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).

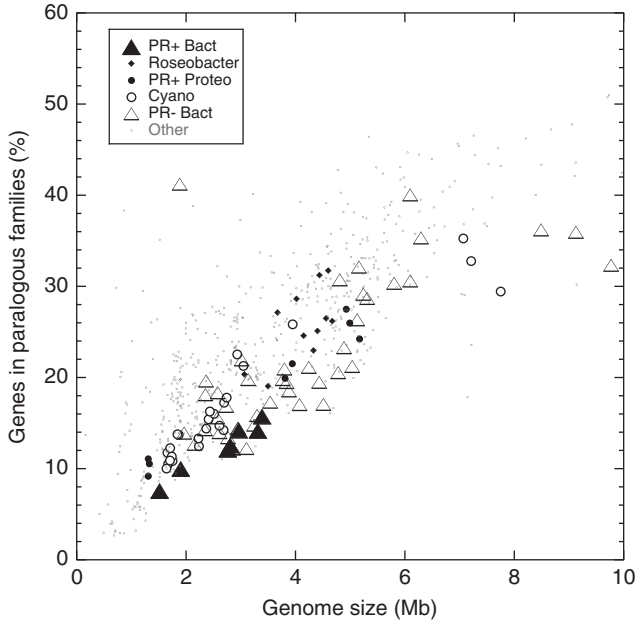

Figure 1 Percentage of genes in paralogous families versus genome size for a number of bacteria. The number of genes in paralogous families was estimated using the BLASTCLUST tool from the NCBI BLAST software (>30% sequence similarity, across >50% of their length and $E<10^{-6}$).

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.

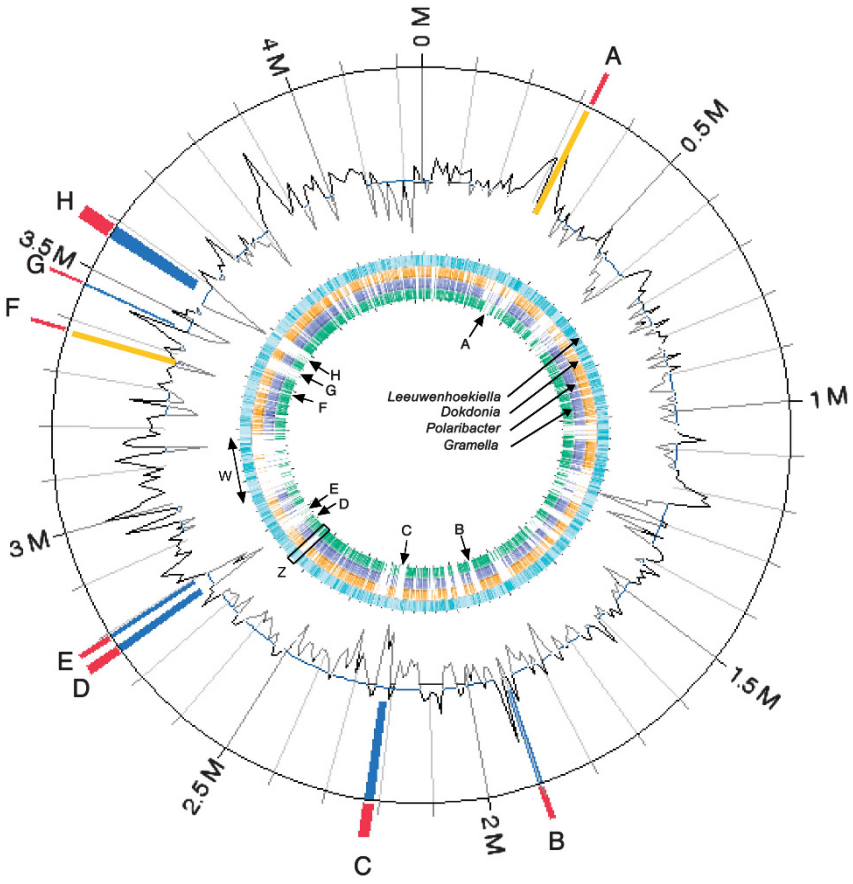

Figure 2 The blue ring shows the genome of *Leeuwenhoekiella blandensis* MED217 ordered from 0 to 4.24 Mb. The next outer ring shows the G+C content along the genome. The outermost ring shows the areas identified as genomic islands with the IslandViewer software. Red, blue and yellow indicate the three 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

