Environmental Microbiology
December 2016, Volume 18 Issue 12 Pages 4456-4470
https://doi.org/10.1111/1462-2920.13429
https://archimer.ifremer.fr/doc/00488/59941/

Polysaccharide utilisation loci of *Bacteroidetes* from two contrasting open ocean sites in the North Atlantic

Bennke Christin M. [1, 6], Krueger Karen [1], Kappelmann Lennart [1], Huang Sixing [2], Gobet Angelique [3], Schueler Margarete [4], Barbe Valerie [5], Fuchs Bernhard M. [1], Michel Gurvan [3, *], Teeling Hanno [1, *], Amann Rudolf I. [1, *]

[1] Max Planck Inst Marine Microbiol, Dept Mol Ecol, Celsiusstr 1, D-28359 Bremen, Germany.
[2] Leibniz Inst DSMZ German Collect Microorganisms &, Inhoffenstr 7B, D-38124 Braunschweig, Germany.
[3] UPMC Univ Paris 06, Sorbonne Univ, Integrat Biol Marine Models Stn Biol Roscoff, CNRS,UMR 8227,CS 90074, F-29688 Roscoff, Bretagne, France.
[4] Univ Bayreuth, Biol Elektronenmikroskopie B1, Univ Str 30, D-95447 Bayreuth, Germany.
[5] CEA, Inst Genom Genoscope, Lab Biol Mol Etud Genom, 2 Rue Gaston Cremieux, F-91057 Evry, France.
[6] Leibniz Inst Balt Sea Res Warnemunde, Seestr 15, D-18119 Rostock, Germany.

* Corresponding authors : Gurvan Michel, email address : gurvan.michel@sb-roscoff.fr ; Hanno Teeling, email address : hteeling@mpi-bre-men.de ; Rudolf I. Amann Ruramann@mpi-bremen.de

Abstract :

Marine Bacteroidetes have pronounced capabilities of degrading high molecular weight organic matter such as proteins and polysaccharides. Previously we reported on 76 Bacteroidetes-affiliated fosmids from the North Atlantic Ocean's boreal polar and oligotrophic subtropical provinces. Here, we report on the analysis of further 174 fosmids from the same libraries. The combined, re-assembled dataset (226 contigs; 8.8 Mbp) suggests that planktonic Bacteroidetes at the oligotrophic southern station use more peptides and bacterial and animal polysaccharides, whereas Bacteroidetes at the polar station (East-Greenland Current) use more algal and plant polysaccharides. The latter agrees with higher abundances of algae and terrigenous organic matter, including plant material, at the polar station. Results were corroborated by in-depth bioinformatic analysis of 14 polysaccharide utilisation loci from both stations, suggesting laminarin-specificity for four and specificity for sulfated xylans for two loci. In addition, one locus from the polar station supported use of nonsulfated xylans and mannans, possibly of plant origin. While peptides likely represent a prime source of carbon for Bacteroidetes in open oceans, our data suggest that as yet unstudied clades of these Bacteroidetes have a surprisingly broad capacity for polysaccharide degradation. In particular, laminarin-specific PULs seem widespread and thus must be regarded as globally important.

Introduction

Members of the *Bacteroidetes* phylum are abundant in marine habitats, both in coastal regions (Alonso *et al.*, 2007; Teeling *et al.*, 2012; Teeling *et al.*, 2016) and in the open ocean (Schattenhofer *et al.*, 2009; Gómez-Pereira *et al.*, 2010). They occur free-living in the water column as well as attached to particles (DeLong *et al.*, 1993; Bennke *et al.*, 2013). Members of the *Bacteroidetes* are known to be involved in the degradation of high molecular weight dissolved organic matter (HMW-DOM), such as polysaccharides and proteins (Cottrell and Kirchman, 2000; Cottrell *et al.*, 2005; Thomas *et al.*, 2011; Fernández-Gómez *et al.*, 2013). For example the analysis of the first genome of a marine representative of the bacteroidetal class *Flavobacteriia*, '*Gramella forsetii'* revealed high numbers of peptidase and glycoside hydrolase (GH) genes and thus a high proteolytic and glycolytic potential (Bauer *et al.*, 2006). Similar adaptations have been found in the genomes of other marine *Flavobacteriia*, such as for *Polaribacter dokdonensis* MED152 (González *et al.*, 2008), *Robiginitalea biformata* HTCC2501 (Oh *et al.*, 2009), *Formosa agariphila* KMM 3901 (Mann *et al.*, 2013), and *Polaribacter* spp. Hel1_33_49 and Hel1_85 (Xing *et al.*, 2015). Metagenomic analyses also support the view of marine *Bacteroidetes* as specialists for HMW-DOM (e.g. Gómez-Pereira *et al.*, 2010; Teeling *et al.*, 2012; Teeling *et al.*, 2016).

The extent to which *Bacteroidetes* specialize on macromolecular substrates varies considerably. This is reflected in a broad spectrum of CAZyme and peptidase gene frequencies in *Bacteroidetes* genomes. Planktonic *Bacteroidetes* such as '*G. forsetii'* KT0803 tend to have lower (40 GH and 116 peptidase genes; Bauer *et al.*, 2006) and algae-associated *Bacteroidetes* such as *F. agariphila* KMM 3901 higher CAZyme and peptidase gene numbers (88 GH and 129 peptidase genes; Mann *et al.*, 2013).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Bacteroidetes of the human gut are particularly CAZyme-rich with an average of around 130 GHs per genome (El Kaoutari *et al.*, 2013).

The capacity of *Bacteroidetes* for the degradation of polysaccharides is often encoded in distinct polysaccharide utilization loci (PULs). PULs are operons or regulons of genes that encode the machinery for the concerted detection, hydrolysis and uptake of a dedicated polysaccharide or class of polysaccharides (e.g. Martens *et al.*, 2011). The starch utilization system (*susA-susG*; *susR*) of the human gut symbiont *Bacteroides thetaiotaomicron* was the first described PUL (Anderson and Salyers, 1989; Shipman *et al.*, 2000). PULs always include an outer membrane transport protein homologous to SusC. This SusC-like protein functions as receptor of the TonB uptake system. In *Bacteroidetes*, this TonB-dependent receptor is usually co-located with an outer membrane lipoprotein homologous to SusD (Reeves *et al.*, 1997; Shipman *et al.*, 2000; Cho and Salyers, 2001; Bjursell *et al.*, 2006; Martens *et al.*, 2011). SusD was shown to bind amylose helices, and to keep starch close to the cell surface of *B. thetaiotaomicron* during degradation (Koropatkin *et al.*, 2008). SusD-like proteins thus define a novel class of carbohydrate-binding proteins and according to current knowledge are unique to the *Bacteroidetes* phylum (Thomas *et al.*, 2011). Within PULs *susC* and *susD* homologs can be co-located with genes coding for glycoside hydrolases, carbohydrate esterases (CEs), carbohydrate binding modules (CBMs), polysaccharide lyases (PLs) and proteins with auxiliary functions (AA). These so-called carbohydrate-active enzymes (CAZymes) are classified in the CAZy database (Cantarel *et al.*, 2009; Lombard *et al.*, 2014). PULs of marine *Bacteroidetes* are frequently found to encode also sulfatases (e.g. Bauer *et al.*, 2006; Thomas *et al.*, 2011; Gómez-Pereira *et al.*, 2012; Mann *et al.*, 2013; Xing *et al.*, 2015), because in contrast to their land plant counterparts polysaccharides from marine algae are often sulfated (e.g. ulvans, agars, carrageenans, porphyran, fucans). So far

most functional studies on PULs have been conducted for land plant polysaccharide-specific PULs in human gut bacteria, for example recently for xyloglucan decomposition by human gut *Bacteroidetes* (Larsbrink *et al.*, 2014). Only few PULs have been characterized for polysaccharides of marine origin, such as agar/porphyran-specific PULs (Hehemann *et al.*, 2012a) that have been laterally transferred from marine *Bacteroidetes* to *Bacteroidetes* of the human gut (Hehemann *et al.*, 2010; Hehemann *et al.*, 2012b) and alginate-specific PULs (Thomas *et al.*, 2012). Notably, alginate induction experiments with *Zobellia galactanivorans* DsiJ^T demonstrated that its alginate-specific PUL is a genuine operon (Thomas *et al.*, 2012). A recent proteomic study on the coastal marine bacteroidetes *'Gramella forsetii*' KT0803 confirmed expression of proteins encoded by a homologous alginate-specific PUL, and identified an additional laminarin-induced PUL (Kabisch *et al.*, 2014). The latter was also shown to be present and inducible by laminarin in a proteomic study of the coastal marine bacteroidetes *Polaribacter* sp. Hel1_33_49 (Xing *et al.*, 2015). These findings notwithstanding, we still have little knowledge on the PUL repertoire and associated degradation potential of marine *Bacteroidetes*, in particular for those thriving in the mostly oligotrophic open oceans.

In order to gain insights into the genetic capacities for polysaccharide degradation of as yet uncultured open ocean *Bacteroidetes*, we constructed fosmid metagenome libraries from two contrasting provinces of the North Atlantic (Fig. S1, Table S1) sampled in late September 2006 (Gómez-Pereira *et al.*, 2010). One library of 35,000 fosmids was constructed from surface water taken in the East Greenland Current (station 3) of the Boreal Polar region (BPLR), and a second of 50,000 fosmids was constructed from surface water collected at station 18 (S18) close to the Azores in the North Atlantic Subtropical region (NAST). Both libraries were screened with a PCR assay targeting the 16S rRNA gene with *Bacteroidetes*-specific primers (CF319 and CF967). A total of 13 (S3)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

and 15 (S18) fosmids with 16S rRNA genes were identified and sequenced (Gómez-Pereira *et al.*, 2012). Subsequently, we end-sequenced 16,938 S3 and 16,255 S18 fosmids (avg. length: 623 bp). Use of combined results from the end-sequences' tetranucleotide frequency analysis and BLAST and HMMer searches of encoded genes allowed identification of additional fosmids of possible bacteroidetal origin. In a previous study we reported on the analysis of the first 76 fully sequenced *Bacteroidetes* fosmids (Gómez-Pereira *et al.*, 2012). This analysis revealed that *Bacteroidetes* from both regions had an unexpectedly high capacity for polymer degradation in view of the overall nutrient depletion of open oceans. The analysis also suggested that *Bacteroidetes* in the more oligotrophic southern region might be more adapted towards the degradation of proteins and peptidoglycan than to polysaccharides of algal origin (Gómez-Pereira *et al.*, 2012).

