The ISME Journal (2013) 7, 1026–1037

www.nature.com/ismej

ORIGINAL ARTICLE

Ecology of marine Bacteroidetes: a comparative genomics approach

Beatriz Fernández-Gómez[1], Michael Richter[2], Margarete Schüler[3], Jarone Pinhassi[4], Silvia G Acinas[1], José M González[5] and Carlos Pedrós-Alió[1]

[1]*Department of Marine Biology and Oceanography, Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Barcelona, Spain;* [2]*Microbial Genomics and Bioinformatics Research Group, Department of Molecular Ecology, Max Planck Institute for Marine Microbiology, Bremen, Germany;* [3]*Department of Molecular Structural Biology, Max Planck Institute for Biochemistry, Martinsried, Germany;* [4]*Centre for Ecology and Evolution in Microbial model Systems, Linnaeus University, Kalmar, Sweden and* [5]*Department of Microbiology, University of La Laguna, La Laguna, Spain*

Bacteroidetes are commonly assumed to be specialized in degrading high molecular weight (HMW) compounds and to have a preference for growth attached to particles, surfaces or algal cells. The first sequenced genomes of marine Bacteroidetes seemed to confirm this assumption. Many more genomes have been sequenced recently. Here, a comparative analysis of marine Bacteroidetes genomes revealed a life strategy different from those of other important phyla of marine bacterioplankton such as Cyanobacteria and Proteobacteria. Bacteroidetes have many adaptations to grow attached to particles, have the capacity to degrade polymers, including a large number of peptidases, glycoside hydrolases (GHs), glycosyl transferases, adhesion proteins, as well as the genes for gliding motility. Several of the polymer degradation genes are located in close association with genes for TonB-dependent receptors and transducers, suggesting an integrated regulation of adhesion and degradation of polymers. This confirmed the role of this abundant group of marine bacteria as degraders of particulate matter. Marine Bacteroidetes had a significantly larger number of proteases than GHs, while non-marine Bacteroidetes had equal numbers of both. Proteorhodopsin containing Bacteroidetes shared two characteristics: small genome size and a higher number of genes involved in CO_2 fixation per Mb. The latter may be important in order to survive when floating freely in the illuminated, but nutrient-poor, ocean surface.

The ISME Journal (2013) **7,** 1026–1037; doi:10.1038/ismej.2012.169; published online 10 January 2013
Subject Category: Integrated genomics and post-genomics approaches in microbial ecology
Keywords: glycoside hydrolase; polymer degradation; polymeric organic matter; protease; proteorhodopsin

Introduction

Members of the phylum Bacteroidetes are the most abundant group of bacteria in the ocean after Proteobacteria and Cyanobacteria (Glöckner *et al.*, 1999; Kirchman, 2002; Amaral-Zettler *et al.*, 2010). They account for a significant fraction of marine bacterioplankton especially in coastal areas, where they represent between 10% and 30% of the total bacterial counts (Alonso-Sáez and Gasol, 2007). They are globally distributed (Pommier *et al.*, 2007) in a variety of marine environments such as coastal, offshore, sediments and hydrothermal vents (Alonso *et al.*, 2007; Pommier *et al.*, 2007). In a recent analysis of metagenomes from the Global Ocean Sampling, Yooseph *et al.* (2010) determined which genomes from cultured microorganisms recruited the most fragments from the metagenomes. *Polaribacter* sp. MED152, one of the Bacteroidetes analyzed here, was among the 10 genomes that recruited the most fragments. Only the genomes of Cyanobacteria and '*Pelagibacter*' were covered at greater depth than MED152. Finally, a CARD-FISH study of Bacteroidetes across the North Atlantic Ocean showed *Polaribacter* to be the most abundant Bacteroidetes in most samples (Gómez-Pereira *et al.*, 2010). This shows that the genus is widely distributed and abundant in the oceans.

The better known members of the Bacteroidetes are specialized in processing polymeric organic matter, particularly in the mammalian gut (for example, *Bacteroides* spp.) or in soils (*Cytophaga*). In aquatic habitats, Bacteroidetes are abundant

Correspondence: C Pedrós-Alió, Marine Biology and Oceanography, Institut de Ciències del Mar, Consejo Superior de Investigaciones Científicas (CSIC), Passeig Marítim de la Barceloneta 37-49, Barcelona, Catalunya 08003, Spain.
E-mail: cpedros@icm.csic.es
Received 6 June 2012; revised 26 October 2012; accepted 10 November 2012; published online 10 January 2013

during and following algal blooms (Pinhassi *et al.*, 2004), showing a preference for consuming polymers rather than monomers (Cottrell and Kirchman, 2000). In the oceans, the main lifestyle of Bacteroidetes is assumed to be attachment to particles and degradation of polymers. This assumption is based on a few analyses of free-living and attached bacterial diversity (DeLong *et al.*, 1993), some microautoradiography experiments (Cottrell and Kirchman, 2000), a correlation of Bacteroidetes abundance with phytoplankton blooms (Pinhassi *et al.*, 2004; Teeling *et al.*, 2012), microcosm experiments in which Bacteroidetes were enriched on organic matter particles (Pedrotti *et al.*, 2009), and microscopic observations of CARD-FISH stained Bacteroidetes associated with the phycosphere of nanoplankton cells (Gómez-Pereira *et al.*, 2010; Teeling *et al.*, 2012).

Thus, Bacteroidetes likely have a different life strategy to that of other marine bacteria such as Alphaproteobacteria and Cyanobacteria. The latter are photoautotrophs, while marine Alphaproteobacteria (at least the most abundant ones) are aerobic heterotrophs that preferentially use monomers and live suspended in the water column. If the preference of Bacteroidetes for polymers and an existence attached to surfaces could be confirmed, their role in the carbon cycle of the oceans would be complementary to that of the other two groups. Analysis of the first marine Bacteroidetes genome, that of '*Gramella forsetii*' KT0803 (Bauer *et al.*, 2006), revealed a genome with a relatively large number of glycoside hydrolases (GHs), peptidases and adhesion proteins, suggesting that '*G. forsetii*' was well adapted to degrade HMW compounds. These traits were also observed in *Polaribacter* sp. MED152, the second free-living marine Bacteroidetes genome analyzed (Gónzalez *et al.*, 2008). The latter authors proposed a dual life strategy for *Polaribacter*. This organism would grow optimally attached to particles or other surfaces where it could move by gliding motility, searching for polymeric substances and degrade them using its large array of peptidases and GHs. When labile organic matter was exhausted on a particle, *Polaribacter* would be forced to float passively in the nutrient-poor water column in search of another particle. Under these conditions, it would use proteorhodopsin (PR) to obtain energy from light and, thus, optimize the use of whatever little organic matter it could find. The genome of another marine flavobacterium (*Dokdonia* sp. MED134; González *et al.*, 2011) suggests that this PR-containing bacterium has similar characteristics to those of *Polaribacter*.

Several more genomes have been sequenced recently and it should now be possible to test whether the assumed role of Bacteroidetes as degraders of polymers can be confirmed. We also wanted to check whether the strategy proposed for *Polaribacter* holds with other PR-containing bacteria. We use four well-characterized strains to generate hypotheses: '*G. forsetii*' KT0803 from the North Sea, and *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *Leeuwenhoekiella blandensis* MED217 from the Mediterranean Sea. *Polaribacter* and *Dokdonia* have a gene coding for PR while the other two do not. And then we tested these hypotheses comparing all the marine Bacteroidetes genomes available and a sample of other bacterial genomes. We analyzed which genetic features were characteristic of Bacteroidetes as a phylum, of marine Bacteroidetes in particular and, finally, we compared the characteristics of bacteria with and without PR.

Materials and methods

Isolation of Bacteroidetes

Polaribacter sp. MED152, *Dokdonia* sp. MED134 and *L. blandensis* MED217 were isolated in 2001 from Northwest Mediterranean Sea surface water (0.5 m depth) collected 1 km off the coast at the Blanes Bay Microbial Observatory (BBMO), Spain (41° 40′ N, 2° 48′ E). All three were isolated on ZoBell agar plates. '*G. forsetii*' KT0803 was isolated in 1999 in the North Sea surface water (1 m depth) collected 1 km off the coast of the island of Helgoland, Germany (54° 09′ N, 7° 52′ E). It was isolated on marine synthetic medium (MPM) plates (Schut *et al.*, 1993).

Genome sequencing and assembly

Whole-genome sequencing of *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *L. blandensis* MED217 was carried out by the J. Craig Venter Institute through the Gordon and Betty Moore Foundation initiative in Marine Microbiology. The genome sequences of *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134 have been published (Gónzalez *et al.*, 2008, 2011). The genome of MED152 has been closed (González *et al.*, unpublished results). The genome of MED134 has one gap with an estimated size of 20 000 nucleotides. Thus, we may be missing around 20 genes at most. The genome sequence of *L. blandensis* MED217 is in one scaffold. Sequencing of '*G. forsetii*' KT0803 was carried out at the Max Planck Institute for Molecular Genetics (Berlin) and the genome has been closed (Bauer *et al.*, 2006).

Gene prediction and annotation

Data mining was carried out with GenDB v2.2 (Meyer *et al.*, 2003), supplemented by the tool JCoast (Richter *et al.*, 2008). For each predicted open reading frame, observations were collected from similarity searches against sequence databases NCBI-nr, Swiss-Prot, KEGG and genomes DB (Richter *et al.*, 2008), and Pfam (Finn *et al.*, 2008) and InterPro (Hunter *et al.*, 2009) were used for proteins. SignalP was used for signal peptide

predictions (Bendtsen *et al*., 2004) and TMHMM for transmembrane helix analysis (Krogh *et al*., 2001). Signal transduction proteins, transporters, domains related to surface adhesion and other genes needed for comparison were found using the Pfam search included in JCoast (Richter *et al*., 2008), with a threshold of $E \leqslant 10^{-4}$.

The annotation of *Polaribacter* sp. MED152 (Gónzalez *et al*., 2008), '*G. forsetii*' KT0803 (Bauer *et al*., 2006) and *Dokdonia* sp. MED134 (González *et al.,* 2011) genomes was manually curated and refined for each open reading frame, while the *L. blandensis* MED217 genome remains automatically annotated.

Genomic islands (GIs) of *Polaribacter* sp. MED152, *Dokdonia* sp. MED134, *L. blandensis* MED217 and '*G. forsetii*' KT0803 were extracted (Fernández-Gómez *et al*., 2012) using IslandViewer database for identification and visualization of GIs (Langille and Brinkman, 2009; Langille *et al*., 2010).

Comparative analysis

Comparative analysis was carried out using JCoast (Richter *et al*., 2008). Paralogous protein families were identified by BLASTP (Altschul *et al*., 1990) all-against-all similarity comparisons at a significance cutoff of 10^{-10} and a coverage of $\geqslant$60%). Extracted proteins were grouped into families and functional proteins were classified according to Pfam and Cluster of Orthologous Genes, COG (Tatusov *et al*., 1997).

Determination of the shared gene content was done by a pair-wise BLASTP all-versus-all search among organisms. Reciprocal best matches were counted by a BLASTP result with expectation values of $E < 10^{-5}$ each, and a query coverage of over 65%.

Phylogenetic analysis

Maximum likelihood trees were constructed with full-length 16S rRNA sequences from 47 Bacteroidetes. After alignment using SINA Aligner, offered by the Silva database (Pruesse *et al*., 2007), the poorly aligned positions and divergent regions were eliminated with Gblocks (Castresana, 2000). The trees were constructed using RAxML server (Stamatakis *et al*., 2008) with the GTR nucleotide substitution model. The online tool Interactive Tree of Life (Letunic and Bork, 2007, 2011) was used for editing the tree.

Accession numbers

The complete '*G. forsetii*' KT0803 sequence is available under GenBank accession number CU207366. *Polaribacter* sp. MED152, *Dokdonia* sp. MED134 and *L. blandensis* MED217 sequences are available under GenBank accession numbers AANA00000000, AAMZ00000000 and AANC00 000000, respectively.

Results and Discussion

Comparison among four marine Bacteroidetes

The basic properties of the four genomes analyzed in detail are shown in Table 1. Several tables in Supplementary Information summarize a comparison of the genes or domains identified in the four genomes concerning nitrogen, phosphorous or sulfur acquisition (Supplementary Tables 1–3), sodium transporters (Supplementary Table 4), one- and two-component systems (Supplementary Tables 5 and 6), adhesion (Supplementary Table 7), paralogous families (Supplementary Table 8) and clusters of polymer degradation genes (Supplementary

Table 1 General features of the four Bacteroidetes genomes analyzed

Characteristic	*Polaribacter MED152*[a]	*Dokdonia MED134*[b]	'*Gramella*' *KT0803*[c]	*Leeuwenhoekiella MED217*[d]
Size (mega base pairs)	2961	3302	3799	4244
G+C content (%)	30.7	38.2	36.6	39.8
Protein-coding genes	2646	3008	3585	3689
Coding density (%)	92.9	89.9	90.2	90.6
Stable RNAs				
rRNA operons	2	2	3	4
tRNA	35	43	44	38
Paralogous families	246	245	356	351
Genes in P. families (%)	27.2	25.2	33.3	32.8
Single copy genes (%)	72.8	74.8	66.6	67.2
Phage integrase-like genes	3	1	2	18
Transposases	0	2	68	24
Conjugative transposon	No	Yes	Yes	Yes
Proteorhodopsin	Yes	Yes	No	No

[a]Manually annotated genome published in Gónzalez *et al.* (2008).
[b]Manually annotated genome published in González *et al.* (2011).
[c]Manually annotated genome published in Bauer *et al.* (2006).
[d]Bacterium described in Pinhassi *et al.* (2006). The automatically annotated genome is available in GenBank.

Table 9). The four genomes were compared pairwise by reciprocal best matches. The two larger genomes shared 2122 orthologous genes while the two smaller genomes shared 1762 (Supplementary Table 10). In all cases, shared genes accounted for between 50% and 59% of the genome.

The four genomes ranged in size between 2.97 and 4.24 Mb (Table 1). The two bacteria without PR (PR− from now on) had larger genomes than the two with PR (PR+ from now on). There were at least three reasons why two of the genomes were larger. First, they had more paralogous families and a larger percent of genes in such families (Table 1; Figure 1). The percentage of genes in paralogous families follows an already recognized linear relationship with genome size for a large selection of bacteria (Figure 1; Pushker *et al.*, 2004; Woyke *et al.*, 2009). Gene duplication and subsequent modifications to carry out novel functions by paralogous proteins is a well-known mechanism of bacterial evolution (review in Andersson and Hughes, 2009).

Second, some genes were present in the larger genomes that code additional functions absent from the smaller ones. The area labeled Z in Figure 2, for example, was present only in the two large genomes. The genes in this area were mostly involved in the metabolism of sugars, particularly arabinose (Supplementary Figure 1): they included the three structural genes of the arabinose operon (*ara*A, *ara*D and *ara*B that are responsible for converting arabinose into D-xylulose-5P), *gal*M (codes an epimerase capable of interconverting L- and D-arabinose), and a few genes related to sugar metabolism, including a sodium/glucose co-transporter. Presence of these genes should allow the bacterium to use arabinose by channeling it to the pentose phosphate pathway (present in the four genomes). In effect, when utilization of arabinose was experimentally tested, '*G. forsetti*' and MED217 could use arabinose while the two other bacteria could not (Francesca Simonato and OI Nedashkovskaya, personal communication).

