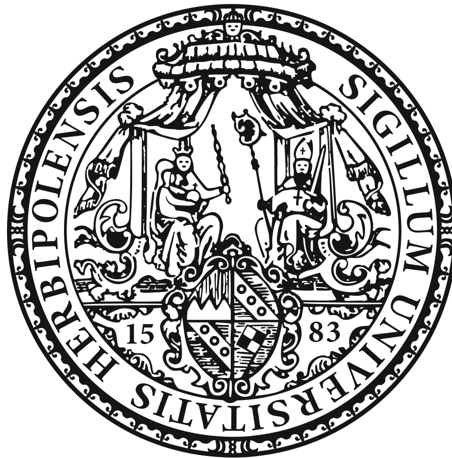


Exploration of microbial diversity and function in Red Sea sponges by deep sequencing

Untersuchungen zur mikrobiellen Diversität und Funktion in Schwämmen aus dem Roten Meer mittels Hochdurchsatz-Sequenzierung

Doctoral thesis for a doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section: Integrative Biology

submitted by

Lucas Moitinho e Silva

from

Cruz das Almas, Brazil

Würzburg, July 2014

Submitted on:

Members of the *Promotionskomitee*:

Chairperson: Prof. Dr. Jörg Schultz

Primary Supervisor: Prof. Dr. Ute Hentschel Humeida

Supervisor (Second): Prof. Dr. Timothy Ravasi

Supervisor (Third): Prof. Dr. Thomas Dandekar

Date of Public Defence: ………………………………………….

Date of Receipt of Certificates: ………………………………………….

Affidavit

I hereby confirm that my Thesis entitled “Exploration of microbial diversity and function in Red Sea sponges by deep sequencing” is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the Thesis.

Furthermore, I confirm that this Thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Place, Date Signature

Acknowledgements

I am profoundly grateful to the following:

Prof. Ute Hentschel for the guidance, believe, teachings, enthusiasm, and integral support along my PhD.

Prof. Timothy Ravasi for the supervision, friendship, and collaboration.

Prof. Thomas Dandekar for the supervision and support.

Prof. Roy Gross for accepting the invitation to examine my Thesis.

My current and former colleagues of Prof. Hentschel's laboratory for sharing their expertise, friendship, daily support, and mostly, for receiving me with open arms. Specially, I am grateful to Paula, Hilde, Sheila, Susanne, Franzi, Odette, Anne, Lorenz, Lars, Volker, Janine, Usama, Kristina, Tine, Eva, Martin, Hannes, Beate, Naseem, Cheng, and Janno.

My colleagues at the Department of Botany II for their agreeable company during these years, specially Jana, Katja, Anton, Wilma, Moni, Andrea, Ulrich, Anna, and Fernando.

The Graduate School of Life Sciences for granting me the fellowship and providing me with important structure for my professional development. Specially, I am grateful to Dr. Gabriele Blum-Oehler, Bianca Putz, and Jennifer Braysher.

My colleagues from KAUST for making my research stay in Saudi Arabia an amazing experience, specially Taewoo, Loqmane, Emily, Feras, Carlo, and Yanal.

Prof. Christian Voolstra for the scientific discussions and warm hospitality and Prof. Michael Berumen for the help during sample collections.

The Coastal and Marine Resources Core Lab and the Biosciences Core Laboratory at KAUST for making sample collections and sequencing possible.

Prof. Georg Krohne, Daniela Bunsen, and Claudia Gehrig at the Department of Electron Microscopy for the excellent expertise in microscopy.

Brigitte Neumann at the Department of Botany I for the friendly guide into the RNA analysis world.

My family and friends for the smiles, care, understanding, and unconditional support; specially, I am deeply grateful to my parents, my brother, and Verena.

Table of contents

Summary....................xi
Zusammenfassung....................xiii
1 Introduction....................16
1.1 Marine sponges....................16
1.2 Sponge morphology and physiology....................17
1.3 Sponge-associated microorganisms....................19
1.3.1 The HMA/LMA dichotomy....................20
1.3.2 Specificity of the association....................22
1.3.3 Stability of the association....................24
1.3.4 Vertical transmission....................25
1.3.5 Physiology....................25
1.3.6 Carbon metabolism....................27
1.3.7 Nitrogen metabolism....................28
1.4 Sponge-isolated bioactive compounds....................30
1.5 The Red Sea sponges: *Stylissa carteri* and *Xestospongia testudinaria*....................31
2 Research aims....................33
3 Materials and methods....................34
3.1 Sample collection....................34
3.2 Laboratory procedures....................36
3.2.1 Transmission electron microscopy....................36
3.2.2 Internal standard synthesis....................37
3.2.3 Nucleic acids isolation....................38
3.2.4 16S rRNA gene amplification and cloning....................39
3.2.5 16S rRNA gene and transcript amplification....................42
3.2.6 Enrichment of prokaryotic mRNA....................45
3.2.7 DNA sequencing....................46
3.3 Bioinformatic analyses....................46
3.3.1 Bioinformatic analysis hardware....................46
3.3.2 Data manipulation and processing....................47
3.3.3 Image, graph, and text editors....................47
3.3.4 Processing of 16S rRNA gene clone sequences....................47

3.3.5 Processing of 16S rRNA gene amplicon sequences...48
3.3.6 Assignment of OTUs to sponge-specific clusters or sponge- and coral-specific clusters..49
3.3.7 Assignment of OTUs to species-specific sequences...50
3.3.8 Analysis of microbial community diversity..50
3.3.9 Analysis of OTU overlap..51
3.3.10 OTU abundance pattern analysis...51
3.3.11 Analysis of transcriptionally active taxa...52
3.3.12 Processing of metatranscriptome sequences...52
3.3.13 Taxonomic annotation of metatranscriptome sequences....................................53
3.3.14 Functional annotation of metatranscriptome sequences.....................................54
3.3.15 Analysis of dissimilarity between replicate metatranscriptomes.........................54
3.3.16 Assembly of metatranscriptomic reads...55
3.3.17 Phylogenetic analysis of gene function...55
4 Results..57
4.1 The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species...57
4.1.1 Abstract...59
4.1.2 Introduction..60
4.1.3 Materials and Methods...62
4.1.4 Results and Discussion...66
4.1.5 Acknowledgements...78
4.1.6 References...78
4.2 Bacterial community profiles in low microbial abundance sponges.............................82
4.3 Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea...93
4.4 GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater...110
4.4.1 Abstract...112
4.4.2 Introduction..112
4.4.3 Methods...114
4.4.4 Results...118
4.4.5 Discussion..125
4.4.6 Acknowledgements..130

4.4.7 Supplementary Figure legends..130
4.4.8 References..130
4.5 Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics..134
4.6 Additional results...151
4.6.1 Clone libraries...151
4.6.2 Phylogenetic analysis of highly expressed key functional genes........................157
5 Discussion..162
5.1 Microbial communities of sponges..162
5.1.1 High-throughput sequencing..162
5.1.2 The HMA/LMA dichotomy...165
5.1.3 Diversity..167
5.1.4 Specificity...168
5.1.5 Composition and transcriptional activity...170
5.2 Functions of sponge microbiomes...173
5.2.1 Meta-omics..173
5.2.2 Environmental microarrays...174
5.2.3 Metatranscriptomics..175
5.2.4 Functional activities of *Stylissa carteri* microbiomes...177
6 Future directions...182
7 Bibliography...184
8 Appendix...205
9 Publications list..207
10 Curriculum Vitae...208

List of figures

Figure 1......18
Figure 2......21
Figure 3......23
Figure 4......28
Figure 5......32
Figure 6......35
Figure 7......71
Figure 8......72
Figure 9......77
Figure 10......119
Figure 11......121
Figure 12......122
Figure 13......123
Figure 14......124
Figure 15......125
Figure 16......152
Figure 17......154
Figure 18......160
Figure 19......170

List of tables

Table 1..36
Table 2..39
Table 3..40
Table 4..41
Table 5..42
Table 6..43
Table 7..44
Table 8..45
Table 9..63
Table 10..67
Table 11..69
Table 12..74
Table 13..153
Table 14..156
Table 15..158
Table 16..166

Summary

Marine sponges (phylum Porifera) are simple, sessile, filter-feeder animals. Microbial symbionts are commonly found in the sponge internal tissue, termed the mesohyl. With respect to the microbial content, sponges are classified as either low-microbial abundance sponges (LMA), or high-microbial abundance sponges (HMA). The HMA/LMA dichotomy was explored in this Thesis using the Red Sea sponges as experimental models. A range of methods encompassing transmission electron microscopy, 16S rRNA gene deep sequencing, and metatranscriptomics was employed towards this goal. Here, particular emphasis was placed on the functional analysis of sponge microbiomes.

The Red Sea sponges *Stylissa carteri, Xestospongia testudinaria, Amphimedon ochracea,* and *Crella cyathophora* were classified as HMA or LMA sponges using transmission electron microscopy. The diversity, specificity, and transcriptional activity of microbes associated with the sponges *S. carteri* (LMA) and *X. testudinaria* (HMA) and seawater were investigated using 16S rRNA amplicon pyrosequencing. The microbial composition of *S. carteri* was more similar to that of seawater than to that of *X. testudinaria*, which is consistent with the observation that the sequence data set of *S. carteri* contained many more possibly seawater sequences (~24%) than the *X. testudinaria* data set (~6%). The most abundant operational taxonomic units (OTUs) were shared between all three sources (*S. carteri*, *X. testudinaria,* seawater), while rare OTUs were unique to any given source. Despite this high degree of overlap, each sponge species contained its own specific microbiota. *S. carteri* microbiomes were enriched of Gammaproteobacteria and members of the genus *Synechococcus* and *Nitrospira*. Enriched members of *X. testudinaria* microbiomes included Chloroflexi, Deferribacteres, and Actinobacteria. The transcriptional activity of sponge-associated microorganisms was assessed by comparing 16S rRNA gene with transcript amplicons, which showed a good correlation.

The microbial functional gene repertoire of sponges and seawater from the Red Sea (*X. testudinaria, S. carteri*) and the Mediterranean (*Aplysina aerophoba, Dysidea avara)* were investigated with the environmental microarray GeoChip 4. Amplicon sequencing was performed alongside in order to assess microbial diversity. The typical microbial diversity patterns characteristic of HMA (abundance of Gammaproteobacteria, Chloroflexi,

Acidobacteria, Deferribacteres, and others) and LMA sponges (abundance of Alpha-, Beta-, Gammaproteobacteria, Cyanobacteria, and Bacteroidetes) were confirmed. The HMA/LMA dichotomy was stronger than any possible geographic pattern based on microbial diversity (amplicon) and functional genes (GeoChip). However upon inspection of individual genes detected by GeoChip, very few specific differences were discernible, including differences related to microbial ammonia oxidation, ammonification (higher gene abundance in sponges over seawater) as well as denitrification (lower gene abundance). Furthermore, a higher abundance of a gene, *pcc*, representative of archaeal autotrophic carbon fixation was noted in sponges over seawater. Thirdly, stress-related genes, in particular those related to radiation, were found in lower abundances in sponge microbiomes than in seawater. With the exception of few documented specific differences, the functional gene repertoire between the different sources appeared largely similar.

The most actively expressed genes of *S. carteri* microbiomes were investigated with metatranscriptomics. Prokaryotic mRNA was enriched from sponge total RNA, sequenced using Illumina HiSeq technology, and annotated with the metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline. High expression of archaeal ammonia oxidation and photosynthetic carbon fixation by members of the genus *Synechococcus* was detected. Functions related to stress response and membrane transporters were among the most highly expressed by *S. carteri* symbionts. Unexpectedly, gene functions related to methylotrophy were highly expressed by gammaproteobacterial symbionts. The presence of seawater-derived microbes is indicated by the phylogenetic proximity of organic carbon transporters to orthologs of members from the SAR11 clade. In summary, the most expressed functions of the *S. carteri*-associated microbial community were revealed and linked to the dominant taxonomic members of the microbiome.

In conclusion, HMA and LMA Red Sea sponges were used as models to gain insights into relevant themes in sponge microbiology, i.e. diversity, specificity, and functional activities. Overall, my Thesis contributes to a better understanding of sponge-associated microbial communities, and the implications of this association to marine ecology.

Zusammenfassung

Marine Schwämme (phylum Porifera) sind einfache, sessile, sich mittels Filtration von Meerwasser ernährende Tiere. Das Schwammgewebe, als Mesohyl definiert, ist häufig durch mikrobielle Symbionten besiedelt. Im Hinblick auf die mikrobielle Abundanz werden Schwämme in sogenannte „low-microbial abundance“ (LMA) oder „high microbial abundance“ (HMA) Kategorien unterteilt. Im Rahmen dieser Doktorarbeit wurde die „HMA/LMA-Dichotomie“ anhand von Schwämmen aus dem Roten Meer untersucht. Verschiedene Methoden, wie die Transmissions-Elektronenmikroskopie, 16S rRNA Gen-Hochdurchsatz-Sequenzierung sowie die Metatranskriptomik kamen zum Einsatz. Ein besonderes Augenmerk wurde auf die funktionale Analyse der Schwammsymbionten gelegt.

Die Schwämme *Stylissa carteri*, *Xestospongia testudinaria*, *Amphimedon ochracea* und *Crella cyathophora* aus dem Roten Meer wurden zunächst mittels Transmissions-Elektronenmikroskopie als HMA oder LMA-Schwämme klassifiziert. Die Diversität, Spezifizität sowie transkriptionelle Aktivität von mit *S. carteri* (LMA), *X. testudinaria* (HMA) oder Meerwasser-assoziierten Mikroorganismen wurde mittels 16S rRNA Gen-Ampliconsequenzierung untersucht. Die mikrobielle Zusammensetzung von *S. carteri* ähnelte mehr der des Meerwassers als der von *X. testudinaria*, was mit dem Befund übereinstimmt, dass der Sequenzdatensatz von *S. carteri* deutlich mehr Meerwassersequenzen enthielt (~ 24%) als der von *X. testudinaria* (6%). Die am häufigsten vorliegenden „operational taxonomic units“ (OTUs) lagen gleichermaßen im Probenmaterial (*S. carteri*, *X. testudinaria*, Meerwasser) vor, während die am wenigsten häufigen OTUs jeweils spezifisch für eine Probe waren. Trotz der hohen Gemeinsamkeiten enthielt jede Schwammart ein eigenes Bakterienprofil. Die Mikrobiome von *S. carteri* waren durch Gammaproteobacteria, sowie Vertreter der Gattung *Synechococcus* und *Nitrospira* angereichert. Häufige Vertreter des *X. testudinaria* Mikrobioms waren Chloroflexi, Deferribacteres und Actinobacteria. Die transcriptionelle Aktivität von Schwamm-assoziierten Mikroorganismen wurde auf der Basis von 16S rRNA-Genen und 16S rRNA-Transkripten verglichen und zeigte eine gute Korrelation.

Das funktionale Genrepertoire von Schwämmen und Meerwasser aus dem Roten Meer (*X. testudinaria*, *S. carteri*) und dem Mittelmeer (*Aplysina aerophoba*, *Dysidea avara*) wurde

mittels des Mikroarrays GeoChip 4 verglichen. Die Ampliconsequenzierung wurde begleitend zur Charakterisierung der mikrobiellen Diversität durchgeführt. Die charakteristischen Diversitätsmuster von HMA-Schwämmen (Gammaproteobacteria, Chloroflexi, Acidobacteria, Deferribacteres und weitere) und LMA-Schwämmen (Alpha-, Beta-, Gammaproteobacteria, Cyanobacteria, Bacteroidetes) konnten bestätigt werden. Die HMA/LMA-Dichotomie war sowohl im Bezug auf die Diversität (Amplikon-Daten) als auch auf das funktionale Genrepertoire (GeoChip) deutlich stärker ausgeprägt als ein mögliches geographisches Muster. Jedoch konnten nach Analyse einzelner Gene nur wenige spezifische Unterschiede definiert werden. Diese betrafen beispielsweise die mikrobielle Ammonium-Oxidation, Ammonifikation (höhere Genabundanz in Schwämmen als Meerwasser) oder die Denitrifikation (niedrigere Genabundanz in Schwämmen als Meerwasser). Weiterhin wurde eine höhere Genabundanz des *pcc* -Gens in Schwämmen als im Meerwasser beschrieben. Dieses Gen gilt als Indikator für archaeale, autotrophe Kohlenstoff-Fixierung. Drittens wurde eine niedrige Genabundanz von Stress-Genen, insbesondere im Bezug auf UV-Strahlung, in Schwämmen als im Meerwasser beobachtet. Von wenigen spezifischen Ausnahmen abgesehen war das funktionale Genrepertoire zwischen dem Probenmaterial jedoch sehr ähnlich.

Die am häufigsten exprimierten Gene des *S. carteri* Mikrobioms wurden mittels Metatranskriptomik untersucht. Zu diesem Zweck wurde die prokaryotische mRNA aus der Gesamt-Schwamm RNA angereichert, mittels Illumina HiSeq-Technologie sequenziert und mittels der MG-RAST-Software annotiert. Die mit am häufigsten exprimierten Gene betrafen die archaeale Ammonium-Oxidation und die photosynthetische Kohlenstoff-Fixierung durch *Synechococcus*. Weiterhin waren Stress- und Membrantransport-relevane Gene mit am häufigsten exprimiert. Unerwartet war der Befund, dass Methylotrophie-verwandte Gene ebenfalls sehr häufig exprimiert wurden. Das Vorliegen von Meerwasserbakterien ist durch die phylogenetische Nähe von organischen Kohlenstoff-Transportern zu Genen der SAR11-Klade belegt. Zusammenfassend wurden in dieser Studie die am häufigsten transkribierten Funktionen des *S. carteri* Mikrobioms beschrieben und bezüglich ihrer taxonomischen Zugehörigkeit analysiert.

Zusammenfassend wurden HMA und LMA-Schwämme aus dem Roten Meer als experimentelle Modellsysteme verwendet, um Einblicke in für die Schwamm-Mikrobiologie relevante Fragen bezüglich der mikrobiellen Diversität, Spezifität und funktionalen Aktivität zu gewinnen. Diese PhD-Arbeit trägt zu einem verbesserten Verständnis der Mikrobiologie

von Schwämmen sowie deren Bedeutung für die marine Ökologie im Allgemeinen bei.

1 Introduction

1.1 Marine sponges

Sponges (phylum Porifera) are sessile animals and populate diverse aquatic habitats, such as tropical seawater (Hooper and Lévi, 1994), mangroves (Engel and Pawlik, 2005), freshwater (Erpenbeck *et al.*, 2011), polar seawater (McClintock *et al.*, 2005), and the deep-sea environment (Witte *et al.*, 1997). Sponges have relevant functional roles (Bell, 2008; de Goeij *et al.*, 2013), including (i) the impact on substrates, such as causing bioerosion (Ginsburg, 1957) and aiding in reef growth processes (Wulff, 1984; Wulff, 2001); (ii) the participation in nutrient cycles (Diaz and Rützler, 2001; Yahel *et al.*, 2003; Fiore *et al.*, 2013), and (iii) the interaction with other organisms (Wulff and Buss, 1979; Ellison *et al.*, 1996; Macdonald *et al.*, 2006; Simister *et al.*, 2012a). Sponges have been economically explored as bath sponges since early Greek civilization (Hooper and van Soest, 2002) and currently represent one of the most prominent marine sources of bioactive compounds (Laport *et al.*, 2009).

Sponges are primitive multicellular eukaryotes, with a fossil record dating back at least 580 million years (Li *et al.*, 1998). It is estimated that over 15.000 sponge species exist worldwide (Hooper and Lévi, 1994; Hooper and van Soest, 2002). The phylum Porifera is formed of 4 classes: (i) Hexactinellida, known as glass sponges, (ii) Demospongiae (demosponges), (iii) Calcarea (calcareous sponges), and (iv) Homoscleromorpha, considered previously as a subclass of Demospongiae (Gazave *et al.*, 2012). Among these, Demospongiae is the most prominent class, containing about 85% of the living sponge species described (Hooper and van Soest, 2002). With regard to sponge phylogeny, a consensus has not been reached whether Porifera is monophyletic, meaning that all species originated from the same ancestor species, or whether it is paraphyletic, in the sense that the common ancestor of all sponge species would also be the ancestor of non-sponge species (Wörheide *et al.*, 2012; Dohrmann and Wörheide, 2013). Sponge paraphyly implies that all higher animals, including humans, are descendants of a sponge-like ancestor (Dohrmann and

Wörheide, 2013).

1.2 Sponge morphology and physiology

Sponges are found in various body forms, including spherical, encrusting, tubular, vase-shaped, or branching shapes. They are also found in various sizes, ranging from millimetres, as observed for the boring sponge *Cliona celata*, to metres in size, as observed for the giant barrel sponge *Xestospongia muta*. Most of sponges are filter-feeders, with the exception of carnivorous species that dwell in nutrient-poor deep-sea environments (Vacelet and Boury-Esnault, 1995; Lee *et al.*, 2012). The general body form of filter-feeder species is specialized in filtering water for nutrition, respiration, and excretion. The incoming water flows through the sponge body via a complex system of water current channels termed the aquiferous system (Brusca and Brusca, 2003). Depending on the aquiferous system design, sponges are classified into: (i) asconoid, when the sponge body has a hollow cylinder shape, whose wall is perforated by small pores; (ii) syconoid, when the sponge body wall is composed of alternating in- and out-pockets that increase the rate between area and volume; or (iii) leuconoid, when the sponge body is formed by water channels composed of many spherical chambers connected by minute capillary-like vessels (Ruppert, 2004) (Figure 1). The aquiferous system of demosponge filter-feeders is leuconoid (Hooper and van Soest, 2002).

The interpretation that sponge cells are organized into tissues is subject to different opinions: while some authors consider that sponges have no tissues (Brusca and Brusca, 2003), others argue that the cell organization of sponges is sufficiently complex to be interpreted as tissues or tissue-like structures (Ruppert, 2004; Leys and Hill, 2012). In this Thesis, the sponge cells are viewed as being organized into tissue as described by Ereskovskii (2010). This interpretation is favoured because of the evidence that sponges possess quite a high level of cellular complexity (Leys *et al.*, 2009). However, it is important to acknowledge that sponge tissues are structurally and functionally simpler, and less specialized than other animals (Ereskovskii, 2010).

The body of sponges is organized into the epithelioid and the connective tissues (Ruppert,

2004) (Figure 1). The epithelioid tissues are bordering structures that separate the connective tissue from the environment. These tissues lack a basal membrane, what is a characteristic of the epithelium of eumetazoans (Ereskovskii, 2010). Epithelioid tissues are classified in the pinacoderm and the choanoderm. In demosponges, the pinacoderm is formed by a layer of cells, the pinacocytes. Pinacocytes are named according to their function or localization as: (i) exopinacocytes, which cover externally the sponges; (ii) basopinacocytes, which attach the sponges to a substrate; or (iii) endopinacocytes, which line the water channels (Ereskovskii, 2010). In some demosponges, the pinacoderm can mediate the body contraction (Nickel *et al.*, 2011). Choanoderm is the other type of epithelioid tissue, which forms chambers of flagellated cells (choanocyte chambers) that generate the water flow through the sponge. Other cell types are also found forming the chambers, such as cells that participate in the regulation of choanocyte flagella beating (central cells) (Reiswig and Brown, 1977) and cells that are placed between the pinacocytes and the choanocytes (cone cells in demosponges) (Langenbruch, 1988).

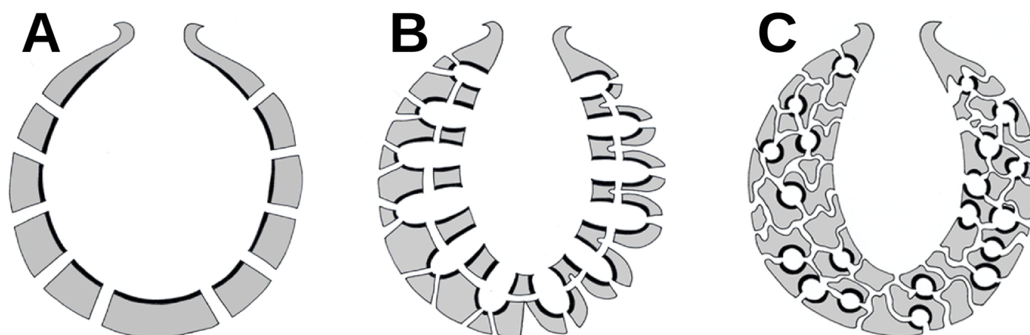


Figure 1. Sponge body structures and tissues. The three different aquiferous system designs of sponges are (A) asconoid , (B) syconoid , and (C) leuconoid . The sponge connective tissue, the mesohyl (in grey), is bordered by the epithelioid tissues, the pinacoderm and the choanoderm (in black). Figure modified from Cavalcanti and Klautau (2011).

The sponge connective tissue (mesohyl) is composed of various cell types and skeleton elements loosely embedded in an extracellular matrix (Ereskovskii, 2010). With regard to the functional specialization of cell groups, the mesohyl can be classified into two tissues analogous of other animals (Korotkova, 1981): (i) the supportive-connective tissue, whose main function is the formation of organic skeleton [e.g. collagen (Borojevic, 1966), spongin

(Garrone, 1978), and glycoprotein matrix (Garrone, 1978; Bonasoro *et al.*, 2001)] and mineral skeleton which in demosponges is composed of siliceous spicules (Simpson and Vaccaro, 1974; Hooper and van Soest, 2002); and (ii) the protective-secretory tissue, whose main functions are protection, storage, secretion of the mesohyl matrix, and transference of food particles and oxygen (Ereskovskii, 2010). The protective-secretory tissue contains various cell types of which abilities include totipotency (Funayama, 2008), storage of toxic metabolites (Uriz *et al.*, 1996), and participation in reproduction (Ereskovskii and Gonobobleva, 2000; Gaino *et al.*, 2006).

As filter-feeders, marine sponges process many thousands litres of seawater a day. The sizes of particles filtered by sponges are estimated to be in the range of 70 µm to 0.3 µm (Reiswig, 1971; Ribes *et al.*, 1999). Therefore, the sponge diet is generally composed of bacteria, nano-, and picoeukaryotes (Reiswig, 1971; Pile *et al.*, 1996; Bell *et al.*, 1999; Ribes *et al.*, 1999; Pile *et al.*, 2003). Smaller organic compounds, such as dissolved organic carbon (DOC), are also taken up during filtration by sponges and/or putatively by their associated microbes (Schmidt, 1970; Reiswig, 1971; Reiswig, 1981; Reiswig, 1990; Yahel *et al.*, 2003; de Goeij *et al.*, 2008b; de Goeij *et al.*, 2013). The food is taken up by choanocytes and pinacocytes via cellular processes such as phagocytosis and pinocytosis (Schmidt, 1970; Ruppert, 2004; Maldonado *et al.*, 2012). The food is then transferred to mesohyl cells via transcytosis and digested by archaeocytes, which are mobile amoeboid mesohyl cells. Then, the products of food digestion are distributed to other mesohyl cells (Maldonado *et al.*, 2012). Finally, undigested particles and excretory products are expelled from sponges through exhalant channels or to the external surface via exocytosis (Schmidt, 1970; Ruppert, 2004; Maldonado et al., 2012).

1.3 Sponge-associated microorganisms

Organisms from the three domains of life (Bacteria, Archaea, and Eukaryota) are symbiotically associated with sponges. Symbiosis is here used in the sense defined by Anton de Bary as “the living together of unlike named organisms”, as quoted by Sapp (1994). Macroscopically, sponges are symbiotically associated with a great diversity of organisms,

including corals (Wulff and Buss, 1979), shrimps (Macdonald *et al.*, 2006), decorator crabs (Stachowicz and Hay, 2000), and mangrove trees (Ellison *et al.*, 1996). Symbiosis is also observed between different sponge species (Wulff, 2008).

The observation of symbiosis between microbes and sponges dates back to late 1930’s (Dosse, 1939). Microbial symbionts are found inside (Vacelet, 1975) and outside of sponge cells, located in the mesohyl (Vacelet, 1975) or in the outer layers of the sponge body (Wilkinson, 1992). To date, more than 28 bacterial phyla including candidate phyla and two archaeal lineages were identified in marine sponges (reviewed by Hentschel *et al*., 2012). The dominant prokaryotic lineages include the archaeal phylum Thaumarchaeota and the bacterial phyla Proteobacteria, Chloroflexi, Actinobacteria, Acidobacteria, Nitrospirae, and candidate phylum Poribacteria (Hentschel *et al.*, 2012). Here, terms that are related to symbiotic microorganisms are generally used in reference to prokaryotic symbionts, except when otherwise indicated.

1.3.1 The HMA/LMA dichotomy

In terms of microbial content, marine sponges are classified as low-microbial-abundance sponges (LMA), also termed as non-bacteriosponges, or high-microbial-abundance sponges (HMA), also termed as bacteriosponges (Vacelet and Donadey, 1977; Reiswig, 1981; Hentschel *et al.*, 2003) (Figure 2). In both types of associations, the microorganisms are generally found extracellularly in the mesohyl. The microbial density of LMA sponges is similar to that of seawater, within the range of 10^5 to 10^6 bacteria/g or mL of sponge. On the other hand, HMA sponges present a much denser microbial community, that ranges from 10^8 to 10^{10} bacteria/g or mL of sponge (Hentschel *et al.*, 2006). In HMA sponges, microbial biomass can represent up to 35% of the animal’s biomass (Hentschel *et al.*, 2012). With respect to the diversity of microbial symbionts, LMA sponge microbiomes are dominated by single or few morphotypes/taxa, while HMA microbiota present a higher diversity of morphotypes/taxa (Vacelet and Donadey, 1977; Kamke *et al.*, 2010; Schmitt *et al.*, 2012b; Giles *et al.*, 2013a). Furthermore sponge-specific clusters are rarely found in LMA sponges (Erwin *et al.*, 2011; Giles *et al.*, 2013a). Interestingly, both types of sponges often co-occur in

the same location (Hentschel *et al.*, 2006).

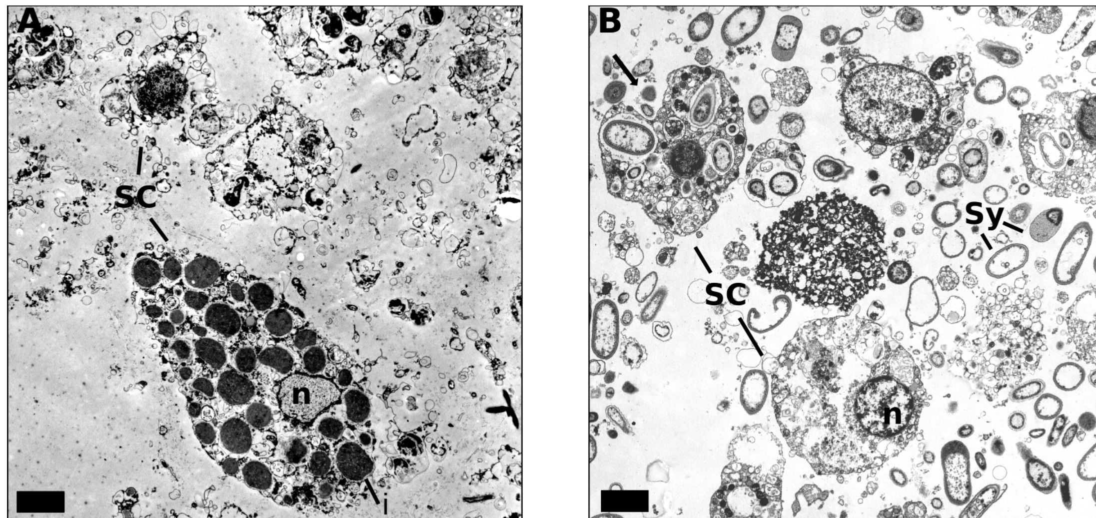


Figure 2. Transmission electron microscopy of the (A) low-microbial-abundance sponge *Stylissa carteri* and the (B) high-microbial-abundance sponge *Xestospongia testudinaria*. The *S. carteri* mesohyl is free of microorganisms. The *X. testudinaria* mesohyl is densely populated by different morphotypes of (Sy) microbial symbionts. (SC) Sponge cells and (n) nucleus are identified. (i) Inclusions are observed in a *S. carteri* cell. The arrow indicates a *X. testudinaria* cell containing microbial symbionts. The scale bars indicate 2 µm.

Morphological and physiological differences were observed between LMA and HMA sponges. For example, the LMA sponges appears to consume less oxygen in comparison with HMA sponges (Reiswig, 1974), as reported for Caribbean sponges under natural conditions. Additionally, LMA sponges mainly feed on particulate organic matter (POM) (Reiswig, 1974; Weisz *et al.*, 2007), while HMA sponges appear to also consume dissolved organic matter (DOM) (Reiswig, 1974). LMA sponges have bigger choanocyte chambers, shorter and wider water canals, and greater pumping rates in comparison with HMA sponges (Vacelet and Donadey, 1977; Boury-Esnault *et al.*, 1990; Weisz *et al.*, 2008). Weisz *et al*. (2008) hypothesized that these morphological and physiological differences resulted from the selection for the LMA type that favours efficient filtration of POM from a large quantity of water and the HMA type that creates an internal environments favourable for high density and diversity of microorganisms. So far, the mechanisms responsible for shaping the HMA/LMA dichotomy are not well understood.

1.3.2 Specificity of the association

The specificity of the association between sponges and microorganisms is a major theme in sponge microbiology. Earlier, Wilkinson (1978) investigated this topic based on metabolic and structural characterization of bacterial strains. Because the great majority of microorganisms observed via microscopy can not be cultured using current techniques (Staley and Konopka, 1985; Amann *et al.*, 1995), this theme is now investigated with molecular tools based on the use of 16S ribosomal RNA (rRNA) gene as a phylogenetic marker, in a culture-independent manner. Based on the phylogenetic analysis of 16S rRNA gene sequences, Hentschel *et al*. (2002) observed that the sponges *Theonella swinhoei* and *Aplysina aerophoba,* which belong to different phylogenetic orders and have a non-overlapping geographic distribution, harbour a uniform and phylogenetically complex microbial population. In order to identify microorganisms specifically associated with sponges, Hentschel *et al*. (2002) defined sponge-specific clusters (SC) as clusters formed by at least three 16S rRNA gene sequences recovered from sponges, but not from other sources, originating from different species and/or from different geographical locations. In a recent revision of the SC concept, more than 7,500 16S rRNA gene sequences derived from sponges were phylogenetically analysed, resulting in the classification of 27% of these sequences into SC (Simister *et al.*, 2012a) (Figure 3). Sponge-specific clusters were identified in at least 14 bacterial phyla, 2 archaeal phyla and in Eukaryota (using 18S rRNA gene as a phylogenetic marker). Some bacterial phyla, including Spirochaetes and the candidate phylum Poribacteria, presented a high proportion of 16S rRNA gene sequences in SC (79% and 92%, respectively). Furthermore, 16S rRNA gene sequences were described to form monophyletic clusters composed exclusively of sponge- and coral-derived sequences, termed as sponge- and coral-specific clusters (SCC) (Taylor *et al.*, 2007; Simister *et al.*, 2012a).

The recent advent of high-throughput sequencing methods showed a different perspective of the sponge-specific clusters concept. These methods allow an unprecedented sequencing depth at relative low costs, with the number of sequences ranging up to millions per sample. In recent studies using 454 amplicon sequencing, the SC-like 16S rRNA gene sequences were recovered from seawater (Webster *et al.*, 2010; Lee *et al.*, 2011) and other non-sponge environments (Taylor *et al.*, 2013), although at low abundances. These findings are in

contrast with the concept of sponge-specific clusters, which was proposed to describe phylogenetically similar sequences recovered from sponges, but not from other sources. Currently, the biological or technical interpretations of the recent findings of SC/SCC-like sequences in non-sponge environments are not fully explored.

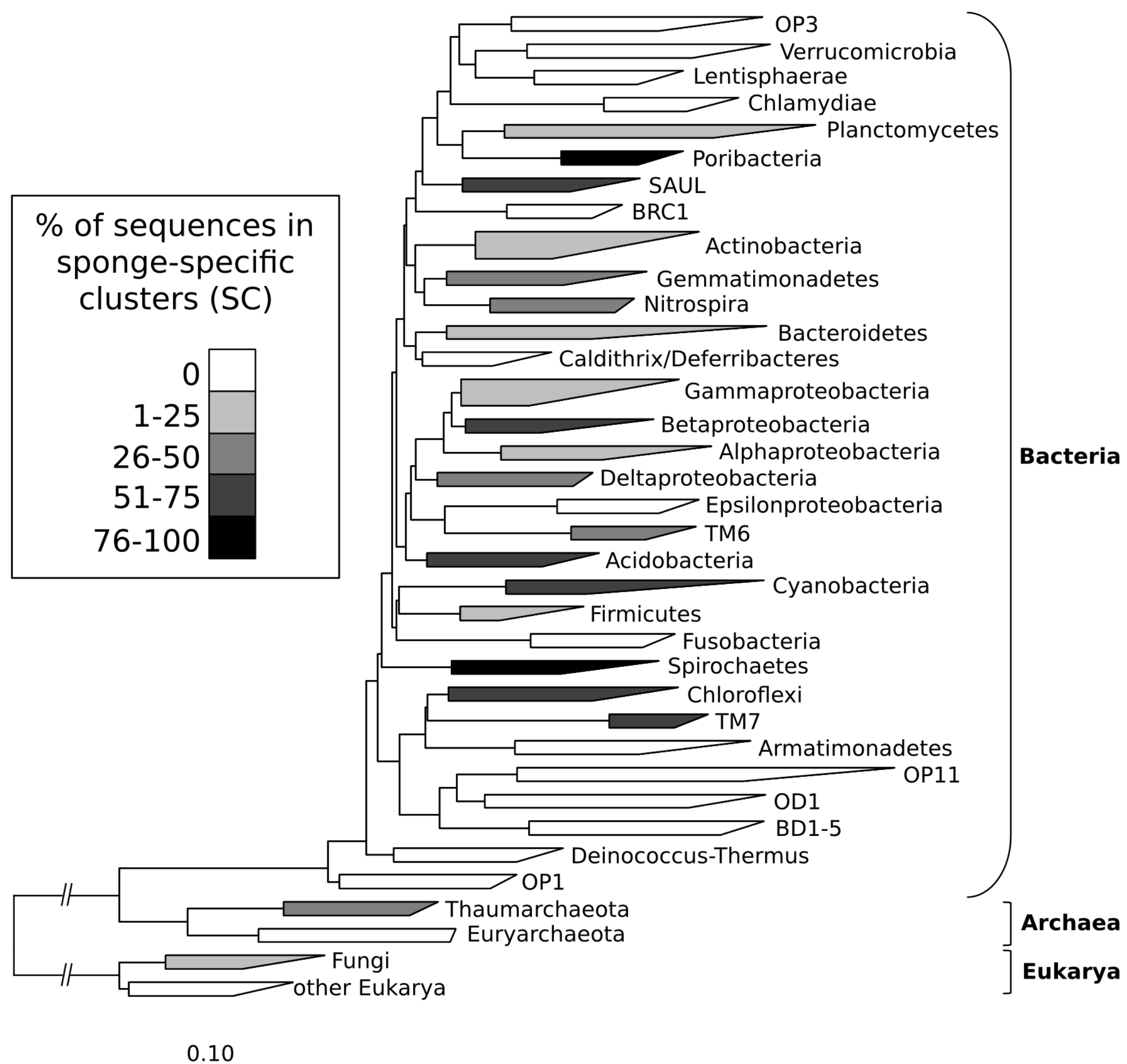


Figure 3. Phylogeny of sponge-associated microorganisms based on the small subunit rRNA genes, the 16S and the 18S. Proportion of sponge-specific clusters (SC) is shown for each representative taxon. Phylogenetic analysis was performed by Simister *et al*. (2012a). Figure modified from Hentschel *et al*. (2012).

It is believed that sponges bear species-specific populations of microbial symbionts. Early, Vacelet and Donadey (1977) discussed whether there are bacteria specifically

associated to a given sponge species based on electron microscopy. Based on the Sanger sequencing of the 16S rRNA gene, several publications reported species-specific lineages in sponge-associated microbiota (Thacker and Starnes, 2003; Erwin *et al.*, 2011; Giles *et al.*, 2013a; Webster *et al.*, 2013). Additionally, studies using 454 pyrosequencing found species-specific microbial populations (Lee *et al.*, 2011; Schmitt *et al.*, 2012a; Schmitt *et al.*, 2012b). Some of these considered up to 70% of the 16S rRNA gene sequences recovered from some sponge microbiomes as species-specific (Schmitt *et al.*, 2012a; Schmitt *et al.*, 2012b). A technical point not discussed in these papers is whether the sequencing depth achieved was enough to claim species-specificity of microbes associated with the investigated sponges.

1.3.3 Stability of the association

Sponge-associated microbes are largely conserved within different individuals of the same species (Taylor *et al.*, 2004; Webster *et al.*, 2013), and are largely stable through time and space. With respect to the temporal stability of sponge microbiomes, two publications reported little variation on the microbial communities of the cold water sponge *Geodia barretti* and the Mediterranean sponge *A. aerophoba* during respectively 8 and 6 months of incubation in aquarium (Hoffmann *et al.*, 2006; Gerçe *et al.*, 2009). The microbial communities of species from the genus *Aplysina* were stable under starvation, antibiotic treatment, and in transplantation experiments (Friedrich *et al.*, 2001; Thoms *et al.*, 2003). Different sponge species harbour different proportions of permanent members of their microbiomes, as indicated by assessment of samples collected during a period of 1 or 2 years (Wichels *et al.*, 2006; White *et al.*, 2012; Björk *et al.*, 2013; Simister *et al.*, 2013). Therefore, the temporal stability of sponge symbionts might be related with the sponge species and/or with the associated microbial species. With regard to the spatial variability, little variation of the microbial community associated with sponges from Antarctica, Australia, and the Mediterranean sea was observed between individuals separated up to hundreds of kilometres of distance (Webster *et al.*, 2004; Taylor *et al.*, 2005; Pita *et al.*, 2013).

1.3.4 Vertical transmission

Vertical transmission of microorganisms is an important mechanism for the conservation and specificity of the sponge microbial consortium. Several publications applying different methodologies, including electron microscopy (Lévi and Porte, 1962; Lévi and Lévi, 1976; Ereskovsky *et al.*, 2005) and fluorescence *in situ* hybridisation (FISH) (Enticknap *et al.*, 2006), showed the presence of microorganisms in sponge reproductive stages. Furthermore, studies based on 16S rRNA gene sequencing demonstrated clearly the transmission of bacteria and archaea lineages from adults to the offspring (Schmitt *et al.*, 2007; Sharp *et al.*, 2007; Schmitt *et al.*, 2008; Steger *et al.*, 2008; Lee *et al.*, 2009; Webster *et al.*, 2010; Gloeckner *et al.*, 2013). Cospeciation, which is a strong evidence of coevolution (Futuyma and Slatkin, 1983), is observed in nature in many ancient relationships between hosts and microbial symbionts (Peek *et al.*, 1998; Noda *et al.*, 2007; Kikuchi *et al.*, 2009). So far, strong evidence for cospeciation of sponges and associated microbial consortia was not found (Schmitt *et al.*, 2012b; Webster *et al.*, 2013). Exceptionally, Montalvo and Hill (2011) found genus-specific lineages of bacteria associated with two geographically distant sponge species, *X. testudinaria* and *X. muta*. The apparent lack of cospeciation between sponges and microbial symbionts is argued as a consequence of the putative acquisition of symbiotic microorganisms from seawater (Taylor *et al.*, 2007; Schmitt *et al.*, 2008; Webster *et al.*, 2010). Even though the combination of vertical and horizontal acquisition of sponge symbionts is the most likely scenario (Taylor *et al.*, 2007; Schmitt *et al.*, 2008), strong experimental evidence was not yet obtained for the horizontal acquisition and maintenance of the acquired microbial populations.

1.3.5 Physiology

The study of microbial functions in sponge symbiosis is still in its infancy due to the fact that the vast majority of sponge-associated microorganisms has not been cultured and due to the lack of a model system for studying this symbiosis (Taylor *et al.*, 2011). However, as reviewed by Taylor *et al*. (2007), several publications have analysed or inferred metabolic

activities of the sponge symbionts from pure-culture studies, analysis of pathways by measurement of specific products and genes, and inference of metabolic potentials based on 16S rRNA gene sequence analysis. The sponge symbionts metabolism reviewed included (i) carbon metabolism, such as heterotrophy, photosynthesis, and methane oxidation; (ii) nitrogen metabolism, including nitrification and nitrogen fixation; (iii) sulfur metabolism, including sulfate reduction and sulfate oxidation; (iv) dehalogenation; and (v) production of secondary metabolites.

Furthermore, the use of multi-“omics” approaches (including single-cell genomics, metagenomics, metatranscriptomics, and metaproteomics) have contributed to build an integrated view of the metabolic pathways postulated to be of relevance in the sponge microbiome. A recent study comparing the functional profiles of the microbes associated with several sponge species with the profile of seawater microbes found evidence of equivalence and evolutionary convergence for the sponge microbial consortia (Fan *et al.*, 2012b). Based on metagenomic data of 6 sponge species, this work presented common gene functions that were enriched in the sponge microbial consortia in comparison with seawater and that were provided by analogous enzymes (Fan *et al.*, 2012b). Most of these functions or related genes were analysed by other authors and include: (i) anaerobic nitrogen metabolism, in particular denitrification (Hoffmann *et al.*, 2009; Liu *et al.*, 2012; Fiore *et al.*, 2013) and ammonia oxidation (Radax *et al.*, 2012b; Fiore *et al.*, 2013); (ii) resistance to stress (Regoli *et al.*, 2000; Liu *et al.*, 2012); (iii) mobile genetic elements, largely transposases (Thomas *et al.*, 2010; Siegl *et al.*, 2011; Liu *et al.*, 2012; Radax *et al.*, 2012b); and (iv) eukaryotic-like proteins that putatively interacts with the host, in particular proteins containing ankyrin repeats and tetratricopeptide repeats (Thomas *et al.*, 2010; Liu *et al.*, 2012; Radax *et al.*, 2012b). The power of “omics” approaches to assess the expression of the functional repertoire of sponge symbionts was revealed by two pioneering publications (Liu *et al.*, 2012; Radax *et al.*, 2012b), leading to new possibilities to explore the functions of sponge symbionts.

Because carbon and nitrogen transformations are relevant in microbial ecology, their ecological implications to the symbiosis will be discussed in detail in the following Sections (1.3.6, 1.3.7).

1.3.6 Carbon metabolism

Carbon fixation is defined as the conversion of inorganic carbon (CO_2) to organic carbon compounds. The fixation of carbon by photosynthetic cyanobacteria is a major source of energy for some sponges (Cheshire and Wilkinson, 1991; Steindler *et al.*, 2002), e.g. the Great Barrier Reef sponge *Phyllospongia lamellosa.* Harbouring photosynthetic cyanobacteria is a successful strategy of sponges to live in tropical reefs, which are commonly oligotrophic areas with high incidence of light (Wilkinson, 1983; Wilkinson, 1987). Cyanobacteria are typically found in the outer layers of sponges, but can also be distributed in the inner core (Hentschel *et al.*, 2006). The most prominent lineage of sponge-associated cyanobacteria is *Synechococcus spongiarum*, of which a sponge-specific cluster contained 245 sequences from 40 sponge species (Simister *et al.*, 2012a). In a proposed mutualistic model (Taylor *et al.*, 2007), cyanobacteria would benefit from a sheltered environment and a constant nutrient supply (Sarà, 1971). The host would benefit from photosynthates, mainly glycerol and glycerol 3-phosphate (Arillo *et al.*, 1993), and UV-protection (Sarà, 1971; Bandaranayake *et al.*, 1996). Furthermore, the host could potentially acquire nutrients via the phagocytosis of cyanobacteria (Usher *et al.*, 2001; Oren *et al.*, 2005; Taylor *et al.*, 2007).

The capability of sponge symbionts to metabolise organic carbon compounds was inferred from publications reporting dissolved organic carbon (DOC) assimilation by sponges (Yahel *et al.*, 2003; de Goeij *et al.*, 2008a). Carbohydrate degradation was proposed as a nutritional basis for the sponge symbiosis with microbes in a recent genomic analysis of single bacterium cells of the candidate phylum Poribacteria (Kamke *et al.*, 2013). Poribacterial cells have the enzymatic repertoire required for nutrition by degradation of DOC components and/or by degradation of compounds from the extracellular matrix of the sponge host (Kamke *et al.*, 2013). Metabolism based on one-carbon organic compounds, in particular methane, was only reported in the association of methanotrophs and deep-sea carnivorous sponges (Vacelet *et al.*, 1995; Vacelet *et al.*, 1996; Vacelet and Boury-Esnault, 2002). The nature of the symbiotic association of sponges with prokaryotes capable of metabolising organic carbon compounds is possibly mutualistic, where the sponge cells nutritionally benefit by the digestion of the symbiont (Vacelet *et al.*, 1996) and the symbionts

benefit from the host micro-environment, or comensalistic, where there is no harm or benefit for the sponges (Kamke *et al.*, 2013).

1.3.7 Nitrogen metabolism

The availability of nitrogen is limited in tropical reefs. Therefore, the efficiency in utilization and recycling different forms of nitrogen is crucial for the evolutionary success of a given microorganism in such oligotrophic environments. As reviewed recently (Fiore *et al.*, 2010), all major pathways of the nitrogen cycle were detected in sponges using several different methodologies (Figure 4) (Weisz *et al.*, 2007; Bayer *et al.*, 2008; Mohamed *et al.*, 2008; Hoffmann *et al.*, 2009; Mohamed *et al.*, 2010; Radax *et al.*, 2012a; Fiore *et al.*, 2013). Sponge symbionts are capable of nitrogen fixation, which is the conversion of nitrogen gas (N_2) into ammonia (NH_3) that is then assimilated into amino acids (Fiore *et al.*, 2010; Barton and Northrup, 2011). Microbial nitrogen fixation was described in sponges (Wilkinson and Fay, 1979) and it is attributed to diverse bacterial taxa (Shieh and Lin, 1994; Mohamed *et al.*, 2008), including cyanobacteria (Wilkinson and Fay, 1979).

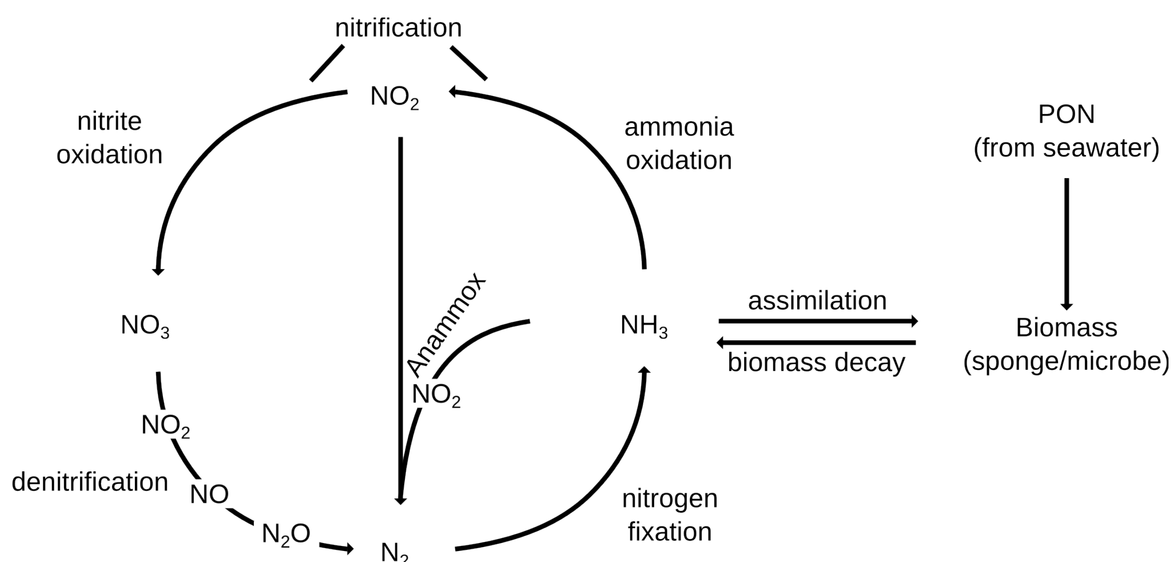


Figure 4. Microbial nitrogen transformations in sponges. Transformations are indicated by arrows. Ammonia (NH_3), nitrite (NO_2), nitrate (NO_3), nitric oxide (NO), nitrous oxide (N_2O), nitrogen gas (N_2), and particulate organic nitrogen (PON). Figure modified from Taylor et al. (2007).

Sponges excrete ammonia as a metabolic waste (Ruppert, 2004), which is potentially toxic. Once outside the sponge cell, ammonia has different fates, including (i) the diffusion to the environmental water through the epithelioid tissues (Ruppert, 2004), (ii) the direct assimilation by symbionts through the glutamine synthase-glutamine oxoglutarate aminotransferase (GS-GOGAT) pathway (Thomas *et al.*, 2010), and (iii) the oxidation by symbionts for energy and electron acquisition. Ammonia monooxygenase genes (*amo*) of archaea (Hallam *et al.*, 2006; Fan *et al.*, 2012b; Radax *et al.*, 2012a) and bacteria (Bayer *et al.*, 2008; Fan *et al.*, 2012b) were detected in sponges, which is an evidence of ammonia-oxidizing capacity of these prokaryotes. Ammonia-oxidizing archaea are the main drivers of ammonia oxidation in at least some sponges, as inferred by the higher abundance of ammonia-oxidizing archaea in comparison with their bacterial counterparts in cold-water sponge species (Radax *et al.*, 2012a). In addition, *amo* was the most transcribed gene by archaeal population in the metatranscriptome study of the cold-water sponge *Geodia barretti*, while bacterial *amo* was not detected in the same study (Radax *et al.*, 2012b). Bacteria capable of anaerobically oxidise ammonia to nitrite (NO_2) and nitrite to nitrogen gas, in a process called anammox (Fiore *et al.*, 2010), were detected in sponges (Mohamed *et al.*, 2010), supporting previous stable isotope measurements that suggested anammox activity (Hoffmann *et al.*, 2009).

Nitrification is defined as the oxidation of ammonia to nitrite and then to nitrate (NO_3) for the purpose of producing energy (Fiore *et al.*, 2010; Barton and Northrup, 2011). Classically, nitrification is considered an aerobic process of which each step is performed by a specialized group of bacteria (Thamdrup, 2012). Due to the recent findings about the ammonia-oxidizing capability of archaea, members of this domain are now assumed to be important players in nitrification (Prosser and Nicol, 2008). Bacteria capable of converting nitrite to nitrate (a process called nitrite oxidation) are commonly found associated with sponges, particularly lineages from the genus *Nitrospira* of the phylum Nitrospirae (Off *et al.*, 2010; Norton and Stark, 2011; Simister *et al.*, 2012a). Furthermore, sponges are widely documented to produce nitrate (Corredor *et al.*, 1988; Diaz and Ward, 1997; Bayer *et al.*, 2007; Jiménez and Ribes, 2007; Bayer *et al.*, 2008; Southwell *et al.*, 2008; Schläppy *et al.*, 2010), which is interpreted as evidence of symbiont-mediated nitrification. Nitrate can be anaerobically reduced to nitrite, to nitric oxide (NO), to nitrous oxide (N_2O), and to nitrogen gas in a process called denitrification (Fiore *et al.*, 2010). Microbial genes (Siegl *et al.*, 2011; Yang and Li, 2012)

and proteins (Liu *et al.*, 2012) that participate in denitrification were identified in sponges, corroborating biochemical evidences of this process (Hoffmann *et al.*, 2009; Schläppy *et al.*, 2010; Fiore *et al.*, 2013).

1.4 Sponge-isolated bioactive compounds

Sponges are among the most prominent marine sources of bioactive compounds. A recent survey estimated that more than 5,300 different products were discovered from marine sponges and their associated microorganisms (Laport *et al.*, 2009). A broad range of substances were recovered from sponges, including terpenoids, alkaloids, peptides, and polyketides (Lejon *et al.*, 2011). There is great economic interest in sponge-derived compounds because of their well-known biotechnological potential. Compounds isolated from marine sponges have several potential applications in biotechnology, including the use as anticancer, antibiotic, antiviral, anti-inflammatory, and antifouling agents (see references within Taylor *et al.*, 2007; Stowe *et al.*, 2011; Li *et al.*, 2013). Ecologically, this plethora of bioactive compounds in sponge is interpreted as a chemical strategy against predation, for competing for space, and for preventing overgrowth by harmful microorganisms (Proksch, 1994).

Due to the high similarity of some sponge-derived compounds to known bacterial secondary metabolites, it is believed that at least some of the bioactive substances are produced by microbial symbionts (Dunlap *et al.*, 2011; Lejon *et al.*, 2011; Wilson *et al.*, 2014). The localisation of secondary metabolites in symbiont cells (Unson *et al.*, 1994) supports this hypothesis. Furthermore, the origin of natural products isolated from sponges was linked to uncultivated symbionts by detection of the corresponding secondary metabolite biosynthetic gene clusters (Piel *et al.*, 2004; Sudek *et al.*, 2007; Wilson *et al.*, 2014). However, the production of bioactive compounds by the host is also frequently the case, as inferred from the localization of such substances in bacteria-free sponge cells (Turon *et al.*, 2000; Salomon *et al.*, 2001).

1.5 The Red Sea sponges: *Stylissa carteri* and *Xestospongia testudinaria*

The Red Sea is an elongated basin that lies between Africa and Asia, representing an economically and environmentally important ecosystem (Zakai and Chadwick-Furman, 2002; El Mamoney and Khater, 2004). It exchanges water at the south with the Gulf of Aden and the Indian Ocean, and connects at the north with the Mediterranean Sea through the Suez Canal (Sofianos and Johns, 2002). Water temperatures are high, around 24 ^{O}C in spring and reach 35 ^{O}C in summer, and there is very little precipitation, ranging from 0.5 to 0.15 m/year (da Silva *et al.*, 1994). Due to strong evaporation (Siddall *et al.*, 2003; Sirocko, 2003), the Red Sea is very saline, where surface water salinity ranges from 36 to 41 practical salinity units (psu) (Edwards, 1987). Furthermore, the Red Sea presents south-north gradients of salinity and surface temperature, where the southern water is warmer and of lower salinity in comparison with the northern water (Ngugi *et al.*, 2012). The Red Sea water is oligotrophic, being deficient of major nutrients, including nitrate, ammonium, phosphate, and silicate (Acker *et al.*, 2008). With regard to biodiversity, the Red Sea presents extensive and diverse coral reef systems, with high levels of species endemism (DiBattista *et al.*, 2013; Nanninga *et al.*, 2014).

The demosponge *S. carteri* (Figure 5A) is a member of the order Halichondrida, family Dictyonellidae. Members of Halichondrida have a widespread geographic distribution, being more common in warmer waters (van Soest *et al.*, 2002). In the Red Sea, *S. carteri* is typically found between 5 to 15 m (Giles *et al.*, 2013b). *S. carteri* is of biotechnological interest as a source of new alkaloids (Eder *et al.*, 1999; Linington *et al.*, 2003) and other bioactive compounds (Kobayashi *et al.*, 2007). *S. carteri* is an LMA sponge (see Figure 2), whose microbiome composition was investigated using Sanger and 454 pyrosequencing methods (Lee *et al.*, 2011; Giles *et al.*, 2013a). According to these studies, a large number of bacterial and archaeal ribo-types (distinct 16S rRNA sequences) are associated with *S. carteri* individuals, although only few lineages dominated the microbial community. The most recovered bacterial 16S rRNA sequences were affiliated to Gammaproteobacteria, while archaeal sequences were at large extent classified to Crenarchaeota – renamed recently to Thaumarchaota (Brochier-Armanet *et al.*, 2008). Furthermore, members of the

gammaproteobacterial population were phylogenetically closely related to bacteria associated with the sponge *Axinella corrugata* (Giles *et al.*, 2013a), which belongs to the same order of *S. carteri*.

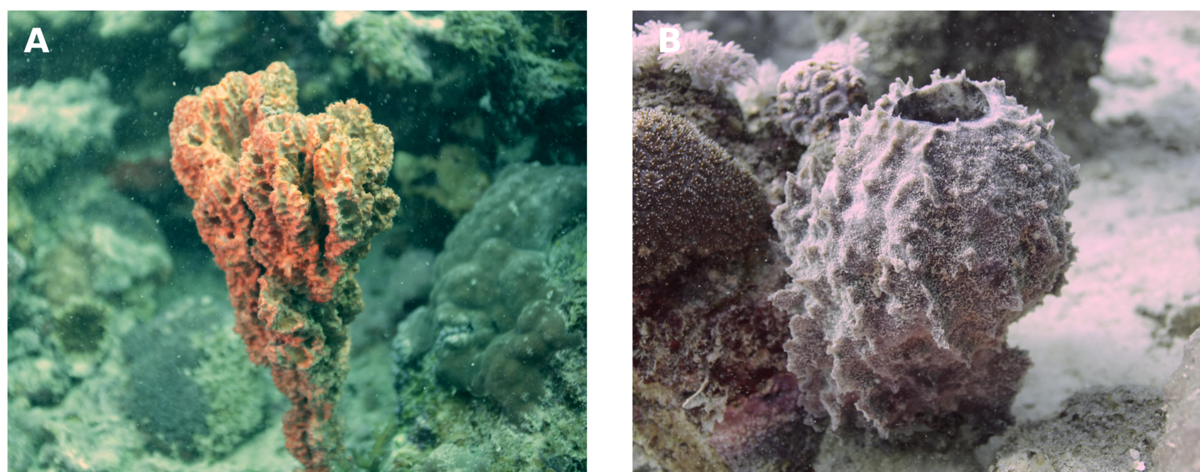


Figure 5. Underwater images of Red Sea sponges. (A) *Stylissa carteri* and (B) *Xestospongia testudinaria* pictures were taken in Fsar reef, at the central coast of Saudi Arabia. Images are courtesy of Michael. L. Berumen, King Abdullah University of Science and Technology, Saudi Arabia.

The giant barrel sponge *X. testudinaria* (Figure 5B) is a member of the order Haplosclerida, suborder Petrosina, family Petrosiidae. Members of this family are widely distributed in the world's oceans (Desqueyroux-Faúndez and Valentine, 2002). In the Red Sea, *X. testudinaria* is often found between 15 to 25 m (Giles E.C., personal communication). Of biotechnological interest is the observation that brominated bioactive compounds and novel Actinobacteria ribo-types were recovered from *X. testudinaria* individuals (Bourguet-Kondracki *et al.*, 1992; Montalvo *et al.*, 2005). *X. testudinaria* is an HMA sponge (see Figure 2), whose microbiome composition was characterized using Sanger and 454 pyrosequencing methods (Lee *et al.*, 2011; Montalvo and Hill, 2011). The *X. testudinaria*-associated microbial community is diverse and dominated by members of the phyla Chloroflexi, Proteobacteria, Firmicutes, Acidobacteria, and Actinobacteria. Members of these phyla are commonly found in sponge-specific clusters (Hentschel *et al.*, 2012; Simister *et al.*, 2012a).

2 Research aims

The aim of my Thesis was to explore the composition and functional activities of microbial communities associated with marine sponges under particular consideration of the HMA/LMA dichotomy. This Thesis has four major objectives:

Study the HMA/LMA dichotomy in Red Sea sponges. The Red Sea sponges were classified according to microbial abundance using transmission electron microscopy, which was carried out for the Red Sea species *Stylissa carteri*, *Xestospongia testudinaria*, *Crella cyathophora*, and *Amphimedon ochracea*. These analyses are part of the electronmicroscopical survey of 56 sponge species (Section 4.1) and the study of the diversity of LMA sponge microbiomes (Section 4.2).

Investigate the microbial community composition of *S. carteri* (LMA) in comparison with *X. testudinaria* (HMA) and to seawater. The composition of microbial communities was investigated with regard to diversity, specificity, and transcriptional activity of their members. I focused on the exploration of important concepts of sponge microbiology, i.e. the HMA/LMA dichotomy and the specificity of the sponge-microbe association. I revisited these concepts using 454 pyrosequencing (Section 4.3) of 16S rRNA genes and transcripts.

Compare the functions of microbial communities associated with sponges and present in seawater. The GeoChip 4, an environmental microarray, was used to investigate the functions of the microbiomes associated with Mediterranean (*Aplysina aerophoba* and *Dysidea avara*) and Red Sea (*Xestospongia testudinaria* and *Stylissa carteri*) sponges, and the seawater. The microbial diversity was characterized alongside using 454 pyrosequencing (Section 4.4).

Investigate the metatranscriptomes of *S. carteri*. The genes transcribed by *S. carteri*-associated microbes were characterized. A protocol to enrich prokaryotic mRNA from total sponge RNA was developed. The genes most expressed by *S. carteri* microbiomes were characterized with respect to their taxonomic and functional classification (Section 4.5).

3 Materials and methods

3.1 Sample collection

Marine sponges were collected by SCUBA diving at Fsar reef (22.23096N; 39.02856E) off the coast of Thuwal, Saudi Arabia (Figure 6). Specimens were collected on two different occasions, hereafter termed Collection 1 and Collection 2 (Table 1). Collections were done in the morning, and samples were taken within a 20 m^2 range. Sponges were collected using a dive knife and placed into zip lock bags containing surrounding seawater. All samples were individually collected and individually processed. Immediately after underwater collection, samples were brought on board the ship for processing. To remove loosely attached particles, sponge pieces were rinsed two times in Ca^{2+}- and Mg^{2+}-free artificial seawater (400 mM NaCl, 27.6 mM Na_2SO_4, 2.3 mM NaHCO3, 8.9 mM KCl, 0.8 mM KBr, 0.4 mM H_3BO_3, 0.15 mM $SrCl_2$, 0.07 mM NaF) (Rottmann *et al.*, 1987). Sponge were cut into 5 mL pieces such that portions of pinacoderm and mesohyl were included in each piece, in sterile Petri dishes containing artificial seawater. Specimens from Collection 1 were placed in 50 mL falcon tubes containing 30 mL RNA*later* solution (Ambion, USA), incubated overnight at 4 ^{O}C, and stored at -80 ^{O}C, according to RNA*later* manufacturer’s instructions. Additionally, 0.5 cm^3 sponge pieces were stored in 2.5% glutaraldehyde (solution in distilled water) for fixation and further processing for transmission electron microscopy (TEM). Sponge material from Collection 1 was preserved in 70% ethanol and used for taxonomic identification by Dr. Rob von Soest, Zoological Museum, University of Amsterdam. Specimens from Collection 2 were collected, processed as in Collection 1, wrapped in aluminium foil, and snap-frozen in liquid nitrogen. The time span between sponge collection and storage was less than 10 min per sample. The total duration of each collection occasion was of less than one hour.

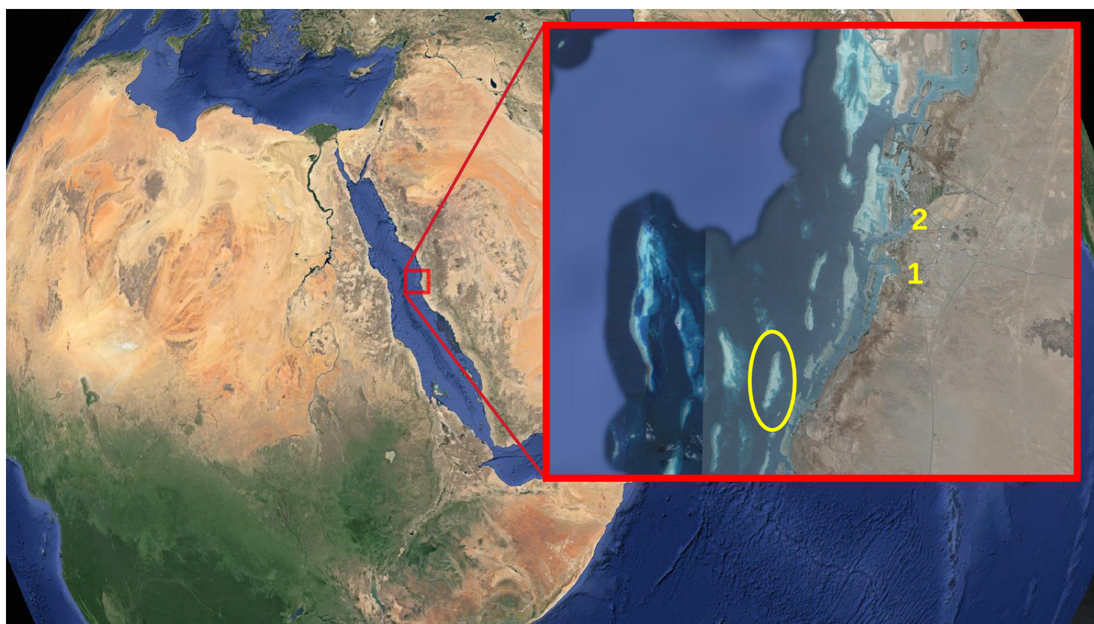


Figure 6. Collection site. The Fsar reef is marked by a circle. (1) The city of Thuwal and the (2) King Abdullah University of Science and Technology campus are marked. Images by Google Earth.

Seawater was sampled at Collection 2. Seawater was sampled in 10 L sterile containers before (WT1), during (WT2) and after (WT3) sponge collection (Table 1). Seawater sample WT1 was filtered on board (7 L), while samples WT2 and WT3 were filtered upon return to the laboratory. Filtering was performed with 0.22 µm hydrophilic Durapore membrane filters (Millipore, USA) and a Masterflex I/P Easy-Load peristaltic pump (Cole-Parmer, USA). The filtration time was approximately 50 min for each replicate. Filters, containing seawater particles above 0.22 µm were placed in 1.5 mL cryotubes and stored at -80 °C.

Table 1. List of samples

Source	Sample identification	Description	Order of collection
Collection 1			
S. carteri	SC1	Replicate 1	1
S. carteri	SC2	Replicate 2	2
S. carteri	SC3	Replicate 3	3
X. testudinaria	XT1	Replicate 1	4
X. testudinaria	XT2	Replicate 2	5
X. testudinaria	XT3	Replicate 3	6
Collection 2			
Seawater	WT1	Replicate 1	1
Seawater	WT2	Replicate 2	5
Seawater	WT3	Replicate 3	9
S. carteri	AS1	Replicate 1	2
S. carteri	AS2	Replicate 2	3
S. carteri	AS3	Replicate 3	4
X. testudinaria	AX1	Replicate 1	6
X. testudinaria	AX2	Replicate 2	7
X. testudinaria	AX3	Replicate 3	8

Collection 1 was performed between 13-14 depth at 30 ^{o}C water temperature, on 02.11.2010.
Collection 2 was performed between 8.5-12 depth at 26 ^{o}C water temperature, on 09.01.2012.

Sample collections were realized under the supervision of the Coastal and Marine Resources Core Lab at the King Abdullah University of Science and Technology (KAUST). The following researchers affiliated with KAUST took part in the sample collections: Timothy Ravasi, Michael L. Berumen, Emily C. Giles, and Feras Lafi.

3.2 Laboratory procedures

3.2.1 Transmission electron microscopy

Transmission electron microscopy (TEM) was performed to assess whether Red Sea

sponges are classified as LMA or HMA (Collection 1, Table 1). Samples fixed in 2.5% glutaraldehyde solution were processed according to Giles *et al*. (2013a). After processing, samples were embedded in Epon 812 (SERVA Electrophoresis, Germany) and sectioned with an Om-U3 ultramicrotome (C. Reichert, Austria). Samples were visualized with an Zeiss EM10 microscope (Zeiss, Germany).