Here we present the analysis of 174 new *bona fide Bacteroidetes* fosmids from the BPLR (S3: 95) and NAST (S18: 79) of the Northern Atlantic, which extends the initial dataset to a total of 250 fosmids. Re-assembly yielded 226 contigs, which we analyzed in terms of peptidases, CAZymes and putative PULs with a special focus on possible polysaccharide substrates, and on the question as to whether differing oceanic provinces select for *Bacteroidetes* clades with different CAZyme repertoires.

Results and Discussion

Characterization of the dataset

The initial 76 fosmids (S3: 40; S18: 36) were pooled with 154 newly sequenced putative *Bacteroidetes* fosmids (S3: 84; S18: 70) and re-assembled. This way, 107 contigs were obtained from S3 and 95 contigs from S18. Some of the new fosmids were selected for sequencing due to high similarity of their end-sequences to previously sequenced fosmids. Thereby it was possible to extend some of the initial fosmids and to obtain assemblies of

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

up to 85.4 kbp (S3) and 72.2 kbp (S18). In addition, 14 putative PULs were retrieved that were not obtained before. Additional fosmids (S3: 11; S18: 9) were selected in order to possibly extend these PUL-carrying fosmids, which after assembly yielded additional 24 contigs (S3: 14; S18: 10). In summary, the total dataset comprised 250 fosmids (S3: 135; S18: 115) that were re-assembled to 226 contigs (S3: 121; S18: 105) of 8.8 Mbp (S3: 4.7 Mbp; S18: 4.1 Mbp).

Based on gene content 96% (S3) and 92% (S18) of the contigs affiliated with *Bacteroidetes.* Thus the error of selecting *Bacteroidetes* fosmids based on information from combined ~1.4 kbp Sanger sequenced end sequences was comparable to the ~5% error of PCR-based screening (Gómez-Pereira *et al.*, 2012). Sequenced non-*Bacteroidetes* contigs affiliated with the *Planctomycetes-Verrucomicrobia-Chlamydia* (PVC)-cluster and with *Proteobacteria*. On class level 95% of the *Bacteroidetes* contigs affiliated with *Flavobacteriia* at S3 and 94% at S18, of which 96% affiliated with the family *Flavobacteriaceae* at both stations. On genus level (Fig. 1) S3 contigs affiliated most frequently with *Polaribacter* (39)%, *Flavobacterium* (13%), *Dokdonia* (6%) and *Gramella* (6%), whereas S18 contigs most frequently affiliated with *Dokdonia* (25%), *Leeuwenhoekiella* (17%), *Flavobacterium* (11%) *Robiginitalea* (8%), *Polaribacter* (7%), and *Croceibacter* (7%). The numbers obtained for *Polaribacter* are in good agreement with *in situ Polaribacter* abundances that were previously determined by catalyzed reporter deposition fluorescence *in situ* hybridization (CARD-FISH) of the same samples (Gómez-Pereira *et al.*, 2010). However, *Gramella* was not detected using CARD-FISH at both stations, and *Leeuwenhoekiella* and *Dokdonia* were only detected at station 18 with abundances below 1% (Gómez-Pereira *et al.*, 2010). Such discrepancies are expected, since our taxonomic affiliations are based on gene BLASTp and HMMer database similarity searches and thus biased towards publically available sequenced *Flavobacteriia*.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Bacteroidetes' peptidases and CAZymes

The numbers of predicted peptidase genes agree by and large with previously published values of Gómez-Pereira *et al.* (2012). The dataset from NAST station 18 had significantly higher peptidase frequencies than the one from BPLR station 3 (Fig. 2A), and in both cases, peptidase frequencies exceeded those of glycolytic CAZymes (GHs, CEs, PLs) by a factor of 1.7 and 2.9, respectively (S3: 25.5 Mbp^{-1} vs. 14.7 Mbp^{-1}; S18: 37.4 Mbp^{-1} vs. 13.1 Mbp^{-1}; Table 1; Fig. 2B, C). Quantitative comparison of peptidase and CAZyme gene numbers based on automatic predictions may contain a certain amount of error due to the involvement of different databases of different sizes and different E-value thresholds. However, our findings are in agreement with the observation that most sequenced genomes of marine *Bacteroidetes* feature higher numbers of peptidase than CAZyme genes (Fernández-Gómez *et al.*, 2013; Xing *et al.*, 2015). Especially planktonic *Bacteroidetes* with small to average-sized genomes tend towards higher peptidase:CAZyme ratios. In contrast, alga-associated *Bacteroidetes* species feature mostly CAZyme-rich large genomes where the combined number of CAZyme genes can exceed those of peptidase genes (Xing *et al.*, 2015). Thus, higher peptidase to CAZyme ratios would be expected for planktonic *Bacteroidetes* from open ocean sites, in particular at a more oligotrophic site like station 18. Our results indicate that proteins (and amino acids) contribute substantially to carbon and nitrogen uptake in open ocean *Bacteroidetes*. The number of peptidase families was also higher at station 18 (S3: 44; S18: 55) as were the frequencies of some individual families such as C40, C44, M38, M42, S09X, S12 and S54, whereas others such as M16B, M50B, S08A and S45 family peptidases were more frequent at S3 (Fig. 2B).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Genes for polysaccharide binding (*susD*) and GH genes exhibited about equal frequencies in both datasets, while TonB-dependent receptor and sulfatase genes were 1.6 to 1.7-fold more frequent in S3 than in S18 contigs (5.1 vs. 3.2 and 6.4 vs. 3.7 Mbp^{-1}; Table 1; Fig. 2A). The latter indicates a higher prevalence of sulfated algal polysaccharides at the polar station S3, which agrees with higher chlorophyll *a* measurements at this station (S3: 0.7 µg l^{-1}; S19: <0.1 µg l^{-1}; Gómez-Pereira *et al.*, 2010). CBM-containing genes were three times more frequent at NAST station 18 than at BPLR station 3 (3.9 vs. 1.3 Mbp^{-1}). Most of them belonged to the peptidoglycan- or chitin-binding CBM50 family (Fig. 2C). This agrees with the analysis of the initial dataset by Gómez-Pereira (2012), who reported higher numbers of peptidoglycan degradation genes at station 18 versus station 3.

Among the most frequent GH families (Fig. 2D), some were about equally represented in contigs from both stations, e.g. GH16 and GH23, whereas for example GH3 (a family comprising diverse functions) and GH92 (a family comprising mostly alpha-mannosidases) were more frequent in the S3 dataset than on S18 contigs (GH3: 1.9 vs. 1.2 Mbp^{-1}; GH92: 1.7 vs. 0.5 Mbp^{-1}). In the less abundant families, GH2 members were found at higher frequencies in S3 contigs (1.1 vs. 0.2 Mbp^{-1}), whereas families GH73, GH5 and GH43 were found at higher frequencies in S18 contigs (GH73: 0.2 vs. 1.0 Mbp^{-1}; GH5: 0.2 vs. 0.7 Mbp^{-1}; GH43: 02. vs. 0.7 Mbp^{-1}). Among the GH families with at least two members in either of the datasets GH106 (0.6 Mbp^{-1}) and GH109 (0.4 Mbp^{-1}) were found only at station 3, GH13 (1.0 Mbp^{-1}) and GH65 (0.5 Mbp^{-1}) only at station 18, and ten GH families were found at both stations (GH2, 3, 5, 10, 16, 23, 43, 73, 92, 113). Within the entire dataset 14 putative PULs were detected (S3: 6 and S18: 8; Fig. 3, 4) comprising 40 GHs of 17 families. In order to gain insights on possible polysaccharide substrates we conducted in-depth manual annotation of these PULs (Table S2).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Commonalities between PULs from both stations

GH16 was among the most frequent GH families in the dataset (Fig. 2D). Five GH16 genes were found in four of the putative PULs, namely on contigs VISS3_015 and VISS3_033 from station S3 (Fig. 3) and VISS18_021 and VISS18_090 from station S18 (Fig. 4). An extracellular location was predicted for all corresponding GH16 proteins (Table S3). The family GH16 comprises various enzymatic activities (Michel *et al.*, 2001; Eklöf *et al.*, 2013; Lombard *et al.*, 2014) and thus phylogenetic analyses are required to determine specificities of GH16 members. Such analyses suggested that the encoded GH16 enzymes are beta-1,3-glucanases usually referred to as laminarinases (Fig. S2). These laminarinases encompass enzymes that act on different types of biologically unrelated beta-1,3-glucans, and only few of these enzymes are specific for genuine laminarin (Labourel *et al.*, 2014). Laminarin (a beta-1,3-glucan with occasional beta-1,6 branching) is the storage polysaccharide of brown algae (Michel *et al.*, 2010) and of diatoms (known as chrysolaminarin; Beattie *et al.*, 1961) and belongs to the most abundant polysaccharides on Earth. Four of the GH16 enzymes (S3: ORFs VISS3_015_23, VISS3_033_04; S18: ORFs VISS18_021_09, VISS18_090_12) belong to a clade that contains ZgLamA$_{GH16}$ from *Z. galactanivorans* DsiJ^{T} (Fig. S2). ZgLamA$_{GH16}$ features an extra loop leading to a bent active site that provides high specificity for genuine algal beta-1,3-glucans (Labourel *et al.*, 2014). The fifth GH16 on contig VISS3_015 (ORF VISS3_015_21) clustered with two functionally uncharacterized GH16 from *Flavobacterium* species. It might be distantly related to the clades containing ZgLamB and ZgLamC (Fig. S2), which do not possess the characteristic ZgLamA$_{GH16}$ loop and act on beta-1,3-glucans and mixed-linkage (beta-1,3-1,4) glucans (Labourel *et al.*, 2015). ORF VISS3_015_21 might encode a similar broad specificity beta-glucanase.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

The predicted GH16 laminarinase genes in all four PULs are each co-located with a GH3 family gene. Two (ORFs VISS3_015_22, VISS3_033_03) of these four GH3 genes were predicted to code for beta-glucosidases and two (ORFs VISS18_021_08, VISS18_090_13) for beta-glycosidases (Fig. S3). These GH3 enzymes likely hydrolyze terminal beta-D-glucosyl residues from oligo-laminarin.

