Vol 464|8 April 2010|doi:10.1038/nature08937

LETTERS

Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota

Jan-Hendrik Hehemann[1,2]†, Gaëlle Correc[1,2], Tristan Barbeyron[1,2], William Helbert[1,2], Mirjam Czjzek[1,2] & Gurvan Michel[1,2]

Gut microbes supply the human body with energy from dietary polysaccharides through carbohydrate active enzymes, or CAZymes[1], which are absent in the human genome. These enzymes target polysaccharides from terrestrial plants that dominated diet throughout human evolution[2]. The array of CAZymes in gut microbes is highly diverse, exemplified by the human gut symbiont *Bacteroides thetaiotaomicron*[3], which contains 261 glycoside hydrolases and polysaccharide lyases, as well as 208 homologues of *susC* and *susD*-genes coding for two outer membrane proteins involved in starch utilization[1,4]. A fundamental question that, to our knowledge, has yet to be addressed is how this diversity evolved by acquiring new genes from microbes living outside the gut. Here we characterize the first porphyranases from a member of the marine Bacteroidetes, *Zobellia galactanivorans*, active on the sulphated polysaccharide porphyran from marine red algae of the genus *Porphyra*. Furthermore, we show that genes coding for these porphyranases, agarases and associated proteins have been transferred to the gut bacterium *Bacteroides plebeius* isolated from Japanese individuals[5]. Our comparative gut metagenome analyses show that porphyranases and agarases are frequent in the Japanese population[6] and that they are absent in metagenome data[7] from North American individuals. Seaweeds make an important contribution to the daily diet in Japan (14.2 g per person per day)[8], and *Porphyra* spp. (nori) is the most important nutritional seaweed, traditionally used to prepare sushi[9,10]. This indicates that seaweeds with associated marine bacteria may have been the route by which these novel CAZymes were acquired in human gut bacteria, and that contact with non-sterile food may be a general factor in CAZyme diversity in human gut microbes.

Marine algae contain sulphated polysaccharides[11] that are absent in terrestrial plants. These unique polymers are used as a carbon source by marine heterotrophic bacteria that produce specific CAZymes[12]. In comparison to the accumulating knowledge on the degradation of plant polysaccharides[13], little is known about the enzymes acting on polysaccharides from marine edible algae such as *Porphyra* (nori), *Ulva* (sea lettuce) or *Undaria* (wakame). This lack of knowledge concerning marine glycoside hydrolases, which impedes their identification in metagenomic studies, was the initial rationale for our search for new glycoside hydrolases in *Z. galactanivorans*. This marine Bacteroidetes was isolated from the red alga *Delesseria sanguinea* for its capacity to degrade agars and carrageenans[14–17]. The recent genome analysis of *Z. galactanivorans* (T.B. *et al.*, manuscript in preparation) showed the existence of five proteins (Zg1017, Zg2600, Zg3376, Zg3628 and Zg3640) that are distant relatives of β-agarases and κ-carrageenases (on average 23% identity with AgaA and CgkA). All of these sequences feature the catalytic signature EXDXXE typical of glycoside hydrolase family 16 (GH16) (Supplementary Fig. 1), but do not possess the critical residues needed for recognition of agarose or κ-carrageenan[12].

To identify their substrate specificity we cloned and expressed these five GH16 genes in *Escherichia coli*. However, only Zg1017 and the catalytic module of Zg2600 were expressed as soluble proteins and could be further analysed (Supplementary Fig. 2). As predicted, these proteins had no activity on commercial agarose (Supplementary Fig. 3) or κ-carrageenan. Consequently, we screened their hydrolytic activity against natural polysaccharides extracted from various marine macrophytes (data not shown). Zg2600 and Zg1017 were found to be active only on extracts from the agarophytic red algae *Gelidium*, *Gracilaria* and *Porphyra*, as shown by the release of reducing ends (Fig. 1a). These enzymatic activities correlate with increasing amounts of 4-linked α-L-galactopyranose-6-sulphate (L6S) in the extracted polysaccharides, as shown by ^{1}H-NMR (Fig. 1b)[18]. This monosaccharide is typical of porphyran, which ideally consists of alternating L6S and 3-linked β-D-galactopyranose units (G)[19] (Fig. 1a and Supplementary Fig. 4a). Occasionally, L6S units are replaced by 4-linked 3,6-anhydro-α-L-galactopyranose units (LA), as found in agarose[20]. Because the extracted porphyran contained about two-thirds L6S–G and one-third LA–G motifs (Fig. 1b), this polysaccharide was purified from LA–G motifs by β-agarase pre-treatment. Zg2600 and Zg1017 showed highest activity on the purified porphyran, whereas agarases were inactive. ^{1}H-NMR analysis of the purified reaction products shows that the major final product of both porphyranases is the disaccharide L6S–G (Supplementary Figs 5 and 6), indicative of cleavage of the β-1,4-glycosidic bond (Supplementary Figs 4b and 6). Therefore, Zg2600 and Zg1017 are the first β-porphyranases described so far—referred to as PorA and PorB, respectively—and they represent a new class of glycoside hydrolases.

