Elham Karimi

Metagenomics and functional genomics of bacterial symbionts of *Spongia* (Porifera, Dictyoceratida) specimens from the Algarvian shore (South Portugal)

Faro, January 2018

Elham Karimi

Metagenomics and functional genomics of bacterial symbionts of *Spongia* (Porifera, Dictyoceratida) specimens from the Algarvian shore (South Portugal)

Doctorate in Marine, Earth and Environmental Sciences
(specialization in **Marine Microbial Ecology**)

Supervised by:
Prof. Dr. Rodrigo da Silva Costa
Prof. Dr. Maria Margarida dos Prazeres Reis

Faro, January 2018

“Metagenomics and functional genomics of bacterial symbionts of *Spongia* (Porifera, Dictyoceratida) specimens from the Algarvian shore (South Portugal)”

This work has not previously been submitted for a degree in any university. To the best of my knowledge, the thesis contains no material previously published or written by another person except where due reference is made in the thesis itself.

©2018 Elham Karimi
The University of Algarve has the perpetual right with no geographical boundaries to archive and publicize this work through printed copies in paper or digitized, or in any other format known or that becomes invented, to broadcast through scientific repositories and to allow its copy and distribution with educational purposes for research, noncommercial, given that its author and editor are credited.

Elham Karimi was awarded an Erasmus Mundus Doctoral scholarship (SALAM) (EMA2 lot7/SALA1206422) to perform the work leading to this.

ACKNOWLEDGMENTS

I was awarded full scholarship support for pursuing my PhD by the Erasmus Mundus Programme (SALAM-lot7) and I am sure that completing this work would have been impossible without such a grant. Thanks to all people who have initiated and worked in this association for helping young science lovers to follow their dreams and passions.

I would also like to sincerely thank many people who made this PhD journey such a rewarding and growing experience.

I would like to greatly thank my supervisors, Prof. Dr. Rodrigo Costa for offering me the opportunity to work on this project and all invaluable advises throughout the thesis, and to Prof. Dr. Margarida Reis for accepting me in UAlg.

I am sincerely grateful to Prof. Dr. Ute Henstchel for hosting me for three months at GEOMAR-Germany and showing me different prospects of doing and collaborating in science.

My special thanks to all co-authors and colleagues contributing in this project: Dr. Jorge M.S. Gonçalves, Dr. Joana R. Xavier, Miguel Ramos, Dr. Beate Slaby, Dr. Cymon Cox, André R. Soares, Dr. Tina Keller-Costa, Dr. Ulisses N. da Rocha, and Dr. Jochen Blom.

While I was working in the lab with the group MicroEcoEvo in CCMAR, I had a great chance to meet nice people who hadn't hesitated to help. I am greatly thankful to Telma Franco for her spectacular help, to Marta Valente for providing excellent Sangar sequencing services. I am also thankful to Gianmaria Califano for his help in the lab at the very beginning of my PhD journey.

I gratefully acknowledge the Research funding grant from FEMS for supporting my stay in Germany to perform part of this project. I would like to thank the Research Unit Marine Microbiology at the Division of Marine Ecology, GEOMAR at Kiel-Germany for all warm hosting and great discussion through seminars and lectures. I learned a lot from you all. Also, thanks to people in institute for Biotechnology and Bioengineering (IBB-IST) in Lisbon for all warm smiles when I had meetings with my supervisor there.

Had I not met decent people; Beate, André, Tabea, and Miguel; this journey would not have been valuable at all. You were wonderful people I met on this journey.

Last but most importantly, many thanks to my husband, Alireza, who has been staying beside me from the beginning of this journey with all unconditional help and love, and big thanks to my family; Maman, Baba, Elahe, Ehsan, Neda and little Artin whom I have always got energy of them from a distance, encouraging me to be patient and hardworking.

ABSTRACT

Sponges are early-branched, filter-feeding metazoans that usually harbor complex microbial communities comprised of diverse "uncultivable" symbiotic bacteria. In this thesis, the functional and taxonomic features of the marine sponge microbiome are determined, using *Spongia officinalis* as model host organism. Emphasis is given to adaptive and functional traits of the profuse and biotechnologically-relevant alphaproteobacterial symbionts of sponges. A metagenomics-centred approach was employed to reveal microbial taxa and genomic signatures enriched in the *Spongia officinalis* endosymbiotic consortium, and thus likely to play pivotal roles in holobiont functioning. Further, a comparative genomics study is presented unveiling the common and specific traits of ten *Alphaproteobacteria* genera isolated from *S. officinalis* with alternative symbiont cultivation methodology. Finally, a sequence composition-dependent binning approach is employed to assemble, from metagenomic sequences, the genome of an uncultured alphaproteobacterial symbiont of *S. officinalis* belonging to the family *Rhodospirillaceae*.

High abundance of polyketide and terpene synthase-, eukaryotic-like protein- (ELPs), type IV secretion system-, plasmid- and ABC transporter-encoding genes, among others, characterized the sponge microbial metagenomes. In contrast, motility and chemotaxis genes were abundant in seawater and sediment microbiomes, but nearly absent in the *S. officinalis* symbiotic consortium. Much higher frequencies of anti-viral CRISPR-Cas and restriction-modification systems, along with much lower viral abundances, were observed in the sponge-associated metagenomes than in the environment and interpreted as true hallmarks of this symbiotic consortium.

In line with outcomes retrieved for the whole symbiotic community, alphaproteobacterial symbionts of marine sponges likely contribute the most to host fitness through nutritional exchange, cell detoxification processes and chemical defense, the latter being theoretically promoted by both polyketide and terpenoid biosynthesis. The several alphaproteobacterial cultures retrieved in this thesis, displaying high natural product biosynthesis capacities, can now be explored in studies aiming at revealing novel biological activities and chemical structures from these symbionts.

Keywords

Porifera, metagenomics, functional genomics, *Alphaproteobacteria*, host-microbe interactions, microbiome.

RESUMO

As esponjas marinhas (filo Porifera) são consideradas um dos mais simples grupos entre os metazoários em função de sua falta de organização em tecidos e órgãos verdadeiros. Porém, estes animais relativamente simples em termos de plano corporal normalmente abrigam comunidades muito complexas de microorganismos. Em função de seu surgimento basal na história evolutiva do planeta, o conhecimento a respeito deste "holobionte", isto é, o consórcio de organismos formado pela esponja marinha hospedeira e todos os seus simbiontes microbianos, possui grande relavância ao avanço da nossa compreensão sobre as interações hospedeiro-microorganismos. Nesta tese de doutoramento, tive como objetivo a determinação das características funcionais e taxonómicas do microbioma das esponjas marinhas no contexto de seu ambiente circundante (água e sedimentos marinhos, e suas respecticvas microbiotas), dando ênfase aos traços adaptativos e funcionais de alfaproteobactérias associadas ao organismo modelo *Spongia officinalis* ("bath sponge"). Para tal, uma abordagem independente de cultivo, centrada em técnicas de metagenómica, foi empregada para revelar grupos taxonómicos e genes microbianos abundantes no consórcio de endosimbiontes da esponja *S. officinalis* e que, desta forma, provavelmente possuem papel importante no funcionamento e homeostase do "holobionte" (**Capítulo 2**). Considerando (1) a abundância, diversidade e plasticidade metabólica de alfaproteobactérias marinhas, (2) a prevalência e ampla distribuição geográfica de alfaproteobactérias especificamente associadas a esponjas marinhas, até então não cultivadas em laboratório, e (3) o facto de que foi possível cultivar muitas estirpes filogeneticamente distintas de alfaproteobactérias associadas a *S. officinalis* através do uso de modificações simples a protocolos convencionais de cultivo bacteriano, o **Capítulo 3** apresenta um estudo de genómica comparativa que revela os atributos funcionais comuns e específicos de dez géneros pertencentes à classe *Alphaproteobacteria* isolados de *S. officinalis* com o emprego de metodologia alternativa. Finalmente, o **Capítulo 4** descreve a utilização de uma abordagem bioinformática de "genomic binning", centrada na composição de nucleótidos, para possibilitar a assemblagem, a partir de sequências metagenómicas do microbioma de *S. officinalis* obtidas no **Capítulo 2**, de um genoma de uma alfaproteobactéria não cultivada em laboratório, pertencente à família *Rhodospirillaceae*. Desta forma, esta tese combina abordagens dependentes e independentes de cultivo visando a obtenção de um conhecimento mais amplo acerca das características genómicas verdadeiramente promovidas no consórcio simbiótico das esponjas marinhas, e revela os traços adaptativos, potencial codificador e as

prováveis funções da complexa comunidade de alfaproteobactérias que habita estes hospedeiros.

A anotação de metagenomas microbianos (**Capítulo 2**) revelou que a comunidade simbiótica de *S. officinalis* distinguiu-se fortemente das comunidades microbianas em água e sedimentos marinhos através de uma maior abundância de linhagens bacterianas até então não cultivadas pertencentes aos filos *Proteobacteria*, *Poribacteria*, *Actinobacteria*, *Gemmatimonadetes*, *Chloroflexi* e *Acidobacteria*. No filo *Proteobacteria*, várias unidades taxonómicas operacionais (OTUs, estabelecidas a 97% de semelhança do gene 16S rRNA) pertencentes às famílias *Rhodobacteraceae* e *Rhodospirillaceae* foram prevalentes nos metagenomas microbianos associados à esponja. Uma alta frequência de genes codificadores de policetídeo-sintases, terpeno-sintases, "eukaryotic-like proteins" (ELPs, p.ex. repetições de sequências de anquirina, tetratricopéptidos e elementos WD40), sistemas tipo IV, plasmídeos, transportadores do tipo ABC e luciferases, entre outros, foi registada no metagenoma microbiano das esponjas. Interessantemente, as comunidades microbianas do sedimento abrigaram maior abundância residual destes (e muitos outros) elementos genéticos claramente "enriquecidos" (ou seja, de maior abundância) nas esponjas marinhas em comparação com as comunidades microbianas da água. Porém, abundâncias muito mais altas de genes envolvidos com quimiotaxia e motilidade microbiana foram obtidas para as comunidades de água e sedimento em comparação com o consórico simbiótico em *S. officinalis*, onde tais elementos genéticos estiveram praticamente ausentes. Fez-se notar também a frequência muito mais alta dos elementos anti-virais "CRISPR-Cas" e de restrição-modificação (R-M), acompanhada por uma redução na abundância de vírus, nos metagenomas microbianos da esponja marinha em comparação com o seu ambiente circundante, e tais elementos foram interpretados como "assinaturas genómicas" fulcrais deste consórcio simbiótico. Em sua totalidade, estes resultados sugerem que, embora os microbiomas inspeccionados sejam altamente contrastantes em termos de taxonomia e função, maiores densidades de partículas e células - factores comuns aos microbiomas de sedimentos e esponjas marinhas - podem possuir algum papel na evolução de traços funcionais convergentes entre ambos. Considerou-se ainda que o consórcio simbiótico das esponjas é formado essencialmente por microorganismos sésseis com papel primordial na troca de nutrientes com o hospedeiro (especialmente no que diz respeito aos metabolismos de azoto e enxofre), além de possuírem mecanismos de defesa anti-viral altamente sofisticados em conjunto com uma alta capacidade de troca génica com outros microorganismos e com o organismo hospedeiro. Este últimos factores são dos que mais contribuem ao carácter verdadeiramente único desta comunidade simbiótica.

Em paralelo à análise metagenómica, uma plataforma alternativa ao cultivo de simbiontes bacterianos, calcada no uso de um meio oligotrófico (MG50) e temperaturas de incubação mais baixas, foi aplicada ao microbioma associado a *S. officinalis* e bem-sucedida no isolamento de uma alta diversidade de géneros de *Alphaproteobacteria* como, por exemplo, *Anderseniella*, *Erythrobacter*, *Labrenzia*, *Loktanella*, *Ruegeria*, *Sphingorhabdus*, *Tateyamaria*, *Pseudovibrio* e dois prováveis novos géneros pertencentes à família *Rhodobacteraceae* **(Capítulo 3)**. A comparação global dos genomas de todas as bactérias mencionadas acima, e a detecção de genes comuns a todos os genomas, veio a demonstrar que a comunidade cultivável de alfaproteobactérias associadas a *S. officinalis* pode contribuir ao aumento do "fitness" do hospedeiro através de mecanismos de detoxificação (p. e.x. remoção de metais pesados e rejeitos metabólicos, degradação de compostos aromáticos e halogenados), fornecimento de vitaminas essenciais, troca de nutrientes (especialmente em relação ao processamento de enxofre e azoto orgânicos) e defesa química (em particular através da biossíntese de policetídeos e terpenóides, muito comum entre estes simbiontes). Em função do potencial codificador muitíssimo diversificado dos dez genomas analisados, foi empregada uma comparação genómica feita com base em anotações ao nível de COGs ("Clusters of Orthologous Groups of Proteins") de forma a revelar padrões de convergência e divergência entre os vários genomas. Este procedimento revelou três grupos genómicos, posteriormente divididos em dois maiores grupos funcionais: genomas peretencentes ao clado *Roseobacter* (Grupo I, GI) *versus* genomas não pertencentes a este clado (Grupo II, G2). Concluiu-se que espécies representativas do Grupo II possuem maior probabilidade em estabelecer relações simbióticas mais próximas com as esponjas marinhas, em função de vários elementos genómicos, como genes codificadores de ELPs, proteínas de adesão e pili - usualmente considerados "factores de simbiose" -, que estiveram presentes em maior abundância nos genomas deste Grupo em comparação com o Grupo I. Particularmente, o género *Anderseniella* apresentou o repertório de traços genotípicos mais claramente associado a uma estratégia de vida simbiótica. Ainda assim, todos os organismos estudados não apresentaram sinais de redução genómica, usualmente considerados indicadores de um modo de vida simbiótico, e de facto possuem um metabolismo de utilização de fontes de carbono altamente versátil, o que sugere que a estratégia adaptativa destas bactérias é bifásica e portanto inclui ambos os estágios livre e de associação ao hospedeiro.

O **Capítulo 4** apresenta um estudo em que sequências metagenómicas do microbioma de *S. officinalis* (**Capítulo 2**) foram utilizadas para reconstruir, através de plataformas sofisticadas de bioinformática, o genoma de um simbionte dominante neste hospedeiro, porém

não cultivado, pertencente à família *Rhodospirillaceae*. Em conjunção com a descrição do genoma obtido, uma abordagem de genómica comparativa foi implementada para determinar os elementos genéticos envolvidos com a adaptação de linhagens pertencentes a esta família a um modo de vida preponderantemente simbiótico. Isto foi alcançado através da determinação de genes significativamente mais/menos abundantes em grupos de genomas de *Rhodospirillaceae* representativos de um modo de vida simbiótico *versus* um modo de vida livre de associação a um hospedeiro ("free-living"). Embora ambos os grupos "simbiótico" e "livre" tenham partilhado muitos genes em comum e demonstrado metabolismo de aquisição e utilização de nutrientes (carbono, azoto, fósforo e enxofre) altamente versátil, os genomas "simbióticos" foram claramente caracterizados pela ausência de genes envolvidos em motilidade e quimiotaxia - de acordo com o observado no microbioma simbiótico "total" (**Capítulo 2**) -, por alta frequência de genes envolvidos com a utilização de enxofre orgânico, e por uma composição distinta de genes codificadores de trasnportadores ABC, metabolitos secundários, sistemas de detoxificação celular e regulação de estresse oxidativo. Em congruência com resultados obtidos para a comunidade simbiótica total (**Capítulo 2**) e muitos dos simbiontes cultivados (**Capítulo 3**), os simbiontes de esponja pertencentes à família *Rhodospirillaceae* provavelmente contribuem mais significativamente ao metabolismo hospedeiro através da troca de nutrientes, processos de detoxificação celular e defesa química, esta última teoricamente promovida pela biossíntese de policetídeos e terpenos. A obtenção, nesta tese, de diversas culturas de alfaproteobactérias potencialmente produtoras destes metaboitos, para além de outros produtos naturais vários, poderá ser explorada em estudos futuros com o objectivo de revelar novas actividades biológicas e estruturas químicas a partir destes simbiontes.

Para contextualizar o trabalho de investigação aqui realizado, e resumir os principais resultados obtidos e suas implicações aos campos da microbiologia de esponjas, interacção hospedeiro-microorganismos e biologia marinha, dois capítulos adicionais foram elaborados (Introducção e Discussão Geral, **Capítulo 1** e **Capítulo 5**, respectivamente) e incorporados à estrutura da tese.

Termos chave

Porifera, metagenómica, genómica funcional, *Alphaproteobacteria*, interacções hospedeiro-microorganismo, microbioma.

TABLE OF CONTENTS

ACKNOWLEDGMENTS I

ABSTRACT III

RESUMO V

TABLE OF CONTENTS IX

LIST OF FIGURES XIII

LIST OF TABLES XV

LIST OF ABBREVIATIONS XVI

CHAPTER 1 GENERAL INTRODUCTION 1

- PREFACE 3
- MARINE SPONGE BIOLOGY AND ECOLOGY 4
- MARINE SPONGES AND THEIR MICROBIAL COMMUNITIES 5
- THE MICROBIOME OF DICTYOCERATIDA SPONGES FROM THE ALGARVE COAST 9
- METAGENOMICS AS A TOOL TO UNDERSTAND MICROBIAL COMMUNITY STRUCTURE AND FUNCTION 12
- RECONSTRUCTION OF GENOMES FROM METAGENOMES 13
- CULTIVATION BIAS IN NATURAL ENVIRONMENTS AND STRATEGIES TO CULTIVATE THE UNCULTIVABLE 15
- GENOME SEQUENCING AND COMPARATIVE GENOMICS TO IDENTIFY THE FUNCTIONAL FEATURES OF SYMBIONT GENOMES 17
- AIMS AND SPECIFIC RESEARCH QUESTIONS 19
  - *Cultivation-independent approaches to extract functional information from yet uncultivated bacterial symbionts* *20*
  - *Alternative isolation and cultivation of sponge-associated bacteria* *21*
- THESIS OUTLINE 23

CHAPTER 2 COMPARATIVE METAGENOMICS REVEALS THE DISTINCTIVE ADAPTIVE FEATURES OF THE *SPONGIA OFFICINALIS* ENDOSYMBIOTIC CONSORTIUM 25

- ABSTRACT 29
- INTRODUCTION 30
- MATERIALS AND METHODS 32
  - *Sampling and sponge identification* *32*
  - *Microbial metagenomic DNA extraction and Next-Generation Sequencing (NGS)* *32*
  - *Metagenome data processing* *33*
  - *Metagenome data analysis* *34*
  - *Alternative analytical pipelines and data validation* *35*
  - *Nucleotide sequence accession numbers* *36*
- RESULTS 36
  - *Sponge identification* *36*
  - *Microbial metagenomes - dataset overview* *37*

Functional and taxonomic ordination.......... 37
16S rRNA gene taxonomic profiling.......... 39
IPR Functional profiling.......... 41
Functional conservation among Spongia officinalis *and other sponge hosts*.......... 43
All domains-all genes taxonomic profiling using MG-RAST.......... 44
DISCUSSION.......... 45
SUPPLEMENTARY MATERIAL.......... 53
ACKNOWLEDGMENTS.......... 53
FUNDING.......... 53
CONFLICT OF INTEREST.......... 53
ETHICS STATEMENT.......... 53
AUTHOR CONTRIBUTIONS.......... 53

CHAPTER 3 FUNCTIONAL GENOMICS OF CULTIVATED SPONGE-ASSOCIATED *ALPHAPROTEOBACTERIA* REVEALS SHARED AND UNIQUE TRAITS UNDERLYING A BIMODAL SYMBIOTIC-FREE-LIVING LIFE-STYLE.......... 55
ABSTRACT.......... 59
INTRODUCTION.......... 60
MATERIAL AND METHODS.......... 62
Sample collection, cultivation of bacteria and phylogenetic analysis.......... 62
Genome sequencing of sponge-associated Alphaproteobacteria.......... 64
Annotation and comparative analysis of genomes.......... 64
Representativeness of cultivated sponge-associated Alphaproteobacteria *across marine biotopes*.......... 65
Nucleotide sequence accession numbers.......... 65
RESULTS.......... 66
Isolation and identification of S. officinalis *associated bacteria*.......... 66
General features of sponge-associated Aphaproteobacteria *genomes*.......... 69
Core- and pan-genome analysis.......... 69
Functional and comparative genomics based on Clusters of Orthologous Groups of Proteins (COGs).......... 71
Shared features of all genomes.......... 72
Carbon, nitrogen, sulfur and phosphorus metabolism.......... 72
Cofactors, vitamins and inorganic ions.......... 72
Defense, antibiotic resistance and reactive oxygen species (ROS) protection.......... 73
Eukaryotic-like proteins (ELPs) encoding genes.......... 73
Roseobacter *versus non-*Roseobacter *genome features*.......... 76
Secondary metabolism.......... 78
Representation of the cultivated Alphaproteobacteria *genomes in marine metagenomes*.......... 79
Embedded description of the Anderseniella *sp. Alg231-50 genome*.......... 79
DISCUSSION.......... 81
Communal metabolite features of cultivable sponge-associated Alphaproteobacteria.......... 82
Antimicrobial agents.......... 83
Representativeness of cultured Alphaproteobacteri*a in the marine sponge microbiome*.......... 84
Genome features of Anderseniell*a sp. Alg231-50*.......... 86
SUPPLEMENTAL MATERIAL.......... 87
ACKNOWLEDGMENTS.......... 87
FUNDING.......... 87
CONFLICT OF INTEREST.......... 87
ETHICS STATEMENT.......... 87
AUTHOR CONTRIBUTIONS.......... 87

CHAPTER 4 METAGENOMIC BINNING REVEALS VERSATILE NUTRIENT CYCLING AND DISTINCT ADAPTIVE FEATURES IN ALPHAPROTEOBACTERIAL SYMBIONTS OF MARINE SPONGES **89**
ABSTRACT 93
INTRODUCTION 94
MATERIALS AND METHODS 96
Sample collection, DNA extraction and sequencing *96*
Metagenome assembly *96*
Genome binning from metagenomes *96*
Genome annotation for comparative genomics *97*
Phylogenomics of sponge-associated Alphaproteobacteria *98*
Geographic distribution of uncultivated, sponge-associated Rhodospirillales *98*
Genome features and life strategy of Rhodospirillaceae *sponge symbionts* *99*
Nucleotide sequence accession numbers *100*
RESULTS 100
Phylogenomics of sponge-associated Alphaproteobacteri*a and relatives* *100*
Core- and pan-genomes of sponge-associated and free-living Rhodospirillaceae *102*
Functional genomics of sponge-associated and free-living Rhodospirillaceae *105*
Adaptations to (sponge) symbiotic and free-living life-styles *109*
DISCUSSION 113
SUPPLEMENTARY DATA 118
ACKNOWLEDGEMENTS 118
FUNDING 118
ETHICAL STATEMENT 118
AUTHOR CONTRIBUTIONS 118

CHAPTER 5 GENERAL DISCUSSION AND OUTLOOK **119**
GENERAL DISCUSSION 121
Comparative metagenomics reveals unique life strategies of the Spongia officinalis *endosymbiotic consortium (Chapter 2)* *123*
The Alphaproteobacteria *symbiotic community in* S. officinalis - *a comparative and functional genomics approach* *125*
Genomes from cultivated bacteria 125
Genome reconstruction of a sponge-specific *Rhodospirillaceae* symbiont 127
CONCLUSIONS 129
FUTURE PERSPECTIVES 132

APPENDICES **135**
APPENDIX I CHAPTER 2 COMPLEMENTARY ANALYSIS AND SUPPLEMENTARY MATERIALS 137
File S1 Preliminary data analyses and data validation with alternative analytical pipelines . *139*
Supplementary Figures and Tables 144
APPENDIX II CHAPTER 3 SUPPLEMENTARY MATERIALS 147
Supplementary Figures 149
Supplementary Tables 151
APPENDIX III CHAPTER 4 COMPLEMENTARY ANALYSIS AND SUPPLEMENTARY MATERIALS 153
File S1 Data analysis 155
Supplementary Methods 155
Supplementary Results 156
Supplementary Figures and Tables 157

BIBLIOGRAPHY **161**

LIST OF FIGURES

Figure 1-1. (**A**) Schematic overview of an asconoid sponge body plan. (**B**) Scan electron microscope (SEM) picture of microbes near the choanocyte chamber. (**C**) fluorescent in situ hybridization of different microbial groups in the sponge mesohyl (Poribacteria, yellow cells; *Nitrospira*, pink cells; *Chloroflexi*, cyan cells; *Deltaproteobacteria*, light (*Deltaproteobacteria*, light green cells; *Gammaproteobacteria*, red cells; *Archaea*, blue cells). Source: Webster and Thomas (2016). 11

Figure 1-2. An Alphaproteobacterium isolated from Spongia (growing on R2A medium) – Cultivation and Photo by E. Karimi. 17

Figure 1-3. Photographs of *Spongia officinalis* in vivo **(A)** and in vitro (**B**). (A) Photograph courtesy of Dr. Jorge Goncalves's team (B) Photograph by E. Karim. 20

Figure 1-4. *Spongia officinalis* sampling site at the Algarve coast, with the exact sampling location marked in red (Copyright © 2017Worksheetworks.com and © Google map). 22

Figure 2-1. Principal Coordinate Analysis (PCoA) of taxonomic (**A**) and functional (**B**) microbial community profiles across biotopes. Community ordinations were based on pairwise Bray-Curtis dissimilarities (Table 2-1) calculated from normalized data, considering oscillations of relative OTU and IPR abundances among samples. Analyses were performed on OTU and IPR community profiles extracted from the 10 metagenomes using the EBI metagenomics (EMG) pipeline. Values on axes denote the extent of variation explained by each principal coordinate, whereas the total variation explained in the ordination space is indicated in the inlet. Significance values result from permutational analysis of variance (PERMANOVA) applied to the corresponding dissimilarity matrices. 38

Figure 2-2. Heat map of the most differentiating microbial phyla across biotopes based on OTU data. Shown are the 17 phyla whose (OTU) relative abundances were found to oscillate the most among biotopes, explaining > 85% of the variation in phylum distributions. The dendrogram clusters phylum entries according to their abundance distributions across biotopes, labeled at the bottom of the diagram. Red squares show higher relative abundance values than the mean, whereas grey squares show lower relative abundance values than the mean. Within each phylum, color intensities are determined as a linear function of the Z-score calculated for each phylum abundance value as the subtraction of that value by the mean divided by the standard deviation around that mean (Z=(x-mean)/sd). SP230-SP233, sponge microbial metagenomes; Sd, sediment metagenomes; Sw, seawater metagenomes. 40

Figure 2-3. Heat map of the most differentiating OTUs across biotopes. Shown are the 31OTUs (97% cut-off) found to oscillate the most among biotopes, explaining > 32 % of the variation in the OTU dataset. Heat map details are as in legend to Figure 2-2. 41

Figure 2-4. Heat map of the 44 most differentiating IPR entries across biotopes. The dendrogram clusters IPR entries according to their abundance distributions across biotopes, labeled at the bottom of the diagram. Heat map details are as in legend to Figure 2-2. 42

Figure 2-5. Abundance distributions of broad functional categories across biotopes. Values on the y-axis represent mean cumulative IPR relative abundances (%) in each biotope ± standard deviations. ABC transporters - 19 IPR entries used in plot construction; Plasmids - 10 IPR entries; Polyketide synthases - 1 IPR entry; Type IV secretion - 6 IPR entries; CRISPR-Cas - 43 IPR entries, Motility - 8 IPR entries involved in gliding and fimbriae-based motility; Flagellum, 56 IPR entries involved in flagellum assembly and motility; chemotaxis - 5 IPR entries; Type II secretion proteins -13 IPR entries; Terpene/Terpenoid biosynthesis - 3 IPR entries; Restriction endonucleases – 68 IPR entries; Sulfatases - 4 IPR entries. All IPR entries can be identified in Appendix I-Table S4. Results of the general test for differences among biotopes are shown at the top of each chart, below the label of each analyzed function. One-Way ANOVA with F statistics results are shown for normally distributed data, whereas ANOVA on Ranks results are shown for data distributions that did not pass normality tests. Bars labeled with different letters

REPRESENT STATISTICALLY DISTINCT BIOTOPES IN TERMS OF IPR RELATIVE ABUNDANCES ACCORDING TO POST-HOC PAIR-WISE TESTS OF SIGNIFICANCE. 45

FIGURE 3-1. MAXIMUM LIKELIHOOD TREE OF ALPHAPROTEOBACTERIA SPECIES BASED ON KIMURA 2-PARAMETER EVOLUTIONARY DISTANCES CALCULATED FOR 16S RRNA GENE SEQUENCES. ALPHAPROTEOBACTERIA STRAINS ISOLATED FROM S. OFFICINALIS ARE SHOWN IN BOLD. NUMBERS OF ISOLATES OBTAINED FROM S. OFFICINALIS THAT BELONG TO THE SAME OTU (100% CUT-OFF) ARE GIVEN IN BRACKETS. CLOSEST NCBI BLASTN HITS AND TYPE STRAINS ((T), IN BOLD) FOR EACH STRAIN ARE SHOWN ON THE TREE. BLUE MARKS SPONGE-ASSOCIATED, ORANGE MARKS INVERTEBRATE ASSOCIATED AND GREEN MARKS MARINE ALGAE-ASSOCIATED CLOSEST NCBI BLASTN HITS AND TYPE STRAINS. STRAINS THAT HAD THEIR GENOME SEQUENCED ARE MARKED WITH AN ASTERISK. BOOTSTRAP VALUES (100 REPETITIONS) ABOVE 70% (0.7) ARE SHOWN ON THE TREE NODES. 68

FIGURE 3-2. NUMBER OF STRAIN-SPECIFIC ("SINGLETON") GENES IN EACH *ALPHAPROTEOBACTERIA* GENOME ANALYZED IN THIS STUDY. 71

FIGURE 3-3. HEATMAP ON THE ABSOLUTE COUNTS OF CODING SEQUENCES CLASSIFIED AS COGS REPRESENTING EUKARYOTIC-LIKE PROTEINS (ELPS) IN THE 10 CULTIVATED *ALPHAPROTEOBACTERIA* GENOMES ANALYZED IN THIS STUDY. THE NUMBERS IN EACH CELL SHOW THE COUNT OF GENES FOR EACH GENOME. THE EMPTY CELLS (DARK BLUE) REPRESENT NO COUNTS (EQUALS ZERO). * COG5424 (PYRROLSO-QUINOLINE QUINONE REPEATS), COG4886 (LEUCINE-RICH REPEATS(LRR)), COG1520 AND COG2319 (WD40 REPEATS), COG0666 (ANKYRIN REPEATS), AND COG0457AND COG0790 (TETRATRICOPEPTIDE REPEATS). 76

FIGURE 3-4. PRINCIPAL COMPONENTS ANALYSIS (PCA) ORDINATION OF THE TEN ALPHAPROTEOBACTERIAL GENOMES ANALYZED IN THIS STUDY BASED ON THEIR FUNCTIONAL PROFILES (I.E. PRESENCE/ABSENCE AND RELATIVE ABUNDANCES OF COG ENTRIES PER GENOME). STRAINS GROUPED BY ELLIPSES HAVE THEIR TAXONOMIC AFFILIATIONS DISCLOSED FOLLOWING THE RESULTS SHOWN IN FIGURE 3-1. NOTE THE CLOSER FUNCTIONAL SIMILARITY BETWEEN MEMBERS OF THE "STAPPIA GROUP" (PSEUDOVIBRIO AND LABRENZIA, FORMALLY BELONGING TO THE FAMILY RHODOBACTERACEAE IN THE ORDER RHODOBACTERALES) TO THE GENUS ANDERSENIELLA (ORDER RHIZOBIALES) THAN TO OTHER GENERA OF THE RHODOBACTERACEAE FAMILY. 77

FIGURE 4-1. PHYLOGENOMIC TREE OF SPONGE-ASSOCIATED ALPHAPROTEOBACTERIA AND CLOSE RELATIVES. THE TREE WAS GENERATED USING PHYLIP WITHIN THE EDGAR ENVIRONMENT USING THE NEIGHBOR JOINING METHOD ON A MATRIX OF KIMURA DISTANCES BETWEEN PREDICTED AMINO ACID SEQUENCES OF PROTEIN-ENCODING GENES IN THE GENOME SEQUENCES. IT CONSISTS OF 35 GENOME ENTRIES (28 FROM ALPHAPROTEOBACTERIA SPECIES). SIXTY GENES COMMON TO ALL GENOMES WERE USED IN TREE CONSTRUCTION. GENOMES IN GREEN WERE ASSEMBLED FROM SPONGE, SEAWATER OR SEDIMENT METAGENOMES, WHEREAS GENOMES IN BLUE REPRESENT ALPHAPROTEOBACTERIAL CULTURES OBTAINED FROM S. OFFICINALIS (CHAPTER 3). GENOMES IN BLACK WERE OBTAINED FROM PUBLIC DATABASES. ENTRIES MARKED IN BOLD CORRESPOND TO GENOME BINS GENERATED IN THIS STUDY. BOOTSTRAP VALUES ABOVE 70% (0.7) ARE SHOWN ON TREE NODES. 102

FIGURE 4-2. PHYLOGENOMIC TREE OF GENOMIC BINS RETRIEVED IN THIS STUDY AND THEIR CLOSEST RELATIVES. THE TREE WAS GENERATED USING PHYLIP WITHIN THE EDGAR ENVIRONMENT USING THE NEIGHBOR JOINING METHOD ON A MATRIX OF KIMURA DISTANCES BETWEEN PREDICTED AMINO ACID SEQUENCES OF PROTEIN-ENCODING GENES IN THE GENOME SEQUENCES. IT CONSISTS OF 14 GENOME ENTRIES (NINE FROM ALPHAPROTEOBACTERIA SPECIES), AND 128 GENES COMMON TO ALL GENOMES WERE USED IN TREE CONSTRUCTION. ENTRIES MARKED IN BOLD CORRESPOND TO GENOME BINS GENERATED IN THIS STUDY. BOOTSTRAP VALUES ABOVE 70% (0.7) ARE SHOWN ON TREE NODES. 103

FIGURE 4-3. GENES SHARED BY AND SPECIFIC TO SYMBIOTIC AND FREE-LIVING RHODOSPIRILLACEAE GENOMES. VENN DIAGRAMS COMPARING THE GENE INVENTORIES OF THREE RHODOSPIRILLACEAE GENOMES RECONSTRUCTED FROM SPONGES ALONG WITH THEIR CLOSEST FREE-LIVING (**A**) AND FURTHER SPONGE-ASSOCIATED (**B**) RELATIVES ARE SHOWN. DIAGRAMS WERE COMPUTED ON EDGAR BASED ON RECIPROCAL BEST BLAST HITS OF THE CODING SEQUENCES PREDICTED BY RAST. FULL NAMES OF SYMBIOTIC AND FREE-LIVING STRAINS ARE AS IN THE FOOTNOTE TO TABLE 4-2. 104

FIGURE 4-4. HEAT MAP OF ABC-TYPE TRANSPORT SYSTEMS SIGNIFICANTLY ENRICHED IN SPONGE SYMBIONT GENOMES. THE HEAT MAP ILLUSTRATES SHIFTS IN THE ABSOLUTE ABUNDANCE OF CDSS ASSIGNED TO EACH COG CATEGORY LISTED IN EACH OF THE SIX GENOMES (THREE SPONGE-ASSOCIATED, THREE FREE-LIVING; SEE TABLE 4-2) USED IN THE COMPARATIVE ANALYSIS. FULL NAMES OF SYMBIOTIC AND FREE-LIVING STRAINS ARE AS IN THE FOOTNOTE TO TABLE 4-2. 109

Figure 4-5. Heat map of COGs in the class "secondary metabolites biosynthesis, transport, and catabolism" found to be significantly enriched in sponge symbiont genomes. The heat map illustrates shifts in the absolute abundance of CDSs assigned to each COG category listed in each of the six genomes (three sponge-associated, three free-living; see Table 4-2) used in the comparative analysis. Full names of symbiotic and free-living strains are as in the footnote to Table 4-2. 110

LIST OF TABLES

Table 2-1. Functional (IPR) and taxonomic (OTU) community dissimilarities calculated between- and within-biotope samples. Shown are average Bray-Curtis dissimilarity values ± standard deviations. Within each row, values tagged with different letters are significantly different ($p < 0.05$) according to One-Way ANOVA, except for the OTU-based comparison between biotopes where non-parametric ANOVA on Ranks was used. 38

Table 3-1. Basic genome features of sponge-associated *Alphaproteobacteria* cultivated in this study. 70

Table 3-2. Restriction-Modification systems identified in the genomes of sponge-associated *Alphaproteobacteria* analyzed in this study. 74

Table 3-3. Distribution of COG entries involved in polyketide biosynthesis across the genomes of sponge-associated *Alphaproteobacteria* analyzed in this study. 75

Table 3-4. Percent alignment of total metagenomic reads from *S. officinalis*, seawater and sediments with the genomes assembled in this study 80

Table 4-1. Basic features of sponge-associated and seawater-derived Alphaproteobacteria genomes assembled in this study. 101

Table 4-2. Genome features of sponge-associated and free-living Rhodospirillaceae. 107

Table 4-3. Nutrient cycling features in sponge-associated and free-living *Rhodospirillaceae* genomes. 108

Table 4-4. COG-based abundance distribution of phage defense mechanisms and eukaryotic-like protein repeats across symbiotic and free-living genomes. 112

LIST OF ABBREVIATIONS

ASW:	Artificial Seawater
ANOVA:	One Way Analysis of Variance
ARPs:	Ankyrin-repeat proteins
BLAST:	Basic Local Alignment Search Tool
CFU:	Colony Forming Unit
CMFASW:	Calcium Magnesium-Free Artificial Seawater
CO1:	Mitochondrial cytochrome oxidase subunit 1
COG:	Clusters of Orthologous Groups
CRISPR:	Clustered Regularly Interspaced Short Palindromic Repeats
DMSO:	Dimethyl sulfoxide
DNA:	Deoxyribonucleic acid
dNTPs:	Deoxynucleoside triphosphates
EDGAR:	Efficient Database framework for comparative Genome Analyses
ELPs:	Eukaryotic-like proteins
EMBL:	European Molecular Biology Laboratory
EMG:	EBI Metagenomics
GYRA:	Computational Cluster Facility in CCMar
HGT:	Horizontal gene transfer
HMA:	High Microbial Abundance
InterPro:	Database of protein families, domains and functional sites
IPR:	InterPro
LMA:	Low Microbial Abundance
LRRs:	Leucine-rich repeat proteins
M5NR:	Non-redundant database
MEGA:	Molecular Evolutionary Genetics Analysis
MG–RAST:	Metagenomic Rapid Annotations using Subsystems Technology
NCBI:	National Center for Biotechnology Information
NGS:	Next–generation sequencing
ORF:	Open reading frame
OTU:	Operational Taxonomic Unit
PCR:	Polymerase Chain Reaction
PKS:	Polyketide synthase
QUAST:	Quality Assessment Tool for Genome Assemblies
R-M:	Restriction modification system
RAST:	Rapid Annotations using Subsystems Technology
RDP:	Ribosomal Database Project
ROS:	Reactive oxygen species
rRNA:	Ribosomal ribonucleic acid
SC:	Sponge-specific bacterial clusters
SCG	Single cell genomics
SIMPER:	Similarity Percentage
T2SS:	Type II secretion systems
TC-DNA:	Total community DNA
TPRs	Tetratricopeptide repeat proteins
UPGMA:	Unweighted Pair Group Method with Arithmetic Mean

Chapter 1

General Introduction

Preface

Life began in the oceans approximately 3.1 to 3.4 billion years ago, based on estimations from microfossils of sulfur-metabolizing bacteria (Wacey et al., 2011). Living organisms are classified into three domains: (i) Eukaryotes and the Prokaryotes which include the domains (ii) Archaea and (iii) Bacteria. However, compelling evidence now exists for the classification of life into only two major domains, the Bacteria and the Archaea, with eukaryotic cells emerging from the latter (Cox et al., 2008; Williams et al., 2013). Sponges (Porifera) are early-derived metazoans whose origin dates back to the Precambrian era (Li et al., 1998). They are simple, benthic invertebrates inhabiting a variety of freshwater and marine habitats, from springs to lakes and from the intertidal zones to the deep seas (Manconi and Pronzato, 2008; Murillo et al., 2016). Also, their biomass regularly surpasses that of reef-building coral species (Diaz and Rützler, 2001). Nowadays, sponges remain prominent organisms in marine benthic communities and their success may result from a highly plastic physiology that entails both body shape acclimatization and symbiotic community resilience/adaptation to changing circumstances (Hentschel et al., 2012). In addition, sponges present a vast and dynamic repository of genetic variability fundamental to their persistence in the oceans throughout evolutionary history. Above all, sponge-associated microbial communities are incredibly diverse and complex. This complexity is similar to that of other diverse symbiotic systems like the human gut microbiome (Arumugam et al., 2011), the rhizosphere microbiome (Berendsen et al., 2012), and the coral microbiome (Ainsworth et al., 2010), compared to other less complex host-microbe associations such as the famous squid- *Vibrio fischeri* symbiosis (Nyholm and McFall-Ngai, 2004) or that of free-living amoebae with *Chlamydia* (Horn and Wagner, 2004).

This chapter is an introduction to the biology and ecology of marine sponges and their remarkable association with a vast diversity of microorganisms. It delineates the most important microbiology and molecular biology techniques applied in this thesis, with emphasis on alternative cultivation strategies to capture novel, as-yet uncultivated symbiont bacteria, and on the use of metagenomics and genomic approaches for a more comprehensive understanding of sponge microbial community functioning. By the end of this chapter, the specific aims and research questions addressed in this thesis are explained, and a brief description of the studied host organism and sampling site is provided.

Marine sponge biology and ecology

Sponges (Phylum Porifera) are sessile, filter feeding organisms that mostly inhabit marine environments. Although they mainly feed by filtering minuscule particles from the water, they may uptake dissolved organic matter as well (de Goeij et al., 2008). Their body contains three matrices: the pinacoderm, the choanoderm and, between them, the mesohyl. The flagellate choanocyte cells create a water flow from the ostia pores (where water is drawn in) through the aquiferous system that spans the sponge body up to a larger exhaling osculum opening (**Figure 1-1A**). However, some carnivorous sponges adapted to the deep seas, classified in the family Cladorhizidae, lack an aquiferous system and instead use their sticky surface to capture small animals (Vacelet, 1995; Hestetun, 2016). The mesohyl contains different cell types, organic collagenous fibers (spongin) and/or an inorganic skeleton of silicon dioxide or calcium carbonate (spicules) (Van Soest et al., 2012). Sponges may assume many different shapes and forms, including erect, encrusting, eroding, cup, and fan- and tree-like, and these are believed to reflect adaptive strategies to different substrate types, depths, light exposure, nutrient availability and hydrodynamics (Huang et al., 2011), among other factors. Sponges are classified into four classes, namely Calcarea (calcareous sponges), Hexactinellida (glass sponges), Demospongiae and Homoscleromorpha (Gazave et al., 2012). Although they have simple body shapes, sponges comprise a highly diversified phylum with approximately 8,500 validly described species to date, with many more yet to be described (Hooper and Van Soest, 2002; Van Soest et al., 2012). Demospongiae is the largest and most diverse class of poriferans. It encompasses about 85% of all extant sponge species (Maldonado, 2009) including three subclasses named Verongimorpha, Keratosa and Hetreroscleromorpha (Morrow and Cárdenas, 2015). The name Demospongiae stems from the Greek *demos* "*people"* and *spongiá "sponge"* which means "the common sponge". The family Spongiidae belongs to the subclass Keratosa which includes the commercial sponges of the genera *Spongia, Hippospongia, Coscinoderma* and *Rhopaloeides*. Singularly within the Porifera, sponges in the subclass Keratosa possess a matrix of protein fibers, instead of spicules, as their primary structural constituents (Becerro, 2012). The Spongiidae family (order Dictyoceratida) consists of six valid genera, namely *Spongia, Hippospongia*, *Coscinoderma, Hyattella, Leiosella, and Rhopaloeides.* The genus *Spongia*, the model sponge host studied this thesis, includes three subgenera called *Spongia*, *Australospongia*, and *Heterofibria*. All six genera of the Spongiidae family vary in form and shape, from encrusting to upright. Most characteristic of Spongiidae is the dense, secondary fiber reticulum that dominates the skeleton. The surface may be heavily armored with an

organized dermal crust of sand, foreign spicules and detritus. The choanocyte chambers of Spongiidae are diplodal (narrow canals for taking into and out), and spherical to oval. In some species, mesohyl and ectosome are supported by heavy deposits of collagen, though this can vary, even between species within the same genus (de Cook and Bergquist, 2002).

Overall, sponge reproduction is quite versatile and known for its sexual and asexual propagation. Asexual reproduction, including fragmentation, budding and gemmulation, can be found in most sponge classes (Ereskovsky and Tokina, 2007). Although sponges lack true reproductive organs, they can reproduce sexually either via hermaphroditism (both sexes in one individual) or gonochorism (each individual organism representing only one sexual state, either male or female). They can either be external fertilizers (oviparous, for example Demospongiae) or internal fertilizers with larvae developing inside the sponge body (viviparous, for example Hexactinellida). Choanocytes can produce sperm which fertilizes the eggs produced by archeocytes. This great variability in reproductive strategies may assist sponges to optimize their population persistence in many different, even hostile environments.

Sponges fulfill different ecological functions in marine ecosystems. Their crucial role in cleaning the water column due to the filtration of dissolved and particulate organic matter is noticeable (Yahel et al., 2007). Moreover, sponges are believed to sustain coral reefs and boost their functioning by taking up dissolved organic matter (DOM) from oligotrophic waters and transforming it in particulate organic matter (POM) promptly consumed by reef organisms (de Goeij et al., 2013). Besides, sponges have been used for different purposes since ancient times, for instance for cleaning and bathing purposes, to this date as possible sources of metabolites applicable in pharmacology and biotechnology (Voultsiadou, 2007; Voultsiadou et al., 2008; Kayal, 2012). Indeed, some bioactive compounds from marine sponges have already passed clinical trials (e.g. Eribulin Mesylate in phase III and Hemiasterlin in phase I (Mayer et al., 2010)) and their antibacterial, antiviral, and antitumoral activities have been extensively documented in recent years (Sipkema et al., 2005; Laport et al., 2009; Anjum et al., 2016).

Marine sponges and their microbial communities

Most symbiont communities are composed of phylogenetically diverse microorganisms and one host individual. The structure and composition of such a community is regulated by different factors, for instance, nutrients provided by the host, chemo-physical characteristics (e.g. pH) of the environment and host individual specific attributes (e.g. power of its immune

system) (Thomas et al., 2016). Nevertheless, the reasons why microbial community compositions differ between different host animals remain largely unidentified (Hacquard et al., 2015). Apart from the evolutionary and ecological importance of sponges, they have received considerable attention due to the distinct microbial communities living with/within them (Hentschel et al., 2012). Sponges harbor bacterial, archaeal and micro-eukaryote phyla. Frequently, bacterial phyla are the paramount groups and can comprise up to 35% of a sponge's wet weight (**Figure 1-1**) (Hentschel et al., 2003; Hill, 2004; Hentschel et al., 2006). To date, up to 52 bacterial phyla (including candidate phyla) were detected in Porifera (Thomas et al., 2016; Slaby et al., 2017), with *Proteobacteria* (mostly *Alpha*- and *Gamma*-), *Actinobacteria, Chloroflexi, Cyanobacteria, Nitrospirae, Crenarchaeota and Poribacteria* being the most dominant ones (Webster and Thomas, 2016). The candidate phylum *Poribacteria* is singularly enriched in sponges (Fieseler et al., 2004) and a versatile carbon utilization metabolism has been suggested for this group based on genomic features uncovered by single cell genomics (Kamke et al., 2013). Among *Proteobacteria*, the class *Alphaproteobacteria* - the major symbiotic taxon addressed in this thesis - represents one of the most diverse and dominant sponge-associated bacterial groups (Enticknap et al., 2006; Schmitt et al., 2007; Simister et al., 2012; Hardoim et al., 2014), with *Rhodobacteraceae* and *Rhodospirillaceae* clades being abundant members of the microbial communities inhabiting Dictyoceratida sponges (Hardoim et al., 2012; Hardoim et al., 2014). *Alphaproteobacteria* species have been detected in both adult and larval samples of marine sponges (Schmitt et al., 2007), suggesting a pattern of vertical symbiont transmission through successive host generations, and thus perhaps a true symbiotic life style for this class in sponges. Besides, a study that compared necrotic sponges with healthy ones pointed out a role of the genus *Pseudovibrio* of the *Alphaproteobacteria* class in host health. *Pseudovibrio* species could not be detected in necrotic sponges while being integral component of healthy specimens (Sweet et al., 2015). Moreover, sponge-associated *Pseudovibrio* spp. have been often reported to display antimicrobial activities *in vitro* (Penesyan et al., 2011; Bondarev et al., 2013; Crowley et al., 2014), suggestive of a possible participation of this genus in host defense. In addition, a previous genome-wide study revealed the potential of a sponge-derived *Pseudovibrio* strain to import and oxidize a wide range of organic and inorganic compounds which can provide carbon, nitrogen, phosphorous, energy, secondary metabolic products and cofactors for the host (Bondarev et al., 2013). In spite of the potential capacities of cultivable sponge-associated bacteria, care should be taken with the interpretation of their actual contribution to host fitness. This is because cultivable sponge symbionts usually correspond to low abundant members of the sponge symbiotic consortium,

whereas the very dominant symbionts have been so far strikingly recalcitrant to cultivation (see e.g. Hardoim et al. (2014)). This aspect is addressed with scrutiny in this thesis by comparing the relative abundances of cultivated and uncultivated alphaproteobacterial symbionts of *Spongia officinalis* using genome-metagenome gene mapping as a proxy for the prevalence of single strains in a complex microbial community.

A meta-analysis of sponge-derived 16S rRNA gene amplicons revealed inter-specific dissimilarities in prokaryote community compositions, with the complexity of sponge symbiont communities ranging from 50 to 3320 operational taxonomic units (OTUs) per host individual. Low intraspecific variabilities in microbial community composition have been interpreted as strong interactions between the host species and its symbionts; while greater variabilities suggested moderate interactions (Thomas et al., 2016). Sponges provide shelter for bacteria against predators and a variety of simple to complex carbon sources, and in turn bacteria are believed to help their hosts through nutrient supply, biosynthesis of essential vitamins, denitrification, chemical defense, prevention of sulfate toxicity and removal of metabolic by-products (Webster and Taylor, 2012; Thomas et al., 2016). Although bacteria comprise a major item of a sponge's diet, it seems that symbionts can be distinguished from edible (food) bacteria. We now know that bacteria use molecular mechanisms to maintain symbiosis with their corresponding sponge hosts. A noticeable mechanism is expression of eukaryotic-like proteins (ELPs) by bacterial symbionts. These are believed to aid the sponge symbionts in the establishment of an intercellular life-style. For instance, genes encoding ankyrin-repeat proteins (ARPs) in sponge symbionts may help them avoiding phagocytosis by the host phagosome by reducing vacuole acidification (Nguyen et al., 2014). Because of their presumed eukaryotic origin, it is believed that bacterial genomes horizontally acquired ELP-encoding genes from their eukaryotic hosts (Fan et al., 2012; Díez-Vives et al., 2016; Reynolds and Thomas, 2016). Therefore, the host has a distinguishing factor to accept them as part of community. To date, several genes encoding ELPs like ankyrin-repeat proteins (ARPs) (Nguyen et al., 2014), tetratricopeptide repeat proteins (TPRs) (Bröms et al., 2006), leucine-rich repeat proteins (LRRs) (Ng and Xavier, 2011), and WD40 domain proteins (Xu and Min, 2011) have been shown to be expressed within the marine sponge microbiome, and they may facilitate bacterial establishment in their hosts (Webster and Thomas, 2016). Furthermore, Type II secretion systems (T2SS), which have the tight adherence locus for biofilm formation, and Type IV secretion systems (T4SS), which have been shown to play a role against zooplankton predation, can equip symbiotic bacteria for staying with the host (Liu et al., 2016).

Besides all above-mentioned features, symbiotic bacterial communities are believed to produce most of the secondary metabolite repertoire of sponges (Fan et al., 2012; Gao et al., 2014; Tian et al., 2014). Sponges have been shown to produce a large variety of biologically active secondary metabolites, which are believed to help them to defend themselves against invading and biofouling microbes as well as predators (Taylor et al., 2007a). Among a range of secondary compounds including terpenoids, peptides, alkaloids (Lejon et al., 2011), polyketides , for example, have received increasing attention in the past few years due to their potential for the development of new drugs, especially due to their often documented antitumor activities. Importantly, many of the polyketides described from sponges thus far have been shown or suggested to be produced by bacterial symbionts rather than the sponge itself. As mentioned above, it is believed that the ecological role of biologically active natural products from sponge symbionts may support the host in defense against natural enemies (Piel et al., 2004), as successfully demonstrated for the bryozoan host *Bugula neritina* (Lopanik et al., 2004). Moreover, it is conceivable that inhibitory compounds may participate in microbe-microbe warfare within the highly dense sponge symbiotic consortium. Additionally, bacterial symbionts of marine sponges may help their hosts not only by producing secondary metabolites like polyketides but also by synthesizing essential vitamin and cofactors including vitamin B (Thomas et al., 2010; Fan et al., 2012; Webster and Thomas, 2016). Animals are unable to synthesize vitamin B which is crucial one and therefore host-associated bacteria are believed to contribute substantially in this regard.

Because of these valuable contributions, the consortium comprising the sponge host and its associated microbes is often regarded as one single functioning unit, called the holobiont (Webster and Taylor, 2012; Webster and Thomas, 2016). To ensure the consistency of key symbiotic partners, poriferans transmit their symbionts to the next sponge generation either vertically or horizontally or through a mixture of both mechanisms. During vertical transmission, specific microbes are passed from the parents to their offsprings while during horizontal transmission certain low abundant microorganisms present in the surrounding water are selectively absorbed by the sponge (Bright and Bulgheresi, 2010). Even though bacteria are also a natural food source for the filter-feeding sponge, a very large number of bacteria colonizes the mesohyl matrix of many demosponges. Because of their high bacterial density, these sponges are also known as “bacteriosponges” or “high microbial-abundance” (HMA) sponges (Reiswig, 1981; Hentschel et al., 2003) in which the bacterial density may reach 10^8 to 10^{10} cells per gram of sponge wet weight (for example, *Ircinia felix*)(Gloeckner et al., 2014). On the other hand, the term “low microbial-abundance” (LMA) applies to all those sponges

with bacterial densities around 10^5-10^6 cells per gram of sponge wet weight (for example, *Dysidea etheria*) which is in the range of bacterial cell densities commonly found in seawater (Hentschel et al., 2006). Although several functional attributes are indeed shared between these two community types (Fan et al., 2012), some questions remain to be answered: why and how do sponges host different abundances of symbionts? Are the symbiotic interactions in HMA sponge species different from those in LMA sponges? So far, investigations comparing the microbial community compositions and functions of HMA *versus* LMA sponges showed detectable variations (Bayer et al., 2014). Moitinho-Silva et al. (2014) detected higher abundance of *Proteobacteria* and *Cyanobacteria* in LMA sponges while the microbiome of HMA sponges was often populated by *Chloroflexi and Poribacteria,* among other phyla. In addition to the most diverse symbiotic prokaryotes (archaea and bacteria) (Hentschel et al., 2003; Taylor et al., 2007a), sponges harbor eukaryotes (Crustacea, Polychaeta, Cnidaria, Nemertean, and Platyhelminthes) which can grow inside or on them. Although most of these eukaryotic associated organisms are opportunistic, sponges act as a substrate for attachment and provide shelter which is crucial for in the early developmental stages of the abovementioned groups (Westinga and Hoetjes, 1981; Wulff, 2006; Padua et al., 2016).

