The ISME Journal (2013) 7, 2287–2300

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 (2013) **7,** 2287–2300; doi:10.1038/ismej.2013.111; published online 11 July 2013
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; published online 11 July 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

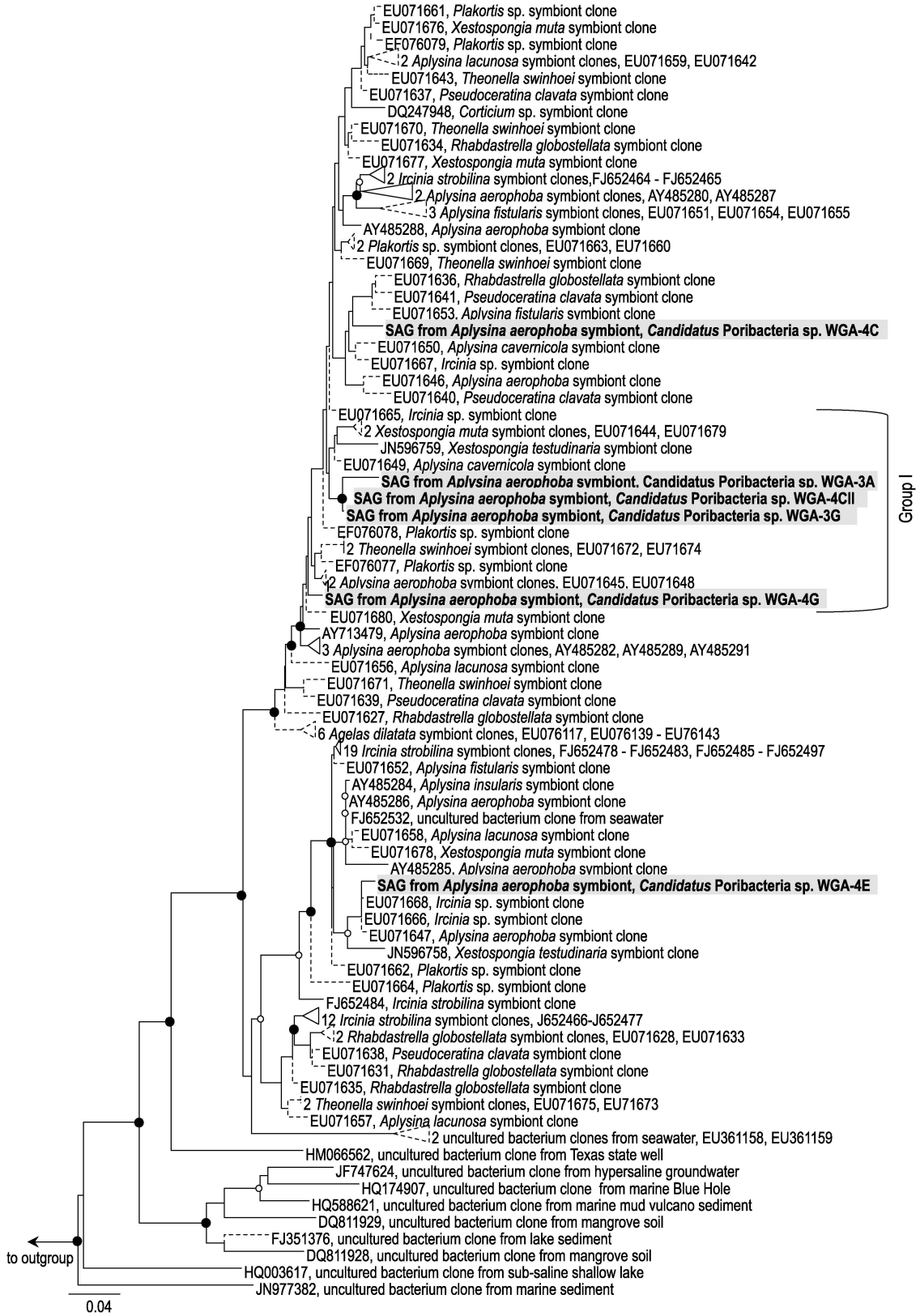

Figure 1 Phylogenetic maximum-likelihood tree based on 16S rRNA genes of the candidate phylum *Poribacteria*. Sequences from poribacterial SAGs are shown in bold and gray shading. The tree was constructed based on long sequences ($\geqslant$1100 nucleotides), shorter sequences were added without changing tree topology and are indicated by dashed lines. Bootstrap support (1000 resamplings) of $\geqslant$90% is shown by filled, and $\geqslant$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 phosphofructokinase (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