CAZyme analysis also suggested xylose-rich polysaccharides as potential substrates at both stations. Particularly interesting in this context is the PUL on the extended S3 contig VISS3_016 (VISS3_016 + VISS3_057; Fig. 3). This PUL harbors a gene that codes for a putative modular enzyme with an N-terminal sulfatase and a C-terminal GH10 family xylanase (ORF VISS3_016_05). The physical link between these two enzymatic activities indicates degradation of a sulfated xylose-rich polysaccharide. This is further supported by presence of two sulfatase genes (ORFs VISS3_016_02, VISS3_016_03**)** and a predicted GH3 family beta-1,4-xylosidase gene (ORF VISS3_016_04; Fig. S3) in this PUL. Xylan metabolism GHs and sulfatases were also predicted in the PUL on contig VISS18_012 (Fig. 4): a GH3 family xylan 1,4-beta-xylosidase (ORF VISS18_012_31; Fig. S3), a GH10 family endo-beta-1,4-xylanase (ORF VISS18_012_17), and a GH43 family beta-xylosidase/alpha-L-arabinofuranosidase (ORF VISS18_012_15). This PUL is more complex than the PUL from VISS3_016, since it codes for additional CAZymes: a GH10 (ORF VISS18_012_32), a GH30 (ORF VISS18_012_22), a PL9 polysaccharide lyase (ORF VISS18_012_18), a CE1 carbohydrate esterase (ORF VISS18_012_09), and no less than six sulfatases (ORFs VISS18_012, _19, _20, _21, _29, _30, _32). Adjacent to the GH43 and GH10 genes a xyloside transporter gene (*xynT*) was predicted, which might transport xyloside into the cytoplasm, where it is further converted to xylulose-5-phosphate via xylose and xylulose. The respective genes *xylA* (xylose-isomerase) and *xylB* (xylulose-kinase) were identified downstream of the PUL. Neutral xylan occurs in red and green

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

algae (Popper *et al.*, 2011), and sulfated xylan has been found in the red macroalga *Palmaria palmata* (Deniaud *et al.*, 2003). Likewise, exopolysaccharides (EPS) of some diatoms and bacteria are sulfated and rich in xylose.

The predicted capacity to decompose laminarin and xylan or xylose-rich polysaccharides agrees with findings of Arnosti *et al.* (2012), who demonstrated the potential for laminarin and xylan hydrolysis using fluorescein-labeled polysaccharides at BPRL station 3 and NAST station 19 of the same cruise (Fig. S1). Laminarin and sulfated xylose-rich polysaccharides are probably more prevalent at the northern BPLR station 3 than at NAST station 18, since S3 is located at the lower border of the East Greenland Current, which transports cold, low saline, but nutrient- and phytoplankton-rich waters from the Arctic Ocean alongside the eastern coast of Greenland southwards (Bersch, 1995), while the NAST is a typical oligotrophic "blue" ocean (Longhurst, 2006).

Additional PULs from BPLR station 3

The extended contig VISS3_113 (VISS3_041 + VISS3_113; 64.5 kbp) encodes a PUL with two predicted GH106 alpha-rhamnosidase genes (ORFs VISS3_113_01, VISS3_113_16). Alpha-L-rhamnosidases are known to cleave terminal alpha-L-rhamnose from cell wall polysaccharides of plants (in rhamnogalacturonans) and green algae (ulvans, a family of sulfated xylorhamnoglucuronans) (Naumoff and Dedysh, 2012; Lahaye and Robic, 2007; Martin *et al.*, 2014; Popper *et al.*, 2011). This PUL also contains two carbohydrate sulfatase genes (ORFs VISS3_113_14, VISS3_113_17) and a non-classified GH (ORF VISS3_113_10), which could act in synergy with the GH106 enzymes in the degradation of sulfated rhamnose-containing polysaccharides.

Contig VISS3_069 has been classified as *Psychroflexus*-related. Known *Psychroflexus* spp. are psychrophilic, colonize surfaces of sea-ice diatoms (Sullivan and Palmisano,

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

1984; Bowman *et al.*, 1998), and use a rather narrow range of substrates, possibly obtained from diatom EPS (Malinsky-Rushansky and Legrand, 1996). Contig VISS3_069 harbors a putative PUL with genes involved in xylan metabolism. There is a putative CE2 family acetyl-xylan esterase (ORF VISS3_069_11), involved in deacetylation of xylans and xylo-oligosaccharides, and a protein of unknown function containing a CBM9, a module which has so far been only found in xylanases (Lombard *et al.*, 2014). This PUL also codes for enzymes that likely take part in degradation of mannan or mannose-rich glycoproteins: two putative GH92 family alpha-1,2-mannosidases (ORFs VISS3_069_22, VISS3_069_28), a putative GH20 hexosaminidase (ORF VISS3_069_14), which may act on N-acetylmannosamine (Senoura *et al.*, 2011), and a GH130 family enzyme (ORF VISS3_069_26). The GH130 family comprises beta-1,4-mannooligosaccharide phosphorylase, an enzyme that is involved in a novel mannan catabolic pathway (Senoura *et al.*, 2011). Since this PUL lacks any obvious endo-polysaccharidase it is probably incomplete. Xylans and mannans are part of plant cell walls, but they are also found in red algae (Popper *et al.*, 2011) and they play an important structural role in diatom cell walls (Hecky *et al.*, 1973). However, land plant biomass is a more likely source for non-sulfated xylans and mannans at the BPLR station 3, since the Arctic Ocean water that is transported with the East Greenland Current towards station S3 is 7-fold to 16-fold richer in terrigenous dissolved organic matter than the Atlantic and Pacific Oceans (Benner *et al.*, 2005) and includes high amounts of land plant material such as driftwood (Hellmann *et al.*, 2013).

Contig VISS3_097 has been classified as *Cellulophaga*-related. *Cellulophaga* species are known to be associated with diatoms and macrophytes from cold marine waters, e.g. *C. algicola* (Bowman, 2000). This PUL on this contig contains a sulfatase (ORF VISS3_097_09), two putative GH92 family alpha-1,2-mannosidases (ORFs

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

VISS3_097_13, VISS3_097_22) and a putative GH32 family levanase (ORF VISS3_097_05). The latter might hydrolyze 2,6-beta-D-fructofuranosidic linkages in 2,6-beta-D-fructans. Fructans (or levans for bacteria) can be synthesized by green algae or bacteria. Bacterial fructans are produced extracellularly and generally composed of beta-2,6-linked fructosyl residues linked to a terminal glucose, such as for example in *Lactobacillus* and *Streptococcus* species (Corrigan and Robyt, 1979; Hendry, 1993; Van Geel-Schutten *et al.*, 1999). Therefore, this PUL might be dedicated to the degradation of sulfated EPS containing mannose and/or fructose residues.

Contig VISS3_052 did not contain the characteristic *susCD* gene pair of PULs, but it might constitute a partial PUL as indicated by the presence of five sulfatases (ORFs VISS3_052, _20, _21, _22, _25, _26), one putative sugar transporter, a carbohydrate kinase, two sugar isomerases and a fructose-1,6-bisphosphate aldolase gene. Two genes (ORFs VISS3_052_12, VIS3_052_24) on this contig share remote similarities with glycoside hydrolases, but no sufficient similarity to be annotated as such.

Additional PULs from NAST station 18

The PUL on contig VISS18_034 encodes two predicted GH13 family alpha-amylases (ORFs VISS18_034_02, VISS18_034_03) from distinct subfamilies (Stam *et al.*, 2006). In phylogenetic reconstruction (Fig. S5), ORF VISS18_034_03 clustered with two GH13_7 subfamily alpha-amylases from *Thermococcus* species, suggesting lateral gene transfer (LGT). In contrast, ORF VISS18_034_02 belongs to the GH13_20 subfamily that includes cyclomaltodextrinase, neopullulanase and maltogenic amylase. In comparison to classical alpha-amylases, these enzymes, and also ORF VISS18_034_02, feature an extra domain that participates in dimer formation (Lee *et al.*, 2002; Stam *et al.*, 2006). Cyclomaltodextrinases effectively hydrolyze cyclomaltodextrin, a circular sugar derived

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

from starch degradation, whereas the degradation of starch and pullulan is less effective (Lee *et al.*, 2002). The PUL also contains a putative GH31 family alpha-glucosidase (ORF VISS18_034_01), which hydrolyzes the oligosaccharides released by alpha-amylases. Therefore, this PUL likely targets an alpha-1,4-glucan, which is for example produced by some bacteria as storage compound during stationary growth (Preiss, 1984; Field *et al.*, 1998). Usage of bacterial polysaccharides by *Bacteroidetes* may be of higher relative importance at NAST station S18 since algae were less abundant than at the BPLR station S3 (Table S1).

The PUL on contig VISS18_001 encodes a predicted GH32 family beta-fructofuranosidase (ORF VISS18_001_23). These enzymes hydrolyze terminal non-reducing beta-D-fructofuranoside residues in beta-D-fructofuranosides (for instance sucrose). The PUL also encodes a putative GH65 family maltose phosphorylase (ORF VISS18_001_20). These enzymes add phosphate to maltose, resulting in D-glucose and beta-D-glucose-1-phosphate.

The PUL on contig VISS18_040 contains enzymes involved in starch and sucrose metabolism, such as a predicted GH65 family maltose phosphorylase (ORF VISS18_040_30) and GH43 family xylosidase/arabinofuranosidase (ORF VISS18_040_21). This PUL also contains a putative GH20 beta-N-acetylhexosaminidase (ORF VISS18_040_01), a GH5_13 glycoside hydrolase that may target beta-mannan (ORF VISS18_040_04; Fig. S4; Aspeborg *et al.*, 2012), a putative GH109 alpha-N-acetylgalactosaminidase (ORF VISS18_040_13), and a GH of unknown specificity (ORF VISS18_040_03). Furthermore, the PUL contains a predicted GH78 family alpha-L-rhamnosidase (ORF VISS18_040_07) and four sulfatases (ORFs VISS18_040_12, _14, _16, _20; family S1 and subfamilies 8, 9, 16, 20; Table S2), similar to the PUL of S3

contigs VISS3_113 + VISS3_041. These enzymes may be involved in the hydrolysis of sulfated polysaccharides (e.g. algal ulvans).

