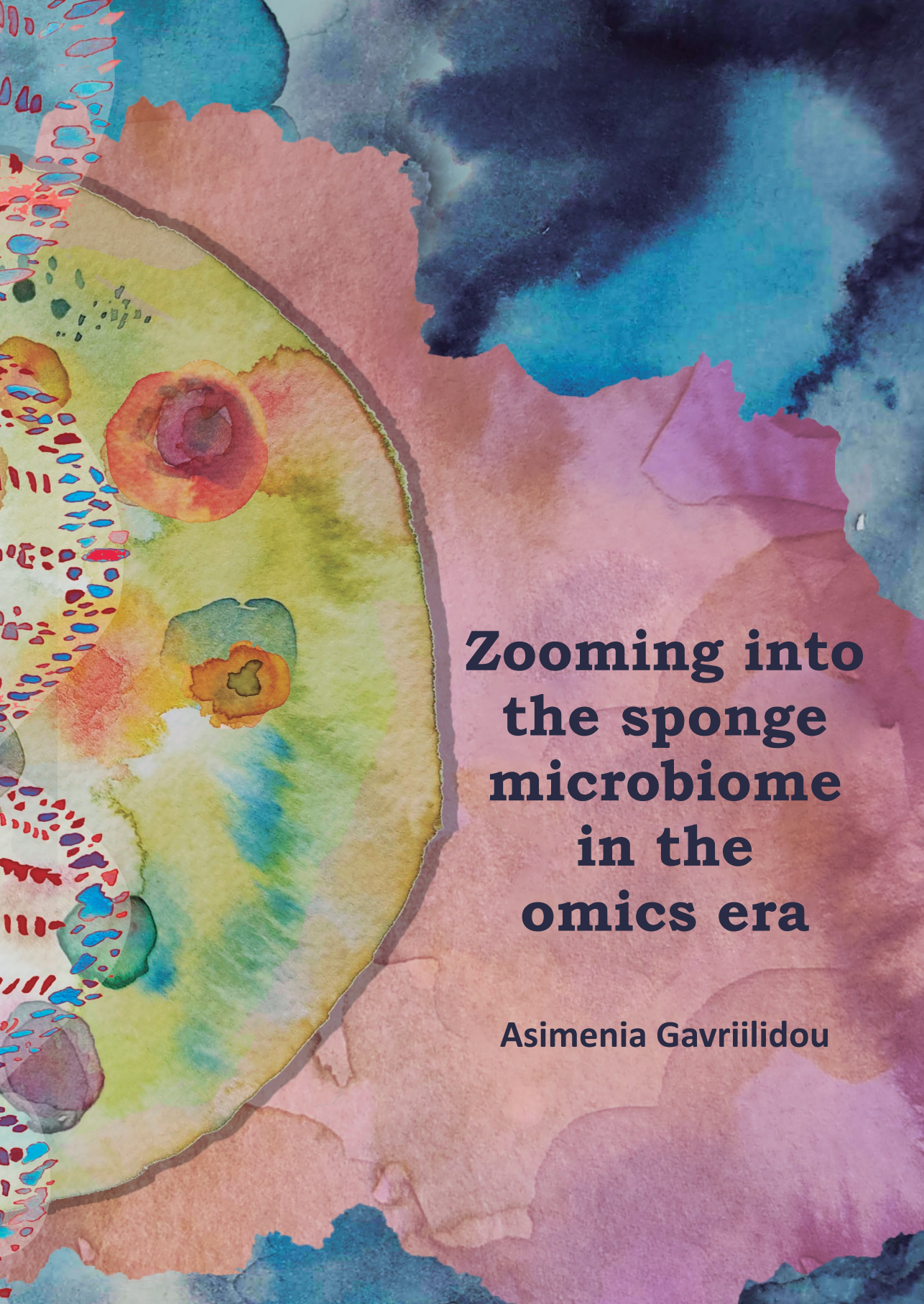


Zooming into the sponge microbiome in the omics era

Asimenia Gavriilidou

Propositions

1. To successfully maintain their residency, sponge symbionts possess a 'defense toolbox' to battle pathogens and host phagocytosis.
   (this thesis)

2. Genome-centric metagenomic methods reveal the hidden diversity and metabolic potential of the sponge microbiome.
   (this thesis)

3. One should always define a scientific question before the design of an experiment.

4. The COVID-19 pandemic reflected how challenging dissemination of science to the public is.

5. Investing time on self-reflection is a key PhD lifehack.

6. Peer review should be included in the curriculum of PhD training programs.

Propositions belonging to the thesis, entitled
"Zooming into the sponge microbiome in the omics era"

Asimenia Gavriilidou
Wageningen, 25 May 2022

Zooming into the sponge microbiome in the omics era

Asimenia Gavriilidou

Thesis committee

Promotors
Dr Detmer Sipkema
Associate professor at the Laboratory of Microbiology
Wageningen University & Research

Prof. Dr Hauke Smidt
Personal chair at the Laboratory of Microbiology
Wageningen University & Research

Co-promotor
Dr Colin J. Ingham
CEO, Hoekmine BV, Utrecht, The Netherlands

Other members
Prof. Dr Michiel Kleerebezem, Wageningen University & Research
Prof. Dr Bas E. Dutilh, Friedrich-Schiller University Jena, Germany
Dr Serina L. Robinson, EAWAG, Dübendorf, Switzerland
Prof. Dr Nicole de Voogd, Naturalis, Leiden, The Netherlands

This research was conducted under the auspices of the Graduate School for Socio-Economic and Natural Sciences of the Environment (SENSE).

Zooming into the sponge microbiome in the omics era

Asimenia Gavriilidou

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Wednesday 25 May 2022
at 1.30 p.m. in the Omnia Auditorium.

Asimenia Gavriilidou

Zooming into the sponge microbiome in the omics era,
226 pages.

PhD thesis, Wageningen University & Research, Wageningen, the Netherlands (2022)
With references, with summary in English

ISBN: 978-94-6447-187-8
DOI: https://doi.org/10.18174/567803

Για την οικογένεια μου...

Table of Contents

Chapter 1

General Introduction

Marine microbial ecology in a nutshell

Microbial life on Earth emerged about 3.8 billion years ago and microbes were the only living organisms populating Earth up to the appearance of eukaryotes 2 billion years later (Blaser et al., 2016; Munn, 2020). Yet, it was not until the 17th century when Antonie van Leeuwenhoek observed 'animalcules' under the microscope that we realized the existence of microbes (Figure 1). Two centuries after their discovery, the so called "Golden Era" of microbiology began with pioneering work by Louis Pasteur, Robert Koch, Ferdinand Cohn and other scientists (Robinson et al., 2010; Berg et al., 2020). The realization that microorganisms are omnipresent in nature forming complex communities often associated with hosts led to the debut of the field of microbial ecology at the end of the 19th century (Robinson et al., 2010; Berg et al., 2020).

While breakthroughs in microbiology started to reveal the richness of the microbial realm over 350 years ago, researchers began to formulate questions about the diversity and role of microbial life in marine environments only recently (Heidelberg et al., 2010). The foundations of marine microbial ecology as a discipline were laid between the 1930s and 1940s by Claude E. ZoBell who was among the first to study marine microbes with a focus on bacterial attachment to surfaces (biofilm formation) (Kirchman and Gasol, 2018). Microbes from the sea were considered as a 'black box' until improvements in microscopy and new visualization techniques during the 1970s to early 1990s increased the understanding of microbial communities and their significance in ocean food webs (Caron, 2005; Kirchman and Gasol, 2018). The first paradigm shift for marine microbial ecologists was the accurate estimation of the total biomass of planktonic microorganisms along with a wealth of information regarding their ecological role in the ocean (Figure 1) (DeLong, 2009). In the 1980s, other landmarks in the field followed namely the discovery of the cyanobacterium *Prochlorococcus* (Chisholm et al., 1988), the most abundant primary producer in the global ocean and the introduction of the term 'microbial loop' (Azam et al., 1983), the carbon route from dissolved organic matter to higher trophic levels via bacteria and protist predators and back (Kirchman and Gasol, 2018).

In the 1990s, a new aspect of marine microbes was unveiled with the analysis of 16S ribosomal RNA (rRNA) gene sequences as phylogenetic markers that showed the preponderance of bacterial lineages very different from the isolates obtained in the laboratory (Giovannoni et al., 1990). Subsequent findings based on the application of the polymerase-chain reaction (PCR) revealed the presence of *Archaea* in the oceans (DeLong, 1992; Fuhrman et al., 1992), previously considered as "extremophiles" or protist symbionts (Kirchman and Gasol, 2018). At the beginning of this century, advances in molecular biology revolutionized marine microbial ecology bringing light to the unprecedented microbial diversity in marine ecosystems. This led to more directed cultivation efforts and the successful isolation and characterization of the first marine bacteria, archaea and viruses (Rappé et al., 2002; DeLong, 2009; Heidelberg et al., 2010).

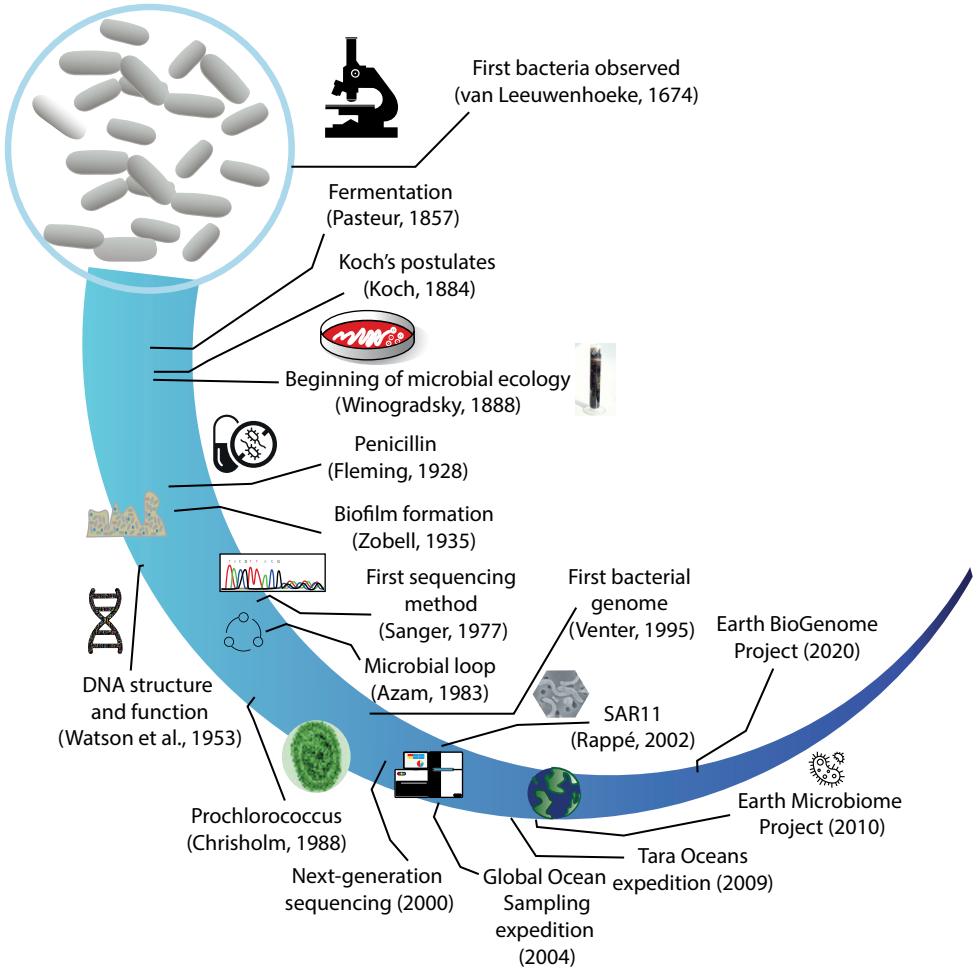


Figure 1. Brief history of scientific breakthroughs in microbiology, molecular biology and marine microbial ecology. Adapted from original artwork of Mariana Ruiz Villarreal (https://commons.wikimedia.org/wiki/File:Timeline_evolution_of_life.svg).

Nowadays, the application of next-generation sequencing technologies and the launch of major sampling expeditions (e.g., Global Ocean Sampling (GOS), Tara Oceans) have resulted in vast amounts of data contributing to major findings, such as the discovery of a new type of phototrophy widespread among marine bacterial taxa (DeLong, 2009; Kirchman and Gasol, 2018). To this point, the microbial ecology of the oceanic surface and coastal environments is relatively well known. However, despite the technological developments and rapid generation of data, there are several understudied habitats within the ocean, such as hydrothermal vents, sediments and microbiomes of marine animals. Exploring these habitats will provide a more comprehensive view on microbial life in the sea and their role in global ecology (Munn, 2020).

Sponges are pore-bearing animals

Sponges are members of the phylum Porifera (from Latin *porus*, pore and *fero*, to bear) and among the most primitive multicellular organisms (Metazoa) (Schutze et al., 1999). Dating back to the Precambrian (Turner, 2021), sponges are true survivors that have evolved into key constituents of marine benthic ecosystems distributed globally from polar to tropical habitats (Hooper et al., 2021). Recent studies support the 'Porifera-first' hypothesis which places sponges as the sister group of all other animals. This is also reflected in their simplistic body architecture as a precursor to complex multicellularity in animals (Nielsen, 2019; Redmond and McLysaght, 2021). Lacking tissues and organs, sponges possess an aquiferous system, which is the basis for their filter-feeding activity. Most sponges secrete a wide variety of spicules (siliceous or calcareous) which function as skeletal components providing structural support (Figure 2) (Taylor et al., 2007a). Epithelial cells called pinacocytes compose the outer layer of the sponge (or pinacoderm) and the internal canals which extend through the pores (ostia) (Simpson, 1984). The second type of cells are known as choanocytes forming specialized chambers inside the sponge where the pumping is generated with the beating of their flagella (Figure 2). Seawater enters from the pores along the canals in the choanocyte chambers where food particles are filtered out and transferred to the connective tissue of the sponge, the mesohyl. This is where another group of cells is present, namely the archaeocytes which are totipotent cells capable of consuming seawater food particles via phagocytosis (Figure 2). After the filtering, water is pumped out of the sponge through openings, called oscula (Simpson, 1984).

Given their high pumping rates (up to 24,000 L per day), sponges can significantly influence pelagic environments due to their suspension feeding and the removal of particles from the water column (Bell, 2008). This is more evident in tropical and cold-water coral reefs where sponges, as efficient filter feeders, convert dissolved organic matter (DOM) to particulate organic matter (POM), mediating the recycling of nutrients within reef food webs ('sponge loop') (de Goeij et al., 2013; Rix et al., 2018; Bart et al., 2021). In order to maintain their dominance in benthic ecosystems, sponges have to deal with spatial competition and predation (Wulff, 2006; Bell, 2008). To succeed in this, they have developed a defensive strategy that involves the production of a vast diversity of biologically active secondary metabolites with antimicrobial, antiviral, cytotoxic and other deterrent effects (Blunt et al., 2017; Blunt et al., 2018). Besides the competitive interactions, sponges are well known for their intimate associations with microbial and macrofaunal communities. Described as 'living hotels', sponges accommodate a large biodiversity of organisms that use them as microhabitats (Bell, 2008). Symbionts vary from facultative sponge-dwelling organisms that also seek protection elsewhere to obligate ones that are either generalists or specialists (Wulff, 2006; Bell, 2008).

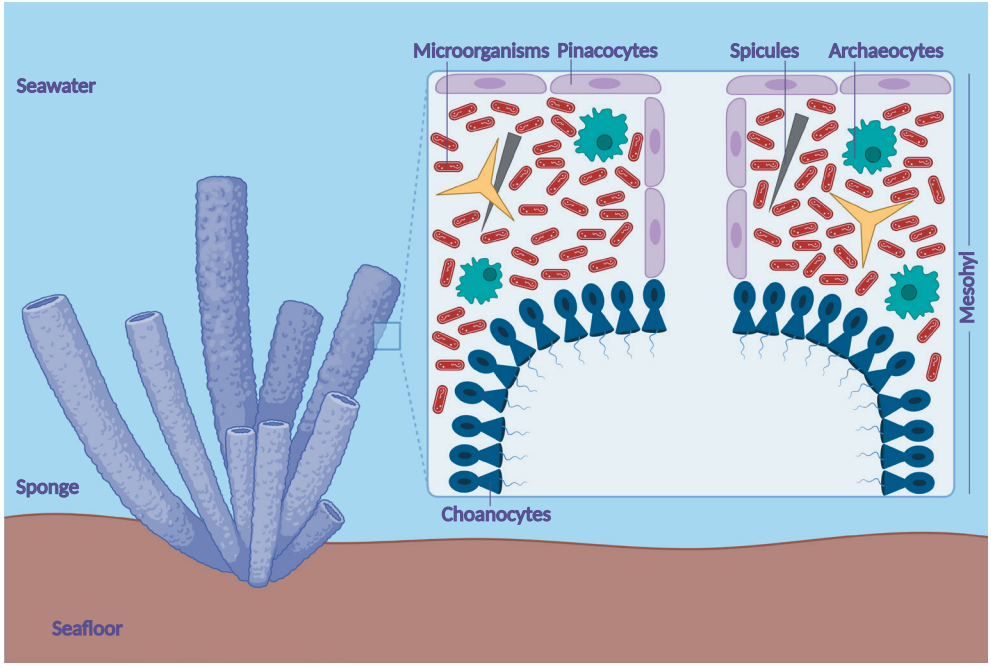


Figure 2. Schematic representation of a sponge on the seafloor and its internal structure (mesohyl). Different types of sponge cells, spicules and microorganisms are displayed.

Sponges ‘n microbes: a symbiosis paradigm

Sponges are known to harbour dense, diverse and yet distinct microbial communities, despite their plain body structure and constant filter-feeding activities (Thomas et al., 2016; Webster and Thomas, 2016). Pioneers in sponge microbiology research described that 40% of the sponge biomass can be comprised of microbial symbionts, thus referring to sponges as ‘bacteriosponges’ (Vacelet and Donadey, 1977; Wilkinson and Fay, 1979; Wilkinson, 1983) or later as ‘microbial fermenters’ (Hentschel et al., 2006). Based on a holistic ecological perspective, sponges are considered ‘holobionts’, a term which describes the host and the associated microbial consortium (Webster and Taylor, 2012; Webster and Thomas, 2016; Pita et al., 2018). Sponge-microbe interactions underpin the evolutionary success of Porifera and represent one of the oldest symbioses between animals and microbes (Hill and Sacristán-Soriano, 2017). There are different scenarios for the symbiont acquisition by the sponge host (Taylor et al., 2007a) with existing evidence for both vertical (from parent sponge to offspring) and horizontal (from surrounding seawater) transmission modes (Webster et al., 2010; Schmitt et al., 2012; Sipkema et al., 2015).

Sponge-associated microorganisms generally reside in the mesohyl around the choanocyte chambers, sometimes within specialized host cells (bacteriocytes) or just below the

pinacoderm (Hentschel et al., 2006; Webster and Thomas, 2016). In some sponge species, microbial populations can reach huge densities of 10^9 cells/cm^3, while other sponges show lower densities roughly equivalent to seawater (10^5-10^6 cells/cm^3) (Hentschel et al., 2006; Schmitt et al., 2012; Webster and Thomas, 2016). These observations led to the classification of sponges into those with high-microbial abundance (HMA) and low-microbial abundance (LMA) (Vacelet and Donadey, 1977; Reiswig, 1981; Hentschel et al., 2003). The HMA-LMA dichotomy has been long considered, although the validity of the concept remains questionable (Hill and Sacristán-Soriano, 2017; Pita et al., 2018). Microbial abundance, diversity (Moitinho-Silva et al., 2017b; Busch, 2021), functional gene content (Bayer et al., 2014) and sponge morphology (Gloeckner et al., 2014; Poppell et al., 2014) have been shown to differ between HMA and LMA sponges. However, there are sponge species hosting an intermediate number of microbes (Gloeckner et al., 2014) while evidence shows a degree of functional relatedness between symbiont communities from divergent hosts, regardless of their HMA-LMA status (Thomas et al., 2010a; Fan et al., 2012). Apparently, there is still a lot to be discovered regarding the symbiont community in the context of this dichotomy.

The sponge microbiome displays remarkable diversity that spans across all major evolutionary lineages of archaea, bacteria and microbial eukaryotes (Taylor et al., 2007b; Taylor et al., 2007a). More than 40 different microbial phyla have been detected in sponges (Thomas et al., 2016). The most dominant bacterial phyla representing the symbiont community are *Proteobacteria*, *Chloroflexota*, *Cyanobacteria*, *Acidobacteriota*, *Actinobacteria*, *Bacteroidota*, *Firmicutes*, *Nitrospirota*, and the candidate phyla PAUC34f and Poribacteria (Thomas et al., 2016; Moitinho-Silva et al., 2017b). Sponge-associated microbial communities are complex and structured as follows: the core microbiota, highly prevalent members in all individuals of a certain species and the variable microbiota, members of which are present only in some individuals (Schmitt et al., 2012; Pita et al., 2018). Generalist symbionts characterize the core sponge microbiome while specialists are under-represented (Thomas et al., 2016). Until recently, it was thought that symbiont communities are 'sponge-specific' (Hentschel et al., 2002; Simister et al., 2012) but it is more likely that they are 'sponge-enriched', as clusters previously defined as sponge-specific are widespread, albeit at low relative abundances, in diverse marine habitats (Taylor et al., 2013; Thomas et al., 2016; Webster and Thomas, 2016). The exact nature of their intimate associations remains unclear and ranges from beneficial (mutualism or commensalism) to harmful for the host (pathogenesis, parasitism and fouling) (Taylor et al., 2007a; Hill and Sacristán-Soriano, 2017). Modern sequencing efforts have revealed that sponges harbor species-specific and relatively stable microbiomes irrespective of their habitat, depth, season or their continuous filter-feeding activity (Thomas et al., 2016; Pita et al., 2018). Yet, there is evidence suggesting that environmental factors such as temperature, salinity and depth influence the composition of the sponge microbiome.

Since cultured representatives of the majority of taxa of the sponge microbiome are lacking, knowledge on the functional roles of the symbionts has been mostly obtained by

culture-independent, omics-based approaches (Hill and Sacristán-Soriano, 2017; Moitinho-Silva et al., 2017b). Shared core functions have been identified between microbiomes of phylogenetically distant sponge species indicating functional convergence (Thomas et al., 2010a; Fan et al., 2012; Ribes et al., 2012). These metabolic features are mainly linked to the acquisition of nutrients (e.g., carbon, nitrogen) present in the host environment, the supply of vitamins to the host, the interactions between symbionts, defense and stress response (Webster and Thomas, 2016; Pita et al., 2018). The majority of symbionts rely on heterotrophy and thus transport and utilize nutrients derived from the filtered seawater or the host itself, for example by degrading complex carbohydrates or the proteoglycan sponge matrix (Kamke et al., 2013; Slaby et al., 2017; Astudillo-Garcia et al., 2018; Bayer et al., 2018; Podell et al., 2018; Sizikov et al., 2020; Robbins et al., 2021; Taylor et al., 2021). Nitrogen metabolism is also evident among sponge-associated microbes due to the large amount of ammonia made available by the host. Symbionts either use ammonia as energy supply or nitrogen source to make organic compounds or assist the sponge in detoxification (Karimi et al., 2019; Engelberts et al., 2020; Moreno-Pino et al., 2020; Taylor et al., 2021). Another potential benefit to the host could be the microbially mediated synthesis of essential vitamins and amino acids the sponge is otherwise not capable of producing (Astudillo-Garcia et al., 2018; Karimi et al., 2019; Engelberts et al., 2020). In order to establish their residency, sponge-dwelling microbes need to protect themselves from host phagocytes and both themselves and the host from viruses, pathogens and toxins (Webster and Thomas, 2016; Hill and Sacristán-Soriano, 2017; Pita et al., 2018). Omics studies have revealed that symbionts encode effective mechanisms against the aforementioned threats, such as eukaryotic-like proteins (ELPs), clustered regularly interspaced short palindromic repeats (CRISPRs), universal stress proteins and toxin-antitoxin systems (TAs) (Burgsdorf et al., 2015; Horn et al., 2016; Karimi et al., 2017; Slaby et al., 2017; Astudillo-Garcia et al., 2018; Bayer et al., 2018; Podell et al., 2018; Moreno-Pino et al., 2020; Sizikov et al., 2020; Robbins et al., 2021). Moreover, sponge-associated microbes contribute to the chemical defense of the host against predators and epibionts via the production of a wide array of biologically active secondary metabolites (Sipkema et al., 2005; Piel, 2009; Indraningrat et al., 2016; Lackner et al., 2017; Schorn et al., 2019). The presence of these functional traits in microbiomes of different sponge species regardless of biogeography implies the adaptation of the associated microbiota towards a successful symbiosis with the host (Thomas et al., 2010a; Fan et al., 2012; Ribes et al., 2012).

What about cultivation? The challenge

Nature hosts an enormous microbial diversity, and yet there is a vast discrepancy between the currently known phyla and those with representative isolates obtained in pure culture (Stewart, 2012; Gutleben et al., 2017). The realization that most microbes seen under the microscope do not produce colonies on traditional cultivation media first occurred in the

1980s and was described as the 'great plate count anomaly' (Staley and Konopka, 1985). This 'unseen majority' (Whitman et al., 1998) was later confirmed with the use of molecular tools, which shed light on the diverse and abundant microbial groups present in numerous niches ranging from soil to human-made systems, sediments, animals and the human body (Stewart, 2012). It accounts for 85 to 99% of all bacteria and archaea and thus, impeding the scientific progress in several aspects, for example in drug biodiscovery (Lok, 2015). Besides the biotechnological and industrial potential of microbes, cultivation is equally significant in order to comprehensively determine their features (metabolism, physiology, cell biology) and understand their role in their respective ecosystems (Lewis and Ettema, 2019; Mu et al., 2020; Lewis et al., 2021). There are many factors that affect the culturability of microorganisms such as finding the appropriate growth conditions and deal with cell dormancy, environmental fluctuations, low abundance and competition (Lewis and Ettema, 2019; Lewis et al., 2021).

A representative example of microbes recalcitrant to cultivation are sponge-associated bacteria. In this case, another level of complexity is added as symbiotic interdependencies need to be taken into account (Schippers et al., 2012). Most conventional cultivation strategies aim at simulating the natural environment by providing the necessary nutrients and growth conditions for the targeted species (Nichols, 2007; Gutleben et al., 2017). There are certain difficulties though, in mimicking the native milieu of the sponge such as the occasional anoxic conditions in the mesohyl when pumping is ceased or the complex interactions between sponges and their microbiomes including cross-feeding of metabolites and signaling (Hoffmann et al., 2006; Sipkema et al., 2009; Sipkema et al., 2011; Pande and Kost, 2017). Many cultivation attempts beyond the traditional plating methods have led to the successful isolation of previously uncultured sponge-associated bacteria. These studies were based on alternative cultivation approaches including the use of diffusion devices (Jung et al., 2014), floating polycarbonate filters (Sipkema et al., 2011), stratified cultivation systems (Gutleben et al., 2020), addition of antibiotics (Versluis et al., 2017b), sponge-derived antimicrobials (Gutleben et al., 2020), arsenic (Keren et al., 2015) or sponge tissue extracts (Webster et al., 2001; Abdelmohsen et al., 2010; Sipkema et al., 2011). In some cases, bacterial recoverability reached 14% of the total diversity (Sipkema et al., 2011). Nevertheless, the majority of the sponge-derived isolates did not reflect the microbial community composition estimated by cultivation-independent methods, reinforcing the gap between the cultured and uncultured microbial fraction (Schippers et al., 2012; Esteves et al., 2016). Therefore, no major phylotypes residing in sponges like the ones belonging to the phyla *Acidobacteriota*, *Chloroflexota*, *Cyanobacteria* or *Ca.* Poribacteria have been amenable to in vitro cultivation (Esteves et al., 2016; Gutleben et al., 2020).

Omics and the data deluge

The advent of high-throughput analytical methods and bioinformatic tools provided increasingly effective ways to circumvent the limitations of cultivation (Gutleben et al., 2017). Current sequencing technologies generate enormous amounts of data, which allow researchers to address fundamental questions about microbial communities from various environmental sources (Eren et al., 2015). The term "omics" includes various disciplines in biological sciences with the suffix -omics which were named according to the type of molecules they are focusing on, such as genes (genomics), RNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) (Vailati-Riboni et al., 2017). Genomics deals with the complete set of genetic information in an organism, transcriptomics does genome-wide expression profiling, proteomics determines function and structure of all sets of proteins, and metabolomics resolves the diversity and abundance of the total cellular metabolites (Zhang et al., 2010; Vailati-Riboni et al., 2017). Integration of the information stemming from the aforementioned omics technologies seems necessary to obtain a global overview of microbial biology (Gutleben et al., 2017). This is defined as the 'multi-omics approach' that can act as a powerful tool for a data-driven, holistic monitoring of functions and dynamics of biological systems in a high-throughput manner (Zhang et al., 2010; Gutleben et al., 2017).

In the late 1990s, shotgun sequencing, particularly in the beginning combined with clone libraries of varying insert size, gave birth to the field of metagenomics which enables direct access to the genetic content of an entire community of organisms inhabiting a certain environment (Handelsman et al., 1998; Rondon et al., 2000; Handelsman, 2004). Since then, metagenomics revolutionized microbial ecology and shed light to the so-called 'microbial dark matter' (Rinke et al., 2013; Lok, 2015). This cultivation-independent method gave insights into the unknown microbial biosphere by revealing entirely new microbial groups otherwise overlooked by rRNA-based or other conventional analyses as well as novel genes and associated functions of even well-known taxa (Baker and Dick, 2013). Comparison between 16S rRNA and shotgun sequencing in microbiome studies has pinpointed several limitations of the rRNA-based strategy such as limited taxonomic resolution among closely related species, proneness to PCR biases and the inability to predict functions. Nevertheless, it always depends on the purpose of the study since 16S rRNA amplicon sequencing can be effectively used as a quick and low-cost method for analyzing the microbial community composition at higher taxonomic levels (Poretsky et al., 2014; Campanaro et al., 2018; Durazzi et al., 2021).

Recent advances in computational tools led to the successful reconstruction of near-complete microbial genomes from environmental metagenomes by a process called binning (Tyson et al., 2004; Luo et al., 2012; Thomas et al., 2012; Alneberg et al., 2018). To date, several binning algorithms have been developed which cluster contigs into groups likely representing genomes of individual species based on sequence composition and/or coverage

(Alneberg et al., 2014; Wu et al., 2014; Kang et al., 2015; Kang et al., 2019). The potential of binning in resolving genomes of uncultivated microbial species paved the way for other computational methods such as comparative genomics, phylogenomics and genome mining that provided a deeper glimpse into the lifestyle and evolution of microbes (Thomas et al., 2012; Baker and Dick, 2013; Eren et al., 2015). In addition, the accelerated discovery rate of yet-uncultured microbial taxa substantially expanded the tree of life (Parks et al., 2017). This led to the realization that there is an urgent need for establishing a nomenclature system to incorporate uncultured taxa based primarily on sequence data (Konstantinidis et al., 2017; Parks et al., 2018; Chuvochina et al., 2019; Murray et al., 2020; Parks et al., 2020; Hugenholtz et al., 2021).

During the past decade, an unprecedented number of genomes has been recovered from numerous environments, including marine sponges. Such data has yielded clues about uncultivated bacterial and archaeal lineages inhabiting different sponge species. Some examples are members of the phyla *Proteobacteria* (Karimi et al., 2019), *Chloroflexota* (Bayer et al., 2018), *Ca.* Poribacteria (Podell et al., 2018), SAUL (Astudillo-Garcia et al., 2018), *Cyanobacteria* (Burgsdorf et al., 2015; Schorn et al., 2019), *Verrucomicrobiota* (Sizikov et al., 2020) and *Thaumarchaeota* (Haber et al., 2020). Hitherto, most omics studies on sponges have focused on resolving the phylogeny and metabolism of specific groups of symbionts (Burgsdorf et al., 2015; Astudillo-Garcia et al., 2018; Bayer et al., 2018; Podell et al., 2018; Schorn et al., 2019; Haber et al., 2020; Sizikov et al., 2020), whereas others provided a more global overview of sponge microbiomes (Slaby et al., 2017; Bayer et al., 2020; Engelberts et al., 2020; Moreno-Pino et al., 2020; Robbins et al., 2021). A lot of attention has been also given to the mining of sponge-associated metagenomes for biosynthetic gene clusters, as marine sponges are among the richest known sources of natural products. Hence, analysis of sponge metagenomes has confirmed the bacterial origin of several biologically active secondary metabolites and revealed many chemically talented sponge symbionts (Wakimoto et al., 2014; Freeman et al., 2016; Nakashima et al., 2016; Lackner et al., 2017; Mori et al., 2018; Tianero et al., 2019; Rust et al., 2020; Storey et al., 2020). Overall, omics-derived knowledge holds great promise for the comprehensive understanding of metazoan host-microbe interactions and for granting access to the biotechnological potential of symbiotic microorganisms.

Research aims and thesis outline

Sponges are among the oldest of all extant animals on the planet and are thus, a success story of the holobiont concept. As introduced in **chapter 1**, microbial ecology effectively began with the discovery that microbes are everywhere and especially in marine environments, both free-living and host-associated. Advances in microscopy and later molecular methods revealed that marine sponges hold a remarkable abundance and diversity of microorganisms inhabiting their mesohyl. This coupled with their significant impact on the marine ecosystem and their extraordinary chemical potential render marine sponges as a promising model system for the study of host-microbe interactions. However, most symbiotic microbes exhibit recalcitrance to in vitro cultivation, which leaves a significant gap of knowledge regarding the relationships between sponges and their microbial symbiotic consortia.

This brings me to the research aim of this thesis, which is to elucidate the phylogeny, primary and secondary metabolism of both cultured and yet-uncultured bacterial lineages representing the sponge microbiome. In this work, we set out to determine functions related to symbiosis and obtain insights into the biosynthetic potential of sponge-dwelling bacteria for secondary metabolites with interesting properties. Overall, the scope of this research was to enhance the understanding of the complex interplay between sponges and their bacterial symbionts by means of cultivation-dependent and cultivation-independent methods.

In **chapter 2**, we explore characteristic biological traits of *Flavobacteriaceae*, the largest family in the phylum *Bacteroidota* in comparison with other major players of the global bacterioplankton. Comparative genomic analysis was performed to investigate features underlying niche adaptation based on genomes of free-living and host-associated bacteria, including seven presumably novel sponge symbionts sequenced here. Genome mining for biosynthetic gene clusters revealed the secondary metabolite repertoire of the studied bacteria. In order to link the genomic properties with observed phenotypes the novel sponge-associated isolates were further subjected to in vitro testing for gliding motility and antimicrobial bioactivity.

Chapter 3 examines the bioactivity of several bacteria isolated from different sponge species against various human pathogens and cancer cells via functional screening. Furthermore, the phylogeny and secondary metabolite biosynthesis potential of the studied strains were assessed based on available genome sequence information. Gene-trait matching analysis followed to identify putative candidate gene clusters or predicted compounds responsible for the observed antimicrobial and anticancer activity.

Chapter 4 describes the phylogenetic structure, habitat distribution and chemical potential of the candidate phylum Tectomicrobia that includes the genus *Candidatus* Entotheonella, one of the most biosynthetically talented sponge symbionts. To investigate the diversity of the candidate phylum Tectomicrobia, phylogenetic analysis was performed using 16S rRNA gene sequences recovered from public databases and generated here. In addition, 16S rRNA

gene amplicons and catalyzed reporter deposition fluorescence in situ hybridization (CARD-FISH) were employed for the identification and description of additional sponge-associated *Ca.* Entotheonella producers of bioactive compounds. Lastly, omics data were used in order to predict functions and estimate the secondary metabolite biosynthetic potential of sponge-associated tectomicrobial representatives.

In **chapter 5**, we resolved the phylogeny and explored the metabolism of sponge-associated members of the candidate phylum Dadabacteria, recently united with the candidate phylum Desulfobacterota. Representatives of this yet-uncultured bacterial lineage have been regularly observed in sponges, but still our knowledge about their role in the sponge microbiome is limited. For this, we performed phylogenomics and metabolic reconstruction to investigate their taxonomy and function. In addition, FISH was applied for description and localization of sponge-associated bacteria belonging to this enigmatic taxonomic group.

In **chapter 6**, the research findings of this thesis are discussed in a broader context compared with the outcomes of previous studies. Moreover, the main conclusions of this research are summarized and future perspectives on how to enhance our current understanding of the host-symbiont interactions are presented.

Chapter 2

Comparative genomic analysis of *Flavobacteriaceae:* insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis

Asimenia Gavriilidou, Johanna Gutleben, Dennis Versluis, Francesca Forgiarini, Mark W.J. van Passel, Colin J. Ingham, Hauke Smidt, Detmer Sipkema

This chapter was published as Gavriilidou, A., *et al.* (2020) *BMC Genomics* 21:569; https://doi.org/10.1186/s12864-020-06971-7

Abstract

Members of the bacterial family *Flavobacteriaceae* are widely distributed in the marine environment and often found associated with algae, fish, detritus or marine invertebrates. Yet, little is known about the characteristics that drive their ubiquity in diverse ecological niches. Here, we provide an overview of functional traits common to taxonomically diverse members of the family *Flavobacteriaceae* from different environmental sources, with a focus on the Marine clade. We include seven newly sequenced marine sponge-derived strains that were also tested for gliding motility and antimicrobial activity. Comparative genomics revealed that genome similarities appeared to be correlated to 16S rRNA gene- and genome-based phylogeny, while differences were mostly associated with nutrient acquisition, such as carbohydrate metabolism and gliding motility. The high frequency and diversity of genes encoding polymer-degrading enzymes, often arranged in polysaccharide utilization loci (PUL), support the capacity of marine *Flavobacteriaceae* to utilize diverse carbon sources. Homologs of gliding proteins were widespread among all studied *Flavobacteriaceae* in contrast to members of other phyla, highlighting the particular presence of this feature within the *Bacteroidetes*. Notably, not all bacteria predicted to glide formed spreading colonies. Genome mining uncovered a diverse secondary metabolite biosynthesis arsenal of *Flavobacteriaceae* with high prevalence of gene clusters encoding pathways for the production of antimicrobial, antioxidant and cytotoxic compounds. Antimicrobial activity tests showed, however, that the phenotype differed from the genome-derived predictions for the seven tested strains. Our study elucidates the functional repertoire of marine *Flavobacteriaceae* and highlights the need to combine genomic and experimental data while using the appropriate stimuli to unlock their uncharted metabolic potential.

Introduction

The family *Flavobacteriaceae* is the largest family of the *Bacteroidetes* phylum, and its members thrive in a wide variety of habitats. These Gram-negative, non-spore forming, rod-shaped, aerobic bacteria are commonly referred to as flavobacteria (Bernardet, 2011; McBride, 2014). The taxonomy of members of the family *Flavobacteriaceae* is considered controversial (Hahnke et al., 2016), with many members that have been renamed and uprooted several times since the first classification (Bernardet and Nakagawa, 2006; McBride, 2014). For the past two decades, the family structure has been relatively stable, even though changes in the taxonomy are still occurring due to the large number of new isolates (McBride, 2014). There has been a vast increase in the number of described genera within the *Flavobacteriaceae* from ten (Jooste and Hugo, 1999) to 158 in the past 20 years. Due to this large number of members, the family has been divided into the following clades: Marine, *Capnocytophaga*, *Flavobacterium*, *Tenacibaculum-Polaribacter* and *Chryseobacterium*, based on 16S ribosomal RNA (rRNA) gene-based phylogenetic analysis (McBride, 2014). Recently the genus *Chryseobacterium* has been reclassified and moved into the family *Weeksellaceae*. *Flavobacteriaceae* are common in terrestrial and freshwater environments and in many cases are numerically predominant in marine habitats (Kirchman, 2002). Within the *Flavobacteriaceae*, bacteria isolated from marine sources are widespread throughout all clades, but a large proportion belongs to the Marine clade (McBride, 2014). To date, marine flavobacteria have been found either free-living or attached to detritus in the water column (DeLong et al., 1993; Bennke et al., 2016). Their lifestyle includes colonization of the surface of algae (Mann et al., 2013), but also close association with invertebrate animals such as sponges (Yoon and Oh, 2012), corals (Sweet et al., 2010) and echinoderms (Romanenko et al., 2007).

A large number of marine flavobacteria degrade high molecular weight macro-molecules, such as complex polysaccharides and proteins, contributing to the carbon turnover in marine environments (Mann et al., 2013; Barbeyron et al., 2016; Bennke et al., 2016). Earlier comparative genomics analyses showed that genomes of marine flavobacteria encode a relatively large number of both glycosyl hydrolases (GHs) and peptidases compared to other marine bacteria. Moreover, a similar number of peptidases was found as compared to other proteolytic specialists, suggesting the important role of flavobacteria in the degradation of both complex carbohydrates and proteins (Fernandez-Gomez et al., 2013). The capacity of flavobacteria to use macromolecules varies considerably, which is reflected by differences in a broad spectrum of enzymes known as carbohydrate-active enzymes (CAZymes) for the whole family. Genes that encode the protein machinery for polysaccharide binding, hydrolysis and transport are often organized in distinct polysaccharide utilization loci (PULs) (Bjursell et al., 2006) that appear unique for the *Bacteroidetes* phylum. The first described PUL was the starch utilization system (Sus) of the human gut symbiont *Bacteroides thetaiotaomicron* (Shipman et al., 2000). PULs in other saccharolytic *Bacteroidetes* encode proteins homologous to SusC and SusD. The SusC-like proteins act as TonB-dependent receptor transporters while the SusD-

like proteins are carbohydrate-binding lipoproteins (Martens et al., 2009). Within PULs, genes encoding these homologs can be found in close proximity to genes coding for CAZymes, such as GHs, polysaccharide lyases (PLs), carbohydrate esterases (CEs) and carbohydrate binding modules (CBMs). Several previous studies have indicated that sequenced genomes of marine *Flavobacteriaceae* feature high proportions of CAZymes and PULs (Fernandez-Gomez et al., 2013; Mann et al., 2013; Xing et al., 2015; Barbeyron et al., 2016; Bennke et al., 2016; Kappelmann et al., 2019), reinforcing their adaptations towards biopolymer degradation.

Besides their specialization as degraders of complex organic matter, many members of the *Bacteroidetes*, and particularly *Flavobacteriaceae*, share another distinct feature, known as 'gliding motility' (McBride and Zhu, 2013). Whilst gliding has been described in bacteria that belong to different taxa, such as *Chloroflexi*, *Cyanobacteria*, *Proteobacteria* and *Bacteroidetes* (Nett and König, 2007; McBride, 2019a), this term has been used loosely and covers multiple molecular mechanisms. *Bacteroidetes* species use their own unique motility machinery that results in rapid gliding movement by pivoting, flipping or crawling over solid surfaces without the aid of flagella or pili (McBride, 2001; McBride and Zhu, 2013). Gliding motility can be observed both microscopically by cells moving on glass slides, as well as on agar by colony spreading (McBride and Zhu, 2013; McBride, 2014). Cell movement is driven by a molecular motor that is composed of several groups of proteins (Gld, Spr and Rem) and powered by the proton motive force (McBride et al., 2009; McBride and Nakane, 2015; Shrivastava et al., 2015). Recent studies revealed that this form of motility involves the rapid movement of adhesins delivered to the cell surface by a novel secretion system, the type 9 secretion system (T9SS), which is highly conserved among *Bacteroidetes* species (Munoz et al., 2020) and unrelated to other known bacterial secretion systems (Sato et al., 2010; McBride and Zhu, 2013; McBride and Nakane, 2015; McBride, 2019a). This system has been extensively studied in the motile aquatic or soil-derived *Flavobacterium johnsoniae* and the non-motile human oral pathogen *Porphyromonas gingivalis* (Lasica et al., 2017). In *F. johnsoniae*, 27 proteins are involved in gliding motility and protein secretion. Orthologs of the *F. johnsoniae* proteins involved in type 9 secretion are also found in *P. gingivalis* (Veith et al., 2017; Johnston et al., 2018). They are common within (McBride and Zhu, 2013), but apparently limited to members of the *Bacteroidetes* (Johnston et al., 2018). Nevertheless, the exact nature of the gliding motility mechanism and its relationship with the T9SS remain under debate.

Gliding motility has previously been linked to the production of secondary metabolites likely due to their common purpose (predation/defense) (Nett and König, 2007). Several antibiotics (β-lactams, quinolones) have been previously isolated from *Flavobacteriaceae* strains (Evans et al., 1978; Cooper et al., 1983; Hida et al., 1985; Kato et al., 1987), with some of them acting against recalcitrant targets such as methicillin-resistant *Staphylococcus aureus* (Funabashi et al., 1993; Katayama et al., 1993). Other bioactive molecules derived from flavobacteria include cell growth-promoting (Imai et al., 1993) and antitumor (Umezawa et al., 1983) compounds, von Willebrand factor receptor antagonists (Kamiyama et al., 1995), protease (Fujita et al., 1994) and topoisomerase (Nemoto et al., 1998) inhibitors, as well as antioxidative and

neuroprotective myxols (Shindo et al., 2007). Even though numerous intriguing molecules are flavobacterial products, the metabolic potential of the family has been poorly investigated. Most of the microbially derived bioactive molecules belong to polyketides, non-ribosomal peptides, saccharides, alkaloids or terpenes and are related through common, highly conserved biosynthetic gene clusters (BGCs) (Milshteyn et al., 2014; Medema and Fischbach, 2015). Computational detection of these BGCs and structural prediction of their products allow microbial genomes to be mined for metabolites (Medema and Fischbach, 2015). Further exploration of the secondary metabolism of marine flavobacteria might therefore unlock a vast resource of novel bioactive compounds.

Here, we performed a comparative genomics analysis to investigate characteristic biological features, such as the complex carbohydrate metabolism and gliding motility mechanism of the *Flavobacteriaceae* family, with a focus on the 'understudied' Marine clade (McBride, 2014). In order to examine the relatedness underlying these properties with other bacterial groups that are abundant and equally important in the marine environment we included genomes of microorganisms belonging to *Cyanobacteria* and *Proteobacteria* phyla. Together with *Bacteroidetes*, they represent the most significant fraction of the global marine bacterioplankton (Kirchman, 2002; Yooseph et al., 2010; Sunagawa et al., 2015b). Moreover, to discern traits that highlight niche-adaptation the same analysis was conducted, comparing host-associated and non-host associated flavobacteria. Of the many recently identified members of the Marine clade of the *Flavobacteriaceae*, only a few have been studied in detail beyond their isolation and initial physiological characterization. In this study, we determined the individual genome sequences of seven presumably novel flavobacteria isolated from marine sponges (Versluis et al., 2017b; Gutleben et al., 2020) and associated their genomic content with phenotypic features in terms of their gliding motility and antimicrobial activity. Finally, in silico genome mining for BGCs was performed to elucidate the secondary metabolic potential of bacteria belonging to dominant marine phyla (*Cyanobacteria* and *Proteobacteria*) and particularly, to members of the *Flavobacteriaceae*.

Methods

Sample Collection and Isolation of Strains

Samples from the sponges *Aplysina aerophoba* and *Dysidea avara* were collected in January 2012 and June 2014, respectively, from Cala Montgó, Spain (42° 06' 52.20" N, 03° 10' 06.52" E) by SCUBA diving at a depth of approximately 12 m (Versluis et al., 2017b; Gutleben et al., 2020). The collection of sponge samples was conducted in strict accordance with Spanish and European regulations within the rules of the Spanish National Research Council with the approval of the Directorate of Research of the Spanish Government. The study was found exempt from ethics approval by the ethics commission of the University of Barcelona since, according to article 3.1 of the European Union directive (2010/63/UE) from the 22/9/2010,

no approval is needed for sponge sacrifice, as they are the most primitive animals and lack a nervous system. Moreover, the collected sponges are not listed in the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). Tissue preparation and cryopreservation were performed as previously described (Sipkema et al., 2011). Cryopreserved samples were stored at -80 °C. Initial cultivation and isolation of the *Flavobacteriaceae* strains from the sponges *Aplysina aerophoba* (DN50, DN105, DN112, Aa_C5, Aa_D4 and Aa_F7) and *Dysidea avara* (Da_B9) were described in Versluis et al. (2017) and Gutleben et al. (2020) (Versluis et al., 2017b; Gutleben et al., 2020) (Table S1).

DNA extraction, Identification and Sequencing

Glycerol stocks of the original strains were initially used as inoculum for regrowth on the original solid isolation media at 20 °C (Table S1). Single colonies were picked and cultured in marine broth 2216 (Difco, Detroit, USA) at 30 °C. Genomic DNA was extracted using the MasterPure DNA Purification Kit (Epicentre, Madison, USA). The quality, purity and concentration of the extracted DNA were estimated by gel electrophoresis, spectrophotometric analysis using a NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and Qubit dsDNA BR Assay kit (Molecular Probes, Life Technologies) used with the DS-11 FX Fluorometer (DeNovix, USA).

To confirm the identity of the strains, 16S ribosomal RNA (rRNA) gene amplicons were generated by PCR using primers 27F (5'-AGAGTTTGGATCMTGGCTCAG-3') and 1492R (5'-CGGTTACCTTGTTACGACTT-3'). The PCR reaction mixture contained 10 µL 5X GoTaq reaction buffer (Promega), 1 µL dNTPs (10 mM), 2.5 µL primer 27F (10 µM), 2.5 µL primer 1492R (10 µM), 0.5 µL GoTaq Polymerase (5 U/µL) (Promega, Madison, WI, USA) and 1 µL of the extracted DNA. Nuclease-free water (Promega, Madison, WI, USA) was added to reach a total reaction volume of 50 µL. The following conditions were used for the bacterial 16S rRNA gene amplification: initial denaturation at 98 °C for 10 min followed by 35 cycles of denaturation at 98 °C for 20 sec, annealing at 52 °C for 20 sec, elongation at 72 °C for 45 sec and a final extension step at 72 °C for 5 min. PCR products were purified using the GeneJET PCR purification kit (Thermo Fisher Scientific, USA) and quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA). The purified PCR products were sent for Sanger sequencing with primers 27F and 1492R (GATC Biotech, Cologne, Germany; now part of Eurofins Genomics Germany GmbH). Trimming (99% good bases, quality value >20, 25-base window) and contig assembly were conducted with DNA Baser 3.5.4.2.

Genome sequencing of strains DN50, DN105 and DN112 was performed using the Illumina MiSeq platform (paired end, 2 × 300 bp reads) (Versluis et al., 2017b). The genomes of strains Da_B9, Aa_C5, Aa_D4 and Aa_F7 were sequenced with Illumina HiSeq (paired end, 2 × 150 bp reads). All genomes were sequenced at GATC Biotech (Konstanz, Germany; now part of Eurofins Genomics Germany GmbH).

Genome Assembly and Quality Control

The quality of the reads was assessed with FASTQC 0.11.4 (Andrews, 2010). Trimmomatic 0.32 was used to remove the Illumina TruSeq adapter sequences and to perform quality filtering (Bolger et al., 2014). A sliding window trimming approach was employed where part of the read in the window (4 bases) was cropped if the average Phred quality in the window was lower than 20. Any raw reads shorter than 20 bases were discarded. Genome sequences generated by the Illumina MiSeq platform were *de novo* assembled with the A5-miseq assembler (version 20160825) (Coil et al., 2015) using default settings. In the case of DN50, the A5-miseq assembler generated a highly fragmented assembly, and the SPAdes 3.11.1 assembler (Bankevich et al., 2012) was used instead. For the HiSeq data, the best k-mer size was automatically selected by KmerGenie 1.6741 (Chikhi and Medvedev, 2014). SPAdes 3.11.1 was employed for the *de novo* assembly of the Illumina HiSeq reads using the selected k-mer. BLASTN (Altschul et al., 1990) with default settings was employed to investigate the assemblies for contamination. All contigs assigned to sequencing artifacts and contamination (e.g. *Enterobacteria* phage phiX174) were discarded prior to downstream analysis. Bowtie2 2.2.5 (Langmead and Salzberg, 2012) was used to map the quality-filtered reads to the assembled contigs resulting in a sequence alignment map (SAM) file. The SAM file was converted into a binary alignment map (BAM) file that was sorted and indexed using SAMtools 0.1.19 (Li et al., 2009). The BAM file was used as input to improve the draft assemblies using Pilon 1.13 (Walker et al., 2014). In addition, coverage per base was calculated using the 'genomecov' command of BedTools 2.17.0 (Quinlan, 2014). The quality of the draft assemblies was evaluated using QUAST 4.6.3 (Gurevich et al., 2013). Completeness and contamination of all analysed genomes were estimated using CheckM 1.07 with the default set of marker genes (Parks et al., 2015).

Genome Annotation and Comparative Genomic Analysis

The draft assemblies of the seven strains sequenced in this study were uploaded to the Integrated Microbial Genomes and Microbiomes (IMG/M version 5.0) system (Chen et al., 2019), and metadata was submitted to the Genomes OnLine Database (GOLD) (Mukherjee et al., 2019). For the comparative analysis, 59 genomes of strains belonging to the phyla *Bacteroidetes* (family *Flavobacteriaceae*), *Cyanobacteria* and *Proteobacteria* (Table S2) were selected that were publicly available at IMG/M (Chen et al., 2019). Protein functional families were automatically assigned via the IMG/M pipeline by comparing predicted proteins to Pfam-A (El-Gebali et al., 2019) Hidden Markov Models using HMMER 3.0b (Finn et al., 2015). Gliding motility and T9SS proteins of the gliding *F. johnsoniae* and non-gliding *P. gingivalis* (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis) were used to identify homologs in the protein sequences of the studied strains by BLASTP searches with an E-value cut-off of 1e-5 using the IMG BLAST Tool. All selected genomes were downloaded via the IMG/M website for further annotation. Predictions of the protein sequences were obtained using Prokka 1.13 (Seemann, 2014). CAZymes were annotated based on HMMER

searches (HMMER 3.0b) (Finn et al., 2015) against the dbCAN database release 6.0 (Lombard et al., 2014). Annotation of PULs was performed using the PULPy pipeline (Stewart et al., 2018). PULs which contained one *susC/susD* gene pair and at least one adjacent gene coding for CAZymes were assigned as "complete". The online server antiSMASH 5.0 (Blin et al., 2019) was used for the identification of secondary metabolite BGCs with "relaxed" detection strictness. The ClusterBlast and KnownClusterBlast modules integrated into antiSMASH 5.0 (Blin et al., 2019) were also used for comparative gene cluster analysis based on the NCBI GenBank (Benson et al., 2013) and the 'Minimum Information about a Biosynthetic Gene Cluster' (MIBiG) (Medema et al., 2015) data standard, respectively.

Phylogenetic Analysis and Data Selection

Taxonomic assignment of the seven newly sequenced isolates was performed by: 1) BLASTN (May 2019) (Altschul et al., 1990) of near full length 16S rRNA gene sequences recovered from this study against the nr/nt NCBI database and 2) single-copy marker gene analysis and placement of genomes in the Genome Taxonomy Database (GTDB) reference tree (Parks et al., 2018) using the GTDB-Tool Kit 1.1.0 (GTDB-Tk) *classify* workflow (Chaumeil et al., 2019). For the reconstruction of the phylogeny, the closest relative of each isolate in NCBI was selected based on the availability of genomes in IMG/M (Chen et al., 2019). Similarly, representatives of the other clades of the family *Flavobacteriaceae* were chosen according to the genome availability in IMG/M (Chen et al., 2019), but also the phylogenetic position of the isolates (based on their 16S rRNA gene sequences) in ARB (Ludwig et al., 2004) using the SILVA SSU Ref NR 99 database (release 132) (Quast et al., 2013) as reference. To compare the functional traits of the *Flavobacteriaceae* members, representatives of two other phyla that are dominant in the marine environment (*Cyanobacteria* and *Proteobacteria*) were included in the analysis and were also used as outgroup. Multiple alignments of the 16S rRNA gene sequences were performed using the SINA Aligner 1.2.11 (Pruesse et al., 2012). A maximum likelihood tree was generated in ARB (Westram et al., 2011) employing RAxML 7.0.3 (Stamatakis et al., 2008) with 1,000 iterations of rapid bootstrapping. Phylogenomic analysis was performed using GTDB-Tk 1.1.0 (Chaumeil et al., 2019). A concatenated amino acid-based phylogeny was reconstructed using the translated amino acid sequences of 120 bacterial marker genes identified in the studied genomes and aligned by GTDB-Tk *identify* and *align* module, respectively. The resulting multiple sequence alignment was used for generating a maximum likelihood protein tree using FastTree 2.1.11 with default parameters (Price et al., 2010). Visualization of both maximum likelihood trees was performed using the Interactive Tree of Life (iTOL) version 3 (Letunic and Bork, 2016).

Phenotypic Assays

Gliding Motility Tests

To determine colony spreading, two different types of marine agar plates were prepared using marine broth 2216 (Difco, Detroit, USA) solidified with either 1% (w/v) or 1.8% (w/v)

of Noble Agar (Sigma-Aldrich, MA, USA). The cells were first grown in marine broth at 30 °C until they reached mid-exponential phase. Subsequently, a 5 µL sample of the cell suspension was spotted at the centre of each test plate using an Eppendorf pipette. After observing growth, the edges of the colonies were viewed with a Zeiss Axio Scope.A1 phase-contrast microscope at a magnification of 10X. Pictures of the colony edges were taken using a Zeiss AxioCam ICc3 attached to the phase-contrast microscope.

Antimicrobial Activity Tests

Disc diffusion assays were carried out according to the Kirby-Bauer susceptibility method (Bauer et al., 1966). To determine the antimicrobial activity of the seven isolates (Table S1), a number of indicator strains were used (Table S3). Prior to the screening, broth cultures of the indicator strains were prepared using 200 µL of the stock cultures to inoculate 3 mL of the respective media and incubated overnight at the corresponding temperatures (Table S3). Subsequently, 200 µL of the cultures (adjusted to OD_{600} 0.5) were uniformly spread on agar plates using a sterile L-shape spreader. The tested isolates were cultured in marine broth 2216 (Difco, Detroit, USA) at 30 °C until they reached stationary phase. After harvesting, the cultures were centrifuged at 1,657 x g for 20 min at room temperature. The supernatant was sterile-filtered using a 0.2 µm syringe filter to remove the bacterial cells. For screening, sterile, 6-mm diameter filter paper discs were impregnated with 60 µL of the cell-free supernatant and air-dried for 1 h. The paper discs were then transferred onto the agar plates covered with a lawn of the indicator strains. As positive controls, antibiotics known to be effective against the indicator strains were used (Table S3). The uninoculated growth medium of each of the tested strains was used as negative control, after receiving the same treatment as described for the cultures of the isolates. The plates were incubated for 24 to 48 h at different temperatures depending on the indicator strains used for the assay. After the incubation, the plates were examined for the presence of inhibition zones that were measured using a calliper. All assays were performed in triplicate.

Statistical analysis

Data were analysed and visualized in R 3.5.0 using vegan 2.5-2 (Oksanen et al., 2018), phyloseq 1.26.1 (McMurdie and Holmes, 2013), ggplot2 3.1.1 (Wickham, 2016), VennDiagram 1.6.20 (Chen, 2018), and ComplexHeatmap (Gu et al., 2016). Functional comparisons between genomes were performed using Pfam annotations as input data. A Bray-Curtis dissimilarity distance matrix was calculated based on the relative abundance (Gene counts/Total genes) of Pfam profiles with the 'vegdist' function in the vegan R package. Variation in the functional profiles was assessed by non-metric multidimensional scaling (NMDS) ordination using the Bray-Curtis dissimilarity matrix with the 'ordinate' function in the phyloseq R package. NMDS plots were created using the 'plot_ordination' function of the ggplot2 R package. The significance of the differences in functional profiles across major taxonomic groups and within the *Flavobacteriaceae* was tested by non-parametric permutational

analysis of variance (PERMANOVA) on the Bray-Curtis dissimilarity matrix using the 'adonis' function in the vegan R package, with the number of permutations set at 999. To rank the Pfams contributing the most to the differentiation of the functional profiles between genomes, Similarity Percentage (SIMPER) analysis was employed with the 'simper' function of the vegan R package. For the pairwise comparisons, only Pfam entries with the highest significant contribution to the dissimilarity are shown (>0.2%, p <0.05).

Results

Genome properties and phylogeny

To determine the genomic characteristics of the seven *Flavobacteriaceae* isolates recently obtained from the sponges *Aplysina aerophoba* and *Dysidea avara* and to allow for comparison with publicly available genomes from other members of the *Flavobacteriaceae*, their genomes were elucidated using Illumina sequencing. Genome sizes of the seven strains ranged from 3.7 to 5.3 Mbp with an average GC content of 38% (Table 1). The assembly generated genomes with 13-83 contigs and N50 ranging from 0.1 to 1.3 Mbp. Estimated completeness was more than 98%, and redundancy was lower than 2% for all assembled genomes. Coverage per base was above 200x for all draft genomes, except for DN50 for which it was 77x. Total gene count varied between 3317 and 4738, while the percentage of coding sequences (CDS) in all the assembled genomes was more than 98%. On average, more than 75% of the total genes could be assigned to protein families (Pfams).

Table 1. Genome properties and quality metrics of the strains sequenced in this study.

	Aa_C5	Aa_D4	Aa_F7	Da_B9	DN50	DN105	DN112
Size (Mbp)	3.7	3.7	4.2	4.3	5.3	4.8	4.7
Contigs	45	13	43	19	38	83	19
%GC	37	37	41	45	39	36	31
N50 (Mbp)	0.8	0.7	0.6	0.6	0.3	0.1	1.3
Per base coverage (x)	1331	357	298	248	77	207	261
Completeness (%)	98.7	98.7	99.3	99.7	99.7	99.7	99.3
Contamination	0.1	0.1	0.2	0.7	1.8	0.6	1.0
Total gene count	3317	3306	3776	4079	4738	4331	4069
CDS genes (%)	98.7	98.7	98.8	98.7	98.7	98.8	98.0
Number of 16S rRNA genes	1	1	2	1	3	1	1
Number of tRNA genes	36	36	37	45	50	40	63
Genes in Pfams (%)	80.9	80.9	77.8	74.8	70.2	73.2	77.8
Isolation source	*Aplysina aerophoba*	*Aplysina aerophoba*	*Aplysina aerophoba*	*Dysidea avara*	*Aplysina aerophoba*	*Aplysina aerophoba*	*Aplysina aerophoba*

Phylogenetic analyses based on 16S rRNA gene as well as concatenated protein sequences both clustered all seven newly sequenced strains within the Marine clade of the family *Flavobacteriaceae* (Figure 1 and S1).

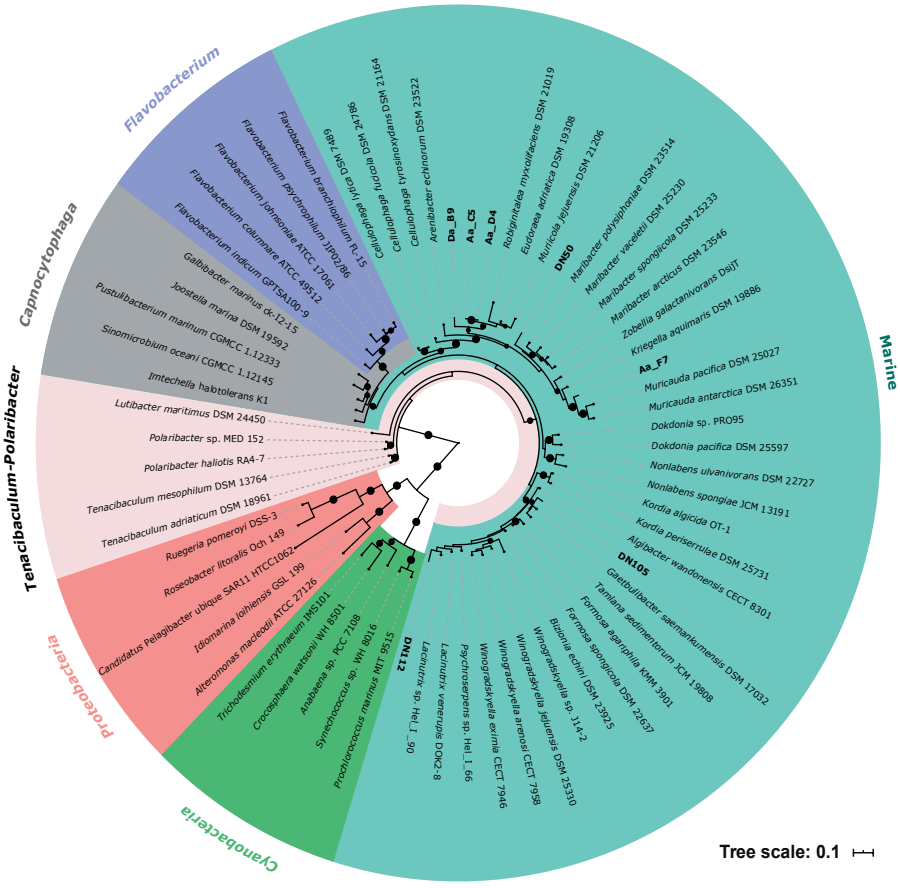


Figure 1. 16S rRNA gene phylogenetic tree of microorganisms whose genomes were compared in this study. Phylogeny was inferred using a maximum likelihood method. Bootstrap values (> 70%) are displayed as black circles on the middle of the branches. Color annotations represent the different *Flavobacteriaceae* clades. Sequences belonging to *Cyanobacteria* and *Proteobacteria* were used as outgroups. Names in bold indicate sequences generated in the current study. The scale bar indicates 0.1 substitutions per site.

All strains isolated in this study formed distinct branches in both phylogenetic trees, except for Aa_D4 and Aa_C5, the 16S rRNA gene sequences of which were 99.6% identical. Similarly, a two-way Amino Acid Identity (AAI) comparison of Aa_D4 and Aa_C5 resulted in 99.8% AAI. Among the newly sequenced isolates, two subclades within the Marine clade were represented, where DN112 and DN105 were phylogenetically adjacent, but distant from the rest of the isolates that grouped together (Figure 1 and S1). A BLASTN search against the nr/

nt NCBI database showed that both Aa_C5 and Aa_D4 were members of the genus *Eudoraea* (Table S4). BLASTN returned *Flagellimonas* sp. as the closest hit to Aa_F7 and *Lacinutrix* sp. to DN112. The closest related sequence to DN50 belonged to an uncultured organism clone, while Da_B9 and DN105 were most closely related to a *Flavobacteriaceae* bacterium, isolated from seamounts and a marine sponge, respectively (Table S4). Taxonomic assignment of the newly sequenced genomes based on the Genome Taxonomy Database (GTDB) showed similar results with the 16S rRNA gene-based taxonomy. According to GTDB taxonomy, none of the strains could be classified to species level. Four of them were classified to genus level and the rest to family level (Table S4).

Functional profiling

Pfam profiles of 66 genomes representing three taxonomic groups (the family *Flavobacteriaceae*, and the phyla *Cyanobacteria* and *Proteobacteria*), all four clades of the *Flavobacteriaceae* (Marine, *Capnocytophaga*, *Tenacibaculum-Polaribacter*, *Flavobacterium*) and two different types of isolation sources (host-associated and non-host associated) were used as input to assess the correlation of taxonomy and life strategy with genome-derived functional traits. *Flavobacteriaceae*, *Cyanobacteria* and *Proteobacteria* had highly divergent functional profiles based on the relative abundances of Pfam annotations (PERMANOVA, p = 0.001) (Figure 2a). *Flavobacteriaceae* and *Cyanobacteria* showed higher overall dissimilarity (57.4%) at the functional level, compared to *Flavobacteriaceae* and *Proteobacteria* (50%). Similarly, functional profiles were significantly different between the four different flavobacterial clades (PERMANOVA, p = 0.001) (Figure 2b). In contrast, no significant differences were observed between host-associated and non-host associated flavobacteria at the functional level (PERMANOVA, p > 0.05) (data not shown).

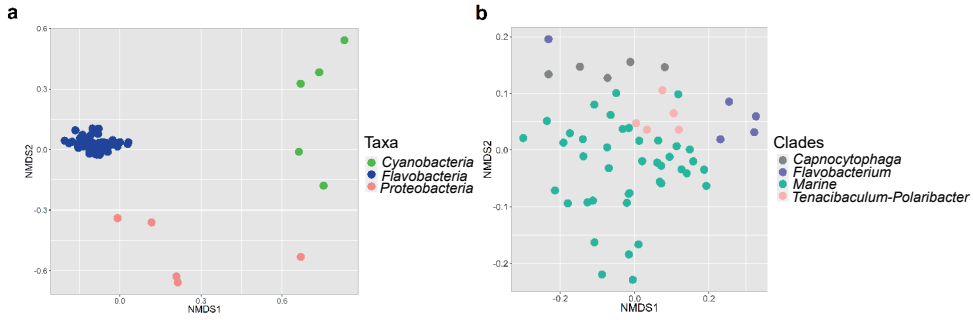


Figure 2. NMDS clustering of genomes based on similarity of functional groups according to Pfam annotations. Genomes are plotted following Bray-Curtis dissimilarity values calculated from Pfam abundance profiles between major taxonomic groups (a) and different *Flavobacteriaceae* clades (b).

