Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria
Elizabeth Ficko-Blean, Aurelie Prechoux, François Thomas, Tatiana Rochat, Robert Larocque, Yongtao Zhu, Mark Stam, Sabine Génicot, Murielle Jam, Alexandra Calteau, et al.

To cite this version:
Elizabeth Ficko-Blean, Aurelie Prechoux, François Thomas, Tatiana Rochat, Robert Larocque, et al.. Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria. Nature Communications, 2017, 8, pp.1685. 10.1038/s41467-017-01832-6. hal-01899178

HAL Id: hal-01899178
https://hal.sorbonne-universite.fr/hal-01899178
Submitted on 19 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Distributed under a Creative Commons Attribution 4.0 International License
Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria

Elizabeth Ficko-Blean, Aurélie Préchoux, François Thomas, Tatiana Rochat, Robert Larocque, Yongtao Zhu, Mark Stam, Sabine Génicot, Murielle Jam, Alexandra Calteau, Benjamin Viart, David Ropartz, David Pérez-Pascual, Gaëlle Correc, Maria Matard-Mann, Keith A. Stubbs, Hélène Rogniaux, Alexandra Jeudy, Tristan Barbeyron, Claudine Médigue, Mirjam Czjzek, David Vallenet, Mark J. McBride, Eric Duchaud & Gurvan Michel

Macroalgae contribute substantially to primary production in coastal ecosystems. Their biomass, mainly consisting of polysaccharides, is cycled into the environment by marine heterotrophic bacteria using largely uncharacterized mechanisms. Here we describe the complete catabolic pathway for carrageenans, major cell wall polysaccharides of red macroalgae, in the marine heterotrophic bacterium Zobellia galactanivorans. Carrageenan catabolism relies on a multifaceted carrageenan-induced regulon, including a non-canonical polysaccharide utilization locus (PUL) and genes distal to the PUL, including a susCD-like pair. The carrageenan utilization system is well conserved in marine Bacteroidetes but modified in other phyla of marine heterotrophic bacteria. The core system is completed by additional functions that might be assumed by non-orthologous genes in different species. This complex genetic structure may be the result of multiple evolutionary events including gene duplications and horizontal gene transfers. These results allow for an extension on the definition of bacterial PUL-mediated polysaccharide digestion.
Carrageenans, alongside agars, are the main cell wall polysaccharides of red macroalgae and play vital roles in the development and physiology of these photosynthetic eukaryotes. These complex polymers consist of D-galactose based units alternatively linked by β-1,4 and α-1,3 linkages. The β-linked unit is either a D-galactose-6-sulfate or a 3,6-anhydro-D-galactose, a bicyclic sugar unique to red macroalgae. Carrageenans are further modified by the presence of various substituents (sulfate, methyl and pyruvate groups). For more than 70 years, carrageenans have been widely used as ingredients in food, personal care, and cosmetic industries due to their gelling and emulsifying properties. In addition, these polysaccharides and their derived oligosaccharides have multiple biological properties and are promising molecules for blue biotechnology. In marine ecosystems carrageenans constitute a huge biomass and thus a precious carbon source for marine heterotrophic bacteria (MHB). However, carrageenan catabolism is largely uncharacterized and only a few enzymes specific for these sulfated galactans are known.

Bacteroidetes are considered key recyclers of marine polysaccharides and notably of carrageenans. Bacteria from this phylum are particularly suitable models for efficiently characterizing complete carbohydrate degradation pathways. Indeed, they have developed multi-component protein systems tailored for sensing, binding, transporting, and degrading specific glycans, as described for the starch-utilization system (Sus)1,12. The genes encoding these proteins are adjacent and co-regulated; these regions are referred to as Polysaccharide Utilization Loci (PULs) (singular) or PULs (plural). Tandem susD-like and susC-like genes, which encode a carbohydrate-binding lipoprotein and a TonB-dependent transporter (TBDT) respectively, are considered as a hallmark of PULs11,12, and the presence of these susCD-like gene pairs is used to identify PULs in Bacteroidetes genomes. Several PULs have been extensively characterized in human intestinal Bacteroidetes,15–18, describing the complex molecular mechanisms behind the glycoside hydrolase (GH) enzymes and carbohydrate-binding proteins involved in the digestion of common dietary polysaccharides. Some previous studies have described the partial characterization of PULs targeting marine polysaccharides, for instance, alginate-specific PULs were shown to be genuine operons and to encode the enzymes responsible for the bioconversion of algal polysaccharides.10 Notably, this bacterium can utilize kappa-carrageenan (KC), iota-carrageenan (IC), and lambda-carrageenan as sole carbon sources and its kappa-carrageenase (CgkA, CAZY family GH16, http://www.cazy.org/23) and iota-carrageenases (CgiA1, CgiA2, and CgiA3, CAZY family GH82) have been studied24-26.

Here we describe in Z. galactanivorans the discovery and the integrative characterization of the complete pathway for utilization of the kappa family carrageenans (containing 3,6-anhydro-D-galactose units, mainly kappa-, iota- and beta-carrageenans25), which is carried out by a complex regulon including a dedicated atypical PUL, lacking the susCD-like pair, and carrageenan-induced loner genes including distal susCD-like pairs. The core of this complex system is conserved in MHB from different phyla. These carrageenan utilization systems appear to display a remarkable plasticity, likely resulting from diverse evolutionary events such as horizontal gene transfers and gene duplications.

Results

The ZGAL_3145-3159 cluster encodes carrageenan-specific enzymes. The annotation of the genome sequence of Z. galactanivorans10 revealed a cluster of genes encoding three GH127 enzymes (ZGAL_3147, ZGAL_3148, and ZGAL_3150), one enzyme assigned to the GH129 family (ZGAL_3152), and three sulfatases belonging to different GH subfamilies (ZGAL_3145, ZGAL_3146, and ZGAL_3151) belonging to the subfamilies S1_19, S1_7, and S1_17, respectively, http://abims.sb-roscoff.fr/sulfatlas/26, suggesting this gene cluster could be a PUL specific for a sulfated algal polysaccharide (Fig. 1). This 15-gene locus notably encodes 4 additional carbohydrate-related enzymes (ZGAL_3153-3156), three cytoplasmic membrane transporters and an araC family transcription factor (ZGAL_3159), but does not include a susCD-like gene pair. The GHs have distant homologs in enteric bacteria: BT1003 (BtGH127) from Bacteroides thetaiotaomicron is a 3-C-carboxy-5-deoxy-1,4-lactose (acetic acid)-hydrolase involved in rhamnogalacturonan-II depolymerization16; BtGH127 from Bifidobacterium longum is an exo-β-1,4-arabinofuranosidase that acts on the plant glycoproteins extensins27; NagBb (BbgGH129) from Bifidobacterium bifidum hydrolyzes alpha-linked N-acetyl-D-galactosamine from intestinal mucin28. To the best of our knowledge L-α-acetic acid, β-L-arabinofuranose and α-N-acetyl-D-galactosamine are not known components of red algae29, suggesting that the GHs from Z. galactanivorans may have new substrate specificities (Supplementary Discussion). To test these hypotheses, we first cloned and overproduced the 11 enzyme-coding genes of this cluster in Escherichia coli BL21(DE3). The resulting recombinant proteins were all soluble and were purified by affinity chromatography (Supplementary Fig. 1), allowing examination of their activities.

In vitro assays of the GH127 and GH129-like enzymes showed no activity on poly- or oligo-saccharides of agarose (uncharged substrate), agar, porphyran, KC or IC (sulfated substrates).

![Fig. 1 Z. galactanivorans PUL involved in carrageenan catabolism. Genes are annotated according to their CAZY family (GH127, GH129-like), sulfatase subfamily (S1_19, S1_7, and S1_17) or enzymatic activity. The names of the genes and their functions are given in Table 1. The acronyms for the gene names are as follows: cgs, CarraGeenan Sulfatase; dag, β-3,6-AnhydroGalactosidase; dau, β-3,6-Anhydrogalactose Utilization; cgr, CarraGeenan Regulator](https://www.nature.com/naturecommunications)
However, the presence of sulfatases in the gene cluster suggested that these GHs could act after the sulfatases and thus require de-sulfated substrates. Since we had already tested agarose, we searched for a natural source of de-sulfated carrageenans. *Furcellaria lumbricalis* is a red macroalga with beta-carrageenan (BC) in its cell wall, primarily composed of kappa-carrabiose and common referred to as furcellaran. No GH activity was detected on furcellaran using fluorophore-assisted carbohydrate electrophoresis (FACE)\(^3\). In order to produce oligosaccharides, furcellaran was treated with the kappa-carrageenase from *Pseudoalteromonas carrageenovora* which cleaves the \(\beta\)-1,4 bond within kappa-carrabiose motifs in an endolytic manner\(^3\). The product oligosaccharides have a neutral 3,6-anhydro-D-galactose (D-AnG) on the non-reducing end and a D-galactose-4-sulfate on the reducing end. These oligosaccharides were purified by size-exclusion chromatography and the fraction containing a majority of hexasaccharides was tested with the GHs (Fig. 2). All the GHs showed activity on the furcellaran hexasaccharide, as demonstrated using FACE. Nonetheless, there appears to be some as yet undetermined differences in specificity between the three GH127 enzymes as the FACE gel shows different intensities and banding patterns (Fig. 2b). ZGAL_3150 (GH127-3) and ZGAL_3152 (GH129-like) were the most active enzymes and appear indistinguishable biochemically based on FACE patterns (Fig. 2c). MALDI-TOF-MS analysis of these enzymatic digests indicated the release of the terminal neutral monosaccharide D-AnG and of a pentasaccharide (Fig. 2e). Thus, these enzymes are exo-lytic for the \(\alpha\) 1,3-(3,6-anhydro)-D-galactosidases which cleave the \(\alpha\)-1,3 linkage between D-AnG and \(\alpha\)-galactose on the non-reducing end, releasing D-AnG and odd-DP (degree of polymerization) oligocarrageenans (Fig. 3). Overall, this describes a novel enzymatic activity, long predicted to be present in nature, but for the first time described here in two families of non-homologous enzymes. The genes of these new GHs have been named: *dagA1* (ZGAL_3147 encoding GH127-1), *dagA2* (ZGAL_3148 encoding GH127-2), *dagA3* (ZGAL_3150 encoding GH127-3), and *dagB* (ZGAL_3152 encoding GH129-like). Furthermore, this discovery strongly supports that the ZGAL_3145-3159 gene cluster is carrageenan-specific.