3.2.2 Internal standard synthesis

Two internal standards (spike-in) were added to the Lysing Matrix E tubes (15 ng each) prior to cell disruption of samples from Collection 1. The spike-ins were artificial mRNAs produced by *in vitro* transcription of two vectors pTXB1 (New England Biolabs, USA) and pFN18A HaloTag T7 Flexi (Promega, USA) (Gifford *et al.*, 2011; Moran *et al.*, 2013; Satinsky *et al.*, 2013). The spike-ins were selected to represent the estimated average prokaryotic gene length of 924 base pairs (bp) (Xu *et al.*, 2006). First, the commercial vectors were linearised with restriction enzymes. The enzyme FastDigest NcoI (Fermentas, Germany) was used to linearise the pFN18A vector and the enzyme FastDigest BamHI (Fermentas, Germany) was used to linearise the pTXB1 vector. Restriction enzyme reactions were carried out for 5 µg of vector and according to the corresponding manufacturer's recommendations. Linearised vectors were purified from circular vectors by loading the products of the restriction enzyme reactions into 1% agarose gel and by cutting the corresponding bands after the electrophoresis run. DNA was purified from agarose gel slices with the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, Germany). Finally, spike-ins were produced by *in vitro* transcription of linearised vectors (1 µg) using the MEGAscript T7 Transcription kit (Ambion, USA) and recovered with the MEGAclear Transcription Clean-Up kit (Ambion, USA) according to manufacturer's recommendations. Spike-ins final base pair lengths were 970 and 917 for pFN18A and pTXB1, respectively.

3.2.3 Nucleic acids isolation

Sponge pieces and seawater filters preserved in RNA*later* or in liquid nitrogen were cut into smaller pieces and placed in ice-cold Lysing Matrix E tubes (MP Biomedicals, USA) containing 600 µL of RLT buffer (Qiagen, Germany) to which 1% 2-mercaptoethanol was added. Each Lysing Matrix E tube was loaded with approximately 200 mg of sponge material or with 1/3 of seawater filter. Sponge pieces were selected such that portions of pinacoderm and mesohyl were included. Lysing Matrix E tubes contain ceramic (1.4 mm) and silica spheres (0.1 mm), as well as glass beads (4 mm) with the purpose of mechanical cell disruption. Cells were disrupted using a homogenizer FastPrep Instrument (MP Biomedicals, USA) for 30 sec, with speed set to 5.5. To obtain sufficient amounts of RNA for metatranscriptomics, 10-15 extractions were performed for each replicate sample.

Total DNA and RNA were co-extracted from cell lysates with the AllPrep DNA/RNA mini kit (Qiagen, Germany) according to the manufacturer's instructions. To avoid RNA degradation, 1 µL of the RNAse inhibitor SUPERase-in (20 u/µL; Ambion, USA) was added to each 20 µL of isolated RNA. Isolated DNA was treated with RNase A (Roche, Germany) by adding 1.5 µL (10 mg/mL) of the nuclease to each 100 µL of isolated DNA. To avoid genomic DNA contamination, RNA was treated with RQ1 RNase-Free DNase (20 u/µL; Promega, Germany) by adding 6 µL of the nuclease to each 50 µL of the isolated RNA solution. Then, RNA was incubated for 1 h at 37 ^{O}C, followed by addition of 6 µL of stop buffer (for each 50µL of reaction; Promega, Germany) and incubation at room temperature for 10 min. Nucleic acids were quantified based on absorption spectroscopy with a NanoDrop 2000c Spectrophotometer (PEQLAB Biotechnologie, Germany). RNA integrity was analysed with an Agilent 2100 Bioanalyzer (Agilent Technologies, USA) or an Experion Automated Electrophoresis System (Bio-Rad, USA). Isolated DNA was stored at -20 ^{O}C and isolated RNA was stored at -80 ^{O}C.

3.2.4 16S rRNA gene amplification and cloning

To obtain a preliminary view of the microbial diversity of *S. carteri* and *X. testudinaria*, the 16S rRNA genes were amplified, cloned and sequenced. Briefly, near full-length 16S rRNA gene was amplified from DNA samples (Collection 1, 30 ng) using the universal primers 27f (5'-GAGTTTGATCCTGGCTCA-3') and 1492r (5'-TACGGYTACCTTGTTACGACT-3') (Lane, 1991). Polymerase chain reaction (PCR) was performed with the DreamTaq Green DNA Polymerase (Fermentas, Germany) following manufacturer's recommendations. The PCR contained 0.4 µM of each primer. The 27f-1492r PCR cycling conditions are described in Table 2. Additionally, a fragment of the 16S rRNA gene was amplified using the degenerated primers 533f and 907r (Simister *et al.*, 2012c) (Table 3). The details of the 533f-907r primer pair and target region are described in the Section 3.2.5. PCR products were purified with the NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL, Germany). The 27f-1492r PCR products were purified directly, while the 533f-907r PCR products were purified after band extraction from agarose gels. Purified PCR products (80 ng) were cloned into the pGEM-T Easy Vector (Promega, Germany) with the T4 DNA Ligase (Promega, Germany) according to manufacturer's instructions. Ligation reaction occurred overnight at 4 ^{O}C.

Table 2. The 27f-1492r PCR cycles

Phase	Temperature (^{O}C)	Time	Cycles
Initial denaturation	96	5'	
Denaturation	96	30''	
Annealing	57	30''	30
Extension	72	90''	
Final extension	72	5'	

Time is reported in minutes (') or seconds ('').

Table 3. The 533f-907r PCR cycles for cloning

Phase	Temperature (^{O}C)	Time	Cycles
Initial denaturation	96	5'	
Denaturation	96	30''	
Annealing	61	30''	30
Extension	72	30''	
Final extension	72	5'	

Time is reported in minutes (') or seconds ('').

Vectors were transformed into competent *Escherichia coli* NovaBlue cells (Inoue *et al.*, 1990) by heat shock, under sterile conditions. 2 µL of the ligation reaction product was added to Eppendorf tubes containing 50 µL of ice-cold competent cells under sterile conditions. These tubes were incubated on ice for 30 min, followed by incubations at 42 ^{O}C for 1 min 15 sec and at 0 ^{O}C for 5 min. After heat shock, transformed competent cells were transferred to Eppendorf tubes containing 1 mL of liquid Luria-Bertani (LB) medium pH 7.0 (Sambrook, 2001) and incubated for 2 h, at 37 ^{O}C. Aliquots of transformed cells of 150 µL and 700 µL were spread using glass beads on LB agar plates (Sambrook, 2001) containing ampicillin (100 µg/mL), isopropyl β-D-1-thiogalactopyranoside (IPTG; 0.5 mM) and 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal; 75 µg/mL). Ampicillin was added to restrict growth to the transformed cells only (cells that contained plasmids), while IPTG and X-gal were added to enable blue-white screening. Plates were incubated overnight at 37 ^{O}C.

White colonies, originating from cells transformed with the pGEM-T Easy Vector that contained an inserted DNA fragment, were screened for the correct insert size. PCR was performed with the vector-specific primers targeting the T7 and the SP6 RNA polymerase promoter sites (respectively 5'-GTAATACGACTCACTATAGGG-3' and 5'-ATTTAGGTGACACTATAG-3') and the DreamTaq Green DNA Polymerase (Fermentas, Germany). The reaction mixture was composed according to the DreamTaq Green DNA Polymerase manufacturer’s instructions. Final primer concentration was 0.4 µM. The T7-SP6 PCR cycling conditions are described in Table 4. For each colony, a piece was picked with a sterile pipette tip and used as template for PCR. The PCR products were analysed by agarose gel electrophoresis. Colonies that resulted in PCR products with the expected insert size were kept in LB agar plates containing ampicillin.

Table 4. The T7-SP6 PCR cycles for vector screening

Phase	Temperature (^{O}C)	Time	Cycles
Initial denaturation	96	5'	
Denaturation	96	30"	
Annealing	45	30"	30
Extension	72	30-90"	
Final extension	72	5'	

Time is reported in minutes (') or seconds (").
The extension times varied according to expected insert size.

In order to sequence the cloned DNA fragment, plasmids were isolated from overnight grown cultures (3 mL, liquid LB medium, 37 ^{O}C, 150 rpm) of randomly selected colonies. The cells suspended in the liquid medium were pelleted in a microcentrifuge (13,000 rpm, 1 min) and resuspended in 150 µL ice-cold alkaline lysis solution I (Table 5). After addition of 150 µL alkaline lysis solution II (Table 5), the suspension was homogenized by inverting the tubes approximately 10 times. The alkaline lysis solution III (150 µL) (Table 5) was added to the tubes, followed by incubation for 30 min on ice. Plasmid precipitation was completed by the addition of 350 µL of ice-cold isopropanol (70%). Plasmid washing and recovery steps were performed according to the protocol for minipreparation of plasmid DNA by alkaline lysis with SDS (Sambrook, 2001).

Table 5. Plasmid DNA minipreparation solutions

Phase	Final concentration
Solution I	
Tris-HCl, pH 7.5	50 mM
EDTA, pH 8	10 mM
RNase A (Roche, Germany)	100 µg/mL
Solution II	
Sodium hydroxide (NaOH)	0.2 N
Sodium dodecyl sulfate (SDS)	1% w/v
Solution III	
Sodium acetate trihydrate	0.408 g/mL (final pH 4.8, adjusted with HCl)

Details of solution preparation are found in the Appendix 1 of Sambrook (2001).

Finally, isolated plasmids containing cloned fragments of the 16S rRNA gene (from base 27 to 1492 or from base 533 to 907; originating from PCR described in Table 2 and Table 3, respectively) were sequenced.

3.2.5 16S rRNA gene and transcript amplification

The amplification of the 16S rRNA genes and transcripts was carried out with the degenerated primers 533f and 907r (Simister *et al.*, 2012c). These primers were designed to flank the hypervariable regions V4 and V5 and to include the gene sequences of the candidate phylum Poribacteria. In order to sequence the amplicons in parallel using 454 pyrosequencing, sequencing adaptors and sample-specific barcodes were added upstream to the primer sequences (Table 6). The reverse primer sequence was the same for all amplification reactions, and it did not contain a barcode sequence.

Table 6. Primers used for amplification of 16S rRNA genes and transcripts with 454 pyrosequencing

WTD1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CGTGTCTCTA**TAGTGCCAGCAGCYGCGGTMA
WTD2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CTCGCGTGTC**TAGTGCCAGCAGCYGCGGTMA
WTD3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TCTCTATGCG**TAGTGCCAGCAGCYGCGGTMA
WTR1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***ACGCGAGTAT**TAGTGCCAGCAGCYGCGGTMA
WTR2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***ACTACTATGT**TAGTGCCAGCAGCYGCGGTMA
WTR3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***ACTGTACAGT**TAGTGCCAGCAGCYGCGGTMA
ASD1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TGATACGTCT**TAGTGCCAGCAGCYGCGGTMA
ASD2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CATAGTAGTG**TAGTGCCAGCAGCYGCGGTMA
ASD3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CGAGAGATAC**TAGTGCCAGCAGCYGCGGTMA
ASR1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CGTAGACTAG**TAGTGCCAGCAGCYGCGGTMA
ASR2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TACGAGTATG**TAGTGCCAGCAGCYGCGGTMA
ASR3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TACTCTCGTG**TAGTGCCAGCAGCYGCGGTMA
AXD1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***ATACGACGTA**TAGTGCCAGCAGCYGCGGTMA
AXD2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TCACGTACTA**TAGTGCCAGCAGCYGCGGTMA
AXD3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***CGTCTAGTAC**TAGTGCCAGCAGCYGCGGTMA
AXR1-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TAGAGACGAG**TAGTGCCAGCAGCYGCGGTMA
AXR2-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***TCGTCGCTCG**TAGTGCCAGCAGCYGCGGTMA
AXR3-533f	*CCATCTCATCCCTGCGTGTCTCCGACTCAG***ACATACGCGT**TAGTGCCAGCAGCYGCGGTMA
Reverse-907r	*CCTATCCCCTGTGTGCCTTGGCAGTCTCAG*ATCCGTCAATTMMYTTGAGTTT

Italic letters in the primers correspond to sequencing adaptors (Roche Titanium Primers – Lib-L), bold letters correspond to sample-specific barcodes, underlined letters correspond to sequence linkers. Degenerated primer sequences are written with normal letters in the 5'-3' direction.

Isolated nucleic acids (Collection 2 samples, Table 7) were used as template for amplification of 16S rRNA genes and transcripts. Firstly, RNA (30ng) was converted into cDNA with the SuperScript III First-Strand Synthesis System (Invitrogen, USA) and random hexamers, according to manufacturer's instructions. PCR and reverse transcription PCR (RT-PCR) were performed with the Qiagen multiplex PCR kit (Qiagen, Germany). PCR was carried out with 30 ng of DNA. RT-PCR was carried out with cDNA (1 µL of the reverse transcription reaction product). Each reaction contained the Q-solution as buffer and 0.4 µM of each primer. PCR program cycles are described in Table 8. PCR with RNA as template was carried out to verify cDNA contamination with genomic DNA. For each starting material, PCR or RT-PCR were realized in triplicates. Triplicates were pooled at equimolar ratios prior sequencing.

Table 7. Starting material for amplification of 16S rRNA genes and transcripts

Source	Sample identification	Material	
		DNA	**RNA**
Seawater	WT1	WTD1	WTR1
Seawater	WT1	WTD1	WTR1
Seawater	WT1	WTD1	WTR1
S. carteri	AS1	ASD1	ASR1
S. carteri	AS2	ASD2	ASR2
S. carteri	AS3	ASD3	ASR3
X. testudinaria	AX1	AXD1	AXR1
X. testudinaria	AX2	AXD2	AXR2
X. testudinaria	AX3	AXD3	AXR3

Table 8. The 533f-907r PCR cycles for pyrosequencing

Phase	Temperature (^{o}C)	Time	Cycles
Initial denaturation	95	15'	
Denaturation	94	30''	
Annealing	62	90''	30
Extension	72	90''	
Final extension	72	10'	

Time is reported in minutes (') or seconds ('').

3.2.6 Enrichment of prokaryotic mRNA

Prokaryotic mRNA was enriched from total RNA isolated from *S. carteri* samples (Collection 1). Prokaryotic mRNA enrichment was also realized with total RNA isolated from *X. testudinaria* samples (Collection 1). *X. testudinaria* metatranscriptomics is expected to be finished after completing my doctoral studies. The prokaryotic mRNA enrichment procedure was performed in two stages, as follows:

Removal of eukaryotic mRNA from total RNA: Eukaryotic mRNA was removed from total sponge RNA (100 µg) based on a modified version of the "2 rounds oligo(dT) selection protocol" of the Poly(A)Purist MAG Kit (Ambion, USA). After the first round of oligo(dT) selection (magnetic beads capture), the non-poly(A) RNA fraction, which contains rRNA and prokaryotic mRNA, was recovered from the Binding Solution with the RNeasy MinElute Cleanup Kit (Qiagen, Germany). The RNeasy MinElute Cleanup Kit does not recover small RNA molecules (< 200 bp), such as tRNA and 5S rRNA. The recovered non-poly(A) RNA fraction was the input for the second round of oligo(dT) selection and non-poly(A) RNA fraction recovery.

Removal of the rRNA subunits of eukaryotes (18S and 28S), bacteria (16S and 23S), and archaea (16S and 23S): The rRNA subunits were removed according to Stewart *et al.* (2010) using the recovered non-poly(A) RNA fraction as template. *S. carteri* DNA was pooled (SC1, SC2, and SC3) and used as template for PCR amplification of rRNA subunit genes. PCR was performed by DreamTaq DNA Polymerase (Fermentas, Germany). Optimal conditions for each rRNA gene were established based on the primer annealing temperatures.

Amplified genes were the templates for synthesis of biotin-labeled anti-sense RNA. The subtraction of rRNA was performed as suggested by the original protocol (Stewart *et al.*, 2010), with 400 ng of starting RNA template and 800 ng of each probe.

In order to confirm rRNA removal, enriched prokaryotic mRNA were analysed with the Experion Automated Electrophoresis System (Bio-Rad, USA). The enriched prokaryotic mRNA (8.5 ng for *S. carteri* and 15 ng for *X. testudinaria* samples) was amplified with the MessageAmp II-Bacteria kit (Ambion, USA).

3.2.7 DNA sequencing

The clone libraries were sequenced with an ABI 377XL automated sequencer and the primer targeting the T7 RNA polymerase promoter site. The amplicons of 16S rRNA genes and transcripts were sequenced with the Roche 454 GS-FLX Titanium platform. Metatranscriptomes were sequenced using Illumina HiSeq 2000 standard protocols, resulting in 101 bp paired end reads with an estimated mean insert size of 280 bp. 454 and Illumina sequencing were performed by the KAUST Genomics Core Lab.

3.3 Bioinformatic analyses

3.3.1 Bioinformatic analysis hardware

Most bioinformatic analyses were performed on a Fujitsu Celsius R920 workstation (Fujitsu, Germany), running the operating system Ubuntu Release 12.04 (precise) 64-bit. The Fujitsu Celsius R920 workstation memory and processor configurations were 251.8 gigabytes and Intel® Xeon(R) CPU E5-2690 0 @ 2.90GHz × 18, respectively.

3.3.2 Data manipulation and processing

Manipulation and processing of data were performed with scripts written in python 2.7.3 (https://www.python.org/) or R 2.14.1 languages (R Development Core Team, 2012). RStudio was the interactive interface for R (https://www.rstudio.com). IPython was the interactive interface for python (Pérez and Granger, 2007).

3.3.3 Image, graph, and text editors

Images and graphs were processed with the GNU image manipulation program (GIMP) v 2.6.12 (https://www.gimp.org) and with the Inkscape vector graphics editor v 0.48 (https://www.inkscape.org). Transmission electron microscopy images were manipulated for adjustment of contrast. This Thesis was written using the office suite LibreOffice v 4.0.3.3 (https://www.libreoffice.org). Bibliography and citation were managed with Zotero v 4.0.17.1 (Roy Rosenzweig Center for History and New Media, USA). Scripts were written using the text editor Kate v 3.8.5 (http://www.kate-editor.org).

3.3.4 Processing of 16S rRNA gene clone sequences

The raw chromatogram trace files were processed with the Staden package (Staden, 1996). Sequences were trimmed to exclude PCR primers and low quality bases (Phred quality score < 20). Taxonomic classification of the sequences was performed with the RDP Classifier (Wang *et al.*, 2007) as implemented in QIIME (Quantitative Insights Into Microbial Ecology) v 1.6 (Caporaso *et al.*, 2010b), with a confidence interval of 0.8. A modified version of the SILVA 108 database was used as reference (Pruesse *et al.*, 2007; Moitinho-Silva *et al.*, 2014a) for taxonomic classification. Basic Local Alignment Search Tool (BLAST) v 2.2.28+ searches ran locally (Altschul *et al.*, 1990) against the National Center for Biotechnology Information (NCBI) nucleotide database (NCBI-nt) downloaded in March 2014.

3.3.5 Processing of 16S rRNA gene amplicon sequences

The raw files generated by 454 pyrosequencing were processed with the QIIME pipeline v 1.4 (Caporaso *et al.*, 2010b). QIIME consists of a suite of bioinformatic tools and scripts for analysing high-throughput amplicon sequencing data. First, sequences in the quality (.qual) and in the nucleotide sequence (.fna) files were sorted according to the barcodes running the script “split_libraries.py”. Additionally, the script “split_libraries.py” allowed for quality processing of raw reads, including (i) the truncation of sequences with a mean quality score below 25 at a 50 bp window; the removal of (ii) the reverse primer sequences, and any subsequent sequence; (iii) unassigned sequences; (iv) sequences with homopolymers longer than 6 nucleotide; (v) sequences with less than 200 bp length; and (vi) sequences that contained any mismatch at the primer or at the barcode.

To correct for sequencing errors, denoising was performed based on the text versions of the standard flowgram format files (.sff.txt) with the standard parameters of the “denoise_wrapper.py” script. Denoiser results were inflated with the “inflate_denoiser_output.py”, which re-integrates centroids and singletons into the pipeline. Prior grouping sequences into operational taxonomic units (OTU), sequences assigned to ASD2 (see Table 7) were removed from downstream analysis because they were contaminated with *X. testudinaria* DNA. Amplicon sequences were grouped into OTUs using the optimal configuration settings of the UCLUST method (Edgar, 2010), as implemented in the “pick_otus.py” script. The sequences were assigned to de novo OTUs based on a 97% similarity threshold. The representative sequence of each OTU was retrieved running the script “pick_rep_set.py”. The representative sequences represented the cluster seeds of the UCLUST method.

Representative sequences were aligned to SILVA 108 (Pruesse *et al.*, 2007) core sequences, which were provided by QIIME. Alignment was performed using PyNAST (Caporaso *et al.*, 2010a), which is an implementation of the Nearest Alignment Space Termination algorithm in python (DeSantis *et al.*, 2006). Sequence that failed to align to SILVA core sequences were removed from downstream analysis. Putative chimeras were

identified among the representative sequences set with ChimeraSlayer (Haas *et al.*, 2011) and with Ccode (Gonzalez *et al.*, 2005) algorithms, as implemented in mothur v 1.25 (Schloss *et al.*, 2009). Chimeras were removed from downstream analysis. OTUs which were composed of a single sequence (singletons) when considering the whole dataset, were excluded. This is common practice when using 454 pyrosequencing for 16S rRNA gene amplicons, because singletons are regarded as sequencing errors (Behnke *et al.*, 2011).

Taxonomic classification of the remaining OTUs was accomplished based on the representative sequence classification. Representative sequences were classified primarily with the RDP Classifier (Wang *et al.*, 2007). Unclassified sequences were classified to the phylum level based on best BLAST hit (e-value < 0.001). Sequences were classified based on the SILVA 108 database, which was modified to re-annotate poribacterial sequences (Moitinho-Silva *et al.*, 2014a). Finally, the OTUs of which representative sequences were classified as eukaryote or chloroplasts were excluded from the dataset. This procedure resulted in an OTU table (Appendix S3 of Moitinho-Silva *et al.* (2014a)) which listed the sequence counts of each OTU in each sample, as well as the taxonomic classification of each OTU.

3.3.6 Assignment of OTUs to sponge-specific clusters or sponge- and coral-specific clusters

Sequences from sponge-specific clusters or sponge- and coral-specific clusters (SC/SCC) were retrieved from the database generated by Simister *et al*. (2012a). This database includes 16S rRNA gene sequences recovered from sponges from several studies. After phylogenetic analysis, Simister *et al*. (2012a) classified part of these sequences into SC/SCC. The Simister *et al*. (2012a) database sequences that were not classified into SC/SCC were considered as non-specific sequences. The 16S rRNA gene 533-907 base pair fragments were extracted from all sequences presented in Simister *et al.* (2012a) database with PrimerProspector v. 1.0.1 (Walters *et al.*, 2011), with 0 penalty for gaps and mismatches on the primers. The 533-907 base pair fragments were clustered using the optimal configuration settings of the UCLUST method (Edgar, 2010), as implemented in QIIME v 1.4. The SC/SCC 16S rRNA

gene 533-907 base pair fragments were only considered as reference for assignment (250 out of 1153) when they did not cluster (> 97% similarity threshold) with the non-specific sequences. This effort was necessary to exclude the SC/SCC 16S rRNA gene 533-907 base pair fragments that were similar to the fragments from non-specific sponge-derived sequences. Finally, the OTUs were assigned by clustering the representative sequences with the selected SC/SCC 16S rRNA gene 533-907 base pair fragments at 97% similarity threshold.

3.3.7 Assignment of OTUs to species-specific sequences

OTU assignment to species-specific sequences was based on published sequences from *S. carteri* (n=67) and *X. testudinaria* (n=414). Sequences of *S. carteri* were generated from Collection 1 samples by Giles *et al*. (2013a). Sequences from *X. testudinaria* were generated from samples collected at Manado Bay, Indonesia in 2003 and 2005 by Montalvo and Hill (2011). The 16S rRNA gene 533-907 base pair fragments were retrieved and OTU assignment were performed as described (Section 3.3.6).

3.3.8 Analysis of microbial community diversity

To assess alpha diversity, the OTU table was rarefied to the smallest sampling effort (32,148; Appendix S4 of Moitinho-Silva *et al.* (2014a)). This procedure is recommended to account for sample-size bias (Gihring *et al.*, 2012). Richness, which is the number of species in a given community, was estimated with the Chao1 and the abundance-based estimator (ACE) (Chao, 1984; Chao and Lee, 1992). Diversity was measured with the Shannon and Simpson indices (Shannon, 1948; Simpson, 1949), which take into account the relative abundance of species. Ecological indices values were calculated based on the OTU table with the vegan package v 2.0-4 in R (Oksanen *et al.*, 2012).

To assess beta diversity, UniFrac analysis (Lozupone and Knight, 2005) was performed

with rarefied datasets as implemented in QIIME v 1.4. UniFrac is a method for computing differences between microbial communities based on phylogenetic information. UniFrac distances were the input for clustering using the unweighted pair-group method using arithmetic averages (UPGMA). Node support was calculated on subsets of the OTU table (jackknife), with 100 repetitions at 10,000 sequences per sample. Subsets of the original OTU table were obtained by summing sequence counts per OTU in order to compare materials and sources with the Bray-Curtis dissimilarity index. Bray-Curtis index values were calculated with the vegan package v 2.0-4 in R (Oksanen *et al.*, 2012).

3.3.9 Analysis of OTU overlap

The OTU overlap among sources (Table 7) was visualized using a Venn diagram. A Venn diagram was obtained based on the rarefied OTU table [Appendix S4 of Moitinho-Silva *et al.* (2014a)]. A data set was obtained by summing the OTU sequence counts from each source. The Venn diagram was generated with the Limma package v 3.12.1 in R (Smyth, 2005).

3.3.10 OTU abundance pattern analysis

To explore OTU abundance patterns, the unsupervised machine learning algorithm Minimum Curvilinear embedding (MCE) (Cannistraci *et al.*, 2010) was applied. MCE was developed for dimension reduction of datasets characterized by low numbers of samples with very high numbers of measures (Cannistraci *et al.*, 2010). To perform MCE dimension reduction, the Euclidean distance between the OTUs was calculated based on their abundance in each sample in the rarefied OTU table. The Euclidean distances were the input to calculate the Minimum Curvilinearity distance matrix. MCE analysis was performed with Matlab (The MathWorks, USA) by our collaborator Carlo Cannistraci, former Post-doc at KAUST. The relative abundance of OTUs was inspected with a heatmap. The order of OTUs in the heatmap was based on the MCE first dimension. The heatmap was generated in R with the gplots v 2.11.0 package (Bolker *et al.*, 2012).

3.3.11 Analysis of transcriptionally active taxa

Taxa sequence counts were calculated by pooling OTU counts according to the deepest taxonomic annotation. Taxon relative abundances in a given sample were calculated by dividing the number of the taxon sequences by the total number of the sequences in the sample data set. R-squared correlation values were calculated with the fitting linear models function in R. The null hypothesis that the relative abundance of sequences assigned to a particular taxon in each cDNA and DNA data set pair is due to their distribution was tested with the two-sided Fisher's exact test (Rivals *et al.*, 2007). Statistical tests were performed with STAMP v 2.0.0 (Parks and Beiko, 2010). P-values were corrected with the Storey's false discovery rate and filtered according to absolute and relative effect size statistics, as described in detail by Moitinho-Silva *et al*. (2014a). The sample ASR2 was omitted from these analyses due to the exclusion of ASD2 derived data.

3.3.12 Processing of metatranscriptome sequences

The raw Illumina reads generated by sequencing the enriched prokaryotic mRNA were initially processed by our collaborators Taewoo Ryu and Loqmane Seridi, at KAUST. The reads were truncated to the first low quality base (Phred quality score > 20). Additionally, sequencing adapters, including partial adapters from the raw reads were removed.

Further processing and annotation of the metatranscriptomic reads were performed with the metagenomics RAST server (MG-RAST) v 3.3.6.1 (Meyer *et al.*, 2008). The MG-RAST is an analysis platform that performs functional assignments of sequences by comparison with protein and nucleotide databases. Processed reads submission was done with the default parameters. Additionally, the quality reads threshold was set to Phred score > 20 and the option to remove artificial replicates (Gomez-Alvarez *et al.*, 2009) was selected. The quality-processed, dereplicated, and screened nucleotide sequences were retrieved from the MG-RAST server (screen.passed.fna files). To identify sequences derived from the two spike-ins

(see Section 3.2.2 for details), the “screen.passed” sequences were mapped against the two spike-in sequences, individually, with the default parameters of the Genious v 6.0.6 software (http://www.geneious.com/). Sequence that were mapped to spike-ins were flagged for removal.

3.3.13 Taxonomic annotation of metatranscriptome sequences

The identified protein features (hereafter named “transcriptional features”) were retrieved from the MG-RAST server (superblat.expand.protein files). Among other BLAT (Kent, 2002) results, these files contain the identifier for the database hit (md5) for each transcriptional feature to a given protein database. The protein databases included in MG-RAST were: eggNOG (Powell *et al.*, 2014), GenBank (Benson, 2004), IMG (Markowitz *et al.*, 2012), KEGG (Kanehisa and Goto, 2000), PATRIC (Wattam *et al.*, 2013), RefSeq (Pruitt *et al.*, 2005), SEED (Overbeek *et al.*, 2005), Swiss-Prot (Gattiker *et al.*, 2003), TrEMBL (Bairoch, 2000), M5NR (Wilke *et al.*, 2012), InterPro (Hunter *et al.*, 2012), and Phantome (http://www.phantome.org). To assess the corresponding domain of life (Eukaryota, Bacteria, and Archaea) of each md5, the M5NR-based annotation was retrieved from the MG-RAST workbench with stringency parameters set to the minimum values (maximum e-values: 1; minimum identity: 1%; minimum alignment length: 1 bp). The M5NR database the reference because it integrates the aforementioned databases (Wilke *et al.*, 2012).

To annotate the transcriptional features according to the domain of life, the best BLAT hit (e-value < 0.001) to any of the protein databases were considered. In case of annotation conflict between databases, the accepted classification was the one supported by the majority of the databases. Transcriptional features that resulted in two or more domains with the same number of supporting databases were classified as tied between these domains. Putative prokaryotic transcriptional features (i.e. those classified to Archaea, to Bacteria, or as a tied between Archaea and Bacteria) were retained for further processing. The taxonomic classification of these putative prokaryotic transcriptional features were retrieved from the lowest common ancestor (LCA) annotation provided by the MG-RAST. The LCA annotation provided by the MG-RAST (superblat.expand.lca files) was only considered when supported

by at least four BLAT hits (e-values < 0.001). Putative prokaryotic transcriptional features that were classified by LCA to Eukaryota or as virus were removed. Putative prokaryotic transcriptional features that were flagged as spike-in (see Section 3.3.12) were removed. After these filtering steps, the remaining features were considered as prokaryotic transcriptional features. Features that were not classified by LCA were considered as unidentified prokaryotic features.

3.3.14 Functional annotation of metatranscriptome sequences

The functional classification of the prokaryotic transcriptional features was done based on the SEED subsystems annotation provided by the MG-RAST (superblat.expand.ontology files). Among other BLAT search results information, these files contain the identifier for the SEED database hit (md5) for each transcriptional feature. The md5 annotations were retrieved from the MG-RAST workbench using the minimum stringency values (maximum e-values: 1; minimum identity: 1%; minimum alignment length: 1 bp). The SEED annotation of the prokaryotic transcriptional features was only considered when supported by a valid BLAT hit (e-value < 0.001). Transcriptional features matching the gene function “Retron-type reverse transcriptase” were regarded as misannotated putative rRNA sequences (Tripp *et al.*, 2011), and thus removed. The remaining annotated prokaryotic transcriptional features were considered for downstream analyses.

3.3.15 Analysis of dissimilarity between replicate metatranscriptomes

Metatranscriptomic data sets were compared with the Bray-Curtis dissimilarity index. The functionally annotated prokaryotic transcriptional feature counts for each data set were rarefied to the minimum feature counts obtained, when considering each sample. The Bray-Curtis index values were calculated with the vegan package v 2.0-4 in R (Oksanen *et al.*, 2012).

3.3.16 Assembly of metatranscriptomic reads

The screened reads (screen.passed.fna files) that were annotated to each selected gene function were de novo assembled. The assembly was performed with Geneious v 6.0.6 using the default parameters, which allowed for 15% maximum gaps per read, a maximum gap size of 2, and 30% of maximum mismatches per read. Match word length was 14 bases and index word length was set to 12. To check for the possibility of chimeric assembly, contigs were also assembled under stringent parameters, which were: no gaps allowed, 1% of maximum mismatches per read, and 100% of overlap between reads. While this procedure gave similar results (Section 4.6.2), reduced bootstrap values for the phylogenetic association were computed; indicating that the assemblies were not prone to chimera formation.

For each gene function, the representative nucleotide consensus sequence was retrieved from the contig that contained the greatest number of reads. The amino acid sequences were deduced with the online tool EMBOSS Transeq (McWilliam *et al.*, 2013). The selected frame was chosen to match the frame indicated by the best BLASTX hit against NCBI non-redundant database (nr).

3.3.17 Phylogenetic analysis of gene function

In order to construct phylogenetic trees, orthologous amino acid sequences were identified in the UniProt Knowledgebase (http://www.uniprot.org/) with BLAST. The top BLAST hits were retrieved for all of the selected gene functions. For the gene functions “copper-containing nitrite reductase” and “methanol dehydrogenase large subunit protein”, orthologous sequences from selected taxa were additionally retrieved, which were Bacteria; and Alpha-, Beta-, and Gammaproteobacteria.

For each gene function, the representative amino acid sequences and orthologs were aligned with MUSCLE (Edgar, 2004). Multiple alignments were cured with Gblocks

(Castresana, 2000). The alignment was performed with the full mode and the curation was accomplished with the less stringent settings on the online server Phylogeny.fr (Dereeper *et al.*, 2008). Substitution models for phylogenetic analysis were selected with the ModelGenerator v 0.85 (Keane *et al.*, 2006). Trees were calculated with the maximum-likelihood algorithm, as implemented in PhyML v 3.0. Branch support was calculated by bootstrap analysis, with 1,000 replicates. Phylogenetic trees were edited with FigTree v 1.4 (http://tree.bio.ed.ac.uk/software/figtree).

4 Results

4.1 The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species

This article is in press in **The Biological Bulletin** and a reprinted is used here with the permission from the Marine Biological Laboratory, Woods Hole, MA and the authors. This article's reference is:

Gloeckner, V., Wehrl, M., **Moitinho-Silva, L.**, Gernert, C., Schupp, P.J., Pawlik, J.R., Lindquist, N.L., Erpenbeck, D., Wörheide, G., Hentschel, U. (2014) The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species. *The Biological Bulletin*. In Press.

The documentation of author's contributions is found in the Appendix.

Running head: Microbial abundances in sponges

The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species

Volker Gloeckner[1*], Markus Wehrl[1*], Lucas Moitinho-Silva[1], Christine Gernert[1], Peter Schupp[2], Joseph R. Pawlik[3], Niels L. Lindquist[4], Dirk Erpenbeck[5], Gert Wörheide[5,6], Ute Hentschel[1]

[*] both authors contributed equally to this work.

[1]Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Julius-von-Sachs Platz 3, 97082 Wuerzburg, Germany

[2]Carl-von-Ossietzky University Oldenburg, Institute for Chemistry and Biology of the Marine Environment (ICBM) Schleusenstr. 1, 26382 Wilhelmshaven, Germany

[3]Department of Biology and Marine Biology, Center for Marine Science, University of North Carolina Wilmington, North Carolina, USA

[4]Institute of Marine Sciences, University of North Carolina at Chapel Hill, North Carolina, USA

[5]Department of Earth and Environmental Sciences, Palaeontology & Geobiology, & GeoBio-Center, Ludwig-Maximilians-Universität München, Richard-Wagner-Str. 10, 80333 München, Germany

[6]SNSB-Bayerische Staatssammlung für Paläontologie und Geologie, Richard-Wagner-Str. 10, 80333 München, Germany

Corresponding author:

Email: ute.hentschel@uni-wuerzburg.de; Tel.: +49-931-31-82581; Fax: +49-931-31-86235

Key words: Sponges, Porifera, microorganism, bacteria, HMA LMA dichotomy, transmission electron microscopy

4.1.1 Abstract

The dichotomy between high microbial abundance (HMA) and low microbial abundance sponges (LMA) has been long recognized. In the present study, 56 sponge species from three different geographic regions (greater Caribbean, Mediterranean, Red Sea) were investigated by transmission electron microscopy for the presence of microorganisms in the mesohyl matrix. Additionally, bacterial enumeration by DAPI-counting was performed on a subset of samples. Of the 56 species investigated, 28 were identified as belonging to the HMA and 28 to the LMA category. While the sponge orders Agelasida and Verongida consisted exclusively of HMA species, the Poecilosclerida were composed only of LMA sponges. Other taxa contained both types of microbial associations (e.g., marine Haplosclerida, Homoscleromorpha, Dictyoceratida) and a clear phylogenetic pattern could not be identified. For a few sponge species, an intermediate microbial load was determined and TEM data did not suffice to reliably determine HMA or LMA status. In order to experimentally determine the HMA or LMA status of a sponge species, we recommend therefore a combination of transmission electron microscopy and 16S rRNA gene sequence data. This study significantly expands previous reports on microbial abundances in sponge tissues and contributes to a better understanding of sponge-microbe symbioses.

4.1.2 Introduction

Sponges (Porifera) represent an evolutionarily ancient phylum with a fossil record dating back to Precambrian times (Li *et al.*, 1998). Today, sponges are important components of the marine benthos and play an important role in the coupling of benthic and pelagic environments owing to their immense filter-feeding capacities (Bell, 2008, de Goeij *et al.*, 2013, Gloeckner, 2013). Within their mesohyl tissues, many sponges harbor a great diversity of symbiotic microorganisms from the three domains of life, i.e. Archaea, Bacteria, and Eukaryota. To date, representatives from more than 28 bacterial phyla (including candidate phyla such as Poribacteria) and two archaeal lineages were identified from marine sponges (Hentschel *et al.*, 2012, Schmitt *et al.*, 2012, Simister *et al.*, 2012). The vast majority of sponge-associated microbes remains uncultivated and is thus functionally largely uncharacterized (Taylor *et al.*, 2007).

The presence of microorganisms in marine sponge tissues has been known for almost a century. Dosse (1939) as well as Levi and Porte (1962) were among the first to describe microorganisms in the sponge mesohyl matrix using transmission electron microscopy. It was soon discovered that while some sponge species harbored dense microbial consortia within their mesohyl tissues, the mesohyl of other species from the same habitat were notably devoid of microorganisms (Reiswig, 1974, Vacelet and Donadey, 1977, Wilkinson, 1978). Accordingly, two general categories were identified which were termed “bacterial sponges” or “non-symbiont harboring, normal sponges” (Reiswig, 1981). Later on, the terms “low microbial abundance” (LMA) and “high microbial abundance” (HMA) sponges were coined to acknowledge the additional presence of archaea in sponge tissues (Hentschel *et al.*, 2003). A typical HMA sponge contains 10^8 to 10^{10} microorganisms/g sponge tissue which can make up to 20 – 35% of the sponge biomass (Hentschel *et al.*, 2012, Reiswig, 1981, Webster *et al.*, 2001), while only 10^5 to 10^6 bacteria/g sponge tissue are found in LMA sponges which is roughly equivalent to the microbial abundances in seawater (Hentschel *et al.*, 2006). This pattern extends to reproductive propagules in that the larvae of HMA sponges contain dense bacterial assemblages at the larval center, while the interior of LMA sponge larvae is largely free of microbes (Ereskovsky and Tokina, 2004, Gloeckner *et al.*, 2013a, Gloeckner *et al.*,

2013b, Maldonado, 2007, Schmitt *et al.*, 2007). Vertical microbial transmission from the parent to the larva, a hallmark of symbiotic host-microbe associations (Bright and Bulgheresi, 2010), is now considered an important and presumably evolutionarily ancient component of HMA sponge symbioses. The microbial community is likely complemented by horizontal acquisition of microbes from seawater although this process has never been demonstrated (Schmitt *et al.*, 2008, Webster *et al.*, 2010).

Besides bacterial abundances, there are also noticeable differences with respect to microbial diversity between HMA and LMA sponges. Several studies, employing a variety of 16S rRNA gene-based methods consistently demonstrated a lower microbial diversity in LMA than in HMA sponges (Erwin *et al.*, 2011, Giles *et al.*, 2013, Gloeckner *et al.*, 2013a, Kamke *et al.*, 2010, Moitinho-Silva *et al.*, 2013, Poppell *et al.*, 2013, Schmitt *et al.*, 2012, Weisz *et al.*, 2007). Each LMA species was dominated by a large clade of Proteobacteria (Alpha-, Beta-, or Gamma-) or Cyanobacteria (genus *Synechococcus*), and there was little overlap between the LMA sponge microbiomes under investigation. Alternatively, the HMA sponge communities showed more phylum-level diversity with Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria, candidate phylum Poribacteria, and other phyla as dominant community members. Differences extend also to the microbial physiology of the respective microbial consortia, particularly with respect to nitrogen metabolism (Bayer *et al.*, 2008, Ribes *et al.*, 2012, Schlappy *et al.*, 2010). Distinct differences were further noted regarding the distribution of certain polyketide synthase genes (supA-PKS) that were found in all HMA but not in the LMA sponges under investigation (Hochmuth *et al.*, 2010). Clearly, much remains to be learned about the metabolism, physiology and function of sponge-associated microbial consortia, particularly in the context of the HMA or LMA dichotomy.

With respect to animal morphology, Vacelet and Donadey (1977) observed early on that the HMA sponges generally display a denser mesohyl, narrower aquiferous canals, and smaller choanoycte chambers than their LMA counterparts. In other words, the HMA sponges appear to be less well “irrigated”. The narrower canals and smaller choanocyte chambers may result in a reduced water flow when compared to LMA sponges (Schlappy *et al.*, 2010, Weisz *et al.*, 2008). In the present study, we revisited the original observations on microbial abundances in sponges by use of electron microscopy, which was complemented by DAPI cell-counting. Altogether, 56 demosponge species from four different geographic locations

were investigated. This survey significantly expands previously published datasets on HMA and LMA sponges (Hentschel *et al.*, 2003, Schmitt, 2007, Schmitt *et al.*, 2008, Weisz *et al.*, 2008) and the findings are interpreted in a taxonomic framework of the animal hosts.

4.1.3 Materials and Methods

Sponge collection

Sponge collections from various locations around the Bahamas (termed “BAH”) were obtained by Scuba diving using the research vessels R/V Seward Johnson and F.G. Walton Smith, Harbor Branch Oceanographic Institution (HBOI), USA, over the years 2003-2013 (Table 9). Sponges were further obtained by Scuba diving in 2004 in Key Largo, Florida, USA (25°01′N, 80°23′W) (“FL”), offshore Rovinj, Croatia (45°08’N; 13°64’E) (“MED”) in 2012, at Souda Bay, Crete, Greece (35°31'N; 24°09'E) (“MED”) in 2013, and at other Mediterranean locations. Sponge samples were further collected at Fsar reef, Thuwal, Red Sea, Saudi Arabia (22°23’N; 39°03’E) (“RS”) in 2010 (Table 9). Three to five sponge individuals were collected per each species and brought to the surface in separate Ziploc bags. The samples were processed as described below within a few hours following sampling.

Table 9. Sponge collection sites

Location	Collection Site	Sponge Species
BAH, Bahamas Islands	Little San Salvador 24˚34.39'N; 75˚58.00'W	*Agelas citrina, Agelas dilatata, Aiolochroia crassa, Siphonodictyon coralliphagum, Iotrochota birotulata* *Plakortis* sp.
	San Salvador 24˚01.14'N; 74˚32.68'W	*Agelas dispar, Aplysina cauliformis var. thick, Aplysina insularis, Verongula gigantea, Ptilocaulis* sp.*, Cliona varians*
	Sweetings Cay 26°36'0'N; 77°52'60'W	*Cribrochalina vasculum, Dysidea etheria*
	Chub Cay 25°23'36.52'N; 77°52'12.68'W	*Aplysina cauliformis var.* thin*, Monanchora arbuscula*
	Grand Bahama Island 25°1.282'N; 77°34.56'W	*Aplysina archeri, Aplysina fistularis, Chalinula molitba*
	Exuma Cay 24°25.642'N;76°40.464'W	*Myrmekioderma gyroderma, Spheciospongia vesparium, Erylus formosus*
	Bimini 25°45.316'N; 79°18.061'W	*Batzella rubra, Cinachyrella alloclada*
	Great Inagua 21°05.945'N; 73°40.216'W	*Plakortis lita, Svenzea zeai*
	Cat Cay 25°31.480'N; 79°17.938'W	*Callyspongia plicifera*
MED, Mediterranean	Rovinj, Croatia 45°05'N, 13°38'E	*Aplysina aerophoba, Chondrosia reniformis, Axinella polypoides, Tethya aurantium, Suberites domuncula, Dysidea avara*
	Marseille, France 43°11'48.92'N; 5°21'48.62'E	*Aplysina cavernicola, Oscarella lobularis*
	Banyuls-sur-Mer, France 42°29'N; 03°08'E	*Crambe crambe*
	Souda, Crete, Greece 36°76.759'N; 24°51.422'E	*Petrosia* sp.*, Acanthella acuta, Axinella cannabina*
FL, Florida	Key Largo, USA 24°56.863'N; 80°27.230'W	*Agelas wiedenmayeri, Xestospongia muta, Ircinia felix, Ircinia strobilina, Smenospongia aurea, Ecytoplasia ferox, Scopalina ruetzleri, Tedania ignis, Mycale laxissima, Niphates erecta, Niphates digitalis, Amphimedon compressa, Callyspongia vaginalis, Aplysina lacunosa*
Red Sea	Fsar Reef, Jeddah, Saudi Arabia 22°23.096'N; 39°02.856'E	*Amphimedon ochracea, Crella cyathophora, Stylissa carteri, Xestospongia testudinaria*

Electron microscopy

Freshly collected sponge material of a few mm^3 in size was fixed in 2.5% glutaraldehyde/ phosphate-buffered saline for 12 hrs, rinsed 3x 20 min in PBS and postfixed in 2% osmium tetroxide/ PBS for 12 hrs. Several pieces per individual (hereafter termed “technical replicates”) were dehydrated in an ethanol series (30, 50, 70, 90, 3x100%), incubated 3x 20 min in propylene oxide and polymerized in Epon 812 (Serva, Germany) for 4 days at 60°C. The embedded sponge pieces were sectioned with an ultramicrotom (OM U3, C. Reichert, Austria). For contrasting, 70-80 nm thick sections were post-stained with 0.5% uranyl acetate in methanol for 10 min and Reynolds lead citrate for 5 min. The resulting sections were investigated by electron microscopy (Zeiss EM 10, Zeiss, Germany). Several different images of three biological specimens were inspected for each species.

Bacterial quantification protocol

A piece of sponge tissue of about 1 g was removed with an ethanol-sterilized scalpel blade and rinsed 3x with 0.2 µm filter-sterilized seawater. The tissue cube was cut so that one side always represented the surface tissue. A 10x dilution was obtained by adding 1 ml sponge tissue to 9 ml of 0.2 µm filter-sterilized seawater. The tissue was homogenized with a mortar and pestle and poured through Nitex (100 µm pore size) to remove unground tissue pieces. The suspension was fixed in paraformaldehyde to a final concentration of 3.7% and stored at 4˚C until use. The tissue remainders were rinsed off the Nitex sheet with 0.5 ml filter-sterilized seawater and homogenized in 4.5 ml of filter-sterilized seawater using a mortar and pestle. The suspension was poured through Nitex, fixed in 3.7% paraformaldehyde final concentration and stored at 4˚C until use. Dilutions ranging from 10^{-1}- 10^{-3} were prepared from each homogenate. One ml of each dilution was stained with 0.7 µg/ ml DAPI (4,6-Diamidino-2-phenylindole) final concentration in the dark for 30 min. The DAPI stock solution (100 µg ml^{-1}) was prepared weekly and stored at 4 ˚C. A volume of 1 ml stained homogenate was added to 9 ml filter-sterilized seawater and filtered onto a black, 25 mm diameter, 0.2 µm polycarbonate membrane (Millipore, Germany) which was supported

by an 0.45 µm cellulose nitrate filter (Schleicher und Schuell, Germany). Vacuum (< 10 cm Hg) was applied carefully with a hand pump. The filters were washed once with filter-sterilized seawater and subsequently rinsed with 3 ml 70% ethanol. The 0.2 µm polycarbonate filter was then mounted with Citifluor (Citifluor Ltd., UK) onto a microscope slide. Bacterial numbers were determined following epifluorescence microscopy using a 100x magnification oil lense (Axiolab Microscope Zeiss, Germany). Three independent specimens were processed for each species. Each sample represents an average bacterial number from 10 different counting fields. For each sample, the number of bacteria and nuclei was counted and the results from the first and second homogenate were summed. The number of cyanobacteria were counted using the red and green fluorescent filter set.

18S rRNA sponge phylogeny

Nearly full-length 18S rRNA gene sequences published in NCBI GenBank (http://www.ncbi.nlm.nih.gov) were analyzed. In few cases, the 18S rRNA gene sequences were unavailable and were therefore substituted by the sequence of a closely related congeneric (and in the case of *Batzella rubra*, by a confamiliar sequence), taking the current changes in demosponge classification into consideration (Redmond *et al.*, 2013). Altogether 45 sponge species were included in the phylogenetic tree and the GenBank accession numbers are provided in Figure 9. The sequences were aligned using the Sponge Genetree Server with 18S rRNA gene secondary structure information included in the analysis (Erpenbeck *et al.*, 2008). Positions that could not be aligned were excluded from further analyses. Maximum-likelihood reconstruction was inferred with RAxML 7.2.5 (Stamatakis, 2006) using the GTRGAMMI model of nucleotide substitution as suggested by jModeltest v0.1.1. (Darriba *et al.*, 2012) under the Akaike Information Criterion (Akaike, 1974) and 100 fast bootstrap replicates. Analyses were performed on the 64-node Linux cluster of the Molecular Geo- and Palaeobiology Labs, LMU Munich.

4.1.4 Results and Discussion

In the present study, 56 sponge species were inspected by transmission electron microscopy for the presence of microorganisms in the mesohyl matrix. Of the 41 sponge species examined from the greater Caribbean (Bahamas and Florida locations, Table 9), 24 were identified as HMA and 17 as LMA (Tables 10, 11). Of the 12 species sampled from the various Mediterranean locations, four were assigned as HMA and eight as LMA. From the Red Sea collection site, one sponge was identified HMA and three as LMA. For previous TEM-based surveys on HMA and LMA sponge patterns, the reader is referred to the literature (Hentschel et al. (2003) and references cited herein, Schmitt et al. (2008), Weisz et al. (2008), Popell et al. (2014)). Judging from our data, the HMA LMA dichotomy may be best described as a continuum with a highly bimodal distribution, in the sense that most investigated species are found at the extreme ends of the continuum.

Table 10. A compilation of high microbial abundance (HMA) sponge species

Species	**Collection site**	**TEM reference**	**Higher Taxon**
Agelas citrina	BAH	Wehrl 2006	Agelasida
Agelas dilatata	BAH	Wehrl 2006	Agelasida
Agelas dispar	BAH	this study	Agelasida
Agelas wiedenmayeri	FL	Wehrl 2006, Schmitt et al. 2008	Agelasida
Aiolochroia crassa	BAH	this study	Verongida
Aplysina aerophoba	MED	Hentschel et al. 2003, Siegl et al. 2008	Verongida
Aplysina archeri	BAH	Wehrl 2006	Verongida
Aplysina cauliformis, thick morphotype	BAH	Wehrl 2006	Verongida
Aplysina cauliformis, thin morphotype	BAH	Wehrl 2006	Verongida
Aplysina cavernicola	MED	Wehrl 2006, Friedrich et al. 1998, 2001	Verongida
Aplysina fistularis	BAH	Wehrl 2006, Gloeckner 2013	Verongida
Aplysina insularis	BAH	Wehrl 2006	Verongida
Aplysina lacunosa	FL	Wehrl 2006	Verongida
Chondrosia reniformis	MED	Wehrl 2006	Chondrosida
Cribrochalina vasculum	BAH	Schiller 2006	(marine) Haplosclerida
Ectyoplasia ferox	FL	Schmitt et al. 2008, Gloeckner et al. 2013	Raspailiidae
Ircinia felix	FL	Schmitt et al. 2007	Dictyoceratida
Ircinia strobilina	FL	Schmitt 2007	Dictyoceratida
Myrmekioderma gyroderma	BAH	Gloeckner 2013	"Halichondrida"
Petrosia sp.	MED	this study	(marine) Haplosclerida
Plakortis lita	BAH	this study	Homoscleromorpha
Plakortis sp.	BAH	Laroche et al. 2007	Homoscleromorpha
Siphonodictyon coralliphagum	BAH	Schiller 2006, Schmitt et al. 2008	(marine) Haplosclerida
Smenospongia aurea	FL	Schmitt et al., 2008, Gloeckner 2013	Dicytoceratida
Spheciospongia vesparium	BAH	this study	"Hadromerida"
Svenzea zeai	BAH	this study	"Halichondrida"
Verongula gigantea	BAH	Wehrl 2006	Verongida
Xestospongia muta	FL	Wehrl 2006, Hentschel et al. 2006	(marine) Haplosclerida

Xestospongia testudinaria	RS	this study	(marine) Haplosclerida

Higher taxon names in quotation marks indicate orders recognized as non-monophyletic which may be subject to emending in the future.
Some TEM data were reported in Master's (Schiller 2006) and PhD theses (Wehrl 2006, Schmitt 2007, Gloeckner 2013).

Table 11. A compilation of low microbial abundance (LMA) sponge species

Species	Collection site	TEM reference	Higher Taxon
Acanthella acuta	MED	this study	"Halichondrida"
Amphimedon compressa	FL	Angermeier et al. 2012	(marine) Haplosclerida
Amphimedon ochracea	RS	this study	(marine) Haplosclerida
Axinella cannabina	MED	this study	"Halichondrida"
Axinella polypoides	MED	Wehrl 2006	"Halichondrida"
Batzella rubra	BAH	Gloeckner 2013	Poecilosclerida (s.s.)
Callyspongia plicifera	BAH	Gloeckner et al. 2013	(marine) Haplosclerida
Callyspongia vaginalis	FL	Schiller 2006, Wehrl 2006	(marine) Haplosclerida
Chalinula molitba	BAH	Schiller 2006, Wehrl 2006	(marine) Haplosclerida
Cinachyrella alloclada	BAH	Gloeckner 2013	Spirophorida
Cliona varians	BAH	Schiller 2006	"Hadromerida"
Crambe crambe	MED	Wehrl 2006	Poecilosclerida (s.s.)
Crella cyathophora	RS	Giles et al. 2012	Poecilosclerida (s.s.)
Dysidea avara	MED	Wehrl 2006	Dictyoceratida
Dysidea etheria	BAH	Schiller 2006	Dictyoceratida
Erylus formosus	BAH	this study	Astrophorida
Iotrochota birotulata	BAH	Wehrl 2006	Poecilosclerida (s.s.)
Monanchora arbuscula	BAH	this study	Poecilosclerida (s.s.)
Mycale laxissima	FL	Wehrl 2006	Poecilosclerida (s.s.)
Niphates digitalis	FL	Schiller 2006, Wehrl 2006	(marine) Haplosclerida
Niphates erecta	FL	Wehrl 2006	(marine) Haplosclerida
Oscarella lobularis	MED	Gloeckner et al. 2013	Homoscleromorpha
Ptilocaulis sp.	BAH	Wehrl 2006	"Halichondrida"
Scopalina ruetzleri	FL	Wehrl 2006, Gloeckner 2013	"Halichondrida"
Stylissa carteri	RS	Giles et al. 2012	"Halichondrida"
Suberites domuncula	MED	Wehrl 2006	"Hadromerida"
Tedania ignis	FL	Schiller 2006, Wehrl 2006	Poecilosclerida (s.s.)

Tethya aurantium	MED	Wehrl 2006	"Hadromerida"

Higher taxon names in quotation marks indicate orders recognized as non-monophyletic and may be subject to emending in the future.
Some TEM data were reported in Master’s (Schiller 2006) and PhD theses (Wehrl 2006, Schmitt 2007, Gloeckner 2013).

The HMA sponge tissues contained dense and morphologically diverse microorganisms that are located largely extracellularly in the mesohyl matrix (Figure 7A-I). In extreme cases, the microbial cells are much more abundant than sponge cells, such as shown here for *Aplysina aerophoba* (Figure 7A) and *Spheciospongia vesparium* (Figure 7B) as well as elsewhere for example for the order Verongida (Friedrich *et al.*, 1999, Vacelet, 1975). Our data support Weisz et al (2008) but disagree with Popell et al. (2013) in that *S. vesparium* is classified as an HMA sponge. While the potential for phenotypic plasticity in microbial amount cannot be ruled out, we have found microbial abundance to be a highly conserved trait on the level of sponge species.

In general, there is a remarkably stable presence of certain bacterial morphotypes, of which some appear to contain intracellular compartments and unusual membrane structures (Figure 7C). Some of these originally described morphotypes (i.e, types A and C according to (Friedrich *et al.*, 1999, Vacelet, 1975) can readily be identified in the present HMA sponge tissue pictures. The amount of bacteria in the HMA sponge tissues can however be variable, ranging from densely packed mesohyl tissues for the orders Verongida (Figure 7A) to moderately dense microbial consortia, such as those of the taxa *Ircinia* or *Agelas* (Figures 7 G, H). However, additional literature reports based on 16S rRNA gene sequencing and inspections of microbial abundances in larval tissues clearly identified these species as HMA sponges (Schmitt *et al.*, 2008, Schmitt *et al.*, 2007, Susanne Schmitt, 2008). Experimental artefacts may have possibly arisen for sponges that were difficult to cut with a diving knife (ie, *Ircinia*), or for sponges having tissues that are less cohesive (ie, *Xestospongia*). However, the consistency between technical replicates (representing samples from the same individual) and between biological replicates (representing samples from different individuals) is remarkably high in our experience.

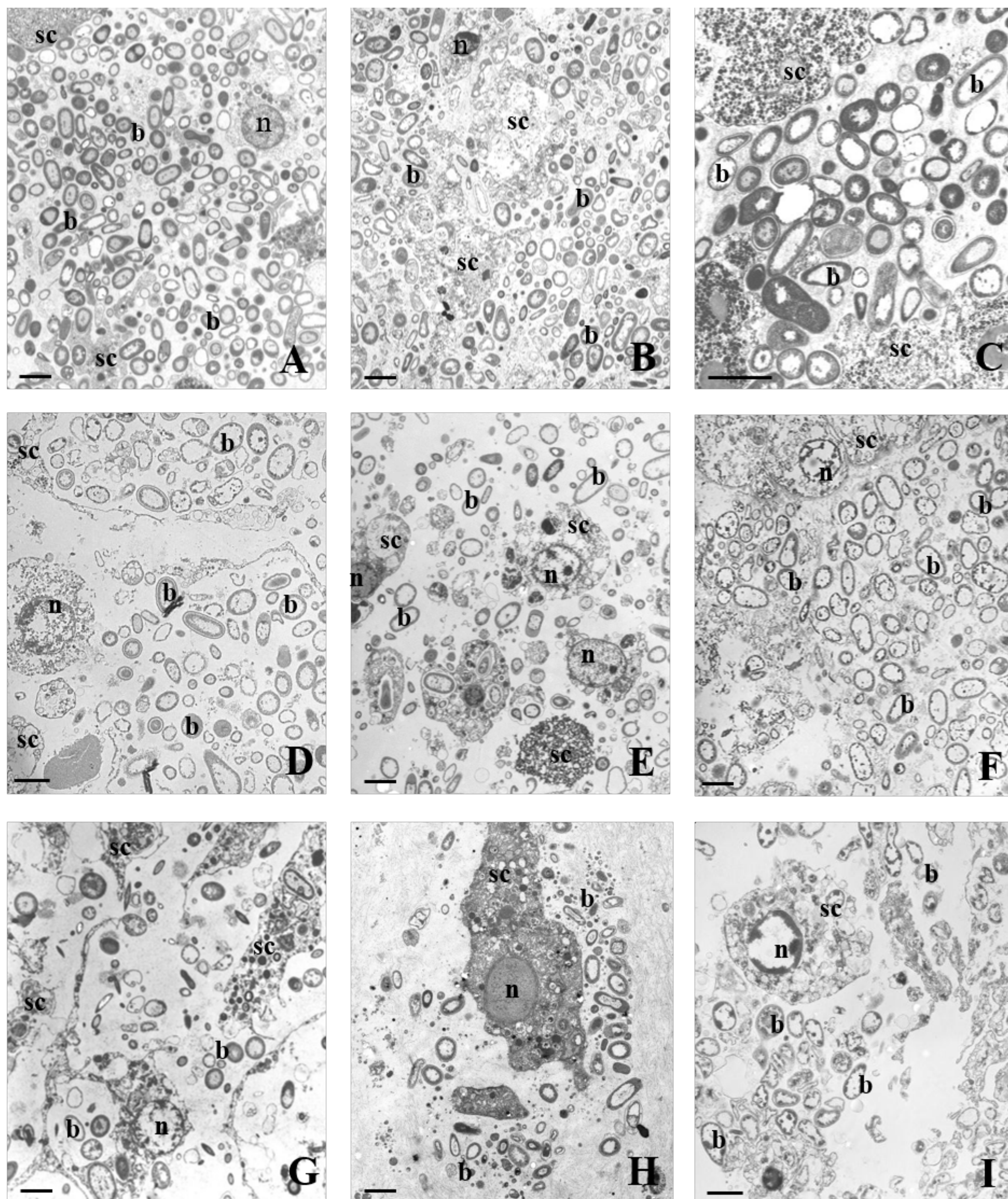


Figure 7. Transmission electron microscopy of selected HMA sponge species. A= *Aplysina aerophoba*, B= *Spheciospongia vesparium*, C= *Aiolochroia crassa*, D= *Petrosia* sp., E= *Xestospongia testudinaria*, F= *Plakortis lita*, G= *Agelas dispar*, H= *Ircinia felix*, I= *Svenzea zeai*. Scale bar 2µm, b=bacteria, n=nucleus, sc=sponge cell.

Relative to HMA sponge species, the mesohyl of LMA sponges was noticeably devoid of microorganisms (Figure 8). Few intracellular bacteria in various stages of digestion were

present that represent likely food bacteria (Figure 8F). Occasionally, bacterial morphotypes were observed in TEM pictures of LMA sponges (Gloeckner *et al.*, 2013a); however, the morphotype diversity appeared unlike that of HMA sponges in that the compartmentalized cells and those with unusual membranes were missing (Friedrich *et al.*, 1999, Vacelet, 1975). Therefore, on the basis of electronmicroscopical observations, the combination of bacterial abundance and morphotype diversity determines whether a sponge species belongs to the HMA or LMA category. However, in some sponges with intermediate bacterial abundances, TEM may not be sufficient to determine whether a given species is HMA or LMA, and in these exceptional cases, additional methods are needed.

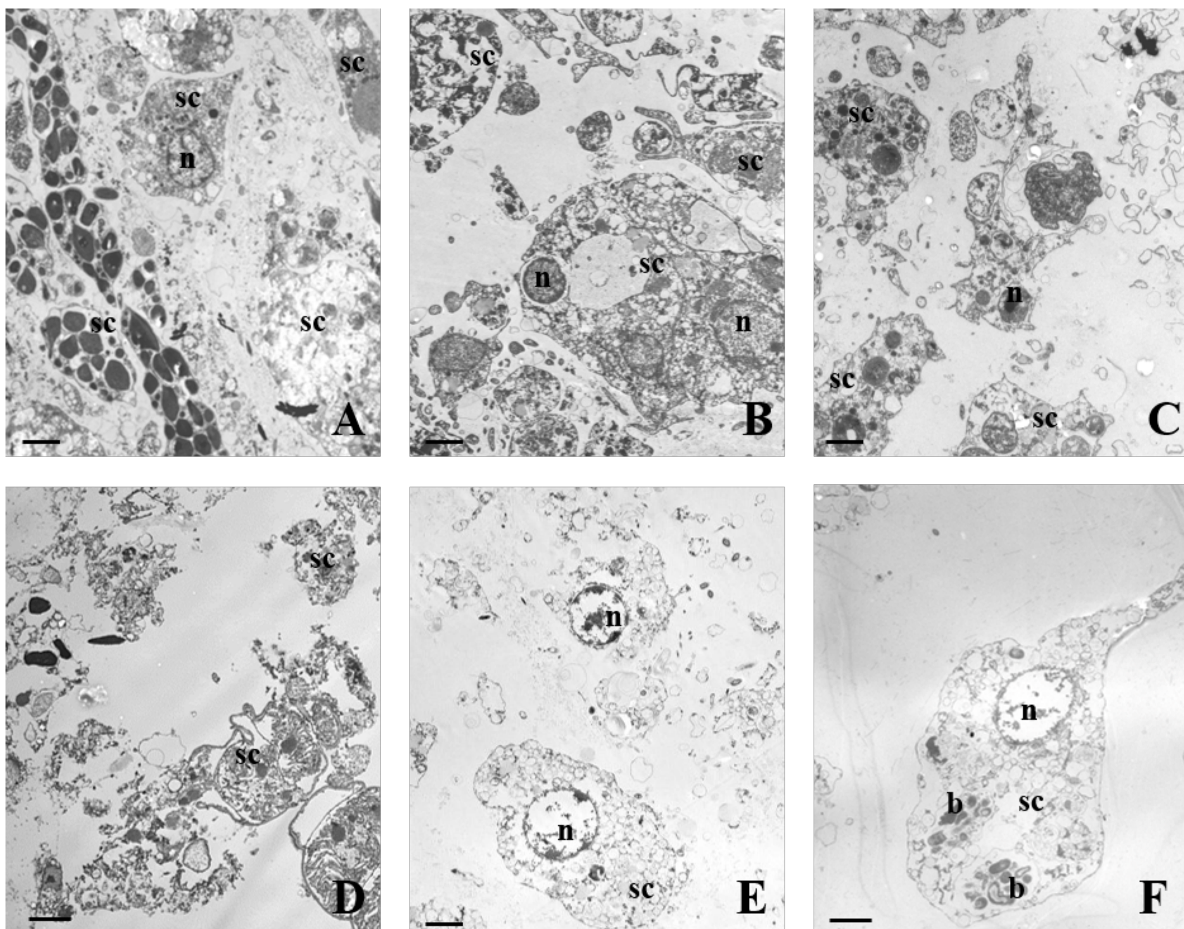


Figure 8. Transmission electron microscopy of selected LMA sponges. A= *Erylus formosus*, B= *Monanchora arbuscula*, C= *Scopalina ruetzleri*, D= *Amphimedon ochracea*, E= *Axinella cannabina*, F= *Acanthella acuta*. Scale bar 2µm, b=bacteria, n=nucleus, sc=sponge cell.

When the status of a sponge species as either HMA or LMA was equivocal on the basis of microscopical observations, an independent line of evidence was sought in microbial enumeration by DAPI-staining (Table 12). Sponge tissue homogenates were used for bacterial and cyanobacterial quantification and numbers were either expressed in total or as ratios relative to the number of sponge nuclei. Further, sponge homogenates were screened for the presence or absence of bacteria, using defined HMA and LMA sponge homogenates as controls. The sponge species subjected to this analysis formed two distinct groups of HMA and LMA sponges, which mirrored the TEM observations. *Siphonodictyon coralliphagum* was the only exception, in which the low bacterial numbers obtained by DAPI staining indicated it was an LMA sponge while the TEM observations identified this species as an HMA sponge (Schmitt, 2007). The DAPI-method can be prone to errors; for example, if the bacterial cells are disrupted upon tissue homogenisation, particularly as a result of the presence of cytotoxic secondary metabolites. However, this problem happened only once in 56 species investigated and appeared to be a rare scenario. The ratio of bacteria per sponge cell nuclei was variable, ranging from 17 to 95 bacterial cells per sponge cell nucleus for *X. muta* and *A. insularis*, respectively, and from 0 to 5 bacterial cells per sponge cell nucleus for the LMA sponges (Table 12). These data support the TEM observations that HMA sponges contain variable amounts of microorganisms in the mesohyl tissues. Major disadvantages of the DAPI-staining method were the high amount of background staining and the limited applicability to sponges with high intrinsic autofluorescence. Also, some sponge tissues were difficult to macerate and yielded clumpy homogenates, making it difficult to determine precise numbers. In these cases, confocal microscopy on sponge tissue sections, as recently employed by Ribes and coworkers (Ribes *et al.*, 2012), is a suitable alternative for microbial visualisation and quantification in sponge tissues.

Several recent publications addressed the HMA-LMA dichotomy by 16S rRNA gene sequencing. Whether obtained by clone libraries (Giles *et al.*, 2013, Kamke *et al.*, 2010), DGGE (Gloeckner *et al.*, 2013a, Poppell *et al.*, 2013, Weisz *et al.*, 2007), terminal restriction fragment length polymorphism (T-RFLP) (Erwin *et al.*, 2011), or amplicon sequencing (Moitinho-Silva *et al.*, 2013, Schmitt *et al.*, 2012), the data consistently revealed a different bacterial composition in LMA sponges than in HMA sponges. Similarly, in comparative DGGEs, the LMA sponges consistently displayed less complex banding patterns than their HMA counterparts (Gerce *et al.*, 2011, Poppell *et al.*, 2013, Weisz *et al.*, 2007). Even small

16S rRNA gene clone libraries were sufficient to detect the major phylogenetic lineages (Moitinho-Silva, Hentschel, unpublished data). 16S rRNA gene data have therefore proven to be very useful to infer the HMA or LMA status of the host sponge.

