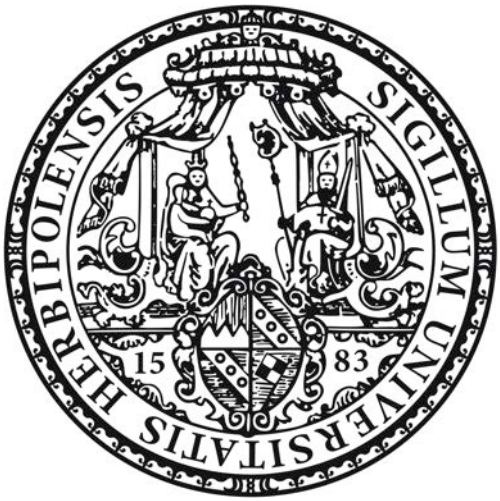
Single-Cell Genomics of the Candidate Phylum Poribacteria



Einzelzell-genomische Analysen des Candidatus Phylums Poribacteria

Dissertation zur Erlangung der Naturwissenschaftlichen Doktorgrades der Julius-Maximilians-Universität Würzburg

vorgelegt von
Janine Kamke
aus
Essen

Würzburg, November 2013

Eingereicht am:

Mitglieder der Prüfungskommission:

Vorsitzender: Prof. Dr. Markus Engstler

1. Gutachter: Prof. Dr. Ute Hentschel Humeida

2. Gutachter: Prof. Dr. Roy Gross

Tag des Promotionskolloquiums:

Doktorurkunde ausgehändigt am:

Eidesstattliche Erklärung

gemäß § 4 Abs. 3 Ziff. 3, 5 und 8
der Promotionsordnung der Fakultät für Biologie der Julius-Maximilians-Universität Würzburg.

Hiermit erkläre ich an Eides statt, die Dissertation „Single-cell genomics of the candidate phylum *Poribacteria*“ eigenständig, d.h. insbesondere selbstständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben. Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Würzburg, November 2013

Janine Kamke

Danksagung

Mein herzlicher Dank richtet sich an folgende Personen:

Prof. Ute Hentschel Humeida, für die Betreuung meiner Doktorarbeit, die vielen damit verbundenen Möglichkeiten und die vielseitige Unterstützung.

Prof. Roy Gross für die freundliche Übernahme des Zweitgutachtens.

Prof. Markus Riederer als Institutsleiter des Lehrstuhls für Botanik II – Ökophysiologie und Vegetationsökologie an der Universität Würzburg für die Bereitstellung des Arbeitsplatzes.

Prof. Thomas Dandekar für die freundliche Beratung.

Meine Kolleginnen und Kollegen in Büro und Labor für die schöne gemeinsame Zeit und die gegenseitige Unterstützung. Besonders Kristina, Volker, Eva, Hilde, Tine, Alex, Susanne, Lucas, Usama, Sheila, Paula und Anne.

Tanja Woyke und ihr Team vom JGI für die fachliche Unterstützung und dass ihr mich so herzlich aufgenommen habt: Chris, Patrick, Alex, Kostas, Natalia, Janie, Shaomei, Julien, Jeff und Sheryl.

Das Team vom Ruder Boskovic Institut in Rovinj für die Bereitstellung ihrer Labore und der Unterstützung beim Probensammeln.

Christian Linden (Universität Würzburg) für die Unterstützung bei Zellsortierungen.

Jörn Piel (ETH Zürich) für die freundliche Zusammenarbeit.

Meinen Eltern dafür, dass sie mir immer ermöglicht haben meinen Weg zu gehen und meine Großeltern für ihre Unterstützung.

Meiner Familie und meinen Freunden.

Und Chris für alles.

Table of contents

Summary

Marine sponges are the most ancient metazoans and of large ecological importance as drivers of water and nutrient flows in benthic habitats. Furthermore marine sponges are well known for their association with highly abundant and diverse microbial consortia. Microorganisms inhabit the extracellular matrix of marine sponges where they can make up to 35% of the sponge's biomass. Many microbial symbionts of marine sponges are highly host specific and cannot, or only in very rare abundances, be found outside of their host environment. Of special interest is the candidate phylum *Poribacteria* that was first discovered in marine sponges and still remains almost exclusive to their hosts. Phylogenetically *Poribacteria* were placed into the *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* superphylum and similarly to many members of this superphylum cell compartmentation has been proposed to occur in members of the *Poribacteria*. The status as a candidate phylum implies that no member of *Poribacteria* has been obtained in culture yet. This restricts the investigations of *Poribacteria* and their interactions with marine sponges to culture independent methods and makes functional characterisation a difficult task.

In this PhD thesis I used the novel method of single-cell genomics to investigate the genomic potential of the candidate phylum *Poribacteria*. Single-cell genomics enables whole genome sequencing of uncultivated microorganisms by singularising cells from the environment, subsequent cell lysis and multiple displacement amplification of the total genomic DNA. This process yields sufficient amounts of DNA for whole genome sequencing and genome analysis. This technique and its relevance for symbiosis studies are discussed in this PhD thesis.

Through the application of single-cell genomics it was possible to increase the number of single-amplified genomes of the candidate phylum *Poribacteria* from initially one to a total of six. Analyses of these datasets made it possible to enhance our understanding of the metabolism, taxonomy, and phylum diversity of *Poribacteria* and thus made these one of the best-characterised sponge symbionts today. The poribacterial genomes represented three phylotypes within the candidate phylum of which one appeared

dominant. Phylogenetic and phylogenomic analyses revealed a novel phylogenetic positioning of *Poribacteria* distinctly outside of the *Planctomycete*, *Verrucomicorbia*, *Chlamydiae* superphylum. The occurrence of cell compartmentation in *Poribacteria* was also revisited based on the obtained genome sequences and revealed evidence for bacterial microcompartments instead of the previously suggested nucleotide-like structures. An extensive genomic repertoire of glycoside hydrolases, glycotransferases, and other carbohydrate active enzymes was found to be the central shared feature between all poribacterial genomes and showed that *Poribacteria* are among those marine bacteria with the largest genomic repertoire for carbohydrate degradation. Detailed analysis of the carbohydrate metabolism revealed that *Poribacteria* have the genomic potential for degradation of a variety of polymers, di- and monosaccharaides that allow these symbionts to feed various nutrient sources accessible through the filter-feeding activities of the sponge host. Furthermore the poribacterial glycobiome appeared to enable degradation of glycosaminoglycan chains, one of the main building blocks of extracellular matrix of marine sponges. Different lifestyles resulting from the poribacterial carbohydrate degradation potential are discussed including the influence of nutrient cycling in sponges, nutrient recycling and scavenging. The findings of this thesis emphasise the long overlooked importance of heterotrophic symbionts such as *Poribacteria* for the interactions with marine sponges and represent a solid basis for future studies of the influence heterotrophic symbionts have on their sponge hosts.

Zusammenfassung

Marine Schwämme sind die ältesten rezenten Vertreter der Metazoen. Durch ihre Lebensweise als Nahrungsfiltrierer und den damit verbundenen Einfluss auf Nährstoffzyklen sind sie von großer ökologischer Relevanz. Des Weiteren zeichnen sich marine Schwämme durch das Zusammenleben mit hoch abundanten und diversen mikrobiellen Konsortien aus. Diese Mikroorganismen finden sich meist in der extrazellulären Matrix des Schwamms und können mehr als 35% der Biomasse ihres Wirtes ausmachen. Viele mikrobielle Symbionten mariner Schwämme sind hochgradig Wirts-spezifisch und können außerhalb des Schwamms, wenn überhaupt, nur in sehr geringer Anzahl gefunden werden. Von besonderem Interesse ist das Candidatus Phylum *Poribacteria*, dessen Vertreter erstmals in Schwämmen detektiert wurden, und bis heute fast ausschließlich in Schwämmen zu finden sind. Phylogenetisch wurden die *Poribacteria* dem *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* (PVC) Superphylum zugeordnet. Einige Vertreter dieses Superphylums zeigen einen kompartimentierten Zellplan auf, eine Eigenschaft, die auch für *Poribacteria* vermutet wird. Der Status der *Poribacteria* als Candidatus Phylum zeigt das Fehlen von Vertretern dieses Phylums in Reinkultur an. Dies beschränkt die Untersuchung von *Poribacteria* auf kultivierungs-unabhängige Methoden, was die funktionelle Charakterisierung dieser Symbionten erheblich erschwert.

In dieser Doktorarbeit wurde das Candidatus Phylum *Poribacteria* mit Hilfe der Einzelzellgenomik untersucht. Diese Methode ermöglicht es aus vereinzelten mikrobiellen Zellen genomische Komplett-DNA zu gewinnen und diese, mit Hilfe der so genannten „multiple displacement amplification“ so hochgradig anzureichern, dass eine Sequenzierung und anschließende Analyse erfolgen kann. Die Anwendung dieser Methode im Allgemeinen und in der Symbiose Forschung wird in dieser Doktorarbeit diskutiert.

Die Einzelzellgenomik ermöglichte die Anzahl poribakterieller Datensätze, von zunächst einem auf sechs Genome zu erhöhen. Die Analyse dieser Genome konnte unser Verständnis vom metabolischen Potential, der Taxonomie und der Diversität innerhalb

dieses Phylums deutlich zu verbessern. Die poribakteriellen Genome beschrieben drei Phylotypen, von denen einer deutlich dominierte. Die phylogenetische Position des Phylums *Poribacteria* wurde außerdem anhand von phylogenetischen und phylogenomischen Berechnungen neu zugeordnet, und resultierte in einer deutlichen Positionierung außerhalb des PVC Superphylums. Weiterhin wurden genomische Hinweise auf einen kompartimentierten Zellplan in *Poribacteria* gefunden. Diese deuten aber nicht, wie vorher vermutet, auf eine Zellkern-ähnliche Struktur hin, sondern auf bakterielle Mikrokompartimente mit noch ungeklärter Funktion. Die Analysen des genomischen Potentials zeigte in allen Datensätzen eine hohe Frequenz von Genen, die für Glycosidasen, Glycosyltransferasen und weiteren Proteinen der so-genannten „carbohydrate active enzymes“ kodieren, was ein ausgeprägtes Vermögen zum Kohlehydratabbau aufzeigt. Das genomische Potential von *Poribacteria* zum Abbau von Kohlehydratpolymeren, Di- und Monosacchariden konnte durch detaillierte Analyse des Kohlehydratmetabolismus genau beschrieben werden. Außerdem schienen *Poribacteria* Glycosaminoglycanketten, die zentrale Bausteine der extrazellulären Matrix des Schwammes sind, abbauen zu können. Die aus dem poribakteriellen Glycobiome resultierenden möglichen Lebensweisen als Wiederverwerter von Nährstoffen, Mitesser, oder auch der Einfluss von *Poribacteria* auf Nährstoffzyklen im Schwamm werden in dieser Doktorarbeit diskutiert. Die Ergebnisse dieser Doktorarbeit machen *Poribacteria* zu den genomische am besten beschriebenen Schwammsymbionten und zeigen die lange übersehene Relevanz heterotropher Symbionten in marinen Schwämmen.

1. General Introduction

1.1. Sponges

Sponges (phylum *Porifera*) are the evolutionary most ancient living metazoans with a fossil record proving their existence at least 580 million years ago (Li et al. 1998). The molecular biomarker 24-isopropylcholestance indicates the existence of sponges back in the Cryogenian period more than 635 million years ago (Love et al. 2008) which is also supported by molecular clock estimates (Erwin et al. 2011). The evolutionary success of sponges is reflected in their widespread occurrence over diverse aquatic habitats. Sponges can be found in freshwater lakes and rivers but most species are found in marine habitats. These occur around the world in from shallow waters to the deep sea as well as from polar regions to tropical reefs (Van Soest et al. 2012). Overall 8,500 valid species descriptions are known and more than double this number of species are suspected to exist (Van Soest et al. 2012). This diversity is also shown in the variety of physical appearances (Fig. 1). Sizes range from extremely large species such as *Xestospongia muta*, often more than one metre in height (McMurray et al. 2008) to millimetre sized sponges encrusting on surfaces of rock or other organisms. An array of colours is known from plain greys to bright reds and yellow, and shapes from compact “ball” structures to largely extended chimney colonies.

1.1.1. Sponge taxonomy

Due to the evolutionary early emergence of sponges, the taxonomy of the *Porifera* is of large importance for the reconstruction of early metazoan evolution. Morphological aspects of body structure as well as spiculae and sponge embryology play a role in classification, while molecular markers become more and more important. The amount and selection of such marker genes lead to different theories of a potential para- or monophyletic origin of *Porifera* and is still controversially discussed (for review see Wörheide et al. 2012; Dohrmann & Wörheide 2013). Within the phylum *Porifera* four extant classes can be found: *Calcarea*, *Demospongiae*, *Hexatinellida*, and *Homoscleromorpha*. The latter was previously regarded as a subclass of the *Demospongiae* and was only recently revealed as a distinct class within the *Porifera* by

molecular investigations (Gazave et al. 2011). Despite this new classification the *Demospongiae* still comprise about 83% of all sponge species (Van Soest et al. 2012).

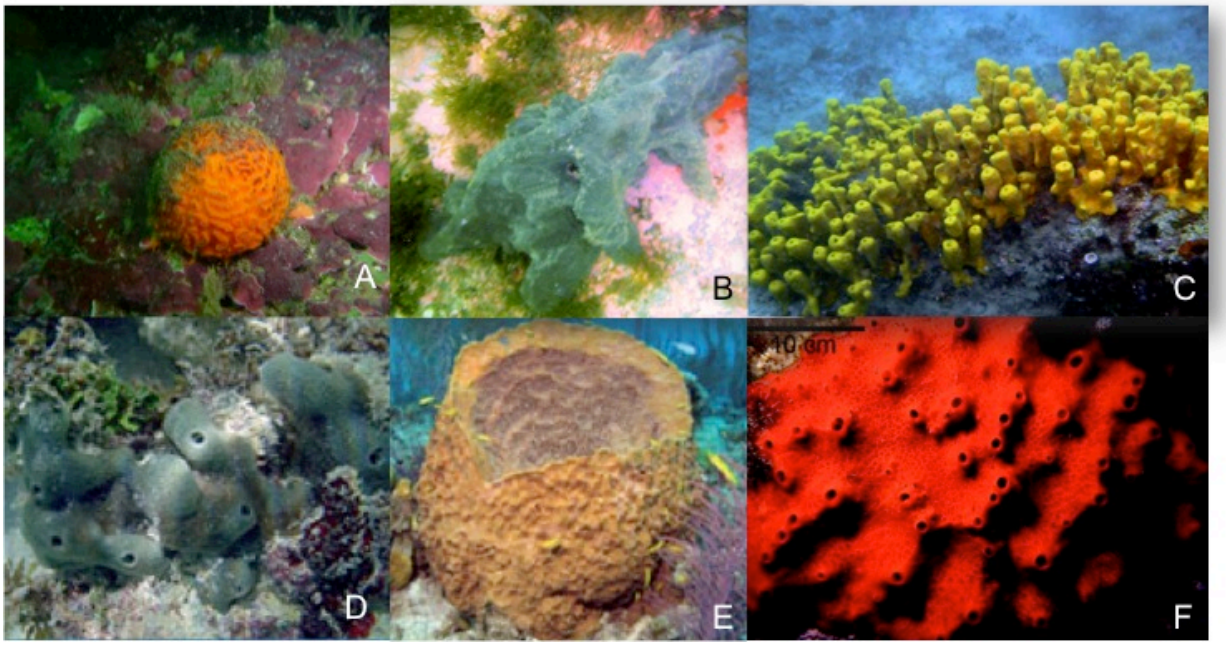


Figure 1: Examples of different sponge species: *Tethya* sp. with an undefined purple encrusting sponge in the background. Photo: J.Kamke (A), *Ancorina alta* Photo: J.Kamke (B), *Aplysina aerophob*a PhotoJ. Kamke, modified with permission from Webster & Taylor (2011) (C), *Amphimedon queenslandica* Photo modified with permission from Hentschel et al. (2012) (D), *Xestospongia muta* Photo modified with permission from Hentschel et al. (2012), and *Tedania digitata* Photo modified with permission from Taylor et al. (2007) (F).

1.1.2. Sponge morphology and mode of life

Sponges are sessile animals that were first to evolve suspension feeding (Ruppert et al. 2004) and their body plan is, although very simple, highly adapted for this process (Hooper & van Soest 2002) (Fig. 2). Sponges lack true tissues or organs as can be found in higher metazoans (Bergquist 1978). An aquiferous system branches through the sponge body which transfers water from small entering pores, so called ostia, through the sponge body into one or several main channels and releases the water through the osculum, the final opening. The outer cell layer (pinacoderm) that separates the inside of the sponge body from the outside is built by so-called pinacocytes. It can be found on the exterior of the sponge and along the aquiferous channels, which lead into chambers of flagellated cells the choanocytes that build the choanoderm. These cells not only drive the water flow through the sponge but also filter out micro-particles from seawater and release them into the mesohyl, an extracellular matrix separating pinacoderm and choanoderm (Fig. 2). Amoeboid, totipotent cells (termed archaeocytes)

in the mesohyl take up these particles by phagocytosis as a food source. The mesohyl is a mostly inorganic network of collagenous fibres (spongin) and siliceous or calcareous spiculae that build a stabilising skeleton. Next to sponge cells that produce these substances (collencytes, sclerocytes, and spongocytes) the mesohyl can also contain high densities of microorganisms permanently inhabiting this extracellular space (section 1.2).

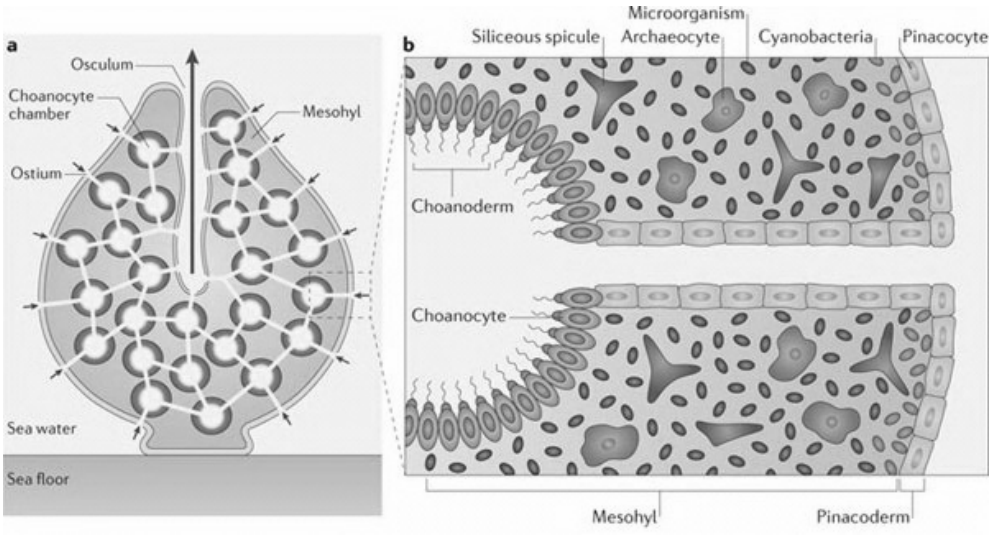


Figure 2: Schematic drawing of the sponge body plan (a) and detailed (b). Figure changed with permission from Hentschel et al. (2012).

During the filtration processes particles smaller than 10 µm are removed from surpassing water (Pile et al. 1996; Reiswig 1971), which is subsequently released almost sterile (Wehrl et al. 2007; Ribes et al. 1999). This efficiency is paired with an enormous filtering capacity, in fact, thousands of litres of water can pass through a sponge per day (Vogel 1977; Bell 2008) and many sponges can filter a volume of water equal to their own body weight within five seconds (Ruppert et al. 2004). Accordingly, sponges have, due to their filtering performance, a large influence on ecological processes in their environments, such as the bentho-pelagic coupling (Bell 2008). It has been shown that sponges can influence trophical interactions and nutrient cycles such as carbon, nitrogen, silicon, and phosphorous (de Goeij et al. 2008; Ribes et al. 2012; Maldonado et al. 2011; Perea-Blázquez et al. 2012; reviewed by Maldonado et al. 2012). The ecological importance of sponges is further enhanced by their abundance in many benthic habitats where they are often dominant species (Hooper & van Soest 2002). Especially in tropical reef ecosystems sponges are largely abundant and diverse (Diaz & Rützler 2001) and it is expected that sponges will in the future play an even

greater role as major reef builders due to the influence of global change on tropical reef systems (Bell et al. 2013).

The sessile lifestyle of sponges resulted in a range of different defence systems that sponges use against predators and pathogens or other “invading” organisms and space competitors. Spiculae serve not only as a stabilising substance to the sponge but are also used as mechanical defence. The well adapted innate immune system of sponges further protects against invading parasites and pathogens (Wiens et al. 2007; Müller & Müller 2003). Moreover, sponges are known as producers of rich sets of bioactive compounds which they not only use as a chemical defence (Pawlik 2011) but which also makes them of high interest from a biotechnological and biomedical point of view (Blunt et al. 2013).

1.2. Sponge microbiology

Sponges are known to be associated with highly abundant and diverse microbial consortia from all three domains of life (*Archaea*, *Bacteria*, and *Eukarya*). These include diatoms, dinoflagellates, macroalgae, and fungi but mostly bacteria and archaea (Taylor et al. 2007), on which I will focus here. The largest amounts of bacterial and archaeal cells can be found extracellular in the mesohyl of many sponge species but also intracellular and even internuclear microorganisms have been described from sponges (Vacelet 1975; Wilkinson 1978b; Friedrich et al. 1999). Various inter- and intraspecific interactions have been reported (reviewd by Taylor et al. 2007). Many microorganisms serve the sponge as mere food source but also commensalistic, mutualistic, and parasitic interactions have been described. For the purpose of this thesis the term “symbiosis” (and “symbiont”) is used to summarise these interactions in its broadest possible definition describing the close association of different phyla or species (Lewin 1982).

1.2.1. Low and high microbial abundance sponges

In early electron microscopy studies of microbial communities in sponges it became apparent that there is a difference between the sponge species and the amount of microorganisms that can be found in their mesohyl (Vacelet & Donadey 1977; Wilkinson 1978a; 1978c). On the one hand, some species show dense amounts of morphologically diverse microorganisms that can make up to 35% of the sponge’s body mass (Vacelet

1975). These species were called “bacteriosponges”. On the other hand, the mesohyl of other sponge species appears largely free of microbial cells. Later sponges were classified into two groups, low (LMA) and high microbial abundance (HMA) sponges (Hentschel et al. 2003). The mesohyl of LMA sponges harbours approximately 10^5-10^6 microbial cells per gram sponge wet weight (Fig. 3), which is within the range of microbial cell density in seawater (Hentschel et al. 2006). In HMA sponges microbial numbers were found to be three to four orders of magnitude higher with up to 10^{10} cells per gram sponge wet weight (Fig.3) (Hentschel et al. 2006). These differences are not only restricted to microbial numbers but also community composition, as microbial communities in LMA sponges appear less diverse and more similar to the communities in seawater (Hentschel et al. 2006; Giles et al. 2012). This was also reflected by the clone libraries of the two sponges *Ancorina alata* (HMA) and *Polymastia* sp. (LMA), where the HMA sponge showed higher diversity and richness of phyla (Kamke et al. 2010). A recent pyrosequencing study also revealed a stronger similarity of the microbial community of the LMA sponge *Stylissa carteri* to seawater than that of the HMA sponge *Xestospongia testudinaria* but also showed that there is an overlap of taxa between both sponge species and the seawater community (Moitinho-Silva et al. 2013).

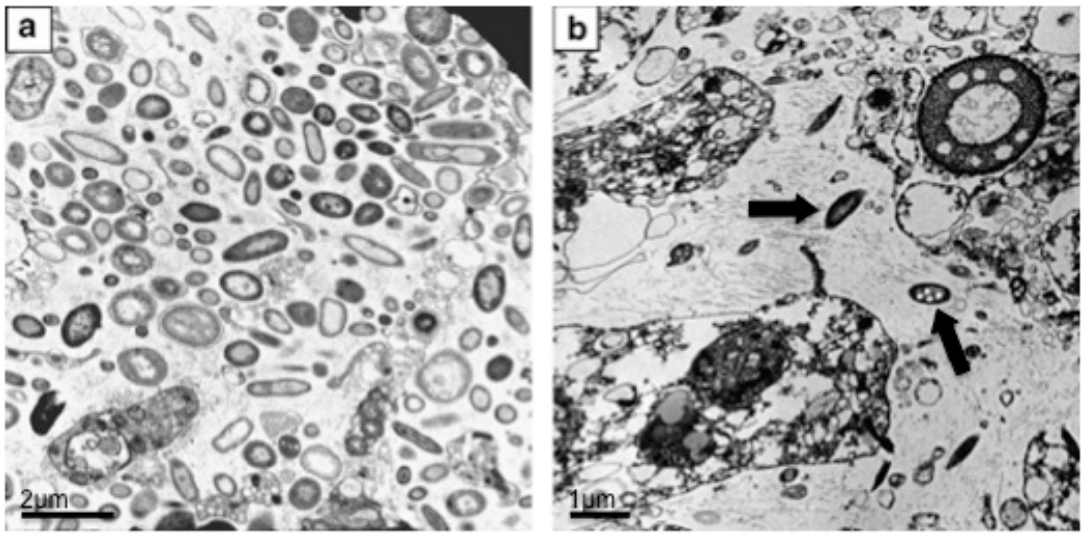


Figure 3: Transmission electron microscopy pictures of the mesohyl of two marine sponge species. The mesohyl of the HMA sponge *Ancorina alata* shows a dense microbial community (a) while in the mesohyl of the LMA sponge *Polymastia* sp. shows only very few microbial cells (indicated by arrows) (b). Picture changed with permission from Kamke et al. (2010).

There has only been little indication as to why some sponge species contain larger amounts of microbes in their mesohyl. An apparent difference between HMA and LMA sponges is expansion of the mesohyl itself, where HMA sponges show a much more

extracellular matrix while the aquiferous system is narrower than in LMA sponges (with comparably small extracellular matrix and an expansive canal system) (Wilkinson 1978a; Weisz et al. 2008; Vacelet & Donadey 1977). This has been connected to slower pumping rates and the need to acquire more dissolved organic matter in HMA sponges compared to LMA sponges (Reiswig 1974; Weisz et al. 2008; Maldonado et al. 2012).

1.2.2. Diversity, specificity, and stability of the sponge microbiome

The diversity of sponge-associated microorganisms is exceptionally high. Today at least 28 bacterial phyla as well as *Thaumarchaeota*, *Euryarchaeota* and eukaryotic microbial caldes have been described from sponges based on cultivation and molecular cloning studies (Hentschel et al. 2012) (Fig. 4). With the emergence of next generation sequencing technologies (amplicon sequencing of the 16S rRNA gene) it also became possible to access rare members of the microbial communities in sponges. This further adds to the diversity and results in more than 30 bacterial phyla (e.g. Webster et al. 2010; Schmitt et al. 2012; Moitinho-Silva et al. 2013; Lee et al. 2010). Among the dominant bacterial phyla in sponges are *Proteobacteria* (*Alpha*-, *Gamma*-, and *Deltaproteobacteria*), *Chloroflexi*, *Actinobacteria*, *Acidobacteria*, *Nitrospirae* and the candidate phylum *Poribacteria* (section 1.2.5) (Hentschel et al. 2012). Apart from *Poribacteria* 12 other candidate phyla have been described from sponges (BD1-5, BRC1, CAB-I, OD1, OP1, OP10, OP11, OP3, OS-K, SBR1093, TM6, and TM7) (Schmitt et al. 2012; 2011; Hentschel et al. 2012). Candidate phyla are of special interest because they contain exclusively uncultured and formally not described organisms that add not only to the phylogenetic diversity but the functional genomic repertoire of the sponge microbiome.

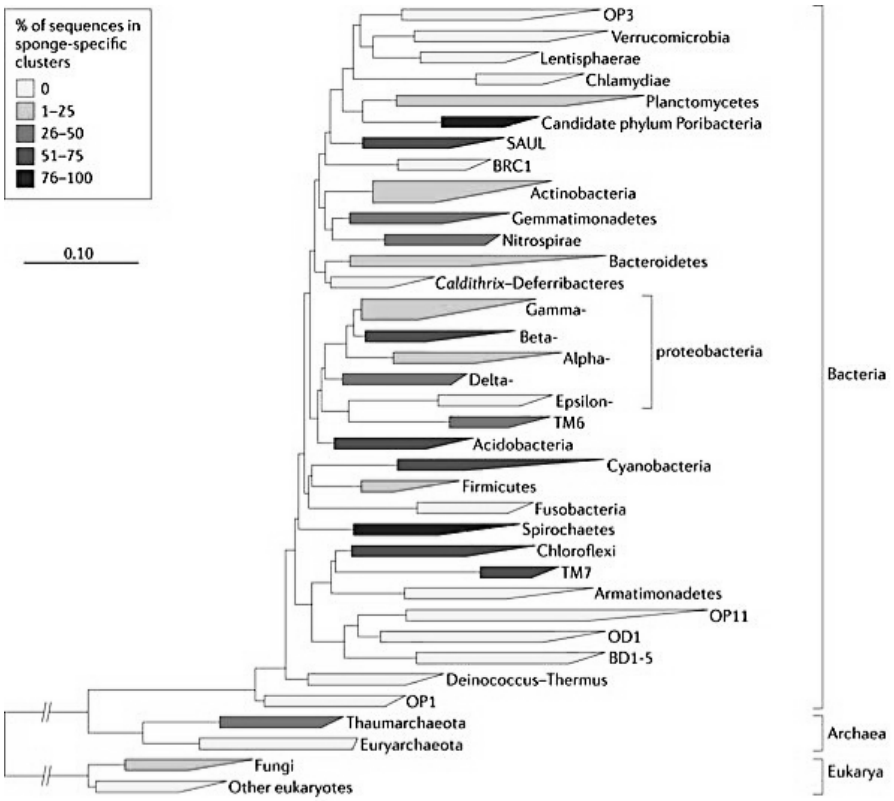


Figure 4: Phylogenetic 16S and 18S rRNA gene based tree showing all bacterial, archaeal and eukaryotic phyla associated with sponges based on Simister et al. (2011) and the degree of sequences falling into sponge specific clusters for each phylum. Picture changed with permission from Hentschel et al. (2012).

The microbial community of sponges is not only extremely diverse but also shows a composition that appears to be specific for these associations. Hentschel et al (2002) originally described a uniform microbial community that was commonly represented by all sponge derived 16S rRNA gene sequences available at the time. This community was shared between different sponge species and sponges from distant geographic locations but was distinctly different from samples from any other environment. Using phylogenetic analysis of the 16S rRNA gene it was shown that many of the bacterial phyla found fall into so-called “sponge-specific clusters” which were defined as clusters of (I) at least three sponge derived sequences that are more closely related to each other than to any sequence from a non-sponge source, (II) contain at least three sequences from different sponge species or from geographically distant locations, and (III) cluster together independently of the treeing method used (Hentschel et al. 2002). Five years later Taylor et al. (2007) showed in a comprehensive review that out of 1700 sponge derived 16S rRNA gene sequences that 32% fell in to monophyletic sponge specific clusters. The existence of these clusters was recently revisited with more than

7500 16S and 18S rRNA gene sponge- derived sequences from bacteria, archaeal and fungi (Simister et al. 2011). The study revealed the presence of sponge specific clusters in 14 bacterial phyla with the most prominent clusters in *Cyanobacteria, Chloroflexi*, *Poribacteria*, *Betaproteobacteria*, and *Acidobacteria* (Fig. 4).

In contrast, 16S rRNA gene amplicon sequencing studies showed the occurrence of seemingly sponge specific microbes in low abundances in other sources such as seawater, sediments, or other marine invertebrates (Taylor et al. 2013; Webster et al. 2010). Furthermore, an analysis of the microbial communities of 32 sponge species from eight different locations around the world showed that on a species level (97% sequence similarity of the 16S rRNA gene) only a very limited number of bacterial operational taxonomic units (OTUs) are present in the majority of all sponge species (Schmitt et al. 2012). This so defined “core community” is seemingly small with only three OTUs represented in more than 70% of all sponge species. In contrast, the majority of all sponge derived OTUs appears specific to a certain sponge species. The authors however observed that the microbial communities, although sponge species, specific are more similar to communities from other sponges than from other environments (Schmitt et al. 2012). Thus, sponges show a distinct microbial community signature that is internally species specific and highly abundant in the host but rare in other environments. Species specificity has also been shown before in three sponge species from Australia, which showed only minimal variability between different specimens of the same species (Taylor et al. 2004b). Similarly, in different species of the sponge genus *Dysidea* specific cyanobacterial symbionts have been detected (Thacker & Starnes 2003). Species-specific microbial communities of sponges have also proven stable over geographic distances. In five sponge species from Antarctica stable communities were found over a radius of ten kilometres (Webster et al. 2004). Taylor et al. (2005) could further show that in the sponge *Cymbastela concentrica* microbial communities remain stable over a 500 km distance. Stability of sponge associated microbial communities on a host specific level was also shown over varying periods of time. The microbial community of the Mediterranean sponge *Aplysina aerophoba* remained stable in captivity during an eleven day experiment (Friedrich et al. 2001). Also over longer periods of time have host specific communities proved to be stable as shown in the sponges *Cymbastela concentrica, Callysponiga* sp., and *Stylinos* sp. that were monitored over days, weeks, and lastly five seasons (Taylor et al. 2004b). Further, Simister et al. (2013) showed using 16S rRNA gene amplicons sequencing techniques

that over a two year sampling survey the two sponges *Ancorina alata* and *Tethya stolonifera* harboured largely stable communities and that variability was only detectable in low abundance OTUs. Finally, sponge associated communities appear to be largely stable even when challenged by differing external influences. Captivity led only to minor variation in the cold water sponge *Geodia barretti* (Hoffmann et al. 2006) and the microbial community of *A. aerophoba* remained stable even in filtered circulating seawater and under treatment with antibiotics (Friedrich et al. 2001). The microbial community of the Great Barrier Reef sponge *Rhopaloeides odorabile* showed to be stable over a period of seven days under elevated nutrient and temperature levels of up to 31°C (Simister et al. 2012b). However in a similar study it was shown that temperatures of 32°C were lethal for *R. odorabile* after 3 days and that the microbial community changed in the necrotic tissues (Simister et al. 2012a). The authors concluded that the host and not the microbial community was temperature sensitive, since microbial communities remained stable in both abundance and activity in healthy tissue parts of the sponge at 32°C. Further studies confirmed stable communities in *R. odorabile* at elevated but sublethal temperatures and a community shift at lethal temperatures (Fan et al. 2013). But it was also shown that the protein expression pattern of the microbial community shifted already at sublethal temperatures (Fan et al. 2013). Whether elevated temperatures influenced the microbial community itself or a change in host function caused the shift is a subject for further studies.

1.2.3. Symbiont acquisition

In the context of sponge specific microbial communities an important question is how these associations are initially established. Different scenarios for the acquisition of microbial communities have been discussed including vertical transmission of symbionts from the adult sponge over eggs and larval stages, environmental transmission through seawater and a combination of the above (Taylor et al. 2007; Schmitt et al. 2012; Webster & Taylor 2011). Vertical transmission is supported by several studies based on electron microscopy, denaturing gradient gel electrophoresis, (DGGE), 16S rRNA clone libraries and amplicon sequencing studies that show the occurrence of the same microbial taxa in adult, larval and embryonic stages (e.g. Schmitt et al. 2008; 2007; Webster et al. 2010; Enticknap et al. 2006; Steger et al. 2008; Gloeckner et al. 2012). This scenario supports the existence of “true symbionts” occurring only in their respective hosts and was suggested to occur for the small core

community detected in 32 sponge species (Schmitt et al. 2012). Coevolution might have occurred between sponges and their microbial symbionts in case of vertical transmission of symbionts over long periods of time, as shown by phylogenetic analyses between different sponge species and their symbionts (Thacker & Starnes 2003; Erpenbeck et al. 2002). At the same time the occurrence of sponge associated microbiota in seawater supports the second possibility of environmental transmission of microbial symbionts. The uptake of rare microbes from seawater which then grow to higher abundances in the sponge mesohyl would explain the occurrence of taxa in seawater that were previously thought to be sponge specific. Schmitt et al (2012) also considered this possibility most likely for the majority of species specific microbial communities found in sponges and thus favours a co-occurrence of vertical and environmental transmission. Until further studies reveal more details about the mode of acquisition of sponge symbionts the combination of vertical and horizontal transmission is considered the most likely one (Hentschel et al. 2012).

Besides research on the initial acquisition of sponge symbionts, the question how these associations can be maintained while located in the mesohyl, the same space where digestion of food bacteria takes place. How do microbial sponge symbionts avoid being digested by their host? It was shown that potential symbionts are not digested when taken up into the mesohyl, in contrast to non-symbiotic bacteria that were largely consumed by sponges (Wilkinson et al. 1984; Wehrl et al. 2007). The mechanisms behind this selective feeding of the sponges are not completely understood. One theory is that the sponge actively identifies the potential symbiont as “non-food” through an epidermal recognition mechanism, while another possibility seems that sponge symbionts posses protective extracellular structures such as enhanced membranes or slime capsules that shield against digestion (Wilkinson et al. 1984). Extended cell walls have indeed been reported from microbial symbionts of marine sponges (Fieseler et al. 2004) but a connection to prevention of phagocytosis by the sponge has not been shown. Much recent attention has been paid to so-called eukaryote like protein domains in respect to symbiont recognition. Repeat domain such as ankyrins, tetratricopeptide repeats, leucin rich repeats, fibronectin-like domains have been found highly abundant and also expressed in sponges (Fan et al. 2012; Thomas et al. 2010a; Liu et al. 2012). These repeat domains are known to mediate protein-protein interactions and have been suggested to aid in the prevention of phagocytosis by the host (Hentschel et al. 2012; Fan et al. 2012). Indeed a recent study showed that an ankyrin gene from sponge

symbionts when expressed in *E.coli* mediate uptake and prevent digestion by amoeba (Nguyen et al. 2013). The exact processes in the sponge environment are still to be determined but are likely to be complex. There is also the possibility of cell signalling being involved. Quorum sensing molecules in the form of acyl homoserine lactone (AHLs) and autoinducer- 2 (AI-2) molecules have been detected from sponge associated microorganism in culture and in vivo (Taylor et al. 2004a; Mohamed et al. 2008a; Zan et al. 2011; Gardères et al. 2012). The AHL system has been further investigated in the alphaproteobacterial sponge symbiont *Rugeria* sp. KLH11 from the sponge M*ycale laxissima*. The authors identified a regulatory system for flagellar biosynthesis and motility that is controlled by AHL quorum sensing molecules in this sponge symbiont (Zan et al. 2012; 2013). The switch between a sessile and motile lifestyle might be an important factor for bacteria penetrating the sponge and it is conceivable that also other microbe-microbe or even microbe-host communications are quorum sensing controlled (Hentschel et al. 2012).

1.2.4. Symbiont metabolism and function

Much effort has been put into investigation the functional relationships between sponges and their microbial symbionts. Yet, we are only beginning to understand how the different partners of this complex system interact and what the functional roles are. Functional characterisation of sponge-associated microorganisms is largely restricted to culture independent methods, due to the fact that the majority of all sponge associated microbes have not been retained in pure culture yet. However, metabolic activity of microbial communities in sponges appears high as indicated by RNA based studies (Kamke et al. 2010; Moitinho-Silva et al. 2013) and several functions of the primary and secondary metabolism of sponge symbionts have been investigated mainly by functional gene and “omics” approaches and physiological examinations of the sponge itself.

The filter feeding activities of sponges supply the animal itself with nutrients and provide a high nutrient environment for associated microorganisms in the mesohyl. This environment appears ideal for heterotrophic microbes to thrive on and the presence of heterotrophic microorganisms in sponges has been reported widely (Maldonado et al. 2012; Taylor et al. 2007; Hentschel et al. 2012). It has been suggested that the uptake of dissolved organic carbon (DOC) by the sponge *Theonella swinhoei* is also mediated

by its bacterial symbionts (Yahel et al. 2003). Bacteria mediated carbon feeding in the coral reef sponge *Haliscara caerulae* has been reported and revealed that at least some of the heterotrophic bacteria in sponges deliver nutrients to their host this way (de Goeij et al. 2008). However, the exact processes of carbon consumption by heterotrophic bacteria and the connection to host carbon feeding are yet to be investigated. It is conceivable that microbes in sponges feed on monosaccharides and mediate digestion of more complex sugars from which the host might also profit. In contrast, carbon autotrophy of sponge-associated microbes has attracted much more attention. Photosynthetic sulphur bacteria, diatoms and dinoflagellates and most dominantly cyanobacteria have been found in sponges with the latter being associated to 90% of all coral reef sponges (Wilkinson 1983; Thacker & Starnes 2003; Taylor et al. 2007). The photosynthetic activities of these organisms can provide the sponge with more than 50% of its required energy as estimated by photosynthesis and respiration rate measurements (Wilkinson 1983). The cyanobacterium *Synechococcus spongiarum* is especially prevalent in sponges and appears to be sponge specific with the largest sponge specific cluster reported, 245 16S rRNA gene sequences from 40 different sponge species (Erwin & Thacker 2007; Simister et al. 2011). These associations are considered truly mutualistic with nutrient exchanges between the sponge and its symbionts that additionally profit from shelter in the sponge body (Brümmer et al. 2008). Apart from phototrophic carbon fixation some sponge symbionts have shown the genomic potential for carbon fixation via the 3-hydroxypropionate cycle as shown for the archaeal symbiont *Crenarchaeum symbiosum* from the sponge *Axinella mexicana* (Hallam et al. 2006b). Furthermore, the microbial metagenome of the sponge *Cymbastela concentrica* contained aerobic type carbon monoxide dehydrogenase genes that were significantly overrepresented in sponge samples in comparison to seawater and thus might indicate the presence of lithoheterotrophic organisms that gain energy by CO oxidation (Thomas et al. 2010a). Indications for further one carbon metabolism in form of methane producing and methane oxidising organisms have also been reported from sponges. Methanotrophic bacteria have been found in the sponge *Arenosclera brasiliensis* (Trindade-Silva et al. 2012) and in a deep-sea carnivorous sponge (Vacelet et al. 1996; 1995). Methanogenic archaea are associated with the Great Barrier Reef sponge *Rhopaloeides odorabile*, which might indicate methane production in anaerobic “tissues” of the sponge (Webster et al. 2001).

Another important and well-investigated nutrient cycle in sponges is the nitrogen cycle. Nitrogen is often limiting in seawater and its supply by symbionts might therefore be important to the sponge. Cyanobacterial and heterotrophic bacterial symbionts of sponges are known to fix inorganic nitrogen in sponges (Wilkinson & Fay 1979; Shieh & Lin 1994; Mohamed et al. 2008b) that can ensure the nitrogen supply to the host. Nitrification is another process mediated by bacterial and archaeal symbionts of sponges in which ammonia is oxidised to nitrite and later to nitrate. Ammonia oxidising archaea and bacteria have been detected in sponges, as well as nitrite oxidisers and the relevant functional genes such as amoA encoding for ammonia monooxygenase (Radax et al. 2012a; Hoffmann et al. 2009; Bayer et al. 2007; 2008; Hallam et al. 2006b; Steger et al. 2008; Diaz & Ward 1997). Furthermore, physiological measurements of nitrate and ammonium release (Bayer et al. 2007; Diaz & Ward 1997; Hoffmann et al. 2009) as well as relevant genes in metagenomic, metatranscriptomic and metaproteomic datasets (Thomas et al. 2010a; Liu et al. 2012; Radax et al. 2012b; Fan et al. 2012) support the presence of active nitrification in sponges. Ammonia is a potentially toxic waste product of the sponge that is detoxified by the microbial symbionts and these benefit from the additional nitrogen sources. Ammonia can be assimilated by sponge symbionts via the glutamine-synthase-glutamine oxoglutarate aminotransferase (GS-GOGAT) pathway (Thomas et al. 2010a) and the uptake and degradation of urea is also supported by genomic data (Hallam et al. 2006a). Fan et al. (2012) found an overrepresentation of glutamate dehydrogenase in three out of six sponge metagenomes and hypothesized that sponge symbionts might also use this enzyme for ammonium assimilation in case of ammonium excess in the sponge.

Interestingly, also anaerobic processes seem to play a role in the nitrogen cycle sponges. Anoxic zones can arise in sponges when the pumping activity stops (Hoffmann et al. 2008). Anaerobic ammonia oxidation (anammox) was measured in the sponge *Geodia barretti* and organisms known for this pathway have been detected in sponges by their 16S rRNA gene sequence (Hoffmann et al. 2009; Mohamed et al. 2010). Indications for denitrifying organisms in sponges were also found (Enticknap et al. 2006). Later denitrification rates were measured in the cold water sponge *G. baretti* and potential denitrifying *Betaproteobacteria* detected by 16S rRNA and functional gene analysis (Hoffmann et al. 2009). Schläppy et al. (2010) further detected denitrification rates in both high and low microbial abundance sponges. Genes for denitrification were also found in the metagenome and metaproteome of *C. concentrica* (Liu et al. 2012).

Further analyses of six metagenomes from different sponge species showed that denitrification is a shared functional feature carried out by sponge associated microbes and it is encoded by different functional genes in the metagenomes of different sponges (Fan et al. 2012).

The sulphur cycle in marine sponges is also dependent on the presence of anoxic zones. Sulphur reduction was measured in *G. barretti* (Hoffmann et al. 2005) and sulphur reducing and oxidising microbes have been detected in several sponge species (Manz et al. 2000; Webster et al. 2001; Hoffmann et al. 2005; 2006; Meyer & Kuever 2008; Taylor et al. 2007). Furthermore, a key enzyme for sulphate reduction and sulphur oxidation, APS reductase (aprA) revealed sulphur-oxidising *Alpha-* and *Gammaproteobacteria* and a sulphate-reducing archaeon in the sponge *Polymastia* cf. *corticata* (Meyer & Kuever 2008). The authors suggested the existence of a sulphur cycle in sponges that is analogous to that occurring in marine oligochaetes (Dubilier et al. 2001).

