

ARTICLE

DOI: 10.1038/s41467-017-01832-6

OPEN

# Carrageenan catabolism is encoded by a complex regulon in marine heterotrophic bacteria

Elizabeth Ficko-Blean<sup>1</sup>, Aurélie Préchoux<sup>1</sup>, François Thomas<sup>1</sup>, Tatiana Rochat<sup>2</sup>, Robert Larocque<sup>1</sup>, Yongtao Zhu<sup>3</sup>, Mark Stam<sup>4</sup>, Sabine Génicot<sup>1</sup>, Murielle Jam<sup>1</sup>, Alexandra Calteau<sup>4</sup>, Benjamin Viart<sup>4</sup>, David Ropartz<sup>5</sup>, David Pérez-Pascual<sup>2</sup>, Gaëlle Correc<sup>1</sup>, Maria Matarad-Mann<sup>1</sup>, Keith A. Stubbs<sup>6</sup>, Hélène Rogniaux<sup>5</sup>, Alexandra Jeudy<sup>1</sup>, Tristan Barbeyron<sup>1</sup>, Claudine Médigue<sup>4</sup>, Mirjam Czjzek<sup>1</sup>, David Vallenet<sup>4</sup>, Mark J. McBride<sup>3</sup>, Eric Duchaud<sup>2</sup> & Gurvan Michel<sup>1</sup>

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.

<sup>1</sup>Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 8227, Integrative Biology of Marine Models, Station Biologique de Roscoff, CS 90074 Roscoff, Bretagne, France. <sup>2</sup>VIM, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France. <sup>3</sup>Department of Biological Sciences, University of Wisconsin-Milwaukee, 53201 Milwaukee, WI, USA. <sup>4</sup>UMR 8030, CNRS, Université Évry-Val-d'Essonne, CEA, Institut de Génomique - Genoscope, Laboratoire d'Analyses Bioinformatiques pour la Génomique et le Métabolisme, F-91000 Évry, France. <sup>5</sup>INRA, UR1268 Biopolymers Interactions Assemblies, F-44316 Nantes, France. <sup>6</sup>School of Molecular Sciences, The University of Western Australia, Crawley, WA 6009, Australia. Elizabeth Ficko-Blean and Aurélie Préchoux contributed equally to this work. Correspondence and requests for materials should be addressed to G.M. (email: [gurvan.michel@sb-roscoff.fr](mailto:gurvan.michel@sb-roscoff.fr))

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 consists of D-galactose based units alternatively linked by  $\beta$ -1,4 and  $\alpha$ -1,3 linkages. The  $\beta$ -linked unit is either a D-galactose-6-sulfate or a 3,6-anhydro-D-galactose, a bicyclic sugar unique to red macroalgae<sup>1</sup>. Carrageenan structures 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<sup>2</sup>. 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<sup>3-7</sup>.

*Bacteroidetes* are considered key recyclers of marine polysaccharides<sup>8-10</sup> and notably of carrageenans<sup>2</sup>. 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)<sup>11,12</sup>. The genes encoding these proteins are adjacent and co-regulated; these regions are referred to as Polysaccharide Utilization Loci (PUL (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 PULs<sup>13</sup>, and the presence of these *susCD*-like gene pairs is used to identify PULs in *Bacteroidetes* genomes<sup>14</sup>. Several PULs have been extensively characterized in human intestinal *Bacteroidetes*<sup>15-18</sup>, 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 alginate into 2-keto-3-deoxy-6-phosphogluconate<sup>19</sup>. The transcription of a PUL in *Bacteroides plebeius* was shown to be induced by porphyran. The PUL-encoded enzymes responsible for the initial endolytic degradation of this sulfated agar (family GH16 and GH86  $\beta$ -porphyranases) were biochemically and structurally characterized<sup>20</sup>. However, such studies, focusing on macroalgal polysaccharides, have not yet reached the level of integrative characterization observed in the field of PULs from human gut *Bacteroidetes*.

The marine *Bacteroidetes* *Zobellia galactanivorans* Dsij<sup>T</sup>, originally isolated from a healthy red macroalga, is a bacterial model

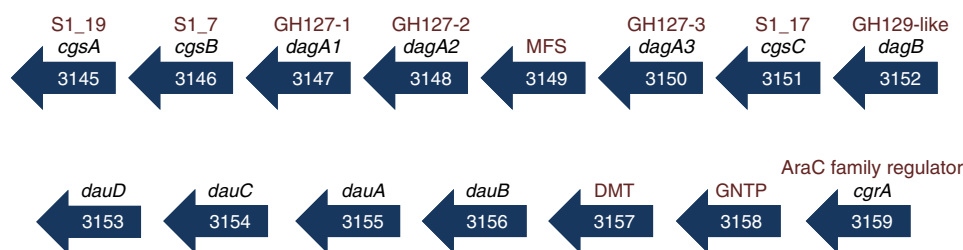
for the bioconversion of algal polysaccharides<sup>10</sup>. Notably, this bacterium can utilize kappa-carrageenan (KC), iota-carrageenan (IC), and lambda-carrageenan as sole carbon sources and its kappa-carrageenase (CgkA, CAZY family GH16, <http://www.cazy.org/><sup>21</sup>) and iota-carrageenases (CgiA1, CgiA2, and CgiA3, CAZY family GH82) have been studied<sup>22-24</sup>.

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-carrageenans<sup>25</sup>), 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.

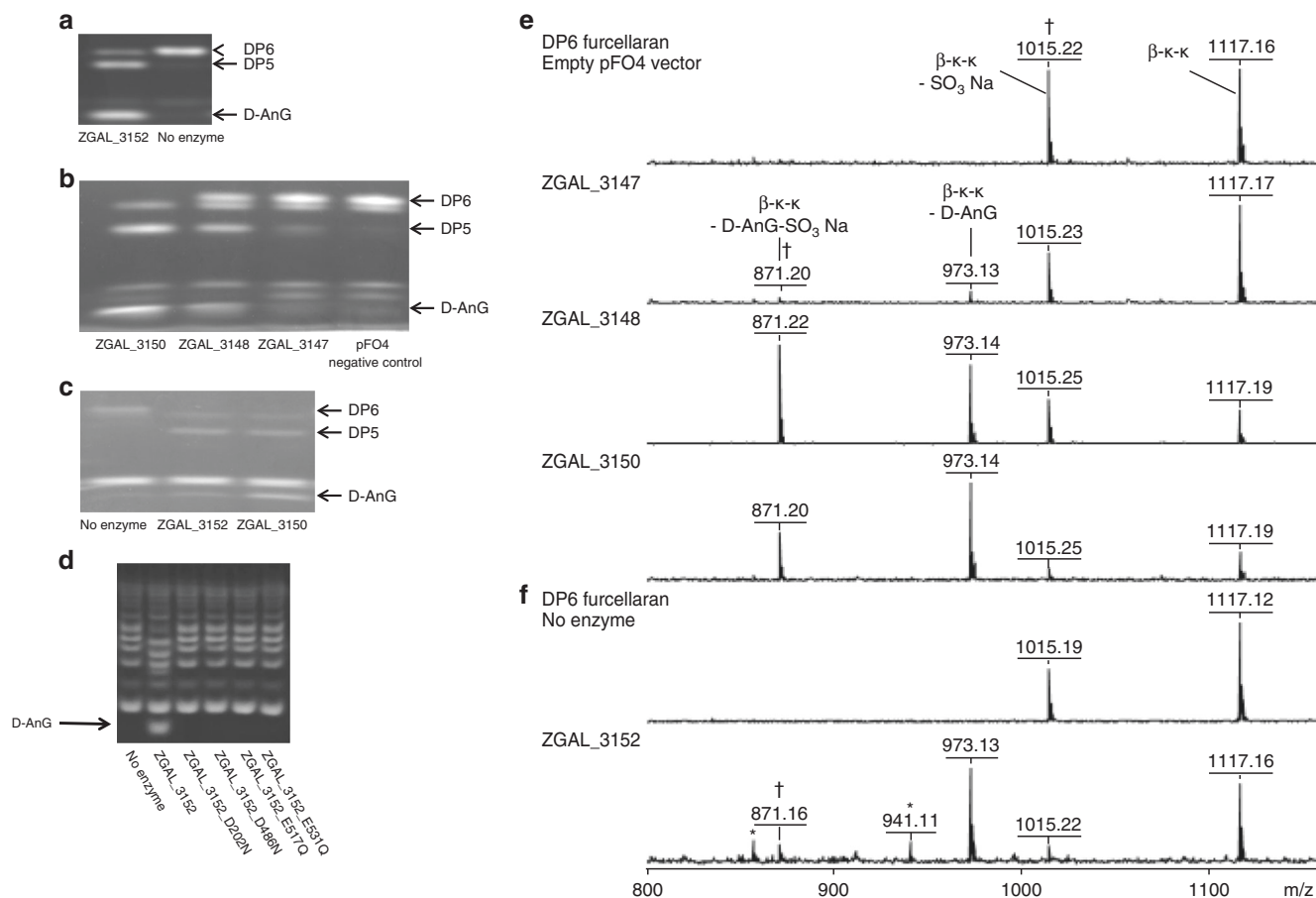
## Results

**The ZGAL\_3145-3159 cluster encodes carrageenan-specific enzymes.** The annotation of the genome sequence of *Z. galactanivorans*<sup>10</sup> revealed a cluster of genes encoding three GH127 enzymes (ZGAL\_3147, ZGAL\_3148, and ZGAL\_3150), one enzyme assigned to the GH129 family<sup>21</sup> (ZGAL\_3152), and three sulfatases belonging to different S1 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/><sup>26</sup>), 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-L-xylose (aceric acid)-hydrolase involved in rhamnogalacturonan-II depolymerization<sup>18</sup>; *BIGH127* from *Bifidobacterium longum* is an exo- $\beta$ -L-arabinofuranosidase that acts on the plant glycoproteins extensins<sup>27</sup>; *NagBb* (*BbGH129*) from *Bifidobacterium bifidum* hydrolyzes alpha-linked N-acetyl-D-galactosamine from intestinal mucin<sup>28</sup>. To the best of our knowledge L-aceric acid,  $\beta$ -L-arabinofuranose and  $\alpha$ -N-acetyl-D-galactosamine are not known components of red algae<sup>29</sup>, 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), agars, porphyrans, KC or IC (sulfated substrates).