Table 12. Quantification of bacteria, cyanobacteria and sponge nuclei by DAPI-staining (per ml^{-1} sponge homogenate)

Species	Bacteria Mean ± SE	Cyanobacteria Mean ± SE	Bacteria total Mean ± SE	Nuclei Mean ± SE	Ratio*
High microbial abundance sponges					
Agelas citrina	$7.0 \times 10^8 \pm 2.1 \times 10^8$	$2.6 \times 10^6 \pm 2.0 \times 10^6$	$7.0 \times 10^8 \pm 2.1 \times 10^8$	$3.2 \times 10^7 \pm 5.7 \times 10^6$	22.16
A.dilatata	$1.9 \times 10^9 \pm 8.7 \times 10^8$	$1.5 \times 10^6 \pm 9.6 \times 10^5$	$1.9 \times 10^9 \pm 8.7 \times 10^8$	$5.7 \times 10^7 \pm 1.7 \times 10^7$	33.40
Aplysina archeri	$1.4 \times 10^9 \pm 1.9 \times 10^7$	$1.1 \times 10^7 \pm 5.8 \times 10^6$	$1.4 \times 10^9 \pm 2.4 \times 10^7$	$3.5 \times 10^7 \pm 1.8 \times 10^7$	39.16
A. insularis	$6.2 \times 10^9 \pm 4.2 \times 10^8$	$2.9 \times 10^8 \pm 7.3 \times 10^7$	$6.5 \times 10^9 \pm 4.9 \times 10^8$	$6.8 \times 10^7 \pm 2.2 \times 10^6$	95.71
Cribochalina vasculum	$5.5 \times 10^9 \pm 2.3 \times 10^9$	$1.1 \times 10^9 \pm 1.7 \times 10^7$	$6.6 \times 10^9 \pm 2.4 \times 10^7$	$2.1 \times 10^8 \pm 3.3 \times 10^7$	31.71
Xestospongia muta	$7.0 \times 10^9 \pm 5.5 \times 10^8$	$1.2 \times 10^9 \pm 2.8 \times 10^8$	$8.2 \times 10^9 \pm 7.7 \times 10^8$	$4.8 \times 10^8 \pm 4.1 \times 10^7$	16.96
Ircinia felix	$1.0 \times 10^9 \pm 7.3 \times 10^7$	$5.7 \times 10^8 \pm 7.1 \times 10^6$	$1.6 \times 10^9 \pm 7.4 \times 10^7$	$6.1 \times 10^7 \pm 1.2 \times 10^7$	25.67
Plakortis sp.	$4.3 \times 10^9 \pm 8.0 \times 10^8$	$3.6 \times 10^7 \pm 1.6 \times 10^7$	$4.3 \times 10^9 \pm 7.9 \times 10^8$	$4.8 \times 10^7 \pm 1.6 \times 10^7$	89.23
Ecytoplasia ferox	$8.7 \times 10^9 \pm 1.7 \times 10^8$	n.d.	$8.7 \times 10^9 \pm 1.7 \times 10^8$	$3.8 \times 10^8 \pm 2.6 \times 10^7$	22.94

Low microbial abundance sponges

Iotrochota birotulata	2.4 x 10^9 ± 2.8 x 10^8	n.d.	2.4 x 10^9 ± 2.8 x 10^8	5.2 x 10^8 ± 4.3 x 10^7	4.67
S. coralliphagum§	2.0 x 10^8 ± 1.5 x 10^7	n.d.	2.0 x 10^8 ± 1.5 x 10^7	2.2 x 10^8 ± 6.2 x 10^7	0.93
Dictyonella funicularis	n.d.	n.d.	n.d.	3.2 x 10^8 ± 1.4 x 10^7	0.00
Tedania ignis	n.d.	n.d.	n.d.	8.9 x 10^8 ± 5.2 x 10^7	0.00
Chalinula molitba	3.5 x 10^6 ± 3.5 x 10^6	n.d.	3.5 x 10^6 ± 3.5 x 10^6	6.1 x 10^8 ± 7.4 x 10^7	0.01
Niphates digitalis	n.d.	n.d.	n.d.	4.6 x 10^8 ± 1.3 x 10^8	0.00
Amphimedon compressa	n.d.	n.d.	n.d.	1.4 x 10^8 ± 2.3 x 10^7	0.00
Callyspongia vaginalis	2.2 x 10^6 ± 1.6 x 10^6	1.8 x 10^6 ± 1.8 x 10^6	4.0 x 10^6 ± 2.0 x 10^6	5.5 x 10^8 ± 1.1 x 10^7	0.01
Ptilocaulis sp.	n.d.	n.d.	n.d.	3.7 x 10^8 ±3.2 x 10^7	0.00
Cliona varians	n.d.	n.d.	n.d.	2.0 x 10^8 ± 2.1 x 10^7	0.00
Dysidea etheria	3.1 x 10^6 ± 1.2 x 10^6	1.3 x 10^6 ± 7.6 x 10^5	4.4 x 10^6 ± 1.9 x 10^6	1.9 x 10^8 ± 8.1 x 10^6	0.02
Caribbean seawater			3.5 x 10^5 ± 5.7 x 10^4		

* Ratio of total bacteria:nuclei

§ For *Siphonodictyon coralliphagum*, the low bacterial bacterial numbers determined by DAPI-staining contradict its status as an HMA sponge, as determined by TEM (Schmitt et al 2008).

The following additional sponges were identified as HMA by qualitative DAPI-screening: *Ectyoplasia ferox, Myrmekioderma gyroderma, Agelas dispar, Ircinia felix, Ircinia strobilina, Smenospongia aurea, Aplysina cauliformis, Aplysina fistularis, Verongula gigantea, Aiolochroia crassa*, *Geodia neptuni, Chondrosia collectrix, Calyx podatypa*. The following additional sponges were identified as LMA by qualitative DAPI-screening: *Erylus formosus, Mycale* (*Arenochalina*) *laxissima, Batzella rubra, Monanchora arbuscula, Scopalina ruetzleri, Callyspongia plicifera, Callyspongia vaginalis,*

Cinachyrella alloclada.

The host phylogeny based on nearly full-length 18S rRNA gene sequences corroborates the current molecular phylogenetic hypotheses of demosponges (e.g., Morrow et al. 2012; see also Redmond et al. (2013) for the most comprehensive 18S rRNA gene phylogeny). However as the new nomenclature is yet to be finalized we refer to taxon names as currently used in the World Porifera Database (van Soest et al., 2014) Several clades were recovered that consisted in the present sample set exclusively of HMA taxa (Verongida, Agelasida (*Agelas*)), while the Poecilosclerida (s.s.) clade with representatives from six different families consisted exclusively of LMA species (Figure 9). For the remaining clades, of which many orders and families are currently under redefinition (see e.g. Redmond *et al.*, 2013), no distinct HMA/ LMA distribution patterns were recognized. Additional and distinct patterns might however be revealed on lower taxonomic levels given a more representative taxonomic sampling. It is noteworthy however that the HMA/ LMA characteristics are conserved in closely related species over time and space when collected from different geographic regions (Montalvo and Hill, 2011, Wilkinson, 1978).

The question remains open as to what causes the HMA-LMA dichotomy. A survey of the literature revealed no apparent correlation with host defense status (Chanas and Pawlik, 1995, Pawlik *et al.*, 1995), with reproductive mode (oviparous vs viviparous), or with ecological parameters as both types of sponges appear in the same habitat in close vicinity (Schiller, 2006). A phylogenetic signature is present only to a moderate extent. At this point it appears most likely that sponge morphology is an important determining factor. The HMA sponges are frequently large and massive and generally have a firm touch and fleshy consistency, while LMA sponges are generally smaller and feel fragile, soft, brittle, or tough (U.H. University of Wuerzburg, personal observations). Similarly, the architecture of the sponge interior plays a determining role (Vacelet and Donadey, 1977, Weisz *et al.*, 2008). A higher choanocyte chamber density in LMA sponges was reported by Poppell (2013) and Schlappy (2010). A current hypothesis states that LMA sponges invest more energy into feeding structures while the nutrition of HMA sponges is supplemented by their microbial symbionts (Poppell *et al.*, 2013). With respect to the postulated role of symbionts in the “sponge loop” (de Goeij *et al.*, 2013), an improved understanding of the different feeding strategies of the sponge “holobiont” is clearly a worthwhile undertaking. It is further safe to speculate that

HMA sponges are morphologically adapted to house microbial consortia within their tissues. In evolutionary terms, the question arises whether the sponges are preconditioned to host microbes or whether the animal tissue morphology is a consequence of containing the microbes within. If preconditioned, it is conceivable that the extracellular matrix (ECM) of HMA sponges may be structurally altered to accommodate the presence of sponge symbionts, and vice versa, that the sponge symbionts may have mechanisms to survive within ECM, modifying it with their activities. Indeed, a recent study provides compelling evidence obtained by single cell genomics that poribacterial sponge symbionts can degrade ECM for nutritional purposes (Kamke *et al.*, 2013).

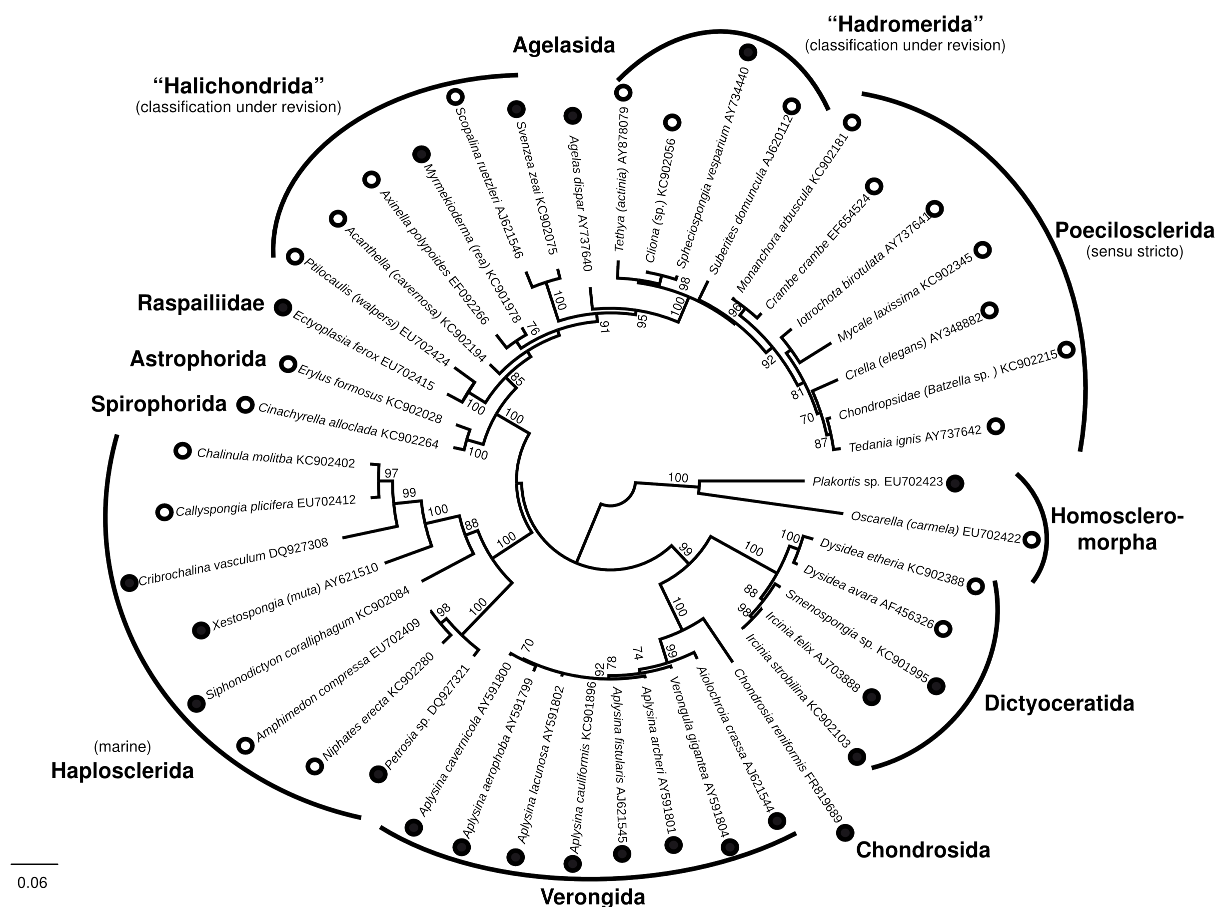


Figure 9. Maximum-likelihood reconstruction of the sponge hosts using almost complete 18S rRNA gene sequences as a phylogenetic marker. Taxon names in parentheses describe the closest related taxon that was used when the 18S rRNA gene sequence of a given host species was unavailable. Number at the branches are bootstrap probabilities > 70%. The taxon names are followed by the NCBI GenBank accession numbers. Black filled dots depict HMA sponges, unfilled dots LMA sponges. Scale bar: substitutions /site.

While the present study is the most comprehensive survey for HMA and LMA patterns to date, more investigations are needed with higher taxonomic depth, including analysis of specimens from other sponge classes besides Demospongiae, and including more locations, such as the Great Barrier Reef, the deep-sea and the polar seas. With regard to methodologies, we propose a combination of transmission electron microscopy and 16S rRNA gene sequence data to reliably determine the HMA or LMA status of the host sponge. The latter is particularly recommended when TEM data yield ambiguous results. The present sponge survey will help pave the way for a functional understanding of the HMA-LMA dichotomy in sponges.

4.1.5 Acknowledgements

We gratefully acknowledge the marine operations personnel at the National Undersea Research Center (Key Largo, Florida), the Ruder Boskovic Institute (Rovinj/Croatia), the Hellenic Center for Marine Research (HCMR, Crete, Greece), and the Coastal and Marine Resources Core Lab at KAUST (Thuwal, Saudi Arabia), in particular T. Ravasi and M. Berumen, for expert help during sponge collection. The crews of the RV Seward Johnson II and F. G. Walton Smith (HBOI) are acknowledged for excellent support during field work. Financial support was provided by the DFG, SFB567-TPC3 to U.H.

4.1.6 References

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* **19**:716-723.

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

Bell, J. J. 2008. The functional roles of marine sponges. *Estuar Coast Shelf S* **79**:341-353.

Bright, M., and S. Bulgheresi. 2010. A complex journey: transmission of microbial symbionts. *Nature Rev. Microbiol.* **8**:218-230.

Chanas, B., and J. R. Pawlik. 1995. Defenses of Caribbean sponges against predatory reef fish II. spicules,

tissue toughness, and nutritional quality. *Mar Ecol Prog Ser* **127**:195-211.

Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. *Nat. Meth.* **9**:772-772.

de Goeij, J. M., D. van Oevelen, M. J. Vermeij, R. Osinga, J. J. Middelburg, A. F. de Goeij, and W. Admiraal. 2013. Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342**:108-110.

Dosse, G. 1939. Bakterien- und Pilzbefunde sowie pathologische und Fäulnis-Vorgänge im Meeres- sowie Süßwasserschwämmen. Untersuchungen mit dem gegenwärtigen Sterben der Badeschwämme in Westindien *Zeitschrift fuer Parasitenkunde* **11**:331-356.

Ereskovsky, A. V., and D. B. Tokina. 2004. Morphology and fine structure of the swimming larvae of *Ircinia oros* (Porifera, Demospongiae, Dictyoceratida). *Invertebr. Reprod. Dev.* **45**:137-150.

Erpenbeck, D., O. Voigt, M. Gultas, and G. Worheide. 2008. The sponge genetree server-providing a phylogenetic backbone for poriferan evolutionary studies. *Zootaxa*:58-60.

Erwin, P. M., J. B. Olson, and R. W. Thacker. 2011. Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PLoS One* **6**:e26806.

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

Gerce, B., T. Schwartz, C. Syldatk, and R. Hausmann. 2011. Differences between bacterial communities associated with the surface or tissue of Mediterranean sponge species. *Microb. Ecol.* **61**:769-782.

Giles, E. C., J. Kamke, L. Moitinho-Silva, M. W. Taylor, U. Hentschel, T. Ravasi, and S. Schmitt. 2013. Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiol. Ecol.* **83**:232-241.

Gloeckner, V. 2013. Untersuchungen zur Diversität, Abundanz und vertikalen Weitergabe von Bakterien in marinen Schwämmen. PhD thesis, University of Wuerzburg, Wuerzburg, Germany.

Gloeckner, V., U. Hentschel, A. V. Ereskovsky, and S. Schmitt. 2013a. Unique and species-specific microbial communities in *Oscarella lobularis* and other Mediterranean *Oscarella* species (Porifera: Homoscleromorpha). *Mar. Biol.* **160**:781-791.

Gloeckner, V., N. Lindquist, S. Schmitt, and U. Hentschel. 2013b. *Ectyoplasia ferox*, an experimentally tractable model for vertical microbial transmission in marine sponges. *Microb. Ecol*. **65**:462-474.

Hentschel, U., L. Fieseler, M. Wehrl, C. Gernert, M. Steinert, J. Hacker, and M. Horn. 2003. Microbial diversity of marine sponges. *Prog. Mol. Subcell. Biol.* **37**:59-88.

Hentschel, U., J. Piel, S. M. Degnan, and M. W. Taylor. 2012. Genomic insights into the marine sponge microbiome. *Nature Rev. Microbiol.* **10**:641-654.

Hentschel, U., K. M. Usher, and M. W. Taylor. 2006. Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.* **55**:167-177.

Hochmuth, T., H. Niederkruger, C. Gernert, A. Siegl, S. Taudien, M. Platzer, P. Crews, U. Hentschel, and J. Piel. 2010. Linking chemical and microbial diversity in marine sponges: possible role for Poribacteria as producers of methyl-branched fatty acids. *Chembiochem* **11**:2572-2578.

Kamke, J., A. Sczyrba, N. Ivanova, P. Schwientek, C. Rinke, K. Mavromatis, T. Woyke, and U. Hentschel.

2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME Journal* **7**:1-14.

Kamke, J., M. W. Taylor, and S. Schmitt. 2010. Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME Journal* **4**:498-508.

Levi, C., Porte, A. 1962. Electron microscopy study of the sponge *Oscarella lobularis* and its amphiblastula larvae. *Cahiers de Biologie Marine* **3**:307-315.

Li, C. W., J. Y. Chen, and T. E. Hua. 1998. Precambrian sponges with cellular structures. *Science* **279**:879-882.

Maldonado, M. 2007. Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. *J Mar. Biol. Assoc. Uk* **87**:1701-1713.

Moitinho-Silva, L., K. Bayer, C. V. Cannistraci, E. C. Giles, T. Ryu, L. Seridi, T. Ravasi, and U. Hentschel. 2013. Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Mol. Ecol.*

Montalvo, N. F., and R. T. Hill. 2011. Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts. *Appl. Environ. Microbiol.* **77**:7207-7216.

Pawlik, J. R., B. Chanas, R. J. Toonen, and W. Fenical. 1995. Defenses of Caribbean sponges against predatory reef fish I: fhemical deterrency. *Mar. Ecol. Prog. Ser.* **127**:183-194.

Poppell, E., J. B. Weisz, L. Spicer, A. Massaro, A. Hill, and M. Hill. 2013. Sponge heterotrophic capacity and bacterial community structure in high- and low-microbial abundance sponges. *Mar. Ecol.* 1-11.

Redmond, N. E., C. C. Morrow, R. W. Thacker, M. C. Diaz, N. Boury-Esnault, P. Cardenas, E. Hajdu, G. Lobo-Hajdu, B. E. Picton, S. A. Pomponi, E. Kayal, and A. G. Collins. 2013. Phylogeny and systematics of demospongiae in light of new small-subunit ribosomal DNA (18S) sequences. *Integr. Comp. Biol.* **53**:388-415.

Reiswig, H. M. 1974. Water transport, respiration and energetics of three tropical marine sponges *Journal of Experimental Mar. Biol. Ecol.* **14**:231-249.

Reiswig, H. M. 1981. Partial carbon and energy budgets of the bacteriosponge *Verongia fistularis* (Porifera: Demospongiae) in Barbados West-Indies *Mar. Biol.* **2**:273-294.

Ribes, M., E. Jimenez, G. Yahel, P. Lopez-Sendino, B. Diez, R. Massana, J. H. Sharp, and R. Coma. 2012. Functional convergence of microbes associated with temperate marine sponges. *Environ. Microbiol.* **14**:1224-1239.

Schiller, R. 2006. Untersuchungen zur Bakterienhaltigkeit in karibischen Schwämmen und ausgewählten Reproduktionsstadien, University of Wuerzburg, Ms thesis,Wuerzburg, Germany.

Schlappy, M. L., S. I. Schottner, G. Lavik, M. M. Kuypers, D. de Beer, and F. Hoffmann. 2010. Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Mar. Biol* .**157**:593-602.

Schmitt, S. 2007. Vertical microbial transmission in Caribbean bacteriosponges. PhD thesis, University of Wuerzburg, Wuerzburg, Germany.

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

Schmitt, S., P. Tsai, J. Bell, J. Fromont, M. Ilan, N. Lindquist, T. Perez, A. Rodrigo, P. J. Schupp, J. Vacelet, N. Webster, U. Hentschel, and M. W. Taylor. 2012. Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME Journal* **6**:564-576.

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

Simister, R. L., P. Deines, E. S. Botte, N. S. Webster, and M. W. Taylor. 2012. Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ. Microbiol.* **14**:517-524.

Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**:2688-2690.

Schmitt, S., Wehrl, M., Lindquist, L., Weisz, J.B., and U. Hentschel. 2008. Morphological and molecular analyses of microorganisms in Caribbean reef adult sponges and in corresponding reproductive material. *Porifera Research Biodiversity, Innovation and Sustainability Custódio MR, Lôbo-Hajdu G, Hajdu E, Muricy G, eds, Série Livros. 28. Museu Nacional, Rio de Janeiro*:561-568.

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

Vacelet, J. 1975. Etude en microscopie electronique de l'association entre bacteries et spongiaires du genre Verongida (Dicytoceratida). *Journal de Microscopie et de Biologie Cellulaire* **23**:271-288.

Vacelet, J., and C. Donadey. 1977. Electron microscope study of the association between some sponges and bacteria *J. Exp. Mar. Biol. Ecol.* **30**:301-314.

Webster, N. S., M. W. Taylor, F. Behnam, S. Lucker, T. Rattei, S. Whalan, M. Horn, and M. Wagner. 2010. Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ. Microbiol.* **12**:2070-2082.

Webster, N. S., R. I. Webb, M. J. Ridd, R. T. Hill, and A. P. Negri. 2001. The effects of copper on the microbial community of a coral reef sponge. *Environ. Microbiol.* **3**:19-31.

Wehrl, M. 2006. Bakterielle Aufnahme, Selektivität und interne Prozessierung bei marinen Schwämmen (Porifera). PhD thesis, University of Wuerzburg, Wuerzburg, Germany.

Weisz, J. B., U. Hentschel, N. Lindquist, and C. S. Martens. 2007. Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Mar. Biol.* **152**:475-483.

Weisz, J. B., N. Lindquist, and C. S. Martens. 2008. Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia* **155**:367-376.

Wilkinson, C. R. 1978. Microbial associations in sponges. III. Ultrastructure of the in situ associations in coral reef sponges. *Mar. Biol.* **49**:177-185.

4.2 Bacterial community profiles in low microbial abundance sponges

This article was published in **FEMS Microbiology Ecology** and a reprint is used here with the permission from John Wiley and Sons and the authors. This article's reference is:

Giles, E.C., Kamke, J., **Moitinho-Silva, L.**, Taylor, M.W., Hentschel, U., Ravasi, T., Schmitt, S. (2013) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*. 83(1), 232–41.

The documentation of author's contributions is found in the Appendix. The supplementary material is found in the CD attached to this PhD Thesis.

FEMS MICROBIOLOGY ECOLOGY

RESEARCH ARTICLE

Bacterial community profiles in low microbial abundance sponges

Emily C. Giles[1], Janine Kamke[2], Lucas Moitinho-Silva[2], Michael W. Taylor[3], Ute Hentschel[2], Timothy Ravasi[1] & Susanne Schmitt[4]

[1]Division of Chemical & Life Sciences and Engineering and Division of Applied Mathematics and Computer Science, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia; [2]Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany; [3]Centre for Microbial Innovation, School of Biological Sciences, The University of Auckland, Auckland, New Zealand; and [4]Department of Earth and Environmental Sciences, Paleontology and Geobiology, Ludwig-Maximilians Universität in Munich, Munich, Germany

Correspondence: Susanne Schmitt, Department of Earth and Environmental Sciences, Paleontology and Geobiology, Ludwig-Maximilians Universität in Munich, Richard-Wagner-Str. 10, 80333 Munich, Germany. Tel.: +49 089 21806706; fax: +49 089 21806601; e-mail: s.schmitt@lrz.uni-muenchen.de

Received 30 May 2012; revised 24 July 2012; accepted 1 August 2012.

DOI: 10.1111/j.1574-6941.2012.01467.x

Editor: Julian Marchesi

Keywords
16S rRNA gene; bacteria; marine sponges; microbial abundance; phylogenetic diversity.

Abstract

It has long been recognized that sponges differ in the abundance of associated microorganisms, and they are therefore termed either 'low microbial abundance' (LMA) or 'high microbial abundance' (HMA) sponges. Many previous studies concentrated on the dense microbial communities in HMA sponges, whereas little is known about microorganisms in LMA sponges. Here, two LMA sponges from the Red Sea, two from the Caribbean and one from the South Pacific were investigated. With up to only five bacterial phyla per sponge, all LMA sponges showed lower phylum-level diversity than typical HMA sponges. Interestingly, each LMA sponge was dominated by a large clade within either *Cyanobacteria* or different classes of *Proteobacteria*. The overall similarity of bacterial communities among LMA sponges determined by operational taxonomic unit and UniFrac analysis was low. Also the number of sponge-specific clusters, which indicate bacteria specifically associated with sponges and which are numerous in HMA sponges, was low. A biogeographical or host-dependent distribution pattern was not observed. In conclusion, bacterial community profiles of LMA sponges are clearly different from profiles of HMA sponges and, remarkably, each LMA sponge seems to harbour its own unique bacterial community.

Introduction

The association of sponges and microbial communities has long been recognized, and extensive efforts have been made to identify these microorganisms and to characterize the association (Taylor *et al.*, 2007; Webster & Taylor, 2012). Overall, more than 7500 sponge-derived 16S rRNA gene sequences are now available in public databases and these sequences are affiliated with 17 formally described phyla and several more candidate phyla (Simister *et al.*, 2012). Sequences were obtained from many studies using clone library (Kennedy *et al.*, 2008; Turque *et al.*, 2008), isolation and cultivation (Zhang *et al.*, 2008; Sipkema *et al.*, 2011), and denaturing gradient gel electrophoresis (DGGE; Mohamed *et al.*, 2008; Anderson *et al.*, 2010) techniques. The application of next-generation sequencing technologies to sponges revealed the presence of even more formally described phyla and candidate divisions (Webster *et al.*, 2010; Lee *et al.*, 2011; Jackson *et al.*, 2012; Schmitt *et al.*, 2012b). Phyla found in many different sponges and represented by many different sequences include *Actinobacteria*, *Acidobacteria*, *Cyanobacteria*, *Chloroflexi*, *Proteobacteria* and the candidate phylum *Poribacteria*.

However, microbial communities can vary considerably among different sponges with respect to both microbial abundance and diversity. Microbial abundance is often estimated by visual inspection of the sponge mesohyl for bacterial cells using transmission electron microscopy (TEM). In an early study, Levi & Porte (1962) were the first to identify granular particles present in the sponge mesohyl as microbial cells. Subsequent microscopy studies systematically investigated sponges for the presence of bacteria, and it was soon realized that differences in

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

microbial abundance exist among sponges (Reiswig, 1974; Vacelet & Donadey, 1977; Wilkinson, 1978). Sponges associated with many microorganisms were termed 'bacterial sponges' (vs. 'corneal sponges'; Bertrand & Vacelet, 1971) or 'bacteriosponges' (vs. 'nonsymbiont-harbouring 'normal' sponges'; Reiswig, 1981) and later 'high microbial abundance (HMA) sponges' (vs. 'low microbial abundance (LMA) sponges'; Hentschel *et al.*, 2003). In one of these early studies, the microbial concentration of the sponge *Aplysina fistularis* was found to be approximately 8×10^{10} cells per mL of tissue, comprising 20% of total specimen volume (Reiswig, 1981). This extraordinarily high number of microorganisms was later confirmed for other HMA sponges (in a range of 10^8–10^{10} bacteria per g or mL of tissue) and was contrasted with a microbial concentration of 10^5–10^6 bacteria per g or mL in LMA sponges (Hentschel *et al.*, 2006). The latter concentration is typical for natural seawater (Hentschel *et al.*, 2006).

LMA and HMA sponges not only differ in microbial abundances but also in microbial diversity. Weisz *et al.* (2007a) used TEM to identify two HMA and one LMA sponge from the Caribbean. Subsequent DGGE analysis revealed multiple bands in HMA sponges, and many of the respective 16S rRNA gene sequences were closely related to previously sponge-derived sequences. In contrast, DGGE from the LMA sponge resulted in few bands and the 16S rRNA gene sequence of the most dominant band showed highest affinity to seawater bacteria (Weisz *et al.*, 2007a). A clear difference in bacterial diversity was also found in a clone library study using the HMA sponge *Ancorina alata* and the LMA sponge *Polymastia* sp. (Kamke *et al.*, 2010). Analysis of a similar number of 16S rRNA gene sequences from both sponges revealed the presence of eight phyla in the HMA sponge but only two phyla in the LMA sponge. This trend was confirmed in a subsequent 454 pyrosequencing study where members of 15 phyla were detected in *A. alata* but members of only eight phyla in *Polymastia* sp. (Schmitt *et al.*, 2012b). The investigation of a single bacterial phylum in three HMA and three LMA sponges resulted in a similar pattern: *Chloroflexi* bacteria are more abundant and diverse in HMA than LMA sponges (Schmitt *et al.*, 2012a).

Further differences between LMA and HMA sponges that often co-occur in the same habitat have been described. For example, while the reproductive stages of HMA sponges generally contained large amounts of microorganisms, as evidenced by TEM, the reproductive stages of LMA sponges generally appeared bacteria-free (Schmitt *et al.*, 2007). Moreover, the sponge metabolic status seems to be influenced by microorganisms only in HMA and not in LMA sponges (Weisz *et al.*, 2007a; Ribes *et al.*, 2012). Microbial abundances might also have an influence on sponge morphology and physiology. A variety of HMA sponges had a denser mesohyl and a more complex aquiferous system with much slower pumping rates than LMA sponges (Weisz *et al.*, 2007b). Finally, LMA and HMA sponges also show differences in their chemistry. A recent study by Hochmuth *et al.* (2010) revealed that *supA* type polyketide synthase (PKS) genes were present in all six HMA sponges but were absent from six LMA sponges. HMA sponges also showed characteristic fatty acid profiles not found in LMA sponges (Hochmuth *et al.*, 2010).

Despite these previous results, there are still many unanswered questions regarding this sharp dichotomy between LMA and HMA sponges. This is also owing to the fact that many studies concentrated either on HMA sponges or did not distinguish between HMA and LMA sponges. In this study, we investigated bacterial communities in the tropical LMA sponges *Crella cyathophora* and *Stylissa carteri* from the Red Sea, in the tropical Caribbean sponges *Callyspongia vaginalis* and *Niphates digitalis*, and in the temperate sponge *Raspailia topsenti* from the South Pacific (New Zealand). We provide detailed data on the diversity and compare the bacterial community among LMA sponges and between bacterial communities in HMA sponges to elucidate the similarity of LMA sponge–microorganism associations.

Materials and methods

Sample collection

The sponges *C. cyathophora* and *S. carteri* were sampled in November 2010 by SCUBA at Fsar Reef in the Red Sea (22°23′N; 39°03′E) at a depth between 13 and 15 m. The sponges *C. vaginalis* and *N. digitalis* were collected in July 2003 off Little San Salvador Island, Caribbean Sea, Bahamas (24°34′N; 75°58′W), at a depth of 12 m. *Raspailia topsenti* was collected in February 2008 at Matheson's Bay on New Zealand's North Island (174°47′S, 36°18′E) at a depth of 3 m. After collection, sponge tissue was immediately washed with calcium–magnesium-free artificial seawater (CMF-ASW) 1–3 times and stored at − 80 °C until further use. Additionally, small sponge tissue pieces of *c.* 0.5 cm^3 from the Red Sea sponges were preserved in 2.5% glutaraldehyde-H_2O_{dd} for further processing by TEM.

Transmission electron microscopy

Fixed samples of Red Sea sponges were washed five times in cacodylate buffer (50 mM, pH 7.2), fixed in 2% osmium tetroxide for 90 min, washed again five times in

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

H_2O_{dd} and incubated overnight in 0.5% uranyl acetate. Following dehydration in an ethanol series (30%, 50%, 70%, 90%, 96% and three times in 100% for 30 min each), the samples were incubated three times for 30 min in 1 × propylene oxide, maintained overnight in 1 : 1 (vol/vol) propylene oxide-Epon 812 (Serva), incubated twice for 2 h in Epon 812 and finally embedded in Epon 812 for 48 h at 60 °C. Samples were then sectioned with an ultramicrotome (OM U3; C. Reichert, Austria) and examined by TEM (Zeiss EM 10; Zeiss, Germany).

DNA extraction and construction of 16S rRNA gene libraries

For the Caribbean and Red Sea sponges (one individual each), an initial cell lysis step was performed with 200–250 mg of tissue (frozen wet weight) from each sponge in 600 µL of 1% 2-mercaptoethanol RLT buffer (Qiagen, Germany) using a FastPrep® Instrument (MP Biomedicals). DNA was then extracted using the Allprep DNA/RNA mini Kit (Qiagen). DNA of *R. topsenti* was extracted from 5 to 6 mg of freeze-dried and ground tissue (separately from three individuals; PCR product was pooled afterwards) using a CTAB-based protocol as described previously (Taylor *et al.*, 2004). Briefly, cells were disrupted by bead-beating in an ammonium acetate extraction buffer containing chloroform/isoamyl alcohol (24 : 1). DNA was precipitated with 3 M sodium acetate and isopropanol, then washed in 70% ethanol, dried and redissolved in H_2O_{dd}. The quality and quantity of extracted DNA was measured using a NanoDrop 8000 spectrophotometer (Thermo Scientific).

A *c.* 1500-bp fragment of the 16S rRNA gene was amplified from all sponge samples using the primer pairs 27f/1492r (Lane, 1991) for the Caribbean and Red Sea sponges and 616V/1492r (Lane, 1991; Juretschko *et al.*, 1998) for *R. topsenti*. Cycling conditions were as follows: initial denaturing step at 96 °C for 4 min, followed by 35 cycles of denaturing at 96 °C for 60 s, primer annealing at 56 °C for 60 s and elongation at 72 °C for 90 s and a final extension step at 72 °C for 5 min. PCR products were cleaned using the QIAquick PCR Purification kit (Qiagen) and cloned either into the pCR® II-TOPO® vector (Invitrogen) or into the pGEM-T-easy vector (Promega). Clone libraries were constructed according to the manufacturer's instructions. Correct-sized inserts were identified by PCR using the vector-specific primer pairs M13f (5′ GTAAAACGACGGCCAG 3′) and M13r (5′ CAGGAAACAGCTATGAC 3′) or SP6 (5′ TAT TTA GGT GAC ACT ATA G 3′) and T7 (5′ TAA TAC GAC TCA CTA TAG GG 3′). PCR products were sequenced by the KAUST Genomics Core Lab on a 3730XL 96 Capillary Sequencing machine (Applied Biosystems) and at Macrogen (South Korea) with the PCR primers used for clone library construction. Poor-quality sequences as well as chimeric sequences detected with Pintail (Ashelford *et al.*, 2005) were removed from the data set. Final sequence data were submitted to the DDBJ/EMBL/GenBank databases under accession numbers JQ062599–JQ062861 and JN850797–JN850860.

Sequence analysis and phylogenetic tree construction

Operational taxonomic units (OTUs) and Chao1 values were calculated at a 98% sequence similarity threshold using Mothur v.1.22.2 (Schloss *et al.*, 2009). For comparison, 85 16S rRNA gene sequences obtained from the New Zealand LMA sponge *Polymastia* sp. (Kamke *et al.*, 2010) were also analysed. To determine the similarity of microbial communities in different environments, hierarchical cluster analysis was performed using an unweighted UniFrac algorithm (Lozupone *et al.*, 2006). 16S rRNA gene sequences from this study, as well as their closest relatives identified by initial BLAST searches (http://blast.ncbi.nlm.nih.gov/Blast.cgi; Altschul *et al.*, 1990), were automatically aligned to a SILVA reference alignment using the SINA Webaligner and merged into the SILVA version 108 database (Ludwig *et al.*, 2004; Pruesse *et al.*, 2007). The alignment was then manually refined using the editor tool in the ARB software package. The alignment is available upon request. Distance (neighbour-joining), maximum parsimony and maximum likelihood trees were calculated in ARB using long ($\geq$ 1200 bp) sequences only. Short sequences were added using the parsimony interactive tool in ARB without changing the tree topology. Phylogenetic consensus trees, using the maximum likelihood tree as a backbone, were manually constructed (Ludwig *et al.*, 1998). Maximum parsimony bootstraps (100 resamplings) were performed to further assess the stability of observed branching patterns.

Results

Sponge microbial abundance classification using TEM

To estimate the abundance of microorganisms in the mesohyl and to group the sponges into either LMA or HMA sponge categories, the mesohyl of three individuals of the Red Sea sponges *C. cyathophora* and *S. carteri* was inspected by TEM (5 fields and at least 2 sections per sponge). All investigated tissue sections were free of microbial cells (Fig. 1). Therefore, *C. cyathophora* and *S. carteri* are classified as LMA sponges. The Caribbean

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

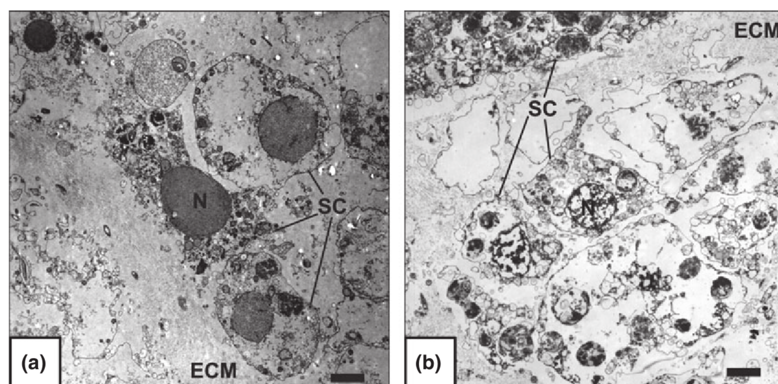


Fig. 1. TEM of the sponge mesohyl of *Crella cyathophora* (a) and *Stylissa carteri* (b). The mesohyl is free of bacterial cells. ECM, extracellular matrix; N, sponge cell nucleus; SC, sponge cell. Scale bar indicates 2 μm.

sponges *C. vaginalis* and *N. digitalis* as well as the New Zealand sponge *R. topsenti* were previously identified as LMA sponges because of the absence of large amounts of bacteria in the mesohyl (Schmitt *et al.*, 2007, 2012a).

Phylum level diversity

A total of 67, 67, 62, 67 and 9 nearly full-length 16S rRNA gene sequences were obtained from the sponges *C. cyathophora*, *S. carteri*, *C. vaginalis*, *N. digitalis* and *R. topsenti*, respectively. In addition, 55 partial 16S rRNA gene sequences were also obtained from *R. topsenti*. Sequences were affiliated with seven different bacterial phyla in total among all sponges, with three to five different phyla within a single sponge (Fig. 2). The phyla *Actinobacteria*, *Nitrospira*, *Planctomycetes* and *Spirochaetes* were represented by only very few sequences (up to four sequences) and were only discovered in a single sponge species each. *Bacteroidetes* were represented by only up to five sequences per sponge but were found in all sponges except *R. topsenti*. Sequences classified as *Cyanobacteria* and *Proteobacteria* were found in all five sponges. Of the detected four classes of *Proteobacteria*, *Alpha-* and *Gammaproteobacteria* were found in four sponges, whereas *Delta-* and *Betaproteobacteria* were found in three and two sponges, respectively. It is apparent that each bacterial community is dominated by a single phylum or proteobacterial class: *Cyanobacteria* in *C. vaginalis*, *Alphaproteobacteria* in *C. cyathophora* and *N. digitalis*, *Betaproteobacteria* in *R. topsenti* and *Gammaproteobacteria* in *S. carteri* (Fig. 2).

Analysis of OTUs within each sponge

Sequences within each sponge were grouped into OTUs with a 98% similarity threshold and 26 (Chao1 value: 78,5), 46 (157), 19 (124), 12 (24) and 18 (63,5) OTUs were obtained for the sponges *C. cyathophora*, *S. carteri*, *C. vaginalis*, *N. digitalis* and *R. topsenti*, respectively

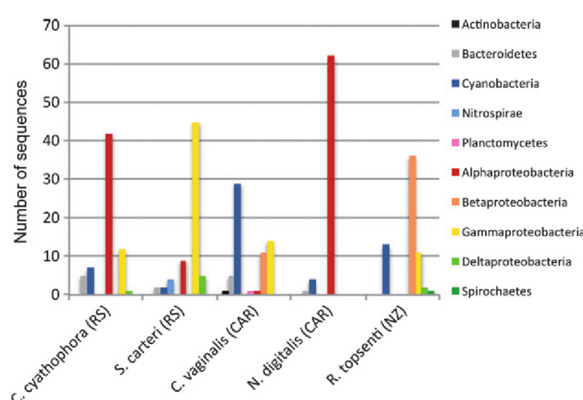


Fig. 2. Bacterial diversity on phylum level (and class level for *Proteobacteria*) within each of the five sponges. The number of sequences within each phylum or proteobacterial class is given. RS, Red Sea; CAR, Caribbean Sea; NZ, New Zealand.

(Fig. 3). All sponges contained mainly singleton OTUs that consisted of only one sequence (Fig. 3). Singleton OTUs comprised 75–81% of all OTUs. Singleton, doubleton (containing two sequences) and tripleton (containing three sequences) OTUs together comprised as much as 96% of all OTUs. Each sponge was dominated by one very large OTU. These biggest OTUs ranged from an OTU containing 10 sequences (15% of all sequences in *S. carteri*) to an OTU containing 54 sequences (81% of all sequences in *N. digitalis*). With the exception of the biggest OTU in *C. vaginalis,* which was affiliated with the *Cyanobacteria*, the biggest OTUs in all other sponges were affiliated with different classes of *Proteobacteria* (Fig. 3). For comparison, 85 16S rRNA gene sequences obtained from the New Zealand LMA sponge *Polymastia* sp. (Kamke *et al.*, 2010) were also analysed. Ninety-three percent of OTUs in *Polymastia* sp. were singletons and 96% were singletons and doubletons. The biggest OTU in *Polymastia* sp. contained 24 sequences (28% of all sequences in *Polymastia* sp.) and was affiliated with *Alphaproteobacteria*.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

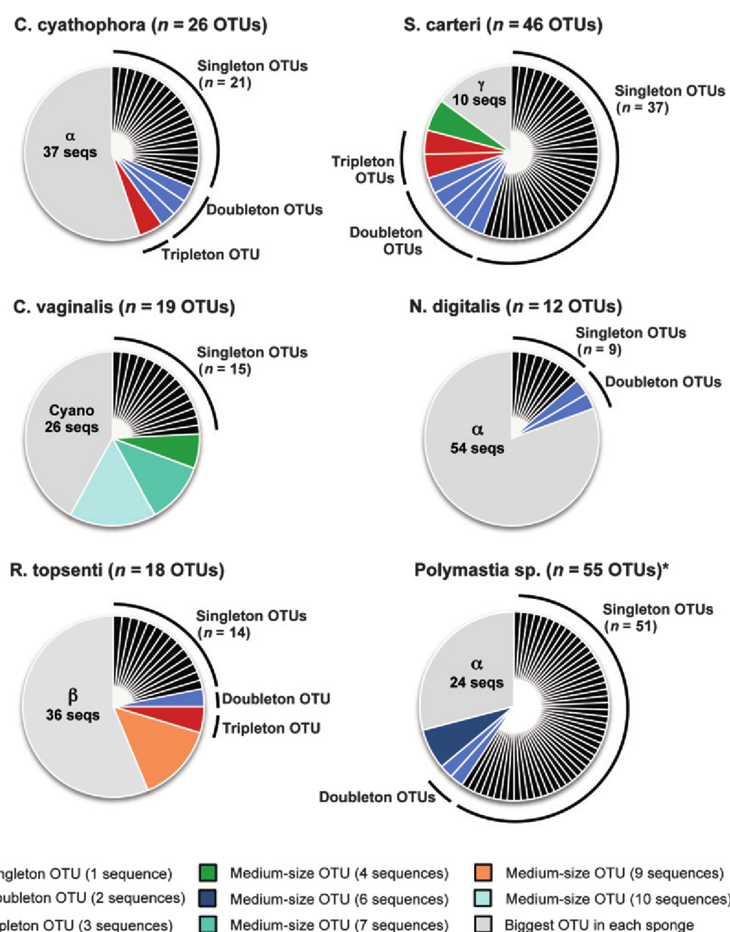


Fig. 3. 98% OTU analysis. The pie charts illustrate the number of OTUs per sponge, and the size of the respective sectors represents the number of sequences within each OTU. Singleton (consisting of one sequence), doubleton (two sequences) and tripleton (three sequences) OTUs are indicated. The biggest OTU (grey colour) in each sponge is highlighted and its size in terms of number of sequences and phylogenetic affiliation are given (α, *Alphaproteobacteria*; β, *Betaproteobacteria*; γ, *Gammaproteobacteria*, Cyano, *Cyanobacteria*). *For comparison, published sequence data from the New Zealand LMA sponge *Polymastia* sp. (N = 85) were also analysed (Kamke *et al.*, 2010).

Phylogenetic tree analysis

All 327 newly obtained 16S rRNA gene sequences were used together with appropriate reference sequences from sponges and other environments to construct phylogenetic trees (Fig. 4, Supporting Information, Fig. S1). A large clade containing 36 sequences from *R. topsenti* within the *Betaproteobacteria* corresponded to the largest OTU in this species (Fig. 3). Large clades were also found for *C. cyathophora*, *N. digitalis* and *S. carteri* within the *Alpha-* and *Gammaproteobacteria* (Fig. 4). Each of these clades contained several of the detected OTUs of the respective sponges. Whereas the clades of *N. digitalis* and *C. cyathophora* had no close similarity to other sponge-derived sequences, the clade of *S. carteri* showed some affinity to sequences derived from the sponge *Axinella corrugata* and the clade of *R. topsenti* fell within a sponge-specific cluster (SSC) (a monophyletic group of only sponge-derived sequences, see Hentschel *et al.*, 2002 for definition). Sequences from *C. vaginalis* were clustered in two smaller clades (11 and 10 sequences) within the *Beta-* and *Gammaproteobacteria*. The betaproteobacterial clade also formed a SSC together with other sponge-derived sequences (Fig. 4). The biggest OTU in *C. vaginalis* was affiliated with the *Cyanobacteria*. Phylogenetic tree analysis showed that this OTU fell within the *Synechococcus* lineage together with sequences from other sponges and environmental sequences (Fig. S1). In addition to the two already mentioned SSCs, two sponge-coral clusters (SCC) (SSCs that in addition contain sequences from corals, see Taylor *et al.*, 2007 for definition) were detected that were affiliated with *Gammaproteobacteria* and *Nitrospira* (Figs 4 and S1).

Similarity of bacterial communities among LMA sponges

When calculating 98% OTUs using 16S rRNA gene sequences from all sponges, 118 OTUs were obtained of which only two OTUs contained sequences from more than one sponge (Fig. 5a). The first comprised one sequence derived from *C. cyathophora* and a second sequence from *S. carteri*. This OTU was affiliated with the *Bacteroidetes*. The second OTU was bigger, with 28

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

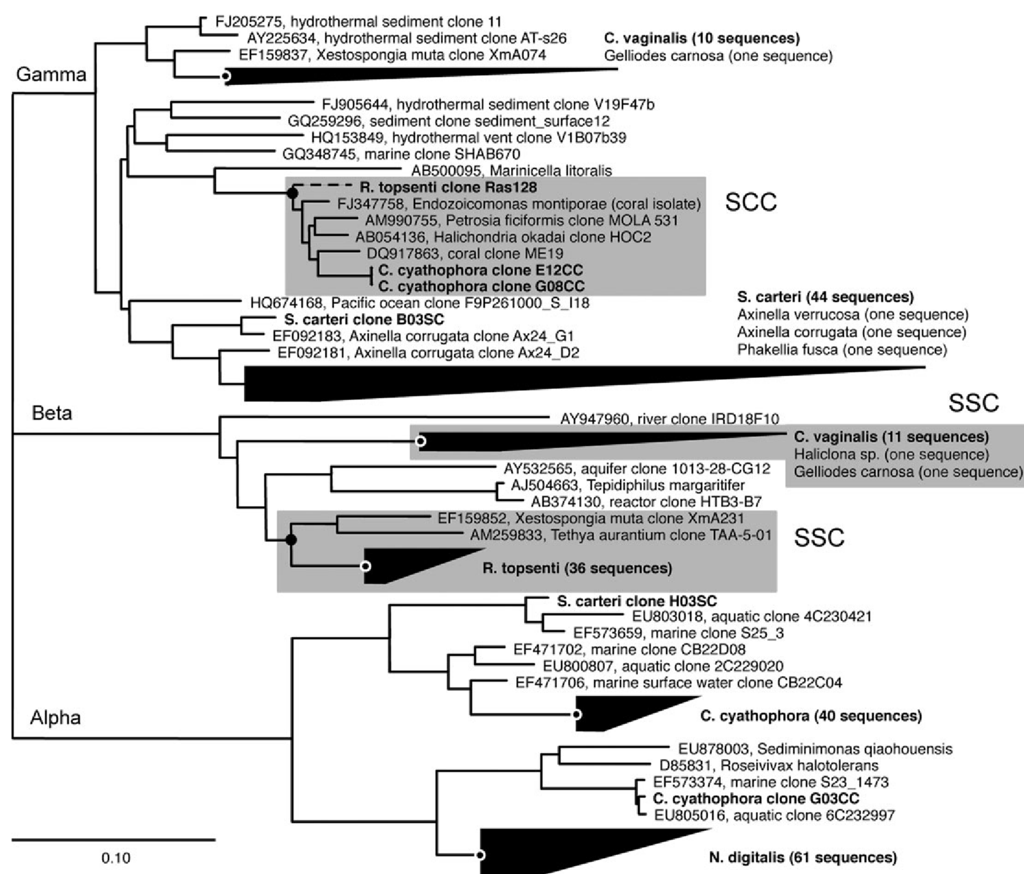


Fig. 4. 16S rRNA gene sequence-based maximum likelihood tree of *Proteobacteria*. Sponge-derived 16S rRNA gene sequences of this study are given in bold. The sponge names outside of wedges indicate species from which the respective sequences were derived. Numbers in parentheses indicate the number of sequences per sponge. Shaded boxes represent SSC and are either labelled as SSC or SCC depending on whether the respective cluster also contains a coral-derived sequence. Dashed line indicates a short sequence (< 1200 bp). Filled circles show bootstrap support of 100% for relevant groups. Scale bar indicates 10% sequence divergence.

sequences from both Red Sea sponges and both Caribbean species (Fig. 5a). It was affiliated with the *Cyanobacteria*.

Total bacterial communities were compared among all LMA sponges of this study using UniFrac analysis (Fig. 5b). For comparison, 85 sequences obtained from the New Zealand LMA sponge *Polymastia* sp. (Kamke *et al.*, 2010) were also included in the analysis. Most similar were the bacterial communities in the Caribbean sponge *N. digitalis* and the Red Sea sponge *C. cyathophora,* whereas the most distant communities were found in both New Zealand species. No obvious correlation with the sampling sites or with host phylogeny was detected (Fig. 5b).

Discussion

Microscopic techniques have long been used to estimate the abundance of bacteria in various sponges (Vacelet & Donadey, 1977; Wilkinson, 1978; Reiswig, 1981). Visual inspection of ultrathin sections of the Red Sea species *C. cyathophora* and *S. carteri* revealed the absence of bacteria in the mesohyl of both sponges (Fig. 1). Similarly, no bacterial cells were previously detected in the Caribbean sponges *C. vaginalis* and *N. digitalis* (Schmitt *et al.*, 2007). *Raspailia topsenti* was identified as LMA sponge because of the absence of large amounts of bacteria in the mesohyl (Schmitt *et al.*, 2012a). However, few bacteria-like structures were found in TEM micrographs, and similar structures were also occasionally detected in other LMA sponges such as *Mycale laxissima* (Schmitt *et al.*, 2007), *Niphates erecta* (Weisz *et al.*, 2007a), *Polymastia janeirensis* and *Hymeniacidon heliophila* (Turque *et al.*, 2008). It is possible that bacteria are generally missed in TEM surveys of LMA sponges because of their low abundances and/or because they only occur at certain sites within the sponge and are not equally distributed throughout the mesohyl as seen in HMA sponges (for

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

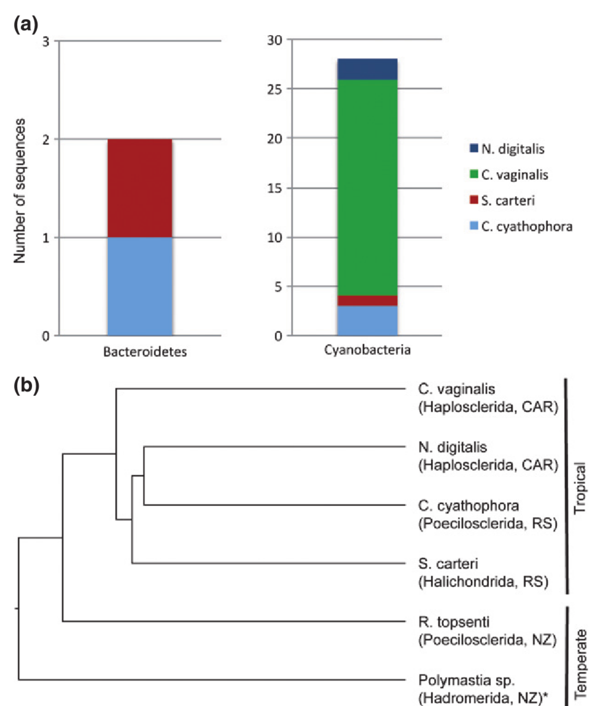


Fig. 5. Bacterial community similarity among LMA sponges. (a), 98% OTU analysis of all new sequences. The two 98% OTUs that contain sequences from different sponges are shown. (b), UniFrac cluster analysis of total bacterial communities. The taxonomic order and sample location of each sponge is given in parenthesis. *For comparison, published sequence data from the New Zealand LMA sponge *Polymastia* sp. ($N = 85$) were also analysed (Kamke *et al.*, 2010). RS, Red Sea; CAR, Caribbean Sea; NZ, New Zealand.

TEM of HMA sponges see e.g. Friedrich *et al.*, 1999; Thoms *et al.*, 2003). Generally, the detected microorganisms in LMA sponges are extracellular and show no signs of digestion, but it remains unclear whether they represent food bacteria, seawater contaminants or long-term sponge associates. However, the different numbers of bacterial cells in LMA sponges relative to the dense microbial communities in HMA sponges are obvious, even to the inexperienced eye.

As expected based on previous reports (Weisz *et al.*, 2007a; Kamke *et al.*, 2010), the diversity of bacterial communities in LMA sponges was lower in comparison with the diversity in HMA sponges. On phylum level, two phyla were found in the LMA sponge *Polymastia* sp., whereas eight phyla were detected in the co-occurring HMA sponge *A. alata* (Kamke *et al.*, 2010). In this study, analysis of a similar number of sequences revealed slightly more (3–5) bacterial phyla in the LMA sponges, but many of these phyla are restricted to a single sponge and are represented by only very few sequences (Fig. 2). Only the phylum *Proteobacteria* was represented by many sequences in all investigated sponges and it was also the dominant phylum in other LMA sponges, for example, *Polymastia* sp. (Kamke *et al.*, 2010) and *H. heliophila* (Erwin *et al.*, 2011). Other typical phyla often found in sponges (Simister *et al.*, 2012) are missing in this study, for example, *Acidobacteria*, *Chloroflexi* and *Gemmatimonadetes*. Representatives of these phyla might therefore be indicative of HMA sponges. Overall, the bacterial community pattern in LMA sponges is characterized by a low phylum-level diversity with usually *Proteobacteria* as the most prominent phylum and the absence of other phyla typical of HMA sponges.

Notable is a single large OTU in the bacterial community of each sponge (Fig. 3) and correspondingly a large clade in the phylogenetic tree (Figs 4 and S1). The biggest OTU in *C. vaginalis* is affiliated with the cyanobacterial *Synechococcus* lineage that was also found in the other Red Sea and Caribbean sponges of this study (Figs 5a and S1). *Synechococcus*-like sequences were recovered from multiple sponges and many of these sequences comprise a large SSC indicating a widespread sponge associate tentatively termed '*Synechococcus spongiarum*' (Usher *et al.*, 2004; Steindler *et al.*, 2005; Erwin & Thacker, 2008). However, sequences obtained in this study did not fall into this sponge-specific clade but instead cluster with environmental sequences (Fig. S1). *Synechococcus* is a very abundant member of the bacterioplankton (Scanlan & West, 2002), and we might have detected a marine free-living form in the LMA sponges.

The largest OTU found in each of the other LMA sponges is affiliated with the *Proteobacteria*. Interestingly, each sponge harbours a different clade within the *Alpha*-, *Beta*- or *Gammaproteobacteria* (Fig. 4). Similarly, dominant proteobacterial phylotypes were also detected in clone library data of other LMA sponges such as *Haliclona* sp. (Sipkema *et al.*, 2009), *H. heliophila* (Erwin *et al.*, 2011), *Ianthella basta* (Luter *et al.*, 2010) and *Polymastia* sp. (Kamke *et al.*, 2010). The dominance of a proteobacterial group seems to be another characteristic of bacterial communities in LMA sponges; however, the underlying reasons remain subject to speculation. It was shown that vertical transmission, as a mechanism to obtain bacteria, seems to occur mainly in HMA sponges (Schmitt *et al.*, 2007). This indicates that LMA sponges may acquire their bacteria mainly from the environment (e.g. seawater, sediment). If these bacteria are taken up by a selective mechanism, then all individuals of a species should be dominated by the same proteobacterium. This seems to be the case in a DGGE analysis of the LMA sponge *N. digitalis* where three individuals showed identical DGGE banding patterns and the respective sequences showed the same closest match in a BLAST search (Weisz *et al.*, 2007a). Alternatively, it is also possible that the

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

largest OTU represents the bacterium that first populated the growing animal. In this case, inoculation occurs by chance and different bacteria may dominate different sponge individuals of the same species. Here, we cannot test this hypothesis because no distinction between individuals was made but this will be an interesting aspect for future research.

In this study, several approaches were used to determine the similarity of bacterial communities among LMA sponges. Both, OTU (Fig. 5a) and UniFrac (Fig. 5b) analyses indicate a small overlap. From 118 OTUs (obtained by analysis of all new sequences in this study), only two contain sequences from more than one sponge (Fig. 5a). The larger OTU within the *Cyanobacteria* is affiliated with the *Synechococcus* lineage (Fig. S1) and might represent a marine free-living form (see discussion above). The second OTU contains only two sequences in total from two sponges and is affiliated with the *Bacteroidetes* (Fig. 5a). These two sequences are very similar to other marine, aquatic and planktonic sequences (Fig. S1) and might also represent a potentially abundant and widespread marine bacterium.

UniFrac analysis was used to compare the entire microbial communities among LMA sponges. Neither a biogeographical pattern according to host sponge distribution (e.g. sponges collected from the same location harbour more similar bacterial communities) nor a distribution according to host phylogeny (e.g. closely related sponges harbour more similar bacterial communities) was observed (Fig. 5b). However, it is interesting that the bacterial communities in the two New Zealand species were most distant. These sponges are the only temperate specimens collected, whereas all other sponges are tropical animals. A similar trend was observed in a recent 454 pyrosequencing study using 32 sponges from eight tropical and temperate locations with most similar bacterial communities (on phylum level) in tropical sponges (Schmitt *et al.*, 2012b). This suggests that environmental factors such as temperature, salinity or nutrient levels might indeed influence the composition of microbial communities in sponges. A combination of environmental factors (shaded vs. nonshaded habitats) and host-related factors (vertical transmission of symbionts) were suggested to structure the complex microbial communities in three HMA *Ircinia* sponges (Erwin *et al.*, 2012). Different environmental and/or host-related factors might also be responsible for the different bacterial community patterns in LMA and HMA sponges.

Another phenomenon found in sponge microbiology is SSCs, that is, monophyletic clusters of only sponge-derived sequences (Hentschel *et al.*, 2002; Taylor *et al.*, 2007; Simister *et al.*, 2012). The number of SSC and SCC detected in this study was relatively low (N = 4; Figs 4 and S1). This is in agreement with previous observations that HMA sponges usually contain many of these clusters but only few are found in LMA sponges (Kamke *et al.*, 2010; Erwin *et al.*, 2011; Schmitt *et al.*, 2012a). It was recently speculated that there might be LMA SSCs (e.g. SSCs that only contain LMA sponge-derived sequences; Schmitt *et al.*, 2012a). This could not be confirmed in this study where three of the four detected SSCs probably contain sequences from both LMA and HMA sponges (Figs 4 and S1). The fact that two SSCs also contain sequences from corals might indicate a wider host-associated distribution of these bacteria than only in sponges. However, this is not a phenomenon typical for LMA sponges but is also well known from HMA sponges although the ecological implications are unclear (Simister *et al.*, 2012). A similarity to other invertebrate-associated bacteria, such as symbionts of ascidians, was not found. Therefore, each LMA sponge seems to harbour a unique bacterial community that is different from other LMA as well as from HMA sponges and also from other marine invertebrates.

In summary, our investigation of bacterial communities in LMA sponges revealed a pattern that is clearly different from bacterial associations with HMA sponges. A low abundance of microorganisms corresponded with a low phylum-level diversity. The community composition of each sponge was dominated by a single clade within either *Cyanobacteria* or different classes of *Proteobacteria*. Typical bacterial phyla of HMA sponges are missing in LMA sponges. Bacterial community similarity between LMA and HMA sponges was very low and might indicate that different environmental- or host-related factors shape bacterial communities in LMA and HMA sponges.

Acknowledgements

This study was supported by DFG grant HE3299/1-3 to U.H. M.W.T. would like to acknowledge funding support from the University of Auckland Faculty Research Development Fund (grant 3609286). E.C.G. and T.R. were supported by King Abdullah University of Science and Technology. L.M.S. was supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Wuerzburg.

References

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

Anderson SA, Northcote PT & Page MJ (2010) Spatial and temporal variability of the bacterial community in different chemotypes of the New Zealand marine sponge *Mycale hentscheli*. *FEMS Microbiol Ecol* **72**: 328–342.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ & Weightman AJ (2005) At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.

Bertrand JC & Vacelet J (1971) L'association entre eponges cornees et bacteries. *CR Hebd Seanc Acad Sci Paris D* **273**: 638–641.

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

Erwin PM, Olson JB & Thacker RW (2011) Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the Northern Gulf of Mexico. *PLoS ONE* **6**: e26806.

Erwin PM, Lopez-Legentil S, Gonzalez-Pech R & Turon X (2012) A specific mix of generalists: bacterial symbionts in Mediterranean *Ircinia* spp. *FEMS Microbiol Ecol* **79**: 619–637.

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). *Mar Biol* **134**: 461–470.

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

Hentschel U, Fieseler L, Wehrl M, Gernert C, Steinert M, Hacker J & Horn M (2003) Microbial diversity of marine sponges. *Marine Molecular Biotechnology* (Mueller W, ed), pp. 59–88. Springer, Berlin.

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

Hochmuth T, Niederkruger H, Gernert C, Siegl A, Taudien S, Platzer M, Crews P, Hentschel U & Piel J (2010) Linking chemical and microbial diversity in marine sponges: possible role for Poribacteria as producers of methyl-branched fatty acids. *ChemBioChem* **11**: 2572–2578.

Jackson SA, Kennedy J, Morrissey JP, O'Gara F & Dobson ADW (2012) Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters. *Microb Ecol* **64**: 105–116.

Juretschko S, Timmermann G, Schmid M, Schleifer KH, Pommerening-Roser A, Koops HP & Wagner M (1998) Combined molecular and conventional analysis of nitrifying bacterium diversity in activated sludge: *Nitrosococcus mobilis* and Nitrospira-like bacteria as dominant populations. *Appl Environ Microbiol* **64**: 3042–3051.

Kamke J, Taylor MW & Schmitt S (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA *vs* 16S rRNA gene comparisons. *ISME J* **4**: 498–508.

Kennedy J, Codling CE, Jones BV, Dobson ADW & Marchesi JR (2008) Diversity of microbes associated with the marine sponge, *Haliclona simulans,* isolated from Irish waters and identification of polyketide synthase genes from the sponge metagenome. *Environ Microbiol* **10**: 1888–1902.

Lane DJ (1991) 16S/23S rRNA sequencing. *Nucleic Acid Techniques in Bacterial Systematics* (Stackebrandt E & Goodfellow M, eds), pp. 115–175. John Wiley and Sons, New York, NY.

Lee OO, Wang Y, Yang JK, 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.

Levi C & Porte A (1962) Electron microscopy study of the sponge *Oscarella lobularis* and its amphiblastula larvae. *Cah Biol Mar* **3**: 307–315.

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

Ludwig W, Strunk O, Klugbauer S, Klugbauer N, Weizenegger M, Neumaier J, Bachleitner M & Schleifer KH (1998) Bacterial phylogeny based on comparative sequence analysis. *Electrophoresis* **19**: 554–568.

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

Luter HM, Whalan S & Webster NS (2010) Exploring the role of microorganisms in the disease-like syndrome affecting the sponge *Ianthella basta*. *Appl Environ Microbiol* **76**: 5736–5744.

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.

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

Reiswig HM (1974) Water transport, respiration and energetics of three tropical marine sponges. *J Exp Mar Biol Ecol* **14**: 231–249.

Reiswig HM (1981) Partial carbon and energy budgets of the bacteriosponge *Verongia fistularis* (Porifera: Demospongiae) in Barbados West-Indies. *Mar Ecol* **2**: 273–294.

Ribes M, Jimenez E, Yahel G, Lopez-Sendino P, Diez B, Massana R, Sharp JH & Coma R (2012) Functional convergence of microbes associated with temperate marine sponges. *Environ Microbiol* **14**: 1224–1239.

Scanlan DJ & West NJ (2002) Molecular ecology oft he marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol* **40**: 1–12.

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

Schmitt S, Wehrl M, Lindquist N, Weisz JB & Hentschel U (2007) Morphological and molecular analyses of microorganisms in Caribbean reef adult sponges and in

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

corresponding reproductive material. *Porifera Research: Biodiversity, Innovation and Sustainability* (Custódio MR, Lôbo-Hajdu G, Hajdu E & Muricy G, eds), pp. 561–568. Série Livros 28, Museu Nacional, Rio de Janeiro.

Schmitt S, Deines P, Behnam F, Wagner M & Taylor MW (2012a) Chloroflexi bacteria are more abundant, diverse, and similar in high than in low microbial abundance sponges. *FEMS Microbiol Ecol* **78**: 497–510.

Schmitt S, Tsai P, Bell J *et al.* (2012b) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564–576.

Simister RL, Deines P, Botte ES, Webster NS & Taylor MW (2012) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ Microbiol* **14**: 517–524.

Sipkema D, Holmes B, Nichols SA & Blanch HW (2009) Biological characterisation of *Haliclona* (*?gellius*) sp.: sponge and associated microorganisms. *Microb Ecol* **58**: 903–920.

Sipkema D, Schippers K, Maalcke WJ, Yang Y, Salim S & Blanch HW (2011) Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge *Haliclona (gellius)* sp. *Appl Environ Microbiol* **77**: 2130–2140.

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

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

Thoms C, Horn M, Wagner M, Hentschel U & Proksch P (2003) Monitoring microbial diversity and natural product profiles of the sponge *Aplysina cavernicola* following transplantation. *Mar Biol* **142**: 685–692.

Turque AS, Cardoso AM, Silveira CB, Vieira RP, Freitas FAD, Albano RM, Gonzalez AM, Paranhos R, Muricy G & Martins OB (2008) Bacterial communities of the marine sponges *Hymeniacidon heliophila* and *Polymastia janeirensis* and their environment in Rio de Janeiro, Brazil. *Mar Biol* **155**: 135–146.

Usher KM, Fromont J, Sutton DC & Toze S (2004) The biogeography and phylogeny of unicellular symbionts in sponges from Australia and the Mediterranean. *Microb Ecol* **48**: 167–177.

Vacelet J & Donadey C (1977) Electron microscope study of the association of some sponges and bacteria. *J Exp Mar Biol Ecol* **30**: 301–314.

Webster N & Taylor MW (2012) Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol* **14**: 335–346.

Webster NS, Taylor MW, Behnam F, Luecker S, Rattei T, Whalan S, Horn M & Wagner M (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12**: 2070–2082.

Weisz JB, Hentschel U, Lindquist N & Martens CS (2007a) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Mar Biol* **152**: 475–483.

Weisz JB, Lindquist N & Martens CS (2007b) Do associated microbial abundances impact marine demosponge pumping rates and tissue densities. *Oecologia* **155**: 367–376.

Wilkinson CR (1978) Microbial associations in sponges. III. Ultrastructure of the *in situ* associations in coral reef sponges. *Mar Biol* **49**: 177–185.

Zhang H, Zhang W, Jin Y, Jin M & Yu X (2008) A comparative study on the phylogenetic diversity of culturable actinobacteria isolated from five marine sponge species. *Antonie Van Leeuwenhoek* **93**: 241–248.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1. 16S rRNA gene sequence-based maximum likelihood tree.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

4.3 Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea

This article was published in **Molecular Ecology** and a reprint is used here with the permission from John Wiley and Sons and the authors. This article's reference is:

Moitinho-Silva, L., Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., Ravasi, T., Hentschel, U. (2014) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology*. 23(6), 1348–63.

The documentation of author's contributions is found in the Appendix. The supplementary material is found in the CD attached to this PhD Thesis.

MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 1348–1363 doi: 10.1111/mec.12365

SPECIAL ISSUE: NATURE'S MICROBIOME

Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea

LUCAS MOITINHO-SILVA,* KRISTINA BAYER,* CARLO V. CANNISTRACI,† EMILY C. GILES,† TAEWOO RYU,† LOQMANE SERIDI,† TIMOTHY RAVASI† and UTE HENTSCHEL*

**Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Julius-von-Sachs Platz 3, 97082 Wuerzburg, Germany, †Division of Biological and Environmental Sciences & Engineering and Division of Applied Mathematics and Computer Science, Computational Biosciences Research Center, King Abdullah University of Science and Technology, Thuwal, 23955-6900, Kingdom of Saudi Arabia*

Abstract

Marine sponges are generally classified as high microbial abundance (HMA) and low microbial abundance (LMA) species. Here, 16S rRNA amplicon sequencing was applied to investigate the diversity, specificity and transcriptional activity of microbes associated with an LMA sponge (*Stylissa carteri*), an HMA sponge (*Xestospongia testudinaria*) and sea water collected from the central Saudi Arabia coast of the Red Sea. Altogether, 887 068 denoised sequences were obtained, of which 806 661 sequences remained after quality control. This resulted in 1477 operational taxonomic units (OTUs) that were assigned to 27 microbial phyla. The microbial composition of *S. carteri* was more similar to that of sea water than to that of *X. testudinaria*, which is consistent with the observation that the sequence data set of *S. carteri* contained many more possibly sea water sequences (~24%) than the *X. testudinaria* data set (~6%). The most abundant OTUs were shared between all three sources (*S. carteri*, *X. testudinaria*, sea water), while rare OTUs were unique to any given source. Despite this high degree of overlap, each sponge species contained its own specific microbiota. The *X. testudinaria*-specific bacterial taxa were similar to those already described for this species. A set of *S. carteri*-specific bacterial taxa related to *Proteobacteria* and *Nitrospira* was identified, which are likely permanently associated with *S. carteri*. The transcriptional activity of sponge-associated microorganisms correlated well with their abundance. Quantitative PCR revealed the presence of Poribacteria, representing typical sponge symbionts, in both sponge species and in sea water; however, low transcriptional activity in sea water suggested that Poribacteria are not active outside the host context.

