

## ORIGINAL ARTICLE

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

Alexander Siegl<sup>1,5</sup>, Janine Kamke<sup>1</sup>, Thomas Hochmuth<sup>2</sup>, Jörn Piel<sup>2</sup>, Michael Richter<sup>3</sup>, Chunguang Liang<sup>4</sup>, Thomas Dandekar<sup>4</sup> and Ute Hentschel<sup>1</sup>

<sup>1</sup>Department of Botany II, Julius-von-Sachs Institute for Biological Sciences, University of Wuerzburg, Wuerzburg, Germany; <sup>2</sup>Kekulé Institute of Organic Chemistry and Biochemistry, University of Bonn, Bonn, Germany; <sup>3</sup>Ribocon GmbH, Bremen, Germany and <sup>4</sup>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<sub>2</sub>-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 tetratricopeptide 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; Musmann *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

<sup>5</sup>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

**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
<i>All contigs (&gt; 200 bp)</i>	
Number of contigs	1597
Number of bases	1 884 861
<i>Large contigs (&gt; 500 bp)</i>	
Number of contigs	554
Number of bases	1 592 880
Average contig size	2875 bp
Largest contig size	46 669 bp
<i>Data set I (contigs with <math>\geq 1</math> partial ORF)</i>	
Number of contigs	455
Number of bases	1 503 489
<i>Data set II (contigs with neither start nor stop codon)</i>	
Number of contigs	1142
Number of bases	381 372

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

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.

## 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  $\mu$ l 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  $\geq 1100$  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

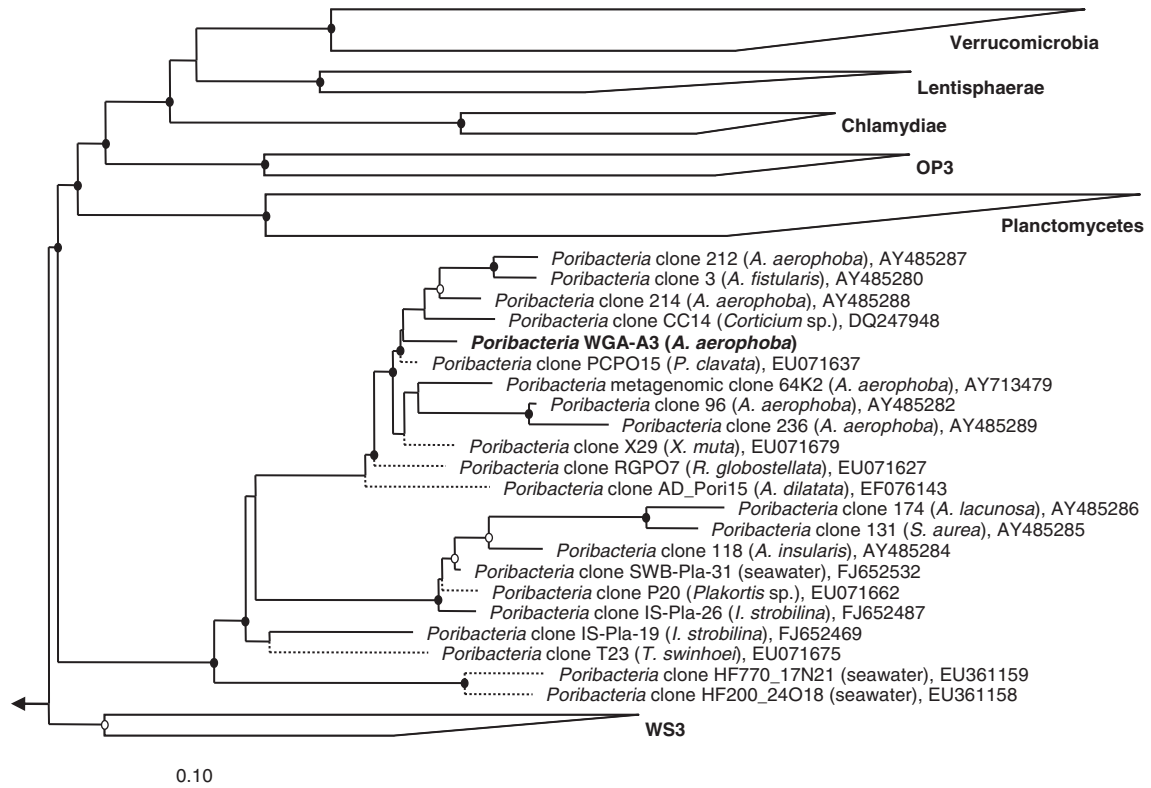
This Whole-Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession ADFK00000000. The version described in this paper is the first version, ADFK01000000.