We solved the crystal structures of PorB and of an inactivated mutant of PorA in complex with a porphyran tetrasaccharide, allowing the identification of key residues for porphyran recognition. The structural differences in β-porphyranases that determine substrate recognition predominantly concern the subsite −2, which binds an L6S unit in PorA and an LA unit in AgaA (Protein Data Bank accession 1URX). The L6S unit is surrounded by W56 (W67, PorB) on one side, R133 on the opposite side and by R59 (R70, PorB) above (Fig. 2a). Whereas W56 stacks to one face of the L6S unit, R59 forms a bidentate hydrogen bond with the hydroxyl group on carbon 2 of the L6S unit and the α-(1,3)-glycosidic-bond oxygen. The sulphate group of the L6S unit is oriented towards the interior of the cleft and protrudes into a positively charged and hydrophobic pocket (Fig. 2b). Critical for porphyran recognition, this pocket leaves space for the bulky sulphate group substituted at carbon 6. In contrast, non-adapted glycoside hydrolases, such as β-agarases, impose high sterical constraints at this position (Supplementary Fig. 7a) rendering productive binding

[1]Université Pierre et Marie Curie, Paris 6, [2]Centre National de la Recherche Scientifique, Végétaux marins et Biomolécules UMR 7139, Station Biologique de Roscoff, Roscoff F 29682, France. †Present address: Biochemistry and Microbiology, University of Victoria, PO Box 3055 STN CSC, Victoria, British Columbia V8W 3PG, Canada.

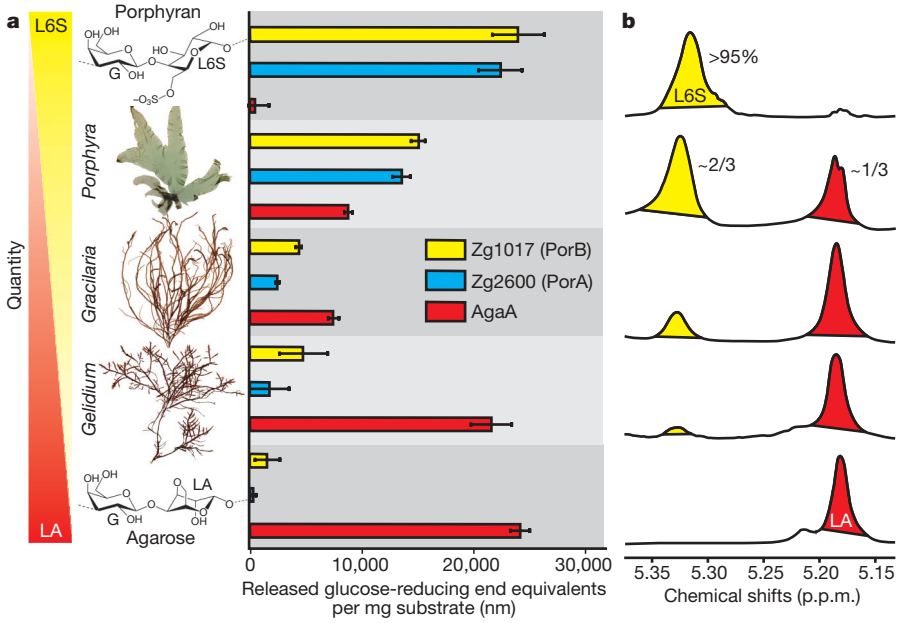

Figure 1 | Activity screening on natural algal polysaccharides reveals porphyranase activity. **a**, PorA and PorB were screened on crude algal extracts and compared to AgaA; the species on which the enzymes were active are displayed. The bars show the amount of produced glucose-reducing end equivalents ($n = 3$, error bars represent mean ± s.d.). The activities on crude extracts are compared with that on pure porphyran (top) and agarose (bottom). **b**, ^{1}H-NMR spectra from the corresponding substrates in **a** show that the enzymatic activity correlates with the quantity of L6S units. Activity of PorA and PorB is highest on pure porphyran (>95% L6S–G).

impossible, as reflected by the biochemical activity of AgaA, which cannot degrade pure porphyran.

This structural information was then exploited for data mining of porphyranases in sequence databases. Six potential β-porphyranases were identified in the GenBank non-redundant database, with identity ranging from 35% to 55% with PorA and PorB. Structure-based sequence alignment indicates that the crucial residues for porphyran recognition are well conserved in all of these GH16 enzymes (Supplementary Fig. 1). Moreover, phylogenetic analysis confirms that these six proteins, together with Zg3376, Zg3628, Zg3640, PorA and PorB, constitute a monophyletic group (clade 2) distinct from the β-agarase and κ-carrageenase subfamilies (Fig. 3; for details see Supplementary Table 2). Therefore, all of these additional proteins are confidently predicted to be β-porphyranases and clade 2 can be defined as a new subfamily within GH16. The genes coding for Zg3376, Zg3628 and Zg3640 from *Z. galactanivorans* are thus named *porC*, *porD* and *porE*, respectively.

All the orthologues of β-porphyranases that we detected by blast analysis in GenBank are encoded by marine bacterial genomes, except for Bp1689 (36%, 35% and 52% identity with PorA, PorB and PorE, respectively) which originates from the human gut bacterium *Bacteroides plebeius*. To date, the literature describes six different strains of this species, all of which were isolated from the intestinal microbiota of Japanese individuals[5] and the type strain was sequenced as part of the Human Microbiome Project (HMP, http://genome.wustl.edu/genomes). Notably, *B. plebeius* also contains a putative β-agarase from GH16, Bp1670 (53% identity with AgaB, Fig. 3). In contrast, none of the 24 other *Bacteroides* genome sequences available at the US National Center for Biotechnology Information (NCBI) contain any β-porphyranase or β-agarase genes. Among these additional species, ten species were isolated from western individuals (eight North American and two French) and six from East Asian individuals (five Japanese and one Japanese Hawaiian). The remaining strains are uncharacterized and were selected by the HMP