Keywords: marine sponges, microbial diversity, pyrosequencing, symbiosis, sponge-specific microbes, transcriptional activity

Received 14 February 2013; revision received 16 April 2013; accepted 18 April 2013

Introduction

Many marine sponges (phylum *Porifera*) host stable associations of symbiotic microbial consortia that are located mainly extracellularly in the mesohyl matrix. The microbial biomass in these sponges is massive and can contribute up to 35% of the animal's biomass (Hentschel *et al.* 2012). Representatives of more than 30 bacterial phyla, candidate phyla and two archaeal lineages have been identified in some sponge species thus far (Webster *et al.* 2010; Schmitt *et al.* 2012a). Remarkably, though, the mesohyl of other sponge species, existing in the same habitat, is virtually devoid of microorganisms. This classification into high microbial abundance

Correspondence: Ute Hentschel, Fax: +49-931-31-86235; E-mail: ute.hentschel@uni-wuerzburg.de and Timothy Ravasi, Fax: +966-2-8020127; E-mail: timothy.ravasi@kaust.edu.sa

© 2013 John Wiley & Sons Ltd

(HMA) or 'bacteriosponges' and low microbial abundance sponges (LMA) or 'nonsymbiont-harbouring sponges' has long been recognized (Vacelet & Donadey 1977; Hentschel *et al.* 2003).

While the microbial diversity of HMA sponges has been studied in great detail (Schmitt *et al.* 2012b), LMA sponges have been investigated far less in this regard. Recent studies have shown that they differ from their HMA counterparts in several aspects. Bacterial abundance is two to five orders of magnitude lower (10^5–10^6 bacteria/g or mL of tissue) than that of HMA sponges (10^8–10^{10} bacteria/g or mL of tissue) and is more in the range of that found in sea water (Hentschel *et al.* 2006). Transmission electron microscopy images have shown that the mesohyl of LMA sponges is essentially devoid of microorganisms (Vacelet & Donadey 1977; Weisz *et al.* 2007; Kamke *et al.* 2010; Giles *et al.* 2013). The larvae of LMA sponge species are, in contrast to those of HMA species (Schmitt *et al.* 2007), also largely bacteria free.

Besides bacterial abundances, there are also noticeable differences with respect to microbial diversity. Several studies, employing a variety of methods (denaturing gradient gel electrophoresis (DGGE), clone library, 454 amplicon sequencing), consistently demonstrated a lower microbial diversity in LMA than in HMA sponges (Weisz *et al.* 2007; Kamke *et al.* 2010; Erwin *et al.* 2011; Gloeckner *et al.* 2013a; Schmitt *et al.* 2012b; Giles *et al.* 2013). Each LMA species was dominated by a large clade of *Cyanobacteria*, *Alpha*-, *Beta*- or *Gammaproteobacteria,* and there was little overlap between the LMA sponge microbiomes under investigation.

To provide insights into the specificity of the host–microbe association, the term sponge-specific cluster (SC) was defined, which refers to a group of at least three 16S rRNA gene sequences that have been recovered from different sponge species and/or from different geographical locations and are more closely related to each other than to any other sequence from non-sponge sources (Hentschel *et al.* 2002). This early concept was revisited using a much larger sequence data set, and sponge specificity or sponge/coral specificity (SC/SCC) was confirmed for one-third of all sequences (Taylor *et al.* 2007; Simister *et al.* 2012c). Although the SC/SCCs are an important concept in sponge microbiology (Hentschel *et al.* 2012), it applies only poorly to LMA sponge species, as SC/SCCs are rarely found in these types of sponges (Erwin *et al.* 2011; Giles *et al.* 2013). In consequence, the microbiome composition of LMA sponges and its variation over space, time and with host phylogeny are currently far from being understood. In this study, we sought to investigate the community composition of an LMA sponge (*Stylissa carteri*) in comparison with an HMA sponge (*Xestospongia testudinaria*) and to sea water by 454 amplicon sequencing. We further included 16S rRNA transcript data as a measure of transcriptional activity. The results provide novel insights into the microbiology of an LMA sponge and support the LMA vs. HMA dichotomy in general. In particular, (i) we show that an LMA sponge harbours its own distinct microbiota and is not merely a reflection of sea water; (ii) we report on the percentages of sea water bacterial sequences in the respective sponge microbiomes; (iii) we show that sequencing depth has profound effects on the percentage of unique vs. shared sequences between sources, which, in return, affects our perception of specificity; (iv) we introduce an analysis based on sequence abundance Minimum Curvilinear embedding (MCE) rather than on presence/absence that was found to be instrumental to recover patterns of host specificity; and (v) we report here that poribacterial symbionts may be present but not active outside the sponge context. Conceptually, as a result of the superior sequencing depth obtained with amplicon sequencing, we propose to change the term 'sponge-specific' to 'sponge-enriched' to account for the fact that sponge-specific microbes also occur in the environment, albeit at much lower numbers.

Materials and methods

Sponge and sea water collection

Sponge samples from the species *Stylissa carteri* (n = 3, termed hereafter AS1, AS2, AS3 to reflect **a**mplicons from ***S****tylissa*) and *Xestospongia testudinaria* (AX1, AX2, AX3 to reflect **a**mplicons from ***X****estospongia*) were collected by SCUBA diving at Fsar reef (22°23′N; 39°03′E) off the coast of Thuwal, Saudi Arabia. Sea water samples were collected during the same dive before (WT1), during (WT2) and after (WT3) collection of sponge samples. The sponge collection occurred at 8.5–12 m depth, on 9 January 2012 in the morning (9–10 am), with water temperature measured at 26 °C. Sponge and sea water samples were collected within a 20 m² range with a dive knife and were processed individually. Immediately after removal, the sponges were placed into zip lock bags containing sea water and brought on board the ship. Sponge samples were cut into 5-mL pieces, rinsed two times with artificial sea water, wrapped in aluminium foil and snap-frozen in liquid nitrogen. Sponge pieces were taken such that portions of pinacoderm, mesohyl and choanoderm were included in each sample. The time span between the first and last sponge sampled was less than 1 h, while the time span between cutting and freezing was less than 10 min. Approximately 7 L of sea water was

© 2013 John Wiley & Sons Ltd

filtered on board (WT1) or upon return to the laboratory within less than 1 h (WT2 and WT3) onto 0.22 μm, hydrophilic Durapore membrane filters (Millipore, USA), using a Masterflex® I/P® Easy-Load® peristaltic pump (Cole-Parmer, USA). The filtering time was approximately 50 min. The frozen filters and sponge pieces were kept at −80 °C until further processing.

Nucleic acid extraction and amplification of 16S rRNA genes and transcripts

DNA and RNA were co-extracted from sponge and sea water samples using the AllPrep DNA/RNA mini kit (QIAGEN, Germany) according to the manufacturer's instructions, after cell lysis according to Giles *et al.* (2013). After digestion of genomic DNA, RNA samples were converted to single-strand cDNA with random hexamers using the SuperScript III First Strand Synthesis System (Invitrogen, USA; Kamke *et al.* 2010). Nucleic acid integrity was analysed by agarose gel electrophoresis for DNA samples and Agilent 2100 Bioanalyzer for RNA samples (Agilent Technologies, USA). DNA concentrations were measured with the NanoDrop 2000c (PEQLAB Biotechnologie GmbH, Germany) or using the Qubit® 2.0 Fluorometer and the Qubit® dsDNA BR and HS Assay Kit (Invitrogen, Germany). DNA was stored at −20 °C until further use. RNA was stored at −80 °C.

The genes and transcripts of 16S rRNA were amplified from sea water, *S. carteri* and *X. testudinaria* DNA (producing the following data sets: WTD1–WTD3, ASD1–ASD3 and AXD1–AXD3, respectively) and cDNA (producing the following data sets: WTR1–WTR3, ASR1–ASR3 and AXR1–AXR3; respectively). PCRs were performed in triplicate using multiplex identifier adaptor-ligated primers. To ensure amplification of 16S rRNA genes or transcripts from members of the candidate phylum *Poribacteria*, we used the primers 533f (5′-GTGCCAGCAGCYGCGGTMA-3′) and 907r (5′-CCG TCAATTMMYTTGAGTTT-3′; Simister *et al.* 2012b). The sequence between 533f and 907r primers includes the hypervariable regions V4 and V5. The PCR was performed using the Qiagen multiplex PCR kit (QIAGEN, Germany), with Q-solution as buffer and 0.4 μM of each primer. The PCR protocol was as follows: initial denaturation at 95 °C for 15 min, followed by 30 cycles of denaturation (94 °C for 30 s), annealing (62 °C for 90 s) and extension (72 °C for 90 s). A final extension for 10 min at 72 °C was performed at the end of the run. The absence of DNA contamination on cDNA samples was confirmed by performing PCR with respective total RNAs as templates. Purified amplicons were pooled at equimolar ratios and sequenced on a Roche 454 GS-FLX Titanium platform. Nucleic acid amplification and sequencing were performed by the KAUST Genomics Core Lab using the Roche 454 GS-FLX (titanium) sequencer system.

Processing of sequences and taxonomic assignment

The 16S rRNA amplicon sequences were processed with the QIIME pipeline v 1.4 (Caporaso *et al.* 2010b). Raw sequences with less than 200 bp length or with mismatches on primer or barcode, or containing a homopolymer longer than 6 nucleotides, or a mean quality score below 25, were discarded. The quality score window was set at 50 bp. Sequences with low-quality score window were truncated to the first base of the window. Denoising of 454 sequences was performed following QIIME standard procedures using denoise_wrapper.py script. The resulting centroids and singletons were re-integrated in the pipeline. Sequences with more than 97% similarity were clustered into operational taxonomic units (OTUs) by UCLUST (Edgar 2010) using optimal configuration to assure the best possible choice of OTUs. Representative sequences of each OTU were used for alignment against SILVA 108 core sequences (Pruesse *et al.* 2007) with PyNAST (Caporaso *et al.* 2010a). Putative chimeras were identified from representative sequences set by ChimeraSlayer (Haas *et al.* 2011) and Ccode (Gonzalez *et al.* 2005) using mothur v 1.25 (Schloss *et al.* 2009). Chimeras were removed from the data set if present in just one data set. Remaining singletons were removed from the whole data set. The remaining OTUs were taxonomically classified by the RDP Classifier (Wang *et al.* 2007) and BLAST (Altschul *et al.* 1990) using the SILVA 108 database as reference (> 97% similarity clustered). Poribacterial sequences obtained from the SILVA database were re-annotated as belonging to the *Poribacteria* phylum instead of the *Firmicutes* phylum, as was the case in the SILVA 108 taxonomy. The primary taxonomic classification was performed by RDP Classifier (0.8 of confidence) up to the genus level. Unclassified OTUs were annotated at phylum level according to the best BLAST hit. Eukaryotic and chloroplast sequences were excluded from downstream analyses. The pyrosequencing data set originating from sample ASD2 was contaminated with *X. testudinaria* DNA and was not considered for analysis.

Assignment of OTU representative sequences

The OTUs were assigned to sponge-specific sequences (Simister *et al.* 2012c) or to those found previously in the same sponge species (Montalvo & Hill 2011; Giles *et al.* 2013). A total of 2500 sequences from sponge-specific clusters or sponge- and coral-specific clusters (SC/SCC) were retrieved from a comprehensive phylogenetic database originating from isolates, clone library or DGGE band sequences (Simister *et al.* 2012c).

© 2013 John Wiley & Sons Ltd

Previously published sequences were obtained for *S. carteri* (67 sequences), from the same location but different collection date (November 2010; Giles *et al.* 2013), and *X. testudinaria* (414 sequences), collected at Manado Bay, Indonesia (01°32′N, 124°55′E) in 2003 and 2005 (Montalvo & Hill 2011). Briefly, the 533–907 base pair fragments of the 16S rRNA reference sequences were extracted using PrimerProspector v 1.0.1 (Walters *et al.* 2011), with 0 penalty for gaps and mismatches on primers. The SC/SCC 16S rRNA gene fragments were only considered as reference for assignment (total of 1153 sequences) when they did not cluster at >97% similarity thresholds with nonspecific sponge-derived sequences from the same database. Thus, a total of 285 SC/SCC sequences representing 26 taxa were omitted (Table S1, Supporting information). All extracted 533–907 base pair fragments from *S. carteri* (*S. carteri* reference sequences) and *X. testudinaria* (*X. testudinaria* reference sequences) studies (Montalvo & Hill 2011; Giles *et al.* 2013) were considered as reference. Finally, the OTUs were assigned when their representative sequences clustered at 97% similarity or higher with reference 16S rRNA gene fragments.

Analysis of community richness, diversity and dissimilarity

The ecological indices were measured on data sets rarefied to the smallest sampling effort as recommended by Gihring *et al.* (2012). Richness was estimated by Chao1 and abundance-based coverage estimator (ACE) indices (Chao 1984; Chao & Lee 1992). The Simpson and Shannon indices were calculated to assess diversity (Shannon 1948; Simpson 1949). All ecological indices were performed using vegan package v 2.0–4 (Oksanen *et al.* 2012) in an R environment. The UniFrac analysis (Lozupone & Knight 2005) was performed on rarefied data sets using the QIIME pipeline. UniFrac results were clustered by an unweighted-pair group method using average linkages (UPGMA). Node support was obtained using subsets of available data (jackknife) based on 100 repetitions at 10 000 sequences per sample. To compare materials and sources, data sets were pooled according to UniFrac clusters, by summing sequences per OTU and compared using the Bray–Curtis dissimilarity index (calculated as in the vegan package; Bray & Curtis 1957).

Analysis and unsupervised exploration of overlap and OTU abundance patterns

To assess the overlap of OTUs among sources, the rarefied data sets were pooled according to their source. The shared and unique OTUs were visualized using Limma package v 3.12.1 (Smyth 2005) in R. The relative abundance of OTUs was obtained by calculating the percentage of sequences assigned to a given OTU in relation to the total sequence count in each source. The Minimum Curvilinear embedding (MCE; Cannistraci *et al.* 2010) algorithm was used for unsupervised exploration and discrimination of OTU abundance patterns. The Minimum Curvilinearity distance matrix—adopted for MCE dimension reduction—was computed starting from the Euclidean distances between the OTU abundances in rarefied data sets (Cannistraci *et al.* 2010). Finally, the relative abundance of OTUs in each data set was inspected using a heatmap ordered according to MCE first dimension (D1). MCE analysis was performed in Matlab and the heatmap was generated using R. Furthermore, detrended correspondence analysis (DCA) was performed based on Bray–Curtis dissimilarities between OTUs in each rarefied source data set. Prior to ordination, the data sets were square root-transformed and Wisconsin double-standardized. Assignments of OTUs to SC/SCC, to *S. carteri* reference sequences and to *X. testudinaria* reference sequences were projected as arrows onto DCA plot. Arrow lengths were scaled according to their correlation. Correlation *P*-values were calculated based on 1000 permutations using coefficient of determination (R-squared). DCA was performed in R using vegan package.

Detection of transcriptionally active taxa

Sequence counts were pooled according to the deepest taxonomical annotation of OTUs. The relative abundance of a given taxon is represented by the number of its sequences in relation to the total number of sequences in the data set. The correlation of means of relative abundances of taxa between data sets was obtained using R-squared. The null hypothesis that relative abundance of sequences assigned to a particular taxon in each cDNA and DNA data set pair (originated from the same sample) is due to their distribution was tested using two-sided Fisher's exact test (Rivals *et al.* 2007) performed using STAMP v2.0.0 (Parks & Beiko 2010). The Storey's false discovery rate was used to correct *P*-values for multiple hypotheses testing. As recommended (Parks & Beiko 2010), taxa with corrected $P < 0.05$ were further filtered for: (i) low number of sequences assigned to the taxa; (ii) an absolute effect size statistic, that is the difference between proportions (DP); and (iii) a relative effect size statistic, that is the ratio of proportions (RP). Thus, the variation in the relative abundance of a taxon between cDNA and DNA was considered statistically significant if: (i) the particular taxon had more than 10 sequence counts in each paired data set; (ii) the DP (calculated as relative

© 2013 John Wiley & Sons Ltd

abundance of taxon sequences in the cDNA data set—relative abundance of taxon sequences in the DNA data set) was larger than 1.5% or the RP (cDNA/DNA) was larger than 0.5; and (iii) the results of all paired comparisons within a source were in agreement. The sample ASR2 was omitted from these analyses due to the exclusion of ASD2 data set.

Quantitative PCR of poribacterial 16S rRNA genes and transcripts

To quantify the poribacterial 16S rRNA gene and transcript numbers in the samples, quantitative PCR (qPCR) assays were performed. The primers were adapted from previously described Poribacteria-specific FISH probes Por389 and Por600 (Fieseler *et al.* 2004). Primer specificity was tested by cloning of PCR products (48 clones) and restriction fragment length polymorphism (RFLP) analysis using two restriction enzymes (which resulted in five different fragment patterns) and sequencing using metagenomic DNA from the Mediterranean sponge *Aplysina aerophoba* as template as described previously (Bayer *et al.* 2008). Briefly, one RFLP pattern represented the majority of clones (91.6%), whereas the other four patterns were represented by one clone each (singletons). Among these singletons, only one did not present high similarity to poribacterial sequence, but rather to an uncultured cyanobacterial clone (EF076240; 93% sequence similarity). Therefore, the poribacterial primers were considered sufficiently specific for qPCR.

Quantitative PCR was performed in a CFX96™ real-time detections system (Bio-Rad, Germany) using the Maxima SYBR Green qPCR Master Mix (Fermentas, Germany) following the manufacturer's instructions. For absolute quantification of poribacterial 16S rRNA genes and transcripts, a serial dilution (1:5) of plasmid aqueous solution (10 ng/μL of tRNA) ranging from 1.65×10^9 to 1.056×10^5 copy numbers was used as external standard. A two-step protocol was carried out as follows: initial denaturation and enzyme activation at 95 °C for 10 min, followed by 40 cycles of denaturation (10 s at 95 °C) and elongation (40 s at 58 °C). Plate reads were taken at the end of each qPCR cycle. For each assay, samples were tested in triplicate. Amplification of specific targets was confirmed by analyses of melt curves (in steps of 0.5 °C for 5 s, with temperatures ranging from 60 °C to 95 °C). Additionally, PCR product sizes were checked on a 2% agarose gel electrophoresis. The qPCR efficiency and gene and transcript copy numbers were calculated using Bio-Rad CFX Manager software (version 1.6.541.1028). The quantification of *Poribacteria* in sea water and sponge nucleic acids (DNA and cDNA) was performed by duplicate qPCR assays with efficiencies of 98.6% and 102.6%. Comparison of 16S rRNA copy estimations among sources was performed by nonparametric Kruskal–Wallis test followed by Dunn's multiple comparisons test using GraphPad Prism version 6.01 for Windows (GraphPad Software, USA).

Results

Richness, diversity and dissimilarity analyses

A total of 887 068 curated denoised sequences were clustered based on >97% sequence similarity into 4089 OTUs. OTUs were excluded if they were identified as chimeric (1814), eukaryotic (20) or chloroplast derived (175). Singletons (2297) and OTUs that did not align with sequences from the SILVA reference database (169) were also excluded. This effort resulted in 1477 microbial OTUs, comprising 91% (806 661) of denoised sequences. Rarefaction curves based on total sequence counts approximate asymptotic behaviour for all data sets, in particular that of *Xestospongia testudinaria*, although additional OTUs might have been discovered with deeper sequencing (Fig. S1, Supporting information).

The OTU richness and diversity estimations were obtained from rarefied data sets to the smallest sampling effort (32 148 sequences in each data set, comprising a total of 1445 OTUs, out of 546 516 sequences). The OTU richness estimators indicated that the majority of the OTUs from sea water, *Stylissa carteri* and *X. testudinaria*, were recovered (Table 1). Overall, sea water presented the highest OTU richness (defined as the number of species in a given community), which was followed by *S. carteri* and *X. testudinaria*. With respect to OTU diversity (combining richness and evenness, taking the relative abundances of species into account), *X. testudinaria* was more diverse than sea water or *S. carteri*. When comparing DNA and cDNA data sets, sea water and *S. carteri* displayed higher richness in DNA than in cDNA data sets, whereas *X. testudinaria* samples showed higher richness in cDNA than in DNA data sets (Table 1, see Table S2, Supporting information for data sets). With respect to OTU diversity, the DNA data sets were more diverse than the corresponding cDNA data sets.

The rarefied data sets were further analysed using weighted UniFrac, clustered by UPGMA and compared by Bray–Curtis dissimilarity index (Fig. 1A). The data sets clustered first according to material (DNA and cDNA) and then to source (sea water, *S. carteri* and *X. testudinaria*). Unweighted UniFrac analysis resulted in the same pattern (data not shown). The Bray–Curtis dissimilarity values between replicates were consistently low. With respect to material, the microbial

© 2013 John Wiley & Sons Ltd

Table 1 Observed number of operational taxonomic units (OTUs) and estimations of richness (Chao1, ACE) and diversity (Shannon, Simpson)

	OTU richness			OTU diversity	
Material	Observed OTUs	Chao1	ACE	Shannon	Simpson
Sea water DNA	695 ± 13	974 ± 17	999 ± 25	3.45 ± 0.05	0.89 ± 0.00
Sea water cDNA	518 ± 72	762 ± 125	776 ± 109	1.63 ± 0.15	0.44 ± 0.04
Stylissa carteri DNA	444 ± 10	685 ± 71	729 ± 83	2.46 ± 0.01	0.78 ± 0.01
S. carteri cDNA	433 ± 24	674 ± 83	699 ± 61	1.95 ± 0.05	0.71 ± 0.01
Xestospongia testudinaria DNA	251 ± 5	297 ± 26	299 ± 25	4.32 ± 0.05	0.98 ± 0.00
X. testudinaria cDNA	274 ± 11	353 ± 35	335 ± 26	3.82 ± 0.02	0.96 ± 0.00

Standard deviations (n = 3, *S. carteri* DNA n = 2) are shown. Values were obtained from rarefied data sets.

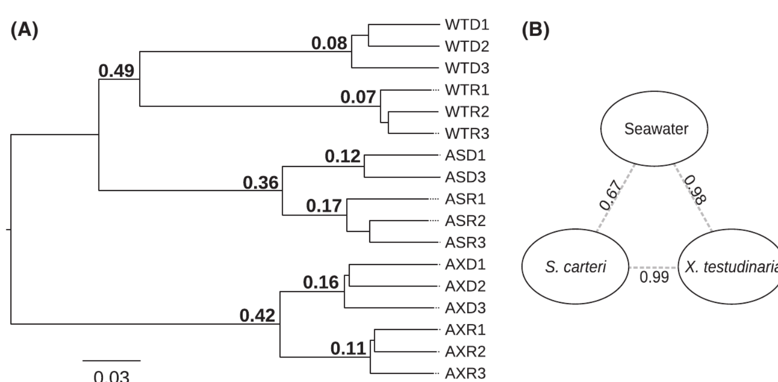


Fig. 1 Distance analysis between communities. (A) unweighted-pair group method using average linkages (UPGMA) analysis of amplicon data sets obtained from microbiota of sea water, *Stylissa carteri* and *Xestospongia testudinaria* based on a weighted UniFrac distance matrix. The data sets were rarefied to the minimum sampling effort. Tree nodes represent jackknife support values >0.8 based on 100 repetitions at 10 000 sequences per sample. The scale bar indicates the distance between clusters in UniFrac units. Bray–Curtis dissimilarity values are shown on the nodes of the tree. (B) Bray–Curtis dissimilarity values between sources.

communities in DNA and cDNA data sets were more dissimilar in sea water (0.5) than in *X. testudinaria* (0.42) or *S. carteri* (0.36). With respect to source, the microbiota of *X. testudinaria* was more dissimilar to either those of sea water (0.98) or of *S. carteri* (0.99; Fig. 1B). In other words, the sea water microbial community was more closely related to that of *S. carteri* (0.67) than to *X. testudinaria* (0.98).

OTU specificity

To analyse the overlap of OTUs among sources, the rarefied data sets were pooled for each source. This resulted in combined data sets for sea water (a total of 1174 OTUs, comprising 192 888 sequences), *S. carteri* (a total of 887 OTUs, comprising 160 740 sequences) and *X. testudinaria* (a total of 424 OTUs, comprising 192 888 sequences). The OTU overlap was visualized by a Venn diagram (Fig. 2A). The majority of OTUs (a total of 836 OTUs, accounting for 99.3% of sequences of the total data set) were shared with at least one other source. Of these, 204 OTUs were shared between all three sources (Fig. 2A), comprising the most abundant OTUs and >90% of sequences in each source (Fig. 2B). On the other hand, the unique OTUs of each source were present at low abundance (<1%; Fig. 2B). It is important to realize that sequencing depth had profound effects on the percentage of unique sequences that decreased with increasing sample size (Fig. 2C).

The patterns of OTU abundances were further investigated in the rarefied data sets, using the MCE algorithm for unsupervised dimension reduction (see Materials and methods for details). The relative distances between OTUs and the presence of patterns were inspected using OTU's ordination according to the MCE first dimension (D1) of embedding (Fig. 3A). This analysis unveiled the presence of two main patterns, sorting the OTUs into two main groups: (i) the leftmost of the ordination grouped OTUs that were highly abundant in sponges but absent from sea water (Fig. 3B); and (ii) the rightmost of the ordination grouped OTUs that were highly abundant in sea water and *S. carteri*

© 2013 John Wiley & Sons Ltd

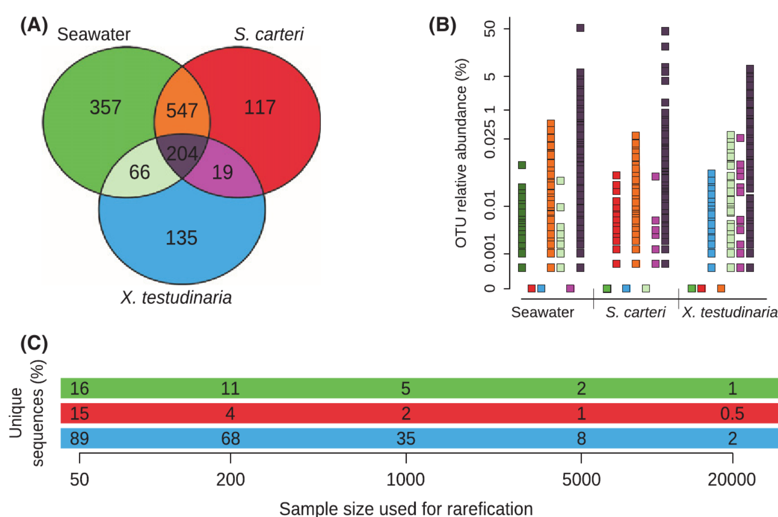


Fig. 2 Analysis of overlap between communities. (A) Venn diagram of operational taxonomic units (OTUs) recovered from sea water, from *Stylissa carteri* and from *Xestospongia testudinaria* data sets. (B) Relative abundance of OTUs in each source. The OTUs were sorted and coloured according to the Venn diagram subgroups. (C) The effects of sequencing depth on the percentage of unique sequences.

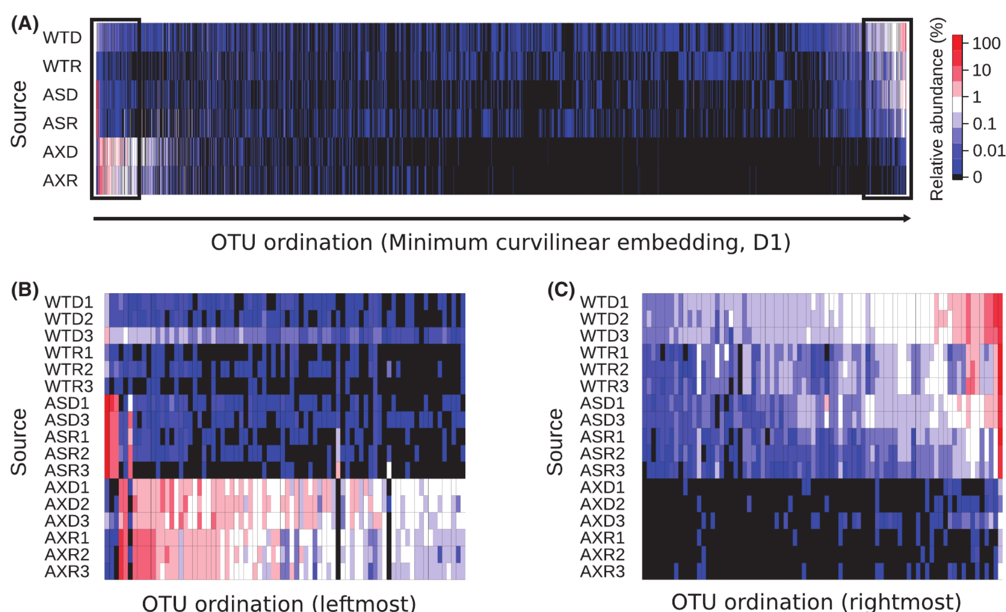


Fig. 3 Analysis of operational taxonomic units (OTU) distribution patterns. (A) Heatmap of relative OTU abundances (%, mean) in each data set. The OTUs were ordered according to the Minimum Curvilinear embedding (MCE) first dimension (D1). (B) Leftmost and (C) rightmost sides of the heatmap (black rectangle) were enlarged. See Materials and methods for source abbreviations.

(Fig. 3C). The closer inspection of OTU patterns of the first group revealed few highly abundant OTUs in *S. carteri* that were present at low abundances or missing from sea water and *X. testudinaria* (Fig. 3B). Furthermore, the pattern revealed many highly abundant OTUs in *X. testudinaria* that were present at low abundances or missing from sea water and *S. carteri*. Interestingly, the DNA of sea water that was sampled after sponge collection (WTD3) as opposed to prior and during sponge collection (WTD1, WTD2) appeared to be enriched in OTUs from the left side of the ordination. Finally, the most abundant OTUs in sea water were also

© 2013 John Wiley & Sons Ltd

highly abundant in *S. carteri* but not in *X. testudinaria* (Fig. 3C).

To assess the fraction of sea water bacteria in the sponge microbiomes, the OTUs on the leftmost side of the MCE first dimension (D1) were taken as being specific to *S. carteri* or *X. testudinaria*. Of 80 OTUs counted each, 6 were specific to *S. carteri* and 74 were specific to *X. testudinaria* (Fig. 3B). These numbers accounted for 65% and 94% of the *S. carteri* or *X. testudinaria* sequences, respectively. In consequence, 35% and 6% are considered as being possibly derived from sea water. Because 11% of the *S. carteri* sequences (referring to the *Synechococcus* clade, see Results section on taxon abundances) could not be identified unambiguously as being sponge- or sea water specific, we conservatively calculate that as much as 24% of the *S. carteri* microbiome may be derived from sea water. To verify whether the recovered OTUs are consistently (across space and time) associated with their sponge hosts, the OTUs were assigned thresholds based on >97% similarity to SC/SCC sequences (Simister *et al.* 2012c), sequences originating from *S. carteri* collected at a different time (*S. carteri* reference sequences; Giles *et al.* 2013) and sequences originating from *X. testudinaria* collected at a different time and location (*X. testudinaria* reference sequences; Montalvo & Hill 2011; Table S3, Supporting information). Based on DCA ordination (Fig. S2, Supporting information), the OTUs assigned to SC/SCC sequences were in agreement with the *X. testudinaria* data sets ($P < 0.001$). Further, the assignment of OTUs to *S. carteri* reference sequences or *X. testudinaria* reference sequences correlated well with the respective sponge data sets ($P < 0.001$).

Taxon abundances

Altogether, 27 bacterial and archaeal phyla were identified in the entire data set. The sea water data sets were dominated by *Proteobacteria* (41% and 17.5% mean sequence abundance in DNA and cDNA data set), *Cyanobacteria* (30.5% and 79%), *Bacteroidetes* (12.7% and 1.3%), *Actinobacteria* (10.0% and 0.3%) and *Deferribacteres* (3.3% and 1.2%). Each of the remaining 21 phyla contributed <1% to the sea water data set sequences (Fig. 4). The five most abundant taxa in sea water DNA data sets were as follows: *Synechococcus* (29.9% mean sequence abundance), the *Candidatus* Pelagibacter of *Alphaproteobacteria* (10.5%), the OCS155 marine group of *Actinobacteria* (9.0%), the unclassified lineage(s) of *Rhodobacteraceae* family (*Proteobacteria*; 4.5%) and the NS2b marine group of *Bacteroidetes* (3.73%; Table 2). The majority of sequences from the most abundant sea water taxa (>1% of relative abundance) were placed at the right side of MCE ordination (Fig. 3, Table S4, Supporting information). As expected, the majority of these sequences were dissimilar to SC/SCC sequences as well as previously published *X. testudinaria* reference sequences, with the only exception observed for

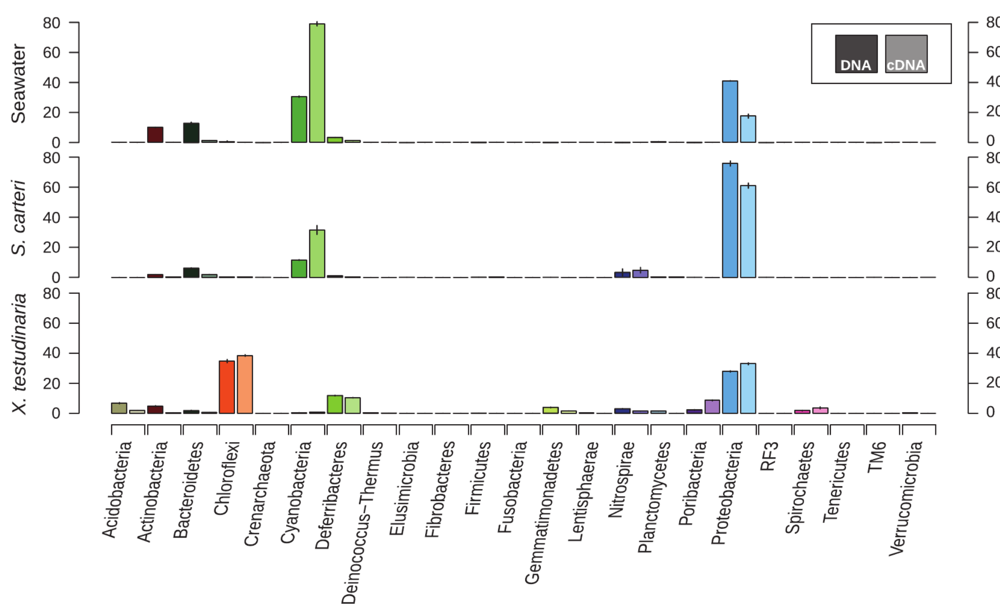


Fig. 4 Phylum distribution in sea water, *Stylissa carteri* and *Xestospongia testudinaria* data sets. The bars represent the number of 16S rRNA sequences (mean ± standard deviation) that were assigned to a given phylum in relation to the total number of sequences in each data set ($n = 3$, *S. carteri* DNA $n = 2$). The 23 phyla found in sponges (of 27 found in total) are shown. 16S rRNA sequences obtained from DNA and cDNA are presented.

© 2013 John Wiley & Sons Ltd

Table 2 The most abundant taxa in DNA data sets of sea water, *Stylissa carteri* and *Xestospongia testudinaria*

Phylum	Taxon†	Relative abundance of taxa (%)*	Relative abundance of assigned sequences within taxa (%)			
			Heatmap position (%)§	SC/SSC	*S. carteri* reference sequences	*X. testudinaria* reference sequences
Sea water						
Cyanobacteria	*Synechococcus*‡	29.91 ± 0.79	Right (98)	—	98.87	98.87
Proteobacteria	*Candidatus* Pelagibacter	10.5 ± 0.65	Right (100)	—	100.00	—
Actinobacteria	OCS155 marine group	9.02 ± 0.4	Right (99)	0.02	—	0.05
Proteobacteria	*Rhodobacteraceae*[f]	4.49 ± 0.04	Right (96)	0.72	—	—
Bacteroidetes	NS2b marine group	3.73 ± 0.29	Right (99)	—	—	—
S. carteri						
Proteobacteria	E01-9C-26 marine group	42.07 ± 1.54	Left (99)	—	99.97	0.01
Proteobacteria	*Gammaproteobacteria*[c]	13.14 ± 4.31	Left (99)	99.19	99.14	—
Cyanobacteria	*Synechococcus*‡	11.33 ± 0.38	Right (99)	—	99.06	99.06
Proteobacteria	*Proteobacteria*[p]	7.98 ± 1	Left (82)	—	82.72	—
Nitrospirae	*Nitrospira*	3.25 ± 2.39	Left (100)	—	99.63	0.37
X. testudinaria						
Chloroflexi	SAR202 clade	21.36 ± 0.78	Left (97)	77.91	—	94.25
Deferribacteres	PAUC34f	11.58 ± 0.41	Left (99)	44.70	—	99.64
Chloroflexi	*Caldilinea*	5.2 ± 1.4	Left (100)	100.00	—	100.00
Actinobacteria	Sva0996 marine group	4.17 ± 0.55	Left (99)	12.42	—	86.84
Chloroflexi	Uncultured (family *Anaerolineaceae*)	3.77 ± 0.48	Left (87)	4.14	—	63.64

*Standard deviations (*S. carteri n* = 2, sea water and *X. testudinaria n* = 3) are shown.
†Taxon annotation represents taxonomic assignment at genus level. Exceptions are indicated for phylum (p), class (c) and family (f) levels.
‡Taxon which 533–907 16S rRNA gene sequence fragments were excluded from SC/SSC reference sequences due to >97% similarity with nonspecific sponge-associated sequence fragments (see Materials and methods and Table S1, Supporting information).
§Heatmap position is in consideration of the analysis presented in Fig. 3. 'Right' and 'Left' comprise the 100 first and last OTUs as determined by the first dimension of Minimum Curvilinear embedding (MCE). 'Left' represents a group of OTUs highly abundant in *S. carteri* or *X. testudinaria*, but not in sea water. 'Right' represents a group of OTUs highly abundant in sea water and *S. carteri*. Middle comprises OTUs between the extremes of the ordination.

Synechococcus. It is important to note that the *Synechococcus* 533–907 base pair fragments from reference 16S rRNA database (Simister *et al*. 2012c) showed >97% similarity between SC/SCC and nonspecific sponge-associated sequences (Table S1, Supporting information). For this reason, these fragments were omitted from the SC/SCC assignment and conclusions cannot be drawn.

The *S. carteri* data sets were dominated by *Proteobacteria* (75.7% and 61% mean sequence abundance in DNA and cDNA data sets), *Cyanobacteria* (11.5% and 31.5%), *Bacteroidetes* (6.0% and 1.9%), *Nitrospirae* (3.3% and 4.7%) and *Actinobacteria* (1.9% and 0.2%; Fig. 4). The other 18 phyla each contributed <1% to the *S. carteri* data set sequences. The five most abundant taxa on the genus- and family level in the *S. carteri* DNA data sets were the E01-9C-26 marine group of *Proteobacteria* (42.0%), the unclassified lineage(s) of *Gammaproteobacteria* (13.1%, best blast hit of the most abundant OTU: JQ062854.1), *Synechococcus* (11.3%), the unclassified lineage(s) of *Proteobacteria* (8%, best blast hit of the most abundant OTU: JQ062836.1) and *Nitrospira* (3.25%; Table 2). With exception of *Synechococcus*, most of the sequences from these taxa were placed at the left side of the heatmap (Fig. 3), corresponding to highly abundant OTUs in *S. carteri*. Furthermore, the majority of these sequences were similar to *S. carteri* reference sequences (>82% of taxon sequences). However, sequences from the unclassified lineage(s) of *Gammaproteobacteria* (99%) were similar to SC/SCC sequences (Table 2). Other abundant taxa in *S. carteri* were placed at the right side of the MCE ordination (Fig. 3, Table S5, Supporting information), in spite of some of the taxa being similar to those obtained previously from *S. carteri* reference sequences.

The most abundant phyla in the *X. testudinaria* data sets were *Chloroflexi* (34.7% and 38.5% mean sequence abundance in DNA and cDNA), *Proteobacteria* (27.8% and 32.3%), *Deferribacteres* (11.7% and 10.3%), *Acidobacteria* (6.7% and 1.9%), *Actinobacteria* (4.7% and 0.3%), *Gemmatimonadetes* (3.8% and 1.5%), *Nitrospirae* (3.0% and

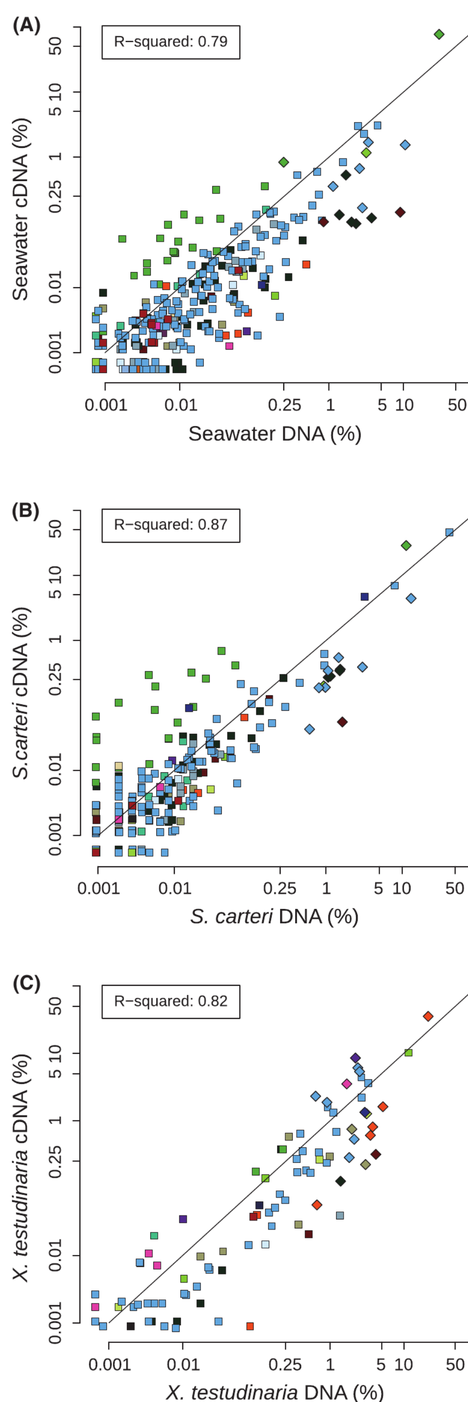
© 2013 John Wiley & Sons Ltd

1.3%), *Poribacteria* (2.2% and 8.5%), *Spirochaetes* (1.7% and 3.5%), *Bacteroidetes* (1.7% and 0.5%) and *Planctomycetes* (1.4% and 0.03%). The other eight phyla each contributed to <1% of the *X. testudinaria* data set sequences (Fig. 4). The five most abundant taxa in the *X. testudinaria* DNA data sets were related to the SAR202 clade of *Chloroflexi* (21.3%), the PAUC34f clade of *Deferribacteres* (11.6%), the *Caldilinea* genus of *Chloroflexi* (5.2%), the Sva0996 marine group of *Actinobacteria* (4.17%) and uncultured bacteria from the *Anaerolineaceae* family of *Chloroflexi* (3.8%; Table 2). The majority of sequences from the most abundant *X. testudinaria* taxa (>1% of relative abundance) were at the left side of the MCE ordination (Fig. 3, Table S6, Supporting information). As expected, the sequences were similar to SC/SCC sequences and to *X. testudinaria* reference sequences.

Transcriptional activities

To explore the relationship between presence and activity, the taxon abundances were compared between DNA and cDNA data sets. In general, the relative abundance of taxa correlated well between data sets, as indicated by high *R*-squared values (0.79 for sea water, 0.87 for *S. carteri* and 0.82 for *X. testudinaria*; $P < 0.01$; Fig. 5). Indeed, only few taxa had significantly different cDNA/DNA ratios (see Materials and methods for selection criteria), of which 15 were in sea water, 14 in *S. carteri* and 19 in *X. testudinaria* (Table S7, Supporting information).

In sea water, two taxa showed significantly higher relative abundances in cDNA than in DNA data sets, these being *Synechococcus* (76.6% in cDNA data sets) and the unclassified lineage(s) of Subsection I of *Cyanobacteria* (0.84%). Another highly abundant taxon, the proteobacterial SAR324 clade (Marine group B), was over-represented in cDNA data sets (2.96%); however, the variation between cDNA and DNA data sets was not significant (Table S7, Supporting information). In *S. carteri*, only *Synechococcus* showed a significantly higher relative abundance in cDNA than in DNA data sets (28.7% mean sequence abundance in cDNA data sets). In *S. carteri*, the E01-9C-26 marine group (45.3%) and *Nitrospira* (4.7%) were over-represented in cDNA data sets, although the variation between cDNA and



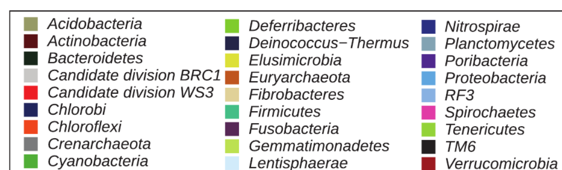


Fig. 5 Relative taxon abundance in amplicon data sets from DNA vs. cDNA plotted in logarithmic scale for (A) sea water, (B) *Stylissa carteri* and (C) *Xestospongia testudinaria* (all $n = 3$, but *S. carteri* $n = 2$). The 16S rRNA sequences were grouped based on the deepest annotated taxon. Taxa with statistically significant variation between DNA and cDNA are represented as diamonds (see Materials and methods for selection criteria). The line represents a 1:1 ratio.

© 2013 John Wiley & Sons Ltd

DNA data sets was not significant (Table S7, Supporting information). In *X. testudinaria*, the *Chloroflexi* SAR202 clade (35.3% in cDNA data sets), *Poribacteria* (8.5%), *Candidatus* Entotheonella (6.1%), *Nitrosococcus* (5.3%), *Spirochaeta* (3.5%), the alphaproteobacterial clade TK34 (2.3%) and the unclassified lineage(s) of the *Gammaproteobacteria* (1.9%) showed significantly higher relative abundances in cDNA than in DNA data sets. Other highly abundant taxa in *X. testudinaria* were over-represented in cDNA data sets, but the variation between cDNA and DNA data sets was not significant (Table S7, Supporting information). These included an uncultured bacteria lineage from the proteobacterial *Desulfurellaceae* family (3.6% in cDNA data sets), the *Defluviicoccus* genus of *Proteobacteria* (4.4%), the GR-WP33-30 clade of *Proteobacteria* (1.3%) and the unclassified lineage(s) of the *Alphaproteobacteria* (1.6%). All other taxa were more abundant in DNA than in cDNA data sets or did not present significant variation between data sets (Table S7, Supporting information).

Quantification of poribacterial 16S rRNA genes (DNA) and transcripts (cDNA) was carried out by qPCR (Fig. 6A). Poribacteria were chosen as an example of an abundant and transcriptionally active taxon of *X. testudinaria* (Fig. 4, Table S7, Supporting information). According to the qPCR estimations on DNA, *X. testudinaria* showed higher 16S rRNA gene copy numbers [$(1.15 \pm 0.85) \times 10^{11}$ copies/g of sponge] than *S. carteri* [$(1.15 \pm 0.96) \times 10^{7}$ copies/g of sponge] or sea water (79.1 ± 47.9 copies/mL; Fig. 6). The mean copy number of the poribacterial 16S rRNA gene in *X. testudinaria* was significantly different only from that of sea water ($P < 0.05$). The transcriptional activity of Poribacteria was assessed by the 16S rRNA transcript/gene ratio. Accordingly, Poribacteria were more active in *X. testudinaria* (377.3 ± 409.5 ratio) than in sea water (58 ± 16.3 ratio) or *S. carteri* (125.3 ± 32.4 ratio), although the difference was only significant in comparison with sea water (Fig. 6 B, $P < 0.05$).

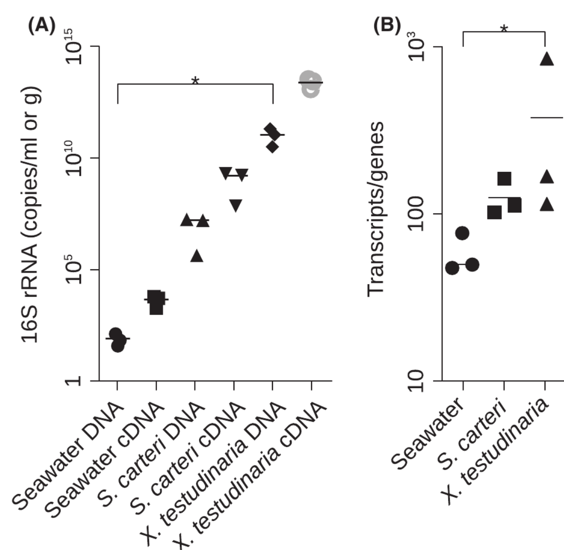


Fig. 6 Presence and activity of members of the candidate phylum *Poribacteria* in sea water, *Stylissa carteri* and *Xestospongia testudinaria*. (A) 16S rRNA copy numbers in DNA and cDNA were estimated by a duplicated qPCR assay using *Poribacteria*-specific primers. Copy numbers are presented for sea water (per mL) and sponge (per g, wet weight) samples. (B) The activity of Poribacteria was inferred calculating the 16S rRNA transcript/gene ratio. Gene copies and transcript/gene ratios were compared among sources using the nonparametric Kruskal–Wallis test followed by Dunn's multiple comparisons test (*$P < 0.05$).

Discussion

Massively parallel pyrosequencing has revolutionized microbial ecology (Sogin *et al.* 2006; Petrosino *et al.* 2009; Kuczynski *et al.* 2012). About a dozen articles have applied this technology to characterize the microbial diversity of marine sponges (Webster *et al.* 2010; Lee *et al.* 2011; Jackson *et al.* 2012; Schmitt *et al.* 2012a,b; Simister *et al.* 2012a,b; White *et al.* 2012; Arellano *et al.* 2013). Here, we employ amplicon pyrosequencing to revisit the concept of HMA and LMA sponges using *Xestospongia testudinaria* (HMA) and *Stylissa carteri* (LMA) as experimental models. Microbial diversity analyses of these sponge species were previously conducted using clone libraries and amplicon pyrosequencing (Lee *et al.* 2011; Montalvo & Hill 2011; Webster *et al.* 2012; Giles *et al.* 2013). These species were chosen here, because they co-inhabit the same reef location, because their status as HMA or LMA is well documented and because we could build on an existing data set, thus giving a perspective into the permanency of the association.

The statistical indicators of diversity are frequently used as a first basis to identify differences in ecological as well as microbial community composition between different samples. One of the most fundamental diversity measures is species richness, which is defined as the number of species found in a given community (Peet 1974; Colwell 2009). In addition, diversity is commonly measured by diversity indices that combine species richness and evenness, taking the relative abundances of species into account (Hurlbert 1971; Colwell 2009). In our study, the sea water microbiota was richest (in that it contained most OTUs); however, the microbiota of *X. testudinaria* was most diverse (in that more OTUs were more evenly distributed). The predicted richness based on Chao1 and ACE estimators resembled previous reports (Lee *et al.* 2011; Montalvo & Hill 2011;

© 2013 John Wiley & Sons Ltd

Ngugi *et al.* 2012), but differed from the predicted higher values for *X. testudinaria* (Lee *et al.* 2011). These differences could conceivably result from the use of different primers and the use of different bioinformatic pipelines. In this context, it is noteworthy that archaea were poorly recovered in our study, unlike in Lee *et al.* (2011), where archaea were abundant in the amplicon libraries.

With the power of deep-sequencing methodologies, it has become possible to obtain a more comprehensive picture of microbial diversity than was possible by clone library construction and Sanger sequencing just some years ago. In consequence, comprehensive 454 sequence data have shown that many of the so-called sponge-specific bacteria do occur in sea water, albeit at very low abundances (Pham *et al.* 2008; Webster *et al.* 2010; Lee *et al.* 2011; Taylor *et al.* 2013). Here, we expand these findings by showing that sponge-specific bacteria are also shared between high and low microbial abundance sponges, in addition to being shared with sea water. Phylogenetic trees from clone library data have already revealed the occasional sequence from an LMA sponge in an SC derived solely from HMA sponges and *vice versa* (Giles *et al.* 2013). In the present study, >99% of all sequences (equalling 836 OTUs) were shared between two sources and >90% of all sequences (equalling 204 OTUs) were shared between all three sources (Fig. 2). However, the numbers of sequences shared between the sources were very different depending on the OTU in question. For example, what was abundant in one source would be considered part of the rare biosphere in the other. In general, it is important to realize that the percentage of unique sequences decreased with increasing sample size (Fig. 2C). In other words, 'everything is found everywhere' (Whitfield 2005) with increasing sequencing depth. The sequencing depth has thus profound impacts on our perception of host specificity.

Our findings along with several other recent publications demonstrating the presence of sponge-specific sequences in the environment (Pham *et al.* 2008; Webster *et al.* 2010; Lee *et al.* 2011; Taylor *et al.* 2013) necessitate a revision of the sponge-specific clusters concept that was originally formulated based on phylogenetic analysis of clone library data (Hentschel *et al.* 2002). Amplicon data, as a consequence of the superior sequencing depth, show rather an enrichment of sequences in sponges in comparison with other habitats. We therefore propose to use the term 'sponge-enriched' or 'sequences that are enriched in sponges' rather than 'sponge-specific' when discussing aspects of host specificity based on amplicon data.

Sponges are by no means sealed entities, and a certain amount of exchange of microorganisms with the environment is not surprising. Microbial release from sponges is conceivable after physical damage caused by storms, wounds caused by predation and specimen collapse resulting from disease [such as during fatal orange band disease of *Xestospongia muta* (Angermeier *et al.* 2011)]. In this context, it is interesting to note that the third sea water replicate that was taken after the sponge samples were collected also contained the highest amount of sponge-specific clades, suggesting symbiont leakage into sea water resulting from tissue injury associated with sampling (Fig. 3B). Furthermore, the expulsion of reproductive material during spawning may result in the leakage of symbionts into the surrounding sea water (Gloeckner *et al.* 2013b). Whether a microbial symbiont can occasionally re-enter the sponge has not yet been investigated but this appears conceivable considering the close proximity of HMA and LMA sponge species on the reef and the very high filtration rates of sponges. Whether or not this process of symbiont mixing between sponges, and more generally horizontal acquisition of symbionts from sea water, is of relevance to sponge symbiosis provides an interesting direction for future research.

The relative abundance and composition of phyla in *X. testudinaria* was consistent with previous reports (Lee *et al.* 2011; Montalvo & Hill 2011; Webster *et al.* 2012), where about 70% of 16S rRNA gene sequences were similar to those recovered from other *X. testudinaria* individuals (Montalvo & Hill 2011). About 60% of sequences, among them the most abundant OTUs (Table 2, Table S6, Supporting information), fell into previously identified sponge-specific clusters or sponge- and coral-specific clusters (SC/SCC; Simister *et al.* 2012c). This finding suggests that the dominant OTUs do account for the sponge-specific communities. This specific association of *X. testudinaria* with its microbiota was confirmed by ordination of OTUs by MCE in a sequence similarity-independent, abundance-based approach (Fig. 3). Sequences representing sea water bacteria were only occasionally recovered, representing less than ~6% of total DNA sequences (Table S6, Supporting information).

The amplicon-sequence-derived profile of *S. carteri* was characteristic of LMA sponges in that comparably low phylum-level diversity was found with dominant phyla being *Proteobacteria*, *Cyanobacteria* and *Nitrospira* (Table 1, Fig. 4). At least 24% of *S. carteri* DNA sequences were derived from sea water, with an occasional presence of *X. testudinaria*-specific clades (Table S5, Supporting information). The identification of specific microbes in *S. carteri* was based on ordination of OTUs by MCE (Fig. 3A). This approach was taken because SC-/SCC-like microbes were present only at low abundances and because sequences previously

© 2013 John Wiley & Sons Ltd

recovered from *S. carteri* (Giles *et al*. 2013) were present in considerable amounts in sea water (Table S3, Supporting information). In consequence, a set of *S. carteri*-specific taxa was identified that was dominated by one specific clade of *Gammaproteobacteria*, E02-9C-26, and accounted for >42% of all amplicon sequences. This and other *S. carteri*-specific clades identified by MCE (a *Nitrospira* clade and an unclassified lineage(s) of *Proteobacteria*, Table 2) were previously recovered from *S. carteri* from the same collection site (Giles *et al*. 2013). This observation lends support to the hypothesis that these clades may be permanently associated with *S. carteri*. Further support for this hypothesis comes from a recent publication which showed that Great Barrier Reef sponges, which are devoid of SC/SCCs, are still capable of maintaining permanent and highly structured microbial communities (Webster *et al*. 2012).

Differences between taxon abundances in DNA and cDNA are generally regarded as a reflection of the differences in metabolic rates between the taxa (Campbell *et al*. 2011; Lanzen *et al*. 2011). Despite some caveats (i.e. differences in 16S rRNA gene copy numbers between species, different transcriptional activities between species at the same growth stage (Campbell *et al*. 2011; Simister *et al*. 2012a), this approach has proven to be of value in sponge microbiology (Kamke *et al*. 2010; Simister *et al*. 2012a). In our study, abundance of different phylotypes inferred from transcripts (cDNA) correlated generally well with the abundance of the corresponding phylotypes (DNA; Fig. 4). *Synechococcus* was by far the most transcriptionally active clade in sea water and also in *S. carteri*. However, due to the lack of resolution on the short 16S rRNA amplicon sequence fragment, we could not clarify if free-living or sponge-specific *Synechococcus* were active in *S. carteri*. In this context, it is interesting to note that a *Stylissa* sp. 445 sponge collected from the Great Barrier Reef was also dominated by *Synechococcus*, as identified by phylogenetic analysis of metagenomic shotgun sequence data (Fan *et al*. 2012). Cyanobacterial transcriptional activity is consistent with the fact that the samples were collected at daytime. On the contrary, in *X. testudinaria*, the typical HMA sponge-associated clades (e.g. *Chloroflexi*, *Poribacteria*, *Candidatus* Entotheonella) displayed significantly higher relative abundances in cDNA over DNA data sets. The over-representation in cDNA over DNA data sets was previously reported for some of these clades, such as those belonging to *Spirochaetes* and *Poribacteria* phyla (Kamke *et al*. 2010; Simister *et al*. 2012a). Experimental qPCR was performed to validate the presence of Poribacteria outside of *X. testudinaria* and to evaluate the activity of this clade outside of the host sponge context. This clade was chosen because it represents a typical sponge symbiont. Our results show that Poribacteria are transcriptionally much less active in sea water in comparison with *X. testudinaria*. Assuming that the variation was not due to extraction bias (von Wintzingerode *et al*. 1997), our result would suggest that although sponge-associated clades are present in sea water in very low concentrations, they may not be metabolically active in sea water and their release from the host may be accidental. This hypothesis is supported by our observation that the sea water collected after sponge collection (WTD3) contained noticeably more sponge-associated lineages than the two water samples collected prior and during sponge collection (WTD1 and WTD2).

Conclusions

The HMA vs. LMA dichotomy was supported using amplicon pyrosequencing. While the microbiome of *Stylissa carteri* was dominated by *Proteobacteria* and *Cyanobacteria*, the microbiome of *Xestospongia testudinaria* contained *Chloroflexi*, *Poribacteria* and *Deferribacteres*, among several others, as indicator phyla. The *S. carteri* data set contained further a higher proportion of possibly sea water-derived bacterial sequences (24%) than the *X. testudinaria* data set (6%). The percentage of sea water bacteria in sponge microbiomes is reported here for the first time, and this number should be taken into consideration when addressing aspects of dynamics and specificity. In spite of this, *S. carteri* clearly harbours a distinct microbiota, which accounts for more than 75% of the sequences recovered. The common notion that LMA sponges contain merely sea water bacteria is thus incorrect. From a methodological point of view, specificity was only revealed upon data analyses of OTU abundance patterns (by MCE ordination), and neither a sequence similarity analysis nor a presence/absence analysis would have been sufficient to unravel this pattern. Pyrosequencing of 16S rRNA genes and transcripts showed that activity was well correlated with the presence of the respective OTUs. Finally, qPCR data verified the presence of *Poribacteria*, representing characteristic symbionts of HMA sponges, also in an LMA sponge and sea water, albeit in much lower concentrations. Although sponge-specific microbes are consistently found in sea water, the transcript/gene ratios indicate that they may not be metabolically active outside the sponge context.

Acknowledgements

We thank the Coastal and Marine Resources Core Lab and the Biosciences Core Laboratory at KAUST. We acknowledge Susanne Schmitt for suggestion of PCR primers and Eva Reisberg for interesting discussions. The insightful comments of three anonymous reviewers have improved the quality of our

© 2013 John Wiley & Sons Ltd

manuscript. LM-S was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Wuerzburg. Additional financial support was provided by KAUST baseline funds to TR.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology*, **215**, 403–410.

Angermeier H, Kamke J, Abdelmohsen UR *et al.* (2011) The pathology of sponge orange band disease affecting the Caribbean barrel sponge *Xestospongia muta*. *FEMS Microbiology Ecology*, **75**, 218–230.

Arellano SM, Lee OO, Lafi FF *et al.* (2013) Deep Sequencing of *Myxilla* (*Ectyomyxilla*) *methanophila*, an epibiotic sponge on cold-seep tubeworms, reveals methylotrophic, thiotrophic, and putative hydrocarbon-degrading microbial associations. *Microbial Ecology*, **65**, 450–461.

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

Bray JR, Curtis JT (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecological Monographs*, **27**, 325–349.

Campbell BJ, Yu L, Heidelberg JF, Kirchman DL (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*, **108**, 12776–12781.

Cannistraci CV, Ravasi T, Montevecchi FM, Ideker T, Alessio M (2010) Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes. *Bioinformatics*, **26**, i531–i539.

Caporaso JG, Bittinger K, Bushman FD *et al.* (2010a) PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, **26**, 266–267.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, **7**, 335–336.

Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*, **11**, 265–270.

Chao A, Lee S-M (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*, **87**, 210–217.

Colwell RK (2009) Biodiversity: concepts, patterns, and measurement. In: *The Princeton Guide to Ecology* (ed Levin SA), pp. 257–263. Princeton Univ. Press, Princeton, New Jersey.

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

Erwin PM, Olson JB, Thacker RW (2011) Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PLoS ONE*, **6**, e26806.

Fan L, Reynolds D, Liu M *et al.* (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the United States of America*, **109**, E1878–E1887.

Fieseler L, Horn M, Wagner M, Hentschel U (2004) Discovery of the novel candidate phylum "*Poribacteria*" in marine sponges. *Applied and Environment Microbiology*, **70**, 3724–3732.

Gihring TM, Green SJ, Schadt CW (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology*, **14**, 285–290.

Giles EC, Kamke J, Moitinho-Silva L *et al.* (2013) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*, **83**, 232–241.

Gloeckner V, Hentschel U, Ereskovsky A, Schmitt S (2013a) Unique and species-specific microbial communities in *Oscarella lobularis* and other Mediterranean *Oscarella* species (Porifera: Homoscleromorpha). *Marine Biology*, **160**, 781–791.

Gloeckner V, Lindquist N, Schmitt S, Hentschel U (2013b) *Ectyoplasia ferox*, an experimentally tractable model for vertical microbial transmission in marine sponges. *Microbial Ecology*, **65**, 462–474.

Gonzalez JM, Zimmermann J, Saiz-Jimenez C (2005) Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics*, **21**, 333–337.

Haas BJ, Gevers D, Earl AM *et al.* (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, **21**, 494–504.

Hentschel U, Hopke J, Horn M *et al.* (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Applied and Environment Microbiology*, **68**, 4431–4440.

Hentschel U, Fieseler L, Wehrl M *et al.* (2003) Microbial diversity of marine sponges. *Progress in Molecular and Subcellular Biology*, **37**, 59–88.

Hentschel U, Usher KM, Taylor MW (2006) Marine sponges as microbial fermenters. *FEMS Microbiology Ecology*, **55**, 167–177.

Hentschel U, Piel J, Degnan SM, Taylor MW (2012) Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*, **10**, 641–654.

Hurlbert SH (1971) The nonconcept of species diversity: a critique and alternative parameters. *Ecology*, **52**, 577–586.

Jackson SA, Kennedy J, Morrissey JP, O'Gara F, Dobson AD (2012) Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters. *Microbial Ecology*, **64**, 105–116.

Kamke J, Taylor MW, Schmitt S (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME Journal*, **4**, 498–508.

Kuczynski J, Lauber CL, Walters WA *et al.* (2012) Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*, **13**, 47–58.

Lanzen A, Jorgensen SL, Bengtsson MM *et al.* (2011) Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiology Ecology*, **77**, 577–589.

Lee OO, Wang Y, Yang J *et al.* (2011) Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *ISME Journal*, **5**, 650–664.

Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environment Microbiology*, **71**, 8228–8235.

Montalvo NF, Hill RT (2011) Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts. *Applied and Environment Microbiology*, **77**, 7207–7216.

© 2013 John Wiley & Sons Ltd

Ngugi DK, Antunes A, Brune A, Stingl U (2012) Biogeography of pelagic bacterioplankton across an antagonistic temperature-salinity gradient in the Red Sea. *Molecular Ecology*, **21**, 388–405.

Oksanen J, Blanchet FG, Kindt R *et al.* (2012) *Vegan: Community Ecology Package*. Available from http://CRAN.R-project.org/package=vegan.

Parks DH, Beiko RG (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*, **26**, 715–721.

Peet RK (1974) The measurement of species diversity. *Annual Review of Ecology and Systematics*, **5**, 285–307.

Petrosino JF, Highlander S, Luna RA, Gibbs RA, Versalovic J (2009) Metagenomic pyrosequencing and microbial identification. *Clinical Chemistry*, **55**, 856–866.

Pham VD, Konstantinidis KT, Palden T, DeLong EF (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. *Environmental Microbiology*, **10**, 2313–2330.

Pruesse E, Quast C, Knittel K *et al.* (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*, **35**, 7188–7196.

Rivals I, Personnaz L, Taing L, Potier MC (2007) Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, **23**, 401–407.

Schloss PD, Westcott SL, Ryabin T *et al.* (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environment Microbiology*, **75**, 7537–7541.

Schmitt S, Wehrl M, Lindquist N, Weisz JB, Hentschel U (2007) Morphological and molecular analyses of microorganisms in Caribbean reef adult sponges and in corresponding reproductive material. In: *Porifera Research: Biodiversity, Innovation and Sustainability* (eds Custódio MR, Lôbo-Hajdu G, Hajdu E & Muricy G), pp. 561–568. Museu Nacional, Rio de Janeiro.

Schmitt S, Hentschel U, Taylor M (2012a) Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia*, **687**, 341–351.

Schmitt S, Tsai P, Bell J *et al.* (2012b) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME Journal*, **6**, 564–576.

Shannon CE (1948) A mathematical theory of communication. *Bell System Technical Journal*, **27**, 379–423.

Simister R, Taylor MW, Tsai P *et al.* (2012a) Thermal stress responses in the bacterial biosphere of the Great Barrier Reef sponge, *Rhopaloeides odorabile*. *Environmental Microbiology*, **14**, 3232–3246.

Simister R, Taylor MW, Tsai P, Webster N (2012b) Sponge-microbe associations survive high nutrients and temperatures. *PLoS ONE*, **7**, e52220.

Simister RL, Deines P, Botte ES, Webster NS, Taylor MW (2012c) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environmental Microbiology*, **14**, 517–524.

Simpson EH (1949) Measurement of diversity. *Nature*, **163**, 688–688.

Smyth GK (2005) Limma: linear models for microarray data. In: *Bioinformatics and Computational Biology Solutions using R and Bioconductor* (eds Gentleman R, Carey V, Dudoit S, Irizarry R & Huber W), pp. 397–420. Springer, New York.

Sogin ML, Morrison HG, Huber JA *et al.* (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 12115–12120.

Taylor MW, Radax R, Steger D, Wagner M (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*, **71**, 295–347.

Taylor MW, Tsai P, Simister RL *et al.* (2013) 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME Journal*, **7**, 438–443.

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.

Walters WA, Caporaso JG, Lauber CL *et al.* (2011) PrimerProspector: *de novo* design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*, **27**, 1159–1161.

Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environment Microbiology*, **73**, 5261–5267.

Webster NS, Taylor MW, Behnam F *et al.* (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*, **12**, 2070–2082.

Webster NS, Luter HM, Soo RM *et al.* (2012) Same, same but different: symbiotic bacterial associations in GBR sponges. *Frontiers in Microbiology*, **3**, 444.

Weisz J, Hentschel U, Lindquist N, Martens C (2007) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Marine Biology*, **152**, 475–483.

White JR, Patel J, Ottesen A *et al.* (2012) Pyrosequencing of bacterial symbionts within *Axinella corrugata* sponges: diversity and seasonal variability. *PLoS ONE*, **7**, e38204.

Whitfield J (2005) Biogeography. Is everything everywhere? *Science*, **310**, 960–961.

von Wintzingerode F, Göbel UB, Stackebrandt E (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews*, **21**, 213–229.

L.M.-S., T.R. and U.H. designed the study. L.M.-S., E.C.G. and T.R. collected samples. L.M.-S., K.B. and E.C.G. performed laboratory research. L.M.-S., C.V.C., T.R. and L.S. performed bioinformatic analyses. L.M.-S. and U.H. wrote the manuscript. All authors contributed to the writing of the manuscript.

Data accessibility

Raw pyrosequencing reads were deposited under the NCBI SRA accession number SRP017932.

© 2013 John Wiley & Sons Ltd

Representative sequences of quality-controlled OTUs are accessible as Supporting information (Appendix S5).

Quality-controlled OTU table and the rarefied quality-controlled OTU table are accessible as Supporting informations (Appendices S3 and S4).

Alignment of representative sequences that are present in the rarefied OTU table and phylogenetic tree used for UniFrac analysis are accessible as Supporting informations (Appendices S1 and S2).

Quantitative PCR results as well as nucleic acid extraction yields are accessible as Supporting informations (Appendices S6 and S7).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Rarefaction curves of 16S rRNA amplicon data sets.

Fig. S2 OTU ordination and assignment to SC/SCC and reference sequences.

Table S1 SC/SCC [1] taxa not used as references for assignment of 16S rRNA amplicons.

Table S2 Observed number of operational taxonomic units (OTUs) from rarified datasets, and estimations of richness (Chao1, ACE) and diversity (Shannon, Simpson).

Table S3 Percentage of SC/SCC and previous studies' sequences in datasets.

Table S4 The most abundant taxa in DNA datasets of sea water.

Table S5 The most abundant taxa in DNA datasets of *Stylissa carteri*.

Table S6 The most abundant taxa in DNA datasets of *Xestospongia testudinaria*.

Table S7 Transcriptional activity of taxa in DNA in sea water, *Stylissa carteri* and *Xestospongia testudinaria*.

Appendix S1 Alignment of representative sequences from quality-controlled, rarified OTUs.

Appendix S2 Phylogenetic tree of representative sequences from quality-controlled, rarified OTUs.

Appendix S3 Tab-delimited file of quality-controlled OTUs.

Appendix S4 Tab-delimited file of quality-controlled, rarified OTUs.

Appendix S5 Representative sequences of quality-controlled OTUs.

Appendix S6 Extraction yields of nucleic acids (ng) from volume (mL) of sea water or wet weight (g) of sponges.

Appendix S7 Quantitative PCR estimation of poribacterial 16S rRNA genes (DNA) and transcripts (cDNA) in sea water, *Stylissa carteri*, and *Xestospongia testudinaria*.