**Fig. 2** Biochemical characterization of the GH127 and GH129-like enzymes. GH127 enzymes (DagA1, DagA2, DagA3; ZGAL_3147, ZGAL_3148, ZGAL_3150) and GH129-like enzymes (DagB, ZGAL_3152). a FACE gel depicting the reaction products after furcellaran (beta-carrageenan) oligosaccharides were treated with pure ZGAL_3152. b FACE gel depicting the reaction products after furcellaran oligosaccharides were treated with soluble lysate of *E. coli* BL21 (DE3) that were transformed with the pFO4 vector alone (negative control) or with the genes of interest cloned into the pFO4 vector. c FACE gels depicting the reaction products after furcellaran oligosaccharides were treated with IMAC-purified ZGAL_3152 and ZGAL_3150. d FACE gels depicting the reaction products after furcellaran oligosaccharides were treated with ZGAL_3152 and four conservative active site mutants. e MALDI MS spectra of the DP6 beta-kappa-kappa oligosaccharide obtained in negative ionization mode after incubation with soluble lysates of BL21(DE3) cells that were transformed with the pFO4 vector only and ZGAL_3147, ZGAL_3148, ZGAL_3150. f Spectrum obtained for DP6 beta-kappa-kappa oligosaccharide incubated with no enzyme and the spectrum of the same sample after treatment with IMAC purified ZGAL_3152. Fragments annotated with a * correspond to HEPES adducts on matrix clusters. DP stands for degree of polymerization.
ZGAL_3145 (S1_19), ZGAL_3146 (S1_7), and ZGAL_3151 (S1_17) were first shown to be active sulfatases using the artificial substrate 4-methylumbelliferyl sulfate. A combination of anion-exchange chromatography (HPLC) and 1H-NMR was used to determine the natural substrates and the regioselectivity of these sulfatases (Supplementary Discussion). As predicted, all these sulfatases are active on carrageenans (Supplementary Fig. 2A, B), ZGAL_3146 (gene named cgsB1) is active on kappa-carrabiose motifs, removing the 4-linked sulfate group from D-galactose to generate beta-carrabiose motifs (Supplementary Figs. 3, 4). ZGAL_3151 (cgsA) removes the 4-linked sulfate group from the galactose moiety of iota-carrabiose motifs, generating alpha-carrabiose motifs (Supplementary Fig. 4). To the best of our knowledge, this is the first time this sulfatase activity has been demonstrated.

The original annotations of the four remaining enzymes were not obviously connected to carrageenan: 2-dehydro-3-deoxy-6-phosphogalactonate aldolase (ZGAL_3153), 2-dehydro-3-deoxy-6-phosphogalactonate aldolase (ZGAL_3154), lactaldehyde dehydrogenase (ZGAL_3155), and aldonic acid dehydratase (ZGAL_3156). However, recent discovery of two enzymes from Vibrio sp. EYJ3 that convert 3,6-anhydro-1-galactose (agar component) into 2-dehydro-3-deoxygalactonate provided new insights. Indeed, Vibrio 3,6-anhydro-1-galactose dehydrogenase (VEJY3_09240) and 3,6-anhydro-1-galactonate cycloisomerase (VEJY3_09370) are distantly related to ZGAL_3155 and ZGAL_3156 (35% and 31% sequence identity, respectively). We thus hypothesized that ZGAL_3155 and ZGAL_3156 could catalyze similar reactions but on the D enantiomer of 3,6-anhydrogalactose. As predicted, ZGAL_3155 oxidizes D-AnG into 3,6-anhydro-D-galactonate in the presence of NAD+ and NADP+, with a 5–6-fold preference for NAD+ (Supplementary Fig. 5A, B). ZGAL_3155 is inactive on D-galactose and is therefore a specific 3,6-anhydro-D-galactose dehydratase (gene named dauA). The resulting 3,6-anhydro-D-galactonate is converted to 2-keto-3-deoxy-6-phosphogalactonate by ZGAL_3156, as measured through the thiobarbituric acid (TBA) assay (Supplementary Fig. 5C). ZGAL_3156 did not demonstrate any activity on D-AnG alone (Supplementary Fig. 5D).
Fig. 5D) and is thus a 3,6-anhydro-\(\beta\)-galactonate cycloisomerase (dauB). In the presence of ATP, ZGAL_3154 phosphorylated the 2-keto-3-deoxy-\(\beta\)-galactonate to 2-keto-3-deoxy-6-phospho-\(\beta\)-galactonate. The activity of this 2-keto-3-deoxy-\(\beta\)-galactonate kinase (dauC) was indirectly measured as a function of the oxidation of NADH (Supplementary Fig. 5E). Finally, ZGAL_3153 catalyzed the conversion of 2-keto-3-deoxy-6-phospho-\(\beta\)-galactonate into \(\beta\)-glyceraldehyde-3-phosphate and pyruvate. The activity of this 2-keto-3-deoxy-\(\beta\)-galactonate aldolase (dauD) was measured both in the forward and reverse direction using the TBA assay (Supplementary Fig. 5F, G). In parallel, and in support of our findings, Lee et al.\textsuperscript{33} recently biochemically characterized homologs of these four enzymes in other marine bacteria.

Crystal structure of two 3,6-anhydro-\(\alpha\)-galactose-related enzymes. Most recombinant proteins were put into crystal trials
in order to deepen our understanding of the structure/function relationship of these new enzymes. We were successful in solving the structures of the first α-1,3-(3,6-anhydro)-β-galactosidase (DagB, ZGAL_3152, GH129-like) and of the first 3,6-anhydro-D-galactonate cycloisomerase (DauB, ZGAL_3156) (pdb id 5opq and 5olc, Supplementary Table 1).

ZGAL_3152 forms a homodimer and the monomer has a complex architecture with an N-terminal distorted beta-sandwich domain (Pro35-Asp299), a central TIM barrel (Tyr300-Lys620) and a C-terminal immunoglobulin-like domain (Glu621-Asp693) (Fig. 4a–d, Supplementary Fig. 6). The dimeric interface is mainly formed by the interaction of long loops from the TIM-barrel (342–357 and 430–459) with the two helices protruding from the N-terminal domain (η3 and α2) of the other monomer. The dimer is also stabilized by the swapping of the C-terminal strand β36. There are two predicted active sites at either side of the base.
of an impressive crevasse that is 60 Å long and 40 Å deep (Fig. 4a, b). Numerous trials for obtaining the structure of ZGAL_3152 complexes were performed, but they were unsuccessful; however, buffer molecules found in the active site of each ZGAL_3152 monomer likely mimic monosaccharide units. In chain C, a Tris, an MDP and a second Tris are bound in the potential subsites –1, +1, and +2, respectively (Fig. 4e). The residues interacting with these buffer molecules (Cys198, Lys208, Asn218, His347, Trp455, Leu536, and Gln566) are well conserved in ZGAL_3152 homologs (Supplementary Fig. 7), suggesting their implication in substrate recognition. Four acidic residues are candidate catalytic residues. Three are located in one monomer from the homodimer (Asp486, Glu517, and Glu531), while Asp202 originates from helix α2 of the second monomer. Asp202, Asp486, and Glu517 are strictly conserved. Site-directed mutagenesis of the four candidates yielded soluble inactive enzymes (Fig. 2d); these residues are thus involved in the catalytic machinery, although the structure of a substrate-enzyme complex would be needed to determine their respective roles. Surprisingly, among the two catalytic residues identified in the GH129 alpha-N-acetylgalactosaminidase NagBb, only Asp435 is conserved in ZGAL_3152 homologs (Asp486 in ZGAL_3152, Supplementary Fig. 7). ZGAL_3152 and NagBb display only 16% sequence identity and therefore, considering this extreme sequence divergence and the non-conservation of the catalytic residues, we propose that ZGAL_3152 homologs do not belong to the GH129 family but rather constitute a new GH family.