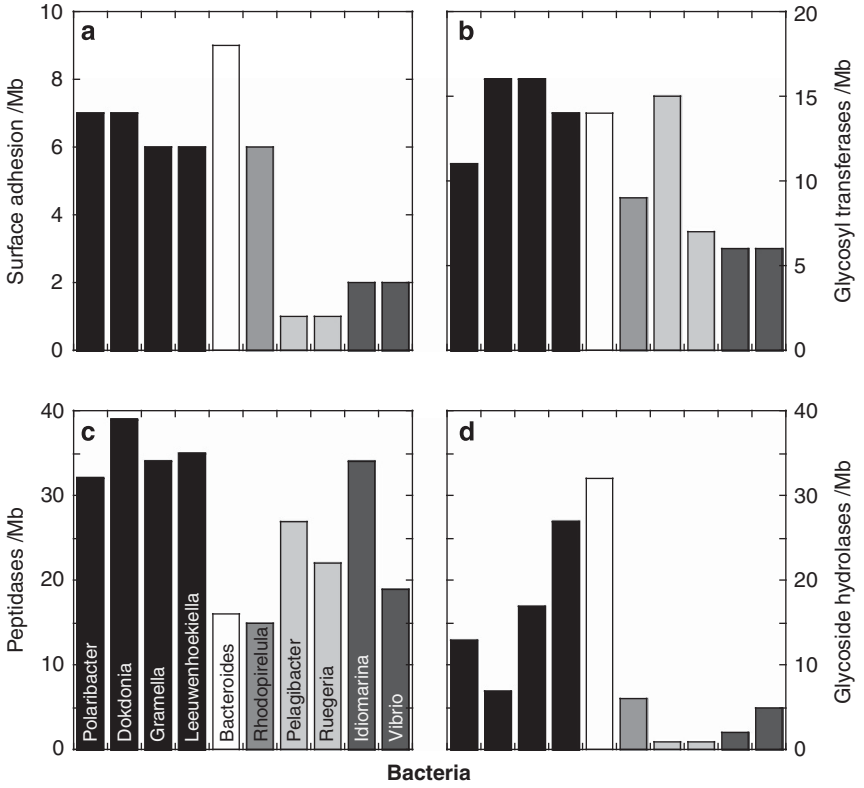

Figure 3 Numbers of different enzymes per megabase of genome for a selection of bacteria: Marine *Bacteroidetes* (black bars) *Polaribacter* sp. MED152, *Dokdonia* sp. MED134, '*Gramella forsetii*' KT0803 and *Leeuwenhoekiella blandensis* MED217; *Bacteroides thetaiotaomicron* (white bar); *Rhodopirellula 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.

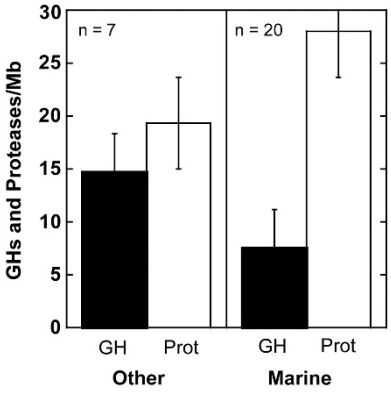

Figure 4 Peptidases and glycoside hydrolases per Mb for marine and non-marine Bacteroidetes.

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.