Annotations did not provide sufficient information for hypotheses on possible substrates for PUL-containing contigs VISS18_065 and VISS18_083 + VISS18_044 (Fig. 4). Contig VISS18_083 + VISS18_044 harbors a putative CBM32-containing GH92 alpha-1,2-mannosidase gene (ORF VISS18_083_32) and a putative CE10 gene, whereas contig VISS18_065 featured a *susC-susD* gene pair, but otherwise no matches to the CAZy database.

Comparative analysis of PULs

Some of the analyzed PUL-containing contigs share regions of high DNA sequence similarity. Such homology was observed not only among PULs from the same sampling station, but also between PULs from both stations. For instance, two regions of contigs VISS3_052 and VISS3_041 + VISS3_113 are highly similar (Fig. S6A). The first (VISS3_052: 12.89-15.82 kbp; VISS3_041 + VISS3_113: 51.28-54.32 kbp) comprises two (L-arabinose isomerase, sulfatase) and the second (VISS3_052: 23.54-33.45 kbp; VISS3_041 + VISS3_113: 12.00-22.81 kbp with a small insertion at 19.54-20.10 kbp) seven genes (membrane protein, L-lactate dehydrogenase, carbohydrate kinase, sugar isomerase, short-chain dehydrogenase, fructose-1,6-bisphosphatase, transcriptional regulator - GntR family). Likewise, contigs VISS3_069 (26.60-28.87 kbp) and VISS3_097 (13.11-15.36 kbp) share a region of high DNA similarity that encodes a GH92 family enzyme (Fig. S6B).

BPLR contigs VISS3_016 + VISS3_057 and the NAST contig VISS18_012 have a highly similar region harboring homologous GH3 family genes, neighbored by three non-homologous sulfatase genes on both contigs (Fig. S6C). Similarly four of the putative

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

laminarinase-containing contigs from both stations (BPLR: VISS3_015, VISS3_033; NAST: VISS18_021, VISS18_090) exhibited a high level of sequence conservation (Fig. S6D).

Comparative genomics suggests that lateral gene transfer (LGT) is frequent among human gut *Bacteroidetes* (Coyne *et al.*, 2014), including exchange of CAZymes and entire PULs (Hehemann *et al.*, 2010; Hehemann *et al.*, 2012b; Martens *et al.*, 2014). We found homologous regions in more than half of the PULs that we analyzed, which suggests that such LGT events might occur also rather frequently among marine *Bacteroidetes*. These analyses also suggest that parts of PULs can be laterally transferred and act as recombination modules in PUL evolution. Whether entire PULs can be laterally transferred from *Bacteroidetes* to non-*Bacteroidetes* is an open question. The fact that *susD* genes have so far only been found in *Bacteroidetes* suggests some type of recombination barrier that prevents establishment of a *Bacteroidetes*-like SusC - SusD interaction in other phyla.

Conclusions

Although a fosmid-based approach is more laborious and comes at a much higher cost per base than a shotgun metagenome approach, it has the advantage of being targeted, since fosmid libraries can be end-sequenced and screened for clones from dedicated taxa, and it is guaranteed to provide sequences that are long enough for the study of larger gene arrangements such as PULs.

In their initial study, Gómez-Pereira *et al.* (2012) concluded that *Bacteroidetes* at the BPLR station S3 were richer in polysaccharide degradation genes than at NAST station S18, who in turn had higher peptidase gene frequencies. Our analysis of the extended fosmid dataset confirmed higher peptidase frequencies at NAST, but higher CAZyme frequencies at the BPLR station could not be substantiated. In the present study, we

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

particularly focused on in-depth manual annotation of CAZymes and CAZyme genome clusters so as to provide a more substantial idea of their activity beyond automated assignment to diverse enzyme families based on bioinformatic tools. Although frequencies of CAZymes were not different between stations, composition of the respective CAZyme sets clearly was, and the prevalence of sulfatases was notably higher at BPLR than at NAST. This agrees with results of Arnosti *et al.* (2012), who found the microbial community at the BPLR station 3 to be capable of degrading the sulfated polysaccharides chondroitin and fucoidan at a faster rate than at the NAST station 19 of the same cruise. Fittingly, co-localizations of sulfatases and GHs were found on six of the 121 S3 contigs (including fosmids S3-860 and S3_DL_C5 reported by Gómez-Pereira *et al.* (2012) not shown in Fig. 3), but only on two of the 105 S18 contigs. Sulfated polysaccharides are produced in large quantities by marine algae, which were more abundant at BPLR station 3 than at NAST station 18 (Table S1). This means that the relative contribution of carbon from amino acids/peptides and from non-algal organic matter was higher at NAST station 18 than at BPLR station 3, which is reflected in (i) higher peptidase frequencies, (ii) higher frequencies of GHs for the hydrolysis of bacterial or animal alpha-1,4-glucans, and (iii) a higher prevalence of CBM50 genes that might cleave either bacterial peptidoglycan or animal chitin. Conversely, higher frequencies of GH92 mannosidases at BPLR station 3 support a higher importance of algal- and (see below) plant-derived polysaccharides at this station.

PUL comparisons indicated that sulfated xylan-rich polysaccharides and algal laminarin are possibly among the more frequent polysaccharide substrates for open ocean *Bacteroidetes*. One PUL at each station contained putative xylan-specific genes and sulfatases, suggesting xylan-rich polysaccharide of marine origin as substrates (e.g. from algal EPS). Four out of the 14 PULs in our dataset were likely laminarin-specific. Similar

PULs have been identified in other members of the *Bacteroidetes* (Kabisch *et al.*, 2014), suggesting that such PULs are widespread and of global importance. This underpins the importance of laminarin as substrate in the marine realm, also in the open ocean.

The dataset presented in this study demonstrates that the CAZyme repertoire of *Bacteroidetes* in open ocean sites such as the BPLR station 3 and the NAST station 18 is diverse and even comprises families such as GH10, 43, 78 and 106 that have been suggested to be characteristic for *Bacteroidetes* feeding on land plant biomass (Kolton et al. 2013). Marine plants and algae produce some polysaccharides that are usually found in terrestrial plants. Rhamnogalacturonans and xylans for example are present in land plant hemicelluloses, but rhamnogalacturonans are also constituents of pectins in marine angiosperms, and xylans have also been found as a form of cell covering in some marine algae (Okuda, 2002) or as cell wall components in some diatoms (Wustman *et al.*, 1998; Murray *et al.*, 2007). Presence of the above mentioned GH families, at least at BPLR station 3, may, however, be most likely explained by the station's location within the East Greenland Current, that transports ample terrigenous organic matter including plant material (Benner *et al.*, 2005; Hellmann *et al.*, 2013). This would also explain the finding of a partial PUL with xylan- and mannan-specific genes without sulfatases at this station.

At oligotrophic open-ocean sites algal and bacterial polysaccharides are produced in much lower amounts than at eutrophic sites. Therefore these energy-rich compounds are particularly valuable at open ocean sites, and consequently heterotrophic bacteria exist at such sites that can consume these polysaccharides when they become available. As in other habitats, the *Bacteroidetes* seem to play a key role in such turnover of complex organic matter also in open oceans. It will be up to future systematic studies to inventory the PUL repertoire of marine *Bacteroidetes* in a comprehensive manner and to explore,

which individual PULs are ubiquitously distributed and thus most important in marine habitats.

Experimental Procedures

Study sites and fosmid library preparation

Samples were taken in the North Atlantic Ocean during the VISION cruise MSM03/01 on board of R/V Maria S. Merian in September 2006 (Fig. S1, Table S1). Fosmid metagenome libraries were constructed from surface water of two contrasting oceanic provinces. Samples from station 3 (S3) were collected in the Boreal Polar province (65°52.64' N, 29°56.54' W) and samples from station 18 (S18) in the North Atlantic Subtropical province (34°04.43' N, 30°00.09' W). Libraries of 35,000 (S3) and 50,000 (S18) fosmids were constructed for both sites. Subsequently, 16,938 (S3) and 16,266 (S18) high-quality end sequences were generated by sequencing inserts from both sites using the Sanger technique. Details have been described elsewhere (Gómez-Pereira *et al.*, 2010; Gómez-Pereira *et al.*, 2012).

Selection and sequencing of fosmids

End sequences were mapped on the 76 previously sequenced fosmids from both libraries in order to detect connecting fosmids. Using a sequence identity threshold of 94.5% or higher, 43 (S3) and 27 (S18) connecting fosmid candidates were identified. Twenty of these had the potential to prolong partial PULs on the previously sequenced fosmids. These were sequenced at LGC Genomics (LGC Genomics GmbH, Berlin, Germany) using the 454 FLX Ti platform and assembled using Newbler. Further 104 putative *Bacteroidetes* fosmids were selected based on BLASTx hits of end-sequences to the NCBI non-redundant protein sequence databases with a rank-based evaluation similar as proposed

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

by Podell and Gasterland (Podell and Gaasterland, 2007), phylogenetic reconstructions based on HMMer 3 searches of all-frame translated end sequences against the Pfam v. 25 database (Krause *et al.*, 2008) and an evaluation of end sequence tetranucleotide usage patterns (Teeling *et al.*, 2004). These fosmids and the remaining 50 connecting fosmid candidates were sequenced at Genoscope (Évry Cedex, France) using the 454 FLX Ti platform and assembled using Newbler as described previously (Gómez-Pereira *et al.*, 2010; Gómez-Pereira *et al.*, 2012). The final dataset comprised 250 fosmids.

Fosmid re-assembly

Fosmid sequences were pooled by station and then re-assembled with SeqMan (Lasergene 8 software suite, DNAStar Inc., Madison, WI, USA). The default setting was used. The assembly quality was checked via the program's strategy view option.

Taxonomic classification

Taxonomic affiliation of sequenced fosmids was done as described for end sequences based on combined analysis for all genes of BLASTp hits to the NCBI non-redundant protein database and HMMer 3 hits to the Pfam v. 25 database (Table S4).