From the 5,173 Pfams detected, 1,648 were present in all three groups (*Flavobacteriaceae*, *Cyanobacteria* and *Proteobacteria*), whereas 1,132 were specific to *Flavobacteriaceae* (Figure 3a). The total number of predicted Pfams in the included genomes of the *Flavobacteriaceae* was 3,843 with 17% of them being specific to the Marine clade. Overall, a high degree of functional conservation, was observed at the Pfam-level among the different clades of the *Flavobacteriaceae*, with 48% of all annotated Pfams being shared by the four clades (Figure 3b). All predicted Pfams that significantly contributed most to the dissimilarity between *Flavobacteriaceae*, *Cyanobacteria* and *Proteobacteria* (SIMPER analysis, > 0.2% contribution, $p < 0.05$) showed higher abundances in *Flavobacteriaceae*. Particularly, most functional attributes strongly selected for in flavobacteria were related to proteins involved in carbohydrate metabolism and transport (Table 2). These include a series of protein domains found in TonB-dependent receptors (pfam13715, pfam07715, pfam00593) and two-component regulatory systems (pfam00072, pfam08281, pfam04542, pfam04397).

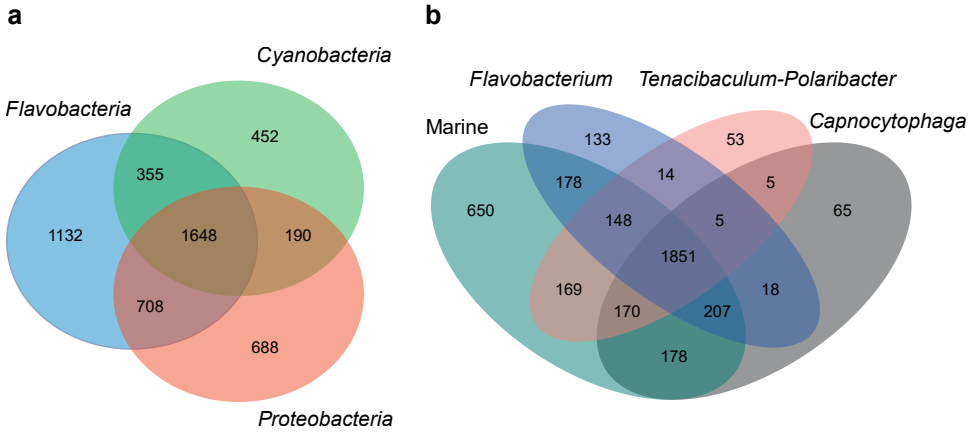


Figure 3. Shared and unique protein families (Pfams) between the analysed genomes. Venn diagrams illustrate shared and unique Pfams amongst *Flavobacteriaceae*, *Cyanobacteria* and *Proteobacteria* (a) and within the *Flavobacteriaceae* (b).

The majority of the genes coding for TonB-dependent receptors were found adjacent to genes that code for SusD/RagB family proteins (pfam07980), which are restricted to the phylum *Bacteroidetes*. Moreover, among the most differentiating Pfam entries was the CHU_C family (pfam13585) that was found only in *Flavobacteriaceae*. This family has been reported as essential for the localization of a cellulase on the cell surface in *Cytophaga hutchinsonii* and the function of this cellulase in crystalline cellulose degradation (Wang et al., 2017). It showed high similarity to the gliding motility-associated C-terminal domain (CTD) (TIGR04131) that is unique to and highly prevalent in the phylum *Bacteroidetes* (Veith et al., 2017). This CTD (type B CTD) has been recently shown to target proteins for secretion by the T9SS (Kulkarni et al., 2019). Similarly, within the *Flavobacteriaceae*, the functional differences between the Marine, the *Capnocytophaga* and the *Flavobacterium* clade were

mainly due to contribution of Pfam entries related to the attachment and degradation of polymeric compounds (Table S5). There were no obvious functions distinguishing the Marine and the *Tenacibaculum-Polaribacter* clades, except for the C-terminal domain of the CHU protein family (pfam13585) that was significantly more abundant in the Marine clade members.

Table 2. Pfams contributing most significantly (> 0.2%, p < 0.05) to differences across major taxonomic groups.

Function name (Pfam ID)	Relative Abundance (%)			Contribution (%)		p-values (< 0.05)	
	F	C	P	F-C	F-P	F-C	F-P
CarboxypepD_reg-like domain (pfam13715)	1.23	0	0.004	0.97	0.97	0.001	0.001
TonB-dependent Receptor Plug Domain (pfam07715)	1.13	0.01	0.44	0.77	0.76	0.001	0.016
TonB dependent receptor (pfam00593)	0.55	0.01	0.29	0.47	0.42	0.001	0.04
Response regulator receiver domain (pfam00072)	1.09	0.85	0.95	0.41	0.49	0.014	0.015
Sigma-70, region 4 (pfam08281)	0.51	0.04	0.11	0.41	0.37	0.001	0.002
LytTr DNA-binding domain (pfam04397)	0.40	0	0.06	0.35	0.34	0.005	0.029
Sigma-70 region 2 (pfam04542)	0.63	0.24	0.21	0.33	0.40	0.004	0.001
C-terminal domain of CHU protein family (pfam13585)	0.38	0	0	0.31	0.35	0.001	0.001

F, *Flavobacteriaceae*; C, *Cyanobacteria*; P, *Proteobacteria*.

Carbohydrate metabolism and transport

The ability to degrade and transform complex carbohydrates was assessed by mining the genomic data for carbohydrate-active enzymes (CAZymes) and PULs. In total, *Flavobacteriaceae* harboured more CAZymes per Mbp (Kruskal-Wallis test, p < 0.001) compared to *Cyanobacteria* and *Proteobacteria* (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis). Comparison of the CAZyme repertoire (average number of CAZymes per Mbp) across the different groups showed significantly higher frequencies of GHs, PLs, CEs and CBMs in *Flavobacteriaceae* (Kruskal-Wallis test, p < 0.001), whereas glycosyl transferases (GTs) were more frequent in *Cyanobacteria* (Kruskal-Wallis test, p < 0.05) (Figure 4a). No statistically significant differences were observed in the frequency of the different CAZyme classes between the different *Flavobacteriaceae* clades (Figure 4b and Additional File 1). In contrast, the number of degradative CAZymes arranged in PULs per Mbp (Kruskal-Wallis test, p < 0.05) and putative PULs per Mbp (Kruskal-Wallis test, p < 0.01) revealed a significant variation in the potential to degrade polysaccharides between the *Flavobacteriaceae* clades (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis).

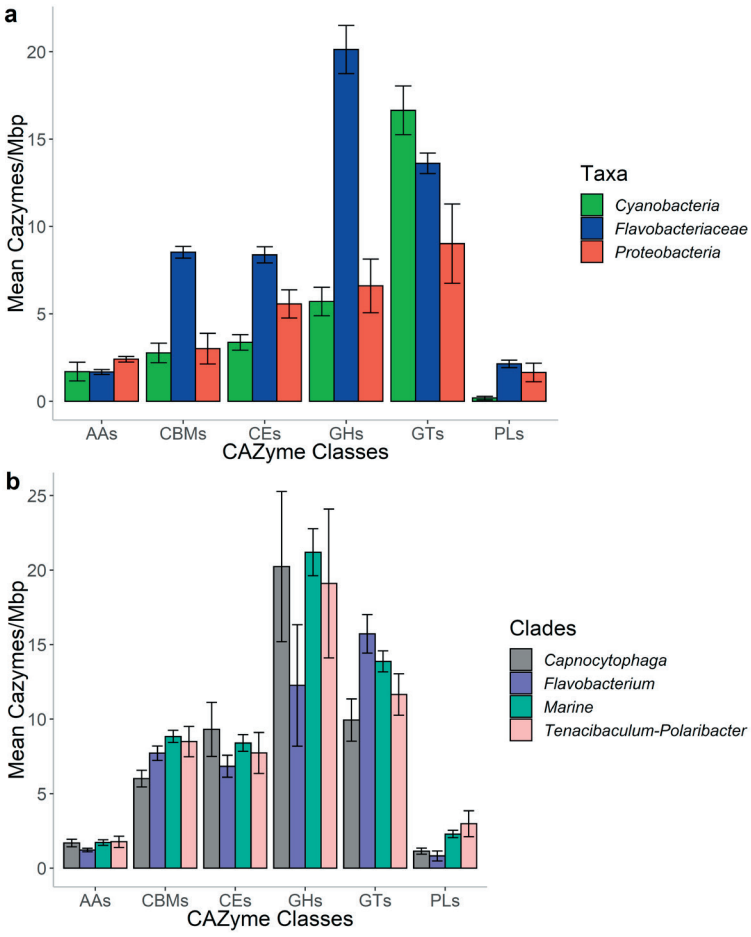


Figure 4. Mean number of CAZymes per Mb in the analyzed genomes. Frequency of CAZymes normalized by total gene counts is shown between different taxonomic groups (a) and *Flavobacteriaceae* clades (b). Error bars represent mean numbers ± standard error. GHs, glycoside hydrolases; PLs, polysaccharide lyases; CEs, carbohydrate esterases; GTs, glycosyl transferases; CBMs, carbohydrate-binding modules.

Strains of the Marine clade possessed an average of 42.4 degradative CAZymes per Mbp of which 14.6% were PUL-associated. Moreover, 3.1 PULs per Mbp were identified in the Marine strains with 5.2% being "complete" (Table 3). The rest of the annotated PULs included only polysaccharide binding proteins (*susC/D* genes), but no CAZyme-encoding genes. Within *Flavobacteriaceae*, *Capnocytophaga* genomes encoded 2.5 times more PUL-associated CAZymes (Kruskal-Wallis test, $p < 0.05$) than genomes belonging to the Marine clade. Similarly, strains of the *Capnocytophaga* clade showed on average the highest frequency of PULs (8.4 PULs per Mbp) while Marine, *Tenacibaculum-Polaribacter* and *Flavobacterium* genomes followed with 3.1, 2.9 and 1.8 PULs per Mbp, respectively (Table 3). In the Marine clade, among the most abundant GH families were GH family 74 (GH74) and GH109 (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis).

GH74 comprises many xyloglucan-hydrolyzing enzymes, while enzymes belonging to GH109 have α-N-acetylgalactosaminidase activity (Carbohydrate Active Enzymes database; http://www.cazy.org/). In the case of CBMs in the Marine clade, most hits belonged to the chitin- or peptidoglycan-binding family (CBM50) and cellulose- and xyloglucan-binding family (CBM44) (Carbohydrate Active Enzymes database; http://www.cazy.org/

Table 3. Comparison of frequencies of degradative CAZymes, PULs and CAZymes in PULs in *Flavobacteriaceae* genomes.

	Flavobacteriaceae			
	M	**C**	**F**	**T**
CAZymes/Mbp	42.4	43.7	28.9	40.1
Total PULs/Mbp	3.1	8.4	1.8	2.9
%Complete PULs/Mbp	5.2	4.7	7.3	5.0
CAZymes in PULs/Mbp	6.2	15.5	4.9	8.5
%CAZymes in PULs/Mbp	14.6	35.1	11.0	19.3

A more detailed table of the PUL-associated CAZymes distribution can be found in Additional File 2 (https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis). M, Marine; C, *Capnocytophaga*; F, *Flavobacterium*; T, *Tenacibaculum-Polaribacter*.

Gliding motility and Type 9 secretion

Based on the studies of (Veith et al., 2017) and (McBride, 2019b), we defined a set of 27 proteins that are related to gliding motility and T9SS, e.g. as reported for *F. johnsoniae* and *P. gingivalis*. These proteins were categorized as i) essential, ii) important, when their absence had a minimal effect but not complete loss of motility and secretion, and iii) non-essential, according to previous reports on *F. johnsoniae* (Veith et al., 2017; McBride, 2019a) (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis). Forty-nine publicly available genomes (Table S2) of gliding and non-gliding members of the different clades of the *Flavobacteriaceae* (Marine, *Capnocytophaga*, *Flavobacterium*, *Tenacibaculum-Polaribacter*) (Figure 5) isolated from different environments (host-associated and non-host associated) were screened for the presence of homologs to these 27 proteins, using amino acid sequences.

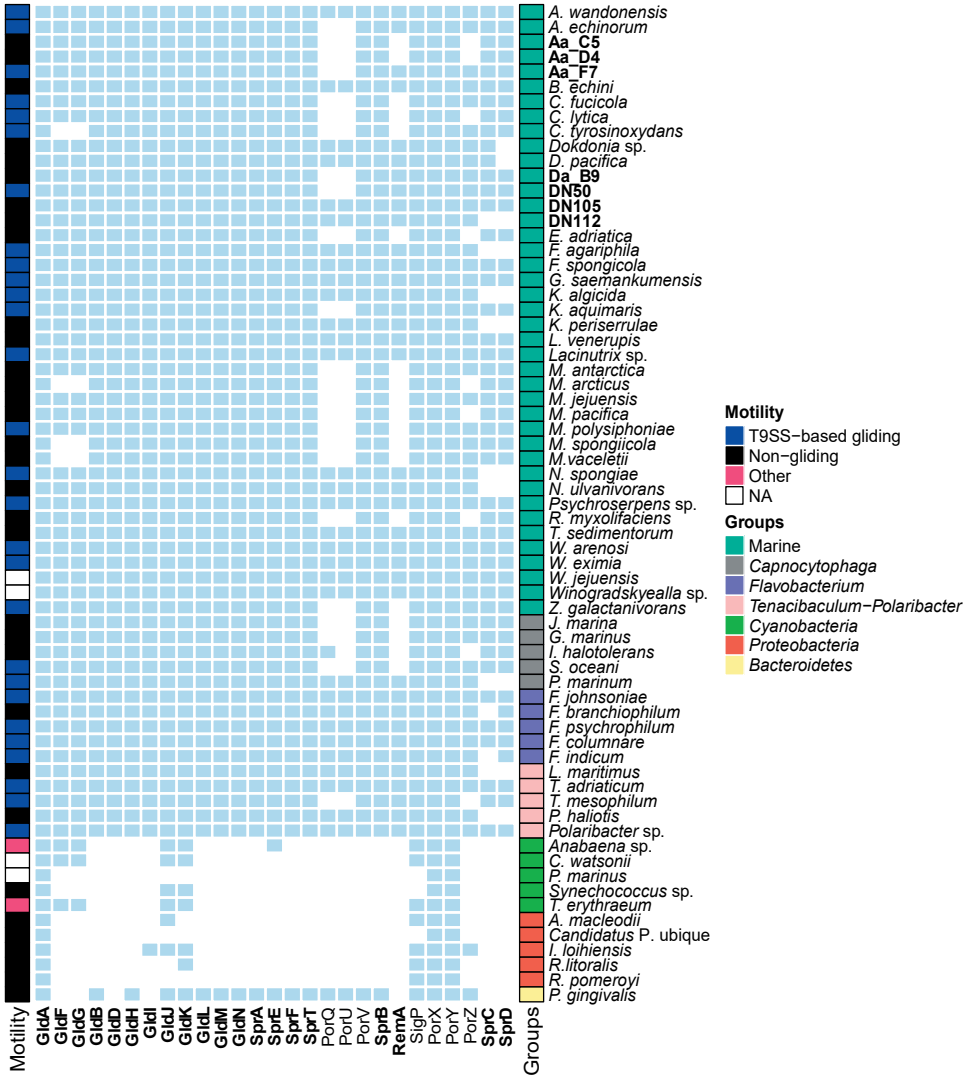


Figure 5. Presence of homologs of gliding motility and T9SS proteins in the predicted proteomes. Presence of homologs is indicated by light blue squares and absence by white squares. Strains sequenced in the framework of this study are indicated in bold. Each species was annotated based on its taxonomy (Groups) and motility as reported in literature. Protein names in bold are considered essential or important for *F. johnsoniae* gliding. The rest of the proteins are either non-essential or their role in *F. johnsoniae* gliding has not been determined (Veith et al., 2017; McBride, 2019a). Other, gliding motility mechanism other than T9SS-based;NA, no information available (see also Additional File 1).

All 29 flavobacteria reported in the literature as 'gliding' had homologs for 20 of the motility and T9SS-related proteins (GldA, GldB, GldF, GldG, GldD, GldH, GldI, GldJ, GldK, GldL, GldM, GldN, SprA, SprE, SprF, SprT, PorV, SigP, PorX and PorY), except for *Cellulophaga tyrosinoxydans* which was lacking GldF and GldG homologs (Figure 5). The 20 proteins encoded by the genomes of all other gliding flavobacteria include all the Gld and Spr proteins that were

considered essential or important for motility in *F. johnsoniae* (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis) (McBride and Zhu, 2013). Homologs to the motility adhesins RemA were not detected in five out of the 29 genomes of gliding members of the *Bacteroidetes* phylum, while SprB homologs were present in all gliding bacteria. In our dataset, there were two strains for which no information regarding motility was available in literature (*Winogradskyella jejuensis* CP32^T and *Winogradskyella* sp. J14-2). Here, we report that both strains have a complete set of homologs to the T9SS-based gliding proteins. For the non-gliding *P. gingivalis*, homologs for the major T9SS components (GldK, GldL, GldM, GldN, SprA, SprE, SprT) were detected in the genome (Figure 5). Four proteins (GldF, GldG, GldD, GldI) that are required for gliding motility of *F. johnsoniae* did not appear to be encoded in the genome of *P. gingivalis*. The same was the case for the motility adhesin RemA and the SprB supporting proteins, SprC and SprD. No obvious differences in the presence of gliding and T9SS proteins were found among genomes of host-associated and non-host-associated strains (data not shown). Similarly, no specific pattern was observed in terms of the gliding motility of the different *Flavobacteriaceae* clades (Figure 5).

Apart from the *Bacteroidetes*, proteins encoded by genomes from two other phyla (*Cyanobacteria* and *Proteobacteria*) were searched by BLASTP (E-value 1e-5) for the presence of homologs of gliding and T9SS-related proteins (Figure 5). In general, homologs to *F. johnsoniae* motility proteins were scarce among these two phyla. From the gliding motility proteins, only GldA homologs were present in all cyanobacterial and proteobacterial genomes. However, GldA is the ATP-binding component of the ABC transporter of the gliding motility apparatus and might share sequence similarity with many other (not necessarily gliding-related) ATP-binding regions of ABC transporters (McBride and Zhu, 2013). *Anabaena* sp. and *Trichodesmium erythraeum* strains were the only cyanobacteria in our dataset previously reported as gliding (Hoiczyk, 2000; Lee et al., 2017), and both genomes were found to encode homologs for five gliding proteins (GldA, GldF, GldG, GldJ, GldK) and three regulatory proteins (SigP, PorX and PorY). The *Anabaena* sp. genome also encoded one Spr ortholog (SprE) that was absent from *T. erythraeum*. Among the genomes tested in this study, those of *Prochlorococcus marinus* and *Candidatus* Pelagibacter ubique encoded the lowest number (three) of homologs to the motility and T9SS proteins (Figure 5). In addition to examining the genomes of the seven flavobacterial sponge isolates, their gliding motility was also tested based on the formation of spreading colonies on agar. All newly sequenced genomes encode homologs for each of the proteins considered essential for gliding motility and type 9 secretion in *F. johnsoniae* (Additional File 1; https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis). However, after examination of the colony morphology, most of the isolated strains (DN105, Da_B9, Aa_C5, Aa_D4, DN112) were non-spreading on agar, and only Aa_F7 and DN50 formed spreading colonies. The edges of the colonies of both Aa_F7 and DN50 formed slender peninsulas spreading on 1% (w/v) agar plates (Figure 6a and b). On 1.8% (w/v) agar plates, DN50 exhibited a colony surface contour pattern probably attributed to gliding motility and secondary colony formation (Figure 6c). This pattern was not observed for any of the other strains.

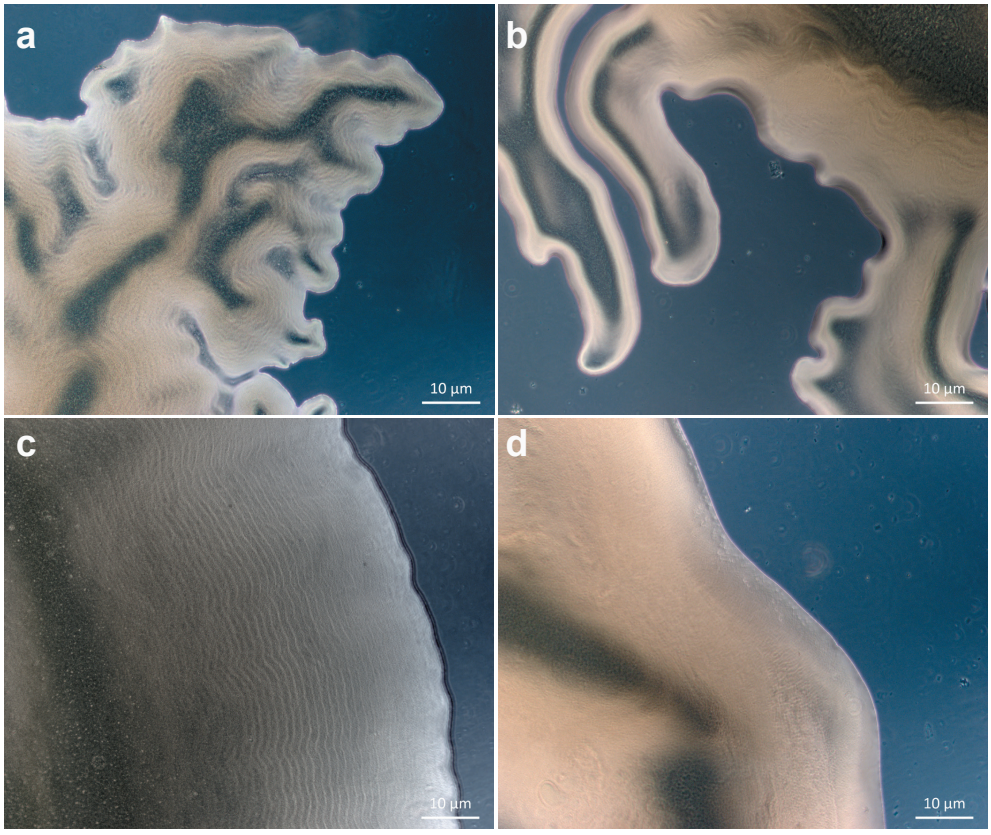


Figure 6. Colony morphologies of the studied flavobacterial strains. a and b) Spreading colonies of strains DN50 and Aa_F7, respectively, that form slender peninsulas on 1% (w/v) agar; c) Surface contour pattern of DN50 colonies on 1.8% (w/v) agar; d) Non-spreading colonies of DN105. Photos were taken with an AxioCam ICc3 attached to a Zeiss Axio Scope.A1 phase-contrast microscope.

Homologs for all gliding and T9SS proteins were identified in the genomes of the spreading strains (Aa_F7 and DN50), except for PorQ and PorU. Consequently, the chromosomal neighbourhood of these genes was investigated for the genomes of Aa_F7 and DN50 and compared to that of *F. johnsoniae* UW101 (Figure 7). In the *F. johnsoniae* genome, *porU* (Fjoh_1556) lies immediately upstream of *porV* (Fjoh_1555) and adjacent to *gldJ* (Fjoh_1557), being encoded on the opposite strand. For Aa_F7 and DN50, the gene arrangement was similar to that found in *F. johnsoniae* with a gene coding for a ligase downstream of *gldJ*. No homolog to *porU* was identified in the genomes of Aa_F7 and DN50, and the genes directly adjacent to *gldJ* were annotated as hypothetical proteins (Figure 7).

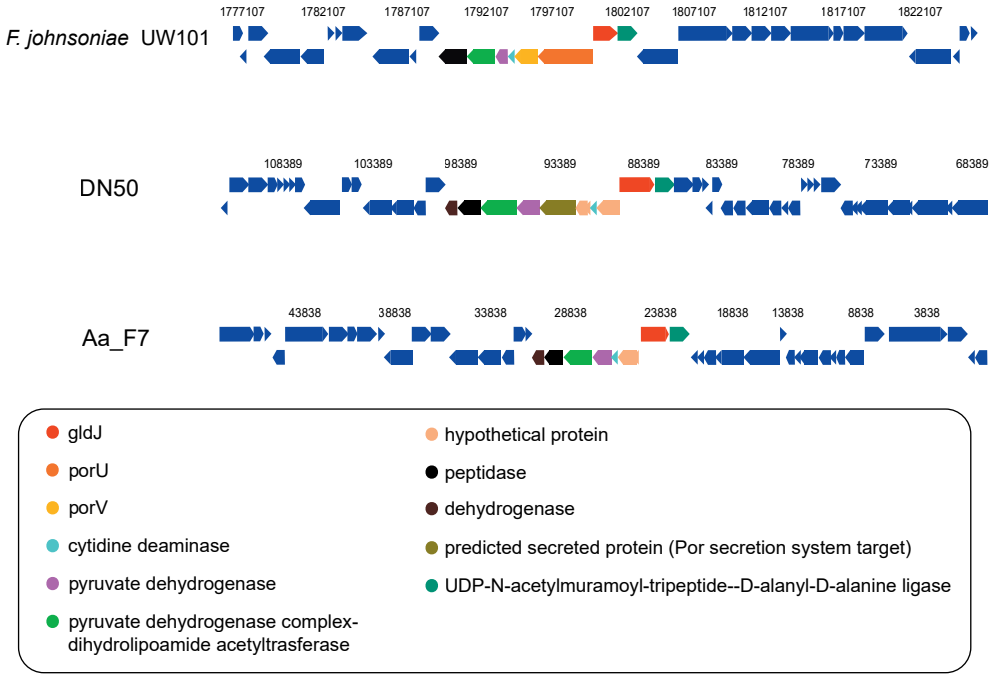


Figure 7. Gene neighbourhoods of *gldJ* in Aa_F7 and DN50 compared to *F. johnsoniae* UW101 genome. Genes of the same colour are from the same orthologous group. Light yellow genes had no Cluster of Orthologous Group (COG) assignment. (Chromosome viewer-IMG/MER)

Sequence alignments of these hypothetical proteins and *F. johnsoniae* PorU by BLASTP showed amino acid identities of 24.4% and 26.3% and a query cover of 8% and 14% for Aa_F7 and DN50, respectively. In the case of *porQ*, it was not possible to use the gene neighbourhood because it is not adjacent to any of the known gliding or T9SS-related genes in *F. johnsoniae*. Instead, a BLASTP search with a lower cut-off (E-value 1e-0) revealed weak hits for PorQ in both Aa_F7 and DN50. Interestingly, in the genome of DN105 all components of *F. johnsoniae* gliding and T9SS machinery were identified, even though it did not form spreading colonies on agar under the examined conditions (Figure 6d). Similarly, *Bizionia echini*, *Lacinutrix venerupis* and *Tamlana sedimentorum* were previously reported as non-motile, but they carry a complete set of orthologous gliding genes in their genomes (Figure 5).

Secondary metabolome

The genomic repertoire for secondary metabolite synthesis of the 66 studied bacteria was assessed using the 'Antibiotics and secondary metabolite analysis shell' (antiSMASH) (Blin et al., 2019). In *Flavobacteriaceae*, the vast majority of BGCs was predicted to encode terpene synthases (44%), followed by type III polyketide synthases (T3PKS) (9%), non-ribosomal peptide synthetases (NRPS) (7%), lanthipeptide (6%) and aryl polyene (6%) clusters. In total, the percentage of genes putatively involved in the biosynthesis of secondary metabolites in

relation to the total number of protein-coding genes in each genome ranged between 0.02 and 0.3%. *Cyanobacteria* showed the highest average number of annotated BGCs (*n* = 7), compared to *Flavobacteriaceae* (*n* = 4) and *Proteobacteria* (*n* = 4) (Table S6). Gene clusters encoding the proteins involved in the biosynthesis of non-ribosomal peptides and polyketides were identified in genomes from all three major groups (Table S6). Terpene BGCs were detected in all analysed genomes and showed the highest relative abundance in *Flavobacteriaceae* and *Cyanobacteria*. In contrast, the most abundant gene cluster in Proteobacteria was affiliated with the biosynthesis of homoserine lactones (HSL), a quorum sensing (QS) molecule in Gram-negative bacteria (Table 4). Interestingly, only one flavobacterial strain (*Muriicola jejuensis* DSM 21206) harboured HSL BGCs.

Table 4. Distribution of BGCs (mean number of BGCs per strain) across major taxonomic groups and different *Flavobacteriaceae* clades.

Gene cluster type (antiSMASH)	Flavobacteriaceae				Cyanobacteria	Proteobacteria
	M	C	F	T		
Terpene	1.63	1.20	1.40	1.20	1.60	0.40
T3PKS	0.39	0	0.20	0.20	0.20	0
NRPS	0.29	0	0.20	0	1.00	0
Lanthipeptide	0.22	0.40	0	0.20	0	0
Aryl polyene	0.17	0	0.60	0.20	0.60	0.20
Aryl polyene-Resorcinol	0.17	0	0.60	0	0	0
Bacteriocin	0.12	0.20	0.20	0	1.40	0.80
NRPS-like	0.05	0.40	0	0	0.60	0
Siderophore	0	0.20	0.60	0.20	0	0.20
Homoserine lactone	0.02	0	0	0	0	1.40
Betalactone	0	0	0.20	0	0	0.40

Only BGC types with mean abundance of more than 0.1 BGCs/strain in the complete dataset are shown. A more detailed table of the BGC distribution can be found in Table S7. M, Marine; C, *Capnocytophaga*; F, *Flavobacterium*; T, *Tenacibaculum-Polaribacter*.

Within the *Flavobacteriaceae*, strains from the Marine and *Flavobacterium* clades contained the largest average number of BGCs (Table S6), whereas on average only 2 BGCs were detected in the *Tenacibaculum-Polaribacter* genomes. In general, the different flavobacterial clades exhibited a similar composition in terms of the types of predicted BGCs being enriched in their genomes, including terpene, lanthipeptide, NRPS and aryl polyene BGCs.

Genome mining for BGCs of the seven newly isolated flavobacteria resulted in the detection of five BGCs, on average, belonging to six different classes based on the antiSMASH database. In total, 63% of the identified BGC encoded proteins were predicted to be involved in pathways for the production of terpenes, lanthipeptides and aryl polyenes. Less abundant but widely distributed across these genomes were bacteriocin, T3PKS and NRPS BGCs. Strain DN50 harboured the largest number of different BGC types and the highest absolute BGC abundance (Figure 8).

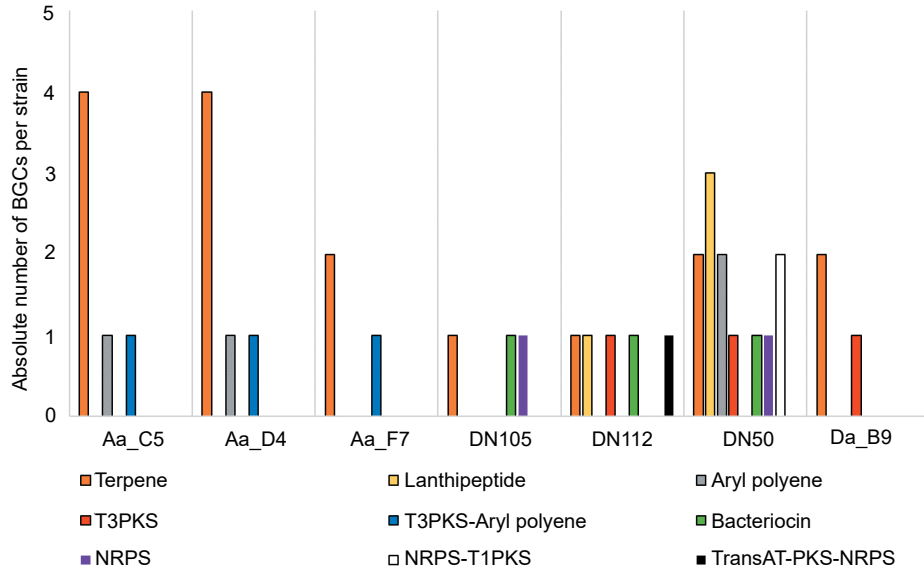


Figure 8. Composition of BGCs identified in the genomes newly sequenced in this study. The absolute number of BGCs per strain assigned to each BGC class is shown. Different colours indicate different BGC classes.

A more detailed comparative analysis of the BGCs showed that the necessary proteins predicted to synthesize carotenoid pigments were present in all seven flavobacterial genomes. Similarly, DN50 harboured a BGC related to the production of flexirubin, another type of bacterial pigment (Reichenbach et al., 1980; Bernardet, 2011; McBride, 2014). To accompany the genome mining results, antimicrobial activity of the seven sponge-associated flavobacterial strains was tested against six indicator strains (*Escherichia coli*, *Bacillus subtilis*, *Staphylococcus simulans*, *Aeromonas salmonicida*, *Candida oleophila*, *Saprolegnia parasitica*) belonging to the three categories Gram-positive, Gram-negative bacteria and fungi-oomycetes. Even though the genome predictions revealed a variety of secondary metabolite BGCs, none of the studied isolates showed antimicrobial activity under the examined conditions (data not shown).

Discussion

The family *Flavobacteriaceae* in the phylum *Bacteroidetes* currently contains more than 150 described genera that are widespread in marine and non-marine ecosystems (Kirchman, 2002). Thriving in such diverse habitats could indicate a high genetic plasticity. Disentangling the phylogeny of the family was controversial and challenging in the past, yet it has remained relatively stable for the past two decades (McBride, 2014). In this study, we followed the definition of flavobacterial clades as proposed by (McBride, 2014) with a focus on the Marine clade. Whole-genome reconstructed phylogeny based on 120 single copy bacterial marker

genes confirmed the 16S rRNA gene-based phylogeny by placing all newly sequenced strains in the Marine clade of the family *Flavobacteriaceae*.

We compared representative genomes of phylogenetically diverse flavobacteria inhabiting various niches to reveal novel insights into the characteristic properties of this group. The functional profiling based on Pfam entries revealed a high degree of distinction between the *Flavobacteriaceae* in comparison with other taxa dominant in marine environments, i.e. *Proteobacteria* and *Cyanobacteria*. This was also true for the genomic content of flavobacteria from different clades. However, comparison of functional profiles of host-associated and non-host-associated flavobacteria did not show significant dissimilarity. These results indicate that the functional traits within the *Flavobacteriaceae* were congruent with both single-copy marker gene- and 16S rRNA gene-based phylogeny (Liu et al., 2019; Silva et al., 2019a), rather than the habitat type, i.e. host-associated or non-host-associated (Silva et al., 2019a). However, it is important to mention that the environmental or organismal source of bacterial strains that have been isolated under laboratory conditions does not prove host-attachment in situ. Moreover, in marine environments a considerable number of host-associations might be coincidental as many microbes show metabolic versatility of a biphasic lifestyle-strategy that includes seawater particles ("free-living") and animal hosts (Wollenberg and Ruby, 2012; Bondarev et al., 2013; Versluis et al., 2018; Karimi et al., 2019), most likely in the search of nutrient sources (Wollenberg and Ruby, 2012). Previous studies provided evidence with respect to the phylogenetic conservation of traits associated with organic substrate utilization in different microorganisms (Martiny et al., 2013; Zimmerman et al., 2013) and particularly in members of the *Flavobacteriaceae* (Liu et al., 2019; Silva et al., 2019a). Interestingly, functional traits appeared to be highly conserved within the different flavobacterial clades (48% shared Pfams), even though the isolation sources of the strains were largely diverse. The key functions strongly enriched in *Flavobacteriaceae*, in general, and also in the Marine clade, were related to nutrient acquisition in regard to complex carbohydrate metabolism as well as gliding motility.

These two features have been repeatedly linked to the important role of *Flavobacteriaceae* in the degradation of high molecular weight organic matter within the marine environment (Kirchman, 2002; Fernandez-Gomez et al., 2013). Investigation of the repertoire of polymer-degrading enzymes among dominant marine bacterial groups revealed that *Flavobacteriaceae* had significantly more GHs, PLs, CEs and CBMs per Mbp compared to *Cyanobacteria* and *Proteobacteria*, supporting their pronounced capacity as polymer degraders and perhaps as key players in ocean carbon cycling (Fernandez-Gomez et al., 2013; Bennke et al., 2016; Kappelmann et al., 2019; Liu et al., 2019; Silva et al., 2019a). The capacity is justified by *Bacteroidetes'* particular genomic content which is enriched in genes encoding highly specific CAZymes, regulatory proteins and transporters arranged in clusters termed PULs that are required for depolymerization of complex polysaccharides (Grondin et al., 2017; Lapebie et al., 2019). Previous analyses of the PUL spectrum in marine *Flavobacteriaceae* revealed a similar average number of degradative CAZymes and PULs

(Kappelmann et al., 2019). Between the different flavobacterial clades, *Capnocytophaga* strains possessed more PULs and PUL-associated CAZymes per Mbp, likely reflecting their increased metabolic capabilities. This was also supported by their slightly larger genomes compared to the Marine clade strains. Genome size could be partially associated with the different ecological niches colonized by *Capnocytophaga* and Marine flavobacteria, as bacteria living in habitats with ample nutrient supply tend to have larger genomes and target more complex substrates (Thomas et al., 2011).

Regarding the function of the predicted CAZymes in the Marine clade genomes, the GH74 family occurs with the highest frequency and contains various enzymes that hydrolyse *β*-1,4 linkages of glucans. Thus, these enzymes assist in the degradation of different oligo- and polysaccharides, including xyloglucans, which is a hemicellulose polysaccharide present in plant cell walls and green algae (Del Bem and Vincentz, 2010). GH109, predicted as *α*-*N*-acetylgalactosaminidase, was the second most frequent GH family in the Marine clade genomes. These enzymes cleave *N*-acetylgalactosamine residues from various substrates such as glycolipids, glycopeptides and glycoproteins. It is highly likely that these substrates are derived from dissolved and/or particulate organic matter found in seawater, but also from the matrix of sponges (Kamke et al., 2013; Bayer et al., 2018). In addition, high frequencies of CBM50 and CBM44 genes were observed in the Marine clade. CBM50 proteins are found attached to enzymes that might cleave either bacterial peptidoglycan or animal chitin, and CBM44 are modules related to enzymes that bind both cellulose and xyloglucan (Bennke et al., 2016; Bayer et al., 2018). Taken together, our results reflect that the Marine clade flavobacteria have the genetic repertoire for utilizing a large diversity of carbon sources derived from algae, plants, bacteria and animals. This feature common in these microbes derived from numerous source habitats underlines the essential role of substrate utilization in the colonization of both host-associated and non-host-associated niches.

Many members of the *Bacteroidetes* and particularly *Flavobacteriaceae* can glide rapidly over surfaces using two novel and intertwined machines, one gliding motor involved in cell movement and one protein secretion system, known as T9SS (McBride and Zhu, 2013; McBride and Nakane, 2015; Johnston et al., 2018; McBride, 2019a). Gliding is common among members of the *Flavobacteriaceae* (Bernardet et al., 2002). This notion could be reinforced by the fact that genomes of all investigated gliding flavobacteria encoded homologs of the necessary components for T9SS-based gliding, with the exception of *Cellulophaga tyrosinoxydans*. This bacterium was lacking the two proteins GldF and GldG, which together with GldA are thought to be the components of an ATP-binding cassette (ABC) transporter involved in flavobacterial gliding (Hunnicutt et al., 2002). According to McBride and Zhu (2013), two other flavobacteria, *Cellulophaga algicola* and *Maribacter* sp. HTCC2170 were also actively gliding even though their genomes were lacking the genes encoding this ABC transporter (McBride and Zhu, 2013). This might suggest either the involvement of another ABC transporter, common in most organisms, or a non-essential role of this ABC transporter in *Bacteroidetes* gliding motility (McBride and Zhu, 2013; Zhu and McBride, 2016; McBride,

2019a). The T9SS exports the motility adhesins RemA and SprB which are propelled along the cell surface by the gliding motor, resulting in cell movement of *F. johnsoniae* (Shrivastava et al., 2012; McBride and Zhu, 2013; McBride and Nakane, 2015). SprB homologs were ubiquitous in gliding flavobacteria, whereas RemA was rare. This might indicate that SprB is more important for *Bacteroidetes* gliding. Additional novel motility adhesins may allow other species to move over diverse surfaces (McBride and Nakane, 2015; Zhu and McBride, 2016).

Gliding motility is considered as an efficient strategy to enhance survival, but the actual role in nature is still unknown. It has been previously suggested that gliding over surfaces facilitates access to nutrient sources (Nett and König, 2007), or is involved in pathogenesis (Sato et al., 2010) or symbiosis (Raina et al., 2019). Accordingly, in both non-host- and host-associated bacteria, the importance of T9SS has been demonstrated for many processes such as polymer degradation, adhesion, motility and virulence (McBride, 2019a). In this study, no specific link was found between host-association and gliding or type 9 secretion. Similarly, no difference was found when comparing the genomic repertoire of members of different taxonomic clades within the *Flavobacteriaceae*. These observations add to the fact that flavobacteria appear to be 'generalists', rather than 'specialists' and this might explain why they thrive in such diverse niches. Especially in the marine environment, the availability of nutrients might drive opportunistic host-microbe interactions, which might explain the presence of flavobacteria in the biomes of numerous marine hosts.

The high degree of conservation of Gld and Spr proteins in all *Flavobacteriaceae*, irrespective of their environmental source or taxonomic clade, supports the notion that T9SS-based gliding motility is widespread among *Bacteroidetes* members (McBride and Zhu, 2013) and especially among *Flavobacteriaceae*. Gliding motility occurs in other bacteria apart from the phylum *Bacteroidetes* (myxobacteria, cyanobacteria, mycoplasmas, etc.) but these employ their own unique machineries substratum-fixed protein complexes (A-motility), type IV pili (T4P) or membrane protrusions (Miyata, 2008; Balagam et al., 2014; Khayatan et al., 2015; Wilde and Mullineaux, 2015). Filamentous cyanobacteria such as *Anabaena* sp. and *Trichodesmium erythraeum* are thought to share a T4P-like nanomotor and polysaccharide secretion that drive and facilitate locomotion (Khayatan et al., 2015; Wilde and Mullineaux, 2015). Nevertheless, homologs to certain core *Bacteroidetes* gliding proteins (GldA, GldF, GldG, GldI, GldJ, GldK, SprE) were found to be encoded in genomes belonging to *Cyanobacteria* or *Proteobacteria*, regardless of their gliding status or mechanism. For example, GldA, GldF and GldG are similar to many putative components of the ABC transporter system, which is ubiquitous in all microorganisms (McBride and Zhu, 2013; Johnston et al., 2018). These proteins might serve another purpose in the respective organisms that might not be linked to gliding motility. In general, the paucity of homologs to *Bacteroidetes* gliding and T9SS proteins among bacteria from other phyla implies the presence of diverse and likely unrelated gliding motility mechanisms (Braun et al., 2005; McBride and Zhu, 2013; McBride, 2019a).

Gliding motility can also result in colony spreading, observed on agar plates as flat, irregularly edged colonies (McBride and Zhu, 2013; McBride, 2014; McBride and Nakane, 2015). The majority of the strains investigated for colony spreading (5 out of 7) were non-spreading on agar under the tested conditions, even though presence of all genes encoding proteins involved in the gliding motility apparatus was predicted by the genomic analysis. Presence of a gene in a genome does not yet imply gene expression. As described above, several bacterial strains that were previously described as 'non-gliding' had all essential genes present in their genomes.

Moreover, previous findings support that formation of non-spreading colonies is not necessarily an indication of absence of gliding motility. For example, *Maribacter sp.* HTCC2170 could glide well on glass but its colonies did not spread on agar (McBride and Zhu, 2013). In addition, it should be highlighted that the growth media used in our colony spreading tests were high in nutrients. Spreading under low nutrient conditions is more effective because non-metabolizable sugars tend to supress the active cell movement on agar media (Wolkin and Pate, 1984; Gorski et al., 1993; McBride, 2014). Thus, these bacteria may be motile in other conditions that were not examined here. On the other hand, the strains Aa_F7 and DN50, both carrying homologs for all T9SS-gliding proteins in their genomes, formed spreading colonies on agar displaying characteristic edges and a surface pattern indicative of secondary colony formation (Fig. 6). In contrast to the other strains, it is possible that these isolates could metabolize all sugars present in the high-nutrient media, thus surpassing the carbohydrate-mediated inhibitory effect on colony spreading. The absence of homologs to PorQ and PorU from Aa_F7 and DN50 genomes did not affect their gliding motility. Both PorQ and PorU are involved in the T9SS substrate attachment to the cell surface in *P. gingivalis* (Glew et al., 2012; Glew et al., 2017) but in *F. johnsoniae* their deletion did not affect gliding motility (Kharade and McBride, 2015). This might be another indication that PorQ and PorU are not essential for full motility (Veith et al., 2017; McBride, 2019a).

A wide variety of bioactive molecules, such as antibiotics, cell growth-promoting, antioxidative and neuroprotective compounds, have been previously isolated from members of the *Flavobacteriaceae* (McBride, 2014). Nevertheless, the metabolic potential of flavobacteria in terms of bioactivity has been underreported in literature. Across taxonomic groups, terpene BGCs were prevalent in the predicted metabolite arsenal of both *Cyanobacteria* and *Flavobacteriaceae*, particularly of the Marine clade. Terpenes are the largest class of small-molecule natural products, found in almost all life forms and performing diverse functions (Tholl, 2006). Previous findings showed that terpene synthases are widely distributed in bacteria, the majority of which are Gram-positive (*Actinomycetales*) but also in Gram-negative bacteria, such as *Cyanobacteria* and *Flavobacteriales* (Yamada et al., 2015). In terms of bioactivity, to date there are only a few reports on terpenes of microbial origin exhibiting antimicrobial (Zhao et al., 2008; Gallucci et al., 2009; Song et al., 2015) and antioxidant activity (Shindo et al., 2007). The almost complete absence of HSL BGCs from the *Flavobacteriaceae* genomes, compared

to their high relative abundance in *Proteobacteria* (Table 4), is in accordance with previous studies on flavobacteria that suggest the existence of alternative QS signalling molecules for *Bacteroidetes* (Harms et al., 2018b; Hudson et al., 2019). For example, strains of *Zobellia galactanivorans* (member of *Flavobacteriaceae*) were found to synthesize another type of QS signalling molecules belonging to dialkylresorcinols, a class of natural products with antibiotic and antioxidant activity (Harms et al., 2018b). In addition, the detection of a HSL-degrading enzyme in the genome of *Z. galactanivorans* (Barbeyron et al., 2016) implies the presence of a communication system distinct from the HSL-based QS system that could function as a defence mechanism with antibiotic effect (Harms et al., 2018b). Within *Flavobacteriaceae*, the secondary metabolite repertoire did not appear to be influenced by phylogeny, as the global composition of predicted BGCs was similar among the clades. Besides terpene synthases, BGCs responsible for the biosynthesis of lanthipeptides were also found in high abundance in flavobacterial genomes. Formerly known as lantibiotics, they are ribosomally synthesized post-translationally modified peptides that belong to class I bacteriocins (Knerr and van der Donk, 2012). It has been previously suggested that bacteriocins play a critical role in mediating microbial community dynamics (Riley and Wertz, 2002; Desriac et al., 2010). Presumably, bacteriocin-producing flavobacteria act as anti-competitors, facilitating the invasion of other closely related species into an established microbial community to successfully colonize a niche. An additional role suggested for bacteriocins is their involvement in host defence, protecting the host from pathogens (Riley and Wertz, 2002).

Genome mining of the BGC arsenal of the sponge-associated strains sequenced in this study revealed homologs to carotenoid-producing BGCs in all seven genomes. Carotenoids are often the most abundant pigments in marine flavobacteria, and their importance lies in their strong antioxidant properties. In flavobacteria, a few structurally rare carotenoids have been identified before, such as saproxanthin and myxol that show significant antioxidative activities and neuroprotective effects (Shindo et al., 2007). Flexirubins represent another type of pigment common in *Bacteroidetes* but not in other bacteria (McBride, 2014). A BGC similar to one shown to encode proteins needed for flexirubin production was identified in strain DN50. Interestingly, such pigments exhibit anticancer and antimicrobial properties (e.g. against *Mycobacterium* sp.) and are considered promising candidates for treatment and prevention of cancer and microbial infections (Mojib et al., 2010; Shim and Liu, 2014).

Even though the genome predictions revealed a large number and variety of secondary metabolite BGCs, none of the studied isolates showed antimicrobial activity under the conditions we examined. The majority of these biosynthetic loci are frequently dormant or expressed at low constitutive levels under laboratory conditions, keeping the true biosynthetic potential of microorganisms hidden and thus hampering the discovery of novel bioactive compounds (Hertweck, 2009; Rutledge and Challis, 2015). This ‘silent’ state can be reversed by inducers of gene expression, such as environmental cues, nutrients or signal molecules (Hertweck, 2009; Rutledge and Challis, 2015). Unlocking the full metabolic potential encoded

by the studied strains might require high-throughput screening of various growth conditions in combination with a large number of indicator strains.

Conclusions

The comparative genomics analysis performed in this study demonstrated that 16S rRNA gene- and single-copy marker gene-based phylogeny, rather than life strategy of the organisms is the main factor correlated to the functional profile of *Flavobacteriaceae*. The traits responsible for the functional divergence between phyla investigated here were found to be associated with gliding motility and nutrient acquisition through the catabolism of carbohydrates. Marine flavobacteria appear to be potent utilizers of a large variety of carbon polymers from algae, bacteria, plants and animals, confirming their role in the ocean carbon cycling as exceptional degraders of particulate organic matter. Additionally, inspection of the gene content revealed the occurrence of homologs for all major components of the T9SS-gliding motility apparatus in all *Flavobacteriaceae* in contrast to members of other phyla (*Cyanobacteria* and *Proteobacteria*) that are known to use different mechanisms for gliding. Phenotypic assays showed the formation of spreading colonies for some of the tested flavobacteria that had the complete set of T9SS-gliding homologs, confirming that not all potentially gliding bacteria form spreading colonies on agar. In terms of their secondary metabolic potential, a large diversity of BGCs was identified in the studied genomes, with terpene BGCs being highly prevalent in both *Flavobacteriaceae* and *Cyanobacteria*. Other BGCs that potentially encode proteins required for the production of compounds with known antimicrobial, antioxidant and anticancer properties were found in *Flavobacteriaceae*. Nevertheless, bioactivity tests did not reflect these genomic findings supporting the fact that the true biosynthetic potential of microorganisms remains hidden due to the 'dormant' state of gene clusters under laboratory conditions. Hence, while this study provides the required broad overview of the genomic content of *Flavobacteriaceae* in terms of their carbon metabolism, gliding motility and secondary metabolite biosynthetic potential, further studies are essential to enhance our current understanding of these distinct features and how flavobacteria implement them in their natural environment.

Acknowledgements

We thank Bart Nijsse, Nikolaos Strepis and Sudarshan Shetty from the Laboratory of Microbiology, Wageningen University & Research, Wageningen, Netherlands for their advice regarding data analysis.

Supplementary Information

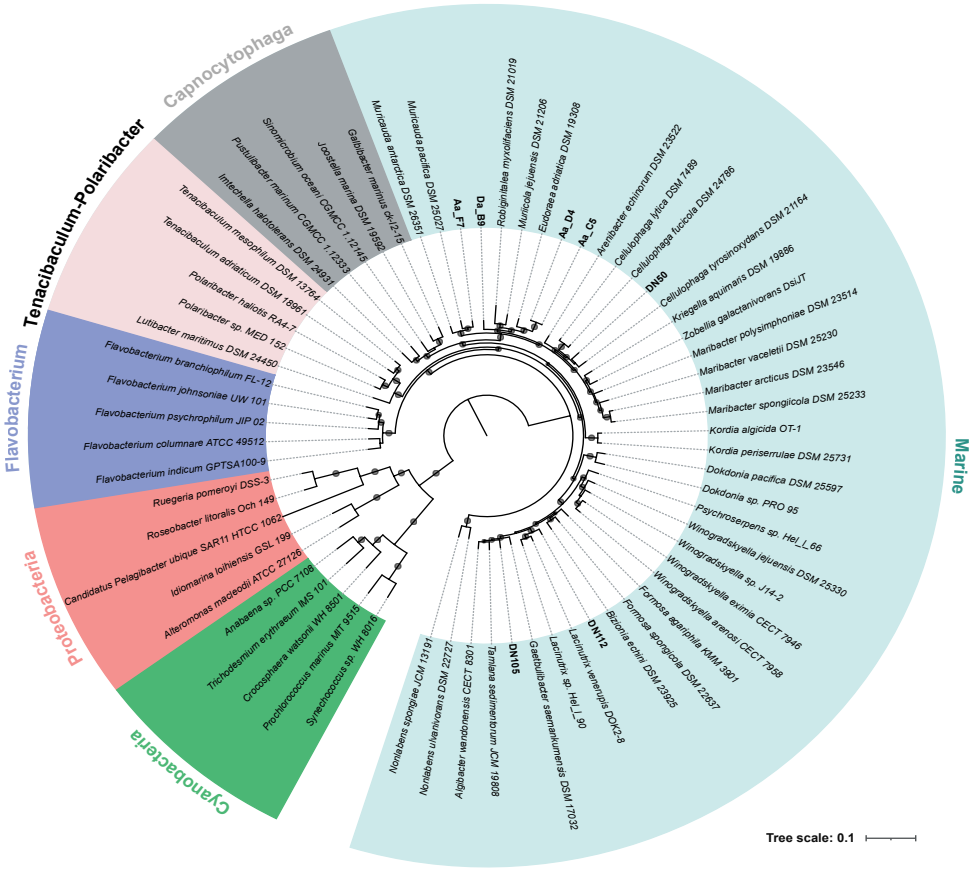


Figure S1. Maximum likelihood tree of the 66 genomes generated in this study based on single-copy marker proteins. Phylogeny was inferred from the concatenation of 120 conserved amino-acid sequences by GTDB-Tk. Black circles in the middle of the branches represent Shimodaira-Hasegawa (SH) likelihood support values. Colour annotations represent the different clades and phyla. Sequences belonging to *Cyanobacteria* and *Proteobacteria* were used as outgroups. Names in bold indicate sequences generated in the present study. Scale bar represents amino acid substitutions per site.

Table S1. Strains and growth media. Details on the preparation of the media and the cultivation conditions can be found in the References section.

Strain ID	Isolation source	Growth Medium	References
DN50	*A. aerophoba*	Mucin & Marine agar	Versluis et al. (2017)
DN105	*A. aerophoba*	Marine agar	Versluis et al. (2017)
DN112	*A. aerophoba*	Marine agar	Versluis et al. (2017)
Da_B9	*D. avara*	Marine agar	unpublished
Aa_F7	*A. aerophoba*	Mucin (50x)	Gutleben et al. (2020)
Aa_C5	*A. aerophoba*	All amino acids (50x)	Gutleben et al. (2020)
Aa_D4	*A. aerophoba*	All amino acids (50x)	Gutleben et al. (2020)

Table S2. Publicly available genomes used in this study (Source: https://gold.jgi.doe.gov/).

IMG Genome Name / Sample Name	IMG Genome ID	Sequencing Status	Genome size (bp)	Gene Count	16SrRNA GenBank accession number	Isolation source
Algibacter wandonensis CECT 8301	2772190795	Draft	4,707,503	4004	KC987358	sediment
Alteromonas macleodii ATCC 27126	2554235747	Finished	4,653,851	3941	CP003841	seawater
Anabaena sp. PCC 7108	2506485002	Permanent Draft	5,886,741	5227	AJWF01000009	seawater
Arenibacter echinorum DSM 23522	2593339296	Permanent Draft	5,193,588	4393	EF536748	sea urchin
Bizionia echini DSM 23925	2622736504	Draft	3,314,721	3072	FJ716799	sea urchin
Candidatus Pelagibacter ubique SAR11 HTCC1062	637000058	Finished	1,308,759	1394	CP000084	seawater
Cellulophaga fucicola DSM 24786	2595699004	Draft	3,922,041	3495	AJ005973	algae
Cellulophaga lytica LIM-21, DSM 7489	649633032	Finished	3,765,936	3358	AB517706	sediment
Cellulophaga tyrosinoxydans DSM 21164	2595698249	Permanent Draft	3,555,791	3209	EU443205	seawater
Crocosphaera watsonii WH 8501	2623620439	Draft	6,291,599	6836	AADV02000003	seawater
Dokdonia pacifica DSM 25597	2724679821	Draft	5,520,745	4968	KP862606	seawater
Dokdonia sp. PRO95	2597489917	Finished	3,305,093	3038	FJ627052	seawater
Eudoraea adriatica DSM 19308	2522572201	Permanent Draft	3,906,474	3579	AM745437	seawater
Flavobacterium branchiophilum FL-15	2561511155	Finished	3,563,292	2925	FQ859183	fish pathogen
Flavobacterium columnare ATCC 49512	2511231122	Finished	3,162,865	2735	CP003222	fish pathogen
Flavobacterium indicum GPTSA100-9	2540341066	Finished	2,993,089	2738	HE774682	spring water
Flavobacterium johnsoniae UW101, ATCC 17061	644736369	Finished	6,096,872	5099	CP000685	soil/freshwater
Flavobacterium psychrophilum JIP02/86	640753027	Finished	2,861,988	2505	AM398681	fish pathogen
Formosa agariphila KMM 3901	2585427664	Permanent Draft	4,228,350	3630	AY187688	algae
Formosa spongicola DSM 22637	2595698203	Permanent Draft	3,155,334	2961	FJ348469	sponge
Gaetbulibacter saemankumensis DSM 17032	2524614665	Permanent Draft	3,089,149	2746	AY883937	sediment
Galbibacter marinus ck-I2-15	2519899576	Permanent Draft	3,572,447	3138	EU928746	sediment
Idiomarina loihiensis GSL 199	2554235415	Finished	2,839,759	2717	CP005964	seawater

Table S2. Continued.

IMG Genome Name / Sample Name	IMG Genome ID	Sequencing Status	Genome size (bp)	Gene Count	16SrRNA GenBank accession number	Isolation source
Imtechella halotolerans K1	2534681666	Permanent Draft	3,086,951	2734	FR774044	brackish water
Joostella marina DSM 19592	2509276026	Permanent Draft	4,508,243	4004	EF660761	seawater
Kordia algicida OT-1	641380434	Permanent Draft	5,019,836	4584	AY195836	seawater
Kordia periserrulae DSM 25731	2734482288	Draft	4,725,576	4128	GU233518	polychaete
Kriegella aquimaris DSM 19886	2622736525	Draft	6,057,242	5014	AB084262	seawater
Lacinutrix sp. Hel_I _90	2582581868	Permanent Draft	3,819,763	3506	JX854138	seawater
Lacinutrix venerupis DOK2-8	2751185745	Draft	3,192,399	2890	CP019352	seawater
Lutibacter maritimus DSM 24450	2622736611	Draft	3,484,703	3147	FJ598048	sediment
Maribacter arcticus DSM 23546	2595698209	Permanent Draft	4,211,145	3760	AY771762	sediment
Maribacter polysiphoniae DSM 23514	2595698208	Permanent Draft	5,129,962	4489	AM497875	algae
Maribacter spongiicola DSM 25233	2731957517	Draft	4,455,565	3894	JX050191	sponge
Maribacter vaceletii DSM 25230	2734482098	Draft	3,889,815	3392	JX050190	sponge
Muricauda antarctica DSM 26351	2619619048	Draft	4,482,831	4102	JN166984	seawater
Muricauda pacifica DSM 25027	2731639122	Draft	4,376,054	4153	JN118551	seawater
Muriicola jejuensis DSM 21206	2724679820	Draft	3,299,848	3040	EU443206	seawater
Nonlabens spongiae JCM 13191	2751185769	Draft	3,393,335	3097	DQ064789	sponge
Nonlabens ulvanivorans DSM 22727	2593339289	Permanent Draft	3,177,440	2918	GU902979	algae
Polaribacter haliotis RA4-7	2788500510	Draft	3,780,569	3361	KX450477	abalone
Polaribacter sp. MED152	638341218	Finished	2,967,100	2723	CP004349	seawater
Prochlorococcus marinus MIT9515	640069324	Finished	1,704,176	1968	CP000552	seawater
Psychroserpens sp. Hel_I_66	2585427602	Permanent Draft	3,842,990	3475	JUGU01000001	seawater
Pustulibacterium marinum CGMCC 1.12333	2663762752	Draft	4,209,902	3796	FPBK01000031	seawater
Robiginitalea myxolifaciens DSM 21019	2636416043	Draft	3,206,962	2897	AB270585	sediment

Table S2. Continued.

IMG Genome Name / Sample Name	IMG Genome ID	Sequencing Status	Genome size (bp)	Gene Count	16SrRNA GenBank accession number	Isolation source
Roseobacter litoralis Och 149	2510065042	Finished	4,745,450	4577	CP002623	macroalgae
Ruegeria pomeroyi DSS-3	637000267	Finished	4,601,053	4355	CP000031	seawater
Sinomicrobium oceani CGMCC 1.12145	2596583572	Draft	5,026,272	4281	FPJE01000054	sediment
Synechococcus sp. WH 8016	2507262052	Finished	2,706,690	3046	AGIK01000004	seawater
Tamlana sedimentorum JCM 19808	2636415452	Draft	3,961,353	3504	AB894238	sediment
Tenacibaculum adriaticum DSM 18961	2756170228	Draft	3,211,939	2987	AM412314	bryozoan
Tenacibaculum mesophilum DSM 13764	2695420951	Draft	3,286,619	3059	AB032501	sponge
Trichodesmium erythraeum IMS101	637000329	Finished	7,750,108	5156	CP000393	seawater
Winogradskyella arenosi CECT 7958	2770939486	Draft	3,675,157	3253	AB438962	sediment
Winogradskyella eximia CECT 7946	2770939573	Draft	4,242,526	3742	AY521225	algae
Winogradskyella jejuensis DSM 25330	2695420983	Draft	3,033,897	2864	JF820844	algae
Winogradskyella sp. J14-2	2751185740	Draft	3,349,669	3049	CP019388	seawater
Zobellia galactanivorans DsijT	2619619092	Draft	5,521,712	4563	AF208293	algae

Table S3. Cultivation conditions of indicator strains and antibiotics used in the antimicrobial activity tests.

Reference strains	Temperature (°C)	Media	Antibiotics (Positive control)
Bacillus subtilis (DSM 402)	30	Nutrient broth	Tetracycline
Staphylococcus simulans (DSM 20037)	37	Tryptic Soy broth (TSB)	Vancomycin
Escherichia coli (MIB, WUR)	37	Luria-Bertani broth (LB)	Tetracycline
Aeromonas salmonicida (DSM 19634)	30	Nutrient broth	Vancomycin
Saprolegnia parasitica (CBS 223.65)	20	Potato Dextrose agar (PDA)	Delvocid
Candida oleophila (DSM 70763)	25	Yeast Malt Broth (YMB)	Nystatin

Table S4. Taxonomic assignment of flavobacterial strains sequenced in this study. Information on the 16S rRNA gene sequence of each strain, BLASTN best hits against nr/nt NCBI database and GTDB-Tk classification.

Strain ID	16S rRNA gene length (bp)	Closest BLASTN hit (Accession number)	ID%	GTDB-Tk classification
Aa_C5	1362	*Eudoraea chungangensis* (NR_148299.1)	99.7	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Eudoraea;s
Aa_D4	1349	*Eudoraea chungangensis* (NR_148299.1)	99.7	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Eudoraea;s
Aa_F7	1257	*Flagellimonas* sp. (MF287795.1)	97.9	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flagellimonas;s
Da_B9	1510	*Flavobacteriaceae* bacterium (MK956921.1)	94.0	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__;s
DN50	1121	Uncultured organism clone (DQ395462.1)	96.6	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__GCA-2746415;s
DN105	1190	*Flavobacteriaceae* bacterium (EU581699.1)	98.8	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Flaviramulus_A;s
DN112	1391	*Lacinutrix* sp. (KX398615.1)	98.2	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__;s

Table S5. Pfam entries most strongly contributing to differentiating genomes from different *Flavobacteriaceae* clades. Pfam entries with the highest significant contribution (> 0.2%, p < 0.05) to the dissimilarity are shown. M, Marine; C, *Capnocytophaga*; T, *Tenacibaculum-Polaribacter*; F, *Flavobacterium*.

Function name (Pfam ID)	Relative Abundance (%)				Contribution (%)			p-values (< 0.05)		
	M	C	T	F	M-C	M-T	M-F	M-C	M-T	M-F
CarboxypepD_reg-like domain (pfam13715)	1.20	1.78	1.28	0.95	1.06	0.51	0.68	0.011	0.999	0.398
TonB-dependent Receptor Plug Domain (pfam07715)	1.08	1.65	1.16	0.97	0.93	0.64	0.57	0.01	0.984	0.731
SusD family (pfam07980)	0.30	0.81	0.26	0.17	0.90	0.36	0.41	0.002	1	0.616
Starch-binding associating with outer membrane (pfam14322)	0.28	0.78	0.24	0.17	0.71	0.38	0.43	0.002	1	0.591
TonB dependent receptor (pfam00593)	0.53	0.90	0.47	0.45	0.62	0.36	0.43	0.01	1	0.438
Sigma-70, region 4 (pfam08281)	0.53	0.62	0.52	0.26	0.42	0.23	0.39	0.04	0.999	0.027
FecR protein (pfam04773)	0.12	0.31	0.08	0.04	0.41	0.17	0.18	0.015	0.977	0.492
Domain of unknown function (DUF4974) (pfam16344)	0.12	0.30	0.08	0.05	0.40	0.17	0.19	0.016	0.978	0.469
Glycosyl hydrolases family 43 (pfam04616)	0.06	0.28	0.02	0.03	0.40	0.10	0.11	0.001	0.998	0.931
Phage integrase family (pfam00589)	0.19	0.32	0.19	0.27	0.33	0.15	0.21	0.015	0.999	0.112
Outer membrane efflux protein (pfam02321)	0.25	0.44	0.28	0.31	0.32	0.13	0.14	0.001	0.993	0.432
Barrel-sandwich domain of CusB or HlyD membrane-fusion (pfam16576)	0.20	0.37	0.22	0.29	0.31	0.16	0.16	0.001	0.676	0.252
AcrB/AcrD/AcrF family (pfam00873)	0.16	0.33	0.19	0.19	0.31	0.13	0.12	0.001	0.888	0.606
ABC transporter (pfam00005)	0.88	0.79	0.86	0.96	0.30	0.22	0.23	0.039	0.955	0.202
Tetratricopeptide repeat (pfam13181)	0.33	0.16	0.27	0.24	0.29	0.20	0.30	0.022	0.938	0.574
Phage integrase SAM-like domain (pfam13102)	0.09	0.23	0.09	0.06	0.28	0.15	0.09	0.003	0.685	0.996
Integrase core domain (pfam00665)	0.03	0.18	0.01	0.11	0.27	0.07	0.16	0.016	0.999	0.141
Glycosyl hydrolases family 2, sugar binding domain (pfam02837)	0.08	0.20	0.06	0.05	0.25	0.17	0.15	0.04	0.733	0.514
Transposase DDE domain (pfam01609)	0.01	0.15	0.00	0.15	0.25	0.02	0.23	0.048	0.999	0.027
Glycosyl hydrolases family 2 (pfam00703)	0.07	0.19	0.06	0.06	0.24	0.17	0.14	0.033	0.669	0.567
IstB-like ATP binding protein (pfam01695)	0.00	0.14	0.00	0.00	0.24	0.00	0.00	0.001	0.999	1
Starch-binding associating with outer membrane (pfam12771)	0.08	0.18	0.08	0.06	0.20	0.10	0.08	0.005	0.968	0.992
Heavy-metal-associated domain (pfam00403)	0.13	0.23	0.19	0.08	0.20	0.16	0.10	0.024	0.409	0.932
C-terminal domain of CHU protein family (pfam13585)	0.41	0.26	0.26	0.37	0.27	0.30	0.19	0.061	0 .045	0.461
Response regulator receiver domain (pfam00072)	1.11	1.01	1.08	1.04	0.46	0.39	0.52	0.529	0.995	0.037
Sigma-70 region 2 (pfam04542)	0.66	0.71	0.60	0.32	0.45	0.23	0.43	0.063	1	0.003
Helix-turn-helix (pfam01381)	0.18	0.27	0.18	0.33	0.21	0.16	0.30	0.386	0.994	0.005
Enoyl-(Acyl carrier protein) reductase (pfam13561)	0.33	0.34	0.27	0.19	0.20	0.21	0.26	0.834	0.938	0.024
Sodium:solute symporter family (pfam00474)	0.18	0.20	0.20	0.05	0.11	0.13	0.20	0.995	0.88	0.001

Table S6. Number of BGCs, average gene counts and % of genes in BGCs per group.

Groups	Total BGCs	Average BGCs	Average genes	% genes in BGCs
Flavobacteriaceae (n = 56)	197	4	4447	0.1
Marine (n = 41)	151	4	3668	0.1
Capnocytophaga (n = 5)	15	3	3591	0.1
Flavobacterium (n = 5)	20	4	3200	0.1
Tenacibaculum-Polaribacter (n = 5)	11	2	3055	0.1
Cyanobacteria (n = 5)	33	7	3397	0.3
Proteobacteria (n = 5)	20	4	3565	0.1

Table S7. Distribution of identified BGCs across different groups and *Flavobacteriaceae* clades.
M, Marine; C, *Capnocytophaga*; F, *Flavobacterium*; T, *Tenacibaculum-Polaribacter*.