© 2013 John Wiley & Sons Ltd

4.4 GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater.

This manuscript was submitted to **FEMS Microbiology Ecology** for publication and a reprint is used here with the permission from the authors. This article's reference is:

Bayer, K.[#], **Moitinho-Silva, L.[#]**, Brümmer, F., Cannistraci, C.V., Ravasi, T., Hentschel, U. (2014) GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater. *FEMS Microbiology Ecology*. Submitted.

[#]Authors contributed equally to this manuscript.

The documentation of author's contributions is found in the Appendix. The supplementary material is found in the CD attached to this PhD Thesis.

Running title: Functional gene repertoire of sponge microorganisms

GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater

Kristina Bayer[1,#], Lucas Moitinho-Silva[1,#], Franz Brümmer[2], Carlo V. Cannistraci[3,4], Timothy Ravasi[3,*] and Ute Hentschel[1,*]

[#]authors contributed equally to this work

[1]Dept. of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany

[2]Biological Institute, Department of Zoology, University of Stuttgart, Stuttgart, Germany

[3]Division of Biological and Environmental Sciences & Engineering and Division of Applied Mathematics and Computer Science, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia

[4]Biomedical Cybernetics Group, Biotechnology Center (BIOTEC), Technische Universität Dresden, Dresden, Germany

*Please address correspondence to: U. Hentschel; Email: ute.hentschel@uni-wuerzburg.de; Tel.: +49-931-31-82581; Fax: +49-931-31-86235; T. Ravasi; Email: timothy.ravasi@kaust.edu.sa; Tel.: +966-2-808-2387; Fax: +966-2-8020127

Keywords: GeoChip, amplicon sequencing, marine sponge, Porifera, functional gene, symbiosis

4.4.1 Abstract

The GeoChip 4 gene array was employed to interrogate the microbial functional gene repertoire of sponges and seawater collected from the Red Sea and the Mediterranean. Complementary amplicon sequencing confirmed the microbial community composition characteristic of high microbial abundance (HMA) and low microbial abundance (LMA) sponges. By use of GeoChip, altogether 20,273 probes encoding for 627 functional genes and representing 16 gene categories were positively identified. Minimum curvilinear embedding (MCE) analyses revealed a clear separation between the samples. The HMA/LMA dichotomy was stronger than any possible geographic pattern, which is shown here for the first time on the level of functional genes. However upon inspection of individual genes, very few specific differences were discernible. These differences were related to microbial ammonia oxidation, ammonification, and archaeal autotrophic carbon fixation (higher gene abundance in sponges over seawater) as well as denitrification and radiation-stress related genes (lower gene abundance in sponges over seawater). Except for few documented specific differences, the functional gene repertoire between the different sources appeared largely similar. This study expands previous reports in that functional gene convergence is not only reported between HMA and LMA sponges but also between sponges and seawater.

4.4.2 Introduction

Marine sponges (phylum Porifera) are important couplers of the benthic and pelagic ecosystems by virtue of their immense filter-feeding capacities (Bell, 2008, de Goeij *et al.*, 2013). Many marine sponges (hereafter referred to as high microbial abundance sponges, HMA) host stable associations of symbiotic microbial consortia within their mesohyl matrix that can contribute up to 35% of the animal's biomass (Hentschel *et al.*, 2012). Representatives of more than 28 bacterial phyla including new candidate phyla such as Poribacteria and Tectomicrobia and two archaeal lineages were identified in marine sponges so far (Webster *et al.*, 2010, Lee *et al.*, 2011, Schmitt *et al.*, 2012, Schmitt *et al.*, 2012,

Simister *et al.*, 2012, Wilson *et al.*, 2014). On the other hand, some sponge species are essentially devoid of microorganisms (hereafter termed low microbial abundance sponges, LMA) and the diversity is largely restricted to Proteobacteria (Alpha-, Gamma-,), Cyanobacteria, and Archaea. The classification into these HMA or LMA sponges has long been recognized (Vacelet & Donadey, 1977, Reiswig, 1981, Hentschel *et al.*, 2003).

Our understanding of microbial functions in sponge symbiosis is still at the beginning due to the fact that the vast majority of sponge-associated microorganisms have not been cultured and due to the lack of an established model system for sponge-microbial symbioses. However, as reviewed (Taylor *et al.*, 2007, Hentschel *et al.*, 2012, Webster & Taylor, 2012), several publications have analysed or inferred metabolic activities of the sponge symbionts, for example as derived from pure-culture studies, analysis of pathways by measurement of specific products and genes, and inference of metabolic potentials based on 16S rRNA gene sequence analysis. A particular emphasis has been placed on the metabolism of carbon, nitrogen, and sulfur as well as secondary metabolism. Recently, the application of “omics” approaches has contributed to build an integrated view of microbial function in the context of the host sponge. A recent metagenomic study reported functional gene convergence in the microbiome of six demosponge species (Fan *et al.*, 2012). Several metatranscriptomic/ metaproteomic studies have provided first insights into the expressed functional gene repertoire of sponge-associated microbial consortia (Liu *et al.*, 2012, Radax *et al.*, 2012, Moitinho-Silva *et al.*, 2014). Thirdly, single cell genomic studies have enabled to correlate specific traits with specific symbiont lineages, thus providing a thought-after link between phylogeny and function (Siegl *et al.*, 2011, Kamke *et al.*, 2013, Wilson *et al.*, 2014). However, much remains to be learned about the functional basis of sponge-microbe interactions.

Functional gene arrays, such as the GeoChip, are one efficient way to assess the functional gene repertoire of environmental microbial samples (Nostrand *et al.*, 2012). Microarrays have been developed to probe a wide diversity of genes in various ecological contexts. For example, functional gene arrays have been useful to establish gene inventories (Mason *et al.*, 2010, Chan *et al.*, 2013), to assess differences between field sites (Kang *et al.*, 2013), and to report shifts in response to environmental parameters such as warming (Yergeau *et al.*, 2012), or following the deep-water Horizon oil spill (Lu *et al.*, 2012). The

GeoChip has so far been applied to natural soils and sediments, and in particular in the context of contamination and bioremediation. Since its invention in 2004, the GeoChip has been constantly improved and updated. The newest version, GeoChip 4, contains 83,992 probes targeting 152,414 genes representing a total 410 gene categories (Tu *et al.*, 2014). In comparison to previous versions, the categories stress response, antibiotic resistance, and bacteriophage genes have been added (Nostrand *et al.*, 2012). In the present study, we employed the GeoChip 4, to our knowledge for the first time, to assess the functional gene repertoire of marine sponge-associated microbiomes.

4.4.3 Methods

Sample collection

Five individuals of the Mediterranean sponges *Aplysina aerophoba* (HMA) and *Dysidea avara* (LMA) were collected in April 2012 by scuba diving at a depth of 10-15 m off the coast at Rovinj, Croatia (45°08'N; 13°64'E). The sponges *Xestospongia testudinaria* (HMA) and *Stylissa carteri* (LMA), (n=5 each) were collected by scuba diving at Fsar reef (22°23'N; 39°03'E) at a depth of 13-15 m off the coast of Thuwal, Saudi Arabia in March 2012. The animals were brought to the surface in plastic bags to avoid contact with air. Small tissue pieces were removed with a sterile scalpel, rinsed in 0.2 µm filtered seawater, immediately frozen in liquid nitrogen and stored at -80 °C until use. Sponge pieces were taken such that portions of pinacoderm and mesohyl were included in each sample. Seawater was sampled using folded plastic canisters (camping store), which were filled during the dives in the vicinity of the sponges. Approximately 7 L of Mediterranean seawater were filtered onto 0.22 µm PES bottle top filters (Merck-Millipore, Germany) using a manually operated vacuum pump. The same volume of Red Sea seawater was filtered onto 0.22 µm hydrophilic Durapore membrane filters (Millipore, USA) using a Masterflex Easy-Load peristaltic pump (Cole-Parmer, USA). The filters were stored at -80 °C until further processing.

DNA extraction

Frozen sponge samples were ground to fine powder using a sterile mortar and pistil. Genomic DNA was extracted using the Allprep DNA/ RNA mini kit (Qiagen, Germany) following the manufacturer's instructions. DNA quality and concentrations were assessed with the NanoDrop 2000c spectrophotometer (Peqlab, Germany) as well as using the Qubit 2.0 fluorometer and the Qubit dsDNA BR Assay Kit (Life Technologies, Germany). DNA was freeze-dried using the ALPHA 1-2 LD lyophilizer (Christ, Germany) and shipped to Glomics Inc. (Norman, USA), where the DNA was processed.

Amplicon sequencing

The 16S rRNA genes were amplified according to Moitinho-Silva *et al*. (2014) with minor modifications. Briefly, PCRs were performed in triplicates using multiplex identifier adaptor-ligated primers 533f (5'-GTGCCAGCAGCYGCGGTMA-3') and 907r (5'-CCGTCAATTMMYTTGAGTTT-3'), (Simister *et al.*, 2012). The sequence region between the 533f and 907r primers includes the hypervariable regions V4 and V5. PCR amplification was performed using the Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific, Germany) with buffer CG. The PCR protocol was as follows: initial denaturation at 98 °C for 3 min, followed by 30 cycles of denaturation (98 °C for 10 s), annealing (65 °C for 30 s), and extension (72 °C for 18 s). An additional elongation step (72 °C for 5 min) was performed at the end of the protocol. Amplified DNA was gel-purified, pooled at equimolar ratios and sequenced on a Roche 454 GS FLX Titanium platform at the KAUST Genomics Core Lab. Raw pyrosequencing reads were deposited under the NCBI SRA accession numbers SRP043983.

Processing of 454 sequences and taxonomic assignment

The 16S rRNA amplicon sequences were processed with the QIIME pipeline v 1.4 (Caporaso *et al.*, 2010) according to Moitinho-Silva *et al*. (2014). Briefly, denoised, quality-curated sequences from Mediterranean samples were pooled with publically available DNA sequences from Red Sea seawater, *X. testudinaria* and *S. carteri* (SRP017932, (Moitinho-Silva *et al.*, 2014). An additional, newly sequenced amplicon dataset of *S. carteri* DNA replicate 2 (ASD2, collected in Moitinho-Silva *et al*. 2014a) was included in the present study. Sequences with > 97% sequence similarity were clustered into operational taxonomic units (OTUs), taxonomically classified and filtered, excluding singletons, chimeras, and eukaryotic sequences.

Ecological indices of amplicon data

Ecological indices and estimators were calculated based on datasets rarefied to minimum number of sequences recovered among all datasets. Richness was estimated by chao1 and ACE (Chao, 1984, Chao A., 1992). Diversity was assessed using Simpson and Shannon indices (Shannon, 1948, Simpson, 1949). Ecological indices were calculated using vegan package v 2.0-4 (Oksanen, *et al.*, 2012) in R. UniFrac analysis (Lozupone & Knight, 2005) was performed for amplicon sequences using the QIIME pipeline (Caporaso *et al.*, 2010) and visualized as PCoA plot. Jackknife support was obtained based on 100 repetitions at 10,000 sequences per sample.

GeoChip 4 loading and data processing

800 ng of each sample was used for directly labeling and hybridized for 16.5 hours. Fluorescence-labeling, array hybridization and scanning were conducted as described (Lu *et al.*, 2012). Data normalization and preliminary analysis (including removal of poor-quality

spots, normalization of spot signal intensity of each spot by mean, removal of spots with low signal intensities based on the signal-to-noise ratio (SNR) as well as removal of outliers) was conducted as described (Wu *et al.*, 2006, He *et al.*, 2007). Probes were assigned as positive when at least three of five (or three of four for Mediterranean seawater) biological replicates showed a positive signal. The normalized signal intensities were calculated as the sum of all probes per gene, divided by the signal intensity of all probes per category, and then averaged across all replicates per sample. Statistical analysis was performed with the non-parametric Mann-Whitney U-test using GraphPad Prism version 6.01 for Windows (GraphPad Software, USA). GeoChip data were analysed in the R environment v 2.15.3 (The R Core Team, 2012).

Minimum curvilinear embedding

The Minimum Curvilinear embedding (MCE) (Cannistraci *et al.*, 2010) algorithm was used for unsupervised exploration and discrimination of samples and gene patterns. The same algorithm was already successfully adopted in environmental microbiology essentially for a similar task (Moitinho-Silva *et al.*, 2014). To explore the relation between samples, the Minimum Curvilinearity distance matrix was computed starting from a table containing gene signal intensities in each sample, which were calculated as the average of normalized probe signal intensities. Normalized probe signal intensities were calculated by dividing the signal intensity of each probe by the total signal intensity of the sample. To explore the relation between genes, the table was transposed and used as input to calculate distances. The Minimum Curvilinearity distance matrices were calculated using the following expression (Cannistraci *et al.*, 2010):

$$MC(x,y) = 1 - Pearson_correlation(Sample_x,\ Sample_y)$$

The calculated distance matrices were adopted for MCE dimension reduction. MCE analysis was performed using the Singular-Value-Decomposition-based algorithm published in Cannistraci *et al.* (2013). MCE analysis was performed both in MATLAB and R, offering the same results. Principal component analysis (PCA) was performed in MATLAB.

4.4.4 Results

Microbial diversity by amplicon sequencing

A total of 234,810 denoised, quality-controlled 16S rRNA gene sequences were generated from three biological replicates each of Mediterranean seawater, *D. avara*, *A. aerophoba*, and one biological replicate of *S. carteri* that had originally been collected (Moitinho-Silva *et al.*, 2014). These sequences were combined with already published 16S rRNA gene sequences from the Red Sea (seawater (n=3), *S. carteri* (n=2), *X. testudinaria* (n=3), (Moitinho-Silva, *et al.*, 2014). This effort resulted in 1,726 OTUs representing a total of 576,444 sequences of which 209,012 sequences were produced in this study.

Rarefaction curves revealed that samples were sequenced with different sampling depth (Suppl. Fig. 1A). Therefore, the datasets were rarefied to the minimum sampling effort (8,925 sequences per sample). Seawater displayed the highest richness at both locations. The term “richness” is based on the number of species in a given community and is judged from the estimators ACE and chao1) (Suppl. Fig. 1B). However, the HMA sponges showed the highest diversity. The term “diversity” takes the relative species abundances into account and is judged from Shannon and Simpson indices. The LMA sponges showed the lowest diversity at each location (Suppl. Fig. 1C).

The distance between the microbiomes under investigation were analyzed by UniFrac and visualized by Principle coordinates analysis (PCoA) plot (Fig. 10A). PCoA of amplicon sequences revealed a clear separation of the three sample types (Fig. 10A). According to PC1 (which explained 43.65 % of the variation), samples were ordered according to source (HMA, LMA, seawater) rather than to geographic location (Mediterranean, Red Sea). The LMA sponge samples (*D. avara* and *S. carteri*) presented a larger variation between sources and were placed between seawater and the HMA samples. PC2 (which explained 36.74 % of the variation) separated *D. avara* from the other sources.

A total of 28 bacterial, archaeal, and candidate phyla were detected in the present study (Fig. 10B, Suppl. Table 1). The microbial diversity of the Red Sea sponges was published in

Moitinho-Silva *et al*. (2014), but is included here for comparative purposes. The phylum level composition of the HMA sponge species *A. aerophoba* and *X. testudinaria* was complex and was composed of *Proteobacteria* (26.6 and 28.0%), *Chloroflexi* (22.9 and 34.5%), *Acidobacteria* (20.8 and 6.8%), *Deferribacteres* (8.7 and 11.7%), *Nitrospirae* (5.1 and 2.9%), *Gemmatimonadetes* (4.2 and 3.8%), *Actinobacteria* (2.9 and 4.6%), *Cyanobacteria* (2.1 and 0.27%), *Spirochaetes* (2.2 and 1.6%), the candidate phylum *Poribacteria* (2.7 and 2.2%), *Bacteroidetes* (1.0 and 1.7%), and *Planctomycetes* (0.02 and 1.4%). Each of the remaining 16 phyla contributed < 1% to the HMA sponge dataset. The LMA sponge species *D. avara* and *S. carteri* were dominated *Proteobacteria* (91.9 and 74.1%), *Cyanobacteria* (3.8 and 12.8%), and *Bacteriodetes* (2.9 and 6.6%) (Fig. 10B, Suppl. Table 1). It is noteworthy, that the proteobacterial community of *D. avara* was dominated by *Beta*- and *Alphaproteobacteria*, while *Gamma*- and *Alphaproteobacteria* were dominant in *S. carteri* (Suppl. Table 1). Each of the remaining 25 phyla contributed < 1% to the LMA sponge dataset. The seawater samples were dominated by *Proteobacteria* (64.3 in Mediterranean and 41.3% in Red Sea seawater), followed by *Cyanobacteria* (15.3 and 30.4%), *Bacteriodetes* (15.5 and 12.7%), *Actinobacteria* (1.3 and 9.8%) and *Verrucomicrobia* (2.3 and 0.08%). Each of the remaining 23 phyla contributed < 1% to the seawater dataset.

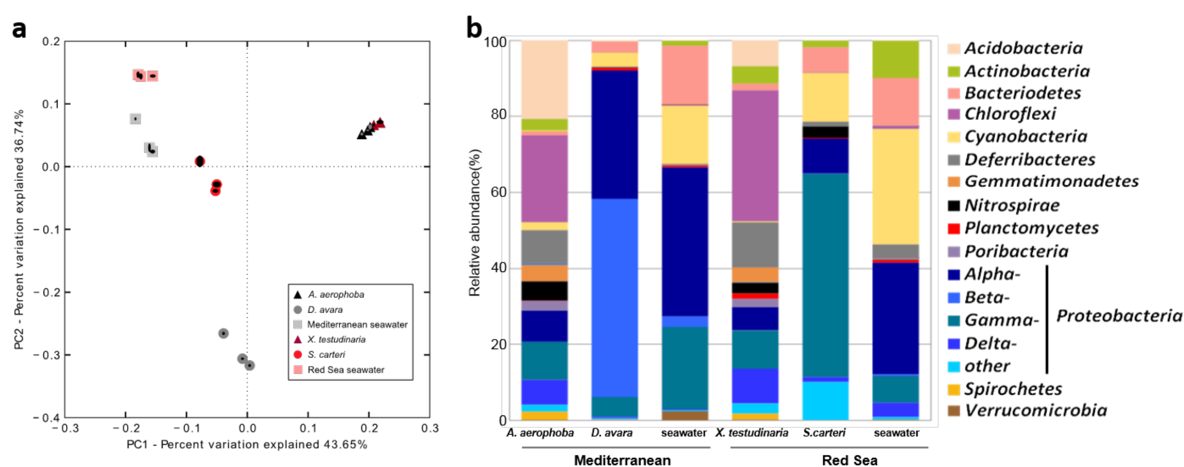


Figure 10. Microbial composition of HMA and LMA sponges as well as of seawater from the Mediterranean and the Red Sea collection site as obtained by 454 pyrosequencing. Relation among samples as performed by principle coordinates analysis (PCoA) based on UniFrac (A), taxon composition of samples (B). Phyla and major classes of Proteobacteria are shown for each source. Taxa which are present at > 1% abundance are shown. See Suppl. Table 1 for details of taxon composition.

Microbial functional analysis by GeoChip

Using the functional GeoChip array 4, a total of 20,273 probes were identified that represented 627 functional genes, grouped altogether in 16 functional gene categories in sponge and seawater samples (Suppl. Figure 1). The majority of positive gene probes belonged to the categories Organic remediation (4577 probes), Stress (4178 probes), Carbon cycling (2327 probes), Metal resistance (2297 probes), Nitrogen (1580 probes), as well as in lesser amounts (< 850 probes) to the categories Antibiotic resistance, Bacteria phage, Bioleaching, Energy Process, Fungi function, Others, Phosphorus, Soil Benefit, Soil borne pathogen, Sulfur, and Virulence (Suppl. Table 2).

The exploration of the samples based on signal intensities of genes by MCE revealed a clear separation (Fig. 11A). The first dimension (horizontal axis) was discriminative between the three sources (HMA, LMA, seawater) and within that, between the locations (Red Sea, Mediterranean) for the sponge samples. The second dimension separated the LMA from HMA and seawater samples. The exploration of signal intensities of genes by MCE (Fig. 11B) unveiled very few discernible patterns. Noticeably, the LMA sponges, in particular *D. avara*, lacked genes (indicated by dark blue boxes). The detected genes in *D. avara* resulted in higher normalized signal intensities. In conclusion, the dataset revealed few specific gene patterns rather than a general overarching trend.

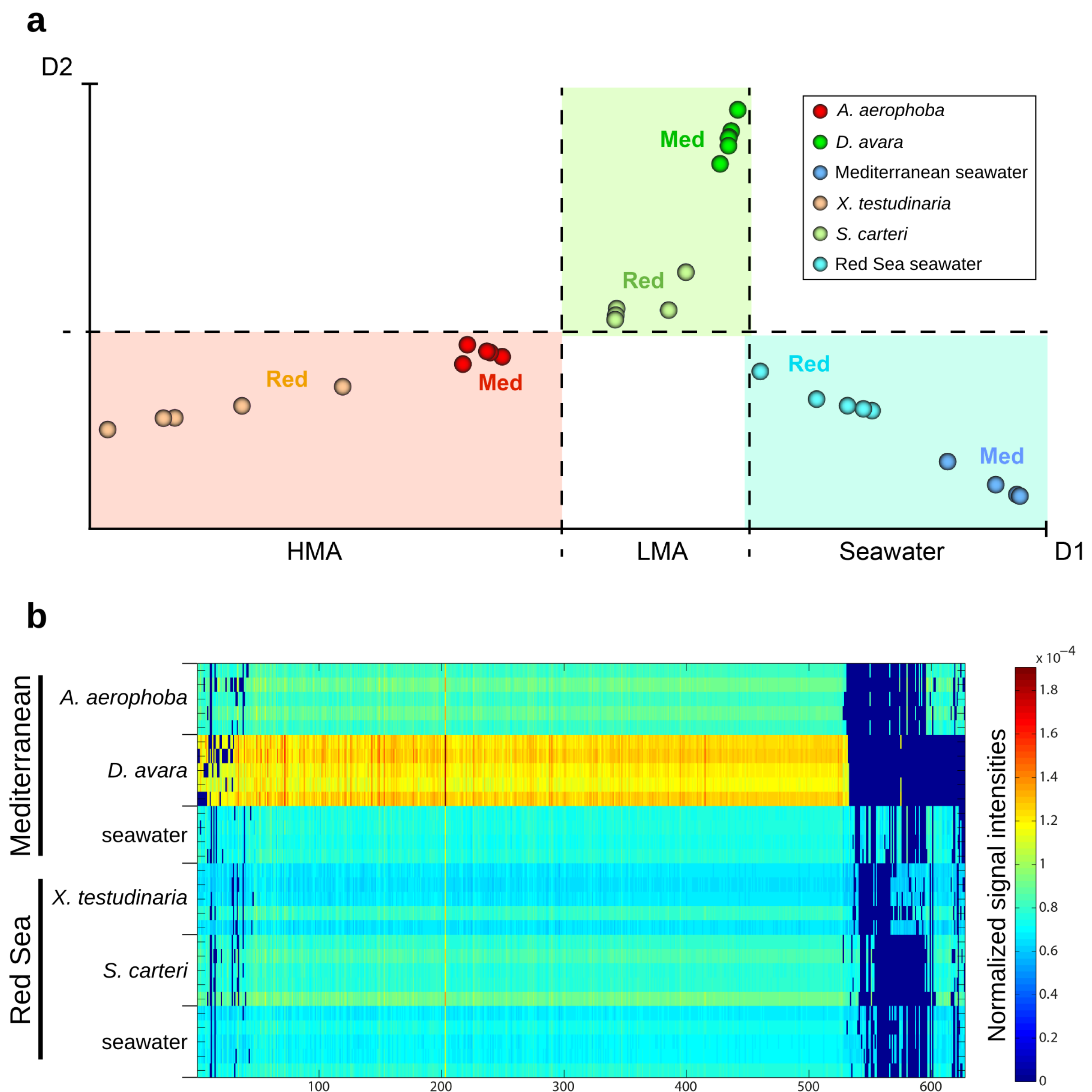


Figure 11. Exploration and visualization of the GeoChip 4 by Minimum Curvilinear Embedding (MCE). The reciprocal relations between the samples are mapped in a two-dimensional space by nonlinear dimension reduction performed by the MCE algorithm. Each spot represents a sample (A). The label "Red" indicates samples collected in the Red Sea; the label "Med" indicates samples collected in the Mediterranean Sea. Heatmap of functional genes' normalized signal intensities. Genes are ordered according to first dimension, D1, of MCE analysis (B).

When the functional genes were inspected by presence/absence analysis, the vast majority (> 88% for the Mediterranean site, >91% for the Red Sea site) of the functional genes was shared between the sources of each collection site (Fig. 12). 0.5-5% of all genes was shared between two sources and only 0.2-4% of genes was unique to any given source. Only a small set of genes (<5%) was unique to any given source. In the following, the gene categories that

were well covered by gene probes (resulting in normalized signal intensity > 5%) and that are meaningful in the context of sponge symbioses are discussed.

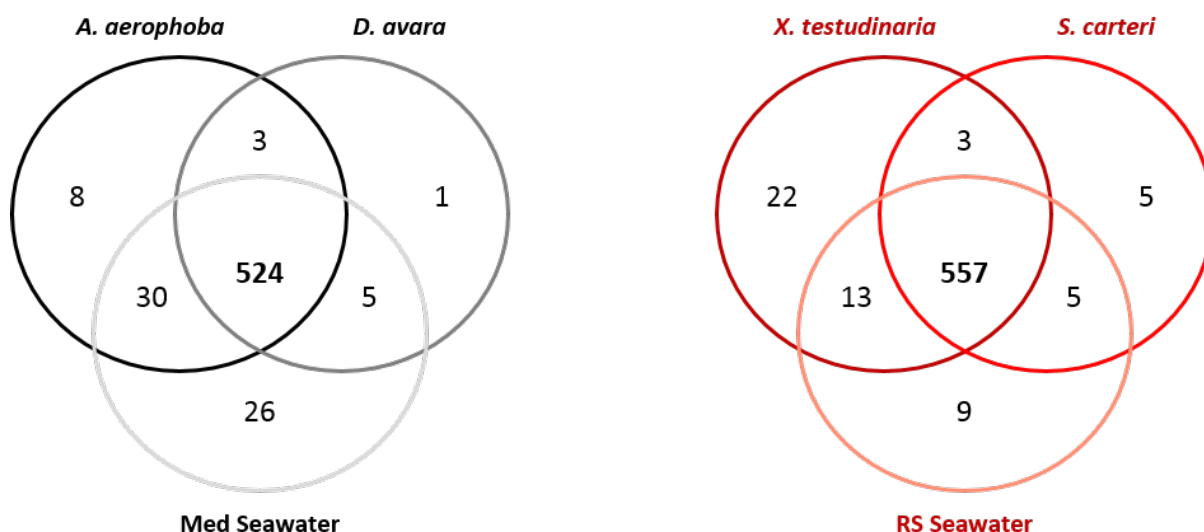


Figure 12. Presence/absence analysis of functional genes identified by the GeoChip. A total of 524 genes (88%) were shared between sources for the Mediterranean site and 557 genes (91%) were shared between sources of the Red Sea site.

With respect to nitrogen metabolism, genes corresponding to nearly all major pathways were identified by GeoChip analyses. These were nitrification, denitrification, ammonification, assimilatory and dissimilatory nitrogen reduction, as well as nitrogen fixation. Genes involved in anaerobic ammonia oxidation (anammox) were noticeably absent. In the following, only the genes that resulted in statistically significantly different signal intensities in either HMA or HMA sponges in comparison to seawater and where the difference was present at both geographic locations are shown (Figs. 13-15). The unit signal intensity is taken hereafter as a proxy for gene abundance. With respect to nitrification (Fig. 13), the sponge samples showed lower archaeal *amoA* gene abundance but higher bacterial *amoA* than seawater. Additionally, the *hao* gene abundance was reduced in sponges over seawater. In terms of denitrification, the sponge samples showed lower amounts of *narG*, *nirS*, and *nosZ* than seawater. The sponge samples showed further higher *nrfA* gene abundance (dissimilatory nitrogen reduction to ammonia) and a higher of *ureC* gene abundance (ammonification) than seawater.

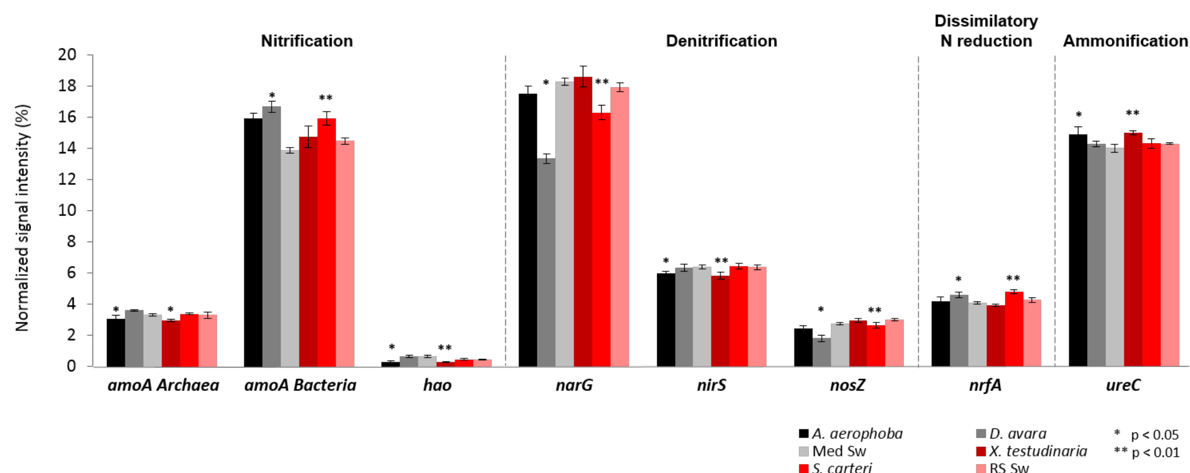


Figure 13. Normalized average signal intensities of genes involved in nitrogen cycling. The microbial processes and corresponding genes are as follows: nitrification (archaeal and bacterial amoA encoding ammonia monooxygenase, hao for hydroxylamine oxidoreductase); denitrification (narG for nitrate reductase, nirS for nitrite reductase, nosZ for nitrate reductase); dissimilatory N reduction to ammonium (nrfA for c-type cytochrome nitrite reductase); ammonification (ureC for urease). Data are presented as the mean ± SE. **: P<0.01, *: P<0.05. "Med Sw" and "RS Sw" stand for Mediterranean and Red Sea seawater.

In terms of carbon cycling (Fig. 14), the *mcrA* gene involved in methane production was reduced in sponges over seawater. The *pcc* gene abundance (Citrate cycle) was higher in sponges, while that of *RuBisCO* (Calvin cycle) gene was lower in sponges over seawater. With respect to carbon degradation, an increased gene abundance encoding for exochitinase, but a decreased gene abundance encoding for xylase was noted in sponges over seawater.

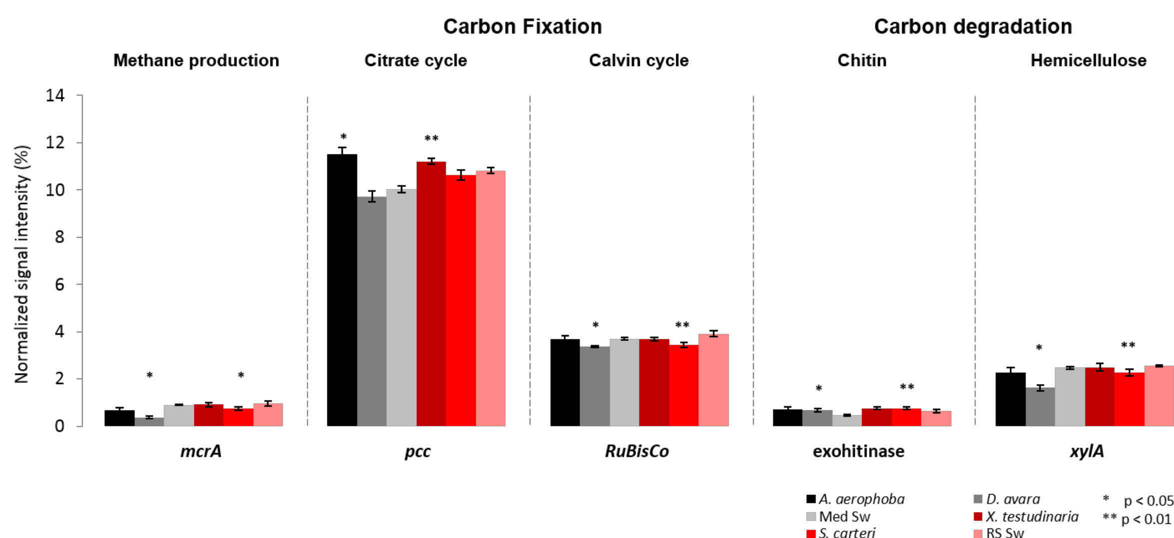


Figure 14. Normalized average signal intensities of gene categories involved in carbon cycling. The microbial processes and corresponding genes are as follows: methane production (*mcrA*, encoding for methyl coenzyme M reductase A); carbon fixation in the citrate cycle (*pcc* for propionyl-CoA carboxylase) and in the Calvin cycle (*RuBisCO* for Ribulose-1, 5-bisphosphate carboxylase/oxygenase); degradation of chitin (with genes encoding for exochitinase) and of hemicellulose (*xylA* for xylanase). The standard nomenclature of the GeoChip 4 was maintained (ref). Data are presented as the mean ± SE. **: P<0.01, *: P<0.05. “Med Sw” and “RS Sw” stand for Mediterranean and Red Sea seawater.

With respect to stress (Fig. 15), the sponge samples showed decreased gene abundance in the category oxygen stress (*ahpC*, *katE*). The *pstS* gene involved in phosphate limitation was increased in sponges over seawater. Noteworthy was the lower gene abundance of *obgE* involved in radiation stress in both sponge samples (HMA, LMA) from both geographic locations.

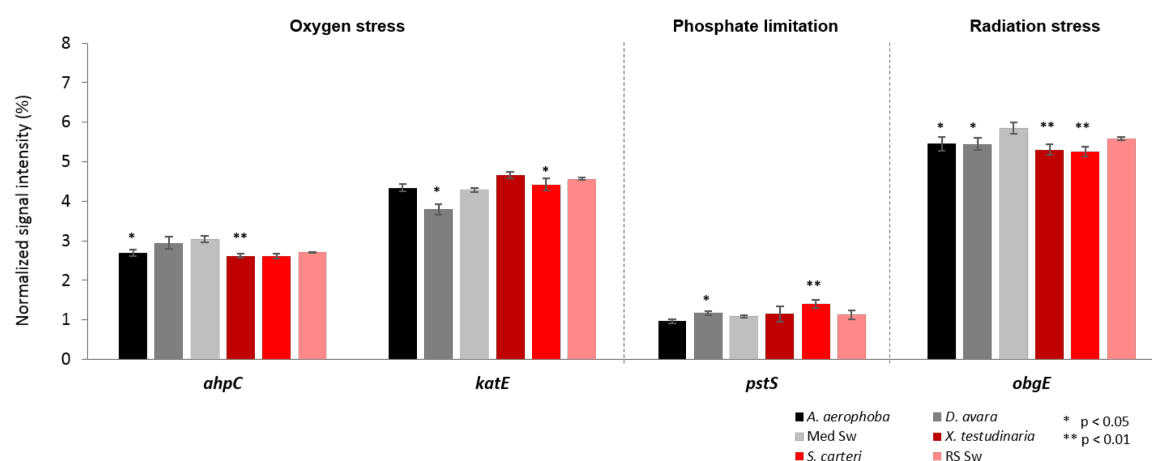


Figure 15. Normalized average signal intensities of gene categories involved in stress. The microbial genes involved in stress response are as follows: *ahpC* encoding for alkyl hydroperoxide reductase and katE involved in oxygen stress; *pstS* encoding for a phosphate binding protein involved in phosphate limitation; *obgE* encoding for a GTPase involved in radiation stress. Data are presented as the mean ± SE. **: P<0.01, *: P<0.05. "Med Sw" and "RS Sw" stand for Mediterranean and Red Sea seawater.

4.4.5 Discussion

The microbial diversity of the investigated sponge species and seawater is fully consistent with previous reports (Schmitt *et al.*, 2012, Schmitt *et al.*, 2012, Moitinho-Silva *et al.*, 2014). The HMA sponges show a more complex phylum-level composition with *Proteobacteria* and *Chloroflexi* as most abundant phyla, followed by *Acidobacteria*, *Deferribacteres*, *Nitrospirae*, *Gemmatimonadetes*, *Actinobacteria*, and candidate phylum Poribacteria. The phylum level composition of the LMA sponge species, *S. carteri*, was dominated by *Proteobacteria*, *Cyanobacteria*, and *Archaea*, as has been reported previously (Giles *et al.*, 2013, Moitinho-Silva *et al.*, 2014). Several additional studies revealed a distinct and different microbial composition in HMA sponges versus LMA sponges (Weisz *et al.*, 2007, Kamke *et al.*, 2010, Erwin *et al.*, 2011, Schmitt *et al.*, 2012, Giles *et al.*, 2013, Gloeckner *et al.*, 2013, Poppell *et al.*, 2013, Moitinho-Silva *et al.*, 2014). However, even though the seawater samples resemble the composition of LMA sponges on the phylum level, the LMA sponges still contain their own characteristic symbiont guilds (Moitinho-Silva *et al.*, 2014), as evidenced here by a dominance of *Betaproteobacteria* in *D. avara* and *Gammaproteobacteria* in *S. carteri*.

The ordination of microbial community diversity by PCoA based on amplicon data

revealed a clear separation of HMA sponges, LMA sponges, and seawater (Fig. 10A). Similarly, when the GeoChip functional data were analysed by MCE, a similar pattern was observed (Fig. 11A). The fact that the HMA LMA pattern was stronger than any geographic bias is remarkable considering that the Mediterranean and Red Sea are distinct locations. To our knowledge, this is the first time that the HMA LMA pattern was identified on the level of microbial functional genes. However, upon ordination by MCE of functional genes, few specific differences could be identified (Fig. 11B).

When the GeoChip data were analysed for the presence/absence of genes, few specific differences could be identified. For instance, the vast majority of genes (> 88% for the Mediterranean site, > 91% for the Red Sea site) were shared between all three sources (Fig. 12). The sets of genes unique to a given source did not reveal any specific genes or pathways, thus the relevance of the unique genes remains unknown (data not shown). It is important to consider, that equal DNA concentrations for each sample were applied to the GeoChip. For example, a much larger volume of seawater was necessary to extract the same amount of DNA. Similarly, much more LMA sponge biomass was necessary to yield the same amount of DNA as from HMA sponges. The results show that the microbial functional gene repertoires of HMA and LMA sponges, as well as of seawater, are largely similar. Any differences appear to be rather due to microbial and/or gene abundances rather than to true differences in the functional gene repertoire.

Functional convergence of microbial sponge symbionts has previously been reported based on comprehensive metagenomic analyses (Fan *et al.*, 2012) and based on nutrient fluxes in combination with phylogenetic analyses of selected sponge symbionts (Ribes *et al.*, 2012). Here we extend previous findings in that functional convergence is also reported between sponges (HMA and LMA) and seawater. Several explanations are possible for this observation. For once, the GeoChip may have only limited applicability to sponge microbiomes in the sense that the truly unique microbial genes may not have been covered. Secondly, owing to the presence of seawater bacteria in sponge microbiomes, which has been shown to amount to 5-24% depending on the HMA or LMA status (Moitinho-Silva *et al.*, 2014), the extracted DNA may have contained a sufficient amount of seawater bacterial DNA to mask true differences between the sponge microbiomes. As a third possible explanation, it may be considered that if seawater was the adaptive driving force, then the functional gene

repertoire is going to be the same whether the microorganisms reside inside or outside of a sponge. In the latter hypothesis, it would thus not be surprising to observe functional gene convergence in different sponge or seawater microbiomes. The few differences that were found and that were of relevance to sponge symbioses are discussed below.

Microbial nitrogen metabolism is a major theme to have emerged out of the past decade of sponge microbiology (Taylor *et al.*, 2007, Hentschel *et al.*, 2012, Webster & Taylor, 2012). Sponges excrete ammonia as a metabolic waste product, which makes them a particularly attractive niche for microorganisms in an otherwise nitrogen-poor marine environment. Accordingly, the abundance of the *amoA* gene, as an indicator for bacterial ammonia oxidation was higher in both HMA and LMA sponges than in seawater (Fig. 13). The archaeal *amoA* was however slightly reduced in HMA sponges and so was the *hao* gene encoding for hydroxylamine oxidase. However, in spite of the higher presence of bacterial *amoA* over archaeal *amoA* gene, metatranscriptome analyses revealed archaeal ammonia monooxygenase to be among the most highly expressed microbial features in *Geodia barretti* and *S. carteri* (Radax, *et al.*, 2012, Moitinho-Silva, *et al.*, 2014). Archaea appear to be the main drivers of ammonia oxidation in many other ecosystems such as seawater (Shi *et al.*, 2011) and soil (Leininger *et al.*, 2006).

Several genes encoding for denitrification were reduced in sponges over seawater (Fig. 13). This observation is consistent with the current perception that sponges have generally aerobic metabolism and that the mesohyl matrix is well oxygenated as a result from the intense pumping activities of the animal. Somewhat contradictory, the *nrfA* gene abundance, encoding for the dissimilatory reduction of nitrate to ammonia, was slightly increased in LMA sponges over seawater. The sponge tissues were shown to turn anaerobic during periods of non-pumping, so denitrification may still be a likely, if only temporary, scenario (Schlappy *et al.*, 2010).

The symbionts may further cover their nitrogen needs from urea hydrolysation. Urea can originate from a variety of sources (Crandall & Teece, 2012), such as by bacterial degradation of nucleic and amino acids and is therefore a likely product to be encountered in the sponge mesohyl. GeoChip analyses revealed higher *ureC* gene abundance in sponges than in seawater (Fig. 13). This gene encoding for a urease subunit and endows the bacteria with the capacity to hydrolyse urea. Urease-encoding gene clusters, urea transporters, and accessory

genes were previously identified in sponge symbiont genomes (Hallam *et al.*, 2006, Siegl *et al.*, 2011). Furthermore, phylogenetic diversity and transcriptional activities of the *ureC* gene was recently documented for the microbial symbionts of *X. testudinaria*, providing strong support for an *in vivo* relevance of microbial urea utilization in sponge symbioses (Su *et al.*, 2013).

Autotrophic carbon fixation is a hallmark of cyanobacteria, including cyanobacterial symbionts of sponges (Taylor *et al.*, 2007). Gene copy numbers for RuBisCO, the key enzyme of cyanobacterial carbon fixation, were however lower in the sponges than by comparison, in seawater (Fig. 14). On the other hand, the key gene, *pcc*, encoding for propionyl-CoA-carboxylase, and representing the recently discovered autotrophic carbon fixation pathway in archaea (Berg *et al.*, 2007) was higher in HMA sponges than in seawater. It appears likely that the wide-spread sponge symbiont *Crenarchaeum symbiosum* is responsible for autotrophic carbon fixation via the 3-hydroxypropionate/4-hydroxybutyrate cycle in sponges (Hallam *et al.*, 2006).

With respect to carbon degradation, only two genes were identified as being statistically different between sponges and seawater (Fig. 14). Higher exochitinase gene abundance underlines the potential of sponge symbionts for chitin degradation. In this context, a glycoside hydrolase (GH74) with potential for chitin deacetylation was recently identified on poribacterial genomes by single cell genomics (Kamke *et al.*, 2013). Alpha-chitin was previously identified as a structural component of the sponge skeleton and may thus likely be encountered in the sponge mesohyl (Ehrlich *et al.*, 2007). The lower abundance of genes encoding for hemicellulose, indicates that hemicellulose degradation is probably not relevant in sponge microbiomes. Similarly, the lower gene abundance of *mcrA*, an indicator gene for anaerobic, methanogenic archaea, was lower in sponges than in seawater, indicating that the potential for methanogenesis in sponges is probably limited.

A consistently lower abundance of genes involved in stress, particularly with respect to oxygen (*ahpC*, *katE*) and radiation stress (*obgE*) was observed in sponges over seawater (Fig. 15). The finding of reduced *obgE* gene levels was observed for HMA and LMA sponges at both locations. A microbial existence within animals is frequently correlated with an increased repertoire of stress protection proteins such as chaperonins (Liu *et al.*, 2012, Fan *et al.*, 2013). Whether the symbionts of sponges are more or less stressedthan their planktonic

counterparts will continue to be an interesting topic for future investigations.

The *pstS* gene, which is involved in phosphate transport, was slightly increased in LMA sponges over seawater, possibly indicating the sponge symbionts are subject to phosphate limitation. In this context it is interesting to note that the proteobacterial glycerol-3-phosphate ABC transporter was among the highly expressed genes in the *S. carteri* transcriptome (Moitinho-Silva *et al.*, 2014). Furthermore, a number of glycerol-3-P ABC-type transporter proteins, termed UgpB, were identified in the metaproteogenome of *C. concentrica* (Liu *et al*., 2012). The Ugp system is thought to be involved in scavenging phosphate-containing compounds (Boos, 1998). Furthermore, single cell genomic analyses revealed an abundance of *phyH* genes in poribacterial symbionts of sponges that should endow the bacteria with the ability to utilize 2-aminoethylphosphonate (2-AEPn) as dissolved organic posphorus source (Kamke *et al.*, 2013). However, in spite of several predictions resulting from omics approaches, functional experimental evidence for phosphorous metabolism is still lacking for sponge symbionts.

In conclusion, the HMA/LMA dichotomy was confirmed by amplicon sequencing for the investigated sponges. Furthermore, GeoChip analyses by MCE showed, for the first time, that the HMA/LMA dichotomy exists also on the functional gene level, even though very few specific differences could be identified. MCE revealed further nonlinear relations between the samples that are only poorly discriminated by conventional methods like principal component analysis (PCA; see Suppl. Fig. 2). Future efforts will be directed at developing algorithms that identify which genes in particular account for the sample separation. With respect to the identified differences in gene repertoire, microbial nitrification and ammonification genes were increased in sponges, while denitrification genes were reduced. With respect to carbon metabolism, a higher abundance of archaeal autotrophic carbon fixation genes was noted in sponges than in seawater. Thirdly, stress genes were found at lower abundances in sponge microbiomes than in seawater. While methodological limitations, such as the applicability of the GeoChip outside of its original “soil” context, cannot be ruled out, it appears nonetheless conceivable, that sponge-associated and seawater microorganisms have most of their functional gene repertoire in common.

4.4.6 Acknowledgements

We thank Martin Pfannkuchen at the Institute Ruder Boskovic (Rovinj, Croatia) as well as the Coastal and Marine Resources Core Lab and the Biosciences Core Laboratory (KAUST, Saudi Arabia) for expert help in sample collection. We thank Zhili He (University of Oklahoma, USA) and Joy D. van Noystrand at Glomics, Inc. for advice. LM-S was supported by a grant of the German *Excellence Initiative* to the Graduate School of Life Sciences, University of Wuerzburg. Additional financial support was generously provided by KAUST baseline funds to TR.

4.4.7 Supplementary Figure legends

Supplementary Figure 1: Rarefaction curves based on a 97% sequence similarity threshold for marine sponge- and seawater-derived 16S rRNA amplicon pyrosequences from the Mediterranean and the Red Sea (A). Aplysina aerophoba (AA1-3), Dysidea avara (DA1-3), Mediterranean seawater (MWT1-3), Xestospongia testudinaria (AXD1-3), Stylissa carteri (ASD1-3), and Red Sea seawater (WTD1-3). Richness (B) and diversity (C) statistics.

Supplementary Figure 2: Principal component analysis (PCA) of the GeoChip 4. Relation among samples were explored by PCA. Samples of HMA (green) and LMA (red) sponges as well as seawater (blue) collected in the Mediterranean or Red Sea are shown.

4.4.8 References

Bell JJ (2008) The functional roles of marine sponges. *Estuarine Coastal and Shelf Science* **79**: 341-353.

Berg IA, Kockelkorn D, Buckel W & Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science* **318**: 1782-1786.

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

Cannistraci CV, Alanis-Lobato G & Ravasi T (2013) Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* **29**: i199-209.

Cannistraci CV, Ravasi T, Montevecchi FM, Ideker T & Alessio M (2010) Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes. *Bioinformatics* **26**: i531-539.

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

Chan Y, Van Nostrand JD, Zhou J, Pointing SB & Farrell RL (2013) Functional ecology of an Antarctic Dry Valley. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 8990-8995.

Chao A (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics* **11**.

Chao A. LSM (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association* **87**: 210-217.

Crandall JB & Teece MA (2012) Urea is a dynamic pool of bioavailable nitrogen in coral reefs. *Coral Reefs* **31**: 207-214.

de Goeij JM, van Oevelen D, Vermeij MJ, Osinga R, Middelburg JJ, de Goeij AF & Admiraal W (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342**: 108-110.

Ehrlich H, Maldonado M, Spindler KD, *et al.* (2007) First evidence of chitin as a component of the skeletal fibers of marine sponges. Part I. Verongidae (demospongia: Porifera). *Journal of experimental Zoology. Part B, Molecular and developmental evolution* **308**: 347-356.

Erwin PM, Olson JB & Thacker RW (2011) Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PloS One* **6**: e26806.

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS & Thomas T (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the United States of America* **109**: E1878-1887.

Fan Y, Thompson JW, Dubois LG, Moseley MA & Wernegreen JJ (2013) Proteomic analysis of an unculturable bacterial endosymbiont (Blochmannia) reveals high abundance of chaperonins and biosynthetic enzymes. *Journal of Proteome Research* **12**: 704-718.

Giles EC, Kamke J, Moitinho-Silva L, Taylor MW, Hentschel U, Ravasi T & Schmitt S (2013) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology* **83**: 232-241.

Gloeckner V, Hentschel U, Ereskovsky AV & Schmitt S (2013) Unique and species-specific microbial communities in *Oscarella lobularis* and other Mediterranean *Oscarella* species (Porifera: Homoscleromorpha). *Marine Biology* **160**: 781-791.

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

He Z, Gentry TJ, Schadt CW, *et al.* (2007) GeoChip: a comprehensive microarray for investigating biogeochemical, ecological and environmental processes. *The ISME Journal* **1**: 67-77.

Hentschel U, Piel J, Degnan SM & Taylor MW (2012) Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology* **10**: 641-654.

Hentschel U, Fieseler L, Wehrl M, Gernert C, Steinert M, Hacker J & Horn M (2003) Microbial diversity of marine sponges. *Prog Mol Subcell Biol* **37**: 59-88.

Kamke J, Taylor MW & Schmitt S (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME Journal* **4**: 498-508.

Kamke J, Sczyrba A, Ivanova N, *et al.* (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME Journal* **7**: 1-14.

Kang S, Van Nostrand JD, Gough HL, He Z, Hazen TC, Stahl DA & Zhou J (2013) Functional gene array-based analysis of microbial communities in heavy metals-contaminated lake sediments. *FEMS Microbiology Ecology* **86**: 200-214.

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. *The ISME Journal* **5**: 650-664.

Leininger S, Urich T, Schloter M, *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442**: 806-809.

Liu M, Fan L, Zhong L, Kjelleberg S & Thomas T (2012) Metaproteogenomic analysis of a community of sponge symbionts. *The ISME Journal* **6**: 1515-1525.

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

Lu Z, Deng Y, Van Nostrand JD, *et al.* (2012) Microbial gene functions enriched in the Deepwater Horizon deep-sea oil plume. *The ISME Journal* **6**: 451-460.

Mason OU, Nakagawa T, Rosner M, *et al.* (2010) First investigation of the microbiology of the deepest layer of ocean crust. *PloS One* **5**: e15399.

Moitinho-Silva L, Seridi L, Ryu T, Voolstra CR, Ravasi T & Hentschel U (2014) Revealing microbial functional activities in the Red Sea sponge Stylissa carteri by metatranscriptomics. *Environmental Microbiology*.

Moitinho-Silva L, Bayer K, Cannistraci CV, *et al.* (2014) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology*.

Nostrand JD, He Z & Zhou J (2012) Use of functional gene arrays for elucidating in situ biodegradation. *Frontiers in Microbiology* **3**: 339.

Oksanen J, Blanchet FG, Kindt R, *et al.* (2012) vegan: Community Ecology Package.

Poppell E, Weisz JB, Spicer L, Massaro A, Hill A & Hill M (2013) Sponge heterotrophic capacity and bacterial community structure in high- and low-microbial abundance sponges. *Marine Ecology* 1-11.

Radax R, Rattei T, Lanzen A, Bayer C, Rapp HT, Urich T & Schleper C (2012) Metatranscriptomics of the marine sponge Geodia barretti: tackling phylogeny and function of its microbial community. *Environmental Microbiology* **14**: 1308-1324.

Reiswig HM (1981) Partial carbon and energy budgets of the bacteriosponge *Verongia fistularis* (Porifera: Demospongiae) in Barbados West-Indies *Marine Biology* **2**: 273-294.

Ribes M, Jimenez E, Yahel G, *et al.* (2012) Functional convergence of microbes associated with temperate marine sponges. *Environmental Microbiology* **14**: 1224-1239.

Schlappy ML, Schottner SI, Lavik G, Kuypers MM, de Beer D & Hoffmann F (2010) Evidence of nitrification

and denitrification in high and low microbial abundance sponges. *Marine Biology* **157**: 593-602.

Schmitt S, Hentschel U & Taylor M (2012) Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia* **687**: 341-351.

Schmitt S, Tsai P, Bell J, *et al.* (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME Journal* **6**: 564-576.

Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* **27**: 379–423.

Shi Y, Tyson GW, Eppley JM & DeLong EF (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal* **5**: 999-1013.

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

Simister RL, Deines P, Botte ES, Webster NS & Taylor MW (2012) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environmental Microbiology* **14**: 517-524.

Simpson EH (1949) Measurement of Diversity. *Nature* **163**: 688-688.

Su J, Jin L, Jiang Q, Sun W, Zhang F & Li Z (2013) Phylogenetically diverse ureC genes and their expression suggest the urea utilization by bacterial symbionts in marine sponge Xestospongia testudinaria. *PloS One* **8**: e64848.

Taylor MW, Radax R, Steger D & Wagner M (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews* **71**: 295-347.

Tu Q, Yu H, He Z, *et al.* (2014) GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Molecular Ecology Resources*.

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.

Webster NS & Taylor MW (2012) Marine sponges and their microbial symbionts: love and other relationships. *Environmental Microbiology* **14**: 335-346.

Webster NS, Taylor MW, Behnam F, *et al.* (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12**: 2070-2082.

Weisz JB, Hentschel U, Lindquist N & Martens CS (2007) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Marine Biology* **152**: 475-483.

Wilson MC, Mori T, Ruckert C, *et al.* (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**: 58-62.

Wu L, Liu X, Schadt CW & Zhou J (2006) Microarray-based analysis of subnanogram quantities of microbial community DNAs by using whole-community genome amplification. *Applied and Environmental Microbiology* **72**: 4931-4941.

Yergeau E, Bokhorst S, Kang S, Zhou J, Greer CW, Aerts R & Kowalchuk GA (2012) Shifts in soil microorganisms in response to warming are consistent across a range of Antarctic environments. *The ISME Journal* **6**: 692-702.

4.5 Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics

This article is in press in **Environmental Microbiology** and is reprinted here with the permission from John Wiley and Sons and the authors. This article's reference is:

Moitinho-Silva, L., Seridi, L., Ryu, T., Voolstra, C.R., Ravasi, T., Hentschel, U. (2014) Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. *Environmental Microbiology*. E-pub ahead of print, doi: 10.1111/1462–2920.12533.

The documentation of author's contributions is found in the Appendix. The supplementary material is found in the CD attached to this PhD Thesis.

environmental microbiology

society for applied microbiology sfam

Environmental Microbiology (2014) doi:10.1111/1462-2920.12533

Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics

Lucas Moitinho-Silva,[1] Loqmane Seridi,[2] Taewoo Ryu,[2] Christian R. Voolstra,[3] Timothy Ravasi[2] and Ute Hentschel[1]*

[1]*Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Julius-von-Sachs Platz 3, 97082 Wuerzburg, Germany.* [2]*Division of Biological and Environmental Sciences and Engineering, Division of Applied Mathematics and Computer Science and* [3]*Red Sea Research Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia.*

Summary

Sponges are important components of marine benthic environments and are associated with microbial symbionts that carry out ecologically relevant functions. *Stylissa carteri* is an abundant, low-microbial abundance species in the Red Sea. We aimed to achieve the functional and taxonomic characterization of the most actively expressed prokaryotic genes in *S. carteri*. Prokaryotic mRNA was enriched from sponge total RNA, sequenced using Illumina HiSeq technology and annotated using the metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline. We detected high expression of archaeal ammonia oxidation and photosynthetic carbon fixation by members of the genus *Synechococcus*. Functions related to stress response and membrane transporters were among the most highly expressed by *S. carteri* symbionts. Unexpectedly, gene functions related to methylotrophy were highly expressed by gammaproteobacterial symbionts. The presence of seawater-derived microbes is indicated by the phylogenetic proximity of organic carbon transporters to orthologues of members from the SAR11 clade. In summary, we revealed the most expressed functions of the *S. carteri*-associated microbial community and linked them to the dominant taxonomic members of the microbiome. This work demonstrates the applicability of metatranscriptomics to explore poorly characterized symbiotic consortia and expands our knowledge of the ecologically relevant functions carried out by coral reef sponge symbionts.

Received 26 February, 2014; revised 15 April, 2014; accepted 15 April, 2014. *For correspondence. E-mail ute.hentschel@uni-wuerzburg.de; Tel. 49-931-31-82581; Fax 49-931-31-86235.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd

Introduction

The use of ‘omics’ to investigate gene expression at RNA and at protein level, respectively metatranscriptomics and metaproteomics, has significantly increased our understanding of microbial community functions in various habitats, including seawater (Frias-Lopez *et al*., 2008; Teeling *et al*., 2012), soil (Leininger *et al*., 2006; Urich *et al*., 2008), marine invertebrates (Kleiner *et al*., 2012; Liu *et al*., 2012; Radax *et al*., 2012b; Sanders *et al*., 2013) and the human gut (Verberkmoes *et al*., 2009; Gosalbes *et al*., 2011). In comparison with proteomics, the RNA-based work has the advantage of using high-throughput sequencing technologies, and it is less technically challenging with respect to molecule extraction, separation and function identification (Moran, 2009). Because the messenger RNA (mRNA) represents a small portion of the total RNA, a common strategy is to physically remove the ribosomal RNA (rRNA) fraction prior to sequencing and to removing annotated rRNA reads from sequence data (Poretsky *et al*., 2009; Stewart *et al*., 2010; Stewart, 2013).

Sponges (phylum Porifera) are important components of marine benthic ecosystems and play important roles in reef consolidation as well as coupling of benthic and pelagic environments. The latter is largely due to their immense filter-feeding capacities and consequent impacts upon coastal food webs and biogeochemical cycles (Bell, 2008; de Goeij *et al*., 2013). Many sponges harbor a great diversity of symbiotic microorganisms from the three domains of life, i.e. Archaea, Bacteria and Eukaryota, within their mesohyl tissues. To date, representatives from more than 28 bacterial phyla (including candidate phyla such as Poribacteria) and two archaeal lineages were identified from marine sponges (Hentschel *et al*., 2012; Schmitt *et al*., 2012; Simister *et al*., 2012b). The sponge-associated microbial communities are enriched in specific lineages (Webster *et al*., 2010), which have been termed ‘sponge-specific’ (Hentschel *et al*., 2002) or more recently ‘sponge-enriched’ (Moitinho-Silva *et al*., 2014), and which are rare in non-sponge

environments (Taylor *et al*., 2013). The vast majority of sponge-associated microbes remains uncultivated and is thus functionally largely uncharacterized (Taylor *et al*., 2007).

There is increasing evidence that some of the functional roles of sponges in benthic environments are carried out by the associated microorganisms, as inferred from the detection of metabolic processes related to carbon, nitrogen and sulfur and the apparent production of chemical defences by these microbes (Taylor *et al*., 2007; Hentschel *et al*., 2012; Wilson *et al*., 2014). Most of these metabolic capabilities were investigated by detecting functional genes and/or by measuring the substrates and products of key enzymatic reactions. Few studies verified whether these functional genes were actually expressed *in situ*, either by assessing the expression levels of mRNA (Mohamed *et al*., 2008; Radax *et al*., 2012a,b) or by detecting the encoded proteins (Liu *et al*., 2012). Only two reports so far employed metaproteomic or metatranscriptomic approaches to characterize the expressed gene repertoire of the whole sponge-associated microbial community without prior targeting of a given pathway (Liu *et al*., 2012; Radax *et al*., 2012b).

This study aimed to characterize the functional gene repertoire that was expressed *in situ* by the microbial consortium associated with the Red Sea sponge *Stylissa carteri*. The Red Sea is a unique ecosystem that is characterized by very high salinities and high temperatures (Ngugi *et al*., 2012). The sponge *S. carteri* was chosen because its microbiome is already well characterized based on 16S rRNA gene clone library and amplicon sequencing (Lee *et al*., 2011; Giles *et al*., 2013; Moitinho-Silva *et al*., 2014). In spite of a large number of bacterial and archaeal ribotypes recovered from *S. carteri* individuals (Lee *et al*., 2011), the community was dominated by only a few lineages (Giles *et al*., 2013; Moitinho-Silva *et al*., 2014). These lineages were transcriptionally active *in situ*, as inferred from the analysis of 16S rRNA transcripts, and belonged mainly to the Gammaproteobacteria and to the genera *Synechococcus* and *Nitrospira* (Moitinho-Silva *et al*., 2014). The present study aimed to explore the dominant, actively transcribed prokaryotic functions within *S. carteri*. A protocol to enrich prokaryotic mRNA from total sponge RNA was developed towards this goal. The microbial community mRNA was sequenced using Illumina HiSeq technology, annotated using the metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) pipeline (Meyer *et al*., 2008), and bioinformatically analysed with respect to taxonomy and function. The results presented here expand our knowledge of the functional roles of microorganisms in sponges, with possible implications for reef ecology.

Results

Metatranscriptome annotation

Approximately 132.78 million paired-end Illumina reads from three *S. carteri* metatranscriptomes (mean of 44.26 ± 3.75 million reads per sample) were submitted to the MG-RAST pipeline (Table 1). The number of reads

Table 1. Summary of sequencing and annotation of *S. carteri* metatranscriptomes.

	Sample 1	Sample 2	Sample 3
Processing of raw sequences			
Raw sequences (paired)	55 384 652	58 715 998	65 978 474
Sequence length of raw sequences	101	101	101
MG-RAST pipeline results			
Submitted sequences	42 787 876	41 471 392	48 520 574
Mean sequence length of submitted sequences	77 ± 22	77 ± 22	77 ± 22
Artificial duplicate sequences	27 213 061	23 350 760	29 756 474
Quality-controlled sequences	9 410 780	12 252 851	11 840 019
Mean sequence length after quality control	87 ± 14	88 ± 14	87 ± 14
Predicted protein features	4 954 546	6 613 923	6 257 089
Predicted rRNA features	2 539 789	3 102 289	3 221 147
Identified protein features	979 442	1 242 433	783 317
Protein features annotated by MG-RAST[a]	568 907	667 826	436 448
Putative prokaryotic transcriptional features[b]	330 094	325 489	252 986
Functionally annotated prokaryotic transcriptional features			
Total	129 094	143 472	92 630
Bacteria (%)	65.44	65.96	68.09
Archaea (%)	6.02	3.83	3.85
Unidentified prokaryotic (%)	28.54	30.20	28.06
Unclassified sequences (%)	0.024	0.015	0.019

Numbers represent sequence or feature counts. Sequence lengths are reported in base pairs.
a. Features with at least one BLAT hit (e-value < 0.001) against any of the MG-RAST protein databases.
b. Protein features that were classified to Bacteria and/or to Archaea based on best BLAT hit (e-value < 0.001) against any of the MG-RAST protein databases. These features were retained for taxonomic classification as provided by the MG-RAST based on the LCA.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

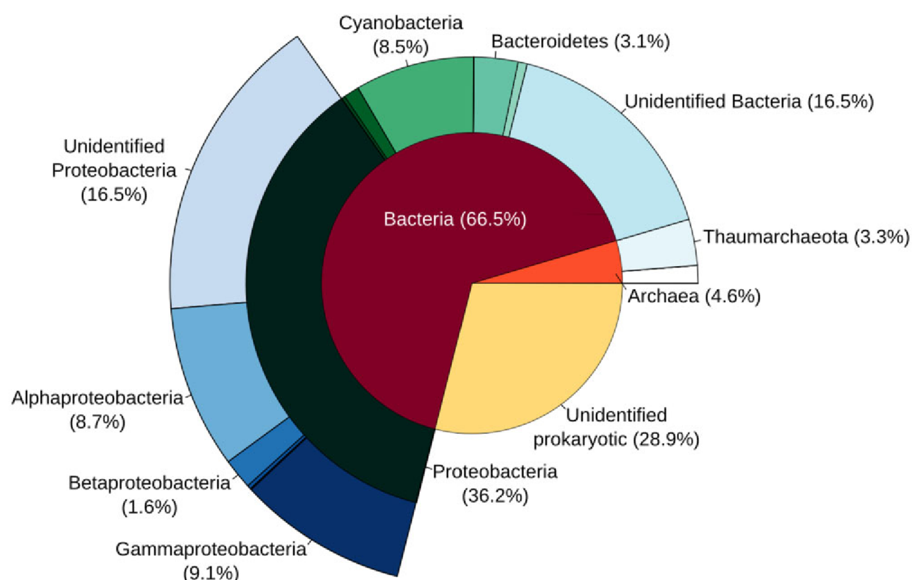


Fig. 1. Taxonomic assignment of *S. carteri* metatranscriptomes. Classification was performed according to MG-RAST-based LCA. Different levels of taxonomic assignment are shown. Taxa representing more than 1.5% of the annotated reads are named. A detailed list of taxonomic assignments at phylum and class level is found in the Supporting Information Table S1.

was reduced to a mean of 11.16 ± 1.53 million reads per biological replicate after quality control steps (Table 1). From these quality-controlled sequences, a mean of 5.94 ± 0.87 million predicted transcriptional features (corresponding to protein features in MG-RAST) and a mean of 2.95 ± 0.36 million rRNA features were obtained per replicate sample. This process resulted in a mean of 1 001 730 ± 230 368 transcriptional features identified per replicate. After a series of feature selections based on taxonomic classifications (see Experimental procedures and Table 1 for details), prokaryotic transcriptional features were functionally annotated based on SEED subsystems, resulting in a mean of 121 732 ± 26 208 per biological replicate. In order to assess the dissimilarity between biological replicates, the functional profiles of the three *S. carteri* metatranscriptomes were compared based on annotated prokaryotic transcriptional feature counts rarefied to the minimal number of features identified in a given replicate, this being 92 630 for sample 3. Accordingly, low variation was observed between replicates as inferred by low Bray–Curtis index values (0.12 ± 0.016).

According to MG-RAST-based lowest common ancestor (LCA) classification of the *S. carteri* metatranscriptomes, 66.5% ± 1.4% of the functionally annotated prokaryotic transcriptional features were assigned to Bacteria, 4.6% ± 1.3% were assigned to Archaea, 28.9 ± 1.1% were considered as unidentified prokaryotic and $1.9 \times 10^{-2} \pm 4.7 \times 10^{-3\%}$ were assigned as ‘unclassified sequences’ (Fig. 1). At phylum level, the majority of the features were assigned to Proteobacteria (36.2 ± 0.73%). Most of the proteobacterial transcriptional features were classified as Gammaproteobacteria (9.1 ± 0.69%), Alphaproteobacteria (8.7 ± 0.26%) and Betaproteobacteria (1.6 ± 0.25%). Approximately half of the proteobacterial transcriptional features could not be assigned to a given class (16.5 ± 0.20%). Other abundant bacterial phyla in *S. carteri* metatranscriptomes were Cyanobacteria (8.5% ± 2.04%) and Bacteroidetes (3.1% ± 0.30%). A large proportion of the transcriptional features were assigned to Bacteria only at the domain level (16.5% ± 0.39%). Among the archaea, the Thaumarchaeota displayed most assigned prokaryotic transcriptional features (3.3 ± 0.72%). The remaining features were scattered among various prokaryotic phyla that individually represented less than 1.5% of the total prokaryotic features (see Supporting Information Table S1 for a complete list of taxonomic assignments at phylum and class level for each biological replicate).

Taxonomic assignment of expressed gene functions

A total of 4111 gene functions were identified in *S. carteri* metatranscriptomes, which were classified into 28 categories (Fig. 2A). The most abundant categories were in the order identified: Clustering-based subsystems (17.2% of annotated prokaryotic transcriptional features), carbohydrates (14.8%), and amino acids and derivatives (12.2%). The most abundant subcategories (Fig. 2B) were in the order identified:

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

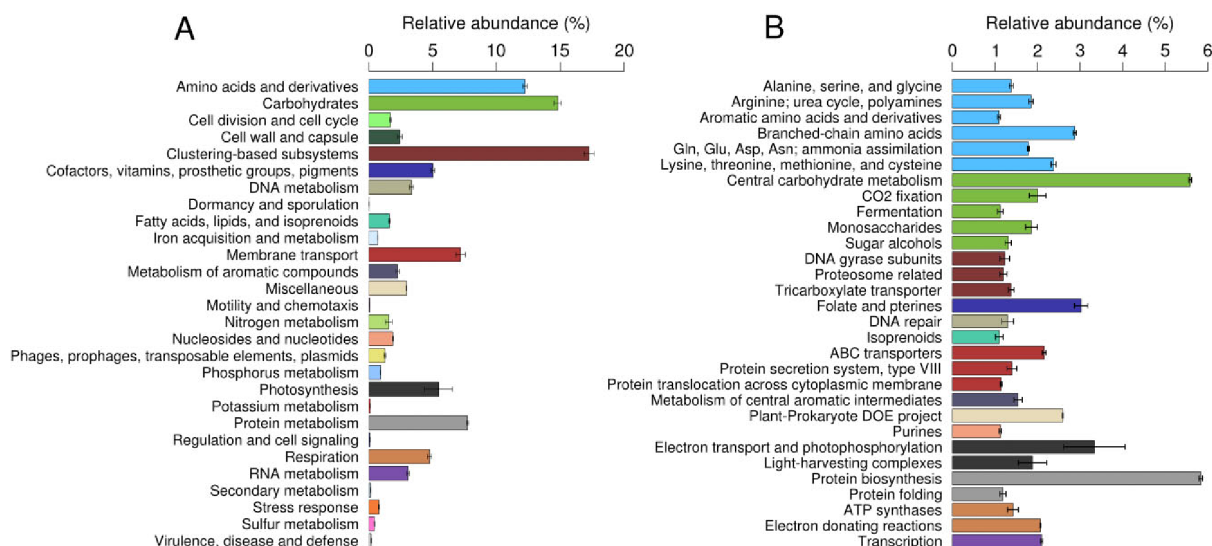


Fig. 2. Functional categories of the *S. carteri* metatranscriptomes. Functional classification of prokaryotic transcriptional features was done based on SEED subsystems. Bars represent percentage of features ($n = 3$, mean ± standard error) that were classified into (A) the first and (B) the second functional category levels. Only the most abundant functions are shown at the second functional category level.

protein biosynthesis (5.8%), central carbohydrate metabolism (5.6%) and electron transport/photophosphorylation (3.3%).