Automated gene prediction and annotation

Gene prediction and annotation of all 226 contigs was done via the RAST server (Aziz *et al.*, 2008). The RAST gene calls and annotations of the included 76 published fosmids differed only marginally from the published ones. Results for all contigs were downloaded and subsequently imported into a local installation of the GenDB (v. 2.2) annotation system (Meyer *et al.*, 2003) for curation. CAZymes were annotated based on HMMER searches against the dbCAN database (Yin *et al.*, 2012), BLASTp (Altschul *et al.*, 1990)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

searches against the CAZy database (Cantarel *et al.*, 2009; Lombard *et al.*, 2014) and HMMER searches against the Pfam v. 28 database (Finn *et al.*, 2010) using E-values derived from manual annotations of test data (Table S5). CAZymes were only annotated when at least two of the three database searches yielded positive results. Peptidases were automatically annotated based on batch BLASTp searches against the MEROPS 9.13 databases (Rawlings *et al.*, 2012) using the default E-value cutoff criterion of 10^{-4}. ABC transporter, TonB-dependent receptor, *susD* genes and sulfatase genes were automatically predicted based on HMMer 3 hits to the Pfam v. 28 database at $E \leq 10^{-5}$ using the following profiles: ABC_tran, ABC_membrane, ABC_membrane_2, ABC_membrane_3, ABC_tran_2, ABC2_membrane, ABC2_membrane_2, ABC2_membrane_3, ABC2_membrane_4, ABC2_membrane_5, ABC2_membrane_6, TonB_dep_Rec, SusD, SusD-like, SusD-like_2, SusD-like_3 and sulfatase. Annotated sequences were deposited at NCBI's Genbank (BioSample accessions SAMN04870880 and SAMN04870884).

Manual CAZyme annotation

CAZymes were identified based on homology with a selected subset of characterized enzymes from each CAZyme family. Initial annotations were validated by BLASTp searches against the UniProtKB/SwissProt database as of February 2014 (The UniProt Consortium, 2014) and HMMer searches against the Pfam v. 27 database (Punta *et al.*, 2012). Each CAZyme was assigned to a CAZY family and, when possible, to an EC number. Abundant glycoside hydrolases from the multi-functional families GH3, GH5, GH13, and GH16 were subjected to an in-depth phylogenetic analysis to determine their substrate-specificities. Experimentally characterized proteins (Table S6) were selected from the CAZy database for each activity within a given GH family and aligned to their

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

contig homologs using MAFFT (FFT-NS-i iterative refinement method; BLOSUM62 amino acid substitution matrix) (Katoh and Standley, 2013). These alignments were used to calculate model tests and maximum likelihood trees with MEGA v. 6.0.6 (Kumar *et al.*, 2004) with bootstrapping (100 resamplings). Annotation of ambiguous proteins was refined based on the proximity to characterized proteins in the phylogenetic trees.

Subcellular locations were predicted using CELLO v. 2.5 (Yu *et al.*, 2006), PSORTb v. 3.0.2 (Yu *et al.*, 2010) and HMMer searches against the TIGRfam profile (Selengut *et al.*, 2007) TIGR04183 (Por secretion system C-terminal sorting domain). Only unambiguous consensus predictions were considered as reliable.

Acknowlegements

We thank the Captain and Crew of the FS Maria S. Merian for their support during cruise MSM03/01, J. Waldmann for bioinformatics and R. Hahnke for comparative PUL alignments. A. Gobet and G. Michel were supported by the National Research Agency of the French Government by the “Blue Enzymes” ANR project (ANR-14-CE19-0020-01). This study was funded by the Max-Planck-Society, the German Science Foundation (DFG) and the FP6 EU program Network of Excellence Marine Genomics Europe.

References

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

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403-410.

Anderson, K.L., and Salyers, A.A. (1989) Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* **171**: 3199-3204.

Arnosti, C., Fuchs, B.M., Amann, R., and Passow, U. (2012) Contrasting extracellular enzyme activities of particle-associated bacteria from distinct provinces of the North Atlantic Ocean. *Front Microbiol* **3**: 425.

Aspeborg, H., Coutinho, P.M., Wang, Y., Brumer, H., and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* **12**: 186.

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A. et al. (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

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.

Beattie, A., Hirst, E.L., and Percival, E. (1961) Studies on the metabolism of the *Chrysophyceae*. Comparative structural investigations on leucosin (chrysolaminarin) separated from diatoms and laminarin from the brown algae. *Biochem J* **79**: 531-537.

Benner, R., Louchouarn, P., and Amon, R.M.W. (2005) Terrigenous dissolved organic matter in the Arctic Ocean and its transport to surface and deep waters of the North Atlantic. *Global Biogeochem Cy* **19:** DOI 10.1029/2004GB002398.

Bennke, C.M., Neu, T.R., Fuchs, B.M., and Amann, R. (2013) Mapping glycoconjugate-mediated interactions of marine *Bacteroidetes* with diatoms. *Syst Appl Microbiol* **36**: 417-425.

Bersch, M. (1995) On the circulation of the northeastern North Atlantic. *Deep Sea Research Part I: Oceanographic Research Papers* **42**: 1583-1607.

Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006) Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* **281**: 36269-36279.

Bowman, J.P. (2000) Description of *Cellulophaga algicola* sp. nov., isolated from the surfaces of Antarctic algae, and reclassification of *Cytophaga uliginosa* (ZoBell and Upham 1944) Reichenbach 1989 as *Cellulophaga uliginosa* comb. nov. *Int J Syst Evol Microbiol* **50 Pt 5**: 1861-1868.

Bowman, J.P., McCammon, S.A., Lewis, T., Skerratt, J.H., Brown, J.L., Nichols, D.S., and McMeekin, T.A. (1998) *Psychroflexus torquis* gen. nov., sp. nov., a psychrophilic species from Antarctic sea ice, and reclassification of *Flavobacterium gondwanense* (Dobson et al. 1993) as *Psychroflexus gondwanense* gen. nov., comb. nov. *Microbiology* **144**: 1601-1609.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* **37**: D233-D238.

Cho, K.H., and Salyers, A.A. (2001) Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* **183**: 7224-7230.

Corrigan, A.J., and Robyt, J.F. (1979) Nature of the fructan of *Streptococcus mutans* OMZ 176. *Infect Immun* **26**: 387-389.

Cottrell, M.T., and Kirchman, D.L. (2000) Natural assemblages of marine proteobacteria and members of the *Cytophaga-Flavobacter* cluster consuming low- and high-molecular-weight dissolved organic matter. *Appl Environ Microbiol* **66**: 1692-1697.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Cottrell, M.T., Yu, L., and Kirchman, D.L. (2005) Sequence and expression analyses of *Cytophaga*-like hydrolases in a Western arctic metagenomic library and the Sargasso Sea. *Appl Environ Microbiol* **71**: 8506-8513.

Coyne, M.J., Zitomersky, N.L., McGuire, A.M., Earl, A.M., and Comstock, L.E. (2014) Evidence of extensive DNA transfer between bacteroidales species within the human gut. *MBio* **5**: e01305-14.

Darling, A.E., Mau, B., and Perna, N.T. (2010) progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**: e11147.

DeLong, E.F., Franks, D.G., and Alldredge, A.L. (1993) Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**: 924-934.

Deniaud, E., Fleurence, J., and Lahaye, M. (2003) Interactions of the mix-linked β-(1,3)/β-(1,4)-D-xylans in the cell walls of *Palmaria palmata* (*Rhodophyta*). *J Phycol* **39**: 74-82.

Eklöf, J.M., Shojania, S., Okon, M., McIntosh, L.P., and Brumer, H. (2013) Structure-function analysis of a broad specificity *Populus trichocarpa* endo-β-glucanase reveals an evolutionary link between bacterial licheninases and plant XTH gene products. *J Biol Chem* **288**: 15786-15799.

El Kaoutari, A., Armougom, F., Gordon, J.I., Raoult, D., and Henrissat, B. (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol* **11**: 497-504.

Fernández-Gómez, B., Richter, M., Schüler, M., Pinhassi, J., Acinas, S.G., González, J.M., and Pedrós-Alió, C. (2013) Ecology of marine *Bacteroidetes*: a comparative genomics approach. *ISME J* **7**: 1026-1037.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Field, C.B., Behrenfeld, M.J., Randerson, J.T., and Falkowski, P. (1998) Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**: 237-240.

Finn, R.D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J.E. et al. (2010) The Pfam protein families database. *Nucleic Acids Res* **38**: D211-22.

Gómez-Pereira, P.R., Fuchs, B.M., Alonso, C., Oliver, M.J., van Beusekom, J.E., and Amann, R. (2010) Distinct flavobacterial communities in contrasting water masses of the North Atlantic Ocean. *ISME J* **4**: 472-487.

Gómez-Pereira, P.R., Schüler, M., Fuchs, B.M., 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.

González, J.M., 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 U S A* **105**: 8724-8729.

Hecky, R.E., Mopper, K., Kilham, P., and Degens, E.T. (1973) The amino acid and sugar composition of diatom cell-walls. *Marine biology* **19**: 323-331.

Hehemann, J.H., Correc, G., Barbeyron, T., Helbert, W., Czjzek, M., and Michel, G. (2010) Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**: 908-912.

Hehemann, J.H., Correc, G., Thomas, F., Bernard, T., Barbeyron, T., Jam, M. et al. (2012a) Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J Biol Chem* **287**: 30571-30584.

Hehemann, J.H., Kelly, A.G., Pudlo, N.A., Martens, E.C., and Boraston, A.B. (2012b) Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc Natl Acad Sci U S A* **109**: 19786-19791.

Hellmann, L., Tegel, W., Eggertsson, Ó., Schweingruber, F.H., Blanchette, R., Kirdyanov, A. et al. (2013) Tracing the origin of Arctic driftwood. *J Geophys Res-Biogeo* **118**: 68-76.

Hendry, G.A.F. (1993) Evolutionary origins and natural functions of fructans-a climatological, biogeographic and mechanistic appraisal. *New Phytol* **123**: 3-14.

Kabisch, A., Otto, A., König, S., Becher, D., Albrecht, D., Schüler, M. et al. (2014) Functional characterization of polysaccharide utilization loci in the marine *Bacteroidetes 'Gramella forsetii'* KT0803. *ISME J* **8**: 1492-1502.

Katoh, K., and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **30**: 772-780.

Kolton, M., Sela, N., Elad, Y., and Cytryn, E. (2013) Comparative genomic analysis indicates that niche adaptation of terrestrial *Flavobacteria* is strongly linked to plant glycan metabolism. *PLoS One* **8**: e76704.

Koropatkin, N.M., Martens, E.C., Gordon, J.I., and Smith, T.J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**: 1105-1115.

Krause, L., Diaz, N.N., Goesmann, A., Kelley, S., Nattkemper, T.W., Rohwer, F. et al. (2008) Phylogenetic classification of short environmental DNA fragments. *Nucleic Acids Res* **36**: 2230-2239.