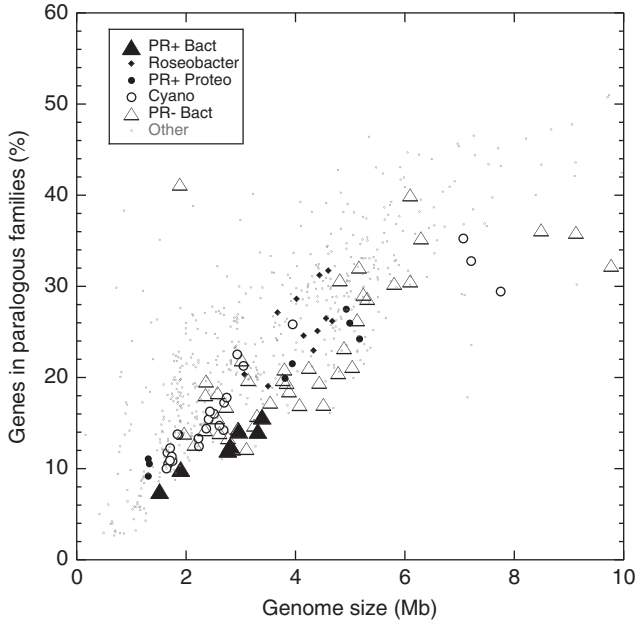


Figure 1 Percentage of genes in paralogous families versus genome size for a number of bacteria. The number of genes in paralogous families was estimated using the BLASTCLUST tool from the NCBI BLAST software (>30% sequence similarity, across >50% of their length and $E<10^{-6}$).

The region labeled W in MED217 is quite extensive and most of it is missing in the other three genomes. This region contained 218 open reading frames. About half of these were hypothetical proteins and 25 more were only identified putatively. Among the remaining open reading frames the number of genes involved in sugar metabolism was remarkable. For example, four copies of beta-galactosidase, one arabinosidase and several regulatory proteins, including two of the arabinose operon, were found. This suggests that MED217 is capable of utilizing a far larger number of sugars than the other three.

Third, the larger genomes also had more mobile elements than the smaller ones (Table 1). Many mobile elements (such as transposases, phage integrases, integrons and conjugative transposons) are usually found next to, or within, GIs, contributing to their mobility. We have recently shown a significant correlation between number of GIs and genome size for 70 marine bacteria ($R=0.53$; $P<0.001$) and this correlation was also observed for the marine Bacteroidetes with a positive and significant correlation ($R=0.78$: $P<0.001$) (Fernández-Gómez *et al.*, 2012). Thus, as expected, we found higher numbers of mobile elements and GIs in the larger genomes. The MED217 genome had eight GIs, with a total of 70.5 kb (1.18% of the genome) and '*G. forsetii*' had five GIs with a total of 105 kb (2.27%). On the other hand, MED134 and MED152 had only two and one GIs (39 and 12 kb, respectively). Some of the genes present in MED217, and missing in the two smaller genomes, were actually located within GIs. Thus, missing areas labeled A to H in Figure 2 corresponded to GIs detected in MED217. Genomic island H, for example, had integrases and transposases at both ends of the island, and Clustered Regularly Interspaced Palindromic Repeats at one end. The latter are repetitive palindromic sequences that, in association with genes *cas*7, *cas*5 and two *cas*3 (also present), serve a defensive function against phages and can be mobilized by horizontal gene transfer (Haft *et al.*, 2005). Most of the annotated genes within the island coded for hypothetical proteins (up to 55%) as previously reported for other marine bacteria (Hsiao *et al.*, 2005). However, a regulatory sigma factor and a putative nitrite reductase were also found in the H genomic island of MED217. Despite the fact that GIs account for only a small part of the differences in size, they may have a very important qualitative importance.

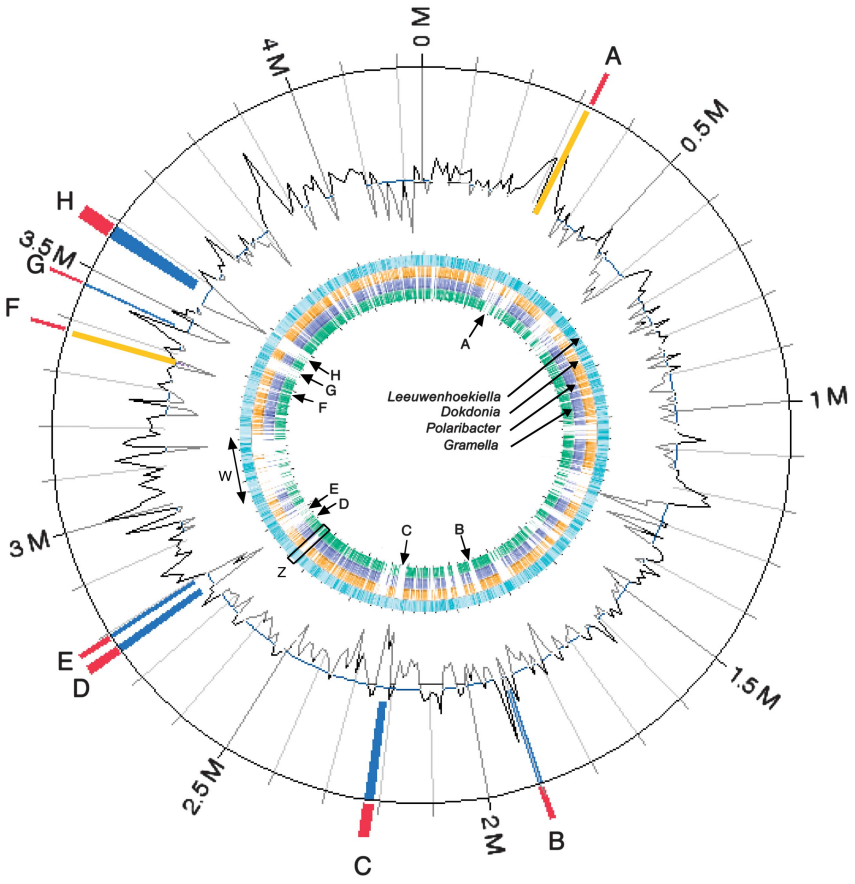


Figure 2 The blue ring shows the genome of *Leeuwenhoekiella blandensis* MED217 ordered from 0 to 4.24 Mb. The next outer ring shows the G+C content along the genome. The outermost ring shows the areas identified as genomic islands with the IslandViewer software. Red, blue and yellow indicate the three different tools used for island prediction in this package. The inner colored rings are the genomes of *Dokdonia* MED134, *Polaribacter* MED152 and '*Gramella forsetii*' KT0803 compared with that of MED217. Each one of these genomes has been compared with that of MED217 by reciprocal best matches and, whenever a match was found, the gene of these genomes was placed next to the position of MED217 best match. That is, these three genomes do not keep their original topology so that genomes can be compared gene to gene. The letters A to H indicate genes present in MED217 and absent from some or all the other genomes, which were identified as genomic islands. Rectangle labeled Z and double arrow labeled W show areas in MED217 that were absent from some of the other genomes, but were not genomic islands.

In summary, the reduced number of paralogous families and genes within such families, the absence of certain metabolic pathways, and the lack of GIs are three important mechanisms accounting for the differences in genome size of these bacteria.

Adhesion and gliding

The four marine Flavobacteria had between 21 and 27 domains involved in surface adhesion (Supplementary Table 7). This represented at least six adhesion genes per Mb (Figure 3a). In contrast, other pelagic bacteria such as the Alphaproteobacteria '*P. ubique*' and *Ruegeria pomeroyi*, and the Gammaproteobacteria *Idiomarina loihiensis* and *Vibrio parahaemolyticus* only had one or two genes per Mb. As could be expected, *Bacteroides thetaiotaomicron* had the most adhesion genes per Mb of all (nine), since it lives attached to the intestinal epithelium of vertebrates (Figure 3a). The marine Bacteroidetes also had two to three times more glycosyl transferases per Mb than the Proteobacteria (with the exception of *P. ubique*, Figure 3b). These proteins are usually positioned in the outer membrane and generate polysaccharides for, for example, attachment. Altogether, this indicates that attachment to particles is an important feature of marine Flavobacteria, while this is not the case for common planktonic Proteobacteria such as '*Pelagibacter*', *Ruegeria* or *Vibrio*.

Most motile bacteria living on surfaces use gliding motility, a name that actually includes several different molecular mechanisms (Jarrell and McBride, 2008). The four marine Flavobacteria had the full complement of 15 genes for gliding motility. Moreover, the three MED strains were shown to have gliding motility under the microscope (OI Nedashkovskaya, personal communication).

Polymer degrading enzymes: peptidases and GHs

We show a comparison of polymer degrading enzymes among a large number of genomes in Supplementary Figure. 2 and a few selected

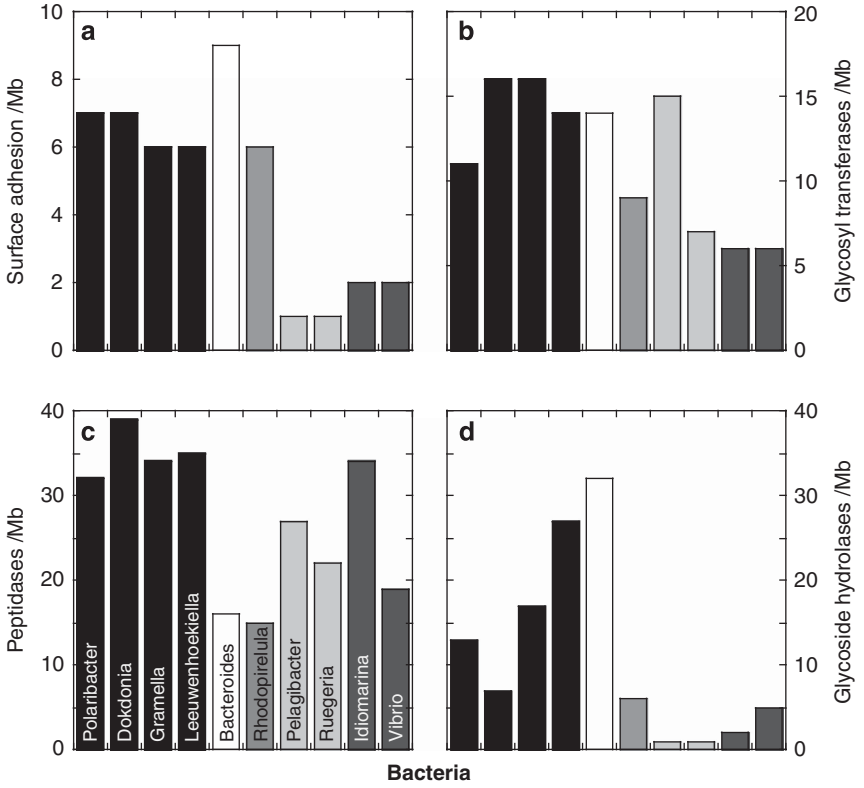


Figure 3 Numbers of different enzymes per megabase of genome for a selection of bacteria: Marine *Bacteroidetes* (black bars) *Polaribacter* sp. MED152, *Dokdonia* sp. MED134, '*Gramella forsetii*' KT0803 and *Leeuwenhoekiella blandensis* MED217; *Bacteroides thetaiotaomicron* (white bar); *Rhodopirellula baltica* (medium gray); the Alphaproteobacteria (light gray), '*Pelagibacter ubique*' and *Ruegeria pomeroyi*; and the Gammaproteobacteria (dark gray) *Idiomarina loihiensis* and *Vibrio parahaemolyticus*. (a) Number of surface adhesion proteins per Mb of genome, (b) glycosyl transferases per Mb, (c) peptidases per Mb and (d) glycoside hydrolases per Mb.

examples in Figures 3c and d. The number of peptidases and GHs increased with the size of the genome in all bacteria (Supplementary Figures 2A and C). Most Bacteroidetes had more of these enzymes than the average bacterium, irrespectively of the genome size (Supplementary Figure 2). This is one of the major observations showing the dedicated role of marine Bacteroidetes as polymer degraders.

In the case of peptidases, the marine Flavobacteria had significantly more genes per Mb than other marine bacteria (Figure 3c; Supplementary Figure 2B). As expected, *B. thetaiotaomicron* had a lower number of peptidases, since it is a polysaccharide specialist (Martens *et al.*, 2009). The relatively high number of peptidases per Mb in '*P. ubique*' is remarkable, especially combined with its low number of GHs, suggesting a preference for proteins over polysaccharides. '*P. ubique*', however, still has less peptidases per megabyte of genome than the marine Bacteroidetes. Finally, *Idiomarina loihiensis* had a large number of peptidases per Mb, in accordance with its putative specialization in the fermentation of amino acids and proteins (Hou *et al.*, 2004). In summary, marine Flavobacteria had a relatively large complement of both GHs and peptidases compared with other marine bacteria and, moreover, they had as many peptidases as other protein specialists.

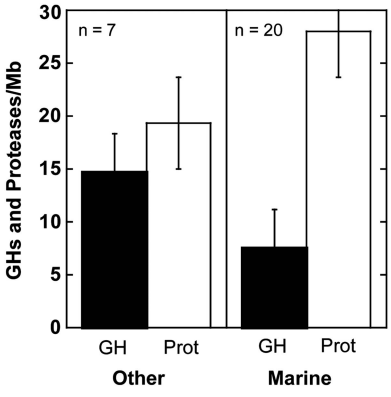


Figure 4 Peptidases and glycoside hydrolases per Mb for marine and non-marine Bacteroidetes.

Marine Flavobacteria also had more GHs per Mb than other planktonic bacteria and the difference was statistically significant (Figure 3d; Supplementary Figure 2D). *Dokdonia* MED134 had a lower number of GHs but still higher than those of other planktonic bacteria such as '*Pelagibacter*', *Ruegeria* or *Idiomarina* (Figure 3d; Supplementary Figure 2D). *B. thetaiotaomicron* had the largest number of GHs per Mb in accordance with its specialization in polysaccharide degradation.

A striking observation was that marine Bacteroidetes had many more peptidases than GHs (Figure 4). This was not the case for the non-marine Bacteroidetes examined. This strongly suggests a specialization of marine Bacteroidetes on the degradation of proteins, which is consistent with experimental studies using microautoradiography (Cottrell and Kirchman, 2000). Gómez-Pereira *et al.* (2012) analyzed Bacteroidetes fosmids from two regions in the N. Atlantic Ocean. They found many polysaccharide degrading enzymes in the phytoplankton-rich polar waters and an abundance of protein degrading enzymes in the subtropical North Atlantic. The latter results are in line with the evidence presented here. The polar zone results, however, suggest the opposite. Isolates from polar waters are needed to test whether the relationship in Figure 4 holds for all marine Bacteroidetes or only for those in temperate zones.

Despite these general characteristics, each one of the four bacteria showed a different suite of enzymes (Supplementary Figure 3), like variations on a shared theme. Probably, in combination with other genes, these differences allow the various species to occupy slightly different niches. We calculated diversity indices for these enzymes and found that the four bacteria had very similar values. In the case of peptidases the indices varied between 3.57 and 3.68 and in the case of GHs between 2.82 and 3.20. These indices show that not only do these bacteria have more peptidases than GHs, but that there is a larger diversity of the former. Thus, the conclusion that protein degradation is the main speciality of marine Bacteroidetes is robust.

Receptors and transporters for high and low molecular weight compounds

If Bacteroidetes are specialized in using HMW compounds, then this should be reflected in a higher proportion of transporters for HMW than for low molecular weight compounds, relative to other bacteria. A fruitful approach to detect important functions in bacteria is to look at the main families of paralogous genes (Supplementary Table 8). The three largest families were the same for the four Flavobacteria: two-component systems (between 30 and 54 genes), TonB-dependent receptors (18–50) and ABC transporters (22–30). These numbers are in striking contrast to those of two common Proteobacteria such as *E. coli* or *Ruegeria pomeroyi*. The latter bacteria had ABC transporters as the largest family, followed by LysR one-component systems. Moreover, the Bacteroidetes had twice the number of TonB-dependent receptors per Mb than the Proteobacteria.