The top 100 most abundant gene functions (representing 38.7% ± 1.9% of annotated prokaryotic transcriptional features) were analysed considering the taxonomic affiliation of the assigned transcriptional features. The most expressed gene functions of *S. carteri* metatranscriptomes were: fap unknown function protein (1.32%), bacterial proteasome-activating AAA-ATPase (PAN) (1.06%) and ammonia monooxygenase (0.96%) (Table 2, Fig. 3A). For some gene functions, the taxonomic affiliation of most of the features remained unresolved at the domain level. These functions were ribulose bisphosphate carboxylase large chain (Fig. 3B, index 22), ribokinase (index 25), branched-chain amino acid ABC transporter, amino acid-binding protein (index 58), 5-carboxymethyl-2-hydroxymuconate delta-isomerase (index 65) and photosynthesis-related genes (indexes 75–78), and respiration (index 92).

Table 2. The 15 most transcribed gene functions in *S. carteri* metatranscriptomes.

Rank	Functional classification	Functional description	Feature abundance (%)[a]	Figure 3 index
1	Membrane Transport	fap unknown function protein	1.32 ± 0.20	62
2	Clustering-based subsystems	bacterial proteasome-activating AAA-ATPase (PAN)	1.06 ± 0.13	45
3	Membrane Transport	ammonia monooxygenase	0.96 ± 0.27	64
4	Clustering-based subsystems	DNA gyrase subunit A (EC 5.99.1.3)	0.91 ± 0.18	42
5	Photosynthesis	photosystem II protein D1 (PsbA)	0.91 ± 0.39	76
6	Fatty Acids, Lipids, and Isoprenoids	gamma-carotene hydroxylase	0.86 ± 0.16	56
7	Carbohydrates	ribokinase (EC 2.7.1.15)	0.86 ± 0.20	25
8	Clustering-based subsystems	TRAP-type C4-dicarboxylate transport system, periplasmic component	0.84 ± 0.1	48
9	Protein Metabolism	chaperone protein DnaK	0.80 ± 0.12	88
10	Carbohydrates	acetate permease ActP (cation/acetate symporter)	0.77 ± 0.18	18
11	Nitrogen Metabolism	copper-containing nitrite reductase (EC 1.7.2.1)	0.77 ± 0.22	68
12	Cofactors, Vitamins, Prosthetic Groups, Pigments	formate dehydrogenase-O, major subunit (EC 1.2.1.2)	0.76 ± 0.12	52
13	RNA Metabolism	DNA-directed RNA polymerase beta subunit (EC 2.7.7.6)	0.75 ± 0.02	96
14	Protein Metabolism	translation elongation factor G	0.74 ± 0.05	83
15	Amino Acids and Derivatives	glutamine synthetase type I (EC 6.3.1.2)	0.70 ± 0.02	9

a. Feature abundance is reported as the percentage of prokaryotic transcriptional features assigned to a given gene function in relation to the total identified prokaryotic features per dataset ($n = 3$, ± standard deviations).

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

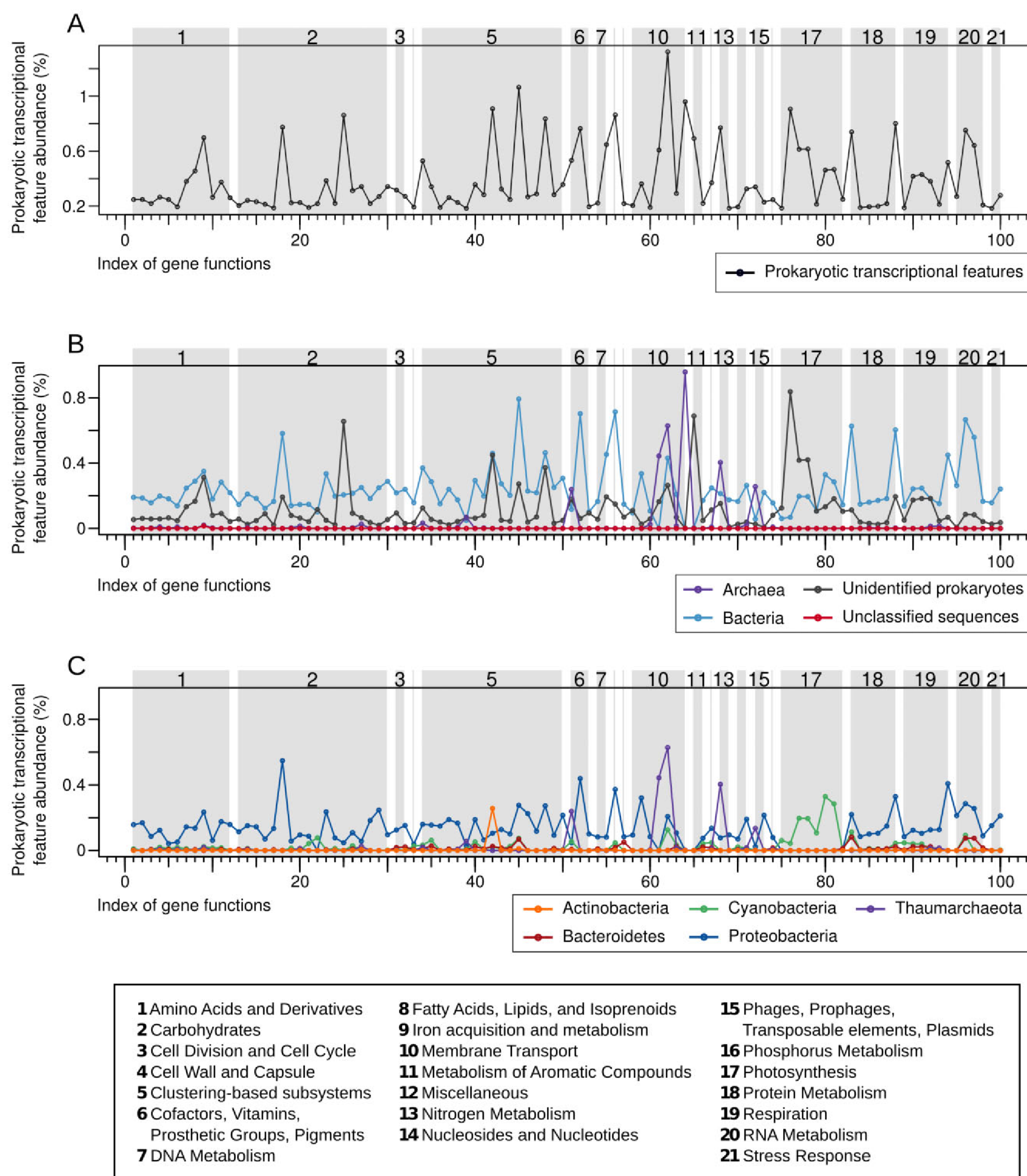


Fig. 3. Taxonomic assignment of the most expressed gene functions in *S. carteri* metatranscriptomes. The top 100 most expressed gene functions were sorted according to their annotation into 21 categories. Prokaryotic transcriptional feature abundance represents the mean percentage between biological replicates ($n = 3$). The list of genes is found in the Supporting Information Table S2.
(A) Transcriptional feature abundance is shown as total abundance, (B) on the domain-level and (C) on the phylum-level.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Most of the highly expressed gene functions were assigned to Bacteria (Fig. 3B, Supporting Information Table S2). Archaea were responsible for only few, but highly expressed functions in the *S. carteri* metatranscriptomes (see Fig. 3B), which were related to membrane transport of proteins (see Fig. 3B, category 10), such as fap unknown function protein (Fig. 3B, index 62) and type III secretion protein HrpG (index 61). However, further inspection of the HrpG nucleotide consensus sequence revealed that the transcripts were likely mis-annotated and represents rather an uncharacterized archaeal protein. Active archaeal ammonia oxidation was indicated by the high expression of archaeal ammonia monooxygenase (index 64) and copper-containing nitrite reductase (figure index 68). Additionally, the heat shock protein 60 family chaperone GroEL (index 72), ferredoxin (index 51) and the probable iron-binding protein from the HesB_IscA_SufA family (index 39) were highly expressed by the archaeal population. More specifically, all these functions were assigned to the phylum Thaumarchaeota, with exception of ammonia monooxygenase of which assignment was only until the domain level (see Fig. 3C).

The top expressed gene functions assigned to bacteria were inspected at phylum level, where Proteobacteria was the most prominent phylum (Fig. 3C). Most of the highly transcribed proteobacterial gene functions were related to carbon metabolism and transport. Proteobacterial gene functions related to single-carbon (C1) metabolism were particularly highly transcribed, as indicated by the expression of proteobacterial formate dehydrogenase-O, major subunit (Fig. 3C, index 52) and methanol dehydrogenase large subunit protein (index 94). Other carbohydrate related gene functions highly expressed by Proteobacteria included ferredoxin-like protein FixX (index 23). Furthermore, periplasmatic transporters of low molecular weight carbon compounds were highly expressed by Proteobacteria in *S. carteri* metatranscriptomes, including acetate permease ActP (cation/acetate symporter) (index 18), TRAP-type C4-dicarboxylate transport system components (indexes 47 and 48) and glycerol-3-phosphate ABC transporter components (indexes 28 and 29). Other highly expressed, transport-related functions included the putative TEGT family carrier/transport protein (index 46), the various polyols ABC transporter/periplasmic substrate-binding protein (index 99), the high-affinity leucine-specific transport system/periplasmic-binding protein LivK (index 59) and ABC transporter/ periplasmic spermidine putrescine-binding protein PotD (index 2). The proteobacterial glutamate synthase (NADPH) large chain (index 8) and glutamine synthetase type I (index 9) were detected among the highly expressed *S. carteri* bacterial gene functions related to amino acids metabolism (see Fig. 3C, category 1).

Proteobacterial gene functions related to stress were among the top expressed functions in *S. carteri* metatranscriptomes. These included rubrerythrin (index 100) and components of chaperone systems, such as the chaperone protein DnaK (index 88), the heat shock protein 60 family co-chaperone GroES (index 87) and the chaperone auxiliary protease ClpB protein (index 50). Gene functions related to intracellular degradation, including the ATP-dependent Clp protease ATP-binding subunit ClpA (index 37) and the bacterial proteasome-activating AAA-ATPase (PAN) (index 45), were also highly transcribed by Proteobacteria. Lastly, other gene functions expressed by proteobacterial populations were abundant in *S. carteri* metatranscriptomes, including gamma-carotene hydroxylase (index 56), DNA-directed RNA polymerase b subunit (indexes 96 and 97), fap unknown function protein (index 62) and translation elongation factor G (index 83).

The phylum Cyanobacteria was second with respect to the top expressed bacterial gene functions in the *S. carteri* metatranscriptomes. The highly expressed cyanobacterial gene functions (Fig. 3C) were related to (i) photosynthesis, such as components of the photosystem II (indexes 75–76), photosystem I (indexes 77–78), phycobilisome (79–81) and ATP synthase chains (indexes 89 to 91); and (ii) CO_2 fixation, including transketolase (index 21) and ribulose bisphosphate carboxylase large chain (index 22). The highest expressed photosynthetic functions were the alpha and beta chains of phycoerythrin (indexes 80 and 81). Other highly expressed cyanobacterial functions in *S. carteri* included ferredoxin (index 51), thioredoxin (index 67), the fap unknown function protein (index 62), SSU ribosomal protein S1p (index 40), the translation elongation factor G (index 83), the bacterial proteasome-activating AAA-ATPase (PAN) (index 45), DNA-directed RNA polymerase beta subunit (index 96) and polyribonucleotide nucleotidyltransferase (index 35).

With respect to the Bacteroidetes, the most prominent gene functions included: translation elongation factor G (Fig. 3C, index 83), DNA-directed RNA polymerase subunits (indexes 96 and 97), bacterial proteasome-activating AAA-ATPase (PAN), (index 45) and TonB-dependent receptor (index 57). A single gene function was highly expressed by Actinobacteria, the DNA gyrase subunit A (index 42). While the DNA gyrase subunit B was among the 100 most transcribed functions in *S. carteri* metatranscriptomes, most of the transcriptional features were classified to Proteobacteria (index 43) rather than Actinobacteria. There is no clear explanation why some genes are coexpressed [phycoerythrin subunits (indices 80, 81); photosystem I subunits (77, 78), and others are not (42, 43)]. The gene functions from other prokaryotic phyla contributed individually $< 0.02\%$ to the top annotated functions.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Table 3. Description of representative assembled reads of selected gene functions.

Taxon	Description	Total abundance rank	Taxon abundance rank	Reads assembled (*n*)	Percentage of reads[b]	Assembly length (bp)	Mean identity (%)[a]
Archaea							
	Ammonia monooxygenase	3rd	1st	3140	88	628	98.87 ± 1.85
	Copper-containing nitrite reductase	11th	4th	1464	50	676	99.50 ± 1.01
Proteobacteria							
	Acetate permease ActP (cation/acetate symporter)	10th	1st	182	6	1462	96.30 ± 4.42
	TRAP-type C4-dicarboxylate transport system, periplasmic component	8th	9th	172	5	941	97.95 ± 3.55
	Formate dehydrogenase-O	12th	2nd	281	10	1330	98.39 ± 2.44
	Methanol dehydrogenase large subunit	24th	3nd	1370	72	1717	99.68 ± 1.02
Cyanobacteria							
	Phycoerythrin alpha chain	26th	1st	312	18	541	99.08 ± 1.46
	Phycoerythrin beta chain	25th	2nd	275	15	531	98.73 ± 2.33

Ranks are based on the number of transcriptional features. The total abundance rank refers to the position of a given gene function in relation to all identified gene functions. The taxon abundance rank refers to the position of a given gene function in relation to all identified functions in the selected taxon.
a. Standard deviations are shown.
b. Percentage of reads is in relation to the assigned reads to a given gene function.

Phylogenetic analysis of highly expressed key gene functions

To obtain better phylogenetic resolution on the taxonomic assignments, the expressed gene functions that were among the most highly expressed and recognized indicator enzymes of primary metabolism were selected for each of the three most abundant community members (Archaea, Proteobacteria and Cyanobacteria). These selected gene functions were related to archaeal nitrogen metabolism (ammonia monooxygenase and copper-containing nitrite reductase), proteobacterial carbon transport (acetate permease ActP and TRAP-type C4-dicarboxylate transport system, periplasmic component) and C1 metabolism (formate dehydrogenase-O, major subunit and methanol dehydrogenase large subunit protein) and cyanobacterial photosynthesis (the alpha and beta chains of phycoerythrin). For each gene function, the representative sequence was retrieved from the consensus nucleotide sequence of the assembly with the greatest number of reads (Table 3). The proportion of reads assembled into a given representative sequence in relation to the total reads of the given gene function varied, ranging from 0.88 to 0.05, in the case of ammonia monooxygenase and TRAP-type C4-dicarboxylate transport system, periplasmic component respectively. While in the first case, the majority of sequences were highly conserved and apparently affiliated with one organism, in the latter case, there appeared to be a higher diversity of transcripts. Furthermore, the mean identity of the assembled sequences to the respective consensus sequence was high (> 96.3%), indicating good agreement between the consensus sequences and reads (Table 3).

The representative sequence of each selected gene function was used for phylogenetic analysis. Overall, the taxonomic assignments of these gene functions were confirmed up to genus level by this approach (Fig. 4), where the representative sequences were phylogenetically related to orthologues from Archaea, from Proteobacteria and from Cyanobacteria. The representative sequences of ammonia monooxygenase (Fig. 4A) and copper-containing nitrite reductase (Fig. 4B) were placed in well supported clades containing sequences from Crenarchaeota and Thaumarchaeota, but the placement into any specific phylum was not supported. The representative sequences of carbon transporters (Fig. 4C–D) were closely related to alphaproteobacterial sequences, falling into a well-supported clade that included sequences from the alphaproteobacterial genus *Pelagibacter* and uncultured Alphaproteobacteria. The representative sequences related to C1 metabolism, particularly the formate dehydrogenase-O, major subunit (Fig. 4E), were closely related to orthologues from Gammaproteobacteria. The representative sequence of methanol dehydrogenase large subunit protein (Fig. 4F) was closely related to an orthologue from the Gammaproteobacterium *Thiorhodococcus drewsii* and fell into a major clade that contained gammaproteobacterial orthologues as well as a well-supported internal cluster of betaproteobacterial orthologues. Finally, both cyanobacterial phycoerythrin subunits were more similar to

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

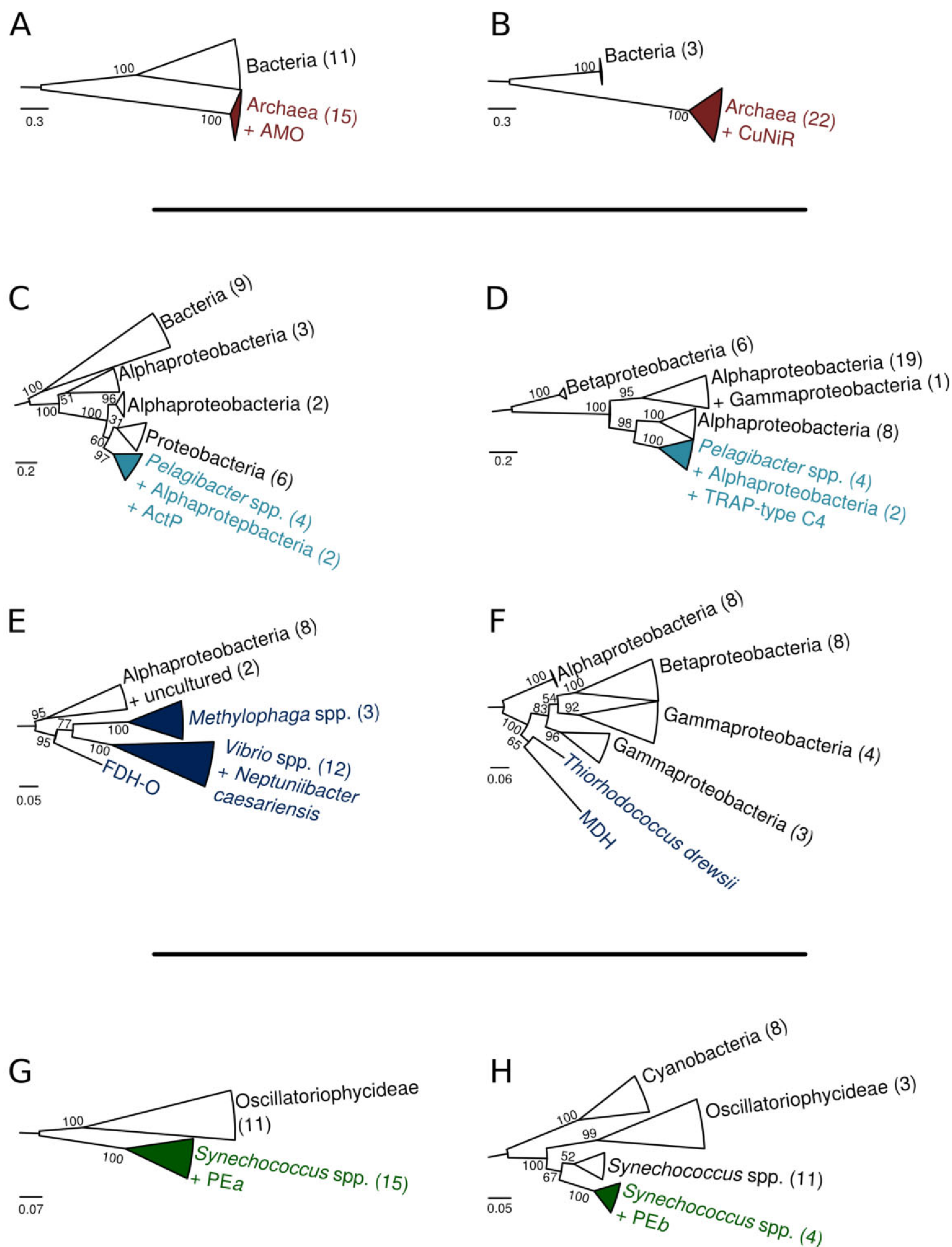


Fig. 4. Phylogeny of the highly expressed key gene functions of the most abundant taxa in *S. carteri* metatranscriptomes. Maximum-likelihood trees were constructed from representative sequences and orthologues that are related to nitrogen metabolism in Archaea: (A) ammonia monooxygenase (AMO) and (B) copper-containing nitrite reductase (CuNiR); to carbon transport in Proteobacteria: (C) acetate permease ActP (ActP) and (D) TRAP-type C4-dicarboxylate transport system, periplasmic component (TRAP-type C4); to C1 metabolism in Proteobacteria: (E) formate dehydrogenase-O, major subunit (FDH-O) and (F) methanol dehydrogenase large subunit protein (MDH); and to photosynthesis in Cyanobacteria, the (G) alpha and (H) beta chains of phycoerythrin (PEa and PEb). Branches were collapsed for a simplified view and named according to the lowest common taxon. Bootstrap values were calculated based on 1000 replications. The scale bars represent the average number of amino acid substitution per site. A list of the orthologous sequences used in the phylogenetic trees is found in the Supporting Information Table S3.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

orthologues from the genus *Synechococcus* than to other cyanobacterial orthologues (Fig. 4G–H).

Discussion

Omics-based approaches have been very successful in revealing important insights into the functional basis of the sponge–microbe association, including primary (Radax *et al.*, 2012b; Kamke *et al.*, 2013) and secondary microbial metabolism (Wilson *et al.*, 2014), as well as into putative mechanisms at the physical sponge–symbiont interphase (Thomas *et al.*, 2010; Nguyen *et al.*, 2014). In the present study, we analysed the functional and taxonomic classification of prokaryotic transcribed genes derived from the Red Sea sponge *S. carteri*. The most abundant transcripts were related to recycling of metabolites, such as ammonia and CO_2, originating from the host sponge. Our findings expand the current knowledge of relevant functions of the sponge microbiome by revealing known (i.e. archaeal ammonia oxidation, cyanobacterial carbon fixation, stress-related functions) and novel (i.e. methylotrophy) functions of tropical reef sponge symbionts.

S. carteri is an LMA sponge (Giles *et al.*, 2013), whose microbiome is rich but uneven, in the sense that a high number of ribotypes was identified of which only few were abundant (Lee *et al.*, 2011; Moitinho-Silva *et al.*, 2014). The LMA sponges represent less complex models for the study of microbial function in comparison to the high-microbial abundance sponges, which contain dense and diverse microbial communities (Weisz *et al.*, 2007; Kamke *et al.*, 2010; Schläppy *et al.*, 2010; Ribes *et al.*, 2012; Schmitt *et al.*, 2012; Giles *et al.*, 2013). On the other hand, the study of LMA microbiomes can be challenging due to the difficulties in obtaining sufficient microbial biomass. For example, only 6% of the volume of the LMA sponge *Dysidea avara* was occupied by microbes (Ribes *et al.*, 2012). Our strategy for enrichment of prokaryotic mRNA consisted of subtracting eukaryotic mRNA and rRNA from total host-symbiont RNA, which allowed us to produce metatranscriptomes that were enriched in prokaryotic functional gene transcripts. This strategy is similar to the protocol from Hampton-Marcell and colleagues (2013), and it should be applicable for the isolation of any host-associated microbial metatranscriptomes. Furthermore, the isolated eukaryotic mRNA would be suitable for sequencing, making it possible to obtain complementary host and symbiont transcriptome pairs.

At higher taxonomic levels, the composition of the *S. carteri* metatranscriptomes was in good agreement with previous data based on 16S rRNA gene sequencing in that it showed a dominance of Proteobacteria as well as an abundance of Cyanobacteria, Bacteroidetes, Nitrospirae, Actinobacteria (Lee *et al.*, 2011; Giles *et al.*, 2013; Moitinho-Silva *et al.*, 2014) and Crenarchaeota (Lee *et al.*, 2011), which were recently renamed to Thaumarchaota (Brochier-Armanet *et al.*, 2008). Of these, only Nitrospirae transcripts were present at rare abundances in the *S. carteri* metatranscriptomes. While Gammaproteobacteria is the most prominent proteobacterial class in the *S. carteri* microbiome by 16S rRNA amplicon sequencing (Lee *et al.*, 2011; Giles *et al.*, 2013; Moitinho-Silva *et al.*, 2014), the analysis of the metatranscriptomes resulted in similar proportions of Alphaproteobacteria and Gammaproteobacteria, each representing ~ 9% of the transcriptional features. These percentages should however not be over-interpreted as about half of the proteobacterial reads could not be placed into a given class (Fig. 1). Furthermore, approximately 40% of the putative prokaryotic transcriptional features were functionally annotated (Table 1). The partial identification and shallow depth of the taxonomic assignment was previously observed in the *Geodia barretti* metatranscriptome (Radax *et al.*, 2012b). This is likely a consequence of the lack of reference genomes and limited sequence length, where greater lengths are expected to provide better taxonomic resolution (Huson *et al.*, 2007).

Ammonia oxidation

Nitrogen metabolism is a major theme in sponge microbiology (Taylor *et al.*, 2007). Ammonia is a common metabolic waste of marine invertebrates (Wang and Douglas, 1998; Davy *et al.*, 2002) and serves as a source of nutrients and energy to sponge symbionts. Accordingly, archaeal ammonia monooxygenase was among the most highly expressed features in the *S. carteri* metatranscriptomes. These results are in agreement with previous findings that archaea dominate the expression of ammonia monooxygenase in different environments, including soil (Leininger *et al.*, 2006), seawater (Shi *et al.*, 2011) and cold water marine sponges (Radax *et al.*, 2012a,b). The archaeal population of *S. carteri* further highly expressed copper-containing nitrite reductase, similar to what has been observed in *G. barretti* where a homologous enzyme was found to be highly transcribed (Radax *et al.*, 2012b). The expression of these genes is in good agreement with the proposed pathway for ammonia oxidation in archaea (Stahl and de la Torre, 2012), where the reduction of nitrite by a copper-dependent nitrite reductase produces NO that serves as electron source for the ammonia monooxygenase. Therefore, archaeal ammonia oxidation in *S. carteri* appears most likely, although the inference of this reaction based solely on the ammonia monooxygenase gene homology and expression may not necessarily be correct (Mussmann *et al.*, 2011; Sayavedra-Soto *et al.*, 2011).

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

In marine sponges, archaeal ammonia oxidation is commonly reported in association with bacterial nitrite oxidation, in a process known as nitrification (Diaz and Ward, 1997; Jiménez and Ribes, 2007; Bayer *et al.*, 2008; Schläppy *et al.*, 2010). Members of the *Nitrospira*, a genus from the nitrite-oxidizing bacterial phylum Nitrospirae, were previously reported to be abundant, enriched and transcriptionally active in *S. carteri* (Moitinho-Silva *et al.*, 2014). However, only a minor portion of the prokaryotic transcriptional features were assigned to Nitrospirae (< 0.03; Supporting Information Table S1) in this study. Interestingly, low abundances of Nitrospirae transcripts were also observed in sludge microbial communities where Nitrospirae genes were abundant (Yu and Zhang, 2012). While the low contribution of Nitrospirae to the *S. carteri* metatranscriptome may indicate low *in situ* activity, it may also reflect the overall transcriptional capacity of this particular phylum. Alternatively, if the proteins or mRNAs are very stable in Nitrospirae, a low transcriptional level may be observed.

Methylotrophy

The high expression of formate dehydrogenase-O and methanol dehydrogenase subunits in the *S. carteri* metatranscriptomes indicates the capability of *S. carteri* symbionts for methylotrophy, which is defined as the ability to acquire energy and biomass from reduced C1 compounds (Chistoserdova *et al.*, 2009). Phylogenetic analyses placed the formate dehydrogenase-O, major subunit and methanol dehydrogenase large subunit protein transcripts close to gammaproteobacterial orthologues, members of which have been previously identified as symbionts of *S. carteri* (Lee *et al.*, 2011; Giles *et al.*, 2013; Moitinho-Silva *et al.*, 2014). The encoded enzymes oxidize formate and methanol and carry out important reactions of methylotrophic metabolism (Anthony, 2004; Chistoserdova *et al.*, 2007; 2009). Accordingly, the high expression of the decarboxylase and methylmalonyl-CoA mutase in *S. carteri* metatranscriptomes (Fig. 3, indexes 26 and 27) that are classified in the C1 assimilatory serine-glyoxylate cycle (Supporting Information Table S2) provides strong evidence for methylotrophic metabolism of the *S. carteri* symbionts. Hence, with the exception of formaldehyde metabolism genes which were identified at low percentage (< 0.2%; Supporting Information Fig. S1), components of all methylotrophy-related modules (Chistoserdova, 2011) were present among the most transcribed functions of the *S. carteri* metatranscriptomes. To date, the association between sponges and methylotrophic bacteria was only described for deep-sea carnivorous sponges (Vacelet *et al.*, 1995; Vacelet and Boury-Esnault, 2002). Our results provide the first evidence of C1 oxidizing symbionts in sponges from a tropical reef. If experimentally confirmed, this novel finding would expand our understanding of the sponge-microbial symbioses, which may have implications for carbon fluxes in the reef ecosystem.

Membrane transporters

The gene functions related to transporter systems of low molecular weight compounds were highly expressed by the proteobacterial population in *S. carteri*. These genes were previously found to be over-represented in the metagenome of the sponge *Cymbastela concentrica* (Thomas *et al.*, 2010), and the related proteins were observed in high abundance in the *C. concentrica* metaproteome, including the LivK from high-affinity leucine-specific transport system and the glycerol-3-P ABC-type transporter UgpB (Liu *et al.*, 2012). The high expression of genes encoding for transporter systems supports the importance of nutrient uptake by sponge symbionts. Indeed, an efficient uptake of dissolved organic matter has been demonstrated in sponges (Reiswig, 1974; Yahel *et al.*, 2003; Weisz *et al.*, 2007; Ribes *et al.*, 2012; de Goeij *et al.*, 2013).

Environmental stress

Marine sponges and their microbial symbionts are likely submitted to various environmental stressors, such as changes in temperature (Webster *et al.*, 2008; Cebrian *et al.*, 2011), UV irradiation (Batel *et al.*, 1998), exposure to reactive oxygen species originating from symbiont photosynthesis (Regoli *et al.*, 2000), disease (Webster, 2007; Luter *et al.*, 2010; Angermeier *et al.*, 2011) and anthropogenic effects, such as increased nutrient and heavy metal concentrations (Simister *et al.*, 2012a). Furthermore, stress related to the sampling strategy cannot be excluded, due to the rapid turnover of mRNA inventories in response to environmental conditions as reviewed by Moran *et al.* (2013). Our results suggest that the proteobacterial symbionts of *S. carteri* have mechanisms in place to cope with stress. The *S. carteri* proteobacterial population showed a high expression of rubrerythrin, which was proposed to be involved in an anaerobic detoxification pathway for oxidative stress (Sztukowska *et al.*, 2002; Weinberg *et al.*, 2004). Furthermore, the high expression by *S. carteri* proteobacterial symbionts of the DnaK/DnaJ and GroES/GroEL chaperone systems, and gene functions related to intracellular protein degradation suggest the capacity to deal with stress-induced protein misfolding (Mogk *et al.*, 2011). In agreement with our results, a high abundance of stress-related genes was

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

found in the metagenome and metaproteome of *C. concentrica* (Thomas *et al.*, 2010; Liu *et al.*, 2012). Sponges were collected at water temperatures of 30°C, which is above the average winter temperature for this location (Raitsos *et al.*, 2013), thus the expression of stress-related genes might also be related to environmental conditions.

Photosynthesis and CO_2 *fixation*

Most of the abundant gene functions assigned to Cyanobacteria in *S. carteri* metatranscriptomes were related to photosynthesis and CO_2 fixation. The majority of Cyanobacteria possess phycobilisomes, which are supramolecular light-harvesting complexes that serve as the primary antenna of photosystem II (Ting *et al.*, 2002). Phycobiliproteins, e.g. phycocyanin and phycoerythrin, are components of the phycobilisomes (Six *et al.*, 2005) and were highly expressed by the cyanobacterial population of *S. carteri*. Phylogenetic analysis of the representative sequences of alpha and beta chains of phycoerythrin (standing for 18% and 15% of the sequences assigned to the respective genes) revealed a close proximity of the expressed genes to orthologues of members from the genus *Synechococcus*. Members of the *Synechococcus* clade were found to be abundant and transcriptionally active in *S. carteri* sponges and in the surrounding seawater (Moitinho-Silva *et al.*, 2014). Our results corroborate the proposed primary symbiotic role of sponge cyanobacteria as sources of fixed carbon for their hosts (Wilkinson, 1983; Arillo *et al.*, 1993; Taylor *et al.*, 2007).

Other functions encoded by symbionts

A TonB-dependent receptor was among the highly transcribed genes affiliated with the phylum Bacteroidetes. TonB-dependent receptors are a family of energy-dependent outer membrane transport proteins (Bradbeer, 1993; Ferguson and Deisenhofer, 2002). Their substrates include iron siderophores, cobalamins (vitamin B12), copper ions, maltodextrin and sucrose (Schauer *et al.*, 2008). TonB-dependent receptors genes are abundant in Bacteroidetes genomes (Fernández-Gómez *et al.*, 2013), and the high expression of these genes observed in *S. carteri* metatranscriptomes confirms their relevance for members of this phylum. With regard to Actinobacteria, most of the features were classified as DNA gyrase subunit A, the relevance of which remains to be investigated.

Functions encoded by seawater bacteria

It was previously estimated that one quarter of the total recovered 16S rRNA gene sequences from *S. carteri* specimens were derived from seawater bacteria (Moitinho-Silva *et al.*, 2014). Therefore, bacterioplankton transcripts were to be expected in the *S. carteri* metatranscriptomes. As an example, the phylogenetic placement of the acetate permease ActP (6% of the reads assigned to the function) and TRAP-type C4-dicarboxylate transport system, periplasmic component (5%) sequences close to orthologues from alphaproteobacterial lineages, including members of the *Pelagibacter* genus, which are part of the highly abundant marine clade SAR11 (Morris *et al.*, 2002), were identified. These findings suggest that seawater microorganisms may continue to be functionally active when inside sponges.

Conclusions

The most expressed gene functions of the *S. carteri* microbiomes were characterized at the transcriptional level and with respect to their taxonomic affiliation. Our analysis revealed the functional processes of sponge symbionts, such as ammonia oxidation by Archaea and photosynthetic carbon fixation by Cyanobacteria, under *in situ* conditions. A high expression of previously reported functions relating to stress and membrane transport was further observed in the *S. carteri* microbiomes. Unexpectedly, methylotrophy-related gene functions were highly expressed by *S. carteri* symbionts and were phylogenetically associated with the dominant gammaproteobacterial population. Our work demonstrates the usefulness of metatranscriptomics to unravel microbial functions within the sponge microcosm.

Experimental procedures

Sample collection

Stylissa carteri specimens (three biological replicates) were collected by SCUBA diving at Fsar reef (22°23′N; 39°03′E), Saudi Arabia. The collection occurred on the morning of 2 November 2010, at 13 to 14 meters depth and at a water temperature of 30°C. Sponges were collected and processed individually as described by Moitinho-Silva *et al.* (2014). Freshly collected sponge pieces (5 ml) were incubated in falcon tubes containing 30 ml RNAlater (Ambion, USA) overnight at 4°C, according to the manufacturer's instructions, and stored at −80°C.

Nucleic acids extraction and prokaryotic mRNA enrichment

Total DNA and RNA were co-extracted from sponge pieces that had been preserved in RNAlater. Sponge pieces of 200 mg wet weight were added to *Lysing Matrix E* tubes (MP Biomedicals, USA) containing 600 µl of 1% 2-mercaptoethanol RLT buffer (Qiagen, Germany). In order to

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

obtain sufficient amounts of RNA, 10–15 extractions were carried out per replicate. The cells were mechanically disrupted by use of the homogenizer FastPrep Instrument (MP Biomedicals, USA) for 30 s at a speed setting of 5.5. Total nucleic acids isolation, nuclease treatment and analyses were performed according to Moitinho-Silva *et al.* (2014). The isolated RNA was maintained in nuclease-free water with the RNAse inhibitor SUPERase-In (1 U μl^{-1}) (Ambion, USA). The integrity of total RNA was inferred using the Experion System (Bio-Rad, USA). Isolated RNA that showed signs of degradation was excluded from further processing. RNA resulting from multiple extractions was pooled per biological replicate according to Poly(A)Purist MAG Kit (Ambion, USA) instructions.

To allow comparison with other metatranscriptomic studies (Moitinho-Silva unpublished data), two internal standards were added to the lysing tubes (15 ng each) prior to cell disruption. The internal standards were artificial mRNAs produced by *in vitro* transcription of two vectors (Moran *et al.*, 2013; Satinsky *et al.*, 2013), pTXB1 Vector (New England Biolabs, USA) and pFN18A HaloTag T7 Flexi Vector (Promega, USA).

In order to enrich for prokaryotic mRNA, eukaryotic mRNA and rRNA were subtracted from total RNA using two RNA sorting protocols based on magnetic beads: Poly(A)Purist MAG Kit (Ambion, USA) and the universal rRNA-subtraction protocol developed by Stewart (2013). In the first protocol, the eukaryotic mRNA was subtracted from total RNA (100 μg) by two rounds of purification using the Poly(A)Purist MAG Kit (Ambion, USA). After each round of magnetic beads capture, the non-poly(A) RNA fraction (the fraction that contained rRNA and prokaryotic mRNA) was recovered from the Binding Solution using the RNeasy MinElute Cleanup Kit (QIAGEN, Germany). The recovered RNA served as starting RNA template for the second protocol, which aimed at the removal of the rRNA subunits of eukaryotes (18S and 28S), bacteria (16S and 23S) and archaea (16S and 23S) according to Stewart *et al.* (2010). For this purpose, co-extracted DNA from the three biological replicates was pooled in a single tube. The pooled DNA was used as template for polymerase chain reaction (PCR) amplification of rRNA subunit genes with DreamTaq DNA Polymerase (Fermentas, Germany). PCR was performed individually for each rRNA subunit gene. The amplified rRNA genes were used as template for synthesis of biotin-labeled anti-sense RNA. The subtraction of rRNA was performed by hybridization to biotin-labeled anti-sense RNA (800 ng of each probe per 400 ng of starting RNA template) followed by streptavidin-coated magnetic bead binding and magnetic separation. The enriched prokaryotic mRNA (8.5 ng) was amplified with MessageAmp II-Bacteria kit (Ambion, USA) according to manufacturer's recommendations. Amplified enriched prokaryotic mRNA was reverse-transcribed to cDNA and sequenced by the KAUST Genomics Core Lab using Illumina HiSeq 2000 standard protocols. The sequencing resulted in paired-end reads (101 bp) with an estimated mean insert size of 280 bp. Mate-pair information was not considered during the annotation pipeline. Raw Illumina reads were deposited under the National Center for Biotechnology Information Sequence Read Archive (NCBI SRA) accession number SRP033297.

Taxonomic and functional annotation of transcripts

The raw Illumina reads containing low quality bases (Phred score < 20, which correspond to error probability 0.01) were truncated to the first base below the cut-off. Sequencing adapters, including partial adapters, were trimmed from raw reads. Resulting read pairs were removed using custom java scripts when containing short reads (< 16 bp). Processed reads were submitted to MG-RAST v. 3.3.6.1 (Meyer *et al.*, 2008) for further processing and annotation using the standard pipeline, with addition of the options for removal of low quality reads (Phred score < 20) and artificial replicates (Gomez-Alvarez *et al.*, 2009). Processed Illumina reads and analyses input files are accessible at MG-RAST under the project ID mgp3802 (static link http://metagenomics.anl.gov/linkin.cgi?project=3802). The quality-processed, de-replicated and screened nucleotide sequences (screen.passed.fna files) were mapped against the two internal standards (see previous section of 'Experimental procedures') using GENEIOUS v 6.0.6 (http://www.geneious.com/) with default parameters. Reads mapped against any of the internal standards were flagged for removal in this study but will be useful for comparison to additional metatranscriptomic data sets that have been generated (Moitinho-Silva, unpublished). Identified protein features (hereafter named 'transcriptional features') were annotated based on their md5 values (identifier for a database hit) that resulted from BLAT searches. For that, the possible md5 values annotations to a given domain of life (Eukaryota, Bacteria, Archaea) were first retrieved from MG-RAST workbench, using M5NR as annotation source, with stringency parameters set to minimum values (maximum e-value of 1, a minimum identity of 1%, and a minimum alignment length of 1). Transcriptional features were taxonomically classified into a domain of life according to the best BLAT hit (e-value < 0.001) against all MG-RAST protein databases (superblat.expand.protein files). In case of annotation conflicts between databases, the accepted classification was the one supported by the majority of the databases. Transcriptional features that resulted in two or more domains with the same number of supporting databases were classified as tied between these domains. Transcriptional features classified to Bacteria, to Archaea or as tied between Bacteria and Archaea were retained for taxonomic classification as provided by the MG-RAST based on the LCA. The MG-RAST-based LCA classification of transcriptional features was only considered when supported by at least four hits (BLAT e-value < 0.001, superblat.expand.lca files). Transcriptional features that were taxonomically classified as eukaryotes or as viruses, or whose reads were flagged as internal standards, were removed. Features that were not classified by MG-RAST-based LCA were considered as unidentified prokaryotic features. Functional classification of prokaryotic transcriptional features was done based on SEED subsystems (BLAT e-value < 0.001, 'superblat.expand.ontology' files) (Overbeek *et al.*, 2005). For that, the possible annotations of md5 values to SEED subsystems were retrieved from MG-RAST workbench using the minimum stringency values. Transcriptional features matching the gene function 'Retron-type reverse transcriptase' were removed, because they were regarded as putative rRNA

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

sequences that had been mis-annotated (Tripp *et al*., 2011). The remaining prokaryotic transcriptional features were considered for analysis.

A matrix containing the gene functions and their prokaryotic transcriptional feature counts in each *S. carteri* replicate was used to calculate Bray–Curtis dissimilarity index (Bray and Curtis, 1957) as in Moitinho-Silva *et al*. (2014).

Phylogeny of selected key gene functions

For each gene function selected, the original screened reads were retrieved (screen.passed.fna files) and *de novo* assembled using GENEIOUS v. 6.0.6 created by Biomatters (http://www.geneious.com). Assembly was carried out using standard parameters, which allowed for 15% maximum gaps per read, a maximum gap size of 2 and 30% of maximum mismatches per read. Match word length was 14 bases, and index word length was set to 12. To check for the possibility of chimera assembly, contigs were also assembled under stringent parameters (i.e. no gaps allowed, 1% of maximum mismatches per read and 100% of overlap between the reads). While this procedure gave similar results, less reads could be assigned. Furthermore, reduced bootstrap values for phylogenetic association were computed. Our analysis indicates that the results are not prone to mistakes by chimeric assembly.

The representative nucleotide consensus sequence was recovered from the contig that contained the greatest number of reads. Deduced amino acid sequences were generated by translating the representative sequences using the frame indicated by the BLASTX best hit against NCBI non-redundant protein database. Orthologous amino acid sequences were retrieved from best BLAST hits against UniProt Knowledgebase (http://www.uniprot.org). For copper-containing nitrite reductase and methanol dehydrogenase large subunit protein, orthologous sequences of selected taxa were also retrieved (Bacteria for the first; and *Alpha*-, *Beta*- and *Gammaproteobacteria* for the latter). Multiple alignments of representative sequences and orthologues were generated using MUSCLE (Edgar, 2004) and cured using GBLOCKS (Castresana, 2000) as implemented in Phylogeny.fr (Dereeper *et al*., 2008). MUSCLE was used in its full mode and Gblocks was performed using the less stringent settings in Phylogeny.fr. Substitution models were selected using MODELGENERATOR v. 0.85 (Keane *et al*., 2006). Maximum-likelihood phylogenetic trees were constructed using PHYML v. 3.0 (Guindon *et al*., 2010). Bootstrap values were calculated with 1000 replicates. Phylogenetic trees were edited using FIGTREE v. 1.4 (http://tree.bio.ed.ac.uk/software/figtree/) and INKSCAPE v. 0.48.4 (http://www.inkscape.org/).

Acknowledgements

We thank the Biosciences Core Laboratory at KAUST for support with sequencing, as well as the Coastal and Marine Resources Core Lab and Prof M. Berumen at KAUST for support with sample collection. B. Neumann (Depart. of Botany I, University of Wuerzburg) is acknowledged for the help with the Experion System. L.M.-S. was supported by a grant of the German Excellence Initiative to the Graduate School of Life Sciences, University of Wuerzburg. Additional financial support was provided by KAUST baseline funds to T.Ra.

References

Angermeier, H., Kamke, J., Abdelmohsen, U.R., Krohne, G., Pawlik, J.R., Lindquist, N.L., and Hentschel, U. (2011) The pathology of sponge orange band disease affecting the Caribbean barrel sponge *Xestospongia muta*. *FEMS Microbiol Ecol* **75:** 218–230.

Anthony, C. (2004) The quinoprotein dehydrogenases for methanol and glucose. *Arch Biochem Biophys* **428:** 2–9.

Arillo, A., Bavestrello, G., Burlando, B., and Sarà, M. (1993) Metabolic integration between symbiotic cyanobacteria and sponges: a possible mechanism. *Mar Biol* **117:** 159–162.

Batel, R., Fafandjel, M., Blumbach, B., Schröder, H.C., Hassanein, H.M.A., Müller, I.M., and Müller, W.E.G. (1998) Expression of the human XPB/ERCC-3 excision repair gene-homolog in the sponge *Geodia cydonium* after exposure to ultraviolet radiation. *Mutat Res* **409:** 123–133.

Bayer, K., Schmitt, S., and 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.

Bell, J.J. (2008) The functional roles of marine sponges. *Estuarine Coast Shelf Sci* **79:** 341–353.

Bradbeer, C. (1993) The proton motive force drives the outer membrane transport of cobalamin in *Escherichia coli*. *J Bacteriol* **175:** 3146–3150.

Bray, J.R., and Curtis, J.T. (1957) An ordination of the upland forest communities of southern Wisconsin. *Ecol Monogr* **27:** 325–349.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., and Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nat Rev Microbiol* **6:** 245–252.

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

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

Chistoserdova, L. (2011) Modularity of methylotrophy, revisited. *Environ Microbiol* **13:** 2603–2622.

Chistoserdova, L., Crowther, G.J., Vorholt, J.A., Skovran, E., Portais, J.-C., and Lidstrom, M.E. (2007) Identification of a fourth formate dehydrogenase in *Methylobacterium extorquens* AM1 and confirmation of the essential role of formate oxidation in methylotrophy. *J Bacteriol* **189:** 9076–9081.

Chistoserdova, L., Kalyuzhnaya, M.G., and Lidstrom, M.E. (2009) The expanding world of methylotrophic metabolism. *Annu Rev Microbiol* **63:** 477–499.

Davy, S.K., Trautman, D.A., Borowitzka, M.A., and Hinde, R. (2002) Ammonium excretion by a symbiotic sponge supplies the nitrogen requirements of its rhodophyte partner. *J Exp Biol* **205:** 3505–3511.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

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

Diaz, M., and Ward, B. (1997) Sponge-mediated nitrification in tropical benthic communities. *Mar Ecol Prog Ser* **156:** 97–107.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32:** 1792–1797.

Ferguson, A.D., and Deisenhofer, J. (2002) TonB-dependent receptors-structural perspectives. *Biochim Biophys Acta* **1565:** 318–332.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedrós-Alió, C. (2013) Ecology of marine Bacteroidetes: a comparative genomics approach. *ISME J* **7:** 1026–1037.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., and Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105:** 3805–3810.

Giles, E.C., Kamke, J., Moitinho-Silva, L., Taylor, M.W., Hentschel, U., Ravasi, T., and Schmitt, S. (2013) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiol Ecol* **83:** 232–241.

de Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., and Admiraal, W. (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342:** 108–110.

Gomez-Alvarez, V., Teal, T.K., and Schmidt, T.M. (2009) Systematic artifacts in metagenomes from complex microbial communities. *ISME J* **3:** 1314–1317.

Gosalbes, M.J., Durbán, A., Pignatelli, M., Abellan, J.J., Jiménez-Hernández, N., Pérez-Cobas, A.E., *et al.* (2011) Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS ONE* **6:** e17447.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59:** 307–321.

Hampton-Marcell, J.T., Moormann, S.M., Owens, S.M., and Gilbert, J.A. (2013) Preparation and metatranscriptomic analyses of host-microbe systems. *Methods Enzymol* **531:** 169–185.

Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J., and Moore, B.S. (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl Environ Microbiol* **68:** 4431–4440.

Hentschel, U., Piel, J., Degnan, S.M., and Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* **10:** 641–654.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17:** 377–386.

Jiménez, E., and Ribes, M. (2007) Sponges as a source of dissolved inorganic nitrogen: nitrification mediated by temperate sponges. *Limnol Oceanogr* **52:** 948–958.

Kamke, J., Taylor, M.W., and Schmitt, S. (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *ISME J* **4:** 498–508.

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K., *et al.* (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* **7:** 2287–2300.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and McLnerney, J.O. (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol Biol* **6:** 29.

Kleiner, M., Wentrup, C., Lott, C., Teeling, H., Wetzel, S., Young, J., *et al.* (2012) Metaproteomics of a gutless marine worm and its symbiotic microbial community reveal unusual pathways for carbon and energy use. *Proc Natl Acad Sci USA* **109:** E1173–E1182.

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

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., *et al.* (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806–809.

Liu, M., Fan, L., Zhong, L., Kjelleberg, S., and Thomas, T. (2012) Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* **6:** 1515–1525.

Luter, H.M., Whalan, S., and Webster, N.S. (2010) Exploring the role of microorganisms in the disease-like syndrome affecting the sponge Ianthella basta. *Appl Environ Microbiol* **76:** 5736–5744.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., 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.

Mogk, A., Huber, D., and Bukau, B. (2011) Integrating protein homeostasis strategies in prokaryotes. *Cold Spring Harb Perspect Biol* **3:** a004366.

Mohamed, N.M., Colman, A.S., Tal, Y., and Hill, R.T. (2008) Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environ Microbiol* **10:** 2910–2921.

Moitinho-Silva, L., Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., *et al.* (2014) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Mol Ecol* **23:** 1348–1363.

Moran, M.A. (2009) Metatranscriptomics: eavesdropping on complex microbial communities. *Microbe* **4:** 329–335.

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.-K., *et al.* (2013) Sizing up metatranscriptomics. *ISME J* **7:** 237–243.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., and Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* **420:** 806–810.

Mussmann, M., Brito, I., Pitcher, A., Sinninghe Damsté, J.S., Hatzenpichler, R., Richter, A., *et al.* (2011) Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci USA* **108:** 16771–16776.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Ngugi, D.K., Antunes, A., Brune, A., and Stingl, U. (2012) Biogeography of pelagic bacterioplankton across an antagonistic temperature-salinity gradient in the Red Sea. *Mol Ecol* **21:** 388–405.

Nguyen, M.T.H.D., Liu, M., and Thomas, T. (2014) Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol Ecol* **23:** 1635–1645.

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.Y., 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.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., and Moran, M.A. (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environ Microbiol* **11:** 1358–1375.

Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., and Schleper, C. (2012a) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges: archaeal nitrification in cold-water sponges. *Environ Microbiol* **14:** 909–923.

Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H.T., Urich, T., and Schleper, C. (2012b) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environ Microbiol* **14:** 1308–1324.

Raitsos, D.E., Pradhan, Y., Brewin, R.J., Stenchikov, G., and Hoteit, I. (2013) Remote sensing the phytoplankton seasonal succession of the Red Sea. *PLoS ONE* **8:** e64909.

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

Reiswig, H.M. (1974) Water transport, respiration and energetics of three tropical marine sponges. *J Exp Mar Bio Ecol* **14:** 231–249.

Ribes, M., Jimenez, E., Yahel, G., Lopez-Sendino, P., Diez, B., Massana, R., *et al*. (2012) Functional convergence of microbes associated with temperate marine sponges. *Environ Microbiol* **14:** 1224–1239.

Sanders, J.G., Beinart, R.A., Stewart, F.J., Delong, E.F., and Girguis, P.R. (2013) Metatranscriptomics reveal differences in in situ energy and nitrogen metabolism among hydrothermal vent snail symbionts. *ISME J* **7:** 1556–1567.

Satinsky, B.M., Gifford, S.M., Crump, B.C., and Moran, M.A. (2013) Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods Enzymol* **531:** 237–250.

Sayavedra-Soto, L.A., Hamamura, N., Liu, C.-W., Kimbrel, J.A., Chang, J.H., and Arp, D.J. (2011) The membrane-associated monooxygenase in the butane-oxidizing Gram-positive bacterium *Nocardioides* sp. strain CF8 is a novel member of the AMO/PMO family. *Environ Microbiol Rep* **3:** 390–396.

Schauer, K., Rodionov, D.A., and 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.

Schläppy, M.-L., Schöttner, S.I., Lavik, G., Kuypers, M.M.M., Beer, D., and Hoffmann, F. (2010) Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Mar Biol* **157:** 593–602.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., *et al*. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6:** 564–576.

Shi, Y., Tyson, G.W., Eppley, J.M., and DeLong, E.F. (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J* **5:** 999–1013.

Simister, R.L., Taylor, M.W., Tsai, P., and Webster, N. (2012a) Sponge-microbe associations survive high nutrients and temperatures. *PLoS ONE* **7:** e52220.

Simister, R.L., Deines, P., Botte, E.S., Webster, N.S., and Taylor, M.W. (2012b) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ Microbiol* **14:** 517–524.

Six, C., Thomas, J.-C., Thion, L., Lemoine, Y., Zal, F., and Partensky, F. (2005) Two novel phycoerythrin-associated linker proteins in the marine cyanobacterium *Synechococcus* sp. strain WH8102. *J Bacteriol* **187:** 1685–1694.

Stahl, D.A., and de la Torre, J.R. (2012) Physiology and diversity of ammonia-oxidizing archaea. *Annu Rev Microbiol* **66:** 83–101.

Stewart, F.J. (2013) Preparation of microbial community cDNA for metatranscriptomic analysis in marine plankton. *Methods Enzymol* **531:** 187–218.

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

Sztukowska, M., Bugno, M., Potempa, J., Travis, J., and Kurtz, D.M., Jr (2002) Role of rubrerythrin in the oxidative stress response of *Porphyromonas gingivalis*. *Mol Microbiol* **44:** 479–488.

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

Taylor, M.W., Tsai, P., Simister, R.L., Deines, P., Botte, E., Ericson, G., *et al*. (2013) 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME J* **7:** 438–443.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., *et al*. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336:** 608–611.

Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., 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.

Ting, C.S., Rocap, G., King, J., and Chisholm, S.W. (2002) Cyanobacterial photosynthesis in the oceans: the origins and significance of divergent light-harvesting strategies. *Trends Microbiol* **10:** 134–142.

Tripp, H.J., Hewson, I., Boyarsky, S., Stuart, J.M., and Zehr, J.P. (2011) Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. *Nucleic Acids Res* **39:** 8792–8802.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., and Schuster, S.C. (2008) Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* **3:** e2527.

Vacelet, J., and Boury-Esnault, N. (2002) A new species of carnivorous deep-sea sponge (Demospongiae: Cladorhizidae) associated with methanotrophic bacteria. *Cahiers de Biologie Marine* **43:** 141–148.

Vacelet, J., Boury-Esnault, N., Fiala-Medioni, A., and Fisher, C.R. (1995) A methanotrophic carnivorous sponge. *Nature* **377:** 296–296.

Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., *et al.* (2009) Shotgun metaproteomics of the human distal gut microbiota. *ISME J* **3:** 179–189.

Wang, J., and Douglas, A.E. (1998) Nitrogen recycling or nitrogen conservation in an alga-invertebrate symbiosis? *J Exp Biol* **201:** 2445–2453.

Webster, N.S. (2007) Sponge disease: a global threat? *Environ Microbiol* **9:** 1363–1375.

Webster, N.S., Cobb, R.E., and Negri, A.P. (2008) Temperature thresholds for bacterial symbiosis with a sponge. *ISME J* **2:** 830–842.

Webster, N.S., Taylor, M.W., Behnam, F., Lucker, S., Rattei, T., Whalan, S., *et al.* (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12:** 2070–2082.

Weinberg, M.V., Jenney, F.E., Jr, Cui, X., and Adams, M.W. (2004) Rubrerythrin from the hyperthermophilic archaeon *Pyrococcus furiosus* is a rubredoxin-dependent, iron-containing peroxidase. *J Bacteriol* **186:** 7888–7895.

Weisz, J., Hentschel, U., Lindquist, N., and Martens, C. (2007) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Mar Biol* **152:** 475–483.

Wilkinson, C.R. (1983) Net primary productivity in coral reef sponges. *Science* **219:** 410–412.

Wilson, M.C., Mori, T., Ruckert, C., Uria, A.R., Helf, M.J., Takada, K., *et al.* (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506:** 58–62.

Yahel, G., Sharp, J.H., Marie, D., Häse, C., and Genin, A. (2003) In situ feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: bulk DOC is the major source for carbon. *Limnol Oceanogr* **48:** 141–149.

Yu, K., and Zhang, T. (2012) Metagenomic and metatranscriptomic analysis of microbial community structure and gene expression of activated sludge. *PLoS ONE* **7:** e38183.

Supporting information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Expressed gene functions involved in methylotrophy. The percentage of features ($n = 3$, mean ± standard error) that were assigned to methylotrophy-related genes are represented by two bars, one for total features and another for Proteobacteria. The background is coloured according to the major methylotrophy modules according to (Chistoserdova, 2011): (i) primary methylated substrate oxidation (red); (ii) formaldehyde handling (blue); (iii) formate dehydrogenase (yellow); and (iv), assimilatory module (serine-glyoxylate cycle), (green).

Table S1. Taxonomic assignment of *S. carteri* metatranscriptomes at phylum and class levels. Percentages in relation to the total annotated prokaryotic transcriptional features are shown for each *S. carteri* biological replicate.

Table S2. Functional and taxonomic classification of the top 100 most expressed gene functions in *S. carteri* metatranscriptomes. Percentages in relation to the total annotated prokaryotic transcriptional features are shown (mean, $n = 3$). Standard deviations are shown for the 'Transcriptional feature' column.

Table S3. Orthologues used for phylogenetic tree construction of selected key gene functions.

© 2014 Society for Applied Microbiology and John Wiley & Sons Ltd, *Environmental Microbiology*

4.6 Additional results

4.6.1 Clone libraries

In order to obtain a preliminary estimation of the microbial diversity within *S. carteri* and *X. testudinaria*, the 16S rRNA gene fragments were amplified using the primer pair 27f-1492r (see Section 3.2.4) from DNA samples pooled according to sponge species (Collection 1). Clones containing near full-length 16S rRNA genes were randomly selected and partially sequenced. This effort resulted in a total of 21 clones (27f-1492r clones) originating from each *S. carteri* and *X. testudinaria* (Table 13). According to similarity searches against the NCBI nucleotide database, all *S. carteri* clone sequences were closer to the sequences derived form *S. carteri* from the study of Giles *et al*. (2013a). This reflect the fact that in this work, where I am a co-author, the exact same samples and primers were used for microbial diversity analysis. The clone sequences that derived from *X. testudinaria* were similar to sequences mainly recovered from marine sponges and to a lesser extend from other sources, including corals (GI: 510829961 and 510829964) and other environments (GI: 392463973).

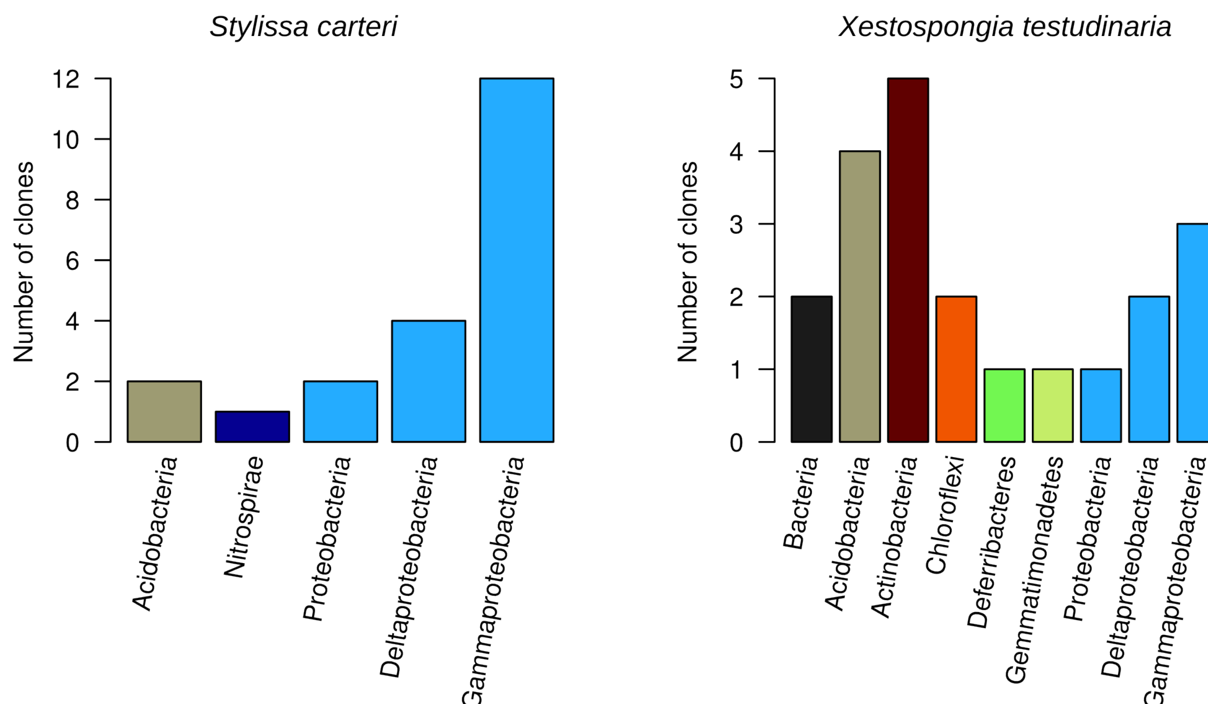


Figure 16. Taxonomic classification of the 27f-1492r clones. DNA from *S. carteri* or *X. testudinaria* was used to clone near full-length 16S rRNA gene sequences. The 27f-1492r clones were partially sequenced and classified according to taxonomy. Taxonomic classification is presented at the phylum level. When phylum level assignment was not supported, kingdom is reported. Proteobacteria are reported at the class level, when the class level assignment was possible.

The 27f-1492r clone sequences were classified according to taxonomy based on SILVA taxonomy (Figure 16). Most of the *S. carteri* clone sequences were classified as Proteobacteria (n=18), specifically Gammaproteobacteria (12). A small portion of *S. carteri* sequences was classified as Acidobacteria (2) or Nitrospirae (1). The *X. testudinaria* sequences were more diverse at phylum level, comprising 6 phyla. Proteobacteria (6), Actinobacteria (5), and Acidobacteria (4) were the phyla with most sequences classified to.

Table 13. The *S. carteri* and *X. testudinaria* 16S rRNA gene clones (27f-1492r) analysis

	BLAST results						Taxonomic classification
Clone	**length (bp)**	**Query GI**	**e-value**	**Align length**	**Ident matches (%)**	**Description**	**Phylum**[1]
SC_1	655	407912998	0	655	100	Unc bac clon F10SC 16S (PG)	Gammaproteobac
SC_10	510	407912979	0	510	99.8	Unc bac clon C12SC 16S (PG)	Gammaproteobac
SC_11	556	407912957	0	556	99.46	Unc bac clon A03SC 16S (PG)	Gammaproteobac
SC_12	298	407912989	3e-145	298	98.32	Unc bac clon G05SC 16S (PG)	Acidobac
SC_13	608	407912982	0	608	99.84	Unc bac clon B09SC 16S (PG)	Gammaproteobac
SC_14	638	407912982	0	638	99.84	Unc bac clon B09SC 16S (PG)	Gammaproteobac
SC_15	622	407912988	0	622	99.52	Unc bac clon F08SC 16S (PG)	Gammaproteobac
SC_16	609	407912982	0	609	100	Unc bac clon B09SC 16S (PG)	Gammaproteobac
SC_17	608	407913000	0	608	100	Unc bac clon A04SC 16S (PG)	Gammaproteobac
SC_18	629	407912973	0	624	100	Unc bac clon E08SC 16S (PG)	Gammaproteobac
SC_19	623	407912975	0	623	100	Unc bac clon G04SC 16S (PG)	Deltaproteobac
SC_2	566	407912957	0	566	99.47	Unc bac clon A03SC 16S (PG)	Gammaproteobac
SC_20	558	407912995	0	559	98.93	Unc bac clon B11SC 16S (PG)	Proteobac
SC_21	511	407912999	0	511	99.61	Unc bac clon G03SC 16S (PG)	Nitrospirae
SC_23	627	407912975	0	627	99.84	Unc bac clon G04SC 16S (PG)	Deltaproteobac
SC_3	424	407912995	0	424	99.53	Unc bac clon B11SC 16S (PG)	Proteobac
SC_4	653	407912975	0	653	100	Unc bac clon G04SC 16S (PG)	Deltaproteobac
SC_5	658	407912975	0	658	100	Unc bac clon G04SC 16S (PG)	Deltaproteobac
SC_6	526	407912989	0	526	100	Unc bac clon G05SC 16S (PG)	Acidobac
SC_7	655	407912998	0	655	100	Unc bac clon F10SC 16S (PG)	Gammaproteobac
SC_9	598	407912982	0	598	99.83	Unc bac clon B09SC 16S (PG)	Gammaproteobac
XT_100	610	290575668	0	609	99.01	Unc sponge bac clon JZ59-42 16S (PG)	Gemmatimonadetes
XT_102	600	350627589	0	606	94.88	Unc delta proteobacterium clon XD1G03 16S (PG)	Proteobac
XT_103	604	510829961	0	604	99.34	Unc bac clon 14F07 16S (PG)	Acidobac
XT_81	372	407728841	0	372	99.46	Unc bac clon TO10-922_C32 16S (PG)	Gammaproteobac
XT_82	578	510829964	0	578	99.83	Unc bac clon 14F10 16S (PG)	Actinobac
XT_83	601	255660752	0	601	99.5	Unc actinobacterium clon OPM51 16S (PG)	Actinobac
XT_84	625	290575665	0	587	97.61	Unc bac clon JZ59-37 16S (PG)	Deltaproteobac
XT_86	624	392463973	0	629	90.46	Unc bac clon BJGMM-1s-275 16S (PG)	Bac
XT_87	521	566553892	0	522	94.64	Unc endophytic bac partial 16S, isolate HsBG30	Gammaproteobac
XT_88	452	511630206	0	449	96.88	Unc bac partial 16S, clon B27/GW947	Gammaproteobac
XT_89	612	253509272	0	613	97.39	Unc Chloroflexus sp. clon XA3D12F 16S (PG)	Chloroflexi
XT_90	629	226443723	0	629	98.73	Unc Acidobac bac clon XA2F05F 16S (PG)	Acidobac
XT_91	628	290575665	0	635	88.19	Unc sponge bac clon JZ59-37 16S (PG)	Bac
XT_92	644	253509279	0	644	99.38	Unc Chloroflexus sp. clon XA3G09F 16S (PG)	Chloroflexi
XT_93	625	510829964	0	625	99.84	Unc bac clon 14F10 16S (PG)	Actinobac
XT_94	633	290575647	0	639	96.56	Unc sponge bac clon JZ59-16 16S (PG)	Acidobac
XT_95	564	210161915	0	565	99.82	Unc Gram-positive bac clon 277AV 16S (PG)	Actinobac
XT_96	639	295639253	0	639	99.84	Unc bac clon IS-67 16S (PG)	Deferribacteres
XT_97	533	290575649	0	533	98.5	Unc sponge bac clon JZ59-18 16S (PG)	Acidobac
XT_98	539	210161915	0	540	99.81	Unc Gram-positive bac clon 277AV 16S (PG)	Actinobac
XT_99	633	407728879	0	633	99.53	Unc bac clon TV10-97_C21 16S (PG)	Deltaproteobac

The *S. carteri* (SC_) and X. testudinaria (XT_) derived clones sequenced are reported. In order to fit to the page, words were abbreviated: **Ind**entical, **unc**ultured, **bac**teria/**bac**terium, **clon**e, **16S** rRNA gene. PG stands for partial gene.

[1]Taxonomic classification are reported at phylum level. When phylum level assignment was not supported, kingdom or root level is reported. Proteobacteria are reported at the class level, when the class level assignment was possible. Names finishing in bacteria were shortened to finish with “bac”.

To inspect for major microbial community shifts due to the use of the primer pair 533f-907r (see Section 3.2.4), clone libraries were constructed from partially amplified 16S rRNA

genes from DNA samples (Collection 1) pooled according to sponge species. Clones containing the V4-V5 hypervariable regions of the 16S rRNA genes were randomly selected and sequenced. This effort resulted in a total of 21 clones (533f-907r clones) originating from each *S. carteri* and *X. testudinaria* (Table 14). According to similarity searches against the NCBI nucleotide database, all but one (SC_54) of the derived clones from *S. carteri* were closer to the sequences deposited by (2013a). *X. testudinaria* clone sequences were similar to sequences mainly recovered from marine sponges. As exceptions, a single clone sequence was most similar to a sequence isolated from coral (clone XT_26, GI: 482678009), a clone sequence (XT_46) aligned partially to an eukaryotic alpha-tubulin gene, and a clone sequence (XT_48) did not result in a positive match against any of the nucleotide sequences at NCBI database.

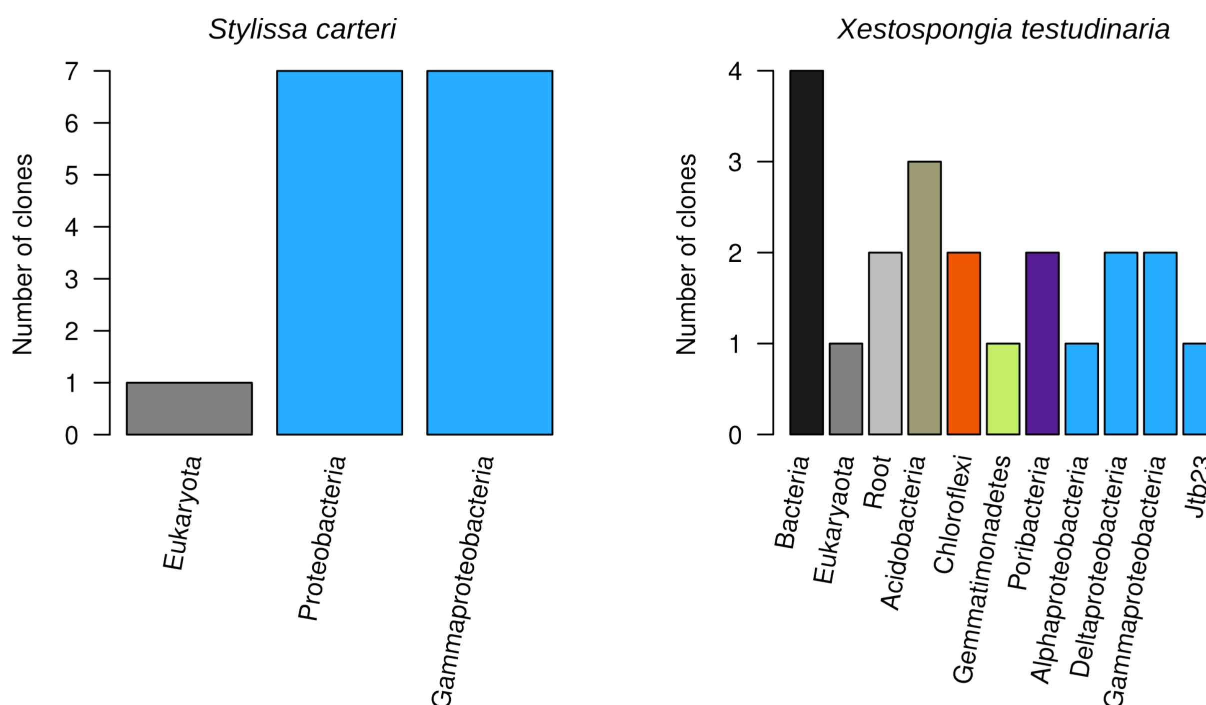


Figure 17. Taxonomic classification of the 533f-907r clones. DNA from *S. carteri* or *X. testudinaria* was used to clone a fragment containing the V4-V5 hypervariable regions of the 16S rRNA gene sequences. The 533f-907r clones were partially sequenced and classified according to taxonomy. Taxonomic classification is presented at the phylum level. When phylum level assignment was not supported, kingdom or Root is reported. Proteobacteria are reported at the class level, when the class level assignment was possible.