ZGAL_3156 folds as a (β/α)-β TIM-barrel (amino acids 137–340) with an α/β lid domain (amino acids 136–341 and 341–377). The crystal structure reveals an octamer, not uncommon in the enolase superfamily (Fig. 5a, Supplementary Fig. 8). A size-exclusion column analysis confirmed that ZGAL_3156 constitutes an octamer in solution (Supplementary Fig. 9). In the enolase superfamily, the active site is located at the interface of the two domains, the lid domain shielding the catalytic machinery from bulk solvent. Two disordered regions (17–26 and 138–143) were not modeled. They are close spatially and constitute the tip of the lid domain (Fig. 5b). Equivalent regions are similarly disordered in the low-resolution structure of the β-galacto-1,4-lactone cycloisomerase AtGCI from Agrobacterium tumefaciens (41% identity, PDB: 4ggb, Fig. 5c), speaking to the flexibility of the lid region in these enzymes. The residues involved in the coordination of the catalytic cation (Asp194, Glu220, and Glu246) and the predicted general base (Lys166) and acid (His296) of AtGCI are strictly conserved between AtGCI and ZGAL_3156 and have similar spatial orientations (Fig. 5e, f). This cation was modeled as an Mg$^{2+}$ in ZGAL_3156 and as a Ca$^{2+}$ in AtGCI. The residues shaping the substrate-binding pocket in AtGCI originate from the TIM-barrel and the lid domain of one monomer and from the loop between helices α2 and α3 of the neighboring monomer (Fig. 5h). While Asp87 and Trp298 are conserved in both enzymes, the other residues are substituted and are thus likely involved in 3,6-anhydro-β-galactone recognition (Fig. 5g).

**PUL-encoded genes are essential for the in vivo utilization of carrageenans.** To investigate in vivo gene function in Z. galactanivorans, we recently developed a genetic technique to construct deletion mutants in this bacterium. Four single deletion mutants (ΔdagA3 (Δzgal_3150, Δgh127-3), ΔdagB (Δzgal_3152, Δgh129-liker), ΔdauA (Δzgal_3155, Δ3,6-anhydro-D-galactose dehydrogenase), ΔcgrA (Δzgal_3159, ΔaraC family regulator)) and a double deletion mutant (ΔdagaA3/ΔdagB) were constructed. They all grew comparably to the wild-type strain in Zobell medium and minimum medium supplemented with agar or galactose (Supplementary Fig. 10A, B). The ΔdagB mutant had little effect on growth relative to wild type in minimal media containing either KC or IC; however, the ΔdagA3 mutant showed a significant growth delay and the double ΔdagA3/ΔdagB mutant abolished growth altogether, confirming the importance of these α-1,3-(3,6-anhydro)-α-galactosidases in carrageenan degradation (Fig. 6a, b). This sharply contrasts with the biochemical characterization of ZGAL_3150 (DagA3, GH127-3) and ZGAL_3152 (DagB, GH129-like), where there was no discernible difference in specificity (Fig. 2c). We attribute the difference in deletion effect may be due to either an unknown difference in specificity between the enzymes or different cellular localizations of the enzymes. Subtle differences between the GH127 and GH129-like enzymes would not be surprising considering the complex, hybrid structure of carrageenans whose regular structures are masked by multiple substituents (e.g., sulfate, methyl, pyruvate). For
instance, such variations in substrate specificities are known in GH16 \(\beta\)-agarases and \(\beta\)-porphyrinases which act on other red algal sulfated galactans. Bioinformatic analyzes predict that ZGAL_3150 and ZGAL_3152 are anchored into the outer membrane and secreted in the periplasm, respectively (Fig. 3). The more likely scenario is that ZGAL_3150 is oriented toward the periplasm. Thus, both enzymes would be localized within the periplasm. Therefore, both enzymes would be localized within the periplasm.

Table 1 Selection of carrageenan-induced genes in Z. galactanivorans

| Locus_tag | Description (gene name/acronym, family) | KC/D-gal log2-FC* | IC/D-gal log2-FC | D-AnG/D-gal log2-FC | padj | padj |
|-----------|--------------------------------------|-----------------|-----------------|-------------------|-----|-----|
| ZGAL_181  | Sulfatase (cgtB2, S1_7)               | 8.4             | 8.0             | 5.1               | 0.76E-34 |
| ZGAL_236  | Kappa-carrageenase (cgpA, GH16)      | 10.6            | 8.7             | 0.8               | 1   |
| ZGAL_1973 | Iota-carrageenase (cgpA3, GH82)      | 7.6             | 8.7             | 0.1               | 1   |
| ZGAL_3150 | Sulfatase (cgsA, S1_19)              | 6.3             | 8.7             | 3.7               | 3.4E-20 |
| ZGAL_3146 | Sulfatase (cgsB1, S1_17)             | 6.8             | 6.0             | 3.9               | 1.20E-33 |
| ZGAL_3147 | \(\alpha\)-3,6-anhydro-\(\alpha\)-galactosidase (dagA1, GH127) | 7.2 | 7.4E-135 | 3.9 | 1.20E-33 |
| ZGAL_3148 | \(\alpha\)-3,6-anhydro-\(\alpha\)-galactosidase (dagA2, GH127) | 6.6 | 2.7E-61 | 3.5 | 6.80E-16 |
| ZGAL_3149 | Sugar permease (MF5)                  | 6.7             | 6.2             | 4.7               | 9.94E-31 |
| ZGAL_3150 | \(\alpha\)-3,6-anhydro-\(\alpha\)-galactosidase (dagA3, GH127) | 8.1 | 2.9E-108 | 7.1 | 3.78E-32 |
| ZGAL_3151 | Sulfatase (cgsC, S1_17)              | 4.8             | 4.1             | 0.7               | 1   |
| ZGAL_3152 | \(\alpha\)-3,6-anhydro-\(\alpha\)-galactosidase (dagB, new GH) | 3.9 | 6.5E-23 | 0.6 | 1   |
| ZGAL_3153 | 2-keto-3-deoxy-D-galactonate aldolase (dagC) | 4.2 | 9.9E-44 | 3.8 | 2.07E-35 |
| ZGAL_3154 | 2-keto-3-deoxy-D-galactonate kinase (dagC) | 3.6 | 2.4E-30 | 3.3 | 1.84E-25 |
| ZGAL_3155 | 3,6-anhydro-D-galactose dehydrogenase (dagA) | 8.1 | 1.3E-80 | 5.6 | 1.40E-37 |
| ZGAL_3156 | 3,6-anhydro-D-galactonate cycloisomerase (dagB) | 9.7 | 2.0E-89 | 5.2 | 1.07E-23 |
| ZGAL_3157 | Sugar/H+ symporter (DMT)             | 8.4             | 8.0             | 7.0               | 3.85E-95 |
| ZGAL_3158 | High-affinity sugar transporter      | 4.3             | 3.0             | 4.1               | 8.6E-34 |
| ZGAL_3159 | Transcriptional regulator (cgrA, AraC family) | 2.3 | 1.8E-14 | 12 | 0.0504 |
| ZGAL_3580 | SusD-like lipoprotein (cgtB)         | 9.1             | 8.3             | 4.2               | 1.74E-24 |
| ZGAL_3581 | SusC-like TonB-dependent receptor (cgtA) | 9.6 | 1.93E-103 | 4.3 | 4.99E-19 |
| ZGAL_3629 | Sulfatase (S1_30)                    | 8.5             | 4.9             | 0.1               | 1   |
| ZGAL_3630 | Sulfatase (S1_28)                    | 7.1             | 4.3             | 0.8               | 1   |
| ZGAL_3631 | Polygalacturonase, (GH28)            | 7.8             | 3.9             | 0.4               | 1   |
| ZGAL_3632 | Polygalacturonase, (GH28)            | 7.8             | 4.3             | 0.3               | 1   |
| ZGAL_3633 | Beta-galactosidase (GH2)             | 9.4             | 6.9             | 1.0               | 1   |
| ZGAL_3634 | Alpha-1,4-fucosidase (GH29)         | 8.3             | 4.8             | 1.8               | 1   |
| ZGAL_3637 | SusC-like TonB-dependent transporter (TDRT) | 9.2 | 1.8E-51 | 0.7 | 1   |
| ZGAL_3638 | SusD-like lipoprotein (SGBP)         | 9.9             | 5.3             | 0.1               | 1   |
| ZGAL_4265 | Iota-carrageenase (cgpA1, GH82)      | 8.7             | 4.7             | 0.0               | 1   |

\*log2-FC: log2-fold change of KC, IC or D-AnG relative to D-gal as unique carbon source. Boldface type indicates significant differences (FWER 5%). Full data set is available in Supplementary Data 1–9.

Complementation experiments restored growth of \(\Delta cgrA\) on KC (Supplementary Fig. 10D). In both sets of complementation experiments the growth was improved relative to the wild-type strain (Supplementary Fig. 10C, D). This is likely because the complemented genes are under the Flavobacterium johnsoniae OmpA promoter and therefore under less stringent transcriptional control than the carrageenan-specific PUL. Phenotyping of \(\Delta cgrA\) on KC or IC solid media indicates a complex mode of regulation (Supplementary Fig. 10E–L, Supplementary Discussion). Finally, the \(cgrA\) deletion had no effect on growth with agar, indicating that this transcription factor is specific for carrageenan catabolism (Supplementary Fig. 10A).