Gene cluster type (antiSMASH)	*Flavobacteriaceae*								*Cyanobacteria*		*Proteobacteria*	
	M		C		F		T					
	Total	%	Total	%	Total	%	Total	%	Total	%	Total	%
Aryl polyene	7	4.6	0	0.0	3	15.0	1	9.1	3	9.09	1	5.0
Aryl polyene-Resorcinol	7	4.6	0	0.0	3	15.0	0	0.0	0	0.00	0	0.0
Aryl polyene-Resorcinol-T3PKS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
Bacteriocin	5	3.3	1	6.7	1	5.0	0	0.0	7	21.21	4	20.0
Bacteriocin-Lanthipeptide	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
Betalactone	0	0.0	0	0.0	1	5.0	0	0.0	0	0.00	2	10.0
Ectoine	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
Hserlactone	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	7	35.0
Ladderane	0	0.0	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
Lanthipeptide	9	6.0	2	13.3	0	0.0	1	9.1	0	0.00	0	0.0
LAP-Bacteriocin-Cyanobactin	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
Lassopeptide	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
Linaridin	2	1.3	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
Microviridin	0	0.0	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
NRPS	12	7.9	0	0.0	1	5.0	0	0.0	5	15.15	0	0.0
NRPS-like	2	1.3	2	13.3	0	0.0	0	0.0	3	9.09	0	0.0
NRPS-like-T1PKS	0	0.0	0	0.0	0	0.0	0	0.0	0	0.00	1	5.0
NRPS-T1PKS	8	5.3	0	0.0	0	0.0	0	0.0	2	6.06	0	0.0
Resorcinol	2	1.3	1	6.7	0	0.0	0	0.0	0	0.00	0	0.0
Siderophore	0	0.0	0	0.0	3	15.0	1	9.1	0	0.00	1	5.0
T1PKS	1	0.7	1	6.7	0	0.0	0	0.0	1	3.03	0	0.0
T1PKS-PUFA-hg1E-KS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
T3PKS	16	10.6	0	0.0	1	5.0	1	9.1	1	3.03	0	0.0
T3PKS-Aryl polyene	8	5.3	0	0.0	0	0.0	1	9.1	0	0.00	0	0.0
Terpene	67	44.4	6	40.0	7	35.0	6	54.5	8	24.24	2	10.0
Terpene-Bacteriocin	0	0.0	0	0.0	0	0.0	0	0.0	1	3.03	0	0.0
TransAT-PKS-like, TransAT-PKS, PKS-like	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0
TransAT-PKS-NRPS	1	0.7	0	0.0	0	0.0	0	0.0	0	0.00	0	0.0

Chapter 3

Bioactivity Screening and Gene-Trait Matching across Marine Sponge-Associated Bacteria

Asimenia Gavriilidou, Thomas Andrew Mackenzie, Pilar Sánchez, José R. Tormo, Colin Ingham, Hauke Smidt and Detmer Sipkema

This chapter was published as Gavriilidou, A., *et al.* (2021) *Mar. Drugs* 2021, 19(2), 75; https://doi.org/10.3390/md19020075

Abstract

Marine sponges harbour diverse microbial communities that represent a significant source of natural products. In the present study, extracts of 21 sponge-associated bacteria were screened for their antimicrobial and anticancer activity, and their genomes were mined for secondary metabolite biosynthetic gene clusters (BGCs). Phylogenetic analysis assigned the strains to four major phyla in the sponge microbiome, namely *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Firmicutes*. Bioassays identified one extract with anti-methicillin resistant *Staphylococcus aureus* (MRSA) activity, and more than 70% of the total extracts had a moderate to high cytotoxicity. The most active extracts were derived from the *Proteobacteria* and *Actinobacteria*, prominent for the production of bioactive substances. The strong bioactivity potential of the aforementioned strains was also evident in the abundance of BGCs, which encoded mainly beta-lactones, bacteriocins, non-ribosomal peptide synthetases (NRPS), terpenes and siderophores. Gene-trait matching was performed for the most active strains, aiming at linking their biosynthetic potential with the experimental results. Genetic associations were established for the anti-MRSA and cytotoxic phenotypes based on the similarity of the detected BGCs with BGCs encoding natural products with known bioactivity. Overall, our study highlights the significance of combining in vitro and in silico approaches in the search of novel natural products of pharmaceutical interest.

Introduction

The fight against cancer and infectious diseases are two of the main challenges the scientific and medical community have been facing during the past decades. A large proportion of compounds for the treatment of human ailments throughout history, from traditional remedies to current-day pharmaceuticals, belonged to natural products (Romano et al., 2017). However, the decline in the discovery rates of new bioactive molecules led to scepticism on the potential of drug development from natural products (Dyshlovoy and Honecker, 2019). Nowadays, in the light of the urgent need for new drugs, particularly those with anticancer and anti-infective properties, increasing attention is returning towards natural products and more specifically, from marine sources (Paterson and Anderson, 2005; Desbois, 2014). The field of marine drug discovery has been growing over the past 20 years, with currently almost 35,000 research articles on natural products of marine origin (Marinlit; https://marinlit.rsc.org/, accessed on 20 November 2020). This is also highlighted by the number of novel secondary metabolites being elucidated per year (1554 in 2018) and a significant number of marine-derived drug candidates under clinical trials or pending approval (Mayer et al., 2010; Hu et al., 2015; Jaspars et al., 2016).

Despite the intensive research effort, the marine environment can be considered rather underexplored for prospecting bioactive molecules in comparison with terrestrial ecosystems (Hughes and Fenical, 2010; Santos et al., 2020a). It encompasses an enormous biodiversity and chemical richness covering approximately 70% of the Earth's surface (Pomponi, 1999). From the marine biosphere, sponges are standing out, often termed as 'chemical factories' or 'gold mines' owing to the production of metabolites with various pharmaceutically interesting biological activities (Sipkema et al., 2005; Perdicaris et al., 2013; Lackner et al., 2017). The ecological role of these molecules is linked to the adaptation to the marine ecosystem where chemical defence is a survival mode for sponges against predators, competitors, fouling organisms and infectious microbes (Paul et al., 2006; Paul et al., 2011).

An integral part of marine sponges is their symbiotic microorganisms which comprise up to 40% of their total volume, forming close associations with the host (Hentschel et al., 2002; Webster and Taylor, 2012), a concept recently defined as 'sponge holobiont' (Pita et al., 2018). Microbial associates have been found to contribute significantly to many functions of the host, including nutrition, health and defence (Taylor et al., 2007a; Webster and Taylor, 2012; Thomas et al., 2016; Pita et al., 2018). In fact, there is a growing amount of experimental evidence that many bioactive substances produced by sponges are of bacterial origin, suggesting sponge-associated bacteria as the real producers rather than the host itself (Piel, 2009; Wilson et al., 2014; Lackner et al., 2017; Morita and Schmidt, 2018; Tianero et al., 2019). Two main findings support this theory, the structural similarity of molecules derived from taxonomically distant taxa and the case of bioactive compounds found in sponges but known to be synthesised by bacteria, such as polyketides and non-

ribosomal peptides (Piel et al., 2004; Moore, 2006; Taylor et al., 2007a; Santos-Gandelman et al., 2014; Waters et al., 2014; Gomes et al., 2016).

More than 60 microbial phyla inhabit marine sponges (Moitinho-Silva et al., 2017b), from which the most predominant ones are the *Proteobacteria*, *Chloroflexi*, *Thaumarchaeota*, *Cyanobacteria*, *Candidatus* Poribacteria, *Acidobacteria* and *Actinobacteria* (Thomas et al., 2016; Moitinho-Silva et al., 2017b). Symbiont-derived compounds have displayed a wide spectrum of biological activities including antimicrobial, anticancer/cytotoxic, anti-inflammatory, antifouling. The majority of sponge-associated microorganisms have been found to synthesise compounds with antimicrobial, antitumor and cytotoxic properties (Thomas et al., 2010b; Mehbub et al., 2014). In the sponge holobiont, the most prominent bacterial producers of antimicrobials are members of the *Actinobacteria* phylum, followed by *Proteobacteria*, *Firmicutes* and *Cyanobacteria* (Thomas et al., 2010b; Santos-Gandelman et al., 2014; Indraningrat et al., 2016; Brinkmann et al., 2017b). The most prolific sources of antimicrobial compounds among bacteria isolated from sponges are the following genera: *Streptomyces*, *Pseudovibrio* and *Bacillus* (Indraningrat et al., 2016). Recently, the study of a novel symbiotic bacterium *Bacillus tequilensis* of the marine sponge *Callyspongia diffusa* led to the discovery of an antibiotic agent active against multidrug-resistant *Staphylococcus aureus* (Kiran et al., 2018). Another recent example was the elucidation of three new flavonoids from sponge-derived *Streptomyces* sp. with antimicrobial activity (Cao et al., 2019). Likewise, most anticancer (i.e., cancer-preventive, antitumor, cytotoxic) molecules produced by sponge-bacteria associations are derived from *Proteobacteria* (*γ-Proteobacteria*), *Actinobacteria* and *Firmicutes* (Thomas et al., 2010b). Some recent examples of natural products isolated from sponge-dwelling bacteria with antitumor and/or cytotoxic potential are nocardiotide A produced by *Nocardiopsis* sp. associated with *Callyspongia* sp. (Ibrahim et al., 2018) and the dipeptide cyclo(-Pro-Tyr) produced by *Bacillus pumilus* associated with *Callyspongia fistularis* (Karanam et al., 2019).

Gaining full access to the actual producers via cultivation and discovering new lead structures are two of the main bottlenecks that exist in the drug discovery pipeline (Lindequist, 2016; Romano et al., 2017). The biosynthetic capacity of even the most well-characterised microbes has been underexplored since a vast number of biosynthetic gene clusters (BGCs) remain cryptic (Reen et al., 2015; Lindequist, 2016; Zhang et al., 2016). A promising strategy to expand our understanding of the true metabolic potential of microorganisms and increase the probability of finding new molecules involves in silico genome mining of cultivable isolates (Bachmann et al., 2014; Liu et al., 2021). In the current study, our aim was to (1) in vitro assess the activity of sponge-associated bacteria against various human pathogens and cancer cells, (2) examine their secondary metabolite biosynthesis potential via genome mining and (3) identify the BGCs potentially involved in the observed bioactivity by gene-trait matching.

Methods

Isolation of Strains and Growth Conditions

Twenty-one bacterial isolates were obtained from six different sponge species (*Aplysina aerophoba*, *Petrosia* ficiformis, *Corticium candelabrum*, *Ircinia* sp., *Chondrilla nucula* and *Acanthella acuta*), in previous studies (Versluis et al., 2017b; Versluis et al., 2018). Cryopreserved glycerol stocks of the strains were initially used as inoculum for regrowth on the original solid isolation media (Table S5) at 20 °C. Single colonies were picked and cultured in 250 mL Erlenmeyer flasks containing 20 mL of the respective liquid media in duplicates. The flasks were incubated at 20 °C and shaken at 150 rpm in the dark for seven days. OD_{600} measurements were taken every 24 h in order to monitor the cell growth. After seven days of incubation and all cultures reaching the stationary phase, the content of the flask was stored at -20 °C until chemical extraction.

Strain Identification and Sanger Sequencing

The identity of the strains was confirmed by 16S ribosomal RNA (rRNA) gene Sanger sequencing. Single colonies were picked, stored in 100 µL nuclease-free water at -20 °C and served as template. 16S rRNA gene amplicons were generated by PCR using primers 27F (5'-AGAGTTTGGATCMTGGCTCAG-3') and 1492R (5'-CGGTTACCTTGTTACGACTT-3') (Lane, 1991). PCR reaction mixture and conditions were the same as described earlier (Gavriilidou et al., 2020). PCR products were purified using the CleanPCR kit (CleanNA, The Netherlands) and quantified using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA). The purified PCR products were sequenced at Eurofins Genomics (Cologne, Germany) with primers 27F and 1492R. Quality control of the raw sequences was done with BioEdit (Hall, 2011) by inspecting the chromatograms. In addition, consensus sequences of the near full-length 16S rRNA gene were obtained by aligning the forward and reverse read with default settings. Taxonomic classification based on the near full-length 16S rRNA gene sequences was performed using the blastn suite (Altschul et al., 1990) against the NCBI nr/nt database (accessed on November 2020), considering as best hits the ones belonging to cultured representatives (Tables 1 and S2).

Chemical Extraction

The culture broths from the Erlenmeyer flasks were extracted with the use of Sepabeads® SP207SS resin (Sorbent Technologies, Inc., Norcross, GA, USA). In particular, 20 mL of acetone was added to 10 mL of culture broth in EPA vials (Dispolab, Someren, The Netherlands). After 2 h of shaking at 220 rpm and room temperature (Kuhner ISF4-X Climo-Shaker, Kuhner Shaker S.A., Barcelona, Spain), the solvent was evaporated under a nitrogen stream overnight. Distilled water until final volume of 10 mL and 3 mL of a suspension of SP207ss resin dissolved in water were added to each sample. The vials were then placed in the incubator shaker at 220 rpm and room temperature for 2 h. A centrifugation step followed

to settle the resin at 3,500 × g for 15 min (Speed Vac Plus SC210A, Savant Instruments, Inc., Holdbrook, NY, USA) and supernatants were discarded. Resin was washed by adding 16 mL of distilled water into each vial, followed by shaking at 220 rpm and room temperature for 2 h. Samples were again centrifuged at 3,500 × g for 15 min and supernatants were discarded. To extract the adsorbed molecules from the resin, 10 mL of acetone was added to the resin for overnight mixing (Kuhner ISF4-X Climo-Shaker) at 20 °C. A last step of centrifugation (at 3,500 × g for 15 min) was performed and the acetone extracts (organic phase) were transferred to glass tubes where 100% DMSO was added and evaporated overnight under a nitrogen stream to 20% DMSO and a concentration of 2×WBE (Whole Broth Equivalent). 500 µL aliquots were prepared for each sample and stored at 4 °C overnight until bioactivity screening.

Bioactivity Screening Assays

Antimicrobial Activity Test

The antimicrobial activity of the crude extracts was assessed against a panel of five different pathogenic microorganisms available from the Culture Collection of Fundación MEDINA (Granada, Spain). Antibacterial activity assays were performed against the Gram-negative *Escherichia coli* ATCC 25922 and *Klebsiella pneumoniae* ATCC 700603 and the Gram-positive methicillin-resistant *Staphylococcus aureus* MRSA MB5393. Standard drugs for the antibacterial bioactivity assays were aztreonam for *E. coli* ATCCC 25922, gentamicin sulfate for *K. pneumoniae* ATCC 700603 and vancomycin for *S. aureus* MRSA MB5393. To assess the antifungal activity, the fungi *Candida albicans* ATCC 64124 and *Aspergillus fumigatus* ATCC 46645 were used. Amphotericin B was used as positive control for the antifungal bioactivity assays. All bioassays were conducted according to previously described methodologies (Monteiro et al., 2012; Audoin et al., 2013; Martin et al., 2013; Santos et al., 2020b). The final volume in the assay was 100 µL. No serial dilutions were used as the testing material was crude extract and not a pure compound. Incubation time was 18 h and temperature 37 °C. The assay was done in duplicate on different days. Absorbance (at 600 nm for bacterial strains and 612 nm for *C. albicans*) or fluorescence (excitation 570 nm, emission 615 nm for resazurin for *A. fumigatus*) were measured using an EnVision 2104 multilabel plate reader (PerkinElmer Inc., Waltham, MA, USA). Data analysis was conducted using the GeneData Screener® software (Genedata, Inc., Basel, Switzerland). Percentage of growth inhibition was calculated as reported by Martin et al. (2013).

Cell Viability Assay

The cytotoxic activity of the extracts was tested by MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assays against the following human cancer cell lines: lung carcinoma A549 (ATCC® CCL-185™), melanoma A2058 (ATCC® CRL-11147™), liver hepatocellular carcinoma HepG2 (ATCC® HB-8065™), breast adenocarcinoma MCF7 (ATCC®

HTB-22™) and pancreas carcinoma MiaPaca-2 (ATCC® CRL-1420™). All cell lines were sourced from the American Type Culture Collection (ATCC, Manassas, VA, USA). The cells were cultured in 96-well plates at a cell density of 10.000 cells/well and maintained at 37 °C, 90% humidity and 5% CO_2. A total of 5 µL of each extract dispensed in 195 µL of medium were used as inoculum for the anticancer assay. The plates were then incubated for 72 h at 37 °C, 90% humidity and 5% CO_2. Methyl methanesulfonate (MMS) 8 mM was used as positive control and 20% DMSO as negative control. A dose-response curve using doxorubicin (chemotherapeutic agent of natural origin) was also included as control, starting at 5 mM with an 8-point serial dilution (1/3). After 72 h treatment, the plates were washed with 100 µL of 1x PBS per well using a Multidrop Combi Reagent Dispenser (Thermo Scientific, Waltham, MA, United States). The MTT solution (tetrazolium dye, 0.5 mg/mL, final concentration 100 µL/well) was then added and the plates were incubated for approximately 3 h at 37 °C. Supernatants were discarded and 100 µL of 100% DMSO was added to each well to dissolve the formazan precipitates. Absorbance levels were measured at 570 nm using a Wallac 1420 Victor2™ Microplate Reader (WALLAC Oy PerkinElmer, Turku, Finland). The resulting data was analysed using the Genedata Screener® software 7.0 (Genedata, Inc., Basel, Switzerland). The test was performed in triplicate.

Genomic DNA Extraction and Whole Genome Sequencing

Overnight liquid bacterial cultures were generated by picking up the same colony used as template for Sanger sequencing and to inoculate the respective media in 50 mL tubes at 20 °C (Table S5). Genomic DNA was extracted from 1 mL of overnight cultures using the MasterPure Gram-Positive DNA Purification Kit (Lucigen, Epicentre). Concentration and purity of the extracted DNA were measured using a Nanodrop 2000c spectrophotometer (Thermo Fisher Scientific, USA) and a Qubit dsDNA BR Assay kit (Molecular Probes, Life Technologies) used with a DS-11 FX Fluorometer (DeNovix, Inc., Wilmington, DE, USA).

Whole genome sequencing of eight strains (Aa3_DN64_1D3, Aa3_Str.68_7G12, Acac_Ps_AB113, Cn_Ps_AB111, Irc_Ps_AB108, Pf1_DN206_4B7, Pf1_Ps_8H04_1, Pf1_Ps_8H06) was performed previously using the Illumina MiSeq platform (paired end, 2 x 300 bp) at GATC Biotech (Konstanz, Germany; now part of Eurofins Genomics Germany GmbH) in a previous study (Versluis et al., 2018). The remaining 13 strains were sequenced in the framework of this study with Illumina HiSeq (paired end, 2 x 150 bp reads) at Novogene Europe (Cambridge, UK).

Bioinformatic Analysis

Quality Control and Genome Assembly

The quality of the Illumina HiSeq reads was investigated with FASTQC 0.11.9 (Andrews, 2010). Adapter removal and read quality filtering was performed with Trimmomatic 0.39 (Bolger et al., 2014) with the following settings: ILLUMINACLIP:/Adapters.fa:2:30:10, HEADCROP:10, MINLEN:40, SLIDINGWINDOW:4:15. Illumina HiSeq reads were de novo assembled with SPAdes 3.14.0 (Bankevich et al., 2012), and contigs less than 1000 bp long were filtered out using reformat.sh from BBTools Suite 38.84 (Bushnell et al., 2017). All draft assemblies were checked for the presence of contigs assigned to sequencing artefacts which were subsequently discarded. Mapping of the quality-filtered reads to the assembled contigs with Bowtie 2.4.1 (Langmead and Salzberg, 2012) using default setting followed. The resulting sequence alignment map (SAM) file was converted into a binary alignment map (BAM) file, which was sorted and then indexed with SAMtools 1.10 (Li et al., 2009). Per base coverage of the draft assemblies was calculated with the 'genomecov' command of BEDTools 2.29.1 (Quinlan, 2014) using the sorted, indexed BAM file as input. Quality of all draft assemblies was assessed with QUAST 5.0.2 (Gurevich et al., 2013) and completeness and contamination were evaluated with CheckM 1.1.2 (Parks et al., 2015) with the default set of marker genes. All relevant information regarding the Illumina MiSeq reads processing and genome assembly are described in Versluis et al. (2018). Raw read sequences and draft genome assemblies generated in this study have been deposited in the European Nucleotide Archine (ENA) under the study accession number PRJEB41620.

Phylogenetic Analysis and Genome Mining

The GTDB-Tool Kit 1.1.0 (GTDB-Tk) (Chaumeil et al., 2019) was used to taxonomically classify all strains based on the presence of single-copy marker genes in their draft assemblies and the placement of their genomes in the Genome Taxonomy Database (GTDB) reference tree (Parks et al., 2018; Parks et al., 2020). The resulting concatenated alignment of the translated amino acid sequences of 120 bacterial marker genes identified in the draft assemblies was used to generate a maximum likelihood tree using FastTree 2.1.11 (Price et al., 2010) with default parameters. Tree visualisation was done with Interactive Tree of Life (iTOL) 3.0 (Letunic and Bork, 2021). For annotation of secondary metabolite Biosynthetic Gene Clusters (BGCs) in the draft assemblies, the online server antiSMASH 5.0 (Blin et al., 2019) was employed with "relaxed" detection strictness and all extra features activated.

Data Visualisation and Availability

Plots were created in R Studio with R 3.5.0 (R Core Team, 2020) using the R package ggplot2 3.3.2 (Wickham, 2016). All codes and data used for the genomic analysis and data visualisation can be found in https://github.com/mibwurrepo/Gavriilidou_et_al_2021_Bioactivity_Screening.

Results

Genome Characteristics

The genomes of 21 strains isolated from six different sponge species (*Aplysina aerophoba*, *Acanthella acuta*, *Corticium candelabrum*, *Chondrilla nucula*, *Ircinia* sp. and *Petrosia ficiformis*) in previous studies (Versluis et al., 2017b; Versluis et al., 2018), were analysed here. Eight genomes (Aa3_DN64_1D3, Aa3_Str.68_7G12, Acac_Ps_AB113, Cn_Ps_AB111, Irc_Ps_AB108, Pf1_DN206_4B7, Pf1_Ps_8H04_1, Pf1_Ps_8H06) were generated by Versluis et al. (2018) and the rest were produced in this study (Table 1). The average number of contigs was 46 and the average coverage per base of the draft assemblies was 219x. The genome size ranged from 2.6 to 7.2 Mbp with a GC content between 32.9 and 72.9% and the average number of genes being 4,654. Genome completeness was 98.7% on average and contamination was less than 1.5% in all cases. More detailed information on the genome assembly metrics can be found in Table S1.

Table 1. Information on taxonomy of sponge-associated isolates according to BLAST searches of their partial 16S rRNA gene sequences against nr/nt NCBI database and genome characteristics. All details on the classification of strains and genome assembly statistics are provided in the supplementary file (Table S1 and S2). Order of strains is in accordance with the phylogenetic tree in Figure 1. Genomes of marked strains were generated in a previous study (Versluis et al., 2018).

Strain ID	Isolation Source	Best BLAST hit	ID%	Genome size (Mbp)	GC content (%)	Total gene count
Aa3_DN55_6A7	*Aplysina aerophoba*	*Bradyrhizobium* sp.	100	7.2	64.6	6,655
Pf1_Ps_8H04_1 [1]	*Petrosia ficiformis*	*Pseudovibrio* sp.	99.9	5.7	48.2	5,174
Pf1_DN206_4B7 [1]	*Petrosia ficiformis*	*Pseudovibrio* sp.	99.85	5.1	52.8	4,692
Irc_Ps_AB108 [1]	*Ircinia* sp.	*Pseudovibrio* sp.	100	5.9	44.6	5,369
Pf1_DN64_8G1	*Petrosia ficiformis*	*Pseudovibrio* sp.	99.93	5.8	51.4	5,275
Aa3_DN64_1D3 [1]	*Aplysina aerophoba*	*Pseudovibrio* sp.	99.93	5.8	50.3	5,239
Pf1_Ps_8H06 [1]	*Petrosia ficiformis*	*Pseudovibrio* sp.	100	6.1	49.7	5,554
Cn_Ps_AB111 [1]	*Chondrilla nucula*	*Pseudovibrio* sp.	99.89	5.9	49.8	5,423
Aa3_Str.68_7G12 [1]	*Aplysina aerophoba*	*Pseudovibrio* sp.	100	5.9	51.0	5,288
Acac_Ps_AB113 [1]	*Acanthella acuta*	*Pseudovibrio* sp.	99.93	5.4	51.0	4,886
Aa3_DN166_3E9_2	*Aplysina aerophoba*	*Ruegeria* sp.	99.93	4.6	56.2	4,558
Pf1_DN81_6F7_2	*Petrosia ficiformis*	*Ruegeria atlantica*	100	4.5	57.9	4,356
Cc1_DN217_4H2	*Corticium candelabrum*	*Microbulbifer echini*	99.85	4.7	49.8	4,229
Aa3_DN138_5C8	*Aplysina aerophoba*	*Acinetobacter radioresistens*	100	3.3	41.4	3,085
Pf1_DN14_7A9_1	*Petrosia ficiformis*	*Psychrobacter celer*	100	2.9	46.8	2,434
Aa3_DN73_5E10	*Aplysina aerophoba*	*Psychrobacter celer*	100	2.6	47.0	2,194
Aa3_DN30_1H2	*Aplysina aerophoba*	*Aquimarina macrocephali*	100	5.4	32.9	4,684
Aa3_DN216_4B10_1	*Aplysina aerophoba*	*Rhodococcus erythropolis*	100	7.1	62.5	6,770
Aa3_DN213_3F7	*Aplysina aerophoba*	*Brevibacterium aurantiacum*	100	4.2	63.0	3,850
Aa3_DN216_4B10_2	*Aplysina aerophoba*	*Janibacter melonis*	99.93	3.4	72.9	3,220
Aa3_DN71_7G3_2	*Aplysina aerophoba*	*Bacillus frigoritolerans*	100	4.9	40.5	4,807

[1] Modified from Versluis et al. (2018).

Strain Identification and Phylogeny

All strains analysed in this study were identified based on both 16S rRNA gene sequences and single-copy marker gene analysis. 16S rRNA gene-based taxonomic assignment of the sponge-associated strains according to the NCBI database was confirmed by the Genome Taxonomy Database (GTDB) (Table 1 and S2) (Parks et al., 2020). Four main clades were formed representing the following phyla: *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Firmicutes* (Figure 1).

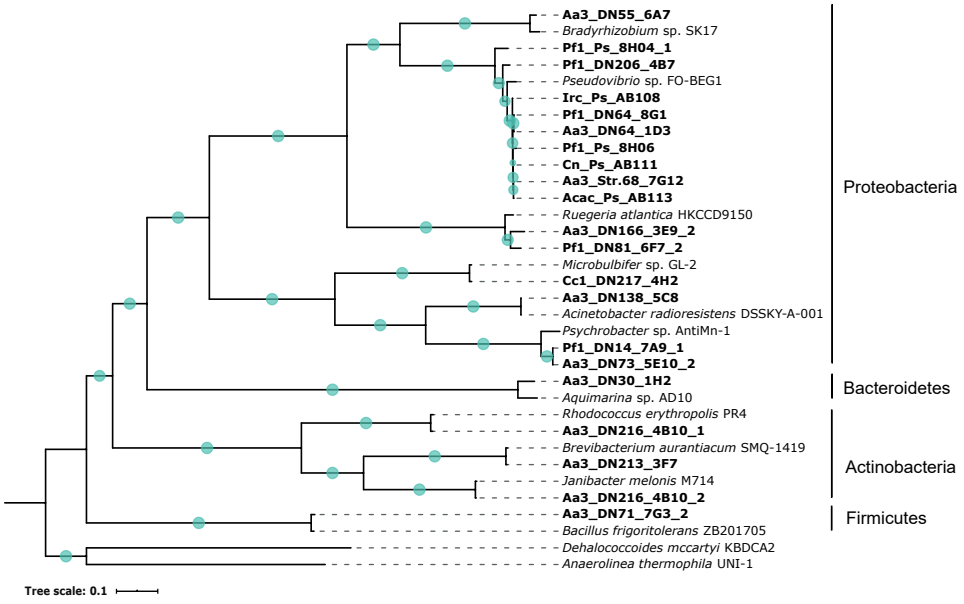


Figure 1. Maximum likelihood tree inferred from a concatenated alignment of 120 conserved amino-acid sequences of bacterial marker genes. Circles on the branches display bootstrap values (>70%). Strains included in the present study are indicated in bold. *Dehalococcoides mccartyi* KBDCA2 and *Anaerolinea thermophila* UNI-1, both members of the *Chloroflexi*, were used as outgroup. The scale bar represents the number of estimated substitutions per site.

Out of the 21 strains, 16 strains belonged to *Proteobacteria*, three strains were classified as *Actinobacteria* and one strain each represented *Bacteroidetes* and *Firmicutes*, respectively. Within the *Proteobacteria*, there were strains most closely related to *Pseudovibrio*, *Bradyrhizobium*, *Ruegeria*, *Microbulbifer*, *Acinetobacter* and *Psychrobacter* strains. In the case of *Actinobacteria*, members of the genera *Rhodococcus*, *Brevibacterium* and *Janibacter* were identified. Moreover, two strains were affiliated with *Aquimarina* and *Bacillus* genera, respectively (Figure 1).

Antimicrobial and Anticancer Activity Screening

Crude extracts of axenic bacterial cultures were tested for their antibacterial activity against a panel of Gram-positive and Gram-negative bacteria (*E. coli*, *K. pneumoniae* and *S. aureus* MRSA). Only one strain (Aa3_DN216_4B10_1) showed significant growth inhibition (61%) of *S. aureus* MRSA. According to both 16S rRNA gene- and whole genome-based taxonomy, the strain Aa3_DN216_4B10_1 was closely related to *Rhodococcus erythropolis*, member of the *Actinobacteria* (Figure 1). The antifungal activity of the extracts was assessed against *C. albicans* and *A. fumigatus*. Growth inhibition of the fungi was not observed for any of the extracts under the tested conditions (data not shown).

The cytotoxicity of the bacterial crude extracts was determined on human skin (A2058), lung (A549), liver (HepG2), breast (MCF7) and pancreas (MiaPaca2) cancer cells. In general, the majority of extracts (76.2%) were mostly effective against HepG2 cells exhibiting moderate to high cytotoxicity (Figure 2). In addition, almost 20% of the extracts resulted in cell death of more than 50% of the A2058, A549 and MiaPaca2 cancer cells, while very weak activity was observed against the MCF7 cell line. Four extracts (Pf1_DN64_8G1, Aa3_DN64_1D3, Aa3_DN73_5E10_2 and Aa3_DN213_3F7) had the highest activity, causing more than 50% of cell death of at least two of the cancer cell lines (Figure 2). These were obtained from isolates belonging to *Proteobacteria* and *Actinobacteria* and more specifically, to the genera *Pseudovibrio*, *Psychrobacter* and *Brevibacterium* (Figure 1). No correlation was observed between phylogenetic proximity of tested strains and their levels of cytotoxicity, except for Aa3_DN64_1D3 and Pf1_DN64_8G1, which both belonged to *Pseudovibrio* and were highly active (Figures 1 and 2).

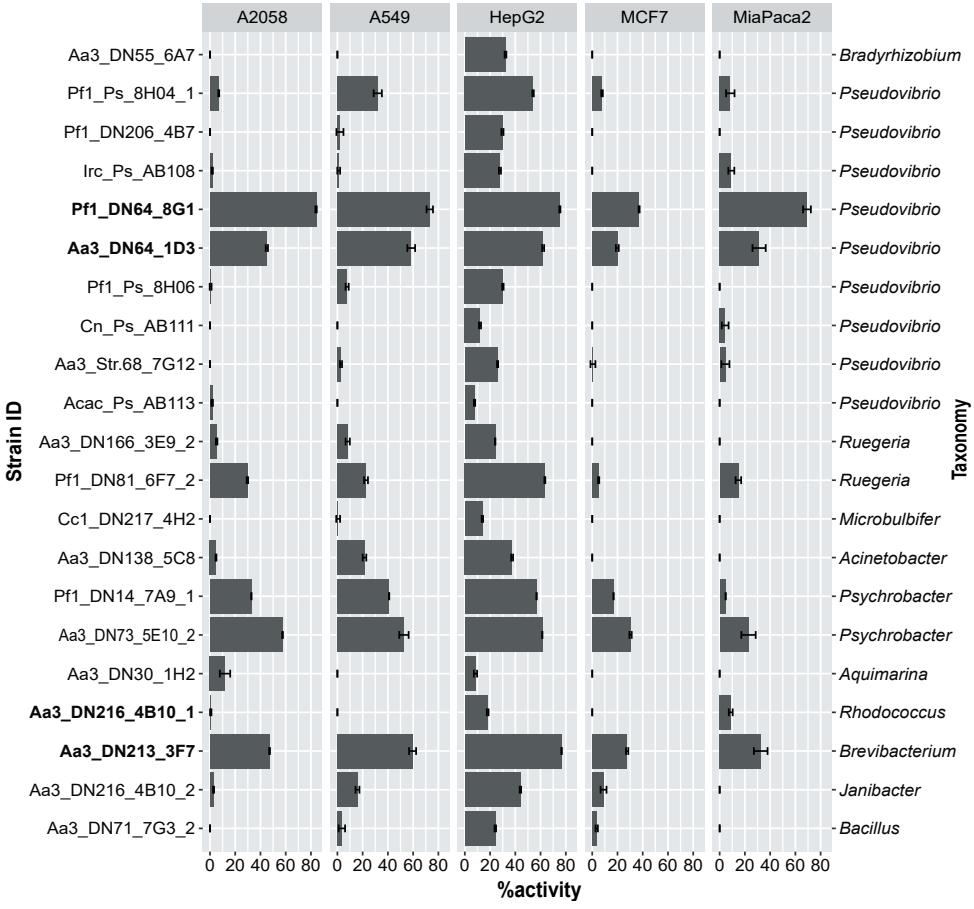


Figure 2. MTT [3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide] assay results. Percentage of activity or cell death of five human cancer cell lines (A2058: melanoma, A549: lung carcinoma, HepG2: hepatocyte carcinoma, MCF7: breast adenocarcinoma and MiaPaca2: pancreas carcinoma) after 72 h of incubation with crude extracts of the tested strains in triplicate. Error bars represent standard errors. Strains are ordered from top to bottom in accordance with the phylogenetic tree in Figure 1 and in bold are the ones selected for gene-trait matching.

Biosynthetic Gene Cluster Profiling

Genomes of the isolates were mined for BGCs, and a total of 153 BGCs belonging to 28 different BGC types as classified by antiSMASH (antibiotics & Secondary Metabolite Analysis Shell) (Blin et al., 2019), were identified. In terms of absolute abundance, the majority of identified BGCs was predicted to encode bacteriocins (n=24), non-ribosomal peptide synthetases (NRPS) (n=19), beta-lactones (n=18), terpenes (n=18) and siderophores (n=10). Similarly, the same categories displayed the highest frequency being detected in almost 50% of the genomes (Figure 3).

Figure 3. Absolute abundance of secondary metabolite biosynthetic gene clusters (BGCs) predicted in the genomes of the strains screened for their bioactivity in the present study. Strains selected for gene-trait matching are marked in bold.

Considering both the genome size and the abundance of BGCs (Table S3), the strains with the highest secondary metabolite biosynthesis potential were the following: Aa3_DN216_4B10_1, Aa3_DN213_3F7, Aa3_DN64_1D3, Pf1_DN64_8G1 and, Aa3_DN30_1H2, all from the *Actinobacteria*, *Proteobacteria* or *Firmicutes*.

Gene-Trait Matching

In silico prediction of the secondary metabolite BGCs was combined with the experimental data obtained from the bioactivity screening assays for Gene-Trait Matching (GTM) aiming at the correlation of genetic features with specific phenotypes. Strains that showed both high secondary metabolite biosynthetic potential and in vitro bioactivity were selected for further analysis.

The antimicrobial activity bioassays showed that from all tested strains, Aa3_DN216_4B10_1 was the only one with antibacterial activity, namely against MRSA. In fact, this strain was

found to harbour the highest number of BGCs and ten-fold more NRPS/NRPS-like BGCs compared to the rest of the isolates (Figure 3 and Table S3). Given the relatively large number of contigs and NRPS/NRPS-like BGCs, further inspection of the BGC regions showed that four out of ten were located at a contig edge. Even though this might indicate the presence of fragmented BGCs across multiple contigs (Blin et al., 2021a), strain Aa3_DN216_4B10_1 showed the highest abundance of NRPS/NRPS-like BGCs among all tested ones. Additionally, the KnownClusterBlast tool of antiSMASH identified three BGCs (BGC 16, BGC 4 and BGC 3) with 100% similarity to known clusters in the Minimum Information about a Biosynthetic Gene (MIBiG) database related to the production of heterobactin A/heterobactin S2 (MIBiG:BGC0000371), rhizomide A (MIBiG:BGC0001758) and branched-chain fatty acids (BCFAs) (MIBiG:BGC0001534) (Table S4). The structure of the aforementioned BGCs are displayed in Figure S1. Based on the antiSMASH analysis, the NRPS cluster predicted to code for heterobactin A/heterobactin S2 consisted of 11 genes responsible for core and additional biosynthesis and transport. It had a well-defined modular structure consisting of all key NRPS components such as adenylation, condensation, thiolation and certain tailoring domains. The NRPS-like BGC potentially encoding the lipopeptide rhizomide A was located at the edge of the contig and contained an adenylation and a thiolation domain. In the case of BGC 3, antiSMASH classified it as bacteriocin type with a predicted core structure for the production of bacteriocins next to the machinery for synthesising BCFAs. Specifically, genes belonging to BGC 3 were 100% similar to the known bkd BGC in *Streptomyces filamentosus* (MIBiG:BGC0001534). Other BGCs homologous to those coding for known antimicrobials, such as chloramphenicol (MIBiG:BGC0000893), kirromycin (MIBiG:BGC0001070) and bacillomycin D (MIBiG:BGC0001090), were also detected in its genome but with lower similarity (<20%) (Table S4).

Comparative analysis of the 21 marine bacterial genomes in terms of their BGCs showed that three strains of the top five with the largest number of BGCs, also had the highest activity against the tested cancer cell lines. According to the experimental results, Pf1_DN64_8G1, Aa3_DN213_3F7 and Aa3_DN64_1D3 induced cell death to all five cancer cell lines with average cytotoxicity of 67.8%, 48.7% and 43.3%, respectively. Pf1_DN64_8G1 and Aa3_DN64_1D3 harboured ten BGCs each, from which only three BGCs had homologs encoding known natural products (Table S4). Two of these clusters were identified as beta-lactone BGCs and showed low similarity to those coding for pseudaminic acid (MIBiG:BGC0001747) (22%) and fengycin (MIBiG:BGC0001095) (13%), compounds known for their cytotoxic activity (Ma et al., 2012; Yin et al., 2013; Kokoulin et al., 2020). On the other hand, seven out of nine BGCs of Aa3_DN213_3F7 gave hits to the MIBiG repository (Table S4). Similarly, four BGCs were related to clusters involved in the production of substances with reported anticancer activity: carotenoids (MIBiG:BGC0000636) (85% similarity) (Chuyen and Eun, 2017; Galasso et al., 2017), ectoines (MIBiG:BGC0000853) (75% similarity) (Sheikhpour et al., 2019) and the siderophore desferrioxamine E (MIBiG:BGC0001478) (50% similarity) (Kalinovskaya et al., 2011) (Table S4).

Discussion

Sponge-associated microbes have already been acknowledged as important members of the 'Natural Products Hall of Fame' for the production of substances with various bioactivities (Brinkmann et al., 2017b). Nevertheless, the majority remains recalcitrant to in vitro cultivation hindering the flow in the marine drug biodiscovery pipeline (Gutleben et al., 2017). Intensive research has focused on screening methods and subsequent elucidation of the chemical molecules. To shed more light on the bioactive potential of microbes, integration of genome mining with activity-driven screenings could uncover the true metabolite arsenal of the microbes of interest (Reen et al., 2015).

In the present study, we initially obtained crude extracts from 21 bacterial strains isolated from marine sponges and determined their antibacterial and cytotoxic activity via high-throughput screening. 16S rRNA gene- and whole genome-based phylogenetic analysis revealed that the tested strains belong mainly to four phyla, *Proteobacteria*, *Actinobacteria*, *Bacteroidetes* and *Firmicutes* (Figure 1). Several of these taxa have been highlighted before for their antimicrobial activity. *Actinobacteria* and *Proteobacteria* are considered the most prolific sources of bioactive secondary metabolites with a wide spectrum of bioactivities among marine microorganisms (Desriac et al., 2013; Graca et al., 2013; Brinkmann et al., 2017a; Khalifa et al., 2019; Lee et al., 2020). However, only one strain, identified as *Rhodococcus erythropolis*, possessed anti-MRSA potential. To the best of our knowledge, no anti-MRSA activity has been previously detected in *Rhodococci* derived from marine sponges. Chelossi et al. (2004) and Abdelmohsen et al. (2014) retrieved *Rhodococcus* strains from *Petrosia ficiformis* and an unidentified marine sponge, respectively, with the ability to inhibit growth of *S. aureus* but not the methicillin-resistant strain. A possible explanation for the general absence of antimicrobial activity of the strains tested here may be the use of crude extracts as testing material and thus, the active principles being present in low concentrations. Moreover, eight of the *Pseudovibrio* strains screened here were selected because they were previously found to be resistant against several antibiotics, such as ampicillin, vancomycin and tetracycline (Versluis et al., 2018). Yet, these antibiotic resistance phenotypes did not match with the antimicrobial activity screening results which could be due to a lack of induction of gene expression under the tested conditions. Diverse fermentation conditions using various nutrients and treatments (e.g. co-culturing) may be of great value in activating silent genes in the search for bioactive secondary metabolites (Reen et al., 2015; Brinkmann et al., 2017a).

The high potential of marine sponge derivatives to inhibit tumour proliferation has led to increasing research efforts toward the discovery of new anticancer compounds (Calcabrini et al., 2017; Zhang et al., 2017; Cetkovic et al., 2018). To date, more than 10% of the screened marine sponges display cytotoxicity against human cancer cell lines (Zhang et al., 2017). Here, the effect of the crude extracts was more evident on HepG2 cells compared to other cancer cell lines (Figure 2). Among the tested strains, we could distinguish four

extracts obtained from strains belonging to the genera *Pseudovibrio*, *Psychrobacter* and *Brevibacterium* that showed the highest activity mainly against A2058, A549 and MiaPaca2 cancer cells. The inhibitory effect on A549 cells by *Pseudovibrio* has previously been reported, and the responsible bioactive natural products were indole alkaloids identified in cultures of *Pseudovibrio denitrificans* (Rodrigues et al., 2017; Romano, 2018). On the other hand, Choi *et al.* Choi et al. (2009) reported *Brevibacterium*-derived compounds causing no induction of cell death on A549, AGC, MCF-7 and HepG2 carcinomas but only weak cytotoxicity against HL-60 cells.

To establish a link between the secondary metabolite BGC repertoire and the growth inhibiting effect against pathogenic bacteria and cancer cell lines, laboratory-based screening methods combined with genome mining were performed followed by GTM analysis. The top five strains in terms of BGC frequency and abundance consisted of three *Proteobacteria* and two *Actinobacteria* strains, confirming their strong bioactivity potential observed before (Naughton et al., 2017; Versluis et al., 2018; Xu et al., 2019; Guerrero-Garzón et al., 2020). Nevertheless, the abundance and diversity of BGCs were not reflected in the phenotype of the strains as only one displayed antibacterial activity under the tested conditions. Anti-MRSA activity shown by the *Rhodococcus* strain (Aa3_DN216_4B10_1) could be explained by the distinctively higher number of NPRS-encoding BGCs in its genome indicating a large specialised secondary metabolite arsenal (Doroghazi and Metcalf, 2013; Ceniceros et al., 2017; Thompson et al., 2020). Two NRPS BGCs had 100% homology with those coding for the siderophore heterobactin A/heterobactin S2 and the lipopeptide rhizomide A, respectively. Siderophores are small molecules that beyond iron acquisition were recently suggested to act as virulence factors and regulators of pathogenicity (Behnsen and Raffatellu, 2016). In addition, Schneider et al. (2020) isolated a siderophore from the co-culture of two marine *Proteobacteria* which displayed species-specific toxicity towards *S. aureus* without affecting MRSA. Likewise, in the case of rhizomide A, antibacterial activity has been observed against several clinically relevant strains, including *S. aureus* but not MRSA (Wang et al., 2018). Another potential candidate responsible for the anti-MRSA activity of the *Rhodococcus* strain might be a putative peptide that contains branched-chain fatty acyl groups encoded by BGC 3 (Figure S1). Part of this BGC is 100% homologous to the bkd BGC in *Streptomyces filamentosus* which encodes the branched-chain α-keto acid dehydrogenase (BKDH), a multi-subunit enzyme complex critical in the synthesis of BCFAs. These fatty acids could serve as precursors in the biosynthesis of antibiotics, such as daptomycin (Luo et al., 2018) which is currently one of the main treatment options for MRSA infections (Kanafani and Corey, 2007; Roch et al., 2017).

Interestingly, the experiments with human cancer cells showed that three out of the five strains with the highest cytotoxicity harboured the largest number of BGCs, suggesting a genotype-phenotype causality. Specifically, the most bioactive extracts belonged to *Pseudovibrio* strains (Pf1_DN64_8G1 and Aa3_DN64_1D3) which carried in their genomes two beta-lactone BGCs similar to those coding for natural products with known anticancer

effect, namely pseudaminic acid and fengycin. Kokoulin et al. (2020) were the first to report a capsular polysaccharide containing pseudaminic acid from another proteobacterium (*Psychrobacter marincola*), which significantly inhibited the growth of HL-60 cells. Fengycins are biosurfactants that have been studied in detail for their bioactivities and considered to be important inhibitors in cancer research (Cheng et al., 2016). Yet, the only marine-derived fengycin with anticancer potential is a fengycin isoform isolated from *Bacillus circulans* with a high efficacy against human colon carcinoma cells (Sivapathasekaran et al., 2010). However, the similarity of the BGCs with those of known activity was relatively low, 22% and 13% for the pseudaminic acid and the fengycin BGC, respectively. This could be explained by the fact that the majority of BGCs detected in the two *Pseudovibrio* genomes did not share any similarity with characterised BGCs highlighting their unknown function in accordance with previous studies (Naughton et al., 2017; Versluis et al., 2018). On the other hand, most BGCs of Aa3_DN213_3F7 had homologs with known bioactivity that confirms the extensive research focus on *Actinobacteria* in terms of their natural products (Graca et al., 2013). The BGCs in the genome of strain Aa3_DN213_3F7 with the highest similarity were related to BGCs coding for carotenoids, ectoines and the siderophore desferrioxamine E. Based on our experimental results, the highest cytotoxicity of Aa3_DN213_3F7 was observed against skin, lung and hepatic carcinomas. Several studies have described the antiproliferative effect of marine carotenoids on different cancer cell lines, such as breast, intestinal, hepatic and leukemic (Chuyen and Eun, 2017). The carotenoid BGC was 85% similar to a BGC identified in *Brevibacterium linens* that encodes a novel lycopene cyclase that catalyses the biosynthesis of β-carotene from lycopene (Krubasik and Sandmann, 2000). B-carotene is a successful carotenoid in the global market with many industrial applications, including prevention of cancer (Galasso et al., 2017). In the case of ectoines, Sheikhpour et al. (2019) showed that ectoine and hydroxyectoine isolated from *Streptomyces* induced apoptosis in lung cancer cells. Another compound involved in the cytotoxicity of Aa3_DN213_3F7 could be related to desferrioxamine E, a synonym for nocardamine. It is a cyclic hydroxamic acid siderophore commercially available as antibiotic, anti-mycobacterial, iron-chelating, antioxidant and anticancer compound. According to previous studies, nocardamine isolated from a novel marine actinobacterium exhibited only antitumor effect but no cytotoxicity against human breast cancer and malignant melanoma cell lines (Kalinovskaya et al., 2011). Nevertheless, cancer cell toxicity high-throughput screening against the Canvass library of natural products revealed the antiproliferative effect of nocardamine on various cancer cell lines, including pancreatic and ovarian (Kearney et al., 2018).

Conclusions

The bioactivity screening and subsequent phylogenetic analysis of the sponge-associated bacteria revealed that the most active extracts both in terms of antibacterial and anticancer activity were affiliated with the phyla *Actinobacteria* and *Proteobacteria*, supporting these two taxonomic groups as prominent sources of bioactive substances. Yet, the antiproliferative impact on human pathogens was less pronounced than that on cancer cells. This is in contrast with the abundance of BGCs detected in the genomes and points out the need of implementing diverse cultivation regimes to trigger the expression of the appropriate genes linked to the production of bioactive molecules. On the other hand, the most cytotoxic strains were among the ones with the highest number of BGCs. GTM analysis revealed several BGCs related to compounds potentially responsible for the induction of cell death of the respective cancer cell lines, facilitating the distinction of favourable candidates. Altogether, these findings highlight the importance of integrating phenotypic assays with genome mining in order to provide insights on the most promising leads for further investigation, such as isolation and identification of the active principles.

Acknowledgements

We thank Fernando Reyes, Carlos Justicia, Angeles Melguizo, Mercedes de la Cruz and Victor González-Menéndez for the technical support (Fundación MEDINA). We thank Jutta Wiese and Johannes Imhoff (GEOMAR) for the bacterial isolates from *Acanthella acuta*, *Chondrilla nucula* and *Ircinia* sp..

Supplementary Information

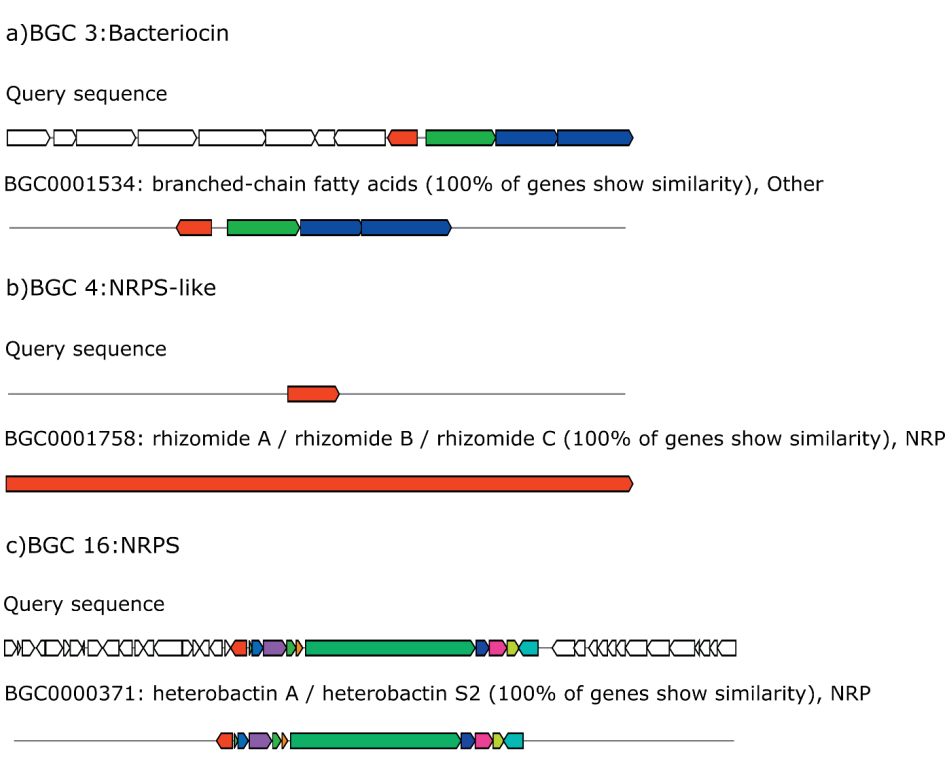


Figure S1. KnownClusterBlast output of biosynthetic gene clusters (BGCs) predicted in strain Aa3_DN216_4B10_1 showing 100% homology to reference BGCs in the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database (Medema et al., 2015). Colours indicate BLAST matches of individual genes between query and reference sequences.

Table S1. Genome assembly metrics of strains included in the present study.

Strain ID	Isolation Source	Contigs	Genome size (Mb)	GC content (%)	N50 (Mb)	Completeness (%)	Contamination (%)	Coverage (X)	Total gene count
Aa3_DN55_6A7	*Aplysina aerophoba*	36	7.2	64.6	0.4	100.0	0.1	31	6655
Pf1_Ps_8H04_1	*Petrosia ficiformis*	18	5.7	48.2	0.5	100.0	0.2	295*	5174
Pf1_DN206_4B7	*Petrosia ficiformis*	52	5.1	52.8	0.4	100	0.8	226*	4692
Irc_Ps_AB108	*Ircinia* sp.	49	5.9	44.6	0.4	99.6	0.0	200*	5369
Pf1_DN64_8G1	*Petrosia ficiformis*	30	5.8	51.4	0.6	99.9	0.3	206	5275
Aa3_DN64_1D3	*Aplysina aerophoba*	22	5.8	50.3	0.5	99.6	0.3	711*	5239
Pf1_Ps_8H06	*Petrosia ficiformis*	40	6.1	49.7	0.4	99.6	0.6	264*	5554
Cn_Ps_AB111	*Chondrilla nucula*	36	5.9	49.8	0.4	99.9	0.0	228*	5423
Aa3_Str.68_7G12	*Aplysina aerophoba*	29	5.9	51.0	0.5	99.7	0.3	138*	5288
Acac_Ps_AB113	*Acanthella acuta*	13	5.4	51.0	0.9	99.6	0.0	191*	4886
Aa3_DN166_3E9_2	*Aplysina aerophoba*	40	4.6	56.2	0.2	99.1	0.0	237	4558
Pf1_DN81_6F7_2	*Petrosia ficiformis*	30	4.5	57.9	0.3	99.9	0.0	165	4356
Cc1_DN217_4H2	*Corticium candelabrum*	108	4.7	49.8	0.1	99.6	1.1	317	4229
Aa3_DN138_5C8	*Aplysina aerophoba*	46	3.3	41.4	0.2	100.0	0.0	310	3085
Pf1_DN14_7A9_1	*Petrosia ficiformis*	51	2.9	46.8	0.1	97.7	0.3	113	2434
Aa3_DN73_5E10_2	*Aplysina aerophoba*	41	2.6	47.0	0.1	92.2	0.3	385	2194
Aa3_DN30_1H2	*Aplysina aerophoba*	39	5.4	32.9	0.2	97.3	1.5	86	4684
Aa3_DN216_4B10_1	*Aplysina aerophoba*	79	7.1	62.5	0.3	99.9	1.1	43	6770
Aa3_DN213_3F7	*Aplysina aerophoba*	60	4.2	63.0	0.2	99.4	0.0	300	3850
Aa3_DN216_4B10_2	*Aplysina aerophoba*	19	3.4	72.9	0.3	99.8	0.0	51	3220
Aa3_DN71_7G3_2	*Aplysina aerophoba*	65	4.9	40.5	0.2	90.4	1.4	112	4807

*adapted from Versluis et al. (2018)

Table S2. Taxonomic classification of studied strains based on 16S rRNA gene sequences (nr/nt database) and marker genes (GTDB database).

Strain ID	16S rRNA gene length (bp)	Best BLAST hit (accession number)	ID%	E-value	GTDB-Tk classification
Aa3_DN55_6A7	1340	*Bradyrhizobium* sp. LM6 (KX774628.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Xanthobacteraceae;g__Bradyrhizobium;s__Bradyrhizobium sp003020075
Pf1_Ps_8H04_1	988	*Pseudovibrio* sp. MaPt6 (JX436420.1)	99.9	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__
Pf1_DN206_4B7	1337	*Pseudovibrio* sp. MA_AMC_33 (MN703974.1)	99.85	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__
Irc_Ps_AB108	1342	*Pseudovibrio* sp. 2011SOCNI15 (KF582860.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Pf1_DN64_8G1	1335	*Pseudovibrio* sp. 2011SOCNI15 (KF582860.1)	99.93	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Aa3_DN64_1D3	1335	*Pseudovibrio* sp. 2011SOCNI15 (KF582860.1)	99.93	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Pf1_Ps_8H06	1325	*Pseudovibrio* sp. 2011SOCNI42 (KF582882.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Cn_Ps_AB111	906	*Pseudovibrio* sp. 2011SOCNI15 (KF582860.1)	99.89	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Aa3_Str.68_7G12	1325	*Pseudovibrio* sp. 2011SOCNI42 (KF582882.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Acac_Ps_AB113	1341	*Pseudovibrio* sp. ESS-18 (MH057247.1)	99.93	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales;f__Stappiaceae;g__Pseudovibrio;s__Pseudovibrio sp900143565
Aa3_DN166_3E9_2	1316	*Ruegeria* sp. 70077 (KX833139.1)	99.93	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Ruegeria;s__
Pf1_DN81_6F7_2	991	*Ruegeria atlantica* DN83_2B6 (KP769432.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Alphaproteobacteria;o__Rhodobacterales;f__Rhodobacteraceae;g__Ruegeria;s__

Table S2. Continued.

Strain ID	16S rRNA gene length (bp)	Best BLAST hit (accession number)	ID%	E-value	GTDB-Tk classification
Cc1_DN217_4H2	1377	*Microbulbifer echini* ROA029 (MT510173.1)	99.85	0.0	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Cellvibrionaceae;g__Microbulbifer;s__
Aa3_DN138_5C8	1396	*Acinetobacter radioresistens* OsEp_Plm_15B15 (MT367790.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Acinetobacter;s__Acinetobacter radioresistens
Pf1_DN14_7A9_1	1373	*Psychrobacter celer* 7A3 (KU525106.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Psychrobacter;s__Psychrobacter sp002810365
Aa3_DN73_5E10_2	1385	*Psychrobacter celer* G205M1 (MH256047.1)	100.0	0.0	d__Bacteria;p__Proteobacteria;c__Gammaproteobacteria;o__Pseudomonadales;f__Moraxellaceae;g__Psychrobacter;s__Psychrobacter sp002810365
Aa3_DN30_1H2	1044	*Aquimarina macrocephali* XH119 (KC178950.1)	100.0	0.0	d__Bacteria;p__Bacteroidota;c__Bacteroidia;o__Flavobacteriales;f__Flavobacteriaceae;g__Aquimarina;s__Aquimarina megaterium
Aa3_DN216_4B10_1	1068	*Rhodococcus erythropolis* KB1 (CP050124.1)	100.0	0.0	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Mycobacteriales;f__Mycobacteriaceae;g__Rhodococcus;s__Rhodococcus erythropolis_D
Aa3_DN213_3F7	1368	*Brevibacterium aurantiacum* SMQ-1419 (CP025333.1)	100.0	0.0	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Actinomycetales;f__Brevibacteriaceae;g__Brevibacterium;s__Brevibacterium aurantiacum
Aa3_DN216_4B10_2	1360	*Janibacter melonis* M0604 (KF924217.1)	99.93	0.0	d__Bacteria;p__Actinobacteriota;c__Actinobacteria;o__Actinomycetales;f__Dermatophilaceae;g__Janibacter;s__Janibacter melonis
Aa3_DN71_7G3_2	1397	*Bacillus frigoritolerans* ZB201705 (CP030063.1)	100.0	0.0	d__Bacteria;p__Firmicutes;c__Bacilli;o__Bacillales;f__Bacillaceae_A;g__Bacillus_X;s__Bacillus_X frigoritolerans

Table S3. Relative abundance of biosynthetic gene cluster (BGC) types identified by antiSMASH (Blin et al., 2019). Numbers were normalized according to the genome size. NRPS; non-ribosomal peptide synthetase.

Strain ID	BGC types						Total BGCs
	Bacteriocin	NRPS	Terpene	Betalactone	Siderophore	Others	
Aa3_DN55_6A7	0.3	0.3	0.0	0.3	0.0	0.0	0.7
Pf1_Ps_8H04_1	0.2	0.0	0.2	0.2	0.0	0.4	0.9
Pf1_DN206_4B7	0.2	0.2	0.2	0.2	0.0	0.4	1.2
Irc_Ps_AB108	0.3	0.2	0.2	0.2	0.2	0.5	1.5
Pf1_DN64_8G1	0.3	0.2	0.2	0.3	0.2	0.7	1.9
Aa3_DN64_1D3	0.3	0.2	0.2	0.3	0.2	0.5	1.7
Pf1_Ps_8H06	0.2	0.2	0.2	0.2	0.0	0.7	1.3
Cn_Ps_AB111	0.3	0.0	0.2	0.2	0.0	0.3	1.0
Aa3_Str.68_7G12	0.2	0.2	0.2	0.2	0.2	0.7	1.5
Acac_Ps_AB113	0.4	0.0	0.2	0.2	0.2	0.4	1.3
Aa3_DN166_3E9_2	0.2	0.0	0.0	0.2	0.0	1.1	1.5
Pf1_DN81_6F7_2	0.2	0.0	0.0	0.2	0.0	0.9	1.3
Cc1_DN217_4H2	0.2	0.0	0.0	0.0	0.2	0.9	1.3
Aa3_DN138_5C8	0.0	0.0	0.0	0.6	0.3	0.3	1.2
Pf1_DN14_7A9_1	0.0	0.0	0.0	0.3	0.0	0.0	0.3
Aa3_DN73_5E10_2	0.0	0.0	0.0	0.4	0.0	0.0	0.4
Aa3_DN30_1H2	0.6	0.4	0.4	0.0	0.0	0.4	1.7
Aa3_DN216_4B10_1	0.1	1.0	0.1	0.0	0.0	1.5	2.8
Aa3_DN213_3F7	0.2	0.7	0.2	0.0	0.2	0.7	2.1
Aa3_DN216_4B10_2	0.0	0.0	0.3	0.0	0.3	0.6	1.2
Aa3_DN71_7G3_2	0.0	0.2	0.4	0.2	0.2	0.4	1.4

Table S4. Information on BGCs detected by antiSMASH (Blin et al., 2019) of strains selected for gene-trait matching and comparison to the Minimum Information about a Biosynthetic Gene cluster (MIBiG) database (Medema et al., 2015).

Strain ID	BGC	BGC type	Most similar known (MIBiG)	Similarity	MIBiG accession (Medema et al., 2015)
Aa3_DN216_4B10_1	1	NRPS, terpene	SF2575	6%	BGC0000269
	2	NRPS	coelichelin	27%	BGC0000325
	3	bacteriocin	branched-chain fatty acids	100%	BGC0001534
	4	NRPS-like	rhizomide A/rhizomide B/ rhizomide C	100%	BGC0001758
	5	NRPS	bacillomycin D	20%	BGC0001090
	6	NRPS	-	-	-
	7	NRPS	rifamorpholine A/rifamorpholine B/rifamorpholine C/rifamorpholine D/ rifamorpholine E	4%	BGC0001759
	8	ectoine	ectoine	75%	BGC0000853
	9	NRPS	erythrochelin	57%	BGC0000349
	10	lanthipeptide	-	-	-
	11	PKS-like, amglyccycl	acarbose	7%	BGC0000691
	12	butyrolactone	-	-	-
	13	NRPS-like	-	-	-
	14	NRPS	chloramphenicol	17%	BGC0000893
	15	NRPS-like	thiotulin	8%	BGC0001193
	16	NRPS	heterobactin A/heterobactin S2	100%	BGC0000371
	17	LAP	diisonitrile antibiotic SF2768	11%	BGC0001574
	18	T1PKS	-	-	-
	19	T1PKS	kirromycin	8%	BGC0001070
	20	terpene	carotenoid	18%	BGC0000633
Pf1_DN64_8G1	1	acyl_amino_acids	-	-	-
	2	NRPS	rimosamide	14%	BGC0001760
	3	betalactone	pseudaminic acid	22%	BGC0001747
	4	T1PKS, T3PKS	-	-	-
	5	bacteriocin	-	-	-
	6	terpene	-	-	-
	7	betalactone	fengycin	13%	BGC0001095
	8	arylpolyene, ladderane	-	-	-
	9	siderophore	-	-	-
	10	bacteriocin	-	-	-

Table S4. Continued.

Strain ID	BGC	BGC type	Most similar known (MIBiG)	Similarity	MIBiG accession (Medema et al., 2015)
Aa3_DN64_1D3	1	betalactone	pseudaminic acid	22%	BGC0001747
	2	NRPS	rimosamide	14%	BGC0001760
	3	acyl_amino_acids	-	-	-
	4	arylpolyene, ladderane	-	-	-
	5	terpene	-	-	-
	6	bacteriocin	-	-	-
	7	siderophore	-	-	-
	8	bacteriocin	-	-	-
	9	betalactone	fengycin	13%	BGC0001095
	10	T3PKS, T1PKS	-	-	-
Aa3_DN213_3F7	1	NRPS	coelibactin	27%	BGC0000324
	2	NRPS	streptobactin	11%	BGC0000368
	3	NRPS-like	-	-	-
	4	siderophore	desferrioxamine E	50%	BGC0001478
	5	terpene	carotenoid	85%	BGC0000636
	6	LAP	corynazolicin	42%	BGC0001174
	7	ectoine	ectoine	75%	BGC0000853
	8	NRPS	caboxamycin	20%	BGC0001444
	9	bacteriocin	-	-	-

Table S5. Growth media and isolation source of the studied bacterial strains. Details on the media preparation can be found in the References mentioned below. MA; Marine Agar, MA/10; Marine Agar (10x diluted), MHA; Mueller-Hinton Agar, MHA/10; Mueller-Hinton Agar (10x diluted).

Strain ID	Isolation source	Medium	Reference
Aa3_DN55_6A7	*Aplysina aerophoba*	MA/10	(Versluis et al., 2017a)
Pf1_Ps_8H04_1	*Petrosia ficiformis*	MA	(Versluis et al., 2018)
Pf1_DN206_4B7	*Petrosia ficiformis*	MA	(Versluis et al., 2017a)
Irc_Ps_AB108	*Ircinia* sp.	MA	(Versluis et al., 2018)
Pf1_DN64_8G1	*Petrosia ficiformis*	MA	(Versluis et al., 2017a)
Aa3_DN64_1D3	*Aplysina aerophoba*	MA	(Versluis et al., 2017a)
Pf1_Ps_8H06	*Petrosia ficiformis*	MA	(Versluis et al., 2018)
Cn_Ps_AB111	*Chondrilla nucula*	MA	(Versluis et al., 2018)
Aa3_Str.68_7G12	*Aplysina aerophoba*	MA	(Versluis et al., 2018)
Acac_Ps_AB113	*Acanthella acuta*	MA	(Versluis et al., 2018)
Aa3_DN166_3E9_2	*Aplysina aerophoba*	MA	(Versluis et al., 2017a)
Pf1_DN81_6F7_2	*Petrosia ficiformis*	MHA	(Versluis et al., 2017a)
Cc1_DN217_4H2	*Corticium candelabrum*	MA	(Versluis et al., 2017a)
Aa3_DN138_5C8	*Aplysina aerophoba*	MA	(Versluis et al., 2017a)
Pf1_DN14_7A9_1	*Petrosia ficiformis*	MHA/10	(Versluis et al., 2017a)
Aa3_DN73_5E10_2	*Aplysina aerophoba*	MHA	(Versluis et al., 2017a)
Aa3_DN30_1H2	*Aplysina aerophoba*	MHA	(Versluis et al., 2017a)
Aa3_DN216_4B10_1	*Aplysina aerophoba*	MHA	(Versluis et al., 2017a)
Aa3_DN213_3F7	*Aplysina aerophoba*	MA	(Versluis et al., 2017a)
Aa3_DN216_4B10_2	*Aplysina aerophoba*	MHA	(Versluis et al., 2017a)
Aa3_DN71_7G3_2	*Aplysina aerophoba*	MHA	(Versluis et al., 2017a)

Chapter 4

Distribution and diversity of *Candidatus* Tectomicrobia, a deep-branching uncultivated bacterial lineage harbouring rich producers of bioactive metabolites

Eike E. Peters, Jackson K. B. Cahn, Asimenia Gavriilidou, Ursula A. E. Steffens, Catarina Loureiro, Michelle A. Schorn, Paco Cárdenas, Nilani Vickneswaran, Philip Crews, Detmer Sipkema, Jörn Piel

Manuscript submitted

Abstract

Genomic and functional analyses of bacteria belonging to the candidate genus 'Entotheonella' have revealed them as the producers of diverse bioactive compounds previously identified from their sponge hosts. Based on genetic distance to other taxa, '*Candidatus* Entotheonella' has been proposed as the first member of a new candidate phylum, 'Tectomicrobia'. Here, we analysed environmental samples and publicly available 16S ribosomal RNA (rRNA) gene sequences to assess the phylogenetic structure and environmental distribution of this as-yet sparsely populated lineage. The data showed that '*Ca.* Entotheonella' and other *'Ca.* Tectomicrobia' were not restricted to marine habitats but also widely distributed among terrestrial locations. Environmental factors, such as water depth and host association were identified to correlate with the phylogenetic patterns of marine phylotypes. The previously described '*Ca.* Entotheonella' lineage could be more accurately divided into at least three different candidate genera with the terrestrial '*Candidatus* Prasianella', the marine '*Candidatus* Thalassonella' and the more widely distributed '*Ca.* Entotheonella' of mixed origin. This first metagenomic and microscopic characterization of '*Ca.* Thalassonella' from a range of sponge hosts did not suggest a role for its members as providers of natural products, despite clear similarities in genome-based predictions of primary metabolism and implied lifestyle with '*Ca.* Entotheonella'. In contrast, within the revised '*Ca.* Entotheonella', the analysis revealed a correlation between the 16S rRNA gene phylogeny and a specific association with sponges and their chemistry, suggesting that this feature might serve as a discovery method to accelerate the identification of new '*Ca.* Entotheonella' producers of bioactive compounds. Application of this strategy led to the identification of the first '*Ca.* Entotheonella' symbiont in a non-lithistid sponge, *Psammocinia* sp., thus indicating a wider host distribution of '*Ca.* Entotheonella'-based chemical symbiosis.

