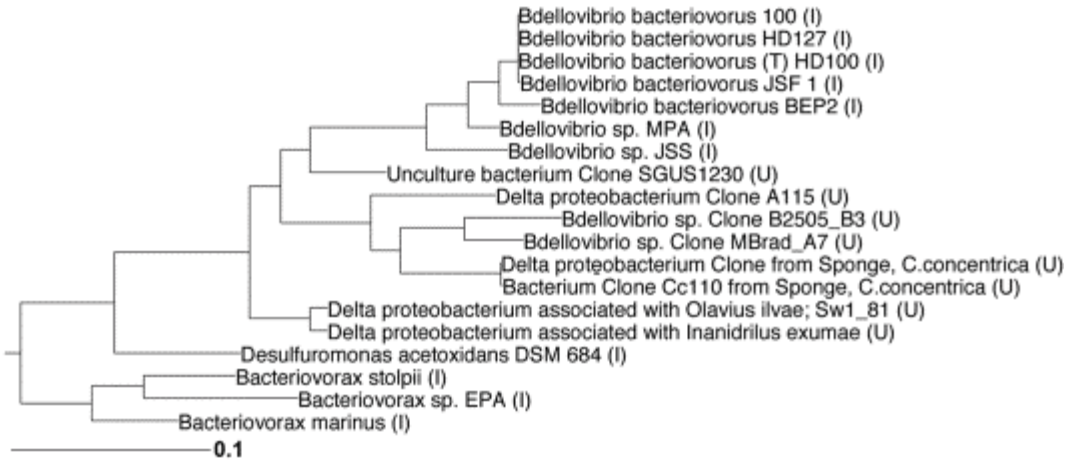


Figure 2.3 Maximum likelihood tree of the Deltaproteobacterium's 16S rRNA gene sequence from *C. concentrica* and related organisms.
The tree was built on 1178 nucleotides of the 16S rRNA gene. An uncultured *euryarchaeote*.sp (AB077227) was used as an outgroup for the analysis (not shown in the tree). The scale bar indicates 0.1 nucleotide changes (10%) per nucleotide position. (U) = uncultured; (I) = isolates.

The sponge-associated Deltaproteobacteria cluster was also distinct from the genera *Desulfuromonas* and *Bacteriovorax*, with dissimilarity levels of 24% and 30%, respectively. The genera *Desulfuromonas* contains aquatic species that are known to grow anaerobically by oxidizing acetate with concomitant reduction of elemental sulfur or Fe (III) (Pfennig and Biebl 1976). *Bacteriovorax* species parasitises other gram-negative bacteria and have a highly motile, free-living, extracellular phase during which they seek out new hosts. The phenotypic characteristics of this genus are identical to that of *Bdellovibrio,* and *Bacteriovorax* was only recently reclassified as a distinct genus based on 16S rRNA gene-based phylogeny (Baer *et al.*, 2000, Baer *et al.*, 2004).

The phylogenetic analysis conducted here showed that the Deltaproteobacterium from *C. concentrica* possibly represents a novel genus or family and, that members of this group of yet-to-be-cultured bacteria are associated with sponges.

2.3.2 Functional genomic comparison of the sponge Deltaproteobacterium with related bacteria

To gain insight into the functional genomic properties of the new Deltaproteobacterium, a partial genome for the organism from the data of the recent metagenomic study (Thomas *et al.,* 2010) was reconstructed. The assembled data represents a composite genome of many related individuals, as seen in other partial genome reconstructions

from metagenomic shotgun data (e.g. (Hallam *et al.,* 2006a, Tyson *et al.,* 2004). Within population variation could exist at the single-nucleotide or gene level. Through binning based on nucleotide signature it was possible to confidently identify 32 genomic scaffolds belonging to the organism, spanning an estimated total of 1.1 Mbp with 582,622 bp of unique sequence. The longest scaffold was 116,573 bp and the average span of scaffold was 35,695 bp. Six hundred and seventy open reading frames in this reconstruction were identified and their functional annotation compared to the completed genomes of *B. bacterivorus* HD100 and *D. acetoxidans* DSM68, the nearest related organisms with available genome sequences.

The sponge Deltaproteobacterium showed an over-representation of the functional COG category moieties of translation and replication, recombination and repair, and an under-representation to signal transduction, compared to *B. bacteriovorus* (Figure 2.4 A). At the individual COG level, specific functions associated with signal transduction mechanisms (COG0642, COG0840, COG0784, COG0644 and COG0745) were underrepresented, while functions associated with ankyrin repeat proteins (COG0666), TPR repeat proteins (COG0790), the 140kD beta subunit of DNA-directed RNA polymerase (COG0085), the 2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component (COG0567), restriction endonuclease related to defense mechanism (COG1403) and D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4, COG2027), were overrepresented (Table 2.2). A similar pattern of functional characteristics was observed when the sponge Deltaproteobacterium was compared to *D. acetoxidans* DSM684 (Figure 2.4 B). The overall functional comparison confirmed that signal transduction was under-represented in the sponge Deltaproteobacterium and translation and replication, recombination and repair were overrepresented. At the individual COG level, specific functions associated with signal transduction mechanisms (COG2199, COG0642, COG2200, COG0840, COG3706, COG0784, COG 2204 and COG 2203) were underrepresented, while functions associated with ankyrin repeat proteins (COG0666) and TPR repeat proteins (COG0790) are overrepresented (Table 2.3). The analysis highlighted the clear functional differences between the sponge Deltaproteobacterium and *B. bacterivorux* and *D. acetoxidans*, respectively, particularly within the function of signal transduction.

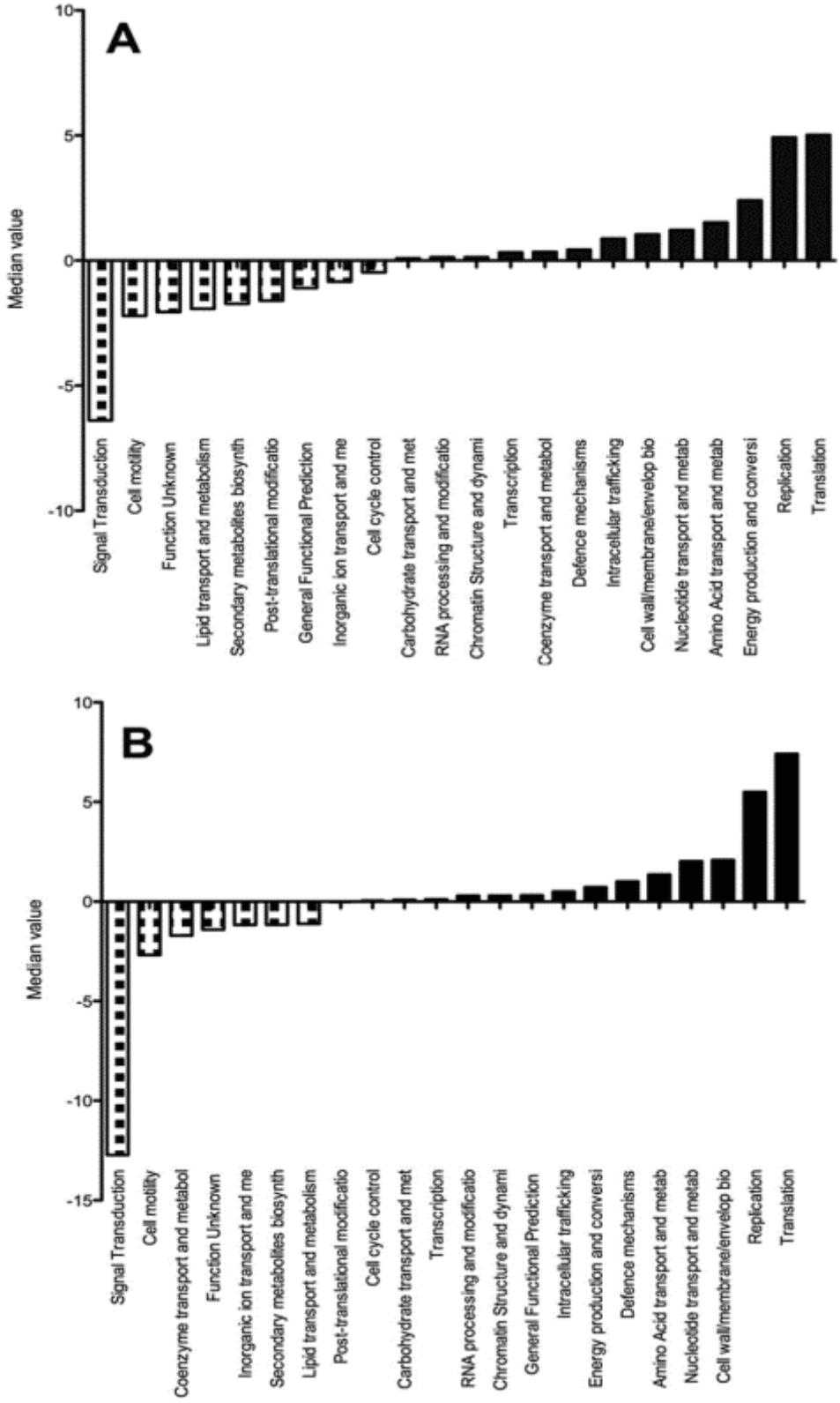


Figure 2.4 Functional genome comparison based on COG categories of Deltaproteobacterium partial genome and the genomes of *B. bacteriovorus* HD100 genome (A) and *D. acetoxidans* DSM684 (B).
The Y-axis indicates the median value with a significant cut-off value of -2 and 2.

Table 2.2 Individual COGs overpresented (⇑) or underpresented (⇓) in Deltaproteobacterium partial genome when compared to *B. bacterivorus* HD100 genome.

COG	Description	Functional category
COG0642	Signal transduction histidine kinase	Signal transduction mechanisms (⇑)
COG0840	Methyl-accepting chemotaxis protein	Cell motility and secretion / Signal transduction mechanisms (⇑)
COG0784	FOG: CheY-like receiver	Signal transduction mechanisms (⇑)
COG0664	cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases	Signal transduction mechanisms (⇑)
COG0745	Response regulators consisting of a CheY-like receiver domain and a winged-helix DNA-binding domain	Signal transduction mechanisms / Transcription (⇑)
COG1028	Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)	Secondary metabolites biosynthesis, transport, and catabolism / General function prediction only (⇑)
COG0666	Ankyrin repeat	General function prediction only (⇓)
COG0790	FOG: TPR repeat, SEL1 subfamily	General function prediction only (⇓)
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	Transcription (⇓)
COG0567	2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes	Energy production and conversion (⇓)
COG1403	Restriction endonuclease	Defense mechanisms (⇓)
COG2027	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)	Cell envelope biogenesis, outer membrane (⇓)

Table 2.3 Individual COGs overpresentend (⇑) or underpresented (⇓) in Deltaproteobacterium partial genome when compared to *D.acetoxidans* DSM684 genome.

COG	Description	Functional category
COG2199	FOG: GGDEF domain	Signal transduction mechanisms (⇑)
COG0642	Signal transduction histidine kinase	Signal transduction mechanisms (⇑)
COG2200	c-di-GMP phosphodiesterase class I (EAL domain)	Signal transduction mechanisms (⇑)
COG0840	Methyl-accepting chemotaxis protein	Cell motility and secretion / Signal transduction mechanisms (⇑)
COG3706	Response regulator containing a CheY-like receiver domain and a GGDEF domain	Signal transduction mechanisms (⇑)
COG0784	FOG: CheY-like receiver	Signal transduction mechanisms (⇑)
COG2204	Response regulator containing CheY-like receiver, AAA-type ATPase, and DNA-binding domains	Signal transduction mechanisms (⇑)
COG2203	FOG: GAF domain	Signal transduction mechanisms (⇑)
COG2801	Transposase and inactivated derivatives	Replication, recombination and repair (⇑)
COG0666	FOG: Ankyrin repeat	General function (⇓)
COG0790	FOG: TPR repeat, SEL1 subfamily	General function (⇓)
COG0085	DNA-directed RNA polymerase, beta subunit/140 kD subunit	Transcription (⇓)
COG0174	Glutamine synthetase	Amino acid transport and metabolism (⇓)
COG0272	NAD-dependent DNA ligase (contains BRCT domain type II)	Replication, recombination and repair (⇓)
COG0567	2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, and related enzymes	Energy production and conversion (⇓)
COG1173	ABC-type dipeptide/oligopeptide/nickel transport systems, permease components	Amino acid transport and metabolism/Inorganic ion transport and metabolism (⇓)
COG1208	Nucleoside-diphosphate-sugar pyrophosphorylase involved in lipopolysaccharide biosynthesis/translation initiation factor 2B, gamma/epsilon subunits (eIF-2Bgamma/eIF-2Bepsilon)	Cell envelope biogenesis, outer membrane / Translation, ribosomal structure and biogenesis (⇓)
COG1403	Restriction endonuclease	Defense mechanisms (⇓)
COG1680	Beta-lactamase class C and other penicillin binding proteins	Defense mechanisms (⇓)
COG2027	D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein 4)	Cell envelope biogenesis, outer membrane (⇓)
COG2805	Tfp pilus assembly protein, pilus retraction ATPase PilT	Cell motility and secretion / Intracellular trafficking and secretion (⇓)

In order to explore the possibilty that certain functions of the Deltaproteobacterium genome have a shared evolutionary history with related genomes, all its proteins were classified taxonomically with a Lowest Common Ancestor (LCA) algorithm implemented in MEGAN (Huson *et al.*, 2007). Taxonomic analysis of 663 proteins

demonstrated that 249 could be assigned to *B. bacteriovorus* HD100, 4 assigned to *D. acetoxidans* DSM684, 185 had no hits in the NR database and the remaining proteins were assigned, at various levels, within the bacteria or to cellular organisms (Table 2.4). The results showed that the Deltaproteobacterium genome shared more conserved proteins with *B. bacteriovorus* HD100 genome, than any other available genome.

Table 2.4 Taxonomical assignment of proteins from the sponge Deltaproteobacterium

Total proteins	663
Assigned proteins	477
Unassigned proteins	1
Proteins with no hits	185
Cellular organisms	475
Bacteria	456
Desulfuromonas acetoxidans DSM 684	4
Bdellovibrio bacteriovorus HD100	249
Eukaryota	9
Archaea	3

Specific functions enriched within the 249 proteins included translation, transcription, replication and cell wall biogenesis, which reflected the evolutionary conserved nature of these putative functions across phylogenetically related organisms. The sponge Deltaproteobacterium and *B. bacteriovorus* also share an “adventurous” gliding motility protein R, which is associated with microorganisms that travel in environments with a low water availability, such as biofilms, microbial mats, and soil (Spormann 1999), and which facilitate host cell infection in some Apicomplexan parasites (Kappe *et al.,* 1999). In addition, there is a shared and conserved Flp pilus assembly protein and a cluster of four proteins that are involved in Type IV pili assembly and translocation (assembly protein TapB, two twitching motility proteins, and a pilin assembly protein PilC). Type IV pili are associated with twitching motility, which has been proposed as important in cell attachment and invasion by Bdellovibrio species, where the force of pili retraction generated by depolymerization could pull the Bdellovibrio cell through its attached prey cell (Lambert *et al.,* 2009). Overall, these observations support the concept that certain aspects of attachment and motility might be shared and conserved between the sponge Deltaproteobacterium and *B. bacteriovorus*.

A number of functions in cell wall and envelope biogenesis were conserved across the sponge Deltaproteobacterium and *B. bacteriovorus*. They included a murein-degrading enzyme (membrane-bound lytic murein transglycosylase D precursor) that plays a role

in recycling of muropeptides during cell elongation and/or cell division. Bdellovibrio species can digest up to 20% of their prey's cell wall after invasion, while keeping their newly formed bdelloplast structure from uncontrolled lysis (Thomashow and Rittenberg 1978). During bacterial predation the amount of digested peptidoglycan would yield enough precursors to form the multiple progeny. Murein biosynthesis requires a crucial disaccharide penta-peptide precursor molecule (UDPMurNAc-pentapeptide) usually formed in the cytoplasm and Bdellovibrio species have therefore retained the six cytoplasmic enzymes (MurA-F), responsible for its *de novo* biosynthesis and periplasmic transport. The sponge Deltaproteobacterium also contains three conserved homologs of these enzymes, namely the UDP-N-acetylmuramoylalanine-D-glutamate ligase (MurD), the UDP-N-acetylmuramoylalanyl-D-glutamyl-2, 6-diaminopimelate-D-alanyl-D-alanine ligase (MurF), and the UDP-N-acetylmuramoylalanyl-D-glutamate-2, 6-diaminopimelate ligase (MurE). Clustered with the ORFs of these proteins is another gene (*mraY*) encoding a phospho-N-acetylmuramoyl-pentapeptide-transferase, which catalyses the first step of lipid cycle reactions and the biosynthesis of the cell wall peptidoglycan. Immediately downstream of this cluster are two genes that are also associated with cell wall biogenesis. Firstly, there is an ORF encoding for the penicillin-binding protein 2B (PBP2), which has transglycosylase activities to catalyze the formation of the beta 1-4 glycosidic links from the lipid-bound precursors to the nascent end of the peptidoglycan strand and which is responsible for bulk peptidoglycan synthesis in other bacteria (Vollmer and Bertsche 2008). The second gene encodes for a peptidoglycan synthetase (FtsI or PBP3), whose homolog in *E. coli* is essential for the septum formation of the murein sacculus and synthesis of cross-linked peptidoglycan from the lipid intermediates (Nakamura *et al.,* 1983). It provides evidence that, besides the putative roles in cell attachment and motility, the putative functional process of cell wall biogenesis may also be an evolutionary conserved function of the sponge Deltaproteobacterium and *B. bacteriovorus*.

2.3.3 The sponge Deltaproteobacterium exhibits a cell-associated life-style

The similarity found in motility, attachment and cell wall biogenesis when compared with *B. bacteriovorus* led to the hypothesis that the sponge Deltaproteobacterium might also form an association (potentially of predatory nature) with other bacterial cells.

Therefore the physical localisation of the Deltaproteobacterium in the sponge by fluorescence *in situ* hybridisation (FISH) was investigated. Using a probe for the domain of bacteria (EUB338), the presence of high numbers of bacteria within the sponge *C. concentrica* was demonstrated (Figure 2.5 A). The probe directed against the Deltaproteobacterium 16S rRNA sequence resulted in the detection of small cells (<1μm in length) in the sponge tissue. These were generally associated with a larger, auto-fluorescent cell type (>3um in length) and only occasionally found as free-living cells (Figure 2.5 B&C). The relative proportion of the Deltaproteobacterium against the total microbiota was estimated to be 6%. The localisation of the Deltaproteobacterium was often at the end of the larger, auto-fluorescent cells.

The auto-fluorescent "host" cells showed green auto-fluorescence when excited with 488nm light, consistent with the fluorescence reported for *cyanobacteria*, diatoms and dinoflagellates (Tang and Dobbs 2007). To determine if the auto-fluorescent cells were eukaryotic (ie. diatoms or dinoflagellets) or bacterial (ie. *cyanobacteria*), a further FISH analysis using a two-probe hybridisation design with a Cy3- labelled probe specific for the Deltaproteobacterium and a FITC-labelled bacteria-specific probe was conducted. The auto-fluorescent cells were found to respond to the general bacterial probe (Figure 2.5 D). Therefore, based on the particular auto-fluorescence and the binding of a bacteria-specific probe, these cells were most likely to be *cyanobacteria*. This finding is consistent with an observation made thirty years ago by Wilkinson (Wilkinson 1979a), who found "Bdellovibrio-like" particles, using electronmicroscopy, within unicellular cyanobacteria in the mesohyl of the two coral reef sponges, *Neofibularia irata* and *Jaspis stelifer.*

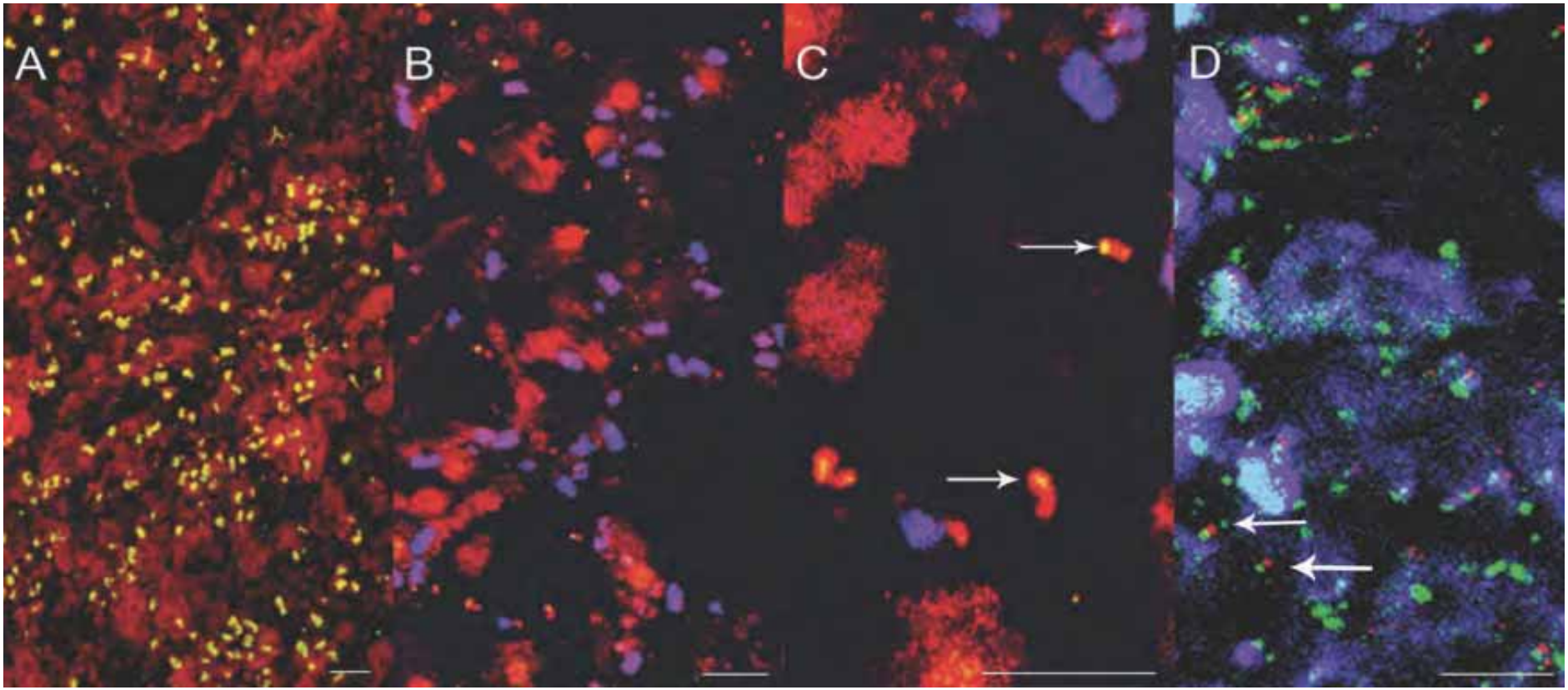


Figure 2.5 Detection of sponge-associated bacteria by fluorescence *in situ* hybridisation (FISH).
A. Section of *C. concentrica* hybridised with a Cy3-labelled Bacteria-specific (EUB338_CY3) probe (yellow) with the auto-fluorescent sponge tissue (red). B. Section of *C. concentrica* hybridised with a Cy3-labelled Deltaproteobacterium-specific probe (yellow), auto-fluorescence background (red) and larger auto-fluorescence eukaryotic cells (blue-purple). Arrows indicate the host-associated nature of the sponge Deltaproteobacterium within auto-fluorescent cells (red). C. Section of *C. concentrica* (enlarged) hybridised with a CY3-labelled Deltaproteobacterium-specific probe (yellow). Arrows indicate the host-associated nature of the sponge Deltaproteobacterium within auto-fluorescent cells (red). D. Section of *C. concentrica* (enlarged) hybridised with a Cy3-labelled Deltaproteobacterium-specific probe (red) and a FITC-labelled Bacteria-specific (EUB338_FITC) probe (green). The auto-fluorescent background of sponge tissues (blue) are also shown. Arrows indicate the associated nature of the sponge Deltaproteobacterium within a putative cyanobacterial cell (green).

2.3.4 Predicted functional properties of the sponge Deltaproteobacterium

To gain further insight into the functional properties of the new Deltaproteobacterium, a careful manual analysis of its partial genome with respect to physiological and ecological functions was performed.

The subunits of the terminal enzyme in the aerobic respiratory chain, the cytochrome C oxidase, in the partial genome (1108814258778_ORF0001) were identified. It is therefore likely that the sponge Deltaproteobacterium is able to grow aerobically, further distinguishing it from the related, strictly anaerobic *Desulfuromonas* species.

A total of 19 proteins had clear functions in membrane transport and could be assigned to four major transporter families (Table 2.5) with the ABC (ATP-binding cassette) superfamily being the largest group. A phosphate transporter (3.A.1.7.2) is encoded by a *pstSCAB* operon and appears to be homologous to the high-affinity phopshate-specific ABC transport system from *Mycobacterium smegmatis* (Gebhard and Cook 2008). A glutathione porter (3.A.1.5.11) is partially encoded by genes similar to *yliA, C* and *D*, which encode ATP-binding cassettes in *E. coli* K-12, for the import of extracellular glutathione and which have been shown to be crucial for growth on glutathione as the sole sulfur source (Suzuki *et al.,* 2005). The Deltaproteobacterium genome also encodes for a tetracycline resistance protein (1108814258826_ORF0009), which functions as a antibiotic efflux transport (metal-tetracycline/H+ antiporter) of the major facilitator superfamily (MFS) (Marger and Saier 1993, Yamaguchi *et al.,* 1990) and which can perform a wide variety of processes, including uptake of essential ions, nutrients and removal of toxic compounds (Griffith *et al.,* 1992). An additional tetracycline efflux/ multidrug resistance protein (1108814257392_ORF0004) indicated that removal of toxins and antibiotics might be a relevant functional aspect of the Deltaproteobacterium.

The Deltaproteobacterium also appeared to maintain features of an essential anabolic capacity, despite its host-associated life-style, and has probably not developed an obligate dependence on nutrients provided by the host cell: It possesses two key enzymes for the chorismate biosynthesis (shikimate synthase: 1108814258173_ORF0043 and 3-phosphoshikimate 1-carboxyvinyltransferase:

1108814258173_ORF0042), which provide intermediates for the synthesis of aromatic amino acids and coenzymes like folic acid, ubiquinone, menaquinone, and enterochelin. Also, proteins involved in menaquinone-8 biosynthesis (menaquinone biosynthesis proteins MenD: 1108814258173_ORF0048; O-succinylbenzoate-CoA ligase MenE: 1108814258778_ORF0042; naphthoate synthase MenB*:* 1108814258807_ORF0012 and a 1,4-dihydroxy-2-naphthoate octaprenyltransferase: 1108814258807_ORF0011) were also identified, indicating the ability of the sponge Deltaproteobacterium to produce this important redox component of the electron transfer chain (Farrand and Taber 1974). Another cofactor which is possibly synthesised is riboflavin (vitamin B2), a precursor for the essential redox-active cofactors FMN and FAD. Three proteins related to riboflavin biosynthesis were found, a 6,7-dimethyl-8-ribityllumazine synthase (1108814258173_ORF0015), a riboflavin biosynthesis protein RibAB (1108814258173_ORF0017) and riboflavin biosynthesis protein RibF (1108814258794_ORF0021).

Table 2.5 Transporter Classification (TC) and functional predictions of putative transport proteins for the sponge *Deltaproteobacterium*.

Transporter classification	Family name	Sequence information	Best BLAST hit in TCDB and comment
2.A.6	**RND superfamily**		
2.A.6.2.22	The conjugated and unconjugated bile (bile-inducible)/multidrug (ethidium, ciprofloxacin, norfloxacin, tetracycline, cefotaxime, rifampicin, erythromycin, chloramphenicol, salicylate; drug-noninducible) efflux pump	SeqName1108814258123ORF0021	CmeB [*Campylobacter jejuni*]
2.A.6.2.22	The MuxABC-OpmB multidrug (aztreonam, macrolides, novobiocin and tetracycline) resistance efflux pump complex (with two RND-type proteins (MuxB and MuxC)), both required for activity	SeqName1108814258123ORF0022	Probable Resistance Nodulation-Cell division (RND) efflux transporter [*Pseudomonas aeruginosa*]
2.A.6.4.1	The secretory accessory proteins, SecDF	SeqName1108814258206ORF0001	SecD [*Escherichia coli*]
2.A.6.4.1	The secretory accessory proteins, SecDF	SeqName1108814258206ORF0002	SecF [*Escherichia coli*]
2.A.33	**The NhaA Na+:H+Antiporter (NhaA) Family**		
2.A.33.1.2	NhaA Na+,K+:H+ antiporter	SeqName1108814258778ORF0014	NhaA [*Vibrio parahaemolyticus*]
3.A.1	**ABC superfamily**		
3.A.1.27.2	The chloroplast lipid (Trigalactosyl acyl glycerol (TDG)) transporter	SeqName1108814257418ORF0017	Tgd1 [*Arabidopsis thaliana* (Mouse-ear cress)]
3.A.1.27.1	The γ -hexachlorocyclohexane (γHCH) uptake permease, LinKLMN (most similar to 3.A.1.12.4, the QAT family)	SeqName1108814257418ORF0018	LinL [*Sphingobium japonicum*]

3.A.1.7.2	Phosphate transporter, PstSCAB	SeqName1108814257581ORF0012	PstB [*Mycobacterium smegmatis*]
3.A.1.7.2	Phosphate transporter, PstSCAB	SeqName1108814257581ORF0013	PstA [*Mycobacterium smegmatis*]
3.A.1.7.2	Phosphate transporter, PstSCAB	SeqName1108814257581ORF0014	PstC [*Mycobacterium smegmatis*]
3.A.1.7.2	Phosphate transporter, PstSCAB	SeqName1108814257581ORF0015	PstS [*Mycobacterium smegmatis*]
3.A.1.5.5	Probable cationic peptide porter (may also take up peptide antibiotics and protamine; implicated in K+ homeostasis) [SapD can stimulate the K+ uptake activities of TrkH and TrkG (TC #2.A.38.1.1) in the presence of ATP]	SeqName1108814258173ORF0037	SapA [*Salmonella typhimurium*]
3.A.1.5.11	Glutathione porter	SeqName1108814258173ORF0038	YliC [*Escherichia coli*]
3.A.1.5.11	Glutathione porter	SeqName1108814258173ORF0039	YliD [*Escherichia coli*]
3.A.1.5.5	Probable cationic peptide porter (may also take up peptide antibiotics and protamine; implicated in K+ homeostasis) [SapD can stimulate the K+ uptake activities of TrkH and TrkG (TC #2.A.38.1.1) in the presence of ATP]	SeqName1108814258613ORF0002	SapA [*Salmonella typhimurium*]
3.A.1.6.1	Sulfate/thiosulfate porter	SeqName1108814258807ORF0008	CysA [*Escherichia coli*]

3.A.1.5.11	Glutathione porter	SeqName1108814258826ORF0004	YliC [*Escherichia coli*]
3.A.1.5.11	Glutathione porter	SeqName1108814258826ORF0005	YliD [*Escherichia coli*]
3.A.1.5.11	Glutathione porter	SeqName1108814258826ORF0047	YliA [*Escherichia coli*]
9.A.40	**The HlyC/CorC (HCC) Family of Putative Transporters**		
9.A.40.3.1	Broad specificity heavy metal transporter, ACDP2 (ancient conserved domain-containing protein-2), CNNM2 or Cyclin2	SeqName1108814258457ORF0005	YrkA [*Bacillus subtilis*]

A total of 6 ORFs, five of which were clustered, encoding the multi-functional ankyrin repeat (AR) protein (COG0666) were found in the partial Deltaproteobacterium (Figure 2.6). The complete genome of *B. bacteriovorus* HD100 only contains two AR-containing proteins. These proteins act as mediators for a diverse range of protein-protein interactions and are found commonly in eukaryotic and viral proteins, but until now rarely in bacteria (Bork 1993, Lux *et al.,* 1990, Mosavi *et al.,* 2004). Eukaryotic-like AR have been reported in obligate intracellular bacteria such as the *Alphaproteobacteria, Wolbachia pipientis* (Iturbe-Ormaetxe *et al.,* 2005) and *Ehrlichia canis* (Mavromatis *et al.,* 2006), as well as in the *Gammaproteobacteria, Legionella pneumophila* (Habyarimana *et al.,* 2008) and *Coxiella butnetii* (Voth *et al.,* 2009). The involvement of AR proteins in host-bacterial associations and survival of the bacteria inside other cells, have been highlighted recently by a mutant study of *L. pneumophila*, which showed that AR proteins cause a significant defect in intracelluar replication (Habyarimana *et al.,* 2008).

Seven tetratricopeptide repeat (TPR) proteins, which similarly mediate protein–protein interactions and the assembly of multi-protein complexes (D'Andrea and Regan 2003), were also identified. TPR-containing proteins have been found to be involved in a variety of biological processes, such as cell cycle regulation, transcriptional control, mitochondrial and peroxisomal protein transport, neurogenesis and protein folding (Blatch and Lassle 1999, Goebl and Yanagida 1991). The genes encoding for the majority of AR and TPR proteins in the Deltaproteobacterium were localised closely together on the genome (Figure 2.6) and could represent a cluster for the mediation of protein-protein interactions within the cell or between cells.

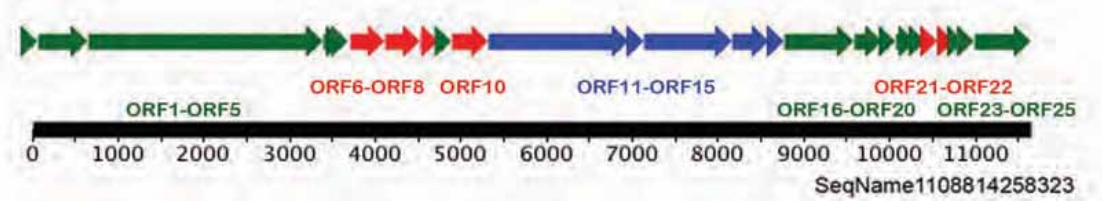


Figure 2.6 Schematic representation of the ORFs encoding AR (blue) and TRP (red) proteins in the Deltaproteobacterium scaffold 1108814258323.
ORFs encoding for other functions are shown in green and the black bar indicates the relative nucleotide position.

2.4 Conclusions

Through culture-independent approaches, a novel Deltaproteobacterium that possibly represents a new genus or family has been identified and characterised. This sponge-associated organism apparently represents only the third group of host-associated Deltaproteobacteria. The sponge Deltaproteobacterium is proposed to live in association with cyanobacteria and likely impacts on the physiology of its photosynthetic host. In turn it could have a consequence for the sponge host, which often harbours photosynthetic symbionts and thereby can acquire photosynthetically fixed carbon (Venn *et al.,* 2008). The functional genomic characterization of the uncultured Deltaproteobacterium discussed above also indicates complex interactions with surrounding cells and environmental milieu. This study offers an insight into multi-level interactions between a sponge, which serves as a host to a cyanobacterium, which in turn is host to a Deltaproteobacterium.

Chapter Three
Metaproteogenomic analysis of the microbial community associated with sponge *Cymbastela concentrica*

3.1 Introduction

Marine sponges represent a significant component of the marine, benthic communities throughout the world. Sponges harbour diverse communities of microorganisms, which often form stable and specific associations with their symbiotic host (Taylor *et al.,* 2007, Webster and Blackall 2008, Zhu *et al.,* 2008). While much progress has been made over the last decade in defining the phylogenetic diversity and patterns of sponge-associated microbial communities (Taylor *et al.,* 2007, Webster and Taylor 2011), information on the function of individual symbionts or the microbial community as a whole is somewhat limited.

Examples where specific members have been assigned functional roles, include cyanobacterial symbionts, which can provide photosynthetically fixed carbon to the sponge host (Wilkinson 1983) and the bacterial production of biologically active metabolites that may play a role in host defense (Schmidt *et al.,* 2000, Unson *et al.,* 1994). The processes of nitrification/denitrification and anaerobic ammonium oxidation (Anammox) have been well investigated in the sponges *Geodia barretti* (Hoffmann *et al.,* 2009), *Dysidea avara* and *Chondrosia reniformis* (Schlappy *et al.,* 2010) using stable isotope experiments, and by the identification of 16S rRNA gene sequences the Anammox process was putatively linked to planctomycetes in a reef sponge (Mohamed *et al.,* 2010) (also reviewed by Webster and Taylor (2011). These approaches require, however, an *a priori* knowledge of the processes performed in the sponge holobiont or the establishment of an irrevocable link between microbial phylogeny and function. Combining these limitations with the inherent difficulty of culturing (potentially obligate) symbionts has meant that there is only a rudimentary understanding of the ecological functions of sponge-associated microorganisms and the nature of the host-symbiont interactions (Webster and Blackall 2008).

The whole-community approaches of metagenomics, metatranscriptomics and metaproteomics provide a promising avenue to explore the function of uncultured organisms and for substantially advancing the field of sponge-microorganism symbiosis research. For example, a recent study in the laboratory on the microbial community associated with the sponge *Cymbastela concentrica* using metagenomics led to the recognition of many novel genomic markers that could provide specific mechanisms for bacteria to persist within, and interact with, their sponge host (Thomas *et al.,* 2010). To further explore the genetic potential offered by metagenomic datasets, metaproteomics is being increasingly employed to describe the expressed protein profile of microbial communities. Low-diversity microbial systems, such as those of acid mine drainage (Ram *et al.,* 2005) and lake water from Antarctica (Ng *et al.,* 2010), but also more complex systems in waste water sludge (Wilmes *et al.,* 2008a, Wilmes *et al.,* 2008b), the human microbiome (Chen *et al.,* 2008), the hindgut microbiome (Burnum *et al.,* 2011) and open ocean (Morris *et al.,* 2010, Sowell *et al.,* 2009, Sowell *et al.,* 2011) have been studied in this way. For these systems, the combination of high-throughput protein mass spectrometry with extensive metagenomic datasets has provided novel and direct insights into functions expressed by microorganisms.

To further the understanding of microbial functions in sponges, an integrated approach of using metagenome sequencing and metaproteomics on the microbial community associated with *C. concentrica*, an abundant marine sponge found in shallow, temperate waters of the Australian east coast was carried out. This sponge contains a stable and diverse microbial community, with predominantly uncultured phylotypes belonging to the *Alpha, Beta, Gammaproteobacteria, Phyllobacteriaceae, Sphingomondales, Piscirickettsiaceae* and *Deltaproteobacteria* amongst others (Thomas *et al.,* 2010). Results obtained here showed the expression of transport functions relevant to host-derived nutrients, aerobic and anaerobic metabolism, stress responses for the adaptation to variable conditions inside the sponge microbial community, as well as proteins that could facilitate a direct molecular interaction between the symbionts and the host. The data also revealed specific protein expression by a *Phyllobacteriaceae* bacterium and a *Nitrosopumilus*-like crenarchaeon, thus linking particular functions to uncultured phylotypes.

3.2 Materials and Methods

3.2.1 Sponge sampling and cell separation

Triplicate samples of *C. concentrica* were collected from Botany Bay near Bare Island, Sydney, Australia (S 33.59.461; E 151.13.946) at 10.00 am on September 15, 2009. Triplicate sponge samples of approximately 45-50 g wet weight were collected by SCUBA diving from a depth of about 7-8m and processed separately. Following collection, the *C. concentrica* tissue was directly transferred into calcium- and magnesium-free seawater (CMFSW; 25 g NaCl, 0.8 g KCl, 1 g Na_2SO_4 , 0.04 g $NaHCO_3$ per 1 L) containing Protease Inhibitor Cocktail VI (PIC; 2 µl per ml; A.G. Scientific, San Diego, US) to preserve the cellular protein content and then stored on ice. Samples were transported on ice to the laboratory at UNSW (approximately 15 minutes of travelling time) for direct and further processing, essentially as described previously (Thomas *et al.,* 2010). Briefly, surface barnacles and other macro-epibionts were physically removed from the sponge using sterile scalpels and forceps. Specimens were then washed twice for five minutes at 200 rpm agitation in CMFSW to remove loosely attached bacterial or particles. The washed sponge was cut into ~1cm^3 cubes and homogenised for 10-15 seconds (discontinuously to avoid foam formation) in 150 ml of fresh CMFSW with PIC (2 µl per ml). The sample was incubated on ice for 30 minutes with agitation at 150 rpm, then filtered through a 125 µm metal sieve into a sterile centrifuge tube and the filtrate centrifuged for 15 minutes at 100 x g at 4^oC to remove remaining sponge cells and tissues. The collected supernatant was centrifuged twice for 15 minutes at 200 x g at 4^oC to remove larger organisms, such as diatoms, from the sample. The resultant supernatant was filtered through a 11 µm filter using vacuum filtration unit and subsequently filtered through a 3 µm filter. The final filtrate was centrifuged for 20 minutes at 16,000 x g at 4^oC and washed twice with CMFSW with PIC. The final bacterial cell pellet was kept at -80^oC until DNA and protein extraction.

3.2.2 DNA sequencing, assembly, filtering and annotation

Shotgun libraries with fragment lengths around 400 bp were produced and sequenced from the extracted DNA using the Roche 454 Titanium platform (performed at the J. Craig Venter Institute, Rockville, USA). Dereplication of sequencing artefacts in the

raw reads was conducted by cd-hit-454 as outlined by Niu *et al.* (2010) with a similarity cut-off of 96% and the shorter replicate read having at least 95% of the length of the longer replicate. Dereplicated reads of each sample were assembled using the Newbler 2.3 assembler (454 Life Science, Branford, CT, USA) with the default settings. Contigs, Singletons and Outliers were pooled and sequences with less than 100 nt were removed.

During microbial cell fractionation, eukaryotic cells and organelles, such as mitochondria and plastids, may not have been sufficiently removed and the metagenomic data may therefore be contaminated with eukaryotic sequences. To remove such "contaminants", sequences were searched after assembly against the National Center for Biotechnology Information (NCBI) NT database (September 15, 2010) and the result files parsed through the last common ancestor algorithm implemented in MEGAN (v3.9) (Huson *et al.,* 2007). All sequences assigned to eukaryotic origin were removed according to the procedure described in Thomas *et al.* (2010).

Open reading frames (ORFs) of coding genes were predicted from the filtered sequences with the MetaGeneAnnotator (Noguchi *et al.,* 2008). This procedure resulted in a total of 2.8 million completed or partial genes. Predicted genes were translated to proteins and searched against the Conserved Domain Database (CDD) (Marchler-Bauer *et al.,* 2011) and the Clusters of Orthologous Group (COG) (Tatusov *et al.,* 2003) with rpsblast, and against the Pfam-A database (v24.0) (Finn *et al.,* 2010) with Hmmer 3, both with an E-value cut-off of 10^{-4}. Proteins with multiple heterogeneous domains were counted separately, while repeats of the same domain in a protein were counted once. Genes were also annotated by the SEED/ Subsystems using the online pipeline MG-RAST (v2) (Meyer *et al.,* 2008) with an E-value cut-off of 10^{-4}. A subset of protein sequences that did not return any hits was subjected to a search against the NCBI non-redundant protein database using BlastP (E-value cut-off of 10^{-4}).

The shotgun sequencing data is available through the Community Cyber infrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) website (http://camera.calit2.net/) under project accession 'CAM_PROJ_BotanyBay'.

3.2.3 Protein extraction and preparation

The cell pellet for each triplicate sample was resuspended separately in 1 ml of lysis buffer containing 10 mM Tris-EDTA (pH 8.0), PIC (2 µl/ml), 0.1% Sodium dodecyl sulfate (SDS; Sigma-Aldrich, Sydney, Australia) and 1mM dithiothreitol (Sigma-Aldrich, Sydney, Australia). The cell sample was then disrupted on ice by sonication with a Branson Sonifer (Danbury, CT, USA) for five cycles of 30 s on a 30% amplitude with 0.5 s on-off pulses. Microscopic analysis showed no intact cells after this lysis step. To desalt the protein sample, the final supernatant was transferred into a 5 kDa cut-off Amicon Ultra-15 filter unit (Milipore, MA, USA) and buffer exchanged with 3 ml of 10 mM Tris-EDTA (pH 8.0) followed by concentration to a smaller volume (~ 200 µl). The final protein concentration of the samples was determined using a bicinchoninic acid protein assay kit (Sigma-Aldrich, Sydney, Australia).

3.2.4 One-Dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1D SDS-PAGE) and in gel trypsin digestion

Protein samples were resuspended in appropriate volumes of SDS-PAGE sample buffer containing 187.5mM Tris (pH6.8), 30% glycerol, 6% SDS, 300mM DTT, 0.03# Bromophenol blue and sterilised MiliQ water. Samples were resolved on a 12% SDS gel using a Mini-PROTEAN system (Bio-Rad, Sydney, Australia), according to the protocol established by Laemmli (1970). The separating gel contained 0.375 M Tris-HCl (pH 8.8), 0.1% SDS, 0.06% ammonium persulphate (APS), 0.06% tetramethylethylenediamine (TEMED) and 12% acrylamide / bis-acrylamide solution (15:1 ratio, Bio-Rad, Sydney, Australia). The stacking gel consisted of 0.125 M Tris-HCl (pH 6.8), 0.3% SDS, 0.1% APS, 0.2% TEMED and 5.5% acrylamide / bis-acrylamide solution. The gel was run at a constant current of 15mA in running buffer containing 0.025 M Tris-HCl (pH 8.3), 0.192 M glycine and 0.1% SDS. The gel was stained with 0.25% Coomassie Blue solution in 50% methanol and 10% acetic acid for at least 4 hours. After staining, the gels were destained in a solution containing 40% methanol and 10% acetic acid for approximately one hour. Profile images were acquired using a conventional digital camera with a white background.

Whole gel lanes were sliced into equal proportions using a sterile gel cutter and each slice was washed with sterile Milli-Q water followed by washing with 25 mM NH_4HCO_3 in acetonitrile, to remove the Coomassie stain. The gel slices were treated separately through a series of reduction, alkylation and dehydration steps before digestion. Briefly, the slices were reduced with 10 mM DTT at 37ºC for 30 minutes, alkylated in 25 mM idoacetamide for 45 minutes and dehydrated in acetonitrile. In gel enzymatic digestion was performed by rehydrating the gel pieces in a buffer containing 80 ng /µl of trypsin (Promega, Sydney, Australia) and 25 mM NH_4HCO_3 at 37^{o}C for 14 hours. Digested peptides were extracted using 1% formic acid and acetonitrile and then dried using a Savant SpeedVac concentrator (Thermo Fisher, Melbourne, Australia).

3.2.5 High-performance liquid chromatography (HPLC) and mass spectrometry

Peptide digests were rehydrated in a buffer containing 1% formic acid and 0.05% heptafluorobutyric acid. Peptides were first separated by nano-LC using an Ultimate 3000 HPLC and autosampler system (Dionex, Amsterdam, Netherlands). Samples (2.5 µl) were concentrated and desalted on a micro C18 precolumn (0.5 mm X 2mm; Michrom Bioresources, Auburn, CA, USA) with H_2O:CH_3CN (98:2, 0.05% trifluoroacetic acid, TFA) at 15 µl / min. After a 4 min wash the precolumn was switched (Valco 10 port valve; Dionex) into line with a fritless nano column (75 µm X ~10 cm), containing C18 media (5 u, 200Å Magic; Michrom) manufactured according to Gatlin *et al.* (1998). Peptides were eluted using a linear gradient of H_2O:CH_3CN (98:2, 0.1% formic acid) to H_2O:CH_3CN (64:36, 0.1% formic acid) at 250 nl min^{-1} over 30 min. High voltage (1800 V) was applied to the low volume tee (Upchurch Scientific, Oak Harbor, WA, USA) and the column tip was positioned ~0.5 cm from the heated capillary (T = 250^{o}C) of an LTQ FT Ultra (Thermo Electron, Bremen, Germany) mass spectrometer. Positive ions were generated by electrospray and the LTQ FT Ultra was operated in data-dependent acquisition mode. A survey scan m/z 350–1750 was acquired in the FT ICR cell (resolution=100,000 at m/z 400, with an initial accumulation target value of 1,000,000 ions in the linear ion trap). Up to the six most abundant ions (4000 counts) with charge states of +2, +3 or +4 were sequentially isolated and fragmented within the linear ion trap using collision-induced dissociation with an activation, q = 0.25 and activation time of 30 ms at a target value of 30 000 ions.

M/Z ratios selected for mass spectrometry–mass spectrometry (MS/MS) were dynamically excluded for 30 s. Peak lists were generated using Mascot Daemon/extract_msn (Matrix Science, Thermo, London, UK) using the default parameters (version 2.2; Matrix Science).

Each peptide sample was subjected to three separate LC-MS/MS analyses, which resulted in a total of 108 sample runs (3 sponge samples, 12 slices per sample and 3 runs per sliced sample)

3.2.6 MS/MS data analysis and database searches

All 108 MS/MS spectra were analyzed using Mascot (version 2.3; Matrix Science, London, UK), Sequest (Thermo Fischer Scientific, San Jose, CA, USA; version 1.0.43.0) and X! Tandem (The GPM, thegpm.org; version 2007.01.01.1). All database searches were performed against a combined search database. This database was generated from all predicted protein sequences from the metagenomic analysis of *C. concentrica* from a previous Sanger-sequencing based study (Thomas *et al.*, 2010) and newly generated metagenomic data (see above). Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 4.0 ppm. Sequest was searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 3.0 ppm. Oxidation of methionine and iodoacetamide derivatives of cysteine were specified in Mascot, Sequest and X! Tandem as variable modifications. Oxidation of methionine, iodoacetamide derivatives of cysteine and acrylamide adduct of cysteine were specified in Mascot as variable modifications.

3.2.7 Protein identification and validation

To discriminate between false-positive and confident peptide matches, spectra for each sponge sample were pooled (36 MS/MS spectra per sample) and loaded into Scaffold (version Scaffold_2_00_05, Proteome Software Inc; Portland, OR, USA) as categorical samples for analysis to validate peptide and protein identifications. Peptide identifications were accepted, if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller *et al.*, 2002). Protein

identifications were accepted, if they could be established at greater than 99.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.*, 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone, were grouped to satisfy the principles of parsimony. The false discovery rate, as estimated by searches against a decoy database, was below 1%.

Data deposition: MS/MS data have been deposited in the PRIDE database (20959–20969). A list of proteins identified from the microbial community in sponge *C. concentrica* is provided in the Appendix 1. Protein databases as well as a full detailed list of identified proteins and peptides in the form of Excel files is available in Appendix Table A1 and A2 in the CD attached in this thesis.

3.2.8 Binning of metagenomic data and comparative genomics

Five dominant bacterial lineages, namely *Phyllobacteriaceae*, *Piscirickettsiaceae*, *Alpha, Beta and Gammaproteobacteria*, *Sphingomondales*, and *Deltaproteobacteria*, were previously constructed based on the binning of the tetranucleotide patterns identified from the metagenomic sequence scaffolds generated by Sanger sequencing (Thomas *et al*., 2010). The original five partial genomes were expanded by adding the original sequences and the NCBI reference genomes as reference genomes into the PhymmBL (Brady and Salzberg 2009) and by classifying all new metagenomic contigs (see above) longer than 1000 nt. Contigs assigned to any of the previously described genome bins (Thomas *et al*., 2010) were then used to expand the partial genomes. To classify the taxonomic origins of the 765 proteins identified in the metaproteome, the protein sequences were mapped back to the five partial genomes based on protein similarity and contig (DNA) identity. For protein similarity, 765 proteins were searched against all the proteins encoded by the partial genomes using BlastP (identity cut-off 95%), and found 65 proteins assigned to *Phyllobacteriaceae*, 5 to *Piscirickettsiaceae*, and 1 to *Alpha*, *Beta*, *Gammaproteobacteria*. For contig identity, the 746 contigs encoding these 765 proteins were extracted, and classified them against the expanded partial genomes using PhymmBL. Again, 65 proteins were assigned to

Phyllobacteriaceae, 17 to *Piscirickettsiaceae*, 15 to *Alpha, Beta, Gammaproteobacteria*, 1 to *Sphingomondales*, and 1 to *Deltaproteobacteria*.

Statistical, pairwise, comparison of the COG category profile and individual COGs between the proteome and partial genome data was performed by re-sampling (n=1,000) of COG subsamples (n=60) as outlined in Lauro *et al.* (2009).

3.2.9 Phylogenetic analysis of 16S rRNA genes and a crenarcheal AmoA

The 16S rRNA gene sequence of the *Phyllobacteriaceae* phylotype was taken from the previously classified partial genome from *C. concentrica* (Thomas *et al.*, 2010), which was also found in the current metagenomic dataset. This sequence was aligned with SINA (v1.2.9) (Pruesse *et al.*, 2007) and inserted it into the latest SILVA SSU Ref tree (version 1.08), using the parsimony function and the parsimony mask (pos_var_Bacteria_94) in the ARB software package (Ludwig *et al.*, 2004). Sequences for laboratory isolates belonging to the *Phyllobacteriaceae* were exported and its common conserved blocks were extracted using Gblocks (version 0.91b) (Castresana 2000). Filtered alignments such as RAxML (Cersion 7.2.8 alpha) (Stamatakis *et al.*, 2008) were used for phylogenetic reconstruction, with default setting and supported by rapid bootstrapping with 1000 resamples. The same procedure was used for the phylogenetic reconstruction of a *Nitrosopumilus*-like crenarchaeon, except that selection of sequences from uncultured strains in the Marine Group I was included (Figure 3.6).

For phylogenetic analysis of the ammonia monooxygenase subunit A (AmoA) protein, homologous sequences were obtained by iterative searches (10X) against the National Center for Biotechnology Information (NCBI) NR database (release September 15, 2010) with an E-value cut-off of 10^{-4}. Redundant sequences were removed using CD-Hit (version 4.3) (Li and Godzik 2006), with a similarity cut-off of 95%. A multiple sequence alignment was conducted using Muscle (version 3.8.31) (Edgar 2004). Representative sequences of bacterial AmoA and methane monooxygenase subunit A (PmoA) were selected according to the representatives used by Hallam *et al*. (2006b), and aligned respectively. The three alignments of archaeal, bacterial AmoA and PmoA were then aligned using a profile-profile alignment in Muscle. Conserved blocks were

selected using Gblocks and phylogeny was construed using RaxML with empirical based frequencies drawn from the alignment.

3.2.10 Fluorescence in situ hybridisation (FISH), probe design and evaluation

Specimens were prepared for microscopy according to the method outlined in Chapter 2. Briefly, hybridisation was performed in a hybrid chamber at 46℃ in hybridisation buffer [0.9 M NaCl, 20 mM Tris-HCl (pH 7.4), 0.01% sodium dodecyl sulfate (SDS) and 30% formamide] for 2 hours. Fluorescent probes were used at a concentration of 5 ng/ µl in hybridisation buffer [0.9 M NaCl, 20 mM Tris-HCl (pH 7.4), 0.01% SDS and 30% formamide]. A specific oligonucleotide probe PHY-CY3 (5'-CTCAATCTCGCGATCTCG-3') was designed and synthesised (Therom Fisher, Germany) to target the *Phyllobacteriaceae* phylotype with the Probe Design program within the ARB software package according to Hugenholtz *et al.* (2002). The target position was from 1258 to 1276 (*E. coli* numbering). A total of 303 partial 16S rRNA gene sequences related to the sponge *Phyllobacteriaceae* were used for the design (Thomas *et al*., 2010). The large multiple-sequence set used in the designing process ensures high specificity of the probe. Probe evaluation was also conducted according to Hugenholtz *et al.* (2002) and the optimal concentration of formamide determined. Slides were incubated in 50 to 100 µl of hybridisation mix dependent on the size of the sections, then carefully rinsed and further incubated at 43℃ in wash buffer [0.2 M Tris-HCl (pH7.4), 5 mM EDTA, 0.01% SDS and 100 mM NaCl, X] for 25 minutes. The wash buffer was carefully rinsed off and the slides air-dried in the dark. The anti-fading agent Citifluor (Citifluor Ltd., London) was mounted on the slides to prevent fluorochrome bleaching. Microscopy analysis was done with an Olympus FV1000 Laser Scanning Microscope (inverted) with excitation wavelength of 543 nm for the Cy3-labeled probe. Colour microscopy images were acquired using the camera attached to the microscope (Olympus, Japan) and processed with Adobe Photoshop CS. Presence and distribution of bacteria in sponge tissues were examined by investigating multiple regions from at least three replicate specimens.