RNA-seq analysis unravels a complex carrageenan-related regulon. RNA-seq expression profiling was performed on Z. galactanivorans grown in minimal media containing D-AnG, KC, and IC, relative to the growth on D-galactose (Table 1, Supplementary Data 1–9). The entire carrageenan-specific PUL was strongly upregulated in both KC and IC with the exception of the transcriptional regulator \(cgrA\) in IC which was at the border of significance. Growth on D-AnG induced 10 genes from the PUL suggesting this monosaccharide, unique to red algae, acts as an effector in regulation. The genes that were induced by D-AnG were specific for oligosaccharide and monosaccharide utilization; no genes were induced for the degradation of carrageenan
polymers. Unexpectedly, numerous stress-related proteins were also induced (e.g., small heat shock protein, universal stress protein, peptide methionine sulfoxide reductase). Together with the observation that the growth with D-AnG induced cell aggregation, this suggests that free D-AnG is not frequent in the natural environment of Z. galactanivorans and that this sugar is normally only an intracellular degradation intermediate and unlikely to be transported by a specific outer membrane transporter.

Numerous genes outside the ZGAL_3145-3159 cluster were strongly upregulated by carrageenans. Significantly, the most induced gene when grown on KC was the kappa-carrageenase cgkA (zgal_236, gh16)\(^{22}\), this gene was also heavily induced by growth on IC (Table 1, Supplementary Data 1–9). The iota-carrageenase genes cgiA1 (zgal_4265, gh82-1) and cgiA3 (zgal_1973, gh82-3)\(^{22,24}\) were upregulated in both KC and IC. Interestingly, cgiA2 (zgal_2155, gh82-2) was the most expressed iota-carrageenase gene in D-galactose but was downregulated in KC and IC. Thus ZGAL_2155 could act as a constitutive sentinel enzyme involved in the initial degradation step releasing signal oligosaccharides inducing the carrageenolytic system. The wild-type kappa-carrageenase CgkA and iota-carrageenase CgiA1 were previously shown to be extracellular enzymes\(^{22,23}\). Thus, the role of these enzymes is probably to generate oligosaccharides for transport by Z. galactanivorans’ SusCD-like transport system. The family S1-7 sulfatase, ZGAL_181, was also among the genes most induced by carrageenans. The corresponding protein displays 65% identity with ZGAL_3146 (CgsB1, S1_7 sulfatase) which desulfates the alpha-carrabiose into beta-carrabiose motifs, suggesting that ZGAL_181 (gene named cgsB2) catalyzes the same reaction. These genes are probably the result of relatively recent gene duplication. Two major players missing in the carrageenan-specific PUL are the archetypal SusC-like and susD-like genes. The most induced gene when grown on IC was the SusC-like TBDT zgal_3581; this gene was also highly induced by growth on KC and to a lesser degree by growth on D-AnG. The adjacent gene encodes a SusD-like lipoprotein (ZGAL_3580) which was also substantially upregulated in all three conditions. This susCD-like gene pair, distal to the PUL, is a good candidate for the outer membrane transport system associated with the ZGAL_3145-3159 cluster. This hypothesis is supported by genomic comparative analyzes, mutant phenotyping and biochemical experiments (see below). Finally, a second PUL is strongly induced in both KC and IC but not D-AnG (ZGAL_3629-3638). This gene cluster encodes four glycoside hydrolases (1 GH2, 1 GH29, and 2 GH28), two sulfatases (S1_28 and S1_30 subfamilies) and a SusCD-like pair.

Characterization of key genes distant from the carrageenan PUL. Complete carrageenan catabolism requires the hydrolysis of the beta-1,4-linkage at the non-reducing end of odd-DP oligo-carrageenans produced by the action of the 3,6-anhydro-D-galactosidases. After beta-galactosidase hydrolysis the resulting even-DP oligosaccharides become again substrates for the 3,6-anhydro-D-galactosidases, and so on until the complete degradation into free d-galactose and 3,6-anhydro-d-galactose. We hypothesized that at least one of the 8 predicted GH2 beta-galactosidases encoded in the genome of Z. galactanivorans was capable of this activity. ZGAL_3633 stood out as the most probable candidate since its gene expression was induced significantly in both iota- and kappa-carrageenans (Table 1). Thus, we cloned ten GH2 constructs from Z. galactanivorans, including ZGAL_3633. All these recombinant enzymes displayed beta-galactosidase activity on an artificial substrate (pNP-beta-D-galactopyranoside, Supplementary Fig. 11). As demonstrated using HPLC (Fig. 7, Supplementary Discussion), two of the GH2s, ZGAL_3633 and ZGAL_4655, are active on odd-DP oligo-carrageenans (furcellaran hydrolyzed by kappa-carrageenase followed by hydrolysis by the 3,6-anhydro-D-galactosidase ZGAL_3152). Both ZGAL_3633 and ZGAL_4655 are constitutively expressed at low levels in Z. galactanivorans in minimum medium supplemented with d-galactose (Supplementary Fig. 7). Sequential digestion of furcellaran oligosaccharides by DagB and GH2 enzymes. HPLC results for the ZGAL_3152 (DagB, GH129-like) and GH2 (ZGAL_3633, ZGAL_4655) sequential enzyme digests on furcellaran oligosaccharides. Oligo-kappa-carrageenan standards (DP2, DP4, DP6, and DP8) are shown in black dotted line on the chromatogram. First, oligo-furcellaran (blue line) was treated with ZGAL_3152 (purple line). This reaction was then stopped by heating at 90 °C for 10 min. Following the inactivation of ZGAL_3152 the oligosaccharides were treated with either ZGAL_3633 (orange line) or ZGAL_3655 (green line).
Data 1–9); however, ZGAL_4655 is not induced by kappa- and iota-carrageenans, suggesting that ZGAL_3633 is the key carrageen-specific beta-galactosidase. Thus, Z. galactanivorans has all the enzyme activities necessary for the complete degradation of kappa family carrageenans, most of them encoded by the ZGAL_3145-3159 cluster and the others by remote, carrageenan-induced genes (GH16: ZGAL_236; GH18: ZGAL_1973, ZGAL_4265; GH2: ZGAL_3633; S1–7: ZGAL_181). The carrageenan gene cluster lacks the susCD-like gene pair found within canonical PULs; however, the susCD-like gene pairs zgal_3580/zgal_3581 and zgal_3637/zgal_3638 are upregulated when grown on iota- and kappa-carrageenans suggesting these are good outer membrane candidates for oligo-carrageenan transport. We succeeded in producing and purifying soluble ZGAL_3580 and ZGAL_3638 (SusD-like proteins) and probed their interaction with red algal cell wall polysaccharides using affinity gel electrophoresis (Fig. 8a). In the native gel without polysaccharide, ZGAL_3580 migrates as a single band while ZGAL_3638 forms a smear. There are changes in the intensity of the different bands corresponding to ZGAL_3638 and all of the polysaccharides tested, suggesting that the polysaccharides may have an effect on the quaternary structure of the protein, but there is no obvious delay in the migration and ZGAL_3638 did not appear to significantly interact with the ligands tested. The absence of ZGAL_3638 specificity for carrageenans and the presence of two putative GH16 beta-porphyranases, ZGAL_3628 (PorD) and ZGAL_3640 (PorE) encoded in the ZGAL_3628-3640 locus suggest that this locus is more likely dedicated to the degradation of another sulfated galactan distinct from carrageenan (e.g., sulfated agars). In contrast, migration of ZGAL_3580 is retarded (as a clear band) by kappa-carrageenan and furcellaran and only slightly by iota carrageenan. No shift is apparent in agar, porphyran or lambda-carrageenan. Therefore, ZGAL_3580 interacts with kappa family carrageenans. Phenotyping experiments on Z. galactanivorans deletion mutant Δzgal_3580/Δzgal_3581 indicate pronounced inhibition of growth relative to wild type on both KC and IC and moderate inhibition of growth on furcellaran (Fig. 8b). The biochemical (Fig. 8a) and the genetic (Fig. 8b) experiments confirm the affinity of ZGAL_3580 and ZGAL_3581 for family kappa carrageenans, but with a difference in preference depending on the chosen method. However, the genetic approach evaluates the combined properties of ZGAL_3580 and ZGAL_3581, whereas the gel shift assay only characterized ZGAL_3580. This suggests that the TBDT ZGAL_3581 has a strong affinity for iota-carrageenan, which compensates for the reduced affinity of the SusD-like protein ZGAL_3580 for this polysaccharide relative to KC. Overall, these results are consistent with ZGAL_3580/ZGAL_3581 being responsible for carrageenan oligosaccharide import, forming part of the carrageenan regulon and thus acting as the distal susCD-like gene pair for the carrageenan PUL. The genes have been named cgTA for the SusC-like TBDT (zgal_3581) and cgTB for the SusD-like lipoprotein (zgal_3580).