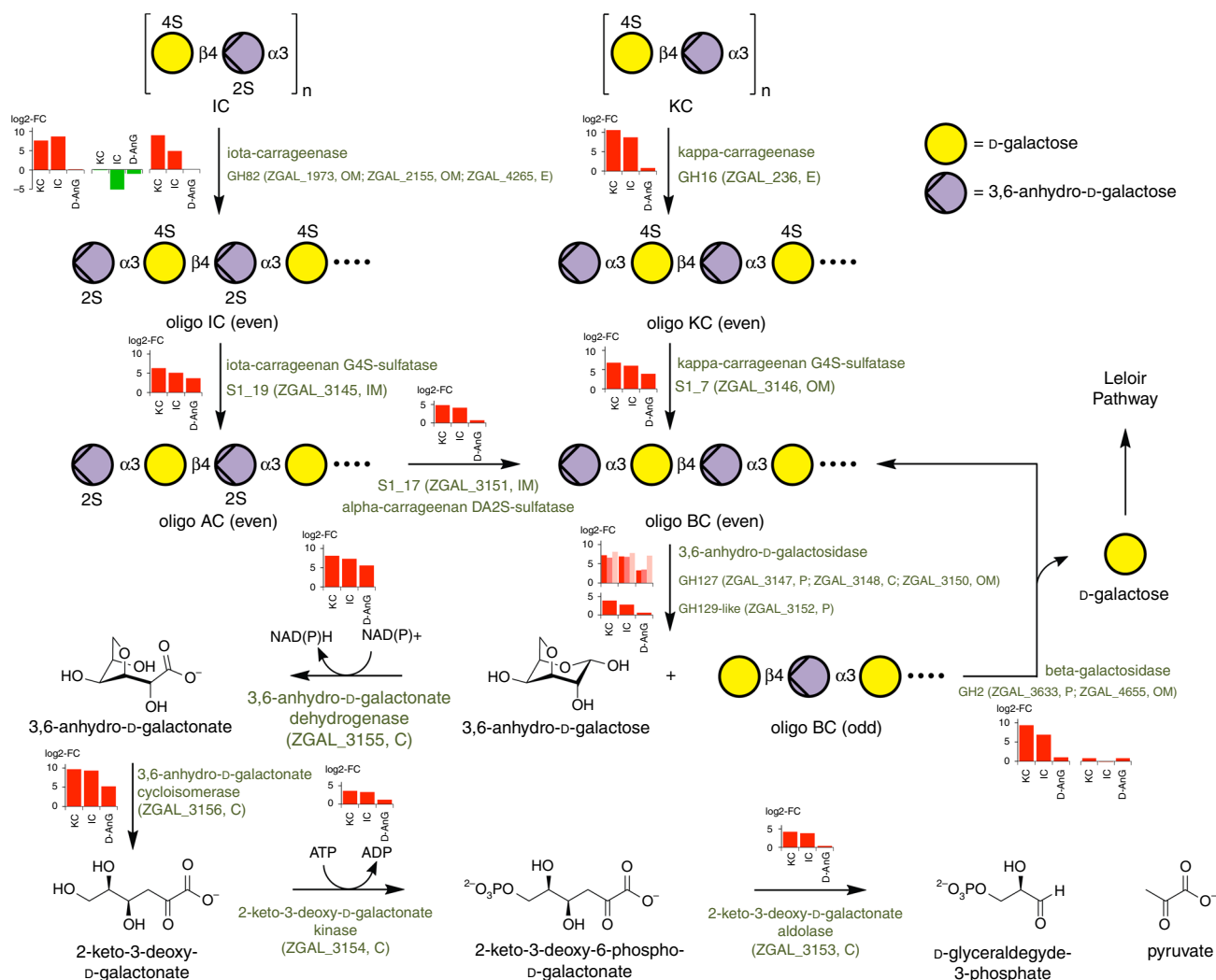
**Fig. 1** *Z. galactanivorans* PUL involved in carrageenan catabolism. Genes are annotated according to their CAZY family (GH127, GH129-like), sulfatase subfamily (S1\_19, S1\_7, and S1\_17) or enzymatic activity. The names of the genes and their functions are given in Table 1. The acronyms for the gene names are as follows: *cgs*, CarraGeenan Sulfatase; *dag*, D-3,6-AnhydroGalactosidase; *dau*, D-3,6-Anhydrogalactose Utilization; *cgr*, CarraGeenan Regulator



**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 a sulfate loss induced by the ionization process. Peaks annotated with a \* correspond to HEPES adducts on matrix clusters. DP stands for degree of polymerization

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 uncharged beta-carrabiose motifs<sup>30</sup>. This hybrid polysaccharide is commonly referred to as furcellaran. No GH activity was detected on furcellaran using fluorophore-assisted carbohydrate electrophoresis (FACE)<sup>31</sup>. 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<sup>3</sup>. 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  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidases which cleave the  $\alpha$ -1,3 linkage between D-AnG and D-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.

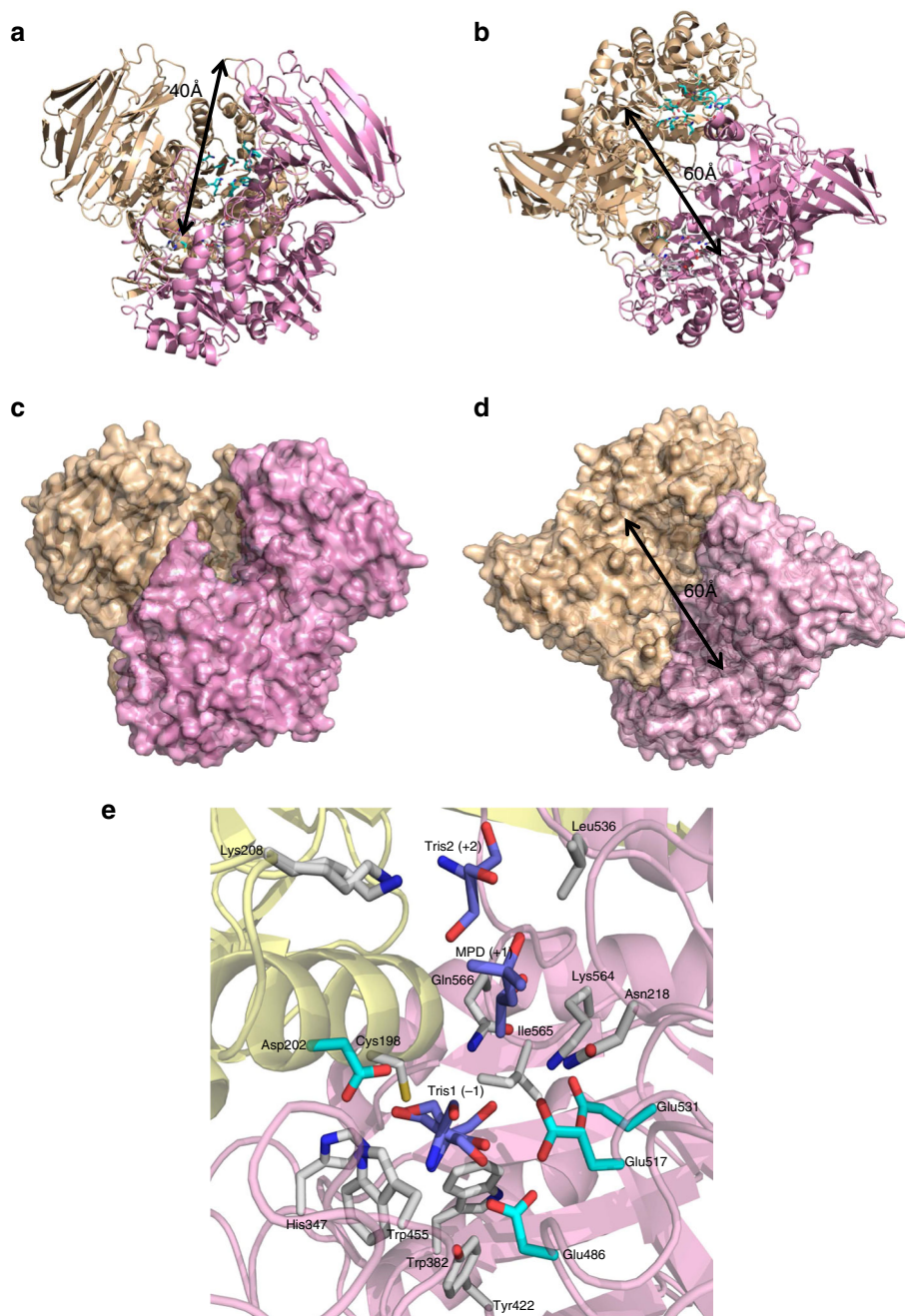


**Fig. 3** Kappa-carrageenan and iota-carrageenan catabolic pathways in *Z. galactanivorans*. Enzymes performing each reaction step are mentioned, together with their encoding genes. For each step, bar plots depict the relative expression of the corresponding gene(s) in kappa-carrageenan (KC), iota-carrageenan (IC), or 3,6-anhydro-D-galactose (D-AnG) compared to D-galactose, as measured by RNA-seq (values are mean of log<sub>2</sub>-fold change,  $n = 3$ ). Predicted enzyme localization is shown next to the gene number: OM, outer membrane; IM, inner membrane; E, extracellular; C, cytoplasmic; P, periplasm. AC stands for alpha-carrageenan and BC for beta-carrageenan. In this schematic, alpha- and beta-carrabiose motifs are considered as intermediary compounds. Nonetheless, alpha- and beta-carrageenans can be naturally found in some red algal species and the catabolic pathway described here is also valid for these natural carrageenans

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 <sup>1</sup>H-NMR was used to determine the natural substrates and the regioselectivity of these sulfatases (Supplementary Discussion). As predicted, all these sulfatases were 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 Fig. 2C). ZGAL\_3145 (*cgsA*) removes the 4-linked sulfate group from the galactose moiety of iota-carrabiose motifs, generating alpha-carrabiose motifs (Supplementary Figs. 3, 4). ZGAL\_3151 (*cgsC*) acts subsequently on the alpha-carrabiose motifs, removing the 2-linked sulfate group from D-AnG to generate beta-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-

deoxygalactonokinase (ZGAL\_3154), lactaldehyde dehydrogenase (ZGAL\_3155), and aldonic acid dehydratase (ZGAL\_3156). However, recent discovery of two enzymes from *Vibrio* sp. *EJY3* that convert 3,6-anhydro-L-galactose (agar component) into 2-dehydro-3-deoxygalactonate<sup>32</sup> provided new insights. Indeed, *Vibrio* 3,6-anhydro-L-galactose dehydrogenase (VEJY3\_09240) and 3,6-anhydro-L-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<sup>+</sup> and NADP<sup>+</sup>, with a 5–6-fold preference for NAD<sup>+</sup> (Supplementary Fig. 5A, B). ZGAL\_3155 is inactive on D-galactose and is therefore a specific 3,6-anhydro-D-galactose dehydrogenase (gene named *dauA*). The resulting 3,6-anhydro-D-galactonate is converted to 2-keto-3-deoxy-D-galactonate 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