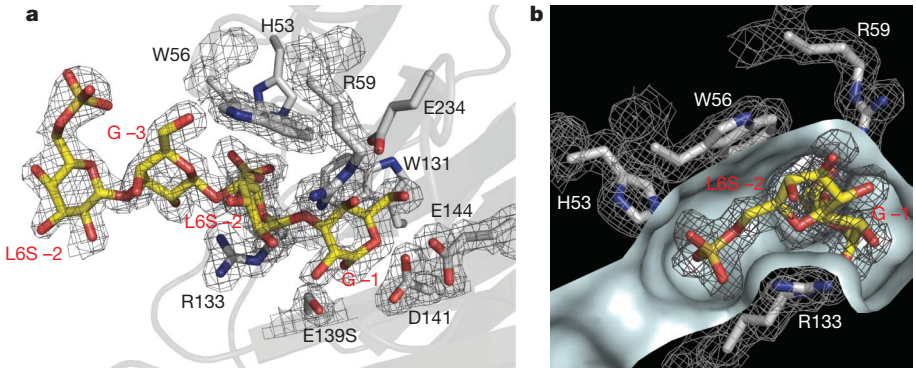

Figure 2 | Structural determinants of porphyran active enzymes. **a**, An active site mutant of PorA was generated by polymerase chain reaction (PCR) mutagenesis and co-crystallized with the purified porphyran tetrasaccharide showing the key residues for porphyran recognition. Four sugar units are defined by electron density. The negative charge of the sulphate group is neutralized with salt bridges from the side chains of H53 and R133. **b**, Surface representation of the substrate binding site, perpendicular to the view in **a**, highlighting the pocket specifically accommodating the sulphate group of the L6S unit bound to binding site −2. The pocket is occupied by a tyrosine residue in β-agarases (Y69 in AgaA), which is situated in the highly divergent amino-terminal region. In both panels final refined electron density ($2F_o - F_c$; grey grid) surrounding the sugar units and the neighbouring amino acids is represented at a 1σ level. The figure was prepared with PyMol.

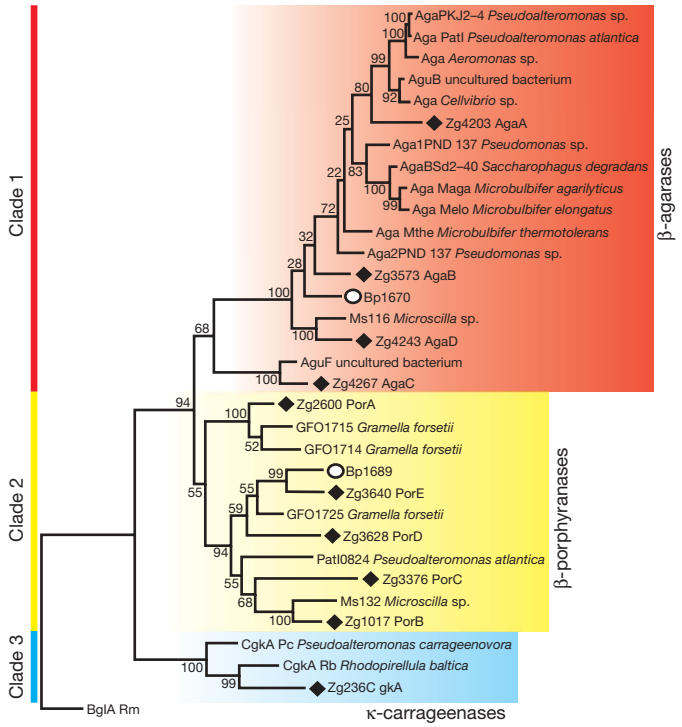

Figure 3 | Phylogenetic analysis of GH16 galactanases reveals porphyranases in many marine bacteria and in the Japanese gut bacterium *B. plebeius*. GH16 enzymes specific for red-algal-sulphated galactans cluster in three clades: clade 1, β-agarases; clade 2, β-porphyranases; and clade 3, κ-carrageenases. Marine bacteria use these enzymes for red-algal galactan degradation. The non-marine exception is the human gut bacterium *B. plebeius*, which contains a β-porphyranase and a β-agarase (white circles) reminiscent of the porphyranolytic system of *Z. galactanivorans* (black diamonds).

(see Supplementary Table 5). Analysis of the *B. plebeius* genome shows that *Bp1670* and *Bp1689* are surrounded by other CAZyme genes (Fig. 4): two glycoside hydrolase family 86 (GH86) β-agarases (*Bp1693* and *Bp1694*), two glycoside hydrolase family 2 (GH2) β-galactosidases (*Bp1672* and *Bp1673*) and a sulphatase (*Bp1701*). This genomic region also contains a gene coding a carbohydrate-binding module family 22 (CBM22) protein (Bp1696), preceded by a *susD*-like gene (*Bp1697*) and its associated TonB-dependent receptor gene (*susC*-like, *Bp1698*). All these carbohydrate-related genes have their closest orthologues in marine bacteria (identity between 48% and 69%), whereas homologous genes from other gut *Bacteroides* species are much more distant (~30% identity). Moreover, six of these genes (*Bp1670*, *Bp1671*, *Bp1689*, *Bp1693*, *Bp1694* and *Bp1696*) have no homologue in other *Bacteroides* genomes. The genes *Bp1696*, *Bp1697* and *Bp1698* also share synteny with two polysaccharide utilization loci (PUL) from *Z. galactanivorans*, which include either *agaB* (*Zg3573*) or *porE* (*Zg3640*). Interspersed between these genes shared with marine bacteria there are genes well conserved with other gut *Bacteroides* (Supplementary Table 3), showing that *B. plebeius* is a regular gut symbiont that received an unusual set of genes most probably by horizontal gene transfer (HGT) from a marine bacterium. The possible mechanism for HGT was identified by analysing the region downstream of *Bp1670*, which contains genes coding for conserved relaxase/mobilization proteins (Bp1662 and Bp1663/Bp1665), required for conjugative DNA transfer[21]. Interestingly, the β-agarase Ms116 and the putative β-porphyranase Ms132 of *Microscilla* sp. PRE1 (Fig. 3) are encoded by a plasmid that contains a relaxase gene (*Ms155*) homologous to *Bp1662*[22]. In conclusion, these results indicate that *B. plebeius* acquired a porphyran utilization locus that originated from an ancestral porphyranolytic marine bacterium, related to the extant marine Bacteroidetes *Z. galactanivorans* and *Microscilla* sp. PRE1.