Based on the biochemical, genetic, and transcriptomic evidence presented here, we have demonstrated for the first time a complete scheme for the catabolism of kappa-, iota- and beta-carrageenans in Z. galactanivorans, from the initial action of the GH16 kappa-carrageenase (CgK) and GH82 iota-carrageenases (CgIa1-3) to the conversion in four steps of D-AnG into α-3,6-anhydro-d-galactose dehydrogenase activity was due to the GH127 family. In clades 3, 4, and 5, the structure of the carrageenan-specific PUL is significantly modified. When considering the 30 bacterial species, the core system is restricted to the genes responsible for the release of D-AnG (dagA) and for its catabolism (dauA, dauB, dauD). This may be due to horizontal gene transfer (HGT) events from Bacteroidetes to other phyla. Somewhat surprising is the lack of dauC in some species. Such bacteria may use a non-

Plasticity of carrageenan utilization systems amends the notion of PUL. We searched the Genbank database for bacteria possessing a potential carrageenan-specific PUL, using BlastP with key enzymes as queries (ZGAL_3150 (DagA3, GH127-3), ZGAL_3155 (DauA, 3,6-anhydro-d-galactose dehydrogenase)). After manual verification of each genomic region, we identified 29 species with a homologous carrageenan-specific PUL (including Tenacibaculum jejune) whose genome sequence has been sequenced by Eric Duchaud’s group and deposited at EMBL in the context of this study). These bacteria belong to four phyla: Bacteroidetes, Proteobacteria, Planctomycetes, and Firmicutes. All these microorganisms originate from marine ecosystems: mostly free-living bacteria isolated from seawater, marine sediments or isolated at the surface of macroalgae, but also gut bacteria from animals feeding on macroalgae (surgel, fish, sea urchin, and abalone). The limits of each PUL were manually refined and 12 clusters of orthologous genes from this PUL were subsequently determined (Supplementary Data 10, 11, Supplementary Fig. 12). Homologs of selected carrageenan-induced genes from Z. galactanivorans were also searched in the 29 bacterial genomes (Supplementary Data 10, 11) with the conservation of these genes evaluated by a Heatmap; based on these conservation profiles, the bacterial species clustered into five main clades (Fig. 9). Clades 1 and 2 include only Bacteroidetes (from different classes) and their PUL organizations are the most similar to that of Z. galactanivorans. Strikingly, several carrageenan-induced genes remote from the Z. galactanivorans carrageenan-specific PUL are found within the carrageenan-specific PULs of other Bacteroidetes. This is notably the case of ZGAL_3581 (cgTA) and ZGAL_3580 (cgTB), which is consistent with the function in Z. galactanivorans of this SusCD-like pair in the import of carrageenan degradation-products. Z. galactanivorans possesses two S1_7 sulfatases (65% sequence identity), one located in the carrageenan PUL (ZGAL_3146) and the other distal to the PUL (ZGAL_181) but forming part of the carrageenan regulon (Table 1). In contrast, the PULs of several Bacteroidetes species contain both orthologous genes of the S1_7 sulfatases (Supplementary Fig. 12), consistent with the hypothesis of recent gene duplication. The GH127 genes are likely another example of gene duplication, since their number varies from 1 to 3 paralogous genes depending on the species. Other Bacteroidetes PULs contain GH16 genes distantly related to zgal_236 (cgKA) and most likely forming a new GH16 subfamily. Within clades 1 and 2 we can define a core carrageenan utilization system which includes dauA, dauB, dauC, dauD, dagA, cgsB, cgrA, cgtA, and cgtB. Unexpectedly, the GH16 kappa-carrageenases and the GH82 iota-carrageenases are not part of the core system. Indeed some species harbor only one type of carrageenase, while others are deprived of any known carrageenases, suggesting that these latter bacteria may have new carrageenase families or that they can only degrade predigested oligosaccharides. Whereas the GH127 genes (dagA) are strictly conserved in the PUL, the GH129-like gene (dagB) is only found in a few species, suggesting that the ancestral α-1,3-(3,6-anhydro)-d-galactosidase activity was due to the GH127 family. When considering the 30 bacterial species, the core system is restricted to the genes responsible for the release of D-AnG (dagA) and for its catabolism (dauA, dauB, dauD). This may be due to horizontal gene transfer (HGT) events from Bacteroidetes to other phyla. Somewhat surprising is the lack of dauC in some species. Such bacteria may use a non-
phosphorylative variant of the Entner-Doudoroff pathway to degrade 2-keto-3-deoxy-D-galactonate, as observed in the archaeon *Picrophilus torridus*. The *susD*-like genes are unique to the *Bacteroidetes* and transfer of the PUL to species belonging to other phyla has resulted in the loss of the *susCD*-like pair. This phenomenon is observed in the clade 3 Gammaproteobacteria (Fig. 9) and was previously identified in the case of alginolytic operons. However, this does not mean that TBDT are absent in the gene clusters of other phyla. For instance, all clade 3 Gammaprotobacteria species have a TBDT gene in their cluster (CATDS2_v1220055 in *Catenovulum agarivorans* DS-2; JRKG01_v1_110122 in *Pseudoalteromonas* sp. PLSV; Patl_0887 in *P. atlantica* T6c; H978DRAFT_1909 in *Alteromonas* sp. ALT199). They are only distantly related to ZGAL_3581, but their location within the cluster strongly suggests that they play a similar role in the import of the carrageenan degradation product.

**Fig. 8** Affinity gel electrophoresis for the SusD-like proteins and phenotyping for the *susCD*-like mutant ΔcgtA-cgtB. **a** BSA and recombinant ZGAL_3580 (CgtB, SusD-like) and ZGAL_3638 (SusD-like) migrated at 60 V over 5 h in native-PAGE gels or in gels containing different red algal cell wall polysaccharides. The delay of migration for ZGAL_3580 (red line) and ZGAL_3638 (green line) was compared between the native gels and the polysaccharide containing gels by comparison with BSA (black line). **b** Growth curves for wild-type *Z. galactanivorans* (black curves) and ΔcgtA-cgtB (red curves) in rich Zobell media and in minimal media containing different carbohydrate substrates.
Such clusters containing polysaccharide-related TBDT genes have already been described in Gammaproteobacteria\[^{38}\]. The presence of carrageenan-specific PULs in gut bacteria from marine herbivorous animals is also reminiscent of the horizontal acquisition of porphyran/agar-related genes in animal and human intestinal symbionts\[^{39}\]. This is quite clear for the surgeonfish symbiont \textit{Epulopiscium} sp. (Fig. 9, clade 3), which is also known to have horizontally acquired \textit{GH}16 porphyranase and \textit{GH}17 genes from marine flavobacteria\[^{40}\]. Phylogenetic analysis of the \textit{GH}127 family unraveled another likely HGT case (Supplementary Fig. 13, Supplementary Data 12). Whereas ZG\_AL\_3148 (DagA2, GH127-2) and ZG\_AL\_3150 (DagA3, GH127-3) cluster only with homologs from marine bacteria harboring the carrageenan-specific PUL, ZG\_AL\_3147 (DagA1, GH127-1) is at the root of a clade including \textit{GH}127 enzymes from human gut \textit{Bacteroides} species. These \textit{dagA}1-like genes are within PULs including \textit{GH} genes at first sight unrelated to carrageenans (e.g. \textit{GH}78, \textit{GH}95, \textit{GH}110; Supplementary Fig. 14). Thus, these \textit{Bacteroides} \textit{GH}127 genes have a marine origin, but have most likely evolved in specificity after their horizontal acquisition.

**Conclusion.** Here we have shown that PUL-like structures (lacking the \textit{susCD}-like pair but maintaining other carrageenan-related genes) are found in bacterial phyla other than \textit{Bacteroides}. Furthermore, the carrageenan utilization system is not static and can be characterized by gene losses and gene acquisitions with a dedicated core 3,6-anhydro-D-galactose metabolism that is conserved within carrageenolytic bacteria. This core system is essential but not sufficient for carrageenan utilization. Missing functions (e.g., carrageenases, some specific sulfatases) may be assumed by non-orthologous genes in different bacterial species. Therefore, polysaccharide utilization pathways are not always conferred by a single locus, even in \textit{Bacteroidetes}, and may consist
of a complex regulon. Moreover, our work experimentally strengthens the recent proposition that the PUL definition should not be restricted to the presence of a ssuCD-like pair and that this PUL notion should be extended to other bacterial phyla.

**Methods**

**Materials.** All materials were obtained from Sigma-Aldrich unless stated otherwise. Bacterial strains and plasmids used for phenotyping and complementation studies are listed in Supplementary Table 4.

**Cloning of target genes.** All the *Z. galactanivorans* genes were cloned and overexpressed as previously described. Briefly, genes were PCR-amplified using the NEB Q5 High-Fidelity DNA Polymerase system (Supplementary Table 2). PCR reactions were done with 30 cycles (denaturation: 95 °C; annealing: 60 °C; elongation: 72 °C) using 0.5 units of enzyme in a total reaction of 50 μL with the primers shown in Supplementary Table 2. Amplicons were cleaned up using the QIAquick PCR Purification Kit (Qiagen) and digested with the appropriate restriction endonucleases. All ligationes were done in the linearized T7 system vector pFO4 (a MCS-modified pET15b) except for ZGAL_3151 which was cloned into pET20b.