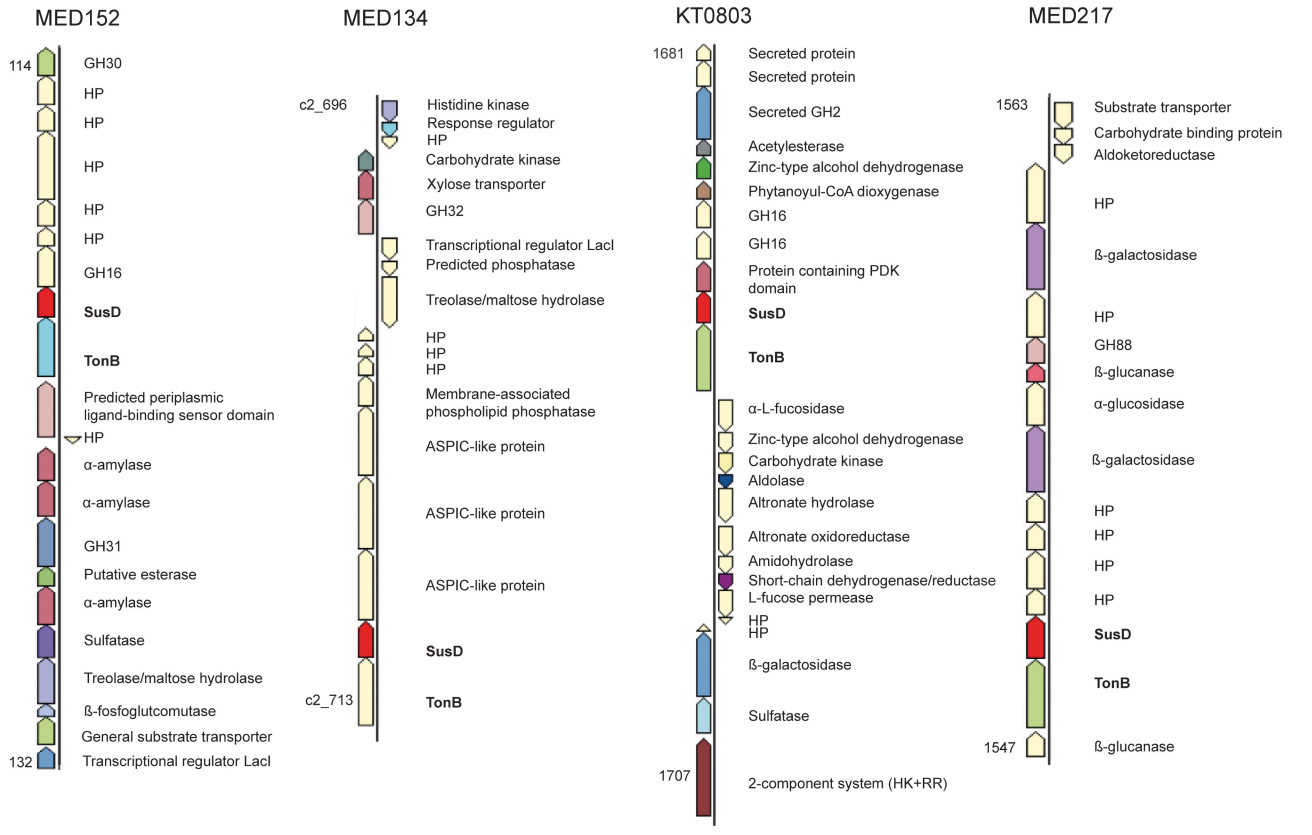


Figure 5 Clusters of genes putatively involved in the attachment and degradation of polymeric compounds and containing TonB-dependent/ligand-gated channel genes and SusD genes. One example from each of the four bacteria is shown. The complete list can be found in Supplementary Table 9. HP, hypothetical protein; HK, histidine kinase; RR, response regulator.

The abundance of TonB-dependent receptors suggests a specialization in degradation of polymers. The four bacteria have between 5 and 21 *susC* plus *susD*-like pairs of genes. In *B. thetaiotaomicron*, SusC is a member of the TonB receptor family specialized in the transport of oligosaccharides from the outer membrane into the periplasmic space. SusD, in turn, is necessary to bind polysaccharides to the outer cell membrane. These two genes alone are enough to account for 60% of the polysaccharide degrading ability of *B. thetaiotaomicron* (Martens *et al.*, 2009). This is likely the role of the many *susC* plus *susD* pairs in the four Bacteroidetes. In effect, these pairs are always next to genes encoding polymer degrading enzymes: sulfatases, amylases, GHs, peptidases and alkaline phosphatases among other enzymes (Figure 5). In *B. thetaiotaomicron*, there are 101 individual pairs of '*susC*-like' and '*susD*-like' genes. This bacterium is a polysaccharide specialist and has a large genome of 6.26 Mb (resulting in 16 *susC–susD* pairs per Mb). In *Flavobacterium johnsoniae*, a soil organism specialized in degradation of polysaccharides, there are 50 TonB-dependent receptors and 10 of these also have the *sus*C plus *susD* combination (McBride *et al.*, 2009). This represents only 1.7 pairs per Mb. In the four marine Bacteroidetes, with much smaller genomes, we found 6, 4, 21 and 17 such pairs in *Polaribacter*, *Dokdonia*, *Leeuwenhoekiella* and 'Gramella' respectively (1.7, 1.2, 4.5 and 5 per Mb, respectively) (Supplementary Table 9). This is in line with the hypothesis of an adaptation to the use of polymers in the marine Bacteroidetes.

Transporters for low molecular weight compounds

Specialization in HMW compounds implies a lower capacity to use low molecular weight compounds. This feature is reflected in a low number of transporters relative to other bacteria (Figure 6). It is striking that the Bacteroidetes clearly follow a pattern different from that of the other bacteria. The slopes of the regressions implied 115 transporter genes per Mb for Gammaproteobacteria ($r^2 = 0.634$, $P < 0.0001$), 67 for Alphaproteobacteria ($r^2 = 0.531$, $P < 0.0001$), 65 for Betaproteobacteria ($r^2 = 0.927$, $P < 0.0001$), 24 for Cyanobacteria ($r^2 = 0.701$, $P < 0.0001$) and 22 for Bacteroidetes ($r^2 = 0.667$, $P < 0.0001$) (Figure 6). The slope for Bacteroidetes was significantly different from all the other except Cyanobacteria ($P = 0.561$). In the case of the latter, a low number of transporters for organic matter is logical since they have an autotrophic way of life. In the case of Bacteroidetes, as shown in the previous section, this low number of transporters for low molecular weight compounds is compensated by transporters for HMW compounds. Since there are both marine and non-marine bacteria in all groups, this feature is clearly a characteristic of Bacteroidetes in general and does not represent an adaptation to marine life. Rather, marine

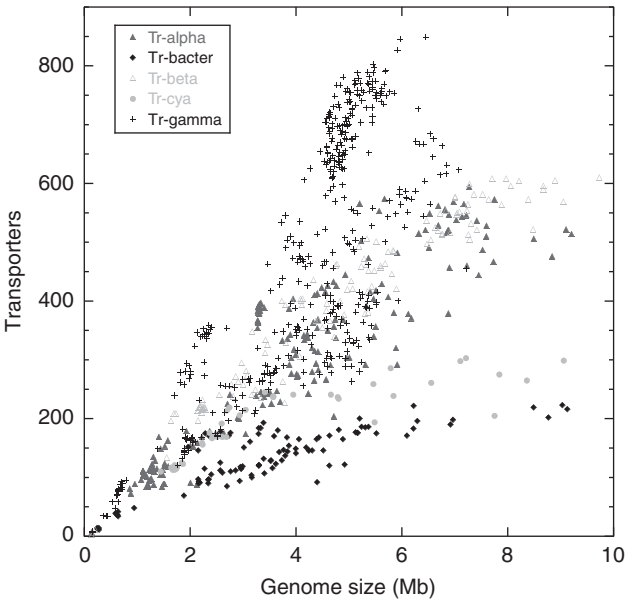


Figure 6 Number of transporters versus genome size for different groups of bacteria: Gammaproteobacteria (crosses), Alphaproteobacteria (solid triangles), Betaproteobacteria (empty triangles), Cyanobacteria (circles) and Bacteroidetes (solid diamonds).

Bacteroidetes must cope with this relatively low number of transporters when living in the ocean. This fact is in agreement with the preference for polymers found in MAR-FISH studies (Cottrell and Kirchman 2000).

PR as a potential adaptation to the oligotrophic surface ocean

Once established that marine Bacteroidetes are well prepared to live on particles using HMW compounds, it is necessary to see how they can survive in the water column when searching for new particles. In *Polaribacter* sp. MED152, it was proposed that PR, in combination with a large suite of light sensing genes and a remarkable number of proteins involved in anaplerotic carbon fixation, represented adaptations to survive in the surface ocean while traveling between particles (Gónzalez *et al.*, 2008). This bacterium has a relatively small genome and it was postulated that it was only able to use a small number of monomeric carbon compounds, a fact consistent with the very low number of transporters compared with other bacteria and with a lower number of sugars used compared with MED217 (O Nedashkovskaya, personal communication).

The presence of PR in the Bacteroidetes is a characteristic widely distributed throughout the different clusters of Flavobacteria (Supplementary Figure 4). Moreover, both PR+ and PR− bacteria are found within the same clusters, such as the Flavobacterium bacterium BAL38 (PR+) and *Flavobacterium psychrophilum* JIP02/86 (PR−). In this section, we examine which of the features

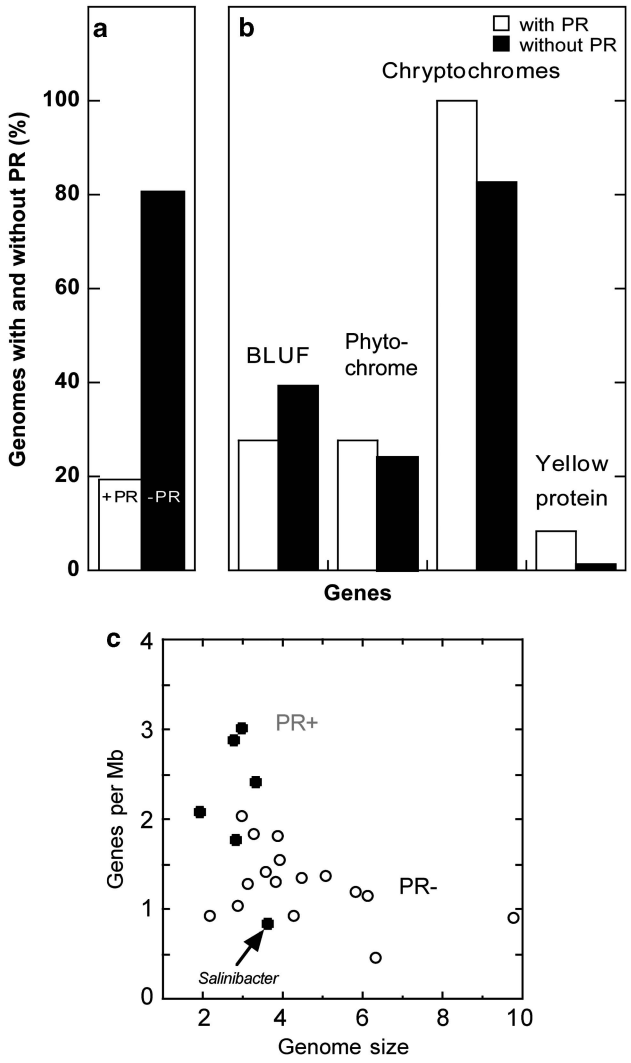


Figure 7 (**a**) Percent of genomes from a collection of 185 marine bacteria with proteorhodopsin (PR+, empty bars) and without it (PR−, filled bars). (**b**) Percent of PR+ or PR− genomes with different light sensing domains. (**c**) Genes involved in anaplerotic fixation and transport of CO_2 in different genomes versus genome size. PR+ (filled symbols) and PR− (empty symbols).

of *Polaribacter* sp. MED152 are consistent with the presence or absence of PR in other Bacteroidetes.

First, let's consider the distribution of light sensing domains in bacteria in general. Figure 7 shows a comparison between 185 genomes of marine bacteria with and without PR. Close to 20% have the PR gene (Figure 7a). The other light-related domains are present in different numbers of genomes. Thus, cryptochromes are widespread both in PR+ and PR− bacteria, while the photoactive yellow protein is present in very few genomes (Figure 7b). In both cases a higher percent of PR+ bacteria have the domain, but the difference is not significant. In conclusion, no significant difference in the number of light sensing domains seems to exist between bacteria with and without PR.

When looking at the distribution of these light sensing domains within PR+ bacteria, however, different patterns emerge (Supplementary Table 11). All of them possess the photolyase class I gene that codes for a widely distributed DNA-repair protein. Two other types of this family of proteins are present in most Gammaproteobacteria and in all the Bacteroidetes. However, they are missing from the two '*Pelagibacter*' genomes analyzed. Finally, two of the four Bacteroidetes have proteins with BLUF and phytochrome domains. This comparison indicates that different PR+ bacteria must use light with different strategies. It is intriguing that even two very closely related bacteria such as the two *Polaribacter* strains have different sensors, while bacteria belonging to different genera, like *Polaribacter* and *Dokdonia*, do share the same light sensing domains. This decoupling between phylogeny and function suggests that lateral gene transfer may be involved in the distribution of PR genes (González *et al.,* 2011).

The second feature hypothesized to enhance survival under oligotrophic conditions was the large number of proteins involved in anaplerotic carbon fixation found in MED152 (Gónzalez *et al*., 2008). We considered several proteins in this group (Supplementary Figure 5): (a) two enzymes involved in the glyoxylate shunt (isocitrate lyase and malate synthase), (b) two enzymes involved in CO_2 fixation with phosphoenolpyruvate (phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase), (c) two enzymes that fix CO_2 with pyruvate (pyruvate carboxylase and malic enzyme) and (d) three proteins involved in bicarbonate acquisition and interconversion (carbonic anhydrase and the two bicarbonate transporters *bic*A and *sbt*A).

PR+ Bacteroidetes had small genomes with a large number of genes involved in these three processes (Figure 7c). Other Bacteroidetes had variable and, in most cases, lower numbers. All the bacteria having >2.0 such genes per Mb had PR. These included *Candidatus* '*Pelagibacter ubique*' and the Gammaproteobacterium HTCC2207 (not shown in the figure). Most other bacteria had a value lower than 1.7. The only PR+ bacterium with values significantly lower than two was *Salinibacter ruber*. The latter, however, is a specialist of hypersaline environments and its life strategy is very different from those of marine bacteria. Thus, the possession of a large suite of genes involved in anaplerotic carbon fixation seems to be linked with the possession of PR in Bacteroidetes (present in five out of six genomes). This, however, does not mean that other bacteria do not have anaplerotic metabolism, only that they do not have as many enzymes capable of fixing CO_2.

In summary, there is a co-occurrence between the PR gene and a large number of anaplerotic CO_2 fixation genes in the genomes of Bacteroidetes. In addition, *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134 have a large number of light sensing/utilization genes. '*Pelagibacter*' shares with the Bacteroidetes the number of anaplerotic CO_2 fixation genes, but not the light sensing and utilizing genes.

Genome size

The two bacteria with PR had smaller genomes than the two without PR (Table 1). We tested whether this was true for all the other available genomes. In effect, all the PR+ Bacteroidetes had genomes smaller than most PR− Bacteroidetes (Figure 1). Among the PR+ Proteobacteria only the '*Pelagibacter*'-like genomes were smaller than those of the PR+ Bacteroidetes, while other PR+ Proteobacteria (for example, Gammaproteobacteria) had larger genomes. The number of known PR+ marine bacteria is still low and this pattern needs to be confirmed with more genomes. However, the environmental Bacteroidetes genomes MS024-2A and MS024-3C also followed this trend (see the two smallest genomes labeled as PR+ Bacteroidetes in Figure 1): they had the PR gene and their genomes were estimated to be smaller than that of MED152, adding strength to the argument of PR+ Bacteroidetes having small genomes.