3.3 Results and Discussions

3.3.1 Overview of metaproteogenomic data

Sequencing of DNA extracted from the microbial communities associated with three samples of *C. concentrica* resulted in a total of 2.8 million unique sequencing reads, which assembled into 988 317 contigs or singletons bigger than 100 nt. Of these, 687 588 passed a filtering procedure for eukaryotic contaminations. For each of the three sponge samples, 342 235, 550 559 and 555 480 proteins (ORFs) were identified respectively, and an average of 235 288, 208 852, 209 592, 284 549 could be annotated to CDD, COG, PFAM and SEED, respectively (Table 3.1).

The three proteomic datasets had a total of 765 non-redundant proteins identified from 5,275 peptide fragments. Taxonomic analysis indicated that 139 proteins were possibly from eukaryotic origin leaving 626 proteins for further functional characterization. Of these 367, 364 and 395 proteins were found in each of the three individual *C. concentrica* samples and 186 proteins were common to all three samples (see Figure 3.1 and Appendix 1). Protein sequences were annotated and clustered into functional categories. It was found that the expressed proteins in the sponge contained a substantial number of proteins (34%) with no assignment, or of hypothetical nature, suggesting the presence of many unrecognised functions in the sponge's microbial community.

Table 3.1 Information for the metagenomic analysis of sponge *C. concentrica*.

Sample	**BBAY40**	**BBAY41**	**BBAY42**
Raw read	678263	1169872	1323699
Average read size (nt)	358	408	393
Unique read	660869	1004075	1111093
Aligned read	347438 (53.8%)	704702 (72.0%)	679422 (63.3%)
Contig > 1000 nt	3389	3246	6925
Contig > 500 nt	11417	11398	26612
Average size of contigs > 500 nt (nt)	1185	1459	1119
N50 size of contigs > 500 nt (nt)	1345	2262	1192
Maximum size of contigs > 500 nt (nt)	28780	323086	39109
Contig > 100 nt	22865	22168	56410
Singleton > 100 nt	275175	251668	360031
Prokaryotic-originned contig and singleton	212117 (71.2%)	200120 (73.1%)	275351 (66.1%)
Unique protein	215220	229113	281686
Total protein	342235	550559	555480
Protein annotated by CDD (1e-4)	149902 (43.8%)	301223 (54.7%)	254738 (45.9%)
Protein annotated by COG (1e-4)	131452 (38.4%)	271121 (49.2%)	223983 (40.3%)
Protein annotated by Pfam (1e-4)	131095 (38.3%)	270007 (49.0%)	227674 (41.0%)
Protein annotated by SEED (1e-4)	182779 (53.4%)	361986 (65.8%)	308882 (55.6%)

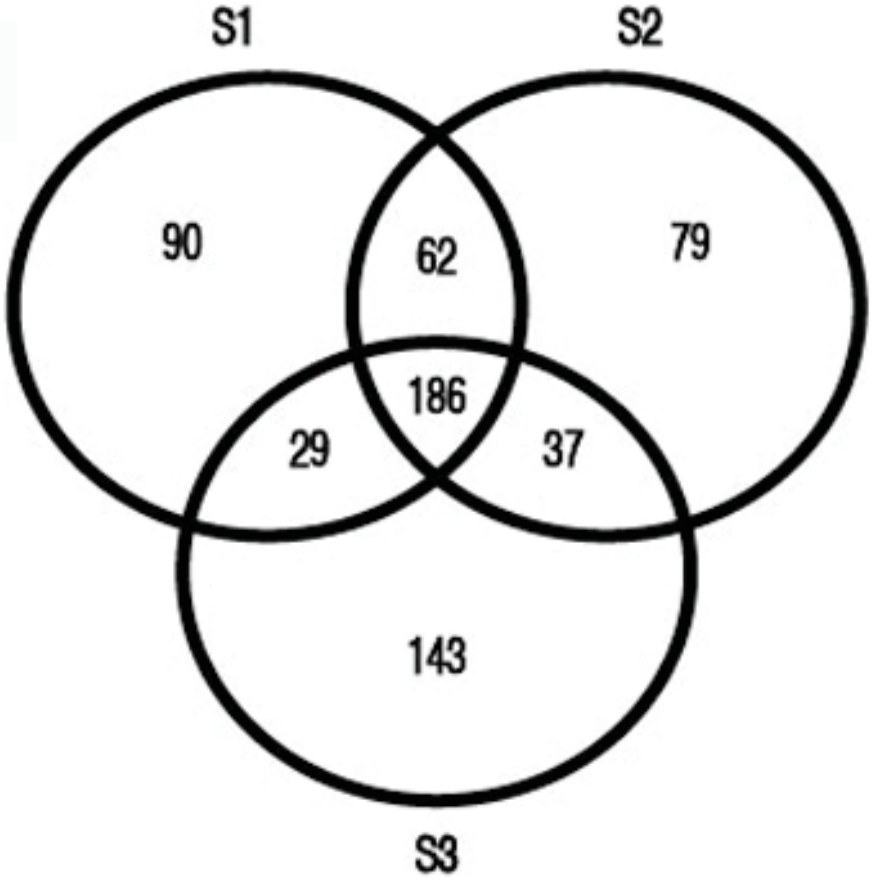


Figure 3.1 Venn diagram showing the distribution of proteins identified across three sponge samples.
S: Sample replicate.

To further compare the expressed functional profile of the microbial communities with their underlying genetic potential, the metaproteomic and metagenomic datasets were analysed based on Clusters of Orthologous Groups (COG) functional categories (Figure 3.2). Specific COGs that were relatively overrepresented in the metaproteome included carbohydrate transport and metabolism, post-translational modification, protein turnover, chaperone functions and signal transduction. A relative underrepresentation was observed for functional groups associated with coenzyme transport and metabolism, transcription, translation, replication recombination and repair, when compared to the metagenome dataset. Further analysis at the individual COG level, indicated an abundance (both in terms of proteins and peptides detected) of chaperonin GroEL (HSP60 family) (COG0459), an array of highly specific transporters (COG0747, COG0834, COG4663, COG0683, COG1653), dehydrogenases with different specificities, FabG (COG1028) and outer membrane receptor proteins CirA, OmpA (COG1629, COG2885) (Figure 3.3). Specific characterization of functions within each

category allowed for specifying physiological properties and activities of the sponge's microbial community.

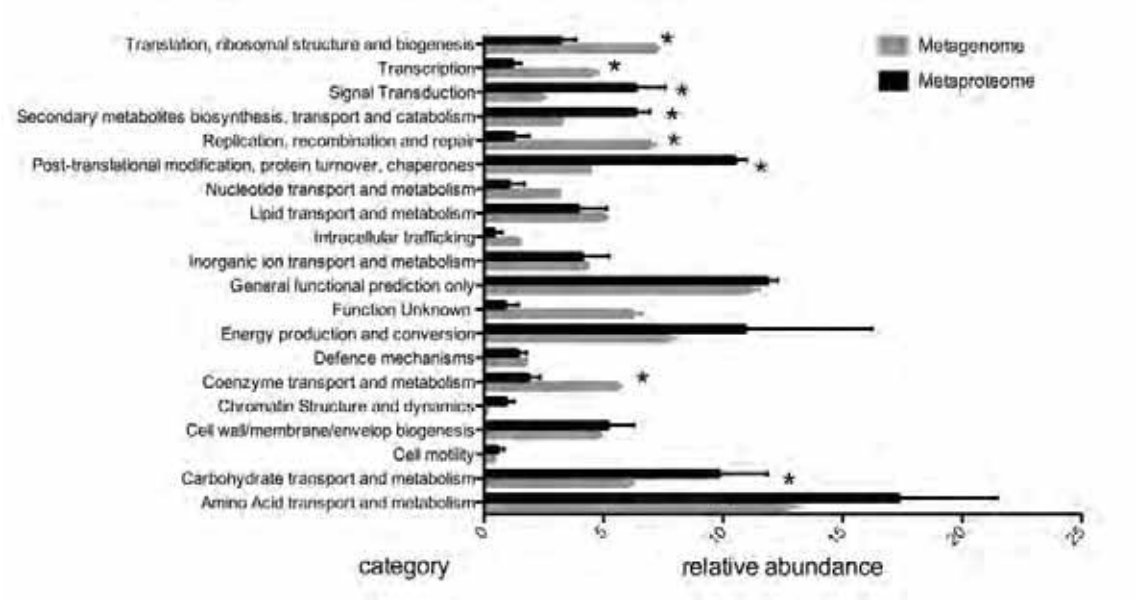


Figure 3.2 Relative abundance of COG categories in the sponge-associated microbial community based on metaproteome and metagenome data.
COG counts were normalised and the percentage of total counts in each COG categories presented above. The error bars show calculated standard variation of triplicate samples and asterisks indicate a statistical significant with a p value <0.05 in a t-test.

3.3.2 Active transport systems involved in nutrient acquisition

The sponge-associated community showed an abundant expression of high-affinity and broad-specificity uptake systems, such as ATP-binding cassette (ABC) transporters and tripartite ATP-independent periplasmic (TRAP) transporters. The most abundant transporter components detected in the metaproteome were periplasmic substrate-binding domains associated with ABC transporters, in particular for amino acids. DppA (COG0747) (Figure 3.3) is the substrate-binding component of the DppABCDEF dipeptide transport system, which has been demonstrated to transport proline-containing dipeptides (Olson *et al.,* 1991). Proline-containing dipeptides have been previously isolated from marine sponges (Liu *et al.,* 2009), however the exact production source (sponge or symbiont) is not known. Nevertheless DppA-type transporters were not observed as abundant in the metaproteome of planktonic bacteria (Sowell *et al.,* 2009, Sowell *et al.,* 2011), highlighting significant nutritional differences in the free-living and sponge-associated environment. Dipeptide transporters of the DppA type have also

been found to be capable of transporting heme and heme precursors (Letoffe *et al.*, 2006), indicating the potential scavenging of these iron-containing compounds from the surrounding or the sponge host. Other abundant proteins detected were HisJ and LivK, which are the periplasmic components of the high-affinity histidine- and leucine-specific transport system, respectively.

The overrepresentation of the COG category for carbohydrate transport and metabolism (Figure 3.2) is principally due to a high number of proteins, which appear to be associated with the glycerol-3-P ABC-type transporter UgpB (COG1653) and the TRAP system DctP (COG1638). UgpB is the periplasmic binding protein of the glycerol-3-phosphate uptake system (Brzoska *et al.*, 1994), which transports glycerol-3-phosphate for use as a carbon source and/or phosphate source. However, glycerol-3-phosphate uptake through the Ugp system is unable to supply sufficient carbon for bacterial growth, but instead only increases the internal phosphate concentration (Boos 1998). Therefore the Ugp system is ideally geared for scavenging phosphate-containing compounds (Boos 1998). The DctP system is well-characterised in *Rhodobacter capsulatus* (Forward *et al.*, 1997) and has also been found in *Wolinella succinogenes (Ullmann et al., 2000)*, where it is responsible for the transport of the C4-dicarboxylates, fumarate and malate (Ullmann *et al.*, 2000). Structurally, the known TRAP substrates are united by the presence of a carboxylate group so that hundreds of organic acids could be potentially transported by the system (Mulligan *et al.*, 2011). Large numbers of evolutionarily diverse TRAP transporters have been found in marine environments, especially in the SAR11 clade (Morris *et al.*, 2002), and this suggests an important role in the transport of diverse substrates.

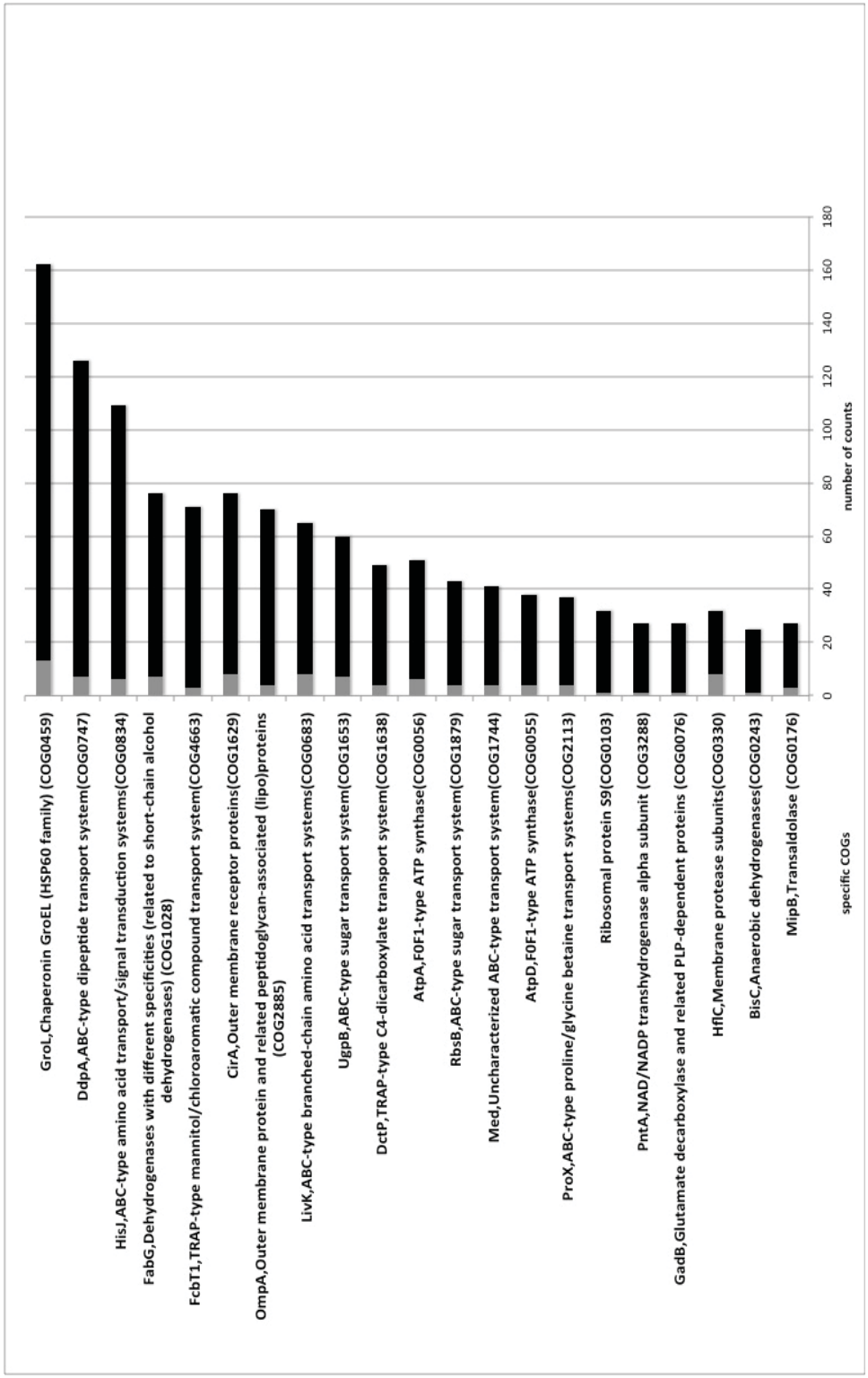


Figure 3.3 Abundance of specific COGs in sponge-associated microbial metaproteome.
Grey bar indicates the absolute counts on the protein level and black bar represents the absolute counts on the peptide level.

Another TRAP transporter expressed in the sponge-associated community belongs to the FcbT1 type, which has been shown in *Comamonas* sp. DJ-12 to transport halogenated, aromatic substrates, such as 4-chlorobenzoate (4-CBA) (Chae *et al.,* 2000). The genetic organization for this TRAP transporter is an operon encoding enzymes responsible for hydrolytic dechlorination of 4-CBA, which can be further metabolised to succinyl-CoA and acetyl-CoA (Nichols and Harwood 1995). FcbT genes can also be induced by benzoate derivatives like 4-bromobenzoate, indicating a wider potential substrate range of aromatic compounds (Chae *et al.,* 2000). Marine holobionts, including algae (Pedersen *et al.,* 1974), jellyfish (White and Hager 1977), polychaetes (Ashworth and Cormier 1967) and sponges (Schmitz and Gopichand 1978) have been recognised as a rich source of naturally occurring halogenated compounds, many of which have antibiotic or antagonistic activities. A number of sponge species, such as *Psammopemma* sp., *Psammaplysilla purpurea*, *Aplysina aerophoba*, and *Dysidea herbacea*, produce brominated aromatic metabolites, including bromoindoles, bromophenol (BP), polybrominated diphenyl ethers, and dibromodibenzo-p-dioxins (Ebel *et al.,* 1997, Gribble 1999, Norte *et al.,* 1988, Utkina *et al.,* 2001). These observations are therefore consistent with the uptake of halogenated aromatic compounds by the sponge's symbionts. Alternatively, the bidirectional nature of TRAP transporters (Poolman and Konings 1993) could facilitate export, making the symbionts the actual producers of the halogenated aromatics. Whichever way, the data show an intimate symbiotic relationship of the sponge with its symbionts through the transport of halogenated aromatic compounds.

TonB-dependent transporters (TBDTs) were also expressed by the sponge-associated microbial community, something that has also been observed for the microbial membrane metaproteome of specimens from the South Atlantic (Morris *et al.,* 2010). Specifically, the outer membrane receptor proteins CirA and OmpA, which utilise a proton motive force to transport nutrients across the outer membrane of Gram-negative bacteria, were detected. Genome studies on bacterioplankton have demonstrated TBDTs to be enriched among marine bacterial species (Giovannoni and Stingl 2007). The transport activities of TBDTs were thought to be restricted to iron complexes (siderophores) and vitamin B12 (cobalamin), but recent experimental and bioinformatic studies indicate that nickel, cobalt, copper, maltodextrins, sucrose, thiamin and chito-oligosaccharides are also suitable substrates (Schauer *et al.,* 2008).

Overall, the sponge-associated community clearly expresses a large number of transporters for the acquisition of various substrates, and in this respect behaves in a similar manner to the planktonic, bacterial communities from oligotrophic open oceans and productive coastal ecosystems (Morris *et al.*, 2010, Sowell *et al.*, 2009, Sowell *et al.*, 2011). Despite these broad similarities, there were clearly subtle differences in transport (e.g. dipeptides, halogenated aromatics) that reflect the nutrients specific for the microhabitats of the sponge.

3.3.3 Stress response

The abundance of expressed proteins associated with post-translational modification, protein turnover and chaperone functions (Figure 3.2) reflected the presence of a high number of chaperone proteins GroEL (HSP60, COG0459), membrane proteases HflC (COG0330) and DnaK (COG0443). These chaperones and proteases are essential for the elimination of denatured or damaged proteins, which could result from stress conditions such as temperature shifts, osmotic pressure, presence of reactive oxygen species and toxic compounds. In addition, the peptide methionine sulfoxide reductase MsrA, which repairs proteins that have been inactivated by oxidation (Ezraty *et al.*, 2005), is expressed. A number of proteins that are annotated to be heme-dependent peroxidases as well as the superoxide dismutase SodA, which eliminates harmful oxidation products like hydrogen peroxide were also detected (Perry *et al.*, 2010). In addition, peroxiredoxin (COG0450) and glutathione-S-transferases (COG0625) were expressed and these proteins might be important to control cytoplasmic redox balance (Hofmann *et al.*, 2002, Vuilleumier 1997). A choline dehydrogenase BetA, which catalyzes the oxidation of choline to glycine betaine was also detected (Landfald and Strom 1986). Betaines are potent and frequently used osmolytes that ensure osmotic balance in the cytoplasm and their production is often induced by osmotic stress.

Stress-related functions have previously been noted to be abundant in the genomes of bacteria associated with *C. concentrica*, compared to planktonic bacteria of the surrounding water (Thomas *et al.*, 2010). The cycling pumping activity of sponges (Vogel 1977) and steep local gradients (Hoffmann *et al.*, 2009), would expose bacteria to variable environmental conditions, in terms of availability of nutrients and electron

acceptors (e.g. oxygen). It appears that symbionts of *C. concentrica* are indeed dealing with such fluctuations in an evolutionary (abundance of genes) and physiological manner (expression of those genes).

3.3.4 Metabolism

The nitrogen metabolism of bacterial and archaeal symbionts is closely linked to the sponge host, which secrctes and accumulates ammonium (Bayer *et al.,* 2008, Hoffmann *et al.,* 2009, Schlappy *et al.,* 2010, Taylor *et al.,* 2007). It is therefore not surprising that the expression of the ammonia monoxygenase membrane-bound subunits β and γ (AmoB and C) and an ammonia transporter (AmtB) were also detected in the microbial community of *C. concentrica* (see Appendix 1). Both sequences were most closely related (BlastN identity: 92%), to those of the marine crenarchaeon *Nitrosopumilus maritimus* (Walker *et al.,* 2010). The genes encoding AmoB and C were adjacent and orientate in opposite transcriptional directions on a contig of the *C. concentrica metagenome* (Figure 3.4a). The contig also contains a gene for the α subunit (*amoA*) and has overall striking synteny with a genomic region of *N. maritimus* (Figure 3.4a, b). Phylogenetic analysis of *amoA* further confirmed the close relationship with the *N. maritimus* (Figure 3.5). Putative nitric oxide reductase subunits (*norQ* and *norD*) are also encoded in the genomic region (Figure 3.4 a, b) and might have a particular role in determining tolerance to nitric oxide under limiting oxygen concentrations or to allow for the use of nitrous oxide as an alternative electron acceptor (Schmidt and Bock 1997).

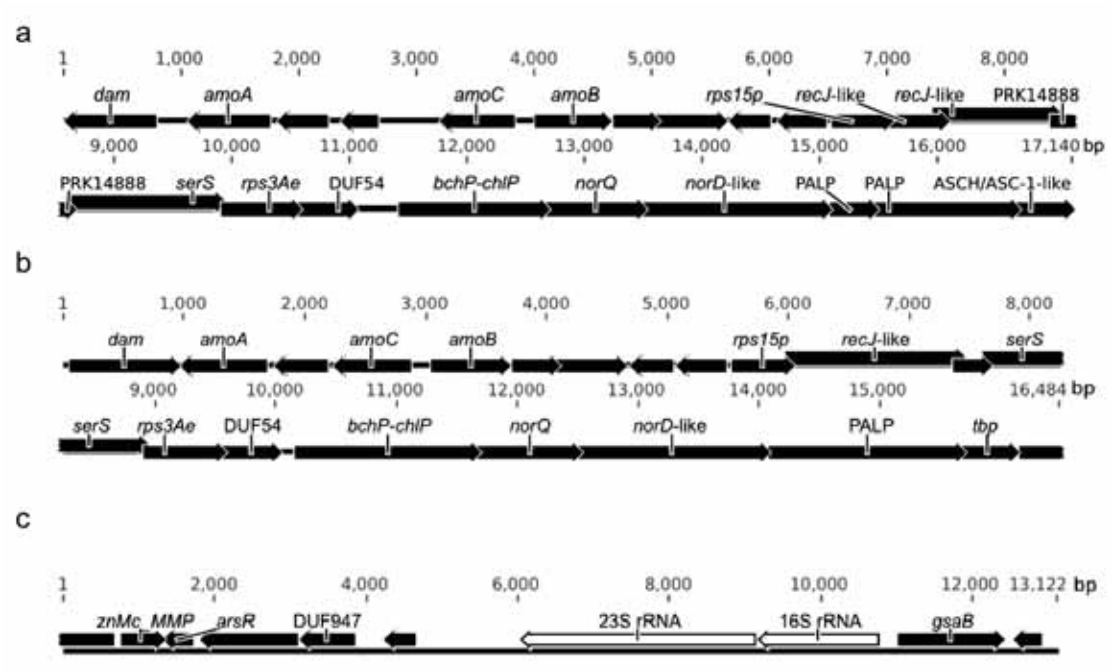


Figure 3.4 Contigs containing an archaeal *amoABC* gene cluster and a crenarchaeal 16S rRNA gene.
A. A contig from the *Nitrosopumilus*-like crenarchaeon in *C. concentrica* containing the *amoABC* gene cluster. B. A part of the *N. maritimus* SCM1 genome containing the *amoABC* gene cluster. C. A contig from the *C. concentrica* metagenome containing a crenarchaeal 16S rRNA gene. Gene annotation is as follows: ZnMc_MMP, Zinc-dependent metalloprotease, matrix metalloproteinase (MMP) sub-family; *arsR*, arsenic resistance operon repressor; DUF947, hypothetical protein with a DUF947 superfamily domain; *gsaB*, glutamate-1-semialdehyde aminotransferase, class III aminotransferase; *dam*, DNA adenine methylase; rps15p, 30S ribosomal protein S15P; PRK14888, PRK14888 superfamily protein; *serS*, Seryl-tRNA synthetase class II; rps3Ae, 30S ribosomal protein S3Ae; DUF54, DUF54 superfamily protein; *bchP-chlP*, geranylgeranyl reductase; *norQ*, moxR-like ATPase, nitric oxide reductase Q protein; *norD*, nitric oxide reductase Q protein; PALP, pyridoxal phosphate-dependent enzyme; tbp, TATA box binding protein; ASCH/ASC-1-like, ASC-1 homology domain, ASC-1-like subfamily. Hypothetical genes are not annotated. Contig layout was generated by Geneious 4.86 (http://www.geneious.com).

N. maritimus belongs to the C1a-α subgroup of crenarcheaeal Marine Group I, which contains many sequences obtained from sponges and plankton (Holmes and Blanch 2007). The phylogeny for the *Nitrosopumilus*-like crenarchaeon in *C. concentrica* was further defined. A 16S rRNA gene sequence in a metagenomic contig (Figure 3.4c) was identified and used to construct the phylogenetic relationship with other selected C1a-α subgroup members. This *C. concentrica*-derived crenarchaeal sequence clustered with a mix of 16S rRNA sequences from sponge-symbionts as well as free-living archaea (Figure 3.6). This shows that aerobic nitrification and transport of ammonia are active in *C. concentrica* and that these functions are likely to be carried out by a *Nitrosopumilus*-like crenarchaeon, which is a representative of common symbiotic and planktonic

archaea. These results also indicated that members of the nitrifying, archaeal clade could exist in either a host-associated or a free-living form, and switch between these stages while associated with the sponge *C.concentrica*. This could offer an explanation why crenarchaeal sequences were not detected in a previous metagenomic analysis of *C. concentrica* (Thomas *et al.,* 2010). The *Nitrosopumilus*-like crenarchaeon may have existed in a free-living form and hence was not captured by the bacterial extraction procedure.

Given that both aerobic and anaerobic conditions might exist within micro-habitats of the sponge tissue, the metaproteomic dataset was investigated for functions related to anaerobiosis. The expression of proteins annotated to COG0076 (GadB; Glutamate decarboxylase and related PLP-dependent proteins) (Appendix 1) part of the glutamate-dependent acid resistance systems (AR2) was identified. AR2 protects cells during anaerobic phosphate-starvation, when glutamate is available, by preventing damage from weak acids produced by carbohydrate fermentation. Although no proteins directly associated with carbohydrate fermentation were detected, acetoacetate decarboxylase (Appendix 1), which is involved in the solventogenesis of the typical fermentation products butyric and acetic acid into acetone and butanol (Schaffer *et al.,* 2002) was expressed. Anaerobic degradation of amines and polyamines may also occur because the expression of crotonobetainyl-CoA hydratase (CaiD) (COG1024) (Appendix 1), which is part of the carnitine degradation pathway (Elssner *et al.,* 2001), was detected. In *E. coli,* this pathway, which includes the dehydration and reduction of L-carnitine to γ-butyrobetaine, is induced during anaerobic growth. The carnitine pathway has also been found to generate the osmoprotectant betaine during anaerobic respiration (Elssner *et al.,* 2001, Kleber 1997, Preusser *et al.,* 1999). Recent analysis of anaerobic carnitine reduction in *E. coli* has also shown that the electron transfer flavoproteins, *fixA* and *fixB* genes are necessary for the transfer of electrons to crotonobetaine reductase (CaiA) (Walt and Kahn 2002) and FixA was found expressed in the sponge community (Appendix 1). Overall the data provide evidence for the existence of both aerobic and anaerobic metabolism in the sponges.

Figure 3.5 Maximum-likelihood tree of archaeal and bacterial AmoA sequences.
The tree is rooted with PmoA AAQ10310, CAE47800 and AAA87220. Bootstrap values of larger than 50% (for 1000 replicates) are only shown in the tree nodes. Bold cases indicate sponge-derived sequences.

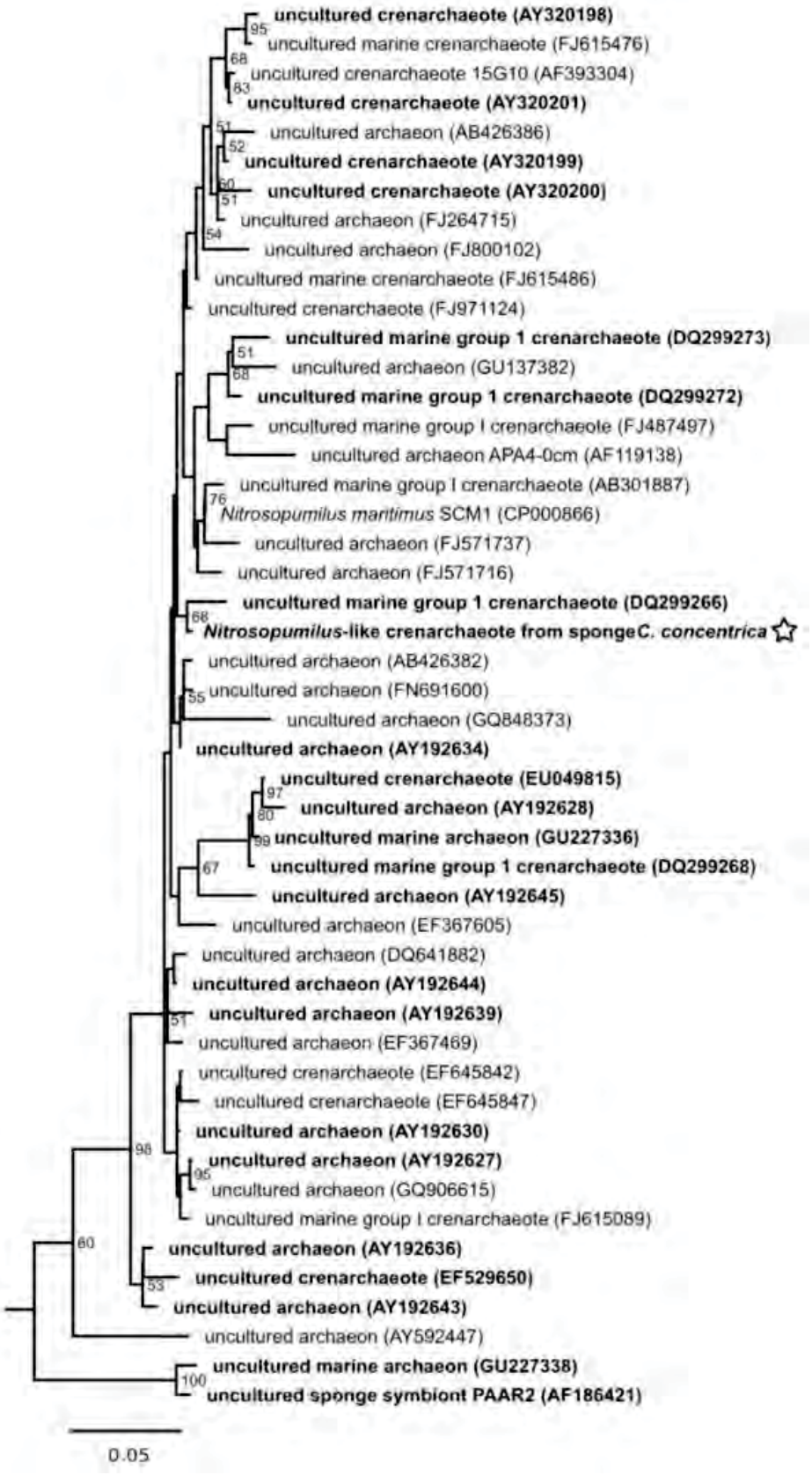


Figure 3.6 16S rRNA gene maximum-likelihood tree of the *Nitrosopumilus*-like crenarchaeon and *N. maritimus* SCM1 with other lineages in C1a-α Marine Group I.
Bootstrap values of larger than 50% (for 1000 replicates) are only shown in the tree nodes. Bold cases indicate sponge-derived sequences.

3.3.5 Molecular symbiont-host interactions

A substantial overrepresentation of ankyrin repeat (AR) and tetratricopeptide (TPR) proteins was observed in a recent metagenomic study of the bacterial communities associated with *C. concentrica* from this laboratory (Thomas *et al.,* 2010). The genes for these eukaryotic-like proteins also clustered in the genome of an uncultured, Deltaproteobacterium of the same sponge (Section 2.3.4). Genes encoding AR proteins have been reported to be abundant in the genomes of obligate and facultative symbionts, such as *Wolbachia pipientis* (Iturbe-Ormaetxe *et al.,* 2005), *Ehrlichia canis* (Mavromatis *et al.,* 2006), *Legionella pneumophila* (Habyarimana *et al.,* 2008) and *Coxiella butnetii* (Voth *et al.,* 2009) and could have a function in mediating host–bacteria associations. This possibility was highlighted recently by a mutant study of *L. pneumophila*, which showed that certain AR proteins controlled the intracellular replication of *L. pneumophila* within the amoebal host (Habyarimana *et al.,* 2008). An AR protein (COG0666) and a TPR protein (PFAM00515) were found in the metaproteomic dataset (Appendix 1) confirming that sponge-associated microorganisms indeed express those proteins, potentially for mediating interactions with the sponge host, as was previously proposed (Thomas *et al.,* 2010). Other proteins with roles in bacteria-eukaryote interactions were also expressed, including proteins with Hep/Hag domains (Appendix 1). The seven-residue repeat domain of Hep/Hag is contained in the majority of the sequences of bacterial hemagglutinins and invasins. The adhesion, YadA was also detected in the metaproteome. This protein has been shown to be responsible for phagocytosis resistance (Nummelin *et al.,* 2004). Sponges are filter feeders, and the presence of such eukaryotic-like proteins and bacteria-eukaryote mediators, suggest that sponge symbionts might use these proteins to escape phagocytosis and/or control their symbiotic relationship with their hosts.

3.3.6 Linking phylotype to function – expression profiling of an uncultured Phyllobacteriaceae phylotype

Binning of previous sequencing data for the metagenome of *C. concentrica* (Thomas *et al.,* 2010) and the current pyrosequencing dataset facilitated the reconstruction of partial genomes of five uncultured sponge symbionts; i.e. phylotypes classified to belong to the *Phyllobacteriaceae, Piscirickettsiaceae, Alpha, Beta, Gammaproteobacteria,*

Sphingomondales and *Deltaproteobacteria* (Thomas *et al.,* 2010) (see also Material and Methods). A total of 65 proteins were assigned to the *Phyllobacteriaceae* genome, 17 to *Piscirickettsiaceae*, 15 to *Alpha, Beta, Gammaproteobacteria*, one to *Sphingomondales* and one to *Deltaproteobacteria*. As the number of hits against the *Phyllobacteriaceae* genome was the highest, the specific dataset for this phylotype was further investigated for its expression profile and genomic features.

Detailed phylogenetic analysis established that the 16S rRNA gene sequence within the *Phyllobacteriaceae* partial genome forms, together with two sequences previously amplified from *C. concentrica* (AY942778 and AY942764), belong to a distinct clade that is deeply branched within the *Phyllobacteriaceae* family (Figure 3.7). A large majority of members of the *Phyllobacteriaceae* are plant-associated and have been well studied with respect to their potential to promote plant growth (Mantelin *et al.,* 2006). They also occupy diverse habitats, such as soil (Jurado *et al.,* 2005), water (Mergaert *et al.,* 2001) and unicellular organisms (Alavi *et al.,* 2001), suggesting a remarkable adaptive capacity to different environmental niches. The sponge sequences of this study are also related to nitrogen-fixing *Mesorhizobium* species and denitrifying *Nitratireductor* species (Labbe *et al.,* 2004), which could imply that the *Phyllobacteriaceae* phylotype in *C. concentrica* is potentially involved in nitrogen metabolism.

The proteome assigned to the *Phyllobacteriaceae* phylotype was compared to its partial genome and found to have over- and underrepresentation of functional COG categories (Figure 3.8a), similar to the overall metaproteome and metagenome comparison (Figure 3.2). COG categories of amino acid transport and metabolism, post-translational modification, protein turnover, chaperone functions and signal transduction were also overrepresented in the expressed proteome data (Figure 3.8a). At the individual COG level, specific transport functions were overrepresented (Figure 3.8b), with both ABC-type transporters and TRAP-type transporters being expressed. Of particular interest was the ABC-type nitrate/sulfonate/bicarbonate transport system TauA, which can import nitrate across the cell membrane. A nitrate reductase gene cluster (*narG*, *narH*, *narI*, and *narY*) was present in the partial genome of the *Phyllobacteriaceae* phylotype, however the expressed proteins NarG and NarY in the metaproteome dataset (Appendix 1), could not be unambiguously assigned to this organism. Nevertheless, the proteomic

data suggests that denitrification is actively expressed and most likely takes place in the anaerobic, inner part of the sponge (Hoffmann *et al.,* 2008). Based on this observation, the physical location of the *Phyllobacteriaceae* phylotype within sponge tissue was investigated using fluorescence *in situ* hybridisation (FISH). No FISH signal was observed near the outer surface of the sponge and the *Phyllobacteriaceae* phylotype was found to reside within the mesohyl, mainly near the chambers, where feeding takes place in sponges (Figure 3.9).

An expressed transposase gene was also found assigned to the *Phyllobacteriaceae* phylotype. Its chromosomal location was flanked by two genes associated with lipid metabolism (cyclopropane-fatty-acyl-phospholipid synthase and acyl-CoA dehydrogenase). Transposons and transposases were found to be abundant in the metagenome of *C. concentrica* and these elements may be an important part of the genomic adaptation process towards a symbiotic relationship between bacteria and host (Thomas *et al.,* 2010). The observation showed that the transposase function is still active and that the transposase genes are not all just remnants of past events of intra- or intercellular horizontal gene transfer.

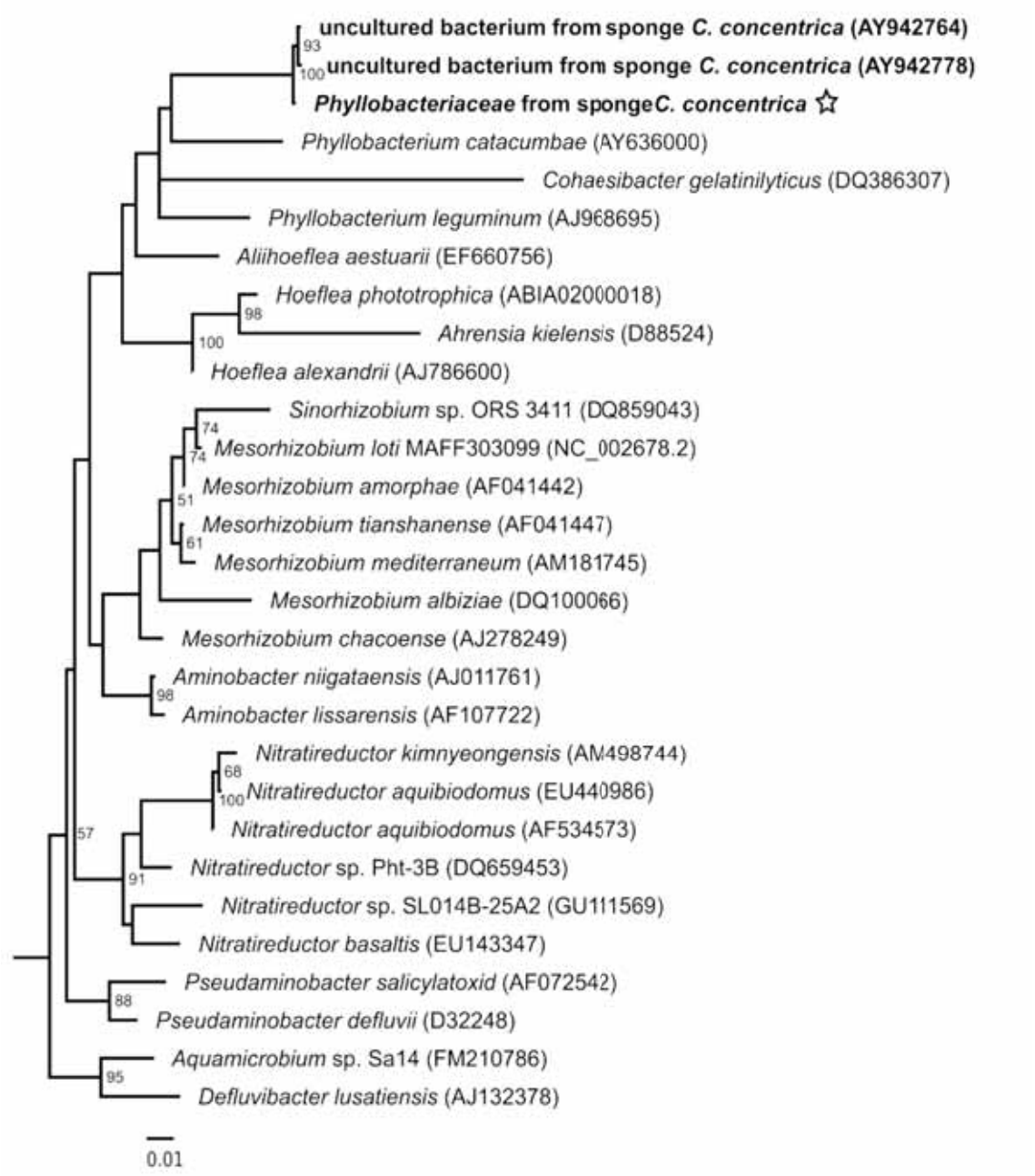


Figure 3.7 Maximum likelihood tree showing the *Phyllobacteriaceae* phylotype and phylogenetic relationship to its closely related neighbour phylotypes.
The tree was built base on the 16S rRNA gene. An uncultured crenarchaeote was used as outgroup for the analysis (not shown in the tree). The scale bar indicates 0.1 nucleotide changes (10%) per nucleotide position.

3.4 Conclusions

This analysis has provided new insights into the activities of sponge-associated microbial communities, and the *Phyllobacteriaceae* phylotype has indicated a clear link between uncultivated phylotypes and functions. For what appears to be the first time,

specific transport functions for typical sponge metabolites (e.g. halogenated aromatics, dipeptides) could be identified. While the co-existence of aerobic and anaerobic phases of the nitrogen cycles has been observed for another sponge system (Hoffmann *et al.*, 2009), the result presented here could assign those functions to specific, expressed proteins and phylotypes in *C. concentrica*. The analysis also indicated the requirement for the microbial communities to respond to variable environmental conditions and hence express an array of stress protection proteins. Finally, molecular interactions between symbionts and their host might also be mediated by a set of expressed eukaryotic-like proteins and cell-cell mediators, and some sponge-associated bacteria (e.g. *Phyllobacteriaceae* phylotype) could be undergoing evolutionary adaptation process to the sponge environment, as evidenced by active mobile genetic elements.

The results presented here have clearly shown that a combined metaproteogenomic approach can provide novel information on the activities and physiology of sponge-associated, microbial communities. This approach has proven not only to be useful for investigating the enormous microbial diversity found in sponges around the world, but also for investigating the functional behaviour of symbiont communities in response to environmental change or host physiology (Webster and Blackall 2008). Such observations will be crucial to understand the dynamic and complex interactions between sponges and their associated microbial symbionts.

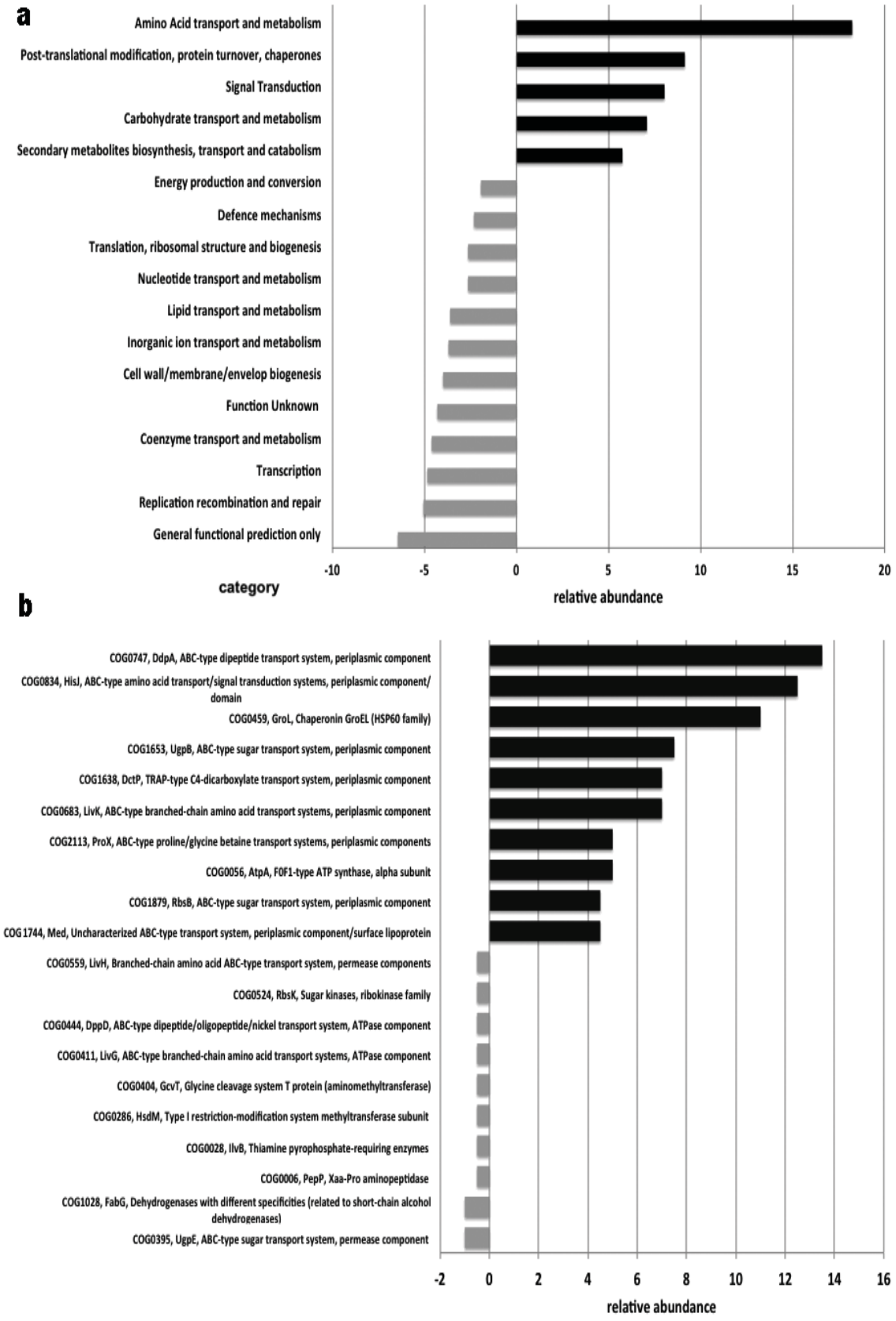


Figure 3.8 Comparison between expressed proteome and partial genome of the *Phyllobacteriaceae* phylotype on the level of COG categories (a) and specific COG level (b). Black and gray bars represent over- and under-representation, respectively, of identified proteins. The X axis displays the median value with significant cut-off value of -1 and 1.

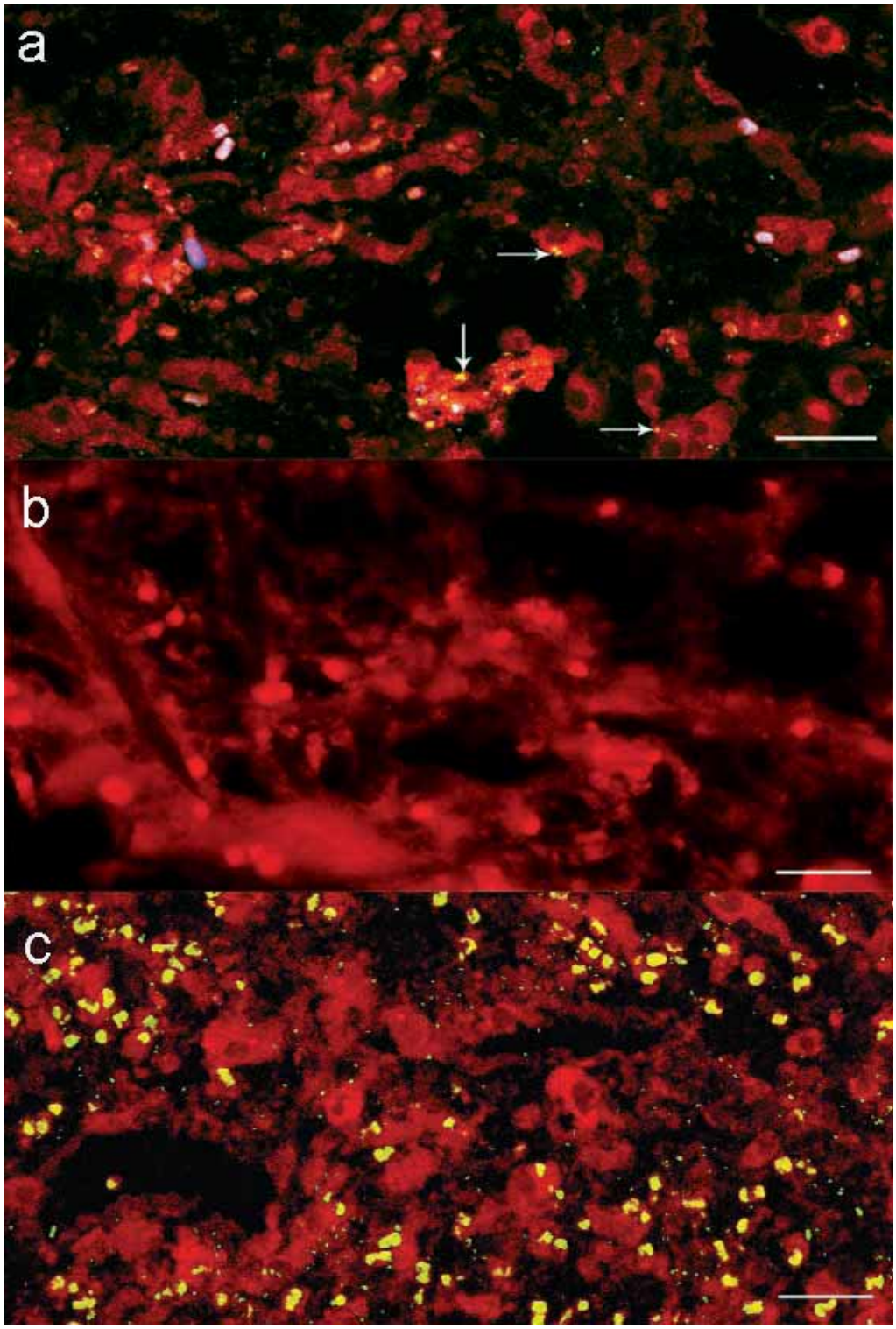


Figure 3.9 Detection of sponge-associated bacteria and the *Phyllobacteriaceae* phylotype by fluorescence *in situ* hybridisation (FISH).
(a) Section of *C. concentrica* hybridised with a Cy3-labelled specific (PHY_Cy3) probe (green or sometimes yellow, if the signal overlapped with the red autofluorescence of the sponge tissue). Arrows indicate the presence of the *Phyllobacteriaceae* phylotype within sponge tissue (mesohyl). (b) Section of non-hybridised *C. concentrica* showing auto-fluorescence background (red). (c) Section of *C. concentrica* hybridised with a Cy3-labelled general bacterial (EUB338_CY3) probe (green) with the auto-fluorescent sponge tissue (red) and larger auto-fluorescence cells (yellow). Scale bar: 20 μm.

Chapter Four
Functional genomic analysis of a dominant *Phyllobacteriaceae* phylotype from the sponge metagenome

4.1 Introduction

Cultivation-independent approaches based on sequencing of DNA have generated a large amount of information about the diversity of microorganisms from natural environments. We now know that only a small fraction of microorganisms from natural environments is amenable to culturing. This is particularly the case for well-studied marine ecosystems, where it is estimated that only 0.01 to 0.1% of marine bacteria produce colonies on standard enrichment plates (Amann *et al.,* 1995). A large discrepancy between the culturable population and the total microbiota was also evident during cultivation studies of sponge-associated microorganisms. As examples, the cultivable microbial diversity from the Great Barrier Reef sponge *Rhopaloeides odorabile* (Webster and Hill 2001) and the Caribbean sponge *Ceratoporella nicholsoni* (Santavy *et al.,* 1990), ranged as low as 0.1-0.23% and 0.15%, respectively. Higher levels of 5% and 11% were reported to be culturable from the deep sea sponge *Scleritoderma* spp. (Olson and McCarthy 2005) and the Mediterranean sponge *Aplysina aerophoba* (Friedrich *et al.,* 2001), respectively, however it stills means that the majority of sponge-associated microorganisms are not amenable to culturing. A recent study using the sponge *C. concentrica* also found that the microbial isolates did not overlap with the community identified using culture-independent techniques such as 16S rRNA gene sequencing (Yung 2010). In addition, the isolates were found to be more closely related to those from seawater and marine invertebrates, suggesting that the majority of the isolates cultured using standard media were not specific to the sponge.

In recent years, a number of studies have been published using culture-independent approaches to study the microbial diversity within sponges. Further accelerating this development, recent reports on the use of the next generation sequencing technologies such as 454-tag pyrosequencing have demonstrated an exceptional microbial diversity

within sponges (Lee *et al.,* 2011, Webster *et al.,* 2009). In particular, sequence data of Great Barrier Reef sponges have revealed high diversity of microbial symbionts within three sponge species and suggested that sponge symbionts are acquired in two evolutionary ways, vertical transmission and acquisition of symbionts from seawater (Webster *et al.,* 2009). Further, novel functional signatures have been identified from recent metagenomics (Thomas *et al.,* 2010) and metaproteomics (Chapter 3) studies on sponge microbial communities.