**Protein production and purification.** In general, *Escherichia coli* BL21(DE3) were transformed with the plasmids containing the gene fragment of interest then grown in the autoinduction Zyp-5052 medium (200 μg mL−1 ampicillin, 20 °C, 72 h). ZGAL_3152 (SeMet) was similarly produced in PASM-5052 medium (1 mg mL−1 ampicillin, until reaching an OD600 of about 0.7. Sulfatase gene expression was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) overnight. All in cases, cells were collected by centrifugation at 3063 g for 30 min. After chemical cell lysis, the lysate was clarified at 13,865 g for 45 min at 4 °C. Using an AKTA FPLC, the supernatant was loaded onto a 5 mL GE His Trap HP column, washed with 20 mM Tris pH 8.0 and 100 mM NaCl and eluted with an increasing gradient of 1–100% 20 mM Tris pH 8.0, 100 mM NaCl, and 1 M imidazole. Fractions containing the protein of interest were pooled and concentrated (MWCO 5 kDa) and then loaded on a Superdex 200 column in 20 mM Tris pH 8.0 and 100 mM NaCl. Fractions were again pooled and concentrated. The ExasyProFLow 300-μL cartridges were used to generate an extinction coefficient for calculation of protein concentration using the A280.

**Marine polysaccharide and oligosaccharide substrates.** Carrageenans poly saccharides were obtained commercially from CP Kelco. The kappa-mu- and kappa-carrageenans were extracted from *Kappaphycus alvarezii*, the iota/nu- and iota-carrageenans from *Euchema denticulatum* and the lambda-carrageenan from tetrasaccharides of *Codium setchbergii*. Fucarcinellans, a beta-carrageenamide from *F. lumbricalis* consisting primarily of beta- and kappa-carrabiose motifs, was obtained as a generous gift from CP Kelco (Brian Rudolph). The exception is alpha- carrageenan which was produced using the native sulfatase from *Pseudoalteromonas atlantica* which is active on IC by using a previously developed protocol. The oligo-lactosylceramide was incubated with purified alpha-carrageenan from *P. atlantica* in 100 mM HEPES pH 7.5, 25 mM NaCl. Oligosaccharides of fucarcinellan, kappa- and iota-carrageenans were produced using the recombinant kappa-carrageenanase from *Pseudoalteromonas carrageenovora* and iota- carrageenase from *Alteromonas fortis*. Carrageenans (0.25%, 5 mL) were treated 48 h at 37 °C with kappa-carrageenan or iota-carrageenan (0.3 mg mL−1, 50 μL) in 100 mM HEPES pH 7.5 and 25 mM NaCl. After checking the total hydrolysis, oligosaccharides mixtures were fractionated by size-exclusion chromatography (SEC). To this end, 5 mL of hydrolysat (concentrated by rotary evaporation at about 5% w/v) were filtered (0.2 μm, Millipore), and injected on a GE Healthcare Superdex 30 prep-grade column (600 × 26 mm i.d.) mounted in series. The elution was conducted at a flow rate of 1 mL min−1 at 20 °C using 50 mM (NH4)2CO3 as the eluent. Oligosaccharides were detected by differential refractometry (Spectra System RI-50, Thermo Separation products) and fractions of 5 mL were collected.

**Sulfatase activity assays.** *HPLC.* Carrageenan-sulfatase activity was measured by high-performance anion-exchange chromatography (HPAEC), according to a protocol adapted from Préchoux et al. Carrageenans solutions (0.5% w/v in H2O mQ) were incubated in presence of purified sulfatases (0.5 mg mL−1) over 20 h at 37 °C in 25 mM Tris pH 8.0, 100 mM NaCl, 0.5 mM CaCl2. For each reaction, a control sample was incubated in similar conditions but with inactivated enzyme (100 °C, 10 min). Reactions were filtered (10 kDa, Amicon Ultra, Millipore) then injected (AS-AP Autosampler) onto an AG11-HC guard column (4 × 50 mm) mounted in series with an AS11-HC anion-exchange column (4 × 250 mm) using an IC5000 system (*Thermo Scientific Dionex*). Elutions for the detection of sulfate were performed with a 12 mM sodium acetate in 12 mM NaOH (Single Pump-5), and the detection of anions was done by a Analytical CD Conductivity Detector associated to a suppressor (ASRS 500, 4 mm) running at 50 mA. Using a standard curve of sulfate and through integration of the peaks, the concentration of sulfate released by the enzymatic reaction was calculated from the difference in the amount of sulfate (retention time at 4 min) between the sample and its associated blanks. For oligosaccharide detection HPAEC analyses were conducted on the same system described for sulfate quantification. Elutions were performed at a flow rate of 0.5 mL min−1 using a NaOH multistep gradient from 8 to 280 mL (40 min). Oligosaccharides were detected by conductivity mode under a current suppression of 30–100%.

**Fluorochrome assisted carbohydrate electrophoresis.** A speed vacuum was used to dry ~50 μg of oligo- polysaccharide reaction volume to be destined for FACE inev. Volume of 2 μL of 0.15 M ANTS (8-amino-3-naphthalene-3,6-disulfonic acid) in a solution of acetic acid and water (3:17) was added to the dried reaction followed by 5 μL of freshly made 1 M sodium cyanoborohydride in DMSO. The samples were incubated overnight at 37 °C in the dark then resuspended in 20 μL of 20% glycerol. Between 2 and 10 μL of sample was loaded onto a 27% poly- acrylamide gel and migrated at 200 V for 2 h in the dark at 4 °C. The gels were visualized under UV light.

**3,6-Anhydro-β-galactose dehydrogenase activity assay.** The enzymatic activity of ZGAL_3155 (DauA, 3,6-Anhydro-β-galactose dehydrogenase) was determined spectrophotometrically as a function of the reduction of NAD+ or NADP+ using a Spark M8 microplate reader (Tecan, France). The reaction mixture was incubated at 25 °C in 25 mM Tris-HCl (pH 7.5) containing 100 mM NaCl and the reaction mixture (250 μL) contained 0.8–4.0 μg of pure recombinant dehydrogenase, 10 mM D-AnG (Dextra) and 1.5 mM NAD+ or NADP+. A negative control was performed by adding 4 μg pure ZGAL_3155 denatured for 10 min at 100 °C, all other conditions were the same within the reaction. An additional control using 10 mM β-galactose instead of D-AnG was also included. The absorbance at 340 nm was followed as a function of time until the absorbance reached a plateau. The activity of the enzyme is expressed as the number of nmoles of NADH produced min−1 assuming the e340 nm = 6220 M−1 cm−1. Each reaction was performed in triplicate. At the end of the reaction, the reaction mixture was denatured for 10 min at 100 °C. The reaction mixture was then centrifuged for 10 min at 29,000g and the supernatant was used as the substrate for the next enzyme, ZGAL_3156.

**3,6-Anhydro-β-galactonate cycloisomerase activity assay.** The enzymatic activity of ZGAL_3156 (DauB, 3,6-Anhydro-β-galactonate cycloisomerase) was determined using the 2-thiobarbituric assay (TBA) method. Volume of 250 μL of the reaction mixture from the preceding enzymatic step (the conversion of D-AnG into 3,6-anhydro-β-galactonate through the action of ZGAL_3155) was incubated with 10–40 μg of pure recombinant ZGAL_3156. The reaction (final volume 300 μL) was performed at room temperature in 25 mM MES pH 6.5 containing 100 mM NaCl (pH 7.5) over 20 min. To measure the activity, aliquots (25 μL) were withdrawn at 0–15 min and the reaction was stopped by the addition of 10 μL of 3% (v/v) acetic acid. The reaction mixture was then centrifuged for 10 min at 29,000g. Volume of 25 μL of the supernatant was then incubated for 20 min at room temperature in the dark with 62.5 μL of 25 mM periodic acid in 250 mM H2SO4 to oxidize 2-keto-3-deoxy-β-galactonate. Oxidation was terminated by the addition of 12.5 μL of 100 mM NADH (0.3% w/v) TBA was then added and the reaction mixture was incubated for 10 min in a boiling water bath. After cooling down to room temperature, a sample of the solution was assayed. Other materials were obtained from Sigma-Aldrich unless stated otherwise. Bacterial strains and plasmids used for phenotyping and complementation studies are listed in Supplementary Table 4.
removed. The color was intensified by the addition of an equal volume of dimepic acid and the absorbance was measured at 549 nm. To produce the substrate for the next reaction, the incubation with ZGAL_3156 was performed for 1 h with 50 µg of enzyme at room temperature. The enzyme was then inactivated for 10 min at 100 °C and centrifuged for 10 min at 29,000 × g and 150 µl of supernatant was used as the substrate for ZGAL_3154.

**2-keto-3-deoxy-D-galactonate kinase activity assay.** The enzymatic activity of ZGAL_3154 (DauD), 2-keto-3-deoxy-D-galactonate kinase) was determined indirectly as a function of the oxidation of NADH using an NADH coupled assay. Reactions were performed in 25 mM MES buffer (pH 6.5). A typical reaction mixture (total volume 193 µl) contained 150 µl of reaction mixture from the preceding step, 2 µg pure recombinant ZGAL_3154, 0.97 mM adenosine-5′-triphosphate (ATP), 9.7 mM MgCl₂ (Acros organics), 0.8 mM Phospho(eno) pyruvic acid tri(cyclohexylammonium) salt, 0.16 mM β-NADH (Applichem), 8 mM KCl and 0.95 µl of mix of pyruvate kinase and lactate dehydrogenase enzymes from rabbit muscle (1:1 Units of pyruvate kinase and 0.8 Units of lactate dehydrogenase, respectively). The reactions were performed at room temperature in 96-wells UV-Star plates (Greiner) and the decrease of absorbance at 340 nm was followed as a function of time until it stabilized. A blank was performed using 2 µg pure ZGAL_3154 denatured for 10 min at 100 °C with all other reaction components the same.