Introduction

The bacterial tree of life contains numerous deep-branching lineages that lack cultivated representatives (Konstantinidis et al., 2017; Parks et al., 2018; Parks et al., 2020; Hugenholtz et al., 2021). Data from 16S rRNA genes and putative environmental genomes, obtained by metagenomic binning or single-cell sequencing, support the existence of dozens of uncultivated phylum-like divisions that are distributed ubiquitously or in more specialized habitats. Among a large and growing list of examples are the Candidate Phyla Radiation (Brown et al., 2015; Hug et al., 2016b) [CPR, previously known as 'Patescibacteria' (Rinke et al., 2013; Parks et al., 2018)], the SAR324 group detected in hydrothermal plumes (Cao et al., 2016; Parks et al., 2020), or '*Candidatus* Poribacteria' present in sponge microbiomes (Kamke et al., 2014). Omics data suggest diverse and intriguing functions for such elusive taxa (Keren et al., 2017; Crits-Christoph et al., 2018), but as experimental validation is usually challenging verified functions remain limited.

In collaborative studies, we recently reported members of the candidate genus 'Entotheonella', first described by researchers at the Scripps Institution of Oceanography (Bewley et al., 1996; Schmidt et al., 2000), as an uncultivated taxon with a remarkably rich specialized metabolism (Wilson et al., 2014; Ueoka et al., 2015; Lackner et al., 2017; Mori et al., 2018). Genome data suggested '*Ca.* Entotheonella' as members of a new candidate phylum, termed 'Tectomicrobia' (Wilson et al., 2014), which was corroborated by a recent reanalysis of the bacterial tree of life based on standardized classification criteria (Parks et al., 2018). All sequenced '*Ca.* Entotheonella' phylotypes with known morphology form multicellular filaments, have large genomes around 10 Mbp, and colonize sponges with which they appear to form mutualistic associations that involve chemical defense (Wakimoto et al., 2014; Wilson et al., 2014; Ueoka et al., 2015; Lackner et al., 2017; Mori et al., 2018) and arsenic and heavy metal detoxification (Keren et al., 2017). Marine sponges are prolific sources of bioactive natural products that may contribute to protecting the sessile animals against predators and epibionts (Pawlik, 2011). In the demosponge *Theonella swinhoei* (order Tetractinellida, suborder Astrophorina, family Theonellidae), containing a particularly rich chemistry, bioinformatic and biochemical data attributed most of the known substances to the two symbiont phylotypes '*Ca.* Entotheonella factor' and '*Candidatus* Entotheonella serta' (Wilson et al., 2014; Mori et al., 2018) (Table 1). These producers generate distinct sets of natural products and, in a mutually exclusive fashion, colonize two different host chemotypes, *T. swinhoei* Y and W, as members of microbiomes comprising numerous other bacteria. In *T. swinhoei* Y, '*Ca.* E. factor' is accompanied by another variant, '*Ca.* E. gemina' (Wilson et al., 2014), while another phylotype is present in a *T. swinhoei* W sampled from the Red Sea. One or more additional '*Ca.* Entotheonella' phylotypes were detected in the theonellid sponge *Discodermia calyx* and collectively linked to the production of three biosynthetically distinct compound families (Table 1) (Kimura et al., 2012; Wakimoto et al., 2014; Nakashima et al., 2016).

Table 1. Sponges and their known or suspected '*Ca.* Entotheonella' symbionts, updated from (Reiter et al., 2020).

Sponge	Source	*'Ca.* Entotheonella' phylotype	Known natural products
T. swinhoei Y	Japan	'*Ca.* E. factor', '*Ca.* E. gemina'	Polytheonamides, onnamides, theopederins, orbiculamides, cyclotheonamides, pseudotheonamides, nazumamide A
T. swinhoei WA	Japan	'*Ca.* E. serta' TSWA1	Misakinolides, theonellamides
T. swinhoei WB	Israel	‹*Ca.* E. serta› TSWB1, '*Ca.* Entotheonella' TSWB2	Swinholides, theonellamides
D. calyx	Japan	'*Ca.* Entotheonella' (one or more phylotypes)	Calyculin, calyxamides, kasumigamide

Besides biosynthetic gene clusters (BGCs) assigned to known sponge natural products, all '*Ca.* Entotheonella' genomes sequenced to date ('*Ca.* E. serta', '*Ca.* E. factor', and '*Ca.* E. gemina') contain multiple biosynthetic loci for as-yet unknown compounds. Enzymatic studies on one of the uncharacterised BGCs suggested that the orphan clusters are likely functional and encode the biosynthesis of previously unknown metabolites (Wilson et al., 2014; Helf et al., 2017; Reiter et al., 2020). Their remarkable metabolic capabilities classify '*Ca.* Entotheonella' as the first uncultivated producer taxon with a chemical richness and variability comparable to important cultivable drug discovery sources, such as filamentous *Actinomycetes*, *Cyanobacteria*, or *Myxobacteria* (Jaspars and Challis, 2014). Their pharmacological potential, intriguing biology, and isolated phylogenetic position warrant further investigations into the distribution and functions of '*Ca.* Tectomicrobia', which currently contain only four members in the Genome Taxonomy Database (GTDB) (Parks et al., 2018).

In this study, we aimed to obtain insights into the phylogenetic structure and environmental distribution of '*Ca.* Tectomicrobia', seeking to address the following three questions: Do further tectomicrobial taxa exist besides '*Ca.* Entotheonella'? If so, what is their chemical potential? Is the association with sponges a general feature of this lineage or can free-living representatives be identified? Our data showed that '*Ca.* Tectomicrobia' contained a wider range of lineages that occur in diverse marine and terrestrial habitats. For one of these, the new sponge-associated candidate genus 'Thalassonella', we presented eight metagenome-assembled genomes (MAGs). Within '*Ca.* Entotheonella', we identified two distinct phylogenetic signatures that suggest the existence of generalist sponge colonizers as well as host-specific members, the latter including all known natural product-rich phylotypes. Based on these data, we tested whether a 16S rRNA gene-based prioritization strategy can pinpoint new '*Ca.* Entotheonella' producers. This strategy resulted in the discovery of a new '*Ca.* Entotheonella' variant in the dictyoceratid sponge *Psammocinia* sp., showing that small molecule-based symbiosis within '*Ca.* Entotheonella' is not restricted to the demosponge order Tetractinellida.

Materials and Methods

Sponge collection and sequencing

Specimens of *Psammocinia* sp. (class Demospongiae, order Dictyoceratida, family Irciniidae) were collected by SCUBA diving at Milne Bay, Papua New Guinea (Zoological Museum of Amsterdam collection number ZMAPOR 19842 = UCSC coll. no. 03526), 12-18 m depth, in December 2003 at 9°37.214‘ S 150°57.332‘ E, 9°14.008‘ S 150°46.782‘ E, 9°14.147‘ S 150°47.173‘ E, and 9°19.868‘ S 150°43.906‘ E as described in Robinson et al. (2007) and preserved in 95% EtOH. Shallow *Geodia barretti* (class Demospongiae, order Tetractinellida, family Geodiidae) were collected by SCUBA diving in 2012 in the Bjørnsund Islands and Lysefjord, western Norway (Cárdenas and Rapp, 2013). Other North Atlantic deep-sea *Geodia* species (*G. atlantica*, *G. barretti*, *G. hentscheli*, *G. macandrewii*, *G. parva*, *G. pachydermata*, *G. phlegraei*) (Cárdenas et al., 2013) were collected between 2004 and 2016 using remotely operating vehicles (ROVs) and trawls/dredges during various cruises in western Norwegian fjords, Kosterfjord (Sweden), Rockall Bank, the Greenland Sea, the Davis Strait, the Flemish Cap, Svalbard and the Galicia Bank. All specimens were preserved in 95% EtOH shortly after collection (Table S1).

Specimens of *G. barretti* (Gb1, Gb2 and Gb4-Gb10) and *G. atlantica* (Ga3) for metagenomic sequencing were collected by dredging onboard R/V Hans Brattström of the University of Bergen from Korsfjord, south of Bergen, Norway (60°8.13’ N, 5°6.7’ E) in September and October 2017. They were chopped and flash frozen with liquid nitrogen upon collection and stored at -80 °C. Specimens (Gb126, Gb278 and Gb305) of *G. barretti* were collected during benthic trawls by the crews of the R/V Pâmiut of the Greenland Institute of Natural Resources during cruises conducted by Fisheries and Oceans Canada in the Davis Strait taken during the same season (Sept-Oct) from 2011–2015 and were stored at -20 °C (Steffen et al.). The *Geodia* samples (Gb1, Gb2, Gb4-Gb10 and Ga3) were crushed in liquid nitrogen to a fine power with pestle and mortar. Two hundred mg of sponge tissue material were disrupted by bead beating using milling balls (5 x 2 mm + 2 x 5 mm) and 2 steps of shaking for 20 s at 4,000 rpm as described in Roume et al. (2013). Tissue lysate was further used for DNA extraction with the AllPrep DNA/RNA/Protein Mini Kit (Qiagen) and sequenced by Novogene (Hong Kong) using the Illumina HiSeq PE150 platform.

Four sponges (DOM10, DOM14B, DOM33, DOM40) were collected by submersible near Portsmouth, Dominica. Sponges were stored in RNALater (Thermo Fisher, USA) at -20 °C until further processing (Table S1). Sponge pieces were defrosted, rinsed with sterile artificial seawater, chopped into small pieces, added to a PowerBead Tube and subjected to shaking three times for 60 s at 6,000 x g using the Precellys 24 Homogenizer (Bertin GmbH, Germany). DNA was subsequently extracted following the standard protocol of the DNeasy PowerSoil Pro Kit (Qiagen). Metagenomic DNA samples were sent to Novogene Europe (United Kingdom) for sequencing using the Illumina Novaseq6000 platform with the PE150 library and sequencing kits.

Soil sampling and sequencing

Soil samples were collected from 19 locations in Germany and Norway (Table S2). To isolate the metagenomic DNA of the samples, the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA) was used. Partial 16S rRNA genes were amplified by a nested PCR approach using the bacterial primers 27F (5'-AGA GTT TGA TCM TGG CTC AG-3') and 1492R (5'-TAC GGY TAC CTT GTT ACG ACT T-3') in a first round of PCR and the '*Ca.* Entotheonella'-specific primers Ento238F (5'-CCG GTC TGA GAT GAG CTT GC-3') (Schmidt et al., 2000) and Ento1442R (5'-TCA CCC CAA TCA CCC CGC-3') (Wilson et al., 2014) in the second round of PCR. The resulting DNA fragments were either sequenced directly or subcloned into pJet1.2 using the CloneJet PCR Cloning Kit (Thermo Scientific, USA) prior to sequencing.

Enrichment of soil bacteria

The collected soil was resuspended in 0.9% (w/v) aqueous NaCl solution using a hand blender. The suspension was stirred with a magnetic stirrer for 20 min then left undisturbed for additional 10 min to allow for settling of particles. The supernatant was decanted through a 32 µm Nytex mesh and centrifuged at 100 × g, 1,000 × g and 4,500 × g to separate the bacteria by their different sedimentation characteristics. Each bacterial fraction was resuspended in 0.9% (w/v) aqueous NaCl. For further separation and to remove remaining soil particles, density gradient centrifugation was employed. A Nycodenz cushion (60% (w/v) SERVA, Heidelberg) was placed beneath the bacterial suspension and centrifuged at 10,000 × g for 20 min at 4 °C. The cell interlayer was transferred to a new tube, washed with phosphate-buffered saline (PBS) and resuspended in 0.9% NaCl. For subsequent CARD-FISH studies, the enriched cells were either fixed in 4% paraformaldehyde (PFA, in PBS) or 100% ethanol at 4 °C overnight and subsequently stored at -20 °C until use.

Isolation of filamentous bacteria from the sponge *Psammocinia* sp.

Filament-enriched bacterial cell fractions from *Psammocinia* sp. were prepared by a modified protocol previously described for *Theonella* sponges (Wilson et al., 2014). Briefly, the sponge tissue was cut into small pieces, immersed in buffered calcium- and magnesium-free artificial sea water (CMF-ASW: 10 mM Tris-Cl, 449 mM NaCl, 9 mM KCl, 33 mM Na_2SO_4, 2 mM $NaHCO_3$) and further processed using a mortar and pestle. The homogenized suspension was subsequently transferred to a Falcon tube and incubated with a mixture of collagenase I and IV (end concentration: 240 µg/ml) for 30 min at 37 °C. After 10-fold dilution with CMF-ASW and addition of 2.5 mM ethylene glycol tetra-acetic acid (EGTA), the solution was incubated on a rotating wheel at 4 °C overnight. Next day, the solution was passed through a 32 µm Nytex mesh and centrifuged at 100 × g for 10 min to pellet filaments. The pellet was washed 3 times with 500 µl CMF-ASW and stored at 4 °C until use.

Phylogenetic analysis

Previously deposited tectomicrobial 16S rRNA gene sequences were retrieved from the GenBank sequence database by BLASTing the full-length 16S rRNA gene sequence of '*Ca.* Entotheonella factor' (KF926817). Hits above a threshold of 75% identity (>1000 sequences) were aligned using MUSCLE (Edgar, 2004), followed by manual alignment correction. A total of 811 sequences were selected to create an alignment with maximal sequence length and without gaps in the region corresponding to positions 337 to 1078 in the *E. coli* homolog. To the dataset were added 187 sequences belonging to known species of *Nitrospirae*, *Deltaproteobacteria*, *Alphaproteobacteria*, *Rokubacteria*, *Acidobacteria* and *Nitrospinae* as outgroups. After a first round of phylogenetic analysis using the Neighbour-Joining method (Nei and Kumar, 2000), the putative tectomicrobial dataset was reduced from 811 to 456 sequences that fulfilled the monophyly criterion. The initial phylogram was generated by the Neighbour-Joining method with 500 bootstrap resamplings using the Jukes-Cantor model to compute evolutionary distances. To further test for a placement within '*Ca.* Tectomicrobia', the phylogeny was also reconstructed with the Maximum Likelihood algorithm with 500 bootstrap resamplings using the Generalized Time Reversible model (Felsenstein, 1985; Nei and Kumar, 2000). A discrete gamma distribution of 0.47 was used to model evolutionary rate differences among sites. The percentage of replicate trees in which the associated taxa clustered together in a bootstrap test of both algorithms are indicated in the final tree when either value was above 50%. All evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013) and visualized using the online tool Interactive Tree Of Life (iTOL) 5.0 (Letunic and Bork, 2016). A global identity matrix was used to calculate the median sequence identities (Hodges-Lehmann median) of all taxa, and a normal approximation was the basis for the upper and lower bounds for the 95% confidence interval.

Metagenome assembly, binning, and bin classification

Adapter removal, quality filtering and normalization was done using the BBtools suite v37.64 (Bushnell et al., 2017) with the following parameters: ktrim=r k=23 mink=7 hdist=1 tpe tbo qtrim=rl trimq=20 ftm=5 maq=20 minlen=50. The reads were normalized for coverage with parameters, target=200 min=3. Filtered reads were assembled with SPAdes v3.12 (Nurk et al., 2017) using the --meta and --only-assembler flags. Contigs were binned and refined using metaWRAP v1.2 (Uritskiy et al., 2018) with minimum completeness of 50%, maximum contamination of 10% (*G. barretti* and *G. atlantica* samples were binned with MaxBin2 2.2.4 and metaBAT2 2.9.1 and DOM samples with MaxBin2 2.2.4, metaBAT2 2.9.1, and CONCOCT 0.4.0). Reassembly was performed using metaWRAP's *reassembly* module (SPAdes 3.10.1). The obtained bins were taxonomically classified using the GTDB-Tool Kit 1.1.0 (GTDB-Tk) *classify* workflow (Chaumeil et al., 2019).

A BLASTN search using the 16S rRNA gene sequences from '*Ca.* E. factor' and '*Ca.* E. serta' as queries was used to detect 16S rRNA gene sequences within the putative '*Ca.* Tectomicrobia' bins. Of the 24 putative '*Ca.* Tectomicrobia' bins, 22 had at least a fragmentary 16S rRNA

gene sequence, and 15 of those had at least 400 bp of overlap with the amplicon generated from our '*Ca.* Entotheonella'-specific primers (Wilson et al., 2014; Mori et al., 2018). These sequences were aligned to the 811 used previously using MUSCLE, followed by trimming of the alignment and manual realignment. The phylogenetic tree was regenerated for analysis using the Maximum Likelihood method as described above.

Phylogenomic and genomic analysis

Phylogenomic trees were constructed in Anvi'o (Eren et al., 2015) using the standard phylogenomics workflow laid out in the tutorial. Briefly, the putative '*Ca.* Tectomicrobia' MAGs were imported into Anvi'o, genes were called with Prodigal (Hyatt et al., 2010), and an HMM profile was created. The protein sequences corresponding to 71 bacterial single-copy genes (Lee, 2019) were extracted, filtered to remove those present in less than 75% of MAGs, and concatenated; these concatenated sequences were then aligned with MUSCLE (Edgar, 2004) and this alignment was used to generate a phylogenomic tree using FastTree v2.1.11 (Price et al., 2010). BGCs were predicted using bacterial AntiSMASH 5.0 with 'relaxed' detection strictness (Blin et al., 2019). The quality of the '*Ca.* Thalassonella' MAGs was estimated with CheckM v1.1.2 (Parks et al., 2015). Only medium-high-quality MAGs (>75% completeness, <5% contamination) were considered for downstream analysis. Genome annotation was performed with Prokka v1.14.6 (Seemann, 2014). Assignment of Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthologs (KO) and metabolic pathway reconstruction were conducted using the KofamKOALA (Aramaki et al., 2020) and BlastKOALA (Kanehisa et al., 2016) online tools with default settings. For annotation with BlastKOALA, the 'species_prokaryotes' KEGG GENES database was used.

Catalyzed reporter deposition fluorescence in situ hybridization (CARD-FISH)

CARD-FISH experiments were performed as described previously (Ueoka et al., 2015). Briefly, *Psammocinia* sp. was flash-frozen in liquid nitrogen and subsequently cut into 30 µm slices using a microtome (Microm, Thermo Fisher, USA). Representative sponge tissue slices were then transferred onto microscopy slides, air-dried for 3 h at room temperature and subsequently fixed in PBS-buffered 4% PFA (4 °C, overnight). After washing with PBS, slices were permeabilized with 10 mg/ml lysozyme (20 mM Tris HCl pH 8.0, 2 mM EDTA, 0.1% Triton X-100) for 30 min at room temperature, washed once with PBS and incubated for 10 min at 4 °C with 1 mg/ml proteinase K (100 mM Tris HCl pH 8.0, 100 mM EDTA, 0.1% Triton X-100). Endogenous peroxidases were subsequently quenched by incubation with 0.01 M HCl for 20 min at room temperature, afterwards washed with PBS, dehydrated in pure EtOH for 3 min at room temperature, and air-dried. The hybridization reaction took place in hybridization buffer [0.9 M NaCl, 20 mM Tris HCl pH 7.6, 10% (wt/vol) dextran sulfate, 0.05% SDS, 1% nucleic acid blocking reagent (Roche, Switzerland), 0.5 mg/mL herring sperm DNA (Sigma)] containing 55% (vol/vol) formamide and 0.5 ng/µL '*Ca.* Entotheonella'-specific horseradish peroxidase (HRP)-coupled probe ESP219 (5'-CCG CAA GCY CAT CTC AGA CC-3'; BioMers) for

4 h at 35 °C. Subsequently, tissue slices were washed once in prewarmed washing buffer (3 mM NaCl, 5 mM EDTA pH 8.0, 20 mM Tris HCl pH 7.6, 0.05% SDS) and once in PBS-T (0.01% Triton X-100) for each 30 min at 37 °C. After equilibration of the probe-delivered HRP in PBS for 20 min, tissue slices were incubated in amplification solution [1× PBS pH 7.6, 2 M NaCl, 20% (wt/vol) dextran sulfate, 0.1% nucleic acid blocking reagent (Roche), 0.0015% H_2O_2] containing Alexa Fluor 647-labeled tyramide (Life Science) for 1 h at 37 °C in the dark. Next, tissue slices were washed 3 times in PBS for 10 min at room temperature, once washed in distilled ice-cold H_2O, dehydrated in pure EtOH, and air-dried. For microscopic analysis, tissue slices were covered with mounting agent (Citifluor Limited, UK) and observed under a Zeiss Axioskop 2 epifluorescence microscope equipped with a 75-W xenon arc lamp (XBO 75) and a 20× Plan Neofluar objective.

Results and Discussion

Analysis of environmental samples for the presence of '*Ca.* Tectomicrobia'

Although our previous analyses had focused on '*Ca.* Tectomicrobia' living in symbiosis with astrophorin sponges, preliminary phylogenetic analyses also suggested the existence of this candidate phylum in other sponges and habitats. In an attempt to uncover further tectomicrobial diversity, we engaged in a targeted sequence prospecting campaign. For this purpose, we collected soil from 19 locations in Germany and Norway (Table S2) and isolated the DNA of enriched bacterial fractions prepared from these samples. To analyse phylogenetic divergence in these samples, we generated 16S rRNA gene sequences by a nested PCR approach previously established for the identification of '*Ca.* Entotheonella' in metagenomic DNA samples (Wilson et al., 2014) or by the use of the more general 16S rRNA gene PCR primers, 27F and 1492R, on enriched bacterial cell pellets. In both cases, target amplicons were obtained and subcloned for sequencing that resulted in a combined 51 new near full-length tectomicrobial 16S rRNA gene sequences. A similar procedure was carried out on *Geodia* sponges collected from the North Atlantic as well as previously collected astrophorin sponges from Japan, generating an additional 34 new tectomicrobial 16S rRNA sequences derived from *Geodia* and 8 from the rest, for a total of 93 (Table S3).

Phylogenetic analyses suggest various marine and terrestrial genus-like lineages in '*Ca.* Tectomicrobia'

For the phylogenetic analyses, public databases were searched for additional potential tectomicrobial 16S rRNA gene sequences to further enrich the dataset. In total, 456 sequences remained that were used in this study, consisting of 363 sequences derived from GenBank and our 93 newly acquired sequences. None of these originated from cultivated bacteria. A previous preliminary phylogenetic analysis of '*Ca.* Tectomicrobia' had suggested the presence of three distinct clades, with the '*Ca.* Entotheonella' lineage harbouring most of the sequences (Wilson et al., 2014). Using our new, expanded dataset, phylograms

inferred by maximum likelihood and neighbour-joining methods instead showed two deep-branching clades with high bootstrap support (Figures 1A and S1-S3). Clade 1 comprised 119 sequences with low sequence divergence (95.4% median sequence identity, MSI), suggesting a single candidate genus, for which we propose the name *'Candidatus* Allonella' (from Greek ἄλλος, other). Clade 2 could be further divided into at least three major subgroups that were well-supported by bootstrap replicates. All sequences previously designated '*Ca.* Entotheonella' (Wilson et al., 2014) fell into one of these three subgroups, along with 11 of the sequences generated in our sequencing campaign. This group, which retains the name '*Ca.* Entotheonella', had an MSI of 96.1%, suggesting a taxonomic ranking of a candidate genus according to the taxonomic thresholds proposed by (Yarza et al., 2014). Subgroup 2 contained 66 marine sponge-derived sequences, 52 of which were obtained from our sampling of deep-sea *Geodia* spp., for which we propose the name '*Ca.* Thalassonella' (Greek: θάλασσα, sea; MSI = 96.6%) . The third genus-like subgroup in Clade 2 emerged with the addition of the new soil-derived 16S rRNA sequences generated in this study, and contained 112 sequences of exclusively terrestrial origin. Because the first soil sample that was positively tested for '*Ca.* Tectomicrobia' originated from a vegetable garden, we propose the name '*Ca.* Prasianella' (ancient Greek, πρασιά, garden bed; MSI = 96.3%). In addition to their distinctive phylogeny, '*Ca.* Prasianella' and '*Ca.* Thalassonella' 16S rRNA gene sequences contained as a diagnostic feature a 25-29 bp insertion in the V3 region, which '*Ca.* Entotheonella' sequences lacked (Figure S4). Two additional smaller subgroups, 't1' and 't2', which contained 10 and 11 sequences of mainly terrestrial origin, respectively, were also present in Clade 2 (Figure S2). However, further sequences are necessary to resolve the phylogenetic relationship between 't1', 't2' and members of '*Ca.* Thalassonella' and '*Ca.* Prasianella'.

Comparison of the groups within Clade 2 shows 85.9-90.1% MSI (Figure 1C), suggesting that Clade 2 represents a candidate order, '*Ca.* Entotheonellales', whereas Clades 1 and 2 had an MSI of 81.1%, suggesting membership in a shared class or phylum according to the thresholds proposed by Yarza et al. (2014). Although our previous efforts to identify reservoirs of '*Ca.* Tectomicrobia' had focused on marine sponges, the current analysis showed that members are widespread in soil. In total, 54.1% of the tectomicrobial sequences derived from soil habitats, 5.0% from freshwater, 1.4% from saltwater and 39.5% from marine sponges (Figure 1B), although this distribution is likely heavily affected by sampling bias.

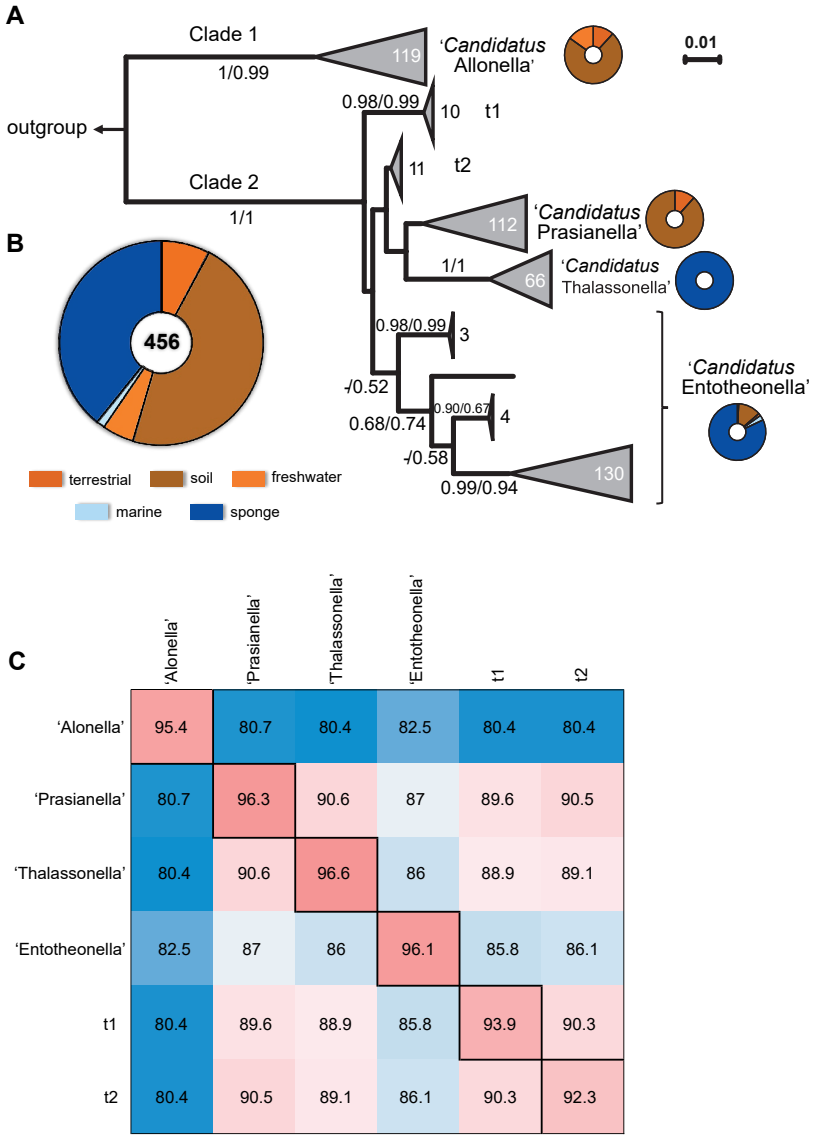


Figure 1. Phylogeny and environmental distribution of the candidate phylum 'Tectomicrobia' based on 16S rRNA gene sequences. A) Phylogenetic tree inferred from 456 tectomicrobial sequences under the neighbour-joining criterion. Support of individual branches by bootstrap values is indicated at the respective nodes for both methods used when either value is above 50% (neighbour-joining/maximum-likelihood). Scale bar, 0.01 changes per nucleotide position. The number of sequences comprising a specific group is shown inside or next to the corresponding clade. The phylum *Nitrospinae* was used here as an outgroup. The environmental distribution of sequences assigned to specific groups is shown as a pie chart next to the corresponding group. B) Pie chart diagram illustrating the environmental distribution of '*Ca.* Tectomicrobia' based on 456 16S rRNA gene sequences. A list of sequences generated in this study can be found in Table S3. C) Median sequence identities within (main diagonal) and between (off-diagonal) the genus-level clades described in this study. A more detailed version of this figure, showing higher and lower level divisions, is provided in Figure S5.

Correlation between phylogeny and environmental distribution within '*Ca.* Tectomicrobia' lineages

All 112 sequences within the genus '*Ca.* Prasianella' originated from terrestrial sources, but no clear further differentiating relationship could be identified between phylogeny and habitat or geographic origin (Figure S3). Similarly, the geographic distribution of '*Ca.* Thalassonella' covered various marine regions including the Pacific, Atlantic, Arctic and Indian Oceans (Figure 2). Among the 66 '*Ca.* Thalassonella' sequences, most originated from various *Geodia* spp. sponges (37 sequences), followed by the 16 sequences from the two *Xestospongia* species, *Xestospongia testudinaria* (Manado, Indonesia) and *Xestospongia muta* (Key Largo, Florida, USA) (Montalvo and Hill, 2011). The large majority of sponges with '*Ca.* Thalassonella' were high-microbial abundance (HMA) sponges (*Geodia* spp., *Xestospongia* spp., *Aplysina* spp., *Ecionemia alata*, *Ircinia strobilina*, *Vaceletia crypta*, *Plakortis halichondrioides*), with one low-microbial abundance (LMA) sponge (*Tedania ignis*) and two species of more uncertain status (*Astrosclera willeyana*, *Haliclona tubifera*) (Simister et al., 2013; Gloeckner et al., 2014; Lavy et al., 2018). '*Ca.* Thalassonella' phylotype clades were strongly based on their host sponge (phylosymbiosis), particularly for *Geodia* and *Xestospongia* sponges for which several '*Ca.* Thalassonella' sequences were available in this study (Figure 2). The '*Ca.* Thalassonella' group also contained a single sequence from a diseased coral (*Montastraea faveolata*) as the only non-sponge-derived representative. However, since a corresponding sequence was not reported from a nearby, healthy coral, it might originated from a contamination rather than the coral microbiome (Kimes et al., 2013).

Although no geographic patterns were observed, sequences from similar water depths clustered together, as apparent from the existence of individual '*Ca.* Thalassonella' *Geodia* subclades comprising sublittoral/fjord (36-200 m), sublittoral/fjord to upper bathyal (36-787 m) and lower bathyal members (688-1462 m) (Figure 2). Thus, water depth, or more likely water masses, along with the sponge host species, appeared to be major determinants influencing the differentiation of '*Ca.* Thalassonella'. The sharing of specific ASVs by different *Geodia* species over large geographical distances could reflect *Geodia* population connectivity (Roberts et al., 2021). For example, *G. macandrewii* and *G. barretti* sharing one ASV between the upper bathyal Svalbard and Davis Strait (~3700 km) may reflect host-microbe connections via the deep-sea East Greenland current, followed by the West Greenland current. Depth has been reported in several studies as a factor that stratifies marine bacterial communities at the species [e.g., *Bacillus cereus* (Liu et al., 2013)), phylum (e.g., SAR11 (Field et al., 1997)], and global level (Sunagawa et al., 2015a). However, more thorough sampling will be required to rule out potential confounding factors.

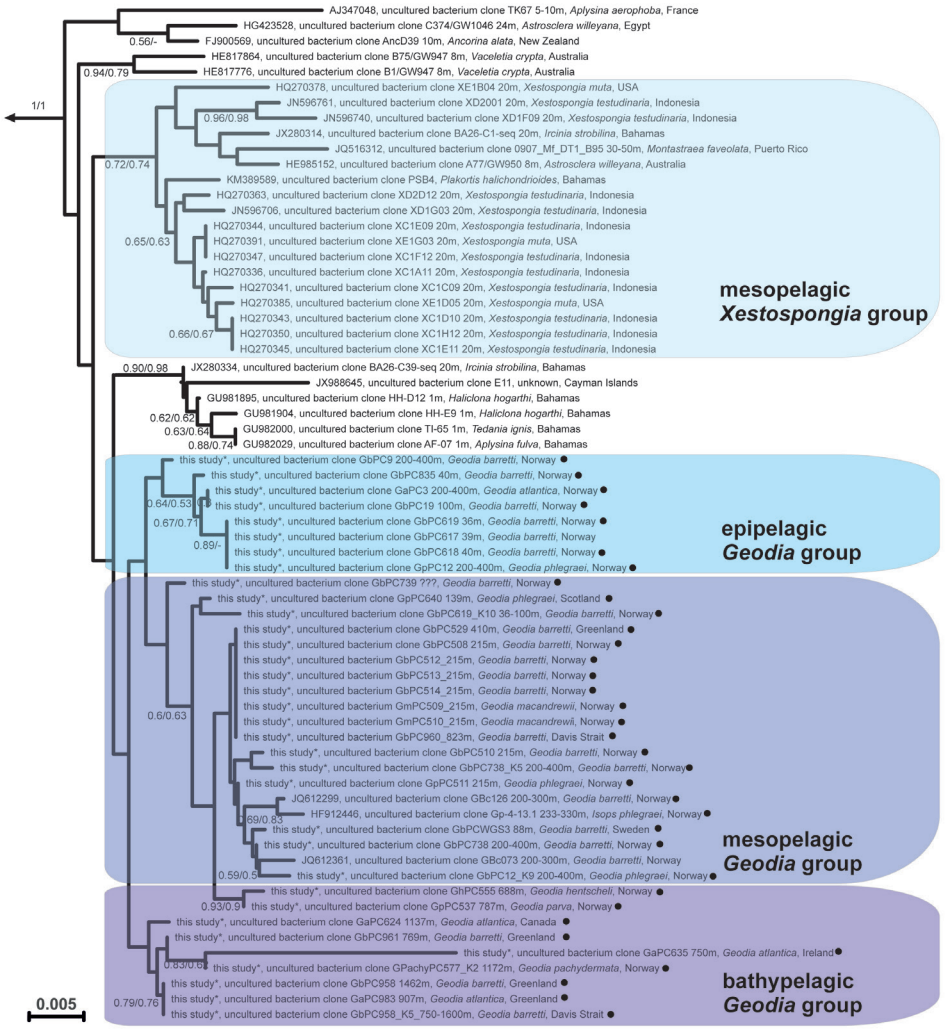


Figure 2. Phylogeny of '*Ca.* Thalassonella'. Detailed view of '*Ca.* Thalassonella subgroup for the tree shown in Figure 1. Bootstrap values for both methods used are given when either value is above 50% (neighbour-joining/maximum-likelihood). Sequences generated in this study are indicated with a black circle. Scale bar, 0.005 changes per nucleotide position.

'*Ca.* Entotheonella' members exhibit two contrasting phylogenetic patterns

Phylogenetic insights were of particular interest for members of '*Ca.* Entotheonella' as rich producers of bioactive natural products in sponges (Kimura et al., 2012; Wakimoto et al., 2014; Wilson et al., 2014; Ueoka et al., 2015; Nakashima et al., 2016; Mori et al., 2018). In contrast to '*Ca.* Thalassonella', only limited correlation was initially apparent between bacterial phylogeny and host affiliation (Figure 3). Also, host sponges here were a mix of HMA (*Theonella*, *Discodermia*, *Aplysina*, *Ircinia*, *Agelas*, *Pseudoceratina purpurea*) and LMA (*Cliona*, *Stylissa*, *Hexadella*, *Dysidea*, *Callyspongia*) sponges (Blanquer et al., 2013; Gloeckner et al., 2014). However, closer analysis revealed two distinct distribution patterns among '*Ca.* Entotheonella' phylotypes. All sequences corresponding to the previously described metabolically talented '*Ca.* Entotheonella' phylotypes fell into one clade (group I in Figure 3, MSI = 96.8%) with relatively high sequence divergence and a branching topology that suggested some degree of specificity with sponge hosts. A contrasting pattern was found in groups IIa (MSI = 98.7%) and IIb (MSI = 99.6%) (Figure 3 and S5), which harboured '*Ca.* Entotheonella' sequences with highly similar 16S rRNA genes that were detected exclusively by PCR in a wide range of sponges. These closely related '*Ca.* Entotheonella' phylotypes are clearly not tied to the phylogeny of their hosts; they might either be host-promiscuous symbionts or non-symbiotic contaminants from seawater. In some sponge species, e.g. *T. swinhoei* or *Discodermia dissoluta*, the same sponge specimen was found to contain sequences from groups I and IIa/IIb. The final '*Ca.* Entotheonella' clade (group IIc) showed slightly higher sequence divergence (MSI = 98.2%) than groups IIa and IIb (Figure 3 and S5). Given the large number of BGCs in each '*Ca.* Entotheonella' genome studied to date and the uniqueness of each BGC inventory (Mori et al., 2018; Reiter et al., 2020), this might suggest a wealth of biosynthetic novelty yet to be discovered. Since group I included all biosynthetically talented '*Ca.* Entotheonella' previously characterized by genomic, microscopic, and spectroscopic methods, we were intrigued by the presence of additional group I members from terrestrial habitats, including soil samples collected at two distant locations in Germany and samples derived from dump layers of the leafcutter ant *Atta colombica*. However, our attempts to mechanically enrich and visualize terrestrial '*Ca.* Entotheonella' were unsuccessful.

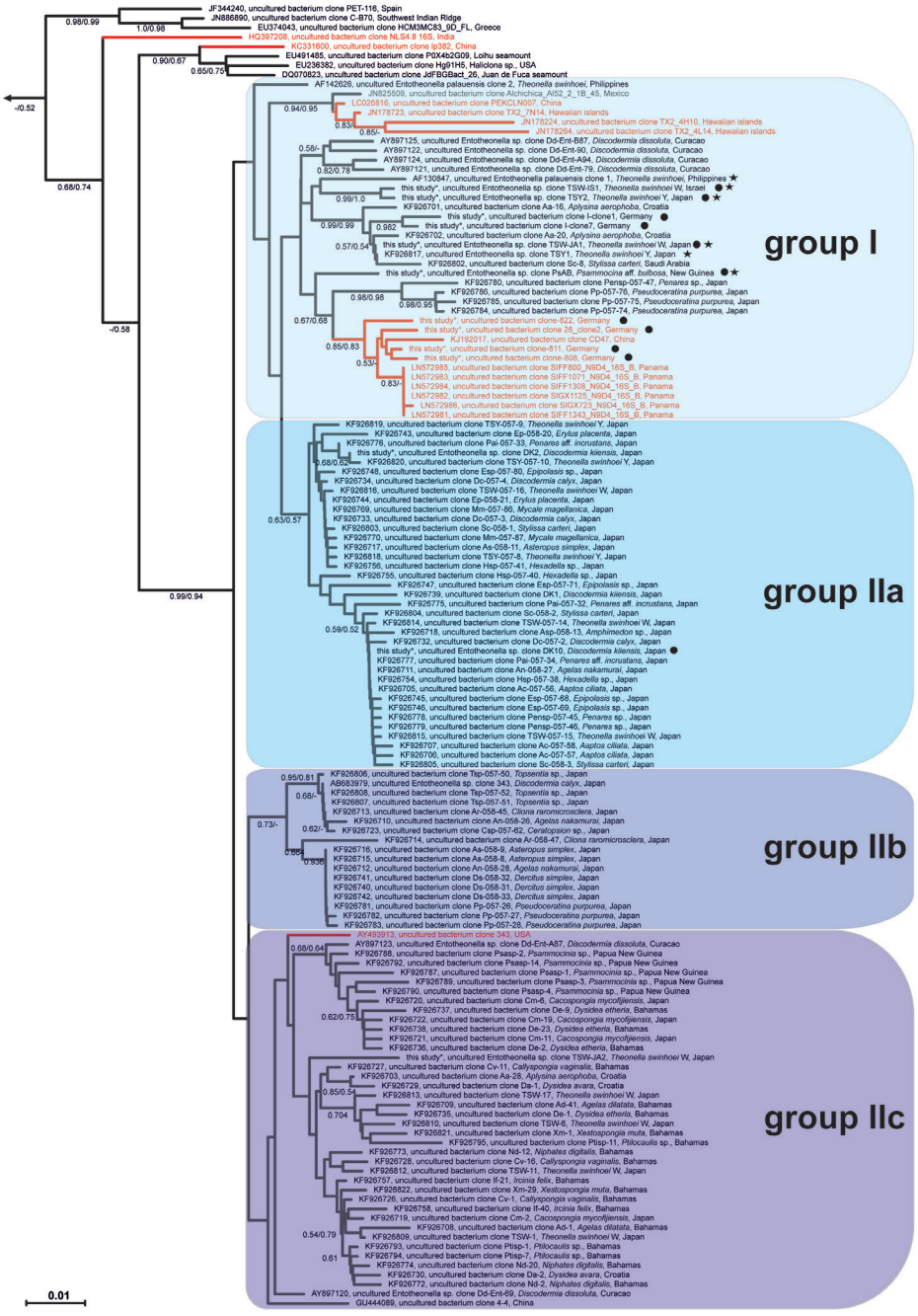


Figure 3. Phylogeny of '*Ca.* Entotheonella'. Detailed view of '*Ca.* Entotheonella' for the tree shown in Figure 1. Bootstrap values are given for both methods used, when either value is above 50% (neighbour-joining/maximum-likelihood). Sequences generated in this study are indicated with a black circle and '*Ca.* Entotheonella' bacteria previously linked to the production of bioactive natural products are highlighted with a black asterisk. Sequences derived from soil samples are highlighted in red. Scale bar, 0.01 changes per nucleotide position.

A potential '*Ca.* Entotheonella' source of the anticancer polyketide psymberin in the sponge *Psammocinia* sp.

The data on group I containing the known symbiotic '*Ca.* Entotheonella' variants raised the question of whether additional sponge-associated producers can be identified in this clade. A promising candidate was a sequence amplified from metagenomic DNA of the sponge *Psammocinia* sp. (order Dictyoceratida, family Irciniidae) with 96.1% identity to the 16S rRNA gene of '*Ca.* Entotheonella factor'. This sponge was previously shown to contain psymberin (Cichewicz et al., 2004; Pettit et al., 2004), a cytotoxin with nanomolar activity against a variety of tumour cell lines (Bielitza and Pietruszka, 2013). Psymberin belongs to the pederin-type family of polyketides that are mostly produced by symbionts from diverse bacterial phyla (Helfrich and Piel, 2016), including onnamides and theopederins in *T. swinhoei* by '*Ca.* E. factor' (Wilson et al., 2014) and mycalamides in the sponge *Mycale hentscheli* by a bacterium of the UBA10353 group (Rust et al., 2020). From a metagenomic DNA library of *Psammocinia* sp. DNA, we had previously isolated the BGC for psymberin, but the producer could not be identified at the time (Fisch et al., 2009). However, re-analysis of the genes surrounding the isolated gene cluster showed high amino acid sequence identity (81-99%) to '*Ca.* Entotheonella' proteins for three gene products (Figure S6B, Table S4), suggesting an '*Ca.* Entotheonella' symbiont as the possible source of psymberin. Initial attempts to detect filamentous bacteria in *Psammocinia* sp. by the established mechanical enrichment technique were unsuccessful, but a modified protocol involving the addition of collagenases and EGTA released copious amounts of multicellular filaments from the sponge extracellular matrix. To further characterize these filaments we performed CARD-FISH experiments using the previously reported '*Ca.* Entotheonella'-specific probe ESP-219 on 30 µm sponge tissue slices (Ueoka et al., 2015). This method resulted in selective labelling of the filamentous bacteria (Figure 4 and S6C), which we subsequently named '*Candidatus* Entotheonella consociata' for its tight association with the sponge matrix (Latin, consociata; associated). The localization of these filaments to the inner pore surfaces of the sponge resembled previous findings in *Theonella* sponges (Ueoka et al., 2015; Lackner et al., 2017). Efforts to directly sequence DNA isolated from the enriched cell pellet were unsuccessful, a phenomenon that we have also encountered for other '*Ca.* Entotheonella'-containing samples (Ueoka et al., 2015; Mori et al., 2018).

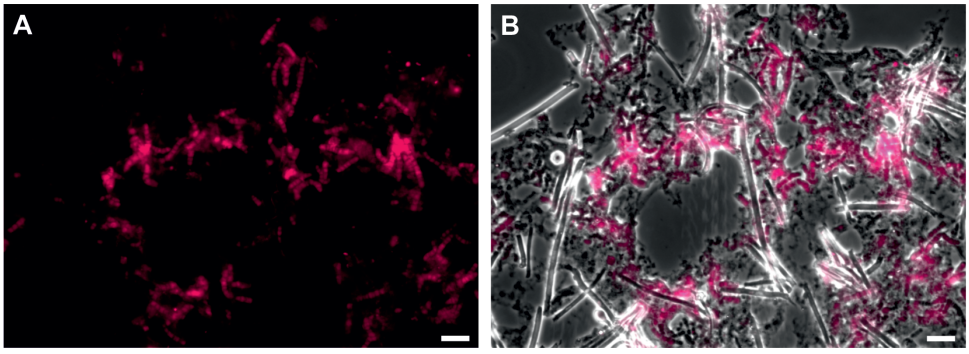


Figure 4. CARD-FISH localization of '*Ca.* E. consociata' in *Psammocinia* sp.. Overlay of a bright-field image of a representative thin slice of *Psammocinia* sp. (A) with a fluorescent image obtained from CARD-FISH labelling of '*Ca*. Entotheonella' (B). Scale bar: 20 µm.

Metagenome-assembled genomes of '*Ca.* Thalassonella' from diverse sponges lack biosynthetic richness

The lack of genomic information about '*Ca.* Tectomicrobia' other than '*Ca.* Entotheonella' is a self-reinforcing problem. Due to the lack of reference genomes against which to compare draft metagenome-assembled genomes (MAGs), these newly-sequenced MAGs could not be unambiguously assigned to the candidate phylum, or to the subordinate taxonomic groups introduced in this study but defined based on 16S rRNA gene relatedness. This can be a challenge for phylogenomic inference tools such as the GTDB (Parks et al., 2018), which assigns taxonomy based on genome similarity to this limited reference set. Several recent metagenomics papers have mentioned the presence of '*Ca.* Tectomicrobia' in metagenomes without addressing how these sequences relate to the available '*Ca.* Entotheonella' data (Miller et al., 2016; Feng et al., 2018; Bezuidt et al., 2020; Robbins et al., 2021; Tong et al., 2021). Because of their conservation, 16S rRNA gene sequences are rarely assembled successfully from metagenomes, preventing a direct comparison of these '*Ca.* Tectomicrobia' MAGs with the 16S rRNA gene-based phylogeny presented here.

Among the metagenomes of 17 sponges recently sequenced by our groups, the GTDB-Tk assigned 24 MAGs with completeness greater than 50% and no more than 10% contamination as belonging to '*Ca.* Tectomicrobia' (Table S1). Crucially, BLAST searches found that 15 of these MAGs contained 16S rRNA gene sequences with length ranging from 399 to 749 bp. These sequences were aligned to those used for phylogenetic analysis in this study. Five of these 15 MAGs clustered with outgroups of '*Ca.* Tectomicrobia', highlighting the challenge of phylogenomic assignment in a sparsely covered region of sequence space, while one diverged early from *'Candidatus* Entotheonellales' but could not be placed into a candidate genus. The remaining nine MAGs fell clearly into the candidate genus 'Thalassonella' (Table S1) (Figure 5A). In conclusion, these nine represent the first genomes from '*Ca.* Tectomicrobia' outside '*Ca.* Entotheonella'.

Finally, a phylogenomic tree was constructed in Anvi'o using these 15 MAGs for which 16S rRNA gene sequences were available, together with three '*Ca.* Entotheonella' MAGs, and three MAGs from phylum *Nitrospinae* as outgroups (Figure 5B). This tree, based on a concatenated alignment of 71 bacterial marker genes, showed remarkable concordance with the 16S rRNA gene-based tree, supporting all assignments at subgenus level within the phylum '*Ca.* Tectomicrobia' . The only exception was Gb4_35, which had an inconclusive '*Ca.* Entotheonellales' designation based on its 16S rRNA gene, but for which phylogenomics revealed that it belongs to the 'sublittoral to upper bathyal' *Geodia* group within '*Ca.* Thalassonella'. Incorporation of putative '*Ca.* Tectomicrobia' MAGs lacking 16S rRNA gene sequences identified an additional seven MAGs as '*Ca.* Thalassonella', and rejected two additional MAGs as being incorrectly assigned to '*Ca.* Tectomicrobia' (Figure S7), but did not uncover any genomes belonging to any of the other candidate '*Ca.* Tectomicrobia' genera, including '*Ca.* Entotheonella' and the yet-elusive '*Ca.* Allonella' and '*Ca.* Prasianella'.

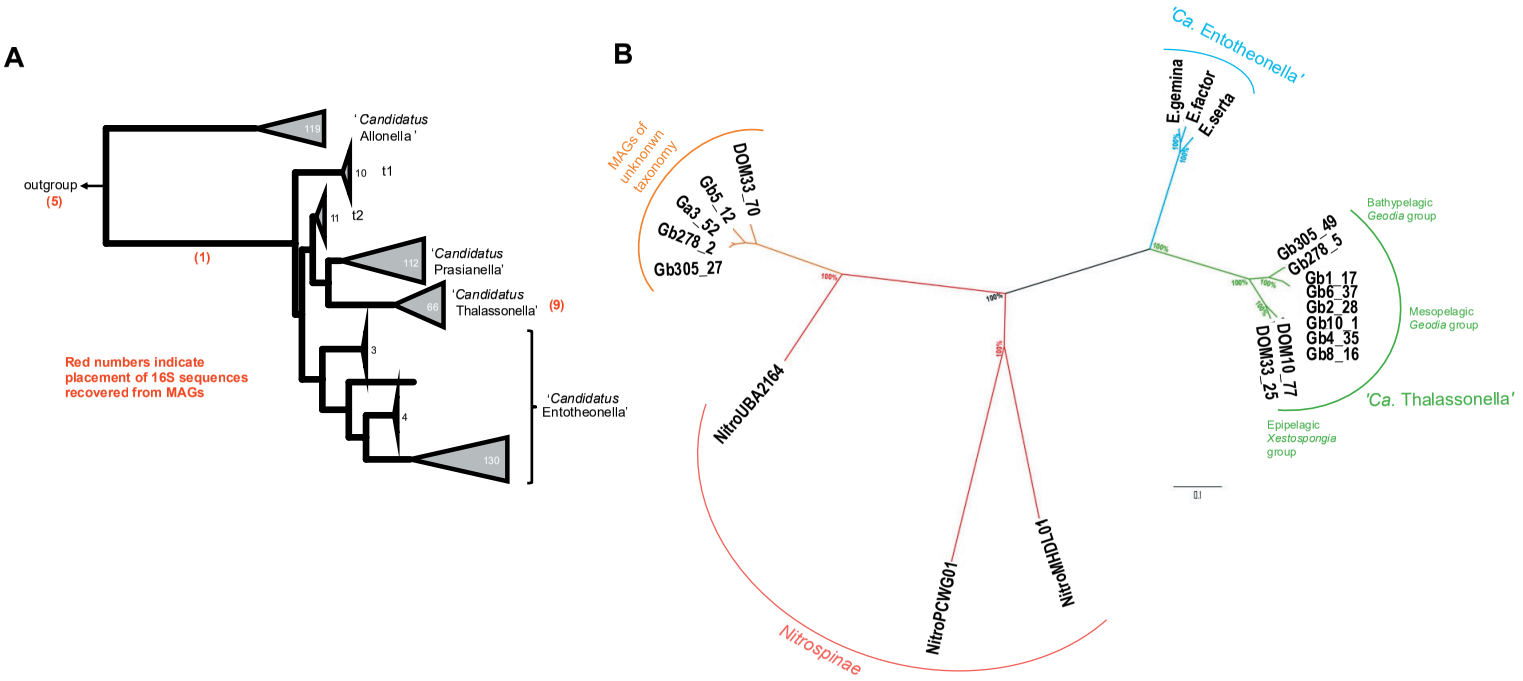


Figure 5. Analysis of putative '*Ca.* Tectomicrobia' members based on shotgun metagenome data. (A) Reproduction of Figure 1A showing the placement of 15 putative '*Ca.* Tectomicrobia' MAGs in red. (B) Unrooted phylogenomic tree of putative '*Ca.* Tectomicrobia' MAGs with 16S rRNA gene sequences including for reference three '*Ca.* Entotheonella' draft genomes ('*Ca.* E. factor', NCBI ID AZHW01; '*Ca.* E. gemina', AZHX01; '*Ca.* E. serta', PPXO01) and three draft genomes from the adjacent bacterial phylum *Nitrospinae* (MHDL01, PCWG01, DCWK01).

To assess the biosynthetic potential of '*Ca.* Thalassonella', all candidate '*Ca.* Tectomicrobia' genomes were analysed using antiSMASH. In contrast to the numerous and largely unique BGCs of '*Ca.* Entotheonella' (Wilson et al., 2014; Mori et al., 2018; Reiter et al., 2020), the '*Ca.* Thalassonella' MAGs possessed just 2-5 BGCs each, nearly all of which were conserved within the genus (Figure 6). Of particular note were two putative terpene BGCs that were also found in the three published '*Ca.* Entotheonella' draft genomes (Wilson et al., 2014; Lackner et al., 2017; Mori et al., 2018). One of these two conserved terpene BGCs was present in all '*Ca*. Thalassonella' MAGs, whereas the other was present in 11 of the 17 (Figure 6). A putative BGC for a modified lipid was present in 15 of 17 '*Ca.* Thalassonella MAGs', and a unimodular type I polyketide synthase (T1PKS) was found exclusively but universally in the 'Mesopelagic *Geodia*' subgenus grouping. Neither the lipid-modifying cluster nor the T1PKS were found in '*Ca.* Entotheonella' genomes, and none of the four conserved '*Ca.* Thalassonella' BGCs were found in the outgroup MAGs. Although the functions of these clusters are unknown, their hypothetical products could be useful molecular biomarkers for the presence of '*Ca.* Tectomicrobia' generally, in the case of the terpenes, or '*Ca.* Thalassonella' specifically, in the case of the other BGCs.

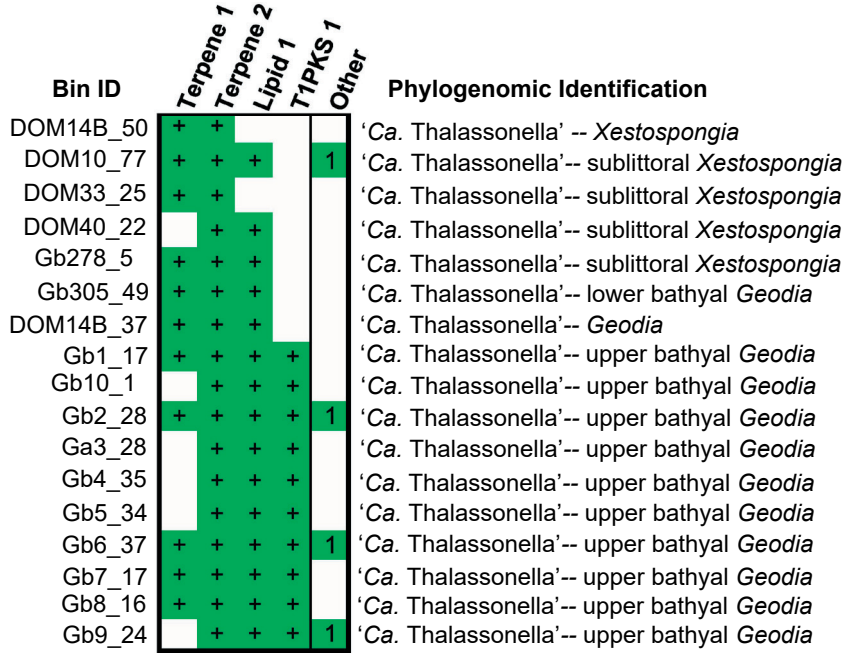


Figure 6. BGC analysis of '*Ca.* Thalassonella' MAGs. Table of conserved BGCs in 17 '*Ca.* Thalassonella'-assigned MAGs, including MAGs assigned on the basis of the phylogenomic tree in Figure S7. MAGs containing an ortholog of a cluster are indicated with a '+', and other, non-conserved clusters are shown in the final column. T1PKS, type I polyketide synthase.

'*Ca.* Thalassonella' members lead a similar lifestyle to '*Ca.* Entotheonella'

Functional characterization and metabolic pathway reconstruction of eight medium-high-quality '*Ca.* Thalassonella' MAGs revealed important clues about their primary metabolism. On average, the analysed MAGs were 87% complete and 2% contaminated with a genome size of 3.6 Mbp and 63% GC content (Table S5). Comparative genomic analysis suggested that '*Ca.* Thalassonella' has a facultative anaerobic, heterotrophic metabolism highly similar to that previously reported for '*Ca.* Entotheonella' (Lackner et al., 2017). The presence of genes coding for respiratory chain and oxygen-tolerant enzymes indicated the capacity of '*Ca.* Thalassonella' to use oxygen as the terminal electron acceptor. Various types of cytochrome c oxidases (low and high O_2 affinity) identified in the genomes support previous findings on the ability of sponge symbionts to survive under different oxygen concentrations prevailing in actively pumping sponges (Moitinho-Silva et al., 2017b; Taylor et al., 2021). Similarly to '*Ca.* Entotheonella', multiple genes encoding CoA-transferases of family III (e.g. formyl-CoA transferase, benzylsuccinate CoA-transferase) and putative pathways for pyruvate fermentation to acetoin and 2,3 butanediol hinted at a facultative anaerobic metabolism (Lackner et al., 2017).

'*Ca.* Thalassonella' MAGs harboured an almost complete set of genes for glycolysis (Embden-Meyerhof-Parnas pathway), the tricarboxylic acid (TCA) cycle and the pentose phosphate pathway (non-oxidative phase) (Table S6). It should be noted that the incompleteness of certain metabolic pathways must be interpreted cautiously as it could be attributed to sequencing or binning artifacts. However, in some cases key enzymes involved in core metabolic pathways might have alternatives performing the same function. For example, '*Ca.* Thalassonella' MAGs were missing one of the key enzymes of the TCA cycle, namely the NADP-dependent isocitrate dehydrogenase [EC: 1.1.1.42], which is found in most prokaryotes, including '*Ca.* Entotheonella' (Lackner et al., 2017). Instead, they contained the eukaryotic-type NAD-dependent isocitrate dehydrogenase [EC:1.1.1.41]. No autotrophic carbon fixation pathways were identified in any of the genomes. A heterotrophic lifestyle was also reflected in the presence of various ATP-binding cassette (ABC) transporters for amino acids, oligosaccharides, lipopolysaccharides and lipoproteins (Table S6).

Gene-based evidence suggested that '*Ca.* Thalassonella' has the capacity of using both methanol and oxalate as energy sources following the same energy acquisition strategy as '*Ca.* Entotheonella'. Methanol is considered a key fuel for marine microorganisms sourced from atmospheric deposition (Yang et al., 2013) and phytoplankton production (Mincer and Aicher, 2016). In the case of oxalate, it has been often observed in marine aerosols (Turekian et al., 2003) and in the form of calcium oxalate in sponges (Cerrano et al., 1999). '*Ca.* Thalassonella' was also predicted to produce several amino acids, cofactors and vitamins (Table S6). This biosynthetic potential has been previously reported for several sponge symbiotic lineages (Lackner et al., 2017; Engelberts et al., 2020; Robbins et al., 2021). Since sponges are not capable of synthesizing several of these compounds (e.g., arginine, histidine, vitamin B12) (Munroe et al., 2019), they are thought to obtain

them either via filter-feeding or from their associated microorganisms (Pita et al., 2018). Our findings suggest that members of the candidate genus 'Thalassonella' carry many traits typically found in sponge symbionts. Further, '*Ca.* Thalassonella' members appeared to follow a similar lifestyle to '*Ca.* Entotheonella' with the exception of the large difference in their BGC potential described above.

Conclusion

In the present study, the phylogenetic analysis of 456 16S rRNA gene sequences, including those retrieved from marine and terrestrial samples obtained by targeted sequence prospecting, demonstrated that the candidate phylum '*Ca.* Tectomicrobia' is composed of a diverse range of phylotypes globally distributed among various habitats. The phylum could be divided into several clades, some of which appeared to correlate with environmental factors such as water depth, or association with marine sponges. The analysis of the '*Ca.* Entotheonella' genus suggested the existence of a small subgroup of '*Ca.* Entotheonella' bacteria in close phylosymbiosis with sponge hosts and a larger group of bacterial phylotypes that were highly similar and showed no co-speciation with the host but rather an apparent biogeographic correlation. All of the known bioactive natural product producers in '*Ca.* Tectomicrobia' fell into the former group. This interconnection offers an interesting possibility for the prioritization of sponges harbouring group I '*Ca.* Entotheonella', and has already led to the discovery of the first potential chemically productive '*Ca.* Entotheonella' species in a non-astrophorin sponge. This symbiont, from *Psammocinia* sp., is the likely producer of the antitumor compound psymberin. Surprisingly, our studies also pointed towards the existence of bacteria within the '*Ca.* Entotheonella' superproducer group I that were not associated with marine sponges, but no identification was possible beyond the 16S rRNA gene amplicon. Further studies will be needed to evaluate whether they are as chemically productive as their sponge-associated counterparts and can be cultivated more easily. In contrast, the first draft genomes from the candidate genus 'Thalassonella' did not reveal a biosynthetic potential resembling that in '*Ca.* Entotheonella', while the primary metabolism of these genera was predicted to be highly similar. However, it remains unknown whether '*Ca.* Entotheonella' represents the only talented genus within '*Ca.* Tectomicrobia' due to the lack of additional tectomicrobial genomes spanning the remaining clades.

Acknowledgements

The authors wish to thank Roy Meoded and Eric Helfrich for support with '*Ca.* Entotheonella' detection, Shirley Pomponi for *D. dissoluta* samples, Ellen Kenchington for the Canadian *G. barretti* samples, the late Hans Tore Rapp for his repeated help to sample the Norwegian fjord *G. barretti* specimens, and Adriaan Schrier for offering his submarine to collect sponges in Dominica. Other sponge collectors: Pilar Ríos (specimens from Spain), Magnus

Tornes (shallow *G. barretti* in Norway), Alexander Plotkin (Svalbard). This research was financially supported by the Gordon and Betty Moore Foundation (#9204, DOI: https://doi.org/10.37807/GBMF9204) and the Swiss National Science Foundation (205320_185077) to JP, the European Commission through Horizon2020 project SponGES (Grant agreement ID: 679849) to DS, PC and AG and the Marie Sklodowska-Curie Individual Fellowship COSMos (Grant agreement ID: 897121) to MAS. This document reflects only the authors' view and the Executive Agency for Small and Medium-sized Enterprises (EASME) is not responsible for any use that may be made of the information it contains.

Supplementary Information

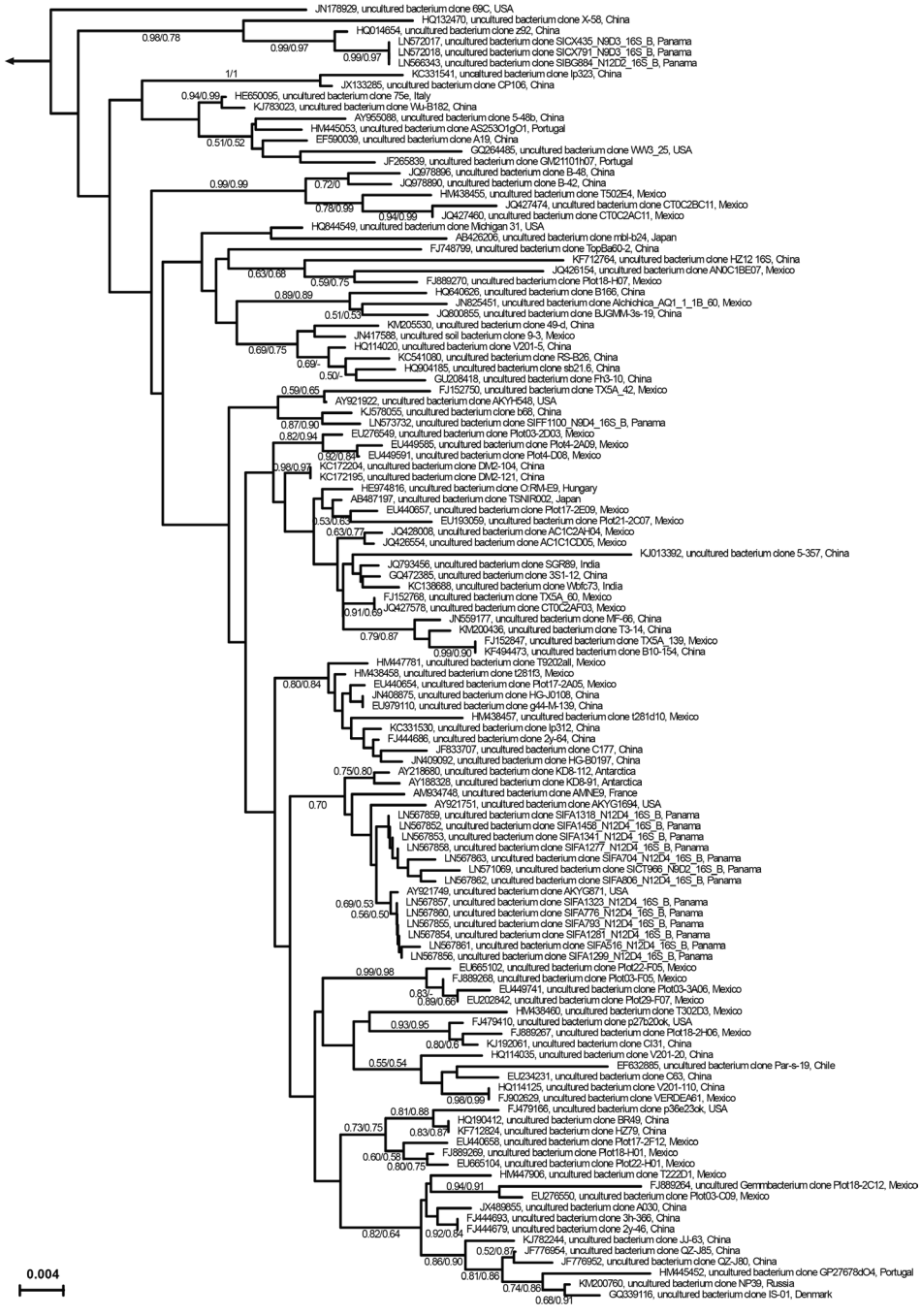


Figure S1. Phylogeny of '*Ca.* Allonella'. Detailed view of '*Ca.* Allonella' genus of the tree shown in Figure 1A. Bootstrap values for both methods used are given when either value was above 50% (neighbour-joining/maximum-likelihood). Scale bar, 0.004 changes per nucleotide position.

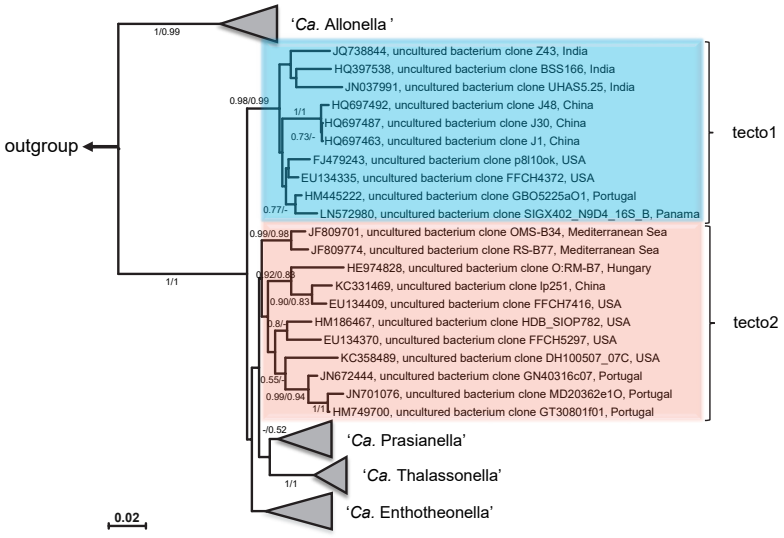


Figure S2. Phylogeny of tecto1 (t1) and tecto2 (t2) subgroups. Detailed view of t1 and t2 subgroup of the tree shown in Figure 1A. Bootstrap values for both methods used are given when either value was above 50% (neighbour-joining/maximum-likelihood). Scale bar, 0.02 changes per nucleotide position.

Figure S3. Phylogeny of '*Ca.* Prasianella'. Detailed view of '*Ca.* Prasianella' genus of the tree shown in Figure 1A. Bootstrap values for both methods used are given when either value was above 50% (neighbour-joining/maximum-likelihood). Sequences generated in this study are indicated with a black circle. Scale bar, 0.01 changes per nucleotide position.

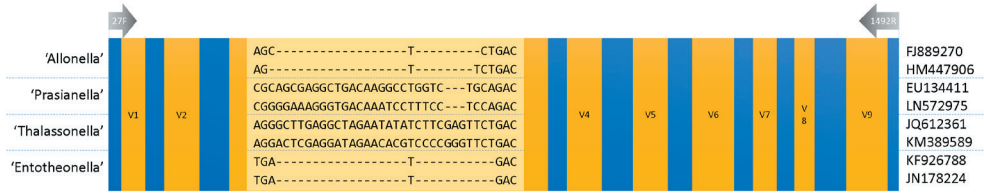


Figure S4. 16S rRNA insertion in variable region 3 (V3). Schematic of the 16S rRNA gene as sequenced by the 27F and 1492R primers, highlighting a 25-29 bp insertion in V3. Variable regions are shown in gold and conserved regions in blue, with two representative sequences shown for each named candidate genus in '*Ca.* Tectomicrobia'.