Apart from these major nutrient cycles some other functions have been described from sponge symbionts. Ahn et al. (2003) postulated that symbionts of marine sponges might be able to degrade halogenated chemicals in sponges. Bayer et al. (2012) showed the presence of novel halogenase gene clusters in the metagenomes of *Aplysina aerophoba* and confirmed their identity by single-cell screenings as of deltaproteobacterial, actinobacterial, and poribacterial origin. Additionally, biosynthesis of cofactors and vitamins have been identified as a shared feature of microbial sponge symbionts (Hentschel et al. 2012). The presence of gene clusters encoding for riboflavin (vitamin B_2), cobalamin (vitamin B_{12}), biotin (vitamin B_7), and thiamine (vitamin B_1) has been reported from several omics studies of individual sponge symbionts or microbial communities of sponges (Hallam et al. 2006a; Thomas et al. 2010a; Liu et al. 2010; Fan et al. 2012). These studies do however not reveal if and how microbial produced vitamins are transferred to the host and these questions therefore remain an interesting topic for further investigations.

Marine sponges are well known as producers of highly diverse secondary metabolites. These substances are used as a natural defence by sponges against predators, space competitors, and harmful microorganisms (Taylor et al. 2007). From a biotechnological and pharmaceutical point of view they are of high interest as they show antimicrobial,

antifungal, antiviral, anti-fouling, anti-inflammatory, and even anticancer activities (Blunt et al. 2013). It was shown in several studies that many of the bioactive metabolites from sponges are actually produced by their microbial symbionts (Thomas et al. 2010b; Piel 2004b). This has drawn much attention to the secondary metabolism of sponge symbionts. The substances produced by these organisms are diverse (Thomas et al. 2010b; Blunt et al. 2013) but in focus are often polyketides and non-ribosomal peptides. Polyketide synthase (PKS) gene clusters have been detected in several metagenomes of marine sponge associated microbial communities. Bacterial PKS genes were found in metagenomes of the sponges *Theonella swinhoei* and *Psammoicinia bulbosa* and are thought to produce the antitumor polyketides onnamide A and psymberin (Fisch et al. 2009; Piel 2004a). Furthermore, a sponge exclusive and ubiquitous monomodular PKS system (Sup-PKS) was found in metagenomic DNA of several sponge species and was predicted to produce methyl-branched fatty acids (Fieseler et al. 2007; Hochmuth & Piel 2009). It was shown that these fatty acids and the relevant PKS system occur exclusively in HMA sponges and also correlate with the presence of the candidate phylum *Poribacteria* (Hochmuth et al. 2010). It was thus suggested that Poribacteria might be the producers of this PKS system, a theory that was supported by linking sup-PKS system genes to a single bacterial cell that was identified as poribacterial based on its 16S rRNA gene sequence (Siegl & Hentschel 2009). Another class of bioactive metabolites from sponge symbionts are non-ribosomal peptides. Non-ribosomal peptide synthases (NRPS) have been identified from metagenomes of sponge microbiota and from bacterial isolates of sponge symbionts (Pimentel-Elardo et al. 2012; Blunt et al. 2013; Piel 2004b). These systems are found foremostly in members of the *Actinobacteria* and especially *Streptomyces*. Single-cell based analysis did also identify a sponge symbiont of the phylum *Chloroflexi* as a producer of non-ribosomal peptides (Siegl & Hentschel 2009).

1.2.5. The candidate phylum *Poribacteria*

In studies of sponge-microbe associations, one bacterial phylum has attracted much attention, namely the candidate phylum *Poribacteria*. Its status as candidate phylum indicates that as of now there are no cultured representatives, which restricts all characterisations to culture independent studies. Nevertheless, *Poribacteria* have many characteristics that make them sponge symbionts of special interest. *Poribacteria* were first discovered in marine sponges (Fieseler et al. 2004) and have since then been

described from various sponge species around the world where they are highly abundant (Lafi et al. 2009; Schmitt et al. 2012; Webster et al. 2010). They are with more than 400 different OTUs (based on 97% sequences similarity) the most diverse and also one of the most abundant phyla in a 16S rRNA gene amplicon sequencing study including 32 sponge species from around the world (Schmitt et al. 2012). Outside of sponges they have only been detected as part of the so-called rare biosphere (Webster et al. 2010; Moitinho-Silva et al. 2013; Taylor et al. 2013; Pham et al. 2008). Thus, *Poribacteria* can still be considered as being largely exclusive to marine sponges and they form some of the largest sponge specific clusters discovered so far (Simister et al. 2011). It appears likely that *Poribacteria* are vertically transmitted between sponge generations, as they have been found in sponge reproductive and embryonic stages in several studies using molecular cloning and amplicon sequencing techniques of the 16S rRNA gene (Schmitt et al. 2008; Webster et al. 2010). However, the potential presence of *Poribacteria* in seawater renders environmental transmission an additional option for these symbionts.

Phylogenetically *Poribacteria* are regarded as part of the *Planctomycete-Verrucomicrobia-Chlamydiae* (PVC) superphylum (Wagner & Horn 2006). Members of this phylum show features of cell compartmentalisation (Santarella-Mellwig et al. 2010; Fuerst 2005). This feature has also been proposed for *Poribacteria* based on ring-shaped fluorescence *in situ* hybridisation (FISH) signals indicating a ribosome-free area in the poribacterial cells and the presence of compartmentalised cells in transmission electron microscopy images of the mesohyl of poribacteria-containing sponges (Fieseler et al. 2004). Furthermore, it has been shown that *Poribacteria* are highly active based on strong fluorescent signals from FISH studies and 16S rRNA transcripts (Fieseler et al. 2004; Kamke et al. 2010). Recent studies revealed also that outside of sponges *Poribacteria* are not only rare in abundance but also transcriptionally inactive while activity inside of the host was confirmed (Moitinho-Silva et al. 2013). Functional information about *Poribacteria* is still scarce. Fieseler et al. (2006) sequenced a 39 kilo-basepair (kbp) long insert of a fosmid clone from an *Aplysina aerophoba* metagenome library that contained a 16S rRNA gene of a poribacterium. Apart from the 16S rRNA gene the fosmid clone sequence contained 26 open reading frames (orfs) encoding for several hypothetical proteins with similarities to genes from the planctomycete *Rhodopirellula baltica*, an undefined oxidoreductase, and eight orfs encoding for transmembrane proteins that might build a transporter (Fieseler et al. 2006). Other

functional information about Poribacteria was obtained by screening the amplified DNA of single cells for functional gene and at the same time identifying phylogenetic origin by 16S rRNA gene analysis (Bayer et al. 2012; Siegl & Hentschel 2009). Thus a PKS system and potential halogenase gene were connected to *Poribacteria* (see section 1.2.4). However, there are still large gaps in our knowledge about the functional repertoire of poribacterial and other symbionts of marine sponges.

1.3. Research objectives

While the microbial diversity of sponge-associated microorganisms is becoming increasingly well understood, we are only beginning to understand the functional aspects of these relationships. The lack of cultivated representatives of the sponge community restricts functional investigation for most sponge symbionts to either physiological investigations of metabolites in sponges or studies based on nucleic acids. Most functional information about the sponge microbiome was obtained by screening for functional genes or whole community analysis of DNA, RNA, or proteins. However, these analyses can give only very limited information about the phylogenetic identity behind the detected functions. Connecting phylogeny with function is however an important component to understand complex interactions of diverse microbial communities such as found in sponges. It allows gaining specific information about target organisms of interest that might get lost in the bulk of information obtained by whole community analyses. Connecting phylogenetic identity with function is therefore an important requirement to fully understand the complex interactions in the sponge microbiome and the roles of individual players in this system.

The aim of this thesis was to investigate the specific functions of the candidate phylum *Poribacteria* using single-cell genomics (see Chapter 2). Genomic characterization on a single-cell level allowed for a connection of functional repertoire and phylogenetic identity not only to species but individual cell level. This made it possible to detect differences in the functional properties of organisms even with identical 16S rRNA gene that would usually be classified as the same organism even though they might represent different ecotypes. This study was focused on the candidate phylum *Poribacteria*, as they represent widespread and abundant symbionts that are almost exclusive to sponges. It was therefore hypothesised that *Poribacteria* show functional traits with importance for this symbiosis. I therefore aimed to investigate the common and individual functions of

Poribacteria by (I) obtaining several single amplified genomes (SAGs) of the candidate phylum *Poribacteria* using fluorescence activated cell sorting (FACS), multiple displacement amplification of the total genomic DNA, PCR based identity screening, and whole genome sequencing using next generation sequencing techniques with subsequent genome analysis (Chapter 2). To cover a large range of the functional potential of *Poribacteria*, I further intended to (II) obtain single-cells representing different groups of poribacterial phylogenetic range based on the 16S rRNA gene. This allowed me to (III) obtain a higher resolution phylogenetic placement of the candidate phylum *Poribacteria* based on sequence alignment of several marker genes compared to 16S rRNA gene based phylogeny. By detecting common functional features between poribacterial genomes I aimed to (IV) identify traits important for the role of *Poribacteria* in this symbiosis. Finally the extensive analysis of several genomes of the candidate phylum *Poribacteria* allowed me to (V) gain a first glimpse of potential novelties of the genomic repertoire of this candidate phylum.

1.4. References

Ahn Y-B, Rhee S-K, Fennell DE, Kerkhof LJ, Hentschel U, Häggblom MM. (2003). Reductive dehalogenation of brominated phenolic compounds by microorganisms associated with the marine sponge *Aplysina aerophoba*. Appl Environ Microb 69:4159–4166.

Bayer K, Scheuermayer M, Fieseler L, Hentschel U. (2012). Genomic mining for novel FADH2-dependent halogenases in marine sponge-sssociated microbial consortia. Mar Biotechnol 15:63–72.

Bayer K, Schmitt S, Hentschel U. (2007). Microbial nitrification in Mediterranean sponges: possible involvement of ammonium-oxidizing *Betaproteobacteria*. *Porifera* research: biodiversity, innovation, sustainability. Série Livros. Museu Nacional, Rio de Janeiro 165–171.

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

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

Bell JJ, Davy SK, Jones T, Taylor MW, Webster NS. (2013). Could some coral reefs become sponge reefs as our climate changes? Glob Chang Biol 19:2613–2624.

Bergquist PR. (1978). Sponges. Hutchinson and Co. Ltd: London.

Blunt JW, Copp BR, Keyzers RA, Munro MHG, Prinsep MR. (2013). Marine natural

products. Nat Prod Rep 30:237–323.

Brümmer F, Pfannkuchen M, Baltz A, Hauser T, Thiel V. (2008). Light inside sponges. J Exp Mar Biol Ecol 367:61–64.

De Goeij JM, Moodley L, Houtekamer M, Carballeira NM, van Duyl FC. (2008). Tracing ^{13}C-enriched dissolved and particulate organic carbon in the bacteria-containing coral reef sponge *Halisarca caerulea*: Evidence for DOM-feeding. Limnol Oceanogr 53:1376–1386.

Diaz MC, Rützler K. (2001). Sponges: An essential component of Caribbean coral reefs. Bul Mar Sci 69:535–546.

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

Dohrmann M, Wörheide G. (2013). Novel scenarios of early animal evolution--is it time to rewrite textbooks? Integr Comp Biol 53:503–511.

Dubilier N, Mülders C, Ferdelman T, de Beer D, Pernthaler A, Klein M, et al. (2001). Access : Endosymbiotic sulphate-reducing and sulphide-oxidizing bacteria in an oligochaete worm. Nature 411:298–302.

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

Erpenbeck D, Breeuwer J, van der Velde H, van Soest R. (2002). Unravelling host and symbiont phylogenies of halichondrid sponges (*Demospongiae*, *Porifera*) using a mitochondrial marker. Mar Biol 141:377–386.

Erwin DH, Laflamme M, Tweedt SM, Sperling EA. (2011). The Cambrian Conundrum: Early divergence and later ecological success in the early history of animals. Science 334:1091-1097.

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

Fan L, Liu M, Simister R, Webster NS, Thomas T. (2013). Marine microbial symbiosis heats up: the phylogenetic and functional response of a sponge holobiont to thermal stress. ISME J 7:991–1002.

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS, et al. (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc Natl Acad Sci USA 109:E1878–87.

Fieseler L, Hentschel U, Grozdanov L, Schirmer A, Wen G, Platzer M, et al. (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Appl Environ Microb 73:2144–2155.

Fieseler L, Horn M, Wagner M, Hentschel U. (2004). Discovery of the novel candidate phylum '*Poribacteria*' in marine sponges. Appl Environ Microb 70:3724–3732.

Fieseler L, Quaiser A, Schleper C, Hentschel U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum *Poribacteria* by environmental genomics. Environ Microbiol 8:612–624.

Fisch KM, Gurgui C, Heycke N, van der Sar SA, Anderson SA, Webb VL, et al. (2009). Polyketide assembly lines of uncultivated sponge symbionts from structure-based gene targeting. Nat Chem Biol 5:494–501.

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

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

Fuerst JA. (2005). Intracellular compartmentation in planctomycetes. Annu Rev Microbiol 59:299–328.

Gardères J, Taupin L, Saïdin JB, Dufour A, Pennec G. (2012). N-acyl homoserine lactone production by bacteria within the sponge *Suberites domuncula* (Olivi, 1792) (*Porifera*, *Demospongiae*). Mar Biol 159:1685–1692.

Gazave E, Lapébie P, Ereskovsky AV, Vacelet J, Renard E, Cárdenas P, et al. (2011). No longer *Demospongiae*: *Homoscleromorpha* formal nomination as a fourth class of *Porifera*. Hydrobiologia 687:3–10.

Giles EC, Kamke J, Moitinho-Silva L, Taylor MW, Hentschel U, Ravasi T, et al. (2012). Bacterial community profiles in low microbial abundance sponges. FEMS Microbiol Ecol 83:232-241.

Gloeckner V, Lindquist N, Schmitt S, Hentschel U. (2012). *Ectyoplasia ferox*, an Experimentally Tractable Model for Vertical Microbial Transmission in Marine Sponges. Microb Ecol 65:462–474.

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

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

Hentschel U, Fieseler L, Wehrl M, Gernert C. (2003). Microbial diversity of marine sponges. Sponges (*Porifera*).

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

Hentschel U, Piel J, Degnan SM, Taylor MW. (2012). Genomic insights into the marine sponge microbiome. Nat Rev Micro 10:641–654.

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

Hochmuth T, Niederkrüger H, Gernert C, Siegl A, Taudien S, Platzer M, et al. (2010). Linking chemical and microbial diversity in marine sponges: Possible role for *Poribacteria* as producers of methyl‐branched fatty acids. ChemBioChem 11:2572–2578.

Hochmuth T, Piel J. (2009). Polyketide synthases of bacterial symbionts in sponges evolution-based applications in natural products research. Phytochemistry 70:1841–1849.

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

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

Hoffmann F, Rapp HT, Reitner J. (2006). Monitoring microbial community composition by fluorescence *in situ* hybridization during cultivation of the marine cold-water sponge *Geodia barretti*. Mar Biotechnol 8:373–379.

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

Hooper JNA, van Soest RMW. (2002). Systema porifera. A guide to the classification of sponges. Plenum Publishers : New York.

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

Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, Hentschel U. (2009). Widespread distribution of *Poribacteria* in *Demospongiae*. Appl Environ Microb 75:5695–5699.

Lee OO, Wang Y, Yang J, Lafi FF, Al-Suwailem A. (2010). Pyrosequencing reveals highly diverse and species-specific microbial communities in sponges from the Red Sea. ISME J 5:650-664.

Lewin RA. (1982). Symbiosis and Parasitism: Definitions and evaluations. BioScience 32:254–260.

Li CW, Chen JY, Hua TE. (1998). Precambrian sponges with cellular structures. Science 279:879–882.

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

Liu MY, Kjelleberg S, Thomas T. (2010). Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. ISME J 5:427–435.

Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, Bradley AS, et al. (2008). Fossil steroids record the appearance of *Demospongiae* during the Cryogenian period.

Nature 457:718–721.

Maldonado M, Navarro L, Grasa A, Gonzalez A, Vaquerizo I. (2011). Silicon uptake by sponges: a twist to understanding nutrient cycling on continental margins. Sci Rep 1:30.

Maldonado M, Ribes M, van Duyl FC. (2012). Nutrient fluxes through sponges: biology, budgets, and ecological implications. Adv Mar Biol 62:113–182.

Manz W, Arp G, Schumann-Kindel G, Szewzyk U, Reitner J. (2000). Widefield deconvolution epifluorescence microscopy combined with fluorescence *in situ* hybridization reveals the spatial arrangement of bacteria in sponge tissue. J Microbiol Meth 40:125–134.

McMurray SE, Blum JE, Pawlik JR. (2008). Redwood of the reef: Growth and age of the giant barrel sponge *Xestospongia muta* in the Florida Keys. Mar Biol 155:159–171.

Meyer B, Kuever J. (2008). Phylogenetic diversity and spatial distribution of the microbial community associated with the Caribbean deep-water sponge *Polymastia* cf. *corticata* by 16S rRNA, aprA, and amoA gene analysis. Microb Ecol 56:306–321.

Mohamed NM, Cicirelli EM, Kan J, Chen F, Fuqua C, Hill RT. (2008a). Diversity and quorum-sensing signal production of *Proteobacteria* associated with marine sponges. Environ Microbiol 10:75–86.

Mohamed NM, Colman AS, Tal Y, Hill RT. (2008b). Diversity and expression of nitrogen fixation genes in bacterial symbionts of marine sponges. Environ Microbiol 10:2910–2921.

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

Moitinho-Silva L, Bayer K, Cannistraci CV, Giles EC, Ryu T, Seridi L, et al. (2013). Specificity and transcriptional activity of microbiota associated with low and high microbial abundance sponges from the Red Sea. Mol Ecol (in press).

Müller WEG, Müller IM. (2003). Origin of the metazoan immune system: identification of the molecules and their functions in sponges. Integr Comp Biol 43:281–292.

Nguyen MTHD, Liu M, Thomas T. (2013). Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol (in press).

Pawlik JR. (2011). The chemical ecology of sponges on Caribbean reefs: natural products shape natural systems. BioScience 61:888–898.

Perea-Blázquez A, Davy SK, Bell JJ. (2012). Estimates of particulate organic carbon flowing from the pelagic environment to the benthos through sponge assemblages. PLoS ONE 7:e29569.

Pham VD, Konstantinidis KT, Palden T, DeLong EF. (2008). Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific subtropical gyre. Environ Microbiol 10:2313–2330.

Piel J. (2004a). Antitumor polyketide biosynthesis by an uncultivated bacterial symbiont of the marine sponge *Theonella swinhoei*. Proc Natl Acad Sci USA 101:16222–16227.

Piel J. (2004b). Metabolites from symbiotic bacteria. Natural product reports.

Pile AJ, Patterson MR, Witman JD. (1996). *In situ* grazing on plankton. Mar Ecol Prog Ser 141:95–102.

Pimentel-Elardo SM, Grozdanov L, Proksch S, Hentschel U. (2012). Diversity of nonribosomal peptide synthetase genes in the microbial metagenomes of marine sponges. Marine Drugs 10:1192–1202.

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

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

Reiswig HM. (1971). Particle feeding in natural populations of three marine demosponges. Biol Bull 141:568–591.

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

Ribes M, Coma R, Gili JM. (1999). Natural diet and grazing rate of the temperate sponge *Dysidea avara* (*Demospongiae*, *Dendroceratida*) throughout an annual cycle. Mar Ecol Prog Ser 176:179–190.

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

Ruppert EE, Fox RS, Barnes RD. (2004). Invertebrate zoology : a functional evolutionary approach. 7 ed. Thomson, Brooks/Cole: Belmont Calif. [u.a.].

Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, Budd A, et al. (2010). The Compartmentalized Bacteria of the *Planctomycetes-Verrucomicrobia-Chlamydiae* Superphylum Have Membrane Coat-Like Proteins. Plos Biol 8:e1000281.

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

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

Schmitt S, Hentschel U, Taylor MW. (2011). Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. Hydrobiologia 687:341–351.

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.

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

Shieh WY, Lin YM. (1994). Association of heterotrophic nitrogen-fixing bacteria with a marine sponge of *Halichondria* sp. B Mar Sci 54:557–564.

Siegl A, Hentschel U. (2009). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. Environ Microbiol Rep 2:507–513.

Simister R, Taylor MW, Rogers KM, Schupp PJ, Deines P. (2013). Temporal molecular and isotopic analysis of active bacterial communities in two New Zealand sponges. FEMS Microbiol Ecol 85:195–205.

Simister R, Taylor MW, Tsai P, Fan L, Bruxner TJ, Crowe ML, et al. (2012a). Thermal stress responses in the bacterial biosphere of the Great Barrier Reef sponge, *Rhopaloeides odorabile*. Environ Microbiol 14:3232–3246.

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

Simister RL, Deines P, Botté ES, Webster NS, Taylor MW. (2011). Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. Environ Microbiol 14: 517-524.

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

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

Taylor MW, Schupp PJ, Baillie HJ, Charlton TS, de Nys R, Kjelleberg S, et al. (2004a). Evidence for acyl homoserine lactone signal production in bacteria associated with marine sponges. Appl Environ Microb 70:4387–4389.

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

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

Taylor MW, Tsai P, Simister RL, 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.

Thacker RW, Starnes S. (2003). Host specificity of the symbiotic cyanobacterium *Oscillatoria spongeliae* in marine sponges, *Dysidea* spp. Mar Biol 142:643–648.

Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A, et al. (2010a).

Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J 4:1557–11567.

Thomas TRA, Kavlekar DP, LokaBharathi PA. (2010b). Marine drugs from sponge-microbe association—a review. Marine Drugs 8:1417–1468.

Trindade-Silva AE, Rua C, Silva GGZ, Dutilh BE, Moreira APB, Edwards RA, et al. (2012). Taxonomic and functional microbial signatures of the endemic marine sponge *Arenosclera brasiliensis*. PLoS ONE 7:e39905.

Vacelet J. (1975). Etude en microscopie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (*Dictyoceratida*). J Microsc Biol Cell 23:271–288.

Vacelet J, Boury-Esnault N, Fiala-Medioni A, Fisher CR. (1995). A methanotrophic carnivorous sponge. Nature 377:296–296.

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

Vacelet J, Fiala-Medioni A, Fisher CR. (1996). Symbiosis between methane oxidizing bacteria and a deep sea carnivorous cladorhizid sponge. Mar Ecol Prog Ser 145: 77-85.

Van Soest RWM, Boury-Esnault N, Vacelet J, Dohrmann M, Erpenbeck D, De Voogd NJ, et al. (2012). Global diversity of sponges (*Porifera*). PLoS ONE 7:e35105.

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

Wagner M, Horn M. (2006). The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotech 17:241–249.

Webster NS, Negri AP, Munro MMHG, Battershill CN. (2004). Diverse microbial communities inhabit Antarctic sponges. Environ Microbiol 6:288–300.

Webster NS, Taylor MW. (2011). Marine sponges and their microbial symbionts: love and other relationships. Environ Microbiol 14:335–346.

Webster NS, Taylor MW, Behnam F, Lücker 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.

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

Wehrl M, Steinert M, Hentschel U. (2007). Bacterial uptake by the marine sponge *Aplysina aerophoba*. Microb Ecol 53:355–365.

Weisz J, Lindquist N, Martens C. (2008). Do associated microbial abundances impact marine demosponge pumping rates and tissue densities? Oecologia 155:367–376.

Wiens M, Korzhev M, Perovic-Ottstadt S, Luthringer B, Brandt D, Klein S, et al. (2007). Toll-like receptors are part of the innate immune defense system of sponges

(*Demospongiae*: *Porifera*). Mol Biol Evol 24:792–804.

Wilkinson CR. (1978a). Microbial associations in sponges. I. Ecology, physiology and microbial populations of coral reef sponges. Mar Biol 49:161–167.

Wilkinson CR. (1978b). Microbial associations in sponges. II. Numerical analysis of sponge and water bacterial populations. Mar Biol 49:169–176.

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

Wilkinson CR. (1983). Net primary productivity in coral-reef sponges. Science 219:410–412.

Wilkinson CR, Fay P. (1979). Nitrogen fixation in coral reef sponges with symbiotic cyanobacteria. Nature 279:527–529.

Wilkinson CR, Garrone R, Vacelet J. (1984). Marine sponges discriminate between food bacteria and bacterial symbionts: Electron microscope radioautography and *in situ* evidence. P Roy Soc Lond B Bio 220:519–528.

Wörheide G, Dohrmann M, Erpenbeck D, Larroux C, Maldonado M, Voigt O, et al. (2012). Deep phylogeny and evolution of sponges (phylum *Porifera*). Adv Mar Biol 61:1–78.

Yahel G, Sharp JH, Marie D, Häse C, Genin A. (2003). *In situ* feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: Bulk DOC is the major source for carbon. Limnol Oceanogr 48:141–149.

Zan J, Cicirelli EM, Mohamed NM, Sibhatu H, Kroll S, Choi O, et al. (2012). A complex LuxR-LuxI type quorum sensing network in a roseobacterial marine sponge symbiont activates flagellar motility and inhibits biofilm formation. Mol Microbiol 85:916–933.

Zan J, Fuqua C, Hill RT. (2011). Diversity and functional analysis of luxS genes in vibrios from marine sponges *Mycale laxissima* and *Ircinia strobilina*. ISME J 5:1505–1516.

Zan J, Heindl JE, Liu Y, Fuqua C, Hill RT. (2013). The CckA-ChpT-CtrA phosphorelay system is regulated by quorum sensing and controls flagellar motility in the marine sponge symbiont *Ruegeria* sp. KLH11. PLoS ONE 8:e66346.

2. Exploring symbioses by single-cell genomics

This article was published in the peer-reviewed journal The Biological Bulletin.

For documentation of individual contributions to this work and consent of all authors for publication in this thesis please refer to appendix A.

Reference: *Biol. Bull.* **223:** 30–43. (August 2012)

Exploring Symbioses by Single-Cell Genomics

JANINE KAMKE[1], KRISTINA BAYER[1], TANJA WOYKE[2], AND UTE HENTSCHEL[1,*]

[1]*Julius-von-Sachs Institute for Biological Sciences, University of Würzburg, Julius-von-Sachs Platz 3, 97082 Würzburg, Germany; and* [2]*Department of Energy Joint Genome Institute, Walnut Creek, California*

Abstract. Single-cell genomics has advanced the field of microbiology from the analysis of microbial metagenomes where information is "drowning in a sea of sequences," to recognizing each microbial cell as a separate and unique entity. Single-cell genomics employs Phi29 polymerase-mediated whole-genome amplification to yield microgram-range genomic DNA from single microbial cells. This method has now been applied to a handful of symbiotic systems, including bacterial symbionts of marine sponges, insects (grasshoppers, termites), and vertebrates (mouse, human). In each case, novel insights were obtained into the functional genomic repertoire of the bacterial partner, which, in turn, led to an improved understanding of the corresponding host. Single-cell genomics is particularly valuable when dealing with uncultivated microorganisms, as is still the case for many bacterial symbionts. In this review, we explore the power of single-cell genomics for symbiosis research and highlight recent insights into the symbiotic systems that were obtained by this approach.

Background

The fields of metagenomics, metatranscriptomics, and metaproteomics, among others collectively referred to as "omics," have made a tremendous impact on symbiosis research (see other articles in this special issue). For many decades, symbiosis research was possible solely by descriptive approaches because neither could the symbionts be cultured (largely unchanged to this day) nor was there experimental access to many symbiotic systems (also largely unchanged). The implementation of cultivation-independent approaches based on 16S rRNA gene sequences thus initiated a major revolution by making it possible to place organisms that were frequently known only by electron microscopy into a phylogenetic context. 16S rRNA gene phylogenies further helped to delineate co-evolution and co-speciation events by comparing host and symbiont phylogenies.

The implementation of omics methods spurred a second wave of information in symbiosis research as it became possible to predict the genomic underpinnings of symbioses. For example, the discovery that many insects had genomically encoded nutritional interdependencies on their symbiotic bacteria represented a milestone discovery (*i.e.,* Gil *et al.*, 2003; Wu *et al.*, 2006; Moran *et al.*, 2008; Wilson *et al.*, 2010). Similarly, metagenomics provided novel insights into chemoautotrophic symbioses, in that sulfur-oxidizing and sulfate-reducing symbionts provide a gutless marine worm host with multiple sources of nutrition (Woyke *et al.*, 2006). Moreover, omics approaches have provided the first glimpse into the functional gene repertoire of marine sponges and their beneficial microbial consortia (Hallam *et al.*, 2006, Thomas *et al.*, 2010; Liu, M., *et al.*, 2011; Liu *et al.*, 2012; Fan *et al.*, 2012).

One recent addition to the omics repertoire is single-cell genomics. It relies on genomic sequence information from individual microbial cells and is entirely cultivation-independent. By use of Phi29 polymerase it is possible to obtain comprehensive genomic information from individual microbial cells—something that to our knowledge is not possible with any other technique to date (Hutchison and Venter, 2006; Binga *et al.*, 2008; Ishoey *et al.*, 2008). Single-cell genomics is especially well suited for symbiosis research in which the vast majority of symbionts have not been cultivated and are thus not accessible by conventional techniques. Here we present a brief overview of the methodology and its current limitations and challenges. We then review the current state of single-cell genomics techniques in symbiosis research using five recently published exam-

Received 11 April 2012; accepted 19 June 2012.
* Author to whom correspondence should be addressed: E-mail: ute.hentschel@uni-wuerzburg.de

Abbreviations: MDA, multiple displacement amplification; SAG, single amplified genome.

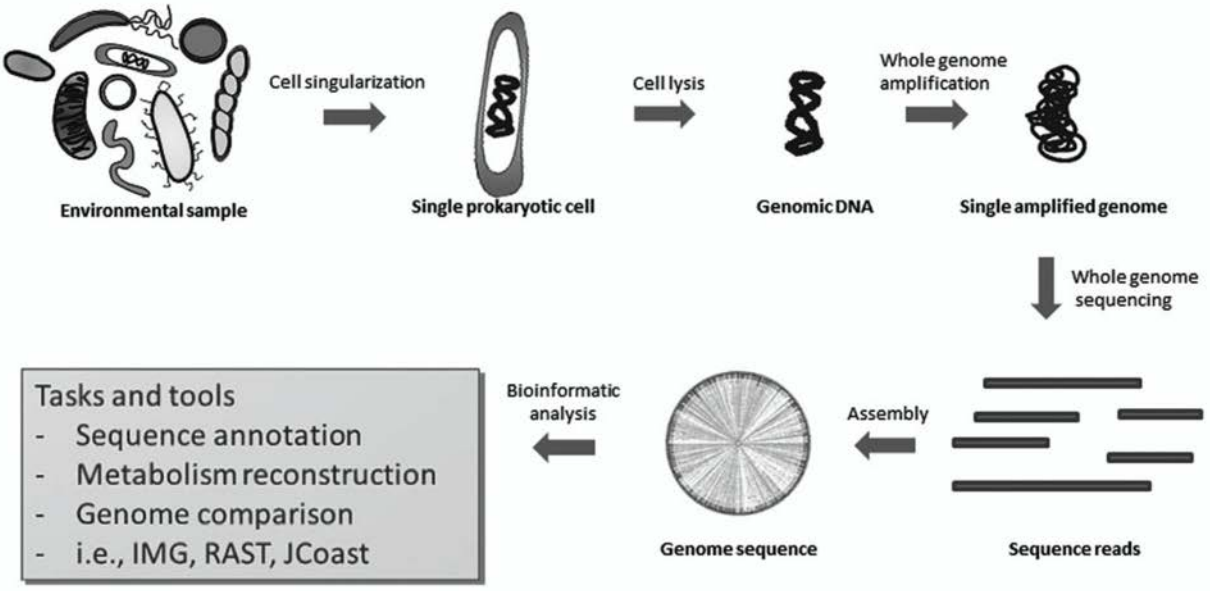


Figure 1. Diagram of the experimental procedures involved in single-cell genomic analysis.

ples. We further discuss how single-cell genomics has been applied to provide genomic insights into secondary metabolism, and we present future prospects of how this technique may lead to further advancements in the field. The application of single-cell genomics to other areas of microbiology is beyond the scope of this paper, and we refer the reader to other recent reviews (Hutchison and Venter, 2006; de Jager and Siezen, 2011; Kalisky and Quake, 2011; Yilmaz and Singh, 2011).

A Laboratory Primer on Single-Cell Genomics

The principle of single-cell genomics is to singularize microbial cells from environmental samples, to access the complete genomic material of a single cell, and to generate sufficient amounts of DNA by amplification for whole-genome sequencing (Fig. 1). The first step is the efficient singularization of the cells, which depends largely on the characteristics of the sample. If enrichments or even pure cultures are available, serial dilution is a possible method (Zhang *et al.*, 2006). One of the most commonly used methods for obtaining single cells from uncultivated microbiota is fluorescence-activated cell sorting (FACS). Microbial cells can be labeled with a fluorescent dye or subjected to fluorescence *in situ* hybridization (FISH) to target intact cells or cells of a distinct phylogenetic affiliation (Podar *et al.*, 2007; Yilmaz *et al.*, 2010). The Single Cell Genomics Center at the Bigelow Laboratory for Ocean Sciences, East Boothbay, Maine, offers services for single-cell sorting and, optionally, also whole-genome amplification, thus making this technique available to any laboratory.

Microfluidic chambers have also proven successful for obtaining single amplified genomes (SAGs) (Marcy *et al.*, 2007a, b; Blainey *et al.*, 2011). This method uses reaction volumes of only 60 nl, which reduces the likelihood of contaminating the sample. Microfluidic devices have also been recently developed for FISH and tyramide-signal-amplification FISH (tsa-FISH) followed by cell sorting *via* flow cytometry directly on the device (Chen *et al.*, 2011; Liu, P., *et al.*, 2011). This approach holds great promise for 16S rRNA gene-based identification of single cells, while bearing low risks of contamination.

Micromanipulation techniques are particularly useful when the target microorganism is morphologically distinct (Ishoey *et al.*, 2008; Woyke *et al.*, 2010). Micromanipulation using microcapillaries has been used successfully in combination with FISH to target phylogenetically distinct cells of interest (Kvist *et al.*, 2007). Other options are micromanipulation by use of optical tweezers or laser capture microdissection. The first method has already been used in combination with microfluidic devices (Pamp *et al.*, 2012), while the latter was applied only to fixed bacterial samples (Klitgaard *et al.*, 2005), thus not permitting subsequent whole-genome amplification procedures.

Following single-cell separation, the next step is cell lysis to provide access to the genomic material, and subsequent multiple displacement amplification (MDA) (Lasken, 2007). MDA relies on the Phi29 polymerase enzyme that

amplifies randomly primed template DNA in an isothermal reaction with very high efficiencies. Because of the strand displacement activity of Phi29 polymerase, newly synthesized DNA becomes directly accessible for the next polymerase molecule, thus resulting in continuous DNA amplification. For a more technical description of the MDA process, laboratory protocols, and a list of commercially available kits, the interested reader is referred to several reviews (Lasken *et al.*, 2005; Silander and Saarela, 2008).

The next step, the whole-genome sequencing of MDA products, has been approached by a variety of sequencing methods. Most studies thus far have relied on 454 pyrosequencing using shotgun or paired end libraries (*e.g.,* Marcy *et al.*, 2007b; Mussmann *et al.*, 2007; Blainey *et al.*, 2011; Siegl *et al.*, 2011). However, hybrid sequencing of 454 and Illumina techniques have been shown to produce better genome coverage than a single sequencing technique (Rodrigue *et al.*, 2009). Combinations of 454 and Sanger sequencing have also been used (Woyke *et al.*, 2009), and the first complete single-cell-derived genome was constructed using Sanger, 454, and Illumina techniques (Woyke *et al.*, 2010). As sequencing techniques are constantly evolving, novel approaches such as the PacBio (Pacific Biosciences of California, Inc.) or IonTorrent (Torrent Systems, Inc.) systems might enable even better draft genome recovery from single cells.

Technical Challenges

While the single-cell technology opens a major window of opportunities into symbioses research, the methodology is still subject to various technical challenges, which are detailed below.

Contamination

Contamination with non-target cells or DNA is one of the major challenges of the single-cell approach. Because single-cell whole-genome amplification *via* MDA is random hexamer primed, any piece of DNA in the reaction mix will co-amplify and compete for amplification with the low fg-range target DNA, if of sufficient length. Contaminating sequences not only reduce sequencing efficiency, but also may significantly confound the analysis of novel single-cell genomes. Contamination can be process-introduced or sample-introduced.

To prevent process contamination, the most stringent decontamination procedures are needed (Table 1). It is best practice to bleach-sterilize work areas and laboratory equipment and UV-irradiate all disposables, as well as buffers and water utilized within the single-cell pre-MDA work flow (Stepanauskas and Sieracki, 2007; Rodrigue *et al.*, 2009). Even with such stringent preventive measures, process-introduced contamination is a rather common phenomenon, largely due to the presence of contaminants such as *Delftia acidovorans* in MDA reagents. As none of the commercial MDA reagents available to date are designed for single-cell applications but rather for the amplification of ng-range DNA or for many hundreds of cells, there has been no incentive or need to provide completely DNA-free reaction components.

Table 1

Current technical challenges and potential solutions in single-cell genomcics

Challenge	Potential Solution
Contamination	• Bleach-sterilize work area and equipment • UV-irradiate all disposables and reagents • Perform two cycles of cell sorting (if possible) • Extensively rinse the microbial cell (*i.e.,* after micromanipulation) • Remove known contaminants (*i.e.,* Delftia, human) by binning methods post-MDA
Limited lysis	• Chemically lyze *via* alkaline solution (KOH) • Enzymatically lyze using lysozyme and/or proteases or custom-made enzyme cocktails • Freeze/thaw, heat, *etc.*
Amplification bias	• Reduce reaction volume, *i.e.,* by addition of crowding agents • Normalize sequencing libraries • Digitally normalize post-sequencing
Chimerism	• Avoid long-mate pair libraries
Fragmented and partial nature of genomes	• Pool individual single amplified genomes representing the same operational taxonomic unit (when applicable)

To circumvent this issue, UV-irradiation of MDA reagents including Phi29 has been used successfully to minimize the co-amplification of free bacterial DNA found in commercial reagents during single-cell MDA (Woyke *et al.*, 2011). Moreover, it is possible to prepare ultra-pure Phi29 in house by using affinity-purification of recombinant Phi29 DNA (Blainey and Quake, 2011).

Sample-introduced contamination represents a slightly different challenge. If single cells are to be isolated using micromanipulation, where individual cells are transferred in rather large volumes, extensive rinsing of single cells by repetitive transfer to clean buffer drops on a slide may be helpful in shedding free DNA or additional small cells. However, this may not suffice, as seen in the *Sulcia* single-cell sequencing project, where more than 40% of the sequence reads were probably host-derived, despite extensive rinsing (Woyke *et al.*, 2010). Isolating single cells using fluorescence-activated cell sorting (FACS) has been shown to eliminate the carry-over of exogenous DNA by minimizing the transfer volume (droplet volume of 10 pl). Rodrigue and colleagues (2009) nicely demonstrated the successful removal of unwanted free DNA by performing two cycles of cell sorting.

Even when maximum preventive measures are taken, a thorough post-MDA quality control is advisable. Direct

Sanger sequencing of 16S rRNA gene PCR products generated for each MDA product has been used to identify the phylogeny of the single-cell genome. In addition, these data can be used to detect a potential contaminant by a thorough review of the Sanger sequencing chromatograms. As the rRNA genes generally represent only a minute fraction of the whole genome (<1%), simply relying on this method has proved rather inefficient. For example, even though no 16S rRNA gene was detected by PCR in the *Sulcia* single-cell project, more than 40% of the sequence data was derived from likely *Delftia* contamination (Woyke *et al.*, 2010). As was expected, the assembled *Delftia* contigs did not encode the 16S rRNA gene.

Post-sequencing, the bioinformatic detection and removal of common contaminants such as *Delftia* and human may be simple, but it can also be challenging if there is no clear discrimination between the target organism and the contaminant. Even when a combination of binning methods such as nucleotide signatures and phylogenetic assignments based on Blast analysis are applied, the distinction between true biological data associated with the single cell (*i.e.*, horizontal gene transfer, plasmids, phage infecting the single cell) and a potential contaminant may be blurred, particularly in highly fragmented assemblies or genomes that lack any near neighbor within the sequence database. For symbiont single-cell genomes inhabiting hosts that do not have a closely related sequenced representative, the confident identification of potential host-derived sequence can be a major challenge.

Limited lysis

Lysis is a key step in single-cell whole-genomes amplification, because it exposes the genomic DNA to make it accessible for amplification. As some cells may harbor only a single copy of their genome, lysis should be gentle so as to fully maintain the integrity of the DNA. Extensive nicking of the genomic DNA and, even more so, introduction of dsDNA breaks leads to complete loss of the linkage information at these genomic sites. Thus the lysis method of choice must be mild, yet harsh enough to enable access to the genomes for the majority of single cells. In an ideal experiment, lysis should be accomplished for each of the single cells isolated, providing access to the genetic make-up for every cell. To date, however, no universal lysis method for all taxa exists.

Cell lysis methods can generally be categorized as chemical, enzymatic, and physical (including acoustic such as sonication, and optical and mechanical), each having their weaknesses and strengths and suitability for prokaryotic single-cell genomics. The currently most common lysis method for a single bacterial cell is chemical lysis *via* alkaline solution (KOH). While easily applicable, this method for opening a cell has a success rate of only about 20% for various environments. Enzymatic lysis methods make use of cell-wall-cleaving enzymes such as lysozyme in combination with proteases (Tamminen and Virta, 2010; Fleming *et al.*, 2011). Although this is a gentle lysis method, vast variations in cell wall properties among different organisms render it unlikely to be universally applicable. Cocktails combining cell-wall-active enzymes with differing specificities such as lysozyme, achromopeptidase, mutanolysin, and lysostaphin may be a viable solution, albeit thorough decontamination may be a necessity. Alternative methods that have been applied or suggested for single-cell genomics include physical methods such as freeze/thawing, heat, and combinations of the various methods discussed here (Kvist *et al.*, 2007; Mussmann *et al.*, 2007; Siegl *et al.*, 2011). For a comprehensive review of single-cell lysis techniques, please refer to Brown and Audet (2008).

Amplification bias

An array of different approaches to reduce amplification bias have emerged over the years, with the most promising result shown by Marcy *et al.* (2007b). The authors demonstrated that a reduction in reaction volumes (60 nl as compared to 50 μl) greatly reduces amplification bias. While sub-microliter-scale amplifications as achieved by microfluidic chips is not practical for many laboratories, crowding agents mimicking smaller reaction volumes, such as PEG and trehalose (Ballantyne *et al.*, 2006; Pan *et al.*, 2008) may be easier to implement into a laboratory's single-cell workflow (Table 1). Additional methods to deal with uneven genome coverage are applied post-MDA. Normalized libraries can be generated using protocols based on subtractive hybridization with the usage of duplex-specific nuclease (Rodrigue *et al.*, 2009; Swan *et al.*, 2011). The major drawback of normalized libraries is that these are labor-intensive and, unless they are adapted to a plate format, not suitable for high-throughput single-cell sequencing projects. Lastly, digital normalization, which informatically reduces the sequence information for over-represented regions of the genome prior to assembly, has proved to be of tremendous value for single-cell sequence data (Rodrigue *et al.*, 2009; A. Sczyrba, University of Bielefeld, pers. comm.)

Chimerism

The nature of MDA introduces chimeric rearrangements, which become apparent after sequencing and are found on the order of one chimeric junction every 20 kbp (Lasken and Stockwell, 2007; Marcy *et al.*, 2007b; Woyke *et al.*, 2009). Although some reports have shown the reduction of chimerism due to S1 nuclease treatment (Zhang *et al.*, 2006), other laboratories have failed to evidence this effect (Woyke *et al.*, 2009). The Lasken laboratory extensively characterized the types of chimera formed during MDA, determining that the majority (~85%) of these rearrange-

ments are inversion/deletion events (Lasken and Stockwell, 2007). Such artifacts can challenge assembly algorithms, and manual assembly curation may be required to break chimeric contigs. It is advisable to avoid long-mate pair libraries, which will provide a high percentage of incorrect pairing information (Table 1).

Fragmented and partial nature of single-cell genomes

Single-cell genomes sequenced to date range from partial genomes of a few hundred kilobases in assembly size (Youssef *et al.*, 2011) to a finished genome (Woyke *et al.*, 2010). The majority of single-cell genomes recovered with today's methodologies will likely be fragmented and partial in nature, rather than resembling a truly complete genome (Table 1). We suspect that the completely finished *Sulcia* single-cell genome will, at least for now, remain an exception rather than the rule (Woyke *et al.*, 2010). This "drafty" nature of most single-cell-derived genomes, in addition to the bias and chimera issues discussed above, makes the bioinformatic analysis of the data less straightforward than sequence data from isolate genomes. As single-cell genomics becomes increasingly popular, various tools specifically designed for genome data from single cells have become available. Over the last few years, several software packages for single-cell assembly have been released that address the problem of highly variable coverage rate in MDA-derived data. SmashCell (Simple Metagenomics Analysis SHell-for sequences from single Cells) is a software framework that combines assembly, gene prediction, and annotation of single-cell data (Harrington *et al.*, 2010). Assemblers that followed were IDBA-UD (Peng *et al.*, 2012) and Velvet-sc. (Chitsaz *et al.*, 2011), and most recently, the novel single-cell-specific assembler called SPAdes, developed by the Pevzner group (Bankevich *et al.*, 2012).

Automatic annotation and its manual refinement can be very challenging when dealing with fragmented single amplified genomes (SAGs), especially for cells with no closely related reference genomes available. Drafty single-cell genomes might not provide the necessary genomic context to securely annotate a gene, or the gene of interest itself is fragmented and thus cannot be annotated with high confidence. This in turn affects the ability to predict the existence of metabolic pathways, as certain key enzymes might be missing from the single-cell genome or are not clearly annotated. Thus tools generally used for comparative genomics are to be used with caution for single-cell data, and the analysis of single cells may be limited. This is a challenge that at present cannot readily be addressed with mere bioinformatics but might be overcome through the steadily increasing amount of available genomic data as well as further improvements in the recovery of single-cell genomes from the environment.