To extend these findings, we searched for homologues of PorA, PorB and Bp1689 in metagenomic data available at NCBI. No porphyranase was found in samples with terrestrial origin. The metagenome data sets obtained by open ocean sampling did not contain porphyranase genes either, but this is probably explained by the lack of red macroalgae in the 'open ocean'. Porphyranase distribution is thus essentially limited to coastal waters and to seaweed-associated bacteria. To test whether porphyranases are common in Japanese gut microbes, we analysed the gut metagenome data obtained from 13 Japanese volunteers (total read length, 726,907,479 base pairs (bp); average length, 840 bp)[6]. Seven potential porphyranases were unambiguously identified in the microbiomes of four people (between 31% and 42% identity with PorB, between 83% and 100% with Bp1689; see Table 1). Six putative β-agarases (GH16) were also detected in four Japanese individuals (between 51% and 59% identity with AgaB, between 96% and 99% with Bp1670). Interestingly, in one Japanese family, the mother and her unweaned baby girl had a microbiota containing porphyranase and agarase genes, suggesting that porphyranolytic gut *Bacteroides* strains may be transmitted between relatives. In contrast, when analysing the gut metagenome data of 18 North American individuals (total read length, 1,830,767,417 bp; average length, 223 bp)[7] no porphyranase or agarase genes were detected, even though this data set is 2.5-fold larger than the

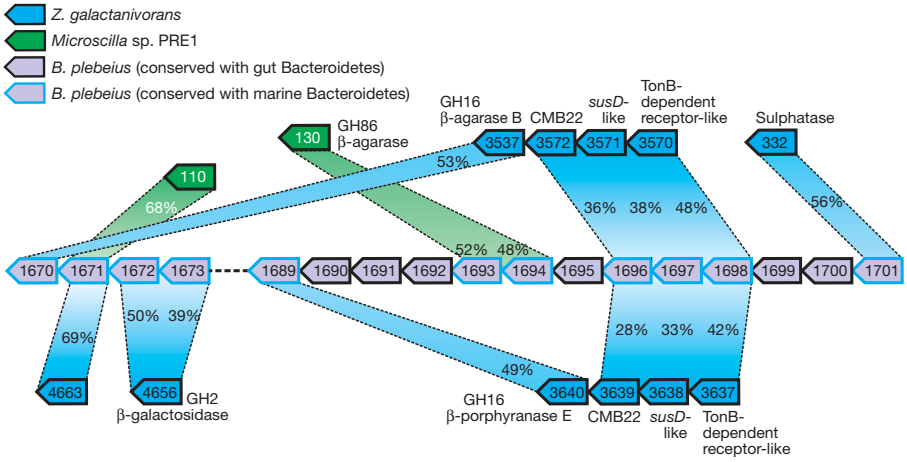

Figure 4 | Upstream and downstream of the porphyranase gene (*Bp1689*), the genome of *B. plebeius* contains carbohydrate-related genes that share highest identity with proteins used for red-algal galactan degradation in two marine Bacteroidetes. Shown is the sequence identity between *B. plebeius*, *Microscilla* sp. PRE1 and *Z. galactanivorans* proteins. Six of these genes (*Bp1670*, *Bp1671*, *Bp1689*, *Bp1693*, *Bp1694* and *Bp1696*) are conserved with marine bacteria, but are absent in genomes of other gut *Bacteroides*.

Table 1 | Presence of porphyranases and agarases in Japanese gut metagenomes

Individual	Query	Porphyranase Bp1689 identity (Zg1017) (%)	Match length (no. of amino acids)	Agarase Bp1670 identity (Zg3573) (%)	Match length (no. of amino acids)
F1-S	F1-S_3191.1_B_pred_2_-_157_654_0_C	83 (42)	165 (109)	-	-
	F1-S_077P07.z_B_pred_2_-_400_864_0_C	84 (32)	139 (150)	-	-
	F1-S_051H15.y_B_pred_2_-_130_390_0_C	93	82	-	-
F2-W (mother)	F2-W_029I11.z_B_pred_1_-_12_764_0_C	85 (39)	235 (162)	-	-
	F2-W_368.1_B_pred_2_-_832_1554_0_C	-	-	96 (51)	222 (202)
F2-Y (child)	F2-Y_2347.1_B_pred_2_+_1383_2351_0_C	87 (37)	-	-	-
	F2-Y_032L13.y_B_pred_1_-_1_614_0_P	94 (31)	184 (207)	-	-
	F2-Y_008O23.y_A_pred_1_+_1_655_1_P	-	-	97 (58)	194 (206)
	F2-Y_399.1_B_pred_5_-_4757_5316_2_P	-	-	97 (59)	185 (205)
In-R	In-R_047A09.g_B_pred_1_+_1_846_0_P	-	-	99 (52)	281 (298)
	In-R_5520.1_A_pred_1_-_43_666_0_C	-	-	98 (59)	207 (209)
In-D	In-D_058K14.b_B_pred_2_-_470_958_0_C	100 (33)	162 (145)	-	-
	In-D_129O10.g_B_pred_1_-_76_909_0_C	-	-	96 (53)	277 (266)

Sequence identities and match lengths are given without parentheses for matches to Bp1689 and Bp1670 and in parentheses for matches to Zg1017 (*Z. galactanivorans* PorB) and Zg3573 (*Z. galactanivorans* AgaB). Queries were run through http://www.metagenome.jp.