	'Allonella'	*Entotheonellaceae*	'Prasianella'	'Thalassonella'	'Entotheonella'	t1	t2	Mesopelagic *Xestospongia* 'Thalassonella'	Epipelagic *Geodia* 'Thalassonella'	Mesopelagic *Geodia* 'Thalassonella'	Bathypelagic *Geodia* 'Thalassonella'	'Entotheonella' group I	'Entotheonella' group IIa	'Entotheonella' group IIb	'Entotheonella' group IIc
'Allonella'	95.4	81.1	80.7	80.4	82.5	80.4	80.4	80.8	80.2	80.2	80.3	82.7	82.3	83.4	82.3
Entotheonellaceae	81.1	88.8	90.1	89.4	87.5	88.5	88.9	89.5	89.9	89.4	89.2	87.1	87.3	87.9	87.3
'Prasianella'	80.7	90.1	96.3	90.6	87.0	89.6	90.5	90.6	90.8	90.3	90.8	86.7	87.0	87.7	86.9
'Thalassonella'	80.4	89.4	90.6	96.6	86.0	88.9	89.1	96.5	96.5	97.2	97.2	85.7	85.6	86.4	86.3
'Entotheonella'	82.5	87.5	87.0	86.0	96.1	85.8	86.1	86.4	86.3	85.8	86.1	96.1	95.9	96.3	96.3
t1	80.4	88.5	89.6	88.9	85.8	93.9	90.3	88.8	89.3	88.7	88.8	85.7	85.8	86.0	85.6
t2	80.4	88.9	90.5	89.1	86.1	90.3	92.3	88.9	89.1	89.4	89.5	85.9	85.8	86.3	86.1
Mesopelagic *Xestospongia* 'Thalassonella'	80.8	89.5	90.6	96.5	86.4	88.8	88.9	98.4	95.6	96.5	97.0	86.1	86.0	86.9	86.6
Epipelagic *Geodia* 'Thalassonella'	80.2	89.4	90.8	96.5	86.3	89.3	89.1	95.6	99.7	96.5	96.7	85.8	85.7	86.5	86.5
Mesopelagic *Geodia* 'Thalassonella'	80.2	89.4	90.3	97.2	85.8	88.7	89.4	96.5	96.5	99.2	98.2	85.4	85.3	86.1	86.1
Bathypelagic *Geodia* 'Thalassonella'	80.3	89.2	90.8	97.2	86.1	88.8	89.5	97.0	96.7	98.2	99.6	85.7	85.4	86.7	86.5
'Entotheonella' group I	82.7	87.1	86.7	85.7	96.1	85.7	85.9	86.1	85.8	85.4	85.7	96.8	95.8	96.2	96.2
'Entotheonella' group IIa	82.3	87.3	87.0	85.6	95.9	85.5	85.8	86.0	85.7	85.3	85.4	95.8	98.7	95.8	95.6
'Entotheonella' group IIb	83.4	87.9	87.7	86.4	96.3	86.0	86.3	86.9	86.5	86.1	86.7	96.2	95.8	99.6	96.9
'Entotheonella' group IIc	82.3	87.3	86.9	86.3	96.3	85.6	86.1	86.6	86.5	86.1	86.5	96.2	95.6	96.9	98.2

Figure S5. Detailed MSI heatmap. Median sequence identities within (main diagonal) and between (off-diagonal) the clades described in this study.

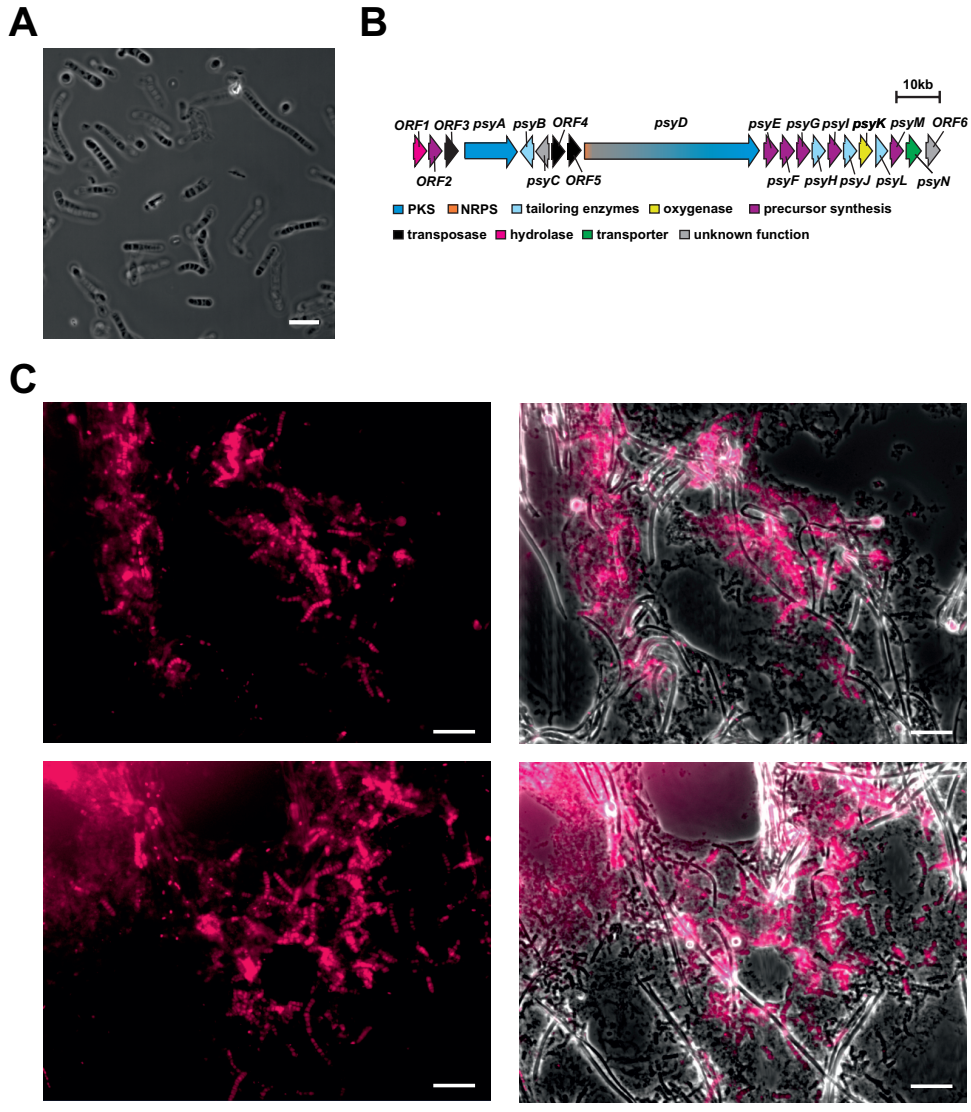


Figure S6. '*Ca.* E. consociata' in *Psammocinia* aff. *bulbosa*. Phase contrast image of cells enriched for filamentous cells and dissociated, prepared from *Psammocina* aff. *bulbosa* mainly showing '*Ca.* E. consociata'. Scale bar 5 µm. (B) Genomic organization of the psy biosynthetic gene cluster. (C) Overlay of a bright-field image of a representative thin slice of *Psammocinia* aff. *bulbosa* (right) with a fluorescent image obtained from CARD-FISH labelling of '*Ca.* Entotheonella' (left). Scale bar: 20 µm

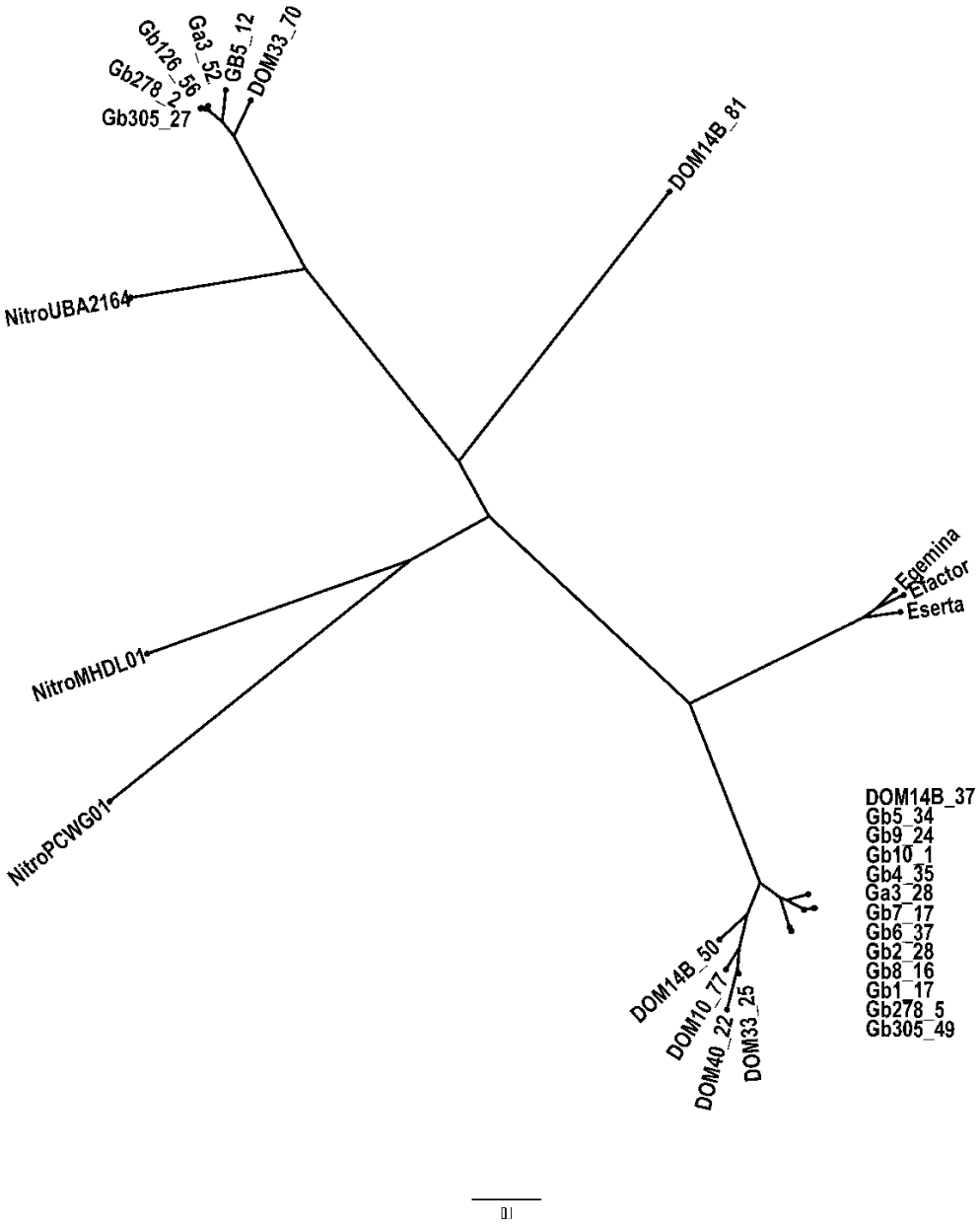


Figure S7. Expanded phylogenomic tree of '*Ca.* Tectomicrobia'. Unrooted Jukes-Cantor phylogenomic tree of putative '*Ca.* Tectomicrobia' MAGs, including those without recovered 16S rRNA gene sequences, plus for reference three '*Ca.* Entotheonella' draft genomes ('*Ca.* E. factor', NCBI ID AZHW01; '*Ca.* E. gemina', AZHX01; '*Ca.* E. serta', PPXO01) and three draft genomes from the adjacent bacterial phylum *Nitrospinae* (MHDL01, PCWG01, DCWK01).

Table S1. Metadata for sponge metagenomes that were used to reconstruct *Ca.* Tectomicrobia composite genomes.

Sponge Host	Location	Sampling Year	Code	Bin	CheckM Completeness / Contamination (%)	16S length	16S ID	Anvi'o ID
Fascaplysinopsis sp.	Portsmouth, Dominica (N15°34'33.24", W61°27'20.16", 101m)	2016	DOM10	DOM10_77	77.58 / 0.00	529	'*Ca.* Thalassonella' -Epipelagic *Xestospongia*	'*Ca.* Thalassonella' -Epipelagic *Xestospongia*
Geodia sp.	Portsmouth, Dominica (N15°34'33.24", W61°27'20.16", 230m)	2016	DOM14B	DOM14B_ 37	63.79 / 0.00			'*Ca.* Thalassonella' -*Geodia*
				DOM14B_50	92.3 / 1.71			'*Ca.* Thalassonella' -*Xestospongia*
				DOM14B_ 81	82.19 / 1.82			Outgroup (other)
Aplysina sp.	Cottage, Dominica (N15°36'54.66", W61°27'48.93", 106m)	2016	DOM33	DOM33_25	76.96 / 0.85	483	'*Ca.* Thalassonella' -Epipelagic *Xestospongia*	'*Ca.* Thalassonella' -Epipelagic *Xestospongia*
				DOM33_70	79.78 / 1.86	496	Outgroup Cluster 1	Outgroup Cluster 1
Desmacella sp.	Connor Bay, Dominica (N15°38'14.13", W61°27'39.13", 146m)	2016	DOM40	DOM40_22	57.75 / 2.59			'*Ca.* Thalassonella' -Epipelagic *Xestospongia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 450m)	2017	Gb1	Gb1_17	92.83 / 1.71	627	'*Ca.* Thalassonella' -Mesopelagic *Geodia*	'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 150m)	2017	Gb2	Gb2_28	91.54 / 1.71	699	'*Ca.* Thalassonella' -Mesopelagic *Geodia*	'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia atlantica	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 150m)	2017	Ga3	Ga3_28	65.51 / 0.00			'*Ca.* Thalassonella' -Mesopelagic *Geodia*
				Ga3_52	59.85 / 0.85	399	Outgroup cluster 1	Outgroup cluster 1
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 150m)	2017	Gb4	Gb4_35	63.79 / 0.00	574	'*Ca.* Tectomicrobia' (singleton, unknown further taxonomy)	'*Ca.* Thalassonella' -Mesopelagic *Geodia*

Table S1. Continued.

Sponge Host	Location	Sampling Year	Code	Bin	CheckM Completeness / Contamination (%)	16S length	16S ID	Anvi'o ID
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 150m)	2017	Gb5	Gb5_34	67.24 / 0.00			'*Ca.* Thalassonella' -Mesopelagic *Geodia*
				Gb5_12	75.96 / 0.85	460	Outgroup cluster 1	Outgroup cluster 1
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 150m)	2017	Gb6	Gb6_37	91.97 / 2.56	747	'*Ca.* Thalassonella' -Mesopelagic *Geodia*	'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 450m)	2017	Gb7	Gb7_17	92.02 / 1.71			'*Ca.* Thalassonella' -- Mesopelagic *Geodia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 450m)	2017	Gb8	Gb8_16	91.97 / 2.56	613	'*Ca.* Thalassonella' -Mesopelagic *Geodia*	'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 450m)	2017	Gb9	Gb9_24	65.51 / 0.00			'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia barretti	Scengsbukt-Korsfjord, Norway (N60°8'8", E5°6'42", 450m)	2017	Gb10	Gb10_1	65.51 / 0.00	496	'*Ca.* Thalassonella' -Mesopelagic *Geodia*	'*Ca.* Thalassonella' -Mesopelagic *Geodia*
Geodia barretti	Davis Strait, Canada (N62°52'15", W58°37'34", 1213m)	2015	Gb126	Gb126_56	68.75 / 0.00			Outgroup cluster 1
Geodia barretti	Davis Strait, Canada (N61°53'36", W60°7'57", 1335m)	2014	Gb278	Gb278_2	67.46 / 0.85	496	Outgroup cluster 1	Outgroup cluster 1
				Gb_278_5	92.27 / 4.27	633	'*Ca.* Thalassonella' -Bathylpelagic *Geodia*	'*Ca.* Thalassonella' -Bathylpelagic *Geodia*
Geodia barretti	Davis Strait, Canada (N61°53'36", W60°7'57", 1437m)	2014	Gb305	Gb305_27	67.62 / 0.00	588		Outgroup cluster 1
				Gb305_49	80.17 / 1.72	638	'*Ca.* Thalassonella' -Bathylpelagic *Geodia*	'*Ca.* Thalassonella' -Bathylpelagic *Geodia*

Table S2. List of soil samples used in this study. ND: Not determined.

Sample	Collection date	Source	Location	Latitude	Longitude
uncultured bacterium clone 22-clone1 *uncultured bacterium clone 22-clone7*	13.07.2012	soil	NO, Geilo	60.546011	8.24134
uncultured bacterium clone 23-clone3	14.07.2012	soil	NO, Drammen	ND	ND
uncultured bacterium clone 25-clone1 *uncultured bacterium clone 25-clone6* *uncultured bacterium clone 25-clone8*	24.07.2012	soil	DE, Meckenheim	50.629311	7.048771
uncultured bacterium clone 26-clone2 *uncultured bacterium clone 26-clone1* *uncultured bacterium clone 26-clone5*	24.07.2012	soil	DE, Meckenheim	50.629152	7.048804
uncultured bacterium clone 5-clone1 *uncultured bacterium clone 5-clone4*	26.05.2012	soil	DE, Bausendorf	50.016165	6.99383
uncultured bacterium clone 6-clone11 *uncultured bacterium clone 6-clone15* *uncultured bacterium clone 6-clone18*	03.06.2012	sediment	DE, Kalenborn	50.557967	7.000637
uncultured bacterium clone 8-clone23 *uncultured bacterium clone 8-clone27* *uncultured bacterium clone 8-clone30* *uncultured bacterium clone 8-clone4* *uncultured bacterium clone 8-clone7* *uncultured bacterium clone-808* *uncultured bacterium clone-811* *uncultured bacterium clone-822*	05.06.2012	soil	DE, Berlin	ND	ND
uncultured bacterium clone 22-clone1	13.07.2012	soil	NO, Geilo	60.546011	8.24134
uncultured bacterium clone A-clone11 *uncultured bacterium clone A-clone2*	08.09.2012	soil	DE, Neuwied	50.438631	7.486529
uncultured bacterium clone B-clone7 *uncultured bacterium clone B-clone8* *uncultured bacterium clone B-clone9*	08.09.2012	soil	DE, Neuwied	50.43848	7.486581

Table S2. Continued.

Sample	Collection date	Source	Location	Latitude	Longitude
uncultured bacterium clone C-clone6 *uncultured bacterium clone C-clone7* *uncultured bacterium clone C-clone8*	08.09.2012	soil	DE, Neuwied	50.427134	7.473694
uncultured bacterium clone D-clone4 *uncultured bacterium clone D-clone6* *uncultured bacterium clone D-clone7*	08.09.2012	soil	DE, Neuwied	50.427196	7.473873
uncultured bacterium clone E-clone1 *uncultured bacterium clone E-clone20* *uncultured bacterium clone E-clone23*	08.09.2012	soil	DE, Neuwied	50.424134	7.475721
uncultured bacterium clone F-clone1 *uncultured bacterium clone F-clone4* *uncultured bacterium clone F-clone9*	08.09.2012	soil	DE, Neuwied	50.424405	7.475612
uncultured bacterium clone H-clone1	09.09.2012	soil	DE, Neuwied	50.425845	7.457446
uncultured bacterium clone I-clone1	09.09.2012	soil	DE, Neuwied	50.426084	7.455733
uncultured bacterium clone J-clone1 *uncultured bacterium clone J-clone3* *uncultured bacterium clone J-clone7*	09.09.2012	soil	DE, Neuwied	50.425897	7.461494
uncultured bacterium clone K-clone1 *uncultured bacterium clone K-clone5* *uncultured bacterium clone K-clone6*	09.09.2012	soil	DE, Neuwied	50.42583	7.46404
uncultured bacterium clone L-clone10 *uncultured bacterium clone L-clone11* *uncultured bacterium clone L-clone2*	09.09.2012	soil	DE, Neuwied	50.425824	7.461559

Table S3. List of (partial) 16S rRNA gene sequences generated in this study.

NCBI accession	Sample	Source	Host	Origin
TBD.	uncultured bacterium clone A-clone15	soil		Germany
TBD.	uncultured bacterium clone F-clone4	soil		Germany
TBD.	uncultured bacterium clone 25-clone1	soil		Germany
TBD.	uncultured bacterium clone D-clone6	soil		Germany
TBD.	uncultured bacterium clone E-clone20	soil		Germany
TBD.	uncultured bacterium clone 26-clone5	soil		Germany
TBD.	uncultured bacterium clone 22-clone7	soil		Norway
TBD.	uncultured bacterium clone 6-clone11	soil		Germany
TBD.	uncultured bacterium clone 25-clone6	soil		Germany
TBD.	uncultured bacterium clone K-clone6	soil		Germany
TBD.	uncultured bacterium clone F-clone9	soil		Germany
TBD.	uncultured bacterium clone 8-clone30	soil		Germany
TBD.	uncultured bacterium clone J-clone3	soil		Germany
TBD.	uncultured bacterium clone 8-clone23	soil		Germany
TBD.	uncultured bacterium clone 26-clone1	soil		Germany
TBD.	uncultured bacterium clone 5-clone4	soil		Germany
TBD.	uncultured bacterium clone J-clone7	soil		Germany
TBD.	uncultured bacterium clone K-clone5	soil		Germany
TBD.	uncultured bacterium clone A-clone2	soil		Germany
TBD.	uncultured bacterium clone C-clone7	soil		Germany
TBD.	uncultured bacterium clone B-clone7	soil		Germany
TBD.	uncultured bacterium clone J-clone1	soil		Germany
TBD.	uncultured bacterium clone 6-clone18	soil		Germany
TBD.	uncultured bacterium clone GaPC3 200-400m	sponge	*Geodia atlantica*	Norway

Table S3. Continued.

NCBI accession	Sample	Source	Host	Origin
TBD.	uncultured bacterium clone GbPC618 40m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC617 39m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC619 36m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC19 100m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC835 40m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC9 200-400m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC739 ???	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GPachyPC577_K2 1172m	sponge	*Geodia pachydermata*	Norway
TBD.	uncultured bacterium clone GbPC958_K5_750-1600m	sponge	*Geodia barretti*	Davis Strait
TBD.	uncultured bacterium clone GbPC619_K10 36-100m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GpPC537 787m	sponge	*Geodia parva*	Norway
TBD.	uncultured bacterium clone GbPCWGS3 88m	sponge	*Geodia barretti*	Sweden
TBD.	uncultured Entotheonella sp. clone DK10	sponge	*Discodermia kiiensis*	Japan
TBD.	uncultured bacterium clone-808	soil		Germany
TBD.	uncultured bacterium clone-811	soil		Germany
TBD.	uncultured bacterium clone 26_clone2	soil		Germany
TBD.	uncultured bacterium clone-822	soil		Germany
TBD.	uncultured Entotheonella sp. clone TSY2	sponge	*Theonella swinhoei Y*	Japan
TBD.	uncultured bacterium clone I-clone1	sponge		Germany
TBD.	uncultured bacterium clone 8-clone4	soil		Germany
TBD.	uncultured bacterium clone E-clone23	soil		Germany
TBD.	uncultured bacterium clone D-clone7	soil		Germany
TBD.	uncultured bacterium clone D-clone4	soil		Germany
TBD.	uncultured bacterium clone 25-clone8	soil		Germany

Table S3. Continued.

NCBI accession	Sample	Source	Host	Origin
TBD.	uncultured bacterium clone 5-clone1	soil		Germany
TBD.	uncultured bacterium clone 22-clone1	soil		Norway
TBD.	uncultured bacterium clone 8-clone7	soil		Germany
TBD.	uncultured bacterium clone E-clone1	soil		Germany
TBD.	uncultured bacterium clone H-clone1	soil		Germany
TBD.	uncultured bacterium clone A-clone11	soil		Germany
TBD.	uncultured bacterium clone C-clone6	soil		Germany
TBD.	uncultured bacterium clone 23-clone3	soil		Norway
TBD.	uncultured bacterium clone F-clone1	soil		Germany
TBD.	uncultured bacterium clone C-clone8	soil		Germany
TBD.	uncultured bacterium clone B-clone8	soil		Germany
TBD.	uncultured bacterium clone B-clone9	soil		Germany
TBD.	uncultured bacterium clone 6-clone15	soil		Germany
TBD.	uncultured bacterium clone L-clone2	soil		Germany
TBD.	uncultured bacterium clone L-clone11	soil		Germany
TBD.	uncultured bacterium clone 8-clone27	soil		Germany
TBD.	uncultured bacterium clone L-clone10	soil		Germany
TBD.	uncultured bacterium clone K-clone1	soil		Germany
TBD.	uncultured bacterium clone GpPC12 200-400m	sponge	*G.phlegraei*	Norway
TBD.	uncultured bacterium clone GaPC635 750m	sponge	*Geodia atlantica*	Ireland
TBD.	uncultured bacterium clone GaPC983 907m	sponge	*Geodia atlantica*	Greenland
TBD.	uncultured bacterium clone GbPC958 1462m	sponge	*Geodia barretti*	Greenland
TBD.	uncultured bacterium clone GaPC624 1137m	sponge	*Geodia atlantica*	Canada
TBD.	uncultured bacterium clone GbPC961 769m	sponge	*Geodia barretti*	Greenland

Table S3. Continued.

NCBI accession	Sample	Source	Host	Origin
TBD.	uncultured bacterium clone GpPC640 139m	sponge	*Geodia phlegraei*	Scotland
TBD.	uncultured bacterium clone GhPC555 688m	sponge	*Geodia hentscheli*	Norway
TBD.	uncultured bacterium clone GbPC738 200-400m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC738_K5 200-400m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC12_K9 200-400m	sponge	*Geodia phlegraei*	Norway
TBD.	uncultured bacterium clone GpPC511 215m	sponge	*Geodia phlegraei*	Norway
TBD.	uncultured bacterium clone GbPC529 410m	sponge	*Geodia barretti*	Greenland
TBD.	uncultured bacterium GbPC960_823m	sponge	*Geodia barretti*	Davis Strait
TBD.	uncultured bacterium GmPC510_215m	sponge	*Geodia macandrewii*	Norway
TBD.	uncultured bacterium GmPC509_215m	sponge	*Geodia macandrewii*	Norway
TBD.	uncultured bacterium GbPC514_215m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium GbPC513_215m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium GbPC512_215m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC508 215m	sponge	*Geodia barretti*	Norway
TBD.	uncultured bacterium clone GbPC510 215m	sponge	*Geodia barretti*	Norway
TBD.	uncultured Entotheonella sp. clone DK2	sponge	*Discodermia kiiensis*	Japan
TBD.	uncultured Entotheonella sp. clone TSW-IS1	sponge	*Theonella swinhoei W*	Israel
TBD.	uncultured bacterium clone I-clone7	soil		Germany
TBD.	uncultured Entotheonella sp. clone TSW-JA1	sponge	*Theonella swinhoei W*	Japan
TBD.	uncultured Entotheonella sp. clone TSW-JA2	sponge	*Theonella swinhoei W*	Japan
TBD.	uncultured Entotheonella sp. clone PsAB	sponge	*Psammocina* aff. *bulbosa*	New Guinea

Table S4. ORFs detected on the loci containing the *psy* genes and their putative functions.

ORF	Proposed Function	Closest Homolog (source organism)	% AA Identity	Accession Number
ORF1	Zn-dependent hydrolase	MBL fold metallo-hydrolase (*Candidatus* Entotheonella palauensis)	81	WP_089939010.1
ORF2	Adenylosuccinate synthetase	Adenylosuccinate synthase (*Candidatus* Entotheonella palauensis)	94	WP_089939008
ORF3	Transposase	IS4 family transposase (*Phormidesmis priestleyi*)	46	WP_073074803.1
PSYA	Trans-AT PKS	PedI (*Pederus fuscipes* symbiont)	41	AAR19304.1
PSYB	Methyltransferase	PedA (*Pederus fuscipes* symbiont)	52	AAS47557.1
PSYC	pedK-like	OnnF (*Candidatus* Entotheonella factor)	42	AAV97874.1
ORF4	Transposase	IS4 family transposase (*Phormidesmis priestleyi*)	46	WP_073074803.1
ORF5	Transposase	Hypothetical protein (*Fimbriiglobus ruber*)	43	WP_088252106.1
PSYD	PKS-NRPS	PedF (*Pederus fuscipies* symbiont)	44	AAS47564.1
PSYE	Phosphoenolpyruvate synthase β and γ subunit	Phosphoenolpyruvate synthase β subunit/ γ subunit (*Discodermia dissoluta* symbiont)	93/92	AAY00048.1/ AAY00046.1
PSYF	Phosphosterase	Phosphoesterase SA1_PKS A (*Discodermia dissoluta* symbiont)	96	AY907538
PSYG	Phosphoenolpyurvate synthase α subunit	Phosphoenolpyruvate synthase α subunit (*Discodermia dissoluta* symbiont)	95	AAY00044.1
PSYH	acyltransferase	Acyltransferase (*Discodermia dissoluta* symbiont)	93	AAY00043.1
PSYI	HMG-CoA-synthase	HMG-CoA-synthase (*Discodermia dissoluta* symbiont)	98	AAY00042.1
PSYJ	Crotonase superfamily	Crotonase superfamily (*Discodermia dissoluta* symbiont)	95	AAY00041.1
PSYK	Flavin-dependent oxygenase	Flavin dependent oxygenase (*Discodermia dissoluta* symbiont)	96	AAY00040.1
PSYL	ACP	Crotonase superfamily (*Discodermia dissoluta* symbiont)	99	AAY00039.1
PSYM	3-oxoacyl ACP synthase	3-oxoacyl ACP synthase (*Discodermia dissoluta* symbiont)	96	AAY00038.1
PSYN	Cation transport ATPase	Cation transport ATPase (*Discodermia dissoluta* symbiont)	95	AAY00037.1
ORF6	unknown	hypothetical protein (*Candidatus* Entotheonella factor)	61	ETW92637.1

Table S5. '*Ca*. Thalassonella' MAG statistics and isolation source information used for prediction of metabolic pathways.

MAG	Source	#Contigs	Genome size (bp)	GC content (%)	N50 (bp)	Completeness (%)	Contamination (%)	#Genes
Gb1_17	*Geodia barretti*	206	3,799,105	62.0	39,894	92.8	1.7	3,680
Gb2_28	*Geodia barretti*	239	3,975,034	62.0	43,132	91.5	1.7	3,846
Gb6_37	*Geodia barretti*	238	3,978,631	61.9	46,811	92.0	2.6	3,887
Gb8_16	*Geodia barretti*	181	3,896,062	61.9	47,852	92.0	2.6	3,744
Gb278_5	*Geodia barretti*	167	3,628,785	62.2	81,152	92.3	4.3	3,494
Gb305_49	*Geodia barretti*	173	3,559,325	62.4	78,347	80.2	1.7	3,428
DOM10_77	*Fascaplysinopsis reticulata*	380	3,220,299	65.8	112,146	77.6	0.0	3,178
DOM33_25	*Aplysina cauliformis*	300	2,731,884	66.6	14,938	77.0	0.9	2,611

Table S6 can be accessed using the following link: https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis

Chapter 5

Candidatus Nemesobacterales, a sponge-specific clade of the candidate phylum Desulfobacterota adapted to a symbiotic lifestyle

Asimenia Gavriilidou, Burak Avci, Catarina Loureiro, Anastasia Galani, Michelle A. Schorn, Thijs J.G. Ettema, Colin J. Ingham, Hauke Smidt, Detmer Sipkema

Manuscript in preparation

Abstract

Members of the candidate phylum Dadabacteria, recently reassigned to the phylum *Candidatus* Desulfobacterota*,* are cosmopolitan in the marine environment, and are found both free-living and associated with hosts that are mainly marine sponges. Yet, these microorganisms are poorly characterised, showing an ambiguous phylogeny and having no cultured representatives. Here, we performed genome-centric metagenomics to elucidate the phylogeny and predict the metabolism of the sponge-associated members of this lineage. Rank-based phylogenomics revealed several new species and a novel family (*Candidatus* Spongomicrobiaceae) within a sponge-specific order, named here *Candidatus* Nemesobacterales (GTDB order RKRQ01). Metabolic reconstruction suggests that *Ca.* Nemesobacterales are aerobic heterotrophs, capable of synthesizing most amino acids, vitamins and cofactors and degrading complex carbohydrates. We also predict functional divergence between sponge- and seawater-associated metagenome-assembled genomes (MAGs). Niche-specific adaptations to the sponge holobiont were evident from significantly enriched genes involved in defense mechanisms against foreign DNA and environmental stressors, host-symbiont interactions and secondary metabolite production. Fluorescent in situ hybridization (FISH) gave a first glimpse of the morphology and lifestyle of a member of *Ca.* Desulfobacterota. The *Candidatus* Nemesobacter rappii was found inside bacteriocytes in the tissue of the sponge *Geodia barretti*. Altogether, this is the first study that sheds light on the enigmatic group *Ca.* Nemesobacterales and their functional characteristics which reflect a symbiotic lifestyle.

Introduction

The candidate phylum Dadabacteria, formerly known as SBR1093, was initially detected in 16S ribosomal RNA (rRNA) gene cloning experiments from activated sludge in sequential batch reactors (Bond et al., 1995). In 2014, the first draft genome was reconstructed from industrial-activated sludge samples using metagenome binning methods (Wang et al., 2014). Hug et al. (2016) later established SBR1093 as a novel phylum under the name 'Dadabacteria' (Hug et al., 2016a). To date, this lineage remains uncultured and understudied displaying an unresolved phylogeny. Waite et al. (2020) recently proposed the assignment of members of the *Candidatus* Dadabacteria to the candidate phylum Desulfobacterota. However, phylogenetic analyses based on concatenated ribosomal proteins did not recover *Ca.* Dadabacteria as monophyletic group with members of the *Ca.* Desulfobacterota phylum (Hug et al., 2016a; Waite et al., 2020) highlighting the lack of confidence in assigning a taxonomic rank to this lineage. Currently, both NCBI (NCBI Taxonomy Browser, June 2021) and SILVA databases (r138.1) contain sequences of this group classified as *Ca.* Dadabacteria. Due to the explosion of novel taxa stemming from genome-centric metagenomics, taxonomic assignment of MAGs has shifted towards whole-genome based approaches to improve the classification of microorganisms recalcitrant to laboratory culture (Konstantinidis et al., 2017; Hugenholtz et al., 2021). In congruence with databases such as the Genome Taxonomy Database (GTDB), the candidate phylum Dadabacteria was united with the phylum *Ca.* Desulfobacterota and remains without designated nomenclature type referred to as 'Desulfobacterota__D' (GTDB releases 95 and 202) (Parks et al., 2020). For the purpose of the comparative genomics study described in this chapter, we decided to follow the classification as proposed by GTDB (Parks et al., 2020). According to the latest version of GTDB (r202), the candidate phylum Dadabacteria is placed within *Ca.* Desulfobacterota as a class-level linage (UBA1144) with three order- (UBA1144, UBA2774 and RKRQ01) and six family-level lineages (UBA1144, UBA2774, CSP1-2, N074bin48, RKRQ01 and GCA-014075295).

Members of this lineage are, currently, all uncultured bacteria, and their MAGs have been recovered from various habitats (free-living and host-associated), such as hydrothermal, terrestrial, human-made ecosystems, seawater and marine sponges (Graham and Tully, 2021). Most marine sponges harbour dense and phylogenetically diverse microbial consortia, which span all domains of life and have tight associations with their hosts (Thomas et al., 2016). These microbe-host associations have been coined as the 'sponge holobiont' (Webster and Taylor, 2012; Pita et al., 2018). Several genome-centric studies have shed light on the functional role of the sponge microbiome highlighting significant attributes of their metabolism, including elemental cycling, nutrient supply and host defense (Diez-Vives et al., 2017; Lackner et al., 2017; Slaby et al., 2017; Astudillo-Garcia et al., 2018; Podell et al., 2019; Engelberts et al., 2020; Sizikov et al., 2020; Robbins et al., 2021; Taylor et al., 2021). However, most of these studies have focused primarily on characterizing highly abundant sponge-associated lineages or the interactions within the holobiont of specific sponge

species. Therefore, less abundant members of sponge microbiomes and their ecological functions have been largely overlooked. Members of *Ca*. Dadabacteria are a perfect example of that. Microbial community profiling of 81 sponge species using 16S rRNA gene amplicon sequencing data showed that *Ca.* Dadabacteria sequences were detected in 78 out of those species, albeit at a low relative abundance (<1%) (Thomas et al., 2016). Despite their ubiquity, no studies have investigated their role in the sponge microbiome.

To address this gap, we resolved the phylogeny of this lineage following a rank-normalised taxonomy by expanding their genomic representation with newly generated genomes recovered from different sponge species. To better understand the ecological importance of this group, the relative abundance of the MAGs was estimated and a meta-analysis was conducted using 16S rRNA gene sequences to assess their environmental distribution. Metabolic reconstruction aimed at elucidating the functional repertoire and symbiotic features of the sponge-derived MAGs. Comparative genomic analysis was further performed among members of this lineage isolated from sponges and seawater to examine potential mechanisms underlying habitat adaptation. Finally, we employed FISH to visualize *Ca.* Nemesobacterales inhabiting sponges for the first time.

Materials and Methods

Sample collection and processing

For metagenomic sequencing, *Geodia barretti* (GB1, GB2 and GB4-10) and *Geodia atlantica* (Ga3) specimens were collected from Korsfjord, Norway (60°8.13’ N, 5°6.7’ E) in September and October 2017 by dredging onboard R/V Hans Brattström of the University of Bergen. In addition, *Petrosia ficiformis* samples (PF4, PF5, PF7, PF9, PF11, PF12) were collected by SCUBA-diving from a semi-submerged cave in Sfakia, Greece (35°12.0123’ N, 24°7.1633’ E). All of the aforementioned samples were dipped into liquid nitrogen and stored at -80 °C. Collection of specimens (GB126, GB278 and GB305) of *G. barretti* was performed by trawling from Davis Strait, Canada (62° 52’ 15.096’’ N, 58° 37’ 34.32’’ W, 62° 31’ 6.24’’ N, 59° 58’ 13.872’’ W, 61° 53’ 36.132’’ N, 60° 7’ 57.612’’ W, respectively) in September and October from 2011 to 2015. The sampling cruises were conducted by Fisheries and Oceans Canada onboard R/V Pâmiut of the Greenland Institute of Natural Resources. These sponge samples were stored at -20 °C until further processing. The DOM sponges (DOM10, DOM14B, DOM33, DOM40 and DOM43) were collected from Dominica, 100-200 m depth near the shore, by a submersible and stored at -20 °C in RNAlater (Thermo Fisher, USA). Samples of *Aplysina aerophoba* were collected and processed as described in (Chaib De Mares et al., 2018). Additional information can be found in Table S1.

For FISH, *G. barretti* sponges were dissected in cubes of ~1 cm^3 and fixed in 4% formaldehyde in Dulbecco’s Phosphate Buffer Saline (DPBS) overnight at 4 °C. Washing steps with DPBS, glycine (50 mM) in DPBS, and MilliQ H_2O followed. All buffers were pre-treated with Diethyl

pyrocarbonate (DEPC) (0.1%) for RNase inhibition. Sponge samples were dehydrated in a 30% and 50% ethanol series and stored in 50% ethanol at -20 °C.

DNA extraction and sequencing

All *G. barretti*, *G. atlantica*, *A. aerophoba* and *P. ficiformis* samples were pulverized in liquid nitrogen using mortar and pestle. For the bead-beating step, 200 mg of sponge tissue were disrupted by 5 stainless steel beads of 2 mm diameter and 2 beads of 5 mm diameter. Shaking followed (2 x 20 s at 4,000 rpm) using a Precellys 24 tissue homogenizer (Bertin GmbH, Germany) according to the protocol of Roume et al. (2013). DNA was then extracted from the tissue lysate using the AllPrep DNA/RNA/Protein Mini Kit (Qiagen, Germany). DNA samples of *G. barretti* and *P. ficiformis* were sent for sequencing at Novogene (Hong Kong) with the Illumina HiSeq PE150 platform. For the *A. aerophoba* samples, sequencing was performed by the Research Group Genome Analytics (GMAK) at DSMZ (Braunschweig, Germany) using the Illumina HiSeq PE100 platform. In addition, DNA from *A. aerophoba* sponges was processed for Pacific Biosciences (PacBio, Menlo Park, CA, USA) long-read sequencing according to Goethe et al. (2020) at DSMZ. For the pre-processing, DOM samples were thawed and rinsed with artificial seawater. The tissue was then added in a PowerBead Tube and shaking followed (3 x 60 s at 6,000 x g) using a Precellys 24 Homogenizer (Bertin GmbH). DNA extraction was performed with the DNeasy PowerSoil Pro Kit (Qiagen, Germany). Sequencing was conducted by Novogene Europe (United Kingdom) using the Illumina Novaseq6000 platform.

Metagenome assembly, binning and classification

Quality control of raw reads was performed with FASTQC 0.11.4 (Andrews, 2010). The BBTools suite 37.64 (Bushnell et al., 2017) was employed for adapter removal and quality filtering (parameters: ktrim=r k=23 mink=7 hdist=1 tpe tbo qtrim=rl trimq=20 ftm=5 maq=20 target=200). Reads were further normalized with BBNorm (target=100 min=5 for *P. ficiformis* and target=200 min=3 for the rest of the samples. Sequencing artifacts and phiX contamination were removed using BBDuk (Bushnell et al., 2017). Filtered reads (*G. barretti*, *P. ficiformis* and DOM sponge samples) were assembled with SPAdes 3.12 (Nurk et al., 2017) with the --meta and --only-assembler modes. For *A. aerophoba*, Illumina HiSeq and PacBio filtered reads were used to make a hybrid assembly with SPAdes 3.12 (--meta and --only-assembler) (Nurk et al., 2017).

Contigs were binned using the metaWRAP 1.2 (Uritskiy et al., 2018) binning module with metaBAT2 2.9.1 (Kang et al., 2019) and MaxBin2 2.2.4 (Wu et al., 2014) for *Geodia*reads and metaBAT2 2.9.1 (Kang et al., 2019), MaxBin2 2.2.4 (Wu et al., 2014) and CONCOCT 0.4.0 (Alneberg et al., 2014) for the rest of the sponge samples. The best version of each bin based on completeness and contamination metrics was selected using metaWRAP's bin refinement module (70% minimum completeness, 5% maximum contamination) in accordance with the Minimum Information about a Metagenome-Assembled Genome (MIMAG) standards

(Bowers et al., 2017). The refined bin set was subjected to individual reassembly (metaWRAP-Reassemble_bins module) with SPAdes 3.10.1 (Bankevich et al., 2012).
Classification of the obtained MAGs was done with the Genome Taxonomy Database-Tool Kit 1.5.0 (GTDB-Tk) *classify* workflow (with reference to GTDB R06-RS202). Only MAGs classified as 'Desulfobacterota__D' were used for downstream analysis. To confirm the genome-based classification, 16S rRNA gene sequences were retrieved from the sponge-associated MAGs with RNAmmer 1.2 (Lagesen et al., 2007). SINA web aligner 1.2.11 (Pruesse et al., 2012) was used to align the MAG-derived 16S rRNA gene sequences which were later imported into ARB 6.0.2 (Westram et al., 2011) containing the SILVA 138 SSU Ref Nr99 database (Quast et al., 2013). The alignment was manually refined in ARB and added by parsimony into the SILVA guide tree, allowing the classification of the sequences of interest and the identification of related sequences that were selected for subsequent phylogenetic analysis.

MAG collection and phylogenetic analysis

To ensure the robust representation of the studied lineage, all genomes publicly available at NCBI (February 2021) under the name *Ca.* Dadabacteria were included in the initial dataset together with MAGs generated here and in previous studies (Table S2 and S3). An additional quality control step was performed with CheckM 1.1.2 (Parks et al., 2015) and thus, the final dataset consisted of MAGs with >80% estimated completeness and <5% contamination.
Phylogenetic inference was performed based on a set of 120 single-copy marker protein sequences identified in the MAGs and aligned using the GTDB-Tk 1.5.0 (Chaumeil et al., 2019) with other members of the Desulfobacterota phylum as the outgroup. The resulting concatenated alignment was trimmed with trimAl 1.4.1 (Capella-Gutierrez et al., 2009) using default settings. A maximum-likelihood tree was generated from the trimmed alignment with IQ-TREE 2.0.6 (Minh et al., 2020) using extended model selection (-m MFP) (Kalyaanamoorthy et al., 2017) with 1000 bootstrap replicates. To confirm the phylogenetic position of the novel genomes, a 16S rRNA gene-based approach was also employed. MAG-derived and closely related 16S rRNA gene sequences identified as described above were aligned using MAFFT 7.453 (Katoh and Standley, 2013). Representative sequences of the *Chloroflexota* phylum were used as outgroup. The alignment was manually trimmed in BioEdit 7.2.5 (Hall, 1999) and resulted in a maximum-likelihood tree built with FastTree 2.1.11 (Price et al., 2010) using default settings. Both trees were annotated and visualised in iTOL 6.3 (Letunic and Bork, 2021) and Inkscape 0.92.3.
In order to assign taxonomic ranks to the newly generated MAGs, we employed a rank normalization approach based on relative evolutionary divergence (RED) which is the metric used for the manual curation of GTDB (Parks et al., 2020; Rinke et al., 2021). First, a phylogenetic tree was inferred from a concatenated alignment of 120 bacterial single-copy marker genes identified in all GTDB genomes spanning the bacterial domain and the ones generated in this study. This was performed with GTDB-Tk 1.5.0 (Chaumeil et al., 2019) *classify* workflow and the tree was built with FastTree 2.1.11 (Price et al., 2010) under the

WAG + GAMMA model with 100 bootstrap replicates. The inferred tree and the taxonomic information obtained from the *classify* command were used as input for PhyloRank 0.1.10 (https://github.com/dparks1134/PhyloRank) in order to calculate the RED of the taxa of interest. Type material was selected according to the quality criteria proposed by (Chuvochina et al., 2019). Presence of 23S, 5S and tRNAs was detected using Barrnap 0.9 (https://github.com/tseemann/barrnap).

Global distribution

The distribution of representatives of the studied lineage across different environments was estimated as described in Zhou et al. (2020). Briefly, the longest 16S rRNA gene sequences acquired from MAGs generated in this study were employed for homology-based search against the 16S rRNA Public Assembled Metagenomes database (01-01-2021) of the Integrated Microbial Genomes and Microbiomes (IMG/M 5.4) system (Chen et al., 2019), using the IMG BLAST Tool (February 2021). BLAST hits with sequence identity greater than 75% (phylum threshold) (Yarza et al., 2014) and the respective metadata (longitude, latitude, habitat type) were downloaded from IMG for further analysis. Map visualization was performed in RStudio 1.4.1106 (RStudio Team, 2020) using R 4.0.3 (R Core Team, 2020) and the R package ggplot2 3.3.3 (Wickham, 2016).

Metabolic reconstruction and comparative genomics

The MAGs were first dereplicated at 95% average nucleotide identity (ANI) with dRep 2.6.2 (Olm et al., 2017) and then their relative abundance in the sponge metagenomes was estimated with CoverM 0.6.1 (https://github.com/wwood/CoverM). Dereplicated MAGs were annotated with EnrichM 0.5.0 (https://github.com/geronimp/enrichM) (*annotate* function) using the following databases: Kyoto Encyclopaedia of Genes and Genomes (KEGG) Orthologies (KOs) for metabolic reconstruction, Pfam and TIGRFAM for detecting proteins of interest. KEGG module completeness was estimated with EnrichM's *classify* function, and only modules of completeness greater than 70% were taken for further consideration. Pfams enriched within and between MAGs from different environments (sponges vs. seawater) were identified with EnrichM's *enrichment* function. Only significantly enriched Pfams ($p<0.05$, after Benjamini-Hochberg correction) were considered for later comparison. Annotation of carbohydrate-active enzymes (CAZymes) was done by HMMER 3.0 (Eddy, 2018) against the dbCAN2 database (r9.0) (Zhang et al., 2018). Secondary metabolite biosynthetic gene clusters were identified in the MAGs with the online tool 'Antibiotics and secondary metabolite analysis shell' (antiSMASH 6.0) (Blin et al., 2021b) applying 'relaxed' detection strictness.

Statistical analysis and data visualization

Data were analysed and visualised with R 4.0.3 (R Core Team, 2020) in RStudio 1.4.1106 (RStudio Team, 2020) using the R packages vegan 2.5.7 (Oksanen et al., 2018) and ggplot2 3.3.3 (Wickham, 2016). To assess the functional variation between sponge and seawater MAGs, Bray-Curtis dissimilarity matrices were generated based on the frequency of KOs, Pfams and TIGRFAMs with the 'vegdist' function of the vegan R package. Differences in the functional profiles between the two groups were evaluated by Permutational Multivariate Analysis of Variance (PERMANOVA) and Non-metric Multidimensional Scaling (NMDS) using the 'adonis2' and the 'metaMDS' function of the vegan R package, respectively. Ordination plots and heatmaps were generated with ggplot2 3.3.3 (Wickham, 2016).

Oligonucleotide probe design

16S rRNA gene sequences recovered from the *Geodia* MAGs and classified as *Ca.* Nemesobacterales (*Candidatus* Nemesobacter rappii) were aligned in SILVA Incremental Aligner (Pruesse et al., 2012) and manually curated in ARB 6.0.6 according to rRNA secondary structure (Ludwig et al., 2004). A phylogenetic tree including all sequences in SILVA SSURef NR 99 138 (Quast et al., 2013) classified as *Ca.* Dadabacteria was reconstructed with the RAxML 7 (Stamatakis, 2006) maximum likelihood method (GTR-GAMMA rate distribution model, rapid bootstrap algorithm, 100 repetitions) using a 50% positional conservation filter for all sequences (Figure S5A). Oligonucleotide probe design was performed in the probe design tool of ARB 6.0.6 using the SILVA database 138 (Figure S5B). Optimum formamide concentration for the probe was determined in silico using the formamide curve generator tool implemented in mathFISH (Yilmaz et al., 2011).

Tissue processing, FISH and microscopy

Processing of sponge tissue was performed following the methods of Radax et al. (2012). Sponge cubes of *G. barretti* were washed briefly with sterile MilliQ H_2O and embedded in cryomedium (KP-CryoCompound, Immunologic, VWR International B.V., The Netherlands) overnight at 4 °C for infiltration. They were then embedded in fresh cryomedium in base moulds to solidify overnight at -80 °C. Longitudinal sections of 5-8 µm were made using a cryostat (Leica CM 3050 S, Leica Biosystems GmbH, Nussloch, Germany) with both chamber and object temperature set at -35 °C. Tissue sections were mounted on polylysine coated glass slides (Thermo Fisher Scientific, Gerhard Menzel B.V. & Co. KG, Germany) and stored at -20 °C.

FISH was performed as described previously (Schimak et al., 2016) with minor modifications. Briefly, the slides were dehydrated in decreasing ethanol series of 95, 80 and 70% (v/v) for 10 min each. They were then washed in 200 mM HCl for 10 min and 20 mM Tris-HCl for 10 min. For permeabilization, the sections were incubated in 1 µg/ml proteinase K (Sigma–Aldrich) solution for 5 min at 37 °C and washed into 20 mM Tris-HCl for 10 min. Hybridization was done using 4-times-Atto594-labelled DADA393 probe (5'-TCA TCC CTC ACG CGA CAT

CGC-3' T) (10 pmol/µl, Biomers) together with unlabelled helper and competitor probes (10 pmol/µl, Biomers) at 46 °C for 3 h. After washing at 48 °C for 10 min, the slides were incubated in 1X phosphate-buffered saline (PBS) for 15 min and shortly dipped into MilliQ water. Sponge tissues were then stained with 4',6-diamidino-2-phenylindole (DAPI, 1 µg/ml) (Thermo Fisher, USA) and mounted in Citifluor (Electron Microscopy Sciences): Vecta Shield (Vector Laboratories) media at a 4:1 v/v. Four-times-Atto488-labelled EUB338 and NON338 probes were used as positive and negative controls, respectively.
Imaging was performed in a Nikon Eclipse Ti2-E inverted microscope equipped with Nikon DS-Qi2 camera and SOLA white light engine. DAPI and FISH signals were illuminated using different fluorescence filter cubes with corresponding excitation and emission spectra. Images were taken with two objectives: a plan-Apochromat 100 x 1.45 oil immersion and a plan-Fluor 10 x 0.3. Maximum intensity projection of multiple images acquired in z-stack were constructed in NIS-Element AR software 5.21.03.

Results

MAGs collection and phylogenomic placement

In total, 29 MAGs obtained from the sponge metagenomes (Table S1) were classified by the GTDB-Tk as 'Desulfobacterota__D'. Accordingly, insertion of the 16S rRNA gene sequences obtained from the respective MAGs by parsimony into the SILVA guide tree (r138) placed them all within *Ca.* Dadabacteria (Quast et al., 2013; Yilmaz et al., 2014). Twenty-eight MAGs passed the quality control with average estimated completeness at 92.2 ± 4.5% and contamination at 1.2 ± 0.6% and were used for subsequent analysis (Table S3). Their predicted genome sizes varied from 1.2 to 1.9 Mbp (average 1.5 ± 0.2 Mbp) with an average GC content of 49.9 ± 2.5%. An additional 92 MAGs were obtained from NCBI and other studies (Table S2). After quality screening, we excluded 47 genomes due to either low completeness (<80%) and/or high contamination (>5%). The final dataset included 73 medium-high quality MAGs according to the MIMAG standards (Bowers et al., 2017) (average completeness 91.0 ± 4.9% and contamination 1.5 ± 1.0%) representing the 'Desulfobacterota__D' (Table S3).
To reconstruct their phylogeny, we first employed a phylogenomic approach using publicly available genomes in addition to MAGs generated in this study (Table S3). Inference of a maximum-likelihood tree on the basis of a concatenated alignment of 120 single-copy bacterial marker proteins supported the GTDB taxonomy of the class UBA1144 into three order-level lineages (UBA1144, UBA2774 and RKRQ01) with high bootstrap values (>98%). MAGs mainly derived from seawater (only one from hydrothermal vents) were assigned to the GTDB order UBA1144, while MAGs from diverse environmental sources, except for seawater and marine sponges comprised the GTDB order UBA2774 (Figure 1). All 49 sponge-associated MAGs (including 28 sequenced here) were placed within the GTDB order RKRQ01 and further clustered into three well supported (bootstrap values >81%) subclades I, II and III (Figure 1), revealing an order-level lineage exclusively comprised of sponge-derived sequences. The 16S rRNA gene tree reflected a similar phylogenetic placement as the genome tree clustering

the gene sequences retrieved from MAGs in three clades within the GTDB order RKRQ01 represented only by MAGs recovered from marine sponges (Figure S1).

Taxonomic rank assignment and proposal of type material

Based on the inferred phylogeny and RED values, the RKRQ01 was best described as an order-level group (median RED = 0.636, median RED for orders = 0.634) falling within the ±0.1 RED interval for all taxa belonging at the same rank (Parks et al., 2018) (Table S4). A putative family (subclade I) was identified that fell within the GTDB order RKRQ01 represented by a single MAG (DOM43_bin18). After bin dereplication at 95% ANI, three novel MAGs were added to GTDB family GCA-014075295, which comprised of a single genus (subclade II) and were resolved into two novel species represented by the MAGs SCOP_bin1 and CLI1_bin1 (Figure 1). In addition, 34 MAGs were placed into GTDB family RKRQ01 (subclade III) representing 14 novel species-level taxa of the GTDB genus RKRQ01 (Table S5). Following the recommendations for defining species and higher ranks of yet-uncultured bacteria using MAGs as type material (Whitman, 2015; Konstantinidis et al., 2017; Chuvochina et al., 2019; Hugenholtz et al., 2021), we propose the following names; *Ca.* Spongomicrobium dominicola sp. nov. (DOM43_bin18), *Ca.* Mycalebacterium zealandia sp. nov. (MH-Pat-all_metabat2_32), *Ca.* Nemesobacter australis sp. nov. (SB0662_bin19) as type species representing subclade I, II and III, respectively. Furthermore, we propose the families *Ca.* Spongomicrobiaceae fam. nov. (subclade I), *Ca.* Mycalebacteriaceae fam. nov. (subclade II), Ca. Nemesobacteraceae fam. nov. (subclade III) and the order *Ca.* Nemesobacterales ord. nov. (Figure 1). More details on the proposed taxonomic outline and etymological description can be found in the Supplementary Information (Table S5).

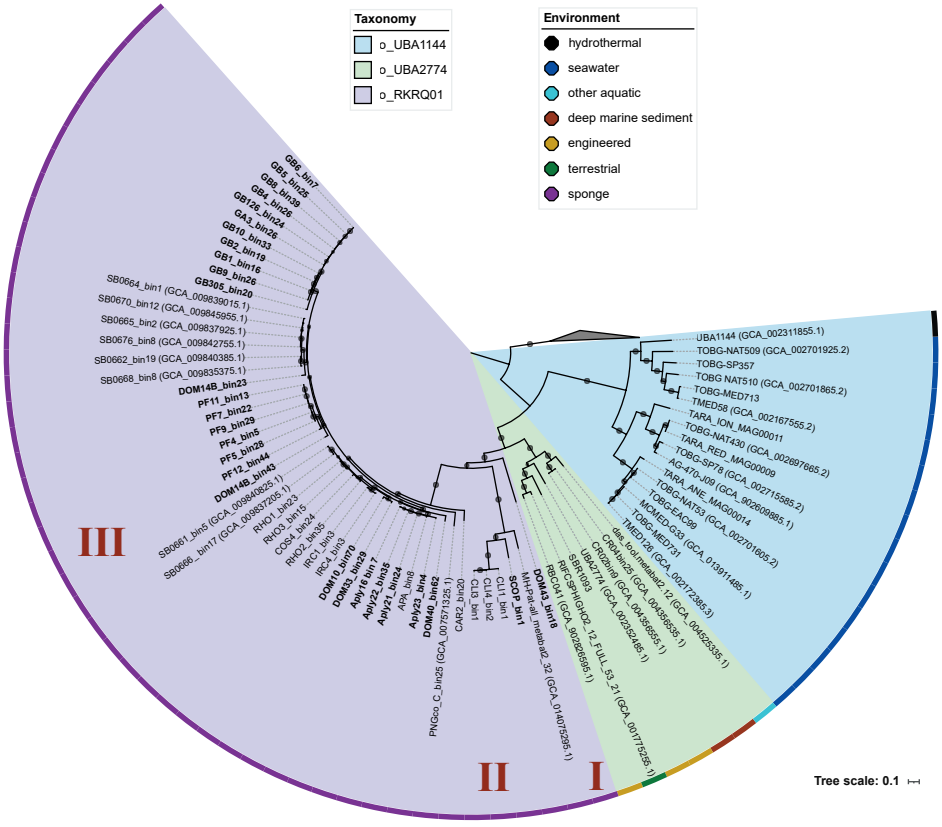


Figure 1. Phylogenomic tree of *Ca.* Dadabacteria (GTDB-Desulfobacterota__D) based on the concatenated alignment of 120 bacterial single-copy marker protein sequences. Outer coloured arcs indicate the environmental source of the MAGs. Different colours on the clades represent assigned taxonomic groups according to GTDB taxonomy. Red numbers show subclades I (*Ca.* Spongomicrobiaceae), II (*Ca.* Mycalebacteriaceae) and III (*Ca.* Nemesobacteraceae). Names in bold represent MAGs generated in this study. Bootstrap values over 80% are shown by grey circles. The collapsed clade indicates the outgroup and includes MAGs representing other members of the *Ca.* Desulfobacterota phylum. The scale bar represents 1 substitution per site.

MAGs relative abundance and global distribution

Read mapping of 27 sponge metagenomes generated here against the *Ca.* Nemesobacterales genomes showed that the dereplicated set (19 MAGs) represented a small fraction of the respective microbial communities accounting for a relative abundance of 0.22% (±0.07%) on average and ranging from 0.002 to 1.29% of the holobiont reads (Table S6). The environmental distribution of *Ca.* Nemesobacterales was investigated by screening all available metagenomes from IMG/M database for the presence of homologous 16S rRNA genes. This search showed 283 unique occurrences according to the sampling location of the metagenomic samples. Members of this lineage were distributed across a wide geographical range and were observed in 29 different types of environments, including

aquatic, terrestrial, hydrothermal, host-associated and engineered systems (Figure 2). The majority of the homologous sequences belonged to metagenomes from marine sponges (23.2%), followed by seawater (13.7%) and marine sediments (5.7%).

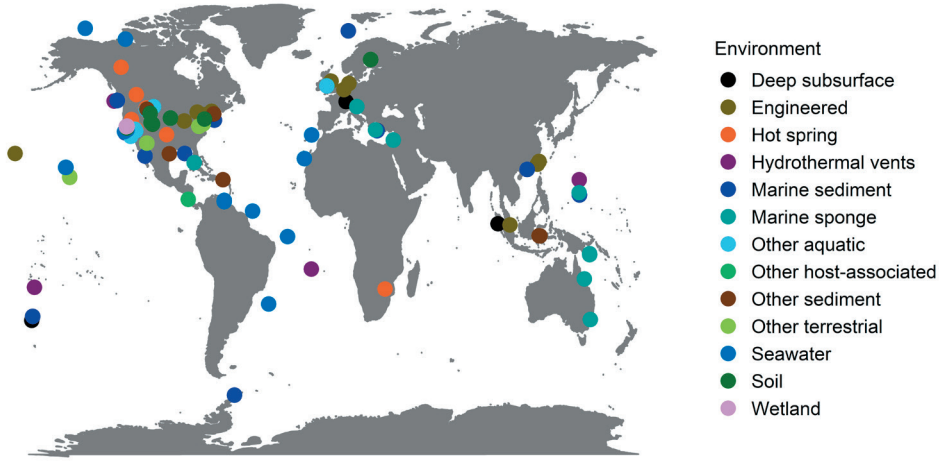


Figure 2. Schematic map representing the global distribution of *Ca.* Nemesobacterales. Members of this lineage are found in diverse environments worldwide, highlighted by different colours.

Niche speciation and functional traits

Comparative genomic analysis was conducted to delineate the functional repertoire of *Ca.* Nemesobacterales using both in-house and publicly available MAGs. Core pathways for carbon metabolism such as glycolysis, tricarboxylic acid cycle (TCA), pentose phosphate pathway (non-oxidative phase) and phosphoribosyl diphosphate (PRPP) biosynthesis pathway were mostly complete, whereas no autotrophic carbon fixation pathways were detected (Table S7). Aerobic respiration is probable and hence, oxygen is likely to be the preferred terminal electron acceptor as no genes for alternatives (e.g., nitrate, nitric oxide, sulfoxides, etc.) were found in the genomes. The majority of MAGs included in our study contained all the essential genes for the biosynthesis of several amino acids, cofactors and vitamins, including proline, lysine, biotin, riboflavin, coenzyme A and heme (Figure 3).

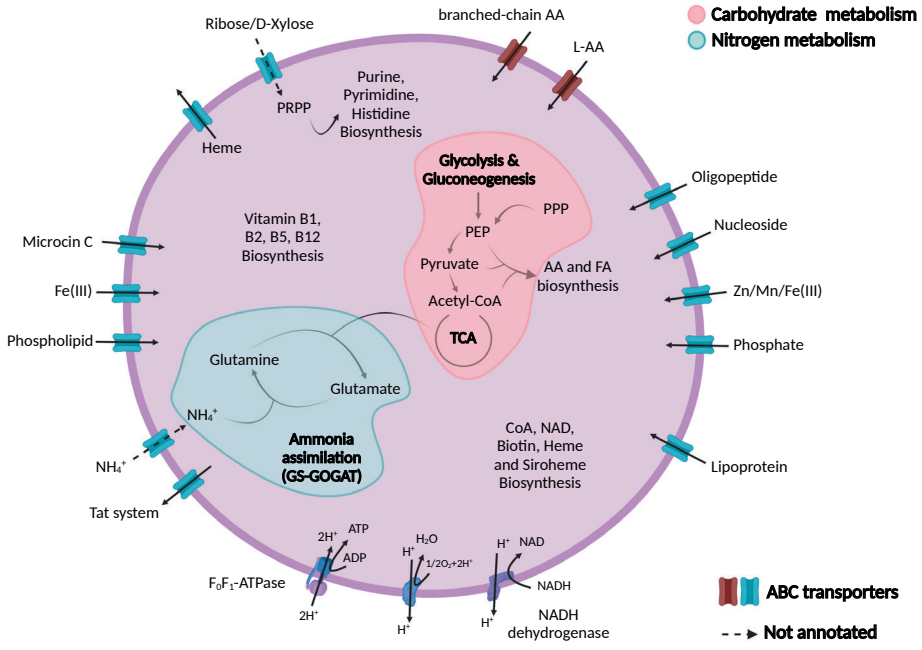


Figure 3. Schematic overview of the inferred metabolism of *Ca.* Nemesobacterales. Pathways affiliated with nitrogen and carbohydrate metabolism are highlighted in different colours. Dashed arrows indicate pathways and transporters for which not all responsible genes were annotated. (Created with https://biorender.com/).

Their putative capacity to degrade carbohydrates was assessed by screening the MAGs for carbohydrate-active enzymes (CAZymes) belonging to the following families: glycoside hydrolases (GHs), carbohydrate esterases (CEs), polysaccharide lyases, carbohydrate binding modules (CBMs) and auxiliary activities . On average, *Ca.* Nemesobacterales possessed 15.8 ± 2.9 CAZymes/Mbp. Overall, 40 different CAZY families were identified. The highest frequency CAZymes normalized for MAG length (average CAZymes/Mbp > 1.0) belonged to GH23, GT2, GT4 and CBM50 (Table S8).

To obtain a better overview of the functions of the host-associated bacteria a set of 19 highly-complete MAGs (>80%) derived from sponges (*Ca.* Nemesobacterales) was compared to 16 seawater-derived MAGs (GTDB order UBA1144). The completeness of the KEGG modules related to carbon and energy metabolism was highly similar between the sponge and seawater MAGs (Figure S2, Table S7). Accordingly, MAGs derived from both habitats seemed to have the potential of synthesizing a similar range of amino acids, cofactors and vitamins. Distinct for some sponge-derived MAGs was the ability to produce cobalamin (vitamin B12) that was absent from the seawater MAGs. Despite their highly similar carbon and amino acid metabolism, annotation of ATP-binding cassette (ABC) transporters showed that only sponge MAGs encoded genes for the transport of amino acids, lipoproteins and

oligopeptides. They shared the same phosphate starvation and nitrogen regulation system, possibly used under nutrient-limiting conditions. In the case of iron, the host-associated MAGs encoded different transport systems, as well as the potential of using the glyoxylate shunt, recently proposed as an acclimation strategy to iron limitation (Koedooder et al., 2018) (Figure S2). CAZyme analysis showed that MAGs from the two different habitats harboured almost the same average number of CAZymes/Mbp (16.2 ± 2.3). Similarly, the most frequent CAZY families (average CAZymes/Mbp > 1) in the seawater MAGs were the same as the ones in the sponge MAGs (Figure 4A). However, the CAZymes repertoire was less diverse for the seawater MAGs. Almost 60% of the total identified CAZY families were detected only in MAGs from sponges whereas only 6 CAZY families were exclusively found in seawater MAGs (Table S8).

Functional profiles based on annotated Pfams, KEGG orthologs (KOs) and TIGRFAMs were used in order to assess correlation between taxonomy (different clades) and niche (sponge vs. seawater) with the predicted traits (Figure S3). MAGs belonging to the different order-level lineages (*Ca.* Nemesobacterales vs. UBA1144) varied significantly at the functional level based on relative abundance analysis of all three annotations (PERMANOVA, $p < 0.001$; Figure 4C and S3). Similarly, highly divergent functional profiles (PERMANOVA, $p < 0.001$; Figure 4C and S3) were observed between bacterial genomes from sponges and seawater. Enrichment analysis showed that sponge-derived MAGs were enriched in Pfams associated with defense mechanisms and symbiosis-related factors (Figure 4B and Table S9).