A current strategy to improve assembly and genome recovery for single cells is the pooling of individual SAGs representing the same operational taxonomic unit (OTU) (Podar *et al.*, 2007; Warnecke and Hugenholtz, 2007; Blainey *et al.*, 2011). This strategy will either await the availability of replicate SAGs within the pool of randomly isolated single cells or require enrichment based, for example, on 16S rRNA probes. Blainey *et al.* (2011) demonstrated elegantly that with the pooling of five SAGs the recovery of more than 95% of a single-cell genome was approached, here that of "*Candidatus* Nitrosoarchaeum limnia SFB1," an ammonia-oxidizing archaeon. Although this is a fine strategy if the environmental population is clonal, it becomes more challenging in a heterogeneous population. Pre-binning of the SAG data using average nucleotide identity (ANI) (Konstantinidis *et al.*, 2006) prior to co-assembly of these datasets may be beneficial in this regard.

Although many laboratories have been working on improving some of the key aspects of single-cell genome sequencing in an attempt to improve the quality of the recovered genomes, there is still major room for progress. A goal to aim for would be the complete recovery of each single cell isolated; small steps toward this goal will aid in moving the field of single-cell genomics to the next level. In the meantime, even a few hundred kilobases from a single cell can be of tremendous value by giving insight into the coding potential of microbial dark matter and providing a long-needed link between phylogeny and function for this uncultured majority.

Application of Single-Cell Genomics to Symbioses

In the following section we introduce representative single-cell genomics studies in a host-associated context. This method has now been applied to a handful of symbiotic systems, including bacterial symbionts of marine sponges, insects (grasshoppers, termites), and vertebrates (mouse, human) (Fig. 2). We give an overview of each experimental system and show what contributions single-cell genomics has made to the corresponding field.

Poribacterial symbiont of marine sponges

Many marine sponges contain massive amounts of microorganisms within their mesohyl matrix, which can contribute up to 35% of the animal's biomass (Hentschel *et al.*, 2006, 2012; Taylor *et al.*, 2007; Webster and Taylor, 2012). Members of at least 30 bacterial phyla and both archaeal lineages were found by high-throughput sequencing technologies within sponge hosts (Webster *et al.*, 2010; Schmitt *et al.*, 2012). The candidate phylum Poribacteria is among the predominant microorganisms in these microbial consortia (Lafi *et al.*, 2009; Schmitt *et al.*, 2011). As with most sponge-associated microorganisms, little is known about the function that *Poribacteria* might play in

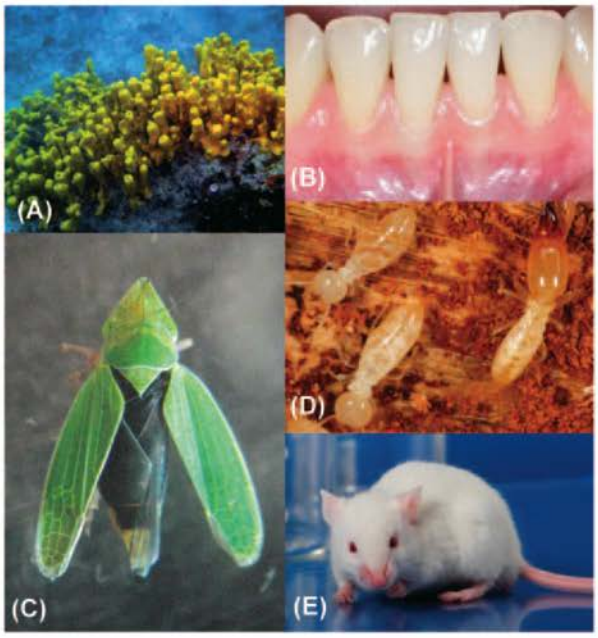


Figure 2. Host animals of symbionts, for which single-cell genomics has been employed: the marine sponge *Aplysina aerophoba* (A), human mouth (B), the green sharpshooter *Draeculacephala minerva* (C), the termite *Reticulitermes speratus* (D), and the mouse *Mus musculus* (E). Photo credits: Janine Kamke, University of Würzburg (A); Martin Linke, Prophylaxepraxis Ahaus (B); Damon Tighe, DOE Joint Genome Institute (C); Yuichi Hongoh, Tokyo Institute of Technology (D); photographer: Valdek Dmytrowski, photo courtesy of Taconic (E).

this symbiosis. This lack is largely attributed to the fact that none of the sponge symbionts have yet been cultured.

Siegl *et al.* (2011) were able to obtain a single poribacterial cell (Poribacteria WGA A3) from the marine sponge *Aplysina aerophoba* by a customized cell-separation protocol followed by FACS sorting (Fig. 2a, Table 2). Through alkaline lysis, MDA, and 454 pyrosequencing, 105 Mbp of raw sequence were recovered, which assembled to a genome size of 1.88 Mbp. While being somewhat flawed by the fragmented nature of many operons, more than 500 contigs contained at least one complete open reading frame that, in combination with the full dataset, led to a comprehensive analysis of the genomic repertoire of a single poribacterium. This study shows that it is possible to arrive at comprehensive genomic information from single cells, even when collected from exceedingly diverse samples, such as marine sponges.

The poribacterial single-cell genome sequence encoded genes involved in glycolysis, TCA cycle, and oxidative phosphorylation (Fig. 3). Furthermore, a purine and pyrimidine metabolism, pathways for canonical amino acids, were identified, rendering auxotrophy an unlikely event (Siegl *et al.*, 2011). Additionally, nitrite assimilation is highly likely to occur in *Poribacteria*, as indicated by the presence of two assimilatory nitrite reductases.

The single-cell genomics study on *Poribacteria* shows potential for degradative metabolism through the presence of several sulfatases, peptidases, and proteins related to

Table 2

A compilation of published reports in which single-cell genomics have been employed in a symbiosis context

Host system	Marine sponge *Aplysina aerophoba*	Sharpshooter *Draeculacephala minerva*	Termite gut protist *Trichonympha agilis*	Human subgingival crevice	Mouse intestine
Authors	Siegl *et al.*, 2011	Woyke *et al.*, 2010	Hongoh *et al.*, 2008	Marcy *et al.*, 2007b	Pamp *et al.*, 2012
Bacterial symbiont	Candidate phylum Poribacteria, WGA A3	Candidate *Sulcia muelleri* DMIN	Candidate phylum Termite group 1, Rs D-17	Candidate phylum TM7	Candidate clade *Arthromitus*, Clostridiaceae
Microbial diversity	Very high	Low	High	High	SFB monocolonized
Cell isolation method	FACS	Micromanipulation	Micromanipulation	Microfluidic chip	Laser tweezers and microfluidic chip
Coverage	~75% of a single genome	Closed single-cell genome	Closed composite genome	Not estimated	~99% of each of five genomes
Genome data	~1.88 Mbp assembled	0.24 Mbp	1.13 Mbp	2.86 Mbp	1.28-1.50 Mbp
Sequencing method	454 pyrosequencing	Sanger sequencing, 454 pyrosequencing, Illumina sequencing	Sanger sequencing, 454 pyrosequencing	454 pyrosequencing	Multiplex 454 pyrosequencing

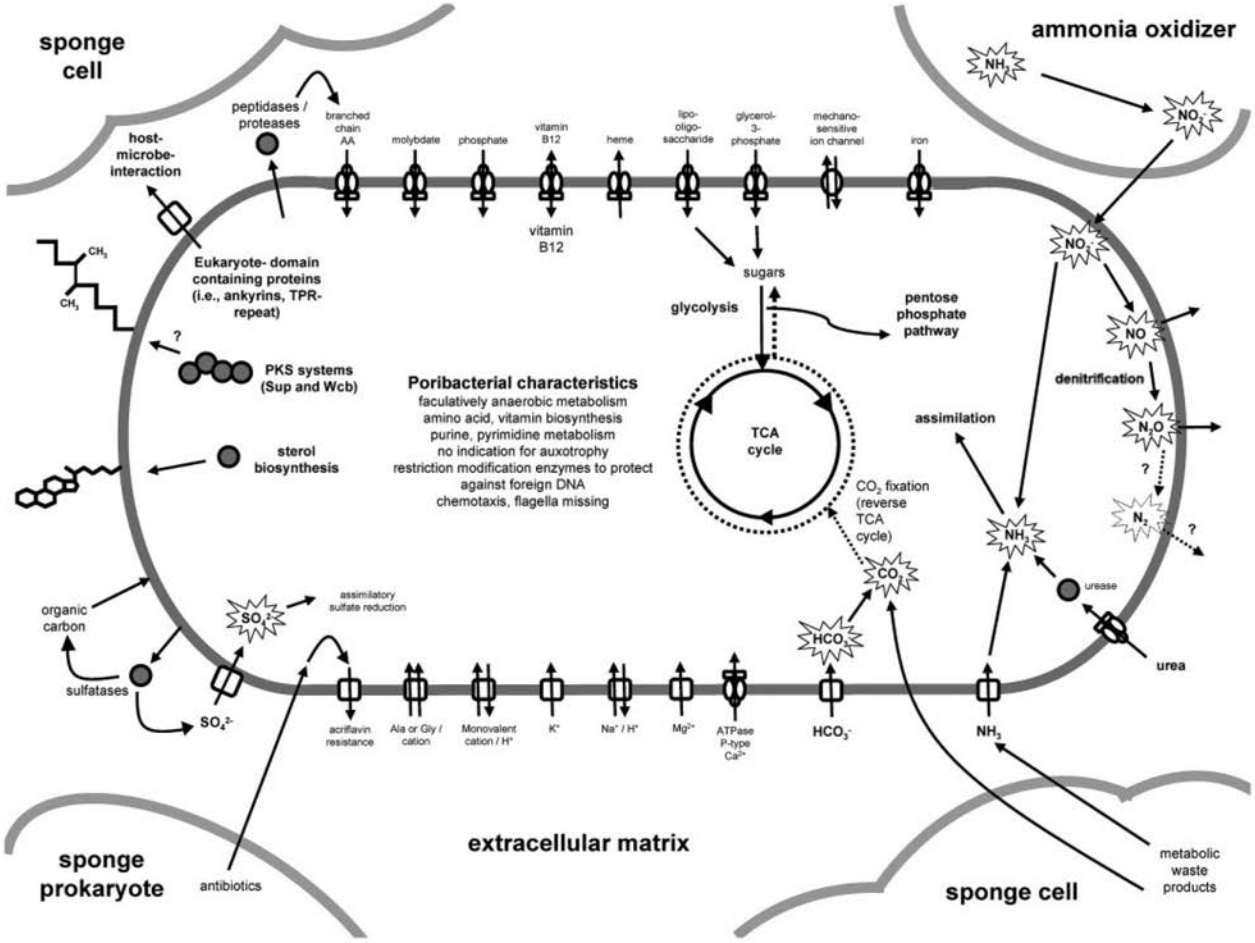


Figure 3. Predicted genomic features of poribacterial symbionts from marine sponges. Modified after Siegl *et al.* (2010), ISME J 5: 61–70, with kind permission by Nature Publishing Group.

N-glycan degradation (Siegl *et al.*, 2011). Degradation of complex substances was also suggested by Liu, M., *et al.* (2011) for another bacterial symbiont of *Cymbastella concentrica.* Furthermore, Hallam *et al.* (2006) reported on several proteases that might degrade extracellular matrix proteins as a defense mechanism on the genome of *C. symbiosum* in *Axinella mexicana.* Whether these degradative enzymes are used in defense or serve a nutritional purpose for the sponge symbionts would be an interesting question for future studies.

There are indications that mechanisms underpinning host-microbe interactions are encoded on the genome of "*Candidatus* Poribacteria WGA A3." These are proteins with eukaryote-like domains, such as ankyrin repeats and tetratricopeptide repeats (TPR), which are likely to mediate protein-protein interactions (Siegl *et al.*, 2011). The latter have been found in all known genomic datasets from prokaryotic sponge symbionts (Hallam *et al.*, 2006; Thomas *et al.*, 2010; Liu, M., *et al.*, 2011), and the expression of proteins with these domains has been confirmed (Liu *et al.*, 2012). Ankyrins are especially likely to be of relevance as they might be involved in the recognition of and protection from host phagocytosis (Liu *et al.*, 2012). This theory is further strengthened by the fact that proteins with ankyrin domains are found in other obligate intracellular pathogenic and symbiotic systems, where they interfere with host cell function (*e.g.,* Mavromatis *et al.*, 2006; Walker *et al.*, 2007; Habyarimana *et al.*, 2008; Voth *et al.*, 2009; Al-Khodor *et al.*, 2010; Murray *et al.*, 2011).

Intracellular bacterial symbiont of a sharpshooter

"*Candidatus* S. muelleri" and "*Candidatus* Baumannia cicadellinicola" are the two obligate symbionts of the green sharpshooter *Draeculacephala minerva* (Fig. 2c). They are localized in the bacteriome, a specialized organelle of the insect that harbors obligate symbionts, and are vertically transmitted *via* the eggs to the next generation (Moran and

Baumann, 2000). *S. muelleri* was found as an intracellular symbiont in several insect species together with *B. cicadellinicola* or other symbiotic bacteria (Takiya *et al.*, 2006; McCutcheon *et al.*, 2009). These insects are feeding on plant sap, which lacks many essential nutrients that are instead supplied by their bacterial symbionts (Redak *et al.*, 2004; McCutcheon *et al.*, 2009). Symbioses with *S. muelleri* are believed to have been established at least 260 million years ago in an ancestor species of today's insects (Moran *et al.*, 2005).

Woyke and coworkers (2010) were able to sequence, nearly base-perfect, the complete genome of *S. muelleri* from *D. minerva* (*S. muelleri* DMIN) (Table 2). They achieved cell separation by dissecting the bacteriome and using light microscopy and micromanipulation to select single cells with the previously described morphology of *S. muelleri*. Through hybrid sequencing of 454 and Sanger techniques it was possible to generate a draft genome, which could be closed with additional Sanger finishing and Illumina polishing. The resulting genome has a size of 243,933 bp. Although the authors faced contamination problems through both MDA reagents and probable exogenous host DNA, they were able to extract *S. muelleri* reads by mapping their data against an existing *S. muelleri* genome from the glassy-winged sharpshooter (*S. muelleri* GWSS) (Wu *et al.*, 2006; McCutcheon and Moran, 2007). The study by Woyke and coworkers (2010) is an example of nearly ideal conditions for a single-cell genomics study because (i) only two bacterial symbionts are housed in the bacteriome, (ii) the target microorganism was morphologically distinct, (iii) the genomes of insect symbionts are extremely reduced and polyploid, thus making genome closure easier to accomplish, and (iv), a reference genome for *Candidatus* Sulcia muelleri was available.

By functional analysis of the genomic data, Woyke *et al.* (2010) found no difference in the metabolic capability encoded in their single-cell-derived genome and available reference sequences (Wu *et al.*, 2006; McCutcheon and Moran, 2007). The *Sulcia* genome encodes first and foremost biosynthesis pathways for eight essential amino acids. The metabolic capacities of *Sulcia* have shown to be complementary with those of the second sharpshooter symbiont "*Candidatus* Baumannia cicadellinicola," which has the genomic potential to produce additional essential amino acids and several vitamins (McCutcheon and Moran, 2007). Comparsion of genes with lower similarity between two *S. muelleri* genomes (DMIN and GWSS) identified potential bacterial surface antigens (Woyke *et al.*, 2010). These proteins might be connected to host specificity, and further investigations could reveal interactions between bacteria and host cells.

To evaluate the quality of the newly obtained genome from the single-cell source, the authors compared it to a metagenomically derived *S. muelleri* genome, which they constructed out of total DNA from 25 bacteriomes from *D. minerva*. The low number of sequence polymorphisms might represent population variations in otherwise conserved data (Woyke *et al.*, 2010). The heterogeneity within the population was further analyzed by detection of single-nucleotide polymorphisms (SNPs) in two independent *S. muelleri* metagenome datasets (a metagenome library of a single bacteriome and the before-mentioned metagenome library of 25 pooled bacteriomes). A low number of SNPs were detected in the latter only. These results suggest a low genetic variety, which is consistent with the stable genome contents and arrangements that have been shown for other primary symbionts of insects (Tamas *et al.*, 2002). *S. muelleri* sequences appear to evolve extremely slowly even when compared to other intracellular symbionts of insects (Takiya *et al.*, 2006).

The study by Woyke *et al.* (2010) proves clearly that obtaining a complete, high-quality genome from a single bacterial cell is possible. Previous data regarding symbiont metabolism were confirmed, and only minor differences between previously sequenced genomes of the same bacterial species were detected, thus supporting the current theory that genetic variety among microbial symbionts is rather low. This analysis reiterates the power of single-cell genomics for the investigation of insect symbiosis.

TG1 symbionts in termite gut protists

Termites are social insects that live mostly on dead plant and wood material. Their ability to digest lignocellulose renders termites and their associated microbial consortia of interest for biofuel production (Weng *et al.*, 2008; Scharf *et al.*, 2011). The eukaryotic and prokaryotic symbionts of termites are phylogenetically diverse, with often several hundred bacterial species being present (reviewed by Hongoh, 2010; Husseneder, 2010). The microbial gut protists are mostly found in lower termites and also harbor bacterial symbionts that aid in the digestion of wood particles (Stingl *et al.*, 2005; Ohkuma, 2008).

Hongoh *et al.* (2008) isolated a single *Trichonympha agilis* protist cell from the termite *Reticulitermes speratus*, *via* micromanipulation (Fig. 2d). They were able to retrieve approximately 10^3 cells of the bacterial phylotype Rs D-17, members of the candidate phylum termite group 1 (TG1). These bacteria are predominant and exclusive to the posterior of the host flagellate *T. agilis*. Multiple displacement amplification enabled the recovery of sufficient DNA for genome sequencing. Through the combination of genomic material from many clonal cells and subsequent whole-genome amplification, it was possible to retrieve a complete, composite genome sequence of Rs D-17 cells, with a size of approximately 1.13 Mbp.

The obtained genome sequence showed evidence for the metabolic adaptation of the bacterial symbiont to an intra-

cellular lifestyle. The Rs D-17 genome still contains several intact pathways for amino acid and cofactor biosynthesis. Several genes belonging to these pathways are duplicates, indicating the importance of those substances for the symbiosis. The eukaryotic partners in this interaction are dependent on the retrieval of amino acids and vitamins from bacterial symbionts, since their lignocellulose-based nutrition does not supply these essential compounds (Husseneder, 2010). In return, the TG1 bacteria are supplied with glucose-6-phosphate (a dominant carbon source) and glutamate (as a nitrogenous compound). Glutamate biosynthesis is disrupted in the bacterial genome but most likely supplied by the protist. The protist might also supply phosphorylated glucose to the symbiont, which helps preserve the symbiont's ATP reservoirs, which are predicted to be produced by fermentation of sugar to acetate. Thus, a complementary metabolism between the symbiotic partners was proposed (Hongoh *et al.*, 2008).

Hongoh *et al.* (2008) further proposed streamlining adaption as a result of reduced genome size, the presence of several pseudogenes, and the presence of duplicated regions of metabolic relevance. This system is thus an excellent subject for further studies of evolutionary pressure on genomes of intracellular symbionts. It would be highly interesting to compare the TG1 symbiont genomes from different protist species, which are known to harbor specific and phylogenetically divergent TG1 bacteria (Stingl *et al.*, 2005; Ohkuma *et al.*, 2007). If genome adaption is still ongoing, it might be possible to investigate both co-evolution with their respective protist hosts and convergent evolution of phylogenetically different TG1 clades.

Candidate phylum TM7 from human gingival crevice

The first study that applied single-cell genomics to host-associated bacteria was conducted on the candidate phylum TM7 from human mouth biofilms (Marcy *et al.*, 2007b). Representatives of the candidate TM7 phylum have been found in biofilms of the subgingvial crevice in healthy humans and also in conjunction with periodonditis (Fig. 2b) (Colombo *et al.*, 2009; Crielaard *et al.*, 2011). The fact that these microorganisms are not highly abundant in the microbiome of the human mouth represents a special challenge when obtaining single cells. Cells were singularized from a biofilm sample using a microfluidic device in combination with light microscopy (Table 2) (Marcy *et al.*, 2007b). This permitted a more targeted selection of cells based on morphological properties. Marcy *et al.* (2007b) sequenced the genomic data from three single TM7 cells (Tm7 a–c), which resulted in datasets of various sizes (TM7 a: 2.86 Mbp assembled data, TM7 b: 10 Mbp unassembled data, TM7 c: 474 kbp assembled data). To ensure analysis with exclusively high-quality data, exclude contamination, and minimize the influence of possible MDA bias, the authors used a very strict quality filtering and binning approach to construct the 963-kbp "TM7 metagenome" out of the combined data of all three cells.

On the basis of this information, the authors predicted the presence of several major pathways in the metagenome, such as glycolysis, TCA cycle, nucleotide biosynthesis, and biosynthesis and salvage pathways for several amino acids (Marcy *et al.*, 2007b). Furthermore, growth of TM7 microorganisms on oligosaccharides and amino acids was indicated, which is consistent with the high nutrient environment of the human mouth. Proteins involved in type IV pilus biosynthesis were also identified, possibly representing a virulence factor (Marcy *et al.*, 2007b). The authors further predicted UDP-*N*-acetylmuramyl tripeptide synthetase to be involved in virulence. In bifidobacteria, this enzyme is involved in peptidoglycan formation, which plays a role in chronic granulomatous inflammation (Simelyte *et al.*, 2003).

This study (Marcy *et al.*, 2007b) is an elegant example of how the single-cell genomics approach can result in novel information about candidate phyla where cultured representatives do not exist. Although the strict binning approach excluded a substantial number of potential TM7 reads, it still provided previously unavailable genomic information about this candidate phylum TM7.

Segmented filamentous bacteria from mouse intestine

Segmented filamentous bacteria (SFB), a specific clade of *Clostridia*, are host-specific symbionts that are present in the lower intestine of many vertebrates. SFB form segmented filaments, which are firmly attached to the epithelia of the host intestine. They are of special interest because they were shown to directly influence the host's immune system (reviewed by Reading and Kasper, 2011). Pamp *et al.* (2012) used a combination of laser tweezers and microfluidic chips to isolate five individual SFB filaments from feces of a SFB-monocolonized mouse (Fig. 2e, Table 2). Data were assembled individually and resulted in greater than 98% genome coverage for each genome. Additionally, two versions of composite genomes including data from all five cells were assembled, one *de novo* and one with an existing reference SFB genome (Prakash *et al.*, 2011).

The presence of nearly complete glycolysis and pentode phosphate pathways indicates metabolic heterotrophy in SFB. Genes involved in the electron transport chain were lacking, leading the authors to postulate substrate-level phosphorylation through phosphoglycerate, pyruvate, and acetate kinases, as well as the production of molecular hydrogen by pyruvate ferredoxin oxidoreductase (Pamp *et al.*, 2012). The genomic potential of SFB also indicates fermentation through several dehydrogenases coupled with substrate oxidation. Extracellular proteases and several

transport systems are also encoded, which might ensure breakdown of larger peptides and uptake of the resulting amino acids. The bacteria seem indeed to be dependent on the uptake of amino acids, vitamins, and co-factors (Pamp *et al.*, 2012). The authors further suggest an anaerobic metabolism for SFB, in which amino acids are taken up and fermented to sustain energy supply, and together with vitamins and co-factors, are used for maintenance of cellular processes.

Four specific protein sequence clusters were identified that distinguish SFBs from other members of the family *Clostridiaceae* 1. Because of their extracellular location and limited catalytic domains, these proteins might be components of surface structures that are involved in niche adaptation (Pamp *et al.*, 2012). The polymorphisms found in the surface proteins suggest the differentiation of SFB among their host animal strains, which may indicate the species (or strain)-specific adaptation of SFB to each host (Kuwahara *et al.*, 2011; Prakash *et al.*, 2011; Sczesnak *et al.*, 2011). Although sequence similarity was quite high between genomes from the same host species (98%–99%), the authors could detect distinct differences between different mice strains, thus supporting the hypothesis of host speciation.

With respect to factors mediating host-microbe interactions, the SFB genomes encode flagella that allow movement within the host mucous layer, or that might be involved in cell adhesion or even cell recognition by the host, which, in turn, can cause an immune response (Pamp *et al.*, 2012). Furthermore, homologs of myosin-cross-reactive antigen (MCRA) were identified, which can trigger an autoimmune disease during *Streptococcus* spp. infections (Wu *et al.*, 2010). MCRAs are also involved in adherence and survival in the host context in other bacteria (O'Flaherty and Klaenhammer, 2010; Malachowa *et al.*, 2011). Additional proteins with potential to mediate microbe-host interactions are ADP-ribosyltransferases (ADPRT). ADPRT are proteins that modify enzymes and are occasionally known as toxins secreted by some pathogenic bacteria. Four types of novel ADPRT that differ in sequence, structure, and location of domains were identified on the SFB genomes (Pamp *et al.*, 2012). Some of these proteins are found close to the phage-related genes and show similarities to toxins from other bacteria.

The SFB genomes also encode for proteins that protect the bacteria from host defenses and ensure their survival. For example, multidrug and toxic compound extrusion (MATE) proteins were found that might protect from antimicrobial substances (Pamp *et al.*, 2012). The authors further identified potential choloylgylcine hydrolases that might aid in the protection from bile acid and O-acetyltransferases and polysaccharide deacetylase with the potential to weaken host lysozyme activity. Furthermore, the genome data suggest protection from oxidative stress by the presence of rubrerythrin and catalases, as well as immune evasion of SFB through catabolism of arginine (Pamp *et al.*, 2012). Finally, the SFB genomes also have the genomic potential to produce a protective extracellular polysaccharide capsule.

This study (Pamp *et al.*, 2012) provided unprecedented insights into the putative mechanisms of interaction between a specific clade of intestinal symbionts and epithelial cells of the mouse intestine. These insights were made possible by having access to an experimentally tractable host model (mouse) in which only the target symbionts were present. The study also benefitted from the availability of SFB reference genomes. As a result, several unique groups of proteins were identified with possible involvement in host-specification, interaction, and symbiont survival. These results have implications for bacterial genome evolution and speciation in the mammalian intestine.

Single-Cell Genomics: Linking Phylogeny With Function

In addition to providing genomic insights into single microbial cells, the method of single-cell genomics can be employed to link phylogenetic identity of a bacterium to a specific function (Swan *et al.*, 2011). Metagenomics has also been successfully used to describe novel microbial gene clusters for secondary metabolites (Kennedy *et al.*, 2010; Piel, 2011). However, in ecosystems with high-microbial diversity, it becomes nearly impossible (the "needle in the haystack" problem) to determine the phylogenetic origin of the sought-after biosynthetic pathway by metagenomics alone (but see exception in Beja *et al.*, 2000). In this context, a combination of single-cell genomics with metagenomics is the best solution (Siegl and Hentschel, 2010; Bayer *et al.*, 2012). The assignment of novel genes and functions to their biological origin is important for heterologous expression studies needed for sustainable metabolite production.

The study by Siegl and Hentschel (2010) aimed to clone secondary metabolite gene clusters from WGA products and to identify the corresponding microbial producers. For this purpose, the microbial consortia of a marine sponge were sorted by FACS and then subjected to WGA. A cosmid library was constructed from the WGA product of a sample containing two bacterial cells, one a member of the candidate phylum Poribacteria and one of a sponge-specific clade of *Chloroflexi*. Library screening led to the genomic characterization of two cosmid clones encoding a polyketide synthase (PKS) and a nonribosomal peptide synthetase (NRPS). PCR screening of WGA products from several additional, FACS-sorted cells supports the assignment of the Sup-PKS gene to *Poribacteria* and the novel NRPS gene to *Chloroflexi*. Here, the single-cell genomics approach has permitted the cloning of entire gene clusters from single

microbial cells of known phylogenetic origin, thus providing a sought-after link between phylogeny and function. One important drawback of this cosmid library approach was, however, the chimeric nature of the cloned WGA product, thus supporting the general impracticability of Phi29-amplified products for heterologous expression studies.

In a recent study, Bayer *et al.* (2012) explored $FADH_2$-dependent halogenase genes in microbial metagenomes and WGA products of FACS-sorted single cells of a marine sponge. Screening of a metagenomic library resulted in four halogenase-bearing clones that could not be taxonomically assigned. In the screened WGA products, 12 reactions were halogenase-positive, representing three distinct clades of these enzymes. For six of these products, a corresponding 16S rRNA gene could be identified for which purity was established by cloning of the respective PCR product and RFLP analysis of at least 32 clones. The WGA was considered to be derived from a single phylotype only if the restriction pattern was uniform. In this determination, deltaproteobacterial, actinobacterial, and poribacterial sponge symbionts were identified as possible producers of the three halogenase clades. The single-cell genomic analysis was the essential technique to allow for the assignment of a given function to specific microbial phylotypes.

Conclusions and Future Perspectives

The advantage of single-cell genomics for obtaining comprehensive information on individual microbial cells is undisputed. These analyses open up the opportunity to investigate sequence heterogeneity within symbionts of a single symbiotic host or host organ (*i.e.,* bacteriome), to delineate genomic differences, and to arrive at hypotheses on genome evolution and microbial speciation using individual microbial cells. Single-cell genomics is furthermore suitable for analysis of symbiosis systems with high microbial diversity in which even the most comprehensive metagenomic sequencing projects cannot address the question of which function belongs to which phylotype, and for investigations of candidate phyla where representatives have not been cultivated. Single-cell genomics has been fully embraced by the scientific community, and the road is clear ahead for it to become a mainstream technique in modern microbiology.

Acknowledgments

Financial support to U. H. was provided by the SFB630-grant TPA5 and by the Bavaria California Technology Center (BaCaTeC). T. W. was funded by the U.S. Department of Energy Joint Genome Institute, Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Literature Cited

Al-Khodor, S., C. T. Price, A. Kalia, and Y. Abu Kwaik. 2010. Functional diversity of ankyrin repeats in microbial proteins. *Trends Microbiol.* **18:** 132–139.

Ballantyne, K. N., R. A. van Oorschot, R. J. Mitchell, and I. Koukoulas. 2006. Molecular crowding increases the amplification success of multiple displacement amplification and short tandem repeat genotyping. *Anal. Biochem.* **355:** 298–303.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski *et al.* 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**: 455–477.

Bayer, K., M. Scheuermayer, L. Fieseler, and U. Hentschel. 2012. Genomic mining for novel $FADH_2$-dependent halogenases in marine sponge-associated microbial consortia. *Mar. Biotechnol.* DOI:10.1007/s10126-012-9455-2.

Beja, O., L. Aravind, E. V. Koonin, M. T. Suzuki, A. Hadd, L. P. Nguyen, S. B. Jovanovich, C. M. Gates, R. A. Feldman, J. L. Spudich, E. N. Spudich, and E. F. DeLong. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. *Science* **289:** 1902–1906.

Binga, E. K., R. S. Lasken, and J. D. Neufeld. 2008. Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J.* **2:** 233–241.

Blainey, P. C., and S. R. Quake. 2011. Digital mda for enumeration of total nucleic acid contamination. *Nucleic Acids Res.* **39:** e19.

Blainey, P. C., A. C. Mosier, A. Potanina, C. A. Francis, and S. R. Quake. 2011. Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6:** e16626.

Brown, R. B., and J. Audet. 2008. Current techniques for single-cell lysis. *J. R. Soc. Interface* **5 (Suppl. 2):** S131–138.

Chen, C. H., S. H. Cho, H. I. Chiang, F. Tsai, K. Zhang, and Y. H. Lo. 2011. Specific sorting of single bacterial cells with microfabricated fluorescence-activated cell sorting and tyramide signal amplification fluorescence in situ hybridization. *Anal. Chem.* **83:** 7269–7275.

Chitsaz, H., J. L. Yee-Greenbaum, G. Tesler, M. J. Lombardo, C. L. Dupont, J. H. Badger, M. Novotny, D. B. Rusch, L. J. Fraser, N. A. Gormley *et al.* 2011. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. *Nat. Biotechnol.* **29**: 915–921.

Colombo, A. P., S. K. Boches, S. L. Cotton, J. M. Goodson, R. Kent, A. D. Haffajee, S. S. Socransky, H. Hasturk, T. E. Van Dyke, F. Dewhirst, and B. J. Paster. 2009. Comparisons of subgingival microbial profiles of refractory periodontitis, severe periodontitis, and periodontal health using the human oral microbe identification microarray. *J. Periodontol.* **80:** 1421–1432.

Crielaard, W., E. Zaura, A. A. Schuller, S. M. Huse, R. C. Montijn, and B. J. Keijser. 2011. Exploring the oral microbiota of children at various developmental stages of their dentition in the relation to their oral health. *BMC Med. Genomics* **4:** 22.

de Jager, V., and R. J. Siezen. 2011. Single-cell genomics: unravelling the genomes of unculturable microorganisms. *Microb. Biotechnol.* **4:** 431–437.

Fan, L., D. Reynolds, M. Liu, M. Stark, S. Kjelleberg, N. S. Webster, and T. Thomas. 2012. Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc. Natl. Acad. Sci. USA* www.pnas.org/cgi/doi/10.1073/pnas.1203287109.

Fleming, E. J., A. E. Langdon, M. Martinez-Garcia, R. Stepanauskas, N. J. Poulton, E. D. Masland, and D. Emerson. 2011. What's new is old: resolving the identity of *Leptothrix ochracea* using single cell genomics, pyrosequencing and fish. *PLoS ONE* **6:** e17769.

Gil, R., F. J. Silva, E. Zientz, F. Delmotte, F. Gonzalez-Candelas, A.

Latorre, C. Rausell, J. Kamerbeek, J. Gadau, B. Holldobler, R. C. van Ham, R. Gross, and A. Moya. 2003. The genome sequence of *Blochmannia floridanus*: Comparative analysis of reduced genomes. *Proc. Natl. Acad. Sci. USA* **100:** 9388–9393.

Habyarimana, F., S. Al-Khodor, A. Kalia, J. E. Graham, C. T. Price, M. T. Garcia, and Y. A. Kwaik. 2008. Role for the ankyrin eukaryotic-like genes of *legionella pneumophila* in parasitism of protozoan hosts and human macrophages. *Environ. Microbiol.* **10:** 1460–1474.

Hallam, S. J., K. T. Konstantinidis, N. Putnam, C. Schleper, Y. Watanabe, J. Sugahara, C. Preston, J. de la Torre, P. M. Richardson, and E. F. DeLong. 2006. Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. *Proc. Natl Acad. Sci. USA* **103:** 18296–18301.

Harrington, E. D., M. Arumugam, J. Raes, P. Bork, and D. A. Relman. 2010. SmashCell: a software framework for the analysis of single-cell amplified genome sequences. *Bioinformatics* **26**: 2979–2980.

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

Hentschel, U., J. Piel, S. M. Degnan, and M. W. Taylor. 2012. Genomic insights into the marine sponge microbiome. *Nat. Rev. Microbiol*. (In press).

Hongoh, Y. 2010. Diversity and genomes of uncultured microbial symbionts in the termite gut. *Biosci. Biotechnol. Biochem.* **74:** 1145–1151.

Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma. 2008. Complete genome of the uncultured termite group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. USA* **105:** 5555–5560.

Husseneder, C. 2010. Symbiosis in subterranean termites: a review of insights from molecular studies. *Environ. Entomol.* **39:** 378–388.

Hutchison, C. A., and J. C. Venter. 2006. Single-cell genomics. *Nat. Biotech.* **24:** 657–658.

Ishoey, T., T. Woyke, R. Stepanauskas, M. Novotny, and R. S. Lasken. 2008. Genomic sequencing of single microbial cells from environmental samples. *Curr. Opin. Microbiol.* **11:** 198–204.

Kalisky, T., and S. R. Quake. 2011. Single-cell genomics. *Nat. Methods* **8:** 311–314.

Kennedy, J., B. Flemer, S. A. Jackson, D. P. H. Lejon, J. P. Morrissey, F. O'Gara, and A. D. Dobson. 2010. Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism. *Mar. Drugs* **8:** 608–628.

Klitgaard, K., L. Molbak, T. K. Jensen, C. F. Lindboe, and M. Boye. 2005. Laser capture microdissection of bacterial cells targeted by fluorescence in situ hybridization. *BioTechniques* **39:** 864–868.

Konstantinidis, K. T., A. Ramette, and J. M. Tiedje. 2006. Toward a more robust assessment of intraspecies diversity, using fewer genetic markers. *Appl. Environ. Microbiol.* **72:** 7286–7293.

Kuwahara, T., Y. Ogura, K. Oshima, K. Kurokawa, T. Ooka, H. Hirakawa, T. Itoh, H. Nakayama-Imaohji, M. Ichimura, K. Itoh, C. Ishifune, Y. Maekawa, K. Yasutomo, M. Hattori, and T. Hayashi. 2011. The lifestyle of the segmented filamentous bacterium: a non-culturable gut-associated immunostimulating microbe inferred by whole-genome sequencing. *DNA Res.* **18:** 291–303.

Kvist, T., B. K. Ahring, R. S. Lasken, and P. Westermann. 2007. Specific single-cell isolation and genomic amplification of uncultured microorganisms. *Appl. Microbiol. Biotechnol.* **74:** 926–935.

Lafi, F. F., J. A. Fuerst, L. Fieseler, C. Engels, W. W. L. Goh, and U. Hentschel. 2009. Widespread distribution of *Poribacteria* in demospongiae. *Appl. Environ. Microbiol.* **75:** 5695–5699.

Lasken, R., A. Raghunathan, T. Kvist, T. Ishøoy, P. Westermann, B. K. Ahring, and R. Boissy. 2005. Multiple displacement amplification from single bacterial cells. Pp. 119–147 in *Whole Genome Amplification*, S. Hughes and R. Lasken, eds. Scion Publishing, Oxfordshire, United Kingdom.

Lasken, R. S. 2007. Single-cell genomic sequencing using multiple displacement amplification. *Curr. Opin. Microbiol.* **10:** 510–516.

Lasken, R. S., and T. B. Stockwell. 2007. Mechanism of chimera formation during the multiple displacement amplification reaction. *BMC Biotechnol.* **7:** 19.

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

Liu, M. Y., S. Kjelleberg, and T. Thomas. 2011. Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. *ISME J.* **5:** 427–435.

Liu, P., R. J. Meagher, Y. K. Light, S. Yilmaz, R. Chakraborty, A. P. Arkin, T. C. Hazen, and A. K. Singh. 2011. Microfluidic fluorescence in situ hybridization and flow cytometry (muflowfish). *Lab Chip* **11:** 2673–2679.

Malachowa, N., P. L. Kohler, P. M. Schlievert, O. N. Chuang, G. M. Dunny, S. D. Kobayashi, J. Miedzobrodzki, G. A. Bohach, and K. S. Seo. 2011. Characterization of a *Staphylococcus aureus* surface virulence factor that promotes resistance to oxidative killing and infectious endocarditis. *Infect. Immun.* **79:** 342–352.

Marcy, Y., T. Ishoey, R. S. Lasken, T. B. Stockwell, B. P. Walenz, A. L. Halpern, K. Y. Beeson, S. M. Goldberg, and S. R. Quake. 2007a. Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3:** 1702–1708.

Marcy, Y., C. Ouverney, E. M. Bik, T. Losekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman, and S. R. Quake. 2007b. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated tm7 microbes from the human mouth. *Proc. Natl. Acad. Sci. USA* **104:** 11889–11894.

Mavromatis, K., C. K. Doyle, A. Lykidis, N. Ivanova, M. P. Francino, P. Chain, M. Shin, S. Malfatti, F. Larimer, A. Copeland *et al.* 2006. The genome of the obligately intracellular bacterium *Ehrlichia canis* reveals themes of complex membrane structure and immune evasion strategies. *J. Bacteriol.* **188:** 4015–4023.

McCutcheon, J. P., and N. A. Moran. 2007. Parallel genomic evolution and metabolic interdependence in an ancient symbiosis. *Proc. Natl. Acad. Sci. USA* **104:** 19392–19397.

McCutcheon, J. P., B. R. McDonald, and N. A. Moran. 2009. Convergent evolution of metabolic roles in bacterial co-symbionts of insects. *Proc. Natl. Acad. Sci. USA* **106:** 15394–15399.

Moran, N. A., and P. Baumann. 2000. Bacterial endosymbionts in animals. *Curr. Opin. Microbiol.* **3:** 270–275.

Moran, N. A., P. Tran, and N. M. Gerardo. 2005. Symbiosis and insect diversification: an ancient symbiont of sap-feeding insects from the bacterial phylum *Bacteroidetes*. *Appl. Environ. Microbiol.* **71:** 8802–8810.

Moran, N. A., J. P. McCutcheon, and A. Nakabachi. 2008. Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* **42:** 165–190.

Murray, J. D., R. R. Muni, I. Torres-Jerez, Y. Tang, S. Allen, M. Andriankaja, G. Li, A. Laxmi, X. Cheng, J. Wen *et al.* 2011. Vapyrin, a gene essential for intracellular progression of arbuscular mycorrhizal symbiosis, is also essential for infection by rhizobia in the nodule symbiosis of *Medicago truncatula*. *Plant J.* **65:** 244–252.

Mussmann, M., F. Z. Hu, M. Richter, D. de Beer, A. Preisler, B. B. Jorgensen, M. Huntemann, F. O. Glockner, R. Amann, W. J. Koopman *et al.* 2007. Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol.* **5:** e230.

O'Flaherty, S. J., and T. R. Klaenhammer. 2010. Functional and phenotypic characterization of a protein from lactobacillus acidophilus involved in cell morphology, stress tolerance and adherence to intestinal cells. *Microbiology* **156:** 3360–3367.

Ohkuma, M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends Microbiol.* **16:** 345–352.

Ohkuma, M., T. Sato, S. Noda, S. Ui, T. Kudo, and Y. Hongoh. 2007. The candidate phylum 'termite group 1' of bacteria: phylogenetic diversity, distribution, and endosymbiont members of various gut flagellated protists. *FEMS Microbiol. Ecol.* **60:** 467–476.

Pamp, S. J., E. D. Harrington, S. R. Quake, D. A. Relman, and P. C. Blainey. 2012. Single cell sequencing provides clues about the host interactions of segmented filamentous bacteria (SFB). *Genome Res.* DOI: 10.1101/gr.131482.111.

Pan, X., A. E. Urban, D. Palejev, V. Schulz, F. Grubert, Y. Hu, M. Snyder, and S. M. Weissman. 2008. A procedure for highly specific, sensitive, and unbiased whole-genome amplification. *Proc. Natl. Acad. Sci. USA* **105:** 15499–15504.

Peng, Y., H. C. Leung, S. M. Yiu, and F. Y. Chin. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* **28**:1420–1428.

Piel, J. 2011. Approaches to capturing and designing biologically active small molecules produced by uncultured microbes. *Annu. Rev. Microbiol.* **65:** 431–453.

Podar, M., C. B. Abulencia, M. Walcher, D. Hutchison, K. Zengler, J. A. Garcia, T. Holland, D. Cotton, L. Hauser, and M. Keller. 2007. Targeted access to the genomes of low-abundance organisms in complex microbial communities. *Appl. Environ. Microbiol.* **73:** 3205–3214.

Prakash, T., K. Oshima, H. Morita, S. Fukuda, A. Imaoka, N. Kumar, V. K. Sharma, S. W. Kim, M. Takahashi, N. Saitou *et al.* 2011. Complete genome sequences of rat and mouse segmented filamentous bacteria, a potent inducer of th17 cell differentiation. *Cell Host Microbe* **10:** 273–284.

Reading, N. C., and D. L. Kasper. 2011. The starting lineup: key microbial players in intestinal immunity and homeostasis. *Front. Microbiol.* **2:** 148.

Redak, R. A., A. H. Purcell, J. R. Lopes, M. J. Blua, R. F. Mizell 3rd, and P. C. Andersen. 2004. The biology of xylem fluid-feeding insect vectors of *Xylella fastidiosa* and their relation to disease epidemiology. *Annu. Rev. Entomol.* **49:** 243–270.

Rodrigue, S., R. R. Malmstrom, A. M. Berlin, B. W. Birren, M. R. Henn, and S. W. Chisholm. 2009. Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS ONE* **4:** e6864.

Scharf, M. E., Z. J. Karl, A. Sethi, and D. G. Boucias. 2011. Multiple levels of synergistic collaboration in termite lignocellulose digestion. *PLoS ONE* **6:** e21709.

Schmitt, S., U. Hentschel, and M. W. Taylor. 2011. Deep sequencing reveals diversity and community structure of complex microbiota in five mediterranean sponges. *Hydrobiologia* **687:** 341–351.

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

Sczesnak, A., N. Segata, X. Qin, D. Gevers, J. F. Petrosino, C. Huttenhower, D. R. Littman, and I. I. Ivanov. 2011. The genome of th17 cell-inducing segmented filamentous bacteria reveals extensive auxotrophy and adaptations to the intestinal environment. *Cell Host Microbe* **10:** 260–272.

Siegl, A., and U. Hentschel. 2010. Pks and nrps gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ. Microbiol. Rep.* **2:** 507–513.

Siegl, A., J. Kamke, T. Hochmuth, J. Piel, M. Richter, C. Liang, T. Dandekar, and U. Hentschel. 2011. Single-cell genomics reveals the lifestyle of poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J.* **5:** 61–70.

Silander, K., and J. Saarela. 2008. Whole genome amplification with phi29 DNA polymerase to enable genetic or genomic analysis of samples of low DNA yield. *Methods Mol. Biol.* **439:** 1–18.

Simelyte, E., M. Rimpilainen, X. Zhang, and P. Toivanen. 2003. Role of peptidoglycan subtypes in the pathogenesis of bacterial cell wall arthritis. *Ann. Rheum. Dis.* **62:** 976–982.

Stepanauskas, R., and M. E. Sieracki. 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc. Natl. Acad. Sci. USA* **104:** 9052–9057.

Stingl, U., R. Radek, H. Yang, and A. Brune. 2005. "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71:** 1473–1479.

Swan, B. K., M. Martinez-Garcia, C. M. Preston, A. Sczyrba, T. Woyke, D. Lamy, T. Reinthaler, N. J. Poulton, E. D. Masland, M. L. Gomez *et al.* 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333:** 1296–1300.