The microbiome of Dictyoceratida sponges from the Algarve coast

The sponge fauna in Algarvian provinces, Portugal, contains diverse representatives of the Dictyoceratida family, a major group of sponges that has been studied previously in regards with their associated microbiomes. Microbial communities associated with sponge species in the Dictyoceratida order, particularly *Sarcotragus spinosulus* and *Ircinia variabilis* collected from coastal waters (Northeast Atlantic) have been characterized by high diversity of bacteria (Hardoim et al., 2012). The phyla *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Proteobacteria (Alphaproteobacteria, Gammaproteobacteria, Deltaproteobacteria)*, *Bacteroidetes*, and the candidate phyla PAUC34f*,* Anck6, and *Poribacteria* were detected as prevailing members of the *S. spinosulus* and *I. variabilis* microbiomes by a series of studies employing different techniques such as PCR-DGGE fingerprinting, 454 pyrosequencing and cultivation–dependent methods (Hardoim et al., 2012; Hardoim et al., 2014; Hardoim and Costa, 2014b). Using a so-called "plate-washing" strategy to characterize the cultivable community of these sponges by next-generation sequencing, Hardoim et al. (2014) revealed that half of the OTUs obtained via cultivation could not be detected via regular cultivation-independent next-generation sequencing (NGS) of 16S rRNA genes PCR-amplified from these samples, whereby a

pronounced dominance of diverse *Alphaproteobacteria* and *Gammaproteobacteria* phylotypes were observed in the cultivation plates (Hardoim et al., 2014). These results were interpreted as indicative of (1) low depth in regular amplicon-based surveys of the sponge-associated microbiome preventing detection of all microbes present in a sample and (2) selective enrichment, on cultivation plates, of low-abundant sponge-associated bacteria which escape usual NGS efforts employed in the characterization of microbial diversity in host-associated samples. The microbial community associated with *S. spinosulus* showed stability over three consecutive years , suggesting that strong selective forces act on the maintenance of this host-microbe relationship across time (Hardoim and Costa, 2014b). In another cultivation-dependent study, Esteves et al. (2013) characterized > 350 bacterial cultures retrieved from *I. variabilis* and *S. spinosulus* by 16S rRNA gene identification and BOX-PCR genotyping. Intriguingly, intraspecific genotypic diversity of dominant, cultivable strains belonging to diverse genera such as *Pseudovibrio*, *Ruegeria* and *Vibrio* was found to vary according to the sponge specimen from where the strains were isolated, suggesting that independent evolutionary trajectories in different host individuals may contribute to genome-wide diversification among closely related bacterial symbionts (Esteves et al., 2013). Moreover, while *Vibrio* spp. were observed to possess the most pronounced *in vitro* antimicrobial activity against model opportunistic bacteria (i.e. *Escherichia coli* and *Staphylococcus aureus*), polyketide synthase (PKS)-encoding genes were highly frequent among *Pseudovibrio* representatives, in agreement with current observations of antimicrobial activities among members of this genus and with genome-based analyses (O'Halloran et al., 2011; Alex and Antunes, 2015; Naughton et al., 2017). Unexpectedly, PKS-encoding genes were found by Esteves et al. (2013) in ten sponge-associated strains belonging to the *Aquimarina* genus, resulting in the first documentation of polyketide biosynthesis potential among cultivable representatives of *Bacteroidetes* phylum.

In summary, a considerable amount of information on the microbiome of Dictyoceratida sponges from the Algarvian shore has accumulated in the past few years. Besides the trends mentioned above on the diversity, cultivability and antimicrobial potential of symbionts, strong evidence has been gathered for host species-specific structuring of bacterial communities among Dictyoceratida sponges from both the Algarvian shore (Hardoim et al., 2012; Hardoim et al., 2014) and the Mediterranena Sea (Erwin et al., 2012a; Pita et al., 2013). In spite of these major outcomes, understanding of the functional attributes of the symbiotic consortium associated with keratose sponges in the Atlanto-Mediterranean zone is limited, and knowledge of symbiotic communities in these host has been mainly restricted to

few species such as *I. variabilis*, *I. oros*, *I. fasciculata* and *S. spinosulus* (Erwin et al., 2012a; Hardoim et al., 2012; Hardoim et al., 2014; Hardoim and Costa, 2014b). In this thesis, advanced metagenomics technologies are employed to unveil the diversity, functionality and adaptive features of uncultivated symbionts of the economically and biotechnologically relevant host *Spongia officinalis*, tackling the functional genomics of a so-far understudied sponge species in the Dictyocertida order. Moreover, genome-wide analyses are employed to uncover the coding potential of a diversified panel of alphaproteobacterial cultures (several of which phylogenetically distinct) retrieved from *S. officinalis* using alternative culture conditions. In the following sections, an overview is given on metagenomics and cultivation approaches used in the analysis of microbial communities in the environment, with particular emphasis on sponge microbiology studies.

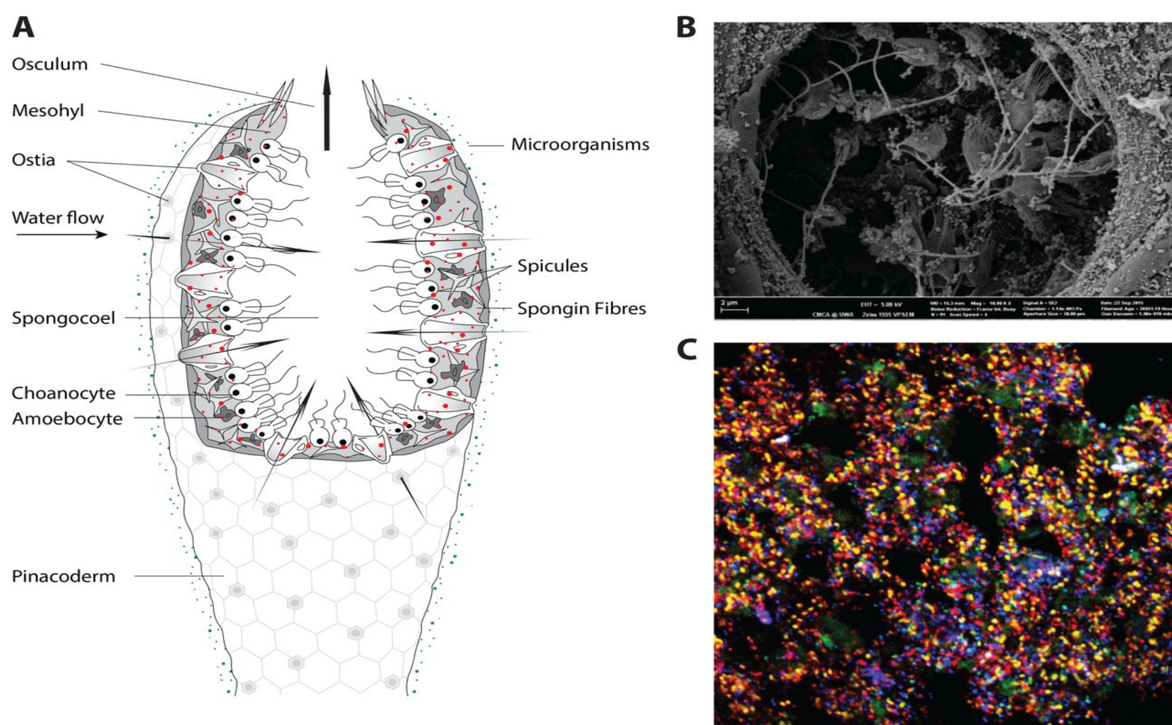


Figure 1-1. (**A**) Schematic overview of an asconoid sponge body plan. (**B**) Scan electron microscope (SEM) picture of microbes near the choanocyte chamber. (**C**) fluorescent in situ hybridization of different microbial groups in the sponge mesohyl (Poribacteria, yellow cells; *Nitrospira*, pink cells; *Chloroflexi*, cyan cells; *Deltaproteobacteria*, light (*Deltaproteobacteria*, light green cells; *Gammaproteobacteria*, red cells; *Archaea*, blue cells). Source: Webster and Thomas (2016).

Metagenomics as a tool to understand microbial community structure and function

Metagenomics is a culture-independent method used to access and explore the diversity and function of microbial communities by directly extracting information from (in an optimal circumstance) all genomes of all microbes present in each environmental sample. Some decades ago, metagenomics approaches have revolutionized microbiology. The term 'metagenome' was first introduced by Handelsman et al. (1998) as "the collective genome of microflora". The technique was developed in the 1990's by two different and independent research teams, one working on genome fragments from planktonic marine archaea and the other one on the chemistry of unknown soil microbes (Stein et al., 1996; Handelsman et al., 1998).

There are two different approaches to metagenomics studies using next-generation sequencing (NGS) technology: 1) the amplicon-based approach focusing on a specific target gene (e.g. rRNA genes) to sequences thousands of copies of, for example, 16S rRNA gene fragments amplified from the numerous different prokaryotes present in diverse microbial communities; 2) the shotgun-based approach that randomly sequences thousands to millions of small gene fragments from entire genomes present in environmental samples (Mineta and Gojobori, 2016). The main advantage of shotgun-based metagenomics is the assessment of whole genomes which allows inferring not only who is present but also what functional processes are prevailing in a given community. Yet, both approaches can be complementary and may be used to answer different scientific questions such as which organisms make part of the communities and which roles they potentially play (Izard and Rivera, 2014). To date, many different - and in part remote - open and host-associated environmental niches have been studied using metagenomics. The use of such cultivation-independent analytical pipelines permits the taxonomic and functional profiling of dominant and rare members of complex microbial consortia, circumventing biases inherent to cultivation-dependent approaches. Importantly, a third strategy worth mentioning is the exploration of metagenomes for novel genes and bioactivities with the use of DNA recombination technology (i.e. cloning) (Van Elsas et al., 2008). Although this approach does not depend strictly on the sheer force of NGS technologies (but can largely benefit from it), it is as well considered a metagenomics-based endeavor since it relies on the insertion of environmental or host-associated DNA (thus, metagenomic DNA fragments) into a cultivable heterologous host (for instance, *E. coli*) and subsequent genotypic and/or phenotypic screenings for desired bioactivities. In fact, some of the most important, recent discoveries within the sponge microbiology field have directly

benefited from this approach. These studies fostered our understanding of the pivotal roles that bacterial symbionts may play in the complex chemistry and defense mechanisms of sponges as the actual producers of manifold biologically active natural products, including a vast range of antitumoral polyketides (Hentschel et al., 2012; Wilson et al., 2014).

Modern metagenomics surveys produce a tremendous amount of raw sequencing data which need to be treated with special processing and analysis pipelines to obtain meaningful, interpretable data outputs. Databases like the Metagenomics Rapid Annotation using Subsystem Technology (MG-RAST) (Dudhagara et al., 2015) and EBI metagenomics (Hunter et al., 2014; Mitchell et al., 2016) archive and contain metagenome data and offer automated pipelines for the analysis of taxonomic and functional contents. So far, several metagenomic profiling studies in sponges have been performed (Thomas et al., 2010; Fan et al., 2012; Rua et al., 2015; Horn, 2017). Some have approached the uniqueness of the marine sponge microbiome by addressing the phylogenetic composition of this symbiotic consortium in comparison with ambient seawater (Fan et al., 2012) and also the potential contribution of sediments as sinks and sources of sponge-associated bacteria (Polónia et al., 2014; Thomas et al., 2016).

Although multiple studies (O'Halloran et al., 2011; Hardoim et al., 2012; Esteves et al., 2013; Hardoim and Costa, 2014b; Steinert et al., 2014) have tried to access sponge-associated microorganisms using cultivation, many of the abundant taxonomic groups and functionally important symbionts remain uncultivable. Functional metagenomics surveys can help to design novel culture media and conditions that meet the specific nutritional and physiological requirements of symbiont bacteria (Gutleben et al., 2017).

Reconstruction of genomes from metagenomes

The recent advance of next generation sequencing technologies and progress in bioinformatics provide feasibility and attainability for accurate genome reconstruction of uncultivated microorganisms. Because a clear majority of sponge symbionts is recalcitrant to cultivation, genome reconstruction can pave the road towards understanding the traits and roles of uncultivated symbionts in their hosts. Up to now, a small number of studies have properly described reconstructed genome sequences for microorganisms out of diverse metagenome data (Iverson et al., 2012; Luo et al., 2012a; Albertsen et al., 2013; Sharon and Banfield, 2013; Nielsen et al., 2014; Burgsdorf et al., 2015; Slaby et al., 2017). Various molecular biology

techniques and bioinformatics tools successfully obtained symbiont genomes from sponges. For example, fosmid library sequencing facilitated the recovery of the draft genome of *Cenarchaeum symbiosum* (an archaeon detected in the marine sponge *Axinella mexicana*) (Hallam et al., 2006). Through single-cell sorting, the first genome of the typically sponge-enriched candidate phylum *Poribacteria* was sequenced (Kamke et al., 2013). Metagenomics shotgun sequencing and subsequent contig binning led to the reconstruction of the draft genomes of a sponge-associated sulfur oxidizing bacterium (Tian et al., 2014) and of *Ca. Synechococcus spongiarium* (Gao et al., 2014; Burgsdorf et al., 2015). In a very recent study, Slaby and her colleagues utilized metagenomic hybrid assembly reads, namely from PacBio and Illumina sequencing technologies, to combine long and short metagenomic reads from the sponge microbiome and thereafter reconstruct the genomes of a great diversity of symbionts. Their method delivered 37 binned genomes belonging to 11 sponge bacterial phyla and two candidate phyla, including *Proteobacteria, Chloroflexi, Acidobacteria, Actinobacteria, Bacteroidetes, Gemmatimonadetes, Deinococcus-Thermus, Nitrospirae, Nitrospinae, Cyanobacteria, Spirochaetes* and the candidate phyla *Poribacteria* and SBR1093 (Slaby et al., 2017). The comparison of binned genomes has greatly improved our understanding about sponge symbionts. The aforementioned studies have shown that genes encoding for "restriction-modification", "toxin-antitoxin" and "replication, recombination and repair" functions are enriched in symbiont genomes (Burgsdorf et al., 2015; Slaby et al., 2017). On the other hand, a lower proportion of genes encoding for "signal transduction mechanisms" were detected in the genomic bins of sponge associated bacteria as compared to similar free-living fellows. This demonstrates that genome reconstructions can contribute significantly to our understanding of sponge-microbe symbiosis. It is worth mentioning here that short-read metagenomic sequencing, which is providing enough sequencing depth by repetition, has its own disadvantage. Short reads create gaps through assembly and therefore, it is hardly possible to obtain a complete genome by pure sheer sequencing and assembly from metagenomes. This disadvantage can be slightly resolved by using long-read sequencing, but still improvements are needed to achieve 100% genome coverage of genomes assembled from short reads binned from complex metagenomes (Slaby, 2017).

Cultivation bias in natural environments and strategies to cultivate the uncultivable

There is still a large gap in cultivating microbes from sponges – and from environmental samples in general - in the light of the existent microbial diversity in open ecosystems as inferred by cultivation-independent surveys using NGS technologies. Cultivating microorganisms is still fundamental to understand the biology and ecology of microbial species. Microbial cultures provide an opportunity to obtain complete, high-quality genome sequences of one single organism and to identify the properties of those organisms that cannot be discerned from genome sequencing alone (Connon and Giovannoni, 2002). Further, ignoring microbial cultivation in favor of the sole use of ribosomal RNA gene approaches can lead to important gaps in microbial community diversity data, because some species in a community sometimes remain undetected in cultivation-independent surveys (Donachie et al., 2007). This can be due to several reasons such as difficulties in cell lysis prior to DNA extraction or the lack of sequencing depth in most microbial diversity NGS surveys. Yet, these bacteria may be obtained on culture plates as cultivation follows different principles. However, cultivation methods are thought to depict only 1% of the total microbial diversity in open ecosystems while molecular methods are often believed to enable full access to so far uncultivable bacteria. In the case of sponge-associated bacteria, it has been shown that between 1% and 14% of the whole community of associated microbes can be cultivated (Webster and Hill, 2001; Olson and McCarthy, 2005; Taylor et al., 2007a; Sipkema et al., 2011; Hardoim et al., 2014). Most bacterial strains in sponges reside inside the mesohyl, where the environment has little similarity to outside surroundings (seawater), and this should be considered when designing any cultivation experiment. Anoxic conditions may occur inside the sponge mesohyl when the sponge temporarily stops to pump water (Hoffmann et al., 2005; Hoffmann et al., 2008). In addition, the iron concentration in the mesohyl may be higher than in the surroundings because of the presence of siderophores (Onuki and Kamino, 2000). Moreover, if the target bacteria need light for growth like *Cyanobacteria*, this should obviously be considered in cultivation and incubation designs. Another aspect to be taken into account is the addition of sponge-derived compounds to the culture medium as a means to support the growth of specific symbiotic bacteria, as attempted in an early cultivation study using lectin, which is in principle a mesohyl component, amendments to the culture medium (Müller et al., 1981). Furthermore, prolonged incubation periods may help slow-growing bacteria to develop (Connon and Giovannoni, 2002), especially if carbon offer is not exceedingly high. Taken together, different circumstances such as the quality and quantity of the substrate, the depletion of some nutrients

or the absence of unknown requirements, viral infections or, enrichment of poisonous products might influence the cultivation of marine bacteria (Eilers et al., 2000), as well as unsuspected mutual dependencies of species living in consortia (Hentschel et al., 2006; Taylor et al., 2007a). Recently, it has been suggested that sample processing of sponge bacterial symbionts may affect bacterial viability after cell detachment from the hosts (Esteves et al., 2016). Therefore, this is also an important obstacle which should be taken into consideration in future, alternative cultivation approaches.

To overcome the difficulty in cultivating sponge bacteria, innovative and promising approaches are required. For example, in a recent study the usage of multiple agar media (with and without antibiotics) combined with the picking of individual colonies and also scraping total bacterial growth from cultivation plates led to the captivation, in the laboratory, of a range of uncultivated sponge bacteria including two new genera and a new species of *Flavobacteriaceae* (Versluis et al., 2017). Likewise, the use of agar plate, liquid, and floating filter culture methods with media containing organic sponge extracts and bacterial signal molecules permitted the isolation of rare bacteria belonging to phyla such as *Planctomycetes*, *Verrucomicroba*, and *Deltaproteobacteria* (Sipkema et al., 2011). Moreover, membrane-based diffusion growth chambers (DGCs, constructed from two combined centrifuge microfilters) have been used to capture sponge symbionts (Steinert et al., 2014). This resulted in the cultivation of fifteen so far ‘uncultivable’ bacteria of the phyla *Bacteroidetes* and *Proteobacteria*. As demonstrated through the abovementioned studies, sponges can be a very rich reservoir for the isolation of novel bacteria as well as novel bioactive compounds. Therefore, finding favorable conditions to grow sponge-associated microbes in the laboratory is a clear interest of the scientific community and the pharmaceutical industry. Although the cultivation process often is time-consuming, laborious and sometimes monotonous, the beauty in color or shape developed by some bacterial species growing on a culture plate can be truly amazing (**Figure 1-2**). Because cultivation allows the discovery of novel gene functions and metabolic processes that ultimately can unveil the role of microbes in their environment, it still is a recommended method by many renowned scientists even in an era of next generation sequencing approaches (Keeling and Campo, 2017). In this thesis, the combination of a low-nutrient medium with lower incubation temperatures and longer incubation periods was employed as a simple and reliable alternative strategy to recover phylogenetically unique bacterial symbionts from *Spongia officinalis*.

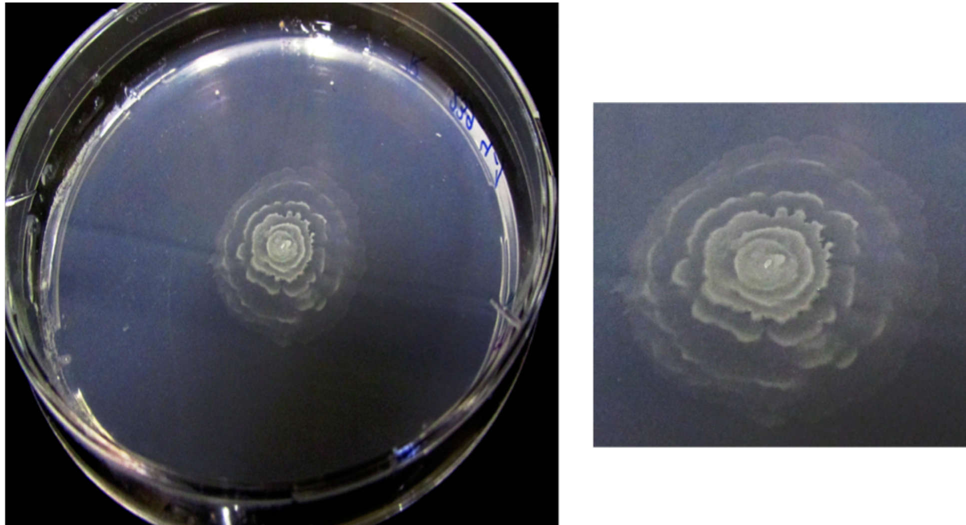


Figure 1-2. An *Alphaproteobacterium* isolated from *Spongia* (growing on R2A medium) –Cultivation and Photo by E. Karimi.

Genome sequencing and comparative genomics to identify the functional features of symbiont genomes

Due to recent technological advances, entire bacterial genomes can nowadays be sequenced quickly and at low costs. In the last decade, the number of complete bacterial genome sequences in public sequence databases has greatly and rapidly increased. For example, the number of bacterial and archaeal genomes sequenced until 2015 was higher than 14,000 genomes (Land et al., 2015).

The majority of bacterial genomes consist of circular chromosomes. They have a single origin of replication and lack structural proteins called histones (proteins that fold eukaryotic DNA). Nevertheless, some bacteria like *Streptomyces coelicolor* (Kieser et al., 1992) and *Agrobacterium tumefaciens* (Allardet-Servent et al., 1993) have linear chromosomes. Genome size, synteny, replicon numbers, and G+C content are common parameters used to describe bacterial genomes (Bentley and Parkhill, 2004). Prokaryotic genome sizes vary across phyla (Bentley and Parkhill, 2004), and different phyla or species of bacteria may demonstrate different patterns of correlation between DNA size and G+C content (Li and Du, 2014). It has further been documented that obligate host-associated bacteria often contain short genomes with low G+C content (McCutcheon and Moran, 2012). The relationship between genome size and content appears to be mainly determined by environmental pressure (Ranea et al., 2004). Basically, the causes for varying genome sizes are related to the genetic information needed to persist in new environments and to population dynamics (Moya et al., 2008). When a bacterium

becomes a symbiont, some genes may become unnecessary or redundant because their functions can be provided by the host organism (Moya et al., 2008). Therefore, reduced genome size may be used as an indicator in the identification of obligate symbiont bacteria (Pérez-Brocal et al., 2006). However, there are far more facultative symbionts interacting with their host which do not necessarily have a small genome. They are rather equipped to be versatile and able to persist and survive in many different environments, for example, *E. coli*, *Vibrio*, *Pseudomonas*, *Burkholderia* and many of the typical model facultative symbionts/opportunistic pathogens whose genomes are not small. In addition to genome reduction, quorum sensing (QS), expression of eukaryotic-like protein (ELPs) and type II and VI secretion systems as adherence factors (Liu et al., 2016) have been suggested to facilitate the establishment of bacteria in their respective hosts as well as bacteria-bacteria interactions (Webster and Thomas, 2016).

Bacterial genome content may be investigated on three different levels, genomic, transcriptomic, and proteomic (Binnewies et al., 2006), and each one can help to understand different aspects of a bacterium's life-style and metabolic capabilities. Each level can significantly improve our knowledge on how bacteria evolve and adapt to best fit to their niches (Ochman and Moran, 2001; Tian et al., 2017). For instance, how new genetic material may be gained via horizontal gene transfer (Moya et al., 2008). Such approaches also provide insights into the molecular mechanisms employed by bacteria to survive and persist in nature, for example, how pathogenic and symbiotic bacteria escape host immune responses when making part of a dense microbial consortium such as the sponge microbiome (Horn et al., 2016). Thus, genomics is a first method (in a series of many) to pursuit a deeper understanding on how bacteria choose and act in their niches, how the host controls its endosymbiotic bacterial communities, how different bacteria interact with each other without losing too much energy in competition, and how bacteria may "convince" their host to keep them as permanent residents. It is worth mentioning here that a genome (genomics) alone can only give limited information on the questions mentioned above. Genomics will only indicate that a bacterium may be able to carry out a certain function, but never if it really does. For any given bacterium or microorganism to carry out a function, the genes need to be also transcribed (transcriptomics) and translated (proteomics). Even after the gene is translated into an actual protein, laboratory tests (mesocosm experiments) are needed to study the function *in vitro* or *in vivo*. Ultimately, to proof functions and/or assign new gene functions, researchers need to create mutants where target genes are knocked-out or expressed in a different bacterium by genetic recombination.

Aims and specific research questions

The general aim of this PhD thesis was to reveal the functional capacities and adaptive strategies of *Spongia officinalis* bacterial symbionts using cultivation-independent (metagenomics and genome reconstruction) and cultivation-dependent (cultivation and whole-genome-sequencing) approaches. In this framework, emphasis was given to the genomic signatures and potential roles played by sponge-associated *Alphaproteobacteria* given (1) the relevance of this bacterial group in marine ecosystems, (2) the paucity of information pertaining to the functional attributes of alphaproteobacterial symbionts of marine sponges and (3) the profuse recovery, by our team, of diverse sponge-derived *Alphaproteobacteria* using alternative cultivation strategies.

Spongia officinalis (Linnaeus, 1759) (**Figure 1-3**) is the first-ever described sponge species, being well known for its commercial use as a bath sponge (Voultsiadou et al., 2008). Its range of occurrence encompasses the Mediterranean Sea (Dailianis et al., 2011) and the northeastern Atlantic Ocean (World Porifera Database, Van Soest et al. (2011)). This species produces many types of structurally diverse metabolites, some of them with pharmaceutical potential (Li et al., 2017). Since *S. officinalis* has the capacity to concentrate all the trace metals present in its surroundings, it is also being used to study the metal availability and load in marine ecosystems (Bauvais et al., 2015). Some specimens have a superficial fiber net supporting the pinacoderm. The whole body is compressible and resilient except where the surface is heavily sand-encrusted (de Cook and Bergquist, 2002). Signs of *S. officinalis* population decline and mortality events have been documented recently due to human impacts on the environment, global warming and invasion of pathogenic microorganisms (Webster, 2007; Garrabou et al., 2009). Despite the biotechnological potential of *S. officinalis* and its associated microbiome, functional information regarding its symbiotic community is rare. Therefore, this thesis makes use of *S. officinalis* as a model organism to approach sponge microbiome functionality, cultivability and adaptive features. The collection of the *S. officinalis* specimens, seawater and sediment samples analyzed in this thesis took place in May 2014 by SCUBA diving at 20 m depth off the coast of Pedra da Greta (36°58'47.2"N, 7°59'20.8"W), Algarve, South Portugal (**Figures 1-4**). This sampling event provided all the biological data used to achieve the specific objectives of this thesis, described below.

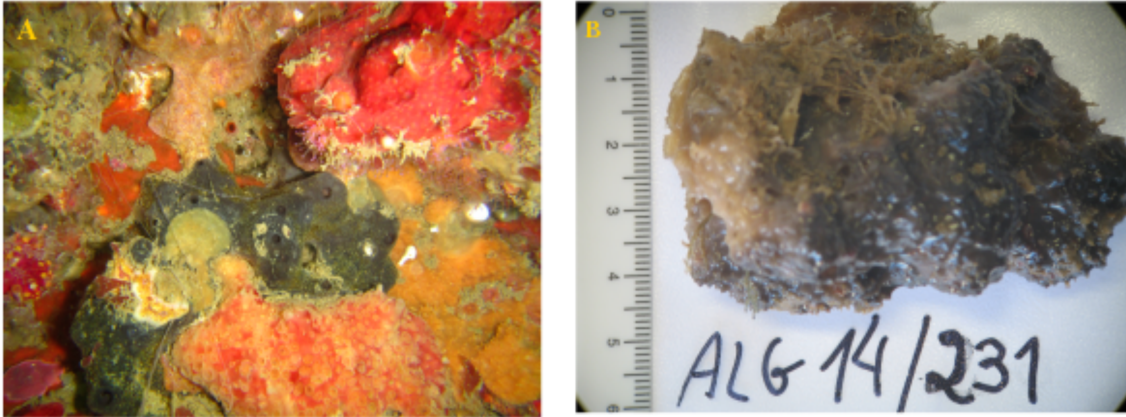


Figure 1-3. Photographs of *Spongia officinalis in vivo* **(A)** and *in vitro* **(B)**. (A) Photograph courtesy of Dr. Jorge Goncalves's team (B) Photograph by E. Karim.

Cultivation-independent approaches to extract functional information from yet uncultivated bacterial symbionts

Metagenomics can be broadly defined as the cultivation-independent analysis of all the DNA content present in a given sample, be it "environmental" or "host-associated". Thus, a comprehensive and well implemented metagenomics-based strategy will optimally enable access to the genomes of all (micro) organisms present in a sample. In this thesis, this strategy was applied to achieve and fulfill the following specific aims and tasks:

1- To uncover the functional and taxonomic structure of *S. officinalis* associated microbial communities, adressing the complementary roles of marine sediments and seawater in contributing to symbiont assembly in this host.

2- To determine the unique genomic features of the *S. officinalis* endosymbiotic consortium and provide an evolutionarily-based interpretation of its pivotal life strategies.

3- To reveal the differential abundance patterns of genomic adaptive features of marine sponge symbionts, such as secondary metabolite biosynthesis, CRISPR-Cas systems, and carbon metabolism traits across *S. oficinalis*, seawater and sediments to better understand the distribution of metabolic resources of biotechnological value across marine biomes.

Because recently-developed bioinformatics techniques provide an opportunity to bin genomes out of metagenome data, in this thesis a sequence composition-dependent binning approach was used to assemble genomes of uncultivable bacteria from *S. officinalis*, seawater

and sediments, placing focus on the functional features and adaptive traits of *Rhodospirillaceae* symbionts (*Alphaproteobacteria*) from marine sponges. This strategy was applied to achieve the following specific objectives:

1. To reconstruct a consensus genome from a dominant, uncultured *Alphaproteobacteria* symbiont (family *Rhodospirillaceae*) of *S. officinalis* using microbial metagenome data derived from this host.
2. To reveal the unique genomic signatures that differentiate symbiotic from free-living *Rhodospirillaceae* lineages by determining their prevailing functions (e.g. Clusters of orthologous protein groups - COGs), highlighting the most discriminatory protein families retrieved from symbiotic vs. free-living phylotypes.
3. To delineate the possible roles played by cultured and uncultured sponge-associated *Alphaproteobacteria* spp. and their likely host fitness-enhancing value within the marine sponge holobiont.

Alternative isolation and cultivation of sponge-associated bacteria

Making parallel use of cultivation and cultivation-independent molecular techniques is the most adequate strategy to fully comprehend the extent of microbial diversity within a community (Donachie et al., 2007), since cultured phylotypes will help to provide a comprehensive genomic sequences database. In this thesis, a dedicated strategy for the isolation of extended bacterial diversity was implemented according to the following specific aims and tasks:

1. Generation of a collection and taxonomic identification of pure bacterial cultures retrieved from the marine sponge *S. officinalis* using alternative cultivation methodologies which favor the growth of difficult-to-cultivate bacteria.

2. Selection of pure bacterial cultures that represent novel species and/or so far rarely cultivated phylotypes for whole genome sequencing, assembly and annotation.

3. Genome-based characterization of the symbiotic living capabilities and metabolic capacities of *Alphaproteobacteria* sponge symbionts, the most prolific and diverse bacterial group cultivated in this thesis with the use of novel methodologies.

Combining information gained from both, genomics and metagenomics, contributes to better understanding of microbial community functioning. Yet scientists still need to sufficiently link these two techniques as “two sides of the same coin” (Gutleben et al., 2017). In this thesis, such as integration is attempted in a comprehensive fashion through a dedicated study of the *Spongia officinalis* microbial metagenome (Chapter 2) coupled to in-depth comparative genomics of ten alphaproteobacterial symbiotic genera isolated from *S. officinalis* using a novel cultivation platform (Chapter 3) and genome reconstruction of a so-far uncultivable alphaproteobacterial symbiont of the family *Rhodospirillaceae* from metagenomic data (Chapter 4), as more accurately described in the outline below.

Figure 1-4. *Spongia officinalis* sampling site at the Algarve coast, with the exact sampling location marked in red (Copyright © 2017Worksheetworks.com and © Google map).

Thesis outline

The first aim of this study (Chapter 2) was to approach 10 microbial metagenomes including four from sponges (*Spongia officinalis*), three from surrounding seawater and three from sediment replicates. Metagenomes were analyzed using EBI and MG-RAST pipelines to compare the sponge, sediment and seawater microbiomes taxonomically and functionally. The adaptive strategies of the marine sponge microbiome were delineated based on functions that were highly enriched or de-selected in this consortium in comparison with seawater and sediment.

The second aim of this thesis (Chapter 3) was to cultivate novel sponge-associated bacteria by utilizing an oligotrophic marine medium, decreased incubation temperature and an extended incubation period to allow slow growing bacterial symbionts to develop into colony forming units (CFUs). Purified colonies were taxonomically identified by 16S rRNA gene sequencing. Ten phylogenetically distinct *Alphaporoteobacteria* strains were selected for genome sequencing to explore their functional characteristics, symbiont features and the presence of genes encoding for the biosynthesis of secondary metabolites.

The third aim (Chapter 4) was to reconstruct the genome of an abundant, uncultivatable alphaproteobacterial symbiont of the family *Rhodospirillaceae* from our model sponge by applying a sequence composition–dependent binning approach. A comparative genomics survey was then undertaken to decipher the adaptive strategies of sponge symbiotic *Rhodospirillaceae* taking the genome composition of close, free-living relatives into account.

A general discussion of the results obtained in thesis and the novel bacteria isolated as well as future perspectives is provided in the last Chapter (Chapter 5).

Chapter 2

Comparative metagenomics reveals the distinctive adaptive features of the *Spongia officinalis* endosymbiotic consortium

Comparative metagenomics reveals the distinctive adaptive features of the *Spongia officinalis* endosymbiotic consortium

Elham Karimi[1], Miguel Ramos[1$], Jorge M.S. Gonçalves[2], Joana R. Xavier[3], Margarida P. Reis[4], Rodrigo Costa[1,5]

[1]Microbial Ecology and Evolution Research Group, Centre of Marine Sciences, Algarve University, Gambelas 8005-139 Faro Portugal

[2]Fisheries, Biodiversity and Conservation Research Group, Centre of Marine Sciences, Algarve University, Gambelas 8005-139 Faro Portugal

[3]Department of Biology and KG Jebsen Centre for Deep-Sea Research, University of Bergen, Thormøhlensgate 53A, 5006 Bergen, Norway

[4]Faculty of Science and Technology, Algarve University, Gambelas 8005-139 Faro Portugal

[5]Institute for Bioengineering and Biosciences (IBB), Department of Bioengineering, IST, Universidade de Lisboa, Lisbon, Portugal

A version of this chapter has been published as an original research article in *Frontiers in Microbiology*, Doi: org/10.3389/fmicb.2017.02499.

Abstract

Current knowledge of sponge microbiome functioning derives mostly from comparative analyses with bacterioplankton communities. We employed a metagenomics-centered approach to unveil the distinct features of the *Spongia officinalis* endosymbiotic consortium in the context of its two primary environmental vicinities. Microbial metagenomic DNA samples (n = 10) from sponges, seawater and sediments were subjected to Hiseq Illumina sequencing (*c.*15 million 100 bp reads per sample). Totals of 10,272 InterPro (IPR) predicted protein entries and 784 rRNA gene operational taxonomic units (OTUs, 97% cut-off) were uncovered from all metagenomes. Despite the large divergence in microbial community assembly between the surveyed biotopes, the *S. officinalis* symbiotic community shared slightly greater similarity ($p < 0.05$), in terms of both taxonomy and function, to sediment than to seawater communities. The vast majority of the dominant *S. officinalis* symbionts (i.e., OTUs), representing several, so-far uncultivable lineages in diverse bacterial phyla, displayed higher residual abundances in sediments than in seawater. CRISPR-Cas proteins and restriction endonucleases presented much higher frequencies (accompanied by lower viral abundances) in sponges than in the environment. However, several genomic features sharply enriched in the sponge specimens, including eukaryotic-like repeat motifs (akyrins, tetratricopeptides, WD-40 and leucine-rich repeats), and genes encoding for plasmids, sulfatases, polyketide synthases, type IV secretion proteins and terpene/terpenoid synthases presented, to varying degrees, higher frequencies in sediments than in seawater. In contrast, much higher abundances of motility and chemotaxis genes were found in sediments and seawater than in sponges. Higher cell and surface densities, sponge cell shedding and particle uptake, and putative chemical signaling processes favoring symbiont persistence in particulate matrices all may act as mechanisms underlying the observed degrees of taxonomic connectivity and functional convergence between sponges and sediments. The reduced frequency of motility and chemotaxis genes in the sponge microbiome reinforces the notion of a prevalent mutualistic mode of living inside the host. This study highlights the *S. officinalis* “endosymbiome” as a distinct consortium of uncultured prokaryotes displaying a likely “sit-and-wait” strategy to nutrient foraging coupled to sophisticated anti-viral defenses, unique natural product biosynthesis, nutrient utilization and detoxification capacities, and both microbe-microbe and host-microbe gene transfer amenability.

Introduction

Sponges (phylum Porifera) rank among the oldest extant metazoans and are distributed worldwide across all oceans and major freshwater bodies, displaying various shapes, sizes and colors, which are possibly influenced by environmental and biotic conditions (Hentschel et al., 2006; Pineda et al., 2015). There are about 8,500 sponge species described to date and likely as many to be described (Van Soest et al., 2012). These sessile, filter-feeding organisms usually shelter dense and complex microbial communities often dominated by diverse, active and phylogenetically distinct bacteria (Taylor et al., 2007a; Kamke et al., 2010; Thomas et al., 2010). Indeed, although sponges intake numerous planktonic microorganisms due to their remarkable filtering activity, their symbiotic communities are taxonomically and functionally different from those found in the water body (Thomas et al., 2010; Costa et al., 2013; Thomas et al., 2016). Until now, little experimental evidence exists for the actual participation of sponge symbionts in contributing to host fitness (Webster and Thomas, 2016). However, sponge-associated microorganisms are believed to benefit their hosts through several, eventually interdependent, mechanisms. These include nutrient provision (e.g. through the synthesis of photosynthates and vitamins (Taylor et al., 2007a; Siegl et al., 2011); *in-host* geochemical cycling (e.g. via nitrification (Bayer et al., 2008; Radax et al., 2012), denitrification (Siegl et al., 2011; Fan et al., 2012), or polyphosphate production (Zhang et al., 2015)); chemical defense (e.g. via the biosynthesis of polyketides (Piel et al., 2004; Wilson et al., 2014); and removal of metabolic by-products such as ammonia (Webster and Taylor, 2012; Webster and Thomas, 2016) and sulfide (Hoffmann et al., 2005). Particularly, the phylogenetic distinctiveness of the marine sponge microbiome and its vast natural product biosynthesis repertoire have both propelled much research interest in this symbiotic relationship (Taylor et al., 2007a; Wilson et al., 2014).

In the last ten years or so, metagenomics (Handelsman, 2001) and single cell genomics (SCG) (Woyke et al., 2009) approaches coupled to next generation sequencing (NGS) technologies have become the tools of trend in the inspection of microbial communities thriving in open and host-associated microniches (Handelsman, 2001; Kennedy et al., 2008; Gilbert and Dupont, 2011; Kumar et al., 2015). Functional gene profiling via shotgun NGS revealed that sponge symbiont communities share a suite of common genetic signatures underlying “specific” adaptive strategies such as high frequencies of eukaryotic-like proteins (ELPs), possibly involved in patterns of host-symbiont recognition, and Clustered Regularly Interspaced Short Palindromic Repeats and associated systems (CRISPR-Cas), that may function as a collective

anti-viral defense system within the sponge symbiotic consortium (Thomas et al., 2010; Fan et al., 2012; Rua et al., 2015; Horn et al., 2016). However, the frequency and abundance of such genetic elements in other marine microhabitats have not yet been fully examined, making it difficult to diagnose them as exclusive adaptive features of marine sponge symbionts. The linkage between identity and function has been now identified for a number of symbiotic lineages, either via SCG or genome binning from metagenomes (Siegl et al., 2011; Moitinho-Silva et al., 2017; Slaby et al., 2017), greatly increasing our knowledge of the potential physiology of particular sponge-enriched lineages belonging e.g. to the *Cyanobacteria*, *Proteobacteria* and *Poribacteria* phyla (Kamke et al., 2013; Gao et al., 2014; Burgsdorf et al., 2015).

In spite of the continued progress enabled by modern cultivation-independent tools, our current understanding of marine sponge microbiome diversity and function mostly derives from comparative studies with the neighboring bacterioplankton (Fan et al., 2012; Rua et al., 2015; Thomas et al., 2016), whereas knowledge of the potential contribution of sediments as sinks and sources of sponge-associated bacteria remains limited. Only recently have studies emerged which investigated sediments in comparative analyses with sponge symbiotic assemblages, using amplicon-based approaches to address the taxonomy and, eventually, *in silico* functional estimates of the examined communities (Polónia et al., 2014; Thomas et al., 2016). Recent evidence suggests that the density and biochemical composition of particles are major drivers of microbial community structure in aquatic microniches (Zhang et al., 2016). Here, we hypothesize that higher particle/surface availability and cell densities may promote the selection of identifiable traits common to sponge-associated and sediment communities not necessarily favored in planktonic settings. To address this hypothesis, in this study we tested whether (1) whole taxonomic and functional profiles and (2) abundance distributions of genotypic traits usually regarded as adaptive features of marine sponge symbionts were significantly different across sponge, sediments and seawater microbial metagenomes.

Spongia officinalis Linnaeus 1759, the first described sponge species, is a canonical bathing sponge (Voultsiadou et al., 2008) displaying widespread occurrence from across the Mediterranean Sea (Dailianis et al., 2011) into the northeastern Atlanctic Ocean and beyond. However, signs of population decline as a consequence of human activity, warming temperatures and bacterial infections have been accumulating in recent years (Webster, 2007; Garrabou et al., 2009). *S. officinalis* belongs to the chemically-rich order Dictyoceratida (Gordaliza, 2010), and as such is the source of diverse biologically-active natural products (Gonzalez et al., 1984; Manzo et al., 2011). In spite of the unequivocal economic and societal

relevance of *S. officinalis*, functional information concerning its symbiotic community is scarce. Here, we employ *S. officinalis* as a model organism to quantitatively address the functional and taxonomic (dis)similarity between sponge, sediment and seawater microbiomes using shotgun metagenomic sequencing. To reveal the distinctive genomic features of the *S. officinalis* symbiotic consortium in the context of its natural environment, we used customized pipelines enabling differential abundance analysis of symbionts (i.e., Operational Taxonomic Units - OTUs - set at 97% 16S rRNA gene similarity) and predicted protein families/domains/sites (i.e. InterPro - IPR - entries) across the studied biotopes. Alternative analytical pipelines were used to verify the consistency of the major trends found, and to compare the *S. offcinalis* microbial metagenome with those of other sponge hosts.

Materials and Methods

Sampling and sponge identification

Sampling of *Spongia officinalis* specimens (*c.* 10 g, n = 4), seawater (2L, n = 3) and sediments (c. 50 g of upper 5 cm layer, n = 3) took place in May 2014 by SCUBA diving at 20 m depth off the coast of Pedra da Greta (36º 58' 47.2N ;7º 59' 20.8W), Algarve, southern Portugal. Seawater samples were taken 1 m above the sponge specimens, while sediment samples were taken 1 m away from the sampled specimens. Underwater procedures and sample transportation were as described previously (Hardoim et al., 2012). Water pH was 8.13, temperature 18º C, and salinity 36.40 ‰. Sponge individuals were identified in the laboratory using standard macro- and microscopic morphological criteria (Hardoim et al., 2012). To aid the traditional identification of the specimens, phylogenetic inference of the subunit I of the mitochondrial cytochrome oxidase (CO1) gene was undertaken. To this end, total community DNA was directly extracted from 0.25g of the inner body of each specimen (see below). Amplification, sequencing and phylogeny of CO1 genes were performed using previously established procedures (Hardoim et al., 2012; Hardoim and Costa, 2014b).

Microbial metagenomic DNA extraction and Next-Generation Sequencing (NGS)

For the analysis of the sponge-associated endosymbiotic community, microbial cell pellets were retrieved from 2.5 g of the inner sponge body as detailed elsewhere (Hardoim et al., 2014). Briefly, cell homogenates obtained from the samples by maceration in calcium/magnesium free artificial seawater (CMFASW) (Garson et al., 1998) were subjected to a differential centrifugation step adapted from earlier protocols (Fieseler et al., 2006; Thomas et al., 2010).

Seawater samples (2L) were passed through 0.22 µM nitrocellulose membranes which were thereafter cut into small pieces, whereas 0.25g of sediment were retrieved from each sample after aseptic sieving (1 mm mesh) and thorough homogenization. All processed samples, including excised sponge pieces used for phylogenetic inference (see above), were stored at -80 ^{0}C prior to total community DNA (TC-DNA) extraction with the UltraClean® Soil DNA isolation kit (MO BIO, Carlsbad, CA, USA) following the manufacturer's instructions. TC-DNA quantity and concentration were determined using the Qubit (Life Technologies Qubit 2.0®) dsDNA HS Assay Kit. Next generation TC-DNA sequencing was performed on an Illumina Hiseq 2500 device at Mr. DNA (Shallowater, TX, USA). DNA libraries were prepared for sequencing using the Nextera DNA Sample preparation kit (Illumina) after the manufacturer's instructions, and sequenced paired end for 200 cycles with sequence depth calibrated at *c.* 15 million 101-bp reads per sample.

Metagenome data processing

Preliminary data processing and analysis revealed that, for most highly ranked taxa (domains, phyla, classes), no sensible changes in microbial community composition could be detected between assembled and unassembled data. However, assembly procedures often reduced considerably the total number of reads that could be used in downstream analysis, especially for sediment samples (**Appendix I- File S1**). Therefore, for the purposes of this study, we primarily employed complementary tools within the Meta-Genome Rapid Annotation using Subsystems Technology server (MG-RAST) v3.0 (Meyer et al., 2008) and the EBI Metagenomics platform (EMG) v2.0 (Mitchell et al., 2016) to deliver accurate metagenomic profiling from unassembled reads, making optimal use of all information generated by our sequencing effort. Prediction of cds, translation into protein sequences and annotation searches were performed using default settings in both MG-RAST and EMG (hard-coded data processing). Briefly, within MG-RAST gene calling was performed using FragGeneScan (Rho et al., 2010), and predicted cds were translated into proteins with clustering at 90% identity level using uclust (Edgar, 2010). Within the EMG pipeline, after quality filtering and length trimming, reads with rRNA sequences were detected using RNA Selector and subjected to taxonomic profiling using QIIME for OTU picking, clustering (at 97% gene similarity) and classification. Reads with rRNA masked were subjected to cd prediction using FragGeneScan, and predicted cds were finally processed with InterProScan for functional annotation against the InterPro database release 31.0, which integrates several protein sequence databases such as Pfam, TIGRFams and PANTHER, among others.

With MG-RAST we extracted an "all domains-all reads" profile of microbiome structure based on all sequences (including "phylogenetic marker" and "functional" genes) that could be assigned a taxonomic origin. Sequencing reads were annotated using the best-hit annotation tool against the M5NR database (Wilke et al., 2012).

The stringency of the BLAST parameter was a maximum e-value of 1e-5, a minimum identity of 60 %, and a minimum alignment length of 15 measured in aa for predicted proteins and in bp for RNA databases. A negligible amount (0.02%) of the reads obtained from *S. officinalis* specimens was assigned as of poriferan origin using MG-RAST, corroborating the efficiency of the microbial enrichment protocol used to process these samples. With the EMG data processing pipeline, we obtained taxonomic and functional profiles of the metagenomes based on 16S rRNA genes (archaeal, bacterial, and microeukaryotic - chloroplast and mitochondrial - operational taxonomic units - OTUs) and InterPro (IPR) protein domain entries, respectively, fetched from the data (Mitchell et al., 2016). Our downstream statistical analyses focused primarily on the OTU and IPR contingency tables delivered using the EMG pipeline given the high dominance of bacterial reads (> 95% of the classifiable reads) verified using MG-RAST, and the possibility to explore the widely integrative, comprehensive and updated InterPro protein sequence database (Finn et al., 2017). Complementary analyses on COG annotations derived from both unassembled and assembled data were performed, and are detailed below.

Metagenome data analysis

The contingency (OTU and IPR) tables extracted from the EMG data processing pipeline were imported into R version 3.2.4 (RCoreTeam, 2015) using the read.delim() function. Since differences in library sizes among samples did not require rarefaction of the data to the least sequenced samples (McMurdie and Holmes, 2014; Weiss et al., 2017) analyses were performed on the full OTU and IPR datasets after Hellinger transformation of the data. This procedure was found to perform better than using relative abundances alone to assess variability in IPR and OTU data across samples by preventing the emergence of false positives and down-weighting the impact of very dominant IPR entries (usually representing primary metabolic traits) on the determination of most differentiating functional attributes among biotopes. Variation in taxonomic (OTU) and functional (IPR) microbial community structures across sediment, seawater, and sponge samples was assessed by principal coordinates analysis (PCoA) using Bray–Curtis dissimilarity matrices as input data within the cmdscale() function in R. Differences were tested for significance by permutational analysis of variance

(PERMANOVA) using the abovementioned matrices and the adonis() function within the VEGAN package, with the number of permutations set at 1000. Similarity Percentage (SIMPER) (Clarke, 1993) analyses were performed with the PAST software v. 3.14 (Hammer et al., 2001) to rank the individual contribution of each annotated OTU and IPR to total data variation in taxonomic and functional profiles, respectively. Analyses performed using Euclidean (instead of Bray-Curtis) distances led to equivalent outcomes and are available on request.

Pairwise tests of significance were run to diagnose differences in IPR, OTU and phylum relative abundances among biotopes with the sim() function in R using Hellinger-normalized data as input. Heat maps were generated to display the top, most differentiating microbial phyla, OTUs and IPR entries (identified via SIMPER analyses) using the heatmap2() function in the gplots package within R. Additionally, we manually inspected all IPR entries oscillating significantly among the biotopes to identify potential "umbrella" functions of likely ecological and evolutionary relevance for sponge microbiome assembly, including traits usually regarded as "specific" genomic signatures of sponge symbionts, and assessed the cumulative contribution of all IPR entries belonging to these so-created, major functional categories in distinguishing between the biotopes. To test whether abundance values of major functional categories assembled manually (**Figure 2-5**) varied significantly among the biotopes, the Shapiro-Wilk statistics was computed in R to inspect the distribution of each measure around means. Thereafter, one-way ANOVA was performed followed by an all pairwise multiple comparison procedure using the Tukey's HSD (honest significant difference) test. For non-normal distributed data, Kruskal-Wallis one-way analysis of variance by ranks was employed followed by post-hoc kruskal nemenyi tests for pairwise multiple comparisons. The same strategy was applied to test for differences among Bray-Curtis dissimilarities between samples, calculated for both IPR and OTU data (**Table 2-1**).

Alternative analytical pipelines and data validation

Besides the core analyses described above using EMG taxonomic and functional profiling of unassembled reads, we performed COG-based annotations of both unassembled and assembled reads on MG-RAST. Assembly of metagenomes was carried out using MetaVelvet (Namiki et al., 2012) with default parameters. Thereafter, assembled and unassembled data were processed within MG-RAST as described above. Predicted protein sequences were searched against the COG database (Tatusov et al., 2003) using a maximum e-value of $1e^{-10}$, minimum identity of 60 %, and minimum alignment length of 15 aa. The resulting COG *vs.* samples tables, for

unassembled and assembled data, were then subjected to ordination analysis using Hellinger transformation and PCoA, as described above, to test whether COG profiles were different according to their origin (that is, *S. officinalis*, seawater and sediments). To contrast the functional profiles retrieved from *S. officinalis* with those obtained for other sponge hosts, we downloaded the COG annotations available on MG-RAST describing the microbiomes of *Rhopaloeides odorabile* (id: mgm4530290.3), *Cymbastela concentrica* (id: mgm4530252.3) and *Cymbastela coralliophila* (id: mgm4530427.3)(Fan et al., 2012) and merged them with COG annotations retrieved in this study in a single file. Only assembled data were used in this comparison. The resulting COG vs. samples matrix was subjected to ordination analysis after Hellinger data transformation as delineated above. Venn diagrams were constructed using Venny 2.1.0 (Oliveros, 2007) to count the number of specific and shared COGs across the four analysed sponge species. Finally, the COGs assigned to the microbiomes of all four sponge species were lumped together and subjected to SIMPER analysis against sediment and seawater metagenomes to rank COG entries contributing the most to differentiate between sponge (all species), sediment and seawater biotopes. All results deriving from these analyses are described in detail as supplementary material (**Appendix I- File S1**).

Nucleotide sequence accession numbers

Sponge CO1 sequences were deposited in the National Center for Biotechnology Information (NCBI) under the accession numbers KX574847 to KX574851. All metagenomes are accessible through the MG-RAST (project ID:13419_021215RCmetagenomes) and EMG platforms (project #ERP012972), and were deposited in the European Nucleotide Archive (ENA) under the accession numbers ERR1103453 to ERR1103462.

Results

Sponge identification

Sponge specimens were identified as *Spongia officinalis* (Linnaeus, 1759) based on macro- and microscopic morphology coupled with phylogenetic inference of the subunit 1 of the mitochondrial cytochrome oxidase (CO1) gene. Analysis of CO1 diversity (**Appendix I- Figure S1**) revealed 100% homology between the nucleotide sequences of our specimens and the Mediterranean ("MEDIT") *S. officinalis* haplotype (GenBank accession no. HQ830362) as defined elsewhere (Dailianis et al., 2011).

Microbial metagenomes - dataset overview

About 15 million paired-end reads (including forward and reverse reads) of 100 nucleotides in length were generated per sample, totaling 15.25 Gb of sequencing information (**Appendix I- Table S1**). Quality filtering and length trimming of reads using the EBI metagenomics pipeline (EMG) resulted in 103,104,001 high-quality reads (averaging 10,310,400 reads per sample) effectively used in downstream analyses (**Appendix I- Table S1**). Overall, 20 - 22% of the reads per sample could be assigned a function (i.e. IPR category) after ORF prediction and annotation with EMG, resulting in 22,156,186 annotated coding sequences (CDs) across the data, which constituted the functional analytical dataset. The number of annotated CDs per sample ranged from 1,808,840 to 2,446,913 reads (**Appendix I- Table S1**) totaling 10,272 IPR domains detected. The taxonomic analytical dataset consisted of 53,551 prokaryotic 16S rRNA gene reads identified from the data using the RNA Selector tool coupled to QIIME-driven operational taxonomic units (OTUs) picking and taxonomic assignment. 16S rRNA gene reads were assigned to 784 operational taxonomic Units (OTUs) in total. Details pertaining to COG annotations performed with MG-RAST can be found in **Appendix I**.

Functional and taxonomic ordination

Principal Coordinates Analysis (PCoA) performed on Bray-Curtis dissimilarity matrices calculated from normalized data revealed that sediments, seawater and *S. officinalis* harbor highly divergent microbial communities at the finest functional (IPR entries) and taxonomic (16S rRNA gene OTUs) levels of resolution (**Figure 2-1**). Sponge and seawater microbial communities presented the highest levels of divergence at both the functional and taxonomic levels, whereas sponge and sediment microbiomes shared the highest extent of functional (IPR) equivalence (**Table 2-1**). Between-biotope community distances were significantly higher than within-biotope distances in all possible combinations (**Table 2-1**) corroborating the consistent trends obtained by ordination analysis (**Figure 2-1**). Highly divergent functional profiles from sediment, seawater and *S. officinalis* microbiomes could as well be depicted using COG annotations of assembled and unassembled data (**Appendix I**). However, the significantly closer similarity between sponges and sediments observed with IPR functional profiling (**Table 2-1**) could not be re-verified employing COG annotations with MG-RAST (see **Appendix I** for details).