Whether the PR+ genomes have experienced streamlining or whether, on the contrary, the PR− genomes have increased in size remains unknown. However, it is tempting to conclude that possession of PR allows bacteria to reduce their genomes. Perhaps, the extra mechanism for energy conservation allows the cells to be less versatile in their carbon source preferences. In other words, they may not need to carry the genes for many different carbon utilization pathways because light energy makes them more independent from organic compounds as energy sources. The small number of transporters (see above) is consistent with this idea.

Unlike the PR+ Bacteroidetes, the PR+ Gammaproteobacteria varied more widely in genome size (Figure 1). It has been suggested that PR has different roles in different marine bacteria (Fuhrman *et al.*, 2008; DeLong and Béjà, 2010). So far, higher cell yields in the light than in the dark have been demonstrated only in *Dokdonia* sp. MED134 (Gómez-Consarnau *et al.*, 2007; Kimura *et al.*, 2011) and in *Polaribacter* sp. MED152 (Férnandez-Gómez *et al.*, in preparation). Enhanced growth in the light has not been seen in '*Pelagibacter*' (Giovannoni *et al.*, 2005) or in Gammaproteobacteria (Stingl *et al.*, 2007). But the usefulness of PR under starvation conditions has been demonstrated in one Gammaproteobacterium (Gómez-Consarnau *et al.*, 2010) and in '*Pelagibacter*' (Steindler *et al.*, 2011). The different genomic characteristics described here are in accordance with these different strategies to use PR.

Conclusions

Our comparative study has shown that the marine Bacteroidetes share the capacity for adhesion and gliding motility, the presence of abundant glycosyl transferases, a large number of polymer degrading enzymes including GHs and, especially, peptidases, and a relatively large number of SusC–SusD pairs associated with many different degrading enzymes. This confirms the role of this abundant group of marine bacteria as degraders of particulate matter, especially of proteins.

The two characteristics shared by all known PR+ Bacteroidetes are small genome size and a higher number of genes involved in CO_2 fixation per Mb than the PR− Bacteroidetes. Interestingly, *Polaribacter* sp. MED152 and *Dokdonia* sp. MED134 shared the same light sensing domains, but those of *Polaribacter irgensii* 23-P were different, despite the latter belonging to the same genus as MED152. Within the PR+ Proteobacteria, analysis of their genomes identified at least two different strategies corresponding to *Pelagibacter*-like bacteria on the one hand, and Gammaproteobacteria on the other. Clearly, possession of PR is used in different ways by different bacteria.

Acknowledgements

BF-G was a recipient of an I3P grant from CSIC. The genomes of the three MED strains were sequenced by the JCVI through the Marine Microbial Initiative of the Gordon and Betty Moore Foundation. We thank Thierry Lombardot for bioinformatic analysis. This work was supported by grants 'GEMMA' CTM2007-63753-C02-01/MAR and 'BBMOGenome' CTM2010-11060-E from the Spanish Ministerio de Ciencia e Innovación.

References

Alonso C, Warnecke F, Amann R, Pernthaler J. (2007). High local and global diversity of Flavobacteria in marine plankton. *Environ Microbiol* **9**: 1253–1266.

Alonso-Sáez L, Gasol J. (2007). Seasonal variations in the contributions of different bacterial groups to the uptake of low-molecular-weight compounds in Northwestern Mediterranean Coastal waters. *Appl Environ Microbiol* **73**: 3528–3535.

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

Amaral-Zettler L, Artigas LF, Baross J, Loka Bharathi PA, Boetius A, Chandramohan D *et al.* (2010). A global census of marine life. In: McIntyre AD (eds) *Life in the World's Oceans*. Blackwell Publishing Ltd.: Chichester, West Sussex, UK, pp. 223–245.

Andersson DI, Hughes D. (2009). Gene amplification and adaptive evolution in bacteria. *Annu Rev Genet* **43**: 167–195.

Bauer M, Kube M, Teeling H, Richter M, Lombardot T, Allers E *et al.* (2006). Whole genome analysis of the marine Bacteroidetes '*Gramella forsetii*' reveals adaptations to degradation of polymeric organic matter. *Environ Microbiol* **8**: 2201–2213.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. (2004). Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* **340**: 783–795.

Castresana J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* **17**: 540–552.

Cottrell MT, Kirchman DL. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-

Flavobacteria cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66**: 1692–1697.

DeLong EF, Béjà O. (2010). The light-driven proton pump proteorhodopsin enhances bacterial survival during tough times. *PLoS Biol* **8**: e1000359.

DeLong EF, Franks DG, Alldredge AL. (1993). Phylogenetic diversity of aggregate-attached vs. free-fiving marine macterial assemblages. *Limnol Oceanogr* **38**: 924–934.

Fernández-Gómez B, Fernàndez-Guerra A, Casamayor EO, González JM, Pedrós-Alió C, Acinas SG. (2012). Patterns and architecture of genomic islands in marine bacteria. *BMC Genomics* **13**: 347–366.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H-R *et al.* (2008). The Pfam protein families database. *Nucleic Acids Res* **36**: D281–D288.

Fuhrman JA, Schwalbach MS, Stingl U. (2008). Proteorhodopsins: an array of physiological roles? *Nat Rev Microbiol* **6**: 488–494.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D *et al.* (2005). Genome streamlining in a cosmopolitan oceanic bacterium. *Science* **309**: 1242–1245.

Glöckner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Appl Environ Microbiol* **65**: 3721–3726.

González JM, Pinhassi J, Fernández-Gómez B, Coll-Llado M, González-Velázquez M, Puigbó P *et al.* (2011). Genomics of the proteorhodopsin-containing marine Flavobacterium *Dokdonia* sp. strain MED134. *Appl Environ Microbiol* **77**: 8676–8686.

Gómez-Consarnau L, Akram N, Lindell K, Pedersen A, Neutze R, Milton DL *et al.* (2010). Proteorhodopsin phototrophy promotes survival of marine bacteria during starvation. *PLoS Biol* **8**: e1000358.

Gómez-Consarnau L, González JM, Coll-Lladó M, Gourdon P, Pascher T, Neutze R *et al.* (2007). Light stimulates growth of proteorhodopsin-containing marine Flavobacteria. *Nature* **445**: 210–213.

Gómez-Pereira PR, Fuchs BM, Alonso C, Oliver M, van Beusekom J, Amann R. (2010). Distribution patterns and diversity of planktonic Flavobacterial clades in contrasting water masses of the North Atlantic Ocean. *ISME J* **4**: 472–487.

Gómez-Pereira PR, Schüler M, Fuchs BM, Bennke C, Teeling H, Waldmann J *et al.* (2012). Genomic content of uncultured Bacteroidetes from contrasting oceanic provinces in the North Atlantic Ocean. *Environ Microbiol* **14**: 52–66.

Gónzalez JM, Fernández-Gómez B, Fernàndez-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M *et al.* (2008). Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). *Proc Natl Acad Sci USA* **105**: 8724–8729.

Haft DH, Selengut J, Mongodin EF, Nelson KE. (2005). A guild of 45 CRISPR-associated (Cas) protein families and multiple CRISPR/Cas subtypes exist in prokaryotic genomes. *PLoS Comput Biol* **1**: e60.

Hou S, Saw JH, Lee KS, Freitas TA, Belisle C, Kawarabayasi Y *et al.* (2004). Genome sequence of the deep-sea gamma-proteobacterium *Idiomarina loihiensis* reveals amino acid fermentation as a source of carbon and energy. *Proc Natl Acad Sci USA* **101**: 18036–18041.

Hsiao WWL, Ung K, Aeschliman D, Bryan J, Finlay BB, Brinkman FSL. (2005). Evidence of a large novel gene pool associated with prokaryotic genomic islands. *PLoS Genetics* **1**: 540–550.

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D *et al.* (2009). InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**: D211–D215.

Jarrell KF, McBride MJ. (2008). The surprisingly diverse ways that prokaryotes move. *Nat Rev Microbiol* **6**: 466–476.

Kimura H, Young CR, Martínez A, DeLong EF. (2011). Light-induced transcriptional responses associated with proteorhodopsin-enhanced growth in a marine flavobacterium. *ISME J* **5**: 1641–1651.

Kirchman D. (2002). The ecology of Cytophaga-Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**: 91–100.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. (2001). Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *J Mol Biol* **305**: 567–580.

Langille MGI, Brinkman FSL. (2009). IslandViewer: an integrated interface for computational identification and visualization of genomic islands. *Bioinformatics* **25**: 664–665.

Langille MGI, Hsiao WWL, Brinkman FSL. (2010). Detecting genomic islands using bioinformatics approaches. *Nat Rev Microbiol* **8**: 373–382.

Letunic I, Bork P. (2007). Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.

Letunic I, Bork P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res* **39**: W475–W478.

Martens EC, Koropatkin NM, Smith TJ, Gordon JI. (2009). Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* **284**: 24673–24677.

McBride MJ, Xie G, Martens EC, Lapidus A, Henrissat B, Rhodes RG *et al.* (2009). Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* **75**: 6864–6875.

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.

Pedrotti ML, Beauvais S, Kerros ME, Iversen K, Peters F. (2009). Bacterial colonization of transparent exopolymeric particles in mesocosms under different turbulence intensities and nutrient conditions. *Aquat Microb Ecol* **55**: 301–312.

Pinhassi J, Bowman JP, Nedashkovskaya OI, Lekunberri I, Gómez-Consarnau L, Pedrós-Alió C. (2006). Leeuwenhoekiella blandensis sp. nov., a genome-sequenced marine member of the family Flavobacteriaceae. *Int J Sys Evol Microbiol* **56**: 1489–1493.

Pinhassi J, Sala MM, Havskum H, Peters F, Guadayol O, Malits A *et al.* (2004). Changes in bacterioplankton composition under different phytoplankton regimens. *Appl Environ Microbiol* **70**: 6753–6766.

Pommier T, Canback B, Riemann L, Bostrom KH, Simu K, Lundberg P *et al.* (2007). Global patterns of diversity and community structure in marine bacterioplankton. *Mol Ecol* **16**: 867–880.

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.

Pushker R, Mira A, Rodríguez-Valera F. (2004). Comparative genomics of gene-family size in closely related bacteria. *Genome Biol* **5**: R27.

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

Schut F, de Vries EJ, Gottschal JC, Robertson BR, Harder W, Prins RA *et al.* (1993). Isolation of typical marine bacteria by dilution culture: growth, maintenance, and characteristics of isolates under laboratory conditions. *Appl Environ Microbiol* **59**: 2150–2160.

Stamatakis A, Hoover P, Rougemont J. (2008). A rapid bootstrap algorithm for the RAxML web servers. *Syst Biol* **57**: 758–771.

Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. (2011). Energy starved *Candidatus* Pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS ONE* **6**: e19725.

Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73**: 2290–2296.

Tatusov RL, Koonin EV, Lipman DJ. (1997). A genomic perspective on protein families. *Science* **278**: 631–637.

Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM *et al.* (2012). Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**: 608–611.

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.

Yooseph S, Nealson KH, Rusch DB, McCrow JP, Dupont CL, Kim M *et al.* (2010). Genomic and functional adaptation in surface ocean planktonic prokaryotes. *Nature* **468**: 60–66.

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

Supplementary Information

Manuscript Number: ISMEJ-11-00825OA

Ecology of marine Bacteroidetes: A comparative genomics approach

Beatriz Fernández-Gómez, Michael Richter, Margarete Schüler, Jarone Pinhassi, Silvia G. Acinas, José M. González, and Carlos Pedrós-Alió

Supplementary Table 1. Enzymes and proteins identified in the four Bacteroidetes involved in nitrogen metabolism (Locus tags for each gene begins with MEDXXX_ for the MED strains and GFO_ for "*Gramella forsetii*", followed by the number in the table).

	Locus tag			
Gene/protein	**MED152**	**MED134**	**KT0803**	**MED217**
Glutamate synthase	05810-05815	13936-13941	2908-2909	10182-10187
Ammonium channel *amt*B	05800 05810	13926	0493	10202
Nitrogen regulatory protein P-II	05805		0494 1188 1225	10197 02305
Glutamine synthetase type I				
Glutamine synthetase type II	05925	03484		04482
Glutamine synthetase type III	05920	03489	0471	04487
Glutamate dehydrogenase	08660	09951	3186	00470
Carbon-nitrogen hydrolase	07120 09410	05909 08166	2116 2172 3530	09030 11799 17720
Allophanate hydrolase	11869-11874	02550-02545	2121-2122	09812-09817
Nitrousoxide reductase (nosZ)			1411	
Nitroreductase	00945 04935 06510	00720 04219	2187 2666	05357 13776
Nitrite and Sulphite reductase	06160		0327	02060
Nitrogen regulatory protein NtrY	09435	288	3216	17875
Dipeptidyl peptidase IV	06250 06600	00860 08711	2972	03390 12504
POT family (Oligopeptide transporter)	06260	00865	2970 2971	03395 03400
AGCS family (Na-Ala symporter)	04880	02355 02490 14567	2366 2391 2396 2661	08225 14120 14125
BCCT family:				
GlyBet			2804	04132
Chol/GlyBet			2552	12719
Urea transporter and assimilation				

Supplementary Table 2. Enzymes and proteins identified in the four Bacteroidetes involved in phosphorus metabolism (Locus tags for each gene begins with MEDXXX_ for the MED strains and GFO_ for "*Gramella forsetii*", followed by the number in the table).

	Locus tag			
Gene/protein	**MED152**	**MED134**	**KT0803**	**MED217**
Alkaline phosphatase (periplasmatic)	12659		2880	03720
Phosphate porin			3192	13149
Phosphate permease	09510	11180	3193	17570
Nucleoside H^+ symporter	05175	09696	1025	00385
Na^+ dependent nucleoside transporter	11959	00115	0290	01910
Nucleoside phosphorylase	06880	00750	3610 1416	16075
Nicotinamide mononucleotide transport	09745	10051	3059	01020 11939
PolyP/ATP NAD kinase	07195	07581	0215	16215
Polyphosphate kinase		04654 06384	1621 1833	06182 10737
Polyphosphate glucokinase	01955	04409	2545	04167
Exopolyphospatase	04590	06373	1835 3499	10732
H^+-translocating pyrophosphatase	11924	00150	0297	01945
Phosphate starvation inducible protein, PhoH	13139	02475	1909	08215
5'-nucleotidase (SurE) Stress condition	06290	00895	2964	03430
Periplasmatic 5'-nucleotidase		02085 02090	0643 0644 1276 1279 1282	08076 08081 05757 05742
Periplasmatic nuclease	00515 10450 11354	07334 08126	2348 2555 2606 3163	13971 01245 05907 05912 17755

Supplementary Table 3. Enzymes and proteins identified in the four Bacteroidetes involved in sulfur metabolism (Locus tags for each gene begins with MEDXXX_ for the MED strains and GFO_ for "*Gramella forsetii*", followed by the number in the table).