## 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  $\mu$ g 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 (Stepanuskas and Sieracki, 2007). The JCoast genomesDB statistics tool (Richter *et al.*, 2008) showed that DNA contamination is  $< 0.1\%$  (1378 bp/1503489 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  $\geq 75\%$  is illustrated as open circles and bootstrap support  $\geq 90\%$  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).

#### 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

**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;  $n = 455$ ).

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

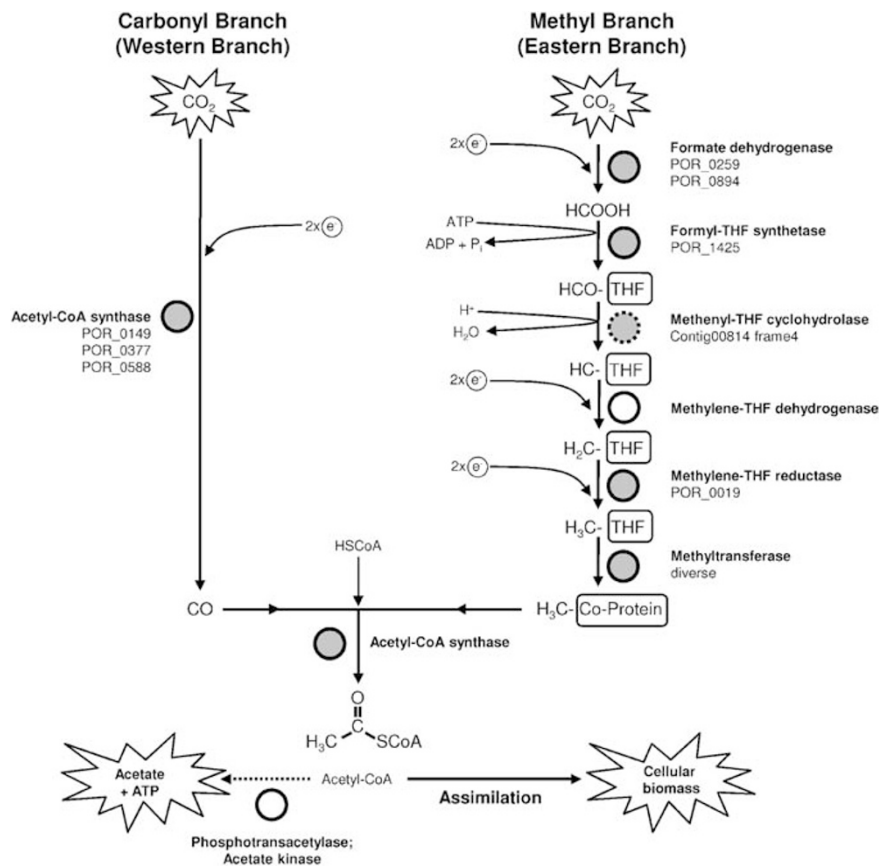
The *Poribacteria* contain the genomic capacity for aerobic, heterotrophic metabolism. Major parts of glycolysis, tricarboxylic acid cycle and the pentose