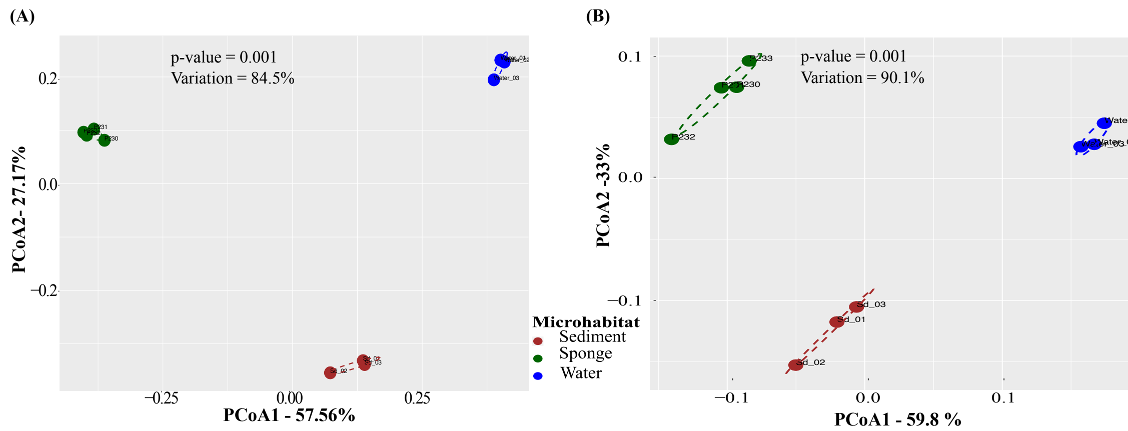


Figure 2-1. Principal Coordinate Analysis (PCoA) of taxonomic (**A**) and functional (**B**) microbial community profiles across biotopes. Community ordinations were based on pairwise Bray-Curtis dissimilarities (Table 2-1) calculated from normalized data, considering oscillations of relative OTU and IPR abundances among samples. Analyses were performed on OTU and IPR community profiles extracted from the 10 metagenomes using the EBI metagenomics (EMG) pipeline. Values on axes denote the extent of variation explained by each principal coordinate, whereas the total variation explained in the ordination space is indicated in the inlet. Significance values result from permutational analysis of variance (PERMANOVA) applied to the corresponding dissimilarity matrices.

Table 2-1. Functional (IPR) and taxonomic (OTU) community dissimilarities calculated between- and within-biotope samples. Shown are average Bray-Curtis dissimilarity values ± standard deviations. Within each row, values tagged with different letters are significantly different ($p < 0.05$) according to One-Way ANOVA, except for the OTU-based comparison between biotopes where non-parametric ANOVA on Ranks was used.

Between	**sponge vs. seawater**	**sponge vs. sediment**	**sediment vs. seawater**
IPRs	0.280 ± 0.020 ^{a}	0.222 ± 0.016 ^{b}	0.252 ± 0.028 ^{c}
OTUs	0.827 ± 0.015 ^{a}	0.718 ± 0.017 ^{b}	0.664 ± 0.030 ^{b}
Within	**sponge**	**seawater**	**sediment**
IPRs	0.091 ± 0.027 ^{a}	0.043 ± 0.002 ^{b}	0.076 ± 0.031^{ab}
OTUs	0.320 ± 0.027 ^{a}	0.196 ± 0.017 ^{b}	0.326 ± 0.047 ^{a}

16S rRNA gene taxonomic profiling

OTUs established at 97% 16S rRNA gene similarity were fetched with the EMG pipeline (see Materials and Methods for details) and used in taxonomic profiling. The *S. officinalis* symbiotic consortium was characterized by a relatively even distribution of diverse and dominant bacterial phyla, namely *Proteobacteria*, *Bacteroidetes*, *Poribacteria*, *Chloroflexi*, *Actinobacteria*, *Acidobacteria* and G*emmatimonadetes*, with 33 prokaryotic phyla (one archaeal, 32 bacterial) being detected across all sponge individuals (**Appendix I- Table S2A**). In contrast, *Proteobacteria* and *Bacteroidetes* dominated the seawater microbiome, followed by *Cyanobacteria* and phytoplankton. The former were also the most abundant phyla in sediments, along with an enormous variety of less abundant groups among which *Planctomycetes*, *Crenarchaeota*, *Actinobacteria*, *Acidobacteria* and *Verrucomicrobia* prevailed. All above-mentioned phyla significantly contributed to data variation among the three inspected biotopes (**Figure 2-2**) ($p < 0.05$, **Appendix I- Table S3A**).

We detected 293, 607 and 341 16S rRNA gene OTUs in sponges, sediments and seawater, respectively (**Appendix I- Table S2B**). Corresponding to 63.8% of all 16S rRNA gene reads from sponges, the ten most abundant OTUs from *S. officinalis* were, without exception, remarkably enriched in the sponge host, showing much lower abundances in the environmental vicinities (**Appendix I- Table S3**). Noteworthy in this regard was OTU 399 belonging to the canonical sponge-enriched phylum *Poribacteria*. It dominated the *S. officinalis* microbiome accounting for 11% of all 16S rRNA genes retrieved from this source, ranking as the second OTU contributing the most to the total phylogenetic divergence computed in the taxonomic dataset (**Appendix I- Table S3B**). The 25 most dominant *S. officinalis* OTUs comprised 86.9% of all sponge-associated 16S rRNA reads. These OTUs encompassed a cocktail of as-yet uncultivable phylotypes in the dominant phyla mentioned above, besides less-abundant lineages belonging to *Nitrospirae* and the candidate groups PAUC39f, SBR1093 and AncK6. All these OTUs could be considered typical *S. officinalis* endosymbionts not only because of their high abundance but also sharp enrichment in numbers within the host in comparison with the environment (**Figure 2-3**) (**Appendix I- Tables S2** and **S3).** Remarkably, this highly selected group of symbionts consistently displayed, with only a few exceptions, greater residual abundances in sediments than in seawater (**Appendix I- Table S3**).

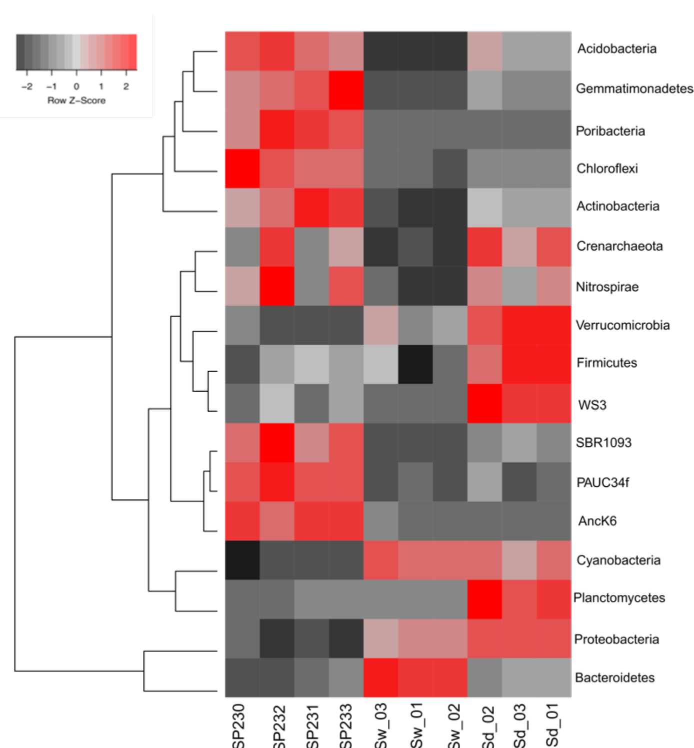


Figure 2-2. Heat map of the most differentiating microbial phyla across biotopes based on OTU data. Shown are the 17 phyla whose (OTU) relative abundances were found to oscillate the most among biotopes, explaining > 85% of the variation in phylum distributions. The dendrogram clusters phylum entries according to their abundance distributions across biotopes, labeled at the bottom of the diagram. Red squares show higher relative abundance values than the mean, whereas grey squares show lower relative abundance values than the mean. Within each phylum, color intensities are determined as a linear function of the Z-score calculated for each phylum abundance value as the subtraction of that value by the mean divided by the standard deviation around that mean (Z=(x-mean)/sd). SP230-SP233, sponge microbial metagenomes; Sd, sediment metagenomes; Sw, seawater metagenomes.

Particularly, OTUs 40 and 37, representing uncultured lineages in the *Acidimicrobiales* (*Actinobacteria*) and Sva0725 (*Acidobacteria*) clades, ranked among the top-25 most abundant OTUs of the complex sediment communities (**Appendix I- Table S3**). In addition, the three most dominant *S. officinalis* gammaproteobacterial symbionts (OTUs 621- order *Chromatiales*, 690 - order *Thiotrichales* and 639 - order HTCC2188) displayed equivalent or even higher abundances in sediments (**Appendix I- Table S3**). Conversely, the very dominant OTUs in seawater, essentially representing a mix of *Flavobacteriia*, *Alphaproteobacteria* and *Gamaproteobacteria* phylotypes, were all markedly de-selected in the sponge host except for OTU 442 (uncultured *Rhodobacteraceae*, the second most abundant seawater phylotype), which was the 5th and 14th most abundant OTU in sediments and sponges, respectively (**Appendix I- Table S3**).

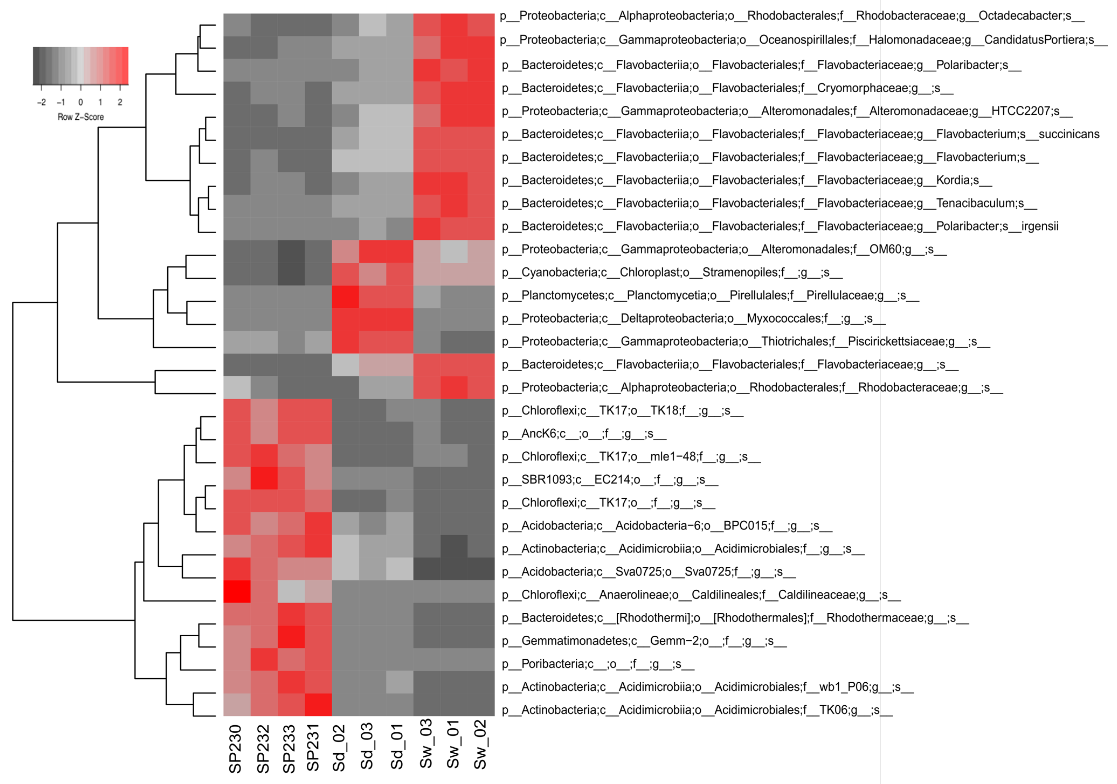


Figure 2-3. Heat map of the most differentiating OTUs across biotopes. Shown are the 31OTUs (97% cut-off) found to oscillate the most among biotopes, explaining > 32 % of the variation in the OTU dataset. Heat map details are as in legend to Figure 2-2.

IPR Functional profiling

From the 10,272 IPRs detected throughout the functional dataset using the EMG data processing pipeline (see Materials and Methods for details), 6046 were present in all biotopes, whereas 234, 695, and 1130 were specific to *S. officinalis*, seawater and sediments, respectively. However, 8325 IPRs displayed significantly different ($p < 0.05$) abundance values (normalized data) among at least two biotopes (**Appendix I- Table S4**), further substantiating the disparate functional assembly among the studied microbiomes (**Figure 2-1**). Due to the high complexity of the functional profiles and the thousands of IPR entries found to vary among biotopes, we used SIMPER analysis to rank those IPRs contributing the most to the total dataset variation (**Appendix I- Table S4**). A heat map of the 44 IPR entries varying the most across the biotopes, found to explain > 5% of (normalized) IPR abundance oscillations altogether, is shown (**Figure 2-4**). This group comprised several IPR entries contrasting the ecological and evolutionary contexts of the surveyed biotopes. Several functional traits strongly selected in the *S. officinalis* microbiome could be pinpointed, the majority of which

showing higher residual abundances in sediments than in seawater. These included a series of eukaryotic-like protein (ELPs) repeats (namely, WD40, leucine-rich, tetratricopeptide and ankyrin repeats, in this order) which remarkably populated the top-oscillating IPRs list (**Figure 2-4** and **Appendix I- Table S4**), along with luciferase-like, TolB-like beta propeller, ABC-transporter type 1, several transposases, and cytochrome P450 and CoA transferase III domain entries, among others (**Figure 2-4**).

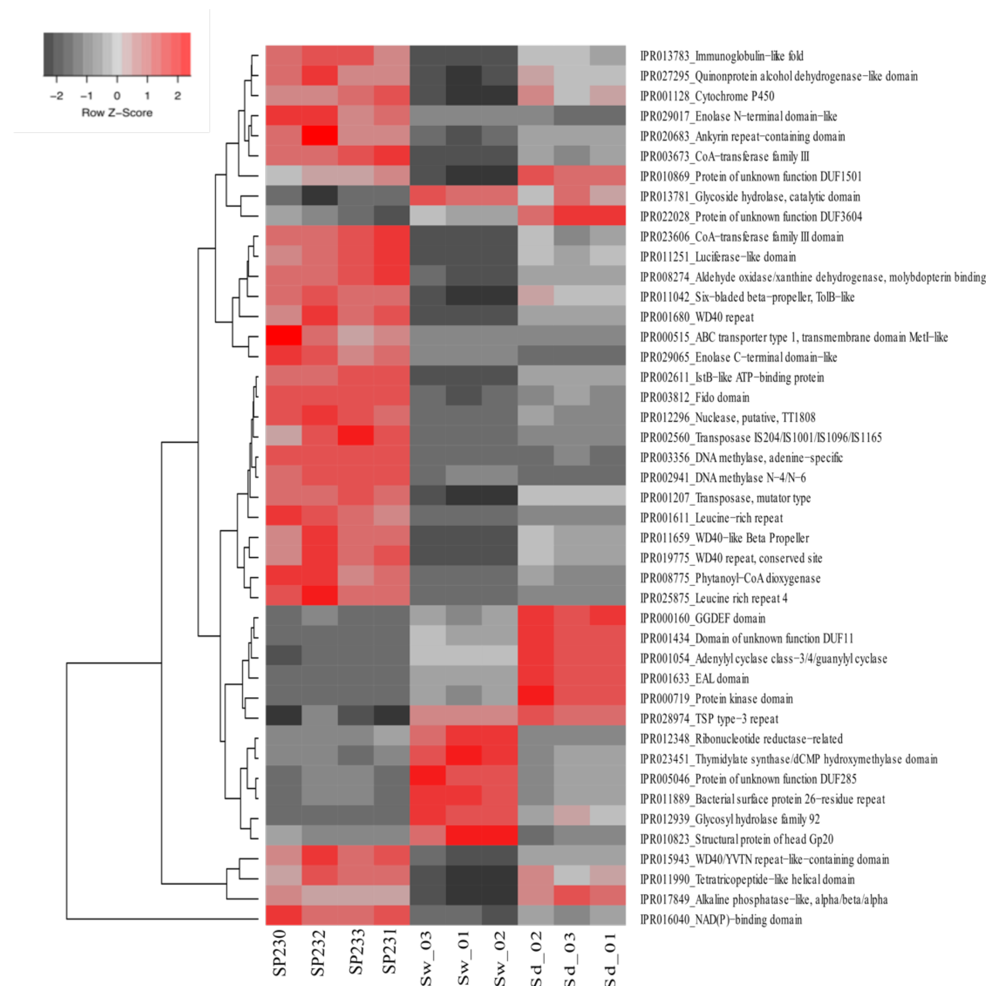


Figure 2-4. Heat map of the 44 most differentiating IPR entries across biotopes. The dendrogram clusters IPR entries according to their abundance distributions across biotopes, labeled at the bottom of the diagram. Heat map details are as in legend to Figure 2-2.

Worth mentioning among IPR entries more abundant in sediments were the GGDEF and EAL domains involved in synthesis and degradation of cyclic di-guanylate (c-di-GMP), known to regulate key cell physiology and life-style features such as motility, biofilm formation and virulence factors. Manual inspection of thousands of IPR entries contributing significantly to data variation (**Appendix I-Table S4**) allowed us to single out a number of “umbrella”

functions (each encompassing several IPR entries) presenting sharply different abundances among the biotopes (**Figure 2-5** and **Appendix I- Figure S2**). This approach clearly depicted the collective enrichment, in the *S. officinalis* microbiome, of all IPR entries classified into the above-mentioned ELPs (**Appendix I- Figure S2**), and those representing the coding of CRISPR-Cas, restriction endonucleases, plasmids, polyketide synthases, terpene/terpenoid synthases, Type IV secretion proteins and ABC transporters (**Figure 2-5**). Most of the observed sponge-enriched functional attributes showed, to varying degrees, significantly higher abundances in sediments than in seawater (**Figures 2-5** and **Appendix I- Figure S2**), except for the ABC transporters category, which includes both import and export transporters, and the restriction endonucleases category, more abundant in seawater than in sediments. Particularly, we uncovered striking diversity of both CRISPR-Cas and restriction endonuclease CDs from the *S. officinalis* microbiome (42 and 50 IPR entries, respectively). Restriction endonuclease reads represented, collectively, about 0.19% of the total number of annotated reads from *S. officinalis*, exceeding the relative abundance of CRISPR-Cas elements (0.11%) in these samples. Highly abundant in both sponge and sediment metagenomes were sulfatases, involved in the utilization of organic sulfated compounds, whereas type II secretion proteins involved in virulence were pronouncedly enriched in sediments metagenomes (**Figure 2-5**). Finally, predicted proteins involved in motility and chemotaxis were much more prevalent in sediments and seawater than in *S. officinalis* (**Figure 2-5**). While gliding and fimbriae types of motility were abundant in seawater, flagellar motility traits were more abundant in sediments.

Functional conservation among Spongia officinalis and other sponge hosts

To verify the extent to which the microbial metagenome of *S. officinalis* resembles those of other sponge hosts regarding their functional attributes, we used MG-RAST to compare the COG profiles obtained in this study (using metagenome reads assembled with MetaVelvet - see **Appendix I**) with those retrieved by Fan et al. (2012) for *Rhopaloeides odorabile*, *Cymbastela concentrica* and *Cymbastela coralliophila*. In spite of the large geographical distance between sampling sites and of the different sampling, sequencing, and data processing methods utilized in both studies, ordination analysis revealed a gradient in COG functional profiles corresponding to the phylogenetic relatedness of the hosts, with *S. officinalis* and *R. odorabile* (order Dictyoceratida) being placed closer to one another in the ordination diagram and farther apart from *C. concentrica* and *C. coralliophila* (order Axinellida) (**Appendix I**). The functional profiles of marine sponges, when pooled into one major group, differed significantly from those of seawater and sediment microbiomes. A high degree of functional

conservation, at the COG-level, was observed among the sponge hosts, with 62.7% of all COGs listed being shared by the four species. Furthermore, sponges were collectively found to share more COGs in common with sediments than with seawater (**Appendix I- File S1**). SIMPER analysis (**Appendix I- Table S5**) revealed that several of the common, enriched sponge symbiont functions were re-verified to contribute sharply to distinguish sponge from seawater and sediment metagenomes as observed in the analysis of *S. officinalis* IPR profiles. Particularly relevant in this regard were restriction-modification systems (i.e. restriction endonucleases), site-specific and adenine-specific DNA methylases, ABC transporters and plasmid-maintenance systems (**Appendix I- Table S5**). As determined in the analysis of IPR profiles, sulfatases were abundant in both sponge and sediment metagenomes, whereas Type II secretion proteins were more abundant in sediments (**Appendix I- Table S5**).

All domains-all genes taxonomic profiling using MG-RAST

Within MG-RAST, we performed a taxonomic assessment, primarily at the domain level, taking all gene reads (and not only 16S rRNA gene reads) that could be taxonomically classified into account, enabling us to determine the distribution of major groups (i.e. domains and viruses) across the biotopes in a more comprehensive fashion. In all biotopes, bacteria were clearly the most dominant group, comprising over 95% of all classifiable gene reads (**Appendix I- Figure S3**). While archaea were less represented in seawater (*c.* 0.18% of classifiable reads) than in sponges (1.6 - 3.9%) and sediments (1.7 - 3.2%), eukaryotic reads were slightly more abundant in the former biotope (3.11 - 4.36% of classifiable reads) than in the latter (2.12 - 2.39% and 2.04 - 2.32% in sponges and sediments, respectively). In spite of their minor representativeness across the entire dataset in terms of read numbers, from among all analyzed groups, viruses were found to oscillate the most in relative abundance among biotopes, displaying up to 13-fold higher abundances in seawater than in sponge and sediment samples (**Appendix I- Figure S3**).

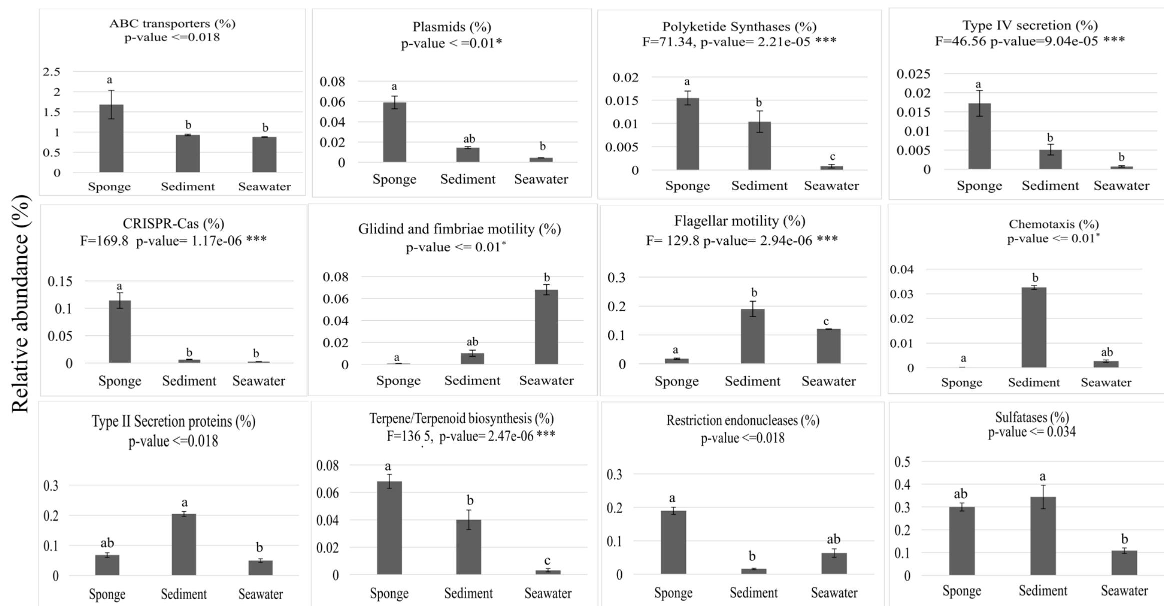


Figure 2-5. Abundance distributions of broad functional categories across biotopes. Values on the y-axis represent mean cumulative IPR relative abundances (%) in each biotope ± standard deviations. ABC transporters - 19 IPR entries used in plot construction; Plasmids - 10 IPR entries; Polyketide synthases - 1 IPR entry; Type IV secretion - 6 IPR entries; CRISPR-Cas - 43 IPR entries, Motility - 8 IPR entries involved in gliding and fimbriae-based motility; Flagellum, 56 IPR entries involved in flagellum assembly and motility; chemotaxis - 5 IPR entries; Type II secretion proteins -13 IPR entries; Terpene/Terpenoid biosynthesis - 3 IPR entries; Restriction endonucleases – 68 IPR entries; Sulfatases - 4 IPR entries. All IPR entries can be identified in Appendix I-Table S4. Results of the general test for differences among biotopes are shown at the top of each chart, below the label of each analyzed function. One-Way ANOVA with F statistics results are shown for normally distributed data, whereas ANOVA on Ranks results are shown for data distributions that did not pass normality tests. Bars labeled with different letters represent statistically distinct biotopes in terms of IPR relative abundances according to post-hoc pair-wise tests of significance.

Discussion

The taxonomic distinctiveness of the *S. officinalis* symbiotic consortium in comparison with those from its neighboring biotopes can be readily observed at the phylum level (**Figure 2-2, Appendix I- Table S2**). Consistent with primer-based studies undertaken for other keratose sponges off the Algarve coast (Hardoim et al., 2014; Hardoim and Costa, 2014b) and also from the Mediterranean Sea (Erwin et al., 2012a; Pita et al., 2013) and the Great Barrier Reef (Webster et al., 2010), this community is primarily made of a complex mix of so-far uncultivable, sponge-enriched symbiotic bacteria. Owing to our comparative experimental design, we gathered compelling evidence for higher sponge symbiont abundances in sediments

than in seawater, revealing an unexpected pattern of distribution of these microorganisms across marine biotopes and extending previous knowledge gained on their occurrence, at very low abundances, in the bacterioplankton (Webster et al., 2010; Webster and Taylor, 2012). Particularly, we identified one possible "generalist par excellence" bacterium in the *Rhodobacteraceae* clade (OTU 442) which, although clearly being a profuse member of seawater communities, likely performs well both in sediments and sponges. Further, the high prevalence of sponge-enriched *Acidobacteria* (especially Sva0725 phylotypes), *Actinobacteria* (*Acidimicrobiales* phylotypes) and *Gammaproteobacteria* (several different orders) in sediments adds further layers of complexity to our understanding of sponge symbiont occurrence in the marine realm. Future cultivation-independent, genome-wide studies targeting the adaptive features of these lineages not only hold promise in revealing their likely roles in the sponge endosymbiotic consortium, but may also improve our view of the genetic traits underpinning the persistence of sponge symbionts in the open environment, and consequently of the evolutionary and ecological forces that mediate the dispersal and community assembly of marine sponge symbionts. However, specific studies aiming at uncovering the potential metabolism, linking identity and function, of foundational sponge-associated bacteria are still relatively scarce. SCG and cultivation-independent genome binning from metagenomes have been proven useful in this regard, unveiling e.g. halogenation capacities within sponge-associated *Chloroflexi*, *Actinobacteria* and *Poribacteria* spp.(Bayer et al., 2013), non-ribosomal peptide biosynthesis potential within the *Chloroflexi* (Siegl and Hentschel, 2010) and multiple adaptive features of the keystone sponge-associate cyanobacterium *Synechococcus spongiarium* (Gao et al., 2014; Burgsdorf et al., 2015). Recently, the ability of several, so-far uncultivable sponge symbionts to utilize carnitine, a quaternary ammonium compound regularly present in the mesohyl matrix of sponges, has been revealed, suggesting parallel adaptation of multiple lineages to a common resource within the *in-spongia* microniche (Slaby et al., 2017). Our taxon-independent, primer-less sequencing approach revealed a pronounced dominance of one *Poribacteria* OTU in *S. officinalis*. It is therefore reasonable to argue that some of the potential metabolic features recently revealed for poribacterial symbionts by means of SCG (Siegl and Hentschel, 2010; Siegl et al., 2011; Kamke et al., 2013) are likely to mediate major bioprocesses and molecular interactions within the *S. officinalis* endosymbiotic consortium. These features include, among others, polyketide biosynthesis capacities (possibly involved in host's chemical defense), a vast, specialized carbohydrate degradation repertoire (considered pivotal to host's nutrient provision), and enrichment of eukaryotic-like repeat proteins (e.g. TRPs, ANKs, LRRs, usually considered to enable

symbionts to evade phagocytosis by the host) all of which could be verified, from our community functional profiles, as characteristic of the *S. officinalis* microbial metagenome. Because these attributes have been commonly verified in diverse sponge symbiont lineages (Slaby et al., 2017), it can be argued that they contribute significantly to the observed difference in taxonomic assembly between sponges, sediments and seawater observed here.

One important finding in this study was the observation that several of the features identified as genomic signatures of the *S. officinalis* microbiome displayed higher abundances in sediments than in seawater. Among these traits we highlight IPR entries underlying the coding of an array of eukaryotic-like proteins (ELPs) or involved in plasmid assembly, stability and conjugative transfer (e.g. plasmid replication, toxin-antitoxin systems and type IV secretion IPRs), secondary/cytotoxic metabolite biosynthesis (e.g. polyketide and terpene/terpenoid synthases, TolB-like and cytochrome P450 IPRs), remediation of oxidative stress (a luciferase-like domain), organic carbon utilization (e.g. sulfatases) and literally hundreds of other individual IPR entries (**Figure 2-5, Appendix I- Figure S2** and **Table S4**). Therefore, some of the features previously regarded as “unique” adaptations of the sponge symbiotic consortium may be well represented in other marine settings. Below, we give emphasis to the abovementioned functions and discuss their patterns of occurrence across bacterial genomes and the marine biotopes studied here.

Inspection of the IPR entries contributing the most to variation in the functional dataset (**Figure 2-5**) revealed the consistent prevalence of ELPs (TRPs, ANKs, LRRs and WD40) among the most sensitive IPRs differentiating the studied biotopes, all of which were enriched in *S. officinalis* (**Appendix I- Figure S2**). The abundance of TRPs and ANKs in sponge microbiomes has been well documented (Thomas et al., 2010; Fan et al., 2012), and a role for these ELPs in preventing phagocytosis of bacterial symbionts by the sponge host has been proposed (Nguyen et al., 2014; Reynolds and Thomas, 2016). In the present study, contrary to previous reports addressing other sponge hosts (Thomas et al., 2010; Fan et al., 2012), WD40 repeats were by far the most abundant ELPs in the *S. officinalis* microbiome, with several entries varying markedly in abundance across the surveyed biotopes (**Figure 2-4**, **Appendix I- Table S4**). WD40 repeats are regarded as prevalent in eukaryotes and uncommon in prokaryotes, and act as a protein-protein or protein-DNA platforms to allow for various protein complex assemblies in cellular metabolism (Xu and Min, 2011; Wang et al., 2015). However, evidence from this study and elsewhere (Díez-Vives et al., 2016; Reynolds and Thomas, 2016) is now accumulating for a broad distribution of these motifs among sponge symbiotic bacteria, offering a new angle from which the spread of these macromolecule network hubs can be seen

throughout the tree of life. Collectively, the presence of ELPs in prokaryotic genomes has been interpreted as suggestive of lateral host-microbe gene transfer given their presumed eukaryotic origin (Horn et al., 2016). Recently, ELPs were shown to be positively expressed in sponge microbial metatranscriptomes (Díez-Vives et al., 2016), supporting their likely importance in mediating cell-cell interactions within the sponge holobiont. In particular, the expression of WD40 repeats was found to be associated with domains of the Tol-dependent translocation system, which is involved in outer membrane integrity, cell invasion and, eventually, pathogenesis of Gram-negative bacteria, suggesting a pivotal role of these repeats in host-microbe interactions (Díez-Vives et al., 2016). The enriched abundance of both WD-40 repeats and one TolB-like domain (IPR011042) in the *S. officinalis* microbial metagenome speaks for distinguishing host colonization capacities and/or virulence potential within this symbiotic consortium.

Polyketides have been intensively studied as sponge-derived natural products whose biosynthesis is primarily mediated by bacteria (Piel, 2002; Piel et al., 2004; Wilson et al., 2014), and are thought to play a role in defense of the sponge host against natural enemies, as demonstrated for the bryozoan host *Bugula neritina* (Lopanik et al., 2004). Terpenes and terpenoids, in their turn, encompass a large class of natural products commonly regarded as of fungal and plant origin whose biosynthesis by bacteria is attracting increasing research interest (Yamada et al., 2015). The most abundant IPR entry related with terpene/terpenoid biosynthesis in *S. officinalis* (IPR008930) corresponds to a family of terpenoid cyclases/protein prenyltransferases responsible for a wide chemodiversity of terpenoid natural products (Christianson, 2017). Considering the broad distribution of terpene/terpenoid systhase genes across bacterial genomes (Yamada et al., 2015), it is tempting to argue that terpenoid biosynthesis in *S. officinalis*, and marine sponges in general, could be as well mediated by bacterial symbionts, emerging as a further mechanism possibly conferring host defense against natural enemies or mediating microbe-microbe interactions within the sponge host. Likewise, cytochrome P450 enzymes (IPR001128) are a superfamily of monooxygenases presenting broad substrate spectrum, being widespread in all domains of life. Particularly in bacteria, they are important in the biosynthesis of secondary metabolites such as erythromycin, and bear potential for applications in synthetic biology and the pharmaceutical industry (Girvan and Munro, 2016). Taken together, these observations suggest high microbially-driven chemical complexity within the *Spongia officinalis* holobiont. Such a vast secondary metabolite repertoire may play pivotal roles in microbiome community assembly, host-symbiont signaling and host defense. Widely known for their key role in bioluminescence, bacterial luciferases are

flavin monooxygenases which incorporate or reduce molecular oxygen in redox reactions, and may have originally evolved as enzymes responsible for reactive oxygen species (ROS) detoxification (Szpilewska et al., 2003). The sharp enrichment of this trait in the *S. officinalis* microbiome, followed by sediments, leads us to posit that these enzymes may primarily act as anti-oxidant agents in these particular settings, along with other anti-oxidant enzymes known to be enriched in sponges such as glutathione peroxidases (Thomas et al., 2010), observed here to possess high abundance in both sponges and sediments (**Appendix I- Table S4**, IPR IPR000889). Enrichment in sulfatases/aryl sulfatases have been suggested as a specialization of marine sponge symbionts enabling them to utilize sulfated polysaccharides from the host's extracellular matrix (Slaby et al., 2017). Sulfatase-encoding genes were abundant not only in the *S. officinalis* (**Appendix I- Table S4, Figure 2-5**) metagenome, but ranked as one major genetic signature of several sponge-associated microbiomes (**Appendix I- Table S5**). Here, we reveal that this trait is equivalently enriched in both sponge and sediment biotopes in comparison with seawater, providing evidence for the common selection of fundamental nutrient acquisition capacities in phylogenetically contrasting microbiomes. In the context of the marine sponge holobiont, sulfatases are supposed to be involved in nutritional exchange between host and microbes, playing a vital role in the cycling of sulfur within the animal.

Altogether, the outcomes delineated above indicate closer resemblance in functional attributes between sponges and sediments than sponges and seawater: a hypothesis corroborated by Bray-Curtis dissimilarity measures calculated for the three biotopes based on the whole array of 10,272 IPR entries uncovered from the data (**Table 2-1**). However, the quantitative trend revealed with whole functional profiles must be considered with caution since statistical significance varied depending on data processing methodology and on the reference database employed (**Appendix I**). Importantly, the level of phylogenetic disparity between all microbiomes was high (**Table 2-1**) in spite of our observation for higher residual symbiont abundances in sediments than in seawater (see above). Therefore, it is likely that surface sediments and endosymbiotic sponge communities, although being chiefly composed by different microbial populations (especially regarding their very dominant members) display a certain degree of independent functional convergence. This prompts us to argue that selective pressures common to particle- and host-associated modes of living constitute an important evolutionary force shaping functional assembly in marine biomes. Several factors, ranging from cell-cell interactions to availability (in quality and quantity) of solid surfaces for cell attachment to modes of symbiont acquisition and release by sponges, may contribute to the observed trends. Microbial cell densities alone, known to be about three orders of magnitude

higher in coastal sediments (Schmidt et al., 1998) and in Dictyoceratida sponges (Hardoim et al., 2012) than in seawater, may be a key factor promoting genetic exchange and adaptive features likely to prevail in the former biotopes. In highly dense circumstances, gene clusters involved in the biosynthesis of natural products such as polyketides and terpenes - or more specifically terpene-quinones very often enriched in Dictyoceratida species such as *S. officinalis* (Gordaliza, 2010; Manzo et al., 2011; Li et al., 2017) - are likely to confer selective advantage to its carriers. Similarly, strategies to neutralize cytotoxic effects are likely to elicit a selective advantage in communities where inhibitory compounds abound. ABC transporters (**Figure 2-5**) comprise a large family of bacterial trans-membrane proteins mediating the import and export of small and large-sized molecules throughout the cell, and may play a fundamental role as detoxifying agents permitting microbial survival in competitive microniches. Particularly, we found permeases within the ABC transporter category (e.g. IPRs 001851, 0038381 and 025966, **Appendix I- Table S4**) with presumed, manifold detoxifying functions commonly abundant in sponges and sediments. Polyketide synthases, type IV secretion and ABC transporter-encoding genes have all been detected on plasmids from several microorganisms (Stinear et al., 2004; Kadlec and Schwarz, 2009; Bruto et al., 2017). The prevalence of these genes along with the higher incidence of plasmid, transposase and ELP-encoding genes (which by themselves speak for greater genetic exchange potential) in the *S. officinalis* microbiome, followed by sediments, hints at a possible convergent selection of these traits in phylogenetically divergent microbial communities. Future studies aiming to define the gene content of the community of circular plasmids present in marine sponges will certainly shed new light on the functional features more likely to traffic about in the mobile gene pool within the Porifera.

Physical connectivity between sponges and sediments, although usually given less importance in microbiology studies than seawater intake via filtering, takes place by the capture of particulate organic matter and particles in suspension by the sponge host (Schönberg, 2016). In addition, loss of sponge cells through shedding and sponge-expelled detritus, both found to be significant processes in sponge cell turnover (Alexander et al., 2014), may act as substantial inputs of sponge-associated microorganisms into superficial sediment layers. Thus, marine sediments may serve as both sources and sinks of sponge-associated microorganisms, but the magnitude and relevance of this exchange remains to be addressed. The moderate abundance of a few dominant sponge symbionts in sediments indicates that these bacterial lineages, if not optimal performers, are capable of persisting - for undetermined periods - at considerable densities in this alternative habitat, thereby enhancing their probability of future lateral

acquisition by the sponge host. Identifying an active role beyond environmental endurance for these lineages in the complex microbiome of marine sediments is challenging. It is known that several factors such as seawater temperature and the composition and age of biofilm and biofouling communities are decisive for invertebrate larval settlement in benthic ecosystems (Hadfield, 2011; Whalan and Webster, 2014). Therefore, it could be argued that increased inter-domain signaling between host larvae and a seeding community of competent sponge associates on particulate/hard substrate - or microbe-microbe signaling in such circumstances - may contribute to higher larval settlement rates in favorable microniches, promoting the selection, on the sea floor, of sponge symbiont lineages able to persist in the open environment.

Contrasting the trends discussed above, microbial genes involved in motility and chemotaxis were altogether more prevalent in seawater and sediment communities, and much less abundant in the sponge host. The ability to move and orchestrate movement in response to chemical cues and gradients are widely acknowledged as imperative mechanisms dictating the distribution of microorganisms in the oceans (Stocker and Seymour, 2012) and as quintessential features of host-associated bacteria (Wadhams and Armitage, 2004; Rawls et al., 2007). Here we show that the *S. officinalis* endosymbiotic consortium displays low abundance of genomic features involved in chemotaxis and flagellar, gliding and fimbriae-based motilities when compared to its surrounding environment, supporting the idea that loss of motility may be common among prevalently vertically transmitted symbionts (Bright and Bulgheresi, 2010). Or, alternatively, for symbionts whose mode of acquisition by the host is rather passive from the microbial standpoint. Particularly relevant in distinguishing sediments from sponges and seawater regarding the regulation of virulence and motility were the higher abundances of GGDEF and EAL protein domains and of Type II secretion proteins in sediments. The above-mentioned domains modulate the concentrations of cellular cyclic-di-GMP, a signaling molecule involved in the regulation of biofilm formation, virulence, motility and cell surface adhesiveness in Gram-negative bacteria (Argenio and Miller, 2004). Indeed, increased cellular c-di-GMP was found to promote Type II secretion activity in *Vibrio cholerae* (Beyhan et al., 2006). Therefore, signal transduction via c-di-GMP and its modulation appears to be a determining factor in shaping the virulome of marine sediments in a singular fashion. Considering the *S. officinalis* endosymbiotic consortium, it is likely that this community essentially consists of “sit-and-wait” performers regarding their nutrient foraging strategies, especially if it is assumed that filtering activity alone is responsible for the total import and distribution of organic carbon and energy into the host. Our dedicated sampling of the inner sponge body disregards the profuse and complex community of epibionts known to populate

the pinacoderm of keratose sponges, where photosynthetic cyanobacteria are favorably selected (Erwin et al., 2012b) and motility and chemotaxis traits may be relevant for colonization and biofouling processes. The consistent trend found here for a primarily heterotrophic, less-motile community of endosymbionts highlights the need of approaching distinct microniches within marine sponges for a better understanding of microbiome spatial distributions and dynamics in these hosts (Webster and Thomas, 2016).

Finally, we detected much higher incidence of CRISPR-Cas and restriction endonucleases in the *S. officinalis* microbiome than in seawater - in accordance with earlier metagenomics surveys (Thomas et al., 2010; Fan et al., 2012; Horn et al., 2016) - and sediments. Therefore, in the context of its two immediate environmental surroundings, the enrichment of both defense mechanisms can be indeed considered a true hallmark of the *S. officinalis* microbiome, and much likely of marine sponges in general. Much has been discussed on the diversity (Fan et al., 2012; Horn et al., 2016) and role of these genetic elements as an efficient, specific anti-phage defense system permitting bacterial survival within the sponge microbial consortium (Thomas et al., 2010; Fan et al., 2012; Horn et al., 2016). In agreement with this hypothesis, we here observed that the relative abundance of both defense systems and of bacteriophages were inversely correlated in *S. officinalis* and seawater, where viral particles were 13-fold more frequent than in sponges and CRISPR-Cas were virtually absent. However, low abundances of both defense systems and of viral DNA were detected in sediments, suggesting that viral populations might be regulated by other mechanisms in these settings rather than high abundance of CRISPR-Cas and R-M systems alone. In this regard, it was curious to note that the diversity and assemblage of restriction endonucleases uncovered from sponges and sediments was fairly comparable, with higher abundances in sponges being the primary factor distinguishing these biotopes (**Appendix I- Table S4**). Future efforts are therefore needed to disentangle the relative forces exerted by CRISPR-Cas and restriction-modification systems on the regulation of viral populations within the Porifera and across marine biomes. To this end, better understanding of the structure of phage communities in host- and particle-associated settings will be much required.

In conclusion, the comprehensive comparative metagenomics strategy employed in this study enabled us to critically assess the distribution of genomic features involved in symbiosis across marine habitats, and to address functional convergence versus divergence in contrasting marine microbial communities more thoroughly. We advocate that such an approach, which in the future shall include the assessment of other invertebrate hosts, is imperative for a holistic

understanding of microbial community dynamics and function in marine sponges and benthic ecosystems at large.

Supplementary Material

The supplementary materials and explanations for this chapter can be found in Appendix I and online at: https://www.frontiersin.org/articles/10.3389/fmicb.2017.02499/full#supplementary-material.

Acknowledgments

The authors thank Alireza Asvadi and Paulo Jorge Moura Pinto da Costa Dias for providing assistance in R coding. André R. Soares and Gianmaria Califano are acknowledged for their generous help during sample processing in the laboratory.

Funding

This work was funded by the Portuguese Foundation for Science and Technology (FCT) through the research grants PTDC/BIA-MIC/3865/2012 and PTDC/MAR-BIO/1547/2014, conceded to RC. EK was supported by a PhD scholarship awarded by the Education, Audiovisual and Culture Executive Agency (European Commission, Erasmus Mundus Programme, Grant EMA2 lot7/SALA1206422). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethics statement

This study relied on *in situ* sampling of microorganisms from marine invertebrates without a nervous system, and as such was exempt from ethical approval procedures according to the current Portuguese legislation (Decreto-Lei nº 113/2013). This study did not occur within privately owned or protected areas. This study did not involve endangered or protected species. The sampling methodology privileged minimally invasive handling procedures, following the guidelines of the European Directive 2010/63/EU.

Author Contributions

RC designed the study; EK, JMSG, JRX, and RC performed the experiments; MPR, JMSG, JX, and RC provided reagents and materials; EK, MR, and RC analyzed the data; EK and RC wrote the main manuscript text and prepared figures and tables. All authors reviewed the manuscript.

Chapter 3

Functional genomics of cultivated sponge-associated *Alphaproteobacteria* reveals shared and unique traits underlying a bimodal symbiotic-free-living life-style

Functional genomics of cultivated sponge-associated *Alphaproteobacteria* reveals shared and unique traits underlying a bimodal symbiotic-free-living life-style

Elham Karimi[1,2]; Tina Keller-Costa[3]; Beate M. Slaby[4]; Cymon J. Cox[2], Ulisses N. da Rocha[5]; Ute Hentschel[4,6], Rodrigo Costa[2,3]

[1]Faculty of Science and Technology, Algarve University, Gambelas 8005-139 Faro, Portugal.

[2]Centre of Marine Sciences, Algarve University, Gambelas 8005-139 Faro, Portugal.

[3]Institute for Bioengineering and Biosciences (iBB), Instituto Superior Técnico (IST), Universidade de Lisboa, 1049-001 Lisbon, Portugal.

[4]RD3 Marine Microbiology, GEOMAR Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany

[5]VU University of Amsterdam, Department of Molecular Cell and Physiology, Amsterdam, The Netherlands

[6]Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany.

A version of this chapter is being prepared to be submitted for publication in an international, peer-reviewed scientific journal.

Abstract

Marine sponges are filter-feeding, early-branched metazoans that usually host complex microbial communities comprised of several, so-far uncultivable symbiotic lineages. In this study, an alternative bacterial cultivation platform using a low-carbon marine medium, lower incubation temperature, and prolonged incubation time, was applied to sample the microbiome of the marine sponge *Spongia officinalis*. The approach led to the laboratory cultivation of diverse *Alphaproteobacteria* genera such as *Anderseniella*, *Erythrobacter*, *Labrenzia*, *Loktanella*, *Ruegeria*, *Sphingorhabdus*, *Tateyamaria*, *Pseudovibrio*, and two likely new genera of *Rhodobacteraceae*. Comparative genomics of representative strains of all above-mentioned genera revealed that the community of cultivable *Alphaproteobacteria* associated with *S. officinalis* could contribute to enhancing host fitness through detoxification mechanisms (e.g. heavy metal and metabolic waste removal, degradation of aromatic and halogenated compounds), provision of essential vitamins and inorganic ions, nutritional exchange (especially regarding the processing of organic sulfur and nitrogen) and chemical defense (through e.g. the biosynthesis of polyketides and terpenoids). COG-based genome annotation was employed to unveil patterns of functional convergence and divergence among the studied strains and genera, revealing three genome clusters which were *a posteriori* approached as two distinct functional groups: *Roseobacter* vs. non-*Roseobacter* genomes. We argue that representative species of the non-*Roseobacter* group were most likely to engage in closer symbiotic interaction with their sponge host than members of the *Roseobacter* group, since genomic features such as eukaryotic-like proteins-, adhesion proteins- and pili-encoding genes - usually regarded as “symbiosis factors” - were more prevalent in the former group. Particularly, the genus *Anderseniella* presented the most remarkable suite of traits underlying symbiotic behavior. Nevertheless, all the organisms inspected did not display signs of genome reduction usually considered indicative of obligate mutualism, and instead possessed highly versatile carbon, nitrogen, phosphorus and sulfur metabolisms underlying biphasic, host-associated / free-living life styles.

Introduction

Investigating sponge microbial symbionts is a fundamental part of todays' marine microbial ecology due to the ecological and biotechnological value which these dense and diverse microbial communities possess (Piel et al., 2004; Piel, 2009; Ebada et al., 2010; Schippers et al., 2012). So far, 52 bacterial phyla have been reported to inhabit sponges via cultivation-independent diversity surveys (Webster and Thomas, 2016) with *Proteobacteria* (mostly *Alpha*- and *Gamaproteobacteria*), *Acidobacteria*, *Actinobacteria*, *Chloroflexi*, *Nitrospirae*, *Cyanobacteria* and the candidate phylum *Poribacteria* being the most abundant (Thomas et al., 2016; Webster and Thomas, 2016). Sponge-associated bacteria engage in nutritional exchange with their hosts and as such are considered to play an important role in benthic biogeochemical cycling (Maldonado et al., 2012). Moreover, they are believed to produce most of the secondary metabolite repertoire of sponges (Piel, 2002; Piel et al., 2004; Fan et al., 2012; Hentschel et al., 2012; Gao et al., 2014; Tian et al., 2014; Wilson et al., 2014).

Alphaproteobacteria display great versatility in their association with multicellular organisms, with interactions ranging from mutualistic over commensal to parasitic and pathogenic (Garrity et al., 2005). Microbial diversity surveys investigating different sponge taxa from various geographic locations and seasons have noted *Alphaproteobacteria* as regular sponge associates (Webster and Hill, 2001; Enticknap et al., 2006; Cleary et al., 2013; Thomas et al., 2016). Particularly, the families *Rhodobacteraceae* and *Rhodospirillaceae* have been found to be dominant members of the marine sponge microbiome and to harbor a high variety of so-far uncultivable lineages likely to be specific to or enriched in these hosts (Simister et al., 2012; Karimi et al., 2017b; Karimi et al., 2018-in press). Genome reconstruction of uncultivated symbionts of the Mediterranean sponge *Aplysina aerophoba* via metagenomic binning revealed common anti-viral defense mechanisms and specialized nutrient acquisition pathways among diverse sponge-associated bacteria (Slaby et al., 2017). *Alphaproteobacteria*, for instance, were enriched in genes encoding for energy production and carnitine metabolism (Slaby et al., 2017). In a recent metagenomic-binning study, we assembled the genome of an uncultivated *Rhodospirillaceae* symbiont of *Spongia officinalis*, revealing taurine import and utilization, lack of motility and chemotaxis and enrichment in glutathione S-transferases as adaptive genomic signatures, among others, of a host-associated life-style within this family (Karimi et al., 2018-in press). Among cultivable (or so-far cultured) sponge-associated *Alphaproteobacteria*, members of the frequently cultivated genus *Pseudovibrio*, for example, are well equipped for a symbiotic lifestyle (Bondarev et al., 2013) and vertically transmitted

through the sponge larvae from parents to their next generation (Enticknap et al., 2006). However, our understanding of the contribution of cultivable sponge-associated bacteria to host health and homeostasis remains hindered by the current lack of information on their densities within the marine sponge microbiome, especially in comparison with those of the dominant and so-far uncultivable symbionts.

Indeed, for decades scientists have tried to cultivate marine sponge-associated bacteria. However, the taxonomic and functional diversity of sponge-derived culture collections is still limited, with 1% to 14% of the total sponge bacterial community estimated to be cultivable using different methods (Webster and Hill, 2001; Olson and McCarthy, 2005; Sipkema et al., 2011; Hardoim and Costa, 2014b). Yet the most abundant bacterial symbionts of sponges in particular remain uncultivated (Taylor et al., 2007a; Hardoim et al., 2014). One complication is that the cultivability of these bacteria is influenced not only by the cultivation method and media, but also by the initial sample processing method (Esteves et al., 2016). To overcome the difficulty in cultivating marine sponge symbionts, alternative strategies have been developed and shown promising results. These include the *in-situ* implantation of nutrient medium-containing diffusion growth chambers (DGCs; (Kaeberlein et al., 2002)) into sponge specimens and their subsequent incubation in the field (Steinert et al., 2014), and the concomitant use of several solid or liquid media (with and without antibiotics) to increase the phylogenetic breadth of the symbiotic bacteria captured in the laboratory (Sipkema et al., 2011; Versluis et al., 2017). In spite of these advances, more attempts to cultivate the "uncultivable" are needed if we are to achieve a comprehensive functional exploration and exploitation of the marine sponge microbiome. Cultured representatives allow full sequencing and precise annotation of bacterial genomes, providing more accurate data than ecogenomic techniques and supporting the analysis and interpretation of environmental sequence data (Rappé, 2013; Gutleben et al., 2017). The effective combination of information from both, cultivation-dependent genomics and cultivation-independent metagenomics can deepen our understanding of microbial community functioning (Gutleben et al., 2017). Cultivated symbionts allow genetic manipulations and physiological, phenotypic characterizations that are essential to assign new proteins and to understand complex metabolic pathways (Gomez-Escribano and Bibb, 2011).

In this study, we hypothesized that changing the solidifying agent agar which may inhibit the growth of certain bacterial taxa (Janssen et al., 2002; Tamaki et al., 2009) to the nontoxic agent gellan gum, combined with a low carbon content in the culture medium and a lower (19°C) incubation temperature with prolonged incubation (8 weeks), could lead to the

isolation of slow-growing, novel sponge-bacterial symbionts. Our culture conditions favored the cultivation of taxonomically diverse *Alphaproteobacteria* strains from *S. officinalis*, which prompted us to investigate the functional features of ten distinct *Alphaproteobacteria* genera, five of which belonging to the *Roseobacter* clade in the family *Rhodobacteraceae*, in more detail. Here, we define the core genomic functions of this *Alphaproteobacteria* consortium, highlight their taxon- and group-specific (i.e. *Roseobacters vs.* "non-*Roseobacters*") particularities and tentatively predict the possible collective role of *Alphaproteobacteria* associated with *S. officinalis.* To assess the relative abundance of these ten *Alphaproteobacteria* symbionts in the *S. officinalis* associated microbial community, we mapped the metagenomic reads of the sponge microbiome (Karimi et al., 2017b) to the ten selected genomes.

Material and methods

Sample collection, cultivation of bacteria and phylogenetic analysis

Four *Spongia officinalis* specimens (Alg230-Alg233, for details see Karimi et al. (2017b) were collected in May 2014 by SCUBA diving at 20 m depth off the coast of Pedra da Greta (36º 58' 47.2N ;7º 59' 20.8W), southern Atlantic Ocean, Portugal, and transported to the laboratory within approximately 1 h in a cooling box. Specimens were processed immediately upon arrival: 2.5 g of the specimens' inner body were cut and macerated with a sterile mortar and pestle in 22.5 mL of calcium and magnesium-free artificial seawater (CMFASW) (for details see Hardoim et al. (2012); Esteves et al. (2013)). The resulting cell suspension was then serially diluted in CMFASW and 100 µL of 10^{-3} to 10^{-8} dilutions were spread on marine gellan gum medium (hereafter called 'MG50') plates in triplicates. The 'MG50' medium was prepared by diluting marine broth (MB; ROTH®) 50 times in artificial seawater (ASW: 23.38 gL^{-1} NaCl, 2.41 gL^{-1} $MgSO_4.7H_2O$, 1.90 gL^{-1} $MgCl_2.6H_2O$, 1.11 gL^{-1} $CaCl_2.2H_2O$, 0.75 gL^{-1} KCl, 0.17 gL^{-1} $NaHCO_3$, final MB-concentration: 0.802 gL^{-1}) and solidified with Phytagel™ (gellan gum; 5 gL^{-1}). All plates were incubated for eight weeks at 19°C. Bacterial growth was monitored weekly and colony forming units (CFUs) counted. Colonies were selected based on their variations in color and shape with the aim to isolate as many different bacterial morphotypes as possible rather than to randomly collect a high number of strains. Nevertheless, highly abundant morphotypes were picked more often to enable access to different bacterial lineages eventually sharing the same colony morphology (see **Appendix II-Table S1**).