Japanese data set (almost tenfold considering average read length)[6]. However, the average read length is shorter in the American data set. To test whether this difference might have an impact on the likelihood of detecting porphyranases and agarases, we searched for homologues of the gene coding for the α-amylase SusA from *B. thetaiotaomicron*, which is well conserved in *Bacteroides* genomes. Using the following criteria (*E*-value $<10^{-15}$, score >50, identity >50%), we found 34 and 306 homologues in the Japanese and American data sets, respectively, confirming that there is no bias due to the read length. Applying the Fischer's exact test[23] to the respective CAZyme counts gives a *P*-value of $P = 1.68 \times 10^{-6}$, indicative of a statistically significant association between populations and the occurrence of porphyranase sequences in gut microbiota. Altogether, analyses of the available genomic and metagenomic data indicate that porphyranase and agarase genes are specifically encountered in Japanese gut bacteria and are probably absent in the microbiome of western individuals.

The detection of this HGT was possible owing to two favourable factors: β-porphyranases are absent in terrestrial microbes, and the transfer was relatively recent compared to the millions of years of mammalian gut microbiome evolution. This is supported by the high sequence identity of the porphyranase and associated genes in *B. plebeius* with genes found in marine Bacteroidetes. In contrast, acquisition of terrestrial plant-specific CAZyme genes probably occurred early in the evolution of herbivorous/omnivorous mammals, and such horizontally acquired genes would be difficult to distinguish from ancestral, vertically transmitted genes. The timing of such a HGT event is, however, difficult to estimate[24]. The question then arises of how the genes described here were acquired by Japanese gut bacteria. Tax records from the eighth century list seaweeds as payments to the Japanese government[9], showing that they had an important role in Japanese culture. Dietary seaweed therefore is the most probable vector for the contact with marine microbes that led to HGT, as the only porphyran source in human nutrition is nori. Traditionally, nori is not roasted and thus contact with associated marine microbes is promoted through Japanese sushi[9]. Consequently, the consumption of food with associated environmental bacteria is the most likely mechanism that promoted this CAZyme update into the human gut microbe.

METHODS SUMMARY

Porphyra umbilicalis algae were collected in spring 2008 near Roscoff in the intertidal zone. One-hundred grams of dry weight material was used for the polysaccharide extraction[25]. This polysaccharide is referred to as porphyran. To remove LA–G motifs, porphyran was pre-digested with an excess of β-agarase B (AgaB)[17]. The resistant fraction was separated from oligosaccharides by gel-filtration. This fraction is referred to as pure porphyran.

Enzyme activity assays were carried out as previously described for β-agarases[16] but at 30 °C, with a porphyranase concentration of ~5 nM and a porphyran solution of 0.125% (w/v). The oligosaccharides produced during enzymatic digestion of all polysaccharides from various algal species were measured by a reducing sugar assay calibrated against a glucose concentration curve and presented as glucose-reducing end equivalents[26].

PorA and PorB were produced by recombinant expression in *E. coli* (see Methods). All enzyme constructs are appended by a His_6-tag for protein purification following standard affinity- and size-exclusion high-performance liquid chromatography (HPLC) methods. The crystal structure of PorA was solved by multiwavelength anomalous diffraction (MAD) on a gold derivative using the program HKL2MAP[27]. A standard mutagenesis kit from Qiagen was used to introduce the E139S mutation into PorA (for primers, see Supplementary Table 4). A complex structure was obtained by co-crystallization with purified porphyran tetraoligosaccharide. Refinement was performed with REFMAC[28] and model building with Coot[29]. The crystal structure of PorB was solved by molecular replacement and using PorA as a search model. All data sets were collected at the European Synchrotron Radiation Facilities.

The purification of porphyran oligosaccharides was carried out by several subsequent liquid chromatography steps. ^{1}H-NMR spectra were recorded and the signals were fully assigned using a complete set of correlation spectra (COSY, HMBC and HMQC). Details of the phylogeny analysis are described in the Methods section.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November 2009; accepted 19 February 2010.

1. Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res.* **37,** D233–D238 (2009).
2. Ley, R. E., Lozupone, C. A., Hamady, M., Knight, R. & Gordon, J. I. Worlds within worlds: evolution of the vertebrate gut microbiota. *Nature Rev. Microbiol.* **6,** 776–778 (2008).
3. Xu, J. *et al.* A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* **299,** 2074–2076 (2003).
4. Martens, E. C., Koropatkin, N. M., Smith, T. J. & Gordon J. I.. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J. Biol. Chem.* **284,** 24673–24677 (2009).
5. Kitahara, M., Sakamoto, M., Ike, M., Sakata, S. & Benno, Y. *Bacteroides plebeius* sp. nov. and *Bacteroides coprocola* sp. nov., isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **55,** 2143–2147 (2005).
6. Kurokawa, K. *et al.* Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res.* **14,** 169–181 (2007).
7. Turnbaugh, P. J. *et al.* A core gut microbiome in obese and lean twins. *Nature* **457,** 480–484 (2009).
8. Fukuda, S. *et al.* Pattern of dietary fiber intake among the Japanese general population. *Eur. J. Clin. Nutr.* **61,** 99–103 (2007).
9. Nisizawa, K., Noda, H., Kikuchi, R. & Watanabe, T. The main seaweed foods in Japan. *Hydrobiologia* **151–152,** 5–29 (1987).
10. Mc Hugh, D. J. in *FAO Fisheries Technical Paper No 441* (FAO, 2003).
11. Kloareg, B. & Quatrano, R. S. Structure of the cell walls of marine algae and ecophysiological functions of the matrix polysaccharides. *Oceanogr. Mar. Biol Annu. Rev.* **26,** 259–315 (1988).
12. Michel, G., Nyvall-Collen, P., Barbeyron, T., Czjzek, M. & Helbert, W. Bioconversion of red seaweed galactans: a focus on bacterial agarases and carrageenases. *Appl. Microbiol. Biotechnol.* **71,** 23–33 (2006).
13. Gilbert, H. J., Stalbrand, H. & Brumer, H. How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr. Opin. Plant Biol.* **11,** 338–348 (2008).
14. Barbeyron, T. *et al. Zobellia galactanovorans* gen. nov., sp. nov., a marine species of *Flavobacteriaceae* isolated from a red alga, and classification of [*Cytophaga*] *uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Zobellia uliginosa* gen. nov., comb. nov. *Int. J. Syst. Evol. Microbiol.* **51,** 985–987 (2001).
15. Barbeyron, T. Gerard, A. Potin, P., Henrissat, B. & Kloareg, B. The κ-carrageenase of the marine bacterium *Cytophaga drobachiensis*. Structural and phylogenetic