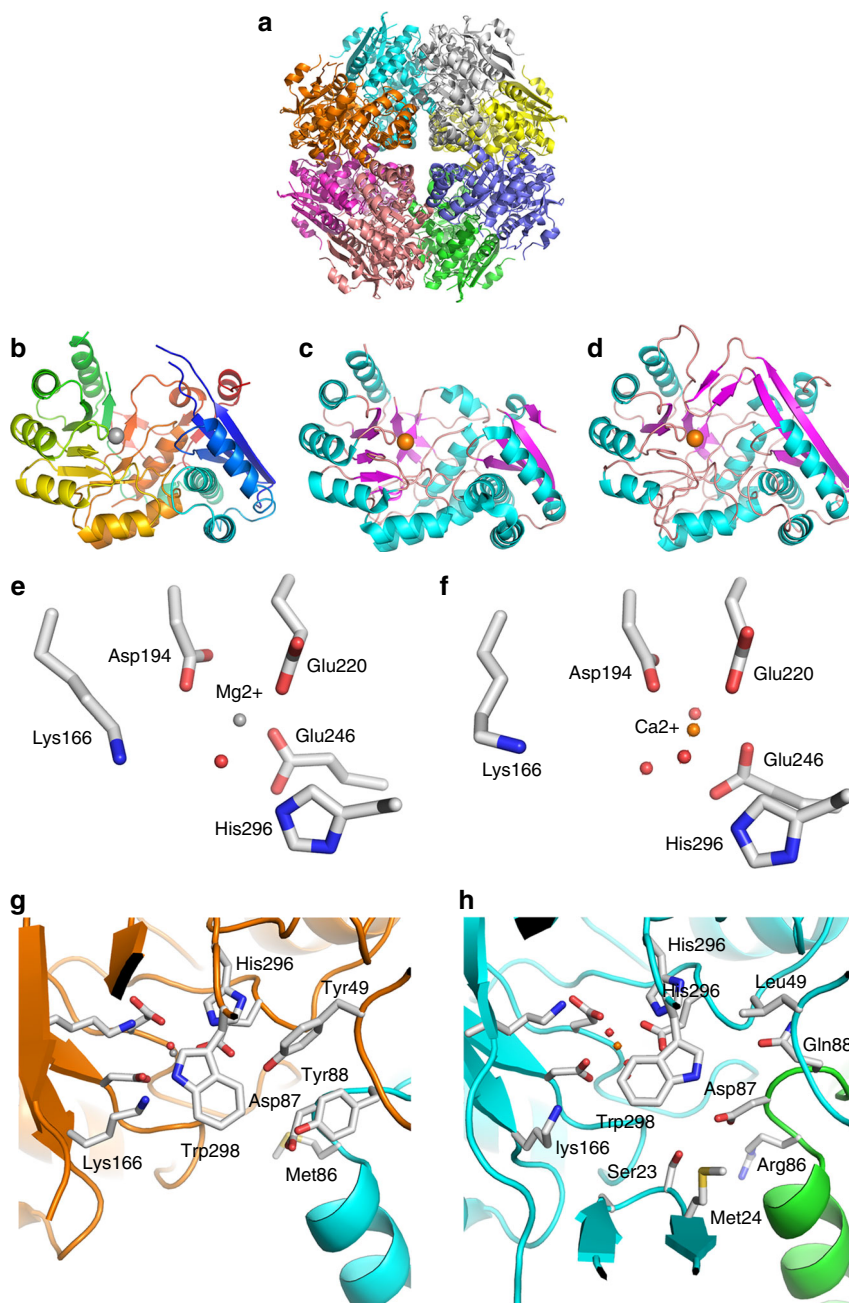


**Fig. 4** X-ray crystal structure of the  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidase DagB. **a** Secondary structure representation of DagB (ZGAL\_3152, GH129-like) showing the depth of the active site crevasse for ZGAL\_3152 (yellow and pink secondary structure representation for each monomer, pdb id 5opq). Active site residues, colored by element, are shown in light blue for one monomer and white for the other. **b** Secondary structure representation showing the width of the active site crevasse. **c** Surface structure representation of the model orientation shown in **a**. **d** Surface structure representation of the model orientation shown in **b**. **e** Active site substructure of the  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidase ZGAL\_3152 (DagB). Three putative subsites are shown as represented by bound Tris1 (subsite -1), MPD (subsite +1) and Tris2 (subsite +2). Four potential catalytically active acidic residues are shown in light blue belonging to two different monomers of ZGAL\_3152. Their site-directed mutagenesis leads to inactive enzymes. The amino acids, colored by element, interact with the Tris and MDP molecules and are conserved at 80% or more within the closest DagB homologs

Fig. 5D) and is thus a 3,6-anhydro-D-galactonate cycloisomerase (*dauB*). In the presence of ATP, ZGAL\_3154 phosphorylated the 2-keto-3-deoxy-D-galactonate to 2-keto-3-deoxy-6-phospho-D-galactonate. The activity of this 2-keto-3-deoxy-D-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-D-galactonate into D-glyceraldehyde-3-phosphate and pyruvate. The activity of this 2-keto-3-deoxy-D-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.<sup>33</sup> recently biochemically characterized homologs of these four enzymes in other marine bacteria.

**Crystal structure of two 3,6-anhydro-D-galactose-related enzymes.** Most recombinant proteins were put into crystal trials



**Fig. 5** X-ray crystal structure of the 3,6-anhydro-D-galactonate cycloisomerase DauB. **a** Cartoon representation of the octamer of DauB (ZGAL\_3156; pdb id 5olc), each chain being colored with a different color. **b** Cartoon representation of the monomer of ZGAL\_3156 (chain A) colored according to a blue-red gradient from N- to C-terminal. **c** Cartoon representation of the low-resolution structure (2.0 Å) of the monomer of the D-galactaro-1,4-lactone cycloisomerase AtGCI from *Agrobacterium tumefaciens* (pdb id 4ggb). The  $\alpha$ -helices and  $\beta$ -strands are colored in magenta and cyan, respectively. **d** Cartoon representation of the high-resolution structure (1.6 Å) of the monomer of AtGCI (pdb id 4hpn) with the entire lid domain visible. **e** and **f** Zoom view of the catalytic machinery of ZGAL\_3156 and AtGCI (pdb id 4hpn), respectively. **g** and **h** Zoom view of the substrate binding pocket of ZGAL\_3156 and AtGCI (pdb id 4hpn), respectively. The chains **h** (in orange) and **b** (cyan) of ZGAL\_3156 are shown **g**. Two equivalent chains of the AtGCI (Cyan and green, respectively) are shown **h**. The crystal corresponding to 4hpn contains only one molecule per asymmetric unit and the biological octamer of AtGCI was thus generated by the crystallographic symmetry

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  $\alpha$ -1,3-(3,6-anhydro)-D-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).

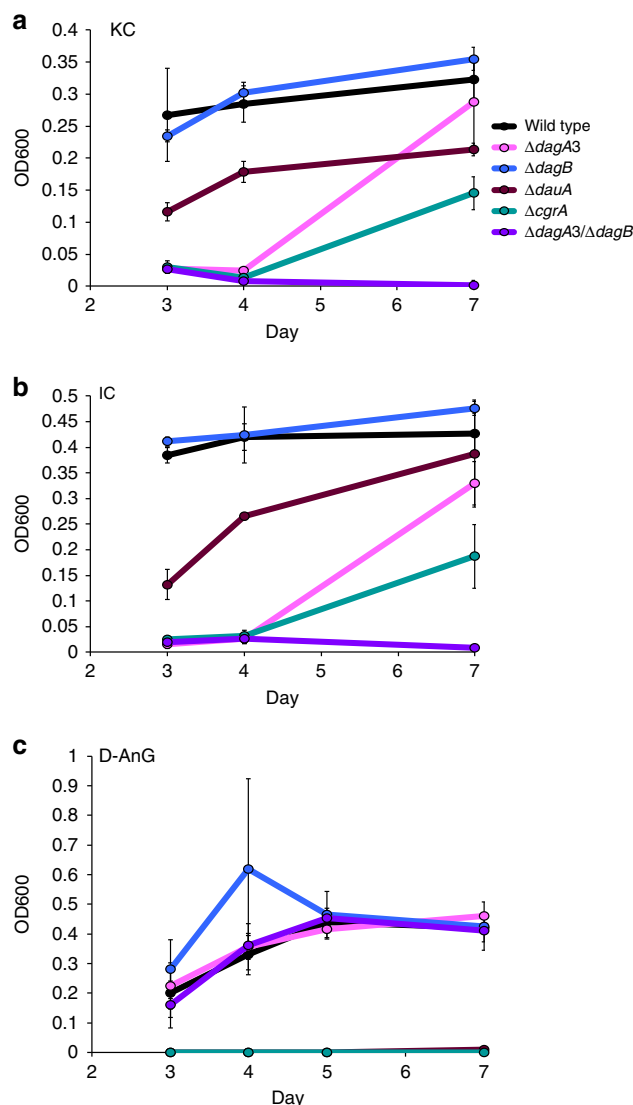
ZGAL\_3152 forms a homodimer and the monomer has a complex architecture with an N-terminal distorted beta-sandwich

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 ( $\eta$ 3 and  $\alpha$ 2) of the other monomer. The dimer is also stabilized by the swapping of the C-terminal strand  $\beta$ 36. 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  $\alpha$ 2 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<sup>28</sup>, 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  $(\beta/\alpha)_7\beta$  TIM-barrel (amino acids 137–340) with an  $\alpha/\beta$  lid domain (amino acids 1–136 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 D-galactaro-1,4-lactone cycloisomerase AtGCI from *Agrobacterium tumefaciens* (41% identity, PDB: 4ggg, Fig. 5c), speaking to the flexibility of the lid region in these enzymes<sup>34</sup>. 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  $\alpha$ 2 and  $\alpha$ 3 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-D-galactonate 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<sup>35</sup>. Four single deletion mutants [ $\Delta dagA3$  ( $\Delta zgal\_3150$ ,  $\Delta gh127-3$ ),  $\Delta dagB$  ( $\Delta zgal\_3152$ ,  $\Delta gh129$ -like),  $\Delta dauA$  ( $\Delta zgal\_3155$ ,  $\Delta 3,6$ -anhydro-D-galactose dehydrogenase),  $\Delta cgrA$  ( $\Delta zgal\_3159$ ,  $\Delta araC$  family regulator)] and a double deletion mutant ( $\Delta dagA3/\Delta dagB$ ) were constructed. They all grew comparably to the wild-type strain in Zobell medium and minimum medium supplemented with agar or



**Fig. 6** Average growth of wild-type *Z. galactanivorans* and deletion mutants. Growth on **a** kappa-carrageenan, **b** iota-carrageenan and **c** 3,6-anhydro-D-galactose (D-AnG) as sole carbon source over a 7-day period. Error bars represent standard error of the mean between three replicates. Legend: *dagA3*,  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidase (*zgal\_3150*, *gh127*); *dagB*,  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidase (*zgal\_3152*, *gh129*-like); *dauA*, 3,6-anhydro-D-galactose dehydrogenase (*zgal\_3155*); *cgrA*, *araC* family transcription factor (*zgal\_3159*)

galactose (Supplementary Fig. 10A, B). The  $\Delta dagB$  mutant had little effect on growth relative to wild type in minimal media containing either KC or IC; however, the  $\Delta dagA3$  mutant showed a significant growth delay and the double  $\Delta dagA3/\Delta dagB$  mutant abolished growth altogether, confirming the importance of these  $\alpha$ -1,3-(3,6-anhydro)-D-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

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

Locus_tag	Description ( <i>gene name/acronym, family</i> )	KC/D-gal		IC/D-gal		D-AnG/D-gal	
		log2-FC*	<i>padj</i>	log2-FC	<i>padj</i>	log2-FC	<i>padj</i>
ZGAL_181	Sulfatase ( <i>cgsB2, S1_7</i> )	<b>8.4</b>	7.47E-94	<b>8.0</b>	5.81E-85	<b>5.1</b>	3.66E-34
ZGAL_236	Kappa-carrageenase ( <i>cgkA, GH16</i> )	<b>10.6</b>	2.68E-163	<b>8.7</b>	7.26E-110	0.8	1
ZGAL_1973	lota-carrageenase ( <i>cgiA3, GH82</i> )	<b>7.6</b>	7.15E-56	<b>8.7</b>	1.12E-72	0.1	1
ZGAL_3145	Sulfatase ( <i>cgsA, S1_19</i> )	<b>6.3</b>	1.7E-65	<b>5.1</b>	1.69E-41	<b>3.7</b>	3.48E-20
ZGAL_3146	Sulfatase ( <i>cgsB1, S1_7</i> )	<b>6.8</b>	1.2E-117	<b>6.0</b>	1.04E-91	<b>3.9</b>	1.20E-33
ZGAL_3147	$\alpha$ -1,3-(3,6-anhydro)-D-galactosidase ( <i>dagA1, GH127</i> )	<b>7.2</b>	7.4E-135	<b>6.9</b>	5.12E-125	<b>3.3</b>	2.67E-25
ZGAL_3148	$\alpha$ -1,3-(3,6-anhydro)-D-galactosidase ( <i>dagA2, GH127</i> )	<b>6.6</b>	2.7E-61	<b>6.8</b>	5.52E-65	<b>3.5</b>	6.80E-16
ZGAL_3149	Sugar permease (MFS)	<b>6.7</b>	3.0E-65	<b>6.2</b>	9.61E-55	<b>4.7</b>	9.94E-31
ZGAL_3150	$\alpha$ -1,3-(3,6-anhydro)-D-galactosidase ( <i>dagA3, GH127</i> )	<b>8.1</b>	2.9E-108	<b>7.8</b>	3.02E-99	<b>7.1</b>	3.78E-82
ZGAL_3151	Sulfatase ( <i>cgsC, S1_17</i> )	<b>4.8</b>	5.4E-44	<b>4.1</b>	1.50E-31	0.7	1
ZGAL_3152	$\alpha$ -1,3-(3,6-anhydro)-D-galactosidase ( <i>dagB, new GH</i> )	<b>3.9</b>	6.5E-23	<b>2.8</b>	5.72E-11	0.6	1
ZGAL_3153	2-keto-3-deoxy-D-galactonate aldolase ( <i>dauD</i> )	<b>4.2</b>	9.9E-44	<b>3.8</b>	2.07E-35	0.4	1
ZGAL_3154	2-keto-3-deoxy-D-galactonate kinase ( <i>dauC</i> )	<b>3.6</b>	2.4E-30	<b>3.3</b>	1.84E-25	1.2	0.1516
ZGAL_3155	3,6-anhydro-D-galactose dehydrogenase ( <i>dauA</i> )	<b>8.1</b>	1.3E-80	<b>7.3</b>	4.01E-65	<b>5.6</b>	1.40E-37
ZGAL_3156	3,6-anhydro-D-galactonate cycloisomerase ( <i>dauB</i> )	<b>9.7</b>	2.0E-89	<b>9.3</b>	1.21E-81	<b>5.2</b>	1.07E-23
ZGAL_3157	Sugar/H <sup>+</sup> symporter (DMT)	<b>8.4</b>	5.8E-140	<b>8.0</b>	3.68E-125	<b>7.0</b>	3.85E-95
ZGAL_3158	High-affinity sugar transporter	<b>4.3</b>	8.1E-39	<b>3.0</b>	4.15E-17	<b>4.1</b>	8.61E-34
ZGAL_3159	Transcriptional regulator ( <i>cgrA, AraC family</i> )	<b>2.3</b>	1.8E-14	<b>1.2</b>	0.0504	-0.5	1
ZGAL_3580	SusD-like lipoprotein ( <i>cgtB</i> )	<b>9.1</b>	2.04E-124	<b>8.3</b>	5.65E-104	<b>4.2</b>	1.17E-24
ZGAL_3581	SusC-like TonB-dependent receptor ( <i>cgtA</i> )	<b>9.6</b>	1.93E-103	<b>9.5</b>	1.31E-99	<b>4.3</b>	4.99E-19
ZGAL_3629	Sulfatase ( <i>S1_30</i> )	<b>8.5</b>	8.26E-55	<b>4.9</b>	2.82E-16	-0.1	1
ZGAL_3630	Sulfatase( <i>S1_28</i> )	<b>7.1</b>	5.54E-48	<b>4.3</b>	6.27E-16	0.8	1
ZGAL_3631	Polygalacturonase, (GH28)	<b>7.8</b>	1.43E-36	<b>3.9</b>	1.70E-07	-0.4	1
ZGAL_3632	Polygalacturonase, (GH28)	<b>7.8</b>	8.43E-20	<b>4.3</b>	1.66E-04	-0.3	1
ZGAL_3633	Beta-galactosidase (GH2)	<b>9.4</b>	1.85E-66	<b>6.9</b>	2.66E-35	1.0	1
ZGAL_3634	Alpha-L-fucosidase (GH29)	<b>8.3</b>	3.04E-28	<b>4.8</b>	9.25E-08	1.8	1
ZGAL_3637	SusC-like TonB-dependent transporter (TBDT)	<b>9.2</b>	1.81E-51	<b>5.1</b>	1.14E-14	0.7	1
ZGAL_3638	SusD-like lipoprotein (SGBP)	<b>9.9</b>	3.92E-53	<b>5.3</b>	8.26E-14	-0.1	1
ZGAL_4265	lota-carrageenase ( <i>cgiA1, GH82</i> )	<b>8.7</b>	2.10E-56	<b>4.7</b>	6.12E-15	0.0	1