The 533f-907r clone sequences were classified according to taxonomy based on SILVA

taxonomy (Figure 17). Most of the *S. carteri* clone sequences were classified as Proteobacteria (n=14), specifically Gammaproteobacteria (7). A single clone sequence of *S. carteri* was classified as from eukaryotic origin, depite its close similarity to a gammaproteobacterium clone (SC_54, Table 14). The *X. testudinaria* sequences were more diverse at phylum level, where 5 phyla were identified. Proteobacteria (6), Acidobacteria (3), Chloroflexi (2) and Poribacteria (2) were the phyla with most sequences classified to. A significant proportion of the sequences were not classified to a given phyla (4). Furthermore, a single clone sequence (XT_48) was identified as eukaryotic, although no similar sequences could be identified at the NCBI nucleotide database. Of the sequences classified as root (2), one (clone XT_26) was similar to an actinobacterium clone sequence, while the other (clone XT_26) aligned partially to an eukaryotic alpha-tubulin gene.

Overall, corresponding microbial communities were inferred from cloning amplified fragments of the 16S rRNA gene with the primer pairs 27f-1492r and 533f-907r. The same major phylotypes were present in *S. carteri* and *X. testudinaria* using the two different primer pairs. The few disagreements observed, e.g. the absence of Nitrospirae sequences among the *S. carteri* sequenced clones or Deferribacteres among the *X. testudinaria* sequenced clones, could be explained by the low number of sequenced clones and the random nature of the clone selection. Furthermore, the amplification of Poribacteria 16S rRNA genes with the primer pairs 533f-907r was confirmed.

Table 14. The *S. carteri* and *X. testudinaria* 16S rRNA gene clones (533f-907r) analysis

		BLAST results					Taxonomic classification
Clone	**length (bp)**	**Query GI**	**e-value**	**Align length**	**Ident matches (%)**	**Description**	**Phylum[1]**
SC_11	338	407913010	5e-174	339	99.71	Unc bac clon A09SC 16S (PG)	Proteobac
SC_12	263	407913010	1e-128	263	98.86	Unc bac clon A09SC 16S (PG)	Proteobac
SC_13	312	407912993	1e-159	312	99.68	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_14	132	155968482	5e-50	126	96.03	Unc bac clon 175T3 16S (PG)	Proteobac
SC_15	356	407912993	3e-166	336	98.51	Unc bac clon A07SC 16S (PG)	Proteobac
SC_16	344	407912993	6e-148	288	100	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_17	418	407912993	0	391	97.44	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_18	260	407912993	8e-131	260	99.62	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_2	339	407913010	8e-167	342	98.25	Unc bac clon A09SC 16S (PG)	Proteobac
SC_20	483	407912993	0	411	99.03	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_3	526	407912993	0	387	100	Unc bac clon A07SC 16S (PG)	Gammaproteobac
SC_4	399	407912979	0	377	98.67	Unc bac clon C12SC 16S (PG)	Proteobac
SC_54	417	297171870	3e-166	417	92.33	Unc gamma proteobacterium HF0500_32L01	Opisthokonta (Eukaryota)
SC_6	368	407913010	0	368	99.46	Unc bac clon A09SC 16S (PG)	Proteobac
SC_8	565	407912993	0	411	99.76	Unc bac clon A07SC 16S (PG)	Gammaproteobac
XT_25	412	440658092	2e-164	391	93.86	Unc marine bacterium clon E12 16S (PG)	Chloroflexi
XT_26	489	482678009	0	376	99.73	Unc actinobacterium clon KX1-51 16S (PG)	Root
XT_27	490	451354037	0	387	99.74	Unc bac clon BA88-C6-seq 16S (PG)	Alphaproteobac
XT_28	528	451353977	0	401	99.5	Unc bac clon BA17-C1-seq 16S (PG)	Bac
XT_29	237	350627498	9e-115	237	98.73	Unc Gemmatimonadetes bac clon XE2B11 16S (PG)	Gemmatimonadetes
XT_30	397	226443709	0	391	99.23	Unc Acidobac bac clon XA2A12F 16S (PG)	Acidobac
XT_32	354	315020323	0	354	100	Unc gamma proteobacterium clon XD2G07 16S (PG)	Gammaproteobac
XT_33	371	511630215	0	371	100	Unc bac partial 16S rRNA gene, clon B33/GW947	Acidobac
XT_34	337	158822000	2e-173	337	99.7	Unc bac isolate DGGE gel band B8-1 16S (PG)	Gammaproteobac
XT_35	501	350627599	0	390	99.74	Unc Chloroflexi bac clon XD1E11 16S (PG)	Chloroflexi
XT_36	517	451353961	0	385	100	Unc bac clon BA102-C41-seq 16S (PG)	Deltaproteobac
XT_37	502	440658092	7e-169	390	94.62	Unc marine bac clon E12 16S (PG)	Bac
XT_39	506	451353977	0	390	100	Unc bac clon BA17-C1-seq 16S (PG)	Bac
XT_42	287	451353926	8e-146	287	99.65	Unc bac clon BA102-C6-seq 16S (PG)	Proteobac
XT_43	497	526299948	0	388	99.48	Unc Poribac bac clon CtgComparison_6 16S (PG)	Poribac
XT_44	474	281495439	0	385	99.48	Unc bac clon AncL36 16S rRNA (PG)	Bac
XT_45	286	451353961	6e-147	286	100	Unc bac clon BA102-C41-seq 16S (PG)	Deltaproteobac
XT_46	480	164508136	6e-40	94	100	Phytophthora cinnamomi tub1 gene for alpha-tubulin	Root
XT_48	507						Eukaryota
XT_49	339	400269018	1e-164	339	97.94	Unc bac clon CYMB_E2 16S (PG)	Acidobac
XT_50	493	269063599	0	387	99.48	Unc Poribac bac partial 16S, clon PC_57	Poribac

The *S. carteri* (SC_) and X. testudinaria (XT_) derived clones sequenced are reported. In order to fit to the page, words were abbreviated: **Ind**entical, **unc**ultured, **bac**teria/**bac**terium, **clon**e, **16S** rRNA gene. PG stands for partial gene.

[1]Taxonomic classification are reported at phylum level. When phylum level assignment was not supported, kingdom or root level is reported. Proteobacteria are reported at the class level, when the class level assignment was possible. Names finishing in bacteria were shortened to finish with “bac”.

4.6.2 Phylogenetic analysis of highly expressed key functional genes

To confirm the phylogenetic inferences made for the most expressed functional genes of *S. carteri* metatranscriptomes (Section 4.5), additional analyses were conducted based on the representative sequences obtained from stringent assembly. The analyses were carried out for the gene functions related to archaeal nitrogen metabolism (ammonia monooxygenase and copper-containing nitrite reductase), proteobacterial carbon transport (acetate permease ActP and TRAP-type C4-dicarboxylate transport system, periplasmic component) and C1 metabolism (formate dehydrogenase-O, major subunit and methanol dehydrogenase large subunit protein), and cyanobacterial photosynthesis (the alpha and beta chains of phycoerythrin). The representative sequences were retrieved from the consensus nucleotide sequence of the contigs with the greatest number of reads. The assembly performed with stringent parameters resulted in contigs with similar or shorter lengths in comparison with the standard assembly (Table 15). Furthermore, contigs with the most assembled reads were less representative in the stringent assembly in comparison with the standard.

Table 15. Comparison between representative sequences obtained from stringent and standard assembly

Gene	Representative sequences*	Stringent parameters		Standard parameters (Section 4.5)§	
		Length (bp)	Total reads covered (%)	Length (bp)	Total reads covered (%)
ammonia monooxygenase	AMO 1	470	44.9	628	88
	AMO 2	268	26.5		
	AMO 3	208	7.0		
copper-containing nitrite reductase	CuNiR 1	676	44.5	676	50
	CuNiR 2	772	11		
	CuNiR 3	627	3.0		
acetate permease ActP	ActP 1	606	4.0	1462	6
	ActP 2	767	3.3		
	ActP 3	825	2.7		
TRAP-type C4-dicarboxylate transport system, periplasmic component	TRAP-type C4 1	961	5.0	941	5
	TRAP-type C4 2	359	3.3		
	TRAP-type C4 3	340	2.9		
formate dehydrogenase-O, major subunit	FDH-O 1	562	6.8	1330	10
	FDH-O 2	1021	5.5		
	FDH-O 3	931	5.3		
methanol dehydrogenase large subunit protein	MDH 1	593	12.2	1717	72
	MDH 2	404	11.6		
	MDH 3	554	10.5		
phycoerythrin alpha chain	PEa 1	524	6.5	541	18
	PEa 2	375	5.5		
	PEa 3	477	4.4		
phycoerythrin beta chain	PEb 1	576	8.0	531	15
	PEb 2	533	6.6		
	PEb 3	467	5.0		

*Representative sequences are the consensus sequences from the three contigs that received the greatest number of reads.
§Representative sequences derived from the assembly using standard parameters are reproduced here for comparison.

The representative sequences of each selected gene function were used for phylogenetic analysis. First, the 20 most similar protein sequences to each of the three translated representative sequences were retrieved from NCBI by BLAST. After multiple alignment and automatic curation, phylogenetic trees were constructed (Figure 18). The archaeal gene

functions were assigned to the phylum Thaumarchaeota (Figure 18 A, B), representing 78% and 44% of the total reads assigned to ammonia monooxygenase and copper-containing nitrite reductase. Additionally, approximately 15% of the reads assigned to copper-containing nitrite reductase were phylogenetically related to Bacteria, instead of Archaea. With respect to the proteobacterial carbon transport genes (Figure 18 C, D), the representative sequences analysed were related to alphaproteobacterial marine bacteria, including members of the groups SAR116 and SAR11. These sequences represented a minimal proportion of the reads assigned to acetate permease ActP (summing 10% of the gene reads) and TRAP-type C4-dicarboxylate transport system, periplasmic component (11%). With regard to the representative sequences of the genes related to C1 metabolism (Figure 18 E, F), the sequences were largely related to gammaproteobacterial homologues, including all representative sequences of formate dehydrogenase-O, major subunit (17%). The two most representative sequences of methanol dehydrogenase large subunit protein (24%) were close related to gammaproteobacterial homologues, while the third most representative sequence (10%) was placed in a branch containing alphaproteobacterial and betaproteobacterial homologues. In comparison with the phylogenetic tree constructed from sequences originated from standard assembly (Section 4.5), this tree presented lower bootstrap values, likely because the assembly resulted in contigs that represent different proportions of the same gene, from which phylogeny was inferred. With respect to the cyanobacterial photosynthesis genes (Figure 18 G, H), the representative sequences of alpha and beta chains of phycoerythrin were all close related to proteins from species of the genus *Synechococcus*, representing 16% and 20% of the reads assigned to each gene.

In summary, these results confirm the phylogenetic inferences made for the most expressed functional genes of *S. carteri* metatranscriptomes (Section 4.5).

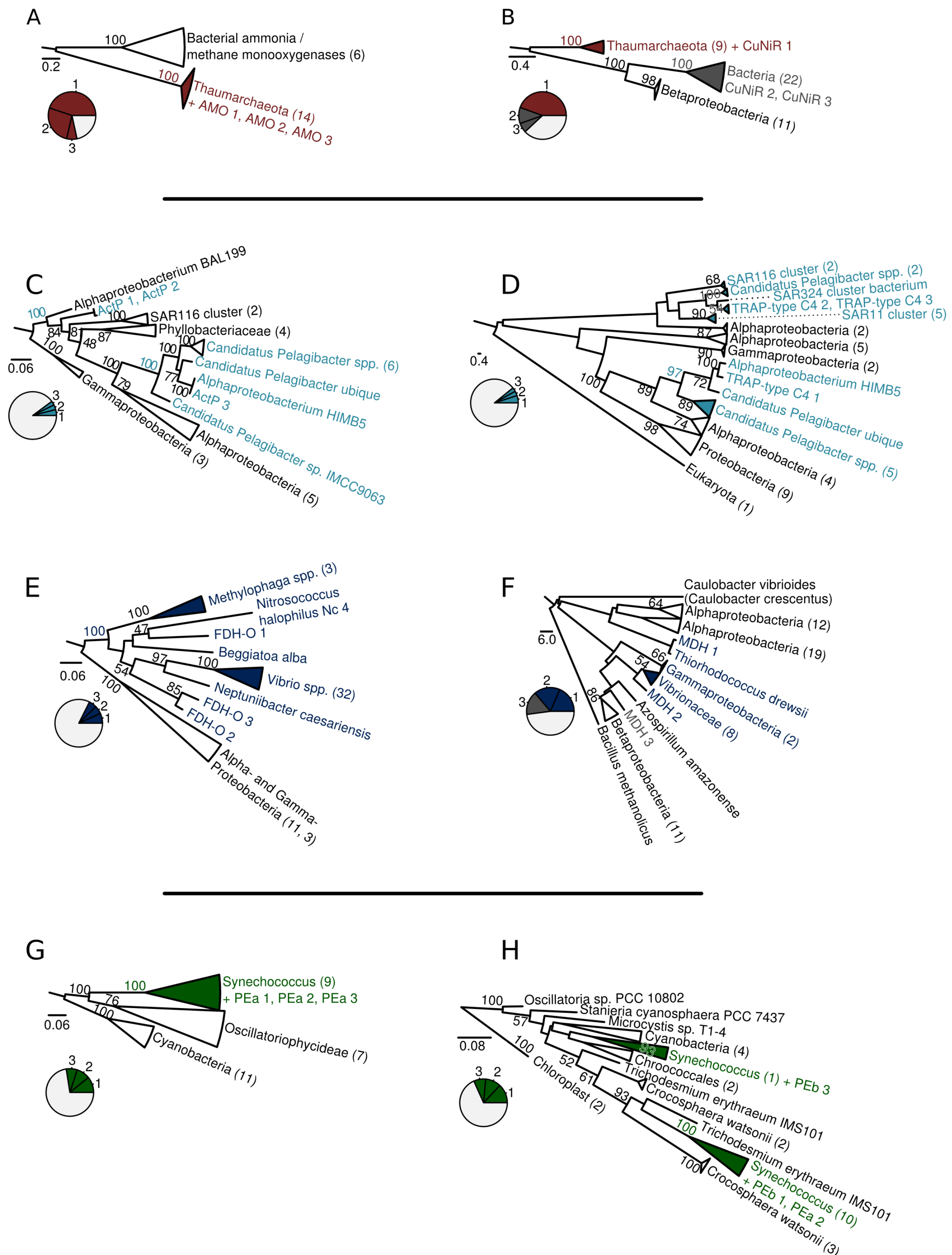


Figure 18. Phylogeny of the highly expressed key gene functions of the most abundant taxa in *S. carteri*. Maximum-likelihood trees were constructed from representative sequences originated from assembly using high stringent parameters and homologues of genes related to nitrogen metabolism in Archaea: (A) ammonia monooxygenase (AMO) and (B) copper-containing nitrite reductase (CuNiR);

to carbon transport in Proteobacteria: (C) acetate permease ActP (ActP) and (D) TRAP-type C4-dicarboxylate transport system, periplasmic component (TRAP-type C4); to C1 metabolism in Proteobacteria: (E) formate dehydrogenase-O, major subunit (FDH-O) and (F) methanol dehydrogenase large subunit protein (MDH); and to photosynthesis in Cyanobacteria, the (G) alpha and (H) beta chains of phycoerythrin (PEa and Peb). Branches were collapsed for a simplified view and named according to the lowest common taxon. Bootstrap values were calculated based on 1000 replications. Bootstrap values lower than 50 are not show. Branches of TRAP-type C4 (D) and MDH (F) trees were transformed proportionally for a better visualization. The scale bars represent the average number of amino acid substitution per site. Pie charts represent the proportion of the representative sequences analysed in relation to the total number of reads assigned to a given gene function.

5 Discussion

5.1 Microbial communities of sponges

5.1.1 High-throughput sequencing

Since the emergence of microbial ecology as a scientific field, which dates back to Antonie van Leeuwenhoek (1632-723) observations of “animalcules” scraped from his teeth, the question of how many different kinds of microorganisms are associated with animal hosts is still to be answered (Robinson *et al.*, 2010). Because the great majority of microorganisms can not cultured using current techniques (Amann *et al.*, 1995), a phenomenon named “the great plate anomaly” (Staley and Konopka, 1985), culture-independent methods using the 16S rRNA gene as a phylogenetic marker have become the common approach to characterize microbial community composition (Tringe and Hugenholtz, 2008). Traditionally, these methods include (i) the construction and Sanger sequencing of nearly full 16S rRNA gene libraries, (ii) denaturing gradient gel electrophoresis (DGGE) profiling, band extraction, and sequencing, (iii) terminal restriction fragment length polymorphism (T-RFLP), and (iv) flourescence *in situ* hybridisation (FISH). Due to the recent advance and access to high-throughput sequencing technologies (e.g. Illumina and 454 pyrosequencing), microbial community analysis now comprises the simultaneous investigation of millions of 16S rRNA gene sequences originating from thousands of samples (Tringe and Hugenholtz, 2008; Bartram *et al.*, 2011). In this way, the use of these technologies represents a revolution in microbial ecology that is expanding our comprehension of microbial diversity (Sogin *et al.*, 2006; Kuczynski *et al.*, 2012). This Thesis is among several recent efforts to analyse sponge microbiomes with high-throughput sequencing technology (Webster *et al.*, 2010; Lee *et al.*, 2011; Jackson *et al.*, 2012; Schmitt *et al.*, 2012a; White *et al.*, 2012; Schmitt *et al.*, 2012b; Simister *et al.*, 2012b; Simister *et al.*, 2012c; Arellano *et al.*, 2013).

Deep sequencing profiling of microbial communities is a fast evolving approach, with

new technologies and methods, including Illumina, IonTorrent, and PacBio, emerging and improving at daily basis (Shokralla *et al.*, 2012; Kozich *et al.*, 2013). Each of these technologies presents their own strengths and drawbacks, and the decision on which technology to choose reflects the balance of multiple factors, including read length, depth, sequence accuracy, usability, and cost (Kuczynski *et al.*, 2012). For example, 454 FLX Titanium can currently yield 10^6 reads per run, with maximum length of 400 bp; while Illumina HiSeq 2000 can yield 10^9 reads per run, with 151 bp length, and a cost relative to 454 FLX Titanium of 0.002 per mega base pairs (Mb) (Kuczynski *et al.*, 2012). To date, 454 pyrosequencing (e.g. Roche 454 GS FLX+ and 454 FLX Titanium) has been the most used high-throughput sequencing method for characterizing the diversity of microbial communities. Because of the availability of protocols for data analysis and better length/depth balance in comparison with other technologies, the diversity of microbial communities of seawater and the sponges *S. carteri* and *X. testudinaria* were investigated using the Roche 454 GS-FLX Titanium platform.

Due to the current limitation of high-throughput sequencing read length obtained by high-throughput sequencing, it is only possible to recover a fragment of the 16S rRNA gene. An average 16S rRNA gene has 1,500 bp length, which comprises sequence regions with different variability rates, named conserved or hypervariable regions (Amann *et al.*, 1995; Van de Peer *et al.*, 1996). To obtain the maximum taxonomic information, hypervariable regions of the 16S rRNA gene are chosen for amplification. These regions are considered as good proxies for the full-length 16S rRNA gene (Liu *et al.*, 2007; Liu *et al.*, 2008), despite the variation on diversity estimations between 16S rRNA gene regions (Wang and Qian, 2009; Engelbrektson *et al.*, 2010; Cai *et al.*, 2013). However, obtaining 16S rRNA gene fragments instead of full-length sequences means having less information and, therefore, less resolution for taxonomic and phylogenetic inferences.

To reduce the amount of sequences to be handled and the impact of sequencing errors in the final analysis, 16S rRNA sequences derived from high-throughput sequencing technologies are normally grouped into operational taxonomic units (OTU). The grouping into OTUs is done based on sequence similarity, without designating a formal taxonomic rank and without requiring phylogenetic analysis (Worden and Not, 2008). In practice, sequence similarity thresholds are used in reference to taxonomic ranks. A cut-off of 97% sequence

similarity is commonly used to represent the species level, while other similarity values are used for genus (95% sequence similarity), family (90%), and order (85%) (Cressman *et al.*, 2010; Webster *et al.*, 2010; Schmitt *et al.*, 2012b). The 97% sequence similarity threshold is based on DNA-DNA hybridization studies and their correspondence to the 16S rRNA gene sequence variability (Stackebrandt and Goebel, 1994). It is important to note that two ribo-types with more than 97% sequence similarity not necessarily represent the same species (Fuhrman and Hagström, 2008). Furthermore, the concept of species in prokaryotes is still subject to debate, which involves genetic and ecological arguments (Doolittle and Zhaxybayeva, 2009; Fraser *et al.*, 2009), and is out of the scope of this Thesis. Pragmatically, the threshold of 97% 16S rRNA gene sequence similarity is widely applied in the recent microbial diversity studies, including those exploring microbiomes associated with sponges (Webster *et al.*, 2010; Kuczynski *et al.*, 2012; Werner *et al.*, 2012; Schmitt *et al.*, 2012b; Simister *et al.*, 2012b; Campbell and Kirchman, 2013; Haegeman *et al.*, 2013). Therefore, the threshold of 97% 16S rRNA gene sequence similarity was used to define OTUs in this Thesis.

The term “rare biosphere” has being recently used to refer to ribo-types recovered in great diversity but in low abundances in studies based on deep sequencing (Sogin *et al.*, 2006; Bowen *et al.*, 2012). Currently, the rare biosphere represents an exciting new frontier in microbial ecology. However, there are some technical issues associated with the quantification and estimation of members of the rare biosphere, including the generation of sequencing errors and chimeric sequences (Engelbrektson *et al.*, 2010; Huse *et al.*, 2010; Haas *et al.*, 2011; Schloss *et al.*, 2011), and the under quantification of low abundant members of microbial communities (Gonzalez *et al.*, 2012). Nevertheless, the rare biosphere was taken in consideration in several microbial diversity studies (Elshahed *et al.*, 2008; Sjöstedt *et al.*, 2012; Hugoni *et al.*, 2013), including sponge-associated microbial communities (Schmitt *et al.*, 2012b; Reveillaud *et al.*, 2014). Because the biological implications of the rare biosphere are still little understood, I aimed to study the abundant members of sponge (and seawater) microbial communities, which are expected to be the major contributors to their respective ecosystems (Pedrós-Alió, 2012).

5.1.2 The HMA/LMA dichotomy

The observation that some marine sponges harbour an abundant microbial population while others are associated with a lesser amount of microbes dates back almost 40 years ago (Reiswig, 1974; Vacelet and Donadey, 1977; Wilkinson, 1978; Reiswig, 1981). In reference to these two types of sponge-microbes association, the terms “high microbial abundance” (HMA) and “low microbial abundance” (LMA) sponges were coined (Hentschel *et al.*, 2003). In this Thesis, the sponges *X. testudinaria* and *Amphimedon ochracea* were classified as being HMA or LMA as part of the most comprehensive survey for HMA and LMA microbial patterns to date (Gloeckner *et al.*, 2014) (Section 4.1). Additionally, I classified the Red Sea sponges *Crella cyathophora* and *S. carteri* as LMA in a study that profiled bacterial communities in LMA sponges (Giles *et al.*, 2013a) (Section 4.2). This study revealed that LMA sponges from different geographic locations are associated with a low phylum-level microbial diversity, a pattern that is clearly different from bacterial associations with HMA sponges. Interestingly, even small 16S rRNA gene clone libraries were sufficient to reveal the different patterns between HMA/LMA microbiomes, by detecting the major phylogenetic lineages of *X. testudinaria* (HMA) and *S. carteri* (LMA) (Section 4.6.1). In conclusion, we recommend a combination of transmission electron microscopy and 16S rRNA gene sequence data in order to experimentally determine the HMA or LMA status of a given sponge. The combination of the molecular and microscopical methods is an important strategy to discern sponge species with an intermediate microbial load, even though the classification of most of the sponges analysed (55 out of 56) was possible based solely on transmission electron microscopy observations.

The dichotomy between HMA/LMA sponges was further supported by the detailed investigation of the diversity, specificity, and transcriptional activities of the microbiomes of *X. testudinaria* and *S. carteri* (Moitinho-Silva *et al.*, 2014a) (Section 4.3). The hypothesis that the LMA microbiome is dominated by the first bacterial population to colonize the growing animal (Giles *et al.*, 2013a) was shown to be unlikely after further investigations (Moitinho-Silva *et al.*, 2014a), because of the high similarities between the microbiomes of different *S. carteri* individuals collected at the same time or collected in two occasions separated by one year. Similar observations for other LMA sponges were reported in other publications (Weisz

et al., 2007; Erwin *et al.*, 2011; Lee *et al.*, 2011; Björk *et al.*, 2013). Altogether, this Thesis contributes significantly to the advance of the knowledge about the HMA/LMA dichotomy (Table 16), and, consequently, the sponge-microbe symbiosis. However, the evolutionary adaptations that led to the HMA/LMA dichotomy remain unclear as well as the implications of these two types of association to the fitness of the involved organisms.

Table 16. Summary of the current knowledge about microbial communities associated with HMA and LMA sponges

Microbiome characteristics	HMA	LMA	Methods*
Abundance	High	Low	**TEM[1]** DAPI-counting[2] Confocal microscopy[3] Real-time PCR (16S/18S rRNA genes ratio)[4]
Richness, assessed by number of OTUs or estimators	Not a very good indicator, because LMA species richness is likely significantly impacted by seawater microbes (Moitinho-Silva *et al.*, 2014a). Can not be directly compared across studies.		**Deep sequencing[5]** Clone libraries[6]
Diversity, assessed by Shannon and Simpson indexes	High	Low (generally)	**Deep sequencing[7]** Clone libraries[8]
Phylum diversity	High phylum-level diversity. Indicator phyla are Chloroflexi and Poribacteria.	Low phylum-level diversity.	**Deep sequencing[5]** **Clone libraries[9]**
Relatedness among other sponge microbiomes	High, generally rich in sponge-specific clusters.	Low, generally poor in sponge-specific clusters.	**Deep sequencing[10]** Clone libraries[11]
Functions	Distinct patterns, few particular functions identified.		**Environmental microarray[12]**
Expressed genes	Poorly characterized. Archaeal ammonia oxidation is a common feature.		**Metatranscriptomics[13]** Metaproteomics[14]

*Methods of selected studies specifically investigating this characteristic. Methods applied in this Thesis are in bold.
[1](Vacelet and Donadey, 1977; Weisz *et al.*, 2007; Giles *et al.*, 2013a; Gloeckner *et al.*, 2014)
[2](Schläppy *et al.*, 2010; Gloeckner *et al.*, 2014)
[3](Ribes *et al.*, 2012)
[4](Bayer *et al.*, 2014)
[5](Lee *et al.*, 2011; Schmitt *et al.*, 2012a; Schmitt *et al.*, 2012b; Blanquer *et al.*, 2013; Cárdenas *et al.*, 2014; Moitinho-Silva *et al.*, 2014a)

[6](Sipkema *et al.*, 2009; Erwin *et al.*, 2011; Montalvo and Hill, 2011; Schmitt *et al.*, 2011)
[7](Blanquer *et al.*, 2013; Moitinho-Silva *et al.*, 2014a)
[8](Erwin *et al.*, 2011; Montalvo and Hill, 2011; Webster *et al.*, 2013)
[9](Section 4.6.1), (Weisz *et al.*, 2007; Kamke *et al.*, 2010; Björk *et al.*, 2013; Giles *et al.*, 2013a; Webster *et al.*, 2013)
[10](Schmitt *et al.*, 2012b; Moitinho-Silva *et al.*, 2014a)
[11](Kamke *et al.*, 2010; Giles *et al.*, 2013a)
[12](Section 4.4)
[13](Radax *et al.*, 2012b; Moitinho-Silva *et al.*, 2014b)
[14](Liu *et al.*, 2012)

5.1.3 Diversity

The diversity of the microbial communities associated with *S. carteri* and *X. testudinaria* was investigated. In my analysis, I consider the fact that the sponges can filter thousands of litres of seawater per day, and, therefore, I also analysed the microorganisms present in the surrounding seawater. The high species richness of *S. carteri* microbial communities in comparison with *X. testudinaria* communities was likely a result of the presence of bacterioplankton, as revealed by the analyses of OTU taxonomic affiliation and abundance patterns using the MCE algorithm. However, diversity indices that accounted for the relative abundance of OTUs revealed that the microbial communities of seawater and *S. carteri* were less diverse in comparison with *X. testudinaria* communities. This Thesis is in agreement with other studies that reported low diverse microbiomes associated with LMA sponges and high diverse microbiomes associated with HMA sponges (Vacelet and Donadey, 1977; Kamke *et al.*, 2010; Schmitt *et al.*, 2012b; Giles *et al.*, 2013a). While the diversity patterns observed in sponges are being approached with different methods, the mechanisms that led to the emergence and maintenance of them are still open questions (Taylor *et al.*, 2007; Hentschel *et al.*, 2012). These questions are potentially answered by experimental approaches that test the influence on the microbial diversity of biological factors, i.e. the host (Wehrl *et al.*, 2007) and the symbionts (Thomas *et al.*, 2010; Nguyen *et al.*, 2014), and structural factors, e.g. the sponge internal structures (Vacelet and Donadey, 1977; Boury-Esnault *et al.*, 1990; Weisz *et al.*, 2008) and the sponge tissue micro-environment (Fiore *et al.*, 2010).

5.1.4 Specificity

The specificity of the symbiosis between marine hosts and microorganisms has been characterized in several models, including the association of microbes with squids (Nyholm and McFall-Ngai, 2003; Nyholm and McFall-Ngai, 2004), corals (Baker, 2003; Silverstein *et al.*, 2012), hydra (Fraune and Bosch, 2007; Bosch, 2012), and sponges (reviewed recently by Hentschel *et al.*, 2012). In sponge microbiology, this theme is interpreted in two different ways: (i) the specificity in relation to the host (sponge-specificity), i.e. when a given microbial lineage is specific to one or more sponge species in comparison with non-sponge environments, such as seawater or other animal; (ii) the specificity in relation to other sponge species (species-specificity), i.e. when a given microbial lineage is associated with a particular sponge species but not with another. The recovery of phylogenetically related microbial lineages from different sponge species sampled worldwide, the so called sponge-specific clusters (SC) (Hentschel *et al.*, 2002; Taylor *et al.*, 2007; Simister *et al.*, 2012a), supports the concept of sponge-specific microbiomes. On the other hand, several studies investigating sponge microbiota reported species-specific microbial lineages (Thacker and Starnes, 2003; Erwin *et al.*, 2011; Lee *et al.*, 2011; Schmitt *et al.*, 2012a; Schmitt *et al.*, 2012b; Giles *et al.*, 2013a; Webster *et al.*, 2013). Although some of these studies have applied high-throughput sequencing for the investigation of sponge microbiomes, the concept of specificity was interpreted from clone library data. Here, the specificity of the symbiosis between sponges and their associated microorganisms was reviewed in the context of high-throughput sequencing technologies for the first time.

The concept of specificity was investigated in this Thesis with respect to the biological factors, sponge- and species-affiliation, and technical factors, sequencing depth, sequence similarity, and OTU abundance. At the maximum sequence depth analysed (32,148 sequences in each dataset), most of the OTUs of the microbiomes studied (i.e. the microbiomes of seawater, *S. carteri*, and *X. testudinaria*) were shared, where more than 90% of the sequences of each microbiome were shared between all three sources and more than 99% between two given sources. In other words, specific OTUs were rare. Moreover, the proportion of unique sequences in each source was dependent of the sequencing depth, where abundant OTUs in a given source started to be found at other sources with the increase of sequence depth. By

performing the MCE analysis, an unsupervised investigation of the OTUs abundance and distribution, it was possible to determine which OTUs were specifically associated with a given source. This analysis was supported by another approach based on the similarity of OTU representative sequences to sequences from the same species but collected in different time and/or location (Montalvo and Hill, 2011; Giles *et al.*, 2013a). In conclusion, it was demonstrated that the patterns of OTU abundance, instead of the "presence or absence" criterion, are instrumental to reveal the specificity of sponge associated consortia in deep sequencing data.

The idea of Lourens Baas Becking of "*everything is everywhere*, but, *the environment selects*" (Baas-Becking, 1934), recently reviewed by De Wit and Bouvier (2006), has greatly influenced modern microbiology (Whitfield, 2005; O'Malley, 2007). While it is accepted that the environment is one of the factors that shape microbial diversity patterns (Martiny *et al.*, 2006; Andersson *et al.*, 2014), the concept that every microbe can be found everywhere was addressed by supporting and contrary publications (reviewed by Zarraonaindia *et al.*, 2013). Deep sequencing approaches have contributed the hypothesis that marine microbes are widespread in a global ocean microbial seed bank and thus the community composition primarily reflects the shifts in the relative abundance of taxa rather than their presence or absence (Gibbons *et al.*, 2013). Interestingly, this seed bank would include those microorganisms specifically associated with sponges (Webster *et al.*, 2010; Taylor *et al.*, 2013), which strengthens the hypothesis of horizontal acquisition of sponge symbionts. High-throughput technologies produce unprecedented sequencing depth that changed our perception of microbiomes (Figure 19). However, the current sequencing technologies have limited taxonomic resolution because of the short sequence length produced (discussed in Section 5.1.1) and the OTU similarity cut-offs applied in its analysis (Hanson *et al.*, 2012). Therefore, previous taxonomic inferences based on curated long-length 16S rRNA gene sequences, such as the concept of SC/SCC (Hentschel *et al.*, 2002), should not be evaluated or validated by deep sequencing results, but taken into account in pursuit of a better understanding of sponge-microbe symbiosis. To account for the occurrence of sponge-specific microbes in the environment, albeit at much lower numbers, we propose to change the term "specific" to "enriched" when investigating aspects of host specificity based on high-throughput sequencing data (Moitinho-Silva *et al.*, 2014a). In this way, my results indicate that both *S. carteri* and *X. testudinaria* harbour microbial communities enriched of

lineages that are particularly abundant in each species (species-enriched) and in other sponges (sponge-enriched or host-enriched).

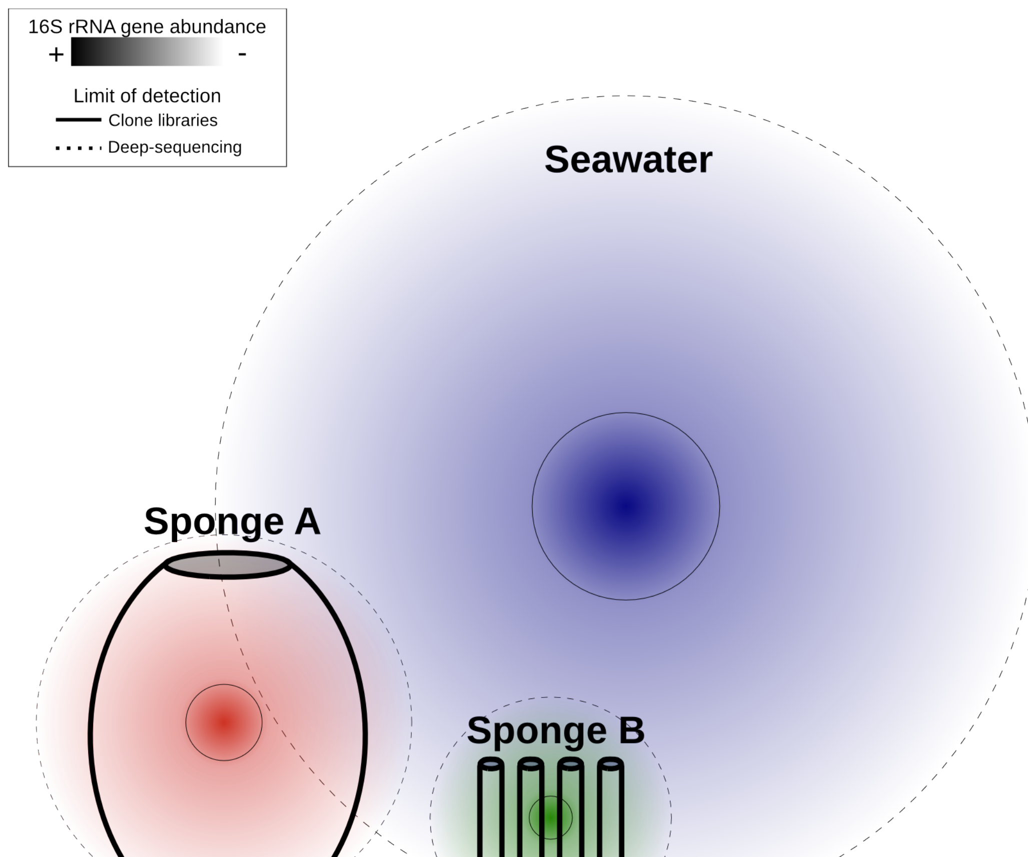


Figure 19. Perception of microbial communities according to sequencing methods. In this scheme, the microbial communities of two hypothetical sponges (A and B) and seawater are represented by different colours. The abundance of 16S rRNA gene sequences as recovered by clone libraries or by deep sequencing is shown by colour intensity. Clone libraries are able to detect a portion of each microbiome, generally the most abundant sequences (continuous line). With increase of sequence depth, enriched 16S rRNA genes of a given microbiome are recovered in other microbiomes (dashed line).

5.1.5 Composition and transcriptional activity

The characterization of the microbial community composition is a logical step towards the understanding of the symbiosis between sponges and microbes. Here, community composition is defined as the identity and the relative abundance of taxa in a given sample

(Hanson *et al.*, 2012). Previous publications have shown that many sponge microbiomes are dominated by members of the archaeal phylum Thaumarchaeota and the bacterial phyla Proteobacteria, Chloroflexi, Actinobacteria, Acidobacteria, Nitrospirae, and candidate phylum Poribacteria (as reviewed by Hentschel *et al.*, 2012). This pattern is characteristic of HMA sponges. As an example of HMA sponge associated microbial communities, the microbiome of *X. testudinaria* was dominated by several phyla, including Proteobacteria, Chloroflexi, Deferribacteres, and Poribacteria. Most of the abundant taxa in *X. testudinaria* were related to SC/SCC. On the other hand, the LMA sponges are dominated by members of few phylum, commonly Proteobacteria or Cyanobacteria, with small overlap between the microbiomes of different species (Giles *et al.*, 2013a). For instance, the microbiome of *S. carteri* was dominated by Gammaproteobacteria, and to a lesser extent Cyanobacteria and Nitrospirae. Among the dominant taxa of *S. carteri* microbiome, only a gammaproteobacterial lineage was related to SC/SCC, representing about 13% of the total sequences recovered for this sponge. The occasional placement of 16S rRNA gene sequences derived from LMA sponges into SC formed by HMA sequences, and vice versa, was already observed (Giles *et al.*, 2013a). My analyses expand this finding by showing that sponge-enriched microbes can be among the dominant members of LMA sponge microbial communities.

Bacteria have the ability to reduce their cellular activity levels in response to environmental stress, starvation, or as a result of reduced growth rates (Kirchman, 2008). Despite the broad investigation of bacterioplankton activity (e.g. Campbell and Kirchman, 2013), only two publications have investigated the activity of sponge symbionts in a community-level approach (Kamke *et al.*, 2010; Simister *et al.*, 2012b). In this Thesis, the transcriptional activity of *S. carteri* and *X. testudinaria* symbionts was assessed based on the differences between taxon abundances in DNA and cDNA data sets. Despite some caveats, which include the differences in 16S rRNA gene copy numbers between species, extraction bias, and the different transcriptional activities between species at the same growth rate, the transcriptional activity can be interpreted as a proxy for cellular activity and potential growth rate (Jones and Lennon, 2010; Campbell *et al.*, 2011; Gaidos *et al.*, 2011; Lanzen *et al.*, 2011). Here, the differences between taxon abundances in DNA and cDNA data sets was interpreted in this manner.

In general, the microbial communities from seawater and from sponges were

transcriptionally active. Consistent with the environmental conditions of the collection, which was conducted on a sunny morning, members of the cyanobacterial genus *Synechococcus* were highly transcriptionally active in both seawater and *S. carteri*. Due to the lack of phylogenetic resolution caused by the sequencing of a short fragment of the 16S rRNA genes, it was not possible to assign *Synechococcus* sequences to SC/SCC. Therefore, it remains unclear whether the active *Synechococcus* lineages in *S. carteri* were free-living or sponge-associated. Both scenarios are possible, since cyanobacterial cells were described to be highly active in both seawater and *S. carteri,* as determined by the analysis of metatranscriptomes from the North Pacific ocean surface (Poretsky *et al.*, 2009) and other *S. carteri* individuals (Section 4.5).

Many of the sponge-enriched clades were found to be transcriptionally active. With regard to *S. carteri*-enriched bacteria, the proteobacterial clade E01-9C-26 and the genus *Nitrospira* were more abundant in cDNA than in DNA libraries, although not statistically significant, likely indicating that they were active. On the other hand, gammaproteobacteria lineages were transcriptionally active on a low level. With respect to *X. testudinaria*, a similar pattern was observed in that some of the most abundant taxa were highly transcriptionally active (e.g. the Chloroflexi SAR202 clade, Poribacteria, the proteobacterial genus *Candidatus* Entotheonella, the proteobacterial genus *Nitrosococcus*, and the Spirochaetes genus *Spirochaeta*), while others were active at a low level (e.g. the Chloroflexi *Caldilinea* genus, the Actinobacteria Sva0996 marine group, and the Gemmatimonadetes BD2-11 terrestrial group). In agreement with my results, other publications have also reported high transcriptional activity for Poribacteria and Spirochaetes (Kamke *et al.*, 2010; Simister *et al.*, 2012b).

The ecological implication of different activity levels observed for sponge associated microorganisms is unclear. Very likely, cellular activity patterns reflect the physiological responses of each ecotype to the particular conditions encountered. In this case, high cellular activity reflects optimal growth conditions, which will lead to higher fitness in relation to competing ecotypes. On the contrary, low cellular activity may indicate a physiological adaptation to unfavourable conditions, such as those putatively encountered by seawater while entering in sponge tissue or when nutrients are scarce. In another sense, the different cellular activities observed between the sponge symbionts may be due to each

microorganism's steady metabolic rate, rather than changes in response to environmental conditions. In this manner, at the time of collection, each symbiont was found in its optimal growth condition. Further experimental approaches, such as monitoring changes in the symbiont transcriptional activity in response to different environmental conditions (Simister *et al.*, 2012b) or nutrient availability (Shao *et al.*, 2014), have the potential to explain the ecological meaning of the different activity levels observed for sponge symbionts.

5.2 Functions of sponge microbiomes

5.2.1 Meta-omics

Meta-omics, where the prefix “meta” refers to microbial communities or environmental samples, have greatly advanced microbial ecology by allowing access to uncultured microbes and approaching microbial functions at the community level (National Research Council (U.S.), 2007; Fritz *et al.*, 2013). The term (or suffix) “omics” defines both the techniques, including methods and analyses, and the research field that comprehend the study of the complete set of a given biological molecule in a particular system, e.g. cell, tissue, or organism (National Research Council (U.S.), 2007). For example, metagenomics refers to the set of all genes from a particular microbial community. Meta-omics allows microbial ecologists to go beyond the basic question “who is there?” by answering the question “what are they doing?”. Ultimately, one of the major aims of meta-omics is to link ecosystem processes to specific microbial populations (Morales and Holben, 2011).

Depending on the target molecule, different meta-omics reveal different functional aspects of the microbial community. Metagenomics assesses the functional potential of the microbial community by characterizing its gene set (DNA level). Metatranscriptomics and metaproteomics assess the active gene functions by characterizing the expression at the RNA and protein levels. Due to the facilities of DNA manipulation and the possibility of recovering the genomic context, meta-omic studies are primarily carried at the genomic level. Consequently, 6,604 publications related to metagenomics are listed in the Pubmed, a

literature database of the NCBI, at the time of writing this Thesis (May 2014), while metaproteomics and metatranscriptomics together sum 476 publications. The analysis of community gene expression have the unique power of revealing functional activities required for a specific condition or habitat. In this way, the study of the gene expression of microbial communities have contributed significantly for the increasing of knowledge about the functional dynamics of various habitats, including seawater (Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2009; Teeling *et al.*, 2012), soil (Leininger *et al.*, 2006; Urich *et al.*, 2008), sponges (Liu *et al.*, 2012; Radax *et al.*, 2012b), and the human gut (Verberkmoes *et al.*, 2009; Franzosa *et al.*, 2014). In this Thesis, insights into the functional basis of the sponge-microbe association were obtained by analysing both the genome and metatranscriptome of marine sponges, using DNA microarrays and high-throughput sequencing.

5.2.2 Environmental microarrays

DNA microarrays have been used to analyse the functional genomic repertoire of complex microbial communities from diverse habitats, including soils, marine sediments, and deep-sea hydrothermal vents (reviewed by He *et al.*, 2012). However, high-throughput sequencing approaches are substituting DNA microarrays in many applications, including metagenomic surveys, due to the technical advances of the former (Ledford, 2008). In comparison with “environmental microarrays” (a term coined by Kjelleberg, 2002), metagenomics based on deep sequencing presents superior sample coverage and is better suited for in-depth studies of unknown microbial community diversity (Roh *et al.*, 2010). On the other hand, environmental microarrays are advantageous because of the superior throughput, smaller costs, and simplicity of the data analysis, as the analysis of signal intensities is less complex than the annotation and sorting of massive sequence reads (Roh *et al.*, 2010).

Here, we used an environmental microarray, the GeoChip 4 (Tu *et al.*, 2014), to assess the functional genomic repertoire of sponge symbionts for the first time. The GeoChip 4 microarray was developed in order to cover biogeochemical processes and microbial responses to environmental perturbations and includes genes involved in several functional

categories such as carbon, nitrogen, sulphur, and phosphorous cycles (Tu *et al.*, 2014). Based on the exploration of the sample patterns by Minimum curvilinear embedding (MCE) (Cannistraci *et al.*, 2010), the samples from the Mediterranean and the Red Sea grouped into three distinct sets, i.e. HMA (*X. testudinaria*, *A. aerophoba*), LMA (*S. carteri*, *D. avara*) and seawater (Section 4.4). However, the analysis carried out to identify the genes responsible for this pattern revealed little differences in the microbial functional gene repertoires of sponges (HMA, LMA) and seawater. This result may reflect the true biological capabilities of these microbial communities, in that the gene intensities analysed in fact correspond to the present genes and their abundances. On the contrary, this result may reflect the limited applicability of this method to these microbial assemblages, in that the genes of sponges and seawater are not well represented by the microarray probes, leading to incomplete or unspecific hybridisations. In conclusion, environmental microarrays that were not projected based on sponge microbiomes, such as the GeoChip, may be more adequate to analyse the functional repertoire of high numbers of sponge microbiomes, where a general approach is favourable in comparison with a detailed investigation of particular gene functions. However, the detailed functional profile of sponge-associated microbial consortia may be better tackled using deep sequencing strategies, due to the possibility of detecting functions that are not represented on the microarray or genes that are not compatible with the microarray probes.

5.2.3 Metatranscriptomics

The work with RNA has the advantage over protein-based approaches of using high-throughput sequencing technologies and because it is less technically challenging with respect to the processes of molecule extraction, separation, and function identification (Moran, 2009). There are two strategies to study metatranscriptomic profiles: (i) the “double-RNA” approach, which aims the sequencing and analysis of both rRNA and mRNA for simultaneous assessment of the transcriptionally active members of the community and their expressed genes (Urich *et al.*, 2008); and (ii) the mRNA-guided approach, which aims the sequencing and analysis of the community mRNA for the characterization of the expressed functional genes. The mRNA-guided approach enriches mRNA from total RNA by physically removing the rRNA fraction prior sequencing and the annotated rRNA reads from sequencing

data (Poretsky *et al.*, 2009; Stewart *et al.*, 2010). This approach has the advantage of concentrating the sequencing and analysis efforts in the mRNA, which represents a small proportion of the total RNA. Here, a strategy was developed to enrich prokaryotic mRNA from total *S. carteri* RNA, which includes the host mRNA and the rRNA from the host and the associated microbial community. For rRNA removal, a protocol based on sample-specific probes (Stewart *et al.*, 2010) was applied, which is suitable for diverse environmental samples. Recently, a similar prokaryotic mRNA enrichment procedure was published (Hampton-Marcell *et al.*, 2013), which uses the Ribo-Zero Magnetic Gold kit (Epidemiology) (Epicentre, USA) for rRNA removal, and, consequently, is limited to a small range of host and bacterial taxa. The method established in this Thesis was successfully carried out to obtain the metatranscriptome of another sponge, *X. testudinaria,* and should be applicable for the isolation of any host-associated microbial metatranscriptome.

Because metatranscriptomics and metagenomics are based on cDNA (DNA synthesized from a mRNA template) and DNA sequencing, they use similar tools and pipelines for the data processing. Metagenomic sequences are frequently assembled to obtain full-length CDS or even genomes of uncultured organisms (Thomas *et al.*, 2012). The assembly procedure is advantageous because assembled reads are longer than single reads, which results in a better functional and taxonomic annotation (Huson *et al.*, 2007). However, the formation of chimeric sequences from the metagenome assembly is of major concern due to its impacts on the annotation (Pignatelli and Moya, 2011). Chimera formation depends on several factors, including genomic coverage, community complexity, read length, and assembly parameters (Pignatelli and Moya, 2011; Fan *et al.*, 2012a). The assembly of metatranscriptomes is not standard procedure. Several metatranscriptomic studies were conducted without the assembly step (Gilbert *et al.*, 2008; Poretsky *et al.*, 2010; Feike *et al.*, 2011; Gifford *et al.*, 2011; Gosalbes *et al.*, 2011), while others have assembled only 16S rRNA transcripts (Urich *et al.*, 2008; Radax *et al.*, 2012a) or the whole dataset (Jang *et al.*, 2012; Mason *et al.*, 2012; Baker *et al.*, 2013). My metatranscriptome analyses were carried out based on unassembled metatranscriptomic sequences, which resulted in the recovery of *S. carteri* metatranscriptomes free from assembly-induced bias, such as chimera formation. However, this strategy may have resulted in a lower rate of annotated reads than if assembly had been conducted (Pignatelli and Moya, 2011). Additionally, sequence assembly was only performed for the highly expressed gene functions, resulting in the recovery of nearly full-length

transcripts. The possibility of chimeric sequences occurrence in this procedure was small because *S. carteri* microbiomes are composed of few highly abundant symbionts which are functionally distinct. Furthermore, it is unlikely that chimeras impacted the phylogenetic inferences made in this Thesis as shown by additional analysis using highly stringent assembly parameters (Section 4.5, 4.6.2). The assembly of the highly expressed genes of *S. carteri* improved their taxonomic classification, allowing for the association of these gene functions to specific symbionts of *S. carteri*.

The metatranscriptome annotation is generally performed based on sequence homology to (meta)genomes (Lesniewski *et al.*, 2012; Gifford *et al.*, 2013; Embree *et al.*, 2014; Sheik *et al.*, 2014) or to sequences from public databases, such as the NCBI non-redundant database (Gifford *et al.*, 2011), the PFAM database (Gilbert *et al.*, 2008), or the MG-RAST server (Feike *et al.*, 2011). Here, *S. carteri* metatranscriptomes were annotated using the MG-RAST server. While this strategy allowed the identification of the transcripts that are similar to known genes, it did not address the expression of the non-conserved genes, including hypothetical CDS. Furthermore, this approach could not reveal the genomic context of the expressed genes. To overcome these issues, parallel sequencing and analysis of the corresponding metagenome can be carried out, what would in addition potentially improve the taxonomic and functional metatranscriptome annotation. However, this parallel approach requires a larger financial budget and increased time of analysis in comparison with the annotation of metatranscriptomes based on public databases. While planning metatranscriptomic experiments, one should balance the benefits and limits of conducting metagenomics in parallel.

5.2.4 Functional activities of *Stylissa carteri* microbiomes

The gene expression profiles of the microbial communities associated with the low-microbial abundance sponge *S. carteri* were characterized with respect to taxonomy and function. The phylum composition of *S. carteri* metatranscriptomes was consistent with the profiles recovered from 16S rRNA gene sequencing (Lee *et al.*, 2011; Giles *et al.*, 2013a; Moitinho-Silva *et al.*, 2014a) in that the bacterial community was dominated by

Proteobacteria, Cyanobacteria, and to a lesser extent, Bacteroidetes, Nitrospirae, and Actinobacteria. Among these, only Nitrospirae transcripts were rare in the *S. carteri* metatranscriptomes, which may indicate that members of this phylum are poorly active *in situ* or that they present low protein turn-over rates. In addition, Thaumarchaota members contributed significantly to the highest expressed prokaryotic genes in *S. carteri*, what is in agreement with previous observations of the presence of archaeal ribo-types in *S. carteri* (Lee *et al.*, 2011). At a lower taxonomic level, Alphaproteobacteria and Gammaproteobacteria were the most abundant classes in *S. carteri* metatranscriptomes and were present at similar proportions. However, Gammaproteobacteria is clearly the dominant class of bacteria according to all studies conducted based on 16S rRNA gene (Lee *et al.*, 2011; Giles *et al.*, 2013a; Moitinho-Silva *et al.*, 2014a). At this taxonomic level, the resolution of the classification was weaker, as inferred from the fact that about half of the proteobacterial reads were not assigned to a given class. The partial identification and shallow depth of taxonomic assignments were previously observed in the metatranscriptome study of the sponge *Geodia barretti* (Radax *et al.*, 2012b) and are likely consequences of the limited sequence length and the lack of reference genomes or metagenomes at databases (Huson *et al.*, 2007; Radax *et al.*, 2012b).

The study of *S. carteri* metatranscriptomes made it possible to assess the *in situ* expression of many gene functions carried out by sponge symbionts (Taylor *et al.*, 2007; Hentschel *et al.*, 2012). Of these, genes coding for enzymes that deal, directly or indirectly, with ammonia were highly expressed by the *S. carteri* microbiota. Most of these genes coded for proteins that participate in pathways of archaeal ammonia oxidation and proteobacterial ammonia assimilation. Ammonia is excreted by sponges and represents a valuable nutritional source for their associated microbiota. Other publications have shown that ammonia-processing genes are present (Hallam *et al.*, 2006; Bayer *et al.*, 2008; Mohamed *et al.*, 2010; Thomas *et al.*, 2010) and highly expressed in sponges (Liu *et al.*, 2012; Radax *et al.*, 2012a; Radax *et al.*, 2012b), confirming the results obtained by measuring intermediate compounds of the nitrogen cycle (e.g. Diaz and Ward, 1997; Bayer *et al.*, 2008; Hoffmann *et al.*, 2009; Ribes *et al.*, 2012). Additionally, these genes were shown to be over-represented in sponge symbionts in relation to free-living seawater microbes, including genes of bacterial ammonia oxidation (Section 4.4), archaeal ammonia oxidation (Fan *et al.*, 2012b), and the gene coding for nitrogen regulatory protein PII (Thomas *et al.*, 2010), a bacterial regulatory protein of

ammonium assimilation. Altogether, these results indicate that ammonia is a key metabolic hub that interfaces sponges and their microbial symbionts.

Genes related to photosynthesis and CO_2 fixation, membrane transporters, and environmental stress were particularly abundant in the *S. carteri* metatranscriptomes. Symbiotic photosynthesis and CO_2 fixation have an ecologically relevant impact on the sponge-host in tropical reefs (Wilkinson, 1983; Cheshire and Wilkinson, 1991; Steindler *et al.*, 2002). Membrane transporters and environmental stress genes were more abundant in sponges-associated microbial communities in comparison with bacterioplankton (Thomas *et al.*, 2010; Fan *et al.*, 2012b). Additionally, these genes were found to be highly expressed by the microbiome of *Cymbastela concentrica* (Liu *et al.*, 2012). Taken together, these results indicate that these functions are particularly related to a sponge-associated life style. However, similar gene functions were highly expressed by different bacterioplankton communities, i.e. genes related with photosynthesis (Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2009; Ottesen *et al.*, 2011; Stewart *et al.*, 2011), diverse membrane transporters (Teeling *et al.*, 2012; Gifford *et al.*, 2013), and stress (Gilbert *et al.*, 2008; Shi *et al.*, 2011). Therefore, these functional categories may not be particularly related to a sponge-associated life style, but rather reflect the diversity of metabolic strategies in the sponge-microbe symbiosis and/or the similar conditions faced by sponge symbionts and marine microorganisms.

The evidence of active methylotrophic metabolism in tropical reef sponges is a novel finding of this Thesis. Methylotrophy is defined as the ability to acquire energy and biomass from reduced one-carbon (C1) compounds, including methane and methanol (Chistoserdova, 2011). Methylotrophic activity in *S. carteri* was inferred by the high expression of gammaproteobacterial formate dehydrogenase-O and methanol dehydrogenase, of which encoded enzymes carry out important reactions in this metabolism (Anthony, 2004; Chistoserdova *et al.*, 2007; Chistoserdova *et al.*, 2009). Additionally, genes related with all the metabolic modules of methylotrophy (Chistoserdova, 2011) were detected in the *S. carteri* metatranscriptomes, including genes that code for enzymes of the formaldehyde metabolism and C1 assimilatory serine-glyoxylate cycle (Section 4.5). The association between sponges and methylotrophic bacteria was only described for deep-sea carnivorous species of the family Cladorhizidae that harbour methanotrophs, a term that refers to the ability to utilize methane (Vacelet *et al.*, 1995; Vacelet and Boury-Esnault, 2002). Because

these sponges lack aquiferous systems, it is hypothesised that they nutritionally benefit from the symbiotic methanotrophs and from predation (Vacelet *et al.*, 1995). In the case of *S. carteri*, it is not clear how exactly these nutritional relationship is established, or if it actually occurs. In this context it is important to know that the central part of the Red Sea is considered a source of atmospheric methanol, based on the annual average of sea-air methanol flux (Millet *et al.*, 2008). Therefore, it is also likely that the gammaproteobacterial symbionts of *S. carteri* metabolise the methanol available from the incoming filtered seawater.

The specific metabolic pathways carried out by the gammaproteobacterial symbionts of *S. carteri* is open to conjecture. In fact, gammaproteobacterial symbionts expressing methanol dehydrogenase genes may not be strict methylotrophs. In other words, gammaproteobacterial symbionts may not use solely C1 compounds to build biomass and obtain energy. For instance, methanol dehydrogenase enzymes do not exclusively oxidise methanol, being capable of transforming a wide range of primary alcohols, including ethanol (Anthony, 1982; Kalyuzhnaya *et al.*, 2008). Secondly, the ability of primarily oxidizing C1 compounds does not necessarily implies in the assimilation of biomass from these compounds, as concluded from the investigation of the C1 metabolism of SAR11 bacteria (Sun *et al.*, 2011), which are abundant members of the surface bacterioplankton (Morris *et al.*, 2002). SAR11 cells produce energy from C1 oxidation, but mainly assimilate biomass from small organic carbon compounds, such as glucose and pyruvate (Sun *et al.*, 2011). Further studies exploring the genomic characteristics and biochemical abilities of *S. carteri* gammaproteobacterial symbionts have the power to elucidate their C1 metabolism. These approaches may expand our understanding of the sponge-microbe symbiosis and its ecological implications for the carbon fluxes in the reef ecosystem.

This Thesis represents the second analysis of sponge metatranscriptomes. Here, we reveal the power of metatranscriptomics to assess the active microbial functions required for living in symbiosis with sponges. In this context, it is important to highlight the distinct perspective obtained by metatranscriptomics in comparison with DNA-based approaches, which is focused on the the presence and abundance of functional genes that are not necessarily expressed. For instance, *Stylissa carteri* microbiomes carry ammonia monooxygenase genes (*amo*) from bacteria and archaea, with a higher abundance of the bacterial genes (Section

4.4). However, *S. carteri* microbiomes expressed exclusively archaeal *amo*; and no bacterial transcripts were detected (Section 4.5). In this context, it would be interesting to expand the current knowledge on the functional equivalence observed at the gene level between sponges (Fan *et al.*, 2012b) or between sponges and seawater (Section 4.4) in light of the gene activity. Overall, my analysis demonstrated the high value of metatranscriptomics to understand the functional roles of microorganisms in sponges.

6 Future directions

The concept of specificity in the association between sponges and microbes was reviewed in this Thesis by use of up-to-date sequencing technologies. Furthermore, the HMA/LMA dichotomy was explored using different technical approaches, including transmission electron microscopy and deep sequencing. Further studies in sponge microbiology should consider these concepts in order to better understand the sponge-microbe symbiosis.

The perspectives of my doctoral Thesis are:

The expansion of genomic databases of sponge-associated microorganisms. Further studies need to be conducted in order to expand the current knowledge of the genomic repertoire of sponge-associated microorganisms. This is a general need of environmental microbiology, and is especially relevant to sponge microbiology. This goal can be obtained by increasing the current sequencing efforts and analysis of sponge-associated metagenomes and genomes. This expansion will be essential to bridge the diversity of the microbial communities associated with sponges and the microbial functions required in this association.

The investigation of sponge microbial diversity in a functional context. Further work should focus on the exploration of ecological observations and theories regarding sponge microbiomes in a functional context. Metatranscriptomics was established as a powerful tool to explore the gene expression of sponge microbiomes. Further experimental studies would benefit from combining metabolic measurements, microbial diversity analysis, and metatranscriptomics. This approach could be instrumental to elucidate the implications of environmental factors (e.g. environmental stressors, niche competition, mutualistic factors, community disturbance) on the metabolism and survivorship of sponge-associated microorganisms, or to investigate the possibility of symbiont horizontal acquisition by the sponge.

The further investigation of methylotrophy in *Stylissa carteri* and other sponges. The evidence of methylotrophy in *S. carteri* symbionts can be further investigated by assessing enzymatic reactions specific of methylotrophy. For instance, methanol dehydrogenase-O gene can be cloned and heterologously expressed towards this purpose. Alternatively, *S. carteri*

tissue could be analysed in order to detect methanol dehydrogenase activity, with basic techniques, such as spectrophotometry, or with advanced methods, such as NanoSIMS. Additionally, it would be relevant to detect the possible sources of methanol and C1 compounds in the environment. This investigation should clarify if sponge methylotrophs are particular to Red Sea reef ecosystems or are expected to occur widely. The latter hypothesis can also tackled by the investigation of genes and transcripts related to methylotrophy in other sponge-associated communities.

The analysis of *Xestospongia testudinaria* metatranscriptomes. In addition to *S. carteri* metatranscriptomes, *X. testudinaria* metatranscriptomes were also isolated and sequenced. Therefore, the analysis of the functional activities of *X. testudinaria* microbiomes is a future perspective resulting from my doctoral Thesis. This analysis can include the comparison between the transcriptional profiles of *X. testudinaria* (HMA) and *S. carteri* (LMA) and the investigation of the gene expressions of sponge-associated symbionts with sequenced genomes, such as Poribacteria.

7 Bibliography

Acker, J., Leptoukh, G., Shen, S., Zhu, T., Kempler, S. (2008) Remotely-sensed chlorophyll a observations of the northern Red Sea indicate seasonal variability and influence of coastal reefs. *Journal of Marine Systems*. 69(3–4), 191–204.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic local alignment search tool. *Journal of Molecular Biology*. 215(3), 403–410.

Amann, R.I., Ludwig, W., Schleifer, K.H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiological Reviews*. 59(1), 143–169.

Andersson, M., Berga, M., Lindström, E.S., Langenheder, S. (2014) The spatial structure of bacterial communities is influenced by historical environmental conditions. *Ecology*. In Press, 140130012013007.

Anthony, C. (1982) *The biochemistry of methylotrophs*. London, United Kingdom: Academic Press.

Anthony, C. (2004) The quinoprotein dehydrogenases for methanol and glucose. *Archives of Biochemistry and Biophysics*. 428(1), 2–9.

Arellano, S.M., Lee, O.O., Lafi, F.F., Yang, J., Wang, Y., Young, C.M., Qian, P.Y. (2013) Deep Sequencing of Myxilla (Ectyomyxilla) methanophila, an epibiotic sponge on cold-seep tubeworms, reveals methylotrophic, thiotrophic, and putative hydrocarbon-degrading microbial associations. *Microbial Ecology*. 65(2), 450–61.

Arillo, A., Bavestrello, G., Burlando, B., Sarà, M. (1993) Metabolic integration between symbiotic cyanobacteria and sponges: a possible mechanism. *Marine Biology*. 117(1), 159–162.

Baas-Becking, L.G.M. (1934) *Geobiologie of inleiding tot de milieukunde*. Den Haag [Netherlands]: W.P. Van Stockum & Zoon N.V.

Bairoch, A. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research*. 28(1), 45–48.

Baker, A.C. (2003) Flexibility and specificity in coral-algal symbiosis: diversity, ecology, and biogeography of *Symbiodinium*. *Annual Review of Ecology, Evolution, and Systematics*. 34(1), 661–689.

Baker, B.J., Sheik, C.S., Taylor, C.A., Jain, S., Bhasi, A., Cavalcoli, J.D., Dick, G.J. (2013) Community transcriptomic assembly reveals microbes that contribute to deep-sea carbon and nitrogen cycling. *The ISME Journal*. 7(10), 1962–1973.

Bandaranayake, W.M., Bemis, J.E., Bourne, D.J. (1996) Ultraviolet absorbing pigments from the marine sponge *Dysidea herbacea*: Isolation and structure of a new mycosporine. *Comparative Biochemistry and Physiology Part C: Pharmacology, Toxicology and Endocrinology*. 115(3), 281–286.

Barton, L., Northrup, D.E. (2011) *Microbial ecology*. 1srt edition. Hoboken, N.J: Wiley-Blackwell.

Bartram, A.K., Lynch, M.D.J., Stearns, J.C., Moreno-Hagelsieb, G., Neufeld, J.D. (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. *Applied and Environmental Microbiology*. 77(11), 3846–3852.

Bayer, K., Kamke, J., Hentschel, U. (2014) Quantification of bacterial and archaeal symbionts in high and low

microbial abundance sponges using real-time PCR. *FEMS Microbiology Ecology*, doi:10.1111/1574–6941.12369.

Bayer, K., Schmitt, S., Hentschel, U. (2007) Microbial nitrification in Mediterranean sponges: possible involvement of ammonia-oxidizing Betaproteobacteria. In *Porifera research: Biodiversity, innovation and sustainability*. Série Livros. Rio de Janeiro: Museu Nacional (Brazil), pp. 165–171.

Bayer, K., Schmitt, S., Hentschel, U. (2008) Physiology, phylogeny and in situ evidence for bacterial and archaeal nitrifiers in the marine sponge *Aplysina aerophoba*. *Environmental Microbiology*. 10(11), 2942–55.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R.R., Stoeck, T. (2011) Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environmental Microbiology*. 13(2), 340–349.

Bell, A.H., Bergquist, P.R., Battershill, C.N. (1999) Feeding biology of *Polymastia croceus*. *Memoirs of the Queensland Museum*. 44, 51–56.

Bell, J.J. (2008) The functional roles of marine sponges. *Estuarine, Coastal and Shelf Science*. 79(3), 341–353.

Benson, D.A. (2004) GenBank. *Nucleic Acids Research*. 33(Database issue), D34–D38.

Björk, J.R., Díez-Vives, C., Coma, R., Ribes, M., Montoya, J.M. (2013) Specificity and temporal dynamics of complex bacteria--sponge symbiotic interactions. *Ecology*. 94(12), 2781–2791.

Blanquer, A., Uriz, M.J., Galand, P.E. (2013) Removing environmental sources of variation to gain insight on symbionts vs. transient microbes in high and low microbial abundance sponges. *Environmental Microbiology*.

Bolker, G.R.W., Bonebakker, L., Gentleman, R., Liaw, W.H.A., Lumley, T., Maechler, M., Magnusson, A., Moeller, S., Schwartz, M., Venables, B. (2012) *gplots: Various R programming tools for plotting data*.

Bonasoro, F., Wilkie, I.C., Bavestrello, G., Cerrano, C., Carnevali, M.D.C. (2001) Dynamic structure of the mesohyl in the sponge *Chondrosia reniformis* (Porifera, Demospongiae). *Zoomorphology*. 121(2), 109–121.

Borojevic, R. (1966) Étude expérimentale de la différenciation des cellules de l'éponge au cours de son développement. *Developmental Biology*. 14(1), 130–153.

Bosch, T.C.G. (2012) What hydra has to say about the role and origin of symbiotic interactions. *The Biological Bulletin*. 223(1), 78–84.

Bourguet-Kondracki, M.L., Rakotoarisoa, M.T., Martin, M.T., Guyot, M. (1992) Bioactive bromopolyacetylenes from the marine sponge *Xestospongia testudinaria*. *Tetrahedron Letters*. 33(2), 225–226.

Boury-Esnault, N., De Vos, L., Donadey, C., Vacelet, J. (1990) Ultrastructure of choanosome and sponge classification. In K. Rützler, ed. *New Perspectives in Sponge Biology*. Washington, D.C.: Smithosonian Institution Press, pp. 237–244.

Bowen, J.L., Morrison, H.G., Hobbie, J.E., Sogin, M.L. (2012) Salt marsh sediment diversity: a test of the variability of the rare biosphere among environmental replicates. *The ISME Journal*. 6(11), 2014–2023.

Brochier-Armanet, C., Boussau, B., Gribaldo, S., Forterre, P. (2008) Mesophilic Crenarchaeota: proposal for a third archaeal phylum, the Thaumarchaeota. *Nature Reviews Microbiology*. 6(3), 245–252.

Brusca, R.C., Brusca, G.J. (2003) *Invertebrates*. 2nd edition. Sunderland, Mass: Sinauer Associates.

Cai, L., Ye, L., Tong, A.H.Y., Lok, S., Zhang, T. (2013) Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. *PloS One*. 8(1), e53649.

Campbell, B.J., Kirchman, D.L. (2013) Bacterial diversity, community structure and potential growth rates along an estuarine salinity gradient. *The ISME Journal*. 7(1), 210–220.

Campbell, B.J., Yu, L., Heidelberg, J.F., Kirchman, D.L. (2011) Activity of abundant and rare bacteria in a coastal ocean. *Proceedings of the National Academy of Sciences of the United States of America*. 108(31), 12776–12781.