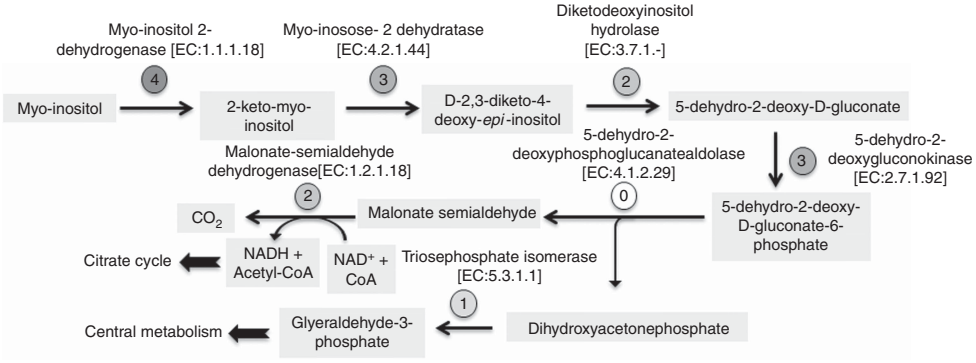

Figure 2 Schematic reconstruction of inositol degradation as encoded on poribacterial SAGs. Numbers within circles represent the number of genomes encoding for the corresponding enzyme.

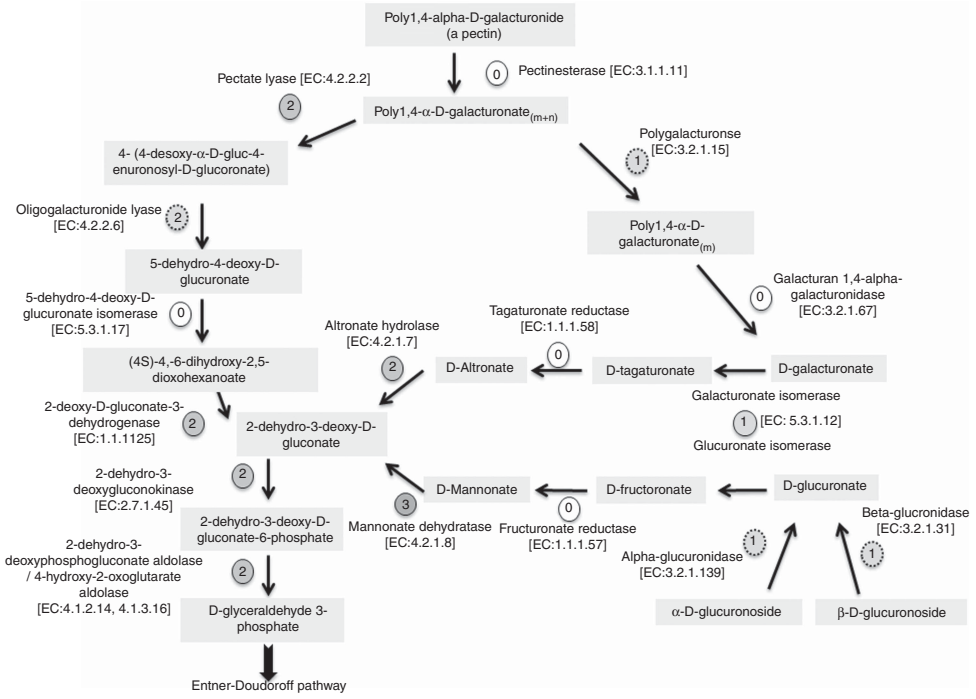

Figure 3 Schematic reconstruction of uronic acid degradation as encoded on poribacterial SAGs. Numbers within circles represent the genomes encoding for the corresponding enzyme. Dashed circles represent manually annotated genes.