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



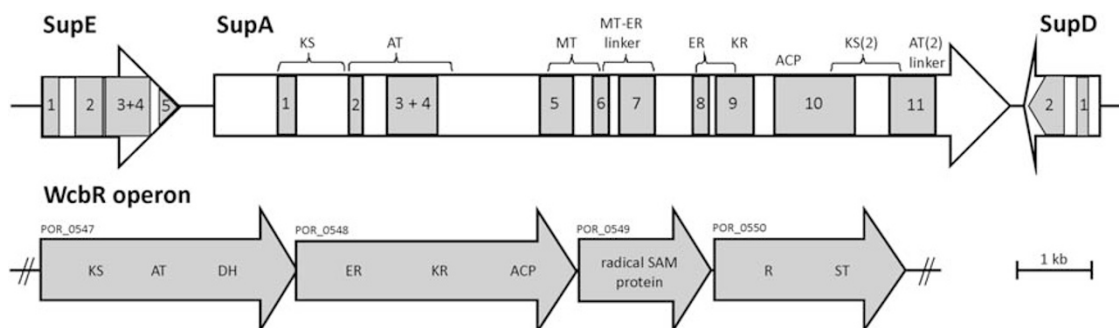
**Figure 2** Reconstruction of the Wood-Ljungdahl pathway for carbon fixation. Filled circles represent complete ORFs and dashed circles gene fragments. Open circles indicate enzymes absent from the data set. Pathway modeled after studies by Drake and Daniel (2004) and Ragsdale and Pierce (2008).

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 methyltransferasing (MT) properties.

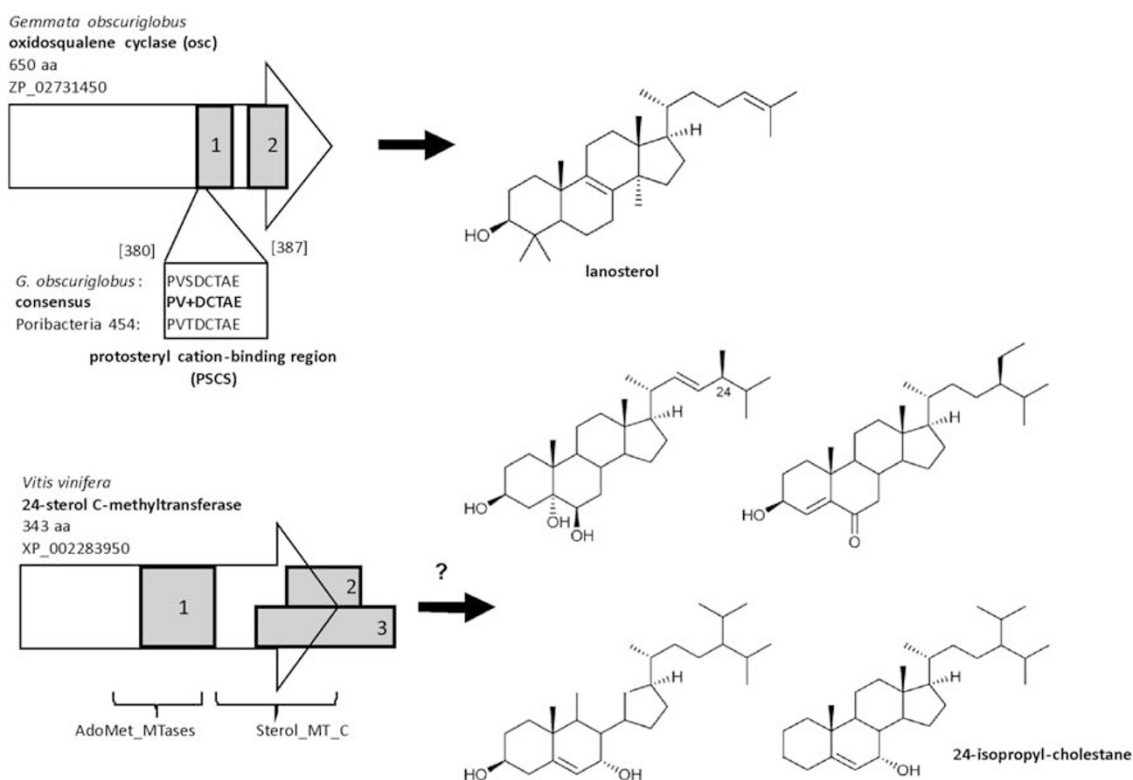
### Sterol biosynthesis

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 protosteroyl cation-binding region (PSCS) (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 protosteroyl 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.