Takiya, D. M., P. L. Tran, C. H. Dietrich, and N. A. Moran. 2006. Co-cladogenesis spanning three phyla: leafhoppers (Insecta : Hemiptera : Cicadellidae) and their dual bacterial symbionts. *Mol. Ecol.* **15:** 4175–4191.

Tamas, I., L. Klasson, B. Canbäck, A. K. Näslund, A.-S. Eriksson, J. J. Wernegreen, J. P. Sandström, N. A. Moran, and S. G. E. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296:** 2376–2379.

Tamminen, M., and M. Virta. 2010. Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells. *Nature Precedings* DOI:http://hdl.handle.net/10101/npre.2010.4761.1

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

Thomas, T., D. Rusch, M. Z. Demaere, P. Y. Yung, M. Lewis, A. Halpern, K. B. Heidelberg, S. Egan, P. D. Steinberg, and S. Kjelleberg. 2010. Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J.* **4:** 1557–1567.

Voth, D. E., D. Howe, P. A. Beare, J. P. Vogel, N. Unsworth, J. E. Samuel, and R. A. Heinzen. 2009. The *Coxiella burnetii* ankyrin repeat domain-containing protein family is heterogeneous, with C-terminal truncations that influence Dot/Icm-mediated secretion. *J. Bacteriol.* **191:** 4232–4242.

Walker, T., L. Klasson, M. Sebaihia, M. J. Sanders, N. R. Thomson, J. Parkhill, and S. P. Sinkins. 2007. Ankyrin repeat domain-encoding genes in the wPip strain of *Wolbachia* from the *Culex pipiens* group. *BMC Biol.* **5:** 39.

Warnecke, F., and P. Hugenholtz. 2007. Building on basic metagenomics with complementary technologies. *Genome Biol.* **8:** 231.

Webster, N. S., and M. W. Taylor. 2012. Marine sponges and their microbial symbionts: love and other relationships. *Environ. Microbiol.* **14:** 335–346.

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

Weng, J. K., X. Li, N. D. Bonawitz, and C. Chapple. 2008. Emerging strategies of lignin engineering and degradation for cellulosic biofuel production. *Curr. Opin. Biotechnol.* **19:** 166–172.

Wilson, A. C. C., P. D. Ashton, F. Calevro, H. Charles, S. Colella, G. Febvay, G. Jander, P. F. Kushlan, S. J. Macdonald, J. F. Schwartz, G. H. Thomas, and A. E. Douglas. 2010. Genomic insight into the amino acid relations of the pea aphid, *Acyrthosiphon pisum*, with its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol. Biol.* **19:** 249–258.

Woyke, T., H. Teeling, N. N. Ivanova, M. Huntemann, M. Richter, F. O. Gloeckner, D. Boffelli, I. J. Anderson, K. W. Barry, H. J.

Shapiro *et al.* 2006. Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443:** 950–955.

Woyke, T., G. Xie, A. Copeland, J. M. González, C. Han, H. Kiss, J. H. Saw, P. Senin, C. Yang, S. Chatterji *et al.* 2009. Assembling the marine metagenome, one cell at a time. *PloS ONE* **4:** e5299

Woyke, T., D. Tighe, K. Mavromatis, A. Clum, A. Copeland, W. Schackwitz, A. Lapidus, D. Wu, J. P. McCutcheon, B. R. McDonald, N. A. Moran, J. Bristow, and J. F. Cheng. 2010. One bacterial cell, one complete genome. *PloS ONE* **5:** e10314.

Woyke, T., A. Sczyrba, J. Lee, C. Rinke, D. Tighe, S. Clingenpeel, R. Malmstrom, R. Stepanauskas, and J.-F. Cheng. 2011. Decontamination of mda reagents for single cell whole genome amplification. *PLoS ONE* **6:** e26161.

Wu, D., S. C. Daugherty, S. E. Van Aken, G. H. Pai, K. L. Watkins, H. Khouri, L. J. Tallon, J. M. Zaborsky, H. E. Dunbar, P. L. Tran, N. A. Moran, and J. A. Eisen. 2006. Metabolic complementarity and genomics of the dual bacterial symbiosis of sharpshooters. *PloS Biol.* **4:** e188.

Wu, H. J., Ivanov, II, J. Darce, K. Hattori, T. Shima, Y. Umesaki, D. R. Littman, C. Benoist, and D. Mathis. 2010. Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* **32:** 815–827.

Yilmaz, S., and A. K. Singh. 2011. Single cell genome sequencing. *Curr. Opin. Biotechnol.* DOI 10.1016/j.copbio.2011.11.018.

Yilmaz, S., M. F. Haroon, B. A. Rabkin, G. W. Tyson, and P. Hugenholtz. 2010. Fixation-free fluorescence in situ hybridization for targeted enrichment of microbial populations. *ISME J.* **4:** 1352–1356.

Youssef, N. H., P. C. Blainey, S. R. Quake, and M. S. Elshahed. 2011. Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone spring, OK, USA). *Appl. Environ. Microbiol.* **77:** 7804–7814.

Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church. 2006. Sequencing genomes from single cells by polymerase cloning. *Nat. Biotechnol.* **24:** 680–686.

3. Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges.

This article was published in the peer-reviewed journal The ISME Journal.

The supplementary information to this article can be accessed at the journal webpage (http://www.nature.com/ismej) and is additionally provided with a CD attached to this PhD thesis.

For documentation of individual contributions to this work and consent of all authors for publication in this thesis please refer to appendix A.

The ISME Journal (2011) 5, 61–70

www.nature.com/ismej

ORIGINAL ARTICLE

Single-cell genomics reveals the lifestyle of *Poribacteria*, a candidate phylum symbiotically associated with marine sponges

Alexander Siegl[1,5], Janine Kamke[1], Thomas Hochmuth[2], Jörn Piel[2], Michael Richter[3], Chunguang Liang[4], Thomas Dandekar[4] and Ute Hentschel[1]

[1]*Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany;* [2]*Kekulé Institute of Organic Chemistry and Biochemistry, University of Bonn, Bonn, Germany;* [3]*Ribocon GmbH, Bremen, Germany and* [4]*Department of Bioinformatics, Biocenter, University of Wuerzburg, Wuerzburg, Germany*

In this study, we present a single-cell genomics approach for the functional characterization of the candidate phylum *Poribacteria*, members of which are nearly exclusively found in marine sponges. The microbial consortia of the Mediterranean sponge *Aplysina aerophoba* were singularized by fluorescence-activated cell sorting, and individual microbial cells were subjected to phi29 polymerase-mediated 'whole-genome amplification'. Pyrosequencing of a single amplified genome (SAG) derived from a member of the *Poribacteria* resulted in nearly 1.6 Mb of genomic information distributed among 554 contigs analyzed in this study. Approximately two-third of the poribacterial genome was sequenced. Our findings shed light on the functional properties and lifestyle of a possibly ancient bacterial symbiont of marine sponges. The *Poribacteria* are mixotrophic bacteria with autotrophic CO_2-fixation capacities through the Wood–Ljungdahl pathway. The cell wall is of Gram-negative origin. The *Poribacteria* produce at least two polyketide synthases (PKSs), one of which is the sponge-specific Sup-type PKS. Several putative symbiosis factors such as adhesins (bacterial Ig-like domains, lamininin G domain proteins), adhesin-related proteins (ankyrin, fibronectin type III) and tetratrico peptide repeat domain-encoding proteins were identified, which might be involved in mediating sponge–microbe interactions. The discovery of genes coding for 24-isopropyl steroids implies that certain fossil biomarkers used to date the origins of metazoan life on earth may possibly be of poribacterial origin. Single-cell genomic approaches, such as those shown herein, contribute to a better understanding of beneficial microbial consortia, of which most members are, because of the lack of cultivation, inaccessible by conventional techniques.
The ISME Journal (2011) **5,** 61–70; doi:10.1038/ismej.2010.95; published online 8 July 2010
Subject Category: microbe–microbe and microbe–host interactions
Keywords: single-cell genomics; sponge; symbiosis; *Poribacteria*; marine

Introduction

Whole-genome amplification (WGA) based on phi29 polymerase-mediated multiple displacement amplification (Dean *et al.*, 2001, 2002) holds great promise in the field of microbial ecology (Hutchison and Venter, 2006; Binga *et al.*, 2008; Ishoey *et al.*, 2008). The multiple displacement amplification reaction results in a billion-fold amplification of DNA even from femtogram amounts, thus making it possible to generate genomic DNA suitable for shotgun sequencing from single microbial cells. This technique is especially well suited for the study of environmental microbial samples in which $<1\%$ of microorganisms have been cultivated, and are thus not accessible by conventional techniques (Rappe and Giovannoni, 2003). Since the introduction of single-cell genomics to microbiology (Raghunathan *et al.*, 2005), few studies have succeeded in the reconstruction of genomes from single uncultivated bacterial cells (Marcy *et al.*, 2007; Mussmann *et al.*, 2007; Rodrigue *et al.*, 2009; Woyke *et al.*, 2009; Woyke *et al.*, 2010). The powerful combination of multiple displacement amplification with new sequencing technologies can thus provide comprehensive genomic insights into the metabolic capacities and environmental adaptations of single microbial cells independent of their cultivation.

Sponges (phylum Porifera) are the most ancient of multicellular animals with a fossil record dating back at least 580 million years (Li *et al.*, 1998). In

Correspondence: U Hentschel, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Julius-von-Sachs-Platz 3, Wuerzburg 97082, Germany.
E-mail: ute.hentschel@uni-wuerzburg.de
[5]Current address: Department of Microbial Ecology, University of Vienna, Althanstr. 14, A-1090 Vienna, Austria.
Received 26 March 2010; revised 17 May 2010; accepted 19 May 2010; published online 8 July 2010

addition to their evolutionary significance, many marine sponges are noteworthy for their intimate associations with diverse and dense microbial communities (Taylor *et al.*, 2007a, b). Members of 22 bacterial phyla, the candidate phylum *Poribacteria*, and two archaeal lineages were identified in sponges so far, many of them forming monophyletic, sponge-specific 16S rRNA sequence clusters (Hentschel *et al.*, 2002; Webster *et al.*, 2009). The majority of these phylogenetically highly diverse and yet sponge-specific microbial consortia remain unculturable. It is generally believed that symbionts contribute significantly to the host's metabolism and biochemical repertoire and possibly also to the evolutionary success of their ancient hosts.

Of special interest is a novel candidate phylum termed '*Poribacteria*' (Fieseler *et al.*, 2004), which belongs to the *Planctomycetes–Verrucomicrobia–Chlamydiae* superphylum (Wagner and Horn, 2006). *Poribacteria* were thus far identified in more than a dozen sponge species from different oceans, suggesting their widespread distribution (Lafi *et al.*, 2009). Vertical transmission, a hallmark of ancient symbioses, of the *Poribacteria* to the next sponge generation through the reproductive stages has been demonstrated (Schmitt *et al.*, 2008; Webster *et al.*, 2009). Although they were long found exclusively in sponges, some recent studies have reported on the presence of poribacterial DNA in seawater (Pham *et al.*, 2008; Webster *et al.*, 2009). Fluorescence *in situ* hybridization studies using *Poribacteria*-specific 16S rRNA-targeting probes showed high abundances of members of this phylum in the mesohyl extracellular matrix and in sponge larvae (Fieseler *et al.*, 2004; Webster *et al.*, 2009). The unusual ring-shaped appearance of poribacterial fluorescence *in situ* hybridization signals is highly suggestive of cell compartmentation. Indeed, nucleoid-containing bacterial cells were previously visualized by electron microscopy studies of the sponge mesohyl, and the presence of DNA in the compartment was confirmed by immunogold labeling (Fuerst *et al.*, 1999). In this context, the finding of membrane coat-like proteins, which were previously considered exclusive to eukaryotes, in several members of the *Planctomycetes–Verrucomicrobia–Chlamydiae* superphylum and their localization to inner membranes of the planctomycete *Gemmata obscuriglobus* is especially noteworthy (Santarella-Mellwig *et al.*, 2010). The first functional genomic information on *Poribacteria* was obtained by library construction from the microbial consortium of the sponge *Aplysina aerophoba* (Fieseler *et al.*, 2006; Siegl and Hentschel, 2009). In the present study, we applied WGA to a single poribacterial cell that was obtained from the sponge *A. aerophoba* by fluorescence-activated cell sorting. This strategy provides nearly 1.9 Mb (Table 1) of poribacterial sequence information, and thus reveals unprecedented insights into the lifestyle of this elusive bacterial phylum. Comprehensive genomic insights regarding poribacterial primary and secondary metabolism, structure of the cell envelope, interaction of *Poribacteria* with the sponge host and the putative significance of *Poribacteria* for metazoan evolution are presented.

Table 1 General information about the 454 FLX run and assembly of a poribacterial SAG

Number of reads	456 103
Average read length	230 bp
Base pairs sequenced	105 233 784
All contigs (>200 bp)	
Number of contigs	1597
Number of bases	1 884 861
Large contigs (>500 bp)	
Number of contigs	554
Number of bases	1 592 880
Average contig size	2875 bp
Largest contig size	46 669 bp
Data set I (contigs with ⩾1 partial ORF)	
Number of contigs	455
Number of bases	1 503 489
Data set II (contigs with neither start nor stop codon)	
Number of contigs	1142
Number of bases	381 372

Abbreviations: ORF, open reading frame; SAG, single amplified genome.

Materials and methods

Sponge microbial community extraction and single-cell sorting

The marine sponge *A. aerophoba* (Verongida, Aplysinidae) was collected in April 2008 by scuba diving offshore Rovinj, Croatia (45°05′N, 13°38′E) at depths of 2–15 m. Sponge-associated prokaryotes were extracted from freshly obtained sponges following an established protocol based on tissue disintegration and differential centrifugation (Fieseler *et al.*, 2004). Purified sponge-associated prokaryotes were subsequently used for single-cell sorting using the fluorescence-activated cell sorting Vantage SE flow cytometer with FACSDiVa option (Becton Dickinson, Heidelberg, Germany) as described previously (Siegl and Hentschel, 2009).

Whole-genome amplification

Fluorescence-activated cell sorting-isolated single microbial cells were disrupted by three cycles of freezing and thawing (−80 and 70 °C; 10 min each) and additional alkaline lysis (Lasken *et al.*, 2005). WGA reactions based on phi29 polymerase-mediated multiple displacement amplification were performed using the REPLI-g Midi kit (Qiagen, Hilden, Germany) according to the manufacturer's recommendations. For the identification of single amplified genomes (SAGs), amplicons were subjected to PCR screenings targeting rRNA genes

(bacterial 16S, archaeal 16S, poribacterial 16S, eukaryotic 18S) with subsequent cloning, restriction fragment length polymorphism and sequencing as described previously (Siegl and Hentschel, 2009).

Pyrosequencing of a SAG
The SAG of a member of the candidate phylum *Poribacteria* was treated with S1-nuclease as recommended (Zhang *et al.*, 2006). Digestion was terminated by adding 2 µl 0.5 M EDTA and heating (at 70 °C for 10 min). The amplified DNA was purified using the QIAamp DNA Mini kit (Qiagen) following the manufacturer's supplementary protocol. Eluted DNA was then subjected to spectrophotometrical DNA quantification. Pyrosequencing was performed by AGOWA GmbH, Berlin, Germany, using the Genome Sequencer FLX System (Roche, Branford, CT, USA). Sequencing reads were assembled using the 454 Newbler assembler version 1.1.03.24.

Sequence annotation
Gene prediction was carried out using the Metagene software (University of Tokyo, Tokyo, Japan; Noguchi *et al.*, 2006). Distributed job processing was performed by a refined version of the GenDB v2.2 system (GenDB, CiBiTec, Bielefeld, Germany) (Meyer *et al.*, 2003). For each predicted open reading frame (ORF), the system retrieves observations from similarity searches against sequence databases NCBI-nr, Swiss-Prot, Kegg (release February 2009), as well as protein family databases Pfam (release 23) and InterPro (release 18, InterProScan 4.3.1). Annotation and data mining was performed by using JCoast (Max Planck Institute for Marine Microbiology, Bremen, Germany), version 1.5 (Richter *et al.*, 2008). Predicted protein-coding sequences were automatically annotated by the MicHanThi software (Max Planck Institute for Marine Microbiology; Quast, 2006). The MicHanThi software predicts gene functions based on similarity searches using the NCBI-nr (including Swiss-Prot) and InterPro database. tRNA genes were identified using tRNAScan-SE (Lowe and Eddy, 1997). The annotation of each protein highlighted within the scope of this study was subjected to manual inspection. For all observations regarding putative protein functions, an *e*-value cutoff of 10^{-4} was considered.

Phylogenetic analysis
Sequence alignment and tree construction were conducted using the ARB program package (Technical University Munich, Munich, Germany; Ludwig *et al.*, 2004) and the SILVA 16S rRNA database, version 93 (Pruesse *et al.*, 2007), with all additionally available poribacterial sequences at the SILVA databases in November 2009. Sequence alignment was performed automatically and manually refined using the ARB alignment tool. Phylogenetic trees were constructed using sequences $\geqslant$1100 bp ($n = 146$) only. Shorter sequences ($n = 8$) were added without changing the tree topology through the parsimony interactive tool in ARB. Trees were rooted with 20 sequences belonging to the *Deltaproteobacteria*. We used three tree construction methods (namely maximum likelihood, maximum parsimony and neighbor joining) in combination with sequence conservation filters for the *Planctomycetes–Verrucomicrobia–Chlamydiae* superphylum and bootstrap analysis (100 resamplings). Trees constructed with different methods were compared and the maximum likelihood tree is shown. All trees showed the same topology on a phylum level.

Accession numbers
This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ADFK00000000. The version described in this paper is the first version, ADFK01000000.

Results and Discussion

SAG sequencing
WGAs of single sponge symbiont cells resulted in the identification of the SAG of a member of the candidate phylum *Poribacteria* (Figure 1). Approximately 28 µg of amplified genomic DNA was obtained as revealed by spectrophotometrical measurement. SAG pyrosequencing generated 105 Mb of raw sequence data, and 1.88 Mb of genomic information organized in 1597 contigs were assembled (Table 1). Two data sets were generated. The first set of 455 contigs harbors at least one ORF per contig and encompasses almost 1.6 Mb of sequence information. A further set of 1142 contigs contains only gene fragments with neither a start nor a stop codon, and therefore these contigs were translated into all six reading frames. References to these partial gene sequences are named by contig name instead of locus tag in this publication.

Contamination is a major concern with the amplification of single genomes. PCR screenings of the amplicon from well A3 using primer pairs targeting different rRNA genes (eukaryotic 18S, bacterial 16S, archaeal 16S; Siegl and Hentschel, 2009) were negative (it must be noted that poribacterial 16S rDNA cannot be amplified with the universal eubacterial primers 27f/1492r because of mismatches in the primer-binding region). After WGA product cleanup, a second round of PCR screenings with the primers 27f/1492r led to the detection of a weak PCR product that was due to *Delftia* DNA, a previously described reagent contaminant (Stepanauskas and Sieracki, 2007). The JCoast genomesDB statistics tool (Richter *et al.*, 2008) showed that DNA contamination is $<0.1\%$ (1378 bp/1503 489 bp) and is therefore considered negligible. The following results further support that only a single poribacterial genome was obtained by

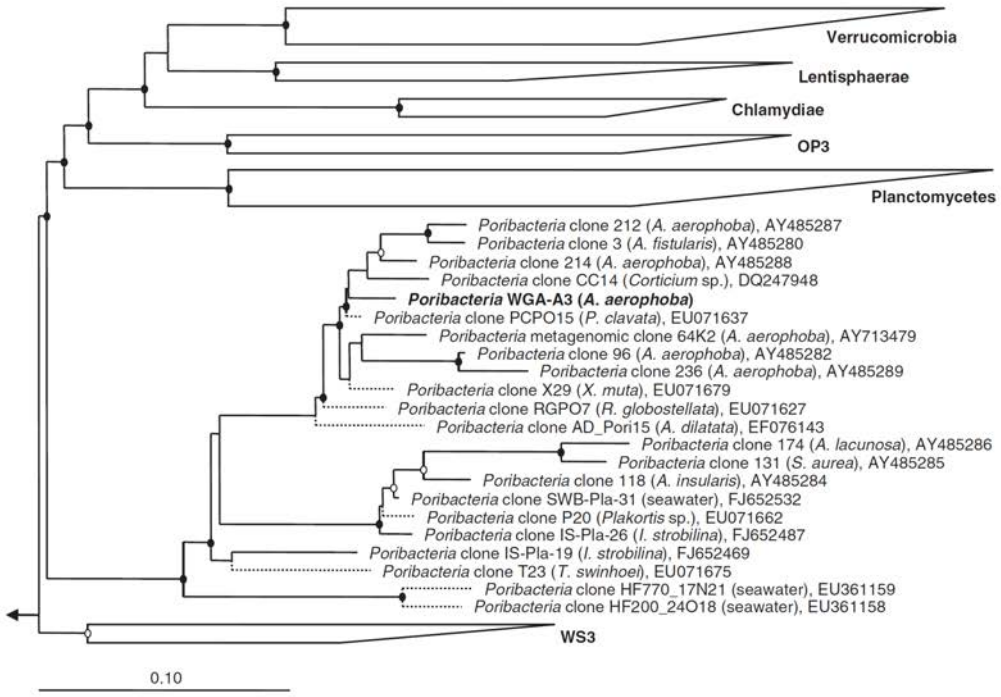


Figure 1 16S rRNA-based phylogenetic tree of the Candidatus phylum *Poribacteria* and the sister phyla within the PVC superphylum. The sequence obtained during this study is shown in bold. Bootstrap support ⩾75% is illustrated as open circles and bootstrap support ⩾90% is indicated by filled circles. Dashed lines indicate sequences <1100 bp, which were added using the parsimony interactive tool in ARB. Sequence sources (marine sponges and seawater, respectively) are given in brackets. Scale bar represents 10% sequence divergence.

WGA: (1) A survey of 55 known single copy genes (Mussmann *et al.*, 2007) showed that none of the 29 detected were present in multiple copies in the 454 data set (Supplementary Table S1). (2) Only a single copy 16S rRNA gene was identified, which might additionally point to a slow growing state of the *Poribacteria*. (3) The calculation of a GC plot with unassembled 454 reads led to a unimodal distribution with a mean of 54%, which is suggestive of a single genotypic sequence source (Supplementary Figure S1).

Table 2 General properties of the assembled *Poribacteria* single amplified genome (SAG)

Genome size	>1 884 861 bp
G+C content	53.41%
Protein coding genes (CDS)	1585 (100%)
Functional conserved proteins	931 (58.7%)
Conserved hypothetical proteins	274 (17.2%)
Hypothetical proteins	380 (23.97%)
Average CDS length	738 bp/246 AA
Coding density	74.72%
Number of tRNAs	19
Number of rRNAs	1 (16S rRNA; partial)

Features were calculated based only on fragments with at least one protein coding gene (CDS; $n = 455$).

General genomic features

The poribacterial genome is at least 1.88 Mb in size (Table 2). Overall, 19 tRNA genes and 14 tRNA synthetase genes and gene fragments were identified with specificities for 14 out of the 20 canonical amino acids, suggesting that approximately two-third of the poribacterial genome sequence was obtained. The average GC content of the poribacterial genome is 53%, which is close to the GC content of the metagenomic poribacterial fosmid clone 64K2 (50%; Fieseler *et al.*, 2006). Altogether, 1585 protein-coding genes were recovered from the 454 assembly, 40% of them with unknown function (Table 2). Approximately 24% are hypothetical proteins without any similarities to database entries (Table 2). This low value reflects their distinct phylogenetic position in the eubacterial tree. Analysis of clusters of orthologous groups (Wheeler *et al.*, 2007) distribution indicated that all three major categories (namely metabolism, cellular processes, information storage and processing) and all groups within these are present in our data set (Supplementary Figure S2).

Primary metabolism

The *Poribacteria* contain the genomic capacity for aerobic, heterotrophic metabolism. Major parts of glycolysis, tricarboxylic acid cycle and the pentose

phosphate pathway (Supplementary Figures S3, S4 and S5), as well as of oxidative phosphorylation (23 complete ORFs) were reconstructed from the 454 data set. Several enzymes involved in purine and pyrimidine metabolism, as well as for all amino-acid biosynthesis pathways are present, thus indicating no auxotrophy at this level. A number of genes coding for degradative metabolism was found. These include nine sulfatases with *N*-acetylglucosamine-6-sulfate, *N*-acetylgalactosamine-6-sulfate, choline, mucine and heparane as putative substrates, as well as several peptidases and other enzymes responsible for the degradation of *N*-glycan. The *Poribacteria* appear as commensalistic symbionts that aid in, and at the same time benefit from the digestion of food particles.

Two intermediate enzymes of the anaerobic respiratory chain, the nitrite reductase (EC 1.7.2.1) and the nitric oxide reductase (EC 1.7.99.7), were identified. These data provide genomic evidence for denitrification in a sponge symbiont, thus endowing the *Poribacteria* with the ability to maintain a functional respiratory chain under oxygen-limiting conditions. Whether the complete reduction from nitrate to N_2 is possible remains to be investigated. With respect to oxygen, the *A. aerophoba* mesohyl is well aerated (Hoffmann *et al.*, 2008). However, microaerobic or even anaerobic microhabitats can principally be generated by active microbial metabolism, that is, when the oxygen demand exceeds the oxygen supply. Furthermore, it has been shown that the *A. aerophoba* mesohyl turns anaerobic during periods of nonpumping. The *Poribacteria* contain two further assimilatory nitrite reductases (EC 1.7.1.4 and 1.7.7.1), thus reflecting their genomic potential to assimilate nitrite, which might be useful under conditions of seasonal ammonia limitation (Bayer *et al.*, 2008). Furthermore, a 10 ORF containing urease gene cluster was identified including various ABC-transporter components, three urease subunits UreA, UreB, UreC, as well as three accessory proteins (Contig c02128). Therefore, the *Poribacteria* are capable of using urea to cover their nitrogen needs. Urea is formed in nature by bacterial degradation of nucleic and amino acids, and is therefore a likely product to be encountered in the sponge mesohyl.

The *Poribacteria* have the genomic repertoire for carbon autotrophy. Autotrophic carbon assimilation has previously been shown for *Cenarchaeum symbiosum*, a crenarchaeote symbiotically associated with axinellid sponges and the only sequenced sponge symbiont genome to date (Hallam *et al.*, 2006a, b). *C. symbiosum* was shown to use a modified 3-hydroxypropionate cycle for carbon

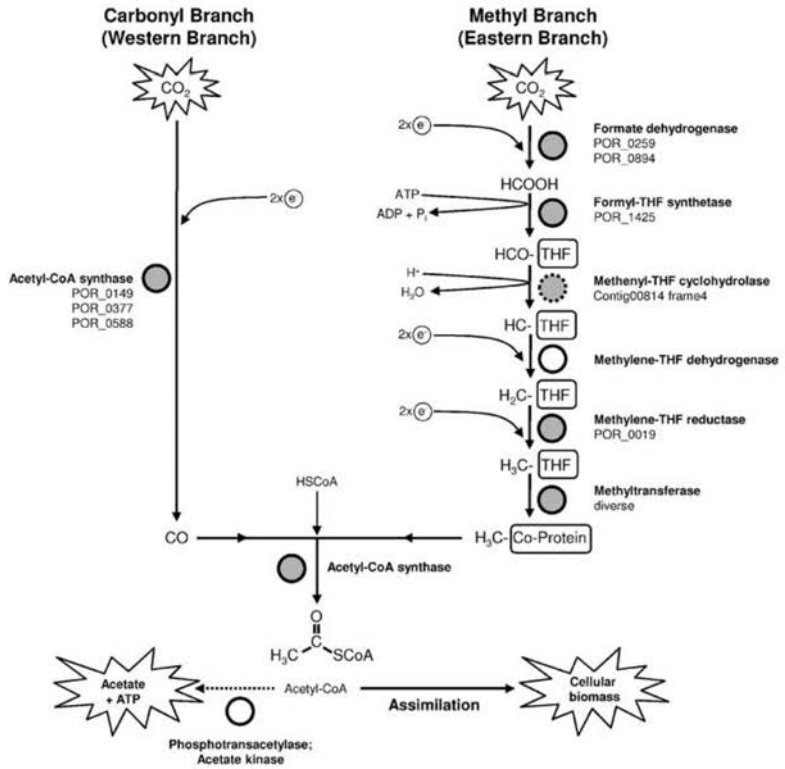


Figure 2 Reconstruction of the Wood–Ljungdahl pathway for carbon fixation. Filled circles represent complete ORFs and dashed circles gene fragments. Open circles indicate enzymes absent from the data set. Pathway modeled after studies by Drake and Daniel (2004) and Ragsdale and Pierce (2008).

fixation (Hallam *et al*., 2006b). Genes encoding for the reductive tricarboxylic acid cycle and the reductive acetyl CoA pathway, also referred to as the Wood–Ljungdahl pathway, were identified on the poribacterial genome (Figure 2). Four partial fragments of the key enzyme of the reductive tricarboxylic acid cycle, citrate lyase (EC number 2.3.3.8), were recovered from the 454 data. All fragments share significant homologies to citrate lyases from *Chlorobi*, which also use the reverse tricarboxylic acid cycle for carbon fixation. The Wood–Ljungdahl pathway constitutes the most basic way of obtaining organic carbon by the joining of two inorganic carbon molecules (Drake and Daniel, 2004; Ragsdale and Pierce, 2008). In evolutionary terms, the Wood–Ljungdahl pathway can probably be regarded as the most ancient metabolic pathway that might already have been present in bacteria inhabiting the anoxic earth several billion years ago (Ragsdale and Pierce, 2008). The key enzyme of the Wood–Ljungdahl pathway, the oxygen-sensitive acetyl-CoA synthase, also referred to as carbon monoxide dehydrogenase, was identified in multiple copies. Further crucial enzymes within the Wood–Ljungdahl pathway are cobalamin-dependent methyltransferases. In total, three of such enzymes were identified in the poribacterial data set (POR_0033, POR_0034 and POR_0549). The best blastp hits of POR_0033 and POR_0034 originate from *Kuenenia stuttgartiensis* and of POR_0549 from *Desulfococcus oleovorans*, both of which are known to be capable of carbon fixation using the Wood–Ljungdahl pathway. All of the three identified methyltransferases possess radical *S*-adenosylmethionine for methylation and B12-binding domains. Interestingly, a gene cluster containing 11 ORFs affiliated with the metabolism of vitamin B12, the most prominent member of cobalamins, was identified on contig c02038. Therefore, it seems likely that the *Poribacteria* are capable of vitamin B12 biosynthesis, which in turn serves as a cofactor for one of the key enzymes (methyltransferase) of the Wood–Ljungdahl pathway.

Polyketide metabolism

The previously known unusually small and phylogenetically distinct Sup-type polyketide synthases (PKSs) are widely distributed in sponge microbiota (Fieseler *et al*., 2007). The poribacterial 454 data set revealed 11 genomic fragments with homologies to the *supA* gene, encoding together ~48% of the entire PKS (Figure 3). Moreover, two accessory members of the Sup-PKS operon were identified in fragments, that is, two *supD* gene fragments and five *supE* gene fragments (Figure 3). The assignment of this elusive Sup-PKS clade to the *Poribacteria*, as has recently been proposed (Siegl and Hentschel, 2009), was thus proven in this study. In addition, a second putative PKS system was identified (Figure 3). Two ORFs (POR_0547 and POR_0548) show high similarities to the lipopolysaccharide type I PKS WcbR from *Nitrosomonas* and *Burkholderia*, as well as to RkpA from *Sinorhizobium fredii* (Parada *et al*., 2006). These PKS types are regularly found in sponge metagenomes (Fieseler *et al*., 2007) and are, in free-living bacteria, usually embedded within lipopolysaccharide gene clusters. Therefore, we can conclude that the *Poribacteria* are the original producers of at least two different types of PKS systems. Interestingly, homologs of both were shown to be major symbiosis and virulence factors in *S. fredii* and *Mycobacterium tuberculosis*, respectively (Minnikin *et al*., 2002; Parada *et al*., 2006), and this raises the intriguing possibility that the poribacterial PKS gene products might fulfill a similarly relevant function in the sponge–microbe interaction.

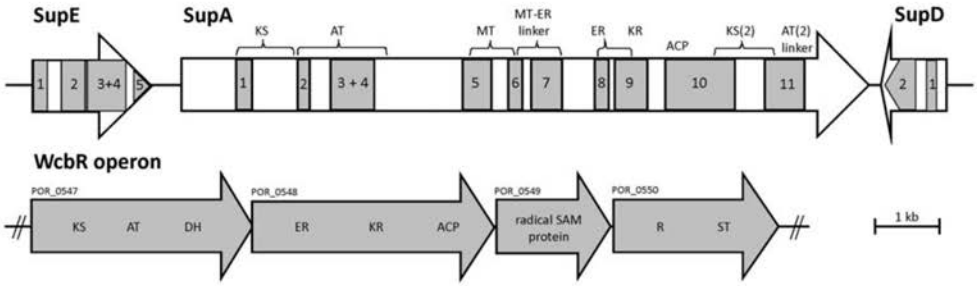


Figure 3 Features of the polyketide metabolism in *Poribacteria*. Display to scale. Abbreviations of PKS functional domains covered: KS, ketosynthase; AT, acyltransferase; DH, dehydratase; ER, enoylreductase; KR, ketoreductase; ACP, acyl carrier protein; R, thioester reductase; ST, sulfotransferase. For both SupE and SupA, two fragments covering the same region were detected (indicated as 3 + 4). This finding could indicate the presence of multiple copies of the Sup genes on the poribacterial genome. Alternatively, a multimodular organization is conceivable like shown for the PKS from an uncultured bacterial symbiont of *Discodermia dissoluta* (SA1_PKSA; AAY00025). Upper part: recovery of gene fragments (gray) affiliated with the Sup-PKS system (SupE, SupA, SupD) mapped against fosmid clone pAPKS18 (white) from *A. aerophoba* (Fieseler *et al*., 2007). Average homologies (positives on the amino-acid level) were as follows: 75% for SupE and 77% for SupD (both compared with pAPKS18 from *A. aerophoba*, Fieseler *et al*., 2007) and 75% for SupA (compared with cosmid clone 1k6 from *A. aerophoba*, Siegl and Hentschel, 2009). Lower part: recovery of a WcbR-like operon, which is a part of the contig c00157 (26,267 bp). POR_0549 is a radical SAM domain protein, and could therefore be an auxiliary PKS domain with methyltransferating (MT) properties.

Sterol biosynthesis

Sterol biosynthesis is almost exclusively restricted to eukaryotes (for exceptions, see Bode *et al.*, 2003; Pearson *et al.*, 2003; Lamb *et al.*, 2007). This evolutionarily ancient process is believed to have evolved very early in earth history shortly after the evolution of oxygenic photosynthesis. The most primitive sterol pathway exists in *G. obscuriglobus* (Love *et al.*, 2009). Our data indicate that *Poribacteria* are capable of sterol formation. Two gene fragments of the key enzyme oxidosqualene cyclase (osc) with homologies to the enzyme from *G. obscuriglobus* were found (Figure 4). Both fragments possess the specific domain TIGR03463 and one of the fragments additionally covers an intact protosteryl cation-binding region (Pearson *et al.*, 2003). As both *G. obscuriglobus* and *Poribacteria*, positioned as sister phyla in phylogenetic trees, most likely possess DNA-containing organelles, it is tempting to speculate that sterols have a key role in the formation of these unusual bacterial structures. In addition, parts of a 24-sterol *C*-methyltransferase were recovered from our data set (Figure 4). Interestingly, these fragments possess homologies only to those of plants and fungi, which is consistent with the finding that C-24 side-chain modifications were so far exclusively found in eukaryotes (Pearson *et al.*, 2003). However, sponges contain a multitude of C-24-alkylated sterols, some of which are established fossil biomarkers for these animals (Figure 4) (Love *et al.*, 2009). One such important molecule is 24-isopropyl-cholestane, which is virtually resistant to bacterial breakdown and has been used to document the presence of *Demospongiae* as early as 635 million years ago (Love *et al.*, 2009). This hydrocarbon is derived from 24-isopropyl-sterols, the biosynthesis of which requires the methylation of the sterol side chain catalyzed by a 24-sterol methyltransferase (Kodner *et al.*, 2008). Interestingly, 24-isopropyl-sterols are so far exclusively found in demosponge groups known to harbor *Poribacteria* (Bergquist *et al.*, 1980). Our findings raise the fascinating possibility that 24-isopropyl steroids may in fact be of poribacterial origin.

Cell envelope

The *Poribacteria* appear to be derived from a Gram-negative ancestor. Besides genes involved in peptidoglycan biosynthesis, lipopolysaccharide biosynthesis genes were identified. The presence of a periplasm is suggested by the observation that five out of six

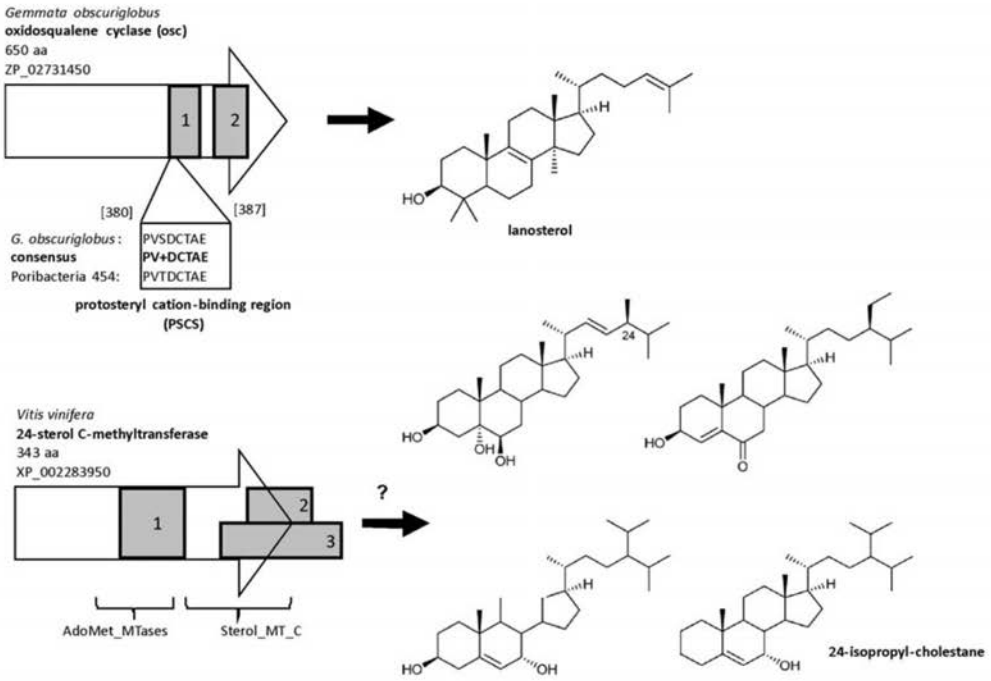


Figure 4 Putative sterol biosynthesis pathway in *Poribacteria*. Upper part: recovery of gene fragments (gray) homologous to oxidosqualene cyclase from *Gemmata obscuriglobus* (white), which is necessary for lanosterol biosynthesis. Average homologies (positives on the amino-acid level) were 57%. Both fragments possess a TIGR03463 domain (2,3-oxidosqualene cyclase). Fragment 1 harbors the protosteryl cation-binding region (PSCS; Bode *et al.*, 2003). Lower part: recovery of gene fragments (gray) homologous to 24-sterol *C*-methyltransferase from *Vitis vinifera* (white). Average homologies (positives on the amino-acid level) were 62%. Fragments 2 and 3 do not share significant similarities on the nucleotide level. This finding could indicate that multiple copies of this gene are present. Conserved domains that are covered by the fragments are displayed. We hypothesize that this enzyme is involved in the formation of the demosponge biomarker 24-isopropyl-cholestane by repeated methylation of carbon C24.

substrate-binding proteins of ABC transporters do not require anchoring. The finding of a TonB-dependent outer membrane receptor (POR_0881), the TonB protein itself (POR_0882) and the preprotein translocase subunit SecA (POR_0883) further indicates the existence of a periplasm. Besides SecA, several additional translocase subunits were found (for example, YajC, SecD, SecF), as well as the Sec-independent Tat-pathway, which is, to our knowledge, known from Gram-negative and Gram-positive bacteria. With respect to cell division, solely the FtsH protease was detected. This finding is in agreement with the genomic survey of the planctomycete *Rhodopirellula baltica* (Glöckner *et al.*, 2003). Finally, genes encoding for flagellar assembly and chemotaxis were entirely missing.

Host–symbiont interaction

Several putative host-interaction factors were recovered, that is, cell surface proteins with bacterial Ig-like domains and laminin G domain proteins, which are possibly involved in adhesion. The existence of eukaryotic domains (such as ankyrin, Sel1, fibronectin type III, leucin-rich repeat) provides further support that interactions between *Poribacteria* and their hosts take place (Taylor *et al.*, 2007a, b). The identification of 23 full or partial genes with tetratricopeptide repeat domains (cl0020) is noteworthy. Interestingly, an abundance of ankyrin and tetratricopeptide repeat proteins was also identified in the microbial metagenome of the sponge *Cymbastela concentrica* (Thomas *et al.*, 2010). These domains are not only involved in eukaryote–prokaryote interactions but also in mitochondrial and peroxisomal transport in eukaryotes. To elucidate the function of these proteins with respect to the poribacterial intracellular compartment and to symbiosis will be a challenge for future studies.

Conclusions

Sequencing of a SAG derived from a member of the candidate phylum *Poribacteria* provided detailed insights into the lifestyle of a possibly ancient bacterium symbiotically associated with marine sponges (Figure 5). With regard to primary

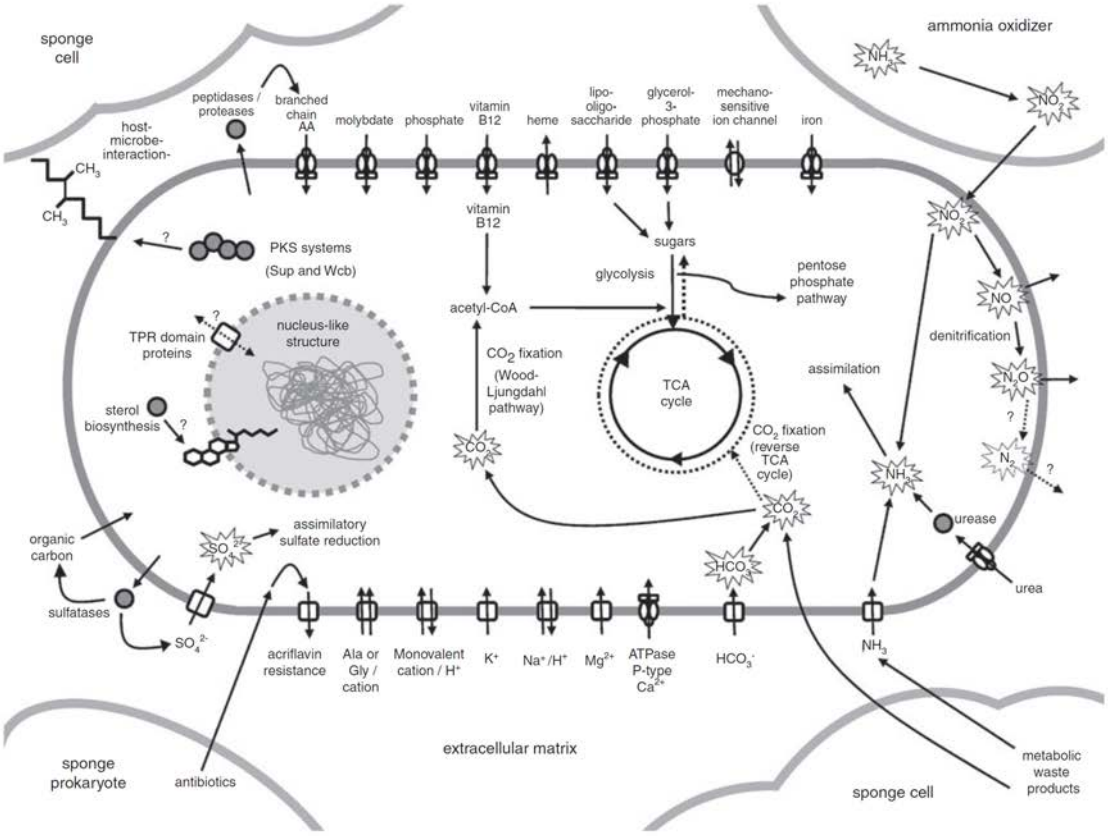


Figure 5 A schematic overview of poribacterial metabolism as deduced from SAG sequencing.

metabolism, the *Poribacteria* are mixotrophic bacteria with autotrophic capabilities of CO_2 fixation through the Wood–Ljungdahl pathway. They are capable of aerobic and anaerobic respiration. The cell wall resembles that of Gram-negative bacteria. Chemotaxis and flagellar genes are absent from our data set. Analysis of secondary metabolism revealed two PKSs that are frequently found in sponge metagenomic libraries, and we can document in this study the first time their unequivocal appointment to *Poribacteria* as the producing microorganism. Several putative symbiosis factors such as adhesins and eukaryotic domain-containing proteins were identified, which may be involved in mediating microbe–sponge interactions. The discovery of sterol biosynthesis genes suggests that sterols may be involved in the function of the poribacterial cellular organelle. Furthermore, the presence of a specific gene encoding for sterol methylation might require reevaluation of fossil sponge biomarkers (24-isopropyl-cholestane) used to date the origins of metazoan life on earth.

Acknowledgements

We gratefully acknowledge C Linden (University of Wuerzburg) for FACS analysis of sponge symbiont cells, R Lasken (C Venter Institute, La Jolla, USA), T Woyke (DOE Joint Genome Institute, Walnut Creek, USA) and B Fartmann (Agowa GmbH, Berlin) for insightful comments. Financial support for this study was provided by the Deutsche Forschungsgemeinschaft grant SFB567-TPC3 to U H and TR34/A5, Da 208/10-1 to CL/TD.