Parallel to the development of culture-independent studies, methods for cultivation, based on modified traditional approaches, have also resulted in the isolation of microorganisms previously considered "unculturable". Perhaps the most significant example was the isolation of the ubiquitous Alphaproteobacterium, *Candidatus Pelagibacter ubique*, using a dilution to extinction method (Rappe *et al.,* 2002). This bacterium has been found in nearly all marine bacterioplankton communities studied, but has long defied cultivation. Natural seawater was used as the growth medium and applied to a high-throughput screening method with specific fluorescent probes. The advantage of using the dilution to extinction method is the removal of subpopulations of fast growing competitors from the sample, and maintaining the numerically dominant slow growing bacteria. The same strategy was also effectively applied to the isolation of *Sphingopyxis alaskensi,* a well-studied oligotrophic bacterium (Eguchi *et al.,* 2001, Ostrowski *et al.,* 2001, Schut *et al.,* 1993). A genome-directed isolation and characterisation approach was also developed using information derived from the massive sequencing data generated from culture-independent methods. Functional information revealed from the sequence data could be used to reconstruct potential functionality, such as metabolic and physiological pathways. These complete or partially complete pathways often provide valuable clues on how selective strategies can be developed to isolate an organism. For example, an autotrophic ammonia-oxidizing marine archaeon *Nitrosopumilus maritimus* SCM-1 was successfully isolated using a specific strategy that excluded other microorganisms with the use of antibiotics (Konneke *et al.,* 2005). The shotgun sequencing study of the acid mine drainage biofilm recovered >95% of the genomes of five dominant organisms (Tyson *et al.,* 2004) and one of the members, *Leptospirillum ferrodiazotrophum* was also isolated by establishing that this organism was capable of fixing nitrogen (Tyson *et al.,* 2005).

Having access to pure cultures of microorganisms representative of a given ecosystem benefits research in several ways. First, it allows for the full genome of the organism to be sequenced. Second, knowledge of the full genome improves the accuracy of functional information derived from culture-independent studies like transcriptomics and proteomics. Third, and most importantly, it allows experiments to be carried out to test the hypotheses derived from genomic data. It is anticipated that characterization of cultures of representative organisms, in combination with culture-independent approaches, could ultimately be the best way to study ecological processes in the environment.

Five dominant sponge symbionts were identified, from the metagenomic analysis of *C. concentrica* (Thomas *et al.,* 2010), and found by metaproteomic analysis to be functionally active in the host sponge (Chapter3). Within these five dominant symbionts, functional expression of a distinct *Phyllobacteriaceae* phylotype was found to be predominantly active. It was clear that specific transporter functions were expressed by this sponge symbiont, reflecting its potential activity in nutrient acquisition. An expressed transposase gene flanked by genes associated with lipid metabolism was also detected and demonstrated that a function in evolutionary adaptation, through exchange of genetic material is on-going among sponge symbionts and likely to be associated with metabolism.

A partial genome of this *Phyllobacteraceae* phylotype was re-constructed from metagenomic data and used to predict aspects of its physiology, metabolism and evolutionary adaptation. It was predicted that this organism would be capable of degrading a number of aromatic compounds and to carry out anaerobic respiration through nitrate reduction. The aim of the work of this chapter was to use the information from the annotation of the genome and to develop a metagenome directed method, based on a dilution to extinction method, to isolate a pure culture of the sponge *Phyllobacteriaceae* phylotype. Preliminary attempts using selective media identified the growth of marine associated isolates. A mixed cell culture was also detected when incubated anaerobically. While a pure culture of the *Phyllobacteriaceae* phylotype was not obtained, the results clearly demonstrated the possibility of isolating ‘unculturable’ microorganisms using the metagenome directed method.

4.2 Materials and Methods

4.2.1 Genomic binning and partial genome reconstruction

Five dominant bacterial partial genomes were previously constructed and expanded (Section 3.2.8) and the partial genome belonging to the *Phyllobacteriaceae* was used. The expansion process using PhymmBL could have falsely binned some non-*Phyllobacteriaceae* phylotype pyrosequencing contigs into this partial genome. To minimise the false-positive rate, a stringent cut-off to the pyrosequencing contigs based on the coverage and GC contents of the Sanger contigs was applied. This cut-off ensured contigs bigger than 3 kb with coverage greater than 66 times and GC content between 57% and 63% were binned to the *Phyllobacteriaceae* phylotype partial genome. Genes were predicted from this refined partial genome using the MetaGeneAnnotator (Noguchi *et al.,* 2006) and translated into protein sequences. To assess completeness, the refined partial genome of *Phyllobacteriaceae* phylotype was searched against the 40 COGs of universal, single-copy genes (Ciccarelli *et al.,* 2006) using rps-Blast, with an E-value cut-off of 10^{-9}.

4.2.2 Partial genome annotation and comparative genomics

Functional annotation of the partial genome was performed with the Rapid Annotations based on the Subsystem Technology (RAST) server (Aziz *et al.,* 2008). The RAST is designed for the rapid "calling" and annotation of the genes in a metagenomic dataset using a "highest confidence first" assignment propagation strategy, based on manually curated subsystems and subsystem-based protein families that automatically guarantee a high degree of assignment consistency (Aziz *et al.,* 2008). The complete predicted protein sequences of *Mesorhizobium loti* MAFF303099 (6,743 proteins) were downloaded from the National Center for Biotechnology Information. The partial protein dataset for the *Phyllobacteiaceae* phylotype and complete protein dataset of *M. loti* were searched with blastP (E-value $<10e^{-5}$), against the Cluster of Orthologous Groups of proteins (COG) database (Tatusov *et al.,* 2003). Proteins with at least 30% identity to an entry were then classified into specific COGs and COG categories. Statistical, pairwise, comparison of the COG profiles between the sponge

Phylobacteriaceae phylotype and *M. loti* was performed by resampling (n=1000) of COG subsamples (n=2000) as outlined in (Lauro *et al.,* 2009).

4.2.3 Sample collection, cell preparation and media preparation

Specimens of the sponge *C. concentrica* were collected from Botany Bay near Bare Island, Sydney Australia (S 33.59.461;E 151.13.946) during February 2010 and transported on ice to the laboratory at UNSW (approximately 15 minutes of travelling time) for immediate processing. The specimens were processed and the bacterial cells extracted as described above (Section 3.2.1), except that the collected supernatants were used immediately for attempts at cultivation. To determine the bacterial cell densities of the extract, direct cell counts were done by 4',6-diamidino-2-phenylindole (DAPI) staining, and counted using a KOVA haemocytometer slide (Hycor, USA), under a fluorescence microscope.

The artificial seawater used for media preparation was prepared in 2 litres batch according to the method of Kester *et al.* (1967). To avoid precipitation of metal ions during autoclaving, the anhydrous salt components (NaCl, Na_2SO_4, KCl, KBr, $NaHCO_3$, KBr, H_3BO_3 and NaF) and hydrated salt components ($MgCl_2$, $CaCl_2$ and $SrCl_2$) were dissolved and autoclaved separately (Table 4.1). To prepare the solution, a litre of each the anhydrous salt solutions and the hydrated salt solutions were used to give a total of two litres. After autoclaving, two solutions were added together. The pH and salinity of the artificial seawater were determined using a conductivity temperature depth (CTD) unit (YSI, Australia). The pH of the water used was 8.0 and the salinity (*S*) was 33 p.s.u (practical salinity unit). Before use, the water was filtered using a 0.22 μm syringe filter (Millipore, Australia).

Table 4.1 Composition of artificial seawater.

	$g \cdot L^{-1}$ stock solution	Final concentration in media (L^{-1})
Salt solution I—*anhydrous salts*		
NaCl	23.926	410.4mM
Na_2SO_4	4.008	28.2mM
KCl	0.667	9.37mM
$NaHCO_3$	0.196	2.18mM
KBr	0.098	840µM
H_3BO_3	0.026	410µM
NaF	0.003	67.7µM
Salt solution II—*hydrated salts*		
$MgCl_2 \cdot 6H_2O$	9.7	52.8mM
$CaCl_2 \cdot 2H_2O$	1.1	10.2mM
$SrCl_2 \cdot 6H_2O$	0.023	82µM

Components prepared according to Kester *et al.* (1967).

A base medium, prepared using sterile artificial seawater (Table 4.2) was used for the preparation of liquid and solid media. All supplements were prepared and filter sterilised with a 0.22µm syringe filter (Millipore, Australia) prior to use. Various filtered sterilised carbon sources were supplemented to make the specific liquid and solid media (Table 4.3). All liquid media were filter sterilised again with a 0.22µm syringe filter (Millipore, Australia). The solid media were prepared by dissolving 1.5% agar in artificial seawater and autoclaving. Filter sterilised nutrient supplements (carbon source, phosphate, nitrogen sources and vitamin mixture), were added to the agar media when it cooled down. Media plates were poured in a class two biological safety cabinet (Email Air-handling, Australia) to avoid contamination.

Table 4.2 Composition of the base medium, used for the preparation of the specific media used in the isolation attempts.

	$g \cdot L^{-1}$ stock solution	Final concentration in media (L−1)
Base media		
1. K_2HPO_4	1.74	10mM
2. NH_4Cl	5.4	100mM
2. Vitamins		
Vitamin working stock	to every 1 litre of base media, add 40ul of stock solution	
Components of the vitamin stock	$mg \cdot 100mL^{-1}$ stock solution	
p-Aminobenzoic acid	0.02	
Nicotinamide	0.034	
Calcium Panotheate	0.02	
Thiamine	0.03	
Pyrodoxin.$2H_2O$	0.01	

The base media and each component were prepared with sterile artificial seawater.

Table 4.3 Carbon sources used in the preparation of liquid and solid media.

	Carbon source	Final concentration in media (L^{-1})
Liquid media		
1. PHY1	Ethanol	2mM
2. PHY2	Methanol	4mM
3. PHY3	Mandelic acid	0.5mM
4. PHY4	Creatine	1mM
Solid media (1.5% agar)		
1. PHY1	Ethanol	2mM
2. PHY2	Methanol	4mM

4.2.4 Isolation, growth conditions and screen of growth using flow Cytometry

The isolation approach was carried out based on the high-throughput method for culturing microorganisms (HTC) (Connon and Giovannoni 2002) (Figure 4.1), with some modifications. Sponge bacteria samples were serially diluted into 2ml of prepared media in a 96-well polystyrene deep well microtiter plate (Eppendorf, Australia) to give final cell concentrations of 1, 10 or 100 cells ml^{-1}. Triplicates of each concentration were made and controls were made using uninoculated media. The same cell dilution

method was used when inoculating cells onto the solid plates. Both liquid and solid media were grown at 20°C both aerobically and anaerobically.

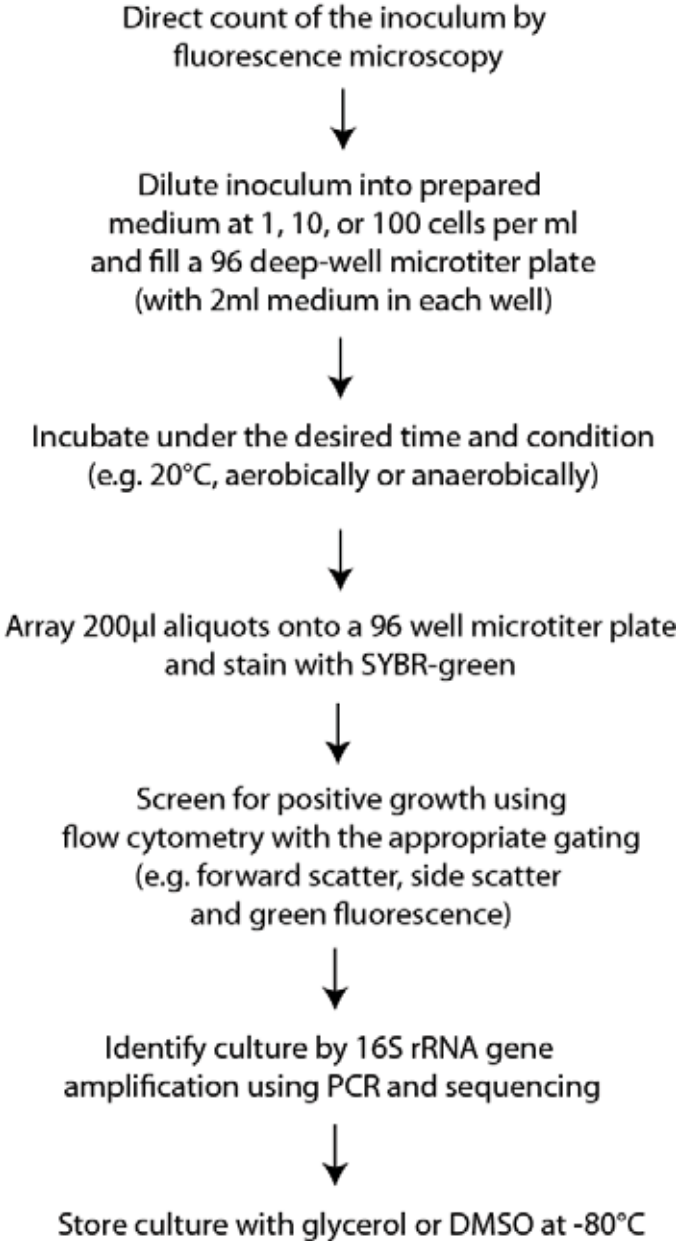


Figure 4.1 Flow chart of the HTC method developed based on Connon and Giovannoni (2002) with modifications.

After incubation for 4 and 12 weeks, 200ul of each culture were directly stained with 1 unit per ml of SYBR-green (Invitrogen, Australia) for 30 minutes in a 96 well plate (Sarstedt, Australia). Inoculum cell populations were counted by flow cytometery (Beckman Coulter Quanta SC MPL, Australia). The size of passing particles was determined using the forward and side scatter based on the scattered light intensity.

Cells were identified based on the fluorescence intensity of the correct size particle, and were turned into digitalised data. Based on the three gates: forward scatter, side scatter, and green fluorescence, the flow cytometer program was able to determine the cell population at the detection limit of 1×10^4 cells mL^{-1}. Appropriate gates were determined using control samples containing the uninoculated media that were stained with SYBR-green. The cell population was scored as positive if the cell concentration was greater than 1×10^5 cells mL^{-1}.

4.2.5 16S rRNA gene amplification

16S rRNA gene amplification of positive samples was carried out using PCR. The amplification method used was based on the colony PCR method, with an extract boiling step. Twenty microlitres of each positive cell sample were denatured at 95 ^oC for five minutes to promote cell lysis. Samples were stored on ice until PCR amplification. The universal 16S rRNA primers 27F and 1492R (Lane 1991), were used to amplify the 16S rRNA gene. The PCR reaction mixture contained 12.5 µL EconoTaq PLUS 2X Master Mix (Lucigen Corporation, Middleton, WI, USA), 400 nM of each primer (Sigma-Aldrich, Castle Hill, NSW, Australia), 2 µL denatured cell sample as template DNA, and Milli-Q water in a final volume of 25 µL. Thermo cycling was carried out with a hot-start at 80 ^oC. Denatured cell samples were placed into the heating block, followed immediately by 30 cycles of denaturation at 94 ^oC (30 seconds), annealing at 54 ^oC (1 minute), and extension at 72 ^oC (2 minutes). A final extension was run at 72^oC for 3 minutes. The sensitivity of this PCR method was determined using *E. coli* cells, and established that PCR products could be amplified from as low as 50 starting cells. A *Phyllobacteriaceae* phylotype specific reverse primer 1178PHY(5'-CTCAATCTCGCGATCTCG-3') was also designed based on the FISH probe described above (Section 3.2.10). It was used in combination with the universal 27F primer for PCR amplification (product size: approx.1100bp). Thermocycling was carried out as described above. PCR products were identified on 1% agarose gel to ensure they were of the correct size (approx. 1450bp or 1100bp). The PCR products were purified with a DNA Clean & Concentrator™-5 kit (Zymo Research Corporation, Irvine, CA) and sequenced bidirectionally using the 27F and 1492R primers (Lane 1991) or 27F and 1178PHY primers by the Ramaciotti Centre at the University of New South Wales

(Sydney, Australia). Sequences were analysed with BLAST searches (http://www.ncbi.nlm.nih.gov/BLAST).

4.3 Results and Discussions

4.3.1 Phylogenetic analysis and reconstruction of the partial genome of the Phyllobacteriaceae phylotype

Detailed phylogenetic analysis carried out previously (Section 3.3.3.6) demonstrated that the OTU assigned to the *Phyllobacteriaceae* fell into a distinct clade that is deeply branched within this family. This clade is distinct from the nitrogen-fixing genera *Mesorhizobium* and nitrate-reducing genera *Nitratireductor*. The genus *Mesorhizobium* includes soil and rhizosphere bacteria of agronomic importance that form nitrogen-fixing symbioses with leguminous plants (Kaneko *et al.,* 2000). The symbiotic functions are carried out through a series of interactions controlled by the exchange of molecular signals between symbiotic bacteria and host plants (Kaneko *et al.,* 2000). The type species within this genus is *M. loti*, which is an important symbiont with several *Lotus* species in soil (Kaneko *et al.,* 2000). *Nitratireductor* is another closely related genus within the *Phyllobacteriaceae* family, named because of its ability to reduce nitrate. The type species *Nitratireductor aquibiodomus* was isolated from a marine denitrification system of the Monteral Biodome in Canada and shown to be capable of reducing nitrate to nitrite (Labbe *et al.,* 2004). Although the sponge *Phyllobacteriaceae* phylotype is distinct from the two neighboring genera, its position on the phylogenetic tree suggests that it could have functions associated with nitrogen metabolism.

After removing the potentially false-positive contigs from the *Phyllobacteriaceae* partial genome, the refined partial genome contained 35 contigs with lengths greater than 3000 nt. It resulted in a total length of 2,716,490 nt, with an average GC content of 59.9%. The refined partial genome was predicted to encode 2857 non-redundant proteins. To evaluate the completeness of this partial genome, 40 COGs of universal, single-copy genes were searched against the partial genome (Ciccarelli *et al.,* 2006). A total of 34 of the 40 single copy COGs were detected, indicating that the partial genome reached a completeness of 85%.

4.3.2 Genome annotation of the Phyllobacteriaceae phylotype partial genome and functional genomic comparison with Mesorhizobium loti MAFF303099

The non-redundant contigs of the partial genome were submitted to the online RAST server (Aziz *et al.,* 2008) and annotated using similarity matching against the subsystems-based annotation (SEED) database (Overbeek *et al.,* 2005). The subsystems were annotated based on biochemical pathways, fragments of pathways and clusters of genes that function together, or any group of genes considered to be functionally related. The result from the RAST functional analysis indicated that a total of 363 subsystems were annotated. As an overview, 14 and 15% of those predicted pathways were functionally associated with metabolism and biosynthesis of amino acid and derivatives and carbohydrates, respectively. Different components of amino acid catabolism and biosynthesis pathways were detected covering most pathways of amino acids (Table 4.4). Functions within carbohydrate metabolism pathways, including glycolysis, gluconeogenesis and the TCA cycle were all detected. Further, the functional analysis predicted that the sponge *Phyllobacteriaceae* phylotype is capable of utilizing a variety of monosaccharides, disaccharides and aminosugars, like chitin and N-acetylglucosamine (Table 4.4). Enzymes responsible for sugar alcohol utilization, including glycerol, glycerol-3-phosphate, erythritol and inositol, and pathways for phosphorus, sulfur and nitrogen assimilation, were also detected (Table 4.1). A nitrogen fixation pathway was not detected, but a step within the nitrate and nitrite assimilation pathway, respiratory nitrate reduction, was found (all subunits of the respiratory nitrate reductase were identified).

Table 4.4 Subsystems associated with metabolism and biosynthesis of carbohydrates, amino acids, nitrogen, sulphur and phosphorus identified within the sponge *Phyllobacteriaceae*-phylotype's partial genome.

Sub-category	Subsystems
Carbohydrates:	
Di- and oligosaccharides	Lactose utilization
One-carbon Metabolism	Serine-glyoxylate cycle
One-carbon Metabolism	One-carbon metabolism by tetrahydropterines
Organic acids	2-methylcitrate to 2-methylaconitate metabolism cluster
Organic acids	Glycerate metabolism
Fermentation	Butanol Biosynthesis
Fermentation	Acetyl-CoA fermentation to Butyrate
Sugar alcohols	Glycerol and Glycerol-3-phosphate Uptake and Utilization
Sugar alcohols	Erythritol utilization
Sugar alcohols	Inositol catabolism
Carbohydrates - no subcategory	VC0266
Monosaccharides	D-ribose utilization
Monosaccharides	Mannose Metabolism
Monosaccharides	D-galactonate catabolism
Monosaccharides	Deoxyribose and Deoxynucleoside Catabolism
Monosaccharides	D-gluconate and ketogluconates metabolism
Monosaccharides	Fructose utilization
Amino acids:	
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamine, Glutamate, Aspartate and Asparagine Biosynthesis
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamate dehydrogenases
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	Glutamine synthetases
Histidine Metabolism	Histidine Degradation
Histidine Metabolism	Histidine Biosynthesis
Arginine; urea cycle, polyamines	Putrescine utilization pathways
Arginine; urea cycle, polyamines	Polyamine Metabolism
Arginine; urea cycle, polyamines	Arginine Biosynthesis extended
Lysine, threonine, methionine, and cysteine	S-methylmethionine
Lysine, threonine, methionine, and cysteine	Methionine Biosynthesis
Lysine, threonine, methionine, and cysteine	Threonine and Homoserine Biosynthesis
Lysine, threonine, methionine, and cysteine	Lysine Biosynthesis DAP Pathway, GJO scratch
Lysine, threonine, methionine, and cysteine	Threonine degradation
Lysine, threonine, methionine, and cysteine	Lysine Biosynthesis DAP Pathway
Amino Acids and Derivatives - no subcategory	Creatine and Creatinine Degradation
Branched-chain amino acids	Branched-Chain Amino Acid Biosynthesis

Branched-chain amino acids	Valine degradation
Branched-chain amino acids	Leucine Biosynthesis
Aromatic amino acids and derivatives	Common Pathway For Synthesis of Aromatic Compounds (DAHP synthase to chorismate)
Aromatic amino acids and derivatives	Chorismate Synthesis
Aromatic amino acids and derivatives	Phenylalanine and Tyrosine Branches from Chorismate
Aromatic amino acids and derivatives	Aromatic amino acid degradation
Aromatic amino acids and derivatives	Aromatic amino acid interconversions with aryl acids
Aromatic amino acids and derivatives	Indole-pyruvate oxidoreductase complex
Aromatic amino acids and derivatives	Chorismate: Intermediate for synthesis of Tryptophan, PAPA antibiotics, PABA, 3-hydroxyanthranilate.
Aromatic amino acids and derivatives	Tryptophan synthesis
Proline and 4-hydroxyproline	Proline Synthesis
Proline and 4-hydroxyproline	A Hypothetical Protein Related to Proline Metabolism
Proline and 4-hydroxyproline	Proline, 4-hydroxyproline uptake and utilization
Alanine, serine, and glycine	Glycine Biosynthesis
Alanine, serine, and glycine	Alanine biosynthesis
Alanine, serine, and glycine	Glycine and Serine Utilization
Alanine, serine, and glycine	Serine Biosynthesis
Alanine, serine, and glycine	Glycine cleavage system
Other metabolism:	
Inorganic sulfur assimilation	Inorganic Sulfur Assimilation
Sulfur Metabolism - no subcategory	Thioredoxin-disulfide reductase
Sulfur Metabolism - no subcategory	Galactosylceramide and Sulfatide metabolism
Organic sulfur assimilation	Alkanesulfonate assimilation
Phosphorus Metabolism - no subcategory	High affinity phosphate transporter and control of PHO regulon
Phosphorus Metabolism - no subcategory	Phosphate metabolism
Nitrogen Metabolism - no subcategory	Ammonia assimilation
Nitrogen Metabolism - no subcategory	Nitrate and nitrite ammonification

Earlier phylogenetic analysis indicated that the sequenced *M. loti* MAFF303099 is phylogenetically more closely related to the sponge *Phyllobacteriaceae* phylotype than to other sequenced bacteria. Therefore the partial genome of the sponge *Phyllobacteriaceae* phylotype was compared with the completed genome of the *M. loti* MAFF303099, based on COG functional categories. The *Phyllobacteriaceae* phylotype showed an overrepresentation of the functional COG categories of amino acid transport and metabolism, replication, recombination and repair, energy production and

conversion and carbohydrate transport and metabolism, and an underrepresentation of those of transcription and signal transduction compared to *M. loti* (Figure 4.2a). At the individual COG level, specific functions associated with transcriptional regulators (LysR COG0583, AcrR, COG1309) and dehydrogenases (FabG, COG 1028) were underrepresented, whereas functions associated with ankyrin repeat (AR) protein (COG0666), type I restriction modification system (HsdM, COG 0286) and TRAP-type C4-dicarboxylate transport system (DctQ, COG1593) were overrepresented (Figure 4.2b). The analysis highlighted clear differences in functional categories between the sponge *Phyllobacteraceae* phylotype and *M. loti* respectively, particularly with respect to amino acid transport, replication and recombination.

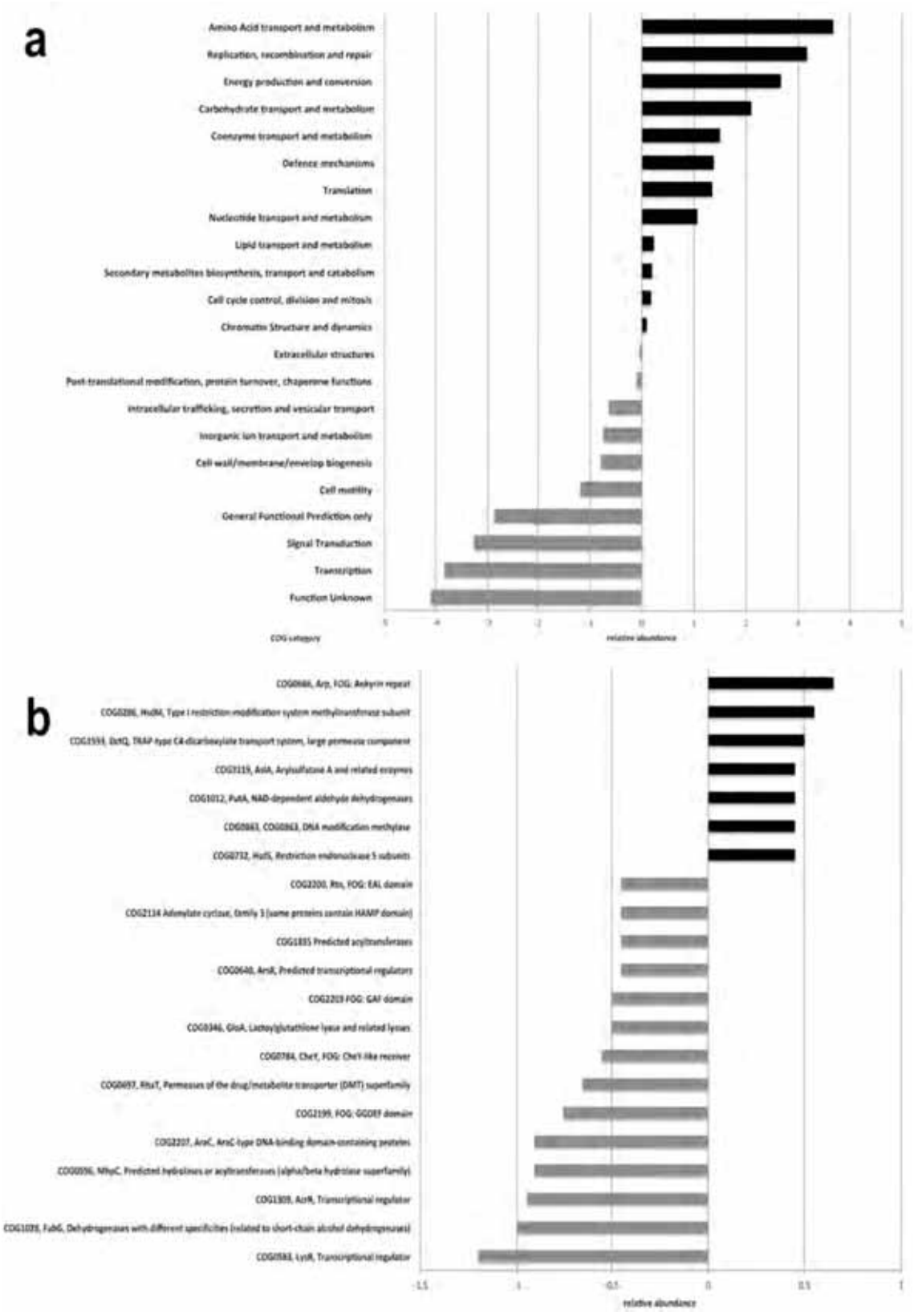


Figure 4.2 Functional genome comparison based on COG categories (a) and specific COGs (b) of the *Phyllobacteriaceae* phylotype partial genome and the genome of *M. loti* MAFF303099. The *X* axis displays the relative abundance with a significant cut-off value of -1 and 1. Black and grey bars represent over- and under-representation, respectively, compared to the sponge *Phyllobacteriaceae.*

4.3.3 Characteristic features of the sponge Phyllobacteriaceae phylotype

Further insights into the physiological potential of the sponge *Phyllobacteriaceae* phylotype were gained from the annotated partial genome. From the annotation data, the immediate observation of the partial genome was that functions specifically associated with motility and chemotaxis were not detected. In contrast, flagellar motility, biosynthesis and assembly were all detected in *M. loti*. Since the partial genome was 85% complete, it indicated that the sponge symbiont is likely to be non-motile. This observation correlates with the fact that the sponge symbiont was found to reside within the mesohyl, in the inner tissue (Chapter 3), where motility would be unexpected due to the porous structure and less crucial as the pumping activity of sponge could deliver nutrient to the symbionts.

4.3.3.1 Utilization of host products as nutrients

Aromatic compounds constitute the second most abundant class of organic substrates and environmental pollutants, and phenylacetate is a major intermediate in bacterial degradation of many aromatic compounds (Luengo *et al.,* 2001, Teufel *et al.,* 2010). Phenylacetate was also found in different types of sponges including *Haliclona cymaeformis* (Wattanadilok *et al.,* 2007) and *Oceanapia sagittaria* (Kijjoa *et al.,* 2007), and the Caribbean sponge, *Calyx* cf. *podatypa* (Adamczeski *et al.,* 1995). Specific genes associated with degradation of phenylacetate were identified from the partial genome of the sponge *Phyllobacteriaceae* phylotype (Figure 4.3b), which demonstrated the potential of using phenylacetate as a nutrient source.

Aromatic compounds in microorganisms are mostly converted to catechol or protocatechuate, where the aromatic ring is subsequently cleaved with the formation of succinate, pyruvate, and/or acetyl coenzymes A (acetyl-CoA), which can then enter the citric acid cycle and provide energy sources for cell growth. Key enzymes involved in protocatechuate degradation via the B-ketoadipate pathways were identified. Once B-ketoadipate is formed, it is converted to succinyl-CoA and acetyl-CoA. The enzyme C-succinyltransferase (pcaF) that completes the last step in the degradation pathway was also identified in the partial genome. Furthermore, enzymes of the degradation pathway of protocatechuate precursors like 4-hydroxymandelate (mandelate degradation) (Figure 4.3a) and salicylate (salicylic acid degradation, naphthalene degradation, data not shown

here) were also identified. These identified degradation pathways all suggested that *Phyllobacteriaceae* is adapted to use different aromatic compounds as carbon sources.

Functional genes encoding enzymatic activity in the shikimic acid pathway (Figure 4.3c) were identified within the partial genome of the sponge *Phyllobacteriaceae* phyolotype. Such a pathway is responsible for the biosynthesis of chorismic acid, which is the direct precursor of many essential metabolites including aromatic amino acids, folate, ubiquinones and other isoprenoid quinones (Knaggs 2003). The pathway is normally absent in metazoans, including humans, which reflects their dietary requirements for aromatic compounds (Zucko *et al.,* 2010). Therefore the detection of the shikimic acid pathway in the *Phyllobacteriace* indicates its potential function in biosynthesis of chorismic acid, which in turn could be used by its host as a metabolite. The examples addressed above suggest that various aromatic compounds or intermediates like phenylacetate, mandalte or shikimate maybe used by the sponge *Phyllobacteriaceae* phyloptype as metabolic substrates.

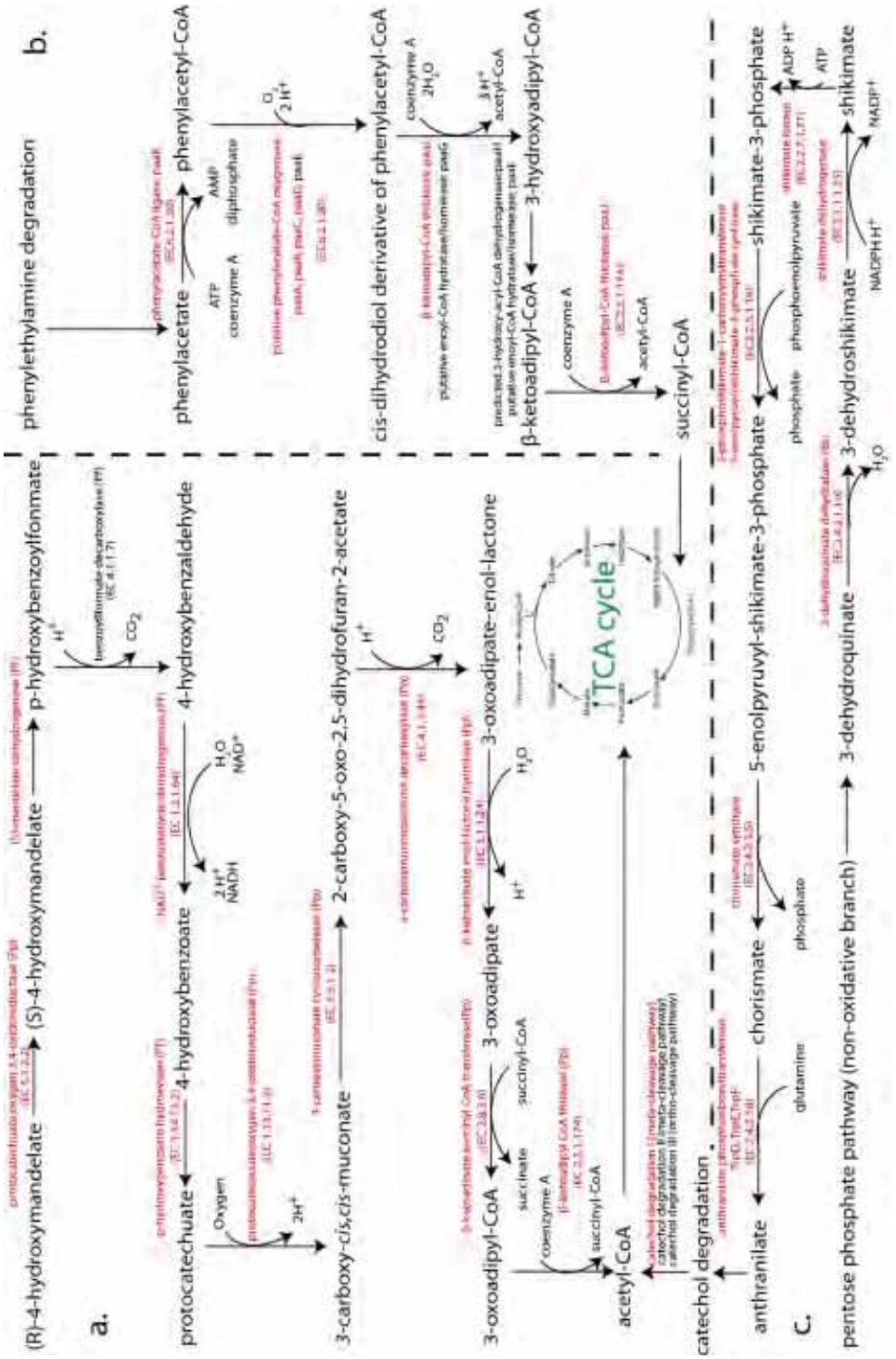


Figure 4.3 Schematic representation of the reactions and enzymes involved in microbial aromatic compound degradation pathways (catabolism). a. 4-hydroxymandelate degradation, b. phenylacetate degradation and c. shikimic acid pathway. The red text indicated the enzymes detected and pathways within the partial genome of the *Phyllobacteriaceae* phylotype.

Creatine phosphate (Roche 1954) and creatinine (Roche *et al.,* 1957) are present in the tissues of sponges, where creatine phosphate is an alternative energy storage molecule that can transfer energetic-phosphate bonds between cells (Wyss and Kaddurah-Daouk 2000). Reversible and non-enzymatic cyclization between creatinine and creatine has been observed in vertebrates (Wyss and Kaddurah-Daouk 2000), and could also be expected to occur in sponge tissue. Two alternative pathways for creatinine degradation were identified in the partial genome (Figure 4.4). In the first pathway, creatinine may be degraded to 1-methylhydantoin and ammonia by a single enzyme which has both cytosine deaminase and creatinine deaminase activity (Kim *et al.,* 1987, Esders and Lynn 1985). By this pathway, creatinine can be used as a nitrogen source, but not as a carbon or energy source. 1-methylhydantoin can be further degraded to N-carbamoylsarcosine and sarcosine by the enzymes 1-methylhydantoine aminohydrolase and N-carbamoylasarcosine amidohydrolase, which were both identified in the partial genome. The second pathway identified metabolises creatine directly to urea and sarcosine through the enzyme creatinase (creatine amidinohydrolase). The sarcosine formed in both degradation pathways can be further degraded to glycine by sarcosine oxidase and sarcosine dehydrogenase.

The cyclization of creatine and creatinine are both pH and temperature dependent (Wyss and Kaddurah-Daouk 2000). Creatine is favored at high pH and low temperatures, whereas creatinine is favoured at elevated temperatures and acidic solutions. It means both forms could be present in the sponge due to environmental changes (e.g. temperature, pH, acidity). The identification of both metabolic pathways indicates that the *Phyllobacteriaceae* may have developed effective scavenging strategies to utilise these carbon sources under different environmental conditions.

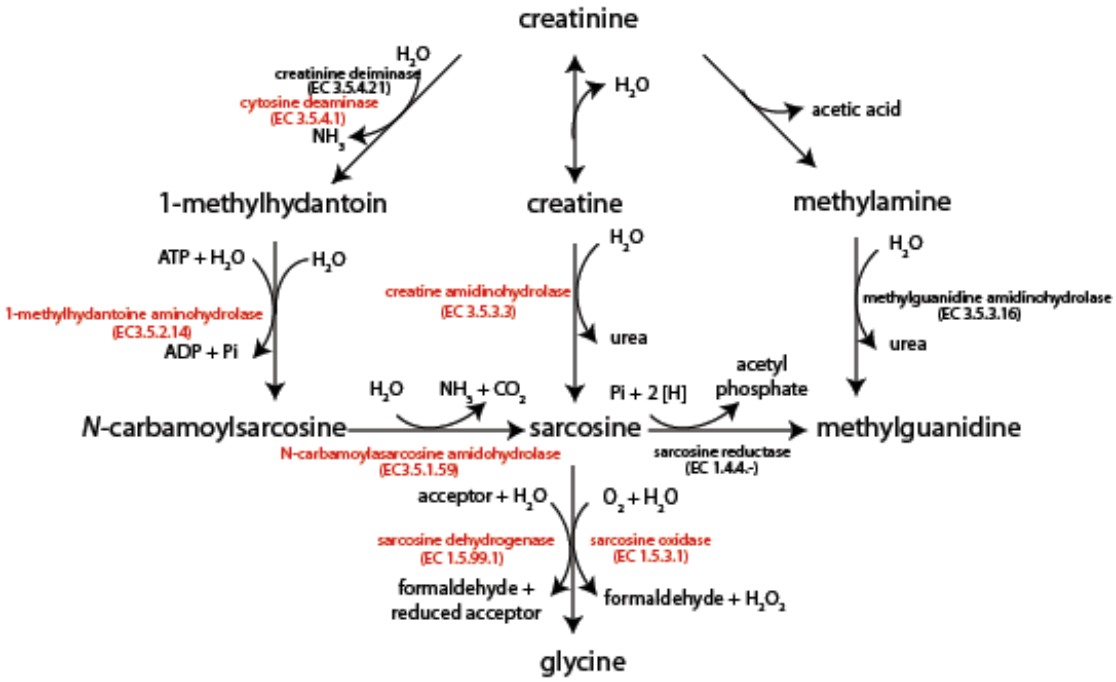


Figure 4.4 Schematic representation of the reactions and enzymes involved in microbial creatine and creatinine degradation pathways. The red text indicated the enzymes detected within the partial genome of *Phyllobacteriaceae* phylotype.

4.3.3.2 Respiration

Denitrification is a microbial process in which oxidised nitrogen compounds are used as alternative electron acceptors for respiration. It is the main biological process responsible for the return of fixed nitrogen to the atmosphere. All subunits of a membrane-associated nitrate reductase (EC1.7.99.4) were identified in the *Phyllobacteriaceae* phylotype partial genome (Table 4.3). The respiratory nitrate reductase usually has three subunits anchored at the cytoplasmic face of the membrane with its active site located in the cytoplasmic compartment (Richardson *et al.*, 2001). A further nitrate/nitrite transporter, which could transport nitrate and nitrite across the cytoplasmic membrane was also identified within the partial genome. No further nitrite reductase (EC1.7.1.4) or assimilatory nitrate reductase (EC 1.7.99.4) were identified within the partial genome, suggesting that nitrite is the end product of this anaerobic respiration. In contrast, assimilatory nitrate reductases and nitrite reductase were identified in the *M. loti* genome, indicating that *M. loti* is capable of using nitrate for assimilation through nitrate and nitrite ammonification. In this system, nitrate is transported across the cytoplasmic membrane by a nitrate ABC transporter, reduced by assimilatory nitrate reductase (EC 1.7.99.4), and further converted to ammonia by nitrite

reductase [NAD(P)H] (EC 1.7.1.4)(Simon 2002). The resulting substrate could then be used for glutamine and glutamate synthesis. Genes for nitrogen fixation identified in the *M. loti* were not identified within the partial genome of the *Phyllobacteriaceae* phylotype, indicating that this symbiont unlikely to be capable of fixing nitrogen.

Table 4.5 Individual nitrate reductase and nitrate transporter identified within sponge *Phyllobacteriaceae*'s partial genome.

Protein sequence name	Annotation
Nitrate reductase:	
Nitra_BBAY40_contig23533.gene_5	COG2181, NarI, Nitrate reductase gamma subunit
Nitra_BBAY40_contig23533.gene_6	COG2180, NarJ, Nitrate reductase delta subunit
Nitra_BBAY40_contig23533.gene_7	COG1140, NarY, Nitrate reductase beta subunit
Nitra_BBAY40_contig23533.gene_8	COG5013, NarG, Nitrate reductase alpha subunit
Nitra_BBAY41_contig16726.gene_2	COG2180, NarJ, Nitrate reductase delta subunit
Nitra_BBAY41_contig16726.gene_3	COG2180, NarJ, Nitrate reductase delta subunit
Nitra_BBAY42_contig27433.gene_1	COG5013, NarG, Nitrate reductase alpha subunit
Nitra_BBAY42_contig27433.gene_2	COG5013, NarG, Nitrate reductase alpha subunit
Nitra_BBAY42_contig27433.gene_3	COG1140, NarY, Nitrate reductase beta subunit
Nitra_BBAY42_contig27433.gene_4	COG2180, NarJ, Nitrate reductase delta subunit
Nitra_BBAY42_contig27433.gene_6	COG2181, NarI, Nitrate reductase gamma subunit
Nitrate transporter:	
Nitra_BBAY41_contig10277.gene_1	COG2223, NarK, Nitrate/nitrite transporter

Fermentation pathways such as for butanol biosynthesis and acetyl-CoA fermentation to butyrate pathways were identified within the partial genome. These observations highlight the fact that the sponge *Phyllobacteriaceae* phylotype is capable of carrying out not only anaerobic respiration through nitrate reduction, but also fermentation for energy production under anoxic conditions. The ubiquinol-cytochrome c reductase (iron-sulfur subunit) and cytochrome C oxidase were also identified from the partial genome, indicating a potential functional ability to carry out aerobic respiration. The potential presence of both aerobic and anaerobic respiratory pathways in the partial genome, indicated that *Phyllobacteriaceae* phylotype could be a facultative anaerobic bacterium. This notion is supported by the observation that the organism was physically localised within the oxygen limited inner spaces in the sponge (e.g. mesohyl; Chapter 3).

4.3.3.3 Defence mechanisms

Microbes have devised various defense strategies against viral predation and exposure to invading nucleic acids. In many bacteria and most archaea (Kunin *et al.,* 2007),

clusters of regularly interspaced short palindromic repeats (CRISPRs) based defense systems provide acquired resistance against invading viruses, possibly acting with an RNA interference-like mechanism to inhibit gene functions of invasive DNA elements (Barrangou *et al.,* 2007, Makarova *et al.,* 2006). CRISPR typically consist of several non-contiguous direct repeats separated by stretches of variable sequences called spacers, which mostly correspond to segments of captured viral sequences. The CRISPR are often adjacent to *cas* domains (CRISPR-associated), which encode a large and heterogenous family of proteins that typically carry functions of nucleases, helicases, polymerases and polynucleotide-binding proteins (Haft *et al.,* 2005). The CRISPRs and *cas* genes form the defense system and the *cas* proteins are thought to be involved in the propagation and functioning of CRISPRs. A number of *cas* genes were identified from the *Phyllobacteriaceae* phylotype partial genome, specifically five *cas* protein families (Csd1, Cas2, Cas3, Cas5 and Cas_CT1975) and 3 unique CRISPR repeat sequences, by searching the adjacent regions of the *cas* genes. This observation explains how sponge symbionts like the *Phyllobacteriaceaea* phylotype are likely to protect themselves and survive under high exposure of invading viral particles brought in by the water pumping activities of sponges.

In addition to the CRISPR/*cas* system, the partial genome also contained a relative high number of DNA modification and restriction systems including restriction endonuclease type I helicase (HsdR, COG0610 COG3096 and HsdS, COG0732) and restriction endonuclease type II methylase (COG1002). Both type I (EC: 3.1.21.3) and II (EC: 3.1.21.4) restriction endonucleases are components of prokaryotic DNA restriction-modification mechanisms, that protect the organism against invading foreign DNA based on different recognition sequence, subunit composition, cleavage position and cofactor requirements (Williams 2003). The presence of restriction endonucleases demonstrated the potential defense mechanism used by *Phyllobacteriaceae* phylotype against foreign DNA derived from outside the sponge's microbial community.

4.3.3.4 Repeat proteins in host-symbiont interactions

The host-symbiont interaction was also reflected through the detection of ankyrin repeat (AR) and tetratricopeptide repeat domain-encoding (TPR) proteins. The AR proteins were found to be overrepresented in the *Phyllobacteriaceae* phylotype partial genome

and this was also observed at the sponge metagenome level (Thomas *et al.*, 2010) and in the genome of a novel *Deltaproteobacteria* member (Chapter2). The repeat protein was not identified within the genome of *M.* loti. A total of 29 AR proteins and 19 TPR proteins were identified. Computational prediction of the subcellular localization of these proteins was carried out using the annotations tools, SignalP4.0 (Petersen *et al.,* 2011) and SecretomeP 2.0 for non-classical protein secretion (Bendtsen *et al.,* 2005). It was found that 8 AR proteins, representing 30% of the total AR proteins detected, had predicted signal peptides predicted. A further 9 AR proteins were predicted to be non-classically secreted proteins (30%). Among the TPR proteins, 6 were predicted to have signal peptides (20%), while none were demonstrated to be non-classically secreted proteins. The high occurrence of AR and TPR proteins having signal peptides for extracellular secretion, suggests that they could interact with other cells and their proteins. Additional detection of non-classically secreted proteins further demonstrated that these proteins could be secreted in multiple ways. AR repeat proteins have been reported to be secreted by the bacteria *Legionella pneumophila* and *Coxiella burnetii* and were found to facilitate host infection and bacterial intracellular survival by mimicking host proteins to modulate host cell functions (Al-Khodor *et al.,* 2008, Alli *et al.,* 2000, Habyarimana *et al.,* 2008, Pan *et al.,* 2008, Price *et al.,* 2010) AR repeat proteins were also present in sponge symbionts in *C. concentrica* (Thomas *et al.,* 2010, Yung *et al.,* 2009) and were recently shown to modulate eukaryotic host cell functions, specifically phagocytosis as seen in experiments with the amoebocyte, *Acanthamoeba castellanii* (Nguyen 2011). This may represent an important mechanism by which symbiotic bacteria avoid digestion and could explain how discrimination between food and symbiont bacteria occurs in the sponge.

4.3.4 Isolation of a sponge Phyllobacteriaceae phylotype

The above analysis indicated that the *Phyllobacteriaceae* phylotype is capable of utilizing a variety of aromatic compounds as nutrient sources and is potentially a facultative anaerobe. This important metabolic information provided clues that could be applied to procedures for isolation of this bacterium and a preliminary cultivation based on the HTC method was carried out in an attempt to achieve its isolation (Section 4.2.3-5).

After incubation at 20°C for 4 weeks, cell growth was checked based on cell population count using flow cytometry (Section 4.2.4). No growth was observed from the liquid media incubated either aerobically or anaerobically. After 12 weeks, growth was detected from a number of anaerobic liquid wells. Two wells, where methanol and ethanol as carbon source was used, showed sequences that are phylogenetically related to *Pseudovibrio ascidiaceicola* (Blastn 96%), belonging to the class of *Alphaproteobacteria* (Fukunaga *et al.,* 2006). The bacterium was isolated from ascidians (sea squirts), which indicated its potential relationship with the marine environment. In an attempt to obtain a pure colony of this organism, the liquid culture was inoculated onto solid media (PHY1, PHY2 with 1.5% agar), but no growth was observed. An interesting observation was made by sequencing DNA from a liquid culture well that contained creatinine as the carbon source. The dendrogram of the initial sequencing data of the 16S rRNA gene appeared noisy and led to the assumption that the culture could contain two or more bacterial species. To test this hypothesis, a *Phyllobacteriaceae* phylotype specific primer was used to target the 16S rRNA gene sequences of the mixed culture. The sequence was then blast searched against the representative 16S rRNA gene sequences of the dominant member of *Phyllobacteriaceae, Sphingomondales, Piscirickettsiaceae, Alpha, Beta, Gammaproteobacteria* from the sponge metagenome data. The result indicated that the isolate was only 89% identical with *Phyllobacteriaceae*, while it was most identical to *Sphingomondales* (Blastn 92%). In order to further separate this mixed culture, the cells were sub-cultured with fresh PHY4 liquid media, the HTC method repeated and incubated at the same conditions. However, no growing cell was detected from the incubated sub-cultures (up to a year).

Pure cultures of members of sponge associated *Phyllobacteriaceae* could not be cultivated using the growth media and conditions employed in this study. Despite the availability of metagenomic sequences that could improve the knowledge of the functional diversity of environmental microbial community, we are still far from measuring and interpreting the full extent of the functional potential of individual microorganisms. The functional potential of an organism is often predicted based on annotation of genes, within a partially functional pathway. Considering that many of the genes stored in the databases have unknown functions or are incorrectly annotated, it is

likely that metagenomes alone cannot offer sufficient information to cultivate a given organism.

The enrichment media and laboratory conditions used could also result in the break down of abiotic interactions which microbes experience in their natural environment. This disruption is often reflected through the use of incorrect concentrations of bio-geochemical factors, including different nutrients in the enrichment media. It is anticipated that symbiotic microorganisms also have developed a dependency on its host organism. This was evident from a previous cultivation study in *R. odorabile,* where the addition of sponge extract promoted growth of several morphotypes not previously or otherwise observed (Webster *et al.,* 2001b).

Most microorganisms in the environment live as part of a community and perform cellular functions in concert with each other. Their interdependencies are reflected in their ability to exchange metabolites, by exchanging dedicated signaling molecules or by competing for limited resources (Brown and Buckling 2008, Nadell *et al.,* 2009, West *et al.,* 2007). For example, using the community culture approach, coupled with dialysis membrane reactors, thermophilic anaerobic glutamate-degrading consortia from anaerobic sludge were successfully enriched (Plugge and Stams 2002). Batch reactors operating with anaerobic-aerobic cycling conditions also allowed the enrichment of mixed microbial sludge communities (Crocetti *et al.,* 2002). Thus, the growth interdependency of environmental microorganisms could help explain why cell growth was observed from a mixed cell culture in this study, which when further diluted and again incubated did not allow for growth to occur.

Culturing remains an important step in the process of understanding the biology and ecology of sponge-associated microbial species. Despite the inability to culture the organism of interest in this study, the data provided some interesting observations. The growth of the mixed culture indicated that the cultivation approach used could promote growth of mixed species. The outcome of these experiments may suggest thatsponge symbionts have developed complex growth interdependence with each other and with their host. This would be an important consideration for the design of future cultivation approaches

4.4 Conclusions

The results presented in this chapter have provided new insights into the specific functions of a sponge *Phyllobacteriaceae* phylotype. Functional signatures associated with survival using host-derived compounds, respiration, defense and host-symbiont interactions have offered important insights into the potential mechanisms this sponge symbiont uses to maintain its symbiotic relationship with its host and to deal with the changing environment. Information on the specific metabolic interdependency between the bacteria and sponge allowed the design of specific media that were used in an attempt to cultivate this bacterium.

The cultivation approach demonstrated the potential use of sequencing information for the development of selective isolation strategies targeting specific microorganisms. It also offered access to those bacteria once considered once 'unculturable'. The combination of this approach and the development of novel culture-independent methods such as metaproteomics would help to design approximate conditions where a significant fraction of microbes may be amenable to culturing, thereby adding more culturable representatives to the major bacterial lineages studied thus far.

Chapter Five

Phylogenetic, functional and protein expression changes as a function of the microbial community in the sponge *Rhopaloeides odorabile* during elevated temperatures

5.1 Introduction

Many environmental factors including climate, anthropogenic pollution, introduced species and nutrient enrichment have all been linked to stress and disease in marine life (Lafferty *et al.,* 2004). Increases in sea surface temperatures and ocean acidification associated with global climate change are predicted to have a significant impact on marine microorganisms and their invertebrate hosts (Hughes *et al.,* 2003, Webster and Hill 2007). It has been demonstrated that elevated seawater temperature can alter the frequency and severity of disease outbreaks in corals by increasing the prevalence and virulence of opportunistic pathogens, facilitating invasion or reducing host resistance and resilience by altering the microbial diversity, function and community dynamics (Sutherland *et al.,* 2004, Vega Thurber *et al.,* 2009).

Shifts in sponge-associated microbial communities in response to stress factors such as Cu^{2+}, temperature and disease correspond to declines in host health and often involve the loss of microbial symbionts and establishment of foreign microbes, including putative pathogens (Lopez-Legentil *et al.,* 2010, Webster *et al.,* 2001a, Webster *et al.,* 2008a, Webster *et al.,* 2008b). The sensitivity of sponge 'holobionts' (sponge host and their diverse microorganisms) to changing conditions makes them ideal indicators for monitoring of environmental change.

Sponges are an important component of marine ecosystems; they are highly diverse, abundant and play critical functional roles such as reef consolidation and bioerosion, provision of habitat for other invertebrates and high filtration efficiencies that couple the benthic and pelagic environments (reviewed in (Bell 2008)). Diverse microbial communities have been identified in sponges, including many lineages, which are

known to be sponge specific (Simister *et al.,* 2011, Taylor *et al.,* 2007, Webster and Blackall 2008, Zhu *et al.,* 2008). A range of beneficial functions has been reported to derive from sponge-associated microorganisms. They include nutrient provision and cycling (Arillo *et al.,* 1993, Wilkinson 1979b, Wilkinson and Garrone 1980), transportation and elimination of waste products (Hoffmann *et al.,* 2005, Kahlert and Neumann 1997), chemical defenses (Hentschel *et al.,* 2001, Schmidt *et al.,* 2000, Unson *et al.,* 1994) and contribution to mechanical structure (Wilkinson 1978a, Wilkinson 1978b). More recently, functions in the nitrogen cycle have been demonstrated from sponges. The processes of nitrification/denitrification and anammox could be assigned to the sponges *Geodia barretti* (Hoffmann *et al.,* 2009), *Dysidea avara* and *Chondrosia reniformis* (Schlappy *et al.,* 2010). Metagenomic analysis of the sponge *Cymbastela concentrica* also revealed high abundance of eukaryotic-like proteins such as the tetratrico peptide repeats (TPR) and ankyrin repeats (AR) proteins from the dominant symbionts (Thomas *et al.,* 2010)(Chapter 2 and 4). These proteins were proposed to play important roles in host-symbiont interactions by potentially mimicking host proteins to modulate host cell functions in sponges (Nguyen 2011).