Kumar, S., Tamura, K., and Nei, M. (2004) MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* **5**: 150-163.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Labourel, A., Jam, M., Jeudy, A., Hehemann, J.H., Czjzek, M., and Michel, G. (2014) The beta-glucanase ZgLamA from *Zobellia galactanivorans* evolved a bent active site adapted for efficient degradation of algal laminarin. *J Biol Chem* **289**: 2027-2042.

Labourel, A., Jam, M., Legentil, L., Sylla, B., Hehemann, J.H., Ferrières, V. et al. (2015) Structural and biochemical characterization of the laminarinase ZgLamCGH16 from *Zobellia galactanivorans* suggests preferred recognition of branched laminarin. *Acta Cryst* **D71**: 173-184.

Lahaye, M., and Robic, A. (2007) Structure and functional properties of ulvan, a polysaccharide from green seaweeds. *Biomacromolecules* **8**: 1765-1774.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O. et al. (2014) A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* **506**: 498-502.

Lee, S.C., Gepts, P.L., and Whitaker, J.R. (2002) Protein structures of common bean (*Phaseolus vulgaris*) alpha-amylase inhibitors. *J Agric Food Chem* **50**: 6618-6627.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* D490-5.

Longhurst, A.R. (2006) Ecological Geography of the Sea.

Malinsky-Rushansky, N.Z., and Legrand, C. (1996) Excretion of dissolved organic carbon by phytoplankton of different sizes and subsequent bacterial uptake. *Oceanographic Literature Review* **11**: 1124.

Mann, A.J., Hahnke, R.L., Huang, S., Werner, J., Xing, P., Barbeyron, T. et al. (2013) The genome of the alga-associated marine flavobacterium *Formosa agariphila* KMM 3901T reveals a broad potential for degradation of algal polysaccharides. *Appl Environ Microbiol* **79**: 6813-6822.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P. et al. (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* **9**: e1001221.

Martens, E.C., Kelly, A.G., Tauzin, A.S., and Brumer, H. (2014) The devil lies in the details: how variations in polysaccharide fine-structure impact the physiology and evolution of gut microbes. *J Mol Biol* **426**: 3851-3865.

Martin, M., Portetelle, D., Michel, G., and Vandenbol, M. (2014) Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. *Appl Microbiol Biotechnol* **98**: 2917-2935.

Meyer, F., Goesmann, A., McHardy, A.C., 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.

Michel, G., Chantalat, L., Duee, E., Barbeyron, T., Henrissat, B., Kloareg, B., and Dideberg, O. (2001) The kappa-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.

Michel, G., Tonon, T., Scornet, D., Cock, J.M., and Kloareg, B. (2010) Central and storage carbon metabolism of the brown alga *Ectocarpus siliculosus*: insights into the origin and evolution of storage carbohydrates in Eukaryotes. *New Phytol* **188**: 67-81.

Murray, A., Arnosti, C., De La Rocha, C., Grosart, H.-P., and Passow, U. (2007) Microbial dynamics in autotrophic and heterotrophic seawater mesocosms. II. Bacterioplankton community structure and hydrolytic enzyme activities. *Aquat Microb Ecol* **49**: 123-141.

Naumoff, D.G., and Dedysh, S.N. (2012) Lateral gene transfer between the *Bacteroidetes* and *Acidobacteria*: the case of α-L-rhamnosidases. *FEBS Lett* **586**: 3843-3851.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Nedashkovskaya, O.I., Kim, S.B., Vancanneyt, M., Snauwaert, C., Lysenko, A.M., Rohde, M. et al. (2006) *Formosa agariphila* sp. nov., a budding bacterium of the family *Flavobacteriaceae* isolated from marine environments, and emended description of the genus *Formosa*. *Int J Syst Evol Microbiol* **56**: 161-167.

Oh, H.M., Giovannoni, S.J., Lee, K., Ferriera, S., Johnson, J., and Cho, J.C. (2009) Complete genome sequence of *Robiginitalea biformata* HTCC2501. *J Bacteriol* **191**: 7144-7145.

Okuda, K. (2002) Structure and phylogeny of cell coverings. *J Plant Res* **115**: 283-288.

Podell, S., and Gaasterland, T. (2007) DarkHorse: a method for genome-wide prediction of horizontal gene transfer. *Genome Biol S* **8**: R16.

Popper, Z.A., Michel, G., Herve, C., Domozych, D.S., Willats, W.G., Tuohy, M.G. et al. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* **62**: 567-590.

Preiss, J. (1984) Bacterial glycogen synthesis and its regulation. *Annu Rev Microbiol* **38**: 419-458.

Punta, M., Coggill, P.C., Eberhardt, R.Y., Mistry, J., Tate, J., Boursnell, C. et al. (2012) The Pfam protein families database. *Nucleic Acids Res* **40**: D290-301.

Rawlings, N.D., Barrett, A.J., and Bateman, A. (2012) MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res* **40**: D343-50.

Reeves, A.R., Wang, G.R., and Salyers, A.A. (1997) Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron*. *J Bacteriol* **179**: 643-649.

Schattenhofer, M., Fuchs, B.M., Amann, R., Zubkov, M.V., Tarran, G.A., and Pernthaler, J. (2009) Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environ Microbiol* **11**: 2078-2093.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Wiley-Blackwell and Society for Applied Microbiology

Selengut, J.D., Haft, D.H., Davidsen, T., Ganapathy, A., Gwinn-Giglio, M., Nelson, W.C. et al. (2007) TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes. *Nucleic Acids Res* **35**: D260-4.

Senoura, T., Ito, S., Taguchi, H., Higa, M., Hamada, S., Matsui, H. et al. (2011) New microbial mannan catabolic pathway that involves a novel mannosylglucose phosphorylase. *Biochem Biophys Res Commun* **408**: 701-706.

Shipman, J.A., Berleman, J.E., and Salyers, A.A. (2000) Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. *J Bacteriol* **182**: 5365-5372.

Stam, M.R., Danchin, E.G., Rancurel, C., Coutinho, P.M., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel* **19**: 555-562.

Sullivan, C.W., and Palmisano, A.C. (1984) Sea Ice microbial communities: distribution, abundance, and diversity of ice bacteria in McMurdo Sound, Antarctica, in 1980. *Appl Environ Microbiol* **47**: 788-795.

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

Teeling, H., Fuchs, B.M., Bennke, C.M., Krüger, K., Chafee, M., Kappelmann, L. et al. (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *eLife* 2016;5:e11888.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Teeling, H., Meyerdierks, A., Bauer, M., Amann, R., and Glöckner, F.O. (2004) Application of tetranucleotide frequencies for the assignment of genomic fragments. *Environ Microbiol* **6**: 938-947.

The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). *Nucleic acids research* **42**: D191-D198.

Thomas, F., Barbeyron, T., Tonon, T., Génicot, S., Czjzek, M., and Michel, G. (2012) Characterization of the first alginolytic operons in a marine bacterium: from their emergence in marine *Flavobacteriia* to their independent transfers to marine *Proteobacteria* and human gut *Bacteroides*. *Environ Microbiol* **14**: 2379-2394.

Thomas, F., Hehemann, J.H., Rebuffet, E., Czjzek, M., and Michel, G. (2011) Environmental and gut *Bacteroidetes*: the food connection. *Front Microbiol* **2**: 93.

Van Geel-Schutten, G.H., Faber, Smit, Bonting, Smith, Ten Brink, B. et al. (1999) Biochemical and structural characterization of the glucan and fructan exopolysaccharides synthesized by the Lactobacillus reuteri wild-type strain and by mutant strains. *Appl Environ Microbiol* **65**: 3008-3014.

Wustman, Lind, Wetherbee, and Gretz (1998) Extracellular matrix assembly in diatoms (*Bacillariophyceae*). Iii. Organization Of fucoglucuronogalactans within the adhesive stalks of achnanthes longipes. *Plant Physiol* **116**: 1431-1441.

Xing, P., Hahnke, R.L., Unfried, F., Markert, S., Huang, S., Barbeyron, T. et al. (2015) Niches of two polysaccharide-degrading *Polaribacter* isolates from the North Sea during a spring diatom bloom. *ISME J* **9**: 1410-1422.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., and Xu, Y. (2012) dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* **40**: W445-51.

Yu, C.S., Chen, Y.C., Lu, C.H., and Hwang, J.K. (2006) Prediction of protein subcellular localization. *Proteins* **64**: 643-651.

Yu, N.Y., Wagner, J.R., Laird, M.R., Melli, G., Rey, S., Lo, R. et al. (2010) PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**: 1608-1615.

Figure Legends

Fig. 1. Genus-level taxonomic affiliation of contigs obtained from re-assembled fosmid sequences of metagenomic libraries from station S3 (121 contigs) and S18 (105 contigs).

Fig. 2. Comparison of selected genes on the 121 BPLR station 3 contigs (4.7 Mbp) and the 105 NAST station 18 contigs (4.1 Mbp). (**A**) genes for ABC-transporters, TonB-dependent receptors (TBDRs), SusD-like proteins, CAZymes, sulfatases and peptidases; (**B**) peptidase families; (**C**) CAZyme classes: glycoside hydrolases (GH), carbohydrate esterases (CE), polysaccharide lyases (PL), carbohydrate-binding modules (CBM) and glycosyltransferases (GT); (**D**) GH families ordered by decreasing averages.

Fig. 3. PUL-containing contigs from BPLR station 3. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

Fig. 4. PUL-containing contigs from NAST station 18. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

Tables

Table 1. Comparison of frequencies of genes involved in organic matter degradation and uptake between the S3 and S18 contig datasets

Supporting Information

Fig. S1. Map of the VISION cruise track with sampled stations S2 - S19 and boundaries of the Boreal Polar (BPLR), Arctic (ARCT), North Atlantic Drift (NADR) and North Atlantic Subtropical (NAST) provinces. Symbol colors represent different water masses as published in Gómez-Pereira *et al.* (2010).

Fig. S2. Unrooted phylogenetic tree of the GH16 family glycoside hydrolases present on contigs VISS3_015, VISS3_033, VISS18_021, and VISS18_090. Phylogenetic trees were calculated using the maximum-likelihood (ML) approach implemented in MEGA v. 6.0.6 (Kumar *et al.*, 2004). Bootstrap values (100 resamplings) are indicated by numbers on the tree. The full listing of the aligned proteins is available as supplementary information. Black diamonds correspond to the GH16 protein sequences from the four contigs. Characterized GH16 laminarinases from *Zobellia galactanivorans* (Labourel *et al.*, 2014) are indicated in bold typeface.