Average CFU counts ranged from 3.0×10^6 (sponge Alg232) to 8.1×10^6 (sponge Alg230) CFUs g^{-1} sponge wet tissue weight. Sponge specimen Alg231 had $6.9\times10^6 \pm 0.00015\times10^6$ CFUs g^{-1} (mean±SE), and was chosen for colony isolation as it showed the greatest variety of morphologically distinct colonies. Here, we benefited from previous knowledge on the (equivalent) functional and taxonomic bacterial diversity present in each sponge specimen, acquired via shotgun metagenome sequencing (Karimi et al., 2017b), to calibrate our sampling effort to cover the total colony morphotype diversity within one specimen (higher morphotype sampling depth) rather than spreading the effort across several specimens, what would likely lead to the retrieval of the same and most abundant phylotypes from different specimens (lower morphotype sampling depth). In total, 48 (many of them morphologically unique) colonies were picked and streaked to purity on MG50 plates. The purified isolates were then grown for 48h in 1:2 diluted marine broth (‘MB2’) and stocked in fresh MB2 supplemented with 20% glycerol at -80°C. For genomic DNA extraction, 2 mL aliquots of the shaken MB2 cultures were centrifuged at 10,000 g for 30 min. Genomic DNA was extracted from the resulting cell pellets using the Wizard® Genomic DNA Purification Kit (Promega, Madison, USA) according to the manufacturer’s instructions. Genomic DNA samples of all isolates were then subjected to 16S rRNA gene amplification and Sanger sequencing for identification as previously described (Esteves et al., 2013). Closest matches to all sequence queries were identified using the BLAST algorithm (December 2016) of the national center for biotechnology information NCBI (Johnson et al., 2008). Taxonomic assignment of bacterial isolates to the genus level was performed using the classifier tool of the ribosomal database project (RDP, release 11, (Cole et al., 2009) as described earlier (Costa et al., 2013; Esteves et al., 2013). Closest 16S rRNA gene sequences from type strains were determined using the RDP sequence match tool. Operational taxonomic units (OTUs) at 100% sequence similarity were assigned by aligning all sequences using the ClustalW algorithm and by calculating a pairwise distance matrix in MEGA7 (Kumar et al., 2016). To construct a phylogenetic tree comprising all *Alphaproteobacteria* (most abundant class of the collection) isolates obtained in this study, and thus more precisely infer which isolates could represent novel bacterial taxa, the 16S rRNA gene sequences of closest matches observed in BLASTN searches and the respective closest *Alphaproteobacteria* type strains found in RDP were included in the alignment procedure. An appropriate evolutionary model was then determined using the ‘find best DNA models’ function of MEGA7. This was the Kimura 2-parameter model with a discrete gamma-distribution and invariable sites (*K2+G+I*). A Maximum Likelihood tree was then determined with bootstrap support using 100 repetitions.

Genome sequencing of sponge-associated **Alphaproteobacteria**

Genomic DNA samples of 10 phylogenetically distinct *Alphaproteobacteria* strains (representing all obtained *Alphaproteobacteria* genera) were sent for genome sequencing on an Illumina MiSeq platform at Mr. DNA (Shallowater, TX, USA). Paired-end libraries (2×301bp) were generated and the genomes were assembled *de novo* into contigs with the NGen DNA assembly software by DNAStar, Inc. as described previously (Karimi et al., 2017a). All contigs of each genome were subjected to a BLAST (NCBI) search via the computational cluster facility 'gyra' (http://gyra.ualg.pt) of the Algarve Centre of Marine Sciences (CCMAR). The extracted BLAST files were then analyzed in MEGAN5 (Huson et al., 2016) to confirm whether the taxonomic affiliation of each contig matched the (16S rRNA gene-based) affiliation of its respective source strain. Contigs found not to fall within the expected taxonomic affiliation of its respective strain and or with less than 1000 bp in length were discarded prior to annotation and downstream comparative analyses.

Annotation and comparative analysis of genomes

Open Reading Frame (ORF) prediction and annotation of the genome sequences were performed using the RAST (Rapid Annotation using Subsystem Technology) prokaryotic genome annotation server (version 2.0) with standard procedures (Aziz et al., 2008). In addition, all genomes were uploaded to the software platform EDGAR 2.0 (Blom et al., 2016) to define core- and pan- genomes of strains in different combinations, and to assess the number of singleton genes for each genome based on the coding sequences (CDSs) predicted using RAST. CDSs were also subjected to annotation based on Clusters of Orthologous Groups of Proteins (COGs) using the on-line server WebMGA (e-value = 0.001)(Wu et al., 2011). Unless otherwise stated, quantitative functional comparisons between the genomes were performed using COG annotations. To this end, the COG profile of each genome was Hellinger-transformed (i.e. square root calculation of the relative abundance of each COG entry in a given genome), after which one COGs *vs.* genomes contingency table merging the functional profiles of all ten genomes into one single file was generated using a customized script (mrg-cog.py)[1] and used as input to ordinate the genomes according to their (COG) functional profiles. This was achieved via principle components analysis (PCA) using the function 'PCA' of the FactoMineR package (Lê et al., 2008), with default parameters, within R version 3.2.4 (RCoreTeam, 2015).

[1] https://github.com/ElhamKarimi/Merge-files-COGs/blob/master/mrg-cog.py

Genomes were clustered into functional groups (essentially, "*Roseobacters*" *vs*. "non-*Roseobacters*") according to trends revealed, after PCA, on the extent of similarity/dissimilarity among them. Thereafter, pair-wise comparisons between functional groups were conducted using White's non-parametric t-test within STAMP v2.0.9 (Parks et al., 2014) to identify COG entries that are differently abundant (i.e. "enriched" or "depleted") between groups of genomes. Moreover, lists of COG entries shared by all genomes belonging to one group (e.g. "*Roseobacters*") and absent in all genomes of the other (e.g. "non-*Roseobacters*"), and vice-versa, were prepared to further explore the typical genomic traits of members of the *Roseobacter* clade in comparison with other alphaproteobacterial species retrieved from *S. officinalis*. Secondary metabolite gene clusters were predicted for each genome with antiSMASH-version3 (Weber et al., 2015).The CRISPRfinder online tool was used with default settings to detect and identify CRISPR repeats and spacer sequences for *Anderseniella* genome (Grissa et al., 2007).

Representativeness of cultivated sponge-associated* Alphaproteobacteria *across marine biotopes

Coverage variations of each *Alphaproteobacteria* genome were inspected by mapping the already available microbial metagenomes from *S. officinalis* (four specimens), surrounding seawater (three replicates) and sediments (three replicates) (Karimi et al., 2017b) against the assembled genome of each bacterium. To this end, the sequencing reads from the replicate metagenome samples within each marine biotope mentioned above were pooled and thereafter aligned to each *Alphaproteobacteria* genome using bowtie2 v. 2.2.6 at default settings (Langmead and Salzberg, 2012). The alignment scores, displayed as proportions of reads in the metagenomes that could be aligned with each single genome, were used as comparative measures of relative abundance of the studied alphaproteobacterial strains across *S. officinalis*, sediments and seawater.

Nucleotide sequence accession numbers

The 16S rRNA gene sequences of the bacterial isolates were deposited at NCBI GenBank under the accession numbers KY363613-KY363636. Assembled genome sequences reported in this study were deposited at the European Nucleotide Archive - European Molecular Biology Laboratory (ENA-EMBL) under the study identification number PRJEB18465 (ERP020395). Genome accession numbers are shown in **Table 3-1**.

Results

Isolation and identification of* S. officinalis *associated bacteria

In total, 48 aerobic, heterotrophic bacterial isolates representing manifold colony morphologies were selected in this study for further genotypic characterization (**Appendix II- Table S1**), with 46 isolates belonging to the phylum *Proteobacteria* and two isolates to the phylum *Actinobacteria* (**Appendix II- Table S2**). Within the *Proteobacteria*, the vast majority of the isolates (41) affiliated with *Alphaproteobacteria* class, while the remainder (5 isolates) was classified as *Gammaproteobacteria*. Isolates in *Alphaproteobacteria* class encompassed three different orders: *Rhizobiales, Sphingomonadales* and *Rhodobacterales*, the latter comprising most isolates (38 strains) (**Figure 3-1, Appendix II- Tables S1** and **S2**). Altogether, twelve formally-recognized bacterial genera and two phylotypes non-classifiable at the genus level were identified. Twenty-eight of *Rhodobacterales* isolates affiliated with the genus *Ruegeria* which was the most abundant genus of the collection and displayed a high degree of intra-generic diversity. Indeed, *Ruegeria* strains grouped into ten distinct OTUs (100% cut-off) across five different *Ruegeria* species (*R. arenilitoris, R. atlantica, R. conchae, R. halocynthiae* and *R. meonggei*). (**Figure 3-1**). Overall, 24 unique 16S rRNA gene OTUs (at 100% sequence similarity cut-off) were observed across the data (**Appendix II- Table S2**). Many of the closest NCBI BLASTn hits and/or type strains to these OTUs originated from various marine sponge species or other invertebrate hosts including corals, bivalves, ascidians, squid and sea urchins. 16S rRNA gene phylogeny revealed that most of the isolates reported in this study affiliated with two subgroups within the *Rhodobacterales* order, namely the *Roseobacter* group containing isolates classified as *Ruegeria, Loktanella, Tateyamaria* and *Rhodobacteraceae* spp. (two strains, see below), and the *Stappia* group containing isolates affiliated with the genera *Pseudovibrio* and *Labrenzia*.

Two isolates (Alg231-04 and Alg231-30) of the *Rhodobacteraceae* family were not classifiable at genus level (**Figure 3-1, Appendix II- Table S2**) and likely represent at least novel bacterial species. Closest type strains *Phaeobacter inhibens* T5 and *Thalassobius aestuarii* JC2049 shared 98% and 97.8% 16S rRNA gene sequence similarity with strains Alg231-04 and Alg231-30, respectively, but phylogenetic analysis showed that these *S. officinalis* isolates clustered separately from both their closest type strains and other *Phaeobacter* and *Thalassobius* representatives (**Appendix II- Figure S1**). In fact, strains Alg231-04 and Alg231-30 grouped with other unclassified *Rhodobacteraceae* and/or uncultivated strains (**Appendix II- Figure S1**), leaving their genus-level taxonomic affiliation

unresolved. In contrast, 16S rRNA gene sequences of *Ruegeria* sp. strain 231-54 and *Pseudovibrio* sp. strain 231-02, representing well studied, sponge-associated cultivable bacteria, shared 100% sequence similarity to their respective closest NCBI BLASTn hits (**Figure 3-1**) which were isolated from *S. officinalis* sampled in the Mediterranean Sea (Bauvais et al., 2015).

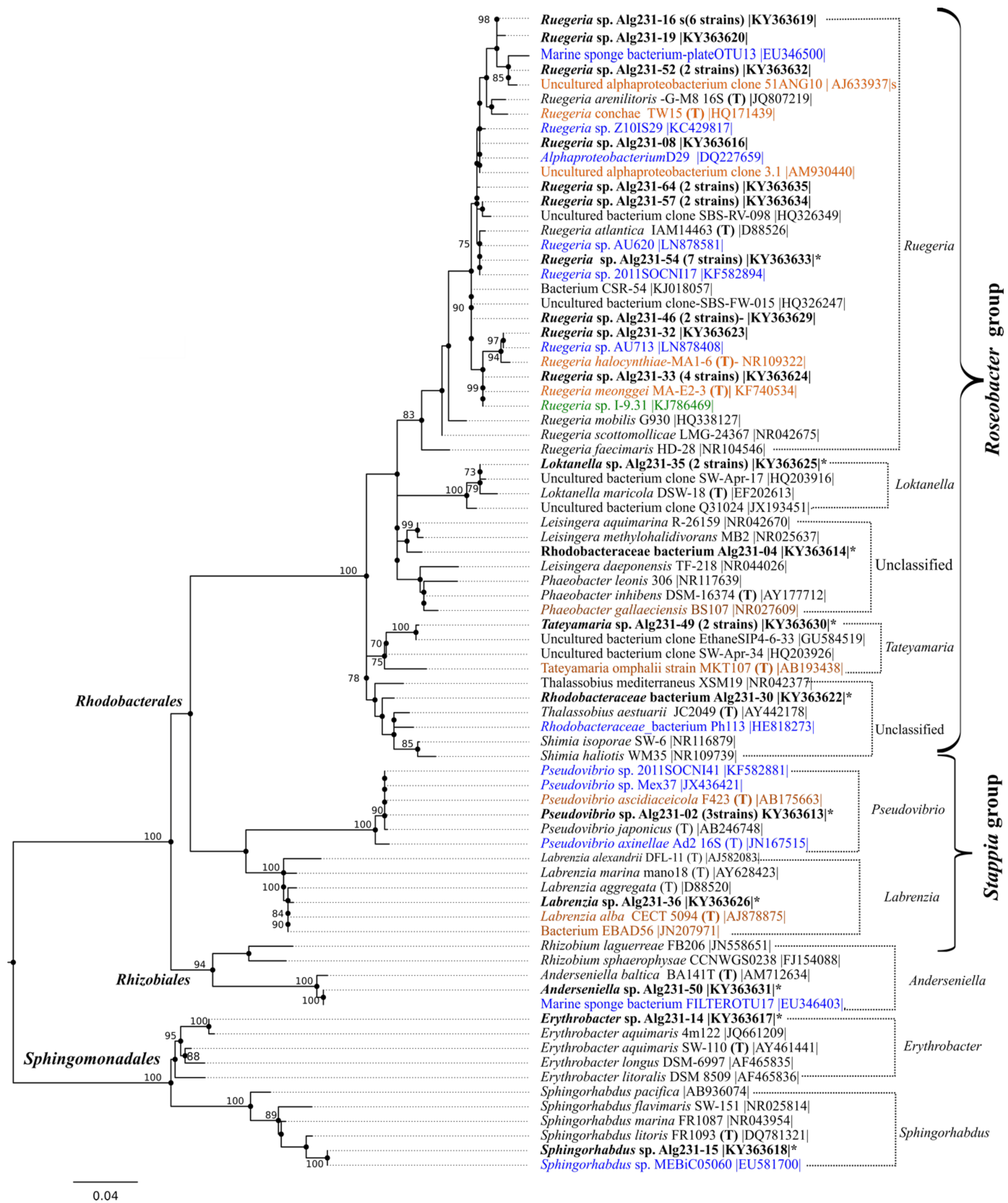


Figure 3-1. Maximum Likelihood tree of *Alphaproteobacteria* species based on Kimura 2-parameter evolutionary distances calculated for 16S rRNA gene sequences. *Alphaproteobacteria* strains isolated from *S. officinalis* are shown in bold. Numbers of isolates obtained from *S. officinalis* that belong to the same OTU (100% cut-off) are given in brackets. Closest NCBI BlastN hits and type strains ((T), in bold) for each strain are shown on the tree. Blue marks sponge-associated, orange marks invertebrate associated and green marks marine algae-associated closest NCBI BLASTN hits and type strains. Strains that had their genome sequenced are marked with an asterisk. Bootstrap values (100 repetitions) above 70% (0.7) are shown on the tree nodes.

General features of sponge-associated* Aphaproteobacteria *genomes

The genomes of ten *Alphaproteobacteria* isolates representing eight formally accepted genera in the *Rhodobacteraceae, Rhodobiaceae, Sphingomonadaceae* and *Erythrobacteraceae* families plus two phylotypes non-classified at the genus level (strains Alg231-04 and Alg231-30 in the *Rhodobacteraceae* family, see above) were fully sequenced and subjected to comparative analyses of phylogeny and function (**Table 3-1**). The size of the assembled alphaproteobacterial genomes ranged between 3.13Mb for *Erythrobacter* sp. Alg231-14 and 7.40Mb for *Labrenzia* sp. Alg231-36. G+C contents varied from 51.3% in *Pseudovibrio* sp. Alg231-02 to 59.6% in the unclassified *Rhodobacteraceae* strain Alg231-04 (**Table 3-1**). The number of coding sequences ranged from 3,139 to 5,120 and the number of RNA gene copies from 41 to 77 including 3 to 12 copies of ribosomal RNA (rRNA) genes (**Table 3-1**).

Core- and pan-genome analysis

To define the core- (i.e. the pool of genes that are common to all analyzed genomes) and pan-genome (the sum of all genes in all analyzed genomes) of the *S. officinalis* associated *Alphaproteobacteria* fully sequenced in this study, the genome of *Labrenzia* sp. Alg231-36 was chosen as reference as it was the largest genome of the collection. The core-genome consisted of 587 genes while the pan-genome comprised 25,449 genes. Genes encoding for ABC transporters, thioredoxins, nitrogen regulation (as an indicator to response in nitrogen limitation), peroxiredoxins, type II/IV secretion systems and glutathione S-transferases (GSTs: isoenzymes required in cellular detoxification) were identified as core genes present in all *Alphaproteobacteria* genomes analyzed here (**Appendix II- Table S3**, see below for further details on core genes). The number of singleton genes (those genes that are unique to each analyzed genome) was calculated for each genome and ranged from 955 singleton genes in *Rhodobacteraceae* bacterium Alg231-04 and 3,193 singleton genes in *Labrenzia* sp. Alg231-36 (**Figure 3-2**). The number of singleton genes correlated to some extent with the phylogenetic position of the isolates; the five *S. officinalis* strains of the *Roseobacter* group had the smallest numbers of singleton genes, followed by the *Sphingomonadales*, the *Rhizobiales* and then the two *Stappia* group isolates *Pseudovibrio* Alg231-02 and *Labrenzia* Alg231-36, which possessed the larger genome as well.

Table 3-1. Basic genome features of sponge-associated *Alphaproteobacteria* cultivated in this study.

Genomes	GC content (%)	Genome size (Mbp)	Total sequence depth (Gbp)	Genome coverage (x)	Coding sequences (CDs)	Number of RNAs	Number of rRNAs	Number of tRNAs	Accession numbers
Anderseniella sp. Alg231-50	57.9	4.61	0.65	143	4,635	45	3	42	LT703003-LT703010
Erythrobacter sp. Alg231-14	56.2	3.13	0.69	221	3,139	44	3	41	LT702999-LT703000
Labrenzia sp. Alg231-36	56.3	7.40	0.93	127	7,706	52	3	49	FREW01000001-FREW01000024
Sphingorhabdus sp. Alg231-15	52.8	3.62	0.47	132	3,702	45	3	42	LT703001-LT703002
Pseudovibrio sp. Alg231-02	51.3	5.96	0.73	124	5,674	77	12	65	FREX01000001-FREX01000026
Rhodobacteraceae bact. Alg231-30	55	4.54	0.76	169	4,604	49	6	43	FREU01000001-FREU01000010
Rhodobacteraceae bact. Alg231-04	59.6	4.81	0.95	198	4,784	61	11	50	FREY01000001-FREY01000029
Ruegeria sp. Alg231-54	56.5	4.92	0.76	155	5,120	50	6	44	FREZ01000001-FREZ01000035
Loktanella sp. Alg231-35	56.8	3.91	1.11	285	4,036	42	3	39	FREV01000001-FREV01000015
Tateyamaria sp. Alg231-49	57.4	4.51	0.77	173	4,793	41	3	38	FRFA01000001-FRFA01000039

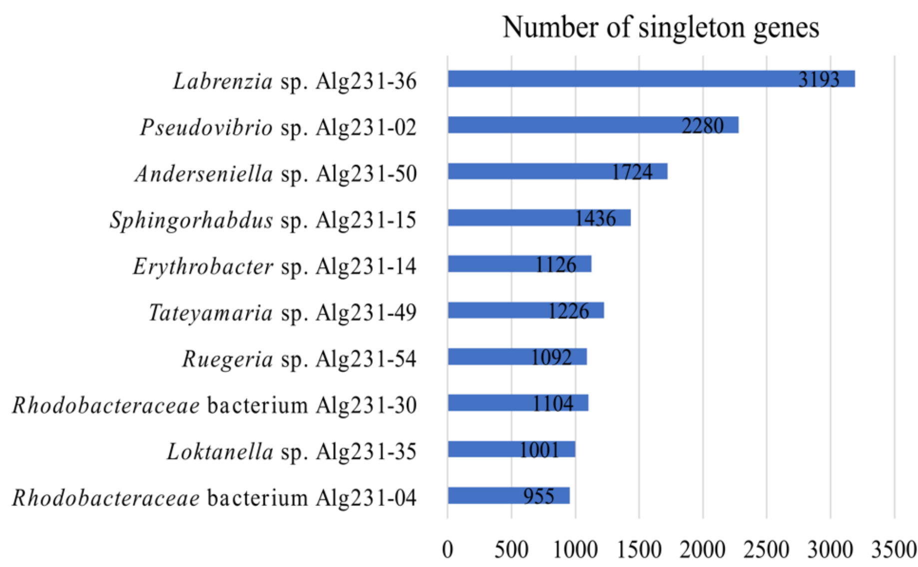


Figure 3-2. Number of strain-specific ("singleton") genes in each *Alphaproteobacteria* genome analyzed in this study.

Functional and comparative genomics based on Clusters of Orthologous Groups of Proteins (COGs)

As a first approach to compare the genomes at a coarse (i.e., COG classes) level, the number of CDSs in each COG class from each of the 10 *Alphaproteobacteria* genomes was summed up; this revealed that the classes 'amino acid transport and metabolism' (E), 'transcription' (K), 'carbohydrate transport and metabolism' (G), and 'energy production and conversion' (C) together with 'general function prediction' (R) and 'function unknown' (S) were the most dominant COG classes of the entire dataset (**Appendix II- Table S4**). Although the rank distribution of COG classes differed somewhat between the individual genomes, the above-mentioned COG classes always prevailed compared to other classes, in each genome (**Appendix II- Table S4**). At the finest level of (COG-based) functional resolution, in total 2,804 individual COG entries were annotated in the 10 genomes, with the number of COGs per genome ranging from 2,309 COGs in Erythrobacter sp. Alg231-14 and 5,625 COGs in *Labrenzia* sp. Alg231-36 (**Appendix II- Table S5**). COG profiling identified 959 COG enteries that were shared among all 10 cultivated *Alphaproteobacteria* genomes (**Appendix II- Table S6**), further detailed below.

Shared features of all genomes

Carbon, nitrogen, sulfur and phosphorus metabolism

All *Alphaproteobacteria* strains had in common several ABC-type transporter-encoding genes for the transport of sugars, dipeptides and branched-chain amino acids. All analyzed genomes had several copies of nitroreductase (COG0778)-encoding genes, an enzyme involved in the reduction of nitrogen-containing aromatic compounds. The nitrogen regulatory protein PII (COG0347), involved in the cell's response to nitrogen source availability, was also present with at least two gene copies in all *Alphaproteobacteria* genomes along with ammonia permease (COG0004; transporter) encoding genes. Several important sulfur metabolic functions were common to the *Alphaproteobacteria* genomes inspected. All contained several gene copies encoding for arylsulfatase A (COG3119), an enzyme breaking down sulfatides thus liberating sulfate, sulfate permeases (COG0659; transporters), sulfur transferases (COG2897), 3'-phosphoadenosine 5'-phosphosulfate (PAPS) 3'-phosphatase (COG1218; enzyme involved in sulfur assimilation and/or sulfate reduction) and sulfite reductases (COG0155; enzymes catalyzing the reduction of sulfite (SO_3^{2-}) to hydrogen sulfide (H_2S)). Further, all cultivated strains harbored several different ABC-type phosphate transporters and phosphate uptake regulators (COG0704). All genomes shared one gene encoding for guanosine polyphosphate pyrophosphohydrolases/synthetases (COG0317; signal transduction mechanisms).

Cofactors, vitamins and inorganic ions

All *Alphaproteobacteria* strains shared several genes encoding for proteins that contain or require B vitamins including thiamine (B1, COG0352), riboflavin (B2, COG0054; COG0307, COG1985), nicotinic acid (B3, COG1057), pyridoxamine phosphate oxidase (B6, COG0259), biotin (B7, COG0340), and cobalamin (B12, COG4547) (**Appendix II- Table S6**). The presence of riboflavin synthase alpha and beta chain and of pyridoxal phosphate biosynthesis protein (PdxJ) encoding genes confirms the potential synthesis of vitamin B2 and B6 by all genomes. Multiple genes in the COG class P "inorganic ion transport and metabolism" were shared among the 10 genomes, ensuring the trafficking of various essential ions including Ca^{2+}/Na^{+}, K^{+}, Fe^{2+}/Zn^{2+}, Fe^{3+}, and Mg/Co/Ni (**Appendix II- Table S6**). In addition, all genomes harbored between one and three arsenate reductase-encoding genes for the reduction of arsenate to arsenite in arsenic detoxification processes.

Defense, antibiotic resistance and reactive oxygen species (ROS) protection

All analyzed *Alphaproteobacteria* genomes shared several gene copies for cation and Na^+ driven multidrug efflux pumps, ABC-type multidrug efflux systems and antimicrobial peptide transport systems (**Appendix II- Table S6**). Hydrolases of the metallo-beta-lactamase superfamily and beta-lactamase class C were collective antibiotic resistance functions, whereby the *Sphingomonadales* strains *Sphingorhabdus* sp. Alg231-15 and *Erythrobacter* sp. Alg231-14 had, with 16 and 11 genes, respectively, the highest gene copy numbers. A gene encoding for an uncharacterized protein (COG1968) conveying resistance against the polypeptide antibiotic bacitracin was also detected (**Appendix II- Table S6).** A shared catalase (peroxidase 1, COG0376) encoding gene could scavenge reactive oxygen species (ROS). Moreover, all genomes were equipped with varied restriction-modification (R-M) systems (i.e. endonucleases) involved in anti-viral defense, but only one single R-M system (COG1403) was common to all of them (**Table 3-2**). Likewise, all genomes possessed genes involved in the biosynthesis of polyketides, but only one COG entry (COG5285) corresponding to a conserved protein domain related with the synthesis of fumonisin was shared by all genomes (**Table 3-3**).

Eukaryotic-like proteins (ELPs) encoding genes

In all ten alphaproteobacterial genomes, genes encoding for eukaryotic like proteins (ELPs) usually regarded to play a role in sponge-microbe interactions including ankyrin repeats (ANKs), tetratricopeptide repeats (TPRs), WD40 proteins, and pyrroloquinoline quinone (PQQ) were identified (**Figure 3-3**). Leucine-rich repeats (LRR), however, were detected only in *Pseudovibrio* sp. Alg231-02. Besides, only *Anderseniella* sp. Alg231-50 possessed all the above-mentioned ELP types. Also, the Anderseniella strain together with *Labrenzia* sp. Alg231-36 possessed the highest numbers of gene copies for the respective ELP motifs (**Figure 3-3**).

Table 3-2. Restriction-Modification systems identified in the genomes of sponge-associated *Alphaproteobacteria* analyzed in this study.

#COG	*An.*[1]	*Ey.*[2]	*Lab.*[3]	*Lak.*[4]	*Ps.*[5]	*R.*4[6]	*R.*30[7]	*Ru.*[8]	*Sp.*[9]	*Ta.*[10]	Description
COG0286	0	0	0	0	1	1	1	0	0	2	Type I restriction-modification system methyltransferase subunit
COG0732	0	0	0	0	1	1	1	0	0	1	Restriction endonuclease S subunits
COG1002	0	0	0	0	0	1	0	0	0	0	Type II restriction enzyme, methylase subunits
COG1403	1	1	1	1	1	2	1	1	1	2	Restriction endonuclease
COG3440	0	0	0	0	1	2	0	0	0	1	Predicted restriction endonuclease
COG4096	0	0	0	0	0	0	1	0	0	0	Type I site-specific restriction-modification system, R (restriction) subunit and related helicases
COG3183	0	0	0	0	0	0	1	0	0	1	Predicted restriction endonuclease
COG3440	0	0	0	0	1	2	0	0	0	1	Predicted restriction endonuclease
COG3587	1	0	0	0	0	0	0	0	0	0	Restriction endonuclease
COG1002	0	0	0	0	0	1	0	0	0	0	Type II restriction enzyme, methylase subunits
COG1401	0	0	0	0	0	0	0	0	0	1	GTPase subunit of restriction endonuclease
COG3587	1	0	0	0	0	0	0	0	0	0	Restriction endonuclease
Total	3	1	1	1	5	10	5	1	1	9	

All the COG entries belong to the Class V " Defense mechanisms "; **1)** *Anderseniella* sp. Alg231-50, **2)** *Erythrobacter* sp. Alg231-14, **3)** *Labrenzia* sp. Alg231-36, **4)** *Loktanella* sp. Alg231-35, **5)** *Pseudovibrio* sp. Alg231-02, **6)** *Rhodobacteraceae* bacterium Alg231-04, **7)** *Rhodobacteraceae* bacterium Alg231-30, **8)** *Ruegeria* sp. Alg231-54, **9)** *Sphingorhabdus* sp. Alg231-15, **10)** *Tateyamaria* sp. Alg231-49.

.

Table 3-3. Distribution of COG entries involved in polyketide biosynthesis across the genomes of sponge-associated *Alphaproteobacteria* analyzed in this study.

# COG	*An.*[1]	*Ey.*[2]	*Lab.*[3]	*Lak.*[4]	*Ps.*[5]	*R.4*[6]	*R.30*[7]	*Ru.*[8]	*Sp.*[9]	*Ta.*[10]	Description
COG2761	1	1	2	2	1	1	2	2	1	3	Predicted dithiol-disulfide isomerase involved in polyketide biosynthesis
COG3315	0	1	0	0	1	0	0	0	0	0	O-Methyltransferase involved in polyketide biosynthesis
COG3319	0	0	1	1	2	0	2	1	0	1	Thioesterase domains of type I polyketide synthases or non-ribosomal peptide synthetases
COG3321	0	0	3	1	2	0	1	0	0	0	Polyketide synthase modules and related proteins
COG5285	5	1	3	3	4	4	3	4	3	5	Protein involved in biosynthesis of mitomycin antibiotics/polyketide fumonisin

All the COG entries belong to the Class Q " Secondary metabolites biosynthesis, transport and catabolism"; **1)** *Anderseniella* sp. Alg231-50, **2)** *Erythrobacter* sp. Alg231-14, **3)** *Labrenzia* sp. Alg231-36**, 4)** *Loktanella* sp. Alg231-35, **5)** *Pseudovibrio* sp. Alg231-02, **6)** *Rhodobacteraceae* bacterium Alg231-04, **7)** *Rhodobacteraceae* bacterium Alg231-30, **8)** *Ruegeria* sp. Alg231-54, **9)** *Sphingorhabdus* sp. Alg231-15, **10)** *Tateyamaria* sp. Alg231-49.

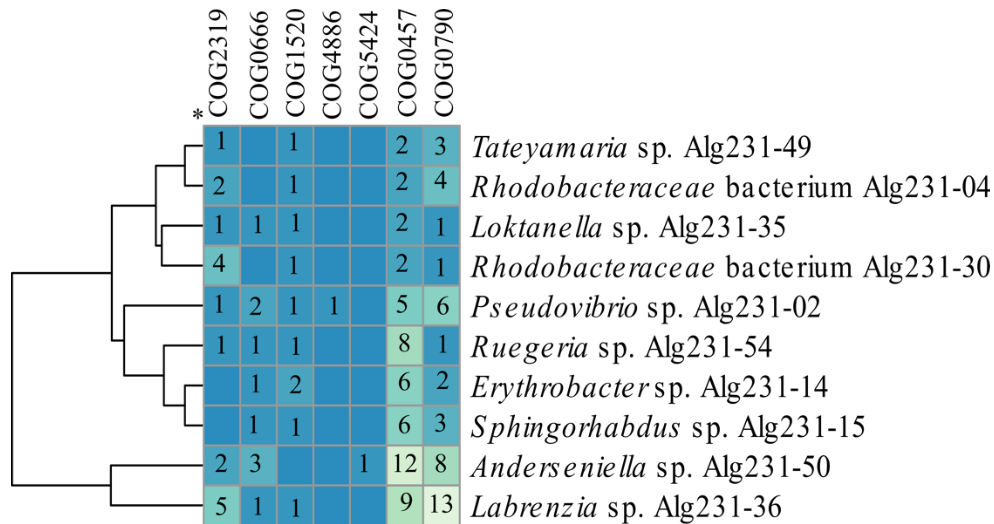


Figure 3-3. Heatmap on the absolute counts of coding sequences classified as COGs representing Eukaryotic-Like Proteins (ELPs) in the 10 cultivated *Alphaproteobacteria* genomes analyzed in this study. The numbers in each cell show the count of genes for each genome. The empty cells (dark blue) represent no counts (equals zero). * COG5424 (pyrrolso-quinoline quinone repeats), COG4886 (Leucine-rich repeats(LRR)), COG1520 and COG2319 (WD40 repeats), COG0666 (Ankyrin repeats), and COG0457and COG0790 (Tetratricopeptide repeats).

Roseobacter *versus non-*Roseobacter *genome features*

To visualize the genetic relatedness or distance between the ten alphaproteobacterial strains, principle components analysis (PCA) was performed based on their COG profiles, considering presence and absence as well as relative abundances of COG entries across all genomes (**Figure 3-4**). Three functional clusters were revealed: (1) five *Roseobacter* group genera/strains, (2) two *Sphingomonadales* genera/strains (*Erythrobacter* and *Sphingorhabdus*), (3) *Anderseniella* sp. Alg231-50 together with *Labrenzia* sp. Alg231-36 and *Pseudovibrio* sp. Alg231-02 (**Figure 3-4**). The third cluster indicates that the genomes of Alg231-36 and Alg231-02 are functionally closer to the *Rhizobiales* strain *Anderseniella* sp. Alg231-50 than to other genera belonging to their current (*Rhodobacterales*/ *Rhodobacteraceae*) taxonomic affiliation (cluster 1) (**Figure 3-4**).

To determine the characteristic genomic traits of the tightly clustering *Roseobacter* group (hereafter called group G1: *Loktanella* sp. Alg231-35, *Rhodobacteraceae* bacterium Alg231-04, *Rhodobacteraceae* bacterium Alg231-30, *Ruegeria* sp. Alg231-54 and *Tateyamaria* sp. Alg231-4), their COG profiles were collectively compared with those of the remaining five alphaproteobacterial genomes (hereafter termed group G2: *Anderseniella* sp.

Alg231-50, *Erythrobacter* sp. Alg231-14, *Labrenzia* sp. Alg231-36, *Pseudovibrio* sp. Alg231-02 and *Sphingorhabdus* sp. Alg231-15.) using a White's non-parametric t-test within STAMP.

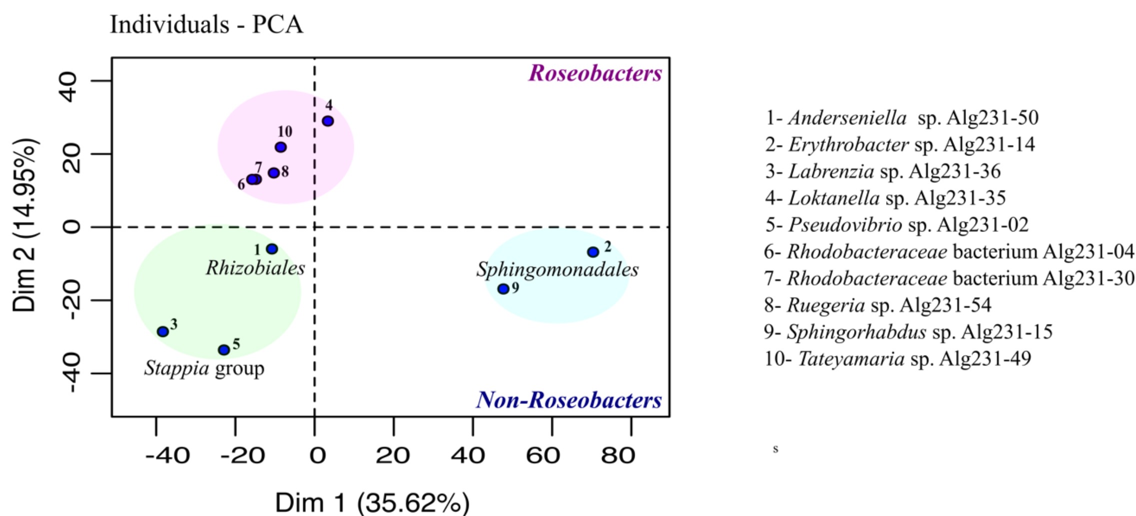


Figure 3-4. Principal Components Analysis (PCA) ordination of the ten alphaproteobacterial genomes analyzed in this study based on their functional profiles (i.e. presence/absence and relative abundances of COG entries per genome). Strains grouped by ellipses have their taxonomic affiliations disclosed following the results shown in Figure 3-1. Note the closer functional similarity between members of the "*Stappia* group" (*Pseudovibrio* and *Labrenzia,* formally belonging to the family *Rhodobacteraceae* in the order *Rhodobacterales)* to the genus *Anderseniella* (order *Rhizobiales*) than to other genera of the *Rhodobacteraceae* family.

The non-parametric t-test showed that 306 COGs were significantly different between the two groups in quantitative terms (**Appendix II- Table S7** and **Figure S2**). Furthermore, in a stringent qualitative comparison, 15 Roseobacter group (G1)-specific COG functions were identified which were present in all G1 genomes and not present in all the other alphaproteobacterial genomes (G2) (**Appendix II- Table S7**). Conversely, 34 COGs were absent in all of the *Roseobacter* group genomes (G1) but present in all non-*Roseobacter* genomes (G2). Six of the unique *Roseobacter* COGs were uncharacterized proteins of unknown function. Other unique *Roseobacter* COGs included a phosphoglyceromutase (COG0696) involved in glycolysis, a phosphatidylglycero phosphatase A (COG1267) involved in lipid metabolism and D-alanyl-D-alanine carboxypeptidase/penicillin binding protein 4 (COG2027) which is susceptible to lactam-antibiotics, a Zn-dependent carboxypeptidase (COG2317), and the DNA-binding protein H-NS(COG2916). Among the unique functions present only in the non-*Roseobacter* genomes (G2) were high abundances of a predicted signal transduction protein (COG5001), a protease II entry (COG1770, protein

catabolism), two inorganic pyrophosphatases (COG0221; COG3808) important in lipid degradation and inorganic phosphate production, and a type IV pili component (COG5461) generally important for adherence, movement and host colonization. COGs related to ABC transporters and sulfate/phosphate metabolism were generally enriched in the *Roseobacter* group (G1). The COG class Q 'Secondary metabolites biosynthesis' was enriched in G1, with only one entry in this class (COG0412, dienelactone hydrolase involved in chlorocatechol degradation) being enriched in G2. Besides these, N-acyl-L-homoserine lactone synthetases (COG3916) were enriched in G1. In contrast, predicted xylanase/chitin deacetylases (COG0726) and ELPs COGs (COG0666, COG0790, and COG0457) were more abundant in G2 (**Figure 3-3** and **Appendix II- Table S7**).

Secondary metabolism

To gain insight into their secondary metabolite production capacities, the ten alphaproteobacterial genomes were screened for the presence of secondary metabolite biosynthetic gene clusters using antiSMASH. All strains except *Rhodobacteraceae* bacterium Alg231-04 and *Erythrobacter* sp. Alg231-14 were found to harbor polyketide synthase (PKS) / non-ribosomal peptide (NRPS) encoding gene clusters via antiSMASH screening, whereby type I (T1PKS) encoding gene clusters were detected in *Pseudovibrio* sp. Alg231-02, *Loktanella* sp. Alg231-35 and *Rhodobacteraceae* bacterium Alg231-30; type III (T3PKS) gene clusters in *Anderseniella* sp. Alg231-50, *Labrenzia* sp. Alg231-36, *Sphingorhabdus* sp. Alg231-15 and *Pseudovibrio* sp. Alg231-02 and NRPS gene clusters in Pseudovibrio sp. Alg231-02, *Labrenzia* Alg231-36 and *Ruegeria* Alg231-54 (**Appendix II- Table S8**). Terpene synthesis encoding gene clusters were detected in eight of the ten strains but not in *Loktanella* sp. Alg231-35 and *Rhodobacteraceae* bacterium Alg231-04 (**Appendix II- Table S8**). In *Erythrobacter* sp. Alg231-14, the terpene gene cluster showed 75% similarity to the astaxanthin-dideoxyglycoside biosynthethic gene cluster. This strain also possessed a lasso-peptide encoding gene cluster. Bacteriocine (peptidic toxins) synthesis gene clusters have been detected for all cultivated strains except *Anderseniella* sp. Alg231-50. However, only *Anderseniella* sp. Alg231-50 harbored a gene cluster encoding for the osmolyte ectoine. Genes encoding for homoserine lactone signaling molecules were identified via antiSMASH in six *Rhodobacterales* strains, namely *Loktanella* sp. Alg231-35, *Rhodobacteraceae* bacterium Alg231-04, *Rhodobacteraceae* bacterium Alg231-30, *Tateyamaria* sp. Alg231-49, *Ruegeria* sp. Alg231-54 and *Labrenzia* sp. Alg231-36 (**Appendix II- Table S8**).

Representation of the cultivated* Alphaproteobacteria *genomes in marine metagenomes

The available shotgun sequenced metagenomes of *S. officinalis* and surrounding seawater and sediment samples (Karimi et al., 2017b) were mapped against the 10 cultivated *Alphaproteobacteria* genomes. The numbers of aligned metagenome reads were generally low for all genomes (**Table 3-4**). Of the 10 cultivated strains, *Anderseniella* sp. Alg231-50 was the most dominant strain in the sponge metagenome followed by *Labrenzia* sp. Alg231-36 and *Ruegeria* Alg231-54, while *Rhodobacteraceae* bacterium Alg231-30 clearly was the least abundant one. All *Alphaproteobacteria* genome reads were somewhat more abundant in the seawater metagenome, followed by sediments and then *S. officinalis* (**Table 3-4**). As a frame of comparison with *Anderseniella* sp. Alg231-50, 7x as many metagenomic reads from *S. officinalis* were found to align with the genome of the dominant and uncultivated *Rhodospirillaceae* symbiont So9, reconstructed from the host's microbial metagenome via genomic binning procedures (Chapter 4 and Karimi et al. (2018-in press)).

Embedded description of the* Anderseniella *sp. Alg231-50 genome

The genome of *Anderseniella* sp. Alg231-50 was further explored for several reasons: first, of the here presented *Alphaproteobacteria* strains, Alg231-50 was the most dominant one in the *S. officinalis* metagenome (**Table 3-4**). Second, many different types of eukaryotic like proteins (ELPs) were present in this genome, some of them even with high copy numbers, suggesting that Alg231-50 is well equipped for symbiotic life inside its host. Third, to the best of our knowledge, there is no *Anderseniella* genome available on public databases yet (stand: 14th of November 2017) and only eight genome sequences exist for the entire *Rhodobiaceae* family but none of them derived from a sponge host. Of the 4,635 CDSs predicted in the *Anderseniella* genome using the RAST server, 1,494 were annotated as encoding for hypothetical proteins, but 4,109 CDSs could be assigned a COG function (**Appendix II- Table S4**). A total of 45 RNA genes were identified including 3 rRNAs and 43 tRNAs (**Table 3-1**). Quite remarkably, this strain shares several genome features in common with the sponge-specific, so-far unculturable *Rhodospirillaceae* symbiont So9 (Karimi et al., 2018-in press), including the potential to degrade aromatic hydrocarbons (e.g. toluene, biphenyl, benzoate, salicylate ester), tolerate heavy metals (copper, cobalt, mercury, chromium) and antibiotics (becta-lactams, fluoroquinolones, colicin E2), and utilize taurine and alkanesulfonates. Moreover, as observed for the symbiont So9, RAST and COG annotations did not reveal any genes encoding for flagellar cell motility and chemotaxis in the *Anderseniella* genome (**Appendix II- Table S5**).

Table 3-4. Percent alignment of total metagenomic reads from *S. officinalis*, seawater and sediments with the genomes assembled in this study.

Genome vs metagenome	Aligned reads[1]	Percent (%)[2]	*Genome vs* metagenome	Aligned reads[1]	Percent (%)[2]
***Anderseniella* sp. Alg231-50**			***Ruegeria* sp. Alg231-54**		
S. officinalis	2610	0.00829	*S. officinalis*	2135	0.00678
Sediment	3264	0.01476	Sediment	6728	0.03043
Seawater	4846	0.0214	Seawater	7022	0.03101
***Erythrobacter* sp. Alg231-14**			***Sphingorhabdus* sp. Alg231-15**		
S. officinalis	1790	0.00568	*S. officinalis*	1708	0.00542
Sediment	2006	0.00907	Sediment	1794	0.00811
Seawater	4197	0.01854	Seawater	4050	0.01789
***Labrenzia* sp. Alg231-36**			***Tateyamaria* sp. Alg231-49**		
S. officinalis	2157	0.00685	*S. officinalis*	2008	0.00638
Sediment	2124	0.00961	Sediment	5485	0.02481
Seawater	4856	0.02145	Seawater	7380	0.03259
***Loktanella* sp. Alg231-35**			***Pseudovibrio* sp. Alg231-02**		
S. officinalis	1911	0.00607	*S. officinalis*	1776	0.00564
Sediment	2445	0.01106	Sediment	1709	0.00773
Seawater	7285	0.03217	Seawater	3913	0.01728
***Rhodobacteraceae* bact. Alg231-04**			***Rhodobacteraceae* bact. Alg231-30**		
S. officinalis	1987	0.00631	*S. officinalis*	236	0.00075
Sediment	3331	0.01507	Sediment	932	0.00422
Seawater	6326	0.02794	Seawater	2200	0.00972

The total number of paired-end sequence reads in the metagenome dataset (Karimi et al., 2017b) were as follows: *S. officinalis* - 31497820; Sediment - 22107730; Seawater - 22641917.

[1]Aligned reads - the number of metagenomic sequence reads from a given environment that aligned with the genome sequence of the respective *Alphaproteobacterium* isolate.

[2] Percent (%) - the percentage of metagenomic sequence reads from a given environment that aligned with the genome sequence of the respective *Alphaproteobacterium* isolate.

Nevertheless, a type IV pili component (COG5461) and a protein required for attachment to host cells (COG5622) has been detected. Furthermore, a highly versatile carbohydrate metabolism and the potential ability to biosynthesize polyphosphates were also inferred for strain Alg231-50. Nitrate (NO_3^-) transporter (COG0600, COG0715, COG1116, COG2223) and nitrate reductase (composing subunit alpha, beta, gamma and delta) (COG5013, COG1140, COG2181, COG2180 and COG3005) genes were as well observed (**Appendix II- Table S5**). Additionally, three possible CRISPR repeats with four spacers in total were predicted for this genome, and polyketides and terpene biosynthesis capacities likely contribute to the secondary metabolite repertoire of the strain (**Appendix II- Table S8**).

Discussion

In this study, we used an oligotrophic medium along with low temperature and prolonged incubation time to attempt the cultivation of sponge–associated bacteria other than those usually retrieved with regular procedures (e.g. Brinkmann et al. (2017); Esteves et al. (2013); O'Halloran et al. (2011); (Kennedy et al., 2009)). Although with 48 isolates the size of our collection was rather small, selecting isolates according to their distinctive morphological characters under the above-mentioned culturing conditions resulted in the collection of 14 bacterial genera, enabling the cultivation of novel species. In comparison, a previous study that collected 327 strains with standard cultivation procedures (i.e. full strength marine agar medium; 25°C incubation over three days and random, quantitative-based colony picking and purification) obtained only slightly more (17) bacterial genera despite the much larger cultivation effort (Esteves et al., 2013). While the latter approach enables quantitative assessments of symbiont diversity and the analysis of genome diversification below species level, the methodology employed here led to the retrieval of a broad phylogenetic panel of isolates, enabling deep comparative genomic assessments among symbionts above the species level.

The here implemented culturing condition was very adequate for the growth of diverse *Alphaproteobacteria* species. These results are in line with the observations of Sipkema et al. (2011) who retrieved a majority of *Alphaproteobacteria* strains from the marine sponge *Haliclona* sp. using diverse oligotrophic media. Typically, most bacteria isolated from sponges to date have been affiliated with the phyla *Actinobacteria, Bacteroidetes, Firmicutes,* and *Proteobacteria* (Muscholl-Silberhorn et al., 2008; Esteves et al., 2013; Hardoim and Costa,

2014b; a). In this study, 24 OTUs were uncovered from 48 isolates of which two OTUs representing *Rhodobacteraceae* phylotypes were regarded as possible candidates for new genera (**Appendix II- Figure S1**). Although isolate 231-04 showed 98% homology with the type strain *Phaeobacter inhibens* T5 isolated from water surface North Sea (Martens et al., 2006) and isolate 231-30 indicated 97% similarity with type strain *Thalassobius aestuarii* JC2049 isolated from tidal surface sediment (Yi and Chun, 2006), both strains were clustered in a different clade after dedicated phylogenetic assessments (**Appendix II- Figure S1**), suggesting that they may represent novel genera, what needs to be confirmed in future phylogenomic assessments including multiple genes from several closely-related genomes.

***Communal metabolite features of cultivable sponge-associated* Alphaproteobacteria**

The functional and comparative genomics assessment performed this study enabled us to delve into common genomic attributes of ten sponge-associated *Alphaproteobacteria*, and to address whether any of these redundant functions may be relevant in improving both host and symbiont fitness within the marine sponge microbiome. One possible, common nutrient cycling role has been predicted through the presence of genes encoding for the nitrogen regulatory protein PII, which regulates glutamine synthetase (GS) activity by activating GS under nitrogen-limiting conditions (de Zamaroczy et al., 1990). However, this appears to be a generalist function important both in host-associated as well as free-living microhabitats. Perhaps more relevant in terms of possible associations between these bacteria and their sponge hosts are the presence of gene copies encoding for arylsulfatases in all genomes, an attribute verified to be enriched in the marine sponge microbiome (Karimi et al., 2017b), underlining therefore one possible role of this pool of *Alphaproteobacteria* species in consuming sulfated polysaccharides, a feature that has also been discussed for uncultivated, widespread *Rhodospirillaceae* species from *S. officinalis* (Karimi et al., 2018-in press and Chapter 4) and revealed to be common among several uncultivated lineages of sponge symbionts (Slaby et al., 2017). Furthermore, we revealed that several genes encoding for vitamin B biosynthesis were shared among the *Alphaproteobacteria* species studied here, suggesting a potential role of members of this class, in general, in providing essential nutrients for host growth and functioning. This hypothesis has been tested in a marine dinoflagellate (*Lingulodinium polyedrum*) and provided evidence that associated *Alphaproteobacteria* nourish vitamins B1 and B12 required by the host for growth (Cruz-López and Maske, 2016). Besides, arsenate reductases encoding genes for reduction of arsenate (As(V)) to arsenite (As(III)) in arsenic detoxification processes highlights the capacity of our isolates to reduce toxic heavy metals (Mukhopadhyay and Rosen, 2002;

Silver and Phung, 2005), a function that has been proved for the sponge symbiont *Entotheonella* sp. which mineralizes arsenic and barium on intracellular vesicles (Keren et al., 2017). Furthermore, genes involved in ABC-type multidrug efflux systems, hydrolases of the metallo-beta-lactamase superfamily, and remediation of ROS stress underline how versatile the mechanisms of cell detoxification employed by these organisms can be (Yung et al., 2011; Santos et al., 2012). Such capabilities may substantially increment bacterial fitness within dense and chemically-rich microbial communities which is the case of the marine sponge microbiome.

Antimicrobial agents

Using antiSMASH, we could detect several antibiotic biosynthetic gene clusters across the studied genomes, in line with accumulating, *in vitro* evidence for mild to high antimicrobial activities by sponge-associated *Alphaproteobacteria* such as *Rugeria*, *Pseudovibrio,* and *Labrenzia* (Hentschel et al., 2001; O'Halloran et al., 2011; Esteves et al., 2013; Graça et al., 2013; Crowley et al., 2014; Naughton et al., 2017). Particularly, both terpene-synthase and polyketide-synthase (PKS) biosynthetic gene clusters were common among the studied strains, each being present in eight out of ten genomes (see **Appendix II- Table S8**). Interestingly, we have previously documented the enrichment of these biosynthetic gene clusters in the microbial metagenome of the *S. officinalis* endosymbiotic consortium (Karimi et al., 2017b), highlighting the potential participation of such usually inhibitory molecules in microbe-microbe antagonistic interactions within the sponge holobiont or in sponge chemical defense. The roles and activities of polyketides from sponge symbiotic bacteria have been largely explored in the last fifteen years (Piel, 2002; Piel et al., 2004; Hentschel et al., 2012; Wilson et al., 2014), and the presence of PKS biosynthetic gene clusters have been consistently documented for culturable, sponge-associated *Pseudovibrio* spp. (Bondarev et al., 2013; Esteves et al., 2013; Naughton et al., 2017). However, much less is known about the potential contribution of bacterial symbionts as producers of terpenoids in marine sponges. Intriguingly, terpenoid biosynthesis has been regularly documented in marine sponges of the order *Dictyoceratida* (Keyzers et al., 2006). Yet the origin of the biosynthesis (host or symbionts) has been, to our knowledge, not addressed so far. Spongian diterpenoids have shown antimicrobial activity against pathogenetic bacteria such as *Pseudomonas aeruginosa* (Keyzers et al., 2006). Dihydrogracilin A, aterpene extracted from the sponge host *Dendrilla membranosa*, has been shown to possess immune modulatory and anti-inflammatory action (Ciaglia et al., 2017). Furthermore, except for *Anderseniella*, all other *Alphaproteobacteria*

strains possessed the potential to produce bacteriocins, proteinaceous toxins synthesized by bacteria usually regarded to inhibit growth of closely related strains and, as such, considered to be major molecules shaping the structure of microbial communities *in situ* (Drider et al., 2016). In this study, we demonstrate that polyketide, terpene and bacteriocin biosynthesis capacities are widespread across diverse sponge-associated *Alphaproteobacteria*, suggesting a pivotal contribution of this group to the chemical complexity, natural product biosynthesis repertoire and taxonomic composition of the marine sponge microbiome.

Finally, we observed ectoine biosynthesis potential exclusively for the *Anderseniella* strain (**Appendix II- Table S8**). Ectoine was detected as protector against osmotic and environmental stresses such as high salinity and cold (Kuhlmann et al., 2011).

Representativeness of cultured* Alphaproteobacteria *in the marine sponge microbiome

We used microbial metagenome-genome mapping as a means to infer abundance relationships between the bacteria analyzed in this study in the marine sponge microbiome. We further contrasted the percentage of metagenome-genome aligned reads obtained for the cultivated symbionts inspected here with that of a dominant, uncultivated alphaproteobacterial symbiont of *S. officinalis*, namely *Rhodospirillaceae* bacterium So9 (**Chapter 4**). The percentage of reads from sponge, sediment and seawater microbial metagenomes which aligned with the genomes of our cultured *Alphaproteobacteria* was generally low. It is possible that technical limitations such as insufficient sequencing depth and/or the usage of only short read lengths which may not align properly with reference genomes (Clooney et al., 2016; Tessler et al., 2017) contribute to an underestimation of relative abundances calculated in this fashion. However, usually higher percentages of aligned metagenome-genome reads were obtained for seawater metagenomes, suggesting that most of the cultures surveyed here may be more abundant in this habitat than in sponges. This reinforces the notion fostered by Hardoim and Costa (2014b) and Montalvo et al. (2014), for instance, that current cultivation attempts of marine symbiotic bacteria tend to enrich and select for low abundant to only moderately abundant members of this consortium. Regardless of possible limitations, the relative abundance of a dominant, uncultured alphaproteobacterium member of the *S. officinalis* microbiome (*Rhodospirillaceae* bacterium So9) was about 7-fold higher than that estimated for the most abundant isolate retrieved in this study, *Anderseniella* sp. strain Alg231-50. Therefore, interpretations regarding the possible ecological roles of the here cultivated symbionts - and essentially all sponge symbiotic bacteria thus far obtained in culture - must be taken with caution. It is felt that a deeper perspective of the true abundance of bacterial

symbionts (both cultivated and uncultivated) in marine sponges is still required for a proper understanding of the relative forces exerted by several bacterial symbionts in holobiont functioning. This perspective cannot be achieved with the sole use of DNA sequencing technologies and, most likely, only a dedicated effort integrating taxon-oriented, high resolution imaging (enabled e.g. by fluorescent in situ hybridization coupled to confocal laser scanning microscopy, FISH-CLSM), symbiont cultivation, and deep microbiome sequencing may altogether enhance our understanding of symbiont abundance ranks in marine sponges, from the very dominant to the very rare bacterial associates. In spite of this, the several common genome features revealed in this study for cultivated and uncultivated bacteria is indicative of functional redundancy and compatible metabolic circuitry among diverse alphaproteobacterial symbionts of marine sponges, substantiating their importance as collective mediators of microbiome functioning and structure in these hosts.