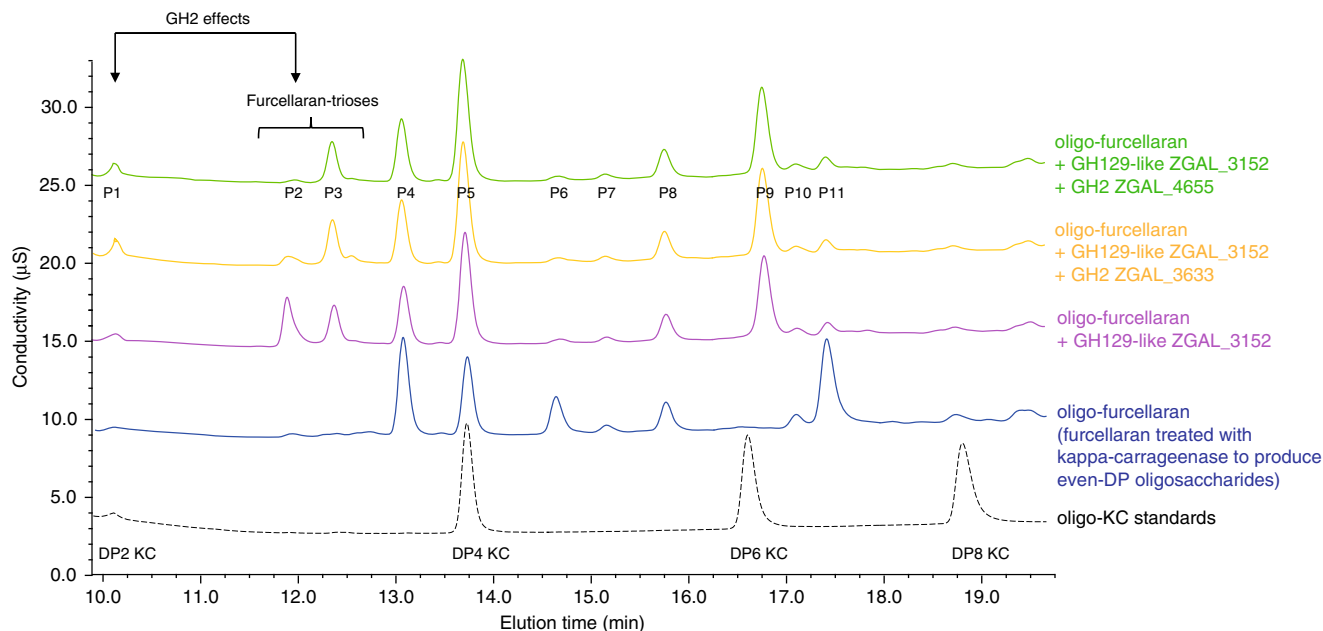
\*Log2-FC: log2-fold change of KC, IC or D-AnG relative to D-gal as unique carbon source. Boldface type indicates significant changes (FWER 5%). Full data set is available in Supplementary Data 1–9

instance, such variations in substrate specificities are known in GH16  $\beta$ -agarases and  $\beta$ -porphyranases which act on other red algal sulfated galactans<sup>36</sup>. 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 in order to sequentially degrade carrageenan oligosaccharides to produce the monosaccharide D-AnG for uptake by the bacterium. Finally, this deletion effect indicates that the two remaining GH127 enzymes are unable to compensate for the loss of ZGAL\_3150 or ZGAL\_3152. The  $\Delta$ *dauA* mutant did not grow on D-AnG and complementation experiments restored growth on this sugar (Fig. 6c, Supplementary Fig. 10C). This indicates that neither ZGAL\_4659 (orthologue of the 3,6-anhydro-L-galactose dehydrogenase VEJY3\_09240, 68% identity) nor any other sugar dehydrogenases from *Z. galactanivorans* could degrade D-AnG and that ZGAL\_3155 (*DauA*, 3,6-anhydro-D-galactose dehydrogenase) is essential for D-AnG catabolism.  $\Delta$ *dauA* had diminished growth on both KC and IC (Fig. 6a, b), suggesting that this mutant is still capable of using the D-galactose units released by carrageenan degradation as a carbon source. The deletion mutant of the AraC family regulator ZGAL\_3159 ( $\Delta$ *cgrA*) showed a significant lag phase and reduced growth on carrageenans (Fig. 6a, b), confirming its importance on the positive regulation of the ZGAL\_3145-3159 cluster function.

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





**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)

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*)<sup>22</sup>; 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*)<sup>23,24</sup> 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<sup>22,23</sup>. 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

Data 1–9); however, ZGAL\_4655 is not induced by kappa- and iota-carrageenans, suggesting that ZGAL\_3633 is the key carrageenan-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; GH82: 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 on 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-porphyranses, 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 carrageenans (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  $\Delta zgal_{3580}/\Delta zgal_{3581}$  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*).

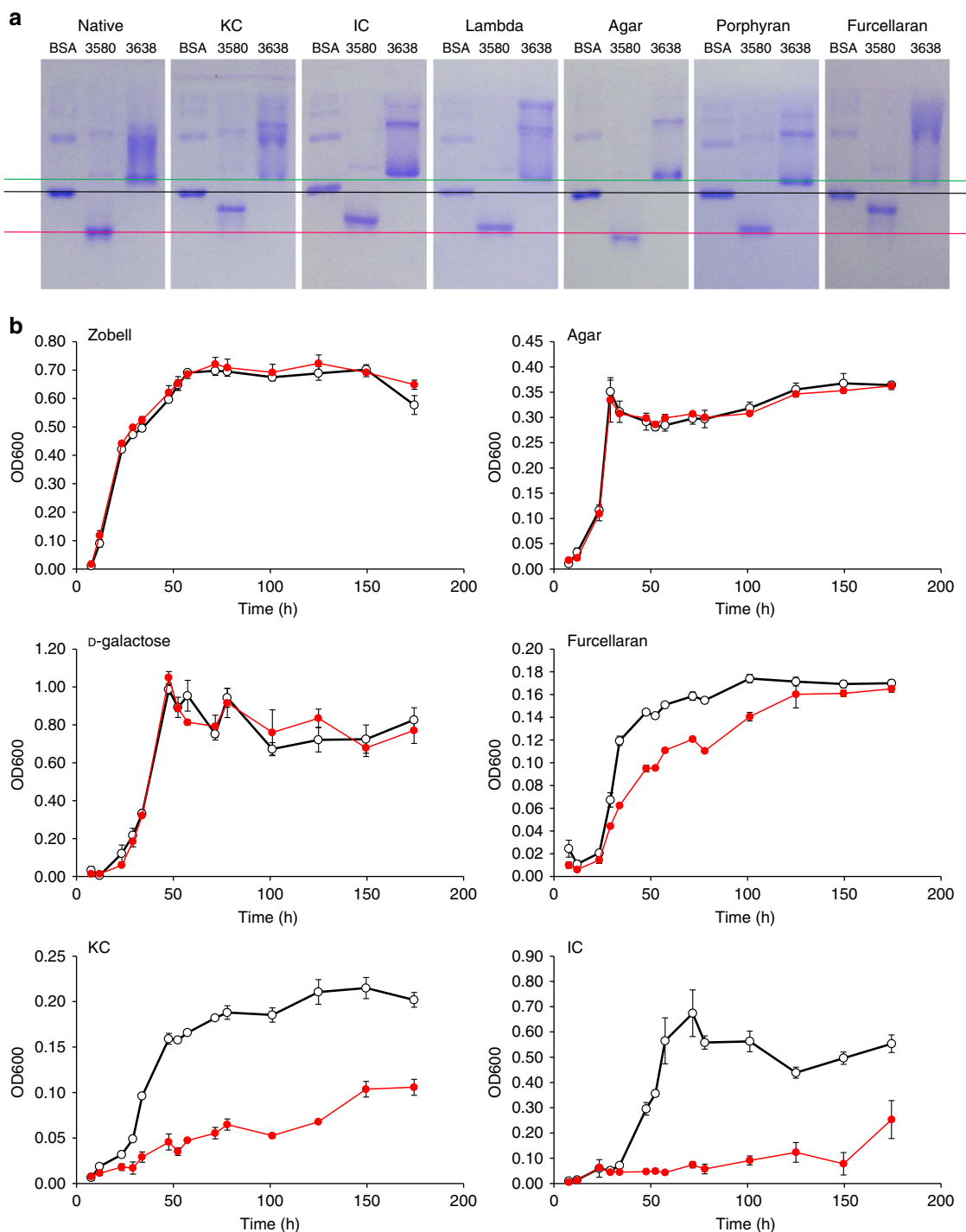
Based on the biochemical, genetic, and transcriptomic evidence presented here, we have demonstrated for the first time a complete scheme for the catabolism of kappa-, iota- and beta-carrageenans in *Z. galactanivorans*, from the initial action of the GH16 kappa-carrageenase (CgkA)<sup>22</sup> and GH82 iota-carrageenases (CgiA1–3)<sup>23,24</sup> to the conversion in four steps of D-AnG into D-glyceraldehyde-3-phosphate and pyruvate (Fig. 3); these latter compounds presumably then enter glycolysis and the citric acid cycle, respectively. The released D-galactose residues are expected to be further converted by the Leloir pathway. The majority of the molecular actors in this pathway and in its regulation are encoded by a discrete genetic locus; however, other key genes are localized remotely on the genome and are part of a

carrageenan-induced gene network (Table 1). These genes are thus part of a complex regulon (referred to as the carrageenan utilization system) and the key genes have been named (Table 1).