Large scale mortality events in sponges have occurred during periods of unseasonably high seawater temperatures and after disease outbreaks (reviewed in (Cebrian *et al.,* 2011, Maldonado *et al.,* 2010, Webster 2007). These die-offs affect not only the survival of localised sponge populations but also the fate of many of the associated marine invertebrates, thus potentially impacting the ecology of entire reef ecosystems (Webster 2007). *In vitro* experiments based on the profiling of the 16S rRNA gene sequences from the microbial community in the Great Barrier Reef (GBR) sponge *Rhopaloeides odorabile* demonstrated that elevated seawater temperature can cause a shift in the community, such that native microbial populations were postulated to be replaced by opportunistic microorganisms, some of which were identified from the decaying tissue of diseased corals (Webster *et al.,* 2008a). It is unclear, however if disease-like symptoms such as necrosis and tissue disintegration are directly caused by external pathogens or by changes in the indigenous symbiont community. To address this issue, high-resolution metagenomic and metaproteomic microbial analysess were undertaken to investigate both the functional and phylogenetic composition of the sponge associated microbial community during thermal stress.

5.2 Materials and Methods

5.2.1 Sampling and experimental setup

A single *R. odorabile* sponge was collected from Davies Reef on the Great Barrier Reef, Australia (18°82'S, 147°65'E) in July 2009 by SCUBA diving. The donor sponge was cut into 48 clones, each approximately 15cm^3 in size and weighing approximately 30 g. These clones were secured in plastic racks and left on the reef to heal for 16 weeks. Clones were then transported to indoor aquaria at the Australian Institute of Marine Science, Townsville, Australia. The water temperature of the aquaria was set to 27 ± 0.5^{o}C to correspond to the ambient water temperature at Davies Reef and clones were acclimatised for 7d. The 48 clones were separated into six tanks each holding eight clones. Each 60 L tank was supplied with 5 µm filtered seawater at a flow rate of 600 ml per minute. Each tank was illuminated under a diel cycle (12:12 h) at 80 µmol quanta m^{-2} s^{-1} using 3ft compact fluorescent tubes tied to overhead lights. Three additional individuals of *R. odorabile* were collected from the same site of the GBR and directly processed for microbial cell enrichment.

5.2.2 Temperature exposure

The thermal stress experiment was commenced at 8 am on 18th November 2009 (day 0). The 6 tanks were randomly assigned to two temperature treatments (T27 and T32) with 3 replicate tanks (a, b and c) in each treatment. Tanks in the T27 treatment were maintained at 27°C throughout the experiment and seawater in the T32 treatment tanks was gradually heated (0.2°C/hr) to the final treatment temperature of 32°C. Temperatures were maintained in each of the treatments until the end of the experiment. Sampling was conducted at the time points listed in Table 5.1. One clone was randomly selected from each tank, photographed and immediately processed for microbial cell enrichment.

Table 5.1 Sample collections in the temperature-shifting experiment.

Time points	Temperatures	
	27°C	32°C
Day 0 (18th Nov 2009)	WT-a	
	WT-b	
	WT-c	
Day 1 (20th Nov 2009)	T27-Day1-a	T32-Day1-a
	T27-Day1-b	T32-Day1-b
	T27-Day1-c	T32-Day1-c
Day 3 (22nd Nov 2009)	T27-Day3-a	T32-Day3-a
	T27-Day3-b	T32-Day3-b
	T27-Day3-c	T32-Day3-c
Day 4 (23rd Nov 2009)	T27-Day4-a	T32-Day4-a
	T27-Day4-b	T32-Day4-b
	T27-Day4-c	T32-Day4-c[1]

[1] WT: Wild type

5.2.3 Microbial cell enrichment and extraction

Following collection, the sponge tissue was directly transferred into calcium- and magnesium-free seawater (CMFSW; 25 g NaCl, 0.8 g KCl, 1 g Na_2SO_4 , 0.04 g $NaHCO_3$ per 1 L) containing Protease Inhibitor Cocktail VI (PIC; 2 µl per ml; A.G. Scientific, San Diego, US) to preserve the cellular protein content and kept on ice or at 4^oC during all further processing. Surface barnacles and other macro-epibionts were physically removed from the sponge using sterile scalpels and forceps. Specimens were washed twice for five minutes at 200 rpm agitation in CMFSW to remove loosely attached bacteria or particles. The washed sponge was cut into ~1cm^3 cubes and homogenised for 10-15 seconds (discontinuously to avoid foam formation), in 150 ml of fresh CMFSW with PIC (2 µl per ml). Samples were incubated on ice for 30 minutes with agitation at 150 rpm and filtered through a 125 µm metal sieve into a sterile centrifuge tube. The filtrate was centrifuged for 15 minutes at 100 x g at 4^oC to remove remaining sponge cells. The collected supernatant was centrifuged twice for 15 minutes at 200 X g at 4^oC to remove unicellular eukaryotic organisms (e.g diatoms) from the sample. The supernatant was vacuum filtered through 11 and 3 µm and filters (Millipore™, Billerica, MA). The final filtrates were centrifuged for 20 minutes at 16,000 X g for 20 minutes at 4^oC and washed twice with CMFSW with PIC. Bacterial cell pellets were kept at -80^oC. DNA was extracted from the cell pellets according to the

procedure described in Thomas *et al* (2010). Protein extraction and preparation was conducted according to the methods described in Chapter 3 (Section 3.2.3)

5.2.4 Terminal Restriction fragment length polymorphism (T-RFLP) analysis

Bacterial 16S rRNA genes were amplified using the primers 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 1492R (5'-TACGGYTACCTTGTTAYGACTT-3'). The 27F primer was fluorescently labelled at the 5' end with 6-carboxyfluorescein (6-FAM). The PCR reaction contained 11 µL EconoTaq PLUS 2X Master Mix (Lucigen Corporation, Middleton, WI, USA), 400 nM of each primer (Sigma-Aldrich, Castle Hill, NSW, Australia), 1 µL 0.2 ng/µL template DNA, and 11 µL Milli-Q water in a final volume of 25 µL. DNA amplifications were performed using an initial denaturation at 94 ^{o}C (3 minutes), followed by 30 cycles of denaturation at 94 ^{o}C (30 seconds), annealing at 54 ^{o}C (1 minute), and extension at 72 ^{o}C (2 minutes). A final extension was run at 72^{o}C for 10 minutes. PCR products were verified on 1% agarose gel to ensure they were of the correct size (approx. 1450bp) and purified using a DNA Clean & Concentrator™- 5 kit (Zymo Research Corporation, Irvine, CA). PCR products were then digested with the restriction enzyme *Rsa*I (GT'AC; 0.25 U/µL; New England Biolabs, Hitchin, Hertfordshire, UK), 2 µL 10X NEB buffer 4, 100 ng DNA product and Milli-Q water used to make a final volume of 20 µL. Digestion reactions were as follows: incubation at 37°C for 4 hours and denaturation at 60°C for 20 minutes. Restriction digests were cleaned using the DNA Clean & Concentrator™- 5 kit (Zymo Research Corporation, Irvine, CA) and then analysed by the Ramaciotti Centre at the University of New South Wales (Sydney, Australia).

Raw data files containing peak information were tabulated in the Peak Scanner™ Software v1.0 (Life Technologies Corporation, Carlsbad, CA) with size standard 'GS1200LIZ'. T-RFLP data were processed and analyzed with the on-line T-REX software (Culman *et al.,* 2009). The data was subjected to quality control procedures: T-RF Alignment (clustering threshold =0.5), and Noise Filtering (peak area, standard deviation multiplier = 2). Data matrix with presence/absence information was generated for analysis in Primer 6 (PRIMER-E Ltd, Lutton, UK).

5.2.5 Metagenomic sequencing read process, assembly, and annotation

Shotgun libraries with fragment lengths around 400 bp were produced and sequenced from the extracted DNA using the Roche 454 Titanium platform (performed at the J. Craig Venter Institute, Rockville, USA). Dereplication of sequencing artefacts in the raw reads were conducted by cd-hit-454 as outlined by Niu *et al* (2010) with a 96% similarity cut-off and the shorter replicate read having at least 95% of the length of the longer replicate. Dereplicated reads of each sample were assembled using the GS *De Novo* Assembler (454 Life Science, Branford, CT, USA) with the default settings. Contigs, singletons and outliers were pooled and sequences less than 100 nt were removed.

To remove any contamination from eukaryotic cells and organelles (mitochondria and plastids) that may have survived the microbial cell fractionation, pooled contigs longer than 100 nt were taxonomically classified based on their sequence composition and similarity by PhymmBL (Brady and Salzberg 2009), according to a custom constructed reference genomic database. Beside the default PhymmBL reference dataset, which included all sequenced prokaryotic genomes in the NCBI RefSeq database (18 Mar 2011), this database also included the genomes of the sponge *Amphimedon queenslandica*, the round worm *Brugia malayi* and *Caenorhabditis briggsae*, the diatom *Blastocystis hominis*, *Thalassiosira pseudonana* and *Phaeodactylum tricornutum*, the hydrozoan *Hydra magnipapillata*, and all sequenced sponge mitochondria. Sequence fragments assigned to eukaryotes above or the ones contained in RefSeq were removed.

Open reading frames (ORFs) of coding genes were predicted from the filtered sequences with the MetaGeneAnnotator (Noguchi *et al.,* 2008). This resulted in a total of 2.8 million complete or partial genes. Predicted genes were translated to proteins and searched against the Clusters of Orthologous Group (COG) (Tatusov *et al.,* 2003) with rpsblast, and against the Pfam-A database (v24.0) (Finn *et al.,* 2010) with Hmmer 3 (http://hmmer.janelia.org/), both with an E-value cut-off of 10^{-10}. Proteins with multiple heterogeneous domains were counted separately, while repeats of the same domain in a protein were counted once. Genes were also annotated by the SEED/Subsystems using the online pipeline MG-RAST (v2) (Meyer *et al.,* 2008) with an E-value cut-off of 10^{-10}.

The shotgun sequencing is available through the Community Cyber infrastructure for Advanced Microbial Ecology Research and Analysis (CAMERA) website (http://camera.calit2.net/) under project accession 'CAM_PROJ_BotanyBay'.

5.2.6 Phylogenetic and diversity analyses.

16S rRNA gene sequences were extracted from the metagenome datasets and phylogeny reconstructed using the strategy outlined in Fan *et al* (2012). Briefly, metagenomic reads were searched using BlastN against an SSU-rRNA database combing the SILVA SSU-Ref and Greengenes databases (DeSantis *et al.,* 2006, Pruesse *et al.,* 2007) and hits with an E-value lower than 10^{-5} and with alignment length >49 bp were retrieved. The total sequence length of the SSU-rRNA gene sequences in each replicate sample were calculated and sequences were pooled and assembled with the GS *de novo* Assembler (454 Life Sciences, Branford, CT) using the "cDNA" option (99% overlapping identity). Contigs longer than 300 nt were aligned with the Mothur aligner (Schloss *et al.,* 2009) to the SILVA SSU seed alignment downloaded from the Mothur website (http://www.mothur.org/). Pairwise distances were calculated and sequences were clustered with the furthest-neighbor algorithm using a distance cut-off of 0.03. Representatives of the new clusters were aligned with the SINA aligner (Pruesse *et al.,* 2007) using Archaea/Bacteria/Eukarya-specific models and inserted into the SILVA SSURef tree using the specific pos_var_Archaea/Bacteria/Eukarya filters of the ARB package (Ludwig *et al.,* 2004). The pairwise phylogenetic distances of the representatives were calculated with an R script (Paradis *et al.,* 2004) and clustered by

Mothur (average linkage and distance cutoff of 0.03). Representatives of these clusters were then defined as the final operational taxonomic units (OTUs) and their abundance in each sample was calculated. Community diversity of each sample was assessed by plotting species richness curves based on OTU numbers and total phylogenetic distance, respectively, using (Caporaso *et al.,* 2010). Taxonomic assignment of the OTUs was manually conducted based on their locations in the SILVA SSURef tree. 16S rRNA gene profiles of the samples were clustered by the weighted-Unifrac algorithm (Lozupone *et al.,* 2011) with 1000 jackknife permutations.

5.2.7 Metaproteomic analysis by tandem mass spectrometry

Peptides for metaproteomic analysis with tandem mass spectrometry were prepared according to the methods outlined above (Section 3.2.4). Briefly, protein samples were re-suspended and resolved using one-dimensional sodium dodecyl sulfate polyacrylamide gel electrophoresis (1D SDS-PAGE). The whole gel lane was sliced into equal proportions and proteins within each slice digested to peptides with trypsin. Peptide digests were rehydrated in a buffer containing 1% formic acid and 0.05% heptafluorobutyric acid. Digests were first separated by nano-LC using an Ultimate 3000 HPLC and autosampler system (Dionex, Amsterdam, Netherlands). Samples (2.5 µl) were concentrated and desalted onto a micro C18 precolumn (0.5 mm X 2mm; Michrom Bioresources, Auburn, CA, USA) with H_2O:CH_3CN (98:2, 0.05% trifluoroacetic acid, TFA) at 15 µl / min. After a 4 min wash the precolumn was switched (Valco 10 port valve; Dionex) into line with a fritless nano column (75 µm X ~10 cm) containing C18 media (5 u, 200Å Magic; Michrom) manufactured according to Gatlin *et al.* (1998). Peptides were eluted using a linear gradient of H_2O:CH_3CN (98:2, 0.1% formic acid) to H_2O:CH_3CN (64:36, 0.1% formic acid) at 250 nl min^{-1} over 30 min. High voltage (1800 V) was applied to the low volume tee (Upchurch Scientific, Oak Harbor, WA, USA) and the column tip was positioned ~0.5 cm from the heated capillary (T = 250^{o}C) of an LTQ FT Ultra (Thermo Electron, Bremen, Germany) mass spectrometer. Positive ions were generated by electrospray and the LTQ FT Ultra was operated in data-dependent acquisition mode. A survey scan m/z 350–1750 was acquired in the FT ICR cell (resolution=100,000 at m/z 400, with an initial accumulation target value of 1,000,000 ions in the linear ion trap). Up to the six most

abundant ions (4000 counts) with charge states of +2, +3 or +4 were sequentially isolated and fragmented within the linear ion trap using collision-induced dissociation with an activation q = 0.25 and activation time of 30 ms at a target value of 30 000 ions. M/Z ratios selected for mass spectrometry–mass spectrometry (MS/MS) were dynamically excluded for 30 s. Peak lists were generated using Mascot Daemon/extract_msn (Matrix Science, Thermo, London, UK) using the default parameters (version 2.2; Matrix Science).

5.2.8 MS/MS data analysis and protein identification

All MS/MS spectra were analyzed using Mascot (version 2.3; Matrix Science, London, UK), Sequest (Thermo Fischer Scientific, San Jose, CA, USA; version 1.0.43.0) and X! Tandem (The GPM, thegpm.org; version 2007.01.01.1). All database searches were performed against the composite database containing all the sequences assembled from the metagenomic analysis of *R. odorabile.* Mascot and X! Tandem were searched with a fragment ion mass tolerance of 0.40 Da and a parent ion tolerance of 4.0 ppm. Sequest was searched with a fragment ion mass tolerance of 0.50 Da and a parent ion tolerance of 3.0 ppm. Oxidation of methionine and iodoacetamide derivative of cysteine were specified in Mascot, Sequest and X! Tandem as variable modifications. Oxidation of methionine, iodoacetamide derivative of cysteine and acrylamide adducts of cysteine were specified in Mascot as variable modifications.

To discriminate between false-positive and confident peptide matches, the software Scaffold (version Scaffold_2_00_05, Proteome Software Inc; Portland, OR) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted, if they could be established at greater than 95.0% probability as specified by the Peptide Prophet algorithm (Keller *et al.,* 2002). Protein identifications were accepted, if they could be established at greater than 99.0% probability and contained at least two identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (Nesvizhskii *et al.,* 2003). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. The false discovery rate, as estimated by searches against a decoy database, was below 1%.

Data deposition: MS/MS data have been deposited in the PRIDE database (20986–20998). Protein databases as well as lists of identified proteins and peptides in form of Excel files are available in the CD attached in this thesis. Proteins identified from the healthy, intermediate and necrotic microbial community samples of the sponge *R. odorabile* are stored in the Appendix Table A3, A4 and A5 respectively. The corresponding peptide identifications are stored in the Appendix Tables A6, A7 and A8 respectively.

5.2.9 Functional comparison of samples

Functional abundance matrices were generated by counting the number of genes (for metagenomic data) or proteins (for metaproteomic data) in each sample that had been assigned to a particular function (e.g. COG, Pfam and Subsystem). The abundance was weighted by the coverage of the genes (calculated during assembly) or proteins (number of peptides identified) assigned to this function. As the average genome sizes can potentially be quite different for metagenomic samples and can thus bias the functional profile comparison (Beszteri *et al.,* 2010), several strategies to predict average genome size (or genome copy) in metagenomic datasets have been proposed (Angly *et al.,* 2009, Frank and Sorensen 2011, Raes *et al.,* 2007). These approaches usually calculate the average coverage of conserved, single-copy genes and use this for normalisation. A similar approach was used here by selecting 38 COGs (namely COG0012, COG0016, COG0048, COG0049, COG0052, COG0080, COG0081, COG0087, COG0088, COG0090, COG0091, COG0092, COG0093, COG0094, COG0096, COG0097, COG0098, COG0099, COG0100, COG0102, COG0103, COG0124, COG0172, COG0184, COG0185, COG0186, COG0197, COG0200, COG0201, COG0202, COG0215, COG0256, COG0495, COG0522, COG0525, COG0533, COG0541, and COG0552) from the 40 universal, single-copy genes for normalisation (Ciccarelli *et al.,* 2006). These 38 COG entries were consistently abundant across all metagenomic samples and thus functional matrices of COG, Pfam and Subsystem annotation counts were normalised by the average abundance of the 38 COG entries in each sample. For proteomic data, weighted sample matrices were standardised to present the proportion of a function in all functions identified in a sample.

Statistical pair-wise comparisons between the healthy / intermediate, intermediate /

necrotic and healthy / necrotic samples were conducted using an R script modified from MetaStats (White *et al.,* 2009). The MetaStats script handles two matrices, the original input counts-table (Ctab) and the generated percentage-table (Ptab). MetaStats uses the Ptab to run a t-test and utilises the Ctab to handle those "spare" counts. A modified script was designed using the sample matrix without normalisation as the Ctab, and the normalised/ standardised matrix (see above) as the Ptab. To ensure only biologically meaningful functional differences are considered in the comparison, statistical significances were established if all of the following criteria were met: 1) the *p*-value was less than 0.05; 2) the group had more than 3 times higher counts of a function than the other group for metagenomic data, and 2 times higher for proteomic data; 3) for the group with higher abundance, the normalised count of the specific function needs to be greater than one per genome for metagenomic data or more than 0.5% abundance for metaproteomic data. Significant functions annotated by COG, Pfam, and Subsystem, respectively, were used for sample clustering using Primer 6. Heatmaps were generated by Cluster 3.0 (de Hoon *et al.,* 2004) and Java TreeView (Saldanha 2004).

5.3 Results and Discussions

5.3.1 Increased temperature causes sponge necrosis

Twenty-one sponge clones were collected at four time points from the aquaria maintained at 27℃ and 32 ℃ (see Table 5.1). Sponges were visually inspected for signs of necrosis and photographed after 4 days. Sponge clones from the 27℃ treatment remained intact without any visible sign of stress, whereas clones from the 32℃ treatment suffered severe necrosis (Figure 5.1). These results are consistent with previous experiments, in which *R. odorabile* clones began to eject small amounts of cellular material and exhibited surface necrosis after 24 h at 33℃ and major tissue necrosis, resulting in protrusion of skeletal fibres, after 3 days (Webster *et al.,* 2008a).

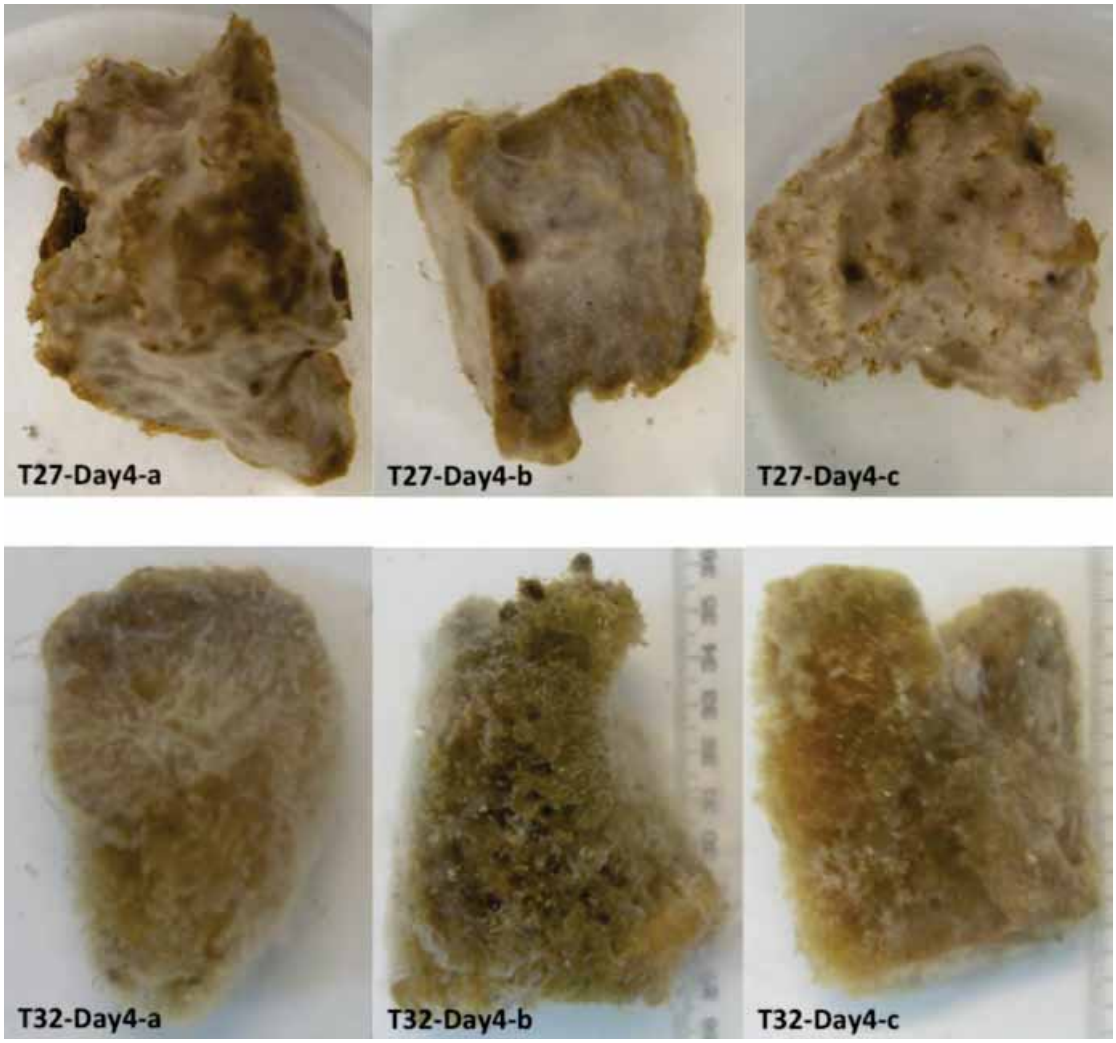


Figure 5.1 Morphological changes observed in sponge clones taken from Day 4.
T27: 27℃ incubation, T32: 32℃ incubation. Sponge tissue remained intact at 27℃ incubation after 4 day, but showed substantial tissue necrosis (de-colourisation and increased skeletal fibres) after incubation at 32℃.

5.3.2 Elevated water temperature causes phylogenetic shifts in the microbial community of Rhopaloeides odorabile

Large shifts in microbial community composition when water temperatures are elevated to 33℃ have previously been observed in *R. odorabile* using DGGE, 16S rRNA clone sequencing and bacterial cultivation (Webster *et al.,* 2008a). For example, a specific culturable *Alphaproteobacterium* symbiont was absent after only 24 hours at 33℃(Webster *et al.,* 2008a). Terminal restriction fragment length polymorphism (T-RFLP) analysis was carried out to examine the microbial communities in all 21 clones collected from the temperature experiment. The result of the analysis was used to investigate the potential linkage between the microbial community shifts with the decline of sponge health (Table 5.1). Sample clustering using Bray-Curtis Similarity

revealed three groups (Figure 5.2). The first cluster contained all clones from the 27℃ treatment, three clones from the 32℃ treatment after 1 day and 2 clones after 3 days. Bacterial communities of those clones shared 60% or higher similarity, with the only exception being sample T32-Day3-a (just below 60%). These clones were categorised as "healthy", as the tissue showed no visible signs of stress. The second group consisted of clones from the 32℃ treatment after 4 days (T32-Day4-b and T32-Day4-c). The similarity between these two clones was greater than 70%, but they were only 50% similar to the healthy group, hence were categorised as "intermediate". The last group consisted of clones from the 32℃ treatment after 3 (T32-Day3-c) and 4 (T32-Day4-a) days. They shared approximately 55% similarity with each other, but only 30% similarity with the other two groups. The deep branching of this group was consistent with the morphological observation of severe tissue loss (Figure 5.1), further supporting the notion that those clones were 'necrotic'. T-RFLP clustering indicated a shift in microbial community composition from healthy sponges to intermediate and eventually to a necrotic tissue state as a function of time at 32℃. A PermANOVA test based on those samples confirmed that both time and temperature are significant factors related to the community shift in *R. odorabile* (p (perm-time)= 0.017 and p (perm-temperature) =0.01, respectively). However the co-effect of both factors was not significant (p = 0.119), indicating that they may act independently.

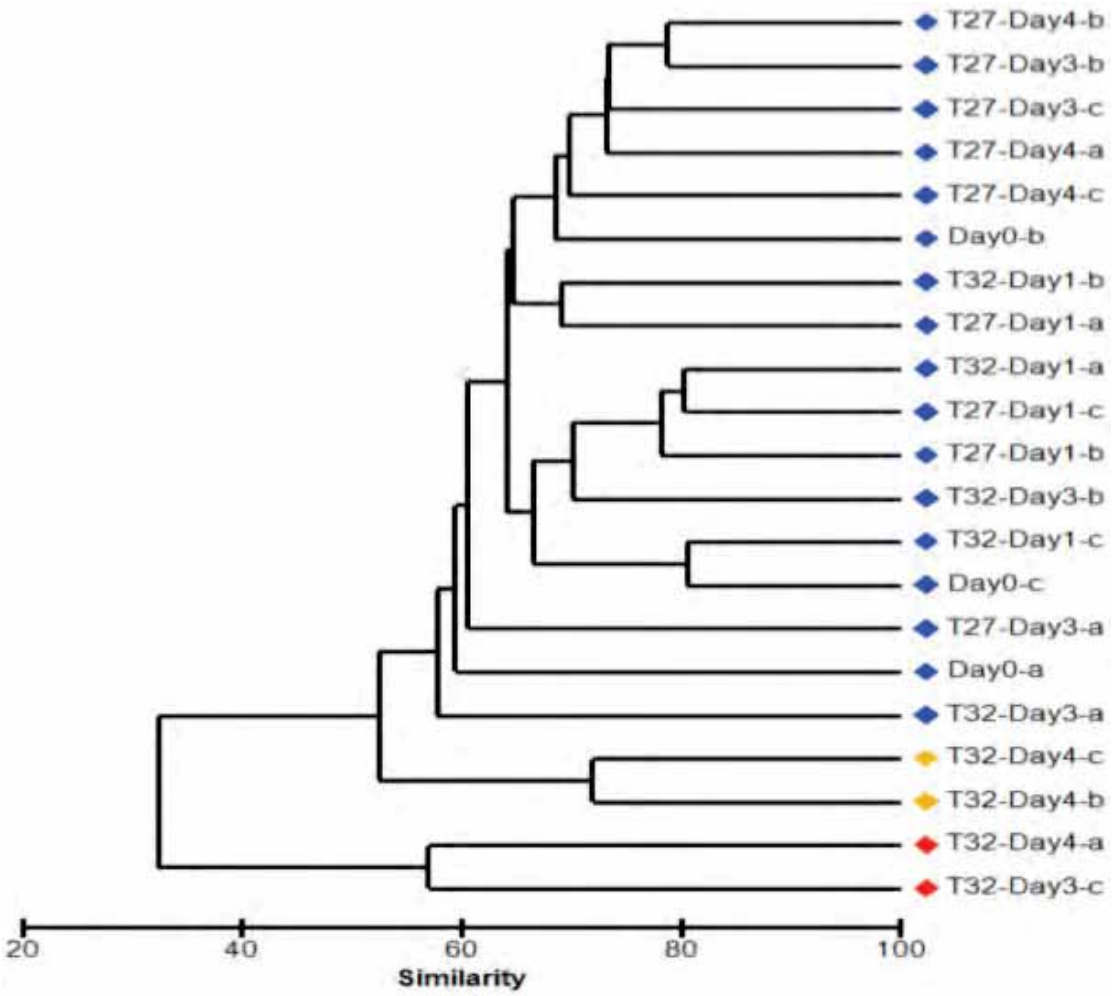


Figure 5.2 Clustering of bacterial communities based on T-RFLP using Bray Curtis similarity.

Clones T27-Day4-a, T27-Day4b, and T27-Day4c from the healthy group, T32-Day4-b and T32-Day4-c from the intermediate group, and T32-Day3-c and T32-Day4-a from the necrotic group, plus three sponges collected directly from the field (wild-type), were selected for metagenomic pyrosequencing to address the phylogenetic and functional composition of microbes within the heat stressed sponges. In total, 5,808,895 unique reads were generated from the 10 samples (Table 5.2). Archaeal and bacterial community profiles were reconstructed from the metagenomic datasets and generated 131 OTUs with a distance cutoff of 0.03. The relative abundance of the 50 most abundant OTUs in all samples was plotted (Figure 5.3).

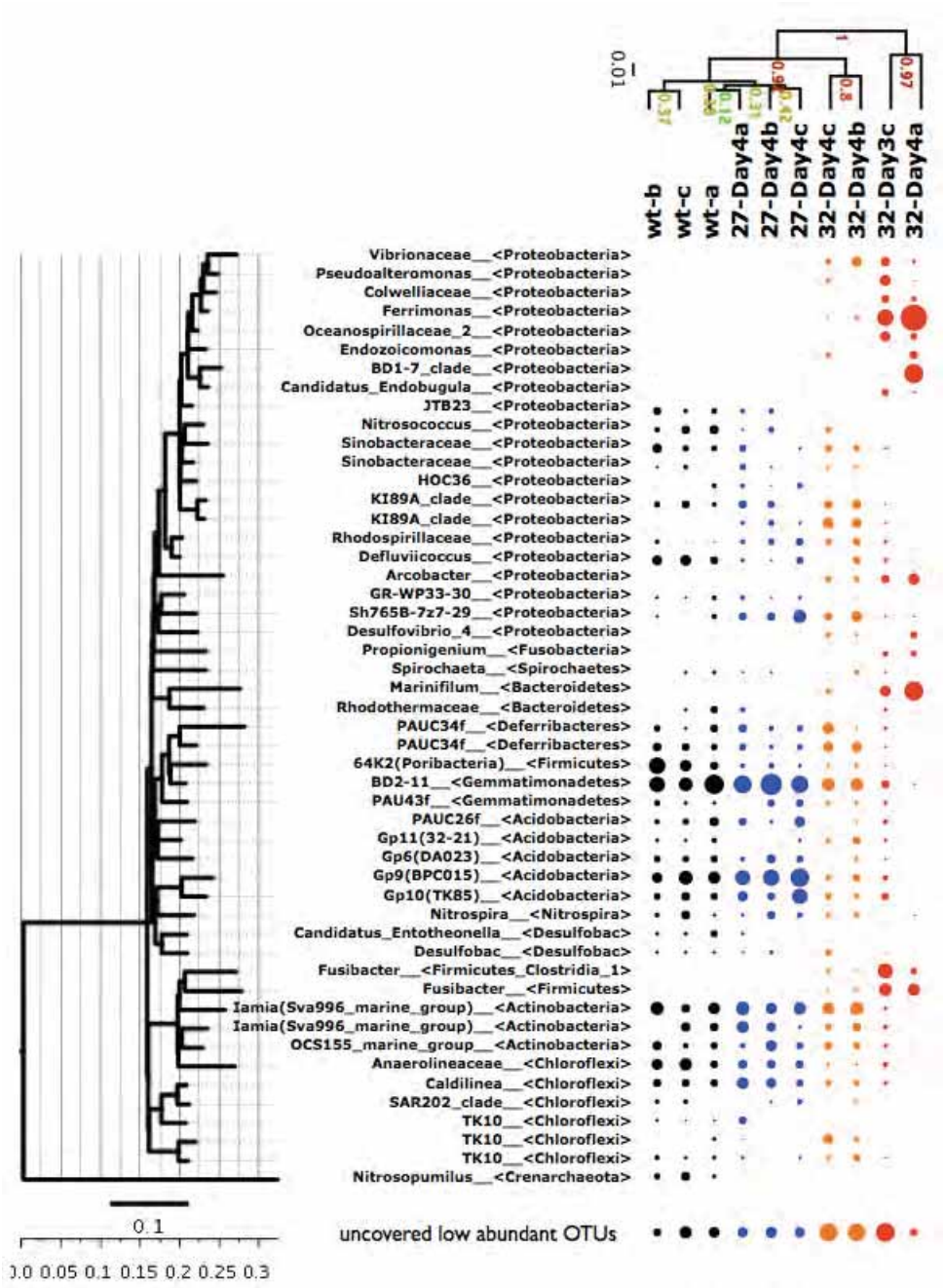


Figure 5.3 Shift in bacterial community through phylogenetic construction from metagenomic dataset.

The 50 most abundant OTUs are plotted and the size of each dot represents its abundance in the bacterial community. The total relative abundance of the low abundant OTUs (outside the top 50) is illustrated at the bottom. The phylogenetic tree of the OTUs is extracted from the entire SILVA SSU Ref tree containing those OTUs and is rooted to the archaeon *Nitrosopumilus*. The phyla of the OTUs are listed in the angular brackets. Samples are clustered according to their community structure using the Unifrac algorithm. Jack-knife confidence levels are given in coloured numbers in the branches. Black represent clones in wild-type group, blue is the healthy group, orange is the intermediate group, and red represents the necrotic group.

Table 5.2 Information for the metagenomic analysis of sponge *R. odorabile* clones.

Sample	WT-a	WT-b	WT-c	T27-Day4-a	T27-Day4-b	T27-Day4-c	T32-Day3-c	T32-Day4-a	T32-Day4-b	T32-Day4-c
Raw read	949133	583576	519285	742998	695928	627630	537954	403759	660167	453805
Average read size (nt)	402.0291645	377.3	401.6	423.4	428.2	343.5	316.1	311	391	368.0881729
Unique read	904146	505357	506761	716090	639263	606860	507711	399034	635356	388317
Aligned read	460513 (50.9%)	198840 (39.4%)	180893 (35.7%)	397269 (55.5%)	334502 (52.3%)	270894 (44.6%)	67199 (13.2%)	225756 (56.6%)	235761 (37.1%)	81113 (20.9%)
Contig > 500 nt	44408	17286	19306	35262	34102	16733	6420	3796	28674	10205
Average size of contigs > 500 (nt)	902	900	836	1002	894	1092	743	2441	784	687
N50 size of contigs > 500 (nt)	909	909	838	1018	895	1178	724	5055	776	664
Maximum size of contigs > 500 (nt)	52481	16414	16419	52143	18553	17003	6062	89781	5441	3749
Contig > 100 nt	71479	29504	32819	54685	55119	28502	11648	5262	46143	18175
Singleton /outlier > 100 nt	416488	284341	296611	306056	291996	317341	407146	153752	382667	282041
Non-eukaryotic contig and singleton/outlier	486090 (99.6%)	310735 (99%)	326136 (99%)	359921 (99.8%)	346460 (99.8%)	345197 (99.8%)	417634 (99.7%)	158279 (99.5%)	428032 (99.8%)	299667 (99.8%)
Unique predicted proteins	597098	367768	394503	450176	433069	401283	475263	186518	518234	356540
Total proteins	828348	457510	480812	645157	609660	515151	505473	254007	636935	396781
Protein annotated by COG ($E<10^{-10}$)	230428 (27.8%)	120728 (26.4%)	123659 (25.7%)	204875 (31.8%)	181301 (29.7%)	134597 (26.1%)	104924 (20.8%)	89516 (35.2%)	175661 (27.6%)	100811 (25.4%)
Protein annotated by Pfam ($E<10^{-10}$)	219205 (26.5%)	115740 (25.3%)	118378 (24.6%)	194148 (30.1%)	173218 (28.4%)	126218 (24.5%)	25965 (5.1%)	85497 (33.7%)	162726 (25.5%)	92386 (23.3%)
Protein annotated by SEED ($E<10^{-10}$)	338201 (40.8%)	180828 (39.5%)	187208 (38.9%)	290960 (45.1%)	266252 (43.7%)	199504 (38.7%)	171954 (34%)	129579 (51%)	255930 (40.2%)	151296 (38.1%)

WT: Wild-type.

The clustering of samples based on OTU abundance from reconstructed 16S rRNA genes generated a similar topology to the clustering of T-RFLPs. The healthy (T27-Day4-a, T27-Day4-b and T27-Day4-c), intermediate (T32-Day3-c and T32-Day4-a) and the necrotic group (T32-Day4-b and T32-Day4-c) were clearly separated from each other and supported by high Jackknife confidence values. The three wild-type samples and the healthy clones formed one clade with only minor differences in OTU abundances (e.g. *Defluviicoccus*, Sh765B-7z7-29, 64K2, and *Nitrosopumilus*). This suggests that *R. odorabile* sponges from the 27℃ treatment have a symbiotic bacterial community consistent with wild sponges (cf. (Webster *et al.*, 2008a)).

Sponges in the intermediate group were related to those of the healthy group, but showed shifts in the bacterial and archaeal community composition towards the necrotic sponges. Specifically, some intermediate samples showed a decreased abundance of certain sponge-associated bacteria (e.g. BD2-11, Gp9 (BPC015), Gp10 (TK85), and *Anaerolineaceae*), while new phylotypes (in particular some *proteobacteria*; Figure 3, orange dots) increased in abundance. This observation may represent an increase in low-abundance, previously undetected sponge-associated bacteria or an invasion and establishment of bacteria from the surrounding seawater. A dramatic community change was evident in the necrotic sponges (Figure 3, red dots) with a marked decrease in bacterial lineages associated with the wild and healthy sponges and an increase in the phylotypes that appeared in low numbers in the intermediate samples. These increased phylotypes include in clone T32-Day4-a, are bacteria belonging to the taxa *Vibrionaceae*, *Pseudoalteromonas*, *Colwelliaceae*, *Ferrimonas*, *Oceanospirillaceae*-2, *Endozoicomonas*, BD107 clade, *Arcobacter*, *Marinifillum*, and *Fusibacter*.

The observed changes in community structure were also reflected in species richness and the proportion of low-abundance community members (Figure 5.3, dots at the lower part for the panel and Figure 5.4). Generally, the intermediate group had the highest species richness, while the wild-type/healthy group had medium species richness. The necrotic clone 32T-Day4-a, having lost most of its symbiotic bacteria (Figure 5.3), had the lowest species richness.

Overall, the results supported a model where the bacterial and archaeal community of *R. odorabile* shifts during exposure to elevated temperatures. The diversity increases in the intermediate state as foreign or low-abundance bacteria begin to establish themselves with the community along with existing symbionts and decreases as sponge symbionts are outcompeted. At this stage specific groups of foreign / low abundant bacteria dominate and appear to establish a new community.

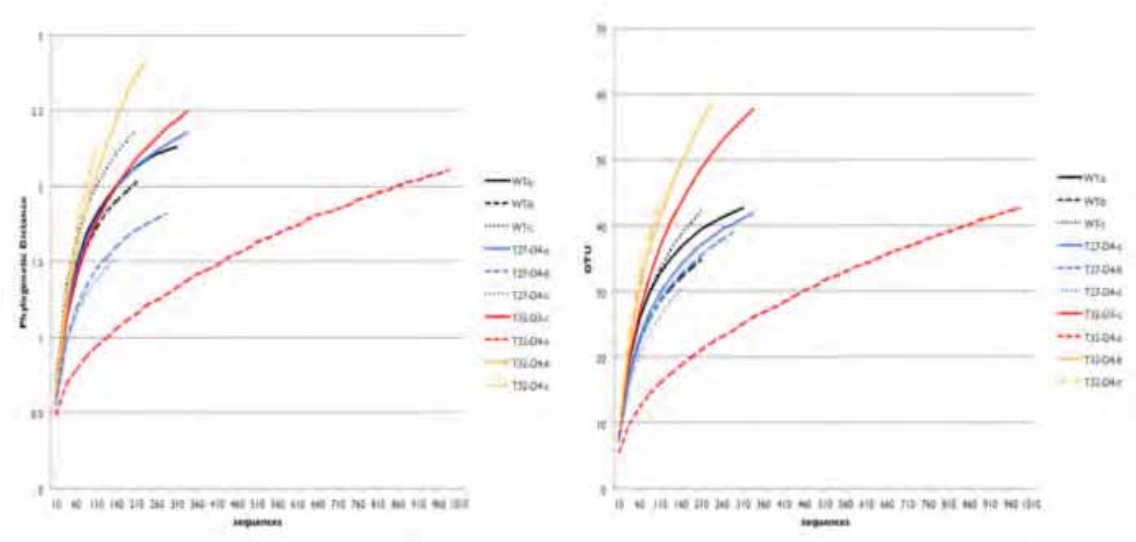


Figure 5.4 Rarefaction plot showing the dynamics of community diversity in different samples. Black, wild-type group; blue, healthy group, yellow, intermediate group and red, necrotic group. Average values of 1,000 replications of Jacknife subsampling were plotted.

5.3.3 Elevated water temperature induces the collapse of host-symbiont interactions

A breakdown of the symbiotic interactions at elevated temperature in *R. odorablie* was indicated from both the state of the sponge (tissue necrosis, Figure 5.1) and its symbiont community (Figure 5.2 and 5.3). Functional metagenome analysis was carried out to further explore this phenomenon using 7 samples (T27-Day4-b, T27-Day4-c, T32-Day3-c, T32-Day4-a, T32-Day4-b, T32-Day4-C), which represented the progression from healthy to necrotic tissue states of sponges. Unique reads in each sample were assembled into contigs resulting in a total of 3,491,775 contigs and unassembled singletons larger than 100 nt (Table 5.2). Of those, 3,478,151 passed the filtering procedure for eukaryotic contamination and a total of 5,329,834 genes could be predicted with 1,466,499, 1,313,482 and 2,171713 being annotated to COGs, Pfam and SEED Subsystem, respectively (Table 5.2). Pair-wise comparison between wild type (wt) / healthy, healthy / intermediate, intermediate / necrotic, and healthy / necrotic

were conducted followed by statistical analysis and filtering to identify those functions that contribute most to the difference between sample types (see Material and Methods). These discriminatory COG, Pfam and SEED functions were plotted in heatmaps for visual representations and used to cluster samples (Figure 5.5).

Consistent with phylogenetic analysis, the necrotic sponges clustered separately from all other samples (which together formed one clade). The functional profile of wild type and sponges from the 27℃ treatment were highly similar and only different in the functions of WD40 repeat (COG2319, PF00400) and a transposase (COG0675, PF01385), which were both more abundant in the wild type group (Figure 5.5). This observation highlights that cultivated *R. odorabile* are representative of wild-type sponges.

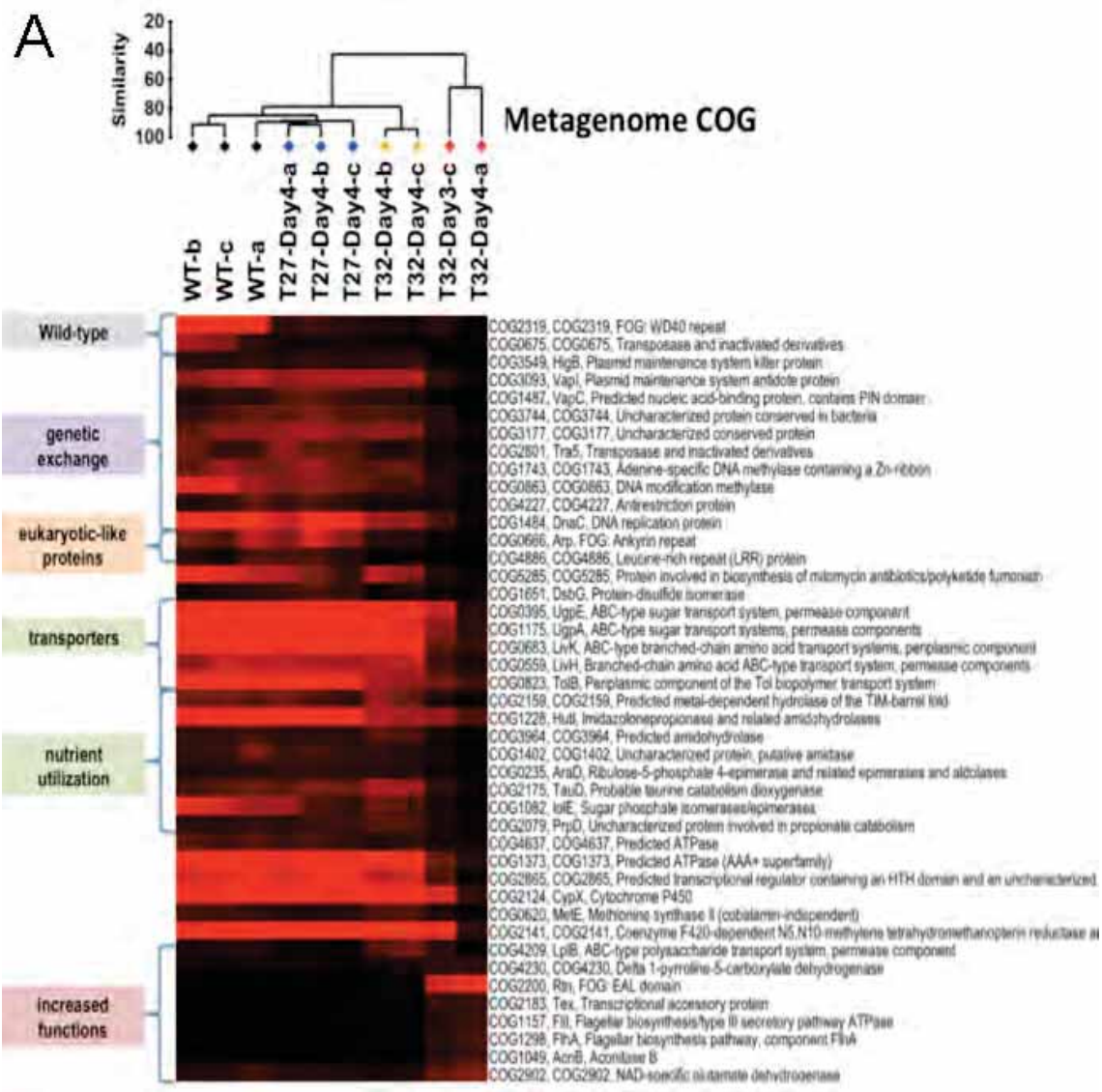
Functions previously described to be abundant in the microbial community of the sponge *C. concentrica* (Thomas *et al.*, 2010) were also predominant in the bacteria and archaea of wild type and healthy *R. odorabile* (Figure 5.5). The functions included mobile genetic elements (plasmids, transposases), restriction-modification systems and CRISPRs, that might facilitate a frequent and selective genetic exchange in the microbial community. Eukaryotic-like proteins (ELPs), which can potentially be used by symbionts to manipulate their host were also found to be abundant in the wild type and healthy *R. odorabile* (cf. (Thomas *et al.*, 2010)). Other functions that were abundant in healthy sponge communities included membrane transport, substrate utilisation, cell signalling and regulation and stress response mechanisms including antibiotic efflux pump and cell-cell adhesion. The latter are likely to be important for supporting and maintaining the sponge-microbe symbiosis.

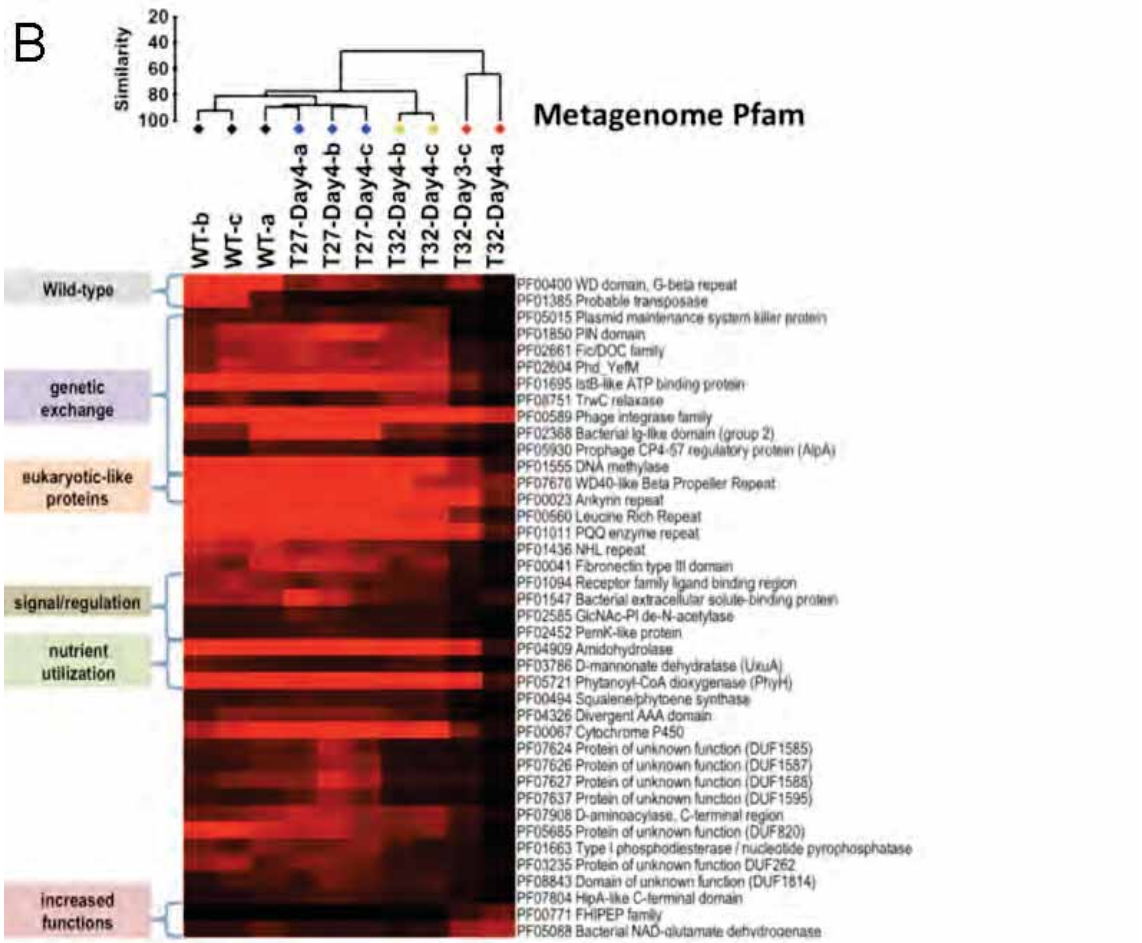
Many of these specific functions were also present in intermediate samples, although some in lower abundance, consistent with the community structure shift revealed by phylogenetic analysis (Figure 5.3).

Metagenomic analysis of the necrotic samples revealed a dramatic shift in the otherwise “characteristic” symbiont profile of the sponge, with the loss of all of the previously described functions. In contrast, necrotic sponges were rich in proteins with EAL

domains and proteins associated with flagellar biosynthesis (Figure 5.5). EAL domain proteins degrade c-di-GMP, an important intracellular signalling molecule that regulates sessile or biofilm lifestyle and virulence traits (Hengge 2009). The over-represented function in flagellar biosynthesis (COG1298, COG1157, COG1049, PF00771) in the necrotic sponge community supported the hypothesis that these bacteria are motile. Similar trends have been reported for stressed corals (Vega Thurber *et al.*, 2009). The necrotic sponge tissue would be a rich source of nutrients and colonising bacteria would benefit from the ability to rapidly reach the nutrient source associated with the decaying sponge cells.

Necrotic sponges had no other notably abundant microbial functions and no pathogenic functions such as virulence proteins were detected. Whilst all healthy sponges shared common microbial functions, few features were shared between the two necrotic individuals. It indicated that diverse microbes could invade or dominate a necrotic sponge. Sponges have a diverse array of natural products, which are widely believed to act as an antimicrobial defence response and it is likely that as a sponge becomes stressed, this defence system breaks down, allowing the establishment of opportunistic microbes, which are normally repelled by the sponge's chemical defence system. It is also possible that the symbionts are particularly sensitive to thermal stress and the loss of these important components of the holobiont leads to a rapid decline in host health.