ELPs are known as symbiotic factors in sponges because of the role they play in the modulation of cellular protein-protein interactions and in the prevention of symbiont phagocytosis by host cells (Díez-Vives et al., 2016; Reynolds and Thomas, 2016). Previous studies have shown that sponge-associated bacterial genomes are enriched in genes encoding for ELPs such as ANKs, TRPs, WD40, and LRR (Thomas et al., 2010; Fan et al., 2012; Liu et al., 2012; Díez-Vives et al., 2016; Karimi et al., 2017b). ELP-encoding genes can be gained from both horizontal and lateral transfer (Lurie-Weinberger et al., 2010). These proteins have been shown to affect phagocytosis of amoeba cells *in vitro* by interrupting phagosome maturation process (Reynolds and Thomas, 2016). Because sponges feed on their trapped bacteria by using specialized archaeocytes (amoeboid-like cells), ELPs might act as molecular signatures enabling bacteria to evade digestion by sponge cells. We found that all *Alphaproteobacteria* genomes analyzed in this study carry ELPs. Non-*Roseobacter* genomes (group G2) had higher ELP counts in comparison with *Roseobacters* (group G1), suggesting higher affinity of members of the former group in establishing favorable or more stable interactions with marine sponges. It remains to be determined whether such molecular signatures may likewise be involved in bacterial adaptation to other sessile marine hosts such as ascidians, corals and bryozoans, therefore supporting a generalist pattern of occurrence of these symbionts across several host organisms, as already documented for *Pseudovibrio* and *Ruegeria* species (see e.g. Keller-Costa et al. (2017)).

Genome features of *Anderseniella* sp. Alg231-50

The 16S rRNA gene of *Anderseniella* sp. strain Alg231-50 shares 100% similarity with the "marine sponge bacterium strain FILTEROTU17", isolated from *Haliclona* sp. (Sipkema et al., 2011). *Anderseniella* species have not been frequently reported from sponges, suggesting recalcitrance to common cultivation procedures. Analysis of the *Anderseniella* genome revealed two COGs (COG5461; Type IV pili component and COG5622; cell attachment protein) involved in adhesion and attachment to host cells. The type IV pilus system (T4PS) is a multifunctional machine which, among other features, promotes adherence to eukaryotic cells (Burrows, 2012). It has been shown that T4PS is important for pathogenic bacteria (Melville and Craig, 2013), for instance, to maximize biofilm formation upon host colonization (Varga et al., 2008). Overall, the presence of adhesion proteins in the *Anderseniella* genome suggests eukaryotic host colonization aptitude consistent with other genomic traits identified in this organism to favor a symbiotic life-style. Another noticeable encoding genes found for this strain were nitrate (NO_3^-) transporter and reductase genes (comprising alpha, beta, gamma and delta subunits) shown to modulate ammonia uptake and utilization via the general nitrogen regulatory system (Ntr) (Moreno-Vivián et al., 1999), suggesting a potential role of strain Alg231-50 strain in metabolic waste (that is, ammonia) removal thereby contributing to host fitness. Furthermore, three CRISPR-Cas were detected in the *Anderseniella* Alg231-50 genome. CRISPRs are adaptive defense systems in bacteria which can memorize any attack from viruses and plasmids based on keeping conserved repeats and different spacer sequences (Barrangou and Marraffini, 2014). Metagenomic studies have shown that the marine sponge micorbiome is enriched in CRISPR-Cas encoding genes (Horn et al., 2016; Karimi et al., 2017b) and, therefore, the presence of these elements in the *Andeseniella* genome is indicative of a close interaction between this symbiont and its sponge host.

In conclusion, the use of simple modifications to regular culture conditions coupled to dedicated genome-wide analysis of marine sponge symbionts enabled unprecedented access to highly versatile metabolisms across diverse *Alphaproteobacteria*. Although the cultivated symbionts reported here clearly do not rank among the most dominant members of the sponge endosymbiotic consortium, they were found to display a multitude of genomic features enabling persistence in this particular microenvironment which have been regularly described as genomic signatures of the marine sponge microbiome. Taken together, the outcomes compiled here contribute to novel insights into the potential roles of alphaproteobacterial communities in mediating molecular interactions and shaping the structure of the marine sponge microbiome. The several, phylogenetically distinct bacterial cultures retrieved in this

study can now be used in the determination of biological activities and natural product biosynthesis in the laboratory.

Supplemental material

The supplementary materials and explanations for this chapter can be found in Appendix II.

Acknowledgments

The authors would like to thank Telma Franco for her great help during sample processing and bacterial DNA extractions, and Marta Valente for Sanger-sequencing of16S rRNA genes.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through the research grants PTDC/BIA-MIC/3865/2012 and PTDC/MAR-BIO/1547/2014 [conceded to RC] and a full PhD scholarship from the Erasmus Mundus Programme/SALAM EMA2 lot7/SALA1206422 [Awarded to EK].

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethics statement

This study relied on *in situ* sampling of microorganisms from marine invertebrates without a nervous system, and as such was exempt from ethical approval procedures according to the current Portuguese legislation (Decreto-Lei nº 113/2013). This study did not occur within privately owned or protected areas. This study did not involve endangered or protected species. The sampling methodology privileged minimally invasive handling procedures, following the guidelines of the European Directive 2010/63/EU.

Author Contributions

EK performed laboratory experiments; RC and UH provided reagents and materials; EK, TKC, BMS, CJC, and UND analyzed the data; EK wrote the first draft of the manuscript and prepared figures and tables. EK and RC wrote the final manuscript text; All authors reviewed the manuscript.

Chapter 4

Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges

Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges

Elham Karimi[1], Beate M. Slaby[2], André R. Soares[3], Jochen Blom[4], Ute Hentschel[2,5], Rodrigo Costa[1,6]

1 Centre of Marine Sciences (CCMAR), Faculty of Science and Technology (FCT), Algarve University, 8005-139 Faro, Portugal.

2 RD3 Marine Microbiology, GEOMAR Helmholtz Centre for Ocean Research Kiel, 24105 Kiel, Germany.

3 Institute of Geography and Earth Sciences, Aberystwyth University, SY23 3DB Aberystwyth, Wales, UK.

4 Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, 35392 Giessen, Germany.

5 Christian-Albrechts-Universität zu Kiel, 24118 Kiel, Germany.

6 Institute for Bioengineering and Biosciences (IBB), Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisbon, Portugal.

A version of this chapter has been accepted for publication in *FEMS Microbiology Ecology*.

Abstract

Marine sponges are early-branched metazoans known to harbor dense and diverse microbial communities. Yet the role of the so far uncultivable alphaproteobacterial lineages that populate these sessile invertebrates remains unclear. We applied a sequence composition–dependent binning approach to assemble one *Rhodospirillaceae* genome from the *Spongia officinalis* microbial metagenome and contrast its functional features with those of closely-related sponge-associated and free-living genomes. Both symbiotic and free-living *Rhodospirillaceae* shared a suite of common features, possessing versatile carbon, nitrogen, sulfur and phosphorus metabolisms. Symbiotic genomes could be distinguished from their free-living counterparts by the lack of chemotaxis and motility traits, enrichment of genes required for the uptake and utilization of organic sulfur compounds - particularly taurine -, higher diversity and abundance of ABC transporters, and a distinct repertoire of genes involved in natural product biosynthesis, plasmid stability, cell detoxification and oxidative stress remediation. These sessile symbionts may more effectively contribute to host fitness via nutrient exchange, and also host detoxification and chemical defense. Considering the worldwide occurrence and high diversity of sponge-associated *Rhodospirillaceae* verified here using a tailored *in silico* approach, we suggest that these organisms are not only relevant to holobiont homeostasis but also to nutrient cycling in benthic ecosystems.

Introduction

Marine sponges contain dense and diverse microbial communities. These sessile filter-feeders have emerged in the evolutionary history of our planet about 600 million years ago, and are known to establish close interactions with prokaryotes. It has been documented that up to 35% of the sponge wet weight can be comprised of bacterial cells (Vacelet and Donadey, 1977). Diverse and usually abundant bacterial communities populate the mesohyl matrix of marine sponges (Taylor et al., 2007b; Webster and Taylor, 2012; Kamke et al., 2014), but much remains to be learned about the actual *in situ* ecological functions of sponge-associated microorganisms and their potential benefits to the sponge host (Kamke et al., 2014; Webster and Thomas, 2016). Molecular evidence suggests that sponge-associated prokaryotes play fundamental roles in nutrient provision, removal of metabolic by-products, chemical defense and shelter from disease (Piel et al., 2004; Hentschel et al., 2006; Thomas et al., 2010; Fan et al., 2012; Wilson et al., 2014; Lackner et al., 2017).

The class *Alphaproteobacteria* (Garrity et al., 2005) ranks among the most widespread and abundant bacterial groups in the oceans, in great extent owing to the high representativeness of members of the complex marine group *Roseobacter* (*Rhodobacteraceae*) in planktonic settings (Giovannoni and Rappé, 2000; Buchan et al., 2005; Wagner-Döbler and Biebl, 2006; Giebel et al., 2011; Simon et al., 2017). Consequently, knowledge of the diversity and function of free-living *Alphaproteobacteria* has grown enormously in recent years (Morris et al., 2002; Luo et al., 2012b; Simon et al., 2017). In contrast, genome-wide analysis of several marine alphaproteobacterial groups other than *Rhodobacteraceae* (e.g. *Rhizobiales*, *Rhodospirillales*, *Rickettsiales*) and of symbiotic *Alphaproteobacteria* remains comparatively scarce in spite of their ever-increasing documentation as dominant players of eukaryote-prokaryote associations in the oceans (Tujula et al., 2009; Simister et al., 2012; Erwin et al., 2014; Pantos et al., 2015). For instance, a *Sulfitobacter* (*Rhodobacteraceae*) symbiont has been shown to promote diatom growth through indole-3-acetic acid (auxin) biosynthesis (Amin et al., 2015), while symbiotic *Roseovarius* (*Rhodobacteraceae*) induces cell elongation and division during the morphogenesis of the green macroalgae *Ulva mutabilis* (Spoerner et al., 2012; Grueneberg et al., 2016). A chemoautotrophic, sulfur-oxidizing symbiont in the order *Rhodospirillaceae* was further found to profusely colonize the mouthless catenulida flatworm *Paracatenula* in an example of intimate and ancient animal-alphaproteobacterial symbiosis (Gruber-Vodicka et al., 2011).

Alphaproteobacteria species represent one of the most widespread, diverse and dominant groups of sponge-associated bacteria (Enticknap et al., 2006; Schmitt et al., 2007; Simister et al., 2012; Hardoim et al., 2014; Thomas et al., 2016). In the last two decades or so, alphaproteobacterial lineages have often been reported as prevailing sponge-associated bacteria both via cultivation-dependent (O'Halloran et al., 2011; Esteves et al., 2013) and cultivation-independent (Taylor et al., 2007a; Hardoim et al., 2009; Simister et al., 2012; Thomas et al., 2016) studies. *Alphaproteobacteria* have been detected in both adult and larval samples of marine sponges (Schmitt et al., 2007), suggesting a pattern of vertical symbiont transmission through successive host generations, and thus an intimate association between these bacteria and their sponge hosts. A dedicated survey of 1,385 sponge-derived alphaproteobacterial 16S rRNA gene sequences revealed that 18% of the entries represented "sponge-specific bacterial clusters" (SC), with >30 alphaproteobacterial SCs being observed (Simister et al., 2012). This suggests that several distinct, deeply branched "sponge-specific" or "enriched" lineages likely representing novel alphaproteobacterial taxa are yet to be fully described.

Due to advances in next generation sequencing technologies and bioinformatics, reconstruction of genomes of uncultivated symbiotic bacteria from metagenomes is now achievable. To date, several studies reported the successful binning of genomes from diverse metagenomic samples (Iverson et al., 2012; Luo et al., 2012a; Albertsen et al., 2013; Sharon and Banfield, 2013; Nielsen et al., 2014; Burgsdorf et al., 2015; Britstein et al., 2016). In this study, we reconstruct the genome of a prevalent and uncultivated alphaproteobacterial symbiont of the *Rhodospirillaceae* family (order *Rhodospirillales*) from microbial metagenomes of the model sponge host *Spongia officinalis*. We reveal the genomic signatures, adaptive features and distinguishing functions of this lineage using an unprecedented comparative genomics endeavor to symbiotic *Rhodospirillaceae* that includes further analysis of closely-related symbiotic genomes binned from the marine sponge *Aplysina aerophoba* (Slaby et al., 2017) and of free-living phylotypes. We further perform an *in silico* assessment of the worldwide abundance and distribution of symbiotic *Alphaproteobacteria*, *Rhodospirillales* and *Rhodospirillaceae* in marine sponges using a recent dataset release (Thomas et al., 2016) to address the implications of the symbionts' metabolism to the functioning of the marine sponge holobiont and of benthic ecosystems at large.

Materials and Methods

Sample collection, DNA extraction and sequencing

S. officinalis specimens (4 biological replicates) and seawater (3 replicates) were collected off the Algarve coast, southern Portugal (Northeast Atlantic, 36º 58' 47.2N ;7º 59' 20.8W) in May 2014, from c. 20 m depth and transported to the laboratory in a cooling box within 1 h. Sampling and sample processing followed the methods described in detail by Hardoim et al. (2014). DNA extraction and Illumina metagenome sequencing procedures were described by Karimi et al. (2017b) and are available as Supporting Information in Appendix III. Briefly, metagenome sequencing of the *S. officinalis* endosymbiotic consortium was performed on microbial cell pellets (MCPs) retrieved from the host samples. To this end, differential centrifugation of homogenates obtained by maceration was carried out as explained previously (Hardoim et al., 2014) and detailed in **Appendix III- File S1**.

Metagenome assembly

To enable higher completeness and coverage of the genomes to be binned from metagenome samples, metagenomic reads from sponge and seawater replicates were pooled per habitat prior to metagenome assembly, which was performed with metaSPAdes of SPAdes 3.9.1 (Nurk et al., 2016). The resulting assemblies were inspected with the metaquast script of QUAST v.4.4 (Gurevich et al., 2013). Only contigs of at least 1,000bp were used for further investigation.

Genome binning from metagenomes

All contigs ≥ 20,000bp in length were split into min. 10,000bp sub-contigs as part of the binning protocol of CONCOCT v. 0.4.0 (Alneberg et al., 2013). The non-concatenated read datasets were mapped to the sub-contigs of the respective metagenome with bowtie2 v. 2.2.6 using default settings (Langmead and Salzberg, 2012). Sorting, indexing, transformation into BAM format, and read depth calculation were performed with Samtools v. 1.2 (Li et al., 2009). A python script (avgcov_from_samtoolsout.py)[1] was used to calculate the average coverage for each contig from the read depth values. For each habitat (sponge and seawater) the coverage tables from every read dataset were merged into one differential coverage table. Binning was performed with CONCOCT v. 0.4.0 (Alneberg et al., 2014) at default settings. After merging sub-contigs back into the original contigs, a fasta file was created for each bin with a python script (mkBinFasta.py)[1].

[1] https://github.com/bslaby/scripts/blob/master/

The completeness of the genomic bins was estimated based on a hmm search (HMMER 3.1b1) against a database of 111 essential single-copy genes (Finn et al., 2011; Alneberg et al., 2013) after prediction of open reading frames (ORFs) with prodigal v. 2.6.1 (Hyatt et al., 2010). Only bins with more than 85% estimated completeness were considered for further analyses. The JSpeciesWS server (Richter et al., 2016) was used to specify the phylogenetic affiliation of all selected bins. After nominating possible target bins (*Alphaproteobacteria*), the reads mapping to the selected bins were assembled *de novo* with SPAdes 3.9.1 and IDBA-UD of IDBA 1.1.3 (Peng et al., 2010; Peng et al., 2012) as an attempt to improve the genome assemblies. The best genome assembly was chosen for each candidate bin by comparing original bins from metagenomic assemblies with re-assemblies from SPAdes and IDBA-UD according to the genome completeness and number of duplicate single-copy genes. The above-described completeness estimation and QUAST (Gurevich et al., 2013) were used to obtain the genome statistics underlying this qualitative comparison.

Genome annotation for comparative genomics

Final draft genomic bins obtained in this study and all other genome sequences, coming from BLAST search and Slaby et al. (2017) study, used in comparative analyses were subjected to open reading frames (ORFs) prediction and subsequent annotation with the Rapid Annotation using Subsystem Technology (RAST) prokaryotic genome annotation server (Aziz et al., 2008; Overbeek et al., 2014), using the “classic RAST” algorithm. Additionally, genome annotation based on Clusters of Orthologous Groups of proteins (COGs) was performed using amino acid fasta files retrieved from RAST in searches against a local version of the COG database downloaded on 2016-02-10 (cdd.tar.gz) from ftp://ftp.ncbi.nih.gov/pub/mmdb/cdd/ (last updated 2015-05-28) (Tatusov et al., 2003), with the Rpsblast+ algorithm of BLAST 2.2.28+. Both RAST and COG-based annotations were employed in a comparison of six genomes comprising sponge-associated (n = 3) and closely-related free-living (n = 3) *Rhodospirillales* bacteria come from BLAST. Finally, EMBL annotation files of our genomic bins and their relatives were exported from RAST and uploaded on EDGAR (Blom et al., 2009; Blom et al., 2016), where comprehensive phylogenomic assessments and pan- and core-genome estimations were carried out.

Phylogenomics of sponge-associated Alphaproteobacteria

Genome-wide phylogenetic inference of alphaproteobacterial species was performed using the genome sequences (1) retrieved in this study (two from seawater, one from *S. officinalis*), (2) from the closest (cultured and uncultured) relatives of the genomes retrieved in this study (11 genomes), (3) from *S. officinalis*-associated *Alphaproteobacteria* obtained on culture medium (Chapter 3)(10 genomes), and (4) from free-living representatives closest to our *S. officinalis*-associated genome (4 genomes). This approach encompassed the analysis of 60 core genes common to all inspected alphaproteobacterial genomes (n = 28). In addition, a detailed analysis of the three bins produced in this study and their very closest matches was carried out. Here, sharper focus was placed on the relationships between the sponge-associated genomes assembled in this study and closest relatives representing (1) genome bins from the sponge host *A. aerophoba* (Slaby et al., 2017), and (2) free-living bacteria. This analysis encompassed 128 genes common to the inspected alphaproteobacterial genomes (n = 9). Phylogenomic trees were computed using PHYLIP within the EDGAR environment, and were based on predicted amino acid sequences of protein-encoding genes in the core genome of each analytical dataset. Sequence alignments were performed using muscle, after which evolutionary distances among sequences were calculated with the Kimura 2-parameter, and trees constructed using the Neighbor-Joining method. Bootstrapping tests of clade robustness were performed with 250 iterations using Consensus tree version 3.69.650. Average nucleotide (ANI) and average amino acid (AAI) identities between genome sequences were computed on EDGAR using BLAST hits between all genes common (core genes) to the genomes used in the abovementioned approaches. Further, Venn diagrams were constructed in EDGAR, considering reciprocal best BLAST hits of the coding sequences (CDSs) predicted with RAST, to determine the number of genes specific and common to sponge-associated and free-living *Rhodospirillales*.

***Geographic distribution of uncultivated, sponge-associated* Rhodospirillales**

We examined the global 16S rRNA gene operational taxonomic units (OTUs) table recently released by Thomas et al. (2016), hereafter called Sponge Microbiome (SM) dataset, to determine the worldwide distribution of *Alphaproteobacteria*, *Rhodospirillales* and *Rhodospirillaceae* phylotypes (i.e. OTUs) across marine sponges and geographic locations. Briefly, the SM dataset describes the prokaryotic diversity found in 804 marine sponge samples (encompassing 81 sponge species) collected (mostly) from shallow waters of the Atlantic, Pacific and Indian Oceans, and of the Mediterranean and Red Seas (Thomas et al., 2016) in addition to 133 and 36 seawater and sediment samples. Besides, 16S rRNA gene sequences

assembled in *Alphaproteobacteria* bins from marine sponges (Slaby et al., 2017) were subjected to phylogenetic inference with the *Rhodospirillales* OTUs present in the SM dataset (Thomas et al., 2016) to determine the geographical range of occurrence and degree of host specificity of the sponge-associated bins examined in this study. A detailed methodological description pertaining to this analysis is provided as Supporting Information (**Appendix III - File S1**).

Genome features and life strategy of* Rhodospirillaceae *sponge symbionts

Two *Alphaproteobacteria* bins assembled from *A. aerophoba* (Slaby et al., 2017) were used along with the *S. officinalis*-derived bin retrieved in this study to form a group of sponge-specific *Rhodospirillaceae*. The three closest relatives to the abovementioned genomes representing free-living *Rhodospirillaceae* were then employed in the comparative scheme. To explore key differences in functional attributes between sponge symbiotic *Rhodospirillaceae* and their closest free-living relatives, three approaches were employed on this set of six genome sequences (three symbiotic *vs*. three free-living). First, RAST annotations were used to (1) obtain an overview of basic genome features (length, GC content, rRNA and tRNA numbers etc) and (2) compare the nutrient metabolism, at the finest level of genetic resolution (i.e. annotated CDSs) between symbiotic and free-living phylotypes. Second, to discover COGs displaying significantly different abundances among symbiotic and free-living genome pools, we carried out a two-sided White´s non-parametric *t*-test (White et al., 2009) in STAMP 2.0.9 (Parks et al., 2014) at the COG-entry level, with a p-value cut-off of 0.05. We enhanced the stringency of the search by considering only those COGs which also displayed differences in mean proportions among pools (symbiotic *vs.* free-living) larger than 0.04%, corresponding to a minimum of 2-fold increase/decrease in COG relative abundances. Finally, we manually performed a Venn diagrams-assisted, qualitative search for life-style-specific COGs present exclusively either in the symbiotic or the in free-living genome pool. Using STAMP 2.0.9, heat maps illustrating the abundance distribution of highly differentiating COGs across symbiotic *vs.* free-living genomes were created, and the unweighted pair group method with arithmetic mean (UPGMA) algorithm was used to construct a dendrogram with a clustering threshold of 0.05. Additionally, the abundance distribution of COGs corresponding to genomic features often reported to be enriched in marine sponge symbionts, i.e. eukaryotic-like proteins (ELPs), clustered regularly interspaced short palindromic repeats (CRISPRs) and restriction modification (R-M) systems, was specifically determined in both genome pools.

Nucleotide sequence accession numbers

Metagenome sequences used for genome binning were deposited by Karimi et al. (2017b) in the European Nucleotide Archive (ENA) under the accession numbers ERR1103453 to ERR1103456 (*S. officinalis* samples) and ERR1103460 to ERR1103462 (seawater samples). The genomic bins retrieved in this study were deposited in ENA under the accession numbers FZLQ01000001-FZLQ01000157 (*Phyllobacteriaceae* bacterium Water-Bin73), FZLR01000001-FZLR01000323 (*Rhodospirillaceae* bacterium Spongia-Bin9) and FZLS01000001-FZLS01000274 (*Rhodobacteraceae* bacterium Water-Bin34).

Results

Phylogenomics of sponge-associated* Alphaproteobacteria *and relatives

We obtained three *Alphaproteobacteria* draft genomes with ≥ 90% completeness: one derived from the marine sponge *S. officinalis* and two from seawater (**Appendix III- Table S1**). Using genome-wide taxonomy, they were assigned to the families *Rhodospirillaceae*, *Rhodobacteraceae* and *Phyllobacteriaceae*, respectively. A summary of their genome features is shown in **Table 4-1**. No genome closely related to the *Rhodospirillaceae* genome from *S. officinalis* (**Table 4-1**, hereafter termed "*Spongia* So9") could be assembled from seawater and sediment metagenomes. Phylogenomic analysis revealed that *Spongia* So9 and genomic bins retrieved from *A. aerophoba* (Slaby et al., 2017) formed a concise phylogenetic clade exclusively comprising so-far uncultivated sponge symbionts. Members of this clade did not resemble any of the diverse, cultivated *Alphaproteobacteria* lineages retrieved recently from *S. officinalis* on culture medium (Chapter 3) (**Figure 4-1**). Closest cultured relatives to the sponge bins constituting the abovementioned clade encompassed terrestrial, plant symbiotic genera such as *Azospirillum* and *Rhodospirillum* (not shown) as well as free-living *Magnetospirillum* species from freshwater, and *Thalassospira* and *Oceanibaculum* species from seawater (**Figure 4-1**), all belonging to the family *Rhodospirillaceae* within the *Alphaproteobacteria*. While bins 65 and 129 from *A. aerophoba* (hereafter termed *Aplysina* Aa65 and *Aplysina* Aa129, respectively) were as well assigned to the family *Rhodospirillaceae* using both 16S rRNA gene and genome-wide phylogenies, the other *A. aerophoba*-associated alphaproteobacterial bins (Slaby et al., 2017) could not be classified to the family level. Because the sponge-associated genomes stood clearly apart from their closest genome matches, forming a sister group to the cultivable *Rhodospirillaceae*

Table 4-1. *Basic features of sponge-associated and seawater-derived* Alphaproteobacteria *genomes assembled in this study.*

Genome	***Spongia* So9[1]**	**Wat34[2]**	**Wat73[3]**
Family	*Rhodospirillaceae*	*Rhodobacteraceae*	*Phyllobacteriaceae*
Completeness estimation (%)	91.9	91.0	91.9
GC content (%)	66.1	37.0	54.5
Genome Size (Mb)	4.1	3.6	1.8
Number of Contigs	323	274	157
Coding sequences	4,171	3,902	1,790
Number of RNAs	42	50	37
5S rRNA	n.d.	1	n.d.
16S rRNA	n.d.	1	1
23S rRNA	n.d.	1	1
tRNAs	42	47	35

[1]*Rhodospirillaceae* bacterium Spongia-Bin9, [2]*Rhodobacteraceae* bacterium Water-Bin34, [3]*Phyllobacteriaceae* bacterium Water-Bin73. n.d., Not detected

species in our phylogenomic survey (**Figure 4-1**), we hereafter refer to this phylogenetic group of symbionts as sponge-enriched *Rhodospirillales* clade (SERC). A considerable degree of evolutionary heterogeneity was depicted within the SERC, with AAI measures ranging from 62.5% to 96.5% for the least (*Aplysina* Aa65 and *Aplysina* Aa95 from *A. aerophoba*) and most similar genomes (*Spongia* So9 and *Aplysina* Aa65) in the group when 60 core-genes were used in the analysis (**Figure 4-1**). Deeper inspection of *Spongia* So9 and its closest *Rhodospirillaceae* relatives from sponges (*Aplysina* Aa65 and *Aplysina* Aa129) and aquatic habitats (*Magnetospirillum magneticum* strain AMB-1 and *Thalassospira australica* strain NP3b2), encompassing thus only five genomes and their 647 core genes (**Figure 4-2**, **Appendix III- Table S2**), revealed that *Spongia* So9 and *Aplysina* Aa65 shared AAI and ANI values of 93.7% and 91.3%, respectively (**Appendix III- Table S2**). These values dropped to 59.5% and 77.8%, respectively, in a comparison between *Spongia* So9 and *M. magneticum* AMB-1; and to 61.2% and 81.1%, respectively, when *Spongia* So9 and *Aplysina* Aa129 were compared (**Appendix III- Table S2**).

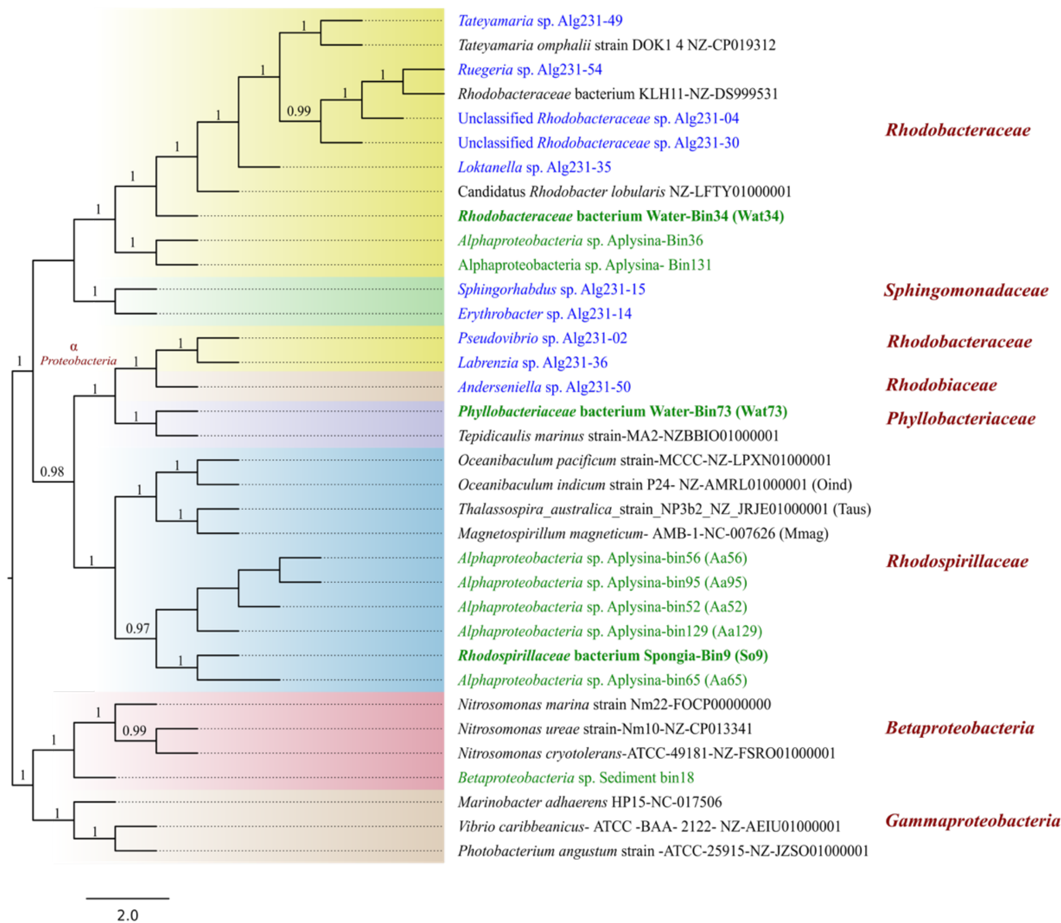


Figure 4-1. Phylogenomic tree of sponge-associated *Alphaproteobacteria* and close relatives. The tree was generated using PHYLIP within the EDGAR environment using the neighbor joining method on a matrix of Kimura distances between predicted amino acid sequences of protein-encoding genes in the genome sequences. It consists of 35 genome entries (28 from *Alphaproteobacteria* species). Sixty genes common to all genomes were used in tree construction. Genomes in green were assembled from sponge, seawater or sediment metagenomes, whereas genomes in blue represent alphaproteobacterial cultures obtained from *S. officinalis* (Chapter 3). Genomes in black were obtained from public databases. Entries marked in bold correspond to genome bins generated in this study. Bootstrap values above 70% (0.7) are shown on tree nodes.

Core- and pan-genomes of sponge-associated and free-living Rhodospirillaceae

The pan-genome of *Spongia* So9 and its closest relatives from sponges and aquatic habitats (**Appendix III- Table S2**) consisted of 13,908 genes, whereas the core-genome consisted of 647 genes (**Figure 4-3A**). *Spongia* So9 shared 3,041 genes with its closest relative *Aplysina* Aa65, supporting the closer relationship between these phylotypes as predicted by phylogenomic inference (**Figures 4-1 and 4-2, Appendix III- Table S2**). A higher proportion of phylotype-specific genes was detected for *Aplysina* Aa129 than for *Spongia* So9 and

Aplysina Aa65, and symbiotic genomes shared more genes with one another than with free-living *Rhodospirillaceae* (**Figure 4-3A**). Also within the SERC a large degree of variability was depicted (**Figure 4-3B**). Although somewhat reduced pan- and core-genomes were observed for this group, amounting to 12,862 and 536 genes, respectively (**Figure 4-3B**), the core/pan-genome ratios calculated for the genome sets analyzed in Figures 3A (symbiotic and free-living) and 3B (symbiotic only) were of similar magnitude (0.047 and 0.042, respectively).

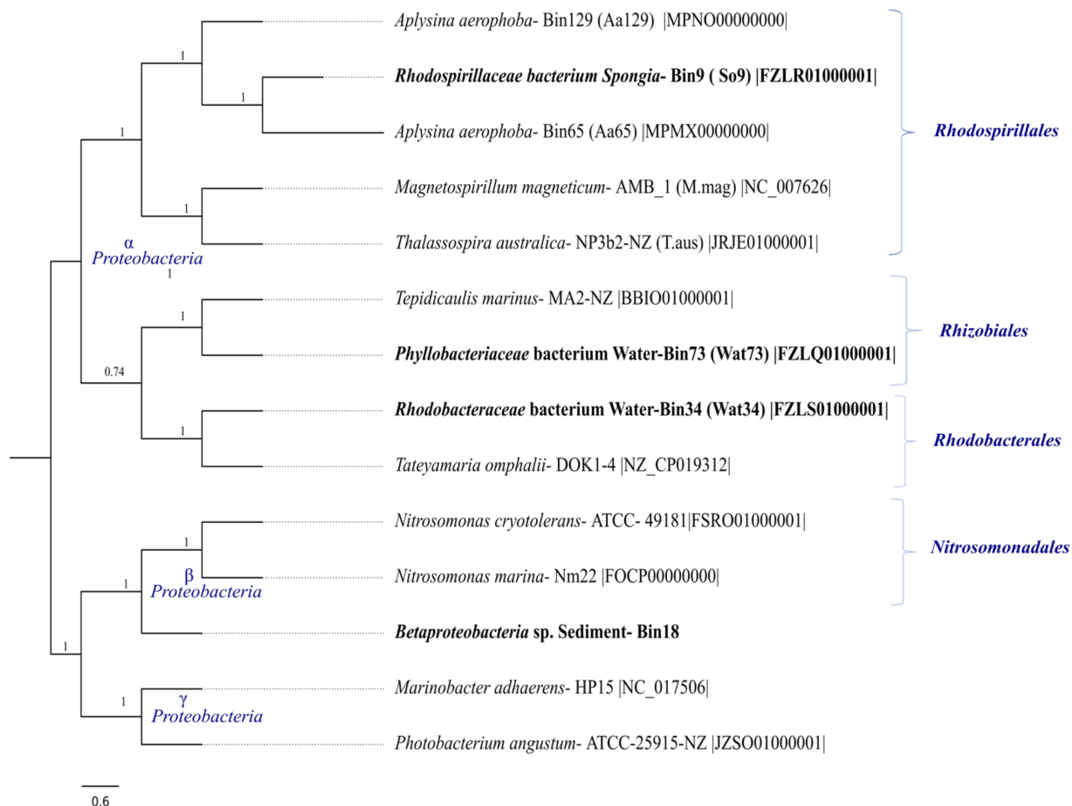


Figure 4-2. Phylogenomic tree of genomic bins retrieved in this study and their closest relatives**.** The tree was generated using PHYLIP within the EDGAR environment using the neighbor joining method on a matrix of Kimura distances between predicted amino acid sequences of protein-encoding genes in the genome sequences. It consists of 14 genome entries (nine from *Alphaproteobacteria* species), and 128 genes common to all genomes were used in tree construction. Entries marked in bold correspond to genome bins generated in this study. Bootstrap values above 70% (0.7) are shown on tree nodes.

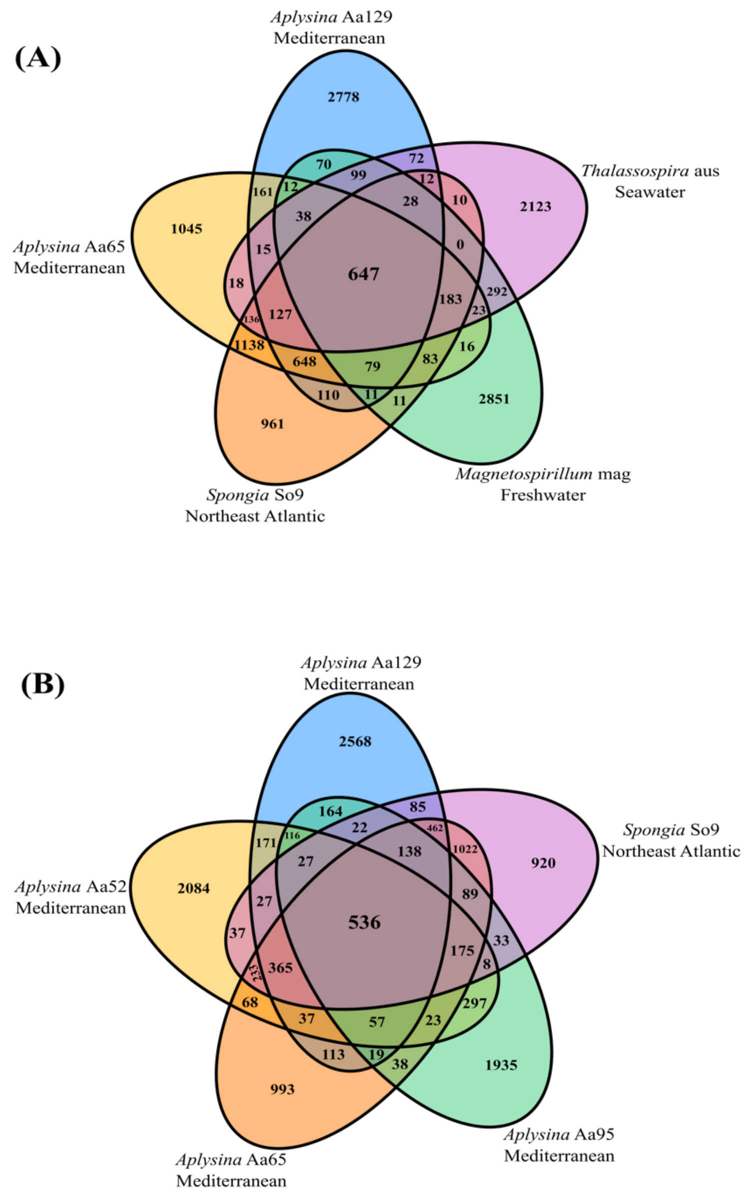


Figure 4-3. Genes shared by and specific to symbiotic and free-living *Rhodospirillaceae* genomes. Venn diagrams comparing the gene inventories of three *Rhodospirillaceae* genomes reconstructed from sponges along with their closest free-living (**A**) and further sponge-associated (**B**) relatives are shown. Diagrams were computed on EDGAR based on reciprocal best BLAST hits of the coding sequences predicted by RAST. Full names of symbiotic and free-living strains are as in the footnote to Table 4-2.

***Functional genomics of sponge-associated and free-living* Rhodospirillaceae**

Table 4-2 summarizes the basic genome features of the symbiotic and free-living *Rhodospirillaceae* genomes inspected in closer detail in this study. We found no evidence for reduced genome sizes within sponge symbionts (**Table 4-2**). GC contents were fairly high, varying from 61.1% (*Spongia* So9) to 69.5% (*Aplysina* Aa129), except for the *T. australica* NP3b2 genome (53.6%). The amount of classifiable CDSs from each genome was quite similar, with 39% to 48% of the predicted CDSs categorized in RAST subsystems, and 70% to 76% of the predicted CDSs assigned a COG entry. We observed a complete lack of genes in the RAST upper category "Motility and Chemotaxis" for the symbiotic genome pool, whereas around one hundred of such genes were found in each genome of the free-living pool (**Table 4-2**).

From both RAST and COG annotations, prediction of a functional Calvin–Benson–Bassham (CBB) cycle (CO_2 fixation via photosynthesis) could only be achieved for *M. magneticum* AMB-1. All symbiotic genomes lacked both ribulose 1,5-bisphosphate carboxylase (RuBisCO) and phosphoribulokinase (prkB) encoding genes, the major indicators of CBB-based photosynthesis in bacteria, although 21 genes involved in photorespiration (oxidative photosynthetic carbon cycle, or C2 photosynthesis), which also requires RuBisCO, could be detected in *Aplysina* Aa129. A chiefly heterotrophic metabolism was instead revealed for all free-living and symbiotic genomes through the presence of complete tricarboxylic and glyoxylate bypass cycles. The symbiotic genomes were characterized by an enrichment of genes involved in fermentative processes (86 to 93 genes in symbiotic vs. 41 to 48 genes in free-living genomes), including acetyl-CoA fermentation to butyrate and butanol biosynthesis. Similarly, genes involved in the degradation of aromatic compounds (119 to 154 genes in symbiotic vs. 23 to 58 genes in free-living genomes) including toluene, biphenyl, benzoate and salicyclate ester, among others, were more abundant in the symbiotic genomes.

Using RAST, versatile nutrient cycling and utilization capacities were uncovered for all phylotypes regardless of life-style (**Table 4-3**). Presumably, the sponge symbionts are capable of importing and utilizing organic sulfur in the form of taurine as suggested by the presence of all required genes *TauABCD*. The sponge symbionts further shared with *M. magneticum* the capacity to assimilate alkanesulfonates, however through different mechanisms: while the free-living bacteria possess the SsuABC system for alkanesulfonates binding and transport, the symbionts rely on alternative ABC-type nitrate/sulfonate/bicarbonate transporters (Pc, Ac and Prc proteins). All symbionts possessed higher numbers of arylsulfatase-encoding genes (hydrolysis of phenol sulfates; alkanesulfonate

metabolism) in comparison with free-living strains (**Appendix III- Table S3**). Congruent with the higher sensitivity of *Rhodospirillaceae* spp. to sulfides, all free-living and symbiotic genomes analyzed lacked the ability to oxidize sulfide to sulfite via the dissimilatory sulfite reductase (Dsr) pathway (a typical feature of purple-sulfur bacteria, including symbionts – see e.g. Tian et al. (2014). Sulfur oxidation capacities observed for *Spongia* So9, *Aplysina* Aa65, *O. indicum* and *M. magneticum* relied rather exclusively on the Sox pathway for thiosulfate oxidation to sulfate. Particularly, *Spongia* So9 and *Aplysina* Aa65 displayed identical and versatile genomic organization concerning their sulfur oxidation coding potential, which consisted of five sulfur oxidation proteins SoxABXYZ, the sulfite and sulfide dehydrogenases SoxD and SoxF, the thioredoxin SoxW and the regulatory protein SoxS, whereas only one Sox-encoding gene (*SoxZ*) was found in the genome of *M. magneticum*. Regarding phosphorus metabolism, all organisms possessed the genes required for the biosynthesis of phosphatase kinases 1 and 2 (ppk1 and ppk2), both of which catalyze the formation of polyphosphate (PolyP) chains through the transfer of the terminal phosphate from ATP. Likewise, all phylotypes used in the comparison were found to be capable of Poly-P breakdown via hydrolysis catalyzed by the exo-polyphosphatase Ppx, and to regulate the uptake of inorganic phosphorus (Pi) via the PHO regulon (**Table 4-3**). Conversely, only the free-living strains contained genes for assimilation and utilization of organic phosphorus in the form of alkylphosphonates (**Table 4-3**). The sponge symbionts possessed furthermore a variety of N-cycling capabilities, sharing with *M. magneticum* AMB-1 the potential to engage in nitrate and nitrite ammonification, ammonia assimilation and denitrification (**Appendix III- Table S3**). However, while the ammonification and ammonia assimilation pathways were similar among all these phylotypes, their involvement in the denitrification process differs. The symbionts reduce nitrate to nitrite via the *NarGHIJ* operon, taking part of the first denitrification step, whereas *M. magneticum* possesses genes conferring nitrite (*Nir*), nitric oxide (*Nor*) and nitrous oxide (*Noz*) reduction capabilities (**Table 4-3**)

Table 4-2. Genome features of sponge-associated and free-living *Rhodospirillaceae*.

Genomic features	***Spongia* So9**	***Aplysina* Aa65**	***Aplysina* Aa129**	***Magnetospirillum magneticum***	***Thalassospira australica***	***Oceanibaculum indicum***
Genome size (Mb)	4.06	4.28	4.83	4.96	4.27	3.95
GC content (%)	61.1	66.2	69.5	65.1	53.6	65.5
Protein coding genes	4,171	4,370	4,907	4,087	4,183	3,797
In Subsystem	1,726 (42%)	1873 (43%)	1,869 (39%)	1,723 (43%)	1,908 (46%)	1,793 (48%)
Not in Subsystem	2,445 (58%)	2,497 (57%)	3,038 (61%)	2,364 (57%)	2,275 (54%)	2,004 (52%)
In COGs	2,899 (70%)	3,083 (71%)	3,450 (70%)	3,119 (76%)	3,143 (75%)	2,883 (76%)
rRNA genes						
5S	n.d.	1	n.d.	2	1	1
16S	n.d.	1	1	n.d.	1	1
23S	n.d.	1	n.d.	2	1	1
tRNA genes	42	46	38	90	56	44
Motility and chemotaxis	0	0	0	112	98	91

Spongia So9, *Rhodospirillaceae* bacterium bin9 from *Spongia officinalis* (this study); *Aplysina* Aa65, *Rhodospirillaceae* bacterium bin65 from *Aplysina aerophoba* (Slaby *et al.*, 2017); *Aplysina* Aa129, *Rhodospirillaceae* bacterium bin129 from *Aplysina aerophoba* (Slaby *et al.*, 2017); Free-living strains: *Magnetospirillum magneticum* strain AMB-1, *Thalassospira australica* strain NP3b2 and *Oceanibaculum indicum* strain P24. n.d.: Not detected.

Table 4-3. Nutrient cycling features in sponge-associated and free-living *Rhodospirillaceae* genomes.

N-cycling category/subcategory/subsystem	*Spongia* So9[1]	*Aplysina* Aa65	*Aplysina* Aa129	*Magnetospirillum magneticum*	*Thalassospira australica*	*Oceanibaculum indicum*
Nitrogen metabolism						
Cyanate hydrolysis	0	0	0	0	0	7
Dissimilatory nitrite reductase	0	0	0	13	0	0
Nitrosative stress	3	1	1	0	1	2
Nitrate and nitrite ammonification	6	6	7	24	6	6
Ammonia assimilation	18	19	26	9	10	9
GS-GOGAT pathway	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Denitrification	4	4	5	31	0	4
Nitrate reductases NarGHIJ	*yes*	*yes*	*yes*	*no*	*no*	*yes*
Nitrite (Nir), nitric oxide (Nor) and nitrous oxide (Nos) reductases	*no*	*no*	*no*	*yes*	*no*	*no*
Others (amidase)	2	2	5	0	0	0
Sulfur metabolism						
Inorganic sulfur assimilation	0	0	0	32	11	0
Release of DMS from DMSP	0	0	0	0	1	0
Sulfur oxidation	13	13	0	1	0	6
Sulfate reduction-associated complexes	0	0	0	10	0	0
Thioredoxin-disulfide reductase	4	5	4	6	9	9
Galactosylceramide and sulfide metabolism	2	2	2	3	0	0
Organic sulfur assimilation and utilization	21	26	51	161	0	9
Taurine utilization	*yes*	*yes*	*yes*	*no*	*no*	*no*
Alkanesulfonates utilization	*yes*	*yes*	*yes*	*yes*	*no*	*yes*
Phosphorus metabolism						
High-affinity PO_4 transporter and PHO regulon	7	8	8	11	9	8
Polyphosphate	4	4	4	3	4	3
Poly-P synthesis (Ppk1-2) and hydrolysis (Ppx)	*yes*	*yes*	*yes*	*yes*	*yes*	*yes*
Alkylphosphonate utilization	0	0	0	4	12	11

[1]Full names of symbiotic and free-living strains are as in the footnote to Table 4-2. Values in cells correspond to number of CDSs classified in each subcategory/subsystem using RAST annotation. The presence of specific genes/complete pathways in selected subsystems is highlighted in italics.

Adaptations to (sponge) symbiotic and free-living life-styles

The genes from the six *Rhodospirillaceae* genomes (**Table 4-2**) were assigned to 2,435 COGs (**Appendix III- Table S3**). Of these, 1,521 COGs (62.5%) occurred in both the sponge-associated and free-living genome pools, whereas 205 and 709 COGs were specific to the sponge-associated and free-living pool, respectively (**Appendix III- Figure S1**). The sponge-associated genome pool contained a higher degree of functional conservation among phylotypes (63.7% of COGs in the core) than the free-living genome pool (50.9%) (**Appendix III- Figure S1**). From a two-sided White's *t*-test, 287 COGs displayed significantly different frequencies among sponge-associated and free-living genome pools (**Appendix III- Figure S2**). Of these, 141 COGs were sponge-enriched and 146 COGs were sponge-depleted (**Appendix III- Tables S4A** and **B**). Sixty-six sponge-enriched and 67 sponge-depleted COGs displayed > 2-fold differences in relative abundance between pools, and were more thoroughly inspected. We fetched 13 COGs representing four categories of ABC-type transport systems enriched in the symbiont genomes. Among these, we highlight the transport system proteins TauABC which mediate taurine uptake (**Table 4-3**), and the leucine-specific transport system proteins LivFGHM (**Figure 4-4**).

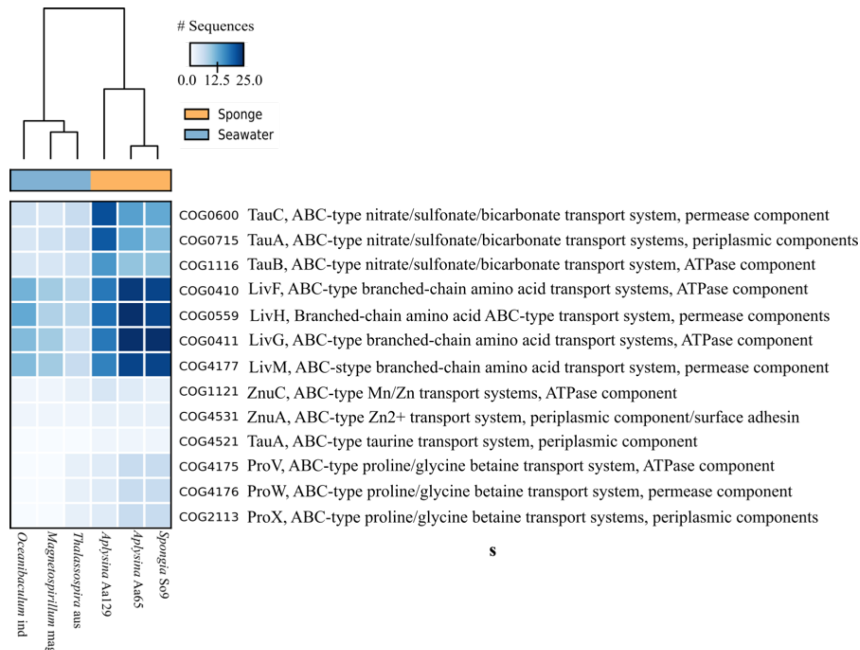


Figure 4-4. Heat map of ABC-type transport systems significantly enriched in sponge symbiont genomes. The heat map illustrates shifts in the absolute abundance of CDSs assigned to each COG category listed in each of the six genomes (three sponge-associated, three free-living; see Table 4-2) used in the comparative analysis. Full names of symbiotic and free-living strains are as in the footnote to Table 4-2.

Furthermore, we found that COG1028 (FabG, dehydrogenases), well represented in both sponge and free-living genomes (**Figure 4-5**), was nevertheless the COG entry most pronouncedly enriched in the symbiotic genome pool (**Appendix III- Table S4**). Specific hydrolases, amidases and dioxygenases (including taurine TauD dioxygenase), along with COG2124 (cytochrome P450 family of oxidative enzymes) and COG5285 (fumonisin polyketide biosynthesis) were further identified as secondary metabolism features (COG class "Q") enriched in sponge symbiont genomes (**Figure 4-5**). Several other COGs notably enriched in the sponge symbiont genomes could be pinpointed (**Appendix III- Table S4**). For instance, COG0346 (lactoylglutathione lyase GloA) and COG0625 (glutathione S-transferase GST) displayed 3 and 2.4-fold increases in abundance in sponge-associated versus free-living genomes, respectively. Also noteworthy were COGs involved in toxin-antitoxin (TA) systems, such as COG3093 (VapI, plasmid maintenance system antidote protein; 7-fold increase in sponge-associated genomes), COG4118 (Phd, antitoxin of TA stability system, 7-fold increase), and COG3549 (HigB plasmid system maintenance protein) (**Appendix III- Figure S2, Table S4**).

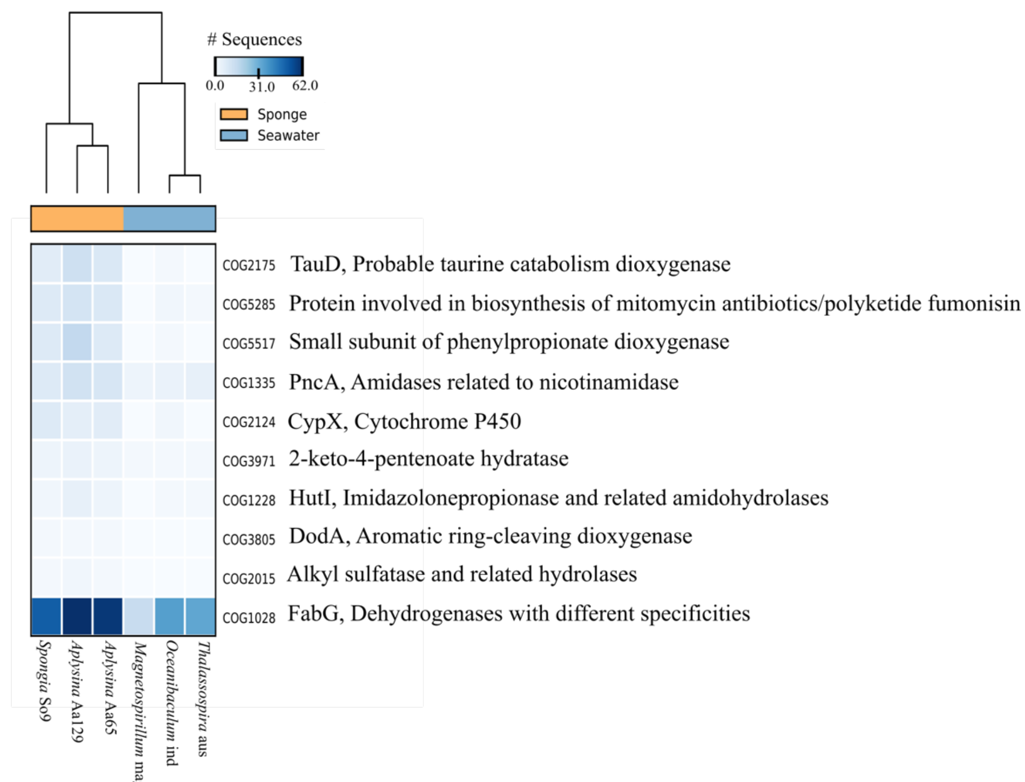


Figure 4-5. Heat map of COGs in the class "secondary metabolites biosynthesis, transport, and catabolism" found to be significantly enriched in sponge symbiont genomes. The heat map illustrates shifts in the absolute abundance of CDSs assigned to each COG category listed in each of the six genomes (three sponge-associated, three free-living; see Table 4-2) used in the comparative analysis. Full names of symbiotic and free-living strains are as in the footnote to Table 4-2.

Further screening for sponge-depleted COGs allowed us to gather evidence for the (complete or nearly complete) loss of motility, chemotaxis and biofilm formation capacities among sponge symbionts (Tables S4). This was reflected by a list of COGs involved in (1) biosynthesis/assembly of chemotaxis proteins (e.g. COG0840, Tar protein) and of response regulators linked to environmental sensing and chemotaxis control (COG0784, Chey-like receiver; COG2204, Atoc response regulator); (2) flagellar motor protein assembly (e.g. COG1360 and COG1291) and flagellin biosynthesis (e.g. COG1344); and (3) synthesis and degradation of c-di-GMPs which act on the regulation of all the above-mentioned traits (e.g. COG2199 and COG2200, c-di-GMP synthetase and phosphodiesterase, respectively). All these COGs were among the most sponge-depleted functions, usually showing no hits in the symbiont genomes but being highly abundant among the free-living bacteria (**Appendix III-Table S4**). Likewise, specific features related with cell membrane composition and integrity (e.g. COG0438, RfaG glycosyltransferase; COG0859, RfaF LPS heptosyltransferase) and iron transport (e.g. COG1269 and COG0848, ferric ion uptake) were noticeably de-selected from the sponge symbiont genomes.

Finally, among the genomic signatures usually reported as diagnostic of a typical sponge symbiotic life-style, CRISPR-Cas and ankyrin repeats, along with one specific R-M system (COG3440), surfaced as elements clearly more frequent in the sponge-associated genomes, whereas no evidence for the enrichment of WD40, TPR and LRR motifs, and other R-M systems in sponge genomes could be gathered (**Table 4-4).**

Table 4-4. COG-based abundance distribution of phage defense mechanisms and eukaryotic-like protein repeats across symbiotic and free-living genomes.