	Locus tag			
Gene/protein	**MED152**	**MED134**	**KT0803**	**MED217**
Sulphatase	00380	14677	0359 1698	01250
Sulphate symporter	09770	07966	3598	00620
Sulphate transporter	09030 12914	10061 14717	3419 2992	00290 13831
Sulphate adenyltransferase	06165-06170	14707-14712	0324-0325	13736-13741
Adenylyl-sulfate kinase	-	-	-	13746
(Phospho)adenylyl-sufate reductase	06175	14702	0323	02055
Adenylylsulfate reductase	-	-	-	-
Sulfite reductase (ferredoxin)	06160		0327	02060
DMSP metabolism	-	-	-	-

Supplementary Table 4. Number of Na^+ transporters identified in the four Bacteroidetes

	Number of ORFs			
Type	**MED152**	**MED134**	**KT0803**	**MED217**
ABC transporter	3	3	2	2
Na^+/H^+ exchanger	3	3	4	3
Na^+/H^+ antiporter:				
NhaA	-	-	1	1
NhaC	2	1	1	1
NhaB	-	-	-	1
Na^+/Ala symporter	1	3	4	3
Na^+/Ca^{+2} exchanger	1	1	2	2
Na^+/SO_4^{2} symporter	1	1	1	1
Na^+:solute symporter family protein (SSF)	2	3	4	6
Na^+/bile acid cotransporter (SBF)	2		1	1
Na^+/HCO_3 (SbtA)	1			
Na^+/HCO_3 (SulP type)	1	1	1	1
Electron chain (A-F subunits)	6	6	6	6
Na^+:dicarboxylate symporters(SDF)	1	1	1	2

Supplementary Table 5. Signal transduction proteins (one-component system) detected in the four Bacteroidetes (E-value ≤10^4)

	Number of ORFs			
Output domains	**MED152**	**MED134**	**KT0803**	**MED217**
HTH_AraC (AraC family protein)	11	12	12	17
GerE (LuxR family protein)	5	1	6	4
AsnC_trans_reg (AsnC/Lrp family protein)	3	5	5	5
GntR (GntR family protein)	2	1	1	1
Crp (Crp family protein)	1	1	2	3
HTH_1	0	0	1	1
HTH_3	8	8	13	15
HTH_5	1	2	3	3
HTH_8	1	1	1	1
HTH_11	1	0	0	1
LytTR	5	1	0	1
MarR (MarR family protein)	2	2	3	4
MerR (MerR family protein)	2	2	2	2
PadR	2	1	1	0
TetR_N (TetR family protein)	2	5	2	4
Trans_reg_C	0	1	0	0
HD	7	7	9	6
Guanylate_cyc	0	1	1	1
Rrf2	1	1	3	1
Input+Output domains				
LacI+Peripla_BP_1 (LacI family protein)	4	1	4	6
HTH_1+LysR LysR family protein	3	1	1	2
Crp+cNMP_binding	1	1	2	1
Fe_de_repress+Fe_dep_repr_C	1	1	1	1
HTH_11+DeoR	0	0	0	1
HTH_AraC+BLUF	1	0	0	0
PAS_3+GerE	0	0	0	1
Total	**64**	**56**	**73**	**81**

Supplementary Table 6. Predicted number of two-component histidine kinases and response regulators. Number is based on hits to specific Pfam profiles at the Microbial Signal Transduction Database (http://genomics.ornl.gov/mist/; E-value ≤10^4). Number per megabase is shown between parentheses.

Feature	**MED152**	**MED134**	**KT0803**	**MED217**
Response regulators by themselves	25 (9)	27 (8)	37 (10)	38 (9)
His kinases by themselves	21 (7)	23 (7)	25 (7)	27 (6)
Hybrid two component systems	5 (2)	5 (2)	7 (2)	11 (3)
Total His kinases	26 (9)	28 (9)	32 (9)	38 (9)
Total	**51**	**55**	**69**	**76**

Supplementary Table 7. Genes and domains with a potential role in adhesion (E-value $\leq 10^4$)

Domain	Number of ORFs			
	MED152	MED134	KT0803	MED217
Fasciclin	2	3	3	5
Thrombospondin type 3 repeat	1	2	2	1
Fibronectin typeIII	1	3	3	1
FG-GAP	4	4	3	7
Cadherin	0	0	1	0
HYR	2	1	4	0
Bac_surface_Ag	8	8	5	8
Big_1	0	0	0	0
Big_3	0	0	0	0
Dockerin_1	0	0	0	0
Cohesin	0	0	0	0
F5_F8_Type_C	3	1	1	5
He_PIG	0	0	1	0
Total	**21**	**22**	**22**	**27**

Supplementary Table 8. The twenty most abundant paralogous families of the four Bacteroidetes and two Proteobacteria

	Pfam model MED152	#	Pfam model MED134	#	Pfam model KT0803	#	Pfam model MED217	#	Pfam model E. *coli*	#	Pfam model S. *pomeroyi*	#
1	**Two component system**	**33**	**Two component system**	**30**	**Two component system**	**44**	**Two component system**	**54**	**ABC_transp**	**77**	**ABC_transp**	**116**
2	**TonBdep_Rec**	**24**	**ABC_transp**	**28**	**TonBdep_Rec**	**39**	**TonBdep_Rec**	**50**	**LysR_substrate**	**33**	**LysR_substrate**	**58**
3	**ABC_transp**	**22**	**TonBdep_Rec**	**18**	**ABC_transp**	**27**	**ABC_transp**	**30**	**Response_reg**	**27**	**adh_short**	**30**
4	adh_short	16	OmpA	13	adh_short	21	adh_short	19	EAL+GerE/MASE+GGDEF	26	Hypothetical	22
5	PKD domain	10	Peptidase_M36 (M14)	11	HlyD	19	SusD_RagB	19	Two component system	23	ECH/ 3HCDH_N	21
6	DEAD	8	Usp	11	SusD_RagB	13	rve	18	Fer4/ Pyr_redox_2	20	DAO/GCV_T	19
7	Sigma_70	6	Secreted protein (containing PKD domain)	10	Protein conteining hyalin/PKD domain	12	HlyD	17	Peripla_BP_1+ LacI	18	AMP_binding	19
8	Acyl_CoA_dh_1	6	Protein containing hyalin/cadherin domains	8	Sigma_70	12	ACR_tran	15	AA_permease	16	Two component system*	17
9	Membrane protein (sensor protein)	6	Aldedh	8	ACR_tran	11	Transposase_11	14	adh_short	16	DctM	17
10	Tranketolase	6	adh_short	8	hypothetical	10	Glyco_hydro_3	12	rve/ Pectactin	14	SBP_bac_5	17
11	GTP_EFTU	6	DEAD	8	Transposase_20	9	Glyco_hydro_43	12	Molybdoprotein	13	Acyl_CoA_dh_1	17
12	NADdependent aldehyde dh	6	Response_reg	7	Phage_integrase	9	Transposase_8	10	ADH_N	13	AnsC_trans_reg	16
13	PKD domain	6	Transket	7	Response_reg	9	Glyco_hydro_2_N	9	HlyD	12	DUF6	16
14	Cys_Met_Meta_PP	5	hypothetical	6	Glycos_transf_1	8	tonBdependent receptor	9	Usher	12	Aminotran_1_2	15
15	hypothetical	5	Peptidase_M56	6	OmpA	8	rve	8	Trasposase_11	11	BPD_transp_1	15
16	SusD_RagB	5	Epimerase	6	rve	7	OmpA	8	Aldedh	11	Ald_Xan_dh_C2	14
17	hypothetical	5	GTP_EFTU	6	MS_channel	7	Transposase_20	8	DeoR+ HTH_DeoR	11	Aldedh	13
18	HTH_AraC	5	Aminotran_1_2	5	DEAD	7	Transposase_8	8	BPD_transp_1	10	BPD_transp_1	11
19	SpoU_methylase	5	Sigma_70	5	E1_E2_ATPase	7	E7_E2_ATPase	7	BPD_transp_1	10	BPD_transp_1	11
20	PseudoU_synth_2	5	Aminotran_3	5	Pyr_redox_2	7	Response_reg	7	PTS_EIIC (EIIA1, EIIB)	10	HTH_AraC	10

Supplementary Table 9. Clusters of genes putatively involved in the attachment and degradation of polymeric compounds for the four Bacteroidetes. Only those that contain the tandem TonB dependent/ligand-gated channel genes and SusD are presented. Notice the amount of genes involved in degradation, adhesion, carbohydrate metabolism, secretion and transport within the clusters. Cells with a minus sign have no Pfam assigned at Evalue ≤ 10^4. Empty cells mean no information available.

ORF MED152	ORF MED152	Annotation	Pfam domain	Direction	SigP	TMHMM
00335	109	Alpha-amylase	Alpha-amylase	⇐	yes	0
00340	110	HP	-	⇐	yes	0
00345	111	HP	-	⇐	no	0
00350	112	Outer membrane protein	**SusD_RagD**	⇐	yes	0
00355	113	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
00360	114	Transcriptional regulator, LacI family	Peripla_BP_1	⇐	no	0
00365	115	Major Facilitator Superfamily	MSF_1	⇒	no	11
00370	116	ß-phosphoglucomutase	Hydrolase	⇒		0
00375	117	Trehalase/maltose phosphorylase	Glyco_hydro_65m	⇒		0
00380	118	N-acetylglycosamine-6-sulfatase	Sulfatase	⇒	no	0
00385	119	Alpha-amylase	Alpha-amylase	⇒		0
00390	120	Putative esterase	Esterase	⇒	yes	0
00395	121	Glycosyl hydrolase, family 31	Glyco_hydro_31	⇒	yes	0
00400	122	Alpha-amylase	Alpha-amylase	⇒	no	0
00405	123	Alpha-amylase	Alpha-amylase	⇒	yes	0
00415	124	Conserved HP	Y_Y_Y	⇒	yes	1
00420	125	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
00425	126	Outer membrane protein	**SusD_RagD**	⇒	yes	0
00430	127	Conserved HP	-	⇒	no	0
00435	128	HP	-	⇒	yes	0
00440	129	Conserved HP	-	⇒	yes	0
00445	130	Conserved HP	-	⇒		0
00450	131	Conserved HP	-	⇒	no	0
00455	132	Glycosyl hydrolase family 30	Glyco_hydro_30	⇒	yes	0
00460	133	Major Facilitator Superfamily	MSF_1	⇒	no	12
05095	1050	Major Facilitator Superfamily	MSF_1	⇐	no	12
05100	1051	Trehalase/maltose hydrolase	Glyco_hydro_65m	⇐	yes	0
05105	1052	HP	Cellulase	⇐	yes	0
05110	1053	Alpha-trehalase	Trehalase	⇐		0
05115	1054	Sugar transporter	MSF_1	⇐	no	12
05120	1055	Sucrose transporter	MSF_1	⇐	no	12
05125	1056	Conserved HP	-	⇐	no	0
05130	1057	Conserved HP	UnbV_ASPIC	⇐	yes	0
05135	1058	Conserved HP	UnbV_ASPIC	⇐	no	0
05140	1059	HP	-	⇐	yes	1
05145	1060	Conserved HP	UnbV_ASPIC	⇐	no	0
05150	1061	Outer membrane protein	**SusD_RagD**	⇐	yes	0
05155	1062	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
05160	1063	Conserved HP	DUF1080	⇐	yes	0
05165	1064	Conserved HP	AP_endonuc_2	⇐	no	0
05170	1065	Oxidoreductase	GFO_IDH_MocA	⇐	no	0

05175	1066	HP	Nuc_H_symport	⇐	no	12
05180	1067	Conserved HP	AP_endonuc_2	⇐	yes	0
05185	1068	Transcriptional regulator, AraC	HTH_AraC	⇐		0
09760	1963	S-adenosylhomocysteine hydrolase	AdoHcyase	⇒		0
09765	1964	Putative sulfite reductase	-	⇒	yes	4
09770	1965	Putative sodium/sulphate symporter	Na_sulph_symp	⇒	no	13
09780	1966	Putative alginate lyase precursor	-	⇒	no	0
09785	1967	Putative alginate lyase precursor	-	⇒	yes	0
09790	1968	Gluconate lyase, SKI family	SKI	⇒		0
09795	1969	6-phosphogluconate dehydrogenase	6PGD	⇒	no	0
09800	1970	Putative Mn+2 Fe+2 transporter	Nramp	⇒	yes	11
09805	1971	Putative alginate lyase precursor	-	⇒	no	0
09810	1972	HP	adh_short	⇒	no	0
09815	1973	HP	-	⇒	yes	0
09820	1974	HP	Hepar_II_III	⇒	no	0
09825	1975	HP	Cupin_2	⇒		0
09830	1976	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
09835	1977	Outer membrane protein	**SusD_RagD**	⇒	no	0
09840	1978	HP	-	⇒	no	0
09845	1979	HP	PKD	⇒	yes	1
09850	1980	Transcriptional regulator, GntR	FCD	⇒		0
09855	1981	Major Facilitator Superfamily protein	MSF_1	⇒	no	11
09860	1982	Short-chain dehydrogenase/reductase	adh_short	⇒	no	0
09865	1983	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
09870	1984	Outer membrane protein	**SusD_RagD**	⇒	yes	0
12434	2482	Putative tRNA/rRNA methyltransferase	SpoU_methylase	⇐		0
12439	2483	HP	-	⇐	no	0
12444	2484	Outer membrane protein	**SusD_RagD**	⇐	yes	0
12449	2485	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
12454	2486	HP	-	⇐	no	0
ORF MED134	**ORF MED134**	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
05189	c1_1083	HP	-	⇐		
05199	c1_1084	HP	Peptidase_M56	⇐	no	4
05204	c1_1085	Transcriptional regulator	Penicillinase_R	⇐		0
05209	c1_1086	Putative hydrolase	Peptidase_M20	⇐	yes	0
05214	c1_1087	Outer membrane protein	**SusD_RagD**	⇐	no	0
05219	c1_1088	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
05224	c1_1089	Putative hydrolase	Peptidase_M20	⇐	yes	0
05229	c1_1090	Probable aminopeptidase	Peptidase_M28	⇐	yes	0
05234	c1_1091	HP	DUF1083	⇐	yes	0
05239	c1_1092	HP	-	⇐	yes	1
05244	c1_1093	HP	-	⇐	yes	0
05249	c1_1094	HP	-	⇐	no	2

05189	c1_1095	UvrABC system protein A	ABC_tran	⇐	no	0
07436	c2_32	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
07441	c2_33	Outer membrane protein	**SusD_RagD**	⇒	yes	0
07446	c2_34	Putative multidrug resistance protein	ACR_tran	⇐	no	14
07451	c2_35	Putative HlyD-like secretion protein	HlyD	⇐	yes	0
07456	c2_36	Putative outer membrane efflux protein	OEP	⇐	yes	0
07461	c2_37	Transcriptional regulator, TetR	TetR_N	⇐		0
10590	c2_668	HP	-	⇐	yes	0
10595	c2_668	HP	-	⇐	no	0
10600	c2_670	HP	MS_channel	⇐	no	5
10605	c2_671	HP	Peptidase_M56	⇐	yes	3
10610	c2_672	Putative antibiotic resistance-related regulatory protein	Penicillinase_R	⇐		0
10615	c2_673	Putative succinyl-diaminopimelate desuccinylase	Peptidase_M20	⇐		0
10620	c2_674	2-keto-3-deoxy-6-phosphogluconate aldolase	Aldolase	⇐		0
10625	c2_675	2-dehydro-3-deoxygluconokinase	Pfk	⇐	no	0
10630	c2_676	Short-chain dehydrogenase/reductase	adh_short	⇐	yes	0
10635	c2_677	Putative hexuronate transport protein	MSF_1	⇐	no	11
10640	c2_678	Transcriptional regulator GntR family	FCD	⇐		0
10645	c2_679	Putative lyase	-	⇐	yes	0
10650	c2_680	Cell surface protein	PKD	⇐	yes	0
10655	c2_681	Outer membrane protein	**SusD_RagD**	⇐	yes	0
10660	c2_682	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
10665	c2_683	Putative pectin degradation protein	Cupin_2	⇐		0
10670	c2_684	Putative chondroitin AC/alginate lyase	-	⇐	no	0
10675	c2_685	Putative chondroitin AC/alginate lyase	-	⇐	yes	0
10680	c2_686	Putative chondroitin AC/alginate lyase	-	⇐	yes	0
10685	c2_687	HP	-	⇐	no	0
10735	c2_696	Putative two-component system sensor (hydrid)	His_kinase	⇒	no	4
10740	c2_697	Response regulator receiver protein	Response_reg	⇒		0
10745	c2_698	HP	-	⇒	yes	0
10750	c2_699	Putative carbohydrate kinase	PfkB	⇐		0
10755	c2_700	Xylose transporter	Sugar_tr	⇐	no	12
10760	c2_701	Glycoside hydrolase, family 32	Glyco_hydro_32N	⇐	no	0
10765	c2_702	Transcriptonal regulator, LacI	Peripla_BP_1	⇒	no	0
10770	c2_703	Beta-phosphoglucomutase	Hydrolase	⇒		0
10775	c2_704	Trehalase/maltose hydrolase	Glyco_hydro_65m	⇒		0
10780	c2_705	HP	-	⇐	no	0
10785	c2_706	HP	Pentapeptide	⇐		0
10790	c2_707	HP	AraC_E_bind	⇐	no	1