### 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 jejuense* 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 (surgeon fish, sea urchin, and abalone). The limits of each PUL were manually refined and 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 5 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 contains 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  $\alpha$ -1,3-(3,6-anhydro)-D-galactosidase activity was due to the GH127 family.

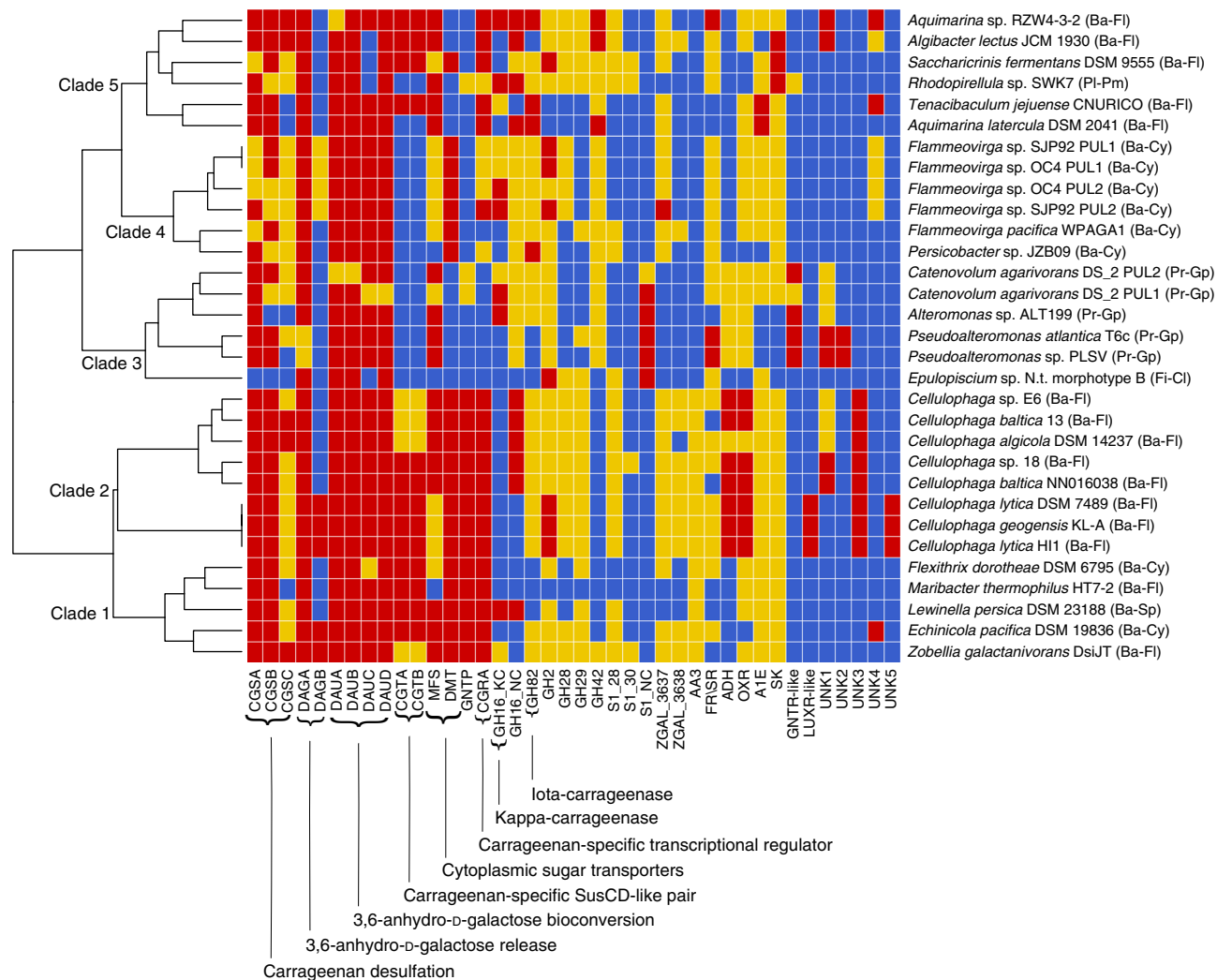
In clades 3, 4, and 5, the structure of the carrageenan-specific PUL is significantly modified. When considering the 30 bacterial species, the core system is restricted to the genes responsible for the release of D-AnG (*dagA*) and for its catabolism (*dauA*, *dauB*, *dauD*). This may be due to horizontal 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-



**Fig. 8** Affinity gel electrophoresis for the SusD-like proteins and phenotyping for the *susCD*-like mutant  $\Delta cgtA-cgtB$ . **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  $\Delta cgtA-cgtB$  (red curves) in rich Zobell media and in minimal media containing different carbohydrate substrates

phosphorylative variant of the Entner-Doudoroff pathway to degrade 2-keto-3-deoxy-D-galactonate, as observed in the archaeon *Picrophilus torridus*<sup>37</sup>. The *susD*-like genes are unique to the *Bacteroidetes*<sup>8</sup> 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<sup>19</sup>. However, this does not mean that TBDT are absent in

the gene clusters of other phyla. For instance, all clade 3 Gammaproteobacteria 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. 9** Conservation of the carrageenolytic PUL organization among bacterial genomes. Putative homologs of selected carrageenan-induced genes from *Z. galactanivorans* were found in 29 other bacterial genomes with the conservation of these genes evaluated by a heatmap. Based on these conservation profiles, the bacterial species cluster into 5 main clades. Red coloring indicates the genes are conserved in a PUL, yellow indicates the genes are conserved elsewhere in the genome and blue indicates the genes are absent in the genome. Gene names for the *Z. galactanivorans* carrageenan-utilization locus are found in Table 1 and the legend of Fig. 1. Abbreviations: CGTA, CarraGeenan Transport A (SusC-like, TonB-dependent transporter); CGTB, CarraGeenan Transport B (SusD-like lipoprotein); GH, glycoside hydrolase; S1, family 1 sulfatase; AA3, auxiliary activity 3; FR/SR, fumarate reductase/succinate dehydrogenase; ADH, aldehyde dehydrogenase; OXR, isoquinoline 1-oxidoreductase alpha subunit; ATE, aldose 1-epimerase; SK, sugar kinase; GNTR-like and LUXR-like, transcription factor families GNTR and LUXR; UNK, unknown function. The phylum and class are indicated beside each organism in parentheses: Ba, *Bacteroidetes*; Pl, *Planctomycetes*; Pr, *Proteobacteria*; Fi, *Firmicutes*; Fl, *Flavobacteriia*; Pm, *Planctomycetia*; Cy, *Cytophagia*; Gp, *Gammaproteobacteria*; Cl, *Clostridia*; Sp, *Sphingobacteriia*

Such clusters containing polysaccharide-related TBDT genes have already been described in Gammaproteobacteria<sup>38</sup>. 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<sup>39</sup>. This is quite clear for the surgeon fish symbiont *Epulopiscium* sp. (Fig. 9, clade 3), which is also known to have horizontally acquired GH16 porphyranase and GH117 genes from marine flavobacteria<sup>40</sup>. Phylogenetic analysis of the GH127 family unraveled another likely HGT case (Supplementary Fig. 13, Supplementary Data 12). Whereas ZGAL\_3148 (DagA2, GH127-2) and ZGAL\_3150 (DagA3, GH127-3) cluster only with homologs from marine bacteria harboring the carrageenan-specific PUL, ZGAL\_3147 (DagA1, GH127-1) is at the root of a clade including GH127 enzymes from human gut *Bacteroides* species. These dagA1-like genes are within PULs including GH genes at first sight unrelated to carrageenans (e.g.

GH78, GH95, GH105; Supplementary Fig. 14). Thus, these *Bacteroides* GH127 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 *susCD*-like pair but maintaining other carrageenan-related genes) are found in bacterial phyla other than *Bacteroidetes*. 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 *Bacteroidetes*, and may consist

of a complex regulon. Moreover, our work experimentally strengthens the recent proposition<sup>41</sup> that the PUL definition should not be restricted to the presence of a *susCD*-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<sup>42</sup>. 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 using 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 ligations 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) cells were transformed with the plasmids containing the gene fragment of interest then grown in the autoinduction Zyp-5052 medium<sup>43</sup> (200 µg mL<sup>-1</sup> ampicillin, 20 °C, 72 h). ZGAL\_3152 (SeMet) was similarly produced in PASM-5052 medium<sup>43</sup>. The sulfatases were produced in 1 L Luria-Bertani medium supplemented with 100 µg/mL ampicillin, until reaching an OD<sub>600</sub> of about 0.7. Sulfatase gene expression was induced with 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) (overnight, 20 °C). In all cases, cells were collected by centrifugation at 3063×g for 30 min. After chemical cell lysis<sup>44</sup>, the lysate was clarified at 13,865×g for 45 min at 4 °C. Using an ÄKTA 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 onto a Superdex S200 column in 20 mM Tris pH 8.0 and 100 mM NaCl. Fractions were again pooled and concentrated. The ExpASY ProtParam tool<sup>45</sup> was used to generate an extinction coefficient for calculation of protein concentration using the A280.

**Marine polysaccharide and oligosaccharide substrates.** Carrageenan polysaccharides were obtained commercially from CP-Kelco. The kappa/mu- and kappa-carrageenans were extracted from *Kappaphycus alvarezii*, the iota/nu- and iota-carrageenans from *Eucheuma denticulatum* and the lambda-carrageenan from tetrasporophytes of *Gigartina skottsbergii*. Fucellaran, a beta-carrageenan 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 using a previously developed protocol<sup>6</sup>. The oligo-iota-carrageenans were incubated with pure sulfatase ZGAL\_3145 to produce oligo-iota/alpha-carrageenans. One volume of oligo-iota-carrageenans (0.5% w/v in H<sub>2</sub>O mQ) was incubated at 37 °C over 48 h with the same volume of the sulfatase at 1.0 mg mL<sup>-1</sup> in 50 mM Tris pH 8.0, 200 mM NaCl, 1 mM CaCl<sub>2</sub>. Oligosaccharides of fucellaran, kappa- and iota-carrageenans were produced using the recombinant kappa-carrageenase from *Pseudoalteromonas carrageenovora*<sup>46</sup> and iota-carrageenase from *Alteromonas fortis*<sup>47</sup>. Carrageenans (0.25%, 5 mL) were treated 48 h at 37 °C with kappa-carrageenase or iota-carrageenase (0.3 mg mL<sup>-1</sup>, 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 hydrolysate (concentrated by rotary evaporation at about 5% w/v) were filtered (0.2 µm, Millipore), and injected on 3 GE Healthcare Superdex 30 prep-grade columns (600 × 26 mm i.d.) mounted in series. The elution was conducted at a flow rate of 1 mL/min at 20 °C using 50 mM (NH<sub>4</sub>)<sub>2</sub>CO<sub>3</sub> 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.<sup>7</sup>. Carrageenan solutions (0.5 % w/v in H<sub>2</sub>O mQ) were incubated in presence of purified sulfatases (0.5 mg mL<sup>-1</sup>) over 20 h at 37 °C in 25 mM Tris pH 8.0, 0.1 M NaCl, 0.5 mM CaCl<sub>2</sub>. For each reaction, a control sample was realized 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 ICS5000 system (Thermo Scientific Dionex). Elutions for the detection of sulfate were performed with isocratic 12 mM NaOH at a flow rate of 1 mL min<sup>-1</sup> (Single Pump-5), and the detection of anions was done by an 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 free sulfate (retention time at 4 min) between samples and their associated blanks. For oligosaccharide detection HPAEC analyzes were conducted on the same system described for sulfate quantification. Elutions were performed at a flow rate of 0.5 mL min<sup>-1</sup> using a NaOH multistep gradient from 8 to 280 mM (40 min). Oligosaccharides were detected by conductivity mode under a current suppression of 50–300 mA.