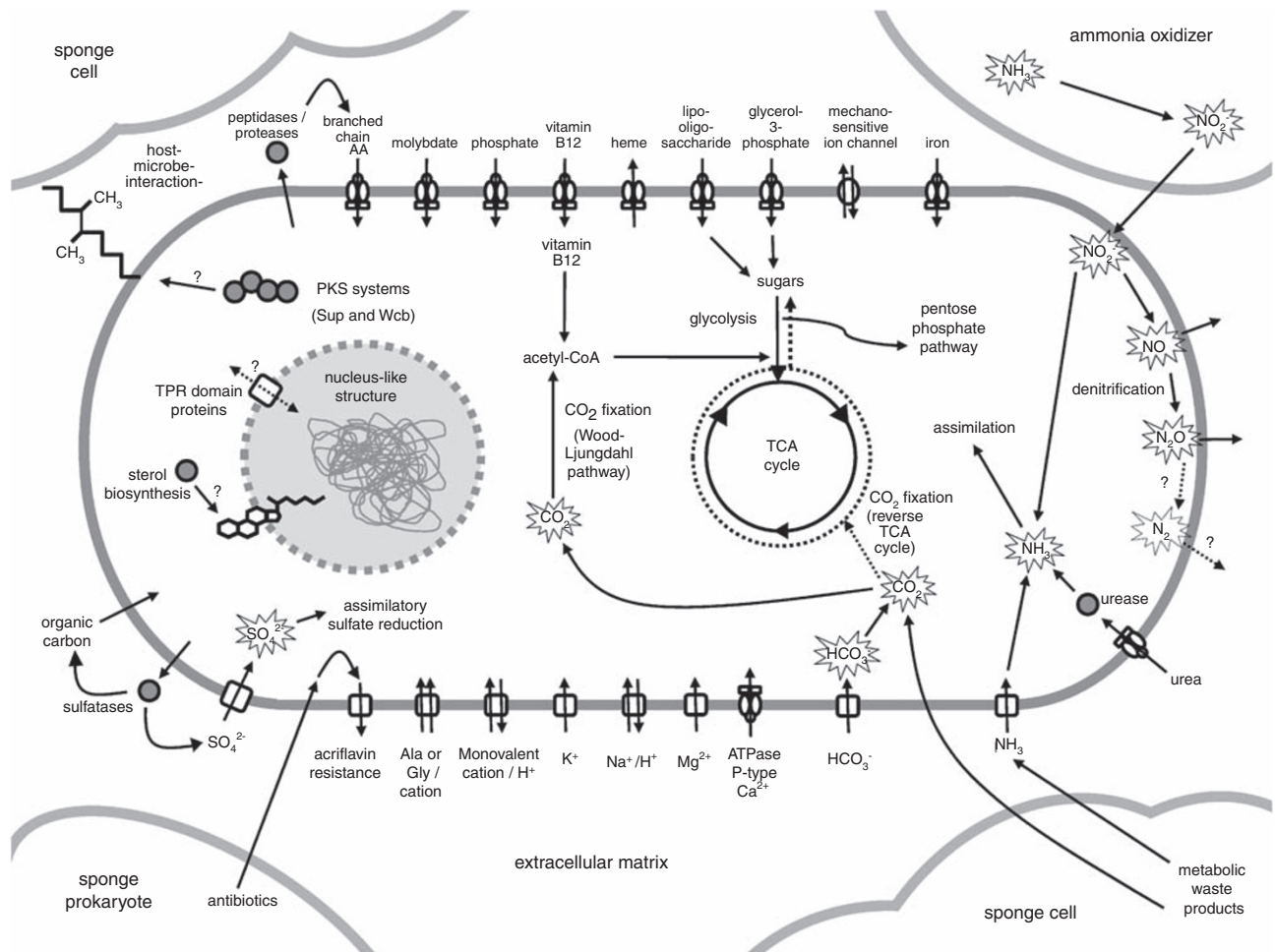
### Host-symbiont interaction

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 (c10020) 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<sub>2</sub> 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<sub>2</sub> 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, Lückner 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>)