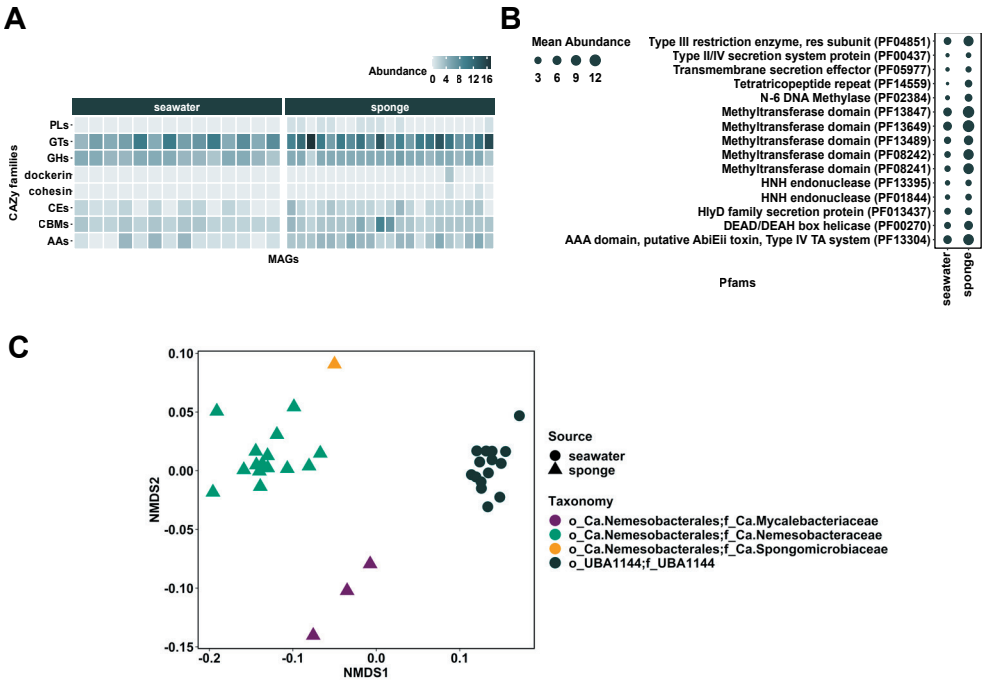


Figure 4. Functional comparison of *Ca.* Dadabacteria (GTDB-Desulfobacterota__D) MAGs associated with marine sponges and seawater. A) Heatmap illustrating the abundance of carbohydrate-active enzymes (CAZymes) families in sponge- and seawater-associated MAGs. PLs, polysaccharide lyases; GTs, glycosyl transferases; GHs, glycoside hydrolases; CEs, carbohydrate esterases; CBMs, carbohydrate-binding modules; AAs, auxiliary activities. B) Mean number of Pfams related to symbiosis factors and defense mechanisms identified in MAGs from each habitat. C) Non-metric multidimensional scaling plot (NMDS) based on Bray-Curtis dissimilarity scores calculated from annotated Pfams. Functional variability is displayed between sponge and seawater MAGs (different shapes) and different taxonomic groups (different colours).

To further investigate the distribution of these functions in the MAGs, screening was done for 230 Pfams encoding restriction-modification systems (RMs), clustered regularly interspaced short palindromic repeat systems (CRISPRs), secretion systems (SSs), toxin-antitoxin systems (TAs), transposases, eukaryotic-like proteins (ELPs) and others (Table S9). In total, 166 Pfams of the aforementioned types were identified either in sponge or seawater MAGs from which 94% were significantly enriched in the *Ca.* Nemesobacterales. Only 26 Pfams were shared between MAGs from both habitats, while 137 Pfams were specific to the sponge symbionts (Table S9). Most Pfams enriched in *Ca.* Nemesobacterales were related to RMs (45 Pfams), ELPs (42 Pfams), TAs (32 Pfams) and SSs (12 Pfams). More specifically, Pfams associated with methyltransferases, helicases, the putative ABI (abortive infection) toxin, type III restriction enzymes, Fic/DOC (filamentation induced by cAMP/death on curing) proteins and tetratricopeptide repeats were among the most abundant ones in *Ca.* Nemesobacterales (average abundance > 5.0). Interestingly, certain Pfams, such as type I restriction enzyme,

leucine rich and tetratricopeptide repeats were overrepresented, even more than 10-fold in sponges compared to seawater (Table S9).

The potential for production of secondary metabolites was also investigated by screening the investigated MAGs for secondary metabolite biosynthetic gene clusters (BGCs). In general, four different BGC types were identified in the genomes from both habitats coding for type I polyketide synthases (T1PKS), type III polyketide synthases (T3PKS) and for enzymes involved in the production of beta-lactones and terpenes (Table S10). Genome mining showed that both BGC abundance and diversity were higher for the sponge-associated MAGs, which harboured on average almost twice as many BGCs (1.11 ± 0.19 BGCs/MAG) compared to the free-living MAGs (0.63 ± 0.15 BGCs/MAG). The distribution of these BGC types was uneven as terpene BGCs were encoded exclusively by seawater MAGs while the rest were mainly found in the sponge-derived MAGs (Figure S4). BGCs coding for beta-lactones and T1PKS showed the highest abundance in *Ca.* Nemesobacterales MAGs (Table S10).

Visualisation of *Ca.* Nemesobacterales

Novel oligonucleotide FISH probes were designed and facilitated for the first time the detection of *Ca.* Nemesobacterales. Selective labelling led to the visualization of *Ca.* N. rappii cells in the tissue of *G. barretti* samples (Table S5). The bacterial cells featured a rod shape with a length of 1.04 + 0.21 µm, and were located both in the sponge mesohyl (Figure 5A) and within sponge cells (Figure 5B-C). *Ca.* Nemesobacter rappii cells often surrounded sponge nuclei resembling bacteriocyte formation (Figure 5B-C).

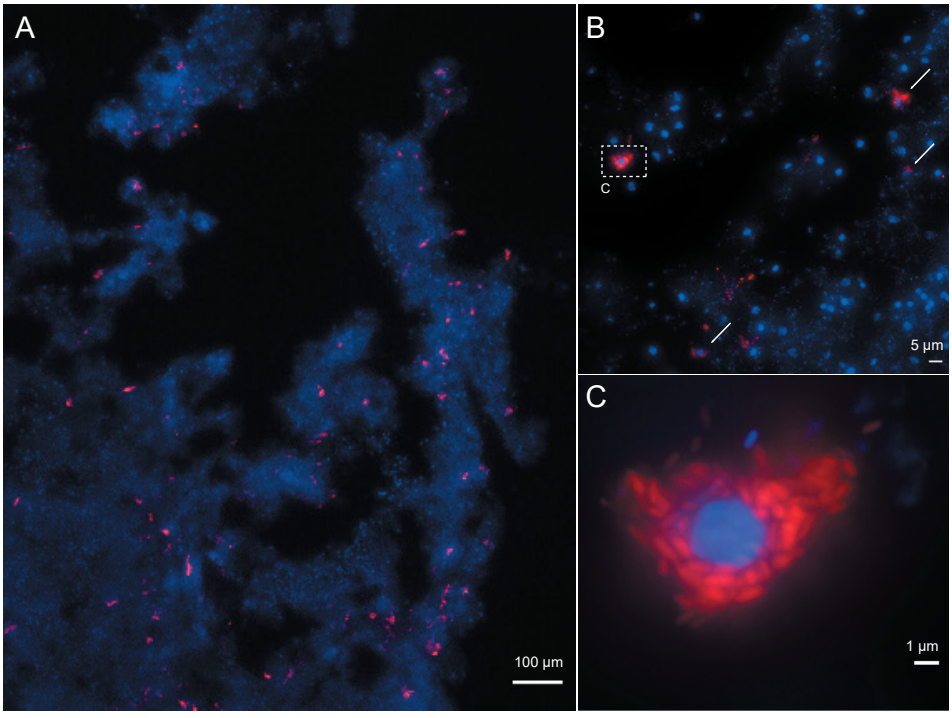


Figure 5. Imaging of *Ca.* N. rappii in sponge tissue. DNA is stained with DAPI (blue) and *Ca.* N. rappii cells are visualized with the 4-times-Alexa594 labelled oligonucleotide probe DADA392 (red). Maximum intensity projection of z-stack fluorescence images that were acquired in (A) 150x, (B) 1000x, and (C) 1500x magnification are shown. (B) Some *Ca.* N. rappii cells cluster around sponge nuclei (arrows), which resemble bacteriocyte formation. The signals shown in dashed rectangle are magnified in (C). Representative images of 47 images taken in three individual experiments using two different sponge specimens are depicted.

Discussion

Members of the 'Desulfobacterota__D' (or 'Dadabacteria') lineage are lacking cultured representatives while their phylogeny and metabolic traits have remained elusive despite the fact that they are ubiquitous. Since the first published MAG in 2014 (Wang et al., 2014), several MAGs have been recovered from a plethora of environmental samples, both free-living and host-associated (Graham and Tully, 2021), especially from marine sponges (Slaby et al., 2017; Engelberts et al., 2020; Robbins et al., 2021). However, to date no information exists on the precise phylogenetic placement and functional characteristics of the sponge-dwelling representatives. To this end, our study incorporated a large number of MAGs, including published and newly determined ones which were assigned to the three existing GTDB orders of 'Desulfobacterota__D' (UBA1144, UBA2774 and RKRQ01). Previous phylogenetic analyses showed three clades representing marine pelagic, hydrothermal and organic carbon-associated systems (incl. two sponge-derived MAGs) (Graham and Tully, 2021). Here, the addition of 49 MAGs affiliated with marine sponges altered the previously

described architecture of the lineage revealing a sponge-specific clade (GTDB order RKRQ01). The remaining MAGs originated mainly from seawater (GTDB order UBA1144) and other marine and non-marine environments (GTDB order UBA2774). Our phylogenomic investigation identified one novel lineage within the sponge-specific order, which based on RED values corresponds to a previously undescribed family, *Ca.* Spongomicrobiaceae. It should be mentioned that *Ca.* Spongomicrobiaceae comprised only one MAG (DOM43_bin18) obtained from a *Scleritoderma* sp. At the species level, several novel species were resolved and affiliated to the other two families, denoted here as *Ca.* Mycalebacteriaceae and *Ca.* Nemesobacteraceae. Thus, our findings considerably expand the 'Desulfobacterota__D' species tree and highlight the current underrepresentation of this lineage.

The relative abundance of *Ca.* Nemesobacterales in metagenomes of different sponge species based on read recruitment analysis ranged between 0.002 to 1.3% with an average of 0.22%. These findings are in line with the relative abundance of 'SBR1093' OTUs (former name of *Ca.* Dadabacteria phylum) in the global sponge microbiome (Thomas et al., 2016). Based on the above numbers and compared to genomic studies on other groups of sponge-associated bacteria (Astudillo-Garcia et al., 2018; Bayer et al., 2018; Sizikov et al., 2020; Taylor et al., 2021), *Ca.* Nemesobacterales are ubiquitous but low in abundance members of the sponge microbiome. Specifically, MAGs of members of *Ca.* Nemesobacterales were recovered here from 27 sponge metagenomes belonging to eight different sponge species collected from different geographical locations while in many cases, the same *Ca.* Nemesobacterales MAG was present in more than one host species (Table S6). This reflects a wide distribution in taxonomically diverse sponge species which might suggest a generalistic role within the sponge microbiome (Thomas et al., 2016). Besides marine sponges, metagenomes from various environments around the world contained homologous sequences of *Ca.* Nemesobacterales, including hydrothermal systems, sediment, seawater, soil and wetlands, further highlighting their broad occurrence at a global scale (Figure 2). Moreover, its members were also detected in tunicates and marine worms, yet the majority of the homologous sequences were sponge-associated indicating their close association and importance of *Ca.* Nemesobacterales within marine sponges.

The sponge holobiont has been substantially studied and thus, sponge-microbe interactions are considered a prime example of symbiosis in the animal realm (Taylor et al., 2007b; Taylor et al., 2007a; Thomas et al., 2010a; Fan et al., 2012; Webster and Taylor, 2012; Webster and Thomas, 2016; Pita et al., 2018). Yet, many bacteria remain undescribed, and their contribution to the host success has not been determined. A wide range of core functions which shape the sponge microbiome have been previously reported concerning nutrient and vitamin metabolism as well as defense features that protect microbial populations from viruses, pathogens and/or the host allowing sustained colonization of the host by the symbionts (Thomas et al., 2010a; Pita et al., 2018). Here, we provide the first insights into the functional potential of the sponge-dwelling *Ca.* Nemesobacterales using genome-centric metagenomics. The metabolism reconstruction showed that carbon acquisition is likely performed via heterotrophy as all major

pathways for carbon metabolism were present. This is consistent with previous analyses of the marine pelagic clade (Graham and Tully, 2021). Even though carbon fixation has been reported before (Wang et al., 2014), no evidence for autotrophy was found in *Ca.* Nemesobacterales. An almost complete electron transport chain for aerobic respiration reflects an aerobic, heterotrophic lifestyle (Figure 3). In contrast to other members of this lineage (Graham and Tully, 2021), *Ca.* Nemesobacterales lack complete pathways for inorganic nitrogen and sulfur metabolism. Therefore, the presence of ABC transporters for amino acids, oligopeptides and lipoproteins might indicate a preference for organic nitrogen and sulfur compounds derived from the host or other members of the microbial community. *Ca.* Nemesobacterales also showed the potential to degrade and transform complex carbohydrates. For example, genes encoding certain glycosylases (GH23) and other related proteins (CBM50) which act on bacterial peptidoglycans or animal chitin were widely present in *Ca.* Nemesobacterales. Genes encoding these enzymes have been previously reported as highly abundant in other sponge symbionts belonging to *Chloroflexota*, *Bacteroidetota*, *Candidatus* Poribacteria (Kamke et al., 2013; Bennke et al., 2016; Bayer et al., 2018; Gavriilidou et al., 2020) and these enzymes are thought to play a role in cell adhesion, aggregation and allorecognition (Kamke et al., 2013).
To investigate any adaptations related to host association, we compared the genomic features of the sponge-derived MAGs to MAGs isolated from seawater samples. *Ca.* Nemesobacterales shared equivalent pathways for carbon and energy metabolism as well as amino acid and cofactor biosynthesis with their seawater counterparts. This supports recent findings on the possible nutrient provisioning of symbionts to the host without being key processes in the holobiont since sponges could also obtain these nutrients through feeding on seawater microbes (Robbins et al., 2021). On the other hand, only sponge-associated MAGs had the ability of synthesizing cobalamin (vitamin B12) which highlights a potential interplay between microbes and host that might concern specific molecules that sponges are not able to synthesize, such as vitamins and amino acids (Thomas et al., 2010a; Kamke et al., 2013; Bayer et al., 2018; Munroe et al., 2019; Engelberts et al., 2020; Haber et al., 2020). MAGs from both habitats showed a similar CAZyme repertoire in terms of abundance. However, *Ca.* Nemesobacterales were predicted to be capable of degrading a greater diversity of complex carbohydrates than their free-living counterparts, probably due to exposure to a larger variety of carbohydrates sourced from the host, other symbionts and/or the incoming seawater (Thomas et al., 2010a).

Further exploration of their functional profiles provided additional support that *Ca.* Nemesobacterales are metabolically divergent from seawater-derived members of the same class (Figure 4C) (Fan et al., 2012; Horn et al., 2016). In fact, genomic signatures significantly enriched among members of *Ca.* Nemesobacterales were related to host-symbiont interactions and prokaryotic defense. Sponge symbionts have been long demonstrated to own a pronounced molecular toolbox of mechanisms in order to maintain a stable relationship with the host by evading phagocytosis, viral infections and coping with

temporal variations of environmental conditions in the sponge niche (Horn et al., 2016; Slaby et al., 2017). It has been proposed that sponge-associated bacteria either evolved these mechanisms in the longstanding process of adaptation to the host environment or acquired them via horizontal gene transfer from the host or other symbionts (Fan et al., 2012; Horn et al., 2016; Reynolds and Thomas, 2016). Here, the majority of enriched Pfams were specific to *Ca.* Nemesobacterales and were affiliated with RMs, TAs and ELPs often prevalent in sponge-associated bacteria (Thomas et al., 2010a; Fan et al., 2012; Horn et al., 2016; Slaby et al., 2017; Astudillo-Garcia et al., 2018; Bayer et al., 2018; Podell et al., 2019; Engelberts et al., 2020; Moreno-Pino et al., 2020; Sizikov et al., 2020; Robbins et al., 2021; Taylor et al., 2021). RMs are mechanisms that contribute to the bacterial defense by recognizing and targeting foreign DNA while TAs are causing cell dormancy or apoptosis in response to environmental stressors and eventually protect the cell population (Makarova et al., 2013; Harms et al., 2018a). For example, a domain belonging to an ATPase associated with the ABI toxin (PF13304), which inhibits viral replication by programming cell death, was the most abundant TA in *Ca.* Nemesobacterales (Haber et al., 2020). Furthermore, ELPs are proteins containing domains of eukaryotic origin in prokaryotes that have been previously proposed to modulate host behaviour by mediating protein-protein interactions and thus, facilitate the establishment of a successful symbiosis (Reynolds and Thomas, 2016). A large number of sponge-associated MAGs carry genes encoding ELPs (Engelberts et al., 2020; Robbins et al., 2021; Taylor et al., 2021) that likely aid them in escaping digestion by sponge cells (Nguyen et al., 2014; Reynolds and Thomas, 2016). Similarly, in this study we show that *Ca.* Nemesobacterales inhabiting phylogenetically diverse sponge species had genes encoding several types of ELPs in their genomes with TPRs showing the highest abundance (Thomas et al., 2010a; Fiore et al., 2015; Diez-Vives et al., 2017). TPR-containing proteins are thought to participate in various cellular processes including virulence of bacterial pathogens (Cerveny et al., 2013). However, the mechanism of interaction between bacterial ELPs and host cells remains unknown. Evidence on co-expression of ELPs with transport systems suggest that they are delivered into the extracellular environment via transporters or secretion systems (Diez-Vives et al., 2017). Recent genomic findings indicate that only a few lineages in the sponge microbiome (e.g., *Gammaproteobacteria*, *Acidobacteriota*, *Gemmatimonadota*) encode genes related to secretion systems (Engelberts et al., 2020; Robbins et al., 2021). Interestingly, *Ca.* Nemesobacterales might be one of these taxa that interact with the sponge host via secretion systems since they were enriched in several Pfams affiliated with type II, III and IV secretion systems.

Another line of prokaryotic defense is the biosynthesis of secondary metabolites. Marine sponges are well known chemical factories and their symbionts are thought to contribute to the production of bioactive molecules in the face of predation, competition and microbial infections (Taylor et al., 2007a; Piel, 2009; Pita et al., 2018). Several representatives of the sponge microbial community have exhibited in vitro bioactivity and possess a wide range of secondary metabolites (Indraningrat et al., 2016; Brinkmann et al., 2017b; Karimi et al.,

2017; Lackner et al., 2017; Bayer et al., 2018; Gavriilidou et al., 2020; Konstantinou et al., 2020; Gavriilidou et al., 2021). *Ca.* Nemesobacterales MAGs carried a greater diversity and abundance of BGCs compared to the seawater-derived MAGs confirming the prominent secondary metabolite biosynthesis potential of sponge-dwelling bacteria. Beta-lactone and T1PKS BGCs were the most predominant in the genomic repertoire of *Ca.* Nemesobacterales. Compounds resulting from these BGCs have shown different properties such as antimicrobial and anticancer activity (Piel et al., 2004; Robinson et al., 2019). Even though their exact ecological function remains rather elusive, it seems that *Ca*. Nemesobacterales are participating in the chemical defense or communication within the sponge holobiont.

Localization of the *Ca.* N. rappii in the tissue of *G. barretti* revealed that the bacterial cells were scattered in the sponge mesohyl and often clustered in vacuole-like structures, probably bacteriocytes which are specialized host cells that accommodate intracellular bacteria such as those previously described in several marine sponge species (Vacelet and Donadey, 1977; Ilan and Abelson, 1995; Maldonado, 2007; Taylor et al., 2021). Bacteriocyte association in eukaryotes has been presumed an event of mutualistic infection between bacterial symbionts and eukaryotic hosts with a yet-unknown physiological role for either of the partners (Moran and Wernegreen, 2000). It has been hypothesized that bacteriocytes are used by the sponge as mechanism to control microbial populations or facilitate certain metabolic exchanges between the symbionts and the host (Maldonado, 2007). These findings further support the aforementioned functional predictions that *Ca.* Nemesobacterales exist in close symbiosis with marine sponges. Based on the observed localization patterns, members of this order in *G. barretti* likely lead an intracellular lifestyle in sponge bacteriocytes. Nevertheless, Taylor et al. (2021) described diversity in cell morphology and localization within the sponge-specific order *Candidatus* Tethybacterales in various sponge species where some symbionts were intracellular and others extracellular. It is highly likely that members of *Ca.* Nemesobacterales might also appear intracellular and others extracellular throughout the mesohyl in the same or different sponge species. Hence, further acquisition of *Ca.* Nemesobacterales genomes in conjunction with experimental data are needed to be able to expand on this hypothesis.

Conclusion

The current study provides unprecedented insights into the phylogeny and metabolism of *Ca.* Nemesobacterales, a sponge-specific order of the newly classified *Ca.* Desulfobacterota phylum. We propose the division of the order *Ca.* Nemesobacterales into three families, namely *Ca.* Nemesobacteraceae, *Ca.* Spongomicrobiaceae (newly described) and *Ca.* Mycalebacteriaceae. Despite their low abundance, *Ca.* Nemesobacterales displayed a wide distribution in different sponge species. Our analysis showed that they can also be found worldwide in diverse environments, albeit with the majority occurring in marine sponges. Metabolism reconstruction revealed that *Ca.* Nemesobacterales can be predicted to be

aerobic, heterotrophic microorganisms with the potential of degrading various complex carbohydrates. Our results highlight that *Ca*. Nemesobacterales are functionally divergent from seawater-associated members of the same lineage. This is mainly attributed to the overrepresentation of several genomic signatures related to nutritional provisioning, host-microbe interactions, phage defense and biosynthesis of bioactive molecules, which clearly reflects a host-associated lifestyle. We infer that *Ca.* Nemesobacterales are in a close relationship with sponges, also indicated by their intracellular existence in the sponge *G. barretti*, which led to a quite distinct gene repertoire and niche-specific adaptations. Future investigations involving additional genomes and metatranscriptomic data as well as localization of other *Ca.* Nemesobacterales in sponges could potentially enhance our understanding of the ecological function of *Ca.* Nemesobacterales in the sponge holobiont.

Acknowledgements

The authors would like to thank the late Hans Tore Rapp for his invaluable help in collecting *Geodia* samples from Norway, Ellen Kenchington for the Canadian *G. barretti* samples, Vasilis Gerovasileiou for performing the sampling of *P. ficiformis* and Andriaan Schrier for supporting the sponge collection in Dominica. Henk Schipper is acknowledged for helping with the sponge tissue processing. We also thank Torsten Thomas for providing us with additional data for our analysis. Maria Chuvochina, Donovan Parks and Philip Hugenholtz are acknowledged for their advice on taxonomy and rank assignment. This research was financially supported by the European Commission through the SponGES project (Grant agreement ID: 679849) to DS and AG and the Marie Sklodowska-Curie Individual Fellowship COSMos (Grant agreement ID: 897121) to MAS.

Supplementary Information

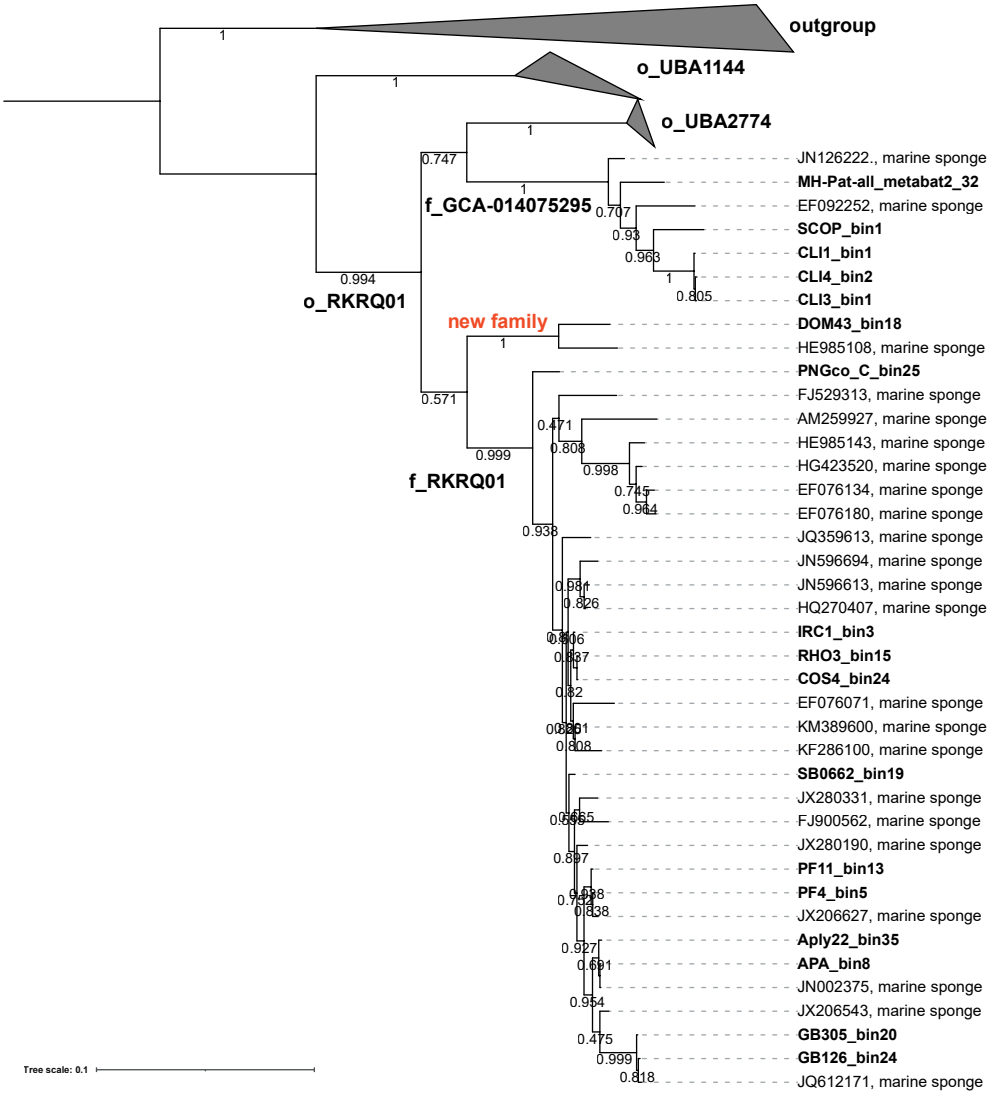


Figure S1. Maximum likelihood tree based on 16S rRNA gene sequences. Names in bold represent MAG-derived sequences. Rank assignment followed GTDB taxonomy. 16S rRNA gene sequences of the *Chloroflexota* phylum were used as outgroup. The scale bar indicates 0.1 substitution per nucleotide position.

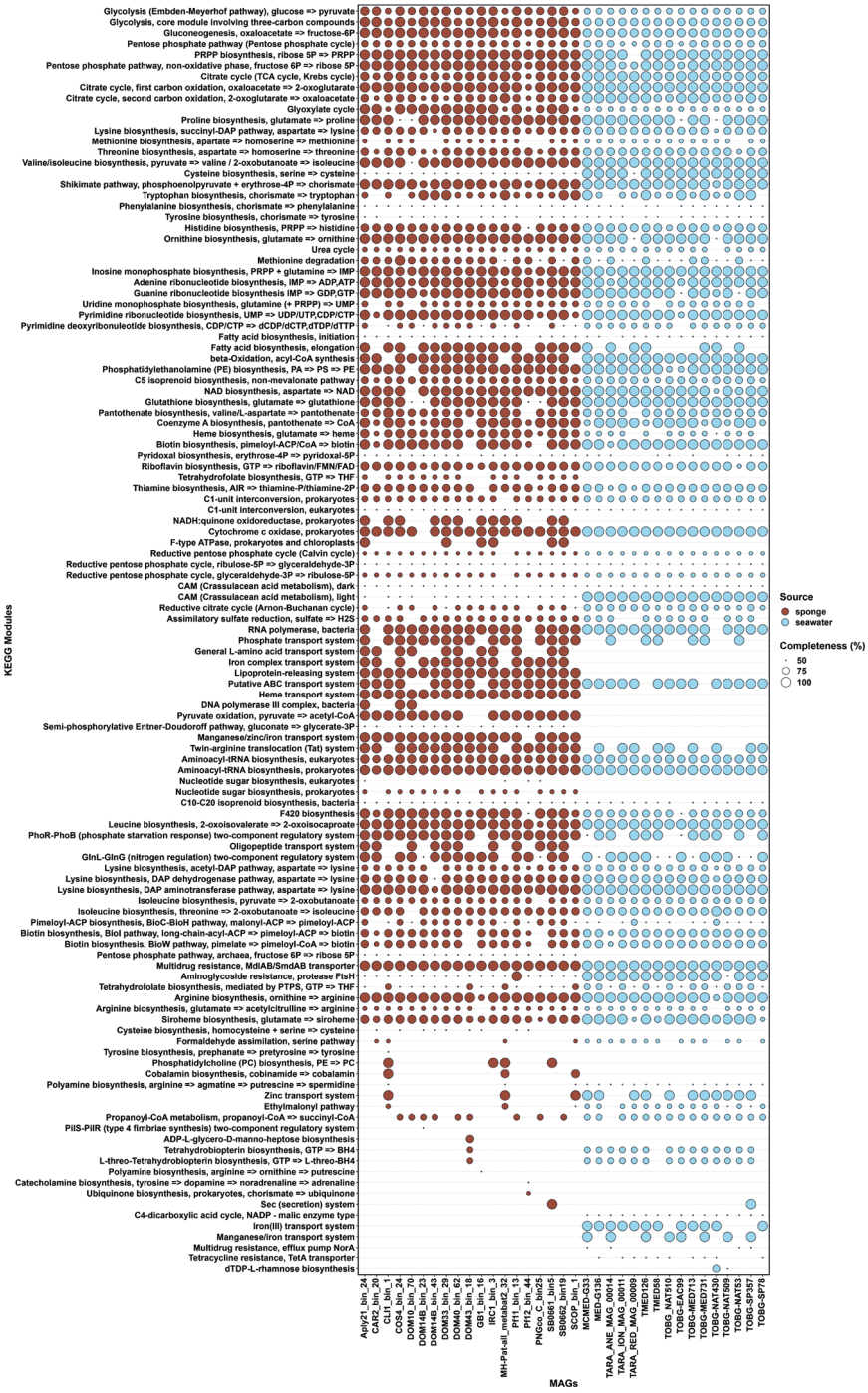


Figure S2. Completeness of KEGG modules annotated in sponge- and seawater-associated MAGs. Colour indicates the environmental source and size of circles the completeness (%) based on the number of steps found to the steps needed for a KEGG module to be complete as annotated by EnrichM.

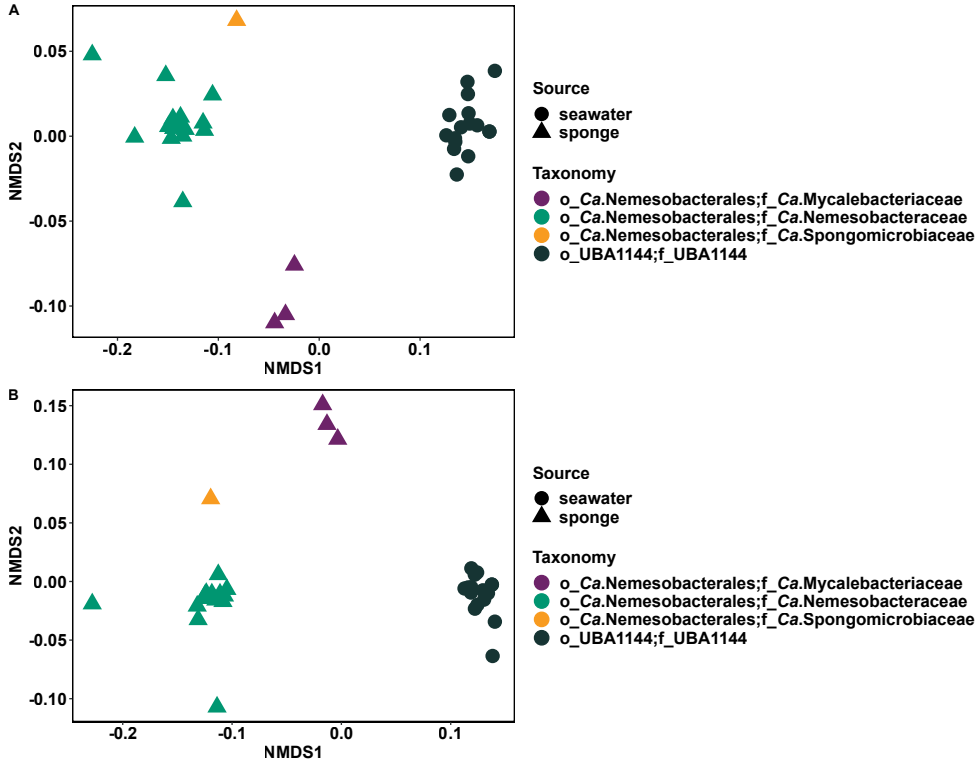


Figure S3. Functional variability of sponge- and seawater-MAGs belonging to different taxonomic groups of the *Ca.* Dadabacteria (GTDB-Desulfobacterota__D). Non-metric multidimensional scaling (NMDS) were created based on Bray-Curtis dissimilarities of A) KEGG orthologs (KOs) and B) TIGRFAMs annotated in the respective MAGs.

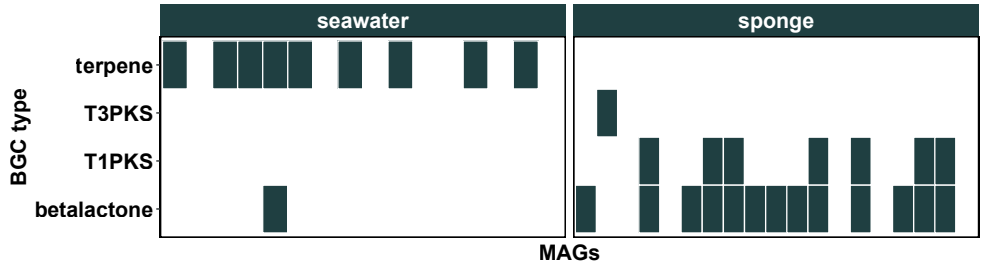
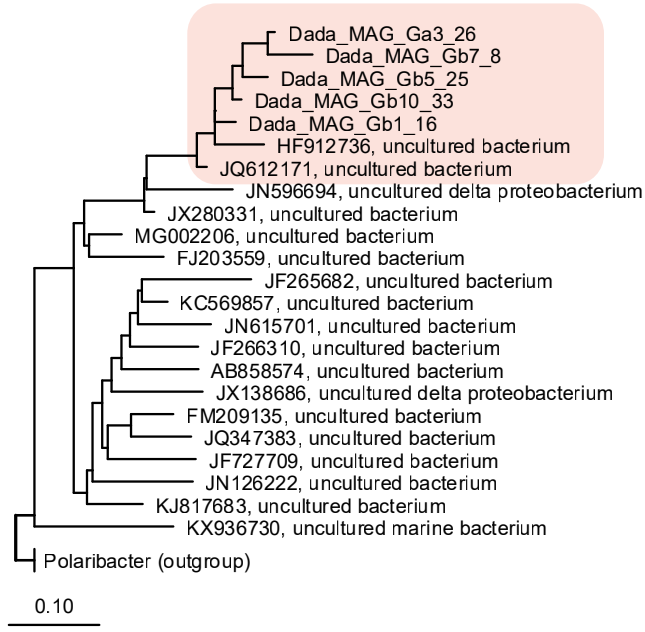


Figure S4. Presence of biosynthetic gene cluster (BGC) families in MAGs derived from sponge and seawater samples.

A



B

Probe	Target	Probe sequence (5 - 3)	Formamide (%)	Reference
DADA393	Dadabacteria	TCA TCC CTC ACG CGA CAT CGC T	35	This study
h1DADA393	Helper for DADA393	GGC TTT CGC CCA TTG CCC AAT AT	35	This study
h2DADA933	Helper for DADA393	CAG AGG TTT ACG ACC CGA AGG C	35	This study
c15DADA393	Competitor for DADA393	TCA TCC CTC ACG CGG CAT CGC T	35	This study
c17DADA393	Competitor for DADA393	TCA TCC CTC ACG CGG CGT CGC T	35	This study
EUB338	Most bacteria	GCT GCC TCC CGT AGG AGT	35	Daims et al., 1999
NON338	Non-sense probe	ACT CCT ACG GGA GGC AGC	35	Amann et al., 1990

Figure S5. Phylogenetic analysis and probe design for the visualization of *Ca.* Nemesobacterales in sponge tissue. (A) Maximum likelihood phylogeny of *Ca.* Dadabacteria (GTDB-'Desulfobacterota__D') 16S rRNA gene sequences and related sequences selected from the SILVA database 138. Probe target group is depicted in red. *Polaribacter* spp. were selected as outgroup. Bar: 0.1 substitutions per nucleotide position. (B) Names, target groups, sequences, and formamide concentrations for newly designed *Ca.* Nemesobacterales probes. Atto488 (green) and Atto594 (red) labelled nucleotides are annotated. Hits between the probe DADA393 and non-target sequences at 0 and 1 mismatch are 3 and 116, respectively.

The supplementary tables are available online at https://github.com/mibwurrepo/Gavriilidou_PhD_Thesis.

Chapter 6

General Discussion

Zooming in: The sponge microbiome

One of the keys to the evolutionary success of the phylum Porifera is the intimate associations with their microbial partners or symbionts (Hill and Sacristán-Soriano, 2017). Throughout this thesis, the term *symbiosis* has been used in a broader sense, to indicate persistent relationships between two or more organisms of different species that live together without considering the exact type of interaction (beneficial, neutral or harmful) (De Bary, 1879; Hill and Sacristán-Soriano, 2017; Pita et al., 2018). In the past decades, most research efforts were directed towards the symbiotic relationships between the host and a single symbiont and hence, knowledge on more complex interactions remain scarce (Apprill, 2020). Marine sponges exemplify such intricate host-associated systems, harboring dense microbial assemblies that place them among the most microbially diverse habitats on Earth (Thompson et al., 2017; Lurgi et al., 2019).

The global sponge microbiome holds an extraordinary diversity with more than 40 reported microbial phyla consisting of both generalist and specialist community members (Thomas et al., 2016). Besides the environment, one of the factors driving the composition and structure of these symbiotic assemblies is host phylogeny (Easson and Thacker, 2014; Thomas et al., 2016; Lurgi et al., 2019). Evidence suggests that phylogenetically related hosts tend to accommodate similar microbiomes, a phenomenon called phylosymbiosis (Lurgi et al., 2019; Lim and Bordenstein, 2020). Our findings were consistent with this pattern as we showed that phylotypes of the newly described genus *Candidatus* Thalassonella of the *Candidatus* Tectomicrobia phylum clustered based on the associated host sponge (**Chapter 4**, Figure 2). Interestingly, a degree of host specificity was also observed for a subgroup of *Candidatus* Entotheonella bacteria, which included all previously known metabolically versatile *Ca.* Tectomicrobia phylotypes (**Chapter 4**, Figure 3). This is further supported by a recent network analysis of the global sponge microbiome which suggested that different hosts in the face of similar environmental threats select for analogous microbiomes equipped with the necessary secondary metabolic functions for them to survive (i.e., chemical defense) (Lurgi et al., 2019). Therefore, sponge hosts and their associated microbial communities are intimately linked in ways that make their study challenging. In **this chapter**, I discuss how the research described in this thesis provided clues on host-symbiont synergism and contributed to the current knowledge of sponge symbiosis.

Charting the unknown

Sponge holobionts (i.e., the host and its microbiota) constitute a significant reservoir of unique microbial diversity (Thomas et al., 2016; Hill and Sacristán-Soriano, 2017), and despite the intensive research efforts of the past decades, there are lineages which remain elusive since they are lacking cultured representatives. So far, most of the focus has been directed towards the dominant or highly specialized members of the sponge microbiota

such as *Proteobacteria* (Karimi et al., 2019; Taylor et al., 2021), *Chloroflexota* (Bayer et al., 2018), *Cyanobacteria* (Burgsdorf et al., 2015; Schorn et al., 2019) and the candidate phylum Poribacteria (Kamke et al., 2014; Podell et al., 2019). This leaves an enormous gap of knowledge regarding the dozens of yet-uncultured microbial lineages inhabiting sponges. Throughout this thesis, we managed to identify several novel sponge-associated taxa based on data from 16S rRNA genes and metagenome-assembled genomes (MAGs) (**Chapter 4** and **5**). More specifically, in **chapter 5** binning of sponge metagenomes and subsequent phylogenomic analysis revealed several draft genomes representing previously undescribed clades within an order exclusively comprising sponge-associated bacteria, namely *Candidatus* Nemesobacterales of the candidate phylum Desulfobacterota. In addition, further diversity in the uncultivated *Ca.* Tectomicrobia phylum was uncovered by combining a targeted amplicon prospecting strategy and phylogenetic analyses of publicly available sequences (**Chapter 4**). Among other previously unknown tectomicrobial taxa, a novel sponge-associated genus (*Ca.* Thalassonella) was identified for which several MAGs were obtained and analyzed (**Chapter 4**). Considering the above, it is apparent that the application of cultivation-independent methods assists in resolving the phylogeny of understudied taxonomic groups and sheds light on the hidden microbial diversity of the sponge holobiont (**Chapter 4** and **5**).

The tree of life has significantly expanded owing to recent advances in high-throughput sequencing and computation (Parks et al., 2018). The explosion of genomic data has raised the need for establishing a standardized classification system for uncultured taxa (Konstantinidis et al., 2017; Murray et al., 2020). Genome-based methods seem to offer a better resolution and an improved phylogenetic signal than the 16S rRNA gene for the construction of a robust taxonomic framework for prokaryotes (Hugenholtz et al., 2021). Inference of phylogenetic trees based on concatenated alignments of single-copy, highly conserved proteins have been extensively applied (**Chapter 2**, **3** and **5**) and are considered as a suitable approach for defining bacterial taxonomy (**Chapter 5**) (Lang et al., 2013; Rinke et al., 2013; Tonini et al., 2015; Parks et al., 2018; Parks et al., 2020; Hugenholtz et al., 2021). Nevertheless, information derived from comparative analyses of 16S rRNA gene sequences also proved to be efficient in the classification of uncultured bacteria (**Chapter 4**) (Yarza et al., 2014). Other classification methods include the use of whole genome similarity measures (e.g., average amino acid or nucleotide identity) or taxonomic rank normalization (e.g., relative evolutionary divergence distances) (**Chapter 5**). Charting the unknown microbial majority in the light of the torrent of genomic data, prerequires a consensus on establishing a unified and comprehensive taxonomic system. Concerted efforts have been made by proposing genomic standards for the description of uncultivated taxa, naming conventions and the designation of type material, and yet a general agreement between taxonomists has not been reached (Konstantinidis et al., 2017; Chuvochina et al., 2019; Murray et al., 2020; Hugenholtz et al., 2021). All of the above have a direct impact on the progress of microbiome research, including the study of the microbial diversity in marine

sponges. In **this chapter**, for consistency purposes I decided to follow the International Code of Nomenclature of Prokaryotes (ICNP) and the bacterial taxonomy as proposed by the Genome Taxonomy Database (GTDB).

A glimpse into the lifestyle of sponge symbionts

Microbial diversity has been a popular topic among sponge microbiologists given the wide application of 16S rRNA gene amplicon sequencing data. However, metagenome-based methods have granted access to functional aspects of the sponge microbiome. In this thesis, we elucidated the metabolic capabilities of different members of the sponge holobiont by metagenome binning, comparative genomic analysis, genome mining and in vitro testing (**Chapter 2**, **3**, **4** and **5**).

Sponge-associated microbes benefit from the generous supply of nutrients derived indirectly from the pumped seawater or produced directly from the host (Taylor et al., 2007a; Hentschel et al., 2012; Pita et al., 2018). Nevertheless, the role of heterotrophy in sponge-dwelling communities remains poorly described even though dissolved organic matter (DOM) is the primary carbon source for sponges and their associated microbiota (Rix et al., 2020; Burgsdorf et al., 2021). DOM represents the largest heterotrophic source of nutrients in the oceans and is comprised of a complex pool of compounds, including glucose, amino acids and algae-produced substances (Rix et al., 2020). Recent quantitative data revealed that heterotrophic microbial symbionts participate actively and majorly in the assimilation of DOM and thus, in the nutrient acquisition of the sponge host (Rix et al., 2020). Furthermore, genomic evidence of features related to carbon metabolism in sponge holobionts revealed high potential in the degradation of complex carbohydrates (Pita et al., 2018). This was highlighted by the prevalence of genes encoding carbohydrate active enzymes (CAZymes) in sponge-associated genomes (**Chapter 2** and **5**). Complex carbohydrates (oligo- and polysaccharides) are ubiquitous and diverse compounds in nature, forming numerous stereochemical combinations. CAZymes are the group of enzymes responsible for the assembly, transformation and catabolism of complex carbohydrates (Lombard et al., 2014). Many members of the sponge microbiota have been reported as highly potent degraders of complex polymers, including members of the phyla *Ca.* Poribacteria (Kamke et al., 2013; Chaib De Mares et al., 2018), *Chloroflexota* (Bayer et al., 2018), SAUL ("sponge-associated unclassified lineage") (Astudillo-Garcia et al., 2018), *Verrucomicrobiota* (Sizikov et al., 2020), *Ca.* Tectomicrobia, *Armatimonadota* and *Firmicutes* (Chaib De Mares et al., 2018). Our analysis revealed two additional sponge-associated bacterial lineages, *Flavobacteriaceae* (**Chapter 2**) and *Candidatus* Nemesobacterales (**Chapter 5**), which hold an expanded genomic repertoire of complex carbohydrate-degrading enzymes. Particularly, the most abundant families of glycoside hydrolases (GH) and associated carbohydrate-binding modules (CBM) were targeting N-acetylgalactosamine (GH109), xyloglucans (GH74 and CBM44), cellulose (CBM44), animal chitin (GH74 and CB50) and bacterial peptidoglycans (GH23 and CBM50)

(**Chapter 2** and **5**). Chitin is a major structural biopolymer of the sponge skeleton and together with spongin (collagen), proteoglycans and glycoproteins form the extracellular matrix of marine sponges (Ehrlich et al., 2007). In addition, N-acetylgalactosamine residues decorate glycoproteins and glycolipids possibly present in the sponge mesohyl, cell walls and organic matter from the filtered seawater (Kamke et al., 2013). Xyloglucans and cellulose are components of plant cell walls and green algae (Bennke et al., 2016). Therefore, our findings highlight the extensive potential of sponge symbionts to breakdown a wide array of complex molecules present in the host matrix, bacterial or sponge cell walls and the seawater in general (**Chapter 2** and **5**). Interestingly, the same CAZy families were also the most abundant in almost all of the polysaccharide-degrading lineages reported before (Kamke et al., 2013; Astudillo-Garcia et al., 2018; Bayer et al., 2018; Chaib De Mares et al., 2018). On the other hand, several sponge symbionts displayed a limited carbohydrate metabolism (Bayer et al., 2020; Knobloch et al., 2020). This metabolic specialization of microbes that represent certain lineages inhabiting a similar environmental niche and encoding equivalent genes for carbohydrate utilization indicates the possible presence of functional guilds within the sponge microbiome (Fan et al., 2012).

The exchange of amino acids, cofactors and vitamins between symbionts and sponge host has been previously identified as another prominent example of the metabolic interdependencies in the sponge holobiont (Figure 1) (Thomas et al., 2010a; Fan et al., 2012b; Hentschel et al., 2012; Webster and Thomas, 2016; Moitinho-Silva et al., 2017a; Pita et al., 2018; Silva et al., 2019b). By means of genome-resolved metagenomics, we detected complete pathways for the biosynthesis of several amino acids (e.g., serine, threonine, proline, ornithine), cofactors (e.g., coenzyme A, heme, molybdenum cofactor) and B vitamins (e.g., riboflavin, pantothenate, biotin) in the studied sponge-associated bacteria (**Chapter 4** and **5**). The biosynthetic potential for similar metabolites has been observed in numerous other sponge symbionts (Lackner et al., 2017; Astudillo-Garcia et al., 2018; Bayer et al., 2020; Engelberts et al., 2020; Knobloch et al., 2020; Robbins et al., 2021; Waterworth et al., 2021), supporting the notion that synthesis of these essential molecules represents a microbially-mediated function common within the sponge microbiome (Hentschel et al., 2012; Pita et al., 2018). In addition, Fiore et al. (2015) observed the expression of genes for the biosynthesis of various amino acids and B vitamins in the symbiont, but not in the sponge metatranscriptome, suggesting a form of synergism between symbiont and host via these metabolites. Accordingly, Munroe et al. (2019) reported arginine as a potential essential amino acid for sponges since increasing arginine concentrations tend to increase the in vitro metabolic activity of sponge cells in the absence of a complete biosynthesis pathway. This might be another indication that sponges receive necessary molecules from external sources, such as their symbionts. Another characteristic example is cobalamin (vitamin B12) which sponges cannot synthesize and likely depend on their symbionts for its acquirement (Thomas et al., 2010a). Indeed, in **chapter 5** a complete biosynthesis pathway for cobalamin was identified only in MAGs associated with sponges rather than in the ones derived from seawater. Apparently, hosts benefit from their symbiotic

relationships since in most cases they are not capable of producing these and other essential compounds and they have been long thought to obtain at least some of them from their microbial residents. In turn, symbiotic microbes may also come out as beneficiaries in this interplay receiving part of their nutrients from the host and ensuring their niche, as hosts might select for producers of necessary molecules (Fan et al., 2012; Hentschel et al., 2012; Webster and Thomas, 2016; Pita et al., 2018).

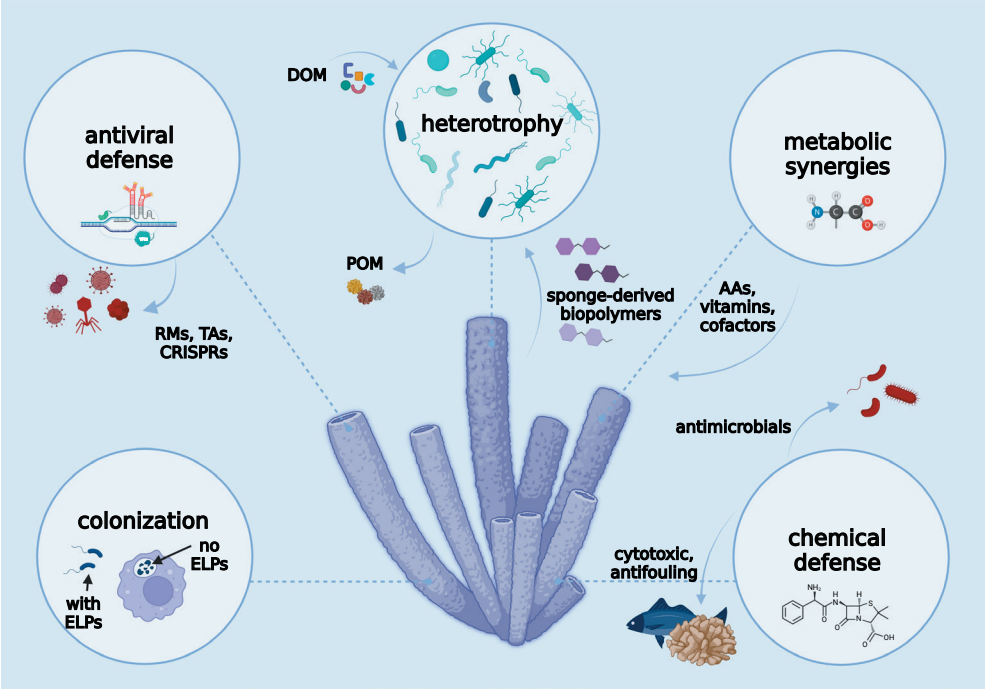


Figure 1. Schematic representation of microbial interactions within the sponge holobiont as inferred from the work performed and discussed in this thesis. RMs; restriction-modification systems, TAs; toxin-antitoxin systems, CRISPRs; clustered regularly interspaced short palindromic repeats, ELPs; eukaryotic-like proteins, DOM; dissolved organic matter, POM; particulate organic matter, AAs; amino acids (Adapted from Robbins et al., 2021). (Created with https://biorender.com/).

These metabolic collaborations can be facilitated by the presence of membrane transporters such as those belonging to the ATP-binding cassette (ABC) transporter superfamily (Pita et al., 2018). ABC transporters are ubiquitous and actively participate in various biological processes ranging from nutrient uptake to detoxification by translocating substrates across cellular membranes (Jones and George, 2004). A diverse array of ABC transporters for ions, phosphate, general L- and branched-chain amino acids, oligopeptides, heme and others were detected in the genomes of both *Ca.* Thalassonella (**Chapter 4**) and *Ca.* Nemesobacterales (**Chapter 5**) reflecting their heterotrophic lifestyle and the dynamic exchange of compounds with the sponge host. It is noteworthy that monosaccharide and/or oligosaccharide ABC transporters were not fully annotated in the analyzed MAGs

(**Chapter 4** and **5**). Differences in sugar transporters have been previously reported between *Candidatus* Tethybacterales families where multiple sugar transporters found abundant in *Ca.* Tethybacteraceae were almost completely absent in *Candidatus* Polydorabacteraceae and *Candidatus* Persebacteraceae (Taylor et al., 2021; Waterworth et al., 2021). This might indicate that symbionts may use various transporters and hence, have different potential in transporting carbohydrates.

Another significant trait of the symbionts that mediates the interactions within the sponge holobiont is the production of biologically active secondary metabolites (Hentschel et al., 2012; Webster and Thomas, 2016; Pita et al., 2018). Displaying a remarkable chemical diversity, these compounds range from terpenoids and alkaloids to peptides and polyketides with diverse bioactivities such as antimicrobial, anticancer and antifouling (Taylor et al., 2007a). Most studies have focused on the biotechnologically relevant biological activities of the metabolites produced by the sponge holobiont, which will be discussed later in this chapter. Yet, unveiling their ecological context remains challenging. Insights into the nature and function of symbiont-derived bioactive compounds have been provided by both culture-dependent and -independent approaches (Florez et al., 2015). For example, isolation of sponge symbionts producing secondary metabolites and subsequent screening assays revealed their antifouling properties since they inhibited the settlement of invertebrate larvae and the biofilm formation of bacteria known to induce this settlement on sponges (Dash et al., 2009). Moreover, numerous substances originating from sponge symbionts have been shown to strongly deter fish or other predators in the marine environment (Florez et al., 2015).

Recent advances in sequencing and bioinformatic tools have made computational mining of genome sequences for biosynthetic gene clusters (BGCs) possible. These BGCs are groups of genes found adjacently in a genome encoding the pathway for the biosynthesis of specialized metabolites (Medema and Fischbach, 2015). One of the many advantages of computational BGC mining in natural product discovery is the ability to tap into the BGC repertoire of previously uncharted resources. Paoli et al. (2021) explored the biosynthetic potential of the global ocean microbiome, unveiling almost 40,000 putative BGCs, most of which belonged to uncharacterized gene cluster families. BGCs with the highest prevalence were ribosomally synthesized and post-translationally modified peptides (RiPPs), aryl polyenes, carotenoids, ectoines and siderophores (Paoli et al., 2021). Throughout this thesis, we investigated the BGC repertoire of various sponge-associated bacteria belonging to diverse phylogenetic groups (**Chapter 2, 3, 4** and **5**) since reports on BGCs in sponge microbiomes are scarce. We showed in **chapters 2, 3** and **4** that terpene BGCs were among the most abundant and prevalent. Interestingly, in **chapter 5** terpene BGCs were only present in seawater-derived MAGs. Even though the exact role of terpenes in the sponge holobiont is unknown, their strong antioxidant potential might imply a quenching and/or scavenging function for reactive oxygen species and thus, protecting both microbes and the host under oxidative stress conditions (Miki et al., 1996; Regoli et al., 2000; Webster, 2007).

Besides epibionts, predators and stress, pathogens are also challenges that sponges have to deal with in their natural environment (Figure 1) (Webster, 2007). One defensive strategy of the sponge symbionts is the production and secretion of compounds with antimicrobial activity. BGCs coding for bacteriocins were prevalent in sponge-associated *Bacteroidota* (*Flavobacteriaceae*) (**Chapter 2** and **3**), *Proteobacteria* and *Actinobacteriota* (**Chapter 3**). Bacteriocins are ubiquitous RiPPs that act as toxins with a relatively narrow killing spectrum, often specific to closely related bacteria. Therefore, it is tempting to speculate that bacteriocin-producing bacteria could have a dual role in the sponge holobiont: 1) reducing competition for resources or preventing host infections by inhibiting the establishment of foreign or pathogenic strains and 2) mediating microbial community interactions as anti-competitors or quorum sensing molecules (Riley and Wertz, 2002; Gillor et al., 2008). Compounds with cytotoxic effects have often been retrieved from sponge symbionts, as well. In our case, almost 80% of the tested sponge-associated isolates displayed medium to high cytotoxicity (**Chapter 3**). An example of symbiosis-related toxicity in nature is the sponge *Terpios hoshinota* and its cyanobacterial symbionts. Production of cytotoxic secondary metabolites combined with the encrusting growth of the sponge led to massive coral mortality (Teruya et al., 2004; Shi et al., 2012). Taken together, the sponge microbiome mediates the interactions of the holobiont in terms of predation, chemical defense and competition and thus, affects the benthic community structure and dynamics.

Dealing with the host and establishing residency

Genomic signatures of sponge-associated bacteria have provided compelling insights into the underlying molecular mechanisms for niche adaptation and host association, since sponge-dwelling bacteria remain largely uncultivable. A common strategy is to compare genomes of sponge- and seawater-derived microbes and look for differences in the genomic content and predicted metabolic pathways (Thomas et al., 2010a; Fan et al., 2012; Slaby et al., 2017; Podell et al., 2019). In **chapter 5**, we observed metabolic divergence between sponge-associated members of *Candidatus* Dadabacteria (now assigned to *Candidatus* Desulfobacterota) and their free-living counterparts. These lineages shared many core metabolic traits related to carbon, energy, amino acid and cofactor metabolism. Yet, sponge-associated MAGs were enriched in genes involved in host colonization and defense, potentially serving as adaptive features to the sponge microenvironment. Moreover, functional clustering showed that phylogeny and environmental source could shape the inferred metabolism of *Ca.* Dadabacteria ('Desulfobacterota__D') (**Chapter 5**). On the other hand, metabolic profiles of *Flavobacteriaceae* diverged among phylogenetic groups but not according to host association (**Chapter 2**). It should be mentioned that flavobacterial genomes studied in **chapter 2** originated from isolates while *Ca.* Dadabacteria ('Desulfobacterota__D') genomes in **chapter 5** were metagenome-assembled. Indeed, isolation of host-associated strains in the laboratory might not confirm their symbiotic relationship with the host in nature

while cultivation-independent studies with direct access to the environmental DNA might give a more realistic view of the interactions. Another assumption is that marine microbes in the quest for nutrients might follow a generalist-like lifestyle and display metabolic versatility colonizing both host-associated and free-living systems (Silva et al., 2019b). This might be true for flavobacteria probably being opportunistic sponge symbionts whereas *Ca.* Nemesobacterales are likely stricter symbionts, since all representative MAGs were exclusively sponge-derived (**Chapter 2 and 5**).

Successful establishment in a host requires certain adaptation mechanisms that confer to symbionts the ability to persist within the host. In the case of sponges, feeding is happening via filtering and subsequent phagocytosis of seawater bacteria and particles and thus, symbiotic microbes need to be distinguished from food and avoid being digested. This is achieved by either recognition of symbionts by the sponge or masking mechanisms of symbionts to evade phagocytic host cells. Several omics surveys have revealed the presence of certain genomic features of the sponge-associated microbes that mediate host-microbe interactions implying a symbiotic lifestyle (Engelberts et al., 2020; Robbins et al., 2021; Taylor et al., 2021). For instance, genes coding for eukaryotic-like proteins (ELPs) could be considered “symbiosis markers”, since they contain repeats found typically in eukaryotes and are simultaneously detected in large abundances in genomes of sponge bacteria (Reynolds and Thomas, 2016). Here, comparison between sponge- and seawater-derived MAGs showed that symbionts of the order *Ca.* Nemesobacterales residing in various sponge species are enriched in ELP-encoding genes (**Chapter 5**). ELPs contain various classes of molecules (e.g., ankyrin-repeat proteins (ARPs) and tetratricopeptide repeats (TPRs)) whose exact function in sponge symbiosis has not been determined yet. Gene expression and phagocytosis assays have indicated that ELPs found in symbionts can interfere with the processing and development of the phagosome (Nguyen et al., 2014; Reynolds and Thomas, 2016), a significant trait to maintain a stable symbiosis.

In addition, consistent with previous findings (Engelberts et al., 2020; Robbins et al., 2021) *Ca.* Nemesobacterales genomes were also enriched in protein domains related with Type II, III and IV secretion systems (**Chapter 5**). Metatranscriptomics has highlighted the cotranscription of genes encoding ELPs with those coding for a wide array of secretion systems (Diez-Vives et al., 2017). These specialized macromolecular machineries are likely responsible for the translocation of ELPs outside of the bacterial cells since their function is to deliver effector proteins into the extracellular space or inject them into a target cell (Costa et al., 2015). Within the sponge host, symbiont ELPs likely modulate phagocytosis thereby mediating host-microbe interactions and promoting colonization (Figure 1).

Another strategy advantageous to a symbiont could be the release of lipoproteins via ABC transporters into the host environment (Taylor et al., 2021). Bacterial lipoproteins have been associated with various functions, including virulence and the ability of pathogens to evade the host defenses, invade and colonize (Kovacs-Simon et al., 2011). All *Ca.* Nemesobacterales genomes contained the complete set of genes of the LolCDE complex (**Chapter 5**) and the ABC

transporter responsible for the translocation of lipoproteins in the outer membrane. Interestingly, these genes were completely absent from the respective seawater MAGs, providing additional support on the relevance of these genomic signatures for a symbiotic lifestyle. Gliding motility might constitute another symbiosis-related trait previously reported as essential for surface attachment and host colonization in several human (Williams et al., 2018) and fish pathogens (Alvarez et al., 2006), as well as plant-root-associated bacteria (Kolton et al., 2014). Even though no specific correlation between gliding motility, type 9 secretion (T9SS) and host-association was observed, all sponge-associated flavobacteria studied in **Chapter 2** had a complete set of gliding and T9SS genes (Figure 5, **Chapter 2**).

Besides host phagocytosis, symbiotic bacteria have to survive the extensive exposure to viruses, potential pathogens and toxins present in the ambient seawater pumped at high rates through the sponge. For this reason, sponge-associated bacteria likely need a suite of immunity mechanisms that will protect them against foreign DNA from other microbes and viruses and facilitate their adaptation to the host environment. Indeed, enrichment of genes associated with prokaryotic defense systems [(e.g., restriction-modification systems (RMs), toxin-antitoxin systems (TAs), clustered regularly interspaced short palindromic repeats (CRISPRs)] in symbiont genomes over seawater genomes has been repeatedly described (**Chapter 5**) (Fan et al., 2012; Horn et al., 2016; Slaby et al., 2017; Astudillo-Garcia et al., 2018; Podell et al., 2019; Moreno-Pino et al., 2020; Sizikov et al., 2020; Robbins et al., 2021). Among the most abundant defense systems in *Ca.* Nemesobacterales were RMs and TAs (**Chapter 5**), in accordance with several symbiont lineages such as *Ca.* Poribacteria (Podell et al., 2019), SAUL (Astudillo-Garcia et al., 2018) and *Verrucomicrobiota* (Sizikov et al., 2020). RMs are ubiquitous in microbes and act by modifying the 'self' DNA and destroying any unmodified foreign DNA (Makarova et al., 2013). Sponge symbionts could use RMs not only to protect themselves against invading genetic material but also to promote horizontal DNA exchange within the sponge holobiont (Astudillo-Garcia et al., 2018). TAs induce programmed cell death or dormancy upon phage infection or as a response to environmental stressors (e.g., amino acid starvation, antibiotics, oxidation) (Harms et al., 2018a). In such manner, symbionts are capable of surviving phage attacks and bioactive metabolites produced by the sponge or other bacteria and eventually form persistent cells. Altogether, the expanded repertoire of defense systems predicted for sponge symbionts underscores their significance in mediating host-microbe interactions and achieving a prosperous symbiosis (Figure 1).

From genotype to phenotype...a long way to go

Advances in sequencing technologies and bioinformatics have facilitated the concept of associating phenotypes to the presence and absence of genes in bacterial genomes (Dutilh et al., 2013). Throughout this thesis, we used extensive sequence data and performed comparative analyses to understand the genotype to an extent that the phenotypic traits can be predicted and hence, elucidate the metabolism of sponge-associated bacteria. In **chapter 2 and 3**, a combination of in vitro testing and whole-genome sequencing of bacterial strains isolated from sponges allowed for gene-trait matching analysis and provided with important hints on the genotype-phenotype association. More specifically, most of the tested isolates showed neither motility on agar (**Chapter 2**) nor antimicrobial activity in vitro, even though they had a complete set of gliding genes and a wide variety of secondary metabolite biosynthetic gene clusters (**Chapter 2 and 3**). Apparently, cultivated microbes under standard laboratory conditions might show no or very low expression of genes or biosynthetic gene clusters (Chiang et al., 2011; Reen et al., 2015). Since an observable phenotypic trait might be the combined result of a gene or multiple genes, abiotic and biotic factors (Orgogozo et al., 2015) applying the proper growth conditions to simulate the natural environment is of vital importance in triggering the expression of these pathways (**Chapter 2**). Moreover, biological systems have multiple levels (from DNA to metabolites) and thus, omics data integration combined with experimental data is needed to identify true associations and obtain a more comprehensive understanding of genotype-phenotype interactions (Ritchie et al., 2015; Prosser, 2020) (Figure 2).

In vitro and in silico methods: joined forces

Marine natural products hold great promise as drug candidates in the face of the antimicrobial resistance crisis and other health care challenges. A considerable research effort has been dedicated to the natural product chemistry of sponges, since they are considered 'gold mines' of bioactive secondary metabolites. Up until recently, many of these natural products were thought to be of sponge origin but increasing evidence implicate the associated bacteria in the production of bioactive compounds. This is attributed to their structural resemblance to bacterial metabolites or because the type of bioactive molecule is commonly found in bacteria and not in animals (Taylor et al., 2007a; Piel, 2009; Uria and Piel, 2009). The first insights into the producing microbes were obtained via cultivation-dependent methods leaving an enormous gap in the real biochemical diversity due to the limited amount of cultivated strains. Cultivation-independent approaches came later to bridge this gap revealing the hidden biosynthetic novelty of the uncultivated fraction of microbes and help in identifying the true producers of the bioactive compounds (Hill and Sacristán-Soriano, 2017).

There are several success stories of describing symbiont-produced chemicals in a cultivation-free manner (Florez et al., 2015). In **chapter 4**, we contributed to one of these stories by

tracing and localizing a putative *Ca.* Entotheonella producer of the anticancer compound psymberin in the tissue of *Psammocinia* sp. This was achieved by means of a 16S rRNA gene-based prioritization strategy and chemically-aided separation of filamentous bacterial cells from the sponge matrix, respectively (**Chapter 4**). Nevertheless, harnessing the biosynthetic potential of already available bacterial isolates is equally important and can bring light to undiscovered capacity (Wilson and Piel, 2013). Genome mining for BGCs and bioactivity screening of sponge-associated strains detected many candidate BGCs potentially involved in the observed anticancer and anti-MRSA activity (**Chapter 3**). Even though circumventing methods that rely on cultivation offers more advantages and fewer limitations, we should always consider cultivation-based approaches as complementary strategies since they can provide useful insights in the prioritization and eventual isolation of promising drug leads.

Challenges and Future perspectives

Sponges can be viewed as evolution incubators since they represent the oldest extant animal-microbe symbiosis encompassing a great microbial biodiversity. Studying sponges and their associates in the omics era has largely enhanced our understanding of the interactions that shape and underpin the sponge holobiont functioning. Most of the generated data is derived from whole genome or lately metagenome sequencing providing important clues about the sponge microbiome composition and metabolic potential. Yet, the obtained information is only a snapshot of 'what is there' and 'what they can do' leaving many fundamental questions unanswered. Future research attempts on integrating high-throughput omics data such as metagenomes, metatranscriptomes, metaproteomes and metabolomes will give additional valuable hints on diverse aspects of complex (micro)biological systems, including sponges. This data integration could also deepen our knowledge of systems as a whole (Figure 2).

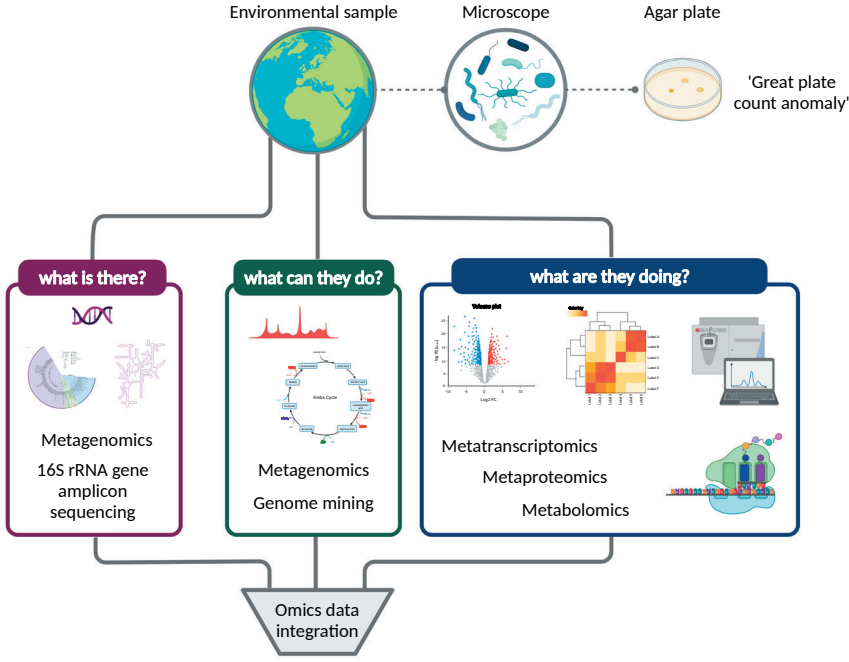


Figure 2. Overview of omics and other data analysis approaches that allow direct access to the environmental sample, bypassing cultivation. Data integration could address significant biological questions about the composition, metabolic potential and actual functions of the studied microbiome. (Created with https://biorender.com/).