<sup>1</sup>H NMR. 800 µL of carrageenan (1% w/v in H<sub>2</sub>O mQ) was incubated at 37 °C with the purified sulfatase (0.5 mg/ml), over several days. After checking for sufficient desulfation by HPAEC, incubation media were freeze-dried, exchanged twice then dissolved in D<sub>2</sub>O (99.9%) to 10 mg mL<sup>-1</sup>. Samples were transferred into 5-mm NMR tubes and <sup>1</sup>H-NMR spectra were recorded at 70 °C, on 500 MHz Bruker AVANCE III HD spectrometer, equipped with an indirect 5-mm gradient probehead TBI <sup>1</sup>H/(BB)/<sup>13</sup>C. Chemical shifts were referenced with respect to trimethylsilylpropionic acid (TSP) used as an external standard, and expressed in ppm. For these experiments, the spectra required 64 scans.

**Glycoside hydrolase activity assays.** Small scale GH127 and GH129-like (ZGAL\_3152) reactions were set up in two formats. Volume of 100 mL of Zyp5052 autoinduction medium was inoculated with *E. coli* BL21(DE3) transformed with the pFO4 vector alone as a control or with the GH clones and grown 72 h at 20 °C. Cells were collected from 2 mL of the cultures, resuspended in 2 mL 50 mM HEPES pH 7.5 and 150 mM NaCl and lysed using a French pressure cell. The lysate was clarified at 13,865×g for 45 min and the supernatant was used for the enzyme activity tests which were set up as follows in a 20 µL reaction volume: 0.5% fucellaran hexasaccharide, 150 mM NaCl, 3.75 mM CaCl<sub>2</sub>, 100 mM HEPES pH 7.5 and 2 µL soluble cell lysate. Alternatively, instead of soluble cell lysate, 1 µL of purified enzyme (1 mg mL<sup>-1</sup>) was added to the same reaction mixture described above. Reactions were left overnight at room temperature. For the GH2 enzymes, the sequential digest reactions were set up with 150 µL 0.25% fucellaran oligosaccharides that had been predigested with the purified GH129-like enzyme, as described above, and then heat denatured for 10 mins in a water bath at 90 °C. The 200 µL reactions were incubated overnight at 37 °C in the following conditions: 5 µg purified GH2, 1 mM CaCl<sub>2</sub>, 1 mM MgCl<sub>2</sub> and 20 mM HEPES pH 7.5.

**Fluorophore assisted carbohydrate electrophoresis.** A speed vacuum was used to dry ~50 µg of oligo- or polysaccharide reaction volume to be destined for FACE<sup>31</sup>. Volume of 2 µL of 0.15 M ANTS (8-aminonaphthalene-1,3,6-trisulfonic 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% polyacrylamide 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-D-galactose dehydrogenase activity assay.** The enzymatic activity of ZGAL\_3155 (DauA, 3,6-Anhydro-D-galactose dehydrogenase) was determined spectrophotometrically as a function of the reduction of NAD<sup>+</sup> or NADP<sup>+</sup> using a Spark 10 M microplate reader (Tecan, France). The reaction was performed 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<sup>+</sup> or NADP<sup>+</sup>. 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 D-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 µmoles of NADH produced min<sup>-1</sup> mg<sup>-1</sup> assuming the ε<sub>340 nm</sub> NADH = 6220 M<sup>-1</sup> cm<sup>-1</sup>. 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 29,000×g and the supernatant was used as the substrate for the next enzyme, ZGAL\_3156.

**3,6-Anhydro-D-galactonate cyclisomerase activity assay.** The enzymatic activity of ZGAL\_3156 (DauB, 3,6-Anhydro-D-galactonate cyclisomerase) was determined using the 2-thiobarbituric assay (TBA)<sup>33,48</sup>. Volume of 250 µL of the reaction mixture from the preceding enzymatic step (the conversion of D-AnG into 3,6-anhydro-D-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 the 20 min time point and the reaction was stopped with 1/10 volume of 12% trichloroacetic acid (TCA). Samples were then centrifuged for 10 min at 29,000×g. Volume of 25 µL of the supernatant was then incubated for 20 min at room temperature in the dark with 62.5 µL 25 mM periodic acid in 250 mM H<sub>2</sub>SO<sub>4</sub> to oxidize 2-keto-3-deoxy-D-galactonate. Oxidation was terminated by the addition of 125 µL 2% (w/v) sodium arsenite in 500 mM HCl. 500 µL 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

removed. The color was intensified by the addition of an equal volume of dimethylsulfoxide 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 (DauC, 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 MgCl<sub>2</sub> (Acros organics), 0.8 mM Phospho(enol) pyruvic acid tri(cyclohexylammonium) salt, 0.16 mM β-NADH (Applichem), 8 mM KCl, and 0.95 µL of mix of pyruvate kinase and lactic 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 read every 5 s with a Spark 10 M microplate reader (Tecan, France). The 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\_3154 incubated 10 min with 5 µg of ZGAL\_3153. Second, synthesis of D-KDPGA from pyruvate and D-glyceraldehyde-3-phosphate or D-glyceraldehyde was tested as follows: 30 µL of MES 25 mM pH 6.5, pyruvate 50 mM and D-glyceraldehyde-3-phosphate or D-glyceraldehyde 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%. The 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 polysaccharides was tested by affinity gel electrophoresis. Kappa-carrageenan (from *Kappaphycus alvarezii*), iota-carrageenan (*Eucheuma denticulum*), lambda-carrageenan (Dupont batch 2321914837), 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 diffusion in the JCSG + screen (Qiagen) for ZGAL\_3156 and the PACT screen (Qiagen) for ZGAL\_3152. Optimized conditions were a 1:1 ratio of 0.14 M Na/K-tartrate, 12% PEG 3350 with 9 mg mL<sup>-1</sup> ZGAL\_3152 and 0.2 M tri potassium citrate, 20% PEG 3350 with 25 mg mL<sup>-1</sup> ZGAL\_3156.

**X-ray crystallography data collection and processing.** SeMet ZGAL\_3152 diffraction data were collected at the ESRF on beamline ID-14-4 at the selenium peak with a wavelength of 0.97936 Å ( $f' = -7.40$ ,  $f'' = 6.38$ ) at 100 K. The cryoprotectant used was 15% MPD in mother liquor. The data were processed using MOSFLM<sup>49,50</sup>, Pointless<sup>51</sup> was used to determine the spacegroup and the data were scaled using SCALA<sup>52</sup> within the CCP4 suite of programs<sup>53</sup>. PrepHADData was used to convert the mtz to SHELXS format and SHELX\_CDE<sup>54</sup> was used to identify the Se subsites through SAD. Finally, CRANK<sup>55</sup> was used to construct the initial model of ZGAL\_3152 through substructure refinement and model building. This model was then used to solve the structure of the native ZGAL\_3152 by molecular replacement using MOLREP<sup>56</sup> (pdb id 5opq, Supplementary Table 1). The ZGAL\_3152 structure went through an iterative process of refinement using REFMAC5<sup>57</sup> and model building using WinCoot<sup>58</sup>. Ramachandran statistics were 96.6% most favored, 3.2% additional allowed and 0.2% disallowed.

Native diffraction data for ZGAL\_3156 was collected on the Proxima2 beamline at SOLEIL using an EIGER detector (Dectris), with a wavelength of 0.98 Å and an oscillation range of 0.1° at 100 K. The crystals were cryoprotected by short immersion in mother liquor supplemented with 15% Glycerol. The data were processed and scaled using XDS and XSCALE<sup>59,60</sup>, Pointless was used to determine the spacegroup<sup>51</sup>. The structure (pdb id 5olc, Supplementary Table 1) was solved by molecular replacement using PHASER<sup>61</sup> and pdb id 4hpn using the biological assembly (octamer) as the primary model. The ZGAL\_3156 structure went through an iterative process of refinement using BUSTER<sup>62</sup> and model building using WinCoot<sup>58</sup>. Ramachandran statistics were 94.4% most favored, 5.3% additional allowed and 0.3% disallowed.

**MALDI-TOF-MS analysis.** MALDI-TOF-MS spectra were acquired in negative ionization mode and reflector detection in the m/z range 500–1500 on an Auto-flexSpeed TOF/TOF mass spectrometer (Bruker Daltonics, Bremen, Germany), equipped with a Smartbeam Laser (355 nm, 1000 Hz). Samples were solubilized in water (500 µg mL<sup>-1</sup>). Volume of 1 µL of the samples were mixed with 1 µL of a DMA-DHB matrix solution prepared as described<sup>63</sup> directly on a polished steel MALDI target plate. Acquisition parameters (laser power, pulsed ion extraction, and so on) were optimized on the sample without any enzyme process. Spectra were recorded using FlexControl 3.4 and processed using FlexAnalysis 3.4 (Bruker Daltonics, Bremen, Germany).

**Construction of *dagA3*, *dagB*, *dauA*, and *cgrA* deletion mutants.** Single deletion mutants for *dagA3* (ZGAL\_3150, GH127-3), *dagB* (ZGAL\_3152, GH129-like), *dauA* (ZGAL\_3155, 3,6-Anhydro-D-galactose dehydrogenase), and *cgrA* (ZGAL\_3159, AraC family regulator), and a double deletion mutant for *dagA3/dagB* were constructed following the previously described method<sup>35</sup>, based on a *sacB* system (Supplementary Tables 3, 4). To delete *dagA3*, a 2.1-kbp fragment including the first 33 bp of *dagA3* and 2,058 bp of upstream sequence was amplified using primers OFT0001 and OFT0003. The fragment was digested with PstI and Sall and ligated into pYT313 that had been digested with the same enzymes, to generate pFT2. A 2.1 kb fragment including the final 33 bp of *dagA3* and 2,013 bp of downstream sequence was amplified using primers OFT0002 and OFT0004. The fragment was cloned into Sall and BamHI sites of pFT2 to generate the *dagA3* deletion construct pFT5. To delete *dagB*, a 2.1-kbp fragment including the first 48 bp of ZGAL\_3152 and 2047 bp of upstream sequence was amplified using primers ORL664 and ORL665. The fragment was digested with BamHI and Sall and ligated into pYT313 that had been digested with the same enzymes, to generate pRF8. A 2.2 kb fragment including the final 3 bp of *dagB* and 2172 bp of downstream sequence was amplified using primers ORL666 and ORL667. The fragment was cloned into Sall and SphI sites of pRF8 to generate the *dagB* deletion construct pRF9. To delete *dauA*, a 2.0-kbp fragment including the first 81 bp of *dauA* and 1887 bp of upstream sequence was amplified using primers ORL670 and ORL671. The fragment was digested with BamHI and Sall and ligated into pYT313 that had been digested with the same enzymes, to generate pRF10. A 2.2 kb fragment including the final 69 bp of *dauA* and 2116 bp of downstream sequence was amplified using primers ORL672 and ORL673. The fragment was cloned into Sall and PstI sites of pRF10 to generate the *dauA* deletion construct pRF11. To delete *cgrA*, a 2.1-kbp fragment including the first 5 bp of *cgrA* and 2030 bp of upstream sequence was amplified using primers ORL676 and ORL677. The fragment was digested with BamHI and Sall and ligated into pYT313 that had been digested with the same enzymes, to generate pRF12. A 2.2 kb fragment including the final 73 bp of *cgrA* and 2064 bp of downstream sequence was amplified using primers ORL678 and ORL679. The fragment was cloned into Sall and PstI sites of pRF12 to generate the *cgrA* deletion construct pRF13. Plasmids pFT5, pRF9, pRF11 and pRF13 were introduced individually into the wild-type *Z. galactanivorans* Dsj1<sup>T</sup> by conjugation from *E. coli* S17\_1 strains. Conjugants with plasmids integrated in the genome were isolated on *Cytophaga*-agar containing 50 µg mL<sup>-1</sup> erythromycin. Single erythromycin-resistant colonies were grown overnight in *Cytophaga* medium in the absence of antibiotics at 30 °C. The cells where a second recombination event resulted in loss of the plasmid were selected on *Cytophaga*-agar containing 5% sucrose. Isolated colonies were checked for erythromycin sensitivity. Deletions were confirmed by PCR on isolated colonies using primer pairs OFT0005-OFT0006 to identify the *dagA3* deletion mutant (mZG\_0026), ORL668-ORL669 to identify the *dagB* deletion mutant (mZG\_0007), ORL674-ORL675 to identify the *dauA* deletion mutant (mZG\_0008) and ORL680-ORL681 to identify the *cgrA* deletion mutant (mZG\_0011). A double deletion mutant for both *dagA3* and *dagB* (mZG\_0047) was constructed by introducing pRF9 into  $\Delta$ *dagA3* by conjugation, followed by sucrose selection and PCR confirmation as described above.