A
Similarity
20
40
60
80
100
Metagenome COG
WT-b
WT-c
WT-a
T27-Day4-a
T27-Day4-b
T27-Day4-c
T32-Day4-b
T32-Day4-c
T32-Day3-c
T32-Day4-a
Wild-type
genetic exchange
eukaryotic-like proteins
transporters
nutrient utilization
increased functions
COG2319, COG2319, FOG: WD40 repeat
COG0675, COG0675, Transposase and inactivated derivatives
COG3549, HigB, Plasmid maintenance system killer protein
COG3093, VapI, Plasmid maintenance system antidote protein
COG1487, VapC, Predicted nucleic acid-binding protein, contains PIN domain
COG3744, COG3744, Uncharacterized protein conserved in bacteria
COG3177, COG3177, Uncharacterized conserved protein
COG2801, Tra5, Transposase and inactivated derivatives
COG1743, COG1743, Adenine-specific DNA methylase containing a Zn-ribbon
COG0863, COG0863, DNA modification methylase
COG4227, COG4227, Antirestriction protein
COG1484, DnaC, DNA replication protein
COG0666, Arp, FOG: Ankyrin repeat
COG4886, COG4886, Leucine-rich repeat (LRR) protein
COG5285, COG5285, Protein involved in biosynthesis of mitomycin antibiotics/polyketide fumonisin
COG1651, DsbG, Protein-disulfide isomerase
COG0395, UgpE, ABC-type sugar transport system, permease component
COG1175, UgpA, ABC-type sugar transport systems, permease components
COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component
COG0559, LivH, Branched-chain amino acid ABC-type transport system, permease components
COG0823, TolB, Periplasmic component of the Tol biopolymer transport system
COG2159, COG2159, Predicted metal-dependent hydrolase of the TIM-barrel fold
COG1228, HutI, Imidazolonepropionase and related amidohydrolases
COG3964, COG3964, Predicted amidohydrolase
COG1402, COG1402, Uncharacterized protein, putative amidase
COG0235, AraD, Ribulose-5-phosphate 4-epimerase and related epimerases and aldolases
COG2175, TauD, Probable taurine catabolism dioxygenase
COG1082, IolE, Sugar phosphate isomerases/epimerases
COG2079, PrpD, Uncharacterized protein involved in propionate catabolism
COG4637, COG4637, Predicted ATPase
COG1373, COG1373, Predicted ATPase (AAA+ superfamily)
COG2865, COG2865, Predicted transcriptional regulator containing an HTH domain and an uncharacterized
COG2124, CypX, Cytochrome P450
COG0620, MetE, Methionine synthase II (cobalamin-independent)
COG2141, COG2141, Coenzyme F420-dependent N5,N10-methylene tetrahydromethanopterin reductase a
COG4209, LplB, ABC-type polysaccharide transport system, permease component
COG4230, COG4230, Delta 1-pyrroline-5-carboxylate dehydrogenase
COG2200, Rtn, FOG: EAL domain
COG2183, Tex, Transcriptional accessory protein
COG1157, FliI, Flagellar biosynthesis/type III secretory pathway ATPase
COG1298, FlhA, Flagellar biosynthesis pathway, component FlhA
COG1049, AcnB, Aconitase B
COG2902, COG2902, NAD-specific glutamate dehydrogenase


B
Similarity
20
40
60
80
100
Metagenome Pfam
WT-b
WT-c
WT-a
T27-Day4-a
T27-Day4-b
T27-Day4-c
T32-Day4-b
T32-Day4-c
T32-Day3-c
T32-Day4-a
Wild-type
genetic exchange
eukaryotic-like proteins
signal/regulation
nutrient utilization
increased functions
PF00400 WD domain, G-beta repeat
PF01385 Probable transposase
PF05015 Plasmid maintenance system killer protein
PF01850 PIN domain
PF02661 Fic/DOC family
PF02604 Phd_YefM
PF01695 IstB-like ATP binding protein
PF08751 TrwC relaxase
PF00589 Phage integrase family
PF02368 Bacterial Ig-like domain (group 2)
PF05930 Prophage CP4-57 regulatory protein (AlpA)
PF01555 DNA methylase
PF07676 WD40-like Beta Propeller Repeat
PF00023 Ankyrin repeat
PF00560 Leucine Rich Repeat
PF01011 PQQ enzyme repeat
PF01436 NHL repeat
PF00041 Fibronectin type III domain
PF01094 Receptor family ligand binding region
PF01547 Bacterial extracellular solute-binding protein
PF02585 GlcNAc-PI de-N-acetylase
PF02452 PemK-like protein
PF04909 Amidohydrolase
PF03786 D-mannonate dehydratase (UxuA)
PF05721 Phytanoyl-CoA dioxygenase (PhyH)
PF00494 Squalene/phytoene synthase
PF04326 Divergent AAA domain
PF00067 Cytochrome P450
PF07624 Protein of unknown function (DUF1585)
PF07626 Protein of unknown function (DUF1587)
PF07627 Protein of unknown function (DUF1588)
PF07637 Protein of unknown function (DUF1595)
PF07908 D-aminoacylase, C-terminal region
PF05685 Protein of unknown function (DUF820)
PF01663 Type I phosphodiesterase / nucleotide pyrophosphatase
PF03235 Protein of unknown function DUF262
PF08843 Domain of unknown function (DUF1814)
PF07804 HipA-like C-terminal domain
PF00771 FHIPEP family
PF05088 Bacterial NAD-glutamate dehydrogenase

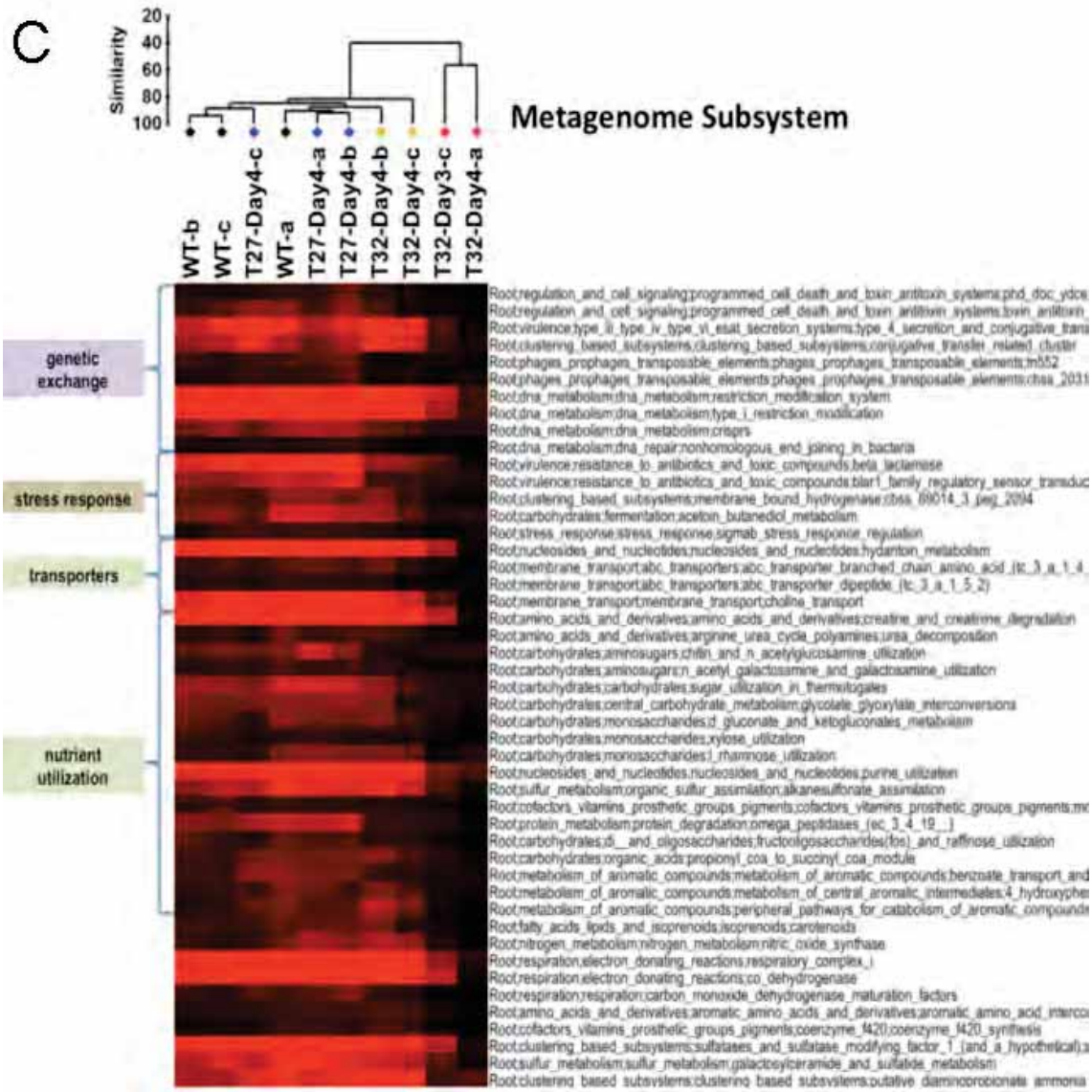


Figure 5.5 Metagenomic functional profiles. A. COG annotation. B. PfamA annotation. C. Subsystem annotation.
Samples are clustered using Bray Curtis similarity and group averages in PRIMER 6 (PRIMER-E Ltd, Lutton, UK). Black dots, the wild-type group. Blue dots, the healthy group. Orange dots, the intermediate group. Red dots, the necrotic group. Heatmaps were plotted according to the abundance of each function (normalised by genome copy) per sample.

5.3.4 Functional dynamics in sponges exposed to elevated water temperature

The phylogenetic and metagenomic analyses showed a clear separation of the necrotic sponges from all other samples. This observation therefore led to the hypothesis that the intermediate state may have undergone substantial expression changes, even without dramatic changes in the community structure, as the intermediate sponges showed some overlap with the necrotic samples. Therefore a metaproteomic approach was used to further compare the changes in expressed functional profiles of the sponge microbial

communities under elevated seawater temperature. Two individual clone samples from each of the healthy (T27-Day4-a and T27-Day4-b), intermediate (T32-Day4-b and T32-Day4-c) and necrotic (T32-Day3-c and T32-Day4-a) groups were selected and their expressed protein profiles characterised using tandem LC-MS/MS. The MS/MS spectra were searched against a database consisting of protein sequences obtained from all 10 metagenomic samples (Section 5.4.4*)*. A total of 809 (3,116 peptides), 781 (3,355 peptides) and 657 (2,279 peptides) proteins from healthy, intermediate and necrotic sponges were identified, respectively.

The differences in protein expression from each sample state were further compared based on the Clusters of Orthologous Groups (COG) functional categories (Figure 5.6). Specific COG functional groups like amino acid transport and metabolism were relatively overrepresented in the healthy group, while post-translational modification was overrepresented in the intermediate group. In contrast, translation, transcription and intracellular trafficking were overrepresented in the necrotic sponges. To establish that the specific proteins that were different between the three groups, the COG, Pfam and SEED annotations of those proteins were extracted and again subjected to pair-wise comparisons (Section 5.3.4). The results were plotted in separate heatmaps representing increased or decreased expression of specific proteins (Figure 5.7).

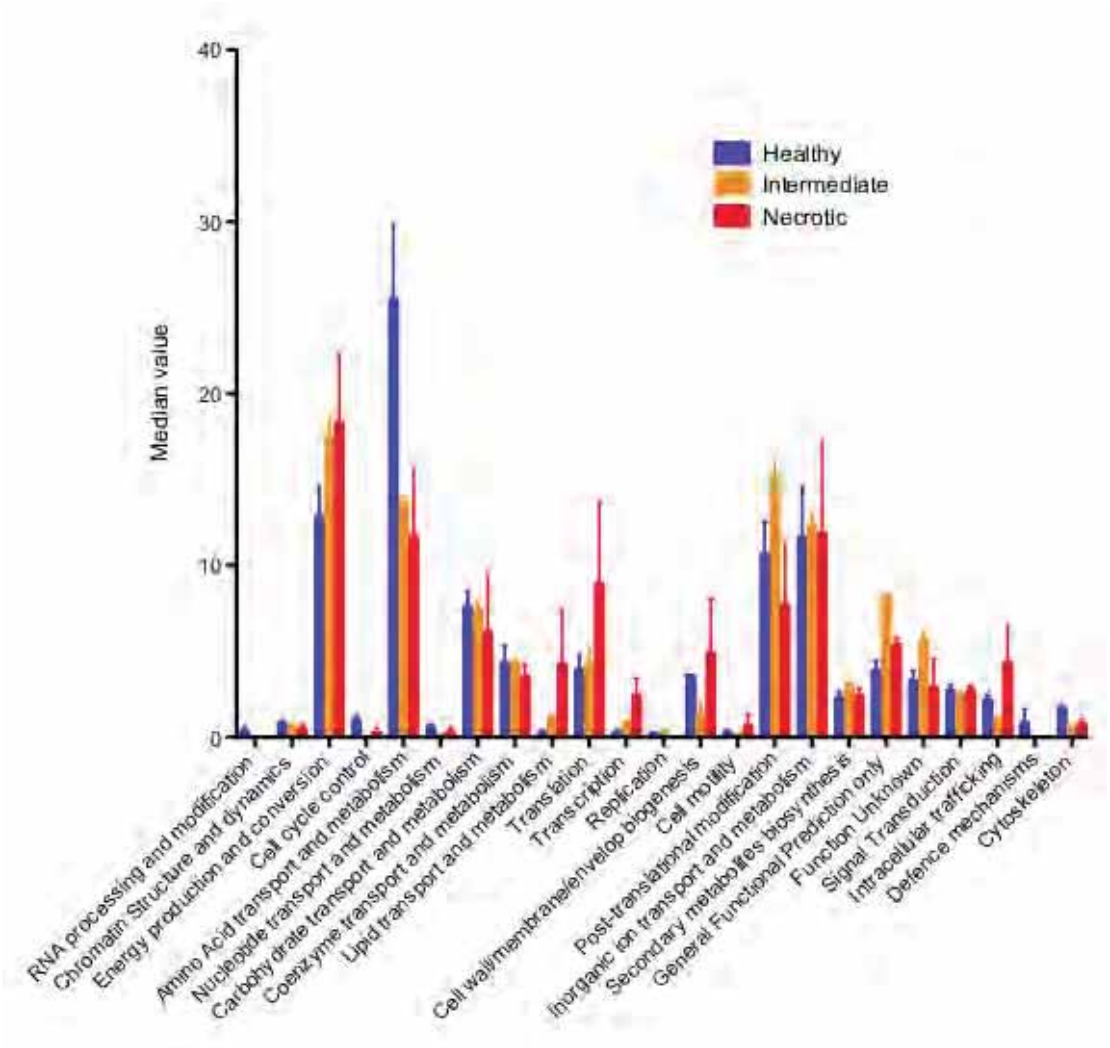


Figure 5.6 Relative abundance of COG categories in healthy (blue), intermediate (orange) and necrotic (red) sponge bacterial communities.
Error bars indicate calculated standard variations of replicates. COG counts were normalised per proteome equivalent and the percentage of total counts in COG is presented here.

5.3.4.1 Changes in active transport systems involved in nutrient acquisition

Abundant transporters were detected in the metagenome of healthy sponges and a number of ABC-type transporters were highly expressed in the healthy sponge metaproteome (Figure 5.7). Two sugar transporters, UgpB (COG1653) for glycerol-3-phosphate and RbsB (COG1879) for ribose, were highly expressed in healthy sponges but decreased markedly in the intermediate group. Glycerol-3-phosphate is used as a phosphate source by many bacteria (Boos 1998) and these transport systems might enable sponge symbionts to scavenge phosphate. UgpB was also highly expressed in the microbial community of *C. concentrica* (Chapter 3). The dramatic reduction in expression of UgpB protein in the intermediate and necrotic sponge communities suggests that the transporter might have an important role in sponge-microbial

symbiosis. The ABC-type transporter Ddp for dipeptides (COG0747, Subsystem: ABC Transporter Dipeptide) and Opp for oligopeptides (Subsystem: ABC Transporter Oligopeptide) were also expressed in the healthy group. The Ddp transporter, responsible for transport of proline-containing dipeptides (Olson *et al.,* 1991) has previously been identified in the metaproteome of the sponge *C. concentrica*, but is missing in the metaproteome of planktonic bacteria (Sowell *et al.,* 2009, Sowell *et al.,* 2011). Furthermore, the membrane-associated protein OmpA was also found to be overrepresented in healthy sponges. This protein is a member of the OmpA-OmpF porin family and OmpA is believed to be a non-specific diffusion channel, allowing various small solutes to cross the outer membrane (Sugawara and Nikaido 1992). Overall, these observations suggest that transporters in nutrient utilisation and signalling are ubiquitous features of sponge symbiotic bacteria and these transporters are highly active in healthy sponges. It is feasible that under thermal stress, they are suppressed, which could lead to nutrient starvation for some bacteria preventing them from carrying out essential cellular functions and further disrupting their association with the host sponge.

5.3.4.2 Changes in active metabolism

Protein functions associated with sugar metabolism were found to be abundantly expressed in healthy sponges, but were mostly reduced in intermediate and necrotic sponges. In particular, the catabolism protein ThuA (PF06283) involved in the utilisation of trehalose was expressed in healthy samples. Trehalose is known to accumulate in a variety of plants and animals, including sponges (e.g. *Suberites domuncula* (Bachinski *et al.,* 1997)), and its intracellular accumulation provides protection against heat and osmotic stress (Dinnbier *et al.,* 1988, Eleutherio *et al.,* 1993). Interestingly, a rapid reduction in trehalose was detected in the marine sponge *S. domuncula* in response to a 10 ℃ increase above ambient seawater temperature (Bachinski *et al.,* 1997). Therefore production, transport and/ or utilisation of trehalose in sponge holobionts could be an important metabolic interaction for symbiosis. A reduced activity in trehalose utilisation in sponge bacteria would have a negative effect on their tolerance to heat shock.

Degradation of aromatic compounds is important for sponge-associated bacteria as it can provide valuable carbon sources (Section 4.3.3.1). Proteins from aromatic

compound degradation pathways were detected in the metaproteomes of all sponges. Specifically, whilst dienelactone hydrolase (COG0412), which is part of chlorocatechol and heterocyclic aromatics degradation pathways (SEED annotation) were detected in the proteome of sponges from all temperature treatments, their expressions were found to be lower in intermediate and necrotic sponges. This again highlights the altered metabolic interactions that occur during tissue necrosis.

5.3.4.3 Changes in stress responses

Sponge symbionts are thought to be exposed to highly variable environmental conditions due to the intermittent pumping activity of the sponge (Vogel 1977) and steep local gradients (Hoffmann *et al.,* 2009). The previous detection of stress-related functions in the metaproteome (Chapter 3) and metagenome of the sponge *C. concentrica* (Thomas *et al.,* 2010) is consistent with the proposal of sponge symbionts having an active stress response. This was also observed by the metaproteomics of *R. odorabile*. For example, the expression of SPFH domain containing proteins potentially involved in phage-infection response (COG0330, COG2268, PF01145, Subsystem: HFL Operon) and proteins in anti-toxic peroxiredoxin (PF10417, Subsystem: Rubrerythrin) were generally overrepresented in the healthy samples. During thermal stress, different stress-related functions were found to be overrepresented. An example was the ATP-dependent molecular chaperone (ClpA, COG0542, PF02861, Subsystem: ClpAS cluster), which has a primary function in the degradation of mis-folded or abnormal proteins. In general, the intermediate samples showed a different protein profile in stress response when compared to the healthy samples and such difference demonstrated a clear link between the responsive functional changes and environmental changes such as elevated water temperature.

5.3.4.4 Sessile versus motile life style

In healthy sponges, functions associated with maintaining cellular integrity and cell division (e.g. membrane associated protein TolQ, COG0811 and LytB, COG0227) were over-represented. TolQ is one of three inner membrane proteins of the Tol-Pal cell envelope complex and is involved in maintaining cell envelope integrity (Walburger *et al.,* 2002), the import of group A colicins (Lazdunski *et al.,* 1998) and in filamentous phage infection (Click and Webster 1997). Recently, the Tol-Pal complex was found to

be part of the cell division machinery and to be required for proper outer-membrane invagination during cell constriction (de Boer *et al.,* 2007).

A number of functions associated with cell adhesion were found to be over-represented in both the healthy and intermediate sponges. They included proteins with fibronectin type III domain (PF00041), which can bind to a variety of substances including collagen, DNA, actin, fibrin and fibronectin receptors on cell surfaces (Dean *et al.,* 1987). Reduced expression of putative adhesins in the necrotic sponge community might relate to a loss of the direct molecular connections between the symbiont and the sponge extracellular matrix, which is known to contain large amounts of collagen (Bergquist 2001). Another cell wall binding protein with repeat structure (PF04122), was found to be overrepresented in the metaproteome of healthy samples. This repeat is found in multiple tandem copies in proteins including amidase enhancers (Kuroda *et al.,* 1992) and adhesins (Waligora *et al.,* 2001). A Cna protein B-type domain (PF05738) was also overrepresented in healthy and immediate sponges. This domain is found in the collagen-binding surface protein of *Staphylococcus aureus* (Deivanayagam *et al.,* 2000). The primary sequence of Cna has a non-repetitive collagen binding A domain, followed by the repetitive B domain. Despite the affinity of the A region for collagen, it is independent of the B region. The B repeat units have been suggested to serve as a 'stalk' that projects the A region from the bacterial surface and thus facilitate bacterial adherence to collagen (Deivanayagam *et al.,* 2000).

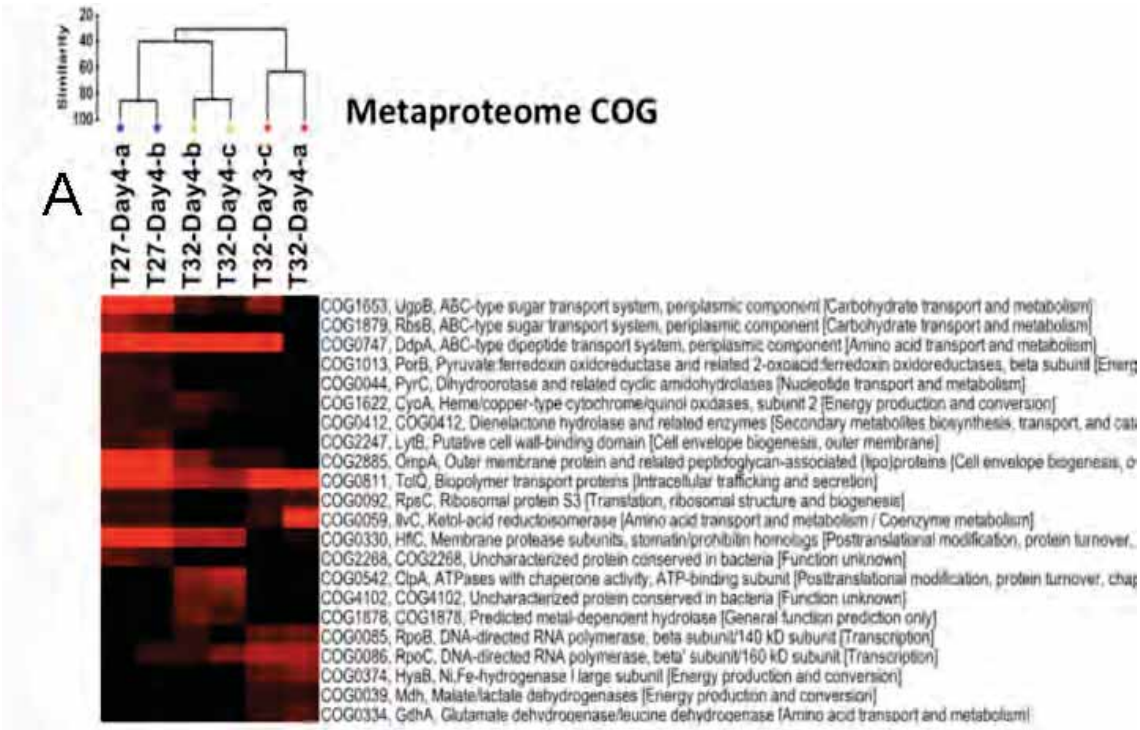
Overall, the results indicated that cell adhesion mechanisms changed from healthy to a necrotic sponge tissue state as a result of increased temperature. These functions may assist the sponge bacteria to adhere to sponge cells and thereby maintain their association with the host. In necrotic sponges these interactions appear to break down, as highlighted by the increased expression of ATP synthase (PF02874) associated with flagella activity in these samples (Figure 5.7). An increased motility function in necrotic sponges is likely due to opportunistic bacteria, as the genomic potential for motility in healthy sponges appears to be limited (Figure 5.5).

5.3.4.5 Active transcription of "invading" bacteria

The shift in microbial community composition between healthy and necrotic sponges was also reflected by general transcriptional and translational activities. In healthy *R.*

odorabile and *C. concentrica* (Chapter 3) communities, those two functions were under-represented in the metagenome compared to the metaproteome, indicating a "static" state of protein expression and regulation. In necrotic *R. odorabile,* transcription-associated functions (e.g. DNA-directed RNA polymerases (COG0085, COG0086, PF04561, PF04563, PF04565, Subsystem: RNA Polymerase Bacterial)) and translation (e.g. RpsC, COG0092) become abundant, consistent with new and invading bacteria actively replication and growing to dominate the community. Eukaryotic-like proteins (ELP) such as NHL repeats (PF01436) are expressed by healthy and intermediate sponge symbiont communities. These ELPs in sponge symbionts may play a role in manipulating their sponge host (Thomas *et al.*, 2010).

Further analysis indicated that no virulence factor was detected from the metaproteome data and combined with the earlier observations of a general decline in genes involved in symbiotic functions and host interactions, it may be suggested that the sponge necrosis was not caused by pathogen infection, but rather by the disturbance and collapse of interactions of the sponge holobiont.



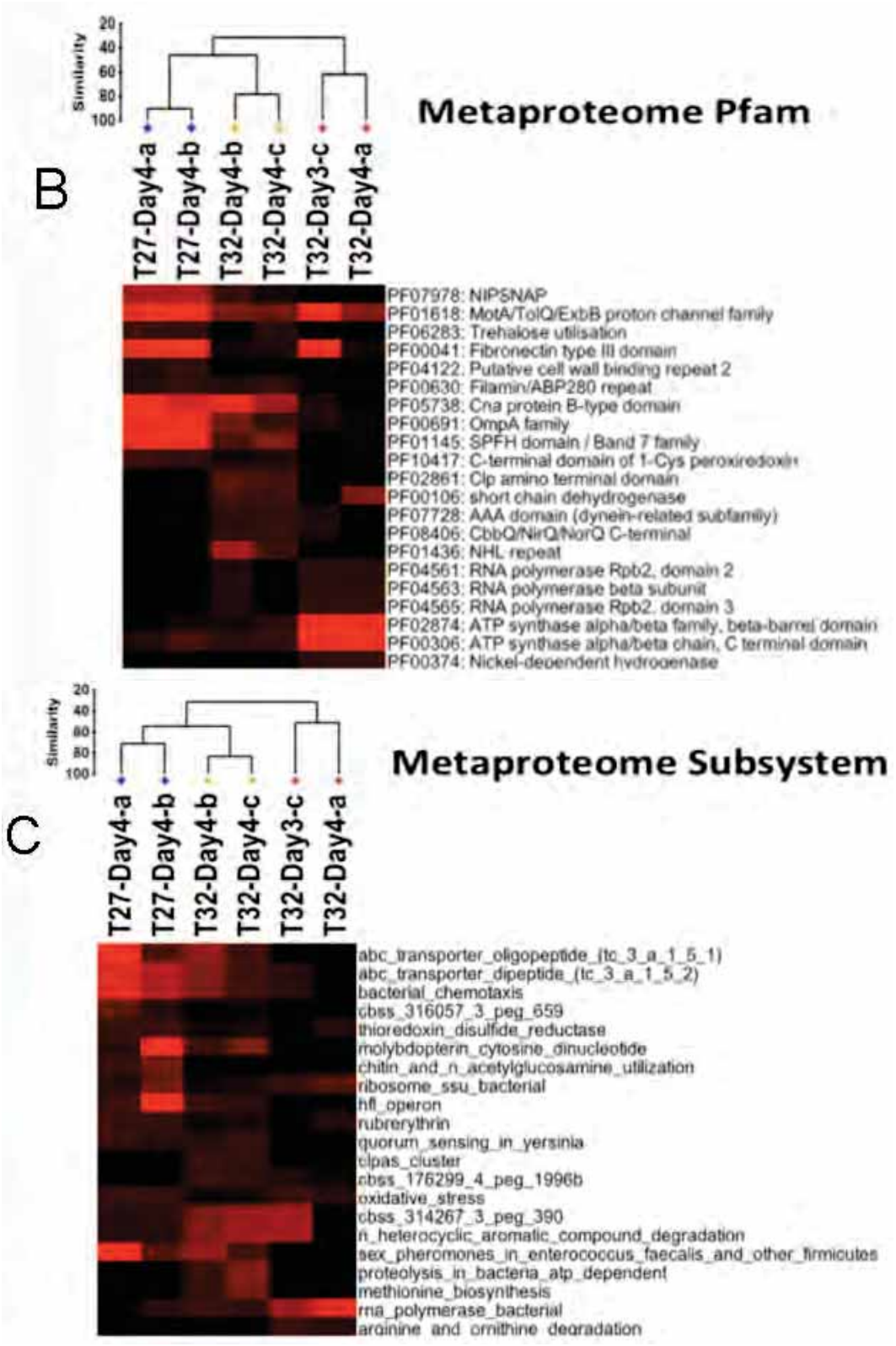


Figure 5.7 Metaproteomics functional profiles.
A. COG annotation. B. PfamA annotation. C. Subsystem annotation. Samples are clustered using the Bray Curtis similarity resemblance and group average in PRIMER 6 (PRIMER-E Ltd, Lutton, UK). Blue dots represent healthy clones, orange dots represent the intermediate clones, and red dots represent necrotic clones. Heatmaps were plotted according to the abundance of each function (standardised to percentage) per sample.

5.4 Conclusions

Pronounced changes in taxonomic composition, species richness, genomic content and expressed functional profiles of sponge bacterial communities occurred under thermal stress. In particular, putatively symbiotic functions, including active transport and metabolism of sponge-specific substrates and functions related to maintaining cellular structure, were identified in healthy sponges at both genomic and proteomic levels. The phylogenetic and genomic composition of the community was stable during early temperature stress, as healthy and intermediate sponges display up to 80% similarity in genome content. However, at the proteome level, only 40% to 55% similarity was observed indicating that symbiont communities within the intermediate sponges had changed their expression profiles. For example, proteins involved in heat stress-responses were abundantly expressed in sponges at this intermediate stage. These results are consistent with a scenario whereby the symbionts still persist in the host, but no longer carry out normal symbiotic functions. Such breakdown in symbiosis would significantly impact the holobiont and thus dramatic phylogenetic and functional changes subsequently occur in the final necrotic stage. Symbiotic functions were lacking from microbial communities of necrotic sponges, which were instead dominated by previously undetected functions such as cell motility and novel metabolic activities. The decline in expression of symbiont functions in stressed sponge sample demonstrated that the sponge necrosis is likely to be caused by the disturbance and collapse of interactions of the sponge holobiont, rather than as a result of pathogen infection.

Chapter Six
General Discussion

6.1 Introduction

Sponges are one of the most ancient multicellular animals (metazoan). They are sessile, filter feeders and represent an important component of benthic ecosystems. Microorganisms form close associations with sponges and can contribute as much as 35% to 60% to the overall biomass (Hentschel *et al.*, 2006, Vacelet and Donadey 1977, Wilkinson 1978b). Sponges have also long been known as a rich source of bioactive secondary metabolites that are of considerable biotechnological and medical interest (Taylor *et al.*, 2007, Wang 2006). Because of these observations, a great deal of interest in sponge research has focused on the analysis of the biodiversity and biogeography of sponge microbial communities. Molecular studies have revealed that diverse and complex microbial assemblages covering 30 bacterial and two archaeal phyla reside in marine sponges (Lee *et al.*, 2011, Taylor *et al.*, 2007, Webster *et al.*, 2009, Zhu *et al.*, 2008). In addition, high similarities among sponge-associated microbial communities have been described, where the identified microbial communities were also phylogenetically distinct to the surrounding seawater. (Hentschel *et al.*, 2002, Taylor *et al.*, 2007). Sponge-associated microorganisms can contribute to the well being of the host though different microprocesses, such as photosynthesis (Erwin and Thacker 2008), nitrification (Lopez-Legentil *et al.*, 2010), sulfate reduction (Hoffmann *et al.*, 2005) and the production of bioactive secondary metabolites that protect the host against environmental perturbations (Faulkner 2001, Flatt *et al.*, 2005, Piel 2004). However, details of the role of microbial symbionts remain largely unknown and there is still a lack of empirical evidence confirming the nature of the symbiotic interactions between sponge-associated microbes and their hosts (Webster and Blackall 2008).

Advances in metagenomic sequencing strategies has revolutionised the field of microbial ecology and have successfully been applied to a variety of environmental samples including oceanic water, soil and the human microbiome (Rusch *et al.*, 2007,

Tringe *et al.*, 2005, Turnbaugh and Gordon 2009). A metagenomic approach was recently applied to an investigation of the microbial community associated with the marine sponge *Cymbastela concentrica* (Thomas *et al.*, 2010). This study revealed distinctive microbial phylotypes and provided valuable insights into potential host-symbiont interactions in marine sponges. In order to extend our understanding of the symbiotic interactions within sponges, this study assessed sponge specific phylotypes that were identified from the metagenomic dataset presented by Thomas *et al.* (2010) (Chapters 2 and 4). From the five dominant phylotypes identified from this metagenomic analysis, two uncultivated phylotypes, one belonging to the *Deltaproteobacteria* and the other to the *Phyllobacteriaceae*, were studied with respect to their potential functions and localisation within the community. Using the reconstructed partial genome, distinct functional properties of the two phylotypes including their lifestyle and interactions with the host were identified. Their localisation within the sponge was also established using fluorescence *in situ* hybridisation (FISH) analysis.

Research conducted as part of this thesis also aimed at improving our understanding of the functional potential of sponge-associated microorganisms in general, by assessing the global functional profile, within the microbial community, through the application of metaproteogenomic analysis (Chapter 3). Such analysis integrates metagenomic data with mass-spectrometry based proteomics, and uses the direct measurement of protein expression to reflect the overall functional activities within the sponge microbial community. The analysis enabled the detection of abundant protein expression in transport function, aerobic and anaerobic metabolism, stress responses as well as identifying proteins that facilitate the host-symbiont interactions (Chapter 3). The result of the analysis also demonstrated that the protein profiles could be linked back to the uncultured *Phyllobacteriaceae* phylotype, which was found to be a dominant phylotype in the metagenome data. The data reflected many adaptation strategies employed by the sponge-associated microorganisms, which potentially aid their persistence in the sponge host.

The differences in protein expression also reflected the functional changes occurring within the microbial community resulting from a fluctuating environment. This was demonstrated by metaproteogenomics of a microbial community in the sponge

Rhopaloeides odorabile under the influence of thermal stress (Chapter 5). Dramatic changes in community structure, species richness, genomic content and functional expression were observed using this combined approach. Based on such observations, it was concluded that a decline in sponge health was caused by the disturbance and the collapse of host-symbiont interactions, rather than disease being caused by pathogenic organisms (Chapter 5).

6.2 Defining the spatial arrangement of sponge microbial communities

The study of microbial ecology focuses on three fundamental questions: (1) what is the structural composition of the microbial community?; (2) what functions are carried out by members of the community?; and (3) how are they spatially arranged in the environmental sample? Metagenomics has the capacity to simultaneously address all three questions. This important tool was therefore utilised in the work described in this thesis and resulted in the identification of two uncultured microbial phylotypes from the sponge *C. concentrica*. Extensive genomic analyses on the partial genome of those two phylotypes were presented in Chapters 2 and 4, respectively; and led to three important findings. First, phylogenetic analysis based on the 16S rRNA gene sequence demonstrated that the Deltaproteobacterium identified was closely associated to the predatory *Bdellovibrio* bacterium. The intracellular lifestyle allows the *Bdellovibrio* bacterium to invade other bacteria, grow and divide inside the prey cell and eventually lyse the host to release more progeny that can attack additional prey (Sockett 2009). Further analysis of the partial genome found that functions associated with cell attachment, motility, elongation and division were conserved across the Deltaproteobacterium and *Bdellovibrio bacteriovorus*, a type strain of the genus *Bdellovibrio*. This second finding indicated that the bacterium could also live an intracellular-like lifestyle in the sponge. The third finding, and perhaps the most important, was the detection of the sponge Deltaproteobacterium through FISH analysis. FISH is a well-developed *in situ* tool based on the complementary binding of a synthetic labeled DNA probe to the target DNA fragments, usually a ribosomal RNA gene such as the 16S rRNA gene (Amann *et al.*, 1995).

6.2.1 Distinct symbiotic life style in the sponge revealed by direct localisation study

FISH analysis demonstrated that the sponge-associated Deltaproteobacterium formed a close association with a larger auto-fluorescent cyanobacterial cell type within the sponge tissue, with a potential intracellular life-style (possibly predatory, Chapter 2). The novel observation represented "a symbiosis within a symbiosis", where the Deltaproteobacterium was hosted by a cyanobacterium that was in turn hosted by the sponge. This observation was consistent with an earlier observation, where “Bdellovibrio-like” particles were identified within unicellular cyanobacteria in the mesohyl of the two coral reef sponges, *Neofibularia irata* and *Jaspis stelifera*, using electron microscopy (Wilkinson 1979a).

Cyanobacterial association with sponges is common. Symbiotic cyanobacteria are abundant within sponges found in shallow reef habitats, where they contribute 25% to 50% of the host's cellular volume (Erwin and Thacker 2007, Rutzler 1990). In addition to their widespread existence in sponges (Erwin and Thacker 2007, Rutzler 1990, Steindler *et al.,* 2005, Thacker 2005, Usher *et al.,* 2004), they also demonstrate functions that are potentially beneficial for the host. These functions include the supply of photosynthetic products to the host (Arillo *et al.,* 1993, Wilkinson and Fay 1979), nitrogen fixation (Wilkinson and Fay 1979), production of secondary metabolites (Flatt *et al.,* 2005) and UV protection (Regoli *et al.,* 2000). Widespread dispersal and horizontal transmission of sponge-associated cyanobacteria in sponges has been suggested (Steindler *et al.,* 2005, Thacker 2005), as well as vertical transmission of these symbionts from parents to offspring (Oren *et al.,* 2005, Usher *et al.,* 2001, Usher *et al.,* 2005). Despite the importance of cyanobacteria in the ecological success of sponges, relatively little is known about how they maintain an association with their host. The novel observation in this thesis of a ‘symbiosis within a symbiosis’ provides a potential explanation for how cyanobacteria maintain their presence in sponge cells. It is anticipated that they achieve this with the help of eukaryotic-like proteins (ELPs), which are expressed and secreted by the associated deltaproteobacterium. The ELPs have been noted in the genomes of obligate and facultative intracellular symbionts of eukaryotic cells (Wu *et al.,* 2004) and found to be abundant and highly conserved in the sequenced genomes of the causative agent of Legionnaires’ disease, *Legionella pneumophila* (Cazalet *et al.,* 2008). Functional analysis of ELPs in *L. pneumophila*

suggests the proteins are likely to interfere with eukaryotic cellular functions and may thus constitute virulence factors (Bruggemann *et al.*, 2006). This prediction was further supported by several follow-up analyses, confirming their implication in virulence, host cell modulation and bacterial survival (Al-Khodor *et al.*, 2008, de Felipe *et al.*, 2008, Habyarimana *et al.*, 2008, Nora *et al.*, 2009, Pan *et al.*, 2008, Sansom *et al.*, 2007, Sansom *et al.*, 2008). Such findings also underline the importance of ELPs in the survival of cyanobacteria in the sponge, where the proteins could be used by the bacteria to avoid digestion from host cells.

6.2.2 Linking localisation of the sponge microbes with their functional potentials

To date, the application of FISH analysis to sponge-associated bacteria has mainly focused on checking the presence of abundant bacteria within environmental samples. For example, FISH analysis was used to confirm the localisation and abundance of the candidate phylum '*Poribacteria*' and nitrite-oxidizing bacteria of the genus *Nitrospira* within sponge microbial communities in a recent 454 amplicon pyrosequencing study (Webster *et al.*, 2009). Advancements in sequencing have also resulted in rapid growth of the field of sponge symbiotic functional analysis, reflected by the whole-genome amplification of the '*Poribacteria*' clade and the sponge specific *Chloroflexi* clade (Siegl and Hentschel 2010, Siegl *et al.*, 2011). Sequencing data demonstrated that the '*Poribacteria*' are potential mixotrophs that can undertake autotrophic carbon fixation through the Wood-Ljungdahl pathway (Siegl *et al.*, 2011). In addition, the '*Poribacteria*' was also predicted to have the ability to maintain anaerobic respiration under oxygen limiting conditions, which would be the case in the microaerobic or anaerobic microhabitats, during active microbial metabolism, and during non-pumping periods in the sponge *Aplysina aerophoba*. Although attempts were made in previous studies to link functional characteristics to physical habitats in sponges, none thus far has attempted to link functional characteristics of bacteria with their actual sponge localisation.

Similar to the '*Poribacteria*', the *Phyllobacteriaceae* phylotype described in Chapter 4 was also predicted to be capable of carrying out anaerobic respiration through nitrate reduction. An additional FISH analysis on this phylotype confirmed its presence within the sponge's inner tissue where oxygen is expected to be limited. This observation not

only consolidated the prediction of the anaerobic capacity of the *Phyllobacteriaceae* in sponges, but also provided evidence to link functional characteristics with physical localisation. The systematic analysis of metagenomic information with FISH data presented here has demonstrated utility for microbial ecology. It enables valuable insight into the phenotypic characteristics of dominant microorganisms by linking together their phylogeny, predicted functions and spatial arrangements.

6.3 Defining the functional potential of sponge microbial communities

Sponges are host to extensive microbial diversity – as outlined by profiling studies involving both culture and culture-independent approaches (Reviewed in Taylor *et al.*, 2007). In building on this key community information recent research has shifted towards attempts to understand the underlying functional potential of sponge microbial communities. This effort led to significant progress being made in the studies of specific symbiotic functions in sponges, with particular attention focused on nitrogen metabolism (Reviewed in Webster *et al.*, 2011). Rapid developments in sequencing technologies have further accelerated research in sponge microbial communities. Both whole-genome amplification and metagenomic analysis have been carried out (Siegl and Hentschel 2010, Siegl *et al.*, 2011, Thomas *et al.*, 2010, Webster *et al.*, 2009), revealing important information about sponge-associated microbial symbionts. Metagenomic data also provide a platform for undertaking other 'meta' based studies, such as metatranscriptomics and metaproteomics. Metaproteomics has recently been developed to characterise functional responses within several environmental microbial communities. This global approach has the potential to reveal the activities of functional cellular processes and ecological processes, occurring within the specific environmental system (Thomas *et al.*, 2007). Data presented in this thesis details a combined functional study using metagenomics and metaproteomics (metaproteogenomics) conducted to investigate the proteins expressed by the indigenous sponge-associated microorganisms in the sponge *C. concentrica* (Chapter 3).

6.3.1 Transporter functions are ubiquitous among sponge microbes

The analysis presented in Chapter 3 revealed active expression of proteins involved in functions associated with transport, stress response, metabolism and the putative function of eukaryotic-like proteins (ELPs). These results are consistent with similar systems presented in the literature, notably the high proportion of proteins associated with active transporter functions, which were detected from metaproteomic studies of microbes from the open sea, e.g. the SAR11 clade from the Sargasso Sea (Sowell *et al.,* 2009, Sowell *et al.,* 2011). Such prevalence of transporter functions was not unexpected with the SAR11 clade, because of the high surface area to cell volume ratio in the small free-living members of the clade. Cyto-electron tomography images established that 25% - 35% of the SAR11 cell volume was periplasmic space (Nicastro *et al.,* 2006). This ensures that the microorganisms can maximise the equilibrium between the cytoplasmic volume and the membrane's capacity to import nutrients for metabolism (Button and Robertson 2000). The optimal design would allow the microorganisms to devote a large proportion of resources to the expression of high affinity uptake systems that target multiple substrates. This ensures the microorganism is better adapted to oligotrophic environments like the open ocean. From a general perspective, the abundant transporter functions detected in the sponge microbial community (Chapter 3), may reflect the probability that sponge-associated microorganisms are also actively scavenging a variety of nutrient substrates, suggesting the possibility that they could also develop a trophic strategy similar to the microbes detected from the open sea (Morris *et al.,* 2010, Sowell *et al.,* 2009, Sowell *et al.,* 2011). In addition, different types of substrate-specific transporters were also expressed by the sponge microorganisms. Typically, transporters for sponge metabolites, such as halogenated aromatics and dipeptides, were identified. This observation indicated that sponge-associated microorganisms devote a substantial resources to their acquisition of host-derived nutrients, providing direct evidence for the existence of a mutualistic relationship between the sponge and its symbionts.

Similarly, transporter functions were also abundant in healthy samples of the sponge *R. odorabile*. In particular, the Ddp transporter, which is responsible for transport of proline-containing dipeptides (Olson *et al.,* 1991). This again highlighted the propensity of sponge-associated bacteria to harbour the transporter function for sponge-derived

substrates. Further, a membrane-associated protein for a non-specific diffusion channel was also detected, likely allowing small solutes to diffuse across the outer membrane. This indicates that the sponge-associated microorganism can also use non-specific diffusion to optimise its nutrient uptake. Overall, the results from the metaproteomic analyses further supported the role of transporter functions for nutrient scavenging as a ubiquitous feature among sponge-associated microorganisms.

The metaproteogenomic approach outlined in this thesis allowed for a direct and parallel assessment of the active fraction of the sponge-associated microorganisms (symbionts), thus extending the information obtained by metagenomic analysis (Thomas *et al.*, 2010) and deep sequencing approaches of PCR amplicon libraries (Webster *et al.*, 2009). In addition, the metaproteomics data also linked the most abundantly translated proteins and the dominant taxonomy from the same metagenome preparation to be established. By mapping the detected proteins to the five partial genomes assembled based on the metagenome data, the *Phyllobacteriaceae* phylotype was shown to express the highest number of proteins, and transporter functions for a variety of substrates were actively expressed. This observation confirmed that the *Phyllobacteriaceae* phylotype is a dominant member of the sponge microbial community and functionally active within the sponge.

6.3.2 Functional responses towards the external environment

In addition to the oligotrophic nature of the marine sponge habitat, the physical environment can also be hostile for the residing microorganisms. The cyclic pumping activity of sponges (Vogel 1977) and steep local gradients (Hoffmann *et al.*, 2009) expose sponge-associated microorganisms to variable environmental conditions.
Stress response functions were identified for symbiotic bacteria associated with both sponges in this study using metaproteomic analyses (Chapters 3 and 5), indicating adaptation to environmental stressors. Abundant proteins associated with post-translational modification, protein turnover, chaperone functions as well as a choline dehydrogenase BetA, which catalyzes the oxidation of choline to glycine betaine as a response to osmotic stress (Landfald and Strom 1986), were detected in the *C. concentrica* metaproteome. In addition, proteins potentially involved in phage-infection

response (COG0330, COG2268, PF01145, Subsystem: HFL Operon) and proteins in anti-toxic peroxiredoxin (PF10417, Subsystem: Rubrerythrin) were detected in the metaproteome of *R. odorabile*. Such observations suggest that microorganisms are capable of dealing with the potentially stressful environment of the host sponge by actively expressing stress response functions and eliminating denatured or damaged proteins. The stress responses demonstrate the utility of metaproteomic analysis in determining the impact of environmental stressors on sponge-associated microorganisms, responses that, in turn affect host-sponge wellbeing. Thus, an integrated metaproteogenomic approach was applied to assess the functional changes in sponge microbial community at elevated water temperatures (Chapter 5).

6.4 Defining environmental stress responses of a sponge microbial community

The term ‘coral holobiont’ was first used to describe the complex entire community of corals, which is comprised of the coral host, the eukaryotic zooxanthellae and the associated microorganisms (Rohwer *et al.,* 2002). It was further suggested that corals could form mutualistic symbiosis with the microorganisms, where the symbiotic microbes contribute to the capacity of the holobiont to adapt or acclimatise to environmental stress (Reshef *et al.,* 2006, van Oppen *et al.,* 2009). A similar term, ‘sponge holobiont’, was proposed to describe the complex community in sponges with particular emphasis on sponge-associated microorganisms (Webster and Taylor 2011). Considering the diversity, abundance and the specific functional capacity of sponge microorganisms, the assessment of their functional responses towards environmental stress should be an important aspect in ‘sponge holobiont’ research.

This thesis details a combined experimental and analytical approach successfully applied for the assessment of changes of the microbial community during environmental stress. No previous studies have systematically analysed changes by simultaneously incorporating information on abundance, phylogenetic composition and functional expression. The data obtained provides strong evidence that alterations in the host-symbiont relationship can lead to a decline in sponge health. This was first shown through the phylogenetic and diversity assessment using T-RFLP and the reconstruction

of 16S rRNA gene sequences from metagenome data. Three distinct physical states were determined and were correlated to the physical appearances of the samples from different stages of the experiment. Distinct community shifts and diversity changes were identified from the three states. A transient state was noted in the intermediate samples, where the community structure was similar to the native microbial community, but yet a distinct shift towards the community characterised in necrotic samples was observed.

The metaproteomic analysis indicated clear changes in function between the three states. COG functional category analysis indicated that while amino acid transport and metabolism were overrepresented in healthy sponge tissue, stress responses such as post-translational modification were found to be overrepresented in intermediate samples. Although the native microbes persisted in the sponge, their functions were not actively expressed; this was particularly the case for functions usually expressed in the healthy host such as transporters and metabolic activities for sugar and phosphate. These putative functions were replaced by an abundant expression of functions associated with heat stress responses. An even more dramatic change was observed in the necrotic samples, where the functional categories associated with translation, transcription and intracellular trafficking were overrepresented. Given that the community structure was also completely different in necrotic samples, it indicated that the 'new' microbial community would be responsible for carrying out such functions, likely following a breakdown in the symbiotic relationship between sponges and its symbionts. A case in point was demonstrated through the reduced expression of functions associated with cell adhesion such as the fibronectin type II domain in necrotic samples. This protein can bind to a variety of substances including collagen, actin and fibrin (Dean *et al.,* 1987) and the sponge symbionts may rely on this function to remain attached to the sponge extracellular matrix as it contains large amounts of collagen (Bergquist 2001). Therefore the reduced expression of this function in the necrotic sponge could cause the loss of the direct molecular connections between symbionts and the sponge extracellular matrix, resulting in the dissociation of symbiotic microorganisms from the sponge tissue.

Previous studies investigating the effect of environment stresses on sponge health have focused on either abiotic or biotic factors, with most work being limited to assessing

changes in microbial community structure and species richness. All these studies detected shifts in the normal microbial communities upon exposure to environmental changes, such as elevated seawater temperatures (Webster *et al.,* 2008a) and exposure to heavy metals (Selvin *et al.,* 2009, Webster *et al.,* 2001a), leading to the conclusion that alterations in host-symbiont community structure caused by environmental influences are correlated with a decline in sponge health, although convincing evidence for the link at a functional level was not demonstrated in these studies.

Studies investigating sponge disease outbreaks have also attempted to link disease state with pathogens (Webster 2007, Webster and Blackall 2008). Only one study has successfully characterised a bacterial pathogen from the sponge *R. odorabile* (Webster *et al.,* 2002). The specific *Alphaproteobacterium* identified was related to the tumor-forming symbionts of *Prionitis sp*. macroalgae, capable of infecting sponges and killing healthy tissue by producing a collagenase enzyme that degrades the spongin fibers (Mukherjee *et al.,* 2009). The results suggested that sponge disease might not be related to pathogens, but rather was due to the breakdown of symbiont-host interactions. The increasing abundance of new phylotypes detected in the intermediate samples could potentially represent 'opportunistic' microbes that are more adaptable to the necrotic sponge tissue and capable of degrading sponge tissue.

A complex interplay between the environment, the abundance and diversity of the microbial community, along with the functional dynamics of its members in response to changes in environmental conditions, was established using the combined approach outlined in Chapter 5. This approach extended beyond the traditional phylogenetic based assessment most often reported in the literature, by adding a comprehensive metaproteomic analysis, effectively revealing actual responses occurring in the community *in situ*.

6.5 Future research directions

The metaproteogenomic studies presented in this thesis provide additional information on the phylogenetic and functional nature of marine sponge-associated microbial communities. They also led to the identification of several key functional aspects such

as transporter functions for sponge-derived substrates, stress responses and ELPs. These functional predictions could be verified using manipulative experiments to place the results in an ecologically meaningful context. This could be achieved by isolating the bacteria belonging to the dominant phylotypes. For instance, the availability of the dominant strains would greatly improve the recombinant approach that was carried out to test the functions of the ankyrin repeats protein in the progression of phagocytosis in sponges (Nguyen 2011). The available strains could provide more extensive functional information and be directly used in testing the predictive function in delaying phagocytosis against amoebocytes.

To further investigate the host-symbiont relationship in sponges, additional studies on the identified transporter functions and metabolism could be undertaken based on expression in a heterologous host. In this case, it would be important to elucidate the specific function of a given transporter for sponge-derived substrates as they could be used to identify the metabolic potential of the target phylotypes, which could aid in pure culture isolation. The details of stress related responses in sponge communities, induced by changes in the environment are also important components of the symbiont-host interaction. Expression studies using RT-PCR provide a feasible way to check those functions directly using samples, such as stress-exposed specimens. Such studies would allow the hypotheses presented in this thesis to be tested and would further enhance our understanding of symbiotic interactions between sponge-associated microorganisms and their hosts.

Although isolation attempts were made in Chapter 4, they were generally not successful. In order to isolate sponge microorganisms, an experimental design based on the considerations addressed in Chapter 4 is required. The genomics-based approach should still be paramount in such future isolation attempts, as this approach has been successfully applied in other systems. For example, the information gained from the metagenomic data associated with nitrogen metabolism was instrumental in the culturing of dominant isolates from the acid mine drainage system (Tyson *et al.,* 2005).

Sequencing the microbial communities associated with marine sponges is expected to continue, with improvements in sequencing technologies, inevitably leading to a much larger sequence databases. One immediate use of such a dataset would be the

identification of functional genes associated with active compounds. Sponges are a rich source of bioactive secondary metabolites, some of which are thought to be produced by the associated microbes (Faulkner 2001, Piel 2004). Both model organisms studied in this thesis have previously been shown to produce bioactive compounds. For example, indole derivatives including bromoindoles have been isolated from marine sponge *R.odorabile* that are known to display anticancer, antibiotic and anti-inflammatory activies (Longeon *et al.,* 2011). In addition, novel hydrolytic enzymes have been identified from marine sponge *C.concentrica* that produce hydrolytic/lipolytic activities and conferred antibacterial properties. More importantly, those novel enzymes were shown produced from bacterial origin (Yung *et al.,* 2011). It is therefore worthwhile to revisit the metagenomic generated data and search for genes or pathways associated with active secondary metabolite production based on the sequence similarity to existing known compounds.

The analysis assessing the changes in community structure, functional genes and proteins during thermal stress was carried out in a controlled environment using small flow-though aquaria. Previously, shifts in microbial communities associated with the sponges *Ircina strobilina*, *Mycale laxissima* and *Clathria prolifera* were identified when they were transferred to and cultivated in aquaculture settings (Isaacs *et al.,* 2009, Mohamed *et al.,* 2008). In another example, the cyanobacterial density in the sponge *Aplysina aerophoba* responded to differences in aquaria light conditions, where the cyanobacterial density increased by 124% with light exposure and decreased to only 1.7% of the original level when in darkness (Kloppel *et al.,* 2008). This observation clearly demonstrates the ability of microbial communities in *ex situ* maintained sponges to adapt to artificial conditions and highlights the changes the community undergoes in response to artificial conditions. It is therefore anticipated that sponge samples maintained in aquaria may not reflect their natural environment counterparts, and hence the changes in functional expression would be expected to differ. Thus it is important to investigate the functional expression of sponge communities *in situ.* To accomplish this task, microbial community protein profiles should be sampled directly from sponges undergoing actual seasonal changes in their natural environment. It is anticipated that differences in functional expression will provide more direct and accurate insights into the adaptive capacity of sponge symbioses and how changes in environmental stress affect the resilience of the host-symbiont interactions.

6.6 Concluding remarks

The work presented in this thesis utilised the approach of metagenomic and metaproteomic analyses to extend the knowledge of the structure, function and dynamics of microbial communities associated with marine sponges. Many interesting aspects of the host-symbiont interactions were revealed and led to questions for further analysis to explore the mechanisms involved. The work presented here highlights the usefulness of the metaproteogenomic approach in microbial ecology as well as the appropriateness of using a marine sponge as a model system for studying eukaryote-prokaryote interactions. In extension, many of the issues discussed and observations described in this thesis are not only relevant to the studies of sponge microbial communities, but also contribute to the advancement of general microbial ecology.

References

Abram F, Enright AM, O'Reilly J, Botting CH, Collins G, O'Flaherty V (2011). A metaproteomic approach gives functional insights into anaerobic digestion. *J Appl Microbiol* **110:** 1550-1560.

Adamczeski M, Reed AR, Crews P (1995). New and known diketopiperazines from the Caribbean sponge, *Calyx cf. podatypa*. *J Nat Prod* **58:** 201-208.

Aebersold R, Mann M (2003). Mass spectrometry-based proteomics. *Nature* **422:** 198-207.

Al-Khodor S, Price CT, Habyarimana F, Kalia A, Abu Kwaik Y (2008). A Dot/Icm-translocated ankyrin protein of *Legionella pneumophila* is required for intracellular proliferation within human macrophages and protozoa. *Mol Microbiol* **70:** 908-923.

Alavi M, Miller T, Erlandson K, Schneider R, Belas R (2001). Bacterial community associated with *Pfiesteria*-like dinoflagellate cultures. *Environ Microbiol* **3:** 380-396.