Adopting a holistic perspective that accounts for the sponge and its symbiotic microbes is of paramount importance in evaluating the holobiont function. This is also highlighted by global initiatives that focus on investigating symbiosis in aquatic systems and dedicate a significant part of their research on eukaryotic hosts, such as those supported by the Gordon and Betty Moore Foundation (https://www.moore.org/initiative-strategy-detail?initiativeId=symbiosis-in-aquatic-systems-initiative) and the Aquatic Symbiosis Genomics Project (McKenna et al., 2021). Furthermore, the study of sponges as model animals may help in acquiring a more concrete and complete picture of the unique set of mechanisms underlying sponge-microbe interactions as previously proposed (Pita et al., 2016; Hentschel, 2021).

Even though available genetic data on sponge hosts has increased during the past decade [reviewed in (Pita et al., 2016)], functional information in the context of symbiosis is still lacking. To fully harness the obtained information, creation of large inventories of biological data derived from both the host and the associated microbes could prove beneficial in addressing queries about the metabolic basis of symbiosis. This might seem daunting in the light of the currently available large-scale datasets. Therefore, new computational methods (e.g., data mining algorithms) need to be developed to cope with the downstream analyses

of the vast amount of generated data. Moreover, the existing data needs to be refined and manually curated as in many cases data registered in public databases (e.g., annotations, sample metadata) is either inaccurate or insufficient.

To narrow the gap between the data inundation and the understanding of the biology behind complex symbiotic systems, it is vital to implement the acquired information in the experimental process. For instance, to alleviate cultivation recalcitrance of the sponge symbionts, a promising strategy would be the application of reverse genomics combined with innovative approaches for targeted cultivation (Cross et al., 2019). First, prioritizing cultivation attempts for key players that will provide important insights into the sponge holobiont or symbionts of high biotechnological value, will maximize effectiveness. Next, information present in near-complete genomes derived from sponge metagenomes could be used to design antibodies that are labelled fluorescently. Antibodies bind specifically to the target bacteria followed by sorting via flow cytometry. Viable cells could then be used for downstream cultivation. Genomic and functional data could help in identifying the proper substrates and growth factors that will lead to the successful isolation and propagation of the culture (Cross et al., 2019; Lewis et al., 2021).

Concluding Remarks

Now zooming out of the sponge microbiome, over these five years of research a shift towards molecular and omics-based techniques has dominated the microbial ecology field since sequencing has become the trend. This is also reflected in the work performed during this thesis. Our findings (**Chapter 2 and 3**) highlight the gap between omics-derived information and experimental testing such as the presence of cryptic genes/pathways or the use of insufficient growth factors to trigger certain phenotypes. Due to the data explosion, there is a tendency of turning the scientific methodology from hypothesis-driven into a more exploratory approach. It is significant to identify a scientific question that would serve as a basis for the design of an experiment in order to prove a hypothesis and explain a biological phenomenon. This, in the framework of exploiting the full potential of the available tools, will increase the probability of unequivocally confirming or refuting the assumed scientific concepts. Indeed, sequence data has to offer a lot of important information, and this is in line with the work presented here (**Chapter 2, 3, 4 and 5**) where omics shed light on novel sponge-associated phylotypes, features that point at a symbiotic lifestyle and an expanded secondary metabolite biosynthesis arsenal with biotechnological potential. In conclusion, despite the fact that the concept of symbiosis existed even before the emergence of metazoans our understanding is still in its infancy. I believe that this thesis played its part in elucidating aspects of sponge-microbe interactions and I truly hope it will be an inspiration for future research endeavors.

Summary

Sponges (phylum Porifera) are the oldest known extant metazoans, the most prolific sources of bioactive molecules in the sea and of great ecological significance in marine benthic communities due to their immense pumping capacity. Marine sponges also represent an excellent example of complex invertebrate holobionts since they live in close associations with their microbial symbionts. The microbial consortia inhabiting sponges are dense and diverse, spanning all three domains of life. Yet, the majority of sponge-associated microorganisms remain uncultured, hampering efforts in determining their interactions with their sponge host. To this end, a broad range of cultivation-independent methods has been developed that provide promising avenues for unravelling important aspects of host-microbe symbiosis. The research described in this thesis aims at enhancing our understanding of the interplay between sponges and bacteria, focusing on both cultured and yet-uncultured sponge-associated lineages. A brief introduction to marine sponges and their symbiotic relationships with microorganisms is given in **chapter 1**. This chapter further outlines the cultivation challenge, how omics-derived information has shed light into sponge-microbe interactions and the motivation behind the research presented here.

More than 40 microbial phyla have been reported as part of the global sponge microbiome. A predominant phylum is the *Bacteroidetes* with most of its members belonging to the family *Flavobacteriaceae*, also known as flavobacteria. In **chapter 2**, different functional traits common among flavobacteria inhabiting various niches, both host-associated and free-living, were investigated. Comparison of their genome-derived metabolic profiles revealed functional divergence between the different phylogenetic groups irrespective of their isolation source. This indicated a putative metabolic versatility for marine flavobacteria as opportunistic sponge symbionts with a dual lifestyle between seawater and host. Marine flavobacteria showed an extensive genetic repertoire for gliding motility, utilization of a wide variety of complex carbohydrates and production of secondary metabolites with interesting bioactivities. Nevertheless, phenotypic assays on gliding motility and antimicrobial activity did not reflect the genomic potential of the bacterial isolates implying the silent state of their genes under the selected in vitro conditions.

To assess the true functional potential of the sponge symbionts, in vitro screening for antimicrobial and cytotoxic activity combined with genome mining for secondary metabolite biosynthetic gene clusters (BGCs) of 21 sponge-derived bacterial isolates was performed, as described in **chapter 3**. The tested isolates represented major sponge-associated taxa with the most active ones belonging to *Proteobacteria* and *Actinobacteria* in accordance with previous studies. Most microbial crude extracts displayed cytotoxicity while one only caused growth inhibition of the methicillin-resistant *Staphylococcus aureus* (MRSA) among a broader test panel of potential bacterial and eukaryotic targets. Moreover, gene-trait matching provided important clues on the genotype-phenotype association by detecting BGCs likely responsible for the observed bioactivity.

Sponges host bacteria which are prominent producers of secondary metabolites but remain recalcitrant to cultivation. To the best of our knowledge, among the most 'chemically talented', yet-uncultured bacteria are the ones belonging to the candidate genus Entotheonella of the candidate phylum Tectomicrobia. In **chapter 4**, we showed that *Ca.* Tectomicrobia inhabit various environmental niches. According to sequencing data, *Ca*. Entotheonella can be divided into three candidate genera including a novel sponge-associated genus *Candidatus* Thalassonella. Metabolic reconstruction predicted that *Ca.* Thalassonella leads a facultatively anaerobic, heterotrophic lifestyle similar to *Ca.* Entotheonella but lacks in the chemical richness. Moreover, this chapter highlights the presence of a 'superproducer' group within *Ca.* Entotheonella in phylosymbiosis with sponge hosts. Further observations regarding this group led to the detection of a putative producer of the polyketide psymberin that displays potent anticancer activity.

Another yet-uncultured bacterial lineage which is understudied even though ubiquitous in sponge microbiomes is the candidate phylum Dadabacteria, recently united with the candidate phylum Desulfobacterota. We were the first to resolve the phylogeny and predict the primary and secondary metabolism of the sponge-associated *Ca.* Dadabacteria ('Desulfobacterota__D') as described in **chapter 5**. Genome reconstruction and phylogenomics uncovered a sponge-specific order (*Ca.* Nemesobacterales) and several novel taxa expanding the genome representation of the sponge-associated *Ca.* Dadabacteria. Functional analysis predicted that *Ca.* Nemesobacterales are aerobic heterotrophs capable of utilizing complex carbohydrates similar to other sponge symbionts. Niche speciation was also evident since functional divergence was predicted between sponge- and seawater-derived metagenome-assembled genomes (MAGs) of the same lineage attributed to distinct genomic features related to symbiosis. *Ca.* Nemesobacterales were significantly enriched in genes associated with host-microbe recognition, adaptation to the host environment and secondary metabolite production reflecting a symbiotic lifestyle. Moreover, fluorescent in situ hybridization was employed to visualise *Ca.* Nemesobacterales for the first time. The visualization of *Candidatus* Nemesobacter rappii revealed their presence inside bacteriocytes in the tissue of the sponge *Geodia barretti*. This further supports the functional predictions that *Ca.* Nemesobacterales live in close association with marine sponges.

To conclude, **chapter 6** discusses the research outcomes and highlights of my zooming into the sponge microbiome in the context of other relevant studies. Research described in this thesis set out to unravel the interactions between sponges and their symbiotic bacteria and bridge the gap between the enormous amount of omics-derived information and our understanding of the sponge holobiont. In conclusion, the sponge microbiome mediates the interactions of the holobiont via heterotrophy, potential metabolite exchange (e.g., amino acids and vitamins) and chemical defense and adapts to the environment of the host with the help of defense systems leading to an intimate symbiotic relationship with the sponge. To this end, future research efforts in sponge-microbe symbiosis should consider joining the forces of cultivation-dependent and -independent approaches since they hold great promise in uncovering true associations in complex living systems, including marine sponges.

References

1. Abdelmohsen, U.R., Bayer, K., and Hentschel, U. (2014) Diversity, abundance and natural products of marine sponge-associated actinomycetes. *Nat Prod Rep* **31**: 381-399.
2. Abdelmohsen, U.R., Pimentel-Elardo, S.M., Hanora, A., Radwan, M., Abou-El-Ela, S.H., Ahmed, S., and Hentschel, U. (2010) Isolation, phylogenetic analysis and anti-infective activity screening of marine sponge-associated actinomycetes. *Mar Drugs* **8**: 399-412.
3. 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 Methods* **11**: 1144-1146.
4. Alneberg, J., Karlsson, C.M.G., Divne, A.M., Bergin, C., Homa, F., Lindh, M.V. et al. (2018) Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. *Microbiome* **6**: 173.
5. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.
6. Alvarez, B., Secades, P., Prieto, M., McBride, M.J., and Guijarro, J.A. (2006) A mutation in *Flavobacterium psychrophilum* tlpB inhibits gliding motility and induces biofilm formation. *Appl Environ Microbiol* **72**: 4044-4053.
7. Andrews, S. (2010) FASTQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc. In.
8. Apprill, A. (2020) The Role of Symbioses in the Adaptation and Stress Responses of Marine Organisms. *Ann Rev Mar Sci* **12**: 291-314.
9. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020) KofamKOALA: KEGG ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* **36**: 2251-2252.
10. Astudillo-Garcia, C., Slaby, B.M., Waite, D.W., Bayer, K., Hentschel, U., and Taylor, M.W. (2018) Phylogeny and genomics of SAUL, an enigmatic bacterial lineage frequently associated with marine sponges. *Environ Microbiol* **20**: 561-576.
11. Audoin, C., Bonhomme, D., Ivanisevic, J., de la Cruz, M., Cautain, B., Monteiro, M.C. et al. (2013) Balibalosides, an original family of glucosylated sesterterpenes produced by the Mediterranean sponge *Oscarella balibaloi*. *Mar Drugs* **11**: 1477-1489.
12. Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyerreil, L.A., and Thingstad, F. (1983) The Ecological Role of Water-Column Microbes in the Sea. *Mar Ecol Prog Ser* **10**: 257-263.
13. Bachmann, B.O., Van Lanen, S.G., and Baltz, R.H. (2014) Microbial genome mining for accelerated natural products discovery: is a renaissance in the making? *J Ind Microbiol Biotechnol* **41**: 175-184.
14. Baker, B.J., and Dick, G.J. (2013) Omic Approaches in Microbial Ecology: Charting the Unknown. *Microbe* **8**: 353-360.
15. Balagam, R., Litwin, D.B., Czerwinski, F., Sun, M., Kaplan, H.B., Shaevitz, J.W., and Igoshin, O.A. (2014) *Myxococcus xanthus* gliding motors are elastically coupled to the

substrate as predicted by the focal adhesion model of gliding motility. *PLoS Comput Biol* **10**: e1003619.

16. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S. et al. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* **19**: 455-477.
17. Barbeyron, T., Thomas, F., Barbe, V., Teeling, H., Schenowitz, C., Dossat, C. et al. (2016) Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* Dsij(T). *Environ Microbiol* **18**: 4610-4627.
18. Bart, M.C., Hudspith, M., Rapp, H.T., Verdonschot, P.F.M., and de Goeij, J.M. (2021) A Deep-Sea Sponge Loop? Sponges Transfer Dissolved and Particulate Organic Carbon and Nitrogen to Associated Fauna. *Front in Mar Sci* **8**.
19. Bauer, A.W., Kirby, W.M., Sherris, J.C., and Turck, M. (1966) Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* **45**: 493-496.
20. Bayer, K., Jahn, M.T., Slaby, B.M., Moitinho-Silva, L., and Hentschel, U. (2018) Marine Sponges as *Chloroflexi* Hot Spots: Genomic Insights and High-Resolution Visualization of an Abundant and Diverse Symbiotic Clade. *mSystems* **3**: e00150-00118.
21. Bayer, K., Moitinho-Silva, L., Brummer, 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**: 832-843.
22. Bayer, K., Busch, K., Kenchington, E., Beazley, L., Franzenburg, S., Michels, J. et al. (2020) Microbial Strategies for Survival in the Glass Sponge *Vazella pourtalesii*. *mSystems* **5**.
23. Behnsen, J., and Raffatellu, M. (2016) Siderophores: More than Stealing Iron. *mBio* **7**.
24. Bell, J.J. (2008) The functional roles of marine sponges. *Estuar Coast Shelf Sci* **79**: 341-353.
25. Bennke, C.M., Kruger, K., Kappelmann, L., Huang, S., Gobet, A., Schuler, M. et al. (2016) Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic. *Environ Microbiol* **18**: 4456-4470.
26. Benson, D.A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., and Sayers, E.W. (2013) GenBank. *Nucleic Acids Res* **41**: D36-42.
27. Berg, G., Rybakova, D., Fischer, D., Cernava, T., Verges, M.C., Charles, T. et al. (2020) Microbiome definition re-visited: old concepts and new challenges. *Microbiome* **8**: 103.
28. Bernardet, J.F. (2011) Family I: *Flavobacteriaceae*. In *Bergey's Manual of Systematic Bacteriology*. Krieg, N.R., Staley, J.T., Brown, D.R., Hedlund, B.P., Paster, B.J., Ward, N.L. et al. (eds). New York: Springer, pp. 106-314.
29. Bernardet, J.F., and Nakagawa, Y. (2006) An Introduction to the Family *Flavobacteriaceae*. In *Prokaryotes*. New York: Springer, pp. 455-480.

30. Bernardet, J.F., Nakagawa, Y., and Holmes, B. (2002) Proposed minimal standards for describing new taxa of the family *Flavobacteriaceae* and emended description of the family. *Int J Syst Evol Microbiol* **52**: 1049-1070.
31. Bewley, C.A., Holland, N., and Faulkner, D.J. (1996) Two classes of metabolites from *Theonella swinhoei* are localized in distinct populations of bacterial symbionts. *Experientia* **52**: 716-722.
32. Bezuidt, O.K.I., Lebre, P.H., Pierneef, R., Leon-Sobrino, C., Adriaenssens, E.M., Cowan, D.A. et al. (2020) Phages Actively Challenge Niche Communities in Antarctic Soils. *mSystems* **5**.
33. Bielitza, M., and Pietruszka, J. (2013) The psymberin story--biological properties and approaches towards total and analogue syntheses. *Angew Chem Int Ed Engl* **52**: 10960-10985.
34. Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006) Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* **281**: 36269-36279.
35. Blanquer, A., Uriz, M.J., and Galand, P.E. (2013) Removing environmental sources of variation to gain insight on symbionts vs. transient microbes in high and low microbial abundance sponges. *Environ Microbiol* **15**: 3008-3019.
36. Blaser, M.J., Cardon, Z.G., Cho, M.K., Dangl, J.L., Donohue, T.J., Green, J.L. et al. (2016) Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges. *mBio* **7**.
37. Blin, K., Shaw, S., Kautsar, S.A., Medema, M.H., and Weber, T. (2021a) The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. *Nucleic Acids Res* **49**: D639-D643.
38. Blin, K., Shaw, S., Kloosterman, A.M., Charlop-Powers, Z., van Wezel, G.P., Medema, M.H., and Weber, T. (2021b) antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* **49**: W29-W35.
39. Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y. et al. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* **47**: W81-W87.
40. Blunt, J.W., Copp, B.R., Keyzers, R.A., Munro, M.H., and Prinsep, M.R. (2017) Marine natural products. *Natl Prod Rep* **34**: 235-294.
41. Blunt, J.W., Carroll, A.R., Copp, B.R., Davis, R.A., Keyzers, R.A., and Prinsep, M.R. (2018) Marine natural products. *Nat Prod Rep* **35**: 8-53.
42. Bolger, A.M., Lohse, M., and Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.
43. Bond, P.L., Hugenholtz, P., Keller, J., and Blackall, L.L. (1995) Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl Environ Microbiol* **61**: 1910-1916.

44. 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**: 2095-2113.
45. Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K. et al. (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**: 725-731.
46. Braun, T.F., Khubbar, M.K., Saffarini, D.A., and McBride, M.J. (2005) *Flavobacterium johnsoniae* gliding motility genes identified by mariner mutagenesis. *J Bacteriol* **187**: 6943-6952.
47. Brinkmann, C., Kearns, P., Evans-Illidge, E., and Kurtböke, D. (2017a) Diversity and Bioactivity of Marine Bacteria Associated with the Sponges *Candidaspongia flabellata* and *Rhopaloeides odorabile* from the Great Barrier Reef in Australia. *Diversity* **9**.
48. Brinkmann, C.M., Marker, A., and Kurtböke, D. (2017b) An Overview on Marine Sponge-Symbiotic Bacteria as Unexhausted Sources for Natural Product Discovery. *Diversity* **9**.
49. Brown, C.T., Hug, L.A., Thomas, B.C., Sharon, I., Castelle, C.J., Singh, A. et al. (2015) Unusual biology across a group comprising more than 15% of domain Bacteria. *Nature* **523**: 208-211.
50. Burgsdorf, I., Sizikov, S., Squatrito, V., Britstein, M., Slaby, B.M., Cerrano, C. et al. (2021) Lineage-specific energy and carbon metabolism of sponge symbionts and contributions to the host carbon pool. *ISME J*: 1-13.
51. 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**: e00391-00315.
52. Busch, K. (2021) Biodiversity, environmental drivers, and sustainability of the global deep-sea sponge microbiome (submitted).
53. Bushnell, B., Rood, J., and Singer, E. (2017) BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One* **12**: e0185056.
54. Calcabrini, C., Catanzaro, E., Bishayee, A., Turrini, E., and Fimognari, C. (2017) Marine Sponge Natural Products with Anticancer Potential: An Updated Review. *Mar Drugs* **15**.
55. Campanaro, S., Treu, L., Kougias, P.G., Zhu, X., and Angelidaki, I. (2018) Taxonomy of anaerobic digestion microbiome reveals biases associated with the applied high throughput sequencing strategies. *Sci Rep* **8**: 1-12.
56. Cao, D.D., Trinh, T.T.V., Mai, H.D.T., Vu, V.N., Le, H.M., Thi, Q.V. et al. (2019) Antimicrobial Lavandulylated Flavonoids from a Sponge-Derived *Streptomyces* sp. G248 in East Vietnam Sea. *Mar Drugs* **17**.
57. Cao, H., Dong, C., Bougouffa, S., Li, J., Zhang, W., Shao, Z. et al. (2016) Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge. *Sci Rep* **6**: 22842.

58. Capella-Gutierrez, S., Silla-Martinez, J.M., and Gabaldon, T. (2009) trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**: 1972-1973.
59. Caron, D. (2005) Marine microbial ecology in a molecular world: what does the future hold? *Sci Mar* **69**: 97-110.
60. Ceniceros, A., Dijkhuizen, L., Petrusma, M., and Medema, M.H. (2017) Genome-based exploration of the specialized metabolic capacities of the genus *Rhodococcus*. *BMC Genomics* **18**: 593.
61. Cerrano, C., Bavestrello, G., Arillo, A., Benatti, U., Bonpadre, S., Cattaneo-Vietti, R. et al. (1999) Calcium oxalate production in the marine sponge *Chondrosia reniformis*. *Mar Ecol Prog Ser* **179**: 297-300.
62. Cerveny, L., Straskova, A., Dankova, V., Hartlova, A., Ceckova, M., Staud, F., and Stulik, J. (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: role in virulence mechanisms. *Infect Immun* **81**: 629-635.
63. Cetkovic, H., Halasz, M., and Herak Bosnar, M. (2018) Sponges: A Reservoir of Genes Implicated in Human Cancer. *Mar Drugs* **16**.
64. Chaib De Mares, M., Jimenez, D.J., Palladino, G., Gutleben, J., Lebrun, L.A., Muller, E.E.L. et al. (2018) Expressed protein profile of a *Tectomicrobium* and other microbial symbionts in the marine sponge *Aplysina aerophoba* as evidenced by metaproteomics. *Sci Rep* **8**: 11795.
65. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2019) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*.
66. Chelossi, E., Milanese, M., Milano, A., Pronzato, R., and Riccardi, G. (2004) Characterisation and antimicrobial activity of epibiotic bacteria from *Petrosia ficiformis* (Porifera, Demospongiae). *J Expl Mar Biol Ecol* **309**: 21-33.
67. Chen, H. (2018) VennDiagram: Generate High-Resolution Venn and Euler Plots. R package version 1.6.20. https://CRAN.R-project.org/package=VennDiagram. In.
68. Chen, I.A., Chu, K., Palaniappan, K., Pillay, M., Ratner, A., Huang, J. et al. (2019) IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* **47**: D666-D677.
69. Cheng, W., Feng, Y.Q., Ren, J., Jing, D., and Wang, C. (2016) Anti-tumor role of *Bacillus subtilis* fmbJ-derived fengycin on human colon cancer HT29 cell line. *Neoplasma* **63**: 215-222.
70. Chiang, Y.-M., Chang, S.-L., Oakley, B.R., and Wang, C.C. (2011) Recent advances in awakening silent biosynthetic gene clusters and linking orphan clusters to natural products in microorganisms. *Curr Opin Chem Biol* **15**: 137-143.
71. Chikhi, R., and Medvedev, P. (2014) Informed and automated k-mer size selection for genome assembly. *Bioinformatics* **30**: 31-37.

72. Chisholm, S.W., Olson, R.J., Zettler, E.R., Goericke, R., Waterbury, J.B., and Welschmeyer, N.A. (1988) A novel free-living prochlorophyte abundant in the oceanic euphotic zone. *Nature* **334**: 340-343.
73. Choi, E.J., Kwon, H.C., Ham, J., and Yang, H.O. (2009) 6-Hydroxymethyl-1-phenazine-carboxamide and 1,6-phenazinedimethanol from a marine bacterium, *Brevibacterium* sp. KMD 003, associated with marine purple vase sponge. *J Antibiot (Tokyo)* **62**: 621-624.
74. Chuvochina, M., Rinke, C., Parks, D.H., Rappe, M.S., Tyson, G.W., Yilmaz, P. et al. (2019) The importance of designating type material for uncultured taxa. *Syst Appl Microbiol* **42**: 15-21.
75. Chuyen, H.V., and Eun, J.B. (2017) Marine carotenoids: Bioactivities and potential benefits to human health. *Crit Rev Food Sci Nutr* **57**: 2600-2610.
76. Cichewicz, R.H., Valeriote, F.A., and Crews, P. (2004) Psymberin, a potent sponge-derived cytotoxin from *Psammocinia* distantly related to the pederin family. *Organic letters* **6**: 1951-1954.
77. Coil, D., Jospin, G., and Darling, A.E. (2015) A5-miseq: an updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31**: 587-589.
78. Cooper, R., Bush, K., Principe, P.A., Trejo, W.H., Wells, J.S., and Sykes, R.B. (1983) Two new monobactam antibiotics produced by a *Flexibacter* sp. I. Taxonomy, fermentation, isolation and biological properties. *J Antibiot (Tokyo)* **36**: 1252-1257.
79. Costa, T.R., Felisberto-Rodrigues, C., Meir, A., Prevost, M.S., Redzej, A., Trokter, M., and Waksman, G. (2015) Secretion systems in Gram-negative bacteria: structural and mechanistic insights. *Nat Rev Microbiol* **13**: 343-359.
80. Crits-Christoph, A., Diamond, S., Butterfield, C.N., Thomas, B.C., and Banfield, J.F. (2018) Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature* **558**: 440-444.
81. Cross, K.L., Campbell, J.H., Balachandran, M., Campbell, A.G., Cooper, C.J., Griffen, A. et al. (2019) Targeted isolation and cultivation of uncultivated bacteria by reverse genomics. *Nat Biotechnol* **37**: 1314-1321.
82. Cárdenas, P., and Rapp, H.T. (2013) Disrupted spiculogenesis in deep-water Geodiidae (Porifera, Demospongiae) growing in shallow waters. *Invertebr Biol* **132**: 173-194.
83. Cárdenas, P., Rapp, H.T., Klitgaard, A.B., Best, M., Thollesson, M., and Tendal, O.S. (2013) Taxonomy, biogeography and DNA barcodes of *Geodia* species (Porifera, Demospongiae, Tetractinellida) in the Atlantic boreo-arctic region. *Zool J Linn Soc* **169**: 251-311.
84. Dash, S., Jin, C., Lee, O.O., Xu, Y., and Qian, P.Y. (2009) Antibacterial and antilarval-settlement potential and metabolite profiles of novel sponge-associated marine bacteria. *J Ind Microbiol Biotechnol* **36**: 1047-1056.
85. De Bary, A. (1879) The phenomenon of symbiosis. *Karl J Trubner, Strasbourg, Germany*.

86. de Goeij, J.M., van Oevelen, D., Vermeij, M.J., Osinga, R., Middelburg, J.J., de Goeij, A.F., and Admiraal, W. (2013) Surviving in a marine desert: the sponge loop retains resources within coral reefs. *Science* **342**: 108-110.
87. Del Bem, L.E., and Vincentz, M.G. (2010) Evolution of xyloglucan-related genes in green plants. *BMC Evol Biol* **10**: 341.
88. DeLong, E.F. (1992) Archaea in coastal marine environments. *Proc Natl Acad Sci U S A* **89**: 5685-5689.
89. DeLong, E.F. (2009) The microbial ocean from genomes to biomes. *Nature* **459**: 200-206.
90. DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924-934.
91. Desbois, A.P. (2014) How might we increase success in marine-based drug discovery? *Expert Opin Drug Discov* **9**: 985-990.
92. Desriac, F., Defer, D., Bourgougnon, N., Brillet, B., Le Chevalier, P., and Fleury, Y. (2010) Bacteriocin as weapons in the marine animal-associated bacteria warfare: inventory and potential applications as an aquaculture probiotic. *Mar Drugs* **8**: 1153-1177.
93. Desriac, F., Jegou, C., Balnois, E., Brillet, B., Le Chevalier, P., and Fleury, Y. (2013) Antimicrobial peptides from marine proteobacteria. *Mar Drugs* **11**: 3632-3660.
94. Diez-Vives, C., Moitinho-Silva, L., Nielsen, S., Reynolds, D., and Thomas, T. (2017) Expression of eukaryotic-like protein in the microbiome of sponges. *Mol Ecol* **26**: 1432-1451.
95. Doroghazi, J.R., and Metcalf, W.W. (2013) Comparative genomics of actinomycetes with a focus on natural product biosynthetic genes. *BMC Genomics* **14**: 611.
96. Durazzi, F., Sala, C., Castellani, G., Manfreda, G., Remondini, D., and De Cesare, A. (2021) Comparison between 16S rRNA and shotgun sequencing data for the taxonomic characterization of the gut microbiota. *Sci Rep* **11**: 1-10.
97. Dutilh, B.E., Backus, L., Edwards, R.A., Wels, M., Bayjanov, J.R., and van Hijum, S.A. (2013) Explaining microbial phenotypes on a genomic scale: GWAS for microbes. *Brief Funct Genomics* **12**: 366-380.
98. Dyshlovoy, S.A., and Honecker, F. (2019) Marine Compounds and Cancer: The First Two Decades of XXI Century. *Mar Drugs* **18**.
99. Easson, C.G., and Thacker, R.W. (2014) Phylogenetic signal in the community structure of host-specific microbiomes of tropical marine sponges. *Front Microbiol* **5**: 532.
100. Eddy, S.R. (2018) HMMER User's Guide: Biological sequence analysis using profile hidden Markov models.
101. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.

102. Ehrlich, H., Krautter, M., Hanke, T., Simon, P., Knieb, C., Heinemann, S., and Worch, H. (2007) First evidence of the presence of chitin in skeletons of marine sponges. Part II. Glass sponges (Hexactinellida: Porifera). *J Exp Zool B Mol Dev Evol* **308**: 473-483.
103. El-Gebali, S., Mistry, J., Bateman, A., Eddy, S.R., Luciani, A., Potter, S.C. et al. (2019) The Pfam protein families database in 2019. *Nucleic Acids Res* **47**: D427-D432.
104. Engelberts, J.P., Robbins, S.J., de Goeij, J.M., Aranda, M., Bell, S.C., and Webster, N.S. (2020) Characterization of a sponge microbiome using an integrative genome-centric approach. *ISME J* **14**: 1100-1110.
105. Eren, A.M., Esen, O.C., Quince, C., Vineis, J.H., Morrison, H.G., Sogin, M.L., and Delmont, T.O. (2015) Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* **3**: e1319.
106. Esteves, A.I., Amer, N., Nguyen, M., and Thomas, T. (2016) Sample Processing Impacts the Viability and Cultivability of the Sponge Microbiome. *Front Microbiol* **7**: 499.
107. Evans, J.R., Napier, E.J., and Fletton, R.A. (1978) G1499-2, a new quinoline compound isolated from the fermentation broth of *Cytophaga johnsonii*. *J Antibiot (Tokyo)* **31**: 952-958.
108. Fan, L., Reynolds, D., Liu, M., Stark, M., Kjelleberg, S., Webster, N.S., and Thomas, T. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci U S A* **109**: E1878-1887.
109. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**: 783-791.
110. Feng, G., Sun, W., Zhang, F., Orlic, S., and Li, Z. (2018) Functional Transcripts Indicate Phylogenetically Diverse Active Ammonia-Scavenging Microbiota in Sympatric Sponges. *Mar Biotechnol (NY)* **20**: 131-143.
111. Fernandez-Gomez, B., Richter, M., Schuler, M., Pinhassi, J., Acinas, S.G., Gonzalez, J.M., and Pedros-Alio, C. (2013) Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J* **7**: 1026-1037.
112. Field, K., Gordon, D., Wright, T., Rappé, M., Urback, E., Vergin, K., and Giovannoni, S. (1997) Diversity and depth-specific distribution of SAR11 cluster rRNA genes from marine planktonic bacteria. *Appl Environ Microbiol* **63**: 63-70.
113. Finn, R.D., Clements, J., Arndt, W., Miller, B.L., Wheeler, T.J., Schreiber, F. et al. (2015) HMMER web server: 2015 update. *Nucleic Acids Res* **43**: W30-38.
114. Fiore, C.L., Labrie, M., Jarett, J.K., and Lesser, M.P. (2015) Transcriptional activity of the giant barrel sponge, *Xestospongia muta* Holobiont: molecular evidence for metabolic interchange. *Front Microbiol* **6**: 364.
115. Fisch, K.M., Gurgui, C., Heycke, N., van der Sar, S.A., Anderson, S.A., Webb, V.L. et al. (2009) Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. *Nat Chem Biol* **5**: 494-501.
116. Florez, L.V., Biedermann, P.H., Engl, T., and Kaltenpoth, M. (2015) Defensive symbioses of animals with prokaryotic and eukaryotic microorganisms. *Nat Prod Rep* **32**: 904-936.

117. Freeman, M.F., Vagstad, A.L., and Piel, J. (2016) Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated 'Entotheonella' bacteria. *Curr Opin Chem Biol* **31**: 8-14.
118. Fuhrman, J.A., McCallum, K., and Davis, A.A. (1992) Novel major archaebacterial group from marine plankton. *Nature* **356**: 148-149.
119. Fujita, T., Hatanaka, H., Hayashi, K., Shigematsu, N., Takase, S., Okamoto, M. et al. (1994) FR901451, a novel inhibitor of human leukocyte elastase from *Flexibacter* sp. I. Producing organism, fermentation, isolation, physico-chemical and biological properties. *J Antibiot (Tokyo)* **47**: 1359-1364.
120. Funabashi, Y., Tsubotani, S., Koyama, K., Katayama, N., and Harada, S. (1993) A New Anti-MRSA Dipeptide, TAN-1057-A. *Tetrahedron* **49**: 13-28.
121. Galasso, C., Corinaldesi, C., and Sansone, C. (2017) Carotenoids from Marine Organisms: Biological Functions and Industrial Applications. *Antioxidants (Basel)* **6**.
122. Gallucci, M.N., Oliva, M., Casero, C., Dambolena, J., Luna, A., Zygadlo, J., and Demo, M. (2009) Antimicrobial combined action of terpenes against the food-borne microorganisms *Escherichia coli*, *Staphylococcus aureus* and *Bacillus cereus*. *Flavour Frag J* **24**: 348-354.
123. Gavriilidou, A., Mackenzie, T.A., Sanchez, P., Tormo, J.R., Ingham, C., Smidt, H., and Sipkema, D. (2021) Bioactivity Screening and Gene-Trait Matching across Marine Sponge-Associated Bacteria. *Mar Drugs* **19**.
124. Gavriilidou, A., Gutleben, J., Versluis, D., Forgiarini, F., van Passel, M.W.J., Ingham, C.J. et al. (2020) Comparative genomic analysis of *Flavobacteriaceae*: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics* **21**: 569.
125. Gillor, O., Etzion, A., and Riley, M.A. (2008) The dual role of bacteriocins as anti- and probiotics. *Appl Microbiol Biotechnol* **81**: 591-606.
126. Giovannoni, S.J., Britschgi, T.B., Moyer, C.L., and Field, K.G. (1990) Genetic diversity in Sargasso Sea bacterioplankton. *Nature* **345**: 60-63.
127. Glew, M.D., Veith, P.D., Chen, D., Gorasia, D.G., Peng, B., and Reynolds, E.C. (2017) PorV is an Outer Membrane Shuttle Protein for the Type IX Secretion System. *Sci Rep* **7**: 8790.
128. Glew, M.D., Veith, P.D., Peng, B., Chen, Y.Y., Gorasia, D.G., Yang, Q. et al. (2012) PG0026 is the C-terminal signal peptidase of a novel secretion system of *Porphyromonas gingivalis*. *J Biol Chem* **287**: 24605-24617.
129. 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. *Biol Bull* **227**: 78-88.
130. Goethe, R., Basler, T., Meissner, T., Goethe, E., Spröer, C., Swiderski, J. et al. (2020) Complete genome sequence and manual reannotation of *Mycobacterium* avium subsp. *paratuberculosis* strain DSM 44135. *Microbiol Resour Announc* **9**: e00711-00720.

131. Gomes, N.G., Dasari, R., Chandra, S., Kiss, R., and Kornienko, A. (2016) Marine Invertebrate Metabolites with Anticancer Activities: Solutions to the “Supply Problem”. *Mar Drugs* **14**.
132. Gorski, L., Godchaux, W.I., and Leadbetter, E.R. (1993) Structural specificity of sugars that inhibit gliding motility of *Cytophaga johnsonae*. *Arch Microbiol* **160**: 121-125.
133. Graca, 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**: e78992.
134. Graham, E.D., and Tully, B.J. (2021) Marine *Dadabacteria* exhibit genome streamlining and phototrophy-driven niche partitioning. *ISME J* **15**: 1248-1256.
135. Grondin, J.M., Tamura, K., Dejean, G., Abbott, D.W., and Brumer, H. (2017) Polysaccharide Utilization Loci: Fueling Microbial Communities. *J Bacteriol* **199**: e00860-00816.
136. Gu, Z., Eils, R., and Schlesner, M. (2016) Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847-2849.
137. Guerrero-Garzón, J.F., Zehl, M., Schneider, O., Rückert, C., Busche, T., Kalinowski, J. et al. (2020) *Streptomyces* spp. From the Marine Sponge *Antho dichotoma*: Analyses of Secondary Metabolite Biosynthesis Gene Clusters and Some of Their Products. *Front Microbiol* **11**.
138. Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072-1075.
139. 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.
140. Gutleben, J., Loureiro, C., Ramirez Romero, L.A., Shetty, S., Wijffels, R.H., Smidt, H., and Sipkema, D. (2020) Cultivation of Bacteria From *Aplysina aerophoba*: Effects of Oxygen and Nutrient Gradients. *Front Microbiol* **11**: 175.
141. Haber, M., Burgsdorf, I., Handley, K.M., Rubin-Blum, M., and Steindler, L. (2020) Genomic Insights Into the Lifestyles of *Thaumarchaeota* Inside Sponges. *Front Microbiol* **11**: 622824.
142. Hahnke, R.L., Meier-Kolthoff, J.P., Garcia-Lopez, M., Mukherjee, S., Huntemann, M., Ivanova, N.N. et al. (2016) Genome-Based Taxonomic Classification of *Bacteroidetes*. *Front Microbiol* **7**: 2003.
143. Hall, T. (2011) BioEdit: An important software for molecular biology. *GERF Bulletin of Biosciences* **2**: 60-61.
144. Hall, T.A. (1999) BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. In *Nucleic Acids Symposium Series*, pp. 95-98.
145. Handelsman, J. (2004) Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669-685.

146. 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**: R245-R249.
147. Harms, A., Brodersen, D.E., Mitarai, N., and Gerdes, K. (2018a) Toxins, Targets, and Triggers: An Overview of Toxin-Antitoxin Biology. *Mol Cell* **70**: 768-784.
148. Harms, H., Klockner, A., Schror, J., Josten, M., Kehraus, S., Crusemann, M. et al. (2018b) Antimicrobial Dialkylresorcins from Marine-Derived Microorganisms: Insights into Their Mode of Action and Putative Ecological Relevance. *Planta Med* **84**: 1363-1371.
149. Heidelberg, K.B., Gilbert, J.A., and Joint, I. (2010) Marine genomics: at the interface of marine microbial ecology and biodiscovery. *Microb Biotechnol* **3**: 531-543.
150. Helf, M.J., Jud, A., and Piel, J. (2017) Enzyme from an Uncultivated Sponge Bacterium Catalyzes S-Methylation in a Ribosomal Peptide. *ChemBioChem* **18**: 444-450.
151. Helfrich, E.J., and Piel, J. (2016) Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat Prod Rep* **33**: 231-316.
152. Hentschel, U. (2021) Harnessing the power of host–microbe symbioses to address grand challenges. *Nat Rev Microbiol*: 1-2.
153. Hentschel, U., Usher, K.M., and Taylor, M.W. (2006) Marine sponges as microbial fermenters. *FEMS Microbiol Ecol* **55**: 167-177.
154. Hentschel, U., Piel, J., Degnan, S.M., and Taylor, M.W. (2012) Genomic insights into the marine sponge microbiome. *Nat Rev Microbiol* **10**: 641-654.
155. Hentschel, U., Hopke, J., Horn, M., Friedrich, A.B., Wagner, M., Hacker, J., and Moore, B.S. (2002) Molecular Evidence for a Uniform Microbial Community in Sponges from Different Oceans. *Appl Environ Microbiol* **68**: 4431-4440.
156. Hentschel, U., Fieseler, L., Wehrl, M., Gernert, C., Steinert, M., Hacker, J., and Horn, M. (2003) Microbial diversity of marine sponges. In *Sponges (Porifera)*. Müller, W.E.G. (ed). Berlin, Germany: Springer Berlin Heidelberg.
157. Hertweck, C. (2009) Hidden biosynthetic treasures brought to light. *Nat Chem Biol* **5**: 450-452.
158. Hida, T., Tsubotani, S., Katayama, N., Okazaki, H., and Harada, S. (1985) Formadicins, new monocyclic beta-lactam antibiotics of bacterial origin. II. Isolation, characterization and structures. *J Antibiot (Tokyo)* **38**: 1128-1140.
159. Hill, M.S., and Sacristán-Soriano, O. (2017) Molecular and functional ecology of sponges and their microbial symbionts. *Climate change, ocean acidification and sponges*: 105-142.
160. Hoffmann, F., Rapp, H.T., and Reitner, J. (2006) Monitoring microbial community composition by fluorescence in situ hybridization during cultivation of the marine cold-water sponge *Geodia barretti*. *Mar Biotechnol (NY)* **8**: 373-379.
161. Hoiczyk, E. (2000) Gliding motility in cyanobacteria: observations and possible explanations. *Arch Microbiol* **174**: 11-17.

162. Hooper, J.N.A., Worheide, G., Hajdu, E., Erpenbeck, D., NJ, D.E.V., and Klautau, M. (2021) Zootaxa 20 years: Phylum Porifera. *Zootaxa* **4979**: 3856.
163. Horn, H., Slaby, B.M., Jahn, M.T., Bayer, K., Moitinho-Silva, L., Forster, F. et al. (2016) An Enrichment of CRISPR and Other Defense-Related Features in Marine Sponge-Associated Microbial Metagenomes. *Front Microbiol* **7**: 1751.
164. Hu, Y., Chen, J., Hu, G., Yu, J., Zhu, X., Lin, Y. et al. (2015) Statistical research on the bioactivity of new marine natural products discovered during the 28 years from 1985 to 2012. *Mar Drugs* **13**: 202-221.
165. Hudson, J., Kumar, V., and Egan, S. (2019) Comparative genome analysis provides novel insight into the interaction of *Aquimarina* sp. AD1, BL5 and AD10 with their macroalgal host. *Mar Genomics* **46**: 8-15.
166. Hug, L.A., Thomas, B.C., Sharon, I., Brown, C.T., Sharma, R., Hettich, R.L. et al. (2016a) Critical biogeochemical functions in the subsurface are associated with bacteria from new phyla and little studied lineages. *Environ Microbiol* **18**: 159-173.
167. Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J. et al. (2016b) A new view of the tree of life. *Nat Microbiol* **1**: 16048.
168. Hugenholtz, P., Chuvochina, M., Oren, A., Parks, D.H., and Soo, R.M. (2021) Prokaryotic taxonomy and nomenclature in the age of big sequence data. *ISME J* **15**: 1879-1892.
169. Hughes, C.C., and Fenical, W. (2010) Antibacterials from the sea. *Chemistry* **16**: 12512-12525.
170. Hunnicutt, D.W., Kempf, M.J., and McBride, M.J. (2002) Mutations in *Flavobacterium johnsoniae* gldF and gldG disrupt gliding motility and interfere with membrane localization of GldA. *J Bacteriol* **184**: 2370-2378.
171. Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., and Hauser, L.J. (2010) Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.
172. Ibrahim, A.H., Attia, E.Z., Hajjar, D., Anany, M.A., Desoukey, S.Y., Fouad, M.A. et al. (2018) New Cytotoxic Cyclic Peptide from the Marine Sponge-Associated *Nocardiopsis* sp. UR67. *Mar Drugs* **16**.
173. Ilan, M., and Abelson, A. (1995) The life of a sponge in a sandy lagoon. *Biol Bull* **189**: 363-369.
174. Imai, S., Fujioka, K., Furihata, K., Fudo, R., Yamanaka, S., and Seto, H. (1993) Studies on cell growth stimulating substances of Low-molecular-weight. Part 3. Resorcinin, a mammalian-cell growth-stimulating substance produced by *Cytophaga johnsonae*. *J Antibiot (Tokyo)* **46**: 1319-1322.
175. Indraningrat, A.A., Smidt, H., and Sipkema, D. (2016) Bioprospecting Sponge-Associated Microbes for Antimicrobial Compounds. *Mar Drugs* **14**.
176. Jaspars, M., and Challis, G. (2014) A talented genus. *Nature* **506**: 38-39.

177. Jaspars, M., De Pascale, D., Andersen, J.H., Reyes, F., Crawford, A.D., and Ianora, A. (2016) The marine biodiscovery pipeline and ocean medicines of tomorrow. *J Mar Biol Assoc U K* **96**: 151-158.
178. Johnston, J.J., Shrivastava, A., and McBride, M.J. (2018) Untangling *Flavobacterium johnsoniae* Gliding Motility and Protein Secretion. *J Bacteriol* **200**: e00362-00317.
179. Jones, P.M., and George, A.M. (2004) The ABC transporter structure and mechanism: perspectives on recent research. *Cell Mol Life Sci* **61**: 682-699.
180. Jooste, P.G., and Hugo, C.J. (1999) The taxonomy, ecology and cultivation of bacterial genera belonging to the family *Flavobacteriaceae*. *Int J Food Microbiol* **53**: 81-94.
181. Jung, D., Seo, E.Y., Epstein, S.S., Joung, Y., Han, J., Parfenova, V.V. et al. (2014) Application of a new cultivation technology, I-tip, for studying microbial diversity in freshwater sponges of Lake Baikal, Russia. *FEMS Microbiol Ecol* **90**: 417-423.
182. Kalinovskaya, N.I., Romanenko, L.A., Irisawa, T., Ermakova, S.P., and Kalinovsky, A.I. (2011) Marine isolate *Citricoccus* sp. KMM 3890 as a source of a cyclic siderophore nocardamine with antitumor activity. *Microbiol Res* **166**: 654-661.
183. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**: 587-589.
184. Kamiyama, T., Umino, T., Satoh, T., Sawairi, S., Shirane, M., Ohshima, S., and Yokose, K. (1995) Sulfobacins A and B, novel von Willebrand factor receptor antagonists. I. Production, isolation, characterization and biological activities. *J Antibiot (Tokyo)* **48**: 924-928.
185. Kamke, J., Sczyrba, A., Ivanova, N., Schwientek, P., Rinke, C., Mavromatis, K. et al. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. *ISME J* **7**: 2287-2300.
186. 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**: e87353.
187. Kanafani, Z.A., and Corey, G.R. (2007) Daptomycin: a rapidly bactericidal lipopeptide for the treatment of Gram-positive infections. *Expert Rev Anti Infect Ther* **5**: 177-184.
188. Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**: 726-731.
189. Kang, D.D., Froula, J., Egan, R., and Wang, Z. (2015) MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* **3**: e1165.
190. Kang, D.D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., and Wang, Z. (2019) MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**: e7359.

191. Kappelmann, L., Kruger, K., Hehemann, J.H., Harder, J., Markert, S., Unfried, F. et al. (2019) Polysaccharide utilization loci of North Sea *Flavobacteriia* as basis for using SusC/D-protein expression for predicting major phytoplankton glycans. *ISME J* **13**: 76-91.
192. Karanam, G., Arumugam, M.K., and Sirpu Natesh, N. (2019) Anticancer Effect of Marine Sponge-Associated *Bacillus pumilus* AMK1 Derived Dipeptide Cyclo (-Pro-Tyr) in Human Liver Cancer Cell Line Through Apoptosis and G2/M Phase Arrest. *Int J Pept Res Ther* **26**: 445-457.
193. Karimi, E., Ramos, M., Goncalves, J.M.S., Xavier, J.R., Reis, M.P., and Costa, R. (2017) Comparative Metagenomics Reveals the Distinctive Adaptive Features of the *Spongia officinalis* Endosymbiotic Consortium. *Front Microbiol* **8**: 2499.
194. Karimi, E., Keller-Costa, T., Slaby, B.M., Cox, C.J., da Rocha, U.N., Hentschel, U., and Costa, R. (2019) Genomic blueprints of sponge-prokaryote symbiosis are shared by low abundant and cultivatable *Alphaproteobacteria*. *Sci Rep* **9**: 1999.
195. Katayama, N., Fukusumi, S., Funabashi, Y., Iwahi, T., and Ono, H. (1993) TAN-1057 A-D, new antibiotics with potent antibacterial activity against methicillin-resistant *Staphylococcus aureus*. Taxonomy, fermentation and biological activity. *J Antibiot (Tokyo)* **46**: 606-613.
196. Kato, T., Hinoo, H., Shoji, J., Matsumoto, K., Tanimoto, T., Hattori, T. et al. (1987) PB-5266 A, B and C, new monobactams: I. Taxonomy, fermentation and isolation. *J Antibiot (Tokyo)* **55**: 135-138.
197. Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.
198. Kearney, S.E., Zahoranszky-Kohalmi, G., Brimacombe, K.R., Henderson, M.J., Lynch, C., Zhao, T. et al. (2018) Canvass: A Crowd-Sourced, Natural-Product Screening Library for Exploring Biological Space. *ACS Cent Sci* **4**: 1727-1741.
199. Keren, R., Lavy, A., Mayzel, B., and Ilan, M. (2015) Culturable associated-bacteria of the sponge *Theonella swinhoei* show tolerance to high arsenic concentrations. *Front Microbiol* **6**: 154.
200. 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.
201. Khalifa, S.A.M., Elias, N., Farag, M.A., Chen, L., Saeed, A., Hegazy, M.F. et al. (2019) Marine Natural Products: A Source of Novel Anticancer Drugs. *Mar Drugs* **17**.
202. Kharade, S.S., and McBride, M.J. (2015) *Flavobacterium johnsoniae* PorV is required for secretion of a subset of proteins targeted to the type IX secretion system. *J Bacteriol* **197**: 147-158.
203. Khayatan, B., Meeks, J.C., and Risser, D.D. (2015) Evidence that a modified type IV pilus-like system powers gliding motility and polysaccharide secretion in filamentous cyanobacteria. *Mol Microbiol* **98**: 1021-1036.

204. Kimes, N.E., Johnson, W.R., Torralba, M., Nelson, K.E., Weil, E., and Morris, P.J. (2013) The *Montastraea faveolata* microbiome: ecological and temporal influences on a Caribbean reef-building coral in decline. *Environ Microbiol* **15**: 2082-2094.
205. Kimura, M., Wakimoto, T., Egami, Y., Tan, K.C., Ise, Y., and Abe, I. (2012) Calyxamides A and B, cytotoxic cyclic peptides from the marine sponge *Discodermia calyx*. *J Nat Prod* **75**: 290-294.
206. Kiran, G.S., Priyadharsini, S., Sajayan, A., Ravindran, A., and Selvin, J. (2018) An antibiotic agent pyrrolo[1,2-a]pyrazine-1,4-dione,hexahydro isolated from a marine bacteria *Bacillus tequilensis* MSI45 effectively controls multi-drug resistant *Staphylococcus aureus*. *RSC Advances* **8**: 17837-17846.
207. Kirchman, D.L. (2002) The ecology of *Cytophaga-Flavobacteria* in aquatic environments. *FEMS Microbiol Ecol* **39**: 91-100.
208. Kirchman, D.L., and Gasol, J.M. (2018) Introduction: The Evolution of Microbial Ecology of the Ocean In *Microbial Ecology of the Oceans*. Kirchman, D.L., and Gasol, J.M. (eds). NJ, USA: John Wiley & Sons, Hoboken, p. 46.
209. Knerr, P.J., and van der Donk, W.A. (2012) Discovery, biosynthesis, and engineering of lantipeptides. *Annu Rev Biochem* **81**: 479-505.
210. Knobloch, S., Jóhannsson, R., and Marteinsson, V.Þ. (2020) Genome analysis of sponge symbiont '*Candidatus* Halichondribacter symbioticus' shows genomic adaptation to a host-dependent lifestyle. *Environ Microbiol* **22**: 483-498.
211. Koedooder, C., Guéneuguès, A., Van Geersdaële, R., Vergé, V., Bouget, F.-Y., Labreuche, Y. et al. (2018) The Role of the Glyoxylate Shunt in the Acclimation to Iron Limitation in Marine Heterotrophic Bacteria. *Front Mar Sci* **5**.
212. Kokoulin, M.S., Kuzmich, A.S., Romanenko, L.A., Chikalovets, I.V., and Chernikov, O.V. (2020) Structure and in vitro Bioactivity against Cancer Cells of the Capsular Polysaccharide from the Marine Bacterium *Psychrobacter marincola*. *Mar Drugs* **18**.
213. Kolton, M., Frenkel, O., Elad, Y., and Cytryn, E. (2014) Potential role of Flavobacterial gliding-motility and type IX secretion system complex in root colonization and plant defense. *Mol Plant Microbe Interact* **27**: 1005-1013.
214. Konstantinidis, K.T., Rossello-Mora, R., and Amann, R. (2017) Uncultivated microbes in need of their own taxonomy. *ISME J* **11**: 2399-2406.
215. Konstantinou, D., Mavrogonatou, E., Zervou, S.K., Giannogonas, P., and Gkelis, S. (2020) Bioprospecting Sponge-Associated Marine Cyanobacteria to Produce Bioactive Compounds. *Toxins (Basel)* **12**.
216. Kovacs-Simon, A., Titball, R.W., and Michell, S.L. (2011) Lipoproteins of bacterial pathogens. *Infect Immun* **79**: 548-561.
217. Krubasik, P., and Sandmann, G. (2000) A carotenogenic gene cluster from *Brevibacterium linens* with novel lycopene cyclase genes involved in the synthesis of aromatic carotenoids. *Mol Gen Genet* **263**: 423-432.

218. Kulkarni, S.S., Johnston, J.J., Zhu, Y., Hying, Z.T., and McBride, M.J. (2019) The Carboxy-Terminal Region of *Flavobacterium johnsoniae* SprB Facilitates Its Secretion by the Type IX Secretion System and Propulsion by the Gliding Motility Machinery. *J Bacteriol* **201**: e00218-00219.
219. Lackner, G., Peters, E.E., Helfrich, E.J., 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**: E347-E356.
220. Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.-H., Rognes, T., and Ussery, D.W. (2007) RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**: 3100-3108.
221. Lane, D.J. (1991) *16S/23S rRNA Sequencing*. New York: John Wiley and Sons.
222. Lang, J.M., Darling, A.E., and Eisen, J.A. (2013) Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. *PLoS One* **8**: e62510.
223. Langmead, B., and Salzberg, S.L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357-359.
224. Lapebie, P., Lombard, V., Drula, E., Terrapon, N., and Henrissat, B. (2019) *Bacteroidetes* use thousands of enzyme combinations to break down glycans. *Nat Commun* **10**: 2043.
225. Lasica, A.M., Ksiazek, M., Madej, M., and Potempa, J. (2017) The Type IX Secretion System (T9SS): Highlights and Recent Insights into Its Structure and Function. *Front Cell Infect Microbiol* **7**: 215.
226. Lavy, A., Keren, R., Yu, K., Thomas, B.C., Alvarez-Cohen, L., Banfield, J.F., and Ilan, M. (2018) A novel *Chromatiales* bacterium is a potential sulfide oxidizer in multiple orders of marine sponges. *Environ Microbiol* **20**: 800-814.
227. Lee, L.H., Goh, B.H., and Chan, K.G. (2020) Editorial: Actinobacteria: Prolific Producers of Bioactive Metabolites. *Front Microbiol* **11**: 1612.
228. Lee, M.D. (2019) GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**: 4162-4164.
229. Lee, M.D., Walworth, N.G., McParland, E.L., Fu, F.X., Mincer, T.J., Levine, N.M. et al. (2017) The *Trichodesmium* consortium: conserved heterotrophic co-occurrence and genomic signatures of potential interactions. *ISME J* **11**: 1813-1824.
230. Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242-245.
231. Letunic, I., and Bork, P. (2021) Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* **49**: W293-W296.
232. Lewis, W.H., and Ettema, T.J.G. (2019) Culturing the uncultured. *Nat Biotechnol* **37**: 1278-1279.
233. Lewis, W.H., Tahon, G., Geesink, P., Sousa, D.Z., and Ettema, T.J.G. (2021) Innovations to culturing the uncultured microbial majority. *Nat Rev Microbiol* **19**: 225-240.

234. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N. et al. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
235. Lim, S.J., and Bordenstein, S.R. (2020) An introduction to phylosymbiosis. *Proc Biol Sci* **287**: 20192900.
236. Lindequist, U. (2016) Marine-Derived Pharmaceuticals - Challenges and Opportunities. *Biomol Ther (Seoul)* **24**: 561-571.
237. Liu, J., Xue, C.X., Sun, H., Zheng, Y., Meng, Z., and Zhang, X.H. (2019) Carbohydrate catabolic capability of a *Flavobacteriia* bacterium isolated from hadal water. *Syst Appl Microbiol* **42**: 263-274.
238. Liu, Y., Lai, Q., Dong, C., Sun, F., Wang, L., Li, G., and Shao, Z. (2013) Phylogenetic diversity of the *Bacillus pumilus* group and the marine ecotype revealed by multilocus sequence analysis. *PLoS One* **8**: e80097.
239. Liu, Y., Makarova, K.S., Huang, W.C., Wolf, Y.I., Nikolskaya, A.N., Zhang, X. et al. (2021) Expanded diversity of Asgard archaea and their relationships with eukaryotes. *Nature* **593**: 553-557.
240. Lok, C. (2015) Mining the microbial dark matter. *Nature* **522**: 270-273.
241. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490-495.
242. Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar et al. (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
243. Luo, C., Tsementzi, D., Kyrpides, N.C., and Konstantinidis, K.T. (2012) Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* **6**: 898-901.
244. Luo, S., Chen, X.A., Mao, X.M., and Li, Y.Q. (2018) Regulatory and biosynthetic effects of the bkd gene clusters on the production of daptomycin and its analogs A21978C1-3. *J Ind Microbiol Biotechnol* **45**: 271-279.
245. Lurgi, M., Thomas, T., Wemheuer, B., Webster, N.S., and Montoya, J.M. (2019) Modularity and predicted functions of the global sponge-microbiome network. *Nat Commun* **10**: 992.
246. Ma, Z., Wang, N., Hu, J., and Wang, S. (2012) Isolation and characterization of a new iturinic lipopeptide, mojavensin A produced by a marine-derived bacterium *Bacillus mojavensis* B0621A. *J Antibiot (Tokyo)* **65**: 317-322.
247. Makarova, K.S., Wolf, Y.I., and Koonin, E.V. (2013) Comparative genomics of defense systems in archaea and bacteria. *Nucleic Acids Res* **41**: 4360-4377.
248. Maldonado, M. (2007) Intergenerational transmission of symbiotic bacteria in oviparous and viviparous demosponges, with emphasis on intracytoplasmically-compartmented bacterial types. *J Mar Biol Assoc U K* **87**: 1701-1713.
249. Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. et al. (2013) The Genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM

3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79**: 6813-6822.

250. Martens, E.C., Koropatkin, N.M., Smith, T.J., and Gordon, J.I. (2009) Complex glycan catabolism by the human gut microbiota: the *Bacteroidetes* Sus-like paradigm. *J Biol Chem* **284**: 24673-24677.
251. Martin, J., da, S.S.T., Crespo, G., Palomo, S., Gonzalez, I., Tormo, J.R. et al. (2013) Kocurin, the true structure of PM181104, an anti-methicillin-resistant *Staphylococcus aureus* (MRSA) thiazolyl peptide from the marine-derived bacterium *Kocuria palustris*. *Mar Drugs* **11**: 387-398.
252. Martiny, A.C., Treseder, K., and Pusch, G. (2013) Phylogenetic conservatism of functional traits in microorganisms. *ISME J* **7**: 830-838.
253. Mayer, A.M., 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**: 255-265.
254. McBride, M.J. (2001) BACTERIAL GLIDING MOTILITY: Multiple Mechanisms for Cell Movement over Surfaces. *Annu Rev Microbiol* **55**: 49-75.
255. McBride, M.J. (2014) The Family *Flavobacteriaceae*. In *The Prokaryotes*. Rosenberg, E., DeLong, E.F., Lory, S., Stackebrandt, E., and Thompson, F. (eds). Berlin, Heidelberg: Springer, pp. 643-676.
256. McBride, M.J. (2019a) *Bacteroidetes* Gliding Motility and the Type IX Secretion System. *Microbiol Spectr* **7**: PSIB-0002-2018.
257. McBride, M.J. (2019b) *Bacteroidetes* Gliding Motility and the Type IX Secretion System. *Microbiol Spectr* **7**.
258. McBride, M.J., and Zhu, Y. (2013) Gliding motility and Por secretion system genes are widespread among members of the phylum *Bacteroidetes*. *J Bacteriol* **195**: 270-278.
259. McBride, M.J., and Nakane, D. (2015) *Flavobacterium* gliding motility and the type IX secretion system. *Curr Opin Microbiol* **28**: 72-77.
260. McBride, M.J., Xie, G., Martens, E.C., Lapidus, A., Henrissat, B., Rhodes, R.G. et al. (2009) Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* **75**: 6864-6875.
261. McKenna, V., Archibald, J.M., Beinart, R., Dawson, M.N., Hentschel, U., Keeling, P.J. et al. (2021) The Aquatic Symbiosis Genomics Project: probing the evolution of symbiosis across the tree of life. *Wellcome Open Research* **6**.
262. McMurdie, P.J., and Holmes, S. (2013) phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* **8**: e61217.
263. Medema, M.H., and Fischbach, M.A. (2015) Computational approaches to natural product discovery. *Nat Chem Biol* **11**: 639-648.

264. Medema, M.H., Kottmann, R., Yilmaz, P., Cummings, M., Biggins, J.B., Blin, K. et al. (2015) Minimum Information about a Biosynthetic Gene cluster. *Nat Chem Biol* **11**: 625-631.
265. Mehbub, M.F., Lei, J., Franco, C., and Zhang, W. (2014) Marine sponge derived natural products between 2001 and 2010: trends and opportunities for discovery of bioactives. *Mar Drugs* **12**: 4539-4577.
266. Miki, W., Otaki, N., Yokoyama, A., and Kusumi, T. (1996) Possible origin of zeaxanthin in the marine sponge, *Reniera japonica*. *Experientia* **52**: 93-96.
267. Miller, I., Lopera, J., Montgomery, K., Puglisi, M., Rose, W., and Kwan, J. (2016) Deconvolution of complete microbial genomes from shotgun metagenomes. *Planta Medica* **82**: SL4.
268. Milshteyn, A., Schneider, J.S., and Brady, S.F. (2014) Mining the metabiome: identifying novel natural products from microbial communities. *Chem Biol* **21**: 1211-1223.
269. Mincer, T.J., and Aicher, A.C. (2016) Methanol production by a broad phylogenetic array of marine phytoplankton. *PloS one* **11**: e0150820.
270. Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., and Lanfear, R. (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**: 1530-1534.
271. Miyata, M. (2008) Centipede and inchworm models to explain *Mycoplasma* gliding. *Trends Microbiol* **16**: 6-12.
272. Moitinho-Silva, L., Diez-Vives, C., Batani, G., Esteves, A.I., Jahn, M.T., and Thomas, T. (2017a) Integrated metabolism in sponge-microbe symbiosis revealed by genome-centered metatranscriptomics. *ISME J* **11**: 1651-1666.
273. Moitinho-Silva, L., Nielsen, S., Amir, A., Gonzalez, A., Ackermann, G.L., Cerrano, C. et al. (2017b) The sponge microbiome project. *Gigascience* **6**: 1-7.
274. Mojib, N., Philpott, R., Huang, J.P., Niederweis, M., and Bej, A.K. (2010) Antimycobacterial activity in vitro of pigments isolated from Antarctic bacteria. *Antonie Van Leeuwenhoek* **98**: 531-540.
275. Montalvo, N.F., and Hill, R.T. (2011) Sponge-associated bacteria are strictly maintained in two closely related but geographically distant sponge hosts. *Appl Environ Microbiol* **77**: 7207-7216.
276. Monteiro, M.C., de la Cruz, M., Cantizani, J., Moreno, C., Tormo, J.R., Mellado, E. et al. (2012) A new approach to drug discovery: high-throughput screening of microbial natural extracts against *Aspergillus fumigatus* using resazurin. *J Biomol Screen* **17**: 542-549.
277. Moore, B.S. (2006) Biosynthesis of marine natural products: macroorganisms (Part B). *Nat Prod Rep* **23**: 615-629.
278. Moran, N.A., and Wernegreen, J.J. (2000) Lifestyle evolution in symbiotic bacteria: insights from genomics. *Trends Ecol Evol* **15**: 321-326.

279. Moreno-Pino, M., Cristi, A., Gillooly, J.F., and Trefault, N. (2020) Characterizing the microbiomes of Antarctic sponges: a functional metagenomic approach. *Sci Rep* **10**: 645.
280. Mori, T., Cahn, J.K.B., Wilson, M.C., Meoded, R.A., Wiebach, V., Martinez, A.F.C. et al. (2018) Single-bacterial genomics validates rich and varied specialized metabolism of uncultivated *Entotheonella* sponge symbionts. *Proc Natl Acad Sci U S A* **115**: 1718-1723.
281. Morita, M., and Schmidt, E.W. (2018) Parallel lives of symbionts and hosts: chemical mutualism in marine animals. *Nat Prod Rep* **35**: 357-378.
282. Mu, D.-S., Ouyang, Y., Chen, G.-J., and Du, Z.-J. (2020) Strategies for culturing active/dormant marine microbes. *Mar Life Sci Technol* **3**: 121-131.
283. Mukherjee, S., Stamatis, D., Bertsch, J., Ovchinnikova, G., Katta, H.Y., Mojica, A. et al. (2019) Genomes OnLine database (GOLD) v.7: updates and new features. *Nucleic Acids Res* **47**: D649-D659.
284. Munn, C.B. (2020) Microbes in the Marine Environment. In *Marine Microbiology: Ecology & Applications*. Munn, C.B. (ed). Boca Raton: CRC Press, p. 436.
285. Munoz, R., Teeling, H., Amann, R., and Rossello-Mora, R. (2020) Ancestry and adaptive radiation of *Bacteroidetes* as assessed by comparative genomics. *Syst Appl Microbiol* **43**: 126065.
286. Munroe, S., Sandoval, K., Martens, D.E., Sipkema, D., and Pomponi, S.A. (2019) Genetic algorithm as an optimization tool for the development of sponge cell culture media. *In Vitro Cell Dev Biol Animal* **55**: 149-158.
287. Murray, A.E., Freudenstein, J., Gribaldo, S., Hatzenpichler, R., Hugenholtz, P., Kampfer, P. et al. (2020) Roadmap for naming uncultivated Archaea and Bacteria. *Nat Microbiol*.
288. Nakashima, Y., Egami, Y., Kimura, M., Wakimoto, T., and Abe, I. (2016) Metagenomic Analysis of the Sponge *Discodermia* Reveals the Production of the Cyanobacterial Natural Product Kasumigamide by 'Entotheonella'. *PLoS One* **11**: e0164468.
289. 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.
290. Nei, M., and Kumar, S. (2000) *Molecular evolution and phylogenetics*: Oxford university press.
291. Nemoto, T., Ojika, M., Takahata, Y., Andoh, T., and Sakagami, Y. (1998) Structures of topostins, DNA topoisomerase I inhibitors of bacterial origin. *Tetrahedron* **54**: 2683-2690.
292. Nett, M., and König, G.M. (2007) The chemistry of gliding bacteria. *Nat Prod Rep* **24**: 1245–1261.
293. Nguyen, M.T., Liu, M., and Thomas, T. (2014) Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. *Mol Ecol* **23**: 1635-1645.

294. Nichols, D. (2007) Cultivation gives context to the microbial ecologist. *FEMS Microbiol Ecol* **60**: 351-357.
295. Nielsen, C. (2019) Early animal evolution: a morphologist's view. *R Soc Open Sci* **6**: 190638.
296. Nurk, S., Meleshko, D., Korobeynikov, A., and Pevzner, P.A. (2017) metaSPAdes: a new versatile metagenomic assembler. *Genome research* **27**: 824-834.
297. Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D. et al. (2018) Vegan: Community Ecology Package. R package version 2.5-2. https://CRAN.R-project.org/package=vegan.
298. Olm, M.R., Brown, C.T., Brooks, B., and Banfield, J.F. (2017) dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* **11**: 2864-2868.
299. Orgogozo, V., Morizot, B., and Martin, A. (2015) The differential view of genotype-phenotype relationships. *Front Genet* **6**: 179.
300. Pande, S., and Kost, C. (2017) Bacterial Unculturability and the Formation of Intercellular Metabolic Networks. *Trends Microbiol* **25**: 349-361.
301. Paoli, L., Ruscheweyh, H.-J., Forneris, C.C., Kautsar, S., Clayssen, Q., Salazar, G. et al. (2021) Uncharted biosynthetic potential of the ocean microbiome. *bioRxiv*.
302. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., and Tyson, G.W. (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**: 1043-1055.
303. Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020) A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**: 1079-1086.
304. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A., and Hugenholtz, P. (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**: 996-1004.
305. Parks, D.H., Rinke, C., Chuvochina, M., Chaumeil, P.A., Woodcroft, B.J., Evans, P.N. et al. (2017) Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* **2**: 1533-1542.
306. Paterson, I., and Anderson, E.A. (2005) The renaissance of natural products as drug candidates. *Science* **310**: 451-453.
307. Paul, V.J., Puglisi, M.P., and Ritson-Williams, R. (2006) Marine chemical ecology. *Nat Prod Rep* **23**: 153-180.
308. Paul, V.J., Ritson-Williams, R., and Sharp, K. (2011) Marine chemical ecology in benthic environments. *Nat Prod Rep* **28**: 345-387.
309. Pawlik, J.R. (2011) The Chemical Ecology of Sponges on Caribbean Reefs: Natural Products Shape Natural Systems. *BioScience* **61**: 888-898.