Fig. S3. Unrooted phylogenetic tree of the GH3 family glycoside hydrolases present on contigs VISS3_015, VISS3_016, VISS3_033, VISS18_012, VISS18_021, and VISS18_090. Phylogenetic trees were calculated using the maximum-likelihood (ML) approach implemented in MEGA v. 6.0.6 (Kumar *et al.*, 2004). Bootstrap values (100 resamplings) are indicated by numbers on the tree. The full listing of the aligned proteins is available as supplementary information. Back diamonds correspond to the GH3 protein

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13429

sequences from the six contigs.

Fig. S4. Unrooted phylogenetic tree of the GH5 family glycoside hydrolases present on contigs VISS18_012 and VISS18_040. Phylogenetic trees were calculated using the maximum-likelihood (ML) approach implemented in MEGA v. 6.0.6 (Kumar *et al.*, 2004). Bootstrap values (100 resamplings) are indicated by numbers on the tree. The full listing of the aligned proteins is available as supplementary information. Black diamonds correspond to the GH5 protein sequences from the two contigs.

Fig. S5. Unrooted phylogenetic tree of the GH13 family glycoside hydrolases present on contig VISS18_034. Phylogenetic trees were calculated following the maximum-likelihood (ML) approach. Phylogenetic trees were calculated using the maximum-likelihood (ML) approach implemented in MEGA v. 6.0.6 (Kumar *et al.*, 2004). Bootstrap values (100 resamplings) are indicated by numbers on the tree. The full listing of the aligned proteins is available as supplementary information. Black diamonds correspond to the GH13 protein sequences from contig VISS18_034.

Fig. S6. Mauve (Darling *et al.*, 2010) alignments of PUL-carrying contigs from station 3 and 18. Locally collinear blocks (LCB) that are similar between contigs are depicted in identical colors and connected by lines. Within each LCB a similarity profile is shown. The height of the profile corresponds to the average level of conservation in the respective region. Genes and their annotations are shown below the LCBs.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13361

Wiley-Blackwell and Society for Applied Microbiology

Table S1. Overview of sampling stations S3 and S18 of the 2006 VISION cruise of the research vessel Maria S. Merian (cruise MSM03/01). Further details are provided in Gómez-Pereira *et al.* (2010) and Gómez-Pereira *et al.* (2012).

Table S2. Carbohydrate-active enzymes in the PUL-containing contigs from stations S3 and S18. Annotations include gene identifiers (locus tags), closest characterized homologues, and EC numbers. The percentage of sequence identity is indicated in parentheses. Annotations in blue indicate genes whose functions have been determined by phylogenetic reconstruction.

Table S3. Predictions of the subcellular locations of the carbohydrate-active enzymes in the PUL-containing contigs from stations 3 and 18.

Table S4. Predicted taxonomic affiliations of the S3 and S18 contigs based on evidences derived from the gene's BLASTp hits to the NCBI nr database and HMMER3 hits to the Pfam v. 25 database.

Table S5. E-value thresholds used for automated CAZyme family detection. Searches were performed against the CAZy database, the dbCAN database and the Pfam database using the indicated E-value thresholds. CAZymes were only annotated when at least two of the three database searches yielded positive results.

Table S6. NCBI GenPept labels and accession numbers of the GH3, GH5, GH13 and GH16 proteins that were used in phylogenetic analyses.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13361

Wiley-Blackwell and Society for Applied Microbiology

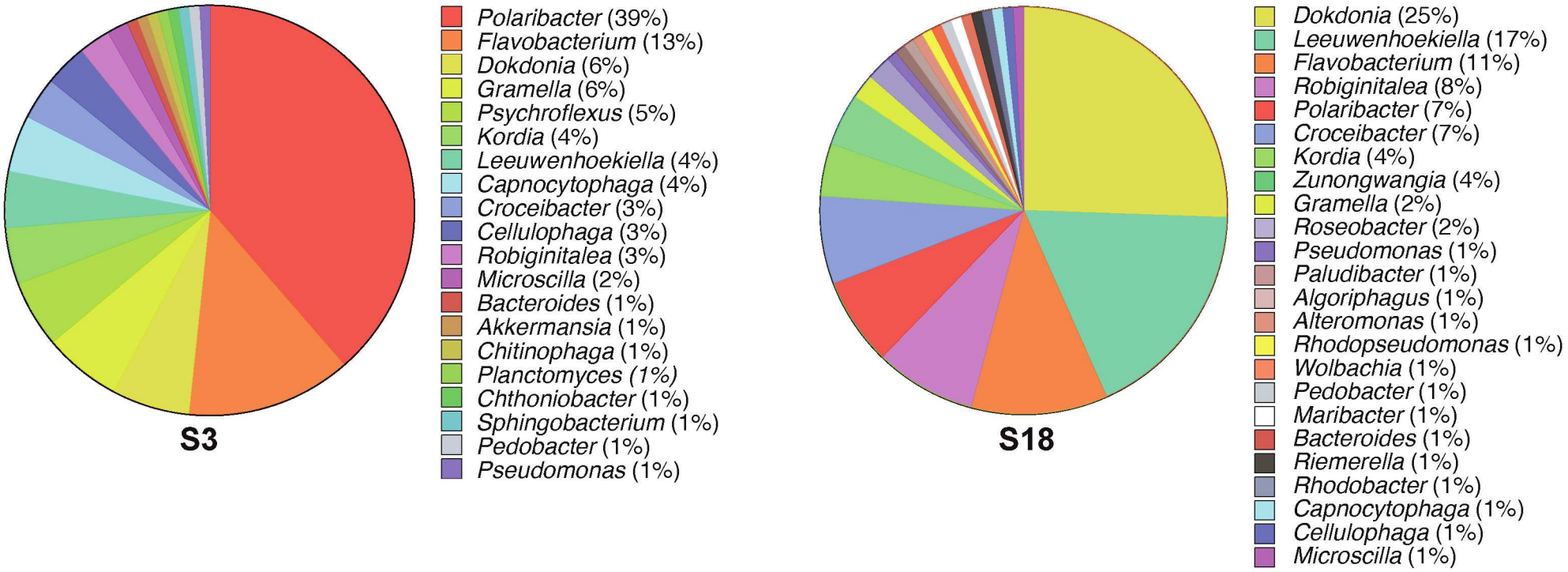

Fig. 1. Genus-level taxonomic affiliation of contigs obtained from re-assembled fosmid sequences of metagenomic libraries from station S3 (121 contigs) and S18 (105 contigs).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13361

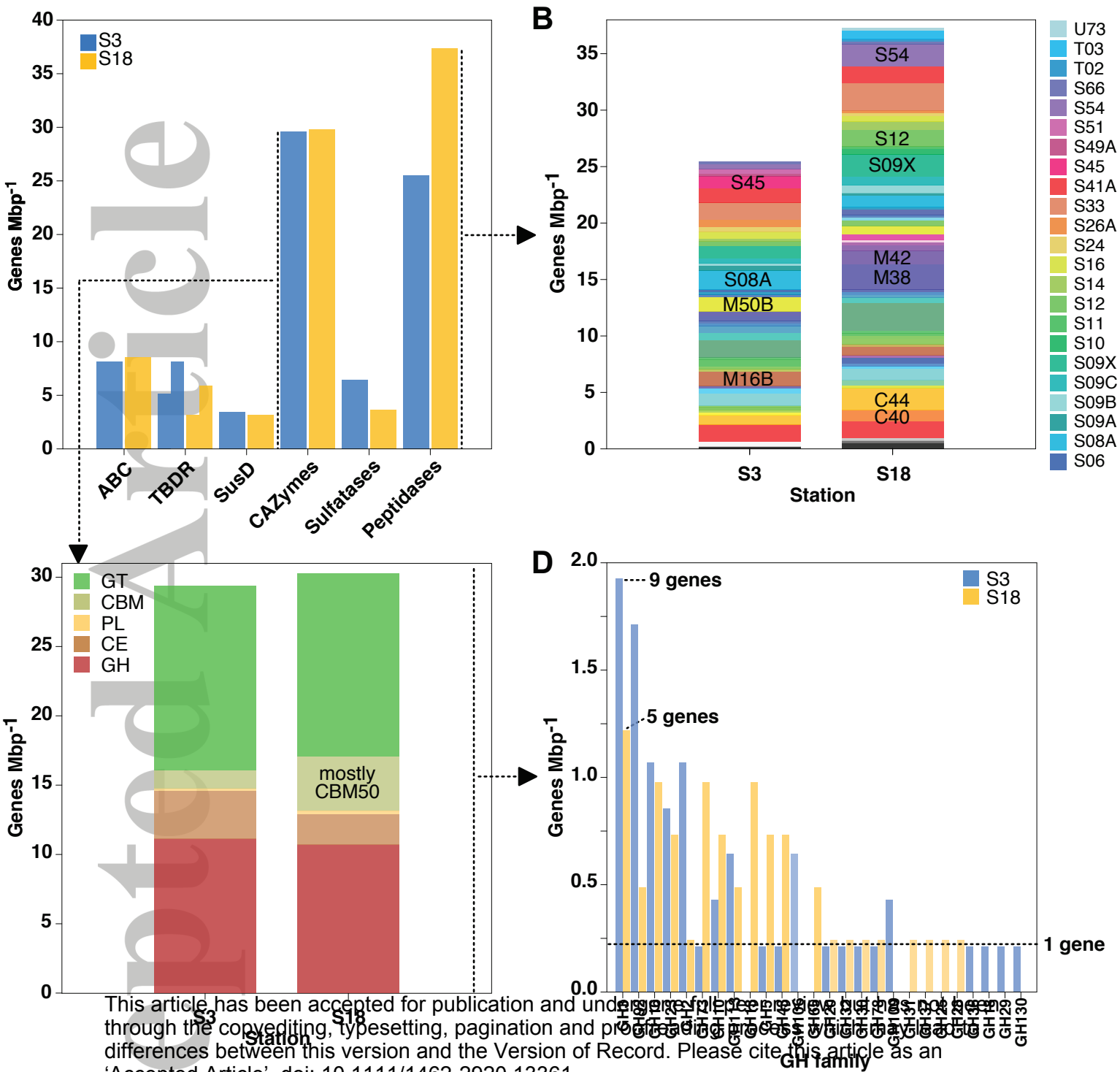

Fig. 2. Comparison of selected genes on the 121 BPLR station 3 contigs (4.7 Mbp) and the 105 NAST station 18 contigs (4.1 Mbp). (**A**) genes for ABC-transporters, TonB-dependent receptors (TBDRs), SusD-like proteins, CAZymes, sulfatases and peptidases (Table 1). Values for TBDRs correspond to two different versions of the TonB_dep_rep Pfam profile (Table 1); (**B**) peptidase families; (**C**) CAZyme classes: glycoside hydrolases (GH), carbohydrate esterases (CE), polysaccharide lyases (PL), carbohydrate-binding modules (CBM) and glycosyltransferases (GT); (**D**) GH families ordered by decreasing averages.