10795	c2_708	HP	-	⇐	yes	0
10800	c2_709	ASPIC-like protein	UnbV_ASPIC	⇐	no	0
10805	c2_710	ASPIC-like protein	UnbV_ASPIC	⇐	no	0
10810	c2_711	ASPIC-like protein	UnbV_ASPIC	⇐	yes	0
10815	c2_712	Outer membrane protein	**SusD_RagD**	⇐	yes	0
10820	c2_713	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
ORF KT0803	**ORF KT0803**	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
0008	9	Major Facilitator Superfamily permease	Sugar_tr	⇒	yes	12
0009	10	Glycosyl hydrolase, family 32	Glyco_hydro_32N	⇒	yes	0
0010	11	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
0011	12	Outer membrane protein	**SusD_RagB**	⇒	yes	0
0012	13	Carbohydrate kinase	PfkB	⇒	no	0
0013	14	Multidrug resistance protein	Multi_Drug_Res	⇒	no	0
0014	15	Metal-dependent membrane protease	Abi	⇒	no	8
0015	16	Hyaluronan synthase	Glycos_transf_2	⇐		5
0016	17	HP	-	⇐		0
0017	18	Membrane protein	-	⇒	no	2
0018	19	Secreted protein	Guanylate_cyc	⇐	yes	1
0019	20	Secreted protein	-	⇐	yes	0
0020	21	Major Facilitator Superfamily permease	MSF_1	⇒	no	11
0021	22	Secreted protein	-	⇒	yes	1
0022	23	Conserved HP	-	⇐	yes	1
0023	24	Secreted protein	-	⇐	yes	0
0024	25	Heavy metal-(Cd/Co/Hg/Pb/Zn)-translocating P-type ATPase	E1_E2_ATPase	⇐		8
0025	26	Secreted protein	Cation_efflux	⇐	no	5
0026	27	Conserved HP	-	⇐		0
0027	28	Fur family transcriptional regulator protein	-	⇐		0
0028	29	HlyD family secretion protein	-	⇐	no	0
0029	30	Heavy metal cation efflux protein	ACR_tran	⇐	no	13
0030	31	HP	-	⇐	no	1
0031	32	Beta-galatosidase	Glyco_hydro_2_N	⇐	yes	0
0032	33	Arabinogalactan 1,4-beta-galactosidase	Glyco_hydro_53	⇐	yes	0
0033	34	Secreted protein	-	⇐	yes	0
0034	35	Outer membrane protein	**SusD_RagB**	⇐	no	0
0035	36	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
0036	37	Two-component system sensor (hydrid)	HATPase_c	⇐	yes	1
0037	38	Membrane protein	-	⇒	no	4
0038	39	Short-chain dehydrogenase/reductase	adh_short	⇒	no	0
0360	357	Sensor/regulator hydrid	GerE	⇒	yes	1
0361	358	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
0362	359	Outer membrane protein	**SusD_RagB**	⇒	no	0
0363	360	Conserved HP	-	⇒	yes	0
0364	361	Conserved HP	-	⇒	no	0
0365	362	Secreted alpha/beta fold	Abhydrolase_2	⇒	yes	1

		hydrolase				
0366	363	Fibronectin III containing domain GH43	F5_F8_type_C	⇐	yes	0
0367	364	Glycosyl hydrolase family 43	Glyco_hydro_3	⇒	yes	1
0368	365	Glycosyl hydrolase family 2	Glyco_hydro_2_N	⇐	yes	0
0369	366	Secreted protein	-	⇐	yes	0
0370	367	Secreted lipase/esterase	-	⇐	yes	0
0688	678	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
0689	679	Outer membrane protein	**SusD_RagB**	⇒	no	0
0690	680	Alpha-L-arabinofuranosidase	Alpha_L_AF_C	⇒	yes	0
0691	681	Arabian endo-1,5-L-arabinosidase	Glyco_hydro_43	⇒	yes	0
0692	682	Conserved HP, membrane or secreted	DUF1680	⇒	yes	1
0693	683	Glycosyl hydrolase, family 43	Glyco_hydro_43	⇒	no	0
0694	684	Arabian endo-1,5-L-arabinosidase	Glyco_hydro_43	⇒	yes	0
0695	685	Alpha-L-arabinofuranosidase	Glyco_hydro_43	⇒	yes	0
0696	686	HP	-	⇒		0
0697	687	Ribulokinase	FGGY_C	⇒		0
0698	688	L-ribulose-5-phsopate 4-epimerase	Aldolase_II	⇒	no	0
0699	689	L-arabinose isomerase	Arabinose_Isome	⇒	no	0
0700	690	Aldose 1-epimerase	Aldose_epim	⇒	no	0
0701	691	Sodium:solute symporter family protein	SSF	⇒	no	14
0702	692	Transaldolase	Transaldolase	⇒	no	0
0703	693	Transketolase N-terminal	Transketolase_N	⇒		0
0704	694	Transketolase C-terminal	Trasnket_pyr	⇒		0
0705	695	Alpha-glucosidase	-	⇒	yes	0
0706	696	Protein containing tetratricopeptide repeats	TRP_2	⇐	no	0
0707	697	HP	-	⇐		1
0708	698	Outer membrane efflux protein	OEP	⇐	no	0
0709	699	ABC-type transporter ATP-binding protein	ABC_tran	⇐	no	0
0710	700	Predicted permease	FtsX	⇐	yes	8
0711	701	Conserved HP	-	⇐	yes	0
0712	702	Predicted permease	FtsX	⇐	no	8
0713	703	Predicted permease	FtsX	⇐	yes	8
0714	704	Predicted permease	FtsX	⇐	no	8
0715	705	Predicted permease	FtsX	⇐	yes	8
0716	706	ABC-type transporter ATP-binding protein	ABC_tran	⇐		0
0717	707	HlyD family secretion protein	-	⇐	yes	1
0718	708	Two-component system response regulator	Sigma54_activat	⇒	no	0
0719	709	Two-component system sensor histidine kinase	HATPase_c	⇒	yes	2
1027	1014	Glucoside 3-dehydrogenase	-	⇒	no	0
1028	1015	Conserved HP	-	⇒	no	0
1029	1016	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1030	1017	Outer membrane protein	**SusD_RagB**	⇒	no	0
1031	1018	Secreted protein	UnbV_ASPIC	⇒	yes	0
1032	1019	Secreted protein	UnbV_ASPIC	⇒	no	1
1033	1020	HP	-	⇒	no	0

1034	1021	HP	AraC_E_bind	⇒	no	1
1148	1135	Secreted alginate lyase-like protein	-	⇒	yes	0
1149	1136	HP	-	⇒	no	0
1150	1137	Secreted protein	-	⇒	yes	0
1151	1138	Secreted protein	Hepar_II_III	⇒		0
1152	1139	Pectin degradation protein	Cupin_2	⇒		0
1153	1140	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
1154	1141	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1155	1142	Protein containing PDK domain	PKD	⇒	no	0
1156	1143	Secreted alginate lyase-like protein	-	⇒	yes	0
1157	1144	Secreted alginate lyase-like protein	-	⇒	yes	0
1158	1145	Trancriptional regulator, GntR	FCD	⇒		0
1159	1146	Major Facilitator Superfamily permease	MSF_1	⇒	no	11
1160	1147	Short-chain dehydrogenase/reductase	adh_short	⇒	no	0
1161	1148	2-dehydro-3-deoxygluconokinase	PfkB	⇒		0
1162	1149	Aldolase	Aldolase	⇒	no	0
1163	1150	Fructose-1,6-bisphosphatse	FBPase	⇒	no	0
1164	1151	Transcriptional regulator, LacI	Peripla_BP_1	⇒	no	0
1165	1152	Phage integrase protein	Phage_integrase	⇒	no	0
1166	1153	Secreted protein	-	⇒	yes	0
1167	1154	Membrane protein	DUF305	⇒	no	3
1254	1241	FerR family protein	FecR	⇒	no	1
1255	1242	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
1256	1243	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1257	1244	Transmembrane spermine/spermidine synthase	Spermine_synth	⇒	no	13
1258	1245	HP	-	⇒	no	0
1259	1246	HP	-	⇒		0
1260	1247	RNA polymerase ECF-type sigma factor	Sigma70_r2	⇒		0
1261	1248	FerR family protein	FecR	⇒		1
1262	1249	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
1263	1250	Outer membrane protein	**SusD_RagB**	⇒	yes	0
1264	1251	Conserved HP, secreted	-	⇒	yes	0
1265	1252	Secreted protein	-	⇒	no	2
1266	1253	Conserved HP, membrane	DUF81	⇒	no	8
1380	1367	Peptidase family 16	Peptidase_M16_C	⇐	no	0
1381	1368	Outer membrane protein	**SusD_RagB**	⇐	no	0
1382	1369	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
1702	1686	L-fucose permease	MSF_1	⇐	no	12
1703	1687	Short-chain dehydrogenase/reductase	adh_short	⇐	no	0
1704	1688	Amidohydrolase family protein	Aminohydro_2	⇐		0
1705	1689	Altronate oxidoreductase	Mannitol_dh_C	⇐		0
1706	1690	Altronate hydrolase	GD_AH_C	⇐		0

1707	1691	Aldolase	Aldolase	⇐		0
1708	1692	2-dehydro-3-deoxyluconokinase	PfkB	⇐		0
1709	1693	Zinc-type alcohol dehydrogenase	ADH_N	⇐		0
1710	1694	Alpha-L-fucosidase	Alpha_L_fucos	⇐	yes	0
1711	1695	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
1712	1696	Outer membrane protein	**SusD_RagB**	⇒	no	0
1713	1697	Protein containing PDK domains	PKD	⇒	no	1
1714	1698	Secreted glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
1715	1699	Membrane or secreted glycosyl hydrolase	CMB_4_9	⇒	yes	0
1716	1700	Phytanoyl-CoA dioxygenase family protein	PhyH	⇒		0
1717	1701	Zinc-type alcohol dehydrogenase	ADH_N	⇒		0
1718	1702	Acetylesterase	Esterase	⇒		0
1719	1703	Secreted glycosyl hydrolase, family 2	Glyco_hydro_2_N	⇒	yes	1
1720	1704	Secreted protein	-	⇒	yes	0
1721	1705	Secreted protein	DUF1080	⇒	yes	0
1722	1706	Nucleotidyltransferase family protein	NTP_transf_2	⇒		0
1723	1707	HP	-	⇒		0
1724	1708	Conserved HP	-	⇒	no	0
1725	1709	Glycosyl hydrolase, family 16	CMB_4_9	⇐	no	0
1726	1710	HP	-	⇐	no	0
1727	1711	Membrane or secreted glycosyl hydrolase	Glyco_hydro_2_N	⇐	yes	0
1728	1712	Secreted glycosyl hydrolase, family 16	Glyco_hydro_16	⇐	yes	0
1729	1713	HP	-	⇒		0
1730	1714	Secreted beta-lactamase family protein	Beta-lactamase	⇒	yes	0
1731	1715	Conserved HP	-	⇐		0
1732	1716	Outer membrane protein	**SusD_RagB**	⇐	no	0
1733	1717	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
1734	1718	FecR family protein	FecR	⇐		1
1735	1719	RNA polymerase ECF-type sigma factor	Sigma70_r4_2	⇐		1
1736	1720	Periplasmatic trehalase-like protein	Trehalase	⇒	yes	0
1737	1721	Aminoglycoside phospho trasferase	Aminotran_3	⇒		0
1738	1722	Beta-galactosidase	Glyco_hydro_2_C	⇒	yes	0
2030	2111	Glyceraldehyde-3-phosphate dehydrogenase A	Gp_dh_C	⇐	no	0
2031	2112	6-phosphofructokinase	PFK	⇐	no	0
2032	2113	Alpha amylase	Alpha_amylase	⇐	no	0
2033	2114	Alpha amylase	Alpha_amylase	⇐	yes	0
2034	2115	Glycoside hydrolase, family 65	Glyco_hydro_65m	⇐		0
2035	2116	Beta-phosphoglucomutase	Hydrolase	⇒	no	0
2036	2117	Major Facilitator Superfamily permease	MSF_1	⇒	no	12
2037	2118	Transcriptional regulator, LacI	Peripla_BP_1	⇒	yes	0

2038	2119	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
2039	2120	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2040	2121	Conserved HP	-	⇒	yes	0
2041	2122	Alpha amylase	Alpha_amylase	⇒	yes	0
2042	2123	Ribonucleotide reductase large subunit	Ribonuc_red_IgC	⇒		0
2844	2814	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	0
2845	2815	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2846	2816	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	4
2847	2817	Conserved HP	-	⇒	yes	0
2848	2818	Secreted protein	-	⇒	yes	0
2849	2819	HP	-	⇒	no	1
2850	2820	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
2851	2821	Outer membrane protein	**SusD_RagB**	⇒	yes	0
2918	2883	HP	-	⇐		0
2919	2884	Conserved HP	-	⇐	yes	0
2920	2885	HP	-	⇐	yes	0
2921	2886	Outer membrane protein	**SusD_RagB**	⇐	no	0
2922	2887	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	1
3462	3418	Two-component system senso (hydrid)	Y_Y_Y	⇒	yes	1
3463	3419	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	yes	1
3464	3420	Outer membrane protein	**SusD_RagB**	⇒	yes	1
3465	3421	Conserved HP	-	⇒	no	0
3466	3422	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	1
3467	3423	Beta-glucosidase	Glyco_hydro_3_C	⇒	yes	0
3468	3424	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
3578	3533	Gliding motility protein	DUF21	⇒	no	3
3579	3534	Gliding motility protein	-	⇒	yes	0
3580	3535	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
3581	3536	Outer membrane protein	**SusD_RagB**	⇒	yes	0
3582	3537	Membrane protein	DUF6	⇐	no	10
3583	3538	Secreted peptidase, family M16	Peptidase_M16_C	⇐	yes	0
3584	3539	Secreted peptidase, family M16	Peptidase_M16_C	⇐	yes	0
ORF MED217	**ORF MED217**	**Annotation**	**Pfam domain**	**Direction**	**SigP**	**TMHMM**
01240	284	Dipeptidyl peptidase 4	Peptidase_S9	⇐		1
01245	285	Endonuclease	Exo_endo_phos	⇐		0
01250	286	N-acetylglucosamien-6-sulfatase	Sulfatase	⇐		0
01255	287	Alpha/beta hydrolase fold	-	⇐	yes	0
01260	288	Glycosyl hydrolase, family 3	Glyco_hydro_3	⇐	yes	0
01265	289	Carboxyesterase	Abhydrolase_2	⇐		0
01270	290	HP	-	⇐		0
01275	291	HP	-	⇐		0
01280	292	Outer membrane protein	**SusD_RagB**	⇐		0
01285	293	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0