**2-keto-3-deoxy-D-galactonate aldolase activity assay.** ZGAL_3153 (DauD), 2-keto-3-deoxy-D-galactonate aldolase activity was verified using the TBA assay (as described above) to measure cleavage or synthesis of 2-keto-3-deoxy-6-phospho-D-galactonate (D-KDPGA). Degradation was measured using 15 µl of product of ZGAL_3153 incubated 10 min in 5 µg of ZGAL_3153. Secondary synthesis of D-KDPGA from fructose-1,6-bisphosphate and D-erythro-lactate was tested as follows: 30 µl of MES 25 mM pH 6.5, pyruvate 50 mM and D-erythro-lactate 20 mM were incubated with 5 µg of ZGAL_3153 for 10 min. Both reactions were done in triplicate at room temperature and stopped by addition of 10% (v/v) of TCA. 12% negative control was done with the same quantity of enzyme, previously boiled for 10 min.

**Affinity gel electrophoresis.** Interaction of the SusD-like proteins ZGAL_3580 and ZGAL_3638 with polyanionic carbohydrates was tested by affinity gel electrophoresis. Kappa-carrageenan (from *Kappaphycus alvarezi*), iota-carrageenan (Eucheuma denticulum), lambda-carrageenan (Dupont batch 232191437), agar (Sigma), porphyran (water extraction from Porphyra sp.) and furcellaran (CP Kelco) were incorporated at a final concentration of 0.21% in native 10% acrylamide gels. ZGAL_3580 (4 µg), ZGAL_3638 (8 µg) and bovine serum albumin (BSA, 6 µg) were loaded into the wells and migrated at 60 V over 5 h. Gels were stained with Coomassie Blue solution.

**Protein crystallization.** Initial hits were obtained using hanging drop vapor diffus-}

**X-ray crystallography data collection and processing.** SeMet ZGAL_3156 and ZGAL_3638 with poly saccharides was tested by affinity gel electrophoresis. Kappa-carrageenan (from *Kappaphycus alvarezi*), iota-carrageenan (Eucheuma denticulum), lambda-carrageenan (Dupont batch 232191437), agar (Sigma), porphyran (water extraction from Porphyra sp.) and furcellaran (CP Kelco) were incorporated at a final concentration of 0.21% in native 10% acrylamide gels. ZGAL_3580 (4 µg), ZGAL_3638 (8 µg) and bovine serum albumin (BSA, 6 µg) were loaded into the wells and migrated at 60 V over 5 h. Gels were stained with Coomassie Blue solution.

**Protein crystallization.** Initial hits were obtained using hanging drop vapor diffus-
by conjugation to obtain the complemented strain AgcA+ CP (mZG_0043).
Control strains with the empty vector PT7356 inserted into the chromosome were generated in the same way. In all cases, cells of integration of the plasmid at the neomycin site were selected on Cytophaga-agar containing 50 µg mL⁻¹ erythromycin and screened by PCR.

Liquid growth tests. Z. galactanivorans strains were routinely grown from glycerol stocks in Zobell medium 2216E (5 g L⁻¹ tryptone, 1 g L⁻¹ yeast extract, filtered seawater)⁴⁻⁸. Before use, all pre-cultures were collected by centrifugation 10 min at 2840×g, washed in 2 volumes of sterile saline solution, and resuspended in sterile saline solution to obtain the OD600. Fifty microliters of bacterial suspension were inoculated into triplicate 40 mL flasks containing 5 mL of Zobell medium or marine mineral medium (MMP) composed for 1 L of 24.7 g NaCl, 6.3 g MgSO₄·7H₂O, 4.6 g MgCl₂·6H₂O, 2 g NH₄Cl, 0.7 g KCl, 0.6 g CaCl₂, 200 mg NaHCO₃, 100 mg K₂HPO₄, 50 mg yeast extract, and 20 mg FeSO₄·7H₂O and supplemented with 0.5% glucose, 0.5% iota-carrageenan, 0.5% galactan (Danisco, 2544-88-20), kappa-carrageenan (Coffoni ×6913) or D-AnG (Dextra, 114 (2007).)

RNA-seq expression profiling. Bacterial strain and culture conditions. The type strain DsijT of Z. galactanivorans was routinely grown in Zobell medium 2216E (Dxco) at 28°C, 170 r.p.m. For transcriptional profiling, cells were cultured in synthetic Marine Mineral Medium (MMP)⁴¹. MMP was supplemented with different carbon sources: d-galactose (Sigma-Aldrich #C0075), iota-carrageenan (Danisco, 2544-88-20), kappa-carrageenan (Goffoni #6913) or D-AnG (Dextra Laboratories #G0002). Briefly, overnight cultures performed in Zobell medium were diluted at OD₆₀₀ 0.05 in triplicate MMP containing 0.5% glucose as C source and incubated at 28°C until reaching the stationary phase. These cultures were then inoculated to 10 mL of MMP containing 0.4 g L⁻¹ d-galactose, D-AnG, kappa- or iota-carrageenan. When cell density reached an OD₆₀₀ of 0.7 (±0.1), bacteria were collected by centrifugation for 3 min at 4°C after addition of ½ volume of frozen killing buffer (20 mM Tris-HCl pH 7.5, 5 mM MgCl₂, 20 mM NaN₃) to the culture sample. The cell pellets were frozen in liquid nitrogen and stored at -80°C until RNA extraction.

RNA extraction. The cell pellets were resuspended into 800 µL of lysis buffer (4 M guanidine thiocyanate, 25 mM sodium acetate pH 5.2, 5 g L⁻¹ N-laurylsarcosinate), immediately mixed to 1 mL hot acid phenol (Sigma #P4682) and incubated at 55°C for 5 min. The aqueous phases were recovered after addition of 1 mL chloroform and centrifugation at 16,000×g during 10 min at room temperature. The samples were extracted at least four times with an equal volume of acid phenolchloroform/I:IA (25:24:1, pH 4.5)) and once with chloroform. Total RNAs were ethanol precipitated at room temperature and the pellets resuspended in RNase-free water. The RNA concentration of samples was measured using a NanoDrop 1000 Spectrophotometer (NanoDrop Technologies, Inc.). The quality of RNA preparations was assessed by capillary electrophoresis using RNA Nano chips with a Bioanalyzer Agilent 2100 (Agilent Technologies, Inc.). The RNA-seq transcriptionic data have been deposited in the GEO database (GEO accession number: GSE101142). The sequence of the Tenacibaculum jejuense strain DsijT was selected for the comparative genomic analysis has been deposited at EMBL (accession: GCA_900198195). All other relevant data are available in this article and its Supplementary Information files, or from the corresponding author upon request.

Received: 16 March 2017 Accepted: 17 October 2017
Published online: 22 November 2017

References
1. Anderson, N. S., Dolan, T. C. & Rees, D. A. Evidence for a common structural pattern in the polysaccharide sulphates of the Rhodophyceae. Nature 205, 1066–1068 (1965).
2. Martin, M., Portetelle, D., Michel, G. & Vandembol, M. Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. Appl. Microbiol. Biotechnol. 98, 2917–2935 (2014).
3. Michel, G. et al. The kappa-carrageenanase 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).
4. Michel, G. et al. The iota-carrageenase of Alteromonas fortis. A beta-helix fold containing enzyme for the degradation of a highly polyanionic polysaccharide. J. Biol. Chem. 276, 40202–40209 (2001).
5. Guibet, M. et al. Degradation of lambda-carrageenan by Pseudoalteromonas carrageenovora lambda-carrageenanase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases. Biochem. J. 404, 101–114 (2007).
6. Prechoux, A., Genicot, S., Rogniaux, H. & Helbert, W. Controlling carrageenans production using a novel formylglycine-dependent sulfatase, an endo-4s iota-carrageenan sulfatase. Mar. Biotechnol. 15, 265–274 (2013).
7. Prechoux, A., Genicot, S., Rogniaux, H. & Helbert, W. Enzyme-assisted preparation of furcellaran-like kappa-/beta-carrageenan. Mar. Biotechnol. 18, 133–143 (2016).
8. Thomas, F., Hehemann, J. H., Rebuffet, E., Cazier, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. Front. Microbiol. 2, 93 (2011).
9. Teeling, H. et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. Science 336, 608–611 (2012).
10. Barberan, T. et al. Habitat-specific selection as a driving force of carbohydrate catabolism in marine heterrotrophic bacteria: example of the model algae-associated bacterium Zobellia galactanivorans DsijT. Environ. Microbiol. 18, 4610–4627 (2016).

RNA-seq data are presented in Table 1, Supplementary Data 2.

PUL identification and comparison. Similar PULs were detected in 29 organisms using synteny results of the MicroScope platform (Supplementary Data 10–11, Supplementary Fig. 12). The proteins from identified PULs and from Z. galactanivorans regulon were then manually grouped in homolog clusters based on sequence similarity. From these 40 clusters, presence/absence of homolog proteins encoded outside the PULs was determined by blastP alignments with at least 35% of sequence identity and 80% of query coverage. Results of genes gathered in a matrix to indicate, for each organism, if a protein cluster homolog is encoded in the PUL (value 2), elsewhere on the genome (value 1) or absent (value 0). From this matrix, a heat map and a hierarchical classification of the organisms were made using heatmap.2 function and the ward.D2 algorithm with Manhattan distances from gplots and hclust R packages, respectively (Fig. 9).