Alli OA, Gao LY, Pedersen LL, Zink S, Radulic M, Doric M *et al.* (2000). Temporal pore formation-mediated egress from macrophages and alveolar epithelial cells by *Legionella pneumophila*. *Infect Immun* **68:** 6431-6440.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215:** 403-410.

Amann R, Snaidr J, Wagner M, Ludwig W, Schleifer KH (1996). *In situ* visualization of high genetic diversity in a natural microbial community. *Journal of Bacteriology* **178:** 3496-3500.

Amann R, Fuchs BM, Behrens S (2001). The identification of microorganisms by fluorescence *in situ* hybridisation. *Curr Opin Biotechnol* **12:** 231-236.

Amann R, Fuchs BM (2008). Single-cell identification in microbial communities by improved fluorescence *in situ* hybridization techniques. *Nat Rev Microbiol* **6:** 339-348.

Amann RI, Krumholz L, Stahl DA (1990). Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. *J Bacteriol* **172:** 762-770.

Amann RI, Ludwig W, Schleifer KH (1995). Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol Rev* **59:** 143-169.

Andrews JH, Harris RF (2000). The Ecology and Biogeography of Microorganisms on Plant Surfaces. *Annu Rev Phytopathol* **38:** 145-180.

Angly FE, Willner D, Prieto-Davo A, Edwards RA, Schmieder R, Vega-Thurber R *et al.* (2009). The GAAS metagenomic tool and its estimations of viral and microbial average genome size in four major biomes. *PLoS Comput Biol* **5:** e1000593.

Arillo A, Bavestrello G, Burlando B, Sara M (1993). Metabolic Integration between Symbiotic Cyanobacteria and Sponges - a Possible Mechanism. *Marine Biology* **117:** 159-162.

Ashworth RB, Cormier MJ (1967). Isolation of 2,6-Dibromophenol from Marine Hemichordate *Balanoglossus Biminiensis*. *Science* **155:** 1558-1559.

Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA *et al.* (2008). The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9:** 75.

Bachinski N, Koziol C, Batel R, Labura Z, Schroder HC, Muller W (1997). Immediate early response of the marine sponge *Suberites domuncula* to heat stress: reduction of trehalose and glutathione concentrations and glutathione *S*-transferase activity. *Journal of Experimental Marine Biology and Ecology* **210:** 129-141.

Baer ML, Ravel J, Chun J, Hill RT, Williams HN (2000). A proposal for the reclassification of *Bdellovibrio stolpii* and *Bdellovibrio starrii* into a new genus, *Bacteriovorax* gen. nov. as *Bacteriovorax stolpii* comb. nov. and *Bacteriovorax starrii* comb. nov., respectively. *International journal of systematic and evolutionary microbiology* **50 Pt 1:** 219-224.

Baer ML, Ravel J, Pineiro SA, Guether-Borg D, Williams HN (2004). Reclassification of salt-water *Bdellovibrio* sp. as *Bacteriovorax marinus* sp. nov. and *Bacteriovorax litoralis* sp. nov. *International journal of systematic and evolutionary microbiology* **54:** 1011-1016.

Banik JJ, Brady SF (2010). Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Curr Opin Microbiol* **13:** 603-609.

Bannister R, Hoogenboom M, Anthony K, Battershill C, Whalan S, Webster N *et al.* (2011). Incongruence between the distribution of a common coral reef sponge and photosynthesis. *Marine Ecology Progress Series* **423:** 95-100.

Bannister RJ, Battershill CN, de Nys R (2010). Demographic variability and long-term change in a coral reef sponge along a cross-shelf gradient of the Great Barrier Reef. *Marine and Freshwater Research* **61:** 389-396.

Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S *et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315:** 1709-1712.

Bayer K, Schmitt S, Hentschel U (2008). Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environ Microbiol* **10:** 2942-2955.

Beja O, Suzuki MT, Koonin EV, Aravind L, Hadd A, Nguyen LP *et al.* (2000). Construction and analysis of bacterial artificial chromosome libraries from a marine microbial assemblage. *Environ Microbiol* **2:** 516-529.

Bell JJ (2008). The functional roles of marine sponges. *Estuar Coast Shelf S* **79:** 341-353.

Belnap CP, Pan C, Denef VJ, Samatova NF, Hettich RL, Banfield JF (2011). Quantitative proteomic analyses of the response of acidophilic microbial communities to different pH conditions. *ISME J* **5:** 1152-1161.

Bendtsen JD, Kiemer L, Fausboll A, Brunak S (2005). Non-classical protein secretion in bacteria. *BMC Microbiol* **5:** 58.

Benndorf D, Balcke GU, Harms H, von Bergen M (2007). Functional metaproteome analysis of protein extracts from contaminated soil and groundwater. *ISME J* **1:** 224-234.

Benndorf D, Vogt C, Jehmlich N, Schmidt Y, Thomas H, Woffendin G *et al.* (2009). Improving protein extraction and separation methods for investigating the metaproteome of anaerobic benzene communities within sediments. *Biodegradation* **20:** 737-750.

Bergquist PR (1978). *Sponges*. Hutchinson and Co. Ltd: London, United Kingdom.

Bergquist PR (2001). *Porifera (Sponges)*. John Wiley & Sons, Ltd.

Bertin PN, Heinrich-Salmeron A, Pelletier E, Goulhen-Chollet F, Arsene-Ploetze F, Gallien S *et al.* (2011). Metabolic diversity among main microorganisms inside an arsenic-rich ecosystem revealed by meta- and proteo-genomics. *ISME J* **5:** 1735-1747.

Beszteri B, Temperton B, Frickenhaus S, Giovannoni SJ (2010). Average genome size: a potential source of bias in comparative metagenomics. *ISME J* **4:** 1075-1077.

Blatch GL, Lassle M (1999). The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays* **21:** 932-939.

Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E *et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31:** 365-370.

Boos W (1998). Binding protein-dependent ABC transport system for glycerol 3-phosphate of *Escherichia coli*. *Methods Enzymol* **292:** 40-51.

Borchiellini C, Manuel M, Alivon E, Boury-Esnault N, Vacelet J, Le Parco Y (2001). Sponge paraphyly and the origin of Metazoa. *J Evolution Biol* **14:** 171-179.

Bork P (1993). Hundreds of ankyrin-like repeats in functionally diverse proteins: mobile modules that cross phyla horizontally? *Proteins* **17:** 363-374.

Brady A, Salzberg SL (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nat Methods* **6:** 673-676.

Brady SF, Chao CJ, Clardy J (2004). Long-chain N-acyltyrosine synthases from environmental DNA. *Appl Environ Microbiol* **70:** 6865-6870.

Brazelton WJ, Baross JA (2009). Abundant transposases encoded by the metagenome of a hydrothermal chimney biofilm. *ISME J* **3:** 1420-1424.

Britschgi TB, Giovannoni SJ (1991). Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing. *Appl Environ Microbiol* **57:** 1707-1713.

Brown SP, Buckling A (2008). A social life for discerning microbes. *Cell* **135:** 600-603.

Bruggemann H, Cazalet C, Buchrieser C (2006). Adaptation of Legionella pneumophila to the host environment: role of protein secretion, effectors and eukaryotic-like proteins. *Current opinion in microbiology* **9:** 86-94.

Brzoska P, Rimmele M, Brzostek K, Boos W (1994). The pho regulon-dependent Ugp uptake system for glycerol-3-phosphate in *Escherichia coli* is trans inhibited by Pi. *Journal of Bacteriology* **176:** 15-20.

Burnum KE, Callister SJ, Nicora CD, Purvine SO, Hugenholtz P, Warnecke F *et al.* (2011). Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *ISME J* **5:** 161-164.

Button DK, Robertson B (2000). Effect of Nutrient Kinetics and Cytoarchitecture on Bacterioplankter Size. *Limnol Oceanogr* **45:** 499-505.

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

Castresana J (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17:** 540-552.

Cazalet C, Jarraud S, Ghavi-Helm Y, Kunst F, Glaser P, Etienne J *et al.* (2008). Multigenome analysis identifies a worldwide distributed epidemic Legionella pneumophila clone that emerged within a highly diverse species. *Genome research* **18:** 431-441.

Cebrian E, Uriz MJ, Garrabou J, Ballesteros E (2011). Sponge mass mortalities in a warming Mediterranean Sea: are cyanobacteria-harboring species worse off? *PLoS One* **6:** e20211.

Chae JC, Kim Y, Kim YC, Zylstra GJ, Kim CK (2000). Genetic structure and functional implication of the fcb gene cluster for hydrolytic dechlorination of 4-chlorobenzoate from *Pseudomonas* sp. DJ-12. *Gene* **258:** 109-116.

Chen J, Ryu S, Gharib SA, Goodlett DR, Schnapp LM (2008). Exploration of the normal human bronchoalveolar lavage fluid proteome. *Proteomics Clin Appl* **2:** 585-595.

Cheshire AC, Wilkinson CR (1991). Modelling the photosynthetic production by sponges on Davies Reef, Great Barrier Reef. *Marine Biology* **109:** 13-18.

Cheshire AC, Wilkinson CR, Seddon S, Westphalen G (1997). Bathymetric and seasonal changes in photosynthesis and respiration of the phototrophic sponge *Phyllospongia lamellosa* in comparison with respiration by the heterotrophic sponge *Ianthella basta* on Davies Reef, Great Barrier Reef. *Marine and Freshwater Research* **48:** 589-599.

Chistoserdova L (2010). Recent progress and new challenges in metagenomics for biotechnology. *Biotechnol Lett* **32:** 1351-1359.

Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P (2006). Toward automatic reconstruction of a highly resolved tree of life. *Science* **311:** 1283-1287.

Click EM, Webster RE (1997). Filamentous phage infection: Required interactions with the TolA protein. *Journal of Bacteriology* **179:** 6464-6471.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al.* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* **37:** 141-145.

Connon SA, Giovannoni SJ (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* **68:** 3878-3885.

Crocetti GR, Banfield JF, Keller J, Bond PL, Blackall LL (2002). Glycogen-accumulating organisms in laboratory-scale and full-scale wastewater treatment processes. *Microbiology* **148:** 3353-3364.

Culman SW, Bukowski R, Gauch HG, Cadillo-Quiroz H, Buckley DH (2009). T-REX: software for the processing and analysis of T-RFLP data. *BMC Bioinformatics* **10:** 171.

D'Andrea LD, Regan L (2003). TPR proteins: the versatile helix. *Trends Biochem Sci* **28:** 655-662.

Dayton PK (1989). Interdecadal variation in an antarctic sponge and its predators from oceanographic climate shifts. *Science* **245:** 1484-1486.

de Boer PAJ, Gerding MA, Ogata Y, Pecora ND, Niki H (2007). The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in E-coli. *Mol Microbiol* **63:** 1008-1025.

de Felipe KS, Glover RT, Charpentier X, Anderson OR, Reyes M, Pericone CD *et al.* (2008). Legionella eukaryotic-like type IV substrates interfere with organelle trafficking. *PLoS Pathog* **4:** e1000117.

de Hoon MJ, Imoto S, Nolan J, Miyano S (2004). Open source clustering software. *Bioinformatics* **20:** 1453-1454.

Dean DC, Bowlus CL, Bourgeois S (1987). Cloning and analysis of the promotor region of the human fibronectin gene. *Proc Natl Acad Sci U S A* **84:** 1876-1880.

Deivanayagam CC, Rich RL, Carson M, Owens RT, Danthuluri S, Bice T *et al.* (2000). Novel fold and assembly of the repetitive B region of the *Staphylococcus aureus* collagen-binding surface protein. *Structure* **8:** 67-78.

Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R *et al.* (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proc Natl Acad Sci U S A* **106:** 16428-16433.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al.* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311:** 496-503.

Denef VJ, VerBerkmoes NC, Shah MB, Abraham P, Lefsrud M, Hettich RL *et al.* (2009). Proteomics-inferred genome typing (PIGT) demonstrates inter-population recombination as a strategy for environmental adaptation. *Environ Microbiol* **11:** 313-325.

Denef VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ, Thomas BC *et al.* (2010). Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. *Proc Natl Acad Sci U S A* **107:** 2383-2390.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al.* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72:** 5069-5072.

Dinnbier U, Limpinsel E, Schmid R, Bakker EP (1988). Transient accumulation of potassium glutamate and its replacement by trehalose during adaptation of growing cells of *Escherichia coli* K-12 to elevated sodium chloride concentrations. *Archives of Microbiology* **150:** 348-357.

Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM *et al.* (2008a). Functional metagenomic profiling of nine biomes. *Nature* **452:** 629-632.

Dinsdale EA, Pantos O, Smriga S, Edwards RA, Angly F, Wegley L *et al.* (2008b). Microbial ecology of four coral atolls in the Northern Line Islands. *PLoS One* **3:** e1584.

Ebel R, Brenzinger M, Kunze A, Gross HJ, Proksch P (1997). Wound activation of protoxins in marine sponge *Aplysina aerophoba*. *J Chem Ecol* **23:** 1451-1462.

Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M *et al.* (2005). Diversity of the human intestinal microbial flora. *Science* **308:** 1635-1638.

Eddy SR (1998). Profile hidden Markov models. *Bioinformatics* **14:** 755-763.

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

Eguchi M, Ostrowski M, Fegatella F, Bowman J, Nichols D, Nishino T *et al.* (2001). *Sphingomonas alaskensis* strain AFO1, an abundant oligotrophic ultramicrobacterium from the North Pacific. *Appl Environ Microbiol* **67:** 4945-4954.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95:** 14863-14868.

Eleutherio EC, Araujo PS, Panek AD (1993). Protective role of trehalose during heat stress in *Saccharomyces cerevisiae*. *Cryobiology* **30:** 591-596.

Elssner T, Engemann C, Baumgart K, Kleber HP (2001). Involvement of coenzyme A esters and two new enzymes, an enoyl-CoA hydratase and a CoA-transferase, in the hydration of crotonobetaine to L-carnitine by *Escherichia coli*. *Biochemistry* **40:** 11140-11148.

Enticknap JJ, Kelly M, Peraud O, Hill RT (2006). Characterization of a culturable alphaproteobacterial symbiont common to many marine sponges and evidence for vertical transmission via sponge larvae. *Appl Environ Microbiol* **72:** 3724-3732.

Erickson KL, Beutler JA, Cardellina IJ, Boyd MR (1997). Salicylihalamides A and B, Novel Cytotoxic Macrolides from the Marine Sponge Haliclona sp. *J Org Chem* **62:** 8188-8192.

Erwin PM, Thacker RW (2007). Incidence and identity of photosynthetic symbionts in Caribbean coral reef sponge assemblages. *J Mar Biol Assoc Uk* **87:** 1683-1692.

Erwin PM, Thacker RW (2008). Phototrophic nutrition and symbiont diversity of two Caribbean sponge-cyanobacteria symbioses. *Mar Ecol-Prog Ser* **362:** 139-147.

Esders TW, Lynn SY (1985). Purification and properties of creatinine iminohydrolase from *Flavobacterium filamentosum*. *J Biol Chem* **260:** 3915-3922.

Ezraty B, Aussel L, Barras F (2005). Methionine sulfoxide reductases in prokaryotes. *Biochim Biophys Acta* **1703:** 221-229.

Fan L, Reynolds D, Liu MY, Stark M, Webster N, Kjelleberg S *et al.* (2012). Surviving as sponge associated miroorganism -- many possibilites on common themes. Submitted.

Farrand SK, Taber HW (1974). Changes in menaquinone concentration during growth and early sporulation in *Bacillus subtilis*. *J Bacteriol* **117:** 324-326.

Faulkner DJ (2001). Marine natural products. *Nat Prod Rep* **18:** 1-49.

Finn RD, Mistry J, Tate J, Coggill P, Heger A, Pollington JE *et al.* (2010). The Pfam protein families database. *Nucleic Acids Res* **38:** D211-222.

Flatt P, Gautschi J, Thacker R, Musafija-Girt M, Crews P, Gerwick W (2005). Identification of the cellular site of polychlorinated peptide biosynthesis in the marine sponge *Dysidea (Lamellodysidea) herbacea* and symbiotic cyanobacterium *Oscillatoria spongeliae* by CARD-FISH analysis. *Marine Biology* **147:** 761-774.

Forward JA, Behrendt MC, Wyborn NR, Cross R, Kelly DJ (1997). TRAP transporters: a new family of periplasmic solute transport systems encoded by the dctPQM genes of *Rhodobacter capsulatus* and by homologs in diverse gram-negative bacteria. *J Bacteriol* **179:** 5482-5493.

Frank JA, Sorensen SJ (2011). Quantitative metagenomic analyses based on average genome size normalization. *Appl Environ Microbiol* **77:** 2513-2521.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al.* (2008). Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci U S A* **105:** 3805-3810.

Friedrich AB, Merkert H, Fendert T, Hacker J, Proksch P, Hentschel U (1999). Microbial diversity in the marine sponge *Aplysina cavernicola* (formerly *Verongia cavernicola*) analyzed by fluorescence *in situ* hybridization (FISH). *Marine Biology* **134:** 461-470.

Friedrich AB, Fischer I, Proksch P, Hacker J, Hentschel U (2001). Temporal variation of the microbial community associated with the mediterranean sponge *Aplysina aerophoba*. *FEMS Microbiol Ecol* **38:** 105-113.

Fukunaga Y, Kurahashi M, Tanaka K, Yanagi K, Yokota A, Harayama S (2006). *Pseudovibrio ascidiaceicola* sp. nov., isolated from ascidians (sea squirts). *International journal of systematic and evolutionary microbiology* **56:** 343-347.

Gatlin CL, Kleemann GR, Hays LG, Link AJ, Yates JR, 3rd (1998). Protein identification at the low femtomole level from silver-stained gels using a new fritless electrospray interface for liquid chromatography-microspray and nanospray mass spectrometry. *Anal Biochem* **263:** 93-101.

Gebhard S, Cook GM (2008). Differential regulation of high-affinity phosphate transport systems of *Mycobacterium smegmatis*: identification of PhnF, a repressor of the phnDCE operon. *J Bacteriol* **190:** 1335-1343.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al.* (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One* **3:** e3042.

Gilbert JA, Thomas S, Cooley NA, Kulakova A, Field D, Booth T *et al.* (2009). Potential for phosphonoacetate utilization by marine bacteria in temperate coastal waters. *Environ Microbiol* **11:** 111-125.

Gilbert JA, Dupont CL (2011). Microbial Metagenomics: Beyond the Genome. *Annu Rev Mar Sci* **3:** 347-371.

Giovannoni S, Stingl U (2007). The importance of culturing bacterioplankton in the 'omics' age. *Nat Rev Microbiol* **5:** 820-826.

Giovannoni SJ, Britschgi TB, Moyer CL, Field KG (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345:** 60-63.

Giovannoni SJ, Bibbs L, Cho JC, Stapels MD, Desiderio R, Vergin KL *et al.* (2005). Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* **438:** 82-85.

Goebl M, Yanagida M (1991). The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem Sci* **16:** 173-177.

Goltsman DS, Denef VJ, Singer SW, VerBerkmoes NC, Lefsrud M, Mueller RS *et al.* (2009). Community genomic and proteomic analyses of chemoautotrophic iron-oxidizing "*Leptospirillum rubarum*" (Group II) and "*Leptospirillum ferrodiazotrophum*" (Group III) bacteria in acid mine drainage biofilms. *Appl Environ Microbiol* **75:** 4599-4615.

Graf J, Ruby EG (1998). Host-derived amino acids support the proliferation of symbiotic bacteria. *Proc Natl Acad Sci U S A* **95:** 1818-1822.

Gribble GW (1999). The diversity of naturally occurring organobromine compounds. *Chem Soc Rev* **28:** 335-346.

Grice EA, Kong HH, Conlan S, Deming CB, Davis J, Young AC *et al.* (2009). Topographical and temporal diversity of the human skin microbiome. *Science* **324:** 1190-1192.

Griffith JK, Baker ME, Rouch DA, Page MG, Skurray RA, Paulsen IT *et al.* (1992). Membrane transport proteins: implications of sequence comparisons. *Curr Opin Cell Biol* **4:** 684-695.

Habyarimana F, Al-Khodor S, Kalia A, Graham JE, Price CT, Garcia MT *et al.* (2008). Role for the Ankyrin eukaryotic-like genes of *Legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ Microbiol* **10:** 1460-1474.

Haft DH, Selengut JD, White O (2003). The TIGRFAMs database of protein families. *Nucleic Acids Res* **31:** 371-373.

Haft DH, Selengut J, Mongodin EF, Nelson KE (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1:** e60.

Hallam SJ, Konstantinidis KT, Putnam N, Schleper C, Watanabe Y, Sugahara J *et al.* (2006a). Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc Natl Acad Sci U S A* **103:** 18296-18301.

Hallam SJ, Mincer TJ, Schleper C, Preston CM, Roberts K, Richardson PM *et al.* (2006b). Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine Crenarchaeota. *PLoS Biol* **4:** e95.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5:** R245-249.

Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68:** 669-685.

Head IM, Saunders JR, Pickup RW (1998). Microbial Evolution, Diversity, and Ecology: A Decade of Ribosomal RNA Analysis of Uncultivated Microorganisms. *Microb Ecol* **35:** 1-21.

Heidelberg KB, Gilbert JA, Joint I (2010). Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microb Biotechnol* **3:** 531-543.

Hengge R (2009). Principles of c-di-GMP signalling in bacteria. *Nature reviews Microbiology* **7:** 263-273.

Hentschel U, Schmid M, Wagner M, Fieseler L, Gernert C, Hacker J (2001). Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges *Aplysina aerophob*a and *Aplysina cavernicola*. *FEMS Microbiol Ecol* **35:** 305-312.

Hentschel U, Hopke J, Horn M, Friedrich AB, Wagner M, Hacker J *et al.* (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl Environ Microbiol* **68:** 4431-4440.

Hentschel U, Usher KM, Taylor MW (2006). Marine sponges as microbial fermenters. *FEMS Microbiol Ecol* **55:** 167-177.

Hewson I, Poretsky RS, Dyhrman ST, Zielinski B, White AE, Tripp HJ *et al.* (2009). Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *ISME J* **3:** 1286-1300.

Hill M, Hill A, Lopez N, Harriott O (2006). Sponge-specific bacterial symbionts in the Caribbean sponge, *Chondrilla nucula* (Demospongiae, Chondrosida). *Marine Biology* **148:** 1221-1230.

Hoffmann F, Larsen O, Thiel V, Rapp HT, Pape T, Michaelis W *et al.* (2005). An anaerobic world in sponges. *Geomicrobiol J* **22:** 1-10.

Hoffmann F, Røy H, Bayer K, Hentschel U, Pfannkuchen M, Brümmer F *et al.* (2008). Oxygen dynamics and transport in the Mediterranean sponge *Aplysina aerophoba*. *Marine Biology* **153:** 1257-1264.

Hoffmann F, Radax R, Woebken D, Holtappels M, Lavik G, Rapp HT *et al.* (2009). Complex nitrogen cycling in the sponge *Geodia barretti*. *Environ Microbiol* **11:** 2228-2243.

Hofmann B, Hecht HJ, Flohe L (2002). Peroxiredoxins. *Biol Chem* **383:** 347-364.

Holmes B, Blanch H (2007). Genus-specific associations of marine sponges with group I crenarchaeotes. *Marine Biology* **150:** 759-772.

Hooper JN, van Soest RWM (2002). *Systema Porifera: a guide to the classification of sponges.* Kluwer Academic/Plenum Publishers: New York.

Huang Y, Lai X, He X, Cao L, Zeng Z, Zhang J *et al.* (2009). Characterization of a deep-sea sediment metagenomic clone that produces water-soluble melanin in *Escherichia coli*. *Mar Biotechnol* **11:** 124-131.

Hugenholtz P, Tyson GW, Blackall LL (2002). Design and evaluation of 16S rRNA-targeted oligonucleotide probes for fluorescence *in situ* hybridization. *Methods Mol Biol* **179:** 29-42.

Hugenholtz P, Tyson GW (2008). Microbiology: metagenomics. *Nature* **455:** 481-483.

Hughes TP, Baird AH, Bellwood DR, Card M, Connolly SR, Folke C *et al.* (2003). Climate change, human impacts, and the resilience of coral reefs. *Science* **301:** 929-933.

Huson DH, Auch AF, Qi J, Schuster SC (2007). MEGAN analysis of metagenomic data. *Genome Res* **17:** 377-386.

Isaacs LT, Kan JJ, Nguyen L, Videau P, Anderson MA, Wright TL *et al.* (2009). Comparison of the Bacterial Communities of Wild and Captive Sponge Clathria prolifera from the Chesapeake Bay. *Mar Biotechnol* **11:** 758-770.

Iturbe-Ormaetxe I, Burke GR, Riegler M, O'Neill SL (2005). Distribution, expression, and motif variability of ankyrin domain genes in *Wolbachia pipientis*. *J Bacteriol* **187:** 5136-5145.

Jeans C, Singer SW, Chan CS, Verberkmoes NC, Shah M, Hettich RL *et al.* (2008). Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community. *ISME J* **2:** 542-550.

Jehmlich N, Kleinsteuber S, Vogt C, Benndorf D, Harms H, Schmidt F *et al.* (2010). Phylogenetic and proteomic analysis of an anaerobic toluene-degrading community. *J Appl Microbiol* **109:** 1937-1945.

Joseph SJ, Hugenholtz P, Sangwan P, Osborne CA, Janssen PH (2003). Laboratory cultivation of widespread and previously uncultured soil bacteria. *Appl Environ Microbiol* **69:** 7210-7215.

Jurado V, Laiz L, Gonzalez JM, Hernandez-Marine M, Valens M, Saiz-Jimenez C (2005). *Phyllobacterium catacumbae* sp. nov., a member of the order 'Rhizobiales' isolated from Roman catacombs. *International journal of systematic and evolutionary microbiology* **55:** 1487-1490.

Kahlert M, Neumann D (1997). Early development of freshwater sponges under the influence of nitrite and pH. *Arch Hydrobiol* **139:** 69-81.

Kan J, Hanson TE, Ginter JM, Wang K, Chen F (2005). Metaproteomic analysis of Chesapeake Bay microbial communities. *Saline Systems* **1:** 7.

Kaneko T, Nakamura Y, Sato S, Asamizu E, Kato T, Sasamoto S *et al.* (2000). Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti*. *DNA Res* **7:** 331-338.

Kappe S, Bruderer T, Gantt S, Fujioka H, Nussenzweig V, Menard R (1999). Conservation of a gliding motility and cell invasion machinery in Apicomplexan parasites. *J Cell Biol* **147:** 937-944.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002). Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74:** 5383-5392.

Kennedy J, Marchesi JR, Dobson AD (2007). Metagenomic approaches to exploit the biotechnological potential of the microbial consortia of marine sponges. *Appl Microbiol Biotechnol* **75:** 11-20.

Kester DR, Duedall IW, Connors DN, Pytkowic.Rm (1967). Preparation of Artificial Seawater. *Limnol Oceanogr* **12:** 176-&.

Kijjoa A, Wattanadilok R, Campos N, Nascimento MS, Pinto M, Herz W (2007). Anticancer activity evaluation of kuanoniamines A and C isolated from the marine sponge *Oceanapia sagittaria*, collected from the Gulf of Thailand. *Mar Drugs* **5:** 6-22.

Kim EY, Oh KH, Lee MH, Kang CH, Oh TK, Yoon JH (2009). Novel cold-adapted alkaline lipase from an intertidal flat metagenome and proposal for a new family of bacterial lipases. *Appl Environ Microbiol* **75:** 257-260.

Kim JM, Shimizu S, Yamada H (1987). Cytosine Deaminase That Hydrolyzes Creatinine to N-Methylhydantoin in Various Cytosine Deaminase-Forming Microorganisms. *Archives of Microbiology* **147:** 58-63.

Klaassens ES, de Vos WM, Vaughan EE (2007). Metaproteomics approach to study the functionality of the microbiota in the human infant gastrointestinal tract. *Appl Environ Microbiol* **73:** 1388-1392.

Klatt CG, Wood JM, Rusch DB, Bateson MM, Hamamura N, Heidelberg JF *et al.* (2011). Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J* **5:** 1262-1278.

Kleber HP (1997). Bacterial carnitine metabolism. *FEMS Microbiol Lett* **147:** 1-9.

Kloppel A, Pfannkuchen M, Putz A, Proksch P, Brummer F (2008). *Ex situ* cultivation of *Aplysina aerophoba* close to *in situ* conditions: ecological, biochemical and histological aspects. *Mar Ecol-Evol Persp* **29:** 259-272.

Knaggs AR (2003). The biosynthesis of shikimate metabolites. *Nat Prod Rep* **20:** 119-136.

Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543-546.

Konstantinidis KT, Braff J, Karl DM, DeLong EF (2009). Comparative metagenomic analysis of a microbial community residing at a depth of 4,000 meters at station ALOHA in the North Pacific subtropical gyre. *Appl Environ Microbiol* **75:** 5345-5355.

Kowalchuk GA, Heidelberg JF, Bailey MJ (2007). The ISME journal: multidisciplinary journal of microbial ecology. *ISME J* **1:** 1-3.

Kunin V, Sorek R, Hugenholtz P (2007). Evolutionary conservation of sequence and secondary structures in CRISPR repeats. *Genome Biol* **8:** R61.

Kunin V, Copeland A, Lapidus A, Mavromatis K, Hugenholtz P (2008). A bioinformatician's guide to metagenomics. *Microbiol Mol Biol Rev* **72:** 557-578, Table of Contents.

Kunze B, Jansen R, Sasse F, Hofle G, Reichenbach H (1995). Chondramides A approximately D, new antifungal and cytostatic depsipeptides from Chondromyces crocatus (myxobacteria). Production, physico-chemical and biological properties. *J Antibiot (Tokyo)* **48:** 1262-1266.

Kuroda A, Rashid MH, Sekiguchi J (1992). Molecular cloning and sequencing of the upstream region of the major Bacillus subtilis autolysin gene: a modifier protein exhibiting sequence homology to the major autolysin and the spoIID product. *J Gen Microbiol* **138:** 1067-1076.

Labbe N, Parent S, Villemur R (2004). *Nitratireductor aquibiodomus* gen. nov., sp. nov., a novel alpha-proteobacterium from the marine denitrification system of the Montreal Biodome (Canada). *International journal of systematic and evolutionary microbiology* **54:** 269-273.

Lacerda CM, Choe LH, Reardon KF (2007). Metaproteomic analysis of a bacterial community response to cadmium exposure. *J Proteome Res* **6:** 1145-1152.

Laemmli UK (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680-685.

Lafferty KD, Porter JW, Ford SE (2004). Are diseases increasing in the ocean? *Annu Rev Ecol Evol S* **35:** 31-54.

Lambert C, Hobley L, Chang CY, Fenton A, Capeness M, Sockett L (2009). A predatory patchwork: membrane and surface structures of *Bdellovibrio bacteriovorus*. *Adv Microb Physiol* **54:** 313-361.

Landfald B, Strom AR (1986). Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. *J Bacteriol* **165:** 849-855.

Lane D (1991). 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M (eds). *Nucleic acid techniques in bacterial systematics.* John Wiley and Sons: Chichester, United Kingdom.

Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S *et al.* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106:** 15527-15533.

Lazarevic V, Whiteson K, Huse S, Hernandez D, Farinelli L, Osteras M *et al.* (2009). Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods* **79:** 266-271.

Lazdunski CJ, Bouveret E, Rigal A, Journet L, Lloubes R, Benedetti H (1998). Colicin import into *Escherichia coli* cells. *Journal of Bacteriology* **180:** 4993-5002.

Lee OO, Chui PY, Wong YH, Pawlik JR, Qian PY (2009). Evidence for vertical transmission of bacterial symbionts from adult to embryo in the Caribbean sponge *Svenzea zeai*. *Appl Environ Microbiol* **75:** 6147-6156.

Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A, Qian PY (2011). Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *ISME J* **5:** 650-664.

Letoffe S, Delepelaire P, Wandersman C (2006). The housekeeping dipeptide permease is the *Escherichia coli* heme transporter and functions with two optional peptide binding proteins. *Proc Natl Acad Sci U S A* **103:** 12891-12896.

Levi C, Porte P (1962). Étude au microscope électronique de l'éponge *Oscarella lobularis* de sa larve amphiblastula. . *Cahiers de Biologie Marine* **3:** 107-315.

Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22:** 1658-1659.

Liu WT, Marsh TL, Cheng H, Forney LJ (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl Environ Microbiol* **63:** 4516-4522.

Liu YH, Yang B, Dong JD, Zhou XF, Yang XW, Lee KJ *et al.* (2009). Proline-Containing Dipeptides from a Marine Sponge of a *Callyspongia* Species. *Helv Chim Acta* **92:** 1112-1117.

Liu Z, Klatt CG, Wood JM, Rusch DB, Ludwig M, Wittekindt N *et al.* (2011). Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J* **5:** 1279-1290.

Lo I, Denef VJ, Verberkmoes NC, Shah MB, Goltsman D, DiBartolo G *et al.* (2007). Strain-resolved community proteomics reveals recombining genomes of acidophilic bacteria. *Nature* **446:** 537-541.

Longeon A, Copp BR, Quevrain E, Roue M, Kientz B, Cresteil T *et al.* (2011). Bioactive indole derivatives from the South Pacific marine sponges Rhopaloeides odorabile and Hyrtios sp. *Mar Drugs* **9:** 879-888.

Lopez-Legentil S, Erwin PM, Pawlik JR, Song B (2010). Effects of sponge bleaching on ammonia-oxidizing Archaea: distribution and relative expression of ammonia monooxygenase genes associated with the barrel sponge *Xestospongia muta*. *Microb Ecol* **60:** 561-571.

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

Ludwig W, Schleifer KH (1999). Phylogeny of Bacteria beyond the 16S rRNA Standard. *Asm News* **65:** 752-757.

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

Luengo JM, Garcia JL, Olivera ER (2001). The phenylacetyl-CoA catabolon: a complex catabolic unit with broad biotechnological applications. *Mol Microbiol* **39:** 1434-1442.

Lux SE, John KM, Bennett V (1990). Analysis of cDNA for human erythrocyte ankyrin indicates a repeated structure with homology to tissue-differentiation and cell-cycle control proteins. *Nature* **344:** 36-42.

Makarova KS, Grishin NV, Shabalina SA, Wolf YI, Koonin EV (2006). A putative RNA-interference-based immune system in prokaryotes: computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct* **1:** 7.

Maldonado M, Sanchez-Tocino L, Navarro C (2010). Recurrent disease outbreaks in corneous demosponges of the genus Ircinia: epidemic incidence and defense mechanisms. *Marine Biology* **157:** 1577-1590.

Mantelin S, Saux MF, Zakhia F, Bena G, Bonneau S, Jeder H *et al.* (2006). Emended description of the genus Phyllobacterium and description of four novel species associated with plant roots: *Phyllobacterium bourgognense* sp. nov., *Phyllobacterium ifriqiyense* sp. nov., *Phyllobacterium leguminum* sp. nov. and *Phyllobacterium brassicacearum* sp. nov. *International journal of systematic and evolutionary microbiology* **56:** 827-839.

Marchler-Bauer A, Lu S, Anderson JB, Chitsaz F, Derbyshire MK, DeWeese-Scott C *et al.* (2011). CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res* **39:** D225-229.

Marger MD, Saier MH, Jr. (1993). A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem Sci* **18:** 13-20.

Markert S, Arndt C, Felbeck H, Becher D, Sievert SM, Hugler M *et al.* (2007). Physiological proteomics of the uncultured endosymbiont of *Riftia pachyptila*. *Science* **315:** 247-250.

Markert S, Gardebrecht A, Felbeck H, Sievert SM, Klose J, Becher D *et al.* (2011). Status quo in physiological proteomics of the uncultured *Riftia pachyptila* endosymbiont. *Proteomics*.

Maron PA, Ranjard L, Mougel C, Lemanceau P (2007). Metaproteomics: a new approach for studying functional microbial ecology. *Microb Ecol* **53:** 486-493.

Mavromatis K, Doyle CK, Lykidis A, Ivanova N, Francino MP, Chain P *et al.* (2006). The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *J Bacteriol* **188:** 4015-4023.

Mergaert J, Boley A, Cnockaert MC, Muller WR, Swings J (2001). Identity and potential functions of heterotrophic bacterial isolates from a continuous-upflow fixed-bed reactor for denitrification of drinking water with bacterial polyester as source of carbon and electron donor. *Syst Appl Microbiol* **24:** 303-310.

Metzker ML (2010). Sequencing technologies - the next generation. *Nat Rev Genet* **11:** 31-46.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al.* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 386.

Mohamed NM, Rao V, Hamann MT, Kelly M, Hill RT (2008). Monitoring bacterial diversity of the marine sponge *Ircinia strobilina* upon transfer into aquaculture. *Appl Environ Microbiol* **74:** 4133-4143.

Mohamed NM, Saito K, Tal Y, Hill RT (2010). Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *ISME J* **4:** 38-48.

Moretti M, Grunau A, Minerdi D, Gehrig P, Roschitzki B, Eberl L *et al.* (2010). A proteomics approach to study synergistic and antagonistic interactions of the fungal-bacterial consortium *Fusarium oxysporum* wild-type MSA 35. *Proteomics* **10:** 3292-3320.

Morris RM, Rappe MS, Connon SA, Vergin KL, Siebold WA, Carlson CA *et al.* (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806-810.

Morris RM, Nunn BL, Frazar C, Goodlett DR, Ting YS, Rocap G (2010). Comparative metaproteomics reveals ocean-scale shifts in microbial nutrient utilization and energy transduction. *ISME J* **4:** 673-685.

Mosavi LK, Cammett TJ, Desrosiers DC, Peng ZY (2004). The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci* **13:** 1435-1448.

Mou X, Sun S, Edwards RA, Hodson RE, Moran MA (2008). Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451:** 708-711.

Mukherjee J, Webster N, Llewellyn LE (2009). Purification and Characterization of a Collagenolytic Enzyme from a Pathogen of the Great Barrier Reef Sponge, *Rhopaloeides odorabile*. *PLoS One* **4**.

Mulligan C, Fischer M, Thomas GH (2011). Tripartite ATP-independent periplasmic (TRAP) transporters in bacteria and archaea. *FEMS Microbiol Rev* **35:** 68-86.

Munro MH, Blunt JW, Dumdei EJ, Hickford SJ, Lill RE, Li S *et al.* (1999). The discovery and development of marine compounds with pharmaceutical potential. *J Biotechnol* **70:** 15-25.

Nadell CD, Xavier JB, Foster KR (2009). The sociobiology of biofilms. *FEMS Microbiol Rev* **33:** 206-224.

Nakamura M, Maruyama IN, Soma M, Kato J, Suzuki H, Horota Y (1983). On the process of cellular division in Escherichia coli: nucleotide sequence of the gene for penicillin-binding protein 3. *Mol Gen Genet* **191:** 1-9.

Nesvizhskii AI, Keller A, Kolker E, Aebersold R (2003). A statistical model for identifying proteins by tandem mass spectrometry. *Anal Chem* **75:** 4646-4658.

Newman DJ, Hill RT (2006). New drugs from marine microbes: the tide is turning. *J Ind Microbiol Biotechnol* **33:** 539-544.

Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA *et al.* (2010). Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. *ISME J* **4:** 1002-1019.

Nguyen T (2011). The Role of Ankyrin Repeat Proteins in the Phagocytosis of Sponge-Associated Bacteria. Honours thesis, University of New South Wales, Sydney.

Nicastro D, Schwartz C, Pierson J, Cho J-C, Giovannoni S, McIntosh J (2006). Three-dimensional Structure of the Tiny Bacterium Pelagibacter ubique Studied by Cryo-electron Tomography. *Microscopy and Microanalysis* **12:** 180-181.

Nichols NN, Harwood CS (1995). Repression of 4-hydroxybenzoate transport and degradation by benzoate: a new layer of regulatory control in the *Pseudomonas putida* beta-ketoadipate pathway. *J Bacteriol* **177:** 7033-7040.

Niu B, Fu L, Sun S, Li W (2010). Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11:** 187.

Noguchi H, Park J, Takagi T (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **34:** 5623-5630.

Noguchi H, Taniguchi T, Itoh T (2008). MetaGeneAnnotator: detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res* **15:** 387-396.

Nora T, Lomma M, Gomez-Valero L, Buchrieser C (2009). Molecular mimicry: an important virulence strategy employed by *Legionella pneumophila* to subvert host functions. *Future Microbiol* **4:** 691-701.

Norte M, Rodriguez ML, Fernandez JJ, Eguren L, Estrada DM (1988). Aplysinadiene and (R,R) 5 [3,5-Dibromo-4-[(2-Oxo-5-Oxazolidinyl)] Methoxyphenyl]-2-Oxazolidinone, 2 Novel Metabolites from *Aplysina aerophoba* - Synthesis of Aplysinadiene. *Tetrahedron* **44:** 4973-4980.

Nummelin H, Merckel MC, Leo JC, Lankinen H, Skurnik M, Goldman A (2004). The Yersinia adhesin YadA collagen-binding domain structure is a novel left-handed parallel beta-roll. *EMBO J* **23:** 701-711.

Olson ER, Dunyak DS, Jurss LM, Poorman RA (1991). Identification and characterization of dppA, an *Escherichia coli* gene encoding a periplasmic dipeptide transport protein. *J Bacteriol* **173:** 234-244.

Olson JB, Lord CC, McCarthy PJ (2000). Improved recoverability of microbial colonies from marine sponge samples. *Microb Ecol* **40:** 139-147.

Olson JB, McCarthy PJ (2005). Associated bacterial communities of two deep-water sponges. *Aquat Microb Ecol* **39:** 47-55.

Olson JB, Kellogg CA (2010). Microbial ecology of corals, sponges, and algae in mesophotic coral environments. *FEMS Microbiol Ecol* **73:** 17-30.

Oren M, Steindler L, Ilan M (2005). Transmission, plasticity and the molecular identification of cyanobacterial symbionts in the Red Sea sponge *Diacarnus erythraenus*. *Marine Biology* **148:** 35-41.

Osinga R, Tramper J, Wijffels RH (1998). Cultivation of marine sponges for metabolite production: applications for biotechnology? *Trends Biotechnol* **16:** 130-134.

Ostrowski M, Cavicchioli R, Blaauw M, Gottschal JC (2001). Specific growth rate plays a critical role in hydrogen peroxide resistance of the marine oligotrophic ultramicrobacterium *Sphingomonas alaskensis* strain RB2256. *Appl Environ Microbiol* **67:** 1292-1299.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al.* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* **33:** 5691-5702.

Pace NR, Stahl DA, Lane DJ, Olsen GJ (1986). The Analysis of Natural Microbial-Populations by Ribosomal-Rna Sequences. *Adv Microb Ecol* **9:** 1-55.

Pace NR (1997). A molecular view of microbial diversity and the biosphere. *Science* **276:** 734-740.

Palenik B, Ren Q, Tai V, Paulsen IT (2009). Coastal Synechococcus metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* **11:** 349-359.

Pan X, Luhrmann A, Satoh A, Laskowski-Arce MA, Roy CR (2008). Ankyrin repeat proteins comprise a diverse family of bacterial type IV effectors. *Science* **320:** 1651-1654.

Paradis E, Claude J, Strimmer K (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* **20:** 289-290.

Park C, Novak JT, Helm RF, Ahn YO, Esen A (2008). Evaluation of the extracellular proteins in full-scale activated sludges. *Water Res* **42:** 3879-3889.

Pedersen M, Saenger P, Fries L (1974). Simple Brominated Phenols in Red Algae. *Phytochemistry* **13:** 2273-2279.

Perry JJ, Shin DS, Getzoff ED, Tainer JA (2010). The structural biochemistry of the superoxide dismutases. *Biochim Biophys Acta* **1804:** 245-262.

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

Pfennig N, Biebl H (1976). *Desulfuromonas acetoxidans* gen. nov. and sp. nov., a new anaerobic, sulfur-reducing, acetate-oxidizing bacterium. *Arch Microbiol* **110:** 3-12.

Piel J (2004). Metabolites from symbiotic bacteria. *Nat Prod Rep* **21:** 519-538.

Plugge CM, Stams AJ (2002). Enrichment of thermophilic syntrophic anaerobic glutamate-degrading consortia using a dialysis membrane reactor. *Microb Ecol* **43:** 378-387.

Poolman B, Konings WN (1993). Secondary solute transport in bacteria. *Biochim Biophys Acta* **1183:** 5-39.

Poretsky RS, Hewson I, Sun S, Allen AE, Zehr JP, Moran MA (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* **11:** 1358-1375.

Powell MJ, Sutton JN, Del Castillo CE, Timperman AI (2005). Marine proteomics: generation of sequence tags for dissolved proteins in seawater using tandem mass spectrometry. *Mar Chem* **95:** 183-198.

Preusser A, Wagner U, Elssner T, Kleber HP (1999). Crotonobetaine reductase from *Escherichia coli* consists of two proteins. *Biochim Biophys Acta* **1431:** 166-178.

Price CT, Al-Khodor S, Al-Quadan T, Abu Kwaik Y (2010). Indispensable role for the eukaryotic-like ankyrin domains of the ankyrin B effector of *Legionella pneumophila* within macrophages and amoebae. *Infect Immun* **78:** 2079-2088.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J *et al.* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35:** 7188-7196.

Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C *et al.* (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464:** 59-65.

Raes J, Korbel JO, Lercher MJ, von Mering C, Bork P (2007). Prediction of effective genome size in metagenomic samples. *Genome Biol* **8:** R10.

Ram RJ, Verberkmoes NC, Thelen MP, Tyson GW, Baker BJ, Blake RC, 2nd *et al.* (2005). Community proteomics of a natural microbial biofilm. *Science* **308:** 1915-1920.

Rappe MS, Connon SA, Vergin KL, Giovannoni SJ (2002). Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418:** 630-633.

Rappe MS, Giovannoni SJ (2003). The uncultured microbial majority. *Annu Rev Microbiol* **57:** 369-394.

Regoli F, Cerrano C, Chierici E, Bompadre S, Bavestrello G (2000). Susceptibility to oxidative stress of the Mediterranean demosponge *Petrosia ficiformis*: role of endosymbionts and solar irradiance. *Marine Biology* **137:** 453-461.

Reiswig HM (1971). Particle feeding in natural populations of three marine demosponges. *The Biological Bulletin* **141:** 568-591.

Reiswig HM (1974). Water transport, respiration and energetics of three tropical marine sponges. *Journal of Experimental Marine Biology and Ecology* **14:** 231-249.

Rendulic S, Jagtap P, Rosinus A, Eppinger M, Baar C, Lanz C *et al.* (2004). A predator unmasked: life cycle of *Bdellovibrio bacteriovorus* from a genomic perspective. *Science* **303:** 689-692.

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

Richardson DJ, Berks BC, Russell DA, Spiro S, Taylor CJ (2001). Functional, biochemical and genetic diversity of prokaryotic nitrate reductases. *Cell Mol Life Sci* **58:** 165-178.

Riesenfeld CS, Schloss PD, Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet* **38:** 525-552.

Roche J (1954). Sur les phosphagenes des esponges. *Comptes Rendus Des Seances De La Societe De Biologie Et De Ses Filiales* **148:** 1541-1543.

Roche J, Vanthoai N, Robin Y (1957). Sur La Presence De Creatine Chez Les Invertebres Et Sa Signification Biologique. *Biochim Biophys Acta* **24:** 514-519.

Rohwer F, Seguritan V, Azam F, Knowlton N (2002). Diversity and distribution of coral-associated bacteria. *Mar Ecol-Prog Ser* **243:** 1-10.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR *et al.* (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66:** 2541-2547.

Rudney JD, Xie H, Rhodus NL, Ondrey FG, Griffin TJ (2010). A metaproteomic analysis of the human salivary microbiota by three-dimensional peptide fractionation and tandem mass spectrometry. *Mol Oral Microbiol* **25:** 38-49.

Rusch DB, Halpern AL, Sutton G, Heidelberg KB, Williamson S, Yooseph S *et al.* (2007). The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. *PLoS Biol* **5:** e77.

Rutzler K (1990). Associations between Caribbean sponges and photosynthetic organisms. In: Rutzler K (ed). *New perspectives in sponge biology: 3rd International Sponge Conferences*. Smithsonian Institute Press: Washington, D.C. pp 544-466.

Saldanha AJ (2004). Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20:** 3246-3248.

Sansom FM, Newton HJ, Crikis S, Cianciotto NP, Cowan PJ, d'Apice AJ *et al.* (2007). A bacterial ecto-triphosphate diphosphohydrolase similar to human CD39 is essential for intracellular multiplication of *Legionella pneumophila*. *Cell Microbiol* **9:** 1922-1935.

Sansom FM, Riedmaier P, Newton HJ, Dunstone MA, Muller CE, Stephan H *et al.* (2008). Enzymatic properties of an ecto-nucleoside triphosphate diphosphohydrolase from *Legionella pneumophila*: substrate specificity and requirement for virulence. *J Biol Chem* **283:** 12909-12918.

Santavy DL, Willenz P, Colwell RR (1990). Phenotypic study of bacteria associated with the caribbean sclerosponge, *Ceratoporella nicholsoni*. *Appl Environ Microbiol* **56:** 1750-1762.

Sara M, Bavestrello G, Cattaneo-Vietti R, Cerrano C (1998). Endosymbiosis in sponges: Relevance for epigenesis and evolution. *Symbiosis* **25:** 57-70.

Schaffer S, Isci N, Zickner B, Durre P (2002). Changes in protein synthesis and identification of proteins specifically induced during solventogenesis in *Clostridium acetobutylicum*. *Electrophoresis* **23:** 110-121.

Schauer K, Rodionov DA, de Reuse H (2008). New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends Biochem Sci* **33:** 330-338.

Schlappy ML, Schottner SI, Lavik G, Kuypers MMM, de Beer D, Hoffmann F (2010). Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology* **157:** 593-602.

Schleper C, DeLong EF, Preston CM, Feldman RA, Wu KY, Swanson RV (1998). Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. *Journal of Bacteriology* **180:** 5003-5009.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al.* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75:** 7537-7541.

Schmidt EW, Obraztsova AY, Davidson SK, Faulkner DJ, Haygood MG (2000). Identification of the antifungal peptide-containing symbiont of the marine sponge *Theonella swinhoei* as a novel delta-proteobacterium, "*Candidatus Entotheonella palauensis*". *Marine Biology* **136:** 969-977.

Schmidt II, Bock E (1997). Anaerobic ammonia oxidation with nitrogen dioxide by *Nitrosomonas eutropha*. *Archives of Microbiology* **167:** 106-111.

Schmidt TM, DeLong EF, Pace NR (1991). Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. *Journal of Bacteriology* **173:** 4371-4378.

Schmitt S, Weisz JB, Lindquist N, Hentschel U (2007). Vertical transmission of a phylogenetically complex microbial consortium in the viviparous sponge *Ircinia felix*. *Appl Environ Microbiol* **73:** 2067-2078.

Schmitt S, Angermeier H, Schiller R, Lindquist N, Hentschel U (2008). Molecular microbial diversity survey of sponge reproductive stages and mechanistic insights into vertical transmission of microbial symbionts. *Appl Environ Microbiol* **74:** 7694-7708.

Schmitz FJ, Gopichand Y (1978). (7e, 13-Xi, 15z)-14,16-Dibromo-7,13,15-Hexadecatrien-5-Ynoic Acid - Novel Dibromo Acetylenic Acid from Marine Sponge *Xestospongia-Muta*. *Tetrahedron Lett* **19:** 3637-3640.

Schneider T, Riedel K (2010). Environmental proteomics: analysis of structure and function of microbial communities. *Proteomics* **10:** 785-798.

Schulze WX, Gleixner G, Kaiser K, Guggenberger G, Mann M, Schulze ED (2005). A proteomic fingerprint of dissolved organic carbon and of soil particles. *Oecologia* **142:** 335-343.

Schut F, de Vries EJ, Gottschal JC, Robertson BR, Harder W, Prins RA *et al.* (1993). Isolation of Typical Marine Bacteria by Dilution Culture: Growth, Maintenance, and Characteristics of Isolates under Laboratory Conditions. *Appl Environ Microbiol* **59:** 2150-2160.

Selvin J, Priya SS, Kiran GS, Thangavelu T, Bai NS (2009). Sponge-associated marine bacteria as indicators of heavy metal pollution. *Microbiol Res* **164:** 352-363.

Sharon I, Battchikova N, Aro EM, Giglione C, Meinnel T, Glaser F *et al.* (2011). Comparative metagenomics of microbial traits within oceanic viral communities. *ISME J* **5:** 1178-1190.

Sharp KH, Eam B, Faulkner DJ, Haygood MG (2007). Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp. *Appl Environ Microbiol* **73:** 622-629.

Shi Y, Tyson GW, DeLong EF (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-269.

Siegl A, Hentschel U (2010). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environmental Microbiology Reports* **2:** 507-513.

Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C *et al.* (2011). Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. *ISME J* **5:** 61-70.

Simister RL, Schmitt S, Taylor MW (2011). Evaluating methods for the preservation and extraction of DNA and RNA for analysis of microbial communities in marine sponges. *Journal of Experimental Marine Biology and Ecology* **397:** 38-43.

Simon C, Daniel R (2009). Achievements and new knowledge unraveled by metagenomic approaches. *Appl Microbiol Biotechnol* **85:** 265-276.

Simon J (2002). Enzymology and bioenergetics of respiratory nitrite ammonification. *FEMS Microbiol Rev* **26:** 285-309.

Singer SW, Chan CS, Zemla A, VerBerkmoes NC, Hwang M, Hettich RL *et al.* (2008). Characterization of cytochrome 579, an unusual cytochrome isolated from an iron-oxidizing microbial community. *Appl Environ Microbiol* **74:** 4454-4462.

Sockett RE (2009). Predatory lifestyle of *Bdellovibrio bacteriovorus*. *Annu Rev Microbiol* **63:** 523-539.

Sogin ML, Morrison HG, Huber JA, Mark Welch D, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc Natl Acad Sci U S A* **103:** 12115-12120.

Sowell SM, Wilhelm LJ, Norbeck AD, Lipton MS, Nicora CD, Barofsky DF *et al.* (2009). Transport functions dominate the SAR11 metaproteome at low-nutrient extremes in the Sargasso Sea. *ISME J* **3:** 93-105.

Sowell SM, Abraham PE, Shah M, Verberkmoes NC, Smith DP, Barofsky DF *et al.* (2011). Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* **5:** 856-865.

Spormann AM (1999). Gliding motility in bacteria: insights from studies of *Myxococcus xanthus*. *Microbiol Mol Biol Rev* **63:** 621-641.

Staley JT, Konopka A (1985). Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol* **39:** 321-346.

Stamatakis A, Hoover P, Rougemont J (2008). A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57:** 758-771.

Steger D, Ettinger-Epstein P, Whalan S, Hentschel U, de Nys R, Wagner M *et al.* (2008). Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environ Microbiol* **10:** 1087-1094.

Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *Journal of Bacteriology* **178:** 591-599.

Steindler L, Ilan M, Beer S (2002). Photosymbiosis in intertidal and subtidal tropical sponges. *Symbiosis* **33:** 263-273.

Steindler L, Huchon D, Avni A, Ilan M (2005). 16S rRNA phylogeny of sponge-associated cyanobacteria. *Appl Environ Microbiol* **71:** 4127-4131.

Stewart FJ, Ottesen EA, DeLong EF (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME J* **4:** 896-907.

Stingl U, Tripp HJ, Giovannoni SJ (2007). Improvements of high-throughput culturing yielded novel SAR11 strains and other abundant marine bacteria from the Oregon coast and the Bermuda Atlantic Time Series study site. *ISME J* **1:** 361-371.

Sugawara E, Nikaido H (1992). Pore-forming activity of OmpA protein of *Escherichia coli*. *J Biol Chem* **267:** 2507-2511.

Sutherland KP, Porter JW, Torres C (2004). Disease and immunity in Caribbean and Indo-Pacific zooxanthellate corals. *Mar Ecol-Prog Ser* **266:** 273-302.