310. Perdicaris, S., Vlachogianni, T., and Valavanidis, A. (2013) Bioactive Natural Substances from Marine Sponges: New Developments and Prospects for Future Pharmaceuticals. *Nat Prod Chem Res* **01**.
311. Pettit, G.R., Xu, J.-P., Chapuis, J.-C., Pettit, R.K., Tackett, L.P., Doubek, D.L. et al. (2004) Antineoplastic Agents. 520. Isolation and Structure of Irciniastatins A and B from the Indo-Pacific Marine Sponge *Ircinia ramosa*. *J Med Chem* **47**: 1149-1152.
312. Piel, J. (2009) Metabolites from symbiotic bacteria. *Nat Prod Rep* **26**: 338-362.
313. Piel, J., Hui, D., Wen, G., Butzke, D., Platzer, M., Fusetani, N., and Matsunaga, S. (2004) Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. *Proc Natl Acad Sci U S A* **101**: 16222–16227.
314. Pita, L., Fraune, S., and Hentschel, U. (2016) Emerging Sponge Models of Animal-Microbe Symbioses. *Front Microbiol* **7**: 2102.
315. Pita, L., Rix, L., Slaby, B.M., Franke, A., and Hentschel, U. (2018) The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome* **6**: 46.
316. Podell, S., Blanton, J.M., Neu, A., Agarwal, V., Biggs, J.S., Moore, B.S., and Allen, E.E. (2019) Pangenomic comparison of globally distributed Poribacteria associated with sponge hosts and marine particles. *ISME J* **13**: 468-481.
317. Pomponi, S.A. (1999) The bioprocess-technological potential of the sea. *J Biotechnol* **70**: 5-13.
318. Poppell, E., Weisz, J., Spicer, L., Massaro, A., Hill, A., and Hill, M. (2014) Sponge heterotrophic capacity and bacterial community structure in high- and low-microbial abundance sponges. *Mar Ecol* **35**: 414-424.
319. Poretsky, R., Rodriguez-R, L.M., Luo, C., Tsementzi, D., and Konstantinidis, K.T. (2014) Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PloS one* **9**: e93827.
320. Price, M.N., Dehal, P.S., and Arkin, A.P. (2010) FastTree 2--approximately maximum-likelihood trees for large alignments. *PLoS One* **5**: e9490.
321. Prosser, J.I. (2020) Putting science back into microbial ecology: a question of approach. *Philos Trans R Soc Lond B Biol Sci* **375**: 20190240.
322. Pruesse, E., Peplies, J., and Glockner, F.O. (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823-1829.
323. Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P. et al. (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.
324. Quinlan, A.R. (2014) BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics* **47**: 11.12.11–11.12.34.
325. R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
326. Radax, R., Rattei, T., Lanzen, A., Bayer, C., Rapp, H.T., Urich, T., and Schleper, C. (2012) Metatranscriptomics of the marine sponge *Geodia barretti*: tackling phylogeny and function of its microbial community. *Environ Microbiol* **14**: 1308-1324.

327. Raina, J.B., Fernandez, V., Lambert, B., Stocker, R., and Seymour, J.R. (2019) The role of microbial motility and chemotaxis in symbiosis. *Nat Rev Microbiol* **17**: 284-294.
328. Rappé, M.S., Connon, S.A., Vergin, K.L., and Giovannoni, S.J. (2002) Cultivation of the ubiquitous SAR11 marine bacterioplankton clade. *Nature* **418**: 630-633.
329. Redmond, A.K., and McLysaght, A. (2021) Evidence for sponges as sister to all other animals from partitioned phylogenomics with mixture models and recoding. *Nat Commun* **12**: 1783.
330. Reen, F.J., Romano, S., Dobson, A.D., and O'Gara, F. (2015) The Sound of Silence: Activating Silent Biosynthetic Gene Clusters in Marine Microorganisms. *Mar Drugs* **13**: 4754-4783.
331. Regoli, F., Cerrano, C., Chierici, E., Bompadre, S., and Bavestrello, G. (2000) Susceptibility to oxidative stress of the Mediterranean demosponge *Petrosia ficiformis*: role of endosymbionts and solar irradiance. *Mar Biol* **137**: 453-461.
332. Reichenbach, H., Kohl, W., Bottger-Vetter, A., and Achenbach, H. (1980) Flexirubin-Type Pigments in *Flavobacterium*. *Arch Microbiol* **126**: 291-293.
333. Reiswig, H.M. (1981) Partial Carbon and Energy Budgets of the Bacteriosponge *Verohgia fistularis* (Porifera: Demospongiae) in Barbados. *Mar Ecol* **2**: 273-293.
334. Reiter, S., Cahn, J.K., Wiebach, V., Ueoka, R., and Piel, J. (2020) Characterization of an Orphan Type III Polyketide Synthase Conserved in Uncultivated "Entotheonella" Sponge Symbionts. *ChemBioChem* **21**: 564.
335. Reynolds, D., and Thomas, T. (2016) Evolution and function of eukaryotic-like proteins from sponge symbionts. *Mol Ecol* **25**: 5242-5253.
336. Ribes, M., Jimenez, E., Yahel, G., Lopez-Sendino, P., Diez, B., Massana, R. et al. (2012) Functional convergence of microbes associated with temperate marine sponges. *Environ Microbiol* **14**: 1224-1239.
337. Riley, M.A., and Wertz, J.E. (2002) Bacteriocins: evolution, ecology, and application. *Annu Rev Microbiol* **56**: 117-137.
338. Rinke, C., Chuvochina, M., Mussig, A.J., Chaumeil, P.A., Davin, A.A., Waite, D.W. et al. (2021) A standardized archaeal taxonomy for the Genome Taxonomy Database. *Nat Microbiol* **6**: 946-959.
339. Rinke, C., Schwientek, P., Sczyrba, A., Ivanova, N.N., Anderson, I.J., Cheng, J.F. et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.
340. Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A., and Kim, D. (2015) Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet* **16**: 85-97.
341. Rix, L., de Goeij, J.M., van Oevelen, D., Struck, U., Al-Horani, F.A., Wild, C., and Naumann, M.S. (2018) Reef sponges facilitate the transfer of coral-derived organic matter to their associated fauna via the sponge loop. *Mar Ecol Prog Ser* **589**: 85-96.

342. Rix, L., Ribes, M., Coma, R., Jahn, M.T., de Goeij, J.M., van Oevelen, D. et al. (2020) Heterotrophy in the earliest gut: a single-cell view of heterotrophic carbon and nitrogen assimilation in sponge-microbe symbioses. *ISME J* **14**: 2554-2567.
343. Robbins, S.J., Song, W., Engelberts, J.P., Glasl, B., Slaby, B.M., Boyd, J. et al. (2021) A genomic view of the microbiome of coral reef demosponges. *ISME J*.
344. Roberts, E., Bowers, D., Meyer, H., Samuelsen, A., Rapp, H., and Cárdenas, P. (2021) Water masses constrain the distribution of deep-sea sponges in the North Atlantic Ocean and Nordic Seas. *Mar Ecol Prog Ser* **659**: 75-96.
345. Robinson, C.J., Bohannan, B.J., and Young, V.B. (2010) From structure to function: the ecology of host-associated microbial communities. *Microbiol Mol Biol Rev* **74**: 453-476.
346. Robinson, S.J., Tenney, K., Yee, D.F., Martinez, L., Media, J.E., Valeriote, F.A. et al. (2007) Probing the bioactive constituents from chemotypes of the sponge *Psammocinia* aff. *bulbosa. J Nat Prod* **70**: 1002-1009.
347. Robinson, S.L., Christenson, J.K., and Wackett, L.P. (2019) Biosynthesis and chemical diversity of β-lactone natural products. *Nat Prod Rep* **20**: 458-475.
348. Roch, M., Gagetti, P., Davis, J., Ceriana, P., Errecalde, L., Corso, A., and Rosato, A.E. (2017) Daptomycin Resistance in Clinical MRSA Strains Is Associated with a High Biological Fitness Cost. *Front Microbiol* **8**: 2303.
349. Rodrigues, A.M.S., Rohée, C., Fabre, T., Batailler, N., Sautel, F., Carletti, I. et al. (2017) Cytotoxic indole alkaloids from *Pseudovibrio denitrificans* BBCC725. *Tetrahedron Lett* **58**: 3172-3173.
350. Romanenko, L.A., Uchino, M., Frolova, G.M., and Mikhailov, V.V. (2007) *Marixanthomonas ophiurae* gen. nov., sp. nov., a marine bacterium of the family *Flavobacteriaceae* isolated from a deep-sea brittle star. *Int J Syst Evol Microbiol* **57**: 457-462.
351. Romano, G., Costantini, M., Sansone, C., Lauritano, C., Ruocco, N., and Ianora, A. (2017) Marine microorganisms as a promising and sustainable source of bioactive molecules. *Mar Environ Res* **128**: 58-69.
352. Romano, S. (2018) Ecology and Biotechnological Potential of Bacteria Belonging to the Genus Pseudovibrio. *Appl Environ Microbiol* **84**.
353. Rondon, M.R., August, P.R., Bettermann, A.D., Brady, S.F., Grossman, T.H., Liles, M.R. et al. (2000) Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl Environ Microbiol* **66**: 2541-2547.
354. Roume, H., Heintz-Buschart, A., Muller, E.E., and Wilmes, P. (2013) Sequential isolation of metabolites, RNA, DNA, and proteins from the same unique sample. *Methods Enzymol* **531**: 219-236.
355. RStudio Team (2020) RStudio: Integrated Development of R. RStudio, PBC, Boston, MA. http://www.rstudio.com/.
356. Rust, M., Helfrich, E.J.N., Freeman, M.F., Nanudorn, P., Field, C.M., Ruckert, C. et al. (2020) A multiproducer microbiome generates chemical diversity in the marine sponge *Mycale hentscheli*. *Proc Natl Acad Sci U S A* **117**: 9508-9518.

357. Rutledge, P.J., and Challis, G.L. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nat Rev Microbiol* **13**: 509-523.
358. Santos, J.D., Vitorino, I., Reyes, F., Vicente, F., and Lage, O.M. (2020a) From Ocean to Medicine: Pharmaceutical Applications of Metabolites from Marine Bacteria. *Antibiotics (Basel)* **9**.
359. Santos, J.D., Vitorino, I., de la Cruz, M., Díaz, C., Cautain, B., Annang, F. et al. (2020b) Diketopiperazines and other bioactive compounds from bacterial symbionts of marine sponges. *Antonie van Leeuwenhoek* **113**: 875-887.
360. Santos-Gandelman, J.F., Giambiagi-deMarval, M., Oelemann, W.M., and Laport, M.S. (2014) Biotechnological potential of sponge-associated bacteria. *Curr Pharm Biotechnol* **15**: 143-155.
361. Sato, K., Naito, M., Yukitake, H., Hirakawa, H., Shoji, M., McBride, M.J. et al. (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. *Proc Natl Acad Sci U S A* **107**: 276-281.
362. Schimak, M.P., Kleiner, M., Wetzel, S., Liebeke, M., Dubilier, N., and Fuchs, B.M. (2016) MiL-FISH: Multilabeled oligonucleotides for fluorescence in situ hybridization improve visualization of bacterial cells. *Appl Environ Microbiol* **82**: 62-70.
363. Schippers, K.J., Sipkema, D., Osinga, R., Smidt, H., Pomponi, S.A., Martens, D.E., and Wijffels, R.H. (2012) Cultivation of sponges, sponge cells and symbionts: achievements and future prospects. *Adv Mar Biol* **62**: 273-337.
364. Schmidt, E.W., Obraztsova, A.Y., Davidson, S.K., Faulkner, D.J., and Haygood, M.G. (2000) Identification of the antifungal peptide-containing symbiont of the marine sponge Theonella swinhoei as a novel δ-proteobacterium, " Candidatus Entotheonella palauensis". *Marine Biology* **136**: 969-977.
365. Schmitt, S., Tsai, P., Bell, J., Fromont, J., Ilan, M., Lindquist, N. et al. (2012) Assessing the complex sponge microbiota: core, variable and species-specific bacterial communities in marine sponges. *ISME J* **6**: 564-576.
366. Schneider, Y., Jenssen, M., Isaksson, J., Hansen, K.O., Andersen, J.H., and Hansen, E.H. (2020) Bioactivity of Serratiochelin A, a Siderophore Isolated from a Co-Culture of Serratia sp. and Shewanella sp. *Microorganisms* **8**.
367. Schorn, M.A., Jordan, P.A., Podell, S., Blanton, J.M., Agarwal, V., Biggs, J.S. et al. (2019) Comparative Genomics of Cyanobacterial Symbionts Reveals Distinct, Specialized Metabolism in Tropical Dysideidae Sponges. *mBio* **10**.
368. Schutze, J., Krasko, A., Custodio, M.R., Efremova, S.M., Muller, I.M., and Muller, W.E. (1999) Evolutionary relationships of Metazoa within the eukaryotes based on molecular data from Porifera. *Proc Biol Sci* **266**: 63-73.
369. Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068-2069.

370. Sheikhpour, M., Sadeghi, A., Yazdian, F., Movafagh, A., and Mansoori, A. (2019) Anticancer and Apoptotic Effects of Ectoine and Hydroxyectoine on Non-Small Cell Lung Cancer cells: An in-vitro Investigation. *Multidisciplinary Cancer Investigation* **3**: 14-19.
371. Shi, Q., Liu, G.H., Yan, H.Q., and Zhang, H.L. (2012) Black disease (Terpios hoshinota): a probable cause for the rapid coral mortality at the northern reef of Yongxing Island in the South China Sea. *Ambio* **41**: 446-455.
372. Shim, J.S., and Liu, J.O. (2014) Recent advances in drug repositioning for the discovery of new anticancer drugs. *Int J Biol Sci* **10**: 654-663.
373. Shindo, K., Kikuta, K., Suzuki, A., Katsuta, A., Kasai, H., Yasumoto-Hirose, M. et al. (2007) Rare carotenoids, (3R)-saproxanthin and (3R,2'S)-myxol, isolated from novel marine bacteria (*Flavobacteriaceae*) and their antioxidative activities. *Appl Microbiol Biotechnol* **74**: 1350-1357.
374. Shipman, J.A., Berleman, J.E., and Salyers, A.A. (2000) Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* **182**: 5365-5372.
375. Shrivastava, A., Lele, P.P., and Berg, H.C. (2015) A rotary motor drives *Flavobacterium* gliding. *Curr Biol* **25**: 338-341.
376. Shrivastava, A., Rhodes, R.G., Pochiraju, S., Nakane, D., and McBride, M.J. (2012) *Flavobacterium johnsoniae* RemA is a mobile cell surface lectin involved in gliding. *J Bacteriol* **194**: 3678-3688.
377. Silva, S.G., Blom, J., Keller-Costa, T., and Costa, R. (2019a) Comparative genomics reveals complex natural product biosynthesis capacities and carbon metabolism across host-associated and free-living *Aquimarina* (*Bacteroidetes*, *Flavobacteriaceae*) species. *Environ Microbiol* **21**: 4002-4019.
378. Silva, S.G., Blom, J., Keller-Costa, T., and Costa, R. (2019b) Comparative genomics reveals complex natural product biosynthesis capacities and carbon metabolism across host-associated and free-living *Aquimarina* (*Bacteroidetes*, *Flavobacteriaceae*) species. *Environ Microbiol*.
379. Simister, R., Taylor, M.W., Rogers, K.M., Schupp, P.J., and Deines, P. (2013) Temporal molecular and isotopic analysis of active bacterial communities in two New Zealand sponges. *FEMS Microbiol Ecol* **85**: 195-205.
380. 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**: 517-524.
381. Simpson, T.L. (1984) *The cell biology of sponges*: Springer Verlag.
382. Sipkema, D., Holmes, B., Nichols, S.A., and Blanch, H.W. (2009) Biological characterisation of *Haliclona* (?*gellius*) sp.: sponge and associated microorganisms. *Microb Ecol* **58**: 903-920.
383. Sipkema, D., Franssen, M.C., Osinga, R., Tramper, J., and Wijffels, R.H. (2005) Marine sponges as pharmacy. *Mar Biotechnol (NY)* **7**: 142-162.

384. 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**: 2130-2140.
385. Sipkema, D., de Caralt, S., Morillo, J.A., Al-Soud, W.A., Sorensen, S.J., Smidt, H., and Uriz, M.J. (2015) Similar sponge-associated bacteria can be acquired via both vertical and horizontal transmission. *Environ Microbiol* **17**: 3807-3821.
386. Sivapathasekaran, C., Das, P., Mukherjee, S., Saravanakumar, J., Mandal, M., and Sen, R. (2010) Marine Bacterium Derived Lipopeptides: Characterization and Cytotoxic Activity Against Cancer Cell Lines. *Int J Pept Res Ther* **16**: 215-222.
387. Sizikov, S., Burgsdorf, I., Handley, K.M., Lahyani, M., Haber, M., and Steindler, L. (2020) Characterization of sponge-associated *Verrucomicrobia*: microcompartment-based sugar utilization and enhanced toxin-antitoxin modules as features of host-associated *Opitutales*. *Environ Microbiol* **22**: 4669-4688.
388. 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*.
389. Song, C., Schmidt, R., de Jager, V., Krzyzanowska, D., Jongedijk, E., Cankar, K. et al. (2015) Exploring the genomic traits of fungus-feeding bacterial genus *Collimonas*. *BMC Genomics* **16**: 1103.
390. Staley, J.T., and Konopka, A. (1985) Measurement of in situ activities of nonphotosynthetic microbes in aquatic and terrestrial habitats. *Ann Rev Microbiol* **39**: 321-346.
391. Stamatakis, A. (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-2690.
392. Stamatakis, A., Hoover, P., and Rougemont, J. (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* **57**: 758-771.
393. Steffen, K., Indraningrat, A.A.G., Erngren, I., Haglöf, J., Becking, L.E., Smidt, H. et al. Go with the flow: oceanographic setting influences the prokaryotic community and metabolome in deep-sea sponges *(submitted)*.
394. Stewart, E.J. (2012) Growing unculturable bacteria. *J Bacteriol* **194**: 4151-4160.
395. Stewart, R.D., Auffret, M.D., Roehe, R., and Watson, M. (2018) Open prediction of polysaccharide utilisation loci (PUL) in 5414 public *Bacteroidetes* genomes using PULpy. *bioRxiv*.
396. Storey, M.A., Andreassend, S.K., Bracegirdle, J., Brown, A., Keyzers, R.A., Ackerley, D.F. et al. (2020) Metagenomic Exploration of the Marine Sponge *Mycale hentscheli* Uncovers Multiple Polyketide-Producing Bacterial Symbionts. *mBio* **11**.
397. Sunagawa, S., Coelho, L.P., Chaffron, S., Roat Kultima, J., Labadie, K., Salazar, G., and al., e. (2015a) Structure and function of the global ocean microbiome. *Science* **348**.

398. Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G. et al. (2015b) Ocean plankton. Structure and function of the global ocean microbiome. *Science* **348**: 1261359.
399. Sweet, M.J., Croquer, A., and Bythell, J.C. (2010) Bacterial assemblages differ between compartments within the coral holobiont. *Coral Reefs* **30**: 39-52.
400. Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol* **30**: 2725-2729.
401. Taylor, J.A., Palladino, G., Wemheuer, B., Steinert, G., Sipkema, D., Williams, T.J., and Thomas, T. (2021) Phylogeny resolved, metabolism revealed: functional radiation within a widespread and divergent clade of sponge symbionts. *ISME J* **15**: 503-519.
402. Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007a) Sponge-associated microorganisms: evolution, ecology, and biotechnological potential. *Microbiol Mol Biol Rev* **71**: 295-347.
403. Taylor, M.W., Hill, R.T., Piel, J., Thacker, R.W., and Hentschel, U. (2007b) Soaking it up: the complex lives of marine sponges and their microbial associates. *ISME J* **1**: 187-190.
404. Taylor, M.W., Tsai, P., Simister, R.L., Deines, P., Botte, E., Ericson, G. et al. (2013) 'Sponge-specific' bacteria are widespread (but rare) in diverse marine environments. *ISME J* **7**: 438-443.
405. Teruya, T., Nakagawa, S., Koyama, T., Arimoto, H., Kita, M., and Uemura, D. (2004) Nakiterpiosin and nakiterpiosinone, novel cytotoxic C-nor-D-homosteroids from the Okinawan sponge *Terpios hoshinota*. *Tetrahedron* **60**: 6989-6993.
406. Tholl, D. (2006) Terpene synthases and the regulation, diversity and biological roles of terpene metabolism. *Curr Opin Plant Biol* **9**: 297-304.
407. Thomas, F., Hehemann, J.H., Rebuffet, E., Czjzek, M., and Michel, G. (2011) Environmental and gut bacteroidetes: the food connection. *Front Microbiol* **2**: 93.
408. Thomas, T., Gilbert, J., and Meyer, F. (2012) Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp* **2**: 3.
409. Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A. et al. (2010a) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J* **4**: 1557-1567.
410. Thomas, T., Moitinho-Silva, L., Lurgi, M., Bjork, J.R., Easson, C., Astudillo-Garcia, C. et al. (2016) Diversity, structure and convergent evolution of the global sponge microbiome. *Nat Commun* **7**: 11870.
411. Thomas, T.R., Kavlekar, D.P., and LokaBharathi, P.A. (2010b) Marine drugs from sponge-microbe association--a review. *Mar Drugs* **8**: 1417-1468.
412. Thompson, D., Cognat, V., Goodfellow, M., Koechler, S., Heintz, D., Carapito, C. et al. (2020) Phylogenomic Classification and Biosynthetic Potential of the Fossil Fuel-Biodesulfurizing *Rhodococcus* Strain IGTS8. *Front Microbiol* **11**: 1417.
413. Thompson, L.R., Sanders, J.G., McDonald, D., Amir, A., Ladau, J., Locey, K.J. et al. (2017) A communal catalogue reveals Earth's multiscale microbial diversity. *Nature* **551**: 457-463.

414. Tianero, M.D., Balaich, J.N., and Donia, M.S. (2019) Localized production of defence chemicals by intracellular symbionts of *Haliclona* sponges. *Nat Microbiol* **4**: 1149-1159.
415. Tong, A.Z., Liu, W., Liu, Q., Xia, G.Q., and Zhu, J.Y. (2021) Diversity and composition of the Panax ginseng rhizosphere microbiome in various cultivation modesand ages. *BMC Microbiol* **21**: 18.
416. Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., and Orti, G. (2015) Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. *PLoS Curr* **7**.
417. Turekian, V.C., Macko, S.A., and Keene, W.C. (2003) Concentrations, isotopic compositions, and sources of size-resolved, particulate organic carbon and oxalate in near-surface marine air at Bermuda during spring. *J Geophys Res Atmos* **108**.
418. Turner, E.C. (2021) Possible poriferan body fossils in early Neoproterozoic microbial reefs. *Nature* **596**: 87-91.
419. Tyson, G.W., Chapman, J., Hugenholtz, P., Allen, E.E., Ram, R.J., Richardson, P.M. et al. (2004) Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.
420. Ueoka, R., Uria, A.R., Reiter, S., Mori, T., Karbaum, P., Peters, E.E. et al. (2015) Metabolic and evolutionary origin of actin-binding polyketides from diverse organisms. *Nat Chem Biol* **11**: 705-712.
421. Umezawa, H., Okami, Y., Kurasawa, S., Ohnuki, T., Ishizuka, M., Takeuchi, T. et al. (1983) Marinactan, antitumor polysaccharide produces by marine bacteria. *J Antibiot (Tokyo)* **XXXVI**: 471-477.
422. Uria, A., and Piel, J. (2009) Cultivation-independent approaches to investigate the chemistry of marine symbiotic bacteria. *Phytochem Rev* **8**: 401-414.
423. Uritskiy, G.V., DiRuggiero, J., and Taylor, J. (2018) MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. *Microbiome* **6**: 1-13.
424. Vacelet, J., and Donadey, C. (1977) Electron microscope study of the association between some sponges and bacteria. *J Exp Mar Biol Ecol* **30**: 301-314.
425. Vailati-Riboni, M., Palombo, V., and Loor, J.J. (2017) What are omics sciences? In *Periparturient diseases of dairy cows*: Springer, pp. 1-7.
426. Veith, P.D., Glew, M.D., Gorasia, D.G., and Reynolds, E.C. (2017) Type IX secretion: the generation of bacterial cell surface coatings involved in virulence, gliding motility and the degradation of complex biopolymers. *Mol Microbiol* **106**: 35-53.
427. Versluis, D., McPherson, K., van Passel, M.W.J., Smidt, H., and Sipkema, D. (2017a) Recovery of Previously Uncultured Bacterial Genera from Three Mediterranean Sponges. *Mar Biotechnol* **19**: 454-468.
428. Versluis, D., McPherson, K., van Passel, M.W.J., Smidt, H., and Sipkema, D. (2017b) Recovery of Previously Uncultured Bacterial Genera from Three Mediterranean Sponges. *Mar Biotechnol* **19**: 454-468.

429. Versluis, D., Nijsse, B., Naim, M.A., Koehorst, J.J., Wiese, J., Imhoff, J.F. et al. (2018) Comparative Genomics Highlights Symbiotic Capacities and High Metabolic Flexibility of the Marine Genus *Pseudovibrio*. *Genome Biol Evol* **10**: 125-142.
430. Waite, D.W., Chuvochina, M., Pelikan, C., Parks, D.H., Yilmaz, P., Wagner, M. et al. (2020) Proposal to reclassify the proteobacterial classes *Deltaproteobacteria* and *Oligoflexia*, and the phylum *Thermodesulfobacteria* into four phyla reflecting major functional capabilities. *Int J Syst Evol Microbiol* **70**: 5972-6016.
431. Wakimoto, T., Egami, Y., Nakashima, Y., Wakimoto, Y., Mori, T., Awakawa, T. et al. (2014) Calyculin biogenesis from a pyrophosphate protoxin produced by a sponge symbiont. *Nat Chem Biol* **10**: 648-655.
432. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S. et al. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* **9**: e112963.
433. Wang, S., Zhao, D., Bai, X., Zhang, W., and Lu, X. (2017) Identification and Characterization of a Large Protein Essential for Degradation of the Crystalline Region of Cellulose by *Cytophaga hutchinsonii*. *Appl Environ Microbiol* **83**: e02270-02216.
434. Wang, X., Zhou, H., Chen, H., Jing, X., Zheng, W., Li, R. et al. (2018) Discovery of recombinases enables genome mining of cryptic biosynthetic gene clusters in *Burkholderiales* species. *Proc Natl Acad Sci U S A* **115**: E4255-E4263.
435. Wang, Z., Guo, F., Liu, L., and Zhang, T. (2014) Evidence of carbon fixation pathway in a bacterium from candidate phylum SBR1093 revealed with genomic analysis. *PLoS One* **9**: e109571.
436. Waters, A.L., Peraud, O., Kasanah, N., Sims, J.W., Kothalawala, N., Anderson, M.A. et al. (2014) An analysis of the sponge *Acanthostrongylophora igens'* microbiome yields an actinomycete that produces the natural product manzamine A. *Front Mar Sci* **1**.
437. Waterworth, S.C., Parker-Nance, S., Kwan, J.C., and Dorrington, R.A. (2021) Comparative genomics provides insight into the function of broad-host range sponge symbionts. *bioRxiv*: 2020.2012. 2009.417808.
438. Webster, N.S. (2007) Sponge disease: a global threat? *Environ Microbiol* **9**: 1363-1375.
439. Webster, N.S., and Taylor, M.W. (2012) Marine sponges and their microbial symbionts: love and other relationships. *Environ Microbiol* **14**: 335-346.
440. Webster, N.S., and Thomas, T. (2016) The Sponge Hologenome. *MBio* **7**: e00135-00116.
441. Webster, N.S., Wilson, K.J., Blackall, L.L., and Hill, R.T. (2001) Phylogenetic diversity of bacteria associated with the marine sponge *Rhopaloeides odorabile*. *Appl Environ Microbiol* **67**: 434-444.
442. Webster, N.S., Taylor, M.W., Behnam, F., Lucker, S., Rattei, T., Whalan, S. et al. (2010) Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol* **12**: 2070-2082.
443. Westram, R., Bader, K., Pruesse, E., Kumar, Y., Meier, H., Glockner, F.O., and Ludwig, W. (2011) ARB: a software environment for sequence data. In *Handbook of Molecular*

Microbial Ecology I: Metagenomics and Complementary Approaches. de Bruijn, F.J. (ed): John Wiley & Sons, pp. 399-406.

444. Whitman, W.B. (2015) Genome sequences as the type material for taxonomic descriptions of prokaryotes. *Syst Appl Microbiol* **38**: 217-222.
445. Whitman, W.B., Coleman, D.C., and Wiebe, W.J. (1998) Prokaryotes: the unseen majority. *Proc Natl Acad Sci U S A* **95**: 6578-6583.
446. Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. ISBN 978-3-319-24277-4. https://ggplot2.tidyverse.org.
447. Wilde, A., and Mullineaux, C.W. (2015) Motility in cyanobacteria: polysaccharide tracks and Type IV pilus motors. *Mol Microbiol* **98**: 998-1001.
448. Wilkinson, C., and Fay, P. (1979) Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. *Nature* **279**: 527-529.
449. Wilkinson, C.R. (1983) Net primary productivity in coral reef sponges. *Science* **219**: 410-412.
450. Williams, C.R., Chen, L., Driver, A.D., Arnold, E.A., Sheppard, E.S., Locklin, J., and Krause, D.C. (2018) Sialylated Receptor Setting Influences Mycoplasma pneumoniae Attachment and Gliding Motility. *Mol Microbiol* **109**: 735-744.
451. Wilson, M.C., and Piel, J. (2013) Metagenomic approaches for exploiting uncultivated bacteria as a resource for novel biosynthetic enzymology. *Chem Biol* **20**: 636-647.
452. Wilson, M.C., Mori, T., Ruckert, C., Uria, A.R., Helf, M.J., Takada, K. et al. (2014) An environmental bacterial taxon with a large and distinct metabolic repertoire. *Nature* **506**: 58-62.
453. Wolkin, R.H., and Pate, J.L. (1984) Translocation of motile cells of the gliding bacterium *Cytophaga johnsonae* depends on a surface component that may be modified by sugars. *J Gen Microbiol* **130**: 2651-2669.
454. Wollenberg, M.S., and Ruby, E.G. (2012) Phylogeny and fitness of *Vibrio fischeri* from the light organs of *Euprymna scolopes* in two Oahu, Hawaii populations. *ISME J* **6**: 352-362.
455. Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A., and Singer, S.W. (2014) MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* **2**.
456. Wulff, J.L. (2006) Ecological interactions of marine sponges. *Can J Zool* **84**: 146-166.
457. Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. et al. (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9**: 1410-1422.
458. Xu, L., Ye, K.X., Dai, W.H., Sun, C., Xu, L.H., and Han, B.N. (2019) Comparative Genomic Insights into Secondary Metabolism Biosynthetic Gene Cluster Distributions of Marine *Streptomyces*. *Mar Drugs* **17**.

459. Yamada, Y., Kuzuyama, T., Komatsu, M., Shin-Ya, K., Omura, S., Cane, D.E., and Ikeda, H. (2015) Terpene synthases are widely distributed in bacteria. *Proc Natl Acad Sci U S A* **112**: 857-862.
460. Yang, M., Nightingale, P.D., Beale, R., Liss, P.S., Blomquist, B., and Fairall, C. (2013) Atmospheric deposition of methanol over the Atlantic Ocean. *Proc Natl Acad Sci U S A* **110**: 20034-20039.
461. Yarza, P., Yilmaz, P., Pruesse, E., Glockner, F.O., Ludwig, W., Schleifer, K.H. et al. (2014) Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. *Nat Rev Microbiol* **12**: 635-645.
462. Yilmaz, L.S., Parnerkar, S., and Noguera, D.R. (2011) mathFISH, a web tool that uses thermodynamics-based mathematical models for in silico evaluation of oligonucleotide probes for fluorescence in situ hybridization. *Appl Environ Microbiol* **77**: 1118-1122.
463. Yilmaz, P., Parfrey, L.W., Yarza, P., Gerken, J., Pruesse, E., Quast, C. et al. (2014) The SILVA and "All-species Living Tree Project (LTP)" taxonomic frameworks. *Nucleic Acids Res* **42**: D643-648.
464. Yin, H., Guo, C., Wang, Y., Liu, D., Lv, Y., Lv, F., and Lu, Z. (2013) Fengycin inhibits the growth of the human lung cancer cell line 95D through reactive oxygen species production and mitochondria-dependent apoptosis. *Anticancer Drugs* **24**: 587-598.
465. Yoon, B.J., and Oh, D.C. (2012) *Spongiibacterium flavum* gen. nov., sp. nov., a member of the family *Flavobacteriaceae* isolated from the marine sponge *Halichondria oshoro*, and emended descriptions of the genera *Croceitalea* and *Flagellimonas*. *Int J Syst Evol Microbiol* **62**: 1158-1164.
466. Yooseph, S., Nealson, K.H., Rusch, D.B., McCrow, J.P., Dupont, C.L., Kim, M. et al. (2010) Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60-66.
467. Zhang, G., Li, J., Zhu, T., Gu, Q., and Li, D. (2016) Advanced tools in marine natural drug discovery. *Curr Opin Biotechnol* **42**: 13-23.
468. Zhang, H., Zhao, Z., and Wang, H. (2017) Cytotoxic Natural Products from Marine Sponge-Derived Microorganisms. *Mar Drugs* **15**.
469. Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z. et al. (2018) dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **46**: W95-W101.
470. Zhang, W., Li, F., and Nie, L. (2010) Integrating multiple 'omics' analysis for microbial biology: application and methodologies. *Microbiology* **156**: 287-301.
471. Zhao, B., Lin, X., Lei, L., Lamb, D.C., Kelly, S.L., Waterman, M.R., and Cane, D.E. (2008) Biosynthesis of the sesquiterpene antibiotic albaflavenone in *Streptomyces coelicolor* A3(2). *J Biol Chem* **283**: 8183-8189.
472. Zhou, Z., Tran, P.Q., Kieft, K., and Anantharaman, K. (2020) Genome diversification in globally distributed novel marine *Proteobacteria* is linked to environmental adaptation. *ISME J* **14**: 2060-2077.

473. Zhu, Y., and McBride, M.J. (2016) Comparative Analysis of *Cellulophaga algicola* and *Flavobacterium johnsoniae* Gliding Motility. *J Bacteriol* **198**: 1743-1754.
474. Zimmerman, A.E., Martiny, A.C., and Allison, S.D. (2013) Microdiversity of extracellular enzyme genes among sequenced prokaryotic genomes. *ISME J* **7**: 1187-1199.

Appendices

Co-author affiliations

Hauke Smidt[1]
Detmer Sipkema[1]
Colin J. Ingham[2]
Mark W.J. van Passel[1,3]
Johanna Gutleben[1]
Dennis Versluis[1]
Francesca Forgiarini[1]
Thomas A. Mackenzie[4]
Pilar Sánchez[4]
José R. Tormo[4]
Eike E. Peters[5]
Jackson K. B. Cahn[5]
Ursula A. E. Steffens[6]
Catarina Loureiro[1]
Michelle A. Schorn[1]
Paco Cárdenas[7]
Nilani Vickneswaran[6]
Philip Crews[8]
Jörn Piel[5]
Burak Avci[1]
Anastasia Galani[1]
Thijs J.G. Ettema[1]

[1] Laboratory of Microbiology, Wageningen University and Research, Stippeneng 4, 6708 WE Wageningen, The Netherlands
[2] Hoekmine BV, 3515 GJ Utrecht, The Netherlands
[3] Ministry of Health, Welfare and Sport, Parnassusplein 5, 2511 VX, The Hague, The Netherlands
[4] Fundación MEDINA, Centro de Excelencia en Investigación de Medicamentos Innovadores en Andalucía, Avda. del Conocimiento 34, 18016 Granada, Spain
[5] Institute of Microbiology, Eidgenössische Technische Hochschule (ETH) Zurich, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland
[6] Kekule Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Strasse 1, 53121 Bonn, Germany.
[7] Pharmacognosy, Department of Pharmaceutical Biosciences, BioMedical Center, Uppsala University, Husargatan 3, 75124 Uppsala, Sweden
[8] Department of Chemistry and Biochemistry, University of California at Santa Cruz, Santa Cruz, California, United States.

Acknowledgements

(Rumour has it that this is the most popular part of a PhD thesis. I hope I won't disappoint the readers and I apologize in advance if I miss someone's name since acknowledging people who have contributed in 5 years of work sometimes can be even more challenging than reporting the work itself.)

It may be a bit cliché, but after half of a decade I can say, with experience now, that PhD life is not about the destination (aka the degree), it's about the journey. So many moments, emotions, challenges...turned out to be lifelong memories. No regrets for me! Looking back, this journey would have never been successful without the help and support of all of you who will probably read these pages and that makes me feel extremely grateful.

Dear **Detmer**, I could not have asked for a better mentor and thank you so much for giving me the opportunity to work with you. You were always there for me when I was freaking out, to calm me down and put me back on track. I will agree with Johanna that you are *our Lord and Savior*. Every time I would talk with you, it was like I was hitting the reset button and finding the confidence and motivation to go on...all this in a magical way. Thank you so much for having faith in me. Your scientific insights and knowledge played a big role in the successful completion of my PhD. It was an honor to be one of your students and I gained so much from you, which made me a better scientist and person and equipped me well for the future. I would also like to thank you for involving me in both the MarPipe and SponGES projects, which gave me the chance to meet all these amazing people, travel around the world and do cool science! I really appreciate it. I am glad that I spent some time with you during these trips, because I realized what a kind, genuine and fun person you are. Of course, I will never forget our adventures during our travels, especially our sampling cruise (ghost-sponge hunting) to the Arctic and when sponges finally appeared in the fjords... they never showed up in Schiphol. Last but not least, I think we mainly share two passions: sponge research and whiskey. I hope we manage to have this whiskey soon!

Dear **Hauke,** thank you for being my promotor for 5 years (minus a week). I still remember your last question during my interview for the PhD position: "How would you describe metagenomics in your own words?". I feel so lucky that I had read the term in Wikipedia and I was able to answer. You are a great supervisor and mentor with an admirable breadth of scientific knowledge. Always kind and eager to guide and support me throughout my PhD. We didn't get to spend a lot of time together, but I really enjoyed our progress meetings and brainstorming/problem-solving sessions with you, Detmer and Colin. Thank you so much for trusting me in this and also involving me in both the Coloured by Flavo and HoloRuminant projects. I am looking forward to continue doing science with you as part of the MolEco group!

Dear **Colin,** thank you very much for giving me the opportunity to become a PhD student since you were the first to contact me for the MarPipe position. It was great to have you as my supervisor! You are a very kind and interesting person and it was a pleasure to work with you. Your passion about research inspired me many times! Your guidance and valuable feedback contributed a lot in my thesis. Even though flavos made my life a bit difficult at the beginning of my PhD, I still find them extremely cool!

I would also like to acknowledge the members of the **PhD thesis committee** for dedicating time in evaluating my research work.

Beyond science, life cannot be sustained without friends. I don't think I would have survived this rollercoaster without my **paranymphs**, **Anastasia**, **Athanasia** and **Kassiani**. Thank you so much for your support, especially during the final stretch of my PhD.

Anastasia, I am so happy that I played a small part in your decision to come to Wageningen for your studies. I still remember our first meeting with Detmer for your MSc thesis. You were so enthusiastic in explaining your own research ideas that I could see you becoming a great scientist! You will probably agree with me that it was fate that brought us together since we have followed almost the same path throughout our academic career with a 'phase delay' of 4 years. I still consider you one of my sponge people even though you jumped to CO stuff. Our discussions (scientific and not) during our 'fresh air' breaks were always helpful, insightful and mostly fun! Thank you for being an amazing labmate and officemate! We shared so many fun moments together working late in the office, singing to Greek songs (Παρακάλα να πεθάνω, Μπήκαν στην πόλη οι οχτροί, κλπ.)! We still need to drink that wine, vinegar by now. Apart from these, you have also been a true friend and family! Thank you for always being there for me, supporting me and allowing me to bully you! Looking forward to many more exciting moments in the next season (6.0)! Σ'αγαπώ πολύ και σ'ευχαριστώ, μικρή μου!

Athanasia (Nancy), you were one of the first people I started hanging out with in Wageningen. I remember the day I met you in Thessaloniki and we found out that both of us would be studying in WUR in a few months! Such a nice coincidence...that turned out to become a strong friendship. You are the voice of reason/the emergency contact/the encyclopedia/the hard drive, always willing to hear me and give me advice. I am also very happy to have you as a labmate in the MolEco group. You are so bright and passionate about your research that I am sure you will have a successful career ahead! Thank you so much for helping me out during my thesis and side projects, explaining to me gut microbiota stuff when I needed it. I love your sense of humor and I have so much fun when I am around you! I hope we share many more memorable moments together in the future! Σ'αγαπώ πολύ και σ'ευχαριστώ, Νανσάκι μου (πιπιπιπιπι)!

Kassiani (Kassi), you know you are in the heart of my PhD life Venn diagram. I remember you coming to Helix after your courses as a MSc student to join our 'fresh air' breaks and complain about our lives and now...our desks are two offices away. I have witnessed your progress as an unofficial supervisor and I am so proud of you! We are so similar and yet so different, but this didn't stop us from building up a strong friendship! Thank you so much for the support and all the times you had to deal with my stress and moodiness! I would also like to thank you for the awesome holidays in Kos! I cherish our moments and I am looking forward to many more! Σ'αγαπώ πολύ και σ'ευχαριστώ, Κασσιανούλα μου!

Being member of the **MolEco** group and the **Laboratory of Microbiology (MIB)** was one of the best parts of my PhD journey. **Willem**, I feel honored to have met you and been part of MIB while you were the chair of the department. PhD meetings with you were a unique experience and I really appreciated your input on my presentations. **Thijs**, you brought about a new era for MIB and I feel grateful that I was part of it during my PhD and from now on as a postdoc. Looking forward to gaining new skills and experiences in the future as part of MIB!

My time in **MIB** would not be that pleasant without my amazing colleagues. Being part of the **MolEco family**, I was lucky enough to interact and connect with so many interesting people from different parts of the world and learn about their cultures. In pre-pandemic times, I really enjoyed our lunch breaks, MolEco cakes, Christmas dinners but also the MIB strategy days, barbeques and parties. For these, I would first like to thank all my **MolEco** buddies. **Steven**, at the beginning you were quite intimidating, but later when I started to get to know you better, I realized how nice and fun you are to be around. You are a great pillar of our lab and I am glad for the times we had together. I envy you for the best costume awards you have won during our MolEco themed dinners. Thank you for answering all of my silly questions! **Ineke**, you are always smiley and helpful. I was happy to come to you every time I needed help and thanks a lot for this! **Merlijn**, you are such a polite, helpful and hard-working person. Thank you for taking care of the lab and sorry for sometimes changing the music in the lab! **Laura**, you started working in MIB during the pandemic so I didn't have the chance to get to know you well. Still, we had a few good laughs at the 5th floor canteen. **Jannie**, I think our friendship kicked-off during the PhD trip and I am so happy that I got to know you better! Thank you very much for all the awesome memories and the support, especially during our final stretch. Can't wait to attend your PhD defense and celebrate with you! **Carrie**, I can hear your voice in my ears right now calling me "Meniaaaaa!" every time you would see me at the corridor. I really enjoyed our talks and had a lot of fun with you. I am glad I had you around during my PhD. **Marina**, my office neighbor/therapist, thank you so much for our long evening talks and your emotional support during the last phase of my PhD. Your advice helped me a lot and you should definitely start charging people for this! **Caifang**, little one, you are the sweetest and so much fun! I am happy I got to know you more and I wish you all the best for your thesis and future! **Martha**, you are so sweet and kind. Thank you for all the complaining-about-our-theses chats and your support. You

always cheered me up! **Ran**, I am glad I got to spend some time with you at the PhD trip. You are so polite and interesting to talk to with a great sense of humor! **Kate**, thanks for all the chit chats we had and I wish you all the best for you and your family - just noticed your belly got bigger! **Chen** and **Taojun**, my office neighbors, you are so kind, hard-working and passionate about your science. When I was coming to the office on the weekends, I knew I would find you there and I was always happy to see you guys. Good luck in finalizing your theses! **Patrick**, we didn't interact a lot but I remember you being a student of Jie and now you are almost a Dr. as well. I appreciate you for always being interested in my presentations, asking questions and giving suggestions. **Janneke** and **Sharon,** I remember when you started working in MIB as MSc students and I saw you developing into great scientists. It was great to know you and I enjoyed our chit chats around Helix! I would also like to acknowledge the PIs of MolEco since they are the nodes of this network: **Hauke**, **Detmer**, **Erwin** and **Clara**, thank you for the discussions and feedback during our MolEco meetings.

MIB is a big family and for this I would also like to acknowledge all my fellow scientists from **BacGen**, **MicFys, MicEvo** and **SSB** for the nice chats and fun moments. **Giannis M.**, you were one of the first people I met here and it was always fun to chat with you. I admire you as a scientist, sooo hard-working and passionate about your research, I am sure you'll do great things in your field! Thanks for being there for me when I needed your advice! Μου λείπεις, μωρή τρέλα! **Nikolas**, you are such a sweet and caring person and always willing to hear other people's problems and offer advice. It was great to know you and spend the beginning of my PhD around you! **Christos (Χρηστάρα),** we started our PhDs around the same time and I witnessed how hard you were working on your project. I am sure you will be a very successful scientist! Thanks for all the discussions and nice moments! **Jueeli**, it was nice to have you around at the beginning of my PhD. You are such a nice and sweet person. Thank you for the support and the moments we shared together! **Despoina (Δεσποινάκι)**, I was there when you first started in the lab as a MSc student and I am happy to see you evolving into a passionate scientist during these years. I got to know you a bit better lately and you are so much fun! **Lot**, I remember we first met when we drove together to Apeldoorn for the A1 SENSE course as young and fresh PhDs. You are such an interesting person to talk to and I always enjoyed our conversations. Thanks a lot for the short therapy sessions at the 5th floor canteen! **Ivette**, thanks a lot for all the lovely moments we shared in USA. You are sweet and funny and I enjoyed our time so much (especially in Amherst - incl. Lot). **Lyon**, I have come to the conclusion that you are the best person to party with! Thank you for all the nice memories, walks, talks and karaoke songs we shared together during our PhD trip! Now that I mentioned parties...**Max,** I have to acknowledge you for being my entry pass to some of the best parties I've been to in Wageningen! **Irene**, thanks for the morning chats at the canteen! **Burak**, you are very passionate about your research and it was a pleasure to collaborate with you during my PhD!

Our lab would not be up and running without the support of the technicians, administration and IT staff. **Steven**, **Ineke**, **Merlijn**, **Laura**, **Hans**, **Rob**, **Philippe**, **Sjon**, **Wim**, **Wilma**, **Tom**, **Iame**, **Anja**, **Heidi**, **Bart** and **Felix** thank you very much for your help and patience!

Special thanks to the **Marines** (aka Detmer's kids) for all the discussions, meetings and fun moments: **Johanna**, **Georg**, **Jie**, **Dat**, **Tika**, **Indra**, **Leire**, **Catarina**, **Michelle**, **Anastasia**, **Valentina**. **Georg**, our times didn't overlap much but it was really nice having you around at the beginning of my PhD. Thanks a lot for the tutorials-legacy you left behind! **Tika**, you are such a kind and sweet person. Thanks for all the nice chats and the amazing dinner you cooked for us at your house (still remember the tempeh!). **Indra**, I really miss you in the lab. You were always fun, teasing people! We had a great time in our trip to Porto for the SponGES project meeting! Thank you for everything. **Jie**, you are such a cool guy with great sense of humor. I've heard you already became rich which doesn't surprise me! Thanks a lot for being around during most of my PhD time. **Dat**, you are one of the most smiley people I know. It was great to have you as a groupmate and officemate. **Valentina**, we haven't spent a lot of time together but I think you are a great addition to our group. Good luck with your thesis!

Catarina, my Geodia buddy, we almost started our PhDs at the same time and I have a vivid memory of you coming tanned from Australia, telling stories about swimming with sharks while complaining about the weather in the Netherlands! You are a very fun and active person which I've always admired in you (since I am super lazy). Thank you for being with me in the lab that summer of DNA/RNA extractions (listening non-stop to the Black Keys album) and for the smooth collaborations in 50% of my thesis! Good luck with finalizing your thesis now and I wish you no more Gb3 nightmares!

Michelle, you are so sweet, kind and always eager to help people. A great scientist with deep knowledge of your topic and an admirable supervisor! Your students are very lucky to have you. I am so happy you joined our group (and office) and for our collaboration in two of my chapters! Thank you very much for all the times you heard me complaining about my thesis and supported me, especially during the final stretch! I would also like to thank you for the awesome dinners, barbeques and all the times you, Mo and Sasha hosted us. You are great hosts!

During my PhD, I had the chance to supervise two very bright, motivated and hard-working MSc students. **Francesca**, you were my first student and we learned many things together since I was at the very beginning of my PhD. Thank you for helping me with the Flavos and I am grateful we included part of your thesis in chapter 2! **Sander**, you were so independent from day 1 with your own ideas and motivated about your project. You taught me a lot about bioinformatics and your metaWRAP work was very important for the Marine group. Thank you both so much for contributing to my thesis!

Teaching plays big part in the WUR PhD life and it's not always fun. In my case, I was again lucky to teach the Microbial Ecology course for 5 years in a row. I really enjoyed my time and especially due to the people who were teaching with me: **Gerben**, **Peng**, **Tika**, **Jie**, **Leire**, **Catarina**, **Prokopis**, **Ruth** and **Michelle**, thank you so much for making the teaching process easier and fun!

One of the highlights of this journey was our **MIB-SSB PhD trip** to Boston and New York. I would first like to acknowledge the members of the PhD trip committee (**Giannis K.**, **Costas**, **Catarina**, **Ivette**, **Lot**, **Enrique**, **Ran**, **Caifang** and **Nong**) for putting in all this effort and time to organize the trip and make sure everything ran smoothly. Special thanks to **Diana** and **Raymond** for being the responsible people accompanying us! I will never forget how much fun we had in room 5015 in Boston and for this I want to thank **Caifang**, **Martha**, **Marina**, **Mamou** and **Ran**. Girls, we had great laughs and it was so nice to get to know you better! **Marina** and **Lyon** you are my karaoke buddies, thanks for the great moments! The TAM, the breakfast in Johnny's Roadside, the nightclub in Soho, the road trip from Boston to NY, the Greek night...thank you all for these amazing memories! <3

Most of you know how emotionally attached I am to my desk and **office 5033**. I've spent a loooot of hours there since I was in front of a screen for most my PhD time. I would like to thank my officemates for all the fun (and not so fun) moments we shared together: **Johanna**, **Loo Wee**, **Sudarshan (Sudi)**, **Jie**, **Dat**, **Leire**, **Kate**, **Anastasia**, **Michelle**, **Yang** and **Nam**. You were always willing to hear me nagging and support me and I feel grateful for this! **Johanna** and **Sudi**, thank you so much for the very warm welcome in the office at the beginning of my PhD...I still remember our first 'fresh air' break! Me and **Anastasia**, often feel nostalgic about our sessions. **Leire**, it was so much fun having you beside me. I miss having you around! **Yang**, it is my pleasure to share the office with you. Good luck with finalizing your thesis! **Nam**, thanks for the nice chats we 've had the last couple of months! And **Johanna**, fyi...the plants are thriving!

Besides the MIB family, I was lucky enough to be part of two consortia during my PhD (**Detmer**, thanks again for this). I would like to thank the **MarPipe family** (**MarPipes** or **MarPipians**) for the fantastic moments we created together during our meetings and trips! Crazy mid-term review meeting in Brussels, seeing Aurora Borealis in the middle of a snowstorm in Tromso, attending boring courses in CNR, partying hard in Napoli, enjoying Belgian beers in Brugge, delicious tapas in Granada...and many more memorable experiences! **Alejandro**, **Anky**, **Arianna**, **Daniel**, **Florent**, **Grant**, **Kevin**, **Jane**, **Maria**, **Sloane**, **Yannik**...we are all so different and yet so similar individuals! Being among the last MarPipians to graduate, I feel so proud seeing you all progress in your own way. Thank you for being awesome science buddies and friends! I hope we see each other soon! **Anky**, **Maraki** (and **Silvia**), κοριτσάρες μου, ευχαριστώ για τις όμορφες στιγμές! Καλή τύχη σε ό,τι κι αν κάνετε! **Grant**, you are so

kind, hard-working and always willing to help! I am so happy to see you evolve into a cool scientist and dad. **Arianna** (Ca___o), thanks for all the fun moments we shared! You will always be my crazy Italian friend. I would also like to acknowledge the rest of the family members, the **MarPipe supervisors,** for the support during this journey (**Rainer**, you are one of a kind!).

I might be a bit biased here but I believe that 'sponge people' are amazing. I have been part of the sponge community since I was a bachelor student, but this became even more evident to me after joining the **SponGES family**. Again, I feel extremely lucky for having the chance to go on a sampling cruise to the Arctic, where I met two of my favorite sponge people. **Vassia**, I think you are one of the few people I know that make me laugh that much (σε λέω λίγο γελοία εδώ, αν το κατάλαβες). I admired your passion about reproduction in sponges and I will never forget your jokes during your presentations. Thank you so much for sharing the cabin with me and tolerating my snoring! **Kathrin (Katerinaki)**, I am so happy I met you at the sampling cruise. You are such a kind and sweet person. Thanks for the great chats and enjoyable moments! I wish Ute would keep you less busy during our meetings, so that we could spend more time together! Girls, I will never forget our sleepless nights processing samples, our moments on deck taking pictures of the midnight sun and our long walk to the beach in Porto. Thank you so much and good luck with everything! I hope I see you soon. **Karin**, we didn't have the chance to spend a lot of time together but we got along from day 1. You are such a passionate and bright scientist and super positive person! Wish you all the best for your future! I would also like to express my gratefulness to the late **Hans Tore** for his invaluable help in sampling the Geodias and, together with **Joana Xavier,** successfully coordinating the SponGES Project.

This is the moment to acknowledge a number of VIPs who made my PhD journey particularly sustainable and delightful during these years in the village...

Johanna, you are the first person I met in Wageningen and in MIB. I still remember my first day in Helix, when you showed me around and introduced me to our officemates. I really need to thank you for granting me the Flavos and we now have a nice chapter together! You helped me so much in my first steps and you were always there for me. Thank you so much for all the support, chats, therapy sessions, laughs, 'fresh air' breaks, droef moments. I miss your energy and smile, dude! I hope I see you soon!

Sudi, you belong to the very few people that fall into multiple 'categories': groupmate, officemate, housemate and friend. You are an amazing scientist with such passion about microbiology that I am sure you will have an exceptional career! You were always there for me, to listen to my problems, answer my questions and give me advice. I really enjoyed our moments together, sharing stories from India and Greece that made me realize how

culturally similar we are! Our 'fresh air' breaks were so fun and inspirational (I am sure Anastasia and Johanna will agree with me). Thank you for teaching me how to cook red lentil dal and for all the delicious meals you shared with me! I wish all the best for you and Diksha and please...you really need to visit Greece!

Leire (darling), you are one of my favorite sponge people. So dedicated and passionate about your research topic! It was so great to have you around during my PhD and I really miss you. You were always sweet and supportive, cheering me up whenever I would feel sad or stressed. Thank you for all the fun moments in the office, lab, gym (!) and the lunch breaks, dinners and drinks we shared together! I wish all the best for you and Miguel! I can't wait to meet the new family member! <3

Ruth (cariño), you are one of the few people who was by my side throughout the whole thing! I admire you so much for your hard work, scientific knowledge and patience! You know better than anyone that a PhD is equal to permanent head damage! Thank you for the support, the nice moments, complaining-about-work sessions, lunch breaks, drinks and deep conversations! Best of luck for your future and I hope I see you soon!

Giannis, I remember like it was yesterday when you met me at the canteen and you asked me "Are you Menia?". From the first moment, you helped me adjust and feel comfortable in the lab! Είσαι από τα πιο φιλότιμα άτομα που γνωρίζω (Sorry for the Greek but there is no English word for 'filotimo'). I knew I always had someone to lean on. You were there to listen, support and help me with whenever I needed it and thank you for this! We miss you around here!

Eleftheria, thank you so much for your support all these years and your patience all these times that we were talking about work and you had to endure it! I am so happy for you guys and I wish all the best for you, Giannis and Iasonas!

Costas (dj Pati), you cheer me up every time I meet you at the corridor or you pass by the office! I am so glad that I got to experience your positive energy, bad jokes and great dancing skills all these years, since we started and are finishing our PhD journeys (almost) at the same time. Thank you so much for your support, especially during the final stretch! Looking forward to celebrating together at your defence! Best of luck for your future career and personal life!

Prokopis, I am so grateful that you were by my side during my PhD time! Thanks a lot for our chats and the times you offered your advice! You always showed understanding since we have many things in common (έχουμε τον ίδιο μπαμπά άλλωστε)! I wish you all the best and soon to be Dr. !

Παιδιά, ευχαριστώ για όλες τις όμορφες στιγμές!

Aristofani (Bro μου), you are one of the kindest and brightest people I 've ever met! You played a big part in maintaining my well-being in Leeuw during my PhD and thank you for this. Always there, the voice of reason, to help me deal with my stress outbreaks and offer a warm, delicious meal cooked by chef Aris! It was an honor to be one of your paranymphs and share this moment with you! Σ'αγαπώ πολύ και σ'ευχαριστώ!

Sara, thank you for being such a dear friend! You are the only person I know who has gone through all the stages of a PhD without being a PhD student. So sweet and caring, always willing to hear me and stand by me! Thank you for all the support, the help during the final stretch (the cover design) and the fun moments we shared during dinners, drinks, quarantines, trips, holidays and many more! Σ'αγαπώ πολύ και σ'ευχαριστώ!

Cri, you made my Mondays way less miserable with your memes and my evenings in Dolder super fun with your jokes! I hope we find a better game to play soon (no more bocce). Thank you for all the nice moments and deep conversations! Looking forward to many more! Love you!

Vivi (Βιβίκα μου), you are my life coach-nutritionist-therapist and I am so grateful we started hanging out this past year. Your big heart, resilience and motivation are admirable and inspired me many times! Thank you so much for always being there for me, supporting me and giving me valuable advice! Σ'αγαπώ πολύ και σ'ευχαριστώ!

I would also like to thank **Iraklis** or **Hercules** (AAAAAAaaaaaaa!), **Thomas** (Θωμά, πότε θα ανοίξουμε αυτό το ταβερνείο επιτέλους;) and **Eugene** (τώρα μπορείς να με λες Δρ. Μένια και όχι κυρα-Μένια) for all the moments we shared as part of the (not-always-functional) **Wageningen family**.

Someone asked me recently if it feels like home when I visit Greece or the Netherlands. I won't disclose my answer but I admit that my life during my PhD would be unbearable without my **Leeuw family**. I have shared this house with so many different people that I might miss some names. Yet, I would like to warmly thank **Aris**, **Fotini**, **Pablo**, **Joan**, **Lorenzo**, **Kassiani (Kassi)**, **Sierra**, **Marion** and **Niamh** for making this house my comfort zone filled with love, care, laughs, music, food and alcohol! **Fotini**, I still remember how happy you were when I told you I am coming to Wageningen! You are the one who introduced me to THE house and thank you so much for this! We've known each other for forever now and we've been through a loooot! Thank you for all the nice memories and your support during my PhD! **Marion**, my guapita, thank you so much for taking care of me, cheering me up and

listening to my problems (PhD-related and not)! Even though we lived together during covid times, we had so many legendary moments! Thank you for tolerating Moody and please stop these transatlantic expeditions! Love you! **Niamh,** the youngest and yet the wisest, you have been such a great addition to the house. You witnessed my transformation into Gollum during the final stretch and still you were there to support me! Thank you for everything, little tsimpouri! Love you!

For some people, PhD life involves a certain amount of drama and in the case you are Greek this is kind of inevitable. **Anastasia**, **Kassiani** and **Nancy**...I would like to thank you from the bottom of my heart for making my daily life less (or more?!) dramatic! I feel like an observer, seeing you growing up, becoming passionate scientists (granny Menia speaking)...I am so proud of you! Thanks again for the immense support throughout my PhD and all the good and crazy times!

Besides all the people surrounding me every day, I was lucky enough to have remote support as well! For this, I would like to thank my girls: **Anastasia**, **Elina**, **Elizabeth**, **Esmy**, **Fani**, **Joanna**, **Myrto** and **Viktoria** and my boy **Konstantinos** for being there for me no matter what! Παλιόγριες και παλιόγερε, έχουμε περάσει τόσα πολλά...χαρές, λύπες, πάρτυ, διακοπές, ορκομωσίες, γάμους, γεννητούρια. Εύχομαι να είμαστε καλά και να περάσουμε ακόμα περισσότερα (κάποιες μου χρωστάτε εκδρομή 3-5 πηγάδια...δε ξεχνώ)! Σας αγαπώ πολύ και σας ευχαριστώ που είστε δίπλα μου όλα αυτά τα χρόνια! Ζω για ένα reunion!

Δε θα έφτανα ποτέ εδώ όμως χωρίς τις θυσίες και την υποστήριξη της οικογένειας μου. Μπαμπά, μαμά, Πέπη, Σουλίκα, Ντίνο, Θείο, Γιάννη, Γιαγιά...είσαστε ο κόσμος μου όλος και το καταφύγιο μου! Σας ευχαριστώ για ΟΛΑ! Κάνατε τα πάντα βάζοντας ο καθένας το λιθαράκι του για να έχω αυτή την πορεία και ξέρουμε όλοι πόσο δύσκολο ήταν αυτό.

Μπαμπά, μαμά, Πέπη...τα λόγια δεν φτάνουν για να εκφράσω την ευγνωμοσύνη μου. Δε σταματήσατε στιγμή να πιστεύετε σε μένα και στις δυνατότητες μου και να μου το υπενθυμίζετε! Σας αγαπώ πολύ (ξέρω δε το λέω συχνά) και σας ευχαριστώ για όλα από τα βάθη της καρδιάς μου!

Παναγιώτη, είσαι ό,τι καλύτερο συνέβη στη ζωή μου τα τελευταία χρόνια. Ραγίζει η καρδιά μου για τις στιγμές που χάνω από τη ζωή σου βλέπωντας σε να μεγαλώνεις μέσα από μια οθόνη. Εύχομαι κάποια στιγμή να επανορθώσω! Όταν μάθεις να διαβάζεις, ελπίζω αυτό το βιβλίο να πέσει στα χέρια σου. Σ'αγαπώ πολύ... η νονά σου.

About the author

Asimenia (Menia) Gavriilidou was born on the 6th of July 1990 in Thessaloniki, Greece. She studied Biology at the Faculty of Sciences at Aristotle University of Thessaloniki, Greece. During her BSc thesis, she studied sponges as ecosystem engineers in submarine caves of the North Aegean Sea. After obtaining her BSc degree in 2013, she continued her studies on Environmental Biology at the University of Crete, Greece where she was the top ranked applicant admitted to the MSc program. Menia specialized in Marine Resources Management and developed a strong interest for Marine Microbiology. Her MSc thesis research focused on the impact of acidification on marine bacteria isolated from the CO_2-venting submarine volcano Kolumbo. This research was conducted at the Environmental Chemistry and Bioanalysis Laboratory at the Institute of Marine Biology, Biotechnology and Aquaculture of the Hellenic Centre of Marine Research in Crete, Greece. After successfully completing the MSc requirements, she graduated in October 2016 and started applying for PhD positions abroad.

In April 2017, Menia was appointed as a PhD fellow in the Molecular Ecology group at the Laboratory of Microbiology at Wageningen University and Research (WUR) under the guidance of Dr. Detmer Sipkema (WUR), Prof. Dr. Hauke Smidt (WUR) and Dr. Colin Ingham (Hoekmine BV). The research work described in this thesis was conducted in the framework of the Marie Skłodowska-Curie Innovative Training Network “MarPipe” and the H2020 EU project “SponGES”. Currently, Menia is employed as a postdoctoral researcher in the Molecular Ecology group at the Laboratory of Microbiology at Wageningen University and Research. Her research focus is on rumen microbiomes and their implication on animal health and their environmental footprint.

List of publications

Tsiola A, Toncelli C, Fodelianakis S, Michoud G, Bucheli TD, **Gavriilidou A**, Kagiorgi M, Kalantzi I, Knauer K, Kotoulas G, Mylona K, Papadopoulou E, Psarra S, Santi I, Tsapakis M, Daffonchio D, Pergantis SA and Pitta P. Low-dose addition of silver nanoparticles stresses marine plankton communities. *Environ. Sci.: Nano*, 2018, 5, 1965-1980. https://doi.org/10.1039/C8EN00195B

Mandalakis M, **Gavriilidou A**, Polymenakou PN, Christakis CA, Nomikou P, Medvecký M, Kilias SP, Kentouri M, Kotoulas G and Magoulas A. Microbial strains isolated from CO2-venting Kolumbo submarine volcano show enhanced co-tolerance to acidity and antibiotics. Mar. Environ. Res. 2019; 144. https://doi.org/10.1016/j.marenvres.2019.01.002

Gavriilidou A, Gutleben J, Versluis D, Forgiarini F, Van Passel MWJ, Ingham CJ, Smidt H and Sipkema D. Comparative genomic analysis of *Flavobacteriaceae*: insights into carbohydrate metabolism, gliding motility and secondary metabolite biosynthesis. *BMC Genomics*. 2020; 21, 569. https://doi.org/10.1186/s12864-020-06971-7

Gavriilidou A, Mackenzie TA, Sánchez P, Tormo JR, Ingham CJ, Smidt H and Sipkema D. Bioactivity Screening and Gene-Trait Matching across Marine Sponge-Associated Bacteria. *Mar. Drugs*. 2021; 19(2):75. https://doi.org/10.3390/md19020075

Busch K, Slaby BM, Bach W, Boetius A, Clefsen I, Colaço A, Creemers M, Cristobo J, Federwisch L, Franke A, **Gavriilidou A**, Hethke A, Kenchington E, Mienis F, Mills S, Riesgo A, Ríos P, Roberts EM, Sipkema D, Pita L, Schupp PJ, Xavier J, Rapp HP, Hentschel U. Biodiversity, environmental drivers, and sustainability of the global deep-sea sponge microbiome. *Manuscript under review*.

Peters EE, Cahn JKB, **Gavriilidou A**, Steffens UAE, Loureiro C, Shorn MA, Cárdenas P, Vickneswaran N, Crews P, Sipkema D, Piel J. Distribution and diversity of *Candidatus* Tectomicrobia, a deep-branching uncultivated bacterial lineage harboring rich producers of bioactive metabolites. *Manuscript submitted*.

Loureiro C, Galani A, **Gavriilidou A**, De Mares MC, Van der Oost J, Medema MH, Sipkema D. Comparative metagenomic analysis of biosynthetic diversity across sponge microbiomes highlights metabolic novelty, conservation and diversification. *Manuscript submitted*.

Gavriilidou A, Avci B, Galani A, Schorn M, Loureiro C, Ingham CJ, Smidt H, Sipkema D. *Candidatus* Nemesobacterales, a sponge-specific clade of the candidate phylum Desulfobacterota adapted to a symbiotic lifestyle. *Manuscript in preparation*.

Netherlands Research School for the
Socio-Economic and Natural Sciences of the Environment

D I P L O M A

for specialised PhD training

The Netherlands research school for the
Socio-Economic and Natural Sciences of the Environment
(SENSE) declares that

Asimenia Gavriilidou

born on 6th July 1990 in Thessaloniki, Greece

has successfully fulfilled all requirements of the
educational PhD programme of SENSE.

Wageningen, 25th of May 2022

Chair of the SENSE board

Prof. dr. Martin Wassen

The SENSE Director

Prof. Philipp Pattberg

The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW)

KONINKLIJKE NEDERLANDSE
AKADEMIE VAN WETENSCHAPPEN

The SENSE Research School declares that **Asimenia Gavriilidou** has successfully fulfilled all requirements of the educational PhD programme of SENSE with a work load of 50.0 EC, including the following activities:

SENSE PhD Courses

- Environmental research in context (2017)
- Research in context activity: 'Organising the General Assembly Meeting of the SponGES H2020 Project' (2018)
- Principles of Ecological Genomics (2018)

Other PhD and Advanced MSc Courses

- MarPipe Course: Biodiscovery with zebrafish, KU Leuven (2017)
- Open & Reproducible Microbiome Data Analysis, Microbiology group, (2018)
- MarPipe Course: Scientific communication and grant writing skills, Consiglio Nazionale delle Ricerche (2018)
- MarPipe Course: Organic structure analysis, Fundación MEDINA (2018)
- MarPipe Course: Bioinformatic tools for new natural products discovery, University of Cork (2019)
- Bioinformatics with Linux and Python, Microbiology group and WIMEK (2020)

External training at a foreign research institute

- MarPipe Secondment: Legal skills, patenting, regulatory affairs and entrepreneurship, ABSint (2019)
- MarPipe Secondment: Bioactivity screening methods, Fundación MEDINA (2020)

Management and Didactic Skills Training

- Supervising two MSc students with thesis (2017-2018)
- Teaching in the BSc course 'Microbiology & Biochemistry' (2017)
- Teaching in the MSc course 'Microbial Ecology' (2017-2021)

Oral Presentations

- *Mining metagenomes of marine sponge Geodia barretti for biosynthetic gene clusters*. Genomics for a Blue Economy, 11-12 December 2019, Naples, Italy

SENSE coordinator PhD education

Dr. ir. Peter Vermeulen

The research described in this thesis was financially supported by the Marie Skłodowska-Curie Actions (MSCA) Innovative Training Network (ITN) Horizon 2020 H2020-MSCA-ITN-2016 “MarPipe” (Grant agreement ID: 721421) and the European Union’s H2020 research and innovation programme “SponGES” (Grant agreement ID: 679849).

Financial support from Wageningen University for printing this thesis is gratefully acknowledged.

Cover idea: Asimenia Gavriilidou & Sara Issa
Cover design: Rebeca Lucía Arredondo Sáinz
Printed by: ProefschriftMaken.nl