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.

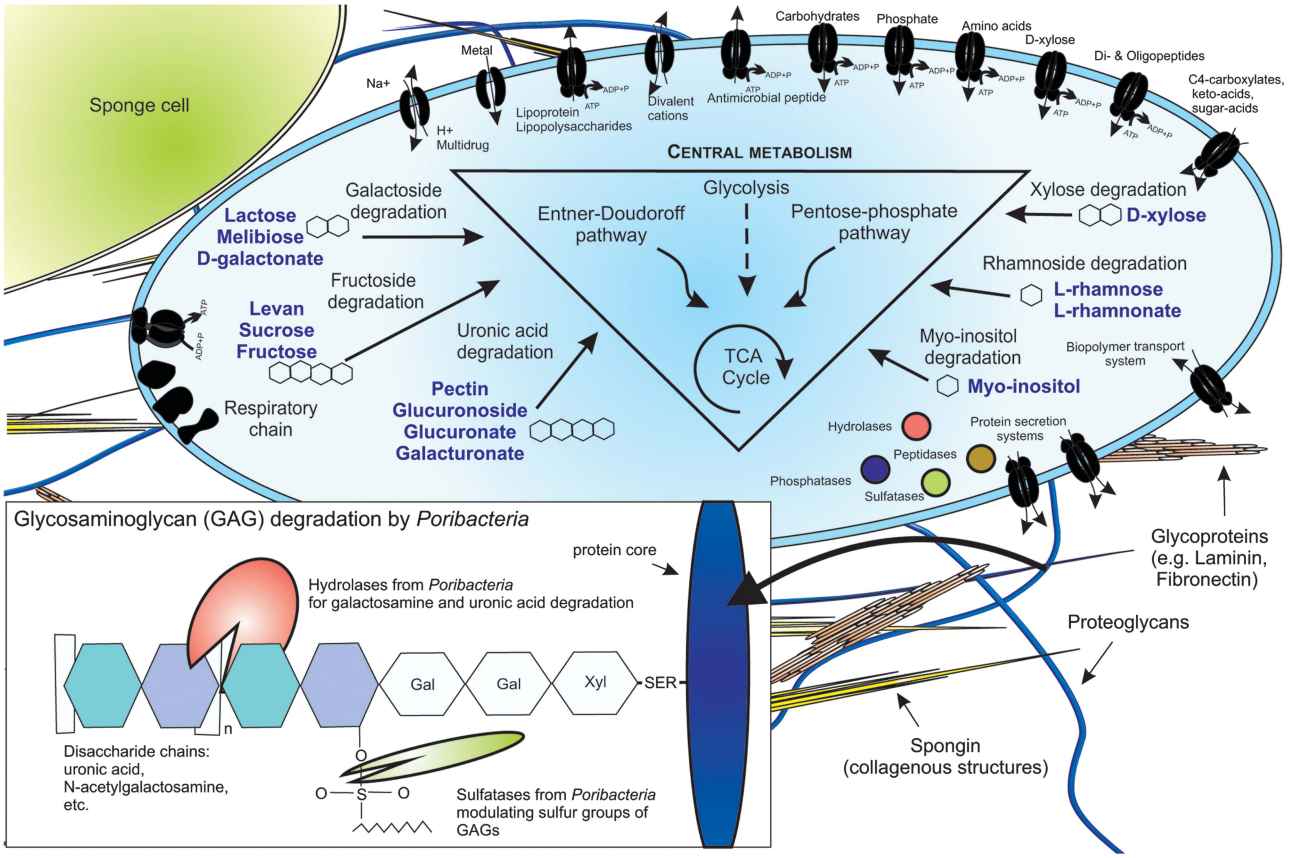

Figure 4 Schematic overview of a poribacterial cell in the sponge extracellular matrix illustrating pathways of carbohydrate metabolism and glycosaminoglycan degradation by poribacterial enzymes. The dashed arrow represents glycolysis that is not supported by the dominant poribacterial phylotype group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