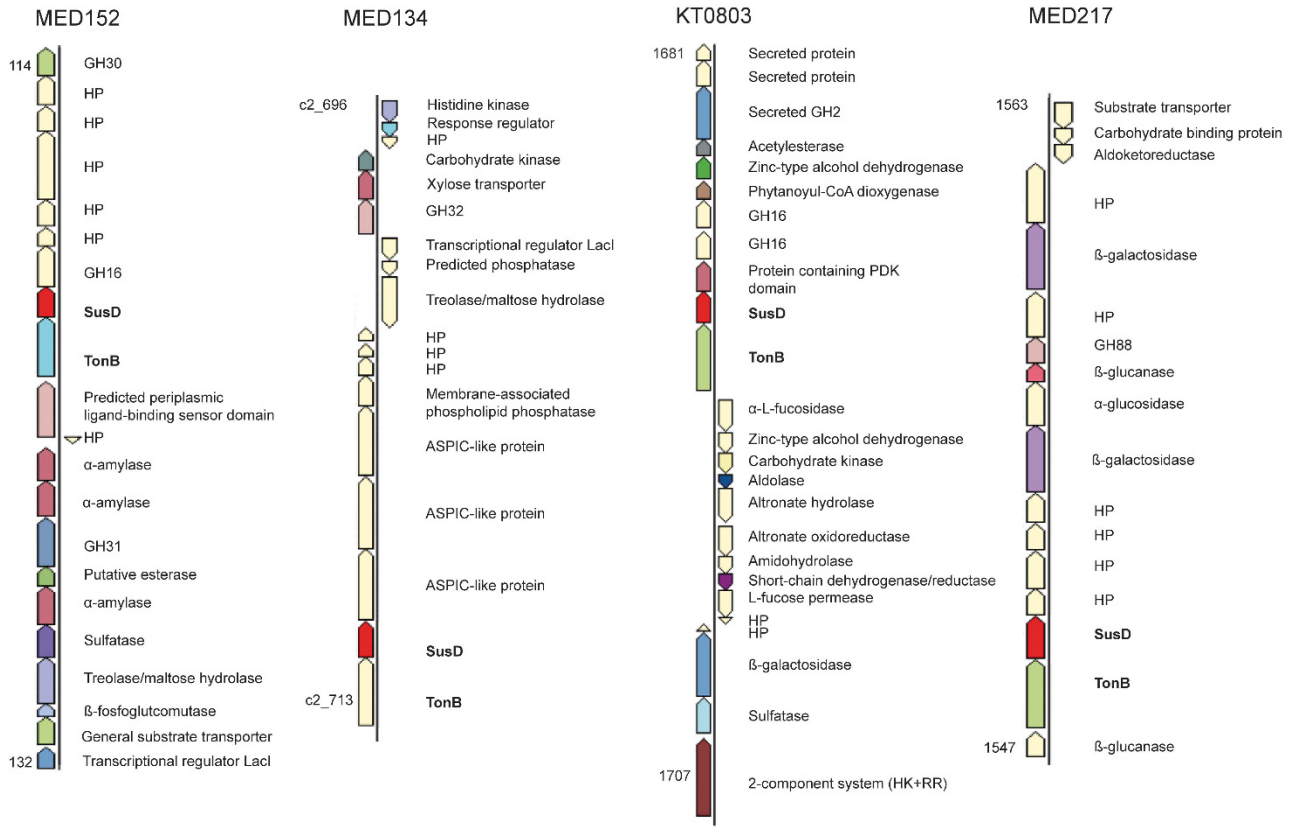

Figure 5 Clusters of genes putatively involved in the attachment and degradation of polymeric compounds and containing TonB-dependent/ligand-gated channel genes and SusD genes. One example from each of the four bacteria is shown. The complete list can be found in Supplementary Table 9. HP, hypothetical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