GH127 phylogeny reconstruction. 549 proteins having a GH127 catalytic module, which aligned on more than 60% of the DBCAN domain, were extracted from the MicroScope platform. Protein sequences were aligned using MAFFT v7.0774, then ambiguous and saturated regions were removed with BMGE v1.12 (with the gap rate parameter set to 0.5)⁴⁶. The best fitting model of amino acid substitution for the GH127 module was selected with ProtTest 4.2.7⁴⁶. A Maximum-Likelihood phylogentic tree was generated with the alignment using PhyML 3.1.0.277 using the LG amino acid substitution model with gamma-distributed rate variation (four categories), estimation of the proportion of invariable sites and exploring tree topologies. 100 bootstrap replicates were performed. The phylogenetic tree was displayed and annotated using the interactive tree of life (ITOL) online tool⁷⁸ (Supplementary Figs. 13, 14, Supplementary Data 12).

Data availability. The coordinates and structure factors for the proteins described above have been deposited in the Protein Data Bank (pdb id: 5opq and 5oel). The RNA-seq transcriptomic data have been deposited in the GEO database (GEO accession number: GSE101142). The sequence of the Tenacibaculum jejuense strain DsijT was selected for the comparative genomic analysis has been deposited at EMBL (accession: GCA_900198195). All other relevant data are available in this article and its Supplementary Information files, or from the corresponding author upon request.
11. Anderson, K. L. & Salyers, A. A. Genetic evidence that outer membrane binding of starch is required for starch utilization by Bacteroides thetaiotaomicron. J. Bacteriol. 171, 3130–3134 (1989).

12. Anderson, K. L. & Salyers, A. A. Biochemical evidence that starch breakdown by Bacteroides thetaiotaomicron involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. J. Bacteriol. 171, 3192–3198 (1989).

13. Bjrussell, M. K., Martens, E. C. & Gordon, J. I. Functional genomic and metagenomic studies of the adaptation of a dominant adult human gut symbiont, Bacteroides thetaiotaomicron, to the suclking period. J. Biol. Chem. 281, 36269–36279 (2006).

14. Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic prediction of polysaccharide utilization loci in Bacteroides species. Bioinformatics 31, 647–655 (2015).

15. Sonnenburg, L. D. et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. Cell 141, 1241–1252 (2010).

16. Larsbrink, J. J. & Alcative genetic locus confers xyloglucan metabolism in select human gut Bacteroides. Nature 506, 498–502 (2014).

17. Cuskin, F. L. et al. Human gut Bacteroides can utilize yeast mannann through a selfish mechanism. Nature 517, 165–169 (2015).

18. Ndeh, D. et al. Complex peptic metabolism in gut bacteria reveals novel catalytic functions. Nature 544, 65–70 (2017).

19. Thomas, F. et al. Characterization of the first algoligin 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 (2012).

20. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize seaweed glycan with carbohydrate-active enzyme updates from extrinsic microbes. Proc. Natl Acad. Sci. USA 109, 19786–19791 (2012).

21. Lombard, V., Sakon, Y. Ramulu, H., Druila, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzyme database (CAZY) in 2013. Nucleic Acids Res. 42, D490–D495 (2014).

22. Barbeyron, T., Gerard, A., Potin, P., Henrissat, B. & Kloareg, B. The kappa-carrageenanase of the marine bacterium Cytophaga drosbichiensis. Structural and phylogenetic relationships within family-16 glycoside hydrolases. Mol. Biol. Evol. 15, 528–537 (1998).

23. Barbeyron, T., Michel, G., Potin, P., Henrissat, B. & Kloareg, B. Iota- carrageenases constitute a novel family of glycoside hydrolases, unrelated to that of kappa-carrageenases. J. Biol. Chem. 275, 35499–35505 (2000).

24. Rebuffet, E. et al. Identification of catalytic residues and mechanistic analysis of the kappa-carrageeansase from Pseudomonas carrageenovorans. Acta Crystallogr. D 55, 918–920 (1999).

25. Michel, G. et al. Expression, purification, crystallization and preliminary X-ray analysis of the iota-carrageeansase from Alteromonas fortis. Acta Crystallogr. D 55, 1690–1695 (1999).

26. Leslie A. G. W. Recent changes to the MOSFLM package for processing and image plate data. Int. CCP4/ESF–EACRM Newsletter. Protein Crystallogr. 26, (1992).

27. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. Acta Crystallogr. D 67, 282–292 (2011).

28. Evans, P. Scaling and assessment of data quality. Acta Crystallogr. D 62, 72–82 (2006).

29. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. Acta Crystallogr. D 50, 760–763 (1994).

30. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. Acta Crystallogr. D 58, 1772–1779 (2002).

31. Ness, S. R., de Graaff, R. A., Abravaya, J. P. & Pannu, N. S. CRANK: new methods for automated macromolecular crystal structure solution. Structure 12, 1753–1761 (2004).

32. Kabsch, W. Xds. J. Appl. Crystallogr. 501 (2010).

33. McCoy, A. J. et al. Phaser crystallographic software. J. Appl. Crystallogr. 46, 2195 (2004).

34. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical and analysis tools on the ExPASy Server. In The Proteomics Protocols Handbook (ed. Walker J. M.) (Humana Press, Totowa, New Jersey, USA, 2005).

35. Michel, G. et al. Expression, purification, crystallization and preliminary X-ray analysis of the iota-carrageeansase from Alteromonas fortis. Acta Crystallogr. D 55, 918–920 (1999).

36. Michel, G. et al. Expression, purification, crystallization and preliminary X-ray analysis of the iota-carrageeansase from Alteromonas fortis. Acta Crystallogr. D 55, 918–920 (1999).

37. Reher, M., Fuhrer, T., Bott, M. & Schonheit, P. The nonphosphorylative pathway of glycate degradation. J. Bacteriol. 192, 964–974 (2010).

38. Studier, F. W. Protein production by auto-induction in high throughput cloning and expression strategies for the post-genomic analysis of marine microorganisms. Microb. Cell Fact. 8, 45 (2009).

39. Ficko-Blean, E. et al. Biochemical and structural investigation of two paralogous aldolases. Bioinformatics 36 (2016).

40. Rebuffet, E. et al. Discovery and structural characterization of a novel glycogen phosphorylase from Picrophilus torridus. J. Bacteriol. 192, 3198–3208 (2010).

41. Tsoukalias, S. et al. Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. PLoS ONE 2, e224 (2007).

42. Hehemann, J. H. et al. Biochemical and structural characterization of the complex agaroxylyl enzyme system from the marine bacterium Zobelia galactanivorans. J. Biol. Chem. 283, 30571–30584 (2012).

43. Reher, M., Fuhrer, T., Bott, M. & Schonheit, P. The nonphosphorylative pathway of glycate degradation. J. Bacteriol. 192, 964–974 (2010).

44. Thomas, F., Barbeyron, T. & Michel, G. Evaluation of reference genes for real-time quantitative PCR in the marine bacterium Zobelia galactanivorans. J. Microbiol. Methods 84, 386–393 (2010).

45. Schneider, R. & Edwards, R. Quality control and preprocessing of metagenome datasets. Bioinformatics 27, 863–864 (2011).
Acknowledgments

We are grateful to Dr Diane Jouanneau for critical discussions on the development of this project and to Dr Lionel Cladière and Ms. Anaïs Naretto for technical assistance. We thank Jonas Collen for taking and sharing with us an image that we submitted for this project and to Dr Lionel Cladière and Ms. Anaïs Naretto for technical assistance. We are grateful to Dr Diane Jouanneau for critical discussions on the development on

Author Contributions

E.F-B. purified and characterized the ZGAL_3633, ZGAL_4655 (GH2), ZGAL_3147, ZGAL_3148, ZGAL_3150 (DagA1-3, GH127-1-3), and ZGAL_3152 (DagB, GH129-like) enzymes, crystallized and solved the structure of ZGAL_3152, and wrote the paper. A.P. and R.L. cloned and expressed all target-genes. A.P. purified and characterized the sulfa- tofano-Nioche, C. et al. Transcriptomic profiling of the oyster pathogen Vibrio splendidus opens a window on the evolutionary dynamics of the small RNA repertoire in the Vibrio genus. RNA 18, 2201–2219 (2012).

70. Rutherford, K. et al. Artemis: sequence visualization and annotation. Bioinformatics 25, 2078–2079 (2009).

71. Varet, H., Brillet-Gueguen, L., Coppee, J. Y. & Dillies, M. A. SARTools: A high-throughput sequencing data. Bioinformatics 31, 166–1669 (2015).

72. Varet, H., Brillet-Gueguen, L., Coppee, J. Y. & Dillies, M. A. SARTools: A DESeq2- and EdgeR-based R pipeline for comprehensive differential analysis of RNA-Seq data. PLoS ONE 11, e0157032 (2016).

73. Vallenet, D. et al. MicroScope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. Nucleic Acids Res. 45, D517–D528 (2016).

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

75. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol. Biol. 10, 210 (2010).

76. Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165 (2011).

77. Guindon, S. et al. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321 (2010).

78. Letunic, I. & Bork, P. Interactive tree of life (iTOl) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res. 44, W242–W245 (2016).