COG	description	*Spongia* So9[1]	*Aplysina* Aa65	*Aplysina* Aa129	*Magnetospirillum magneticum*	*Thalassospira australica*	*Oceanibaculum indicum*
CRISPR-Cas systems							
COG1203	CRISPR-associated helicase Cas3	0	1	1	0	0	0
COG1343	CRISPR-associated protein Cas2	0	1	0	0	0	0
COG1518	CRISPR-associated protein Cas1	0	1	1	0	0	0
COG3512	CRISPR-associated protein, Cas2 homolog	1	0	0	0	0	0
Total CDSs in COGs		1	3	2	0	0	0
Restriction-modification systems							
COG0286	HsdM, Type I restriction-modification system methyltransferase subunit	3	1	2	0	2	1
COG0732	HsdS, Restriction endonuclease S subunits	0	1	1	0	1	1
COG1002	Type II restriction enzyme, methylase subunits	2	0	0	1	0	0
COG1403	McrA, Restriction endonuclease	1	2	2	1	1	1
COG1715	Mrr, Restriction endonuclease	1	0	1	0	0	1
COG3440	Predicted restriction endonuclease	2	2	1	0	0	0
COG4096	HsdR, Type I site-specific restriction-modification system, R (restriction) subunit	0	1	0	0	1	1
Total CDSs in COGs		9	7	7	2	5	5
Ankyrin repeats							
COG0666	Arp, FOG: Ankyrin repeat	1	1	0	0	0	0
Tetratricopeptide repeats							
COG0457	NrfG, FOG: TPR repeat	1	2	1	6	1	2
COG0790	FOG: TPR repeat, SEL1 subfamily	2	3	4	7	3	2
Total CDSs in COGs		3	5	5	13	4	4
WD40 repeats							
COG1520	FOG: WD40-like repeat	1	1	1	2	1	1
COG2319	FOG: WD40 repeat	0	1	2	0	0	1
Total CDSs in COGs		1	2	3	2	1	2

[1]Full names of symbiotic and free-living strains are as in the footnote to Table 2. Values correspond to the number of CDSs classified in each COG category.

Discussion

Since the earliest molecular-based studies of the marine sponge microbiome (e.g. Hentschel et al. (2002)), evidence has been accumulating for a central relevance of *Alphaproteobacteria* species as abundant and diverse members of this distinct symbiotic consortium (Taylor et al., 2007b; Schmitt et al., 2012). To this date, several lineages within the class have been broadly recognized as either sponge-specific (Hardoim et al., 2009; Hardoim et al., 2012; Simister et al., 2012) or sponge-enriched (Taylor et al., 2013), and have been suggested to be vertically transmitted across successive host generations (Schmitt et al., 2007). Current availability of large-scale, next-generation sequencing data permits standardized, molecular-based inspection of the diversity and abundance of microorganisms in nature at unprecedented levels of resolution and comprehensiveness. In this study, screening of the SM dataset (Thomas et al., 2016) was instrumental to solidify the status of the *Alphaproteobacteria* as major players of the marine sponge microbiome at a global scale due to their considerable relative abundance across several hosts. Furthermore, we unveiled large, previously unrecognized phylotype diversification within sponge-associated *Rhodospirillales* and *Rhodospirillaceae*, with up to > 500 OTUs assigned to these groups worldwide (**Appendix III- File S1**). Inspection of so-far uncultivated, sponge-specific *Rhodospirillaceae* lineages via genome binning allowed us to address the evolution and adaptive features of this unique group of symbionts in detail, contributing with a growing body of contemporary surveys to deepen our understanding of the likely roles of several sponge symbiotic lineages in a cultivation-independent manner (Tian et al., 2014; Burgsdorf et al., 2015; Moitinho-Silva et al., 2017; Slaby et al., 2017).

Intriguingly, our phylogenomic assessment revealed an unanticipated, reasonably large degree of diversification within the studied sponge-associated genomic bins. We gathered evidence for both genome conservation of *Rhodospirillaceae* symbionts from different hosts and biogeographical settings (*Spongia* So9 and *Aplysina* Aa65) and genome diversification among *Rhodospirillaceae* and closely related symbionts inhabiting the same host at a single location (all *A. aerophoba* genomic bins) (**Figures 4-1** and **4-3**). In fact, the pattern of gene gain and loss gleaned here for the symbiont genomes was of the same magnitude as that observed for an equally-sized group of symbiont and free-living *Rhodospirillaceae* genomes (**Figure 4-3**), and much larger than that observed by Burgsdorf *et al.* (2015) for a group of four closely-related "*Candidatus* Synechococcus spongiarum" (*Cyanobacteria*) phylotypes, with core/pan-genome ratios of 0.042 and 0.26 obtained for the *Rhodospirillaceae* (this study) and

"*Ca.* Synechococcus spongiarum" symbiont groups (Burgsdorf et al., 2015), respectively. This is consistent with a picture of inter-species and inter-general diversification, to the least, within the pool of sponge-derived, alphaproteobacterial genome bins analyzed here (**Figures 4-1** and **4-3A**). Comparatively, a striking core/pan-genome ratio of 0.94 was obtained for six closely-related strains of the human pathogen *Klebsiella pneumoniae*, with values dropping to 0.79 and 0.72 when strains from subspecies *K. pneumoniae ozaenae* and *K. rhinoscleromatis* were subsequently added to the analysis (Caputo et al., 2015). In contrast, the core/pan-genome ratio of 20 *Escherichia coli* strains was found to be 0.11 (Touchon et al., 2009). Altogether, these data suggest high genome-wide rearrangement (and eventual adaptive irradiation) within sponge symbionts sharing common ancestry both below (Gao et al., 2014; Burgsdorf et al., 2015) and above (this study) species level. Although biogeographical isolation and host species-driven selection could be evoked as main forces underlying this pattern, we advocate that much of the observed variation may as well result from (i) microniche partitioning across environmental gradients within the host and (ii) independent host colonization events leading to unique host persistence trajectories and genetic exchange with vicinal sponge symbionts. These emerge as more plausible causes of the divergence observed e.g. between *Aplysina* Aa65 and *Aplysina* Aa129, both from the same host and location (Slaby et al., 2017). When core- and pan-genomes were delineated on the basis of annotated COGs instead of predicted CDSs, much higher degrees of conservation were observed among the sponge-associated genomes (**Appendix III- Figure S1**). Although this hints at a considerable degree of functional equivalence among the studied symbionts, still a large proportion of CDSs could not be assigned to COG categories (or RAST subsystems), making it challenging to infer the complete functional divergence that derives from the gene composition heterogeneity observed among the strains.

As usual for members of the family *Rhodospirillaceae* (Baldani et al., 2014), the sponge-associated genomes had an elevated GC content. Correlations between DNA length and GC content are known to vary across the tree of life depending on the studied taxa (Li and Du, 2014). Nevertheless, it has been postulated that obligate host-associated bacteria usually contain small genomes with low GC content (McCutcheon and Moran, 2012). Our data stands in sharp contrast with this perspective (**Table 4-2**), and this opposition could be interpreted as an indication for a recently-established rather than ancient association. Recent associations are usually regarded as examples of facultative, secondary symbioses whereas ancient associations would correlate with obligate, primary symbioses (Moya et al., 2008). While the functional commonalities found between the sponge-associated and free-living phylotypes studied here

are supportive of a recent association hypothesis, it is rather difficult to favor the facultative symbiosis picture given the virtual inexistence, based on 16S rRNA gene assessments (**Appendix III- Table S5**), of the symbiotic lineages in sediments and seawater. In fact, several adaptive features to symbiotic living, detailed below, have been uncovered, which distinguish these symbionts from their free-living relatives. We moreover perceived that sponge-depleted COGs more often represented complete loss of function in the sponge-associated genome pool (e.g. absence of several motility and chemotaxis COGs) whereas sponge-enriched COGs more often resulted from quantitative differences in abundance of orthologous genes present in both pools. Assuming that *Rhodospirillaceae* sponge symbionts evolved from free-living ancestors, adaptation to the sponge host involved transition from motility to sessility resulting from the disposal of costly cell appendages (Martínez-García et al., 2014) - and their coding DNA - of pivotal relevance to the fitness of several particle- and host-associated bacteria (see e.g. Utada et al. (2014); Rossez et al. (2015)).We speculate that such a transition might take place quickly once the symbionts have accrued the necessary genetic machinery, and ensuing metabolic circuitry, that make them apt as true members of the sponge endosymbiotic consortium. In this context, the large metabolic versatility typical of both marine *Rhodospirillaceae* and the sponge symbiotic consortium would provide the evolving symbionts with a formidable gene repertoire with which fine adjustment to the novel conditions could be promoted. Availability of diverse carbon and nutrient sources as a consequence of the host's filtering activity (Maldonado et al., 2012) would drive the maintenance of large and versatile genomes for optimal nutrient acquisition. Mechanisms used by bacteria to silence foreign DNA with GC percentages lower than their own (Navarre et al., 2006) could act as a "purifying" selective force keeping GC contents high as observed for the symbionts inspected in this study. This process could be of relevance in the marine sponge microbiome given the high exposure to viral particles and horizontal gene transfer proneness within this consortium (Taylor et al., 2007a; Fan et al., 2012).

In terms of nutrient metabolism, one feature obviously enriched in the symbiont genomes regards their capability to import and utilize organic sulfur compounds. Particularly, the complete set of all proteins required for taurine-specific import (TauABC) and desulfonation (TauD) was markedly more abundant in the symbiotic genomes (**Figure 4-4, Appendix III- Table S4**). This indicates that taurine (2-aminoethanesulfonic acid) metabolism constitutes an important adaptive feature of sponge symbiotic *Rhodospirillaceae*. Taurine occurs naturally as an abundant compound in the tissues of a wide range of animals, from marine invertebrates (Allen and Garrett, 1971) to mammals (Schuller-Levis and Park, 2003),

and potential for taurine utilization was recently reported for an uncultivated sponge symbiont in the order *Chromatiales* (*Gammaproteobacteria*) (Gauthier et al., 2016). Taurine-conjugated fatty acids and N-acyl taurines from diverse marine sponges such as *Axinella* (Huang et al., 2013), *Callyspongia* (Huang et al., 2015), *Ircinia* (Emura et al., 2006) and *Geodia* (Olsen et al., 2016) have been increasingly reported in recent years, providing evidence for diverse and available organic sulfonated compounds that may be utilized by sponge symbionts. TauD desulfonates taurine as an alternative sulfur source to sulfite, succinate and 2-aminoacetaldehyde. Through the action of a sulfite oxidase (e.g. COG2041), the potentially toxic sulfite can be oxidized to sulfate, which can be easily excreted or incorporated into cellular components. In addition, the higher frequency of arylsufatase A (COG3119)-encoding genes in the sponge-associated genome pool highlights the potential of these *Rhodospirillaceae* spp. to utilize sulfated polysaccharides - known to be present in marine sponges (Vilanova et al., 2009) -, a feature that has been suggested to be of relevance for diverse sponge-associated bacteria (Slaby et al., 2017). We observed that our symbionts may be capable of utilizing carnitine - a quaternary ammonium compound present in the mesohyl matrix of sponges - in agreement with Slaby et al. (2017) who reported that many of the typical sponge symbionts may possess this ability. However, carnitine usage was not verified in this study to be a distinguishing feature of symbiotic *Rhodospirillaceae* (White´s p-value > 0.05; see COG1804 in Table S3 for details) as observed for sulfonated (e.g. taurine) and sulfated organic compounds. We moreover found that the sponge-associated *Rhodospirillaceae* are capable of transforming both organic (ammonification) and inorganic (nitrate reduction) forms of nitrate. The latter reaction, which in our survey was an exclusive feature of the sponge-associated genomes, corresponds to the first step of the denitrification process, and has been considered a signature feature of marine sponge metagenomes (see e.g. Fan et al. (2012)). This suggests that symbiotic *Rhodospirillaceae* spp. may rank as one major mediator of denitrification processes within the marine sponge holobiont, whereas coupling between ammonification and ammonium assimilation processes may enable these organisms to obtain constant supplies of nitrogen that can be incorporated into cell biomass. Similarly, *Rhodospirillaceae* may be key players in poly-P formation within marine sponges, a process postulated to significantly contribute to phosphorous sequestration in benthic ecosystems (Zhang et al., 2015).

While fine-tuning their nutrient-scavenging repertoire and metabolism to establish themselves within the marine sponge holobiont, concomitant selective pressures might have acted on membrane and cell wall composition, and on trans-membrane protein diversity, of the evolving *Rhodospirillaceae* spp. symbionts. This is reflected by the clear-cut enrichment of

several and specific ABC transport systems (**Figure 4-4**) along with the acute reduction of glycosyltransferases (e.g. RfaF and RfaG) involved in outer membrane biogenesis in the symbiotic genome pool. Besides using ATP hydrolysis to pump molecules across cellular membranes (Wilkens, 2015), ABC transporters may also be involved in the regulation of osmolarity and membrane integrity (Higgins, 2001). Several ABC importers have been identified as key factors in the acquisition of essential nutrients and energy for growth (Skaar et al., 2004; Cui and Davidson, 2011), and indeed the number and variety of transporter motifs correlate with the life-style of microorganisms (Ren and Paulsen, 2007). Apart from the taurine system highlighted above, particularly intriguing was the selection, in the symbiotic genomes, of the full high-affinity branched-chain amino acid transport system LivFGHMK, involved in leucine uptake and metabolism (**Figure 4-4, Appendix III- Table S4**). Evidence has been gathered for the spread of this transport system from alphaproteobacterial (*Rhodospirilalles* and *Rhodobacterales* species) to gammaproteobacterial (*Chromatiales* species) sponge symbionts via horizontal gene transfer (Tian et al., 2014; Gauthier et al., 2016). Along with our findings, this illustrates how positive, vertical selection of nutritional factors in a given lineage may spread laterally across other genomes in a community setting, a feature that may be promoted in highly dense and integrative symbiotic consortia.

Two COGs involved with the metabolism of the antioxidant glutathione were remarkably enriched in the sponge-associated genomes (**Appendix III- Table S3**). While the glutathione lyase GloA (COG0346) plays a role in the detoxification of methylglyoxal in bacterial cell metabolism (MacLean et al., 1998), glutathione S-transferases (GSTs, COG0625) are involved in different metabolic processes including protection against oxidative stress, cellular detoxification and antimicrobial drug resistance (Guengerich et al., 1996; Allocati et al., 2009). Bacterial GSTs also may ensure the correct folding, synthesis and degradation of enzyme complexes (Vuilleumier, 1997). The much greater frequency of GloA and GST-encoding genes in the symbiotic genomes therefore hints at oxidative stress and antimicrobial agents as key selective forces shaping the evolution of *Rhodospirillaceae* sponge symbionts, and likely of sponge-associated bacteria in general. Taken together, our data suggest that the major contribution of these symbionts to host fitness relate with nutritional exchange and host detoxification (through e.g. ammonium assimilation). Considering their versatile nutrient assimilation profiles, participation in numerous geochemical processes, and widespread occurrence and abundance in association with marine sponges, we posit that these so-far uncultivated symbiotic bacteria could be

relevant not only to sponge holobiont homeostasis but also to the functioning of reef ecosystems.

Supplementary data

Supplementary materials and explanations for this chapter can be found in Appendix III and on digital format of the thesis (CD).

Acknowledgements

The authors acknowledge Jorge M.S. Gonçalves and Joana R. Xavier for their invaluable help in sampling procedures and sponge identification, respectively.

Funding

This work was supported by the Portuguese Foundation for Science and Technology (FCT) through the research grants [PTDC/BIA-MIC/3865/2012 and PTDC/MAR-BIO/1547/2014 to RC], Federation of European Microbiological Societies (FEMS) for conceding a research grant [FEMS-RG-2015-0115 to EK] to support her stay at GEOMAR to accomplish this work, and the European Union's Horizon 2020 research and innovation program under grant agreement [679849 ('SponGES') to BMS and UH]. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Ethical Statement

For this study, *in situ* sampling of microorganisms from marine invertebrates without a nervous system was performed, and as such the work was exempt from ethical approval procedures according to the current Portuguese legislation (Decreto-Lei nº 113/2013). The research did not occur within privately owned or protected areas, neither it did involve endangered or protected species. The sampling methodology privileged minimally invasive procedures in accordance with the guidelines of the European Directive 2010/63/EU.

Author Contributions

RC designed the study draft; EK, BMS, RC and UH designed the final study; RC and UH provided resources and materials; EK and RC performed the experiments; EK, BMS, ARS, JB and RC analyzed the data; EK and RC wrote the main manuscript text and EK prepared figures and tables. All authors reviewed the manuscript.

Chapter 5

General Discussion and Outlook

General discussion

Sponges carry complex communities of symbiotic microbes believed to provide many benefits to their hosts through different ways (Lackner et al., 2017). Although sponges feed on microbes, it is now known that several microbial lineages have adapted to live within their sponge hosts by preventing being ingested while concomitantly performing diverse functions suited to this particular microenvironment (Nguyen et al., 2014; Slaby et al., 2017). Besides, compelling evidence has been gathered for vertical transmission of symbiotic bacteria through sponge generations (Schmitt et al., 2007; Lee et al., 2009; Bright and Bulgheresi, 2010), suggesting that these bacteria play crucial roles in sponge physiology. Moreover, microbial community composition within a given host species has been documented to be stable across geographical distances (see e.g Pita et al. (2013)) and time (Hardoim and Costa, 2014b), favoring the notion of regulatory mechanisms promoting sponge holobiont homeostasis regardless of environmental circumstances. In contrary, it has been often observed that the sponge-associated symbiotic consortium is likely to change considerably according to the host species (Hardoim et al., 2012; Pita et al., 2013; Thomas et al., 2016), indicating that host physiology and modes of living (e.g. erect, eroding, encrusting, with thick or thin mesohyl) play a major role in shaping the structure of its own microbiome.

It is now clear that sponges harbor remarkable microbial diversity which involves common and specific microbial populations (Thomas et al., 2016). It is also evident that sponge symbiotic bacteria have evolved specialized metabolic traits (such as the ability to import and utilize sulfated polyssacharides) that allow them to thrive in particular microniches within their hosts (Slaby et al., 2017). However, we current lack in-depth knowledge of (1) the precise functional capacity of each of the diverse symbiotic bacteria that inhabit sponges (the diversity-function linkage problem, that is: who does what?), (2) the extent to which these functional traits may contribute to host fitness and (3) how each of the major symbiotic lineages interact and communicate with their hosts to establish and maintain their symbiosis. This thesis approaches the above-mentioned issues by revealing the unique genetic signatures of the microbiome associated with the economically important model organism *Spongia officinalis* (see below) and by delving specifically into the functional traits and adaptive features of cultivated and uncultivated alphaproteobacterial symbionts inhabiting this host.

The marine sponge microbiome comprises several so-far uncultivated, yet highly abundant, lineages of symbiotic bacteria within manifold phyla such as *Acidobacteria*,

Actinobacteria, *Chloroflexi*, *Poribacteria*, *Proteobacteria* and *Tectomicrobia*, among others (Taylor et al., 2007a; Simister et al., 2012; Wilson et al., 2014; Thomas et al., 2016). Within the phylum *Proteobacteria*, many previous reports have unveiled a diverse range of sponge-specific or sponge-enriched alphaproteobacterial lineages likewise recalcitrant to cultivation to this date (Hardoim et al., 2009; Simister et al., 2012; Naim et al., 2014; Thomas et al., 2016). Revealing the precise functional features of these manifold difficult-to-cultivate symbiotic bacteria, fostering knowledge of their metabolism, potential use in biotechnology, and contribution as host fitness-enhancing factors is one of the major challenges of today's sponge microbiology research. Previous research performed by the Microbial Ecology and Evolution Research Group (MicroEcoEvo) at CCMar has consistently reported on the prevalence of so-far uncultivatable alphaproteobacterial lineages belonging to the families *Rhodospirillaceae* and *Rhodobacteraceae* in Dictyoceratida sponges (i.e. keratose sponges possessing a protein fibers skeleton) from the Algarve shore (Hardoim et al., 2012; Hardoim et al., 2014; Hardoim and Costa, 2014a), prompting our team to approach this relevant sponge-associated group in more detail. In this context, this thesis combines both cultivation-independent and alternative cultivation-dependent procedures to advance our knowledge of the coding potential of phylogenetically distinct sponge-associated symbionts, resulting in an unprecedented comparative genomics endeavor revealing the functional and adaptive attributes of diverse *Alphaproteobacteria* species found in association with *Spongia officinalis*.

Although much research has addressed different sponge species to explore or describe their associated microbial communities (Hentschel et al., 2012; Costa et al., 2013; Hardoim and Costa, 2014b; Horn et al., 2016; Thomas et al., 2016; Moitinho-Silva et al., 2017; Slaby et al., 2017), it is clear that further knowledge must be fostered for a number of other important sponge species that may serve as key model organisms in the study of host-microbe relationships, especially if we consider that there are more than 8,500 sponge species described to date and as many yet to be described (Van Soest et al., 2012). One compelling case is that of *Spongia officinalis* Linnaeus 1759, the first-ever described sponge species. *S. officinalis* possesses an Atlanto-Mediterranean geographical range and is well-known for its historical use by humans as a bath sponge (Dailianis et al., 2011). Several biologically-active natural products, particularly terpenoids, have been consistently obtained from *S. officinalis* (Gonzalez et al., 1984; Manzo et al., 2011), highlighting its potential use as sources of novel biotechnological appliances. However, whether such compounds are effectively produced by the sponge itself or by its symbionts has not yet been disclosed. In spite of the acknowledged societal, economical and biotechnological relevance of *Spongia officinalis*, knowledge of its

associated microbial community is scarce and apparently restricted, to the best of our knowledge, to one inaugural study performed by Bauvais et al. (2015). Using PCR-DGGE fingerprinting, these authors revealed that the structure of the *S. officinalis* symbiotic consortium diverged from that of seawater communities. They moreover inspected the heavy metal resistance capacity of several *S. officinalis*-associated bacterial cultures, documenting tolerance to nickel, copper, zinc or lead for some of the alphaprotobacterial taxa investigated in this thesis such as *Pseudovibrio* and *Ruegeria* (Bauvais et al., 2015).

Given the abovementioned, the work compiled in this thesis dissects the *S. officinalis* microbiome in an unprecedented fashion by (1) unveiling its functional and taxonomic diversity using high-end, cultivation-independent metagenome sequencing technologies, (2) revealing the coding potential of hard-to-culture and phylogenetically unique alphaproteobacterial isolates obtained from this host, and (3) disclosing the adaptive features and functional attributes of a so-far uncultivable alphaproteobacterial symbiont using *in silico* genome reconstruction via binning of metagenomic reads.

The aim of this section is to integrate the most relevant outcomes from the three original research chapters presented in this thesis to achieve a systematic understanding of the *S. officinalis* holobiont (and of the marine sponge holobiont at large) functioning and homeostasis, focusing on the participation of, and roles played by, symbiotic *Alphaproteobacteria* species in this dynamics.

Comparative metagenomics reveals unique life strategies of the **Spongia officinalis** *endosymbiotic consortium (Chapter 2)*

Next generation sequencing (NGS) technologies have experienced dramatic improvements in throughput and final data quality in the past few years. Consequently, these techniques have been essential for the development of the microbial ecology field, which intrinsically deems with complex systems composed by thousands (if not hundreds of thousands) of invisible organisms, several of which (so-far) uncultivable in the laboratory and interacting with one another in multiple and often unknown ways. Nowadays, metagenomics and genomics-based procedures, from DNA extraction to NGS to bioinformatics methodologies, are being used as tools of trend in this area. These approaches have accordingly paved the way to access the diversity and function of sponge-associated microbial communities, with pioneering shotgun metagenomic sequencing studies revealing for the first time a series of so-called “genetic signatures” underlying microbial adaptation to live within sponges (Thomas et al., 2010; Fan

et al., 2012). Among these, the enrichment of CRISPR-Cas and R-M systems involved in antiviral defense, and of a range of eukaryotic-like proteins (ELPs) possibly mediating molecular interactions between host and symbionts, emerged as pivotal features apparently "unique" to sponge-associated microbiomes. In this thesis, the distribution of these and several other genetic elements found to be enriched in sponge-associated microbial consortia was diligently verified across our model sponge host *S. officinalis* and its environmental surroundings using shotgun metagenome NGS (**Chapter 2**). Although one challenging aspect of these approaches is their dependence on available reference genomes for precise annotation of protein functions from gene sequences (Nielsen et al., 2014), they provide an alternative route with which the functionality of (particularly) uncultivable microorganisms (the case of most of the dominant sponge symbionts) can be assessed. As such, metagenomic NGS is substantially contributing to advance our understanding of the functions of the vast uncultivable majority that makes-up the microbiome of marine sponges (Siegl et al., 2011; Fan et al., 2012; Wilson et al., 2014; Burgsdorf et al., 2015; Slaby et al., 2017).

Taking advantage of this approach, we could reveal, in the *S. officinalis* microbiome, multiple functional traits/adaptive features promoted in this particular consortium in comparison with seawater and sediment microbial metagenomes. These include genes encoding for polyketide and terpene/terpenoid biosynthesis, a vast carbohydrate degradation repertoire, the presence of eukaryotic-like proteins, biosynthesis of CRISPR-Cas and R-M endonucleases, Type IV secretion proteins, ABC transporters and plasmids. Some of these functions were found to be enriched in the microbial metagenomes of other sponge hosts, being generally interpreted as "specific" genomic signatures of sponge-associated bacteria (Fan et al., 2012; Horn et al., 2016; Thomas et al., 2016). However, previous studies delivering metagenomic functional and taxonomic profiling of marine sponge microbiomes have only contrasted this consortium with the vicinal seawater, what is understandable given the fact that sponges pump thousand liters of water per day into their bodies, and thus planktonic microbial communities inevitably make part of sponge microbiome dynamics, probably as source of both food and symbiotic bacteria. Nevertheless, our study in **Chapter 2** demonstrates that several of the functional traits and bacterial taxa enriched in the sponge microbiome possess higher residual abundances in marine sediments than in seawater. We evoke sponge cell shedding and higher availability of solid surfaces and microbial cell densities in sediments and sponges (than in seawater) as plausible explanatory mechanisms underlying the observed, closer functional resemblance between sponge and sediment microbiomes in comparison with sponge and seawater microbiomes. Our observation bears implications to our understanding of the

evolutionary mechanisms involved in bacterial adaptation to sponge and eukaryotic hosts in general, and the patterns observed here should be verified in other sponge hosts as well as in other symbiotic systems such as corals, ascidians, bryozoans etc.

From the taxonomic standpoint, most of the abundant endosymbiotic bacteria detected in *S. officinalis* were found to possess higher relative abundances in sediment than in seawater microbiomes, further strengthening the notion that the participation of the former habitat as source and sink of marine sponge symbionts shall not be undervalued. The taxonomic approach employed in this study was fundamental to reveal the diversity of *Alphaproteobacteria* lineages associated with *S. officinalis*, whereby uncultivated members of the families *Rhodobacteraceae* and *Rhodospirillaceae* were clearly the most prevalent. As observed for the other symbiotic lineages, *S. officinalis*-associated *Alphaproteobacteria* were usually more abundant in sediments than in seawater, with the exception of one particular *Rhodobacteraceae* lineage (OTU 442) found to be highly abundant in seawater and very well represented both in the sponge and in sediments (Appendix I -Table S3B). In the following chapters, this thesis approaches the phylogeny, evolution and function of sponge-associated *Alphaproteobacteria*, including *Rhodobacteraceae* and *Rhodospirillaceae* species, in a cultivation-dependent and -independent manner, respectively, using comprehensive functional and comparative genomics.

The* Alphaproteobacteria *symbiotic community in* S. officinalis *- a comparative and functional genomics approach

Genomes from cultivated bacteria

In spite of the large amount of information obtained from novel, cultivation-independent molecular techniques, microbial cultivation, especially if providing alternatives to more traditional protocols, can be a powerful complementary method leading to the retrieval and laboratory maintenance of novel, readily identifiable organismal biomass that can be promptly described and categorized, both genotypically and phenotypically. Therefore, using alternative strategies to cultivate sponge symbiotic bacteria, and thus helping to overcome the inherent difficulties in cultivating these microorganisms, was one of the main aims of this PhD thesis. In **Chapter 3**, we coupled alternative cultivation to full genome sequencing of sponge symbiotic bacteria, aiming at linking bacterial identity and function for a broad phylogenetic range of symbionts retrieved *in vitro*. For example, some typical features of sponge symbionts such as the production of bioactive secondary metabolites, which is believed to play an

important role in host defense (Hentschel et al., 2012), can be predicted from genome-wide analyses and subsequently straightforwardly tested in the laboratory. In **Chapter 3**, we adapted a general culture isolation approach in an attempt to access as much phylogenetic diversity of slow-growing bacteria as possible, optimally including rare and abundant symbionts, which have not been cultivated so far. Although this approach was very helpful for the retrieval of so-far understudied as well as novel bacterial species, it was considered to fail in enabling the laboratory domestication of highly abundant sponge-associated bacterial lineages, a difficult-to-overcome problem inherent to all attempts to cultivate marine sponge symbionts so far (Sipkema et al., 2011; Hardoim et al., 2014). Instead, the approach, which relied on the usage of lower incubation temperatures, lower amounts of organic carbon provided in the medium and longer incubation periods, enriched almost exclusively for rarer sponge-associated bacteria usually detected superficially using regular cultivation-independent NGS, as observed previously by our team (Hardoim et al., 2014). Interestingly, most of the isolates retrieved belonged to the phylum *Alphaproteobacteria* spanning several different families within this class, allowing us to deeply interrogate the genomes of cultivable sponge-associated *Alphaproteobacteria* in a comprehensive fashion, emphasizing on functional attributes of members of the highly versatile *Roseobacter* clade (family *Rhodobacteraceae*), and to discriminate between diverse alphaproteobacterial taxa on the basis of their genotypic features. Taxonomically, the panel of isolates subjected to genome sequencing comprised representatives of well-studied sponge associates with sequenced genomes such as *Ruegeria* (Zan et al., 2012) and *Pseudovibrio* (Bondarev et al., 2013; Alex and Antunes, 2015) as well as less-studied alphaproteobacterial taxa such as *Anderseniella*, *Labrenzia*, *Sphingorhabdus*, *Loktanella*, *Tateyamaria* and two *Rhodobacteraceae* lineages non-classifiable at the genus level, thus contributing with highly novel genomes and cultured bacterial biomass to be deeply explored in future studies (see below). In summary, we found that genes enconding for N-acyl-L-homoserine lactone synthetases were enriched in *Roseobacter* clade genomes in comparison with non-*Roseobacters*, in agreement with the known quorum-sensing capacities reported for members of this clade (see e.g. Zan et al. (2012) for an example from a sponge symbiont). Conversely, genes encoding for type IV pilus biosynthesis (involved in host colonization and adherence) and ELPs (involved in host-microbe molecular interactions) were more frequent in non-*Roseobacter* genomes, leading us to argue that the latter group may comprise bacterial taxa more efficient in engaging in symbiotic relationships than *Roseobacter* clade lineages well known for their occurrence as free-living, planktonic bacteria (**Chapter 3**). Moreover, we observed that nearly all isolates (eight out of ten) were found to be putative producers of

terpenoids, leading us to hypothesize that terpenoid biosynthesis, frequently reported for *Dictyoceratida* sponges, may in fact be boosted by their alphaproteobacterial symbionts.

Putative novel bacterial species are regularly isolated from sponges (Lee et al., 2006; O'Halloran et al., 2011; Jackson et al., 2015). As suggested above, in **Chapter 3** we described the isolation of two likely novel species in the family *Rhodobacteraceae*. Also, strain *Anderseniella* Alg231-50 is, to the best of our knowledge, the first representative of the *Anderseniella* genus with a full genome sequenced. Indeed, nearly all information about this genus on databases is based on 16S rRNA sequences with only one strain of marine organism (soft coral mucus) reported so far (stand: 14th of November 2017). From all the alphaproteobacterial lineages submitted to full genome sequencing, *Anderseniella* Alg231-50 displayed the highest representativeness in *Spongia officinalis* according to the metagenome-genome mapping approach used in our study to infer relative abundances in the total sponge microbial community. This suggests that, among the cultures studied here, this strain may correspond to the most relevant in ecological terms. Interestingly, all of the isolated strains did not rank among highly dominant taxa in seawater, sediments or sponge microbiomes, reinforcing the notion raised by Hardoim et al. (2014) that current sponge symbiont cultivation platforms usually promote the enrichment of rare to only moderately abundant bacteria on culture plates, what somehow can be explored as an useful alternative to shed light on the coding potential of microbial dark matter from sponge microbiomes and other marine settings. Indeed, much of the metabolism and genome architecture of the ten sponge-associated alphaproteobacteria studied here can be explored in future studies aiming at promoting or confirming, in the laboratory, the activities inferred from genome mining. This approach seems to be particularly promising regarding the *in vitro* biosynthesis of natural products such as terpenes and polyketides, and their corresponding inhibitory activities, from these cultures.

Genome reconstruction of a sponge-specific Rhodospirillaceae *symbiont*

With the latest progress in sequencing technologies and the availability of straightforward and simple programs, genome binning and mining procedures have been ever increasing in speed and reliability. As many sponge-associated microbes cannot be simply cultivated with existing microbial isolation procedures, we benefited from this *in silico* approach to reconstruct the genome of an abundant *Alphaproteobacteria* (family *Rhodospirillaceae*) symbiont of *Spongia officinalis* from our sponge-associated microbial metagenomes (**Chapter 4)**. By analyzing the binned genome (termed "So9" in **Chapter 4**) using a comparative genomics approach that

included other sponge-associated *Rhodospirillaceae* genomes built *in silico* (Slaby et al., 2017) along with the genomes of free-living relatives, we could verify and confirm, in the So9 genome and its closest sponge-associated relatives, the presence of several adaptive features reported in **Chapter 2** as genetic signatures of the overall sponge microbiome. Therefore, as a consequence of our study in **Chapter 4**, we now know that some of these features are typical of the genome of symbiotic *Alphaproteobacteria* in the *Rhodospirillaceae* family. These include the presence of polyketide synthase, terpenoid synthase, sulfatase and ABC transporter encoding genes, for instance, which were enriched in *Rhodospirillaceae* genomes and could be verified in our metagenomes in **Chapter 2.** Besides, the lack of motility and chemotaxis traits observed in **Chapter 2** for the total microbial community could be verified again among these symbiotic *Alphaproteobacteria*, in a remarkable example of gene and cell accessory disposal underlying the adaptation of bacteria to a host-associated life-style involving the switching from a motile to a sessile state. We moreover suggest that fine-tuning of nutrient metabolism has been a key aspect in the evolutionary history of sponge-specific *Rhodospirillaceae* lineages given the enrichment, in their genomes, of genes involved in the import and metabolism of organic sulfur compounds such as taurine.

In **Chapter 4** we further revealed, using an *in silico* approach, the widespread occurrence of diverse *Rhodospirillaceae* species across several sponge hosts and geographic locations (Atlantic Ocean, Mediterranean Sea, Pacific Ocean), further strengthening the relevance and likely stability of this particular symbiotic relationship. Finally, the recovery of this draft *Rhodospirillaceae* genome allowed us to gain insights into the adaptation and metabolic versatility of these yet-uncultured endosymbiotic bacteria in *S. officinalis*, and it is likely that the functional traits revealed in this study may be conserved among prevalent *Rhodospirillaceae* species inhabiting other sponge hosts.

Conclusions

Collectively, the findings of this thesis illustrate:

Chapter 2: "Comparative metagenomics reveals the distinctive adaptive features of the **Spongia officinalis** ***(Porifera, Dictyoceratida) endosymbiotic consortium"***

1- Microbial diversity and function in *S. officinalis* sharply differs from those of vicinal sediments and seawater, highlighting the uniqueness of the *S. officinalis* endosymbiotic consortium and of marine sponges in general (through comparative analyses, performed in this study, between the microbiomes of *S. officinalis* and those of other sponge hosts).

2- In spite of the striking divergence in microbiome structure mentioned above, closer functional and phylogenetic resemblance was found between sponge and sediment than between sponge and seawater microbial communities, an observation that bears implications to our understanding of the evolution of sponge-microbe and host-microbe symbiotic relationships in marine ecosystems.

3- Among literally thousands of individual protein domains/functions (that is, IPR entries) found to oscillate significantly in relative abundance across *S. officinalis*, sediment and seawater microbiomes, IPR entries collectively involved in the coding CRISPR-Cas, restriction endonucleases, plasmids, polyketide synthases, terpene/terpenoid synthases, Type IV secretion proteins and ABC transporters were highlighted as some of the most significant contributors shaping the distinct functionality of the *S. officinalis* endosymbiotic consortium.

4- The distinctiveness of the *S. officinalis* symbiotic consortium could be readily observed at a very coarse level of taxonomic resolution, since the relative abundance of > 20 microbial phyla changed significantly across sponge, sediments and seawater. Among these, the bacterial phyla *Acidobacteria*, *Gemmatimonadetes*, *Poribacteria*, *Chloroflexi*, *Actinobacteria* and *Nitrospirae*, and the candidate phyla SBR1093, PAUC34f and AncK6, were pronouncedly enriched in the sponge samples.

***Chapter 3: "Functional genomics of cultivated sponge-associated* Alphaproteobacteria *reveals shared and unique traits underlying a bimodal symbiotic-free-living life-style*"**

5- The use of an oligotrophic medium (MG50) along with low-temperature (18°C) and prolonged incubation (8 weeks) permitted the cultivation of diverse sponge-associated *Alphaproteobacteria* not usually retrieved in most traditional cultivation attempts, likely by suppressing/preventing overgrowth of copiotrophic bacteria such as *Vibrio* spp. that tend to dominate culture plates rich in organic carbon.

6- Genes encoding for ABC transporters, thioredoxins, nitrogen regulation, peroxiredoxins, type II/IV secretion systems and glutathione S-transferases were identified as core genes present in all analyzed *Alphaproteobacteria*, suggesting functional redundancy within a broad phylogenetic panel of bacterial symbionts of sponges, what ultimately may confer functional resilience to the *S. officinalis* holobiont.

7- At the COG-level, genes encoding for arylsulfatase A (an enzyme breaking down sulfatides), sulfate permeases, sulfur transferases, 3'-phosphoadenosine 5'-phosphosulfate (PAPS) 3'-phosphatase and sulfite reductases were detected as common features of all cultured *Alphaproteobacteria*.

8- Genes encoding for vitamin B biosynthesis were shared by the cultivated *Alphaproteobacteria* symbionts of *S. officinalis*. The presence of riboflavin synthases (alpha and beta chains) and of pyridoxal phosphate biosynthesis protein-encoding genes validate the synthesis of vitamin B among these symbionts.

9- Terpene synthesis and polyketide synthase (PKS)/non-ribosomal peptide (NRPS) encoding gene clusters were found for the vast majority (eight out of ten) of the cultivated *Alphaproteobacteria* cultures, revealing their potential to contribute to the complex biochemistry known for keratose sponges and, eventually, to host defense against natural enemies.

Chapter 4: "Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges"

10- A *Rhodospirillaceae* sp. draft genome with ≥ 90% completeness was assembled from the microbial metagenome (**Chapter 2**) of the marine sponge *S. officinalis*, allowing the appraisal of the functional attributes of a so-far uncultivable, dominant sponge-associated (alpha-) bacterium in this thesis.

11- The genome reconstructed from *S. officinalis* (*Spongia* So9) and the genomic bin retrieved from the marine sponge *A. aerophoba (Aplysina* Aa65) were found to be

highly similar in spite of their geographical distance and host origin, sharing AAI and ANI values of 93.7% and 91.3%, respectively.

12- Symbiotic *Rhodospirillaceae* genomes were characterized by an enrichment of genes involved in fermentative processes, degradation of aromatic compounds, importing and utilizing organic sulfur in the form of e.g. taurine (*TauABCD* enconding genes) and sulfatides (arylsulfatase-encoding genes), and toxin-antitoxin (TA) systems.

13- Evidence for the (complete or nearly complete) loss of motility, chemotaxis and biofilm formation capacities among sponge *Rhodospirillaceae* symbionts.

14- CRISPR-Cas and ankyrin repeats, along with one specific R-M system arose as anti-viral genetic elements more frequent in sponge-associated than free-living *Rhodospirillaceae* genomes.

Altogether, these results underline that the use of multiple methods is worthwhile to increase our understanding of the function and to describe the life-strategies of specific groups of host-associated bacteria, fostering knowledge of their potential contributions as individual symbionts to host fitness, the possibilities of inter-dependent metabolism among members of one particular phylogenetic clade (i.e., *Alphaproteobacteria*), and the extent to which diverse lineages within one such clade display functional redundancy, an aspect that may bear important implications to holobiont dynamics and homeostasis. In the specific context of the marine sponge-microbe association, such a taxon-oriented endeavor helps disentangling the microbial identity-function (who does what?) conundrum typical of highly diverse microbial communities often consisting of several so-far uncultivable lineages. In this regard, particularly noteworthy was the congruence between a taxon-idenpendent, open-metagenomics sequencing study (**Chapter 2**) and the following taxon-oriented surveys implemented in this thesis (**Chapters 2** and **3**), identifying in individual alphaproteobacterial symbionts several genotypic features found to characterize the uniqueness of the collective marine sponge microbiome. Indeed, combining metagenomics and genomics approaches with microbial cultivation holds much promise in the exploitation of the metabolism and biotechnological potential of marine sponge microorganisms, therefore scientists still need to sufficiently associate these two techniques as “two sides of the same coin” (Gutleben et al., 2017).

Future perspectives

While this thesis has applied a combination of cultivation-independent and cultivation-dependent approaches to better explore and understand the marine sponge microbiome using one thus-far understudied yet relevant model organism, providing new insights into the structure and function of the *S. officinalis* microbiome and the possible roles of *Alphaproteobacteria* in sponges, still many more aspects remain to be addressed for a more accurate picture of the *S. officinalis* microbial consortium and its functional attributes, and obviously microbial communities inhabiting sponges in general. Considering the particular approach applied in this thesis, which describes one attempt - among manifold other possibilities - to increase the cultivability of marine sponge associates, it is clear that more research is needed to fully elucidate the coding potential of other cultivated or yet-to-be cultivated bacterial genomes, with the ultimate goal of incrementing our capacity not only to read the potential roles that symbionts may play in this particular association, but also to advance our understanding of bacterial metabolism. In this context, a shift in research effort and investment, from the simple identification of bacterial cultures using 16S rRNA gene sequencing to full genome sequencing, as attempted in **Chapter 3**, can substantially improve our knowledge of the genetic diversity and metabolism of such cultures, with direct consequences to a new perspective of host-microbe and microbe-microbe molecular interactions and signaling mechanisms within this system. If geared with novel symbiont cultivation platforms enabling manipulation of an ever-increasing phylogenetic breadth of microorganisms in the laboratory, full genome sequencing of symbionts will sharply enlarge the comprehensiveness of high-quality genomic databases, facilitating laboratory- and *in silico*-based studies of marine microorganisms.

Moreover, the data published in this thesis open new perspectives to the formulation and design of new media and alternative cultivation attempts for capturing more sponge-associated bacteria besides the *Alphaproteobacteria* lineages described here. Members of other so-far uncultivated phyla from sponges such as *Acidobacteria* and *Actinobacteria* (which we have shown to be particularly high abundant in *S. officinalis*) have been successfully retrieved from other environmental matrices such as soils and seawater, and adjustments in methodology can be made towards the retrieval of symbiotic lineages within such groups using specific, metagenomics-derived information gained here and in past studies. Different culture conditions using e.g. media enriched with organic sulfur compounds whose efficient assimilation and

utilization seem to be a common feature of several so-far uncultivable sponge symbionts, in combination with different incubation temperatures, may improve efficiency in capturing further sponge-associated bacterial lineages. This way, continuous (alternative) cultivation attempts may significantly contribute to lead, in the long-term, to a clearer understanding of the role and metabolism of the typically hard-to-culture, yet highly abundant bacterial symbionts of sponges. Bearing this in mind, during the course of the research leading to this thesis additional cultivation experiments have been performed. The data retrieved have been partially analyzed in what will constitute a future scientific article deriving from this research yet not included in the main body of the present thesis. These cultivation experiments have been done using two media named VXA (double-strength VL55 medium (Sait et al., 2002)) and 1/10R2A (low nutrient medium) which succeeded in enabling access to further so-far uncultivated bacterial symbionts of sponges belonging to the groups *Alphaproteobacteria*, *Gammaproteobacteria*, and *Actinobacteria* from *Spongia* sp. specimens. This has encouraged us to deposit more 15 genomes in public databases that will be addressed in a future study in a comparative fashion, as performed in **Chapter 3**. Nevertheless, as mentioned above the genomics and metagenomics datasets generated in this study equipped us with an enormous amount of detailed information with which many more suitable media and incubation conditions can be designed.

Besides, some further ideas could be tested, from the technical standpoint, which I would have considered for designing future experiments. For example, it could be interesting to use third generation sequencing technologies such as Pacific Biosciences' (PacBio) single molecule realtime (SMRT) sequencing (Eid et al., 2009), which delivers very long reads and has been shown to generate suitable data for the reconstruction of genomes from metagenomes using sequence-dependent binning approaches (Slaby, 2017; Slaby et al., 2017). Applying technologies enabling the retrieval of long reads can for instance increase the success rates in reconstructing the genomes of other important, uncultivated bacteria which are not that prevalent in marine sponges or other symbiotic systems in general.

Moreover, in as far as innovative studies relying on genome and metagenome sequencing can increase our knowledge of complex microbiomes, current sponge microbiology research can largely benefit from metatranscriptomics (see. e.g. Kamke et al. (2010) and Moitinho-Silva et al. (2014)) and experimental transcriptomics (from cultivated symbionts) surveys to better identify those taxa and functions truly performing and being expressed *in situ* and in controlled microcosms. The fact that scarce are the studies that so far made full use of these possibilities is to some extent suggestive of the technical challenges that need to be

overcome to increase the reliability of the data generated using these approaches. Eventually and likewise, the future application of proteomics technologies bears promise in fostering knowledge of actively expressed functions in the marine sponge microbiome, permitting analyses of this truly unique symbiotic system with ever-increasing levels of detail (Horgan and Kenny, 2011).

Appendices

Appendix I

Chapter 2 complementary analysis and supplementary materials

File S1

Preliminary data analyses and data validation with alternative analytical pipelines

Metagenome data assembly

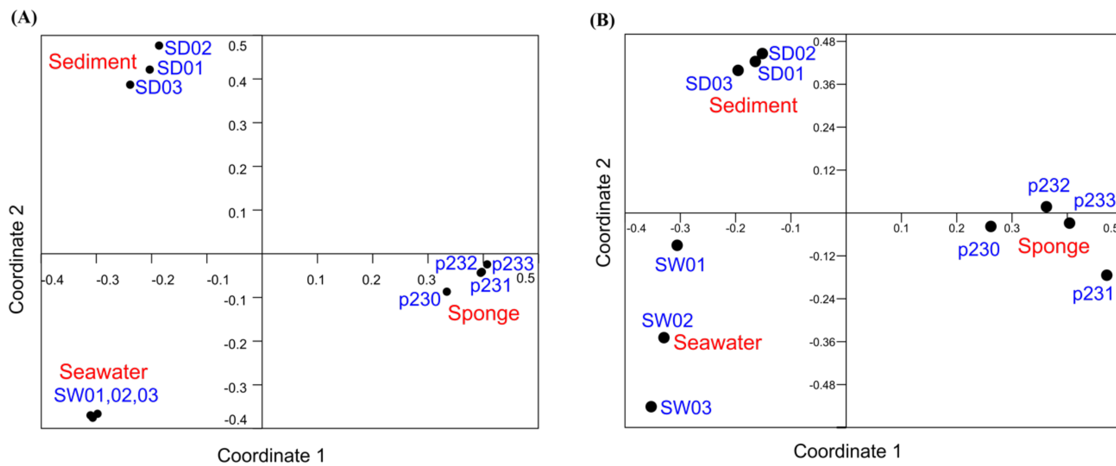
Metagenome assembly was performed using MetaVelvet (Namiki et al., 2012) under default settings. This procedure enabled us to obtain higher numbers of assembled nucleotides from complex sediment metagenomes than assembly using IDBA_UD, whereas performance of both assemblers was comparable for seawater and sponge metagenomes (data not shown). Still, because of the low proportions of reads actually used in the generation of contigs, and the low average contig length obtained for sediment samples (**Table I-FS1**), we opted to perform our core comparative metagenome analysis using unassembled data (see Figures. 1 to 5 in the main text, and Figures. S2 and S3 in the supplements). Metagenome assemblies obtained with MetaVelvet were subjected to COG-based annotations. While COG profiles from both unassembled and assembled data were used to (1) verify whether patterns described for IPR profiles using unassembled data could be reproduced using other analytical pipelines, COG profiles from assembled data only were used to (2) compare the functional profiles obtained for the metagenomes analysed in this study with those published by Fan et al. (2012) to screen for common functions (i.e. COG entries) across different sponge hosts (see below).

Appendix I- Table FS 1. MetaVelvet assembly of *Spongia officinalis* (SP), sediment (Sd) and seawater microbial metagenome samples.

Sample	# contigs	Reads used (%)	N50 (bp)	Largest contig (bp)	Mean contig (bp)	Total assembly (bp)
SP230	122 435	51.51	631	45 968	510	62 460 207
SP231	31 651	52.63	1 615	42 085	1 128	35 694 965
SP232	152 240	51.75	385	70 453	393	59 774 639
SP233	91 672	53.69	842	84 626	570	52 254 976
Sd_01	1 437 527	19.10	97	1 203	76	109 116 486
Sd_02	2 097 956	26.43	83	1 006	65	135 483 210
Sd_03	2 858 973	28.35	88	1 151	60	171 780 628
Seawater_01	759 222	68.95	320	22 151	102	77 133 569
Seawater_02	75 462	35.63	600	18 025	320	24 150 712
Seawater_03	68 013	32.23	666	35 041	344	23 372 397

COG annotations of unassembled and assembled reads

Within MG-RAST, we performed COG-based annotations for both the (1) unassembled dataset (also used in IPR annotations with the EBI metagenomics pipeline (EMG) - see main text) and (2) the assembled dataset obtained with MetaVelvet (**Table AS1.1**). In both cases, Principal Coordinates Analysis (PCoA) performed on the functional profiles recovered the overall trend observed with the analysis of InterPro (IPR) functional categories (see main text, Figure 1), depicting highly contrasting, significantly different microbiomes in terms of function ($p < 0.0006$) across the three biotopes (**Figure FS1.1**). Particularly, the sponge symbiotic consortium was found to significantly differ from seawater and sediment microbiomes for both assembled and unassembled datasets, whereas no significant, pairwise differences between seawater and sediment microbiomes were found (**Figure FS1.1**). In contrast with results obtained using the IPR database (main text), COG annotations did not reveal the specific pattern of closer functional resemblance between sponges and sediments than between sponges and seawater ($p > 0.05$ for differences between Bray-Curtis dissimilarities). Thus, this particular outcome may slightly shift depending on the data processing pipelines and databases being used. Differently from the IPR annotation using EMG, COG annotations for both assembled and unassembled reads resulted in skewed numbers of reads with assigned functions among the different samples. Although the data transformation procedure used in this study corrects quite well for highly skewed data, it is important to consider this aspect when interpreting the results retrieved with COG annotations. Further, for all samples analyzed, much higher numbers of reads could be assigned functions using the IPR database in comparison with the COG database (22,156,186 vs 4,559,625 reads with function across the whole unassembled dataset, respectively). Likewise, the total number of IPR entries uncovered from the whole unassembled dataset was as well much higher than the total number of COGs (10,272 IPR vs. 2497 COG entries). Altogether, these outcomes suggest that the use of the EMG processing pipeline resulted in a more refined annotation of our data due to both (1) higher equitability among the total number of reads analysed in each sample (see **Table S1**) and (2) much higher numbers of annotated reads computed along with higher diversity of functions (i.e. IPR entries) retrieved for all samples, substantiating our choice to use this particular analysis in our main results. Yet the COG annotations were very useful to contrast our data with COG-based profiles obtained previously for other sponge hosts, such as *Rhopaloides odorabile*, *Cymbastela concentrica* and *Cymbastela coralliophila*, all characterized by Fan et al. (2012).

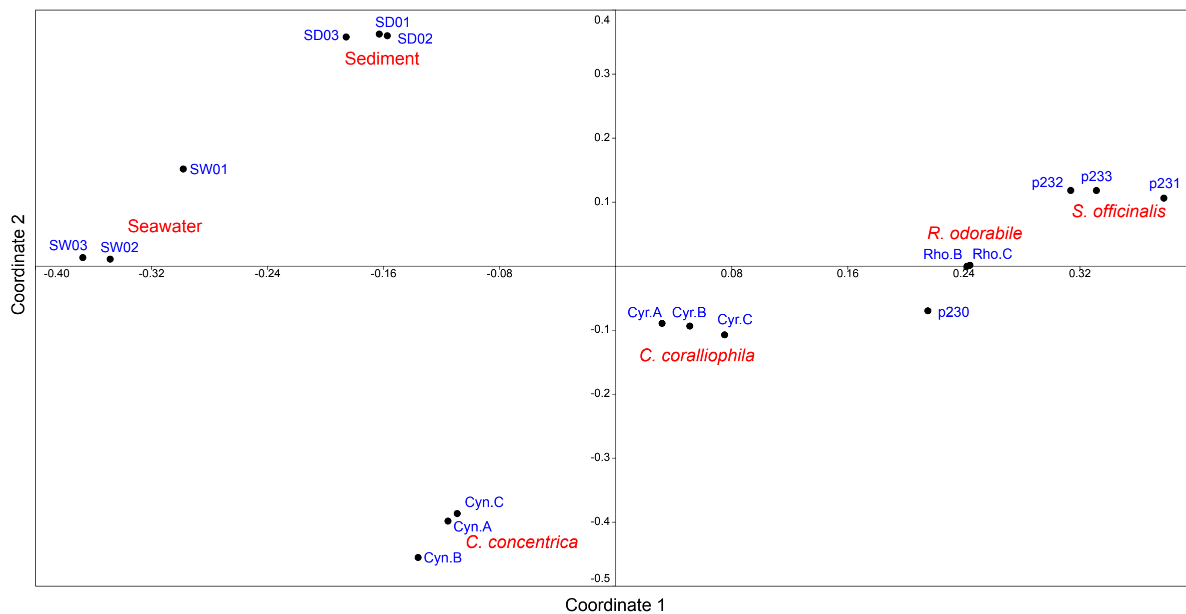


Appendix I- Figure FS 1. Principal Coordinate Analysis (PCoA) of functional microbial community profiles across biotopes based on COG annotations of unassembled (A) and assembled (B) metagenomes. Community ordinations were based on pairwise Bray-Curtis dissimilarities calculated from normalized data, considering oscillations of relative COG abundances among samples. Analyses were performed on COG community profiles extracted from the metagenomes using MG-RAST. The first and second coordinates explain 39.7% and 36.1% (A) and 51.2% and 24.6% (B) of the total dataset variation within unassembled and assembled metagenomes, respectively. Significance values resulting from permutational analysis of variance (PERMANOVA) applied to the corresponding dissimilarity matrices are as follows. Overall differences among groups: $p = 0.0004$ and 0.0005 for unassembled and assembled metagenomes, respectively. Pairwise significances: sponges were found to be different from seawater and sediment metagenomes in both datasets, with p values < 0.03 and < 0.04 for unassembled and assembled data, respectively. No significant difference was found between sediment and seawater functional profiles in both datasets ($p > 0.05$).

Contrasting functional profiles of different sponge hosts

As mentioned above, functional COG profiles retrieved for metagenomes assembled in this study were compared with those retrieved by Fan et al. for the sponge hosts *Rhopaloeides odorabile* (belonging to the order Dictyoceratida, as *Spongia officinalis*), *Cymbastela concentrica* and *Cymbastela coralliophila* (belonging to the order Axinellida). In spite of the differences in sampling and sample processing procedures, next-generation sequencing methodology and throughput, data processing, and size of the metagenome libraries produced in this and in the Fan et al. (2012) studies, PCoA on Hellinger transformed data (for both non-rarefied and rarefied datasets) revealed a interesting gradient in functional profiles resembling the taxonomic relatedness of the sponge hosts, whereby *S. officinalis* and *R. odorabile* shared greater similarities with one another than with either of the *Cymbastela* hosts (**Figure FS-2**). When lumped together as one single group, marine sponge microbial metagenomes were found

to possess functional profiles significantly different from those retrieved for the sediment and seawater metagenomes sequenced in this study (**Figure FS-2**). Here, the same trend observed for IPR profiles retrieved from unassembled reads could be gathered: marine sponges had, collectively, significantly greater similarity with sediments (average Bray-Curtis dissimilarity: 23.4%) than with seawater metagenomes (average Bray-Curtis dissimilarity: 26.1%) (p = 0.0042). However, it is worth mentioning, as explained above, that the different data generation methods and analytical pipelines may influence these results to some extent.