relationships within family-16 glycoside hydrolases. *Mol. Biol. Evol.* **15,** 528–537 (1998).
16. Jam, M. *et al.* The endo-β-agarases AgaA and AgaB from the marine bacterium *Zobellia galactanivorans*: two paralogue enzymes with different molecular organizations and catalytic behaviours. *Biochem. J.* **385,** 703–713 (2005).
17. Allouch, J. *et al.* The three-dimensional structures of two β-agarases. *J. Biol. Chem.* **278,** 47171–47180 (2003).
18. Maciel, J. S. *et al.* Structural characterization of cold extracted fraction of soluble sulfated polysaccharide from the red seaweed *Gracilaria birdiae*. *Carbohydr. Polym.* **71,** 559–565 (2008).
19. Turvey, J. R. & Rees, D. A. Isolation of L-galactose-6-sulphate from a seaweed polysaccharide. *Nature* **189,** 831–832 (1961).
20. Anderson, N. S. & Rees, D. A. Porphyran — a polysaccharide with a masked repeating structure. *J. Chem. Soc.* 5880–5887 (1965).
21. Garcillán-Barcia, M. P., Francia, M. V. & de la Cruz, F. The diversity of conjugative relaxases and its application in plasmid classification. *FEMS Microbiol. Rev.* **33,** 657–687 (2009).
22. Zhong, Z. *et al.* Sequence analysis of a 101-kilobase plasmid required for agar degradation by a *Microscilla* isolate. *Appl. Environ. Microbiol.* **67,** 5771–5779 (2001).
23. Fisher, R. A. On the interpretation of χ^2 from contingency tables, and the calculation of P. *J. R. Stat. Soc.* **85,** 87–94 (1922).
24. Kuo, C. H. & Ochman, H. Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol. Direct* **4,** 35 (2009).
25. Ishihara, K., Oyamada, C., Matsushima, R., Murata, M. & Muraoka, T. Inhibitory effect of porphyran, prepared from dried ''nori'', on contact hypersensitivity in mice. *Biosci. Biotechnol. Biochem.* **69,** 1824–1830 (2005).
26. Kidby, D. K. & Davidson, D. J. A convenient ferricyanide estimation of reducing sugars in the nanomole range. *Anal. Biochem.* **55,** 321–325 (1973).
27. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with *SHELX* programs. *J. Appl. Cryst.* **37,** 843–844 (2004).
28. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53,** 240–255 (1997).
29. Potterton, L. *et al.* Developments in the *CCP*4 molecular-graphics project. *Acta Crystallogr. D* **60,** 2288–2294 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank B. Kloareg and C. de Vargas for critical discussions and reading of the manuscript and M. Jam, A. Jeudy and D. Freudenreich for technical assistance. The 'Marine Plants and Biomolecules' laboratory is funded by the French national research centre (Centre National de la Recherche Scientifique) and the University Marie Curie; J.-H.H. was supported by a European Marie Curie PhD grant; this work was also funded by the 'Region Bretagne' through the program Marine 3D. G.M. was supported by the GIS 'Genomique Marine' and the French Research Ministry (ACI Young Researcher). We thank the beamline scientists and staff at the European Synchrotron Radiation Facilities for technical support during data collections, the NMR Service, University Bretagne Occidentale, for access to the Bruker NMR spectrometer and Genoscope for sequencing the *Z. galactanivorans* genome.

Author Contributions J.-H.H. cloned, purified and crystallized the enzymes and extracted polysaccharides; J.-H.H. and M.C. collected data and solved the crystal structures; G.C. and J.-H.H. purified and characterized oligosaccharides; G.C. and W.H. performed the NMR analysis; G.M., T.B. and J.-H.H. performed the bioinformatic analysis; M.C., T.B., G.M. and J.-H.H. designed the study; J.-H.H., M.C. and G.M. analysed the data and wrote the paper. All authors discussed the results and commented on the manuscript.

Author Information Atomic coordinates and structure factors have been deposited at the Protein Data Bank under accession codes 3ILF (PorA_E139S) and 3JUU (PorB). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to M.C. (czjzek@sb-roscoff.fr) or G.M. (gurvan@sb-roscoff.fr).