**Construction of *cgtA-cgtB* double gene deletion mutant.** To delete *cgtA-cgtB* [*zgal\_3581* (*susC*-like)-*zgal\_3580* (*susD*-like)], a 2.1-kbp fragment including the first 54 bp of *cgtA* and 2060 bp of upstream sequence was amplified using primers 2088 and 2089 (Supplementary Tables 3, 4). The fragment was digested with SacI and SpeI and ligated into pYT354 that had been digested with the same enzymes, to generate pYT381. A 2.2 kb fragment including the final 54 bp of *cgtB* and 2,169 bp of downstream sequence was amplified using primers 2090 and 2091. The fragment was cloned into SpeI and SphI sites of pYT381 to generate the *cgtA-cgtB* deletion construct pYT382. Primers 2102 and 2103 were used to identify the *cgtA-cgtB* deletion mutant (mZG\_0054).

**Complementation of *dauA* and *cgrA* deletion mutants.** Deletion mutants  $\Delta$ *dauA* and  $\Delta$ *cgrA* were complemented by ectopic plasmid integration at a neutral site, as previously described<sup>35</sup> (Supplementary Tables 3 and 4). To complement  $\Delta$ *dauA*, promoterless *dauA* was amplified with primers 1978 and 1979 and cloned into the XbaI and SphI sites of pYT356 to generate pYT360. Plasmid pYT360 was inserted into  $\Delta$ *dauA* by conjugation to obtain the complemented strain  $\Delta$ *dauA* + CP (complementation plasmid) (mZG\_0039). To complement  $\Delta$ *cgrA*, promoterless *cgrA* was amplified with primers 1980 and 1981 and cloned into the XbaI and SphI sites of pYT356 to generate pYT361. Plasmid pYT361 was inserted into  $\Delta$ *cgrA*

by conjugation to obtain the complemented strain  $\Delta cgrA + CP$  (mZG\_0043). Control strains with the empty vector pYT356 inserted into the chromosome were generated in a similar way. In all cases, cells with integration of the plasmid at the neutral site were selected on *Cytophaga*-agar containing  $50 \mu\text{g mL}^{-1}$  erythromycin and screened by PCR.

**Liquid growth tests.** *Z. galactanivorans* strains were routinely grown from glycerol stocks in Zobell medium 2216E ( $5 \text{ g L}^{-1}$  tryptone,  $1 \text{ g L}^{-1}$  yeast extract, filtered seawater)<sup>64</sup>. Before use, all pre-cultures were collected by centrifugation 10 min at  $2844\times g$ , washed in 2 volumes of sterile saline solution, and resuspended in sterile saline solution to the same OD<sub>600</sub>. Fifty microliters of bacterial suspension were inoculated into triplicate 40 mL flasks containing 5 mL of Zobell medium or marine mineral medium (MMM) composed for 1 L of  $24.7 \text{ g NaCl}$ ,  $6.3 \text{ g MgSO}_4 \cdot 7\text{H}_2\text{O}$ ,  $4.6 \text{ g MgCl}_2 \cdot \text{H}_2\text{O}$ ,  $2 \text{ g NH}_4\text{Cl}$ ,  $0.7 \text{ g KCl}$ ,  $0.6 \text{ g CaCl}_2$ ,  $200 \text{ mg NaHCO}_3$ ,  $100 \text{ mg K}_2\text{HPO}_4$ ,  $50 \text{ mg}$  yeast extract, and  $20 \text{ mg FeSO}_4 \cdot 7\text{H}_2\text{O}$ <sup>19</sup> and supplemented with  $4 \text{ g L}^{-1}$  kappa-carrageenan, iota-carrageenan, agar, D-galactose or D-AnG. Growth at  $20^\circ\text{C}$  under  $160 \text{ r.p.m.}$  was followed by measuring the absorbance at  $600 \text{ nm}$  in a Spark 10 microplate reader (Tecan) on  $200 \mu\text{L}$  aliquots. Erythromycin ( $10 \mu\text{g mL}^{-1}$ ) was added to all experiments with strains mZG\_0031, mZG\_0037, mZG\_0039, mZG\_0041, and mZG\_0043 (Supplementary Table 4).

**Solid growth tests.** Degradation of carrageenans was tested on either ZoBell or MMM medium solidified with  $10 \text{ g L}^{-1}$  kappa-carrageenan or  $20 \text{ g L}^{-1}$  iota-carrageenan. Two microliters of cell suspension prepared as described above were spotted in the center of the Petri dish, and incubated at  $30^\circ\text{C}$ . Degradation was evidenced by the formation of a hole for kappa-carrageenan or liquefaction for iota-carrageenan.

**RNA-seq expression profiling.** *Bacterial strain and culture conditions.* The type strain Dsij<sup>T</sup> of *Z. galactanivorans* was routinely grown in Zobell medium 2216E (Difco) at  $28^\circ\text{C}$ ,  $170 \text{ r.p.m.}$  For transcriptome profiling, cells were cultivated in synthetic Marine Mineral Medium (MMM)<sup>65</sup>. MMM was supplemented with different carbon sources: D-galactose (Sigma-Aldrich #G0750), iota-carrageenan (Danisco, 2544-88-20), kappa-carrageenan (Coffoni  $\times 6913$ ) or D-AnG (Dextra Laboratories #G0002). Briefly, overnight cultures performed in Zobell medium were diluted at OD<sub>600</sub> 0.05 in triplicate MMM containing 0.5% glucose as C source and incubated at  $28^\circ\text{C}$  until reaching the stationary phase. These cultures were then used to inoculate  $10 \text{ mL}$  of MMM containing  $0.4 \text{ g L}^{-1}$  D-galactose, D-AnG, kappa- or iota-carrageenan. When cell density reached an OD<sub>600</sub> of  $0.7 (\pm 0.1)$  bacteria were collected by centrifugation for 3 min at  $4^\circ\text{C}$  after addition of  $\frac{1}{2}$  volume of frozen killing buffer ( $20 \text{ mM Tris-HCl}$  pH 7.5,  $5 \text{ mM MgCl}_2$ ,  $20 \text{ mM NaN}_3$ ) to the culture sample. The cell pellets were frozen in liquid nitrogen and stored at  $-80^\circ\text{C}$  until RNA extraction.

**RNA extraction.** The cell pellets were resuspended into  $800 \mu\text{L}$  of lysis buffer ( $4 \text{ M}$  guanidine thiocyanate,  $25 \text{ mM}$  sodium acetate pH 5.2,  $5 \text{ g L}^{-1}$  N-laurylsarcosinate), immediately mixed to  $1 \text{ mL}$  hot acid phenol (Sigma #P4682) and incubated during 5 min at  $65^\circ\text{C}$  for efficient cell lysis. The aqueous phases were recovered after addition of  $1 \text{ mL}$  chloroform and centrifugation at  $16,000\times g$  during 10 min at room temperature. The samples were extracted at least four times with an equal volume of acid phenol:chloroform:IAA ( $25:24:1$ , (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, Palo Alto, USA). Total RNA extracts were treated using DNase I (Qiagen) to remove residual genomic DNA.

**mRNA enrichment, RNA-seq library preparation and sequencing.** mRNA enrichment was performed using the Illumina Ribo-Zero rRNA Removal Kit (Bacteria). The rRNA-depleted samples were purified using a RNeasy MinElute Cleanup Kit (Qiagen) and efficient rRNA depletion was confirmed by Bioanalyzer 2100. Strand-specific RNA-seq libraries were generated with the ScriptSeq V2 RNA-Seq Library Preparation Kit (Illumina), according to the manufacturer's instructions. Briefly,  $2\text{--}4 \mu\text{g}$  of RNA samples were fragmented, di-tagged cDNAs were synthesized by successive random priming with terminal-tagging oligos and then purified with the Ampure bead XP system. Enrichment of purified cDNAs was done by no more than 14 PCR cycles. The sequencing was performed with a NextSeq 500 platform using a single-end 75 bases run. A total of 3.5–11 million passing filters reads were obtained per sample.

**Read mapping and differential expression analysis.** Sequencing reads were pre-processed for trimming of adapter sequences with cutadapt-1.9.1, then using PRINSEQ<sup>66</sup> for read quality with filtering (-min\_len 20 -min\_qual\_mean 20 -min\_qual\_score 10) and trimming (trim\_qual\_right 10) options. Mapping was performed on *Z. galactanivorans* Dsij<sup>T</sup> reference genome (retrieved from MicroScope "zobellia\_gal\_DsijT\_v2"; Refseq NC\_015844.1) by Bowtie2 software with the very-sensitive option<sup>67</sup>. Alignments files were then converted to BAM files using SAMtools<sup>68</sup>, log2-genome coverage files were computed as described<sup>69</sup> and these expression profiles can be visualized in the Artemis viewer<sup>70</sup>. The number of reads mapping to each predicted CDS was calculated for the 12 data sets by HTSeq-

count<sup>71</sup> with the -m union option. Differential expression analysis (normalization and statistical tests) was performed using the SARTools with DESeq2 software<sup>72</sup> and a bonferroni *p*-value adjustment was used to correct for multiple testing. Genes with an adjusted *p*-value  $< 0.05$  were considered as differentially expressed. 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<sup>73</sup> (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 amino acid identity and 80% coverage threshold. Results were then 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 v7.307<sup>74</sup>, then ambiguous and saturated regions were removed with BMGE v1.12 (with the gap rate parameter set to 0.5)<sup>75</sup>. The best fitting model of amino acid substitution for this data set was selected with ProtTest v3.4.2<sup>76</sup>. A Maximum-Likelihood phylogenetic tree was generated with the alignment using PhyML 3.1.0.2<sup>77</sup> 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 tool<sup>78</sup> (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 5olc). The RNA-seq transcriptome data have been deposited in the GEO database (GEO accession number: GSE101142). The sequence of the *Tenacibaculum jejuense* genome (used in 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

- Anderson, N. S., Dolan, T. C. & Rees, D. A. Evidence for a common structural pattern in the polysaccharide sulphates of the *Rhodophyceae*. *Nature* **205**, 1060–1062 (1965).
- Martin, M., Portetelle, D., Michel, G. & Vandenberg, M. Microorganisms living on macroalgae: diversity, interactions, and biotechnological applications. *Appl. Microbiol. Biotechnol.* **98**, 2917–2935 (2014).
- Michel, G. et al. 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 (2001).
- 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).
- Guibert, M. et al. Degradation of lambda-carrageenan by *Pseudoalteromonas carrageenovora* lambda-carrageenase: a new family of glycoside hydrolases unrelated to kappa- and iota-carrageenases. *Biochem. J.* **404**, 105–114 (2007).
- Prechoux, A., Genicot, S., Rogniaux, H. & Helbert, W. Controlling carrageenan structure using a novel formylglycine-dependent sulfatase, an endo-4S-iota-carrageenan sulfatase. *Mar. Biotechnol.* **15**, 265–274 (2013).
- Prechoux, A., Genicot, S., Rogniaux, H. & Helbert, W. Enzyme-assisted preparation of furcellaran-like kappa-/beta-carrageenan. *Mar. Biotechnol.* **18**, 133–143 (2016).
- Thomas, F., Hehemann, J. H., Rebuffet, E., Czjzek, M. & Michel, G. Environmental and gut bacteroidetes: the food connection. *Front. Microbiol.* **2**, 93 (2011).
- Teeling, H. et al. Substrate-controlled succession of marine bacterioplankton populations induced by a phytoplankton bloom. *Science* **336**, 608–611 (2012).
- Barbeyron, T. et al. Habitat and taxon as driving forces of carbohydrate catabolism in marine heterotrophic bacteria: example of the model algae-associated bacterium *Zobellia galactanivorans* DsijT. *Environ. Microbiol.* **18**, 4610–4627 (2016).