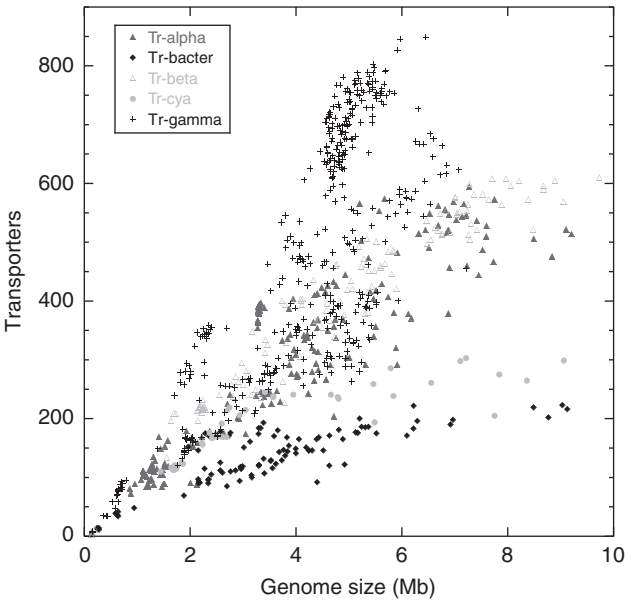

Figure 6 Number of transporters versus genome size for different groups of bacteria: Gammaproteobacteria (crosses), Alphaproteobacteria (solid triangles), Betaproteobacteria (empty triangles), Cyanobacteria (circles) and Bacteroidetes (solid diamonds).

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

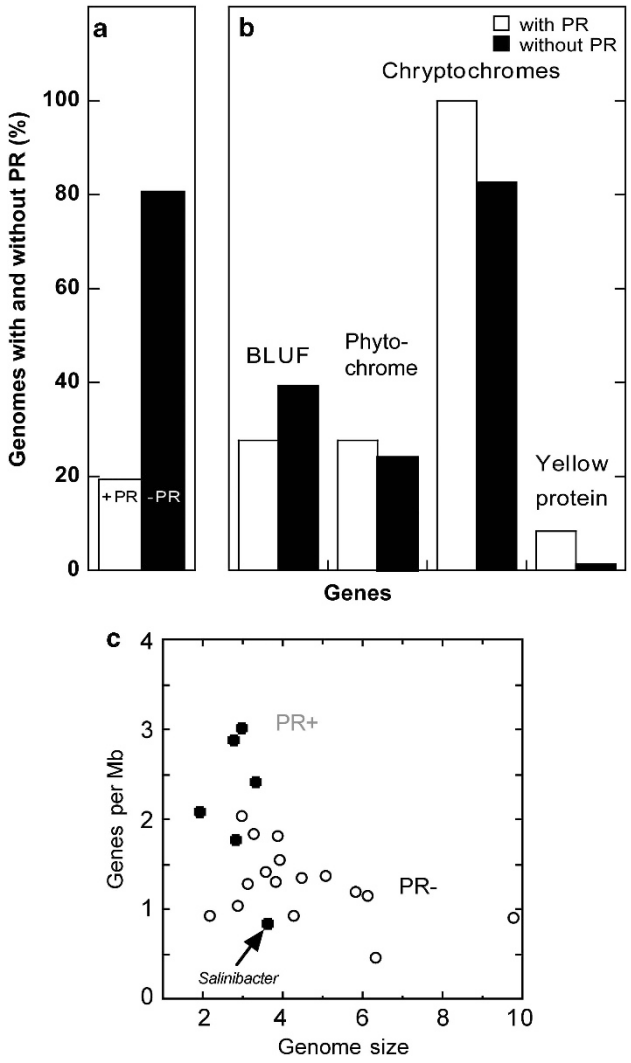

Figure 7 (**a**) Percent of genomes from a collection of 185 marine bacteria with proteorhodopsin (PR+, empty bars) and without it (PR−, filled bars). (**b**) Percent of PR+ or PR− genomes with different light sensing domains. (**c**) Genes involved in anaplerotic fixation and transport of CO_2 in different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