METHODS

Activity screening on algal extracts. Algae of species (*Ectocarpus siliculosus*, *Delesseria sanguinea*, *Zostera marina*, *Porphyra umbilicalis*, *Gracilaria gracilis*, *Gelidium spinosum*, *Chondrus crispus* and *Laurencia pinnatifida*) were all collected at low tide near Roscoff, extensively washed with fresh water, shock frozen with liquid nitrogen and ground in a mortar until a fine powder was obtained. One-hundred grams of dry weight material was used for the polysaccharide extraction[25]. The algal powder was treated with 80% ethanol, centrifuged and the supernatant was discarded. Ethanol extraction was repeated until the supernatant remained colourless. The obtained algal solid fraction was pelleted, frozen and subsequently dried by lyophyllization. For the enzymatic activity tests, 50 mg of algal powder was resuspended in 1 ml distilled water and extracted for 1 h at 98 °C. The suspension was centrifuged for 5 min at 3,000*g* and the supernatant was transferred into a new tube. This extraction was repeated twice and the three extraction supernatants were finally combined. The dried mass was recovered for each extract to calculate the normalized enzyme activity and to record ^{1}H-NMR-spectra.

One-hundred microlitres of algal extract was incubated with ~1 μg of enzyme at 30 °C in 300 mM NaCl and 10 mM sodium acetate, 16 mM tri-sodium citrate (pH 7.2) until complete digestion could be obtained. Complete digestion was monitored by further adding of enzyme and re-measuring the amount of released oligosaccharides. The samples were centrifuged for 30 min at 3,000*g*, to spin down non-digested gel particles, and re-equilibrated for 20 min at 20 °C. Soluble reaction products in the supernatant were quantified by a reducing sugar assay[26]. The amount of reducing ends produced by the enzymes was calculated with a glucose calibration curve and the obtained results are presented as glucose-reducing end equivalents (Fig. 1). The final values were calculated by subtracting control levels measured for solutions with no enzyme added (data not shown). The calibration curve with 50–300 μg ml^{-1} of glucose was used to calculate the amount of reducing ends as glucose-reducing end equivalents.

Medium throughput cloning and expression strategy. The five new GH16 enzymes from *Z. galactanivorans* were selected for recombinant expression. Two enzymes (Zg2600 and Zg3628) are multimodular and consist of catalytic modules appended by carbohydrate-binding modules (CBMs)[30] or modules of unknown function. From these multimodular enzymes only the catalytic modules were amplified by PCR and subsequently cloned. Cloning strategy and primers of all targets are given in Supplementary Table 4.

All targets were amplified in parallel by PCR with a set of forward and reverse primers that were calculated to have a melting temperature of 70 °C. Purified PCR products were digested with a mix of the corresponding restriction enzyme/buffer system (NEB) for 3 h at 37 °C. After digestion, the PCR products were again purified with a standard DNA purification kit and finally eluted with 25 μl of H_2O. The PCR products were ligated with an equally digested pET15b-derived plasmid (gift from M. Cygler), which was additionally dephosphorylated. For the transformation, chemically competent DH5α *E. coli* cells were used following standard procedures.

The recombinant soluble expression of the different constructs was first screened in small scale. The obtained clones were transformed into recombinant Bl21 (DE3) cells and then used to inoculate a 3 ml Luria–Bertani-broth preculture, supplemented with 100 μg ml^{-1} of ampicillin in deep-well format. The preculture was incubated overnight at 37 °C and 10 μl were transferred for inoculation of 2 ml ZYP-5052 medium (ampicillin, 100 μg ml^{-1})[31]. The expression test cultures were incubated at 20 °C for three days and harvested by centrifugation. The cells were lysed with 500 μl of lysis buffer composed of 50 mM Tris, pH 8, 300 mM NaCl, 1 mg ml^{-1} lysozyme and DNase. The cell lysate was centrifuged and the supernatant separated for batch affinity purification with hypercel nickel resin. The remaining cell pellet, which contained the insoluble expressed fraction, was extracted with 6 M urea for 30 min and centrifuged. The supernatant was stored for SDS–polyacrylamide gel electrophoresis (PAGE) analysis. The cell lysate supernatant was used for batch affinity purification and 50 μl of nickel charged hypercel resin was added to the supernatant. After incubation for 20 min, the supernatant was discarded and the pellet washed with 500 μl buffer (lysis buffer without lysozyme and DNase) followed by further centrifugation. The bound proteins were eluted by adding 500 mM imidazole solution. Fractions of inclusion bodies, soluble cell lysate and batch-purified proteins were analysed on 12% SDS–PAGE (data not shown). The expression procedures were then up-scaled to 1 litre cultures for protein production before structural and biochemical analysis.

Crystallization and structure solution. PorA was crystallized by the hanging-drop method in 18–20% PEG 4,000, 0.2 M ammonium sulphate and 0.1 M sodium acetate, pH 4.6, at 20 °C. Drops contained 2 μl of 2.6 mg ml^{-1} protein solution plus 1 μl of crystallization solution and were equilibrated against 500 μl crystallization solution. A heavy atom derivative was obtained by soaking a native crystal for 24 h in mother liquor supplemented with 5 mM potassium tetra-chloro-aureate (III) and 5% glycerol. Before flash freezing at 100 K the crystal was back-soaked in mother liquor with 10% glycerol for ~1 min. MAD data were collected on beamline ID23I. Four data sets were collected from one heavy atom derivative crystal, around the LIII gold absorption edge. Anomalous data were processed with XDS and scaled with XSCALE[32]. The heavy atom substructure solution was determined with SHELXD[33] followed by density modification with SHELXE as part of HKL2MAP[27] and the resulting electron density map was improved by solvent flattening with the program DM (CCP4)[34]. The crystallization conditions for the PorA_E139S complex structure were adapted from the native protein and 5–10 mM of neo-porphyrotetraose was added. For these crystals the data collection was carried out on beamline BM30A. Data were processed with Mosflm[35] and SCALA (CCP4)[34]. Refinement was performed with REFMAC[28] and model building with Coot[29].