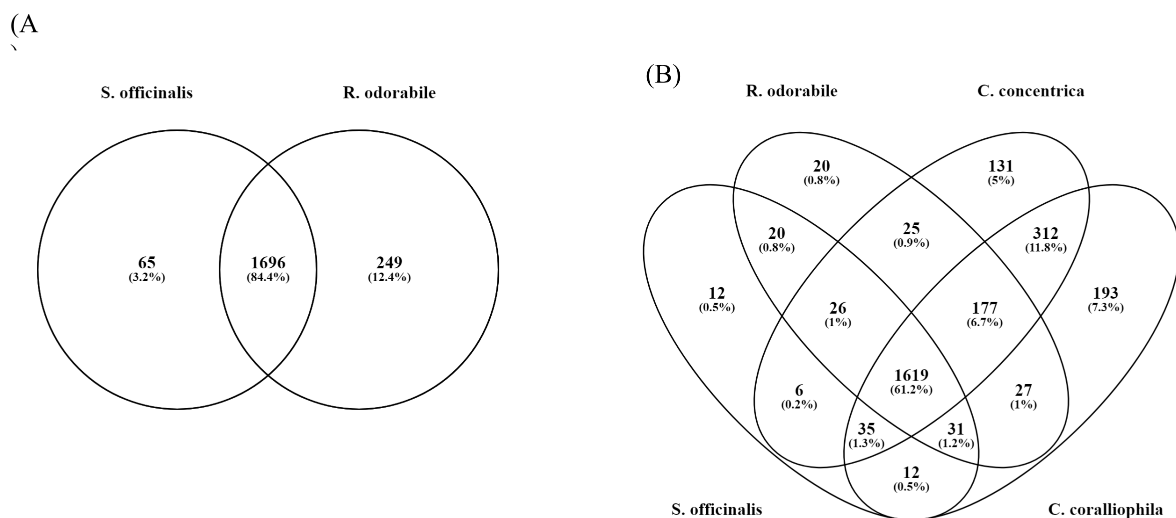


Appendix I- Figure FS 2. Principal Coordinate Analysis (PCoA) of COG functional profiles obtained for microbial metagenomes assembled in this study and those obtained for other sponge hosts. Community ordinations were based on pairwise Bray-Curtis dissimilarities calculated from normalized (Hellinger-transformed, non-rarefied) data, considering oscillations of relative COG abundances among samples. Ordination using data normalization after rarefying the metagenome libraries (standardization of all samples to the least sequenced sample) revealed the same trends as ordination using normalization on non-rarefied libraries libraries. Analyses were performed on COG community profiles extracted from the corresponding metagenomes using MG-RAST. *Rhopaloeides odorabile* (Rho), *Cymbastela concentrica* (Cyn) and *Cymbastela coralliophila* (Cyr) microbial metagenomes (Fan et al., 2012) were used in a comparative analysis against the COG-annotations retrieved in this study from S. officinalis, sediment and seawater metagenome assemblies. The first and second coordinates explain 34.3% and 17.7% of the total dataset variability, respectively. Significance values resulting from permutational analysis of variance (PERMANOVA) applied to the corresponding dissimilarity matrix are as follows. Overall differences among groups (all sponges vs. seawater vs. sediments): p = 0.0001. Pairwise significances: sponges were found to be different from seawater (p = 0.0057) and sediment (p = 0.0153) metagenomes, while no significant difference was found between sediment and seawater functional profiles (p = 0.309).

COGs specific to and shared by S. officinalis and other sponge hosts

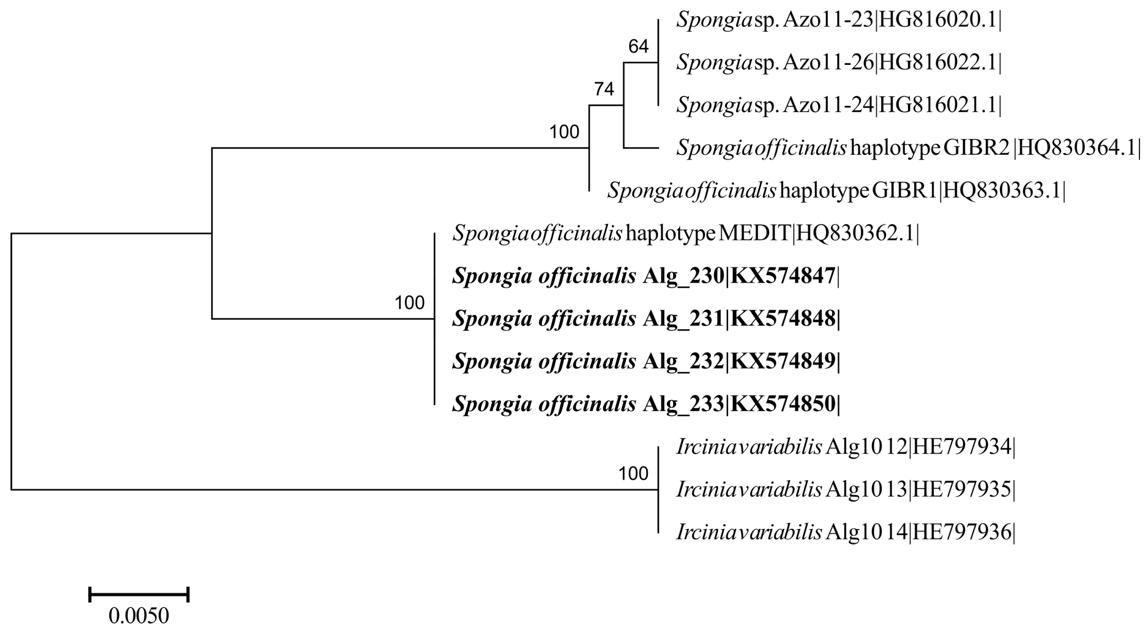
Independently of the quantitative assessments highlighted above, our comparative scheme enabled us fetch those COG entries shared by and specific to each of the sponge metagenome

libraries analysed. In line with PCoA results (**Figure FS-2**), we found that *S. officinalis* had more COGs in common with *R. odorabile* than with *C. concentrica* and *coralliophila* (**Figure FS-3**). Likewise, sponges altogether possessed more COGs in common with sediments than with seawater (data not shown). The functional core of the four sponge species was high (1691 COGs), representing 61.2% of all COGs identified in these metagenomes and revealing a considerable extent of functional convergence not only across a wide host phylogeny spectrum (as observed by Fan et al., 2012), but also geographical distances. Interestingly, SIMPER analysis of COG profiles listed for all four sponges together against sediment and seawater metagenomes revealed several sponge-enriched functions ranking as the most differentiating among biotopes (**Table S5**). Remarkably frequent among such top COG entries were type I and II restriction-modification systems identified here and by Fan et al. (2012) as sponge microbiome genetic signatures. Also, several of the observations made for IPR functional profiles obtained from unassembled reads could be revisited in this analysis, such as the higher abundance of ankyrin, tetratricopeptide, leucine-rich and WD-40 repeats in the sponge metagenomes, followed by sediments, as well as the distribution of polyketide, plasmid stabilization systems, ABC transporters and cytochrome P450 predicted functions, for instance, which followed the same trends observed for IPR annotations (**Table S5**).

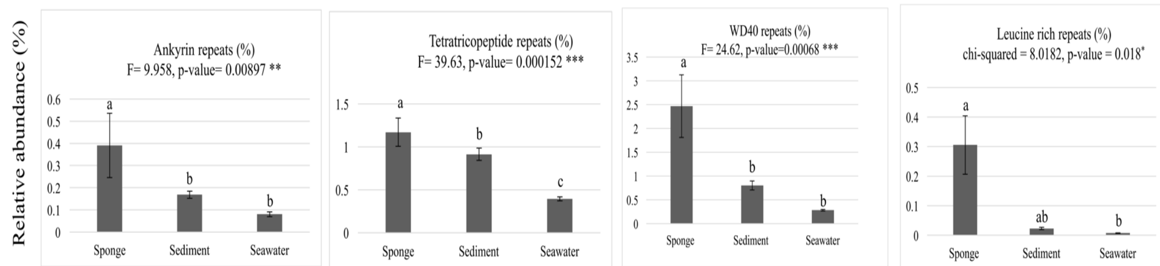


Appendix I- Figure FS 3.COGs shared by and specific to *Spongia officinalis* and Rhopaloeides odorabile (**A**) and *S. officinalis, R. odorabile*, *Cymbastela concentrica* and *Cymbastela coralliophila* (**B**). Results derive from COG annotations of assembled metagenomes retrieved in this study (*S. officinalis*) and by Fan et al. (2012) (*R. odorabile, C. concentrica* and *C. coralliophora*).

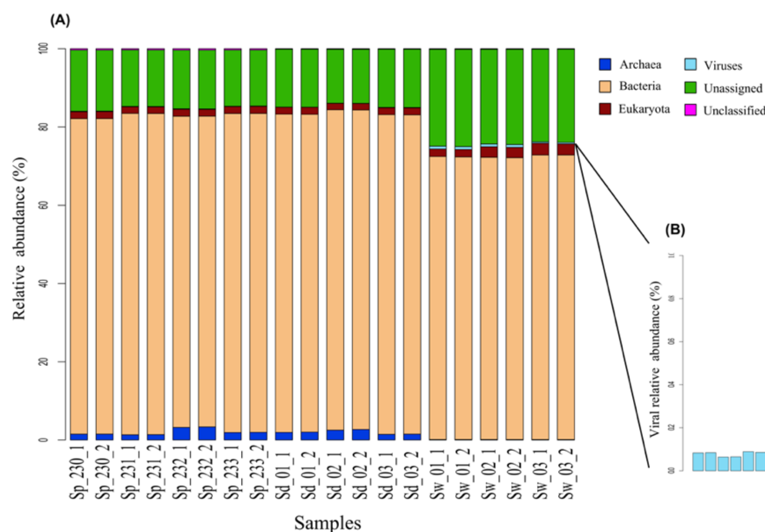
Supplementary Figures and Tables



Appendix I-Figure S1. Cytochrome oxidase I (COI) gene-based phylogenetic inference of sponge specimens examined in this study and their closest relatives**.** All sequences were aligned within the software package MEGA7. The evolutionary history was inferred using the Maximum Likelihood method based on the Kimura 2-parameter model. The tree for the heuristic search was obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. A discrete Gamma distribution was used to model evolutionary rate differences among sites. Percent bootstrap values greater than 70% are shown. The tree is rooted to the *Rhopaloeides* genus and sponges from the present study are shown in bold.



Appendix I -Figure S2. Abundance distributions of ELPs (TRPs, ANKs, LRRs and WD40) across biotopes. Values on the y-axis represent mean cumulative IPR relative abundances (%) in each biotope ± standard deviations. Ankyrin repeats - 3 IPR entries used in plot construction; Tetratricopeptide repeats - 10 IPR entries; WD40 repeats - 4 entries; leucine-rich repeats - 5 IPR entries. Results of the general test for differences among biotopes (One-Way ANOVA) are shown at the top of each chart, below the label of each analyzed function. Bars labeled with different letters represent statistically distinct biotopes in terms of IPR relative abundances according to pair-wise tests of significance.



Appendix I-Figure S3. Relative abundance and distribution of microbial domains (**A**) and viruses (**B**) across the biotopes based on best-hit classifications using MG-RAST

Appendix I-Table S1. Number of sequence reads per quality control steps using the EBI metagenomics (EMG) pipeline (v. 2.0).

x	a	b
a		
b		

Appendix I-Table S2. 16S rRNA gene-based distribution of microbial phyla (A) and OTUs (B) across biotopes

x	a	b
a		
b		

Appendix I-Table S3. Most differentiating microbial phyla (A) and OTUs (B) among biotopes.

x	a	b
a		
b		

Appendix I-Table S4. Most differentiating IPR entries among biotopes.

x	a	b
a		
b		

Appendix I-Table S5. Most differentiating COG entries among biotopes, with "sponges" representing functional profiles of S. officinalis, *R. odorabile*, *C. concentrica* and *C. coralliophila*.

x	a	b
a		
b		

***Above files are available on digital format due to their large sizes.**

Appendix II

Chapter 3 supplementary materials

Supplementary Figures

(A)

231_04	**Closest type strain (RDP, ≥1200bp); accession number**	**Similarity %**
1	*Phaeobacter inhibens* strain T5 (T); (NR_042761)	98.06
2	*Phaeobacter gallaeciensis* strain BS107 (T); (NR_027609)	97.98
3	*Leisingera methylohalidivorans* strain MB2 (T); (NR_025637)	97.69
4	*Leisingera aquimarina* strain R-26159 (T); (NR_042670)	97.57
5	*Ruegeria scottomollicae* LMG 24367 (T); (AM905330)	97.28
6	*Leisingera caerulea* LMG 24369 (T); (AM943630)	97.25
	Closest hit on NCBI BLASTN; accession number	
1	*Rhodobacteraceae* bacterium ACEMC 26-3; (FM163007)	99.88
2	Uncultured bacterium clone OS3BR21; JN233117	99.42
3	*Phaeobacter sp.* P97; (KX163077)	98.68
4	*Phaeobacter inhibens* strain DSM17395; (CP002976)	98.63
5	*Phaeobacter gallaeciensis* strain 2.10; (CP002972)	98.63
6	*Phaeobacter sp.* P104; (KX163079)	98.61
7	*Leisingera methylohalidivorans* strain MB2; (NR_121711)	98.02
8	*Phaeobacter* sp. strain 8-1; (AJ536670)	97.61

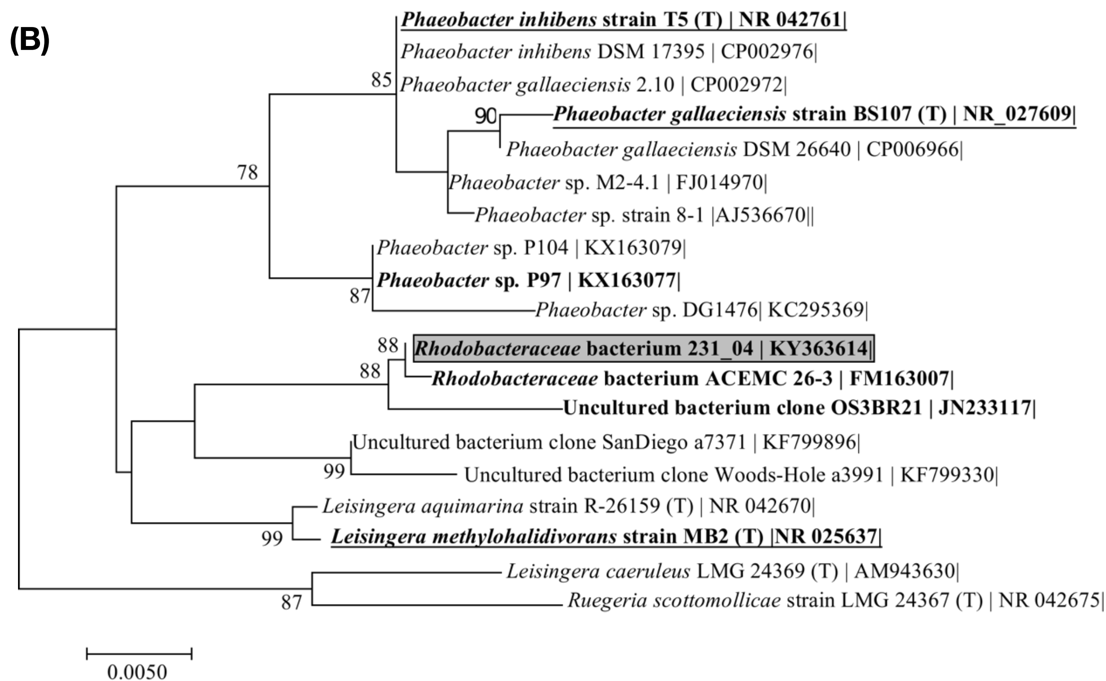


Figure S1.1. (A) The six closest type strains (T) (RDP Sequence match) and the eight closest NCBI BlastN hits to *Rhodobacteraceae* bacterium 231-04 are shown with their sequence similarity values. **(B)** 16S rRNA gene phylogeny of *Rhodobacteraceae* bacterium 231-04 (highlighted in grey) and close relatives based on the Maximum Likelihood method using the Kimura 2-parameter model. The top-three closest type strains (T) are highlighted in bold and underlined. The top-three closest NCBI BlastN hits are marked in bold. The tree with the highest log likelihood (-1688.3662) is shown. One hundred replicates were run to bootstrap the tree. The percentage of trees in which the associated taxa clustered together is shown next to the branches (60% cut-off). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.0558)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 51.9478% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 19 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 817 positions in the final dataset.

(A)

231_30	**Closest type strain (RDP, ≥1200bp); accession number**	**Similarity %**

1	*Thalassobius aestuarii* JC2049 (T); (AY442178)	97.8
2	*Shimia marina CL-TA03 (T);* (AY962292)	97.69
3	*Thalassobius mediterraneus* CECT 5383 (T); (AJ878874)	97.15
4	*Shimia haliotis* WM35 (T); (KC196071)	97.14
5	*Leisingera aquimarina* LMG 24366T (T); (AM900415)	96.49
6	*Leisingera methylohalidivorans* MB2 (T); (AY005463)	96.06

	Closest hit on NCBI BLASTN; accession number	
1	*Rhodobacteraceae* bacterium Ph113; (HE818273)	98.57
2	*Shimia sagamensis* strain JAMH 011; (NR_137204)	98.35
3	*Rhodobacteraceae* bacterium 2tb2; (FJ952817)	98.13
4	Uncultured bacterium clone B12_10.3_2; (FJ716880)	98.13
5	Uncultured bacterium clone BF5_1108; (KC307193)	98.02
6	*Alphaproteobacterium* C32; (AB302373)	98.02

(B)

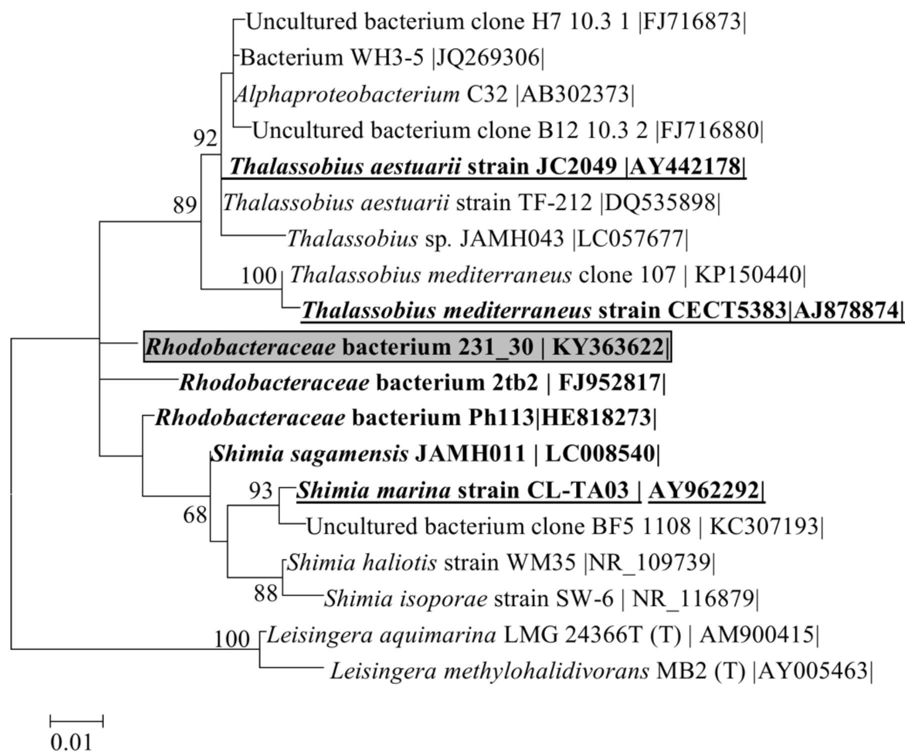
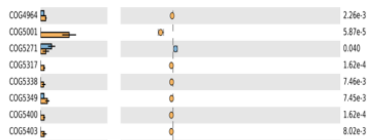


Figure S1.2. (A) The six closest type strains (T) (RDP Sequence match) and the six closest NCBI BlastN hits to *Rhodobacteracae* bacterium 231-30 are shown with their sequence similarity values. **(B)** 16S rRNA gene phylogeny of *Rhodobacteraceae* bacterium 231-30 (highlighted in grey) and close relatives based on the Maximum Likelihood method using the Tamura-Nei model. The top three closest type strains (T) are highlighted in bold and underlined. The top three closest NCBI BlastN hits are marked in bold. The tree with the highest log likelihood (-2050.0808) is shown. One hundred replicates were run to bootstrap the tree. The percentage of trees in which the associated taxa clustered together is shown next to the branches (60% cut-off). A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+*G*, parameter = 0.2182)). The rate variation model allowed for some sites to be evolutionarily invariable ([+*I*], 75.8138% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 19 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 906 positions in the final dataset.

Appendix II- Figure S1. The closest type strains (T) (RDP Sequence match) and the closest NCBI BlastN hits to the *Rhodobacteraceae* bacterium 231-04 and 231-30.



Appendix II- Figure S2. Bar plots of COG entries displaying significantly different relative abundances in *Roseobacter* clade (Group 1) versus non-*Roseobacter* clade (Group 2) genomes*.
*This figure was too large for paper formatting but can be found in digital format (CD).

Supplementary Tables

Appendix II- Table S1. Colony morphology and classification of bacterial isolates obtained on MG50 medium from Spongia officinalis.

x	a	b
a		
b		

Appendix II- Table S2. 16S rRNA gene-based taxonomic affiliation of the 48 Spongia officinalis isolates retrieved on MG50 medium.

Phylum	Class	Order	Family	Genus	N° isolates	N° OTUs (100%)
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Ruegeria*	28	10
Proteobacteria	*Gammaproteobacteria*	*Vibrionales*	*Vibrionaceae*	*Vibrio*	4	2
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Pseudovibrio*	3	1
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	Unclassified *Rhodobacteraceae* Alg231-04	1	1
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	Unclassified *Rhodobacteraceae* Alg231-30	1	1
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Tateyamaria*	2	1
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Loktanella*	2	1
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Dermacoccaceae*	*Dermacoccus*	1	1
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Erythrobacteraceae*	*Erythrobacter*	1	1
Proteobacteria	*Alphaproteobacteria*	*Rhodobacterales*	*Rhodobacteraceae*	*Labrenzia*	1	1
Proteobacteria	*Gammaproteobacteria*	*Alteromonadales*	*Shewanellaceae*	*Shewanella woodyi*	1	1
Proteobacteria	*Alphaproteobacteria*	*Sphingomonadales*	*Sphingomonadaceae*	*Sphingorhabdus*	1	1
Proteobacteria	*Alphaproteobacteria*	*Rhizobiales*	*Rhodobiaceae*	*Anderseniella*	1	1
Actinobacteria	*Actinobacteria*	*Actinomycetales*	*Nocardioidaceae*	*Nocardioides*	1	1
2 phyla	**3 classes**	**6 orders**	**8 families**	**12 classified + 2 unclassified**	**48**	**24**

Appendix II- Table S3. List of protein encoding-genes shared between the 10 Alphaproteobacteria genomes.

x	a	b
a		
b		

Appendix II- Table S4. Number of protein-encoding sequences (CDSs) present in each Clusters of Orthologous Groups of Proteins (COG) class for each of the 10 *Alphaproteobacteria* genomes.

x	a	b
a		
b		

Appendix II- Table S5. Number of protein-encoding sequences (CDSs) classified into Clusters of Orthologous Groups of Proteins (COG) for each of the 10 Alphaproteobacteria genomes analysed in this study.

x	a	b
a		
b		

Appendix II-Table S6. Clusters of Orthologous Groups (COG) entries shared between the ten alphaproteobacterial genomes analysed in this study.

x	a	b
a		
b		

Appendix II- Table S7. COG entries displaying significantly different relative abundances in *Roseobacter* clade (Group 1) versus non-*Roseobacter* clade (Group 2) genomes. (**B)** COG entries absent in all *Roseobacter* genomes and present in all non-*Roseobacter* genomes. (**C**) COG entries present in all *Roseobacter* genomes and absent in all non-*Roseobacter* genomes**.**

x	a	b
a		
b		

Appendix II- Table S8. Antibiotics and Secondary Metabolite Analysis Shell (antiSMASH) results. Shown are all secondary metabolite biosynthetic gene clusters found on each *Alphaproteobacteria* genome by the antiSMASH search tool.

x	a	b
a		
b		

***Above tables and figure are only available on digital format due to their large sizes.**

Appendix III

Chapter 4 complementary analysis and supplementary materials

File S1 Data analysis

Supplementary Methods

Sponge and seawater sample processing

Briefly, 2.5 g of the inner body of each sponge specimen were cut and macerated with sterile mortar and pestle containing 22.5 mL of calcium/magnesium free artificial seawater (CMFASW) (Garson et al., 1998). The resulting homogenates were subjected to a differential centrifugation step (Hardoim et al., 2014) to discard host-derived cells and retrieve microbial cell pellets (MCPs) from the host samples. MCPs were stored at -80 ºC until total community DNA (TC-DNA) extraction. Two liters of seawater from each replicate were filtered through a 0.22 µm nitrocellulose membrane (Merck Millipore, Billerica, MA, USA) which was then stored at -80 ºC until TC-DNA extraction. TC-DNA was extracted from MCPs and nitrocellulose membrane filters with the UltraClean® Soil DNA isolation kit (Mo Bio, Carlsbad, CA, USA) according to the manufacturer's protocol. TC-DNA concentrations were determined with the Qubit (Life Technologies Qubit 2.0®) dsDNA HS Assay Kit. TC-DNA sequencing was carried out on an Illumina Hiseq 2500 apparatus at Mr. DNA (Shallowater, TX, USA). Briefly, sequencing DNA libraries were prepared using the Nextera DNA Sample preparation kit (Illumina) following the manufacturer's instructions, and sequenced paired end for 200 cycles with depth calibrated at *c.* 15 million 101bp reads per sample (Chapter 2) (Karimi et al., 2017b)

Geographic distribution of uncultivated, sponge-associated Rhodospirillales

All OTUs assigned to *Alphaproteobacteria*, *Rhodospirillales* and *Rhodospirillaceae* in the SM dataset by any of the three taxonomies employed in the analysis, namely Greengenes, RDP and SILVA, were retrieved from the list of representative OTU sequences delivered by Thomas et al. (2016). Thereafter, a customized R script designed to merge this list of classifications with the general SM OTU table was used to create a specific sample vs. OTUs table containing only (and all) alphaproteobacterial OTUs. The ascertainment of the distribution and relative abundance of the three target taxa across (1) all sponge samples, (2) sponge taxonomic orders and (3) geographical locations was accomplished by merging the customized OTU table with accompanying metadata released by Thomas *et al.* (2016). Furthermore, 16S rRNA gene sequences assembled in *Alphaproteobacteria* bins from marine sponges (Slaby et al., 2017) were subjected to phylogenetic inference with the

Rhodospirillales OTUs present in the SM dataset (Thomas et al., 2016).The closest SM OTU to our 16S rRNA gene queries was thereafter subjected to an *in silico* analysis of worldwide distribution and abundance across sponge hosts to delineate the likely degree of host fidelity and dispersal patterns of the sponge-associated bins examined closely in this study.

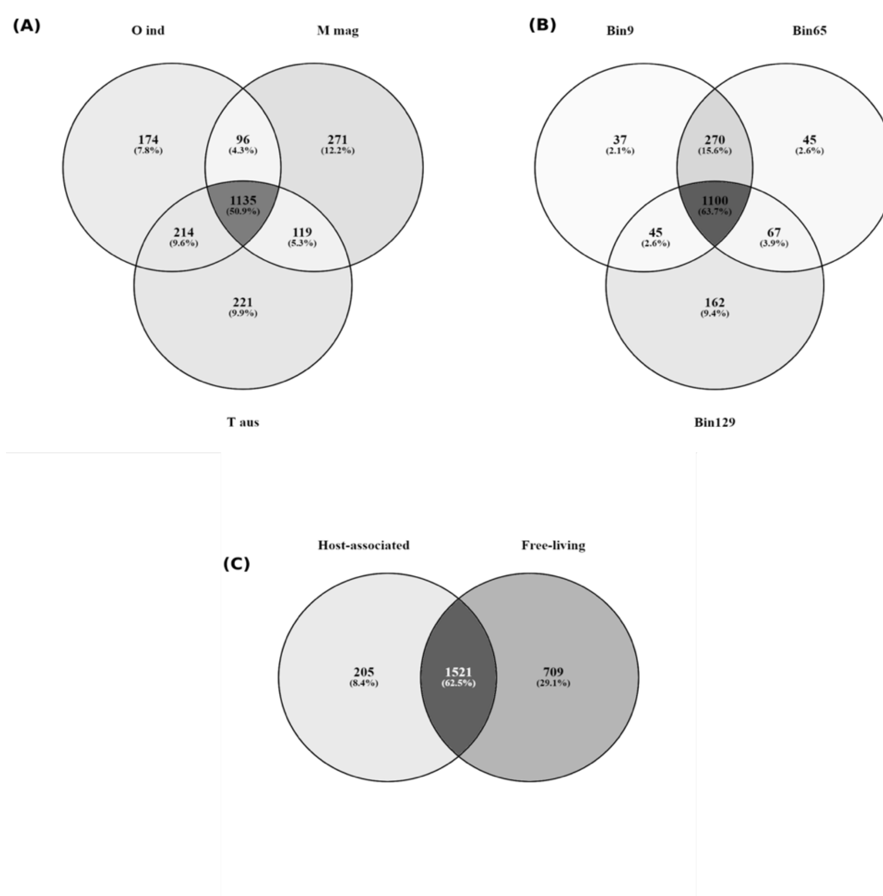
Supplementary Results

Distribution of Rhodospirillaceae species across marine sponges

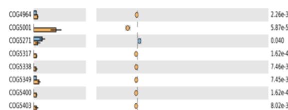
Depending on the database used in taxonomic assignments, the total number of 16S rRNA gene OTUs (from c. 3600 to 4150) and reads (from 1,181,054 to 1,434,120) assigned to the class *Alphaproteobacteria* represented around 8-9.5% of the total microbiome make-up of marine sponges within the SM dataset, in terms of both richness (OTU numbers) and abundance (read numbers) (**Table S5**). Using the RDP and Greengenes databases, the family *Rhodospirillaceae* approached 11% and 19% of the total *Alphaproteobacteria* richness (OTUs) and abundance (reads) uncovered in the SM dataset, respectively. These numbers dropped to around 3.5% (richness) and 4.5% (abundance) when taxonomy was assigned with the SILVA database (**Table S5**), found to label as "unclassified" several OTUs assigned as *Rhodospirillaceae* by the former databases. In any case, many hundreds of *Rhodospirillaceae* and closely related OTUs were found to occur in association with marine sponges worldwide. 16S rRNA gene sequences of phylotypes Aa65 and Aa129 were assigned as *Rhodospirillaceae* by the three databases (data not shown). Specifically, the 16S rRNA gene of phylotype Aa65 from *Aplysina aerophoba* - the closest relative to phylotype So9 assembled in this study - shared greatest relatedness with OTU 003276 from the SM dataset. This OTU was found in 204 of the 804 sponge samples examined by (Thomas et al., 2016), amounting to 4,801 reads. Only one read belonging to OTU 003276 was found in a sediment sample (out of 36), whereas no reads from this OTU could be found in the seawater samples (n = 133) from the SM dataset. We noticed that OTU 003276 alone accounted for *c.* 23% and 8.7% of all *Rhodopirillaceae* reads detected in sponge species of the order Dictyoceratida (to which *S. officinalis* belongs) and Verongiida, and that its representativeness in the Mediterranean Sea is likely correlated with high sampling effort of such host species in this zone. Indeed, OTU 003276 was a particularly dominant alphaproteobacterium of *Aplysina aerophoba* (Verongiida, Croatia) and *Ircinia variabilis* (Dictyoceratida, Spain) microbial communities, being also commonly found in the

Mediterranean dictyoceratiids *Sarcotragus fasciculatus*, *Ircinia oros* and *Ircinia strobilina* (**Table S5B**).

Supplementary Figures and Tables



Appendix III- Figure S1. Specificity and sharedness of COG entries among free-living and symbiotic *Rhodospirillaceae* genomes**.** COGs specific and common to individual free-living (A) and sponge-associated (B) genomes are shown along with COGs specific and common to free-living and symbiotic genomes pools (C), each encompassing the three individual genomes shown in (A) and (B), respectively. Full names of symbiotic and free-living strains are as in the footnote to Table 4-2.



Appendix III- Figure S2. Bar plots of sponge-enriched or sponge-depleted COGs (n = 287) displaying statistically significant differences in abundance among sponge-associated (orange) and free-living (blue) *Rhodospirillaceae* genome pools are shown. Error bars indicate within-group standard deviations. All presented categories passed a corrected p-value of 0.05 in White´s non-parametric t-test. This figure was too large for paper formatting but can be found in digital format (CD).

Appendix III- Table S1. *Alphaproteobacteria* genome bins produced in this study and selected for further analysis after assembly improvements attempted with Spades and IDBA-UD.

Samples	Raw sum of ORF*	#Dup. **	Sum	Completeness estimation (%)***	Selected for further analysis	# contigs	# contigs (≧1000 bp)	Largest contig	Total length	N50
Spongia										
Bin9 (So9)[1]	108	5	103	92.79%		470	434	64,933	4,034,471	14,271
Bin9_re-assembly SPAdes	102	0	102	91.89%	✓	281	273	94,913	4,052,441	24,896
Bin9_re-assembly IDBA-UD	100	1	99	89.19%		333	333	68,041	4,222,698	19,238
Water										
Bin34 (Wat34)[2]	104	3	101	90.99%	✓	274	274	195,517	3,617,015	30,799
Bin34_re-assembly SPAdes	103	4	99	89.19%		573	439	113,781	3,634,468	14,420
Bin34_re-assembly IDBA-UD	103	4	99	89.19%		761	491	98,106	3,677,683	13,031
Bin73 (Wat73)[3]	105	3	102	91.89%	✓	157	157	60,648	17,925,66	20,176
Bin73_re-assembly SPAdes	105	3	102	91.89%		140	137	78,859	1,784,018	22,544
Bin73_re-assembly IDBA-UD	104	3	101	90.99%		243	223	36,324	1,777,061	14,660

* Number of essential single-copy genes out of 111, ** number of duplicate single-copy genes, *** Genome completeness percentage (sum of essential coding genes minus the number of duplicates divided by 111).[1]*Rhodospirillaceae* bacterium Spongia_Bin9, [2]*Rhodobacteraceae* bacterium Water_Bin34, [3]*Phyllobacteriaceae* bacterium Water_Bin73.

Appendix III- Table S2. Average amino acid identity (AAI, A) and average nucleotide identity (ANI, B) measures among sponge-associated and free-living *Rhodospirillaceae.*

(A) AAI

Genomes	So9	Aa65	Aa129	Taus	Mmag
Spongia So9	100	93.71	61.2	57.18	59.52
Aplysina Aa65	93.71	100	60.69	57.24	59.35
Aplysina Aa129	61.2	60.69	100	58.2	60.47
Magnetospirillum mag	57.16	57.38	58.2	100	60.53
Thalassospira aus	59.52	59.35	60.47	60.53	100

(B) ANI

Genomes	Taus	Mmag	Aa129	So9	Aa65
Thalassospira aus	100	78.77	82.32	76.67	77.15
Magnetospirillum mag	79.18	100	77.12	77.57	77.73
Aplysina Aa129	75.13	77.97	100	81.11	81.56
Spongia So9	74.59	77.79	81.36	100	91.3
Aplysina Aa65	76.73	78.39	80.85	91.36	100

***Full names of genomes are as in the footnote to Table 4-2.**

Appendix III- Table S3. COG annotation of sponge-associated (highlighted in green) and free-living *Rhodospirillaceae* genomes.

x	a	b
a		
b		

Appendix III- Table S4. COGs enriched (A) and depleted (B) in sponge-associated Rhodospirillaceae.

x	a	b
a		
b		

Appendix III- Table S5. Worldwide abundance, distribution and richness of Alphaproteobacteria, Rhodospirillales and Rhodospirillaceae OTUs (A) and of SM OTU 003276 (B) in marine sponges.

x	a	b
a		
b		

*** Above tables can be found in digital format (on CD) due to their large sizes.**

Bibliography

Ainsworth, T.D., Thurber, R.V., and Gates, R.D. (2010). The future of coral reefs: a microbial perspective. *Trends Ecol Evol* 25(4)**,** 233-240. doi: 10.1016/j.tree.2009.11.001.

Albertsen, M., Hugenholtz, P., Skarshewski, A., Nielsen, K.L., Tyson, G.W., and Nielsen, P.H. (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotech* 31(6)**,** 533-538. doi: 10.1038/nbt.2579.

Alex, A., and Antunes, A. (2015). Whole genome sequencing of the symbiont *Pseudovibrio* sp. from the intertidal Marine Sponge *Polymastia penicillus* revealed a gene repertoire for host-switching permissive lifestyle. *Genome Biology and Evolution* 7(11)**,** 3022-3032. doi: 10.1093/gbe/evv199.

Alexander, B.E., Liebrand, K., Osinga, R., van der Geest, H.G., Admiraal, W., Cleutjens, J.P.M., et al. (2014). Cell turnover and detritus production in marine sponges from tropical and temperate benthic ecosystems. *PLoS. ONE* 9(10)**,** e109486. doi: 10.1371/journal.pone.0109486.

Allardet-Servent, A., Michaux-Charachon, S., Jumas-Bilak, E., Karayan, L., and Ramuz, M. (1993). Presence of one linear and one circular chromosome in the Agrobacterium tumefaciens C58 genome. *J. Bacteriol.* 175(24)**,** 7869-7874.

Allen, J.A., and Garrett, M.R. (1971). "Taurine in marine invertebrates," in *Adv. Mar. Biol.,* eds. F.S. Russell & M. Yonge. Academic Press), 205-253.

Allocati, N., Federici, L., Masulli, M., and Di Ilio, C. (2009). Glutathione transferases in bacteria. *FEBS J* 276(1)**,** 58-75. doi: 10.1111/j.1742-4658.2008.06743.x.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat Meth* 11(11)**,** 1144-1146. doi: 10.1038/nmeth.3103.

Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., et al. (2013). CONCOCT: clustering contigs on coverage and composition. *arXiv preprint arXiv:1312.4038*.

Amin, S.A., Hmelo, L.R., van Tol, H.M., Durham, B.P., Carlson, L.T., Heal, K.R., et al. (2015). Interaction and signalling between a cosmopolitan phytoplankton and associated bacteria. *Nature* 522(7554)**,** 98-101. doi: 10.1038/nature14488.

Anjum, K., Abbas, S.Q., Shah, S.A.A., Akhter, N., Batool, S., and Hassan, S.S.u. (2016). Marine sponges as a drug treasure. *Biomolecules & Therapeutics* 24(4)**,** 347-362. doi: 10.4062/biomolther.2016.067.

Argenio, D.A., and Miller, S.I. (2004). Cyclic di-GMP as a bacterial second messenger. *Microbiology* 150(8)**,** 2497-2502. doi: 10.1099/mic.0.27099-0.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., et al. (2011). Enterotypes of the human gut microbiome. *Nature* 473**,** 174. doi: 10.1038/nature09944.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., et al. (2008). The RAST server: rapid annotations using subsystems technology. *BMC Genomics* 9(1)**,** 1. doi: 10.1186/1471-2164-9-75.

Baldani, J.I., Videira, S.S., dos Santos Teixeira, K.R., Reis, V.M., de Oliveira, A.L.M., Schwab, S., et al. (2014). "The family *Rhodospirillaceae*," in *The Prokaryotes: Alphaproteobacteria and Betaproteobacteria,* eds. E. Rosenberg, E.F. DeLong, S. Lory, E. Stackebrandt & F. Thompson. (Berlin, Heidelberg: Springer Berlin Heidelberg), 533-618.

Barrangou, R., and Marraffini, L.A. (2014). CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell* 54(2)**,** 234-244. doi: 10.1016/j.molcel.2014.03.011.

Bauvais, C., Zirah, S., Piette, L., Chaspoul, F., Domart-Coulon, I., Chapon, V., et al. (2015). Sponging up metals: bacteria associated with the marine sponge Spongia officinalis. *Mar. Environ. Res.* 104**,** 20-30. doi: 10.1016/j.marenvres.2014.12.005.

Bayer, K., Moitinho-Silva, L., Brümmer, F., Cannistraci, C.V., Ravasi, T., and Hentschel, U. (2014). GeoChip-based insights into the microbial functional gene repertoire of marine sponges (high microbial abundance, low microbial abundance) and seawater. *FEMS Microbiol. Ecol.* 90(3)**,** 832-843. doi: 10.1111/1574-6941.12441.

Bayer, K., Scheuermayer, M., Fieseler, L., and Hentschel, U. (2013). Genomic mining for novel FADH2-dependent halogenases in marine sponge-associated microbial consortia. *Mar. Biotechnol* 15(1)**,** 63-72. doi: 10.1007/s10126-012-9455-2.

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(11)**,** 2942-2955. doi: 10.1111/j.1462-2920.2008.01582.x.

Becerro, M. (2012). *Advances in sponge science: phylogeny, systematics, ecology.* UK: Academic Press.

Bentley, S.D., and Parkhill, J. (2004). Comparative genomic structure of prokaryotes. *Annu Rev Genet* 38**,** 771-791.

Berendsen, R.L., Pieterse, C.M.J., and Bakker, P.A.H.M. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci* 17(8)**,** 478-486. doi: 10.1016/j.tplants.2012.04.001.

Beyhan, S., Tischler, A.D., Camilli, A., and Yildiz, F.H. (2006). Transcriptome and phenotypic responses of Vibrio cholerae to increased cyclic di-GMP level. *J. Bacteriol.* 188(10)**,** 3600-3613. doi: 10.1128/JB.188.10.3600-3613.2006.

Binnewies, T.T., Motro, Y., Hallin, P.F., Lund, O., Dunn, D., La, T., et al. (2006). Ten years of bacterial genome sequencing: comparative-genomics-based discoveries. *Funct. Integr. Genomics* 6(3)**,** 165-185. doi: 10.1007/s10142-006-0027-2.

Blom, J., Albaum, S.P., Doppmeier, D., Pühler, A., Vorhölter, F.-J., Zakrzewski, M., et al. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinform.* 10(1)**,** 154. doi: 10.1186/1471-2105-10-154.

Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., et al. (2016). EDGAR 2.0: an enhanced software platform for comparative gene content analyses. *Nucleic Acids Res* 44**,** W22–W28. doi: 10.1093/nar/gkw255.

Bondarev, V., Richter, M., Romano, S., Piel, J., Schwedt, A., and Schulz-Vogt, H.N. (2013). The genus Pseudovibrio contains metabolically versatile bacteria adapted for symbiosis. *Environ. Microbiol.* 15(7)**,** 2095-2113. doi: 10.1111/1462-2920.12123.

Bright, M., and Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nat Rev Micro* 8(3)**,** 218-230. doi: 10.1038/nrmicro2262.

Brinkmann, C., Kearns, P., Evans-Illidge, E., and Kurtböke, D. (2017). Diversity and Bioactivity of Marine Bacteria Associated with the Sponges Candidaspongia flabellata and Rhopaloeides odorabile from the Great Barrier Reef in Australia. *Diversity* 9(3)**,** 39.

Britstein, M., Devescovi, G., Handley, K.M., Malik, A., Haber, M., Saurav, K., et al. (2016). A New N-Acyl Homoserine Lactone Synthase in an Uncultured Symbiont of the Red Sea Sponge Theonella swinhoei. *Appl. Environ. Microbiol.* 82(4)**,** 1274-1285. doi: 10.1128/aem.03111-15.

Bröms, J.E., Edqvist, P.J., Forsberg, Å., and Francis, M.S. (2006). Tetratricopeptide repeats are essential for PcrH chaperone function in Pseudomonas aeruginosa type III secretion. *FEMS Microbiol Lett* 256(1)**,** 57-66. doi: 10.1111/j.1574-6968.2005.00099.x.

Bruto, M., James, A., Petton, B., Labreuche, Y., Chenivesse, S., Alunno-Bruscia, M., et al. (2017). *Vibrio crassostreae*, a benign oyster colonizer turned into a pathogen after plasmid acquisition. *ISME J* 11(4)**,** 1043-1052. doi: 10.1038/ismej.2016.162.

Buchan, A., González, J.M., and Moran, M.A. (2005). Overview of the marine Roseobacter lineage. *Appl. Environ. Microbiol.* 71(10)**,** 5665-5677. doi: 10.1128/AEM.71.10.5665–5677.2005.

Burgsdorf, I., Slaby, B.M., Handley, K.M., Haber, M., Blom, J., Marshall, C.W., et al. (2015). Lifestyle evolution in cyanobacterial symbionts of sponges. *Mbio* 6(3)**,** e00391-00315. doi: 10.1128/mBio.00391-15.

Burrows, L.L. (2012). Prime time for minor subunits of the type II secretion and type IV pilus systems. *Mol. Microbiol.* 86(4)**,** 765-769. doi: 10.1111/mmi.12034.

Caputo, A., Merhej, V., Georgiades, K., Fournier, P.-E., Croce, O., Robert, C., et al. (2015). Pan-genomic analysis to redefine species and subspecies based on quantum discontinuous variation: the Klebsiella paradigm. *Biology Direct* 10(1)**,** 55. doi: 10.1186/s13062-015-0085-2.

Christianson, D.W. (2017). Structural and chemical biology of terpenoid cyclases. *Chem Rev* 117(17)**,** 11570-11648. doi: 10.1021/acs.chemrev.7b00287.

Ciaglia, E., Malfitano, A.M., Laezza, C., Fontana, A., Nuzzo, G., Cutignano, A., et al. (2017). Immuno-Modulatory and Anti-Inflammatory Effects of Dihydrogracilin A, a Terpene Derived from the Marine Sponge Dendrilla membranosa. *International Journal of Molecular Sciences* 18(8)**,** 1643. doi: 10.3390/ijms18081643.

Clarke, K.R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18(1)**,** 117-143. doi: 10.1111/j.1442-9993.1993.tb00438.x.

Cleary, D.F.R., Becking, L.E., Voogd, N.J.d., Pires, A.C.C., Polónia, A.R.M., Egas, C., et al. (2013). Habitat- and host-related variation in sponge bacterial symbiont communities

in Indonesian waters. *FEMS Microbiol. Ecol.* 85(3), 465-482. doi: 10.1111/1574-6941.12135.

Clooney, A.G., Fouhy, F., Sleator, R.D., O' Driscoll, A., Stanton, C., Cotter, P.D., et al. (2016). Comparing Apples and Oranges?: Next Generation Sequencing and Its Impact on Microbiome Analysis. *PLoS. ONE* 11(2), e0148028. doi: 10.1371/journal.pone.0148028.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., et al. (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res* 37(suppl 1), D141-D145.

Connon, S.A., and Giovannoni, S.J. (2002). High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68(8), 3878-3885. doi: 10.1128/AEM.68.8.3878–3885.2002.

Costa, R., Keller-Costa, T., Gomes, N.C., da Rocha, U.N., van Overbeek, L., and van Elsas, J.D. (2013). Evidence for selective bacterial community structuring in the freshwater sponge *Ephydatia fluviatilis*. *Microb. Ecol.* 65(1), 232-244. doi: 10.1007/s00248-012-0102-2.

Cox, C.J., Foster, P.G., Hirt, R.P., Harris, S.R., and Embley, T.M. (2008). The archaebacterial origin of eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 105(51), 20356-20361. doi: 10.1073/pnas.0810647105.

Crowley, S.P., O'Gara, F., O'Sullivan, O., Cotter, P.D., and Dobson, A.D.W. (2014). Marine *Pseudovibrio* sp. as a novel source of antimicrobials. *Mar Drugs* 12(12), 5916-5929. doi: 10.3390/md12125916.

Cruz-López, R., and Maske, H. (2016). The Vitamin B(1) and B(12) Required by the Marine Dinoflagellate Lingulodinium polyedrum Can be Provided by its Associated Bacterial Community in Culture. *Front Microbiol* 7, 560. doi: 10.3389/fmicb.2016.00560.

Cui, J., and Davidson, A.L. (2011). ABC solute importers in bacteria. *Essays Biochem.* 50, 85-99. doi: 10.1042/bse0500085.

Dailianis, T., Tsigenopoulos, C.S., Dounas, C., and Voultsiadou, E. (2011). Genetic diversity of the imperilled bath sponge *Spongia officinalis* Linnaeus, 1759 across the Mediterranean Sea: patterns of population differentiation and implications for taxonomy and conservation. *Mol. Ecol.* 20(18), 3757-3772. doi: 10.1111/j.1365-294X.2011.05222.x.

de Cook, S.C., and Bergquist, P.R. (2002). "Family Spongiidae Gray, 1867," in *Systema Porifera: A Guide to the Classification of Sponges,* eds. J.N.A. Hooper, R.W.M. Van Soest & P. Willenz. (Boston, MA: Springer US), 1051-1060.

de Goeij, J.M., van den Berg, H., van Oostveen, M.M., Epping, E.H.G., and van Duyl, F.C. (2008). Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. *Mar. Ecol. Prog. Ser.* 357, 139-151. doi: 10.3354/meps07403.

de Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., et al. (2013). Surviving in a marine desert: The sponge loop retains resources within coral reefs. *Science* 342(6154), 108-110. doi: 10.1126/science.1241981.

de Zamaroczy, M., Delorme, F., and Elmerich, C. (1990). Characterization of three different nitrogen-regulated promoter regions for the expression of glnB and glnA in

Azospirillum brasilense. *Molecular & general genetics : MGG* 224(3), 421-430. doi: 10.1007/bf00262437.

Diaz, M.C., and Rützler, K. (2001). Sponges: an essential component of Caribbean coral reefs. *Bull. Mar. Sci.* 69(2), 535-546.

Díez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., and Thomas, T. (2016). Expression of eukaryotic-like protein in the microbiome of sponges. *Mol. Ecol.* 26(5), 1432-1451. doi: 10.1111/mec.14003.

Donachie, S.P., Foster, J.S., and Brown, M.V. (2007). Culture clash: challenging the dogma of microbial diversity. *ISME J* 1(2), 97-99.

Drider, D., Bendali, F., Naghmouchi, K., and Chikindas, M.L. (2016). Bacteriocins: Not Only Antibacterial Agents. *Probiotics and Antimicrobial Proteins* 8(4), 177-182. doi: 10.1007/s12602-016-9223-0.

Dudhagara, P., Bhavsar, S., Bhagat, C., Ghelani, A., Bhatt, S., and Patel, R. (2015). Web Resources for Metagenomics Studies. *Genomics Proteomics Bioinformatics* 13(5), 296-303. doi: 10.1016/j.gpb.2015.10.003.

Ebada, S.S., Lin, W., and Proksch, P. (2010). Bioactive sesterterpenes and triterpenes from marine sponges: occurrence and pharmacological significance. *Mar Drugs* 8(2), 313-346.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19), 2460-2461. doi: 10.1093/bioinformatics/btq461.

Eid, J., Fehr, A., Gray, J., Luong, K., Lyle, J., Otto, G., et al. (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323(5910), 133-138. doi: 10.1126/science.1162986.

Eilers, H., Pernthaler, J., Glöckner, F.O., and Amann, R. (2000). Culturability and In Situ Abundance of Pelagic Bacteria from the North Sea. *Appl. Environ. Microbiol.* 66(7), 3044-3051.

Emura, C., Higuchi, R., and Miyamoto, T. (2006). Irciniasulfonic acid B, a novel taurine conjugated fatty acid derivative from a Japanese marine sponge, *Ircinia* sp. *Tetrahcdron* 62(24), 5682-5685. doi: 10.1016/j.tet.2006.03.087.

Enticknap, J.J., Kelly, M., Peraud, O., and Hill, R.T. (2006). Characterization of a culturable alphaproteobacterial symbiont common to many marine sponges and evidence for vertical transmission via sponge larvae. *Appl. Environ. Microbiol.* 72(5), 3724-3732. doi: 10.1128/AEM.72.5.3724-3732.2006.

Ereskovsky, A.V., and Tokina, D.B. (2007). Asexual reproduction in homoscleromorph sponges (Porifera; Homoscleromorpha). *Mar Biol* 151(2), 425-434. doi: 10.1007/s00227-006-0439-5.

Erwin, P.M., López-Legentil, S., González-Pech, R., and Turon, X. (2012a). A specific mix of generalists: bacterial symbionts in Mediterranean *Ircinia* spp. *FEMS Microbiol. Ecol.* 79(3), 619-637. doi: 10.1111/j.1574-6941.2011.01243.x.

Erwin, P.M., López-Legentil, S., and Turon, X. (2012b). Ultrastructure, molecular phylogenetics, and chlorophyll a content of novel cyanobacterial symbionts in temperate sponges. *Microb. Ecol.* 64(3), 771-783. doi: 10.1007/s00248-012-0047-5.

Erwin, P.M., Pineda, M.C., Webster, N., Turon, X., and López-Legentil, S. (2014). Down under the tunic: bacterial biodiversity hotspots and widespread ammonia-oxidizing archaea in coral reef ascidians. *ISME J.* 8(3), 575-588. doi: 10.1038/ismej.2013.188.

Esteves, A.I.S., Amer, N., Nguyen, M., and Thomas, T. (2016). Sample Processing Impacts the Viability and Cultivability of the Sponge Microbiome. *Front Microbiol* 7, 499. doi: 10.3389/fmicb.2016.00499.

Esteves, A.I.S., Hardoim, C.C.P., Xavier, J.R., Goncalves, J.M., and Costa, R. (2013). Molecular richness and biotechnological potential of bacteria cultured from Irciniidae sponges in the north-east Atlantic. *FEMS Microbiol. Ecol.* 85(3), 519-536. doi: 10.1111/1574-6941.12140.

Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., et al. (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. U.S.A.* 109(27), E1878-E1887. doi: 10.1073/pnas.1203287109.

Fieseler, L., Horn, M., Wagner, M., and Hentschel, U. (2004). Discovery of the Novel Candidate Phylum “Poribacteria” in Marine Sponges. *Appl. Environ. Microbiol.* 70(6), 3724-3732. doi: 10.1128/AEM.70.6.3724-3732.2004.

Fieseler, L., Quaiser, A., Schleper, C., and Hentschel, U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum *Poribacteria* by environmental genomics. *Environ. Microbiol.* 8(4), 612-624. doi: 10.1111/j.1462-2920.2005.00937.x.

Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., et al. (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res* 45(D1), D190-D199. doi: 10.1093/nar/gkw1107.

Finn, R.D., Clements, J., and Eddy, S.R. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* 39(suppl_2), W29-W37. doi: 10.1093/nar/gkr367.

Gao, Z.-M., Wang, Y., Tian, R.-M., Wong, Y.H., Batang, Z.B., Al-Suwailem, A.M., et al. (2014). Symbiotic adaptation drives genome streamlining of the cyanobacterial sponge symbiont “Candidatus Synechococcus spongiarum”. *MBio* 5(2), e00079-00014. doi: 10.1128/mBio.00079-14.

Garrabou, J., Coma, R., Bensoussan, N., Bally, M., Chevaldonné, P., Cigliano, M., et al. (2009). Mass mortality in Northwestern Mediterranean rocky benthic communities: effects of the 2003 heat wave. *Glob. Change. Biol* 15(5), 1090-1103. doi: 10.1111/j.1365-2486.2008.01823.x.

Garrity, G.M., Bell, J.A., and Lilburn, T. (2005). "Class I. Alphaproteobacteria class. nov," in *Bergey's Manual® of Systematic Bacteriology*. Springer), 1-574.

Garson, M., Flowers, A.E., Webb, R.I., Charan, R.D., and McCaffrey, E.J. (1998). A sponge/dinoflagellate association in the haplosclerid sponge *Haliclona* sp.: cellular origin of cytotoxic alkaloids by Percoll density gradient fractionation. *Cell Tissue Res* 293(2), 365-373. doi: 10.1007/s004410051128.