Cannistraci, C.V., Ravasi, T., Montevecchi, F.M., Ideker, T., Alessio, M. (2010) Nonlinear dimension reduction and clustering by Minimum Curvilinearity unfold neuropathic pain and tissue embryological classes. *Bioinformatics*. 26(18), i531–9.

Caporaso, J.G., Bittinger, K., Bushman, F.D., DeSantis, T.Z., Andersen, G.L., Knight, R. (2010a) PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics*. 26(2), 266–7.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Pena, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinsky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. (2010b) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*. 7(5), 335–6.

Cárdenas, C.A., Bell, J.J., Davy, S.K., Hoggard, M., Taylor, M.W. (2014) Influence of environmental variation on symbiotic bacterial communities of two temperate sponges. *FEMS Microbiology Ecology*. 88(3), 516–527.

Castresana, J. (2000) Selection of Conserved Blocks from Multiple Alignments for Their Use in Phylogenetic Analysis. *Molecular Biology and Evolution*. 17(4), 540–552.

Cavalcanti, F.F., Klautau, M. (2011) Solenoid: a new aquiferous system to Porifera. *Zoomorphology*. 130(4), 255–260.

Chao, A. (1984) Nonparametric estimation of the number of classes in a population. *Scandinavian Journal of Statistics*. 11, 265–270.

Chao, A., Lee, S.-M. (1992) Estimating the number of classes via sample coverage. *Journal of the American Statistical Association*. 87, 210–217.

Cheshire, A.C., Wilkinson, C.R. (1991) Modelling the photosynthetic production by sponges on Davies Reef, Great Barrier Reef. *Marine Biology*. 109(1), 13–18.

Chistoserdova, L. (2011) Modularity of methylotrophy, revisited: modularity of methylotrophy, revisited. *Environmental Microbiology*. 13(10), 2603–2622.

Chistoserdova, L., Crowther, G.J., Vorholt, J.A., Skovran, E., Portais, J.-C., Lidstrom, M.E. (2007) Identification of a fourth formate dehydrogenase in *Methylobacterium extorquens* AM1 and confirmation of the essential role of formate oxidation in methylotrophy. *Journal of Bacteriology*. 189(24), 9076–9081.

Chistoserdova, L., Kalyuzhnaya, M.G., Lidstrom, M.E. (2009) The expanding world of methylotrophic metabolism. *Annual Review of Microbiology*. 63, 477–499.

Corredor, J.E., Wilkinson, C.R., Vicente, V.P., Morell, J.., Otero, E. (1988) Nitrate release by Caribbean reef sponges. *Limnology and Oceanography*. 33, 114–120.

Cressman, M.D., Yu, Z., Nelson, M.C., Moeller, S.J., Lilburn, M.S., Zerby, H.N. (2010) Interrelations between the microbiotas in the litter and in the intestines of commercial broiler chickens. *Applied and Environmental Microbiology*. 76(19), 6572–6582.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.-F., Guindon, S., Lefort, V.,

Lescot, M., Claverie, J.-M., Gascuel, O. (2008) Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research*. 36(suppl 2), W465–W469.

DeSantis, T.Z., Jr, Hugenholtz, P., Keller, K., Brodie, E.L., Larsen, N., Piceno, Y.M., Phan, R., Andersen, G.L. (2006) NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Research*. 34(Web Server issue), W394–399.

Desqueyroux-Faúndez, R., Valentine, C. (2002) Family Petrosiidae Van Soest, 1980. In *Systema Porifera A Guide to the Classification of Sponges*. Springer, pp. 906–917.

Diaz, M.C., Rützler, K. (2001) Sponges: An essential component of Caribbean coral reefs. *Bulletin of Marine Science*. 69(2), 535–546.

Diaz, M., Ward, B. (1997) Sponge-mediated nitrification in tropical benthic communities. *Marine Ecology Progress Series*. 156, 97–107.

DiBattista, J.D., Berumen, M.L., Gaither, M.R., Rocha, L.A., Eble, J.A., Choat, J.H., Craig, M.T., Skillings, D.J., Bowen, B.W. (2013) After continents divide: comparative phylogeography of reef fishes from the Red Sea and Indian Ocean. *Journal of Biogeography*. 40(6), 1170–1181.

Dohrmann, M., Wörheide, G. (2013) Novel scenarios of early animal evolution—Is it time to rewrite textbooks? *Integrative and Comparative Biology*. 53(3), 503–511.

Doolittle, W.F., Zhaxybayeva, O. (2009) On the origin of prokaryotic species. *Genome Research*. 19(5), 744–756.

Dosse, G. (1939) Bakterien- und Pilzbefunde sowie pathologische und Fäulnisvorgänge in Meeres- und Süsswasserschwämmen. *Zeitschrift für Parasitenkunde*. 11(2-3), 331–356.

Dunlap, W.C., Long, P.F., Jaspars, M. (2011) Biomedicinals from the microbial metagenomes of marine invertebrates. In *Handbook of Molecular Microbial Ecology II*. John Wiley & Sons, Inc., pp. 517–544.

Eder, C., Proksch, P., Wray, V., Steube, K., Bringmann, G., van Soest, R.W.M., Sudarsono, Ferdinandus, E., Pattisina, L.A., Wiryowidagdo, S., Moka, W. (1999) New alkaloids from the indopacific sponge *Stylissa carteri*. *Journal of Natural Products*. 62(1), 184–187.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*. 32(5), 1792–1797.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 26(19), 2460–2461.

Edwards, F.J. (1987) Climate and oceanography. In *Key environments: Red Sea*. Oxford: Pergamon, pp. 45–68.

Ellison, A.M., Farnsworth, E.J., Twilley, R.R. (1996) Facultative mutualism between red mangroves and root-fouling sponges in Belizean mangal. *Ecology*. 77(8), 2431.

El Mamoney, M.H., Khater, A.E.M. (2004) Environmental characterization and radio-ecological impacts of non-nuclear industries on the Red Sea coast. *Journal of Environmental Radioactivity*. 73(2), 151–168.

Elshahed, M.S., Youssef, N.H., Spain, A.M., Sheik, C., Najar, F.Z., Sukharnikov, L.O., Roe, B.A., Davis, J.P., Schloss, P.D., Bailey, V.L., Krumholz, L.R. (2008) Novelty and uniqueness patterns of rare members of the soil biosphere. *Applied and Environmental Microbiology*. 74(17), 5422–5428.

Embree, M., Nagarajan, H., Movahedi, N., Chitsaz, H., Zengler, K. (2014) Single-cell genome and metatranscriptome sequencing reveal metabolic interactions of an alkane-degrading methanogenic community.

The ISME Journal. 8(4), 757–767.

Engelbrektson, A., Kunin, V., Wrighton, K.C., Zvenigorodsky, N., Chen, F., Ochman, H., Hugenholtz, P. (2010) Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *The ISME Journal*. 4(5), 642–647.

Engel, S., Pawlik, J. (2005) Interactions among Florida sponges. I. Reef habitats. *Marine Ecology Progress Series*. 303, 133–144.

Enticknap, J.J., Kelly, M., Peraud, O., Hill, R.T. (2006) Characterization of a culturable alphaproteobacterial symbiont common to many marine sponges and evidence for vertical transmission via sponge larvae. *Applied and Environmental Microbiology*. 72(5), 3724–3732.

Ereskovskii, A.V. (2010) *The comparative embryology of sponges*. Dordrecht; New York: Springer.

Ereskovskii, A.V., Gonobobleva, E.L. (2000) New data on embryonic development of *Halisarca dujardini* Johnston, 1842 (Demospongiae, Halisarcida). *Zoosystema*. 22(2), 355–368.

Ereskovsky, A., Gonobobleva, E., Vishnyakov, A. (2005) Morphological evidence for vertical transmission of symbiotic bacteria in the viviparous sponge *Halisarca dujardini* Johnston (Porifera, Demospongiae, Halisarcida). *Marine Biology*. 146(5), 869–875.

Erpenbeck, D., Weier, T., de Voogd, N.J., Wörheide, G., Sutcliffe, P., Todd, J.A., Michel, E. (2011) Insights into the evolution of freshwater sponges (Porifera: Demospongiae: Spongillina): barcoding and phylogenetic data from Lake Tanganyika endemics indicate multiple invasions and unsettle existing taxonomy. *Molecular Phylogenetics and Evolution*. 61(1), 231–236.

Erwin, P.M., Olson, J.B., Thacker, R.W. (2011) Phylogenetic diversity, host-specificity and community profiling of sponge-associated bacteria in the northern Gulf of Mexico. *PLoS One*. 6(11), e26806.

Fan, L., McElroy, K., Thomas, T. (2012a) Reconstruction of ribosomal RNA genes from metagenomic data. *PloS One*. 7(6), e39948.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., Thomas, T. (2012b) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proceedings of the National Academy of Sciences of the United States of America*. 109(27), E1878–87.

Feike, J., Jurgens, K., Hollibaugh, J.T., Kruger, S., Jost, G., Labrenz, M. (2011) Measuring unbiased metatranscriptomics in suboxic waters of the central Baltic Sea using a new in situ fixation system. *The ISME Journal*. 6(2).

Fiore, C.L., Baker, D.M., Lesser, M.P. (2013) Nitrogen biogeochemistry in the Caribbean sponge, Xestospongia muta: A Source or sink of dissolved inorganic nitrogen? M. R. Mormile, ed. *PLoS One*. 8(8), e72961.

Fiore, C.L., Jarett, J.K., Olson, N.D., Lesser, M.P. (2010) Nitrogen fixation and nitrogen transformations in marine symbioses. *Trends in Microbiology*. 18(10), 455–463.

Franzosa, E.A., Morgan, X.C., Segata, N., Waldron, L., Reyes, J., Earl, A.M., Giannoukos, G., Boylan, M.R., Ciulla, D., Gevers, D., Izard, J., Garrett, W.S., Chan, A.T., Huttenhower, C. (2014) Relating the metatranscriptome and metagenome of the human gut. *Proceedings of the National Academy of Sciences of the United States of America*.

Fraser, C., Alm, E.J., Polz, M.F., Spratt, B.G., Hanage, W.P. (2009) The bacterial species challenge: making sense of genetic and ecological diversity. *Science*. 323(5915), 741–746.

Fraune, S., Bosch, T.C.G. (2007) Long-term maintenance of species-specific bacterial microbiota in the basal

metazoan *Hydra*. *Proceedings of the National Academy of Sciences*. 104(32), 13146–13151.

Frias-Lopez, J., Shi, Y., Tyson, G.W., Coleman, M.L., Schuster, S.C., Chisholm, S.W., Delong, E.F. (2008) Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America*. 105(10), 3805–3810.

Friedrich, A.B., Fischer, I., Proksch, P., Hacker, J., Hentschel, U. (2001) Temporal variation of the microbial community associated with the mediterranean sponge *Aplysina aerophoba*. *FEMS Microbiology Ecology*. 38(2-3), 105–115.

Fritz, J.V., Desai, M.S., Shah, P., Schneider, J.G., Wilmes, P. (2013) From meta-omics to causality: experimental models for human microbiome research. *Microbiome*. 1(1), 14.

Fuhrman, J.A., Hagström, A. (2008) Bacterial and archaeal community structure and its patterns. In D. L. Kirchman, ed. *Microbial Ecology of the Oceans*. Hoboken, N.J: Wiley-Blackwell, pp. 45–90.

Funayama, N. (2008) Stem cell system of sponge. In T. G. Bosch, ed. *Stem Cells*. Springer Netherlands, pp. 17–35.

Futuyma, D.J., Slatkin, M. (1983) *Coevolution*. Sinauer Associates Incorporated.

Gaidos, E., Rusch, A., Ilardo, M. (2011) Ribosomal tag pyrosequencing of DNA and RNA from benthic coral reef microbiota: community spatial structure, rare members and nitrogen-cycling guilds. *Environmental Microbiology*. 13(5), 1138–1152.

Gaino, E., Scalera Liaci, L., Sciscioli, M., Corriero, G. (2006) Investigation of the budding process in Tethya citrina and *Tethya aurantium* (Porifera, Demospongiae). *Zoomorphology*. 125(2), 87–97.

Garrone, R. (1978) *Phylogenesis of connective tissue: morphological aspects and biosynthesis of sponge intercellular matrix*. Basel ; New York: S. Karger.

Gattiker, A., Michoud, K., Rivoire, C., Auchincloss, A.H., Coudert, E., Lima, T., Kersey, P., Pagni, M., Sigrist, C.J.A., Lachaize, C., Veuthey, A.L., Gasteiger, E., Bairoch, A. (2003) Automated annotation of microbial proteomes in SWISS-PROT. *Computational Biology and Chemistry*. 27(1), 49–58.

Gazave, E., Lapébie, P., Ereskovsky, A.V., Vacelet, J., Renard, E., Cárdenas, P., Borchiellini, C. (2012) No longer Demospongiae: Homoscleromorpha formal nomination as a fourth class of Porifera. *Hydrobiologia*. 687(1), 3–10.

Gerçe, B., Schwartz, T., Voigt, M., Rühle, S., Kirchen, S., Putz, A., Proksch, P., Obst, U., Syldatk, C., Hausmann, R. (2009) Morphological, bacterial, and secondary metabolite changes of *Aplysina aerophoba* upon long-term maintenance under artificial conditions. *Microbial Ecology*. 58(4), 865–878.

Gibbons, S.M., Caporaso, J.G., Pirrung, M., Field, D., Knight, R., Gilbert, J.A. (2013) Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences of the United States of America*. 110(12), 4651–4655.

Gifford, S.M., Sharma, S., Booth, M., Moran, M.A. (2013) Expression patterns reveal niche diversification in a marine microbial assemblage. *The ISME journal*. 7(2), 281–298.

Gifford, S.M., Sharma, S., Rinta-Kanto, J.M., Moran, M.A. (2011) Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *The ISME Journal*. 5(3), 461–72.

Gihring, T.M., Green, S.J., Schadt, C.W. (2012) Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology*. 14(2), 285–90.

Gilbert, J.A., Field, D., Huang, Y., Edwards, R., Li, W., Gilna, P., Joint, I. (2008) Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PloS One*. 3(8), e3042.

Giles, E.C., Kamke, J., Moitinho-Silva, L., Taylor, M.W., Hentschel, U., Ravasi, T., Schmitt, S. (2013a) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*. 83(1), 232–41.

Giles, E.C., Saenz-Agudelo, P., Berumen, M.L., Ravasi, T. (2013b) Novel polymorphic microsatellite markers developed for a common reef sponge, *Stylissa carteri*. *Marine Biodiversity*. 43(3), 237–241.

Ginsburg, R.N. (1957) Early diagenesis and lithification of shallow-water carbonate sediments in south Florida. *Special Publications of the Society of Palaeontology and Minerals*. 5, 80–100.

Gloeckner, V., Lindquist, N., Schmitt, S., Hentschel, U. (2013) *Ectyoplasia ferox*, an experimentally tractable model for vertical microbial transmission in marine sponges. *Microbial Ecology*. 65(2), 462–74.

Gloeckner, V., Wehrl, M., Moitinho-Silva, L., Gernert, C., Schupp, P.J., Pawlik, J.R., Lindquist, N.L., Erpenbeck, D., Wörheide, G., Hentschel, U. (2014) The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species. *The Biological Bulletin*. In Press.

De Goeij, J.M., van den Berg, H., van Oostveen, M., Epping, E., van Duyl, F. (2008a) Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. *Marine Ecology Progress Series*. 357, 139–151.

De Goeij, J.M., Moodley, L., Houtekamer, M., Carballeira, N.M., van Duyl, F.C. (2008b) Tracing 13C-enriched dissolved and particulate organic carbon in the bacteria-containing coral reef sponge Halisarca caerulea: Evidence for DOM-feeding. *Limnology and Oceanography*. 53(4), 1376–1386.

De Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., Admiraal, W. (2013) Surviving in a marine desert: The sponge loop retains resources within coral reefs. *Science*. 342(6154), 108–110.

Gomez-Alvarez, V., Teal, T.K., Schmidt, T.M. (2009) Systematic artifacts in metagenomes from complex microbial communities. *The ISME Journal*. 3(11), 1314–1317.

Gonzalez, J.M., Portillo, M.C., Belda-Ferre, P., Mira, A. (2012) Amplification by PCR artificially reduces the proportion of the rare biosphere in microbial communities. *PLoS One*. 7(1), e29973.

Gonzalez, J.M., Zimmermann, J., Saiz-Jimenez, C. (2005) Evaluating putative chimeric sequences from PCR-amplified products. *Bioinformatics*. 21(3), 333–7.

Gosalbes, M.J., Durbán, A., Pignatelli, M., Abellan, J.J., Jiménez-Hernández, N., Pérez-Cobas, A.E., Latorre, A., Moya, A. (2011) Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One*. 6(3), e17447.

Haas, B.J., Gevers, D., Earl, A.M., Feldgarden, M., Ward, D.V., Giannoukos, G., Ciulla, D., Tabbaa, D., Highlander, S.K., Sodergren, E., Methe, B., DeSantis, T.Z., Petrosino, J.F., Knight, R., Birren, B.W. (2011) Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*. 21(3), 494–504.

Haegeman, B., Hamelin, J., Moriarty, J., Neal, P., Dushoff, J., Weitz, J.S. (2013) Robust estimation of microbial diversity in theory and in practice. *The ISME Journal*. 7(6), 1092–1101.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y., Sugahara, J., Preston, C., de la Torre, J., Richardson, P.M., DeLong, E.F. (2006) Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proceedings of the National Academy of Sciences of the United States of America*.

103(48), 18296–18301.

Hampton-Marcell, J.T., Moormann, S.M., Owens, S.M., Gilbert, J.A. (2013) Preparation and metatranscriptomic analyses of host-microbe systems. *Methods in enzymology*. 531, 169–185.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., Martiny, J.B.H. (2012) Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*. 10(7), 497–506.

Hentschel, U., Fieseler, L., Wehrl, M., Gernert, C., Steinert, M., Hacker, J., Horn, M. (2003) Microbial diversity of marine sponges. *Progress in Molecular and Subcellular Biology*. 37, 59–88.

Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J., Moore, B.S. (2002) Molecular evidence for a uniform microbial community in sponges from different oceans. *Applied and Environmental Microbiology*. 68(9), 4431–40.

Hentschel, U., Piel, J., Degnan, S.M., Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nature Reviews Microbiology*. 10(9), 641–54.

Hentschel, U., Usher, K.M., Taylor, M.W. (2006) Marine sponges as microbial fermenters. *FEMS Microbiology Ecology*. 55(2), 167–77.

He, Z., Van Nostrand, J.D., Zhou, J. (2012) Applications of functional gene microarrays for profiling microbial communities. *Current Opinion in Biotechnology*. 23(3), 460–466.

Hoffmann, F., Radax, R., Woebken, D., Holtappels, M., Lavik, G., Rapp, H.T., Schläppy, M.-L., Schleper, C., Kuypers, M.M.M. (2009) Complex nitrogen cycling in the sponge *Geodia barretti*. *Environmental Microbiology*. 11(9), 2228–2243.

Hoffmann, F., Rapp, H., Reitner, J. (2006) Monitoring microbial community composition by Fluorescence In Situ Hybridization during cultivation of the marine cold-water sponge *Geodia barretti*. *Marine Biotechnology*. 8(4), 373–379.

Hooper, J.N.A., van Soest, R.W.M. (2002) Class Demospongiae Sollas, 1885. In *Systema Porifera A Guide to the Classification of Sponges*. Springer, pp. 15–51.

Hooper, J.N.., Lévi, C. (1994) Biogeography of Indo-west Pacific sponges: Microcionidae, Raspailiidae, Axinellidae. In *Sponges in time and space*. Rotterdam: Balkema, pp. 191–212.

Hugoni, M., Taib, N., Debroas, D., Domaizon, I., Jouan Dufournel, I., Bronner, G., Salter, I., Agogué, H., Mary, I., Galand, P.E. (2013) Structure of the rare archaeal biosphere and seasonal dynamics of active ecotypes in surface coastal waters. *Proceedings of the National Academy of Sciences of the United States of America*. 110(15), 6004–6009.

Hunter, S., Jones, P., Mitchell, A., Apweiler, R., Attwood, T.K., Bateman, A., Bernard, T., Binns, D., Bork, P., Burge, S., de Castro, E., Coggill, P., Corbett, M., Das, U., Daugherty, L., Duquenne, L., Finn, R.D., Fraser, M., Gough, J., Haft, D., Hulo, N., Kahn, D., Kelly, E., Letunic, I., Lonsdale, D., Lopez, R., Madera, M., Maslen, J., McAnulla, C., McDowall, J., McMenamin, C., Mi, H., Mutowo-Muellenet, P., Mulder, N., Natale, D., Orengo, C., Pesseat, S., Punta, M., Quinn, A.F., Rivoire, C., Sangrador-Vegas, A., Selengut, J.D., Sigrist, C.J.A., Scheremetjew, M., Tate, J., Thimmajanarthanan, M., Thomas, P.D., Wu, C.H., Yeats, C., Yong, S.-Y. (2012) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*. 40(D1), D306–D312.

Huse, S.M., Welch, D.M., Morrison, H.G., Sogin, M.L. (2010) Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environmental Microbiology*. 12(7), 1889–1898.

Huson, D.H., Auch, A.F., Qi, J., Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome*

Research. 17(3), 377–386.

Inoue, H., Nojima, H., Okayama, H. (1990) High efficiency transformation of *Escherichia coli* with plasmids. *Gene*. 96(1), 23–28.

Jackson, S.A., Kennedy, J., Morrissey, J.P., O'Gara, F., Dobson, A.D. (2012) Pyrosequencing reveals diverse and distinct sponge-specific microbial communities in sponges from a single geographical location in Irish waters. *Microbial Ecology*. 64(1), 105–16.

Jang, H.B., Kim, Y.K., Del Castillo, C.S., Nho, S.W., Cha, I.S., Park, S.B., Ha, M.A., Hikima, J.-I., Hong, S.J., Aoki, T., Jung, T.S. (2012) RNA-seq-based metatranscriptomic and microscopic investigation reveals novel metalloproteases of *Neobodo* sp. as potential virulence factors for soft tunic syndrome in *Halocynthia roretzi*. *PloS One*. 7(12), e52379.

Jiménez, E., Ribes, M. (2007) Sponges as a source of dissolved inorganic nitrogen: Nitrification mediated by temperate sponges. *Limnology and Oceanography*. 52(3), 948–958.

Jones, S.E., Lennon, J.T. (2010) Dormancy contributes to the maintenance of microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*. 107(13), 5881–5886.

Kalyuzhnaya, M.G., Hristova, K.R., Lidstrom, M.E., Chistoserdova, L. (2008) Characterization of a novel methanol dehydrogenase in representatives of *Burkholderiales*: implications for environmental detection of methylotrophy and evidence for convergent evolution. *Journal of Bacteriology*. 190(11), 3817–3823.

Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K., Woyke, T., Hentschel, U. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *The ISME Journal*.

Kamke, J., Taylor, M.W., Schmitt, S. (2010) Activity profiles for marine sponge-associated bacteria obtained by 16S rRNA vs 16S rRNA gene comparisons. *The ISME Journal*. 4(4), 498–508.

Kanehisa, M., Goto, S. (2000) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*. 28(1), 27–30.

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., Mclnerney, J.O. (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC evolutionary biology*. 6, 29.

Kent, W.J. (2002) BLAT - The BLAST-Like Alignment Tool. *Genome Research*. 12(4), 656–664.

Kikuchi, Y., Hosokawa, T., Nikoh, N., Meng, X.-Y., Kamagata, Y., Fukatsu, T. (2009) Host-symbiont co-speciation and reductive genome evolution in gut symbiotic bacteria of acanthosomatid stinkbugs. *BMC Biology*. 7(1), 2.

Kirchman, D.L. (2008) *Microbial ecology of the oceans*. 2nd edition. Hoboken, N.J: Wiley-Blackwell.

Kjelleberg, S. (2002) Environmental biotechnology. *Current Opinion in Biotechnology*. 13(3), 199–203.

Kobayashi, H., Kitamura, K., Nagai, K., Nakao, Y., Fusetani, N., van Soest, R.W.M., Matsunaga, S. (2007) Carteramine A, an inhibitor of neutrophil chemotaxis, from the marine sponge *Stylissa carteri*. *Tetrahedron Letters*. 48(12), 2127–2129.

Korotkova, G.P. (1981) *Origin and evolution of ontogenesis*. Leningrad: Leningrad University Press.

Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., Schloss, P.D. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina

sequencing platform. *Applied and Environmental Microbiology*. 79(17), 5112–5120.

Kuczynski, J., Lauber, C.L., Walters, W.A., Parfrey, L.W., Clemente, J.C., Gevers, D., Knight, R. (2012) Experimental and analytical tools for studying the human microbiome. *Nature Reviews Genetics*. 13, 47–58.

Lane, D.J. (1991) 16S/23S rRNA sequencing. In E. Stackebrandt & M. Goodfellow, eds. *Nucleic acid techniques in bacterial systematics*. London, United Kingdom: Wiley, pp. 115–175.

Langenbruch, P.-F. (1988) Body structure of marine sponges. *Zoomorphology*. 108(1), 13–21.

Lanzen, A., Jorgensen, S.L., Bengtsson, M.M., Jonassen, I., Ovreas, L., Urich, T. (2011) Exploring the composition and diversity of microbial communities at the Jan Mayen hydrothermal vent field using RNA and DNA. *FEMS Microbiology Ecology*. 77(3), 577–89.

Laport, M., Santos, O., Muricy, G. (2009) Marine sponges: Potential sources of new antimicrobial drugs. *Current Pharmaceutical Biotechnology*. 10(1), 86–105.

Ledford, H. (2008) The death of microarrays? *Nature*. 455(7215), 847–847.

Lee, O.O., Chui, P.Y., Wong, Y.H., Pawlik, J.R., Qian, P.-Y. (2009) Evidence for vertical transmission of bacterial symbionts from adult to embryo in the Caribbean sponge *Svenzea zeai*. *Applied and Environmental Microbiology*. 75(19), 6147–6156.

Lee, O.O., Wang, Y., Yang, J., Lafi, F.F., Al-Suwailem, A., Qian, P.Y. (2011) Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. *The ISME Journal*. 5(4), 650–64.

Lee, W.L., Reiswig, H.M., Austin, W.C., Lundsten, L. (2012) An extraordinary new carnivorous sponge, *Chondrocladia lyra*, in the new subgenus *Symmetrocladia* (Demospongiae, Cladorhizidae), from off of northern California, USA. *Invertebrate Biology*. 131(4), 259–284.

Leininger, S., Urich, T., Schloter, M., Schwark, L., Qi, J., Nicol, G.W., Prosser, J.I., Schuster, S.C., Schleper, C. (2006) Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature*. 442(7104), 806–809.

Lejon, D.P.H., Kennedy, J., Dobson, A.D.W. (2011) Identification of novel bioactive compounds from the metagenome of the marine sponge *Haliclona simulans*. In *Handbook of Molecular Microbial Ecology II*. John Wiley & Sons, Inc., pp. 553–562.

Lesniewski, R.A., Jain, S., Anantharaman, K., Schloss, P.D., Dick, G.J. (2012) The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *The ISME Journal*. 6(12), 2257–2268.

Lévi, C., Lévi, P. (1976) Embryogenese de *Chondrosia reniformis* (Nardo), demosponge ovipare, et transmission des bacteries symbiotiques. *Annales des Sciences Naturelles - Zoologie et Biologie animale*. 18, 367–380.

Lévi, C., Porte, A. (1962) Etude au microscope électronique de l'éponge *Oscarella lobularis* Schmidt et de sa larve amphiblastula. *Cahiers de Biologie Marine*. 3, 307–315.

Leys, S.P., Hill, A. (2012) The physiology and molecular biology of sponge tissues. *Advances in Marine Biology*. 62, 1–56.

Leys, S.P., Nichols, S.A., Adams, E.D.M. (2009) Epithelia and integration in sponges. *Integrative and Comparative Biology*. 49(2), 167–177.

Li, C.W., Chen, J.Y., Hua, T.E. (1998) Precambrian sponges with cellular structures. *Science*. 279(5352), 879–82.

Linington, R.G., Williams, D.E., Tahir, A., van Soest, R.W.M., Andersen, R.J. (2003) Latonduines A and B, new alkaloids isolated from the marine sponge *Stylissa carteri*: structure elucidation, synthesis, and biogenetic implications. *Organic Letters*. 5(15), 2735–2738.

Liu, M., Fan, L., Zhong, L., Kjelleberg, S., Thomas, T. (2012) Metaproteogenomic analysis of a community of sponge symbionts. *The ISME Journal*. 6(8), 1515–1525.

Liu, Z., DeSantis, T.Z., Andersen, G.L., Knight, R. (2008) Accurate taxonomy assignments from 16S rRNA sequences produced by highly parallel pyrosequencers. *Nucleic Acids Research*. 36(18), e120.

Liu, Z., Lozupone, C., Hamady, M., Bushman, F.D., Knight, R. (2007) Short pyrosequencing reads suffice for accurate microbial community analysis. *Nucleic Acids Research*. 35(18), e120.

Li, Y.-X., Himaya, S., Kim, S.-K. (2013) Triterpenoids of marine origin as anti-cancer agents. *Molecules*. 18(7), 7886–7909.

Lozupone, C., Knight, R. (2005) UniFrac: a new phylogenetic method for comparing microbial communities. *Applied and Environmental Microbiology*. 71(12), 8228–35.

Macdonald, K.S., Rios, R., Duffy, J.E. (2006) Biodiversity, host specificity, and dominance by eusocial species among sponge-dwelling alpheid shrimp on the Belize Barrier Reef. *Diversity and Distributions*. 12(2), 165–178.

Maldonado, M., Ribes, M., van Duyl, F.C. (2012) Nutrient fluxes through sponges. In *Advances in Marine Biology*. Elsevier, pp. 113–182.

Markowitz, V.M., Chen, I.-M.A., Chu, K., Szeto, E., Palaniappan, K., Grechkin, Y., Ratner, A., Jacob, B., Pati, A., Huntemann, M., Liolios, K., Pagani, I., Anderson, I., Mavromatis, K., Ivanova, N.N., Kyrpides, N.C. (2012) IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Research*. 40(Database issue), D123–129.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L., Horner-Devine, M.C., Kane, M., Krumins, J.A., Kuske, C.R., Morin, P.J., Naeem, S., Ovreås, L., Reysenbach, A.-L., Smith, V.H., Staley, J.T. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*. 4(2), 102–112.

Mason, O.U., Hazen, T.C., Borglin, S., Chain, P.S.G., Dubinsky, E.A., Fortney, J.L., Han, J., Holman, H.-Y.N., Hultman, J., Lamendella, R., Mackelprang, R., Malfatti, S., Tom, L.M., Tringe, S.G., Woyke, T., Zhou, J., Rubin, E.M., Jansson, J.K. (2012) Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *The ISME Journal*. 6(9), 1715–1727.

McClintock, J.B., Amsler, C.D., Baker, B.J., van Soest, R.W.M. (2005) Ecology of antarctic marine sponges: an overview. *Integrative and Comparative Biology*. 45(2), 359–368.

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., Lopez, R. (2013) Analysis tool web services from the EMBL-EBI. *Nucleic Acids Research*. 41(Web Server issue), W597–600.

Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E.M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., Edwards, R.A. (2008) The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9(1), 386.

Millet, D.B., Jacob, D.J., Custer, T.G., de Gouw, J.A., Goldstein, A.H., Karl, T., Singh, H.B., Sive, B.C., Talbot, R.W., Warneke, C., Williams, J. (2008) New constraints on terrestrial and oceanic sources of atmospheric methanol. *Atmospheric Chemistry and Physics*. 8(23), 6887–6905.

Mohamed, N.M., Colman, A.S., Tal, Y., Hill, R.T. (2008) Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. *Environmental Microbiology*. 10(11), 2910–2921.

Mohamed, N.M., Saito, K., Tal, Y., Hill, R.T. (2010) Diversity of aerobic and anaerobic ammonia-oxidizing bacteria in marine sponges. *The ISME Journal*. 4(1), 38–48.

Moitinho-Silva, L., Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., Ravasi, T., Hentschel, U. (2014a) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology*. 23(6), 1348–63.

Moitinho-Silva, L., Seridi, L., Ryu, T., Voolstra, C.R., Ravasi, T., Hentschel, U. (2014b) Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. *Environmental Microbiology*. E-pub ahead of print, doi: 10.1111/1462–2920.12533.

Montalvo, N.F., Hill, R.T. (2011) Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts. *Applied and Environmental Microbiology*. 77(20), 7207–16.

Montalvo, N.F., Mohamed, N.M., Enticknap, J.J., Hill, R.T. (2005) Novel actinobacteria from marine sponges. *Antonie van Leeuwenhoek*. 87(1), 29–36.

Morales, S.E., Holben, W.E. (2011) Linking bacterial identities and ecosystem processes: can ‘omic’ analyses be more than the sum of their parts? *FEMS Microbiology Ecology*. 75(1), 2–16.

Moran, M.A. (2009) Metatranscriptomics: Eavesdropping on Complex Microbial Communities. *Microbe*. 4, 329–335.

Moran, M.A., Satinsky, B., Gifford, S.M., Luo, H., Rivers, A., Chan, L.-K., Meng, J., Durham, B.P., Shen, C., Varaljay, V.A., Smith, C.B., Yager, P.L., Hopkinson, B.M. (2013) Sizing up metatranscriptomics. *The ISME Journal*. 7(2), 237–243.

Morris, R.M., Rappé, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., Giovannoni, S.J. (2002) SAR11 clade dominates ocean surface bacterioplankton communities. *Nature*. 420(6917), 806–810.

Nanninga, G.B., Saenz-Agudelo, P., Manica, A., Berumen, M.L. (2014) Environmental gradients predict the genetic population structure of a coral reef fish in the Red Sea. *Molecular Ecology*. 23(3), 591–602.

National Research Council (U.S.) (2007) *The new science of metagenomics: revealing the secrets of our microbial planet*. Washington, DC: National Academies Press.

Ngugi, D.K., Antunes, A., Brune, A., Stingl, U. (2012) Biogeography of pelagic bacterioplankton across an antagonistic temperature-salinity gradient in the Red Sea. *Molecular Ecology*. 21(2), 388–405.

Nguyen, M.T.H.D., Liu, M., Thomas, T. (2014) Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Molecular Ecology*. 23(6), 1635–1645.

Nickel, M., Scheer, C., Hammel, J.U., Herzen, J., Beckmann, F. (2011) The contractile sponge epithelium sensu lato - body contraction of the demosponge *Tethya wilhelma* is mediated by the pinacoderm. *Journal of Experimental Biology*. 214(10), 1692–1698.

Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K., Hongoh, Y., Constantino, R., Uys, V., Zhong, J., Kudo, T., Ohkuma, M. (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology*. 16(6), 1257–1266.

Norton, J.M., Stark, J.M. (2011) Regulation and measurement of nitrification in terrestrial systems. In *Methods in Enzymology*. Burlington: Academic Press, pp. 343–368.

Nyholm, S.V., McFall-Ngai, M.J. (2003) Dominance of *Vibrio fischeri* in secreted mucus outside the light organ of *Euprymna scolopes*: the first site of symbiont specificity. *Applied and Environmental Microbiology*. 69(7),

3932–3937.

Nyholm, S.V., McFall-Ngai, M.J. (2004) The winnowing: establishing the squid-vibrio symbiosis. *Nature Reviews Microbiology*. 2(8), 632–642.

Off, S., Alawi, M., Spieck, E. (2010) Enrichment and physiological characterization of a novel *Nitrospira*-like bacterium obtained from a marine sponge. *Applied and Environmental Microbiology*. 76(14), 4640–4646.

Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Stevens, M.H.H., Wagner, H. (2012) vegan: Community Ecology Package.

O'Malley, M.A. (2007) The nineteenth century roots of 'everything is everywhere'. *Nature Reviews Microbiology*. 5(8), 647–651.

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(1), 35–41.

Ottesen, E.A., Marin, R., Preston, C.M., Young, C.R., Ryan, J.P., Scholin, C.A., Delong, E.F. (2011) Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *The ISME Journal*. 5(12).

Overbeek, R., Begley, T., Butler, R.M., Choudhuri, J.V., Chuang, H.-Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E.D., Gerdes, S., Glass, E.M., Goesmann, A., Hanson, A., Iwata-Reuyl, D., Jensen, R., Jamshidi, N., Krause, L., Kubal, M., Larsen, N., Linke, B., McHardy, A.C., Meyer, F., Neuweger, H., Olsen, G., Olson, R., Osterman, A., Portnoy, V., Pusch, G.D., Rodionov, D.A., Rückert, C., Steiner, J., Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V. (2005) The Subsystems Approach to Genome Annotation and its Use in the Project to Annotate 1000 Genomes. *Nucleic Acids Research*. 33(17), 5691–5702.

Parks, D.H., Beiko, R.G. (2010) Identifying biologically relevant differences between metagenomic communities. *Bioinformatics*. 26(6), 715–21.

Pedrós-Alió, C. (2012) The Rare Bacterial Biosphere. *Annual Review of Marine Science*. 4, 449–466.

Peek, A.S., Feldman, R.A., Lutz, R.A., Vrijenhoek, R.C. (1998) Cospeciation of chemoautotrophic bacteria and deep sea clams. *Proceedings of the National Academy of Sciences of the United States of America*. 95(17), 9962–9966.

Van de Peer, Y., Chapelle, S., De Wachter, R. (1996) A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Research*. 24(17), 3381–3391.

Pérez, F., Granger, B.E. (2007) IPython: a System for Interactive Scientific Computing. *Computing in Science and Engineering*. 9(3), 21–29.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., Matsunaga, S. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proceedings of the National Academy of Sciences of the United States of America*. 101(46), 16222–16227.

Pignatelli, M., Moya, A. (2011) Evaluating the fidelity of de novo short read metagenomic assembly using simulated data. *PloS One*. 6(5), e19984.

Pile, A.J., Grant, A., Hinde, R., Borowitzka, M.A. (2003) Heterotrophy on ultraplankton communities is an important source of nitrogen for a sponge-rhodophyte symbiosis. *The Journal of Experimental Biology*. 206(Pt 24), 4533–4538.

Pile, A.J., Patterson, M.R., Witman, J.D. (1996) In situ grazing on plankton <10 μm by the boreal sponge

Mycale lingua. *Marine Ecology Progress Series*. 141, 95–102.

Pita, L., Turon, X., López-Legentil, S., Erwin, P.M. (2013) Host rules: spatial stability of bacterial communities associated with marine sponges (<i>Ircinia</> spp.) in the Western Mediterranean Sea. *FEMS Microbiology Ecology*. 86(2), 268–276.

Poretsky, R.S., Hewson, I., Sun, S., Allen, A.E., Zehr, J.P., Moran, M.A. (2009) Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology*. 11(6), 1358–1375.

Poretsky, R.S., Sun, S., Mou, X., Moran, M.A. (2010) Transporter genes expressed by coastal bacterioplankton in response to dissolved organic carbon. *Environmental Microbiology*. 12(3), 616–627.

Powell, S., Forslund, K., Szklarczyk, D., Trachana, K., Roth, A., Huerta-Cepas, J., Gabaldón, T., Rattei, T., Creevey, C., Kuhn, M., Jensen, L.J., von Mering, C., Bork, P. (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Research*. 42(Database issue), D231–239.

Proksch, P. (1994) Defensive roles for secondary metabolites from marine sponges and sponge-feeding nudibranchs. *Toxicon*. 32(6), 639–655.

Prosser, J.I., Nicol, G.W. (2008) Relative contributions of archaea and bacteria to aerobic ammonia oxidation in the environment. *Environmental Microbiology*. 10(11), 2931–2941.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B.M., Ludwig, W., Peplies, J., Glockner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research*. 35(21), 7188–96.

Pruitt, K.D., Tatusova, T., Maglott, D.R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*. 33(suppl 1), D501–D504.

Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., Schleper, C. (2012a) Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges: Archaeal nitrification in cold-water sponges. *Environmental Microbiology*. 14(4), 909–923.

Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H.T., Urich, T., Schleper, C. (2012b) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environmental Microbiology*. 14(5).

R Development Core Team (2012) *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

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(3), 453–461.

Reiswig, H.M. (1990) *In situ* feeding in 2 shallow-water hexactinellid sponges. In *New Perspectives in Sponge Biology*. Washington, DC: Smithsonian Institution Press, pp. 504–510.

Reiswig, H.M. (1981) Partial carbon and energy budgets of the bacteriosponge *Verohgia fistularis* (Porifera: Demospongiae) in Barbados. *Marine Ecology*. 2(4), 273–293.

Reiswig, H.M. (1971) Particle feeding in natural populations of three marine demosponges. *The Biological Bulletin*. 141(3), 568–591.

Reiswig, H.M. (1974) Water transport, respiration and energetics of three tropical marine sponges. *Journal of Experimental Marine Biology and Ecology*. 14(3), 231–249.

Reiswig, H.M., Brown, M.J. (1977) The central cells of sponges. *Zoomorphologie*. 88(2), 81–94.

Reveillaud, J., Maignien, L., Murat Eren, A., Huber, J.A., Apprill, A., Sogin, M.L., Vanreusel, A. (2014) Host-specificity among abundant and rare taxa in the sponge microbiome. *The ISME Journal*.

Ribes, M., Coma, R., Gili, J.-M. (1999) Natural diet and grazing rate of the temperate sponge *Dysidea avara* (Demospongiae, Dendroceratida) throughout an annual cycle. *Marine Ecology Progress Series*. 176, 179–190.

Ribes, M., Jiménez, E., Yahel, G., López-Sendino, P., Diez, B., Massana, R., Sharp, J.H., Coma, R. (2012) Functional convergence of microbes associated with temperate marine sponges. *Environmental Microbiology*. 14(5), 1224–1239.

Rivals, I., Personnaz, L., Taing, L., Potier, M.C. (2007) Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*. 23(4), 401–7.

Robinson, C.J., Bohannan, B.J.M., Young, V.B. (2010) From structure to function: the ecology of host-associated microbial communities. *Microbiology and Molecular Biology Reviews*. 74(3), 453–476.

Roh, S.W., Abell, G.C.J., Kim, K.-H., Nam, Y.-D., Bae, J.-W. (2010) Comparing microarrays and next-generation sequencing technologies for microbial ecology research. *Trends in Biotechnology*. 28(6), 291–299.

Rottmann, M., Schröder, H.C., Gramzow, M., Renneisen, K., Kurelec, B., Dorn, A., Friese, U., Müller, W.E. (1987) Specific phosphorylation of proteins in pore complex-laminae from the sponge *Geodia cydonium* by the homologous aggregation factor and phorbol ester. Role of protein kinase C in the phosphorylation of DNA topoisomerase II. *The EMBO Journal*. 6(13), 3939.

Ruppert, E.E. (2004) *Invertebrate zoology: a functional evolutionary approach*. 7th edition. Belmont, CA: Thomson-Brooks/Cole.

Salomon, C., Deerinck, T., Ellisman, M., Faulkner, D. (2001) The cellular localization of dercitamide in the Palauan sponge *Oceanapia sagittaria*. *Marine Biology*. 139(2), 313–319.

Sambrook, J. (2001) *Molecular cloning: a laboratory manual*. 3rd edition. Cold Spring Harbor, N.Y: Cold Spring Harbor Laboratory Press.

Sapp, J. (1994) *Evolution by association: a history of symbiosis*. New York: Oxford University Press.

Sarà, M. (1971) Ultrastructural aspects of the symbiosis between two species of the genus *Aphanocapsa* (Cyanophyceae) and *Ircinia variabilis* (Demospongiae). *Marine Biology*. 11(3), 214–221.

Satinsky, B.M., Gifford, S.M., Crump, B.C., Moran, M.A. (2013) Use of internal standards for quantitative metatranscriptome and metagenome analysis. *Methods in Enzymology*. In press.

Schläppy, M.-L., Schöttner, S.I., Lavik, G., Kuypers, M.M.M., Beer, D., Hoffmann, F. (2010) Evidence of nitrification and denitrification in high and low microbial abundance sponges. *Marine Biology*. 157(3), 593–602.

Schloss, P.D., Gevers, D., Westcott, S.L. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PloS One*. 6(12), e27310.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J., Weber, C.F. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. 75(23), 7537–41.

Schmidt, I. (1970) Phagocytose et pinocytose chez les Spongillidae. *Zeitschrift für vergleichende Physiologie*. 66(4), 398–420.

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. *Applied and Environmental Microbiology*. 74(24), 7694–708.

Schmitt, S., Deines, P., Behnam, F., Wagner, M., Taylor, M.W. (2011) Chloroflexi bacteria are more diverse, abundant, and similar in high than in low microbial abundance sponges. *FEMS Microbiology Ecology*. 78(3), 497–510.

Schmitt, S., Hentschel, U., Taylor, M. (2012a) Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia*. 687(1), 341–351.

Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N., Perez, T., Rodrigo, A., Schupp, P.J., Vacelet, J., Webster, N., Hentschel, U., Taylor, M.W. (2012b) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *The ISME Journal*. 6(3), 564–76.

Schmitt, S., Weisz, J.B., Lindquist, N., Hentschel, U. (2007) Vertical transmission of a phylogenetically complex microbial consortium in the viviparous sponge *Ircinia felix*. *Applied and Environmental Microbiology*. 73(7), 2067–2078.

Shannon, C.E. (1948) A mathematical theory of communication. *The Bell System Technical Journal*. 27, 379–423.

Shao, Y., Arias-Cordero, E., Guo, H., Bartram, S., Boland, W. (2014) In vivo Pyro-SIP assessing active gut microbiota of the cotton leafworm, *Spodoptera littoralis* M. J. Horsburgh, ed. *PLoS One*. 9(1), e85948.

Sharp, K.H., Eam, B., Faulkner, D.J., Haygood, M.G. (2007) Vertical transmission of diverse microbes in the tropical sponge *Corticium* sp. *Applied and Environmental Microbiology*. 73(2), 622–629.

Sheik, C.S., Jain, S., Dick, G.J. (2014) Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. *Environmental Microbiology*. 16(1), 304–317.

Shieh, W.Y., Lin, Y.M. (1994) Association of heterotrophic nitrogen-fixing bacteria with a marine sponge of *Halichondria* sp. *Bulletin of marine science*. 54(2), 557–564.

Shi, Y., Tyson, G.W., Eppley, J.M., DeLong, E.F. (2011) Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *The ISME Journal*. 5(6), 999–1013.

Shokralla, S., Spall, J.L., Gibson, J.F., Hajibabaei, M. (2012) Next-generation sequencing technologies for environmental DNA research: next-generation sequencing for environmental data. *Molecular Ecology*. 21(8), 1794–1805.

Siddall, M., Rohling, E.J., Almogi-Labin, A., Hemleben, C., Meischner, D., Schmelzer, I., Smeed, D.A. (2003) Sea-level fluctuations during the last glacial cycle. *Nature*. 423(6942), 853–858.

Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., Dandekar, T., Hentschel, U. (2011) Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *The ISME Journal*. 5(1), 61–70.

Da Silva, A.M., Young, C.C., Levitus, S. (1994) *Atlas of Surface Marine Data 1994*. Silver Spring, Maryland.

Silverstein, R.N., Correa, A.M.S., Baker, A.C. (2012) Specificity is rarely absolute in coral-algal symbiosis: implications for coral response to climate change. *Proceedings of the Royal Society B: Biological Sciences*. 279(1738), 2609–2618.

Simister, R.L., Deines, P., Botte, E.S., Webster, N.S., Taylor, M.W. (2012a) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environmental Microbiology*. 14(2), 517–24.

Simister, R.L., Taylor, M.W., Tsai, P., Fan, L., Bruxner, T.J., Crowe, M.L., Webster, N. (2012b) Thermal stress responses in the bacterial biosphere of the Great Barrier Reef sponge, *Rhopaloeides odorabile*. *Environmental Microbiology*. 14(12), 3232–46.

Simister, R.L., Taylor, M.W., Tsai, P., Webster, N. (2012c) Sponge-microbe associations survive high nutrients and temperatures. *PLoS One*. 7(12), e52220.

Simister, R., Taylor, M.W., Rogers, K.M., Schupp, P.J., Deines, P. (2013) Temporal molecular and isotopic analysis of active bacterial communities in two New Zealand sponges. *FEMS Microbiology Ecology*. 85(1), 195–205.

Simpson, E.H. (1949) Measurement of Diversity. *Nature*. 163, 688–688.

Simpson, T.L., Vaccaro, C.A. (1974) An ultrastructural study of silica deposition in the freshwater sponge *Spongilla lacustris*. *Journal of Ultrastructure Research*. 47(3), 296–309.

Sipkema, D., Holmes, B., Nichols, S.A., Blanch, H.W. (2009) Biological characterisation of *Haliclona* (?*gellius*) sp.: sponge and associated microorganisms. *Microbial Ecology*. 58(4), 903–20.

Sirocko, F. (2003) Ups and downs in the Red Sea. *Nature*. 423(6942), 813–814.

Sjöstedt, J., Koch-Schmidt, P., Pontarp, M., Canbäck, B., Tunlid, A., Lundberg, P., Hagström, A., Riemann, L. (2012) Recruitment of members from the rare biosphere of marine bacterioplankton communities after an environmental disturbance. *Applied and Environmental Microbiology*. 78(5), 1361–1369.

Smyth, G.K. (2005) Limma: linear models for microarray data. In R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, & W. Huber, eds. *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. New York: Springer, pp. 397–420.

Van Soest, R.W.M., Erpenbeck, D., Alvarez, B. (2002) Family Dictyonellidae Van Soest, Diaz & Pomponin, 1990. In *Systema Porifera A Guide to the Classification of Sponges*. Springer, pp. 773–786.

Sofianos, S.S., Johns, W.E. (2002) An Oceanic General Circulation Model (OGCM) investigation of the Red Sea circulation, 1. Exchange between the Red Sea and the Indian Ocean. *Journal of Geophysical Research*. 107(C11)(107(C11)), 11.

Sogin, M.L., Morrison, H.G., Huber, J.A., Mark Welch, D., Huse, S.M., Neal, P.R., Arrieta, J.M., Herndl, G.J. (2006) Microbial diversity in the deep sea and the underexplored ‘rare biosphere’. *Proceedings of the National Academy of Sciences of the United States of America*. 103(32), 12115–20.

Southwell, M.W., Popp, B.N., Martens, C.S. (2008) Nitrification controls on fluxes and isotopic composition of nitrate from Florida Keys sponges. *Marine Chemistry*. 108(1–2), 96–108.

Stachowicz, J.J., Hay, M.E. (2000) Geographic variation in camouflage specialization by a Decorator Crab. *The American naturalist*. 156(1), 59–71.

Stackebrandt, E., Goebel, B.M. (1994) Taxonomic note: A place for DNA-DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. *International Journal of Systematic Bacteriology*. 44(4), 846–849.

Staden, R. (1996) The staden sequence analysis package. *Molecular Biotechnology*. 5(3), 233–241.

Staley, J.T., Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology*. 39, 321–346.

Steger, D., Ettinger-Epstein, P., Whalan, S., Hentschel, U., de Nys, R., Wagner, M., Taylor, M.W. (2008)

Diversity and mode of transmission of ammonia-oxidizing archaea in marine sponges. *Environmental Microbiology*. 10(4), 1087–1094.

Steindler, L., Beer, S., Ilan, M. (2002) *Photosymbiosis in intertidal and subtidal tropical sponges*. Philadelphia, PA, USA: Balaban.

Stewart, F.J., Ottesen, E.A., DeLong, E.F. (2010) Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME Journal*. 4(7), 896–907.

Stewart, F.J., Sharma, A.K., Bryant, J.A., Eppley, J.M., DeLong, E.F. (2011) Community transcriptomics reveals universal patterns of protein sequence conservation in natural microbial communities. *Genome Biology*. 12(3), R26.

Stowe, S.D., Richards, J.J., Tucker, A.T., Thompson, R., Melander, C., Cavanagh, J. (2011) Anti-biofilm compounds derived from marine sponges. *Marine Drugs*. 9(12), 2010–2035.

Sudek, S., Lopanik, N.B., Waggoner, L.E., Hildebrand, M., Anderson, C., Liu, H., Patel, A., Sherman, D.H., Haygood, M.G. (2007) Identification of the putative bryostatin polyketide synthase gene cluster from ‘Candidatus *Endobugula sertula*’, the uncultivated microbial symbiont of the marine bryozoan *Bugula neritina*. *Journal of natural products*. 70(1), 67–74.

Sun, J., Steindler, L., Thrash, J.C., Halsey, K.H., Smith, D.P., Carter, A.E., Landry, Z.C., Giovannoni, S.J. (2011) One carbon metabolism in SAR11 pelagic marine bacteria. *PloS One*. 6(8), e23973.

Taylor, M.W., Hill, R.T., Hentschel, U. (2011) Meeting report: 1st international symposium on sponge microbiology. *Marine biotechnology (New York, N.Y.)*. 13(6), 1057–1061.

Taylor, M.W., Radax, R., Steger, D., Wagner, M. (2007) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiology and Molecular Biology Reviews*. 71(2), 295–347.

Taylor, M.W., Schupp, P.J., Dahllöf, I., Kjelleberg, S., Steinberg, P.D. (2004) Host specificity in marine sponge-associated bacteria, and potential implications for marine microbial diversity. *Environmental Microbiology*. 6(2), 121–130.

Taylor, M.W., Schupp, P.J., De Nys, R., Kjelleberg, S., Steinberg, P.D. (2005) Biogeography of bacteria associated with the marine sponge *Cymbastela concentrica*. *Environmental Microbiology*. 7(3), 419–433.

Taylor, M.W., Tsai, P., Simister, R.L., Deines, P., Botte, E., Ericson, G., Schmitt, S., Webster, N.S. (2013) ‘Sponge-specific’ bacteria are widespread (but rare) in diverse marine environments. *The ISME Journal*. 7(2), 438–43.

Teeling, H., Fuchs, B.M., Becher, D., Klockow, C., Gardebrecht, A., Bennke, C.M., Kassabgy, M., Huang, S., Mann, A.J., Waldmann, J., Weber, M., Klindworth, A., Otto, A., Lange, J., Bernhardt, J., Reinsch, C., Hecker, M., Peplies, J., Bockelmann, F.D., Callies, U., Gerdts, G., Wichels, A., Wiltshire, K.H., Glockner, F.O., Schweder, T., Amann, R. (2012) Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science*. 336(6081), 608–611.

Thacker, R.W., Starnes, S. (2003) Host specificity of the symbiotic cyanobacterium Oscillatoria spongeliae in marine sponges, *Dysidea* spp. *Marine Biology*. 142(4), 643–648.

Thamdrup, B. (2012) New pathways and processes in the global nitrogen cycle. *Annual Review of Ecology, Evolution, and Systematics*. 43(1), 407–428.

Thomas, T., Gilbert, J., Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*. 2(1), 3.

Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A., Heidelberg, K.B., Egan, S., Steinberg, P.D., Kjelleberg, S. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *The ISME Journal*. 4(12), 1557–67.

Thoms, C., Horn, M., Wagner, M., Hentschel, U., Proksch, P. (2003) Monitoring microbial diversity and natural product profiles of the sponge *Aplysina cavernicola* following transplantation. *Marine Biology*. 142(4), 685–692.

Tringe, S.G., Hugenholtz, P. (2008) A renaissance for the pioneering 16S rRNA gene. *Current Opinion in Microbiology*. 11(5), 442–446.

Tripp, H.J., Hewson, I., Boyarsky, S., Stuart, J.M., Zehr, J.P. (2011) Misannotations of rRNA can now generate 90% false positive protein matches in metatranscriptomic studies. *Nucleic Acids Research*. 39(20), 8792–8802.

Tu, Q., Yu, H., He, Z., Deng, Y., Wu, L., Van Nostrand, J.D., Zhou, A., Voordeckers, J., Lee, Y.-J., Qin, Y., Hemme, C.L., Shi, Z., Xue, K., Yuan, T., Wang, A., Zhou, J. (2014) GeoChip 4: a functional gene-array-based high-throughput environmental technology for microbial community analysis. *Molecular Ecology Resources*, n/a–n/a.

Turon, X., Becerro, M.A., Uriz, M.J. (2000) Distribution of brominated compounds within the sponge *Aplysina aerophoba*: coupling of X-ray microanalysis with cryofixation techniques. *Cell and Tissue Research*. 301(2), 311–322.

Unson, M.D., Holland, N.D., Faulkner, D.J. (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), 1–11.

Urich, T., Lanzén, A., Qi, J., Huson, D.H., Schleper, C., Schuster, S.C. (2008) Simultaneous assessment of soil microbial community atructure and function through analysis of the meta-transcriptome N. Ward, ed. *PLoS One*. 3(6), e2527.

Uriz, M.J., Becerro, M.A., Tur, J.M., Turon, X. (1996) Location of toxicity within the Mediterranean sponge *Crambe crambe* (Demospongiae: Poecilosclerida). *Marine Biology*. 124(4), 583–590.

Usher, K., Kuo, J., Fromont, J., Sutton, D. (2001) Vertical transmission of cyanobacterial symbionts in the marine sponge *Chondrilla australiensis* (Demospongiae). *Hydrobiologia*. 461(1-3), 9–13.

Vacelet, J. (1975) Etude en microscopie electronique de l’association entre bacteries et spongiaires du genre *Verongia* (Dictyoceratida). *Journal de Microscopie et Biologie Cellulares*. 23, 271–288.

Vacelet, J., Boury-Esnault, N. (2002) A new species of carnivorous deep-sea sponge (Demospongiae: Cladorhizidae) associated with methanotrophic bacteria. *Cahiers de Biologie Marine*. 43(2), 141–148.

Vacelet, J., Boury-Esnault, N. (1995) Carnivorous sponges. *Nature*. 373(6512), 333–335.

Vacelet, J., Boury-Esnault, N., Fiala-Medioni, A., Fisher, C.R. (1995) A methanotrophic carnivorous sponge. *Nature*. 377(6547), 296–296.

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.

Vacelet, J., Fiala-Médioni, A., Fisher, C., Boury-Esnault, N. (1996) Symbiosis between methane-oxidizing bacteria and a deep-sea carnivorous cladorhizid sponge. *Marine Ecology Progress Series*. 145, 77–85.

Verberkmoes, N.C., Russell, A.L., Shah, M., Godzik, A., Rosenquist, M., Halfvarson, J., Lefsrud, M.G., Apajalahti, J., Tysk, C., Hettich, R.L., Jansson, J.K. (2009) Shotgun metaproteomics of the human distal gut

microbiota. *The ISME Journal*. 3(2), 179–189.

Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N., Knight, R. (2011) PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics*. 27(8), 1159–61.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R. (2007) Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*. 73(16), 5261–5267.

Wang, Y., Qian, P.-Y. (2009) Conservative fragments in bacterial 16S rRNA genes and primer design for 16S ribosomal DNA amplicons in metagenomic studies. *PloS One*. 4(10), e7401.

Wattam, A.R., Abraham, D., Dalay, O., Disz, T.L., Driscoll, T., Gabbard, J.L., Gillespie, J.J., Gough, R., Hix, D., Kenyon, R., Machi, D., Mao, C., Nordberg, E.K., Olson, R., Overbeek, R., Pusch, G.D., Shukla, M., Schulman, J., Stevens, R.L., Sullivan, D.E., Vonstein, V., Warren, A., Will, R., Wilson, M.J.C., Yoo, H.S., Zhang, C., Zhang, Y., Sobral, B.W. (2013) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*.

Webster, N.S., Luter, H.M., Soo, R.M., Botté, E.S., Simister, R.L., Abdo, D., Whalan, S. (2013) Same, same but different: symbiotic bacterial associations in GBR sponges. *Frontiers in Microbiology*. 3.

Webster, N.S., Negri, A.P., Munro, M.M.H.G., Battershill, C.N. (2004) Diverse microbial communities inhabit Antarctic sponges. *Environmental Microbiology*. 6(3), 288–300.

Webster, N.S., Taylor, M.W., Behnam, F., Lucker, S., Rattei, T., Whalan, S., Horn, M., Wagner, M. (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environmental Microbiology*. 12(8), 2070–82.

Wehrl, M., Steinert, M., Hentschel, U. (2007) Bacterial uptake by the marine sponge *Aplysina aerophoba*. *Microbial Ecology*. 53(2), 355–65.

Weisz, J.B., Lindquist, N., Martens, C.S. (2008) Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? *Oecologia*. 155(2), 367–76.

Weisz, J., Hentschel, U., Lindquist, N., Martens, C. (2007) Linking abundance and diversity of sponge-associated microbial communities to metabolic differences in host sponges. *Marine Biology*. 152(2), 475–483.

Werner, J.J., Koren, O., Hugenholtz, P., DeSantis, T.Z., Walters, W.A., Caporaso, J.G., Angenent, L.T., Knight, R., Ley, R.E. (2012) Impact of training sets on classification of high-throughput bacterial 16s rRNA gene surveys. *The ISME Journal*. 6(1), 94–103.

White, J.R., Patel, J., Ottesen, A., Arce, G., Blackwelder, P., Lopez, J.V. (2012) Pyrosequencing of bacterial symbionts within *Axinella corrugata* sponges: diversity and seasonal variability. *PLoS One*. 7(6), e38204.

Whitfield, J. (2005) Is everything everywhere? *Science*. 310(5750), 960–961.

Wichels, A., Würtz, S., Döpke, H., Schütt, C., Gerdts, G. (2006) Bacterial diversity in the breadcrumb sponge *Halichondria panicea* (Pallas). *FEMS Microbiology Ecology*. 56(1), 102–118.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E.M., Kyrpides, N., Mavrommatis, K., Meyer, F. (2012) The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinformatics*. 13(1), 141.

Wilkinson, C.R. (1987) Interocean differences in size and nutrition of coral reef sponge populations. *Science*. 236(4809), 1654–1657.

Wilkinson, C.R. (1978) Microbial associations in sponges. II. Numerical analysis of sponge and water bacterial populations. *Marine Biology*. 49(2), 169–176.

Wilkinson, C.R. (1983) Net primary productivity in coral reef sponges. *Science*. 219(4583), 410–412.

Wilkinson, C.R. (1992) Symbiotic interactions between marine sponges and algae. In *Algae and Symbioses: Plants, Animals, Fungi, Viruses, Interactions Explored*. Bristol: Biopress, pp. 112–151.

Wilkinson, C.R., Fay, P. (1979) Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature*. 279(5713), 527–529.

Wilson, M.C., Mori, T., Rückert, C., Uria, A.R., Helf, M.J., Takada, K., Gernert, C., Steffens, U.A.E., Heycke, N., Schmitt, S., Rinke, C., Helfrich, E.J.N., Brachmann, A.O., Gurgui, C., Wakimoto, T., Kracht, M., Crüsemann, M., Hentschel, U., Abe, I., Matsunaga, S., Kalinowski, J., Takeyama, H., Piel, J. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature*. 506(7486), 58–62.

De Wit, R., Bouvier, T. (2006) 'Everything is everywhere, but, the environment selects'; what did Baas Becking and Beijerinck really say? *Environmental Microbiology*. 8(4), 755–758.

Witte, U., Brattegard, T., Graf, G., Springer, B. (1997) Particle capture and deposition by deep-sea sponges from the Norwegian-Greenland Sea. *Marine Ecology Progress Series*. 154, 241–252.

Worden, A.Z., Not, F. (2008) Ecology and diversity of picoeukaryotes. In D. L. Kirchman, ed. *Microbial Ecology of the Oceans*. Hoboken, N.J: Wiley-Blackwell, pp. 159–206.

Wörheide, G., Dohrmann, M., Erpenbeck, D., Larroux, C., Maldonado, M., Voigt, O., Borchiellini, C., Lavrov, D.V. (2012) Deep phylogeny and evolution of sponges (phylum Porifera). In *Advances in Marine Biology*. Elsevier, pp. 1–78.

Wulff, J.L. (2001) Assessing and monitoring coral reef sponges: Why and how? *Bulletin of Marine Science*. 69(2), 831–846.

Wulff, J.L. (2008) Collaboration among sponge species increases sponge diversity and abundance in a seagrass meadow. *Marine Ecology*. 29(2), 193–204.

Wulff, J.L. (1984) Sponge-mediated coral reef growth and rejuvenation. *Coral Reefs*. 3(3), 157–163.

Wulff, J.L., Buss, L.W. (1979) Do sponges help hold coral reefs together? *Nature*. 281(5731), 474–475.

Xu, L., Chen, H., Hu, X., Zhang, R., Zhang, Z., Luo, Z.W. (2006) Average gene length is highly conserved in prokaryotes and eukaryotes and diverges only between the two kingdoms. *Molecular Biology and Evolution*. 23(6), 1107–1108.

Yahel, G., Sharp, J.H., Marie, D., Häse, C., Genin, A. (2003) In situ feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: Bulk DOC is the major source for carbon. *Limnology and Oceanography*. 48(1), 141–149.

Yang, Z., Li, Z. (2012) Spatial distribution of prokaryotic symbionts and ammoxidation, denitrifier bacteria in marine sponge *Astrosclera willeyana*. *Scientific Reports*. 2(528).

Zakai, D., Chadwick-Furman, N.E. (2002) Impacts of intensive recreational diving on reef corals at Eilat, northern Red Sea. *Biological Conservation*. 105(2), 179–187.

Zarraonaindia, I., Smith, D.P., Gilbert, J.A. (2013) Beyond the genome: community-level analysis of the microbial world. *Biology & Philosophy*. 28(2), 261–282.

8 Appendix

Statement of individual author contributions and of legal second publication rights

Publication (complete reference): **Moitinho-Silva, L.**, Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., Ravasi, T., Hentschel, U. (2014) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. *Molecular Ecology*. 23(6), 1348–63.

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	**LMS**	UH	TRa		
Data Collection	**LMS**	ECG	TRa		
Data Analysis and Interpretation	**LMS**	KB	CVC	TRy	LS
Manuscript Writing	**LMS**	UH	TRa		

Explanations (if applicable):

Publication (complete reference): **Moitinho-Silva, L.**, Seridi, L., Ryu, T., Voolstra, C.R., Ravasi, T., Hentschel, U. (2014) Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. *Environmental Microbiology*. E-pub ahead of print, doi: 10.1111/1462–2920.12533.

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	**LMS**	UH	TRa	CRV	
Data Collection	**LMS**	TRy	LS	TRa	
Data Analysis and Interpretation	**LMS**	LS	TRy		
Manuscript Writing	**LMS**	UH	TRa	CRV	

Explanations (if applicable):

Publication (complete reference): Bayer, K.[#], **Moitinho-Silva, L.[#]**, Brümmer, F., Cannistraci, C.V., Ravasi, T., Hentschel, U. (2014) GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater. *FEMS Microbiology Ecology*. Manuscript submitted.

Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	KB[#]	**LMS[#]**	UH	TRa	
Data Collection	KB[#]	**LMS[#]**	FB	TRa	
Data Analysis and Interpretation	KB[#]	**LMS[#]**	CVC	UH	TRa
Manuscript Writing	UH	KB[#]	**LMS[#]**		

Explanations (if applicable): [#]Authors contributed equally to this work.

Publication (complete reference): Giles, E.C., Kamke, J., **Moitinho-Silva, L.**, Taylor, M.W., Hentschel, U., Ravasi, T., Schmitt, S. (2013) Bacterial community profiles in low microbial abundance sponges. *FEMS Microbiology Ecology*. 83(1), 232–41.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	SS	UH	MWT	ECG	
Data Collection	ECG	JK	**LMS**		
Data Analysis and Interpretation	SS	ECG	JK	**LMS**	
Manuscript Writing	SS	UH	MWT	TR	

Explanations (if applicable):

Publication (complete reference): Gloeckner, V., Wehrl, M., **Moitinho-Silva, L.**, Gernert, C., Schupp, P.J., Pawlik, J.R., Lindquist, N.L., Erpenbeck, D., Wörheide, G., Hentschel, U. (2014) The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species. *The Biological Bulletin*. In Press.					
Participated in	**Author Initials,** Responsibility decreasing from left to right				
Study Design	UH	VG	MW		
Data Collection	VG	MW	UH	**LMS**	CG
Data Analysis and Interpretation	VG	MW	UH	GW	DE
Manuscript Writing	UH	VG	**LMS**		

Explanations (if applicable):

I confirm that I have obtained permission from both the publishers and the co-authors for legal second publication.

I also confirm my primary supervisor's acceptance.

Lucas Moitinho e Silva   25.06.2014   Wuerzburg, Germany

Signature

9 Publications list

Publications derived from the doctoral research

1. Bayer, K., **Moitinho-Silva, L.**, Brümmer, F., Cannistraci, C.V., Ravasi, T., Hentschel, U. (2014) GeoChip-based insights into the microbial functional gene repertoire of marine sponges (HMA, LMA) and seawater. **FEMS Microbiology Ecology**. Submitted.
2. **Moitinho-Silva, L.**, Seridi, L., Ryu, T., Voolstra, C.R., Ravasi, T., Hentschel, U. (2014) Revealing microbial functional activities in the Red Sea sponge *Stylissa carteri* by metatranscriptomics. **Environmental Microbiology**. E-pub ahead of print, doi: 10.1111/1462–2920.12533.
3. Gloeckner, V., Wehrl, M., **Moitinho-Silva, L.**, Gernert, C., Schupp, P., Pawlik, J.R., Lindquist, N.L., Erpenbeck, D., Wörheide, G., Hentschel, U. (2014) The HMA-LMA dichotomy revisited: an electronmicroscopical survey of 56 sponge species. **The Biological Bulletin**. In press.
4. **Moitinho-Silva, L.**, Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., Ravasi, T., Hentschel, U. (2014) Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. **Molecular Ecology**. 23(6), 1348–1363.
5. Harjes, J., Ryu, T., Abdelmohsen, U.R., **Moitinho-Silva, L.**, Horn, H., Ravasi, T., Hentschel, U. (2014) Draft genome sequence of the antitrypanosomally active sponge-associated bacterium *Actinokineospora* sp. strain EG49. **Genome Announcements**. 2(2), e00160–14.
6. Giles, E.C., Kamke, J., **Moitinho-Silva, L.**, Taylor, M.W., Hentschel, U., Ravasi, T., Schmitt, S. (2013) Bacterial community profiles in low microbial abundance sponges. **FEMS Microbiology Ecology**. 83(1), 232–241.

Publications derived from prior research

7. **Moitinho-Silva, L.**, Kondo, M.Y., Oliveira, L.C.G., Okamoto, D.N., Paes, J.A., Machado, M.F.M., Veronez, C.L., Motta, G., Andrade, S.S., Juliano, M.A., Ferreira, H.B., Juliano, L., Gouvea, I.E. (2013) *Mycoplasma hyopneumoniae* in vitro peptidase activities: identification and cleavage of kallikrein-kinin system-like substrates. **Veterinary Microbiology**. 163(3-4), 264–273.
8. **Moitinho-Silva, L.**, Heineck, B.L., Reolon, L.A., Paes, J.A., Klein, C.S., Rebelatto, R., Schrank, I.S., Zaha, A., Ferreira, H.B. (2012) *Mycoplasma hyopneumoniae* type I signal peptidase: expression and evaluation of its diagnostic potential. **Veterinary Microbiology**. 154(3-4), 282–291.
9. Lorena, V.M.B., Verçosa, A.F.A., Machado, R.C.A., **Moitinho-Silva, L.**, Cavalcanti, M.G.A., Silva, E.D., Ferreira, A.G.P., Correa-Oliveira, R., Pereira, V.R.A., Gomes, Y.M. (2008) Cellular immune response from Chagasic patients to CRA or FRA recombinant antigens of *Trypanosoma cruzi*. **Journal of Clinical Laboratory Analysis**. 22(2), 91–98.

10 Curriculum Vitae