References

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

Bergquist PR, Hofheinz W, Oesterhelt G. (1980). Sterol composition and the classification of the Demospongiae. *Biochem System Ecol* **8**: 423–435.

Binga EK, Lasken RS, Neufeld JD. (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* **2**: 233–241.

Bode HB, Zeggel B, Silakowski B, Wenzel SC, Reichenbach H, Müller R. (2003). Steroid biosynthesis in prokaryotes: identification of myxobacterial steroids and cloning of the first bacterial 2,3(S)-oxidosqualene cyclase from the myxobacterium *Stigmatella aurantiaca*. *Mol Microbiol* **47**: 471–481.

Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P *et al*. (2002). Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci USA* **99**: 5261–5266.

Dean FB, Nelson JR, Giesler TL, Lasken RS. (2001). Rapid amplification of plasmid and phage DNA using Phi 29 DNA polymerase and multiply-primed rolling circle amplification. *Genome Res* **11**: 1095–1099.

Drake HL, Daniel SL. (2004). Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res Microbiol* **155**: 869–883.

Fieseler L, Hentschel U, Grozdanov L, Schirmer A, Wen G, Platzer M *et al*. (2007). Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. *Appl Environ Microbiol* **73**: 2144–2155.

Fieseler L, Horn M, Wagner M, Hentschel U. (2004). Discovery of the novel candidate phylum '*Poribacteria*' in marine sponges. *Appl Environ Microbiol* **70**: 3724–3732.

Fieseler L, Quaiser A, Schleper C, Hentschel U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum *Poribacteria* by environmental genomics. *Environ Microbiol* **8**: 612–624.

Fuerst JA, Webb RI, Garson MJ, Hardy L, Reiswig HM. (1999). Membrane-bounded nuclear bodies in a diverse range of microbial symbionts of Great Barrier Reef sponges. *Mem Queensland Mus* **44**: 193–203.

Glöckner FO, Kube M, Bauer M, Teeling H, Lombardot T, Ludwig W *et al*. (2003). Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci USA* **100**: 8298–8303.

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

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

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

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

Hutchison III CA, Venter JC. (2006). Single-cell genomics. *Nat Biotechnol* **24**: 657–658.

Ishoey T, Woyke T, Stepanauskas R, Novotny M, Lasken RS. (2008). Genomic sequencing of single microbial cells from environmental samples. *Curr Opin Microbiol* **11**: 198–204.

Kodner RB, Summons RE, Pearson A, King N, Knoll AH. (2008). Sterols in a unicellular relative of the metazoans. *Proc Natl Acad Sci USA* **105**: 9897–9902.

Lafi FF, Fuerst JA, Fieseler L, Hentschel U. (2009). Widespread distribution of poribacteria in *Demospongiae*. *Appl Environ Microbiol* **75**: 5695–5699.

Lamb DC, Jackson CJ, Warrilow AG, Manning NJ, Kelly DE, Kelly SL *et al*. (2007). Lanosterol biosynthesis in the prokaryote *Methylococcus capsulatus*: insight into the evolution of sterol biosynthesis. *Mol Biol Evol* **24**: 1714–1721.

Lasken R, Raghunathan A, Kvist T, Ishøy T, Westermann P, Ahring BK *et al*. (2005). Multiple displacement amplification of genomic DNA. In: Hughes, S, Lasken, R (eds). *Whole Genome Amplification: Methods Express*. Scion Publishing: Banbury, pp 119–147.

Li CW, Chen JY, Hua TE. (1998). Precambrian sponges with cellular structures. *Science* **279**: 879–882.

Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, Bradley AS *et al.* (2009). Fossil steroids record the appearance of *Demospongiae* during the Cryogenian period. *Nature* **457**: 718–721.

Lowe TM, Eddy SR. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

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.

Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL *et al.* (2007). Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3**: 1702–1708.

Meyer F, Goesmann A, McHardy AC, Bartels D, Bekel T, Clausen J *et al.* (2003). GenDB—an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**: 2187–2195.

Minnikin DE, Kremer L, Dover LG, Besra GS. (2002). The methyl-branched fortifications of *Mycobacterium tuberculosis*. *Chem Biol* **9**: 545–553.

Mussmann M, Hu FZ, Richter M, de Beer D, Preisler A, Jørgensen BB *et al.* (2007). Insights into the genome of large sulfur bacteria revealed by analysis of single filaments. *PLoS Biol* **5**: e230.

Noguchi H, Park J, Takagi T. (2006). MetaGene: prokaryotic gene finding from environmental genome shotgun sequences. *Nucleic Acids Res* **33**: 5623–5630.

Parada M, Vinardell JM, Ollero FJ, Hidalgo A, Gutiérrez R, Buendía-Clavería AM *et al.* (2006). *Sinorhizobium fredii* HH103 mutants affected in capsular polysaccharide (KPS) are impaired for nodulation with soybean and *Cajanus cajan*. *Mol Plant Microbe Interact* **19**: 43–52.

Pearson A, Budin M, Brocks JJ. (2003). Phylogenetic and biochemical evidence for sterol synthesis in the bacterium *Gemmata obscuriglobus*. *Proc Natl Acad Sci USA* **100**: 15352–15357.

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

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

Quast C. (2006). MicHanThi—design and implementation of a system for the prediction of gene functions in genome annotation projects. Diploma thesis, University of Bremen.

Raghunathan A, Ferguson Jr HR, Bornarth CJ, Song W, Driscoll M, Lasken RS. (2005). Genomic DNA amplification from a single bacterium. *Appl Environ Microbiol* **71**: 3342–3347.

Ragsdale SW, Pierce E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO_2 fixation. *Biochim Biophys Acta* **1784**: 1873–1898.

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

Richter M, Lombardot T, Kostadinov I, Kottmann R, Duhaime MB, Peplies J *et al.* (2008). Jcoast—a biologist-centric software tool for data mining and comparison of prokaryotic (meta)genomes. *BMC Bioinformatics* **9**: 177.

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR, Chisholm SW. (2009). Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS One* **4**: e6864.

Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, Budd A *et al.* (2010). The compartmentalized bacteria of the *Planctomycetes-Verrucomicrobia-Chlamydiae* superphylum have membrane coat-like proteins. *PloS Biol* **8**: e1000281.

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

Siegl A, Hentschel U. (2009). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Environ Microbiol Reports*; e-pub ahead of print 31 July 2009, doi: 10.1111/j.1758-2229.2009.00057.x.

Stepanauskas R, Sieracki ME. (2007). Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *Proc Natl Acad Sci USA* **104**: 9052–9057.

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

Taylor MW, Thacker RT, Hentschel U. (2007b). Genetics. Evolutionary insights from sponges. *Science* **316**: 1854–1855.

Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, Halpern A *et al.* (2010). Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. *ISME J*; e-pub ahead of print 3 June 2010, doi:10.1038/ismej.2010.74.

Wagner M, Horn M. (2006). The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**: 241–249.

Webster NS, Taylor MW, Behnam F, Lücker S, Rattei T, Whalan S *et al.* (2009). Deep sequencing reveals exceptional diversity and modes of transmission for bacterial sponge symbionts. *Environ Microbiol*; e-pub ahead of print 29 September 2009, doi: 10.1111/j.1462-2920.2009.02065.x.

Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V *et al.* (2007). Database resources of the National Center for Biotechnology. *Nucleic Acids Res* **35**: D5–D12.

Woyke T, Tighe D, Mavromatis K, Clum A, Copeland A, Schackwitz W *et al.* (2010). One bacterial cell, one complete genome. *PLoS One* **5**: e10314.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.

Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW *et al.* (2006). Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24**: 680–686.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)

4. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges.

This article has been published in the peer-reviewed journal The ISME Journal and is available electronically ahead of print (doi:10.1038/ismej.2013.111.).

The supplementary information to this article can be accessed at the journal webpage (http://www.nature.com/ismej) and is additionally provided with a CD attached to this PhD thesis.

For documentation of individual contributions to this work and consent of all authors for publication in this thesis please refer to appendix A.

The ISME Journal (2013), 1–14

www.nature.com/ismej

ORIGINAL ARTICLE

Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges

Janine Kamke[1], Alexander Sczyrba[2,3], Natalia Ivanova[2], Patrick Schwientek[2], Christian Rinke[2], Kostas Mavromatis[2], Tanja Woyke[2] and Ute Hentschel[1]

[1]*Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany;* [2]*Department of Energy Joint Genome Institute, Walnut Creek, CA, USA and* [3]*Center for Biotechnology, Bielefeld University, Bielefeld, Germany*

Many marine sponges are hosts to dense and phylogenetically diverse microbial communities that are located in the extracellular matrix of the animal. The candidate phylum *Poribacteria* is a predominant member of the sponge microbiome and its representatives are nearly exclusively found in sponges. Here we used single-cell genomics to obtain comprehensive insights into the metabolic potential of individual poribacterial cells representing three distinct phylogenetic groups within *Poribacteria*. Genome sizes were up to 5.4 Mbp and genome coverage was as high as 98.5%. Common features of the poribacterial genomes indicated that heterotrophy is likely to be of importance for this bacterial candidate phylum. Carbohydrate-active enzyme database screening and further detailed analysis of carbohydrate metabolism suggested the ability to degrade diverse carbohydrate sources likely originating from seawater and from the host itself. The presence of uronic acid degradation pathways as well as several specific sulfatases provides strong support that *Poribacteria* degrade glycosaminoglycan chains of proteoglycans, which are important components of the sponge host matrix. Dominant glycoside hydrolase families further suggest degradation of other glycoproteins in the host matrix. We therefore propose that *Poribacteria* are well adapted to an existence in the sponge extracellular matrix. *Poribacteria* may be viewed as efficient scavengers and recyclers of a particular suite of carbon compounds that are unique to sponges as microbial ecosystems.

The ISME Journal advance online publication, 11 July 2013; doi:10.1038/ismej.2013.111
Subject Category: Microbe-microbe and microbe-host interactions
Keywords: marine sponge; symbiont; carbohydrate degradation; extracellular matrix; single-cell genomics

Introduction

Marine sponges (phylum *Porifera*) are the most ancient extant metazoans with fossil records, indicating their emergence more than 600 million years ago (Love *et al.*, 2009). These animals are sessile filter feeders with an enormous filtering capacity that is known to affect nutrient concentrations in the surrounding environment (Gili and Coma, 1998; Maldonado *et al.*, 2005, 2012). In addition to their evolutionary and ecological significance, sponges have attracted recent scientific attention owing to their specific and unique microbiology (Hentschel *et al.*, 2012). The microbial biomass in sponges is located in the extracellular matrix, the so-called 'mesohyl', and can make up 35% of the sponge body mass (Vacelet, 1975). Collectively, representatives of more than 30 bacterial phyla and both archaeal lineages have so far been found in sponges from various geographic locations (Webster *et al.*, 2010; Schmitt *et al.*, 2012; Simister *et al.*, 2012). The microbial diversity of marine sponges is well investigated (Taylor *et al.*, 2007), and the collective repertoire of 'omics' approaches has been instrumental to shed light on the functional genomic traits of the collective sponge microbiome (Thomas *et al.*, 2010; Fan *et al.*, 2012; Liu *et al.*, 2012; Radax *et al.*, 2012). However, community-wide approaches do not provide sufficient information about functions of specific symbiont clades. Providing a thorough understanding of symbiont function is further complicated by the fact that many sponge symbiont lineages remain uncultivated, such as for the many candidate phyla found in these animals (Schmitt *et al.*, 2012).

One such candidate phylum, termed Candidate phylum *Poribacteria*, was originally discovered and

Correspondence: U Hentschel, Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Julius-von-Sachs-Platz 3, Wuerzburg 97082, Germany.
E-mail: ute.hentschel@uni-wuerzburg.de
Received 26 February 2013; revised 4 June 2013; accepted 4 June 2013

described in marine sponges (Fieseler *et al.*, 2004). *Poribacteria* are widely distributed and highly abundant in sponge species around the world (Fieseler *et al.*, 2004; Lafi *et al.*, 2009; Schmitt *et al.*, 2011, 2012), and also occur freely in seawater, albeit at very low abundances (Pham *et al.*, 2008; Webster *et al.*, 2010; Taylor *et al.*, 2013). They were shown to be affiliated with the *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* (PVC) superphylum (Wagner and Horn, 2006). *Poribacteria* are vertically transmitted via reproductive stages (Schmitt *et al.*, 2008, Webster *et al.*, 2010). Owing to their abundance and diversity within sponges, *Poribacteria* can be regarded as a model sponge symbiont.

Single-cell genomics has become the most useful tool to investigate the genomic repertoire of distinct uncultivated microbial symbionts (Kamke *et al.*, 2012) and other microorganisms (Woyke *et al.*, 2009; Yoon *et al.*, 2011; Stepanauskas, 2012). It also has previously been successfully applied to a poribacterial cell (Siegl *et al.*, 2011). Unlike other 'omics' approaches, this method can connect phylogenetic identity with the functional potential of uncultivated microbial organisms, even from high diversity environments. Here we used single-cell genomics to analyze five poribacterial cells from the Mediterranean sponge *Aplysina aerophoba*, expanding the existing data set from one (Siegl *et al.*, 2011) to a total of six poribacterial single-amplified genomes (SAGs). We provide an in-depth genomic analysis of one of the main symbiont lineages in the complex microbiota of marine sponges. The property of carbohydrate degradation emerged as the most common feature among the analyzed genomes. We therefore focused on carbohydrate degradation potential of *Poribacteria* in this study and discussed the results in context of a nutritional basis of the sponge–microbe symbiosis.

Materials and methods

Sample collection and processing

Samples of the marine sponge *A. aerophoba* were collected in September 2009 by scuba diving to a depth of 5–12 m at the Coast of Rovinj, Croatia (45°08′N, 13°64′E). The animals were transported to the University of Wuerzburg (Wuerzburg, Germany) and kept in seawater aquaria until further processing within 1 week of collection. Fresh sponge samples were used for extraction of sponge-associated prokaryotes using an established protocol of tissue disruption, density centrifugation and filtration by Fieseler *et al.* (2004).

Single-cell sorting, whole genome amplification and PCR screening

Single-cell isolations were conducted with freshly extracted and purified sponge-associated prokaryotes using the fluorescence-activated cell sorting Vantage SE flow cytometer with FACSDiVa option (Becton Dickinson, Heidelberg, Germany) as described previously (Siegl and Hentschel, 2010). For cell lysis and whole genome amplification (WGA) by multiple displacement amplification, we followed the same procedure as Siegl *et al.* (2011). To identify phylogenetically WGA products and check for possible contamination, we screened the WGA products obtained by polymerase chain reaction (PCR) using 16S or 18S rRNA gene primers targeting *Eubacteria*, *Archaea*, *Poribacteria* and *Eukaryotes*, as described previously (Siegl and Hentschel, 2010). *Poribacteria*-positive WGA products were additionally screened with the degenerated PCR primer pair 27f-B and 1492r-B (Cho and Giovannoni, 2004) that covers *Eubacteria* more broadly and enables to obtain longer 16S rRNA gene sequences from *Poribacteria*. Subsequent cloning of PCR products, restriction fragment length polymorphism analysis and Sanger sequencing was conducted to confirm the presence of a single cell using the same procedures as Siegl and Hentschel (2010).

WGA products that originated from single poribacterial cells were then subjected to another round of multiple displacement amplification under the same conditions as stated above. Before genome sequencing, we conducted a S1 nuclease treatment and DNA purification as described by Siegl *et al.* (2011). Five SAGs were selected in the PCR screening process and subjected to whole genome sequencing: *Candidatus* Poribacteria WGA-3G, WGA-4C, WGA-4CII, WGA-4E and WGA-4G (hereafter referred as 3G, 4C, 4CII, 4E and 4G, respectively). These were complemented by one poribacterial SAG sequence from an earlier study by Siegl *et al.* (2011), *Candidatus* Poribacteria WGA-A3 (hereafter referred as 3A). The existing assembly for this SAG was used for annotation and further analyses as described below.

Genome sequencing, assembly and annotation

A detailed description of all steps of genome sequencing, assembly, annotation and quality checks can be found in Supplementary Text S1. Briefly, a combination of Illumina and 454 pyrosequencing was conducted for double displacement amplification products of SAGs 4C, 4E and 4G at LGC Genomics (Berlin, Germany) and the DOE Joint Genome Institute (JGI, Walnut Creek, CA, USA). SAGs 3G and 4CII were also sequenced at JGI using Illumina HiSeq2000 technology only. Illumina sequences were normalized using DUK, a filtering program developed at JGI, and used for assembly including 454 reads (if available). For Illumina/454 hybrid assemblies, a combination of Velvet (Zerbino and Birney, 2008), Allpaths-LG (Zerbino and Birney, 2008) and the 454 Newbler assembler (Roche/454 Life Sciences, Branford, CT, USA) was used. For Illumina assemblies, we used the programs Velvet and Allpaths-LG. All assemblies were submitted to the IMG/ER annotation pipeline (Markowitz *et al.*, 2009) for gene prediction and automatic

functional annotation. All data sets were quality checked and screened for possible contamination (Supplementary Table S1). Manual curation and functional analyses were conducted within the IMG/MER system (Markowitz *et al*., 2012a), unless stated otherwise.

The Whole Genome Shotgun projects were deposited at DDBJ/EMBL/GenBank under the accession nos. ASZN00000000 (3G), APGO00000000 (4C), ASZM 00000000 (4CII), AQTV00000000 (4E) and AQPC00 000000 (4G). The versions described in this paper are versions ASZN01000000 (3G), APGO01000000 (4C), ASZM01000000 (4CII), AQTV01000000 (4E) and AQPC01000000 (4G). Raw data, genome assemblies and annotations can also be accessed under the IMG software system (http://img.jgi.doe.gov) under genome IDs 2 265 129 006–2 265 129 011. Additionally included 16S rRNA gene sequences for phylogenetic analysis from PCR screenings were submitted to GenBank under accession numbers KC713965–KC713966.

ANI and tetranucleotide frequency analysis
Average nucleotide identities based on BLAST (ANIb) and tetranucleotide frequencies were estimated using the JSpecies software (v.1.2.1; http://www.imedea.uib.es/jspecies/about.html) with default parameters (Richter and Rosselló-Móra, 2009).

Genome completeness estimation
Genome size and completeness were estimated using two conserved single copy gene sets that have been determined from all bacterial ($n = 1516$) and all archaeal ($n = 111$) finished genome sequences in the IMG database (Markowitz *et al*., 2012b). The sets consist of 138 bacterial and 162 archaeal conserved single copy genes that occurred only once in at least 90% of all genomes by analysis of an abundance matrix based on hits to the protein family database (Punta *et al*., 2012). HMMs of the identified protein families were used to search both, all SAG assemblies and all combined assemblies by means of the HMMER3 software (Finn *et al*., 2011; http://hmmer.janelia.org/help). Resulting best hits above precalculated cutoffs were counted and the completeness was estimated as the ratio of conserved single-copy gene to total conserved single-copy genes in the set after normalization to 90%. Thereafter, the estimated complete genome size was calculated by division of the estimated genome coverage by the total assembly size.

Phylogenetic analysis
Poribacterial 16S rRNA gene sequences were aligned using the SINA aligner (Pruesse *et al*., 2012) and manually checked in the ARB software package (v.5.3; http://www.arb-home.de/) (Ludwig *et al*., 2004). The SILVA 16S rRNA database version 111 (Pruesse *et al*., 2012) was used for selection of reference sequences plus additional poribacterial 16S rRNA gene sequences in the SILVA database in January 2013. Only sequences $\geqslant 1100$ bp were used to construct a maximum-likelihood bootstrap tree with 1000 resamplings with the RAxML software (v.7.2.8; http://www.exelixis-lab.org/) (Stamatakis, 2006). The resulting tree was reimported into ARB and short poribacterial sequences ($\leqslant 1099$ bp) were added without changing tree topology using the parsimony interactive tool in ARB.

Screening for carbohydrate-active enzymes
Protein sequences of poribacterial genomes were screened against the HMM profile-based database of carbohydrate-active enzymes obtained from dbCAN (Yin *et al*., 2012) in December 2012 using hmmsearch in the HMMER software package (v.3.0; http://hmmer.janelia.org/help) (Finn *et al*., 2011). Results were filtered using an *e*-value cutoff $< 10^{-5}$. In addition, all returned hits were manually evaluated based on their functional annotation in IMG/MER and excluded in case of conflicting results. Comparison between glycoside hydrolase (GH)-encoding genes (E.C.: 3.2.1.x) between *Poribacteria* and all free-living planktonic organisms available in the IMG software system in May 2013 was conducted additionally using functional annotation tools in IMG.

Results and discussion

General genomic features

SAG sequencing. Final genome assembly sizes for the poribacterial cells ranged from 0.19 to 5.44 Mbp (Table 1). For genomes 3G, 4C and 4E, genome recovery was large enough to estimate genome coverage of 98.54%, 38.36% and 58.20%, respectively, whereas the largely fragmented assemblies of 3A, 4CII and 4G did not permit for genome size estimation. The estimated poribacterial genome sizes ranged from 4.25 to 6.27 Mb (Table 1) and do not suggest genome size reduction. The guanine–cytosine content ranged from 47% to 50%, with the exception of genome 4C (41%) (Table 1). Protein coding genes accounted for 95.5–99.6% of the retrieved genomes (Table 1). Approximately 30% of these could not be functionally assigned (50% for genome 3A). The investigated genomes encoded for only one copy of the 16S rRNA gene, which is consistent with previous reports (Fieseler *et al*., 2006; Siegl *et al*., 2011).

Definition of phylotypes. Phylogenetic analysis of nearly full-length 16S rRNA gene sequences of *Poribacteria* showed that three of the six analyzed SAGs (3A, 3G and 4CII) clustered closely together, with genome 4G also in close proximity (>97%

Table 1 Summary of assembly and genome statistics of poribacterial genomes

	Group I				*4C*	*4E*
	3A	*3G*	*4CII*	*4G*		
Assembly size (bp)	414 219	5 441 554	543 453	189 191	1 629 923	3 647 669
Estimated genome size (bp)	N/A	5 521 899	N/A	N/A	4 249 040	6 267 358
Estimated genome recovery (%)	N/A	98.54	N/A	N/A	38.36	58.20
Number of contigs	157	286	44	15	276	521
Largest contig size (bp)	13 447	227 865	92 961	33 129	76 460	69 798
Sequencing effort (Mbp)						
454 FLX	105	N/A	N/A	N/A	N/A	N/A
454 FLX Titanium	N/A	N/A	N/A	97.1	153.6	74.4
Illumina GA IIx	N/A	N/A	N/A	5800	N/A	6800
Illumina HI Seq 2000	N/A	1410	780	N/A	2300	N/A
GC content (%)	49	48	48	47	41	50
Protein CDs						
No	503	4772	473	170	1618	3281
%	99.60	99.00	98.95	95.51	99.02	98.86
Protein coding genes with function prediction						
No	256	3228	305	112	950	2376
%	50.89	67.64	64.48	65.88	58.71	72.42
Protein coding genes without function prediction						
No	247	1544	168	58	668	905
%	49.11	32.36	35.52	34.12	41.29	27.58
rRNAs						
No	1	2	3	1	3	4
%	0.20	0.04	0.63	0.56	0.18	0.12
tRNAs						
No	1	43	2	7	13	31
%	0.20	0.89	0.42	3.93	0.80	0.93

Abbreviations: CD, coding genes; GC, guanine–cytosine content; NA, not applicable; rRNA, ribosomal RNA; tRNA, transfer RNA.

sequence similarity), whereas the other two genomes (4C and 4E) each fell separate from this group (Figure 1). Average nucleotide identity and tetranucleotide frequency analysis confirmed a closer relationship between SAGs 3A, 3G, 4CII and 4G than to the other two SAGs (Supplementary Table S2 and Supplementary Figure S1), and they were therefore defined as one phylotype named group I (Figure1). The other two SAGs each represent a separate phylotype. Group I represents, however, a 'composite phylotype' as the values for tetranucleotide frequency and average nucleotide identity analysis are under the defined thresholds of 0.99 for tetranucleotide frequency and 95–96% ID for average nucleotide identity (Richter and Rosselló-Móra, 2009).

GHs and other CAZymes

The ability of *Poribacteria* to degrade and transform complex carbohydrates was assessed by screening genome data against the dbCAN (Yin *et al*., 2012) and classified according to the carbohydrate-active enzymes (CAZy) database (Cantarel *et al*., 2009). Most poribacterial hits matched GHs and glycosyl transferase, whereas carbohydrate binding modules, carbohydrate esterases and polysaccharide lyases made up for a smaller proportion (Supplementary Table S3, Supplementary Figure S2). GH and glycosyl transferase also showed the highest diversity with up to 18 different GHs and 18 glycosyl transferases as detected in SAG 3G. We detected 17, 76, 21, 14 and 63 hits to GH families in SAGs 3A, 3G, 4C, 4CII and 4E, respectively (fasta file available in Supplementary Information). No GHs were found in 4G, which is however largely incomplete. GH frequencies of 1.6%, 1.3% and 1.9% genes/genome were estimated for 3G, 4C, and 4E, respectively. In a recent study by Martinez-Garcia *et al*. (2012), marine *Verrucomicrobia* genomes contained approximately 0.9–1.2% genes encoding for GHs as compared with the average of 0.2% in other bacteria. Considering slight differences in screening methods between our study and that by Martinez-Garcia *et al*. (2012), the poribacterial GH frequency was found to be similar to that of *Verrucomicrobia*. Poribacterial genomes also show an almost linear correlation between the number of GHs and genome coverage with $R^2 = 0.9525$ (Supplementary Figure S3). Similar to *Verrucomicrobia*, these findings could indicate a specialization of *Poribacteria* towards carbohydrate degradation.

We additionally compared the poribacterial GH frequencies with those of all available finished genomes of marine planktonic bacteria in IMG ($n = 102$, May 2013) and also the nearly closed *Verrucomicrobia* single-cell genome AAA168-F10 by Martinez-Garcia *et al*. (2012) based on annotation

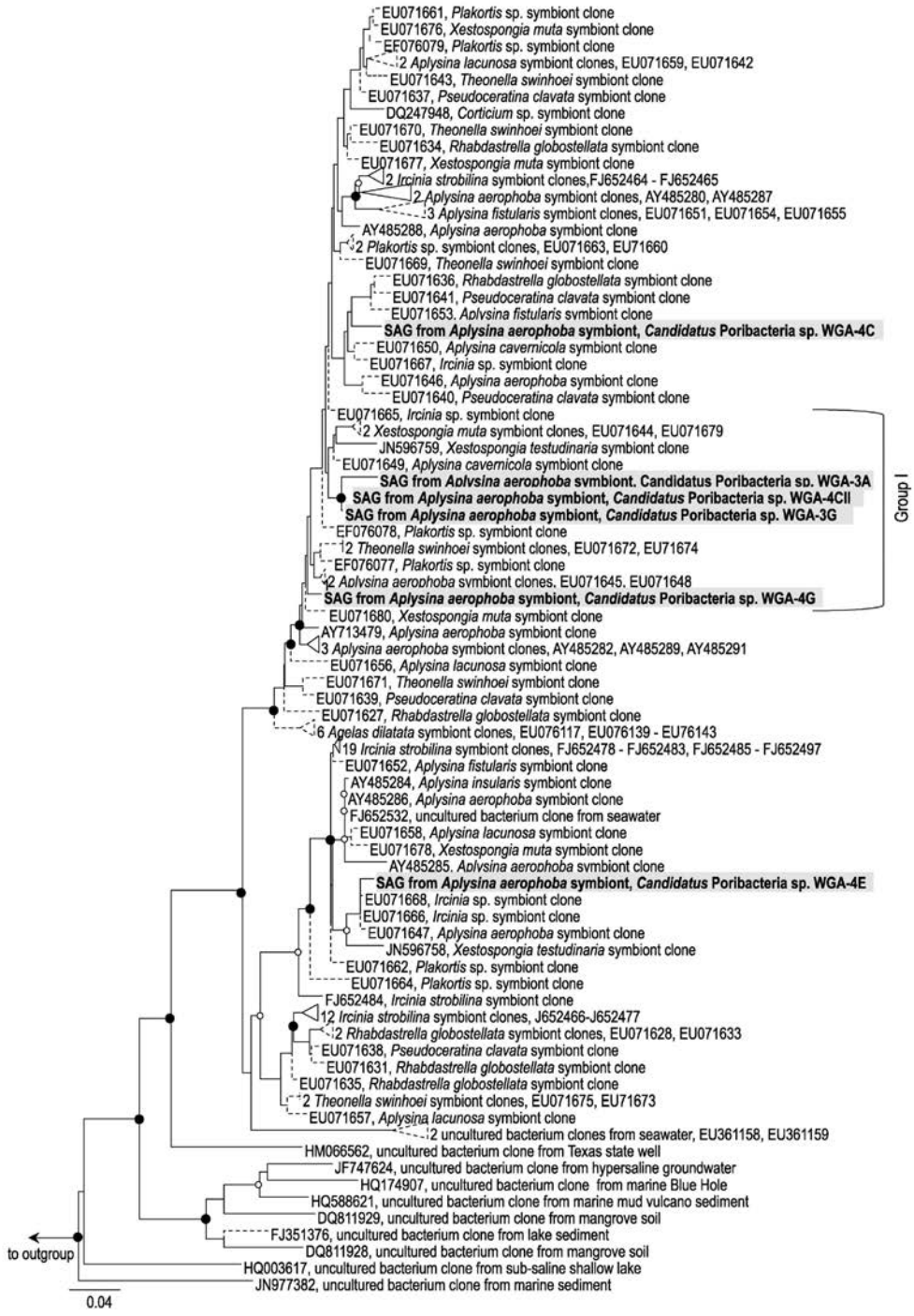


Figure 1 Phylogenetic maximum-likelihood tree based on 16S rRNA genes of the candidate phylum *Poribacteria*. Sequences from poribacterial SAGs are shown in bold and gray shading. The tree was constructed based on long sequences (⩾1100 nucleotides), shorter sequences were added without changing tree topology and are indicated by dashed lines. Bootstrap support (1000 resamplings) of ⩾90% is shown by filled, and ⩾75% by open circles. The outgroup consisted of several *Spirochaetes* sequences. Scale bar represents 4% sequence divergence.

of EC numbers 3.2.1.x (see CAZY database). GH frequencies based on this analysis were on average 0.32% for planktonic bacteria, 0.24% for *Verrucomicrobia* AAA168-F10 and 0.08–0.42% for poribacterial genomes. A *Poribacteria*-specific set of enzymes was however not identified, which may be because of the fact that the most dominant GH families in *Poribacteria* are not accessible in the IMG system because of the lack of EC number annotations available for these families.

Overall, 22 different GH families were identified in all poribacterial genomes (Table 2 and Supplementary Figure S4). Glycoside hydrolase family 109 (GH109) was the most abundant family of GHs. Most of the poribacterial GH109 proteins are annotated as predicted dehydrogenases and related proteins. An α-*N*-acetylgalactosaminidase activity is described for GH109, although more functions may be assigned to this family (Henrissat 'GH109' in CAZypedia; URL: http://www.cazypedia.org, accessed

Table 2 Overview of glycoside hydrolase families in *Poribacteria*

GH family	*Known activities*	*Potential activities in Poribacteria*	*Putative substrates*	*Group I*			*4C*	*4E*
				3A	*3G*	*4CII*		
GH2	β-Galactosidase, β-glucuronidase, β-mannosidase, others	β-Galactosidase/ β-glucuronidase	β-D-galactoside β-D-glucuronic acid (glycosaminoglycans/ mucopolysaccharides)			1		
GH4	α-Glucosidase, α-galactosidase, α-glucuronidase, other	α-Galactosidase (EC: 3.2.1.22)	Melibiose					1
GH5	Cellulase, β-1,3-glucosidase, many others (55 subfamilies, 20 experimentally defined enzyme functions)	—	β-Linked oligo- and polysaccharides and glycoconjugates					1
GH13	α-Amylase and related enzymes	1,4-α-Glucan branching enzyme (EC: 2.4.1.18), α-amylase	α-Glycoside linkages, (1–4)-α-D-glucosidic linkages in polysaccharides 1,4-α-D-glucan	1	2			
GH23	G-type lysozyme, peptidoglycan lyase, peptidoglycan-lytic transglycosylase	Soluble lytic murein transglycosylase and related proteins (EC: 3.2.1.–)	Peptidoglycan, β-1,4-linkage between *N*-acetylmuramyl and *N*-acetylglucosaminyl residues in peptidoglycan		3	2	1	1
GH32	Levanase, invertase, others	β-Fructofuronidase (EC: 3.2.1.26)	Sucrose		3			6
GH33	Sialidase or neuraminidase (EC: 3.2.1.18); *trans*-sialidase (EC: 2.4.1.–); 2-keto-3-deoxynononic acid sialidase (EC: 3.2.1.–)	Exo-α-sialidase, neuraminidase (EC: 3.2.1.18)	Neuraminic acids, α-(2→3)-, α-(2→6)-, α-(2→8)-glycosidic linkages of terminal sialic acid residues in oligosaccharides, glycoproteins, glycolipids, colominic acid	1	2		1	14
GH36	α-Galactosidase, α-*N*-acetylgalactosaminidase	α-Galactosidase	Terminal α-galactosyl moieties from glycolipids and glycoproteins		1		1	
GH50	β-Agarases	—	Cleave β-1,4 glycosidic bonds of agarose, releasing neoagaro-biose, -tetraose and -hexaose		1			
GH51	Endoglucanase, α-L-arabinofuranosidase	α-L-arabinofuranosidase	α-L-arabinofuranosides, α-L-arabinans containing (1,3)- and/or (1,5)-linkages, arabinoxylans and arabinogalactans		1		1	
GH67	α-Glucuronidase, xylan α-1,2-glucuronidase	α-Glucuronidase (EC: 3.2.1.139)	Glucuronic acid appended to the C_2-OH of xylose at the non-reducing end of xylooligosaccharides		1			
GH74	Endoglucanase, oligoxyloglucan reducing end-specific cellobiohydrolase, xyloglucanase	Some with BNR (bacterial nuraminidase/ASP-box repeat)	β-1,4-Linkages of various glucans	5	25	7	6	4
GH76	1,6-α-Mannosidase	—	(1→6)-α-D-mannosidic linkages in unbranched (1→6)-mannans		1		1	1
GH88	D-4,5 unsaturated β-glucuronyl hydrolase	—	Release of 4-deoxy-4(5)-unsaturated D-glucuronic acid from oligosaccharides produced by polysaccharide lyases		1		1	1
GH93	Exo-α-L-1,5-arabinase	—	Release of arabinobiose from the non-reducing end of α-1,5-L-arabinan		1			
GH95	α-L-fucosidase, α-1,2-L-fucosidase	α-L-fucosidase (EC:3.2.1.51)]	Fuco-α1-2-Gal linkages attached at the non-reducing ends of oligosaccharides		1		1	
GH105	Unsaturated rhamnogalacturonyl hydrolase	—	Ramnosylgalacturan degradation					1
GH106	α-L-rhamnosidase	—	Hydrolysis of terminal non-reducing α-L-rhamnose residues in α-L-rhamnosides		1			
GH109	α-*N*-acetylgalactosaminidase	—	*N*-acetylgalactosamine linkage in glycoproteins	2	29	4	8	33
GH127	β-L-arabinofuranosidase	—	Release of L-arabinose from specific disaccharides and glycoconjugates	1	1			
GH129	α-*N*-acetylgalactosaminidase	—	Mucin-type glyoproteins		1			
GH130	1-β-D-mannopyranosyl-4-D-glucopyranose: phosphate α-D-mannosyltransferase	—	Mannan catabolism		1			

Abbreviation: GH, glycoside hydrolase.

December 2012). α-*N*-acetylgalactosaminidases cleave *N*-acetylgalactosamine residues from glycoproteins and glycolipids. The two most apparent sources are the sponge glycoconjugates or glycoproteins of cell walls, and organic matter (both dissolved and particulate) from seawater. The dominance of GH109 is in agreement with the poribacterial ability to use uronic acids as a carbon source (see below), as both point towards degradation of glycoproteins in sponge matrix components.

The family with the second largest amount of poribacterial hits, GH74, contains many xyloglucan-hydrolyzing enzymes with diverse functions (Yaoi and Ishida 'GH74' in CAZypedia; URL: http://www.cazypedia.org, accessed December 2012). Generally, these enzymes act on the β-1,4-linkages of glucans and might degrade several different substrates including poly- and oligosaccharides of all kinds of organisms. Some GH74 hits from *Poribacteria* genomes encode for proteins with the potential for chitin deacetylation. Chitin is a polysaccharide of *N*-acetyl-D-glucosamine entities and represents a major component of the skeleton of *A. aerophoba* (Ehrlich *et al.*, 2010). Glucans also occur in other glycoconjugates, some of which are responsible for self–non-self recognition in sponges (reviewed by Fernandez-Busquets and Burger, 2003). Xyloglucan-degrading activity may also be inferred from the poribacterial GH74 proteins. Xyloglucans are a major component of plant cell walls and also occur in some green algae (Del Bem and Vincentz, 2010).

GH family GH23 also appears to be widely present in *Poribacteria*. Proteins of bacterial origin in this family are peptidoglycan lyases that cleave the β-1,4-linkage between *N*-acetylmuramyl and *N*-acetyl-glucosaminyl residues in peptidoglycan (Clarke 'GH23' in CAZypedia, URL: http://www.cazypedia.org, accessed December 2012). Poribacterial proteins with hits to this family showed similarity to lytic transglycosylases of other bacteria and contained protein family domain Pfam01464-transglycosylase SLT domain, making it likely that these proteins indeed act as peptidoglycan-lytic transglycosylases.

Of further interest are families GH32 and GH33, which are represented by several hits in genome 4E. Most hits to GH32 were annotated as β-fructofuranosidase (EC: 3.2.1.26), a central enzyme for sucrose degradation. Most hits on poribacterial genomes to GH family GH33 were annotated as neuraminidase/sialidase-like enzyme, exo-α-sialidase (EC: 3.2.1.18), or have high similarities to genes encoding for proteins with this function. This enzyme cleaves terminal sialic acid residues from oligosaccharides, glycoproteins and glycolipids. Sialic acids are *N*-or *O*-substituted derivates of neuraminic acid, most often *N*-acetylneuraminic acid.

Carbon degradation pathways

Single-cell genomic analyses identified *Poribacteria* as aerobic, heterotrophic organisms. Genes for the major central pathways, such as glycolysis, pentose phosphate pathway, tricarboxylic acid cycle, the non-phosphorylative or semiphosphorylative Entner–Doudoroff pathway and oxidative phosphorylation were largely present (Supplementary Table S4 and Figure 4). Group I might utilize glucose completely over the ED pathway, as none of the genomes belonging to this phylotype encoded for phospho-fructokinase (EC: 2.7.1.11) or fructose bisphophate aldolase (EC: 4.1.2.13). This suggests that glycolysis is missing in this phylotype. Instead, group I genome 3G encoded for enzymes enabling transfer of glucose into glucuronate by glucose 1-dehydrogenase (EC: 1.1.1.47) and gluconolactonase (EC: 3.1.1.17) to then be degraded by the ED pathway. The phylogenetically more distant genome 4E encoded for both complete glycolysis and ED pathway. Apart from glucose, *Poribacteria* seem to be able to use a variety of additional carbon sources (Supplementary Table S5).

Galactoside, fructoside, xyloside and rhamnoside degradation. Poribacterial genomes showed the potential for degradation of galactoside polymers such as melibiose and lactose. Oxidative degradation of galactosides was supported as well as parts of the Lelior pathway (Supplementary Figure S5 and Supplementary Table S5). The potential for degradation of the fructoside polymers levan and sucrose was also encoded in poribacterial genomes (Supplementary Figure S6) and additionally genes coding for enzymes involved in degradation of D-xylose over the xylose isomerase pathway (Supplementary Figure S6). The genomic potential for rhamnoside degradation was shown by genes relevant for the L-rhamnose isomerase pathway and oxidative L-rhamnose degradation, which was strongly supported by multiple gene copies encoding for enzymes involved in this pathway (Supplementary Figure S7 and Supplementary Table S5). For a detailed description of these pathways in *Poribacteria*, please refer to Supplementary Text S2.

The substrates of these degradation pathways can generally be found in oligo- and polysaccharides of glycoconjugates or biopolymers of various organisms, and especially in the cell walls of many plants and bacteria (Sutherland, 1985; Rehm, 2010; Ray *et al.*, 2011; Visnapuu *et al.*, 2011; Singh *et al.*, 2012). Therefore, they might be freely available in the sponge mesohyl as a result of sponge feeding on bacteria and microalgae.

Inositol degradation. A nearly complete inositol dehydrogenase pathway was present in group I genomes, as well as in genome 4E, and partially in 4C. This pathway degrades myoinositol to glyceraldehyde-3-phosphate, which is further used in the central metabolism (Figure 2). Inositol phosphates are cell wall compounds in all eukaryotes and archaea, and are rarely found in bacteria (Michell, 2011). In addition, phosphorylated inositol is a

precursor for several important lipid molecules including sphingolipids, ceramides and glycosylphosphatidylinositol anchors (Reynolds, 2009), as well as many stress-protective solutes of eukaryotes (Michell, 2011). Inositol phosphate is part of the signal transduction in sponges, as shown in *Geodia cydonium*, where production of inositol triphosphate increases after sponge cell aggregation (Müller *et al*., 1987). It seems therefore likely that either the sponge itself or eukaryotic microorganisms could be a source of inositol for *Poribacteria*. Inositol degradation has been reported from a variety of other bacteria (Fry *et al*., 2001; Yoshida *et al*., 2008, 2012; Kohler *et al*., 2010) including *Sinorhizobium* symbionts of soybean where this pathway was shown to provide a competitive advantage in the plant rhizosphere (Galbraith *et al*., 1998). We hypothesize that *Poribacteria* not only use myoinositol as a carbon source but also as an agent for regulating metabolic functions involved in sponge–microbe symbiosis.

Uronic acid degradation. Analysis of the poribacterial genomes further revealed the presence of several genes connected to uronic acid degradation (Figure 3 and Supplementary Table S5). The degradation of polymers such as pectin or pectin-like glycoconjugates likely occurs in all investigated phylotypes as indicated by the presence of genes encoding for polygalacturonase (EC: 3.2.1.15) in genome 4E and pectate lyase (EC: 4.2.2.2) in group

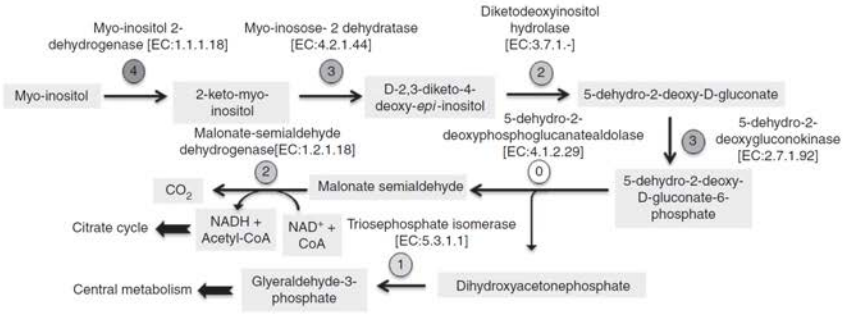


Figure 2 Schematic reconstruction of inositol degradation as encoded on poribacterial SAGs. Numbers within circles represent the number of genomes encoding for the corresponding enzyme.

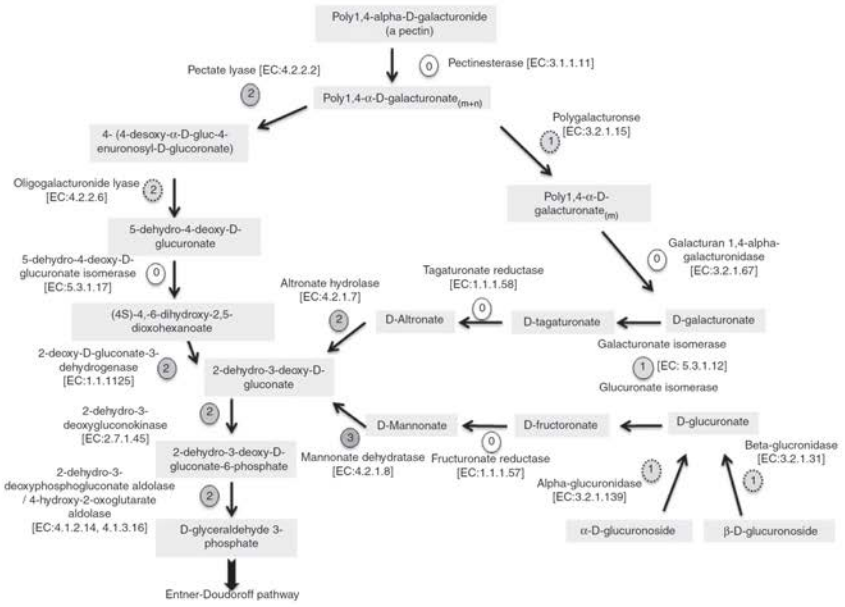


Figure 3 Schematic reconstruction of uronic acid degradation as encoded on poribacterial SAGs. Numbers within circles represent the genomes encoding for the corresponding enzyme. Dashed circles represent manually annotated genes.

I genome 3G and genome 4C. This was further supported by the presence of genes encoding for enzymes participating in 5-dehydro-4-deoxy-D-glucuronate degradation such as oligogalacturonide lyase (EC: 4.2.2.6), 2-deoxy-D-glucconate 3-dehydrogense (EC: 1.1.1.125) and 2-dehydro-3-desoxy-D-glucokinase (EC: 2.7.1.45) in SAG group I genomes and partially in genome 4C. Furthermore, glucuronoside polymers appear to be degradable by *Poribacteria* as indicated by the presence of α- and β-D-glucuronosidases (EC: 3.2.1.139; EC: 3.2.1.31) in group I genomes 3G and 4CII, respectively. The occurrence of genes encoding for galacturonate isomerase (EC: 5.3.1.12), alteronate hydrolase (EC: 4.2.1.7) and mannonate dehydratase (EC: 4.2.1.8) in high copy number on several poribacterial genomes of the phylotypes represented by group I and 4E points towards the possibility of galacturonate and glucuronate catabolism. The products of these degradation steps could then enter the ED pathway via 2-dehydro-3-desoxyphophogluconate aldolase (EC: 4.1.2.14; 4.1.3.16), which was found in group I genomes and in genome 4E.