Suzuki H, Koyanagi T, Izuka S, Onishi A, Kumagai H (2005). The yliA, -B, -C, and -D genes of *Escherichia coli* K-12 encode a novel glutathione importer with an ATP-binding cassette. *J Bacteriol* **187:** 5861-5867.

Tang YZ, Dobbs FC (2007). Green autofluorescence in dinoflagellates, diatoms, and other microalgae and its implications for vital staining and morphological studies. *Appl Environ Microbiol* **73:** 2306-2313.

Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV *et al.* (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* **4:** 41.

Taylor MW, Schupp PJ, Dahllof I, Kjelleberg S, Steinberg PD (2004). Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environ Microbiol* **6:** 121-130.

Taylor MW, Schupp PJ, de Nys R, Kjelleberg S, Steinberg PD (2005). Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*. *Environ Microbiol* **7:** 419-433.

Taylor MW, Radax R, Steger D, Wagner M (2007). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71:** 295-347.

Teeling H, Meyerdierks A, Bauer M, Amann R, Glockner FO (2004). Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6:** 938-947.

Teufel R, Mascaraque V, Ismail W, Voss M, Perera J, Eisenreich W *et al.* (2010). Bacterial phenylalanine and phenylacetate catabolic pathway revealed. *Proc Natl Acad Sci U S A* **107:** 14390-14395.

Thacker RW (2005). Impacts of shading on sponge-Cyanobacteria symbioses: A comparison between host-specific and generalist associations. *Integr Comp Biol* **45:** 369-376.

Thomas T, Egan S, Burg D, Ng C, Ting L, Cavicchioli R (2007). Integration of genomics and proteomics into marine microbial ecology. *Marine Ecology Progress Series* **332:** 291-299.

Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A *et al.* (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* **4:** 1557-1567.

Thomashow MF, Rittenberg SC (1978). Intraperiplasmic growth of *Bdellovibrio bacteriovorus* 109J: solubilization of *Escherichia coli* peptidoglycan. *J Bacteriol* **135:** 998-1007.

Toyoda A, Iio W, Mitsumori M, Minato H (2009). Isolation and identification of cellulose-binding proteins from sheep rumen contents. *Appl Environ Microbiol* **75:** 1667-1673.

Tringe SG, von Mering C, Kobayashi A, Salamov AA, Chen K, Chang HW *et al.* (2005). Comparative metagenomics of microbial communities. *Science* **308:** 554-557.

Tripp HJ, Bench SR, Turk KA, Foster RA, Desany BA, Niazi F *et al.* (2010). Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464:** 90-94.

Turnbaugh PJ, Gordon JI (2009). The core gut microbiome, energy balance and obesity. *J Physiol* **587:** 4153-4158.

Turon X, Galera J, Uriz MJ (1997). Clearance rates and aquiferous systems in two sponges with contrasting life-history strategies. *The Journal of Experimental Zoology* **278:** 22-36.

Tyson G, Chapman J, P. H, Allen E, Ram R, Richardson P *et al.* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37-43.

Tyson GW, Lo I, Baker BJ, Allen EE, Hugenholtz P, Banfield JF (2005). Genome-directed isolation of the key nitrogen fixer *Leptospirillum ferrodiazotrophum* sp. nov. from an acidophilic microbial community. *Appl Environ Microbiol* **71:** 6319-6324.

Ullmann R, Gross R, Simon J, Unden G, Kroger A (2000). Transport of C(4)-dicarboxylates in *Wolinella succinogenes*. *J Bacteriol* **182:** 5757-5764.

Unson MD, Holland ND, Faulkner DJ (1994). A Brominated Secondary Metabolite Synthesized by the Cyanobacterial Symbiont of a Marine Sponge and Accumulation of the Crystalline Metabolite in the Sponge Tissue. *Marine Biology* **119:** 1-11.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Usher KM, Kuo J, Fromont J, Sutton DC (2001). Vertical transmission of cyanobacterial symbionts in the marine sponge *Chondrilla australiensis* (Demospongiae). *Hydrobiologia* **461:** 15-23.

Usher KM, Toze S, Fromont J, Kuo J, Sutton DC (2004). A new species of cyanobacterial symbiont from the marine sponge *Chondrilla nucula*. *Symbiosis* **36:** 183-192.

Usher KM, Sutton DC, Toze S, Kuo J, Fromont J (2005). Inter-generational transmission of microbial symbionts in the marine sponge *Chondrilla australiensis* (Demospongiae). *Marine and Freshwater Research* **56:** 125-131.

Utkina NK, Denisenko VA, Scholokova OV, Virovaya MV, Gerasimenko AV, Popov DY *et al.* (2001). Spongiadioxins A and B, two new polybrominated dibenzo-p-dioxins from an Australian marine sponge *Dysidea dendyi*. *J Nat Prod* **64:** 151-153.

Vacelet J, Donadey C (1977). Electron microscope study of the association between some sponges and bacteria. *Journal of Experimental Marine Biology and Ecology* **30:** 301-314.

van Oppen MJH, Leong JA, Gates RD (2009). Coral-virus interactions: A double-edged sword? *Symbiosis* **47:** 1-8.

Vega Thurber R, Willner-Hall D, Rodriguez-Mueller B, Desnues C, Edwards RA, Angly F *et al.* (2009). Metagenomic analysis of stressed coral holobionts. *Environ Microbiol* **11:** 2148-2163.

Venn A, Loram J, Douglas A (2008). Photosynthetic symbioses in animals. *J Exp Bot* **59:** 1069-1080.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

VerBerkmoes NC, Denef VJ, Hettich RL, Banfield JF (2009a). Systems biology: Functional analysis of natural microbial consortia using community proteomics. *Nat Rev Microbiol* **7:** 196-205.

Verberkmoes NC, Russell AL, Shah M, Godzik A, Rosenquist M, Halfvarson J *et al.* (2009b). Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3:** 179-189.

Visick KL, McFall-Ngai MJ (2000). An exclusive contract: specificity in the *Vibrio fischeri-Euprymna scolopes* partnership. *Journal of Bacteriology* **182:** 1779-1787.

Vogel S (1977). Current-induced flow through living sponges in nature. *Proc Natl Acad Sci U S A* **74:** 2069-2071.

Vollmer W, Bertsche U (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in *Escherichia coli*. *Biochim Biophys Acta* **1778:** 1714-1734.

Voth DE, Howe D, Beare PA, Vogel JP, Unsworth N, Samuel JE *et al.* (2009). The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J Bacteriol* **191:** 4232-4242.

Vuilleumier S (1997). Bacterial glutathione S-transferases: what are they good for? *J Bacteriol* **179:** 1431-1441.

Walburger A, Lazdunski C, Corda Y (2002). The Tol/Pal system function requires an interaction between the C-terminal domain of TolA and the N-terminal domain of TolB. *Mol Microbiol* **44:** 695-708.

Waligora AJ, Hennequin C, Mullany P, Bourlioux P, Collignon A, Karjalainen T (2001). Characterization of a cell surface protein of Clostridium difficile with adhesive properties. *Infect Immun* **69:** 2144-2153.

Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al.* (2010). *Nitrosopumilus maritimus* genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine *crenarchaea*. *Proc Natl Acad Sci U S A* **107:** 8818-8823.

Walsh DA, Zaikova E, Howes CG, Song YC, Wright JJ, Tringe SG *et al.* (2009). Metagenome of a versatile chemolithoautotroph from expanding oceanic dead zones. *Science* **326:** 578-582.

Walt A, Kahn ML (2002). The fixA and fixB genes are necessary for anaerobic carnitine reduction in *Escherichia coli*. *J Bacteriol* **184:** 4044-4047.

Wang G (2006). Diversity and biotechnological potential of the sponge-associated microbial consortia. *J Ind Microbiol Biotechnol* **33:** 545-551.

Wang HB, Zhang ZX, Li H, He HB, Fang CX, Zhang AJ *et al.* (2011). Characterization of metaproteomics in crop rhizospheric soil. *J Proteome Res* **10:** 932-940.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73:** 5261-5267.

Ward DM, Weller R, Bateson MM (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature* **345:** 63-65.

Wattanadilok R, Sawangwong P, Rodrigues C, Cidade H, Pinto M, Pinto E *et al.* (2007). Antifungal activity evaluation of the constituents of *Haliclona baeri* and *Haliclona cymaeformis*, collected from the Gulf of Thailand. *Mar Drugs* **5:** 40-51.

Webster N, Hill RT (2007). Vulnerability of marine microbes on the Great Barrier Reef to climate change. In: Johnson HE, Marshall PA (eds). *Climate Change and the Great Barrier Reef.* Great Barrier Reef Marine Park Authority and the Australian Greenhouse Office: Townsville.

Webster NS, Hill RT (2001). The culturable microbial community of the Great Barrier Reef sponge *Rhopaloeides odorabile* is dominated by an alpha-Proteobacterium. *Marine Biology* **138:** 843-851.

Webster NS, Webb RI, Ridd MJ, Hill RT, Negri AP (2001a). The effects of copper on the microbial community of a coral reef sponge. *Environ Microbiol* **3:** 19-31.

Webster NS, Wilson KJ, Blackall LL, Hill RT (2001b). Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl Environ Microbiol* **67:** 434-444.

Webster NS, Negri AP, Webb RI, Hill RT (2002). A spongin-boring alpha-proteobacterium is the etiological agent of disease in the Great Barrier Reef sponge *Rhopaloeides odorabile*. *Mar Ecol-Prog Ser* **232:** 305-309.

Webster NS (2007). Sponge disease: a global threat? *Environ Microbiol* **9:** 1363-1375.

Webster NS, Blackall LL (2008). What do we really know about sponge-microbial symbioses? *ISME J* **3:** 1-3.

Webster NS, Cobb RE, Negri AP (2008a). Temperature thresholds for bacterial symbiosis with a sponge. *ISME J* **2:** 830-842.

Webster NS, Xavier JR, Freckelton M, Motti CA, Cobb R (2008b). Shifts in microbial and chemical patterns within the marine sponge *Aplysina aerophoba* during a disease outbreak. *Environ Microbiol* **10:** 3366-3376.

Webster NS, Taylor MW, Behnam F, Lucker S, Rattei T, Whalan S *et al.* (2009). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol*.

Webster NS, Cobb RE, Soo R, Anthony SL, Battershill CN, Whalan S *et al.* (2011). Bacterial community dynamics in the marine sponge *Rhopaloeides odorabile* under in situ and ex situ cultivation. *Mar Biotechnol (NY)* **13:** 296-304.

Webster NS, Taylor MW (2011). Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol* doi: 10.1111/j.1462-2920.2011.02460.x.

Wehrl M, Steinert M, Hentschel U (2007). Bacterial uptake by the marine sponge *Aplysina aerophoba*. *Microb Ecol* **53:** 355-365.

West SA, Diggle SP, Buckling A, Gardner A, Griffins AS (2007). The social lives of microbes. *Annu Rev Ecol Evol S* **38:** 53-77.

Whipps JM, Hand P, Pink D, Bending GD (2008). Phyllosphere microbiology with special reference to diversity and plant genotype. *J Appl Microbiol* **105:** 1744-1755.

White JR, Nagarajan N, Pop M (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comput Biol* **5:** e1000352.

White RH, Hager LP (1977). Occurrence of Fatty-Acid Chlorohydrins in Jellyfish Lipids. *Biochemistry* **16:** 4944-4948.

Whitman WB, Coleman DC, Wiebe WJ (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95:** 6578-6583.

Wilkins MJ, Verberkmoes NC, Williams KH, Callister SJ, Mouser PJ, Elifantz H *et al.* (2009). Proteogenomic monitoring of Geobacter physiology during stimulated uranium bioremediation. *Appl Environ Microbiol* **75:** 6591-6599.

Wilkinson C (1979a). *Bdellovibrio*-like Parasite of Cyanobacteria symbiotic in Marine Sponges. *Archives of Microbiology* **123:** 101-103.

Wilkinson CR (1978a). Microbial Associations in Sponges .I. Ecology, Physiology and Microbial-Populations of Coral-Reef Sponges. *Marine Biology* **49:** 161-167.

Wilkinson CR (1978b). Microbial associations in sponges. II. Numerical analysis of sponge and water bacterial populations. *Marine Biology* **49:** 169-176.

Wilkinson CR (1979b). Nutrient translocation from symbiotic cyanobacteria to coral reef sponges. In: Levi C, Boury-Esnault N (eds). *Biologie des spongiaires*. Colloques Internationaux du Centre National de la Recherche Scientifique: Paris, France.

Wilkinson CR, Fay P (1979). Nitrogen-Fixation in Coral-Reef Sponges with Symbiotic Cyanobacteria. *Nature* **279:** 527-529.

Wilkinson CR (1980). Nutrient Translocation from Green Algal Symbionts to the Fresh-Water Sponge *Ephydatia-Fluviatilis*. *Hydrobiologia* **75:** 241-250.

Wilkinson CR, Garrone R (1980). Nutrition of marine sponges. Involevement of symbiotic bacteria in the uptake of dissolved carbon. In: Smith DC, Tiffon Y (eds). *Nutrition in the Lower Metazoa.* Pergamon Press,: Oxford, UK. pp 157-161.

Wilkinson CR (1983). Net primary productivity in coral reef sponges. *Science* **219:** 410-412.

Wilkinson CR, Garrone R, Vacelet J (1984). Marine Sponges Discriminate between Food Bacteria and Bacterial Symbionts - Electron-Microscope Autoradiography and *In situ* Evidence. *P Roy Soc Lond B Bio* **220:** 519-528.

Wilkinson CR (1987). Interocean Differences in Size and Nutrition of Coral-Reef Sponge Populations. *Science* **236:** 1654-1657.

Wilkinson CR, Cheshire AC (1989). Patterns in the distribution of sponge population across the central Great Barrier Reef. *Coral Reefs* **8:** 127-134.

Williams RJ (2003). Restriction endonucleases: classification, properties, and applications. *Mol Biotechnol* **23:** 225-243.

Wilmes P, Bond PL (2004). The application of two-dimensional polyacrylamide gel electrophoresis and downstream analyses to a mixed community of prokaryotic microorganisms. *Environ Microbiol* **6:** 911-920.

Wilmes P, Andersson AF, Lefsrud MG, Wexler M, Shah M, Zhang B *et al.* (2008a). Community proteogenomics highlights microbial strain-variant protein expression within activated sludge performing enhanced biological phosphorus removal. *ISME J* **2:** 853-864.

Wilmes P, Wexler M, Bond PL (2008b). Metaproteomics provides functional insight into activated sludge wastewater treatment. *PLoS One* **3:** e1778.

Woese CR (1987). Bacterial evolution. *Microbiol Rev* **51:** 221-271.

Woyke T, Xie G, Copeland A, Gonzalez JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4:** e5299.

Wu M, Sun LV, Vamathevan J, Riegler M, Deboy R, Brownlie JC *et al.* (2004). Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol* **2:** E69.

Wyss M, Kaddurah-Daouk R (2000). Creatine and creatinine metabolism. *Physiol Rev* **80:** 1107-1213.

Xie W, Wang F, Guo L, Chen Z, Sievert SM, Meng J *et al.* (2011). Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. *ISME J* **5:** 414-426.

Yamaguchi A, Udagawa T, Sawai T (1990). Transport of divalent cations with tetracycline as mediated by the transposon Tn10-encoded tetracycline resistance protein. *J Biol Chem* **265:** 4809-4813.

Yung PY, Burke C, Lewis M, Egan S, Kjelleberg S, Thomas T (2009). Phylogenetic screening of a bacterial, metagenomic library using homing endonuclease restriction and marker insertion. *Nucleic Acids Res* **37:** e144.

Yung PY (2010). Metagenomics of the microbial communities associated with the marine sponge *Cymbastela concentrica*. PhD thesis, University of New South Wales, Sydney.

Yung PY, Burke C, Lewis M, Kjelleberg S, Thomas T (2011). Novel antibacterial proteins from the microbial communities associated with the sponge Cymbastela concentrica and the green alga Ulva australis. *Appl Environ Microbiol* **77:** 1512-1515.

Zabriskie TM, Klocke JA, Ireland CM, Marcus AH, Molinski TF, Faulkner DJ. (1986). Jaspamide, a modified peptide from a Jaspis sponge, with insecticidal and antifungal activity. *Journal of the American Chemical Society* **108:** 3123-3124.

Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* **12:** 172-191.

Zhu P, Li Q, Wang G (2008). Unique microbial signatures of the alien Hawaiian marine sponge *Suberites zeteki*. *Microb Ecol* **55:** 406-414.

Zucko J, Dunlap WC, Shick JM, Cullum J, Cercelet F, Amin B *et al.* (2010). Global genome analysis of the shikimic acid pathway reveals greater gene loss in host-associated than in free-living bacteria. *BMC Genomics* **11:** 628.

Appendix

Appendix 1. Proteins identified from the metaproteomic analysis of the microbial community associated with the sponge *C. concentrica*.

See Appendix Table A1 and A2 in the CD for the full detail of protein and peptide identifications.

Protein ID	Protein Accession Name	No. Peptides	Blast	COG	Pfams	CDD	Functional cluster
1	lcl|sequenceName|1108430966485|locusTag|ORF0001	6		COG0039:Malate/lactatedehydrogenases			TCA
2	BBAY41_GBSR4EV01AXUVE.gene_1	38			PF03098.8, Animal haem peroxidase		Stress response
4	BBAY40_F51NMB401C7CF8.gene_1	7				TIGR02734, crtl_fam, phytoene desaturase.	Coenzyme biosynthesis
5	BBAY40_F51NMB401DZQJG.gene_1	17		COG3621, Patatin [General function prediction only].			General function
6	BBAY41_GBSR4EV02H0JUT.gene_1	2			PF00538.12, linker histone H1 and H5 family		Nucleosome structure
9	BBAY40_contig05126.gene_2	4		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
11	BBAY41_contig13737.gene_1	2		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport
12	BBAY40_F51NMB401CISX6.gene_1	9			PF05658.7, Hep_Hag		Repeat protein
14	BBAY41_GBSR4EV02JE9R4.gene_2	10		#N/A			Uncharacterized
15	BBAY42_GBUJSMQ01CR141.gene_1	18		COG0176, MipB, Transaldolase [Carbohydrate transport and metabolism].			Carbohydrate metabolism
16	BBAY42_GBUJSMQ01CW781.gene_1	15		COG0224, AtpG, F0F1-type ATP synthase, gamma subunit [Energy production and conversion].			ATP synthesis coupled proton transport
17	lcl|sequenceName|1108814257657|locusTag|ORF0095	2		COG1704:Uncharacterized ACR			Uncharacterized
18	BBAY40_contig07843.gene_4	67		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
19	BBAY42_contig50386.gene_1	6			PF00102.20, Protein-tyrosine phosphatase		General function
20	BBAY40_contig01885.gene_1	12		COG0050, TufB, GTPases - translation elongation factors [Translation, ribosomal structure and biogenesis].			Translation
21	BBAY42_GBUJSMQ01C7K96.gene_2	4		#N/A			Uncharacterized
23	BBAY42_contig18749.gene_1	6			PF01764.18, Lipase (class 3)		Lipid metabolism
24	BBAY42_GBUJSMQ02GODUT.gene_2	8			PF03895.8, YadA-like C-terminal region		Repeat protein
25	lcl|sequenceName|1108814258041|locusTag|ORF0016	3		COG0459:Chaperonin GroEL(HSP60family)			Protein folding
26	BBAY40_contig16058.gene_1	2		#N/A			Uncharacterized
27	BBAY41_contig09751.gene_1	11		#N/A			Uncharacterized

28	BBAY41_GBSR4EV02HMO14.gene_1	4			PF01067.15, Calpain large subunit, domain III		Signalling
30	BBAY42_GBUJSMQ02FRT29.gene_1	3		COG3203, OmpC, Outer membrane protein (porin) [Cell envelope biogenesis, outer membrane].			Transport
31	BBAY42_GBUJSMQ02JSQ57.gene_1	10			PF01404.12, Ephrin receptor ligand binding domain		Signalling
33	lcl|sequenceName|1106850648422|locusTag|ORF0001	10		COG1028:Dehydrogenases with different specificities (related to short-chain alcoho ldehydrogenases)			Hypothetical
34	lcl|sequenceName|1108383314823|locusTag|ORF0001	5		#N/A			Uncharacterized
36	BBAY42_contig10940.gene_1	26		COG0076, GadB, Glutamate decarboxylase and related PLP-dependent proteins [Amino acid transport and metabolism].			Amino acid metabolism
38	BBAY40_F51NMB401EGTL9.gene_1	2		#N/A			Transport
39	BBAY42_GBUJSMQ02IC6JT.gene_1	9	hypothetical protein BRAFLDRAFT_67691 [Branchiostoma floridae]gb|EEN62449.1| hypothetical protein BRAFLDRAFT_67691 [Branchiostoma floridae]				Hypothetical
40	BBAY42_GBUJSMQ02G8T30.gene_2	6		#N/A			Uncharacterized
41	BBAY41_GBSR4EV02JQDWJ.gene_2	2		#N/A			Uncharacterized
42	BBAY40_contig00476.gene_1	11		COG4663, FcbT1, TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component			Transport
43	BBAY40_contig22406.gene_3	16		COG1744, Med, Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein			Transport
45	lcl|sequenceName|1106850806698|locusTag|ORF0001	16	PREDICTED:similartoneurocalcin[Daniorerio],2.79527E-93"/anno2="Neurocalcin-delta-Gallusgallus(Chicken),2.85306E-94"/anno3="LOC556178;similartoneurocalcin,1.85749E-93"				Signalling
46	BBAY42_GBUJSMQ02FGGV6.gene_1	2		#N/A			Uncharacterized
47	BBAY42_GBUJSMQ01A5UID.gene_1	17		#N/A			Uncharacterized
48	BBAY42_GBUJSMQ01AN4FP.gene_1	2		#N/A			Uncharacterized
50	BBAY40_contig15260.gene_1	20		COG0666, Arp, FOG: Ankyrin repeat [General function prediction only].			Ankyrin repeat
51	BBAY40_contig00255.gene_10	12		COG1744, Med, Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein			Transport
55	BBAY42_GBUJSMQ01BEWM3.gene_1	11			PF05817.7, Oligosaccharyltransferase subunit Ribophorin II		Translation
56	lcl|sequenceName|1108814261193|locusTag|ORF0001	11	putative unknown membrane associated protein [Oceanobactersp.RED65]gi|94427095|gb|EAT12076.1|				Transport
57	lcl|sequenceName|1108457449915|locusTag|ORF0001	6		#N/A			Uncharacterized
58	lcl|sequenceName|1108472215206|locusTag|ORF0002	6	hypotheticalphageprotein[Sodalisglossinidiusstr.'morsitans']gi|85059067|ref|YP_454769.1|hypotheticalproteinSG1089[Sodalisglossinidiusstr.'morsitans']				Hypothetical
60	lcl|sequenceName|1108814259223|locusTag|ORF0001	6	hypotheticalproteinROS217_03820[Roseovariussp.217]gi|85670327|gb|EAQ25188.1|hypotheticalproteinROS217_03820				Hypothetical
61	BBAY40_F51NMB401CWDM4.gene_1	18	conserved hypothetical protein [Xanthomonas fuscans subsp. aurantifolii str. ICPB 10535]gb|EFF47327.1| conserved hypothetical protein [Xanthomonas fuscans subsp. aurantifolii str.				Hypothetical
62	BBAY42_contig05234.gene_1	21	cd01903, Ntn_AC_NAAA, AC_NAAA This conserved domain includes two closely related proteins, acid ceramidase (AC, also known as N-acylsphingosine amidohydrolase), and N-acylethanolamine-hydrolyzing acid amidase (NAAA). AC catalyzes the hydrolysis of ceramide to sphingosine and fatty acid.				Signalling
63	BBAY41_GBSR4EV01EWSOT.gene_1	4		#N/A			Uncharacterized

64	BBAY42_GBUJSMQ01DHNYK.gene_1	8		COG2274, SunT, ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain			Transport
65	BBAY40_F51NMB401CRJYI.gene_1	4		COG3203, OmpC, Outer membrane protein (porin) [Cell envelope biogenesis, outer membrane].			Transport
66	BBAY41_contig04643.gene_1	5		#N/A			Uncharacterized
68	BBAY42_GBUJSMQ01D90ZD.gene_1	10	cd03406, Band_7_3, A subgroup of the band 7 domain of flotillin (reggie) like proteins. This subgroup contains proteins similar to stomatin, prohibitin, flotillin, HlfK/C and podicin.				Signalling
69	lcl|sequenceName|1108814260130|locusTag|ORF0002	6		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain			Transport
70	BBAY42_GBUJSMQ02HVHAQ.gene_1	15			PF00217.12, ATP:guanido phosphotransferase, C-terminal catalytic domain		ATP synthesis coupled proton transport
71	BBAY41_GBSR4EV02IU16O.gene_1	6		#N/A			Uncharacterized
72	BBAY40_contig03099.gene_1	3	hypothetical protein BRAFLDRAFT_122528 [Branchiostoma floridae]gb|EEN68300.1| hypothetical protein BRAFLDRAFT_122528 [Branchiostoma floridae]				Hypothetical
73	BBAY41_GBSR4EV01AYJHB.gene_2	7		#N/A			Uncharacterized
74	BBAY41_GBSR4EV01CIURK.gene_1	2		COG0214, SNZ1, Pyridoxine biosynthesis enzyme [Coenzyme metabolism].			Coenzyme biosynthesis
75	BBAY40_F51NMB402IDX2V.gene_1	2			PF05658.7, Hep_Hag		Repeat protein
76	lcl|sequenceName|1108814260597|locusTag|ORF0001	17		COG0055:F0F1-type ATP synthase beta subunit,7.7E-241"			ATP synthesis coupled proton transport
77	lcl|sequenceName|1108431156874|locusTag|ORF0001	17		COG2036:Histones H3 and H4			Nucleosome structure
78	BBAY41_contig12395.gene_2	21		COG0747, DdpA, ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism].			Transport
81	lcl|sequenceName|1108814262305|locusTag|ORF0001	15		COG2885:Outermembraneproteinandrelatedpeptidoglycan-associated(lipo)proteins,5.0E-40"			Transport
82	lcl|sequenceName|1108814258830|locusTag|ORF0025	2		COG0747:ABC-typedipeptide/oligopeptide/nickeltransportsystems,periplasmiccomponents,3.6E-102			Transport
86	BBAY40_F51NMB401B0D7B.gene_1	7		#N/A			Uncharacterized
88	lcl|sequenceName|1108814258838|locusTag|ORF0067	24		COG2885:Outermembraneproteinandrelatedpeptidoglycan-associated(lipo)proteins,7.2E-28"			Transport
89	BBAY42_GBUJSMQ01D1K1Q.gene_1	5		COG0214, SNZ1, Pyridoxine biosynthesis enzyme [Coenzyme metabolism].			Coenzyme biosynthesis
90	BBAY40_contig04124.gene_1	4		COG4889, COG4889, Predicted helicase [General function prediction only].			Transcription
92	BBAY42_GBUJSMQ01ECYYE.gene_1	8		COG1274, PckA, Phosphoenolpyruvate carboxykinase (GTP) [Energy production and conversion].			Carbohydrate metabolism
93	BBAY40_F51NMB401BPIJS.gene_1	4		#N/A			Uncharacterized
95	BBAY40_contig01153.gene_3	3		#N/A			Uncharacterized
96	BBAY42_contig14405.gene_1	2			PF05724.4, Thiopurine S-methyltransferase (TPMT)		Amino acid metabolism
97	BBAY40_F51NMB402G3TBL.gene_1	18		#N/A			Uncharacterized

98	BBAY42_GBUJSMQ02IQ4PK.gene_1	31			PF00753.20, Metallo-beta-lactamase superfamily		Electron transport
99	lcl|sequenceName|1108814257657|locusTag|ORF0047	31		COG0103:RibosomalproteinS9,2.4E-77"/			Translation
103	lcl|sequenceName|1108814258772|locusTag|ORF0016	13		COG0840:Methyl-acceptingchemotaxisprotein,6.1E-67"			Signalling
104	BBAY41_GBSR4EV02I36PM.gene_1	2			PF03098.8, Animal haem peroxidase		Stress response
105	BBAY40_F51NMB401ADKLF.gene_1	7		#N/A			Uncharacterized
106	BBAY42_GBUJSMQ02IPBZ1.gene_1	4		COG3621, COG3621, Patatin [General function prediction only].			General function
107	BBAY41_GBSR4EV02H4TA2.gene_1	2		COG0055, AtpD, F0F1-type ATP synthase, beta subunit [Energy production and conversion].			ATP synthesis coupled proton transport
109	BBAY42_GBUJSMQ02GQX7T.gene_1	26		#N/A			Uncharacterized
110	lcl|sequenceName|1108430643219|locusTag|ORF0001	52		#N/A			Uncharacterized
111	BBAY41_contig24081.gene_2	2			PF00560.26, Leucine Rich Repeat		Repeat protein
112	BBAY40_F51NMB401DMI2R.gene_1	14		#N/A			Uncharacterized
113	BBAY40_contig24467.gene_7	2		COG0747, DdpA, ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism].			Transport
114	BBAY42_GBUJSMQ01COIXS.gene_1	2		#N/A			Uncharacterized
115	BBAY40_F51NMB402F8WMF.gene_2	4		#N/A			Uncharacterized
116	BBAY42_contig25887.gene_4	10			PF02530.7, Porin subfamily		Transport
117	BBAY41_GBSR4EV02HKHHZ.gene_1	5		#N/A			Uncharacterized
119	BBAY40_contig11103.gene_14	45		COG0747, DdpA, ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism].			Transport
120	BBAY40_contig07676.gene_2	25	hypothetical protein CENSYa_1696 [Cenarchaeum symbiosum A]gb|ABK78312.1| hypothetical protein CENSYa_1696 [Cenarchaeum symbiosum A]				Hypothetical
121	BBAY42_contig14970.gene_1	7		COG4206, BtuB, Outer membrane cobalamin receptor protein [Coenzyme metabolism].			Transport
122	BBAY42_GBUJSMQ01B8YEE.gene_1	6		#N/A			Uncharacterized
123	BBAY40_contig00005.gene_4	4		COG2113, ProX, ABC-type proline/glycine betaine transport systems, periplasmic components			Transport
124	BBAY40_F51NMB401DW239.gene_1	6	cd03039, GST_N_Sigma_like, GST_N family, Class Sigma_like; composed of GSTs belonging to class Sigma and similar proteins, including GSTs from class Mu, Pi and Alpha. GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.				Stress response
125	BBAY40_F51NMB401DL3AX.gene_1	5		#N/A			Uncharacterized
126	BBAY41_GBSR4EV01D96IP.gene_1	4		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
127	BBAY40_contig09308.gene_1	22		#N/A			Uncharacterized
128	BBAY41_GBSR4EV02JH54T.gene_1	6		COG2274, SunT, ABC-type bacteriocin/lantibiotic exporters, contain an N-terminal double-glycine peptidase domain			Transport

131	BBAY40_F51NMB401BUTJK.gene_1	6		#N/A			Uncharacterized
132	BBAY41_GBSR4EV01AJ2CV.gene_1	2		COG1159, Era, GTPase [General function prediction only].			General function
133	lcl|sequenceName|1108814258115|locusTag|ORF0008	14	hypotheticalproteinBpro_5280[Polaromonassp.JS666]gi|91700968|gb|ABE47139.1|hypotheticalproteinBpro_5280[Polaromonassp. JS666],7.01005E-21"				Hypothetical
134	BBAY42_contig44375.gene_2	4		COG0214, SNZ1, Pyridoxine biosynthesis enzyme [Coenzyme metabolism].			Coenzyme biosynthesis
135	BBAY42_GBUJSMQ01DAX0Y.gene_1	4	PRK10933, PRK10933, trehalose-6-phosphate hydrolase; Provisional.				Stress response
137	BBAY40_F51NMB401A3GQN.gene_1	15			PF03098.8, Animal haem peroxidase		Stress response
138	BBAY42_contig03867.gene_3	6			PF06314.4, Acetoacetate decarboxylase (ADC)		General metabolism
139	lcl|sequenceName|1108814257723|locusTag|ORF0011	7	COG0811:Biopolymertransportproteins,"MotA/TolQ/ExbBprotonchannel[gammaproteobacteriumKT71]gi|88699322|gb|EAQ96436.1|MotA/TolQ/ExbBprotonchannel[gammaproteobacteriumKT71],7.34318E-109"/anno3="TonBsystembiopolymertransportcomponent,7.79227E-102"				Transport
140	BBAY40_F51NMB401AV2E2.gene_1	2		#N/A			Uncharacterized
141	lcl|sequenceName|1108814258080|locusTag|ORF0005	8		#N/A			Uncharacterized
143	BBAY41_GBSR4EV01D7GIW.gene_1	2			PF00305.12, Lipoxygenase		Lipid metabolism
144	BBAY40_F51NMB402GWHJF.gene_1	6		#N/A			Uncharacterized
145	BBAY40_F51NMB401CTBOJ.gene_1	6		#N/A			Uncharacterized
146	BBAY41_GBSR4EV01BKPF2.gene_1	5	hypothetical protein Slin_4258 [Spirosoma linguale DSM 74]gb|ADB40242.1| conserved hypothetical protein [Spirosoma linguale DSM 74]				Hypothetical
147	BBAY41_GBSR4EV01B4AGD.gene_1	9		#N/A			Uncharacterized
148	BBAY40_F51NMB401BJ3GX.gene_1	24	dual oxidase 1 [Strongylocentrotus purpuratus]ref|XP_001176709.1| PREDICTED: similar to dual oxidase 1 isoform 2 [Strongylocentrotus purpuratus]				General function
149	BBAY40_F51NMB401BE0EQ.gene_1	2		#N/A			Uncharacterized
150	BBAY40_F51NMB401D80I4.gene_1	2		COG0590, CumB, Cytosine/adenosine deaminases			Nucletotide metabolism
151	lcl|sequenceName|1108814257948|locusTag|ORF0009	6		COG0330:Membraneproteasesubunits,stomatin/prohibitinhomologs,2.4E-70"			Proteolysis
152	BBAY40_F51NMB401B64MF.gene_1	25		#N/A			Uncharacterized
153	BBAY40_contig02689.gene_2	6		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
155	BBAY40_F51NMB401A8AMR.gene_1	2		#N/A			Uncharacterized
156	BBAY42_GBUJSMQ01CQGSR.gene_1	6		#N/A			Uncharacterized
157	BBAY42_GBUJSMQ01EH3SS.gene_1	4		#N/A			Uncharacterized
158	BBAY42_GBUJSMQ02IPW9H.gene_1	21			PF03098.8, Animal haem peroxidase		Stress response
159	BBAY41_contig13100.gene_2	6	COG1744, Med, Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein [General function prediction only].				Transport

160	lcl|sequenceName|1108814245591|locusTag|ORF0001	5	OmpA/MotB [Rhodospirillum rubrum ATCC 11170]emb|CAB99313.1| porin-assoziated protein (PAP) [Rhodospirillum rubrum]gb|ABC24122.1| OmpA/MotB [Rhodospirillum rubrum ATCC 11170]				Transport
161	BBAY41_GBSR4EV01BI3RS.gene_1	19		COG1028, FabG, Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)			Lipid biosynthesis
162	BBAY40_F51NMB402IGEKD.gene_1	2			PF02807.8, ATP:guanido phosphotransferase, N-terminal domain		ATP synthesis coupled proton transport
163	BBAY42_GBUJSMQ02FIT0G.gene_1	14		#N/A			Uncharacterized
164	lcl|sequenceName|1108814238418|locusTag|ORF0001	11	F0F1 ATP synthase subunit alpha [Myxococcus xanthus DK 1622]gb|ABF91104.1| ATP synthase F1, alpha subunit [Myxococcus xanthus DK 1622]				ATP synthesis coupled proton transport
165	BBAY42_contig53441.gene_1	22		#N/A			Uncharacterized
166	lcl|sequenceName|1106850573238|locusTag|ORF0001	4	proteinofunknownfunctionDUF1214[NitrobacterhamburgensisX14]gi|91799705|gb|ABE62080.1|proteinofunknownfunctionDUF1214[NitrobacterhamburgensisX14				Uncharacterized
167	BBAY42_GBUJSMQ02HE38D.gene_1	12	ADP/ATP carrier protein [Haliotis diversicolor]				ATP synthesis coupled proton transport
168	BBAY41_GBSR4EV01D5QFE.gene_1	35		#N/A			Uncharacterized
169	BBAY42_GBUJSMQ02IVUKF.gene_1	4	hypothetical protein [Tuber melanosporum Mel28]emb|CAZ79861.1| unnamed protein product [Tuber melanosporum]				Hypothetical
170	BBAY40_contig22747.gene_12	39		COG4663, FcbT1, TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component			Transport
171	BBAY42_GBUJSMQ02H936A.gene_1	6		COG0571, Rnc, dsRNA-specific ribonuclease [Transcription].			Transcription
172	BBAY41_contig01910.gene_3	6		COG0479, FrdB, Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit			TCA
173	BBAY40_F51NMB401AF3GE.gene_1	2		#N/A			Uncharacterized
174	lcl|sequenceName|1108472235523|locusTag|ORF0001	2	Y4oD[Rhizobiumsp.NGR234]gi|2496724|sp|P55589|Y4OD_RHISNHypothetical23.5kDaproteiny4oDgi|2182554|gb|AAB91797.1|Y4oD[Rhizobiumsp.NGR234],5.15075E-46"/anno2="Hypothetical23.5kDaproteiny4oD-Rhizobiumsp.(strainNGR234),4.02532E-47"				Hypothetical
176	BBAY42_GBUJSMQ02JYS1M.gene_1	10			PF05739.12, SNARE domain		General function
177	BBAY40_F51NMB402HJVNR.gene_1	6		#N/A			Uncharacterized
178	BBAY42_GBUJSMQ02GCUVN.gene_1	4	hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]ref|YP_241226.1| hypothetical protein XC_0115 [Xanthomonas campestris pv. campestris str. 8004]ref|YP_001901525.1| hypothetical protein xccb100_0119 [Xanthomonas campestris pv. campestris str. B100]gb|AAM39429.1| hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]gb|AAY47206.1				Hypothetical
179	BBAY41_GBSR4EV02IAN63.gene_1	4		#N/A			Uncharacterized
180	BBAY41_GBSR4EV02HGJ8Z.gene_1	4		#N/A			Uncharacterized
181	BBAY42_GBUJSMQ02HRRDG.gene_1	2		#N/A			Uncharacterized
183	BBAY40_F51NMB401CIQ5Z.gene_1	12			pfam00017, SH2, SH2 domain.		Signalling
184	BBAY42_contig44260.gene_1	16		COG0702, COG0702, Predicted nucleoside-diphosphate-sugar epimerases			Transport
185	BBAY42_GBUJSMQ02IVQIH.gene_1	2		#N/A			Uncharacterized
186	BBAY41_contig23431.gene_1	3		COG1744, Med, Uncharacterized ABC-type transport system, periplasmic component/surface lipoprotein			Transport

188	BBAY40_contig00081.gene_1	12		COG1638, DctP, TRAP-type C4-dicarboxylate transport system, periplasmic component			Transport
189	BBAY41_GBSR4EV02FODEH.gene_1	4		#N/A			Uncharacterized
190	BBAY40_F51NMB401D0MJY.gene_1	14		COG0515, SPS1, Serine/threonine protein kinase			Signalling
191	BBAY40_F51NMB401E1Y98.gene_1	5		#N/A			Uncharacterized
192	BBAY42_GBUJSMQ01DPL55.gene_1	6			PF00071.15, Ras family		Signalling
194	BBAY42_GBUJSMQ02H11R2.gene_2	9		#N/A			Uncharacterized
195	BBAY42_GBUJSMQ02G19KN.gene_1	7		#N/A			Uncharacterized
196	BBAY40_contig19464.gene_11	19		COG1638, DctP, TRAP-type C4-dicarboxylate transport system, periplasmic component			Transport
198	BBAY42_contig01089.gene_1	8		#N/A			Uncharacterized
199	BBAY40_F51NMB401BKT9D.gene_1	4		#N/A			Uncharacterized
200	BBAY40_F51NMB401EL24X.gene_1	3		#N/A			Uncharacterized
202	BBAY40_contig10228.gene_1	9		COG0605, SodA, Superoxide dismutase [Inorganic ion transport and metabolism].			Stress response
204	lcl|sequenceName|1108814258486|locusTag|ORF0014	2		#N/A			Uncharacterized
207	BBAY42_GBUJSMQ02ISOX4.gene_1	18		#N/A			Uncharacterized
208	lcl|sequenceName|1108385086626|locusTag|ORF0002	5		#N/A			Uncharacterized
209	lcl|sequenceName|1108814258328|locusTag|ORF0010	5		COG1629:Outermembranereceptorproteins,mostly Fetransport			Transport
210	BBAY40_F51NMB402HDD9L.gene_1	2		#N/A			Uncharacterized
212	BBAY40_contig14943.gene_1	6			PF09459.3, Ethylbenzene dehydrogenase		Nitrate reduction
213	BBAY41_GBSR4EV02FJUE6.gene_1	2			PF07715.8, TonB-dependent Receptor Plug Domain		Transport
214	BBAY40_contig07886.gene_4	5		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
215	BBAY40_contig06170.gene_1	3		COG1652, XkdP, Uncharacterized protein containing LysM domain [Function unknown].			Uncharacterized
216	BBAY42_GBUJSMQ02GP6Y5.gene_1	9		#N/A			Uncharacterized
218	BBAY40_contig17557.gene_1	2		COG4771, FepA, Outer membrane receptor for ferrienterochelin and colicins [Inorganic ion transport and metabolism].			Transport
219	BBAY42_GBUJSMQ01AVX7F.gene_1	8		COG0538, Icd, Isocitrate dehydrogenases [Energy production and conversion].			General metabolism
220	BBAY41_contig09582.gene_3	74		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain			Transport
221	BBAY40_contig08171.gene_14	9		COG0443, DnaK, Molecular chaperone [Posttranslational modification, protein turnover, chaperones].			Protein folding

223	BBAY42_GBUJSMQ01DMW9M.gene_1	4		#N/A			Uncharacterized
224	lcl|sequenceName|1106850535982|locusTag|ORF0001	10		#N/A			Uncharacterized
225	BBAY40_F51NMB402F8T39.gene_1	7		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
226	BBAY41_GBSR4EV02JE0NW.gene_1	2		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain			Transport
228	BBAY42_GBUJSMQ02HZWL8.gene_2	2		COG0443, DnaK, Molecular chaperone [Posttranslational modification, protein turnover, chaperones].			Protein folding
229	lcl|sequenceName|1106851064362|locusTag|ORF0001	13		#N/A			Uncharacterized
230	BBAY41_contig01209.gene_2	2		COG2113, ProX, ABC-type proline/glycine betaine transport systems, periplasmic components			Transport
232	BBAY41_contig13368.gene_1	15		COG0747, DdpA, ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism].			Transport
233	lcl|sequenceName|1108814238677|locusTag|ORF0001	15	molybdopterin dehydrogenase FAD-binding [Desulfovibrio sp. FW1012B]gb|EFC19207.1| molybdopterin dehydrogenase FAD-binding [Desulfovibrio sp. FW1012B]				Electron transport
234	BBAY41_contig00310.gene_4	9		COG1638, DctP, TRAP-type C4-dicarboxylate transport system, periplasmic component			Transport
235	lcl|sequenceName|1108814248139|locusTag|ORF0004	2	chaperonin GroEL [Sphingobium chlorophenolicum L-1]				Protein folding
236	BBAY42_contig10537.gene_2	2		COG0581, PstA, ABC-type phosphate transport system, permease component [Inorganic ion transport and metabolism].			Transport
237	BBAY40_contig00384.gene_1	6		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport
238	BBAY40_contig00237.gene_1	5		#N/A			Uncharacterized
239	BBAY40_F51NMB402IH9JJ.gene_1	5		#N/A			Uncharacterized
240	BBAY42_contig04577.gene_1	11		COG1629, CirA, Outer membrane receptor proteins, mostly Fe transport [Inorganic ion transport and metabolism].			Transport
241	BBAY41_contig22397.gene_11	10		COG0687, PotD, Spermidine/putrescine-binding periplasmic protein [Amino acid transport and metabolism].			Transport
243	BBAY42_contig44679.gene_15	42		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
244	BBAY42_GBUJSMQ01CKTZP.gene_1	7	PRK05899, PRK05899, transketolase; Reviewed.				Carbohydrate metabolism
245	BBAY42_contig05002.gene_1	4		COG1132, MdlB, ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms].			Transport
246	lcl|sequenceName|1108814257736|locusTag|ORF0005	4		COG0056:F0F1-typeATPsynthasealphasubunit,0.0"			ATP synthesis coupled proton transport
247	lcl|sequenceName|1108814258375|locusTag|ORF0007	3		COG1629:Outermembranereceptorproteins,mostlyFetransport,1.5E-28"			Transport
248	BBAY41_contig15022.gene_25	2				PF02384.9, N-6 DNA Methylase	Transcription
249	lcl|sequenceName|1108814257319|locusTag|ORF0040	4		COG0459:ChaperoninGroEL(HSP60family),0.0"			Protein folding
250	BBAY42_GBUJSMQ01CL8GY.gene_1	26		COG3288, PntA, NAD/NADP transhydrogenase alpha subunit [Energy production and conversion].			Electron transport
251	BBAY41_GBSR4EV02GL5LO.gene_1	7		COG0055, AtpD, F0F1-type ATP synthase, beta subunit [Energy production and conversion].			ATP synthesis coupled proton transport

252	BBAY40_F51NMB402GR25P.gene_1	11		COG5126, FRQ1, Ca2+-binding protein (EF-Hand superfamily)			Signalling
253	BBAY42_GBUJSMQ02HMLLG.gene_1	2		COG1132, MdlB, ABC-type multidrug transport system, ATPase and permease components [Defense mechanisms].			Transport
254	lcl|sequenceName|1108814258114|locusTag|ORF0001	26		COG1629:Outermembranereceptorproteins,mostlyFetransport,1.8E-50"/			Transport
255	BBAY40_contig09017.gene_5	13		COG1879, RbsB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
256	BBAY40_contig08746.gene_6	19		COG0450, AhpC, Peroxiredoxin [Posttranslational modification, protein turnover, chaperones].			Stress response
257	BBAY42_contig11017.gene_1	7		COG0656, ARA1, Aldo/keto reductases, related to diketogulonate reductase [General function prediction only].			General function
258	lcl|sequenceName|1108814239459|locusTag|ORF0003	27	aldehyde oxidase and xanthine dehydrogenase molybdopterin binding [Desulfovibrio sp. FW1012B]gb|EFC19208.1| aldehyde oxidase and xanthine dehydrogenase molybdopterin binding [Desulfovibrio sp. FW1012B]				General function
259	lcl|sequenceName|1108814239459|locusTag|ORF0002	8	molybdopterin dehydrogenase FAD-binding [Desulfovibrio sp. FW1012B]gb|EFC19207.1| molybdopterin dehydrogenase FAD-binding [Desulfovibrio sp. FW1012B]				Electron transport
260	BBAY40_contig11789.gene_1	6		#N/A			Uncharacterized
262	lcl|sequenceName|1108814258804|locusTag|ORF0055	2		COG1538:Outermembraneprotein,1.2E-68"/anno5="OutermembraneproteintolCprecursor-Vibriocholerae,3.3547E-47"			Transport
263	BBAY41_GBSR4EV01DLXB3.gene_1	6		#N/A			Uncharacterized
264	lcl|sequenceName|1106850404455|locusTag|ORF0001	6	putativehaemagglutinin/invasin,2.78089E7"/anno2="outermembraneprotein[Erythrobactersp.NAP1]gi|85689536|gb|EAQ29539.1|outermembraneprotein[Erythrobacte rsp.NAP1],2.22413E-10"				Repeat protein
266	BBAY42_GBUJSMQ02JKUFE.gene_1	9	hypothetical protein NB231_01653 [Nitrococcus mobilis Nb-231]gb|EAR20515.1| hypothetical protein NB231_01653 [Nitrococcus mobilis Nb-231]				Hypothetical
267	BBAY42_GBUJSMQ02GRATV.gene_1	21		#N/A			Uncharacterized
268	BBAY42_GBUJSMQ02I8618.gene_1	29	hypothetical protein Slin_4258 [Spirosoma linguale DSM 74]gb|ADB40242.1| conserved hypothetical protein [Spirosoma linguale DSM 74]				Hypothetical
269	lcl|sequenceName|1108814240168|locusTag|ORF0008	6	TonB-dependent receptor domain protein [marine gamma proteobacterium HTCC2148]gb|EEB77521.1| TonB-dependent receptor domain protein				Transport
271	BBAY42_contig29876.gene_1	8		COG0451, WcaG, Nucleoside-diphosphate-sugar epimerases			Transport
272	BBAY40_F51NMB402G4KXR.gene_2	9		#N/A			Uncharacterized
273	BBAY41_GBSR4EV01CCPQ0.gene_1	4		#N/A			Uncharacterized
274	BBAY40_contig00255.gene_6	7		COG2358, Imp, TRAP-type uncharacterized transport system, periplasmic component [General function prediction only].			Transport
275	lcl|sequenceName|1108814239595|locusTag|ORF0001	14	putative porin [Aurantimonas manganoxydans SI85-9A1]gb|EAS48596.1| putative porin [Aurantimonas manganoxydans SI85-9A1]				Transport
277	BBAY40_contig11418.gene_1	18		COG4663, FcbT1, TRAP-type mannitol/chloroaromatic compound transport system, periplasmic component			Transport
278	BBAY42_GBUJSMQ01AX5D7.gene_2	13			PF00305.12, Lipoxygenase		Lipid biosynthesis
279	BBAY40_contig19696.gene_1	7	PRK05035, PRK05035, electron transport complex protein RnfC; Provisional.				Electron transport
280	BBAY42_GBUJSMQ01DOG5K.gene_1	2		#N/A			Uncharacterized
281	BBAY41_GBSR4EV02IV8MD.gene_1	40	hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]ref|YP_241226.1| hypothetical protein XC_0115 [Xanthomonas campestris pv. campestris str. 8004]ref|YP_001901525.1| hypothetical protein xccb100_0119 [Xanthomonas campestris pv. campestris str. B100]gb|AAM39429.1| hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]gb|AAY47206.1				Hypothetical

282	BBAY42_GBUJSMQ01ECAA0.gene_1	34	hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]ref|YP_241226.1| hypothetical protein XC_0115 [Xanthomonas campestris pv. campestris str. 8004]ref|YP_001901525.1| hypothetical protein xccb100_0119 [Xanthomonas campestris pv. campestris str. B100]gb|AAM39429.1| hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]gb|AAY47206.1				Hypothetical
283	BBAY42_GBUJSMQ02HGXPF.gene_1	2		#N/A			Uncharacterized
284	BBAY42_GBUJSMQ01E1DC7.gene_1	2		#N/A			Uncharacterized
285	BBAY40_F51NMB402GL73L.gene_1	14		COG2226, UbiE, Methylase involved in ubiquinone/menaquinone biosynthesis [Coenzyme metabolism].			Coenzyme biosynthesis
286	BBAY42_GBUJSMQ02F79CH.gene_1	2		#N/A			Uncharacterized
287	BBAY41_GBSR4EV01BSRWE.gene_1	2			PF02157.8, Cation-dependent mannose-6-phosphate receptor		Transport
288	BBAY40_F51NMB401B5FLK.gene_1	3			PF00672.18, HAMP domain		Signalling
289	BBAY41_GBSR4EV02HAIZM.gene_1	20		#N/A			Uncharacterized
290	BBAY40_F51NMB402IAP4J.gene_1	4		#N/A			Uncharacterized
291	BBAY42_GBUJSMQ01DB72D.gene_1	24	cd00249, AGE, AGE domain; N-acyl-D-glucosamine 2-epimerase domain; Responsible for intermediate epimerization during biosynthesis of N-acetylneuraminic acid. Catalytic mechanism is believed to be via nucleotide elimination and readdition and is ATP modulated.				General metabolism
292	BBAY40_F51NMB402H6123.gene_1	4		#N/A			Uncharacterized
293	lcl|sequenceName|1108814258832|locusTag|ORF0354	6		Q2G5N5|ATPB_NOVAD ATP synthase subunit beta			ATP synthesis coupled proton transport
295	BBAY41_contig21595.gene_5	26			PF02530.7, Porin subfamily		Transport
299	BBAY42_GBUJSMQ02F45EO.gene_1	6			PF02450.8, Lecithin:cholesterol acyltransferase		General metabolism
301	BBAY42_GBUJSMQ01D8U85.gene_1	2		COG1574, COG1574, Predicted metal-dependent hydrolase with the TIM-barrel fold [General function prediction only].			General function
302	BBAY41_GBSR4EV02IYO8X.gene_1	10		COG0435, ECM4, Predicted glutathione S-transferase [Posttranslational modification, protein turnover, chaperones].			Stress response
303	BBAY40_F51NMB402IK93J.gene_1	7		#N/A			Uncharacterized
304	BBAY42_GBUJSMQ01DEHQO.gene_1	6			PF11932.1, Protein of unknown function (DUF3450)		Transport
305	BBAY42_contig39467.gene_1	2		#N/A			Uncharacterized
306	lcl|sequenceName|1108814240260|locusTag|ORF0001	6	hypothetical protein ELI_04080 [Erythrobacter litoralis HTCC2594]gb|ABC62908.1| hypothetical protein ELI_04080 [Erythrobacter litoralis HTCC2594]				Hypothetical
307	BBAY41_contig03454.gene_1	16			PF00036.25, EF hand		Signalling
308	BBAY42_GBUJSMQ02H6KOM.gene_1	7		#N/A			Uncharacterized
309	lcl|sequenceName|1108814258799|locusTag|ORF0004	3	hypotheticalproteinEBAC000-63A02.42[unculturedbacterium442],5.21071E-120"				Hypothetical
311	BBAY42_GBUJSMQ01CH5TK.gene_1	9		COG0022, AcoB, Pyruvate/2-oxoglutarate dehydrogenase complex, dehydrogenase (E1) component, eukaryotic type, beta subunit [Energy production and conversion].			TCA
312	BBAY42_contig35640.gene_2	25	conserved hypothetical protein, secreted [Candidatus Poribacteria sp. WGA-A3]gb|EFC34519.1| conserved hypothetical protein, secreted [Candidatus Poribacteria sp. WGA-A3]				Hypothetical