01545	344	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
01550	345	Outer membrane protein	**SusD_RagB**	⇒		0
01555	346	Alkaline phosphatase	Alk_phosphatase	⇒	yes	0
01560	347	Alkaline phosphatase	Alk_phosphatase	⇒		0
01565	348	Endo-1,4-beta-xylanase	Glyco_hydro_43	⇐	yes	0
01570	349	TonB dependent receptor	**TonB_dep_Reg**	⇒	yes	0
01575	350	HP	-	⇒		0
01580	351	Two-component sensor histidine kinase	HATPase_c	⇐		2
01585	352	Transcriptional regulator	Sigma54_activat	⇐		0
01590	353	HP	-	⇒		1
01595	354	ABC transporter, APT-binding protein	ABC_tran	⇒		0
01600	355	Predicted permease	FtsX	⇒		8
01605	356	Predicted permease	FtsX	⇒	yes	8
01610	357	Predicted permease	FtsX	⇒		8
01615	358	Predicted permease	FtsX	⇒	yes	8
01620	359	Predicted permease	FtsX	⇒		8
01625	360	Predicted permease	FtsX	⇒	yes	8
01630	361	ABC transporter, AT-binding protein	ABC_tran	⇒		0
01635	362	Outer membrane efflux protein	OEP	⇒		0
01640	363	Pirin family protein	Pirin	⇒		0
01645	364	Bacterial regulatory protein	GerE	⇒	no	0
01650	365	HP	Peptidase_S41	⇒		0
01655	366	Two-component sensor histidine kinase	HATPase_c	⇒		1
02945	614	ASPIC-like protein	UnbV_ASPIC	⇐		0
02950	615	ASPIC-like protein	UnbV_ASPIC	⇐		0
02955	616	ASPIC-like protein	UnbV_ASPIC	⇐		0
02960	617	Outer membrane protein	**SusD_RagB**	⇐	no	0
02965	618	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
03285	683	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
03290	684	Outer membrane protein	**SusD_RagB**	⇒		0
03295	685	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	0
03300	686	Outer membrane protein	**SusD_RagB**	⇒	no	0
03305	687	HP	**-**	⇒	0	0
04542	922	HP	DUF718	⇐		0
04547	923	HP	-	⇐		0
04552	924	Sulfatase	Sulfatase	⇐		0
04557	925	Outer membrane protein	**SusD_RagB**	⇐		0
04562	926	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
04567	927	Transforming growth factor	Fasciclin	⇐		0
04572	928	Transforming growth factor	Fasciclin	⇐		0
04577	929	Outer membrane protein	**SusD_RagB**	⇐	yes	1
04582	930	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	no	0
04587	931	Transforming growth factor	Fasciclin	⇐		
05257	1069	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
05262	1070	Outer membrane protein	**SusD_RagB**	⇒	yes	0

05267	1071	HP	-	⇒		0
05272	1072	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒		0
05277	1073	HP	-	⇒		0
05282	1075	Beta-glucosidase	Glyco_hydro_3C	⇒		0
05287	1076	Glycosyl hydrolase, family 16	Glyco_hydro_16	⇒	yes	0
05902	1193	Endonuclease	Exo_endo_phos	⇐		0
05907	1194	Endonuclease	Exo_endo_phos	⇐		0
05912	1195	Outer membrane protein	**SusD_RagB**	⇐		0
05917	1196	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
05922	1197	Two-component system sensor histidine kinase	HATPase_c	⇐		1
05927	1198	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
05932	1199	Outer membrane protein	**SusD_RagB**	⇒		0
05937	1200	HP	**-**	⇒		0
05942	1201	HP	**-**	⇒		0
05947	1202	Xylosidase	Glyco_hydro_43	⇒	no	0
06606	1327	Alpha/beta hydrolase fold	Abhydrolase_1	⇒		8
06611	1328	Bacterial regulatory family LuxR protein	GerE	⇒		0
06616	1329	Probable TonB-dependent outer membrane receptor	-	⇒		0
06621	1330	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒	no	1
06626	1331	Outer membrane protein	**SusD_RagB**	⇒		0
06631	1332	Carboxyesterase	COesterase	⇒		0
06636	1333	Alpha-L-rhamnosidase	Bac_rhamnosid	⇒		0
06641	1334	Two-component system response regulator	Response_reg	⇒		0
06646	1335	Two-component system sensor histidine kinase	His_kinase	⇐	no	2
06651	1336	Phosphatase	Metallophos	⇐		1
06656	1337	Beta-glucuronidase	Glyco_hydro_2N	⇒		1
06661	1338	Hydrolase	Lactamase_B	⇐		0
07701	1543	Outer membrane protein	**SusD_RagB**	⇐		0
07706	1544	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
07711	1545	Putative beta-xylosidase	Glyco_hydro_43	⇐		0
07716	1546	Two-component system sensor histidine kinase	HATPase_c	⇒		0
07721	1547	HP	Glyco_hydro_43	⇐		0
07726	1548	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
07731	1549	Outer membrane protein	**SusD_RagB**	⇐		0
07736	1550	HP	-	⇐		0
07741	1551	HP	-	⇐	yes	0
07746	1552	HP	-	⇐		0
07751	1553	HP	-	⇐		0
07756	1554	Beta-galactosidase	Glyco_hydro_2C	⇐		0
07761	1555	Alpha-galactosidase	-	⇐	yes	0
07766	1556	Putative secreted hydrolase	Glyco_hydro_16	⇐		0
07771	1557	Putative glycosyl hydrolase	Glyco_hydro_88	⇐	yes	0
07776	1558	HP	-	⇐		0
07781	1559	Beta-galactosidase	Glyco_hydro_2C	⇐		0
07786	1560	HP	-	⇐		1
07791	1561	Aldo/keto reductase family	Aldo_ket_red	⇐	no	0

		protein				
07796	1562	Endo-1,4-beta-xylanase A	-	⇒	no	0
07801	1563	Putative membrane protein	MSF_1	⇒		12
07851	1572	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
07856	1573	Outer membrane protein	**SusD_RagB**	⇒	no	0
09957	1998	HP	**-**	⇐	no	0
09962	1999	HP	**-**	⇐	yes	0
09967	2000	Outer membrane protein	**SusD_RagB**	⇐		0
09972	2001	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
09977	2002	HP	-	⇐		0
09982	2003	Two-component system (hybrid)	HATPase_c	⇐	no	1
11554	2318	Aldolase 1-epimerase	Aldolase_epim	⇒		0
11559	2319	L-arabinose isomerase	Arabinose_isome	⇐		0
11564	2320	Sodium/glucose contransporter 1	SSF	⇐	no	14
11569	2321	L-ribulose-5-phosphate 4-epimerase	Aldolase_II	⇐		0
11574	2322	L-ribulokinase	FGGY_C	⇐		0
11579	2323	Alpha-N-arabinofuranosidase A	Alpha_L_AF_C	⇐		0
11584	2324	HP	DUF1680	⇐		0
11589	2325	Arabinan endo-1,5-alpha-L-arabinosidase A	Glyco_hydro_43	⇐		0
11594	2326	Xylosidase/arabinosidase	Glyco_hydro_44	⇐	yes	0
11599	2327	Arabinan endo-1,5-alpha-L-arabinosidase A	Glyco_hydro_43	⇐		0
11604	2328	Alpha-L-arabinosidase	Alpha_L_AF_C	⇐		0
11609	2329	ASPIC-like protein	UnbV_ASPIC	⇐		0
11614	2330	Outer membrane protein	**SusD_RagB**	⇐	no	0
11619	2331	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
12894	2587	Mannonate dehydratase	UxuA	⇐		0
12899	2588	Short-chain dehydrogenase/reductase	adh_short	⇐		0
12904	2589	Outer membrane protein	**SusD_RagB**	⇐		0
12909	2590	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	yes	0
12914	2591	Trascriptional regulator, LacI	Peripla_BP_1	⇐		0
12919	2592	Hypotethical oxidoreductase	GFO_IDH_MocA	⇐		0
12924	2593	HP	Aminohydro_2	⇐		0
12929	2594	Two-component system sensor histidine kinase	HATPase_c	⇒	no	2
12934	2595	Probable transcriptional regulator	Response_reg	⇒		0
12939	2596	Xylosidase/arabinosidase	Glyco_hydro_43	⇒		0
13289	2665	Glucokinase	ROK	⇐		1
13294	2666	Esterase	Abhydrolase_3	⇐		0
13299	2667	Putative alpha-glucosidase	-	⇐		0
13304	2668	HP	Glyco_hydro_92	⇐	yes	0
13309	2669	HP	DUF1237	⇐		0
13314	2671	HP	F5_F8_type_C	⇐	no	0
13319	2672	Outer membrane protein	**SusD_RagB**	⇐		0

13324	2673	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		0
(…)	(…)	HP	-	⇐		
13364	2681	Putative alpha-1,2-mannosidase	Glyco_hydro_92	⇐		0
13369	2682	L-fucose-proton symporter	MSF_1	⇐		12
13379	2683	Putative esterase	Esterase	⇐		1
13384	2684	Arylsufatase	Sulfatase	⇐	no	0
13394	2685	Beta-galactosidase	Glyco_hydro_2	⇐		0
13399	2686	Beta-galactosidase	Glyco_hydro_2_N	⇐	0	
13404	2687	Two-component system (hybrid)	HATPase_c	⇒	yes	1
13409	2688	Rhamnogalacturonan acetylesterase precursor	Lipase_GDSL	⇒	yes	0
13414	2689	Beta-galactosidase	Glyco_hydro_2_N	⇐		1
13419	2690	HP	Lipase_GDSL	⇐		0
13424	2691	Acetylesterase	DUF303	⇐		0
13429	2692	Arabinose metabolism trascriptional repressor	GntR	⇐		0
13434	2693	Hypothetical oxidoreductase	adh_short	⇒		0
13439	2694	Putative sugar isomerase	-	⇒		0
13444	2695	Glycerol kinase	FGGY_N	⇒		0
13449	2696	L-lactate dehydrogenase	FMN_dh	⇒		0
13454	2697	Transmembrane protein	-	⇒		11
13459	2698	Exo-poly-alpha-D-galacturonidase precurson	Glyco_hydro_28	⇒		0
13464	2699	HP	-	⇒		1
13469	2700	HP	-	⇒	yes	0
13474	2701	HP	-	⇒		0
13479	2702	HP	-	⇒		0
13484	2703	Arabinosidase	-	⇐		0
13489	2704	Xylanase	Peptidase_S9	⇐	yes	1
13494	2705	Beta-galactosidase	Glyco_hydro_2_N	⇐		0
13499	2706	Glycosyl hydrolase, family 88	Glyco_hydro_88	⇐		0
13504	2707	Conserved HP	Glyco_hydro_2_N	⇐	yes	0
13509	2708	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇒		0
13514	2709	Outer membrane protein	**SusD_RagB**	⇒		0
13519	2710	Two-component system sensor histidine kinase	HATPase_c	⇒		0
13524	2711	HP	-	⇒		0
13529	2712	HP	-	⇒		0
13536	2714	HP	Response_reg	⇐		0
13541	2715	Two-component system sensor histidine kinase	HATPase_c	⇐		0
13546	2716	HP	-	⇐		0
13551	2717	Aminotransferase	Aminotran_1_2	⇐		0
13556	2718	HP	-	⇒		0
13561	2719	Transcriptional regulator, AraC	HTH_AraC	⇒		0
13566	2720	HP	-	⇒		5
13571	2721	HP	-	⇒		1
13576	2722	HP	Peptidase_S9	⇐		0
13581	2723	Outer membrane protein	**SusD_RagB**	⇐	no	0
13586	2724	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐		1
17130	3416	Ribonucleoside-diphosphate reductase alpha chain	Ribonuc_red_IgC	⇐		0
17135	3417	Ribonucleoside-diphosphate reductase beta chain	ATP-cone	⇐	no	0

17145	3418	Alpha-glucosidase	Glyco_hydro_31	⇐		0
17150	3419	Alpha-amylase	Alpha_amylase	⇐	yes	0
17155	3420	Alpha-amylase	Alpha_amylase	⇐	yes	0
17160	3421	HP	-	⇐		0
17165	3422	Outer membrane protein	**SusD_RagB**	⇐		0
17170	3423	TonB-dependent outer membrane receptor	**TonB_dep_Reg**	⇐	no	1
17175	3424	Serine/threonine-protein kinase	-	⇐		1

64

65

Supplementary Table 10. Genes shared between each pair of the four Bacteroidetes genomes analyzed based on reciprocal best matches.

	KT0803	**MED134**	**MED152**
MED217	2122	1985	1741
KT0803		1888	1692
MED134			1762

Supplementary Table 11. Bacteria in which a PR coding gene has been identified and distribution of particular light-sensing domains (a black square means presence, a minus sign absence). BLUF: Blue-light sensing using flavin; Pchr: Phytochrome; AnCchr: Animal cryptochrome and (6-4) photolyase family; DASH: Cryptochrome DASH family; Photoly: Photolyase class I.

Organism	**BLUF**	**Pchr**	**AnCchr**	**DASH**	**Photoly**
		Bacteroidetes			
Polaribacter sp. MED152	■	■	■	■	■
Polaribacter irgensii 23-P	-	-	■■	■	■
Dokdonia sp. MED134	■	■	■	■	■
F. bacterium BAL38	-	-	■	■	■
		Alphaproteobacteria			
Ca. "P. ubique" HTCC1002	-	-	-	-	■
Ca. "P. ubique" HTCC1062	-	-	-	-	■
Octadecabacter antarcticus 307	-	-	■	-	■
		Betaproteobacteria			
Methylophilales HTCC2181	-	-	-	■	■
		Gammaproteobacteria			
Vibrio campbelli ATCC BAA-1116	-	-	■	■	■
Vibrio sp. S14	-	-	■	■	■
HTCC2143	-	-	■	■	■
HTCC2207	-	-	-	■	■
Photobacterium sp. SKA34	-	-	■	■	■
Marinobacter sp. ELB17	-	-	■■	■	■

Supplementary Figure 1. Genes identified in section Z of the *Leeuwenhoekiella blandensis* MED217 genome. They include the structural genes of the arabinose operon (red), plus other genes involved in sugar metabolism (blue and green) and transport (*yidK*).