Uronic acids are sugar acids that can be found in various biopolymers of plant, animal or bacterial origin (Sutherland, 1985; Rehm, 2010). The occurrence of uronic acids in glycosaminoglycans (GAGs) is especially worth noting, because the extracellular matrix of sponges is largely constructed by these polymers (Fernandez-Busquets and Burger, 2003). Therefore, *Poribacteria* should literally be submerged in uronic acid-containing substances. It is known that GAGs of sponges are different from those of higher animals (Misevic and Burger, 1993), with the main components being fucose, glucuronic acid, mannose, galactose, *N*-acetylglucosamine and sulfate in sponge GAGs (Misevic and Burger, 1986, 1993; Misevic *et al*., 1987).

Transporters

The poribacterial genomes code for a range of transporters representing different families (Table 3 and Figure 4), and here we concentrate on those involved in carbohydrate metabolism. Typical sugar transport systems/phosphotransferase systems were missing from all poribacterial genomes. Genes coding for proteins of the tripartite ATP-independent periplasmatic transporter family were found on all three phylotypes, which are often involved in organic acid transport such as C4-dicarboxylates, keto-acids and sugar acids (*N*-acetyl neuraminic acids, sialic acid). The most dominant transporter family in all poribacterial genomes was the ATP-binding cassette superfamily. ATP-binding cassette transporters were found for a variety of broader substrate categories such as amino acids, di- and oligopeptides, carbohydrates, lipoproteins, metal ions and systems involved in cell protection and competition with other organisms (Supplementary Table S6). Most of the detected carbohydrate ATP-binding cassette transporters were simple or multiple sugar transporters and did not show any further specification. However, we detected a system of D-xylose transport on group I genome 3G and transporters for maltose, malto-oligosaccharides, arabinose and lactose on several other genomes besides 3G.

Sulfatases

Genome analysis of the six poribacterial genomes revealed a total of 103 genes coding for sulfatases (Table 4). Individually, group I genomes coded for 57, genome 4C for 25 and genome 4E for 21 sulfatase genes. Genes coding for choline sulfatase (betC; EC: 3.1.6.6) were identified in group I genomes, and 4E. This enzyme transforms choline sulfate to choline

Table 3 Transporters classes on poribacterial genomes

Function ID	*Transporter classification*	*Group I*				*4C*	*4E*
		3A	*3G*	*4CII*	*4G*		
TC: 1.A.30	H^+- or Na^+-translocating bacterial flagellar Mot/Exb superfamily	0	5	3	0	1	4
TC: 2.A.55	Nramp family	0	2	0	1	0	1
TC: 2.A.56	TRAP-T family	0	7	0	0	3	2
TC: 2.A.64	Tat family	0	3	2	0	0	2
TC: 2.A.66	MOP flippase superfamily	0	6	0	1	2	4
TC: 2.A.76	RhtB family	0	0	0	0	0	1
TC: 3.A.1	ABC superfamily	12	174	19	5	44	95
TC: 3.A.15	Outer membrane protein secreting MTB family	0	6	0	0	0	2
TC: 3.A.5	Sec family	0	7	1	1	2	5
TC: 3.A.7	Type IV (conjugal DNA–protein transfer or VirB) secretory pathway (IVSP) family	0	1	0	0	0	1
TC: 4.C.1	Proposed FAT family	1	2	0	0	0	1
TC: 5.A.4	Prokaryotic SDH family	0	2	0	0	0	3
TC: 9.B.22	Leukotoxin secretion MorC family	1	0	0	0	0	2

Abbreviations: ABC, ATP-binding cassette; FAT, fatty acid transporter; MOP, m/oligosaccharidyl-lipid/polysaccharide; MorC, morphogenesis protein C; Mot/Exb, Motor/ExbBD outer membrane transport energizer; MTB, main terminal branch; Nramp, metal ion (Mn^{2+}-iron) transporter; RhtB, resistance to homoserine/threonine; SDH, succinate dehydrogenase; Sec, general secretory pathway; Tat, twin arginine targeting; TRAP-T, tripartite ATP-independent periplasmic transporter.

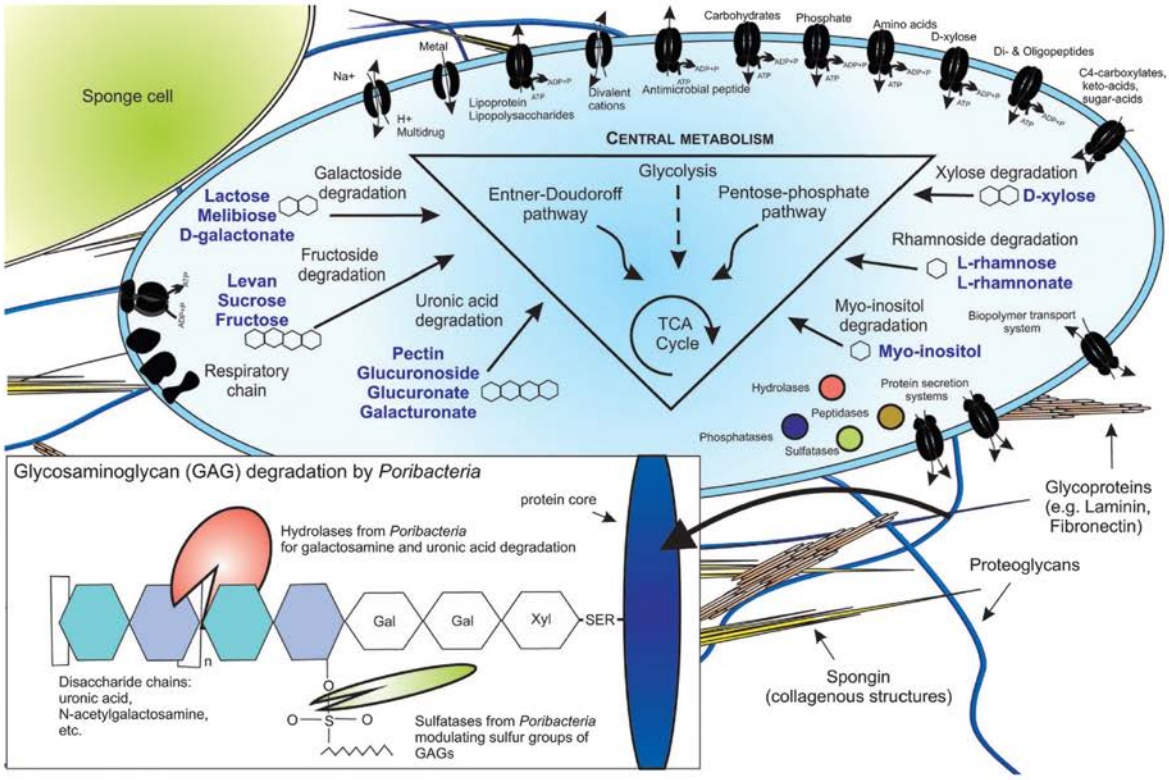


Figure 4 Schematic overview of a poribacterial cell in the sponge extracellular matrix illustrating pathways of carbohydrate metabolism and glycosaminoglycan degradation by poribacterial enzymes. The dashed arrow represents glycolysis that is not supported by the dominant poribacterial phylotype group I.

Table 4 Sulfatase genes on poribacterial genomes

Sulfatase type	*KO term*	*KO name*	*Function*	*Number of genes on genome*					
				Group I				*4C*	*4E*
				3A	*3G*	*4CII*	*4G*		
Total		N/A		8	45	3	1	25	21
Arylsulfatase	K01130	aslA	EC: 3.1.6.1	—	1	—	—	—	1
Cerebroside-sulfatase	K01134	ARSA	EC: 3.1.6.8	—	1	—	—	—	—
Choline-sulfatase	K01133	E3.1.6.6, betC	EC: 3.1.6.6	—	4	—	1	3	3
Iduronate-2-sulfatase	K01136	IDS	EC: 3.1.6.13	—	—	—	—	1	—
N-acetyl-galactosamine-4-sulfatase	K01135	ARSB	EC: 3.1.6.12	—	1	—	—	1	—
N-sulfoglucosamine-sulfohydrolase	K01565	SGSH	EC: 3.10.1.1	—	1	—	—	—	—
Uncharacterized	K01138	—	EC: 3.1.6.-	1	4	2	—	1	1
Not determined		N/A		7	33	1		19	16

Abbreviations: KO, KEGG orthology; N/A, not applicable.

and then to glycine betaine, which is often used as an osmoprotectant (Le Rudulier *et al.*, 1984) or can be degraded via glycine and serine to pyruvate. This degradative pathway is largely present on genome 4E, whereas glycine degradation was encoded in all three phylotypes. As choline-*O*-sulfate is synthesized by different microorganisms (Spencer and Harada, 1960; Fitzgerald and Luschinski, 1977; Rivoal and Hanson, 1994), it is likely available as carbon and sulfur substrate for *Poribacteria*.

Two genes coding for arylsulfatase (EC: 3.1.6.1) were detected in group I genomes 3G and 4E. General substrates for this type of enzyme are phenol sulfates, but the exact kind of phenol sulfate is difficult to determine because of high similarities between different types of substrates. Genome 3G

codes also for a cerebroside-sulfatase (EC: 3.1.6.8), which also hydrolyzes phenol sulfates, ascorbate 2-sulfate and galactose-3-sulfate residues in lipids. These enzymes might provide *Poribacteria* with an organic sulfur source from hydrolyzed sulfate esters in the absence of inorganic sulfate. *Poribacteria* have the genomic potential for assimilatory sulfate reduction over adenosine-5′-phosphosulfate and 3′-phosphoadenosin-5′-phosphosulfate and subsequent cysteine synthesis (data not shown). The use of organic sulfur compounds under sulfur-limiting conditions has been shown in various bacteria (Kertesz, 2000). The presence of sulfated lipids and polysaccharides in the sponge extracellular matrix is well documented (Zierer and Mourao, 2000; Vilanova *et al.*, 2009), and it may thus serve as a possible carbon and sulfur source.

Group I and genome 4C both encode for *N*-acetylgalactosamine-4-sulfatase (EC: 3.1.6.12), which lyses the sulfate groups of *N*-acetylgalactosamine-4-sulfate from chondroitin and dermatan sulfate. Chondroitin is known to be part of sponge GAG chains, and *N*-acetylgalactosamine was found in GAG chains and glycoproteins of sponges (Fernandez-Busquets and Burger, 2003). In addition, genome 4C codes for iduronate-2-sulfatase (EC: 3.1.6.13) and SAG group I genome 3G for *N*-sulfoglucosamine sulfohydrolase (EC: 3.10.1.1); however, their substrates are yet to be identified in sponges.

Symbiotic heterotrophy in sponges

Symbioses frequently have a nutritional basis, such as nitrogen fixation in *Rhizobium*–legume symbioses (Lodwig *et al.*, 2003), supplementation of amino acids in *Buchnera* symbionts of aphids (Ramsey *et al.*, 2010; Hansen and Moran, 2011), chemoautotrophy in marine mussels and worms (Woyke *et al.*, 2006; Petersen *et al.*, 2011) or photosynthesis in symbionts of corals or ascidians (Weis and Allemand, 2009; Schnitzler and Weis, 2010; Donia *et al.*, 2011). Here we present bacterial heterotrophy, that is, the ability to utilize diverse carbon sources, as a potential functional basis for the interaction of microbial symbionts with sponges. The role of heterotrophy in symbiosis in general has so far been underestimated. Only recently, studies of gut microbiomes of human, ruminants or termites, and other terrestrial hosts (Warnecke *et al.*, 2007; Hess *et al.*, 2011; Zhu *et al.*, 2011; Schloissnig *et al.*, 2013) have received major attention. Here we show that symbiotic heterotrophy could also be of relevance in a marine host, using poribacterial symbionts of marine sponges as an example. Whether the host gains benefit from the heterotrophic metabolism of its symbionts is an interesting question for future investigations.

The carbohydrate degradation potential of *Poribacteria* opens up various nutritional sources that are taken up by the host's extensive filtration activities and transported into the sponge interior. Sponges feed on dissolved and particulate organic matter (Yahel *et al.*, 2003; De Goeij *et al.*, 2008) as well as heterotrophic bacteria and eukaryotes (Maldonado *et al.*, 2012; Perea-Blázquez *et al.*, 2012). It has further been shown that sponge feeding on dissolved and particulate organic matter from bacterial and algal sources can be mediated by bacterial symbionts (De Goeij *et al.*, 2008). Here we provide the genomic background and detailed carbon degradation pathways behind this ecological observation. We could show that *Poribacteria* have the potential to catabolize various substrates that can be found in, for example, cell wall components, glycoproteins and polysaccharides from marine algae and prokaryotes (Rehm, 2010; Jiao *et al.*, 2011).

The carbon degradation repertoire of *Poribacteria* would further be consistent with degradation of compounds from the extracellular matrix of the sponge host (Figure 4). Our hypothesis is supported by (i) the high abundance of GH families with acetylgalactosaminidase or sialidase activities (and several GHs consistent with this hypothesis), (ii) the ability to degrade uronic acids, frequent components of glycoconjugates and especially GAGs, as well as (iii) the presence of sulfatases with GAGs as the specific substrate. Remarkably, *Poribacteria* are significantly enriched in high microbial abundance over low microbial abundance of sponges or seawater (Schmitt *et al.*, 2012; Taylor *et al.*, 2013) One major difference between high microbial abundance and low microbial abundance sponges (apart from the amount and diversity of microorganisms itself) is that high microbial abundance sponges contain a notably expanded mesohyl matrix as compared with their low microbial abundance counterparts (Hentschel *et al.*, 2003; Weisz *et al.*, 2008). The extracellular matrix may thus provide both habitat and nutrient source to *Poribacteria.* As sponges are known to continuously remodel matrix components (Bond, 1992), it is likely that polymers that become available in this process, serve as nutrient substrates for *Poribacteria.* This process is unlikely harmful to the host as no signs of tissue destruction in healthy *A. aerophoba* sponges can be observed neither in the natural environment nor by transmission electron microscopy of the mesohyl matrix (Vacelet, 1975; Friedrich *et al.*, 2001)

The bacteria–sponge interaction may go beyond nutrition and may have also a mechanistic basis. In this context, the so-called 'sponge aggregation factor' is of relevance, which shows a proteoglycan-like structure and has a major role in cell-specific aggregation cell–matrix connections and adhesion processes (Müller and Zahn, 1973; Misevic and Burger, 1993; Fernandez-Busquets and Burger, 2003). It might be conceivable that *Poribacteria* influence the function of the sponge aggregation factor through glycosylase, glyocotransferase activities, or by modification of sulfate groups on the

mucopolysaccharide chains. The poribacterial potential to influence sialylation by GH family 33 enzymes might also affect this process. It has been shown that sialic acid residues are an important component of glycoproteins and mucopolysaccharides on sponge cell surfaces (Garrone *et al.*, 1971) and that these also have a major role in sponge cell aggregation (Müller *et al.*, 1977). It might thus be possible that poribacterial enzymes interfere with mechanistic processes related to adhesion, aggregation and self–non-self recognition.

Conflict of Interest

The authors declare no conflict of interest.

Acknowledgements

We gratefully acknowledge the marine operations personnel at the Ruder Boskovic Institute (Rovinj/Croatia) for the help during sponge collection, Kristina Bayer for logistical support, C Linden (University of Würzburg) for FACS analysis of sponge symbiont cells and Michael Richter (Ribocon GmbH, Bremen) for useful bioinformatic advice. LGC Genomics (Berlin) is acknowledged for excellent customer services. Financial support to UH was provided by the SFB630-grant TPA5, the SFB567-grant TPC3 and by the Bavaria California Technology Center (BaCaTeC). TW, CR, PS, NI and KM were funded by the US Department of Energy Joint Genome Institute, Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

Bond C. (1992). Continuous cell movements rearrange anatomical structures in intact sponges. *J Exp Zool* **263**: 284–302.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233–D238.

Cho JC, Giovannoni SJ. (2004). Cultivation and growth characteristics of a diverse group of oligotrophic marine *Gammaproteobacteria*. *Appl Environ Microbiol* **70**: 432–440.

De Goeij JM, Moodley L, Houtekamer M, Carballeira NM, Van Duyl FC. (2008). Tracing ^{13}C-enriched dissolved and particulate organic carbon in the bacteria-containing coral reef sponge *Halisarca caerulea*: evidence for DOM feeding. *Limnol Oceanogr* **53**: 1376–1386.

Del Bem LEV, Vincentz MG. (2010). Evolution of xyloglucan-related genes in green plants. *BMC Evol Biol* **10**: 341.

Donia MS, Fricke WF, Partensky F, Cox J, Elshahawi SI, White JR *et al.* (2011). Complex microbiome underlying secondary and primary metabolism in the tunicate–*Prochloron* symbiosis. *Proc Natl Acad Sci USA* **108**: E1423–E1432.

Ehrlich H, Ilan M, Maldonado M, Muricy G, Bavestrello G, Kljajic Z *et al.* (2010). Three-dimensional chitin-based scaffolds from *Verongida* sponges (*Demospongiae*: *Porifera*). Part I. Isolation and identification of chitin. *Int J Biol Macromol* **47**: 132–140.

Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, Webster NS *et al.* (2012). Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. *Proc Natl Acad Sci USA* **109**: E1878–E1887.

Fernandez-Busquets X, Burger MM. (2003). Circular proteoglycans from sponges: first members of the spongican family. *Cell Mol Life Sci* **60**: 88–112.

Fieseler L, Horn M, Wagner M, Hentschel U. (2004). Discovery of the novel candidate phylum '*Poribacteria*' in marine sponges. *Appl Environ Microbiol* **70**: 3724–3732.

Fieseler L, Quaiser A, Schleper C, Hentschel U. (2006). Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum *Poribacteria* by environmental genomics. *Environ Microbiol* **8**: 612–624.

Finn RD, Clements J, Eddy SR. (2011). HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res* **39**: W29–W37.

Fitzgerald JW, Luschinski PC. (1977). Further studies on the formation of choline sulfate by bacteria. *Can J Microbiol* **23**: 483–490.

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

Fry J, Wood M, Poole PS. (2001). Investigation of myo-inositol catabolism in *Rhizobium leguminosarum* bv. *viciae* and its effect on nodulation competitiveness. *Mol Plant Microbe Interact* **14**: 1016–1025.

Galbraith MP, Feng SF, Borneman J, Triplett EW, de Bruijn FJ, Rossbach S. (1998). A functional myo-inositol catabolism pathway is essential for rhizopine utilization by *Sinorhizobium meliloti*. *Microbiol-Sgm* **144**: 2915–2924.

Garrone R, Thiney Y, Pavans de Ceccatty M. (1971). Electron microscopy of a mucopolysaccharide cell coat in sponges. *Experientia* **27**: 1324–1326.

Gili JM, Coma R. (1998). Benthic suspension feeders: their paramount role in littoral marine food webs. *Trends Ecol Evol* **13**: 316–321.

Hansen AK, Moran NA. (2011). Aphid genome expression reveals host–symbiont cooperation in the production of amino acids. *Proc Natl Acad Sci USA* **108**: 2849–2854.

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

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

Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G *et al.* (2011). Metagenomic discovery of biomass-degrading genes and genomes from cow rumen. *Science* **331**: 463–467.

Jiao G, Yu G, Zhang J, Ewart HS. (2011). Chemical structures and bioactivities of sulfated polysaccharides from marine algae. *Mar Drugs* **9**: 196–223.

Kamke J, Bayer K, Woyke T, Hentschel U. (2012). Exploring symbioses by single-cell genomics. *Biol Bull* **223**: 30–43.

Kertesz MA. (2000). Riding the sulfur cycle-metabolism of sulfonates and sulfate esters in Gram-negative bacteria. *FEMS Microbiol Rev* **24**: 135–175.

Kohler PRA, Zheng JY, Schoffers E, Rossbach S. (2010). Inositol catabolism, a key pathway in *Sinorhizobium meliloti* for competitive host nodulation. *Appl Environ Microbiol* **76**: 7972–7980.

Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, Hentschel U. (2009). Widespread distribution of *Poribacteria* in *Demospongiae*. *Appl Environ Microbiol* **75**: 5695–5699.

Le Rudulier D, Strom AR, Dandekar AM, Smith LT, Valentine RC. (1984). Molecular biology of osmoregulation. *Science* **224**: 1064–1068.

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

Lodwig EM, Hosie AH, Bourdes A, Findlay K, Allaway D, Karunakaran R *et al.* (2003). Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature* **422**: 722–726.

Love GD, Grosjean E, Stalvies C, Fike DA, Grotzinger JP, Bradley AS *et al.* (2009). Fossil steroids record the appearance of *Demospongiae* during the Cryogenian period. *Nature* **457**: 718–721.

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.

Maldonado M, Carmona MC, Velasquez Z, Puig A, Cruzado A, López A *et al.* (2005). Siliceous sponges as a silicon sink: an overlooked aspect of benthopelagic coupling in the marine silicon cycle. *Limnol Oceanogr* **50**: 799–809.

Maldonado M, Ribes M, van Duyl FC. (2012). Nutrient fluxes through sponges: biology, budgets, and ecological implications. *Adv Mar Biol* **62**: 113–182.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Grechkin Y *et al.* (2012a). IMG/M: the integrated metagenome data management and comparative analysis system. *Nucleic Acids Res* **40**: D123–D129.

Markowitz VM, Chen IM, Palaniappan K, Chu K, Szeto E, Grechkin Y *et al.* (2012b). IMG: the Integrated Microbial Genomes database and comparative analysis system. *Nucleic Acids Res* **40**: D115–D122.

Markowitz VM, Mavromatis K, Ivannova NN, Chen IM, Chu K *et al.* (2009). IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* **25**: 2271–2278.

Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PS, Reitenga KG *et al.* (2012). Capturing single cell genomes of active polysaccharide degraders: an unexpected contribution of Verrucomicrobia. *PLoS One* **7**: e35314.

Michell RH. (2011). Inositol and its derivatives: their evolution and functions. *Adv Enzyme Regul* **51**: 84–90.

Misevic GN, Burger MM. (1986). Reconstitution of high cell binding affinity of a marine sponge aggregation factor by cross-linking of small low affinity fragments into a large polyvalent polymer. *J Biol Chem* **261**: 2853–2859.

Misevic GN, Finne J, Burger MM. (1987). Involvement of carbohydrates as multiple low affinity interaction sites in the self-association of the aggregation factor from the marine sponge Microciona prolifera. *J Biol Chem* **262**: 5870–5877.

Misevic GN, Burger MM. (1993). Carbohydrate–carbohydrate interactions of a novel acidic glycan can mediate sponge cell adhesion. *J Biol Chem* **268**: 4922–4929.

Müller WE, Arendes J, Kurelec B, Zahn RK, Müller I. (1977). Species-specific aggregation factor in sponges. Sialyltransferase associated with aggregation factor. *J Biol Chem* **252**: 3836–3842.

Müller WE, Rottmann M, Diehl-Seifert B, Kurelec B, Uhlenbruck G, Schroder HC. (1987). Role of the aggregation factor in the regulation of phosphoinositide metabolism in sponges. Possible consequences on calcium efflux and on mitogenesis. *J Biol Chem* **262**: 9850–9858.

Müller WEG, Zahn RK. (1973). Purification and characterization of a species-specific aggregation factor in sponges. *Exp Cell Res* **80**: 95–104.

Perea-Blázquez A, Davy SK, Bell JJ. (2012). Estimates of particulate organic carbon flowing from the pelagic environment to the benthos through sponge assemblages. *PLoS One* **7**: e29569.

Petersen JM, Zielinski FU, Pape T, Seifert R, Moraru C, Amann R *et al.* (2011). Hydrogen is an energy source for hydrothermal vent symbioses. *Nature* **476**: 176–180.

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

Pruesse E, Peplies J, Glockner FO. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* **28**: 1823–1829.

Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C *et al.* (2012). The Pfam protein families database. *Nucleic Acids Res* **40**: D290–D301.

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

Ramsey JS, MacDonald SJ, Jander G, Nakabachi A, Thomas GH, Douglas AE. (2010). Genomic evidence for complementary purine metabolism in the pea aphid, *Acyrthosiphon pisum*, and its symbiotic bacterium *Buchnera aphidicola*. *Insect Mol Biol* **19**: 241–248.

Ray B, Bandyopadhyay SS, Capek P, Kopecký J, Hindák F, Lukavský J. (2011). Extracellular glycoconjugates produced by cyanobacterium *Wollea saccata*. *Int J Biol Macromol* **48**: 553–557.

Rehm BHA. (2010). Bacterial polymers: biosynthesis, modifications and applications. *Nat Rev Microbiol* **8**: 578–592.

Reynolds TB. (2009). Strategies for acquiring the phospholipid metabolite inositol in pathogenic bacteria, fungi and protozoa: making it and taking it. *Microbiology* **155**: 1386–1396.

Richter M, Rosselló-Móra R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci USA* **106**: 19126–19131.

Rivoal J, Hanson AD. (1994). Choline-*O*-sulfate biosynthesis in plants (identification and partial characterization of a salinity-inducible choline sulfotransferase from species of *Limonium* (*Plumbaginaceae*). *Plant Physiol* **106**: 1187–1193.

Schloissnig S, Arumugam M, Sunagawa S, Mitreva M, Tap J, Zhu A *et al.* (2013). Genomic variation

landscape of the human gut microbiome. *Nature* **493**: 45–50.

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

Schmitt S, Hentschel U, Taylor MW. (2011). Deep sequencing reveals diversity and community structure of complex microbiota in five Mediterranean sponges. *Hydrobiologia* **687**: 341–351.

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.

Schnitzler CE, Weis VM. (2010). Coral larvae exhibit few measurable transcriptional changes during the onset of coral–dinoflagellate endosymbiosis. *Mar Genom* **3**: 107–116.

Siegl A, Hentschel U. (2010). PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. *Env Microbiol Rep* **2**: 507–513.

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

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

Singh RP, Shukla MK, Mishra A, Reddy CR, Jha B. (2012). Bacterial extracellular polymeric substances and their effect on settlement of zoospore of *Ulva fasciata*. *Colloids Surf B* **103**: 223–230.

Spencer B, Harada T. (1960). The role of choline sulphate in the sulphur metabolism of fungi. *Biochem J* **77**: 305–315.

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

Stepanauskas R. (2012). Single cell genomics: an individual look at microbes. *Curr Opin Microbiol* **15**: 613–620.

Sutherland IW. (1985). Biosynthesis and composition of Gram-negative bacterial extracellular and wall polysaccharides. *Annu Rev Microbiol* **39**: 243–270.

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

Taylor MW, Tsai P, Simister RL, 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.

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

Vacelet J. (1975). Etude en microscopie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (*Dictyoceratida*). *J Microsc Biol Cell* **23**: 271–288.

Vilanova E, Coutinho CC, Mourao PA. (2009). Sulfated polysaccharides from marine sponges (*Porifera*): an ancestor cell–cell adhesion event based on the carbohydrate–carbohydrate interaction. *Glycobiol* **19**: 860–867.

Visnapuu T, Mardo K, Mosoarca C, Zamfir AD, Vigants A, Alamäe T. (2011). Levansucrases from *Pseudomonas syringae* pv. tomato and *P. chlororaphis* subsp. *aurantiaca*: substrate specificity, polymerizing properties and usage of different acceptors for fructosylation. *J Biotechnol* **155**: 338–349.

Wagner M, Horn M. (2006). The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. *Curr Opin Biotechnol* **17**: 241–249.

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT *et al.* (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.

Webster NS, Taylor MW, 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* **8**: 2070–2082.

Weis VM, Allemand D. (2009). Physiology. What determines coral health? *Science* **324**: 1153–1155.

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

Woyke T, Teeling H, Ivanova NN, Huntemann M, Richter M, Gloeckner FO *et al.* (2006). Symbiosis insights through metagenomic analysis of a microbial consortium. *Nature* **443**: 950–955.

Woyke T, Xie G, Copeland A, González JM, Han C, Kiss H *et al.* (2009). Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.

Yahel G, Sharp JH, Marie D, Hase C, Genin ARN. (2003). *In situ* feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: bulk DOC is the major source for carbon. *Limnol Oceanogr* **48**: 141–149.

Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**: W445–W451.

Yoon HS, Price DC, Stepanauskas R, Rajah VD, Sieracki ME, Wilson WH *et al.* (2011). Single-cell genomics reveals organismal interactions in uncultivated marine protists. *Science* **332**: 714–717.

Yoshida K, Sanbongi A, Murakami A, Suzuki H, Takenaka S, Takami H. (2012). Three inositol dehydrogenases involved in utilization and interconversion of inositol stereoisomers in a thermophile, *Geobacillus kaustophilus* HTA426. *Microbiology* **158**: 1942–1952.

Yoshida K, Yamaguchi M, Morinaga T, Kinehara M, Ikeuchi M, Ashida H *et al.* (2008). Myo-inositol catabolism in *Bacillus subtilis*. *J Biol Chem* **283**: 10415–10424.

Zerbino DR, Birney E. (2008). Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* **18**: 821–829.

Zhu L, Wu Q, Dai J, Zhang S, Wei F. (2011). Evidence of cellulose metabolism by the giant panda gut microbiome. *Proc Natl Acad Sci USA* **108**: 17714–17719.

Zierer MS, Mourao PA. (2000). A wide diversity of sulfated polysaccharides are synthesized by different species of marine sponges. *Carbohydr Res* **328**: 209–216.

Supplementary Information accompanies this paper on The ISME Journal website (http://www.nature.com/ismej)

5. Novel features of the candidate phylum *Poribacteria* revealed by single-cell genomic analysis

---

Authors

Janine Kamke[1], Christian Rinke[2], Patrick Schwientek[2], Kostas Mavromatis[2], Natalia Ivanova[2], Alexander Sczyrba[2, 3], Tanja Woyke[2], and Ute Hentschel[1]*

Authors affiliations:

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

[2]Department of Energy Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, United States of America

[3]Center for Biotechnology, Bielefeld University, Universitätsstraße 27, D-33615 Bielefeld, Germany

This article was submitted to the peer-reviewed journal Plos One and is currently under review.

The supplementary information to this article is provided with a CD attached to this PhD thesis.

For documentation of individual contributions to this work and consent of all authors for publication in this thesis please refer to appendix A.

Abstract

The candidate phylum *Poribacteria* is one of the most dominant and widespread members of the microbial communities residing within marine sponges. Cell compartmentalization had been postulated along with their discovery about a decade ago and their phylogenetic association to the *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* superphylum was proposed soon thereafter. In the present study we revised these features based on genomic data obtained from six poribacterial single cells. We propose that *Poribacteria* form a distinct monophyletic phylum contiguous to the PVC superphylum together with other candidate phyla. Our genomic analyses supported the possibility of cell compartmentalization in form of bacterial microcompartments. Further analyses of eukaryote-like protein domains stressed the importance of such proteins with features including tetratricopeptide repeats, leucin rich repeats as well as low density lipoproteins receptor repeats, the latter of which are reported here for the first time from a sponge symbiont. Finally, examining the most abundant protein domain family on poribacterial genomes revealed diverse phyH family proteins, some of which may be related to dissolved organic posphorus uptake.

Introduction

Single-cell genomics is a powerful tool to describe genomes of as yet uncultivated organisms from diverse environments [1,2]. Recently it allowed a first glimpse into the vast functional diversity represented by genomes of previously largely uncharacterized candidate phyla [3]. In this study we investigated six single amplified genomes (SAGs) of a symbiont of marine sponges, the candidate phylum *Poribacteria*. *Poribacteria* were first discovered as highly abundant symbionts of marine sponges [4] and as of now lack any cultivated representatives. With the advent of next generation sequencing technologies they were also detected in seawater albeit in low abundances [5-7]. *Poribacteria* are one of the most predominant taxa inhabiting the extracellular matrix (mesohyl) of sponge species around the world [8-10]. These symbionts are vertically transmitted over larval stages from the adult sponge to the next generation [6,11]. Initially, the candidate phylum *Poribacteria* showed a moderate phylogenetic relationship to *Planctomycetes*, *Verrucomicrobia*, and *Chlamydiae* (PVC superphylum) based on monophyletic clustering in 16S rRNA gene analysis [4]. Later, *Poribacteria* were classified as members of the PVC superphylum although the exact position within the superphylum could not be completely resolved [12]. Similar to some members of the PVC superphylum *Poribacteria* were also suspected to have a compartmentalized cell plan [4]. In this study we revisited the features of phylogeny and cell compartmentalization based on the sequence data of six single-cell derived genomes from the candidate phylum *Poribacteria*. We further reveal a large abundance and diversity of eukaryote-like domain containing proteins as well as phyH-like proteins in *Poribacteria*.

Materials and Methods

Genome annotation and analysis

Six poribacterial single-cell genome sequences were included in this study, these being *Candidatus Poribacteria* WGA 3A, 3G, 4C, 4CII, 4E and 4G with Genbank accession numbers ADFK02000000, ASZN01000000, APGO01000000, ASZM01000000, AQTV01000000, AQPC01000000, respectively. These genomes were previously obtained by our group from uncultivated bacteria inhabiting the marine sponge *Aplysina aerophoba* by fluorescence activated cell sorting (FACS), multiple

displacement amplification (MDA), and next generation sequencing [13,14]. For a detailed description of all steps from sample collection to genome assembly and annotation please refer to Kamke et al. [14]. Genome sequences were automatically annotated via the IMG pipeline [15] and manually curated in IMG/MER. All analyses were conducted using the tools in IMG/MER unless further specified.

Clustering analysis of PhyH family genes

For clustering of pfam 05721-PhyH family proteins we used the fastclust algorithm in usearch [16] with an identity cutoff of 60% amino acid id.

Phylogenetic 16S rRNA gene analysis

Sequences for 16S rRNA gene based phylogenetic analysis were selected from the SILVA 16S rRNA database version 108 [17] in the ARB software package (V5.3) [18]. All poribacterial 16S rRNA sequences (≥1100 bp) available in GenBank by June 2013 and the 16S rRNA sequences of poribacterial single-cell genomes were included. Additional sequences for the candidate phyla *Aerophobetes* (CD12) and *Hydrogenedentes* (NKB19) were obtained by blast searches [19] of reference sequences (accession number JN675971 for CD12 and CR933119 for NKB19) against Genbank nr/nt database in June 2013 and selecting the 100 best hits with >75% sequence ID and sequence length ≥1100bp. All sequence added to the original database were aligned using the SINA aligner [20] and included into the ARB database for further manual refinement. Alignments were exported from ARB for phylogenetic tree construction using RAxML (v7.3.2) [21]. Maximum likelihood trees were constructed using sequences ≥1100bp only and 50% conservation filters. Bootstrap analysis was carried out with 500 resamplings. Trees were reimported into ARB and sequences < 1100bp were added to the tree using the parsimony interactive tool in ARB without changing tree topology.

Phylogenetic analysis of 83 bacterial marker protein sequences

For the calculation of the bacterial phylogenetic tree we followed the procedure described by Rinke et al. [3] based on a custom marker set of 83 bacteria specific markers (Suppl. Table S1) described in the study. Briefly, single-cell genome assemblies of *Poribacteria* were translated into all six reading frames and marker genes were detected and aligned with hmmsearch and hmmalign included in the HMMER3 package [22] using HMM profiles obtained from phylosift

(http://phylosift.wordpress.com/). Extracted marker protein sequences were used to build concatenated alignments of up 83 markers per genome. Alignments were included into the database constructed by Rinke and coworkers [3] and reference sequences were selected for phylogenetic tree construction. Phylogenetic inference methods used were the maximum likelihood based FastTree2 [23] and a custom RAxML bootstrap script originally provided by Christian Goll and Alexandros Stamatakis (Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Germany) and modified by Douglas Jacobsen (Bioinformatics Computing Consultant, LBNL, Berkeley, USA). The script requires two input files, the alignment file as PHYLIP format and a starting tree calculated by RAxML-Light [24]. The script workflow is briefly summarized as follows: First RAxML version 7.3.5 [21] creates bootstrap replicates of the multiple sequence alignments and stepwise addition order parsimony trees as starting points for the maximum likelihood search, based on user defined rate heterogeneity and substitution models. Next RAxML-Light [24] is run on every bootstrap replicate. After all RAxML-Light runs are finished the resulting replicate trees are fed into RAxML to calculate the bootstrap support values which are drawn upon the starting tree. The rate heterogeneity and amino acid evolution models used were GAMMA and LG for the custom RAxML bootstrap script, and CAT approximation with 20 rate categories and Jones-Taylor-Thorton (JJT) for FastTree2. To evaluate the robustness of the protein trees we used seven different out-group taxon configurations (Table 1).

Results and Discussion

Phylogenetic revision of Poribacteria

Analysis of phylogenetic interferences of up to 83 marker genes (hereafter termed phylogenomic analyses) showed that all poribacterial SAGs clustered, with 100% bootstrap support in all our tree calculations, in a monophyletic group distinct to the PVC superphylum (Table 1, Fig. 1). *Poribacteria* SAGs clustered with the recently proposed phyla *Aerophobetes* (CD12) and/ or *Hydrogenedentes* (NKB19) [3] in most of our phylogenomic calculations (Table 1). This loosely affiliated clade, including other phyla such as *Elusimicrobia*, formed in some tree calculations a sister clade to the PVC superphylum (Fig. 1). Phylogenetic analysis of the 16S rRNA gene supported

monophyletic clustering of *Poribacteria* with strong bootstrap support (Fig. 2). However, phylogenetic placement based on the 16S rRNA gene did not show the direct grouping

Table 1. Summary of phylogenetic inference results from all phylogenomic tree calculations.

Inference [1]	Species [2]	Por BS[3]	Sistergroup[4]	Clade members (BS)[5]	Outgroup[6]
Fasttree, CAT, JTT	2311	100%	Hydrogenedentes (NKB19)	Poribacteria, Hydrogenedentes (NKB19), Aerophobetes (CD12) (100%)	all bacteria
Fasttree, CAT, JTT	316	100%	Aerophobetes (CD12) + Hydrogenedentes (NKB19)	Poribacteria, Hydrogenedentes (NKB19), Aerophobetes (CD12) (100%)	Spirochaetes, Alpha- & Betaproteobacteria, Firmicutes, Cyanobacteria, Elusimicrobia
Fasttree, CAT, JTT	310 (noS)	100%	Chloroflexi	Poribacteria, Chloroflexi, Hydrogenedentes (NKB19), Aerophobetes (CD12) (87%)	Spirochaetes
Fasttree, CAT, JTT	312 (noS)	100%	Hydrogenedentes (NKB19)	Poribacteria, Hydrogenedentes (NKB19), Aerophobetes (CD12), Elusimicrobia (71%)	Spirochaetes, Alpha-, Beta-, & Gammaproteobacteria
Fasttree, CAT, JTT	306 (noS)	100%	Aerophobetes (CD12)	Poribacteria, Aerophobetes (CD12) (100%)	Spirochaetes
RAxML, GAMMA, LG	312 (noS)	100%	Hydrogenedentes (NKB19)	Poribacteria, Hydrogenedentes (NKB19), Aerophobetes (CD12), Elusimicrobia (45%)	Spirochaetes, Alpha-, Beta-, & Gammaproteobacteria

[1] Inference method, rate categories, and substitution model
[2] number of species in tree. Single sequences which did not belong to any main clades were removed before tree calculations where indicated (noS = no Singletons).
[3] Bootstrap support for the phylum Poribacteria
[4] sistergroup to the phylum Poribacteria
[5] sistergroup to the phylum Poribacteria
[6] phyla added as outgroups for tree calculation

with *Aerophobetes* (CD12) and/ or *Hydrogenedentes* (NKB19) (Fig. 2). Instead *Poribacteria* were placed (bootstrap support 91%) separately within a larger cluster of other phyla including the PVC superphylum as well as the candidate phylum WS3, recently renamed as *Latescibacteria* [3] and a monophyletic lineage previously described as “sponge associated unclassified lineage” (SAUL) [9].

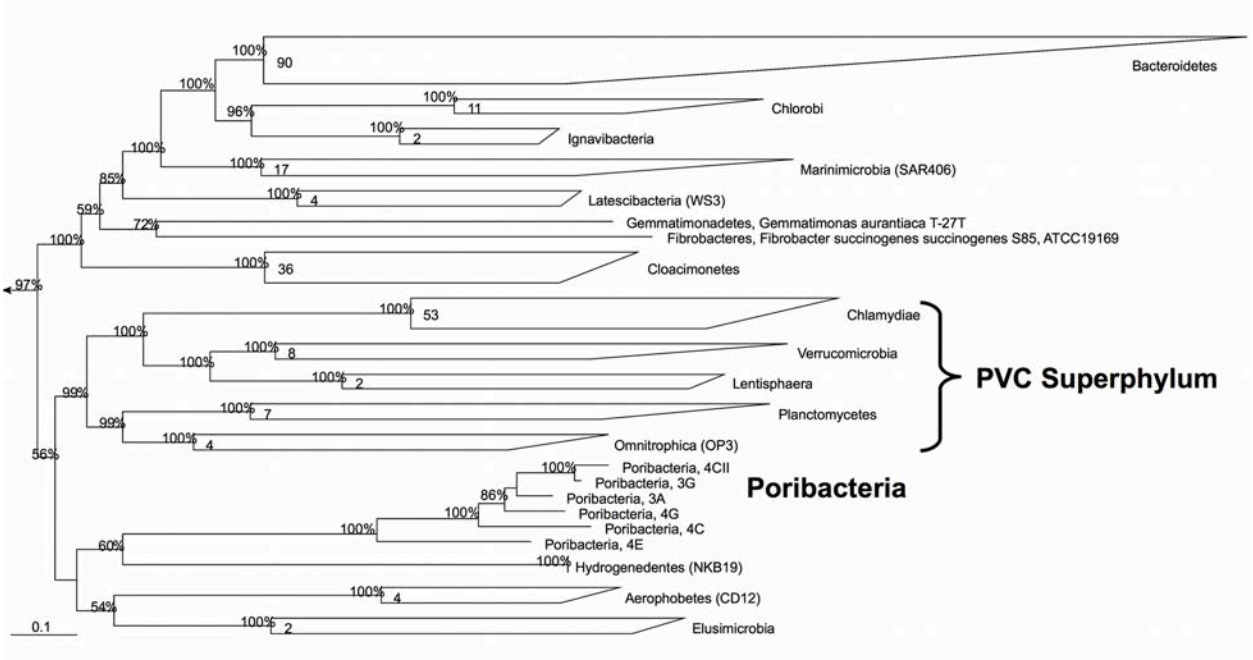


Figure 5: Phylogenomic tree based on a concatenated alignment of up to 83 genes illustrating the phylogenetic position of the candidate phylum Poribacteria. Bootstrap value (100 resamplings) are shown on tree nodes where support ≥ 50%. Number of genomes per group is displayed in group boxes. Outgroup consists of several species of Spirochaetes and Gammaproteobacteria. The scale bar represents 10% sequence divergence.

The inconsistency between phylogenomic and the 16S rRNA gene-based phylogeny might be due to the relatively low resolution provided by the single marker gene (16S rRNA) analysis compared to multiple genes analysis as has been suggested previously [3,25]. On the other hand the phylogenomic analysis, limited to the relatively small amount of draft reference sequences available at the time of analysis, might not be able to properly resolve the general placement of the phylum. We expect that the position of the *Poribacteria* in the tree of life will be further refined as more genome sequences of *Poribacteria* and of other candidate phyla become available. Importantly, the phylogenetic analyses performed in this study (whether 16S rRNA gene or marker genes based) did not support a clustering of *Poribacteria* with the PVC superphylum, which is in contrast to what was suggested earlier [12].

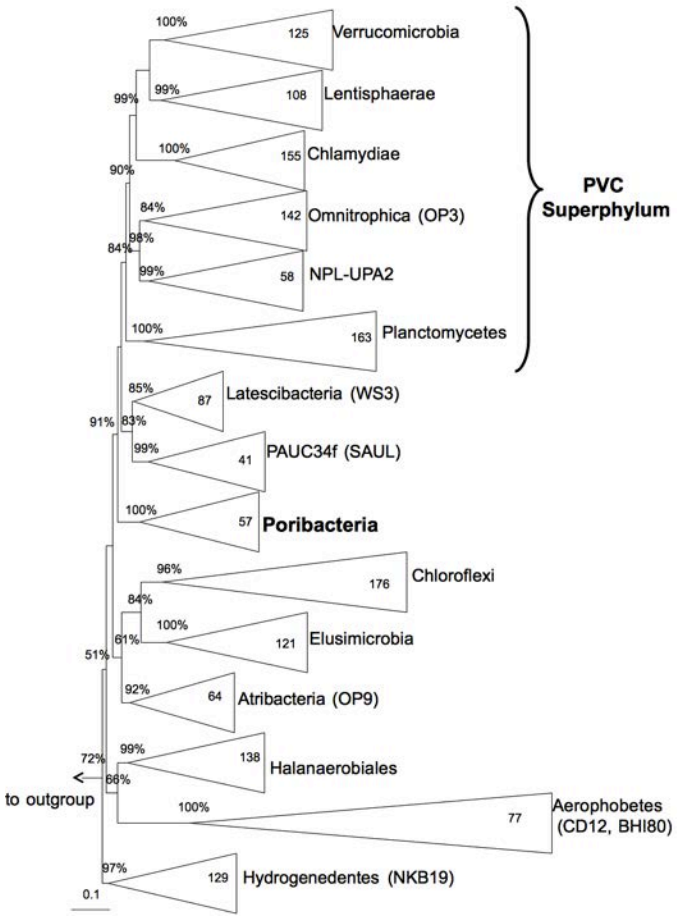


Figure 6: 16S rRNA gene based maximum likelihood tree illustrating the phylogenetic position of the candidate phylum *Poribacteria*. Bootstrap values (500 resamplings) ≥ 50% are shown on tree nodes. Numbers of sequences included per group is shown in group boxes. Outgroup consists of 80 sequences belonging to the *Bacteroidetes*. Scale bar represents 10% sequences divergence.