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**, 3199–3204 (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. Bjursell, M. K., Martens, E. C. & Gordon, J. I. 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 (2006).
14. Terrapon, N., Lombard, V., Gilbert, H. J. & Henrissat, B. Automatic prediction of polysaccharide utilization loci in *Bacteroidetes* species. *Bioinformatics* **31**, 647–655 (2015).
15. Sonnenburg, E. D. et al. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell* **141**, 1241–1252 (2010).
16. Larsbrink, J. et al. A discrete genetic locus confers xyloglucan metabolism in select human gut *Bacteroidetes*. *Nature* **506**, 498–502 (2014).
17. Cuskin, F. et al. Human gut *Bacteroidetes* can utilize yeast mannan through a selfish mechanism. *Nature* **517**, 165–169 (2015).
18. Ndeh, D. et al. Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
19. Thomas, F. et al. 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 (2012).
20. Hehemann, J. H., Kelly, A. G., Pudlo, N. A., Martens, E. C. & Boraston, A. B. Bacteria of the human gut microbiome catabolize red seaweed glycans with carbohydrate-active enzyme updates from extrinsic microbes. *Proc. Natl Acad. Sci. USA* **109**, 19786–19791 (2012).
21. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).
22. Barbeyron, T., Gerard, A., Potin, P., Henrissat, B. & Kloareg, B. The kappa-carrageenase of the marine bacterium *Cytophaga drobachiensis*. 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 family GH82 iota-carrageenases. *Biochemistry* **49**, 7590–7599 (2010).
25. Craigie J. Cell Walls. in *Biology of the red algae* (eds Cole K., Sheath R.). (Cambridge University Press, Cambridge, 1990).
26. Barbeyron, T. et al. Matching the diversity of sulfated biomolecules: creation of a classification database for sulfatases reflecting their substrate specificity. *PLoS ONE* **11**, e0164846 (2016).
27. Fujita, K., Takashi, Y., Obuchi, E., Kitahara, K. & Saganuma, T. Characterization of a novel beta-L-arabinofuranosidase in *Bifidobacterium longum*: functional elucidation of a DUF1680 protein family member. *J. Biol. Chem.* **289**, 5240–5249 (2014).
28. Kiyohara, M. et al. Alpha-N-acetyl galactosaminidase from infant-associated bifidobacteria belonging to novel glycoside hydrolase family 129 is implicated in alternative mucin degradation pathway. *J. Biol. Chem.* **287**, 693–700 (2012).
29. Popper, Z. A. et al. Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu. Rev. Plant Biol.* **62**, 567–590 (2011).
30. Correc, G. et al. Comparison of the structures of hybrid  $\kappa$ - $\beta$ -carrageenans extracted from *Furcellaria lumbricalis* and *Tichocarpus crinitus*. *Carbohydr. Polym.* **88**, 31–36 (2012).
31. Jackson, P. The use of polyacrylamide-gel electrophoresis for the high-resolution separation of reducing saccharides labelled with the fluorophore 8-aminonaphthalene-1,3,6-trisulphonic acid. Detection of picomolar quantities by an imaging system based on a cooled charge-coupled device. *Biochem. J.* **270**, 705–713 (1990).
32. Yun, E. J. et al. The novel catabolic pathway of 3,6-anhydro-L-galactose, the main component of red macroalgae, in a marine bacterium. *Environ. Microbiol.* **17**, 1677–1688 (2015).
33. Lee, S. B., Kim, J. A. & Lim, H. S. Metabolic pathway of 3,6-anhydro-D-galactose in carrageenan-degrading microorganisms. *Appl. Microbiol. Biotechnol.* **100**, 4109–4121 (2016).
34. Vetting, M. W., Bouvier, J. T., Gerlt, J. A. & Almo, S. C. Purification, crystallization and structural elucidation of D-galactaro-1,4-lactone cycloisomerase from *Agrobacterium tumefaciens* involved in pectin degradation. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **72**, 36–41 (2016).
35. Zhu, Y. et al. Genetic analyses unravel the crucial role of a horizontally acquired alginate lyase for brown algal biomass degradation by *Zobellia galactanivorans*. *Environ. Microbiol.* **19**, 2164–2181 (2017).
36. Hehemann, J. H. et al. Biochemical and structural characterization of the complex agarolytic enzyme system from the marine bacterium *Zobellia galactanivorans*. *J. Biol. Chem.* **287**, 30571–30584 (2012).
37. Reher, M., Fuhrer, T., Bott, M. & Schönheit, P. The nonphosphorylative Entner-Doudoroff pathway in the thermoacidophilic euryarchaeon *Picrophilus torridus* involves a novel 2-keto-3-deoxygluconate-specific aldolase. *J. Bacteriol.* **192**, 964–974 (2010).
38. Blanvillain, S. et al. Plant carbohydrate scavenging through tonB-dependent receptors: a feature shared by phytopathogenic and aquatic bacteria. *PLoS ONE* **2**, e224 (2007).
39. Hehemann, J. H. et al. Transfer of carbohydrate-active enzymes from marine bacteria to Japanese gut microbiota. *Nature* **464**, 908–912 (2010).
40. Rebuffet, E. et al. Discovery and structural characterization of a novel glycosidase family of marine origin. *Environ. Microbiol.* **13**, 1253–1270 (2011).
41. Grondin J. M., Tamura K., Dejean G., Abbott D. W. & Brumer H. Polysaccharide utilization loci: fuelling microbial communities. *J. Bacteriol.* doi:10.1128/JB.00860-16.
42. Groisillier, A. et al. MARINE-EXPRESS: taking advantage of high throughput cloning and expression strategies for the post-genomic analysis of marine organisms. *Microb. Cell Fact.* **9**, 45 (2010).
43. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
44. Ficko-Blean, E. et al. Biochemical and structural investigation of two paralogous glycoside hydrolases from *Zobellia galactanivorans*: novel insights into the evolution, dimerization plasticity and catalytic mechanism of the GH117 family. *Acta Crystallogr. D* **71**, 209–223 (2015).
45. Gasteiger E., et al. Protein identification and analysis tools on the ExPASy Server. in *The Proteomics Protocols Handbook* (ed. Walker J. M.) (Humana Press, Totowa, New Jersey, USA, 2005).
46. Michel, G. et al. Expression, purification, crystallization and preliminary x-ray analysis of the kappa-carrageenase from *Pseudoalteromonas carrageenovora*. *Acta Crystallogr. D* **55**, 918–920 (1999).
47. Michel, G. et al. Expression, purification, crystallization and preliminary X-ray analysis of the iota-carrageenase from *Alteromonas fortis*. *Acta Crystallogr. D* **56**, 766–768 (2000).
48. Buchanan, C. L., Connaris, H., Danson, M. J., Reeve, C. D. & Hough, D. W. An extremely thermostable aldolase from *Sulfolobus solfataricus* with specificity for non-phosphorylated substrates. *Biochem. J.* **343**, 563–570 (1999). Pt 3.
49. Powell, H. R. The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. D* **55**, 1690–1695 (1999).
50. Leslie A. G. W. Recent changes to the MOSFLM package for processing film and image plate data. *Jnt. CCP4/ESF-EACBM Newslett. Protein Crystallogr.* **26**, (1992).
51. Evans, P. R. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. D* **67**, 282–292 (2011).
52. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
53. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
54. Schneider, T. R. & Sheldrick, G. M. Substructure solution with SHELXD. *Acta Crystallogr. D* **58**, 1772–1779 (2002).
55. Ness, S. R., de Graaff, R. A., Abrahams, J. P. & Pannu, N. S. CRANK: new methods for automated macromolecular crystal structure solution. *Structure* **12**, 1753–1761 (2004).
56. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
57. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. *Acta Crystallogr. D* **60**, 2184–2195 (2004).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
59. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
60. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D* **66**, 133–144 (2010).
61. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
62. Smart, O. S. et al. Exploiting structure similarity in refinement: automated NCS and target-structure restraints in BUSTER. *Acta Crystallogr. D* **68**, 368–380 (2012).
63. Ropartz, D. et al. Performance evaluation on a wide set of matrix-assisted laser desorption ionization matrices for the detection of oligosaccharides in a high-throughput mass spectrometric screening of carbohydrate depolymerizing enzymes. *Rapid Commun. Mass. Spectrom.* **25**, 2059–2070 (2011).
64. ZoBell, C. Studies on marine bacteria I The cultural requirements of heterotrophic aerobes. *J. Mar. Res.* **4**, 42–75 (1941).
65. Thomas, F., Barbeyron, T. & Michel, G. Evaluation of reference genes for real-time quantitative PCR in the marine flavobacterium *Zobellia galactanivorans*. *J. Microbiol. Methods* **84**, 61–66 (2011).
66. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).



67. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
68. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
69. Toffano-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* **16**, 944–945 (2000).
71. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with 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**, e0157022 (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).

## Acknowledgments

We are grateful to Dr Diane Jouanneau for critical discussions on the development on this project and to Dr Lionel Cladière and Ms. Anaïs Naretto for technical assistance. We thank Jonas Collén for taking and sharing with us an image that we submitted for consideration as a featured image on the journal's homepage. We thank the local contact support on the beamlines ID-14-4 and Proxima2 at the ESRF (Grenoble, France) and SOLEIL (Paris, France) with special thanks to Gavin Fox for his help with data treatment. This work has benefited from the platforms and expertise of the High-throughput Sequencing Platform of I2BC (Paris, France). We thank the INRA MIGALE bioinformatics platform (<http://migale.jouy.inra.fr>) for providing computational resources, help and support. G.M., E.D. and C.M. acknowledge support from the Agence Nationale de la Recherche (ANR) with regard to the “Blue Enzymes” project (reference ANR-14-CE19-0020-01). G.M. and M.C. are also grateful to ANR for its support with regards to the investment expenditure program IDEALG (<http://www.idealg.ueb.eu/>), grant agreement No. ANR-10-BTBR-04). M.J.M. thanks the National Science Foundation for the grant MCB-1516990. K.A.S. thanks the Australian Research Council (FT100100291) and the Centre for Microscopy, Characterisation and Analysis at The University of Western Australia.

## 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 sulfatases and GH2 enzymes and assisted in writing the paper. M.J. characterized the SusD-like proteins. F.T., R.L., Y.T., T.B. and M.J.McB. performed the genetic experiments and the phenotyping of the *Zobellia* mutant strains. F.T. assisted in writing the paper. T.R., D. P.-P. and E.D. performed all the transcriptomics experiments and analyzes. T.R. assisted in writing the paper. M.S., A.C., B.V., C.M. and D.V. performed all the bioinformatics analyzes. M.S., A.C. and D.V. assisted in writing the paper. S.G. and M.J. purified and characterized ZGAL\_3155 (DauA), ZGAL\_3156 (DauB), ZGAL\_3154 (DauC), and ZGAL\_3153 (DauD) and assisted in writing the paper. S.G. was involved in the characterization of the GH2 enzymes. D.R. and H.R. performed the mass spectrometry experiments. G.C. produced and purified the fucellaran oligosaccharides. M.M. performed the site-directed mutagenesis of ZGAL\_3152. K.A.S. synthesized initial potential substrates for ZGAL\_3152 and assisted in writing the paper. M.J. crystallized ZGAL\_3156 and A.J. assisted with the crystallization experiments. M.C. collected crystallography data and assisted in writing the paper. G.M. designed research, determined the structure of ZGAL\_3156 and wrote the paper with E.F.-B.

## Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-01832-6.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017