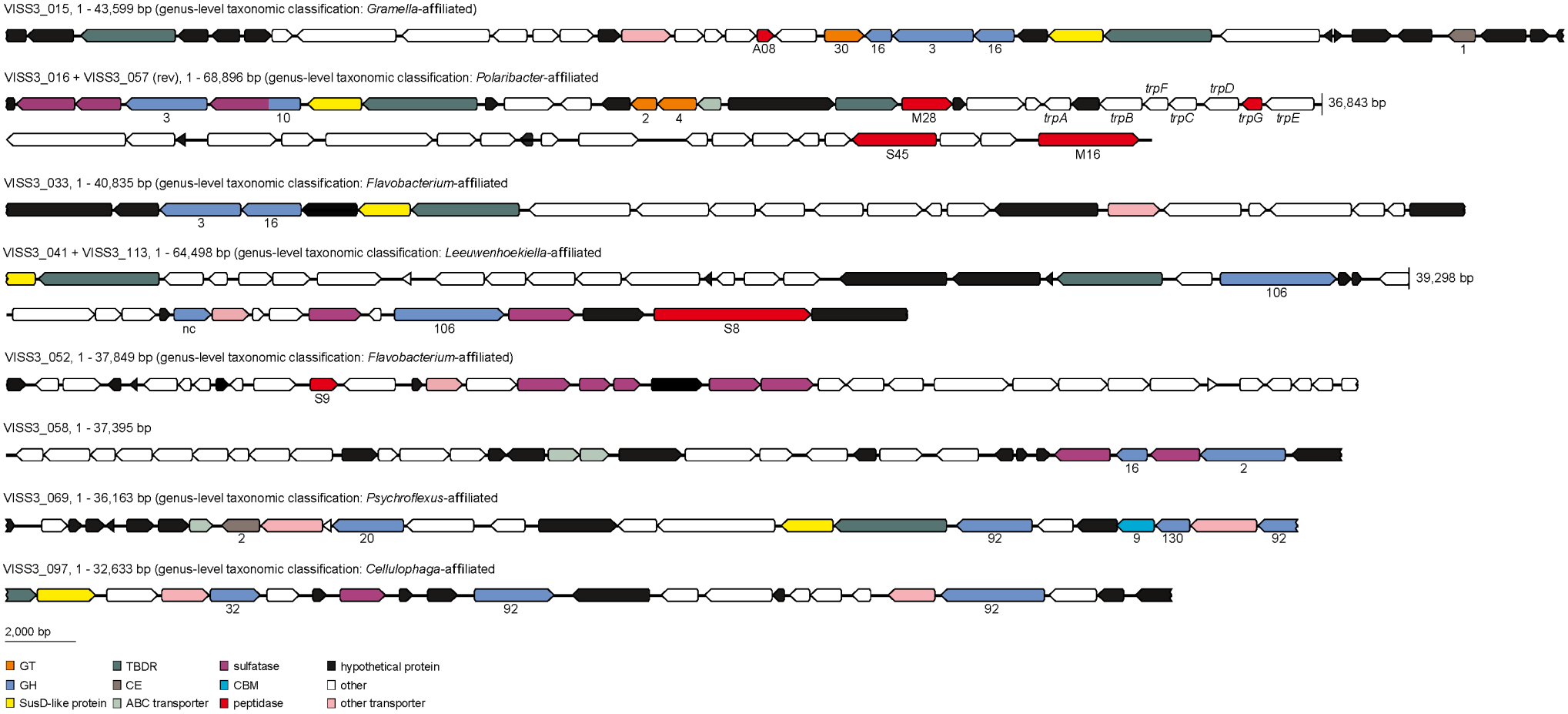

Fig. 3. PUL-containing contigs from BPLR station 3. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

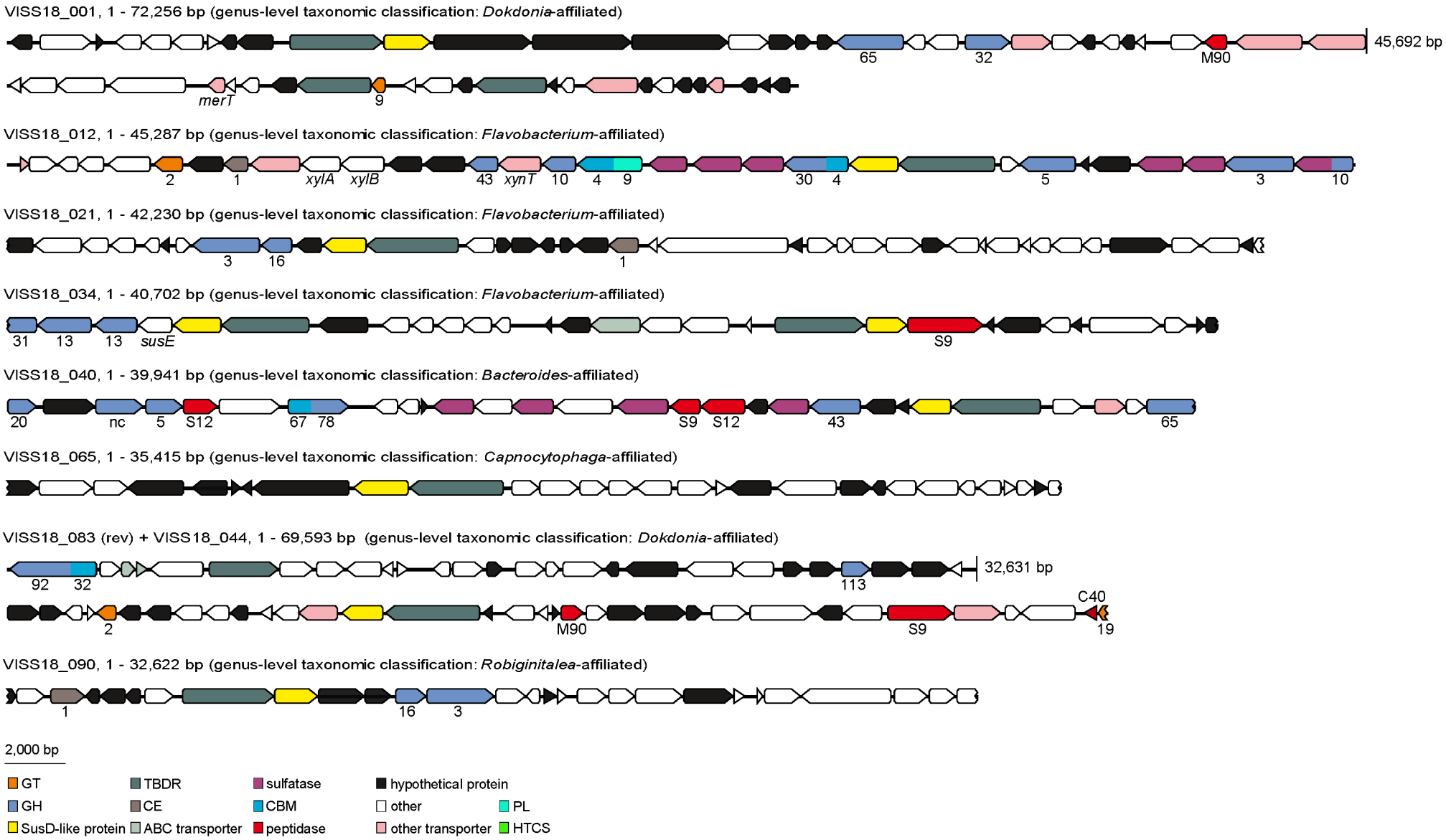

Fig. 4. PUL-containing contigs from NAST station 18. Names, presumed taxonomic affiliations, lengths and the gene contents are provided for each contig.

Table 1. Comparison of frequencies of genes involved in organic matter degradation and uptake between the S3 and S18 contig datasets.

genes / Mbp	S3	S18
ABC transporters genes*	8.1	8.5
TonB-dependent receptor genes*	5.1 (8.1)	3.2 (5.9)
susD genes*	3.4	3.2
CAZymes**	29.5	29.7
GHs	11.1	10.7
CBMs	1.3	3.9
CEs	3.4	2.2
PLs	0.2	0.2
GTs	13.5	12.7
GHs + CEs + CLs	14.7	13.1
sulfatase genes*	6.4	3.7
peptidase genes***	25.5	37.4

* HMMer 3 searches against the Pfam v. 28 database with E ≤ 10^{-5}. Values for ABC transporter and *susD* genes were determined by combining all genes with hits to any of the following profiles: ABC_tran, ABC_membrane, ABC_membrane_2, ABC_membrane_3, ABC_tran_2, ABC2_membrane, ABC2_membrane_2, ABC2_membrane_3, ABC2_membrane_4, ABC2_membrane_5, ABC2_membrane_6, as well as SusD, SusD-like, SusD-like_2, and SusD-like_3. Values for TonB-dependent receptor genes were determined using the TonB_dep_rec profiles from October 2014 and 2012 (values in brackets). Both TonB_dep_rec profiles predicted lower numbers as was suggested by BLAST-based database similarity searches.

** Combined results of BLASTp searches against the CAZy database, HMMer searches against the Pfam v. 28 database and the dbCAN database with manually adjusted E-value thresholds (Teeling *et al.*, 2016).

*** Batch BLASTp searches against the MEROPS 9.13 database with E≤ 10^{-4} (the database is small, hence E≤ 10^{-4} is considered significant).

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as an 'Accepted Article', doi: 10.1111/1462-2920.13361