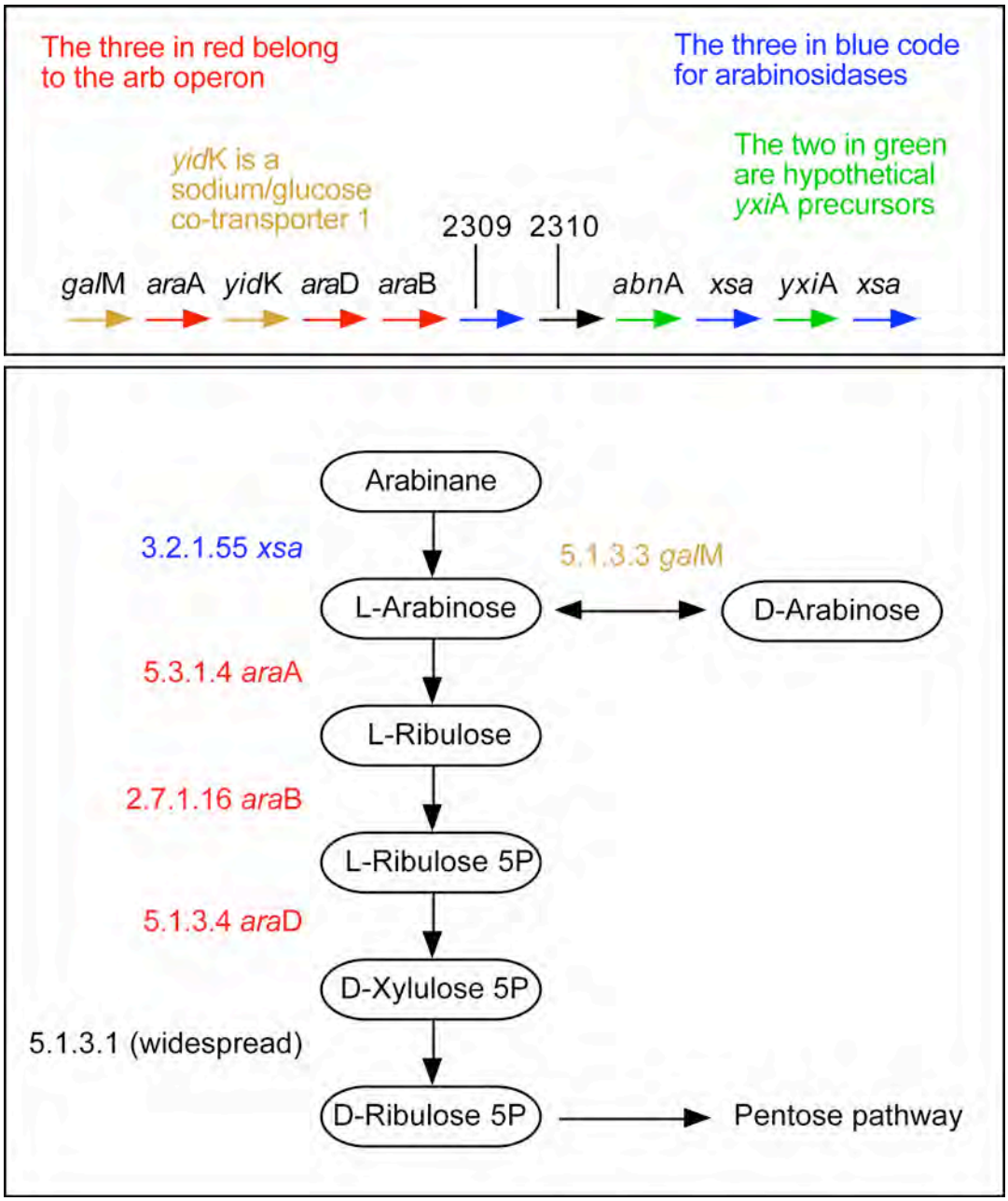
Supplementary Figure 2. A. Number of peptidases versus genome size. B. Peptidases per Mb of genome. C. Glycosyl hydrolases versus genome size. D. Glycosyl hydrolases per Mb. PR+ Bacteroidetes are in purple, PR- Bacteroidetes in blue.

Supplementary Figure 3. Diversity of peptidases (A) and glycoside hydrolases (B) in the four marine Bacteroidetes. Bray-Curtis distances among the peptidases and GHs of each bacterium were calculated. Interestingly, the distances were larger for the GHs (0.33 to 0.68) than for the proteases (0.20 to 0.27). For proteases, the four bacteria were essentially equidistant. For the GHs, MED152 and MED134, with the smaller genomes, were more similar among them (0.33 between them versus 0.54 to 0.68 with the other two bacteria), while MED217 and "G. forsetii" KT0803, with larger genomes, were closer between them (0.40) than to the other two bacteria. Therefore, despite having more and more diverse proteases than GHs, the four bacteria differed more in the latter.

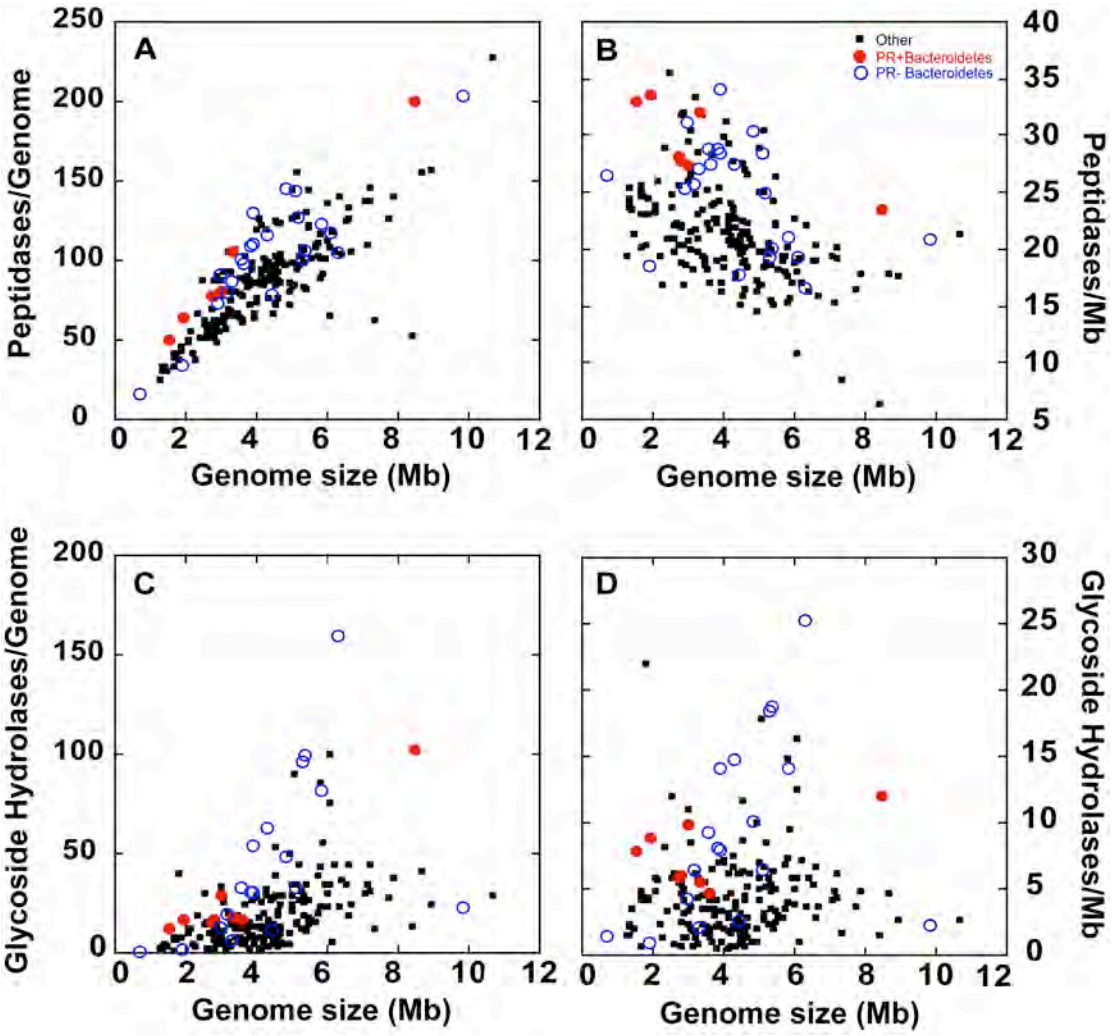
Supplementary Figure 4. Phylogenetic tree of the *Bacteroidetes* emphasizing the marine bacteria (blue), those with the whole genome sequenced (green), with PR (red) and those belonging to the culture collection from the Blanes Bay Microbial Observatory (yellow).

Supplementary Figure 5. Enzymes involved in CO_2 fixation, including the anaplerotic metabolism.

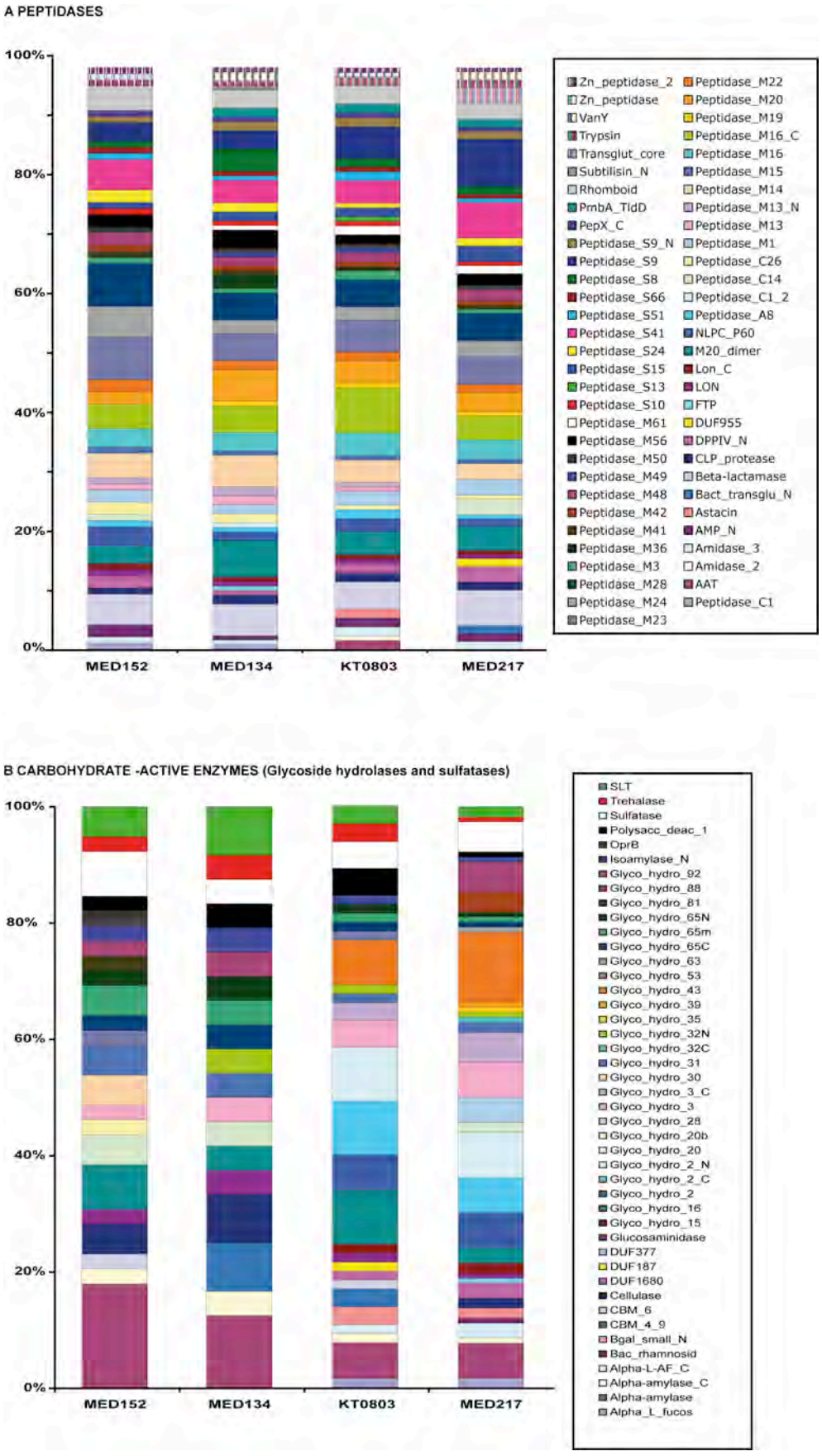
Suppl. Figure 1



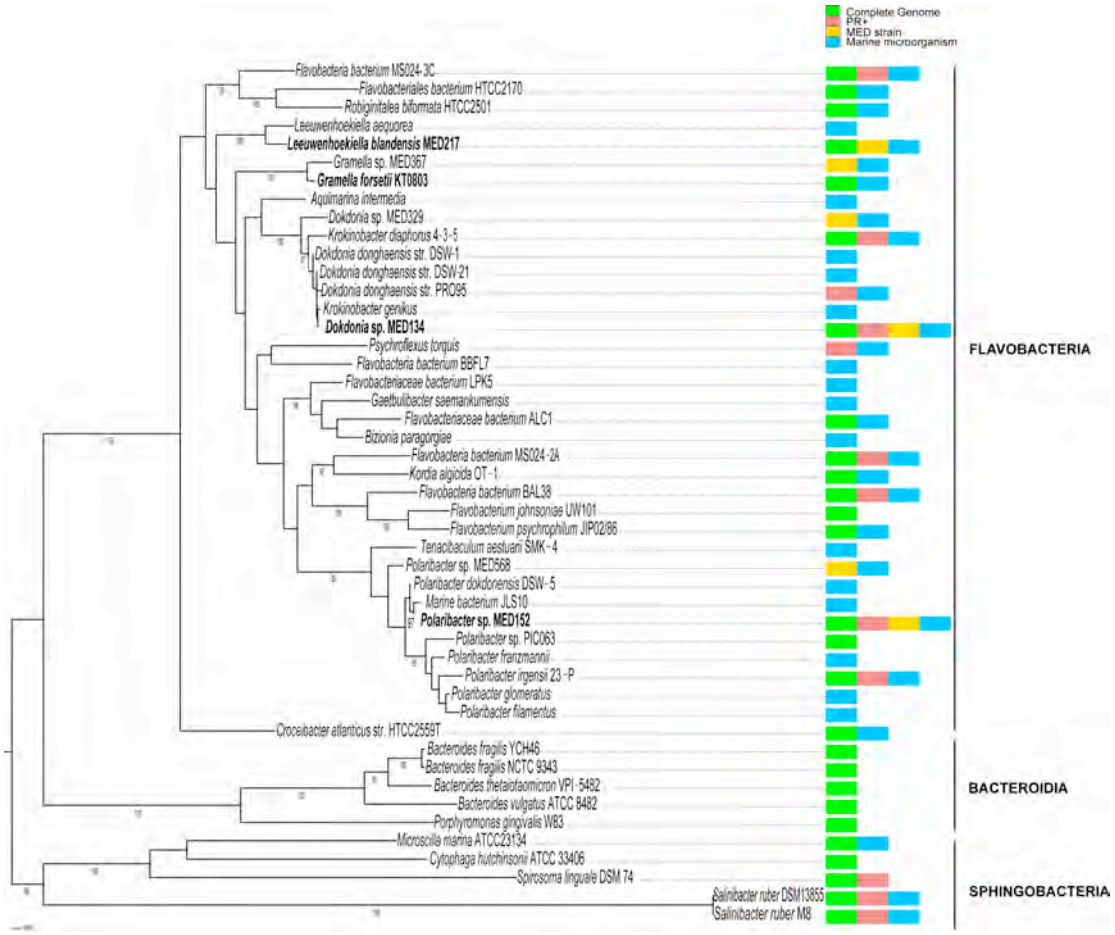
Suppl. Figure 2



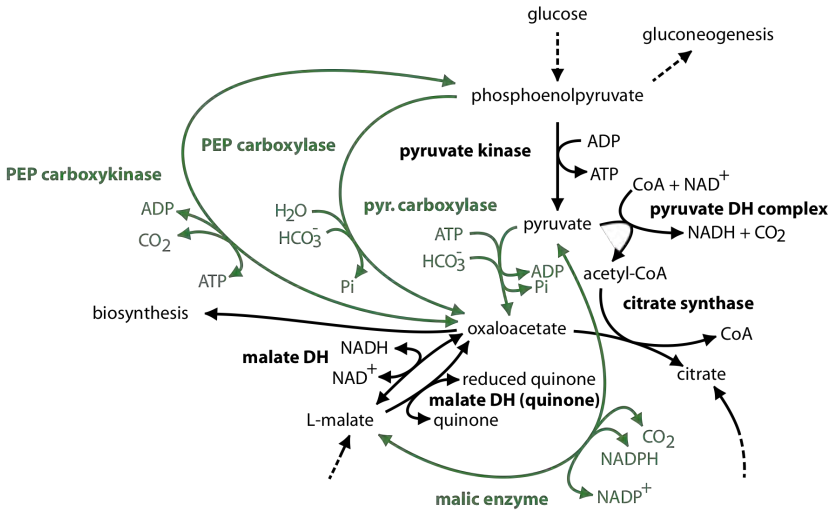
Suppl. Figure 3



Suppl. Figure 4



145 **Suppl. Figure 5**



146

147