Previous studies based on concatenated alignments of protein data [26,27], also showed the phylogenetic position of *Poribacteria* outside the PVC superphylum. However, these studies included only one poribacterial genome sequence available at that time, *Candidatus* Poribacteria sp. WGA A3, which was later shown to be flawed by contaminating DNA [14]. Besides phylogenetic analysis, Gupta and coworkers [26] described a conserved hypothetical protein as a potential marker for the PVC superphylum, which was however missing from all poribacterial SAGs analysed here. This lack of a PVC marker protein provides further support for the independent phylogenetic position of *Poribacteria*.

Genomic evidence for microcompartments

Cell compartmentalization is one characteristic that has been proposed for *Poribacteria* based on ring shaped fluorescence in situ hybridization (FISH) signals and the electron

microscopic observations of compartmentalized prokaryotic cells in the mesohyl of the sponge *Aplysina aerophoba* [4]. The observed structures appeared similar to those described for many members of the *Planctomycetes* [29] and most specifically for *Gemmata obscuriglobus* where the compartment was proposed to be a nucleus-like structure [30]. The existence of cell compartmentalization in members of the PVC superphylum was later connected to the occurrence of membrane coat like proteins encoded on the genome [31]. To further investigate the possibility of cell compartmentalization in *Poribacteria*, we searched poribacterial SAGs for possible genomic evidence of such features. We were not able to find membrane coat like proteins or any genomic indication of large cell compartments. This is in accordance with a recent study which challenged the concept of the existence of these compartments even in other bacteria and confutes the existence of a nucleus-like structure in *G. obscuriglobus* [32].

Our analysis did reveal evidence for a possible occurrence of bacterial microcompartments (BMCs) in *Poribacteria*. Four of six poribacterial SAGs encoded for genes with hits to either one of two pfam domains namely, pfam00936 BMC or pfam03319 EutN CcmL (Table 2).

Table 2: BMC shell protein markers on poribacterial SAGs.

Function ID	Name	3G	4CII	4E	3A
pfam00936	BMC	1	1	2	0
pfam03319	EutN CcmL	3	0	3	1

These domains are considered markers for BMC shell proteins. Specifically, we identified three regions with conserved genomic structure between different poribacterial SAGs (groups A-C) (Fig 3) that encoded for genes with these domains. A fourth region (group D) was identified on SAG 4E with two BMC shell proteins enclosing a set of 21 genes (Fig. 3).

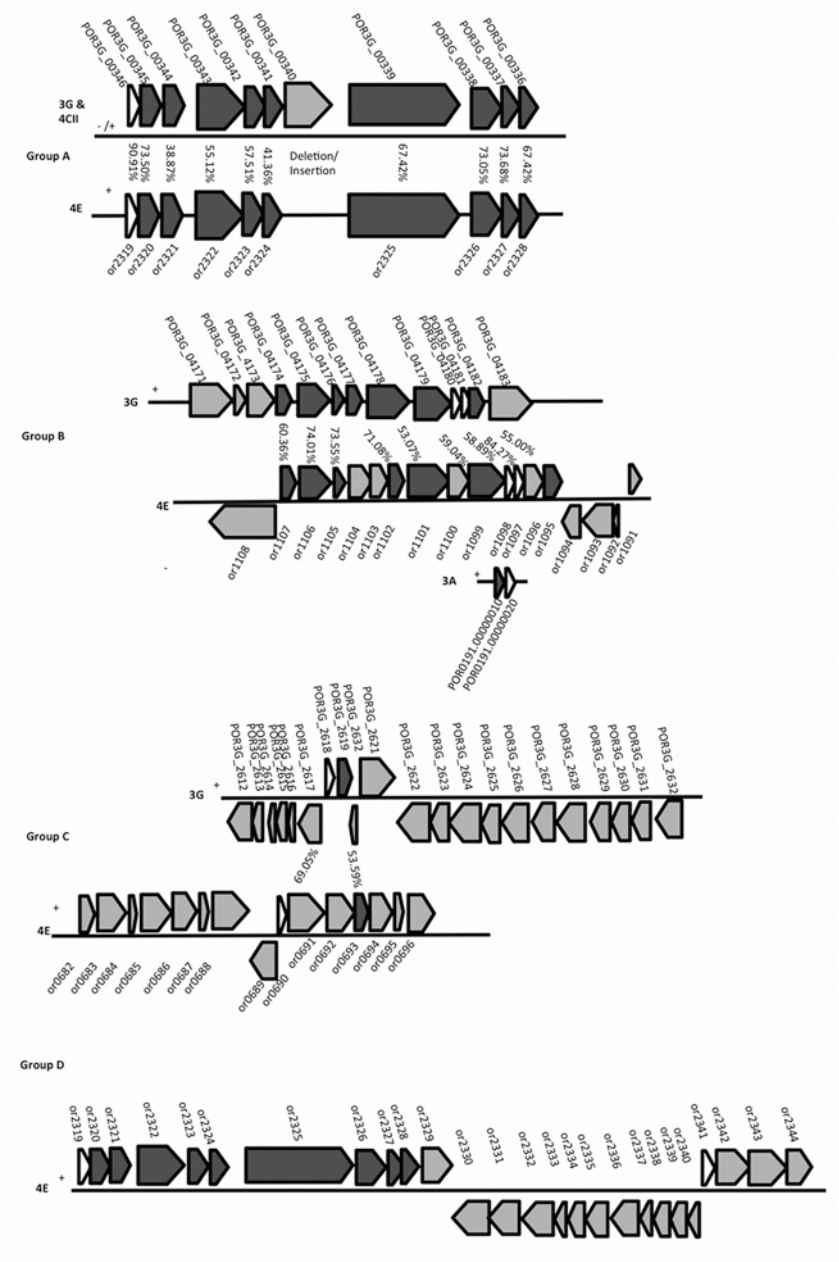


Figure 7: Schematic view of poribacterial BMC shell protein groups. For a better overview all genes are displayed in 5'-3' direction of the BMC shell protein gene. The actual strand orientation might be different and is indicated by plus or minus signs. Genes are shown with locus taq and amino acid identities based on IMG/MER homology searches are shown between genes where applicable. BMC shell protein genes are shown in white, other genes with homologies between different SAGs are shown in dark grey, other genes are shown in light grey.

A detailed description of these groups can be found in the supplemental material Text S1. BMCs are proteinaceous structures that
enclose sets of enzymes of diverse functions performing a chain of reactions within the compartment [33]. BMC shell functions have been described as concentrating enzymes and substrates together to increase reaction efficiency, protection of e.g. oxygen sensitive enzymes, enclosure of toxic or volatile metabolites that are produced/consumed by enzymes in the shell, and concentrating metabolites to increase efficiency [33,34]. Kerfeld et al. [33] suggested that at least two (or more) pfam00936 domain proteins and one pfam03319 domain protein might be required as building blocks of functional BMCs. Out of all poribacterial SAGs only 4E encoded for more than one pfam00936 domain and, with the exception of SAG 4CII, all poribacterial SAGs encoded for a higher number of pfam03319 than pfam00936 domains (Table 2). This is unusual when compared to most other BMC shell protein studied to date (Suppl. Table S2). *Poribacteria,* together with *Planctomycetes*, the candidate phylum *Atribacteria* (OP9), and some additional phyla (Suppl. Table S2), appear to be among the few exceptions containing more pfam03319 than pfam00936 domains.

The so far best described BMC functions are the carboxysome and BMCs containing enzymes for ethanolamine or propanediol utilization. Bioinformatic analysis of all available BMC shell protein encoding genomic regions at the time by Yeates et al. [35] revealed that functional proteins within the BMCs are often encoded in close proximity of the BMC shell proteins and identified a set of functions regularly occurring with BMC shell proteins. However, the genes in poribacterial BMC clusters did not show direct similarities to any of these previously described functions but some genes in poribacterial BMC clusters give an indication of potential functions. It is noteworthy that many of the described enzymatic reactions in previously described BMCs are co-factor dependent (often vitamin B12), and that the co-factor biosynthesis genes were often found in close proximity to BMC shell protein genes [33]. In poribacterial BMC group B we found genes for riboflavin (vitamin B2) biosynthesis, which might indicate a riboflavin dependent process occurring in poribacterial BMCs. Riboflavin is a major cofactor in many processes of the energy metabolism. To our knowledge riboflavin biosynthesis genes have so far not been described from other BMC shell gene clusters. Further investigations will reveal, whether there are indeed BMCs with riboflavin dependent reactions. Furthermore, poribacterial BMC gene clusters show similar regulatory systems to previously described clusters. A recent study by Jorda et al. [36] identified

clusters of BMCs shared between different organisms by comparing similarities of genes in the genomic neighborhoods of BMC shell proteins. They identified two BMC clusters that are characterized by a two-component regulatory system with a signal transduction histidine kinase and response regulator receiver [36]. Poribacterial BMC clusters appear to similarly regulated, since we also detected genes of a two-component regulatory system in three out of four described poribacterial BMC groups (see Fig. 3, and supplemental text S1). However, none of the functional genes on poribacterial BMC clusters showed similarities to those on the clusters described by Jorda et al. [36] and therefore the true functions of poribacterial BMCs remain to be investigated.

It is suspected that novel BMC functions will be revealed in the future [36] especially from genomes with a more scattered operon structure [33]. This might also be the case for *Poribacteria* where the identified genomic regions with BMC shell protein genes (group A-C) appear scattered across the genome. For example, the different BMC shell protein genes (with pfam00936 and pfam03319) are generally in different genomic regions on poribacterial genomes and not encoded together within one region, as it is the case for many so far functionally characterized BMC types [33]. Functional components of poribacterial BMCs might therefore also be encoded on different genomic regions. Alternatively, the existence of only one pfam00936 domain and the occurrence of transposase genes in BMC gene clusters B and D (see text S1) might indicate lack of function [33]. Future efforts are needed to resolve this issue for *Poribacteria*.

Eukaryote-like repeat proteins

Eukaryote-like repeat domain containing proteins have received much recent attention in sponge microbiology and their involvement in mediating host-microbe interactions has been postulated. Especially ankyrin (ANK) and tetratricopeptide repeats (TPR) have been in focus of such investigations [37-39]. To examine the role of these domains on poribacterial SAGs we searched for proteins with pfam hits to repeat and eukaryote-like domains in the IMG/MER database and also compared these to all finished genomes of free-living marine bacteria available in the IMG database in July 2013 (n=98). We detected 41 such domains on poribacterial SAGs. The majority of these showed a higher domain frequency per total genes on at least one poribacterial SAG when compared to the average frequency of this domain on genomes of free-living

marine bacteria (Fig. 4, Table S7). For 14 pfam domains the frequency on poribacterial genomes was even higher than the maximum frequency of this domain on the genome of any free-living marine bacterium. Many domains occurred simultaneously on the same genes with a total of 668 domains in all poribacterial SAGs on 490 encoded proteins (3A: 15 domains on 11genes, 3G: 335 domains on 240 genes, 4C: 95 domain on 75 genes, 4CII: 24 domains on 16 genes, 4E: 181 domains on 135 genes, and 4G: 17 domains on 8 genes).

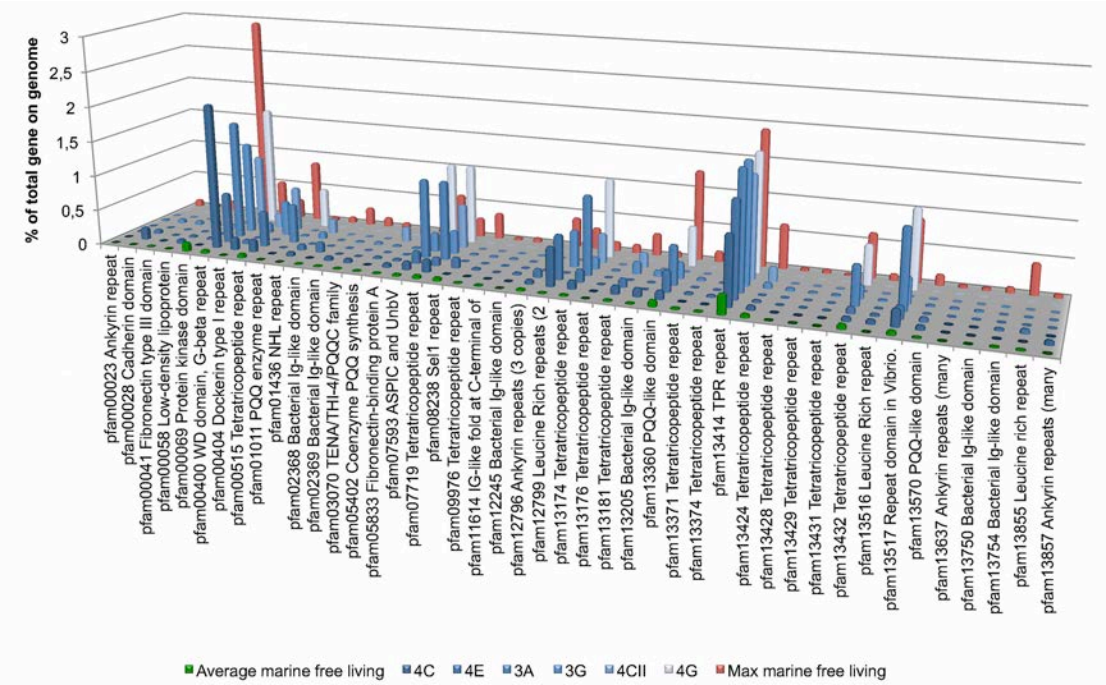


Figure 8: Bar plot showing frequency of eukaryote-like pfam domains found on poribacterial SAGs in comparison to the average and maximum frequency on all finished genomes of marine free-living bacteria available in IMG in July 2013.

Among the most abundant domains were TPRs with pfams 013414, 00515, 07719, 13432, 13174, and 13181, which were also represented by eight other pfams (13424, 13374, 13371, 09976, 13431, 13429, 13428, and 13176) but in lower abundances. We were also able to find Sel1 repeat like proteins domains encoded on poribacterial SAGs 3G and 4E (0.02 and 0.15% of total genes, respectively) which have a similar structure to TPRs [40]. In total TPRs represented the highest frequency of repeat domains on poribacterial SAGs. Furthermore WD40 domains (pfam00400) were highly abundant on poribacterial SAGs, as well as two-copy leucin rich repeats (LRR) (pfam 12799), and the VCBS domain (pfam 13517) which is a domain found in high numbers in the genera *Vibrio*, *Colwellia*, *Bradyrhizobium* and *Shewanella*. Pfam domain 07593- ASPIC and

UNbV was also present on several poribacterial SAGs in multiple copies. ANK repeat domains were detected (pfam 12796, 13637, 13857, and 00023) in lower numbers on a total of 14 genes on SAGs 3G, 4C and 4E (Table S7). The frequency of genes with pfam domains representing ankyrin repeats was often higher than average compared to the genomes of free-living marine bacteria (Table S7).

The occurrence of low-density lipoprotein (LDL) receptor repeat class B domains (pfam00058) on poribacterial genomes seemed noteworthy. We found these domains on one gene in each SAG 4C and 4E as well as on five genes in SAG 3G. Outside of *Poribacteria* this domain has only been found in proteins of 14 bacterial genomes but not in archaeal genomes publically available at the IMG/MER database in July 2013. Most of these bacterial hits however do not show the tandem repeats that are characteristic for this domain in eukaryotes. Such tandem repeats were only detected in the poribacterial proteins and proteins of four other bacterial genomes. Amongst these were free-living marine cyanobacteria (*Cyanothece* species, *Pleurocapsa* sp. PCC 7327), the marine deep sea piezophile *Mortiella* sp. PE36, and the strictly anaerobic bacterium *Paludibacter propionicigenes* WB4, DSM 17365. The LDL receptor is best described in mammals where they transport ligands into the cell for degradation by lysosomes and plays a role in cholesterol homeostasis [41]. The LDL repeat domain class B is part of the region of the LDL receptor which is responsible for ligand release and receptor recycling [42]. Virtually nothing is known about such domains in bacteria and it remains to be investigated whether there is a real connection to eukaryotic domains.

Although the limited data did not allow for any functional assignments of the LDL receptor genes, a role on the cell surface seems very likely in *Poribacteria* since all of the discovered genes with these domains had predicted transmembrane helices (TMHs) (~86%) with the majority of the protein located outside of the cell or signal peptides (SPs) (~14%). TMHs and SPs were also frequently predicted on genes representing other eukaryote-like proteins of *Poribacteria* (Suppl. Table S8 and S9). High abundances (≥50% of genes with this pfam) of either TMHs or SPs were found on genes also encoding for bacterial Ig like domain protein genes, PQQ enzyme repeat containing genes, fibronectin type III domain and cadherin domain genes. Also genes with some of the pfams domains representing LRR and TPRs showed strong representation of TMH and SPs. Additionally, many poribacterial eukaryote-like domain genes (especially WD40 repeats) encoded for a domain belonging to the Por secretion

system C-terminal sorting domain family (TIGR04183) (Suppl. Table S9), which is characteristic of proteins with outer membrane locations. Such secretion systems are often used by pathogens to transport virulence factors or carbohydrate degrading enzymes of polysaccharide degrading gut bacteria [43-45] . In *Poribacteria* this secretion system might represent an additional way to transport eukaryote-like domain containing proteins.

Our findings support previous reports of repeat and eukaryote-like domains being highly abundant in symbionts of marine sponges. The identification of proteins with these domains from the microbial communities of the sponge *Cymbastella concentrica* by ways of metaproteogenomics [46] pointed towards an active functional role of these proteins. ANK domain proteins of sponge symbionts have been suspected to be involved in preventing phagocytosis by the sponge host as in analogy to similar functions of ANK domain proteins in bacterial pathogens *Legoniella pneumophila* or *Coxiella burnetti* [39,47]. Indeed, in a recent paper Nguyen et al. [48] were able to show that ANK proteins from a marine sponge symbiont that were expressed in *E.coli* prevent phagocytosis of the bacterial cells by amoeba. The authors suggested this to be a function of sponge symbionts to avoid digestion by their host [48]. Thus, poribacterial ANK proteins may also facilitate similar functions.

LRRs have been found in proteins of pathogenic bacteria such as *Yersinia* species where LRRs are part of important virulence factors [49] or *Listeria monocytogenes* which encodes for LRR containing protein InlB that aids in host cell invasion [50]. Also TPRs were shown to be involved in different functions of pathogenesis [51] and fibronectin domains were shown to play a role in host-pathogen interactions as well, although in this case bacterial proteins bind to the fibronectin domains of the host protein [52,53]. It would be interesting to explore whether bacterial fibronectin domains might be used in a similar way. Furthermore, fibronectin III domains have been found in polysaccharide degrading extracellular enzymes of *Clostridium thermocellum* [54]. Hentschel et al. [47] speculated that such functions in sponge symbionts could be connected to interactions with molecules of the sponge host extracellular matrix and our recent investigations of poribacterial carbohydrate degradation potential [14] support this hypothesis. However, at the current stage, we are just beginning to decipher the real functions of eukaryote-like proteins in *Poribacteria*. As many of these proteins may not be located outside of the poribacterial cell, as indicated by the large amount of

proteins detected without TMHs or SPs (Table S9), it appears likely that at least some may mediate intracellular protein-protein interactions.

High abundance of phyH -domain containing proteins

Among poribacterial genomes we found a remarkably high occurrence of genes encoding for proteins with pfam domain pfam05721-phyH (Suppl. Table S10). This pfam describes a protein family containing eukaryotic phytanoyl-CoA dioxygenase proteins, ectoine hydroxylases from eukaryotes and prokaryotes, and several bacterial deoxygenases of mostly unknown function (http://pfam.sanger.ac.uk/family/PF05721). These proteins are Fe(II) and 2-oxoglutarate dependent oxygenases that catalyze a wide range of oxidative reactions Among bacterial phyH genes are some potentially involved in quorum sensing [55,56], synthesis of the compatible solute 5-hydroxyectoine [57], and utilization of phosphorous sources [58,59]. We screened for this domain in all genomes publically available in the IMG/MER database in July 2013. All poribacterial genomes showed a frequency of more than 1.9% genes with this domain per total number of genes (Suppl. Table S10). All other genomes available in the database at the time (independent of its domain Bacteria, Archaea, or Eukaryota) showed a frequency of less than 0.049% of genes with this domain per total genes. This large abundance of genes belonging to the same pfam family might indicate an importance of the related functions for *Poribacteria*.

A clustering analysis of poribacterial sequences showed that there was large diversity amongst poribacterial phyH family genes with 305 sequences clustering in 193 clusters with 60% aa id threshold (Suppl. Table S11). For the majority of poribacterial genes with this domain a reliable functional annotation could not be made. Best homologies were usually between genes of poribacterial SAGs, despite the high diversity indicated by the clustering analysis. Some of the poribacterial phyH family genes also showed homology to another uncharacterized deoxygenase encoded on the first genome fragment sequence from a poribacterial metagenome clone 64K2 [60]. This might indicate *Poribacteria*-specific functions within the phyH family.

Although the majority of poribacterial phyH genes remained without further functional characterization, we were still able to make functional predictions in some cases. Poribacterial SAGs 3G and 4E encoded for phyH genes (OID 2265144857 and 2265139858, respectively) with homologies (40% aa id each) to a 2-

aminoethylphosphonate (2-AEPn) utilization gene (phnY) for which function was experimentally proven [59]. These poribacterial genomes also encoded directly upstream of this gene for a protein of the HD phosphohydrolase family (phnZ) (OID 2265144856 in 3G and 2265139857 in 4E), which is the only other gene involved in this 2-AEPn utilization pathway [59]. Both poribacterial genomes further encoded for another predicted phosphohydrolase downstream of the previously described genes with as of yet unknown function in this pathway. 2-AEPn is assumed to be one of the biggest sources of dissolved organic phosphorous in the oceans [61,62] and represents an alternative phosphorous source to the often limited dissolved inorganic phosphorous. The use of dissolved organic phosphorous i.e. phosphonates by many marine bacteria has been described before [63,64]. Phosphonates such as 2-AEPn are found largely in phospholipids of marine invertebrates including sponges and are also produced by some marine bacteria [65-68]. Therefore organic phosphorous sources should be largely available in the sponge mesohyl and the ability to utilize 2-AEPn as a phosphorous source might therefore be a competitive advantage. The presence of both genes identified as essential for 2-AEPn utilization [59] indicated the presence of this pathway also in *Poribacteria* and elucidated one possible function of phyH superfamily genes in this candidate phylum.

Concluding remarks

Our study demonstrates the power of single-cell genomics to reveal features of candidate phyla, which could, to our knowledge, not be characterized in any other way. Here we show by use of phylogenetic and phylogenomic analyses that *Poribacteria* are not members of the PVC superphylum, but rather form a distinct monophyletic phylum in close proximity. We provide genomic evidence for bacterial microcompartments in *Poribacteria* that show very little similarity to any previously described BMCs. Further novel functions might be hidden in the various eukaryote-like protein domains, which may be involved in mediating host-microbe interactions within the sponge holobiont. The high abundance of diverse phyH-domain containing proteins points to important and potentially specific functions in *Poribacteria*. Most of these functions remain to be revealed in future studies but some show the genomic potential for organic phosphorous utilization. Our analyses show how genome sequences can help to revisit past hypotheses and at the same time open the way for new investigations by revealing

novel functional features. Challenges for the future will be to experimentally demonstrate function and to ultimately understand the implications for symbiosis.

Acknowledgements

Financial support to U.H. was provided by the SFB630- grant TPA5, the SFB567-grant TPC3, and by the Bavaria California Technology Center (BaCaTeC). T.W., C.R., P.S., N.I and K.M. were funded by the U.S. Department of Energy Joint Genome Institute, Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

1. Kamke J, Bayer K, Woyke T, Hentschel U (2012) Exploring symbioses by single-cell genomics. Biol Bull 223: 30–43.

2. Stepanauskas R (2012) Single cell genomics: an individual look at microbes. Curr Opin Microbiol 15: 613–620.

3. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. Nature 499: 431–437.

4. Fieseler L, Horn M, Wagner M, Hentschel U (2004) Discovery of the novel candidate phylum "*Poribacteria*" in marine sponges. Appl Environ Microb 70: 3724–3732.

5. Pham VD, Konstantinidis KT, Palden T, DeLong EF (2008) Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific subtropical gyre. Environ Microbiol 10: 2313–2330.

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

7. Taylor MW, Tsai P, Simister RL, Deines P, Botte E, et al. (2013) "Sponge-specific" bacteria are widespread (but rare) in diverse marine environments. ISME J 7: 438–443.

8. Lafi FF, Fuerst JA, Fieseler L, Engels C, Goh WWL, et al. (2009) Widespread distribution of *Poribacteria* in *Demospongiae*. Appl Environ Microb 75: 5695–5699.

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

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

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

12. Wagner M, Horn M (2006) The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotech 17: 241–249.

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

14. Kamke J, Sczyrba A, Ivanova N, Schwientek P, Rinke C, et al. (2013) Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. ISME J. In press. doi:10.1038/ismej.2013.111.

15. Markowitz VM, Kyrpides NC (2007) Comparative genome analysis in the integrated microbial genomes (IMG) system. Methods Mol Biol 395: 35–56.

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

17. Quast C, Pruesse E, Yilmaz P, Gerken J (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41: D590–D596.

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

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

20. Pruesse E, Peplies J, Glöckner FO (2012) SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28: 1823–1829.

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

22. Finn RD, Clements J, Eddy SR (2011) HMMER web server: interactive sequence similarity searching. Nucleic Acids Res 39: W29–W37.

23. Price MN, Dehal PS, Arkin AP (2010) FastTree 2--approximately maximum-

likelihood trees for large alignments. PLoS ONE 5: e9490. doi:10.1371/journal.pone.0009490.

24. Stamatakis A, Aberer AJ, Goll C, Smith SA, Berger SA, et al. (2012) RAxML-Light: a tool for computing terabyte phylogenies. Bioinformatics 28: 2064–2066.

25. Wu D, Hugenholtz P, Mavromatis K, Pukall R, Dalin E, et al. (2009) A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. Nature 462: 1056–1060.

26. Gupta RS, Bhandari V, Naushad HS (2012) Molecular signatures for the PVC clade (*Planctomycetes*, *Verrucomicrobia*, *Chlamydiae*, and *Lentisphaerae*) of bacteria provide insights into their evolutionary relationships. Front Microbio 3. doi: 10.3389/fmicb.2012.00327.

27. Lang J, Lang J, Lang J, Darling A, Darling A, et al. (2013) Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices. PLoS ONE 8: e62510. doi:10.1371/journal.pone.0062510.

28. Simister RL, Deines P, Botté ES, Webster NS, Taylor MW (2011) Sponge-specific clusters revisited: a comprehensive phylogeny of sponge-associated microorganisms. Environ Microbiol 14: 517–524.

29. Fuerst JA (2005) Intracellular compartmentation in planctomycetes. Annu Rev Microbiol 59: 299–328.

30. Fuerst JA, Webb RI (1991) Membrane-bounded nucleoid in the eubacterium *Gemmatata obscuriglobus*. Proc Natl Acad Sci USA 88: 8184–8.

31. Santarella-Mellwig R, Franke J, Jaedicke A, Gorjanacz M, Bauer U, et al. (2010) The compartmentalized bacteria of the *Planctomycetes-Verrucomicrobia-Chlamydiae* superphylum have membrane coat-like proteins. Plos Biol 8: e1000281. doi:10.1371/journal.pbio.1000281.t002.

32. Santarella-Mellwig R, Pruggnaller S, Roos N, Mattaj IW, Devos DP (2013) Three-dimensional reconstruction of bacteria with a complex endomembrane system. Plos Biol 11: e1001565. doi:10.1371/journal.pbio.1001565.s014.

33. Kerfeld CA, Heinhorst S, Cannon GC (2010) Bacterial Microcompartments. Annu Rev Microbiol 64: 391–408.

34. Yeates TO, Crowley CS, Tanaka S (2010) Bacterial microcompartment organelles: protein shell structure and evolution. Annu Rev Biophys 39: 185–205.

35. Yeates TO, Thompson MC, Bobik TA (2011) The protein shells of bacterial microcompartment organelles. Curr Opin Struc Biol 21: 223–231.

36. Jorda J, Lopez D, Wheatley NM, Yeates TO (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci 22: 179–195.

37. Thomas T, Rusch D, DeMaere MZ, Yung PY, Lewis M, et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J 4: 1557–11567.

38. Siegl A, Hentschel U (2009) PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. Environ Microbiol Rep 2: 507–513.

39. Fan L, Reynolds D, Liu M, Stark M, Kjelleberg S, et al. (2012) Functional equivalence and evolutionary convergence in complex communities of microbial sponge symbionts. Proc Natl Acad Sci USA 109: E1878–E1887.

40. Mittl PR, Schneider-Brachert W (2007) Sel1-like repeat proteins in signal transduction. Cell Signal 19: 20–31.

41. Hussain MM, Strickland DK, Bakillah A (1999) The mammalian low-density lipoprotein receptor family. Annu Rev Nutr 19: 141–172.

42. Davis CG, Goldstein JL, Südhof TC, Anderson RG, Russell DW, et al. (1987) Acid-dependent ligand dissociation and recycling of LDL receptor mediated by growth factor homology region. Nature 326: 760–765.

43. Sato K, Naito M, Yukitake H, Hirakawa H, Shoji M, et al. (2010) A protein secretion system linked to bacteroidete gliding motility and pathogenesis. Proc Natl Acad Sci USA 107: 276–281.

44. Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, et al. (2011) The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. PLoS ONE 6: e18814. doi:10.1371/journal.pone.0018814.

45. Nguyen KA, Travis J, Potempa J (2007) Does the importance of the C-terminal residues in the maturation of RgpB from *Porphyromonas gingivalis* reveal a novel mechanism for protein export in a subgroup of gram-negative bacteria? J Bacteriol 189: 833–843.

46. Liu MY, Kjelleberg S, Thomas T (2010) Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. ISME J 5: 427–435.

47. Hentschel U, Piel J, Degnan SM, Taylor MW (2012) Genomic insights into the marine sponge microbiome. Nat Rev Micro 10: 641–654.

48. Nguyen MTHD, Liu M, Thomas T (2013) Ankyrin-repeat proteins from sponge symbionts modulate amoebal phagocytosis. Mol Ecol. In press. doi:10.1111/mec.12384.

49. Viboud GI, Bliska JB (2005) Yersinia outer proteins: role in modulation of host cell signaling responses and pathogenesis. Annu Rev Microbiol 59: 69–89.

50. Bierne H, Cossart P (2002) InlB, a surface protein of Listeria monocytogenes that behaves as an invasin and a growth factor. J Cell Sci.115: 3357–67.

51. Cerveny L, Straskova A, Dankova V, Hartlova A, Ceckova M, et al. (2013) Tetratricopeptide repeat motifs in the world of bacterial pathogens: Role in virulence mechanisms. Infect Immun 81: 629–635.

52. Schwarz-Linek U, Werner JM, Pickford AR (2003) Pathogenic bacteria attach to human fibronectin through a tandem beta-zipper. Nature 423: 177–181.

53. Schwarz-Linek U, Höök M, Potts JR (2004) The molecular basis of fibronectin-mediated bacterial adherence to host cells. Mol Microbiol 52: 631–641.

54. Kataeva IA, Seidel RD, Shah A, West LT, Li XL, et al. (2002) The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. Appl Environ Microb 68: 4292–4300.

55. Schofield CJ, McDonough MA (2007) Structural and mechanistic studies on the peroxisomal oxygenase phytanoyl-CoA 2-hydroxylase (PhyH). Biochem Soc Trans 35: 870–875.

56. Hao Y, Winans SC, Glick BR, Charles TC (2010) Identification and characterization of new LuxR/LuxI-type quorum sensing systems from metagenomic libraries. Environ Microbiol 12: 105–117.

57. Reuter K, Pittelkow M, Bursy J, Heine A, Craan T, et al. (2010) Synthesis of 5-hydroxyectoine from ectoine: Crystal structure of the non-heme Iron(II) and 2-oxoglutarate-dependent dioxygenase EctD. PLoS ONE 5: e10647. doi:10.1371/journal.pone.0010647.

58. White AK, Metcalf WW (2007) Microbial metabolism of reduced phosphorus compounds. Annu Rev Microbiol 61: 379–400.

59. Martinez A, Tyson GW, DeLong EF (2010) Widespread known and novel phosphonate utilization pathways in marine bacteria revealed by functional screening and metagenomic analyses. Environ Microbiol 12: 222–238.

60. Fieseler L, Quaiser A, Schleper C, Hentschel U (2006) Analysis of the first genome fragment from the marine sponge-associated, novel candidate phylum *Poribacteria* by environmental genomics. Environ Microbiol 8: 612–624.

61. Kolowith LC, Ingall ED, Benner R (2001) Composition and cycling of marine organic phosphorus. Limnol Oceanogr 46: 309–320.

62. Clark LL, Ingall ED, Benner R (1999) Marine organic phosphorus cycling: novel insights from nuclear magnetic resonance. Am J Sci 299: 724–737.

63. Villarreal-Chiu JF, Quinn JP, McGrath JW (2012) The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. Front Microbio 3. doi: 10.3389/fmicb.2012.00019.

64. Dyhrman ST, Chappell PD, Haley ST, Moffett JW, Orchard ED, et al. (2006) Phosphonate utilization by the globally important marine diazotroph *Trichodesmium*. Nature 439: 68–71.

65. Mukhamedova KS, Glushenkova AI (2000) Natural Phosphonolipids. Chemistry of Natural Compounds 36: 329–341.

66. Dembitsky VM, Rezanka T, Srebnik M (2003) Lipid compounds of freshwater

sponges: family Spongillidae, class Demospongiae. Chemistry and Physics of Lipids 123: 117–155.

67. Dembitskii VM (1988) Lipids of marine origin. IV. 1,2-Di-O-alkylglycerophospho- and -phosphonolipids from the marine sponge *Ectyodoryx kovdaicum*. Chem Nat Compd 24: 642–643.

68. Dyhrman ST, Benitez-Nelson CR, Orchard ED, Haley ST, Pellechia PJ (2009) A microbial source of phosphonates in oligotrophic marine systems. Nature Geosci 2: 696–699.

6. General Discussion

6.1 A retrospective of the candidate phylum *Poribacteria*

In this PhD thesis I used single-cell genomics to analyse the genomic potential of the candidate phylum *Poribacteria*. These studies are an example of the power of single-cell genomics to gain detailed information about a candidate phylum for which most previous information was based on the 16S rRNA gene. From their initial discovery (Fieseler et al. 2004) our knowledge of *Poribacteria* has constantly grown. The poribacterial single-cell genomics data have significantly improved and revised some of the earlier discoveries about *Poribacteria* and thus made this candidate phylum one of the best characterised sponge symbionts today. Here I review how the single-cell genomics studies of this PhD thesis contributed to our knowledge of *Poribacteria*.

6.1.1. Phylogeny

Upon discovery (Fieseler et al. 2004) *Poribacteria* were classified as part of the PVC superphylum. This was further supported by an additional phylogenetic analysis of the members of the PVC superphylum (Wagner and Horn, 2006). In chapter 5 of this thesis phylogenetic and phylogenomic analyses made it possible to re-resolve the taxonomic structure of *Poribacteria* on the phylum level and refuted the initial classification by showing a distinct position outside of PVC superphylum. However, it was not possible to place the candidate phylum *Poribacteria* with high confidence in context to other phyla, which is likely due to the fact that insufficient genomic information is currently available from relevant phyla that might cluster with *Poribacteria*. The phylogenomic tree (Fig.1 Chapter 5) as well as the 16S rRNA gene based phylogenetic tree (Fig. 2, Chapter 5) indicated that the positioning of *Poribacteria* might be influenced by the availability of data from other candidate phyla such as *Hydrogenedentes* or SAUL. It is thus very likely that with more genomic data becoming available the phylogenetic positioning of *Poribacteria* will have to be revisited.

6.1.2. Diversity and abundance

Poribacteria have been described as highly abundant symbionts of many marine sponges (Fieseler et al. 2004; Lafi et al. 2009). In addition to the high abundance a large intra-phylum diversity became apparent, especially with the emergence of 16S rRNA gene amplicon sequencing studies, which revealed more than 400 OTUs based on 97% 16S rRNA gene sequence similarity (Schmitt et al. 2012). Currently there is no analysis of the individual abundance of different poribacterial phylotypes. The findings of this PhD thesis indicate that not all poribacterial phylotypes are equally abundant within marine sponges. The six obtained poribacterial single cells were classified based on their 16S rRNA gene phylogeny, average nucleotide identity (ANI), and tetra nucleotide frequency profile (Chapter 4) into three different phylotypes. Four out of six SAGs grouped together into one phylotype that was termed group I. Additionally the 16S rRNA gene obtained from a fosmid clone of an *A. aerophoba* metagenome bank (accession number AY713479) (Fieseler et al. 2006) would also fall into group I (Fig.1, Chapter 4). The strong representation of group I phylotypes indicates that this might be the dominant phylotype in *A. aerophoba*. Since the retrieval of single-cells from microbial communities is independent of PCR biases this might reflect a true difference in the abundances of poribacterial phylotypes, assuming cell lysis efficiency is equal between different poribacterial phylotypes. Further investigation will need to reveal abundance patterns of different poribacterial phylotypes in sponges and sponge species. This information would be useful to interpret how much of the genomic information obtained by single-cell genomics is related to what proportion of the poribacterial community and if it can be referred to other sponge species.

6.1.3. Cell compartmentation

Cell compartmentation in the form of a double membrane bound nucleoid-like structure was described from symbionts of different marine sponges (Fuerst et al. 1998). Such microbial cells were also found in the mesohyl of the *A. aerophoba* and six different morphotypes were described (Fieseler et al. 2004). The simultaneous occurrence of ring-shaped *Poribacteria*-specific FISH signals, which suggested a ribosome-free area in the cell centre, indicated the phylogenetic identity of these cells to be members of the candidate phylum *Poribacteria* (Fieseler et al. 2004). Single-cell genomics analysis did not reveal any indication of such a cell compartment (Chapter 5). However, the analyses showed the genomic evidence for the presence a different kind of cell

compartment in the form of bacterial microcompartments. The physical presence of such microcompartments remains to be revealed in symbionts of marine sponges by electron microscopy. Since these compartments are usually only synthesised in the presence of the substrates that are metabolised or degraded in the compartment (Kerfeld et al. 2010), an absence of such compartments in current electron microscopy studies does not necessarily imply a general absence of these compartments. Furthermore the poribacterial microcompartment most likely represents a novel type of these structures (see Chapter 5) and it is possible that poribacterial compartments show a different shape to those of other bacteria and have thus not been recognised as bacterial microcompartments.

6.1.4. Similarity to fosmid clone 64K2

Fieseler and colleagues (2006) sequenced the first poribacterial genome fragment from a fosmid library containing a 16S rRNA gene and 27 functional orfs. This fosmid sequence encoded for a phyH-like protein domain. Genes with this domain were found to be the most abundant functional genes on poribacterial SAGs (Chapter 5). The majority of the poribacterial phyH genes, also including the phyH gene on the fosmid clone, could not yet be further functionally classified. This shows that there is still a lack of suitable reference sequences for functional gene characterisation in the current databases. Especially for candidate phyla genomic information is still rare and therefore functional analysis can be difficult.

Another similarity between the fosmid clone sequence and the poribacterial SAGs were the characteristics of the 16S rRNA gene. All poribacterial SAGs and the fosmid sequence encoded for only one 16S rRNA gene. Additionally in all sequences except for SAG 4E, which represents a potentially rare phylotype, the 16S rRNA gene was unlinked to the remaining rRNA genes. The presence of unlinked ribosomal RNA genes is not restricted to *Poribacteria* but was reported from several other phyla (Liesack and Stackebrandt, 1989; Bensaadi-Merchermek et al. 1995; Ruepp et al. 2000; Boyer et al. 2001; Rurangirwa et al. 2002; Tamas et al. 2002; Glöckner et al. 2003; Henne et al. 2004). In *Poribacteria* it might be indicative for specific poribacterial phylotypes including the dominant group I. Furthermore a single copy of the 16S rRNA has been related to a slower response to the availability of growth substrates and therefore a slower growth in general (Klappenbach et al. 2000). With the constant food supply in

the sponge mesohyl (Chapter 4) *Poribacteria* might not need to respond quickly to nutrient availability and therefore a potentially slow growth rate might not be a disadvantage.

6.1.5. Secondary metabolism

Previous studies showed the presence of a unusually small polyketide synthase (PKS) system in several metagenomes of marine sponges which was subsequently termed sponge ubiquitous PKS (Sup-PKS) and was predicted to produce methyl-branched fatty acids (Fieseler et al. 2007). *Poribacteria* were suggested to be producers of this type of PKS system as their presence correlated with the occurrence of the relevant genes in the sponge metagenomes and the detection of the relevant methyl-branched fatty acids in the different sponge species (Hochmuth et al. 2010). Furthermore a whole genome amplification product of a FACS sorted single microbial cell from the microbial community of *A. aerophoba* tested PCR positive for a poribacterial 16S rRNA gene and for the genes of the Sup-PKS system (Siegl and Hentschel, 2009) and thus strengthened the hypothesis that Sup-PKS was *Poribacteria* derived. This PKS system was also encoded together with a second, previously undescribed PKS system on the sequence fragments of the first poribacterial SAG WGA A3 (Chapter 3). Due to contamination problems that were only later detected in this dataset (Chapter 4) and the fragmented nature of the relevant reads it was not possible to identify the phylogenetic origin of these PKS systems and the association to *Poribacteria* in dataset WGA A3 could not be further supported. In contrast to previous findings, none of the five newly obtained poribacterial SAGs encoded any PKS system. Therefore the Sup-PKS system cannot with certainty be regarded as of poribacterial origin and additional studies are needed to resolve this issue in the future.

Similarly the genome fragments encoding for potential sterol biosynthesis proteins could also not be taxonomically assigned after the detection of contaminating reads in dataset WGA A3. During this PhD thesis primers based on the sterol biosynthesis gene fragments were designed and used to screen an *A. aerophoba* metagenome fosmid library to identify further genes related to this pathway. A fosmid clone was identified that contained full sequences for both sterol biosynthesis genes found in fragments on dataset WGA A3, the oxidosqualene cyclase and the 24-sterol C-methyltransferase (unpublished data). However, neither one of these genes, nor any of the other genes on

the fosmid clone sequence showed any homologies to genes on the other five poribacterial SAGs. This indicated that the sterol biosynthesis genes discovered in WGA A3 are not likely to be of poribacterial origin. However, the fact that these genes were found in both the metagenomic fosmid library and dataset WGA A3 points towards an origin of another sponge symbiont that might be related to *Candidatus* Nitrospira defluvii, as indicated by homologies of the contaminating DNA in WGA A3 (Chapter 4). Therefore the hypothesis that bacterial symbionts might be the producers of 24-isopropyl-cholestane, a fossil biomarker used to estimate the evolutionary origin of *Porifera* (Love et al. 2008) might still be valid.

6.1.6. Shared functional features within *Poribacteria*

Comparative genome analysis of the six poribacterial SAGs revealed a strong potential for carbohydrate degradation in *Poribacteria* (Chapter 4). This appeared to be the main functional feature shared between all detected phylotypes and pointed to many potential nutrient sources for *Poribacteria*. The detailed analysis of the poribacterial SAGs also showed that different poribacterial phylotypes encoded for specific functional pathways within the carbon metabolism. For example, the genomic potential for glycolysis was only encoded on SAG 4E while group I genomes showed no such genomic potential but instead encoded for enzymes to directly transfer glucose into the Entner-Doudoroff pathway (Chapter 4). Similarly, all poribacterial phylotypes showed the genomic potential for uronic acid polymer degradation but the different phylotypes encoded for different variants of the metabolic capacities to achieve this. SAG 4C and group I genomes encoded for the pectate lyase pathway while SAG 4E encoded for polygalacturonase and subsequent enzymes (Chapter 4). These observations are similar to what was found for nitrogen metabolism in metagenomes of microbial communities of six different sponge species (Fan et al. 2012). Genes for pathways such as denitrification and ammonium oxidation were encoded on the genomic data of microbial communities of six different sponge species but the there was a difference in pathway variants between the datasets of different hosts. This led to the suggestion that there is a large degree of functional equivalence in the nitrogen cycle of marine sponges (Fan et al. 2012). Ribes et al. (2012) came to similar conclusions about the nitrogen cycle in two Mediterranean sponges. The authors measured similar rates for nitrogen cycling but also detected a large variation in the phylogenetic origin of the relevant functional genes between the microbial communities of the two sponges. The

poribacterial data indicate that there might also be a degree of functional equivalence for carbohydrate metabolism. This shows that functional equivalence in the sponge microbiome does not only occur between communities of different host species but also within one sponge species and between members of the same bacterial phylum.

6.2. Ecological considerations

The main functional discovery of the single-cell genomics study of *Poribacteria* is their diverse and complex genomic potential for the degradation of organic substrates, especially carbohydrates. The poribacterial glycobiome was revealed in Chapter 4. It offers several possible nutrient sources for *Poribacteria* as a sponge symbiont from inside and outside the host environment. Furthermore this PhD thesis showed that *Poribacteria* also have the capacity to use organic sulphur (Chapter 4) and phosphate sources (Chapter 5).

Assuming that there is a degree of functional equivalence in microbial communities of marine sponges, as indicated by metagenomic and physiological studies (Ribes et al. 2012; Fan et al. 2012), the functional profile of *Poribacteria* that was discovered during this PhD thesis might reflect general functional traits of heterotrophic microorganisms in sponges. This might especially be the case for HMA sponges to which *Poribacteria* are almost exclusive (Hochmuth et al. 2010) and where they are one of the dominant phyla in the microbial communities (Schmitt et al. 2012). It is conceivable that, with the majority of microbial symbionts in marine sponges assumed to be heterotrophs (Taylor et al. 2007), and with microbial symbionts taking up as much as 35% of the sponge's biomass (Vacelet, 1975), the functional repertoire of these organisms largely influences the metabolic processes and therefore also nutrient cycling in marine sponges. If the microbial symbionts of marine sponges indeed have an influence on nutrient cycling in the host, then it might be possible to see a difference in nutrient cycling of HMA and LMA sponges.