Gauthier, M.-E.A., Watson, J.R., and Degnan, S.M. (2016). Draft genomes shed Light on the dual bacterial symbiosis that dominates the microbiome of the coral reef sponge

Amphimedon queenslandica. *Frontiers in Marine Science* 3, 196 doi: 10.3389/fmars.2016.00196.

Gazave, E., Lapébie, P., Ereskovsky, A.V., Vacelet, J., Renard, E., Cárdenas, P., et al. (2012). No longer Demospongiae: Homoscleromorpha formal nomination as a fourth class of Porifera. *Hydrobiologia* 687(1), 3-10. doi: 10.1007/s10750-011-0842-x.

Giebel, H.-A., Kalhoefer, D., Lemke, A., Thole, S., Gahl-Janssen, R., Simon, M., et al. (2011). Distribution of Roseobacter RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. *ISME J.* 5(1), 8-19. doi: 10.1038/ismej.2010.87.

Gilbert, J.A., and Dupont, C.L. (2011). Microbial metagenomics: beyond the genome. *Annu. Rev. Mar. Sci.* 3, 347-371. doi: 10.1146/annurev-marine-120709-142811.

Giovannoni, S., and Rappé, M. (2000). "Evolution, diversity, and molecular ecology of marine prokaryotes," ed. D.L. Kirchman. (New York: Wiley-Liss Inc), 47-84.

Girvan, H.M., and Munro, A.W. (2016). Applications of microbial cytochrome P450 enzymes in biotechnology and synthetic biology. *Curr Opin Chem Biol* 31(Supplement C), 136-145. doi: 10.1016/j.cbpa.2016.02.018.

Gloeckner, V., Wehrl, M., Moitinho-Silva, L., Gernert, C., Schupp, P., Pawlik, J.R., et al. (2014). The HMA-LMA Dichotomy Revisited: an Electron Microscopical Survey of 56 Sponge Species. *The Biological Bulletin* 227(1), 78-88. doi: 10.1086/BBLv227n1p78.

Gomez-Escribano, J.P., and Bibb, M.J. (2011). Engineering Streptomyces coelicolor for heterologous expression of secondary metabolite gene clusters. *Microbial Biotechnology* 4(2), 207-215. doi: 10.1111/j.1751-7915.2010.00219.x.

Gonzalez, A.G., Estrada, D.M., Martin, J.D., Martin, V.S., Perez, C., and Perez, R. (1984). New antimicrobial diterpenes from the sponge *Spongia officinalis*. *Tetrahcdron* 40(20), 4109-4113. doi: 10.1016/0040-4020(84)85092-9.

Gordaliza, M. (2010). Cytotoxic terpene quinones from marine sponges. *Mar Drugs* 8(12), 2849-2870. doi: 10.3390/md8122849.

Graça, A.P., Bondoso, J., Gaspar, H., Xavier, J.R., Monteiro, M.C., de la Cruz, M., et al. (2013). Antimicrobial Activity of Heterotrophic Bacterial Communities from the Marine Sponge Erylus discophorus (Astrophorida, Geodiidae). *PLoS. ONE* 8(11), e78992. doi: 10.1371/journal.pone.0078992.

Grissa, I., Vergnaud, G., and Pourcel, C. (2007). CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35(suppl_2), W52-W57.

Gruber-Vodicka, H.R., Dirks, U., Leisch, N., Baranyi, C., Stoecker, K., Bulgheresi, S., et al. (2011). Paracatenula, an ancient symbiosis between thiotrophic *Alphaproteobacteria* and catenulid flatworms. *Proc. Natl. Acad. Sci. U.S.A.* 108(29), 12078-12083. doi: 10.1073/pnas.1105347108.

Grueneberg, J., Engelen, A.H., Costa, R., and Wichard, T. (2016). Macroalgal morphogenesis induced by waterborne compounds and bacteria in coastal seawater. *PLoS. ONE* 11(1), e0146307. doi: 10.1371/journal.pone.0146307.

Guengerich, F.P., Gillam, E.M., and Shimada, T. (1996). New applications of bacterial systems to problems in toxicology. *Crit Rev Toxicol* 26(5)**,** 551-583.

Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8)**,** 1072-1075. doi: 10.1093/bioinformatics/btt086.

Gutleben, J., Chaib De Mares, M., van Elsas, J.D., Smidt, H., Overmann, J., and Sipkema, D. (2017). The multi-omics promise in context: from sequence to microbial isolate. *Crit. Rev. Microbiol.***,** 1-18. doi: 10.1080/1040841X.2017.1332003.

Hacquard, S., Garrido-Oter, R., Gonzalez, A., Spaepen, S., Ackermann, G., Lebeis, S., et al. (2015). Microbiota and Host Nutrition across Plant and Animal Kingdoms. *Cell Host Microbe* 17(5)**,** 603-616. doi: 10.1016/j.chom.2015.04.009.

Hadfield, M.G. (2011). Biofilms and marine invertebrate larvae: what bacteria produce that larvae use to choose settlement sites. *Annu. Rev. Mar. Sci.* 3**,** 453-470. doi: 10.1146/annurev-marine-120709-142753.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y.-i., Sugahara, J., et al. (2006). Genomic analysis of the uncultivated marine crenarchaeote Cenarchaeum symbiosum. *Proc. Natl. Acad. Sci. U.S.A.* 103(48)**,** 18296-18301. doi: 10.1073/pnas.0608549103.

Hammer, Ø., Harper, D., and Ryan, P. (2001). PAST-palaeontological statistics, ver. 1.89. *Palaeontol. Electron* 4(9).

Handelsman, J. (2001). "Metagenomics and Microbial Communities," in *eLS*. John Wiley & Sons, Ltd).

Handelsman, J., Rondon, M.R., Brady, S.F., Clardy, J., and Goodman, R.M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* 5(10)**,** R245-R249. doi: 10.1016/S1074-5521(98)90108-9.

Hardoim, C.C.P., Cardinale, M., Cúcio, A.C.B., Esteves, A.I.S., Berg, G., Xavier, J.R., et al. (2014). Effects of sample handling and cultivation bias on the specificity of bacterial communities in keratose marine sponges. *Front Microbiol* 5(611)**,** 611. doi: 10.3389/fmicb.2014.00611.

Hardoim, C.C.P., and Costa, R. (2014a). Microbial communities and bioactive compounds in marine sponges of the family irciniidae-a review. *Mar Drugs* 12(10)**,** 5089-5122. doi: 10.3390/md12105089.

Hardoim, C.C.P., and Costa, R. (2014b). Temporal dynamics of prokaryotic communities in the marine sponge Sarcotragus spinosulus. *Mol. Ecol.* 23(12)**,** 3097-3112. doi: 10.1111/mec.12789.

Hardoim, C.C.P., Costa, R., Araujo, F.V., Hajdu, E., Peixoto, R., Lins, U., et al. (2009). Diversity of bacteria in the marine sponge *Aplysina fulva* in Brazilian coastal waters. *Appl. Environ. Microbiol.* 75(10)**,** 3331-3343.

Hardoim, C.C.P., Esteves, A.I., Pires, F.R., Goncalves, J.M., Cox, C.J., Xavier, J.R., et al. (2012). Phylogenetically and spatially close marine sponges harbour divergent bacterial communities. *PLoS. ONE* 7(12)**,** e53029. doi: 10.1371/journal.pone.0053029.

Hentschel, U., Fieseler, L., Wehrl, M., Gernert, C., Steinert, M., Hacker, J., et al. (2003). "Microbial Diversity of Marine Sponges," in *Sponges (Porifera),* ed. W.E.G. Müller. (Berlin, Heidelberg: Springer Berlin Heidelberg), 59-88.

Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J., et al. (2002). Molecular evidence for a uniform microbial community in sponges from different oceans. *Appl. Environ. Microbiol.* 68(9)**,** 4431-4440. doi: 10.1128/aem.68.9.4431-4440.2002.

Hentschel, U., Piel, J., Degnan, S.M., and Taylor, M.W. (2012). Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* 10(9)**,** 641-654. doi: 10.1038/nrmicro2839.

Hentschel, U., Schmid, M., Wagner, M., Fieseler, L., Gernert, C., and Hacker, J. (2001). Isolation and phylogenetic analysis of bacteria with antimicrobial activities from the Mediterranean sponges Aplysina aerophoba and Aplysina cavernicola. *FEMS Microbiol. Ecol.* 35(3)**,** 305-312. doi: 10.1111/j.1574-6941.2001.tb00816.x.

Hentschel, U., Usher, K.M., and Taylor, M.W. (2006). Marine sponges as microbial fermenters. *FEMS Microbiol. Ecol.* 55(2)**,** 167-177. doi: 10.1111/j.1574-6941.2005.00046.x.

Hestetun, J.T. (2016). *Carnivorous sponges of the Atlantic and Arctic Oceans. Phylogeny, taxonomy, distribution and microbial associations of the Cladorhizidae (Demospongiae, Poecilosclerida).* Doctoral, University of Bergen.

Higgins, C.F. (2001). ABC transporters: physiology, structure and mechanism–an overview. *Res Microbiol* 152(3)**,** 205-210.

Hill, R. (2004). Microbes from marine sponges: a treasure trove of biodiversity for natural products discovery. *AT Bull (ed.), Microbial diversity and bioprospecting. ASM Press, Washington, DC***,** 177-190.

Hoffmann, F., Larsen, O., Thiel, V., Rapp, H.T., Pape, T., Michaelis, W., et al. (2005). An anaerobic world in sponges. *Geomicrobiol. J* 22(1-2)**,** 1-10. doi: 10.1080/01490450590922505.

Hoffmann, F., Røy, H., Bayer, K., Hentschel, U., Pfannkuchen, M., Brümmer, F., et al. (2008). Oxygen dynamics and transport in the Mediterranean sponge Aplysina aerophoba. *Mar Biol* 153(6)**,** 1257-1264. doi: 10.1007/s00227-008-0905-3.

Hooper, J.N., and Van Soest, R.W. (2002). "Systema Porifera. A guide to the classification of sponges," in *Systema Porifera*. Springer), 1-7.

Horgan, R.P., and Kenny, L.C. (2011). 'Omic' technologies: genomics, transcriptomics, proteomics and metabolomics. *The Obstetrician & Gynaecologist* 13(3)**,** 189-195. doi: 10.1576/toag.13.3.189.27672.

Horn, H. (2017). *Analysis and interpretation of (meta-) genomic data from host-associated microorganisms.* Ph.D, uni-wuerzburg.

Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Förster, F., et al. (2016). An enrichment of CRISPR and other defense-related features in marine sponge-associated microbial metagenomes. *Front Microbiol* 7**,** 1751. doi: 10.3389/fmicb.2016.01751.

Horn, M., and Wagner, M. (2004). Bacterial Endosymbionts of Free-living Amoebae1. *J. Eukaryot. Microbiol.* 51(5), 509-514. doi: 10.1111/j.1550-7408.2004.tb00278.x.

Huang, R., Chen, Y., Zhou, X., Yang, X., and Liu, Y. (2015). A New N-Acyl taurine from the south China sea marine sponge *Callyspongia* sp. *Chem Nat Compd* 51(3), 540-541. doi: 10.1007/s10600-015-1335-3.

Huang, R., Peng, Y., Zhou, X., Yang, X., and Liu, Y. (2013). A new taurine derivative from South China Sea marine sponge *Axinella* sp. *Nat. Prod. Res.* 27(17), 1537-1541. doi: 10.1080/14786419.2012.733389.

Huang, Z., Brooke, B., and Li, J. (2011). Performance of predictive models in marine benthic environments based on predictions of sponge distribution on the Australian continental shelf. *Ecol. Inform.* 6(3), 205-216. doi: 10.1016/j.ecoinf.2011.01.001.

Hunter, S., Corbett, M., Denise, H., Fraser, M., Gonzalez-Beltran, A., Hunter, C., et al. (2014). EBI metagenomics—a new resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res* 42(D1), D600-D606. doi: 10.1093/nar/gkt961.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016). MEGAN Community Edition - interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput Biol* 12. doi: 10.1371/journal.pcbi.1004957.

Hyatt, D., Chen, G., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11. doi: 10.1186/1471-2105-11-119.

Iverson, V., Morris, R.M., Frazar, C.D., Berthiaume, C.T., Morales, R.L., and Armbrust, E.V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science* 335(6068), 587-590. doi: 10.1126/science.1212665.

Izard, J., and Rivera, M. (2014). *Metagenomics for Microbiology.* Academic Press.

Jackson, S.A., Kennedy, J., Morrissey, J.P., apos, Gara, F., and Dobson, A.D.W. (2015). Maribacter spongiicola sp. nov. and Maribacter vaceletii sp. nov., isolated from marine sponges, and emended description of the genus Maribacter. *Int. J. Syst. Evol. Microbiol* 65(7), 2097-2103. doi: doi:10.1099/ijs.0.000224.

Janssen, P.H., Yates, P.S., Grinton, B.E., Taylor, P.M., and Sait, M. (2002). Improved Culturability of Soil Bacteria and Isolation in Pure Culture of Novel Members of the Divisions Acidobacteria, Actinobacteria, Proteobacteria, and Verrucomicrobia. *Appl. Environ. Microbiol.* 68(5), 2391-2396. doi: 10.1128/AEM.68.5.2391-2396.2002.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T.L. (2008). NCBI BLAST: a better web interface. *Nucleic Acids Res* 36(suppl 2), W5-W9.

Kadlec, K., and Schwarz, S. (2009). Novel ABC transporter gene, vga (C), located on a multiresistance plasmid from a porcine methicillin-resistant *Staphylococcus aureus* ST398 strain. *Antimicrob. Agents. Chemother.* 53(8), 3589-3591. doi: 10.1128/AAC.00570-09.

Kaeberlein, T., Lewis, K., and Epstein, S.S. (2002). Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. *Science* 296(5570), 1127-1129. doi: 10.1126/science.1070633.

Kamke, J., Rinke, C., Schwientek, P., Mavromatis, K., Ivanova, N., Sczyrba, A., et al. (2014). The candidate phylum poribacteria by single-cell genomics: New insights into phylogeny, cell-compartmentation, eukaryote-Like repeat proteins, and other genomic features. *PLoS. ONE* 9(1)**,** e87353. doi: 10.1371/journal.pone.0087353.

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(12)**,** 2287-2300. doi: 10.1038/ismej.2013.111.

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(4)**,** 498-508. doi: 10.1038/ismej.2009.143.

Karimi, E., Gonçalves, J.M.S., Reis, M., and Costa, R. (2017a). Draft Genome Sequence of Microbacterium sp. Strain Alg239_V18, an Actinobacterium Retrieved from the Marine Sponge *Spongia* sp. *Genome Announcements* 5(3). doi: 10.1128/genomeA.01457-16.

Karimi, E., Ramos, M., Gonçalves, J.M.S., Xavier, J.R., Reis, M.P., and Costa, R. (2017b). Comparative Metagenomics Reveals the Distinctive Adaptive Features of the *Spongia officinalis* Endosymbiotic Consortium. *Front Microbiol* 8(2499). doi: 10.3389/fmicb.2017.02499.

Karimi, E., Slaby, B.M., Soares, A.R., Blom, J., Hentschel, U., and Costa, R. (2018-in press). Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS. (Fed. Eur. Microbiol. Soc). Microbiol. Ecol.*

Kayal, E. (2012). *The evolution of the mitochondrial genomes of calcareous sponges and cnidarians.* . PhD Thesis, Iowa State University, .

Keeling, P.J., and Campo, J.d. (2017). Marine Protists Are Not Just Big Bacteria. *Curr. Biol.* 27(11)**,** R541-R549. doi: 10.1016/j.cub.2017.03.075.

Keller-Costa, T., Eriksson, D., Gonçalves, J.M.S., Gomes, N.C.M., Lago-Lestón, A., and Costa, R. (2017). The gorgonian coral Eunicella labiata hosts a distinct prokaryotic consortium amenable to cultivation. *FEMS Microbiol. Ecol.* 93(12)**,** fix143-fix143. doi: 10.1093/femsec/fix143.

Kennedy, J., Baker, P., Piper, C., Cotter, P.D., Walsh, M., Mooij, M.J., et al. (2009). Isolation and Analysis of Bacteria with Antimicrobial Activities from the Marine Sponge Haliclona simulans Collected from Irish Waters. *Mar. Biotechnol* 11(3)**,** 384-396. doi: 10.1007/s10126-008-9154-1.

Kennedy, J., Marchesi, J.R., and Dobson, A.D. (2008). Marine metagenomics: strategies for the discovery of novel enzymes with biotechnological applications from marine environments. *Microb Cell Fact* 7**,** 27. doi: 10.1186/1475-2859-7-27.

Keren, R., Mayzel, B., Lavy, A., Polishchuk, I., Levy, D., Fakra, S.C., et al. (2017). Sponge-associated bacteria mineralize arsenic and barium on intracellular vesicles. *Nat. Commun.* 8**,** 14393. doi: 10.1038/ncomms14393.

Keyzers, R.A., Northcote, P.T., and Davies-Coleman, M.T. (2006). Spongian diterpenoids from marine sponges. *Nat. Prod. Rep.* 23(2)**,** 321-334. doi: 10.1039/B503531G.

Kieser, H.M., Kieser, T., and Hopwood, D.A. (1992). A combined genetic and physical map of the Streptomyces coelicolor A3 (2) chromosome. *J. Bacteriol.* 174(17)**,** 5496-5507.

Kuhlmann, A.U., Hoffmann, T., Bursy, J., Jebbar, M., and Bremer, E. (2011). Ectoine and Hydroxyectoine as Protectants against Osmotic and Cold Stress: Uptake through the SigB-Controlled Betaine-Choline- Carnitine Transporter-Type Carrier EctT from Virgibacillus pantothenticus. *J. Bacteriol.* 193(18)**,** 4699-4708. doi: 10.1128/JB.05270-11.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.***,** msw054.

Kumar, V., Maitra, S.S., and Shukla, R.N. (2015). Environmental metagenomics: the data assembly and data analysis perspectives. *J. Inst. Eng. (India): Ser. A* 96(1)**,** 71-83. doi: 10.1007/s40030-014-0102-y.

Lackner, G., Peters, E.E., Helfrich, E.J.N., and Piel, J. (2017). Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl. Acad. Sci. U.S.A.* 114(3)**,** E347-E356. doi: 10.1073/pnas.1616234114.

Land, M., Hauser, L., Jun, S.-R., Nookaew, I., Leuze, M.R., Ahn, T.-H., et al. (2015). Insights from 20 years of bacterial genome sequencing. *Funct. Integr. Genomics* 15(2)**,** 141-161. doi: 10.1007/s10142-015-0433-4.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4)**,** 357-359. doi: 10.1038/nmeth.1923.

Laport, M.S., Santos, O.C.S., and Muricy, G. (2009). Marine Sponges: Potential Sources of New Antimicrobial Drugs. *Curr Pharm Biotechnol* 10(1)**,** 86-105. doi: 10.2174/138920109787048625.

Lê, S., Josse, J., and Husson, F. (2008). FactoMineR: an R package for multivariate analysis. *Journal of statistical software* 25(1)**,** 1-18.

Lee, O.O., Chui, P.Y., Wong, Y.H., Pawlik, J.R., and 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. doi: 10.1128/AEM.00023-09.

Lee, O.O., Lau, S.C., Tsoi, M.M., Li, X., Plakhotnikova, I., Dobretsov, S., et al. (2006). Shewanella irciniae sp. nov., a novel member of the family Shewanellaceae, isolated from the marine sponge Ircinia dendroides in the Bay of Villefranche, Mediterranean Sea. *Int. J. Syst. Evol. Microbiol* 56(Pt 12)**,** 2871-2877. doi: 10.1099/ijs.0.64562-0.

Lejon, D.P., Kennedy, J., and Dobson, A.D. (2011). Identification of novel bioactive compounds from the metagenome of the marine sponge Haliclona simulans. *Handbook of Molecular Microbial Ecology II: Metagenomics in Different Habitats***,** 553-562.

Li, C.-W., Chen, J.-Y., and Hua, T.-E. (1998). Precambrian sponges with cellular structures. *Science* 279(5352)**,** 879-882.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* 25(16)**,** 2078-2079. doi: 10.1093/bioinformatics/btp352.

Li, J., Gu, B.-B., Sun, F., Xu, J.-R., Jiao, W.-H., Yu, H.-B., et al. (2017). Sesquiterpene quinones/hydroquinones from the marine sponge *Spongia pertusa Esper*. *J. Nat. Prod.* 80(5), 1436-1445. doi: 10.1021/acs.jnatprod.6b01105.

Li, X.-Q., and Du, D. (2014). Variation, evolution, and correlation analysis of C+G content and genome or chromosome size in different kingdoms and phyla. *PLoS. ONE* 9(2), e88339. doi: 10.1371/journal.pone.0088339.

Liu, F., Li, J., Feng, G., and Li, Z. (2016). New Genomic Insights into "Entotheonella" Symbionts in Theonella swinhoei: Mixotrophy, Anaerobic Adaptation, Resilience, and Interaction. *Front Microbiol* 7(1333). doi: 10.3389/fmicb.2016.01333.

Liu, M., Fan, L., Zhong, L., Kjelleberg, S., and Thomas, T. (2012). Metaproteogenomic analysis of a community of sponge symbionts. *ISME J* 6(8), 1515-1525. doi: 10.1038/ismej.2012.1.

Lopanik, N., Lindquist, N., and Targett, N. (2004). Potent cytotoxins produced by a microbial symbiont protect host larvae from predation. *Oecologia* 139(1), 131-139. doi: 10.1007/s00442-004-1487-5.

Luo, C., Tsementzi, D., Kyrpides, N.C., and Konstantinidis, K.T. (2012a). Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* 6(4), 898-901. doi: 10.1038/ismej.2011.147.

Luo, H., Löytynoja, A., and Moran, M.A. (2012b). Genome content of uncultivated marine Roseobacters in the surface ocean. *Environ. Microbiol.* 14(1), 41-51. doi: 10.1111/j.1462-2920.2011.02528.x.

Lurie-Weinberger, M.N., Gomez-Valero, L., Merault, N., Glöckner, G., Buchrieser, C., and Gophna, U. (2010). The origins of eukaryotic-like proteins in Legionella pneumophila. *Int J Med Microbiol* 300(7), 470-481. doi: 10.1016/j.ijmm.2010.04.016.

MacLean, M.J., Ness, L.S., Ferguson, G.P., and Booth, I.R. (1998). The role of glyoxalase I in the detoxification of methylglyoxal and in the activation of the KefB K+ efflux system in Escherichia coli. *Mol. Microbiol.* 27(3), 563-571. doi: 10.1046/j.1365-2958.1998.00701.x.

Maldonado, M. (2009). Embryonic development of verongid demosponges supports the independent acquisition of spongin skeletons as an alternative to the siliceous skeleton of sponges. *Biol. J. Linn. Soc.* 97(2), 427-447. doi: 10.1111/j.1095-8312.2009.01202.x.

Maldonado, M., Ribes, M., and van Duyl, F.C. (2012). "Nutrient fluxes through sponges: biology, budgets, and ecological implications," in *Adv. Mar. Biol.,* eds. M.A. Becerro, M.J. Uriz, M. Maldonado & X. Turon. Academic Press), 113-182.

Manconi, R., and Pronzato, R. (2008). Global diversity of sponges (Porifera: Spongillina) in freshwater. *Hydrobiologia* 595(1), 27-33.

Manzo, E., Ciavatta, M.L., Villani, G., Varcamonti, M., Sayem, S.M., van Soest, R., et al. (2011). Bioactive terpenes from *Spongia officinalis*. *J. Nat. Prod.* 74(5), 1241-1247. doi: 10.1021/np200226u.

Martens, T., Heidorn, T., Pukall, R., Simon, M., Tindall, B.J., and Brinkhoff, T. (2006). Reclassification of Roseobacter gallaeciensis Ruiz-Ponte et al. 1998 as Phaeobacter gallaeciensis gen. nov., comb. nov., description of Phaeobacter inhibens sp. nov., reclassification of Ruegeria algicola (Lafay et al. 1995) Uchino et al. 1999 as

Marinovum algicola gen. nov., comb. nov., and emended descriptions of the genera Roseobacter, Ruegeria and Leisingera. *Int. J. Syst. Evol. Microbiol* 56(6)**,** 1293-1304. doi: doi:10.1099/ijs.0.63724-0.

Martínez-García, E., Nikel, P.I., Chavarría, M., and de Lorenzo, V. (2014). The metabolic cost of flagellar motion in *Pseudomonas putida* KT2440. *Environ. Microbiol.* 16(1)**,** 291-303. doi: 10.1111/1462-2920.12309.

Mayer, A.M.S., Glaser, K.B., Cuevas, C., Jacobs, R.S., Kem, W., Little, R.D., et al. (2010). The odyssey of marine pharmaceuticals: a current pipeline perspective. *Trends Pharmacol Sci* 31(6)**,** 255-265. doi: 10.1016/j.tips.2010.02.005.

McCutcheon, J.P., and Moran, N.A. (2012). Extreme genome reduction in symbiotic bacteria. *Nat Rev Micro* 10(1)**,** 13-26.

McMurdie, P.J., and Holmes, S. (2014). Waste not, want not: why rarefying microbiome data is inadmissible. *PLoS Comput Biol* 10(4)**,** e1003531. doi: 10.1371/journal.pcbi.1003531.

Melville, S., and Craig, L. (2013). Type IV Pili in Gram-Positive Bacteria. *Microbiology and Molecular Biology Reviews : MMBR* 77(3)**,** 323-341. doi: 10.1128/MMBR.00063-12.

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 Bioinform.* 9(1)**,** 386. doi: 10.1186/1471-2105-9-386.

Mineta, K., and Gojobori, T. (2016). Databases of the marine metagenomics. *Gene* 576(2, Part 1)**,** 724-728. doi: 10.1016/j.gene.2015.10.035.

Mitchell, A., Bucchini, F., Cochrane, G., Denise, H., Hoopen, P.t., Fraser, M., et al. (2016). EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res* 44(D1)**,** D595-D603. doi: 10.1093/nar/gkv1195.

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(6)**,** 1348-1363. doi: 10.1111/mec.12365.

Moitinho-Silva, L., Diez-Vives, C., Batani, G., Esteves, A.I.S., Jahn, M.T., and Thomas, T. (2017). Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J* 11(7)**,** 1651-1666. doi: 10.1038/ismej.2017.25.

Montalvo, N.F., Davis, J., Vicente, J., Pittiglio, R., Ravel, J., and Hill, R.T. (2014). Integration of culture-based and molecular analysis of a complex sponge-associated bacterial community. *PLoS. ONE* 9(3)**,** e90517. doi: 10.1371/journal.pone.0090517.

Moreno-Vivián, C., Cabello, P., Martínez-Luque, M., Blasco, R., and Castillo, F. (1999). Prokaryotic nitrate reduction: molecular properties and functional distinction among bacterial nitrate reductases. *J. Bacteriol.* 181(21)**,** 6573-6584.

Morris, R.M., Rappe, M.S., Connon, S.A., Vergin, K.L., Siebold, W.A., Carlson, C.A., et al. (2002). SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420(6917)**,** 806-810.

Morrow, C., and Cárdenas, P. (2015). Proposal for a revised classification of the Demospongiae (Porifera). *Front. Zool* 12(1)**,** 7. doi: 10.1186/s12983-015-0099-8.

Moya, A., Pereto, J., Gil, R., and Latorre, A. (2008). Learning how to live together: genomic insights into prokaryote-animal symbioses. *Nat. Rev. Genet.* 9(3)**,** 218-229. doi: 10.1038/nrg2319.

Mukhopadhyay, R., and Rosen, B.P. (2002). Arsenate reductases in prokaryotes and eukaryotes. *Environ. Health Perspect.* 110(Suppl 5)**,** 745-748.

Müller, W., Zahn, R., Kurelec, B., Lucu, C., Müller, I., and Uhlenbruck, G. (1981). Lectin, a possible basis for symbiosis between bacteria and sponges. *J. Bacteriol.* 145(1)**,** 548-558.

Murillo, F.J., Kenchington, E., Lawson, J.M., Li, G., and Piper, D.J.W. (2016). Ancient deep-sea sponge grounds on the Flemish Cap and Grand Bank, northwest Atlantic. *Mar Biol* 163.

Muscholl-Silberhorn, A., Thiel, V., and Imhoff, J.F. (2008). Abundance and bioactivity of cultured sponge-associated bacteria from the Mediterranean Sea. *Microb. Ecol.* 55(1)**,** 94-106.

Naim, M.A., Morillo, J.A., Sørensen, S.J., Waleed, A.A.-S., Smidt, H., and Sipkema, D. (2014). Host-specific microbial communities in three sympatric North Sea sponges. *FEMS Microbiol. Ecol.* 90(2)**,** 390-403. doi: 10.1111/1574-6941.12400.

Namiki, T., Hachiya, T., Tanaka, H., and Sakakibara, Y. (2012). MetaVelvet: an extension of Velvet assembler to de novo metagenome assembly from short sequence reads. *Nucleic Acids Res.* 40**,** e155-e155. doi: 10.1093/nar/gks678.

Naughton, L.M., Romano, S., O'Gara, F., and Dobson, A.D.W. (2017). Identification of Secondary Metabolite Gene Clusters in the Pseudovibrio Genus Reveals Encouraging Biosynthetic Potential toward the Production of Novel Bioactive Compounds. *Front Microbiol* 8(1494). doi: 10.3389/fmicb.2017.01494.

Navarre, W.W., Porwollik, S., Wang, Y., McClelland, M., Rosen, H., Libby, S.J., et al. (2006). Selective silencing of foreign DNA with low GC Content by the H-NS protein in *Salmonella*. *Science* 313(5784)**,** 236-238. doi: 10.1126/science.1128794.

Ng, A., and Xavier, R.J. (2011). Leucine-rich repeat (LRR) proteins: Integrators of pattern recognition and signaling in immunity. *Autophagy* 7(9)**,** 1082-1084. doi: 10.4161/auto.7.9.16464.

Nguyen, M.T.H.D., Liu, M., and Thomas, T. (2014). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol. Ecol.* 23(6)**,** 1635-1645. doi: 10.1111/mec.12384.

Nielsen, H.B., Almeida, M., Juncker, A.S., Rasmussen, S., Li, J., Sunagawa, S., et al. (2014). Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat Biotech* 32(8)**,** 822-828. doi: 10.1038/nbt.2939.

Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P. (2016). MetaSPAdes: a new versatile de novo metagenomics assembler. *arXiv preprint arXiv:1604.03071*.

Nyholm, S.V., and McFall-Ngai, M. (2004). The winnowing: establishing the squid-vibrio symbiosis. *Nat Rev Micro* 2(8)**,** 632-642.

O'Halloran, J., Barbosa, T., Morrissey, J., Kennedy, J., O'Gara, F., and Dobson, A. (2011). Diversity and antimicrobial activity of *Pseudovibrio* spp. from Irish marine sponges. *J. Appl. Microbiol.* 110(6)**,** 1495-1508.

Ochman, H., and Moran, N.A. (2001). Genes Lost and Genes Found: Evolution of Bacterial Pathogenesis and Symbiosis. *Science* 292(5519)**,** 1096-1099. doi: 10.1126/science.1058543.

Oliveros, J.C. (2007). "VENNY. An interactive tool for comparing lists with Venn diagrams".).

Olsen, E.K., Søderholm, K.L., Isaksson, J., Andersen, J.H., and Hansen, E. (2016). Metabolomic profiling reveals the N-Acyl-taurine geodiataurine in extracts from the marine sponge *Geodia macandrewii* (Bowerbank). *J. Nat. Prod.* 79(5)**,** 1285-1291. doi: 10.1021/acs.jnatprod.5b00966.

Olson, J.B., and McCarthy, P.J. (2005). Associated bacterial communities of twodeep-water sponges. *Aquat. Microb. Ecol.* 39(1)**,** 47-55.

Onuki, H., and Kamino, K. (2000). Bacterial growth stimulation with exogenous siderophore and synthetic N-acyl homoserine lactone autoinducers under iron-limited and low-nutrient conditions. *Appl. Environ. Microbiol.* 66(7)**,** 2797-2803.

Overbeek, R., Olson, R., Pusch, G.D., Olsen, G.J., Davis, J.J., Disz, T., et al. (2014). The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res* 42(D1)**,** D206-D214. doi: 10.1093/nar/gkt1226.

Padua, A., Lanna, E., and Klautau, M. (2016). Macrofauna inhabiting the sponge< span class=" italic"> Paraleucilla magna(Porifera: Calcarea) in Rio de Janeiro, Brazil. *J. Mar. Biol. Assoc. U. K.* 96(3)**,** 605-614. doi: 10.1017/S0025315412001804.

Pantos, O., Bongaerts, P., Dennis, P.G., Tyson, G.W., and Hoegh-Guldberg, O. (2015). Habitat-specific environmental conditions primarily control the microbiomes of the coral *Seriatopora hystrix*. *ISME J* 9(9)**,** 1916-1927. doi: 10.1038/ismej.2015.3.

Parks, D.H., Tyson, G.W., Hugenholtz, P., and Beiko, R.G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics* 30(21)**,** 3123-3124. doi: 10.1093/bioinformatics/btu494.

Penesyan, A., Tebben, J., Lee, M., Thomas, T., Kjelleberg, S., Harder, T., et al. (2011). Identification of the Antibacterial Compound Produced by the Marine Epiphytic Bacterium Pseudovibrio sp. D323 and Related Sponge-Associated Bacteria. *Mar Drugs* 9(8)**,** 1391. doi: 10.3390/md9081391.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2010). "IDBA – a practical iterative de Bruijn graph de novo assembler," in *Research in Computational Molecular Biology: 14th Annual International Conference, RECOMB 2010, Lisbon, Portugal, April 25-28, 2010. Proceedings,* ed. B. Berger. (Berlin, Heidelberg: Springer Berlin Heidelberg), 426-440.

Peng, Y., Leung, H.C.M., Yiu, S.M., and Chin, F.Y.L. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinforma Oxf Engl* 28. doi: 10.1093/bioinformatics/bts174.

Pérez-Brocal, V., Gil, R., Ramos, S., Lamelas, A., Postigo, M., Michelena, J.M., et al. (2006). A small microbial genome: the end of a long symbiotic relationship? *Science* 314(5797)**,** 312-313. doi: 10.1126/science.1130441.

Piel, J. (2002). A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. U.S.A.* 99(22)**,** 14002-14007. doi: 10.1073/pnas.222481399.

Piel, J. (2009). Metabolites from symbiotic bacteria. *Nat. Prod. Rep.* 26(3)**,** 338-362. doi: 10.1039/b703499g.

Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., et al. (2004). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci USA* 101(46)**,** 16222-16227. doi: 10.1073/pnas.0405976101.

Pineda, M.C., Duckworth, A., and Webster, N. (2015). Appearance matters: sedimentation effects on different sponge morphologies. *J. Mar. Biol. Assoc. U. K.* 96(02)**,** 481-492. doi: 10.1017/s0025315414001787.

Pita, L., Turon, X., López-Legentil, S., and Erwin, P.M. (2013). Host rules: spatial stability of bacterial communities associated with marine sponges (*Ircinia* spp.) in the Western Mediterranean Sea. *FEMS Microbiol. Ecol.* 86(2)**,** 268-276. doi: 10.1111/1574-6941.12159.

Polónia, A.R., Cleary, D.F., Duarte, L.N., de Voogd, N.J., and Gomes, N.C. (2014). Composition of archaea in seawater, sediment, and sponges in the kepulauan seribu reef system, Indonesia. *Microb. Ecol.* 67(3)**,** 553-567. doi: 10.1007/s00248-013-0365-2.

Radax, R., Hoffmann, F., Rapp, H.T., Leininger, S., and Schleper, C. (2012). Ammonia-oxidizing archaea as main drivers of nitrification in cold-water sponges. *Environ. Microbiol.* 14(4)**,** 909-923. doi: 10.1111/j.1462-2920.2011.02661.x.

Ranea, J.A., Buchan, D.W., Thornton, J.M., and Orengo, C.A. (2004). Evolution of protein superfamilies and bacterial genome size. *J. Mol. Biol.* 336(4)**,** 871-887. doi: 10.1016/j.jmb.2003.12.044.

Rappé, M.S. (2013). Stabilizing the foundation of the house that ‘omics builds: the evolving value of cultured isolates to marine microbiology. *Curr Opin Microbiol* 16(5)**,** 618-624. doi: 10.1016/j.mib.2013.09.009.

Rawls, J.F., Mahowald, M.A., Goodman, A.L., Trent, C.M., and Gordon, J.I. (2007). In vivo imaging and genetic analysis link bacterial motility and symbiosis in the zebrafish gut. *Proc. Natl. Acad. Sci. U.S.A.* 104(18)**,** 7622-7627. doi: 10.1073/pnas.0702386104.

RCoreTeam (2015). R (Version 3.2. 3)[Computer software]. *Vienna, Austria: R Foundation for Statistical Computing*.

Reiswig, H.M. (1981). Partial Carbon and Energy Budgets of the Bacteriosponge Verohgia fistularis (Porifera: Demospongiae) in Barbados. *Mar. Ecol.* 2(4)**,** 273-293. doi: 10.1111/j.1439-0485.1981.tb00271.x.

Ren, Q., and Paulsen, I.T. (2007). Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *J. Mol. Microbiol. Biotechnol.* 12(3-4)**,** 165-179. doi: 10.1159/000099639.

Reynolds, D., and Thomas, T. (2016). Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol. Ecol.* 25(20), 5242-5253. doi: 10.1111/mec.13812.

Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res* 38(20), e191-e191. doi: 10.1093/nar/gkq747.

Richter, M., Rosselló-Móra, R., Oliver Glöckner, F., and Peplies, J. (2016). JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32(6), 929-931. doi: 10.1093/bioinformatics/btv681.

Rossez, Y., Wolfson, E.B., Holmes, A., Gally, D.L., and Holden, N.J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. *PLoS Path* 11(1), e1004483. doi: 10.1371/journal.ppat.1004483.

Rua, C.P., Gregoracci, G.B., Santos, E.O., Soares, A.C., Francini-Filho, R.B., and Thompson, F. (2015). Potential metabolic strategies of widely distributed holobionts in the oceanic archipelago of St Peter and St Paul (Brazil). *FEMS Microbiol. Ecol.* 91(6), fiv043. doi: 10.1093/femsec/fiv043.

Sait, M., Hugenholtz, P., and Janssen, P.H. (2002). Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environ. Microbiol.* 4(11), 654-666.

Santos, A.L., Gomes, N.C.M., Henriques, I., Almeida, A., Correia, A., and Cunha, Â. (2012). Contribution of reactive oxygen species to UV-B-induced damage in bacteria. *J Photochem Photobiol B: Biol* 117(Supplement C), 40-46. doi: 10.1016/j.jphotobiol.2012.08.016.

Schippers, K.J., Sipkema, D., Osinga, R., Smidt, H., Pomponi, S.A., Martens, D.E., et al. (2012). 6 Cultivation of Sponges, Sponge Cells and Symbionts: Achievements and Future Prospects. *Adv. Mar. Biol.* 62, 273.

Schmidt, J.L., Deming, J.W., Jumars, P.A., and Keil, R.G. (1998). Constancy of bacterial abundance in surficial marine sediments. *Limnol. Oceanogr* 43(5), 976-982. doi: 10.4319/lo.1998.43.5.0976.

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(3), 564-576. doi: 10.1038/ismej.2011.116.

Schmitt, S., Weisz, J.B., Lindquist, N., and Hentschel, U. (2007). Vertical transmission of a phylogenetically complex microbial consortium in the viviparous sponge *Ircinia felix*. *Appl. Environ. Microbiol.* 73(7), 2067-2078. doi: 10.1128/aem.01944-06.

Schönberg, C.H.L. (2016). Happy relationships between marine sponges and sediments – a review and some observations from Australia. *J. Mar. Biol. Assoc. U. K.* 96(2), 493-514. doi: 10.1017/S0025315415001411.

Schuller-Levis, G.B., and Park, E. (2003). Taurine: new implications for an old amino acid. *FEMS Microbiol Lett* 226(2), 195-202. doi: 10.1016/S0378-1097(03)00611-6.

Sharon, I., and Banfield, J.F. (2013). Genomes from metagenomics. *Science* 342(6162), 1057-1058. doi: 10.1126/science.1247023.

Siegl, A., and Hentschel, U. (2010). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ Microbiol Rep* 2(4)**,** 507-513. doi: 10.1111/j.1758-2229.2009.00057.x.

Siegl, A., Kamke, J., Hochmuth, T., Piel, J., Richter, M., Liang, C., et al. (2011). Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges. *ISME J* 5(1)**,** 61-70. doi: 10.1371/journal.pone.0087353.

Silver, S., and Phung, L.T. (2005). Genes and enzymes involved in bacterial oxidation and reduction of inorganic arsenic. *Appl. Environ. Microbiol.* 71(2)**,** 599-608. doi: 10.1128/AEM.71.2.599–608.2005.

Simister, R.L., Deines, P., Botte, E.S., Webster, N.S., and Taylor, M.W. (2012). Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. *Environ. Microbiol.* 14(2)**,** 517-524. doi: 10.1111/j.1462-2920.2011.02664.x.

Simon, M., Scheuner, C., Meier-Kolthoff, J.P., Brinkhoff, T., Wagner-Dobler, I., Ulbrich, M., et al. (2017). Phylogenomics of Rhodobacteraceae reveals evolutionary adaptation to marine and non-marine habitats. *ISME J* 11(6)**,** 1483–1499. doi: 10.1038/ismej.2016.198.

Sipkema, D., Osinga, R., Schatton, W., Mendola, D., Tramper, J., and Wijffels, R.H. (2005). Large-scale production of pharmaceuticals by marine sponges: Sea, cell, or synthesis? *Biotechnol Bioeng* 90(2)**,** 201-222. doi: 10.1002/bit.20404.

Sipkema, D., Schippers, K., Maalcke, W.J., Yang, Y., Salim, S., and Blanch, H.W. (2011). Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge Haliclona (gellius) sp. *Appl. Environ. Microbiol.* 77(6)**,** 2130-2140. doi: 10.1128/AEM.01203-10.

Skaar, E.P., Humayun, M., Bae, T., DeBord, K.L., and Schneewind, O. (2004). Iron-source preference of *Staphylococcus aureus* infections. *Science* 305(5690)**,** 1626-1628. doi: 10.1126/science.1099930.

Slaby, B.M. (2017). *Exploring the microbiome of the Mediterranean sponge Aplysina aerophoba by single-cell and metagenomics.* PhD thesis, Wurzburg University.

Slaby, B.M., Hackl, T., Horn, H., Bayer, K., and Hentschel, U. (2017). Metagenomic binning of a marine sponge microbiome reveals unity in defense but metabolic specialization. *ISME J* 11**,** 2465–2478. doi: 10.1038/ismej.2017.101.

Spoerner, M., Wichard, T., Bachhuber, T., Stratmann, J., and Oertel, W. (2012). Growth and thallus morphogenesis of *Ulva mutabilis* (Chlorophyta) depends on a combination of two bacterial species excreting regulatory factors. *J. Phycol.* 48(6)**,** 1433-1447. doi: 10.1111/j.1529-8817.2012.01231.x.

Stein, J.L., Marsh, T.L., Wu, K.Y., Shizuya, H., and DeLong, E.F. (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J. Bacteriol.* 178(3)**,** 591-599.

Steinert, G., Whitfield, S., Taylor, M.W., Thoms, C., and Schupp, P.J. (2014). Application of diffusion growth chambers for the cultivation of marine sponge-associated bacteria. *Mar Biotechnol (NY)* 16(5)**,** 594-603. doi: 10.1007/s10126-014-9575-y.

Stinear, T.P., Mve-Obiang, A., Small, P.L.C., Frigui, W., Pryor, M.J., Brosch, R., et al. (2004). Giant plasmid-encoded polyketide synthases produce the macrolide toxin of *Mycobacterium ulcerans*. *Proc Natl Acad Sci USA* 101(5)**,** 1345-1349. doi: 10.1073/pnas.0305877101.

Stocker, R., and Seymour, J.R. (2012). Ecology and physics of bacterial chemotaxis in the ocean. *Microbiol. Mol. Biol. Rev.* 76(4)**,** 792-812. doi: 10.1128/MMBR.00029-12.

Sweet, M., Bulling, M., and Cerrano, C. (2015). A novel sponge disease caused by a consortium of micro-organisms. *Coral Reefs* 34(3)**,** 871-883. doi: 10.1007/s00338-015-1284-0.

Szpilewska, H., Czyż, A., and Wgrzyn, G. (2003). Experimental evidence for the physiological role of bacterial luciferase in the protection of cells against oxidative stress. *Curr Microbiol* 47(5)**,** 379-382. doi: 10.1007/s00284-002-4024-y.

Tamaki, H., Hanada, S., Sekiguchi, Y., Tanaka, Y., and Kamagata, Y. (2009). Effect of gelling agent on colony formation in solid cultivation of microbial community in lake sediment. *Environ. Microbiol.* 11(7)**,** 1827-1834. doi: 10.1111/j.1462-2920.2009.01907.x.

Tatusov, R.L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., et al. (2003). The COG database: an updated version includes eukaryotes. *BMC Bioinform.* 4**,** 41. doi: 10.1186/1471-2105-4-41.

Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007a). Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol. Mol. Biol. Rev.* 71(2)**,** 295-347. doi: 10.1128/MMBR.00040-06.

Taylor, M.W., Thacker, R.W., and Hentschel, U. (2007b). Evolutionary insights from sponges. *Science* 316(5833)**,** 1854-1855. doi: 10.1126/science.1144387.

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(2)**,** 438-443. doi: 10.1038/ismej.2012.111.

Tessler, M., Neumann, J.S., Afshinnekoo, E., Pineda, M., Hersch, R., Velho, L.F.M., et al. (2017). Large-scale differences in microbial biodiversity discovery between 16S amplicon and shotgun sequencing. *Sci. Rep.* 7(1)**,** 6589. doi: 10.1038/s41598-017-06665-3.

Thomas, T., Moitinho-Silva, L., Lurgi, M., Björk, J.R., Easson, C., Astudillo-García, C., et al. (2016). Diversity, structure and convergent evolution of the global sponge microbiome. *Nat. Commun.* 7**,** 11870. doi: 10.1038/ncomms11870.

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(12)**,** 1557-1567. doi: 10.1038/ismej.2010.74.

Tian, R.-M., Zhang, W., Cai, L., Wong, Y.-H., Ding, W., and Qian, P.-Y. (2017). Genome Reduction and Microbe-Host Interactions Drive Adaptation of a Sulfur-Oxidizing Bacterium Associated with a Cold Seep Sponge. *mSystems* 2(2)**,** e00184-00116. doi: 10.1128/mSystems.00184-16.

Tian, R.M., Wang, Y., Bougouffa, S., Gao, Z.M., Cai, L., Bajic, V., et al. (2014). Genomic analysis reveals versatile heterotrophic capacity of a potentially symbiotic sulfur-oxidizing bacterium in sponge. *Environ. Microbiol.* 16(11)**,** 3548-3561.

Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., et al. (2009). Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet* 5(1)**,** e1000344. doi: 10.1371/journal.pgen.1000344.

Tujula, N.A., Crocetti, G.R., Burke, C., Thomas, T., Holmstrom, C., and Kjelleberg, S. (2009). Variability and abundance of the epiphytic bacterial community associated with a green marine Ulvacean alga. *ISME J* 4(2)**,** 301-311.

Utada, A.S., Bennett, R.R., Fong, J.C.N., Gibiansky, M.L., Yildiz, F.H., Golestanian, R., et al. (2014). *Vibrio cholerae* use pili and flagella synergistically to effect motility switching and conditional surface attachment. 5**,** 4913. doi: 10.1038/ncomms5913.

Vacelet, J. (1995). Carnivorous sponges. *Nature* 373(6512)**,** 333-335.

Vacelet, J., and Donadey, C. (1977). Electron microscope study of the association between some sponges and bacteria. *J. Exp. Mar. Biol. Ecol.* 30(3)**,** 301-314. doi: 10.1016/0022-0981(77)90038-7.

Van Elsas, J.D., Costa, R., Jansson, J., Sjöling, S., Bailey, M., Nalin, R., et al. (2008). The metagenomics of disease-suppressive soils – experiences from the METACONTROL project. *Trends Biotechnol* 26(11)**,** 591-601. doi: 10.1016/j.tibtech.2008.07.004.

Van Soest, R.W.M., Boury-Esnault, N., Vacelet, J., Dohrmann, M., Erpenbeck, D., De Voogd, N.J., et al. (2012). Global diversity of sponges (Porifera). *PLoS. ONE* 7(4)**,** e35105. doi: 10.1371/journal.pone.0035105.

Van Soest, R.W.M.B.-E., N. Hooper, J.N.A. Rützler, K., de Voogd, N.J.A., B.; Hajdu, E., Pisera, A.B., Manconi, R.S., C. Klautau, M. , Picton, B.K., M. Vacelet, J., Dohrmann, M., et al. "World Porifera database. Accessed at http://www.marinespecies.org/porifera on 2017-12-16".).

Varga, J.J., Therit, B., and Melville, S.B. (2008). Type IV Pili and the CcpA Protein Are Needed for Maximal Biofilm Formation by the Gram-Positive Anaerobic Pathogen Clostridium perfringens. *Infect Immun* 76(11)**,** 4944-4951. doi: 10.1128/IAI.00692-08.

Versluis, D., McPherson, K., van Passel, M.W.J., Smidt, H., and Sipkema, D. (2017). Recovery of Previously Uncultured Bacterial Genera from Three Mediterranean Sponges. *Mar. Biotechnol*. doi: 10.1007/s10126-017-9766-4.

Vilanova, E., Coutinho, C.C., and Mourão, P.A.S. (2009). Sulfated polysaccharides from marine sponges (Porifera): an ancestor cell–cell adhesion event based on the carbohydrate–carbohydrate interaction. *Glycobiology* 19(8)**,** 860-867. doi: 10.1093/glycob/cwp059.

Voultsiadou, E. (2007). Sponges: an historical survey of their knowledge in Greek antiquity. *J. Mar. Biol. Assoc. U. K.* 87(6)**,** 1757-1763. doi: 10.1017/S0025315407057773.

Voultsiadou, E., Vafidis, D., and Antoniadou, C. (2008). Sponges of economical interest in the Eastern Mediterranean: an assessment of diversity and population density. *J. Nat. Hist.* 42(5-8)**,** 529-543. doi: 10.1080/00222930701835506.

Vuilleumier, S. (1997). Bacterial glutathione S-transferases: what are they good for? *J. Bacteriol.* 179(5), 1431-1441.

Wacey, D., Kilburn, M.R., Saunders, M., Cliff, J., and Brasier, M.D. (2011). Microfossils of sulphur-metabolizing cells in 3.4-billion-year-old rocks of Western Australia. *Nature Geoscience* 4(10), 698-702. doi: 10.1038/ngeo1238.

Wadhams, G.H., and Armitage, J.P. (2004). Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* 5(12), 1024-1037. doi: 10.1038/nrm1524.

Wagner-Döbler, I., and Biebl, H. (2006). Environmental biology of the marine *Roseobacter* lineage. *Annu. Rev. Microbiol.* 60(1), 255-280. doi: doi:10.1146/annurev.micro.60.080805.142115.

Wang, Y., Hu, X.-J., Zou, X.-D., Wu, X.-H., Ye, Z.-Q., and Wu, Y.-D. (2015). WDSPdb: a database for WD40-repeat proteins. *Nucleic Acids Res* 43(D1), D339-D344. doi: 10.1093/nar/gku1023.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H.U., Bruccoleri, R., et al. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Res* 43(W1), W237-W243.

Webster, N., and Hill, R. (2001). The culturable microbial community of the Great Barrier Reef sponge Rhopaloeides odorabile is dominated by an α-Proteobacterium. *Mar Biol* 138(4), 843-851. doi: 10.1007/s002270000503.

Webster, N.S. (2007). Sponge disease: a global threat? *Environ. Microbiol.* 9(6), 1363-1375. doi: 10.1111/j.1462-2920.2007.01303.x.

Webster, N.S., and Taylor, M.W. (2012). Marine sponges and their microbial symbionts: love and other relationships. *Environ. Microbiol.* 14(2), 335-346. doi: 10.1111/j.1462-2920.2011.02460.x.

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(8), 2070-2082. doi: 10.1111/j.1462-2920.2009.02065.x.

Webster, N.S., and Thomas, T. (2016). The sponge hologenome. *MBio* 7(2), e00135–00116. doi: 10.1128/mBio.00135-16.

Weiss, S., Xu, Z.Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5(1), 27. doi: 10.1186/s40168-017-0237-y.

Westinga, E., and Hoetjes, P.C. (1981). The intrasponge fauna of Spheciospongia vesparia (Porifera, Demospongiae) at Curaçao and bonaire. *Mar Biol* 62(2), 139-150. doi: 10.1007/bf00388176.

Whalan, S., and Webster, N.S. (2014). Sponge larval settlement cues: the role of microbial biofilms in a warming ocean. *Sci. Rep.* 4, 4072. doi: 10.1038/srep04072.

White, J.R., Nagarajan, N., and Pop, M. (2009). Statistical methods for detecting differentially abundant features in clinical metagenomic samples. *PLoS Comp Biol* 5(4), e1000352. doi: 10.1371/journal.pcbi.1000352.

Wilke, A., Harrison, T., Wilkening, J., Field, D., Glass, E.M., Kyrpides, N., et al. (2012). The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools. *BMC Bioinform.* 13(1), 141. doi: 10.1186/1471-2105-13-141.

Wilkens, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Reports* 7, 14. doi: 10.12703/P7-14.

Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231. doi: 10.1038/nature12779.

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(7486), 58-62. doi: 10.1038/nature12959.

Woyke, T., Xie, G., Copeland, A., González, J.M., Han, C., Kiss, H., et al. (2009). Assembling the marine metagenome, one cell at a time. *PLoS. ONE* 4(4), e5299. doi: 10.1371/journal.pone.0005299.

Wu, S., Zhu, Z., Fu, L., Niu, B., and Li, W. (2011). WebMGA: a customizable web server for fast metagenomic sequence analysis. *BMC Genomics* 12(1), 444.

Wulff, J.L. (2006). Ecological interactions of marine sponges. *Can. J. Zool.* 84(2), 146-166.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. *Protein Cell* 2(3), 202-214. doi: 10.1007/s13238-011-1018-1.

Yahel, G., Whitney, F., Reiswig, H.M., Eerkes-Medrano, D.I., and Leys, S.P. (2007). In situ feeding and metabolism of glass sponges (Hexactinellida, Porifera) studied in a deep temperate fjord with a remotely operated submersible. *Limnol. Oceanogr* 52(1), 428-440. doi: 10.4319/lo.2007.52.1.0428.

Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-ya, K., Omura, S., Cane, D.E., et al. (2015). Terpene synthases are widely distributed in bacteria. *Proc. Natl. Acad. Sci. U.S.A.* 112(3), 857-862. doi: 10.1073/pnas.1422108112.

Yi, H., and Chun, J. (2006). Thalassobius aestuarii sp. nov., isolated from tidal flat sediment. *The Journal of Microbiology* 44(2), 171-176.

Yung, P.Y., Burke, C., Lewis, M., Kjelleberg, S., and Thomas, T. (2011). Novel antibacterial proteins from the microbial communities associated with the sponge Cymbastela concentrica and the green alga Ulva australis. *Appl. Environ. Microbiol.* 77(4), 1512-1515. doi: 10.1128/AEM.02038-10.

Zan, J., Cicirelli, E.M., Mohamed, N.M., Sibhatu, H., Kroll, S., Choi, O., et al. (2012). A complex LuxR–LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. *Mol. Microbiol.* 85(5), 916-933. doi: 10.1111/j.1365-2958.2012.08149.x.

Zhang, F., Blasiak, L.C., Karolin, J.O., Powell, R.J., Geddes, C.D., and Hill, R.T. (2015). Phosphorus sequestration in the form of polyphosphate by microbial symbionts in marine sponges. *Proc. Natl. Acad. Sci. U.S.A.* 112(14), 4381-4386. doi: 10.1073/pnas.1423768112.

Zhang, Y., Xiao, W., and Jiao, N. (2016). Linking biochemical properties of particles to particle-attached and free-living bacterial community structure along the particle density gradient from freshwater to open ocean. *J Geophys Res Biogeosci* 121(8)**,** 2261-2274. doi: 10.1002/2016JG003390.