PorB was crystallized within one-to-two weeks by the sitting-drop method with a crystallization solution composed of 33–35% PEG 1,000, 0.4 M lithium sulphate and 0.1 M MES buffer at pH 6.0. The sitting drops were formed by mixing 2 μl of protein solution (5.6 mg ml^{-1}) and 1 μl of crystallization solution, and equilibrated against a 500 μl crystallization solution at 20 °C). For cryo-conservation, crystals were brought to 10% glycerol in 5% increments in the crystallization solution. After solving the phases by molecular replacement, the model building and refinement were carried out as described for PorA. For all data sets, 5% of the observations were flagged as free and used to monitor refinement procedures and a summary for data collection and final refinement statistics are given in Supplementary Table 1.

Purification of the oligosaccharides by preparative size-exclusion chromatography. The material for oligosaccharide purification was obtained by incubation at 30 °C of porphyran with 2.6 μg ml^{-1} of PorA in water.

The purification of porphyran oligosaccharides was carried out by preparative size-exclusion chromatography with three Superdex 30 (26/60) columns in series, integrated on a HPLC system liquid injector/collector (Gilson). The freeze-dried hydrolysis product was dissolved in deionized water at a concentration of 4% (w/v). After filtration, 4 ml of sample was injected and eluted with 50 mM ammonium carbonate at a flow rate of 1.5 ml min^{-1}. Fractions of oligosaccharides that could be assigned to distinct peaks were collected, pooled and analysed by HPLC (data not shown) and fluorophore-assisted PAGE and the fractions were freeze-dried by lyophyllization for further analyses by nuclear magnetic resonance (Supplementary Fig. 6).

Fluorophore-assisted carbohydrate electrophoresis. Fractions obtained after separation of oligosaccharides were analysed by fluorophore-carbohydrate–PAGE[36]. Aliquots of 100 μl were labelled with 2 μl AMAC (2-amnoacridone) solution or with 2 μl ANTS (8-aminonaphthalene-1,3,6-trisulphonate) solution. After dissolution of the oligosaccharide pellets, 5 μl of a 1 M sodium cyanoborohydride in dimethylsulphoxide (DMSO) was added and the mixture was incubated for 16 h at 37 °C. The samples were analysed on a 30% polyacrylamide running gel with a 4% stacking gel.

Nuclear magnetic resonance spectroscopy. All samples were solubilized in D_2O and exchanged twice, before being analysed on a Bruker 500 MHz NMR spectrometer. For the polysaccharides extracted from algae that were used as substrates for enzymatic assays, ^{1}H-NMR spectra were recorded at 70 °C using 64 scans. For oligosaccharides, ^{1}H-NMR spectra were recorded at 25 °C using 16 scans (Supplementary Fig. 6). Chemical shifts are expressed in p.p.m. by reference to an external standard TSP (trimethylsilylpropionic acid). The NMR signals of the purified oligosaccharides were fully assigned using a complete set of correlation spectra: COSY (doublequantum-filtered correlation spectroscopy), HMBC (heteronuclear multiple bond correlation) and HMQC (heteronuclear single quantum correlation).

Sequence and phylogeny analyses. Multiple sequence alignments were generated using MAFFT with the iterative refinement method and the scoring matrix Blosum62[37]. The structure-based multiple alignment of the β-porphyranases was displayed using the program ESPript[38]. For the phylogeny analyses of the GH16 enzymes, the sequences were selected from the CAZY database. The MAFFT alignment of these sequences was manually refined using Bioedit (T. Hall), on the basis of the superposition of the structure of the κ-carrageenase of *Pseudoalteromonas carrageenovora*[39], the β-agarase AgaA[17] and the β-porphyranases PorA and PorB from *Z. galactanivorans*. Phylogenetic trees were derived from this refined alignment using the maximum likelihood method with the program PhyML[40]. The reliability of the trees was always tested by bootstrap analysis using 100 resamplings of the data set. The trees were displayed with MEGA 3.1[41].

30. Boraston, A. B., Bolam, D. N., Gilbert, H. J. & Davies, G. J. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem. J.* 382, 769–781 (2004).
31. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* 41, 207–234 (2005).
32. Kabsch, W. Evaluation of single-crystal X-ray-diffraction data from a position-sensitive detector. *J. Appl. Cryst.* 21, 916–924 (1988).
33. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* 58, 1772–1779 (2002).

34. Potterton, E., Briggs, P., Turkenburg, M. & Dodson, E. A graphical user interface to the CCP4 program suite. *Acta Crystallogr. D* **59,** 1131–1137 (2003).
35. Leslie, A. G. W. & Powell, H. R. in *Evolving Methods for Macromolecular Crystallography* Vol. 245 41–51 (Springer, 2007).
36. Starr, C. M., Masada, R. I., Hague, C., Skop, E. & Klock, J. C. Fluorophore-assisted carbohydrate electrophoresis in the separation, analysis, and sequencing of carbohydrates. *J. Chromatogr. A* **720,** 295–321 (1996).
37. Katoh, K., Misawa, K., Kuma, K.-i. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30,** 3059–3066 (2002).
38. Gouet, P., Robert, X. & Courcelle, E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res.* **31,** 3320–3323 (2003).
39. Michel, G. *et al.* The κ-carrageenase of *P. carrageenovora* features a tunnel-shaped active site: a novel insight in the evolution of Clan-B glycoside hydrolases. *Structure* **9,** 513–525 (2001).
40. Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52,** 696–704 (2003).
41. Kumar, S., Tamura, K. & Nei, M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief. Bioinform.* **5,** 150–163 (2004).