Sponges feed on both particulate and dissolved organic matter (POM and DOM, respectively) (Pile et al. 1996; Ribes et al. 1999; Yahel, 2003; de Goeij et al. 2008; Maldonado et al. 2012). Sponges feeding on bulk DOM was discovered in the marine sponge *Theonella swinheoei* (Yahel, 2003), which covers the majority of its carbon needs with dissolved organic carbon (DOC). Later De Goeij et al. (2008) showed that

other sponge species also feed on DOC and rely to a similar extent on it as their major carbon source. Feeding on DOC appeared to be the missing link between an often reported imbalance between particulate organic carbon (POC) ingestion and respiration rates where the respiration rate was often higher than could be accounted for by POC ingestion (Reiswig, 1974; Reiswig, 1981; Thomassen and Riisgard, 1995; Hadas et al. 2009). For some sponge species this imbalance appeared to be larger than for others. In a recent review Maldonado et al. (2012) compared all metabolic measurement data of sponges available in literature at the time in order to account for differences in the carbon balance of LMA and HMA sponges. This comparison revealed that imbalances between POC ingestion and respiration rate, as well as DOC uptake levels, have been observed in representative species of both LMA and HMA sponges. However, the authors also found the imbalanced carbon budget in HMA sponges to be larger than in LMA sponges and that HMA sponges may rely more strongly on DOC uptake (Maldonado et al. 2012). Although this literature review hinted towards a difference in the carbon flows of LMA and HMA sponges it was restricted in the sense that the microbial abundance and the status of LMA or HMA sponge was often not resolved for the sponge species in the reviewed literature.

Ribes et al. (2012) directly compared uptake and release rates of DOC, dissolved organic nitrogen (DON), as well as inorganic nitrogen (NH_4^+, NO_3^-, NO_2), and phosphate from the two HMA sponges *Agelas oroides* and *Chondrosia reniformis*, and the LMA sponge *Dysidea avara* that occur simultaneously in the same habitat. The measurements in this study confirmed the difference between LMA and HMA sponges in terms of nutrient cycling. Both investigated HMA sponges showed strong and significant ingestion of DOC and ammonium while the LMA sponge *D. avara* did not show any significant uptake of any of the measured substances (Ribes et al. 2012). The results of this study support the theory that microbes inhabiting HMA sponges might be involved in the uptake of organic nutrients. Another study showed the ingestion of ^{13}C labelled DOC by the marine sponge *Halisarca caerule* directly and mediated through bacterial symbionts as shown by the detection of bacteria-specific fatty acids in the host that contained the labelled carbon (de Goeij et al. 2008). This is direct evidence that DOC is not only ingested by the sponge itself but also its bacterial symbionts. The poribacterial genome sequences encode for the capacity to carry out relevant processes for the uptake and processing of DOC by sponge symbionts. Thus it seems highly likely that *Poribacteria* indeed have an influence at least on carbon cycling in

marine sponges even though direct evidence for the *in situ* reaction occurring in *Poribacteria* remains a subject for future studies.

The importance to further investigate processes in sponges that are mediated by heterotrophic microorganisms, such as *Poribacteria*, becomes apparent when considering the large ecological implications that are connected to these processes. Nutrient cycling in marine sponges is one of the most important topics in sponge ecology. Several studies have shown that sponges influence diverse nutrient cycles through their filter feeding activities and the relevance for the bentho-pelagic coupling (Gili and Coma, 1998; Ribes et al. 1999; reviewed by Maldonado et al. 2012). Recent research further stressed the importance of sponge nutrient cycling for tropical reef ecosystems. De Goeij and colleagues (2013) showed that sponges are mainly responsible for taking up DOM and making it accessible to higher trophical levels by reintroducing a substantial proportion back into the ecosystem as detritus. This process was termed the sponge loop and is supposed to be a major driver of the DOM cycling which is the main source of energy and nutrients in tropical reef systems (De Goeij et al. 2013). These findings are of great significance since they explain how in the oligotrophic environment of tropical reef nutrients get reintroduced into higher trophical levels. In pelagic environments this is mediated by the microbial loop in which heterotrophic bacteria incorporate DOM and therefore make it again accessible for higher trophical levels (Azam et al. 1983; Azam et al. 1994). The microbial loop alone could, however, not account for DOM recycling in tropical reefs (de Goeij and van Duyl, 2007) and only with the discovery of the sponge loop can DOM recycling be explained for these environments (De Goeij et al. 2013). It was suggested that the sponge loop exists in temperate or cold-water environments as well (De Goeij et al. 2013) and therefore the influence of sponges on nutrient cycles on a global scale could be even larger. Considering the metabolic interactions mediated by microbial symbionts in marine sponges that were discussed above it is conceivable that the sponge loop is at least partially a microbial loop within the sponge host. The high abundance of microorganisms in the sponge and the high concentration of ingested DOM might make the nutrient recycling process more effective and could explain why, in the tropical reef investigated by De Goeij et al. (2013), the sponge loop accounted for a much larger proportion of carbon transfer from DOM than the microbial loop.

6.3. Future directions

6.3.1. Current potential and future directions for single-cell genomics in sponge microbiology

In chapter two of this thesis, single-cell genomics was introduced and some studies using this technique in context to symbiosis were described. In this section I will discuss the potential for the application of single-cell genomics in sponge microbiology and possible variations for future studies.

Despite many efforts the majority of the sponge microbiota is still not available in culture today. This explains why only a handful of genomes from sponge symbionts are available in the current databases and finding suitable reference genomes is a challenge for community-wide studies. Some studies successfully obtained genome sequences from metagenomic data. Binning of metagenome data from the microbial community of *C. concentrica* led to the recovery of partial genome sequences from five *Proteobacteria* (*Bdellovibrionales*, *Phyllobacteriaceae*, *Sphingomonadales, Piscirickettsiaceae*, and *Gamma-proteo I*) (Thomas et al. 2010). However, only little functional information was gained from these draft genomes and the majority of the data obtained in this study was not functionally assigned (Thomas et al. 2010). The draft genome of a deltaproteobacterial symbiont of the marine sponge *C. concentrica* was also obtained by this binning approach and thoroughly analysed (Liu et al. 2010). While this approach provided quality information about a sponge-associated microorganism the authors were restricted in the choice of organism to work with by the data becoming available through the binning process. Thus it is arguable how important the characterised organism really is for the symbiosis or if it, in the worst-case scenario, might represent a seawater-derived organism that is not involved in the symbiosis. Hallam et al. (2006; 2006) successfully obtained a composite genome of *Crenarachaeum symbiosum*, the main archaeal symbiont of the sponge *A. mexicana*, from a fosmid library. The success of these studies was based on the fact the *C. symbiosum* is the dominant organism in this host. It makes up to 65% of the sponge's microbial biomass based on FISH analysis (Preston et al. 1996). Therefore the construction of a *C. symbiosum* enriched fosmid library was possible (Schleper et al. 1997; Schleper et al. 1998; Hallam, Mincer, et al. 2006; Hallam, Konstantinidis, et al. 2006). In most sponge species the microbial diversity is, however, much higher (Taylor et al. 2007; e.g. Webster et al. 2010; Schmitt et al. 2012) and dominance of certain

organisms can only be found in LMA sponges (Giles et al. 2013). Thus obtaining enough genomic information from one organism from a metagenome of an HMA sponge would be much more difficult and excludes this option for most sponge symbionts.

The single-cell genomic approach used in this PhD thesis provided the possibility to obtain genomic information of previously defined target organisms even from sponges with high microbial diversity. The candidate phylum *Poribacteria* as a target organism ensured obtaining genomic information of a widespread symbiont of HMA sponges with high microbial diversity. Single-cell genomics would also allow for the selection of organisms that are important but rare members of the microbial community. Such organisms would likely be overlooked in the large datasets obtained by metagenomics and therefore single-cell derived data would be a useful addition to metagenomics. Generally single-cell genomics could provide the genome sequences that are much needed as a reference for community-wide studies. The majority of the genomic data available today represents only limited amount of phyla from a limited amount of environments (Wu et al. 2009; Rinke et al. 2013). These datasets are therefore not necessarily the best references for annotation of genomic data from other environments such as the sponge microbiome where the microbial community is very diverse, yet specific for its environment and also contains several previously undescribed candidate phyla (Schmitt et al. 2012).

A recent study by Rinke et al. (2013) showed how powerful single-cell genomics can be in obtaining genomic information of candidate phyla from diverse environmental samples. This study significantly broadened our knowledge about bacterial and archaeal phylogeny and metabolism by defining novel phyla and the taxonomic placement as well as revealing novel functional features such as archaeal sigma factors or new assignments of stop codons in bacterial translation. Furthermore the study showed that the genomic information of candidate phyla significantly improves phylogenetic assignments of metagenome reads (Rinke et al. 2013). Similarly successful was the application of single-cell genomics for the candidate phylum *Poribacteria* in this thesis. Not only was it possible to refine the phylogenetic placement of *Poribacteria* (Chapter 5) but also showed the analysis of the metabolic potential for the first time with detailed insights into the functions of heterotrophic microbes in marine sponges (Chapter 4). Obtaining more reference genomes in future studies would further

broaden our knowledge of the sponge microbiome and might therefore reveal novel functional features essential for this symbiosis. Next to the many candidate phyla that were found in marine sponges, members of the so-called core community i.e. *Proteobacteria* and *Chloroflexi* (Schmitt et al. 2012) could also be interesting target organisms for genome sequencing.

With further development and decreasing cost of genome sequencing projects, large scale sequencing of several phylogenetically diverse single-cell genomes from the sponge microbe could provide a variety of reference genomes and at the same time build a strong genomic reference library for future studies. This approach was successfully conducted for microbial communities from the mesopelagic oceans (Swan et al. 2011) and in combination with metagenomics for freshwater bacterioplankton (Martinez-Garcia et al. 2012). In both studies, single-cell genomics revealed the important players for chemotrophic carbon fixation in the dark ocean and for photo-heterotrophy as well as chemoautotrophy in freshwater lakes, respectively. Applying such large scale sequencing strategies to sponge microbiota could help to obtain phylogenetic information about the functional features discovered in community-wide studies of metagenomics, metatranscriptomics and metaproteomics (e.g. Thomas et al. 2010; Radax et al. 2012; Fan et al. 2012; Liu et al. 2012) as well as refine the functional information obtained in these studies.

The functional profile of the microbial community of *C. concentrica* was studied using metagenomics and metaproteomics (Thomas et al. 2010; Liu et al. 2012). These studies were mainly based on differential abundance of clusters of orthologous genes (COGs) between the sponge-derived and seawater-derived metagenomic data and between sponge-derived microbial metagenome and metaproteome data. This functional approach based on COGs works well finding differences between compared datasets but could benefit from a more in-depth analysis of the pathways and genes behind the broadly defined COGs. The products of many functional genes are often involved in several different reactions and pathways and the contigs of metagenome studies alone do not provide enough information and genomic context to identify the true function behind the detected genes, as they are often fragmented or represent composite assemblies and not genomic information from the same organism (Fig. 1).

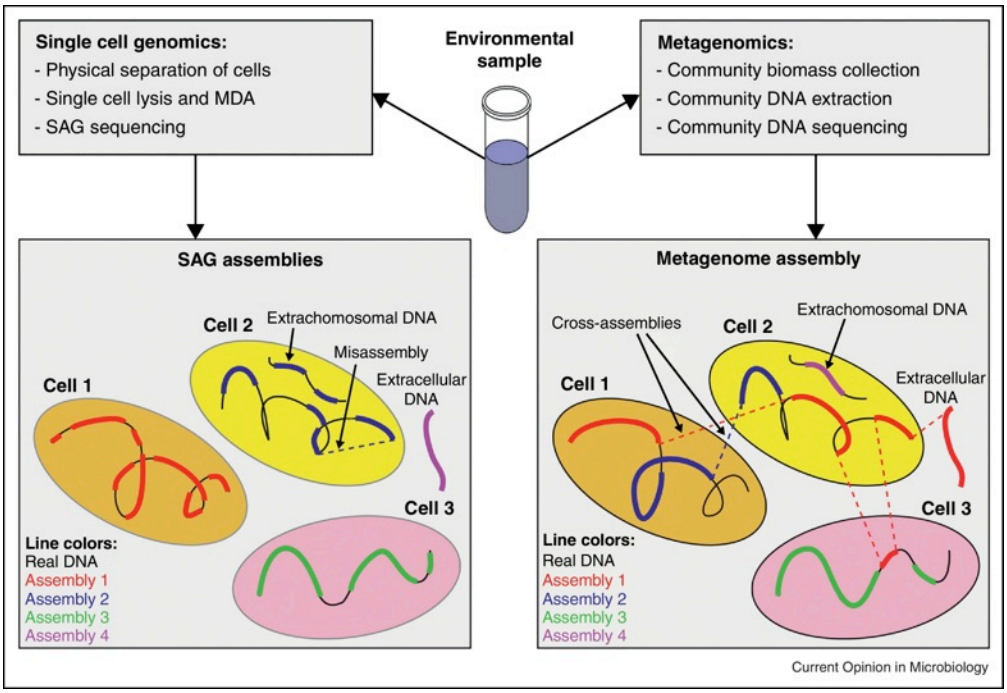


Figure 1: Comparison of single-cell genomics (SCG) and metagenomics assemblies. Dashed lines indicate false assembly contiguity. While in SCG assemblies all resulting contigs originate from DNA that was present in an individual cell, metagenomic assemblies are often consensuses from a multitude of cells. Picture used with permission from Stepanauskas (2012).

A comparison with genome data from abundant organisms in the environment would provide additional detail to the analysis. Knowing the full genomic context is essential for accurate prediction of genes and pathways as a functioning and active enzyme complex or signalling pathway is only formed by gene products that actually are in contact with each other within the same organism (Blainey, 2013). Metagenomics lacks the detailed genomics resolution and therefore would benefit from additional single-cell genomics data.

Single-cell genomics of sponge symbionts could also complement metagenomic studies by adding a better connection between phylogeny and function. Fan et al. (2012) successfully identified shared functional features between metagenomes of the microbial communities of six sponge species. However, they also found that the functional profiles were to some degree host-specific. The possibility of connecting phylogeny and function would add further value to this analysis as it would allow identification of the taxa responsible for the shared features among the microbiota of different sponge species and therefore reveal taxa with special importance for the relationship to the sponge. However, linking phylogeny and function can be difficult in

metagenome studies of the sponge microbiome because of the many uncultivated members of these communities and very little relevant genome sequences available in the current databases. Radax et al. (2012) restrained from tracing the phylogenetic origin of the functional metatranscriptome data of *G*. *barretti* completely, since the genome sequences needed for the correct assignment of the obtained mRNA data were not available and simulation analysis of taxonomic assignments resulted in either false or no phylogenetic assignment for most sequences. These studies show that community-wide approaches of sponge microbiota could well benefit from the availability of genome sequences from representatives of abundant sponge symbionts. In order to fully understand the complex systems of sponge-microbe associations it is not only important to detect functional features of the microbial communities but also identify the individual players and how they influence each other. Here single-cell genomics provides a tool to achieve this level of detail.

The future potential for single-cell genomics in sponge microbiology goes beyond supplying further reference genomes for community-wide studies. Single-cell genomics provides access to all DNA molecules within the analysed cell which can, next to the main chromosome(s), also include extra chromosomal DNA such as plasmids, DNA of symbionts or pathogens, and DNA from food sources or organelles (Stepanauskas, 2012) (Fig. 1). Investigating such molecules allows an insight into genomic interactions of the target organism. A recent study by Yoon et al. (2011) showed, by using single-cell genomics, the interactions of marine picobiliphytes with bacteria, viruses, and phages. In an environment with many diverse players, such as in marine sponges, this approach could gain useful information about inter-species relationships and give genomic information beyond microbial interactions with the primary sponge host. Eukaryotic microbes are abundant in sponges. Dinoflagellates, diatoms, or other microalgae as well as fungi have been reported (reviewed by Taylor et al. 2007). These organisms can serve as a food source for the sponges but also symbiotic interactions have been reported (Taylor et al. 2007; Weisz et al. 2010). Such organisms could be potential targets for future single-cell genomic studies to reveal interactions between the different members of the sponge microbiome in more detail.

Of great interest in sponge microbiology is the influence of viruses and phages. Viruses and virus-like particles have been reported from marine sponges (Vacelet and Gallissian, 1978; Hadas et al. 2006; Claverie et al. 2009) and recently molecular cloning

studies showed phage transmitted horizontal gene transfer in the microbial community of the marine sponge *Hymeniacidon perlevis* (Harrington et al. 2012). Clustered regularly interspaced short palindromic repeats (CRISPRs) and CRISPR-associated proteins (CAS) represent a prokaryotic adaptive immune system against integration of extra-chromosomal DNA. These genomic elements have been found in high abundance and overrepresented in sponge metagenomes compared to seawater metagenomic data (Thomas et al. 2010; Fan et al. 2012). This stressed the potential influence of phages on the microbial community of sponges. Yet little is known about viruses and phages in sponges and their influence on the microbial community. Webster et al. (2011) identified this subject as “of fundamental importance to understanding the ecology and evolution of sponge symbioses” and in need of “increased research efforts”. Single-cell genomics can help to reveal viral interactions in the microbial communities of marine sponges. Individual viral particles were sequenced using a method called single-viral genomics where viral particles from a mixed assemblage were singularised by FACS and amplified using phi29 transmitted MDA (Allen et al. 2011). The application of this method to samples with high microbial abundance and diversity, such as the sponge microbiome, will likely require much optimisation and is still restricted to DNA viruses but can be a useful method to obtain genomic information of viruses and phages. Sequencing single bacterial or eukaryotic microbes can also give access to viral and phage genomic information, when found in the host cell and at the same time provides information about the host itself. Targeted sorting of infected cells might be possible by applying specific cell stains and subsequent single-cell sorting as it was demonstrated for microalgae (Martínez Martínez et al. 2011). Further developments of this and other methods such as single-cell transcriptomics (Tang et al. 2011; Kang et al. 2011) and metabolomics (Heinemann and Zenobi, 2011; Yanes, 2013) that are currently being established will enlarge the possibilities for analysis of microbial communities in marine sponges and elsewhere.

6.3.2. Further investigations of the candidate phylum *Poribacteria*

The investigation of the candidate phylum *Poribacteria* has come a long way from their first discovery and studies describing abundance and diversity patterns (Fieseler et al. 2004; Schmitt et al. 2012). This represents the first step in the cultivation-independent repertoire of methods used to answer the central questions in microbial ecology (Fig. 2). Genomic characterisation of *Poribacteria* by metagenomics (Fieseler et al. 2004) and

single-cell genomics (Siegl and Hentschel, 2009; this PhD thesis) represents the second step in this process. Investigating the genomic repertoire of the candidate phylum *Poribacteria* has strongly improved our understanding of the taxonomy, inter-phylum diversity and metabolic potential of these sponge symbionts. The datasets obtained in this PhD thesis will serve as an important basis for future investigations of poribacterial metabolic activity and *in situ* functioning (Fig. 2). *Poribacteria* were shown to be highly abundant and active symbionts of marine sponges based on 16S rRNA gene and 16S rRNA data, respectively (Fieseler et al. 2004; Kamke et al. 2010; Schmitt et al. 2012; Moitinho-Silva et al. 2013). The newly obtained genome data of *Poribacteria* makes it now possible to investigate poribacterial gene and transcript abundance in metagenome and metatranscriptome data. Mapping metagenomic reads, termed fragment recruitment (Rusch et al. 2007), from different sponge species and non-sponge sources to poribacterial genomes could reveal the degree of the genomic representation of *Poribacteria* in these environments beyond the 16S rRNA gene level. This could help to solve questions about poribacterial distribution in and outside the sponge and might show distribution patterns of the different poribacterial phylotypes in different sponge species.

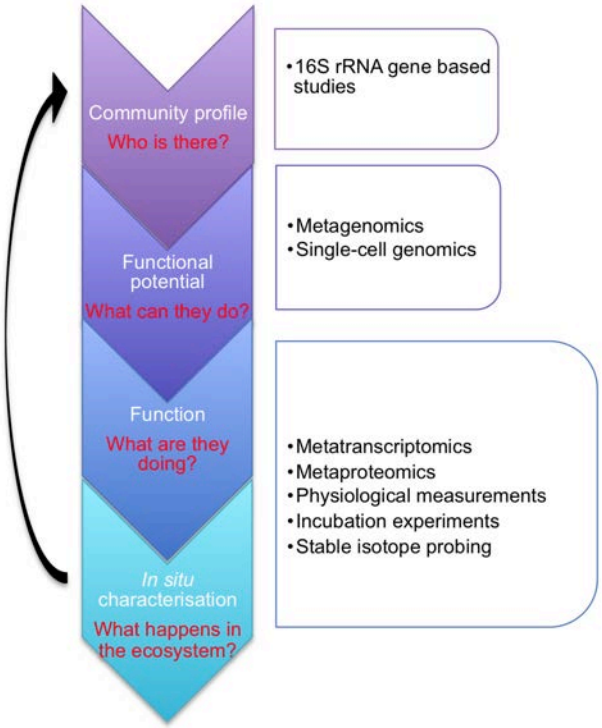


Figure 2: Schematic overview of the central questions in microbial ecology research and some of the methods used to answer them. The order and combination of approached is independent on the individual study and often interdependent.

Fragment recruitment was applied successfully before to investigate the genomic distribution of SAGs in the pelagic and coastal areas (Woyke et al. 2009; Swan et al. 2013). Through this approach the importance of the organisms represented by the obtained SAGs was shown and correlations to abiotic factors such as temperature and latitude were revealed (Woyke et al. 2009; Swan et al. 2013). In addition to metagenomic fragments recruitment, this approach could be used in combination with metatranscriptomic datasets to show metabolic activity. For *Poribacteria* this might reveal which functions are truly important for living within the host matrix and might indicate whether feeding on matrix substances indeed takes place. The availability of suitable metagenome and metatranscriptome datasets is essential for the success of fragment recruitment analyses. So far these dataset are limited to a few species (Thomas et al. 2010; Radax et al. 2012; Fan et al. 2012; Liu et al. 2012; Fan et al. 2013) but with further development and decreasing cost of sequencing techniques new datasets could soon become available.

Sequencing of poribacterial SAGs showed the existence of a possible dominant taxonomic unit that was defined as group I. Four out of six SAGs were associated to this group. Sequence similarity within the group was high and within the 97% 16S rRNA gene similarity threshold for definition of an OTU. Especially SAG 4CII and 3G showed very similar characteristics with 99.9% sequence id based on the 16S rRNA gene and tetra nucleotide frequency of 0.99. Genome comparison between these two SAGs showed a large degree of genome synteny (e.g. Chapter 5 microbial microcompartments). It would be interesting to further investigate genomic structures between members of group I and especially SAG 3G and 4CII. Comparison of genome structure might reveal important aspects of poribacterial population structure. The degree of polymorphism as well as the insertion and deletion of genes might give information about genome evolution within this symbiont. For pathogenic bacteria population structures can reveal important information about virulence factors and evolution patterns towards pathogenicity (Achtman, 2004; Wilson, 2012). Such analyses might therefore reveal similar features for symbiosis factors in *Poribacteria*.

The single-cell genomics data of *Poribacteria* might also be helpful for the development of cultivation methods for these and other heterotrophic sponge symbionts. Current culturing methods for sponge-associated microbiota have been shown to be insufficient to isolate any relevant proportion of the microbial communities detected with culture-

independent techniques. This is especially the case in HMA sponges where none of the taxa detected with 16S rRNA amplicons sequencing are represented in culture (Schippers et al. 2012). Even though culture-independent methods for characterisation of bacteria are constantly improved and further developed, it is still highly desirable to obtain organisms in culture as it would allow to study physiological processes in those organisms and reactions to different culture conditions (Schippers et al. 2012). SAG analysis indicated that *Poribacteria* might grow on a variety of carbon substrates. This could be tested by growth experiments on complex media that contain carbon sources specific to the poribacterial glycobiome revealed in this thesis (Chapter 4) or using minimal media and adding substrates separately to single out important nutrients. However, precise predictions for poribacterial growth media are not completely possible because of the large percentage of genes and especially glycoside hydrolase genes that could not functionally be characterised. These might have large influence on poribacterial growth conditions. Metatranscriptomic data of sponges hosting *Poribacteria* might be useful to identify genes actively transcribed by *Poribacteria* and might therefore narrow down potential growth substrates. A similar approach was used by Bomar and colleagues (2011) who identified mucin as the main nutrient for the dominant bacterial symbiont of the medical leech *Hirudo verbena* by metatranscriptome analysis. Adding mucin to a culture medium enabled the authors to grow this symbiont for the first time. A similar approach might also be possible for *Poribacteria* even though the sponge microbiome is much more diverse than the microbial community of *H. verbena*. The poribacterial SAG sequences would allow binning out relevant metatranscriptomic reads that could then be used to identify potential growth substrates.

Finally the metabolic potential that was encoded on poribacterial genomes could be used as a starting point for functional studies to validate the hypotheses of poribacterial carbohydrate degradation *in situ*. Feeding experiments with different carbohydrate sources and HMA sponges with high abundance of *Poribacteria* might reveal which carbon substrates are really taken up in sponges containing *Poribacteria*. This could be compared to LMA sponges without a relevant abundance of *Poribacteria* or LMA species where metabolic inactivity of *Poribacteria* was shown (Moitinho-Silva et al. 2013). Testing different carbohydrates as substrates might reveal whether there is a preference for specific nutrient sources and whether polymers are in fact degraded even in the presence of di-or monosaccharides. Such experiments could be conducted

under controlled conditions in aquaria or even in field experiments. Such experiments were already successfully applied to measure uptake and ingestion of DOC and other carbon substrates in sponges (e.g. Yahel et al. 2003; de Goeij et al. 2008; De Goeij et al. 2008) and could be modified with substrates likely to digested by *Poribacteria* based on their genomic profile. Furthermore stable isotope probing (SIP) techniques such as DNA and RNA-SIP would allow for detection of those bacteria that metabolised a labelled substrate based on the occurrence the labelled isotope in their DNA or RNA, respectively (Whiteley et al. 2006; Friedrich, 2006). Analyses of the 16S rRNA gene or transcript would then allow for phylogenetic identification. This technique could prove the uptake of labelled substrates by *Poribacteria* or other symbionts. Furthermore, a time series of such an experiment might reveal the fate of these substrates after initial ingestion and therefore offer an insight into trophic interactions in the sponge. This could answer the question if *Poribacteria* are simple food scavengers or aid in the digestion of food particles. Similar experiments have been successful for other heterotrophic bacteria. The trophic interaction between starch degrading microorganisms of the human colon were revealed using RNA-SIP (Kovatcheva-Datchary et al. 2009) and the same technique was used to identify the active heterotrophic organisms in the pelagic redoxclines of the Baltic Sea (Berg et al. 2013). The application of these techniques to the sponge-microbe ecosystem would be the next step in answering the central questions of microbial ecology in sponges in general and specific for *Poribacteria* by following steps from microbial abundance and diversity (e.g. Fieseler et al. 2004; Schmitt et al. 2012) to genomic potential (Fieseler et al. 2006; Siegl and Hentschel, 2009), and finally revealing functional activity *in situ* (Fig.2).

6.4. References

Achtman, M. (2004) Population structure of pathogenic bacteria revisited. Int J Med Microbiol 294: 67–73.

Allen, L.Z., Ishoey, T., Novotny, M.A., McLean, J.S., Lasken, R.S., and Williamson, S.J. (2011) Single virus genomics: a new tool for virus discovery. PLoS ONE 6: e17722.

Azam, F., Fenchel, T., Field, J.G., Gray, J.S., Meyer-Reil, L.A., and Thingstad, F. (1983) The ecological role of water-column microbes in the sea. Mar Ecol Prog Ser 10: 257–263.

Azam, F., Smith, D.C., Steward, G.F., and Hagström, A. (1994) Bacteria-organic matter coupling and its significance for oceanic carbon cycling. Microb Ecol 28: 167–179.

Bensaadi-Merchermek, N., Salvado, J.C., Cagnon, C., Karama, S., and Mouchès, C. (1995) Characterization of the unlinked 16S rDNA and 23S-5S rRNA operon of *Wolbachia pipientis*, a prokaryotic parasite of insect gonads. Gene 165: 81–86.

Berg, C., Beckmann, S., Jost, G., Labrenz, M., and Jürgens, K. (2013) Acetate-utilizing bacteria at an oxic-anoxic interface in the Baltic Sea. FEMS Microbiol Ecol 85: 251–261.

Blainey, P.C. (2013) The future is now: single-cell genomics of *Bacteria* and *Archaea*. FEMS Microbiol Rev 37: 407–427.

Bomar, L., Maltz, M., Colston, S., and Graf, J. (2011) Directed culturing of microorganisms using metatranscriptomics. mBio 2: e00012–11.

Boyer, S.L., Flechtner, V.R., and Johansen, J.R. (2001) Is the 16S-23S rRNA internal transcribed spacer region a good tool for use in molecular systematics and population genetics? A case study in cyanobacteria. Mol Biol Evol 18: 1057–1069.

Claverie, J.-M., Grzela, R., Lartigue, A., Bernadac, A., Nitsche, S., Vacelet, J., et al. (2009) *Mimivirus* and *Mimiviridae*: Giant viruses with an increasing number of potential hosts, including corals and sponges. J Invertebrate Pathol 101: 172–180.

De Goeij, J.M., and van Duyl, F.C. (2007) Coral cavities are sinks of dissolved organic carbon (DOC). Limnol Oceanogr 52: 2608–2617.

De Goeij, J.M., Moodley, L., Houtekamer, M., Carballeira, N.M., and van Duyl, F.C. (2008) Tracing ^{13}C-enriched dissolved and particulate organic carbon in the bacteria-containing coral reef sponge *Halisarca caerulea*: Evidence for DOM-feeding. Limnol Oceanogr 53: 1376–1386.

De Goeij, J.M., van den Berg, H., van Oostveen, M.M., Epping, E., and Van Duyl, F.C. (2008) Major bulk dissolved organic carbon (DOC) removal by encrusting coral reef cavity sponges. Mar Ecol Prog Ser 357: 139–151.

De Goeij, J.M., van Oevelen, D., Vermeij, M.J.A., Osinga, R., Middelburg, J.J., de Goeij, A.F.P.M., and Admiraal, W. (2013) Surviving in a Marine Desert: The Sponge Loop Retains Resources Within Coral Reefs. Science 342: 108–110.

Fan, L., Liu, M., Simister, R., Webster, N.S., and Thomas, T. (2013) Marine microbial symbiosis heats up: the phylogenetic and functional response of a sponge holobiont to thermal stress. ISME J 7: 991–1002.

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 USA 109: E1878–87.

Fieseler, L., Hentschel, U., Grozdanov, L., Schirmer, A., Wen, G., Platzer, M., et al. (2007) Widespread occurrence and genomic context of unusually small polyketide synthase genes in microbial consortia associated with marine sponges. Appl Environ Microb 73: 2144–2155.

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

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

Friedrich, M.W. (2006) Stable-isotope probing of DNA: insights into the function of uncultivated microorganisms from isotopically labeled metagenomes. Curr Opion Biotech 17: 59–66.

Fuerst, J.A., Webb, R.I., Garson, M.J., Hardy, L., and Reiswig, H.M. (1998) Membrane-bounded nucleoids in microbial symbionts of marine sponges. FEMS Microbiol Lett 166: 29–34.

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

Gili, J.M., and Coma, R. (1998) Benthic suspension feeders: their paramount role in littoral marine food webs. Trends Ecol Evol 13: 316–321.

Glöckner, F.O., Kube, M., Bauer, M., Teeling, H., Lombardot, T., Ludwig, W., et al. (2003) Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. Proc Natl Acad Sci USA 100: 8298–8303.

Hadas, E., Marie, D., Shpigel, M., and Ilan, M. (2006) Virus predation by sponges is a new nutrient-flow pathway in coral reef food webs. Limnol Oceanogr 51: 1548–1550.

Hadas, E., Shpigel, M., and Ilan, M. (2009) Particulate organic matter as a food source for a coral reef sponge. J Exp Biol 212: 3643–3650.

Hallam, S.J., Konstantinidis, K.T., Putnam, N., Schleper, C., Watanabe, Y.-I., Sugahara, J., et al. (2006) Genomic analysis of the uncultivated marine crenarchaeote *Cenarchaeum symbiosum*. Proc Natl Acad Sci USA 103: 18296–18301.

Hallam, S.J., Mincer, T.J., Schleper, C., Preston, C.M., Roberts, K., Richardson, P.M., and DeLong, E.F. (2006) Pathways of carbon assimilation and ammonia oxidation suggested by environmental genomic analyses of marine crenarchaeota. Plos Biol 4: e95.

Harrington, C., Del Casale, A., Kennedy, J., Neve, H., Picton, B.E., Mooij, M.J., et al. (2012) Evidence of bacteriophage-mediated horizontal transfer of bacterial 16S rRNA genes in the viral metagenome of the marine sponge *Hymeniacidon perlevis*. Microbiology 158: 2789–2795.

Heinemann, M., and Zenobi, R. (2011) Single cell metabolomics. Curr Opion Biotech 22: 26–31.

Henne, A., Brüggemann, H., Raasch, C., Wiezer, A., Hartsch, T., Liesegang, H., et al. (2004) The genome sequence of the extreme thermophile *Thermus thermophilus*. Nat Biotechnol 22: 547–553.

Hochmuth, T., Niederkrüger, H., Gernert, C., Siegl, A., Taudien, S., Platzer, M., et al. (2010) Linking chemical and microbial diversity in marine sponges: Possible role for *Poribacteria* as producers of methyl-branched fatty acids. ChemBioChem 11: 2572–

2578.

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

Kang, Y., Norris, M.H., Zarzycki-Siek, J., Nierman, W.C., Donachie, S.P., and Hoang, T.T. (2011) Transcript amplification from single bacterium for transcriptome analysis. Genome Res 21: 925–935.

Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010) Bacterial Microcompartments. Annu Rev Microbiol 64: 391–408.

Klappenbach, J.A., Dunbar, J.M., and Schmidt, T.M. (2000) rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microb 66: 1328–1333.

Kovatcheva-Datchary, P., Egert, M., Maathuis, A., Rajilić-Stojanović, M., de Graaf, A.A., Smidt, H., et al. (2009) Linking phylogenetic identities of bacteria to starch fermentation in an in vitromodel of the large intestine by RNA-based stable isotope probing. Environ Microbiol 11: 914–926.

Lafi, F.F., Fuerst, J.A., Fieseler, L., Engels, C., Goh, W.W.L., and Hentschel, U. (2009) Widespread distribution of *Poribacteria* in *Demospongiae*. Appl Environ Microb 75: 5695–5699.

Liesack, W., and Stackebrandt, E. (1989) Evidence for unlinked rrn operons in the planctomycete *Pirellula marina*. J Bacteriol 171: 5025–5030.

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

Liu, M.Y., Kjelleberg, S., and Thomas, T. (2010) Functional genomic analysis of an uncultured delta-proteobacterium in the sponge *Cymbastela concentrica*. ISME J 5: 427–435.

Love, G.D., Grosjean, E., Stalvies, C., Fike, D.A., Grotzinger, J.P., Bradley, A.S., et al. (2008) Fossil steroids record the appearance of *Demospongiae* during the Cryogenian period. Nature 457: 718–721.

Maldonado, M., Ribes, M., and van Duyl, F.C. (2012) Nutrient fluxes through sponges: biology, budgets, and ecological implications. Adv Mar Biol 62: 113–182.

Martinez-Garcia, M., Swan, B.K., Poulton, N.J., Gomez, M.L., Masland, D., Sieracki, M.E., and Stepanauskas, R. (2012) High-throughput single-cell sequencing identifies photoheterotrophs and chemoautotrophs in freshwater bacterioplankton. ISME J 6: 113–123.

Martínez Martínez, J., Poulton, N.J., Stepanauskas, R., Sieracki, M.E., and Wilson, W.H. (2011) Targeted sorting of single virus-infected cells of the coccolithophore *Emiliania huxleyi*. PLoS ONE 6: e22520.

Moitinho-Silva, L., Bayer, K., Cannistraci, C.V., Giles, E.C., Ryu, T., Seridi, L., et al. (2013) Specificity and transcriptional activity of microbiota associated with low and high

microbial abundance sponges from the Red Sea. Mol Ecol In press.

Pile, A.J., Patterson, M.R., and Witman, J.D. (1996) *In situ* grazing on plankton. Mar Ecol Prog Ser 141: 95–102.

Preston, C.M., Wu, K.Y., Molinski, T.F., and DeLong, E.F. (1996) A psychrophilic crenarchaeon inhabits a marine sponge: *Cenarchaeum symbiosum* gen. nov., sp. nov. Proc Natl Acad Sci USA 93: 6241–6246.

Radax, R., Rattei, T., Lanzén, 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.

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

Reiswig, H.M. (1981) Partial carbon and energy budgets of the bacteriosponge *Verohgia fistularis* (*Porifera*: *Demospongiae*) in Barbados. Mar Ecol 2: 273–293.

Ribes, M., Coma, R., and Gili, J.M. (1999) Natural diet and grazing rate of the temperate sponge *Dysidea avara* (*Demospongiae*, *Dendroceratida*) throughout an annual cycle. Mar Ecol Prog Ser 176: 179–190.

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

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.

Ruepp, A., Graml, W., Santos-Martinez, M.L., Koretke, K.K., Volker, C., Mewes, H.W., et al. (2000) The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. Nature 407: 508–513.

Rurangirwa, F.R., Brayton, K.A., McGuire, T.C., Knowles, D.P., and Palmer, G.H. (2002) Conservation of the unique rickettsial rRNA gene arrangement in *Anaplasma*. Int J Syst Evol Microbiol 52: 1405–1409.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., et al. (2007) The Sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific. Plos Biol 5: e77.

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.

Schleper, C., DeLong, E.F., Preston, C.M., Feldman, R.A., Wu, K.-Y., and Swanson, R.V. (1998) Genomic analysis reveals chromosomal variation in natural populations of the uncultured psychrophilic archaeon *Cenarchaeum symbiosum*. J Bacteriol 87: 5003-5009.

Schleper, C., Swanson, R.V., Mathur, E.J., and DeLong, E.F. (1997) Characterization of

a DNA polymerase from the uncultivated psychrophilic archaeon *Cenarchaeum symbiosum*. J Bacteriol 179: 7803-7811.

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.

Siegl, A., and Hentschel, U. (2009) PKS and NRPS gene clusters from microbial symbiont cells of marine sponges by whole genome amplification. Environ Microbiol Rep 2: 507–513.

Stepanauskas, R. (2012) Single cell genomics: an individual look at microbes. Curr Opin Microbiol 15: 613–620.

Swan, B.K., Martinez-Garcia, M., Preston, C.M., Sczyrba, A., Woyke, T., Lamy, D., et al. (2011) Potential for chemolithoautotrophy among ubiquitous *Bacteria* lineages in the Dark Ocean. Science 333: 1296–1300.

Swan, B.K., Tupper, B., Sczyrba, A., Lauro, F.M., Martinez-Garcia, M., González, J.M., et al. (2013) Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. Proc Natl Acad Sci USA 110: 11463–11468.

Tamas, I., Klasson, L., Canbäck, B., Näslund, A.K., Eriksson, A.-S., Wernegreen, J.J., et al. (2002) 50 million years of genomic stasis in endosymbiotic bacteria. Science 296: 2376–2379.

Tang, F., Lao, K., and Surani, M.A. (2011) Development and applications of single-cell transcriptome analysis. Nature Meth 8: 6-11.

Taylor, M.W., Radax, R., Steger, D., and Wagner, M. (2007) Sponge-associated microorganisms: Evolution, ecology, and biotechnological potential. Microbiol Mol Biol R 71: 295–347.

Thomas, T., Rusch, D., DeMaere, M.Z., Yung, P.Y., Lewis, M., Halpern, A., et al. (2010) Functional genomic signatures of sponge bacteria reveal unique and shared features of symbiosis. ISME J 4: 1557–11567.

Thomassen, S., and Riisgard, H.U. (1995) Growth and energetics of the sponge *Halichondria panicea*. Mar Ecol Prog Ser 128: 239–246.

Vacelet, J. (1975) Etude en microscopie électronique de l'association entre bactéries et spongiaires du genre *Verongia* (*Dictyoceratida*). J Microsc Biol Cell 23: 271–288.

Vacelet, J., and Gallissian, M.-F. (1978) Virus-like particles in cells of the sponge *Verongia cavernicola* (*Demospongiae*, *Dictyoceratida*) and accompanying tissues changes. J Invertebrate Pathol 31: 246–254.

Wagner, M., and Horn, M. (2006) The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opion Biotech 17: 241–249.

Webster, N.S., and Taylor, M.W. (2011) Marine sponges and their microbial symbionts: love and other relationships. Environ Microbiol 14: 335–346.

Webster, N.S., Taylor, M.W., Behnam, F., Lücker, 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.

Weisz, J.B., Massaro, A.J., Ramsby, B.D., and Hill, M.S. (2010) Zooxanthellar symbionts shape host sponge trophic status through translocation of carbon. Biol Bull 219: 189–197.

Whiteley, A.S., Manefield, M., and Lueders, T. (2006) Unlocking the “microbial black box” using RNA-based stable isotope probing technologies. Curr Opion Biotech 17: 67-71.

Wilson, D.J. (2012) Insights from genomics into bacterial pathogen populations. PLoS pathog 8: e1002874.

Woyke, T., Xie, G., Copeland, A., González, J.M., Han, C., Kiss, H., et al. (2009) Assembling the marine metagenome, one Cell at a time. PLoS ONE 4: e5299.

Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., et al. (2009) A phylogeny-driven genomic encyclopaedia of *Bacteria* and *Archaea*. Nature 462: 1056–1060.

Yahel, G., Sharp, J.H., Marie, D., Häse, C., and Genin, A. (2003) *In situ* feeding and element removal in the symbiont-bearing sponge *Theonella swinhoei*: Bulk DOC is the major source for carbon. Limnol Oceanogr 48: 141–149.

Yanes, O. (2013) Metabolomics: Playing piñata with single cells. Nat Chem Biol 9: 471–473.

Yoon, H.S., Price, D.C., Stepanauskas, R., Rajah, V.D., Sieracki, M.E., Wilson, W.H., et al. (2011) Single-cell genomics reveals organismal interactions in uncultivated marine protists. Science 332: 714–717.

Appendix A

Author's contributions and permissions

Statement of individual author contributions:

Kamke J, Bayer K, Woyke T, Hentschel U (2012). Exploring Symbioses by Single-Cell Genomics. *The Biological Bulletin* **223:** 30-43.									
Participated in	Author Initials								
Manuscript Design	JK	UH							
Manuscript writing	JK	KB	TW	UH					

I confirm that I have obtained permission from the co-authors see signature below:

Bayer, K.	
Woyke, T.	
Hentschel, U.	

Statement of individual author contributions:

Siegl A, Kamke J, Hochmuth T, Piel J, Richter M, Liang C, Dandekar T, Hentschel U (2011). Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. *ISME J* **5:** 61-70.									
Participated in	Author Initials								
Study Design	AS	UH							
Data collection	AS								
Data analysis and interpretation	AS	JK	TH	JP	MR	CL	TD	UH	
Manuscript writing	AS	UH	JK						

I confirm that I have obtained permission from the co-authors see signature below:

Siegl, A.	
Hockmuth, T.	
Piel, J.	
Richter, M.	
Liang, C.	
Dandekar, T.	
Hentschel, U.	

Statement of individual author contributions:

Kamke J, Sczyrba A, Ivannoca N, Schwientek P, Rinke C, Mavromatis K, Woyke T, Hentschel U (2013). Single cell genomics analysis reveals complex carbon degradation patterns in poribacterial symbionts of marine sponges. The ISME Journal									
Participated in	Author Initials								
Study Design	JK	UH	TW						
Data collection	JK								
Data analysis and interpretation	JK	AS	NI	PS	CR	KM	TW	UH	
Manuscript writing	JK	UH	AS	TW	PS				

I confirm that I have obtained permission from the co-authors see signature below:

Sczyrba, A.	
Ivannova, N.	
Schwientek, P.	
Rinke, C.	
Mavromatis, K.	
Woyke, T.	
Hentschel, U.	

Statement of individual author contributions:

Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivannoca N,Sczyrba A, Woyke T, and Hentschel U (2013). Novel features of the candidate phylum Poribacteria revealed by single-cell genomic analysis Submitted to PlosOne									
Participated in	Author Initials								
Study Design	JK	UH	TW						
Data collection	JK								
Data analysis and interpretation	JK	CR	NI	PS	AS	KM	TW	UH	
Manuscript writing	JK	UH	CR						

I confirm that I have obtained permission from the co-authors see signature below:

Sczyrba, A.	
Ivannova, N.	
Schwientek, P.	
Rinke, C.	
Mavromatis, K.	
Woyke, T.	
Hentschel, U.	

List of publications

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

2. **Siegl A**, **Kamke J**, **Hochmuth T**, **Piel J**, **Richter M**, **Liang C**, **Dandekar T**, **Hentschel U**. 2011. Single-cell genomics reveals the lifestyle of Poribacteria, a candidate phylum symbiotically associated with marine sponges. The ISME Journal **5**:61–70.

3. **Angermeier H**, **Kamke J**, **Abdelmohsen UR**, **Krohne G**, **Pawlik JR**, **Lindquist NL**, **Hentschel U**. 2011. The pathology of sponge orange band disease affecting the Caribbean barrel sponge *Xestospongia muta*. FEMS Microbiology Ecology **75**:218–230.

4. **Kamke J**, **Bayer K**, **Woyke T**, **Hentschel U**. 2012. Exploring symbioses by single-cell genomics. The Biological Bulletin. **223**:30–43.

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

6. **Kamke J**, **Sczyrba A**, **Ivanova N**, **Schwientek P**, **Rinke C**, **Mavromatis K**, **Woyke T**, **Hentschel U**. 2013. Single-cell genomics reveals complex carbohydrate degradation patterns in poribacterial symbionts of marine sponges. The ISME Journal. DOI:10.1038/ismej.2013.111 Advance online publication ahead of print.

7. **Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, Woyke T, Hentschel U**. 2013. Novel features of the candidate phylum Poribacteria revealed by single-cell genomic analysis. Submitted to Plos One.