313	BBAY42_GBUJSMQ02JUDI5.gene_1	2		#N/A			Uncharacterized
316	BBAY40_contig08910.gene_9	16		COG1879, RbsB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
317	lcl|sequenceName|1108814235370|locusTag|ORF0002	7	HflC protein [marine gamma proteobacterium HTCC2148]gb|AAS07891.1| HflC protein [uncultured marine bacterium 463]gb|EEB79428.1| HflC protein [marine gamma proteobacterium HTCC2148]				Proteolysis
318	lcl|sequenceName|1108457275420|locusTag|ORF0002	6		COG1132:ABC-typemultidrug/protein/lipidtransportsystem,ATPasecomponent,7.2E-36"			Transport
319	BBAY41_contig00127.gene_4	14		COG2113, ProX, ABC-type proline/glycine betaine transport systems, periplasmic components			Transport
320	BBAY42_GBUJSMQ01DMFI3.gene_2	4			PF02450.8, Lecithin:cholesterol acyltransferase		General metabolism
322	BBAY40_F51NMB402JT9W5.gene_2	13		COG1028, FabG, Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)			Lipid biosynthesis
323	BBAY42_contig38217.gene_5	7	hypothetical protein NIDE2749 [Candidatus Nitrospira defluvii]emb|CBK42455.1| exported protein of unknown function [Candidatus Nitrospira defluvii]				Hypothetical
324	BBAY42_GBUJSMQ02GJTSU.gene_1	7			PF01593.17, Flavin containing amine oxidoreductase		Ammonia related
326	BBAY41_contig00671.gene_1	5		#N/A			Uncharacterized
329	BBAY40_F51NMB402JVEWQ.gene_1	2			PF07715.8, TonB-dependent Receptor Plug Domain		Transport
333	lcl|sequenceName|1108458138393|locusTag|ORF0001	4	aggregationfactorprotein3,formD[Microcionaprolifera],3.57147E-27"				Repeat protein
334	BBAY40_contig14445.gene_2	4		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
335	BBAY42_GBUJSMQ01A413M.gene_1	13	PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]ref|XP_001194750.1| PREDICTED: hypothetical protein [Strongylocentrotus purpuratus]				Hypothetical
337	lcl|sequenceName|1108457233736|locusTag|ORF0001	9		#N/A			Uncharacterized
338	lcl|sequenceName|1106850427075|locusTag|ORF0001	7		#N/A			Uncharacterized
339	BBAY41_GBSR4EV01EETCQ.gene_1	6		COG0056, AtpA, F0F1-type ATP synthase, alpha subunit [Energy production and conversion].			ATP synthesis coupled proton transport
340	BBAY42_GBUJSMQ01B8RWS.gene_1	8		#N/A			Uncharacterized
341	lcl|sequenceName|1108814239573|locusTag|ORF0001	8	amino acid ABC transporter, ATP-binding protein [uncultured marine bacterium Ant4D3]				Transport
342	BBAY41_GBSR4EV02IR1OR.gene_1	12	hypothetical protein XCC0110 [Xanthomonas campestris pv. campestris str. ATCC 33913]ref|YP_241226.1| hypothetical protein XC_0115 [Xanthomonas campestris pv. campestris str. 8004]ref|YP_001901525.1| hypothetical protein xccb100_0119 [Xanthomonas campestris pv. campestris str. B100]				Hypothetical
344	BBAY40_F51NMB402FGQUQ.gene_1	2			PF06980.4, Protein of unknown function (DUF1302)		Uncharacterized
345	BBAY40_F51NMB402HEAPI.gene_1	4		#N/A			Uncharacterized
346	BBAY42_contig39393.gene_1	23	cd00249, AGE, AGE domain; N-acyl-D-glucosamine 2-epimerase domain; Responsible for intermediate epimerization during biosynthesis of N-acetylneuraminic acid. Catalytic mechanism is believed to be via nucleotide elimination and readdition and is ATP modulated.				General metabolism
347	BBAY42_GBUJSMQ01DQVIU.gene_1	10			PF00305.12, Lipoxygenase		Lipid biosynthesis
348	BBAY42_GBUJSMQ01AOE7S.gene_1	2		COG0174, GlnA, Glutamine synthetase [Amino acid transport and metabolism].			Ammonia related

349	BBAY41_GBSR4EV01CXCKV.gene_1	5	N-acyl-D-glucosamine 2-epimerase [Syntrophobacter fumaroxidans MPOB]gb|ABK16058.1| N-acyl-D-glucosamine 2-epimerase [Syntrophobacter fumaroxidans MPOB]				Carbohydrate metabolism
350	lcl|sequenceName|1108814234517|locusTag|ORF0012	2	gamma-glutamyl phosphate reductase [marine gamma proteobacterium HTCC2207]gb|EAS47156.1| gamma-glutamyl phosphate reductase [marine gamma proteobacterium HTCC2207]				Amino acid biosynthesis
351	lcl|sequenceName|1108814258830|locusTag|ORF0133	32		COG0747:ABC-typedipeptide/oligopeptide/nickeltransportsystems,periplasmiccomponents,2.8E-34"/			Transport
352	BBAY40_contig08314.gene_2	8		COG1879, RbsB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
353	BBAY40_contig23972.gene_3	13		COG2113, ProX, ABC-type proline/glycine betaine transport systems, periplasmic components			Transport
354	BBAY41_GBSR4EV01B6TJW.gene_1	24		COG4122, COG4122, Predicted O-methyltransferase [General function prediction only].			Stress response
355	BBAY41_GBSR4EV01BF0P9.gene_1	13	Ca2+-triggered coelenterazine-binding protein 2 [Renilla muelleri]				General function
356	BBAY40_F51NMB401AO4WP.gene_1	4		#N/A			Uncharacterized
357	BBAY41_GBSR4EV01BOEL6.gene_2	2		#N/A			Amino acid biosynthesis
359	BBAY42_GBUJSMQ02H0O4N.gene_1	10		#N/A			Uncharacterized
361	BBAY40_contig01192.gene_1	4		COG0554, GlpK, Glycerol kinase [Energy production and conversion].			Lipid metabolism
362	BBAY42_GBUJSMQ02I15UA.gene_1	6	dual oxidase 1 [Lytechinus variegatus]				General function
363	BBAY41_GBSR4EV01BB0KU.gene_1	10		#N/A			Uncharacterized
364	BBAY40_F51NMB402GZFBU.gene_1	4			PF01593.17, Flavin containing amine oxidoreductase		Ammonia related
366	BBAY42_GBUJSMQ01ESNUH.gene_1	9		#N/A			Uncharacterized
367	BBAY40_F51NMB401BLTTQ.gene_1	2		#N/A			Uncharacterized
368	BBAY42_GBUJSMQ01DAJLJ.gene_1	2		#N/A			Uncharacterized
369	BBAY42_GBUJSMQ01EDK68.gene_1	4			PF02798.13, Glutathione S-transferase, N-terminal domain		Signalling
370	BBAY40_contig10450.gene_5	5		COG1638, DctP, TRAP-type C4-dicarboxylate transport system, periplasmic component			Transport
371	BBAY42_GBUJSMQ01BXT3M.gene_1	6			PF00036.25, EF hand		Signalling
372	BBAY40_contig10771.gene_12	11		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
373	BBAY40_contig16508.gene_1	11			PF00106.18, short chain dehydrogenase		General function
376	BBAY41_GBSR4EV02FQCTR.gene_2	2		#N/A			Uncharacterized
378	BBAY40_F51NMB401CSUJU.gene_1	2		#N/A			Uncharacterized
379	BBAY40_contig08449.gene_2	6		COG1960, CaiA, Acyl-CoA dehydrogenases [Lipid metabolism].			Lipid metabolism
380	BBAY42_GBUJSMQ02JW4Z5.gene_1	2		COG0013, AlaS, Alanyl-tRNA synthetase [Translation, ribosomal structure and biogenesis].			Translation

381	BBAY41_GBSR4EV02GEC3O.gene_1	25		#N/A			Uncharacterized
382	BBAY40_F51NMB402FPWG7.gene_1	10	PREDICTED: hypothetical protein [Hydra magnipapillata]				Hypothetical
383	BBAY42_GBUJSMQ01CZ6RZ.gene_1	8	cd03134, GATase1_PfpI_like, A type 1 glutamine amidotransferase (GATase1)-like domain found in PfpI from Pyrococcus furiosus. A type 1 glutamine amidotransferase (GATase1)-like domain found in PfpI from Pyrococcus furiosus.				General function
385	lcl|sequenceName|1108814260774|locusTag|ORF0002	4	Porin,alphaproteobacteriatype[Mesorhizobiumsp.BNC1]gi|69279677|ref|ZP_00614483.1|Porin,alphaproteobacteriatype[Mesorhizobiumsp.BNC1],8.93824E-25"/anno3="porin,6.75423E-25"				Transport
386	BBAY40_contig00230.gene_2	8		COG3181, COG3181, Uncharacterized protein conserved in bacteria [Function unknown].			Uncharacterized
387	BBAY40_F51NMB401AUA8U.gene_1	2		COG0625, Gst, Glutathione S-transferase [Posttranslational modification, protein turnover, chaperones].			Stress response
388	BBAY40_F51NMB401B2CGI.gene_1	6			PF00102.20, Protein-tyrosine phosphatase		General function
390	lcl|sequenceName|1106105419241|locusTag|ORF0002	4		COG0702:Predictednucleoside-diphosphate-sugarepimerases,2.0E-7"			Carbohydrate metabolism
391	BBAY40_contig02875.gene_6	5		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain			Transport
392	BBAY40_contig23972.gene_19	8		COG0055, AtpD, F0F1-type ATP synthase, beta subunit [Energy production and conversion].			ATP synthesis coupled proton transport
393	BBAY40_contig10654.gene_1	16		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
394	lcl|sequenceName|1108431148252|locusTag|ORF0001	6	"HypotheticalproteinCBG18635[Caenorhabditisbriggsae],1.98538E-18"/anno2="CathepsinB-likecysteineproteinase3precursor-Caenorhabditiselegans,1.73865E-18"/anno3="cpr-3;CysteinePRoteaserelated[EC:3.4.22.1];K01363cathepsinB,1.47838E-17"				Hypothetical
395	BBAY40_F51NMB402IKUXT.gene_1	2		#N/A			Uncharacterized
396	lcl|sequenceName|1108814258930|locusTag|ORF0001	2	hypotheticalprotein,1.25761E45"/anno2="hypotheticalproteinZMO1522[Zymomonasmobilissubsp.mobilisZM4]gi|56543992|gb|AAV90146.1|conservedhypotheticalprotein[Zymomonasmobilissubsp.mobilisZM4],1.89253E-45"				Hypothetical
397	BBAY42_GBUJSMQ02FLWUA.gene_1	2			PF01731.13, Arylesterase		Stress response
398	BBAY41_GBSR4EV01CKB0X.gene_1	4		COG5648, NHP6B, Chromatin-associated proteins containing the HMG domain [Chromatin structure and dynamics].			Nucleosome assembly
401	BBAY40_F51NMB401CTI4W.gene_1	9	PREDICTED: similar to chloride intracellular channel 5 [Hydra magnipapillata]				Transport
402	BBAY40_F51NMB401EHWSE.gene_1	18		COG2885, OmpA, Outer membrane protein and related peptidoglycan-associated (lipo)proteins			Transport
403	BBAY41_GBSR4EV02FMVLD.gene_1	2	hypothetical protein VspiD_04590 [Verrucomicrobium spinosum DSM 4136]				Hypothetical
405	BBAY40_F51NMB401DIBQP.gene_1	4		COG0154, GatA, Asp-tRNAAsn/Glu-tRNAGln amidotransferase A subunit and related amidases			Translation
406	lcl|sequenceName|1108430362061|locusTag|ORF0001	13		COG1028:Dehydrogenaseswithdifferentspecificities(relatedtoshort-chainalcoholdehydrogenases),1.1E-11"			Lipid biosynthesis
407	BBAY42_GBUJSMQ02GRISS.gene_1	5		COG3869, COG3869, Arginine kinase [Amino acid transport and metabolism].			Amino acid biosynthesis
408	BBAY42_GBUJSMQ01BUZPM.gene_1	3			PF05649.6, Peptidase family M13		General metabolism
409	BBAY40_contig09800.gene_16	2		COG3181, COG3181, Uncharacterized protein conserved in bacteria [Function unknown].			Uncharacterized
411	BBAY40_contig06814.gene_15	10		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport

412	BBAY40_F51NMB401AP13O.gene_1	7		#N/A			Uncharacterized
413	BBAY41_contig03700.gene_1	5		COG1100, COG1100, GTPase SAR1 and related small G proteins [General function prediction only].			General function
415	BBAY41_GBSR4EV02G60AG.gene_1	2		#N/A			Uncharacterized
416	BBAY40_F51NMB401C0AW2.gene_1	10		COG5026, COG5026, Hexokinase [Carbohydrate transport and metabolism].			Carbohydrate metabolism
417	lcl|sequenceName|1108814262327|locusTag|ORF0002	2		COG3106:PredictedATPase,6.2E-76"			Uncharacterized
418	BBAY42_contig41288.gene_1	2		#N/A			Uncharacterized
419	BBAY42_contig47154.gene_2	2			PF00036.25, EF hand		Signalling
420	BBAY42_GBUJSMQ01D8O1H.gene_1	4	PREDICTED: similar to multidrug resistance protein 1; P-glycoprotein [Monodelphis domestica]				Transport
421	lcl|sequenceName|1108814258101|locusTag|ORF0002	2		COG0507:ATP-dependentexoDNAse(exonucleaseV),alphasubunit-helicasesuperfamilyImember,1.1E-89"			Transcription
427	BBAY42_GBUJSMQ02G8T30.gene_1	8		#N/A			Uncharacterized
428	BBAY40_contig25205.gene_10	8		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
429	BBAY42_GBUJSMQ02HMT3F.gene_1	23	hypothetical protein BRAFLDRAFT_129071 [Branchiostoma floridae]gb|EEN51030.1| hypothetical protein BRAFLDRAFT_129071 [Branchiostoma floridae]				Hypothetical
430	lcl|sequenceName|1108814242899|locusTag|ORF0001	19	cytochrome c oxidase subunit II [Topsentia ophiraphidites]gb|ABW83861.1| cytochrome c oxidase subunit 2 [Topsentia ophiraphidites]				Electron transport
431	BBAY41_GBSR4EV02ISMU4.gene_2	9		COG2885, OmpA, Outer membrane protein and related peptidoglycan-associated (lipo)proteins			Transport
432	lcl|sequenceName|1108814258778|locusTag|ORF0031	2		COG0050:GTPases-translationelongationfactors,4.3E-236"			Translation
433	BBAY40_F51NMB402IDEMO.gene_2	4		#N/A			Uncharacterized
435	BBAY42_contig35368.gene_1	9	secreted protein [Candidatus Poribacteria sp. WGA-A3]gb|EFC34149.1| secreted protein [Candidatus Poribacteria sp. WGA-A3]				Other
436	BBAY41_GBSR4EV02GWVDP.gene_1	9		#N/A			Uncharacterized
437	BBAY41_GBSR4EV01DPCVG.gene_2	8			PF09380.3, FERM C-terminal PH-like domain		Other
439	BBAY40_F51NMB402J1982.gene_1	7		#N/A			Uncharacterized
440	BBAY40_F51NMB401EN5VN.gene_2	2		#N/A			Uncharacterized
441	BBAY42_GBUJSMQ01DSN5Y.gene_1	2		#N/A			Uncharacterized
443	BBAY41_GBSR4EV01BI9MJ.gene_2	6		#N/A			Uncharacterized
444	BBAY42_GBUJSMQ01CY4NP.gene_1	9	hypothetical protein BRAFLDRAFT_129071 [Branchiostoma floridae]gb|EEN51030.1| hypothetical protein BRAFLDRAFT_129071 [Branchiostoma floridae]				Hypothetical
445	lcl|sequenceName|1108814258013|locusTag|ORF0003	8	TonB-dependent receptor [marine gamma proteobacterium HTCC2143]gb|EAW31665.1| TonB-dependent receptor [marine gamma proteobacterium HTCC2143]				Transport
446	BBAY40_contig07288.gene_7	20		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport

447	BBAY40_contig05614.gene_1	2		COG0649, NuoD, NADH:ubiquinone oxidoreductase 49 kD subunit 7 [Energy production and conversion].		Electron transport
448	BBAY40_contig06466.gene_2	2		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].		Protein folding
449	BBAY42_GBUJSMQ02H68SZ.gene_1	6		PF02197.10, Regulatory subunit of type II PKA R-subunit		General function
452	BBAY42_contig26892.gene_10	7		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component		Transport
453	BBAY41_GBSR4EV01DUOW0.gene_1	11		COG4221, COG4221, Short-chain alcohol dehydrogenase of unknown specificity [General function prediction only].		Lipid biosynthesis
454	BBAY40_contig23575.gene_2	6		COG3211, PhoX, Predicted phosphatase [General function prediction only].		Scavenging
455	BBAY40_F51NMB401EKBA1.gene_1	3		COG4993, Gcd, Glucose dehydrogenase [Carbohydrate transport and metabolism].		Carbohydrate metabolism
456	lcl|sequenceName|1108814260902|locusTag|ORF0001	12		COG1629:Outermembranereceptorproteins,mostlyFetransport,4.7E-29"		Transport
457	BBAY42_GBUJSMQ02H4T36.gene_1	5		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].		Protein folding
458	BBAY41_GBSR4EV01CHPC0.gene_1	6		COG3588, COG3588, Fructose-1,6-bisphosphate aldolase [Carbohydrate transport and metabolism].		Carbohydrate metabolism
459	BBAY42_contig19791.gene_7	29		COG0056, AtpA, F0F1-type ATP synthase, alpha subunit [Energy production and conversion].		ATP synthesis coupled proton transport
460	lcl|sequenceName|1108814257826|locusTag|ORF0002	4		COG0499:S-adenosylhomocysteine hydrolase,2.3E294"		Amino acid biosynthesis
461	BBAY40_contig03129.gene_1	13		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain		Transport
463	lcl|sequenceName|1108814258788|locusTag|ORF0005	6		COG2211:Na+/melibiosesymporterandrelatedtransporters,7.1E-32"/		Transport
464	BBAY42_GBUJSMQ02JPZIO.gene_1	4		COG0563, Adk, Adenylate kinase and related kinases [Nucleotide transport and metabolism].		Nucletotide biosynthesis
465	lcl|sequenceName|1108457314776|locusTag|ORF0001	3		COG1100:GTPases,1.7E-22"/		Signalling
466	lcl|sequenceName|1108814257542|locusTag|ORF0009	2		COG0069:Glutamatesynthasedomain2,0.0"/		Amino acid biosynthesis
467	BBAY41_GBSR4EV01CJ05Q.gene_1	2		PF00240.16, Ubiquitin family		General function
468	BBAY40_contig10597.gene_1	2		COG0260, PepB, Leucyl aminopeptidase [Amino acid transport and metabolism].		Amino acid metabolism
469	lcl|sequenceName|1106850431099|locusTag|ORF0001	2	HypotheticalproteinZC178.1[Caenorhabditiselegans]gi|17566580|ref|NP_505242.1|ZC178.1[Caenorhabditiselegans],3.6842E-7"/anno2="ZC178.1,2.4482E-7"			Hypothetical
470	BBAY40_contig12441.gene_2	2		COG0234, GroS, Co-chaperonin GroES (HSP10) [Posttranslational modification, protein turnover, chaperones].		Protein folding
472	BBAY40_contig07924.gene_2	2		COG0176, MipB, Transaldolase [Carbohydrate transport and metabolism].		Carbohydrate metabolism
473	BBAY40_contig25241.gene_2	2		COG4147, DhlC, Predicted symporter [General function prediction only].		Transport
474	BBAY42_GBUJSMQ02HXXY1.gene_1	3		COG1574, COG1574, Predicted metal-dependent hydrolase with the TIM-barrel fold [General function prediction only].		General function
475	BBAY40_F51NMB401CZZ8X.gene_1	2		COG0114, FumC, Fumarase [Energy production and conversion].		TCA
476	BBAY41_GBSR4EV02GJ1NH.gene_1	2		COG0451, WcaG, Nucleoside-diphosphate-sugar epimerases		Transport

477	BBAY40_F51NMB402HR5WC.gene_1	4		#N/A			Uncharacterized
478	BBAY42_GBUJSMQ01BZ8BK.gene_1	2		#N/A			Uncharacterized
479	BBAY40_F51NMB401CCDU0.gene_1	2		COG0044, PyrC, Dihydroorotase and related cyclic amidohydrolases [Nucleotide transport and metabolism].			Nucletotide biosynthesis
480	BBAY40_F51NMB401DHLKG.gene_1	2		COG1960, CaiA, Acyl-CoA dehydrogenases [Lipid metabolism].			Lipid metabolism
481	BBAY41_GBSR4EV02HW9LF.gene_1	3		COG0834, HisJ, ABC-type amino acid transport/signal transduction systems, periplasmic component/domain			Transport
486	BBAY42_GBUJSMQ01AUSEP.gene_2	2			PF10240.2, Protein of unknown function (DUF2464)		Uncharacterized
487	BBAY40_contig02383.gene_2	2		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
488	BBAY40_contig16462.gene_1	2		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
489	lcl|sequenceName|1106850359979|locusTag|ORF0001	4		COG2303:Cholinedehydrogenaseandrelatedflavoproteins,1.2E-56"/anno4="betA:cholinedehydrogenase,1.7E-13"			Stress response
490	BBAY42_contig48964.gene_2	2		#N/A			Uncharacterized
492	BBAY42_GBUJSMQ02IE0NK.gene_1	2			PF04030.7, D-arabinono-1,4-lactone oxidase		Stress response
493	BBAY40_contig03067.gene_1	7			PF03945.7, delta endotoxin, N-terminal domain		General function
495	BBAY40_F51NMB402F0CFI.gene_1	2		#N/A			Uncharacterized
496	BBAY42_contig37676.gene_5	2		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport
497	BBAY42_contig46774.gene_1	3			PF03727.9, Hexokinase		Carbohydrate metabolism
498	BBAY42_contig06095.gene_1	2		#N/A			Uncharacterized
499	BBAY40_F51NMB402JTM6P.gene_1	2	hypothetical protein P700755_22735 [Psychroflexus torquis ATCC 700755]				Hypothetical
500	BBAY41_GBSR4EV02H6OA5.gene_2	2		#N/A			Uncharacterized
501	BBAY41_contig00362.gene_1	6			PF07963.5, Prokaryotic N-terminal methylation motif		Transport
502	BBAY41_contig00175.gene_18	9		COG3203, OmpC, Outer membrane protein (porin) [Cell envelope biogenesis, outer membrane].			Transport
504	BBAY41_contig16108.gene_44	3		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
505	BBAY40_F51NMB402J2O5E.gene_1	2		#N/A			Uncharacterized
506	BBAY40_F51NMB401EG1T1.gene_1	4			PF00503.13, G-protein alpha subunit		Other
507	BBAY41_contig03092.gene_1	5		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport
508	BBAY42_GBUJSMQ01AIBB9.gene_1	6		COG0112, GlyA, Glycine/serine hydroxymethyltransferase [Amino acid transport and metabolism].			Amino acid biosynthesis
509	BBAY42_GBUJSMQ01BRRVG.gene_1	8	PRK06210, PRK06210, enoyl-CoA hydratase; Provisional.				Lipid metabolism

511	BBAY41_contig05585.gene_1	2		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
512	BBAY40_F51NMB402I5JLT.gene_1	2		#N/A			Uncharacterized
513	BBAY42_GBUJSMQ01DKWOA.gene_1	2		#N/A			Uncharacterized
514	BBAY41_contig14505.gene_5	9			PF04744.5, Monooxygenase subunit B protein		Ammonia related
515	BBAY41_contig14505.gene_4	2			PF04896.5, Ammonia monooxygenase/methane monooxygenase, subunit C		Ammonia related
517	BBAY41_contig01100.gene_1	2			PF00394.15, Multicopper oxidase		General function
518	BBAY42_GBUJSMQ02F4QAF.gene_1	3		#N/A			Uncharacterized
519	BBAY40_F51NMB402GFDC8.gene_1	2		#N/A			Uncharacterized
521	BBAY41_GBSR4EV02H42H5.gene_1	2	conserved hypothetical protein, secreted [Candidatus Poribacteria sp. WGA-A3]gb|EFC34519.1| conserved hypothetical protein, secreted [Candidatus Poribacteria sp. WGA-A3]				Hypothetical
522	BBAY42_GBUJSMQ01DFM2T.gene_1	3		COG1024, CaiD, Enoyl-CoA hydratase/carnithine racemase [Lipid metabolism].			Lipid metabolism
523	lcl|sequenceName|1108385168727|locusTag|ORF0001	4		COG0176:Transaldolase,1.4E-47"/anno3="Transaldolase-Synechococcussp.(strainWH8102),7.5563E-64"			General metabolism
524	BBAY42_GBUJSMQ02HZHQ7.gene_1	2		#N/A			Uncharacterized
525	BBAY41_GBSR4EV02ISX54.gene_2	2		#N/A			Uncharacterized
527	BBAY41_GBSR4EV02G05A1.gene_1	4		COG5265, ATM1, ABC-type transport system involved in Fe-S cluster assembly, permease and ATPase components			Transport
528	BBAY42_GBUJSMQ01E25FS.gene_1	3		#N/A			Uncharacterized
529	BBAY41_contig17961.gene_1	6	galactose-binding protein [Carcinoscorpius rotundicauda]				Transport
530	BBAY40_F51NMB402F5ERS.gene_1	12		#N/A			Uncharacterized
531	BBAY40_F51NMB402G9U0R.gene_1	12		#N/A			Uncharacterized
532	BBAY41_contig04627.gene_1	4		COG0190, FolD, 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase			Coenzyme biosynthesis
533	lcl|sequenceName|1106850496623|locusTag|ORF0001	6		COG0610:RestrictionenzymestypeIhelicasesubunitsandrelatedhelicases,5.0E-40"			Restriction modification
534	BBAY40_F51NMB401DQAPZ.gene_1	2			PF02530.7, Porin subfamily		Transport
537	BBAY40_contig10966.gene_6	6		COG0715, TauA, ABC-type nitrate/sulfonate/bicarbonate transport systems, periplasmic components			Transport
538	BBAY41_contig15144.gene_7	6		COG2132, SufI, Putative multicopper oxidases [Secondary metabolites biosynthesis, transport, and catabolism].			Nitrate reduction
539	BBAY40_contig09661.gene_23	6		COG0092, RpsC, Ribosomal protein S3 [Translation, ribosomal structure and biogenesis].			Translation
540	BBAY41_contig11758.gene_17	2		COG5256, TEF1, Translation elongation factor EF-1alpha (GTPase) [Translation, ribosomal structure and biogenesis].			Translation
542	lcl|sequenceName|1106851029035|locusTag|ORF0001	5		COG1028:Dehydrogenaseswithdifferentspecificities(relatedtoshort-chainalcoholdehydrogenases),2.6E-7"/			Lipid biosynthesis

543	BBAY40_F51NMB402JGTER.gene_1	5		COG0656, ARA1, Aldo/keto reductases, related to diketogulonate reductase [General function prediction only].			General metabolism
545	BBAY40_contig11939.gene_1	2		COG1185, Pnp, Polyribonucleotide nucleotidyltransferase (polynucleotide phosphorylase)			Translation
546	BBAY40_F51NMB401D2Q76.gene_1	2		COG0004, AmtB, Ammonia permease [Inorganic ion transport and metabolism].			Transport
547	BBAY40_contig08045.gene_1	2		COG3666, COG3666, Transposase and inactivated derivatives [DNA replication, recombination, and repair].			Transposases
548	BBAY41_GBSR4EV01C3EOQ.gene_1	4		COG0065, LeuC, 3-isopropylmalate dehydratase large subunit [Amino acid transport and metabolism].			Amino acid biosynthesis
549	BBAY42_GBUJSMQ01AREZZ.gene_1	2		COG1028, FabG, Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)			Lipid biosynthesis
550	BBAY40_F51NMB401C3KZG.gene_1	2			PF00929.17, Exonuclease		Transcription
551	BBAY42_contig48654.gene_1	2		#N/A			Uncharacterized
553	BBAY41_contig06701.gene_2	2		COG1879, RbsB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
554	BBAY42_contig06367.gene_1	2		COG0664, Crp, cAMP-binding proteins - catabolite gene activator and regulatory subunit of cAMP-dependent protein kinases [Signal transduction mechanisms].			Stress response
556	BBAY41_GBSR4EV01BAJXL.gene_1	2	hypothetical protein NAEGRDRAFT_80317 [Naegleria gruberi]gb|EFC42703.1| hypothetical protein NAEGRDRAFT_80317 [Naegleria gruberi]				Hypothetical
557	BBAY41_GBSR4EV01BDSRM.gene_1	2		#N/A			Uncharacterized
558	BBAY41_contig09176.gene_3	2		COG3968, COG3968, Uncharacterized protein related to glutamine synthetase [General function prediction only].			Amino acid biosynthesis
559	lcl|sequenceName|1108472340186|locusTag|ORF0001	2	HypotheticalPE-PGRSfamilyproteinRv1325c/MT1367precursor[MagnetospirillummagneticumAMB-1]gi|83309124|ref|YP_419388.1|hypotheticalproteinamb0025[MagnetospirillummagneticumAMB-1],5.40468E-7"				Hypothetical
560	BBAY40_F51NMB402FKMP8.gene_1	4			PF11901.1, Protein of unknown function (DUF3421)		Uncharacterized
561	BBAY41_contig04755.gene_1	2		#N/A			Uncharacterized
562	BBAY40_F51NMB401D2BNU.gene_1	2			PF03098.8, Animal haem peroxidase		Stress response
563	BBAY40_contig08682.gene_2	2		COG0583, LysR, Transcriptional regulator [Transcription].			Transcription
565	BBAY40_F51NMB402HDPWY.gene_1	6		COG1629, CirA, Outer membrane receptor proteins, mostly Fe transport [Inorganic ion transport and metabolism].			Transport
567	BBAY40_F51NMB401DD7D4.gene_1	2		COG0260, PepB, Leucyl aminopeptidase [Amino acid transport and metabolism].			Amino acid metabolism
568	BBAY42_contig05527.gene_2	2		#N/A			Uncharacterized
569	BBAY40_F51NMB402I8BMW.gene_1	4			PF03098.8, Animal haem peroxidase		Stress response
570	BBAY42_GBUJSMQ01BBPO0.gene_2	4		#N/A			Uncharacterized
571	BBAY40_F51NMB402IVPQN.gene_1	4		#N/A			Uncharacterized
573	BBAY40_F51NMB401AP7V1.gene_2	4		#N/A			Uncharacterized

574	BBAY41_GBSR4EV02HFUF2.gene_1	2		#N/A			Stress response
575	BBAY40_contig01731.gene_6	2		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
576	BBAY42_GBUJSMQ02FKPQZ.gene_1	6		COG0050, TufB, GTPases - translation elongation factors [Translation, ribosomal structure and biogenesis].			Translation
577	lcl|sequenceName|1108814260851|locusTag|ORF0001	6		#N/A			Uncharacterized
579	BBAY41_GBSR4EV01CU5FJ.gene_1	4		#N/A			Uncharacterized
581	lcl|sequenceName|1108431096587|locusTag|ORF0001	4	SRCR12;scavengerreceptorcysteine-richproteintype12,9.07897E-21"/anno3="Neurotrypsinprecursor-Saguinuslabiatus(Red-chestedmustachedtamarin),2.92197E-16"				Scavenging
582	BBAY42_GBUJSMQ02JQC0V.gene_1	2			PF00043.18, Glutathione S-transferase, C-terminal domain		Amino acid biosynthesis
583	lcl|sequenceName|1108814244010|locusTag|ORF0002	2	Peroxidase [Nitrococcus mobilis Nb-231]gb|EAR20476.1| Peroxidase [Nitrococcus mobilis Nb-231]				Electron transport
584	BBAY41_GBSR4EV02I6RXS.gene_1	2		#N/A			Uncharacterized
585	BBAY40_F51NMB402H38PX.gene_1	5		COG2086, FixA, Electron transfer flavoprotein, beta subunit [Energy production and conversion].			Electron transport
586	BBAY40_contig13810.gene_4	3		COG1653, UgpB, ABC-type sugar transport system, periplasmic component [Carbohydrate transport and metabolism].			Transport
587	lcl|sequenceName|1108472215973|locusTag|ORF0001	4		#N/A			Uncharacterized
588	BBAY40_contig01731.gene_5	5		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
590	BBAY41_GBSR4EV02H2SEP.gene_1	2			PF00056.16, lactate/malate dehydrogenase, NAD binding domain		TCA
592	BBAY40_F51NMB401EFQMS.gene_1	2		#N/A			Uncharacterized
593	BBAY42_GBUJSMQ01CP2F1.gene_1	2			PF07494.4, Two component regulator propeller		General function
594	BBAY41_GBSR4EV01ESNT1.gene_1	2			PRK14479, PRK14479, dihydroxyacetone kinase; Provisional.		General metabolism
595	BBAY40_contig22747.gene_7	3		COG1432, COG1432, Uncharacterized conserved protein [Function unknown].			Uncharacterized
596	BBAY40_F51NMB401AYKJ1.gene_1	2		COG4770, COG4770, Acetyl/propionyl-CoA carboxylase, alpha subunit [Lipid metabolism].			Lipid metabolism
597	BBAY41_GBSR4EV01AJ6RX.gene_1	2		#N/A			Uncharacterized
598	BBAY40_F51NMB401EQ099.gene_1	2		COG0563, Adk, Adenylate kinase and related kinases [Nucleotide transport and metabolism].			Nucletodie metabolism
599	BBAY42_contig00435.gene_1	7		COG1028, FabG, Dehydrogenases with different specificities (related to short-chain alcohol dehydrogenases)			General metabolism
601	BBAY42_GBUJSMQ01D59I8.gene_1	4		COG0686, Ald, Alanine dehydrogenase [Amino acid transport and metabolism].			Electron transport
602	BBAY41_GBSR4EV01D94J0.gene_1	4	pfam00503, G-alpha, G-protein alpha subunit. G proteins couple receptors of extracellular signals to intracellular signaling pathways. The G protein alpha subunit binds guanyl nucleotide and is a weak GTPase.				Signalling
603	BBAY40_F51NMB401DCC4Q.gene_1	2		#N/A			Uncharacterized
605	BBAY41_GBSR4EV01AYP9J.gene_2	2		COG0034, PurF, Glutamine phosphoribosylpyrophosphate amidotransferase [Nucleotide transport and metabolism].			Nucletotide biosynthesis

606	BBAY40_contig14457.gene_3	2		COG1162, COG1162, Predicted GTPases [General function prediction only].			General function
607	BBAY41_GBSR4EV01BALUP.gene_1	2	cd00288, Pyruvate_Kinase, Pyruvate kinase (PK): Large allosteric enzyme that regulates glycolysis through binding of the substrate, phosphoenolpyruvate, and one or more allosteric effectors. Like other allosteric enzymes, PK has a high substrate affinity R state and a low affinity T state.				General metabolism
608	BBAY40_F51NMB402JNB0O.gene_1	2		COG0324, MiaA, tRNA delta(2)-isopentenylpyrophosphate transferase [Translation, ribosomal structure and biogenesis].			Translation
609	BBAY40_F51NMB401BFSUM.gene_1	2			PF00494.12, Squalene/phytoene synthase		Coenzyme biosynthesis
610	BBAY42_GBUJSMQ02HLIU0.gene_1	2		#N/A			Uncharacterized
611	BBAY40_F51NMB402HTHLI.gene_1	2		#N/A			Uncharacterized
612	BBAY40_contig10771.gene_15	2		COG3839, MalK, ABC-type sugar transport systems, ATPase components [Carbohydrate transport and metabolism].			Transport
613	BBAY42_GBUJSMQ02JDE4R.gene_1	2		#N/A			Uncharacterized
614	BBAY42_contig01792.gene_4	2		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
615	BBAY41_GBSR4EV02I29PL.gene_1	2		COG1448, TyrB, Aspartate/tyrosine/aromatic aminotransferase [Amino acid transport and metabolism].			Transport
616	BBAY41_GBSR4EV01DQM7Z.gene_1	2		COG5126, FRQ1, Ca2+-binding protein (EF-Hand superfamily)			Signalling
617	lcl|sequenceName|1108814248159|locusTag|ORF0011	2	3-hydroxyacyl-CoA dehydrogenase type II [gamma proteobacterium HdN1]emb|CBL44686.1| 3-hydroxyacyl-CoA dehydrogenase type II [gamma proteobacterium HdN1]				Electron transport
618	BBAY42_GBUJSMQ01ARRMU.gene_1	2		#N/A			Uncharacterized
619	BBAY42_GBUJSMQ01DFE84.gene_1	3		COG1022, FAA1, Long-chain acyl-CoA synthetases (AMP-forming) [Lipid metabolism].			Lipid metabolism
620	BBAY42_GBUJSMQ01CDMLQ.gene_1	2	cd03185, GST_C_Tau, GST_C family, Class Tau subfamily; GSTs are cytosolic dimeric proteins involved in cellular detoxification by catalyzing the conjugation of glutathione (GSH) with a wide range of endogenous and xenobiotic alkylating agents, including carcinogens, therapeutic drugs, environmental toxins and products of oxidative stress.				Stress response
621	BBAY42_GBUJSMQ02HAGOL.gene_1	3		#N/A			Uncharacterized
622	BBAY42_contig29876.gene_2	2	PLN02650, PLN02650, dihydrokaempferol 4-reductase.				Coenzyme biosynthesis
624	BBAY41_GBSR4EV01DP64B.gene_1	2		#N/A			Uncharacterized
625	BBAY42_contig10867.gene_6	9		COG5013, NarG, Nitrate reductase alpha subunit [Energy production and conversion].			Nitrate reduction
626	BBAY40_contig04989.gene_3	12		COG1140, NarY, Nitrate reductase beta subunit [Energy production and conversion].			Nitrate reduction
627	BBAY41_GBSR4EV01A09FB.gene_1	2		#N/A			Uncharacterized
628	BBAY41_contig04906.gene_1	4		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
629	BBAY40_F51NMB401A9CIX.gene_1	2		#N/A			Uncharacterized
630	BBAY42_contig11234.gene_1	4			PF01234.10, NNMT/PNMT/TEMT family		Amino acid metabolism

632	BBAY42_GBUJSMQ01EZDS3.gene_2	2		#N/A			Uncharacterized
633	BBAY42_GBUJSMQ02H8IG7.gene_1	2	hypothetical protein BRAFLDRAFT_93248 [Branchiostoma floridae]gb|EEN62467.1| hypothetical protein BRAFLDRAFT_93248 [Branchiostoma floridae]				Hypothetical
635	BBAY40_F51NMB402IPDTO.gene_1	2			PF00106.18, short chain dehydrogenase		General metabolism
636	BBAY40_F51NMB402HCLGN.gene_1	2		#N/A			Uncharacterized
637	BBAY40_contig04989.gene_2	24		COG0243, BisC, Anaerobic dehydrogenases, typically selenocysteine-containing [Energy production and conversion].			Nitrate reduction
638	BBAY41_GBSR4EV01D75KK.gene_1	2		COG0377, NuoB, NADH:ubiquinone oxidoreductase 20 kD subunit and related Fe-S oxidoreductases			Electron transport
639	BBAY42_GBUJSMQ02GXO4U.gene_1	3		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
640	BBAY40_F51NMB401BYNRS.gene_1	2		#N/A			Uncharacterized
641	BBAY42_GBUJSMQ01CXAI2.gene_1	2		COG1012, PutA, NAD-dependent aldehyde dehydrogenases [Energy production and conversion].			General metabolism
642	BBAY40_contig06055.gene_1	2	secreted protein [Candidatus Poribacteria sp. WGA-A3]gb|EFC34148.1| secreted protein [Candidatus Poribacteria sp. WGA-A3]				Other
643	BBAY41_contig10304.gene_1	2		COG0220, COG0220, Predicted S-adenosylmethionine-dependent methyltransferase [General function prediction only].			General function
644	BBAY42_GBUJSMQ01EWC3Q.gene_1	2		COG1012, PutA, NAD-dependent aldehyde dehydrogenases [Energy production and conversion].			General metabolism
645	BBAY42_GBUJSMQ01C1JH9.gene_1	2		COG0225, MsrA, Peptide methionine sulfoxide reductase [Posttranslational modification, protein turnover, chaperones].			Protein folding
646	BBAY40_F51NMB402F5N4M.gene_1	2		COG0747, DdpA, ABC-type dipeptide transport system, periplasmic component [Amino acid transport and metabolism].			Transport
647	BBAY42_GBUJSMQ01EZFTD.gene_1	2	cd03858, M14_CP_N-E_like, Carboxypeptidase (CP) N/E-like subfamily of the M14 family of metallocarboxypeptidases (MCPs). The M14 family are zinc-binding CPs which hydrolyze single, C-terminal amino acids from polypeptide chains, and have a recognition site for the free C-terminal carboxyl group, which is a key determinant of specificity.				General metabolism
648	BBAY40_F51NMB402H4E88.gene_1	2		#N/A			Uncharacterized
649	lcl|sequenceName|1108814257855|locusTag|ORF0023	2		COG1629:Outermembranereceptorproteins,mostlyFetransport,9.7E-64"			Transport
650	BBAY41_contig16108.gene_38	2		COG1205, COG1205, Distinct helicase family with a unique C-terminal domain including a metal-binding cysteine cluster [General function prediction only].			Transcription
651	BBAY41_GBSR4EV02GKD6Q.gene_1	2		COG1025, Ptr, Secreted/periplasmic Zn-dependent peptidases, insulinase-like			Protein folding
652	BBAY40_contig11288.gene_2	2		#N/A			Uncharacterized
653	BBAY41_GBSR4EV01EXMDT.gene_1	3		#N/A			Uncharacterized
654	BBAY42_GBUJSMQ01C568I.gene_1	3		COG0323, MutL, DNA mismatch repair enzyme (predicted ATPase) [DNA replication, recombination, and repair].			Transcription
655	BBAY41_GBSR4EV02JI0O7.gene_1	3	TIGR03466, HpnA, hopanoid-associated sugar epimerase. The sequences in this family are members of the pfam01370 superfamily of NAD-dependent epimerases and dehydratases typically acting on nucleotide-sugar substrates. The genes of the family modeled here are generally in the same locus with genes involved in the biosynthesis and elaboration of hopene, the cyclization product of the polyisoprenoid squalene.				General metabolism
656	BBAY42_contig03867.gene_4	2		#N/A			Uncharacterized

657	BBAY42_contig06399.gene_2	2			PF01055.19, Glycosyl hydrolases family 31		Carbohydrate metabolism
658	lcl|sequenceName|1108431016963|locusTag|ORF0001	2	Acyl-CoA-bindingprotein-Gallusgallus(Chicken),6.98508E-24"/anno3="DBI;diazepambindinginhibitor(GABAreceptormodulator,acyl-CoenzymeAbindingprotein);K08762diazepam-bindinginhibitor(GABAreceptormodulator,acyl-CoA-bindingprotein),1.70825E-22"				Other
659	BBAY40_contig14396.gene_3	2		COG0724, COG0724, RNA-binding proteins (RRM domain) [General function prediction only].			General function
660	BBAY42_contig07033.gene_1	2		#N/A			Uncharacterized
661	BBAY41_contig18873.gene_1	2		#N/A			Uncharacterized
662	BBAY42_contig48784.gene_1	2		#N/A			Uncharacterized
663	BBAY41_contig00921.gene_3	2		COG0017, AsnS, Aspartyl/asparaginyl-tRNA synthetases [Translation, ribosomal structure and biogenesis].			Translation
664	BBAY42_GBUJSMQ01CFI12.gene_2	3			PF01929.10, Ribosomal protein L14		Translation
665	BBAY42_contig51593.gene_1	3			PF00515.21, Tetratricopeptide repeat		Repeat protein
666	lcl|sequenceName|1106850853373|locusTag|ORF0001	2		COG0056:F0F1-typeATPsynthasealphasubunit,3.4E-226"/anno3="atpA:ATPsynthaseF1,alphasubunit,2.3E-137"			ATP synthesis coupled proton transport
667	BBAY42_GBUJSMQ01EL4BR.gene_1	2			PF00018.21, SH3 domain		Signalling
668	BBAY42_GBUJSMQ02JPTYA.gene_1	2	PREDICTED: similar to hypoxia up-regulated 1 [Ciona intestinalis]				Stress response
669	BBAY42_GBUJSMQ01D99QX.gene_2	2		COG0479, FrdB, Succinate dehydrogenase/fumarate reductase, Fe-S protein subunit			General metabolism
670	BBAY41_GBSR4EV02J4X6Z.gene_1	2		#N/A			Uncharacterized
671	BBAY42_GBUJSMQ01AZM0N.gene_1	3			PF08953.4, Domain of unknown function (DUF1899)		Uncharacterized
672	BBAY41_contig13851.gene_3	2		COG0056, AtpA, F0F1-type ATP synthase, alpha subunit [Energy production and conversion].			ATP synthesis coupled proton transport
673	BBAY42_GBUJSMQ01BZAH6.gene_1	12		COG0462, PrsA, Phosphoribosylpyrophosphate synthetase			Transport
674	BBAY42_GBUJSMQ01EV60I.gene_1	2		COG3517, COG3517, Uncharacterized protein conserved in bacteria [Function unknown].			Uncharacterized
675	BBAY40_F51NMB401A5849.gene_1	2	hypothetical protein Nmar_0344 [Nitrosopumilus maritimus SCM1]gb|ABX12240.1| hypothetical protein Nmar_0344 [Nitrosopumilus maritimus SCM1]				Hypothetical
676	BBAY42_GBUJSMQ02JHYIU.gene_1	2		COG0312, TldD, Predicted Zn-dependent proteases and their inactivated homologs [General function prediction only].			General function
677	BBAY42_GBUJSMQ01CDN54.gene_1	2	PRK08912, PRK08912, hypothetical protein; Provisional.				Hypothetical
678	lcl|sequenceName|1106626940957|locusTag|ORF0001	3		COG1629:Outermembranereceptorproteins,mostlyFetransport,4.3E-19"			Transport
679	BBAY41_contig18906.gene_1	2		#N/A			Uncharacterized
680	lcl|sequenceName|1108814262756|locusTag|ORF0002	2		COG1082:Sugarphosphateisomerases/epimerases,1.2E-10"			Carbohydrate metabolism
682	BBAY41_GBSR4EV02G596R.gene_1	2		#N/A			Uncharacterized
683	BBAY42_contig19676.gene_1	2		#N/A			Uncharacterized

684	BBAY40_F51NMB401B3BTI.gene_1	2		#N/A			Uncharacterized
685	BBAY42_GBUJSMQ02GNGO9.gene_1	3		COG0538, Icd, Isocitrate dehydrogenases [Energy production and conversion].			General metabolism
686	BBAY40_F51NMB402IX42M.gene_1	2		#N/A			Uncharacterized
687	BBAY42_contig03683.gene_9	2		#N/A			Uncharacterized
688	lcl|sequenceName|1106850405396|locusTag|ORF0001	2		#N/A			Uncharacterized
689	BBAY40_F51NMB402JUW19.gene_1	2		#N/A			Uncharacterized
690	BBAY42_GBUJSMQ02G2ZT6.gene_1	2		#N/A			Uncharacterized
691	BBAY40_contig00646.gene_2	2		COG0112, GlyA, Glycine/serine hydroxymethyltransferase [Amino acid transport and metabolism].			Amino acid biosynthesis
692	BBAY40_contig00892.gene_1	2	TIGR01783, TonB-siderophor, TonB-dependent siderophore receptor. This subfamily model encompasses a wide variety of TonB-dependent outer membrane siderophore receptors.				Transport
693	BBAY42_contig00052.gene_10	2	lysine-arginine-ornithine-binding periplasmic protein [Labrenzia alexandrii DFL-11]gb|EEE46774.1| lysine-arginine-ornithine-binding periplasmic protein [Labrenzia alexandrii DFL-11]				Transport
694	BBAY40_F51NMB401AEAG0.gene_1	4		#N/A			Uncharacterized
695	BBAY42_contig09429.gene_1	2			PRK00179, pgi, glucose-6-phosphate isomerase; Reviewed.		General metabolism
696	BBAY41_contig02176.gene_1	2		#N/A			Uncharacterized
698	BBAY42_GBUJSMQ01DUX9X.gene_1	2		#N/A			Uncharacterized
699	BBAY42_GBUJSMQ02HBBO7.gene_1	2		#N/A			Uncharacterized
700	BBAY42_GBUJSMQ01C042D.gene_1	2		#N/A			Uncharacterized
701	BBAY42_GBUJSMQ01DDUEE.gene_1	2		COG0330, HflC, Membrane protease subunits, stomatin/prohibitin homologs			Proteolysis
702	BBAY42_contig36463.gene_3	2		#N/A			Uncharacterized
703	BBAY42_GBUJSMQ01A1SFG.gene_1	2		#N/A			Uncharacterized
704	BBAY42_GBUJSMQ01BJUYN.gene_1	2		#N/A			Uncharacterized
706	lcl|sequenceName|1108472303538|locusTag|ORF0001	2	hypothetical protein Nmar_0992 [Nitrosopumilus maritimus SCM1]gb|ABX12888.1| hypothetical protein Nmar_0992 [Nitrosopumilus maritimus SCM1]				Hypothetical
707	BBAY42_GBUJSMQ01B4UEW.gene_1	3	hypothetical protein VspiD_04590 [Verrucomicrobium spinosum DSM 4136]				Hypothetical
710	BBAY40_F51NMB401BYUBS.gene_1	3		#N/A			Uncharacterized
711	BBAY41_GBSR4EV01C01Q9.gene_1	2		COG1574, COG1574, Predicted metal-dependent hydrolase with the TIM-barrel fold [General function prediction only].			General function
712	BBAY40_contig23972.gene_21	2		COG0056, AtpA, F0F1-type ATP synthase, alpha subunit [Energy production and conversion].			ATP synthesis coupled proton transport
714	lcl|sequenceName|1106850569797|locusTag|ORF0001	2	hypotheticalproteinFrancci3_2726[Frankiasp.CcI3]gi|86568277|gb|ABD12086.1|hypotheticalproteinFrancci3_2726[Frankiasp.CcI3],7.36191E-7"/anno2="hypotheticalprotein,1.28739E-7"				Hypothetical

715	BBAY42_contig02827.gene_2	2		COG4771, FepA, Outer membrane receptor for ferrienterochelin and colicins [Inorganic ion transport and metabolism].			Transport
716	BBAY42_GBUJSMQ02F6SCB.gene_1	2		#N/A			Uncharacterized
717	lcl|sequenceName|1108814258491|locusTag|ORF0006	2		COG0533:Metal-dependentproteaseswithpossiblechaperoneactivity,2.1E-148"			Protein folding
718	BBAY42_GBUJSMQ01D2ELE.gene_1	2		#N/A			Uncharacterized
719	BBAY42_contig36399.gene_7	2	hypothetical protein NIDE3308 [Candidatus Nitrospira defluvii]emb|CBK42996.1| conserved exported protein of unknown function [Candidatus Nitrospira defluvii]				Hypothetical
720	BBAY41_contig18563.gene_1	2		#N/A			Uncharacterized
723	BBAY41_GBSR4EV02FJ69S.gene_1	3		COG4993, Gcd, Glucose dehydrogenase [Carbohydrate transport and metabolism].			Carbohydrate metabolism
724	BBAY40_contig00088.gene_17	2		COG5517, COG5517, Small subunit of phenylpropionate dioxygenase			General metabolism
725	BBAY42_contig16059.gene_1	3		#N/A			Uncharacterized
726	BBAY42_contig37953.gene_2	5	hypothetical protein Nwat_0215 [Nitrosococcus watsoni C-113]gb|ADJ27186.1| conserved hypothetical protein [Nitrosococcus watsoni C-113]				Hypothetical
727	BBAY40_contig14296.gene_1	5		COG0683, LivK, ABC-type branched-chain amino acid transport systems, periplasmic component			Transport
728	BBAY40_contig02750.gene_2	3		COG1061, SSL2, DNA or RNA helicases of superfamily II [Transcription / DNA replication, recombination, and repair].			Transcription
729	BBAY41_contig10991.gene_20	2		COG0136, Asd, Aspartate-semialdehyde dehydrogenase [Amino acid transport and metabolism].			Amino acid biosynthesis
730	BBAY42_contig37424.gene_3	2		COG1309, AcrR, Transcriptional regulator [Transcription].			Transcription
731	BBAY41_contig12190.gene_3	2		COG0187, GyrB, Type IIA topoisomerase (DNA gyrase/topo II, topoisomerase IV), B subunit			Transcription
732	BBAY40_contig08115.gene_1	2		COG1012, PutA, NAD-dependent aldehyde dehydrogenases [Energy production and conversion].			General metabolism
734	BBAY42_contig13032.gene_2	2		#N/A			Uncharacterized
735	BBAY42_GBUJSMQ02IZC79.gene_1	2		#N/A			Uncharacterized
736	lcl|sequenceName|1108814246906|locusTag|ORF0001	3	TonB-dependent receptor [Caulobacter sp. K31]gb|ABZ70979.1| TonB-dependent receptor [Caulobacter sp. K31]				Transport
737	BBAY41_GBSR4EV02FHOE7.gene_1	2		#N/A			Uncharacterized
738	BBAY42_GBUJSMQ02HGWDN.gene_1	3		COG3243, PhaC, Poly(3-hydroxyalkanoate) synthetase [Lipid metabolism].			General function
741	BBAY40_F51NMB402JN1JK.gene_1	2		#N/A			Uncharacterized
742	BBAY41_GBSR4EV02HNWSK.gene_1	2		#N/A			Uncharacterized
744	lcl|sequenceName|1108814258813|locusTag|ORF0030	2		COG0281:Malicenzyme,1.2E-257"/anno3="NADP-dependentmalicenzyme-Rhizobiummeliloti(Sinorhizobiummeliloti)			Nucletodie metabolism
745	BBAY42_contig13835.gene_2	2		COG0845, AcrA, Membrane-fusion protein [Cell envelope biogenesis, outer membrane].			Transport
746	BBAY40_F51NMB401A190W.gene_2	2		#N/A			Uncharacterized

749	BBAY41_GBSR4EV02F0LZK.gene_1	2	PREDICTED: similar to Aldehyde dehydrogenase 1 family, member L2 isoform 2 [Strongylocentrotus purpuratus]ref|XP_001176706.1| PREDICTED: similar to Aldehyde dehydrogenase 1 family, member L2 isoform 2 [Strongylocentrotus purpuratus]				General metabolism
750	BBAY42_contig09305.gene_2	2		COG0776, HimA, Bacterial nucleoid DNA-binding protein [DNA replication, recombination, and repair].			Transcription
751	BBAY42_GBUJSMQ02GPTJ8.gene_1	4		COG0114, FumC, Fumarase [Energy production and conversion].			TCA
753	BBAY40_F51NMB401CBF76.gene_1	2	PRK08277, PRK08277, D-mannonate oxidoreductase; Provisional.				General metabolism
754	BBAY42_contig10792.gene_1	2		COG0459, GroL, Chaperonin GroEL (HSP60 family) [Posttranslational modification, protein turnover, chaperones].			Protein folding
755	BBAY41_contig06139.gene_1	2		#N/A			Uncharacterized
756	BBAY41_GBSR4EV02IJ3EU.gene_1	2		COG0499, SAM1, S-adenosylhomocysteine hydrolase [Coenzyme metabolism].			Amino acid biosynthesis
757	lcl|sequenceName|1108814259580|locusTag|ORF0002	2		#N/A			Uncharacterized
758	BBAY42_GBUJSMQ01AMSHF.gene_1	2			PF00378.13, Enoyl-CoA hydratase/isomerase family		Lipid metabolism
759	BBAY41_GBSR4EV01D59DD.gene_1	2		#N/A			Uncharacterized
760	BBAY42_GBUJSMQ01D1BC4.gene_1	2		COG1022, FAA1, Long-chain acyl-CoA synthetases (AMP-forming) [Lipid metabolism].			Coenzyme biosynthesis
762	BBAY42_GBUJSMQ02IUAUA.gene_1	2		PTZ00069, PTZ00069, 60S ribosomal protein L5; Provisional.			Translation
763	lcl|sequenceName|1106850786386|locusTag|ORF0001	2		COG0855:Polyphosphatekinase,4.1E-115"/anno3="polyphosphatekinase[EC:2.7.4.1];			Scavenging
764	BBAY40_contig12889.gene_2	2		COG0451, WcaG, Nucleoside-diphosphate-sugar epimerases			Transport
765	lcl|sequenceName|1108814257924|locusTag|ORF0020	2	Hep/Hagrepeatprotein,1.41034E-31"/anno2="hypotheticalproteinBpse4_03003057[Burkholderiapseudomallei406e],5.54379E-31"				Repeat protein
766	BBAY41_GBSR4EV01B1JZB.gene_1	2		COG0711, AtpF, F0F1-type ATP synthase, subunit b [Energy production and conversion].			ATP synthesis coupled proton transport
767	lcl|sequenceName|1106850471259|locusTag|ORF0001	2	NicotinamideN-methyltransferase-Susscrofa(Pig),7.31985E-23"/anno3="PREDICTED:similartonicotinamideN-methyltransferase[Bostaurus],8.50375E-23"				Electron transport